

DISSERTAÇÃO DE MESTRADO Nº 887

**PROCEDIMENTO PARA DETECÇÃO DE FALSEAMENTO VIA AMOSTRAGEM
NÃO UNIFORME**

Fabio Mafra Kunoh

DATA DA DEFESA: 25/09/2015

Universidade Federal de Minas Gerais

Escola de Engenharia

Programa de Pós-Graduação em Engenharia Elétrica

**PROCEDIMENTO PARA DETECÇÃO DE FALSEAMENTO VIA
AMOSTRAGEM NÃO UNIFORME**

Fabio Mafra Kunoh

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Prof. Leonardo Antônio Borges Tôrres

Belo Horizonte - MG

Setembro de 2015

**"Procedimento para Detecção de Falseamento via Amostragem
Não Uniforme"**

Fabio Mafra Kunoh

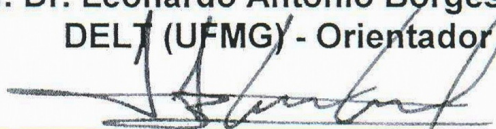
Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica.

Aprovada em 25 de setembro de 2015.

Por:



Prof. Dr. Leonardo Antônio Borges Tôres
DELT (UFMG) - Orientador



Prof. Dr. Luis Antonio Aguirre
DELT (UFMG)



Prof. Dr. Adriano Vilela Barbosa
(UFMG)



Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes
DELT (UFMG)

K96p

Kunoh, Fábio Mafra.

Procedimento para detecção de falseamento via amostragem não uniforme [manuscrito] / Fábio Mafra Kunoh. – 2015.
xvii, 68 f., enc.: il.

Orientador: Leonardo Antônio Borges Tôrres.
Coorientador: Luis Antonio Aguirre.

Dissertação (mestrado) Universidade Federal de Minas Gerais,
Escola de Engenharia.

Bibliografia: f. 66-68.

1. Engenharia elétrica - Teses. 2. Amostragem - Teses.
3. Processamento de sinais - Teses. I. Tôrres, Leonardo Antônio Borges.
II. Aguirre, Luis Antonio. III. Universidade Federal de Minas Gerais. Escola
de Engenharia. IV. Título.

CDU: 621.3(043)

Área de Concentração: Sinais e Sistemas
Linha de Pesquisa: Controle, Automação e Robótica
Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG Brasil



Procedimento para Detecção de Falseamento via Amostragem Não Uniforme

Fábio Mafra Kunoh

Dissertação submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do grau de Mestre em Engenharia Elétrica.

Orientadores: Dr. Leonardo Antônio Borges Tôrres
Dr. Luis Antonio Aguirre

Belo Horizonte, 25 de Setembro de 2015

Agradecimentos

Aos meus amigos que sempre estiveram presentes ajudando não somente com os desafios acadêmicos, mas também trazendo alegria sempre que presentes.

Aos meus professores e orientadores Leonardo Tôrres e Luis Aguirre que não só me guiaram nessa jornada de forma ética e profissional, mas também com um cuidado que me fez sentir como se fosse parte de uma família.

Aos meus pais e irmão que sempre me deram todo o suporte e apoio para perseguir o meu próprio caminho.

À Nanda que, com muito amor, nunca deixou de acreditar em mim e não permitiu que eu deixasse também.

Epígrafe

“A verdadeira viagem de descobrimento não consiste em procurar novas paisagens, e sim em ter novos olhos.”

Marcel Proust

Conteúdo

Agradecimentos	ii
Resumo	vii
Abstract	viii
Lista de Tabelas	x
Lista de Figuras	xv
Lista de Símbolos	xvi
Lista de Abreviações	xvii
1 Introdução	1
1.1 Motivação	2
1.2 Objetivos	3
1.3 Estrutura da dissertação	4
2 Amostragem Não Uniforme no Domínio do Tempo	6
2.1 Falseamento (<i>Aliasing</i>)	6
2.2 Determinação dos Instantes de Amostragem Não Uniformes	7
2.2.1 Amostragem com <i>Jitter</i> Estocástico (<i>SJS</i>)	7

2.2.2	Sorteio Aleatório de Instante de Amostragem	10
2.2.3	Processo Aditivo Aleatório (<i>ARS</i>)	10
2.2.4	Decimação Aleatória	11
3	Análise da Amostragem Não Uniforme no Domínio da Frequência	12
3.1	A Transformada Discreta Não Uniforme de Fourier (NDFT)	12
3.2	Ruído da Linha de Base	14
3.2.1	Transformada Discreta de Fourier Revisitada	14
3.2.2	Transformada Discreta Não Uniforme de Fourier	17
3.3	Influência das Variáveis de Amostragem no Ruído da Linha de Base	19
3.3.1	Influência da Taxa de amostragem Média	19
3.3.2	Influência do Tamanho da Janela Amostrada	21
3.3.3	Influência dos Componentes de Frequência de um Sinal	23
3.3.4	Limite na Amostragem <i>Missing Data</i>	23
3.3.5	<i>Jitter</i> e Limitações da NDFT na Amostragem Não Uniforme	25
4	Detecção Automática de Falseamento	27
4.1	Características Estocásticas dos Tipos de Amostragem Não Uniforme	27
4.1.1	Função Densidade de Probabilidade de t_k	28
4.2	Estimadores do Espectro	29
4.3	O Teste de Hipótese	35
4.3.1	Função Densidade de Probabilidade das Grandezas Estimadas	35
4.3.2	Teste de Hipóteses Bidimensional	36

4.3.3	Algoritmo para Detecção de Falseamento	38
5	Resultados Numéricos	41
5.1	Simulações	41
5.1.1	Primeiro Teste	41
5.1.2	Teste em Número Elevado de Sinais	48
5.2	Detecção de Frequências Acima de $f_s/2$	54
6	Conclusões	63
6.1	Pesquisas Futuras	64
	Bibliografia	68

Resumo

O teorema da amostragem estabelece um limite inferior para a taxa de amostragem necessária para garantir a perfeita representação de um sinal contínuo por uma sequência enumerável de amostras uniformemente espaçadas no tempo. Quando o limite não é atendido, ocorre superposição espectral, fenômeno conhecido como falseamento ou *aliasing*. A amostragem não uniforme não está sujeita ao limite estabelecido pelo teorema da amostragem, mas é influenciada por outras limitações de caráter prático. Este trabalho investiga algumas dessas limitações e propõe um procedimento para verificar se um sinal com conteúdo espectral desconhecido está, após a amostragem uniforme, falseado ou não. O procedimento é avaliado por meio de simulações, imaginando a implementação do algoritmo em processadores digitais com restrições de memória pré-estabelecidas. Os resultados indicam que é possível detectar, com uma alta taxa de sucesso, o falseamento, desde que alguns requisitos de tamanho amostral e razão sinal-ruído sejam atendidos.

Abstract

The sampling theorem determines a lower bound for the sampling frequency necessary for the perfect representation of a continuous signal by a countable set of uniformly spaced time samples. If such a bound is not satisfied the resulting spectrum is said to be aliased. Nonuniform sampling is not bound by the sampling theorem, although it is influenced by some practical limitations. This work investigates some of these limitations and proposes a procedure to check if a signal with unknown spectrum is aliased after uniform sampling. Such procedure is evaluated by means of simulation and is discussed in this work in the context of an implementation in digital signal processors with memory restrictions. Results show that alias detection is possible, with a high rate of success, if sample size and signal to noise ratio requirements are met.

Lista de Tabelas

5.1	Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Sem Vazamento Espectral	50
5.2	Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas - Sem Vazamento Espectral	50
5.3	Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Com Vazamento Espectral	50
5.4	Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas - Com Vazamento Espectral	51
5.5	Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Pós Janelamento	51
5.6	Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas Pós Janelamento	51
5.7	Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Pós Filtragem de R_Y	54
5.8	Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas - Pós Filtragem de R_Y . . .	54
5.9	Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Sem Vazamento Espectral . .	59

5.10	Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Sem Vazamento Espectral	60
5.11	Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Com Vazamento Espectral . .	60
5.12	Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Com Vazamento Espectral	60
5.13	Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Pós Janelamento	61
5.14	Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Pós Janelamento	61
5.15	Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Pós Filtragem	61
5.16	Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Com Vazamento Espectral	61

Lista de Figuras

1.1	<i>Visão geral da arquitetura de um possível instrumento digital no qual o algoritmo de detecção automática de falseamento poderia ser executado.</i>	4
2.1	<i>Falseamento em sinal temporal. Sinais $x(t) = \cos(2\pi 15t)$ (linha contínua) e $y(t) = \cos(2\pi 10t)$ (linha pontilhada) sendo amostrados nos mesmos instantes temporais.</i>	7
2.2	<i>Variação do instante de amostragem no eixo temporal - Método SJS.</i>	8
2.3	<i>Transformadas de Fourier para sinal amostrado com método SJS, para diferentes limitações de τ_k. $T_J = 10$ s. $\bar{f}_s = 10$ Hz. $x(t) = \cos(2\pi 7kT)$. A) $\tau_k = 0$, B) $\tau_k \leq T/40$, C) $\tau_k \leq T/20$, D) $\tau_k \leq T/8$, E) $\tau_k \leq T/4$, F) $\tau_k \leq T/2$.</i>	9
3.1	<i>Sinal $X(nT_0) = \cos(2\pi nT_0) \cos(2\pi nT_0)$ amostrado uniformemente.</i>	15
3.2	<i>Sinal $X(nT_0) = \cos(2\pi nT_0) \cos(6\pi nT_0)$ amostrado uniformemente.</i>	16
3.3	<i>Sinal $X(nT_0) = \cos(2\pi nT_0) \cos(6\pi nT_0)$, amostrado não uniformemente.</i>	18

- 3.4 $|X(\omega_m)|$ para diferentes valores de \bar{f}_s . a) 64 amostras/segundo.
 b) 256 amostras/segundo. c) 1024 amostras/segundo. d) 4096
 amostras/segundo. Sinal avaliado: $x(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 =$
 16 Hz , $f_2 = 210 \text{ Hz}$, $f_3 = 876 \text{ Hz}$; $T_J = 1 \text{ s}$. Método de amos-
 tragem: Sorteio Aleatório. 20
- 3.5 $|X(\omega_m)|$ de sinal amostrado com $\bar{f}_s = 256$ amostras/segundo.
 Sinal avaliado: $x(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 16 \text{ Hz}$, $f_2 = 210 \text{ Hz}$, $f_3 =$
 876 Hz ; $T_J = 1 \text{ s}$. Método de amostragem: Sorteio Aleatório. . . 21
- 3.6 $|X(\omega_m)|$ para diferentes tamanhos de janela temporal. a) T_J
 $= 1$ segundo. b) $T_J = 3$ segundos. c) $T_J = 6$ segundos. d) T_J
 $= 12$ segundos. Sinal avaliado: $x(t) = \sum_{i=1}^5 \cos(2\pi f_i t)$, $f_1 =$
 5 Hz , $f_2 = 30 \text{ Hz}$, $f_3 = 48 \text{ Hz}$, $f_4 = 80 \text{ Hz}$, $f_5 = 250 \text{ Hz}$; $\bar{f}_s =$
 256 Hz . Método de amostragem: Sorteio Aleatório. 22
- 3.7 $|X(\omega_m)|$ para diferentes sinais. a) $x(t) = \cos(2\pi f_1 t)$, b) $x(t) =$
 $\sum_{i=1}^3 \cos(2\pi f_i t)$, c) $x(t) = \sum_{i=1}^5 \cos(2\pi f_i t)$, d) $x(t) = \sum_{i=1}^7 \cos(2\pi f_i t)$,
 $f_1 = 72 \text{ Hz}$, $f_2 = 53 \text{ Hz}$, $f_3 = 112 \text{ Hz}$, $f_4 = 5 \text{ Hz}$, $f_5 = 47 \text{ Hz}$, $f_6 =$
 21 Hz , $f_7 = 98 \text{ Hz}$. $T_J = 1 \text{ s}$ $\bar{f}_s = 256 \text{ Hz}$. Método de amos-
 tragem: Sorteio Aleatório. 24
- 3.8 NDFT do sinal $x(t) = \cos(2\pi 300t)$ amostrado não uniforme-
 mente via M.D., a partir de um sinal com $f_s = 1000$ amos-
 tras/segundo e taxa de decimação = 10. 25
- 4.1 Valor Esperado $|\mu_W(f)|$ (linha cheia) e a margem de confiança
 $|\mu_W \pm \sqrt{R_W}|$ (linha tracejada) de $W(f)$. $N=16$ 34
- 4.2 Valor Esperado $|\mu_W(f)|$ (linha cheia) e a margem de confi-
 ança $|\mu_W \pm \sqrt{R_W}|$ (linha pontilhada) de $W(f)$. A) $N = 128$
 amostras. B) $N = 512$ amostras 34

- 4.3 *Distribuição Gaussiana bivariada e, em vermelho, elipse que corresponde ao nível de confiança de 95%.* 37
- 4.4 *Distribuição Gaussiana bivariada, em que componentes real e imaginária não estão correlacionadas e, em vermelho, elipse que corresponde ao nível de confiança de 95%.* 38
- 4.5 *Fluxograma do procedimento de detecção de falseamento. . . .* 40
- 5.1 $|\widehat{Y}(f)|$. $T_J = 0,25\text{ s}$ $f_s = 1024\text{ Hz}$. $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32\text{ Hz}$, $f_2 = 208\text{ Hz}$, $f_3 = 604\text{ Hz}$ 42
- 5.2 *Valor esperado $|\mu_Y(f)|$ e banda de confiança $|\mu_Y \pm \sqrt{R_Y}|$, calculados usando Equações 4.14 e 4.15. $T_J = 0,25\text{ s}$, $\bar{f}_s = 1024\text{ Hz}$, método ARS: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32\text{ Hz}$, $f_2 = 208\text{ Hz}$, $f_3 = 604\text{ Hz}$* 43
- 5.3 *Detalhe da Figura 5.2: $|\mu_Y(f)|$ e $|\mu_Y \pm \sqrt{R_Y}|$ do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32\text{ Hz}$, $f_2 = 208\text{ Hz}$, $f_3 = 604\text{ Hz}$* 43
- 5.4 *Valor esperado $|\mu_Y^{MC}(f)|$ e banda de confiança $|\mu_Y^{MC} \pm \sqrt{R_Y^{MC}}|$, determinados via simulação Monte Carlo. Obtido usando 1000 realizações. $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32\text{ Hz}$, $f_2 = 208\text{ Hz}$, $f_3 = 604\text{ Hz}$. $T_J = 0,25\text{ s}$ $\bar{f}_s = 1024\text{ Hz}$, método ARS.* 44
- 5.5 *Detalhe da Figura 5.4: $|\mu_Y^{MC}(f)|$ e $|\mu_Y^{MC} \pm \sqrt{R_Y^{MC}}|$ do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32\text{ Hz}$, $f_2 = 208\text{ Hz}$, $f_3 = 604\text{ Hz}$* 44
- 5.6 $|\widetilde{Y}(f)|$ obtida a partir de 256 amostras do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32\text{ Hz}$, $f_2 = 208\text{ Hz}$, $f_3 = 604\text{ Hz}$. $T_J = 0,25\text{ s}$ $\bar{f}_s = 1024\text{ Hz}$, método ARS. 45
- 5.7 *Resultado do Teste de Hipótese em que o sinal $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32\text{ Hz}$, $f_2 = 208\text{ Hz}$, $f_3 = 604\text{ Hz}$, amostrado a $\bar{f}_s = 1024\text{ Hz}$, é avaliado utilizando-se $\mu_Y(f)$ e $\sigma_{Y^{MC}}(f)^2$* 46

- 5.8 *Resultado do Teste de Hipótese em que o sinal $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32$ Hz, $f_2 = 208$ Hz, $f_3 = 604$ Hz, amostrado a $\bar{f}_s = 1024$ Hz é avaliado utilizando-se $\mu_Y^{MC}(f)$ e $\sigma_{YMC}(f)^2$ 47*
- 5.9 *Comparação entre resultados de Teste de Hipótese. Sinal avaliado: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32$ Hz, $f_2 = 208$ Hz, $f_3 = 604$ Hz, amostrado a $\bar{f}_s = 1024$ Hz. Linha contínua azul: Teste é avaliado utilizando-se $\mu_Y(f)$ e $\sigma_{YMC}(f)^2$. Linha tracejada verde: Teste é avaliado utilizando-se $\mu_Y^{MC}(f)$ e $\sigma_{YMC}(f)^2$. . . 48*
- 5.10 *Transformadas do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 40,6$ Hz, $f_2 = 416,8$ Hz, $f_3 = 643,2$ Hz, de uma realização com $N = 512$ Amostras e $SNR = 20$ dB. a) $|\hat{Y}(f)|$, $f_s = 1024$ Hz. b) $|\tilde{Y}(f)|$, $\bar{f}_s = 1024$ Hz. 52*
- 5.11 *Resultado do teste de hipótese para o sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 40,6$ Hz, $f_2 = 416,8$ Hz, $f_3 = 643,2$ Hz, de uma realização com $N = 512$ Amostras e $SNR = 20$ dB. 53*
- 5.12 *$|\hat{Y}(f)|$ obtida a partir de 256 amostras do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 344$ Hz, $f_2 = 712$ Hz, $f_3 = 1296$ Hz. $f_s = 1024$ Hz, $SNR = 20$ dB. 56*
- 5.13 *$|\tilde{Y}(f)|$ obtida a partir de 256 amostras do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 344$ Hz, $f_2 = 712$ Hz, $f_3 = 1296$ Hz. $\bar{f}_s = 1024$ Hz, $SNR = 20$ dB, método ARS 56*
- 5.14 *$|\hat{Y}(f)|$ expandida até f_{final} , obtida a partir de 256 amostras do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 344$ Hz, $f_2 = 712$ Hz, $f_3 = 1296$ Hz $f_s = 1024$ amostras/segundo, $SNR = 20$ dB. 57*
- 5.15 *$\mu_W(f)$ obtida a partir de 256 amostras - $\bar{f}_s = 1024$ 58*
- 5.16 *$\mu_Y(f)$ e margem de confiança expandidas até f_{final} . Obtida a partir de 256 amostras do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 =$*

- 344 Hz, $f_2 = 712$ Hz, $f_3 = 1296$ Hz, $\bar{f}_s = 1024$ amostras/segundo,
SNR = 20dB. 58
- 5.17 Resultado do teste de hipótese do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 =$
 344 Hz, $f_2 = 712$ Hz, $f_3 = 1296$ Hz, $\bar{f}_s = 1024$ amostras/segundo,
SNR = 20dB. 59

Lista de Símbolos

$x(t)$	Função matemática no domínio do tempo;
$X(f)$	Função matemática no domínio da frequência;
ω	Frequência angular;
Ω_0	Frequência angular fundamental;
T	Período de amostragem;
f_s	Frequência de amostragem;
\bar{f}_s	Frequência média de amostragem;
T_j	Duração da janela de amostragem;
$E[]$	Operador Esperança Matemática;
$\text{Cov}[]$	Operador Covariância;
$\varphi_m(f)$	Função Característica;
$W(f)$	Janela Estocástica de Frequência;
$\mu_Y(f)$	Valor Esperado do Vetor Aleatório Y;
$R_Y(f, \psi)$	Matriz de Covariância do Vetor Aleatório Y;
$\hat{Y}(f)$	FFT de sinal amostrado uniformemente;
$\tilde{Y}(f)$	NDFFT de sinal amostrado não uniformemente;

Lista de Abreviações

<i>DFT</i>	<i>Discrete Fourier Transform</i>
<i>FFT</i>	<i>Fast Fourier Transform</i>
<i>NDFT</i>	<i>Non-Uniform Discrete Fourier Transform</i>
<i>NUT-DFT</i>	<i>Non-Uniform Time Discrete Fourier Transform</i>
<i>RMS</i>	<i>Root Mean Square</i>
<i>SJS</i>	<i>Stochastic Jitter Sampling</i>
<i>ARS</i>	<i>Additive Random Sampling</i>
<i>MD</i>	<i>Missing Data</i>
<i>SNR</i>	<i>Signal to Noise Ratio</i>

Introdução

O mundo é composto de um enorme número de atividades que podem ser interpretadas como sinais. Sinais podem ser processados via eletrônica analógica, mas, para diversas aplicações, são evidentes as vantagens de sua digitalização. Ainda assim, muitas aplicações ainda lidam com sinais em sua forma analógica, devido a limitações existentes na mudança desses para o domínio digital. A um dos principais fenômenos limitadores da digitalização é dado o nome de falseamento da informação ou *Aliasing*, que surge quando, devido a uma amostragem com taxa indevidamente baixa, ocorre superposição espectral.

O teorema da amostragem de *Nyquist-Shannon* estabelece que um sinal analógico, limitado em banda, ao ser amostrado em intervalos regulares, pode ser perfeitamente recuperado a partir de suas amostras, desde que a frequência de amostragem seja maior que o dobro da maior frequência do sinal. De forma a atender o critério de largura de banda e evitar esse tipo de corrupção de sinal, a prática geralmente adotada é a utilização de um filtro analógico passa-baixas, chamado de filtro *anti-aliasing*, que descarta a informação de alta frequência, permitindo apenas a análise e operação de conteúdos espectrais inferiores a um certo limite.

Apesar de ser utilizada por padrão na maioria dos sinais avaliados, a amostragem uniforme não é indispensável para a análise desses no domínio da frequência. Como detalhadamente descrito por Bilinskis (2007), existindo a possibilidade de amostrar um sinal de forma não uniforme, mesmo que seu conteúdo espectral exceda várias vezes a taxa média de amostragem, não ocorrerá o falseamento. Isso permite uma taxa de amostragem média e armazenamento menos exigentes no caso da amostragem não uniforme em comparação à uniforme, ainda que seja necessário o armazenamento dos instantes temporais das amostras. Entretanto, sinais que são amostrados com intervalo entre amostras assumindo valores temporais distintos, geralmente, são problemáticos, uma vez que as ferramentas tradicionais de análise não são próprias para o tratamento desse tipo de dados. Para lidar com sinais

amostrados não uniformemente, existem ferramentas não convencionais específicas não muito difundidas nas áreas de Engenharia. Entretanto, tais ferramentas possuem limitações distintas da transformada uniforme, geralmente decorrentes de definições de caráter prático.

