

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO DE MATEMÁTICA

MARILA TORRES DE AGUIAR

**OTIMIZAÇÃO IRRESTRITA: ASPECTOS TEÓRICOS E
COMPUTACIONAIS**

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO

2015

MARILA TORRES DE AGUIAR

**OTIMIZAÇÃO IRRESTRITA: ASPECTOS TEÓRICOS E
COMPUTACIONAIS**

Trabalho de Conclusão de Curso apresentado ao Departamento de Matemática da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Licenciada em Matemática”.

Orientadora: Profa. Dra. Elenice Weber Stieglmeier

CORNÉLIO PROCÓPIO

2015

TERMO DE APROVAÇÃO

MARILA TORRES DE AGUIAR

OTIMIZAÇÃO IRRESTRITA: ASPECTOS TEÓRICOS E COMPUTACIONAIS

Trabalho de Conclusão de Curso apresentado ao Departamento de Matemática da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Licenciada em Matemática” – Área de Concentração: Licenciatura em Matemática.

BANCA EXAMINADORA

Prof^a. Dr^a. Elenice Weber Stiegelmeier
Universidade Tecnológica Federal do Paraná

Prof^a. Dr^a. Glaucia Maria Bressan
Universidade Tecnológica Federal do Paraná

Prof^a. Dr^a. Michele Cristina Valentino
Universidade Tecnológica Federal do Paraná

Cornélio procópio, 01 de Junho de 2015.

A Folha de Aprovação assinada encontra-se na Coordenação do Curso.

A minha família por acreditar e apoiar os meus sonhos.

Em especial meus pais que tanto fizeram por mim.

AGRADECIMENTOS

Primeiramente a Deus, por estar presente em todos os momentos da minha vida, por guiar os meus caminhos e me conceder sabedoria e saúde.

A minha família, em especial meus pais Liliana e Aguiar, pelo carinho, amor, suporte emocional e apoio incondicional.

Agradeço as professoras Glaucia Bressan e Michele Valentino por aceitarem ser banca na defesa deste trabalho.

Agradeço profundamente a minha orientadora Elenice Weber Stiegelmeier pela atenção, compreensão, ensinamentos, por toda a paciência, dedicação, carinho, nervosismo produtivo e por ter sido um pouco minha mãe também. Por todas as horas dedicadas nesse período de sua vida para me ensinar, que acreditou que eu era uma pessoa capaz, por me cobrar, questionar, ajudar a conduzir meus objetivos de forma que eu tivesse ainda mais força para alcançá-los e por abrir novas portas para o meu sonho. Agradeço por ter percebido os meus anseios em aprender, desafiar e dar o suporte suficiente para que eu pudesse aprender coisas novas. E ainda, por ter sido um espelho na minha vida em muitos aspectos.

A todos os meus amigos e colegas do curso e de Iniciação Científica pela companhia, amizade, aprendizados e troca de experiências, em especial Nayara Zebediff, Pedro Mazini e Roberta Almeida por terem me oferecido amor, carinho, sorrisos, abraços e porque foram e são parte de minha família ao longo desses anos de convivência. Ao José Fields pela ajuda na realização e ao Lucas Fidelis pelo apoio e atenção durante o desenvolvimento deste trabalho.

RESUMO

AGUIAR, Marila Torres de. OTIMIZAÇÃO IRRESTRITA: ASPECTOS TEÓRICOS E COMPUTACIONAIS. 41 f. Trabalho de Conclusão de Curso – Departamento de Matemática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2015.

Neste trabalho são abordados problemas de otimização em que todas as funções usadas para defini-los são continuamente diferenciáveis e não lineares, ou seja, são problemas de programação não linear (PNL). O caso particular abordado é o problema irrestrito. São estudadas algumas situações que garantem a existência de um minimizador e, em seguida, são discutidas as condições de otimalidade para o problema de otimização irrestrito. Para a obtenção da solução numérica são apresentadas estratégias de resolução baseadas no método iterativo do gradiente.

Palavras-chave: Programação Não Linear. Otimização Irrestrita. Método do Gradiente.

ABSTRACT

AGUIAR, Marila Torres de. UNCONSTRAINED OPTIMIZATION: THEORETICAL AND COMPUTATIONAL ASPECTS. 41 f. Trabalho de Conclusão de Curso – Departamento de Matemática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2015.

In this work, we consider unconstrained optimization problems, with nonlinear continuously differentiable function, that are, Nonlinear Programming Problems (NLPs). The existence of a minimizer is studied at first, next optimization conditions are discussed for the unconstrained optimization problem. Solving strategies are applied, based on the Gradient method, in order to obtain numerical solution.

Keywords: Nonlinear Programming Problem. Unconstrained Optimization. Gradient Method.

LISTA DE FIGURAS

| | | | |
|-----------|---|---|----|
| FIGURA 1 | – | Curvas de nível do problema 2. | 26 |
| FIGURA 2 | – | Curvas de níveis. | 29 |
| FIGURA 3 | – | Exemplo da trajetória “zig-zag” do método do gradiente. | 29 |
| FIGURA 4 | – | (a) Função Parabolóide. (b) Curvas de níveis. | 34 |
| FIGURA 5 | – | (a) Evolução da módulo da função Parabolóide. (b) Evolução do gradiente ... | 35 |
| FIGURA 6 | – | (a) Função Rosenbrock. (b) Curvas de níveis. | 35 |
| FIGURA 7 | – | (a) Variação da função Rosenbrock. (b) Variação da norma do gradiente. | 36 |
| FIGURA 8 | – | (a) Função Freudenstein and Roth. (b) Função de Baele’s | 37 |
| FIGURA 9 | – | (a) Variação da função Freudenstein and Roth. (b)Variação da norma do gradiente. | 38 |
| FIGURA 10 | – | (a) Variação da função Beale’s. (b) Variação da norma do gradiente. | 38 |

LISTA DE TABELAS

| | | |
|----------|---|----|
| TABELA 1 | – Problemas selecionados para análise. | 32 |
| TABELA 2 | – Resultados obtidos com a rotina do gradiente, Algoritmo 1. | 33 |
| TABELA 3 | – Resultados obtidos com a rotina do Matlab <i>fminunc</i> | 33 |
| TABELA 4 | – Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função Parabolóide. | 34 |
| TABELA 5 | – Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função de Rosenbrock. | 36 |
| TABELA 6 | – Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função Freudenstein and Roth. | 37 |
| TABELA 7 | – Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função de Beale's. | 38 |

SUMÁRIO

| | | |
|----------|---------------------------------------|-----------|
| 1 | INTRODUÇÃO | 10 |
| 2 | FERRAMENTAS MATEMÁTICAS | 13 |
| 2.1 | NOÇÕES DE TOPOLOGIA NO \mathbb{R}^N | 13 |
| 2.2 | RESULTADOS DA ÁLGEBRA LINEAR | 14 |
| 2.3 | RESULTADOS DO CÁLCULO DIFERENCIAL | 16 |
| 3 | OTIMIZAÇÃO IRRESTRITA | 19 |
| 3.1 | PROGRAMAÇÃO MATEMÁTICA | 19 |
| 3.2 | OTIMIZAÇÃO IRRESTRITA | 20 |
| 3.2.1 | Existência de soluções | 20 |
| 3.2.2 | Condições de Otimalidade | 21 |
| 3.2.3 | Uma aplicação: funções quadráticas | 25 |
| 4 | MÉTODO DO GRADIENTE | 27 |
| 4.1 | MÉTODO DO GRADIENTE | 27 |
| 4.2 | DETALHES DE IMPLEMENTAÇÃO | 28 |
| 4.2.1 | Método da Seção Áurea | 31 |
| 5 | TESTES COMPUTACIONAIS | 32 |
| 5.1 | RESULTADOS | 32 |
| 6 | CONSIDERAÇÕES FINAIS | 39 |
| | REFERÊNCIAS | 40 |

1 INTRODUÇÃO

Otimização é um tema muito presente no dia a dia acadêmico e industrial. Vários campos da ciência fazem uso das ferramentas de otimização com o objetivo de ajudar na tomada de decisão. Dentre eles, pode-se citar, agricultura, finanças, transporte, processos químicos, produtivos, recursos naturais, ambientais e energéticos, entre outros (ARENALES et al., 2007; RIBEIRO; KARAS, 2013).

A otimização refere-se ao estudo de problemas que buscam otimizar alguma variável através de uma escolha sistemática, deste modo, otimizar significa encontrar a melhor maneira de se fazer algo, dada uma medida do que é ser melhor. Nesse processo, o objetivo é minimizar ou maximizar a função custo ou lucro em determinado processo.

A descrição matemática dos problemas de otimização é feita por meio de modelos de otimização. Um modelo é a representação simplificada de uma situação real. O conceito de simplificação inerente aos modelos está relacionado ao fato de que, dada a complexidade da realidade, é praticamente impossível e/ou economicamente inviável incluir na representação do problema todas as variáveis que podem interferir no resultado do fenômeno que se está estudando. Assim, o modelo abrange apenas as variáveis mais relevantes e que exercem maior impacto sobre o problema (GOLDBARG; LUNA, 2005).

Os modelos matemáticos usados em otimização seguem em geral um padrão que é composto por uma função objetivo, um critério de otimização a minimizar (min) ou maximizar (max), e um conjunto de restrições que devem ser satisfeitas, expressas na forma de equações ou inequações, e ainda, por fim, a definição do conjunto de domínio das variáveis. Dependendo da forma do funcional objetivo e das restrições, os modelos matemáticos são classificados em lineares e não lineares.

Neste trabalho será abordado apenas os modelos não lineares, ou seja, os modelos matemáticos que no processo da modelagem apresentam aspectos que prejudicam a linearidade, seja nas variáveis ou em uma de suas equações. A não linearidade pode aparecer na função objetivo, a qual pode ser não linear e/ou no conjunto de restrições, sendo este formado por

equações ou inequações não lineares. Outra classe importante dentro da otimização está relacionada com as restrições, quando o problema de programação não linear (PNL) apresenta um conjunto de restrições, têm-se o caso da otimização restrita, caso contrário, tem-se o problema de otimização irrestrita.

