

Ministério da Educação
Universidade Federal do Agreste
de Pernambuco

Relatório Final
Métodos de Otimização

Ministério da Educação
Universidade Federal do Agreste
de Pernambuco

Relatório Final
Métodos de Otimização

Mateus Baltazar de Almeida
Matheus Machado Vieira
Orientador: Gersonilo Oliveira da Silva

Sumário

1	Métodos Matemáticos de Otimização	7
1.1	O Conceito de Otimização	7
1.2	Otimização de Funções à Uma Variável Real	8
1.3	Programando o Método	10
1.4	Otimização de Funções à Várias Variáveis	12
2	Métodos Clássicos de Otimização	13
2.1	O Método de Newton	13
2.1.1	Entendendo o Método	13
2.1.2	Encontrando Mínimos	15
2.2	Outros Métodos	16
2.2.1	Métodos de Gradiente Descendente	16
2.2.2	Simplex	16
2.3	Programando os Métodos	17
2.4	O Método de Newton para Várias Variáveis	19
3	Os Métodos Modernos de Otimização	21
3.1	Breve Relato Histórico	21
3.2	Métodos de Um	21
3.2.1	O Método - Uma breve descrição	21
3.2.2	Exemplos Aplicações	21
3.2.3	Possíveis Aplicações	21
3.3	Métodos de Dois	21
3.3.1	O Método - Uma breve descrição	21
3.3.2	Exemplos Aplicações	21
3.3.3	Possíveis Aplicações	21
3.4	Um com o outro	21
4	Aplicações à Mecânica Celeste	23
4.1	Entendendo o Problema de N Corpos	23
4.2	A Otimização na Mecânica	23
4.3	Resultados Numéricos	23

5	Demais Resultados	25
5.1	Outros Resultados	25
5.1.1	25
	Bibliografia	27

Capítulo 1

Métodos Matemáticos de Otimização

1.1 O Conceito de Otimização

Diz-se otimização, o processo que tem como objetivo encontrar condições que minimizam ou maximizam algo (seja energia, tempo, dinheiro, etc). Sendo este, muitas vezes um trabalho árduo, custoso.

Dessa maneira, na matemática, tal processo é amplamente utilizado quando busca-se valores em conjunto um A (que pode ter restrições), com o objetivo de encontrar uma solução ótima (a melhor resposta para o problema), aplicando os valores de A em uma função objetivo predefinida.

Podendo assim, ser representada da forma como a seguir.

Dada uma função

$$f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}. \quad (1.1)$$

A maximização pode ser definida como; a busca pelo elemento $x^* \in A$, que satisfaz:

$$f(x^*) \geq f(x), \quad (1.2)$$

para todo $x \in A$.

A minimização pode ser definida como; a busca pelo elemento $x^* \in A$, que satisfaz:

$$f(x^*) \leq f(x), \quad (1.3)$$

para todo $x \in A$.

Com isso, podemos agora entender como esse processo pode ser custoso. Iniciando com o fato de que existem os pontos máximos e mínimos (pontos críticos¹), locais e globais, na imagem das funções. Sendo os pontos críticos locais, aqueles que não são os menores ou maiores valores para a minimização e maximização, respectivamente. E os pontos globais, aqueles que representam o menor ou maior valor na imagem das funções, para a minimização e maximização, respectivamente.

¹Iremos denotar como pontos críticos apenas aqueles onde a função tem derivada e esta se anula.

Criando, desse modo, uma certa incerteza ao encontrar um valor crítico numa função, já que, é estritamente difícil saber se o ponto crítico encontrado é local ou global. Como pode-se perceber na Figura 1.1.

