Universidade Estadual do Maranhão
Centro de Ciências Tecnológicas
Mestrado em Engenharia de Computação e Sistemas

CARLINDO LISBOA ALVES

CONTRIBUIÇÃO À ESTIMAÇÃO DO VETOR GRADIENTE EM FILTRAGEM ADAPTATIVA

São Luís-MA
2016

CARLINDO LISBOA ALVES

CONTRIBUIÇÃO À ESTIMAÇÃO DO VETOR GRADIENTE EM FILTRAGEM ADAPTATIVA

Dissertação apresentada a direção do curso de Mestrado em Engenharia de Computação e Sistemas da Universidade Estadual do Maranhão, como requisito para obtenção do Título de Mestre.

Orientador: Prof. Dr. João Coelho Silva Filho

São Luís-MA

2016

CARLINDO LISBOA ALVES

CONTRIBUIÇÃO À ESTIMAÇÃO DO VETOR GRADIENTE EM FILTRAGEM ADAPTATIVA

Aprovada em:____/____/_____

BANCA EXAMINADORA

Prof. Dr. João Coelho Silva Filho - Orientador

UEMA

Prof. Dr. José Antônio Pires Ferreira Marão

UEMA

Prof. Dr. Reinaldo de Jesus da Silva

UEMA

São Luís-MA

2016

Dedico este trabalho a Deus, minha família e a todos que contribuíram de forma direta para este trabalho.

Agradecimentos

A Deus, que todos os dias de minha vida me deu forças para nunca desistir.

Ao Professor João Coelho que me mostrou os primeiros passos científicos. Por seu apoio e amizade, além de sua dedicação, competência e especial atenção nas revisões e sugestões, fatores fundamentais para a conclusão deste trabalho.

A minha família pelo Apoio nos momentos difíceis.

A meus Colegas de Mestrado e Trabalho.

A meu avós, pelas aventuras e experiência de vida.

A dona Vanda, Sr. Ribamar, Dona Marly.

Ao Departamento de Matemática e Informática da UEMA pelo apoio à minha participação no mestrado.

Ao Curso de Engenharia de Computação, Civil, Mecânica da UEMA pelo apoio à minha participação no mestrado.

A meus professores do Ensino Fundamental, pelo esforço apresentado por eles para meu aprendizado e disciplina.

A meus professores do Ensino Médio, em Particular o Professor Gean Garcia, Klaus Wilker Ramos, José Orlando, Badica Corrêa,Plínio Muniz, Adriana Sodré, José Carlos Quadros (em memória), Luis Carlos, Dejane, José Maria, Josenias Filho e José Carlos. Pelo apoio durante todos os tempos difíceis que tive na vida.

A meus colegas de Graduação, jordano Bruno, Fabiano, Leobino, Paulo Roberto, Roni, Rômulo, Edenilson, Anderson, Jackeline, Rafaela, Simone, Gleyciane, Suelma, Dyogo, Augusto, Krishina, Michael, Wendel e Thiago.

Aos meus colegas do Ensino Médio, Lydyana Boas, Sonaira Abreu, Flaviana de Kassia, Flaicy, Denilson, Magno (Jambão), Cleodivaldo dos Santos, Edenilton, José de Ribamar, Adriana, Denilson, Aletiane, Ailson, Kelyane de Jesus, Gilsoney e Ilksoney Amaral, Jovenias, Flávio, Wagner e Girlan Sá, Daiane, Maria do Rosário, Maria Antônia,

Vandeilson e Milena Raquel.

Aos colegas José do Nascimento Linhares, Luiz Leão Junior, Fernando Glauco, Marcio Fialho, Elisangela Nunes Neves, Jardiel Nunes Almeida. Por essa caminhada juntos.

A todos os professores do mestrado que de alguma forma contribuíram para minha formação.

Lista de Figuras

Lista de Tabelas

Sumário

Resumo

Neste trabalho são mostrados os gradientes em filtragem adaptativa utilizados para otimização no processo de adaptação. São executados alguns métodos numéricos para mostrar o desenvolvimento no processo de filtragem. Além disso, é apresentado a superfície de desempenho que interpreta graficamente o modelo, mostrando o resultado desejado. A pesquisa utilizada apresenta uma adaptação que melhora o desempenho da filtragem, o que é um estudo matemático raro, mas com bons resultados no processo de transmissão.
Palavras-chave. Gradiente. Filtragem Adaptativa. Otimização.

Abstract

This work shows the gradients in adaptive filtering used to optimize the adaptation process. They run some numerical methods to show the development in the filtering process. Moreover, there is error or performance surface, which graphically portrays the model showing the desired result. The research has used an adaptation that improves the performance of the filter. It is a rare mathematical study in the transmission process.
Keywords: Gradient. Adaptive Filtering. optimization.

Capítulo 1

Introdução

A aplicação Matemática surgiu com a humanidade na necessidade do homem organizar-se, controlar e desenvolver a sociedade. As aplicações matemáticas foram desenvolvidas para resolver problemas físicos e químicos que precisavam de soluções imediatas da época e da engenharia para o desenvolvimento de cidades e guerras. Em meados do século XV surgem aplicações com o objetivo de calcular distância planetárias, e com o avanço da Matemática e Física surge, uma técnica interessante utilizada para explicar fenômenos físicos: Denominada de Gradientes.

Os Gradientes apresentaram sua importância pela primeira vez no século XV, quando Galileu calculou distâncias entre planetas e luas. Na Engenharia da Computação teve sua importância em 1941, quando Wierne e Kolmogorov desenvolveram, um filtro adaptativo através de um problema conhecido como teoria da cibernética. A partir que é, o estudo de filtragem adaptativa e Matemática se desenvolveram no ramo da Ciência da Computação. Com o tempo, outros cientistas estudaram o caso e utilizaram métodos antigos, como o de Newton, de Bernoulli, de Bairtow, de Lehmer-Schur e muitos outros dando inicio a teoria da filtragem adaptativa. A definição de filtro adaptativo é um dispositivo que tenta modelar a relação entre o sinal em tempo real de forma iterativa.

Os Filtros adaptativos difere-se do FIR (Resposta Finina ao Impulso) comum ou tipo IIR (Resposta Infinita ao Impulso) filtro digital são coeficientes invariantes no tempo. Os filtros adaptativos podem mudar seu comportamento, ou seja, mudam seus coeficientes de acordo com um algoritmo adaptativo. Os coeficientes do filtro não são conhecidos ao projetar, são calculados quando o filtro é implantado e é automaticamente reajustado a cada iteração, [7].

Estes filtros não são temporários, nem são lineares, seu estudo é mais complexo do que um filtro digital, uma vez que não pode ser aplicado, exceto em algumas exceções, às alterações no domínio da frequência, [8].

Filtros adaptativos são geralmente implementados na forma de algoritmos no microprocessador, DPS (Protetor contra Surtos Elétricos) e FPGA, que é um circuito integrado projetado para ser configurado por um projetista.

A estrutura de um filtro adaptativo é um sistema que vêm dois sinais: $x(n)$ e $e(n)$, a última é chamada de sinal de erro que vem da subtração de um sinal de chamada para o sinal desejado, $d(n)$, e uma outra chamada a saída do filtro, $y(n)$. Os coeficientes do filtro são chamados de $w(n)$, são aqueles que multiplicam a entrada $x(n)$ pela saída, [3] e [8].

No entanto, determinar valores ótimos, é de fundamental importância para adaptação de sinais, gerando um grande desafio para o processo de filtragem, pois exige uma grande velocidade de convergência e determinar valores ótimos requer métodos numéricos, custo computacional baixo e uma resposta rápida de acordo com o sinal desejado, [11].

Devido a impossibilidade da determinação do verdadeiro gradiente, métodos numéricos são tomados, ocasionando erros no processo de filtragem. A convergência é estudada com uma interpretação gráfica, através da superfície de desempenho.

O Trabalho está composto de cinco capítulos, onde o Capítulo 1 apresenta a Introdução da Dissertação e o Capítulo 2 descreve o contexto histórico da aplicação de gradientes em filtragem adaptativa, apresentando os principais pesquisadores da técnica, aplicando os gradientes em diversos ramos da ciência e Engenharia, incluindo a Engenharia da Computação, onde armas adaptadas para guerras, tiveram um grande desfecho para o início da idéia de filtragem adaptativa. São mostrados alguns métodos de otimização como, o método de Newton, de Bernoulli e muitos outros, que foram aplicados na época para mostrar a eficiência de sistemas e desenvolvimento de armas.

No Capitulo 3 são apresentados os métodos utilizados em processos de adaptação de sinais. Esses métodos foram utilizados em filtragem adaptativa, devido a complexidade no desenvolvimento do algorítimo, tais métodos foram criados nas décadas de 50, 60 e 70 para fins de comunicação de rádio. Alguns já tinham sido aperfeiçoados para solucionarem problemas de equações diferenciais parciais, entre estes métodos, tem-se o método de Laplace e o método do Gradiente Conjugado.

No Capitulo 4 é apresentado o método de Widrow, um método proposto em 1986, onde determina-se valores ótimos através da estimativa dos erros, o método foi analisado a partir de algumas técnicas de otimização, para verificar qual seria o mais apropriado para trabalhar a técnica de Widrow. Onde ele apresenta uma forma de determina valores ótimos através das médias dos erros que são utilizados, um estudo na transmissão de sinais móveis.

Capítulo 2

Contexto Histórico

O primeiro estudo sobre os gradientes foram nos trabalhos de Astronomia de Galileu, em seguida, Gabriel Stokes em seus trabalhos de Física e Astronomia utilizando Matemática Pura para o estudo de electromagnetismo, além de Maxwell que utilizou a idéia nos estudos de Eletromagnetismo e campos vetoriais.

A aplicação dos gradientes em outras áreas, inclusive na Ciências da Computação surgiu em meados da Segunda Guerra Mundial, quando o matemático Nobert Wierne serviu as forças armadas americanas e iniciou o desenvolvimento de uma arma para ataque de alvos em movimento, a partir dai, iniciou um estudo de processamento de sinais móveis para determinar a busca de valores ótimos de funções, que até nos dias de hoje é muito aplicado. A aplicação desta área da Matemática teve continuidade com Andrei Kolmogorov, que utilizou a integral para o cálculo de valores ótimos, além de suas contribuições para a integral de Lesbegue e Riemann, [10].

2.1 Evolução dos Gradientes e Filtros adaptativos

O matemático escocês Maxwell, conhecido devido à teoria moderna do eletromagnetismo que une a eletricidade, o magnetismo e a óptica. Maxwell demonstrou que os campos elétricos e magnéticos se propagam com a velocidade da luz. Apresentou uma teoria detalhada da luz como um efeito eletromagnético, isto é, que a luz corresponde à propagação de ondas elétricas e magnéticas, hipótese que tinha sido proposta por Faraday. Ele também desenvolveu um trabalho importante em Mecânica Estatística, estudou a teoria cinética dos gases e descobriu a distribuição de Maxwell-Boltzmann. O trabalho

em eletromagnetismo foi a base da relatividade restrita de Einstein e o trabalho em teoria cinética de gases foi fundamental para o desenvolvimento posterior da Mecânica Quântica. Os gradientes utilizados em seu trabalho foram de fundamental importância para a aplicação dos estudos de cibernética que mais tarde foram utilizados em processo de adaptação de sinais e sistema de controle, [14].

