



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA
MESTRADO ACADÊMICO EM ENGENHARIA DE TELEINFORMÁTICA

RAFAEL DE CARVALHO BLUHM

**SINAIS COMPLEXOS IMPRÓPRIOS: FILTRAGEM ADAPTATIVA E GRADIENTE
NATURAL**

FORTALEZA

2023

RAFAEL DE CARVALHO BLUHM

SINAIS COMPLEXOS IMPRÓPRIOS: FILTRAGEM ADAPTATIVA E GRADIENTE
NATURAL

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia. Área de Concentração: Sinais e Sistemas.

Orientador: Prof. Dr. Charles Casimiro Cavalcante.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

B62s Bluhm, Rafael de Carvalho.

Sinais Complexos Impróprios: Filtragem Adaptativa e Gradiente Natural / Rafael de Carvalho Bluhm. – 2023.

138 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2023.

Orientação: Prof. Dr. Charles Casimiro Cavalcante.

1. Sinais Complexos. 2. Amplamente Linear. 3. Filtragem Adaptativa. 4. Gradiente Natural. I. Título.

CDD 621.38

RAFAEL DE CARVALHO BLUHM

SINAIS COMPLEXOS IMPRÓPRIOS: FILTRAGEM ADAPTATIVA E GRADIENTE
NATURAL

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia. Área de Concentração: Sinais e Sistemas.

Aprovada em: 10/02/2023.

BANCA EXAMINADORA

Prof. Dr. Charles Casimiro Cavalcante (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. João César Moura Mota
Universidade Federal do Ceará (UFC)

Prof. Dr. Marcello Luiz Rodrigues de Campos
Universidade Federal do Rio de Janeiro (UFRJ)

À minha mãe maravilhosa, minha família linda,
minha noiva querida, meus prezados amigos,
professores e todos aqueles que me apoiaram e
acreditaram em mim.

AGRADECIMENTOS

Primeiramente, dedico esta dissertação à Deus. Agradeço à Ele esta enorme oportunidade apesar de todas as dificuldades.

Um agradecimento especial à minha mãe Silésia, extraordinária mulher, pela paciência e suporte. À toda minha família, tios em especial, por todo apoio e puxões de orelhas. À minha amada noiva Anélia pela paciência e apoio, por também ter uma fé inabalável.

Ao professor Charles, pelo apoio e paciência, minha eterna gratidão. Esta tese também é dedicada à todos os professores do GTel e do Deti, sempre solícitos e atenciosos. Um agradecimento especial a todos os colegas do Gpsi e SPIRAL pelo encorajamento e ajuda. Gratidão em especial à UFC, uma instituição exemplar.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001

"A mente que se abre a uma nova ideia jamais
voltará ao seu tamanho original." (Albert
Einstein)

RESUMO

Esta dissertação estuda o comportamento de sinais complexos e sua simetria (ou a falta desta) com a performance em filtros adaptativos. O estudo utiliza as seguintes técnicas de filtragem linear: o algoritmo WLLMS (*Widley Linear Least Mean Square* ou LMS aumentado), o método de Newton e o método do gradiente natural aumentados. O método adaptativo WLLMS é baseado em sua versão complexa usual LMS. Enquanto esta última processa o sinal por meio de uma estrutura linear, a versão aumentada utiliza o sinal e sua cópia conjugada numa estrutura dita *amplamente* linear. O critério de otimização utilizado é o erro quadrático médio, cujo gradiente fornece o ajuste dos coeficientes dos filtros. O método WLLMS provou-se, em muitos casos, adequado em fornecer menores erros com velocidades de convergência similares às aquelas apresentadas por sua versão linear LMS. Outra técnica, baseada no método de Newton usual, é o método de Newton aumentado. Desenvolvido de forma a utilizar a inversa da matriz hessiana aumentada como fator de passo no algoritmo do gradiente. O modelo estimado foi construído utilizando o lema da inversão de matrizes. O critério de otimização é o mesmo utilizado no caso WLLMS, o erro quadrático médio. O método de Newton aumentado se mostrou bastante satisfatório em fornecer menores erros. No pior dos casos este apresentou uma performance equivalente ao do método usual. Por último é apresentado o chamado Gradiente Complexo Aumentado construído como uma notação complexa do caso real. Utilizando métricas reais como plano de fundo é possível utilizar notação complexa com os mesmos resultados do caso real.

Palavras-chave: sinais complexos; amplamente linear; filtragem adaptativa; gradiente natural.

ABSTRACT

This dissertation studies the behavior of complex signals and its symmetry (or the lack of it) with performance in adaptive filters. The study uses the following linear filtering techniques: the WLLMS algorithm (*Widely Linear Least Mean Square* or augmented LMS), the augmented Newton's method and the augmented natural gradient method. WLLMS adaptive method is based on its usual LMS complex version. While the this last processes the signal through a linear structure, the augmented version uses the signal and its conjugated copy in a so-called *widely* linear structure. The optimization criterion is the mean squared error, whose gradient is used to adjust the filters coefficients. The WLLMS method proved, in many cases, adequate in providing smaller errors with convergence speeds similar to its linear version, LMS. Another technique, based on the usual Newton method, is the augmented Newton method. Developed in order to use the inverse of the augmented hessian matrix as a step factor in the gradient algorithm. The estimated model was constructed using the matrix inversion lemma. The optimization criterion is the same used in the WLLMS case, the mean squared error. The augmented Newton method proved to be quite satisfactory in providing smaller errors. In the worst case, this presented a performance equivalent to that of the usual LMS method. Finally, the so-called *Augmented Complex Gradient*, constructed as a complex real-case notation, is presented. Using real metrics at background, it is possible to use complex notation with the same results as in the real case.

Keywords: complex signals; widely linear; adaptive filtering; natural gradient.

LISTA DE FIGURAS

Figura 1 – Representação gráfica do plano complexo para um dado número complexo $z = x + jy$ e seu conjugado	22
Figura 2 – Representação gráfica do polar plano complexo para um dado número complexo $z = \rho(\cos \theta + j \sin \theta)$ e sua forma de Euler	23
Figura 3 – Representação do mapeamento feito pela função complexa $f(z) = e^z$	24
Figura 4 – Solução harmônica sem perdas ($\alpha = 0$) e unidirecional para as equações de Maxwell	36
Figura 5 – Transformada de Fourier, que decompõe a função (dada em vermelho) em suas componentes harmônicas (em azul)	37
Figura 6 – Perfil de escoamento laminar indicando o campo de velocidades ao passar por um perfil circular	38
Figura 7 – Exemplos de um sinal $f(t) = \exp(-\alpha \cdot t)$ em (a) e em (b) como uma versão amostrada do caso contínuo com $T = 1$ unidade de tempo	40
Figura 8 – Representação gráfica de um sinal do tipo $s(t) = e^{j\omega t} = \cos \omega t + j \sin \omega t$	41
Figura 9 – Parte real de um sinal discreto do tipo $s[n] = \cos(\alpha n) + j \sin(\alpha n)$, ou seja $\text{Re}\{s[n]\} = \cos(\alpha n)$, para um período de amostragem de $T = \pi/20$, intervalo de tempo de $0 \leq t \leq 6\pi$ e $\alpha = 1$	42
Figura 10 – Um exemplo de sinal periódico discreto do tipo $s[n] = \cos n + j \sin n$ para 95 amostras	42
Figura 11 – Curvas de nível para distribuição gaussiana bivariada de média zero	62
Figura 12 – Curvas de nível para distribuição gaussiana bivariada de média zero	65
Figura 13 – Perfis de processos GWSS que modelam sinais	66
Figura 14 – Exemplo de sinal u (contínuo ou discreto) filtrado por um filtro indicado por w , resultando em um sinal de saída v	68
Figura 15 – Esquema de um filtro transversal discreto IIR com $M \times N$ coeficientes	69
Figura 16 – Atrasador unitário, que recebe o símbolo “ z^{-1} ” devido à transformada Z associada ao atraso unitário	69
Figura 17 – Esquema de um filtro transversal discreto IIR em termos de sua função de transferência.	70
Figura 18 – Esquema de um filtro transversal discreto FIR de tamanho M	71

Figura 19 – Esquema de um filtro transversal discreto FIR com comparação da saída com o sinal desejado	72
Figura 20 – Exemplo para o método iterativo do gradiente para seis iterações.	75
Figura 21 – Configuração para um discreto amplamente linear. O sinal é dividido de forma que a parte não conjugada passa por um filtro $\mathbf{w}$ e seu conjugado por $\mathbf{h}$. Em seguida os sinais são adicionadas em uma única saída $v[n]$	79
Figura 22 – Perfil e erro quadrático médio para um sinal gaussiano circular	87
Figura 23 – Sinal $\rho_s = 0,3$, $\varphi_s = 0$ e ruído circular para os algoritmos linear e aumentado	88
Figura 24 – Sinal $\rho_s = 0,9$, $\varphi_s = 0$ e ruído circular para os algoritmos linear e aumentado	88
Figura 25 – Sinal circular e ruído não-circular com $\rho_N = 0,2$ para os algoritmos linear e aumentado	89
Figura 26 – Erro médio quadrático para um sinal e ruído não-circulares com $\rho_s = 0.9$, $\varphi_s = -\pi/3$, $\rho_N = 0.2$ e $\varphi_N = \pi/4$	90
Figura 27 – Realizações dos algoritmos para o caso 4QAM	91
Figura 28 – Método de Newton para encontro de raízes para uma iteração em um dado instante i . A linha hachurada representa a derivada da função no ponto x_i indicada por $f'(x_i)$	94
Figura 29 – Método de Newton para o sinal circular e ruído circular com o perfil mostrado em (a). Nessa simulação $\rho_s = \rho_N = 1/\sqrt{2}$ (circular) e $\varphi_s = \varphi_N = 0$. Em (b) $\lambda = 0.9$ (valor usual), (c) $\lambda = 0.6$ e d) $\lambda = 0.99$	110
Figura 30 – Método de Newton para o sinal circular e ruído não circular com o perfil mostrado em (a). Nessa simulação $\rho_n = 0.2$ (não-circular) e $\varphi_N = \pi/4$. . .	111
Figura 31 – Realizações dos algoritmos para o caso 4QAM	112
Figura 32 – Exemplo de variedade com mapeamento em duas regiões que possuem uma interseção	119
Figura 33 – Função definida sobre M restrita à curva mapeada pela carta, para uma parametrização φ	122

LISTA DE TABELAS

Tabela 1 – Derivadas para funções de primeira ordem e quadrática. As linhas indicam a função e as colunas o tipo de derivada calculada	55
--	----

LISTA DE SÍMBOLOS

$j = \sqrt{-1}$	Unidade imaginária
$\mathbb{R}$	Corpo dos números reais
$\mathbb{C}$	Corpo dos números complexos
$T_p M$	Tangente à uma variedade num dado ponto (pode não ser especificado)
z	minúscula, um valor escalar $\in \mathbb{R}$ ou $\mathbb{C}$
z^*	Conjugado de um número, vetor ou matriz complexa qualquer z
$\mathbf{z}$	Em negrito, um vetor $\in \mathbb{R}^n$ ou $\mathbb{C}^n$ de um espaço de dimensão n
$\mathbf{Z}$	Em negrito maiúsculo uma matriz $\in \mathbb{R}^{m \times n}$ ou $\mathbb{C}^{m \times n}$ de um espaço de dimensão $m \times n$
$\underline{\mathbf{Z}}$	Variável complexa aumentada $\in \mathbb{C}_*^n$
z^T	Transposto de um vetor ou matriz qualquer z
z^H	Transposto conjugado de um vetor ou matriz qualquer z
$ z $	Norma euclidiana de um escalar z .
$\ \mathbf{z}\ $	Norma euclidiana de um vetor $\mathbf{z}$.
$E\{\}$	Operador esperança

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Contribuições	17
1.2	Organização	18
2	FUNDAMENTOS DO CÁLCULO $\mathbb{C}\mathbb{R}$	20
2.1	Análise Complexa de Funções Holomorfas	21
2.1.1	<i>Propriedades dos números complexos</i>	<i>21</i>
2.1.2	<i>Derivada Complexa</i>	<i>23</i>
2.1.3	<i>Integração Holomorfa Complexa</i>	<i>27</i>
2.2	Derivadas de Wirtinger para Pares Escalares	27
2.2.1	<i>Operadores descritos por meio de transformações de coordenadas</i>	<i>27</i>
2.2.2	<i>Operadores descritos pela diferencial de uma função real</i>	<i>28</i>
2.2.3	<i>Isomorfismos importantes</i>	<i>29</i>
2.2.4	<i>Aplicações da derivadas de Wirtinger</i>	<i>29</i>
2.3	Cálculos de Integrais	32
2.4	Exemplos da representação complexa	33
2.4.1	<i>Soluções harmônicas para a equação de onda</i>	<i>34</i>
2.4.2	<i>A transformada de Fourier</i>	<i>36</i>
2.4.3	<i>Dinâmica dos Fluidos</i>	<i>38</i>
2.5	Sinais e sua representação complexa	39
2.5.1	<i>Sinais reais</i>	<i>39</i>
2.5.2	<i>Sinais complexos</i>	<i>40</i>
3	VETORES E TRANSFORMAÇÕES AMPLAMENTE LINEARES . . .	43
3.1	Transformações lineares	43
3.2	Relação linear entre variáveis real e complexa	45
3.3	Vetores e transformações amplamente lineares	45
3.4	Produtos internos de vetores aumentados	48
3.5	Formas bilineares	49
3.6	Derivadas de vetores e matrizes complexas	49
3.6.1	<i>Exemplos do uso dos operadores complexos</i>	<i>51</i>
3.6.1.1	<i>Derivadas para funções do tipo $f(\mathbf{z}, \mathbf{z}^*) = \mathbf{a}^H \mathbf{z}$</i>	<i>51</i>

3.6.1.2	<i>Derivadas para funções do tipo $g(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{a}$</i>	52
3.6.1.3	<i>Derivadas para a forma quadrática $h(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{R} \mathbf{z}$</i>	53
3.6.1.4	<i>Resumo geral das regras de derivação vetorial</i>	55
3.7	Nomenclatura de Vetores e Matrizes	56
4	ESTATÍSTICA COMPLEXA	58
4.1	Variáveis Aleatórias Complexas	58
4.1.1	<i>Variáveis aleatórias unidimensionais reais</i>	58
4.1.2	<i>Variáveis bidimensionais e Complexas</i>	59
4.2	Circularidade e Propriedade	59
4.2.1	<i>Circularidade de distribuições gaussianas</i>	60
4.2.1.1	<i>Caso bidimensional</i>	60
4.2.1.2	<i>Caso vetorial</i>	62
4.3	Processos Aleatórios Complexos	63
4.4	Modelagem de sinais complexos	64
5	FILTROS DISCRETOS COMPLEXOS	68
5.1	Filtros discretos IIR	69
5.2	Filtros discretos FIR	70
5.3	Modelagem de filtros transversais discretos.	71
5.3.1	<i>Filtragem ótima (de Wiener)</i>	73
5.3.2	<i>O método do gradiente (Steepest Descent)</i>	74
5.3.3	<i>O método dos Mínimos Quadrados (LMS)</i>	75
5.3.4	<i>Fator de passo</i>	76
5.4	Filtros Amplamente Lineares	78
5.4.1	<i>Modelagem</i>	78
5.4.2	<i>Coefficientes ótimos</i>	80
5.4.3	<i>Ortogonalidade aumentada</i>	81
5.4.4	<i>Solução Iterativa da Descida mais Íngreme</i>	82
5.4.5	<i>Fator de passo em filtros aumentados</i>	83
5.4.6	<i>Aproximação WLLMS</i>	84
5.4.7	<i>Realizações WLLMS e LMS comparadas</i>	85
5.4.7.1	<i>Modelagem, escolha de parâmetros e alguns resultados</i>	86
5.4.7.2	<i>Exemplos com sinais 4QAM</i>	89

6	MÉTODO DE NEWTON AUMENTADO	92
6.1	Séries de Taylor	92
<i>6.1.1</i>	<i>A aproximação de primeira ordem</i>	<i>92</i>
<i>6.1.2</i>	<i>Duas ou mais variáveis</i>	<i>92</i>
6.2	Raízes de uma função pelo método de Newton	93
6.3	Método de Newton para otimização	95
<i>6.3.1</i>	<i>Otimização para funções de uma variável</i>	<i>95</i>
<i>6.3.2</i>	<i>Otimização para funções de múltiplas variáveis</i>	<i>96</i>
<i>6.3.3</i>	<i>Métodos comparados</i>	<i>98</i>
<i>6.3.4</i>	<i>Expressão real composta</i>	<i>98</i>
<i>6.3.5</i>	<i>Expressão complexa aumentada</i>	<i>99</i>
<i>6.3.5.1</i>	<i>Exemplo de aplicação na função custo complexa.</i>	<i>100</i>
<i>6.3.6</i>	<i>Relação entre as expressões real composta e a complexa aumentada</i>	<i>101</i>
6.4	Método para expressões complexas compostas	102
<i>6.4.0.1</i>	<i>Exemplo de aplicação na função custo complexa amplamente linear.</i>	<i>104</i>
6.5	Estimação da matriz inversa	105
<i>6.5.1</i>	<i>O lema de inversão de matrizes para a autocorrelação aumentada</i>	<i>106</i>
<i>6.5.2</i>	<i>Método a partir do ponto ótimo</i>	<i>107</i>
6.6	Realizações com o algoritmo de Newton	108
<i>6.6.1</i>	<i>Sinal aleatório mais ruído</i>	<i>109</i>
<i>6.6.2</i>	<i>Exemplos com sinais 4QAM</i>	<i>110</i>
7	GRADIENTE NATURAL COMPLEXO	113
7.1	Conceito geral	113
<i>7.1.1</i>	<i>Coordenadas generalizadas e métrica</i>	<i>114</i>
<i>7.1.2</i>	<i>O vetor gradiente em coordenadas generalizadas</i>	<i>115</i>
<i>7.1.3</i>	<i>O método do gradiente natural</i>	<i>117</i>
<i>7.1.3.1</i>	<i>Exemplo de aplicação</i>	<i>117</i>
7.2	Conceito do gradiente complexo natural	118
<i>7.2.1</i>	<i>Variedades riemannianas</i>	<i>118</i>
<i>7.2.2</i>	<i>Espaço tangente</i>	<i>119</i>
<i>7.2.2.1</i>	<i>Vetores no $\mathbb{R}^n$</i>	<i>120</i>
<i>7.2.2.2</i>	<i>Vetores sobre a variedade</i>	<i>121</i>

7.2.3	<i>Métricas sobre a variedade</i>	123
7.2.4	<i>Bases complexificadas</i>	124
7.2.5	<i>Representação com coeficientes complexos</i>	124
7.2.6	<i>Representação aumentada</i>	125
7.2.7	<i>A estrutura complexa</i>	126
7.2.8	<i>Representação de formas diferenciais complexas</i>	127
7.2.9	<i>Produto interno e métrica</i>	128
7.2.10	<i>O co-gradiente</i>	129
7.2.11	<i>O gradiente complexo aumentado</i>	130
7.2.12	<i>Aplicações do gradiente natural complexo</i>	131
7.2.12.1	<i>Representação cartesiana</i>	131
7.2.12.2	<i>Representação polar</i>	132
7.2.12.3	<i>Representação oblíqua</i>	134
7.2.13	<i>Considerações sobre o método</i>	135
8	CONCLUSÃO	136
	REFERÊNCIAS	137

1 INTRODUÇÃO

O processamento de sinais cresce a cada dia com o desenvolvimento de equipamentos cada vez mais robustos e modernos e com isso estudo de novas técnicas matemáticas de aprendizado e adaptação se torna uma necessidade na vida de um engenheiro do ramo. A presente dissertação apresenta uma introdução ao estudo dos sinais complexos e alguns métodos de processamento adaptativo em torno destes. Técnicas baseadas na minimização do erro médio quadrático e correções dadas pela métrica são os elementos chave do presente trabalho.

Áreas como radar, biomedicina, eletromagnetismo etc, apresentam dados que podem ser tratados em notação complexa devido à possibilidade de representação em duas dimensões. O tratamento complexo então possibilita uma notação mais simples ao descrever fase e magnitude (ou parte real e imaginária) simultaneamente. Em processamento de sinais, o *design* de filtros que levam em conta a correlação que existe entre as diferentes componentes do sinal, motivam a necessidade dessa notação. Essa medida de similaridade pode ser associada a noção de circularidade ou não circularidade (relacionada com a impropriedade da distribuição, como será visto posteriormente).

Uma pergunta então pode ser feita: com base na simetria (ou assimetria) presente nos dados, é possível algum ganho no desempenho considerando (ou mesmo ignorando) estas propriedades? A resposta a isto é o cerne dessa dissertação. A técnica utilizada para evidenciar isso foi baseada em simulações que, na maioria das vezes, indicam uma melhora no desempenho (em termos de minimização do erro quadrático médio) ao se levar em conta características de impropriedade.

1.1 Contribuições

Essa dissertação tem como contribuição maior ser um compêndio sobre a teoria adaptativa de sinais amplamente lineares, reunindo alguns tópicos da área escritos em língua portuguesa. Os principais temas abordados são:

- i) WLLMS;
- ii) Método de Newton aumentado;
- iii) Gradiente natural aumentado.

Estes tópicos são os guias da escrita do presente trabalho.

Diversas publicações demonstram a relevância do tema. Desde as clássicas, como

em (Mandic; Goh, 2009) e (Schreier; Scharf, 2010), que discorrem extensivamente sobre temas correlatos. Outras que: estenderam o assunto para tópicos como estimação de sinais do tipo em (Picinbono; Chevalier, 1995) ou na descrição de dados como (Ollila, 2008) etc.. Recentemente o tema vem sendo amplamente utilizado, seja com teoria da informação (Qiu; Qian; Wang, 2022), com teoria de grafos (Amar; Routtenberg, 2022), em processamento de sinais mQAM (Yadav *et al.*, 2022) ou mesmo entre vários outros temas e áreas. O autor publicou o artigo (Bluhm; Cavalcante, 2022) juntamente com seu orientador que engloba o item (iii) da presente dissertação, tratando sobre o desenvolvimento da representação do gradiente natural complexo.

1.2 Organização

Esta dissertação é organizada da seguinte maneira:

- **Capítulo 2** - *Fundamentos do Cálculo $\mathbb{C}\mathbb{R}$*

Apresenta o embasamento matemático teórico básico do chamado cálculo $\mathbb{C}\mathbb{R}$ e conceitos básicos de análise complexa com intuito de apresentar as derivadas de Wirtinger. O cálculo integral complexo é visto, utilizando as expressões $\mathbb{C}\mathbb{R}$ obtidas. Também são expostos alguns exemplos de empregos da notação complexa fora da alçada de processamento de sinais. O capítulo termina com a apresentação de sinais em sua notação complexa.

- **Capítulo 3** - *Vetores e transformações amplamente lineares*

Neste capítulo o assunto abordado é o das transformações amplamente lineares. Estas são importantes na caracterização de filtros ditos amplamente lineares bem como na descrição vetorial e matricial aumentadas, derivadas de vetores, formas quadráticas, que serão importantes na otimização de funções custo com parâmetros complexos.

- **Capítulo 4** - *Estatística Complexa*

É uma capítulo de definições importantes. Descreve o necessário sobre estatística complexa para definição dos conceitos de circularidade e não-circularidade (relacionados com os respectivos conceitos de propriedade e improriedade), importantes na definição estatística de sinais complexos. Também é apresentada a modelagem dos sinais utilizados de acordo com essas propriedades.

- **Capítulo 5** - *Filtros Discretos Complexos*

Apresenta a modelagem de filtros lineares usualmente em teoria de Sinais e Sistemas. São descritos os modelos IIR e em seguida os transversais FIR. Neste capítulo é mostrado o *design* do filtro amplamente linear, construído a partir da sua versão linear FIR, do qual é

derivado o método WLLMS.

– **Capítulo 6** - *Método de Newton Aumentado*

A partir do estudo de séries de Taylor, o método de Newton é utilizado para otimização. O mesmo procedimento é feito com a versão em variáveis complexa da série de Taylor sobre uma função real, desta vez resultando no método de newton em sua forma aumentada.

– **Capítulo 7** - *Gradiente Natural Complexo*

Este capítulo apresenta uma inovação na escrita, nos moldes do cálculo $\mathbb{C}\mathbb{R}$, do chamado gradiente natural. Utilizando espaços tangentes à variedade expressa em bases complexificadas é desenvolvido o gradiente natural complexo aumentado.

– **Capítulo 8** - *Conclusão*

O conteúdo deste capítulo é o sobre encerramento dos tópicos presentes neste documento.

O próximo capítulo introduz o cálculo complexo, um importante ponto de partida e utilizado nas descrição de sinais complexos e funções custo de filtros que utilizam esses sinais.

2 FUNDAMENTOS DO CÁLCULO $\mathbb{C}\mathbb{R}$

O estudo de sinais complexos depende de um bom entendimento da análise complexa (cujo corpo é representado por $\mathbb{C}$). Três casos principais de aplicações em funções que utilizam variáveis em $\mathbb{C}$ utilizados no presente trabalho são:

- (i) a descrição estatística em termos de parâmetros complexos;
- (ii) a otimização de funções reais (não holomorfas) escritas em termos de variáveis complexas;
- (iii) a escrita de gradientes em notação complexa.

Essas descrições que utilizam variáveis complexas, como será visto, são uma contrapartida linear de suas versões com variáveis reais, i.e: uma função de variável real pode ser escrita em notação complexa, mesmo que a função em si não seja necessariamente complexa. Tais variáveis são relacionadas por uma transformação linear. Com isso é possível lidar com funções do tipo

$$f : \mathbb{C} \rightarrow \mathbb{R} . \quad (2.1)$$

Esse tipo de função é especialmente importante em otimização (caso (ii) e caso (iii)). Isso ocorre devido a impossibilidade de ordenamento do corpo complexo, tornando funções reais representantes ideais de funções custos a serem otimizadas (estas levam a valores reais e são arbitrariamente escolhidas para fornecer um valor mínimo ou máximo local). Funções do tipo (2.1) também podem ser utilizadas para descrever densidades de probabilidade sem necessariamente as otimizar (caso (i)).

Este capítulo busca destacar, de forma concisa e resumida, o conhecimento matemático básico requerido para a leitura sequencial da presente dissertação. Começando pela análise do cálculo complexo de funções holomorfas, como visto nos cursos de variável complexa ministrados nos cursos de graduação em engenharias, com as definições de derivada e integração complexa. Em sequência, é feita uma introdução ao cálculo de funções não-holomorfas, apresentando as expressões das derivadas de Wirtinger, com seus desdobramentos no emprego dessa nova notação como, por exemplo, integrais de linhas e elementos de área. Por último são apresentados exemplo de aplicações de variáveis complexas em outras áreas para em seguida serem abordadas na descrição de sinais.

2.1 Análise Complexa de Funções Holomorfas

Aqui é apresentado um resumo sobre números complexos, ferramenta fundamental nesta dissertação. Em seguida, o conceito de funções holomorfas é apresentado. Embora estas não sejam utilizadas diretamente na otimização de funções do tipo (2.1), elas são o ponto de partida e representam certos tipos de transformações entre variáveis complexas.

2.1.1 Propriedades dos números complexos

Um número complexo é um par ordenado munido de um produto especial. Seja o corpo complexo descrito por $\mathbb{C}$, existe um isomorfismo com o plano cartesiano descrito por $\mathbb{C} \approx \mathbb{R}^2$. Este é descrito como um par ordenado com as devidas propriedades definidas: soma

$$(a, b) + (c, d) = (a + c, b + d) , \quad (2.2)$$

produto complexo (responsável pela riqueza do corpo),

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc) \quad (2.3)$$

e a multiplicação por um escalar $c \in \mathbb{R}$

$$c \cdot (a, b) = (c \cdot a, c \cdot b) . \quad (2.4)$$

Com essa definição, a unidade imaginária é dada por $j = (0, 1)$, pois $(0, 1) \cdot (0, 1) = (-1, 0)$, possibilitando o cálculo de raízes de números negativos, bem como trabalhar em um espaço similar ao $\mathbb{R}^2$ mas com a vantagem de uma notação mais enxuta. O produto (2.3) possibilita um vasto conjunto de propriedades interessantes¹. O par complexo costuma ser representado por $(a, b) \equiv a + jb$.

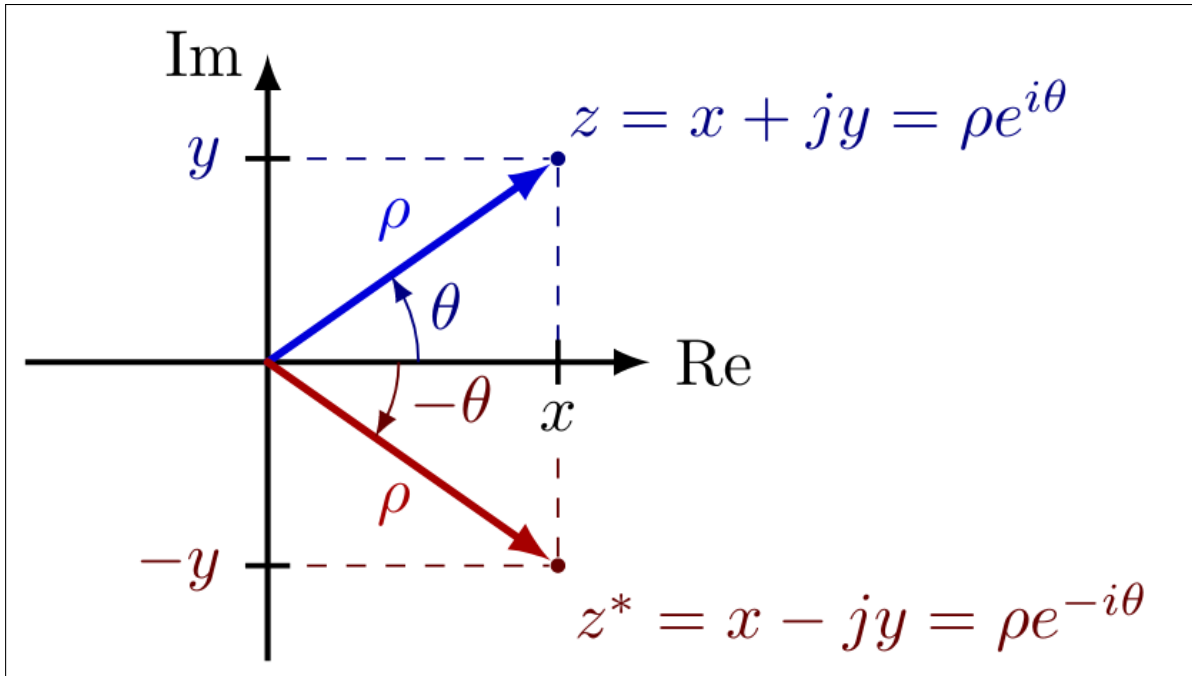
É importante salientar as propriedades geométricas de se trabalhar com os números complexos e seu produto: no caso real a presença do sinal negativo indica uma mudança de sentido de 180° do dado valor no eixo real, como exemplo seja $a \in \mathbb{R}$ então $-1 \cdot a = -a$, indica uma mudança de sentido no eixo real, e o valor -1 representa essa rotação. A unidade complexa, como resultado de duas ações que deve resultar na rotação em 180° (-1) é então escolhida como um valor que rotacione um dado valor em 90° , representada por j . Isso torna impossível

¹ Ainda há possibilidade de trabalhar com quatérnions, que seriam números isomorfos ao $\mathbb{R}^4$ com propriedades similares, mas esse tipo de abordagem não será feita neste trabalho, mas aumenta as dimensões na qual se pode trabalhar, ao mesmo tempo que se mantém as boas propriedades de um número complexo de alta dimensão.

a representação dos complexos em uma só dimensão: é necessário adicionar uma componente ortogonal, de forma que a unidade imaginária seja lateral e ortogonal. Devido a isso, tem-se a impossibilidade de ordenamento.

Dado um número complexo $z = x + jy$, a parte real é descrita por $\text{Re}\{z\} = x$ e a parte imaginária por $\text{Im}\{z\} = y$. Ao valor $z^* = x - jy$ dá-se o nome de conjugado de z , este desempenha um papel central na análise complexa. Também é importante notar que $\mathbb{R}$ está contido nos complexos visto que $x + j0 \in \mathbb{C}$ pode representar o valor real x . A Figura 1 a seguir demonstra graficamente tais propriedades. Da mesma forma que o par ordenado (x, y) no plano

Figura 1 – Representação gráfica do plano complexo para um dado número complexo $z = x + jy$ e seu conjugado



Fonte: elaborada pelo autor.

admitem a representação polar (ρ, θ) , um dado complexo $z = x + jy$ também pode ser posto em tal forma

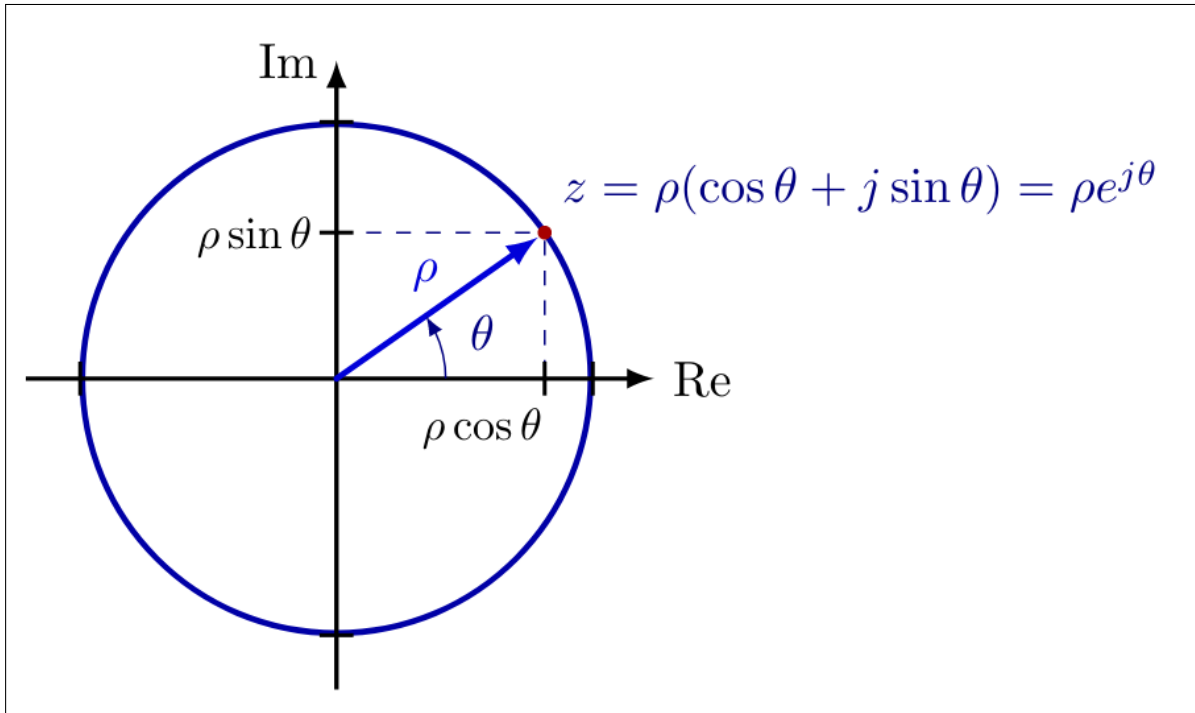
$$z = \rho (\cos \theta + j \sin \theta) . \quad (2.5)$$

com $\rho = \sqrt{x^2 + y^2}$, $\cos \theta = x/\rho$ e $\sin \theta = y/\rho$ como indicado na Figura 2. A notação polar é relacionada diretamente com outra notação particularmente útil vista a seguir.

Dada a exponencial representada em série de potências, sua extensão complexa para um valor $z \in \mathbb{C}$ é

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} . \quad (2.6)$$

Figura 2 – Representação gráfica do polar plano complexo para um dado número complexo $z = \rho(\cos \theta + j \sin \theta)$ e sua forma de Euler



Fonte: elaborada pelo autor.

Para (2.6) considere o valor $z = j\theta$,

$$e^{j\theta} = \sum_{k=0}^{\infty} \frac{(j\theta)^k}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k}}{(2k)!} + j \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k+1}}{(2k+1)!} . \quad (2.7)$$

O primeiro somatório corresponde ao cosseno e o segundo ao seno. A expressão simplificada:

$$e^{j\theta} = \cos \theta + j \sin \theta , \quad (2.8)$$

permite representar um número complexo por $z = \rho e^{j\theta}$ (veja novamente a Figura 2). Essa representação tira proveito das propriedades da exponencial para produtos e quocientes.

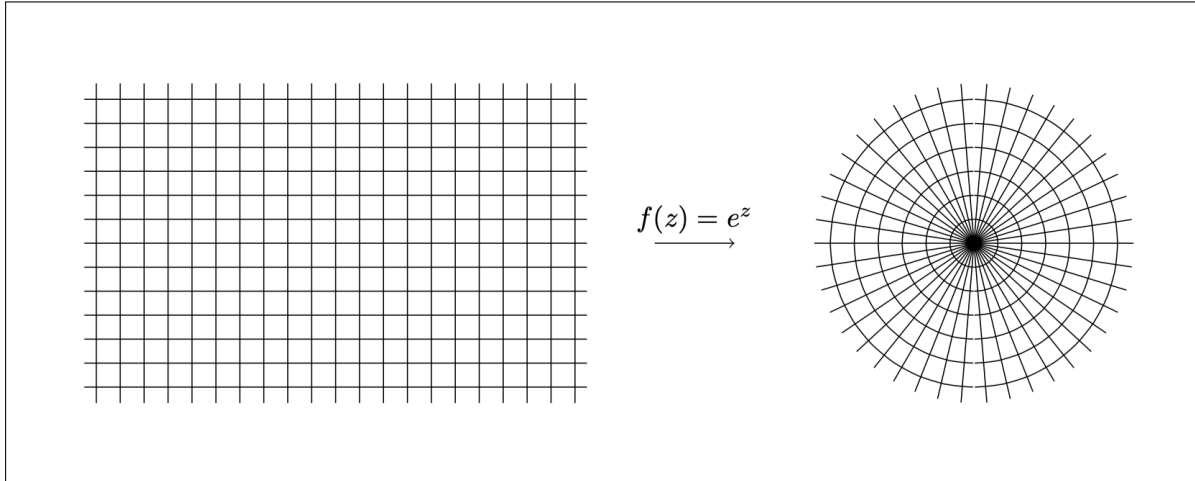
2.1.2 Derivada Complexa

Uma função complexa é uma transformação entre planos complexos. Dado um número complexo $z = x + jy \in \mathbb{C}$ com $x, y \in \mathbb{R}$, uma função complexa escalar é uma aplicação $f(z) = u + jv : \mathbb{C} \rightarrow \mathbb{C}$ tal que

$$f(x, y) = u(x, y) + jv(x, y).$$

com $u, v \in \mathbb{R}$. Escrevendo de forma compacta por $f(z) = w \in \mathbb{C}$, esta pode ser vista como um mapa do tipo $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. Um exemplo é mostrado na Figura 3, que exibe como as linhas

Figura 3 – Representação do mapeamento feito pela função complexa $f(z) = e^z$



Fonte: elaborada pelo autor.

retangulares são transformadas nas curvas radiais e concêntricas pela função dada.

Ao se trabalhar com funções complexas, conceitos básico de diferenciabilidade devem ser abordados, dentre eles os conceitos de limites, continuidade e derivada. Nestas, conceitos de limites e continuidade são similares à suas contrapartes em $\mathbb{R}^2$. É definida a derivada complexa em torno de um dado ponto z (Brown; Churchill, 2013) como o limite

$$f'(z) = \lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} . \quad (2.9)$$

Para que a expressão anterior seja bem definida em um ponto dado esta precisa ser independente da direção em que $\Delta z = \Delta x + j\Delta y$ se aproxima de z . Isso pode ser feito tomando o limite em (2.9) ora com $\Delta x \rightarrow 0$ para $\Delta y = 0$, ora o contrário. Também pode-se utilizar o ponto de vista de operadores lineares. Dado um ponto fixado $z_0 = x_0 + jy_0$, tal operador é

$$f'(z_0) \cdot h = \lim_{t \rightarrow 0} \frac{f(z_0 + th) - f(z_0)}{t} , \quad (2.10)$$

para algum $h = r + js$ ($\Delta z = t \cdot h$). Expandindo a equação anterior para $f = u + jv$, é obtida a expressão:

$$f' \cdot h = \frac{\partial u}{\partial x} r + \frac{\partial u}{\partial y} s + j \frac{\partial v}{\partial x} r + j \frac{\partial v}{\partial y} s . \quad (2.11)$$

O produto $f' \cdot h = f'(z_0) \cdot h$ deve satisfazer a multiplicação complexa em (2.3). Escreva o produto em termos de suas partes real e imaginária com $f' = f'_x + j \cdot f'_y$, então

$$f' \cdot h = f'_x \cdot r - f'_y \cdot s + j \cdot f'_y \cdot r + j \cdot f'_x \cdot s . \quad (2.12)$$

As expressões (2.11) e (2.12) são iguais se

$$\begin{cases} f'_x = \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \\ f'_y = \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \end{cases} . \quad (2.13)$$

Estas são denominadas de condições de Cauchy-Riemann, expressas por

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \text{ e } \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} . \quad (2.14)$$

Uma função é dita *holomorfa* quando satisfaz as condições 2.14. Essas condições impõem uma estrutura bastante restritiva sobre funções complexas.

Utilizando essas definições, pode-se reescrever a derivada complexa de funções holomorfas da forma:

$$f'(z) = f'_x(x,y) + j \cdot f'_y(x,y) = \frac{\partial u}{\partial x} + j \frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - j \frac{\partial u}{\partial y} , \quad (2.15)$$

Com a derivada calculada em um ponto qualquer $z \in \mathbb{C}$. Essa derivada, um operador linear que age sobre valores complexos por meio de um produto complexo, resulta em uma derivada direcional na direção desse valor dado. O resultado pode ser visualizado pela forma matricial da equação (2.11),

$$f' \cdot h = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} .$$

Como o operador é válido para qualquer valor $h \in \mathbb{C}$, este em sua forma matricial (Greene; Krantz; Society, 2006) é

$$f' = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x} & -\frac{\partial v}{\partial x} \\ \frac{\partial v}{\partial x} & \frac{\partial u}{\partial x} \end{bmatrix} .$$

Nesta forma é fácil visualizar que a derivada complexa em termos das componentes real e imaginária possui a mesma estrutura exibida na representação matricial de números complexos, que tem a forma:

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = a\mathbf{I} + b\mathbf{J} . \quad (2.16)$$

A unidade imaginária na forma matricial é dada pela regra:

$$\mathbf{J}^2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} . \quad (2.17)$$

Note a equivalência nas representações discutidas até o momento:

$$a + ib \approx \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \quad (2.18)$$

O símbolo $\approx$ denota o isomorfismo e deixa claro que a derivada de uma função holomorfa é também um número complexo que equivale a um operador linear.

A seguir serão examinados dois exemplos de derivação, um para uma função holomorfa e outro para uma não-holomorfa, para ilustrar essa distinção de maneira mais concreta.

Exemplo de uma função holomorfa: Seja uma função $f(z) = z^2$. Esta pode ser escrita como $f(z) = (x + jy)^2 = x^2 - y^2 - j2xy$, com $u = x^2 - y^2$ e $v = 2xy$. As condições (2.14)

$$\frac{\partial u}{\partial x} = 2x, \quad \frac{\partial v}{\partial y} = 2x, \quad \frac{\partial u}{\partial y} = -2y \quad \text{e} \quad \frac{\partial v}{\partial x} = 2y,$$

são satisfeitas. Logo $f(z)$ possui derivada complexa por (2.15) em todo $\mathbb{C}$ que vale

$$f'(z) = \frac{\partial u}{\partial x} + j \frac{\partial v}{\partial x} = 2x + j2y = 2z.$$

Regra similar àquela para o caso real para função do tipo $f(x) = x^2$ em que $f'(x) = 2x$.

Exemplo de uma função não holomorfa: Um caso trivial e interessante de uma função que não é holomorfa é o da função $f(z) = zz^* = |z|^2 = x^2 + y^2$. Como $u(x, y) = x^2 + y^2$ e $v(x, y) = 0$,

$$\frac{\partial u}{\partial x} = 2x, \quad \frac{\partial v}{\partial y} = 0, \quad \frac{\partial u}{\partial y} = 2y \quad \text{e} \quad \frac{\partial v}{\partial x} = 0.$$

A função não possui, portanto, derivadas complexas (2.9) em nenhum ponto. Esse exemplo é interessante por mostrar uma função do tipo $f : \mathbb{C} \rightarrow \mathbb{R}$, e a impossibilidade de diferenciação complexa usual devido às restrições impostas por (2.14).

De forma geral, uma função do tipo $f : \mathbb{C} \rightarrow \mathbb{R}$, que não seja uma função constante, não possui derivada complexa, visto que $v = 0$ para um dado $u(x, y)$ qualquer, quebrando a regra imposta por (2.14). Essas condições de derivação vistas até aqui são muito rígidas, não permitindo qualquer diferenciação de um funcional real em relação à parâmetros complexos, indesejável para funções de otimização que utiliza tais parâmetros. Para contornar esse problema utiliza-se um conceito um pouco diferente de derivadas, tratando os números complexos e seu par conjugado correspondente como uma transformação de coordenadas, como será visto adiante. Antes disso, são mostrados alguns resultados de integração para funções holomorfas a seguir.

2.1.3 Integração Holomorfa Complexa

A estrutura rígida de funções holomorfas também resultam em integrações mais fáceis de se trabalhar. Seja uma função complexa, que satisfaça as condições de Cauchy-Riemann (2.14), o teorema de Cauchy para integrais (Brown; Churchill, 2013) é dado por

$$\oint_C f(z)dz = 0 .$$

A expressão é válida pra qualquer caminho simplesmente conexo fechado C para o caso em que $f(z)$ não possua nenhum ponto singular dentro da região definida por este caminho. Para $f(z)$ holomorfa e com um ponto singular $z = z_0$ no interior de C a integração é dada por:

$$\oint_C \frac{f(z)}{z - z_0} dz = 2\pi j f(z_0) .$$

Essas expressões são importantes em cálculos de caminhos e de mudanças de variáveis e permitem lidar facilmente com essas funções bem comportadas, bastando o conhecimento do comportamento destas sobre o contorno C para determinar qualquer propriedade no interior desse contorno.

2.2 Derivadas de Wirtinger para Pares Escalares

Em otimização de funções reais com valores complexos podem ser utilizadas as chamadas derivadas (ou operadores) de Wirtinger (Adali; Haykin, 2010) descritos nessa seção. Esses operadores permitem o cálculo de derivadas para funções não holomorfas do tipo (2.1), comumente encontradas em otimização utilizando sinais complexos, desempenhando um papel central em análise de funções custo com parâmetros complexos.

O operadores são descritos em termos de transformações entre pares de escalares. Essas podem ser estudadas do ponto de vista de **transformações de coordenadas** ou pela **diferencial de uma função real** (que equivalem, respectivamente, à notação contravariante e covariante). No próximo capítulo é feita uma pequena generalização para dimensões maiores, o caso de vetores e transformações lineares.

