Universidade Federal do Amazonas
Pró-reitoria de Pesquisa e Pós-graduação
Departamento de Apoio à Pesquisa
Programa de Bolsas de Iniciação Científica

Relatório Final

Projeto PIBIC: PIB - E/023/2011

Otimização de Funções de Várias Variáveis Reais:

Métodos Interativos Clássicos

Orientador *Bolsista*

Bolsista:	Suellen Brasil da Silva - CNPq
Orientador:	Professor Dr. Nilomar Vieira de Oliveira
Vigência da Bolsa:	01/08/2011 a 31/07/2012
Unidade Executora do Projeto:	Instituto de Ciências Exatas Departamento de Matemática

Manaus - 2012

Conteúdo

1 RESUMO **3**

2 INTRODUÇÃO **4**

3 OBJETIVOS **5**

3.1 Objetivo Geral 5

3.2 Objetivos Específicos 5

4 METODOLOGIA **6**

5 RESULTADOS **7**

5.1 Diferenciabilidade em $\mathbb{R}$ 7

5.2 Regras de derivação 8

5.3 Alguns teoremas importantes 9

5.4 Diferenciabilidade em $\mathbb{R}^n$ 13

5.5 Formas Quadrática 17

5.6 Funções cônvexas e côncavas 19

5.7 Máximos e mínimos de funções de várias variáveis 21

5.8 Métodos unidimensionais de Otimização 22

6 CONCLUSÃO **29**

7 CRONOGRAMA **30**

8 REFERÊNCIAS **31**

1 RESUMO

Este projeto dará continuidade ao estudo da otimização de funções de n variáveis, iniciado através do estudo de análise matemática no espaço $\mathbb{R}^n$.Daí será dada ênfase especial aos métodos clássicos de otimização iterativos: Métodos de Newton, Gradientes e algumas variações. A metodologia utilizada consistiu primeiramente em pesquisa bibliográfica e estudo dirigido, embasando-se nos estudos sistemáticos e específicos em alguns tópicos de análise na reta e no $\mathbb{R}^n$ fundamentados para o amadurecimento de idéias afim de desenvolver uma linguagem matemática, conciliando a escrita, a descrição e o racioncínio lógico. Ao longo destes estudos dirigidos, foram realizados seminários onde os tópicos mencionados acima foram expostos e tanto o conteúdo quanto a didática e metodologia do ensino foram avaliados pelo orientador. Todos os resultados estudados foram demostrados com rigor científico que a pesquisa matemática impõe.

2 INTRODUÇÃO

Neste relatório, consta o resultado do projeto de iniciação científica em matemática, onde o bolsista submeteu-se ao estudo de otimização de funções de n variáveis , uma das subsáreas da análise Matemática.

Os tópicos abordados no projeto, bem como seus teoremas, demonstrações e definições mais importantes estarão descritos aqui de forma clara e concisa, para um bom entendimento do leitor. Também consta no relatório o cronograma de execução, as atividades desenvolvidas e as obras consultadas para o estudo perante o projeto.

A Iniciação Científica em Matemática Pura, ao contrário das outras ciências, não ensina o bolsista a fazer pesquisa matemática mas sim prepara o aluno para alcançar este nível. A Matemática requer um árduo trabalho de formação afim de que se possa usar seus conhecimentos para produzir pesquisa. Neste projeto, a bolsista recebe orientações e treinamento no amadurecimento dos conteúdos de otimização e da linguagem científica formal escrita e falada.

3 OBJETIVOS

3.1 Objetivo Geral

O projeto de Iniciação Científica em Matemática visa uma melhor qualificação dos profissionais formados nas áreas de exatas, possibilitando aos alunos um aprendizado mais profundo e rigoroso sobre os mais variados relacionados a área em questão, incentivando-os a buscarem sempre a aumentar o conhecimento, não só aquele referente ao projeto em si, mas também aquele adquirido fora dele, durante a sua graduação, e não deixá-lo estagnado. Tal projeto visa também gerar a possibilidade de que num futuro mais próximo os seus participantes tenham maturidade acadêmica suficiente para engajar-se num curso de pós-graduação e/ou uma carreira trabalhista na educação, no comércio ou na indústria como um profissional altamente qualificado.

O objetivo desta iniciação científica é abrir um espaço para que os alunos sejam orientados e formados em um nível bem superior a média nacional. Dado o grau de abertura deste projeto, uma vez que envolve indiscriminadamente os principais cursos de ciências exatas e tecnológicas, uma oportunidade dada aos alunos, professores e futuros pesquisadores para que aprendam e aperfeiçoem os seus conhecimentos matemáticos necessários para suas próprias atividade.

3.2 Objetivos Específicos

Neste projeto será estudado inicialmente alguns tópicos de análise na reta e no $\mathbb{R}^n$ visando introduzir conceitos necessários para a melhor visualização dos metódos clássicos de otimização iterativos: Métodos de Newton, Gradientes e algumas variações.

Quanto a Análise Matemática na reta serão estudados tópicos relacionados a diferenciabilidade em $\mathbb{R}$ e referente a análise no $\mathbb{R}^n$ serão abordados temas pertinentes a diferenciabilidade no $\mathbb{R}^n$.

4 METODOLOGIA

O projeto inicia-se com o estudo de diferenciabilidade e integrabilidade em $\mathbb{R}$ e diferenciabilidade em $\mathbb{R}^n$. Estes tópicos, que embora básicos, são importantes para o bolsista e o ajudarão a compreender os próximos tópicos, que são dos metódos clássicos de otimização iterativos: Métodos de Newton, Gradientes e algumas variações.Tais conhecimentos foram absorvidos através de estudos individuais, seminários e resoluções de exercícios.

Depois do estudo desses tópicos de análisen reta e no $\mathbb{R}^n$, o orientando passa a estudar tópicos dos metódos clássicos de otimização iterativos: Métodos de Newton, Gradientes e algumas variações. Enfocando os seguintes tópicos: funções convexas, funções quadráticas definidas positivas, métodos unidimensionais de otimização, convergência local e convegência global, método do gradiente, métodos quase-Newton e métodos de otimização para funções não-diferenciáveis,como estudo dirigido pelo orientador. Para alcançar êxito nos tópicos estudados, fez-se estudos individuais e depois discursões com o orientador para sua melhor compreensão e, em seguinda, espondo-o através de seminários, além de exercícios que propiciam um melhor entendimento.

5 RESULTADOS

5.1 Diferenciabilidade em $\mathbb{R}$

Definição

Sejam $f : X \longrightarrow \mathbb{R}$ e $a \in X \cap X'$. A derivada da função f no ponto a é o limite

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}.$$

se o limite acima existir, diz-se que f é *derivável no ponto a*. Quando existe a derivada $f'(x)$ em todos os pontos $x \in X \cap X'$ diz-se que a função $f : X \longrightarrow \mathbb{R}$ é *derivável* no *conjunto* X e obtém-se uma nova função $f' : X \cap X' \longrightarrow \mathbb{R}$, $x \mapsto f'(x)$, chamada de *derivada*.