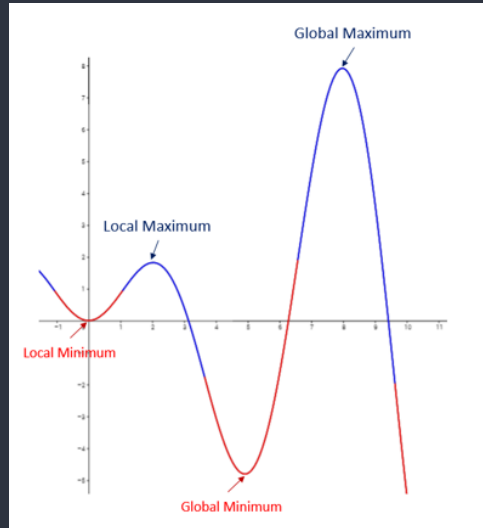


Figura 1.1: Exemplo de pontos críticos locais e globais indicados no gráfico de uma função.
Essa imagem foi coletada de <https://www.onlinemathlearning.com>.

Como na grande maioria dos casos não é possível conhecer antecipadamente todo o espaço gerado pela função, a busca pelo ótimo pode resultar em pontos críticos locais. Donde, dependendo do problema, tais resultados chegam a ser satisfatórios. No entanto, quando o objetivo é estritamente encontrar os máximos ou mínimos globais, o trabalho necessário acaba sendo mais custoso, pelo fato de que há a incerteza da existência de mais pontos na grande maioria dos problemas.

Ademais, o conjunto A , o qual contém os valores aplicáveis no problema, podem respeitar restrições, fazendo com que o campo de respostas do problema, não seja contínuo.

1.2 Otimização de Funções à Uma Variável Real

Evidentemente, as funções possuem as variáveis dependentes (que representam o objeto da otimização) e as variáveis independentes (que suas grandezas podem ser selecionadas), podemos denotar que, para a equação

$$y = f(x), \quad (1.4)$$

quando buscamos otimizá-la, temos como objetivo encontrar valores, x , que quando aplicados à $f(x)$, temos o mínimo ou máximo valor y (seja ele local, ou preferencialmente, global).

Partindo dessa perspectiva, acaba surgindo a necessidade de utilizar algum recurso para encontrar os pontos críticos. E nesse sentido, pode-se utilizar a técnica de **derivação**, que oferece o recurso de identificar tais pontos.

Podemos observar um fator interessante; por exemplo, quando a função está num ponto máximo, existem duas possibilidades, a primeira, a função para de crescer e, em seguida, torna-se indefinida; e a segunda possibilidade, quando a função para de crescer e começa a decrescer. Com isso, é importante ressaltar que por definição, quando a derivada (taxa de variação) em um ponto é positiva, a função cresce, e quando é negativa a função decresce. Conclui-se que, quando a taxa de variação é zero, a função ou para de crescer ou de decrescer, o que configura um ponto de máximo ou de mínimo. Mas também, é quando pode ocorrer o que denominamos ponto de sela (é um comportamento errático da função, que precisa de uma análise mais refinada). Ou em última instância, quando a função é constante.

Com o uso da derivada, podemos pensar num método de otimização bastante simples, considerando $f(x)$ a função que queremos otimizar e $f'(x)$ sua função derivada, podemos dizer que o conjunto O , abaixo definido, possui todos os ótimos locais e globais de $f(x)$:

$$O := \{f(x) | f'(x) = 0\}. \quad (1.5)$$

Portanto, aplicando um filtro em O para obter o máximo e mínimo do conjunto, acabamos por obter o máximo e mínimo de $f(x)$:

$$\max(f(x)) = \max(O), \quad (1.6)$$

$$\min(f(x)) = \min(O). \quad (1.7)$$

Então podemos perceber três problemas:

- Determinar como encontrar os pontos onde a derivada se anule, isto é, os pontos críticos;
- Determinar se temos de fato todos os pontos;
- Classificação dos pontos críticos, que nos revela quais são os máximos e os mínimos, e, ademais, da sua relevância quanto a serem locais ou globais.

Consideraremos, por agora, apenas o problema de encontrar os pontos críticos.

A taxa de variação de uma função, $y = f(x)$, em relação a x , é dada pela relação $\Delta y / \Delta x$. Sendo o cálculo da derivada demonstrado pela Figura 1.2, o que nos leva ao fato de que os pontos críticos provém do conjunto solução da equação $f'(x) = 0$, como definido na equação 1.5. Além disso, caso a função esteja definida num intervalo fechado, temos pelo Teorema do Valor Absoluto, a garantia que ela atingirá o valor dos seus pontos extremos, ou seja, seus valores de máximo e mínimo serão atingidos naquele intervalo. O que culmina nossa busca pelos pontos críticos.