2.2.1 Operadores descritos por meio de transformações de coordenadas

Dada uma variável complexa $z = x + jy$ e sua conjugada $z^* = x - jy$, em que (x, y) é o par cartesiano, é possível tratar o par de variáveis complexas como uma transformação de

coordenadas $(z(x, y), z^*(x, y))$ que admite inversa $(x(z, z^*), y(z, z^*))$. Aplicando a regra da cadeia usual do cálculo diferencial para uma função dada por $f = f(x(z, z^*), y(z, z^*))$,

$$\begin{aligned}\frac{\partial f}{\partial z} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial z} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial z}, \\ \frac{\partial f}{\partial z^*} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial z^*} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial z^*}.\end{aligned}$$

O mesmo pode ser feito para a inversa. Escrevendo $f(z, z^*) = g(z(x, y), z^*(x, y))$ a fim de facilitar a visualização, pode-se também tomar-se as derivadas em termos de x e y , de forma que,

$$\begin{aligned}\frac{\partial g}{\partial x} &= \frac{\partial g}{\partial z} \frac{\partial z}{\partial x} + \frac{\partial g}{\partial z^*} \frac{\partial z^*}{\partial x}, \\ \frac{\partial g}{\partial y} &= \frac{\partial g}{\partial z} \frac{\partial z}{\partial y} + \frac{\partial g}{\partial z^*} \frac{\partial z^*}{\partial y}.\end{aligned}$$

Em qualquer um dos casos anteriores, é possível obter os operadores

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - j \frac{\partial}{\partial y} \right) \quad \text{e} \quad \frac{\partial}{\partial z^*} = \frac{1}{2} \left(\frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right). \quad (2.19)$$

Estes operadores independem da função f escolhida e podem ser utilizados para funções não necessariamente holomorfas, visto que não foi cobrado nenhum comportamento especial da função f utilizada. As expressões acima são chamadas de derivadas de Wirtinger, e formam uma extensão natural da mudança de sistema coordenado para funções reais que dependam de números complexos.

2.2.2 Operadores descritos pela diferencial de uma função real

O resultado da seção anterior pode ser obtido também por meio da diferencial (formas diferenciais) de uma função real que utiliza valores complexos. Isso é possível com uso dos incrementos $dz = dx + jdy$, $dz^* = dx - jdy$ e da expressão

$$df = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial z^*} dz^*. \quad (2.20)$$

Esta expressão é escolhida arbitrariamente para equivaler ao resultado real, ou seja,

$$df = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial z^*} dz^* = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy. \quad (2.21)$$

Após a substituição dos incrementos dx , dy em termos de dz e dz^* em (2.21), a expressão para a diferencial é dada por:

$$df = \frac{\partial f}{\partial x} \left(\frac{dz + dz^*}{2} \right) + \frac{\partial f}{\partial y} \left(\frac{dz - dz^*}{2j} \right). \quad (2.22)$$

A expressão em (2.20) pode então ser expressa em termos das derivadas parciais que dependam de (x, y) como

$$df = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial z^*} dz^* = \frac{1}{2} \left(\frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right) dz + \frac{1}{2} \left(\frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right) dz^* . \quad (2.23)$$

O resultado é igual àquele obtido em (2.19), demonstrando a equivalência de representação. O estudo de formas generalizadas e coordenadas gerais será aprofundado posteriormente.

2.2.3 Isomorfismos importantes

O corpo complexo, por natureza admite um isomorfismo natural com o plano cartesiano, denotado por $\mathbb{C} \approx \mathbb{R}^2$. Isso representa a equivalência entre números complexos e pares ordenados no plano, com o uso do produto complexo em (2.3). Outro isomorfismo mostrado foi o de matrizes do tipo (2.16) e os complexos. A seção anterior adicionou mais um: o par $(z, z^*) \in \mathbb{C}_*^2$. Isso significa que é possível representar um número complexo $z \in \mathbb{C}$ como um par $(z, z^*) \in \mathbb{C}_*^2$.

As representações e seus isomorfismos não garantem que qualquer função que atua sobre qualquer um desses espaços seja ou não holomorfa, só mostram a equivalência entre as variáveis utilizadas. A seguir é aplicado o uso dos operadores (2.19) para funções holomorfas e não holomorfas.

2.2.4 Aplicações da derivadas de Wirtinger

Para verificar como as condições de Cauchy-Riemann (2.14) considere a função $h(z) = g(z, z^*) = f(x, y) = u + jv$. Em seguida aplique a essa função o operador conjugado dado em (2.19), o resultado deve ser:

$$\frac{\partial g}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right) = \frac{1}{2} \left[\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} + j \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \right] .$$

Para que a condição (2.14) seja satisfeita a função dada deve satisfazer a condição

$$\frac{\partial g}{\partial z^*} = 0 . \quad (2.24)$$

A condição acima é uma forma equivalente do teorema de Cauchy-Riemann na notação de Wirtinger. Para que uma função seja holomorfa esta deve ser independente (explicitamente) do

seu complexo conjugado². Da mesma forma, a derivada não conjugada de (2.19), é dada por:

$$\frac{\partial g}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right) = \frac{\partial u}{\partial x} + j \frac{\partial v}{\partial x} = h'(z). \quad (2.25)$$

A expressão 2.25 obtida acima é a derivada complexa para funções holomorfas. De maneira resumida, para uma dada função arbitrária, esta é holomorfa se (2.24) é satisfeita. Se, e somente se, esta for holomorfa a derivada complexa é expressa por (2.25).

Exemplo de função holomorfa: Seja a função $f(z) = g(z, z^*) = z^2$ já utilizada, holomorfa por não depender da variável z^* . De fato, por (2.24)

$$\frac{\partial g}{\partial z^*} = \frac{\partial(z^2)}{\partial z^*} = 0.$$

Isso significa que g satisfaz as condições de Cauchy-Riemann. A derivada complexa é dada por:

$$\frac{\partial g}{\partial z} = 2z,$$

equivalente a $f' = 2x + j2y$, de acordo com o resultado já encontrado em exemplo similar.

Um ponto crucial é que os operadores (2.19) do cálculo $\mathbb{C}\mathbb{R}$ formam uma extensão do cálculo complexo usual, permitindo também o uso em funções não holomorfas, facilitando a derivação em funções escritas da forma $g(z, z^*)$, já que o par (z, z^*) é visto como um par de variáveis quaisquer, tratadas como se fossem independentes entre si. Como exemplo, uma função não holomorfa do tipo $f : \mathbb{C} \rightarrow \mathbb{R}$ é visto a seguir.

Exemplo de função não holomorfa: Como já visto anteriormente a função $f(z) = g(z, z^*) = zz^* = |z|^2 = x^2 + y^2$ não possui uma derivada complexa definida por (2.9), visto que essa não satisfaz a condição (2.24). De fato,

$$\frac{\partial g}{\partial z^*} = z \neq 0, \text{ verdade somente para } z = 0.$$

É possível o cálculo da derivada em relação a z (derivada não-conjugada) expressa por

$$\frac{\partial g}{\partial z} = z^*.$$

Nesse caso a derivada acima não representa (2.9), pois a função não é holomorfa. É possível a derivação de acordo com as mesmas equações em termos de $\partial/\partial x$ e $\partial/\partial y$.

² A forma de escrever funções holomorfas somente como função de z , como $h(z)$, já evidencia esse comportamento: uma função holomorfa não depende explicitamente de seu conjugado. Não está errado escrever $h(z) = zz^*$, como a operação de conjugação implícita, mas isso quebra o ponto de vista de z^* como uma nova variável ao lado de z , bem como devido a não linearidade da operação e, por isso, será evitada aqui.

Os operadores (2.19) permitem a otimização em relação ao par (z, z^*) , mesmo para funções que não possuem a derivada complexa usual, como mostrado no exemplo anterior.

Em termos de cálculos de pontos críticos (pontos em que a função dada possui derivadas nulas) os operadores equivalem às condições de otimizações reais. Dada uma função f real, tanto a expressão

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right) = 0, \quad (2.26)$$

quanto

$$\frac{\partial f}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right) = 0, \quad (2.27)$$

representam expressões para o ponto crítico de f , implicitamente equivalentes à $\partial/\partial x = 0$ e $\partial/\partial y = 0$.

Também é importante verificar a extensão natural que se faz com a regra da cadeia. Dada uma função composta $f(z, z^*) = h(g(z, z^*), g^*(z, z^*))$, pelas definições de derivação vistas anteriormente é possível escrever:

$$\frac{\partial f}{\partial z^*} = \frac{\partial h(g, g^*)}{\partial z^*} = \frac{1}{2} \left(\frac{\partial h(g, g^*)}{\partial x} + j \frac{\partial h(g, g^*)}{\partial y} \right). \quad (2.28)$$

Pelas regras usuais do cálculo, cada termo em relação à coordenada real pode ser escrito de forma compacta (de forma similar para a outra componente)

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial g} \frac{\partial g}{\partial x} + \frac{\partial h}{\partial g^*} \frac{\partial g^*}{\partial x}.$$

Substituindo em (2.28), e agrupando os termos comuns:

$$\frac{\partial f}{\partial z^*} = \frac{\partial h}{\partial g} \left[\frac{1}{2} \left(\frac{\partial g}{\partial x} + j \frac{\partial g}{\partial y} \right) \right] + \frac{\partial h}{\partial g^*} \left[\frac{1}{2} \left(\frac{\partial g^*}{\partial x} + j \frac{\partial g^*}{\partial y} \right) \right].$$

A partir desse resultado é possível escrever a regra da cadeia de funções no cálculo $\mathbb{C}\mathbb{R}$, permitindo o cálculo sobre mudança de variáveis, dada pela expressão:

$$\frac{\partial f}{\partial z^*} = \frac{\partial h}{\partial g} \frac{\partial g}{\partial z^*} + \frac{\partial h}{\partial g^*} \frac{\partial g^*}{\partial z^*}.$$

O conceito de derivada no cálculo $\mathbb{C}\mathbb{R}$ também pode ser estendida para cálculo de múltiplas variáveis, como será visto mais adiante, para vetores formados por pares vetoriais com seu conjugado.

2.3 Cálculos de Integrais

Com regras de derivações definidas, é interessante perguntar se é possível realizar algum cálculo integral sobre funções não holomorfas escritas na notação $\mathbb{CR}$ apresentada. Considere uma região simples e fechada A no $\mathbb{R}^2 \approx \mathbb{C} \approx \mathbb{C}_*^2$, para o cálculo de integrais sobre tais regiões, utiliza-se a relação:

$$\iint_A f(x,y) dx dy = \frac{1}{2j} \oint_{\partial A} F(z, z^*) dz . \quad (2.29)$$

A expressão acima utiliza a definição:

$$\frac{\partial F(z, z^*)}{\partial z^*} \equiv f(x, y) .$$

É importante tomar nota que para uma dada função $F(z, z^*) \triangleq F = U + jV$, é possível escrever esta em termos de (z, z^*) ou (x, y) , por meio de uma mudança de variáveis. A integral de linha então pode ser escrita como

$$\frac{1}{2j} \oint_{\partial A} F dz = \frac{1}{2j} \oint_{\partial A} (U + jV)(dx + jdy) = \frac{1}{2j} \left[\oint_{\partial A} (U dx - V dy) + \oint_{\partial A} (V dx + U dy) \right] .$$

Utilizando o teorema Green no plano

$$\frac{1}{2j} \oint_{\partial A} F dz = \frac{1}{2j} \iint \left(-\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} \right) dx dy + \frac{1}{2} \iint \left(\frac{\partial U}{\partial x} - \frac{\partial V}{\partial y} \right) dx dy$$

e expandindo a expressão da derivada em relação a variável conjugada por:

$$\frac{\partial F}{\partial z^*} = \frac{1}{2} \left(\frac{\partial F}{\partial x} + j \frac{\partial F}{\partial y} \right) = \frac{1}{2} \left[\frac{\partial U}{\partial x} - \frac{\partial V}{\partial y} + j \left(\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y} \right) \right] ,$$

é possível escrever a integral de forma simplificada por:

$$\frac{1}{2j} \oint_{\partial A} F dz = \iint_A f(x,y) dx dy . \quad (2.30)$$

A expressão acima é uma forma do teorema de Green para cálculos para funções ditas não holomorfas, permitindo a troca de uma integral de caminho por uma de área sobre uma dada região, relacionando a forma usual de integração dupla com sua equivalente $\mathbb{CR}$. Também demonstra a necessidade de apenas uma variável de integração nessa última, dada a dependência entre o número complexo e seu conjugado. Tal dependência poderia ser questionada na formulação das derivadas, mas estas se tratam de elementos infinitesimais, linearmente independentes. O mesmo não ocorre em caminhos de integração.

É possível relacionar a mudança de variável de forma que o elemento diferencial de área seja dado por

$$dxdy = Jdzdz^*,$$

através do jacobiano da transformação, dado por

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial z^*} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial z^*} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2j} & -\frac{1}{2j} \end{vmatrix} = \frac{j}{2}.$$

Isso permite reescrever a equação (2.30) como³

$$\frac{1}{2j} \oint_{\partial A} F dz = \iint_A f(x,y) dxdy = \frac{j}{2} \iint_{A_{\mathbb{C}}} f(z,z^*) dzdz^* . \quad (2.31)$$

Esse tipo de relação permite escrever e calcular integrais, sobre domínios puramente complexos, e a relação que existe entre tais integrais, bem como relacionar funções em suas diferentes representações.

Este tópico encerra a teoria de variáveis complexas e a introdução ao cálculo $\mathbb{CR}$. A próxima seção exhibe exemplos de aplicações gerais de variáveis complexas.

2.4 Exemplos da representação complexa

O processamento de sinais é beneficiado extensivamente pelo uso de números complexos, mesmo que sua natureza seja real. Tal fato ocorre em diversas áreas como no eletromagnetismo, em mecânica dos fluidos etc. Mesmo em matemática pura o cálculo complexo facilita a resolução de integrais reais, utilizando-se de suas extensões analíticas complexas, de mais fácil integração, para se obter o resultado procurado no corpo dos números reais.

Quase a totalidade de livros introdutórios de processamento de sinais já utilizam sinais complexos em seus capítulos iniciais. Isso não acontece por acaso: tal notação traz um ganho considerável de abstração e redução nos cálculos, mostrando o poder que dos complexos. Infelizmente, ao mesmo tempo que há tal direcionamento, tais conteúdos não destacam algumas propriedades importantes que os sinais complexos possuem devido à sua própria natureza matemática: a *circularidade* (ou a falta desta).

A presente seção capítulo traz exemplos gerais de emprego de notação complexa que descrevem problemas gerais. Também é feita uma introdução à teoria de sinais e sua notação complexa.

³ Equivalem a $\frac{j}{2} dz \wedge dz^* = dx \wedge dy$, (com $\wedge$ representando o produto exterior entre as diferenciais complexas).

2.4.1 Soluções harmônicas para a equação de onda

A seguir veremos como as equações de Maxwell em sua forma diferencial em coordenadas cartesianas (Griffiths, 2013)

$$\nabla \cdot \mathbf{D} = \rho, \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (2.32)$$

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (2.33)$$

possuem soluções harmônicas determinadas considerando os campos como exponenciais complexas. Para o conjunto de equações acima, ρ representa a densidade de carga do meio, o vetor $\mathbf{E}$ o campo elétrico e $\mathbf{B}$ o vetor campo magnético. Além disso, $\mathbf{J} = \sigma \mathbf{E}$ é chamada de a densidade de corrente. As relações entre os campos são dadas por $\mathbf{B} = \mu \mathbf{H}$ e $\mathbf{D} = \epsilon \mathbf{E}$, com tais constantes supostas conhecidas. O operador diferencial utilizado em coordenadas cartesianas é

$$\nabla = \frac{\partial}{\partial x_1} \mathbf{e}_1 + \frac{\partial}{\partial x_2} \mathbf{e}_2 + \frac{\partial}{\partial x_3} \mathbf{e}_3. \quad (2.34)$$

As operações $\cdot$ e $\times$ aqui representam o produto escalar e o produto vetorial respectivamente. Os vetores $\mathbf{e}_1$, $\mathbf{e}_2$ e $\mathbf{e}_3$ são os vetores unitários nas direções x, y e z respectivamente. Pressupõe-se que o leitor já esteja familiarizado com tais operações e notações dos cursos iniciais de cálculo.

Para a solução harmônica considere o problema para o campo elétrico cartesiano $\mathbf{E} = (E_1, E_2, E_3) = E_1 \mathbf{e}_1 + E_2 \mathbf{e}_2 + E_3 \mathbf{e}_3$. Tal campo pode depender da posição no espaço $\mathbf{r} = (x_1, x_2, x_3) = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + x_3 \mathbf{e}_3$ e do tempo simultaneamente, sendo expresso por $\mathbf{E} = \mathbf{E}(\mathbf{r}, t)$. O mesmo ocorre com o campo magnético $\mathbf{B} = \mathbf{B}(\mathbf{r}, t)$.

A fim de resolver o conjunto de equações de Maxwell para o campo elétrico, parte-se da expressão para o rotacional ($\nabla \times$) em (2.32). Aplicando novamente o rotacional, considerando os parâmetros isométricos e constantes:

$$\nabla \times \nabla \times \mathbf{E} = -\nabla \times \frac{\partial \mathbf{B}}{\partial t} = -\mu \nabla \times \frac{\partial \mathbf{H}}{\partial t}. \quad (2.35)$$

Ao utilizar a identidade vetorial $\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$ no resultado anterior chega-se a

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu \frac{\partial}{\partial t} (\nabla \times \mathbf{H}). \quad (2.36)$$

Substituindo a expressão do rotacional para o campo magnético em (2.33), e considerando o meio livre de quaisquer cargas $\nabla \cdot \mathbf{E} = 0$, além de linear, homogêneo, isotrópico, a expressão pode ser escrita como:

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu \frac{\partial}{\partial t} \left(\sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t} \right),$$

$$\nabla^2 \mathbf{E} = \mu \sigma \frac{\partial \mathbf{E}}{\partial t} + \mu \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} . \quad (2.37)$$

A expressão (2.37) acima é uma equação diferencial em termos do campo elétrico. A partir dela é possível encontrar as soluções harmônicas no tempo, também chamadas de estacionárias periódicas. Essas soluções estão intimamente ligadas ao uso da exponencial complexa (solução harmônica), que facilitam a resolução. Para isso considere o campo elétrico da forma complexa $\hat{\mathbf{E}}(\mathbf{r}, t) = \tilde{\mathbf{E}}(\mathbf{r})e^{j\omega t}$. As derivadas em relação ao tempo são

$$\frac{\partial \hat{\mathbf{E}}}{\partial t} = j\omega \tilde{\mathbf{E}}e^{j\omega t}$$

e

$$\frac{\partial^2 \hat{\mathbf{E}}}{\partial t^2} = -\omega^2 \tilde{\mathbf{E}}e^{j\omega t} .$$

A derivada segunda em relação à posição é

$$\nabla^2 \hat{\mathbf{E}} = (\nabla^2 \tilde{\mathbf{E}}) e^{j\omega t} .$$

Substituindo as expressões anteriores para a derivada na equação (2.37):

$$\nabla^2 \tilde{\mathbf{E}}e^{j\omega t} = (\mu \sigma j\omega \tilde{\mathbf{E}} - \mu \epsilon \omega^2 \tilde{\mathbf{E}}) e^{j\omega t} .$$

A expressão anterior pode ser simplificada caso seja posta na forma

$$\nabla^2 \tilde{\mathbf{E}} - \gamma^2 \tilde{\mathbf{E}} = 0 , \quad (2.38)$$

com $\gamma^2 = j\omega\mu(\sigma + j\omega\epsilon)$. É importante notar que a expressão (2.38) não possui derivadas temporais. A solução plana monocromática para um campo polarizado em x_1 , que viaja na direção do eixo x_3 é $\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{E}(x_3)\mathbf{e}_1$, pois satisfaz a equação diferencial:

$$\frac{\partial^2 \tilde{E}}{\partial x_3^2} - \gamma^2 \tilde{E} = 0 . \quad (2.39)$$

A solução geral para a expressão acima é da forma $\tilde{E} = E_1 e^{\gamma x_3} + E_0 e^{-\gamma x_3}$. Tomando $\gamma = \alpha + j\beta$, tem-se $\tilde{E} = E_1 e^{\alpha x_3} e^{j\beta x_3} + E_0 e^{-\alpha x_3} e^{-j\beta x_3}$, cuja parte real é dada por $\mathbf{E} = \text{Re}\{\tilde{E}e^{j\omega t}\mathbf{e}_1\}$,

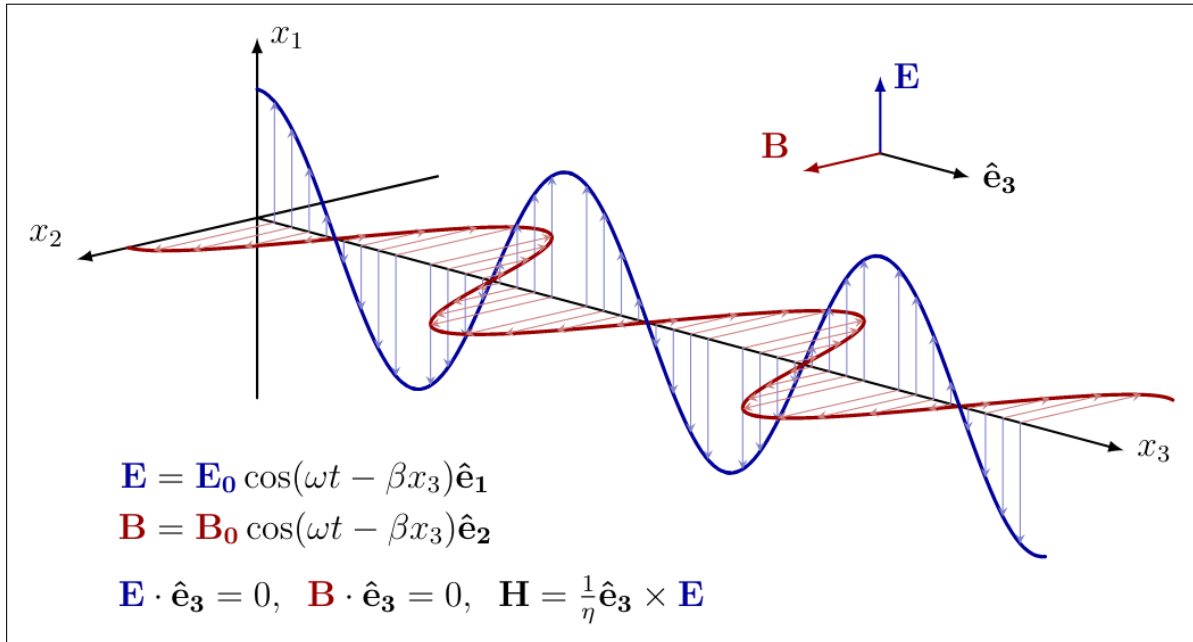
$$\mathbf{E} = [E_1 e^{\alpha x_3} \cos(\omega t + \beta x_3) + E_0 e^{-\alpha x_3} \cos(\omega t - \beta x_3)] \mathbf{e}_1 , \quad (2.40)$$

que é a solução geral para as condições dadas: é também uma solução harmônica em relação à posição com a presença de atenuação. A solução sem atenuação $\alpha = 0$ para uma onda unidirecional $A = 0$ é

$$\mathbf{E} = E_0 \cos(\omega t - \beta x_3) \mathbf{e}_1 . \quad (2.41)$$

Procedendo de forma semelhante para o campo magnético verifica que a solução é da mesma forma com a relação entre as amplitudes dada por $E_0/H_0 = \eta$. A Figura 4 demonstra as soluções para o caso sem atenuação das equações de Maxwell.

Figura 4 – Solução harmônica sem perdas ($\alpha = 0$) e unidirecional para as equações de Maxwell



Fonte: elaborada pelo autor.

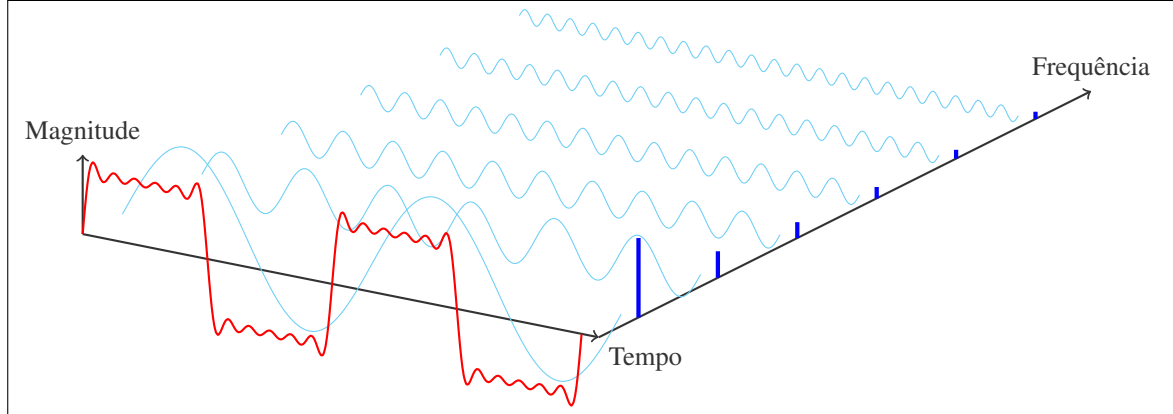
2.4.2 A transformada de Fourier

Como outro exemplo temos uma ferramenta matemática bastante utilizada em processamento de sinais: a transformada de Fourier. Esta é uma transformada integral que utiliza como núcleo a forma de Euler para números complexos, dada por (Oppenheim; Willsk; Young, 1983)

$$F(\omega) = \mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt . \quad (2.42)$$

Com isso uma função dada pode ser decomposta em componentes harmônicas, como indicado na Figura 5, indicando a decomposição da função representada em vermelho em suas componentes harmônicas (em azul cada).

Figura 5 – Transformada de Fourier, que decompõe a função (dada em vermelho) em suas componentes harmônicas (em azul)



Fonte: elaborada pelo autor.

Com uso das propriedades de derivação da transformada, dadas por:

$$\mathcal{F} \left\{ \frac{\partial f(\mathbf{r}, t)}{\partial t} \right\} = j\omega F(\mathbf{r}, \omega)$$

e

$$\mathcal{F} \left\{ \frac{\partial^2 f(\mathbf{r}, t)}{\partial t^2} \right\} = -\omega^2 F(\mathbf{r}, \omega) ,$$

é possível a resolução de uma equação do tipo (2.37). No caso de uma função real, $f(\mathbf{r}, t)$ expressão dada por

$$\nabla^2 f(\mathbf{r}, t) = \mu\sigma \frac{\partial f(\mathbf{r}, t)}{\partial t} + \mu\epsilon \frac{\partial^2 f(\mathbf{r}, t)}{\partial t^2} . \quad (2.43)$$

tem sua transformada calculada por

$$\mathcal{F} \left\{ \nabla^2 f(\mathbf{r}, t) = \mu\sigma \frac{\partial f(\mathbf{r}, t)}{\partial t} + \mu\epsilon \frac{\partial^2 f(\mathbf{r}, t)}{\partial t^2} \right\} ,$$

que equivale a

$$\nabla^2 F(\mathbf{r}, \omega) = \mu\sigma j\omega F(\mathbf{r}, \omega) - \mu\epsilon \omega^2 F(\mathbf{r}, \omega) . \quad (2.44)$$

Essa expressão é similar àquela encontrada em (2.38), desta vez com $\gamma^2 = j\omega\mu(\sigma + j\omega\epsilon)$.

Dessa forma:

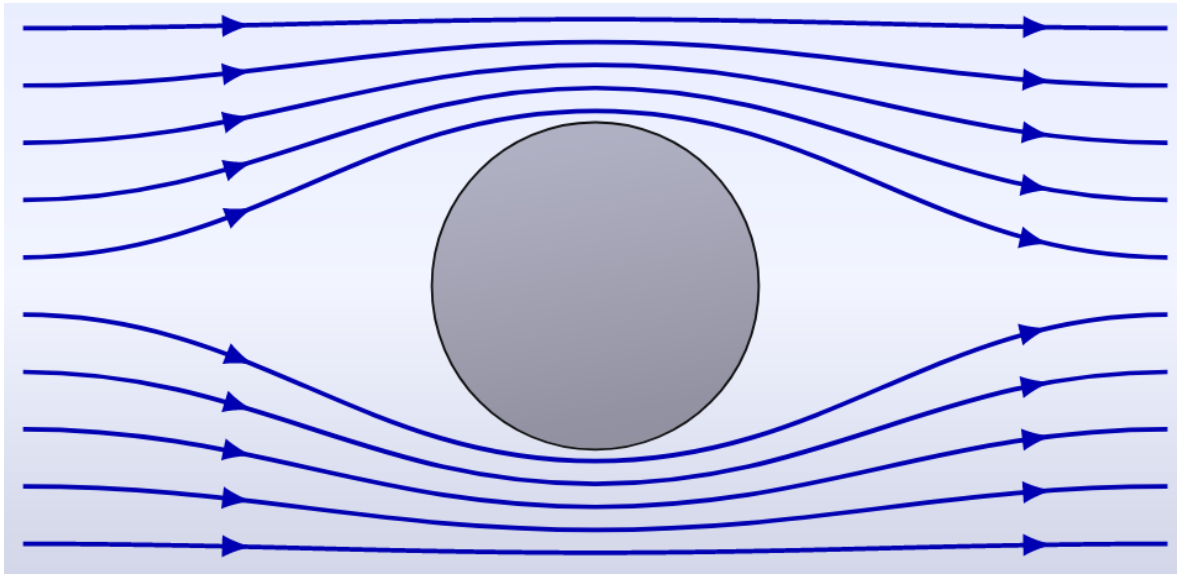
$$\nabla^2 F(\mathbf{r}, \omega) - \gamma^2 F(\mathbf{r}, \omega) = 0 . \quad (2.45)$$

A comparação entre a expressão (2.45) e (2.38) mostra que considerar a solução harmônica no tempo como método de solução equivale a encontrar as decomposições periódicas da equação diferencial dada.

2.4.3 Dinâmica dos Fluidos

O uso de variáveis complexas e funções holomorfas auxilia até no estudo de perfis bi-dimensionais de escoamentos laminares, comuns no estudo de dinâmica dos fluidos. A Figura 6 mostra um desses perfis onde as linhas azuis indicam o campo de velocidades do fluido na presença de um perfil circular. O uso de variáveis complexas na análise de escoamentos desse

Figura 6 – Perfil de escoamento laminar indicando o campo de velocidades ao passar por um perfil circular



Fonte: elaborada pelo autor.

tipo é possível devido às equações de Cauchy-Riemann representarem as mesmas condições assumidas para o escoamento, como será demonstrado a seguir.

Dado o campo de velocidades plano $V = (u, v)$, que descreve a velocidade em cada direção coordenada, pode-se aplicar duas condições ao fluido: $\nabla \cdot V = 0$ (fluxo sem sumidouros ou fontes) e $\nabla \wedge V = 0$ (fluxo irrotacional) (Acheson, 1990). Com isso, o campo de velocidade exhibe intrinsecamente as relações:

$$\begin{aligned}\nabla \cdot V &= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \\ \nabla \wedge V &= \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0.\end{aligned}$$

Tais condições são exatamente as condições (2.13) para uma função holomorfa do tipo $f = u - jv$, que recebe o nome de velocidade complexa. Definir um potencial F cuja derivada seja igual à velocidade⁴ facilita a solução. Esse campo potencial é uma função holomorfa $F = \phi + j\psi$ dada

⁴ A existência de tal potencial é garantida pela condição $\nabla \wedge V = 0$.

pela relação:

$$\frac{dF}{dz} = f. \quad (2.46)$$

A velocidade do fluido é encontrada tomando a norma

$$|V| = |f| = \left| \frac{dF}{dz} \right|. \quad (2.47)$$

A função de campo pode ser integrada, definindo caminhos descritos pelo escoamento.

2.5 Sinais e sua representação complexa

A teoria de sinais e sistemas lineares não poderia subestimar a importância da representação por números complexos (ou vetores). Os sinais são essenciais para a transmissão de informações, visto que representam conjuntos de valores que codificam dados a serem comunicados. Esta seção explora as representações de sinais tanto reais quanto complexos em contextos contínuos e discretos, com foco particular nesta última modalidade.

2.5.1 Sinais reais

Estes utilizam valores reais para representar a informação. No caso contínuo, um sinal real escalar é uma função do tipo

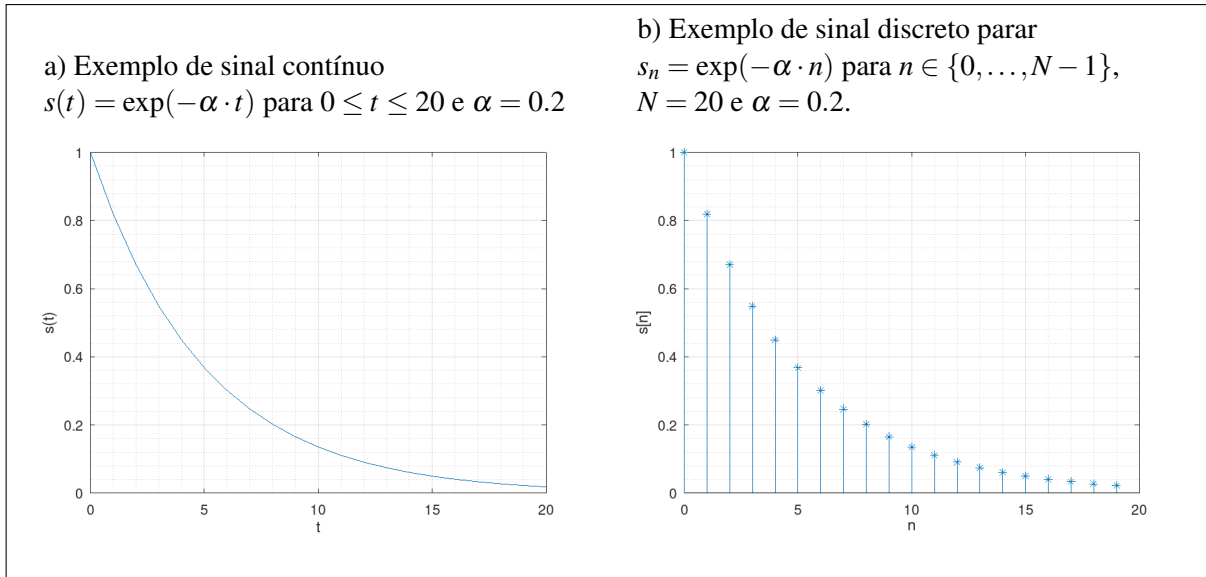
$$s(t) : I \subset \mathbb{R} \rightarrow \mathbb{R} \quad (2.48)$$

em que $t \in I$ representa o tempo que flui continuamente em um dado intervalo I . Um sinal escalar discreto é uma sequência do tipo

$$s[n] =: I \subset \mathbb{Z} \rightarrow \mathbb{R}, \quad (2.49)$$

para um intervalo I de tempo discreto n . A seguinte notação será adotada: sinais contínuos dependerão de tempo por “()” ao passo que os discretos por “[]” ou por um índice s_n . Os sinais podem ter uma duração finita ou não, por isso a escolha de um intervalo para representá-los. A Figura 7 ilustra dois exemplos de sinais em um intervalo específico, representando tanto o caso contínuo quanto o discreto de uma função exponencial. O sinal discreto é uma versão amostrada do sinal contínuo, onde valores igualmente espaçados no tempo T (período de amostragem) são selecionados a partir do sinal original, conforme a equação $s[n] = s(nT)$.

Figura 7 – Exemplos de um sinal $f(t) = \exp(-\alpha \cdot t)$ em (a) e em (b) como uma versão amostrada do caso contínuo com $T = 1$ unidade de tempo



Fonte: elaborada pelo autor.

2.5.2 Sinais complexos

Como extensão do caso real, um sinal complexo escalar real é uma função do tipo

$$s(t) : I \subset \mathbb{R} \rightarrow \mathbb{C}, \quad (2.50)$$

resultando em uma parametrização da forma $s(t) = x(t) + jy(t)$, com $x(t)$ e $y(t)$ reais. De forma análoga, um sinal complexo escalar discreto é uma função que depende de valores inteiros de tempo de forma que

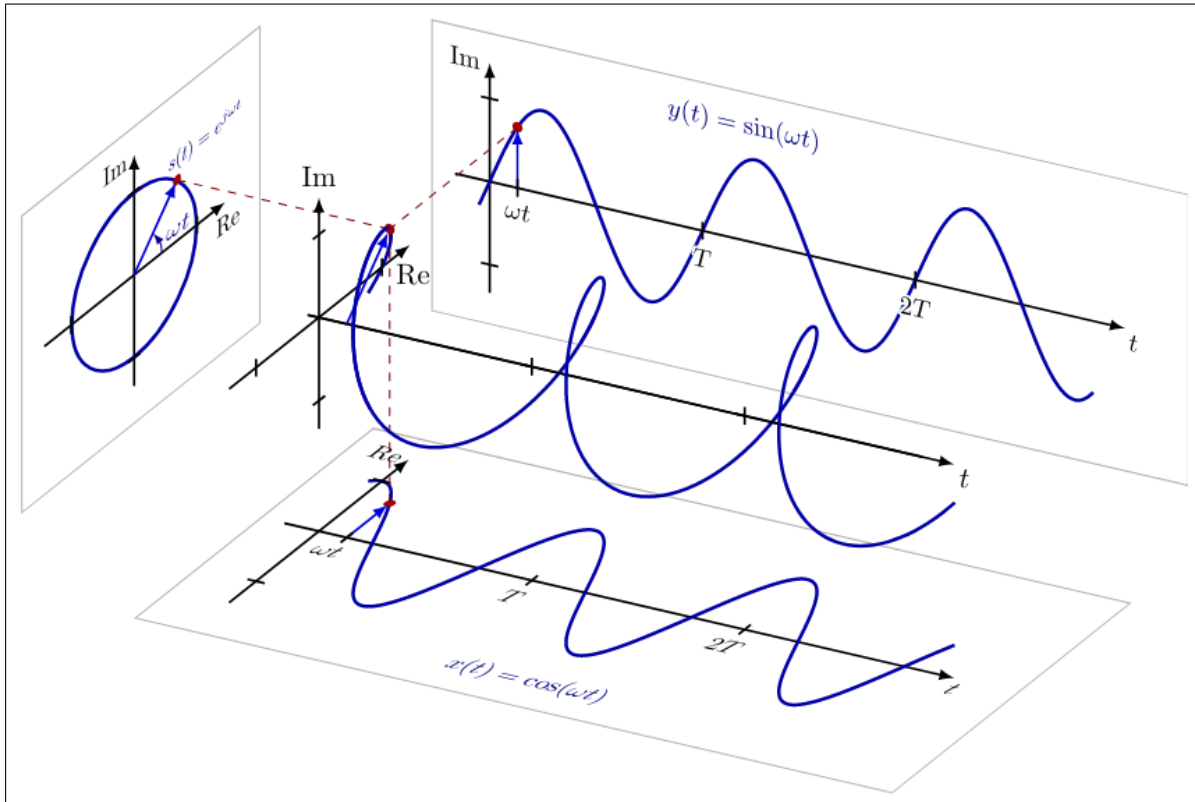
$$s[n] : I \subset \mathbb{Z} \rightarrow \mathbb{C}, \quad (2.51)$$

com a forma cartesiana complexa $s[n] = x[n] + jy[n]$, para $x[n]$ e $y[n]$ discretos.

Um exemplo de sinal contínuo complexo é $s(t) = \cos \omega t + j \sin \omega t$, que pode ser visualizado na Figura 8 a seguir. Na figura é possível visualizar o sinal ao longo do tempo como uma helicóide, o perfil da trajetória que corresponde ao plano complexo e os perfis correspondentes aos eixos real e imaginário em função do tempo.

Como exemplo de sinal discreto, seja $s[n] = \cos(\alpha n) + j \sin(\alpha n)$ para um dado intervalo. A parte real do sinal está representado na Figura 9 que representa a componente real e sua dependência do tempo discreto n . A Figura 10 (a) mostra o perfil do sinal dado e em (b) seu comportamento com o tempo.

Figura 8 – Representação gráfica de um sinal do tipo $s(t) = e^{j\omega t} = \cos \omega t + j \sin \omega t$



Fonte: elaborada pelo autor.

A vantagem da utilização da forma complexa em vez de um par cartesiano para sinais reside na capacidade de empregar a forma de Euler:

$$s[n] = r[n] \cos \theta[n] + jr[n] \sin \theta[n] = r[n] e^{j\theta[n]}. \quad (2.52)$$

Isso simplifica drasticamente os cálculos devido às propriedades da exponencial. Por exemplo, uma modulação (multiplicação de sinais) é facilmente representada por:

$$ae^{j\alpha} \cdot be^{j\beta} = ab \cdot e^{j(\alpha+\beta)}. \quad (2.53)$$

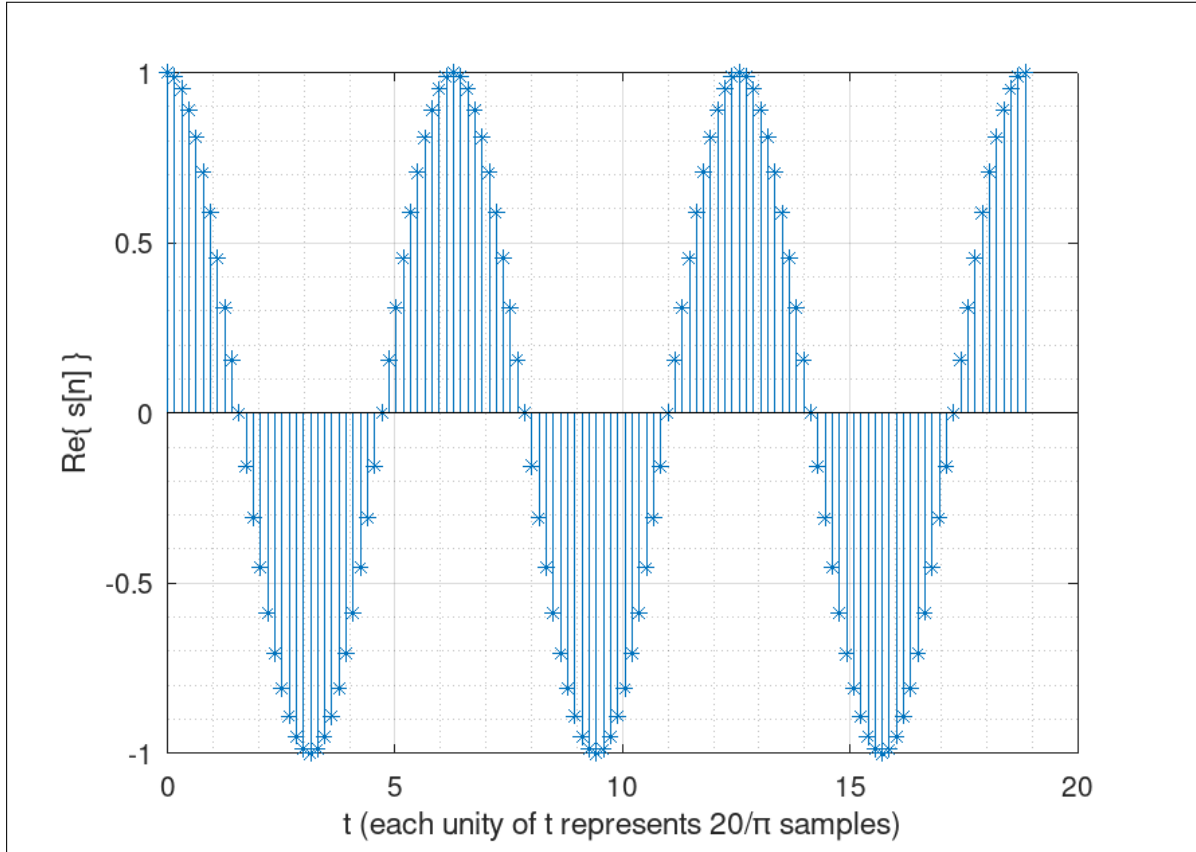
Além disso, a integração e diferenciação de exponenciais obedecem a resultados simples, como:

$$\frac{d(e^{j\alpha})}{d\alpha} = je^{j\alpha}. \quad (2.54)$$

Exponenciais oferecem uma abordagem mais conveniente ao utilizar a transformada de Fourier ou a transformada Z (ou a transformada de Laplace no caso contínuo). Essa é a principal razão pela qual os sinais são introduzidos como valores complexos em livros introdutórios de processamento de sinais e sistemas.

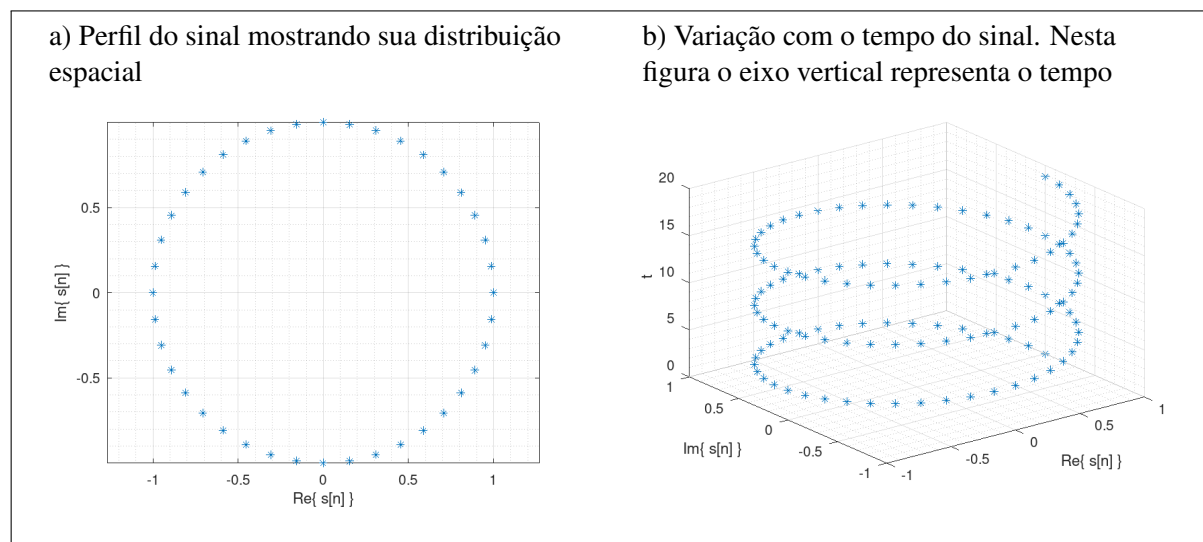
O próximo capítulo explorará a parte algébrica desta dissertação, proporcionando uma descrição abrangente de filtros amplamente lineares em conjunto com os sinais complexos discutidos anteriormente.

Figura 9 – Parte real de um sinal discreto do tipo $s[n] = \cos(\alpha n) + j \sin(\alpha n)$, ou seja $\text{Re}\{s[n]\} = \cos(\alpha n)$, para um período de amostragem de $T = \pi/20$, intervalo de tempo de $0 \leq t \leq 6\pi$ e $\alpha = 1$



Fonte: elaborada pelo autor.

Figura 10 – Um exemplo de sinal periódico discreto do tipo $s[n] = \cos n + j \sin n$ para 95 amostras



Fonte: elaborada pelo autor.

3 VETORES E TRANSFORMAÇÕES AMPLAMENTE LINEARES

Transformações lineares são de grande importância em diversas áreas do conhecimento. Elas permitem, em um certo intervalo, aproximar o comportamento de sistemas mais complicados. Sem elas as descrições e cálculos exigiriam estudo mais amplo, um maior poder computacional, enquanto modelos lineares permitem uma boa aproximação e simplicidade. Esse capítulo introduz de forma concisa essas transformações seguida de sua versão amplamente linear. Em sequência são apresentadas as relações entre produtos internos e formas quadráticas. Por último é introduzida a diferenciação de vetores e matrizes em variáveis complexas.