As não linearidades englobadas em um modelo matemático estão dentro de duas principais categorias, que são

- Relações observadas empiricamente, tais como variações não proporcionais em custos, resultados de processos e características de qualidade.
- Relações deduzidas estruturalmente, que englobam fenômenos físicos, deduzidos matematicamente e regras administrativas.

Todos os fenômenos físicos devem ser levados em consideração na elaboração do modelo matemático, que, por sua vez, influenciam no nível de complexidade do método de solução empregado. A forma matemática que a função objetivo a ser minimizada (ou maximizada) irá tomar, bem como as restrições e o tipo de variável, irão definir os diversos modelos de otimização.

Outra forma de classificar os problemas de otimização é pela dimensão: unidimensional e multidimensionais. Os problemas unidimensionais envolvem funções que dependem de uma única variável dependente, assim, pode-se dizer, que a busca do ótimo consiste em subir e descer picos ou vales. Já os problemas multidimensionais, envolvem funções que dependem de duas ou mais variáveis dependentes, portanto, a otimização multidimensional pode ser visualizada como a busca de picos ou vales. Entretanto, não se está restrito a andar em uma única direção, em vez disso, a topografia deve ser examinada para se alcançar a meta de forma eficiente. Algumas referências para este assunto podem ser encontradas em Clarke Yu. S. Ledyayev e Wonnenski (1998), Izmailov e Solodov (2009), Ribeiro e Karas (2013) e Vinter (2000).

Diante da importância que os problemas de otimização apresentam na solução de problemas, na física, na matemática, nas engenharias, entre outras, o estudo dos modelos de otimização tornam-se fundamentais para a tomada de decisão em determinados processos, bem como os estudos de estratégias de resolução numérica.

Diversos autores tem estudado e proposto modelos de otimização com o objetivo de reduzir os custos de produção, algumas referências sobre este assunto são Stiegelmeier et al. (2010), Cacho (1998), Kotani et al. (2009, 2011) e Jones e Cacho (2000). Em Stiegelmeier et al. (2010) foi proposto um modelo de otimização dinâmico não linear com o objetivo de minimizar os custos de produção em um lavoura a partir da aplicação de controle químico usando

programação não linear. Em Cacho (1998) e Jones e Cacho (2000) foram desenvolvidas diversas formas de controle para maximizar os lucros em uma determinada estação ou ano e, conseqüentemente, diminuir os custos em um sistema de colheita usando programação dinâmica. Já, Kotani et al. (2009) e Kotani et al. (2011) propõe um modelo dinâmico de manejo de plantas daninhas a partir da remoção de plantas daninhas na tentativa de minimizar as perdas associadas á presenças das plantas invasoras. Outra problema usado na agricultura onde a otimização está presente é no controle de populações de pestes, Rafikov e Balthazar (2005) e Christiaans e Pething (2007) apresentam modelos populacionais baseados no controle de populações.

Portanto, visando buscar estratégias de resolução de problemas de otimização não linear, no presente trabalho será abordado o problema de programação não linear (PNL) para variáveis contínuas. O caso particular abordado é o problema irrestrito, ou seja, problemas de otimização sem restrições. Serão descritas algumas situações que garantem a existência de um minimizador e, em seguida, discutidas as condições de otimalidade para o problema de otimização irrestrita. Resultados numéricos serão obtidos via método do gradiente, o qual consiste em procurar o mínimo na direção de maior taxa de decrescimento da função objetivo a partir de uma solução inicial dada.

O presente texto esta dividido em 6 capítulos. O Capítulo 2, seguinte a esta introdução, apresenta definições básicas sobre topologia no \mathbb{R}^n e alguns resultados importantes da Álgebra Linear e Cálculo Diferencial importantes para o desenvolvimento da teoria. O Capítulo 3 introduz a teoria de otimização, onde são descritos o problema de otimização irrestrito, aspectos teóricos sobre existência de soluções e condições de otimalidade para problemas irrestritos e, ainda, são apresentados alguns problemas para exemplificar o uso da teoria. No Capítulo 4 são descritos os aspectos computacionais baseados no método do gradiente, o qual será utilizado para a resolução numérica. O Capítulo 5 apresenta os resultados computacionais obtidos a partir da implementação do método do gradiente, a análise da convergência e eficiência do método a partir de testes computacionais realizados com o auxílio da rotina *fminunc* presente no toolbox do software Matlab. Finalmente, no Capítulo 6 são apresentadas as considerações finais sobre o presente trabalho.

2 FERRAMENTAS MATEMÁTICAS

Nesta seção serão abordados conceitos da Topologia no \mathbb{R}^n , da Álgebra Linear e do Cálculo Diferencial considerados fundamentais para o desenvolvimento do trabalho. Algumas referências sobre estes assuntos são Lima (1999) e Boldrini (1998), Guidorizzi (2001), respectivamente.

2.1 NOÇÕES DE TOPOLOGIA NO \mathbb{R}^N

Primeiramente destaca-se aspectos relevantes da teoria de conjuntos, que têm fundamental importância para o desenvolvimento deste trabalho.

Definição 1. Um ponto a é ponto interior do conjunto $X \subset \mathbb{R}^n$ se existe $\varepsilon > 0$ tal que $B(a, \varepsilon) \subset X$ em que $B(a, \varepsilon) = \{x \in \mathbb{R}^n / |x - a| < \varepsilon\}$, é uma bola de centro a e raio ε .

Definição 2. O conjunto dos pontos interiores de X é chamado de interior do conjunto X e denotado por $\text{int } X$.

Definição 3. Um conjunto $A \subset \mathbb{R}^n$ é dito aberto se $A = \text{int } A$, ou seja, todos os pontos de A são interiores a A .

Definição 4. Um ponto a é ponto aderente do conjunto $X \subset \mathbb{R}^n$ quando a é limite de alguma sequência de pontos $x_n \in X$.

Definição 5. O fecho de um conjunto X é o conjunto \bar{X} formado por todos os pontos aderentes do conjunto X .

Definição 6. Um conjunto X é dito fechado quando $X = \bar{X}$, ou seja, todo ponto aderente de X pertence a X .

Definição 7. Um conjunto $X \subset \mathbb{R}^n$ chama-se compacto quando é limitado e fechado.

Definição 8. Uma cobertura do conjuntos $X \subset \mathbb{R}^n$ é uma família $C = (C_\lambda)_{\lambda \in L}$ de conjuntos $C_\lambda \subset \mathbb{R}^n$, tais que $X \subset \bigcup_{\lambda \in L} C_\lambda$, isto é, para todo $x \in X$ existe algum $\lambda \in L$ tal que $x \in C_\lambda$.

Definição 9. Suponha $x_1 \neq x_2 \in \mathbb{R}^n$. Os pontos descritos por

$$\lambda x_1 + (1 - \lambda)x_2 = x_2 + \lambda(x_1 - x_2),$$

com $\lambda \in \mathbb{R}^n$, descrevem a reta que passa por x_1 e x_2 . Para $0 < \lambda < 1$, a combinação descreve o segmento de reta entre x_1 e x_2 .

Definição 10. Um conjunto C é convexo se o segmento de reta entre dois pontos quaisquer do conjunto estiver em C , isto é, para quaisquer $x_1, x_2 \in C$ e $\lambda \in [0, 1]$,

$$\lambda x_1 + (1 - \lambda)x_2 \in C.$$

Definição 11. Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : C \rightarrow \mathbb{R}$ é convexa se

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

para todo $x, y \in C$ e $0 \leq \theta \leq 1$.

Definição 12. Em termos precisos, $f : X \rightarrow \mathbb{R}$ é contínua em $a \in X$ quando, $\forall \varepsilon > 0$, exista $\delta > 0$ tal que $x \in X$ e $|x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$. Quando f é contínua em cada ponto do domínio, f é contínua.

Definição 13. Uma função é dita de classe C^2 quando as derivadas parciais de segunda ordem são contínuas.

Definição 14. Uma função é suave sempre que suas derivadas parciais de qualquer ordem estão definidas e são contínuas.

2.2 RESULTADOS DA ÁLGEBRA LINEAR

No desenvolvimento da teoria de otimização é fundamental ter em mente alguns conceitos da álgebra linear como por exemplo, a matriz simétrica e matriz definida positiva, as quais são usadas para definir as condições de otimalidade. Conforme apresentado a seguir.

Definição 15. Considere a matriz real $A = (a_{ij})$ de ordem $m \times n$. Denomina-se transposta de A e indica-se por A^t a seguinte matriz $n \times m$: $A^t = (b_{ji})$ em que $b_{ji} = a_{ij}, \forall i = (1, 2, \dots, m)$ e $j = 1, 2, \dots, n$.

Definição 16. Uma matriz quadrada é dita simétrica se $A = A^t$, ou seja, quando $a_{ij} = a_{ji}$, para todo i e j com $i = 1, \dots, m$ e $j = 1, \dots, n$.

Definição 17. O núcleo de uma matriz quadrada $A \in M^{n \times n}$, denotado por $N(A)$, é um subconjunto de \mathbb{R}^n formado por todas as soluções do sistema homogêneo $Ax = 0$, ou seja, $N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$.

Definição 18. Seja $A \in M^{n \times n}$ uma matriz simétrica. Diz-se que A é definida positiva quando $x^T Ax > 0$, para todo $x \in \mathbb{R}^n \setminus \{0\}$. Tal propriedade é denotada por $A > 0$. Se $x^T Ax \geq 0$, para todo $x \in \mathbb{R}^n$, A é dita semi-definida positiva, fato este denotado por $A \geq 0$.

Definição 19. Os menores principais A_k da matriz A são as matrizes definidas pelo determinante, denotado por \det , das k primeiras linhas e colunas de A , ou seja, considere

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

então,

$$A_1 = a_{11},$$

$$A_2 = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

e

$$A_3 = \det(A).$$

Para o caso geral, tem-se:

$$A_k = \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix},$$

com $k = 1, 2, \dots, n$.

Definição 20. Considere $A \in \mathbb{R}^{n \times n}$, $u \in \mathbb{R}^n$ não-nulo é autovetor de A se existe $\lambda \in \mathbb{R}$ tal que:

$$Au = \lambda u$$

Neste caso, u é autovetor de A associado ao autovalor λ .

Definição 21. Uma matriz real simétrica é definida positiva se e somente se todos seus autovalores são positivos.