Partindo desse pressuposto, é nítido que a derivada de uma função é também uma função. Portanto, fica evidente que uma função pode ser derivada mais de uma vez, sendo essas derivadas, denominadas de ‘primeira derivada’, ‘segunda derivada’

$$\frac{dy}{dx} = f'(x) = y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Figura 1.2: Derivada: taxa de variação dy/dx .
Essa imagem foi coletada de <https://brasilescola.uol.com.br>.

e por aí em diante. De modo que, a segunda derivada é a derivada da primeira derivada. Concluindo-se que dada a função $f(x)$, sua primeira derivada é $df/dx = p(x)$, e sua segunda derivada $dp/dx = s(x)$.

Agora vamos tratar da classificação dos pontos críticos.

É sabido que a primeira derivada representa a taxa de variação instantânea de um ponto na curva, e, que, a segunda derivada proporciona informações complementares, como por exemplo, se é um ponto de máximo, mínimo ou inflexão, de modo que, ela determina a concavidade da curva naquele ponto. Como pode ser visto na Figura 1.3. E, de modo construtivo, devemos ressaltar que é importante o estudo do gráfico das derivadas de uma função, pois é necessariamente através do uso da interpretação do comportamento da primeira derivada e/ou da segunda derivada, que obtemos a classificação dos pontos críticos seguindo os critérios que são apontados na Figura 1.3.

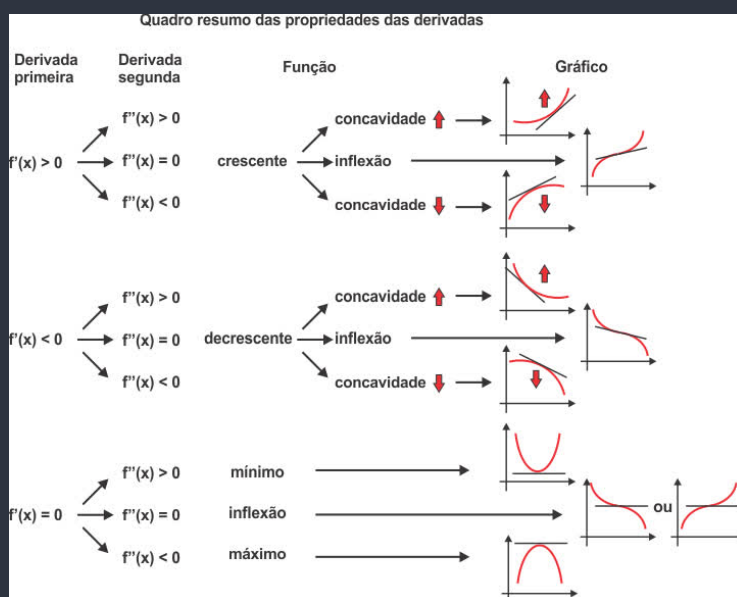


Figura 1.3: Propriedades e relação da primeira e segunda derivada.
Essa imagem foi coletada de <https://www.alfaconnection.pro.br>.

1.3 Programando o Método

A própria definição da derivada já nos oferece uma visão simples de como ela pode ser implementada num programa de computador. Mas, com alguns nuances que devem ser levados em consideração, segue a tal implementação:

```
pub fn derive1x1_v1<F>(f: &F, x: f64) -> f64
where
    F: Fn(f64) -> f64,
{
    ( f(x + h) - f(x) ) / h
}
```

Essa implementação é ingênua no que diz respeito a precisão da operação. O uso, depende da aplicação, caso se queira prezar por velocidade de cálculo e muito pouco sobre precisão, então provavelmente, essa solução seja boa o suficiente. O problema com a precisão desta implementação é que não se é levado em consideração o aspecto infinitesimal da variação (Δx na Figura 1.2, ou a variável h na implementação em Rust), já que é por este aspecto que definimos a derivada. Não temos como ter uma variável com tal propriedade num programa de computador. Como consequência, $f(x+h) - f(x)$ já é calculada de forma falha, entregando uma variação diferente da que seria quando se tem uma variável infinitesimal. O que de fato é entregue, é a variação um pouco mais à frente do ponto esperado.