3.1 Transformações lineares

Os sistemas lineares desempenham um papel crucial na análise de sistemas, sendo introduzidos aqui através dos operadores lineares. Esses operadores são fundamentais para representar transformações entre espaços vetoriais, podendo ser expressos de forma matricial quando as bases dos espaços envolvidos são selecionadas. Uma transformação linear é um conceito matemático que mapeia vetores de um espaço $\mathcal{U}$ para outro espaço $\mathcal{V}$. Denotada como $A : \mathcal{U} \rightarrow \mathcal{V}$, uma transformação linear é definida de modo que A obedeça às regras da linearidade, que são:

$$A(\mathbf{u} + \mathbf{w}) = A(\mathbf{u}) + A(\mathbf{w}), \quad (3.1)$$

chamada de regra da adição, e

$$A(a\mathbf{u}) = aA(\mathbf{u}), \quad (3.2)$$

chamada de regra da multiplicação por um escalar. Essas duas regras devem ser válidas para todo $\mathbf{u}, \mathbf{w} \in \mathcal{U}$ e $a \in \mathcal{F}$, o corpo sobre o qual o espaço vetorial está definido. Fixados os vetores bases de $\mathcal{U}$ e $\mathcal{V}$, o operador A pode ser representado unicamente por meio de uma matriz, determinada pela atuação do operador sobre os vetores base do espaço considerado. Sejam $\{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset \mathcal{U}$ e $\{\mathbf{v}_1, \dots, \mathbf{v}_m\} \subset \mathcal{V}$ bases de $\mathcal{U}$ e $\mathcal{V}$ respectivamente, tem-se

$$A(\mathbf{u}_k) = \sum_{l=1}^m a_{kl} \mathbf{v}_l, \quad (3.3)$$

que permite determinar os elementos a_{kl} . Estes podem ser dispostos em uma matriz $m \times n$ e representa o operador A nas bases utilizadas,

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} .$$

Sejam os espaços vetoriais $\mathcal{U}_k$ em que os vetores $\mathbf{u}_k \in \mathcal{U}_k$, para $k \in 1, \dots, N$, são definidos sobre o corpo $\mathcal{F}$. Uma transformação A sobre produtos de espaços do tipo

$$A : \mathcal{U}_1 \times \cdots \times \mathcal{U}_N \rightarrow \mathcal{F} , \quad (3.4)$$

é chamada de forma multilinear se esta atua da forma $A(u_1, \dots, u_N) = a \in \mathcal{F}$ e satisfaz as condições das equações (3.1) e (3.2) em cada um dos operandos.

As transformações lineares, juntamente com a definição de soma $(A + B)(u) = A(u) + B(u)$ e de multiplicação por um escalar $(aA)(u) = a(A(u))$ formam um espaço vetorial. Os funcionais lineares, dado pelo conjunto de formas lineares (3.4)

$$L : \mathcal{U}_1 \times \cdots \times \mathcal{U}_N \rightarrow \mathcal{F} , \quad (3.5)$$

são um exemplo desse espaço, chamado de dual do espaço dado. A seguir um funcional importante é apresentado em forma de exemplo.

Exemplo de forma bilinear (matriz métrica): Dado o funcional linear $g : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, a matriz associada a este é chamada de matriz métrica. Sejam $\mathbf{u} = \sum_k u_k \mathbf{e}_k$ e $\mathbf{v} = \sum_l v_l \mathbf{e}_l$ pertencentes a $\mathbb{R}^n$ com $\mathbf{e}_k$ como base. Utilizando as propriedades da linearidade tem-se

$$g(\mathbf{u}, \mathbf{v}) = \sum_{k,l} u^k v^l g(\mathbf{e}_k, \mathbf{e}_l) = \sum_{k,l} u^k v^l g_{kl} .$$

Os coeficientes $g_{kl} = g(\mathbf{e}_k, \mathbf{e}_l)$ (como atuam nos elementos de base do funcional) definem a forma bilinear dada. No caso em que $g_{kl} = \delta_{kl}$, com

$$\delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases} , \quad (3.6)$$

a forma define o produto interno canônico em $\mathbb{R}^n$. Isso conclui o exemplo.

Funcionais lineares são importantes pois com eles se definem as formas diferenciais, co-vetores sobre uma dado espaço.

Operadores lineares são utilizados nesse trabalho tanto para relacionar espaços reais e complexos quanto para a linearização de espaços mais gerais por meio de diferenciação. O primeiro caso será visto a seguir, enquanto o segundo será conduzido junto do estudo de variedades diferenciáveis posteriormente.

3.2 Relação linear entre variáveis real e complexa

O Capítulo 2 mostra como números complexos escalares se relacionam com sua parte real e imaginária, com a definição do par $(z, z^*) \in \mathbb{C}_*^2$. Essas novas variáveis permitem o uso do cálculo $\mathbb{C}\mathbb{R}$ (a diferenciação sobre funções complexas, não necessariamente holomorfas) resultando nos operadores (2.19). Esse capítulo lida com esse tipo de relação entre as coordenadas do ponto de vista de transformações lineares, relacionando variáveis reais retangulares com variáveis complexas, a partir da relação matricial (Schreier; Scharf, 2010)

$$\begin{bmatrix} z \\ z^* \end{bmatrix} = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (3.7)$$

A matriz

$$\mathbf{T} = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} \quad (3.8)$$

representa a transformação $\mathbb{R}^2 \rightarrow \mathbb{C}_*^2$ dada que levam os vetores canônicos $\mathbf{e}_1 = [1 \ 0]^T$ e $\mathbf{e}_2 = [0 \ 1]^T$ nos vetores complexos $[1 \ 1]^T$ e $[j \ -j]^T$, respectivamente. A matriz $\mathbf{T}$ possui inversa dada por:

$$\mathbf{T}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -j & j \end{bmatrix}. \quad (3.9)$$

Essa inversa obedece a relação $\mathbf{T}^{-1} = \frac{1}{2}\mathbf{T}^H$. O sobrescrito H indica o hermitiano (transposto conjugado) da matriz ou vetor. Espaços na forma do par (z, z^*) são chamados *aumentados* e indicam a representação complexa de um espaço real (x, y) . A seção a seguir generaliza o esse conceito para o caso de dimensões maiores que dois.

3.3 Vetores e transformações amplamente lineares

Espaços complexificados sempre terão dimensão par por construção. As transformações amplamente lineares são uma extensão natural do caso anterior, relacionadas com vetores

aumentados do tipo

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \\ z_1^* \\ \vdots \\ z_n^* \end{bmatrix}. \quad (3.10)$$

Existem $(2n)!$ arranjos possíveis para a construção de um vetor aumentado. Uma abordagem comumente utilizada é representada por $[z_1, z_1^*, \dots, z_n, z_n^*]^T$, alternando entre cada elemento e seu conjugado. Embora seja possível adotar outras construções, este trabalho dará preferência à forma (3.10) acima, pois nela é viável agrupar as n primeiras entradas $[z_1, \dots, z_n]^T$ como $\mathbf{z}$ e as n entradas restantes $[z_1^*, \dots, z_n^*]^T$ como $\mathbf{z}^*$, simplificando a notação.

No caso complexo (não-aumentado) há um isomorfismo entre espaços vetoriais complexos e reais. Dados os vetores $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ e $\mathbf{w} = \mathbf{u} + j\mathbf{v}$ com $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, estes podem ser relacionados por uma transformação dita $\mathbb{C}$ -linear da forma

$$\mathbf{w} = \mathbf{A}\mathbf{z}, \quad (3.11)$$

em que $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ e $\mathbf{A} \in \mathbb{C}^{n \times n}$. Esta transformação é isomorfa à sua versão $\mathbb{R}$ -linear, por meio de vetores reais compostos

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \mathbf{A}_{\mathbb{R}} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix},$$

$\mathbf{A} = \mathbf{A}_r + j\mathbf{A}_i$ então

$$\mathbf{A}_{\mathbb{R}} = \begin{bmatrix} \mathbf{A}_r & -\mathbf{A}_i \\ \mathbf{A}_i & \mathbf{A}_r \end{bmatrix}_{2n \times 2n}.$$

A transformação acima recebe o nome de composta descomplexificada de $\mathbf{A}$ (Lima, 2008).

O par complexo aumentado $(\mathbf{z}, \mathbf{z}^*)$, representado por

$$\underline{\mathbf{z}} = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} = \begin{bmatrix} \mathbf{z}^T & \mathbf{z}^H \end{bmatrix}^T, \quad (3.12)$$

será denotado pelo sublinhado $\underline{\mathbf{z}} \in \mathbb{C}_*^{2N}$, como em (Schreier; Scharf, 2010). É possível relacionar um vetor $\mathbb{R}$ -linear (aumentado real) e seu equivalente complexo aumentado, de forma semelhante

a (3.8) com uso de vetores compostos

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix},$$

em que $\mathbf{I}$ indica a matriz identidade $n \times n$. A matriz composta acima é uma transformação linear e permite transitar entre a representação aumentada real e a aumentada complexa. Representa-se essa matriz por

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix}_{2n \times 2n}, \quad (3.13)$$

com $\mathbf{T}\mathbf{T}^H = \mathbf{T}^H\mathbf{T} = 2\mathbf{I}_{2n \times 2n}$. Dada uma transformação linear real composta $\mathbf{M}_{2n \times 2n}$

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix},$$

sua representação complexa aumentada é

$$\underline{\mathbf{w}} = \mathbf{T}\mathbf{M} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \frac{1}{2}\mathbf{T}\mathbf{M}\mathbf{T}^H \underline{\mathbf{z}} = \underline{\mathbf{H}} \underline{\mathbf{z}}.$$

A matriz $\underline{\mathbf{H}} = \frac{1}{2}\mathbf{T}\mathbf{M}\mathbf{T}^H$, de dimensão $2n \times 2n$, é chamada de matriz aumentada de $\mathbf{M}$. Note que

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}$$

com $\mathbf{M}_{kl}$ para $k, l = \{1, 2\}$, representa uma transformação linear qualquer, e não um isomorfismo entre uma transformação de espaços $\mathbb{C}$ -lineares e espaços $\mathbb{R}$ -lineares. Caso essa matriz represente tal isomorfismo tem-se

$$\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{A}^* \end{bmatrix},$$

levando à representação complexa usual (3.11). Caso seja uma representação de uma transformação qualquer

$$\underline{\mathbf{H}} = \frac{1}{2}\mathbf{T}\mathbf{M}\mathbf{T}^H = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^* & \mathbf{H}_1^* \end{bmatrix} \quad (3.14)$$

levando à representação *amplamente linear*

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{w}^* \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^* & \mathbf{H}_1^* \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix}, \quad (3.15)$$

em forma compacta expressa por

$$\underline{\mathbf{w}} = \underline{\mathbf{H}} \underline{\mathbf{z}} . \quad (3.16)$$

É interessante notar que a primeira linha da expressão (3.15) fornece a forma não prolixa da transformação amplamente linear

$$\mathbf{w} = \mathbf{H}_1 \mathbf{z} + \mathbf{H}_2 \mathbf{z}^* . \quad (3.17)$$

A segunda linha é o conjugado da primeira, devido à construção empregada. A relação entre a versão composta e aumentada da matriz é dada por

$$\mathbf{H}_1 = \frac{1}{2} [\mathbf{M}_{11} + \mathbf{M}_{22} + j(\mathbf{M}_{21} - \mathbf{M}_{12})]$$

e

$$\mathbf{H}_2 = \frac{1}{2} [\mathbf{M}_{11} - \mathbf{M}_{22} + j(\mathbf{M}_{21} + \mathbf{M}_{12})] ,$$

de acordo com (3.14), permitindo a relação entre as formas lineares usuais e as amplamente lineares.

3.4 Produtos internos de vetores aumentados

É possível relacionar produtos internos em sua forma real composta e sua forma complexa aumentada. Sejam os vetores compostos reais:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_i \end{bmatrix} \text{ e } \mathbf{y} = \begin{bmatrix} \mathbf{y}_r \\ \mathbf{y}_i \end{bmatrix} , \quad (3.18)$$

onde $\mathbf{x}_r, \mathbf{x}_i, \mathbf{y}_r, \mathbf{y}_i \in \mathbb{R}^n$. Esses vetores compostos são relacionados com os vetores complexos por $\mathbf{u} = \mathbf{x}_r + j\mathbf{x}_i$, $\mathbf{v} = \mathbf{y}_r + j\mathbf{y}_i \in \mathbb{C}^n$. Com isso a relação entre produtos internos canônicos real e complexo será dada por:

$$\mathbf{x}^T \mathbf{y} = \text{Re}\{\mathbf{u}^H \mathbf{v}\} .$$

Já entre os vetores aumentados e real é expressa como:

$$\frac{1}{2} \underline{\mathbf{u}}^H \underline{\mathbf{v}} = \frac{1}{2} \mathbf{x}^T \mathbf{T}^H \mathbf{T} \mathbf{y} = \mathbf{x}^T \mathbf{y} .$$

A relação entre vetores aumentados e real é dada por $\underline{\mathbf{u}} = \mathbf{T}\mathbf{x}$ e $\underline{\mathbf{v}} = \mathbf{T}\mathbf{y}$. A relação entre as expressões para o produto interno podem ser colocadas de forma geral por

$$\mathbf{x}^T \mathbf{y} = \text{Re}\{\mathbf{u}^H \mathbf{v}\} = \frac{1}{2} \underline{\mathbf{u}}^H \underline{\mathbf{v}} , \quad (3.19)$$

A expressão representa a equivalência entre os produtos internos em uma só expressão.

3.5 Formas bilineares

Com o auxílio do produto interno discutido na seção anterior, torna-se possível construir formas bilineares, as quais desempenham um papel crucial em estatísticas de segunda ordem, métricas de espaços mais gerais, entre outros contextos. Dados os vetores compostos $\mathbf{x}$ e $\mathbf{y}$ conforme definidos em (3.18), uma forma bilinear é representada por um funcional m da forma:

$$m(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{M} \mathbf{y} .$$

A matriz $\mathbf{M}$ é uma matriz simétrica dada e caracteriza a operação. É possível escrever a equação anterior em termos de vetores e matrizes complexificadas por meio da matriz (3.14), com isso:

$$\mathbf{x}^T \mathbf{M} \mathbf{y} = \mathbf{x}^T \frac{\mathbf{T}^H \mathbf{T}}{2} \mathbf{M} \frac{\mathbf{T}^H \mathbf{T}}{2} \mathbf{y} = \frac{1}{2} (\mathbf{T} \mathbf{x})^H \frac{\mathbf{T} \mathbf{M} \mathbf{T}^H}{2} \mathbf{T} \mathbf{y} ,$$

que resulta na relação

$$\mathbf{x}^T \mathbf{M} \mathbf{y} = \frac{1}{2} \underline{\mathbf{u}}^H \underline{\mathbf{H}} \underline{\mathbf{v}} . \quad (3.20)$$

Na forma acima as matrizes se relacionam por $\underline{\mathbf{H}} = (\mathbf{T} \mathbf{M} \mathbf{T}^H)/2$, semelhante ao resultado (3.14) pois o produto interno é também uma forma bilinear. As matrizes aumentadas usadas em formas bilineares são importantes em aproximações de segunda ordem de funções vetoriais ou funções custos de filtros lineares.

3.6 Derivadas de vetores e matrizes complexas

O Capítulo 2 mostra como formar derivadas a partir de valores escalares complexos. Nessa seção será apresentada uma forma de compactar a notação e escrever as derivadas em função de operadores vetoriais. Para uma função do tipo $f(\mathbf{z}) : \mathbb{C}^N \rightarrow \mathbb{C}$, com $\mathbf{z} = (z_1, z_2, \dots, z_N) \in \mathbb{C}^N$, as derivadas são definidas como vetores linha¹ como:

$$\frac{\partial f}{\partial \mathbf{z}} = \left[\frac{\partial f}{\partial z_1} \quad \dots \quad \frac{\partial f}{\partial z_N} \right] \quad \text{e} \quad \frac{\partial f}{\partial \mathbf{z}^*} = \left[\frac{\partial f}{\partial z_1^*} \quad \dots \quad \frac{\partial f}{\partial z_N^*} \right] \in \mathbb{C}^N . \quad (3.21)$$

Essas são as formas vetoriais das derivadas de Wirtinger, em que cada elemento é dado pela expressão (2.19), formando um co-vetor. Co-vetores serão descritos como vetores linha no

¹ As derivadas serão consideradas linha para corresponderem às componentes de uma forma diferencial.

presente trabalho². Pode-se definir os operadores como vetores colunas pela notação

$$\frac{\partial f}{\partial \mathbf{z}^T} = \begin{bmatrix} \frac{\partial f}{\partial z_1} \\ \vdots \\ \frac{\partial f}{\partial z_N} \end{bmatrix} \quad \text{e} \quad \frac{\partial f}{\partial \mathbf{z}^H} = \begin{bmatrix} \frac{\partial f}{\partial z_1^*} \\ \vdots \\ \frac{\partial f}{\partial z_N^*} \end{bmatrix} \in \mathbb{C}^N. \quad (3.22)$$

Estes não devem ser confundidos com o gradiente em notação complexa: representam a escrita em coluna dos operadores definidos em (3.21), de acordo com as relações

$$\left(\frac{\partial f}{\partial \mathbf{z}} \right)^T = \frac{\partial f}{\partial \mathbf{z}^T} \quad (3.23)$$

e

$$\left(\frac{\partial f}{\partial \mathbf{z}^*} \right)^T = \frac{\partial f}{\partial \mathbf{z}^H}, \quad (3.24)$$

para f real. Em relação às operações de conjugação, como hermitiano, é possível relacioná-las diretamente, ainda com f real, por

$$\left(\frac{\partial f}{\partial \mathbf{z}} \right)^H = \frac{\partial f}{\partial \mathbf{z}^H}, \quad f \in \mathbb{R}.$$

A derivada de funções vetoriais do tipo $\mathbf{f}(\mathbf{z}) : \mathbb{C}^N \rightarrow \mathbb{C}^M$ são definidas por

$$\frac{\partial \mathbf{f}}{\partial \mathbf{z}^*} = \left[\frac{\partial f_i}{\partial z_j^*} \right]_{M \times N},$$

para as i -ésimas linhas e j -ésimas colunas. É interessante notar que a operação de derivação sempre aumenta o posto em relação a função operada. Para matrizes como $\mathbf{A}(z) = [A_{ij}(w)] \in \mathbb{C}^{M \times N}$ a derivada é calculada elemento a elemento por

$$\frac{\partial \mathbf{A}}{\partial z^*} = \left[\frac{\partial A_{ij}}{\partial z^*} \right]_{M \times N}.$$

Caso a matriz seja a variável de derivação para uma função do tipo $f : \mathbb{C}^{M \times N} \rightarrow \mathbb{C}$ a derivada é definida como

$$\frac{\partial f}{\partial \mathbf{Z}^*} = \left[\frac{\partial f}{\partial Z_{ij}^*} \right]_{M \times N}, \quad \mathbf{Z} \in \mathbb{C}^{M \times N}.$$

Com essas definições é possível o cálculo de derivadas de vetores nas formas citadas nesse capítulo sem a necessidade de se trabalhar com as partes reais e imaginárias, tanto no caso de funções holomorfas como não holomorfas, diretamente em notação complexa. Alguns exemplos importantes de derivação são mostrados a seguir

² Um operador linha e pertence ao espaço dual. O gradiente é a versão vetorial do operador dado como será visto adiante.

3.6.1 Exemplos do uso dos operadores complexos

A fim de exemplificar e adiantar alguns dos cálculos presentes nesta dissertação, foram escolhidos os exemplos mais comuns e suas derivadas dadas pelas expressões (3.21) e (3.22) vistos na seção anterior. Tais expressões serão úteis nas equações de otimização presentes em filtros complexos que possuem função real a ser minimizada com coeficientes complexos. Os dois primeiros exemplos são produtos internos enquanto o último é uma forma quadrática geral.

3.6.1.1 Derivadas para funções do tipo $f(\mathbf{z}, \mathbf{z}^*) = \mathbf{a}^H \mathbf{z}$.

A função dada é complexa e representa o produto interno hermitiano, podendo ser escrita nas formas matricial e em função de seus elementos como

$$f(\mathbf{z}, \mathbf{z}^*) = \mathbf{a}^H \mathbf{z} = \begin{bmatrix} a_1^* & \dots & a_N^* \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} = \sum_{k=1}^N a_k^* z_k, \quad (3.25)$$

para um dado vetor complexo $\mathbf{a}$ constante. A partir das relações (3.21) para o cálculo da derivada não conjugada, a derivada para um elemento genérico $m \in \{1, \dots, N\}$ é dada por:

$$\frac{\partial f}{\partial z_m} = \frac{\partial}{\partial z_m} \left(\sum_{k=1}^N a_k^* z_k \right) = \sum_{k=1}^N a_k^* \frac{\partial z_k}{\partial z_m}.$$

Como $\partial z_k / \partial z_m = \delta_{km}$, a expressão anterior pode ser simplificada para

$$\frac{\partial f}{\partial z_m} = \sum_{k=1}^N a_k^* \delta_{km} = a_m^*. \quad (3.26)$$

Então com os elementos dados pela expressão (3.26) é possível construir o vetor linha de derivadas como

$$\frac{\partial f}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial f}{\partial z_1} & \dots & \frac{\partial f}{\partial z_N} \end{bmatrix} = \begin{bmatrix} a_1^* & \dots & a_N^* \end{bmatrix}. \quad (3.27)$$

A derivada da expressão em relação à variável vetorial não conjugada pode ser escrita de forma vetorial como segue:

$$\frac{\partial(\mathbf{a}^H \mathbf{z})}{\partial \mathbf{z}} = \mathbf{a}^H. \quad (3.28)$$

Essa expressão indica a derivada da função f em relação ao vetor $\mathbf{z}$, resultando no vetor transposto conjugado de $\mathbf{a}$.

Derivando mesma expressão (3.25) agora em relação a variável conjugada:

$$\frac{\partial f}{\partial z_m^*} = \frac{\partial}{\partial z_m^*} \left(\sum_{k=1}^N a_k^* z_k \right) = \sum_{k=1}^N a_k^* \frac{\partial z_k}{\partial z_m^*} = 0 .$$

Dessa vez a derivada da expressão em relação à variável vetorial conjugada é escrita de forma vetorial por:

$$\frac{\partial(\mathbf{a}^H \mathbf{z})}{\partial \mathbf{z}^*} = 0 . \quad (3.29)$$

Utilizando as relações (3.23) e (3.24), tem-se as derivadas em suas versões colunas

$$\frac{\partial(\mathbf{a}^H \mathbf{z})}{\partial \mathbf{z}^T} = \frac{\partial(\mathbf{z}^T \mathbf{a}^*)}{\partial \mathbf{z}^T} = \mathbf{a}^* \quad (3.30)$$

e

$$\frac{\partial(\mathbf{a}^H \mathbf{z})}{\partial \mathbf{z}^H} = 0 . \quad (3.31)$$

Com isso foi descrito as possíveis derivadas vetoriais complexas de primeira ordem para uma função do tipo $f = \mathbf{a}^H \mathbf{z}$.

3.6.1.2 Derivadas para funções do tipo $g(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{a}$.

Reescrevendo novamente a função em termos de seus elementos

$$g(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{a} = \begin{bmatrix} z_1^* & \dots & z_N^* \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} = \sum_{k=1}^N z_k^* a_k . \quad (3.32)$$

Com isso é possível proceder de forma semelhante aos cálculos anteriores. A derivada do m -ésimo elemento não conjugado individual é:

$$\frac{\partial g}{\partial z_m} = \frac{\partial}{\partial z_m} \left(\sum_{k=1}^N z_k^* a_k \right) = \sum_{k=1}^N a_k \frac{\partial z_k^*}{\partial z_m} = 0 .$$

Na forma vetorial

$$\frac{\partial(\mathbf{z}^H \mathbf{a})}{\partial \mathbf{z}} = 0 . \quad (3.33)$$

A derivada em relação ao conjugado é dada por:

$$\frac{\partial g}{\partial z_m^*} = \frac{\partial}{\partial z_m^*} \left(\sum_{k=1}^N z_k^* a_k \right) = \sum_{k=1}^N a_k \frac{\partial z_k^*}{\partial z_m^*} = a_m .$$

Já na forma vetorial expressa por:

$$\frac{\partial g}{\partial \mathbf{z}^*} = \begin{bmatrix} \frac{\partial g}{\partial z_1^*} & \dots & \frac{\partial g}{\partial z_N^*} \end{bmatrix} = \begin{bmatrix} a_1 & \dots & a_N \end{bmatrix}$$

De forma compacta a derivada em relação ao conjugado vetorial é

$$\frac{\partial(\mathbf{z}^H \mathbf{a})}{\partial \mathbf{z}^*} = \mathbf{a}^T. \quad (3.34)$$

Como regra geral a derivada eliminaria o termo conjugado, mas o operador deve ter o formato linha. É preciso deixar a função com o termo $\mathbf{z}^*$ em forma de coluna, longe da operação de transposição. Isso pode ser feito reescrevendo $\mathbf{z}^H \mathbf{a} = \mathbf{a}^T \mathbf{z}^*$ e então

$$\frac{\partial(\mathbf{z}^H \mathbf{a})}{\partial \mathbf{z}^*} = \frac{\partial(\mathbf{a}^T \mathbf{z}^*)}{\partial \mathbf{z}^*} = \mathbf{a}^T.$$

Novamente através das expressões (3.23) e (3.24), as derivadas em versão coluna são

$$\frac{\partial(\mathbf{z}^H \mathbf{a})}{\partial \mathbf{z}^T} = 0$$

e a versão conjugada

$$\frac{\partial(\mathbf{z}^H \mathbf{a})}{\partial \mathbf{z}^H} = \mathbf{a}. \quad (3.35)$$

Observe que, como a derivada em relação a variável conjugada (3.35) não se anula, a função $g = \mathbf{z}^H \mathbf{a}$ não é holomorfa.

É interessante observar que tanto $\mathbf{a}^H \mathbf{z}$ e $\mathbf{z}^H \mathbf{a}$ são funções que levam de valores complexos a valores complexos: a primeira é uma função holomorfa de acordo com (3.31). Individualmente ambas expressões não podem representar uma função real mas a soma $\mathbf{a}^H \mathbf{z} + \mathbf{z}^H \mathbf{a} = 2\text{Re}\{\mathbf{a}^H \mathbf{z}\}$ em geral é utilizada em funções de valores complexos no corpo dos reais.

3.6.1.3 Derivadas para a forma quadrática $h(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{R} \mathbf{z}$.

Esse exemplo é importante pois apresenta a derivação de uma expressão quadrática. Dada uma matriz $\mathbf{R} = [r_{kl}]_{N \times N}$ com $r_{kl} \in \mathbb{C}$, colocando na forma vetorial e em notação de índice, a função pode ser escrita como:

$$h = \mathbf{z}^H \mathbf{R} \mathbf{z} = \begin{bmatrix} z_1^* & \dots & z_N^* \end{bmatrix} \begin{bmatrix} r_{11} & \dots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{N1} & \dots & r_{NN} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} = \sum_{k,l=1}^N z_k^* r_{kl} z_l. \quad (3.36)$$

Nesse caso, o m -ésimo elemento da derivada é

$$\frac{\partial h}{\partial z_m} = \frac{\partial}{\partial z_m} \left(\sum_{k,l=1}^N z_k^* r_{kl} z_l \right) = \sum_{k,l=1}^N z_k^* r_{kl} \delta_{lm} = \sum_{k=1}^N z_k^* r_{km} .$$

Com cada elemento individual é possível construir a forma matricial da expressão por:

$$\frac{\partial h}{\partial \mathbf{z}_m} = \left[\sum_{k=1}^N z_k^* r_{k1} \quad \dots \quad \sum_{k=1}^N z_k^* r_{kN} \right] = \begin{bmatrix} z_1^* & \dots & z_N^* \end{bmatrix} \begin{bmatrix} r_{11} & \dots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{N1} & \dots & r_{NN} \end{bmatrix} , \quad (3.37)$$

que equivale a forma vetorial compacta

$$\frac{\partial(\mathbf{z}^H \mathbf{R} \mathbf{z})}{\partial \mathbf{z}} = \mathbf{z}^H \mathbf{R} . \quad (3.38)$$

Para o cálculo da derivada conjugada o procedimento é similar. A derivada do m -ésimo elemento é

$$\frac{\partial h}{\partial z_m^*} = \frac{\partial}{\partial z_m^*} \left(\sum_{k,l=1}^N z_k^* r_{kl} z_l \right) = \sum_{k,l=1}^N \delta_{mk} r_{kl} z_l = \sum_{l=1}^N r_{ml} z_l ,$$

que resulta no vetor linha

$$\frac{\partial h}{\partial \mathbf{z}_m^*} = \left[\sum_{l=1}^N r_{1l} z_l \quad \dots \quad \sum_{l=1}^N r_{Nl} z_l \right] . \quad (3.39)$$

Note que esse vetor linha é o mesmo resultante da operação

$$\left[\sum_{l=1}^N r_{1l} z_l \quad \dots \quad \sum_{l=1}^N r_{Nl} z_l \right] = \left(\begin{bmatrix} r_{11} & \dots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{N1} & \dots & r_{NN} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} \right)^T = \mathbf{z}^T \mathbf{R}^T .$$

Com isso a expressão para a derivada é

$$\frac{\partial \mathbf{z}^H \mathbf{R} \mathbf{z}}{\partial \mathbf{z}^*} = \mathbf{z}^T \mathbf{R}^T . \quad (3.40)$$

Este resultado pode ser obtido seguindo uma regra geral na qual primeiro isolamos o termo $\mathbf{z}^*$ expressando-o como $h = \mathbf{z}^H \mathbf{R} \mathbf{z} = \mathbf{z}^T \mathbf{R}^T \mathbf{z}^*$, e em seguida, tomamos a derivada, eliminando o termo conjugado (variável presente no denominador). Em termos vetoriais, isso é representado por:

$$\frac{\partial(\mathbf{z}^T \mathbf{R}^T \mathbf{z}^*)}{\partial \mathbf{z}^*} = \mathbf{z}^T \mathbf{R}^T .$$

Utilizando as equações (3.23) e (3.24), as derivadas no formato coluna da forma quadrática são dadas por:

$$\frac{\partial(\mathbf{z}^H \mathbf{R} \mathbf{z})}{\partial \mathbf{z}^T} = \frac{\partial(\mathbf{z}^T \mathbf{R}^T \mathbf{z}^*)}{\partial \mathbf{z}^T} = \mathbf{R}^T \mathbf{z}^*. \quad (3.41)$$

e sua versão conjugada é dada por:

$$\frac{\partial(\mathbf{z}^H \mathbf{R} \mathbf{z})}{\partial \mathbf{z}^H} = \mathbf{R} \mathbf{z}. \quad (3.42)$$

Da mesma forma que os produtos internos, esta forma quadrática é uma função de valores complexos em complexos. No entanto, caso a matriz $\mathbf{R}$ utilizada seja simétrica, os valores resultantes da forma quadrática serão reais.

3.6.1.4 Resumo geral das regras de derivação vetorial

A partir dos exemplos anteriores é possível agrupar os resultados de acordo com a Tabela 1, apresentada a seguir. Uma regra de derivação notável é a seguinte: ao calcular a

Tabela 1 – Derivadas para funções de primeira ordem e quadrática. As linhas indicam a função e as colunas o tipo de derivada calculada

1	2	3	4	5
	$\frac{\partial}{\partial \mathbf{z}}$	$\frac{\partial}{\partial \mathbf{z}^*}$	$\frac{\partial}{\partial \mathbf{z}^T}$	$\frac{\partial}{\partial \mathbf{z}^H}$
$\mathbf{a}^H \mathbf{z}$	$\mathbf{a}^H$	0	$\mathbf{a}^*$	0
$\mathbf{z}^H \mathbf{a}$	0	$\mathbf{a}^T$	0	$\mathbf{a}$
$\mathbf{z}^H \mathbf{R} \mathbf{z}$	$\mathbf{z}^H \mathbf{R}$	$\mathbf{z}^T \mathbf{R}^T$	$\mathbf{R}^T \mathbf{z}^*$	$\mathbf{R} \mathbf{z}$

Fonte: elaborado pelo autor.

derivada, elimina-se da função a variável do mesmo tipo que está presente no denominador, sendo necessário reescrever a função de modo a evidenciar corretamente os tipos envolvidos. É importante observar que, devido às expressões (3.23) e (3.24), as colunas 2 e 3 correspondem às versões transpostas das colunas 4 e 5, respectivamente. Essa regra se aplica igualmente no caso de derivadas vetoriais, como por exemplo, para $\mathbf{f} = [f_1 \cdots f_N] = \mathbf{z}^H \mathbf{R}$, um vetor linha. A derivada é dada por:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{z}^H} = \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{z}^H} & \cdots & \frac{\partial f_N}{\partial \mathbf{z}^H} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial z_1^*} & \cdots & \frac{\partial f_N}{\partial z_1^*} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial z_N^*} & \cdots & \frac{\partial f_N}{\partial z_N^*} \end{bmatrix}, \quad (3.43)$$

onde o m -ésimo elemento da função é $f_m = \sum_i z_i^* r_{im}$. A derivada de cada termo é dada por:

$$\frac{\partial f_m}{\partial z_k^*} = \frac{\partial}{\partial z_k^*} \left(\sum_i z_i^* r_{im} \right) = r_{km} . \quad (3.44)$$

Substituindo a expressão (3.43) em (3.44), é possível isolar a matriz $\mathbf{R}$, ao eliminar a variável $\mathbf{z}^H$, como expresso em:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{z}^H} = \begin{bmatrix} r_{11} & \cdots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{N1} & \cdots & r_{NN} \end{bmatrix} = \mathbf{R} .$$

Essas derivadas serão fundamentais na otimização de funções quadráticas comuns em filtros adaptativos, como veremos posteriormente.

3.7 Nomenclatura de Vetores e Matrizes

Para garantir clareza na descrição dos vetores e matrizes utilizados neste trabalho, é útil abrir uma breve seção sobre a nomenclatura adotada. As diferentes formas de organização são resumidas nos seguintes casos:

- **Real:** Representado por $\mathbf{x} \in \mathbb{R}^N$, que é o vetor real padrão.
- **Composto (real):** Denotado pelo vetor

$$\mathbf{c} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathbb{R}^{2N} ,$$

onde $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ são vetores reais.

- **Complexo:** Representado por $\mathbf{z} \in \mathbb{C}^N$.
- **Aumentado:** Utilizado para vetores como

$$\underline{\mathbf{z}} = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} ,$$

e para matrizes como

$$\underline{\mathbf{M}} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{N}^* & \mathbf{M}^* \end{bmatrix} .$$

Normalmente, vetores e matrizes aumentadas são representados com um sublinhado, a menos que seja explicitamente declarado, como no Capítulo 7.

– **Composto (complexo):** Representado por

$$\mathbf{z}_c = \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} \in \mathbb{C}^{2N},$$

onde $\mathbf{z}, \mathbf{w} \in \mathbb{C}^N$ são vetores complexos. Similarmente ao caso anterior, o composto também pode ser colocado na forma aumentada, como $\underline{\mathbf{z}}_c$. Isso é conhecido como **composto aumentado**, que essencialmente duplica a dimensão do caso composto.

Assim, encerramos este capítulo, cujo objetivo foi estabelecer o fundamento algébrico essencial para compreender sinais complexos. No próximo capítulo, exploraremos conceitos fundamentais de estatística complexa, os quais se apoiarão em alguns dos conceitos apresentados aqui.

4 ESTATÍSTICA COMPLEXA

O estudo estatístico desempenha um papel crucial na análise de sinais. Na realidade, todos os sinais do mundo real enfrentam a presença inevitável de ruídos, interferências e atrasos, todos eles provenientes de fontes diversas. Este capítulo explora ferramentas estatísticas com o objetivo de analisar características de não-circularidade, como será detalhado posteriormente.

4.1 Variáveis Aleatórias Complexas

Variáveis aleatórias reais fornecem a base para a compreensão e definição de variáveis aleatórias complexas. Assim, uma introdução a essas variáveis é realizada com o intuito de estabelecer notações e definições gerais sobre o assunto. Primeiramente, o conceito de variáveis reais escalares é definido. Em seguida, essas variáveis são abordadas como variáveis bidimensionais conjuntas, uma vez que as variáveis complexas são definidas da mesma forma.

4.1.1 Variáveis aleatórias unidimensionais reais

Uma variável aleatória real é uma função que mapeia resultados possíveis de um conjunto de eventos em valores reais. Mais precisamente, é uma função do tipo:

$$X : \Omega \rightarrow \mathbb{R} \quad (4.1)$$

onde Ω é o conjunto dos eventos possíveis, e X a função arbitrária que mapeia cada $\omega \in \Omega$. Como exemplo trivial, são considerados os possíveis resultados ao lançar uma moeda, representados pelo conjunto $\Omega = \{\omega_1 = \text{cara}, \omega_2 = \text{coroa}\}$. Pode ser atribuída uma função (de escolha arbitrária) que mapeia esse conjunto de possibilidades para o conjunto dos números reais, exemplificado abaixo:

$$X(\omega) = \begin{cases} 0, & \omega = \text{cara} \\ 1, & \omega = \text{coroa} . \end{cases} \quad (4.2)$$

É importante observar que este mapeamento poderia ter sido escolhido de forma diferente; o essencial é que tal escolha seja biunívoca. De maneira geral, o conjunto Ω é frequentemente omitido, e trabalha-se com a variável aleatória X . A cada evento (valor da variável) está associada uma probabilidade, e consequentemente, a cada variável aleatória também.

4.1.2 Variáveis bidimensionais e Complexas

Para iniciar o estudo de sinais complexos, é necessário estabelecer a definição de uma variável aleatória complexa, representada por $Z = X + jY : \Omega \rightarrow \mathbb{C}$, onde Ω é um espaço de eventos complexos e tanto X quanto $Y : \Omega \rightarrow \mathbb{R}$ são variáveis aleatórias reais definidas anteriormente. Devido à impossibilidade de ordenação dos números complexos, as probabilidades e densidades são calculadas como variáveis conjuntas.

Considerando X e Y como variáveis contínuas unidimensionais, a função de distribuição cumulativa é definida por

$$F_Z(z) \triangleq F_{X,Y}(x,y) = \Pr(X \leq x, Y \leq y)$$

enquanto a função de densidade de probabilidade é expressa como

$$f_Z(z) \triangleq f_{X,Y}(x,y) = \frac{\partial^2 F_Z(z)}{\partial x \partial y} .$$

A partir destas definições, podemos calcular as médias de Z , dadas por

$$E\{Z\} = E\{X\} + jE\{Y\} .$$

No caso discreto, em que X e Y são variáveis aleatórias discretas, a média de Z é calculada como

$$E\{Z\} = \sum \sum z p_{X,Y}(x,y)$$

onde $p_{X,Y}(x,y) = \Pr(X = x, Y = y)$ representa a distribuição de massa conjunta.

Além disso, é importante mencionar a possibilidade do caso misto, no qual uma das variáveis é discreta e a outra é contínua. No entanto, este caso não será abordado nesta dissertação.

4.2 Circularidade e Propriedade

Dois conceitos fundamentais na análise de sinais complexos são descritos de forma geral a seguir. Esses conceitos servirão como base para a construção de definições relacionadas à modelagem de sinais complexos circulares e não circulares.

- **Circularidade** (associada ao conceito de não-circularidade): Uma variável complexa $Z = X + jY$ é considerada circular se a distribuição de probabilidade de Z for a mesma

que a distribuição de sua versão rotacionada $e^{j\phi}Z$ para todo $\phi \in [-\pi, \pi)$. Esta definição requer que todos os momentos associados à distribuição sejam conhecidos e iguais nas partes reais e imaginárias para cada ordem de momento, além de exigir que os momentos cruzados sejam nulos. A não-circularidade é o oposto dessa definição.

- **Propriedade** (associada ao conceito de impropriedade): Considera apenas estatísticas de segunda ordem e permite uma caracterização simples de distribuições gaussianas bivariadas (sejam escalares ou vetoriais). A propriedade é essencialmente a mesma definição de circularidade, porém levando em conta apenas os momentos de segunda ordem.

Para o caso de distribuições gaussianas (que possam ser deslocadas para a origem, ou que tenham média nula), circularidade e propriedade são conceitos equivalentes.

4.2.1 Circularidade de distribuições gaussianas

Nesta seção do trabalho, abordamos as definições de circularidade e distribuição gaussiana, começando com o caso bivariado e, em seguida, avançando para o caso vetorial. Essa abordagem visa consolidar as definições anteriores de propriedade/impropriedade, que serão utilizadas ao longo do trabalho.

4.2.1.1 Caso bidimensional

Para caracterizar distribuições gaussianas os momentos de segunda ordem são requeridos (conhecendo-se a média). Essa seção mostra como se dá essa caracterização para o caso bi-dimensional, com uso de variáveis aleatórias complexas, relacionando também com a descrição real.

Uma distribuição gaussiana (normal) com média nula e matriz de covariância dada pelas seguintes esperanças:

$$\mathbf{\Sigma}_{XY} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \quad (4.3)$$

é representada pela notação $\mathcal{N}(0, \mathbf{\Sigma}_{XY})$, onde $\sigma_X^2 = E\{X^2\}$, $\sigma_Y^2 = E\{Y^2\}$ e $\sigma_{XY} = E\{XY\}$ são os momentos de segunda ordem. O coeficiente de correlação é definido como:

$$\rho_r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2} \sqrt{\sigma_Y^2}} \quad (4.4)$$

onde $|\rho_r| \leq 1$, representando a relação entre as variâncias e a covariância. A função de densidade que descreve a distribuição gaussiana conjunta com média zero e caracterizada pela matriz de

covariância real (4.3) é dada por:

$$f_{XY}(x,y) = \frac{1}{\pi(\det[\mathbf{\Sigma}_{xy}])^{1/2}} \exp \left\{ -\frac{1}{2} [x \ y] \mathbf{\Sigma}_{xy}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \right\} . \quad (4.5)$$

Uma distribuição gaussiana é dita própria se $\sigma_{XY} = 0$ e $\sigma_X^2 = \sigma_Y^2$.

É possível representar os momentos de segunda ordem também com uso de variáveis complexas. Seja a variável aleatória $Z = X + jY$, com X e Y variáveis gaussianas unidimensionais, a variância complexa é

$$r_{ZZ^*} = E\{ZZ^*\} = E\{|Z|^2\} = \sigma_X^2 + \sigma_Y^2 .$$

Apenas com o valor de r_{ZZ^*} não é possível determinar os valores individuais para σ_X^2 ou σ_Y^2 separadamente. Nem mesmo o valor de σ_{XY} é mencionado nessa expressão. A introdução de mais uma equação permite determinar todos os momentos reais. Esta é denominada pseudo-covariância, dada por:

$$r_{ZZ} = E\{Z^2\} = \sigma_X^2 - \sigma_Y^2 + j2\sigma_{XY} .$$

Com ambas as equações em mãos, é possível determinar todos os valores reais que caracterizam a distribuição gaussiana bidimensional. É importante notar que a pseudo-covariância pode ser um número complexo, enquanto a covariância é sempre real. Podemos definir o fator de circularidade complexo, $\rho_c = r_{ZZ}/r_{ZZ^*}$ como em (Li; Adali, 2012), o qual está relacionado com o coeficiente de correlação (4.4) através de (Ollila, 2008). A relação entre esses dois coeficientes é expressa por:

$$\rho_r = \frac{\text{Im}\{\rho_c\}}{\sqrt{1 - \text{Re}\{\rho_c\}}} .$$

A partir das definições anteriores, tomando o fato que a matriz de covariância real (4.3) pode ser construída da forma composta por

$$\mathbf{\Sigma}_{XY} = E \left\{ \begin{bmatrix} X \\ Y \end{bmatrix} \cdot \begin{bmatrix} X & Y \end{bmatrix} \right\} ,$$

é possível construir a matriz de correlação aumentada

$$\underline{\mathbf{\Sigma}} = E \left\{ \begin{bmatrix} Z \\ Z^* \end{bmatrix} \cdot \begin{bmatrix} Z & Z^* \end{bmatrix} \right\} = \begin{bmatrix} r_{ZZ^*} & r_{ZZ} \\ (r_{ZZ})^* & (r_{ZZ^*})^* \end{bmatrix} . \quad (4.6)$$

A distribuição (4.5) em termos de variáveis aumentada é escrita como

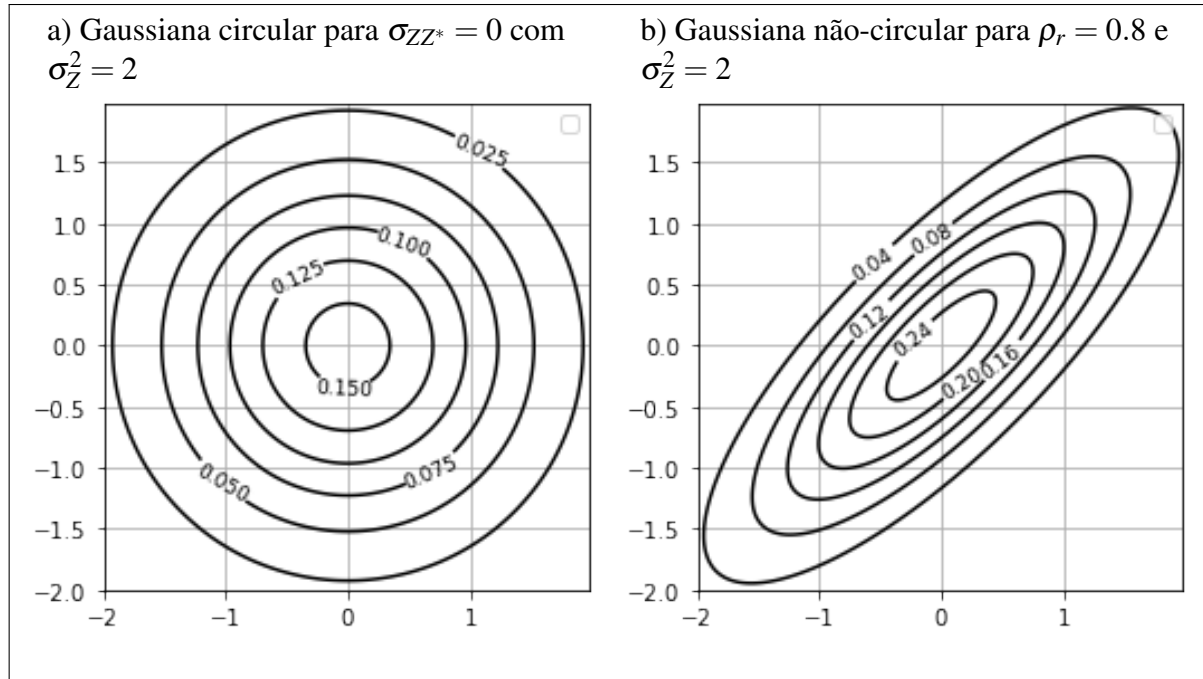
$$f_{ZZ^*}(z, z^*) = \frac{1}{\pi(\det[\underline{\Sigma}])^{1/2}} \exp \left\{ -\frac{1}{2} [z^* \ z] \underline{\Sigma}^{-1} \begin{bmatrix} z \\ z^* \end{bmatrix} \right\}, \quad (4.7)$$

evidenciando o uso de coordenadas complexas para representação da gaussiana real. O caso circular, em que $r_{ZZ} = 0$, implica diretamente nas condições $\sigma_{XY} = 0$ e $\sigma_X^2 = \sigma_Y^2$ e a densidade é escrita como (Amblard; Gaeta; Lacoume, 1996):

$$f_{ZZ^*}(z, z^*) = \frac{1}{\pi\sigma_z^2} \exp \left\{ -\frac{1}{\sigma_z^2} |z|^2 \right\}.$$

A notação aumentada da distribuição permite o uso da álgebra e diferenciação complexas, utilizando os operadores como (3.21). No entanto, ao contrário, a integração não pode ser diretamente realizada em (z, z^*) por falta de definição de caminhos em $\mathbb{C}_*^2$. As curvas de nível para as distribuições (4.5) ou (4.7) podem ser vistas na Figura 11, em (a) qualquer rotação passa

Figura 11 – Curvas de nível para distribuição gaussiana bivariada de média zero



Fonte: elaborada pelo autor.

desapercebida enquanto em (b) observa-se que qualquer rotação em torno da origem leva à uma distribuição diferente.