Definição 22. *As seguintes propriedades são condições necessárias e suficientes para que uma matriz A seja positiva definida.*

- $x^T Ax > 0, \forall x \neq 0$.
- *Todos os autovalores de matriz A são positivos.*
- *Todas os menores principais da matriz A possuem determinantes positivos.*
- *Todos os pivôs são positivos e não é preciso, teoricamente, fazer trocas de linhas na eliminação gaussiana em A .*

2.3 RESULTADOS DO CÁLCULO DIFERENCIAL

Algumas ferramentas do Cálculo Diferencial são fundamentais para o desenvolvimento do trabalho. Destaca-se os conceitos relacionados ao gradiente de uma função e sua matriz Hessiana, os quais são utilizados no desenvolvimento dos métodos numéricos.

O conceito de gradiente de uma função será muito usado nesse trabalho, por isso, vale destacar algumas propriedades, tais como:

1. O gradiente é a direção de crescimento ou decréscimo da função.
2. O gradiente é a direção de crescimento ou decréscimo mais rápido.
3. O gradiente é perpendicular à curva de nível da função.

Definição 23. *Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função de classe C^2 . O gradiente de f é definido por*

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

e a Hessiana de f é dada por

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}.$$

Teorema 1. (Taylor de primeira ordem)

Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável e $\bar{x} \in \mathbb{R}^n$. Então, pode-se escrever

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + r(x)$$

com

$$\lim_{x \rightarrow \bar{x}} \frac{r(x)}{\|x - \bar{x}\|^2} = 0.$$

Demonstração. O polinômio $p_1(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x})$ é uma aproximação linear para f em torno do ponto \bar{x} e é chamado polinômio de Taylor de ordem 1 da função. Dentre todos os polinômios de grau menor ou igual a 1, ele é o único que satisfaz $p(\bar{x}) = f(\bar{x})$ e $\nabla p(\bar{x}) = \nabla f(\bar{x})$. O limite nulo significa que para x próximo de \bar{x} o resto $r(x)$ é muito pequeno e vai para zero mais rápido que $\|x - \bar{x}\|$. \square

Teorema 2. (Taylor de segunda ordem)

Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função duas vezes diferenciável e $\bar{x} \in \mathbb{R}^n$, então

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(\bar{x})(x - \bar{x}) + r(x)$$

com

$$\lim_{x \rightarrow \bar{x}} \frac{r(x)}{\|x - \bar{x}\|^2} = 0.$$

Demonstração. Analogamente, o polinômio

$$p_2(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(\bar{x})(x - \bar{x})$$

é uma aproximação quadrática para f em torno do ponto \bar{x} e é chamado polinômio de Taylor de ordem 2 da função.

Dentre todos os polinômios de grau menor ou igual a 2, ele é o único que satisfaz

$$p(\bar{x}) = f(\bar{x}), \nabla p(\bar{x}) = \nabla f(\bar{x}) \text{ e } \nabla^2 p(\bar{x}) = \nabla^2 f(\bar{x}).$$

O limite nulo significa que para x próximo de \bar{x} o resto $r(x)$ é muito pequeno e vai para zero mais rápido que $\|x - \bar{x}\|^2$. \square

Teorema 3. (Teorema de Weierstrass)

Sejam $D \subset \mathbb{R}^n$ um conjunto compacto não-vazio e $f : D \rightarrow \mathbb{R}$ uma função contínua. Então, a função f assume valor máximo e mínimo global em D .

Demonstração. Deve-se primeiro mostrar que, se $D \subset \mathbb{R}^n$ é compacto e f é contínua, então $f(D)$ é compacto. Seja (y_n) uma sequência em $f(D)$, então cada $y_n = f(x_n)$ para algum $x_n \in D$. Como D é compacto, existe uma subsequência (x_{n_k}) de (x_n) tal que $x_{n_k} \rightarrow x \in D$. Como f é contínua, $f(x_{n_k}) \rightarrow f(x) \in f(D)$, logo, $f(D)$ é compacto.

Agora, sendo $f(D)$ compacto, da Definição 7, sabe-se que $f(D)$ é fechado e limitado, assim, admitirá supremo e ínfimo. Sejam, $m = \inf\{f(x)/x \in D\}$, $M = \sup\{f(x)/x \in D\} \in f(D)$, assim, existem x_0 e $x_1 \in D$ tais que $m = f(x_0)$ e $M = f(x_1)$.

Logo,

$$f(x_0) \leq f(x) \leq f(x_1),$$

para todo $x \in D$.

Portanto, problemas de minimizar e de maximizar f em um conjunto D têm soluções globais. □

3 OTIMIZAÇÃO IRRESTRITA

Nesta seção será apresentada a formulação matemática dos problemas de otimização não linear, com destaque para problema de otimização irrestrito. Resultados de existência de soluções e as condições necessárias e suficientes de otimalidade para essa classe de problemas são apresentados. Ainda, são descritos alguns exemplos para mostrar a aplicação das condições apresentadas.

3.1 PROGRAMAÇÃO MATEMÁTICA

De forma geral, pode-se dizer que otimização consiste em encontrar pontos de mínimo ou de máximo de uma função sobre um conjunto de restrições. O problema pode ser expresso matematicamente como um problema de programação matemática da seguinte forma:

$$\begin{aligned} & \min f(x) \\ & \text{sujeito a} \\ & g_i(x) = a_i, \quad i = 1, 2, \dots, p \\ & h_j(x) \geq b_j, \quad j = 1, 2, \dots, n \end{aligned} \tag{1}$$

em que $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ e $h_j : \mathbb{R}^n \rightarrow \mathbb{R}$.

No problema expresso em (1) defini-se x como um vetor n -dimensional de variáveis, $f(x)$ o funcional objetivo ou a função que deve ser maximizada ou minimizada, $g_i(x)$ restrições expressas por igualdade, $h_j(x)$ restrições expressas por desigualdades e a_i e b_j constantes. Note que um problema de maximizar f pode ser substituído por minimizar $-f$.

Problemas de otimização em geral podem ser escritos usando a formulação (1), chamada de formulação padrão. Uma distinção importante entre os problemas do tipo (1) é conforme a forma do funcional objetivo $f(x)$ e suas restrições. Nesse contexto, se $f(x)$ não for linear e/ou as restrições forem não lineares, tem-se o caso da programação não linear. Além disso,

se existe pelo menos uma restrição, tem-se problemas de otimização restrito, caso contrário, tem-se o problema irrestrito. Algumas referências para este assunto são Clarke Yu. S. Ledyayev e Woiwinski (1998), Izmailov e Solodov (2009) e Ribeiro e Karas (2013).

3.2 OTIMIZAÇÃO IRRESTRITA

Problemas irrestritos surgem de modelos de problemas que naturalmente não possuem restrições. Em alguns casos, o problema teria restrições, mas elas podem ser ignoradas. Em outros casos, problemas originalmente restritos têm as restrições transformadas em um termo de penalidade na função objetivo, tornando-se irrestritos.

Em minimização irrestrita, tem-se por objetivo minimizar uma função de variáveis reais a um valor real sem que haja restrições nos valores das variáveis, ou seja:

$$\min f(x), \quad x \in \mathbb{R}^n, \quad (2)$$

em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função suave.

Para a obtenção da solução do problema descrito em (2) é necessário utilizar uma condição necessária de otimalidade, a qual fornece uma candidata à solução analítica para o problema e, portanto, exata.

3.2.1 EXISTÊNCIA DE SOLUÇÕES

A seguir são descritas as condições para a existência de soluções para o problema irrestrito (2).

Considere um conjunto $D \subset \mathbb{R}^n$ e uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$. O problema de minimizar f descrito em (2) pode ser reescrito como:

$$\min f(x), \quad x \in D, \quad (3)$$

onde D é o conjunto viável do problema. O objetivo do problema (3) é encontrar o ponto $x \in D$ tal que o valor de f seja o menor possível.

Definição 24. Um ponto $x^* \in D$ é um minimizador global de (3), se

$$f(x^*) \leq f(x) \quad \forall x \in D. \quad (4)$$

Para determinar o minimizador global do problema (3) a condição (4) deve ser satisfeita, porém, em muitos casos, se tem apenas o conhecimento local de f . Nesse caso, o conjunto de soluções viáveis fica restrito a uma vizinhança de x^* e, com isso, pode-se obter apenas minimizadores locais de f .

Definição 25. Um ponto $x^* \in D$ é um minimizador local fraco do problema (3), se existe uma vizinhança U de x^* tal que

$$f(x^*) \leq f(x) \quad \forall x \in D \cap U, \quad (5)$$

em que a vizinhança U é um conjunto aberto que contém x^* .

Definição 26. Um ponto $x^* \in D$ é um minimizador local forte ou minimizador local estrito do problema (3), se existe uma vizinhança U de x^* tal que

$$f(x^*) < f(x) \quad \forall x \in U, \quad (6)$$

com $x \neq x^*$.

Observa-se, pelas definições anteriores, que para determinar um ponto de mínimo x^* do problema (3) é necessário fazer uma inspeção dos pontos do domínio de f e verificar se nenhum deles possui valor de função menor. No entanto, quando uma função é suave, existem condições suficientes de otimalidade que garantem a existência de um minimizador de f , como por exemplo, o Teorema de Weierstrass, descrito na Seção 2.3, o qual garante a existência de um máximo ou mínimo global em um conjunto D compacto.

3.2.2 CONDIÇÕES DE OTIMALIDADE

Considere o problema de minimização irrestrita (3), repetido aqui para facilitar o entendimento:

$$\min f(x), \quad x \in \mathbb{R}^n, \quad (8)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Para que um dado $x^* \in \mathbb{R}^n$ seja minimizador local do problema (8) algumas condições devem ser satisfeitas. Condições estas chamadas de condições necessárias de otimalidade.

Teorema 4. (*Condição necessária de primeira ordem*)

Considere a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* é um minimizador local irrestrito de f , então

$$\nabla f(x^*) = 0. \quad (9)$$

Definição 27. Um ponto $x^* \in \mathbb{R}^n$ que satisfaz (9) é dito ponto crítico ou estacionário da função f .

