UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Geraldo Henrique Menezes Ferreira

Condições de otimalidade em programação não linear

Juiz de Fora
2024

Geraldo Henrique Menezes Ferreira

Condições de otimalidade em programação não linear

Dissertação apresentada ao Programa de Pós-graduação em Matemática da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Matemática. Área de concentração: Matemática Aplicada

Orientador: Prof. Dr. Wilhelm Passarella Freire

Juiz de Fora

2024

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Ferreira, Geraldo Henrique.
Condições de otimalidade em programação não linear / Geraldo Henrique Menezes Ferreira. – 2024.
77 f. : il.

Orientador: Wilhelm Passarella Freire
Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-graduação em Matemática, 2024.

1. Otimização. 2. Condições de Otimalidade. 3. Multiplicadores de Lagrange. I. Freire, Wilhelm, orient. II. Título.

Geraldo Henrique Menezes Ferreira

Condições de Otimalidade em Programação Não Linear

Dissertação apresentada ao Programa de Pós-graduação em Matemática da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Matemática. Área de concentração: Matemática Aplicada

Aprovada em 06 de setembro de 2024.

BANCA EXAMINADORA

Prof. Dr. Wilhelm Passarella Freire - Orientador

Universidade Federal de Juiz de Fora

Prof. Dr. Hernando José Rocha Franco

Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais

Prof. Dr. Sandro Rodrigues Mazorche

Universidade Federal de Juiz de Fora

Juiz de Fora, 10/09/2024.

Documento assinado eletronicamente por **Sandro Rodrigues Mazorche**, **Professor(a)**, em 10/09/2024, às 10:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **HERNANDO JOSE ROCHA FRANCO**, **Usuário Externo**, em 10/09/2024, às 11:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Wilhelm Passarella Freire**, **Professor(a)**, em 11/09/2024, às 09:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1976003** e o código CRC **C7317BFE**.

Documento assinado digitalmente
GERALDO HENRIQUE MENEZES FERREIRA
Data: 19/11/2024 14:26:41-0300
Verifique em https://validar.iti.gov.br

AGRADECIMENTOS

Agradeço a Deus em primeiro lugar. Por tudo em minha vida.

À minha esposa e à minha família por todo o apoio e paciência.

Ao meu orientador Wilhelm, pela dedicação, paciência, conselhos e ensinamentos durante a construção deste trabalho.

Aos meus amigos, pelo suporte e companheirismo durante todo o curso.

A todos os professores que contribuiram para minha formação, à CAPES por fomentar o apoio da pesquisa e pelo apoio financeiro e à Universidade Federal de Juiz de Fora, pela oportunidade.

A todos, muito obrigado!

RESUMO

A otimização é uma área da Matemática cujo o objetivo é estudar problemas que consistem em maximizar ou minimizar funções. Um problema de otimização pode ser restrito ou irrestrito, linear ou não linear, convexo ou não convexo, diferenciável ou não diferenciável. Este trabalho tem como objetivo estudar as condições de otimalidade em programação não linear, para problemas restritos e irrestritos. Para problemas de otimização sem restrição uma condição necessária para que um ponto seja um candidato a minimizador da função objetivo é que este seja um ponto crítico, além disso é possível classificar a natureza desse ponto com a análise da matriz Hessiana de f. Para problemas com restrições, de igualdade ou desigualdade, as condições de otimalidade são apresentadas através da teoria dos Multiplicadores de Lagrange.

Palavras-chave: Otimização. Condições de Otimalidade. Multiplicadores de Lagrange.

ABSTRACT

Optimization is an area of Mathematics whose objective is to study problems that consist of maximizing or minimizing functions. An optimization problem can be restricted or unrestricted, linear or nonlinear, convex or nonconvex, differentiable or nondifferentiable. This work aims to study the optimality conditions in nonlinear programming, for restricted and unrestricted problems. For optimization problems without restriction, a necessary condition for a point to be a candidate to minimize the objective function is that it be a critical point. In addition, it is possible to classify the nature of this point with the analysis of the Hessian matrix of f. For problems with restrictions, of equality or inequality, the optimality conditions are presented through the theory of Lagrange Multipliers.

Keywords: Optimization. Optimality Conditions. Lagrange multipliers.

SUMÁRIO

1 INTRODUÇÃO . . . 7
2 DEFINIÇÕES GERAIS . . . 9
3 OTIMIZAÇÃO SEM RESTRIÇÕES . . . 14
3.1 FUNÇÕES DE UMA VARIÁVEL . . . 14
3.2 FUNÇÕES DE VÁRIAS VARIÁVEIS . . . 16
4 OTIMIZAÇÃO COM RESTRIÇÕES . . . 21
4.1 OTIMIZAÇÃO COM RESTRIÇÕES DE IGUALDADE . . . 21
4.1.1 Condição necessária de 1ª ordem . . . 22
4.1.2 Condição necessária e suficiente de 2ª ordem . . . 26
4.2 OTIMIZAÇÃO COM RESTRIÇÕES DE DESIGUALDADE . . . 38
4.2.1 Condição necessária de 1ª ordem . . . 40
4.2.2 Condição necessária e suficiente de 2ª ordem . . . 43
4.3 OTIMIZAÇÃO COM RESTRIÇÕES DE IGUALDADE E DESIGUALDADE 48
4.3.1 Condição necessária de 1ª ordem . . . 49
4.3.2 Condição necessária e suficiente de 2ª ordem . . . 51
5 CONDIÇÕES DE OTIMALIDADE APLICADAS EM ALGORITMOS . . . 54
5.1 MÉTODOS DE DIREÇÕES VIÁVEIS . . . 54
5.1.1 Método do gradiente e Método de Cauchy . . . 55
5.1.2 Método multidimensional de Newton . . . 56
5.1.3 Direções conjugadas e Método quasi-Newton . . . 57
5.1.4 Método DFP - Davidon, Fletcher e Powell . . . 58
5.2 MÉTODO DE ZOUTENDIJK . . . 59
5.3 PROGRAMAÇÃO LINEAR SUCESSIVA . . . 61
5.4 ALGORITMO PENLAB . . . 62
5.5 MÉTODO DE PONTOS INTERIORES DE NEWTON . . . 63
5.6 ALGORITMO DE DIREÇÕES VIÁVEIS E PONTOS INTERIORES PARA PROBLEMAS COM RESTRIÇÕES - FDIPA . . . 63
5.7 MÉTODO DE FUNÇÃO DE BARREIRA . . . 67
5.8 MÉTODO DE REESCALONAMENTO NÃO LINEAR BASEADO NA FUNÇÃO BARREIRA LOGARÍTMICA MODIFICADA . . . 69
5.9 FERRAMENTAS COMPUTACIONAIS . . . 71
5.9.1 FMINCON - MATLAB . . . 71
5.9.2 CVX - MATLAB . . . 72
5.9.3 SciPy - optimize . . . 72
6 CONCLUSÃO . . . 74
REFERÊNCIAS . . . 76

1 INTRODUÇÃO

Vários problemas vindos de áreas como Economia, Engenharia, Computação, entre outras, são formulados como problema de otimização. A Otimização é uma área da Matemática cujo objetivo é estudar problemas que consistem em maximizar ou minimizar funções onde as variáveis satisfazem determinadas restrições. Essencialmente, a Otimização busca encontrar os pontos de um conjunto $\Omega \in \mathbb{R}^n$, chamado conjunto viável, onde a função objetivo $f : \Omega \to \mathbb{R}$ assume seus valores de máximo ou mínimo, levando em consideração suas restrições, que são na forma de igualdades ou desigualdades.

Um problema de otimização pode ser modelado de acordo com as características das funções envolvidas, deste modo, os problemas podem ser irrestritos, quando a função objetivo não possui nenhuma restrição em seu conjunto viável, ou restritos, quando há restrição no conjunto viável, isto é, pontos fora dessa região não serão considerados. Além disso, o problema pode ser linear, quando função objetivo e as restrições são funções lineares. Esses problemas geralmente são resolvidos usando métodos como o Simplex ou Programação Linear (24). Não linear, quando função objetivo ou as restrições são não lineares exigindo métodos mais complexos para obtenção do valor ótimo, como algoritmos genéticos, métodos de gradiente, ou métodos de otimização global. O problema também pode ser convexo ou não convexo, diferenciável quando a função possui derivada em todos os pontos do seu domínio ou não diferenciável quando a função não possui derivada em alguns pontos do seu domínio. Neste caso, utilizam-se técnicas de Análise Convexa no lugar do Cálculo Diferencial (25).

O método multiplicadores de Lagrange fundamental na otimização matemática, geralmente utilizada para encontrar máximos e mínimos de uma função sujeita a restrições. De forma resumida, o método dos multiplicadores de Lagrange permite, de certo modo, transformar um problema de otimização com restrição em um problema irrestrito, que é mais simples de resolver. Com o auxílio dos multiplicadores, é possível apresentar condições de primeira ordem para problemas de otimização com restrições, esta condição, de certo modo, fornece os pontos críticos do problema, ou seja, os possíveis candidatos a máximo ou mínimo do problema proposto.

Neste trabalho vamos abordar as técnicas dos multiplicadores de Lagrange com a finalidade de encontrar a solução de um problema de otimização não linear e diferenciável no espaço $\mathbb{R}^n$, apresentando as condições de otimalidade para estes problema. Estudaremos separadamente o caso de um problema sem restrições, onde $\Omega = \mathbb{R}^n$, o caso com restrições de igualdade, restrições de desigualdade e por fim problemas com restrições de igualdade e desigualdade, conhecido como restrição mistas. Em seguida, apresentamos alguns algoritmos clássicos da literatura e softwares matemáticos que utilizam às condições de otimalidade. O objetivo principal desse trabalho é apresentar de forma simples e objetiva as condições de otimalidade para problemas de programação não linear com auxílio dos multiplicadores de Lagrange.

Este trabalho contém 6 capítulos, onde o capítulo 1 é a introdução. No capítulo 2 são apresentadas algumas definições e resultados matemáticos fundamentais para o entendimento de todo o texto. No capítulo 3, são abordados os problemas de otimização sem restrições, são apresentadas as definições de valor ótimo, mínimo local (global), ponto crítico, entre outros, fundamentais em todo o estudo. O capítulo 4 apresenta os problemas de otimização com restrições, onde é introduzido o conceito dos multiplicadores de Lagrange e as técnicas para encontrar e classificar os candidatos à solução do problema, com o auxílio da função Lagrangeana Clássica. Os algoritmos e softwares matemáticos que possuem relação com as condições de otimalidade estudadas são apresentados no capítulo 5.

Por fim, apresentamos as considerações finais no capítulo 6. Os resultados teóricos apresentados no trabalho são resultados clássicos da teoria de otimização, suas demonstrações podem ser encontradas nos textos referência.

2 DEFINIÇÕES GERAIS

Neste capítulo listaremos algumas definições e resultados úteis para os demais capítulos.

Definição 2.0.1. *Sejam $v_1, ..., v_n \in V$ vetores não nulos, e os escalares $\alpha_1, ..., \alpha_n \in \mathbb{R}$, não simultaneamente nulos. Se*

$$\alpha_1 v_1 + ... + \alpha_n v_n = 0,$$

*dizemos que os vetores $v_1, ..., v_n \in V$ são **Linearmente Dependentes (LD)**, por outro lado, se*

$$\alpha_1 v_1 + ... + \alpha_n v_n = 0 \Longrightarrow \alpha_1 = ... = \alpha_n = 0,$$

*então os vetores $v_1, ..., v_n \in V$ são ditos **Linearmente Independentes (LI)***

Definição 2.0.2. *Seja A uma matriz $m \times n$, definimos o seu posto como sendo o número de linhas não nulas de sua forma escalonada.*

Definição 2.0.3 (Núcleo e Imagem de uma Matriz)**.** *Seja A uma matriz $m \times n$. O conjunto*

$$N(A) = \{v \in \mathbb{R}^n \mid Av = 0\},$$

é chamado Núcleo de A. O Núcleo de A é um subespaço de $\mathbb{R}^n$.

O conjunto

$$Im(A) = \{y \in \mathbb{R}^m \mid Av = y,\ v \in \mathbb{R}^n\},$$

é chamado Imagem de A. O conjunto Imagem de A é um subespaço de $\mathbb{R}^m$.

Definição 2.0.4. *Seja A uma matriz de ordem $n \times n$ simétrica. A é dita*

1. *Positiva semidefinida se $x \cdot Ax \geq 0$, $\forall x \in \mathbb{R}^n$;*
2. *Positiva definida se $x \cdot Ax > 0$, $\forall x \in \mathbb{R}^n - \{0\}$;*
3. *Negativa semidefinida se $x \cdot Ax \leq 0$, $\forall x \in \mathbb{R}^n$;*
4. *Negativa definida se $x \cdot Ax < 0$, $\forall x \in \mathbb{R}^n - \{0\}$;*
5. *Indefinida se $x \cdot Ax > 0$ para algum $x \in \mathbb{R}^n$ e $y \cdot Ay < 0$ para algum $y \in \mathbb{R}^n$.*

Teorema 2.0.1. *Seja A uma matriz simétrica, então:*

1. *A é positiva definida* $\iff$ *os autovalores de A são positivos;*

2. *A é positiva semidefinida* $\iff$ *os autovalores de A são não negativos;*

3. *A é negativa definida* $\iff$ *os autovalores de A são negativos;*

4. *A é negativa semidefinida* $\iff$ *os autovalores de A são não positivos;*

5. *A é indefinida* $\iff$ *A tem autovalores positivos e negativos.*

Definição 2.0.5. *Uma **função** $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ é uma regra que associa a cada ponto $x \in X$ um único ponto*

$$y = f(x) = (f_1(x), f_2(x), ..., f_m(x)).$$

*As funções $f_1, f_2, ..., f_m : X \subset \mathbb{R}^n \to \mathbb{R}$ chamam-se as **funções coordenadas** de f.*

Definição 2.0.6. *Dizemos que uma função $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ é contínua no ponto $a \in X$, quando para cada $\epsilon > 0$ dado, pode-se obter um $\delta > 0$ tal que:*

$$x \in X,\ ||x - a|| < \delta \Longrightarrow ||f(x) - f(a)|| < \epsilon.$$

Definição 2.0.7. *Sejam $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ e $A \subset \mathbb{R}^n$ um conjunto aberto. A **derivada direcional** de f no ponto $a \in A$ na direção de um vetor $v \neq 0$ é definida por*

$$\frac{\partial f}{\partial v}(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}.$$

Se $v = e_i = (0, 0, ..., 1, ..., 0)$ temos que

$$\frac{\partial f}{\partial e_i}(a) = \frac{\partial f}{\partial x_i}(a) = \lim_{t \to 0} \frac{f(a + te_i) - f(a)}{t},$$

que é da derivada parcial de f em relação a x_i no ponto a.

Definição 2.0.8. *Dizemos que uma função $f : X \subset \mathbb{R}^n \to \mathbb{R}$ é de classe C^n em X e escrevemos $f \in C^n(X)$ se f possui todas as derivadas parciais de ordem n contínuas em X.*

Escrevemos $f \in C^0(X)$ para dizer que f é contínua em X.

Definição 2.0.9. *Sejam $f : A \to \mathbb{R}^m$ e $A \subset \mathbb{R}^n$ um conjunto aberto. A aplicação f é diferenciável no ponto $a \in A$ se existe uma transformação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ tal que*

$$f(a + v) = f(a) + Tv + r(v) \text{ com } \lim_{v \to 0} = \frac{r(v)}{||v||} = 0.$$

A transformação linear T é a derivada de f no ponto a e possui, em relação as bases canônicas de $\mathbb{R}^n$ e $\mathbb{R}^m$, uma matriz $m \times n$ chamada matriz jacobiana de f em a.

Definição 2.0.10. *A **matriz Jacobiana** de f em a é dada por*

$$Jf(a) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \dfrac{\partial f_1}{\partial x_2}(a) & \dots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \dfrac{\partial f_2}{\partial x_1}(a) & \dfrac{\partial f_2}{\partial x_2}(a) & \dots & \dfrac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(a) & \dfrac{\partial f_m}{\partial x_2}(a) & \dots & \dfrac{\partial f_m}{\partial x_n}(a) \end{bmatrix}.$$

Definição 2.0.11. *No caso em que $m = 1$, a matriz jacobiana de $f : A \subset \mathbb{R}^n \to \mathbb{R}$ em a se reduz ao vetor $1 \times n$ denominado **vetor gradiente** de f em a*

$$\nabla f(a) = \left[\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\right] = Jf(a).$$

Em relação ao vetor gradiente podemos destacar algumas propriedades.

1. o vetor gradiente aponta na direção de maior crescimento da função localmente.
2. entre todas as direções possíveis, o vetor gradiente de f, aponta na direção de maior taxa de crescimento de f;
3. o vetor gradiente de f no ponto a é ortogonal à superfície de nível que passa por a.

Definição 2.0.12. *Seja $f : A \subset \mathbb{R}^n \to \mathbb{R}$ uma aplicação duas vezes diferenciável em A aberto. Então a derivada $f''(a) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ com respeito à base canônica pode ser representada por uma matriz*

$$Hf(a) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \dots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \dfrac{\partial^2 f}{\partial x_2 \partial x_2}(a) & \dots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(a) & \dfrac{\partial^2 f}{\partial x_n \partial x_2}(a) & \dots & \dfrac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{bmatrix}.$$

*A matriz acima é chamada **matriz Hessiana** de f em a.*

Teorema 2.0.2 (Regra da Cadeia)**.** *Sejam $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^n$ abertos e $f : X \to \mathbb{R}^n$, $g : Y \to \mathbb{R}^p$ diferenciáveis nos pontos $a \in X$, $b = f(a) \in Y$, com $f(X) \subset Y$.*

Então $g \circ f : X \to \mathbb{R}^p$ é diferenciável no ponto a e

$$J(g \circ f)(a) = Jg(b) \cdot Jf(a)$$

Resumidamente: a derivada da aplicação composta é a composta das derivadas.

Definição 2.0.13 (Fórmula de Taylor com resto de Lagrange)**.** *Sejam $f : A \subset \mathbb{R}^n$ um conjunto aberto e $f : A \to \mathbb{R}$ de classe C^2 em A. Dados $a, x \in A$ existe z no segmento $[a, x] = \{z \in \mathbb{R}^n \mid z = a + t(x - a),\ 0 \leq t \leq 1\}$, tal que*

$$f(x) = f(a) + \nabla f(a) \cdot (x - a) + \frac{1}{2}(x - a) \cdot Hf(z)(x - a).$$

Teorema 2.0.3 (Teorema da Função Inversa)**.** *Sejam $f : A \to \mathbb{R}^n$ uma função de classe C^1 no aberto $A \subset \mathbb{R}^n$ e $a \in A$ tal que $det Jf(a) \neq 0$. Então existem abertos $U, V \subset \mathbb{R}^n$ tais que*

$$a \in U,\ f(a) \in V,\ U \subset A,\ f(U) = V \text{ e } f : U \to V \text{ é injetora.}$$

A função $f^{-1} : V \to U$, localmente inversa de f é também de classe C^1 e

$$Jf^{-1}(f(a))Jf(a) = I.$$

Teorema 2.0.4 (Teorema da Função Implícita)**.** *Sejam $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ uma função de classe C^1 e $(a, b) = (a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_m) \in \mathbb{R}^{n+m}$ tal que $F(a, b) = 0$.*

Suponhamos que $det J_y F(a, b) \neq 0$ onde $J_y F(a, b)$ é a submatriz da $JF(x, y)$ cujas colunas são correspondentes às derivadas parciais em relação a $y_1, y_2, ..., y_m$.

Existem um conjunto aberto $U \in \mathbb{R}^n$ e uma função $f : U \to \mathbb{R}^m$ de classe C^1 tais que

1. *$a \in U$ e $f(a) = b$;*
2. *$y = f(x)$ e $F(x, y) = F(x, f(x)) = 0$, $\forall x \in U$;*
3. *$Jf(x) = -\left[J_y F(x, f(x))\right]^{-1} J_x F(x, f(x))$, $\forall x \in U$.*

Definição 2.0.14. *Um vetor $d \in \mathbb{R}^n$ é chamado direção de descida de f no ponto $\bar{x}$, se existe um $\delta > 0$ tal que*

$$f(\bar{x} + \lambda d) < f(\bar{x}), \quad \text{para todo} \quad \lambda \in (0, \delta].$$

Se f é diferenciável, a condição

$$\nabla f(\bar{x})^t d < 0,$$

é suficiente para que d seja uma direção de descida de f em $\bar{x}$.

Definição 2.0.15. *Um conjunto $X \subset \mathbb{R}^n$ é dito* ***convexo*** *se, dados dois pontos quaisquer $x, y \in X$, o segmento*

$$(1-t)x + ty \in X, \quad \forall t \in [0,1].$$

Definição 2.0.16. *Seja $f : X \subset \mathbb{R}^n \to \mathbb{R}$ onde X é convexo. Dizemos que f é* ***convexa*** *se,*

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y), \quad \forall x, y \in X \quad e \quad t \in [0,1].$$

Quando a desigualdade é estrita dizemos que f é ***estritamente convexa****. Se $-f$ é convexa, dizemos que f é côncava.*

Teorema 2.0.5 (Weierstrass)**.** *Seja $f : X \subset \mathbb{R}^n \to \mathbb{R}$ contínua e X compacto. Então existem $x_0, x_1 \in X$ tais que*

$$f(x_0) \leq f(x) \leq f(x_1), \quad \forall x \in X.$$

3 OTIMIZAÇÃO SEM RESTRIÇÕES

Neste capítulo vamos abordar problemas de otimização que envolvem funções de uma e também de várias variáveis. Estudaremos problemas irrestritos e os mecanismos que nos fornecem condições para resolver estes problemas.

3.1 FUNÇÕES DE UMA VARIÁVEL

Nesta seção vamos tratar o problema de otimização sem restrições que envolve funções de uma variável. Consideraremos o seguinte problema de otimização

$$(P)\begin{cases} \text{Minimizar} \quad f(x) \\ x \in \mathbb{R} \end{cases}$$

Queremos encontrar, caso existam, números reais para os quais o valor da função objetivo seja o menor possível quando comparado com os valores calculados em vizinhanças desses pontos ou mesmo nos demais pontos de $\mathbb{R}$.

Definição 3.1.1. *O valor ótimo de (P) é o real $\bar{x}$ tal que*

$$\bar{x} = inf\{f(x) \mid x \in \mathbb{R}\}$$

quando o ínfimo existe.