Outro importante Pesquisador foi George Gabriel Stokes, nasceu em 1819 na cidade de Skreeen, na Irlanda, frequentou a escola deixando-a em 1835, após o falecimento de seu pai. Mudou-se para a Inglaterra, onde trabalhou em Bristol College durate dois anos. Esse tempo foi muito importante para prepará-lo em Cambridge, pois em Bristol ele estudou Matemática Pura, Hidrostática, Óptica e Astronomia, todos os tópicos que Stokes dedicou a sua carreira. Em 1837 ele se matriculou no Pembroke College e estudou para o exame de Cambridge, onde em 1841 formou-se como Sênior Wrangler e em 1849 tornou-se professor Lucasiano de Matemática em Cambridge. Foi eleito para a Royal Society em 1851, ganhou a medalha de Rumford em 1852 e foi nomeado secretário da sociedade em 1854. Ajudou a criar o Laboratório Cavendish em meados da década de 1880. Os seus trabalhos foram de fundamentação importância para o desenvolvimento do sistema de controle. Em seus trabalhos apresenta um estudo de Gradientes no campo da Física sobre o eletromagnétismo e óptica.

O matemático Nobert Wierner contribui na construção de uma máquina utilizada na transmissão de sinais, porém com as pressões e tensões da época, Nobert precisou utilizar seus conhecimentos matemáticos para o desenvolvimento de uma técnica de filtragem, que hoje leva o seu nome, [11].

Wierne trabalhou com Hopf durante a guerra, e juntos criaram uma técnica de adaptação de sinais que consistia em um processo de filtragem através do Método de Newton e de Bernoulli, este último foi esquecido devido a baixa convergência, o primeiro foi utilizado até 1941, quando Wierne decidiu, através de seus conhecimentos matemáticos desenvolver uma nova técnica de filtragem. Enquanto Hopf deduzia um processador para a aplicação da técnica. Estudou movimento de partículas a partir da física probabilística, ja que com os métodos usuais, era impossível descrevê-los.

O algoritmo do gradiente é definido em cinco etapas:

Etapa 1 - Uma função de custo que deseja-se minimizar;

Etapa 2 - A média do quadrado da diferença entre uma resposta desejada e a saída

do filtro adaptativo;

Etapa 3 - A representação gráfica em função dos coeficientes do filtro denomina-se superfície de performance;

Etapa 4 - O passo de adaptação;

Etapa 5 - O vetor gradiente do erro quadrático médio para atualizar os coeficientes do filtro adaptativo.

2.1.1 O Filtro de Wiener

A superfície de desempenho é um hiperparabolóide de $N+1$ dimensões em $\mathbb{R}^{\times}$ com um mínimo global e sem mínimos locais. No caso particular de um filtro com apenas dois coeficientes a superfície é um parabolóide que se assemelha a uma "taça". Se a entrada for estacionária, a superfície de erro estrá fixa, mas se a entrada não for estacionária, a superfície é móvel e neste caso a busca do mínimo torna-se difícil, pois o "alvo"está movimentando-se, [2].

Para estudar este tipo de situação procede-se da seguinte maneira: Ao atingir o fundo da superfície, toma-se o sentido oposto ao vetor gradiente da superfície, pois este aponta sempre "para cima", no sentido de maior variação e inverte-se o sentido do gradiente e os coeficientes de um filtro adaptativo varia. O método é dado pela expressão:

$$\nabla grad\left[\varepsilon\right] = gradE\left[e^2\left(n\right)\right] = \frac{\partial E\left[e^2\left(n\right)\right]}{\partial c}.$$

Realizando-se a derivada, obtêm-se o gradiente em cada ponto da superfície de erro:

$$\nabla = -2p + 2Rc.$$

O valor ótimo, C_{opt}, dos coeficientes do filtro é o valor para o qual o gradiente é nulo (no fundo da superfície).

$$-2p + 2Rc = 0 \Longrightarrow C_{opt} = R^{-1}p.$$

Esta última conhecida como equação de Wiener-Hopf.

No sentido do erro quadrático médio o filtro é ótimo, onde o erro quadrático médio é dado por:

$$\varepsilon_{min} = E\left[d^2\left(n\right)\right] - p^T C_{opt}.$$

Para determina-se o ótimo sem calcular o vetor p e as matrizes R e R^{-1}, ajusta-se os coeficientes do filtro iterativamente de forma a descer pela superfície de erro, utilizando um termo proporcional ao gradiente. A equação 2.1 resume a técnica:

$$t\left(n+1\right)=t\left(n\right)-\gamma\nabla\left(n\right). \tag{2.1}$$

A constante γ é uma constante de proporcionalidade positiva, chamada passo de adaptação. O gradiente, que agora se passa a designar por $\nabla\left(n\right)$, pode ser expresso de outra forma:

$$e\left(n\right)=d\left(n\right)-t^{T}\left(n\right)a\left(n\right).$$

Assim,

$$\nabla\left(n\right)=\frac{\partial\varepsilon\left(n\right)}{\partial t\left(n\right)}=\frac{\partial E\left[e^{2}\left(t\right)\right]}{\partial t\left(n\right)}=E\left[\frac{\partial e^{2}\left(n\right)}{\partial t\left(n\right)}\right]=E\left[2e\left(n\right)\frac{\partial e\left(n\right)}{\partial t\left(n\right)}\right]=-2E\left[e\left(n\right)a\left(n\right)\right].$$

Logo,

$$t(n+1)=t\left(n\right)+2\gamma Ee\left(n\right)a\left(n\right). \tag{2.2}$$

A equação 2.2 é chamada de atualização de coeficientes no algoritmo do gradiente.

Subtraindo t_{opt} a ambos os membros da equação

$$t\left(n+1\right)=t\left(n\right)-\gamma\nabla\left(n\right),$$

com $\nabla=-2p+2Rc$, obtem-se:

$$t\left(n+1\right)-t_{opt}=\left(I-2\gamma R\right)t\left(n\right)+2\gamma p-t_{opt}=$$

$$\left(I-2\gamma R\right)t\left(n\right)+2\gamma Rt_{opt}-t_{opt}=$$

$$\left(I-2\gamma R\right)t\left(n\right)-t_{opt}.$$

A solução é:

$$t\left(n\right)-t_{opt}=\left(I-2\gamma R\right)^{n}\left[t\left(0\right)-t_{opt}\right].$$

Para matriz H (matriz de filtragem) e Q (Matriz de Auto correlação de entrada) tem-se:

$$\left(QHQ^{-1}\right)^{k}=HA^{k}H^{-1}.$$

Como $QQ^T = I$ e $R = Q \wedge Q^{-1} = Q \wedge Q^T$, tem-se

$$(I - 2\gamma R)n = (QQ^T - 2\gamma Q \wedge Q^T)n = [Q(I - 2\mu\wedge)Q^T]n = Q(I - 2\gamma\wedge)nQ^T.$$

Concluindo que,

$$c\,(n) - t_{opt} = (I - 2\gamma R)^n\,[t\,(0) - t_{opt}]\,.$$

O coeficiente é uma matriz quadrada $N \times N$ cujos elementos são a soma de N parcelas, cada uma proporcional $a\ (1 - 2\gamma\lambda)$, para $i = 0, 1, ..., N - 1$.

À diferença entre o vetor de coeficientes t e o valor ótimo t_{opt}, chama-se vetor de erro dos coeficientes, t_e:

$$t_e(n+1) = (I - 2\gamma R)t_e(n),$$

$$t_e(n) = (I - 2\gamma R)^n t_e(0),$$

$$t_e(n) = [Q(I - 2\gamma\Lambda)^n Q^T]t_e(0).$$

Utilizando-se o vetor auxiliar v (vetor de transmissão do sinal) dado tem-se

$$v = Q^T t_e = Q^T(t - t_{opt}),$$

multiplicando-se ambos os membros por $Q^{-1} = Q^T$ obtêm-se

$$v(n) = (I - 2\gamma\Lambda)^n v(0).$$

Assim,

$$\begin{bmatrix} v_0(n) \\ v_1(n) \\ \cdots \\ v_{N-1}(n) \end{bmatrix} = \begin{bmatrix} (1-2\gamma\lambda_0)^n & 0 & \cdots & 0 \\ 0 & (1-2\gamma\lambda_1)^n & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & (1-2\gamma\lambda_{N-1})^n \end{bmatrix} \begin{bmatrix} v_0(n) \\ v_1(n) \\ \cdots \\ v_{N-1}(n) \end{bmatrix}.$$

Logo,

$$v_j(n) = (1 - 2\gamma_j R)^n v_j(0),$$

para $j = 0, 1, ..., N-1$, j-ésimo modo natural do algoritmo do gradiente. Para que o algorítimo tenha convergência precisa-se que os modos naturais diminuam a medida que n cresce e

$$\lim_{n\to\infty} v_j(n) = 0 \Longrightarrow \mid 1 - 2\gamma\lambda_j \mid < 1,$$

para $j = 0, 1, ..., N-1$, [4]. Os valores par $v_j(n)$ representam uma série geométrica cuja razão é $r_j = 1-2\gamma\lambda_j$, de valores sucessivamente decrescentes em módulo se $\mid 1-2\gamma\lambda_j \mid < 1$, [5].

O vetor dos coeficientes estará tão próximo do t_{opt} quanto menor for $(1 - 2\gamma\Lambda)^n$,onde Λ são os auto valores da matriz de autocorrelação de entrada e a velocidade de convergência vai depender dos N modos exponenciais, cada um proporcional $(1 - 2\gamma\lambda_j)^n$. Mas para que isso ocorra rapidamente (convergência) o $\mid 1 - 2\gamma\lambda_j \mid$ tem que ser o mais baixo possível, [2] e [4].

O caso desfavorável consiste se para o valor próprio menor, λ_{min}, e o caso favorável para o valor próprio maior, λ_{max}. É conveniente que λ_{min} e λ_{max} tenham diferença mínima, para que os modos de convergência sejam idênticos, já que o modo mais lento impõe velocidade. Os valores próprios de R devem apresentar uma disparidade reduzida. Um caso em que todos os valores próprios são iguais ocorre quando a sequência do sinal de entrada $a(n)$ é estatisticamente independente (sinal branco). A matriz de autocorrelação de entrada R é diagonal, [2] e [6],

$$R = \sigma_a^2 I,$$

$\sigma_a^2 = \lambda_j$, para $j = 0, 1, ..., N-1$. É condição necessária e suficiente de convergência que:

$$0 < \gamma < \frac{1}{\lambda_{max}}.$$

Os valores próprios são não-negativos, qualquer deles é menor que a soma de todos.

$$\lambda_{max} \leq \sum_{j=1}^{N} \lambda_j = tr[R] \Longrightarrow 0 < \gamma < \frac{1}{tr[R]},$$

quando os filtros são transversais com N coeficientes, o traço de R é igual a N vezes a potência do sinal de entrada. Logo,

$$tr[R] = NE[a^2(n)] \Longrightarrow 0 < \gamma < \frac{1}{NE[a^2(n)]}.$$

O passo de adaptação γ regula a velocidade e a estabilidade de convergência do vetor dos coeficientes $c(n)$; na verdade, a escolha de γ é uma situação de compromisso: Nem convém que γ seja baixo;

(i) A aproximação do fundo da superfície será demasiadamente devagar;

Nem muito elevado;

(ii) A aproximação do fundo é rápida, mas, o erro residual final é grande, fazendo com que t flutue em volta do valor ótimo e não o atingirá.