4.2.1.2 Caso vetorial

Para um dado vetor aleatório $\mathbf{Z} = \mathbf{X} + j\mathbf{Y} \in \mathbb{C}^N$ com média nula, onde cada componente é uma variável aleatória real, podemos escrever de forma compacta a matriz de correlação

aumentada utilizando o vetor aumentado $\underline{\mathbf{Z}} = (\mathbf{Z} \ \mathbf{Z}^*)$ (3.12). De forma similar à (4.6), temos:

$$\underline{\mathbf{R}} = E\{\underline{\mathbf{Z}}\underline{\mathbf{Z}}^H\} = \begin{bmatrix} E\{\mathbf{Z}\mathbf{Z}^H\} & E\{\mathbf{Z}\mathbf{Z}^T\} \\ E\{\mathbf{Z}\mathbf{Z}^T\}^* & E\{\mathbf{Z}\mathbf{Z}^H\}^* \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}}^* & \mathbf{R}^* \end{bmatrix}, \quad (4.8)$$

onde $\underline{\mathbf{R}}$ é a matriz de correlação aumentada, com dimensão $2N \times 2N$. As componentes de (4.8), de dimensão $N \times N$ cada, escritas em termos das componentes reais do vetor complexo, fornecem a covariância:

$$\mathbf{R} = E\{\mathbf{X}\mathbf{X}^T\} + E\{\mathbf{Y}\mathbf{Y}^T\} + j(E\{\mathbf{Y}\mathbf{X}^T\} - E\{\mathbf{X}\mathbf{Y}^T\})$$

$$\mathbf{R} = \mathbf{R}_{XX} + \mathbf{R}_{YY} + j(\mathbf{R}_{YX} - \mathbf{R}_{XY}),$$

e a pseudo-covariância

$$\tilde{\mathbf{R}} = \mathbf{R}_{XX} - \mathbf{R}_{YY} + j(\mathbf{R}_{XY} + \mathbf{R}_{YX}).$$

Para o caso em que o vetor é próprio, tem-se $\tilde{\mathbf{R}} = 0$, por conseguinte,

$$\mathbf{R}_{XX} = \mathbf{R}_{YY} \quad \text{e}$$

$$\mathbf{R}_{XY} = -\mathbf{R}_{YX}.$$

Com isso a caracterização de segunda ordem é coberta, com a utilização de variáveis aumentadas, que equivalem às suas versões reais por meio das expressões anteriores.

4.3 Processos Aleatórios Complexos

É possível caracterizar processos gaussianos discretos que variam no tempo por meio da autocorrelação. Esta é uma medida de similaridade do sinal com ele mesmo que pode ser feita por uma convolução discreta com seu conjugado. A pseudo-autocorrelação faz o mesmo mas sem a operação de conjugação. Dado um valor escalar $Z[n]$ cada elemento da autocorrelação é dado por

$$r[k, m] = E\{Z[k+m]Z^*[k]\} - E\{Z[k+m]\}E\{Z^*[k]\}$$

e a pseudo-autocorrelação por

$$\tilde{r}[k, m] = E\{Z[k+m]Z[k]\} - E\{Z[k+m]\}E\{Z[k]\}.$$

Se $\tilde{r}[k, m] = 0$ para todos os pares $[k, m]$, o processo é considerado próprio. Quando a média é zero, a autocorrelação coincide com a covariância do processo. Para um processo WSS (*wide sense stationary*, estacionário em sentido amplo) de média nula, podemos escrever $r[k, m] = r[m]$ e $\tilde{r}[k, m] = \tilde{r}[m]$, tornando o processo dependente apenas do intervalo de tempo entre as amostras. No caso de um processo modelado por um ruído gaussiano WSS, temos a propriedade de que $r[m] = E\{Z[m]Z^*[m]\} = r_{ZZ^*}$ e $\tilde{r}[m] = EZ[m]Z[m] = r_{ZZ}$ para todo m . A Figura 12 a seguir ilustra o comportamento para dois processos gaussianos de média nula. É possível observar que a pseudo-autocorrelação se torna irrelevante à medida que o sinal exibe uma maior simetria circular. A Figura 12 (a) exibe o perfil associado a um processo gaussiano circular, mostrando mesma variância para as partes real e imaginária. Em (b) tem-se o histograma do processo, seguida por (c) a autocorrelação e a pseudo-autocorrelação em (d). A Figura 12 (c) exibe um processo não circular, em que tanto a parte real, quanto a imaginária são gaussianas mas com variâncias diferentes, além de uma rotação. Da mesma forma, é possível analisar os parâmetros com destaque a pseudo-autocorrelação, que agora possui comportamento similar àquele exibido pela correlação, destacando a circularidade do sinal.

4.4 Modelagem de sinais complexos

Um sinal complexo pode ser modelado a partir de um processo aleatório complexo. Os sinais discretos presentes nesta dissertação serão modelados de acordo com a equação

$$Z[n] = e^{j\varphi} \left(\sqrt{1 - \rho^2} X[n] + j\rho Y[n] \right). \quad (4.9)$$

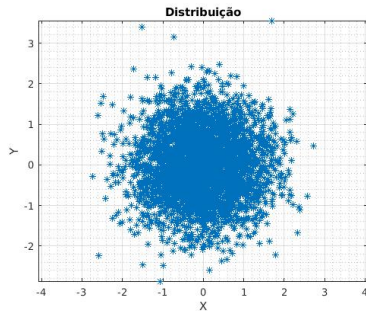
Nesse modelo as funções $X[n]$ e $Y[n]$ serão processos WSS escalares, cada um com de média zero e variância σ_X^2 e σ_Y^2 , em geral descorrelacionados. O ruído será modelado de forma semelhante, com fator de escala logarítmico ou linear. O valor de $0 < \rho < 1$, é a razão de perfil, relaciona o quanto proporcionalmente cada componente é levada em conta. A variável φ é o quanto esse perfil é rotacionado.

Um caso que será recorrente no presente trabalho é o de um sinal gaussiano WSS (GWSS) dado pelo modelo (4.9), para o caso em que $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (em geral, unitárias) variâncias de processos descorrelacionados. A relação entre a razão do perfil ρ e o fator de circularidade complexo ρ_c é dado por

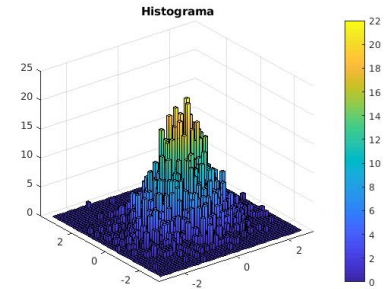
$$\rho_c = \frac{r_{ZZ}}{r_{ZZ^*}} = (1 - 2\rho^2)e^{j2\varphi},$$

Figura 12 – Curvas de nível para distribuição gaussiana bivariada de média zero

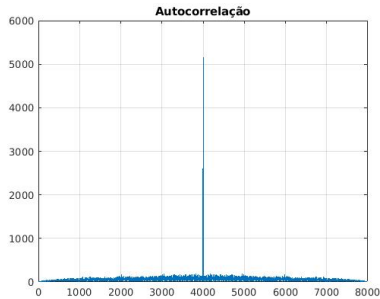
(a) Distribuição gaussiana circular



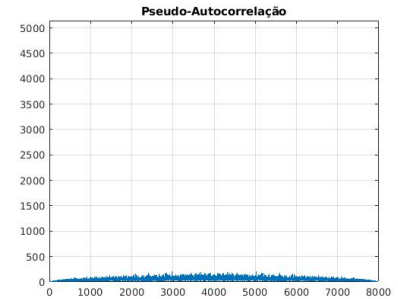
(b) Histograma do processo dado em (a)



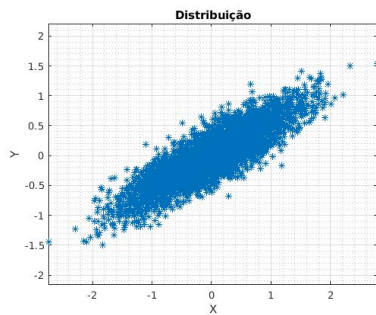
(c) Autocorrelação do processo dado em (a)



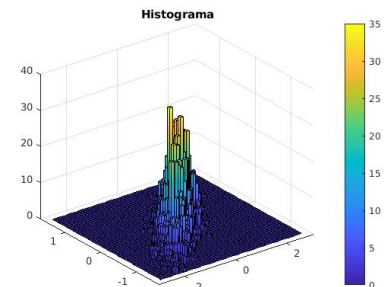
(d) Pseudo-autocorrelação do processo dado (a)



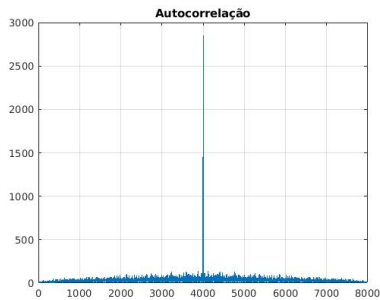
(e) Distribuição gaussiana não circular



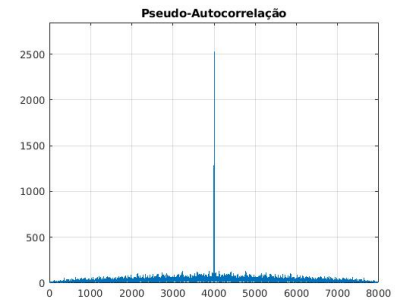
(f) Histograma do processo dado em (e)



(g) Autocorrelação do processo dado (e)



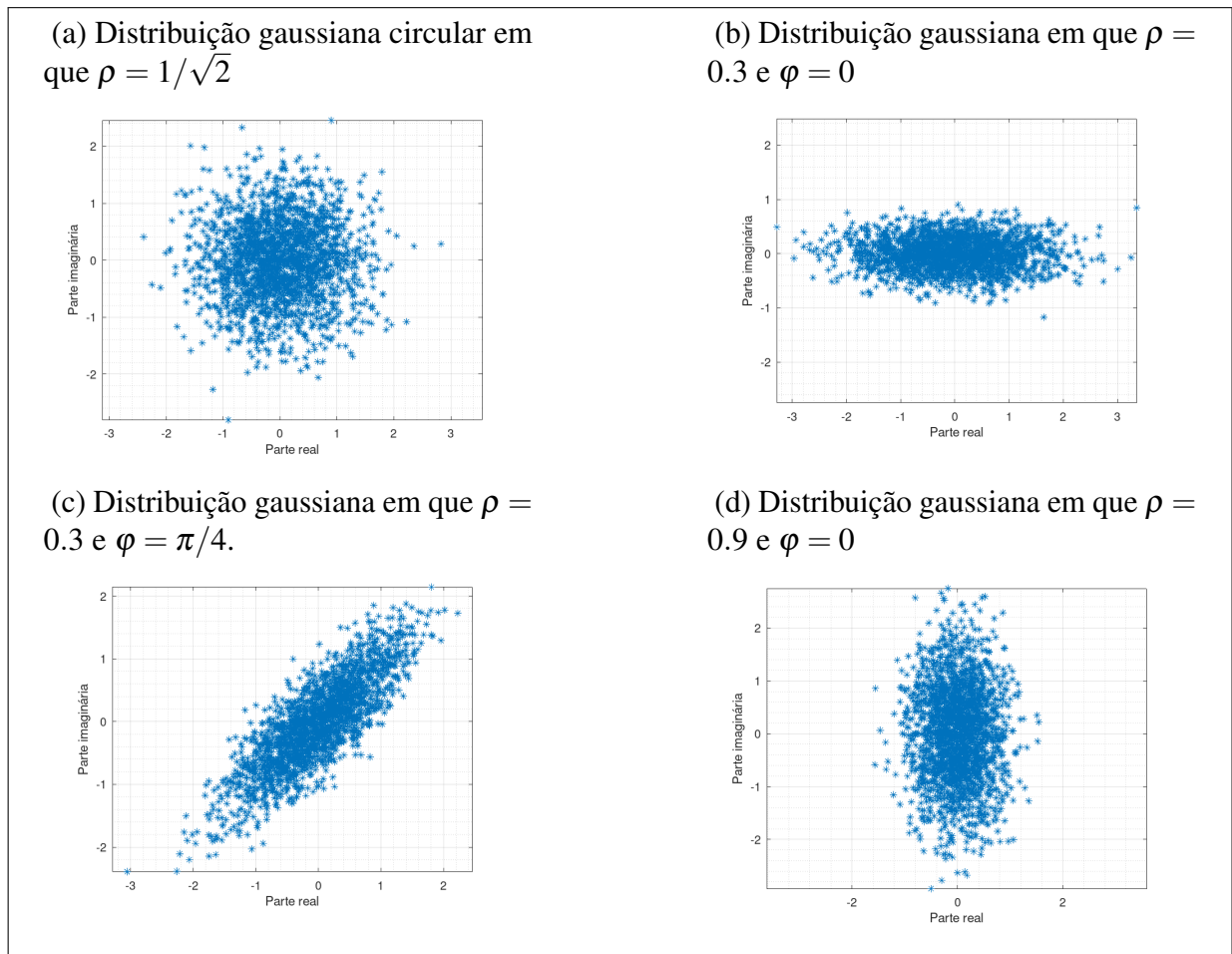
(h) Pseudo autocorrelação do processo dado (e)



Fonte: Elaborada pelo autor.

a partir das pseudo-variâncias $r_{ZZ} = (1 - 2\rho^2)e^{j2\varphi}\sigma^2$ e da variância $r_{ZZ^*} = \sigma^2$, obtidas diretamente do modelo do sinal (4.9). Observe que para que se tenha a circularidade nesse caso então $\rho_c = 0$, equivalente a $r_{ZZ} = 0$. Esse valor de propriedade leva à uma razão de perfil de $\rho = 1/\sqrt{2}$ para todo ângulo φ , nesse caso as componentes serão multiplicadas pelo mesmo fator e serão descorrelacionadas entre si. A Figura 13 ilustra com diferentes valores de fatores de perfis e rotação. Nela os processos são mostrados em perfil, em que fica mais evidente capturar

Figura 13 – Perfis de processos GWSS que modelam sinais



Fonte: Elaborada pelo autor.

as propriedades de circularidade (ou não-circularidade). Quanto mais próximo ρ fica de zero, maior a predominância da energia da parte real, Figura 13 (b). De forma contrária, quanto mais próxima da unidade, mais o perfil tem sua energia na parte imaginária como mostra a Figura 13 (d). O parâmetro ρ , um escalar não consegue rotacionar o sinal, por isso o perfil em (b) é rotacionado pela exponencial complexa em $\pi/4$ em (c).

Este capítulo encerra o tratamento básico que é requerido para a construção da teoria estatística envolvida no estudo de sinais complexos. Noções de variáveis complexas e

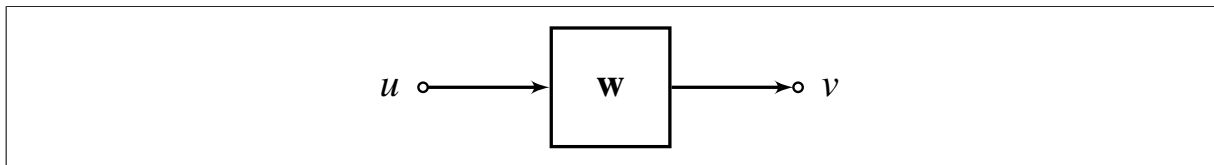
circularidade, e suas relações com momentos de segunda ordem como a pseudo-correlação, serão fundamentais no desenvolvimento de técnicas de filtragens de sinais complexos que serão assunto do próximo capítulo.

5 FILTROS DISCRETOS COMPLEXOS

A abordagem da filtragem complexa é motivada pela necessidade de comparar e destacar semelhanças com filtros amplamente lineares. Essa comparação permite uma análise mais abrangente de construções, expressões e simulações entre os dois tipos de filtros. Neste capítulo, são apresentados os principais conceitos relativos à filtragem discreta de sinais complexos.

A filtragem desempenha um papel fundamental no processamento de sinais. Matematicamente, um filtro é uma função que recebe um sinal como entrada e produz um sinal de saída processado de forma arbitrária. O sinal de saída geralmente mantém o mesmo formato do sinal de entrada, seja contínuo ou discreto. Além disso, é possível encontrar situações em que um tipo de sinal resulta em outro: por exemplo, um amostrador converte um sinal contínuo em um sinal discreto. Inversamente, por meio da reconstrução de sinal, é possível converter um sinal discreto em um contínuo¹. Como modelo geral de filtro, dado um sinal qualquer u como

Figura 14 – Exemplo de sinal u (contínuo ou discreto) filtrado por um filtro indicado por w , resultando em um sinal de saída v



Fonte: elaborada pelo autor.

entrada uma operação de filtragem é do tipo $v = w(u)$ para cada instante, contínuo ou discreto. A operação é indicada pelo diagrama na Figura 14. O processamento de sinais é complexo quando os sinais considerados u e v , bem como os coeficientes do filtro w tem parâmetros complexos, caso considerado no presente trabalho. Também não há perda de generalização ao aplicar a teoria de sinais complexos ao caso real, por isso a escolha imediata em se trabalhar dessa forma.

Esta dissertação dará predileção aos sinais do tipo discreto, devido à possibilidade de simulação computacional. Não obstante, a partir desse ponto serão considerados somente os sinais como complexos discretos, a menos que expresse o contrário. O mesmo vale para os filtros discretos considerados, que terão natureza linear e o número finito de coeficientes (valores multiplicativos). Os filtros discretos se enquadram em duas categorias de resposta: os que tem uma resposta infinita ao impulso (IIR) e os de resposta finita ao impulso (FIR), este último um caso particular do primeiro, ambos abordados a seguir.

¹ Existem também filtros cujo sinal de entrada e resposta são de naturezas diferentes.

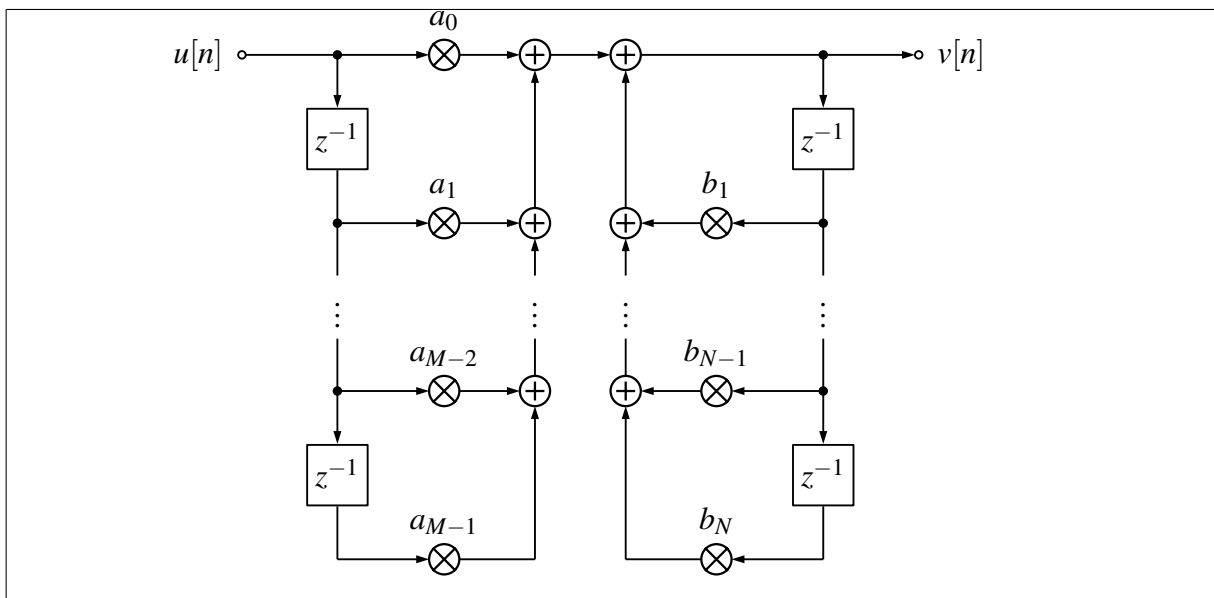
5.1 Filtros discretos IIR

Um filtro linear de resposta infinita é aquele que depende recursivamente de saídas passadas bem como de uma combinação linear de entradas atrasadas, ou seja, tem a forma matemática

$$v[n] = \sum_{k=0}^{M-1} a_k u[n-k] + \sum_{k=1}^N b_k v[n-k] . \quad (5.1)$$

Essa equação permite a representação do filtro de acordo com a Figura 15 a seguir, que explicita como o sinal é tratado pela equação dada. Nessa representação os blocos indicados pela legenda

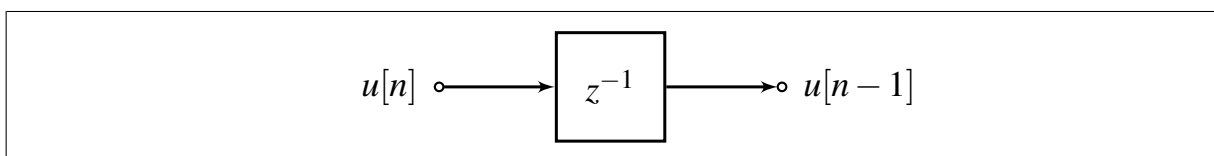
Figura 15 – Esquema de um filtro transversal discreto IIR com $M \times N$ coeficientes



Fonte: elaborada pelo autor.

z^{-1} são chamados atrasadores unitários, devido à propriedade do atraso dado pela transformada Z. Cada atrasador unitário funciona como indica a Figura 16, resultando em uma versão passada do

Figura 16 – Atrasador unitário, que recebe o símbolo “ z^{-1} ” devido à transformada Z associada ao atraso unitário



Fonte: elaborada pelo autor.

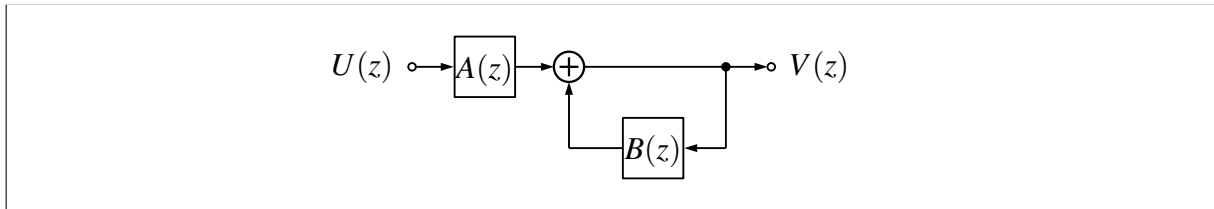
sinal em uma unidade de tempo discreta. Voltando à construção presente na Figura 15, observe que a mesma possui somente três estruturas fundamentais na filtragem: somadores, atrasadores e

multiplicadores. Aplicando a transformada Z na expressão (5.1) tem-se a resposta ao impulso dada por:

$$H(z) = \frac{V(z)}{U(z)} = \frac{\sum_{k=0}^{N-1} a_k z^{-k}}{1 + \sum_{k=1}^M b_k z^{-k}} = \frac{A(z)}{1 + B(z)} \quad (5.2)$$

para $A(z) = \sum_{k=0}^{N-1} a_k z^{-k}$ e $B(z) = \sum_{k=1}^M b_k z^{-k}$. O filtro é recursivo pois possui realimentação, permitindo resposta infinita mesmo para entrada finita (Oppenheim; Willsky; Young, 1983). Isso fica evidente na representação em termos de suas funções de transferências (típica em teoria de controle) como mostrado na Figura 17. Em geral, filtros IIR conseguem boas respostas

Figura 17 – Esquema de um filtro transversal discreto IIR em termos de sua função de transferência.



Fonte: elaborada pelo autor.

mesmo com poucos coeficientes, em contrapartida são propensos a divergirem, mesmo para sinais limitados. Filtros IIR devem ser bem projetados em resposta em frequência, com cuidado especial à critérios de estabilidade.

5.2 Filtros discretos FIR

Os filtros empregados nas simulações deste estudo são classificados como FIR (*Finite Impulse Response*), uma categoria específica de filtros IIR. Nos filtros FIR, os coeficientes associados à retroalimentação das saídas são todos iguais a zero, ou seja, $b_k = 0$, para todo $k \in [1, N]$. Aplicando essa condição na equação 5.1 e renomeando as constantes como $a_k = w_k^*$ (observe que antecipadamente considerar os coeficientes como conjugados não afeta a solução do filtro, apenas resulta no conjugado da solução final), obtemos a expressão temporal do filtro FIR:

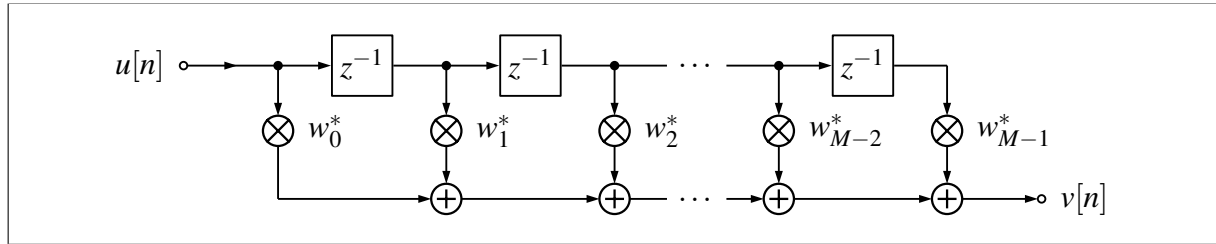
$$v[n] = \sum_{k=0}^{M-1} w_k^* u[n-k]. \quad (5.3)$$

Expressando de forma compacta em notação vetorial:

$$v[n] = \mathbf{w}^H \mathbf{u}[n], \quad (5.4)$$

onde $\mathbf{w} = [w_0 \dots w_{M-1}]^T$ é um vetor composto pelos coeficientes do filtro, e $\mathbf{u}[n] = [u[n] \dots u[n - M + 1]]^T$ é o vetor das amostras do sinal no instante n . O número de coeficientes M de um filtro também é conhecido como *tamanho* ou *comprimento* do filtro. A configuração típica do filtro é ilustrada na Figura 18. Por possuir uma construção simples sem realimentação o filtro FIR é mais

Figura 18 – Esquema de um filtro transversal discreto FIR de tamanho M .



Fonte: elaborada pelo autor.

fácil de se trabalhar, sem necessidade de qualquer análise no domínio da frequência. Como será visto a seguir, essa configuração permite a fácil construção de soluções no domínio do tempo, ótimas ou adaptativas. Esse tipo de filtro também é chamado de transversal, devido à sua forma quando posto como mostrado na Figura 18.

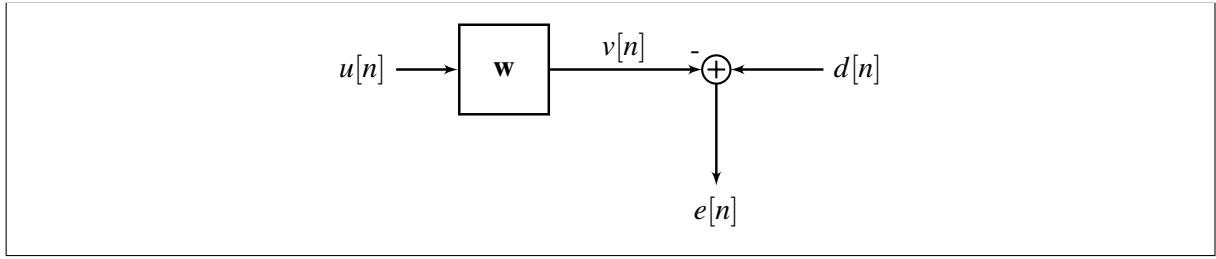
5.3 Modelagem de filtros transversais discretos.

O projeto de filtragem se dá ao escolher os valores dos coeficientes sob certas condições. O método básico será o de construir uma função de custo, que é a esperança do erro quadrático médio e por meio desta, dois principais caminhos podem ser tomados: encontrar o ponto ótimo da função diretamente (solução de Wiener) ou usar a expressão para buscar um método iterativo para estimar tal ponto ótimo (adaptativo). Em ambos os casos como ocorre a minimização do erro quadrático médio, esse critério é chamado de MSE (*Mean Square Error*).

A configuração básica de uso do filtro é mostrada na Figura 19 a seguir, em que a caixa indicada por $\mathbf{w}$ é a configuração transversal FIR presente na Figura 18, que possui comprimento M . O sinal de entrada $u[n]$ representa o sinal de treinamento (ou desejado) $d[n]$, possivelmente contaminada por ruídos, que pode ter passado por algum canal e/ou muitas vezes atrasada. Esse sinal de entrada é posto na forma vetorial de acordo com (5.4), fornecendo uma saída $v[n] = \mathbf{w}^H \mathbf{u}[n]$. O sinal de erro é dado por $e[n] = d[n] - v[n]$, leva à expressão do chamado custo (representada pela letra $\mathcal{J}$) expressa por

$$\mathcal{J} = E\{|e[n]|^2\} = E\{e[n]e^*[n]\}. \quad (5.5)$$

Figura 19 – Esquema de um filtro transversal discreto FIR com comparação da saída com o sinal desejado



Fonte: elaborada pelo autor.

Com uso das definições, a expressão do custo é expandida de

$$\mathcal{J} = E\{(d[n] - \mathbf{w}^H \mathbf{u}[n]) (d[n] - \mathbf{w}^H \mathbf{u}[n])^*\}$$

para a forma

$$\mathcal{J} = E\{d[n]d^*[n]\} - E\{d[n]\mathbf{u}^H[n]\}\mathbf{w} - \mathbf{w}^H E\{d^*[n]\mathbf{u}[n]\} + \mathbf{w}^H E\{\mathbf{u}[n]\mathbf{u}^H[n]\}\mathbf{w}.$$

Renomeando as esperanças na equação acima, temos:

A variância escalar do sinal desejado (considerado de média nula) é denotada por $\sigma_d^2 = E\{d[n]d^*[n]\} = E\{|d[n]|^2\}$. O vetor de correlação cruzada entre o sinal desejado e o sinal recebido é representado por:

$$\mathbf{p} = E\{d^*[n]\mathbf{u}[n]\} = \begin{bmatrix} E\{d^*[n]u[n]\} \\ \vdots \\ E\{d^*[n]u[n-M+1]\} \end{bmatrix}_{M \times 1}.$$

Enquanto a matriz de correlação do sinal de entrada é expressa como:

$$\mathbf{R} = E\{\mathbf{u}[n]\mathbf{u}^H[n]\} = \begin{bmatrix} E\{u[n]u^*[n]\} & \cdots & E\{u[n]u^*[n-M+1]\} \\ \vdots & \ddots & \vdots \\ E\{u[n-M+1]u^*[n]\} & \cdots & E\{u[n-M+1]u^*[n-M+1]\} \end{bmatrix}_{M \times M},$$

Com essas definições, a expressão pode ser simplificada em sua forma vetorial como:

$$\mathcal{J} = \sigma_d^2 - \mathbf{p}^H \mathbf{w} - \mathbf{w}^H \mathbf{p} + \mathbf{w}^H \mathbf{R} \mathbf{w}, \quad (5.6)$$

que representa a equação da função custo na filtragem complexa convencional. Esta forma quadrática é determinada por uma matriz simétrica semi-positiva definida $\mathbf{R}$, permitindo o cálculo de um mínimo global. As estatísticas em σ_d^2 , $\mathbf{p}$ e $\mathbf{R}$ são características do sinal de treinamento e do sinal recebido, tornando $\mathbf{w}$ a variável ajustável no filtro em função de algum critério. Duas formas principais de ajuste serão consideradas adiante.

5.3.1 Filtragem ótima (de Wiener)

A função custo definida anteriormente em (5.6) levanta a questão de quais conjuntos de coeficientes $\mathbf{w}$ minimizariam essa função. Esses conjuntos representariam os melhores coeficientes para minimizar o erro quadrático médio. A resposta é encontrada ao buscar os pontos críticos da função $\mathcal{J}$ em relação aos coeficientes do filtro. Essa tarefa pode ser simplificada utilizando a Tabela 1, que fornece as derivadas para expressões de primeira e segunda ordens. Aplicando essas derivadas à função custo, chegamos à expressão:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}^H} = -\mathbf{p} + \mathbf{R}\mathbf{w}. \quad (5.7)$$

Essa expressão é função dos coeficientes $\mathbf{w}$. A solução para essa expressão de primeiro grau fornece os M coeficientes ótimos:

$$\mathbf{w}^{\text{otp}} = \mathbf{R}^{-1}\mathbf{p}. \quad (5.8)$$

A expressão acima representa a solução ótima: o conjunto estacionário de coeficientes para o filtro que minimiza o erro médio quadrático. Embora forneça diretamente os melhores coeficientes nesse aspecto, a expressão envolve uma inversão matricial e depende de estatísticas teóricas, o que pode ser impraticável em implementações reais. Uma abordagem viável para contornar essa questão será discutida no final desta seção.

É importante observar a relação entre o erro e o sinal no gradiente (5.7). Podemos expressá-la em termos da esperança, da seguinte forma:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}^H} = -\mathbf{E}\{d^*[n]\mathbf{u}[n]\} + \mathbf{E}\{\mathbf{u}[n]\mathbf{u}^H[n]\mathbf{w}\}.$$

Colocando o vetor $\mathbf{u}[n]$ em evidência na equação anterior:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}^H} = -\mathbf{E}\mathbf{u}[n] \underbrace{(d[n] - \mathbf{u}^H[n]\mathbf{w})}_{=e[n]}.$$

Portanto, a relação entre o erro e o sinal recebido é dada por:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}^H} = -\mathbf{E}\mathbf{u}[n]e^*[n]. \quad (5.9)$$

Esta equação mostra que a melhor escolha dos coeficientes do filtro faz com que o erro seja ortogonal ao sinal recebido, indicando que $\mathbf{E}\{\mathbf{u}[n]e^*[n]\} = 0$ para uma solução ótima.

Observe que a regra de inversão matricial presente em (5.8) requer as estatísticas para o vetor de correlação e para a matriz de correlação. Essas estatísticas não são possíveis de

obter no mundo real, pois requerem um número infinito de realizações devido ao operador de esperança. Nesse caso, pode-se estimar o vetor de correlação cruzada como:

$$\hat{\mathbf{p}} = \frac{1}{N} \sum_{n=1}^N d^*[n] \mathbf{u}[n] , \quad (5.10)$$

e a matriz de correlação por:

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^N \mathbf{u}[n] \mathbf{u}^H[n] . \quad (5.11)$$

Aqui, N é o tamanho da amostra de sinal a ser comparado com o sinal desejado (o valor também pode ser definido de outra forma, desde que tenha um número de amostras suficiente). Ambas as estimativas podem ser utilizadas para calcular o ponto ótimo estimado:

$$\hat{\mathbf{w}}^{\text{otp}} = \hat{\mathbf{R}}^{-1} \hat{\mathbf{p}} , \quad (5.12)$$

que representa uma solução mais palpável na prática do que aquela presente em (5.8), mas ainda utiliza a inversão matricial da correlação.

5.3.2 O método do gradiente (*Steepest Descent*)

A fim de evitar o problema computacional relacionado à inversão matricial presente na solução ótima (5.8), é utilizado um método iterativo bastante conhecido como o método do gradiente. O método consiste em ir aproximando o próximo ponto em termos do anterior, de forma que

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \Delta \mathbf{w}[n] ,$$

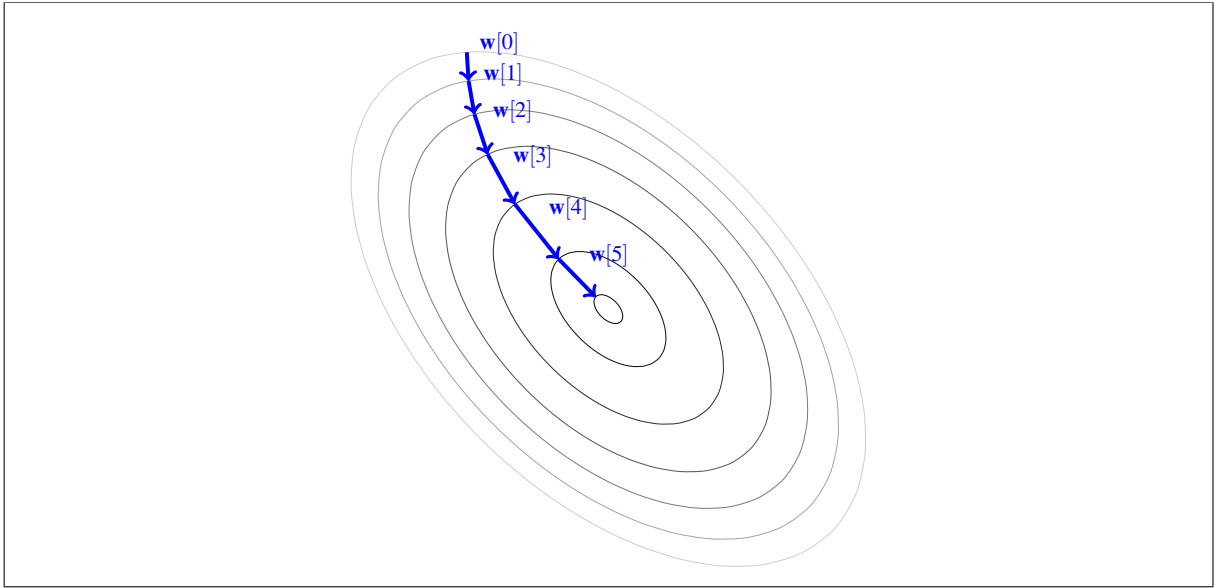
para um dado ponto inicial $\mathbf{w}[0]$ escolhido arbitrariamente. Para o método em questão, a variação nos coeficientes $\Delta \mathbf{w}[n]$ é dada em termos do chamado gradiente (5.7), calculado para cada ponto n e ponderado por um valor escalar $\mu > 0$, chamado de fator de passo. A expressão iterativa pode ser escrita por

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \mu \frac{\partial \mathcal{J}}{\partial \mathbf{w}^H} [n] . \quad (5.13)$$

Tal método iterativo se baseia no fato do vetor gradiente apontar para a direção de máximo crescimento (e com a conseguinte mudança de sinal, resultar na busca por um ponto de menor valor em cada iteração). Utilizando o resultado (5.9) é possível obter a forma:

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu E\{\mathbf{u}[n] e^*[n]\} \quad (5.14)$$

Figura 20 – Exemplo para o método iterativo do gradiente para seis iterações.



Fonte: elaborada pelo autor.

em termos do erro $e[n] = d[n] - v[n]$. O método pode ser visualizado para curvas de níveis como mostrado na Figura 20, considerando o tamanho do filtro de $M = 2$. Essas curvas de nível são as projeções da superfície parabólica da função $\mathcal{J}$ sobre o plano definido pelos coeficientes do filtro. Em cada ponto o vetor gradiente é computado utilizando a expressão (5.7), dado pela seta em azul, ponderado pelo fator de passo (que pode inclusive ser variável). O método ocorre até um certo número predefinido de iterações baseado em algum critério arbitrário.

5.3.3 O método dos Mínimos Quadrados (LMS)

Também chamado de método do gradiente estocástico, o método LMS (*Least Mean Square*) é uma aproximação para o método do gradiente mostrado na seção anterior, expresso pela equação (5.13). Apesar do método do gradiente evitar a inversão matricial ao utilizar um método iterativo, ele ainda depende de estatísticas que são impossíveis de se obter no mundo real pela própria definição de esperança. O método LMS usa estimativas para os valores de esperanças presentes em (5.13), tomando como valor de aproximação uma amostra em cada instante. Para o cálculo do vetor de correlação entre sinal e sinal desejado, o valor é estimado por

$$\hat{\mathbf{p}}[n] = d^*[n]\mathbf{u}[n], \quad (5.15)$$

e para a matriz de correlação

$$\hat{\mathbf{R}}[n] = \mathbf{u}[n]\mathbf{u}^H[n] . \quad (5.16)$$

Diferente das estimativas em (5.10) e (5.11), as estimativas instantâneas acima não precisam possuir um *rank* exigido para inversão, dessa forma cada estimativa pode ser tomada no ponto dado. Essa escolha, em termos do instante presente, permite uma melhor caracterização do método LMS, em que o lado direito da expressão (5.17) a seguir não depende de quaisquer valores passados. Essas aproximações estatísticas levam a uma estimativa para o gradiente sem cálculo da esperança, dado por

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}^H}[n] = -d^*[n]\mathbf{u}[n] + \mathbf{u}[n]\mathbf{u}^H[n]\mathbf{w} . \quad (5.17)$$

Por conseguinte o método de busca do ponto ótimo pode ser expresso através de (5.13) por

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu(d^*[n]\mathbf{u}[n] - \mathbf{u}[n]\mathbf{u}^H[n]\mathbf{w}[n]) .$$

Como o erro pode ser expresso por $e^*[n] = d^*[n] - \mathbf{u}^H[n]\mathbf{w}[n]$, é possível escrever a expressão em termos deste

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\mathbf{u}[n]e^*[n] . \quad (5.18)$$

A expressão é similar à (5.14) com a devida aproximação para $E\{\mathbf{u}[n]e^*[n]\}$ como $\mathbf{u}[n]e^*[n]$ em cada ponto. Essa regra de atualização é usada computacionalmente para o método LMS.

5.3.4 Fator de passo

Uma breve discussão sobre o fator de passo é apresentada aqui. Optar por um fator de passo muito pequeno pode resultar em um número excessivo de iterações. Por outro lado, um fator excessivamente grande pode impedir a convergência do algoritmo. Para evitar escolhas inadequadas, esta seção indica a faixa de valores que o fator de passo deve assumir.

Considere a equação de incremento baseada na expressão (5.7) para $\mathbf{w}[n]$, um ponto num dado instante

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu(\mathbf{p} - \mathbf{R}\mathbf{w}[n]) .$$

Subtraindo o valor $\mathbf{w}^{\text{opt}}$ fixo de cada membro

$$\begin{aligned}\mathbf{w}[n+1] - \mathbf{w}^{\text{opt}} &= \mathbf{w}[n] + \mu \mathbf{p} - \mu \mathbf{R} \mathbf{w}[n] - \mathbf{w}^{\text{opt}} \\ &= (\mathbf{I} - \mu \mathbf{R}) \mathbf{w}[n] - \mathbf{w}^{\text{opt}} + \underbrace{\mu \mathbf{R} \mathbf{R}^{-1}}_{\mathbf{w}^{\text{opt}}} \mathbf{p} \\ &= (\mathbf{I} - \mu \mathbf{R}) \mathbf{w}[n] - (\mathbf{I} - \mu \mathbf{R}) \mathbf{w}^{\text{opt}},\end{aligned}$$

para a matriz identidade $\mathbf{I}$ de tamanho $M \times M$. Simplificando mais a expressão anterior:

$$\mathbf{w}[n+1] - \mathbf{w}^{\text{opt}} = (\mathbf{I} - \mu \mathbf{R})(\mathbf{w}[n] - \mathbf{w}^{\text{opt}}). \quad (5.19)$$

Isso equivale a uma translação com o ponto ótimo como origem. A matriz de correlação poder ser posta em forma diagonal, i.e escrita por $\mathbf{R} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H$ com $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_M\}$ para $\mathbf{Q}$ a matriz de autovetores unitária. A equação (5.19), com a variável auxiliar $\boldsymbol{\eta}[n] = \mathbf{Q}^H(\mathbf{w}[n+1] - \mathbf{w}^{\text{opt}})$, se torna:

$$\boldsymbol{\eta}[n+1] = (\mathbf{I} - \mu \mathbf{\Lambda}) \boldsymbol{\eta}[n]. \quad (5.20)$$

Como a matriz $\mathbf{\Lambda}$ é diagonal, para cada componente i vale a relação:

$$\eta_i[n+1] = (1 - \mu \lambda_i) \eta_i[n].$$

Para que se tenha uma condição de estabilidade $|1 - \mu \lambda_i| < 1$, juntamente com a condição de não estacionaridade

$$0 < \mu < \frac{2}{\lambda_i},$$

para todos i autovalores. O passo escolhido é determinado pelo autovalor que faz com que a divisão seja a menor possível, ou seja, o maior valor entre os autovalores $\lambda_{\max} = \max\{\lambda_1, \dots, \lambda_M\}$.

Com isso a escolha teórica ideal para um fator de passo adequado se encontra no intervalo

$$0 < \mu < \frac{2}{\lambda_{\max}}.$$

Usar o fator de passo dado por essa expressão envolve a diagonalização da matriz de correlação, um processo dispendioso. Por isso utiliza-se a propriedade que o traço de uma matriz, invariante às transformações unitárias, de forma que $\lambda_{\max} \leq \text{tr}\{\mathbf{R}\}$ e

$$0 < \mu < \frac{2}{\text{tr}\{\mathbf{R}\}}. \quad (5.21)$$

Como $\mathbf{R}$ é uma esperança, utiliza-se uma estimativa da correlação do sinal (5.11) em seu lugar para aplicações práticas.

Este capítulo desempenha um papel crucial ao fornecer uma visão geral do que será abordado nos filtros amplamente lineares. O próximo capítulo demonstrará a equivalência de muitas das expressões aqui apresentadas com o caso expandido, proporcionando uma comparação entre ambos.

5.4 Filtros Amplamente Lineares

As seções anteriores introduziram os conceitos fundamentais dos filtros lineares. Partindo do modelo discreto transversal, derivamos o critério de minimização do erro quadrático médio (MSE). O filtro utilizado possui uma estrutura representada na Figura 18, com uma saída definida como $v[n] = \mathbf{w}^H \mathbf{u}[n]$, caracterizando uma relação sesquilinear expressa por um produto interno hermitiano.