Definição 3.1.2. *Seja uma função $f : \mathbb{R} \to \mathbb{R}$. Diremos que $\bar{x}$ é mínimo **local** de f, se existe $\epsilon > 0$ tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in (\bar{x} - \epsilon, \bar{x} + \epsilon).$$

*Diremos que $\bar{x}$ é mínimo **global** de f, se*

$$f(\bar{x}) \leq f(x), \quad \forall x \in \mathbb{R}.$$

*Se as igualdades anteriores tornam-se estritas, dizemos que $\bar{x}$ é mínimo local (global) **estrito**.*

Analogamente, as definições de máximo são obtidas invertendo os sinais das desigualdades acima.

Teorema 3.1.1. *Seja $f : X \subset \mathbb{R}^n \to \mathbb{R}$ uma função convexa, então cada ponto de mínimo local de f é também mínimo global.*

Exemplo 3.1.1. *(28) A função $f(x) = cos(x)$, com $x \in [0, 2\pi]$ possui um valor mínimo em $\bar{x} = \pi$, pois*

$$cos(\pi) \leq cos(x), \quad \forall x \in [0, 2\pi].$$

Definição 3.1.3. *Seja $f : \mathbb{R} \to \mathbb{R}$. Um ponto $\bar{x}$ é chamado* ***ponto crítico*** *de f, se*

$$f'(\bar{x}) = 0.$$

O próximo teorema fornece uma condição necessária para que $\bar{x}$ seja um mínimo local de f.

Teorema 3.1.2 (Condição necessária de 1ª ordem)**.** *Uma condição necessária para que $\bar{x}$ seja um mínimo local da função $f : \mathbb{R} \to \mathbb{R}$, $f \in C^1(\mathbb{R})$, é*

$$f'(\bar{x}) = 0.$$

A condição apresentada no teorema não é suficiente, ou seja, se $f'(\bar{x}) = 0$, a função f pode ter ou não um extremo local no ponto $\bar{x}$. De fato, considere a função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^3$, o ponto crítico de f ocorre em $\bar{x} = 0$, porém a função não possui valores mínimos nem máximos.

Além disso, quando $f'(\bar{x})$ não existe, $f(x)$ pode ter ou não um extremo local em $\bar{x}$. É importante notar que, se $\bar{x}$ é um minimizador de f, então $\bar{x}$ é um ponto crítico de f.

Para funções de classe $C^2(\mathbb{R})$, é possível enunciar **Condiçãos necessárias e suficiente de 2ª ordem**.

Teorema 3.1.3 (Condição necessária de 2ª ordem)**.** *Uma condição necessária para que $\bar{x}$ seja um mínimo local da função $f : \mathbb{R} \to \mathbb{R}$, $f \in C^2(\mathbb{R})$ onde $f'(\bar{x}) = 0$, ou seja, $\bar{x}$ é um ponto crítico de f, é:*

$$f''(\bar{x}) \geq 0.$$

Teorema 3.1.4 (Condições suficientes de 2ª ordem)**.** *Sejam $f : \mathbb{R} \to \mathbb{R}$, $f \in C^2(\mathbb{R})$ e $\bar{x}$ um ponto crítico de f, isto é, $f'(\bar{x}) = 0$, então:*

i se $f''(\bar{x}) > 0$, então $\bar{x}$ é um mínimo local de f;

ii se $f''(\bar{x}) = 0$, nada se pode afirmar.

Vimos que, quando a condição $f'(\bar{x}) = 0$ é satisfeita, o sinal de $f''(\bar{x})$ é suficiente para estabelecer se o ponto crítico do problema é um mínimo local. Vale ressaltar que a condição de primeira ordem é necessária para que se possa aplicar os demais testes.

Exemplo 3.1.2. *(22) Considere o seguinte problema de otimização sem restrições:*

$$\textit{Minimizar}\, f(x) = (x^2 - 1)^3.$$

Vamos calcular a derivada de f

$$f'(x) = 6x(x^2 - 1)^2,$$

os possíveis minimizadores do problema são os pontos $\bar{x}$ tais que $f'(\bar{x}) = 0$, logo

$$6x(x^2 - 1)^2 = 0.$$

Resolvendo a equação acima obtemos as seguintes soluções $\bar{x} = 0$, $\bar{x} = 1$ e $\bar{x} = -1$ que são os candidatos a solução do problema. Vamos usar o resultado apresentado no Teorema 3.1.4 para classificar os valores obtidos. A derivada segunda de f é dada por:

$$f''(x) = 6(x^2 - 1)^2 + 24x^2(x^2 - 1) = (x^2 - 1)(30x^2 - 6).$$

Substituindo os valores de $\bar{x}$ em f'' obtemos

$$f''(0) = 6,$$

$$f''(1) = 0,$$

$$f''(-1) = 0.$$

Portanto o ponto $\bar{x} = 0$ é um minimizador do problema. Além disso, podemos observar na Figura 3.1.2 que os pontos $\bar{x} = 1$ e $\bar{x} = -1$ não são pontos de mínimo (nem de máximo).

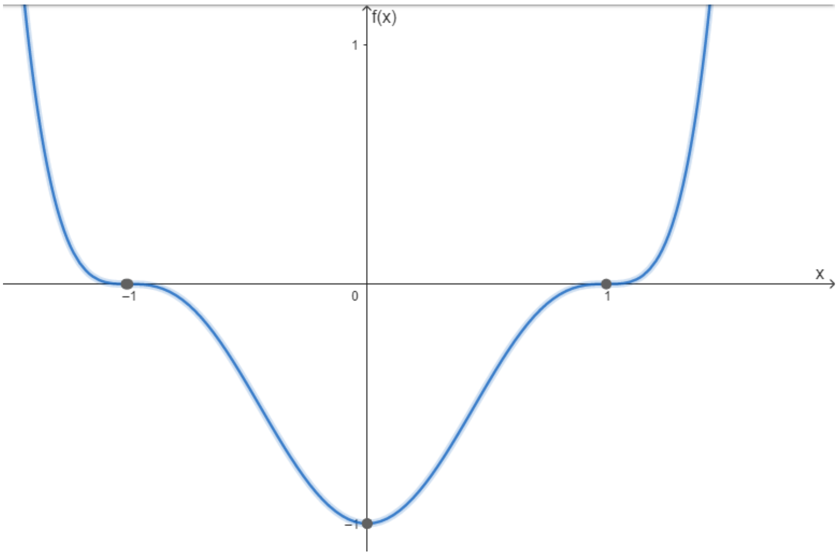
Quando uma função é convexa, as condições de otimalidade tornam-se mais simples, pois as condições de 2ª ordem nos fornecem a convexidade local da função, além disso, tratando-se de funções convexas, um mínimo local será também um mínimo global.

3.2 FUNÇÕES DE VÁRIAS VARIÁVEIS

Nesta seção vamos tratar o problema de otimização (P) sem restrições que envolvem funções de várias variáveis. Neste caso a função objetivo está definida da seguinte forma

$$f : \mathbb{R}^n \to \mathbb{R}.$$

Assim como vimos na seção anterior, queremos encontrar, caso exista, um ponto $\bar{x} \in \mathbb{R}^n$ no qual o valor da função objetivo seja o menor possível quando comparado com os valores calculados nos demais pontos de $\mathbb{R}^n$.



– Figura 3.1 - Interpretação Geométrica do Exemplo (Fonte: Criado pelo Autor)

Definição 3.2.1. *Seja uma função $f : \mathbb{R}^n \to \mathbb{R}$. Diremos que $\bar{x}$ é mínimo* ***local*** *de f, se existe $\epsilon > 0$ tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in B(\bar{x}, \epsilon).$$

Diremos que $\bar{x}$ é mínimo ***global*** *de f, se*

$$f(\bar{x}) \leq f(x), \quad \forall x \in \mathbb{R}^n.$$

Se as igualdades anteriores tornam-se estritas, dizemos que $\bar{x}$ é mínimo local (global) ***estrito****.*

Analogamente, as definições de máximo são obtidas invertendo os sinais das desigualdades acima.

Definição 3.2.2. *Seja $f : \mathbb{R}^n \to \mathbb{R}$. Um ponto $\bar{x}$ é chamado* ***ponto crítico*** *de f, se*

$$\nabla f(\bar{x}) = 0.$$

Definição 3.2.3. *Seja uma função $f : \mathbb{R}^n \to \mathbb{R}$, dizemos que $\bar{x} \in \mathbb{R}^n$ é um* ***ponto de sela*** *de f se existem vetores $u, v \in \mathbb{R}^n$ tais que*

$$f(\bar{x} + tu) > f(\bar{x}) \quad e \quad f(\bar{x} + tv) < f(\bar{x}),\ t \in (\alpha, \beta) \subset \mathbb{R}.$$

O resultado apresentado a seguir nos fornece uma condição necessária para que $\bar{x}$ seja um minimizador de $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$.

Teorema 3.2.1 (Condição necessária de 1ª ordem)**.** *Uma condição necessária para que $\bar{x}$ seja um mínimo local da função $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$, é*

$$\nabla f(\bar{x}) = 0.$$

Assim como no caso de funções de uma variável, o teorema garante que, se f tem um extremo local em $\bar{x}$, este é um ponto crítico. No entanto, nem todos os pontos críticos dão origem a um extremo da função objetivo.

Exemplo 3.2.1. *Considere o seguinte problema de otimização*

$$\textit{Minimizar}\, f : \mathbb{R}^2 \to \mathbb{R}, \textit{onde}\, f(x, y) = x^2 - y^2.$$

Note que $\nabla f(x, y) = (2x, -2y)$, e o único ponto crítico de f é o ponto $(0, 0)$. Observe que este não é um ponto mínimo local de f. De fato, sejam $u = (1, 0)$ e $v = (0, 1)$. assim

$$f((0,0) + t \cdot u) = f((0,0) + (t,0) = f(t,0) = t^2 > 0 = f(0,0),$$
$$f((0,0) + t \cdot v) = f((0,0) + (0,t) = f(0,t) = -t^2 < 0 = f(0,0).$$

Portanto o ponto $(0, 0)$ é um ponto de sela.

O exemplo anterior ilustra o fato que o ponto crítico da função não é nem máximo e nem mínimo local. Através da análise da Hessiana de f, é possível indicar a natureza dos pontos críticos da função, e assim apresentar **Condições necessárias e suficientes de 2ª ordem** para funções de várias variáveis.

Teorema 3.2.2 (Condição necessária de 2ª ordem)**.** *Uma condição necessária para que $\bar{x}$ seja um mínimo local da função $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(\mathbb{R}^n)$ onde $\nabla f(\bar{x}) = 0$, é que:*

$$Hf(\bar{x}) \textit{ seja positiva semidefinida.}$$

Teorema 3.2.3 (Condições suficientes de 2ª ordem)**.** *Sejam $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(\mathbb{R}^n)$ e $\bar{x} \in \mathbb{R}^n$ um ponto crítico de f, isto é $\nabla f(\bar{x}) = 0$, então:*

i se $Hf(\bar{x})$ é positiva definida, então $\bar{x}$ é um mínimo local estrito de f;

ii se $Hf(\bar{x})$ é semidefinida, nada se pode afirmar.

Teorema 3.2.4. *Sejam $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(\mathbb{R}^n)$ e $\bar{x} \in \mathbb{R}^n$ um ponto crítico de f, isto é $\nabla f(\bar{x}) = 0$. Se $Hf(\bar{x})$ é indefinida, então $\bar{x}$ é um ponto de sela de f.*

Exemplo 3.2.2. *(28) Considere o seguinte problema de otimização*

$$\textit{Minimizar}\, f(x,y) = x^4 + y^4 - 4xy + 1.$$

Os pontos críticos de f são dados pelo sistema de equações

$$\nabla f(x,y) = (4x^3 - 4y, 4y^3 - 4x) = (0,0).$$

Resolvendo o sistema acima, obtemos os seguintes pontos como candidatos a minimizadores do problema, $(0,0), (1,1)$ e $(-1,-1)$. Vamos analisar a natureza dos pontos utilizando a Hessiana.

$$Hf(x,y) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x \partial x}(x,y) & \dfrac{\partial^2 f}{\partial x \partial y}(x,y) \\ \dfrac{\partial^2 f}{\partial y \partial x}(x,y) & \dfrac{\partial^2 f}{\partial y \partial y}(x,y) \end{bmatrix} = \begin{bmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{bmatrix}.$$

Para o ponto crítico $(0,0)$ temos

$$Hf(0,0) = \begin{bmatrix} 0 & -4 \\ -4 & 0 \end{bmatrix}$$

Os autovalores da matriz $Hf(0,0)$ são dados por $\lambda_1 = 4$ e $\lambda_2 = -4$, e portanto a matriz é indefinida e pelo Teorema 3.2.4, $(0,0)$ é um ponto de sela. Para os pontos críticos $(1,1)$ e $(-1,-1)$ temos

$$Hf(1,1) = Hf(-1,-1) = \begin{bmatrix} 12 & -4 \\ -4 & 12 \end{bmatrix}$$

Os autovalores da matriz são dados por $\lambda_1 = 8$ e $\lambda_2 = 16$ e, portanto, a matriz é positiva definida. Logo, de acordo com a Condição suficiente de segunda ordem, os pontos $(1,1)$ e $(-1,-1)$ são pontos de mínimo do problema apresentado.

A Hessiana da função objetivo nos permite realizar diversas análises, e assim, encontrar outras ferramentas que possibilitam solucionar o problema de otimização. Considere os seguintes resultados.

Definição 3.2.4. *Seja A uma matriz simétrica de ordem n. Seja Δ_p o determinante da submatriz de A formada pelas p primeiras linhas de A e p primeiras colunas de A, com $1 \leq p \leq n$. Δ_p é chamado o menor principal de ordem p de A.*

Teorema 3.2.5. *Sejam A uma matriz simétrica de ordem n e Δ_p o menor principal de ordem p de, então:*

(1) A é positiva definida $\Longleftrightarrow \Delta_p > 0,\ p = 1, 2, ..., n$

(2) A é negativa definida $\Longleftrightarrow (-1)^p \Delta_p > 0,\ p = 1, 2, ..., n$

(3) A é uma matriz semidefinida positiva $\Longleftrightarrow \Delta_p > 0,\ p = 1, 2, ..., n-1,$ *e* $\Delta_n = 0$*;*

(4) A é uma matriz semidefinida negativa $\Longleftrightarrow (-1)^p \Delta_p > 0, p = 1, 2, ..., n-1,$ *e* $\Delta_n = 0.$

Exemplo 3.2.3. *(7) Considere o seguinte problema de otimização*

$$\textit{Minimizar}\, f(x, y, z) = 2x^2 + xy + 4y^2 + xz + z^2 + 2.$$

Vamos calcular as condições de primeira ordem, isto é, $\nabla f(x, y, z) = (0, 0, 0)$*, onde*

$$\nabla f(x, y, z) = (4x + y + z, x + 8y, x + 2z).$$

Resolvendo a equação acima obtemos como solução única e possível minimizador do problema de otimização o ponto $(0, 0, 0)$*. Como a condição de primeira ordem é apenas necessária, vamos aplicar o Teorema 3.2.2. A Hessiana de f é dada por*

$$Hf(x, y, z) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x \partial x}(x, y, z) & \dfrac{\partial^2 f}{\partial x \partial y}(x, y, z) & \dfrac{\partial^2 f}{\partial x \partial z}(x, y, z) \\ \dfrac{\partial^2 f}{\partial y \partial x}(x, y, z) & \dfrac{\partial^2 f}{\partial y \partial y}(x, y, z) & \dfrac{\partial^2 f}{\partial y \partial z}(x, y, z) \\ \dfrac{\partial^2 f}{\partial z \partial x}(x, y, z) & \dfrac{\partial^2 f}{\partial z \partial y}(x, y, z) & \dfrac{\partial^2 f}{\partial z \partial z}(x, y, z) \end{bmatrix} = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 8 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

Os menores principais da matriz são

$$\Delta_1 = 4, \quad \Delta_2 = 31, \quad \Delta_3 = 54.$$

Portanto, pelo Teorema 3.2.5, podemos concluir que a Hessiana de f é positiva definida, e por fim $(0, 0, 0)$ *é um minimizador da função.*

Podemos resumir os resultados apresentados nesse capítulo da seguinte maneira: dada uma função objetivo f, a condição necessária para que um ponto $\bar{x}$ seja um candidato a mínimo do problema de otimização (P) é $\nabla f(\bar{x}) = 0$, no qual chamamos de condições de primeira ordem. Sendo atendida esta condição, para que o ponto seja um ponto de mínimo do problema é suficiente que $\Delta_p > 0$, com $p = 1, 2, ..., n$, onde Δ_p são os menores principais da Hessiana de f.

4 OTIMIZAÇÃO COM RESTRIÇÕES

No capítulo anterior, estudamos as condições necessárias e suficientes para um problema de otimização sem restrições com uma função objetivo de uma variável, onde a região viável Ω era $\mathbb{R}$ e para o caso de uma função de várias variáveis onde Ω era $\mathbb{R}^n$. O fato da função objetivo possuir ou não extremos depende, é claro, da expressão de f, porém a região viável Ω possui influência direta nos resultados.

As principais referências deste capítulo são (1, 2, 3, 5, 27).

4.1 OTIMIZAÇÃO COM RESTRIÇÕES DE IGUALDADE

Veremos agora o caso onde a região viável é constituída de restrições de igualdade. Para isso, considere o seguinte problema de otimização

$$(P)\begin{cases}\text{Minimizar} \quad f(x)\\ \text{sujeito a} \quad x \in \Omega = \{x \in \mathbb{R}^n \mid h_i(x) = 0\}\\ \text{com} \quad i = 1, 2, ...m.\end{cases} \tag{4.1}$$

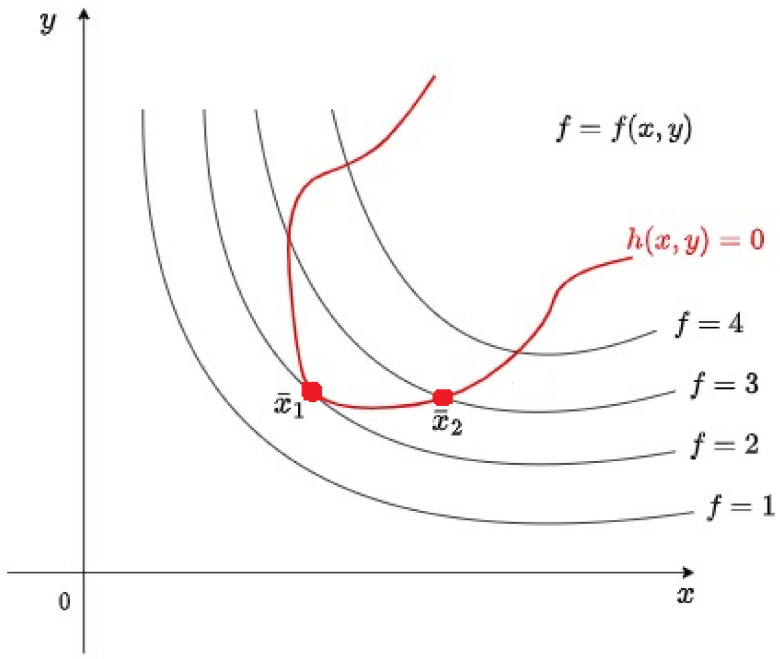
Onde $f : \mathbb{R}^n \to \mathbb{R}$ e $h : \mathbb{R}^n \to \mathbb{R}^m$, $h(x) = (h_1(x), h_2(x), ..., h_m(x))$, são no mínimo de classe $C^1(\mathbb{R}^n)$.

Para ilustrar, vamos começar com um problema de otimização de duas variáveis, ou seja, encontrar os extremos de f quando o ponto (x, y) pertencer à curva de nível $h(x, y) = 0$, a Figura 4.1 mostra o conjunto restrição Ω com diversas curvas de nível de $f(x, y) = c$, com $c = 1, 2, 3, 4$. Para minimizar f sujeita a $h(x, y) = 0$ precisamos determinar o menor valor de c, tal que a curva de nível $f(x, y) = c$ intercepte não transversalmente $h(x, y) = 0$. A Figura 4.1 mostra que isso ocorre quando essas curvas têm uma reta tangente comum.

Podemos observar pela Figura 4.1 que o ponto $\bar{x}_2$ não pode ser um mínimo local de f, pois existem curvas de nível que interceptam a região viável arbitrariamente próximo de $\bar{x}_2$, tanto pela direita quanto pela esquerda. Portanto, existem pontos à direita de $\bar{x}_2$, na região viável, onde o valor da função objetivo é maior que no ponto $\bar{x}_2$, analogamente, existem pontos à esquerda de $\bar{x}_2$ onde o valor da função objetivo é menor que no ponto $\bar{x}_2$.

Por outro lado, observando a figura, podemos notar que a curva de nível $f(x, y) = 2$ que passa pelo ponto $\bar{x}_1$ é tangente à curva de nível h, além disso, as demais curvas de nível de f que tocam a região viável estão acima de $c = 2$, isto é, o valor de f no ponto $\bar{x}_1$ será sempre menor que os valores da função avaliados nos demais pontos de Ω. Logo, concluímos que $\bar{x}_1$ é um mínimo de f restrito a h.

De forma resumida, temos que um ponto $\bar{x}$ é **extremo** de f sujeito à restrição h, então a curva de nível de f que passa por $\bar{x}$ é tangente à curva de nível $h = 0$. Além



– Figura 4.1 - Interpretação Geométrica do Problema com Restrições

disso, os vetores $\nabla f(\bar{x})$ e $\nabla h(\bar{x})$ devem ser paralelos em $\bar{x}$, de fato, se as curvas de nível são tangentes em $\bar{x}$, as retas normais ao ponto $\bar{x}$ devem ser as mesmas. Desse modo, $\nabla h(\bar{x})$ deve ser perpendicular a curva de nível $h = 0$, analogamente, $\nabla f(\bar{x})$ também é perpendicular a curva de nível $h = 0$. Portanto, existe um número real λ tal que

$$\nabla f(\bar{x}) = -\lambda \cdot \nabla h(\bar{x}). \tag{4.2}$$

O número real λ é denominado multiplicador de Lagrange. A seguir apresentamos uma **condição necessária de 1ª ordem** para o problema com restrições de igualdade, recorrendo a esses multiplicadores.

4.1.1 Condição necessária de 1ª ordem

A ideia central deste método é transformar o problema de otimização (4.1) restrito em um problema de otimização sem restrições, e assim, aplicar os resultados já vistos no capítulo anterior.

Definição 4.1.1. *Sejam uma superfície $\Omega \in \mathbb{R}^n$ e um ponto $x \in \Omega$. O plano tangente T à superfície Ω é o conjunto de todos os vetores tangentes em x às curvas diferenciáveis que passam por x.*

Definição 4.1.2. *Um ponto $\bar{x} \in \Omega$ que satisfaz a restrição $h(x) = 0$ é dito ser um ponto regular da restrição se os vetores gradientes*

$$\nabla h_1(\bar{x}), \nabla h_2(\bar{x}), \cdots, \nabla h_m(\bar{x}),$$

são linearmente independentes.