O foco deste trabalho reside no estudo e desenvolvimento de um procedimento que, utilizando dados amostrados, tanto de forma uniforme quanto não uniforme no tempo, seja capaz de realizar uma detecção automática de falseamento.

1.1 Motivação

A amostragem não uniforme de sinais é um tópico que vem sendo estudado há várias décadas, sendo um dos trabalhos mais antigos na área o desenvolvido por Shapiro e Silverman (1960), que apresentaram tipos de amostragem não uniforme e técnicas para estimação da densidade de potência espectral. Entretanto, esta é uma área de pesquisa muito ativa, não só no estudo direto do assunto, como também em diversas aplicações, como exemplo, ótica (Xu et al., 2015), ressonância magnética (Hoch et al., 2014) e comunicações (Chen et al., 2014), dentre muitas outras.

Publicações relacionadas à amostragem não uniforme possuem diversos focos, sendo muito comum encontrar trabalhos relacionados com reconstrução de sinais (Maymon et al., 2011; Venkamani et al., 2001; Aldroubi e Gröchenig, 2001), desenvolvimento de transformadas rápidas (Kunis e Potts, 2008; Mednieks, 1999), o estudo de diferentes tipos de amostragem e estimação espectral (Bland e Tarczynski, 1997; Boyi e Deyun, 2012).

Dois pontos, entretanto, não possuem um desenvolvimento muito claro nas publicações encontradas e são focos desta dissertação. O primeiro está relacionado com um ruído espectral que surge sempre que se calcula a transformada de um sinal amostrado de forma não uniforme para o domínio da frequência. É comum tal aspecto ser citado, porém não é muito comum encontrar referências que expliquem sua origem e quais fatores influenciam a variância de tal ruído.

O segundo ponto não muito comum na literatura é a abordagem da amostragem não uniforme do ponto de vista de instrumentação e seus procedimentos para detecção de falseamento. Os principais trabalhos com esse foco foram desenvolvidos por Artyukh et al. (1997a), Artyukh et al. (1997b), Mednieks et al. (1997) e Artyukh et al. (2001), mas ainda assim, o trata-

mento é superficial e não há uma descrição de um procedimento que seja detalhado, com descrições de etapas e de decisões bem definidas, o que é o foco deste trabalho. Os quatro trabalhos citados descrevem ferramentas de um mesmo instrumento, chamado *DASP-Lab System*. Os três primeiros são focados na descrição da interface e nos módulos necessários para a implementação desse instrumento, que funciona a partir de uma conexão com um computador. Nos trabalhos são detalhadas as características do conversor analógico-digital para que, a partir de uma taxa de amostragem média de 100 MHz, fosse possível analisar sinais de largura de banda até 1,2 GHz. São também descritos os requerimentos de amostragem, como o menor intervalo permitido entre amostras e o número necessário de amostras para que sinal de interesse não desapareça em níveis elevados de ruído, porém não há aprofundamento matemático que justifique os valores recomendados. O quarto trabalho citado, ainda descrevendo o mesmo instrumento, é um pouco mais detalhado na descrição das características do processamento de sinais DASP (*Digital Alias-free Signal Processing*), mas ainda assim, seu conteúdo é, em boa parte, uma revisão dos demais artigos citados.

Um último ponto a ser citado é que a transformada de Fourier para dados amostrados de maneira não uniforme tem como resultado, para cada uma das frequências, uma variável aleatória. Apesar de ser um conceito claro, são raros os trabalhos que obtêm resultados a partir do desenvolvimento das características estocásticas do sinal obtido. Somente nos trabalhos Tarczynski e Dongdong (2005) e Eng et al. (2008) são desenvolvidas expressões para a avaliação da variância do resultado da transformada de dados não-uniformemente amostrados, sendo o segundo uma das principais referências utilizadas neste trabalho.

1.2 Objetivos

Este trabalho visa o desenvolvimento de um procedimento, passível de implementação em um instrumento, que seja capaz de realizar uma detecção automática de falseamento da informação de um sinal. O intuito é a possível implementação em um instrumento que seja capaz de alternar entre os modos de amostragem uniforme e amostragem não uniforme e, nesse segundo modo, executar o algoritmo de detecção.

Um dos principais focos deste trabalho é a consideração de uma estrutura prática, em que se tem recursos limitados como, por exemplo, uma

quantidade de memória finita. Neste trabalho serão realizadas simulações que consideram diferentes tamanhos de memória para o armazenamento das amostras. Um diagrama ilustrativo da possível ferramenta é apresentado na Figura 1.1.

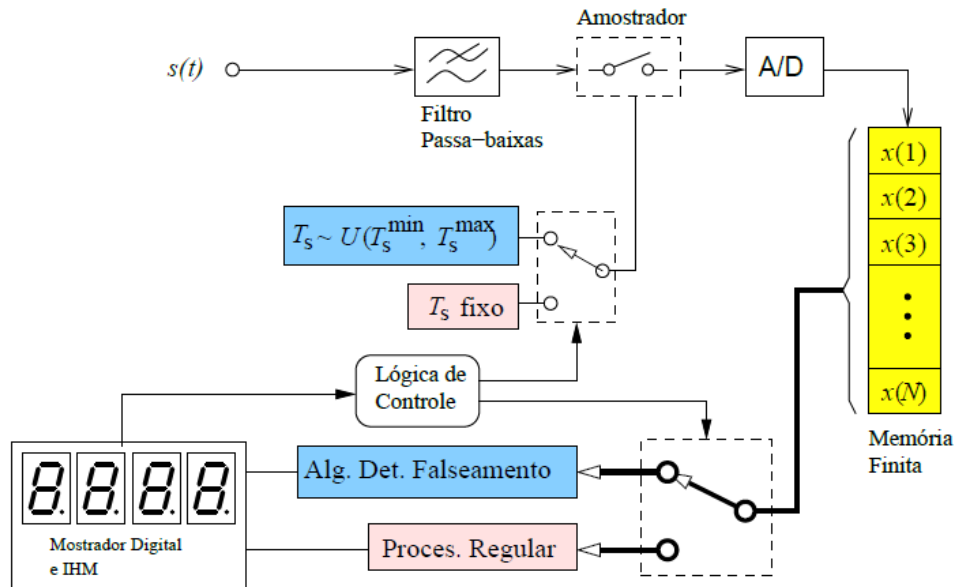


Figura 1.1: *Visão geral da arquitetura de um possível instrumento digital no qual o algoritmo de detecção automática de falseamento poderia ser executado.*

1.3 Estrutura da dissertação

Esta dissertação está organizada da seguinte maneira. No Capítulo 2, são apresentados conceitos referentes à amostragem não uniforme no domínio do tempo. É realizada uma revisão do conceito de falseamento e diferentes tipos de amostragem não uniforme são apresentados. No Capítulo 3, são apresentados conceitos no domínio frequência. A transformada não uniforme de Fourier é apresentada, assim como o ruído da linha de base inerente a sinais amostrados não uniformemente, a sua origem e quais variáveis têm impacto sobre a sua variância. No Capítulo 4, o procedimento proposto é apresentado. Inicialmente, são desenvolvidas as equações dos estimadores de valor esperado e matriz de covariância da transformada de Fourier de sinais amostrados não

uniformemente. Em seguida, é detalhado o teste de hipótese bidimensional utilizado para uma possível rejeição, devido a falseamento, da transformada de Fourier avaliada. No Capítulo 5 as simulações realizadas para avaliar o procedimento proposto são apresentadas e, finalmente, no Capítulo 6, as considerações finais e propostas de pesquisas futuras são descritas.

Amostragem Não Uniforme no Domínio do Tempo

Como este trabalho visa o desenvolvimento de um procedimento que analisa o conteúdo espectral de um sinal amostrado de forma não uniforme, para o seu entendimento, é necessário o desenvolvimento de conceitos tanto no domínio do tempo quanto da frequência. Os conceitos relacionados ao domínio do tempo serão tratados primeiro. Neste capítulo serão descritos os conceitos relacionados à definição de falseamento e a técnicas de amostragem não uniforme.

2.1 Falseamento (*Aliasing*)

Alias é um termo da língua inglesa semelhante ao termo pseudônimo na língua portuguesa. É utilizado quando uma pessoa é conhecida por um nome alternativo que não o seu oficial. Porém, ao invés de ser utilizado para preservar a identidade de alguém, sua função é mais próxima da expressão “também conhecido como”.

Como foi mencionado no capítulo anterior, é necessário, ao amostrar um sinal uniformemente, que a frequência de amostragem seja superior ao dobro da maior frequência do sinal. Tal condição é necessária para que a energia de uma componente de alta frequência não seja interpretada como tendo frequência inferior.

Como exemplo, na Figura 2.1 é apresentado um sinal de 15 Hz que é amostrado a uma taxa de 25 amostras por segundo. Nesse caso, como a frequência de amostragem não é superior a 30 amostras por segundo (o dobro de 15 Hz), o efeito de falseamento ocorre e esse sinal amostrado passa a apresentar uma componente espectral erroneamente associada a frequência de 10 Hz.

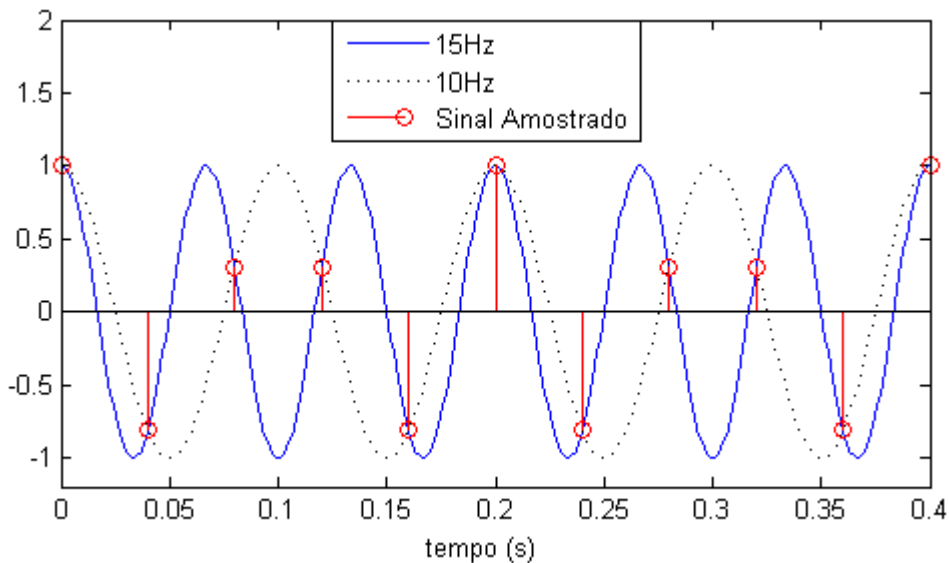


Figura 2.1: *Falseamento em sinal temporal.*

Sinais $x(t) = \cos(2\pi 15t)$ (linha contínua) e $y(t) = \cos(2\pi 10t)$ (linha pontilhada) sendo amostrados nos mesmos instantes temporais.

2.2 Determinação dos Instantes de Amostragem Não Uniformes

Um dos pontos mais importantes no tratamento de dados amostrados de forma não uniforme é a determinação dos instantes de tempo em que ocorrem as medidas. Em muitos casos, não existe o controle dessa variável e os dados precisam ser processados a partir de coletas que ocorreram somente em momentos possíveis. Neste trabalho será considerado que é possível o controle desta variável e que, ou estes instantes de amostragem podem ser definidos antes mesmo do início da coleta, ou pelo menos a característica da distribuição estatística que define a amostragem pode ser determinada ou conhecida. Existem diversas formas de realizar a distribuição de instantes de amostragem e algumas delas serão discutidas ao longo desta seção.

2.2.1 Amostragem com *Jitter* Estocástico (*SJS*)

Considere uma situação em que se tem amostragem uniforme com frequência f_s . As amostras do sinal estarão sempre espaçadas de um intervalo de

valor $T = 1/f_s$, ocorrendo então nos instantes de tempo $0, T, 2T, 3T, \dots$. Uma das formas de se gerar uma distribuição de instantes aleatórios para a realização de uma amostragem não uniforme é utilizar Amostragem com Jitter Estocástico (*Stochastic Jitter Sampling* ou *SJS*). Esta permite a variação dos instantes em torno de valores de referência, que são os instantes da amostragem uniforme. A essa variação do instante de amostragem é dado o nome de *Sampling Noise* (τ_k). Dessa forma, as amostras passam a ocorrer em $0 + \tau_0, T + \tau_1, 2T + \tau_2, 3T + \tau_3, \dots$. Sendo assim, é preciso definir o quão grande pode ser τ_k . A limitação $|\tau_k| \leq T/2$, determina que um instante de amostragem não pode ocorrer antes do instante anterior, nem depois do seguinte. Como exemplo, a Figura 2.2 ilustra um eixo em que os instantes de amostragem podem variar com $|\tau_k| \leq T/4$. As seções mais claras do eixo são os locais nos quais é possível ocorrer amostragem.

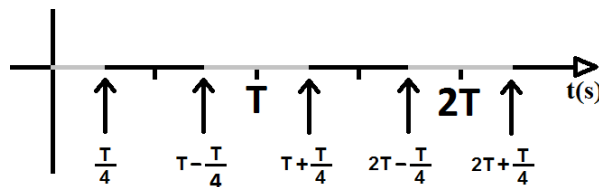


Figura 2.2: Variação do instante de amostragem no eixo temporal - Método SJS.

Utilizando uma ferramenta computacional adequada, que será posteriormente apresentada na Seção 3.1, simulações foram realizadas para diferentes limites para τ_k . A Figura 2.3 mostra a transformada de Fourier de uma senoide de 7 Hz amostrado a uma taxa média (\bar{f}_s) de 10 amostras por segundo com diferentes limites para τ_k . A duração da janela de amostragem foi $T_J = 10$ s.

Para os resultados apresentados com rótulos de *A* a *F*, os limites para τ_k foram definidos como $0, T/40, T/20, T/8, T/4$ e $T/2$. A simulação *A* é como uma amostragem padrão, em que se encontram raias em valores de frequência iguais a $nf_s \pm 7$ Hz. À medida que o intervalo de valores possíveis de τ_k aumenta, a frequência verdadeira do sinal não é mais apresentada como raias falseadas de alta frequência. Na simulação *F*, em que é possível que se tenha amostras em qualquer valor temporal que ainda produza uma sequência monotonicamente crescente de valores de tempo, o resultado obtido não apresenta raias falseadas. É possível observar que, nestas simulações, o valor de 10 amostras por segundo não respeitaria o teorema da amostragem, se as

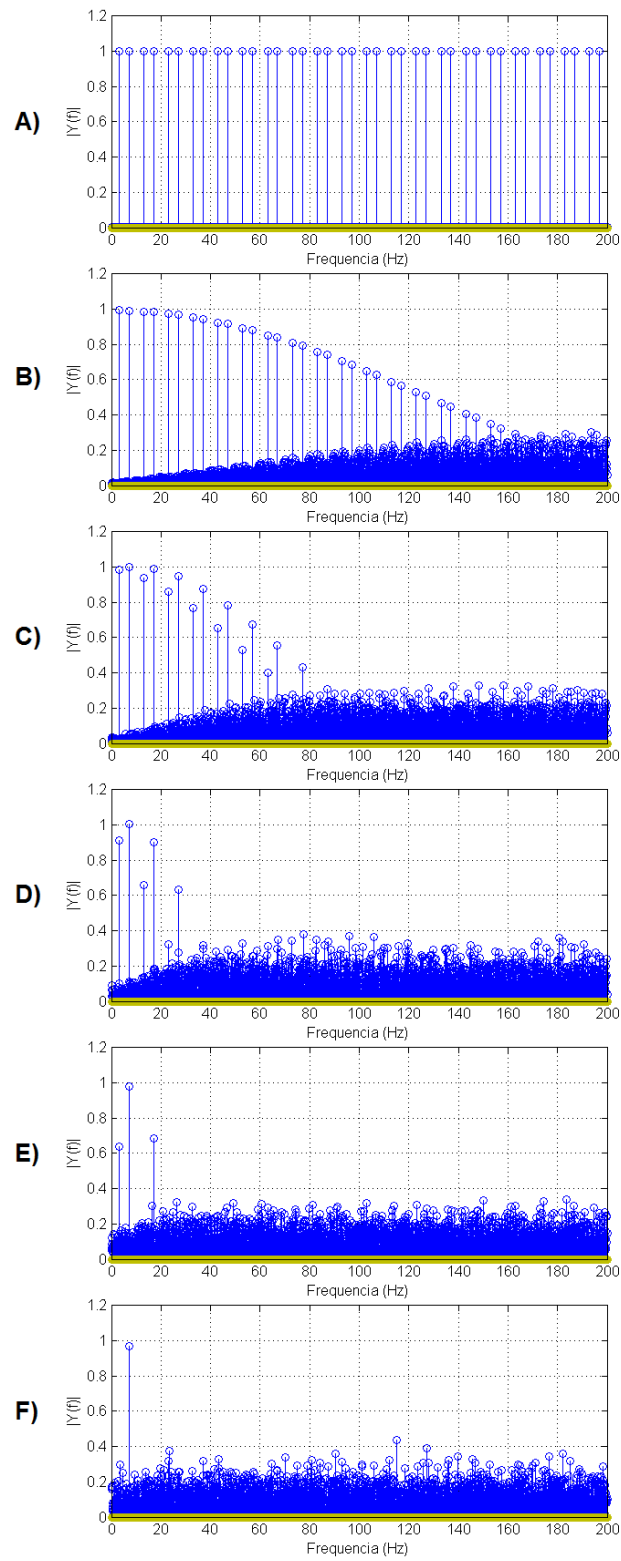


Figura 2.3: Transformadas de Fourier para sinal amostrado com método SJS, para diferentes limitações de $|\tau_k|$. $T_J = 10$ s. $\bar{f}_s = 10$ Hz.

$$x(t) = \cos(2\pi 7kT).$$

$$A) |\tau_k| = 0, \quad B) |\tau_k| \leq T/40, \quad C) |\tau_k| \leq T/20,$$

$$D) |\tau_k| \leq T/8, \quad E) |\tau_k| \leq T/4, \quad F) |\tau_k| \leq T/2.$$

amostras fossem equidistantes. Ainda assim, o valor de 7Hz é avaliado em sua posição correta. Um ponto determinante a ser observado é o *ruído da linha de base* que surge à medida que os sinais falseados desaparecem. Esse efeito surge sempre que os dados amostrados estão distribuídos de forma não uniforme no tempo, independente do método de amostragem, e é um dos fatores mais importantes desta dissertação (se não o mais importante), o que leva à necessidade de uma série de cuidados que serão discutidos ao longo de todo este trabalho.

2.2.2 Sorteio Aleatório de Instante de Amostragem

A determinação dos instantes de amostragem descritos nesta sub-seção é a mais simples e rende resultados semelhantes ao método descrito na Seção 2.2.1 quando é definido o limite $|\tau_k| < T/2$. Ela parte do pressuposto que os instantes de amostragem podem e precisam ser todos definidos anteriormente à coleta de dados. Para uma coleta que ocorra em uma janela de amostragem com duração de T_J segundos, define-se uma taxa de amostragem média, definindo-se o número de amostras medidas N , pela expressão $\bar{f}_s = N/T$. Os instantes de amostragem t_k são sorteados a partir de uma distribuição de probabilidade uniforme $U(0,T)$ e são posteriormente ordenados.

Apesar de esse ser o método de implementação mais simples para simulações, em aplicações reais ele não é muito prático uma vez que é preciso saber exatamente durante quanto tempo ocorrerá a amostragem antes mesmo que ela comece. Ele pode ser utilizado para simulações com o objetivo de obter familiaridade com resultados relacionados à amostragem não uniforme, mas para o desenvolvimento mais aprofundado, os demais métodos serão priorizados.

2.2.3 Processo Aditivo Aleatório (*ARS*)

Para determinadas aplicações, é interessante que os instantes de amostragem possam ser definidos em tempo real e uma forma de obtê-los à medida que são necessários, simultaneamente a uma coleta, por exemplo, é o Processo Aditivo Aleatório (*ARS: Additive Random Sampling*), em que os instantes de amostragem são definidos como:

$$t_k = t_{k-1} + \tau_k, \quad k = 0, 1, 2, \dots \quad (2.1)$$

Dessa forma, um instante de amostragem é definido ao se adicionar uma variável aleatória (*Sampling Noise*) ao instante anterior. A amostragem uniforme pode ser, também, considerada um caso específico desse modelo.

Esse será o principal método utilizado neste trabalho, uma vez que as características estocásticas das variáveis (instante de amostragem e *sample noise*) são bem conhecidas e facilmente obtidas. Conforme é discutido em (Papenfuß et al., 2003), existem diversas justificativas para a utilização do método de amostragem não uniforme *ARS*, principalmente quando se considera os métodos de amostragem que possam ser implementados em aplicações práticas. Mais detalhes sobre o método serão apresentados posteriormente.

2.2.4 Decimação Aleatória

A decimação é um processo em que se reduz a taxa de amostragem de um sinal, descartando-se uma parcela do total de amostras coletadas. Geralmente, tal processo é realizado considerando apenas uma a cada N amostras, mantendo-se uniforme o intervalo entre amostras consideradas, obtendo-se então um fator de decimação igual a N . Se tal processo não for precedido por uma filtragem *anti-aliasing* pode ocorrer o surgimento de frequências falseadas, uma vez que a taxa de amostragem diminui com o descarte de amostras.

