

UNIVERSIDADE FEDERAL DE SERGIPE - UFS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA - CCET
DEPARTAMENTO DE FÍSICA - DFI

TRABALHO DE CONCLUSÃO DE CURSO - TCC

**PRINCÍPIOS DE ANÁLISE DE FOURIER APLICADOS A
PERIODOGRAMAS: O CASO LOMB-SCARGLE**

Submetido ao Departamento de Física da Universidade Federal de Sergipe como pré-requisito para se obter o grau de Licenciado em Física.

Jailton dos Santos Filho

São Cristóvão, março de 2020.

**PRINCÍPIOS DE ANÁLISE DE FOURIER APLICADOS A
PERIODOGRAMAS: O CASO LOMB-SCARGLE**

Jailton dos Santos Filho

Orientador: Prof. Dr. Raimundo Lopes de Oliveira Filho

Banca examinadora:

Prof. Dr. Raimundo Lopes de Oliveira Filho

Prof. Dr. Osmar de Souza e Silva Júnior

Prof. Dr. Luciano Casarini

“ Não há linguagem mais universal e mais simples,
mais livres de erros e de obscuridades, isto é,
mais digna de expressar as relações invariáveis das
coisas naturais [...] [A matemática] parece ser uma
faculdade da mente humana destinada a suplementar
a brevidade da vida e a imperfeição dos sentidos.”

Joseph Fourier,
Teoria analítica do calor,
Discurso preliminar (1822)

PRINCÍPIOS DE ANÁLISE DE FOURIER APLICADOS A PERIODOGRAMAS: O CASO LOMB-SCARGLE

Jailton dos Santos Filho

Resumo

Alguns fenômenos astrofísicos apresentam variabilidade periódica em suas propriedades físicas que são analisadas por meio de algoritmos e ferramentas matemático-computacionais, dentre elas o periodograma. Neste trabalho é feita uma revisão das técnicas matemáticas implementadas no desenvolvimento de um dos periodogramas mais usados em Astrofísica: o periodograma Lomb-Scargle. Ao final deste documento é apresentado um protótipo que simula um pulsar, que é uma estrela da qual se observa luz na forma de pulsos regulares e que pode ter seu período de rotação determinado através das técnicas aqui apresentadas.

1 Introdução

O desenvolvimento da representação de funções por meio de séries infinitas em termos de senos e cossenos levou Joseph Fourier (1768 – 1830) a elaborar uma análise que representasse funções periódicas (PUPIN, 2011). A abrangência dos resultados obtidos por meio desse estudo tornou relevante o trabalho desenvolvido, o qual posteriormente passou a ser chamado de Análise de Fourier. Esse tratamento, quando implementado em algoritmos computacionais, pode ser usado na análise de séries temporais astronômicas com o fim de verificar se existe uma modulação regular com um período característico.

Dentre as ferramentas computacionais que são baseadas na Análise de Fourier, estão os periodogramas. O uso dos periodogramas é comumente aplicado no estudo de variações temporais do brilho de fontes astrofísicas. Neste trabalho apresentamos o alicerce e o desenvolvimento matemático do algoritmo Lomb-Scargle, que é um dos periodogramas mais utilizados em Astrofísica.

2 Análise de Fourier

O avanço no entendimento de funções periódicas ocorrido entre os séculos XVIII e XIX foi relevante para o progresso das ciências exatas. A busca por representações alternativas de funções periódicas foi conduzida por matemáticos como Leonhard Euler, Jean d'Alembert, Daniel Bernoulli e Jean-Baptiste Joseph, o Barão de Fourier (PUPIN, 2011). Notório se tornou o trabalho de Fourier, com o desenvolvimento de uma expansão em séries infinitas para representar funções periódicas em termos de funções trigonométricas, com base em senos e cossenos. Por conta da relevância desse trabalho, esse desenvolvimento é denominado Séries de Fourier.

2.1 Funções periódicas

Uma função é periódica, por definição, se existir P tal que:

$$f(x) = f(x + P) \quad (1)$$

para todo $x + P$ no domínio de f .

Consideremos o caso em que P é o menor número real positivo que faça valer a equação anterior, de modo a introduzir o conceito de período de uma função. Nessa condição, P , denominado período fundamental, representa um incremento na variável da função tal que haja, para qualquer valor de x pertencente ao seu domínio, repetição no valor da função (SANTOS, 2004). Sendo P uma medida de repetibilidade de uma função, e partindo-se de um dado valor de x , completa-se o primeiro ciclo da função quando x evoluir para o equivalente a $x + P$, o segundo ciclo para o caso de $x + 2P$, e assim sucessivamente. Logo, para múltiplos inteiros de P tem-se o mesmo efeito e portanto é possível generalizar a Eq. 1:

$$f(x) = f(x + nP) \quad (2)$$

Os múltiplos inteiros do período fundamental da função são chamados de harmônicos. O primeiro harmônico é definido para $n = 1$, o segundo para $n = 2$, e assim por diante. A partir de P define-se Frequência (F), tal que $F = 1/P$. Dessa forma, F é uma medida da fração do ciclo decorrido um inteiro da unidade usada para medir o período. Por exemplo, seja x uma medida temporal em segundos, e P igual a $2s$. Neste caso a frequência é $0,5 s^{-1}$, ou $0,5 Hz$, o que significa que 50% do ciclo foi completado no decorrer de 1 segundo.

Quando o período em questão for o fundamental, P , daqui em diante denotado por T , tem-se dele a frequência fundamental:

$$F_0 = \frac{1}{T} \quad (3)$$

Para o tratamento de Séries de Fourier é necessário definir Frequência Angular (ω), por ser apropriada no tratamento de funções trigonométricas que são empregadas nas expansões desenvolvidas ao longo deste trabalho:

$$\omega = 2\pi F \quad (4)$$

Vale ressaltar que o número 2π vem da associação da repetibilidade de funções trigonométricas com tal valor.

Comparando as Equações 3 e 4, temos que a frequência angular fundamental é:

$$\omega_0 = \frac{2\pi}{T} \quad (5)$$

2.2 Séries de Fourier

2.2.1 Séries de Fourier para funções periódicas em intervalos iguais a 2π definidas em \mathbb{R}

Fourier tratou de encontrar uma expansão de uma função em séries infinitas utilizando as funções seno e cosseno. É sabido que as funções seno e cosseno são periódicas. ou seja, podem ser expressas por:

$$\sin(x) = \sin(x + P) \quad (6)$$

$$\cos(x) = \cos(x + P) \quad (7)$$

ou ainda por:

$$\sin(\omega x) = \sin(\omega x + P) \quad (8)$$

$$\cos(\omega x) = \cos(\omega x + P) \quad (9)$$

sendo P o período de cada função. Além disso, essas funções possuem período fundamental $T = 2\pi$, e portanto:

$$\omega_0 = \frac{2\pi}{T} = \frac{2\pi}{2\pi} = 1.$$

Então, para a função seno:

$$\sin(\omega_0 x) = \sin(\omega_0 x + P) = \sin(x + P) = \sin x.$$

E para a função cosseno:

$$\cos(\omega_0 x) = \cos(\omega_0 x + P) = \cos(x + P) = \cos x.$$

Desse modo, utilizamos o período fundamental das funções seno e cosseno para representar a expansão em série de Fourier para funções de período 2π , dada por:

$$f(x) = \sum_{n=0}^{\infty} [a_n \cos(nx) + b_n \sin(nx)] \quad (10)$$

em que:

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx \quad (11)$$

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx \quad (12)$$

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx) dx \quad (13)$$

são os coeficientes da expansão.

Para uma melhor compreensão do tratamento matemático no estudo em questão, faz-se necessário demonstrar cada um dos coeficientes supracitados seguindo a ordem na qual eles foram apresentados. Essa demonstração está apresentada no Apêndice A.

2.2.2 Séries de Fourier para funções periódicas em intervalos iguais a 2π definidas em \mathbb{C}

Seja $z = a + bi$ um número complexo tal que $Re(z) = a$ é a parte real e $Im(z) = b$ é a parte imaginária de z . Sua representação gráfica pode ser feita num sistema 2D ortogonal representando as partes imaginária e real no que é conhecido como plano complexo, como mostrado na Figura 1.

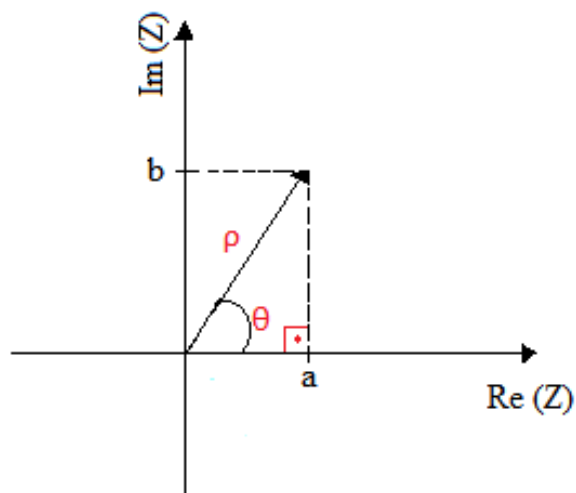


Figura 1: Plano Complexo.

Nesta representação, o ângulo θ é definido por:

$$\theta = \tan^{-1} \frac{b}{a}. \quad (14)$$

Sendo ρ o módulo de z , podemos relacionar cada parte do número complexo com as seguintes identidades trigonométricas:

$$\begin{aligned} a &= \rho \cos \theta \\ b &= \rho \sin \theta. \end{aligned}$$

Então,

$$z = \rho (\cos \theta + i \sin \theta). \quad (15)$$

Tomando a derivada de z em relação a θ , temos:

$$\frac{dz}{d\theta} = \rho (-\sin \theta + i \cos \theta).$$

Sendo $i^2 = -1$, e colocando a unidade imaginária em evidência, segue:

$$\frac{dz}{d\theta} = i\rho(\cos \theta + i \sin \theta)$$

$$\frac{dz}{d\theta} = iz.$$

Isso resulta numa equação diferencial ordinária de primeira ordem separável:

$$\frac{dz}{z} = i d\theta$$

$$\int \frac{dz}{z} = \int i d\theta$$

$$\ln z = i\theta + C.$$

Podemos assumir, sem perda de generalidade, que a constante C resultante da integração anterior tem valor nulo. Com isso,

$$z = e^{i\theta}.$$

De acordo com a Eq. 15, temos por consequência:

$$e^{i\theta} = \rho(\cos \theta + i \sin \theta).$$

Leonhard Euler particularizou o resultado acima usando o módulo do número complexo igual a 1, isto é:

$$e^{i\theta} = \cos \theta + i \sin \theta \quad (\text{ou o equivalente, } e^{-i\theta} = \cos \theta - i \sin \theta).$$

que é chamada de exponencial complexa, ou relação de Euler para números complexos na forma polar. Uma vez definida a exponencial complexa, podemos encontrar algumas relações para as funções seno e cosseno como feito a seguir.

Para qualquer n inteiro, podemos escrever:

$$e^{in\theta} = \cos(n\theta) + i \sin(n\theta).$$

Reescrevendo a relação de Euler, substituindo θ pelo número real x , temos:

$$e^{inx} = \cos(nx) + i \sin(nx).$$

Buscando uma relação para a função seno, segue que,

$$i \sin(nx) = e^{inx} - \cos(nx).$$

Isto é,

$$2i \sin(nx) - i \sin(nx) = e^{inx} - \cos(nx)$$

$$2i \sin(nx) = e^{inx} - e^{-inx}.$$

Portanto,

$$\sin(nx) = \frac{e^{inx} - e^{-inx}}{2i}. \quad (16)$$

Com manipulação similar, tem-se que:

$$\cos(nx) = \frac{e^{inx} + e^{-inx}}{2}. \quad (17)$$

A Eq. 10 pode ser reescrita como,

$$f(x) = a_0 + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]. \quad (18)$$

Inserindo as identidades 17 e 18 na expansão, temos:

$$\begin{aligned} f(x) &= a_0 + \sum_{n=1}^{\infty} \left[a_n \left(\frac{e^{inx} + e^{-inx}}{2} \right) + b_n \left(\frac{e^{inx} - e^{-inx}}{2i} \right) \right] \\ f(x) &= \underbrace{a_0}_{=C_0} + \sum_{n=1}^{\infty} \left[\underbrace{\left(\frac{a_n}{2} + \frac{b_n}{2i} \right)}_{=C_n} e^{inx} + \underbrace{\left(\frac{a_n}{2} - \frac{b_n}{2i} \right)}_{=C_{-n}} e^{-inx} \right] \end{aligned}$$

em que C_0 , C_n e C_{-n} são constantes complexas.

Note que o número inteiro n pode assumir valores negativos por conta do sinal negativo da unidade imaginária. Desse modo, para que seja possível inserir as constantes complexas em função de uma única soma infinita em termos de n , é preciso que n varie de menos infinito a mais infinito. Isto é,

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{inx}$$

que é chamada de Expansão em Série de Fourier na forma complexa, na qual:

$$\begin{aligned} C_0 &= a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx \\ C_n &= \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx \end{aligned}$$

O coeficiente C_0 é definido de modo similar ao feito para o coeficiente a_0 da Eq. 10 como mostrado no Apêndice A. A demonstração da constante complexa C_n encontra-se no Apêndice B.