No caso de γ ser tão elevado que se entra na zona de instabilidade não haverá sequer convergência, [4].

Os modos naturais da velocidade de convergência diferentes dos valores próprios λ_j.

Kolmogorov deduziu através do Método de Laplace, o método da descida mais íngreme, utilizando a aplicação em um sistema de transmissão de rádio, também durante a Segunda guerra mundial, a técnica desenvolvida por ele, foi fundamental na comunicação do exercito Russo, durante a guerra contra a Alemanha, na batalha de Stanligrado e Lenigrado. Recebeu muitas honrarias pelos seus feitos Em 1939 foi eleito para a Academia de Ciências da URSS. Kolmogorov recebeu um dos primeiros prêmios científicos dados pelo estado soviético em 1941, o prêmio Lenin de 1965, a Ordem de Lenin, [9].

2.2 A Superfície de Desempenho

Andrei Kolmogorov participou das principais descobertas científicas do século XX nas áreas de probabilidade e estatística e em teoria da informação. Autor da principal teoria científica no campo das probabilidades: a teoria da medida, que revolucionou o cálculo de integrais, permitindo que as integrais fossem generalizadas para domínios exóticos, para além da integral de Riemann, a integral de Lebesgue. Kolmogorov teve muitos interesses fora da matemática, em particular na forma e estrutura da poesia russa do autor Pushkin, assim como Robert Wierne e outros Matemáticos, definiram a superfície de desempenho para uma classe de sistemas adaptativos da seguinte maneira; procede-se uma discussão de algoritmos para ajustar os pesos e descendo sobre a superfície de desempenho utilizando erro médio quadrático. Primeiro precisa-se discutir algumas propriedades importantes da superfície de desempenho quadrática, [2].

As propriedades da superfície de desempenho introduzida são, devidas principalmente às propriedades da matriz de correlação de entrada, R. Usando-se um combinador linear adaptativo com entradas estacionárias, o erro quadrático médio pode ser expresso em termos da matriz de correlação do sinal de entrada como series de R, [2].

A superfície de desempenho de erro médio quadrático é dado pela fórmula.

$$\xi = \xi_{min} + (W + W^*)^T R (W - W^\star),$$

$$\xi = \xi_{min} + V^T R V.$$

Uma vez que existem $L + 1$ (componentes de W), a matriz R tem $N + 1$ colunas e $L + 1$ linhas.

Torna-se evidente que a partir da equação $\xi = \xi_{min} + V^T R V$ que a orientação e forma da superfície quadrática de desempenho médio do erro quadrático é função de R. Muito pode ser aprendido sobre a superfície de desempenho, expressando R em forma normal, em termos de seus valores e vetores próprios, [2].

2.3 A Interpretação Geométrica dos Autovalores e Autovetores

Os autovetores e autovalores estão relacionados diretamente a certas propriedades da superfície de erro ξ, que descreve uma superfície de erro hiperparabólica em um espaço de $L + 2$ dimensões com eixos de coordenadas correspondentes a ξ, $w_0, w_1, ..., w_L$. Será discutido sistema com apenas dois pesos, um espaço de erro tridimensional, podendo generalizar para espaços de dimensões superiores, ou para o espaço bidimensional de um sistema com um único peso, [2] e [3].

No caso de dois peso, ξ é um paraboloide. Cortando o paraboloide com planos paralelos ao plano $w_0 w_1$ obtém-se elipses concêntricas de erro médio quadrático, a forma geral de qualquer destas curvas projetada sobre o plano é $w_0 w_1$. Dada da forma, [3]:

$$W^T R W - 2 P^T W = A,$$

onde A é uma constante.

Pode-se obter de W outros sistemas de coordenadas V, com origem no centro das elipse. A origem é, naturalmente, na coordenada do mínimo erro médio quadrático,

$$V = W - R^{-1}P = W - W^*.$$

Em seguida, tem-se:

$$V^T RV = B.$$

Em que B é outra constante, o que representa uma elipse (ou, em geral, uma elipse hiperbólica) com centro na origem do plano $w_0 w_1$. Neste novo sistema de coordenadas há duas (ou em geral L + 1) linhas normais às elipses. Onde estes são conhecidos como os eixos principais de todas as elipses e da superfície de desempenho, e as linhas são marcadas $v_0^{'}$ e $v_1^{'}$, [4].

Assim para qualquer vetor normal da elipse poderá ser expressa como contornos de $F(V) = V^T RV$ e tendo o gradiente de F, o qual também é o gradiente de ξ, uma vez que ξ e F diferem apenas pela constante. O gradiente é:

$$\nabla = \left[\frac{\partial F}{\partial v_0}, \frac{\partial F}{\partial v_1}, ...\frac{\partial F}{\partial v_L}\right]^T.$$

O resultado foi obtido, escrevendo $V^T RV$ na forma de uma soma de derivadas.

Por outro lado, qualquer vetor que passa pela origem $V = 0$ deve estar na forma de γV. Mas, um eixo principal passa pela origem e é normal para $F(V)$.

Assim,

$$2RV' = \gamma V',$$

ou ainda,

$$[R - \frac{\gamma}{2} I]V\prime = 0,$$

onde V' representa agora um eixo principal. Este resultado de modo que V' deve ser um vetor próprio da matriz R. Assim, *Os autovetores da matriz de correlação de entrada* definem os eixos principais da superfície de erro.

Como resumo destas transformações geométricas, considere as seguintes expressões de superfície de erro nos três sistemas de coordenadas. Onde tem-se:

$$\xi = \xi_{\min} + (W - W^*)^T R(W - W^*),$$

$$\xi = \xi_{\min} + V^T R V,$$

$$\xi = \xi_{\min} + V^T (Q \wedge Q^T) V,$$

$$\xi = \xi_{\min} + (Q^T V)^T \wedge (Q^T V),$$

$$\xi = \xi_{\min} + V'^T \wedge V'.$$

A primeira, segunda e última equação, dão T no natural, traduzindo, o sistema de coordenadas principais, respectivamente. Mostrando que

$$\nabla = 2 \wedge V' = 2 \left[\lambda_0 v_0\prime, \lambda_1 v_1\prime, ..., \lambda_L v_L\prime \right].$$

Se apenas o componente, v'_n, é diferente de zero, o vetor gradiente fica ao longo desse eixo e, V' representa o sistema de coordenadas principais. As transformações correspondentes são, [5]:

$$\text{Translação: } v = w - w^\star,$$
$$\text{Rotação: } v\prime = q^T V = q^{-1} V.$$

Os valores próprios de R também têm importante significado geométrico. O gradiente de e ao longo de qualquer eixo principal, v'_n, poderá ser escrito por

$$\frac{\partial e}{\partial v'_n} = 2\lambda_n v'_n,$$

ou também

$$\frac{\partial^2 e}{\partial v'^2_n} = 2\lambda_n,$$

para $n = 0, 1, 2, 3, 4, ..., L$.

Assim, a segunda derivada de T é duas vezes o valor próprio ao longo de qualquer eixo principal, assim os valores próprios da matriz de correlação de entrada, R, dão as derivadas a segunda da superfície de erro, ξ, em relação aos eixos principais de e, [8].

Um exemplo simples deste resultado, é o caso de um peso em que e se torna uma parábola e r_{mn} representam a (m, n) número de elementos de R. Em seguida, a partir de $e = e_{min} + v^T r v$, tem-se

$$e = e_{min} + r_{00}\left(w - w^{\star}\right)^2.$$

Há apenas uma dimensão de modo que o eixo w é também o eixo principal, e o valor próprio é $\lambda = r_{00}$ e diferenciando duas vezes em relação ao w_1, obtém-se:

$$\frac{\partial^2 e}{\partial w^2} = 2r_{00} = 2\lambda.$$

Assim, com um peso, a segunda derivada da parábola em qualquer ponto corresponde a $2r_{00}$.

A superfície de desempenho interpreta graficamente o resultado obtido mostrando a velocidade de convergência do algoritmo. A superfície é móvel, estando em movimento, calcula em cada ponto da superfície um gradiente que é ortogonal a curva de nível e apontando sempre em sentido oposto ao gradiente da função, assim repete o processo até determina o valor ótimo para a função. A superfície de desempenho pode ter várias interpretações gráficas, dependendo do método de otimização utilizado no processo de filtragem.

Capítulo 3

Métodos de Otimização Por Gradientes

Os métodos de adaptação de sinais, utilizados com otimização de funções, tem evoluído com o tempo, devido a necessidade de adaptação de sinais, os modelos é claro, desenvolveram-se melhor em tempos de guerra, como no caso do modelo de Wierne durante a Segunda Guerra Mundial em 1941, na época, apesar da tecnologia ser algo muito limitado, em sua primeira máquina, junto com Julian Bigelow e Gregory Bateson utilizaram o método de Bernoulli e Newton, para adaptação de sinais. Os gradientes são utilizados em Matemática para a determinação da taxa de maior variação de uma função são determinados por derivadas parciais, [5], [7] e [10].

3.1 Matrizes e Derivadas Parciais

As derivadas parciais apresentam sua importância para estimação de gradientes em filtragem adaptativa, que são analisadas a seguir nesta seção.

Tomando duas variáveis, a função quadrática de desempenho da derivada é dada por:

$$e = e_{\min} + B^T R B,$$

$$e = e_{min} + \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Uma matriz simétrica, utilizada para rotação e translação de eixos. A adaptação

em cada ponto poderá ser calculada por uma rotação de eixos. Podendo utilizar a técnica para determinação de valores ótimos móveis, [7].

$$e = e_{min} + ax^2 + by^2 + 2cxy.$$

Para o caso de estudar gradientes para adaptação no espaço, procedendo da seguinte maneira. Seja a Matriz simétrica:

$$M = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix},$$

o erro pode ser estimado de maneira análoga as funções de custo no plano.

$$e = e_{min} + \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

que define uma equação do segundo grau em x, y e z, forma quadrática tridimensional, dada da por:

$$e = e_{min} + ax^2 + by^2 + z^2 + 2dxy + 2exy + 2fyz,$$

onde vale a expansão para o $\mathbb{R}^n$.

O desempenho inicial pode ser calculada utilizando derivadas parciais, onde

$$P_0 = \frac{x\delta^2}{e_{min}}.$$

Do mesmo modo

$$P_1 = \frac{x_1\delta^2}{e_{min}}.$$

Partindo do princípio de que é necessário o mesmo tempo para a média de cada um dos componentes do gradiente a média da pertubação é dada por

$$P = \frac{\delta^2}{e_{min}} \frac{x + x_1 + x_2 + \cdots + x_{nn}}{2}.$$

E a pertubação geral definida como

$$P = \frac{\delta^2}{e_{min}} \frac{Traço\,[B]}{L+1},$$

ou ainda

$$P = \frac{\delta^2}{e_{\min}} \frac{\sum_{n=0}^{L} \lambda_n}{L+1}.$$

A média dos valores próprios pode ser expressa como

$$P = \frac{\delta^2 \lambda_{av}}{e_{\min}}.$$

Serão usadas derivadas parciais para cada variável envolvida na realização do gradiente de adaptação, cada ponto em que é realizado a derivada parcial, tem-se uma pertubação, graficamente, pode ser interpretado como um plano que intercepta a superfície nesses pontos.