Uma maneira de se obter uma amostragem não uniforme é realizando um sorteio aleatório que define quais dados serão descartados e quais serão mantidos. Dessa forma, a taxa de amostragem média diminui, porém, se a probabilidade de descarte das amostras é uniforme, como foi observado empiricamente através de simulações, o limite de *Nyquist* da amostra original se mantém. Tal efeito é discutido na Seção 3.3.4. A esse método, também é dado o nome de *Missing Data* (MD) e esse processo pode ser considerado um caso particular do método *ARS*, em que τ_k possui uma distribuição de probabilidade discreta.

Análise da Amostragem Não Uniforme no Domínio da Frequência

As transformadas tradicionais não são capazes de avaliar o conteúdo espectral de sinais que tenham sido amostrados de maneira não uniforme no tempo. Para isso, é necessário a utilização de uma transformada que considere o intervalo não uniforme entre as amostras. Neste capítulo é descrita a transformada adequada para esse tipo de sinal e algumas de suas propriedades.

3.1 A Transformada Discreta Não Uniforme de Fourier (NDFT)

A análise de sinais é uma arte que evoluiu muito ao longo do último século, com a criação de diferentes ferramentas e tipos de abordagem. A análise de sinais no domínio da frequência é uma das ferramentas de maior potencial, permitindo o entendimento e solução de diversos tipos de desafios. Dentre as diversas transformadas integrais, uma das mais utilizadas é conhecida como a transformada de Fourier:

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt, \quad (3.1)$$

em que ω representa a frequência angular avaliada e $j = \sqrt{-1}$. Essa é uma transformada que opera funções analíticas e contínuas no tempo. Porém, no mundo digital, os sinais são transformados para o domínio discreto via amostragem e quantização e estes são operados pela transformada discreta de Fourier (DFT):

$$X(jk\Omega_0) = \sum_{n=0}^{N_0-1} x(nT_0)e^{-jk\Omega_0nT_0}, \quad k = 0, 1, \dots, N_0 - 1. \quad (3.2)$$

A Expressão 3.2 descreve a transformação de um sinal periódico e discreto $x(nT_0)$ que possui N_0 pontos, uniformemente distribuídos. O intervalo entre amostras é dado por T_0 e a transformada é avaliada em múltiplos inteiros da frequência fundamental Ω_0 . Com essa operação obtém-se, a partir de um sinal periódico e discreto, um resultado também periódico e discreto e com N_0 pontos.

O intuito deste trabalho é o estudo de sinais em que as amostras não estão espaçadas uniformemente. Para isso, é necessário utilizar uma ferramenta adequada, conhecida como transformada discreta não uniforme de Fourier (NDFT ou NUT-DFT)(Bland et al., 1996)

$$X(\omega_m) = \sum_{n=0}^{N-1} x(t_n)e^{-j\omega_m t_n}, \quad t_n \in \mathfrak{R}, \quad (3.3)$$

em que N indica o número de amostras avaliadas, ω_m a frequência angular avaliada, e t_n o instante de amostragem. Sua aparência lembra muito a DFT descrita anteriormente. Sua diferença reside nos valores de tempo em que são localizadas as amostras e nos valores de frequência nos quais é calculada a transformada, que podem ser definidos de forma arbitrária. Tanto os valores de ω_m quanto t_n podem assumir qualquer valor, sendo a definição destes um ponto crucial de projeto. Sendo assim, a DFT pode ser considerada um caso particular da NDFT, em que os pontos estão distribuídos sempre à mesma distância no eixo temporal. Um ponto importante a ser observado é que, para um sinal com N amostras, avaliado em M frequências, a ordem de complexidade desta transformada é $O(MN)$ (Kunis e Potts, 2008). Existem formas de se obter transformadas rápidas para dados distribuídos de forma não uniforme no tempo, conforme é discutido em (Potts et al., 2001), porém, com o ganho de velocidade, há um custo de precisão. Neste trabalho, a expressão (3.3) será utilizada. Ainda assim, o uso de tal ferramenta gera diversos efeitos que serão discutidos ao longo deste trabalho.

3.2 Ruído da Linha de Base

Como foi apresentado na Seção 2.2.1, ao se calcular a transformada de Fourier de um sinal amostrado de forma não uniforme, surge um ruído da linha de base que polui todo o espectro. Tal ruído surge, não só no método apresentado (SJS), como em qualquer sinal amostrado de forma não uniforme. Será apresentado nas seções seguintes que tal ruído é função, não somente do sinal amostrado, como também dos parâmetros de amostragem. Esta seção tem como objetivo apresentar a origem de tal ruído e quais parâmetros tem impacto direto em sua instensidade.

Para melhor apresentar os argumentos que explicam o surgimento do ruído da linha de base é preciso voltar às definições da transformada discreta de Fourier e da transformada discreta não uniforme de Fourier.

3.2.1 Transformada Discreta de Fourier Revisitada

Como apresentado na Seção 3.1, a DFT é definida como:

$$X(k\Omega_0) = \sum_{n=0}^{N_0-1} x(nT_0)e^{-jk\Omega_0nT_0}, \quad k = 0, 1, \dots, N-1. \quad (3.4)$$

Tal expressão, a partir da identidade de Euler, pode ser expandida para:

$$X(k\Omega_0) = \sum_{n=0}^{N_0-1} x(nT_0)(\cos(k\Omega_0nT_0) + j \operatorname{sen}(k\Omega_0nT_0)). \quad (3.5)$$

Para facilitar a explicação, considere que o sinal amostrado seja:

$$x(nT_0) = \cos(\omega_p nT_0 + \phi_p). \quad (3.6)$$

Assim, temos:

$$X(k\Omega_0) = \sum_{n=0}^{N_0-1} [\cos(\omega_p nT_0 + \phi_p) \cos(k\Omega_0nT_0) + j \cos(\omega_p nT_0 + \phi_p) \operatorname{sen}(k\Omega_0nT_0)]. \quad (3.7)$$

Tal equação pode ser analisada em duas situações distintas: Quando $\omega_p = k\Omega_0$ e quando $\omega_p \neq k\Omega_0$. Quando a frequência do sinal (ω_p) é diferente da frequência avaliada ($k\Omega_0$), ambos os termos do somatório são equivalentes a ondas

moduladas e, desde que a janela temporal avaliada seja equivalente a uma quantidade inteira de períodos do sinal, seu somatório será nulo. Quando $\omega_p = k\Omega_0$ os produtos dentro do somatório somente serão nulos em condições específicas de fase distintas. O somatório do produto de cossenos é nulo somente quando a fase do sinal for $\phi_p = N\pi + \pi/2$, e o somatório do produto de cosseno por seno somente é nulo quando $\phi_p = N\pi$. Esse é o motivo pelo qual só aparecem raias no espectro nas frequências em que existem componentes do sinal. Para ilustrar os conceitos apresentados, as Figuras 3.1 e 3.2 são apresentadas.

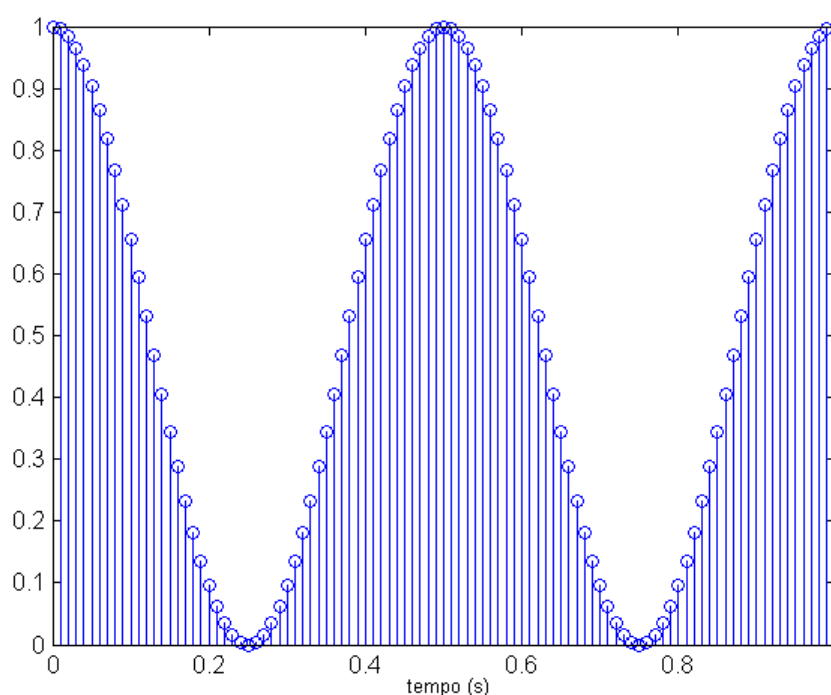


Figura 3.1: Sinal $X(nT_0) = \cos(2\pi nT_0) \cos(2\pi nT_0)$ amostrado uniformemente.

A Figura 3.1 representa a amostragem do produto de dois sinais cossenos de mesma frequência. Ao se reescrever a expressão (3.7) como

$$X(k\Omega_0) = \sum_{n=0}^{N_0-1} \cos(\omega_p nT_0 + \phi_p) \cos(k\Omega_0 nT_0) + \sum_{n=0}^{N_0-1} j \cos(\omega_p nT_0 + \phi_p) \sin(k\Omega_0 nT_0), \quad (3.8)$$

tal operação é avaliada dentro de seu primeiro somatório, quando $\omega_p = \Omega_0$ e $\phi_p = n2\pi$. Pela identidade

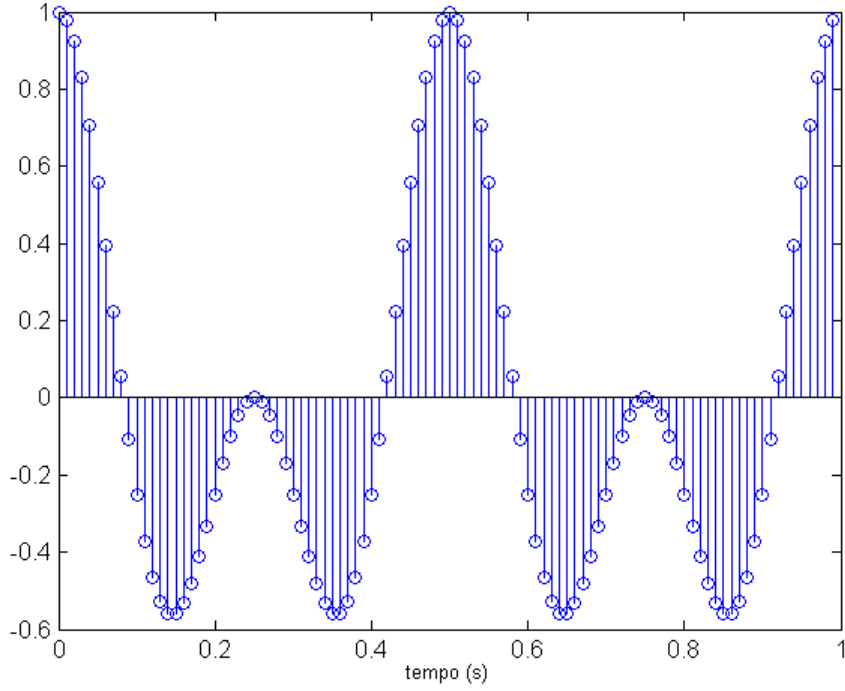


Figura 3.2: Sinal $X(nT_0) = \cos(2\pi nT_0) \cos(6\pi nT_0)$ amostrado uniformemente.

$$\cos(2\pi nT_0) \cos(2\pi nT_0) = \cos^2(2\pi nT_0) = \frac{\cos(4\pi nT_0) + 1}{2}, \quad (3.9)$$

conclui-se que o somatório de tal função sempre será maior que zero. Para a segunda metade da Expressão 3.8, temos:

$$\sum_{n=0}^{N_0-1} j \cos(\omega_p nT_0 + \phi_p) \text{sen}(k\Omega_0 nT_0) \quad (3.10)$$

que é o somatório de um sinal que é periódico de média nula e que, desde que $\phi_p = n\pi$ e tal somatório seja avaliado em um número inteiro de períodos, seu resultado será nulo.

No caso em que $\phi_p \neq n\pi$, o valor do primeiro somatório da Expressão 3.8 será inferior à situação onde $\phi_p = n\pi$, porém o valor do segundo somatório

será maior, fazendo com que a relação

$$\begin{aligned} & \left| \sum_{n=0}^{N_0-1} \cos(\omega_p n T_0) \cos(k \Omega_0 n T_0) \right. \\ & \quad \left. + \sum_{n=0}^{N_0-1} j \cos(\omega_p n T_0) \text{sen}(k \Omega_0 n T_0) \right| \\ = & \left| \sum_{n=0}^{N_0-1} \cos(\omega_p n T_0 + \phi_p) \cos(k \Omega_0 n T_0) \right. \\ & \quad \left. + \sum_{n=0}^{N_0-1} j \cos(\omega_p n T_0 + \phi_p) \text{sen}(k \Omega_0 n T_0) \right| \end{aligned} \quad (3.11)$$

seja válida para qualquer valor de ϕ_p . Assim conclui-se que, desde que avaliado em um número inteiro de períodos de um sinal,

$$\begin{aligned} X(k \Omega_0) = \sum_{n=0}^{N_0-1} [\cos(\omega_p n T_0 + \phi_p) \cos(k \Omega_0 n T_0) \\ + j \cos(\omega_p n T_0 + \phi_p) \text{sen}(k \Omega_0 n T_0)] > 0, \quad \omega_p = k \Omega_0. \end{aligned} \quad (3.12)$$

A Figura 3.2 representa a amostragem do produto de senoides de diferentes frequências. Como resultado, obtém-se um sinal periódico de média nula. Tal operação ocorre, em ambos os termos da Expressão 3.8, quando $\omega_p \neq k \Omega_0$. Então, desde que o sinal seja amostrado em uma janela temporal equivalente a um número inteiro de períodos deste, obtém-se:

$$\begin{aligned} X(k \Omega_0) = \sum_{n=0}^{N_0-1} [\cos(\omega_p n T_0 + \phi_p) \cos(k \Omega_0 n T_0) \\ + j \cos(\omega_p n T_0 + \phi_p) \text{sen}(k \Omega_0 n T_0)] = 0, \quad \omega_p \neq k \Omega_0. \end{aligned} \quad (3.13)$$

3.2.2 Transformada Discreta Não Uniforme de Fourier

Assim como a DFT, a NDFT pode ser expandida utilizando-se a identidade de Euler. Seguindo passos equivalentes aos da seção anterior, chega-se à seguinte equação para a transformada não uniforme de um sinal senoidal puro não uniformemente amostrado:

$$X(\omega_m) = \sum_{n=0}^{N_0-1} [\cos(\omega_p t_n + \phi_p) \cos(\omega_m t_n) + j \cos(\omega_p t_n + \phi_p) \text{sen}(\omega_m t_n)]. \quad (3.14)$$

Como foi citado anteriormente, a DFT é um caso particular desta transformada. Então, desde que o sinal temporal seja amostrado uniformemente, as conclusões obtidas na seção anterior são válidas para esta transformada. Porém, se o sinal tiver sido amostrado de forma não uniforme, a afirmação

sobre o valor nulo da transformada, quando a frequência avaliada ω_m é diferente da frequência do sinal ω_p , não é mais válida. Isso pois, uma vez que os instantes de avaliação da função de interesse são aleatórios, não é possível garantir que o somatório do produto da função amostrada pela senoide de interesse seja nulo. De forma a exemplificar o conceito exposto, é apresentado, na Figura 3.3, um sinal, que é o produto de duas senoídes de frequências diferentes, amostrado não uniformemente.

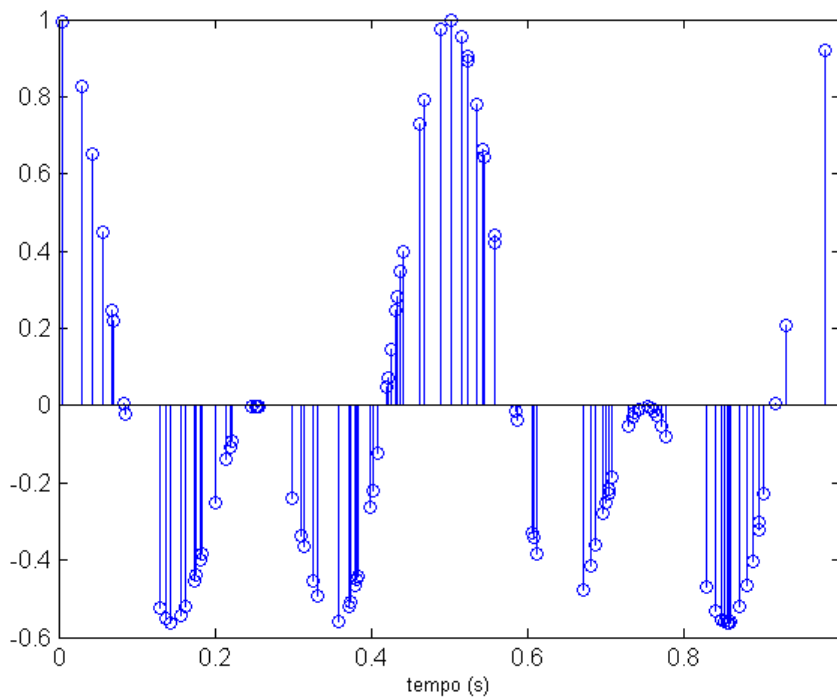


Figura 3.3: Sinal $X(nT_0) = \cos(2\pi nT_0) \cos(6\pi nT_0)$, amostrado não uniformemente.

O sinal aqui avaliado é o mesmo da Figura 3.2, porém, como o intervalo entre amostras é não uniforme, ao se somar todas as amostras, somente em casos extremamente específicos e muito raros, o somatório destes valores será nulo, mesmo que se avalie um número inteiro de períodos do sinal de interesse.

3.3 Influência das Variáveis de Amostragem no Ruído da Linha de Base

Como foi já observado, ao se tratar dados amostrados de forma não uniforme para o cálculo da transformada de um sinal, surge um ruído da linha de base que polui todo o espectro. A existência desse ruído é um forte limitador da capacidade de análise desta ferramenta, uma vez que sinais de interesse, por não terem amplitude suficiente para se destacarem, podem ser confundidos com ele. A variância desse ruído é influenciada por diversos parâmetros que devem ser definidos na coleta de dados. Esta seção tem por objetivo citar quais características, tanto do sinal quanto do tipo de amostragem, influenciam esse ruído. Aqui, as conclusões obtidas de forma empírica serão citadas de forma mais superficial, sendo que o desenvolvimento matemático que as explica será apresentado no Capítulo 4.

3.3.1 Influência da Taxa de amostragem Média

O primeiro parâmetro a ser discutido é a taxa média de amostragem de um sinal, que tem influência direta no ruído da linha de base. Para ilustrar o impacto dessa variável, testes foram executados nos quais foram calculadas transformadas com diferentes valores de taxa de amostragem. O sinal avaliado foi:

$$x(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 16 \text{ Hz}, f_2 = 210 \text{ Hz}, f_3 = 876 \text{ Hz}; \quad (3.15)$$

de forma que foi utilizada janela de amostragem de 1 segundo de duração. Os resultados são apresentados na Figura 3.4.

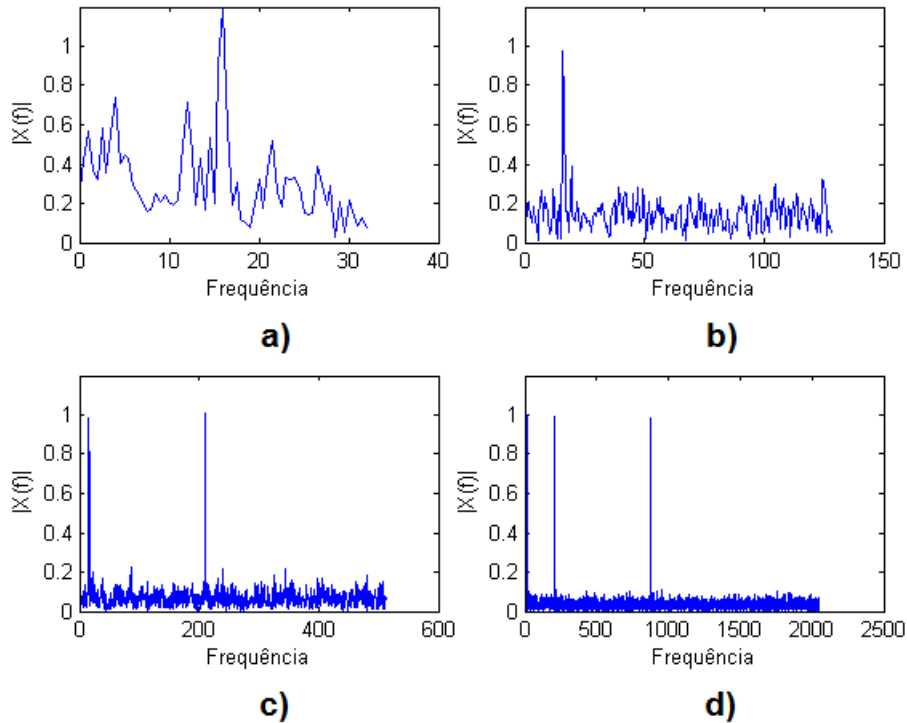


Figura 3.4: $|X(\omega_m)|$ para diferentes valores de \bar{f}_s .
 a) 64 amostras/segundo. b) 256 amostras/segundo.
 c) 1024 amostras/segundo. d) 4096 amostras/segundo.
 Sinal avaliado: $x(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$,
 $f_1 = 16$ Hz, $f_2 = 210$ Hz, $f_3 = 876$ Hz;
 $T_J = 1$ s. Método de amostragem: Sorteio Aleatório.