2.2.3 Séries de Fourier para funções reais de período fundamental T

As relações 8 e 9 representam a periodicidade das funções seno e cosseno. Uma vez que ω_0 representa a frequência angular fundamental, então para funções que possuem período

fundamental T , temos:

$$\sin\left(\frac{2\pi}{T}x\right) = \sin\left(\frac{2\pi}{T}x + P\right) \quad (19)$$

$$\cos\left(\frac{2\pi}{T}x\right) = \cos\left(\frac{2\pi}{T}x + P\right) \quad (20)$$

assim, a Eq. 18 pode ser reescrita como:

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi}{T}nx\right) + b_n \sin\left(\frac{2\pi}{T}nx\right) \right]$$

sendo:

$$a_0 = \frac{1}{T} \int_0^T f(x) dx \quad (21)$$

$$a_n = \frac{2}{T} \int_0^T f(x) \cos\left(\frac{2\pi}{T}nx\right) dx \quad (22)$$

$$b_n = \frac{2}{T} \int_0^T f(x) \sin\left(\frac{2\pi}{T}nx\right) dx \quad (23)$$

são os coeficientes da expansão (para demonstrações, ver Apêndice C).

2.2.4 Séries de Fourier para funções de variável complexa de período fundamental T

A dedução da expansão de funções com período fundamental T em Séries de Fourier, definidas em \mathbb{C} , é feita de forma semelhante a empregada para o caso de funções com período fundamental igual a 2π , mostrado na Seção 2.2.2. Por esse motivo, os detalhes da demonstração não serão apresentados. O resultado obtido é:

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{i\frac{2\pi}{T}nx}$$

em que

$$C_n = \frac{1}{T} \int_0^T f(x) e^{i\frac{2\pi}{T}nx} dx$$

Demonstração do coeficiente C_n :

É sabido que

$$C_n = \frac{a_n}{2} + \frac{b_n}{2i}$$

uma vez que

$$a_n = \frac{2}{T} \int_0^T f(x) \cos\left(\frac{2\pi}{T}nx\right) dx$$

$$b_n = \frac{2}{T} \int_0^T f(x) \sin\left(\frac{2\pi}{T}nx\right) dx$$

Então,

$$C_n = \frac{1}{T} \int_0^T f(x) \cos\left(\frac{2\pi}{T}nx\right) dx + \frac{1}{Ti} \int_0^T f(x) \sin\left(\frac{2\pi}{T}nx\right) dx$$

como

$$\frac{1}{i} = \frac{i}{i^2} = -i$$

então,

$$C_n = \frac{1}{T} \int_0^T f(x) \cos\left(\frac{2\pi}{T}nx\right) dx - \frac{i}{T} \int_0^T f(x) \sin\left(\frac{2\pi}{T}nx\right) dx$$

$$C_n = \frac{1}{T} \int_0^T f(x) \left[\cos\left(\frac{2\pi}{T}nx\right) - i \sin\left(\frac{2\pi}{T}nx\right) \right] dx$$

Portanto,

$$C_n = \frac{1}{T} \int_0^T f(x) e^{-i\frac{2\pi}{T}nx} dx$$

q.e.d.

2.2.5 Transformada de Fourier

É comum a investigação de fenômenos físicos passar por etapas de modelagem que demandam aplicações de técnicas matemáticas, estatísticas e computacionais. Deve-se buscar o tratamento que melhor se ajuste para uma descrição mais precisa das observações de tais fenômenos. Uma das ferramentas matemáticas mais poderosas e que se adaptam a uma gama considerável de situações envolvendo problemas físicos é a Transformada de Fourier.

As transformações integrais transformam uma equação diferencial em uma equação algébrica (SILVA JR., 2009). É comum fenômenos naturais serem descritos por equações diferenciais, parciais ou ordinárias. Nessa situação, uma abordagem promissora é expressar o problema em uma equação algébrica de mais fácil manipulação, resolver tal equação e relacionar o resultado com a equação diferencial que rege o problema. Essa manipulação matemática é o alvo da *transformação integral*, e que é definida como um operador linear $\mathbf{T} : f \rightarrow F$ como

$$F(k) = \mathbf{T} f(x) = \int_a^b N(k, x) f(x) dx .$$

Para essa representação temos que $N(k, x)$ é o núcleo da transformação e $F(k)$ é a transformada de $f(x)$ pela ação do operador \mathbf{T} . Geralmente, k e $x \in \mathbb{R}$, e $x \in [a, b]$.

Para o caso em que o núcleo da transformação é dado por:

$$N(k, x) \equiv \frac{e^{-ikx}}{\sqrt{2\pi}}$$

temos, de acordo com a definição supracitada:

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-ikx} f(x) dx \quad (24)$$

que é a transformada de Fourier para uma função $f(x)$.

A transformada de Fourier descreve padrões na observação de um conjunto de dados. É importante notar que ela, tal como definida na Eq. 24 atua sobre um conjunto de valores contínuos. Para uma melhor compreensão do fenômeno a serem estudados na próxima seção, faz-se necessário avaliar o comportamento discreto de certas amostras. Neste sentido, utiliza-se comumente uma variante conhecida como Transformada Discreta de Fourier (TDF). Vejamos.

Uma vez que o núcleo da transformada de Fourier depende de uma função exponencial, a menos de uma constante, pode-se assumir sem perda de generalidade, que a Eq. 24 pode ser escrita como

$$F(k) = \int_{-\infty}^{+\infty} e^{-ikx} f(x) dx \quad (25)$$

A exponencial no integrando acima uma contribuição do tipo senoidal.

Uma variante chamada TDF é definida para o caso de conjunto de dados discretos:

$$X_k = \sum_{n=0}^{N-1} X_n e^{-ikx} \quad (26)$$

em que X_m é um número complexo cujo módulo está associado ao número inteiro n que representa os valores discretos atribuídos a somatória. Essa relação pode ser percebida considerando que

$$\begin{aligned} e^{-ikx} &\rightarrow e^{-ib_n} \\ X_k &\rightarrow X_m \end{aligned}$$

então,

$$X_m = \sum_{n=0}^{N-1} X_n e^{-ib_n} = X_0 e^{-ib_0} + X_1 e^{-ib_1} + \dots + X_n e^{-ib_{N-1}}.$$

Usando a fórmula de Euler, $e^{-ix} = \cos x - i \sin x$, tem-se que,

$$X_m = X_0[\cos(b_0) - i \sin(b_0)] + X_1[\cos(b_1) - i \sin(b_1)] + \dots$$

que é um número complexo e como tal pode ser expresso por $X_m = A_m + iB_m$, sendo $\rho = \sqrt{A_m^2 + B_m^2}$ o módulo de X_m .

Existe um número complexo correspondente a cada valor inteiro avaliado discretamente. Assumindo o caso de análise por evolução temporal, a descrição gráfica pode ser feita em função de θ (Eq. 14) e do módulo do número complexo como está apresentado na Figura 2. O ângulo θ corresponde nesta representação ao período de uma senóide.

É possível aplicar a TDF para tratar a análise de variabilidade temporal de um conjunto de amostras observadas. Valores discretos são os registros mais comuns em Astrofísica e em geral é através deles que se faz, por exemplo, estudos de variabilidade temporal do brilho em sistemas astrofísicos.

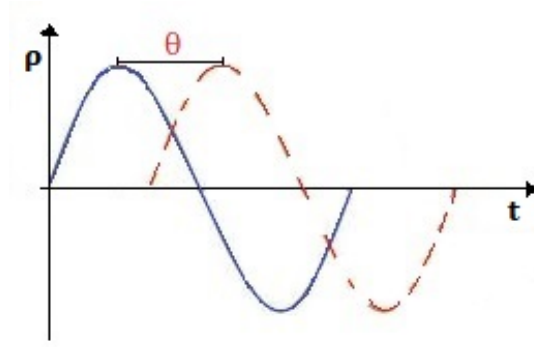


Figura 2: Módulo de X_m versus tempo de observação.

3 Periodograma

Periodograma é uma ferramenta matemático-computacional usada para identificar e caracterizar sinais periódicos em um conjunto de dados (VANDERPIAS, 2018). Neste trabalho é feita uma visita e detalhamento do desenvolvimento matemático do periodograma Lomb-Scargle (SCARGLE, 1982). Com base em técnicas de Fourier, esse é um periodograma largamente usado em Astrofísica. Vale ressaltar que este documento tem como referência primária o texto elaborado por (SCARGLE, 1982), reapresentando suas análises e resultados, inclusive figuras, e de outras referências bibliográficas conforme indicação ao longo do texto.

Matematicamente, a função que representa o periodograma é comumente definida por:

$$P_x(\omega) = \frac{1}{N_0} |FT_x(\omega)|^2 \quad (27)$$

na qual $FT_x(\omega)$ é a TDF definida para um conjunto arbitrário de dados $X(t_j)$, dada por:

$$FT_x(\omega) = \sum_{j=1}^{N_0} X(t_j) e^{-i\omega t_j} \quad (28)$$

assumindo X como uma variável física medida num intervalo de tempo t_j e expressa por:

$$X(t_j) ; j = 1, 2, \dots, N_0. \quad (29)$$

3.1 O caso Lomb-Scargle

Para investigar um conjunto de dados é preciso considerar que existem casos em que se tem cobertura temporal regular, de modo contínuo, e casos nos quais a cobertura é irregular. A análise de séries temporais, portanto, requer um tratamento específico para diferentes tipos de observações. Como consequência, existem técnicas utilizadas no intuito de determinar se há periodicidade associada aos dados investigados.

Jefrey D. Scargle, em 1982, propôs uma modificação na definição clássica do periodograma a fim de obter um comportamento estatístico simples. Sua intenção foi estender o tratamento empregado para a análise espectral de dados uniformemente espaçados aos casos nos quais os sinais não são espaçados igualmente. A ideia não foi propor uma nova técnica de análise, mas avaliar estatisticamente o acordo entre a análise e a realidade do resultado após aplicação do periodograma ao conjunto de dados. Neste sentido, (SCARGLE, 1982) adaptou a definição da referida técnica para o caso de acúmulo de dados sem interrupção temporal irregular. Tais modificações mostraram que (i) o comportamento estatístico do periodograma para espaçamentos desiguais é essencialmente idêntico para o caso em que se observam amostras igualmente espaçadas; (ii) a análise do periodograma é exatamente equivalente ao ajuste de dados de mínimos quadrados para ondas senoidais; (iii) a translação temporária é invariante (SCARGLE, 1982).

A soma de um sinal e erros observacionais aleatórios pode ser definida como:

$$X_i = X(t_i) = X_s(t_i) + R(t_i) \quad (30)$$

em que

$X_s(t_i)$ está relacionado com a detecção de sinais estritamente periódicos; $R(t_i)$ é referente ao ruído contido na medição dos sinais. O ruído é portanto, uma contribuição adicional indesejada de uma fonte que esteja sendo observada.

Para se obter um tratamento estatístico mais simples, foi assumido que o ruído $R(t_i)$ é normalmente distribuído com média zero e variância constante (σ_0^2). Tanto o ruído quanto a variância estão relacionados ao mesmo conjunto de dados com seus elementos expressos em função do tempo, $[X(t_i) \text{ sendo } i = 1, 2, \dots, N_0]$.

Vale ressaltar que, para cada tempo de observação se tem uma contribuição diferente do ruído, assim pode-se assumir que $R(t_i)$ é estatisticamente independente de $R(t_j)$. Em uma dada observação, o que se deseja é verificar se existe periodicidade no sinal detectado, considerando a contribuição do ruído existente no conjunto de dados.

De acordo com as equações 27 e 28, temos:

$$P_x(\omega) = \frac{1}{N_0} |FT_x(\omega)|^2$$

$$P_x(\omega) = \frac{1}{N_0} \left| \sum_{j=1}^{N_0} X(t_j) e^{-i\omega t_j} \right|^2. \quad (31)$$

Note que a função periodograma possui uma dependência com o módulo quadrado da TDF. Esta, por sua vez nos dá uma função que possui partes real e imaginária. Assim, sabendo que a definição matemática do módulo quadrado de uma função nos \mathbb{C} é:

$$|\psi(x)|^2 = \psi(x)^* \psi(x)$$

Segue que

$$|FT_x(\omega)|^2 = FT_x(\omega)^* FT_x(\omega)$$

A partir da relação de Euler para números complexos apresentada na seção anterior, temos que:

$$e^{-i\omega t} = \cos(\omega t) - i \sin(\omega t)$$

Então,

$$P_x(\omega) = \frac{1}{N_0} \left| \sum_{j=1}^{N_0} X(t_j) e^{-i\omega t_j} \right|^2.$$

Temos então que efetuar o produto da TDF com seu complexo conjugado, que é obtido multiplicando o sinal da parte imaginária por -1 . Assim,

$$\begin{aligned} P_x(\omega) &= \frac{1}{N_0} \left[\left(\sum_{j=1}^{N_0} X(t_j) e^{i\omega t_j} \right) \left(\sum_{j=1}^{N_0} X(t_j) e^{-i\omega t_j} \right) \right] \\ P_x(\omega) &= \frac{1}{N_0} \left[\left(\sum_{j=1}^{N_0} X(t_j) (\cos \omega t_j + i \sin \omega t_j) \right) \left(\sum_{j=1}^{N_0} X(t_j) (\cos \omega t_j - i \sin \omega t_j) \right) \right] \end{aligned}$$

Sabendo que,

$$(a + b)(a - b) = a^2 - b^2$$

vem,

$$\begin{aligned} P_x(\omega) &= \frac{1}{N_0} \left[\left(\underbrace{\sum_{j=1}^{N_0} X(t_j) \cos \omega t_j}_{=a} + \underbrace{\sum_{j=1}^{N_0} X(t_j) i \sin \omega t_j}_{=b} \right) \left(\underbrace{\sum_{j=1}^{N_0} X(t_j) \cos \omega t_j}_{=a} - \underbrace{\sum_{j=1}^{N_0} X(t_j) i \sin \omega t_j}_{=b} \right) \right] \\ P_x(\omega) &= \frac{1}{N_0} \left[\left(\sum_{j=1}^{N_0} X(t_j) \cos \omega t_j \right)^2 - \left(\sum_{j=1}^{N_0} X(t_j) i \sin \omega t_j \right)^2 \right]. \end{aligned}$$

Uma vez que $X_i = X(t_i)$, temos:

$$P_x(\omega) = \frac{1}{N_0} \left[\left(\sum_j X_j \cos \omega t_j \right)^2 + \left(\sum_j X_j \sin \omega t_j \right)^2 \right]. \quad (32)$$

Essa equação será chamada de *periodograma clássico* ao longo deste trabalho.