Este capítulo introduz a estrutura amplamente linear (aumentada), cuja saída é definida como $v[n] = \mathbf{w}^H \mathbf{u}[n] + \mathbf{h}^H \mathbf{u}^*[n]$. A partir desta estrutura, realizamos a modelagem do sistema. Todo esse procedimento é conduzido por meio de uma nova função custo associada

5.4.1 Modelagem

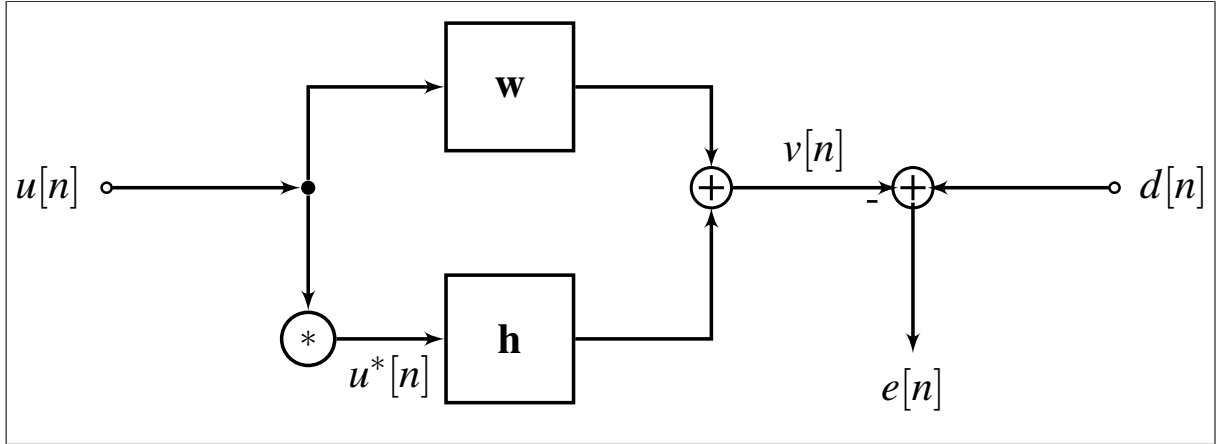
A estrutura de um filtro amplamente linear segue exatamente a construção matemática descrita pela equação (3.17), necessitando de dois filtros. Como indica a Figura 21 cada filtro de coeficientes $\mathbf{w}$ e $\mathbf{h}$ recebe um sinal discreto $u[n]$ e seu conjugado, respectivamente. Os coeficientes podem ser estimados com base na minimização do erro quadrático médio da expressão amplamente linear associada à um sinal desejado (ou de treinamento) $d[n]$. Cada filtro descrito por $\mathbf{w}$ e $\mathbf{h}$ possui configuração discreta transversal como exibido na Figura 18. Pela modelagem amplamente linear, a saída do filtro é expressa por:

$$v[n] = \mathbf{w}^H \mathbf{u}[n] + \mathbf{h}^H \mathbf{u}^*[n] . \quad (5.22)$$

Quando esta saída é subtraída do sinal de treinamento, o sinal resultante é chamado de sinal de erro $e[n]$. O objetivo é minimizar o erro quadrático médio:

$$E\{|e[n]|^2\} = E\{|d[n] - v[n]|^2\} .$$

Figura 21 – Configuração para um discreto amplamente linear. O sinal é dividido de forma que a parte não conjugada passa por um filtro $\mathbf{w}$ e seu conjugado por $\mathbf{h}$. Em seguida os sinais são adicionadas em uma única saída $v[n]$



Fonte: elaborada pelo autor.

Cada filtro é dado por um vetor de coeficientes complexos da forma:

$$\mathbf{w} = \begin{bmatrix} w_0 \\ \vdots \\ w_{M-1} \end{bmatrix} \text{ e } \mathbf{h} = \begin{bmatrix} h_0 \\ \vdots \\ h_{M-1} \end{bmatrix} \in \mathbb{C}^M ,$$

O sinal recebido pelo filtro é da forma

$$\mathbf{u}[n] = \begin{bmatrix} u[n] \\ u[n-1] \\ \vdots \\ u[n-M+1] \end{bmatrix} \in \mathbb{C}^M .$$

Pelo esquema do filtro amplamente linear esse sinal de entrada é dividido em duas partes: uma versão sem alterações é processada por M coeficientes de $\mathbf{w}$ e outra versão conjugada é filtrada pelos M coeficientes de $\mathbf{h}$. O objetivo, similar a que foi feito anteriormente, é encontrar o conjuntos de coeficientes $\mathbf{w}$ e $\mathbf{h}$ que minimizem $\mathcal{J} = E\{|e[n]|^2\}$, escrita em termos dos coeficientes complexos do filtro. Desenvolvendo a expressão para o erro quadrático médio,

$$\begin{aligned} \mathcal{J} &= E\{|e[n]|^2\} = E\{(d[n] - v[n])(d^*[n] - v^*[n])\} \\ &= E\{(d[n] - \mathbf{w}^H \mathbf{u}[n] - \mathbf{h}^H \mathbf{u}^*[n])(d^*[n] - \mathbf{u}^H[n] \mathbf{w} - \mathbf{u}^T[n] \mathbf{h})\} , \end{aligned}$$

que pode ser colocada na forma

$$\begin{aligned} \mathcal{J} &= E\{|d[n]|^2\} - E\{\mathbf{u}^H[n] d[n]\} \mathbf{w}_1 - E\{\mathbf{u}^T[n] d[n]\} \mathbf{w}_2 - \mathbf{w}_1^H E\{\mathbf{u}[n] d^*[n]\} - \mathbf{w}_2^H E\{\mathbf{u}^*[n] d^*[n]\} \\ &\quad + \mathbf{w}_1^H E\{\mathbf{u}[n] \mathbf{u}^H[n]\} \mathbf{w}_1 + \mathbf{w}_1^H E\{\mathbf{u}[n] \mathbf{u}^T[n]\} \mathbf{w}_2 + \mathbf{w}_2^H E\{\mathbf{u}^*[n] \mathbf{u}^H[n]\} \mathbf{w}_1 + \mathbf{w}_2^H E\{\mathbf{u}^*[n] \mathbf{u}^T[n]\} \mathbf{w}_2 . \end{aligned}$$

É possível representar a equação anterior de forma mais compacta, facilitando a visualização. Isso pode ser feito renomeando as esperanças como $\sigma_d^2 = E\{|d[n]|^2\}$, $\mathbf{p} = E\{\mathbf{u}[n]d^*[n]\}$, $\mathbf{q} = E\{\mathbf{u}[n]d[n]\}$, $\mathbf{R} = E\{\mathbf{u}[n]\mathbf{u}^H[n]\}$ e $\tilde{\mathbf{R}} = E\{\mathbf{u}[n]\mathbf{u}^T[n]\}$, em que a última expressão é a matriz de pseudo auto-correlação. A função custo pode ser reescrita como:

$$\mathcal{J} = \sigma_d^2 - \mathbf{p}^H \mathbf{w} - \mathbf{q}^T \mathbf{h} - \mathbf{w}^H \mathbf{p} - \mathbf{h}^H \mathbf{q}^* + \mathbf{w}^H \mathbf{R} \mathbf{w} + \mathbf{w}^H \tilde{\mathbf{R}} \mathbf{h} + \mathbf{h}^H \tilde{\mathbf{R}}^* \mathbf{w} + \mathbf{h}^H \mathbf{R}^* \mathbf{h}. \quad (5.23)$$

Em forma matricial composta complexa a expressão anterior é expressa como:

$$\mathcal{J} = \sigma_d^2 - \begin{bmatrix} \mathbf{p} \\ \mathbf{q}^* \end{bmatrix}^H \begin{bmatrix} \mathbf{w} \\ \mathbf{h} \end{bmatrix} - \begin{bmatrix} \mathbf{w} \\ \mathbf{h} \end{bmatrix}^H \begin{bmatrix} \mathbf{p} \\ \mathbf{q}^* \end{bmatrix} + \begin{bmatrix} \mathbf{w} \\ \mathbf{h} \end{bmatrix}^H \begin{bmatrix} \mathbf{R} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}}^* & \mathbf{R}^* \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{h} \end{bmatrix}.$$

Definindo o vetor de correlação cruzada composto² por:

$$\mathbf{p}_c[n] = E\{d^*[n] \begin{bmatrix} \mathbf{u}[n] \\ \mathbf{u}^*[n] \end{bmatrix}\} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q}^* \end{bmatrix},$$

o vetor complexo composto como:

$$\mathbf{w}_c = \begin{bmatrix} \mathbf{w} \\ \mathbf{h} \end{bmatrix},$$

e a matriz de correlação aumentada por:

$$\underline{\mathbf{R}} = \begin{bmatrix} \mathbf{R} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}}^* & \mathbf{R}^* \end{bmatrix},$$

é possível escrever a relação simplificada

$$\mathcal{J} = \sigma_d^2 - \mathbf{p}_c^H \mathbf{w}_c - \mathbf{w}_c^H \mathbf{p}_c + \mathbf{w}_c^H \underline{\mathbf{R}} \mathbf{w}_c. \quad (5.24)$$

A expressão anterior, que representa o custo do erro quadrático médio para o filtro amplamente linear, possui a mesma forma da equação (5.6), com o devido emprego de vetores compostos e da correlação aumentada.

5.4.2 Coeficientes ótimos

Para encontrar o valor dos coeficientes ótimos (solução de Wiener) calculam-se os pontos críticos da função custo. Derivando a expressão (5.23) com as derivadas (3.21) e calculando os valores nos quais essas se anulam:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}^H} = 0 \quad \text{e} \quad \frac{\partial \mathcal{J}}{\partial \mathbf{h}^H} = 0.$$

² Esse vetor representa a correlação (em versão aumentada) entre o sinal desejado e o sinal armazenado no filtro em um dado instante.

Essas derivadas resultam nas equações com os coeficientes ótimos:

$$\begin{cases} \frac{\partial J}{\partial \mathbf{w}^H} = -\mathbf{p} + \mathbf{R}\mathbf{w}^{\text{opt}} + \tilde{\mathbf{R}}\mathbf{h}^{\text{opt}} = 0 \\ \frac{\partial J}{\partial \mathbf{h}^H} = -\mathbf{q}^* + \tilde{\mathbf{R}}^*\mathbf{w}^{\text{opt}} + \mathbf{R}^*\mathbf{h}^{\text{opt}} = 0 \end{cases} . \quad (5.25)$$

As equações anteriores estão acopladas. Resolvendo estas, os seguintes coeficientes ótimos são obtidos:

$$\mathbf{w}^{\text{opt}} = [\tilde{\mathbf{R}}^{-1}\mathbf{R} - \mathbf{R}^{*-1}\tilde{\mathbf{R}}^*]^{-1}(\tilde{\mathbf{R}}^{-1}\mathbf{p} - \mathbf{R}^{*-1}\mathbf{q}^*)$$

$$\mathbf{h}^{\text{opt}} = [\mathbf{R}^{-1}\tilde{\mathbf{R}} - \tilde{\mathbf{R}}^{*-1}\mathbf{R}^*]^{-1}(\mathbf{R}^{-1}\mathbf{p} - \tilde{\mathbf{R}}^{*-1}\mathbf{q}^*) .$$

Escrevendo de forma compacta a Equação (5.25):

$$\underline{\mathbf{R}}\mathbf{w}_c = \begin{bmatrix} \mathbf{R} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}}^* & \mathbf{R}^* \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{h} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q}^* \end{bmatrix} .$$

A notação de vetor complexo aumentado como em (5.24) permite escrever a solução (5.25) de maneira mais curta da forma:

$$\mathbf{w}_c^{\text{opt}} = \underline{\mathbf{R}}^{-1}\underline{\mathbf{p}}. \quad (5.26)$$

Esta tem a mesma forma da solução usual linear complexa e exprime a solução ótima amplamente linear³.

5.4.3 Ortogonalidade aumentada

Uma propriedade importante é a ortogonalidade das soluções em relação ao erro, o que simplifica expressões e facilita o trabalho com subespaços. Considerando o sinal de saída com os coeficientes ótimos:

$$E\{\mathbf{u}[n]e_{\text{opt}}^*[n]\} = E\{\mathbf{u}[n](d^*[n] - v_{\text{opt}}^*[n])\} = E\{\mathbf{u}[n]d^*[n]\} - E\{\mathbf{u}[n]v_{\text{opt}}^*[n]\} ,$$

onde $y^*[n] = \mathbf{u}^H[n]\mathbf{w}^{\text{opt}} + \mathbf{u}^T[n]\mathbf{h}^{\text{opt}}$. Portanto,

$$E\{\mathbf{u}[n]e^*[n]\} = \mathbf{p} - E\{\mathbf{u}[n](\mathbf{u}^H[n]\mathbf{w}^{\text{opt}} + \mathbf{u}^T[n]\mathbf{h}^{\text{opt}})\} = \mathbf{p} - E\{\mathbf{u}[n]\mathbf{u}^H[n]\}\mathbf{w}^{\text{opt}} - E\{\mathbf{u}[n]\mathbf{u}^T[n]\}\mathbf{h}^{\text{opt}}$$

$$E\{\mathbf{u}[n]e^*[n]\} = \mathbf{p} - \mathbf{R}\mathbf{w}^{\text{opt}} - \tilde{\mathbf{R}}\mathbf{h}^{\text{opt}} = 0$$

³ Uma dedução mais compacta dos gradientes poderia ser feita escrevendo diretamente da equação (5.24) e a derivando em relação ao conjugado do vetor aumentado, mas aqui tentou-se explicitar as relações lineares cruzadas entre coeficientes e mostrar diretamente a forma da matriz $\underline{\mathbf{R}}$.

conforme a primeira equação em (5.25). Para o sinal conjugado, o raciocínio é similar:

$$E\{\mathbf{u}^*[n]e^*[n]\} = E\{\mathbf{u}^*[n](d^*[n] - y^*[n])\} = E\{\mathbf{u}^*[n]d^*[n]\} - E\{\mathbf{u}^*[n]y^*[n]\} ,$$

$$E\{\mathbf{u}^*[n]e^*[n]\} = \mathbf{q}^* - E\{\mathbf{u}^*[n](\mathbf{u}^H[n]\mathbf{w}^{\text{opt}} + \mathbf{u}^T[n]\mathbf{h}^{\text{opt}})\}$$

$$E\{\mathbf{u}^*[n]e^*[n]\} = \mathbf{q}^* - E\{\mathbf{u}^*[n]\mathbf{u}^H[n]\}\mathbf{w}^{\text{opt}} - E\{\mathbf{u}^*[n]\mathbf{u}^T[n]\}\mathbf{h}^{\text{opt}} = \mathbf{q}^* - \tilde{\mathbf{R}}^*\mathbf{w}^{\text{opt}} - \mathbf{R}^*\mathbf{h}^{\text{opt}} = 0 ,$$

de acordo com a expressão (5.25), que demonstra a ortogonalidade também em relação ao conjugado do sinal, indicando que o erro não só é ortogonal ao sinal original, como também ao seu conjugado.

5.4.4 Solução Iterativa da Descida mais Íngreme

A equação (5.26) fornece de forma compacta a solução ótima para encontrar os coeficientes do filtro. No entanto, essa abordagem pode ser computacionalmente custosa devido à necessidade de inversão matricial. Para contornar esse problema, o método utilizado será o do gradiente estocástico, que utiliza a função custo definida em (5.23). As derivadas em (5.25) fornecem os vetores de coeficientes da seguinte forma:

$$\frac{\partial J}{\partial \mathbf{w}^H} = -\mathbf{p} + \mathbf{R}\mathbf{w} + \tilde{\mathbf{R}}\mathbf{h} \quad \text{e} \quad \frac{\partial J}{\partial \mathbf{h}^H} = -\mathbf{q}^* + \tilde{\mathbf{R}}^*\mathbf{w} + \mathbf{R}^*\mathbf{h}$$

Essas derivadas indicam a direção de maior crescimento da função custo. A partir da solução linear (Nossek, 2015) e utilizando as equações acima, é possível determinar uma solução amplamente linear similar. Em cada iteração, esse método busca o próximo ponto com base no gradiente. Para um dado par de passos μ_w e μ_h apropriados e para valores iniciais $\mathbf{w}[0]$ e $\mathbf{h}[0]$ a primeira iteração é dada como:

$$\mathbf{w}[1] = \mathbf{w}[0] + \mu_w \Delta \mathbf{w} \quad \text{e} \quad \mathbf{h}[1] = \mathbf{h}[0] + \mu_h \Delta \mathbf{h} ,$$

em que $\Delta \mathbf{w}$ e $\Delta \mathbf{h}$ são os gradientes, com sinais invertidos. A expressão para o próximo passo é

$$\mathbf{w}[1] = \mathbf{w}[0] + \mu_w (\mathbf{p} - \mathbf{R}\mathbf{w}[0] - \tilde{\mathbf{R}}\mathbf{h}[0])$$

$$\mathbf{h}[1] = \mathbf{h}[0] + \mu_h (\mathbf{q}^* - \tilde{\mathbf{R}}^*\mathbf{w}[0] - \mathbf{R}^*\mathbf{h}[0]) ,$$

que independe de inversões matriciais. Por indução, a equação geral para um número arbitrário de $n + 1$ passos em termos da iteração anterior é:

$$\begin{cases} \mathbf{w}[n+1] &= \mathbf{w}[n] + \mu_w (\mathbf{p} - \mathbf{R}\mathbf{w}[n] - \tilde{\mathbf{R}}\mathbf{h}[n]) \\ \mathbf{h}[n+1] &= \mathbf{h}[n] + \mu_h (\mathbf{q}^* - \tilde{\mathbf{R}}^*\mathbf{w}[n] - \mathbf{R}^*\mathbf{h}[n]) \end{cases} . \quad (5.27)$$

Essa expressão é conhecida como solução iterativa da descida mais íngreme. Caso os passos sejam iguais $\mu_w = \mu_h = \mu$ é possível representar a solução como definido em (5.26). O caso amplamente linear é similar ao caso usual, apenas com o uso dos vetores aumentados e a matriz de correlação aumentada.

5.4.5 Fator de passo em filtros aumentados

A partir da expressão da solução iterativa (5.27), e de maneira semelhante ao que foi feito no caso complexo não-aumentado em (5.19), podemos escrever:

$$\mathbf{w}_c[n+1] = (\mathbf{I} - \mu \mathbf{R}) \mathbf{w}_c[n] + \mu \mathbf{p},$$

onde $\mathbf{I}$ é uma matriz de dimensão $2M \times 2M$. Utilizando uma transformação de coordenadas para os coeficientes do filtro, dada por $\mathbf{c}_c[n+1] = \mathbf{w}_c[n+1] - \mathbf{w}_c^{\text{opt}}$, a expressão pode ser reescrita como:

$$\mathbf{c}_c[n+1] = (\mathbf{I} - \mu \mathbf{R}) \mathbf{c}_c[n]$$

$$\mathbf{Q}^H \mathbf{c}_c[n+1] = (\mathbf{Q}^H \mathbf{Q} - \mu \mathbf{Q}^H \mathbf{Q} \mathbf{\Lambda}) \mathbf{Q}^H \mathbf{c}_c[n]$$

onde $\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H = \mathbf{R}$ representa a decomposição em autovalores. Tomando $\mathbf{Q}^H \mathbf{c}_c[n] = \eta_c[n]$, a expressão se torna:

$$\eta_c[n+1] = (\mathbf{I} - \mu \mathbf{\Lambda}) \eta_c[n].$$

Com isso, podemos determinar o intervalo para o fator de passo aceitável, levando em consideração a condição de estabilidade $|1 - \mu \lambda_k| < 1$, onde $\mathbf{\Lambda} = \text{diag}(\lambda_k)_{k=1}^{2(M+1)}$. Assim:

$$0 < \mu < \frac{2}{\lambda_{\max}}. \quad (5.28)$$

O valor escalar $\lambda_{\max}$ é o maior autovalor da matriz de covariância aumentada. Devido à complexidade computacional envolvida no cálculo de $\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H = \mathbf{R}$, emprega-se o mesmo artifício apresentado em (5.21). Neste caso, é utilizado o traço da matriz de covariância aumentada $\mathbf{R}$ para determinar uma estimativa do maior autovalor, utilizando a seguinte relação:

$$0 < \mu < \frac{2}{\text{tr}(\mathbf{R})}. \quad (5.29)$$

É importante destacar que o passo necessário para a abordagem da aproximação amplamente linear deve, em geral, ser menor do que o utilizado na abordagem linear padrão. Isso se deve ao fato de que o traço da matriz de covariância aumentada $\mathbf{R}$ na forma (5.29) é geralmente maior do que o traço da matriz de covariância original $\mathbf{R}$.

5.4.6 Aproximação WLLMS

Com base nas expressões da descida mais íngreme vistas anteriormente, a solução iterativa dos mínimos quadrados amplamente lineares (WLLMS - widely linear least mean square) fundamenta-se na estimativa, a cada passo, das matrizes de correlação e pseudocorrelação, assim como na estimação das correlações cruzadas entre o sinal e o sinal desejado (Nossek, 2015). Dessa forma:

$$\mathbf{R}[n] = \mathbf{u}[n]\mathbf{u}^H[n]$$

$$\tilde{\mathbf{R}}[n] = \mathbf{u}[n]\mathbf{u}^T[n]$$

$$\mathbf{p}[n] = \mathbf{u}[n]d^*[n]$$

$$\mathbf{q}^*[n] = \mathbf{u}^*[n]d^*[n]$$

.

Tais valores são estimativas instantâneas para cada passo dado. Substituindo cada uma dessas equações nas expressões em (5.27), obtém-se:

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu_w \mathbf{u}[n] (d^*[n] - \mathbf{u}^H[n]\mathbf{w}[n] - \mathbf{u}^T[n]\mathbf{h}[n])$$

$$\mathbf{h}[n+1] = \mathbf{h}[n] + \mu_h \mathbf{u}^*[n] (d^*[n] - \mathbf{u}^H[n]\mathbf{w}[n] - \mathbf{u}^T[n]\mathbf{h}[n]) ,$$

em que a expressão comum entre parenteses é o conjugado do erro amplamente linear dada por

$$e^*[n] = d^*[n] - v^*[n] = d^*[n] - \mathbf{u}^H[n]\mathbf{w}[n] - \mathbf{u}^T[n]\mathbf{h}[n] .$$

É possível escrever as expressões escrevendo em termos do erro como

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu_w \mathbf{u}[n]e^*[n]$$

$$\mathbf{h}[n+1] = \mathbf{h}[n] + \mu_h \mathbf{u}^*[n]e^*[n] .$$

Com o uso dos vetores compostos e novamente com um passo comum é possível simplificar a expressão acima para a equação:

$$\mathbf{w}_c[n+1] = \mathbf{w}_c[n] + \mu \mathbf{u}[n]e^*[n] . \quad (5.30)$$

Essa última tem a mesma forma do caso complexo usual. Lembrando que o passo comum é um escalar escolhido de acordo com a expressão (5.29).

5.4.7 Realizações WLLMS e LMS comparadas

A fim de ilustrar os resultados obtidos anteriormente, na presente seção é feito um conjunto de simulações numéricas para a análise de convergência e de desempenho de filtros amplamente lineares sobre aqueles lineares a fim de comparar resultados. O algoritmo do LMS aumentado (WLLMS), dado a partir da equação (5.30) é mostrado no Algoritmo 1 a seguir. Como entradas, o algoritmo recebe o sinal desejado e o sinal $\mathbf{u}$ em sua totalidade, bem como o

Algoritmo 1: Algoritmo do método WLLMS

Entrada: u , d , μ , e o número de coeficientes M

início

Seja Q = comprimento total do sinal d ou u .

Inicialize $\mathbf{w}_c$ com comprimento $2M$ e $n = 0$

Inicialize $\mathbf{e} = [e[0], \dots, e[Q-1]]^T$ com tamanho Q .

para $M-1 \leq n \leq Q-1$ **faça**

$\underline{\mathbf{u}} = [u[n], \dots, u[n-M+1], u^*[n], \dots, u^*[n-M+1]]^T$

$e[n] = d[n] - \mathbf{w}_c^H \underline{\mathbf{u}}$

$\mathbf{w}_c \leftarrow \mathbf{w}_c + \mu \underline{\mathbf{u}} e^*$

$n \leftarrow n + 1$

fim

fim

passo e o tamanho do filtro. A cada iteração o erro é armazenado em um vetor a ser retornado. A partir desse vetor de erros é calculada uma média por várias realizações do algoritmo anterior, utilizando a aproximação de Monte Carlo. A saída do algoritmo pode ser o vetor $\mathbf{e}$. Ele contém os erros calculados para cada amostra do sinal.

O algoritmo LMS complexo tem o mesmo formato do mostrado anteriormente, bastando tomar $\underline{\mathbf{u}} = \mathbf{u}$ como o vetor de entrada e $\mathbf{w}_c = \mathbf{w}$ com tamanho M , como pode ser visto no Algoritmo 2. Uma das vantagens de se utilizar filtros amplamente lineares reside no fato de que as formas dos algoritmos se mantêm equivalentes. Isso também ocorre no Método de Newton que será no capítulo seguinte.

Algoritmo 2: Algoritmo LMS complexo usual

Entrada: u, d, μ , e o número de coeficientes M .

início

Seja Q = comprimento total do sinal d ou u .

Inicialize $\mathbf{w}$ comprimento M e $n = 0$

Inicialize $\mathbf{e} = [e[0], \dots, e[Q-1]]^T = 0$ de tamanho Q .

para $M-1 \leq n \leq Q-1$ **faça**

$\mathbf{u} = [u[n], \dots, u[n-M+1]]^T$

$e[n] = d[n] - \mathbf{w}^H \mathbf{u}$

$\mathbf{w} \leftarrow \mathbf{w} + \mu \mathbf{u} e^*$

$n \leftarrow n + 1$

fim

fim

5.4.7.1 Modelagem, escolha de parâmetros e alguns resultados

A primeira parte de simulações utiliza como sinal desejado uma composição feita a partir de ruídos gaussianos WSS. O sinal utilizado foi o modelo gaussiano apresentado em (4.9) dado por:

$$d[n] = e^{j\varphi_S} \left(\sqrt{1 - \rho_S^2} X_S[n] + j\rho_S Y_S[n] \right). \quad (5.31)$$

O ruído também foi modelado da mesma forma. A relação com o sinal desejado foi linear, feita pela relação (SNR), de forma que:

$$N[n] = \frac{e^{j\varphi_N} \left(\sqrt{1 - \rho_N^2} X_N[n] + j\rho_N Y_N[n] \right)}{SNR}. \quad (5.32)$$

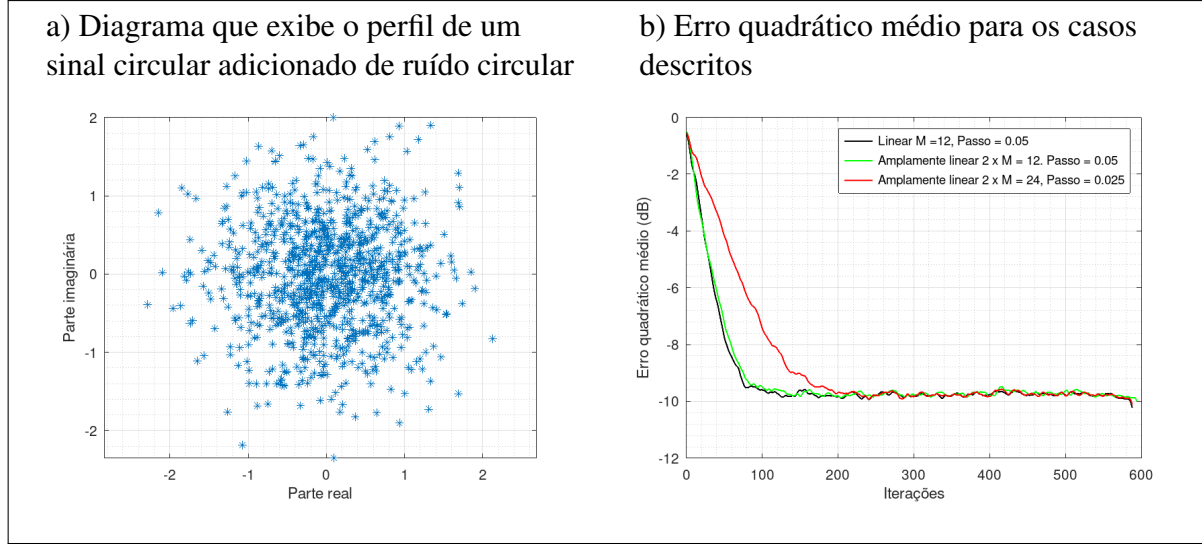
As componentes $X_S[n], Y_S[n], X_N[n]$ e $Y_N[n]$ foram modeladas como gaussianas de média nula e variância unitária, descorrelacionadas entre si. Os valores de $\rho_S, \rho_N, \varphi_S$ e φ_N são os valores em (4.9) para o sinal e o ruído respectivamente.

O algoritmo foi utilizado para minimização de ruído, sobre um canal sem perdas ou distorções (unitário) de forma que $u[n] = d[n] + N[n]$. A métrica utilizada foi o erro quadrático médio, em função das iterações. Para estimar a média desse erro foi utilizada a aproximação de Monte Carlo sobre o Algoritmo 1. Em todos os casos o algoritmo é comparado com sua versão complexa usual o Algoritmo 2, submetido à mesma estimação de Monte Carlo.

Para o caso circular alguns valores foram atribuídos afim de verificar como seria feita a melhor comparação. A Figura 22 ilustra a primeira escolha, que será a escolha do tamanho do

filtro para o sinal gaussiano média nula e variância unitária de suas componentes reais, também circular com $\rho = 1/\sqrt{2}$. O ruído também foi modelado circular com SNR = 10. Convém notar o

Figura 22 – Perfil e erro quadrático médio para um sinal gaussiano circular



Fonte: elaborada pelo autor.

comportamento para tamanhos de filtros e passos diferentes.

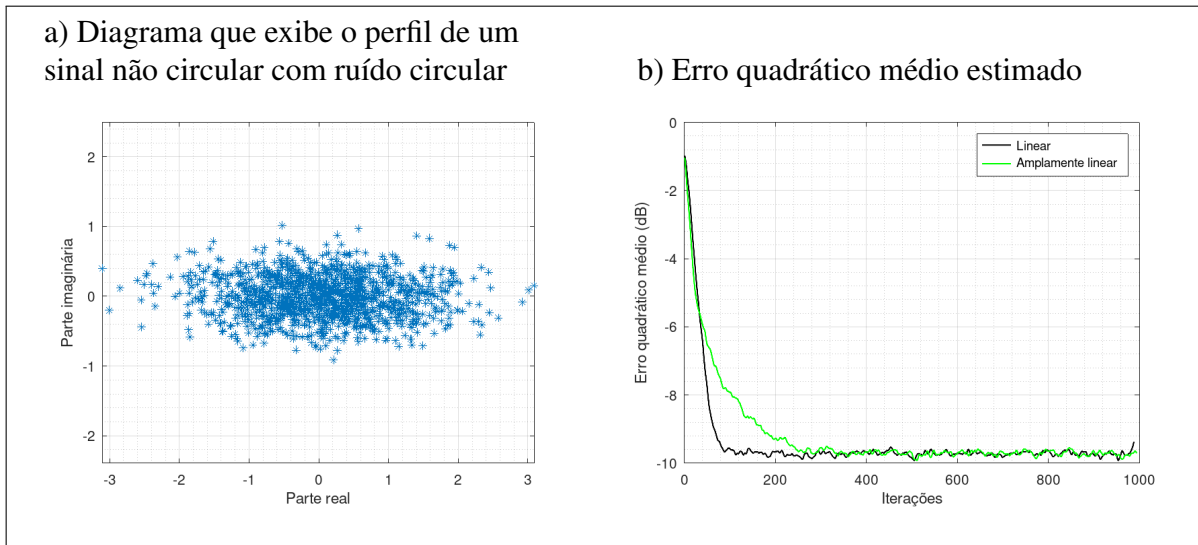
O caso em que o passo é igual para ambos e o tamanho do filtro total é o mesmo, parece ser a melhor forma de se comparar ambos algoritmos, pois no caso circular ambos algoritmos devem ser equivalentes. A Figura 22 (b) mostra o caso linear (complexo usual), com tamanho de filtro $M_{\text{linear}} = 12$, equivale a um amplamente linear de mesmo passo, com o tamanho total dos dois filtros $M_{\text{aumentada}} = 2M_{\text{individual}}$ igual ao caso usual linear $M_{\text{aumentada}} = M_{\text{linear}} = 12$ (cada filtro separadamente tem $M_{\text{individual}} = 6$). O passo, determinado por (5.21) e (5.29), é aproximadamente o mesmo, pois a matriz de autocorrelação no primeiro caso tem dimensão $M_{\text{linear}} \times M_{\text{linear}}$, mesma dimensão da matriz aumentada $2M_{\text{individual}} \times 2M_{\text{individual}}$.

Para o caso em que o tamanho de cada filtro, no caso aumentado, é igual ao tamanho do filtro linear, $M_{\text{individual}} = M_{\text{linear}}$ (implicando em um filtro duas vezes maior que o usual $M_{\text{aumentada}} = 2M_{\text{linear}}$), o passo deve ser dividido pela metade pois o traço agora é o dobro. Caso o passo fosse o mesmo dos outros casos, $\mu = 0.05$, o filtro apresentaria um maior erro que o exibido, apesar da convergência mais rápida. Ainda de acordo com a Figura 22 a escolha do caso $M_{\text{linear}} = M_{\text{aumentada}}$ de mesmo passo parece ser o melhor comparativo nesse caso e também será utilizada para comparar o desempenho de sinais não circulares.

Utilizando a mesma expressão de sinal e ruído anterior foram realizadas as simulações para diferentes casos de ρ_S , φ_S , ρ_N e φ_N .

O primeiro caso, exibido na Figura 23, tem $\rho_s = 0,3$ e $\varphi_s = 0$ e o ruído circular. O

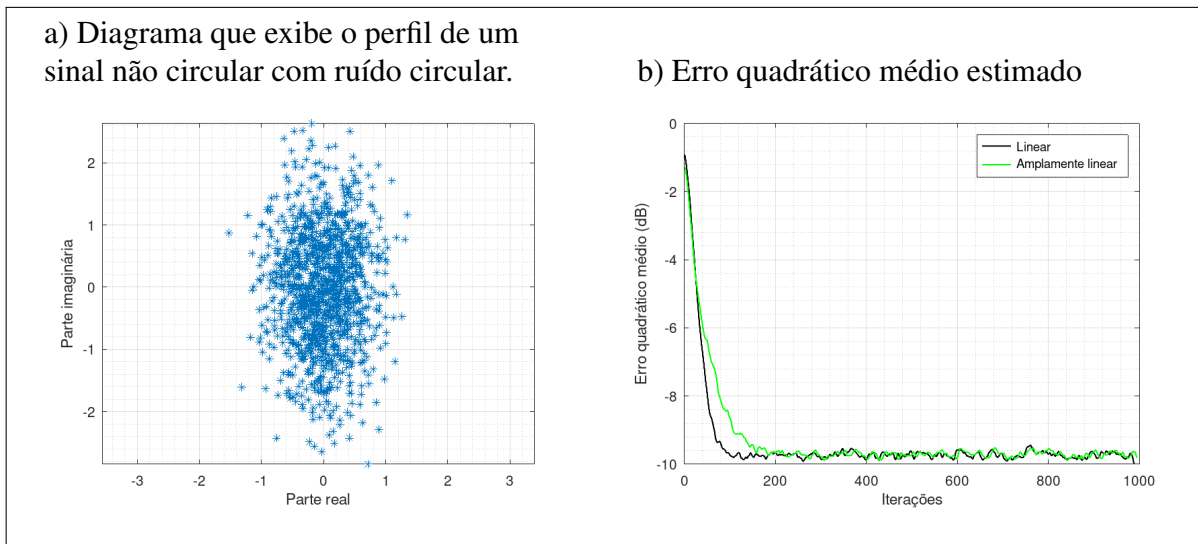
Figura 23 – Sinal $\rho_s = 0,3$, $\varphi_s = 0$ e ruído circular para os algoritmos linear e aumentado



Fonte: elaborada pelo autor.

passo utilizado foi o mesmo dos exemplos anteriores e se trata das mesmas configurações de filtro que os utilizados. Da mesma forma para $\rho_s = 0,9$, $\varphi_s = 0$ e mesmo ruído do caso anterior, o erro é mostrado na Figura 24. Ambos resultados mostram que mesmo para perfis não circulares

Figura 24 – Sinal $\rho_s = 0,9$, $\varphi_s = 0$ e ruído circular para os algoritmos linear e aumentado

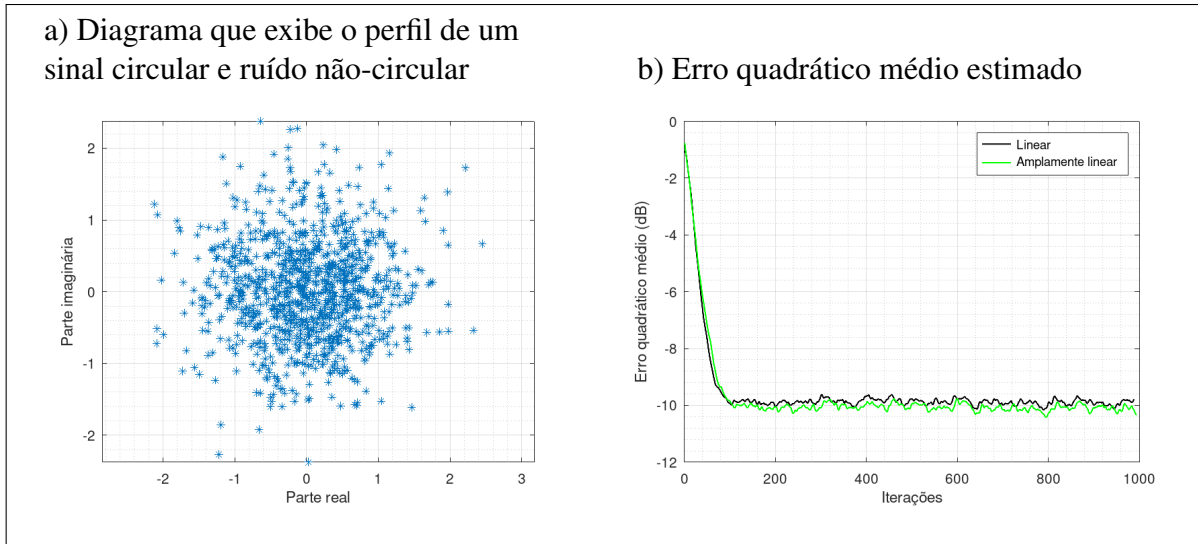


Fonte: elaborada pelo autor.

de sinal desejado, a performance é parecida nos casos usual e aumentado, sendo um pouco melhor em velocidade de convergência para o caso linear. O mesmo ocorre para algum $\varphi_s \neq 0$ no caso de ruído circular.

O algoritmo aumentado começa a fazer diferença na presença do ruído não-circular. A Figura 25 abaixo mostra o comportamento para $\rho_N = 0.2$ e $\varphi_N = 0$, com sinal desejado

Figura 25 – Sinal circular e ruído não-circular com $\rho_N = 0,2$ para os algoritmos linear e aumentado



Fonte: elaborada pelo autor.

circular.

Outro resultado é mostrado na Figura 26 a seguir, onde o sinal e o ruído agora não são mais circulares. O sinal tem o fator $\rho_s = 0.9$ com ângulo $\varphi_s = -\pi/3$. O ruído $\rho_N = 0.2$ e $\varphi_N = \pi/4$.

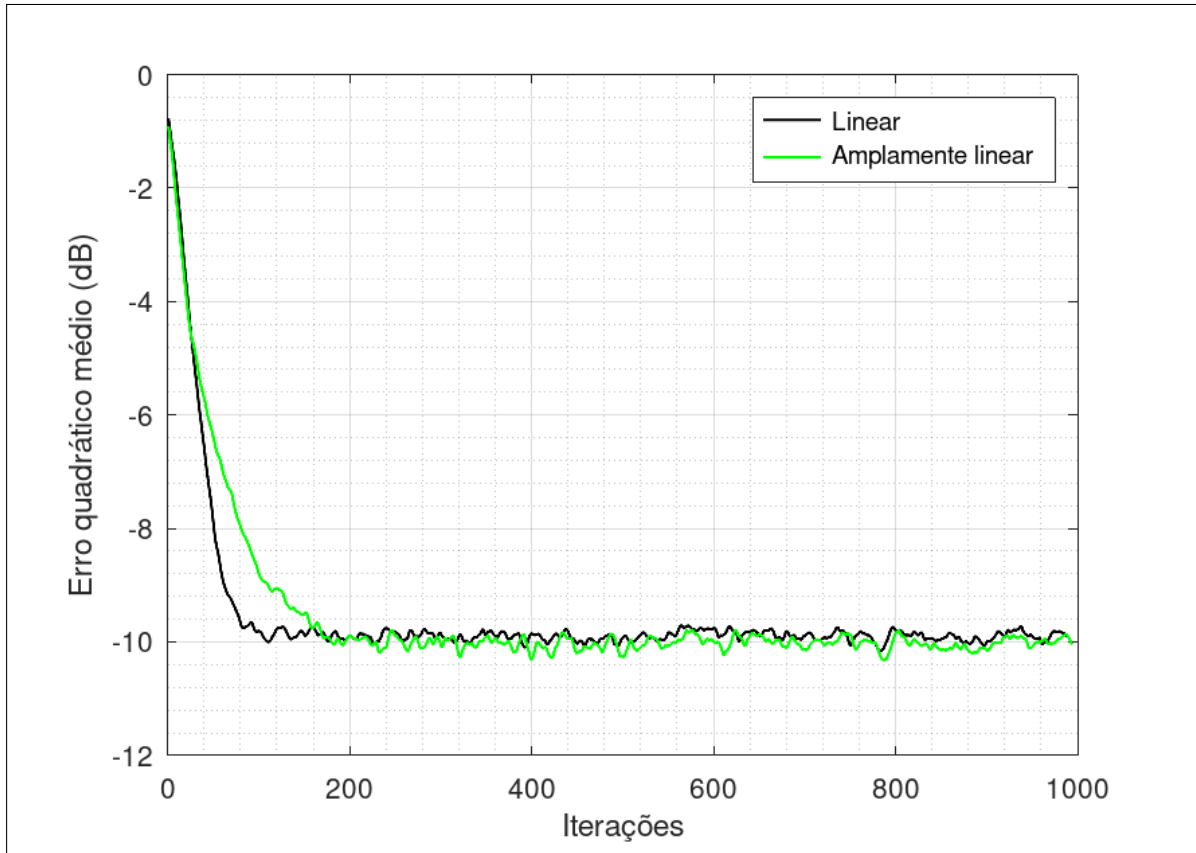
Esses exemplos mostram que, de forma geral, o erro se torna um pouco menor com o algoritmo aumentado para ruídos não circulares e sinal circular. Nos demais casos tem uma performance bem similar ao complexo usual com uma demora maior em convergência. A presença de fatores de passo menores faz com que esses casos demorem ainda mais à alcançar a convergência não influenciando tanto na melhoria do erro.

5.4.7.2 Exemplos com sinais 4QAM

O caso de sinais 4QAM para sistemas de comunicações é analisado nesta seção. O sinal desejado utilizado é uma constelação dado pelos pontos $\{(1, j), (1, -j), (-1, j), (-1, -j)\}$. O ruído é modelado ainda por (5.32), com mesmas especificações gerais dos casos anteriores: passo $\mu = 0,05$, tamanhos de filtro $M_{\text{linear}} = M_{\text{aumentado}} = 12$ e SNR = 10. Os casos são ilustrados na Figura 27 com as descrições das demais variáveis.

No caso 4QAM, há sempre um pequeno ganho em performance com velocidade de

Figura 26 – Erro médio quadrático para um sinal e ruído não-circulares com $\rho_s = 0.9$, $\varphi_s = -\pi/3$, $\rho_N = 0.2$ e $\varphi_N = \pi/4$

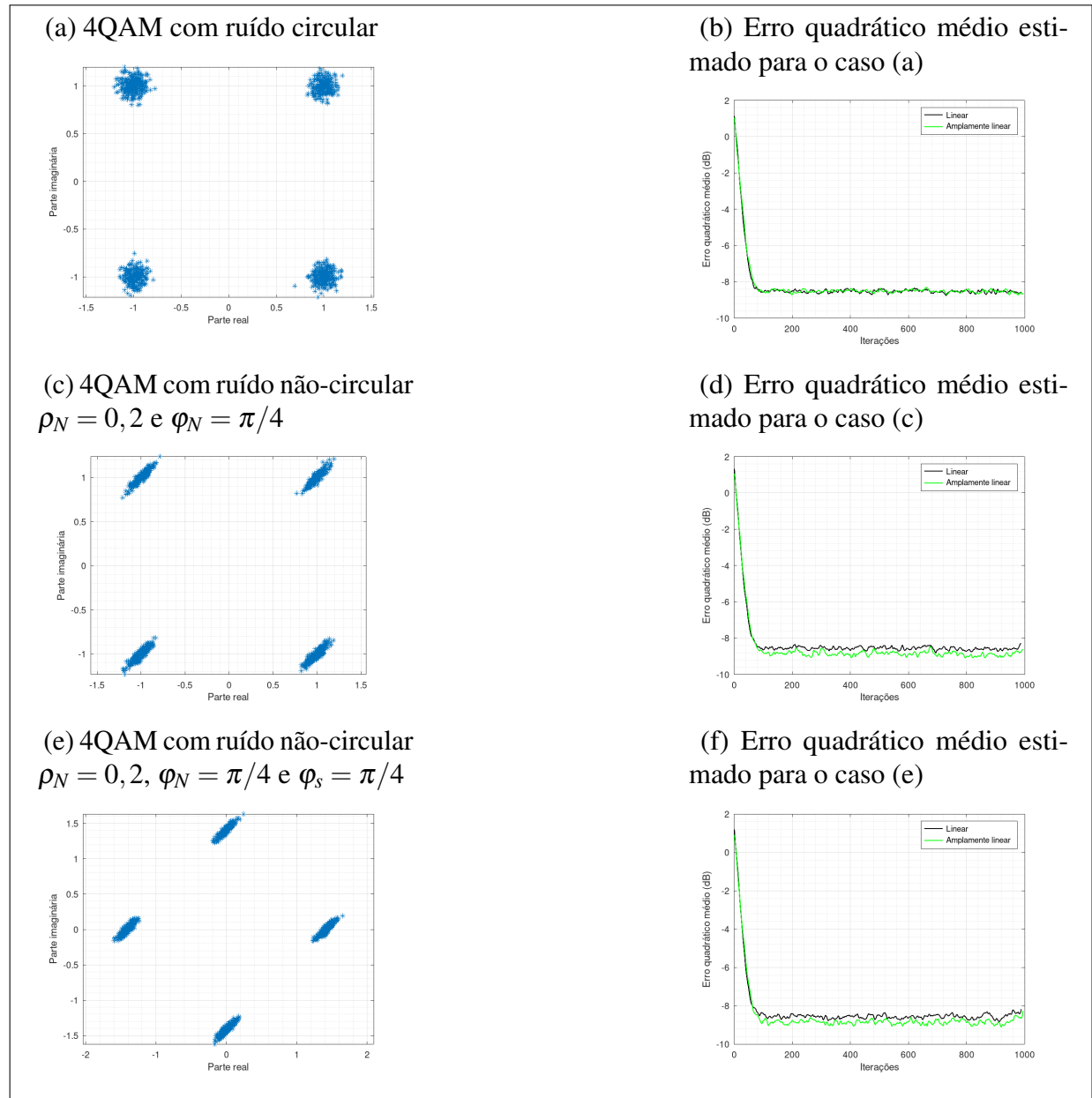


Fonte: elaborada pelo autor.

convergência similar em comparação com o sinal de treinamento gaussiano. Em geral sinais não gaussianos periódicos terão menor erro quando utilizam o algoritmo aumentado.

Este capítulo conclui a discussão sobre algoritmos complexos e amplamente lineares em suas variantes LMS. No próximo capítulo, aproveitaremos os conceitos e estruturas aprendidos até o momento para apresentar o Método de Newton. Este método oferece uma convergência mais rápida em comparação com os algoritmos LMS, embora isso venha acompanhado de um aumento no custo computacional.

Figura 27 – Realizações dos algoritmos para o caso 4QAM



Fonte: Elaborada pelo autor.