Exemplo 1 Considere a função $f : \mathbb{R} \to \mathbb{R}$, definida por

$$f(x) = x^2.$$

Ache a derivada de f no ponto $x = x_0$.

Solução Vamos calcular o limite que define a derivada de uma função em um ponto x_0. Temos

$$\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x\to x_0} \frac{x^2 - x_0^2}{x - x_0} = 2x_0.$$

Portanto, $f'(x_0) = 2x_0$. Como podemos fazer isto para qualquer ponto $x \in \mathbb{R}$ dado arbitrariamente no domínio da função f, concluímos que $f'(x) = 2x$.

Teorema 1

A fim de que $f : X \longrightarrow \mathbb{R}$ seja derivável no ponto $a \in X \cap X'$ é necessário e suficiente que exista $c \in \mathbb{R}$ tal que $a + h \in X \Rightarrow f(a+h) = f(a) + c \cdot h + r(h)$, onde

$$\lim_{h\to 0} \frac{r(h)}{h} = 0. \text{ No caso afirmativo, tem-se que } c = f'(a).$$

Demonstração

Seja $W = \{h \in \mathbb{R}; a+h \in X\}$. Então $0 \in W \cap W'$.
Supondo que $f'(a)$ exista, definimos $r : W \longrightarrow \mathbb{R}$ pondo $r(h) = f(a+h) - f(a) - f'(a) \cdot h$. Então

$$\frac{r(h)}{h} = \frac{f(a+h) - f(a)}{h} - f'(a),$$

logo $\lim_{h\to 0}\frac{r(h)}{h}=0$. A condição é, portanto, necessária. Reciprocamene, se vale a condição, então $\frac{r(h)}{h}=\frac{[f(a+h)-f(a)]}{h}-c$, logo o $\lim_{h\to 0}\frac{(f(a+h)-f(a))}{h}-c=\lim_{h\to 0}\frac{r(h)}{h}=0$, portanto $f'(a)$ existe e é igual a c.■

5.2 Regras de derivação

Teorema 2

Sejam f e g deriváveis em a e seja c uma constante. Então as funções $f+g$, $c\cdot f$ e $f\cdot g$ são deriváveis em a e têm-se

1. $(f+g)'(a)=f'(a)+g'(a)$
2. $(cf)'(a)=cf'(a)$
3. $(f\cdot g)'(a)=f'(a)g(a)+f(a)g'(a)$
4. $\left(\frac{f}{g}\right)'(a)=\frac{f'(a)g(a)-f(a)g'(a)}{[g(a)]^2}$ *com* $g(a)\neq 0$

Demonstrações

$$
\begin{aligned}
1.(f+g)'(a) &= \lim_{x\to a}\frac{[f(x)+g(x)]-[f(a)+g(a)]}{x-a}\\
&= \left[\lim_{x\to a}\frac{f(x)-f(a)}{x-a}+\frac{g(x)-g(a)}{x-a}\right]\\
&= f'(a)+g'(a)\\
&\text{logo}\quad (f+g)'(a)=f'(a)+g'(a)\blacksquare
\end{aligned}
$$

$$
\begin{aligned}
2.(cf)'(a) &= \lim_{x\to a}\frac{cf(x)-cf(a)}{x-a}\\
&= \lim_{x\to a}\frac{f(x)-f(a)}{x-a}\\
&= cf'(a),\ \text{ ou seja } (cf)'(a)=cf'(a)\blacksquare
\end{aligned}
$$

$$
\begin{aligned}
3.(f\cdot g)'(a) &= \lim_{x\to a}\frac{f(x)g(x)-f(a)g(a)}{x-a}\\
&= \lim_{x\to a}\frac{f(x)g(x)-f(a)g(x)+f(a)g(x)-f(a)g(a)}{x-a}\\
&= \lim_{x\to a}\left[\frac{f(x)-f(a)}{x-a}\cdot g(x)+f(a)\cdot\frac{g(x)-g(a)}{x-a}\right]\\
&= f'(a)g(a)+f(a)g'(a).\blacksquare
\end{aligned}
$$

$$
4.\left(\frac{f}{g}\right)'(a) = \lim_{x\to a}\frac{\frac{f(x)}{g(x)}-\frac{f(a)}{g(a)}}{x-a} = \lim_{x\to a}\frac{f(x)g(a)-f(a)g(x)}{x-a}\cdot\frac{1}{g(x)g(a)}.
$$

somando e subtraindo $f(a)g(a)$ ao numerador resulta

$$
\left(\frac{f}{g}\right)'(a) = \lim_{x\to a}\left[\frac{f(x)-f(a)}{x-a}\cdot g(a)-f(a)\cdot\frac{g(x)-g(a)}{x-a}\right]\cdot\frac{1}{g(x)g(a)}.
$$

e portanto,

$$
\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a)-f(a)g'(a)}{[g(a)]^2}.\blacksquare
$$

5.3 Alguns teoremas importantes

Teorema 3 **(Regra da Cadeia)**

Sejam $f : X \to \mathbb{R}, g : Y \to \mathbb{R}, a \in X \cap X', b \in Y \cap Y', f(X) \subset Y$ e $f(a) = b$. Se f é derivável no ponto a e g é derivável no ponto b então $g \circ f : x \longrightarrow \mathbb{R}$ é derivável no ponto a, com $(g\circ f)'(a) = g'(f(a))\cdot f'(a)$.

Demonstração

Seja uma sequência de pontos $x_n \in X - \{a\}$ com $\lim x_n = a$ e ponhamos $y_n = f(x_n)$, logo $\lim y_n = b$. Sejam $\mathbb{N}_1 = \{n \in \mathbb{N}; f(x_n) \neq f(a)\}$ e $\mathbb{N}_1 = \{n \in \mathbb{N}; f(x_n) = f(a)\}$. Se $n \in \mathbb{N}_1$ então $y_n \in Y - b$ e

$$
\frac{g(f(x_n))-g(f(a))}{x_n-a} = \frac{g(y_n)-g(b)}{y_n-b}\cdot\frac{f(x_n)-f(a)}{x_n-a}
$$

Portanto, se $\mathbb{N}_1$ é infinito, tem-se $\lim\limits_{n\in\mathbb{N}_1}\dfrac{[g(f(x_n))-g(f(a))]}{x_n-a}=g'(f(a))\cdot f'(a)$. Se $\mathbb{N}_2$ é infinito tem-se $\lim\limits_{n\in\mathbb{N}_2}\dfrac{[f(x_n)-f(a)]}{x_n-a}=0$, logo $f'(a)=0$. Ainda neste caso, tem-se $\lim\limits_{n\in\mathbb{N}_2}\dfrac{[g(f(x_n))-g(f(a))]}{x_n-a}=0=g'(f(a))\cdot f'(a)$. Como $\mathbb{N}=\mathbb{N}_1\cap\mathbb{N}_2$, resulta daí que, em qualquer hipótese, vale

$$\lim_{n\in\mathbb{N}}\frac{g(f(x_n))-g(f(a))}{x_n-a}=g'(f(a))\cdot f'(a). \blacksquare$$

Teorema 4

Seja $f:(a,b)\to\mathbb{R}$ uma função diferenciável em $x_0\in(a,b)$ então f é contínua é x_0.