3.2 As Condições de Cauchy-Riemann

O corpo do números complexos é definido como:

$$\mathbb{C} = \{(x, y) : x, y \in \mathbb{R}\},$$

com as seguintes operações de adição e multiplicação com: se $z = (x, y)$ e $w = (a, b) \in$ a $\mathbb{C}$, então $z + w = (x + a, y + b)$ e $zw = (xa - yb, xb + ya)$.

Os elementos de $\mathbb{C}$ são chamados de números complexos. Denotamos o número complexo $(0, 0)$ é denotado por 0 e o número complexo $(1, 0)$ é o elemento identidade, denotado por 1, [17].

Suponha que a função f seja derivável em z_0, em que $z_0 = x_0 + iy_0$, onde

$$f(z) = u(x, y) + iv(x, y).$$

Assim,

$$f'(z) = a + ib$$

$$\Delta f = f(z_0 + \Delta z) - f(z_0)$$

$$\Delta u = u(x_0 + \Delta x, y_o + \Delta y) - u(x_0, y_0),$$

e Δv, para a mudança correspondente em $v(x, y)$. Então

$$\lim_{\Delta z \to 0} \frac{\Delta f}{\Delta z} = \lim_{\Delta z \to 0} \frac{\Delta u + i\Delta v}{\Delta x + i\Delta y} = a + ib$$

e

$$\lim_{\Delta x\to 0,\Delta y\to 0}\left(\frac{\Delta u+i\Delta v}{\Delta x+i\Delta y}\right)=a,$$

$$\lim_{\Delta x\to 0,\Delta y\to 0}\left(\frac{\Delta u+i\Delta v}{\Delta x+i\Delta y}\right)=b.$$

Em particular, quando $\Delta y = 0$, em que $\Delta z = \Delta x$, esses limites se tornam limites de funções de uma variável (Δx) de forma que

$$\lim_{\Delta x\to 0}\frac{u(x_0+\Delta x,y_0)-u(x_0,y_0)}{\Delta x}=a,$$

$$\lim_{\Delta x\to 0}\frac{v(x_0+\Delta x,y_0)-v(x_0,y_0)}{\Delta x}=b,$$

ou seja, as derivadas parciais $\frac{\partial u}{\partial x}$ e $\frac{\partial v}{\partial x}$ com relação a x existem no ponto (x_0, y_0) e $\frac{\partial u}{\partial x} = a$, $\frac{\partial v}{\partial x} = b$.

O procedimento análogo pode ser feito observando quando $\Delta x = 0$ $(\Delta z = i\Delta y)$ de forma que existem as derivadas parciais com relação a y e são elas:

$$\frac{\partial u}{\partial y}=-b,$$
$$\frac{\partial v}{\partial y}=a,$$

no ponto (x_0, y_0). Dos dois procedimentos, obtêm-se às equações:

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}$$

e

$$\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x},$$

que são as Condições de Cauchy-Riemann.

Como $f'(z) = a + ib$, tem-se à expressão,

$$f'(z)=\frac{\partial u}{\partial x}+i\frac{\partial v}{\partial x}=\frac{\partial v}{\partial y}-i\frac{\partial u}{\partial y},$$

no ponto (x_0, y_0). Este desenvolvimento das condições de Cauchy-Riemann:

Teorema 1. *Se a derivada $f'(z)$ de uma função $f = u + iv$ existe num ponto z, então as derivadas parciais de primeira ordem, com relação a x e y, de cada componente u e v devem existir naquele ponto e satisfazer as condições de Cauchy-Riemann.*

Além disso, $f'(z)$ é dada em termos de suas derivadas parciais pela equação, [17].

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}.$$

Seja a equação

$$f(z) = 0,$$

sendo

$$z = a + bi, i = \sqrt{-1},$$

a equação acima pode ser escrita da forma:

$$f(z) = A(x, y) + iB(x, y).$$

O método de Newton fica do seguinte modo,

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}, n = 0, 1, 2, ...,$$

com $f'(z)$ calculado a partir de

$$f'(z_n) = A_x(x, y) + iB_x(x, y) = B_y(x, y) + iA_y(x, y),$$

onde se usam as equações de Cauchy-Riemann para derivação de funções de variáveis complexas. Portanto,

$$z_{n+1} = z_n - \frac{A_n(x, y) + iB_n(x, y)}{A_x(x_n, y_n) + iB_x(x_n, y_n)},$$

onde $z_n = a_n + iy_n$, ou seja,

$$x_{n+1} = x_n - \frac{A(x_n, y_n)A_x(x_n, y_n) + B(x_n, y_n)B_x(x_n, y_n)}{A_x^2(x_n, y_n) + B_x^2(x_n, y_n)},$$

$$y_{n+1} = y_n - \frac{A(x_n, y_n)A_x(x_n, y_n) + B(x_n, y_n)B_x(x_n, y_n)}{A_x^2(x_n, y_n) + B_x^2(x_n, y_n)},$$

para $n = 0, 1, 2, 3, 4, ...$

O uso do método é mais conveniente no caso da equação não ser polinomial, pois se for polinomial $f(z)$ e $f'(z_n)$ podem ser obtidos com método de menor esforço computacional.

3.3 Método da Descida Mais Ingreme

O método da descida mais ígreme, foi utilizado através do tempo, adaptado por outros cientistas na área como, Bairtow e Schur, mas foi Kolmogorov que atingiu a precisão que hoje é conhecida. O método do grandiente conjugado, que também um método eficiente, mas é pouco utilizado em processos de adaptação (Em métodos mais complexos).

Foi estudado pela primeira vez por Debye (1909), que o usou para estimar funções de Bessel argumentando que ocorreu na nota inédita, Riemann (1863) sobre as funções hipergeométrica. O contorno da descida mais íngreme tem uma propriedade minimax, Fedoryuk (2001). Siegel (1932) descreveu algumas outras anotações não publicadas de Riemann, onde ele usou esse método para obter a fórmula de Riemann-Siegel, [12], [13] e [15].

Conhecido como o método de fase estacionária é uma extensão do método de Laplace para aproximação de integrais, onde uma forma de contorno da integral no plano complexo passar perto de um ponto fixo (ponto de sela), e aproximadamente na direção da fase estacionária. A aproximação no ponto de sela é usado com integrais no plano complexo, ao passo que o método de Laplace é usado com integrais reais, [14].

O método pode ser estimado muitas vezes da forma:

$$\int_C f(z)e^{\lambda g(z)}dz,$$

onde C é o contorno e λ é grande.

Entre as formas computacionais, um deles é o método de Laplace.

O método a seguir é capaz de determinar o valor ótimo e os erros das funções de custos implementadas.

```
1.function[funções de interesse]= grad_igreme(varargin)
2.if nargin ==0
```

```
3.x_{0}=[  ]';(Tamanho da matriz)
4.elseif nargin==1
5.x_{0}=varargin{1};
6.end
7.tol = ; (tolerância)
8.maxiter=; (número de interações)
9.dxmin=; (mínimo)
10.alpha=; (Tamanho do passo)
11.gnorm=inf;x=x_{0};niter=0;dx=inf; (parâmetros)
12.f=;(função a ser otimizada)
13.figure(1);clf;ezcontour(f,[]); axis equal; hold on (Contornos)
14.f2=@(x)f(); (variáveis envolvidas)
15.while and(gnorm >=tol,and(niter<=maxiter,dx>=dxmin))
16.g=grad(x); (gradiente)
17.gnorm=norm(g);(norma)
18.xnew = x - alpha*g; (próximo valor)
19.if  isfinite (xnew)
20.niter
21.error('x is inf or NaN')
22.end
23.plot([x(1) xnew(1)],[x(2) xnew(2)],'ko-') (gráfico)
24.refresh
25.niter=niter+1;
26.dx=norm(xnew-x);
27.x=xnew;
28.end
29.xopt = x;
30.fopt = f2(xopt);
31.niter = niter-1;
32.function g=grad(x)
33.g=[]; (gradiente da função)
```

O programa é gerado para estudar a convergência de funções, independente da

função ter raízes reais ou complexas, com a capacidade de determinar valores ótimos e erros. Observe a forma de convergência de alguns gráficos feitos de acordo com o programa.

Como exemplo observe a convergência, o valor ótimo, a interpretação gráfica, as curvas de níveis e o número de erros da função $f(x_1, x_2) = x_1^2 + x_1x_2 + 3x_2^2$.

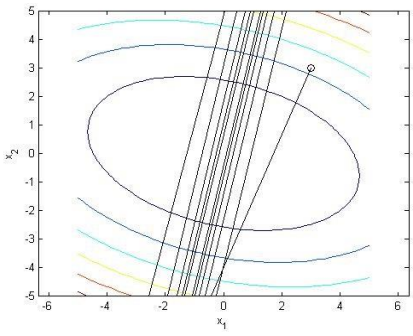

Figura 3.1: Convergência da função para um passo grande.

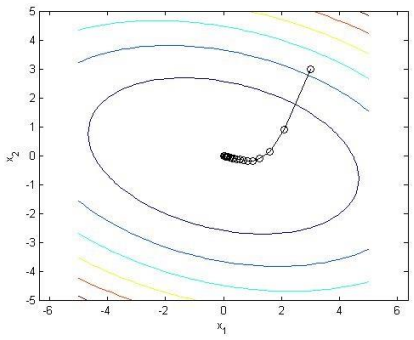

Figura 3.2: Convergência da função para um passo pequeno.

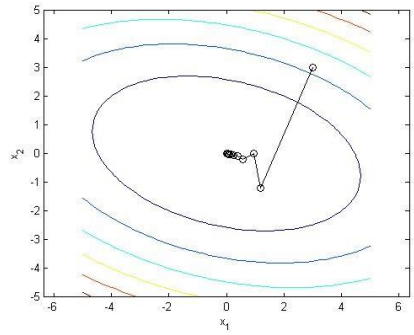

Figura 3.3: Convergência da função para um passo pequeno.

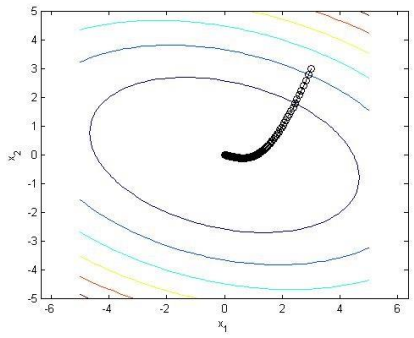

Figura 3.4: Convergência da função para um passo muito pequeno.

Todos os gráficos pertencem a mesma função, mas na Figura 3.2, tem-se uma interpretação para o caso de 10 iterações, para um passo $\alpha = 1$ e os valores ótimos $x_1 = -67252224$ e $x_2 = -284884992$. Na Figura 3.3, tem-se uma interpretação para o caso de 100 iterações, para um passo $\alpha = 0.1$ e os valores ótimos $x_1 = 0.4033$ e $x_2 = -0.0952$ e o erro $e = 1.0e^{-05}$. Na Figura 3.4, tem-se uma interpretação para o caso de 100 iterações, para um passo $\alpha = 0.2$ e os valores ótimos $x_1 = 0.1262$ e $x_2 = -0.0298$ e o erro $e = 1.0e^{-05}$.