6 MÉTODO DE NEWTON AUMENTADO

6.1 Séries de Taylor

O método de Newton é uma abordagem bem estabelecida que utiliza aproximações lineares para encontrar raízes de equações. Essa técnica se baseia na série de Taylor para aproximar o valor de uma função em torno de um ponto específico. Embora a série de Taylor exata possa exigir uma soma infinita para uma precisão completa, uma aproximação útil pode ser obtida com um número finito de termos. A expressão geral é dada por:

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a) (x-a)^n . \quad (6.1)$$

Essa fórmula nos permite encontrar o valor de $f(x)$ a partir de qualquer ponto a , desde que as derivadas de f sejam conhecidas no ponto a . Quanto mais termos (derivadas) forem incluídos na série, mais precisa será a aproximação. Por outro lado, uma quantidade limitada de termos (derivadas de ordens superiores) resultará em uma estimativa menos precisa.

6.1.1 A aproximação de primeira ordem

Uma aproximação de primeira ordem, que considera apenas as potências unitárias na série de Taylor, é dada por:

$$f(x) \approx f(a) + f'(a)(x-a). \quad (6.2)$$

Essa forma simplificada é amplamente empregada em diversas áreas onde a velocidade dos cálculos é mais importante do que uma precisão extrema. Uma característica interessante dessa aproximação é que ela pode ser utilizada de forma recursiva para obter resultados que seriam de ordens superiores por meio de sucessivas iterações de primeira ordem. Essencialmente, essa é a ideia fundamental por trás do método de Newton.

6.1.2 Duas ou mais variáveis

Para duas variáveis a série de Taylor para uma dada função em torno de um ponto (a, b) é expressa por

$$f(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{1}{n!m!} \frac{\partial^{n+m} f(a, b)}{\partial x^n \partial y^m} (x-a)^n (y-b)^m . \quad (6.3)$$

O resultado anterior pode ser estendido para diversas variáveis em torno do ponto $\mathbf{x} = (x_1, \dots, x_n)$, que terá uma aproximação de segunda ordem em notação vetorial dada por:

$$f(\mathbf{x} + \mathbf{h}) \approx f(\mathbf{x}) + \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) \cdot \mathbf{h} + \langle \mathbf{h}, H(f(\mathbf{x})) \cdot \mathbf{h} \rangle, \quad (6.4)$$

em que o operador

$$\frac{\partial}{\partial \mathbf{x}} = \left[\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right] \quad (6.5)$$

é a versão real do operador descrito pela equação (3.21), o vetor que representa o incremento $\mathbf{h} = (h_1, \dots, h_n)$ e a matriz hessiana real é

$$H(f(\mathbf{x})) = \left[\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right]_{n \times n}, \quad (6.6)$$

que pode ser descrito em termos do operador em (6.5), como será visto adiante.

6.2 Raízes de uma função pelo método de Newton

Uma aproximação da série de Taylor (6.1) pode ser usada para buscar uma raiz da função $f(x)$. Para um dado ponto inicial arbitrário x_0 , a aproximação de primeira ordem é dada por

$$f(x_0 + h_0) \approx r = f(x_0) + f'(x_0)h_0$$

para um incremento $h_i = x_{i+1} - x_i$ conforme indicado na Figura para $i = 0$, a primeira iteração. De acordo com essa aproximação linear, o incremento pode ser escrito por

$$h_0 = -\frac{f(x_0)}{f'(x_0)},$$

que fornece a expressão do ponto seguinte

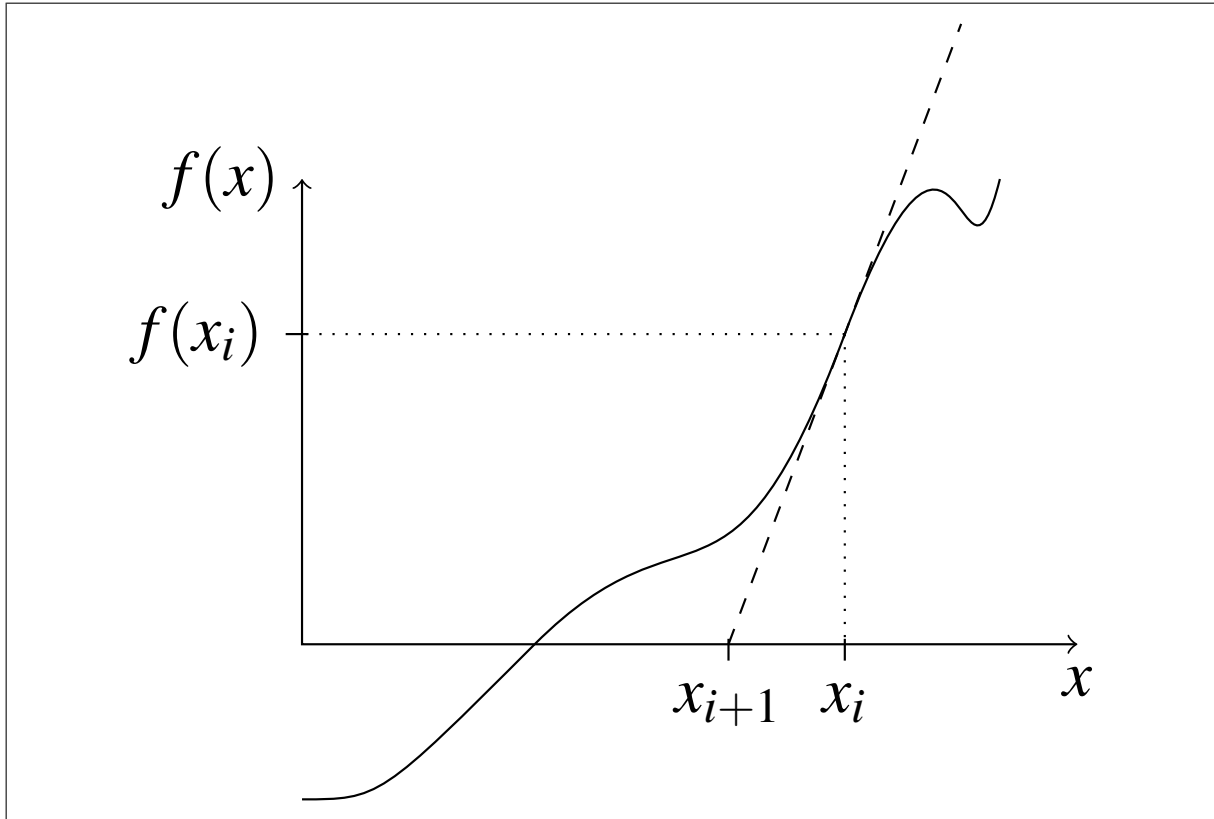
$$x_1 = x_0 + h_0 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Seguindo esse procedimento até uma iteração para um dado ponto particular na k -ésima iteração

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}. \quad (6.7)$$

A equação acima (6.7) é a forma geral de solução iterativa de busca de raízes que utiliza aproximações de primeira ordem. No caso de um funcional de mais de uma variável tem-se, $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$, e esse método não funciona, pois terá uma só equação de recursão para múltiplas

Figura 28 – Método de Newton para encontro de raízes para uma iteração em um dado instante i . A linha hachurada representa a derivada da função no ponto x_i indicada por $f'(x_i)$



Fonte: elaborada pelo autor.

variáveis. Em contrapartida, pode ser utilizado para funções vetoriais do tipo $\mathbf{f}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ como em

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix},$$

para $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ componentes de $\mathbf{f}(\mathbf{x})$. A expressão de aproximação para um vetor $\mathbf{f}$ coluna pelo teorema de Taylor em torno de um dado ponto $\mathbf{x}$ por meio de (6.3) é

$$\mathbf{r} = \mathbf{f}(\mathbf{x}) + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}) \cdot \mathbf{h}, \quad (6.8)$$

em que o ponto “ $\cdot$ ” indica o produto matricial. A equação recursiva de diferenças, que representa a aproximação de primeira ordem, é escrita por

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}_i) \right]^{-1} \cdot \mathbf{f}(\mathbf{x}_i).$$

O termo da derivada (Jacobiano) acima tem representação matricial

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f_l}{\partial x_m} \right]_{n \times n}, \quad (6.9)$$

num dado ponto $\mathbf{x}$ (l representa a linha e m a coluna). É importante frisar que o jacobiano não é dado por multiplicação matricial do operador com a função em forma vetorial: para se obter a mesma matriz jacobiana que esteja de acordo com (6.9), a operação equivalente utilizando o produto matricial seria

$$\left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right)^T = \frac{\partial}{\partial \mathbf{x}^T} \cdot \mathbf{f}^T = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} \cdot \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}, \quad (6.10)$$

ou da forma

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \mathbf{f} \cdot \frac{\partial}{\partial \mathbf{x}},$$

com o operador atuando da direita para esquerda sobre a função pelo produto matricial. Observe que essa forma de se representar operadores diferenciais em formas de vetores envolve um cuidado especial em representar de forma correta as operações componente a componente.

6.3 Método de Newton para otimização

6.3.1 Otimização para funções de uma variável

Ao utilizar o método de Newton considerando a derivada da função ao invés da mesma, obtém-se o ponto ótimo da função. Dessa forma, dada uma função $f : \mathbb{R} \rightarrow \mathbb{R}$, a raiz a ser calculada iterativamente é a de sua derivada. Isto equivale a utilizar (6.7) para a função $g(x) = f'(x)$, em que a expressão resultante é

$$x_{i+1} = x_i - \frac{g(x_i)}{g'(x_i)} = x_i - \frac{f'(x_i)}{f''(x_i)}. \quad (6.11)$$

A expressão acima equivale à otimização do passo para a aproximação de segunda ordem $q = f(x) + f'(x)h + 1/2 f''(x)h^2$ (a letra q evidencia a aproximação quadrática). Então a derivada

$$\frac{dq}{dh} = f'(x) + f''(x)h = 0 \quad (6.12)$$

implica no passo ótimo

$$h = -\frac{f'(x)}{f''(x)}, \quad (6.13)$$

mesmo dado na expressão (6.11). Os dois procedimentos anteriores são equivalentes ao próximo: o cálculo direto do ponto crítico da aproximação linear $r = f(x) + f'(x)h$, em relação à variável utilizada

$$\frac{dr}{dx} = f'(x) + f''(x)h = 0, \quad (6.14)$$

resultando no mesmo passo de (6.13) ou (6.11), mostrando a equivalência de possíveis deduções para o fator de passo da otimização.

6.3.2 Otimização para funções de múltiplas variáveis

Para múltiplas variáveis o procedimento é semelhante. Seja uma função escalar $f : \mathbb{R}^n \rightarrow \mathbb{R}$, a qual se buscam os pontos críticos locais. Para isso utiliza-se uma função g que seja a derivada da função f , de forma que

$$\mathbf{g}(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}(\mathbf{x}) \quad \cdots \quad \frac{\partial f}{\partial x_n}(\mathbf{x}) \right],$$

agora é a função vetorial a ter os zeros encontrados. Para a função $\mathbf{g} = \mathbf{f}^T = \partial f / \partial \mathbf{x}$, um vetor linha, basta transpor a equação (6.8), que fornece a expressão correspondente

$$\mathbf{r}^T = \mathbf{f}^T + \mathbf{h}^T \cdot \frac{\partial \mathbf{f}^T}{\partial \mathbf{x}^T} = \mathbf{g} + \mathbf{h}^T \cdot \frac{\partial \mathbf{g}}{\partial \mathbf{x}^T}. \quad (6.15)$$

A expressão acima, com a devida substituição de $\mathbf{g} = \partial f / \partial \mathbf{x}$, possui o termo

$$\frac{\partial \mathbf{g}}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial \mathbf{g}}{\partial x_1} \\ \vdots \\ \frac{\partial \mathbf{g}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial}{\partial x_1} \frac{\partial f}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} \frac{\partial f}{\partial x_n} \end{bmatrix},$$

representando o hessiano já definido em (6.6). Este pode ser posta em forma de produto matricial por meio de

$$\frac{\partial \mathbf{g}}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} \end{bmatrix} f = \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f,$$

que permite a escrita de (6.15)

$$\mathbf{r}^T = \frac{\partial f}{\partial \mathbf{x}} + \mathbf{h}^T \cdot \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f. \quad (6.16)$$

A equação (6.16) acima pode ser transposta

$$\mathbf{r} = \frac{\partial f}{\partial \mathbf{x}^T} + \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f \cdot \mathbf{h}, \quad (6.17)$$

onde o hessiano mantém sua forma devido à simetria (por construção) do operador

$$\frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial x_1} & \dots & \frac{\partial}{\partial x_n} \end{bmatrix} f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}. \quad (6.18)$$

Os pontos críticos são os zeros da função vetorial $\mathbf{r}$ definidos pela aproximação de primeira ordem (hiperplano)

$$\frac{\partial f}{\partial \mathbf{x}^T} + \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f \cdot \mathbf{h} = 0. \quad (6.19)$$

O resultado (6.17) também pode ser obtido pelo seguinte procedimento: dada $f : \mathbf{R}^n \rightarrow \mathbf{R}$, a aproximação linear da função é

$$r = f + \frac{\partial f}{\partial \mathbf{x}} \cdot \mathbf{h}. \quad (6.20)$$

Tomando a derivada para cada componente, colocando estas em notação coluna e tornando tal construção igual a zero, tem-se o sistema

$$\frac{\partial r}{\partial \mathbf{x}^T} = \frac{\partial f}{\partial \mathbf{x}^T} + \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f \cdot \mathbf{h} = 0. \quad (6.21)$$

Em quaisquer dos casos, o fator de passo para otimização é dado por

$$\mathbf{h} = - \left[\frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f \right]^{-1} \cdot \frac{\partial f}{\partial \mathbf{x}^T}, \quad (6.22)$$

para um dado ponto $\mathbf{x}$. Por isso a expressão para a equação de diferenças que representa o método será

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \left[\frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} f \right]^{-1} \cdot \frac{\partial f}{\partial \mathbf{x}^T}$$

ou

$$\mathbf{x}_{i+1} = \mathbf{x}_i - [H(f)]^{-1} \cdot \frac{\partial f}{\partial \mathbf{x}^T}. \quad (6.23)$$

6.3.3 Métodos comparados

Apesar de parecerem bem diferentes as equações do fator de passo unidimensional e multidimensional se confundem. A expressão para o fator de passo para o caso unidimensional, dada por

$$h = - [f''(x)]^{-1} f'(x) ,$$

é a mesma apresentada em (6.22). Basta considerar $x = (x_1, \dots, x_n)$ e $h = (h_1, \dots, h_n)$, com o hessiano igual à derivada segunda $f''(x) = H(f(x))$ e a derivada equivale ao gradiente da função.

6.3.4 Expressão real composta

Esta seção estende a escrita anterior da série de Taylor para as representações aumentadas do tipo $f(\mathbf{x}, \mathbf{y}) : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ para multivariáveis do tipo $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ e seus equivalentes por mudanças de variáveis para uma função do mesmo tipo $f(\mathbf{z}, \mathbf{z}^*)$ com $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ e $\mathbf{z}^* = \mathbf{x} - j\mathbf{y}$. Começando com real composto, para apresentar a notação composta utilizada, e após para o caso complexo aumentado.

Uma função do tipo $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ representa um caso específico de nomeação de uma variável $\mathbb{R}^{2n}$, em que esta é dividida em dois conjuntos de variáveis, posteriormente justificado pela construção de vetores complexos. A aproximação de primeira ordem para este tipo de função, de acordo com a expansão de Taylor, é descrita por

$$f(\mathbf{x} + \mathbf{h}, \mathbf{y} + \mathbf{k}) \approx r = f(\mathbf{x}, \mathbf{y}) + \frac{\partial f}{\partial \mathbf{x}} \cdot \mathbf{h} + \frac{\partial f}{\partial \mathbf{y}} \cdot \mathbf{k} \quad (6.24)$$

onde $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_n)$ serão os vetores que representam o ponto dado. As derivadas individuais para cada conjunto de variáveis é

$$\frac{\partial f}{\partial \mathbf{x}} = \left[\frac{\partial f}{\partial x_1} \quad \dots \quad \frac{\partial f}{\partial x_n} \right] \text{ e } \frac{\partial f}{\partial \mathbf{y}} = \left[\frac{\partial f}{\partial y_1} \quad \dots \quad \frac{\partial f}{\partial y_n} \right]$$

Com essa aproximação linear é possível usar algum dos métodos da seção anterior, de forma a encontrar o passo que fornece o ponto crítico local e otimiza a função. O método escolhido é o de encontrar os zeros para as derivadas pela diferenciação de (6.24) em relação aos operadores colunas $\partial/\partial \mathbf{x}^T$ e $\partial/\partial \mathbf{y}^T$, versões reais de (3.22), de acordo com as regras da seção anterior tem-se

$$\begin{aligned} \frac{\partial r}{\partial \mathbf{x}^T} &= \frac{\partial f}{\partial \mathbf{x}^T} + \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial f}{\partial \mathbf{x}} \cdot \mathbf{h} + \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial f}{\partial \mathbf{y}} \cdot \mathbf{k} , \\ \frac{\partial r}{\partial \mathbf{y}^T} &= \frac{\partial f}{\partial \mathbf{y}^T} + \frac{\partial}{\partial \mathbf{y}^T} \cdot \frac{\partial f}{\partial \mathbf{x}} \cdot \mathbf{h} + \frac{\partial}{\partial \mathbf{y}^T} \cdot \frac{\partial f}{\partial \mathbf{y}} \cdot \mathbf{k} . \end{aligned}$$

Reescrevendo as expressões acima em notação matricial composta por

$$\begin{bmatrix} \frac{\partial}{\partial \mathbf{x}^T} \\ \frac{\partial}{\partial \mathbf{y}^T} \end{bmatrix} r = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}^T} \\ \frac{\partial}{\partial \mathbf{y}^T} \end{bmatrix} f + \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial}{\partial \mathbf{y}} \\ \frac{\partial}{\partial \mathbf{y}^T} \cdot \frac{\partial}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{y}^T} \cdot \frac{\partial}{\partial \mathbf{y}} \end{bmatrix} f \cdot \begin{bmatrix} \mathbf{h} \\ \mathbf{k} \end{bmatrix}, \quad (6.25)$$

em notação compacta na notação utilizada como

$$\frac{\partial r}{\partial \mathbf{c}^T} = \frac{\partial f}{\partial \mathbf{c}^T} + \frac{\partial}{\partial \mathbf{c}^T} \cdot \frac{\partial}{\partial \mathbf{c}} f \cdot \mathbf{p} \quad (6.26)$$

onde $\mathbf{c}^T = [\mathbf{x} \ \mathbf{y}]$, $\partial/\partial \mathbf{c} = [\partial/\partial \mathbf{x} \ \partial/\partial \mathbf{y}]$ and $\mathbf{p}^T = [\mathbf{h} \ \mathbf{k}]$. Essa notação é equivalente (6.22) para o dobro de variáveis utilizadas, mas na forma (6.25) fica evidente cada vetor de variável utilizado. Ao calcular a solução para (6.26) é obtido o passo de busca para o ponto crítico, dado por

$$\mathbf{p} = - \left[\frac{\partial}{\partial \mathbf{c}^T} \cdot \frac{\partial}{\partial \mathbf{c}} f \right]^{-1} \cdot \frac{\partial f}{\partial \mathbf{c}}.$$

6.3.5 Expressão complexa aumentada

A partir da expressão real composta (6.24), é possível realizar a mudança de coordenadas para $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ e $\mathbf{z}^* = \mathbf{x} - j\mathbf{y}$. Substituindo o operador (3.21), a seguir

$$\frac{\partial f}{\partial \mathbf{z}} = \frac{1}{2} \left(\frac{\partial f}{\partial \mathbf{x}} - j \frac{\partial f}{\partial \mathbf{y}} \right) \text{ and } \frac{\partial f}{\partial \mathbf{z}^*} = \frac{1}{2} \left(\frac{\partial f}{\partial \mathbf{x}} + j \frac{\partial f}{\partial \mathbf{y}} \right),$$

na expressão de Taylor expressa em termos do par $(\mathbf{z}, \mathbf{z}^*)$, leva à representação em primeira ordem

$$r = f(\mathbf{x}, \mathbf{y}) + \frac{\partial f}{\partial \mathbf{x}} \cdot \mathbf{h} + \frac{\partial f}{\partial \mathbf{y}} \cdot \mathbf{k} = f(\mathbf{z}, \mathbf{z}^*) + \left(\frac{\partial f}{\partial \mathbf{z}} + \frac{\partial f}{\partial \mathbf{z}^*} \right) \cdot \mathbf{h} + \left(j \frac{\partial f}{\partial \mathbf{z}} - j \frac{\partial f}{\partial \mathbf{z}^*} \right) \cdot \mathbf{k}.$$

Agrupando os termos e utilizando o incremento complexo $\boldsymbol{\delta} = \mathbf{h} + j\mathbf{k}$ a expressão anterior pode ser posta em notação complexificada como

$$r = f(\mathbf{z}, \mathbf{z}^*) + \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^*.$$

Diferenciando a expressão anterior em relação ao operador conjugado coluna

$$\frac{\partial r}{\partial \mathbf{z}^H} = \frac{\partial f}{\partial \mathbf{z}^H} + \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^*, \quad (6.27)$$

também em relação ao operador não-conjugado

$$\frac{\partial r}{\partial \mathbf{z}^T} = \frac{\partial f}{\partial \mathbf{z}^T} + \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^*. \quad (6.28)$$

Note que as expressões (6.27) e (6.28) são equivalentes: uma é o conjugado da outra. Isso ocorre devido ao formato complexo aumentado, que relaciona pares reais com pares conjugados.

Agrupando essas expressões em um vetor composto e igualando ao vetor nulo, tem-se a forma matricial

$$\begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}^H} \\ \frac{\partial f}{\partial \mathbf{z}^T} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}^H} \\ \frac{\partial f}{\partial \mathbf{z}^T} \end{bmatrix} + \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial}{\partial \mathbf{z}} f & \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial}{\partial \mathbf{z}^*} f \\ \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial}{\partial \mathbf{z}} f & \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial}{\partial \mathbf{z}^*} f \end{bmatrix} \cdot \begin{bmatrix} \underline{\boldsymbol{\delta}} \\ \underline{\boldsymbol{\delta}}^* \end{bmatrix} = \mathbf{0} . \quad (6.29)$$

A forma acima foi escolhida pois, definindo $\underline{\boldsymbol{\delta}} = \begin{bmatrix} \boldsymbol{\delta}^T & \boldsymbol{\delta}^H \end{bmatrix}^T$ e o operador complexo aumentado como o vetor linha

$$\frac{\partial}{\partial \underline{\mathbf{z}}} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} & \frac{\partial}{\partial \mathbf{z}^*} \end{bmatrix} , \quad (6.30)$$

a equação em (6.29), que fornece o passo para otimização, pode ser posta em uma forma simplificada

$$\frac{\partial f}{\partial \underline{\mathbf{z}^H}} + \frac{\partial}{\partial \underline{\mathbf{z}^H}} \cdot \frac{\partial}{\partial \underline{\mathbf{z}}} f \cdot \underline{\boldsymbol{\delta}} = \mathbf{0} . \quad (6.31)$$

Outra forma de se escrever a expressão acima é tomar o conjugado, em cada termo, que fornece

$$\frac{\partial f}{\partial \underline{\mathbf{z}^T}} + \frac{\partial}{\partial \underline{\mathbf{z}^T}} \cdot \frac{\partial}{\partial \underline{\mathbf{z}^*}} f \cdot \underline{\boldsymbol{\delta}}^* = \mathbf{0} . \quad (6.32)$$

Com isso, fica evidente que derivada pode ser escrita em relação ao passo aumentado $\underline{\boldsymbol{\delta}}$ ou seu conjugado. Com as expressões anteriores é possível o cálculo do passo de otimização isolando-se $\underline{\boldsymbol{\delta}}$ em (6.31) ou $\underline{\boldsymbol{\delta}}^*$ em (6.32).

6.3.5.1 Exemplo de aplicação na função custo complexa.

Dada uma função custo como em (5.6) de valor real com variável vetorial complexa

$$\mathbf{z} = [z_1 \ \cdots \ z_N]^T \in \mathbb{C}^N$$

$$\mathcal{J} = \sigma^2 - \mathbf{p}^H \mathbf{z} - \mathbf{z}^H \mathbf{p} + \mathbf{z}^H \mathbf{R} \mathbf{z} ,$$

quer-se determinar qual regra terá o passo pelo método de Newton. Para isso, pela Tabela 1 a derivada não conjugada será

$$\frac{\partial \mathcal{J}}{\partial \mathbf{z}} = -\mathbf{p}^H + \mathbf{z}^H \mathbf{R} .$$

Pela mesma regra, a derivada do vetor linha acima em relação ao operador hermitiano elimina os valores que possuam tal variável, de forma que

$$\frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial}{\partial \mathbf{z}} \mathcal{J} = \mathbf{R} .$$

De maneira semelhante, para

$$\frac{\partial \mathcal{J}}{\partial \mathbf{z}^*} = -\mathbf{p}^T + \mathbf{z}^T \mathbf{R}^T ,$$

implica em

$$\frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial}{\partial \mathbf{z}^*} \mathcal{J} = 0 .$$

Portanto a expressão (6.27) para o ponto ótimo neste exemplo dado é

$$\frac{\partial f}{\partial \mathbf{z}^H} + \mathbf{R} \cdot \boldsymbol{\delta} = 0 \quad (6.33)$$

Em forma aumentada seria

$$\begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}^H} \\ \frac{\partial f}{\partial \mathbf{z}^T} \end{bmatrix} + \begin{bmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{R}^* \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\delta}^* \end{bmatrix} = 0 .$$

devido à propriedade de construção da matriz de correlação $\mathbf{R}^T = \mathbf{R}^*$. Devido a expressão ser diagonal, a expressão aumentada acima é meramente ilustrativa. A solução pode ser encontrada isolando-se o fator de passo em (6.33)

$$\boldsymbol{\delta} = -\mathbf{R}^{-1} \cdot \frac{\partial f}{\partial \mathbf{z}^H} .$$

Utilizando o fator de passo dado é possível escrever a regra de construção

$$\mathbf{z}_{n+1} = \mathbf{z}_n - \mathbf{R}^{-1} \cdot \frac{\partial f}{\partial \mathbf{z}^H} \Big|_{\mathbf{z}_n} \quad (6.34)$$

6.3.6 Relação entre as expressões real composta e a complexa aumentada

É visto nessa seção que, a relação entre as versões real composta (6.26) e as expressões complexas aumentadas (6.31) e (6.32), relacionam-se através da matriz de transformação entre variáveis (3.13). Essa relação se dá pela forma (6.30) e as derivada reais, como descreve a equação matricial

$$\begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{y}} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} & \frac{\partial}{\partial \mathbf{z}^*} \end{bmatrix} \cdot \underbrace{\begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix}}_{=\mathbf{T}} .$$

Calculando o transposto do hermitiano da expressão anterior,

$$\frac{\partial}{\partial \mathbf{c}^T} = \mathbf{T}^T \frac{\partial}{\partial \underline{\mathbf{z}}^T} \text{ e } \frac{\partial}{\partial \mathbf{c}^H} = \mathbf{T}^H \frac{\partial}{\partial \underline{\mathbf{z}}^H} , \quad (6.35)$$

com

$$\frac{\partial}{\partial \underline{\mathbf{c}}^T} = \frac{\partial}{\partial \underline{\mathbf{c}}^H}.$$

A partir da expressão real (6.26), usando a propriedade relacionada à hermitiana em (6.35),

$$\mathbf{T}^H \cdot \left(\frac{\partial f}{\partial \underline{\mathbf{z}}^H} + \frac{\partial}{\partial \underline{\mathbf{z}}^T} \cdot \frac{\partial}{\partial \underline{\mathbf{z}}} f \cdot \mathbf{T} \cdot \mathbf{p} \right) = \mathbf{0}, \quad (6.36)$$

com $\mathbf{T} \cdot \mathbf{p} = \underline{\boldsymbol{\delta}}$. O termo dentro do parêntesis é exatamente a expressão em (6.31), confirmando assim a relação entre a forma real e a obtida pela derivada hermitiana. O caso transposto pode ser obtido de forma semelhante, também a partir de (6.26) com auxílio de (6.35)

$$\mathbf{T}^T \cdot \left[\frac{\partial f}{\partial \underline{\mathbf{z}}^T} + \frac{\partial}{\partial \underline{\mathbf{z}}^T} \cdot \left(\frac{\partial}{\partial \underline{\mathbf{z}}} f \cdot \mathbf{T} \right)^* \cdot \mathbf{p} \right] = \mathbf{0}. \quad (6.37)$$

A igualdade

$$\left(\frac{\partial}{\partial \underline{\mathbf{z}}} f \cdot \mathbf{T} \right)^* = \left(\frac{\partial}{\partial \underline{\mathbf{c}}} f \right)^* = \frac{\partial}{\partial \underline{\mathbf{c}}} f$$

dentro da expressão (6.37) é possível pela natureza real da forma composta. Reescrevendo então (6.37) com a observação anterior

$$\mathbf{T}^T \cdot \left(\frac{\partial f}{\partial \underline{\mathbf{z}}^T} + \frac{\partial}{\partial \underline{\mathbf{z}}^T} \cdot \frac{\partial}{\partial \underline{\mathbf{z}}^*} f \cdot \mathbf{T}^* \cdot \mathbf{p} \right) = \mathbf{0}. \quad (6.38)$$

para $\mathbf{T}^* \cdot \mathbf{p} = \underline{\boldsymbol{\delta}}^*$. O termo dentro do parêntesis representa a equação (6.32), dessa vez evidenciando a relação entre a forma real e a obtida pela derivada transposta.

6.4 Método para expressões complexas compostas

As expressões desenvolvidas até aqui são uteis para lidar com funções do tipo $f(\mathbf{z}, \mathbf{z}^*)$, como é o caso da função custo de filtros complexos não amplamente-lineares. Para expressões aumentadas com o dobro de variáveis, o procedimento é semelhante ao que foi feito anteriormente. Pode ser atribuído o dobro de dimensão às variáveis já utilizadas, e cada uma dessas “dividida” ao meio em duas novas variáveis. Também pode-se considerar construir o mesmo método para o dobro de variáveis em cada expressão, como foi calculado na seção anterior. Isso será importante para a expressão do custo de filtros amplamente lineares de variáveis vetoriais $\mathbf{z}, \mathbf{w} \in \mathbb{C}^N$, do tipo $f(\mathbf{z}, \mathbf{w}, \mathbf{z}^*, \mathbf{w}^*) = f(\mathbf{z}_c, \mathbf{z}_c^*)$, para $\mathbf{z}_c = [\mathbf{z}^T \ \mathbf{w}^T]^T$.

Partindo da expressão real para a aproximação de primeira ordem

$$r = f(\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v}) + \frac{\partial f}{\partial \mathbf{x}} \cdot \mathbf{h} + \frac{\partial f}{\partial \mathbf{y}} \cdot \mathbf{k} + \frac{\partial f}{\partial \mathbf{u}} \cdot \mathbf{m} + \frac{\partial f}{\partial \mathbf{v}} \cdot \mathbf{n}.$$

para $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} \in \mathbf{R}^n$. Tomando as derivadas e agrupando as variáveis como $\tilde{\mathbf{x}} = (\mathbf{x}, \mathbf{y})$ e $\tilde{\mathbf{u}} = (\mathbf{u}, \mathbf{v})$, bem como os fatores de passo como $\tilde{\mathbf{h}} = (\mathbf{h}, \mathbf{k})$ e $\tilde{\mathbf{m}} = (\mathbf{m}, \mathbf{n})$,

$$\begin{bmatrix} \frac{\partial r}{\partial \tilde{\mathbf{x}}^T} \\ \frac{\partial r}{\partial \tilde{\mathbf{u}}^T} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial \tilde{\mathbf{x}}^T} \\ \frac{\partial f}{\partial \tilde{\mathbf{u}}^T} \end{bmatrix} + \begin{bmatrix} \frac{\partial}{\partial \tilde{\mathbf{x}}^T} \frac{\partial}{\partial \tilde{\mathbf{x}}} f & \frac{\partial}{\partial \tilde{\mathbf{x}}^T} \frac{\partial}{\partial \tilde{\mathbf{u}}} f \\ \frac{\partial}{\partial \tilde{\mathbf{u}}^T} \frac{\partial}{\partial \tilde{\mathbf{x}}} f & \frac{\partial}{\partial \tilde{\mathbf{u}}^T} \frac{\partial}{\partial \tilde{\mathbf{u}}} f \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{h}} \\ \tilde{\mathbf{m}} \end{bmatrix}.$$

Essa expressão é exatamente a descrita em (6.25). Para o caso complexificado pela mudança de variáveis dada por

$$\frac{\partial f}{\partial \mathbf{z}} = \frac{1}{2} \left(\frac{\partial f}{\partial \mathbf{x}} - j \frac{\partial f}{\partial \mathbf{y}} \right) \text{ e } \frac{\partial f}{\partial \mathbf{z}^*} = \frac{1}{2} \left(\frac{\partial f}{\partial \mathbf{x}} + j \frac{\partial f}{\partial \mathbf{y}} \right),$$

para o primeiro conjunto e

$$\frac{\partial f}{\partial \mathbf{w}} = \frac{1}{2} \left(\frac{\partial f}{\partial \mathbf{u}} - j \frac{\partial f}{\partial \mathbf{v}} \right) \text{ and } \frac{\partial f}{\partial \mathbf{w}^*} = \frac{1}{2} \left(\frac{\partial f}{\partial \mathbf{u}} + j \frac{\partial f}{\partial \mathbf{v}} \right),$$

para o segundo. A aproximação linear pode ser reescrita como

$$r = f + \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^* + \frac{\partial f}{\partial \mathbf{w}} \cdot \boldsymbol{\epsilon} + \frac{\partial f}{\partial \mathbf{w}^*} \cdot \boldsymbol{\epsilon}^*,$$

com $\boldsymbol{\delta} = \mathbf{h} + j\mathbf{k}$ e $\boldsymbol{\epsilon} = \mathbf{m} + j\mathbf{n}$. Calculando as derivadas para cada variável e igualando todas ao vetor nulo

$$\begin{aligned} \frac{\partial r}{\partial \mathbf{z}^T} &= \frac{\partial f}{\partial \mathbf{z}^T} + \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^* + \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial f}{\partial \mathbf{w}} \cdot \boldsymbol{\epsilon} + \frac{\partial}{\partial \mathbf{z}^T} \cdot \frac{\partial f}{\partial \mathbf{w}^*} \cdot \boldsymbol{\epsilon}^* = \mathbf{0} \\ \frac{\partial r}{\partial \mathbf{z}^H} &= \frac{\partial f}{\partial \mathbf{z}^H} + \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^* + \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial f}{\partial \mathbf{w}} \cdot \boldsymbol{\epsilon} + \frac{\partial}{\partial \mathbf{z}^H} \cdot \frac{\partial f}{\partial \mathbf{w}^*} \cdot \boldsymbol{\epsilon}^* = \mathbf{0} \\ \frac{\partial r}{\partial \mathbf{w}^T} &= \frac{\partial f}{\partial \mathbf{w}^T} + \frac{\partial}{\partial \mathbf{w}^T} \cdot \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial}{\partial \mathbf{w}^T} \cdot \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^* + \frac{\partial}{\partial \mathbf{w}^T} \cdot \frac{\partial f}{\partial \mathbf{w}} \cdot \boldsymbol{\epsilon} + \frac{\partial}{\partial \mathbf{w}^T} \cdot \frac{\partial f}{\partial \mathbf{w}^*} \cdot \boldsymbol{\epsilon}^* = \mathbf{0} \\ \frac{\partial r}{\partial \mathbf{w}^H} &= \frac{\partial f}{\partial \mathbf{w}^H} + \frac{\partial}{\partial \mathbf{w}^H} \cdot \frac{\partial f}{\partial \mathbf{z}} \cdot \boldsymbol{\delta} + \frac{\partial}{\partial \mathbf{w}^H} \cdot \frac{\partial f}{\partial \mathbf{z}^*} \cdot \boldsymbol{\delta}^* + \frac{\partial}{\partial \mathbf{w}^H} \cdot \frac{\partial f}{\partial \mathbf{w}} \cdot \boldsymbol{\epsilon} + \frac{\partial}{\partial \mathbf{w}^H} \cdot \frac{\partial f}{\partial \mathbf{w}^*} \cdot \boldsymbol{\epsilon}^* = \mathbf{0}, \end{aligned}$$

e reorganizando tudo em notação matricial

$$\begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}^H} \\ \frac{\partial f}{\partial \mathbf{w}^H} \\ \frac{\partial f}{\partial \mathbf{z}^T} \\ \frac{\partial f}{\partial \mathbf{w}^T} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}^H} \\ \frac{\partial}{\partial \mathbf{w}^H} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} & \frac{\partial}{\partial \mathbf{w}} \end{bmatrix} \\ \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}^T} \\ \frac{\partial}{\partial \mathbf{w}^T} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} & \frac{\partial}{\partial \mathbf{w}} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}^*} \\ \frac{\partial}{\partial \mathbf{w}^*} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}^*} & \frac{\partial}{\partial \mathbf{w}^*} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\epsilon} \\ \boldsymbol{\delta}^* \\ \boldsymbol{\epsilon}^* \end{bmatrix} = \mathbf{0}.$$

A expressão pode ser escrita de maneira curta atribuindo às variáveis um formato composto complexo $\mathbf{z}_c = (\mathbf{z}, \mathbf{w})$, bem como ao fator de passo $\boldsymbol{\delta}_c = (\boldsymbol{\delta}, \boldsymbol{\varepsilon})$, de forma que a derivada é dada por

$$\frac{\partial}{\partial \mathbf{z}_c} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} & \frac{\partial}{\partial \mathbf{w}} \end{bmatrix}$$

e a expressão composta é

$$\begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}_c^H} \\ \frac{\partial f}{\partial \mathbf{z}_c^T} \end{bmatrix} + \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}_c^H} \cdot \frac{\partial}{\partial \mathbf{z}_c} & \frac{\partial}{\partial \mathbf{z}_c^H} \cdot \frac{\partial}{\partial \mathbf{z}_c^*} \\ \frac{\partial}{\partial \mathbf{z}_c^T} \cdot \frac{\partial}{\partial \mathbf{z}_c} & \frac{\partial}{\partial \mathbf{z}_c^T} \cdot \frac{\partial}{\partial \mathbf{z}_c^*} \end{bmatrix} f \cdot \begin{bmatrix} \boldsymbol{\delta}_a \\ \boldsymbol{\delta}_a^* \end{bmatrix} = \mathbf{0} . \quad (6.39)$$

Note que o resultado é o mesmo para (6.29), como se espera caso as variáveis sejam agrupadas da forma proposta. Outras formas de escrita para a expressão acima são possíveis, mas essa a conveniência desta é mostrada no exemplo a seguir.

6.4.0.1 Exemplo de aplicação na função custo complexa amplamente linear.

De forma semelhante ao que já foi feito, para a função custo amplamente linear (5.24) equivalente à expressão

$$\mathcal{J} = \sigma^2 - \mathbf{p}_c^H \mathbf{z}_c - \mathbf{z}_c^H \mathbf{p}_c + \mathbf{z}_c^H \mathbf{R} \mathbf{z}_c , \quad (6.40)$$

para $\mathbf{R}$ a matriz de correlação aumentada. Tem-se então as derivadas

$$\frac{\partial \mathcal{J}}{\partial \mathbf{z}_c} = -\mathbf{p}_c^H + \mathbf{z}_c^H \mathbf{R} ,$$

$$\frac{\partial}{\partial \mathbf{z}_c^H} \cdot \frac{\partial}{\partial \mathbf{z}_c} \mathcal{J} = \mathbf{R}$$

e

$$\frac{\partial}{\partial \mathbf{z}_c^H} \cdot \frac{\partial}{\partial \mathbf{z}_c^*} \mathcal{J} = 0 .$$

A expressão da primeira linha em (6.39) é

$$\frac{\partial f}{\partial \mathbf{z}_c^H} + \mathbf{R} \cdot \boldsymbol{\delta}_c = 0 ,$$

fornecendo o passo composto

$$\boldsymbol{\delta}_c = -\mathbf{R}^{-1} \cdot \frac{\partial f}{\partial \mathbf{z}_c^H} .$$

O passo fornece a aproximação dada por

$$\mathbf{z}_c[n+1] = \mathbf{z}_c[n] - \underline{\mathbf{R}}^{-1} \cdot \left. \frac{\partial f}{\partial \mathbf{z}_c^H} \right|_{\mathbf{z}_c[n]},$$

em forma matricial expandida

$$\begin{bmatrix} \mathbf{z}[n+1] \\ \mathbf{w}[n+1] \end{bmatrix} = \begin{bmatrix} \mathbf{z}[n] \\ \mathbf{w}[n] \end{bmatrix} - \begin{bmatrix} \mathbf{R} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}}^* & \mathbf{R}^* \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}^H} \\ \frac{\partial f}{\partial \mathbf{w}^H} \end{bmatrix}. \quad (6.41)$$

Note que aqui há a necessidade do cálculo da matriz inversa em quaisquer dos casos, tornando o algoritmo inviável em termos de um bom desempenho. A próxima seção apresenta uma solução engenhosa para o cálculo da matriz inversa para o caso da correlação (quando o hessiano é a função custo complexa ou complexa aumentada).

6.5 Estimação da matriz inversa

O maior problema do método de Newton, e suas “versões” descritas até aqui, é a presença da inversão do hessiano (matriz de correlação no caso destas aplicadas às funções custo dos filtros vistos anteriormente). Esse problema é semelhante ao que ocorre na busca dos coeficientes ótimos pelo método de Wiener. Para isso, nessa seção será apresentado o método iterativo que estima a inversa da matriz de correlação aumentada, ponto a ponto, para aproximações aplicadas às funções custo (onde o hessiano é uma matriz hermitiana devido à forma quadrática associada).

Considere a matriz que representa a correlação aumentada

$$\underline{\mathbf{R}} = \begin{bmatrix} \mathbf{R} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}}^* & \mathbf{R}^* \end{bmatrix},$$

lembrando que $\mathbf{R}$ é a autocorrelação e $\tilde{\mathbf{R}}$ é a pseudo-autocorrelação. A inversa pode ser calculada de maneira exata por (Schreier; Scharf, 2010)

$$\underline{\mathbf{R}}^{-1} = \begin{bmatrix} (\mathbf{R} - \tilde{\mathbf{R}}(\mathbf{R}^*)^{-1}\tilde{\mathbf{R}}^*)^{-1} & -(\mathbf{R} - \tilde{\mathbf{R}}(\mathbf{R}^*)^{-1}\tilde{\mathbf{R}}^*)^{-1}\tilde{\mathbf{R}}(\mathbf{R}^*)^{-1} \\ -(\mathbf{R}^* - \tilde{\mathbf{R}}^*\mathbf{R}^{-1}\tilde{\mathbf{R}})^{-1}\tilde{\mathbf{R}}^*\mathbf{R}^{-1} & (\mathbf{R}^* - \tilde{\mathbf{R}}^*\mathbf{R}^{-1}\tilde{\mathbf{R}})^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{N}^* & \mathbf{M}^* \end{bmatrix} = \underline{\mathbf{M}} \quad (6.42)$$

que possui a mesma estrutura aumentada presente em (3.14). Essa deve ser a matriz a ser calculada caso se queira uma solução exata.

6.5.1 O lema de inversão de matrizes para a autocorrelação aumentada

É possível atribuir, a cada instante discreto, um valor estimado para a inversa da matriz de correlação aumentada. Para isso seja a estimativa da matriz a expressão de diferenças dada por

$$\underline{\mathbf{R}}[n+1] = \lambda \underline{\mathbf{R}}[n] + \underline{\mathbf{u}}[n+1] \underline{\mathbf{u}}^H[n+1] \quad (6.43)$$

onde $0 < \lambda \leq 1$ é chamado de fator de esquecimento. Para estimar a inversa em cada instante é utilizada a forma (6.43), utilizando o chamado lema de inversão de matrizes. O tema é desenvolvido da seguinte forma: Sejam $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ e $\mathbf{D}$ matrizes da mesma ordem que $\underline{\mathbf{R}}$, a expressão dada por

$$\mathbf{A} = \mathbf{B} + \mathbf{C} \mathbf{D} \mathbf{C}^H$$

tem inversa dada por

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1} \mathbf{C} [\mathbf{C}^H \mathbf{B}^{-1} \mathbf{C} + \mathbf{D}^{-1}]^{-1} \mathbf{C}^H \mathbf{B}^{-1}. \quad (6.44)$$

Com isso, a partir das mudanças de variáveis para

$$\begin{cases} \mathbf{A} = \underline{\mathbf{R}}[n+1] \\ \mathbf{B} = \lambda \underline{\mathbf{R}}[n] \\ \mathbf{C} = \underline{\mathbf{u}}[n+1] \\ \mathbf{D} = 1 \end{cases}, \quad (6.45)$$

é possível reescrever (6.43) por

$$\underline{\mathbf{R}}^{-1}[n+1] = \frac{1}{\lambda} \left[\underline{\mathbf{R}}^{-1}[n] - \frac{\underline{\mathbf{R}}^{-1}[n] \underline{\mathbf{u}}[n+1] \underline{\mathbf{u}}^H[n+1] \underline{\mathbf{R}}^{-1}[n]}{\lambda + \underline{\mathbf{u}}^H[n+1] \underline{\mathbf{R}}^{-1}[n] \underline{\mathbf{u}}[n+1]} \right] \quad (6.46)$$

para um valor inicial arbitrário de $\underline{\mathbf{R}}^{-1}[0]$ (como sugestão pode-se utilizar a identidade). Essa forma é a mesma que é utilizada no caso complexo usual, mas com as devidas trocas de matrizes e vetores para o caso aumentado. De forma explícita, renomeando os valores $\mathbf{r}[n+1] = \underline{\mathbf{u}}[n+1] \underline{\mathbf{u}}^H[n+1]$ e $\tilde{\mathbf{r}}[n+1] = \underline{\mathbf{u}}[n+1] \underline{\mathbf{u}}^T[n+1]$, é possível escrever a inversa da correlação aumentada como:

$$\underline{\mathbf{R}}^{-1}[n+1] = \begin{bmatrix} \mathbf{M}[n+1] & \mathbf{N}[n+1] \\ \mathbf{N}^*[n+1] & \mathbf{M}^*[n+1] \end{bmatrix}$$

$$= \frac{1}{\lambda} \left\{ \begin{bmatrix} \mathbf{M}[n] & \mathbf{N}[n] \\ \mathbf{N}^*[n] & \mathbf{M}^*[n] \end{bmatrix} - \frac{\begin{bmatrix} \mathbf{M}[n] & \mathbf{N}[n] \\ \mathbf{N}^*[n] & \mathbf{M}^*[n] \end{bmatrix} \begin{bmatrix} \mathbf{r}[n+1] & \tilde{\mathbf{r}}[n+1] \\ \tilde{\mathbf{r}}^*[n+1] & \mathbf{r}^*[n+1] \end{bmatrix} \begin{bmatrix} \mathbf{M}[n] & \mathbf{N}[n] \\ \mathbf{N}^*[n] & \mathbf{M}^*[n] \end{bmatrix}}{\lambda + \begin{bmatrix} \mathbf{u}^H[n+1] & \mathbf{u}^T[n+1] \end{bmatrix} \begin{bmatrix} \mathbf{M}[n] & \mathbf{N}[n] \\ \mathbf{N}^*[n] & \mathbf{M}^*[n] \end{bmatrix} \begin{bmatrix} \mathbf{u}[n+1] \\ \mathbf{u}^*[n+1] \end{bmatrix}} \end{bmatrix} \right\}.$$

Os termos (submatrizes) presentes da matriz inversa em (6.42) são

$$\mathbf{M}[n+1] = \frac{1}{\lambda} \left\{ \mathbf{M}[n] + \frac{1}{t[n]} \left(\mathbf{M}[n] \mathbf{r}[n+1] \mathbf{M}[n] + \mathbf{M}[n] \tilde{\mathbf{r}}[n+1] \mathbf{N}^*[n] \right. \right. \\ \left. \left. + \mathbf{N}[n] \tilde{\mathbf{r}}^*[n+1] \mathbf{M}[n] + \mathbf{N}[n] \mathbf{r}^*[n+1] \mathbf{N}^*[n] \right) \right\}$$

e

$$\mathbf{N}[n+1] = \frac{1}{\lambda} \left\{ \mathbf{N}[n] + \frac{1}{t[n]} \left(\mathbf{N}[n] \tilde{\mathbf{r}}^*[n+1] \mathbf{N}[n] + \mathbf{N}[n] \mathbf{r}^*[n+1] \mathbf{M}^*[n] \right. \right. \\ \left. \left. + \mathbf{M}[n] \mathbf{r}[n+1] \mathbf{N}[n] + \mathbf{M}[n] \tilde{\mathbf{r}}[n+1] \mathbf{M}^*[n] \right) \right\},$$

para os valores iniciais atribuídos à $\mathbf{M}[0]$ e $\mathbf{N}[0]$, e com o termo real auxiliar

$$t[n] = \lambda + 2\text{Re} \left\{ \mathbf{u}^H[n+1] \mathbf{M}[n] \mathbf{u}[n+1] + \mathbf{u}^H[n+1] \mathbf{N}[n] \mathbf{u}^*[n+1] \right\}.$$

A expressão (6.46) é a estimativa da autocorrelação aumentada e é uma estimativa instantânea discreta, usada de forma iterativa. Uma expressão para a matriz de correlação $\mathbf{R}$ (não aumentada) pode ser feita da mesma forma, basta reutilizar o lema da inversão de matrizes do início ou tomar $\tilde{\mathbf{R}} = 0$ na forma aumentada do lema.