À medida que a taxa de amostragem média se eleva, a variância do ruído cai de forma assintótica. Nas quatro execuções, o sinal avaliado foi o mesmo (somatório de senoides de 16 Hz, 210 Hz e 876 Hz) e é interessante reafirmar que, mesmo nas situações em que a taxa média de amostragem é baixa, as componentes de alta frequência não geram falseamento, inequivocamente identificado a partir de um pico (um máximo local) claramente presente. Como padrão, os limites da transformada se estenderam até metade da frequência média de amostragem, porém, como o falseamento não ocorre nestes testes, o limite é uma variável de projeto e pode ser estendido a valores muito superiores aos utilizados. Para reforçar esse ponto, o limite do gráfico da Figura 3.4b foi estendido até 2048 Hz e é apresentado na Figura 3.5.

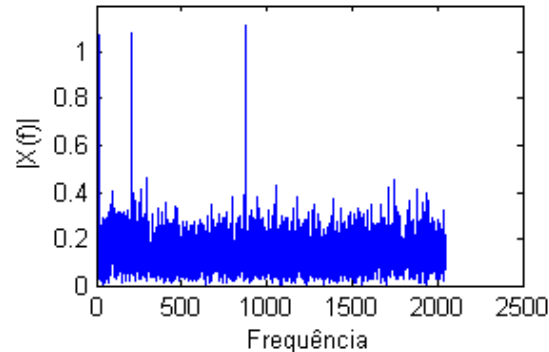


Figura 3.5: $|X(\omega_m)|$ de sinal amostrado com $\bar{f}_s = 256$ amostras/segundo.

$$\text{Sinal avaliado: } x(t) = \sum_{i=1}^3 \cos(2\pi f_i t),$$

$$f_1 = 16 \text{ Hz}, f_2 = 210 \text{ Hz}, f_3 = 876 \text{ Hz};$$

$$T_J = 1 \text{ s. Método de amostragem: Sorteio Aleatório.}$$

Observa-se que, mesmo a taxa média de amostragem sendo apenas 256 amostras por segundo, a componente na frequência 876Hz é identificada corretamente.

Finalmente, cabe citar que estes são resultados de uma variável aleatória e que, a cada realização obtém-se valores diferentes para a NDFFT. Entretanto, como o intuito desta seção é ilustrativo, esta subseção e as seguintes apresentam resultados de execuções que poderiam ser distintos, porém são bem representativos em relação ao conceito que se deseja firmar.

3.3.2 Influência do Tamanho da Janela Amostrada

Para o segundo parâmetro a ser discutido, os testes são semelhantes aos citados na Seção 3.3.1. O seguinte sinal é avaliado:

$$x(t) = \sum_{i=1}^5 \cos(2\pi f_i t), \quad f_1 = 5 \text{ Hz}, f_2 = 30 \text{ Hz}, f_3 = 48 \text{ Hz}, \quad (3.16)$$

$$f_4 = 80 \text{ Hz}, f_5 = 250 \text{ Hz};$$

amostrado a uma taxa média de 256 amostras por segundo. A variável desse teste foi a duração temporal da janela avaliada e os resultados são apresentados na Figura 3.6.

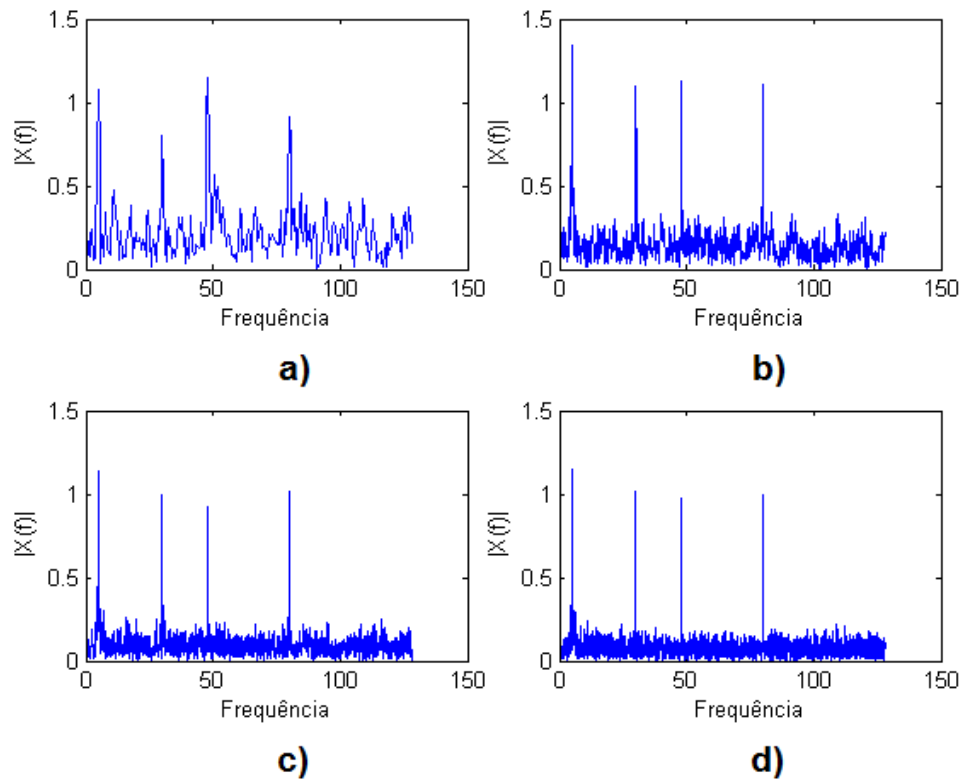


Figura 3.6: $|X(\omega_m)|$ para diferentes tamanhos de janela temporal.

a) $T_J = 1$ segundo. b) $T_J = 3$ segundos.

c) $T_J = 6$ segundos. d) $T_J = 12$ segundos.

Sinal avaliado: $x(t) = \sum_{i=1}^5 \cos(2\pi f_i t)$,

$f_1 = 5$ Hz, $f_2 = 30$ Hz, $f_3 = 48$ Hz, $f_4 = 80$ Hz, $f_5 = 250$ Hz;

$f_s = 256$ Hz. Método de amostragem: Sorteio Aleatório.

O comportamento observado aqui é semelhante ao encontrado na avaliação anterior. À medida que a duração da janela avaliada aumenta, a resolução espectral cresce (seu valor é o inverso da duração da janela) e a variância do ruído da linha de base decresce. Aqui, a componente de 250Hz, como era esperado, não gera componente devido a falseamento e o limite superior do espectro, novamente, foi definido arbitrariamente.

Os resultados aqui apresentados ilustram comportamentos muito semelhantes aos da Seção 3.3.1. A variância de ruído se eleva à medida que, ou a taxa de amostragem média, ou a duração temporal do sinal avaliado decresce. Uma variável que avalia conjuntamente esses dois parâmetros é a quantidade de amostras utilizadas para a avaliação da NDFT. Em ambos os testes ci-

tados, o ruído se eleva, à medida que o número de amostras utilizadas no cálculo da transformada não uniforme decresce.

3.3.3 Influência dos Componentes de Frequência de um Sinal

O terceiro parâmetro a ser discutido, foi avaliado por testes semelhantes aos das Seções 3.3.1 e 3.3.2. Os testes aqui descritos, no entanto, são os únicos que avaliam características, não da amostragem escolhida, mas sim do sinal de interesse. Como os sinais simulados são sempre uma soma de senoides de diferentes frequências, foi percebido que sinais que possuem mais componentes de frequência possuem um ruído da linha de base mais proeminente. A Figura 3.7 apresenta o gráfico das transformadas de sinais que possuem, respectivamente, 1, 3, 5 e 7 componentes em frequência.

As características de amostragem foram iguais em todos os testes: 1 segundo de duração e uma taxa média de 256 amostras por segundo.

O aumento da variância do ruído da linha de base com o aumento do número de componentes de frequência do sinal é comportamento muito interessante e o mais surpreendente dentre os testes apresentados. Aqui, percebe-se uma notável diferença do nível do ruído, sendo que as características de amostragem são as mesmas. No Capítulo 4, quando um desenvolvimento matemático mais detalhado é apresentado, a razão desse comportamento, um tanto quanto surpreendente, ficará claro.

3.3.4 Limite na Amostragem *Missing Data*

Para a obtenção da transformada de Fourier de um sinal, a partir de amostras não uniformes, foi mencionada a importância da hipótese de que amostras podem ser coletadas em qualquer instante de tempo, evitando assim o falseamento. Porém, em aplicações práticas, nem sempre intervalos entre amostras podem assumir qualquer valor. No caso de uma amostragem do tipo *MD*, apesar de as amostras serem selecionadas aleatoriamente, seus intervalos são sempre múltiplos de um intervalo base (que foi definido na coleta em taxa elevada). Nesse caso, o Limite de Nyquist que já existia, baseado nesse intervalo mínimo, se mantém. Para exemplificar o conceito, o sinal

$$x(t) = \cos(2\pi 300t), \quad (3.17)$$

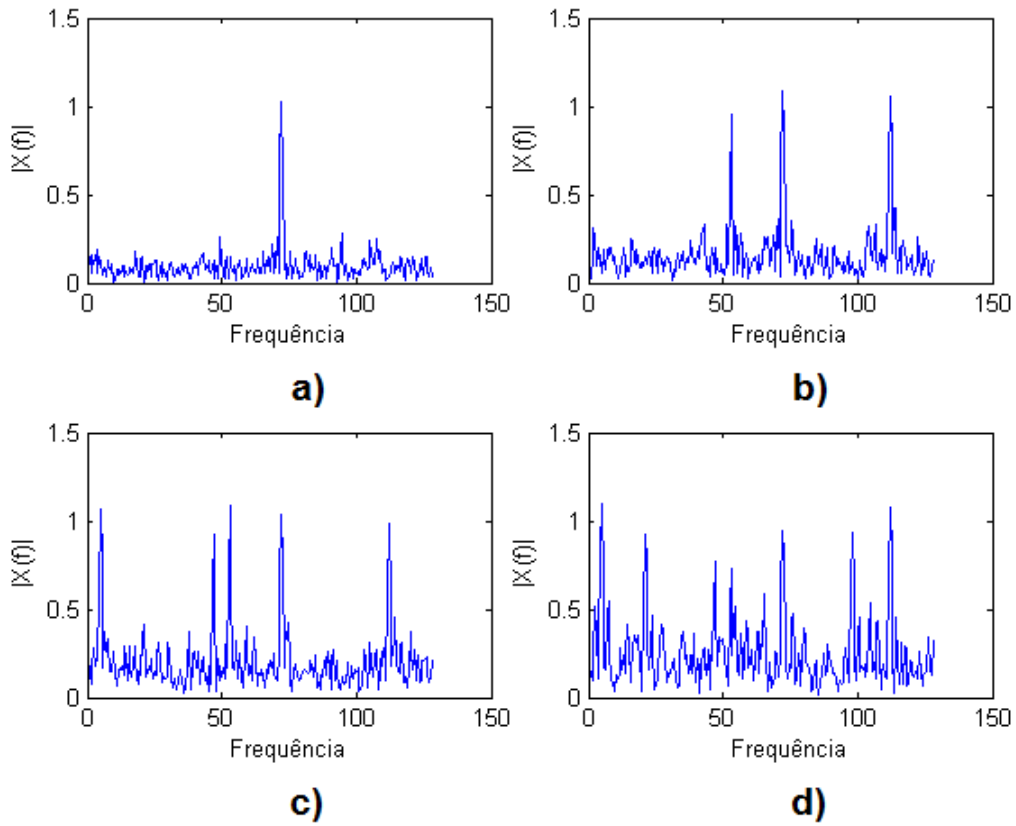


Figura 3.7: $|X(\omega_m)|$ para diferentes sinais.

a) $x(t) = \cos(2\pi f_1 t)$,

b) $x(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$,

c) $x(t) = \sum_{i=1}^5 \cos(2\pi f_i t)$,

d) $x(t) = \sum_{i=1}^7 \cos(2\pi f_i t)$,

$f_1 = 72 \text{ Hz}, f_2 = 53 \text{ Hz}, f_3 = 112 \text{ Hz}, f_4 = 5 \text{ Hz},$

$f_5 = 47 \text{ Hz}, f_6 = 21 \text{ Hz}, f_7 = 98 \text{ Hz}.$

$T_J = 1 \text{ s} \quad f_s = 256 \text{ Hz}.$

Método de amostragem: Sorteio Aleatório.

foi amostrado uniformemente a uma taxa de 1000 amostras por segundo. Em seguida, foi aplicada uma decimação aleatória com uma taxa de decimação de 10, ou seja, foi considerado apenas 1/10 das amostras, selecionadas aleatoriamente, gerando então um novo sinal com taxa de amostragem média de 100 amostras por segundo. A NDFT calculada do sinal é apresentada na Figura 3.8

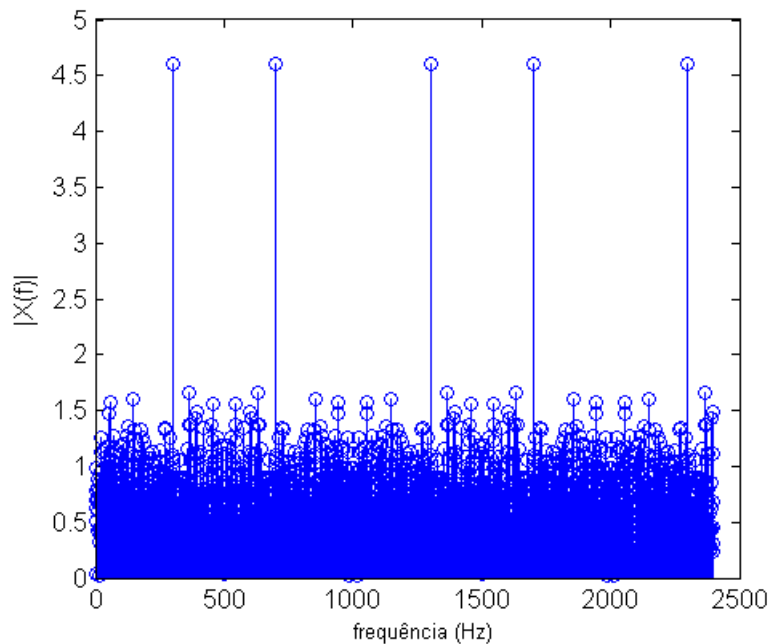


Figura 3.8: NDFT do sinal $x(t) = \cos(2\pi 300t)$ amostrado não uniformemente via M.D., a partir de um sinal com $f_s = 1000$ amostras/segundo e taxa de decimação = 10.

Como pode ser observado, ocorre um espelhamento em múltiplos da metade da frequência de amostragem máxima original (500 Hz), assim como antes da decimação. Isso ocorre devido à impossibilidade de existência de amostras em valores de tempo que sejam frações de 1ms.

3.3.5 Jitter e Limitações da NDFT na Amostragem Não Uniforme

Considerando os conceitos apresentados até então, a análise em frequência de sinais amostrados de forma não uniforme parece apresentar uma capacidade de se estender a frequências arbitrariamente altas, entretanto, na

prática, limites reais existem e estão relacionados com a precisão de medida dos instantes de amostragem. Medidas de tempo estão sempre sujeitas a erro e, à discrepância entre os valores reais e os valores considerados nos instantes de amostragem, é dado o nome de *Jitter*. Quanto menor for a incerteza associada às medidas de tempo consideradas para cada amostra, maior será a banda válida de uma NDFt calculada. Tal assunto é foco de estudo de várias pesquisas e mais informações sobre o assunto podem ser lidas em (Tarczynski e Allay, 2004).

Detecção Automática de Falseamento

Ao se analisar o conteúdo espectral de um sinal, obtém-se uma nova visão de da informação contida naqueles dados. A transformada de Fourier, há décadas, é utilizada para a análise do espectro de sinais e foi visto que a transformada não uniforme de Fourier consegue ser, até elevados valores de frequência, imune ao falseamento, porém ao custo do surgimento de um ruído que é consequência de vários fatores.

O ruído da linha de base é função de variáveis aleatórias, sendo que algumas delas são definidas ao se escolher o método de amostragem e suas características. Sendo assim, é esperado que seja, de alguma forma, possível conhecer características estocásticas desse ruído. Como foi detalhadamente desenvolvido por Eng et al. (2008), é possível, uma vez conhecidas as características de amostragem aleatória, calcular estimadores de valor esperado e variância de um espectro.

A contribuição original deste trabalho está em, uma vez calculadas as transformadas uniforme e não uniforme de Fourier, a realização de testes de inferência estatística que utilizem os estimadores citados para a avaliação e possível rejeição os resultados obtidos, devido à suspeita de falseamento.

Neste capítulo, as etapas mais importantes do procedimento de detecção de falseamento são desenvolvidas. São detalhadas a obtenção dos dois primeiros momentos centrais estocásticos do espectro e a realização do teste de hipótese do espectro avaliado.

4.1 Características Estocásticas dos Tipos de Amostragem Não Uniforme

Foi visto anteriormente que existem diversas formas de se definir a forma como os instantes de amostragem não uniforme são gerados. Um parâmetro

muito importante é o *sampling noise* (τ_k) e, a partir da função densidade de probabilidade dessa variável, é possível desenvolver equações de um estimador do valor esperado do espectro de um sinal amostrado. Por ter uma formulação matemática mais próxima da possibilidade de uma aplicação prática, deste momento em diante, o raciocínio será desenvolvido apenas para o método de amostragem não uniforme *ARS* (Seção 2.2.3).

4.1.1 Função Densidade de Probabilidade de t_k

O conhecimento das características de τ_k é essencial para descrever o ruído da linha de base. Para isso, o primeiro passo é a definição da função de probabilidade (função densidade de probabilidade (*PDF*), no caso contínuo, função massa de probabilidade (*PMF*), no caso discreto) associada ao k -ésimo instante de amostragem (t_k):

$$p_k(t) = \begin{cases} dP(t_k \leq t)/dt, & \text{Variável aleatória contínua.} \\ P(t_k = t), & \text{Variável aleatória discreta.} \end{cases} \quad (4.1)$$

Em que a função P indica probabilidade ou densidade de probabilidade. Para o desenvolvimento a seguir, será considerado somente o caso contínuo. As funções densidade de probabilidade de τ_k e t_k estão diretamente relacionadas. Como uma continuação das definições apresentadas na Seção 2.2, se esse raciocínio for desenvolvido para o método de amostragem não uniforme *ARS* teremos as expressões a seguir:

$$\begin{aligned} t_k &= t_{k-1} + \tau_k, & t_0 &= 0, \\ \tau_k &\in (T_S^{min}, T_S^{max}), & E[\tau_k] &= T, \\ p_k(t) &= p_\tau \star \dots \star p_\tau = p_\tau^{(k)}(t), \end{aligned} \quad (4.2)$$

em que $p_\tau \star p_\tau$ representa a autoconvolução da *PDF* de τ_k .

Tarczynski e Allay (2004) apresentam o conceito de que é possível obter uma estimativa mais precisa do espectro ao se adotar um τ_k que seja obtido de uma distribuição de probabilidade variante no tempo, mas tal ideia não será adotada pois, como será discutido na Seção 4.2, o procedimento apresentado neste trabalho depende de que a distribuição seja invariante.

4.2 Estimadores do Espectro

Como foi desenvolvido em (Eng et al., 2008), é possível, utilizando-se o Teorema de Bayes, desenvolver a expressão para função densidade de probabilidade do espectro de um sinal. Basta que se conheça as características da amostragem adotada. Entretanto, como a expressão resultante é virtualmente impossível de ser resolvida, tal abordagem se reduz a discussões teóricas.

Ainda no artigo citado, é desenvolvido de forma bem detalhada como, a partir da função característica da expressão de amostragem não uniforme, chega-se ao valor esperado e à matriz de covariância do espectro de um sinal. Aqui será apresentado o desenvolvimento com a sequência de passos que levam aos estimadores desejados.

A transformada não uniforme de Fourier, que é calculada a partir de uma quantidade finita de N amostras, pode ser representada como:

$$Y(f|t_1^N) = \sum_{k=1}^N y_k w_k(f|t_k), \quad (4.3)$$

em que $w_k(f|t_1^N)$ é uma ponderação a ser definida e

$$t_1^N = (t_k)_{k=1}^N \quad (4.4)$$

denota a sequência de valores temporais em que são realizadas as medidas. $Y(f|t_1^N)$ é uma aproximação da transformada determinística

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} dt, \quad (4.5)$$

cuja inversa é dada por:

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft} df. \quad (4.6)$$

Considerando este fato, duas funções de ponderação podem ser usadas em (4.3), a saber:

$$w_k(f|t_k) = (\tau_k)^n e^{-2j\pi ft_k} \Rightarrow \begin{cases} w_k(f|t_k) = e^{-j2\pi ft_k}, & n = 0, \text{ DirT}, \\ w_k(f|t_k) = \tau_k e^{-j2\pi ft_k}, & n = 1, \text{ RieT}, \end{cases} \quad (4.7)$$

em que o intervalo de tempo decorrido entre duas amostras é usado na chamada *transformada de Riemann* (RieT), ou não é usado na chamada *transformada de Dirichlet* (DirT). A variável $n \in \{0, 1\}$ indica a transformada utilizada. No caso de instantes temporais uniformemente espaçados, temos o caso específico da transformada discreta de Fourier:

$$Y_{DTFT}(f) = T \sum_{k=1}^N s(kT) e^{-j2\pi f k T}, \quad (4.8)$$

que corresponde a utilizar como ponderação $w_k(f|kT) = T e^{-j2\pi f k T}$, sendo que fator de escala T nem sempre é utilizado na literatura.

Para o melhor esclarecimento de como τ_k tem importância crucial nas propriedades estocásticas da transformada não uniforme de Fourier, será utilizada a definição de função característica, dada por:

$$\varphi_m(f) \triangleq E_{\tau_k}[(\tau_k)^n e^{-j2\pi f \tau_k}], \quad n = 0, \text{ ou } n = 1, \quad (4.9)$$

em que τ_k é o intervalo entre amostras e n é um parâmetro que assume valores diferentes quando se utiliza transformada RieT ou DirT.