Demonstração

Seja $x_0\in(a,b)$, vamos mostrar que temos $\lim\limits_{x\to x_0}f(x)=f(x_0)$. Equivalemente,

$$\begin{aligned}\lim_{x\to x_0}(f(x)-f(x_0)) &= \lim_{x\to x_0}\left(\frac{f(x)-f(x_0)}{x-x_0}(x-x_0)\right)\\ &= \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}\cdot\lim_{x\to x_0}(x-x_0)=0,\end{aligned}$$

pois, observando a última igualdade e considerando a hipótese do teorema, f é diferenciável em x_0, logo existe o primeiro limite, o qual é igual a $f'(x_0)$, e o $\lim_{x\to x_0}(x-x_0)=0$. Portanto, o produto desses limites é zero.$\blacksquare$

A recíproca desse teorema é falsa, conforme podemos comprovar a seguir.

Exemplo 2 Considere a função $f:\mathbb{R}\to\mathbb{R}$, definida por $f(x)=|x|$. Mostre que a função $f(x)=|x|$ é contínua em $x=0$, mas não existe a derivada de f nesse ponto.

Solução O jeito mais simples de mostrar essa afirmação, é através do cálculo dos limites laterais de

$$\lim_{x\to 0}\frac{f(x)-f(0)}{x-0}$$

que define a derivada de uma função, pois se esse limite existir e, conseqüentemente, a derivada da função no ponto $x=0$, então os limites laterais, pela esquerda e pela direita, devem ser iguais. Mas

$$\lim_{x\to 0,x<0}\frac{|x|-0}{x-0}=\lim_{x\to 0,\,x<0}-1=-1\neq\lim_{x\to 0,\,x>0}\frac{|x|-0}{x-0}=\lim_{x\to 0,x>0}1=1.$$

Portanto, a $f(x)=|x|$ não é diferenciável no ponto $x=0$

Teorema 5

Se f é diferenciável num intervalo aberto (a,b) e se f atinge seu máximo ou mínimo num ponto $c \in (a,b)$ então $f'(c) = 0$.

Demonstração

Suponhamos que f atinge seu máximo em c. Isto é, $f(x) \leq f(c)$ para todo $x \in (a,b)$. Seja (x_n) uma seqüência convergindo para c tal que $a < x_n < c$ para todo n. (Como $a < c$ podemos, por exemplo, tomar $x_n = c - \frac{1}{n}$ para n suficientemente grande.) Então, como f é diferenciável em c, temos que a seqüência

$$\frac{f(x_n) - f(c)}{x_n - c}$$

converge para $f'(c)$. Podemos notar que cada termo nesta seqüência de quocientes é não-negativo, pois $f(x_n) \leq f(c)$ e $x_n < c$. Assim, $f'(c) \geq 0$.

Da mesma forma, seja (y_n) uma seqüência convergindo para c tal que $c < y_n < b$ para todo n. Então os termos da seqüência

$$\frac{f(y_n) - f(c)}{y_n - c}$$

são todos não-positivos, pois $f(y_n) \leq f(c)$ e $y_n > c$. Como a seqüência dos quocientes novamente converge para $f'(c)$, devemos ter $f'(c) \leq 0$.

Portanto, concluímos que $f'(c) = 0$.

O caso em que f atinge um mínimo em c é demonstrado de maneira análoga. ■

Teorema 6 **(Teorema de Rolle)**

Se f é uma função contínua em $[a,b]$ e diferenciável em (a,b) e $f(a) = f(b) = 0$ então existe ao menos um ponto $c \in (a,b)$ tal que $f'(c) = 0$.

Demonstração

Como f é contínua e $[a,b]$ é um conjunto compacto, temos que existem x_1 e x_2 em $[a,b]$ tais que $f(x_1) \leq f(x) \leq f(x_2)$ para todo $x \in [a,b]$. Se x_1 e x_2 forem ambos as extremidades do intervalo $[a,b]$ então $f(x) = f(a) = f(b)$ para todo $x \in [a,b]$. Neste caso, f é a função constante e $f'(x) = 0$ para *todos* $x \in (a,b)$. Caso contrário, f atinge um máximo ou um mínimo em algum ponto $c \in (a,b)$. Então, pelo teorema anterior, $f'(c) = 0$. ■

Teorema 7-Teorema do valor médio

Se f é uma função contínua em $[a,b]$ e diferenciável em (a,b) então existe ao menos um ponto $c \in (a,b)$ tal que

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

Demonstração

Vamos definir uma função h que satisfaz as hipóteses do teorema de Rolle. Para este fim, considere, inicialmente, a função $g : [a,b] \to \mathbb{R}$, definida por

$$g(x) = \frac{f(b)-f(a)}{b-a}(x-a) + f(a), \qquad \text{para todo } x \in [a,b].$$

Então a função $h = f - g$ é contínua em $[a,b]$ e diferenciável em (a,b). Como $h(a) = h(b) = 0$, h satisfaz as hipóteses do teorema de Rolle. Assim, existe um ponto $c \in (a,b)$ tal que $h'(c) = 0$. Logo,

$$0 = h'(c) = f'(c) - g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a}. \qquad ■$$

Fórmula de Taylor (LEMA)

Seja $r : J \longrightarrow \mathbb{R}^n$ vezes derivável no ponto $0 \in J$. A fim de que seja $r^{(i)}(0) = 0$ para $i = 0,1,2,...,n$ é necessário e suficiente que $\lim_{h \to 0} \frac{r(h)}{h^n} = 0$.

Teorema 8 (Fórmula de Taylor infinitesimal)

Seja $f : I \longrightarrow \mathbb{R}$ n vezes derivável no ponto $a \in I$. A função $r : J \longrightarrow \mathbb{R}$, definida no intervalo $J = \{h \in \mathbb{R}; a+h \in I\}$ pela igualdade

$$f(a+h) = f(a) + f'(a) \cdot h + \frac{f''(a)}{2} \cdot h^2 + ... + \frac{f^{(n)}(a)}{n!} \cdot h^n + r(h),$$

cumpre $\lim_{h\to 0}\frac{r(h)}{h^n}=0$. *Reciprocamente, se* $p(h)$ *é um polinômio de grau* $\leq n$ *tal que* $r(h)=f(a+h)-p(h)$ *cumpre* $\lim_{h\to 0}\frac{r(h)}{h^n}=0$ *então* $p(h)$ *é o polinômio de Taylor de ordem* n *de* f *no ponto* a*, isto é,*

$$p(h)=\sum_{i=0}^{n}\frac{f^i(a)}{i!}\cdot h^i.$$

Demonstração

A função r, definida pela fórmula de Taylor, é n vezes derivável no ponto 0 e tem derivadas nulas nesse ponto, até a ordem n. Assim pelo lema anterior, vale $\lim_{h\to 0}\frac{r(h)}{h^n}=0$. Reciprocamente, se $r(h)=f(a+h)-p(h)$ é tal que $\lim_{h\to 0}\frac{r(h)}{h^n}=0$ então, novamente pelo lema anterior, as derivadas de r no ponto 0 são nulas até a ordem n, logo $p^{(i)}(0)=f^{(i)}(a)$ para $i=0,1,2,...,n$ ou seja, p(h) é o polinômio de Taylor de ordem n da função f no ponto a.