A partir daí podemos reformular, considerando esse deslocamento, redefinindo da seguinte forma:

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h}. \quad (1.8)$$

Que é equivalente à:

$$f'(x) = \frac{f(x+h) - f(x-h)}{(x+h) - (x-h)}. \quad (1.9)$$

Usaremos essa segunda formulação por motivos de custo de multiplicação de de números reais se comparado a somas e subtrações, além da possível perda de precisão. Assim, conseguimos compensar o tal deslocamento numa nova função insignificamente mais custosa, e mais precisa. Como vemos no código a seguir:

```
pub fn derive1x1_v2<F>(f: &F, x: f64) -> f64
where
    F: Fn(f64) -> f64,
{
    let x1 = x - h;
    let x2 = x + h;

    let y1 = f(x1);
    let y2 = f(x2);
    return (y2 - y1) / (x2 - x1);
}
```


Capítulo 2

Métodos Clássicos de Otimização

2.1 O Método de Newton

2.1.1 Entendendo o Método

O Método de Newton, foi desenvolvido com o objetivo de encontrar estimativas para as raízes de uma função. De modo que, a execução do método é feita de forma iterativa, repetindo sempre o mesmo processo, atualizando o mesmo valor.

Este método faz uso do recurso de derivação, existindo, uma relação muito forte com o ângulo da reta tangente ao ponto, na função. Ademais, é importante ressaltar, que é necessário um palpite inicial, que represente o valor de x , no qual, partindo desse valor, será buscado a raiz. E com isso, vamos entender como o dado método funciona.

Partindo do princípio do método, o objetivo, será utilizar a reta tangente a um ponto, para gerar valores cada vez mais próximos da raiz daquela função. De modo que, será analisado a interseção da reta tangente com o eixo das abscissas. E sendo, a diferença do valor x da entrada da função com o valor da fração entre a função e sua derivada no mesmo ponto, o novo valor de entrada na função, construindo assim várias iterações, gerando valores cada vez mais próximos da raiz. Como será visto, a seguir:

Já sabido que a equação da reta é dada como:

$$(y - y_0) = m(x - x_0). \quad (2.1)$$

E levando em conta que temos como objetivo encontrar x , que é um ponto sobreposto no eixo x , podemos considerar $y = 0$, logo:

$$-y_0 = m(x - x_0). \quad (2.2)$$

Com isso, sabemos que y_0 é a imagem da função $f(x_0)$, e m representa o ângulo da reta tangente ao ponto x_0 , ou seja: $m = f'(x_0)$. Desenvolvendo essa equação, temos:

$$-f(x_0) = f'(x_0)(x - x_0), \quad (2.3)$$

$$-f(x_0) = f'(x_0)x - f'(x_0)x_0, \quad (2.4)$$

$$0 = f'(x_0)x - f'(x_0)x_0 + f(x_0), \quad (2.5)$$

$$0 = x - x_0 + \frac{f(x_0)}{f'(x_0)}, \quad (2.6)$$

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}. \quad (2.7)$$

E com isso, exemplificando em um gráfico, temos que, sendo $f(x) = -x^2 + 2x$, o palpite inicial $x_0 = 1.5$, $P1$ sendo o ponto que representa x_0 aplicado a função f e x_1 a interseção da reta tangente à $P1$ com o eixo x .