A próxima função a ser definida é o núcleo de *Dirichlet* normalizado, também chamada de *aliased sinc function*.

$$d_N^{ARS}(f) \triangleq \frac{1 - \varphi_0(f)^N}{1 - \varphi_0(f)} = \frac{1 - E_{\tau_k}[e^{-j2\pi f \tau_k}]^N}{1 - E_{\tau_k}[e^{-j2\pi f \tau_k}]}. \quad (4.10)$$

Mesclando as definições apresentadas nas Equações 4.3 e 4.6, obtêm-se:

$$\begin{aligned} Y(f) &= \sum_{k=1}^N y_k w_k(f|t_k) = \sum_{k=1}^N \left(\int_{-\infty}^{\infty} S(\psi) e^{j2\pi \psi t_k} d\psi \right) w_k(f|t_k) \\ &= \int_{-\infty}^{\infty} S(\psi) W(f, \psi) d\psi, \end{aligned} \quad (4.11)$$

em que

$$W(f, \psi) = \sum_{k=1}^N e^{j2\pi \psi t_k} w_k(f|t_k), \quad (4.12)$$

e a definição da ponderação $w_k(f|t_k)$ em (4.7) implica em

$$W(f, \psi) = W(f - \psi) \triangleq \sum_{k=1}^N (\tau_k)^n e^{-j2\pi(f-\psi)t_k}. \quad (4.13)$$

Pode-se concluir, a partir de (4.11) e (4.13), que a transformada de um sinal amostrado não uniformemente pode ser interpretada como uma convolução do valor esperado da janela estocástica de frequência ($W(f)$) com a transformada de Fourier regular de um sinal. E, de forma similar, a matriz de covariância de um espectro é obtida pela convolução dupla do espectro do sinal de interesse ($S(f)$) com a matriz de covariância de $W(f)$.

Chega-se então a duas das equações mais importantes deste capítulo:

$$\mu_Y(f; S) = S(f) \star \mu_W(f), \quad (4.14)$$

$$R_Y(f, \psi; S) = S(f) \star \star R_W(f, \psi), \quad (4.15)$$

a definição dos estimadores de valor esperado e de matriz de covariância de um espectro¹, sendo que

$$\mu_W(f) \triangleq \text{E}[W(f)], \quad (4.16)$$

e

$$R_W(f, \psi) \triangleq \text{Cov}[W(f), W(\psi)]. \quad (4.17)$$

A prova da obtenção do valor esperado é obtida a partir de (4.12):

$$\begin{aligned} \mu_Y(f; S) &= \text{E}[Y(f)] \\ &= \text{E}\left[\int_{-\infty}^{\infty} S(\phi)W(f - \phi)d\phi\right] \\ &= \int_{-\infty}^{\infty} S(\phi)\text{E}[W(f - \phi)]d\phi \\ &= S \star \mu_W(f) \end{aligned}$$

e de forma similar, a matriz de covariância é obtida da seguinte maneira:

¹ Foram utilizadas as definições de convolução e convolução dupla dadas por $f \star g(x) = \int f(\xi)g(x - \xi)d\xi$ e $f \star \star h(x, y) = \int \int f(\xi)h(x - \xi, y - \eta)f^*(\eta)d\xi d\eta$.

$$\begin{aligned}
R_Y(f, \psi; S) &= \text{Cov}[Y(f), Y(\psi)] \\
&= \text{E}[(Y(f) - \mu_Y(f))(Y(\psi) - \mu_Y(\psi))^*] \\
&= \text{E}\left[\left(\int_{-\infty}^{\infty} S(\phi)(W(f - \phi) - \mu_W(f - \phi))d\phi\right) \right. \\
&\quad \left. \times \left(\int_{-\infty}^{\infty} S(\eta)(W(\psi - \eta) - \mu_W(\psi - \eta))d\eta\right)^* \right] \\
&= \int \int_{-\infty}^{\infty} S(\phi)\text{E}[(W(f - \phi) - \mu_W(f - \phi)) \\
&\quad \times (W(\psi - \eta) - \mu_W(\psi - \eta))^*]S(\eta)^*d\phi d\eta \\
&= S \star \star R_W(f, \psi).
\end{aligned}$$

A matriz de covariância obtida acima é instrumental para se determinar a variabilidade esperada para os valores obtidos, no domínio da frequência, ao se computar o espectro de um mesmo sinal, a partir de diferentes realizações de amostragem não uniforme.

É importante notar que as expressões mostradas acima, originalmente deduzidas e publicadas em (Eng et al., 2008), apontam para uma interpretação do espectro esperado, que seria obtido a partir de um número infinito de realizações de amostragem não uniforme seguida do cômputo da NDFt para um mesmo sinal $s(t)$, como o resultado de uma convolução entre o espectro $S(f)$ desse sinal original e um sinal de ponderação $W(f)$. O sinal de ponderação, por sua vez, depende das características estocásticas do processo de amostragem não uniforme. Portanto, conclui-se que o espectro esperado é função não apenas das características estocásticas do processo de amostragem, *mas também da distribuição espectral de potência do sinal que se está sendo amostrado.*

De forma a facilitar a exibição das equações para obtenção dos estimadores de valor esperado e matriz de covariância de $W(f)$, a seguir, se encontram as expressões resumidas

$$\mu_W(f) = \varphi_n(f)d_N(f) \quad (4.18)$$

$$R_W(f, \psi) = \Phi_n(f, \psi)^T D_N(f, \psi) - \mu_W(f) - \mu_W(\psi)^*, \quad (4.19)$$

sendo que as expressões explícitas estão apresentadas no quadro a seguir.

Expressões para o Cálculo de Média e Covariância de $W(f)$

$$\begin{aligned}
\varphi(f) &= \text{E} [e^{-j2\pi t\tau_k}]. \\
d_N^{ARS}(f) &= \begin{cases} \frac{1-\varphi(f)^N}{1-\varphi(f)}, & \varphi(f) \neq 1, \\ N, & \varphi(f) = 1. \end{cases} \\
\Phi^{ARS}(f, \psi) &= \begin{pmatrix} \varphi(f)\varphi(f-\psi) \\ \varphi(f-\psi) \\ \varphi(-\psi)\varphi(\psi-f) \end{pmatrix} \\
D_N^{ARS}(f, \psi) &= \begin{pmatrix} \sum_{k=1}^N \sum_{l=1}^{k-1} \varphi(f)^{k-l-1} \varphi(f-\psi)^{l-1} \\ d_N^{ARS}(f-\psi) \\ \sum_{k=1}^N \sum_{l=1}^{k-1} \varphi(-\psi)^{k-l-1} \varphi(\psi-f)^{l-1} \end{pmatrix}
\end{aligned}$$

Uma forma interessante de se interpretar as expressões é que a convolução com a janela estocástica de frequência introduz o ruído da linha de base que polui o espectro de um sinal amostrado não uniformemente. No caso ideal, $W(f)$ seria um impulso unitário que, ao convoluir com qualquer sinal, o retornaria perfeitamente, sem alterações. A Figura 4.1 apresenta um exemplo que, para $W(f)$ obtido por amostragem não uniforme, é apresentado o valor esperado e a margem de confiança referente a um desvio padrão. Nesse exemplo, o sinal é amostrado utilizando o método ARS e possui 16 amostras. Fica claro que são os lóbulos laterais de $W(f)$ que provocam o ruído da linha de base que surge no espectro estimado. Para maior clareza do argumento, foi escolhido uma situação na qual a incerteza associada ao sinal é elevada, consequência de um número de amostras consideradas muito baixo. À medida que o número de amostras se eleva, tal gráfico se aproxima da forma de um impulso unitário e sua incerteza cai assintoticamente. A Figura 4.2 ilustra $W(f)$ nas situações onde se tem 128 amostras (Figura 4.2-A) e 512 amostras (Figura 4.2-B).

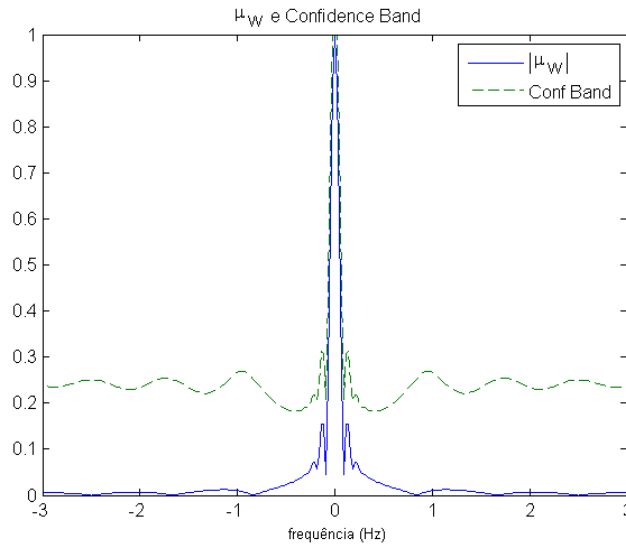


Figura 4.1: Valor Esperado $|\mu_W(f)|$ (linha cheia) e a margem de confiança $|\mu_W \pm \sqrt{R_W}|$ (linha tracejada) de $W(f)$. $N=16$

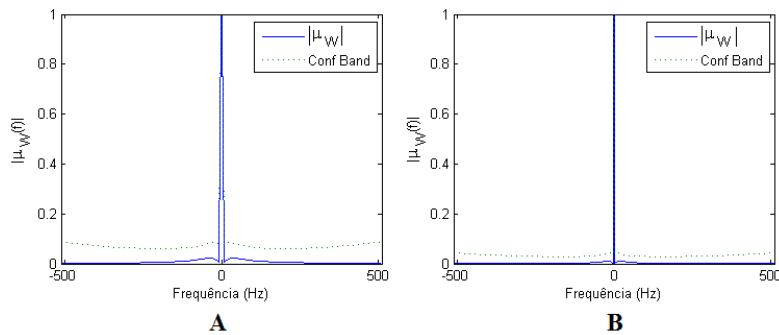


Figura 4.2: Valor Esperado $|\mu_W(f)|$ (linha cheia) e a margem de confiança $|\mu_W \pm \sqrt{R_W}|$ (linha pontilhada) de $W(f)$.
A) $N = 128$ amostras. B) $N = 512$ amostras

Um ponto muito importante a ser ressaltado é a natureza limitada dos estimadores apresentados. Se utilizado um número muito reduzido de amostras, os erros de estimação podem inviabilizar a obtenção de resultados que justifiquem o procedimento. Tal análise é detalhadamente discutida no Capítulo 5.

4.3 O Teste de Hipótese

Como foi concluído na Seção 4.2, um ponto de grande dificuldade foi atingido. Para a obtenção dos estimadores de espectro gerado por amostragem não uniforme, é preciso o conhecimento prévio do espectro verdadeiro do sinal de interesse (vide expressões (4.14) e (4.15)). Uma forma de abordar esse problema, se o objetivo da utilização da transformada não uniforme de Fourier (NDFT) for a obtenção de um espectro livre de falseamento, é usar como estimativa do espectro verdadeiro aquele obtido a partir da transformada rápida de Fourier (FFT) calculada para uma realização de amostragem *uniforme* do sinal.

Inicialmente, como hipótese nula, considera-se que os valores obtidos na FFT calculada a partir de uma amostragem uniforme são fidedignos. Se tal premissa estiver correta, a estimativa $\hat{\mu}_Y$, obtida a partir da expressão (4.14), será, também, fidedigna. Avalia-se, então, a NDFT de uma amostragem não uniforme e confronta-se o resultado com a estimativa $\hat{\mu}_Y$ obtida anteriormente, através de um teste de hipótese em que se considera margens de confiança associadas à estimativa da matriz de covariância calculada como mostrado na expressão (4.15).

4.3.1 Função Densidade de Probabilidade das Grandezas Estimadas

Como foi visto na Seção 4.2, é possível estimar o espectro de um sinal amostrado de forma não uniforme. A partir disso, parece ser possível realizar um teste de hipótese envolvendo o espectro obtido e sua matriz de covariância.

Entretanto, existe um ponto que exige cuidado: ao se calcular a transformada não uniforme de Fourier de um sinal, é obtido, para cada frequência, um valor complexo. Tanto a parte real quanto a parte imaginária possuem distribuições independentes e normais (Eng et al., 2008). Ao se avaliar o módulo destes valores complexos, os valores obtidos passam a se comportar como variáveis aleatórias com função densidade de probabilidade do tipo *Rice* (Eng et al., 2008). Realizar testes de hipótese em distribuições assimétricas não é uma atividade simples, uma vez que, para cada valor de frequência, haverá uma distribuição com parâmetros diferentes que torna os cálculos muito particulares. No entanto, existe uma saída mais simples. Como o espectro complexo possui duas componentes com distribuições Gaussianas, é possível

realizar o teste de hipótese não em relação ao módulo de cada valor obtido a partir da *NDFT* calculada, mas sim no plano complexo. Nesse caso, ao invés da obtenção de um intervalo unidimensional que garanta, por exemplo, 95% de significância, será obtida uma elipse no plano complexo.

4.3.2 Teste de Hipóteses Bidimensional

Como foi explicado na Seção 4.3.1, de forma viabilizar o teste de hipótese, ao invés da realização de um teste em que o limite que define a significância possui apenas uma dimensão, será realizado um teste que considera uma distribuição Gaussiana bivariada.

Variáveis aleatórias de duas dimensões podem possuir distribuições com características diferentes, dependendo do eixo avaliado. Para distribuições Gaussianas bivariadas, é possível definir elipses que envolvam, em média, determinadas porcentagens de valores, em torno do valor mais provável, obtidos a partir da distribuição. Como exemplo, a Figura 4.3 apresenta uma distribuição Gaussiana bidimensional, obtida de uma variável complexa com partes real (X_R) e parte imaginária (X_{Im}). É apresentada também, em vermelho, uma elipse que indica uma região em que se tem 95% de chance de se encontrar um valor obtido a partir da distribuição Gaussiana.

Na Figura 4.3, é apresentado um exemplo de distribuição que trata um caso mais abrangente, no qual existe correlação entre as partes real e imaginária da variável. Nesse caso, os elementos da matriz de covariância não pertencentes à diagonal principal possuem valores não nulos e a elipse está inclinada em relação ao eixo horizontal.

A transformada não uniforme de Fourier fornece, para cada frequência, uma variável aleatória complexa. Como são geradas por funções ortogonais, as partes real e imaginária da variável são não correlacionadas. Dessa forma, em uma situação ideal, os elementos fora da diagonal principal da matriz de covariância dos dados possuem valores nulos e a elipse correspondente estaria alinhada com a direção horizontal. Considerando-se uma situação prática, como a *NDFT* é avaliada a partir de uma quantidade finita de valores, nem sempre esta situação ideal é obtida e os elementos não pertencentes à diagonal principal da matriz de covariância podem apresentar pequenos valores numéricos. Entretanto, à medida que o número amostras avaliadas cresce, os valores destes elementos decrescem tornando-se algumas ordens de grandeza inferiores aos elementos pertencentes à diagonal principal da matriz de co-

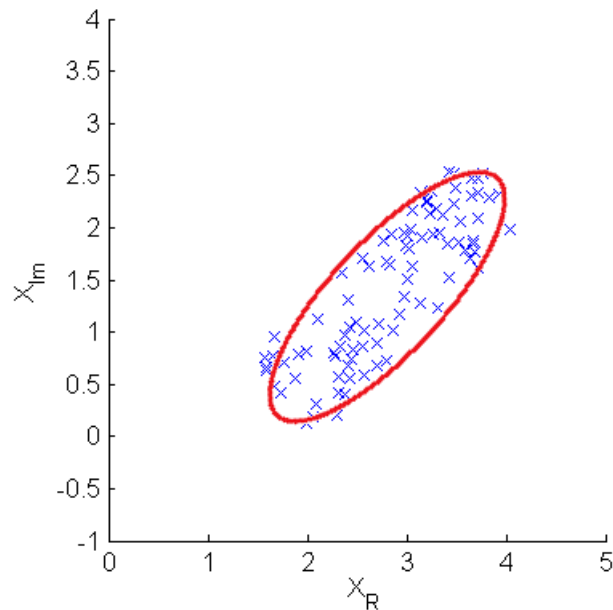


Figura 4.3: *Distribuição Gaussiana bivariada e, em vermelho, elipse que corresponde ao nível de confiança de 95%.*

variância. Um exemplo de distribuição Gaussiana bivariada, obtida a partir de uma variável complexa com partes real (X_R) e parte imaginária (X_{Im}) não correlacionadas e elipse que determina uma região de 95% de probabilidade de se encontrar valores oriundos dessa distribuição é apresentada na Figura 4.4.

Para uma variável complexa com parte real (X_R) e parte imaginária (X_{Im}) independentes e, considerando que ambas as partes possuem distribuições Gaussianas, suas médias são dadas por $\overline{X_R}$ e $\overline{X_{Im}}$ e suas variâncias dadas por σ_R^2 e σ_{Im}^2 . A equação da elipse que representa uma região, em torno da média, com determinada probabilidade de se encontrar valores oriundos de uma distribuição é dada por:

$$\left(\frac{X_R - \overline{X_R}}{\sigma_R}\right)^2 + \left(\frac{X_{Im} - \overline{X_{Im}}}{\sigma_{Im}}\right)^2 = s, \quad (4.20)$$

em que s tem seu valor definido a partir da probabilidade desejada. Por exemplo, para um nível de significância de 95%, $s = 5,991$ e para 99%, $s = 9,210$. Para demais valores, tabelas de probabilidade podem ser encontradas (Tong, 1990).

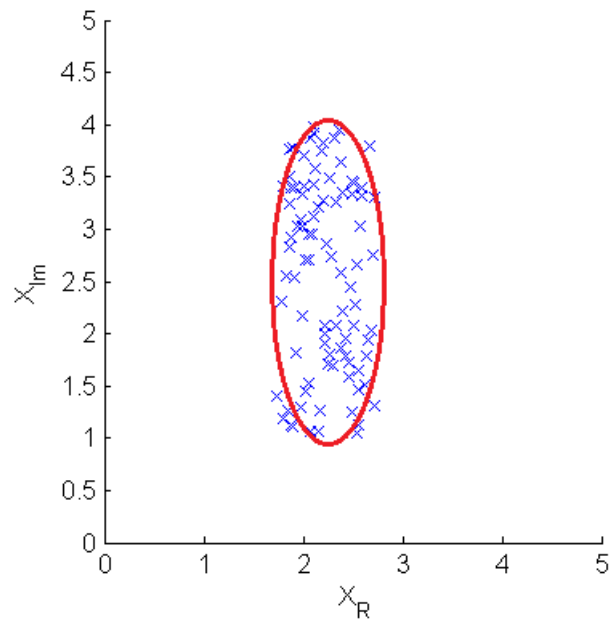


Figura 4.4: *Distribuição Gaussiana bivariada, em que componentes real e imaginária não estão correlacionadas e, em vermelho, elipse que corresponde ao nível de confiança de 95%.*

Como foi definido nas Equações (4.14) e (4.15), é possível se ter, para cada frequência da *NDFT* de um sinal, o valor esperado do espectro e a variância correspondente. Dessa forma, utilizando-se essa informação, juntamente com a Equação (4.20), é possível realizar testes de hipótese para cada uma das frequências desejadas. Sendo assim, a avaliação do espectro é realizada por meio de testes de hipótese para cada uma de suas raias, em que a hipótese nula é que os valores obtidos a partir dos estimadores são fidedignos e, conseqüentemente, os valores obtidos a partir da transformada de Fourier do sinal uniformemente amostrado também são.

4.3.3 Algoritmo para Detecção de Falseamento

Para a avaliação do algoritmo proposto, serão feitos testes que realizam as operações descritas neste capítulo. De forma a ter uma visão mais clara do procedimento, esta seção tem como intuito descrever seus passos.

Inicialmente é preciso definir o sinal (ou o conjunto de sinais) avaliado. Nos testes descritos no Capítulo 5, os sinais avaliados são obtidos através da

seguinte expressão:

$$y(t) = \sum_{i=1}^n \cos(2\pi f_i t), \quad (4.21)$$

de forma que número n de componentes de frequência do sinal é definido de forma diferente, dependendo do teste avaliado.

Em sequência, o sinal avaliado é amostrado de forma uniforme e sua FFT é calculada, obtendo-se então $\hat{Y}(f) = \text{FFT}\{y(t)\}$.

Uma vez obtido $\hat{Y}(f)$ e definidas as características de amostragem não uniforme, são calculados os valores de μ_Y e R_Y usando as expressões (4.14) e (4.15).

O sinal, então, é amostrado não uniformemente e sua NDFT é calculada (expressão (3.3)), obtendo-se então $\tilde{Y}(f) = \text{NDFT}\{y(t)\}$.

Finalmente o $\tilde{Y}(f)$ e μ_Y são confrontados em um teste de hipótese que avalia se há diferenças significativas entre eles. A expressão (4.20), que realiza essa avaliação, pode ser reescrita então da seguinte forma:

$$\left(\frac{\Re\{\tilde{Y}(f)\} - \Re\{\mu_Y(f)\}}{\sigma_{\Re\{Y(f)\}}} \right)^2 + \left(\frac{\Im\{\tilde{Y}(f)\} - \Im\{\mu_Y(f)\}}{\sigma_{\Im\{Y(f)\}}} \right)^2 = s, \quad (4.22)$$

em que as variâncias da parte real σ_{\Re}^2 e da parte imaginária σ_{\Im}^2 de $Y(f)$ são obtidas a partir do estimador R_Y e, como os testes terão uma significância de 95%, $s = 5,991$.

A Figura 4.5 apresenta o fluxograma que resume o processo descrito.