5.4 Diferenciabilidade em $\mathbb{R}^n$

Definição

Derivadas parciais

Seja $f:U\longrightarrow\mathbb{R}$ uma função definida no aberto $U\subset\mathbb{R}^n$. Para cada $i=1,...,n$, a *i-ésima derivada parcial* de f no ponto $a=(a_1,...,a_n)\in U$ é o número

$$\frac{\partial f}{\partial x_i}(a)\lim_{t\to 0}\frac{f(a+te_i)-f(a)}{t}=\lim_{t\to 0}\frac{f(a_1,...,a_i+t,...,a_n)-f(a)}{t},$$

caso este limite exista. Como U é aberto, podemos achar $\delta>0$ tal $a+te_i\in U$ para todo $t\in(\delta,-\delta)$.Então está bem definido o caminho retílineo $\lambda:(\delta-\delta)\to U,\lambda(t)=a+te_i$. A definição acima diz que $\frac{\partial f}{\partial x_i}(a)(f\circ\lambda)'(0)=$ derivada, no ponto $t=0$, da função real $f\circ\lambda:(-\delta,\delta)\to\mathbb{R}$.

Funções de classe C^1

Seja $f:U\longrightarrow\mathbb{R}$ uma função que possui as n derivadas parciais em todos os pontos do aberto $U\subset\mathbb{R}^n$. Ficam então definidas n funções

$$\frac{\partial f}{\partial x_1},...,\frac{\partial f}{\partial x_n}:U\longrightarrow\mathbb{R},\text{ onde }\frac{\partial f}{\partial x_i}:x\longrightarrow\frac{\partial f}{\partial x_i}(x)$$

Se estas funções forem contínuas em U, diremos que f é uma *função de classe* C^1 e dizemos que $f \in C^1$.

Teorema 1

Toda função $f : U \longrightarrow \mathbb{R}$ *de classe* C^1 *é diferenciável.*

Demostração

Para simplificar, suporemos $U \subset \mathbb{R}^2$.O caso geral trata-se apenas de uma notação mais elaborada.

Vamos fixar um ponto $c = (a,b) \in U$ e tomar $v = (h,k)$ tal que $c + v \in B \subset U$, onde B é uma bola com centro em c. Seja

$$r(c) = f(a+h, b+k) - f(a,b) - \frac{\partial f}{\partial x} \cdot h - \frac{\partial f}{\partial y} \cdot k,$$

onde as derivadas são calculadas no ponto $c = (a,b)$. Podemos escrever

$$r(v) = f(a+h, b+k) - f(a, b+k) + f(a, b+k) - f(a,b) - \frac{\partial f}{\partial x} \cdot h - \frac{\partial f}{\partial x} \cdot k.$$

Pelo teorema do valor médio de uma variável real, existem $\theta_1, \theta_2 \in (0,1)$ tais que

$$r(v) = \frac{\partial f}{\partial x}(a + \theta_1 h,\ b+k) \cdot h + \frac{\partial f}{\partial y}(a,\ b+\theta_2) \cdot k - \frac{\partial f}{\partial x} \cdot h - \frac{\partial f}{\partial y} \cdot k.$$

logo

$$\frac{r(v)}{\|v\|} = \left[\frac{\partial f}{\partial x}(a+\theta_1 h,\ b+k) - \frac{\partial f}{\partial x}(a,\ b)\right] \frac{h}{\sqrt{h^2+k^2}} + \left[\frac{\partial f}{\partial y}(a,\ b+\theta_2 k) - \frac{\partial f}{\partial y}(a,\ b)\right] \frac{k}{\sqrt{h^2+k^2}}.$$

Quando $v \longrightarrow 0$ os termos dentro do colchete acima tendem a zero, pela continuidade das derivadas $\dfrac{\partial f}{\partial x}$ e $\dfrac{\partial f}{\partial y}$. Além disso, os termos fora dos colchetes tem valor absoluto ≤ 1. Portanto $\lim\limits_{v \longrightarrow 0} \dfrac{r(v)}{\|V\|} = 0$ e então f é diferenciável. $\blacksquare$

Teorema 2

Seja $f : U \longrightarrow \mathbb{R} \longrightarrow$ *diferenciável no aberto* $U \subset \mathbb{R}^n$*, com* $a \in U$*. Dado o vetor* $v = (\alpha_1, ..., \alpha_n)$*, se* $\lambda : (-\delta,\ \delta) \longrightarrow U$ *é qualquer caminho diferenciável tal que* $\lambda(0) = a$ *e* $\lambda'(0) = v$*, tem-se*

$$(f \circ \lambda)'(0) = \langle grad f(a),\, v\rangle \frac{\partial f}{\partial v}(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) \cdot \alpha_i.$$

basta aplicar a fórmula diretamente

$$(f \circ \lambda)' = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot \frac{d\lambda_i}{dt},$$

observando que , para $\lambda(t) = (\lambda_1(t), ..., \lambda_n(t))$, tem-se $\alpha_i = \dfrac{d\lambda_i}{dt}(0)$.

Notarmos ainda que $\dfrac{\partial f}{\partial v}(a) = (f \circ \lambda)'(0)$ com $\lambda(t) = a + tv$, pois $\lambda'(0) = v$.

Teorema do valor médio

Dada $f : U \longrightarrow \mathbb{R}$ *direnciável no aberto* $U \subset \mathbb{R}^n$ *, se o segmento de reta* $[a, a+v]$ *estiver contido em* U *então existe* $\theta \in (0,1)$ *tal que*

$$f(a+v) - f(a) = \frac{\partial f}{\partial v}(a + \theta v) = \langle grad f(a+\theta v),\, v\rangle$$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a+\theta v) \cdot \alpha_i$$

onde $v = \alpha_1, ..., \alpha_n$*.*

Com efeito, considerando o caminho retílineo $\lambda : [0,1] \longrightarrow U$, dado por $\lambda(t) = a + tv$, vemos que $f(a+v) - f(a) = (f \circ \lambda)(1) - (f \circ \lambda)(0)$. Pelo teorema do valor médio para funções de uma variável real, existe $\theta \in (0,1)$ tal que $(f \circ \lambda)(1) - (f \circ \lambda)(0) = (f \circ \lambda)'(\theta)$. Pela regra da cadeia,

$$(f \circ \lambda)'(\theta) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a+\theta v) \cdot \alpha_i = \frac{\partial f}{\partial v}(a+\theta v) = \langle grad f(a+\theta v),\, v\rangle . \blacksquare$$