Teorema 4.1.1. *Se $\bar{x} \in \Omega$ é regular, então o plano tangente a Ω em $\bar{x}$ é dado por*

$$T_{\bar{x}} = N(\nabla h(\bar{x})) = \{v \in \mathbb{R}^n \mid \nabla h(\bar{x})v = 0\} \tag{4.3}$$

Teorema 4.1.2 (Condição necessária de 1ª ordem)**.** *Seja $\bar{x} \in \Omega$ um mínimo local de f sujeita a restrição $h(x) = 0$. Suponha que $\bar{x}$ seja um ponto regular das restrições. Então existe m multiplicadores de Lagrange $\lambda = (\lambda_1, \lambda_2, \cdots \lambda_m)$, tais que:*

$$\begin{cases} \nabla f(\bar{x}) + \lambda \cdot \nabla h(\bar{x}) = 0 \\ \\ h(\bar{x}) = 0. \end{cases} \tag{4.4}$$

Vale ressaltar que os multiplicadores de Lagrange $\lambda = (\lambda_1, ..., \lambda_m)$ apresentados no Teorema são números reais, e m é a quantidade de restrições associadas ao problema de otimização.

Para minimizar uma função em um problema sem restrições, como visto na seção anterior, o primeiro passo é encontrar seus pontos críticos. A introdução do multiplicador de Lagrange, em um certo sentido, transforma o problema de otimização da função f com n variáveis e com m restrições em um problema com $n + m$ variáveis e irrestrito.

Exemplo 4.1.1. *(5) Considere o seguinte problema de otimização*

$$\begin{cases} \textit{minimizar} \quad f(x, y) = -x^2 y \\ \textit{s. a} \quad (x, y) \in \Omega = \{(x, y) \in \mathbb{R}^2 \mid h(x, y) = 2x^2 + y^2 = 3\}. \end{cases}$$

Observe que f e h são de classe C^1, o Teorema 4.1.2 exige que o ponto $(\bar{x}, \bar{y})$ seja regular, note que

$$\nabla h(x, y) = (4x, 2y),$$

o que implica $\nabla h(x, y) = (4x, 2y) = (0, 0) \iff (x, y) = (0, 0)$. Porém o ponto $(0, 0)$ não pertence a região viável, visto que $h(0, 0) = 0 < 3$. Deste modo, para todos os pontos da região viável temos $\nabla h(\bar{x}, \bar{y}) \neq (0, 0)$. O sistema

$$\begin{cases} \nabla f(\bar{x}) + \lambda \cdot \nabla h(\bar{x}) = 0 \\ \\ h(\bar{x}) = 0, \end{cases}$$

pode ser escrito como

$$\begin{cases} \dfrac{\partial f}{\partial x}(\bar{x},\bar{y}) + \lambda\dfrac{\partial h}{\partial x}(\bar{x},\bar{y}) = 0 \\ \dfrac{\partial f}{\partial y}(\bar{x},\bar{y}) + \lambda\dfrac{\partial h}{\partial y}(\bar{x},\bar{y}) = 0 \\ h(x,y) = 2x^2 + y^2 = 3 \end{cases} \Longrightarrow \begin{cases} -2xy + 4x\lambda = 0 \\ -x^2 + 2y\lambda = 0 \\ 2x^2 + y^2 = 3 \end{cases} \Longrightarrow \begin{cases} 2xy - 4x\lambda = 0 \quad (1) \\ x^2 - 2y\lambda = 0 \quad (2) \\ 2x^2 - y^2 = 3 \quad (3) \end{cases}$$

Da equação (1) *temos*

$$2x(y - 2\lambda) = 0,$$

logo $x = 0$ *ou* $y - 2\lambda = 0$. *Se* $x = 0$, *então*

$$2x^2 + y^2 = 3 \Longrightarrow y^2 = 3 \Longrightarrow y = \pm\sqrt{3}.$$

Deste modo, em (2), *podemos concluir que* $\lambda = 0$. *Por fim, podemos concluir que os pontos*

$$(0, -\sqrt{3}, 0) \text{ e } (0, +\sqrt{3}, 0),$$

são soluções do sistema acima.

Por outro lado, se $y - 2\lambda = 0$ *temos que* $y = 2\lambda$ *e então* $y^2 = 4\lambda^2$. *Da equação* (2) *temos*

$$x^2 - 2y\lambda \Longrightarrow x^2 = 4\lambda^2,$$

substituindo essas informações em (3), *obtemos* $8\lambda^2 + 4\lambda^2 = 3$, *ou seja,*

$$\lambda^2 = \frac{1}{4} \Longrightarrow \lambda = \pm\frac{1}{2}.$$

Se $\lambda = \frac{1}{2}$, *então* $y = 1$ *e* $x^2 = 1$, *isto é,* $x = \pm 1$. *Se* $\lambda = -\frac{1}{2}$, *então* $y = -1$ *e* $x = \pm 1$. *Portanto, os pontos*

$$(-1, 1, 1/2), (1, 1, 1/2), (-1, -1, -1/2), (1, -1, -1/2),$$

também são solução do sistema.

Pelo Teorema 4.1.2, qualquer ponto de mínimo problema de otimização deve ser solução do sistema associado às condições de primeira ordem, portanto os pontos que procuramos estão entre

$$(-1, 1), (1, 1), (-1, -1), (1, -1), (0, \sqrt{3}) \quad e \quad (0, -\sqrt{3}).$$

Como o problema apresenta uma função contínua definida em um conjunto compacto (Weierstrass), podemos substituir os pontos encontrados na função objetivo para encontrar o mínimo

$$f(0,\sqrt{3}) = f(0,-\sqrt{3}) = 0$$

$$f(1,1) = f(-1,1) = -1$$

$$f(-1,-1) = f(1,-1) = 1$$

Portanto, os pontos $(1,1)$ *e* $(-1,1)$ *são mínimos globais do problema de otimização.*

O exemplo anterior nos mostra uma aplicação direta do Teorema 4.1.2, é importante ressaltar que as soluções do sistema associado as condições de primeira ordem são possíveis candidatos a mínimos do problema inicial, porém a condição não nos fornece nenhuma informação sobre a natureza destes pontos. Considere o exemplo a seguir.

Exemplo 4.1.2. *(7) Encontre os extremos do seguinte problema de otimização*

$$\begin{cases} minimizar \quad f(x,y) = -xy \\ s.\ a \quad (x,y) \in \Omega = \{(x,y) \in \mathbb{R}^2 \mid h(x,y) = x+y = 6\}. \end{cases}$$

Para aplicar o Teorema, vimos que o ponto $(\bar{x},\bar{y}) \in \Omega$ *deve ser regular. Observe que,*

$$\nabla h(x,y) = (1,1).$$

Note que $\nabla h(x,y) \neq (0,0)$ *para todo ponto em* $\mathbb{R}^2$*. O sistema associado as condições de primeira ordem são dados por:*

$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) + \lambda \dfrac{\partial h}{\partial x}(x,y) = 0 \\ \dfrac{\partial f}{\partial y}(x,y) + \lambda \dfrac{\partial h}{\partial y}(x,y) = 0 \\ h(x,y) = x+y = 6 \end{cases} \implies \begin{cases} -y+\lambda = 0 \quad (1) \\ -x+\lambda = 0 \quad (2) \\ x+y = 6 \quad (3) \end{cases}$$

Das equações (1) *e* (2) *temos* $x = y = \lambda$*, substituindo na equação* (3) *obtemos* $2\lambda = 6$ *o que implica* $\lambda = 3$ *e por fim temos* $x = y = 3$*.*

Portanto, o unico ponto que satisfaz as condições de primeira ordem é o ponto $(3,3,3)$*.*

O teorema de Weierstrass diz que uma função contínua f definida em um conjunto compacto possui pelo menos um máximo global e pelo menos um mínimo global. Porém, o teorema não se aplica nesse caso, diferente do exemplo anterior. A seguir, vamos desenvolver uma maneira de avaliar a natureza do ponto $(3,3,3)$.

4.1.2 Condição necessária e suficiente de 2ª ordem

Como visto nos exemplos anteriores, os resultados apresentados garantem que, se um ponto é extremo local da função objetivo f então este é solução do sistema 4.4, associado ao problema original através dos multiplicadores de Lagrange. Contudo, nada podemos afirmar sobre a natureza desses pontos. Para os problemas de otimização sem restrição, na seção 3.1, estudamos ferramentas que permitem classificar a natureza das soluções obtidas nas condições de primeira ordem.

Através da Hessiana do sistema 4.4 podemos realizar um estudo semelhante ao que foi feito no capítulo anterior, e assim apresentar condições necessárias e suficientes de segunda ordem.

Teorema 4.1.3 (Condição necessária de 2ª ordem)**.** *Uma condição necessária para que o ponto regular $\bar{x} \in \Omega$ seja mínimo local de f, onde $\bar{x}$ satisfaz a condição de primeira ordem 4.4 é que a matriz*

$$H(\lambda, \bar{x}) = Hf(\bar{x}) + \lambda \cdot Hh(\bar{x}), \tag{4.5}$$

seja ***positiva semidefinida*** *em $T_{\bar{x}} = N(\nabla h(\bar{x}))$.*

Retornemos ao Exemplo 3.2.1, minimizar $f(x, y) = x^2 - y^2$, vimos que o ponto $(0, 0)$ é o único ponto crítico, porém este é um ponto de sela. Considere agora o seguinte problema de otimização

$$\begin{cases} \text{Minimizar} & f(x,y) = x^2 - y^2 \\ \text{s. a} & h(x,y) = y = 0. \end{cases} \tag{4.6}$$

Note que $\nabla h(x, y) = (0, 1) \neq (0, 0)$ para todo ponto em $\mathbb{R}^2$. As condições de primeira ordem são dadas por

$$\begin{cases} 2x = 0 \\ -2y = 0 \\ y = 0, \end{cases}$$

que possui como solução o ponto $(\lambda, x, y) = (0, 0, 0)$. Como f está restrita à $y = 0$, então $f(x, 0) = x^2$ e, obviamente, o ponto $(0, 0)$ é solução do problema. Observe que $H(\lambda, \bar{x}) = Hf(\bar{x}) + \lambda \cdot Hh(\bar{x})$ é dada por

$$H(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix},$$

além disso $N(\nabla h(x, y)) = \{(x_0, y_0) \in \mathbb{R}^2 \mid (0, 1) \cdot (x_0, y_0) = 0\}$. Deste modo, concluímos que

$$(x_0, 0) \cdot H(0,0)(x_0, 0) = 2x_0^2 > 0,\ \forall (x_0, 0) \in N(\nabla h(x,y)) - \{(0,0)\}, \tag{4.7}$$

a matriz $H(\lambda, x, y)$ é positiva definida em $N(\nabla h(x,y))$ e o ponto $(0,0)$ satisfaz as condições necessárias de segunda ordem. Retornaremos nesse exemplo futuramente para destacar a importância desta condição.

Devemos notar que os testes fornecidos até agora, para problemas com restrições, são ferramentas que permitem encontrar pontos críticos que podem ser, ou não, mínimos do problema proposto.

No capítulo anterior, vimos que, se a Hessiana do problema é positiva definida, então temos um mínimo local estrito, com o auxílio dos multiplicadores de Lagrange é possível desenvolver um teste semelhante.

Consideremos o caso para uma função de duas variáveis, minimizar $f(x,y)$ sujeita à restrição $(x,y) \in \Omega = \{(x,y) \in \mathbb{R}^2 \mid h(x,y) = 0\ \}$, onde f e h são no mínimo de classe $C^2(\mathbb{R}^2)$. Defina a função

$$\mathbb{L}(\lambda, x, y) = f(x,y) + \lambda \cdot h(x,y), \tag{4.8}$$

e suponha que a condição de primeira ordem 4.1.2 seja satisfeita para o ponto $(\lambda, \bar{x}, \bar{y})$, isto é,

$$\begin{cases} \nabla f(\bar{x}, \bar{y}) + \lambda \cdot \nabla h(\bar{x}, \bar{y}) = 0 \\ h(\bar{x}, \bar{y}) = 0. \end{cases}$$

Note que o conjunto Ω é uma curva de nível, estamos interessados na variação da função objetivo f ao longo dessa curva. Como $(\lambda, \bar{x}, \bar{y})$ é um ponto regular, então os vetores $\nabla h(\bar{x}, \bar{y})$ são *l.i.* e, pelo Teorema da Função Implícita 2.0.4 é possível escrever y em função de x em uma vizinhança de $\bar{x}$. Suponha que $y = \phi(x)$ seja o gráfico de Ω, deste modo, a função objetivo $f(x,y)$ restrita a $(x,y) \in \Omega$, pode ser escrita como $f(x,y) = f(x, \phi(x))$. Em certo sentido, este é um problema de uma variável, sendo assim, é possível realizar um teste parecido com o que foi feito no capítulo anterior.

Utilizando a regra da cadeia na função $f(x,y) = f(x, \phi x)$, temos

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{d\phi}{dx}, \tag{4.9}$$

a segunda derivada é dada por

$$\frac{d^2 f}{dx^2} = \frac{\partial^2 f}{\partial x^2} + 2\frac{\partial^2 f}{\partial x \partial y}\frac{d\phi}{dx} + \frac{\partial^2 f}{\partial y^2}\left(\frac{d\phi}{dx}\right)^2 + \frac{\partial f}{\partial y}\frac{d^2\phi}{dx^2}. \tag{4.10}$$

A restrição $h(x,y)=0$ pode ser escrita como $h(x,\phi(x))=0$, diferenciando ambos os lados da igualdade em relação a x temos

$$\frac{\partial h}{\partial x}+\frac{\partial h}{\partial y}\frac{d\phi}{dx}=0,$$

daí,

$$\frac{d\phi}{dx}=-\frac{\partial h/\partial x}{\partial h/\partial y} \tag{4.11}$$

e

$$\frac{d^2\phi}{dx^2}=-\frac{1}{\partial h/\partial y}\left[\frac{\partial^2 h}{\partial x^2}-2\frac{\partial^2 h}{\partial x\partial y}\frac{\partial h/\partial x}{\partial h/\partial y}+\frac{\partial^2 h}{\partial y^2}\left(\frac{\partial h/\partial x}{\partial h/\partial y}\right)^2\right]. \tag{4.12}$$

Substituindo a Equação (4.11) na Equação (4.9) temos

$$\frac{df}{dx}=\frac{\partial f}{\partial x}-\frac{\partial f}{\partial y}\frac{\partial h/\partial x}{\partial h/\partial y}, \tag{4.13}$$

e substituindo as Equações (4.11) e (4.12) na Equação (4.10) temos

$$\begin{aligned}\frac{d^2 f}{dx^2}=-\frac{1}{(\partial h/\partial y)^2}\quad&\left\{\left[\frac{\partial^2 f}{\partial x^2}-\frac{\partial f/\partial y}{\partial h/\partial y}\frac{\partial^2 h}{\partial x^2}\right]\left(\frac{\partial h}{\partial y}\right)^2-2\left[\frac{\partial^2 f}{\partial x\partial y}-\frac{\partial f/\partial y}{\partial h/\partial y}\frac{\partial^2 h}{\partial x\partial y}\right]\frac{\partial h}{\partial x}\frac{\partial h}{\partial y}\right.\\&\left.+\left[\frac{\partial^2 f}{\partial y^2}-\frac{\partial f/\partial y}{\partial h/\partial y}\frac{\partial^2 h}{\partial y^2}\right]\left(\frac{\partial h}{\partial x}\right)^2\right\}.\end{aligned} \tag{4.14}$$

Sabemos que a condição de primeira ordem é satisifeita no ponto $(\lambda,\bar{x},\bar{y})$, isto é,

$$\nabla f(\bar{x},\bar{y})+\lambda\cdot\nabla h(\bar{x},\bar{y})=0,$$

logo temos que $\dfrac{\partial f}{\partial x}=-\lambda\dfrac{\partial h}{\partial x}$ e $\dfrac{\partial f}{\partial y}=-\lambda\dfrac{\partial h}{\partial y}$. Deste modo a Equação (4.13) no ponto $\bar{x}$ pode ser reescrita como

$$\left.\frac{df}{dx}\right|_{\bar{x}}=\left.\frac{\partial f}{\partial x}\right|_{\bar{x}}+\lambda\left.\frac{\partial h}{\partial x}\right|_{\bar{x}}=0,$$

finalmente, a Equação (4.14) no ponto $\bar{x}$ é dada por

$$\left.\frac{d^2 f}{dx^2}\right|_{\bar{x}}=-\frac{1}{(\partial h/\partial y)^2}\left[\frac{\partial^2\mathbb{L}}{\partial x^2}\left(\frac{\partial h}{\partial y}\right)^2-2\frac{\partial^2\mathbb{L}}{\partial x\partial y}\frac{\partial h}{\partial x}\frac{\partial h}{\partial y}+\frac{\partial^2\mathbb{L}}{\partial y^2}\left(\frac{\partial h}{\partial y}\right)^2\right]$$

$$\left.\frac{d^2 f}{dx^2}\right|_{\bar{x}}=-\frac{1}{(\partial h/\partial y)^2}\begin{vmatrix}0 & -\dfrac{\partial h}{\partial x} & -\dfrac{\partial h}{\partial y}\\ -\dfrac{\partial h}{\partial x} & \dfrac{\partial^2\mathbb{L}}{\partial x^2} & \dfrac{\partial^2\mathbb{L}}{\partial x\partial y}\\ -\dfrac{\partial h}{\partial y} & \dfrac{\partial^2\mathbb{L}}{\partial x\partial y} & \dfrac{\partial^2\mathbb{L}}{\partial y^2}\end{vmatrix} \tag{4.15}$$

onde $\mathbb{L}$ é a função auxiliar definida em (4.8). O determinante 3×3 encontrado na Equação (4.15) é chamado de **Hessiano orlado** ou **Hessiano com borda**, o seu sinal é oposto ao de $\frac{d^2 f}{dx^2}$. Portanto, caso o sinal do Hessiano orlado seja negativo, o ponto é um mínimo local, caso seja positiva, um máximo local, e caso o valor seja zero, o teste é inconclusivo. Podemos resumir o que foi dito no seguinte resultado:

Teorema 4.1.4. *(Condição suficiente de 2ª ordem) Considere o problema de minimizar $f(x, y)$ sujeito a $(x, y) \in \Omega = \{(x, y) \in \mathbb{R}^2 \mid h(x, y) = 0\}$, onde f e h são no mínimo de classe $C^2(\mathbb{R}^2)$, além disso, considere que a condição de primeira ordem seja satisfeita no ponto $(\lambda, \bar{x}, \bar{y})$. Forme a função auxiliar $\mathbb{L} = f(x, y) + \lambda \cdot h(x, y)$ e o determinante da Hessiana orlada*

$$|\bar{H}| = \begin{vmatrix} 0 & -\dfrac{\partial h}{\partial x} & -\dfrac{\partial h}{\partial y} \\ -\dfrac{\partial h}{\partial x} & \dfrac{\partial^2 \mathbb{L}}{\partial x^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x \partial y} \\ -\dfrac{\partial h}{\partial y} & \dfrac{\partial^2 \mathbb{L}}{\partial x \partial y} & \dfrac{\partial^2 \mathbb{L}}{\partial y^2} \end{vmatrix} \tag{4.16}$$

1. *Se $|\bar{H}| > 0$, então $(\lambda, \bar{x}, \bar{y})$ é um máximo local;*
2. *Se $|\bar{H}| < 0$, então $(\lambda, \bar{x}, \bar{y})$ é um mínimo local;*
3. *Se $|\bar{H}| = 0$, o teste é inconclusivo e o ponto $(\lambda, \bar{x}, \bar{y})$ pode ser um mínimo, máximo ou nenhum dos dois.*

Condisidere novamente o problema 4.6, isto é,

$$\begin{cases} \text{Minimizar} & f(x, y) = x^2 - y^2 \\ \text{s. a} & h(x, y) = y = 0. \end{cases}$$

a função auxiliar é dada por

$$\mathbb{L}(\lambda, x, y) = x^2 - y^2 + \lambda \cdot y,$$

e o determinante Hessiano orlado tomará a seguinte fomra

$$|\bar{H}|(\lambda, x, y) = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{vmatrix} = 0. \tag{4.17}$$

Observe que, se utilizássemos apenas o resultado apresentado no Teorema 4.1.4, o teste seria inconclusivo. Porém, por meio da condição necessária de segunda ordem, é possível obter a solução do problema.

Neste exemplo, lidamos com um caso simples em $\mathbb{R}^2$, com o intuito de enfatizar a importância da condição 4.1.3. Contudo, como essa condição envolve o núcleo da jacobiana das restrições, que é um subespaço, e o teste acaba exigindo um esforço computacional considerável. Como será discutido no próximo capítulo, essa complexidade limita a utilização desta condição em algoritmos.

Exemplo 4.1.3. *Retornemos ao Exemplo (4.1.2), dado por*

$$\begin{cases} \textit{minimizar} \quad f(x,y) = -xy \\ \textit{s. a} \quad (x,y) \in \Omega = \{(x,y) \in \mathbb{R}^2 \mid h(x,y) = x + y = 6\}. \end{cases}$$

Vimos que o único ponto que satisfaz a condição de primeira ordem é o ponto $(3,3,-3)$*, vamos usar o Teorema (4.1.4) para classificar o ponto encontrado. A função auxiliar do problema é definida por*

$$\mathbb{L} = -xy + \lambda \cdot (x + y - 6),$$

o determinante hessiano orlado do problema é dado por

$$|\bar{H}| = \begin{vmatrix} 0 & -\dfrac{\partial h}{\partial x} & -\dfrac{\partial h}{\partial y} \\ -\dfrac{\partial h}{\partial x} & \dfrac{\partial^2 \mathbb{L}}{\partial x^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x \partial y} \\ -\dfrac{\partial h}{\partial y} & \dfrac{\partial^2 \mathbb{L}}{\partial x \partial y} & \dfrac{\partial^2 \mathbb{L}}{\partial y^2} \end{vmatrix} = \begin{vmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{vmatrix} = -2 < 0.$$

Portanto, podemos afirmar que o ponto $(3,3,3)$ *é um mínimo local do problema.*

Para que as condições de primeira e segunda ordem vistas até aqui possam ser aplicadas, é necessário que o ponto candidato a mínimo do problema seja regular. A linearidade independente dos gradientes das restrições garante que os multiplicadores de Lagrange existam. Deste modo, caso um ponto não seja regular, não podemos aplicar os resultados vistos, pois os teoremas dependem da regularidade do ponto. Considere o exemplo a seguir

Exemplo 4.1.4.

$$\begin{cases} \textit{Minimizar} \quad f(x,y) = -y \\ \textit{s. a} \quad (x,y) \in \Omega = \{(x,y) \in \mathbb{R}^2 \mid h(x,y) = x^2 + y^2 = 0\}. \end{cases} \tag{4.18}$$

Para aplicarmos os resultados vistos até aqui, precisamos satisfazer as condições de regularidade, para este caso temos

$$\nabla h(x,y) = (2x, 2y) = (0,0) \Longleftrightarrow (x,y) = (0,0).$$

Podemos notar que a restrição $h(x,y) = x^2 + y^2 = 0$ possui solução única, $x = 0$ e $y = 0$. Como o ponto $(0,0)$ não é regular não podemos aplicar os teoremas vistos no decorrer do capítulo, sendo necessário realizar outra análise do problema.