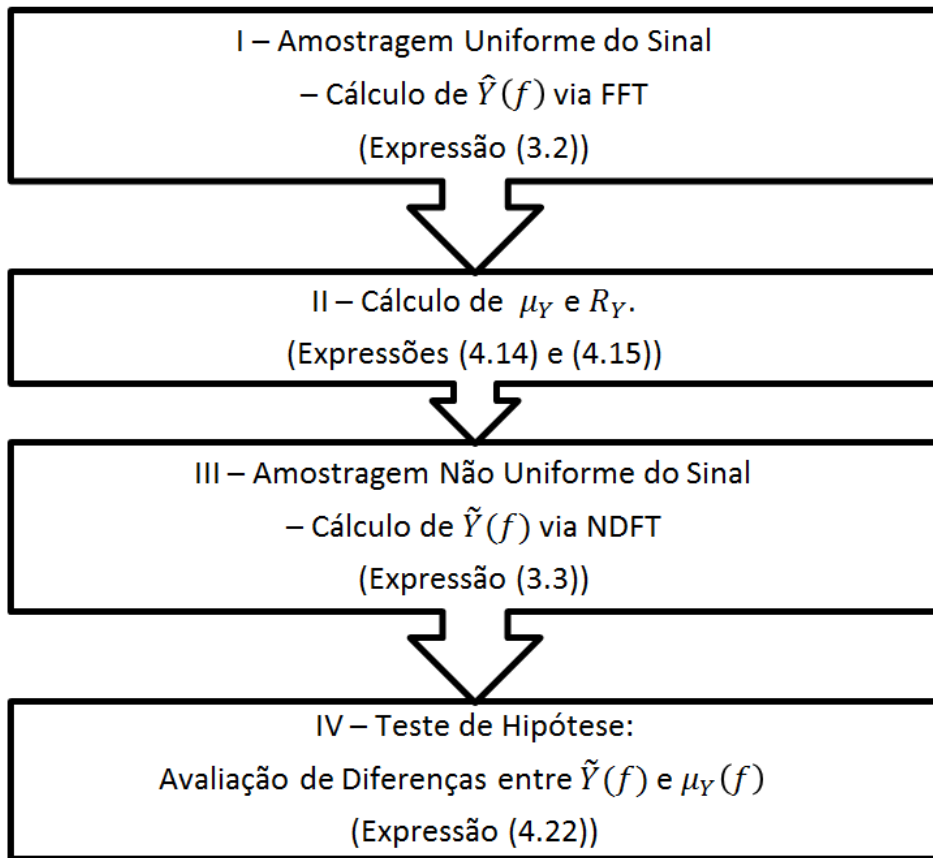


Figura 4.5: *Fluxograma do procedimento de detecção de falseamento.*

A primeira simulação, apresentada na Seção 5.1.1, é um exemplo em que os passos e resultados obtidos para um determinado sinal são apresentados com o intuito de descrever e exemplificar o procedimento aqui descrito.

Resultados Numéricos

Ao longo do Capítulo 4, foi visto que é possível, a partir dos dados de amostragem e do resultado obtido por FFT, estimar qual será o valor esperado e a matriz de covariância para a transformada de Fourier de um sinal amostrado de forma não uniforme. Foi discutido que tais dados são suficientes para a avaliação da validade dos picos de frequência obtidos pela FFT. Entretanto, como foi discutido na Seção 4.2, o erro de estimação pode inviabilizar o procedimento, dependendo do número de amostras consideradas.

Neste capítulo serão apresentados resultados de simulações com diferentes objetivos. Inicialmente serão confrontados resultados obtidos pelos estimadores e pelo método de Monte Carlo e, em seguida, serão avaliados métodos que dependem somente dos estimadores do espectro para um grande número de sinais.

5.1 Simulações

5.1.1 Primeiro Teste

O primeiro teste realizado tem, como intuito, apresentar um exemplo dos passos descritos nessa seção. Será apresentado também, para esse exemplo, o impacto do erro de estimação. Para isso, na etapa do teste de hipótese, serão utilizados os valores de espectro médio (μ_Y^{MC}) e matriz de covariância (R_Y^{MC}) obtidos pelo método de Monte Carlo. Esses resultados serão comparados com um teste de hipótese que utilize os estimadores μ_Y e R_Y obtidos por meio das Equações 4.14 e 4.15.

Para o teste, o sinal avaliado foi:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 32 \text{ Hz}, f_2 = 208 \text{ Hz}, f_3 = 604 \text{ Hz}, \quad (5.1)$$

sendo amostrado a uma taxa de 1024 amostras/segundo durante 0,25s. A

Figura 5.1 apresenta o módulo da FFT do sinal.

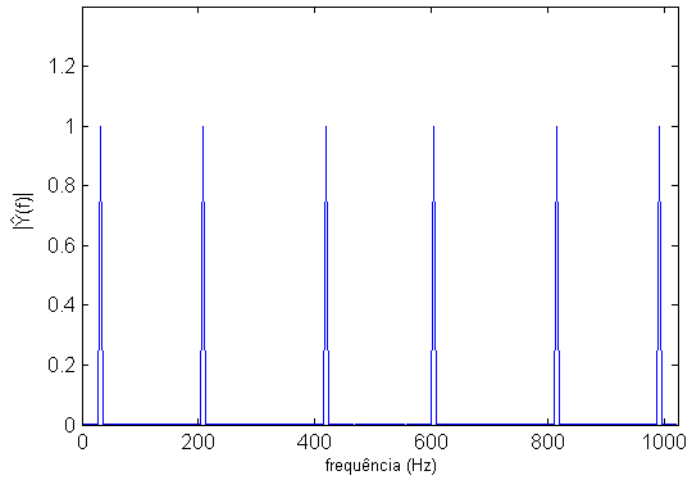


Figura 5.1: $|\hat{Y}(f)|$. $T_J = 0,25 s$ $f_s = 1024 Hz$.
 $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32 Hz$, $f_2 = 208 Hz$, $f_3 = 604 Hz$.

Ao longo deste capítulo, os gráficos serão exibidos com as frequências apresentadas no intervalo $(0 \bar{f}_s)$. Sendo assim, na Figura 5.1, considerando a metade inferior do gráfico, apenas dois picos representam adequadamente as frequências do sinal: 32 Hz e 208 Hz, sendo que o pico em 420 Hz é a representação de um pico falseado.

O mesmo sinal (Equação 5.1) é reamostrado, porém de forma não uniforme. As características de taxa de amostragem média de 1024 amostras/segundo e janela de dados de 0,25s são mantidas. Utilizando as Equações 4.14 e 4.15, encontra-se o valor esperado e a covariância para cada frequência do espectro. Tais informações são apresentadas na Figura 5.2, por meio do valor esperado do espectro e de uma banda de confiança dada por um desvio padrão.

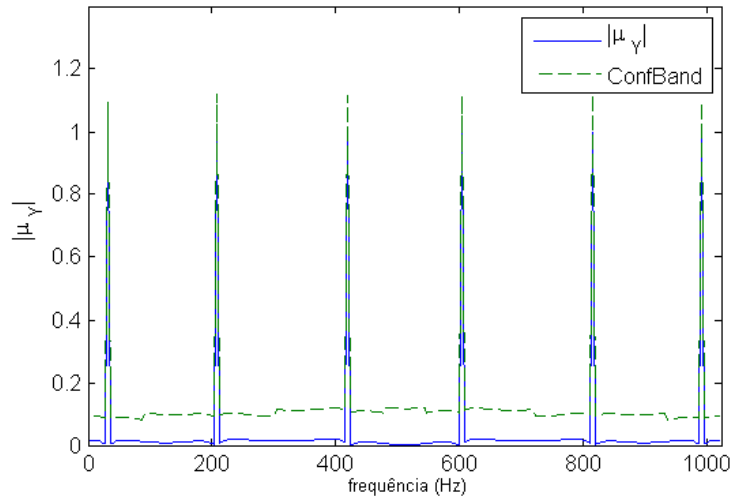


Figura 5.2: Valor esperado $|\mu_Y(f)|$ e banda de confiança $|\mu_Y \pm \sqrt{R_Y}|$, calculados usando Equações 4.14 e 4.15. $T_J = 0,25$ s, $\bar{f}_s = 1024$ Hz, método ARS:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 32 \text{ Hz}, f_2 = 208 \text{ Hz}, f_3 = 604 \text{ Hz}.$$

Para facilitar a visualização, um dos picos da Figura 5.2 é mostrado em detalhe na Figura 5.3:

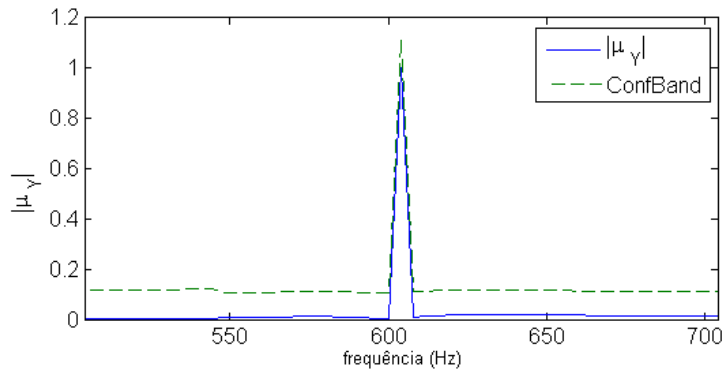


Figura 5.3: Detalhe da Figura 5.2: $|\mu_Y(f)|$ e $|\mu_Y \pm \sqrt{R_Y}|$ do sinal:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 32 \text{ Hz}, f_2 = 208 \text{ Hz}, f_3 = 604 \text{ Hz}.$$

Para efeitos de comparação, o mesmo gráfico da Figura 5.2 é calculado, porém via Monte Carlo. Neste método são realizadas 1000 amostragens não uniformes diferentes do mesmo sinal e, para cada uma é calculada $\tilde{Y}(f)$. $\mu_Y^{MC}(f)$ é obtido avaliando-se a média dos resultados de $\tilde{Y}(f)$ e R_Y^{MC} , sua

variância.

O resultado é apresentado na Figura 5.4

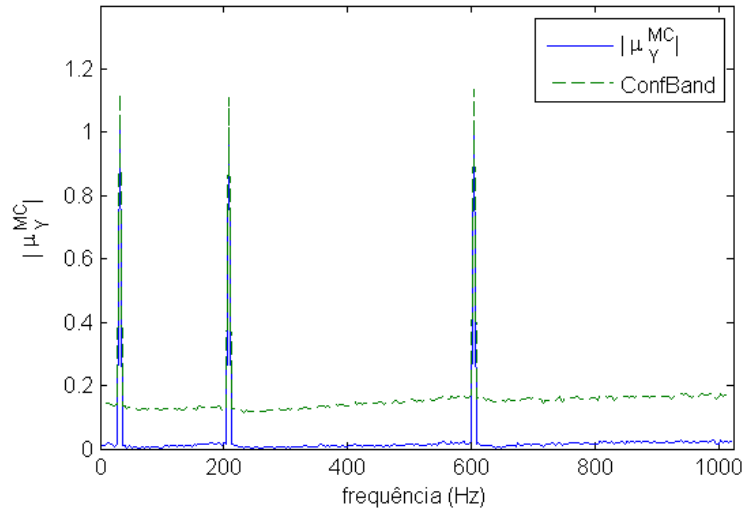


Figura 5.4: Valor esperado $|\mu_Y^{MC}(f)|$ e banda de confiança $|\mu_Y^{MC} \pm \sqrt{R_Y^{MC}}|$, determinados via simulação Monte Carlo. Obtido usando 1000 realizações. $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32$ Hz, $f_2 = 208$ Hz, $f_3 = 604$ Hz. $T_J = 0,25$ s $f_s = 1024$ Hz, método ARS.

Para facilitar a visualização, um dos picos da Figura 5.4 é mostrado em detalhe na Figura 5.5:

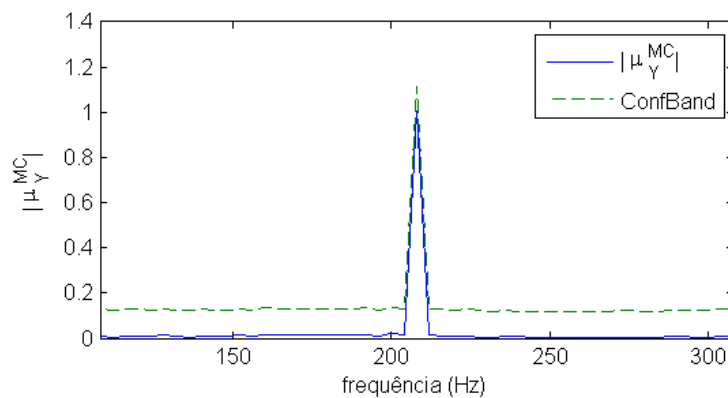


Figura 5.5: Detalhe da Figura 5.4: $|\mu_Y^{MC}(f)|$ e $|\mu_Y^{MC} \pm \sqrt{R_Y^{MC}}|$ do sinal: $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32$ Hz, $f_2 = 208$ Hz, $f_3 = 604$ Hz.

Como o gráfico da Figura 5.4 não foi obtido usando a FFT do sinal, as frequências acima de 512Hz não aparecem na forma de sinais falseados. Observa-se que, desconsiderando isso, os valores esperados e as bandas de confiança mostrados na Figura 5.2 são semelhantes.

O sinal é então amostrado então de forma não uniforme e seu espectro é apresentado na Figura 5.6.

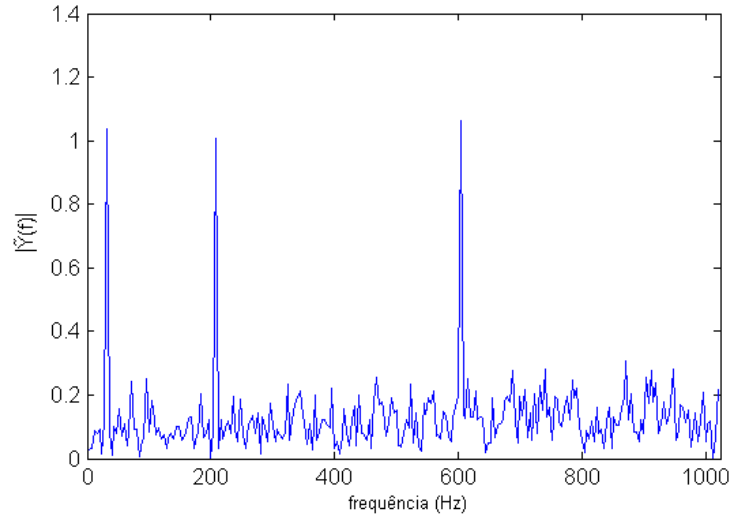


Figura 5.6: $|\tilde{Y}(f)|$ obtida a partir de 256 amostras do sinal:
 $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32$ Hz, $f_2 = 208$ Hz, $f_3 = 604$ Hz.
 $T_J = 0,25$ s $\bar{f}_s = 1024$ Hz, método ARS.

Para a realização do teste de hipótese, será utilizada a Expressão (4.20), de forma a confrontar o $\tilde{Y}(f)$ com o $\mu_Y(f)$. Porém, como foi explicado, para um primeiro teste, será confrontado o resultado obtido pela NDFT com o estimador de valor esperado obtido pela Expressão (4.14), utilizando como referência a variância de cada frequência obtida pelo método de Monte Carlo. Em resumo, será utilizada a expressão

$$\left(\frac{\Re \{ \tilde{Y}(f) \} - \Re \{ \mu_Y(f) \}}{\sigma_{\Re\{Y^{MC}(f)\}}} \right)^2 + \left(\frac{\Im \{ \tilde{Y}(f) \} - \Im \{ \mu_Y(f) \}}{\sigma_{\Im\{Y^{MC}(f)\}}} \right)^2 = s, \quad (5.2)$$

em que $\sigma_{\Re\{Y^{MC}(f)\}}^2$ representa a variância da parte real de $Y(f)$ no método de Monte Carlo e $\sigma_{\Im\{Y^{MC}(f)\}}^2$ tem o mesmo significado, porém para a parte imaginária.

Na Expressão 5.2, como o valor da variável s depende do nível de significância do teste, para a apresentação do resultado, foram escolhidos os valores 5,991 e 9,210, que representam 95% e 99%, respectivamente. Como o eixo de frequências avaliadas é extenso, a apresentação dos valores obtidos para cada uma das frequências, por meio de tabelas, não forneceria uma visão clara do resultado. Porém, como para cada teste o resultado é um valor escalar, é possível apresentar o resultado de todas as frequências em um único gráfico. A Figura 5.7 apresenta o resultado dos testes de hipótese que avaliam, para cada frequência, se há diferenças significativas entre $\tilde{Y}(f)$ e $\mu_Y(f)$. Os valores que superam os limites apresentados são as frequências nas quais rejeita-se a hipótese nula de que não há diferenças significativas e, conseqüentemente, indica as frequências nas quais rejeita-se os valores obtidos para $\hat{Y}(f)$ como sendo válidos.

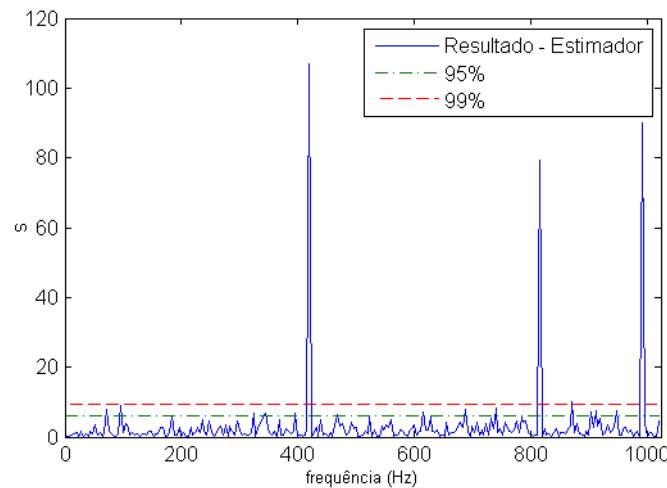


Figura 5.7: *Resultado do Teste de Hipótese em que o sinal*

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 32 \text{ Hz}, f_2 = 208 \text{ Hz}, f_3 = 604 \text{ Hz},$$

amostrado a $f_s = 1024 \text{ Hz}$, é avaliado utilizando-se $\mu_Y(f)$ e $\sigma_{YMC}(f)^2$.

Fica muito claro que os três picos que representam as frequências falseadas (420 Hz, 816 Hz e 992 Hz) obtiveram um valor muito acima dos limites referentes aos níveis de significância de 95% e 99%. Algumas das demais frequências foram rejeitadas ao se considerar a significância de 95% e isso ocorre devido a serem pontos em que o erro associado ao cálculo de $\tilde{Y}(f)$ foi muito elevado, porém, como tais valores não apresentaram valores significativos no cálculo de $\hat{Y}(f)$, sabe-se que não existe energia nestes valores

de frequência. Cabe lembrar que, nesta primeira simulação, o sinal avaliado é completamente livre de ruído na análise da FFT e, em uma simulação mais real, a situação poderia ser diferente. Tal discussão é apresentada nas próximas seções.

Um ponto interessante a ser apresentado é como seria o gráfico obtido, se, ao invés da utilização do valor esperado obtido por estimador, fosse utilizado o valor médio obtido pelo método de Monte Carlo. No caso, a expressão que representaria o teste seria:

$$\left(\frac{\Re \{ \tilde{Y}(f) \} - \Re \{ \mu_Y^{MC}(f) \}}{\sigma_{\Re \{ Y^{MC}(f) \}}} \right)^2 + \left(\frac{\Im \{ \tilde{Y}(f) \} - \Im \{ \mu_Y^{MC}(f) \}}{\sigma_{\Im \{ Y^{MC}(f) \}}} \right)^2 = s. \quad (5.3)$$

Tal resultado é apresentado na Figura 5.8.

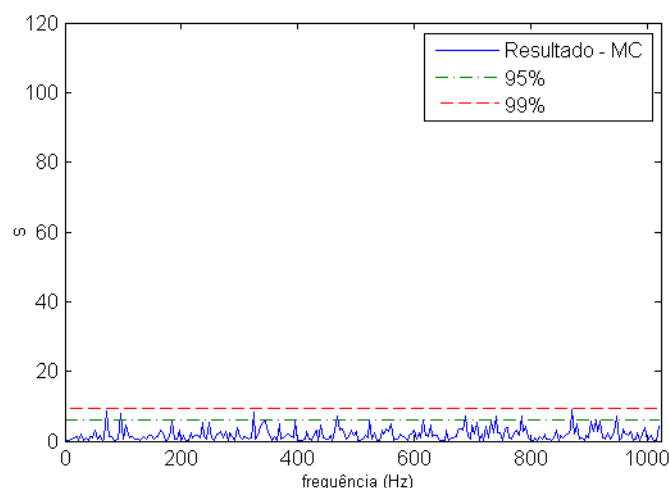


Figura 5.8: *Resultado do Teste de Hipótese em que o sinal*

$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32$ Hz, $f_2 = 208$ Hz, $f_3 = 604$ Hz, amostrado a $f_s = 1024$ Hz é avaliado utilizando-se $\mu_Y^{MC}(f)$ e $\sigma_{Y^{MC}}(f)^2$.

A ausência de picos que ultrapassem a linha tracejada vermelha, indica que não é possível afirmar que existem diferenças significativas entre $\tilde{Y}(f)$ e $\mu_Y^{MC}(f)$ com um nível de significância de 99%. Cabe salientar que, exceto pelos picos que representam as frequências falseadas, a Figura 5.8 é praticamente idêntica à Figura 5.7. Para explicitar o ponto, o resultado dos dois testes foi sobreposto e apresentado na Figura 5.9.