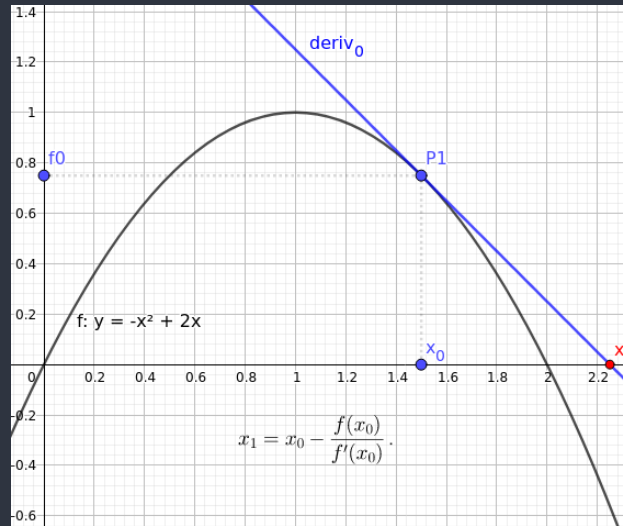


Figura 2.1: Primeira iteração do Método de Newton.

E a partir do dado gráfico, agora podemos entender melhor como a iteração funcionará, pois, o valor gerado, x_1 , será aplicado no mesmo método. E com isso, podemos construir a seguinte equação:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}. \quad (2.8)$$

Desse modo, gera-se uma sequência $\{x_k\}$, donde, esta, converge para a raiz da função. E nesse sentido, vejamos a próxima iteração (Figura 2.2) do exemplo mostrado na Figura 2.1.

E assim, podemos notar a aproximação de x_k para a raiz de f , observando o ponto x_2 .

Ademais, é válido ressaltar que o denominador da equação 2.8 tem que ser diferente de 0 (ou seja, a reta tangente num ponto P não pode ser paralela ao eixo

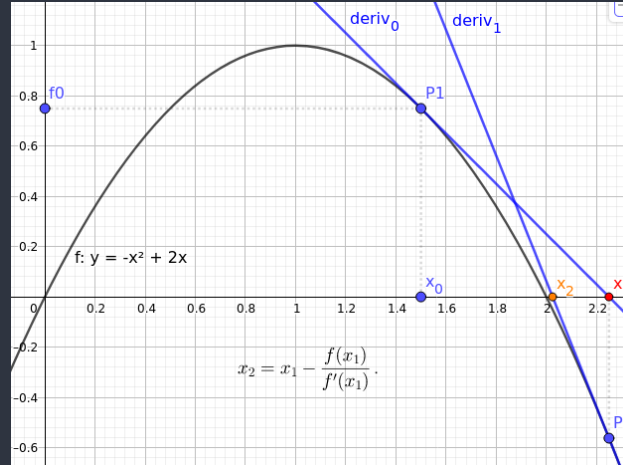


Figura 2.2: Segunda iteração do Método de Newton.

x). No entanto, caso contrário, isso significa que a dada função não possui raiz na proximidade daquele ponto.

2.1.2 Encontrando Mínimos

É dessa maneira que, agora, podemos utilizar este método para encontrar mínimos de uma função. Vimos que, o Método de Newton calcula raízes, e desse modo, vinculando o que foi estudado no Capítulo 1, sabemos que, os mínimos de uma função podem ser representados como raízes de sua derivada.

Então, dado que as raízes de $f(x)$ podem ser geradas a partir da equação 2.8, temos que: considerando $g(x)$ uma função duas vezes derivável, e tendo como objetivo encontrar seus pontos de mínimo, pode-se utilizar o Método de Newton para resolver tal problema da seguinte forma:

$$x_{k+1} = x_k - \frac{g'(x_k)}{g''(x_k)}. \quad (2.9)$$

Encontrando então as raízes aproximadas de sua derivada, e considerando estas como x^* , temos então que:

$$g'(x^*) = 0. \quad (2.10)$$

O método é simples, entrega muitas vezes ótimos locais próximos ao ponto inicial, mas tem seu destaque, que é ser, facilmente computável.

Problemas de maximização podem ser vistos sob o seguinte olhar:

$$\max(f(x)) = \min(-1 * f(x)) \quad (2.11)$$

Que com isto podem ser otimizados pelo Método de Newton também.