Como queremos minimizar $f(x,y) = -y$ e a restrição exige $y = 0$, deste modo a única solução para o problema de otimização é o ponto $(0,0)$, onde

$$f(0,0) = -0 = 0.$$

Como dito anteriormente, a condição de regularidade garante a existência dos multiplicadores de lagrange, podemos observar na Figura 4.2 que não existe $\lambda \nabla h(x,y)$ na solução do problema. De fato, o sistema associado as condições de primeira ordem do problema 4.18 é dado por

$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) + \lambda \dfrac{\partial h}{\partial x}(x,y) = 0 \\ \dfrac{\partial f}{\partial y}(x,y) + \lambda \dfrac{\partial h}{\partial y}(x,y) = 0 \\ h(x,y) = x^2 + y^2 = 0 \end{cases} \Longrightarrow \begin{cases} 2x\lambda = 0 \quad (1) \\ 1 + 2y\lambda = 0 \quad (2) \\ x^2 + y^2 = 0 \quad (3) \end{cases}$$

O sistema acima não possui solução, deste modo não existe λ tal que

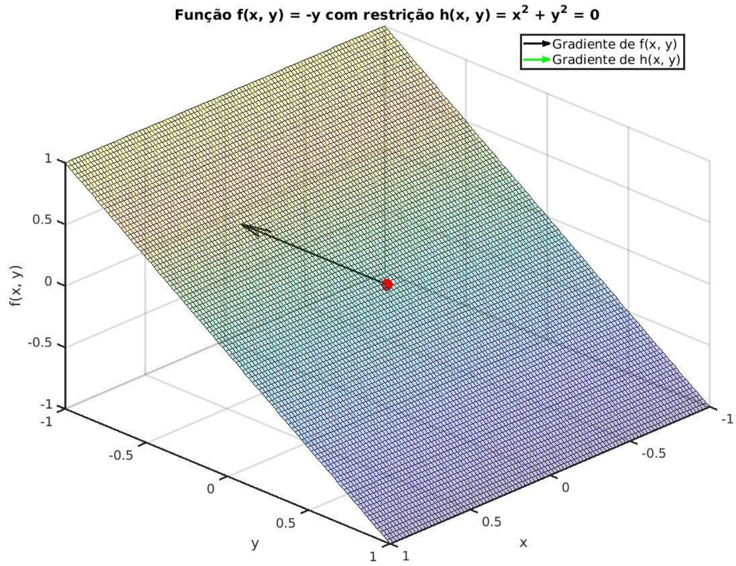
$$\nabla f(\bar{x}, \bar{y}) + \lambda \cdot \nabla h(\bar{x}, \bar{y}) = 0.$$

O exemplo anterior ilustra o caso onde a solução do problema era um ponto não regular. Vimos que se fez necessário buscar outra maneira de se resolver o problema. A regularidade permite tratar o problema de otimização de maneira mais simples e eficiente, pois facilita a análise local do problema e a aplicação de algoritmos de otimização. A presença de pontos irregulares pode complicar a formulação do problema e exigir técnicas mais sofisticadas para lidar com singularidades.

É possível generalizar a condição de segunda para o caso de n variáveis, considere o problema inicial (P)

$$\begin{cases} \text{Minimizar} \quad f(x) \\ \text{sujeito a} \quad x \in \Omega = \{x \in \mathbb{R}^n \mid h_i(x) = 0\}, \\ \text{com} \quad i = 1, 2, ...m, \end{cases}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ e $h : \mathbb{R}^n \to \mathbb{R}^m$, $h(x) = (h_1(x), h_2(x), ..., h_m(x))$, são no mínimo de classe $C^2(\mathbb{R}^n)$. Neste caso, a função auxiliar e a matriz Hessiana orlada serão definidas, respectivamente, da seguinte maneira



– Figura 4.2 - Interpretação Geométrica do do Exemplo 4.1.4

$$\mathbb{L} = f(x) + \lambda \cdot h(x), \text{ com } \lambda = (\lambda_1, ..., \lambda_m)$$

e

$$\bar{H} = \begin{bmatrix} 0 & -\dfrac{\partial h}{\partial x_1} & -\dfrac{\partial h}{\partial x_2} & \cdots & -\dfrac{\partial h}{\partial x_n} \\ -\dfrac{\partial h}{\partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_n} \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ -\dfrac{\partial h}{\partial x_n} & \dfrac{\partial^2 \mathbb{L}}{\partial x_n \partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 \mathbb{L}}{\partial x_n^2} \end{bmatrix}. \tag{4.19}$$

Assim como no capítulo anterior, a análise dos pontos obtidos na condição de primeira ordem será feita através dos menores principais da matriz Hessiana orlada, onde

$$|\bar{H}_2| = \begin{vmatrix} 0 & -\dfrac{\partial h}{\partial x_1} & -\dfrac{\partial h}{\partial x_2} \\ -\dfrac{\partial h}{\partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_2} \\ -\dfrac{\partial h}{\partial x_2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_2 \partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_2^2} \end{vmatrix}, \quad |\bar{H}_3| = \begin{vmatrix} 0 & -\dfrac{\partial h}{\partial x_1} & -\dfrac{\partial h}{\partial x_2} & -\dfrac{\partial h}{\partial x_3} \\ -\dfrac{\partial h}{\partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_3} \\ -\dfrac{\partial h}{\partial x_2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_2 \partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_2^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_2 \partial x_3} \\ -\dfrac{\partial h}{\partial x_3} & \dfrac{\partial^2 \mathbb{L}}{\partial x_3 \partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_3 \partial x_2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_3^2} \end{vmatrix}$$

Deste modo temos que $|\bar{H}_i|$, $(i = 1, 2, ..., n)$ são os menores principais da Hessiana orlada, além disso $|\bar{H}_n| = |\bar{H}|$. Assim como no caso sem restrições, para que um ponto seja mínimo local, é suficiente que a matriz Hessiana orlada seja positiva definida.

Teorema 4.1.5. *Considere o problema de minimizar $f(x)$ sujeito a $x \in \Omega = \{x \in \mathbb{R}^n \mid h(x) = 0\}$, onde $f : \mathbb{R}^n \to \mathbb{R}$ e $h : \mathbb{R}^n \to \mathbb{R}^m$, $h(x) = (h_1(x), h_2(x), ..., h_m(x))$, são no mínimo de classe $C^2(\mathbb{R}^n)$. Suponha que o ponto $(\lambda, \bar{x})$ satisfaça a condição de primeira ordem e forme a função auxiliar $\mathbb{L} = f(x) + \lambda \cdot h(x)$. Seja $|\bar{H}_i|$, $(i = 1, 2, ..., n)$ os menores principais da Hessiana orlada $|\bar{H}|$, se*

- $|\bar{H}_2| < 0$, $|\bar{H}_3| < 0$,...,$|\bar{H}_n| < 0$*, então a matriz é positiva definida, e temos um ponto de mínimo local;*
- $|\bar{H}_2| > 0$, $|\bar{H}_3| < 0$, $|\bar{H}_4| > 0$, ...,$(-1)^n|\bar{H}_n| > 0$*, então a matriz é negativa definida, e temos um ponto de máximo local.*

Note que a matriz Hessiana orlada nada mais é que a matriz Hessiana da função objetivo acrescida de uma linha e uma coluna, formadas por 0 e as derivadas parciais da restrição. Para o caso de várias restrições, serão acrescentadas linhas e colunas conforme a quantidade de restrições. Suponha que o problema contenha n variáveis e m restrições $(m < n)$, para simplificar a notação, considere que $h^j(x) = 0, (j = 1, ..., m)$, e $h_i^j = \dfrac{\partial h^j}{\partial x_i}$ sejam as restrições do problema e suas derivadas parciais, respectivamente. A função auxiliar toma a forma

$$\mathbb{L} = f(x) + \sum_{j=1}^{m} \lambda_j \cdot h^j. \tag{4.20}$$

Deste modo o hessiano orlado aparecerá da seguinte maneira

$$|\bar{H}| = \begin{vmatrix} 0 & 0 & \cdots & 0 & | & h_1^1 & h_1^2 & \cdots & h_n^1 \\ 0 & 0 & \cdots & 0 & | & h_1^2 & h_2^2 & \cdots & h_n^2 \\ - & - & - & - & - & - & - & - & - \\ 0 & 0 & \cdots & 0 & | & h_1^m & h_2^m & \cdots & h_n^m \\ - & - & - & - & - & - & - & - & - \\ h_1^1 & h_1^2 & \cdots & h_1^m & | & \mathbb{L}_{11} & \mathbb{L}_{12} & \cdots & \mathbb{L}_{1n} \\ h_2^1 & h_2^2 & \cdots & h_2^m & | & \mathbb{L}_{21} & \mathbb{L}_{22} & \cdots & \mathbb{L}_{2n} \\ - & - & - & - & - & - & - & - & - \\ h_n^1 & h_n^2 & \cdots & h_n^m & | & \mathbb{L}_{n1} & \mathbb{L}_{n2} & \cdots & \mathbb{L}_{nn} \end{vmatrix} \tag{4.21}$$

onde os símbolos $\mathbb{L}_{ij}$ representam as derivadas parciais da função auxiliar. Vários menores principais podem ser formados através da matriz acima. O menor principal que tiver $\mathbb{L}_{22}$ como último elemento de sua diagonal principal será denotado por $|\bar{H}_2|$, ao incluir uma linha e uma coluna, obtemos $|\bar{H}_3|$ e assim sucessivamente. A condição suficiente de segunda ordem para o determinante acima é estabelecida em termo dos sinais dos últimos $(n-m)$ menores principais da matriz, sendo que, para um mínimo local, é suficiente que os menores principais tenham o sinal de $(-1)^m$.

Existe uma maneira de escrever as condições de primeira e segunda ordem para o problema (P) em termos da função auxiliar usada anteriormente, para isso precisamos da seguinte definição:

Definição 4.1.3. *A função lagrangiana, ou lagrangiano, associada ao problema de otimização (P) é a função $\mathbb{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ definida por*

$$\mathbb{L}(\lambda, x) = f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x). \tag{4.22}$$

Vamos apresentar as condições necessárias de primeira ordem para um problema de otimização com restrições de igualdade em termos da lagrangiana. Ao calcularmos as derivadas parciais da função, obtemos:

$$\begin{cases} \dfrac{\partial \mathbb{L}}{\partial x_1}(\lambda, x) = \dfrac{\partial f}{\partial x_1}(x) + \lambda \dfrac{\partial h}{\partial x_1}(x) \\ \vdots \\ \dfrac{\partial \mathbb{L}}{\partial x_n}(\lambda, x) = \dfrac{\partial f}{\partial x_n}(x) + \lambda \dfrac{\partial h}{\partial x_n}(x) \\ \dfrac{\partial \mathbb{L}}{\partial \lambda}(\lambda, x) = h(x). \end{cases} \tag{4.23}$$

Além disso, se o ponto (λ, x) atende as condições de primeira ordem apresentadas no Teorema 4.1.2, então este é ponto crítico da função lagrangiana (4.22), isso significa que podemos reescrever as **condições necessárias de primeira ordem** em termos da função lagrangiana.

Teorema 4.1.6 (Condições necessárias de 1ª Ordem). *Seja $\bar{x} \in \Omega$ e suponha $\nabla h(\bar{x})$ sejam l.i. ($\bar{x}$ é um ponto regular). Se $\bar{x}$ é um extremo local de f, então existem m multiplicadores de Lagrange $\lambda = (\lambda_1, ..., \lambda_m)$, tal que $(\lambda, \bar{x})$ é ponto crítico do lagrangiano 4.22, isto é,*

$$\begin{cases} \dfrac{\partial \mathbb{L}}{\partial x_1}(\lambda, \bar{x}) = 0 \\ \vdots \\ \dfrac{\partial \mathbb{L}}{\partial x_n}(\lambda, \bar{x}) = 0 \\ \dfrac{\partial \mathbb{L}}{\partial \lambda_1}(\lambda, \bar{x}) = 0 \\ \vdots \\ \dfrac{\partial \mathbb{L}}{\partial \lambda_m}(\lambda, \bar{x}) = 0. \end{cases} \tag{4.24}$$

Observe na definição da função lagrangiana que λ multiplica apenas as restrições do problema, isso significa que as equações

$$\frac{\partial \mathbb{L}}{\partial \lambda_i}(\lambda, \bar{x}) = 0 \text{ com } i = 1, ..., m,$$

são equivalentes as equações de restrição.

Teorema 4.1.7 (Condição suficiente de 2ª ordem). *Considere o problema de minimizar $f(x)$ sujeito a $x \in \Omega = \{x \in \mathbb{R}^n \mid h(x) = 0\}$, onde $f : \mathbb{R}^n \to \mathbb{R}$ e $h : \mathbb{R}^n \to \mathbb{R}^m$, $h(x) = (h_1(x), h_2(x), ..., h_m(x))$, são no mínimo de classe $C^2(\mathbb{R}^n)$. Suponha que o ponto $(\lambda, \bar{x})$ satisfaça a condição de primeira ordem e forme a função lagrangiana $\mathbb{L} = f(x) + \lambda \cdot h(x)$. Seja $|\bar{H}_i|$, $(i = 1, 2, ..., n)$ os menores principais da Hessiana orlada*

$$|\bar{H}| = \begin{vmatrix} 0 & -\dfrac{\partial h}{\partial x_1} & -\dfrac{\partial h}{\partial x_2} & \cdots & -\dfrac{\partial h}{\partial x_n} \\ -\dfrac{\partial h}{\partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_n} \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ -\dfrac{\partial h}{\partial x_n} & \dfrac{\partial^2 \mathbb{L}}{\partial x_n \partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 \mathbb{L}}{\partial x_n^2} \end{vmatrix}. \tag{4.25}$$

Se

- $|\bar{H}_2| < 0$, $|\bar{H}_3| < 0$,...,$|\bar{H}_n| < 0$, *então a matriz é positiva definida, e temos um ponto de mínimo local;*
- $|\bar{H}_2| > 0$, $|\bar{H}_3| < 0$, $|\bar{H}_4| > 0$, ...,$(-1)^n|\bar{H}_n| > 0$, *então a matriz é negativa definida, e temos um ponto de máximo local.*

Exemplo 4.1.5. *Considere o problema de otimização*

$$\begin{cases} minimizar \quad f(x,y,z) = x + 2y + 2z \\ s.\ a \quad (x,y,z) \in \Omega = \begin{cases} (x,y,z) \in \mathbb{R}^3 \mid h_1(x,y,z) = x^2 + y^2 = 25; \\ (x,y,z) \in \mathbb{R}^3 \mid h_2(x,y,z) = x + y + z = 0.\} \end{cases} \end{cases}$$

Para um problema de otimização com apenas uma restrição, vimos que um ponto $\bar{x}$ *que satisfaz a restrição* $h_i(\bar{x}) = 0$, $(i = 1, 2)$, *é regular se* $\nabla h_1(\bar{x})$ *e* $\nabla h_2(\bar{x})$ *são l.i. Para o caso com mais de uma restrição, basta verificar se o posto da matriz escalonada equivalente à matriz jacobiana de h no ponto x é igual ao número de restrições, o que é equivalente a dizer que os vetores gradientes de h são linearmente independentes.*

A matriz jacobiana $Jh(x,y,z)$ *é dada por*

$$\begin{bmatrix} 2x & 2y & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Se $x = 0$, *então* $y^2 = 25$ *e assim* $y = \pm 5$. *A matriz jacobiana toma a seguinte forma*

$$\begin{bmatrix} 0 & \pm 10 & 0 \\ 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & \pm 10 & 0 \end{bmatrix},$$

que tem posto igual a 2. Caso $x \neq 0$, *podemos multiplicar* $2x$ *pela linha dois da matriz jacobiana e subtrair na linha um, esse resultado substituímos na linha dois. Deste modo a jacobiana será escrita da seguinte maneira*

$$\begin{bmatrix} 2x & 2y & 0 \\ 0 & 2y-2x & -2x \end{bmatrix}.$$

que também tem posto igual a 2. Portanto todos os pontos da região viável irão satisfazer a condição de regularidade.

A função lagrangiana do problema é definida da seguinte maneira

$$\mathbb{L}(\lambda_1, \lambda_2, x, y, z) = x + 2y + 2z + \lambda_1 \cdot (x^2 + y^2 - 25) + \lambda_2 \cdot (x + y + z),$$

calculando as derivadas parciais de primeira ordem obtemos o seguinte sistema

$$\begin{cases} \dfrac{\partial \mathbb{L}}{\partial x} = 1 + 2x\lambda_1 + \lambda_2 = 0 \\ \dfrac{\partial \mathbb{L}}{\partial y} = 2 + 2y\lambda_1 + \lambda_2 = 0 \\ \dfrac{\partial \mathbb{L}}{\partial z} = 2 + \lambda_2 = 0 \\ \dfrac{\partial \mathbb{L}}{\partial \lambda_1} = x^2 + y^2 - 25 = 0 \\ \dfrac{\partial \mathbb{L}}{\partial \lambda_2} = x + y + z = 0. \end{cases}$$

Resolvendo o sistema associado as condições de primeira ordem obtemos os seguintes pontos críticos

$$\left(5, 0, -5, \frac{1}{10}, -2\right) \; e \; \left(-5, 0, 5, -\frac{1}{10}, -2\right).$$

Com as derivadas parciais de segunda ordem obtemos a Hessiana orlada

$$\bar{H} = \begin{bmatrix} 0 & 0 & 2x & 2y & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 2x & 1 & -2\lambda_1 & 0 & 0 \\ 2y & 1 & 0 & -2\lambda_1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

O problema inicial possui três variáveis ($n = 3$) e duas restrições ($m = 2$), a condição de segunda ordem será dada através da análise do último menor principal, uma vez que ($n - m = 3 - 2 = 1$), sendo que, para um mínimo local basta que o sinal do determinante seja igual ao de $(-1)^m = 1$, ou seja, positivo.

Para o ponto $\left(5, 0, -5, \frac{1}{10}, -2\right)$ *temos*

$$|\bar{H}| = \begin{vmatrix} 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 10 & 1 & 1/5 & 0 & 0 \\ 0 & 1 & 0 & 1/5 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{vmatrix} = -20 < 0,$$

deste modo, temos um máximo local. Para o ponto $\left(-5, 0, 5, \dfrac{-1}{10}, -2\right)$

$$|\bar{H}| = \begin{vmatrix} 0 & 0 & -10 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ -10 & 1 & -1/5 & 0 & 0 \\ 0 & 1 & 0 & -1/5 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{vmatrix} = 20 > 0$$

onde encontramos um valor de mínimo local.

4.2 OTIMIZAÇÃO COM RESTRIÇÕES DE DESIGUALDADE

Na seção anterior, estudamos problemas de otimização com restrições de igualdade, vimos que, para encontrar os candidatos à solução do problema, em uma região viável, basta encontrar as soluções do sistema de equações associado ao problema, levando em conta as condições de regularidade. Entretanto, a grande maioria dos problemas aplicados a engenharia ou à economia são modelados de tal forma que a região viável é obtida através do uso de desigualdades. Para simplificar a leitura, iremos adotar h para as restrições de igualdade e λ como os multiplicadores de h, e denotar g como restrições de desigualdade e μ como os multiplicadores de g. Considere o seguinte problema de minimização

$$(P)\begin{cases} \text{Minimizar} \quad f(x) \\ \text{sujeito a} \quad x \in \Omega = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0,\} \\ \text{com} \quad j = 1, 2, ...p. \end{cases} \tag{4.26}$$

Onde $f : \mathbb{R}^n \to \mathbb{R}$ e $g : \mathbb{R}^n \to \mathbb{R}^p$, $g(x) = (g_1(x), g_2(x), ..., g_p(x))$, são no mínimo de classe $C^2(\mathbb{R}^n)$.

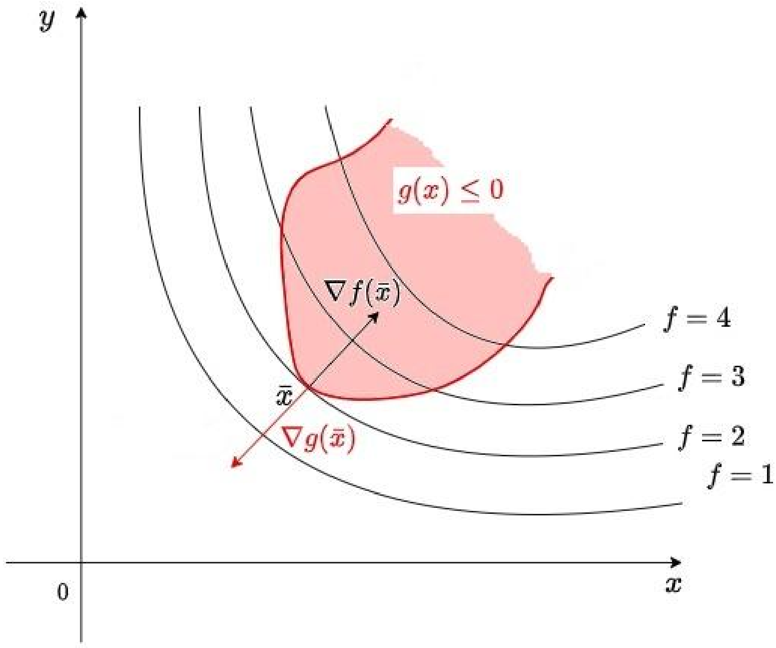
Assim como no caso com restrições de igualdade, queremos encontrar os minimizadores de f na região viável, caso existam.