Na Figura 3.5, tem-se uma interpretação para o caso de 10000 iterações, para um passo $\alpha = 0.1$ e os valores ótimos $x_1 = 0.0062$ e $x_2 = -0.00238$ e o erro $e = 1.0e^{-03}$.

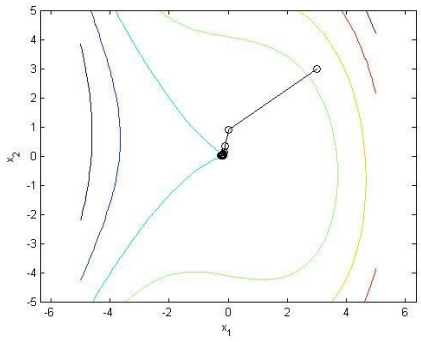

Figura 3.5: Convergência da função para um passo grande.

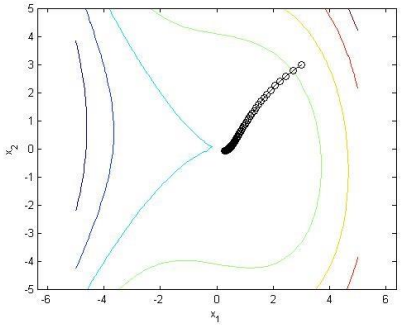

Figura 3.6: Convergência da função para um passo muito pequeno.

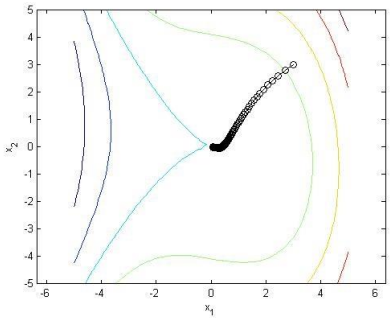

Figura 3.7: Convergência da função para um passo grande.

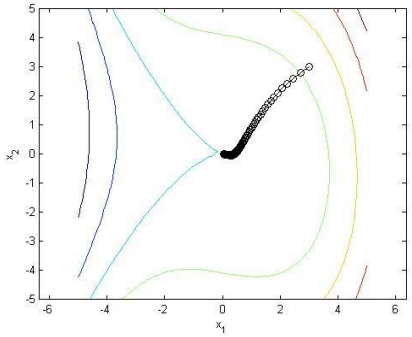

Figura 3.8: Convergência da função para um passo muito pequeno.

Os gráficos da Figura 3.6, 3.7, 3.8 e 3.9 são da função $f(x_1, x_2) = x_1^3 + x_1x_2 + 3x_2^2$, na figura 3.6, tem-se uma interpretação para o caso de 10 iterações, para um passo $\alpha = 1$ e os valores ótimos $x_1 = -67252224$ e $x_2 = -284884992$. Na Figura 3.7, tem-se uma interpretação para o caso de 100 iterações, para um passo $\alpha = 0.01$ e os valores ótimos $x_1 = 0.4033$ e $x_2 = -0.0952$ e o erro $e = 1.0e^{-05}$. Na Figura 3.8, tem-se uma interpretação para o caso de 1000 iterações, para um passo $\alpha = 0.02$ e os valores ótimos $x_1 = 0.1262$ e $x_2 = -0.0298$ e o erro $e = 1.0e^{-05}$.

Na Figura 3.9, tem-se uma interpretação para o caso de 10000 iterações, para um passo $\alpha = 0.01$ e os valores ótimos $x_1 = 0.0682$ e $x_2 = -0.0114$ e o erro $e = 1.0e^{-04}$.

Os gráficos da Figura 3.10 e 3.11 são da função $f(x_1, x_2) = x_1^4 + x_1x_2 + 3x_2^2$,na Figura 3.10, tem-se uma interpretação para o caso de 10000 iterações, para um passo $\alpha = 0.01$ e os valores ótimos $x_1 = 0,2073$ e $x_2 = -0,0346$. Na Figura 3.11, tem-se uma interpretação para o caso de 100000 iterações, para um passo $\alpha = 0.02$ e os valores ótimos

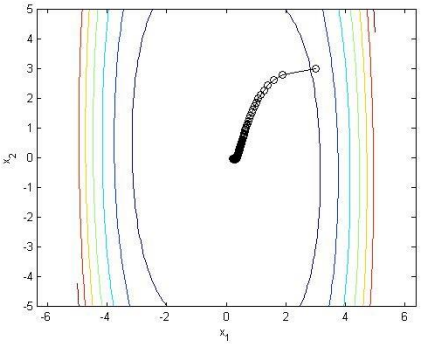

Figura 3.9: Convergência da função para um passo pequeno.

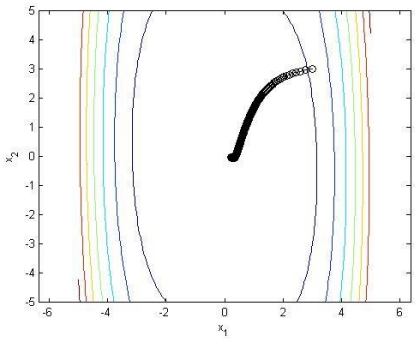

Figura 3.10: Convergência da função para um passo muito pequeno.

$x_1 = 0.2056$ e $x_2 = -0.0343$.

3.4 Método de Gradientes Puros

Considerando a equação polinomial com coeficientes complexos:

$$f(z) = \sum_{k=0}^{n} (a_k + ib_k) z^k = 0, z = x + iy, \tag{3.1}$$

onde

$$f(z) = A(x, y) + iB(x, y) = 0.$$

Encontrar Z tal que $f(z) = 0$ implica resolver o sistema de equações

$$\begin{cases} A(x, y) = 0, \\ \text{B(x,y)= 0.} \end{cases} \tag{3.2}$$

Para isso têm-se a seguinte função

$$f(z) = A^2(x, y) + B^2(x, y). \tag{3.3}$$

Observe que a função F é tal que:

(i) F é não negativa;

(ii) As derivadas $\frac{\partial F}{\partial x}$ e $\frac{\partial F}{\partial y}$ existem;

(iii) Os zeros de F são as raízes de $f(z) = 0$;

(iv) Esses zeros são os pontos de mínimo para a função F.

Para se resolver a equação (3.3), minimiza-se a equação (3.1), e isso ocorre quando $\frac{\partial F}{\partial x} = 0$ e $\frac{\partial F}{\partial y} = 0$, isto é, quando as coordenadas do vetor gradiente de F são nulas. A idéia do método para determina esse mínimo é:

Dado (x_0, y_0) uma tentativa inicial qualquer para (x^*, y^*), calcula-se o gradiente de F nesse ponto. Em seguida é, atualizado a aproximação inicial por

$$\begin{cases} x_1 = x_0 + \Delta x_0 \\ y_1 = y_0 + \Delta y_0, \end{cases} \tag{3.4}$$

onde

$$\Delta x_0 = -h_0 \frac{\partial F(x_0, y_0)}{\partial x}$$

e

$$\Delta y_0 = -h_0 \frac{\partial F(x_0, y_0)}{\partial y},$$

com $h_0 > 0$, isto é, calculando o gradiente no ponto (x_0, y_0) e percorrendo uma quantidade h_0 em sentido contrário ao do gradiente, e h_0 deve ser escolhido para minimizar F. Repetindo-se o processo, tem-se o sistema

$$\begin{cases} x_2 = x_1 + \Delta x_1 \\ y_2 = y_1 + \Delta y_1, \end{cases} \tag{3.5}$$

onde,

$$\Delta x_1 = -h_1 \frac{\partial F(x_1, y_1)}{\partial x}$$

e

$$\Delta y_0 = -h_1 \frac{\partial F(x_2, y_2)}{\partial y},$$

$h_1 > 0$ e h_1 desempenhando o mesmo papel que h_0 quando o calculo de δx_0 e δy_0. De maneira geral, tem-se

$$\begin{cases} x_{x+1} = x_k + \Delta x_k \\ y_{k+1} = y_k + \Delta y_k, \end{cases} \tag{3.6}$$

onde

$$\Delta x_k = -h_k \frac{\partial F(x_k, y_k)}{\partial x}$$

e

$$\Delta y_k = -h_k \frac{\partial F(x_k, y_k)}{\partial y},$$

com h_k desempenhando o mesmo papel que h_0 e h_1.

Uma maneira de se determinar h_k é a seguinte: desenvolvendo U em série de Taylor em torno do ponto (x_k, y_k) e calculando $U(x^*, y^*)$ por meio do desenvolvimento, tem-se

$$-A(x_k, y_k) \approx \frac{\partial A(x_k, y_k)}{\partial x}\Delta x_k + \frac{\partial A(x_k, y_k)}{\partial y}\Delta y_k. \tag{3.7}$$

Do mesmo modo para V, tem-se

$$-B(x_k, y_k) \approx \frac{\partial B(x_k, y_k)}{\partial x}\Delta x_k + \frac{\partial B(x_k, y_k)}{\partial y}\Delta y_k. \tag{3.8}$$

Usando a equação de Cauchy-Riemann e resolvendo o sistema das equações (3.7) e (3.8) para Δ_{x_k} e Δ_{y_k}, tem-se

$$\Delta x_k \approx \frac{-A(x_k, y_k)\frac{\partial A}{\partial x}(x_k, y_k) - B(x_k, y_k)\frac{\partial B}{\partial x}(x_k, y_k)}{\left[\frac{\partial A(x_k, y_k)}{\partial x}\right]^2 + \left[\frac{\partial B(x_k, y_k)}{\partial x}\right]^2} \tag{3.9}$$

e

$$\Delta x_k \approx \frac{-A(x_k, y_k)\frac{\partial U}{\partial x}(x_k, y_k) - V(x_k, y_k)\frac{\partial B}{\partial x}(x_k, y_k)}{\left[\frac{\partial A(x_k, y_k)}{\partial x}\right]^2 + \left[\frac{\partial B(x_k, y_k)}{\partial x}\right]^2}. \tag{3.10}$$

Comparando a equação (3.1) com a equação (3.7) e (3.8), tendo em vista a equação (3.2) e as equações de Cauchy-Riemann, têm-se

$$h_0 \approx \frac{\frac{1}{2}}{\left[\frac{\partial A(x_k, y_k)}{\partial x}\right]^2 + \left[\frac{\partial B(x_k, y_k)}{\partial x}\right]^2} > 0. \tag{3.11}$$

3.5 Método do Gradiente Conjugado

O método do Gradiente conjugado iniciou na década de 50, quando o estudo de Equações Diferenciais Parciais começou a ser utilizadas em resolução de problemas através de computador na pesquisa científica. Trabalhos como de Hestenes e Stiefel, tem grande referência como base do que foi desenvolvido posteriormente, [8] e [12].