Teorema 5. (Condição necessária de segunda ordem)

Considere que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja duas vezes diferenciável. Se x^* é um minimizador local irrestrito de f , então vale (9) e a matriz Hessiana de f no ponto x^* é semi-definida positiva, isto é,

$$x^T \nabla^2 f(x^*) x \geq 0, \quad (10)$$

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

Teorema 6. (Condição suficiente de segunda ordem)

Considere que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja duas vezes diferenciável na vizinhança aberta de x^* . Se x^* é um ponto estacionário da função f e $\nabla^2 f(x^*)$ é definida positiva, então x^* é minimizador local estrito de f .

O Teorema 6 fornece as condições suficientes para que um ponto x^* seja minimizador local de f .

O resultado a seguir caracteriza minimizadores locais e globais quando a função objetivo é convexa.

Teorema 7. Se $f : D \rightarrow \mathbb{R}$ uma função convexa no conjunto convexo $D \subset \mathbb{R}^n$, então todo minimizador local x^* de f em D é um minimizador global de f nesse conjunto. Se f é diferenciável, então, todo ponto estacionário x^* é um minimizador global de f .

A demonstração dos Teoremas acima encontram-se em Izmailov e Solodov (2009).

No caso de problemas de maximização, resultados análogos dos Teoremas 5 e 6 são obtidos invertendo a desigualdade (10) e considerando a matriz Hessiana definida negativa.

A seguir são apresentados alguns exemplos com o objetivo de aplicar as condições de otimalidade descritas acima.

Exemplo 1 (Função de Rosenbrock) Considere a função

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

Mostre que $x^* = [1 \ 1]^T$ é o único minimizante local de f .

Solução: O processo de resolução consiste em verificar a existência de pontos críticos, x^* , a partir da Definição 27 e, ainda, verificar se esses pontos são minimizadores locais, segundo o Teorema 5, ou seja, $\nabla^2 f(x^*)$ é definida positiva.

A partir do cálculo do gradiente de f , tem-se:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 200(x_2 - x_1^2)(-2x_1) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

$$\nabla f(x) = 0 \Leftrightarrow \begin{cases} -2(1 - x_1) = 0 \Rightarrow x_1 = 1 \\ x_2 - x_1^2 \Rightarrow x_2 = x_1^2 \end{cases},$$

ou seja, $\nabla f(x) = 0 \Leftrightarrow x_1 = 1$ e $x_2 = 1$. Portanto, $x^* = [1 \ 1]$ é o único ponto estacionário de f .

Em seguida, verificando se a matriz é definida positiva, tem-se:

$$\nabla^2 f(x^*) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_1 x_2}(x_1, x_2) \\ \frac{\partial^2 f}{\partial x_2 x_1}(x_1, x_2) & \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) \end{bmatrix},$$

onde $\frac{\partial^2 f}{\partial x_1^2} = 1200x_1^2 + 2 - 400x_2$, $\frac{\partial^2 f}{\partial x_2^2} = 200$ e $\frac{\partial^2 f}{\partial x_1 x_2} = \frac{\partial^2 f}{\partial x_2 x_1} = -400x_1$.

Obtêm-se, assim,

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 1200x_1^2 + 2 - 400x_2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}.$$

$$\nabla^2 f(1, 1) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix},$$

Como os autovetores de $\nabla^2 f(1, 1)$ são positivos, então $\nabla^2 f(1, 1)$ é definida positiva. Portanto, o ponto x^* é um ponto estacionário de f , ou seja, $\nabla f(x^*) = 0$, é minimizador de f uma vez que as condições dos Teorema 5 e 6 são satisfeitas.

Exemplo 2 (Parabolóide) Considere a função

$$f(x) = (x_1 + 1)^2 + (x_2 + 2)^2.$$

Mostre que $x^* = [-1 \ -2]^T$ é o único minimizante local de f .

Solução: O processo de resolução consiste em verificar a existência de pontos críticos, x^* , a partir da Definição 27 e, ainda, verificar se esses pontos são minimizadores locais, segundo o Teorema 5, ou seja, $\nabla^2 f(x^*)$ é definida positiva.

A partir do cálculo do gradiente de f , tem-se:

$$\nabla f(x) = \begin{bmatrix} 2(x_1 + 1) \\ 2(x_2 + 2) \end{bmatrix}$$

com

$\nabla f(x) = 0 \Leftrightarrow x_1 = -1$ e $x_2 = -2$ e, então, $x^* = [-1 \ -2]$ é o único ponto estacionário de f .

Em seguida, verificando se a matriz é definida positiva, tem-se:

$$\nabla^2 f(x^*) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_1 x_2}(x_1, x_2) \\ \frac{\partial^2 f}{\partial x_2 x_1}(x_1, x_2) & \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) \end{bmatrix},$$

com

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial^2 f}{\partial x_2^2} = 1 \text{ e } \frac{\partial^2 f}{\partial x_1 x_2} = \frac{\partial^2 f}{\partial x_2 x_1} = 0.$$

$$\text{Assim, } \nabla^2 f(x_1, x_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Substituindo o ponto crítico:

$$x^* = [-1 \ -2] \Rightarrow \nabla^2 f(-1, -2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Como $x^T A x = [x_1 \ x_2] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2 > 0, \forall x_1 \text{ e } x_2 \in \mathbb{R}^n - \{0\}$, pela Definição 18, tem-se que $A > 0$.

Portando, $A = \nabla^2 f(-1, -2)$ é definida positiva e x^* é minimizador local de f uma vez que as condições dos Teorema 5 e 6 são satisfeitas.

Exemplo 3 (Minimização) Encontrar o ponto de mínimo da função

$$\min f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2.$$

Solução: O processo de resolução desse exemplo será dado em duas etapas: a primeira consiste em verificar, segundo o Teorema 4, os valores de x_1 e x_2 tais que

$$\nabla f(x_1, x_2) = 0 \Leftrightarrow \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow x_1 = 1 \text{ e } x_2 = 2.$$

Portanto, $\nabla f(x_1, x_2) = 0$, então $x_1 = 1$ e $x_2 = 2$.

A segunda parte consiste em verificar, segundo o Teorema 5, que $\nabla^2 f(x^*) > 0$.

$$\nabla^2 f(x^*) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} > 0.$$

Desta forma $a = 2 > 0$, $d = 2 > 0$ e $ac - bd = 3 > 0$.

Assim, todos os menores principais são positivos, então pela Definição 22, $\nabla^2 f(x^*)$ é definida positiva.

Logo, $x_1^* = 1$ e $x_2^* = 2$ são os pontos de mínimo global estrito.

3.2.3 UMA APLICAÇÃO: FUNÇÕES QUADRÁTICAS

Considere a função

$$ax^2 + 2bxy + cy^2 + dx + ey + g = 0 \quad (11)$$

a qual pode ser descrita na forma matricial como

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + g = 0.$$

Seja $x = \begin{bmatrix} x \\ y \end{bmatrix}$, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $m = \begin{bmatrix} d & e \end{bmatrix}$ e $w = g$, então a equação quadrática passa a ser

$$x^T Ax + b^T x + c = 0. \quad (12)$$

Definição 28. Considere $A \in (\mathbb{R}^{n \times n})$ uma matriz simétrica, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$. Defina-se função quadrática $f: \mathbb{R}^n \rightarrow \mathbb{R}$ por $f(x) = \frac{1}{2}x^T Ax + b^T x + c$.

Note que, a Hessiana dessa função é exatamente a matriz A . A forma quadrática deve ser multiplicada por $\frac{1}{2}$ para simplificar os cálculos devido a derivada (derivada de ordem 2).

Para exemplificar, considere os problemas a seguir.

Problema 1 Considere o problema de minimização irrestrito (3), com $f(x) = x^T Ax + b^T x$, uma função quadrática, onde

$$A = \begin{bmatrix} 6 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 4 \end{bmatrix}.$$

Solução: O processo de resolução desse problema será dado em duas etapas: a primeira consiste em verificar, segundo o Teorema 6, se a matriz Hessiana, A , é definida positiva. Para isso, segundo Definição 22, se os menores principais de uma matriz são positivos, então, A é definida positiva, logo, $A_1 = 6 > 0$, $A_2 = 11 > 0$ e $A_3 = 42 > 0$. A segunda etapa consiste em verificar se x^* é um ponto estacionário de f , ou seja, $\nabla f(x^*) = 0$, então, tem-se

$$f(x) = x^T A x + b^T x \Leftrightarrow (Ax^2 + bx)$$

$$\nabla f(x) = 2Ax + b \Rightarrow 2Ax + b = 0 \Rightarrow x^* = -\frac{1}{2}A^{-1}b.$$

Logo, $x^* = -\frac{1}{2}A^{-1}b$ é o ponto de mínimo local estrito, pois a condição de suficiência assegura que $f(x) > f(x^*)$, com x em uma vizinhança.

Problema 2 Considere $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tal que

$$f(x_1, x_2) = x^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2.$$

A partir do cálculo do gradiente, tem-se:

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$\nabla f(0, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

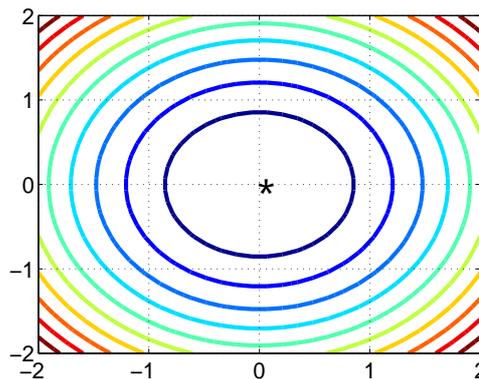


Figura 1: Curvas de nível do problema 2.

A Figura 1 ilustra as curvas de nível da função f . Note que as curvas de nível são circunferências de centro comum no ponto $P_1 = (0, 0)$. A derivada da função nesse ponto é nula, portanto é ponto estacionário. O gradiente aponta a direção de maior crescimento da função, por isso, ao analisar a Figura 1 fica evidente que o ponto P_1 é ponto de mínimo global da função.

4 MÉTODO DO GRADIENTE

Esta seção descreve o método do gradiente usado na resolução numérica de problemas de minimização de funções não lineares. O método do gradiente consiste em procurar o mínimo na direção de maior taxa de decrescimento da função objetivo a partir de uma solução inicial dada. Algumas referências sobre o método do gradiente são Izmailov e Solodov (2012) e Chapra e Canale (2008).