A Eq. 32 é válida para qualquer valor de frequência (ω). Dessa equação, percebe-se que se X contém um componente senoidal de frequência ω_0 , então para valores de frequências próximo ω_0 , os fatores $X(t)$ e $\exp(-i\omega t)$ estão em fase, contribuindo significativamente para que sejam alcançados valores maiores no somatório. Para outros valores de ω os termos da somatória se alternam positiva e negativamente, com isso alguns fatores se cancelam, o que resulta em menores valores alcançados no somatório (SCARGLE, 1982). Quando as observações ao longo de um intervalo de tempo Δt forem igualmente espaçadas,

consideramos $\Delta t = 1$; $t_j = j$; $X_j = X(t_j)$ (SCARGLE, 1982), fazendo com que a função clássica do periodograma seja escrita como:

$$P_x(\omega) = \frac{1}{N_0} \left| \sum_{j=1}^{N_0} X_j e^{-ij\omega} \right|^2 \quad (33)$$

que é comumente avaliada para um conjunto de frequências naturais dado por:

$$\omega_n = \frac{2\pi n}{T} ; n = -\frac{N_0}{2}, \dots, \frac{N_0}{2} \quad (34)$$

sendo T o intervalo de tempo total necessário para se detectar padrões de periodicidade nos sinais observados. O uso desse conjunto de frequências naturais nos permite escrever uma forma mais explícita para a função clássica do periodograma. Isto é,

$$\begin{aligned} P_x(\omega_n) &= \frac{1}{N_0} |FT_x(\omega_n)|^2 \\ P_x(\omega_n) &= \frac{1}{N_0} \left| \sum_{n=1}^{N_0/2} X_n e^{-i\omega_n} \right|^2 \end{aligned} \quad (35)$$

Vale ressaltar que as variáveis (frequência, período, amplitude) dadas por $P_x(\omega_n)$ são independentes umas das outras, enquanto que as variáveis para $P_x(\omega)$, em frequências intermediárias, são dependentes.

Ainda considerando o conjunto de frequências anteriormente apresentado, para $n = 1$, se tem a mais baixa frequência na qual há um limite mínimo para cobertura temporal para que seja possível detectar uma variabilidade de período T .

Tomando, da relação apresentada na Eq. 34, o valor de uma frequência como sendo

$$N = \frac{N_0}{2} \quad (36)$$

então, percebemos que:

$$\omega_N = \frac{2\pi N}{T} = \frac{\pi N_0}{T} \quad (37)$$

que é chamada *frequência de Nyquist*, a maior frequência sobre a qual há um limite mínimo para cobertura temporal para que seja possível detectar uma variabilidade de período T . Neste caso, está sendo considerado o curto intervalo de amostragem $\Delta t = T/N_0$.

A frequência fundamental para os dados não uniformemente espaçados é basicamente inalterada, desde que seja considerado o intervalo T . Porém, o significado dessa frequência é mais sutil. Uma vez que a maior frequência na qual existe um limite mínimo que possibilite a detecção da variabilidade em um determinado período é $\pi/\Delta T$, qual seria o valor médio do intervalo de tempo mais apropriado que pode ser usado para generalizar essa definição? Seria algébrico, geométrico, harmônico? Nessa situação, a melhor maneira de escolher o conjunto de frequências é considerando a função de resposta espectral, algumas vezes chamada de *janela espectral*.

O periodograma é uma função da fase de ondas senoidais. A janela espectral, por sua vez, descreve a resposta da análise de todo conjunto de dados disponível para uma única frequência, o que também justifica a escolha do tratamento matemático através da TDF. Lomb (1976) derivou expressões para o periodograma aplicado a um sinal senoidal. A expressão exata não satisfazia as condições iniciais que permitiriam uma resposta significativa para análise dos dados (um dos motivos pelos quais é necessário propor uma modificação na função periodograma clássica, que será abordado mais adiante). Mas quando avaliada sobre fase e simplificada com as devidas aproximações, o periodograma clássico mostrado na Eq. 32 se reduz a:

$$P_s(\omega) = [W(\omega - \omega_s) + W(\omega + \omega_s)]^2 \quad (38)$$

em que P_s é o periodograma devido a onda senoidal de frequência ω_s e W é a TDF da janela de observação do domínio temporal, dada por:

$$W(\omega) = \frac{1}{N_0} \left[\sum_{j=1}^{N_0} e^{i\omega t} \right] \quad (39)$$

Deeming (1975, eq.[8]) mostrou que o valor esperado para o periodograma clássico é igual a convolução da potência do espectro real com a função de *janela espectral* para a amostragem particular usada. Isto é,

$$G(\omega) = |W(\omega)|^2 \quad (40)$$

Em outras palavras, a *função de janela* é uma parte da função de Green.

Da Eq. 38, temos:

$$P_s(\omega) = [W(\omega - \omega_s)]^2 + 2W(\omega - \omega_s)W(\omega + \omega_s) + [W(\omega + \omega_s)]^2 \quad (41)$$

em que ω_s é a frequência deslocada, ou seja, que possui um valor maior ou menor em relação ao sinal original. Assim, o primeiro e o terceiro termos representam uma função que tem sua origem deslocada por ω_s e $-\omega_s$, respectivamente.

Quando, utilizando 38, se detecta uma variabilidade considerável no brilho do objeto observado e graficamente é notado um pico extremamente estreito no periodograma proveniente da potência do objeto astrofísico, sendo $\omega = 0$ e ω_s não tão pequeno, então para $\omega > 0$ o segundo e terceiro termos da Eq. 41 são desprezíveis, e a resposta básica é $G(\omega - \omega_s)$. Consequentemente G é chamada de *periodogram window*, ou *janela espectral*.

a) Limitações ao uso do Periodograma Clássico

O problema estatístico mais relevante no uso do periodograma clássico é que a resposta da função $P(\omega)$ gera uma detecção considerável do ruído na medição dos sinais, e estes sinais não diminuem em amplitude com o aumento da cobertura temporal. Quando X está associado a uma distribuição normal, nos casos em que a contribuição do ruído é significativa, a variância relativa $\sigma_p / \langle P \rangle$ tem valor constante (Bartlett 1950 e referências a ele; Richards 1967), não importando o quão longa seja a cobertura temporal da observação.

No resultado de um periodograma estão presentes contribuições associadas a ruídos no conjunto de dados. A respeito da Eq. 33, Richards (1967) considera que “*exceto nos casos de pouco interesse, o a função periodograma clássico não tem aplicação para cálculos práticos... Repetidas experiências mostraram que, com um sinal de ruído ... 33 gera um espectro muito diferente do esperado para $P(\omega)$, independentemente de quão extenso T for tomado ou quão pequeno Δt for escolhido.*” Tukey (1967) faz considerações similares acerca dos espectros, “*se tratarmos de problemas envolvendo a superposição de alguns fenômenos periódicos simples, como fazem os astrônomos interessados em estrelas binárias ou problemas similares, podemos aprender muito a partir do periodograma.*”

Faz-se necessário buscar uma relação entre o sinal e o ruído. Para isso é proposta uma modificação para a variável anteriormente definida na Eq. 29, que passa a ser:

$$X(t_i) = X_0 \sin(\omega_0 t_i + \phi) + R(t_i); \quad i = 1, 2, \dots, N_0. \quad (42)$$

Verifica-se a existência de um valor de potência que está associado ao valor da frequência do sinal periódico (P_X), como também um valor de potência que está associado a contribuição do ruído (P_R).

Partindo da definição clássica do periodograma mostrada na Eq. 32, podemos encontrar uma representação para o valor esperado de P_X .

$$\begin{aligned} P_x(\omega)^* P_x(\omega) &= \frac{1}{N_0} \left[\sum_{j=1}^{N_0} X(t_j) [\cos \omega t_j - i \sin \omega t_j] \right] \frac{1}{N_0} \left[\sum_{j=1}^{N_0} X(t_j) [\cos \omega t_j + i \sin \omega t_j] \right] \\ P_X &= \frac{1}{N_0^2} \left[(X^2(t_j) \cos^2 \omega t_j)^2 + (X^2(t_j) \sin^2 \omega t_j)^2 \right] \\ P_X &= \frac{1}{N_0^2} \left[X^2(t_j) \underbrace{[\cos^2 \omega t_j + \sin^2 \omega t_j]}_{=1} \right] \\ P_X &= \frac{X^2(t_j)}{N_0^2}. \end{aligned}$$

Já que $X(t_j) = X_0$ e $N_0 = 2N$, temos:

$$\begin{aligned} P_X &= \frac{X_0^2}{4N^2} \\ P_X &= \frac{1}{N^2} \left(\frac{X_0}{2} \right)^2 \end{aligned} \quad (43)$$

Analogamente, o valor esperado para a P_R é:

$$P_R = R^* R = \langle R^2 \rangle$$

Uma vez que segue uma distribuição cuja média é nula, o valor de P_X pode ser representado por sua variância:

$$P_R = \sigma_0^2 \quad (44)$$

Desse modo, a expressão desejada, isto é, a relação sinal-ruído pode ser expressa como o quociente entre as Equações 43 e 44:

$$P = \frac{P_X}{P_R} = \frac{1}{N^2} \left(\frac{X_0}{2} \right)^2 \frac{1}{\sigma_0^2} \quad (45)$$

A existência de uma frequência ω_0 associada a uma periodicidade num conjunto de dados implica na existência de outras frequências associadas a ela que podem eventualmente serem detectadas. A potência do periodograma pode ser medida, conseqüentemente, para frequências diferentes de ω_0 . Esse é um problema a ser contornado, chamado de *spectral leakage* proveniente do uso da função periodograma clássica. Pode-se calcular, por exemplo, a potência no periodograma quando a cobertura temporal sobre a qual os dados são amostrados é finita e quando o tamanho finito do intervalo de tempo entre duas amostras for avaliado.

Em ambos os casos tem-se um fenômeno bem conhecido, denominado *aliasing*. A forma no qual ele é estudado torna-o sensível a uma uniformidade na amostragem. Ou seja, qualquer variação no espaçamento dos dados influencia a detecção do fenômeno. Alguns trabalhos na literatura sustentam essa afirmação (e.g., Beutler 1966, 1970; Marsy and Lui 1975, 1977; Kar, Hornkohl, and Farmer 1981; Ludeman 1981). Há trabalhos nos quais os efeitos no fenômeno *aliasing* de duas perturbações diferentes a partir da mesma amostragem são discutidos. Por exemplo, Beutler (1970), Sturrock (1981) e Shiub (1981) discutem os efeitos em exclusões de amostras aleatórias e independentes, e amostragens irregulares, nas quais as coberturas temporais são aleatoriamente perturbadas sobre valores igualmente espaçados.

Para o caso em que se deseja recuperar o sinal numa faixa limitada, a recuperação pode ser alcançada com amostragens irregulares desde que a taxa de amostragem média supere a taxa Nyquist (i.e., a média do número de amostras por unidade de tempo tem de ser maior que o sinal que contém a maior frequência). No entanto, Beutler (1966) apresenta um exemplo de que a recuperação da amostragem é possível mesmo se a taxa de amostragem média for menor do que a taxa Nyquist. Outro caso de interesse é que o espectro a ser recuperado pode ser estimado mesmo se a amostragem não for recuperada - apenas a ordem das amostras precisa ser retida (Beutler 1970). Neste caso, o fato de o intervalo de tempo ser hipoteticamente infinito impossibilita a recuperação da amostragem temporal. Pelo fato da amostragem ser avaliada em um intervalo de tempo infinito, não se pode obter resultados práticos a partir do espectro estimado.

O fenômeno (*aliasing*) pode ser contornado em algumas situações nas quais as amostragens são tipicamente irregulares e que são usuais em observações astronômicas. Contudo uma regularidade na sua periodicidade pode ser forçada, distanciando os valores de frequências próximas a ω_0 , fazendo com que haja uma similaridade com o *aliasing* (ver Figura 3, reproduzida de (SCARGLE, 1982)). Alguns problemas podem surgir na análise da observação, visto que mesmo quando tal periodicidade (com uma regularidade forçada) não está presente, a amostragem tende a ser semirregular, isto é, entre duas amostras uniforme e aleatoriamente espaçadas (SCARGLE, 1982).

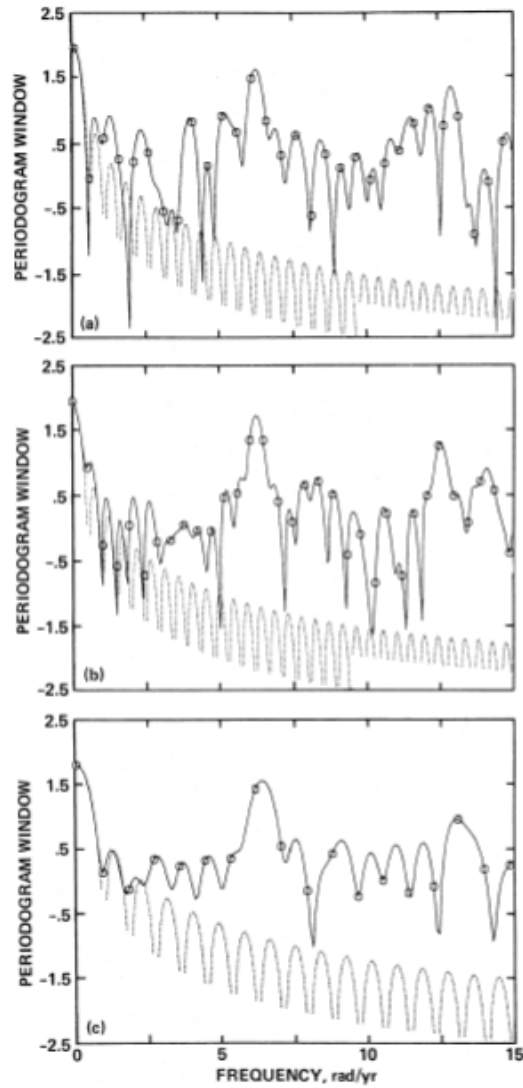


Figura 3: Logaritmo do periodograma clássico obtido a partir dos dados astronômicos das estrelas: (a) G96-45, (b) G146-72 e (c) Wolf-1062. Em cada caso, a linha sólida é a janela calculada a partir das fórmulas clássicas (Equações 38 e 40) enquanto a linha pontilhada mostra por comparação a função de janela para espaçamentos iguais com o mesmo intervalo de tempo total e o mesmo número de pontos. Os símbolos octagonais abertos são a função de janela avaliada no conjunto de frequências definidas na Eq. 36, enquanto a curva sólida é exagerada pelo fator 10 relativo a esse conjunto. (Figura reproduzida de (SCARGLE, 1982)).