Suponha que $\bar{x}$ seja uma solução do problema de otimização, isto é, suponha que $\bar{x} \in \Omega$ é um ponto mínimo de f na região viável, neste caso temos duas possibilidades: $g_j(\bar{x}) = 0$ ou $g_j(\bar{x}) < 0$. Se $g_j(\bar{x}) = 0$ é dito que a restrição está **ativa** no ponto $\bar{x}$, este é um caso semelhante ao problema com uma restrição de igualdade, deste modo, como $\bar{x}$ é ponto de mínimo de f em Ω, então vale o Teorema 4.1.2, logo

$$\nabla f(\bar{x}) + \mu \cdot \nabla g(\bar{x}) = 0.$$

Aqui vale uma condição extra, $\mu \geq 0$, onde $\mu = (\mu_1, ..., \mu_p)$. Usaremos uma interpretação geométrica para justificar esse fato. Se $\bar{x}$ é um mínimo local de f e satisfaz as condições de primeira ordem, logo é um ponto regular, isto é, $\nabla g_j(\bar{x}) \neq 0$. Deste modo, os vetores gradientes $\nabla f(\bar{x})$ e $\nabla g(\bar{x})$ são paralelos. Uma vez que $g(\bar{x}) = 0$, o ponto $\bar{x}$ está na fronteira da região viável, assim o vetor gradiente de f em $\bar{x}$ deve apontar para o interior da região viável (pois o vetor gradiente aponta a direção de maior crescimento da função no ponto), por outro lado, o vetor gradiente de g em $\bar{x}$ deve apontar para o exterior de região viável.

Por fim, podemos concluir que os vetores paralelos $\nabla f(\bar{x})$ e $\nabla g(\bar{x})$ possuem sentidos opostos, sendo assim, $\nabla f(\bar{x}) = -\mu \nabla g(\bar{x})$ com $\mu \geq 0$, a Figura 4.2 ilustra tudo o que foi dito.



– Figura 4.3 - Interpretação Geométrica para $\mu \geq 0$

Para o caso $g_j(\bar{x}) < 0$, dizemos que a restrição está **inativa** no ponto $\bar{x}$, geometricamente temos que o ponto $\bar{x}$ estará no *interior* de Ω. Observe que estamos em um caso semelhante ao que foi apresentado na Seção 3.1, onde procuramos extremos no interior da região viável, sendo assim, se $\bar{x}$ é um extremo local de f em Ω então $\bar{x}$ deve ser um ponto crítico de f.

Podemos resumir o que foi dito do seguinte modo, se a restrição está **ativa**, então

$$\begin{cases} \nabla f(\bar{x}) + \mu \cdot \nabla g(\bar{x}) = 0, \\ \mu \geq 0, \\ g(\bar{x}) = 0, \end{cases} \tag{4.27}$$

Se a restrição está **inativa**, então

$$\begin{cases} \nabla f(\bar{x}) = 0, \\ g(\bar{x}) < 0. \end{cases} \tag{4.28}$$

A condição dada por **Karush-Kuhn-Tucker (KKT)** unifica os dois casos, para isso considere o seguinte sistema

$$\begin{cases} \nabla f(\bar{x}) + \mu \cdot \nabla g(\bar{x}) = 0, \\ \mu \cdot g(\bar{x}) = 0, \\ \mu \geq 0, \\ g(\bar{x}) \leq 0. \end{cases} \tag{4.29}$$

4.2.1 Condição necessária de 1ª ordem

Apresentaremos uma **condição necessária de 1ª ordem** para o caso com restrições de desigualdade.

Teorema 4.2.1 (Condições Necessárias de 1ª Ordem)**.** *Seja $\bar{x} \in \Omega$ um mínimo local de f sujeita a restrição $g(x) \leq 0$. Suponha que $\bar{x}$ seja um ponto regular das restrições. Então existe p multiplicadores de Lagrange $\mu = (\mu_1, ..., \mu_p)$, tais que:*

$$\begin{cases} \nabla f(\bar{x}) + \mu \cdot \nabla g(\bar{x}) = 0, \\ \mu \cdot g(\bar{x}) = 0, \\ \mu \geq 0, \\ g(\bar{x}) \leq 0. \end{cases} \tag{4.30}$$

Novamente, os multiplicadores de Lagrange $\mu = (\mu_1, ..., \mu_p)$ apresentados no Teorema são números reais, e p é a quantidade de restrições associadas ao problema de otimização.

O sistema 4.30 pode ser escrito de forma equivalente através da função lagrangiana 4.22.

$$\begin{cases} \dfrac{\partial \mathbb{L}}{\partial x_1}(\mu, \bar{x}) = 0, ..., \dfrac{\partial \mathbb{L}}{\partial x_n}(\mu, \bar{x}) = 0, \\ \dfrac{\partial \mathbb{L}}{\partial \mu_1}(\mu, \bar{x}) \geq 0, ..., \dfrac{\partial \mathbb{L}}{\partial \mu_p}(\mu, \bar{x}) \geq 0, \\ \mu \cdot g(\bar{x}) = 0, \\ \mu \geq 0. \end{cases} \tag{4.31}$$

O sistema 4.24 apresentado na seção anterior e o sistema 4.31 possuem alguns pontos em comum, mas também é possível destacar algumas diferenças. Observe que nos dois casos é usado o mesmo lagrangiano $\mathbb{L}$, além disso, suas derivadas parciais em relação a x devem ser iguais a zero. Como não é possível saber se a restrição está ativa ou não, a condição

$$\frac{\partial \mathbb{L}}{\partial \lambda}(\lambda, \bar{x}) = 0$$

usada para o caso de problemas com restrições de igualdade, é inválida para o caso onde há restrições de desigualdade, pois esta condição é equivalente a $g(x) = 0$. Essa condição é substituída pelas condições a seguir

$$\mu \cdot g(\bar{x}) = 0 \quad \text{e} \quad \frac{\partial \mathbb{L}}{\partial \mu}(\mu, \bar{x}) \geq 0,$$

sendo que a segunda nada mais é que a própria restrição apresentada no problema original. Quanto a condição de regularidade, vimos que esta condição é verificada apenas quando a restrição está ativa. O multiplicador de Lagrange λ pode ser negativo ou postivo, no caso de uma restrição de igualdade, porém para um problema de otimização com restrições de desigualdades, o multiplicador μ deve ser não negativo.

Exemplo 4.2.1. *(5) Considere o seguinte problema*

$$\begin{cases} \text{minimizar} \quad f(x, y) = xy \\ \text{sujeito a} \quad g(x, y) = x^2 + y^2 \leq 1. \end{cases}$$

A condição de regularidade nos diz que o gradiente de g deve ser linearmente independente para cada candidato a solução do problema de otimização, no qual a restrição g está ***ativa****. Note que, se $g(x, y) = x^2 + y^2 = 1$ então x e y não podem ser simultaneamente nulos, como*

$$\nabla g(x, y) = (2x, 2y)$$

temos que $\nabla g(x, y) \neq (0, 0)$ para todo (x, y) onde g está ativa. Escrevendo o sistema de equações

$$\begin{cases} \nabla f(\bar{x}) + \mu \cdot \nabla g(\bar{x}) = 0, \\ \mu \cdot g(\bar{x}) = 0, \\ \mu \geq 0, \\ g(\bar{x}) \leq 0, \end{cases} \tag{4.32}$$

as Condições de 1ª ordem para esse problema são dadas por

$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) + \mu \dfrac{\partial g}{\partial x}(x,y) = y + 2\mu x = 0, \\ \dfrac{\partial f}{\partial y}(x,y) + \mu \dfrac{\partial g}{\partial y} = x + 2\mu y = 0, \\ \mu \cdot [x^2 + y^2 - 1] = 0, \\ \mu \geq 0, \\ x^2 + y^2 \leq 1. \end{cases}$$

Note que, se $\mu = 0$, então

$$y + 2\mu x = 0 \Longrightarrow y = 0$$
$$x + 2\mu y = 0 \Longrightarrow x = 0$$

e o ponto $(0,0,0)$ satisfaz as condições de 1ª ordem. Se $\mu > 0$ então

$$\mu = -\frac{y}{2x} = -\frac{x}{2y} \tag{4.33}$$
$$x^2 = y^2 \Longrightarrow x = \pm y.$$

Como $\mu > 0$, da equação $\mu \cdot [x^2 + y^2 - 1] = 0$ temos que

$$x^2 + y^2 = 1 \Longrightarrow x = y = \pm\frac{1}{\sqrt{2}}$$

Substituindo os valores de x e y na Equação 4.33 obtemos os seguintes candidatos a minimizadores do problema

$$(0,0,0), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{2}\right), \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{2}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{2}\right) \quad e \quad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2}\right).$$

A hipótese $\mu \geq 0$ elimina os pontos onde os multiplicadores de Lagrange são negativos, restando apenas três pontos que atendem todas as condições do Teorema 4.2.1:

$$(0,0,0), \quad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2}\right) \quad e \quad \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{2}\right)$$

Como a função f é contínua e a região viável é um conjunto compacto, pelo Teorema de Weierstrass, f possui máximos e mínimos globais, além disso, qualquer minimizador do problema de otimização deve ser solução do sistema de equações associados as condições de primeira ordem, deste modo, para encontrar o mínimo da função, basta calcular a função objetivo nos três pontos e selecionar o de menor valor. Portanto temos

$$f(0,0) = 0$$
$$f\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = f\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = -\frac{1}{2}.$$

Finalmente, podemos concluir que

$$\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \quad e \quad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

são mínimos globais do problema.

A interpretação da condição $\mu \cdot g(\bar{x}) = 0$ do Teorema 4.2.1 é feita da seguinte maneira: se $\mu > 0$ e $g = 0$ então a restrição está ativa e assim é possível tratá-la como uma restrição de igualdade em vez de uma restrição de desigualdade, o que é mais simples de trabalhar.

4.2.2 Condição necessária e suficiente de 2ª ordem

Teorema 4.2.2 (Condição necessária de 2ª ordem). *Uma condição necessária para que o ponto regular $\bar{x} \in \Omega$ seja mínimo local de f, onde $\bar{x}$ satisfaz a condição de primeira ordem é que a matriz*

$$H(\mu, \bar{x}) = Hf(\bar{x}) + \mu \cdot Hg(\bar{x}), \tag{4.34}$$

seja ***positiva semidefinida*** *em $T_{\bar{x}} = N(\nabla g(\bar{x}))$.*

Conforme visto no decorrer desta seção, a restrição do problema pode estar ativa ou inativa em $\bar{x}$, ou seja, dado um ponto $(\mu, \bar{x})$ que satisfaça a condição de primeira ordem, caso a restrição esteja ativa, essa será tratada como uma restrição de igualdade e assim ficam válidos os Teoremas 4.1.4 e 4.1.5 da seção anterior, onde a função lagrangiana será

$$\mathbb{L} = f(x) + \mu \cdot g(x), \tag{4.35}$$

e a matriz Hessiana orlada será

$$\bar{H} = \begin{bmatrix} 0 & -\dfrac{\partial g}{\partial x_1} & -\dfrac{\partial g}{\partial x_2} & \cdots & -\dfrac{\partial g}{\partial x_n} \\ -\dfrac{\partial g}{\partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1^2} & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 \mathbb{L}}{\partial x_1 \partial x_n} \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ -\dfrac{\partial g}{\partial x_n} & \dfrac{\partial^2 \mathbb{L}}{\partial x_n \partial x_1} & \dfrac{\partial^2 \mathbb{L}}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 \mathbb{L}}{\partial x_n^2} \end{bmatrix}. \tag{4.36}$$

Por outro lado, caso a restrição esteja inativa, os multiplicadores de Lagrange dessas restrições devem ser nulos, logo essas restrições desaparecem da Hessiana orlada 4.36 associada ao problema. Neste caso, ficam válidos os resultados apresentados no Capítulo 3.

Para o caso de várias restrições, serão acrescentadas linhas e colunas conforme a quantidade de restrições, conforme anteriormente. Suponha que o problema contenha n variáveis e p restrições ativas, para simplificar a notação, considere que $g^j(x) = 0, (j = 1, ..., p)$, e $g_i^j = \dfrac{\partial g^j}{\partial x_i}$ sejam as restrições ativas do problema e suas derivadas parciais, respectivamente. A função lagrangiana toma a forma

$$\mathbb{L} = f(x) + \sum_{j=1}^{p} \mu_j \cdot g^j. \tag{4.37}$$

Deste modo o hessiano orlado aparecerá da seguinte maneira

$$|\bar{H}| = \begin{vmatrix} 0 & 0 & \cdots & 0 & | & g_1^1 & g_1^2 & \cdots & g_n^1 \\ 0 & 0 & \cdots & 0 & | & g_1^2 & g_2^2 & \cdots & g_n^2 \\ - & - & - & - & - & - & - & - & - \\ 0 & 0 & \cdots & 0 & | & g_1^p & g_2^p & \cdots & g_n^p \\ - & - & - & - & - & - & - & - & - \\ g_1^1 & g_1^2 & \cdots & g_1^p & | & \mathbb{L}_{11} & \mathbb{L}_{12} & \cdots & \mathbb{L}_{1n} \\ g_2^1 & g_2^2 & \cdots & g_2^p & | & \mathbb{L}_{21} & \mathbb{L}_{22} & \cdots & \mathbb{L}_{2n} \\ - & - & - & - & - & - & - & - & - \\ g_n^1 & g_n^2 & \cdots & g_n^p & | & \mathbb{L}_{n1} & \mathbb{L}_{n2} & \cdots & \mathbb{L}_{nn} \end{vmatrix}. \tag{4.38}$$

A condição suficiente de segunda ordem para o determinante acima é análoga ao caso de restrições de igualdade, dado em termo dos sinais dos últimos $(n-p)$ menores principais da matriz, sendo que, para um mínimo local, é suficiente os menores principais tenham o sinal de $(-1)^p$.

Exemplo 4.2.2. *Considere o problema*

$$\begin{cases} \textit{Minimize} \quad f(x,y,z) = \dfrac{1}{2}(x^2+y^2+z^2) \\ \textit{sujeito a} \quad g(x,y,z) = x+y+z \leq -3. \end{cases}$$

A função lagrangiana do problema acima é

$$\mathbb{L}(\mu,x,y,z) = \frac{1}{2}(x^2+y^2+z^2) + \mu \cdot (x+y+z+3).$$

Temos que verificar a condição de regularidade para o caso onde a restrição esteja ativa, isto é, quando $x+y+z=-3$, note que

$$\nabla g(x,y,z) = (1,1,1),$$

deste modo temos $\nabla g(x,y,z) \neq (0,0,0)$ para todos os pontos onde g está ativa. As condições necessárias de primeira ordem são dadas por

$$\begin{cases} x+\mu = 0, \\ y+\mu = 0, \\ z+\mu = 0, \\ \mu \cdot (x+y+z+3) = 0, \\ \mu \geq 0, \\ x+y+z \leq -3. \end{cases}$$

Se $\mu = 0$, então $x=y=z=0$ e o ponto $(0,0,0,0)$ é um candidato a solução do problema. Porém, observe que, quando $\mu = 0$ temos que $x+y+z > -3$, portanto não podemos ter $x=y=z=0$, e este ponto é excluído.

Agora, resolvendo o problema para $\mu > 0$ chegamos a conclusão que $x=y=z=-1$ e $\mu = 1$, logo, $(-1,-1,-1,1)$ é o único ponto que satisfaz as condições de primeira ordem do problema.

A Hessiana orlada do problema é dada por

$$\bar{H}(x,y,z,\mu) = \begin{bmatrix} 0 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix},$$

e $|\bar{H}| = -3 < 0$, *deste modo, o ponto é um mínimo local do problema.*

Para um problema com n restrições e p desigualdades, os determinantes da Hessiana orlada serão conforme (4.38), sendo que, para um mínimo local, é suficiente os últimos $(n-p)$ menores principais tenham o sinal de $(-1)^p$.

Exemplo 4.2.3. *(27) Considere o seguinte problema de otimização*

$$\begin{cases} min & f(x,y) = -(x^2+y^2) \\ s.\ a & g_1(x,y) = 2x+y \leq 2 \\ & g_2(x,y) = -x \leq 0 \\ & g_3(x,y) = -y \leq 0 \end{cases} \tag{4.39}$$

Vamos verificar a condição de regularidade para o caso onde temos restrições ativas.

1. *Para uma restrição ativa, temos*

   - *se* $g_1 = 2$, *então* $\nabla g_1(x,y) = (2,1)$;
   - *se* $g_2 = 0$, *então* $\nabla g_1(x,y) = (-1,0)$;
   - *se* $g_3 = 0$, *então* $\nabla g_1(x,y) = (0,-1)$.

2. *Para duas restrições ativas, temos*

   - *se* $g_1 = 2$, *e* $g_2 = 0$, *então*

   $$\begin{bmatrix} \nabla g_1(x,y) \\ \nabla g_2(x,y) \end{bmatrix} = \begin{matrix} (2,1) \\ (-1,0) \end{matrix} \sim \begin{matrix} (2,1) \\ \left(0, \frac{1}{2}\right) \end{matrix}$$

   que tem posto 2;

   - *se* $g_1 = 2$, *e* $g_3 = 0$, *então*

   $$\begin{bmatrix} \nabla g_1(x,y) \\ \nabla g_3(x,y) \end{bmatrix} = \begin{matrix} (2,1) \\ (0,-1) \end{matrix}$$

   que tem posto 2;

   - *se* $g_2 = 0$ *e* $g_3 = 0$

   $$\begin{bmatrix} \nabla g_2(x,y) \\ \nabla g_3(x,y) \end{bmatrix} = \begin{matrix} (-1,0) \\ (0,-1) \end{matrix}$$

   que tem posto 2;

3. *note que não é possível que as três restrições estejam ativas, pois teríamos* $2x+y=2$ *com* $x=0$ *e* $y=0$, *um absurdo.*

Desta maneira, todos os pontos da região viável satisfazem as condições de regularidade. A função lagrangiana associada ao problema é dada por

$$\mathbb{L} = -(x^2+y^2)+\mu_1(2x+y-2)+\mu_2(-x)+\mu_3(-y) = -(x^2+y^2)+\mu_1(2x+y-2)-\mu_2 x-\mu_3 y$$

O sistema associado as condições de primeira ordem é

$$\begin{cases} \dfrac{\partial \mathbb{L}}{\partial x} = -2x + 2\mu_1 - \mu_2 = 0 & (1) \\ \dfrac{\partial \mathbb{L}}{\partial y} = -2y + \mu_1 - \mu_3 = 0 & (2) \\ \mu_1(2x + y - 2) = 0 & (3) \\ \mu_2 x = 0 & (4) \\ \mu_3 y = 0 & (5) \\ \mu_1 \geq 0,\ \mu_2 \geq 0,\ \mu_3 \geq 0. \end{cases}$$

Vamos resolver o sistema acima, de $\mu_2 x = 0$ *temos* $x = 0$ *ou* $\mu_2 = 0$*, se* $x = 0$ *então de* (1) *e* (3) *temos* $\mu_2 = 2\mu_1$ *e* $\mu_1(y-2) = 0$ *e assim temos* $y = 2$ *ou* $\mu_1 = 0$*. Se* $y = 2$ *então* $\mu_3 = 0$*, substituindo esses valores em* (2) *obtemos* $\mu_1 = 4$ *e finalmente concluímos que* $\mu_2 = 8$*, portanto o ponto* $(0, 2, 4, 8, 0)$ *é uma solução do sistema.*

Da igualdade $\mu_1(y-2) = 0$*, caso* $\mu_1 = 0$*, então encontramos o ponto* $(0, 0, 0, 0, 0)$ *como solução.*

De (5)*, se* $y = 0$*, então* $\mu_3 = \mu_1$ *e* $\mu_1(2x-2) = 0$ *o que implica* $\mu_1 = 0$ *ou* $x = 1$*, se considerarmos* $\mu_1 = 0$ *encontramos a solução* $(0, 0, 0, 0, 0)$ *novamente. Para* $x = 1$*, temos* $(1, 0, 1, 0, 1)$ *como solução.*

Por fim, se $x \neq 0$ *e* $y \neq 0$ *então* $\mu_2 = \mu_3 = 0$*, de* (1) *e* (2) *obtemos* $x = \mu_1$ *e* $y = \dfrac{\mu_1}{2}$*. Substituindo em* (3) *encontramos* $\mu_1 = \dfrac{4}{5}$*, deste modo, o ponto* $\left(\dfrac{4}{5}, \dfrac{8}{5}, \dfrac{4}{5}, 0, 0\right)$ *também é uma solução.*

A condição de primeira ordem apresenta quatro pontos como possíveis minimizadores do problema de otimização,

$$(0,0,0,0,0);\ (0,2,4,8,0);\ (1,0,1,0,1) \text{ e } \left(\frac{4}{5}, \frac{8}{5}, \frac{4}{5}, 0, 0\right).$$

Utilizaremos a Hessiana orlada para classificar a natureza dos pontos encontrados. A matriz associada ao problema é dada por

$$\bar{H} = \begin{bmatrix} 0 & 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & -2 & 0 \\ 1 & 0 & 1 & 0 & -2 \end{bmatrix}$$

como o problema apresenta três restrições e duas variáveis, precisamos calcular apenas o último menor princpal da matriz, nesse caso, temos $|\bar{H}| = -2$*, além disso* $(-1)^p = (-1)^3 = -1$*, e ambos os números possuem o mesmo sinal, logo podemos concluir que os pontos encontrados são mínimos locais do problema.*

Substituindo os pontos

$$(0,0),\ (0,2),\ (1,0)\ e\ \left(\frac{4}{5}, \frac{8}{5}\right)$$

na função objetivo entramos

$$f(0,0) = 0;$$
$$f(0,2) = -4;$$
$$f(1,0) = -1;$$
$$f\left(\frac{4}{3}, \frac{2}{3}\right) = -\frac{48}{25}.$$

E finalmente, podemos concluir que o ponto $(0,2)$ *é um mínimo global.*

4.3 OTIMIZAÇÃO COM RESTRIÇÕES DE IGUALDADE E DESIGUALDADE

O caso geral de um problema de minimização com restrições de igualdade e desigualdade, também chamado de restrições mistas, reúne toda a teoria das seções anteriores. Considere o seguinte problema de otimização

$$(P) = \begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & x \in \Omega \, \{x \in \mathbb{R}^n \mid h_i(x) = 0,\ g_j(x) \leq 0\} \\ \text{com} & i = 1, 2, \cdots, m, \\ & j = 1, 2, \cdots, p. \end{cases} \tag{4.40}$$

Onde $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ e $g : \mathbb{R}^n \to \mathbb{R}^p$, com $h(x) = (h_1(x), h_2(x), ..., h_m(x))$ e $g(x) = (g_1(x), g_2(x), ..., g_p(x))$ são no mínimo de classe $C^2(\mathbb{R}^n)$.

4.3.1 Condição necessária de 1ª ordem

A condição necessária de primeira ordem para problemas com restrições mistas, também conhecidas como condições de Karush-Kuhn-Turcker, reúne os Teoremas 4.1.2 e 4.2.1.