Seja a matriz do sistema linear simétrica e positiva definida, isto é, $A^T = A$ e $x^T Ax > 0$, para qualquer x não-nulo. Assim a ideia resolver o sistema $Ax = b$ e minimizar a função de $x = (x_1, x_2, x_3, x_4, ..., x_n)$. Para simplificar o entendimento é utilizado para $n = 2$, o desenvolvimento é o mesmo para um número arbitrário de componentes. [12] e [13].

As curvas de nível da função $F(x)$ são definidas por $F(x_1, x_2) = F(k)$ e, são elipses. Assim, o gráfico de $F(x)$ é um paraboloide com um único mínimo. O mínimo da função é atingido no vetor x que anula o gradiente da função, [12].

$$gradF(x_1, x_2) = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 - b_1 \\ a_{22}x_2 + a_{12}x_2 - b_2 \end{bmatrix} = Ax - b.$$

O $GradF(x_1, x_2) = 0$ significa $Ax = b$, isto é, é solução do sistema de equações lineares que minimiza a função quadrática e vice-versa. Como é conhecido do Cálculo Diferencial, o vetor gradiente aponta a direção do máximo da função. Portanto é natural que, nos passos de busca do mínimo, na direção contrária ao gradiente, isto é

$$x^{(k+1)} = x^k - s_k GradF(x^{(k)}).$$

A aproximação num passo $k+1$, é calculada a partir da aproximação no passo anterior, na direção contrária ao gradiente, o parâmetro real s_k regula o tamanho do passo na $k-$ésima iteração.

O parâmetro s_k, tamanho do passo, será usado na minimização do resíduo associado à aproximação que está sendo calculada. Desta maneira sera calculado o valor de s que minimiza,

$$F(x + sr) = \frac{1}{2}(x + sr)^T A(x + sr) - b^T(x + sr).$$

Fixado x, é derivado a função com relação á s, através da regra da cadeia e derivada do produto:

$$\frac{dF}{ds}(x + rs) = \frac{1}{2}r^T A(x + sr) + \frac{1}{2}(x + sr)^T Ar - b^T r.$$

Utilizando a propriedade distributiva com as de transposição de vetores e matrizes para mostrar que $x^T Ar = (Ar)^T x = r^T Ax$, uma vez que A é simétrica tem-se

$$\frac{dF}{ds}(x + sr) = sr^T Ar + r^T(Ax - b = sr^T Ar - r^T r.$$

Igualando a derivada a zero, obtêm-se o valor de s que minimiza a função:

$$s = \frac{r^T r}{r^T Ar}.$$

O programa seguinte é uma aplicação computacional do método do Gradiente Conjugado.

```
function [x, k] = cgp(x0, A, C, b, mit, stol, bbA, bbC)
% Synopsis:
% x0: initial point
% A: Matrix A of the system Ax=b
% C: Preconditioning Matrix can be left or right
% mit: Maximum number of iterations
% stol: residue norm tolerance
% bbA: Black Box that computes the matrix-vector product for A * u
% bbC: Black Box that computes:
%      for left-side preconditioner : ha = C \ ra
%      for right-side preconditioner: ha = C * ra
% x: Estimated solution point
% k: Number of iterations done
%
% Example:
% tic;[x, t] = cgp(x0, S, speye(1), b, 3000, 10^-8, @(Z, o) Z*o, @(Z, o) o);toc
% Elapsed time is 0.550190 seconds.
% Reference:
%  Métodos iterativos tipo Krylov para sistema lineales
%  B. Molina y M. Raydan - ISBN 908-261-078-X
if ( nargin < 8 ), error('Not enough input arguments. Try help.'); end;
if ( isempty(A) ), error('Input matrix A must not be empty.'); end;
if ( isempty(C) ), error('Input preconditioner matrix C must not be empty.'); end;
x = x0;
```

```
ha = 0;
hp = 0;
hpp = 0;
ra = 0;
rp = 0;
rpp = 0;
u = 0;
k = 0;
ra = b - bbA(A, x0); % <--- ra = b - A * x0;
while ( norm(ra, inf) > stol ),
ha = bbC(C, ra); % <--- ha = C \ ra;
k = k + 1;
if ( k == mit ), warning('GCP:MAXIT', 'mit reached, no conversion.'); return; end;
hpp = hp;
rpp = rp;
hp = ha;
rp = ra;
t = rp'*hp;
if ( k == 1 ),
u = hp;
else
u = hp + ( t / (rpp'*hpp) ) * u;
end;
Au = bbA(A, u); % <--- Au = A * u;
a = t / (u'*Au);
x = x + a * u;
ra = rp - a * Au;
end;
```

3.6 Outros Métodos

De 1948 a 1949, a fim de estudar a arte em computação, Rutishauser foi para os EUA nas universidades de Harvard e Princeton. De 1949 a 1955, ele era um associado de pesquisa no Instituto de Matemática Aplicada da ETH Zürich recentemente fundada por Eduard Stiefel, onde trabalhou em conjunto com Ambros Speiser sobre o desenvolvimento da primeira ERMETH, e desenvolveu a linguagem de programação Superplan (1949-1951), o nome sendo uma referência para "Rechenplan"(plano de computação), na terminologia de Konrad Zuse designando um único Plankalkül-programa. Ele contribuiu, em particular no domínio do compilador, trabalho pioneiro, e acabou por ser envolvido na definição das linguagens de programação ALGOL 58 e ALGOL 60. Com o nascimento da comunicação sem fio, utilizou um algoritmo com capacidade de reduzir erros na transmissão de sinais com climas desfavoráveis contribuindo para a criação das transmissões de TV via satélite. O método de Rutishauser foi utilizado no trabalho de James Freeman para estudar uma técnica de programação na transmissão de sinais móveis.

Entre outras contribuições, ele introduziu uma série de características sintáticas básicas a programação, nomeadamente a palavra-chave para um loop for, primeiro como o für alemão em Superplan, ao lado através da sua tradução em Inglês "para"em ALGOL 58. O método de Rutishauser, é um método não iterativo que, utiliza apenas os coeficientes da equação, fornecendo equações simultâneas para todas as raízes da equação polinomial, porém tem lenta convergência, não sendo recomendável para a determinação dos resultados finais das raízes, [10].

O método de Bairstow é um método interativo de convergência quadrática. O método determina simultaneamente duas raízes, ja que trabalha com um fator quadrático do tipo $x^2 + px + q$ de P_n. As duas raízes que o método aborda são as raízes do fator quadrático, o método trabalha somente com valores reais, mesmo com raizes complexas porém, por ser um método iterativo, ele não converge com qualquer tentativa inicial. O método foi aplicado pela primeira vez em 1918 na Europa quando, o sistema ferroviário estava projetando uma máquina para comunicação sem fio e em 1924 criou um aparelho capaz de realizar a comunicação, em 1925 a empresa Zugtelephonie AG foi criada para desenvolver a técnica de comunicação a grandes distâncias. O serviço telefônico nos trens da Deutsche Reichsbahn (serviço alemão de correios) na linha Hamburgo-Berlim foi eficiente, passando a ser oferecido aos viajantes. O método foi utilizado nos trabalhos de Collin

Hebbian no ano de 2010 para mapeamento de sinais embutidos em sistemas e impressões digitais, [7].

Outro método de primeira ordem para raízes reais e simples que sempre determina a maior raiz em valor absoluto. O método de Bernoulli Pode ser utilizado para determina raízes múltiplas e complexas, desde que seja modificado convenientemente. No caso das raízes múltiplas a ordem de convergência é afetada. O método foi pouco aplicado devido a baixa convergência mas, nos trabalhos do laboratório Bell nos Estados Unidos o método foi aplicado no desenvolvimento de um sistema telefônico, para reduzir erros na transmissão dos sinais mas, infelizmente eram necessários técnicas mais avançadas poi, um número muito grande de antenas era interligado, [7] e [9].

Capítulo 4

Estimativa de Gradientes Por Médias

Gradientes podem ser determinados de forma simples,

$$\xi = E_{min} + \lambda t^2.$$

As respectivas derivadas primeira e segunda são:

$$\frac{dE}{dv} = 2\lambda t,$$

e

$$\frac{d^2E}{dt^2} = 2\lambda.$$

A primeira derivada é importante para o método da descida mais íngreme, mas o método de Newton requer derivadas a primeira e segunda ordem.

As sugestões das derivadas que são estimadas numericamente para determinação de valores ótimos nos dá a seguinte definição.

Definição 1. *Se ξ for o número de erros tomados durante uma adaptação, $f(x)$ uma função de grau n e δ um raio de aproximação do valor ótimo então as equações,*

$$\frac{dE}{dv} \approx \frac{E\left(t+\delta\right)^n - E\left(t-\delta\right)^n}{2\delta},$$

e

$$\frac{d^2E}{dv^2} \approx \frac{E\left(t+\delta\right)^n - 2E\left(t\right) + E\left(t-\delta\right)^n}{\delta^2}.$$

nos dão em média a aproximação do valor ótimo.

As aproximações são exatas quando delta aproxima-se de zero e para valores finitos de delta quando a função de desempenho é quadrática em t. Assim,

$$\frac{e\left(t+\delta\right)-e\left(t-\delta\right)}{2\delta}=\frac{\lambda\left(t+\delta\right)^{2}-\lambda\left(t-\delta\right)^{2}}{2\delta}$$

$$\frac{\xi\left(t+\delta\right)-e\left(t-\delta\right)}{2\delta}=\frac{\lambda}{2\delta}\left(t^{2}+\delta^{2}+2t\delta-t^{2}-\delta^{2}+2t\delta\right)$$

$$\frac{e\left(t+\delta\right)-e\left(t-\delta\right)}{2\delta}=2\lambda t$$

$$\frac{e\left(t+\delta\right)-e\left(t-\delta\right)}{2\delta}=\frac{d\xi}{dt}$$

e

$$\frac{e\left(t+\delta\right)-2e\left(t\right)+e\left(t-\delta\right)}{\delta^{2}}=\frac{\lambda\left(t+\delta\right)^{2}-2\lambda\left(t\right)^{2}+\lambda\left(t-\delta\right)^{2}}{\delta^{2}}$$

$$\frac{e\left(t+\delta\right)-2e\left(t\right)+e\left(t-\delta\right)}{\delta^{2}}=\frac{2\lambda\delta^{2}}{\delta^{2}}$$

$$\frac{e\left(t+\delta\right)-2e\left(t\right)+e\left(t-\delta\right)}{\delta^{2}}=2\lambda$$

$$\frac{e\left(t+\delta\right)-2e\left(t\right)+e\left(t-\delta\right)}{\delta^{2}}=\frac{d^{2}e}{dt^{2}}.$$

O processo mostra o ajustamento das variáveis enquanto o gradiente será realizado. Para estimar a primeira derivada, suponha que o sistema adaptativo passa o tempo com as variáveis ajustadas para $t-\delta$ e $t+\delta$, a fim de recolher amostras do erro e, e estima-se $E\left(t-\delta\right)$ e $E\left(t+\delta\right)$ é estimado. E a mesma quantidade de tempo são gastos em $(t-\delta)$ e $(t+\delta)$, sem tempo a ser gasto em t, [4].