Em ambos os problemas (ruído na resposta da função $P(\omega)$ e *spectral leakage*), a crítica mais bem estabelecida é que o uso das técnicas apresentadas é equivalente a estimar uma tendência nos parâmetros do domínio espectral (IONIDES, 2016). Um exemplo, chamado de *windowing* ou *tapering*, é a multiplicação dos dados pela função que se aproxima de zero ao final do intervalo da amostragem. Outro exemplo é o tratamento análogo da função de autocorrelação. Por meios elementares pode-se mostrar que ambos os tratamentos empregados para os problemas possuem correspondência com a convolução do espectro e a função *janela espectral* (Harris 1978). Tal convolução reduz a variância da amostragem, que está relacionada com a potência, $\sigma_p / \langle P \rangle$, sendo ela uma medida do espectro de potência do periodograma. Analogamente, o *spectral leakage* pode ser controlado, porque a *função de janela* pode ser escolhida, tal que as amplitudes das frequências próximas a ω_0 são reduzidas. Vários tipos de *funções de janela* podem ser propostas, testadas e usadas. Harris (1978) apresenta gráficos de superposições de frequências próximas a ω_0 para quarenta e cinco tipos de funções. Assim, o que apresenta ser uma medida diferente é na realidade equivalente a função de autocorrelação (Richards 1967).

É interessante notar que todas essas técnicas podem ser aplicadas ao periodograma para ambos os tipos de amostras, isto é, para dados uniforme e não uniformemente espaçados (ver Thompson 1971). Uma desvantagem em qualquer redução se dá no fato de que os valores do espectro para potências diferentes de ω_0 não mais são independentes, de modo que as propriedades estatísticas necessitam de um tratamento mais sutil para que se possa obter as curvas de luz desejadas.

Uma abordagem diferente ao problema *leakage* é tentar removê-lo a partir do espectro. Várias técnicas foram sugeridas para esse fim (e.g., Wehlau and Leung 1964; Fitch and Wehlau 1965; Gray and Desikachary 1973; Meisel 1978; Swan 1981). Uma vez que o ruído observacional não pode ser evitado, o objetivo dessas técnicas de deconvolução é, portanto, diminuir sua amplificação. No entanto, a maioria delas falha nessa tentativa.

Por fim, os problemas estatístico e *leakage*, provém da impossibilidade de estimar os parâmetros no domínio espectral. Além disso, o tratamento estatístico não é tão severo como habitualmente se propunha, para os casos em que o sinal é estritamente periódico. O periodograma, mais precisamente a função periodograma clássico (Eq. 32) para amostras desigualmente espaçadas não é necessariamente a melhor ferramenta para tratar desse problema em especial. Porém a simplicidade do comportamento estatístico do periodograma torna-o útil quando se deseja avaliar a confiabilidade da detecção de um sinal que apresenta periodicidade.

b) Uma Nova Definição do Periodograma

Um resultado de interesse na análise de séries temporárias se obtém quando X é um ruído puramente avaliado numa distribuição normal, o que torna a potência P_X distribuída de forma exponencial. No caso de amostragens que estão espaçadas de maneira uniforme, o tratamento estatístico utilizado no periodograma é simples e comumente utilizado (e.g. Groth 1975; eq. [13] para $n = 1$ apresentada nesse mesmo artigo). A equação para tratar os casos nos quais as amostras são igualmente espaçadas tem um tratamento estatístico mais simples em comparação ao caso em que as amostras são espaçadas de maneira não uniforme (Eq. 32). O que se deseja com uma nova definição do periodograma é obter uma distribuição exponencial equivalente em ambos os casos.

Seja a seguinte TDF:

$$FT_x(\omega) = \sqrt{\frac{N_0}{2}} \sum_{j=1}^{N_0} X(t_j) [A \cos \omega t_j + i B \sin \omega t_j]. \quad (46)$$

Neste caso, A e B são funções desconhecidas de ω que podem depender da amostragem t_j . O periodograma correspondente é:

$$\begin{aligned} P_x(\omega) &= \frac{1}{N_0} |FT_x(\omega)|^2 \\ P_x(\omega) &= \frac{1}{N_0} [FT_x(\omega)^* FT_x(\omega)] \\ P_x(\omega) &= \frac{1}{N_0} \left[\left(\sqrt{\frac{N_0}{2}} \sum_j A X_j \cos \omega t_j \right)^2 + \left(\sqrt{\frac{N_0}{2}} \sum_j B X_j \sin \omega t_j \right)^2 \right] \\ P_x(\omega) &= \left(\frac{A^2}{2} \right) \left[\sum_j X_j \cos \omega t_j \right]^2 + \left(\frac{B^2}{2} \right) \left[\sum_j X_j \sin \omega t_j \right]^2 \end{aligned} \quad (47)$$

Sendo que, se:

$$A = B = \sqrt{\frac{2}{N_0}}$$

então, as Equações 46 e 47 se reduzem as fórmulas clássicas do periodograma.

A justificativa básica por trás dessas definições é que para amostragens iguais FT_X corresponde à TDF (e no limite $\Delta t \rightarrow 0, N \rightarrow \infty$, é proporcional a Transformada de Fourier); o mesmo raciocínio pode ser empregado para P_X , a potência do espectro. Mas essa redução não é única: existem outras escolhas que devem ser impostas para determinar A e B . Em particular, a distribuição estatística da nova definição do periodograma será feita de forma análoga ao caso no qual as amostras são igualmente espaçadas. Isso pode ser alcançado com simples escolhas de A e B .

Seja o caso em que X é dado apenas pela contribuição do ruído, tal que esteja seguindo uma distribuição normal¹ com média zero e variância constante σ_0^2 . Nessa condição é possível expressar o ruído (X) e em função uma combinação linear de variáveis normais aleatórias e independentes entre si, da seguinte forma:

$$C(\omega) = A \sum_j X(t_j) \cos \omega t_j \quad (48)$$

No entanto uma combinação linear de variáveis que segue uma distribuição normal também é normal (Parzen 1962, §3.4, Theorem 4A, p. 90). A média de C é nula ($\langle C \rangle = 0$) e sua a

¹Uma distribuição normal, ou gaussiana, é uma função de densidade de probabilidade utilizada em estatística para descrever fenômenos naturais representados por variáveis aleatórias. O gráfico (histograma) proveniente dos valores médios da distribuição tem como característica principal que o valor da área sob a curva referente à essa função é constante e igual a 1. (N.E.)

variância é dada por:

$$\begin{aligned}\sigma_c^2 &= \langle C^2(\omega) \rangle \\ \sigma_c^2 &= A^2 \sum_j \sum_k \langle X(t_j) X(t_k) \rangle \cos \omega t_j \cos \omega t_k\end{aligned}\quad (49)$$

Como os termos cruzados ($i \neq j$) são eliminados devido a independência entre as variáveis anteriormente assumida, temos:

$$\begin{aligned}\sigma_c^2 &= A^2 \sum_j \langle R R \rangle \cos^2 \omega t_j \\ \sigma_c^2 &= A^2 \sum_j R^2 \cos^2 \omega t_j\end{aligned}$$

Uma vez que a variância do ruído R é σ_0^2 , temos:

$$\sigma_c^2 = A^2 \sigma_0^2 \sum_j \cos^2 \omega t_j \quad (50)$$

Analogamente, podemos definir outra combinação linear valendo-se das condições anteriormente citadas, a fim de relacionar essa nova expressão com a distribuição 47. Considerando

$$S(\omega) = B \sum_j X(t_j) \cos \omega t_j \quad (51)$$

podemos, com a mesma argumentação que resultou na Eq. 48, expressar:

$$\sigma_s^2 = B^2 \sigma_0^2 \sum_j \sin^2 \omega t_j \quad (52)$$

Consequentemente, a Eq. 47 pode ser escrita como:

$$\begin{aligned}P_x(\omega) &= \left(\frac{A^2}{2}\right) \left[\sum_j X_j \cos \omega t_j\right]^2 + \left(\frac{B^2}{2}\right) \left[\sum_j X_j \sin \omega t_j\right]^2 \\ P_x(\omega) &= \left(\frac{1}{2}\right) [C^2(\omega) + S^2(\omega)]\end{aligned}\quad (53)$$

Isto significa que a função periodograma é uma soma dos quadrados de duas distribuições normais, de média zero para variáveis aleatórias. Essa soma tem uma distribuição exponencial de probabilidade, mas *apenas se as variâncias das duas variáveis normais são iguais*. Esse resultado é bem conhecido (e.g., Papoulis 1965, § 7.1, example 7-7, pp.194-195). Designando X e Y como as variáveis aleatórias, de variância σ_1 e σ_2 , respectivamente, podemos aplicar o método de Papoulis (1965). Nesse método, essas variáveis estão relacionadas como $Z = Z^2 + Y^2$, tal que Z está distribuído de acordo com:

$$P_Z(z) = \frac{\exp(-z/2\sigma_2^2)}{2\sigma_1\sigma_2} G\left[z/4 \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\right] \quad (54)$$

sendo $G(x) = \exp(-x) I_0(x)$.

A função de Bessel modificada de primeira espécie e ordem m pode ser expressa por:

$$I_m(x) = \sum_{k=0}^{\infty} \frac{1}{k!(k+m)!} \left(\frac{x}{2}\right)^{2k+m} \quad (55)$$

Para $m = 0$, temos $I_0 = 1$. Assim, o termo que multiplica a exponencial anteriormente apresentada é, portanto, a função de Bessel modificada de ordem zero.

Sendo

$$x = (z/4) \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right)$$

então, para $\sigma_1 = \sigma_2 = \sigma$, temos:

$$\begin{aligned} G(x) &= \exp\left((-z/4) \left(\frac{1}{\sigma^2} + \frac{1}{\sigma^2}\right)\right) \underbrace{I_0}_{=1} \\ G(x) &= \exp(-z/2\sigma^2) \end{aligned}$$

Assim, a Eq. 54 passa a ser:

$$P_Z(z) = \frac{\exp(-z/2\sigma^2)}{2\sigma^2} \exp(-z/2\sigma^2) \quad (56)$$

que é a equação conhecida para a soma dos quadrados de duas variáveis normais de mesma variância na forma de distribuição exponencial.

O resultado para o caso em que $\sigma_1 \neq \sigma_2$ é diferente do apresentado na Eq. 54 para a condição de $\sigma_1 = \sigma_2$. Neste caso, ao invés de forçamos a igualdade das variâncias, vamos reescrever as Equações 48 e 51, escolhendo funções para A e B de tal maneira que possamos obter $\sigma_c = \sigma_s = \sigma_0$. Tais escolhas foram

$$A(\omega) = Q(\omega) \left(\sum_j \cos^2 \omega t_j \right)^{-1/2} \quad (57)$$

e

$$B(\omega) = Q(\omega) \left(\sum_j \sin^2 \omega t_j \right)^{-1/2}, \quad (58)$$

em que $Q(\omega)$ é uma função arbitrária de ω . Substituindo 57 e 58 em 47, temos que:

$$\begin{aligned} P_x(\omega) &= \left(\frac{1}{2}\right) \left[\sum_j A X_j \cos \omega t_j \right]^2 + \left(\frac{1}{2}\right) \left[\sum_j B X_j \sin \omega t_j \right]^2 \\ P_x(\omega) &= \left(\frac{1}{2}\right) \left[\left(\frac{\sum_j X_j \cos \omega t_j}{Q(\omega) \left(\sum_j \cos^2 \omega t_j \right)^{1/2}} \right)^2 + \left(\frac{\sum_j X_j \sin \omega t_j}{Q(\omega) \left(\sum_j \sin^2 \omega t_j \right)^{1/2}} \right)^2 \right] \end{aligned}$$

Para que a função periodograma (P) tenha o mesmo valor médio como no caso igualmente espaçado, isto é, quando se tem amostras avaliadas numa cobertura temporal regular de

modo contínuo, é preciso impor a condição de que $Q(\omega) = 1$. Assim,

$$P_x(\omega) = \left(\frac{1}{2}\right) \left[\frac{\left(\sum_j X_j \cos \omega t_j\right)^2}{\sum_j \cos^2 \omega t_j} + \frac{\left(\sum_j X_j \sin \omega t_j\right)^2}{\sum_j \sin^2 \omega t_j} \right] \quad (59)$$

e que resulta na distribuição de probabilidade esperada.

Se $A = B = \sqrt{2/N_0}$, a Eq. 59 se torna equivalente às Equações 38 e 33. Como os valores de A e B estão relacionados com o valor da frequência, outros resultados para A e B são obtidos de acordo com a escolha de ω . Por exemplo, tem-se para o caso de amostras que não são uniformemente espaçadas que $A \neq B$ mesmo quando $\omega = \omega_n$. Todavia, pode-se considerar $A \approx B \approx (2/N_0)^{1/2}$ para boa parte dos valores escolhidos para a frequência ².