4.1 MÉTODO DO GRADIENTE

Em muitos casos, se torna difícil e até mesmo impraticável, obter uma candidata a solução analiticamente devido à complexidade das expressões envolvidas. Em casos como esses, precisa-se utilizar métodos numéricos, com os quais é possível a obtenção de uma solução aproximada.

Para a resolução numérica do problema de PNL irrestrito descrito na seção anterior, será utilizado o método clássico do gradiente, também chamado método de Cauchy ou método de máxima descida. A escolha desse método foi baseada na sua importância na resolução de problemas de minimização de funções e por ser uma estratégia baseada no método de descida Izmailov e Solodov (2012).

O método do gradiente é um processo iterativo que a cada etapa faz uma busca linear na direção do vetor gradiente da função objetivo no ponto corrente e , consiste em procurar o mínimo na direção de maior taxa de decrescimento da função objetivo a partir de uma solução inicial x_0 dada.

Como o nome implica, o método do gradiente usa explicitamente informação sobre a derivada da função f para gerar um algoritmo eficiente na localização do minimizador. Uma das vantagens desse método é que a informação da derivada da função permite alcançar o extremo com menor número de avaliações da função, ou seja, melhor eficiência computacional.

O método gradiente requer como entrada uma solução inicial x_0 e um procedimento

que calcula o gradiente da função a ser minimizada ou maximizada. Basicamente, o algoritmo pode ser descrito como um processo iterativo em que novos pontos x_{k+1} são gerados a partir do ponto atual x_k

$$x_{k+1} = x_k + \alpha_k d_k \quad (13)$$

onde d_k é a direção de busca linear, α_k representa o tamanho do passo a ser dado na direção de busca e k é o número de iterações.

O algoritmo, então, é dividido em duas etapas, a determinação da direção de busca d_k e a avaliação do parâmetro α_k . Para isso, é necessário utilizar o método da Seção Áurea, usada como estratégia na escolha do passo ótimo e, também, o Critério de Armijo, importante na determinação do minimizador x^* descritos em Izmailov e Solodov (2012).

4.2 DETALHES DE IMPLEMENTAÇÃO

A estratégia adotada para resolver o problema irrestrito (2) é, a partir de uma aproximação inicial $x_k \in \mathbb{R}^n$ da solução do problema, encontrar um ponto $x_{k+1} \in \mathbb{R}^n$ tal que

$$f(x_{k+1}) < f(x_k). \quad (14)$$

Tendo como objetivo encontrar o mínimo de $f(x)$, então, pode-se tomar como estratégia a direção contrária a do gradiente, ou seja, $-\nabla f(x)$ a direção onde $f(x)$ mais decresce. A proposição a seguir enuncia esta relação.

Proposição 1. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável. Então, pode-se afirmar que*

- i. o vetor gradiente ∇f aponta para uma direção onde f é crescente,*
- ii. dentre todas as direções onde f cresce, a direção do gradiente ∇f é a de crescimento mais rápido,*
- iii. o gradiente ∇f é perpendicular a superfície de nível que passa por x .*

A partir da Proposição 1, será apresentado o método iterativo do gradiente para localizar o mínimo de uma função, utilizando o vetor gradiente.

Para exemplificar o uso das superfícies de nível, considere a função $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, onde D é um retângulo em \mathbb{R}^2 . O problema de minimizar f pode ser enunciado como minimizar $f(x)$ na caixa $D = [a \ b] \times [c \ d]$ (veja Figura 2).

A Figura 2 ilustra o caso em que f é um parabolóide. Observe que cada círculo no plano xy corresponde a superfície de nível.

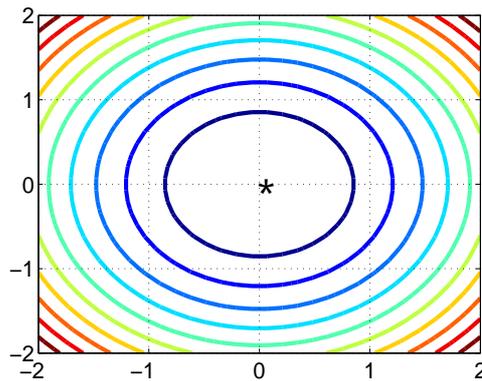


Figura 2: Curvas de níveis.

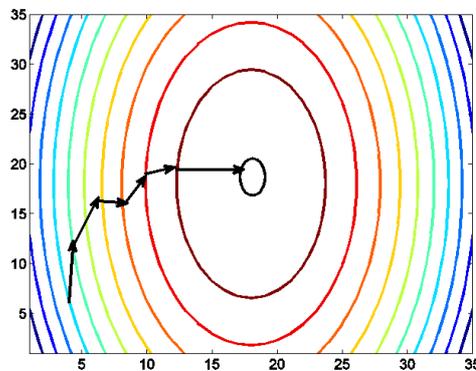


Figura 3: Exemplo da trajetória “zig-zag” do método do gradiente.

Seja x_k um ponto qualquer em $D \in \mathbb{R}^n$ e suponha $\nabla f(x_0) \neq 0$. Pelo item ii. da Proposição 1 a escolha mais óbvia de uma direção de descida é dada por $d_k = -\nabla f(x_k)$. Então, a estratégia é caminhar nessa direção para encontrar o mínimo de f . Com isso, obtêm-se uma sequência $\{x_k\}$ para curvas de nível da função $f(x)$ associadas a valores cada vez menores, fazendo com que a sequência assim gerada possa convergir a uma solução do problema (em geral, só há garantia de convergência para pontos estacionários). Uma característica desse tipo de problema é que as direções utilizadas nas iterações subsequentes são ortogonais. Por isso, a sequência gerada $\{x_k\}$ pelo método se aproxima da solução seguindo uma trajetória “zig-zag”. A Figura 3 ilustra um exemplo das trajetórias que o método do gradiente adota.

Definida a escolha da direção de descida como sendo o anti-gradiente $-\nabla f(x_k)$, precisa-se definir o comprimento do passo α_k nessa direção. O comprimento do passo é calculado observando o comportamento da função f ao longo da semi-reta a partir de x_k na direção d_k . Observe que, pelo fato de d_k pertencer ao conjunto de todas as direções de descida da função f no ponto x , um número $\alpha_k > 0$ satisfazendo (14) sempre existe. A escolha de α_k é feita por uma minimização local de $f(x_k + \alpha_k \nabla f(x))$, com o uso da função auxiliar $g_k(\alpha_k) = f(x_k + \alpha_k \nabla f(x))$.

Observe que $g_k(\alpha_k)$ é uma função de uma única variável. Então, pode-se fazer uso do método da seção áurea para encontrar o seu mínimo. Ao aplicar o método da seção áurea na função $g_k(\alpha_k)$ deve-se tomar o cuidado de restringir α_k , de tal forma que o vetor $x_k + \alpha_k d_k$ permaneça em D . Assim, ao se minimizar a função $g_k(\alpha_k)$, encontra-se os valores ótimos de α_k para cada valor de x_k em cada iteração. Detalhes do método da Seção Áurea são descritos em Izmailov e Solodov (2012).

O processo segue até que a distância entre os pontos x_k e x_{k+1} seja menor que um certo δ fixado como critério de parada, nesse caso será dado por $\|x_{k+1} - x_k\| < \delta$. Também, pode-se adotar como critério de parada o próprio vetor gradiente, da seguinte forma $\|\nabla f(x_k)\| < \delta$.

A sequência $\{x_k\}$ gerada pelo método do gradiente converge quando o tamanho do passo α_k é calculado pela busca exata, nesse caso, faz-se uso do método da Seção Áurea para encontrar o passo ótimo em cada iteração k .

O Algoritmo 1 descreve o procedimento iterativo adotado em que novos pontos x_{k+1} são gerados a partir do ponto atual x_k , baseado no método do gradiente (13), usando as estratégias estabelecidas anteriormente.

Algoritmo 1: Rotina do Gradiente

Entrada: $x_0 \in \mathbb{R}^n$, $f(x)$, $\delta > 0$

Saída: Solução ótima x^*

- 1 Inicialize $k = 0$;
 - 2 REPITA enquanto $\|x_{k+1} - x_k\| > \delta$;
 - 3 $g_k = \nabla f(x)$;
 - 4 $d_k = -g_k$;
 - 5 $\alpha_k =$ mínimo da seção áurea tal que $f(x_k + \alpha_k d_k) < f(x_k)$;
 - 6 $x_{k+1} = x_k + \alpha_k d_k$;
 - 7 $k = k + 1$;
 - 8 **retorna**;
 - 9 Solução ótima = x_k^*
-

O Algoritmo 1 foi implementado em linguagem de programação com apoio do software matemático Matlab. No algoritmo implementado, a determinação do tamanho do passo α_k é obtido pela estratégia da Seção Áurea. Para validar os resultados, foram feitas comparações de diversos problemas de PNL da literatura, usando como método de resolução o método do gradiente estudado e o Optimization Toolbox do Matlab, o qual é uma coleção de funções que estendem a capacidade do Matlab em resolver problemas de otimização.

4.2.1 MÉTODO DA SEÇÃO ÁUREA

O método da Seção Áurea visa a otimização unidimensional. Em vista disso, ele é frequentemente aplicado juntamente como o método do gradiente, pois, qualquer que seja a dimensão inicial do problema, quando o método já calculou a direção de busca, d_k , então, é necessário obter o passo ótimo, α_k .

Supondo que f seja unimodal. Seja $f(x)$ uma função contínua com um único mínimo no intervalo $[a, b]$. O método da Seção Áurea consiste em criar uma sequência $I = x_n$ que converge para o mínimo da função, sendo $I = \frac{a+b}{2}$.

O Algoritmo 2 descreve a rotina da Seção Áurea utilizada para encontrar o passo ótimo, α_k , o qual é necessário para a implementação do método do gradiente.