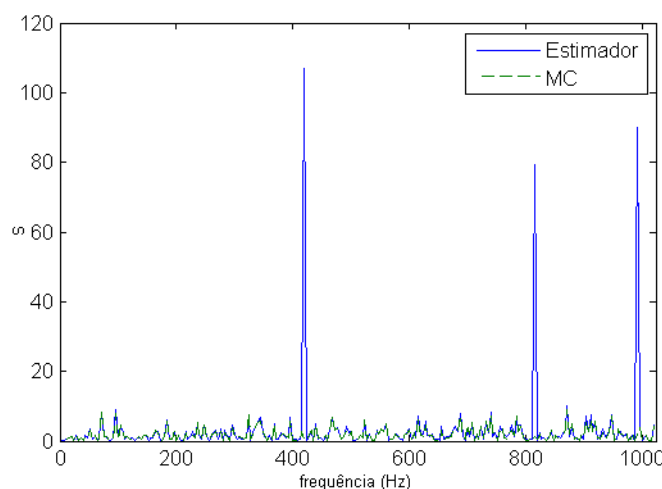


Figura 5.9: *Comparação entre resultados de Teste de Hipótese. Sinal avaliado:*
 $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 32$ Hz, $f_2 = 208$ Hz, $f_3 = 604$ Hz,
 amostrado a $\bar{f}_s = 1024$ Hz.
 Linha contínua azul: Teste é avaliado utilizando-se $\mu_Y(f)$ e $\sigma_{Y^{MC}}(f)^2$.
 Linha tracejada verde: Teste é avaliado utilizando-se $\mu_Y^{MC}(f)$ e $\sigma_{Y^{MC}}(f)^2$.

Aqui, fica claro que, apesar do estimador de covariância do espectro ser problemático, o estimador de valor esperado é bem robusto e pode ser utilizado para testes de hipótese com quantidade informação e de ruído semelhante ao teste realizado.

5.1.2 Teste em Número Elevado de Sinais

Na Seção 5.1.1, foi realizado um teste que ilustra a sequência de passos do procedimento proposto nesse trabalho. Nele, observou-se que é possível detectar, por meio de uma avaliação objetiva, frequências falseadas no resultado da FFT. Entretanto, é desejável realizar a avaliação de resultados de uma forma que seja diferente, pelo menos, em dois quesitos: O primeiro ponto é a utilização de resultados obtidos via Monte Carlo. Como a proposta do procedimento envolve uma quantidade de memória limitada, não faz sentido uma ferramenta que dependa da realização de dezenas ou centenas de medições. O segundo ponto é a realização de uma quantidade maior de sinais avaliados. A partir de um número maior de sinais variados, com ruídos diferentes a cada realização, é possível ter uma visão crítica mais bem fundamentada do procedimento. Assim sendo, serão realizados testes de hipótese

considerando os estimadores μ_Y e R_Y , calculados a partir das convoluções da FFT do sinal com μ_W e R_W . A expressão aqui utilizada então será:

$$\left(\frac{\Re \{ \tilde{Y}(f) \} - \Re \{ \mu_Y(f) \}}{\sigma_{\Re\{Y(f)\}}} \right)^2 + \left(\frac{\Im \{ \tilde{Y}(f) \} - \Im \{ \mu_Y(f) \}}{\sigma_{\Im\{Y(f)\}}} \right)^2 = s, \quad (5.4)$$

sendo $\sigma_{\Re\{Y(f)\}}^2$ e $\sigma_{\Im\{Y(f)\}}^2$, respectivamente, a variância da parte real e a variância da parte imaginária da transformada, obtidas do estimador $R_Y(f)$ (Equação 4.15).

Tal avaliação será importante para ver o impacto da utilização dos estimadores $\mu_Y(f)$ e $R_Y(f)$. Para os testes, serão gerados sinais aleatórios definidos por:

$$\begin{aligned} x(t) &= \sum_{i=1}^3 \text{sen}(2\pi f_i t) + e(t), \\ t &\in \{t_1, t_2, \dots, t_N\}, \end{aligned} \quad (5.5)$$

em que f_i é uma variável aleatória sorteada de uma distribuição uniforme $U[0, \bar{f}_s]$, e $e(t)$ é um ruído gaussiano branco (AWGN). Serão realizados testes que levarão em conta dois fatores, cada um com três níveis: Tamanho de memória de 128, 256 e 512 posições para o armazenamento do sinal e relação sinal ruído¹ de 20 dB, 0 dB e -20 dB.

As simulações realizadas consistem em, para cada uma das nove combinações de fatores, avaliar cinco testes, cada um composto por cem realizações. Nas simulações, tanto os sinais quanto o ruído assumem valores diferentes a cada realização. Como os valores de f_i têm 50% de chance ser maior que $\frac{1}{2}\bar{f}_s$ e aparecer como falseamento em μ_Y , avaliou-se a capacidade do algoritmo de detectá-las corretamente, caso ocorram.

No primeiro teste, com o intuito de eliminar a possibilidade de ocorrência de vazamento espectral (avaliado posteriormente), criou-se a limitação que só seriam possíveis frequências sorteadas que tivessem uma quantidade inteira de períodos na janela temporal avaliada. A Tabela 5.1 apresenta os resultados desse primeiro teste, em que os valores representam a porcentagem de acertos ao se detectar as frequências falseadas.

Como era esperado, os resultados foram melhores para situações de valores de SNR elevados e maior número de amostras.

¹ A razão sinal-ruído (SNR) é definida como a razão, em dB, entre a potência do sinal de interesse e a potência do ruído. $\text{SNR} = 10 \log_{10} \left(\frac{P_{\text{sinal}}}{P_{\text{ruído}}} \right)$.

Tabela 5.1: Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Sem Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(87 ± 3)%	(63 ± 4)%	(22 ± 2)%
256	(94 ± 2)%	(83 ± 2)%	(20 ± 2)%
512	(97 ± 1)%	(94 ± 1)%	(22 ± 2)%

Outro resultado obtido foi a porcentagem de frequências verdadeiras que o algoritmo acusou, inadequadamente, como falseamento, apresentado na Tabela 5.2:

Tabela 5.2: Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas - Sem Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(6 ± 1)%	(7 ± 2)%	(18 ± 3)%
256	(4 ± 1)%	(4 ± 1)%	(18 ± 3)%
512	(3 ± 1)%	(2 ± 1)%	(16 ± 2)%

Nesta segunda avaliação, observou-se que muitas das frequências consideradas falseadas erroneamente foram assim avaliadas devido a casos específicos de sinais avaliados. Nestes, a raia falseada de um sinal de alta frequência se posicionava exatamente na posição de um sinal de baixa frequência. Ainda assim, os resultados do teste se comportam como era esperado. Sinais com menos ruído e mais amostras são avaliados com uma taxa de sucesso mais elevada.

Os mesmos resultados são agora calculados, porém sem a preocupação de evitar o vazamento espectral. Nestes casos, são permitidos, dentro do limite $[0, \overline{f_s}]$, qualquer valor de frequência. A Tabela 5.3 apresenta a taxa de sucesso ao se identificar corretamente as frequências falseadas.

Tabela 5.3: Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Com Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(25 ± 2)%	(24 ± 1)%	(21 ± 2)%
256	(21 ± 2)%	(21 ± 2)%	(19 ± 1)%
512	(22 ± 2)%	(22 ± 2)%	(19 ± 4)%

Observa-se uma queda drástica na capacidade do algoritmo de identificar

as frequências falseadas, o que mostra o quão impactante é o vazamento espectral. A Tabela 5.4 apresenta, nesse caso, a porcentagem de frequências verdadeiras que foram erroneamente consideradas falseadas.

Tabela 5.4: Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas - Com Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	$(2 \pm 1)\%$	$(2 \pm 1)\%$	$(18 \pm 2)\%$
256	$(0,5 \pm 1)\%$	$(0,5 \pm 1)\%$	$(18 \pm 1)\%$
512	$(0,5 \pm 1)\%$	$(0,5 \pm 1)\%$	$(16 \pm 1)\%$

Aqui, foi encontrado um caso interessante pois os resultados são melhores que os apresentados no caso sem vazamento. Isso ocorre pois, nesta situação, a porcentagem de sinais compostos de frequências cujas raias falseadas coincidem com raias verdadeiras é significativamente menor.

Com o intuito de reduzir o efeito negativo causado pelo vazamento espectral, os testes, agora, foram refeitos porém utilizando a técnica de janelamento nos dados obtidos por amostragem uniforme e não uniforme. A janela utilizada em ambos os tipos de amostragem é a janela de *Hanning*. Os resultados são apresentados nas Tabelas 5.5 e 5.6

Tabela 5.5: Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Pós Janelamento

Amostras	20dB	0dB	-20dB
128	$(61 \pm 3)\%$	$(60 \pm 2)\%$	$(31 \pm 2)\%$
256	$(59 \pm 2)\%$	$(60 \pm 2)\%$	$(33 \pm 4)\%$
512	$(60 \pm 2)\%$	$(59 \pm 3)\%$	$(32 \pm 2)\%$

Tabela 5.6: Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas Pós Janelamento

Amostras	20dB	0dB	-20dB
128	$(3 \pm 1)\%$	$(7 \pm 2)\%$	$(28 \pm 3)\%$
256	$(2 \pm 1)\%$	$(4 \pm 1)\%$	$(28 \pm 3)\%$
512	$(1 \pm 1)\%$	$(2 \pm 1)\%$	$(28 \pm 3)\%$

Apesar da clara melhora dos resultados obtidos com a técnica de janelamento, as taxas de acerto ainda ficaram abaixo de 65%.

Ao analisar os sinais avaliados, um a um, no último teste, em procura do motivo da baixa taxa de sucesso, foi possível observar um ponto muito interessante. Observe na Figura 5.10 as transformadas uniforme e não uniforme de um dos sinais considerados no teste cujos resultados são apresentados na Tabela 5.5.

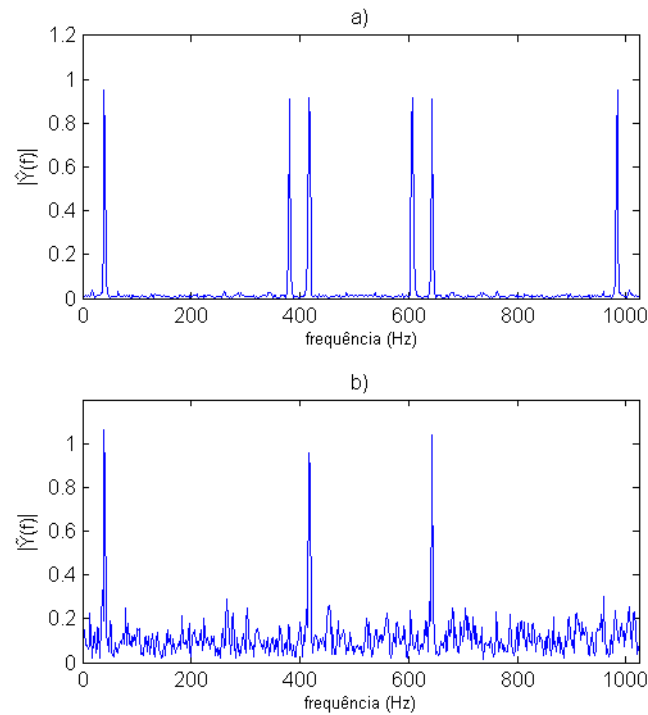


Figura 5.10: *Transformadas do sinal:*

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t),$$

$$f_1 = 40,6 \text{ Hz}, f_2 = 416,8 \text{ Hz}, f_3 = 643,2 \text{ Hz},$$

de uma realização com $N = 512$ Amostras e $SNR = 20dB$.

$$a) |\widehat{Y}(f)|, f_s = 1024 \text{ Hz}.$$

$$b) |\widetilde{Y}(f)|, \bar{f}_s = 1024 \text{ Hz}.$$

A partir de uma análise visual, fica claro, das seis raias encontradas na FFT, quais das três são frequências verdadeiras e quais são falseadas. Ainda assim, ao se realizar os testes de hipóteses, nenhuma das raias de $|\widehat{Y}(f)|$ é rejeitada. O motivo da não rejeição das raias se deve aos valores mal condicionados obtidos ao se calcular a matriz de covariância de cada frequência. A Figura 5.11 apresenta os valores obtidos no teste de hipótese desse sinal.

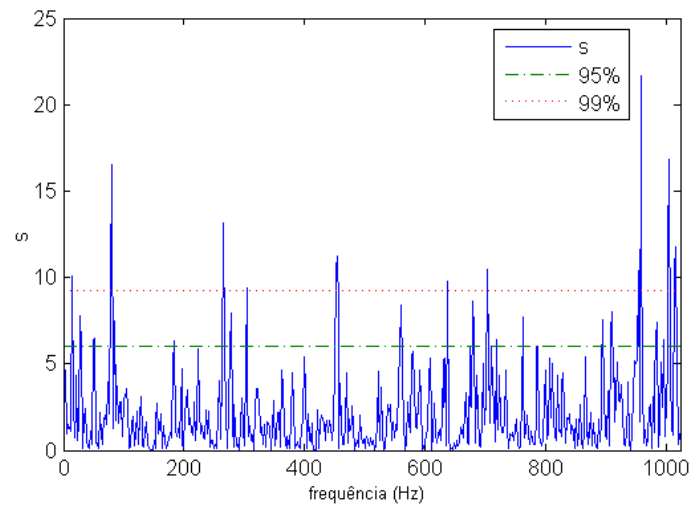


Figura 5.11: Resultado do teste de hipótese para o sinal:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t),$$

$$f_1 = 40,6 \text{ Hz}, f_2 = 416,8 \text{ Hz}, f_3 = 643,2 \text{ Hz},$$

de uma realização com $N = 512$ Amostras e $SNR = 20dB$.

Apesar da grande diferença entre $\tilde{Y}(f)$ e $\mu_Y(f)$, os valores de variância nas frequências falseadas são tão elevados que o parâmetro s avaliado no teste de hipótese não assume valores maiores que 5,991. Para as frequências falseadas, 380,8 Hz, 643,2 Hz e 983,4 Hz, os valores do parâmetro s foram, respectivamente, 4,49, 0,03 e 3,91.

Como visto nos resultados apresentados, uma grande limitação ocorre devido à utilização de estimadores no cálculo da matriz de covariância dos valores de frequências do espectro. Tal ponto deve ser tratado com muita atenção e é interessante que algum tratamento de dados possa ser realizado de forma a obter resultados mais satisfatórios.

Como um teste da possibilidade do tratamento das matrizes de covariância obtidos para o espectro, foi realizado mais um teste. É possível observar nas Figuras 5.2 e 5.4 que as bandas de confiança, obtidas a partir de $|\mu_Y \pm \sqrt{R_Y}|$, apesar de ter uma leve oscilação, não possuem valores muito discrepantes de um valor de covariância médio do espectro. Para avaliar se a manipulação dos dados contidos em R_Y gera uma melhora nos resultados,

o teste de hipótese foi modificado para:

$$\left(\frac{\Re \{ \tilde{Y}(f) \} - \Re \{ \mu_Y(f) \}}{\frac{1}{N_0} \sum_{i=1}^{N_0} (\sigma_{\Re\{Y(f_i)\}})} \right)^2 + \left(\frac{\Im \{ \tilde{Y}(f) \} - \Im \{ \mu_Y(f) \}}{\frac{1}{N_0} \sum_{i=1}^{N_0} (\sigma_{\Im\{Y(f_i)\}})} \right)^2 = s, \quad (5.6)$$

de forma que, ao realizar o teste de hipótese, ao avaliar cada frequência da transformada, ao invés de se considerar a variância somente daquela determinada frequência, é considerada uma média das variâncias de todas as N_0 frequências (inclusive, dos picos). De um ponto de vista, é como se, ao se obter os gráficos apresentados nas Figuras 5.2 e 5.4, à linha tracejada verde, fosse aplicado um filtro passa-baixas, de forma que ela passasse a ter um valor contínuo.

Os resultados destes testes são apresentados nas Tabelas 5.7 e 5.8

Tabela 5.7: Porcentagem de Sucesso no Teste de Hipótese na Detecção de Raias Falseadas - Pós Filtragem de R_Y

Amost.	20dB	0dB	-20dB
128	(99,1 ± 0,4)%	(59,5 ± 1,8)%	(31,2 ± 1,7)%
256	(99,8 ± 0,3)%	(99,2 ± 0,6)%	(30,3 ± 1,8)%
512	(99,9 ± 0,2)%	(99,9 ± 0,2)%	(33,9 ± 1,5)%

Tabela 5.8: Porcentagem de Erro no Teste de Hipótese por Interpretar Raias Verdadeiras como Falseadas - Pós Filtragem de R_Y

Amost.	20dB	0dB	-20dB
128	(6,5 ± 1,5)%	(7,5 ± 1,7)%	(28,4 ± 3,4)%
256	(3,5 ± 1,1)%	(11,7 ± 1,9)%	(26,3 ± 2,6)%
512	(1,3 ± 0,2)%	(10,1 ± 1,5)%	(24,9 ± 1,4)%

Observou-se uma melhora significativa nos resultados, indicando que tal filtragem é adequada para o tipo de sinal avaliado. Para os sinais que possuem uma SNR de 20dB, nos três testes foi obtido uma detecção adequada de frequências falseadas acima de 99%.

5.2 Detecção de Frequências Acima de $f_s/2$

Até esta seção, os testes tiveram o intuito de avaliar se as frequências obtidas em espectros gerados por FFT eram falseadas. Os métodos utiliza-

dos podem ser considerados de custo muito elevado para um resultado que poderia ser obtido de outras maneiras. Por exemplo, num caso mais simples, para detecção poderia ser utilizado um filtro *anti-alias* ou então, em uma solução mais complexa, poderia ser realizada a lenta variação da taxa de amostragem uniforme e a observação de deslocamento das raias no espectro indicaria falseamento. Entretanto, desenvolvimento apresentado pode ser um pontapé inicial na criação de ferramentas que possam superar problemas mais complexos de solução desconhecida.

O teste descrito nesta seção tem como objetivo avaliar a presença de energia em frequências que possam ser muitas vezes maiores que $\frac{1}{2}f_s$. Esse é um teste inicial e seu desenvolvimento mais detalhado será considerado parte das possibilidades de pesquisas futuras.

O teste de avaliação de frequências muito superiores ao limite de *Nyquist* é semelhante ao teste apresentado na Seção 5.1.2. A sequência de passos será descrita a seguir, ilustrada por gráficos gerados em uma de suas realizações.

Como nos testes anteriores, inicialmente o sinal é amostrado uniformemente e sua FFT é calculada. O sinal no exemplo apresentado é

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 344 \text{ Hz}, f_2 = 712 \text{ Hz}, f_3 = 1296 \text{ Hz}, \quad (5.7)$$

em que $f_s = 1024$ amostras/segundo e $\text{SNR} = 20\text{dB}$. O módulo de sua FFT é apresentado na Figura 5.12

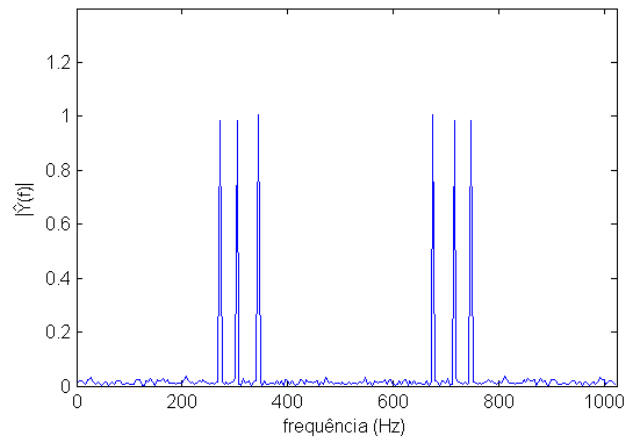


Figura 5.12: $|\hat{Y}(f)|$ obtida a partir de 256 amostras do sinal:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 344 \text{ Hz}, f_2 = 712 \text{ Hz}, f_3 = 1296 \text{ Hz}.$$

$$f_s = 1024 \text{ Hz}, \text{ SNR} = 20 \text{ dB}.$$

Em seguida, o sinal é amostrado de forma não uniforme, e sua NDFT é calculada, estendendo-se o espectro até a frequência de interesse, f_{final} . Neste exemplo, a $\tilde{Y}(f)$ será calculada até a frequência de 2048 Hz. O gráfico exemplo da realização é apresentado na Figura 5.13.

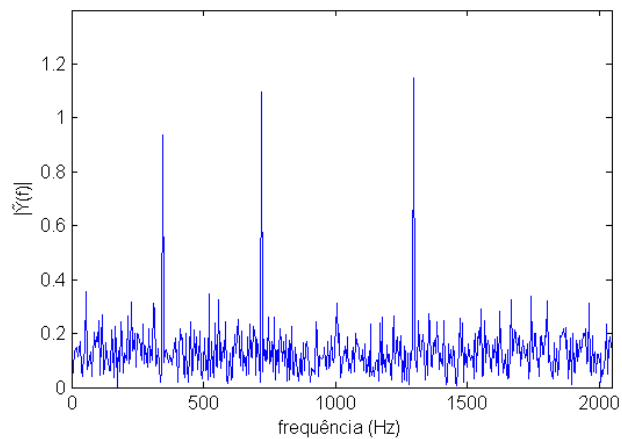


Figura 5.13: $|\tilde{Y}(f)|$ obtida a partir de 256 amostras do sinal:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 344 \text{ Hz}, f_2 = 712 \text{ Hz}, f_3 = 1296 \text{ Hz}.$$

$$\bar{f}_s = 1024 \text{ Hz}, \text{ SNR} = 20 \text{ dB}, \text{ método ARS}$$

Como a hipótese nula é definida como $\hat{Y}(f)$ sendo fidedigna, é esperado que o sinal avaliado não possua energia em frequências superiores a $\frac{1}{2}\bar{f}_s$. A

FFT calculada de um sinal amostrado adequadamente, mesmo que livre de falseamento, gera um sinal, no domínio da frequência, periódico com picos por toda sua extensão. No caso da NDFFT, o sinal obtido, no domínio da frequência, não é periódico e pode ser estendido a frequências arbitrariamente altas. Neste exemplo, ao se avaliar $\mu_Y(f)$, como esse é um estimador de valor esperado de um espectro amostrado não uniformemente, é esperado que seu resultado, assim como o resultado de $\tilde{Y}(f)$, não possua um comportamento periódico. Então, para sua obtenção, é preciso que as frequências maiores que $\frac{1}{2}\overline{f_s}$ de $\hat{Y}(f)$ sejam desconsideradas.