A distribuição de probabilidade descrita pela Eq. 59 é obtida também por um ajuste de mínimos quadrados diferente do apresentado por Lomb (1976) e por Kar, Hornkhol e Farmer (1981). Lomb (1976) mostra expressões das quais é possível obter a equação para a modificação do periodograma utilizando a análise de mínimos quadrados. O resultado obtido é a expressão em fase ³ na resposta da onda senoidal:

$$\begin{aligned} 2P_s(\omega) &= \frac{(\sum \cos \omega_s t_j \cos \omega t_j)}{(\sum \cos^2 \omega_s t_j) (\sum \cos^2 \omega_s t_j)} + \frac{(\sum \cos \omega_s t_j \sin \omega t_j)}{(\sum \cos^2 \omega_s t_j) (\sum \sin^2 \omega_s t_j)} \\ &+ \frac{(\sum \sin \omega_s t_j \cos \omega t_j)}{(\sum \sin^2 \omega_s t_j) (\sum \cos^2 \omega_s t_j)} + \frac{(\sum \sin \omega_s t_j \sin \omega t_j)}{(\sum \sin^2 \omega_s t_j) (\sum \sin^2 \omega_s t_j)} \end{aligned} \quad (60)$$

Aparentemente não é possível reescrever a Eq. 60 a partir de 38, então a *função de janela* sob a forma da Eq. 40 não pode ser definida.

Consequentemente, a redefinição sugerida tem a forma da Eq. 59:

$$P_x(\omega) = \left(\frac{1}{2}\right) \left[\frac{\left(\sum_j X_j \cos \omega (t_j - \tau)\right)^2}{\sum_j \cos^2 \omega (t_j - \tau)} + \frac{\left(\sum_j X_j \sin \omega (t_j - \tau)\right)^2}{\sum_j \sin^2 \omega (t_j - \tau)} \right], \quad (61)$$

em que τ é o tempo de retardo específico para cada frequência, definido como:

$$\tan(2\omega\tau) = \left(\frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j} \right) \quad (62)$$

A Eq. 61 é, portanto, uma versão levemente modificada do periodograma. As Figuras 4 e 5, reproduzidas de (SCARGLE, 1982), comparam resultados do periodograma Lomb-Scargle obtidos antes e após a modificação aqui referida.

Utiliza-se a Eq. 61 por conta de seu comportamento estatístico simples dado pela distribuição exponencial de probabilidade. Além disso, será abordado a seguir, sua equivalência com o ajuste de mínimos quadrados de ondas senoidais, tal como a possibilidade

²Como a frequência ω é o argumento das funções sin e cos, os valores escolhidos para ela têm, necessariamente, de satisfazer as condições trigonométricas vinculadas à periodicidade dessas funções. (N.E.)

³Quando pulsos ondulatórios apresentam superposição construtiva durante sua propagação, de modo que haja uma acréscimo no valor da amplitude resultante. (N.E.)

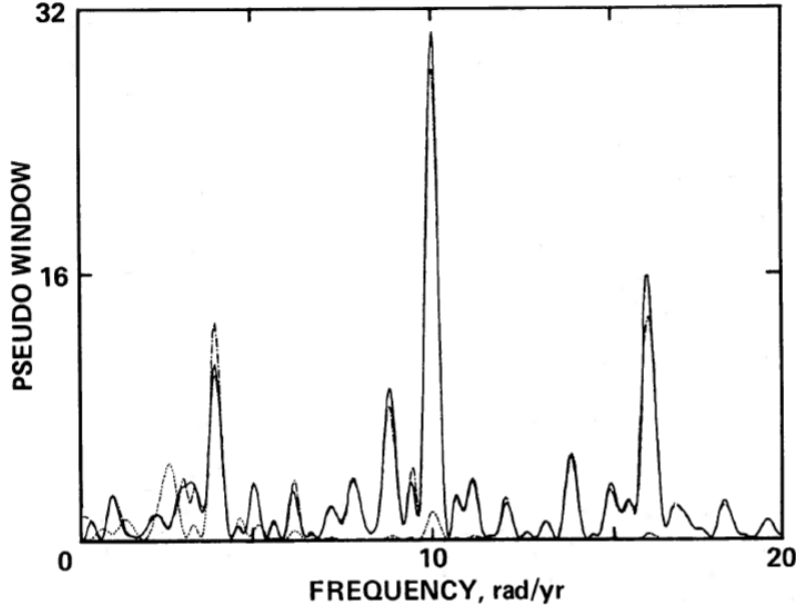


Figura 4: Comparação do periodograma clássico (*linha sólida*), obtida usando a Eq. 39, com a versão modificada (*linha pontilhada*) dada pela Eq. 61. Como foi explicado, uma vez que a resposta espectral do periodograma modificado não pode ser escrita na forma da Eq. 38, estritamente falando a *função de janela* não pode ser definida. Portanto, o que é mostrado aqui é uma *pseudowindow*, mais especificamente a resposta para altas frequências senoidais ($\omega_s = 10$), com amostragem como na Figura 3a. A linha pontilhada é o periodograma clássico calculado para os mesmos dados alternados em espaços iguais. (Figura reproduzida de (SCARGLE, 1982)).

de reduzi-la à forma clássica descrita pela equação Eq. 33 se o espaçamento for igual e se a translação temporal for invariante, ratificam a sua utilidade.

A invariância à translação temporal (*time-translation invariance*) é uma propriedade útil que se observa na definição clássica do periodograma. Nas Equações 39 e 40, percebe-se que se existe um deslocamento no tempo original, isto é $t_j \rightarrow t_j + T_0$ para todo j , então o periodograma permanece inalterado, uma vez que existe simplesmente um fator na fase $\exp(i\omega T_0)$ de módulo unitário contido no valor absoluto. Essa propriedade, todavia, não é verificada na forma modificada do periodograma descrita pela Eq. 61. Então é preciso encontrar uma maneira de restaurar a invariância, o que pode ser feito de várias formas. Uma delas é inserir o termo de retardo τ na Eq. 62 como mostramos anteriormente, definindo-o como:

$$\tau = \frac{1}{2\omega} \tan^{-1} \left[\frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j} \right] \quad (63)$$

Note que se $t_j \rightarrow t_j + T_0$, então $\tau \rightarrow \tau + T_0$ fazendo com que T_0 se anule no argumento de cada função na Eq. 61,

$$\cos \omega (t_j - \tau) = \cos \omega (t_j + T_0 - [\tau + T_0])$$

e,

$$\sin \omega (t_j - \tau) = \sin \omega (t_j + T_0 - [\tau + T_0])$$

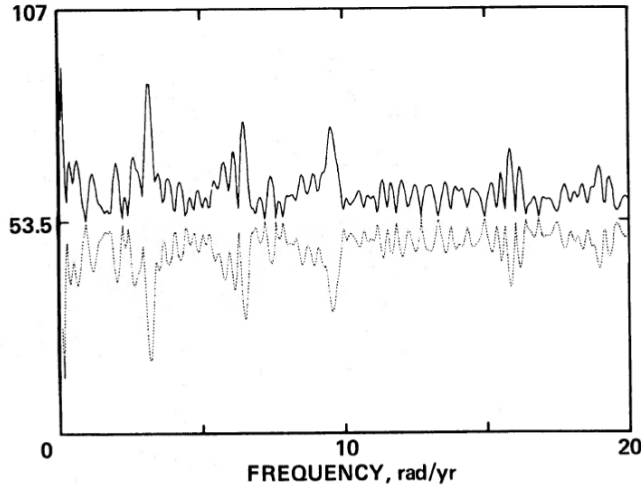


Figura 5: As expressões $\sum_j \cos^2 \omega t_j$ (acima) e $\sum_j \sin^2 \omega t_j$ (abaixo). Elas estão contidas nos denominadores da Eq. 61. O periodograma clássico, Eq. 32 é obtido se seus denominadores estão aproximados com a constante $N_0/2$, ponto médio do eixo das ordenadas. Os tempos $\{t_j\}$ são os mesmos da Figura 3a. (Figura reproduzida de (SCARGLE, 1982)).

Conseqüentemente, se obtém a invariância à translação temporal. Além disso, o comportamento estatístico simples virtualmente idêntico para amostragens iguais, dado pela distribuição exponencial, não é alterado com essa modificação. Desse modo, fica justificada a escolha da Eq. 61 como uma nova definição do periodograma.

A Eq. 61 pode ser reduzida à forma clássica, Eq. 32, se $\tau \equiv 0$. Note que a periodicidade da função arcotangente se anula em 2π , o que resulta num problema a ser contornado. Para sanar esta dificuldade é necessário impor τ continuamente como função de ω e usar resolução com frequências suficientemente altas tal que nenhuma variação na fase seja perdida. Vale ressaltar também que

$$\lim_{\omega \rightarrow 0} \tau(\omega) = \frac{1}{N_0} \sum_{j=1}^{N_0} t_j = \langle t \rangle, \quad (64)$$

o que permite contornar a indeterminação no segundo termo da Eq. 64. Uma vez que τ , A e B podem ser calculados para qualquer amostragem, seja igual ou desigualmente espaçada. A dependência de N_0 para calcular o tempo se dá da mesma maneira que no periodograma clássico.

O periodograma, tal como foi apresentado, é equivalente a análise de mínimos quadrados harmônicos. Essa relação pode ser percebida considerando algumas definições.

Seja $X_f(t)$ a variável aleatória, dada por:

$$X_f(t) = A \cos \omega t + B \sin \omega t \quad (65)$$

Para minimizar a média quadrada da diferença entre os dados analisados, temos:

$$E(\omega) = \sum_{j=1}^{N_0} [X(t_j) - X_f(t_j)]^2 \quad (66)$$

sendo $X(t_j)$ a já conhecida variável definida como a soma de sinais aleatórios na Eq. 29.

Desde que A e B sejam inseridos de maneira linear, em contraste à frequência ω , elas podem ser determinadas pelas técnicas padrões de mínimos quadrados. O valor maximizado da soma dos quadrados é:

$$\Delta E(\omega) = \sum_{j=1}^{N_0} [X(t_j)]^2 - E_{MIN}(\omega) \quad (67)$$

sendo uma função de ω que pode ser determinada gráfica ou numericamente.

Os detalhes desse procedimento podem ser encontrados em Lomb (1976), que generaliza a Eq. 65 para

$$X_f(t) = A \cos \omega(t - \tau) + B \sin \omega(t - \tau) \quad (68)$$

se, e somente se,

$$\tau = \frac{1}{2\omega} \tan^{-1} \left[\frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j} \right]$$

for solução para a equação

$$\sum_{j=1}^{N_0} \cos \omega(t - \tau) + \sin \omega(t - \tau) = 0$$

Assim, a Eq. 67 se reduz à:

$$\Delta E(\omega) = \left[\frac{\left(\sum_j X_j \cos \omega(t_j - \tau) \right)^2}{\sum_j \cos^2 \omega(t_j - \tau)} + \frac{\left(\sum_j X_j \sin \omega(t_j - \tau) \right)^2}{\sum_j \sin^2 \omega(t_j - \tau)} \right] \quad (69)$$

que é, apesar de ser derivada por procedimentos diferentes, a mesma equação proposta na definição do periodograma modificado. Percebe-se que o valor máximo no periodograma se dá quando a mesma frequência que minimiza a soma de quadrados dos resíduos para os dados da onda senoidal for alcançada. Dessa maneira, qualquer resultado teórico obtido para o periodograma pode ser empregado igualmente à análise do ajuste de mínimos quadrados, e vice-versa.

3.1.1 Estatística do periodograma

Esta seção trata da distribuição estatística de variável aleatória X e da função periodograma P_X . A ideia é mostrar de que maneira foi possível estabelecer a existência de um sinal estritamente periódico em relação a presença de ruído, através de um comportamento estatístico simples. No caso em questão, essa contribuição resulta em flutuações no espectro que descreve o comportamento de uma fonte periódica.

a) A Distribuição de $P(\omega)$

As flutuações no espectro de potência de uma fonte periódica podem ser causadas pela contribuição de uma fonte não periódica? Essa pergunta pode ser respondida usando uma

abordagem probabilística, considerando o caso no qual X é dado apenas em função da contribuição do ruído. Isso implica que P_X passa a ser uma distribuição de probabilidade de variáveis aleatórias.

A probabilidade de que é possível encontrar um valor menor ou igual a X numa função de distribuição é chamada de *função de distribuição cumulativa* (CDF, seguindo o acrônimo em inglês). O uso dessa distribuição de probabilidades é justificado por algumas razões. Por exemplo, as estimativas da CDF podem ser construídas sem agrupamento de dados ⁴ (Lomb 1976). Seja $p(z)$ a distribuição diferencial de probabilidade da CDF. Pelo fato de $p(z)$ possuir dependência na seleção arbitrária dos números de *bins*, sempre vai existir perda de informação na detecção dos sinais periódicos. Quando se deseja obter o valor máximo sobre um conjunto de frequência, pode-se utilizar uma outra característica de $p(z)$: a CDF do máximo de um conjunto de frequências de variáveis aleatórias é igual ao produto das CDFs das variáveis (SCARGLE, 1982).

A distribuição diferencial de probabilidade pode ser expressa por:

$$P_Z(z)dz = \Pr(z < Z < z + dz) = \exp(-z) dz \quad (70)$$

A variância do ruído P será a partir de agora medida em unidade de σ_0^2 . A CDF associada à distribuição diferencial é:

$$F_Z(z) = \Pr\{Z < z\} = \int_0^z P_Z(z') dz' = \int_0^z e^{-z'} dz' = 1 - \exp(-z) \quad (71)$$

A Eq. 71 apresenta uma desvantagem no uso da CDF: as subtrações entre várias distribuições ⁵ tendem a desaparecer na integração. Consequentemente todas as CDFs tendem a ser similares, as distribuições diferenciais são mais distintas, apresentando termos que contribuem durante todo o processo de integração.