A fórmula de Taylor (LEMA 1)

Seja $r : B \longrightarrow \mathbb{R}$ *de classe* C^2 *na bola aberta* $B \subset \mathbb{R}^n$*, de centro* 0*.*
Se $r(0) = \dfrac{\partial f}{\partial x_i}(0) = \dfrac{\partial^2 r}{\partial x_i \partial x_j}(0) = 0$ *para quaiquer* $i, j = 1, ..., n$*, então* $\lim\limits_{v \to 0} \dfrac{r(v)}{\|v\|^2} = 0$*.*

Demostração

Seja $r : B \longrightarrow \mathbb{R}$ uma função de classe C^1, isso quer dizer que é diferenciável, que se anula com todas as suas derivadas $\dfrac{\partial r}{\partial x_i}$ no ponto $v = 0$. Pelo teorema do valor médio, para cada $v = (\alpha_1, ..., \alpha_n) \in B$ existe θ tal que $0 < \theta < 1$ e

$$r(v) = \sum_{i=1}^{n} \frac{\partial r}{\partial x_i}(\theta v) \cdot \alpha_i, \text{ logo } \frac{r(v)}{\|v\|^2} = \sum_{i=1}^{n} \frac{\frac{\partial r}{\partial x_i}(\theta v)}{\|\theta v\|} \cdot \frac{\theta \alpha_i}{\|v\|}.$$

Como cada derivada parial $\dfrac{\partial r}{\partial x_i}$ se anula, justamente com todas as suas derivadas $\dfrac{\partial^2 r}{\partial x_j x_i}$ no ponto 0, resulta da observação inicial que

$$\lim_{v \to 0} \left[\frac{\frac{\partial r}{\partial x_i}(\theta v)}{\|\theta v\|} \right] = 0 \text{ para todo } i = 1, ..., n.$$

Além disso, cada quociente $\dfrac{\theta \alpha_i}{\|v\|}$ tem valor absoluto ≤ 1. Assim $\lim\limits_{v \to 0} \dfrac{r(v)}{\|v\|^2} = 0$.■

Teorema (fórmula de Taylor)

Seja $f : U \longrightarrow \mathbb{R}$ de clsse C^2 no aberto $U \subset \mathbb{R}^n$.Fixando $a \in U$, para todo $v = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$ tal que $a + v \in U$, escrevemos

$$f(a+v) - f(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot \alpha_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} \cdot \alpha_i \alpha_j + r(v),$$

as derivadas sendo calculadas no ponto a. Então $\lim\limits_{v \to 0} \dfrac{r(v)}{\|v\|^2} = 0$.

Demostração

Do lema 1 Iremos demonstrar que

$$r(v) = f(a+v) - f(a) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} \cdot \alpha_i \alpha_j$$

se anula, quando suas derivadas parciais de primeira e segunda ordem, no ponto $v = 0$.

Primeiro vamos começar lembrando que as variáveis independentes são as coordenadas $\alpha_1, ..., \alpha_2$ de v. É em relação a elas que as derivadas parciais de r devem ser tomadas, mesmo continuando escrevendo $\frac{\partial r}{\partial x_i}$ e $\frac{\partial^2 r}{\partial x_i \partial x_j}$. Observemos também que, no somatório duplo que ocorre na definição de $r(v)$, cada par de variáveis α_i e α_j aparece em duas parcelas iguais, a saber, $\frac{\partial^2 f}{\partial x_j \partial x_i} \cdot \alpha_j \alpha_i$ e $\frac{\partial^2 f}{\partial x_i \partial x_j} \cdot \alpha_i \alpha_j$. Levando isto em conta, temos:

$$\frac{\partial r}{\partial x_j}(v) = \frac{\partial f}{\partial x_j}(a+v) - \frac{\partial f}{\partial x_j}(a) - \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \cdot \alpha_i.$$

Derivando outra vez, temos:

$$\frac{\partial^2 r}{\partial x_i \partial x_j}(v) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a+v) - \frac{\partial^2 f}{\partial x_i \partial x_j}(a).$$

Consequentemente $r(0) = 0$, $\frac{\partial r}{\partial x_i}(0) = 0$ e $\frac{\partial^2 f}{\partial x_i \partial x_j}(0) = 0$ para quaisquer $i, j = 1, ..., n$ ■.

5.5 Formas Quadrática

Uma *forma quadrática* $H : \mathbb{R}^n \longrightarrow \mathbb{R}$ é uma função cujo valor no vetor $v = (\alpha_1, ..., \alpha_2)$ é ln $\sum_{i,j=1}^{n} h_{i,j} \alpha_i \alpha_j$, onde $[h_{i,j}]$ é uma matriz simétrica $n \times n$. O valor da forma quadrática H no vetor v será indicado com a notação $H \cdot v^2$ .Portanto

$$H \cdot v^2 = \sum_{i,j=1}^{n} h_{ij} \alpha_i \alpha_j \text{ quando } v = (\alpha_1, ..., \alpha_n)$$

Se $t \in \mathbb{R}$ então $H \cdot (tv)^2 = t^2 \cdot (H \cdot v^2)$.

A forma quadrática H chama-se *não-negativa* quando $H \cdot v^2 > 0$ para todo $v \neq 0$ em $\mathbb{R}^n$ e *indefinida* quando existem $v, w \in \mathbb{R}^n$ tais que $H \cdot v^2 > 0$ e $H \cdot w^2 < 0$. De modo análogo se definem forma quadrática *negativa* e *não-negativa*. Quando H é positiva ou negativa, diz-se que ela é *definida*.

Hessiana

Dada a função $f : U \longrightarrow \mathbb{R}$, de classe C^2 no aberto $U \subset \mathbb{R}^n$, a *forma quadrática hessiana* $H(x) = (Hf)(x)$ de f no ponto $x \in U$ é aquela cuja matriz é $[h_{ij}] = \left[\dfrac{\partial^2 f}{\partial x_i \partial x_j}(x)\right]$. Assim, para todo $v = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$, tem-se

$$H(x) \cdot v^2 = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \cdot \alpha_i \alpha_j$$

A forma hessiana é usada para determinar a natureza dos pontos críticos da função f.

Matrizes definidas e semidefinidas

Definição:

Sejam A uma matriz $n \, x \, n$ e $Q(h) = h^T \cdot A \cdot h$ a forma quadrática associada.

1. A matriz A (ou a forma quadrática Q) é *positiva definida* se

$$Q(h) = h^T \cdot A \cdot h > 0$$

   para todo $h \neq 0$ em $\mathbb{R}^n$

2. A matriz A (ou a forma quadrática Q) é *positiva semidefinida* se

$$Q(h) = h^T \cdot A \cdot h \geq 0$$

   para todo $h \neq 0$ em $\mathbb{R}^n$.

3. A matriz A (ou a forma quadrática Q) é *negativa definida* se

$$Q(h) = h^T \cdot A \cdot h < 0$$

   para todo $h \neq 0$ em $\mathbb{R}^n$.