Teorema 4.3.1 (Condições de Karush-Kuhn-Turcker (KKT)). *Sejam f, g e h funções de classe C^1 e seja $\bar{x} \in \Omega$ um minimizador do problema de otimização. Caso alguma restrição esteja ativa em $\bar{x}$, vamos renomeá-las de forma que sejam as l primeiras, ou seja, $g_1, \cdots, g_l, \cdots g_p$. Suponha que $\bar{x}$ seja um ponto regular, isto é, o posto da matriz jacobiana*

$$\begin{bmatrix} \nabla h_1(\bar{x}) \\ \nabla h_2(\bar{x}) \\ \vdots \\ \nabla h_m(\bar{x}) \\ \nabla g_1(\bar{x}) \\ \nabla g_2(\bar{x}) \\ \vdots \\ \nabla g_l(\bar{x}) \end{bmatrix}_{(m+l)\times n}$$

é igual a $m + l$, onde m são as restrições de igualdade e l o número de restrições ativas em $\bar{x}$.

Então existem $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)$ e $\mu = (\mu_1, \mu_2, \cdots, \mu_p)$ multiplicadores de lagrange tais que

$$\begin{cases} \nabla f(\bar{x}) + \lambda \cdot \nabla h(\bar{x}) + \mu \cdot \nabla g(\bar{x}) = 0, \\ h(\bar{x}) = 0 \\ \mu \cdot g(\bar{x}) = 0 \\ \mu \geq 0 \\ g(\bar{x}) \leq 0. \end{cases} \tag{4.41}$$

Podemos escrever esse sistema, de modo equivalente, através da função lagrangiana

$$\mathbb{L}(\lambda, \mu, x) = f(x) + \lambda \cdot h(x) + \mu \cdot g(x). \tag{4.42}$$

O sistema toma a seguinte forma

$$\begin{cases} \dfrac{\partial \mathbb{L}}{\partial x_1}(\lambda,\mu,\bar{x}) = 0, ..., \dfrac{\partial \mathbb{L}}{\partial x_n}(\lambda,\mu,\bar{x}) = 0, \\ \dfrac{\partial \mathbb{L}}{\partial \lambda_1}(\lambda,\mu,\bar{x}) = 0, ..., \dfrac{\partial \mathbb{L}}{\partial \lambda_m}(\lambda,\mu,\bar{x}) = 0, \\ \dfrac{\partial \mathbb{L}}{\partial \mu_1}(\lambda,\mu,\bar{x}) \geq 0, ..., \dfrac{\partial \mathbb{L}}{\partial \mu_p}(\lambda,\mu,\bar{x}) \geq 0, \\ \mu \cdot g(\bar{x}) = 0, \\ \mu \geq 0. \end{cases} \tag{4.43}$$

Note que, os problemas de otimização com restrições de igualdade ou desigualdade são casos particulares do caso acima.

Exemplo 4.3.1. *Considere o seguinte problema de otimização*

$$\begin{cases} \textit{minimizar} \quad f(x,y) = -x + y^2 \\ \textit{sujeito a} \quad h(x,y) = x^2 + y^2 = 4 \\ \qquad\qquad g_1(x,y) = -x \leq 0 \\ \qquad\qquad g_2(x,y) = -y \leq 0 \end{cases}$$

Vamos verificar os casos de regularidade dos pontos que satisfazem as restrições ativas. Para uma única restrição ativa, temos um caso único:

$$h(x,y) = 4 \textit{ onde } \nabla h(x,y) = (2x, 2y) = (0,0) \iff x = y = 0, \textit{ mas como}$$
$$h(x,y) = x^2 + y^2 = 4, \textit{ temos que } x \textit{ e } y \textit{ não podem ser simultaneamente nulos.}$$

Como a restrição h está sempre ativa não existem pontos no conjunto viável onde as restrições g_1 e g_2 sejam as únicas restrições ativas. Para o caso onde há duas restrições ativas, temos, $h(x,y) = 4$ e $g_1(x,y) = 0$, com $y \geq 0$, deste modo o único ponto que satisfaz a condição $h(x,y) = 4$ é o ponto $(0,2)$. Sendo assim, a matriz

$$\begin{bmatrix} \nabla h_1(0,2) \\ \nabla g_1(0,2) \end{bmatrix} = \begin{matrix} (0,4) \\ (-1,0) \end{matrix} \sim \begin{matrix} (-1,0) \\ (0,4) \end{matrix}$$

tem posto igual a 2.

Outro caso para duas restrições ativas é $h(x,y) = 4$ e $g_2(x,y) = 0$, com $x \geq 0$, e o único ponto que satisfaz a condição $h(x,y) = 4$ é o ponto $(2,0)$. Daí

$$\begin{bmatrix} \nabla h_1(2,0) \\ \nabla g_1(2,0) \end{bmatrix} = \begin{matrix} (4,0) \\ (0,-1) \end{matrix}$$

tem posto igual a 2.

Finalmente, não podemos ter o caso onde há três restrições ativas, de fato, pois se $h(x,y) = 4$, $g_1(x,y) = 0$ e $g_2(x,y) = 0$ para algum ponto (x,y), então teríamos $x = 0$, $y = 0$ e $x^2 + y^2 = 4$, o que é um absurdo.

As condições de KKT do problema são dados por

$$\begin{cases} -1 - 2\lambda_1 x + \mu_1 = 0, \\ 2y - 2\lambda_1 y + \mu_2 = 0, \\ x^2 + y^2 = 4, \\ \mu_1 x = 0, \\ \mu_2 y = 0, \\ \mu_1,\ \mu_2 \geq 0, \\ x,\ y \geq 0. \end{cases}$$

O sistema acima possui uma única solução, quando $\lambda_1 = 1/4$, $\mu_1 = \mu_2 = 0$, logo o ponto obtido é $(x,y) = (2,0)$. O teorema de Karush-Kuhn-Tucker garante que qualquer solução de um problema de otimização deve ser solução do sistema de equações associados às condições de primeira ordem. Logo, como a função é contínua, a região viável é um conjunto compacto e existe um único ponto que satisfaz as condições de primeira ordem. Este é a solução do problema de otimização.

4.3.2 Condição necessária e suficiente de 2ª ordem

Da mesma forma como nas seções anteriores, vamos apresentar uma condição de segunda ordem para problemas com restrições mistas, através do estudo da Hessiana do problema.

Teorema 4.3.2 (Condição necessária de 2ª ordem)**.** *Uma condição necessária para que o ponto regular $\bar{x} \in \Omega$ seja mínimo local de f, onde $\bar{x}$ satisfaz a condição de primeira ordem é que a matriz*

$$H(\lambda, \mu, \bar{x}) = Hf(\bar{x}) + \lambda \cdot Hh(\bar{x}) + \mu \cdot Hg(\bar{x}), \tag{4.44}$$

seja ***positiva semidefinida*** *em $T_{\bar{x}}$, ou seja, $v \cdot H(\lambda, \mu, \bar{x})v \geq 0$, $\forall v \in T_{\bar{x}}$.*

Como foi visto no decorrer do capítulo, a condição suficiente de segunda ordem para problemas com restrições mistas faz uso da função lagrangiana e da matriz Hessiana orlada associada ao problema, que nada mais é que a combinação das funções 4.20 e 4.37 e das matrizes 4.21 e 4.38.

$$\mathbb{L} = f(x) + \sum_{i=1}^{m} \lambda_i \cdot h^i + \sum_{j=1}^{p} \mu_j \cdot g^j \tag{4.45}$$

onde m são as restrições de igualdade e p o número de restrições ativa para o problema com restrições de desigualdade.

$$|\bar{H}| = \begin{vmatrix} 0 & ... & 0 & | & h_1^1 & ... & h_n^1 & g_1^1 & ... & g_n^1 \\ 0 & ... & 0 & | & h_1^2 & ... & h_n^2 & g_1^2 & ... & g_n^2 \\ - & - & - & - & - & - & - & - & - & - \\ 0 & ... & 0 & | & h_1^m & ... & h_n^m & g_1^p & ... & g_n^p \\ - & - & - & - & - & - & - & - & - & - \\ h_1^1 & \cdots & h_n^1 & g_1^1 & \cdots & g_n^1 & | & \mathbb{L}_{11} & \cdots & \mathbb{L}_{1n} \\ h_2^1 & \cdots & h_2^m & g_2^1 & \cdots & g_2^p & | & \mathbb{L}_{21} & \cdots & \mathbb{L}_{2n} \\ - & - & - & - & - & - & - & - & - & - \\ h_n^1 & \cdots & h_n^m & g_n^1 & \cdots & g_n^p & | & \mathbb{L}_{n1} & \cdots & \mathbb{L}_{nn} \end{vmatrix} \tag{4.46}$$

Note que o Hessiana orlado foi subdividido em quatro blocos, o bloco superior esquerdo é composto apenas por 0, o bloco inferior direito contém o hessiano simples, os outros dois blocos, contendo as derivadas parciais das restrições possuem uma relação de imagem especular, fazendo referência a diagonal principal, resultando em um arranjo simétrico entre os elementos da hessiana orlada.

- **Condição suficiente de 2ª ordem**

A condição suficiente de segunda ordem para o determinante acima é semelhante aos casos anteriores, dado em termo dos sinais dos últimos $n - (m + p)$ menores principais da matriz, sendo que, para um mínimo local, é suficiente que os menores principais tenham todos o mesmo sinal, a saber $(-1)^{(m+p)}$. Note que, um número par ou ímpar de restrições no problema faz grande diferença, uma vez que (-1) elevado a uma potência par é positivo, para o caso de uma potência ímpar, o sinal será mantido.

As condições de segunda ordem estabelecidas nesse capítulo possibilitam analisar a natureza de um ponto crítico, que satisfaz a condição de primeira ordem, indicando se este é um ponto de máximo ou mínimo local, porém nada podemos afirmar sobre a globalidade do ponto.

Para questões de globalidade, a convexidade desempenha papel fundamental, sendo possível apresentar um teorema que garante a globalidade do ponto.

Teorema 4.3.3. *Sejam f, g e h funções de classe $C^2(\mathbb{R}^n)$, definidas em um aberto convexo $U \in \mathbb{R}^n$. Considere o problema de minimizar $f(x)$ sujeito a $\Omega = \{x \in \mathbb{R}^n \,|\, h(x) = 0, g(x) \leq 0\}$ e suponha que $(\bar{\lambda}, \bar{\mu}, \bar{x})$ satisfaça as condições de 1^a ordem. Se*

- *a função f é côncava em U;*
- *as funções $h_1, ..., h_m$ são lineares;*
- *as funções $g_1, ..., g_p$ são convexas em U;*

então $(\bar{\lambda}, \bar{\mu}, \bar{x})$ é um mínimo global de f em Ω.

5 CONDIÇÕES DE OTIMALIDADE APLICADAS EM ALGORITMOS

Conforme visto no capítulo anterior, as condições de KKT são fundamentais para o estudo de problemas de otimização não linear com restrições. Essas condições generalizam os multiplicadores de Lagrange e permitem lidar tanto com restrições de igualdade quanto de desigualdade. Nesse capítulo vamos estudar algoritmos que utilizam as condições de KKT, veremos que essas condições formam a base de muitos métodos numéricos para otimização convexa e não convexa.

5.1 MÉTODOS DE DIREÇÕES VIÁVEIS

Os métodos de direções viáveis são métodos que, em geral, resolvem problemas de programação não linear, movendo-se de um ponto factível para outro ponto factível melhorado.

Consideremos o seguinte problema de otimização

$$(P)\begin{cases}\text{Minimizar} \quad f(x)\\ x \in \mathbb{R}^n\end{cases}$$

onde f é no mínimo de classe C^2. A estratégia dos métodos de direções viáveis tem em comum uma iteração que consiste nos seguintes passos:

- Passo 1: Para um ponto x^k, uma direção de descida estrita $d^k \in \mathbb{R}^n$ de f no ponto x^k é determinada;
- Passo 2: Determine um $\lambda_k \in \mathbb{R}$ tal que

$$f(x^k + \lambda_k d^k) < f(x^k),$$

  (geralmente temos $\lambda_k > 0$).
- Passo 3: Definir $x^{k+1} = x^k + \lambda_k d^k$.

O valor λ_k é chamado de tamanho do passo, e sua determinação é chamada de busca linear na direção d^k. Quando λ_k é dobtido resolvendo o problema

$$\begin{cases}\min \quad f(x^k + \lambda d^k)\\ \lambda \in \mathbb{R}_+,\end{cases}$$

temos uma busca linear exata.

Os métodos apresentados a seguir são de extrema importância para a área da programação, especialmente otimização. Além disso, é frequentemente demonstrado na literatura que métodos desse tipo geralmente convergem para pontos KKT.

5.1.1 Método do gradiente e Método de Cauchy

Uma direção de descida estrita $d^k \in \mathbb{R}^n$ de f no ponto x^k é obtida escolhendo

$$d^k = -\nabla f(x^k) \neq 0. \tag{5.1}$$

Este vetor pode ser interpretado como a direção de descida máxima. A prova para essa afirmação pode ser encontrada em (30).

Se λ_k for obtido através da busca linear exata, então temos o método do gradiente clássico. Os passos do algoritmo estão definidos a seguir

(i) Escolha $x^0 \in \mathbb{R}^n$ e defina $k = 0$;

(ii) Calcular $d^k = -\nabla f(x^k) \neq 0$;

(iii) Se $d^k = 0$ (ou $|d^k| \leq \epsilon$), pare. Caso contrário, vá para o próximo passo;

(iv) Determine uma solução λ_k do problema $\min_{\lambda \in \mathbb{R}_+} f(x^k + \lambda d^k)$;

(v) Defina $x^{k+1} = x^k + \lambda_k d^k,\ k = k + 1$ e retorne ao passo (ii).

Note que a condição de parada $d^k = 0$ pode ser interpretada como a condição de primeira ordem.

O Método de Cauchy (11) é um método utilizado para problema de otimização em $\mathbb{R}^n$. O método é um processo que a cada iteração realiza uma busca na direção oposta do vetor gradiente da função objetivo.

Os passos do algoritmo são descritos a seguir:

- **Método de Cauchy**

  (i) Dado: $x_0 \in \mathbb{R}^n,\ k = 0$;

  (ii) REPITA enquanto $\nabla f(x_k) \neq 0$;

  (iii) Defina $d = -\nabla f(x_k)$;

  (iv) Obtenha $t_k > 0$ tal que $f(x_k + t_k d_k) < f(x_k)$;

  (v) Faça $x_{k+1} = x_k + t_k d_k$

  (vi) Defina $k = k + 1$.

A diferença entre o Método de Cauchy e o Método do gradiente está no passo (iv), onde que, no algortimo de Cauchy realiza uma busca linear exata, o trabalho (11) mostra que o algoritmo de Cauchy, com o tamanho do passo t_k obtido no passo (iv), é globalmente convergente, ou seja, para qualquer sequência (x_k) e qualquer ponto $\bar{x}$ de (x_k), obtém-se que $\bar{x}$ é estacionário, isto é, $\bar{x}$ é ponto crítico.

5.1.2 Método multidimensional de Newton

O Método clássico de Newton (30), pode ser generalizado e utilizado em problemas de otimização sem restrições. Para que isso possa ser feito, basta substituir a primeira e segunda derivadas, $f'(x^k)$ e $f''(x^k)$, pelo gradiente de f, no ponto x^k, isto é, $\nabla f(x^k)$ e pela hessiana $Hf(x^k)$, respectivamente.

Essas substituições resultam na seguinte fórmula de recorrência

$$x^{k+1} = x^k - (Hf(x^k))^{-1}\nabla f(x^k), \tag{5.2}$$

onde $f \in C^2(\mathbb{R}^n)$. Os passos do Método são descritos a seguir

(i) Escolha $x^0 \in \mathbb{R}^n$ e defina $k = 0$;

(ii) Calcular $g^k = -\nabla f(x^k)$;

(iii) Se $g^k = 0$ (ou $|g^k| \leq \epsilon$), pare. Caso contrário, vá para o próximo passo;

(iv) Defina $x^{k+1} = x^k - (Hf(x^k))^{-1}g^k$, $k = k + 1$ e retorne ao passo (ii).

Novamente, o critério de parada utilizado no algoritmo é $d^k = 0$. Para o Método multidimensional de Newton, não é necessário o cálculo da inversa da hessiana de f no passo (iv), o ponto x^{k+1} é obtido pela igualdade $x^{k+1} = x^k + d^k$, onde d^k é solução do sistema de equações

$$Hf(x^k)d = -g^k. \tag{5.3}$$

O algoritmo utiliza a hessiana da função objetivo, no capítulo anterior apresentamos as condições necessárias e suficientes de segunda ordem, que dependem da hessiana da função lagrangeana, porém, a aplicação de métodos computacionais que utilizam a hessiana podem ser tornar inviáveis, dependendo da quantidade de variáveis do problema, uma vez que o cálculo da hessiana e a resolução de sistemas do tipo $Hf(x^k)d = -g^k$ exigem um trabalho computacional considerável.

Se a matriz $Hf(x^k)$ for positiva definida para todo k o método de Newton é um método descendente com direção $d^k = -(Hf(x^k))^{-1}\nabla f(x^k)$ e passo constante, de tamanho $\lambda_k = 1$. Caso λ_k seja determinado conforme a seção anterior, obtemos o Método multidimensional com busca linear exata, que é semelhante ao anterior e possui o mesmo critério de parada

(i) Escolha $x^0 \in \mathbb{R}^n$ e defina $k = 0$;

(ii) Calcular $g^k = -\nabla f(x^k)$;

(iii) Se $g^k = 0$ (ou $|g^k| \leq \epsilon$), pare. Caso contrário, vá para o próximo passo;

(iv) Calcule d^k (solução do sistema 5.3);

(v) Determine uma solução λ_k do problema $\min_{\lambda \in \mathbb{R}_+} f(x^k + \lambda d^k)$;

(vi) Defina $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ e retorne ao passo (ii).

Existem outras variações do método de Newton, como, por exemplo, substituir a matriz hessiana do passo (iv) por uma matriz definida positiva simétrica $B_k = Hf(x^k) + \sigma I$ com $\sigma > 0$. Essa mudança resulta no Método de Newton modificado, que não será abordado neste trabalho, uma vez que a ideia central é apresentar a aplicação dos resultados do Capítulo 4 em algoritmos.

5.1.3 Direções conjugadas e Método quasi-Newton

Nas duas seções anteriores apresentamos dois métodos utilizados para otimização sem restrições, o Método de Newton, geralmente, é um método que converge de forma mais rápida que o método dos gradientes, porém o método utiliza a matriz hessiana de f, além disso, precisa encontrar a solução do sistema 5.3, o que requer bastante trabalho computacional. O método do gradiente, apesar de convergir de forma mais lenta, tem como vantagem exigir apenas as primeiras derivadas parciais da função objetivo.

Os dois métodos, Direções conjugadas e Métodos quasi-Newton, são procedimentos que buscam combinar as vantagens de ambos. A construção matemática de cada algoritmo pode ser encontrada em (30). Como estamos tratando de algoritmos clássicos da literatura, acreditamos que apresentar tais construções nesse trabalho seria mero exercício de repetição.

- **Direções conjugadas**

    (i) Escolha $x^0 \in \mathbb{R}^n$ e calcule $g^0 = \nabla f(x^0)$;

    (ii) Se $g^0 = 0$, (ou $|g^0| \leq \epsilon$) pare. Caso contrário defina $d^0 = -g^0$, $k = 0$ e vá para o próximo passo;

    (iii) Determine λ_k por meio da busca linear exata;

    (iv) Defina $x^{k+1} = x^k + \lambda_k d^k$;

    (v) Calcule $g^{k+1} = \nabla f(x^{k+1})$;

    (vi) Se $g^{k+1} = 0$ (ou $|g^{k+1}| \leq \epsilon$) pare. Caso contrário vá para o próximo passo;

    (vii) Se $k < n - 1$, então vá para o passo (viii). Se $k = n - 1$, então vá para o passo (x);

    (viii) Calcule $\beta_k = \dfrac{g^{(k+1)T} g^{k+1}}{g^{kT} g^k}$;

(ix) Defina $d^{k+1} = -g^{k+1} + \beta_k d^k$, $k = k+1$ e vá para o passo (iii);

(x) Defina $x^0 = x^n$, $d^0 = -g^n$, $k = 0$ e vá para o passo (iii).

- **Método quasi-Newton**

  O método quasi-Newton é um procedimento da forma

$$x^0 \in \mathbb{R}^n,\ x^{k+1} = x^k + \lambda_k d^k, \tag{5.4}$$

onde λ_k é determinado por busca linear exata. A direção de busca é obtida calculando $d^k = -M_k g^k$, onde $M \in \mathbb{R}^{n\times n}$ é uma matriz simétrica positiva definida que deve ser calculada em cada passo. Os passos para construção da matriz M estão disponíveis em (30). Como M_k é uma matriz positiva definida, então d^k é uma direção descendente. O algoritmo quasi-Newton é semelhante ao método de Newton com busca linear exata, para cada n passos do algoritmo quasi-Newton, temos, aproximadamente um passo do método de Newton, deste modo, a condição de parada no passo (ii) é análoga para ambos os casos, e por fim este é mais um algoritmo clássico da literatura que utiliza as condições de otimalidade vistas no capítulo anterior.

5.1.4 Método DFP - Davidon, Fletcher e Powell

O método DFP, apresentado por Davidon, Fletcher e Powell, é um procedimento que combina os métodos quasi-Newton e de direções conjugadas. A estrutura básica do método é idêntica à do método de direção conjugada. O algoritmo realiza buscas nas direções

$$d^0, d^1, \cdots, d^{n-1}$$

onde $d^0 = -\nabla f(x^0)$ e $d^k = M_k g^k$ para $k \geq 1$. Quando um ciclo de n iterações é concluído, um novo ciclo é iniciado, definindo $x^0 = x^n$.

Podemos ressaltar mais alguns aspectos importantes sobre os métodos apresentados até o momento. Conforme dito anteriormente, o método do gradiente converge para a solução de forma mais lenta, porém é mais estável e possui aplicação computacional mais fácil, quando comparado com o método de Newton. Por sua vez, o método de Newton converge para a solução de forma mais rápida, mas exige extenso trabalho computacional por iteração, por conta da necessidade do uso da matriz hessiana. Para o caso onde a função possui muitas variáveis, o método torna-se inviável.

Vimos no capítulo anterior, as Condições necessárias e suficientes de segunda ordem para problemas de otimização irrestritos ou com restrições, depende diretamente do cálculo da hessiana da função objetivo ou da função lagrangiana, esses resultados teóricos são fundamentais para a área, mas conforme dito, tanto seu cálculo analítico, quanto o cálculo numérico podem ser extremamente complexos. Em testes numéricos extensivos, o método desenvolvido por Broyden, Fletcher, Goldfarb e Shannon (BFGS), disponível em (21), apresenta bons resultados. O método tem como base o método DFP, onde, para tornar o método mais estável, é realizada uma substituição no décimo passo do algoritmo DFP, novamente, o critério de parada não é alterado.