A quantidade associada ao ruído, $\Pr\{Z > z\} = \exp(-z)$, fornece o valor da mais alta potência avaliada em uma frequência pré-selecionada. O aumento nesse valor da potência observada (na realidade, a potência do sinal e a relação ruído-potência) pode ser detectado devido a flutuação do ruído.

O maior valor de potência do espectro de potência, isto é, o maior pico, para o qual $Z = \max_n P(\omega_n)$ dentro do conjunto de frequências (Eq. 36). Assim, a propriedade multiplicativa anteriormente mencionada produz, para este caso:

$$\Pr\{Z > z\} = \exp(-z) \quad (72)$$

Como $F_Z(z) = 1 - \exp(-z)$, temos utilizando a Eq. 71 que

$$\begin{aligned} \Pr\{Z > z\} &= 1 - F_Z(z) \\ \Pr\{Z > z\} &= 1 - [1 - \exp(-z)]^N \end{aligned} \quad (73)$$

⁴Durante a uma detecção de sinais periódicos é preciso particionar os dados de acordo com as informações que eles carregam. Neste caso, o número de *bins* está associado ao agrupamento de dados que representa a contribuição do ruído na observação (N.E)

⁵Neste caso, é considerada apenas a distribuição de probabilidades que representa a potência, mais precisamente a o fator de subtração (sinal negativo) entre as mesmas durante o processo de integração. (N.E.)

Além de ser possível obter o valor máximo sobre um conjunto de frequências, uma outra vantagem no uso da CDF é que grande parte das quantidades secundárias de detecção, como falsa detecção e probabilidades de sinais perdidos, que são relevantes neste contexto, podem ser inferidos diretamente no gráfico da CDF (Figura 6).

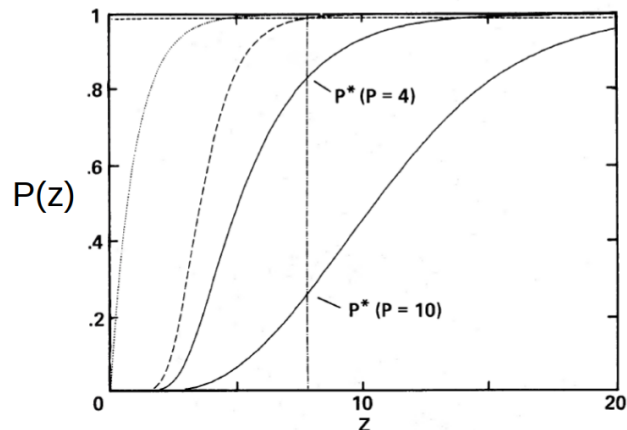


Figura 6: A linha pontilhada é a função de distribuição cumulativa CDF para a potência associada a um conjunto de frequências, no caso em que X é dado em função apenas do ruído (Eq. 71). (Figura reproduzida de (SCARGLE, 1982)).

Na Figura 6 é destacado o valor tal que a probabilidade de se detectar uma flutuação do ruído é superior a z . Tal valor é definido como p_0 e é referido como taxa de falso alarme desejada. O valor de z no qual a curva atinge $1 - p_0$ é chamado de z_0 . De modo análogo, a curva tracejada é a CDF para o máximo sobre $N = 25$ frequências, ainda que somente tratando o ruído (Eq. 73). O motivo pelo qual o valor de z , indicado na figura pela linha vertical tracejada, ser chamado de limiar de detecção é que as potências do sinal acima desse limiar são falsas apenas para uma fração de p_0 . As linhas sólidas são CDFs para a potência máxima quando os sinais estão presente, sendo a mais acima com $P = 4$ e a inferior com $P = 10$. Assim, existe um limiar de detecção z_0 tal que um sinal é associado a uma periodicidade quando sua potência p_0 superar tal valor. Trata-se de uma análise probabilística, e como tal existe uma probabilidade de não se detectar tais valores limiares que é dada pela CDF avaliada em z_0 .

Ao examinar um grande número de frequências e selecionar o maior valor de P , verifica-se que a Eq. 73 contém uma *statistical penalty*². No artigo base deste documento, (SCARGLE, 1982), é destacado que experimentos independentes, para o conjunto de frequências N foram realizados e, mesmo quando pequena probabilidade de sucesso para um só valor selecionado tenha sido observada, a chance de detecção de um deles torna-se muito alta se N for grande o suficiente (geralmente N tende ao infinito). (SCARGLE, 1982) argumenta ainda que é possível mostrar que o valor esperado como o maior no espectro de potências do ruído sobre o conjunto de frequências N é:

$$\langle Z(\max) \rangle = \sum_{k=1}^N 1/k \quad (74)$$

²Lacuna estatística (N.E.)

que é um série bem conhecida, e que diverge ao logaritmo de N .

Para o caso em que $P_s(\omega)$ é uma potência unicamente dependente do sinal periódico e sendo a relação sinal-ruído definida como $P = P_S/P_R$, a CDF para a relação sinal-ruído de Z é (Groth (1975), eq. [16] com $n = 1$):

$$F_Z(z) = 1 - \exp[-(z + p)]\phi(z, p) \quad (75)$$

em que ϕ é uma integral da função de Bessel que tem a seguinte representação em série:

$$\phi(x, y) = \sum_{m=0}^{\infty} \sum_{k=0}^m \frac{x^k y^m}{k! m!} \quad (76)$$

Se o sinal periódico está associado a uma frequência dentro do conjunto de frequências N , então a CDF de $Z = \max_n P(\omega_n)$ é:

$$F_N(z) = [1 - \exp(-z)]^{N-1} F_Z(z) \quad (77)$$

em que os valores de $P(\omega_n)$ são independentes. O primeiro termo da Eq. 77 possui dependência apenas com a contribuição do ruído para o espectro de potências presente no conjunto das frequências $N - 1$ (uma vez que o sinal periódico está relacionado com uma frequência do conjunto N), enquanto o segundo termo é desconhecido.

b) A Distribuição Associada de $P(\omega)$ e $P(\omega')$

É necessário destacar a relação de $P(\omega_n)$ entre os casos nos quais as amostras são igual e desigualmente espaçadas. Quando os dados são espaçados uniformemente os valores de $P(\omega_n)$ são independentes entre si, de acordo com o conjunto de frequências naturais, de modo que é possível aplicar as condições de ortogonalidade relacionadas a funções trigonométricas presentes na equação do periodograma (Eq. 36). Matematicamente, pode-se explicar que para os dados igualmente espaçados tem-se que a distribuição dada pelas funções seno e cosseno respeitam a condição de ortogonalidade. Enquanto que, para as amostras não uniformemente espaçadas, essa ortogonalidade desaparece. Não há uma maneira a partir da qual se recupere a ortogonalidade sem correlacionar as frequências, que devem permanecer independentes. Essa relação entre as frequências invalida o propósito da análise espectral, que não é um procedimento a ser aqui discutido.

Porém, se o intervalo de frequência varrido for escolhido de tal modo que o grau de dependência entre as potências for consideravelmente pequeno, é possível achar uma correlação entre $P(\omega)$ e $P(\omega')$ através dos seus respectivos índices, ω e ω' . Lomb (1976) mostrou que a quantidade relevante que resulta dessa correspondência é a *função de janela* mostrada na Eq. 46. Esta função contém informações relevantes sobre dependências e correlações.

Para que os valores assumidos por ω_n não se correlacionem, é preciso que a função $G(\omega)$ seja descrita por valores de n avaliados no conjunto de frequências naturais. A justificativa de tal suposição deve-se ao tratamento previsto no processo Gaussiano, pois nele a falta de correlação implica independência. Isso implica numa situação com ampla variedades de amostras no caso de $G(\omega)$ possuir valores nulos, relativamente pequenos ou que são aproximadamente espaçado de forma igual. Isso pode ser notado nas Figuras 5 e 6,

apresentadas por (SCARGLE, 1982). Tais valores nulos compreendem um conjunto de frequências naturais nas quais o periodograma é avaliado.

c) *A Probabilidade Falso Alarme* ⁶

Existe um valor de flutuação no ruído z_0 , o qual pode-se detectar uma periodicidade associado a um valor da potência p_0 .

Das Equações 71 e 73 temos que:

$$- \exp(-z) = 1 - [1 - \exp(-z)]$$

A distribuição exponencial de z_0 está relacionada com o número de frequência avaliado, da seguinte maneira:

$$\exp(-z_0) = (1 - p_0)^{\frac{1}{N}} \quad (78)$$

Assim,

$$\begin{aligned} - \exp(-z_0) &= 1 - [1 - \exp(-z_0)] \\ - \exp(-z_0) &= 1 - [1 - (1 - p_0)^{\frac{1}{N}}] \\ z_0 &= \ln[1 - (1 - p_0)^{\frac{1}{N}}] \end{aligned} \quad (79)$$

que é o valor limite para a potência que se desejava. A probabilidade de falso alarme é dada pelo termo p_0 e N é o número de frequências avaliadas no valor máximo do periodograma. p_0 é um número fixado para que se possa obter o valor da potência limiar desejada. Tal valor também está associado com a eficiência de detecção do sinal registrado. Por exemplo, seja $p_0 = 0,01$. Então,

$$z_0 \approx \ln(N/p_0) = \ln(N) + 4,6$$

Vale ressaltar que o valor da potência está associado a um valor da frequência. Seja $N = 30$, por exemplo, o valor escolhido. Consequentemente, $z_0 \approx 8$. Assim, Z (máximo valor de potência no periodograma), deve ser maior do que 8 para que se tenha o registro de um sinal com pelo menos 99% de probabilidade de detecção.

De acordo com a equação seguinte, que provém da relação sinal-ruído (Eq. 45),

$$P = \frac{1}{N^2} \left(\frac{X_0}{2\sigma_0} \right)^2$$

o valor da amplitude do sinal diminui com o aumento da frequência, quando consideramos $X_0/\sigma_0 = 1$.

⁶Aproximação usada para estimar a intensidade de um pico que mede o ruído de um conjunto de dados que não contém sinais periódicos (VANDERPIAS, 2018). (N.E.)

d) A Eficiência de Detecção

Na seção anterior foi visto que é possível detectar um sinal não periódico por meio da intensidade do pico relacionado ao ruído num conjunto de dados. O valor que representa o limite para tal detecção é definido, de acordo com a amostragem avaliada, e portanto não tem um valor único. Para a análise dos dados avaliados neste trabalho definiu-se um limite dado pelo valor da potência z_0 (que está relacionado ao valor p_0 - que é a probabilidade de falso alarme). No entanto a probabilidade de falso alarme não é suficiente para calcular a taxa de detecção de um sinal periódico dentro do conjunto de dados avaliado (VANDERPIAS, 2018). Desse modo, é preciso calcular a eficiência de detecção desse sinal. Nesta situação, usa-se a CDF avaliada em z_0 para calcular a probabilidade dessa detecção. Isto é,

$$p^*(N, P) = \Pr\{miss\} = F_N(z_0) = (1 - p_0)^{1-1/N} \{1 - \exp[-(z_0 + P)] \phi(z_0, P)\} \quad (80)$$

sendo $p_0 = 0,01$, escolhido para que seja possível detectar pelo menos 99% de probabilidade, ϕ e z_0 são dados pelas Equações 76 e 79, respectivamente. O fator $(1 - p_0)^{1-1/N}$ varia lenta e monotonicamente de 1 a $1 - p_0 = 0,99$, com N indo de 1 ao ∞ . A contribuição desse fator constante na última equação é desprezível para o resultado que se deseja obter e pode ser omitido sem perda de generalidade:

$$p^*(N, P) = \{1 - \exp[-(z_0 + P)] \phi(z_0, P)\} \quad (81)$$

Tendo N e P como parâmetros, a equação acima pode ser usada para comparar a eficiência de detecção do conjunto de dados. A Eq. 81 é definida como a probabilidade de detectar um sinal de potência P , ou seja:

$$DE = 1 - p^*(N, P) \quad (82)$$

sendo o sinal de potência dado por:

$$P = \left\{ [\ln(N/p_0)]^{1/2} - C \right\}^2 \quad (83)$$

uma vez que P não depende somente da amplitude do sinal, X_0 , mas também dos parâmetros observacionais σ_0 e N , como foi visto anteriormente.

Na Figura 7, apresentada por (SCARGLE, 1982), estão mostrados os resultados a partir do que foi discutido sobre a eficiência de detecção e taxa de falso alarme mostrando como eles se relacionam com a CDF da potência com e sem a presença do sinal. Outrossim, a taxa da relação sinal-ruído deve ser relativamente alta enquanto que a detecção limiar ($\Pr\{miss\}$) é baixa.