O movimento de x_k dentro da sequência, é determinado pela relação das quantidades e propriedades, que tanto a primeira quanto a segunda derivada oferecem. As quantidades, determinam a velocidade do movimento, e os sinais, indicam a direção

do movimento. De certa forma, podemos interpretar o movimento da sequência $\{x_k\}$ como instantes do movimento de uma bola numa ladeira, que no começo de sua descida é acelerada, e, conforme chega ao plano no fim da ladeira, começa a reduzir sua velocidade, até supostamente chegar no ponto mais baixo.

2.2 Outros Métodos

2.2.1 Métodos de Gradiente Descendente

O método de Newton na verdade abriu espaço para uma família de métodos similares, que como visto, podem tanto ser métodos para encontrar raízes de funções quando máximos/mínimos. São estes chamados de métodos quasi-Newton, que se olharmos tomando como o ponto de partida o método de Newton, eles oferecem maior flexibilidade em como se deseja gastar os recursos de precisão e velocidade computacional.

O método de Newton, que encontra o $\min(f(x))$ através de uma sequência $\{x_k\}$:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}. \quad (2.12)$$

Pode ser reescrito da seguinte forma:

$$x_{k+1} = x_k - \frac{1}{f''(x)} * f'(x_k). \quad (2.13)$$

Como sabemos que quem determina a direção onde a função é $f'(x)$, não é completamente necessário o uso de $\frac{1}{f''(x)}$. O que a maioria dos métodos quasi-Newton fazem é substituir por aproximações boas o suficientes:

$$x_{k+1} = x_k - \alpha * f'(x_k). \quad (2.14)$$

Onde α satisfaz:

$$f(x_k - \alpha * f'(x_k)) < f(x_k). \quad (2.15)$$

A partir da equação 2.14, podemos escolher um α de modo que é menos custoso encontrar esse α do que calcular a segunda derivada da função objetivo, ou até podemos nem calcular nada, basta considerar um α fixo e tão pequeno que quando a restrição apresentada em 2.15 não for cumprida, temos uma aproximação boa o suficiente para o mínimo da função.

2.2.2 Simplex

O método Simplex pode ser considerado classificado como clássico por ser um dos métodos mais famosos. Formulado por George B. Dantzig, fruto de uma sugestão de outro homem, T. S. Motzkin, que contribuiu para diversas áreas da matemática.

Os problemas que o método Simplex resolvem fazem parte de um conjunto de problemas de Otimização Linear (ou Programação Linear), os quais se restringem a apenas funções lineares. Além disso, tais problemas normalmente acompanham restrições sobre as entradas da função a ser otimizada.

Um problema de otimização linear pode ser resumido em:

$$Z = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad (2.16)$$

Restrições também lineares para a função objetivo:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\ &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m \\ x_1 \geq 0, x_2 \geq 0, \dots, x_n &\geq 0 \end{aligned} \quad (2.17)$$

O tipo de problema que o Simplex resolve se desenvolve como a seguir:

$$\max\{c^t x \mid Ax \leq b, x \geq 0\} \quad (2.18)$$

A forma de operação desse método é semelhante ao que uma pessoa comum faria quando se deparasse com o problema de forma gráfica. O problema, sendo completamente linear, gera retas, tais retas acabam por formar regiões de possíveis soluções, e dessas regiões queremos saber qual a melhor solução. Para isso bastaria procurar dentro de tal região onde o valor da função Z é o maior possível. Uma vez a região sendo construída por retas, ela acaba por ser convexa, como um polígono no plano ou um poliedro no espaço, o que a acaba por facilitar a busca pelo máximo dentro dessa região uma vez que basta olha os vértices de tal região.

Repara-se que o método Simplex não se utiliza de artifícios e ferramentas do Cálculo como nos outros métodos anteriormente apresentados, mas sim formas diferentes de analisar o problema, que por certos aspectos é simples. O método completo é constituído por operações em uma matriz específica (Tableau) que representa o problema, o qual não é necessário ser apresentado aqui por se distanciar da "família" de métodos de otimização que é apresentado neste documento.