6.5.2 Método a partir do ponto ótimo

É possível a obtenção do método de Newton, já considerando a função custo, através dos coeficientes ótimos compostos,

$$\mathbf{w}_c = \underline{\mathbf{R}}^{-1} \mathbf{p}_c, \quad (6.47)$$

isso é possível, mas particulariza a solução do método para a função custo. A aproximação para o valor da correlação cruzada composta será

$$\mathbf{p}_c[n+1] = \lambda \mathbf{p}_c[n] + \underline{\mathbf{u}}[n+1] d^*[n+1] \quad (6.48)$$

de forma semelhante ao que foi feito para a correlação em (6.43). Aplicando a relação ótima, no instante posterior $\mathbf{w}_c[n+1] = \underline{\mathbf{R}}^{-1}[n+1] \mathbf{p}_c[n+1]$ tem-se a expressão:

$$\mathbf{w}_c[n+1] = \mathbf{w}_c[n] + \underline{\mathbf{R}}^{-1}[n+1] \underline{\mathbf{u}}[n+1] e_p^*[n+1], \quad (6.49)$$

para a estimativa de $\underline{\mathbf{R}}^{-1}[n+1]$ dada pelo lema da inversão de matrizes (6.46) e o erro conjugado aumentado escalar dito *a priori*

$$e_p^*[n+1] = d^*[n+1] - \mathbf{u}^H[n+1]\mathbf{w}[n] - \mathbf{u}^T[n+1]\mathbf{h}[n] .$$

A estimativa da correlação também pode ser utilizada diretamente na expressão (6.41), pois o gradiente pode ser estimado como o produto do sinal pelo erro. Essa seção somente apresenta a dedução simplificada para o caso particular amplamente linear.

6.6 Realizações com o algoritmo de Newton

O algoritmo de Newton aumentado, analisado nas seções anteriores utilizando o lema de inversão de matrizes aumentado (6.46) para aproximação da inversa da correlação, é mostrado no Algoritmo 3 a seguir. Nele a cada iteração a matriz é estimada, com base num dado valor inicial, servindo como parâmetro multiplicativo do gradiente. Lembrando que esse algoritmo é válido para filtros amplamente lineares como mostrado na Figura 21. A forma é similar àquela

Algoritmo 3: Método de Newton aumentado

Entrada: u , d , e o número de coeficientes M .

início

Seja Q = comprimento total do sinal d ou u .

Inicialize $\mathbf{w}_c$ comprimento $2M$ e $n = 0$

Inicialize $\mathbf{e} = [e[0], \dots, e[Q-1]]^T = 0$ de tamanho Q .

Tome $0 < \lambda < 1$, e inicialize $\underline{\mathbf{R}}^{-1}$.

para $M-1 \leq n \leq Q-1$ **faça**

$\underline{\mathbf{u}} = [u[n], \dots, u[n-M+1] \ u^*[n], \dots, u^*[n-M+1]]^T$

$\underline{\mathbf{R}}^{-1} \leftarrow \frac{1}{\lambda} \left[\underline{\mathbf{R}}^{-1} - \frac{\underline{\mathbf{R}}^{-1} \underline{\mathbf{u}} \underline{\mathbf{u}}^H \underline{\mathbf{R}}^{-1}}{\lambda + \underline{\mathbf{u}}^H \underline{\mathbf{R}}^{-1} \underline{\mathbf{u}}} \right]$

$e = d[n] - \mathbf{w}_c^H \underline{\mathbf{u}}$

$\mathbf{w}_c \leftarrow \mathbf{w}_c + \underline{\mathbf{R}}^{-1} \underline{\mathbf{u}} e^*$

$n \leftarrow n + 1$

fim

fim

apresentada pelo complexo usual, que no caso seriam com as matrizes de correlação, sinal e coeficientes não aumentados, como mostra o Algoritmo 4. Esta seção apresentará alguns exemplos de desempenho do método de Newton aumentado comparado à sua versão usual a seguir.

Algoritmo 4: Método de Newton linear

Entrada: u , d , e o número de coeficientes M .

início

Seja Q = comprimento total do sinal d ou u .

Inicialize $\mathbf{w}$ comprimento M e $n = 0$

Inicialize $\mathbf{e} = [e[0], \dots, e[Q-1]]^T = 0$ de tamanho Q .

Tome $0 < \lambda < 1$, e inicialize $\mathbf{R}^{-1}$.

para $M-1 \leq n \leq Q-1$ **faça**

$\mathbf{u} = [u[n], \dots, u[n-M+1]]^T$

$\mathbf{R}^{-1} \leftarrow \frac{1}{\lambda} \left[\mathbf{R}^{-1} - \frac{\mathbf{R}^{-1} \mathbf{u} \mathbf{u}^H \mathbf{R}^{-1}}{\lambda + \mathbf{u}^H \mathbf{R}^{-1} \mathbf{u}} \right]$

$e = d[n] - \mathbf{w}^H \mathbf{u}$

$\mathbf{w} \leftarrow \mathbf{w} + \mathbf{R}^{-1} \mathbf{u} e^*$

$n \leftarrow n + 1$

fim

fim

6.6.1 Sinal aleatório mais ruído

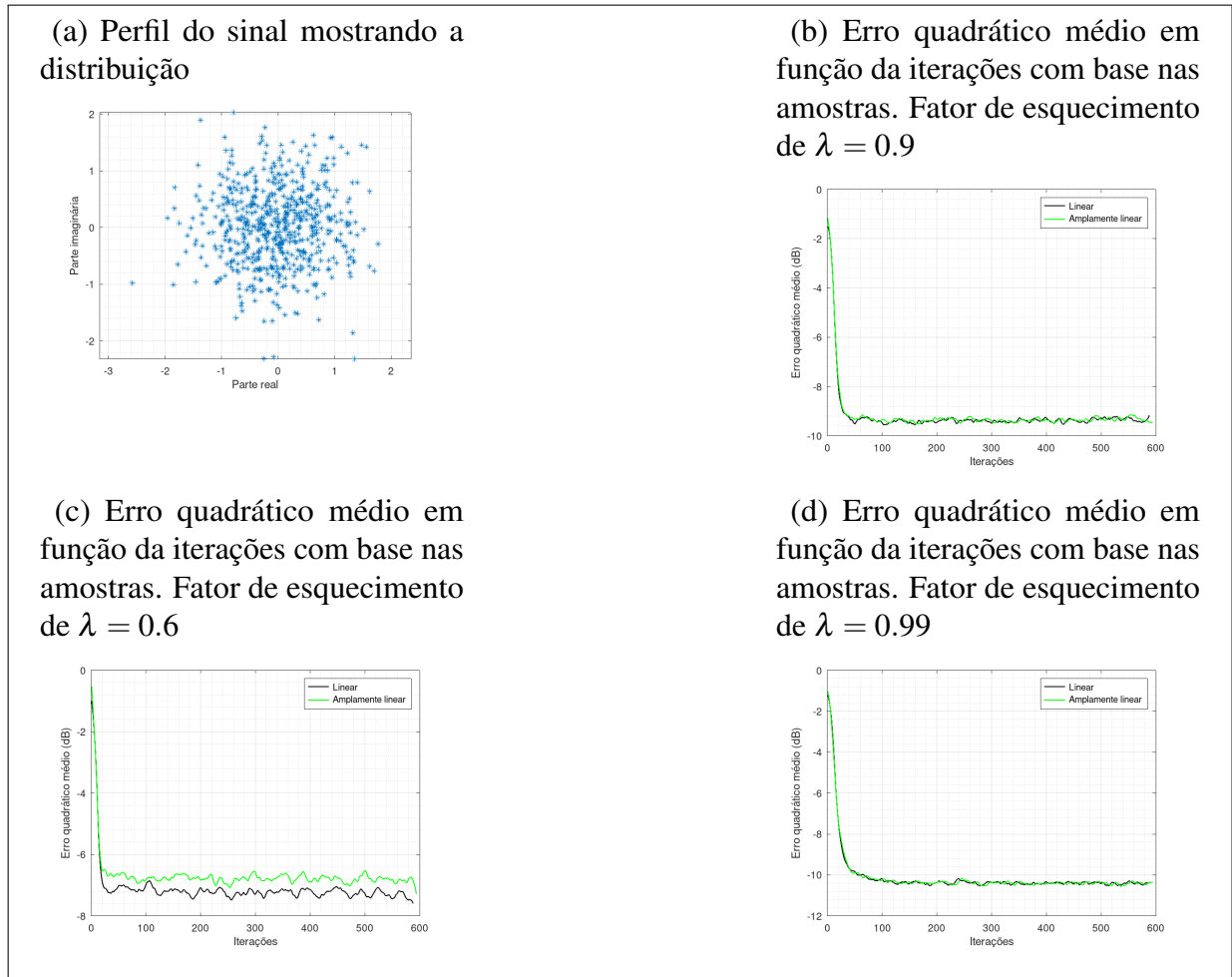
O primeiro caso é de uma sinal com ruído gaussiano modelado de forma que o sinal desejado é dado pelo mesmo modelo já utilizado em (5.31), referente à equação (4.9). Da mesma forma, o ruído de acordo com a expressão (5.32), com a presença de um fator divisor de SNR (relação sinal ruído linear).

A métrica utilizada foi também a do erro quadrático médio das amostras. Nela a média foi estimada por meio de uma recursão de Monte Carlo. Os tamanhos de filtro padrão foram: $M_{\text{linear}} = 12$ para o complexo usual e $M_{\text{aumentado}}$ para o caso aumentado ($M_{\text{individual}} = 6$), exceto quando o exemplo deixar explícito. Em todos os casos o valor de do sinal em relação ao ruído é dado por $SNR = 10$.

O primeiro exemplo de um sinal circular com ruído circular é mostrado na Figura 29, que apresenta a performance em relação à escolha do fator de esquecimento λ . Nesse caso os algoritmos complexo usual e aumentado tem performance similar para valores de λ próximos de 0,9. O método aumentado é mais sensível à escolha desse fator, apresentando maior erro com a diminuição deste em comparação com o usual. Um fator de esquecimento menor também provoca mais flutuações do erro quadrático médio.

O segundo caso é um sinal circular e ruído não circular, com $\rho_N = 0.2$ e rotação de $\varphi_N = \pi/4$, como ilustrado na Figura 30 (a). Esse exemplo exhibe variações nos tamanhos

Figura 29 – Método de Newton para o sinal circular e ruído circular com o perfil mostrado em (a). Nessa simulação $\rho_s = \rho_N = 1/\sqrt{2}$ (circular) e $\varphi_s = \varphi_N = 0$. Em (b) $\lambda = 0.9$ (valor usual), (c) $\lambda = 0.6$ e d) $\lambda = 0.99$



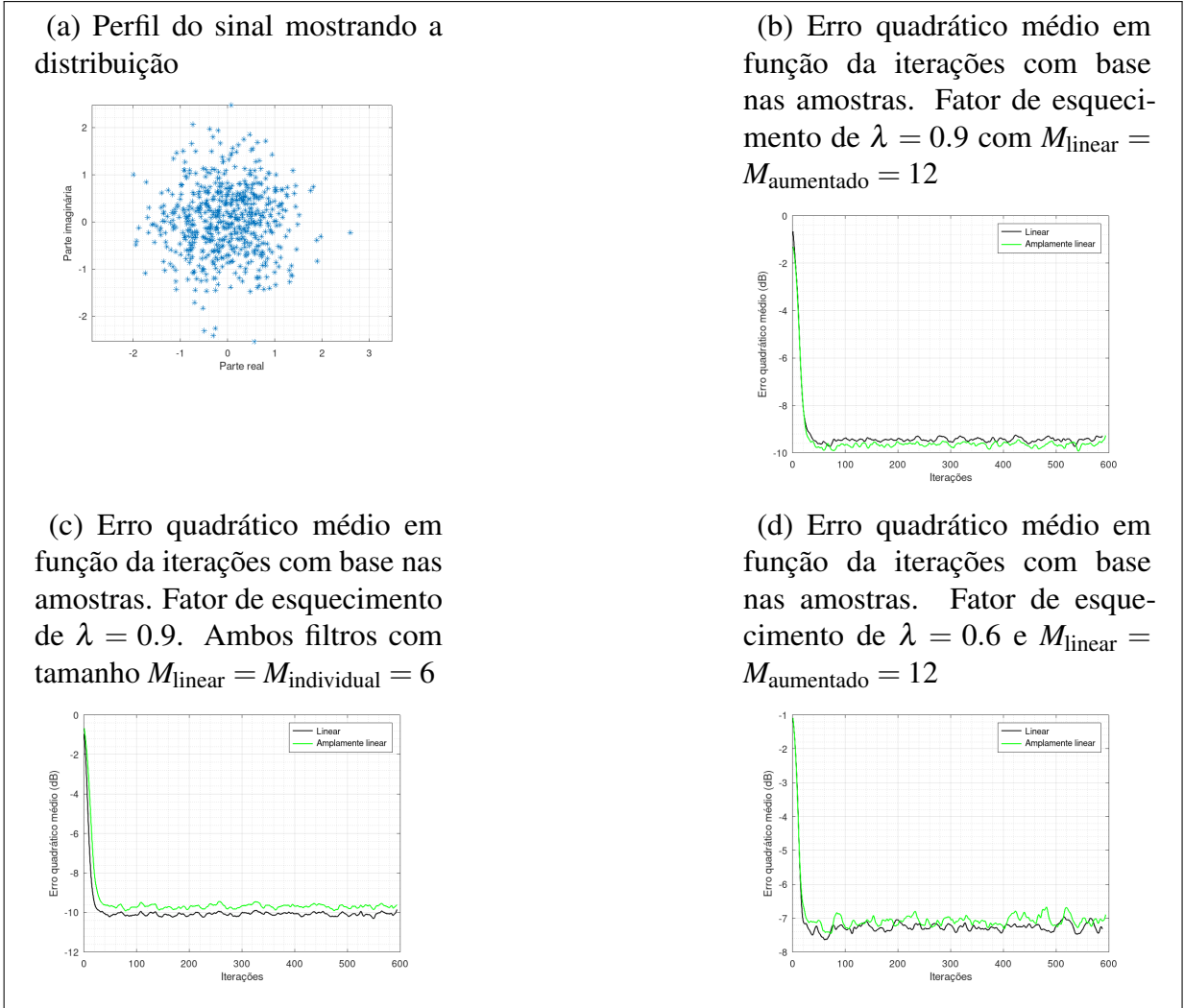
Fonte: Elaborada pelo autor.

do filtro aumentado. Nesse caso há um pequeno ganho na aproximação por usar o algoritmo aumentado a Figura 30 (b). Na Figura 30 (c) e (d) foram utilizados tamanhos de filtros diferentes para comparação.

6.6.2 Exemplos com sinais 4QAM

Da mesma forma que no caso do gradiente aumentado, foi testado o sinal do tipo 4QAM mais ruído dado por (5.32). Os resultados também são similares, como mostra a Figura 31. Em todos os casos, não há ganho em se utilizar o algoritmo aumentado para sinais circulares, dado um mesmo número de coeficientes em ambos os filtros (o que era esperado, vista que a matriz de pseudo-correlação é nula neste caso). No algoritmo de Newton não há, praticamente, perda de velocidade de convergência em comparação com o caso linear, indicando uma pequena

Figura 30 – Método de Newton para o sinal circular e ruído não circular com o perfil mostrado em (a). Nessa simulação $\rho_n = 0.2$ (não-circular) e $\varphi_N = \pi/4$

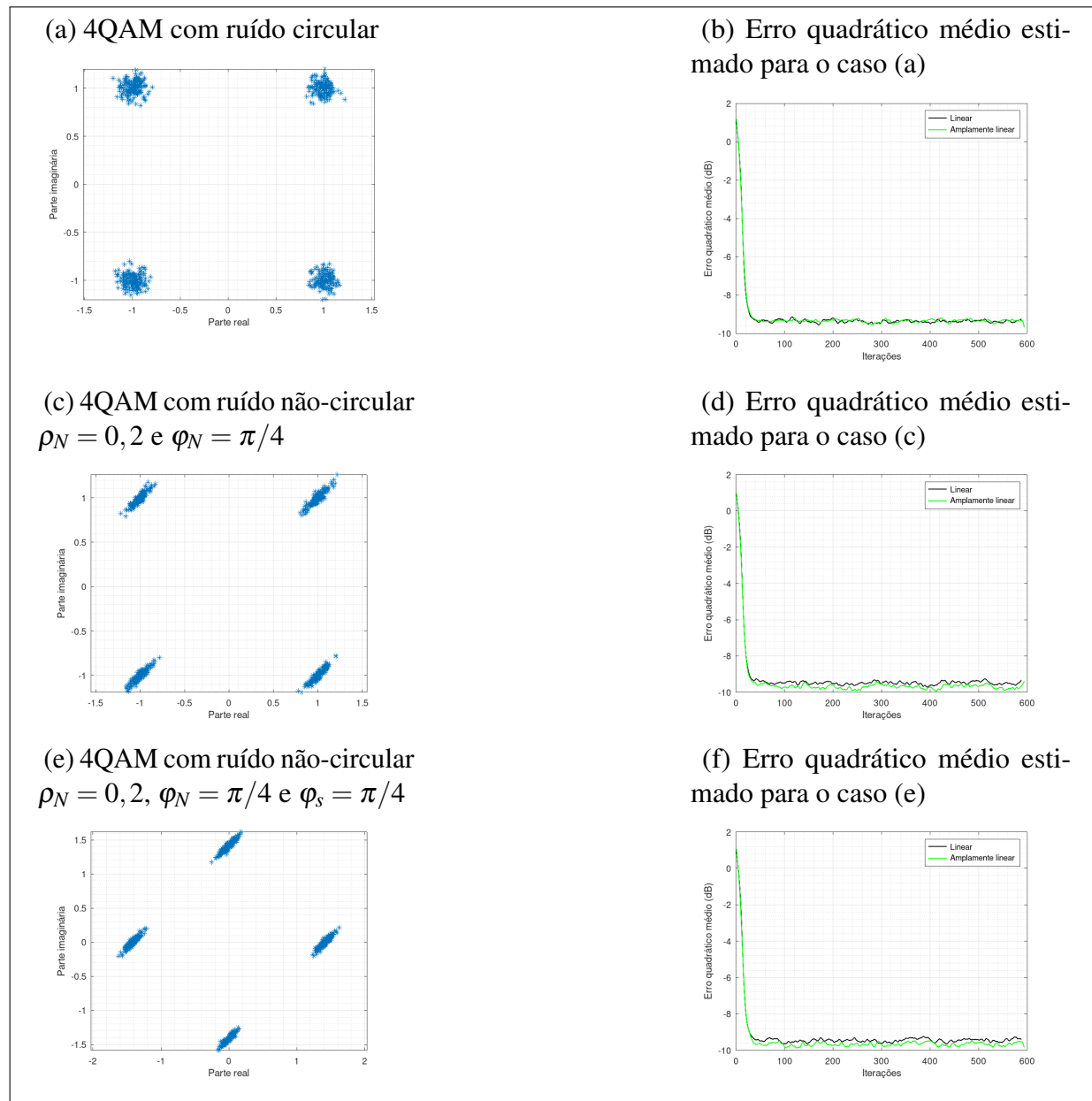


Fonte: Elaborada pelo autor.

melhora (como visto na Figura 31 nos casos (d) e (f)) ou, no pior dos casos, equivalência como na mesma figura em (b).

Este capítulo conclui a análise por meio de testes simulados desses algoritmos. Com isso, foi possível estudar o algoritmo de Newton em sua forma complexa e também em sua versão aumentada. No próximo capítulo, abordaremos o algoritmo natural em sua forma aumentada, explorando como ele pode ser expresso na forma complexa.

Figura 31 – Realizações dos algoritmos para o caso 4QAM



Fonte: Elaborada pelo autor.

7 GRADIENTE NATURAL COMPLEXO

Este capítulo apresenta uma extensão dos espaços tangentes reais empregados para obter uma expressão generalizada do gradiente complexo natural. É proposta uma expressão válida para um sistema de coordenadas generalizado numa base complexa. A equivalência com a representação real é feita através da estrutura complexa que os espaços reais podem possuir, permitindo a mudança da base utilizada. Exemplos típicos de aplicações são apresentados, ilustrando a abordagem proposta.

É sempre possível manter todos os cálculos em notação real, tratando valores complexos como valores reais bidimensionais e depois usar uma métrica real, como em (Nitta, 2014), mas torna a representação matemática menos natural e mais complicada do que os cálculos na representação complexa. O procedimento adotado será o uso do cálculo de Wirtinger para determinar a métrica aumentada complexa, para qualquer sistema de coordenadas de dimensão par, mantendo a correspondência entre o caso real e o caso aumentado complexo.

Por fim, para mostrar a equivalência entre as representações, são apresentados três exemplos: em coordenadas cartesianas, em cilíndricas e em oblíquas. Na primeira, a presença de um fator de escala permite comparar, por exemplo, o algoritmo gradiente descendente em domínios reais e complexos ao longo do eixo real (ambos são iguais quando os sinais reais são aplicados) o que geralmente não ocorre em vários estudos, como (Kreutz-Delgado, 2009) e (Brandwood, 1983), mas com uso aplicação acertado em (Mandic; Goh, 2009), (Monzingo; Miller, 1980) e (Widrow; McCool; Ball, 1975). Os dois últimos exemplos demonstram a relação entre a notação complexa aumentada e os resultados de gradiente natural em (Amari; Douglas, 1998).

7.1 Conceito geral

O método do gradiente foi aplicado extensivamente nos capítulos 5.4 e 6. Este, para um espaço $\mathbf{x} \in \mathbb{R}^N$, tem a forma

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mu_k \mathbf{B}^{-1} \nabla f|_{\mathbf{x}_k}, \quad (7.1)$$

onde $\mathbf{B} \in \mathbb{R}^{N \times N}$ é alguma matriz simétrica não singular que pode ou não variar ponto a ponto. O método do gradiente utilizado foi o qual $\mathbf{B} = \mathbf{I}$, a matriz identidade. O algoritmo de Newton o caso em que $\mathbf{B} = \mathbf{H}$, o hessiano. O fator de passo também pode ser variável (que não foi considerado no presente trabalho por estar apresentando algoritmos gerais).

O método do gradiente natural (Amari; Douglas, 1998) é o caso em que a matriz $\mathbf{B}$ é a métrica das coordenadas utilizadas, em forma matricial denotada por $\mathbf{G}$. A nomenclatura *natural* presente no nome, decorre da matriz métrica¹ aparecer naturalmente quando se muda de sistema de coordenadas, como uma propriedade do gradiente.

7.1.1 Coordenadas generalizadas e métrica

O gradiente natural utiliza o inverso da métrica como correção, a fim de melhorar o desempenho da aproximação como descrito em (Amari; Douglas, 1998). Com isso em mente, essa seção discorre resumidamente sobre todo ferramental necessário para a definição da matriz métrica e a justificativa de a utilizar.

Seja um vetor de posição $\mathbf{r} = \mathbf{r}(u_1, \dots, u_N)$ definido no $\mathbf{R}^N$ (ou de forma similar $x_k = x_k(u_1, \dots, u_N)$ para uma mudança de coordenadas do sistema euclidiano para os com coordenadas u_k). A partir do vetor incremental de deslocamento

$$d\mathbf{r} = \sum_k \frac{\partial \mathbf{r}}{\partial u_k} du_k ,$$

pode-se definir a base das coordenadas dada pelo vetor

$$\mathbf{e}_k = \frac{\partial \mathbf{r}}{\partial u_k} , \quad (7.2)$$

linearmente independentes. A métrica aparece quando se quer fazer medidas de comprimentos (e/ou ângulos). O elemento de comprimento ao quadrado g (chamado de métrica) é

$$\begin{aligned} g = \langle d\mathbf{r}, d\mathbf{r} \rangle &= \left\langle \sum_k \frac{\partial \mathbf{r}}{\partial u_k} du_k, \sum_l \frac{\partial \mathbf{r}}{\partial u_l} du_l \right\rangle \\ &= \sum_{k,l} \langle \mathbf{e}_k, \mathbf{e}_l \rangle du_k du_l = \sum_{k,l} g_{kl} du_k du_l . \end{aligned}$$

O termo

$$g_{kl} = \langle \mathbf{e}_k, \mathbf{e}_l \rangle , \quad (7.3)$$

é a matriz métrica do novo sistema coordenado, uma correção nas medidas provocadas pelo uso de outro sistema de coordenadas. O produto interno do espaço euclidiano $\mathbb{R}^N$ induz essa nova métrica². A métrica g é um tensor de segunda ordem que atua em vetores resultando em

¹ A matriz métrica é a matriz que representa a forma bilinear da métrica.

² A matriz métrica será utilizada como sinônimo de métrica.

um escalar, no caso euclidiano $g : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$, i.e para os vetores $\mathbf{v} = \sum v_k \mathbf{e}_k$ e $\mathbf{w} = \sum w_l \mathbf{e}_l$ a métrica atua sobre esses vetores resultando em:

$$g(\mathbf{v}, \mathbf{w}) = \sum_{kl} g_{kl} v_k w_l.$$

Isso ocorre devido a definição de base dual como: $du_k(\mathbf{e}_i) = \delta_{ki}$, em que $\delta_{ki} = 1$ se $k = i$ e zero caso contrário.

A simetria do produto interno implica em uma métrica igualmente simétrica. Em sua representação matricial, isso é expresso por:

$$\mathbf{G} = \begin{bmatrix} g_{11} & \cdots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{N1} & \cdots & g_{NN} \end{bmatrix}. \quad (7.4)$$

Se $g_{kl} = \delta_{kl}$ o sistema de coordenadas é euclidiano. Para $g_{kl} = 0$ com $k \neq l$, é dito ortogonal.

7.1.2 O vetor gradiente em coordenadas generalizadas

Com o auxílio do produto interno é fácil transitar entre formas diferenciais e vetores. A forma mais importante na análise do vetor gradiente é a diferencial (ou derivada) de uma aplicação $f : \mathbb{R}^N \rightarrow \mathbb{R}$, dada por

$$df = \frac{\partial f}{\partial u_1} du_1 + \cdots + \frac{\partial f}{\partial u_N} du_N$$

e representa o incremento na função f dada por cada incremento das coordenadas u_k , mas está expresso nas bases du_k , bases duais de $\mathbf{e}_k$. Isso implica que df pode atuar sobre vetores $\mathbf{v} = v_1 \mathbf{e}_1 + \cdots + v_N \mathbf{e}_N$, resultando em um valor escalar, em termos de suas componentes como

$$df(\mathbf{v}) = \sum_k \frac{\partial f}{\partial u_k} du_k(\mathbf{v}) = \sum_k \frac{\partial f}{\partial u_k} v_k.$$

A métrica pode ser imaginada da mesma forma, já que está escrita também na base dual. Esta atuando sobre um par de vetores $\mathbf{v} = v_1 \mathbf{e}_1 + \cdots + v_N \mathbf{e}_N$ e $\mathbf{w} = w_1 \mathbf{e}_1 + \cdots + w_N \mathbf{e}_N$ é

$$g(\mathbf{v}, \mathbf{w}) = \sum_{k,l} g_{kl} du_k(\mathbf{v}) du_l(\mathbf{w}) = \sum_{k,l} g_{kl} v_k w_l,$$

representa o produto interno entre os vetores dados, pois

$$\langle \mathbf{v}, \mathbf{w} \rangle = \left\langle \sum_k v_k \mathbf{e}_k, \sum_l w_l \mathbf{e}_l \right\rangle = \sum_{k,l} g_{kl} v_k w_l,$$

de acordo com (7.3). Então tem-se a propriedade $g(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle$, na qual o produto euclidiano é utilizado nas novas coordenadas (ou novo espaço, conceito de métrica induzida) e atua sobre as bases não necessariamente ortonormais. Em termos de coordenadas locais (em termos das componentes em forma vetorial) a operação é dada por

$$g(\mathbf{v}, \mathbf{w}) = \begin{bmatrix} v_1 & \cdots & v_N \end{bmatrix} \cdot \begin{bmatrix} g_{11} & \cdots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{N1} & \cdots & g_{NN} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \mathbf{v}^T \mathbf{G} \mathbf{w} .$$

A matriz $\mathbf{G}$ acima representa uma forma positiva definida pois

$$\mathbf{v}^T \mathbf{G} \mathbf{v} = \sum_{k,l} v_k v_l \langle \mathbf{e}_k, \mathbf{e}_l \rangle = \sum_{k,l} \langle v_k \mathbf{e}_k, v_l \mathbf{e}_l \rangle = \left\langle \sum_k v_k \mathbf{e}_k, \sum_l v_l \mathbf{e}_l \right\rangle = \left\| \sum_k v_k \mathbf{e}_k \right\|^2 \geq 0.$$

O gradiente, indicado por ∇f , o qual é uma incógnita até aqui, seria o equivalente vetorial de df . Estes podem ser associados pelo uso do produto interno (métrica) da forma

$$df(\mathbf{v}) = g(\nabla f, \mathbf{v}) = \langle \nabla f, \mathbf{v} \rangle .$$

Expressando as operações em coordenadas locais, com δf representando as componentes de df na base dada:

$$\delta f^T \mathbf{v} = \nabla f^T \mathbf{G} \mathbf{v} \rightarrow \delta f^T = \nabla f^T \mathbf{G} .$$

A expressão é válida para todo vetor $\mathbf{v}$ dado. O termo

$$\nabla f = \mathbf{G}^{-1} \delta f ,$$

representa o vetor gradiente em termos das componentes da forma diferencial. Esse gradiente (agora chamado de gradiente natural pois é corrigido pela métrica) é expresso em termos de componentes por

$$\begin{bmatrix} (\nabla f)_1 \\ \vdots \\ (\nabla f)_N \end{bmatrix} = \mathbf{G}^{-1} \cdot \begin{bmatrix} \frac{\partial f}{\partial u_1} \\ \vdots \\ \frac{\partial f}{\partial u_N} \end{bmatrix} . \quad (7.5)$$

Os termos $(\nabla f)_k = \sum_l g_{kl} \partial f / \partial u_l$ acima se referem a k -ésima componente do gradiente que será dado por

$$\nabla f = \left(\sum_l g_{1l} \frac{\partial f}{\partial u_l} \right) \mathbf{e}_1 + \cdots + \left(\sum_l g_{Nl} \frac{\partial f}{\partial u_l} \right) \mathbf{e}_N .$$

O uso do gradiente acima para o método do gradiente (7.1) é o chamado *método do gradiente natural*, será visto a seguir.

7.1.3 O método do gradiente natural

Esta seção apresenta o método para o caso real (como foi deduzido nas seções anteriores), bem como alguns exemplos. Pela expressão (7.1) e utilizando a matriz $\mathbf{B} = \mathbf{G}$ como a métrica, a regra de aproximação para um passo fixo é

$$\mathbf{u}[n+1] = \mathbf{u}[n] - \mu \mathbf{G}^{-1} \delta f. \quad (7.6)$$

em que o inverso da métrica é calculado a cada ponto. Em termos de notação, o vetor δf que possui componentes de df , nada mais é que a derivada em formato coluna

$$\delta f = \frac{\partial f}{\partial \mathbf{u}^T} = \begin{bmatrix} \frac{\partial f}{\partial u^1} \\ \vdots \\ \frac{\partial f}{\partial u^N} \end{bmatrix}.$$

Similar ao caso complexo em (3.22), aqui real.

7.1.3.1 Exemplo de aplicação

O exemplo, retirado de (Amari; Douglas, 1998) utiliza coordenada polar (caso bi-dimensional). O vetor posição escrito como um par ordenado é dado por $\mathbf{r}(\rho, \varphi) = (\rho \cos \varphi, \rho \sin \varphi)$. A função custo dada no exemplo é

$$\mathcal{J} = \frac{1}{2} \left[(\rho \cos \varphi - 1)^2 + \rho^2 \sin^2 \varphi \right].$$

Os vetores base, necessários para a construção da matriz métrica, são dados por meio das derivadas

$$\begin{cases} \frac{\partial \mathbf{r}}{\partial \rho} = (\cos \varphi, \sin \varphi) \triangleq \mathbf{e}_\rho \\ \frac{\partial \mathbf{r}}{\partial \varphi} = \rho(-\sin \varphi, \cos \varphi) \triangleq \mathbf{e}_\varphi. \end{cases}$$

A métrica, que independe da escolha da função custo, construída a partir das relações acima é

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ 0 & \rho^2 \end{bmatrix}, \quad (7.7)$$

sua inversa é facilmente calculada e posta no termo incremental do gradiente

$$\nabla f = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\rho^2} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \rho} \\ \frac{\partial f}{\partial \varphi} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial \rho} \\ \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \end{bmatrix}.$$

A expressão anterior pode ser escrita como

$$\nabla f = \frac{\partial f}{\partial \rho} \mathbf{e}_\rho + \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \mathbf{e}_\varphi = \frac{\partial f}{\partial \rho} \hat{\mathbf{e}}_\rho + \frac{1}{\rho} \frac{\partial f}{\partial \varphi} \hat{\mathbf{e}}_\varphi ,$$

em que a presença do acento circunflexo em $\mathbf{e}_k$ indica um vetor base de módulo unitário. Utilizando as derivadas acima na função custo proposta, a função de atualização do vetor de coordenadas $\mathbf{w}[n] = [\rho[n] \ \varphi[n]]^T$ num dado instante discreto é

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \mu \begin{bmatrix} \rho[n] - \cos \varphi[n] \\ \frac{\sin \varphi[n]}{\rho[n]} \end{bmatrix} . \quad (7.8)$$

A inversa da métrica só precisa ser calculada um vez, se tornando uma matriz a ser calculada ponto a ponto, ainda sendo possível estimá-la em dimensões mais altas. Versões modernas do gradiente natural utilizam como métrica a matriz de Fisher (Amari; Nagaoka, 2000).

7.2 Conceito do gradiente complexo natural

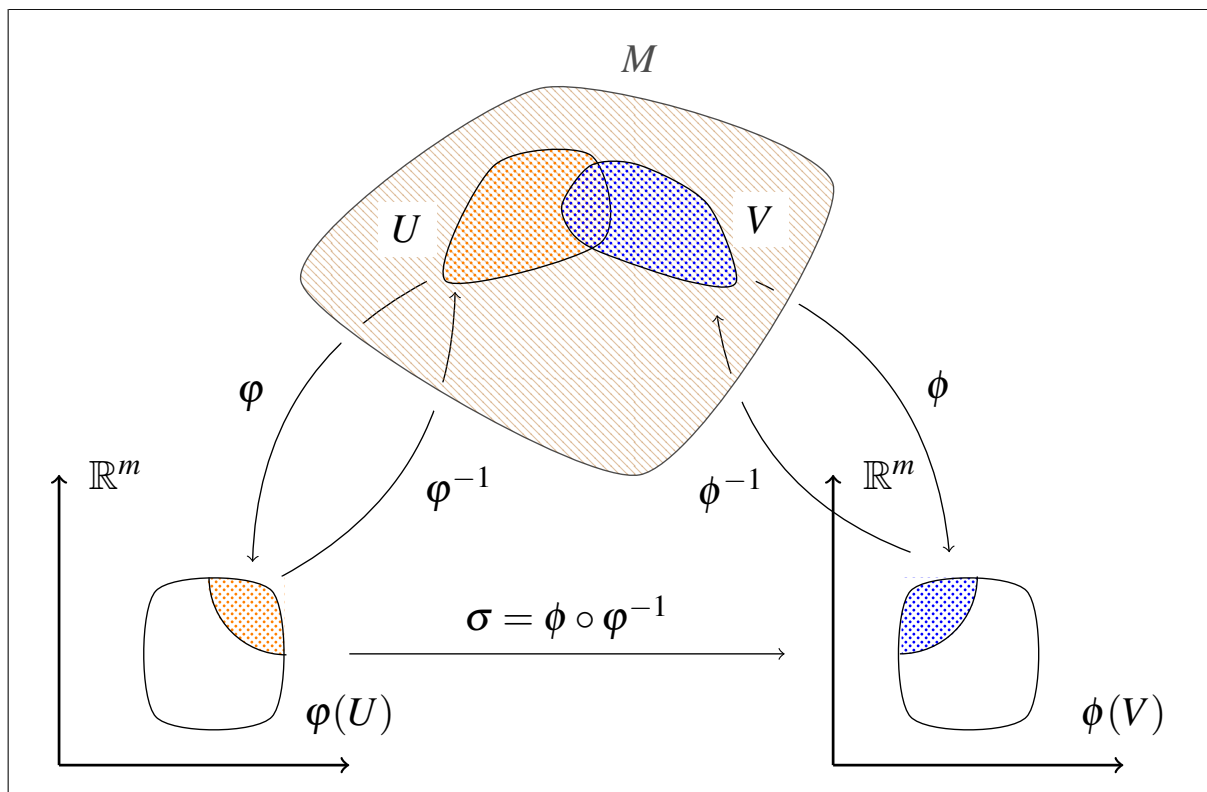
O gradiente complexo natural surge da escrita direta, em variáveis complexas do caso real. É uma extensão dos resultados anteriores no espaço aumentado, permitindo a escrita em termos da variável e sua conjugada. A partir dessa seção serão introduzidos conceitos a fim de generalizar os resultados anteriores, começando pela generalização de espaços para aqueles de dimensões e formatos mais gerais, chamados de variedades. Em seguida estas serão expressas por meio de coordenadas complexas e terão o gradiente calculado. Por fim, o gradiente complexo natural é apresentado representando o gradiente natural escrito inteiramente na notação de Wirtinger.

7.2.1 Variedades riemannianas

O gradiente natural no plano pode ser visto do ponto de vista de transformações de coordenadas como feito na seção anterior. Aqui esse conceito é generalizado para um espaço mais geral que é uma variedade riemanniana, como será visto a seguir.

Pode-se imaginar uma variedade M (diferenciável) de dimensão m como um equivalente à uma hiper-superfície diferenciável sem necessariamente estar contida em um espaço ambiente (mesmo que esteja, o estudo independe do espaço suporte). Uma variedade pode ser localmente mapeada por um espaço euclidiano de mesma dimensão (abertos na variedade são

Figura 32 – Exemplo de variedade com mapeamento em duas regiões que possuem uma interseção



Fonte: elaborada pelo autor.

homeomórficas a abertos no espaço euclidiano³). Para que uma variedade seja dita riemanniana é preciso que esta tenha uma métrica g em todo ponto, e que esta varie de maneira suave ponto a ponto (Carmo, 2008). Uma variedade riemanniana é representada como (M, g) .

Uma carta é definida como um par $\{U, \phi\}$, de um aberto U na variedade M mapeada por ϕ , chamado de sistema de coordenadas locais: funções diferenciáveis invertíveis. Um conjunto de cartas locais cobrindo a variedade é um atlas. É possível relacionar diferentes cartas, que possuam interseções, como mostra a Figura 32, por meio de uma mudança de coordenadas $\sigma : \phi(U \cap V) \rightarrow \phi(U \cap V)$. A inversa de um mapeamento ϕ^{-1} é chamada parametrização local para $\phi(U)$ em M .

7.2.2 Espaço tangente

Para espaços mais abstratos é necessária a definição de vetor tangente de forma um pouco mais geral que o apresentado na expressão em (7.2). Ao invés de utilizar vetores os imaginando como “setas no espaço” o estudo de variedades define os vetores como derivadas.

³ Isto significa que um aberto sobre a variedade pode ser deformado e dobrado (sem fazer cantos ou rasgos) resultando em um aberto no espaço euclidiano.

Isso é possível devido ao isomorfismo que existe entre as bases no $\mathbb{R}^n$ e operadores diferenciais, como será visto a seguir, primeiramente para o caso euclidiano e em seguida sobre variedades.

7.2.2.1 Vetores no $\mathbb{R}^n$

Vetores a partir da origem no espaço euclidiano são expressos em termos do vetores base como

$$\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i$$

em que $\mathbf{e}_k$ são os vetores base canônicos. Pode-se utilizar uma base a partir de um dado ponto p tal que $\mathbf{v}_p = \sum_{i=1}^N v_i \mathbf{e}_i|_p$ (chamado espaço afim) na qual as coordenadas v_k são medidas a partir do ponto p . Em ambos os casos o vetor pode ser expresso meio de um vetor coluna $\mathbf{v} = \mathbf{v}_p = [v_1 \cdots v_N]^T$. Esse é o ponto de vista de vetores como setas orientadas, podendo ser transladadas até a origem sem perda de generalização.

Por outro lado, considere uma função $f : \mathbb{R}^N \rightarrow \mathbb{R}$ restrita a uma curva parametrizada num dado intervalo $\alpha : (-\varepsilon, \varepsilon) \subset \mathbb{R} \rightarrow \mathbb{R}^N$. Tome o ponto no intervalo aberto considerado $0 \in (-\varepsilon, \varepsilon)$ de forma que $\alpha(0) = p$. A derivada da função restrita à curva é

$$\left. \frac{d(f \circ \alpha(t))}{dt} \right|_{t=0} = \left. \frac{df(x_1(t), \dots, x_N(t))}{dt} \right|_{t=0},$$

em que o símbolo “ $\circ$ ” indica operação de composição de funções. Por se tratar de uma composição sobre o espaço euclidiano pode-se escrever a expressão anterior como

$$\left. \frac{d(f \circ \alpha(t))}{dt} \right|_{t=0} = \sum_{k=1}^N x'_k(0) \left. \frac{\partial f}{\partial x_k} \right|_p = \left(\sum_{k=1}^N x'_k(0) \left. \frac{\partial}{\partial x_k} \right|_p \right) f. \quad (7.9)$$

Tomando a curva como $x_k(t) = p_k + tv_k$, o operador (termo entre parêntesis do lado direito da expressão (7.9) acima) pode ser posto na forma

$$D_v = \sum_{k=1}^N v_k \left. \frac{\partial}{\partial x_k} \right|_p,$$

representando o operador derivada direcional no ponto p que atua sobre funções suaves do tipo $f : \mathbb{R}^N \rightarrow \mathbb{R}$. Podemos então identificar o vetor a partir de um dado ponto p no $\mathbb{R}^N$ pelo isomorfismo⁴ $\mathbf{v}_p \approx D_v$, de forma que o vetor dado pode ser expresso por

$$\mathbf{v}_p = \sum_{k=1}^N v_k \left. \frac{\partial}{\partial x_k} \right|_p.$$

⁴ O isomorfismo entre vetores e operadores permite escrita das bases como derivadas parciais (operadores diferenciais). Isso é útil em variedades mais abstratas na qual não é possível definir vetores como setas no espaço.

O vetor, como escrito do ponto de vista da expressão acima, é um operador diferencial e atua sobre funções reais no ponto dado, resultando em um valor escalar. O fato de vetores agora serem representados por operadores diferenciais permite o uso de propriedades como: a regra do produto $\mathbf{v}_p(fg) = \mathbf{v}_p(f)g(p) + f(p)\mathbf{v}_p(g)$, bem como definições de linearidade como $(\mathbf{v}_p + \mathbf{w}_p)f = \mathbf{v}_p(f) + \mathbf{w}_p(f)$ e $(k\mathbf{v}_p)f = k\mathbf{v}_p(f)$, para $\mathbf{w}_p$ também um vetor a partir do ponto p .

7.2.2.2 Vetores sobre a variedade

Sobre a notação utilizada: Desta seção em diante os vetores definidos sobre variedades usarão letras maiúsculas sem negrito. Vetores minúsculos em negrito se referirão ao formato coluna (ou linha no caso de co-vetores) das componentes dada uma base. Isso é conveniente para distinguir um vetor escrito em conjunto com sua base de sua representação como vetor coluna (somente componentes) i.e, dado um vetor

$$V = \sum_{k=1}^N v_k \frac{\partial}{\partial x_k} ,$$

o vetor coluna

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix} ,$$

representa as componentes de V na base dada.

Para uma variedade M de dimensão N , pode-se agir de maneira semelhante ao caso $\mathbb{R}^N$, definindo desta vez uma função escalar sobre a variedade $f : M \rightarrow \mathbb{R}$. O espaço tangente pode ser definido restringindo a função a uma curva sobre M .