Sendo assim, antes de se avaliar os estimadores de espectro esperado e sua matriz de covariância, $\hat{Y}(f)$ é considerada nula no intervalo $[\frac{1}{2}\overline{f_s} \ f_{final}]$. O gráfico exemplo da realização é apresentado na Figura 5.14.

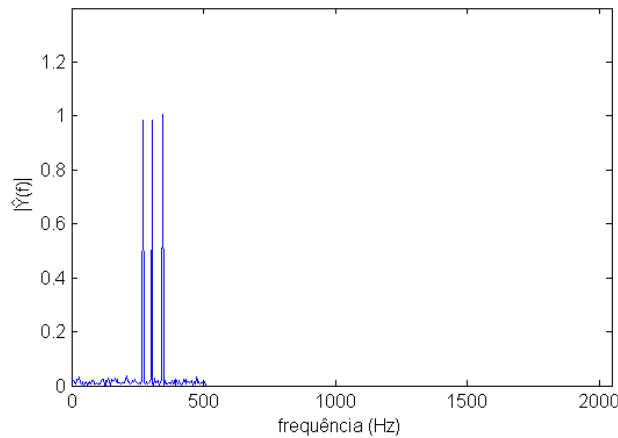


Figura 5.14: $|\hat{Y}(f)|$ expandida até f_{final} , obtida a partir de 256 amostras do sinal:
 $y(t) = \sum_{i=1}^3 \cos(2\pi f_i t)$, $f_1 = 344$ Hz, $f_2 = 712$ Hz, $f_3 = 1296$ Hz
 $f_s = 1024$ amostras/segundo, $SNR = 20$ dB.

Então, a partir das Equações (4.14) e (4.15), os estimadores de valor esperado e matriz de covariância do espectro, obtidos por amostragem não uniforme, são avaliados. Um cuidado é necessário nesse ponto, pois, ao se avaliar os valores de μ_W e R_W , é preciso que estes sejam calculados, pelo menos, no intervalo $[-f_{final} \ f_{final}]$. A Figura 5.15 ilustra μ_W e sua margem de confiança calculados para o exemplo de realização e a Figura 5.16, μ_Y e sua margem de confiança.

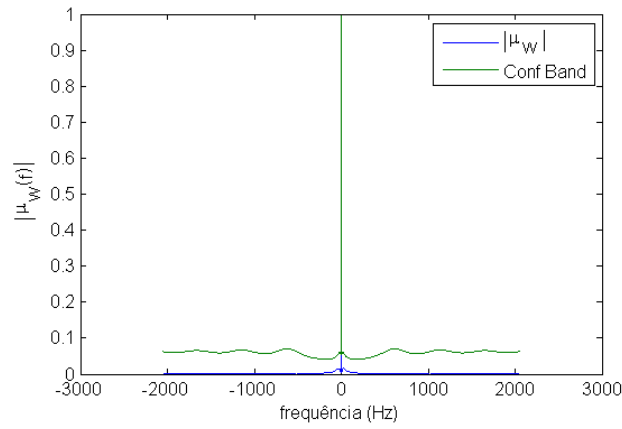


Figura 5.15: $\mu_W(f)$ obtida a partir de 256 amostras - $\bar{f}_s = 1024$.

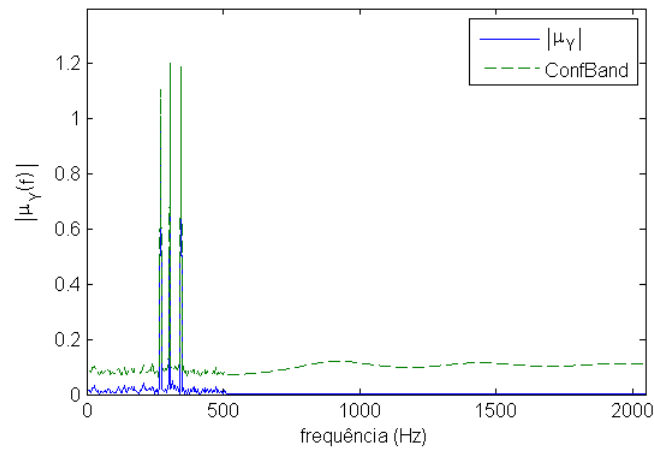


Figura 5.16: $\mu_Y(f)$ e margem de confiança expandidas até f_{final} .

Obtida a partir de 256 amostras do sinal:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 344 \text{ Hz}, f_2 = 712 \text{ Hz}, f_3 = 1296 \text{ Hz},$$

$$\bar{f}_s = 1024 \text{ amostras/segundo}, \text{ SNR} = 20\text{dB}.$$

Então, finalmente, confronta-se $\tilde{Y}(f)$ e $\mu_Y(f)$, da mesma forma como foi feito na Seção 5.1.2. O resultado obtido para os testes de hipótese são apresentados na Figura 5.17, lembrando que os valores que superam os limites apresentados são as frequências nas quais a $\hat{Y}(f)$ expandida é rejeitada.

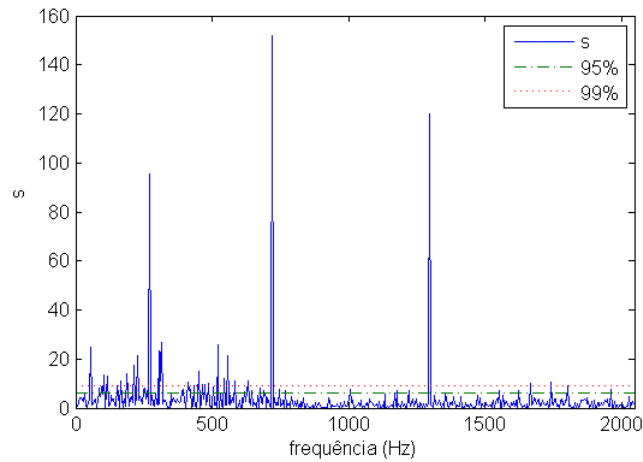


Figura 5.17: Resultado do teste de hipótese do sinal:

$$y(t) = \sum_{i=1}^3 \cos(2\pi f_i t), \quad f_1 = 344 \text{ Hz}, f_2 = 712 \text{ Hz}, f_3 = 1296 \text{ Hz},$$

$$\bar{f}_s = 1024 \text{ amostras/segundo}, \text{ SNR} = 20\text{dB}.$$

O teste aqui descrito, assim como na seção anterior, foi realizado para uma grande quantidade de sinais, com o intuito de avaliar sua taxa de sucesso para sinais compostos de frequências aleatórias. Em todos eles, os sinais foram gerados da mesma forma que na Seção 5.1.2, porém com a limitação de suas componentes não serem maiores que 2048 Hz, sendo $f_s = 1024$.

No primeiro teste, foi avaliado o sucesso em detecção de falseamento sendo, também semelhante à seção anterior, criado a limitação de o sinal avaliado ser composto somente de frequências que não geram vazamento espectral. O resultado é apresentado na Tabela 5.9:

Tabela 5.9: Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Sem Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(68 ± 2)%	(52 ± 2)%	(18 ± 2)%
256	(72 ± 3)%	(65 ± 3)%	(20 ± 2)%

Como era esperado, tanto o número de amostras quando a razão sinal-ruído afetaram os resultados, não sendo obtidas taxas de sucesso que chegassem a 75%.

Novamente, foi calculada a porcentagem de frequências verdadeiras que o algoritmo acusou, inadequadamente, como falseamento. Esse resultado é

apresentado na Tabela 5.10:

Tabela 5.10: Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Sem Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(2 ± 1)%	(2 ± 1)%	(5 ± 1)%
256	(1 ± 1)%	(1 ± 1)%	(5 ± 1)%

Os valores aqui obtidos, exceto para sinais com SNR=-20dB, foram inferiores a 3%. Como as possibilidades de componentes em frequência são maiores que as do teste similar realizado na Seção 5.1.2, a limitação relacionada a vazamento espectral não prejudicou os resultados.

O segundo teste é similar ao primeiro, porém sem a preocupação de evitar o vazamento espectral. A Tabela 5.11 apresenta a taxa de sucesso ao se identificar corretamente as frequências falseadas.

Tabela 5.11: Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Com Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(19 ± 2)%	(19 ± 3)%	(18 ± 3)%
256	(15 ± 3)%	(18 ± 2)%	(19 ± 2)%

Observa-se, também nos testes desta seção, uma queda drástica na capacidade do algoritmo de identificar as frequências falseadas devido ao vazamento espectral.

A Tabela 5.12 apresenta a porcentagem de frequências verdadeiras que foram erroneamente consideradas falseadas.

Tabela 5.12: Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Com Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(1 ± 1)%	(1 ± 1)%	(7 ± 2)%
256	(1 ± 1)%	(1 ± 1)%	(6 ± 1)%

Novamente, exceto para os valores SNR=-20dB, a porcentagem de erro ficou abaixo de 3%.

Os teste foram refeitos, utilizando a técnica de janelamento nos dados amostrados. Assim como no teste de detecção de falseamento, a janela utilizada em ambos os tipos de amostragem é a janela de *Hanning*. Os resultados são apresentados nas Tabelas 5.13 e 5.14

Tabela 5.13: Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Pós Janelamento

Amostras	20dB	0dB	-20dB
128	(29 ± 2)%	(29 ± 2)%	(19 ± 1)%
256	(29 ± 2)%	(30 ± 2)%	(21 ± 3)%

Tabela 5.14: Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Pós Janelamento

Amostras	20dB	0dB	-20dB
128	(2 ± 1)%	(5 ± 2)%	(21 ± 3)%
256	(1 ± 1)%	(2 ± 1)%	(20 ± 2)%

O janelamento dos dados amostrados, assim como na Seção 5.1.2 gerou uma pequena melhora dos resultados. Ainda assim a taxa de sucesso da detecção de falseamento, não superou 30% em nenhum dos testes.

O último teste, que realiza janelamento do sinal, foi refeito, porém, para cada frequência, foi considerado o valor médio das covariâncias do espectro, assim como apresentado na Expressão 5.6. Os resultados são apresentados nas Tabelas 5.7 e 5.16

Tabela 5.15: Porcentagem de Sucesso na Detecção de Raias Falseadas no Teste de Frequências Elevadas - Pós Filtragem

Amostras	20dB	0dB	-20dB
128	(73 ± 2)%	(66 ± 3)%	(21 ± 1)%
256	(75 ± 2)%	(74 ± 3)%	(24 ± 3)%

Tabela 5.16: Porcentagem de Erro no Teste de Frequências Elevadas por Interpretar Raias Verdadeiras como Falseadas - Com Vazamento Espectral

Amostras	20dB	0dB	-20dB
128	(3 ± 1)%	(5 ± 1)%	(6 ± 1)%
256	(2 ± 1)%	(4 ± 2)%	(7 ± 1)%

Assim como na Seção 5.1.2, a consideração de um valor médio de covariância para todas as frequências gerou resultados significativamente melhores. Entretanto ainda assim, os resultados não chegaram a uma taxa de sucesso acima de 95% como na seção citada. Por tratar os dados e partir de premissas diferentes, era esperado que os resultados desta seção se comportassem de forma diferente e, como será discutido no Capítulo 6, o estudo e desenvolvimento deste procedimento é uma oportunidade de pesquisas futuras.

Conclusões

Este trabalho teve como objetivo a utilização de uma amostragem não uniforme para a detecção de falseamento. Essa meta foi alcançada com sucesso. Uma transformada adequada foi estudada e implementada corretamente e suas características foram investigadas e entendidas. Ficou claro o motivo pelo qual ocorre o ruído da linha de base no resultado da NDFT, como descrito na Seção 3.2, e ficou claro como determinadas variáveis afetam sua variância.

A definição do procedimento para detecção de falseamento se tornou possível graças as expressões que permitiram estimar o valor esperado da transformada de Fourier de dados não uniformemente espaçados no tempo e sua matriz de covariância, descritos na Seção 4.2. Em relação ao procedimento proposto, dois grandes desafios foram superados. O primeiro sendo a dependência dos estimadores em relação ao espectro verdadeiro do sinal, que foi superada com a premissa de que a transformada dos dados uniformemente espaçados era fidedigna. E o segundo sendo a realização de um teste de hipótese com uma distribuição assimétrica, que foi superado com a mudança da avaliação para dentro do plano complexo, como descrito na Seção 4.3.

O objetivo de desenvolver um procedimento para realizar automaticamente a detecção de falseamento em sinais foi desenvolvido com sucesso. Os resultados apresentados na Seção 5.1.2 mostraram que para sinais livres de vazamento espectral, nas condições apresentadas, é possível ter taxas de sucesso de detecção de falseamento acima de 90%. Também foi mostrado o quão prejudicial, para o procedimento, pode ser o vazamento espectral, que fez com que as taxas de sucesso ficassem abaixo de 25% e que o tratamento dos dados a partir de técnicas de janelamento elevaram os resultados para valores próximos de 60%. Finalmente, os resultados dessa seção mostraram o quão importante é o tratamento dos dados obtidos para $R_Y(f)$, que por meio de uma simples operação de média, para os sinais avaliados, foram obtidos taxas de sucesso acima de 99% para mais de metade dos cenários avaliados, como é mostrado na Tabela 5.7.

Dois pontos requerem atenção especial por serem críticos nos resultados obtidos. Inicialmente, é preciso ter muito cuidado ao lidar com os estimadores de matriz de covariância do espectro. Ao lidar com uma quantidade finita de amostras, ficou claro que, das etapas realizadas, essa é a que mais sofre com a limitação indicada.

O segundo ponto é o custo computacional do procedimento. Todo o trabalho foi desenvolvido tendo em vista limitação de memória. Entretanto, a realização de convoluções duplas e de determinados laços duplos no cálculo das grandezas estimadas podem parecer de custo computacional muito elevado para a obtenção de resultados que poderiam ser obtidos de outras formas. Ainda assim, vale lembrar que as etapas que exigem mais poder computacional se encontram dentro do cálculo de $W(f)$ e esse é função somente das características de amostragem não uniforme. Ou seja, uma vez pré-determinadas as características de amostragem não uniforme, o cálculo de $W(f)$ pode ser realizado em qualquer plataforma, tendo seus resultados exportados para o dispositivo de implementação. No entanto, existem outras maneiras de lidar com detecção de falseamento de desenvolvimento mais simples e custo computacional muito mais baixo, como, por exemplo, a variação da taxa de amostragem e a observação do possível deslocamento dos picos de energia no espectro do sinal avaliado. Porém, como foi visto no último teste da dissertação, o método de detecção de falseamento é um passo inicial que pode levar ao desenvolvimento de ferramentas capazes de realizar avaliações de conteúdo espectral em frequências muito superiores ao limite de *Nyquist*.

Finalmente, cabe ressaltar que todos os resultados obtidos foram gerados a partir de um ambiente de alto nível de controle das variáveis tanto de interesse quanto espúrias. A coleta de dados gerados em situação real implementada em *hardware* seria uma excelente avaliação da capacidade prática dos procedimentos propostos.

6.1 Pesquisas Futuras

A partir dos resultados obtidos neste trabalho, surgem oportunidades de pesquisas futuras, aqui descritas.

A primeira sugestão está relacionada com os sinais utilizados nas simulações apresentadas. Deseja-se, futuramente, realizar estudos de como as decisões tomadas influenciam a análise de sinais diferentes, como por exemplo sinais caóticos que, apesar de possuir um comportamento imprevisível,

possuem conteúdo espectral bem característico, permitindo uma análise detalhada.

A segunda sugestão de estudo é referente ao tratamento do estimador de matriz de covariância. O único teste realizado com esta variável foi uma simples média de seus valores. Deseja-se realizar o estudo de diferentes filtros e tratamento desses dados, como por exemplo, a promediação dos dados, em busca de resultados ainda melhores.

A terceira sugestão está em dar sequência ao desenvolvimento do procedimento para avaliação de conteúdo espectral em frequências muito superiores ao limite de *Nyquist*. Essa é uma possibilidade de pesquisa futura certamente interessante, uma vez que foi observado, nas simulações, que as decisões tomadas no procedimento de detecção de falseamento em frequências sub-*Nyquist* não tiveram o mesmo impacto nessas avaliações em frequências elevadas. Apesar das consequências geradas pelos efeitos de vazamento espectral, janelamento e avaliação da média de $R_Y(f)$ serem semelhantes nos dois casos, nos testes frequências elevadas, não foram obtidos resultados tão bons quanto os valores de taxa de sucesso maiores que 95% apresentados na Tabela 5.7.

A quarta sugestão é a implementação em hardware do procedimento proposto para testes que envolvam coleta de dados gerados em situações práticas. Tal avaliação será interessante para verificar capacidade de detecção de falseamento em um ambiente onde não se tem controle de todas as variáveis.

Bibliografia

- Aldroubi, A. e Gröchenig, K. (2001). Nonuniform sampling and reconstruction in shift invariant spaces. *SIAM Review*, 43:585 – 620.
- Artyukh, Y., Bilinkis, I., Greitans, M., e Vedin, V. (1997a). Signal digitizing and recording in the DASP-lab system. *Proceedings of the 1997 International Workshop on Sampling Theory and Applications*, 1:357 – 360.
- Artyukh, Y., Bilinkis, I., Greitans, M., Mednieks, I., e Rybakov, A. (2001). DASP-lab system - a demonstrator of the new DSP technology. *Automatic Control and Computer Science*, 6:3 – 21 , Number 1.
- Artyukh, Y., Mednieks, V., e Vedin, V. (1997b). Virtual oscilloscope of the DASP-lab system. *Proceedings of the 1997 International Workshop on Sampling Theory and Applications*, 1:375 – 378.
- Bilinkis, I. (2007). *Digital Alias-Free Signal Processing*. John Wiley & Sons, 1st edition.
- Bland, D., Laakso, T., e Tarczynski, A. (1996). Analysis of algorithms for nonuniform-time Discrete Fourier Transform. *Circuits and Systems, 1996. ISCAS '96., Connecting the World., 1996 IEEE International Symposium*, 2:453 – 456.
- Bland, D. e Tarczynski, A. (1997). Optimum nonuniform sampling sequence for alias frequency supression. *IEEE International Symposium on Circuits and Systems*, 1:2693 – 2696.
- Boyi, N. e Deyun, X. (2012). Optimal nonuniform sampling for system identification on sparsely sampled data. *Chinese Journal of Electronics*, 21:292 – 298.

- Chen, Y., Goldsmith, A., e Eldar, Y. (2014). Channel capacity under sub-Nyquist nonuniform sampling. *IEEE Transactions on Information Theory*, 60:4739 – 4756, Number 8.
- Eng, F., Gunnarsson, F., e Gustafsson, F. (2008). Frequency domain analysis of signals with stochastic sampling times. *IEEE Transactions on Signal Processing*, 56(7):3089 – 3099.
- Hoch, J., Maciejewski, M., Mobli, M., Schuyler, A., e Stern, A. (2014). Nonuniform sampling and maximum entropy reconstruction in multidimensional NMR. *Accounts of Chemical Research*, 47:708 – 717, Number 2.
- Kunis, S. e Potts, D. (2008). Time and memory requirements of the nonequispaced fft. *Sampling Theory in Signal and Image Processing*, (7):77 – 100.
- Maymon, S., Oppenheim, A., e Fellow, L. (2011). Sinc interpolation of nonuniform samples. *IEEE Transactions on Signal Processing*, 59:4745 – 4758, Number 10.
- Mednieks, I. (1999). Methods for spectral analysis of nonuniformly sampled signals. *Proceedings of the 1999 International Workshop on Sampling Theory and Applications*, 1:190 – 193.
- Mednieks, I., Mikelsons, A., e Mucenieks, G. (1997). Virtual power analyser of the DASP-lab system. *Proceedings of the 1997 International Workshop on Sampling Theory and Applications*, 1:371 – 374.
- Papenfuß, F., Artyukh, Y., Boole, E., e Timmerman, D. (2003). Optimal sampling functions in nonuniform sampling driver designs to overcome the Nyquist limit. *Acoustics, Speech, and Signal Processing*, 6:257 – 260.
- Potts, D., Steidl, G., e Tasche, M. (2001). Fast Fourier transform for nonequispaced data: A tutorial. *Modern Sampling Theory: Mathematics and Applications*, 3:247–270.
- Shapiro, H. e Silverman, R. (1960). Alias-free sampling of random noise. *Journal of the Society for Industrial and Applied Mathematics*, 8:225 – 236.

-
- Tarczynski, A. e Allay, N. (2004). Spectral analysis of randomly sampled signals: Suppression of aliasing and sampler jitter. *IEEE Transactions On Signal Processing*, 52(12):3324 – 3334.
- Tarczynski, A. e Dongdong, Q. (2005). Optimal random sampling for spectrum estimation in DASP applications. *International Journal of Applied Mathematics and Computer Science*, 15(4):463 – 469.
- Tong, Y. (1990). *The Multivariate Normal Distribution*. Springer, 1st edition.
- Venkamani, R., Bresler, Y., e Fellow, L. (2001). Optimal sub-nyquist nonuniform sampling and reconstruction for multiband signals. *IEEE Transactions on Signal Processing*, 49:2301 – 2313.
- Xu, S., Chai, Y., Hu, Y., Jiang, C., e Li, Y. (2015). Reconstruction of digital spectrum from periodic nonuniformly sampled signals in offset linear canonical transform domain. *Optics Communications*, 348:59 – 65,.