2.3 Programando os Métodos

Método de Newton

A forma mais simples e mais útil de implementar o método de Newton é na forma de busca das raízes, que uma vez implementada, só precisamos por como

entrada a primeira e a segunda derivada da função que desejamos minimizar, já que o método não precisa saber qual a função de fato. A seguir temos a implementação na linguagem de programação Rust:

```
pub fn newton1x1<F>(funcao_derivada: &F, x: f64) -> (usize, f64)
where
    F: Fn(f64) -> f64,
{
    let mut entrada_atual = x;
    let maximo_iteracoes = 100;

    for iteracao_atual in 1..=maximo_iteracoes {
        let diferenca: f64 =
            funcao_derivada(entrada_atual.clone())
            /
            derive1x1(&funcao_derivada, &entrada_atual);

        println!("diferenca: {}", diferenca);
        entrada_atual -= diferenca;

        if diferenca.abs() < 0.0000001 {
            return (iteracao_atual, entrada_atual);
        }
    }

    return (maximo_iteracoes, entrada_atual);
}
```

Os parâmetros da função são:

- Uma função $f : \mathbb{R} \rightarrow \mathbb{R}$
- Uma entrada x sendo o chute inicial do ótimo.

A função "derive1x1" recebe como parâmetro uma função e um ponto, tendo como retorno a derivada da função entregue, no ponto especificado. Restringindo-se à funções do tipo $f : \mathbb{R} \rightarrow \mathbb{R}$, donde deve ser escrita pelo usuário como bem entender.

Método do Gradiente Descendente

asd asdasd

```
// Sendo  $f(x) = x^4 - 3x^3 + 2$ 
#include <bits/stdc++.h>
#define decimal long double
using namespace std;
```

```
// Calcula a derivada da funcao
decimal df(decimal x) {
    return 4 * pow(x, 3) - 9 * pow(x, 2);
}
int main() {
    // Palpite inicial
    decimal x_proximo = 5.5;
    // Variavel para iterar
    decimal x_atual;
    // Quantidade maxima de iteracoes
    int iteracoes = 100000;
    decimal alpha = 0.00001;
    // Precisao para condicao de parada
    decimal precisao = 0.000000001;
    // Comeca as iteracoes
    while(iteracoes > 0) {
        x_atual = x_proximo;
        x_proximo = x_atual - (alpha * df(x_atual));
        decimal precisao_atual = x_atual - x_proximo;
        // Se atingir a precisao desejada
        if(abs(precisao_atual) < precisao)
            break; // Para as iteracoes
        // Conclui uma iteracao
        iteracoes -= 1;
    }
    cout << "x_que_minimiza_f(x) = " << x_proximo << endl;
    // Saida do algoritmo:
    // —> x que minimiza  $f(x) = 2.25$ 
    return 0;
}
```

2.4 O Método de Newton para Várias Variáveis

Capítulo 3

Os Métodos Modernos de Otimização

3.1 Breve Relato Histórico

3.2 Métodos de Um

3.2.1 O Método - Uma breve descrição

3.2.2 Exemplos Aplicações

3.2.3 Possíveis Aplicações

3.3 Métodos de Dois

3.3.1 O Método - Uma breve descrição

3.3.2 Exemplos Aplicações

3.3.3 Possíveis Aplicações

3.4 Um com o outro

Capítulo 4

Aplicações à Mecânica Celeste

4.1 Entendendo o Problema de N Corpos

4.2 A Otimização na Mecânica

4.3 Resultados Numéricos

Capítulo 5

Demais Resultados

5.1 Outros Resultados

5.1.1

Referências Bibliográficas

- [1] *Garcia, Arnaldo. Elementos de Álgebra. Rio de Janeiro. Associação instituto Nacional de Matemática Pura e Aplicada, 2003.*
- [2] *Lima, Elon Lajes Álgebra Linear. 7ª Edição. Rio de Janeiro; IMPA, 2008.*
- [3] *Halmos, Paul R. Espaços Vetoriais de Dimensão Finita. Tradução [de] Guilherme de la Penha. Rio de Janeiro: Campus, 1978.*