Algoritmo 2: Rotina da Seção Áurea

Entrada: a, b, δ
Saída: $I = \alpha_k$

- 1 Inicialize $k = 0$;
- 2 REPITA enquanto $|b - a| > \delta$;
- 3 $x_a = b - 0.618(b - a)$;
- 4 $x_b = a + 0.618(b - a)$;
- 5 Se $f(x_a) < f(x_b)$;
- 6 $b = x_b$;
- 7 $x_b = a + 0.618(b - a)$;
- 8 Senão $a = x_a$;
- 9 $x_a = b - 0.618(b - a)$;
- 10 $I = \frac{x_a + x_b}{2}$ **retorna;**

Detalhes do Método da Seção Áurea podem ser encontrados em Izmailov e Solodov (2012).

5 TESTES COMPUTACIONAIS

Nesta seção serão apresentados os resultados computacionais baseados no estudo de funções da literatura, com o objetivo de se verificar a eficiência e convergência do método do gradiente estudado usando como ferramenta o software matemático Matlab.

5.1 RESULTADOS

Os testes computacionais são ferramentas que auxiliam na investigação da eficiência dos algoritmos na resolução de problemas de acordo com os critérios escolhidos. Além disso, os critérios de parada são essenciais para constatar se um ponto estacionário foi alcançado ou se o algoritmo fracassou após atingir um número máximo de iterações.

Para a realização dos testes computacionais foi realizada a escolha de alguns problemas da literatura a fim de verificar a eficiência e velocidade de convergência do método do gradiente descrito na Seção 4. Para validar os resultados utiliza-se a rotina de minimização *fminunc* presente no Optimization Toolbox do Matlab.

Na Tabela 1 apresenta-se os problemas da literatura selecionados para os testes computacionais, presentes em Moré et al. (1981).

Tabela 1: Problemas selecionados para análise.

| $f(x_1, x_2)$ | Função |
|---|-----------------------|
| $f_1 = (x_1 + 1)^2 + (x_2 + 2)^2$ | Parabolóide |
| $f_2 = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ | Rosenbrock |
| $f_3 = (-13 + x_1 + ((5 - x_2)x_2 - 2)x_2)^2 + (-29 + x_1) + ((x_2 + 1)x_2 - 14)x_2)^2$ | Freudenstein and Roth |
| $f_4 = (1.5 - x_1(1 - x_2))^2 + (2.25 - x_1(1 - x_2^2))^2 + (2.625 - x_1(1 - x_2^3))^2$ | Beale's |

Com o auxílio do software matemático Matlab foi implementado o método do gradiente descrito no Algoritmo 1 para minimizar funções $f(x_1, x_2)$ considerando o problema irrestrito (2). O objetivo nesse momento é verificar se o método esta convergindo ao ponto ótimo utilizando a busca linear na direção d_k com escolha do passo ótimo α_k em cada iteração k . Onde

$0 < \alpha_k < 1$ é obtido pelo algoritmo da Seção Áurea.

O critério de parada adotado foi $\|x_{k+1} - x_k\| < \delta$ ou atingir um número máximo de iterações, nesse caso, fixou-se 1000 iterações e $\delta = 0.0001$. Denota-se x_0 ponto inicial, k número de iterações, x^* minimizador e $f(x^*)$ valor da função em x^* .

Tabela 2: Resultados obtidos com a rotina do gradiente, Algoritmo 1.

| Função | x_0 | k | x^* | $f(x^*)$ |
|-------------------------|----------|-----|-----------------|-------------------------|
| 1 Parabolóide | [4 5] | 3 | [-1.000 -1.999] | 1.351×10^{-10} |
| 2 Rosenbrock | [0.5 2] | 36 | [1.027 1.055] | 7.383×10^{-4} |
| 3 Freudenstein and Roth | [0.5 -2] | 351 | [11.359 -0.899] | 4.898×10^1 |
| 4 Beale's | [1 1] | 95 | [2.992 0.497] | 1.029×10^{-5} |

Na Tabela 2 são apresentados os resultados obtidos usando a rotina do gradiente descrita no Algoritmo 1. Optou-se por utilizar o Optimization Toolbox do MatLab afim de comparar os resultados obtidos e verificar a eficiência do método do gradiente. O Optimization Toolbox do MatLab, consiste em uma coleção de funções que estendem a capacidade do Matlab em resolver problemas de otimização já pré-definidas. Para a resolução de problemas irrestritos o Matlab possui a rotina minimização *fminunc*, usada para resolver problemas de otimização sem restrições. A Tabela 3 ilustra resultados obtidos com a rotina do Matlab *fminunc* usando as funções selecionadas para análise.

Tabela 3: Resultados obtidos com a rotina do Matlab *fminunc*.

| Função | x_0 | k | x^* | $f(x^*)$ |
|-------------------------|----------|-----|-----------------|-------------------------|
| 1 Parabolóide | [4 5] | 2 | [-1.000 -2.000] | 1.806×10^{-13} |
| 2 Rosenbrock | [0.5 2] | 11 | [1.000 1.000] | 1.888×10^{-11} |
| 3 Freudenstein and Roth | [0.5 -2] | 19 | [11.41 -0.89] | 4.898×10^1 |
| 4 Beale's | [1 1] | 15 | [3 0.5] | 2.35×10^{-12} |

Ao se analisar os resultados obtidos na Tabela 2, observa-se que o método do gradiente proposto resolveu todos os problemas de forma satisfatória, ou seja, todas as funções convergiram, uma vez que a sequência gerada x_k está se aproximando do mínimo de $f(x_k)$. Para validar estes resultados foram resolvidos os problemas propostos usando a rotina *fminunc* presente no Matlab, com resultados reportados na Tabela 3. A partir da comparação dos resultados obtidos nas Tabelas 2 e 3, conclui-se que o método do gradiente convergiu satisfatoriamente em todos os problemas analisados, uma vez que ponto ótimo foi alcançado.

Para analisar os resultados obtidos foram utilizadas as curvas de níveis da função com o uso da ferramenta gráfica do Matlab. A seguir são descritos os resultados e gráficos obtidos para cada uma das funções estudadas.

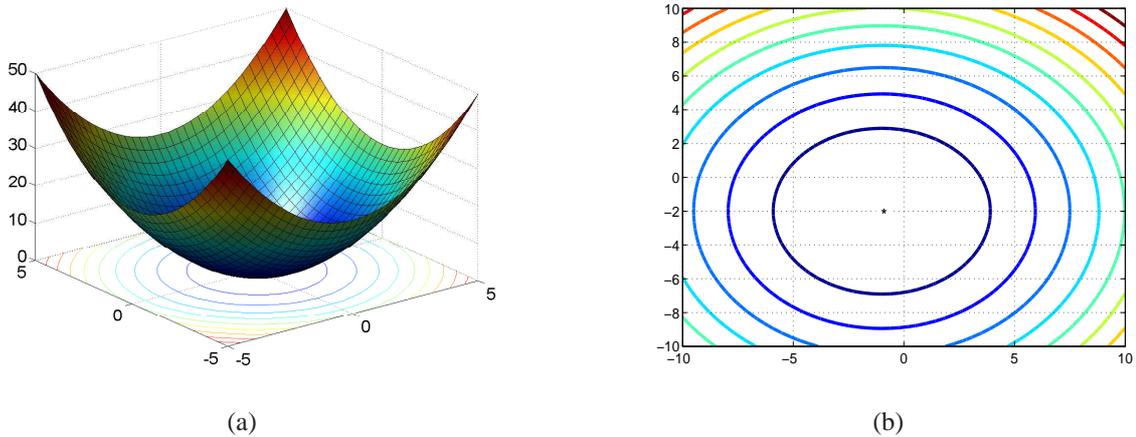


Figura 4: (a) Função Parabolóide. (b) Curvas de níveis.

A Figura 4 ilustra o comportamento da função Parabolóide $f(x) = (x_1 + 1)^2 + (x_2 + 2)^2$. Pode-se observar que existe um ponto de mínimo em $x^* = [-1 \ -2]^T$. Pela análise das curvas de níveis é possível verificar que realmente a função Parabolóide apresenta um minimizador em $[-1 \ -2]^T$. Testando diferentes condições iniciais, percebe-se que a função Parabolóide converge rapidamente para o ponto x^* , e ainda, pode-se concluir que o ponto x^* é um minimizador global.

Testando a rotina do gradiente apresentada no Algoritmo 1 para minimizar a função de Parabolóide e escolhendo $x_0 = [4 \ 5]^T$, o problema é resolvido com apenas 3 iterações, cujos valores da função e da norma do gradiente são dados na Tabela 4. A Figura 5 mostra os gráficos gerados pela rotina do gradiente, para a variação da função e da norma do gradiente, denotado por $|f(x_k)|$ e $\|g\|$, respectivamente, onde a norma do gradiente é descrita como $\|g\| = \sqrt{(x_1)^2 + (x_2)^2}$, ao longo das iterações. Foi utilizado a escala logarítmica no eixo vertical para uma melhor visualização.

Tabela 4: Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função Parabolóide.

| k | f | $\ g\ $ |
|---|--------------------------|---------|
| 0 | 74.00000×10^1 | 6.4031 |
| 1 | 6.17760×10^8 | 2.2363 |
| 2 | 1.35133×10^{10} | 2.2360 |

A função de Rosenbrock $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ é ilustrada na Figura 6(a). Observa-se que essa função apresenta um minimizante global em $x^* = [1 \ 1]^T$. A Figura 6(b) ilustra as curvas de níveis da função de Rosenbrock.