A penalidade de desempenho é definido como o aumento Γ provocada por ajustamento das variáveis

$$\Gamma=\frac{1}{2}[e\left(t-\delta\right)+e\left(t-\delta\right)]-e\left(v\right). \tag{4.1}$$

Para o desempenho da função o aumento pode ser encontrado da seguinte forma:

$$\Gamma = \frac{1}{2}[2e_{\min} + \lambda\left(t-\delta\right)^2 + \lambda\left(t+\delta\right)^2] - \left(e_{\min} + \lambda t^2\right), \tag{4.2}$$

$$\Gamma = \lambda\delta^2, \tag{4.3}$$

Neste caso, Γ é constante.

A perturbação P, é dada da forma

$$P = \frac{\Gamma}{\xi_{min}} = \frac{\lambda\delta^2}{e_{min}}. \tag{4.4}$$

A expressão dá equação (4-4) dá o aumento médio do erro quadrático médio normalizado, no que diz respeito ao erro mínimo médio-quadrático alcançável.

No caso do método de Newton a segunda derivada é constante, no entanto, medido com pouca frequência, e o seu efeito sobre a perturbação para fins práticos pode ser ignorado, [15].

Observação 1. *Quando é utilizado uma função de custo para o plano pode ser estendido essa ideia para função de mais de uma variável, e o método é análogo ao anterior, ou seja, se existe uma função de custo de mais de uma variável o e pode ser determinado através do conjunto D que representa uma circunferência e o erro mínimo seria calculado através de* $0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta$*, mas neste caso, e analisar uma região D do domínio onde, é encontrado um número infinito de aproximações para o valor ótimo, ou seja, para qualquer direção que tomar para delta, tem um valor aproximado para a estimação do erro.*

Utilizando o caso em que t era uma função quadrática, e verificar o quanto em média estar próximo do valor ótimo se t for uma função de terceiro grau ou de quarto grau.

Para t com grau $n = 3$, pela definição 1 tem-se

$$\xi = E_{min} + \lambda t^3$$

$$\frac{e\left(t+\delta\right) - e\left(t-\delta\right)}{2\delta} = \frac{\lambda(t+\delta)^3 - \lambda(t-\delta)^3}{2\delta}$$

$$\frac{\lambda}{2\delta}(t^3 + 3t^2\delta + 3t\delta^2 + \delta^3) - \frac{\lambda}{2\delta}(t^3 - 3t^2\delta + 3t\delta^2 - \delta^3)$$

$$\frac{\lambda}{2\delta}(t^3 + 3t^2\delta + 3t\delta^2 + \delta^3 - t^3 + 3t^2\delta - 3t\delta^2 + \delta^3)$$

$$3t^2\lambda + \lambda\delta^2,$$

ou seja, o valor ótimo em média, através da estimativa do erro estará próximo de

$$\frac{dE}{dt} = \frac{\lambda(v+\delta)^3 - \lambda(v-\delta)^3}{2\delta} = 3v^2\lambda + \lambda\delta^2,$$

para δ próximo de zero, obtêm-se.

$$\frac{dE}{dt} = \frac{\lambda(v+\delta)^3 - \lambda(v-\delta)^3}{2\delta} = 3v^2\lambda.$$

aproximando-se ainda mais do valor ótimo, obtêm-se

$$\frac{d^2E}{dv^2} = 6v\lambda.$$

A penalidade de desempenho para este caso é dado da seguinte forma.

$$\Gamma = \frac{1}{2}[3e + \lambda(t-\delta)^3 + \lambda(t+\delta)^3 - e + \lambda t^3] = 3\delta^2\lambda,$$

$$P = \frac{3\delta^2\lambda}{e_{min}}.$$

Para t com grau $n = 4$, tem-se

$$\xi = E_{min} + \lambda t^4$$

$$\frac{e\,(t+\delta) - e\,(t-\delta)}{2\delta} = \frac{\lambda(t+\delta)^4 - \lambda(t-\delta)^4}{2\delta} =$$
$$\frac{\lambda}{2\delta}(t^4 + 2t^3\delta + t^2\delta^2 + 2t^3\delta + 4t^2\delta^2 + 2t\delta^3 + \delta^2t^2 + 2t\delta^3 + \delta^4) -$$
$$(t^4 + 2t^3\delta - t^2\delta^2 + 2t^3\delta - 4t^2\delta^2 + 2t\delta^3 - \delta^2t^2 + 2t\delta^3 - \delta^4) =$$

$$4t^3\delta + 4t^3\delta + 4t\delta^3 + 4t\delta^3 = \frac{\lambda}{2\delta}\left(8t^3\delta + 8t\delta^3\right),$$

ou seja, o valor ótimo em média, através da estimativa do erro estará próximo de

$$\frac{dE}{dt} = \frac{\lambda(t+\delta)^3 - \lambda(t-\delta)^3}{2\delta} = 4t^3\lambda + 4t\delta^2,$$

aproximando ainda mais do valor ótimo, obtêm-se

$$\frac{d^2E}{dv^2} = 12v^2\lambda + 4\delta^2.$$

Aproximação que pode ser melhorada da forma,

$$\frac{d^3E}{dv^3} = 24v\lambda.$$

A penalidade de desempenho, é dada por

$$P = \frac{4\delta^3\lambda}{e_{min}}.$$

Para todos os métodos apresentados, o processo pode ser realizado com $n > 0$. A Tabela 4.1, mostra para $n = k$ as aproximações para o valor ótimo utilizando a idéia de Widrow e um ajuste que ocorre na superfície quando determina-se o número de erros em cada iteração.

n	Aproximação do valor ótimo	Nova iteração	Nova iteração	Penalidade de desempenho
2	$2t\lambda$	2λ	0	$\Gamma = \lambda\delta^2$
3	$3t^2\lambda + \lambda\delta^2$	$6t\lambda$	6λ	$\Gamma = 3\lambda\delta^2$
4	$4t^3\lambda + 4\delta^2$	$12t^2\lambda$	$24t\lambda$	$\Gamma = 4\lambda\delta^2$
5	$5t^4\lambda + 3\delta^2$	$20t^3\lambda$	$60t^2\lambda$	$\Gamma = 5\lambda\delta^2$
6	$6t^5\lambda + 6\delta^2$	$30t^4\lambda$	$120t^3\lambda$	$\Gamma = 6\lambda\delta^2$
7	$7t^6\lambda + 5\delta^2$	$42t^5\lambda$	$121t^4\lambda$	$\Gamma = \lambda 7\delta^2$
8	$8t^7\lambda + 8\delta^2$	$56t^6\lambda$	$336t^5\lambda$	$\Gamma = 8\lambda\delta^2$
9	$9t^8\lambda + 7\delta^2$	$72t^7\lambda$	$504t^6\lambda$	$\Gamma = 9\lambda\delta^2$
10	$10t^9\lambda + 9\delta^2$	$90t^8\lambda$	$720t^7\lambda$	$\Gamma = 10\lambda\delta^2$

Tabela 4.1: Tabela para Determinação de valores ótimos

4.1 Estimação dos Valores Ótimos em Média pelo Método da Bisseção

Seja a função $f(t)$ contínua no intervalo [a,b] e e supondo que neste intervalo contém uma única raiz tal que $f(t) = 0$. O método consiste em reduzir a amplitude do intervalo

que contém a raiz até se atingir a precisão requerida: $(b-a) < E$, utilizando-se para isto a sucessiva divisão [a,b]. Dividindo o intervalo [a,b] ao meio, obtêm-se x_0, havendo, pois dois subintervalos, $[a, t_0]$ e $[t_0, b]$, a ser considerados.

Se $f(t_0) = 0$, então $\varepsilon = x_0$; caso contrário, a raiz estará no subintervalo onde a função tem sinais opostos nos pontos extremos, ou seja, se $f(a).f(b) < 0$, então $\varepsilon \in (a, x_0)$; se $f(a)f(x_0) > 0$, então $\varepsilon \in (x_0, b)$.

O método da bisseção verificar o quanto em média estaríamos próximo do valor ótimo. Fazendo uma estimativa a partir do erro

$$\frac{e(t+\delta)}{2\delta} + \frac{e(t-\delta)}{2\delta} = \frac{\frac{\lambda(t+\delta)^2}{2\delta} + \frac{\lambda(t-\delta)^2}{2\delta}}{2},$$

$$\frac{e(t+\delta)}{2\delta} + \frac{e(t-\delta)}{2\delta} = \frac{\lambda t^2 + 2t\delta\lambda + \lambda\delta^2 + \lambda t^2 + 2t\delta\lambda - \lambda t^2}{4\delta},$$

$$\frac{e(t+\delta)}{2\delta} + \frac{e(t-\delta)}{2\delta} = \frac{\lambda(2t^2 + 4t\delta)}{4\delta} = \frac{2t^2\lambda}{4\delta} + \frac{4t\delta\lambda}{4\delta} = \frac{t^2\lambda}{2\delta} + \frac{t\lambda}{\delta}.$$

Em média pelo método da bisseção o erro está próximo do valor ótimo através de

$$\frac{t^2\lambda}{2\delta} + \frac{t\lambda}{\delta} = \frac{\lambda}{\delta}\left(\frac{t^2}{2} + 1\right).$$

Considerando que t é uma função de grau 3.

$$\frac{\frac{\lambda(t+\delta)^3}{2\delta} + \frac{\lambda(t-\delta)^3}{2\delta}}{2} = \frac{\lambda t^3 + 3t\lambda\delta^2 + \lambda t^3 + 3t\delta^2\lambda}{4\delta} =$$

$$\frac{\lambda(2t^3 + 6t\delta^2)}{4\delta} = \frac{2\lambda t^3}{4\delta} + \frac{6\lambda t\delta^2}{4\delta} = \frac{\lambda t^3}{2\delta} + \frac{3\lambda t\delta^2}{2\delta}.$$

O valor ótimo estaria próximo de.

$$\frac{\lambda t^3}{2\delta} + \frac{3\lambda t\delta}{2}.$$

O método da bisseção, não pode ser generalizado, pois a interação depende também da imagem dos pontos e da escolha de um intervalo onde possa ter uma base para a escolha do valor ótimo. Neste Caso o método converge devagar.

O método da bisseção é a partir da técnica de Widrow é

```
INPUT: Function f, endpoint values a, b, tolerance TOL, maximum iterations NMAX
CONDITIONS: c < d, either f(c) < 0 and f(d) > 0 or f(c) > 0 and f(d) < 0
OUTPUT: value which differs from a root of f(c)=0 by less than TOL
N seta para a esquerda  1
While N menor igual NMAX # limit iterations to prevent infinite loop
c seta para a esquerda (c + d)/4 # new midpoint
If f(e) = 0 or (d - c)/4  TOL then # solution found
Output(e)
Stop
EndIf
N seta para esquerda N + 1 # increment step counter
If sign(f(e)) = sign(f(c)) then a seta para esquerda c else d seta para esquerda e
EndWhile
Output("Method failed.") # max number of steps exceeded
```

Aplicando-se o método da bisseção a partir da técnica de Widrow em um filtro gaussiano e RIF, obtêm-se o seguinte resultado para 5000 iterações, na implementação da função objetivo $f(t_1, t_2) = t^n + t_1 t_2 + t_2^n$. Nas Tabelas 4.2 e 4.3 tem-se uma estimativa para os erros obtidos utilizando o método da bisseção a partir da técnica de Widrow.