Seja um curva suave $\alpha(t)$ sobre a variedade M . A curva pode ser parametrizada no intervalo aberto $(-\varepsilon, \varepsilon)$ de forma que $\alpha(0) = p \in U \subset M$. O vetor tangente é um operador (Carmo, 2008) definido por

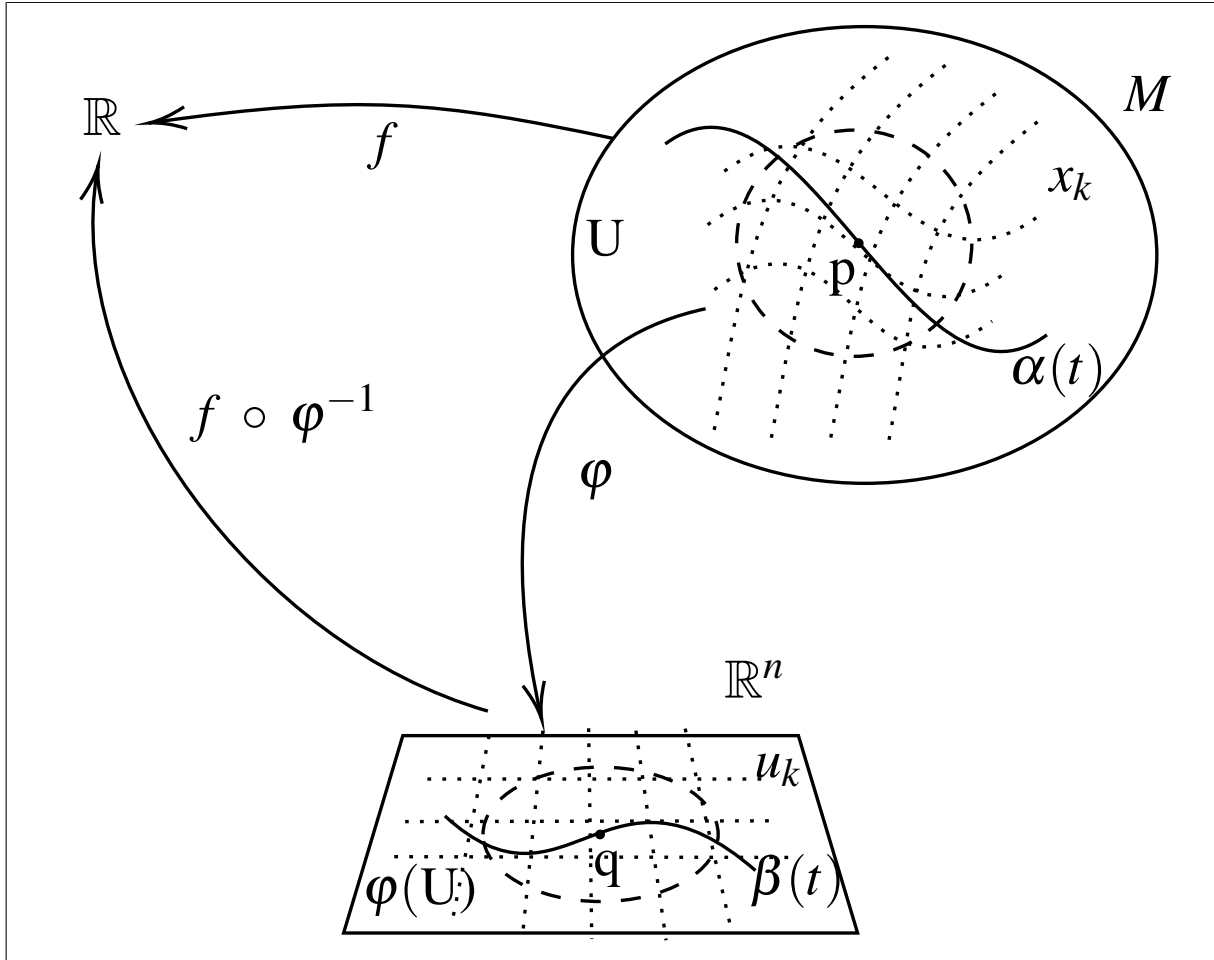
$$V_p(f) \triangleq \left. \frac{d}{dt}(f \circ \alpha(t)) \right|_{t=0} . \quad (7.10)$$

O termo $\alpha'(0) = V_p$ representa a derivada da curva num dado ponto, atuando sobre funções definidas em M .

A curva sobre a variedade em termos da carta (parametrização) é

$$\alpha(t) = \varphi^{-1} \circ \beta(t) ,$$

Figura 33 – Função definida sobre M restrita à curva mapeada pela carta, para uma parametrização φ



Fonte: elaborada pelo autor.

dados $\beta(t) = (u_1(t), \dots, u_N(t)) \in \varphi(U)$ a curva em função das coordenadas da carta com $\beta(t_0) = q = \varphi(p)$. A derivada (7.10) pode ser desenvolvida por

$$V_p(f) = \frac{d}{dt} f(\alpha(t)) \Big|_{t_0} = \frac{d(f \circ \varphi^{-1} \circ \beta(t))}{dt} \Big|_{t_0} = \sum_{i=1}^N u'_i(t_0) \frac{\partial(f \circ \varphi^{-1})}{\partial u_i} \Big|_{\varphi(p)}, \quad (7.11)$$

pois $f \circ \varphi^{-1} : \mathbb{R}^N \rightarrow \mathbb{R}$, representa a função por meio da carta, como mostra a Figura 33, em que essa função leva diretamente da carta aos reais.

Como as coordenadas da carta buscam representar (ou mapear) as coordenadas sobre a variedade x_i , estas são escolhidas de forma que utilizem os mesmos coeficientes da variedade. Definindo $x_i(p) = u_i \circ \varphi(p)$ mantém cada coordenada independente uma da outra, atuando como um mapa identidade para os coeficientes (coordenadas) e com isso

$$x_i(\alpha(t)) = u_i \circ \varphi \circ \alpha(t) = u_i(u_1(t), \dots, u_n(t)) = u_i(t),$$

que equivale à $u_i(t) = x_i(t)$ como coeficientes a serem utilizados. A derivada parcial em (7.11) pode ser definida então da forma

$$\left. \frac{\partial}{\partial x_i} \right|_p f := \left. \frac{\partial (f \circ \varphi^{-1})}{\partial u_i} \right|_{\varphi(p)} .$$

Com isso, da mesma forma que no caso $\mathbb{R}^N$, a expressão (7.11) fornece o vetor tangente expresso por bases de operadores diferenciais por

$$V_p = \sum_{i=1}^n v_i \left. \frac{\partial}{\partial x_i} \right|_p ,$$

para $v_i = x'_i(0)$. Como regra geral pode-se usar a carta diretamente como sinônimo da variedade, como apresentado em (Carmo, 2008). O fato da curva sobre a variedade não ter sido especificada permite determinar uma base associada ao espaço tangente no ponto dado como

$$\left\{ \left. \frac{\partial}{\partial x_1} \right|_p, \dots, \left. \frac{\partial}{\partial x_n} \right|_p \right\} . \quad (7.12)$$

Essa base permite a construção do espaço tangente $T_p M$ no ponto: um espaço vetorial linear, isomorfo à $\mathbb{R}^N$. Quando não for necessário especificar o ponto dado é comum escrever as bases sem e subscrito p bem como o espaço tangente como TM .

7.2.3 Métricas sobre a variedade

Os vetores tangentes em uma variedade riemanniana possibilitam a definição de uma métrica, que por sua vez determina um produto interno $\langle \cdot, \cdot \rangle$ no espaço tangente TM em um ponto específico. Essa métrica, expressa em termos dos parâmetros (ou coordenadas) da variedade, é representada por

$$g_{kl} = \left\langle \left. \frac{\partial}{\partial x_k} \right|, \left. \frac{\partial}{\partial x_l} \right| \right\rangle , \quad (7.13)$$

Para que a variedade seja considerada riemanniana, a métrica deve ser suave e diferenciável em todos os pontos. É importante ressaltar que a métrica pode ser derivada de um espaço ambiente no qual a variedade está imersa ou através de medidas de comprimento intrínsecas à própria variedade, utilizando conceitos de formas diferenciais. .

As seções seguintes se concentram em obter a forma do gradiente tangente sobre uma variedade expressa por meios de coordenadas complexas aumentadas.

7.2.4 Bases complexificadas

As definições vistas até aqui permitem construir um espaço complexo a partir de um espaço real. Isso se torna claro quando se considera as bases de um espaço tangente TM^{2N} para algum ponto dado (mas não especificado) de uma variedade diferenciável de dimensão par M^{2N} .

As bases são

$$\left\{ \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N}, \frac{\partial}{\partial y_1}, \dots, \frac{\partial}{\partial y_N} \right\}, \quad (7.14)$$

em coordenadas da variedade $(x_1, \dots, x_N, y_1, \dots, y_N)$. Reescrevendo as bases, com o auxílio dos operadores (3.21),

$$\left\{ \frac{\partial}{\partial z_k}, \frac{\partial}{\partial z_k^*} \right\} = \left\{ \frac{1}{2} \left(\frac{\partial}{\partial x_k} - j \frac{\partial}{\partial y_k} \right), \frac{1}{2} \left(\frac{\partial}{\partial x_k} + j \frac{\partial}{\partial y_k} \right) \right\}. \quad (7.15)$$

pois as bases reais são construídas sobre funções definidas sobre a variedade. Essas bases complexas têm como duais as bases $\{dz_1, \dots, dz_N, dz_1^*, \dots, dz_N^*\}$, onde $dz_k = dx_k + jdy_k$ e $dz_k^* = dx_k - jdy_k$ para $k = \{1, \dots, n\}$. Isso pode ser verificado diretamente pela construção das bases complexas, i.e

$$dz_m \left(\frac{\partial}{\partial z_n} \right) = \delta_{mn}, \quad dz_m \left(\frac{\partial}{\partial z_n^*} \right) = 0$$

$$dz_m^* \left(\frac{\partial}{\partial z_n} \right) = 0, \quad dz_m^* \left(\frac{\partial}{\partial z_n^*} \right) = \delta_{mn},$$

em que $\delta_{mn} = 1$ se $m = n$, zero para todos outros casos.

Será visto a seguir a construção de dois tipos de espaços usando (7.15): a representação real (chamada complexificada) que possui uma construção particular (aumentada) e a outra que possui coeficientes complexos utilizando a base real (7.14). Esta última equivale ao uso de coeficientes complexos e base complexa. O próximo tópico começará por expor a representação complexa sobre bases reais (extensão complexa), que é o caso mais geral, seguido do caso aumentado, sobre o qual se dará atenção.

7.2.5 Representação com coeficientes complexos

A representação utilizando coeficientes complexos sobre bases reais, em um dado ponto de uma variedade $2N$ -dimensional, fornece um espaço tangente complexo. Sejam os coeficientes complexos $a_k = a_k^r + ja_k^i$, $b_k = b_k^r + jb_k^i \in \mathbb{C}$, um vetor tangente U em um ponto

pode ser escrito como

$$U = \sum_{k=1}^N \left(a_k \frac{\partial}{\partial x_k} + b_k \frac{\partial}{\partial y_k} \right) .$$

Mudando de bases de acordo com (7.15), o mesmo vetor pode ser escrito por

$$U = \sum_{k=1}^N \left(\lambda_k \frac{\partial}{\partial z_k} + \mu_k \frac{\partial}{\partial z_k^*} \right) ,$$

com coeficientes

$$\begin{cases} \lambda_k = a_k^r - b_k^i + j(a_k^i + b_k^r) \\ \mu_k = a_k^r + b_k^i + j(a_k^i - b_k^r) . \end{cases} \quad (7.16)$$

Esses coeficientes, também complexos, agora utilizam a base complexa e pode ser posto na forma de vetor coluna por

$$\mathbf{u} = \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} , \quad (7.17)$$

onde $\boldsymbol{\lambda} = [\lambda_1 \dots \lambda_n]^T$ e $\boldsymbol{\mu} = [\mu_1 \dots \mu_n]^T$. Esse vetor irá pertencer ao espaço linear $T_{\mathbb{C}}M = TM \otimes \mathbb{C}$, isomórfico à $\mathbb{C}^{2N} \approx \mathbb{R}^{4N}$.

7.2.6 Representação aumentada

Para um vetor tangente real $V \in TM \approx \mathbb{R}^{2N}$, sua representação na base tangente (7.14) é

$$V = \sum_{k=1}^N \left(r_k \frac{\partial}{\partial x_k} + s_k \frac{\partial}{\partial y_k} \right) ,$$

para $r_k, s_k \in \mathbb{R}$, coeficientes reais. Este mesmo vetor na base complexa é dado por

$$V = \sum_{k=1}^N \left[(r_k + js_k) \frac{\partial}{\partial z_k} + (r_k - js_k) \frac{\partial}{\partial z_k^*} \right] .$$

Utilizando um novo coeficiente complexo $\eta_k = r_k + js_k$, o vetor real acima, quando escrito em base complexa, é equivalente à um vetor aumentado

$$V = \sum_{k=1}^N \left(\eta_k \frac{\partial}{\partial z_k} + \eta_k^* \frac{\partial}{\partial z_k^*} \right) . \quad (7.18)$$

O vetor (7.18) representado em coordenadas locais pelo vetor coluna é

$$\mathbf{v} = \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{bmatrix} , \quad (7.19)$$

para $\boldsymbol{\eta} = [\eta_1 \dots \eta_N]^T$. Vetores nas formas (7.18) e (7.19) são os já conhecidos *aumentados*, e representam um espaço real de dimensão de mesmo tamanho, ao contrário de (7.17), um espaço de dimensão dobrada.

7.2.7 A estrutura complexa

É importante tomar nota sobre a representação de um vetor somente na base não-conjugada $\partial/\partial z_k$, como

$$U' = \sum_{k=1}^n \lambda_k \frac{\partial}{\partial z_k}.$$

Essa representação é dada pelo espaço tangente holomórfico $T_{\mathbb{C}}^{1,0}M$. O correspondente conjugado possui vetores do tipo,

$$U'' = \sum_{k=1}^n \mu_k \frac{\partial}{\partial z_k^*},$$

descrito por $T_{\mathbb{C}}^{0,1}M$ com as derivadas conjugadas $\partial/\partial z_k^*$ como base. Observe o fato que um vetor geral complexo é dado por $T_{\mathbb{C}}M = T_{\mathbb{C}}^{1,0}M \oplus T_{\mathbb{C}}^{0,1}M$.

A possibilidade de escrita de um vetor real utilizando somente a base não-conjugada (ou base holomorfa) é possível quando associada à uma estrutura complexa J , um operador que tem a seguinte propriedade

$$\begin{cases} J \frac{\partial}{\partial x_k} = \frac{\partial}{\partial y_k} \\ J \frac{\partial}{\partial y_k} = -\frac{\partial}{\partial x_k} \end{cases}, \quad (7.20)$$

com $J^2 = -I$, a identidade. Esse operador leva a uma regra geral mnemônica para os produtos dos coeficientes $1 \cdot j = j \cdot 1 = J$ e $j \cdot j = -1$, de forma que é possível a representação de vetores reais somente pela componente holomorfa. Seja $V' \in T_{\mathbb{C}}^{1,0}M$, essa representação, com auxílio de tal estrutura leva a

$$\begin{aligned} V' &= \sum_{k=1}^n \eta_k \frac{\partial}{\partial z_k} = \sum_{k=1}^n \frac{1}{2} (1r_k + js_k) \cdot \left(1 \frac{\partial}{\partial x_k} - j \frac{\partial}{\partial y_k} \right) \\ &= \sum_{k=1}^n \frac{1}{2} \left(r_k \frac{\partial}{\partial x_k} - r_k J \frac{\partial}{\partial y_k} + s_k J \frac{\partial}{\partial x_k} + s_k \frac{\partial}{\partial y_k} \right), \end{aligned}$$

com auxílio das regras de multiplicação dos coeficientes,

$$V' = \sum_{k=1}^n \left(r_k \frac{\partial}{\partial x_k} + s_k \frac{\partial}{\partial y_k} \right) = V.$$

Este resultado evidencia a representação somente em base holomorfa, com a devida presença da estrutura complexa, que leva esta à forma real (7.18). A diferença de notação ora em letra maiúscula para vetores expressos em sua base, ora em notação negrita na forma matricial tem uma razão: enquanto para a representação aumentada a condição $V^* = V$ é verdade (as bases também conjugam), a notação coluna (matricial), em geral $\mathbf{v}^* \neq \mathbf{v}$.

7.2.8 Representação de formas diferenciais complexas

Formas diferenciais formam um espaço vetorial, definidos no dual de espaços tangentes (são então construídas pelas bases duais dx_k, dy_k ou suas versões complexas). As chamadas 1-formas equivalem à vetores e agem sobre estes, resultando em valores escalares. Da mesma maneira que vetores, estas podem ser estendidas por coeficientes complexos. Elas serão representadas por letras gregas minúsculas sem índices e apresentadas na forma complexa a seguir.

Sejam os coeficientes complexos $\bar{a}_k = \bar{a}_k^r + j\bar{a}_k^i$ e $\bar{b}_k = \bar{b}_k^r + j\bar{b}_k^i$ ambas em $\mathbb{C}$, estendendo uma 1-forma real com esses coeficientes tem-se

$$v = \sum_{k=1}^N (\bar{a}_k dx_k + \bar{b}_k dy_k) ,$$

com cada elemento de base complexo $dz_k = dx_k + jdy_k$ e seu conjugado. É possível escrever a expressão acima como

$$v = \frac{1}{2} \sum_{k=1}^N (\bar{\mu}_k dz_k + \bar{\lambda}_k dz_k^*) , \quad (7.21)$$

para os coeficientes

$$\begin{cases} \bar{\lambda}_k = \bar{a}_k^r - \bar{b}_k^i + j(\bar{a}_k^i + \bar{b}_k^r) \\ \bar{\mu}_k = \bar{a}_k^r + \bar{b}_k^i + j(\bar{a}_k^i - \bar{b}_k^r) . \end{cases} \quad (7.22)$$

A fim de se ter o produto hermitiano quando em notação aumentada, a aplicação da 1-forma requer a conjugação (incluindo as bases). Para um vetor $Q = \sum_{k=1}^N (\delta_k \partial / \partial z_k + \varepsilon_k \partial / \partial z_k^*)$ no espaço tangente, a aplicação da conjugada da forma (7.21) é

$$v^*(Q) = \frac{1}{2} \sum_{k=1}^N (\bar{\lambda}_k^* \delta_k + \bar{\mu}_k^* \varepsilon_k) = \frac{1}{2} \begin{bmatrix} \bar{\lambda}^* & \bar{\mu}^* \end{bmatrix} \begin{bmatrix} \delta \\ \varepsilon \end{bmatrix} \quad (7.23)$$

para os vetores linha $\bar{\lambda} = [\bar{\lambda}_1 \cdots \bar{\lambda}_N^*]$ e $\bar{\mu} = [\bar{\mu}_1 \cdots \bar{\mu}_N]$. Esta expressão em coordenadas locais é representada pelo vetor linha

$$\mathbf{v} = \frac{1}{2} \bar{\mathbf{u}} ,$$

com $\bar{\mathbf{u}} = [\bar{\boldsymbol{\lambda}} \quad \bar{\boldsymbol{\mu}}]$. Em breve será possível relacionar esta forma com o vetor (7.17) pela métrica definida sobre a variedade.

Para o caso complexo aumentado pode-se usar a construção anterior com $\bar{\boldsymbol{\lambda}} = \bar{\boldsymbol{\alpha}}$ e $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\alpha}}^*$, denominando a forma aumentada como ω . Aplicando a forma a um vetor tangente aumentado é

$$\omega^*(V) = \frac{1}{2} \bar{\mathbf{w}}^* \mathbf{v}, \quad (7.24)$$

em formato matricial. Também é possível representar toda a operação anterior em coordenadas reais com coeficientes reais para $\omega = \sum_{k=1}^N (\bar{p}_k dx_k + \bar{q}_k dy_k)$ e $V = \sum_{k=1}^N (r_k \partial/\partial x_k + s_k \partial/\partial y_k)$, versões na base real dos operadores aumentados anteriores, resultando no valor real

$$\omega^*(V) = \begin{bmatrix} \bar{\mathbf{p}} & \bar{\mathbf{q}} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} = \bar{\mathbf{p}}\mathbf{r} + \bar{\mathbf{q}}\mathbf{s}, \quad (7.25)$$

devido à $\omega^* = \omega$, para a base real com coeficientes reais. Isso também vale para a representação aumentada, onde $\bar{\alpha}_k = \bar{p}_k + j\bar{q}_k$ and $\eta_k = r_k + js_k$,

$$\omega^*(V) = \frac{1}{2} \begin{bmatrix} \bar{\boldsymbol{\alpha}}^* & \bar{\boldsymbol{\alpha}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \bar{\boldsymbol{\eta}} \end{bmatrix} = \Re \{ \bar{\boldsymbol{\alpha}}^* \boldsymbol{\eta} \} = \bar{\mathbf{p}}\mathbf{r} + \bar{\mathbf{q}}\mathbf{s}. \quad (7.26)$$

Os resultados (7.25) e (7.26) mostram a equivalência entre essas representações: na forma aumentada, a ação de uma 1-forma em um vetor tangente aumentado resulta em um valor real da mesma maneira que no caso real. A escolha da ação do conjugado de uma forma implica a escolha de um produto interno sesquilinear no segundo operador e isso permite a notação complexa matricial no caso aumentado.

7.2.9 Produto interno e métrica

A fim de relacionar a métrica com a aplicação de formas diferenciais é necessário a construção de um produto interno aumentado baseado no produto interno hermitiano. Considere uma forma sesquilinear aumentada no segundo operando h , que é aplicada aos vetores tangentes $V = \sum (\eta_k \partial/\partial z_k + \eta_k^* \partial/\partial z_k^*)$ e $W = \sum (\alpha_k \partial/\partial z_k + \alpha_k^* \partial/\partial z_k^*)$ para os coeficientes $\eta_k, \alpha_k \in \mathbb{C}$. Pelas propriedades da linearidade

$$h(W, V) = \sum_{i,k} \left\langle \alpha_i \frac{\partial}{\partial z_i} + \alpha_i^* \frac{\partial}{\partial z_i^*}, \eta_k \frac{\partial}{\partial z_k} + \eta_k^* \frac{\partial}{\partial z_k^*} \right\rangle,$$

em que $\langle \cdot, \cdot \rangle$ é o produto interno hermitiano (conjuga o segundo operando) que representa h . Em notação matricial como

$$h(W, V) = \begin{bmatrix} \boldsymbol{\alpha}^H & \boldsymbol{\alpha}^T \end{bmatrix} \cdot \tilde{\mathbf{H}} \cdot \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{bmatrix} = \mathbf{w}^H \tilde{\mathbf{H}} \mathbf{v} ,$$

para a matriz composta

$$\tilde{\mathbf{H}} = \begin{bmatrix} \left[\left\langle \frac{\partial}{\partial z_i^*}, \frac{\partial}{\partial z_k^*} \right\rangle \right] & \left[\left\langle \frac{\partial}{\partial z_i^*}, \frac{\partial}{\partial z_k} \right\rangle \right] \\ \left[\left\langle \frac{\partial}{\partial z_i}, \frac{\partial}{\partial z_k^*} \right\rangle \right] & \left[\left\langle \frac{\partial}{\partial z_i}, \frac{\partial}{\partial z_k} \right\rangle \right] \end{bmatrix} , \quad (7.27)$$

em que o índice i representa a coluna e k a linha. Esta matriz se relaciona com uma dada métrica real

$$\mathbf{G} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}^T & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \left[\left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_k} \right\rangle \right] & \left[\left\langle \frac{\partial}{\partial y_i}, \frac{\partial}{\partial x_k} \right\rangle \right] \\ \left[\left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial y_k} \right\rangle \right] & \left[\left\langle \frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_k} \right\rangle \right] \end{bmatrix}$$

para $\mathbf{P} = \mathbf{P}^T$, $\mathbf{R} = \mathbf{R}^T$ e $\mathbf{G} = \mathbf{G}^T$, por expansão direta de (7.27). Para $\mathbf{H} = 2\tilde{\mathbf{H}}$, tem-se que

$$\mathbf{H} = \frac{1}{2} \begin{bmatrix} \mathbf{P} + \mathbf{R} + j(\mathbf{Q}^T - \mathbf{Q}) & \mathbf{P} - \mathbf{R} + j(\mathbf{Q}^T + \mathbf{Q}) \\ \mathbf{P} - \mathbf{R} - j(\mathbf{Q}^T + \mathbf{Q}) & \mathbf{P} + \mathbf{R} - j(\mathbf{Q}^T - \mathbf{Q}) \end{bmatrix} . \quad (7.28)$$

Essa última expressão demonstra a relação entre a forma bilinear e a sua versão real com presença da métrica. Observe que numa dada base

$$h(W, V) = \langle W, V \rangle = \frac{1}{2} \mathbf{w}^H \mathbf{H} \mathbf{v} , \quad (7.29)$$

é válida para vetores em sua forma aumentada e é a versão de (3.19) para vetores em coordenadas gerais. A métrica é importante pois permite relacionar vetores com seus duais. Comparando (7.29) com a expressão (7.24) pode-se perceber que a aplicação de uma forma pode ser representada por um produto interno hermitiano

$$\omega^*(V) = \langle W, V \rangle . \quad (7.30)$$

A métrica $\mathbf{H}$ preserva as distâncias reais quando se trabalha em notação aumentada, como indicado em (7.25) e (7.26).

7.2.10 O co-gradiente

A diferencial de uma função resulta em uma 1-forma, usualmente esse processo de diferenciação chamado de derivada exterior. Esta, em coordenadas complexas, pode ser expressa

por

$$df = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}} & \frac{\partial f}{\partial \mathbf{z}^*} \end{bmatrix} \cdot \begin{bmatrix} d\mathbf{z} \\ d\mathbf{z}^* \end{bmatrix}. \quad (7.31)$$

para os vetores linha $\partial f / \partial \mathbf{z} = [\partial f / \partial z_1 \dots \partial f / \partial z_n]$, $\partial f / \partial \mathbf{z}^* = [\partial f / \partial z_1^* \dots \partial f / \partial z_n^*]$ e para $d\mathbf{z} = [dz_1 \dots dz_n]^T$ e seu conjugado. Essa forma diferencial operando em um dado vetor V , como dado em (7.18) ou (7.19), resulta na expressão

$$df(V) = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}} & \frac{\partial f}{\partial \mathbf{z}^*} \end{bmatrix} \cdot \begin{bmatrix} d\mathbf{z} \\ d\mathbf{z}^* \end{bmatrix} V = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}} & \frac{\partial f}{\partial \mathbf{z}^*} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{bmatrix} \quad (7.32)$$

para $d\mathbf{z}(V) = \boldsymbol{\eta}$ e $d\mathbf{z}^*(V) = \boldsymbol{\eta}^*$ em forma vetorial. Lembrando que para uma função real f a conjugada da derivada holomorfa é

$$\left(\frac{\partial f}{\partial \mathbf{z}} \right)^* = \frac{\partial f^*}{\partial \mathbf{z}^*} = \frac{\partial f}{\partial \mathbf{z}^*}. \quad (7.33)$$

Nesse caso a forma diferencial equivale à

$$df(V) = \frac{\partial f}{\partial \mathbf{z}} \boldsymbol{\eta} + \left(\frac{\partial f}{\partial \mathbf{z}} \boldsymbol{\eta} \right)^* = 2\text{Re} \left\{ \frac{\partial f}{\partial \mathbf{z}} \boldsymbol{\eta} \right\}.$$

7.2.11 O gradiente complexo aumentado

A propriedade (7.29) permite a representação do chamado gradiente de uma função f (aqui representado por ∇f), que deverá ser a versão vetorial da já conhecida forma diferencial df . Ressaltando o uso da função real f e do vetor aumentado V (e sua forma vetorial coluna $\mathbf{v}$). A derivada direcional é

$$df(V) = \langle \nabla f, V \rangle = \frac{1}{2} (\nabla f)^H \mathbf{H} \cdot \mathbf{v}. \quad (7.34)$$

com $\mathbf{H}$ a métrica em (7.28). A forma df , representada por um vetor linha, já é escalonada e conjugada em sua própria definição. Nomeando o operador co-gradiente aumentado como $\boldsymbol{\delta} f = [\partial f / \partial \mathbf{z} \quad \partial f / \partial \mathbf{z}^*]$, a expressão para (7.30) pode ser escrita em coordenadas locais como

$$\boldsymbol{\delta} f \cdot \mathbf{v} = \frac{1}{2} (\nabla f)^H \mathbf{H} \cdot \mathbf{v},$$

válida para qualquer entrada vetorial $\mathbf{v}$ que representa o vetor V . Isolar o termo ∇f leva à expressão proposta para o gradiente complexo aumentado, dada a seguir por

$$\nabla f = 2\mathbf{H}^{-1}(\boldsymbol{\delta} f)^H. \quad (7.35)$$

Este gradiente depende de ambas derivadas não-conjugada e conjugada: fato de fácil verificação ao se expandir a expressão em suas componentes

$$\nabla f \equiv \begin{bmatrix} \nabla f_{\mathbf{z}} \\ \nabla f_{\mathbf{z}^*} \end{bmatrix} = 2\mathbf{H}^{-1} \begin{bmatrix} \left(\frac{\partial f}{\partial \mathbf{z}}\right)^H \\ \left(\frac{\partial f}{\partial \mathbf{z}^*}\right)^H \end{bmatrix} = 2\mathbf{H}^{-1} \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}^H} \\ \frac{\partial}{\partial \mathbf{z}^T} \end{bmatrix} f, \quad (7.36)$$

onde os índices $\mathbf{z}$ e $\mathbf{z}^*$ se referem às componentes das bases. A relação $\overline{\nabla f_{\mathbf{z}}} = \nabla f_{\mathbf{z}^*}$ se observa para funções reais. Devido ao formato aumentado da matriz $\mathbf{H}$ dada explicitamente em (7.28), de acordo com (Schreier; Scharf, 2010), sua inversa também manterá esse formato (por verificação algébrica). Nomeando essa inversa como

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{N}^* & \mathbf{M}^* \end{bmatrix},$$

a primeira componente do gradiente complexo natural aumentado é o valor dado pela equação

$$\nabla f_{\mathbf{z}} = 2\mathbf{M} \cdot \frac{\partial f}{\partial \mathbf{z}^H} + 2\mathbf{N} \cdot \frac{\partial f}{\partial \mathbf{z}^T}, \quad (7.37)$$

que é a componente do chamado *gradiente complexo natural*. Note que existem duas definições equivalentes devido à construção da forma aumentada: o *gradiente natural complexo aumentado* dado pela expressão (7.35) e o *gradiente natural complexo*, uma componente do anterior, retratado em (7.37).

7.2.12 Aplicações do gradiente natural complexo

A fim de demonstrar a equivalência entre o caso aumentado e real, três exemplos do uso do gradiente são dados nesta seção. O primeiro caso envolve o uso em coordenadas cartesianas, que leva a um fator de escala. Em seguida é apresentada a expressão para o caso polar. Por último um caso oblíquo bidimensional é descrito por apresentar uma métrica real não diagonal.

7.2.12.1 Representação cartesiana

Para a métrica cartesiana $\mathbf{G} = \mathbf{I}$ (a matriz identidade) em $\mathbb{R}^{2N}$, o gradiente complexo natural (componente da base não conjugada) é

$$\nabla f_{\mathbf{z}} = 2 \frac{\partial f}{\partial \mathbf{z}^H}.$$

Esse gradiente pode ser usado, para pequenos fatores de passo, como incremento por

$$\mathbf{z}_{k+1} = \mathbf{z}_k - \mu_k 2 \frac{\partial f}{\partial \mathbf{z}^H}, \quad (7.38)$$

para um fator de escala real μ_k . Pode-se verificar a equivalência deste com o caso real $\mathbb{R}^{2N}$ tomando $\mathbf{z} = \mathbf{x} + j\mathbf{y} \in \mathbb{C}^N$

$$\mathbf{x}_{k+1} + j\mathbf{y}_{k+1} = \mathbf{x}_k + j\mathbf{y}_k + \mu_k 2 \left[\frac{1}{2} \left(\frac{\partial f}{\partial \mathbf{x}^T} + j \frac{\partial f}{\partial \mathbf{y}^T} \right) \right],$$

que resulta em $\mathbf{x}_{k+1} = \mathbf{x}_k + \mu_k \nabla_{\mathbf{x}} f$ e $\mathbf{y}_{k+1} = \mathbf{y}_k + \mu_k \nabla_{\mathbf{y}} f$, para $\nabla_{\mathbf{x}} = \partial / \partial \mathbf{x}^T$ e $\nabla_{\mathbf{y}} = \partial / \partial \mathbf{y}^T$ devido a métrica em coordenadas cartesianas ser a identidade.

7.2.12.2 Representação polar

Considerando um exemplo de coordenadas polares em que a métrica é dada por $\mathbf{G} = \text{diag}\{1, \rho^2\}$, para um par de coordenadas dado por (ρ, φ) para $z = \rho + j\varphi$. Por substituição direta em (7.28) em seguida de (7.35) tem-se a expressão

$$\begin{bmatrix} \nabla f_z \\ \nabla f_{z^*} \end{bmatrix} = 2\mathbf{H}^{-1} \begin{bmatrix} \frac{\partial}{\partial z^*} \\ \frac{\partial}{\partial z} \end{bmatrix} = 2 \left(\frac{1}{2} \begin{bmatrix} 1 + \frac{1}{\rho^2} & 1 - \frac{1}{\rho^2} \\ 1 - \frac{1}{\rho^2} & 1 + \frac{1}{\rho^2} \end{bmatrix} \right) \begin{bmatrix} \frac{\partial}{\partial z^*} \\ \frac{\partial}{\partial z} \end{bmatrix}.$$

A primeira componente da expressão aumentada acima é o gradiente natural complexo

$$\nabla f_z = \left(1 + \frac{1}{\rho^2} \right) \frac{\partial f}{\partial z^*} + \left(1 - \frac{1}{\rho^2} \right) \frac{\partial f}{\partial z}, \quad (7.39)$$

com $\rho = \text{Re}\{z\}$ e a presença de ambas derivadas

$$\frac{\partial f}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial \rho} + j \frac{\partial f}{\partial \varphi} \right)$$

e

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial \rho} - j \frac{\partial f}{\partial \varphi} \right)$$

em termos das coordenadas polares. Essas expressões em (7.39) resultam na expressão em termos das coordenadas dadas

$$\nabla f_z = \frac{\partial f}{\partial \rho} + j \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi}. \quad (7.40)$$

A expressão acima, representa a componente da base não conjugada do gradiente complexo, expressa em sua respectiva base por

$$\nabla_{f_z} \frac{\partial}{\partial z} = \left(\frac{\partial f}{\partial \rho} + j \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \right) \cdot \frac{1}{2} \left(\frac{\partial}{\partial \rho} - j \frac{\partial}{\partial \varphi} \right) .$$

A fim de verificar que a expressão acima equivale à forma real polar do gradiente, faz-se uso da estrutura complexa (7.20). Com isso, para obter a expressão em função de ρ e φ , desenvolvendo esta resulta em

$$\begin{aligned} \nabla_{f_z} \frac{\partial}{\partial z} &= \frac{1}{2} \left(\frac{\partial f}{\partial \rho} \frac{\partial}{\partial \rho} - \frac{\partial f}{\partial \rho} J \frac{\partial}{\partial \varphi} + \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} J \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \frac{\partial}{\partial \varphi} \right) \\ &= \frac{\partial f}{\partial \rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \frac{\partial}{\partial \varphi} , \end{aligned}$$

a expressão para o gradiente real em coordenadas polares, para bases não normalizadas $\{\partial/\partial \rho, \partial/\partial \varphi\}$ como em (Amari; Douglas, 1998). A expressão para o gradiente natural complexo

$$z_{k+1} = z_k - \mu_k \nabla f_z ,$$

dado em (7.39) possui passo de atualização é

$$\delta z = -\mu \nabla f_z ,$$

para um dado instante discreto k . Como o gradiente pode ser expresso por (7.40) , então o passo $\delta z = \delta \rho + j \delta \varphi$ pode ser apresentado também como

$$\delta \rho + j \delta \varphi = -\mu \left(\frac{\partial f}{\partial \rho} + j \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \right) ,$$

que é a versão real implícita na notação complexa, como é o caso de coordenadas aumentadas (7.39).

É importante destacar a diferença entre o uso anterior, já em coordenadas polares, para o caso em que $z = x(\rho, \varphi) + jy(\rho, \varphi) = \rho \cos \varphi + j \rho \sin \varphi$ que é uma representação cartesiana com uso de coordenadas polares. Nesse caso pode-se utilizar o termo para a componente do gradiente

$$\nabla_z f = \frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} ,$$

que tem a mudança de coordenada para o caso complexo pelo jacobiano

$$\begin{bmatrix} \nabla_z \\ \nabla_{z^*} \end{bmatrix} = \begin{bmatrix} \cos \varphi + j \sin \varphi & -\rho \sin \varphi + j \cos \varphi \\ \cos \varphi - j \sin \varphi & -\rho \sin \varphi - j \rho \cos \varphi \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \rho} \\ \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \end{bmatrix} .$$

A expressão anterior de forma compacta é

$$\begin{bmatrix} \nabla_z \\ \nabla_{z^*} \end{bmatrix} = \frac{\partial(z, z^*)}{\partial(x, y)} \cdot \frac{\partial(x, y)}{\partial(\rho, \varphi)} \cdot \nabla_{(\rho, \varphi)} f = \frac{\partial(z, z^*)}{\partial(\rho, \varphi)} \cdot \nabla_{(\rho, \varphi)} f .$$

indicando uma transformação de coordenadas polares para complexas, por meio de uma coordenada auxiliar retangular. A diferença se torna mais evidente ao se escrever o elemento diferencial em cada caso. Enquanto nesta $dz = dx(\rho, \varphi) + jdy(\rho, \varphi)$, a forma proposta de atualização já traz as coordenadas diretamente como $dz = d\rho + j d\varphi$.

7.2.12.3 Representação oblíqua

Todos os exemplos anteriores foram feitos a partir de casos ortogonais, com ($\mathbf{Q} = 0$ na matriz (7.28)). Este exemplo lida com o caso bi-dimensional (u, v) , para $z = u + jv$ em que $Q \neq 0$. A matriz métrica real pode ser escrita por

$$\mathbf{G} = \begin{bmatrix} 1 & \cos \varphi \\ \cos \varphi & 1 \end{bmatrix} ,$$

equivalendo à uma representação $(u + v \cos \varphi, v \sin \varphi)$ para um valor fixo de φ . A partir desta pode-se calcular $\mathbf{H}$ utilizando (7.28) e, em seguida, computar o gradiente complexo aumentado por (7.36)

$$\begin{bmatrix} \nabla f_z \\ \nabla f_{z^*} \end{bmatrix} = 2 \cdot \left(\frac{1}{2} \begin{bmatrix} 2 & 2j \cos \varphi \\ -2j \cos \varphi & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \frac{\partial f}{\partial z^*} \\ \frac{\partial f}{\partial z} \end{bmatrix} .$$

O resultado da primeira linha fornece a componente do gradiente complexo

$$\nabla f_z = \frac{2}{\sin^2 \varphi} \left(\frac{\partial f}{\partial z^*} - j \cos \varphi \frac{\partial f}{\partial z} \right) . \quad (7.41)$$

Procedendo da mesma forma que no exemplo anterior, pela expansão das expressões das bases e uso da estrutura complexa

$$\nabla f_z = \frac{1}{\sin^2 \varphi} \left[\frac{\partial f}{\partial u} - \cos \varphi \frac{\partial f}{\partial v} + j \left(-\cos \varphi \frac{\partial f}{\partial u} + \frac{\partial f}{\partial v} \right) \right] ,$$

a versão utilizando as coordenadas reais. Aqui o incremento para pequenos deslocamentos, proveniente do elemento diferencial, é $\delta z = \delta u + i \delta v$ e tem-se as regras equivalentes e atualização dadas por

$$\delta z = -\mu \nabla f_z$$

para (7.41), ou por

$$\delta u + i\delta v = -\mu \frac{1}{\sin^2 \varphi} \left[\frac{\partial f}{\partial u} - \cos \varphi \frac{\partial f}{\partial v} + j \left(-\cos \varphi \frac{\partial f}{\partial u} + \frac{\partial f}{\partial v} \right) \right]. \quad (7.42)$$

O resultado em (7.42) indica que as componentes real e imaginária do gradiente natural complexo em termos das variáveis (u, v) são exatamente o resultado do caso real bidimensional: expressá-lo como (7.41) representa o mesmo gradiente escrito em variáveis e bases complexas.

7.2.13 Considerações sobre o método

Como demonstra o presente capítulo, juntamente com as aplicações descritas na seção 7.2.12, o gradiente natural complexo é corrigido pela inversa da métrica e possui componentes conjugada e não-conjugada de derivadas. Isto foi possível ao definir o gradiente como a versão vetorial da derivada exterior complexa de uma função real.

Mesmo no caso cartesiano, a matriz leva a um fator de escala de 2. Esse fator não surte efeito em casos em cálculos de pontos críticos (pontos em que o gradiente é igualado a zero), como em $\partial f / \partial z^* = 0$ para uma função real, mas pode ser importante na taxa de atualização quando utiliza um gradiente. Em (Brandwood, 1983) há uma certa confusão sobre o fator de escala do caso retangular estar associado à alguma escolha como “ $\partial z / \partial z = 2$ ”, enquanto na verdade fator multiplicativo vem da métrica $\tilde{\mathbf{H}}$ (dada $\mathbf{G}$ cartesiana). A relação $\partial z / \partial z = 1$ é sempre verdade como pode ser verificada em (7.15).

Para os casos polar e oblíquo, foi visto que as derivadas $\partial f / \partial z^*$ e $\partial f / \partial z$ estão presentes na expressão do gradiente. Fato semelhante ao método de Newton; este para a inversa do Hessiano, assim para o gradiente natural para a inversa da métrica mas ambas com a forma

$$\nabla f_{\mathbf{z}} = \mathbf{A} \cdot \frac{\partial f}{\partial \mathbf{z}^H} + \mathbf{B} \cdot \frac{\partial f}{\partial \mathbf{z}^T},$$

para matrizes $\mathbf{A}$ e $\mathbf{B}$. O gradiente complexo é a expressão que preserva distâncias (métrica) para qualquer sistema de coordenadas escolhido.

8 CONCLUSÃO

Este trabalho teve como objetivo o estudo de sinais complexos levando em conta as não-circularidades presentes em seus perfis. A partir de uma introdução teórica matemática em variáveis complexas, foi apresentado o cálculo $\mathbb{CR}$, que é uma poderosa ferramenta na análise de funções de otimização com parâmetros complexos. Logo após, foi apresentado o conceito de transformações amplamente lineares, juntamente com algumas propriedades de espaços vetoriais aumentados. Uma breve caracterização matemática sobre variáveis aleatórias complexas e sua relação com o conceito de não circularidade foi fornecida, sendo importante na caracterização de processos estatísticos que regem o estudo do presente trabalho, com uso em conceitos de filtragem amplamente linear.

Foi possível modelar filtros ótimos amplamente lineares (5.26), que possuem a mesma forma matemática daquela exibida em filtros lineares convencionais. O algoritmo WLLSM (5.30) foi apresentado, mantendo uma escrita similar ao padrão linear quando em notação aumentada. Em sequência, foi destacado que a vantagem da descrição amplamente linear não é apenas ter uma forma similar à linear, mas também apresentar um pequeno acréscimo na melhora do desempenho (em relação ao erro), demonstrados com auxílio de simulações.

O método de Newton aumentado foi apresentado e comparado com o método usual complexo. O método foi adaptado com auxílio do lema da inversão de matrizes, para estimar a inversa da correlação, exibindo uma leve melhora no desempenho em relação à minimização do erro quadrático médio em todos os casos simulados em comparação ao método complexo usual.

Por último, foi proposto o algoritmo do gradiente natural aumentado, que é a representação do gradiente natural em forma complexa, permitindo a manutenção da escrita complexa utilizando uma métrica aumentada (que representa uma métrica real). Algumas realizações nessa notação foram tomadas como exemplo, demonstrando a equivalência de representações.

REFERÊNCIAS

- ACHESON, D. **Elementary Fluid Dynamics**. Clarendon Press, 1990. (Elementary Fluid Dynamics). ISBN 9780198596608. Disponível em: <https://books.google.com.br/books?id=gIZXvwEACAAJ>.
- ADALI, T.; HAYKIN, S. **Adaptive Signal Processing: Next Generation Solutions**. New York: Hoboken, NJ, Wiley, 2010.
- AMAR, A.; ROUTTENBERG, T. Widely-linear mmse estimation of complex-valued graph signals. **arXiv:2208.10588**, 2022.
- AMARI, S.; DOUGLAS, S. Why natural gradient? In: **Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)**. [S. l.: s. n.], 1998. v. 2, p. 1213–1216 vol.2.
- AMARI, S.; NAGAOKA, H. **Methods of Information Geometry**. American Mathematical Society, 2000. (Translations of mathematical monographs). ISBN 9780821843024. Disponível em: <https://books.google.com.br/books?id=vc2FWS07wLUC>.
- AMBLARD, P.; GAETA, M.; LACOUME, J. Statistics for complex variables and signals — part i: Variables. **Signal Processing**, v. 53, n. 1, p. 1–13, 1996. ISSN 0165-1684. Disponível em: <https://www.sciencedirect.com/science/article/pii/0165168496000710>.
- BLUHM, R. C.; CAVALCANTE, C. C. Wirtinger calculus and complex natural gradient algorithm. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS. **Anais [...]**. Sta. Rita do Sapucaí, MG: SBrT, 2022.
- BRANDWOOD, D. A complex gradient operator and its application in adaptive array theory. **IEE Proceedings H (Microwaves, Optics and Antennas)**, v. 130, p. 11–16(5), February 1983. ISSN 0143-7097. Disponível em: <https://digital-library.theiet.org/content/journals/10.1049/ip-h-1.1983.0004>.
- BROWN, J. W.; CHURCHILL, R. V. **Complex Variables and Applications**. New York: McGraw-Hill Education, 2013. ISBN 0073383171.
- CARMO, M. P. do. **Geometria riemanniana**. New York: IMPA, 2008. ISBN 0073383171.
- GREENE, R.; KRANTZ, S.; SOCIETY, A. M. **Function Theory of One Complex Variable**. [S. l.]: American Mathematical Society, 2006. (Graduate studies in mathematics). ISBN 9780821839621.
- GRIFFITHS, D. J. **Introduction to electrodynamics**. [S. l.]: Pearson, 2013.
- KREUTZ-DELGADO, K. **The Complex Gradient Operator and the CR-Calculus**. 2009.
- LI, X.-L.; ADALI, T. Complex-valued linear and widely linear filtering using mse and gaussian entropy. **Trans. Sig. Proc.**, IEEE Press, v. 60, n. 11, p. 5672–5684, nov 2012. ISSN 1053-587X. Disponível em: <https://doi.org/10.1109/TSP.2012.2210889>.
- LIMA, E. L. **Álgebra Linear**. Rio de Janeiro: IMPA, 2008. ISBN 978-85-244-0089-6.

MANDIC, D.; GOH, V. **Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models**. Wiley, 2009. (Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control). ISBN 9780470742631. Disponível em: <https://books.google.com.br/books?id=MaW8MalkztUC>.

MONZINGO, R.; MILLER, T. **Introduction to adaptive arrays**. München: Wiley-Interscience, 1980.

NITTA, T. Natural gradient descent for training stochastic complex-valued neural networks. **International Journal of Advanced Computer Science and Applications**, The Science and Information Organization, v. 5, n. 7, 2014. Disponível em: <http://dx.doi.org/10.14569/IJACSA.2014.050729>.

NOSSEK, J. A. **Adaptive and Array Signal Processing**. München: Technische Universität München, 2015.

OLLILA, E. On the circularity of a complex random variable. **IEEE Signal Processing Letters**, v. 15, p. 841–844, 2008.

OPPENHEIM, A.; WILLISKY, A.; YOUNG, I. **Signals and Systems**. [S. l.]: Prentice-Hall, 1983. (Prentice-Hall signal processing series). ISBN 9780138097318.

PICINBONO, B.; CHEVALIER, P. Widely linear estimation with complex data. **IEEE Transactions on Signal Processing**, v. 43, n. 8, p. 2030–2033, 1995.

QIU, C.; QIAN, G.; WANG, S. Widely linear maximum complex correntropy criterion affine projection algorithm and its performance analysis. **IEEE Transactions on Signal Processing**, v. 70, p. 3540–3550, 2022.

SCHREIER, P. J.; SCHARF, L. L. **Statistical Signal Processing of Complex-Valued Data**. Cambridge: Cambridge University Press, 2010. ISBN 978-0-511-67772-4.

WIDROW, B.; MCCOOL, J.; BALL, M. The complex lms algorithm. **Proceedings of the IEEE**, v. 63, n. 4, p. 719–720, 1975.

YADAV, R.; VENKATASUBRAMANI, L. N.; KOILPILLAI, R. D.; VENKITESH, D. Widely linear filtering for multiimpairment compensation in dispersion managed mqam modulated optical systems. **IEEE Access**, v. 10, p. 73278–73293, 2022.