5.2 MÉTODO DE ZOUTENDIJK

Vamos apresentar o Método de direções viáveis de Zoutendijk (1, 13), a cada iteração o método gera uma nova direção de busca melhorada, e então otimiza ao longo dessa direção.

Inicialmente, consideremos o caso onde a região viável é definida por um sistema de restrições lineares, deste modo, o problema de otimização será do tipo

$$\begin{cases} \text{Minimizar } \ f(x) \\ \qquad \text{s. a. } \ Ax \leq b \\ \qquad\qquad Qx = q, \end{cases} \tag{5.5}$$

onde A é uma matriz $m \times n$, Q é uma matriz $l \times n$, b e q são vetores. Para este caso, d é uma direção viável melhorada se $A_1 d \leq 0$, $Qd = 0$ e $\nabla f(x)^t d < 0$. A prova dos resultados e todos os detalhes do Método de Zoutendijk estão disponíveis em (1).

Dado um ponto viável x, e para um vetor diferente de zero, um método para encontrar uma direção viável melhorada d, é minimizar $\nabla f(x)^t d < 0$ sujeito às restrições $A_1 d \leq 0$ e $Qd = 0$. Entretanto, se $\bar{d}$ é um vetor tal que $\nabla f(x)^t \bar{d} < 0$, $A_1 \bar{d} \leq 0$ e $Q\bar{d} = 0$ aconteça, então o valor objetivo ótimo do problema anterior é $-\infty$, considerando $\lambda \bar{d}$ onde $\lambda \to \infty$. Assim, uma restrição que limita o vetor d ou a função objetivo deve ser introduzida. A seguir são apresentados três problemas para gerar uma direção viável melhorada.

$$P_1 \begin{cases} \text{Minimizar } \ \nabla f(x)^t d < 0 \\ \qquad \text{s. a. } \ A_1 d \leq 0 \\ \qquad\qquad Qd = 0 \\ \qquad\qquad -1 \leq d_j \leq 1 \text{ para } j = 1, \cdots, n. \end{cases} \tag{5.6}$$

$$P_2 \begin{cases} \text{Minimizar } \ \nabla f(x)^t d < 0 \\ \qquad \text{s. a. } \ A_1 d \leq 0 \\ \qquad\qquad Qd = 0 \\ \qquad\qquad d^t d \leq 1. \end{cases} \tag{5.7}$$

$$P_3 \begin{cases} \text{Minimizar } \nabla f(x)^t d < 0 \\ \text{s. a. } A_1 d \leq 0 \\ Qd = 0 \\ \nabla f(x)^t d \geq -1. \end{cases} \tag{5.8}$$

Como $d = 0$ é solução viável para cada um dos problemas acima, e como seu valor objetivo é igual a 0, o valor ótimo dos problemas P_1, P_2 e P_3 não pode ser positivo. Se o valor mínimo da função objetivo dos problemas acima for negativo, então uma direção viável melhorada é encontrada. Por outro lado, se o valor mínimo da função objetivo for igual a zero, então x é um ponto KKT.

O algoritmo de Zoutendijk para minimizar funções diferenciáveis com restrições lineares é dado do seguinte modo:

- **Método de Zoutendijk**

(i) Etapa de inicialização: Encontre uma solução viável x_1 com $Ax_1 \leq b$ e $Qx_1 = q$. Seja $k = 1$ e vá para a Etapa principal;

(ii) Etapa principal: Dado x_k, suponha que A^t e b^t são decompostos em (A_1^t, A_2^t) e (b_1^t, b_2^t), de modo que $A_1 x_k = b_1$ e $A_2 x_k < b_2$. Seja d uma solução ótima para o seguinte problema (os problemas P_2 ou P_3 poderiam ser usados)

$$\begin{cases} \text{Minimizar } \nabla f(x)^t d < 0 \\ \text{s. a. } A_1 d \leq 0 \\ Qd = 0 \\ -1 \leq d_j \leq 1 \text{ para } j = 1, \cdots, n. \end{cases}$$

Se $\nabla f(x_k)^t d_k = 0$, pare; x_k é um ponto KKT, com as variáveis duais do problema anterior fornecendo os multiplicadores de lagrange correspondentes. Caso contrário, vá para o próximo passo.

(iii) Seja λ_k uma solução ótima para o seguinte problema de busca linear

$$\begin{aligned} \text{Minimizar} \quad & f(x_k + \lambda d_k) \\ \text{s. a} \quad & 0 \leq \lambda \leq \lambda_{max}, \end{aligned}$$

(para determinar $\lambda_{max} ver(1)$). Seja $x_{k+1} = x_k + \lambda_k d_k$, um novo conjunto de restrições em x_{k+1} e atualize A_1 e A_2. Substitua k por $k+1$ e vá para o passo (II).

O Método de Zoutendijk para problema contendo restrições não lineares não envolve as condições de KKT, mas as condições de otimalidade de Fritz John, que não foram abordados neste trabalho, deste modo, tal método não será abordado.

5.3 PROGRAMAÇÃO LINEAR SUCESSIVA

Os algoritmos apresentados anteriormente vimos que a cada iteração é resolvido um problema de programação linear de busca de direção com base em aproximações de primeira ordem, e em seguida é realizada uma busca nessa direção. Conceitualmente, isso é semelhante às abordagens de programa linear sucessivo (Successive Linear Program SLP). A cada iteração k, um programa linear de busca de direção é formulado, com base em aproximações de primeira ordem para as funções objetivo e de restrição, além de limites de passo apropriados ou restrições de região de confiança nos componentes de direção (1). Se $d_k = 0$ resolve o problema, então a iteração x_k é ótima para a aproximação de primeira ordem. Além disso, essa solução é um ponto KKT e encerramos o procedimento. Caso contrário, faça uma nova iteração tomando $x_{k+1} = x_k + d_k$ ou rejeita esta iteração e reduz os limites do passo, e então o processo. A decisão sobre aceitar ou rejeitar a nova iteração é tipicamente feita com base em uma função de mérito moldada em torno da função de penalidade l_i, que é dada por

$$F_E(x) = f(x) + \mu \cdot \left[\sum_{i=1}^{m} \max\{0, g_i(x)\} + \sum_{i=1}^{l} |h_i(x)| \right]. \tag{5.9}$$

A abordagem foi introduzida por Griffith e Stewart da Shell Development Company em 1961, é amplamente utilizada nas indústrias de petróleo e química. A vantagem desse tipo de método é sua facilidade e robustez na implementação para problemas de larga escala.

O Algoritmo de Programação Linear Sucessiva de Penalidade (PSLP) é um método SLP que emprega a função de penalidade l_i no problema de busca de direção em si, e possui boas propriedades de robustez e convergência. O problema a ser considerado para esse algoritmo é do tipo:

$$\begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & h_i(x) = 0, \quad i = 1, \cdots l, \\ & g_j(x) \leq 0, \quad j = 1, \cdots m, \\ & x \in \Omega = \{x : Ax \leq b\}, \end{cases} \tag{5.10}$$

onde todas as funções são continuamente diferenciáveis, $x \in \mathbb{R}^n$ e as restrições lineares que definem o problema foram todas acomodadas no conjunto Ω.

A análise matemática por trás do método é extensa e acreditamos que escreve-la neste texto foge do objetivo do capítulo. Sugerimos ao leitor que consulte (1).

O principal conceito que desejamos destacar nesse método é a utilização da condição suficiente de segunda ordem, apresentada no **Teorema 10.3.1 item a** do livro (1). Até o momento, os algoritmos apresentados utilizam, em sua maioria, as condições de primeira ordem de KKT. É importante ressaltar que, mesmo com sua complexidade computacional, alguns algoritmos utilizam os resultados de segunda ordem apresentados no capítulo anterior.

Existem outros métodos que têm como base o algoritmo SLP e utilizam as condições de KKT, como, por exemplo a **Programação Quadrática Sucessiva SQP ou Abordagem Lagrangeana Projetada, Método de Rosen o Método de Zangwill** entre outros, para mais detalhes sobre esses métodos ver (1, 13).

5.4 ALGORITMO PENLAB

O algoritmo PENLAB (8, 12) é um pacote de software de código aberto para otimização linear, não linear, e otimização semidefinida não linear. O algoritmo é escrito inteiramente em MATLAB e permite que os usuários editem e o desenvolvam ainda mais. Particularmente, ele é adequado para fins de ensino e pesquisa e para testar novas ideias algorítmicas.

O algoritmo tem como principal objetivo resolver problemas de otimização não linear, sujeito a restrições de desigualdade e igualdade não lineares e desigualdades matriciais não lineares.

O problema pode ser formulado da seguinte maneira:

$$\begin{cases} \text{Minimizar} & f(x,Y), \quad \text{com} \quad x \in \mathbb{R}^n, Y_1 \in S^{p1}, \cdots, Y_k \in S^{pk}, \\ \text{sujeito a} & g_i(x,Y) \leq 0, \quad i = 1, \cdots, m_g \\ & h_i(x,Y) = 0, \quad i = 1, \cdots, m_h \\ & A_i(x,Y) \preceq 0, \quad i = 1, \cdots, m_A \\ & \lambda_i I \preceq Y_i \preceq \bar{\lambda}_i I, \quad i = 1, \cdots, k \end{cases} \tag{5.11}$$

onde $x \in \mathbb{R}^n$ é o vetor de variáveis, $Y_1 \in S^{p1}, \cdots, Y_k \in S^{pk}$ são as vairáveis matriciais, denotando-se por $Y = [Y_1, \cdots Y_k]$, f_i, g_i e h_i são funções de classe C^2 de $\mathbb{R}^n \times S^{p1} \times \cdots \times S^{pk} \to \mathbb{R}$ e $\lambda_i, \bar{\lambda}_i$ são os limites inferior e superior, respectivamente, dos autovalores de Y_i, $A_i(x,Y)$ são os mapeamentos duas vezes diferenciáveis de $\mathbb{R}^n \times S^{p1} \times \cdots \times S^{pk} \to S^{pA_i}$, sendo pA_i para $i = 1, \cdots, m_A$ inteiros positivos.

O algoritmo PENLAB combina ideias dos métodos de penalidade (exterior) e de barreira (interior) com o método do Lagrangiano aumentado. O algoritmo também utiliza os multiplicadores de lagrange clássicos que vimos no capítulo anterior, sendo que no primeiro passo do algoritmo pretende-se encontrar a solução de um sistema linear do tipo

$$\nabla_{x\,Y} \mathbb{L}(x, Y, \mu, \Xi, U, \bar{U}, v, \pi, \Pi) = 0 \tag{5.12}$$

$$h(x,Y) = 0 \tag{5.13}$$

onde os parâmetros $\mu, \Xi, U, \bar{U}$ são multiplicadores de lagrange fixos e π, Π são parâmetros de penalidade. Note que o sistema acima é análogo às condições de primeira ordem apresentadas anteriormente. Para a construção matemática do método e informações para download do software, ver (8, 12).

5.5 MÉTODO DE PONTOS INTERIORES DE NEWTON

O trabalho (10) apresenta uma formulação e teoria do Método de pontos interiores de Newton. Em um primeiro momento, o trabalho detalha o método de ponto interior primal-dual para problemas lineares e faz algumas comparações com o Método de Newton amortecido aplicado às condições de KKT. O objetivo central do artigo é transferir da programação linear uma formulação viável de um método de ponto interior para os problemas de programação não linear.

Os autores ressaltam que, para atingir esse objetivo, as condições de KKT perturbadas e o método de função de barreira logarítimica são essenciais. Deste modo, nas Seções 2, 3 e 4, o artigo apresenta os problemas de programação linear e não linear, as condições de KKT para cada caso e a formulação das condições de KKT perturbadas, além disso, alguns algoritmos são descritos, como, por exemplo, o método de Newton padrão e o método de Ponto interior.

Na Seção 5 é demonstrada a convergência local do problema, já a convergência global é apresentada na Seção 6. É importante ressaltar que o principal Teorema do trabalho afirma que, sobre certas hipóteses, é claro, o algoritmo converge para um ponto KKT do problema de otimização. As aplicações numéricas são descritas na Seção 7 de (10).

Conforme dito no decorrer deste capítulo, as condições de otimalidade apresentadas no capítulo anterior são fundamentais na otimização, seja para problemas restritos ou irrestritos. Novamente, não vimos uma aplicação direta das condições de segunda ordem em um método numérico, reforçando que tais condições são de difícil aplicação numérica, devido à sua complexidade computacional.

5.6 ALGORITMO DE DIREÇÕES VIÁVEIS E PONTOS INTERIORES PARA PROBLEMAS COM RESTRIÇÕES - FDIPA

Nessa seção, vamos apresentar o FDIPA (18) para problemas de otimização não linear com restrições de desigualdade, isto é, o problema (P) será do tipo

$$(P)\begin{cases}\text{Minimizar} \quad f(x)\\ \text{sujeito a} \quad x \in \Omega = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0, \}\\ \text{com} \quad j = 1, 2, ...p.\end{cases} \tag{5.14}$$

Onde $f : \mathbb{R}^n \to \mathbb{R}$ e $g : \mathbb{R}^n \to \mathbb{R}^p$, $g(x) = (g_1(x), g_2(x), ..., g_p(x))$, são no mínimo de classe $C^2(\mathbb{R}^n)$.

A função lagrangeana associada ao problema é

$$\mathbb{L}(\mu, x) = f(x) + \mu \cdot g(x), \tag{5.15}$$

cuja matriz hessiana é

$$H(\mu, \bar{x}) = Hf(\bar{x}) + \mu \cdot Hg(\bar{x}). \tag{5.16}$$

A condição de KKT para o problema (P), onde $\bar{x}$ é um ponto regular, é dado por

$$\begin{cases} \nabla f(\bar{x}) + \mu \cdot \nabla g(\bar{x}) = 0, \quad (1) \\ \mu \cdot G(\bar{x}) = 0, \quad (2) \\ \mu \geq 0, \quad (3) \\ g(\bar{x}) \leq 0, \quad (4) \end{cases} \tag{5.17}$$

onde $G(\bar{x})$ é uma matriz diagonal tal que $G(\bar{x})_{ii} = g_i(\bar{x})$. O FDIPA é um método que converge globalmente para os pontos de KKT, o trabalho (18) mostra que o método, além de ser eficiente, é de fácil aplicação, robusto e eficiente.

A partir de um ponto inicial, que está no interior da região viável, o algoritmo apresentado gera sequências $\{x^k\}$ de pontos, com valores decrescentes quando avaliados na função objetivo, e convergindo para um conjunto de pontos KKT do problema.

A cada iteração, uma direção de busca d é determinada resolvendo um sistema de equações, que é uma direção de descida para a função objetivo e também uma direção viável de Ω. Uma busca linear é então realizada para garantir que o novo ponto seja interior e o objetivo, seja mais baixo.

O sistema de equações (1) e (2) é resolvido em $(\bar{x}, \mu)$ por meio de iterações de ponto fixo, de tal que forma que a cada iteração as equações (3) e (4) sejam sempre satisfeitas, dessa forma sempre são obtidas as soluções de KKT do problema.

Uma iteração de Newton para a solução de (1) e (2) começando em $(\bar{x}, \mu)$ define uma nova estimativa, $(\bar{x}_\alpha, \mu_\alpha)$, resolvendo o seguinte problema

$$\begin{bmatrix} B & \nabla g(x) \\ \Lambda \nabla g^T(x) & G(x) \end{bmatrix} \begin{bmatrix} \bar{x}_\alpha - \bar{x} \\ \mu_\alpha - \mu \end{bmatrix} = - \begin{bmatrix} \nabla f(\bar{x}) + \mu \cdot \nabla g(\bar{x}) \\ \mu \cdot G(\bar{x}) \end{bmatrix}, \tag{5.18}$$

onde $B = H(\bar{x}, \mu)$ e Λ é uma matriz diagonal tal que $\Lambda_{ii} = \mu_i$. É importante ressaltar que a matriz B deve ser positiva definida para a garantia da convergência global do problema. O sistema 5.18 é modificado, com o intuito de se obter uma melhor estimativa para o ponto $(\bar{x}, \mu)$, deste modo uma nova direção é definida

$$d_\alpha = x_\alpha - x. \tag{5.19}$$

Assim, o sistema 5.18 é reescrito da seguinte maneira

$$Bd_\alpha + \mu_\alpha \nabla g(x) = -\nabla f(x) \tag{5.20}$$

$$\Lambda g^T(x) d_\alpha + \mu_{\alpha i} G(x) = 0, \tag{5.21}$$

que fornece uma nova direção de descida e uma nova estimativa para μ. Das equações 5.20 e 5.21, obtém-se que $d_\alpha = 0$ em um ponto crítico. Entretanto, d_α pode não ser uma direção viável, uma vez que, à medida que qualquer restrição tende a zero, a equação 5.21 tende a uma direção tangente da região viável. Note que a equação 5.21 pode ser reescrita da seguinte maneira

$$\mu_i \nabla g_i^T(x) d_\alpha + \mu_{\alpha i} g_i(x) = 0, \quad ,1,...,p \tag{5.22}$$

o que implica que

$$\mu_i \nabla g_i^T(x) d_\alpha = 0,$$

para todo i tal que $g_i(x) = 0$.

Para contornar esse problema, define-se o seguinte sistema linear em d de $\bar{\mu}$:

$$Bd + \bar{\mu} \nabla g(x) = -\nabla f(x), \tag{5.23}$$

$$\Lambda g^T(x) d + \bar{\mu} G(x) = -\rho \Lambda \omega, \tag{5.24}$$

obtido pela adição de um vetor negativo do lado direito da igualdade de 5.21, onde os escalares ρ e $\omega \in \mathbb{R}^p$ são positivos e $\bar{\mu}$ é uma nova estimativa de μ. Agora, a equação 5.24 é equivalente a

$$\mu_i \nabla g_i^T(x) d + \bar{\mu}_i g_i(x) = -\rho \mu_i \omega_i, \quad ,1,...,p$$

e

$$\nabla g_i^T(x) d = -\rho \omega_i < 0,$$

para as restrições ativas, assim d é uma direção viável.

O termo $-\rho\Lambda\omega$ produz uma deflexão em d_α na direção do interior de Ω. Finalmente, como a deflexão de d_α é proporcional a ρ e como d_α é uma direção de descida, ao estabelecer limites superiores para ρ, é possível garantir que d também seja uma direção de descida. Como $d_\alpha^T \nabla f(x) < 0$ os limites são obtidos impondo

$$d^T \nabla f(x) \leq \varepsilon d_\alpha^T \nabla f(x) \quad \varepsilon \in (0,1), \tag{5.25}$$

o que implica $d^T\nabla f(x) < 0$. A condição 5.25.

Observe que d pode ser obtido resolvendo o sistema

$$Bd_\beta + \mu_\beta \nabla g(x) = 0, \tag{5.26}$$

$$\Lambda \nabla g^T(x) d_\beta + \mu_\beta G(x) = -\Lambda\omega, \tag{5.27}$$

e calculando $d = d_\alpha + \rho d_\beta$. Para que $d(x)$ seja um campo de direção uniformemente viável, ρ também deve ser limitado por baixo. As ideias mostradas acima formam uma base para os métodos iterativos que estamos estudando.

A seguir, apresentaremos o algoritmo base para problemas de otimização com restrições de desigualdade que converge globalmente para os pontos de KKT do problema. O algoritmo é escrito da seguinte forma:

Algoritmo 5.6.1. *(FDIPA)*

Paramêtros

$$\varepsilon \in (0,1),\ \eta \in (0,1),\ \varphi > 0 \text{ e } \nu \in (0,1).$$

Dados iniciais

$$\bar{x} \in \Omega^0, 0 < \mu \in \mathbb{R}^p,\ B \in \mathbb{R}^{n\times n} \text{ simétrica, positiva definida, e } 0 < \omega \in \mathbb{R}^p.$$

- *Passo 1: Cálculo de uma direção de busca*

  (i) Calcule (d_α, μ_α) resolvendo o sistema linear

$$Bd_\alpha + \mu_\alpha \nabla g(x) = -\nabla f(x) \tag{5.28}$$

$$\Lambda g^T(x) d_\alpha + \mu_{\alpha i} G(x) = 0, \tag{5.29}$$

$$\text{Se } d_\alpha = 0, \text{ pare.}$$

  (ii) Calcule (d_β, μ_β) resolvendo o sistema linear

$$Bd_\beta + \mu_\beta \nabla g(x) = 0, \tag{5.30}$$

$$\Lambda \nabla g^T(x) d_\beta + \mu_\beta G(x) = -\Lambda\omega, \tag{5.31}$$

  (iii) Se $d_\beta^T \nabla f(x) > 0$, então

$$\rho = min\{\varphi \|d_\alpha\|_2^2; (\varepsilon - 1) d_\alpha^T \nabla f(x) / d_\beta^T \nabla f(x)\},$$

  caso contrário

$$\rho = \varphi \|d_\alpha\|_2^2.$$

  (iv) Calcule $d = d_\alpha + \rho d_\beta$ e $\bar{\mu} = \mu_\alpha + \rho\mu_\beta$

- *Passo 2: Encontre o primeiro termo* t *da sequência* $\{1, v, v^2, v^3, \cdots\}$ *que satisfaz*

$$f(x + td) \leq f(x) + t\eta\nabla f^T(x)d$$

e

$$g_i(x + td) < 0, \quad se \quad \bar{\mu}_i \geq 0,$$

$$g_i(x + td) < g_i(x), \quad se \quad \bar{\mu}_i < 0.$$

- *Passo 3: Atualizações*

  (i) Defina $x = x + td$ *e defina novos valores para* $\omega > 0$ *e* $\mu > 0$ *e* B *simétrica e positiva definida.*

  (ii) Vá para o Passo 1.

A prova de convergência, a inclusão de restrições de igualdade e testes numéricos são encontrados em (18).

O algoritmo FDIPA é bastante utilizado em problemas de otimização com restrições, vimos que o método utilizada das condições de KKT apresentadas no capítulo anterior para definir a direção de busca, além disso, podemos notar também que as condições necessárias e suficientes de segunda ordem não são aplicadas nesse algoritmo.

O trabalho (8) apresenta variações do algoritmo FDIPA, o FDIPA-SDP, FDIPA-SDP-NS e FDIPA-GSDP, que utilizam as condições de KKT para determinar a direção de busca. Além disso, os algoritmos utilizam o mesmo critério de parada do passo 1.