4. A matriz A (ou a forma quadrática Q) é *negativa semidefinida* se

$$Q(h) = h^T \cdot A \cdot h \geq 0$$

   para todo $h \neq 0$ em $\mathbb{R}^n$.

5. A matriz A (ou a forma quadrática Q) é *indefinida* se existem h e k em $\mathbb{R}^n$ tais que

$$Q(h) = h^T \cdot A \cdot h > 0 \text{ e } Q(k) = k^T \cdot A \cdot k > 0$$

5.6 Funções cônvexas e côncavas

Definição:

1. Dizemos que uma função $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ definida em um subconjunto convexo U de $\mathbb{R}^n$ é *cônvexa* se, somente se,

$$f((1-t)\cdot p + t\cdot q) \leq (1-t)\cdot f(p) + t\cdot f(q),$$

para todo $p,\ q \in U$ e todo $t \in [0,1]$.

2. Dizemos que uma função $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ definida em um subconjunto convexo U de $\mathbb{R}^n$ é *côncava* se, somente se,

$$f((1-t)\cdot p + t\cdot q) \geq (1-t)\cdot f(p) + t\cdot f(q),$$

para todo $p,\ q \in U$ e todo $t \in [0,1]$.

Observações importantes:

1. Dizer que uma função f é convexa em um conjunto convexo U, isto é, dizer que f satisfaz a expressão

$$f((1-t)\cdot p + t\cdot q) \leq (1-t)\cdot f(p) + t\cdot f(q),$$

para todo $p,q \in U$ e todo $t \in [0,1]$, significa dizer que o segmento de reta secante que passa pelos pontos $(p, f(p))$ e $(q, f(q))$ sempre está *acima* ou *coincide* com o gráfico de f.

2. Da mesma forma dizer que uma função f é côncava em um conjunto convexo U, isto é, dizer que f satisfaz a expressão

$$f((1-t)\cdot p + t\cdot q) \geq (1-t)\cdot f(p) + t\cdot f(q),$$

para todo $p,q \in U$ e todo $t \in [0,1]$ é nada mais que dizer que o segmento de reta secante que passa pelos pontos $(p, f(p))$ e $(q, f(q))$ sempre está *abaixo* do gráfico de f.

Teorema 1

Seja $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função de classe C^1 definida em um subconjunto convexo U de $\mathbb{R}^n$.

1. f é uma função cônvexa em U se, e somente se,

$$\text{f}(q) \geq f(p) + Df(p) \cdot (q - p),$$

   para todo $p, q \in U$.

2. f é uma função côncava em U se, e somente se,

$$f(q) \leq f(p) + Df(p) \cdot (q - p),$$

   para todo $p, q \in U$.

Corolário 1

Sejam $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função de classe C^1 definida em um subconjunto convexo U de $\mathbb{R}^n$ e $p \in U$ um ponto crítico de f.

1. Se f é uma função cônvexa em U, então p é um ponto de mínimo global de f em U.

2. Se f é uma função côncava em U, então p é um ponto de máximo global de f em U.

Teorema 2

Seja $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função de classe C^2 definida em um subconjunto convexo e aberto U de $\mathbb{R}^n$.

1. f é uma função cônvexa em U se, e somente se, a matriz hessiana $D^2 f(p)$ é positiva semidefinida para todo $p \in U$.

2. f é uma função côncava em U, se somente se, a matriz hessiana $D^2 f(p)$ é negativa semidefinida para todo $p \in U$.

Exemplo 1

Considere a função $z = f(x,y) = x^4 + x^2y^2 + y^4 - 3x - 8y$ definida em $\mathbb{R}^2$. A matriz hessiana de f é dada por

$$D^2 f(x,y) = \begin{bmatrix} 12x^2 + 2y^2 & 4xy \\ 4xy & 2x^2 + 12y^2 \end{bmatrix}$$

Uma vez que os dois menores principais de ordem 1,

$$12x^2 + 2y^2 \ \text{ e } \ 2x^2 + 12y^2,$$

e o menor principal de ondem 2,

$$24x^4 + 132x^2y^2 + 24y^4,$$

de $D^2 f(x,y)$ são maiores ou iguais a zero para todo (x,y) em $\mathbb{R}^2$, segue-se pelo teorema anterior número 2 que f é uma função cônvexa em $\mathbb{R}^2$.

Exemplo 2

Considere a função $z = f(x,y) = xy$ definida em $\mathbb{R}^2$. A matriz hessiana de f é dada por

$$D^2 f(x,y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Uma vez que o menor principal de ordem 2 de $D^2 f(x,y)$ é igual a -1 para todo $(x,y) \in \mathbb{R}^2$, segue-se pelo teorema 2 que f não é uma função cônvexa e nem uma função côncava em $\mathbb{R}^2$.

5.7 Máximos e mínimos de funções de várias variáveis

Definição:

Considere uma função de n variáveis

$$f : D_f \subset \mathbb{R}^n \longrightarrow \mathbb{R}$$

1. Dizemos que $p \in D$ é um *ponto de máximo global* de f em D se

   $$f(x) \leq f(p)$$

   para todo $x \in D$. Neste caso, o número real $f(p)$ é denominado o *valor máximo* de f.

2. Dizemos que $p \in D$ é um ponto *ponto de mínimo global* de f em D se
$$f(x) \geq f(p)$$
para todo $x \in D$. Neste caso, o número real $f(p)$ é denominado o *valor mínimo* de f.

3. Dizemos que $p \in D$ é um *ponto de máximo local* de f em D se existe uma bola aberta B (p, r) de centro em p e raio $r > 0$ tal que
$$f(x) \leq f(p)$$
para todo $x \in B(p,\ r) \cap D$

4. Dizemos que $p \in D$ é um ponto de *ponto de mínimo local* de f em D se existe uma bola aberta B (p, r) de centro em p e raio $r > 0$ tal que
$$f(x) \geq f(p)$$
para todo $x \in B(p,\ r) \cap D$

5. Dizemos que $p \in D$ é um ponto de *extremo global* de f em D se p é um ponto de máximo global ou um ponto de mínimo global de f em D.

6. Dizemos que $p \in D$ é um *extremo local* de f em D se p é um ponto de máximo local ou um ponto de mínimo local de f em D.

Observação:

A função f é chamada de *função-objetivo* e o conjunto D de *conjunto admissível* do problema de otimização.

5.8 Métodos unidimensionais de Otimização

O estudo de métodos de otimização unidimensionais sem restrições (OUSR) sao necessário por quatro motivos:

1. Alguns problemas sem restrição são também unidimensionais.
2. Em muitos problemas as restrições podem ser incorporadas à função objetivo reduzindo, desta forma, o problema a uma variável.
3. As técnicas para problemas multidimensionais com e sem restrições geralmente envolvem repetir inúmeras vezes a resolução de um problema unidimensional sem restrições.
4. Ajuda na solução de problemas de otimização cujo intervalo de busca (região viável) não é conhecido.