3.1.2 Maximização da eficiência de detecção

É possível maximizar a eficiência da detecção utilizando-se dos resultados anteriores. Para isso vamos considerar a potência do sinal P , restringindo sua relação apenas em função das frequências N e também que a variância σ_0^2 é proporcional a N . Assim, é verdade que para um conjunto de frequências N_1 há uma variância correspondente σ_1^2 . Analogamente, para um conjunto N_2 existe uma variância σ_2^2 . Considerando o caso em que $N_2 < N_1$, temos $\sigma_2 < \sigma_1$. Supondo que o ruído para cada conjunto de frequência esteja seguindo uma

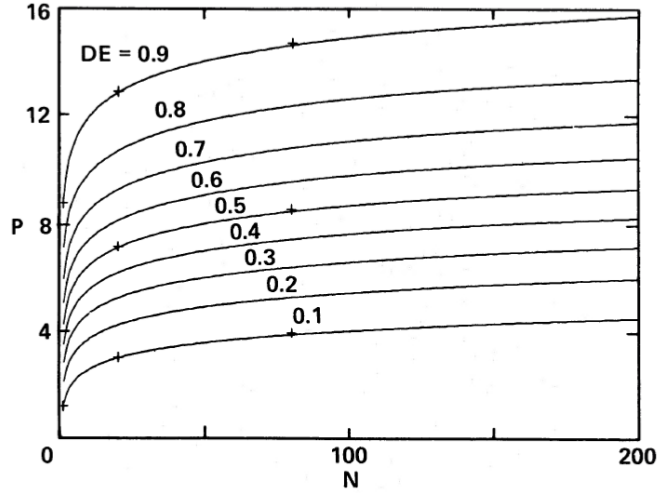


Figura 7: Contornos de eficiência de detecção ($DE = 1 - p^*$), como funções do sinal de potência P e o número de amostras, N . As curvas que contém o valor de DE são dadas a partir da equação 79, enquanto que os pontos marcados com o sinal ” + ” são calculados conforme as equações 72, 76 e 78. (Figura reproduzida de (SCARGLE, 1982)).

distribuição gaussiana, foi possível derivar uma expressão a partir da relação sinal-ruído (Eq. 45), tal que

$$P = N \left(\frac{1}{\sigma_0} \right)^2 \quad (84)$$

a partir de P1 e P2, temos

$$\frac{N_1}{\sigma_1^2} = \frac{N_2}{\sigma_2^2} \quad (85)$$

e conseqüentemente,

$$\sigma_2 = \left(\frac{N_2}{N_1} \right)^{1/2} \sigma_1 \quad (86)$$

De acordo com a Eq. 81, p^* aumenta monotonicamente com N , e para z_0 , $F_N(z_0)$ é monotonicamente crescente para z_0 . Isso era de se esperar, uma vez que todas as CDFs são crescentes em virtude da definição na Eq. 71. Da Eq. 79 pode ser visto que z_0 é uma função crescente de N . Igualmente, analisando a Eq. 80, percebe-se que p^* é uma função crescente. Conseqüentemente, o fato de $\sigma \propto N$ implica que, a probabilidade de detecção de um sinal periódico aumenta. Tal comportamento foi justificado quando foi analisada a lacuna espectral usando a Eq. 73. Valores muito pequenos de N significam que poucas tentativas de maximizar a eficiência de detecção do sinal periódico estão sendo realizadas, tornando o maior valor do pico no espectro de potências, que mede a variação da componente do ruído, mais relevante na análise do sinal.

Não se pode esquecer a relação entre os dados e o intervalo de amostragem efetivo, Δt . O limiar de detecção aumenta em função da frequência e a eficiência de detecção pode ser maximizada calculando a média dos dados mesmo quando se tem apenas dois pontos. Porém, há um problema: quanto mais dados são analisados, num dado intervalo de tempo, menor será a *frequência de Nyquist* (ver Eq. 37). A diminuição da frequência que define um limite mínimo para a cobertura temporal, torna mais difícil determinar a frequência característica de uma eventual oscilação periódica no conjunto de dados. Todavia, é preciso considerar esse efeito quando não se tem informações acerca do sinal de frequência,

realizando uma investigação mais extensiva dos dados.

Existe uma relação entre o número médio de dados avaliados e o número de frequências, sendo $N_0/2$ o fator limite que elimina a correlação entre as frequências avaliadas no periodograma $P(\omega)$. O número de frequências deve ser muito menor do que $N_0/2$. Tal restrição não modificaria a potência do sinal, mas reduziria o valor de z_0 de acordo com a Eq. 79. Portanto, sempre que N decai abaixo de $N_0/2$, a eficiência de detecção melhora conforme a Eq. 81. Isso só pode ser feito se, e somente se, os sinais que não nos interessam estiverem presentes na gama de frequências ignoradas, ou seja, acima de $N_0/2$. Para um N fixo, verifica-se que a Eq. 83 é uma função crescente de N_0/σ_0^2 , o que torna a potência do sinal maximizada. Assim, pode-se confirmar que quanto maior for o conjunto de dados, maior será a eficiência de detecção.

A verificação acerca da condição de crescimento permanente da P em relação a N não poder ser analisada tão facilmente, pois os resultados dependem de quão crescente é a função. A Figura 7 mostra as curvas geradas a partir dos tratamentos impostos e que tinham o fim de maximizar a eficiência.

3.1.3 Comparação entre as formas clássica e modificada do Periodograma

a) O Periodograma Clássico

Para uma detecção precisa de um sinal senoidal, foi preciso utilizar a *função de janela* definida na Eq. 38. Com isso foi possível obter a resposta do periodograma clássico, Eq. 32, apenas em termos do tempo observacional t_j . (SCARGLE, 1982) utilizou os dados obtidos por um programa de paralaxe ⁷ usado no *U.S. Naval Observatory* para construir os gráficos apresentados na Figura 3. Tal Figura mostra o comportamento gráfico mediante o tratamento empregado através da *função de janela*. (SCARGLE, 1982) argumenta que *a observação foi feita pelo Dr. Harrington Kindly a partir de três estrelas nas quais a perturbação devido ao companheiro invisível* ⁸ *foi detectada durante a execução do programa* (Behall and Harrington 1976; Harrington 1977). É nítida a diferença entre as janelas das três estrelas, apesar das técnicas observacionais empregadas para cada uma delas serem similares. O modelo padrão "sin c^2 " ⁹ foi empregado e sua aplicação gerou resultados similares às técnicas empregadas pelo periodograma modificado no tratamento de casos igualmente espaçados. Observe que todas as três possuem um pico central estreito, cuja largura é da ordem de $2\pi/T$, chamado de *main lobe*. Para o primeiro ponto no qual a função "sin c^2 " se anula, uma das *janelas* se anula e as outras duas têm pontos de mínimo. Podemos considerar que todas elas possuem frequências próximas e igualmente espaçadas com maiores amplitudes àquelas que decaem rapidamente com a função "sin c^2 ", mesmo sabendo que as frequências próximas (sidelobes) diferem de estrela para estrela. Há uma

⁷Desvio aparente na posição de um objeto (estrelas, por exemplo) quando visto por observadores em locais distintos. (N.E.)

⁸O nosso sistema solar é uma raridade no universo por muitos fatores, inclusive pelo fato de possuir uma única estrela orbitada pelos planetas. Geralmente, as estrelas se encontram orbitando uma companheira e ambas girando em torno do centro de massa dos seus respectivos sistemas. Aliás, na constelação da Ursa Maior, localizada a 250 mil anos-luz da terra, há de um sistema solar com cinco sóis. (N.E.)

⁹Na teoria de processamento digital, a função *sinus cardinalis* é definida como: $\text{sin } c = \frac{\sin(\pi x)}{\pi x}$, sendo π o fator de normalização. Percebe-se que a integral sobre o eixo real tem valor constante igual a 1, já que $\lim_{x \rightarrow 0} \text{sin } c = 1$. (N.E.)

diferença entre a *real função de janela* e o modelo dado pela função "sin c^2 ", principalmente pela presença de um maior pico em qualquer um dos lados do *main lobe*, que o desloca 1 ciclo por ano ($\omega = 2\pi$ radianos por ano) (Scargle, 1982). Outrossim, a função "sin c^2 " possui uma particularidade que a difere da *real função de janela*, chamada de séries de picos largos em ambos os lados do *main lobe*, que desaparece por meio de integração múltipla de 1 ciclo por intervalo de amostragem (i.e., *frequência de Nyquist*). Vale ressaltar que os conceitos de características estritamente periódicas e o de deslocamento de frequências (*aliasing*) já foram tratados na Seção 3.1.

b) O Periodograma Modificado

Passemos agora à comparação entre as definições clássica e modificada do periodograma (Equações 61 e 62, respectivamente). Como vimos na Seção 3.1, se tratando de uma entrada senoidal, não é possível obter a resposta do periodograma, utilizando a Eq. 38, pois essa equação não é compatível com a redução apresentada na definição modificada (Eq. 61). No entanto, quando se é considerada uma frequência específica, numa amostra particular de um conjunto temporal t_i notou-se que é possível calcular a resposta de tal comportamento senoidal. O sinal será dado por:

$$X_s = \sin \omega_s t_i + aR_i \quad (87)$$

em que o termo aR_i representa o ruído. Apresentamos aqui uma reprodução do processo realizado por (SCARGLE, 1982) para produzir o ruído, o qual foi gerado por adição de cinco variáveis aleatórias distribuídas uniformemente no intervalo $(-\frac{1}{2}, \frac{1}{2})$, o que resultou numa rvariável aleatória pseudo-Gaussiana de variância 0,64.

Na Figura 4, apresentada por (SCARGLE, 1982), nota-se o periodograma modificado para o caso com a mesma amostragem da Figura 3a, usando os parâmetros $\omega_s = 10$ e $a = 0$. Partindo da hipótese de que a frequência é alta o suficiente, contendo apenas os termos positivos do conjunto de frequências naturais, percebe-se que o periodograma se torna muito eficiente. Nota-se também uma semelhança entre as curvas, mostrando que a diferença entre o periodograma modificado e a definição clássica não é tão grande. Essa simetria é provavelmente devido a sobreposição dos sinais de frequências (cálculos análogos com sinais de frequências maiores geram comportamentos muitos mais simétricos).

Expomos aqui a análise feita por (SCARGLE, 1982), a acerca da Figura 4, na qual é mostrado o periodograma para os mesmos dados uniformemente espaçados e linearmente interpolados. O resultado apresentado indica que tal interpolação não é boa para sinais de altas frequências, pois a maioria as oscilações podem não ser detectadas. Portanto, não é de se admirar que os espectros de dados interpolados não tenham informações sobre o pico que deveria existir por conta das oscilações presentes na detecção. Já para as frequências próximas a ω_o , a interpolação apresentada é válida na descrição do comportamento, mas dificilmente seria usada se uma forma alternativa fosse viável.

A Figura 5 mostra um outro resultado da versão modificada do periodograma. Variações nos valores da frequência entre 10% e 20% são típicas de uma amostragem moderadamente irregular (SCARGLE, 1982).

c) Ruído nos dados

A Figura 8 apresentada por (SCARGLE, 1982) e adaptada para este documento mostra o periodograma modificado (Eq. 61) com a inserção de ruído no tratamento dos dados considerado na situação ilustrada pela Figura 4.

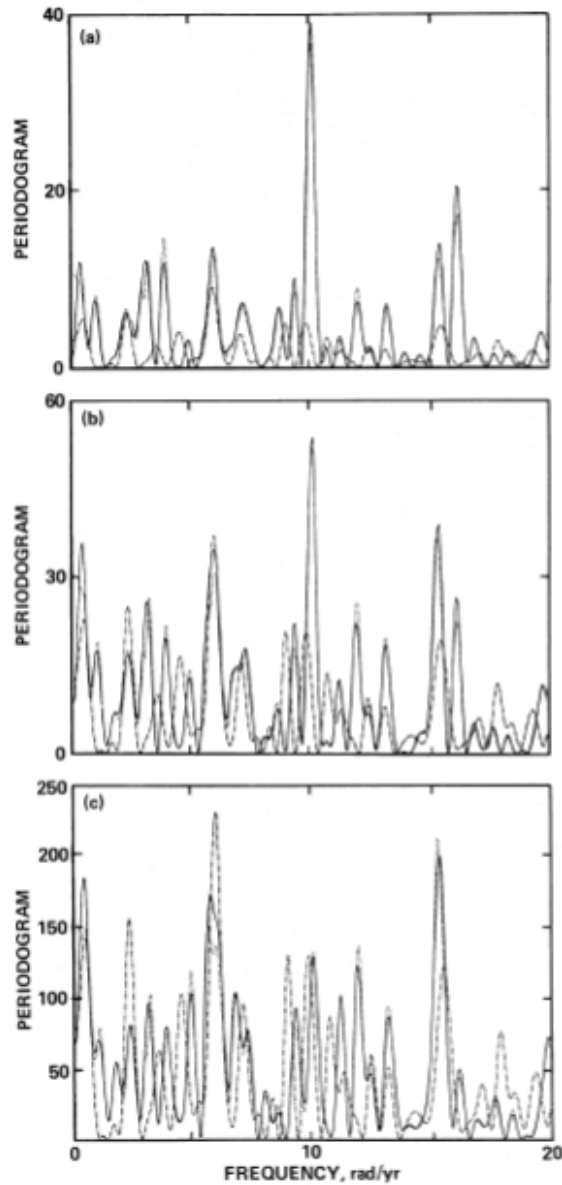


Figura 8: *Pseudowindows* determinadas por análises de dados artificiais, como na Figura 4, mas com a adição do ruído de acordo com a Eq. 87 com (a) $a = 2$, (b) $a = 4$ e (c) $a = 10$ (as relações sinal-ruído correspondentes são 16,4 e 0.65). A linha sólida é o periodograma clássico, e a linha pontilhada é o periodograma modificado; a linha tracejada é o periodograma clássico em função apenas do ruído. (Figura reproduzida de (SCARGLE, 1982)).

Da expressão proveniente da relação sinal-ruído apresentada na Seção 3.1.1, temos

$$P = N_0 \left(\frac{X_0}{2\sigma_0} \right)^2 \quad (88)$$

A interpretação que decorre da equação acima, assim como os valores a seguir, seguem a análise feita por (SCARGLE, 1982). Consequentemente os resultados apresentados são uma reprodução do que fora obtido no artigo base deste documento. Para $X_0 = 1$, $N_0 = 107$ e $\sigma_0 = 0.64$ tem-se que os níveis para a relação sinal-ruído são: $(a) = 65$, $(b) = 2.6$ e $(c) = 0.65$. Com isso nota-se que o sinal é claramente detectado em (a) , enquanto em (b) o ruído é tal que a detecção poderia ser alcançada apenas para um limite bastante amplo (e assim, considerando a probabilidade de falso alarme). Tomando $N_0 = 107$ na Eq. 73 foi obtido o valor que corresponde a uma detecção de 97,0% de confiança. Isso é consequência do quociente entre a potência do pico pela potência do ruído que resulta em 8.2 na Figura 8. Então, foi possível detectar a amplitude 1 para uma onda senoidal na presença do ruído de variância maior que 2,5 caso tal detecção possa ser utilizada. O resultado dessa análise mostra também que o periodograma pode ser utilizado para detectar um sinal periódico mesmo quando a amostra contém ruído. Na Figura 8c a detecção não pode ser feita, pois a relação entre as potências gera um valor 5, o que corresponde a uma probabilidade de 50% não sendo possível assim diferir o ruído de um sinal periódico.