5.7 MÉTODO DE FUNÇÃO DE BARREIRA

O Método de função de barreira é uma abordagem de otimização não linear que, em geral, transforma um problema restrito em um problema irrestrito, ou em uma sequência de problemas irrestritos. Essas funções estabelecem uma barreira para dificultar a aproximação da solução da fronteira da região viável (Ω). A ideia principal é manter a solução sempre no interior de Ω a cada processo iterativo. O problema de barreira será formulado a seguir, para isso considere o seguinte problema de otimização

$$(P)\begin{cases}\text{Minimizar} \quad f(x)\\ \text{sujeito a} \quad x \in \Omega = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0,\}\\ \text{com} \quad j = 1, 2, ...p.\end{cases} \tag{5.32}$$

Onde $f : \mathbb{R}^n \to \mathbb{R}$ e $g : \mathbb{R}^n \to \mathbb{R}^p$, $g(x) = (g_1(x), g_2(x), ..., g_p(x))$, são no mínimo de classe $C^2(\mathbb{R}^n)$.

Problema de Barreira

Encontrar:

$$\begin{cases} Inf\ \theta(\mu) \\ \text{sujeito a } \mu > 0 \end{cases} \tag{5.33}$$

onde $\theta(\mu) = inf\{f(x) + \mu B(x) \ : \ g(x) < 0,\ x \in \Omega\}$. Nesse caso, B é uma função de barreira não negativa e contínua na região $\{x : g(x) < 0\}$, e se aproxima de ∞ à medida que o limite da região $\{x : g(x) \leq 0\}$ se aproxima do interior. A função de barreira B é definida do seguinte modo

$$B(x) = \sum_{i=1}^{m} \phi[g_i(x)], \tag{5.34}$$

onde ϕ é uma função univariada que é contínua sobre $\{y : y < 0\}$ e satisfaz

$$\phi(y) \geq 0, \text{ se } y < 0, \quad \text{e} \quad \lim_{y \to 0^-} \phi(y) = \infty. \tag{5.35}$$

Para mais detalhes e exemplos, ver (1).

Os autores de (1) mostram que, sobre certas condições, inclusive a de regularidade que apresentamos no capítulo anterior, o Método de função de barreira produz uma sequência de estimativas de Multiplicadores de Lagrange que convergem para um conjunto ótimo de multiplicadores.

A seguir apresentaremos os passos do algoritmo utilizando a função de barreira 5.34 para o problema de programação não linear (P).

- **Método de função de barreira**

(i) Etapa de inicialização: Seja $\epsilon > 0$, um escalar de terminação, e escolha um ponto $x_1 \in \Omega$ com $g(x_1) < 0$. Seja $\mu_1 > 0$, $\beta \in (0, 1)$, defina $k = 1$ e vá para a etapa principal

(ii) Etapa principal: Começando com x_k, resolva o seguinte problema:

$$\text{Minimize } f(x) + \mu_k B(x)$$

$$\text{sujeito a } g(x) < 0$$

$$x \in \Omega.$$

Seja x_{k+1} uma solução ótima e vá para o próximo passo.

(iii) Se $\mu_k B(x_{k+1}) < \epsilon$, pare. Caso contrário, faça $\mu_{k+1} = \beta\mu_k$, $k = k + 1$, e vá para o passo (ii).

O método de função de barreira para resolver problemas de otimização não linear restritos possui algumas dificuldades computacionais. Inicialmente, a busca deve começar com um ponto $x \in \Omega$ com $g(x) < 0$, como já vimos, encontrar tal ponto pode não ser uma tarefa fácil, por fim, devido a estrutura da função de barreira B, e para valores pequenos de μ, a maioria das técnicas de busca pode encontrar problemas de condicionamento e dificuldades com erros de arredondamento ao resolver o problema de minimizar $f(x) + \mu B(x)$. Para mais detalhes, ver (1).

5.8 MÉTODO DE REESCALONAMENTO NÃO LINEAR BASEADO NA FUNÇÃO BARREIRA LOGARÍTMICA MODIFICADA

Nessa seção apresentamos o Método de Reescalonamento não linear baseado na função logarítmica modificada (26). O método de reescalonamento baseado na função barreira logarítmica modificada que apresentamos a seguir é desenvolvido no trabalho (26) e tem como base os métodos da Lagrangiana aumentada não quadrática.

Os métodos de reescalonamento visam transformar o problema de otimização original, reescalonando a função objetivo e/ou as restrições e assim obter outro problema, equivalente ao problema inicial, que possui a mesma solução.

A transformação não linear realizada depende de um parâmetro positivo, e baseia-se em uma função suave, denominada função de reescalonamento. Os seguintes passos fazem parte do método de reescalonamento: minimização irrestrita da função Lagrangiana clássica do problema equivalente e atualizações das variáveis primais e duais, ou seja, os Multiplicadores de Lagrange.

A função de reescalonamento citada acima pertence a uma classe de funções suaves (Ψ) e satisfaz as seguintes propriedades:

1. $\Psi(0) = 0, \ \Psi'(0) = 1,$
2. $\Psi'(t) > 0,$
3. $\Psi''(t) < 0,$
4. $\lim_{t \to +\infty} \Psi'(t) = 0.$

Onde $\Psi \in C^2$ definida no intervalo $(a, +\infty)$, $-\infty \leq a < 0$, tal que $\Psi(a) = -\infty$ e $\Psi'(a) = +\infty$.

Para cada função Ψ definida, origina-se um tipo de método de reescalonamento não linear. O método que apresentado no trabalho (26) utiliza a função definida da seguinte maneira

$$\Psi(t) = \ln(t+1), \quad \text{onde} \quad \Psi : \mathbb{R} \to \mathbb{R}. \tag{5.36}$$

O método considera o seguinte problema de otimização

$$\begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & h_i(x) = 0, \quad i = 1, K, m \\ & g_j(x) \geq 0, \quad j = 1, K, r. \end{cases} \tag{5.37}$$

Onde $f : \mathbb{R}^n \to \mathbb{R}$, $h_i : \mathbb{R}^n \to \mathbb{R}$ e $g_j : \mathbb{R}^n \to \mathbb{R}$, são no mínimo de classe $C^1(\mathbb{R}^n)$.

Para iniciar a construção do algoritmo é preciso obter o problema equivalente de 5.37. Inicialmente, são inseridas as variáveis de excesso s_j nas restrições de desigualdade no problema, obtendo:

$$\begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & h_i(x) = 0, \quad i = 1, K, m \\ & g_j(x) - s_j = 0, \quad j = 1, K, r \\ & s_j \geq 0, \quad j = 1, K, r. \end{cases} \tag{5.38}$$

A seguir, o sistema 5.38 é reescalonado e parametrizado, mediante um parâmetro de relaxação positivo c, as restrições de desigualdade do problema 5.38 resultando no seguinte problema equivalente:

$$\begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & h_i(x) = 0, \quad i = 1, K, m \\ & g_j(x) - s_j = 0, \quad j = 1, K, r \\ & c^{-1} \ln(cs_j + 1) \geq 0 \quad j = 1, K, r. \end{cases} \tag{5.39}$$

O trabalho mostra que as alterações realizadas no problema original não alteram a região viável.

O primeiro passo do algoritmo é minimizar o problema 5.39 considerando μ fixo, isso é feito através das condições necessárias de primeira ordem, que para este problema podem ser escritas da seguinte maneira

$$\nabla \mathbb{L}(x, s, \lambda, \pi, \mu) = 0. \tag{5.40}$$

Com a solução do sistema obtem-se as direções que atualizarão x, s, λ e π. Para os multiplicadores de lagrange μ_j, a seguinte regra é utilizada

$$\mu_j^{k+1} = \mu_j^k / (c^k s_j^k + 1), \quad j = 1, K, r.$$

O algoritmo é constituído por dois *loops*, sendo eles, o loop externo e o loop interno. O loop interno possui os seguintes passos: resolução do sistema 5.40, determinação do tamanho do passo e atualização das variáveis. Já o loop externo controla a convergência global do problema através das condições de KKT para o problema 5.38, além disso, o loop é responsável por atualizar o multiplicador μ^k.

Sejam os pontos $x^k, s^k, \lambda^k, \pi^k$ e μ^k obtidos na k-ésima iteração do algoritmo, as condições de KKT utilizadas como critério de parada para o loop externo são

- $\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k h_i(x^k) + \sum_{j=1}^{r} \pi_j^k g_j(x^k) - \sum_{j=1}^{r} \mu_j^k s_j^k = 0$
- $h_i(x^k) = 0, \quad i = 1, K, m$

  $g_j(x^k) - s_j^k = 0 \quad j = 1, K, r$

  $s_j^k \geq 0 \quad j = 1, K, r$
- $\mu_j^k \geq 0 \quad j = 1, ..., r$
- $\mu_j^k g_j(x^k) \geq 0 \quad j = 1, ..., r.$

O algoritmo finaliza as condições de KKT forem satisfeitas. Para mais detalhes do loop interno e aplicações numéricas do algoritmo ver (26).

O algoritmo apresentado nessa seção utiliza fortemente as condições de KKT vistas no capítulo anterior, assim como o algoritmo FDIPA, as condições de segunda ordem não aplicadas no contexto computacional.

5.9 FERRAMENTAS COMPUTACIONAIS

No decorrer das seções anteriores, estudamos alguns métodos numéricos que utilizam, de algum modo, os resultados estudados no capítulo anterior. O uso de ferramentas computacionais é fundamental para a implementação de tais métodos, além disso, existem alguns softwares que também fazem uso dos resultados vistos anteriormente. Softwares como o MATLAB e bibliotecas de código aberto como o SciPy em Python desempenham um papel essencial ao fornecer plataformas robustas e flexíveis para otimização. Nesta seção vamos apresentar alguns software de grande utilidade na programação.

5.9.1 FMINCON - MATLAB

O fmincon (find minimum of constrained nonlinear multivariable function (23)) é uma função do MATLAB, que está integrado na parte do Optimization Toolbox, usada para minimização de funções não lineares com várias variáveis, sujeitas a restrições. É projetada para resolver problemas de programação não linear, considerando restrições de igualdade e desigualdade, bem como limites superiores/inferiores para as variáveis.

Como um software de aplicação para solução de problemas de otimização não linear com restrições, o FMINCON utiliza algoritmos que, na prática, buscam soluções que satisfaçam as condições de KKT. Ele faz isso internamente, calculando derivadas, Lagrangianos e trabalhando com multiplicadores de Lagrange de maneira implícita. O software possui vários algoritmos em sua base de dados, e conforme o método de otimização escolhido, o FMINCON resolve uma série de subproblemas de otimização que aproximam as condições de KKT.

Entre os métodos disponíveis no software que estão diretamente ligados aos resultados apresentados no capítulo anterior podemos destacar: fmincon Trust Region Reflective Algorithm, fmincon Active Set Algorithm e fmincon Interior Point Algorithm que trabalha diretamente com as condições de segunda ordem associada a função lagrangeana, deste modo, destacamos o seguinte passo do algoritmo, denominado Passo Direto

- **Passo Direto**

$$H = \nabla^2 f(x) + \sum_i \lambda_i \nabla^2 g_i(x) + \sum_j \mu_j \nabla^2 h_j(x). \tag{5.41}$$

Conforme podemos ver em (23), por padrão, o algoritmo primeiro tenta dar um passo direto, pois possui uma convergência mais rápida, porém esse é o passo computacionalmente mais complexo e custoso.

5.9.2 CVX - MATLAB

O sistema CVX é um software desenvolvido na plataforma MATLAB e tem como finalidade a construção e solução de problemas de otimização convexa disciplinada. O CVX tem suporte para uma série de tipos de problemas padrão, incluindo programas lineares e quadráticos, programas de cone de segunda ordem, programas semidefinidos e programação não linear restritos, além disso, o CVX também pode resolver problemas de otimização convexa mais complexos, como, por exemplo, problemas que envolvam funções não diferenciáveis.

Um ponto importante que destacamos no software CVX é que, em termos de função lagrangeana, todo problema de otimização convexa quando é resolvido pelo CVX tem automaticamente o seu problema dual resolvido. Além disso, o software possui comandos específicos a fim de declarar as variáveis duais, ou multiplicadores de lagrange. Deste modo, o software garante a solução de um problema de otimização convexa através das condições KKT.

Para aplicações do software em problemas de otimização, ver (15, 16).

5.9.3 SciPy - optimize

O SciPy (Scientific Python) é um software de código aberto para a linguagem de programação Python, utilizada em áreas como matemática e engenharia. O software fornece funções e algoritmos avançados para computação, permitindo o desenvolvimento de aplicações complexas em áreas como álgebra linear, otimização, integração, estatísticas e resolução de equações diferenciais.

O SciPy possui um módulo específico para a área de otimização, chamada SciPy - optimize, esse módulo fornece algoritmos para resolver problemas de otimização não lineares (com suporte para algoritmos de otimização local e global), programação linear, mínimos quadrados restritos e não lineares, descoberta de raiz e ajuste de curva. Entre os métodos disponíveis no software podemos destacar o Método multidimensional de Newton, Direções conjugadas e Método quasi-Newton, que estudamos nesse capítulo.

Existem vários métodos numéricos que podemos utilizar em problemas de otimização, e seria impossível citar todos nesse texto. Estamos interessados em métodos que, de algum modo, utilizam os resultados teóricos do capítulo anterior em algum passo da iteração. Dito isto, é importante destacar que existem algoritmos que não utilizam o que foi nenhum resultado apresentado, como, por exemplo o **Método de busca de Coordenadas Cíclicas**, que não utiliza as derivadas da função objetivo, e os **Algoritmos Genéticos (AG)**, que se diferem dos algoritmos de otimização clássicos, em geral, os algoritmos genéticos trabalham com uma codificação do conjunto de parâmetros, com um espaço de busca onde estão todas as possíveis soluções do problema. Além disso os AGs utilizam como critério de parada um número pré-determinado de passos nas iterações. Os AG são bastante eficientes na busca de soluções ótimas em uma grande variedade de problemas, porém não utilizam de nenhuma das condições de otimalidade deste trabalho. Para mais informações sobre Algoritmos Genéticos, ver (14, 19).

6 CONCLUSÃO

Apresentamos neste trabalho as condições de otimalidade para problemas de otimização restritos e irrestritos. Vimos que, para problemas de otimização sem restrições com funções de uma variável, uma condição necessária para que o ponto seja um minimizador do problema é que este, seja um ponto crítico, vimos que é possível classificar a natureza deste ponto crítico com a segunda derivada, que chamamos de condição de segunda ordem, em que, basta que a derivada segunda seja positiva, para se ter um mínimo. Para o caso onde a segunda derivada é nula, o teste é inconclusivo. Para funções de várias variáveis, o teste de segunda ordem é feito pelo estudo da matriz Hessiana de f, no qual, para um mínimo local, é suficiente que a matriz seja positiva definida. Tais resultados, apresentados no Capítulo 3, apesar de simples, são úteis para o entendimento do texto que se seguiu, uma vez que, conforme visto, a ideia central da teoria dos multiplicadores de Lagrange é transformar um problema com n variáveis e m restrições em um problema com $n + m$ variáveis e irrestrito.

Transformar um problema restrito em um problema irrestrito pode ser feito com o auxílio da função lagrangiana clássica, que combina a função objetivo e as restrições do problema em uma única expressão. Ao calcular as derivadas parciais de primeira ordem dessa nova função, encontramos um sistema de equações associado às condições de primeira ordem, onde os pontos que são soluções do sistema são pontos críticos da lagrangiana, e assim, possíveis minimizadores do problema original. Outro fator importante ligado a função lagrangiana é a possibilidade de classificação dos pontos críticos através do determinante Hessiano Orlado. No Capítulo 4 apresentamos uma maneira de se obter esse determinante, ressaltando a importância do cálculo para a análise dos pontos obtidos na resolução do sistema de primeira ordem. Além disso, não encontramos na literatura uma apresentação mais explícita e organizada do cálculo do hessiano orlado.

O método dos multiplicadores de Lagrange, na teoria, mostra-se bastante eficaz para encontrar os extremos de uma função, uma vez que as restrições estão bem definidas e o método sempre segue um conjunto fixo de passos matemáticos, em um processo bem estruturado. Obviamente, o método depende da diferenciabilidade das funções e da existência de soluções viáveis. Se essas condições forem satisfeitas, o procedimento e o resultado serão determinados sem aleatoriedade, diferentemente de métodos estocásticos ou heurísticos.

Apesar de ser uma ferramenta para o cálculo das soluções de um problema de otimização e fornecer a possibilidade da análise teórica dos problemas, para situações onde são apresentados problemas com muitas variáveis e restrições, o sistema de equações derivado da lagrangiana pode se tornar extremamente complexo, exigindo técnicas numéricas avançadas, o que pode ser computacionalmente inviável, além disso, o método exige que as funções sejam no mínimo C^2.

Finalmente, como discutido no Capítulo 5, algoritmos como o Método do Gradiente, Método de Cauchy, Método de Zoutendijk, o FDIPA e suas variações, entre outros, utilizam em seus passos as condições de primeira ordem para a resolução de problemas de programação não linear. Esses métodos se destacam pela eficiência na convergência e pela menor complexidade computacional. No entanto, as condições de segunda ordem, apesar de serem capazes de fornecer resultados mais precisos e robustos, sua aplicação é menos frequente, devido ao seu elevado custo computacional. Por fim, a aplicação de métodos que se baseiam nessas condições é mais apropriada para problemas onde o tempo de processamento não é um elemento crucial, mas ainda enfrenta desafios significativos em contextos que exigem soluções rápidas e de baixo custo computacional.

REFERÊNCIAS

1 BAZAARA, M.; SHERALI, H. D.; SHETTY, C. M. **Nonlinear Programming:** Theory and Algorithms 3. ed. New Jersey: Winley-Interscience, 2006.

2 BERNARDELLO, A et al. **Matemática para Economistas** con Microsoft Excel y Matlab. 1. ed. Buenos Aires: Omicron System, 2004.

3 BERTSEKAS, P., D.; **Nonlinear Programming** Second Edtion. Athena Scientific. Belmont, Massachusetts, 1995.

4 BERTSIMAS, D.; TSITSIKLIS J, N. **Introduction to Linear Optimization** 4. ed. Athena Scientific. Belmont, Massachusetts, 1997

5 BORTOLOSSI, H. J. **Cálculo Diferencial a Várias Variáveis.** Uma Introdução à Teoria de Otimização Ed. PUC-Rio. São Paulo, Loyola 2002.

6 CHEN, Y.**A duality theory with zero Gap for Nonlinear Programming.** Department of Computer Science and Engineering, Washington University in St. Louis. Saint Louis, 2007.

7 CHIANG, A. C., WAINWRIGHT, K. **Matemática para Economistas.** Elsevier Editora: Rio de Janeiro, 2005.

8 CÓRDOVA, Elmer Giovanni Bazán. **ALGORITMOS NUMÉRICOS PARA PROGRAMAÇÃO SEMIDEFINIDA E APLICAÇÕES EM OTIMIZAÇÃO ESTRUTURAL**. Tese (Doutorado em Engenharia Mecânica) Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia Mecânica, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

9 DONÁ, de D. **Lagrangian Duality**. The Unisersity of Newcastle, 2006.

10 EL-BAKRY, A., S.; TAPIA, R., A.; TSUCHIYA, T.; ZHANG, Y. On the Formulation and Theory of the Newton Interior-Point Method for Nonlinear Programming. JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 89, No. 3, pp. 507-541, 1996

11 FERRAZ, B., A. **Métodos computacionais de otimização**. Ponta Grossa: Atena Editora, 2021.

12 FIALA, J. KOCVARA, M. STINGL, M. **PENLAB: A MATLAB solver for nonlinear semidefinite optimization**. Mathematical Programming manuscript, ArXiv e-prints, 2013.

13 FRITZSCHE, H. **Programação não-linear:** análise e métodos. Edgard Blucher Ltda - Universidade de São Paulo, São Paulo, 1978.

14 GOLDBERG, D., E., **Genetic Algorithms in Search Optimization and Machine Learning. John Wiley & Sons**. United States of America, 1997.

15 GRANT, M.; BOYD, S.; CVX: Matlab software for disciplined convex programming, Disponível: https://cvxr.com/cvx/doc/index.html

16 GUIMARÃES, D. A.; FLORIANO, G. H. F.; CHAVES, L. S.; **A Tutorial on the CVX System for Modeling and Solving Convex Optimization Problems**. IEEE LATIN AMERICA TRANSACTIONS, VOL. 13, NO. 5, MAY 2015.

17 HERNANDO, José Rocha Franco. **Método de Direções Interiores ao Epígrafo - IED para Otimização não Diferenciável e não Convexa via Dualidade Lagrangeana:** Estratégias para Minimização da Lagrangeana Aumentada. 2018. Tese (Doutorado em Modelagem Computacional) - Programa de Pós-graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora, Minas Gerais, 2018.

18 HERSKOVITS, J., **Feasible Direction Interior-Point Technique for Nonlinear Optimization**. JOTA- Journal of Optimization Theory and Applications, v. 99, pp. 121–146, 1998.

19 HOLLAND, J., H., **Adaptation in Natural and Artficial Systems**.University of Michigan Press, 1975.

20 JERROLD E., TROMBA, A. **Vector calculus.** W. H. Freeman and Company Publishers: Nova Iorque, 2012.

21 LUENBERGER, D., G., **Linear and Nonlinear Programming**. Reading, Massachusetts: Addison-Wesley Publishing Company, 2003.

22 MAHEY, P. **Programação Não-Linear.** Introdução à Teoria e aos Métodos. Rio de Janeiro: Campus, 1987.

23 MathWorks. fmincon. Disponível em http://www.mathworks.com/help/optim/ug/fmincon.html.

24 RIBEIRO, A., KARAS, E., W., **Otimização Contínua:** aspectos teóricos computacionais. Cengage Learning, São Paulo, Brazil, 2013.

25 ROCKAFELLAR, R., T., **Convex analysis**. Princiton university press, 2015.

26 SILVA, I. C. R., JÚNIOR, S. A., **RESULTADOS NUMÉRICOS DO MÉTODO DE REESCALONAMENTO NÃO LINEAR BASEADO NA FUNÇÃO BARREIRA LOGARÍTMICA MODIFICADA**. In. Simpósio Brasileiro de Pesquisa Operacional, XLII., 2010, Rio Grande do Sul. Bento Gonçalves: Universidade Federal de Santa Maria.

27 SIMON, P. C., BLUME, L. **Mathematics for Economists.** W. W. Norton & Company, Inc. London, 1994.

28 STEWART, J. **Cálculo:** Volume 1. 7ª ed. São Paulo: Cengage Learning, 2013.

29 ZÖRNIG, P. **Introdução à programação não linear**. Brasília: UNB, 2011.

30 ZÖRNIG, P. **Nonlinear Programming:** An introduction. Berlin/Boston: Walter de Gruyter, 2014.