Podemos classificar os métodos numéricos para resolução dos problemasde OUSR em

1. Métodos diretos (MD's): procedimentos de busca do ponto ótimo através da comparação direta do valor assumido pela funçao objetivo em uma seqüência de pontos viáveis. Não envolve o cálculo de derivadas analíticas ou numéricas.
2. Métodos Indiretos (MI's): utiliza as condições necessárias para existênciade pontos estacionários, ou seja, usa a derivada para cálcular o ponto ótimo. .

Método de Newton no $\mathbb{R}^n$

Seja o problema de resolver o sistema de equações não-lineares $F(x) = 0$, onde

$$F(x) = (f_1(x)\ f_2(x)\ .....f_n(x))$$

com $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

No qual a matriz jacobiana da função F é denotada por:

$$J(x) = (\bigtriangledown f_1'(x)\ \ \bigtriangledown f_2'(x)\ \ ....\bigtriangledown f_n'(x))^t.$$

Um dos métodos para resolver $F(x) = 0$ é o método de Newton. Esse método é baseado na solução aproximada de problemas.

O k-ésimo problema é dado pela aproximação $F(x)$ pela série de Taylor em torno do ponto x_k:

$$F(x) \approx L_k(x) = F(x_k) + J(x_k)(x - x_k).$$

O ponto seguinte x_{k+1} é a solução de:

$$L_k(x) = F(x_k) + J(x_k)(x - x_k) = 0$$

.

Se $J(x_k)$ é não singular temos que $L_{k(x)}$ = 0 tem uma única solução, portanto o método de Newton em cada interação k consiste em resolver o seguinte sistema linear:

$$J(x_k)S_k = -F(x_k)$$

$$x_{k+1} = x_k + S_k$$

Dessa forma o processo gera uma sequência que converge para a solução do problema.

As vantagens do método de Newton são:

1. O procedimento tem convergência quadrática local se $\nabla^2 \neq 0$ no ponto ótimo.
2. Para uma função objetivo quadrática o mínimo é obtido em uma iteração, pois a expansão e truncamento da série de Taylor da função $f(x)$ até o termo da 2^a derivada (inclusive) é exata.
3. Converge rapidamente quando a estimativa inicial é boa.

As desvantagens do método de Newton são:

1. Só se aplica a funções onde existam ∇f e ∇^2
2. Deve-se calcular a cada iteração ∇f e ∇^2
3. Sensível à estimativa inicial.
4. Se $\nabla^2 \rightarrow 0$ a convergência é lenta.
5. Se existe mais de um extremo o método pode não convergir para o ponto desejado ou pode oscilar.

Método do Gradiente

Derivada Direcional

Sejam $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função definida em um subconjunto $D \in \mathbb{R}^n$, p um ponto interior de D e v um vetor em $\mathbb{R}^n$. A *derivada direcional* de f no ponto p na direção de v é definida por

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

caso o limite exista.

Vetor gradiente

Considere uma função escalar $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ de clsse C^1 e p um ponto do domínio D de f.

$$\nabla f(p) = \left(\frac{\partial f}{\partial x_1}(p), \frac{\partial f}{\partial x_2}(p), \dots, \frac{\partial f}{\partial x_n}(p) \right).$$

O *campo gradiente* de f é a função vetorial

$$\nabla f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$p \longmapsto \nabla f(p) = \left(\frac{\partial f}{\partial x_1}(p), \frac{\partial f}{\partial x_2}(p), \dots, \frac{\partial f}{\partial x_n}(p) \right).$$

Algumas propriedades do gradiente

Fixando $a \in U$, suporemos que $f \in C^1$ e que grad $f(a) \neq 0$. Então:

1. O gradiente aponta para uma direção segundo a qual a função é crescente;
2. Dentre todas as direções ao longo das quais f cresce, a direção do gradiente é a de crescimento mais rápido;
3. O gradiente de f no ponto a é ortogonal ao conjunto de nível de f que passa por a.

No caso do método do gradiente (ou declive máximo), a direção de descida escolhida é

$$r^{(n)} = -\nabla f(x^{(n)}) = b - Ax^{(n)}.$$

que neste caso de sistemas lineares é também designado por resíduo. Resta encontrar o valor n que minimiza f, de entre os possíveis valores $x^{(n)} + r^{(n)}$. Encontramos o ponto de mínimo, derivando f (nesses pontos) em ordem a α , ou seja,

$$\begin{aligned}
\frac{d}{d\alpha}\, f(x^{(n)} + \alpha r^{(n)}) &= \nabla f(x^{(n)} + \alpha r^{(n)}) \cdot r^{(n)} \\
&= (b^{(n)} - A(x^{(n)} + \alpha r^{(n)}) \cdot r^{(n)}) \\
&= (b - Ax^{(n)}) \cdot r^{(n)} - \alpha r^{(n)} \cdot r^{(n)} \\
&= = r^{(n)} \cdot r^{(n)} - \alpha r^{(n)} \cdot r^{(n)}.
\end{aligned}$$

Assim, o valor mínimo n será obtido com o zero da derivada,

$$r^{(n)} \cdot r^{(n)} - \alpha^{(n)} A r^{(n)} \cdot r^{(n)} = 0 \Leftrightarrow \alpha^{(n)} = \frac{r^{(n)} . r^{(n)}}{(A\ \alpha r^{(n)}\ \cdot\ r^{(n)})}.$$

Em conclusão, dado um vetor inicial $x^{(0)}$, o método do gradiente resume-se à iteração

$$x^{(n+1)} = x^{(n)} + \frac{r^{(n)}}{\cdot} r^{(n)} r^{(n)} \cdot A r^{(n)}\ r^{(n)} \text{ com } r^{(n)} = b - Ax^{(n)}$$

Um critério de paragem consiste em exigir que $\|r^{(n)}\|^2 = r^{(n)} \cdot r^{(n)} < \varepsilon$, com ε pequeno, notando que isso implica que $Ax^{(n)}$ é próximo de b.

Observação: Aqui o método do gradiente está aplicado ao caso de sistemas lineares, em que a função f é apenas auxiliar, nem tão pouco aparece na expressão do método. No entanto, este método pode ser utilizado para minimizar funções diferenciáveis, nesse caso mais geral, $r^{(n)} = -\nabla f(x^{(n)})$.

Métodos Quase-Newton

Os Métodos quase-Newton necessitam apenas que o gradiente da função objetivo esteja disponível em cada iteração. Ao medir as mudanças no gradiente de uma iteração para outra, eles tentam construir um modelo para que a função objetivo seja boa o bastante para produzir convergencia superlinear.

Por não necessitar de segundas derivadas, os métodos quase-Newton podem ser até mais eficientes do que o método de Newton em alguns casos.