Testando a rotina do gradiente apresentada no Algoritmo 1 para minimizar a função

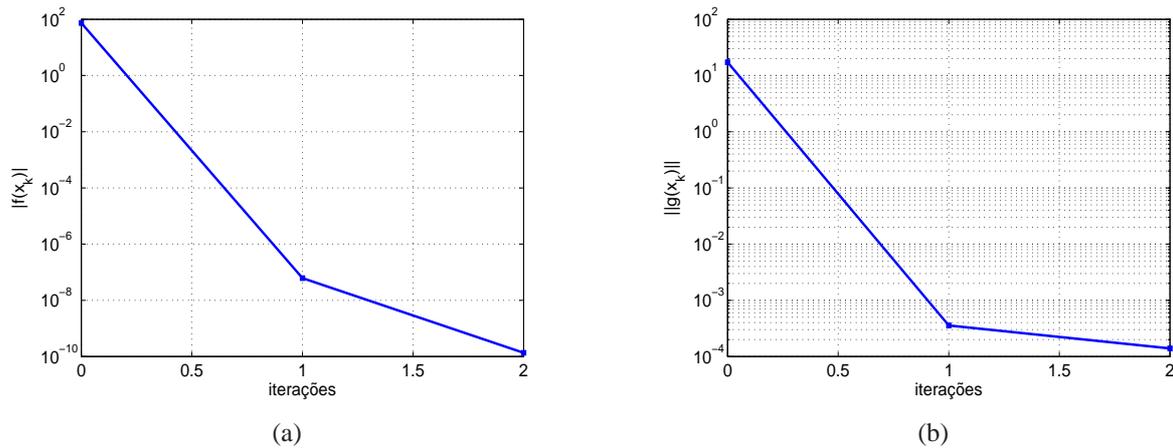


Figura 5: (a) Evolução da módulo da função Parabolóide. (b) Evolução do gradiente

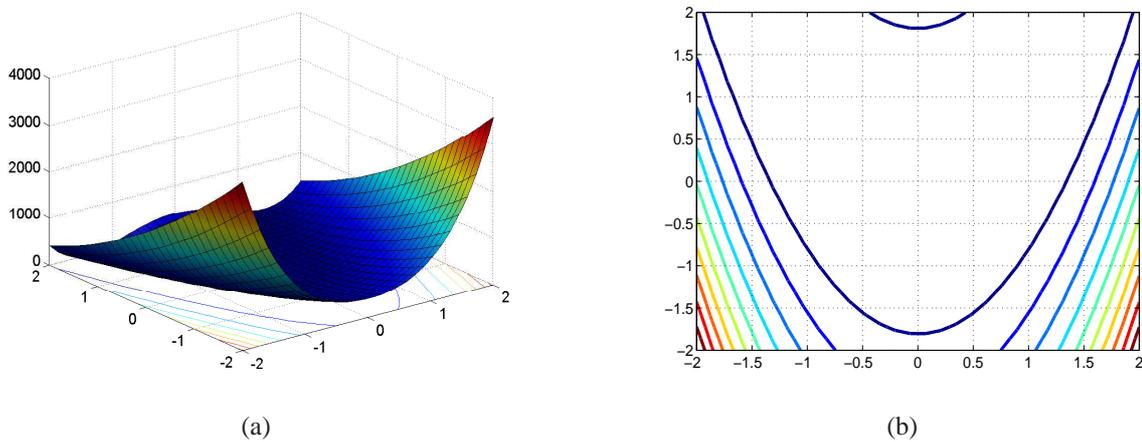


Figura 6: (a) Função Rosenbrock. (b) Curvas de níveis.

de Rosenbrock a partir da ponto $x_0 = [0.5 \ 2]^T$, o problema é resolvido com 36 iterações, cujos valores da função e da norma do gradiente obtidos são dados na Tabela 5. A Figura 7 mostra os gráficos gerados pela rotina, com a variação da função e da norma do gradiente ao longo das iterações.

Uma observação importante para esse tipo de problema é que as estratégias adotadas podem corresponder a um mínimo local. Esse fato pode ser observado na função de Rosenbrock, a qual possui um único minimizante global $x^* = [1 \ 1]^T$ uma vez que que a condição suficiente de segunda ordem é satisfeita, ou seja, a matriz Hessiana de $f(x)$ no ponto x^* é definida positiva. Tomando $x_0 = [1.2 \ 1.2]^T$, observe que o método do gradiente convergiu para o ponto $x^* = [1 \ 1]^T$ com 43 iterações. Porém, tomando-se $x_0 = [-1 \ 2]^T$, o método do gradiente não convergiu para o minimizante, pois x_0 não se encontra na bacia de atração do mínimo global. Esse fato, mostra que o método do gradiente é sensível as condições iniciais quando

utiliza-se funções não-convexas.

Tabela 5: Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função de Rosenbrock.

| k | f | $\ g\ $ |
|----------|---------------------------------|---------------------------------|
| 0 | $3.065000000000000 \times 10^2$ | $0.020600000000000 \times 10^2$ |
| 1 | $0.000252895781668 \times 10^2$ | $0.017738881884331 \times 10^2$ |
| 2 | $0.000251902396495 \times 10^2$ | $0.017736543282180 \times 10^2$ |
| \vdots | \vdots | \vdots |
| 34 | $0.000008703813623 \times 10^2$ | $0.014727695784690 \times 10^2$ |
| 35 | $0.000007390570247 \times 10^2$ | $0.014725671613558 \times 10^2$ |

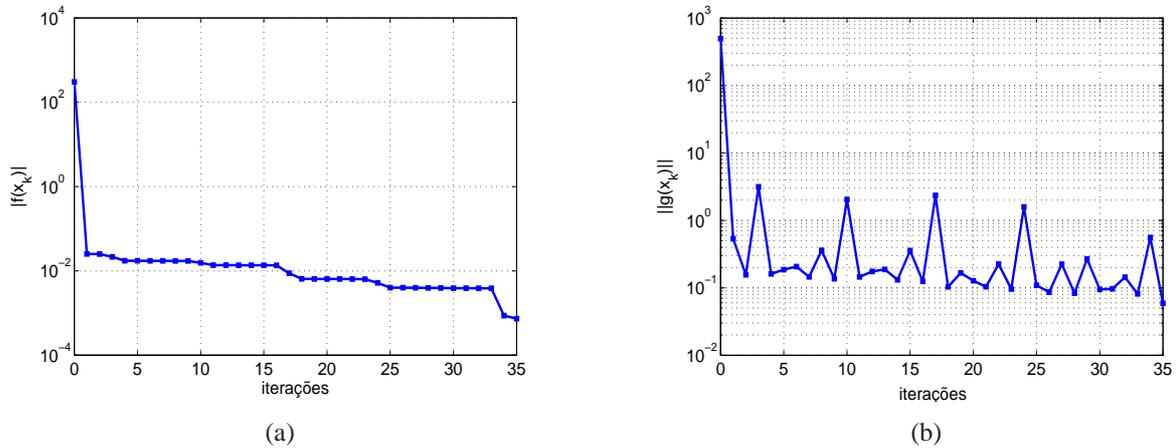


Figura 7: (a) Variação da função Rosenbrock. (b) Variação da norma do gradiente.

A comportamento da função de Freudenstein and Roth é ilustrada na Figura 8(a) e da função de Beale's na Figura 8(b). Vale ressaltar que a função de Freudenstein and Roth apresenta um minimizador global em $x^* = [5 \ 4]^T$ com valor da função objetivo $f(x_1, x_2) = 0$ e um minimizador local em $x^* = [11.41 \ -0.8986]^T$ com valor da função objetivo $f(x_1, x_2) = 48.9842$. Já a função de Beale's apresenta um único minimizador global em $x^* = [3 \ 0.5]^T$ com valor da função objetivo $f(x_1, x_2) = 0$.

Testando a rotina do gradiente para minimizar a função de Freudenstein and Roth a partir do ponto $x_0 = [0.5 \ -2]^T$, o problema é resolvido com 351 iterações, cujos valores da função e da norma do gradiente são reportados na Tabela 6. A Figura 9 mostra os gráficos gerados pela rotina com a variação da função e da norma do gradiente ao longo das iterações, respectivamente. Observe que neste caso o método do gradiente obteve como minimizador o ponto $x^* = [11.359 \ -0.899]^T$, com a tolerância desejada, que corresponde ao ótimo local. Testando para diferentes pontos iniciais x_0 o método não convergiu para o ótimo global, nesse caso o método do gradiente apenas obteve o minimizador local. Isso mostra que o método do

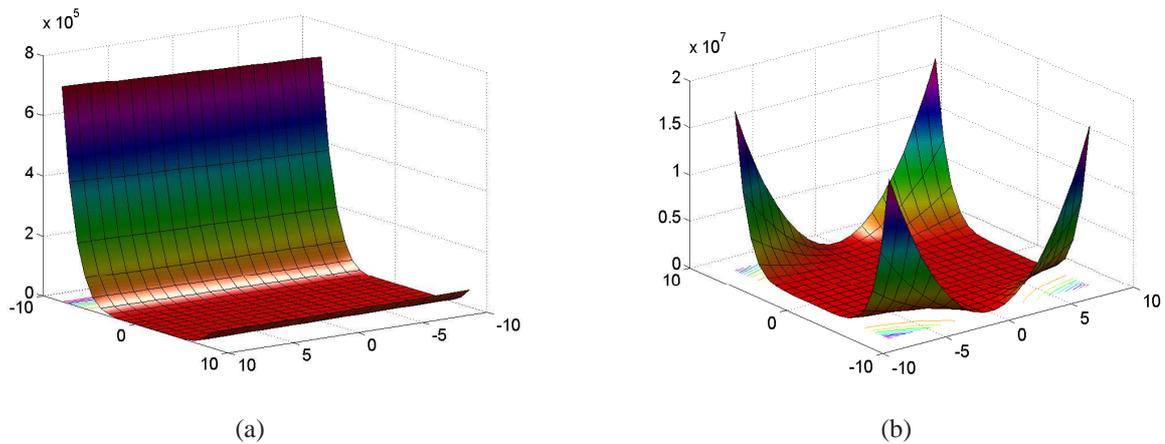


Figura 8: (a) Função Freudenstein and Roth. (b) Função de Baele's

Tabela 6: Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função Freudenstein and Roth.

| k | f | $\ g\ $ |
|----------|----------------------------------|---------------------------------|
| 0 | $0.4005000000000000 \times 10^3$ | $1.272187182797976 \times 10^3$ |
| 1 | $0.031652629717962 \times 10^3$ | $0.034477639375029 \times 10^3$ |
| 2 | $0.081050487552624 \times 10^3$ | $0.008174433481515 \times 10^3$ |
| \vdots | \vdots | \vdots |
| 349 | $0.048985840050098 \times 10^3$ | $0.000930914655352 \times 10^3$ |
| 350 | $0.048985406217656 \times 10^3$ | $0.000058338482239 \times 10^3$ |

gradiente possui algumas condições que devem ser melhoradas para um melhor desempenho do método.