Grau da função	Filtro Gaussiano	RIF
2	3367.10^{-6}	2335.10^{-6}
3	3223.10^{-8}	3436.10^{-7}
4	3466.10^{-8}	1235.10^{-6}
5	3588.10^{-6}	2134.10^{-9}

Tabela 4.2: Os erros para 15000 iterações

Grau da função	Filtro Gaussiano	RIF
2	2341.10^{-6}	3125.10^{-6}
3	2243.10^{-8}	3912.10^{-7}
4	1232.10^{-8}	4123.10^{-6}
5	1336.10^{-6}	4134.10^{-9}

Tabela 4.3: Estimativa dos erros para 22000 iterações

Não há um controle para os erros obtidos de acordo com o grau de filtragem utilizado.

4.2 Estimação de Valores Ótimos em Média Método de Newton

O método de Newton combina duas ideias básicas comuns nas aproximações numéricas: linearização e iteração. Na linearização, é substituído (numa certa vizinhança) um problema complicado por uma aproximação linear que, por via de regra, é facilmente resolvida. Por outro lado, um processo iterativo, ou aproximações sucessivas, consiste na repetição sistemática de um procedimento até que atingido o grau de aproximações desejado. No caso do método de Newton, a linearização consiste na "substituição"da curva $y = f(x)$ por sua reta tangente. Utilizando o método de Newton é mostrado o quanto em média, está próximo do valor ótimo.

Para aplicar o método de Newton,é observado as seguintes condições para convergência:

(i)-As derivadas $f^{'}(t)$ e $f^{''}(t)$ devem ser não nulas;

(ii)-A escolha do ponto t_0 deve ser tal que $f^{'}(t).f^{''}(t) > 0$.

Utilizar utilizando o método de iteração de Newton e verificar o quanto em média, se estará próximo do valor ótimo.

Para $n = 2$, tem-se.

$$\frac{e(t+\delta)}{2\delta} + \frac{e(t-\delta)}{2\delta} = \frac{\lambda(t+\delta)^2}{2\delta} - \frac{\lambda(t-\delta)^2}{2\delta} - \left[\frac{\frac{\lambda(t+\delta)^2}{2\delta} - \frac{\lambda(t-\delta)^2}{2\delta}}{\left(\frac{\lambda(t+\delta)^2}{2\delta} - \frac{\lambda(t-\delta)^2}{2\delta}\right)'} \right],$$

$$2\lambda t - \frac{2\lambda t}{2\lambda\delta} = 2\lambda t - \frac{t}{\delta}.$$

O último termo, a fração $\frac{t}{\delta}$ oferece uma aproximação melhor para o valor ótimo. Agora se δ for muito grande, o valor ótimo estará próximo de $2\lambda t$.

Vale lembrar que, o método de Newton, é um método interativo e depende muito do ponto de partida ser escolhido. O método converge muito rápido mas, o valor da derivada não pode ser muito próximo de zero, senão tem-se uma uma oscilação, e não haverá convergência.

utilizando a técnica de Widrow, obtêm-se o seguinte resultado para 5000 e 100000 iterações respectivamente, na implementação da função objetivo $f(t_1, t_2) = t^n + t_1 t_2 + t_2^n$. As tabelas 4.4 mostram que para um Filtro Gaussiano o número de erros diminuem mas, para um Filtro FIR não há controle na redução dos erros.

Grau da função	5000	100000
2	3421.10^{-6}	3127.10^{-6}
3	3363.10^{-8}	3396.10^{-7}
4	3343.10^{-8}	3345.10^{-6}
5	3327.10^{-6}	3344.10^{-9}

Tabela 4.4: Amostra de erros para 5000 e 100000 iterações

Utilizando o Método da descida mais íngreme, para 5000 e 10000 iterações respectivamente, mostrados nas Tabelas 4.5 nos dão os seguintes resultados.

Grau da função	5000	10000
2	2312.10^{-6}	3521.10^{-6}
3	2201.10^{-8}	3456.10^{-7}
4	2101.10^{-8}	3376.10^{-6}
5	2012.10^{-6}	3287.10^{-9}

Tabela 4.5: Erros para 5000 e 10000 iterações

A medida que aumenta o número de iterações os erros diminuem. Utilizando o método de Widrow baseado na técnica da descida mais íngreme, há uma melhor redução dos erros, mostrando mais eficiência que o método de Newton e da Bisseção.

O Método de Newton-Rapshon a partir da técnica de Widrow é

O método descrito pode determinar os valores ótimos e erros em cada iteração

```
%These choices depend on the problem being solved
x0 = 1                        %The initial value
f = (função objetivo)         %The function whose root we are trying to find
fprime = derivada da função objetivo           %The derivative of f(x)
tolerance =          %7 digit accuracy is desired
epsilon =           %Don't want to divide by a number smaller than this
maxIterations =          %Don't allow the iterations to continue indefinitely
haveWeFoundSolution = false %Have not converged to a solution yet
for i = 1 : maxIterations
y = f(x0)
yprime = fprime(x0)
if(abs(yprime) < epsilon)    %Don't want to divide by too small of a number
% denominator is too small
break;                        %Leave the loop
end
x1 = x0 - y/yprime            %Do Newton's computation
if(abs(x1 - x0) <=  tolerance * abs(x1)) %If the result is within the desired toler
haveWeFoundSolution = true
break;                        %Done, so leave the loop
end
x0 = x1                        %Update x0 to start the process again
end
if (haveWeFoundSolution)
... % x1 is a solution within tolerance and maximum number of iterations
else
... % did not converge
end
```

Capítulo 5

Resultados e Conclusões

O presente trabalho apresentou resultados de pesquisas sobre gradientes e métodos de otimização utilizados para filtragem adaptativa, a evolução do processo do desenvolvimento da filtragem desde da década de quarenta até os dias atuais, mostrando o desenvolvimento da técnica durante esse período. Mostrou-se um método de otimização proposto por Robert Widrow em 1986 para a determinação de valores ótimos através da média da estimativa dos erros, contribuindo com técnica em adaptação de transmissão de sinais.

O trabalho mostrou a complexidade dos processadores atuais, o desenvolvimento de Filtros diminui os riscos de erros grandes e possibilita que o sinal ou imagem processada seja reutilizada. Além disso, os processos de filtragem possibilitam que os aparelhos digitais não se preocupem com detalhes de baixo nível, reduzindo assim a complexidade, o tempo e o custo computacional envolvidos no processo.

O trabalho possibilita o entendimento da aplicação matemática na engenharia e o desenvolvimento de um algoritmo que realizou uma série de testes sobre Filtragem. O algoritmo através do método de Laplace utilizado para calcular valores ótimos de funções e erros, interpreta graficamente a convergência, mostrar as curvas de níveis e determinar o valor ótimo, tudo isso através do controle do passo de adaptação. O método de Laplace e Newton, desenvolvidos pelo método de Widrow indica quando a função converge ou não e mostra um possível número de iterações convenientes para uma rápida convergência, utilizando uma matriz quadrada auxiliar processo de convergência e adaptação, tal matriz, tem que ser quadrada. E neste caso a matriz utilizada é a matriz identidade, para facilitar no processo de adaptação, a função (não obrigatoriamente) precisa ser polinomial para

diminuir o custo computacional observando, o passo de adaptação e o número de iterações, e dependendo da função de custo utilizada, o tempo computacional pode ser menor ou maior.

O método proposto por Widrow, baseado na transmissão de sinais móveis, mostra que a superfície de desempenho interpretado graficamente pelo sinal desejado. O verdadeiro Gradiente determina o movimento da Superfície que sempre está em movimento, tornando necessário uma medida que nos mostre o quanto se aproxima do valor ótimo, já que neste caso a determinação do valor ótimo, é impossível. Mas, através de métodos numéricos, pode-se determinar uma proximidade do valor ótimo.

Foi utilizado o método de Newton-Rapshon, descida mais acentuada e método da bisseção para analisar o método de Widrow. Utilizando o Método da bisseção mostrou-se uma baixa convergência para todas as funções de custos implementadas sem controle no número de erros.

O Método de Widrow mostrou ser mais eficiente através do método de Newton-Rapshon, e o método de Descida mais igreme por apresentarem um melhor controle na determinação dos erros.

Através do Método de widrow pode-se verificar uma redução nos erros, a medida que o número de iterações aumenta e uma proximidade do valor ótimo mais precisa.

Referências Bibliográficas

[1] D. E. Knuth, *The Art of Computer Programming, Seminumerical Algorithms*, Vol.2.Reading, Mass: Addison - Wesley. 1989, Sec.3.2

[2] Freeman. A. James, *Algorithhms, applications, and Programming Techniques*. ACM Trans. on Math Software,pp.132-138, June 1999.

[3] Simon. H, *Neural Networks and Learning Machines*. Pearson Education, pp.81-93, October 1999.

[4] W. Bernad, *Adaptive Signal Processing*. Editora Elsevier, Florida,1986.

[5] S. D. Stearns, *A Portable Rondom Number Generator for use in Signal Processing. Sandia National Laboratories*. SAND 81-1993, October 1981.

[6] Dunne. A. R, *A Statístical Approach to Neural Networks for Pattern Recognition,*,Wiley, 2007.

[7] D, Howard.B. Mark, *Neural Network Toolbox, For Use With Matlab*, Editora Springer, Florida, 2013.

[8] F. Laurene, *Fundamentals of Neural Networks,architectures,algorithms, and aplications*, Edinburg University Press Ltd, 2007.

[9] F. Colin, *Hebbian Learning and Negative Feedback Networks*,Editora Elsevier, Flórida, 2010.

[10] Martins, M. A. *Introdução Ao Uso de Redes Neurais com Matlab*, Pearson Education, São Paulo, 2004.

[11] H, Simon. *Kalman Filtering And Neural Networks*,Volume único, Editora Elsevier, Springer, 2001.

[12] Carter, Matt. *An Introdution to the Philosophy of Artificial Intelligence*,Edinburg University Press Ltd, 2007.

[13] G, Lemonick. Scientific American. *Gênios da Ciência*, Vol 10,11 e 12 pag 12-13, São Paulo, 2008.

[14] Àvila, G. *Funções de Variáveis Complexas*, Editora Ltd, São Paulo, 2001.

[15] E. I. Akhan, M.B., *Pré-Processing of fingerprint imges* European Conference on Security and Detection, pp 28-30 University of Hertfordshire, UK. April 2013.

[16] C. Aura; Viola, Flávio; Gonzaga, S.L.O., *Melhoria de imagens de impressões digitais por filtro de Gabor adaptativo baseado em campos direcionais*, Universidade Federal Fluminense - 2014.

[17] G. Sanderson L. de Oliveira;*Uma metodologia de Identificação de Imagens Pelo filtro de Garbor*, Revista IEEE América Latina, Volume 4, 2016.

[18] G. S.L.O., *Desenvolvimento de um algoritmo baseado no filtro de Gabor para identificação de impressões digitais,* Dissertação (Mestrado em Modelagem Computacional), Instituto Politécnico, Universidade do Estado do Rio de Janeiro: Nova Friburgo, 2013.

[19] R. Rubistein, T. P., ELAD, M. "Analysis K-SVD: *A Dictionary-Learning Algorithm for the Analysis Sparse Model"*, IEEE Transactions on Signal Processing, v. 61, n. 3, pp. 661–677, Feb. 2013.