Seja o seguinte *problema* :

$$\text{minimizar } f(x) \qquad (1)$$

onde a $f \in C^2, x \in \mathbb{R}^n$ e $F \in \mathbb{R}^{n\times n}$ a hessiana de $f(F\ \nabla^2\ f)$

O método de Newton modificado consiste em encontrar um novo ponto a cada iteração da seguinte forma:

$$x_{k+1} = x_k - t_k S_k\, \nabla\, f(x_k) \qquad (2)$$

onde $S \in \mathbb{R}^{n\times n}$ é uma matriz simétrica , $\nabla f \in \mathbb{R}^n$ o gradiente da função no ponto e t_k é escolhido de tal forma que minimize $f(x_{k+1})$. Se $S = F^{-1}$ for a inversa da Hessiana temos o método de Newton e se $S = I$, onde I é a matriz identidade, nos temos o *steepest descent*.

Através dos métodos Quase-Newton é possível obter uma aproximaçao $H \in \mathbb{R}^{n\times n}$ da inversa da matrix Hessiana, ao invés da exata exigida nos métodos de Newton. Essa aproximação é feita obedecendo a condição secante descrita a seguir:

$$H_{k+1}\, y_k = S_k \qquad (3)$$

Onde, de acordo com o problema sem restrições 1, temos os seguintes vetores S_k e y_k:

$$S_k = x_k - x_{k+1} \qquad (4)$$

$$y_k = \nabla f(x_k) - \nabla f(k+1) \qquad (5)$$

1) **Método Quase-Newton do tipo DFP**

Nos anos 50, a necessidade de um algoritmo que acelerasse as iterações durante a resolução de problemas de minimização do tipo (1), isto é, que resolvesse rapidamente com custo computacional reduzido fez com que Davidon desenvolvesse o primeiro algoritmo Quase-Newton que deu origem ao DFP.

DFP (Davidon, Fletcher e Powell), um dos primeiros métodos a construir uma aproximação da inversa da hessiana, foi originalmente proposto por Davidon em 1959, e posteriormente desenvolvido por Fletcher e Powell em 1963.

Seja $H \in \mathbb{R}^{n\times n}$ a aproximação inversa da hessiana. Entao, em cada interação temos a seguinte atualização:

$$H_{k+1} = H_k + \frac{S_k S_k^t}{S_k^t y_k} - \frac{H_k y_k y_k^t H_k}{y_k^t H_k y_k} \qquad (6)$$

Se consideramos a matriz B a aproximação da Hessiana de tal forma que $B = H^{-1}$,temos em (7) seguinte Condição Secante, também conhecida como Condição Quase-Newton.

$$B_{k+1} S_k = y_k \qquad (7)$$

onde (3) e (7) são iguais.

Assim podemos considerar a seguinte atualização na matriz B

$$B_{k+1} = B_k - \frac{B_k S_k S_k^T B_k}{S_k B_k S_k^T} + \frac{y_k y_k^T}{y_k^T S_k} \qquad (8)$$

A equação (8) é denominada atualização BFGS da matriz B.

2) **Método Quase-Newton do tipo BFGS**

O mais popular dos métodos Quase-Newton é o BFGS, denominado dessa maneira para referir-se aos idealizadores da técnica: Broyden, Fletcher, Goldfarb e Shanno. A atualização BFGS da matriz B_{k+1} é apresentada em (8), onde S_k e y_k são descritos em (3) e (4) , respectivamente.

Em problemas sem restrições, a atualização BFGS produzirá uma matriz B_{k+1} simétrica definida positiva sempre que a matriz B_k também seja definida positiva e que se verifique, além da Condição Quase-Newton (7), a seguinte condição de curvatura:

$$S_k^T y_k > 0 \qquad (9)$$

Em problemas com restrições, o vetor y_k é obtido da seguinte maneira:

$$y_k = \nabla_x \iota(x_{k+1},\, \lambda_{k+1},,\, \mu_{k+1}) - \nabla_x \iota(x_k,\, \lambda_k,\, \mu_k) \qquad (10)$$

onde ι é o lagrangeano da função-objetivo.

No entanto, em problemas com restrições, a Hessiana exata do problema não é necessariamente definida positiva na solução. Portanto, nesses casos, nem sempre é possível garantir que a matriz B obtida através da atualização BFGS seja definida positiva. Para superar essa dificuldade, Powell propôs uma modificação da atualização BFGS.

Ele sugeriu que:

$$s_k^T y_k < 0.2 s_k^T B_K s_k \qquad (11)$$

então calcula-se ϕ para obter um novo y_k , mantendo-se o mesmo s_k.

$$\phi = \frac{0.8 s_k^T B_K s_k}{s_k^T B_k s_k - s_k^T y_k} \qquad (12)$$

O novoy_k é obtido da seguinte maneira:

$$y_k = \phi y_k + (1 - \phi) B_k s_k \qquad (13)$$

Desvantagens do Método Quase-Newton

Esse método tem duas desvantagens em relação ao método de Newton:

1. . Necessidade de definir o passo.
2. Uma função deve ser avaliada a mais a cada iteração

6 CONCLUSÃO

Diante dos resultados obtidos acima, que foram devidamente selecionados e estudados por métodos didáticos, seminários e aulas especiais de esclarecimento com professor orientador, o orientando conseguiu formular uma generalização que os descreva do modo mais simples e claro possível, de modo que suas dificuldades que estavam aparecendo, conforme às apresentações foram se tornando mais clara e sempre ciente de que todas as soluções dependiam de seu próprio desempenho.

O projeto em si está trazendo muitos benefícios diante das deficiências que tinha, fazendo com que enxergasse suas aptidões e seus erros e, que ao longo desta trajetória foi desenvolvido uma linguagem mais sucinta conciliando a escrita, a descrição e o raciocínio lógico dos resultados, tendendo para um encadeamento lógico das proposições e na análise das propriedades mais relevantes dos objetos que foram estudados.

A escolha dos tópicos visou um equilíbrio entre a estrutura lógica do assunto e a utilidade em possíveis aplicações relacionadas com o projeto, visto que sua existência neste, é de fundamental importância dentro do corpo do projeto.

7 CRONOGRAMA

O desenvolvimento do projeto obedece o seguinte cronograma:

Atividades	Ago	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul
Diferenciabilidade em $\mathbb{R}$	X											
Diferenciabilidade em $\mathbb{R}^n$		X										
Funções convexas			X									
Formas quadráticas definidas positivas				X								
Métodos unidimensionais de otimização					X							
Convergência local e convergência global						X						
Métodos de Newton no $\mathbb{R}^n$							X	X				
Método do gradiente								X	X			
Método do quase-Newton									X	X		
Métodos de otimização para funções não-diferenciáveis											X	X
Apresentação de Seminários				X								
Elaboração do Relatório Parcial						X						
Relatório Final												X

8 REFERÊNCIAS

Referências

[2] E. L. Lima, Álgebra Linear(Quarta edição).Coleção Matemática Universitária, IMPA, CNPq, 2000. C

[2] Elon Lages, 1929- Análise real , Vol. 1. 10.ª edição. Rio de Janeiro: IMPA

[3] Elon Lages, 1929- Análise real, Vol. 2. 10.ª edição. Rio de Janeiro: IMPA

[4] Humberto josé Bortolossi, 2012- Cálculo Diferencial a Várias Variáveis, uma introdução à otimização, 4.ª edição. Rio de Janeiro: PUC