No caso da função de Beale's, o problema é resolvido com 95 iterações, tomando como ponto inicial $x_0 = [1 \ 1]^T$ o método convergiu para o ponto ótimo $x^* = [2.992 \ 0.497]^T$, com a precisão desejada. Os valores da função e da norma do gradiente são reportados na Tabela 7. Na Figura 10 são apresentados os resultados obtidos via rotina com a variação da função e da norma do gradiente ao longo das iterações, respectivamente, mostrando um bom desempenho do método uma vez que o valor da função está tendendo ao valor mínimo de f e a norma do gradiente tende a zero. Vale ressaltar, que considera-se como critério de parada $\|x_{k+1} - x_k\| < \delta$, o qual poderia ser modificado como sendo a norma do gradiente sem perda de generalidade.

Portanto, pode-se verificar que cada mínimo local de uma função diferenciável possui uma bacia de atração. Se o método do gradiente for iniciado em um ponto x_0 não situado na bacia de atração do mínimo global podem ocorrer duas situações, o método do gradiente converge para um mínimo local associado a bacia de atração que estiver o ponto x_0 ou caso x_0 não esteja associado a nenhuma bacia de atração, então o método do gradiente não converge.

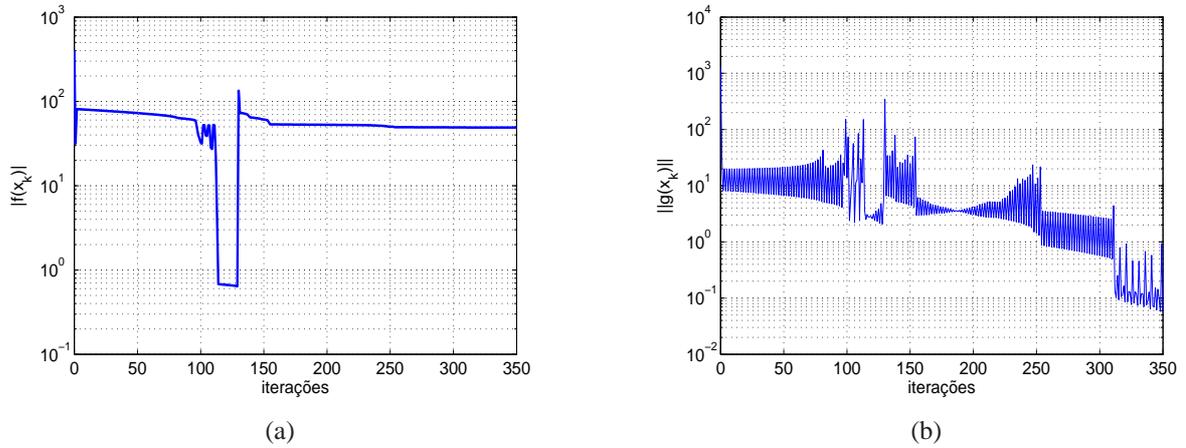


Figura 9: (a) Variação da função Freudenstein and Roth. (b) Variação da norma do gradiente.

Tabela 7: Valores da função e da norma do gradiente obtidos com a rotina do gradiente para a função de Beale's.

| k | f | $\ g\ $ |
|----------|-----------------|------------------|
| 0 | 14.203125000000 | 27.7534252725253 |
| 1 | 4.3685273537189 | 6.47938073465913 |
| 2 | 1.2405573559748 | 3.53502843421325 |
| \vdots | \vdots | \vdots |
| 93 | 0.0000130769958 | 0.01831385026192 |
| 94 | 0.0000103401623 | 0.00265280832215 |

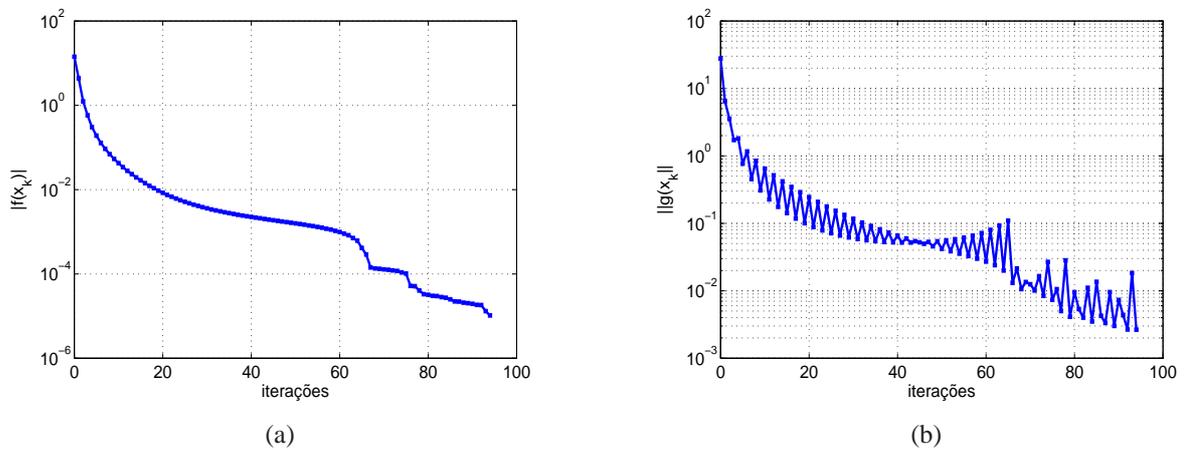


Figura 10: (a) Variação da função Beale's. (b) Variação da norma do gradiente.

6 CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentado o estudo de problemas de otimização irrestrita. Apresentou-se o desenvolvimento teórico das condições de otimalidade para problemas de otimização irrestrito, bem como, o estudo das condições de existência de soluções, além de apresentar o métodos iterativos do gradiente utilizado para obter as soluções numéricas.

Vale destacar que os modelos de otimização irrestrita, surgem muitas vezes de algoritmos de problemas restritos, os quais podem ser modelados como problemas irrestritos a partir da penalização das restrições e, com isso, podendo ser resolvido pelas definições apresentadas no decorrer do trabalho. Portanto, o estudo de problemas de otimização irrestritos torna-se fundamental para auxiliar na tomada de decisão de determinados processos.

Este trabalho proporcionou a relação entre os problemas irrestritos e sua representação, afim de minimizar, obtendo soluções ótimas, através do método do gradiente, implementado no Matlab.

O método do gradiente é de primeira ordem, e em geral, é pouco eficiente para resolver problemas práticos. Mesmo assim, o estudo desse método é um passo importante para se entender os fundamentos da otimização computacional.

O método do gradiente implementado apresentou convergência satisfatória uma vez que seus resultados foram validados comparado com a rotina *fminunc* presente no Matlab. Outro aspecto analisado foi a sensibilidade as condições iniciais a partir do estudo das bacias de atração, pois se o método do gradiente for iniciado longe de uma bacia de atração do mínimo local, o método pode divergir e, ainda, vale destacar que para problemas que apresentam mais de um minimizador, o método do gradiente irá encontrar o minimizador local, devido a fragilidade do método implementado.

REFERÊNCIAS

- ARENALES, M. et al. **Pesquisa Operacional**. Rio de Janeiro: Elsevier, 2007.
- BOLDRINI, J. L. **Algebra Linear**. São Paulo: Harper & Row do Brasil, 1998.
- CACHO, O. J. Solving bioeconomic optimal control models numerically. In: **Bioeconomics Workshop post-Australian Agricultural and Resource Economics Society Conference**. New South Wales, Australia: [s.n.], 1998. p. 13–26.
- CHAPRA, S. C.; CANALE, R. P. **Métodos Numéricos para Engenharia**. São Paulo: MacGraw-Hill, 2008.
- CHRISTIAANS, T. E. T.; PETHING, R. Optimal pest control in agriculture. **Journal of Economic Dynamics and Control**, v. 31, n. 12, p. 3965–3985, 2007.
- CLARKE YU. S. LEDYAEV, R. J. S. F. H.; WONLENSKI, P. R. **Nonsmooth Analysis and Control Theory**. New York: Springer-Verlag, 1998.
- GOLDBARG, H. P.; LUNA, L. **Otimização Combinatória e Programação Linear**. Rio de Janeiro: Elsevier, 2005.
- GUIDORIZZI, H. L. **Um curso de cálculo Volume 1**. Rio de Janeiro: LTC, 2001.
- IZMAILOV, A.; SOLODOV, M. **Otimização Volume 1. Condições de otimalidade, elementos de análise convexa e dualidade**. Rio de Janeiro: IMPA, 2009.
- IZMAILOV, A.; SOLODOV, M. **Otimização Volume 2. Métodos Computacionais**. Rio de Janeiro: IMPA, 2012.
- JONES, R.; CACHO, O. J. A dynamic optimisation model of weed control. In: **44th Annual Conference of the Australian Agricultural and Resource Economics**. Sydney, Australia: [s.n.], 2000. p. 1–17.
- KOTANI, K.; KAKINAKA, M.; MATSUDA, H. Dynamic economic analysis on invasive species management: Some policy implications of catchability. **Mathematical Biosciences**, v. 220, n. 1, p. 1–14, 2009.
- KOTANI, K.; KAKINAKA, M.; MATSUDA, H. Optimal invasive species management under multiple uncertainties. **Mathematical Biosciences**, v. 233, n. 1, p. 32–46, 2011.
- LIMA, E. L. **Curso de Análise Volume 1**. São Paulo: SBM, 1999.
- MORÉ, J. J.; GARBOW, B. S.; HILLSTROM, K. E. Testing unconstrained optimization software. **ACM Transactions on Mathematical Software**, v. 7, n. 1, p. 17–41, 1981.
- RAFIKOV, M.; BALTHAZAR, J. M. Optimal pest control problem in population dynamics. **Computational and Applied Mathematics**, v. 24, n. 1, p. 65–81, 2005.

RIBEIRO, A. A.; KARAS, E. W. **Otimização Contínua: Aspectos Teóricos e Computacionais**. São Paulo: Cengage Learning, 2013.

STIEGELMEIER, E. W. et al. Modelo de otimização da aplicação de herbicida para o controle de plantas daninhas considerando a evolução da resistência. In: **XVIII Congresso Brasileiro de Automática, CBA**. Porto Alegre, RS: [s.n.], 2010. p. 886–893.

VINTER, R. **Optimal Control**. New York: Birkhauser, 2000.