4 Construção de um protótipo que simula o movimento de um Pulsar

Uma vez compreendido o tratamento matemático usado na definição do periodograma, é possível utilizar essa ferramenta para analisar a curva de luz gerada pelo protótipo (apresentado nas Figuras 9, 10, 11). Com o entendimento computacional necessário para a aplicação dessa técnica, trabalhos posteriores podem identificar se a emissão proveniente do protótipo é detectada de maneira regular, a fim de compará-la com o que se observa em um objeto astrofísico típico, como por exemplo um Pulsar. Um pulsar é uma estrela de nêutrons que emite energia pulsada de forma regular, que é parte remanescente de uma estrela muito mais massiva que o Sol, extremamente densa, com uma altíssima velocidade de rotação, a qual gera um intenso campo magnético, o qual capta elétrons restringindo-os a percorrermos trajetórias, tais que a emissão energética resultante é direcionada especificamente para os espaço interestelar. Tal fenômeno, quando observado, se assemelha ao efeito de iluminação orientada de um farol (SAGAN, 1998).

A construção do protótipo foi proposta com o objetivo de criar um vínculo prático e experimental com o teoria abordada ao longo deste trabalho. Neste sentido, a elaboração do projeto e o entendimento do fenômeno físico a ele associado foram conceitos relevantes no desenvolvimento deste documento. Construiu-se um experimento que representa um pulsar, que tem uma fração de sua emissão pulsada decorrente de interação das linhas de campo magnético com um meio material local e desalinhamento dos eixos de rotação e de orientação de campo magnético. Por conta do controle da inclinação dos eixos magnético e rotacional, foi possível detectar sinais periódicos por meio de um fotodetector acoplado ao equipamento. A seguir são mostradas algumas figuras do dispositivo.

Na Figura 9, eixo de rotação gira pela ação do motor alimentado por uma fonte de 5 V. Sendo que, por conta da resistência do próprio material, a velocidade de giro é mantida de acordo com a tensão aplicada. Os botões prestes a serem acionados, fazem a ligação com o motor e com a lâmpadas de led soldadas na extremidades do eixo horizontal.

Com base no arranjo mostrado na Figura 11 é possível caracterizar a fração de energia pulsada, incluindo a determinação do período de pulsação e modulação da curva de luz como função da inclinação relativa entre os eixos magnético e rotacional.

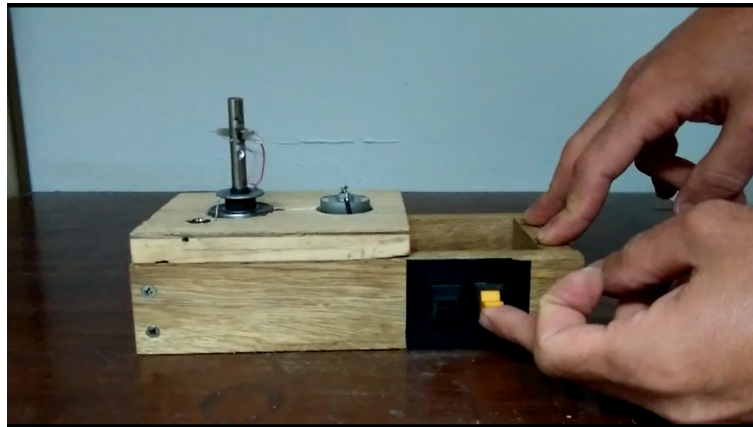


Figura 9: Protótipo. Destaque para o eixo de rotação (vertical), eixo de campo magnético (horizontal) e motor de alimentação.

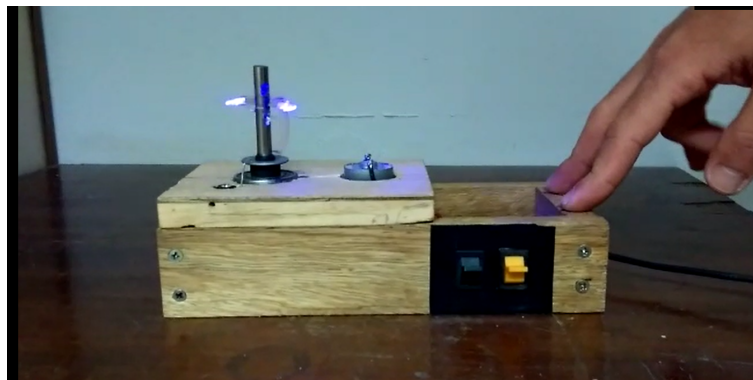


Figura 10: As lâmpadas de led estão acesas e o motor alimentado. Nessa situação, o giro está acontecendo e a emissão pode ser detectada.

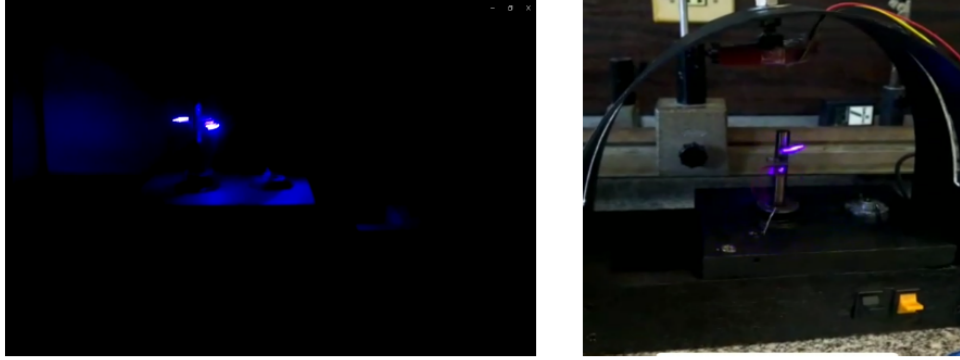


Figura 11: À esquerda: Luz ambiente desligada de modo a perceber com maior precisão a emissão luminosa proveniente do eixo horizontal. À direita: Fórmica em forma de parábola acoplada ao dispositivo, servindo de suporte para o fotodetector que está no centro da concavidade.

5 Considerações finais

O periodograma Lomb-Scargle (Lomb 1976; Scargle 1982) é um algoritmo frequentemente utilizado devido a sua aplicabilidade em situações nas quais se deseja encontrar uma periodicidade em um determinado conjunto de dados. Em especial, a comunidade astronômica faz uso dessa ferramenta uma vez que fenômenos periódicos são comuns no estudo do Cosmo. Vimos que a análise de Fourier é a base matemática necessária para o desenvolvimento do periodograma e que o caso Lomb-Scargle é geralmente apresentado como uma técnica que permite detectar o componente periódico em um conjunto de amostras desigualmente espaçadas (VANDERPIAS, 2018).

Tendo em vista a variedade de métodos utilizados para analisar séries temporais astronômicas (e.g. Ajuste de mínimos quadrados e Aproximação Baresiana), por que usar o caso específico do periodograma Lomb-Scargle? Podemos responder essa pergunta argumentando que o mesmo seja possivelmente a técnica mais conhecida para analisar a periodicidade de dados espaçados de maneira desigual não só na Astronomia, mas em outros ramos da ciência. Porém existe uma razão mais sutil por trás dessa escolha: o fato do seu fundamento matemático ser construído através da análise de Fourier, como também sua correspondência com o método de ajuste de mínimos quadrados (apresentado no final da seção 3.1). Outrossim, existem outras aplicações para o tipo de análise utilizada no caso Lomb-Scargle. Por exemplo, o algoritmo pode ser estendido para medir decaimentos de sinais, detectar sinais não estacionários e sinais de frequências múltiplas (VANDERPIAS, 2018).

A aplicação do periodograma Lomb-Scargle no estudo do protótipo apresentado na seção 4 pode servir de base para trabalhos futuros. Vimos que, com base no arranjo experimental mostrado na seção anterior pode-se detectar sinais periódicos provenientes da variação do brilho do objeto, de modo que é possível determinar o seu período de pulsação.

Por fim, o método de análise de séries temporais apresentado neste documento contém características que fazem com que seu uso possa ser expandido em estudos práticos e teóricos em diversos campos da ciência. Além disso, o vínculo com a Astrofísica torna possível a descrição matemática de fenômenos que são cruciais para o entendimento da

evolução cósmica e o desenvolvimento da ciência moderna.

6 Apêndices

A Obtenção dos coeficientes da Expansão em Série de Fourier para funções periódicas de intervalos iguais a 2π definida em \mathbb{R}

$a_0; n = 0$

Sejam f e g duas funções distintas, definidas no intervalo $[\alpha, \beta]$. O produto escalar entre elas pode ser definido por:

$$(f, g) = \int_{\alpha}^{\beta} f(x) g(x) dx$$

No caso em questão, o conjunto de funções que podem ser calculadas na expansão é:

$$\{1, \cos(x), \sin(x), \cos(2x), \sin(2x), \dots\}$$

E o intervalo considerado é $[0, 2\pi]$

O conjunto supracitado é ortogonal, isto é, as funções a ele pertencente são independentes, significando que o produto escalar entre pares de funções distintas desse conjunto é nulo (CHURCHILL, 1978). Então, efetuando o produto escalar em ambos os lados da expansão com a função constante 1, temos:

$$f(x) = \sum_{n=0}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]$$

$$(f(x), 1) = a_0 (\cos(0), 1) + b_0 (\sin(0x), 1) + a_1 (\cos(x), 1) + b_1 (\sin(x), 1) + \dots$$

Aplicando a definição de produto escalar entre pares distintas de funções de um conjunto ortogonal, percebe-se que o primeiro termo, a_0 , é o único que não será nulo nesse somatório, uma vez que $\cos(0) = 1$. Desse modo,

$$\int_0^{2\pi} f(x) dx = a_0 (\cos(0), 1)$$

$$\int_0^{2\pi} f(x) dx = a_0 (1, 1)$$

$$\int_0^{2\pi} f(x) dx = a_0 \int_0^{2\pi} dx$$

$$\int_0^{2\pi} f(x) dx = a_0 2\pi$$

Portanto,

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx$$

q.e.d.

$a_n; n \geq 1$:

Procedendo analogamente à demonstração anterior, faz-se necessário agora efetuar o produto escalar em ambos os lados da expansão em série de Fourier com a função $\cos(mx)$.

$$\begin{aligned} f(x) &= \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)] \\ (f(x), \cos(mx)) &= \sum_{n=1}^{\infty} [a_n (\cos(nx), \cos(mx)) + b_n (\sin(nx), \cos(mx))] \end{aligned}$$

Da definição de ortogonalidade citada anteriormente, percebe-se que o produto escalar entre os pares de funções iguais dessa expansão só será diferente de zero quando $m = n$. Então,

$$(f(x), \cos(mx)) = a_m (\cos(mx), \cos(mx))$$

Consequentemente,

$$\begin{aligned} (f(x), \cos(mx)) &= \int_0^{2\pi} a_m (\cos(nx) \cos(mx)) dx = a_m \int_0^{2\pi} \cos^2(mx) dx \\ &= a_m \int_0^{2\pi} \frac{1}{2} [1 + \cos(2mx)] dx \\ &= a_m \int_0^{2\pi} \frac{1}{2} dx + \underbrace{a_m \int_0^{2\pi} \cos(2mx) dx}_0 \end{aligned}$$

Então,

$$\begin{aligned} \int_0^{2\pi} f(x) \cos(mx) dx &= \frac{1}{2} a_m 2\pi \\ \int_0^{2\pi} f(x) \cos(mx) dx &= a_m \pi \end{aligned}$$

Como $m = n$, temos, portanto:

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx$$

q.e.d.

Mostrar-se-á que $\forall m \neq n$,

$$\int_0^{2\pi} \cos(mx) \cos(nx) dx = 0$$

Prova:

Considere as seguintes relações trigonométricas:

$$\cos(a + b) = \cos a \cdot \cos b - \sin a \cdot \sin b \quad (89)$$

$$\cos(a - b) = \cos a \cdot \cos b + \sin a \cdot \sin b \quad (90)$$

Somando (15) e (16), segue:

$$\begin{aligned} \cos(a + b) + \cos(a - b) &= 2 \cos a \cdot \cos b \\ \cos a \cdot \cos b &= \frac{1}{2} [\cos(a + b) + \cos(a - b)] \end{aligned}$$

Assim,

$$\begin{aligned} \int_0^{2\pi} \cos(mx) \cos(nx) dx &= \int_0^{2\pi} \frac{1}{2} [\cos(mx + nx) + \cos(mx - nx)] dx \\ &= \int_0^{2\pi} \frac{1}{2} [\cos(x(m + n)) + \cos(x(m - n))] dx \\ &= \frac{1}{2} \int_0^{2\pi} \cos(x(m + n)) dx + \frac{1}{2} \int_0^{2\pi} \cos(x(m - n)) dx \\ &= \frac{1}{2} \left[\underbrace{\frac{1}{m + n} \sin(x(m + n))}_0 \right]_0^{2\pi} + \left[\underbrace{\frac{1}{m - n} \sin(x(m - n))}_0 \right]_0^{2\pi} \end{aligned}$$

Portanto,

$$\int_0^{2\pi} \cos(nx) \cos(mx) dx = 0; \forall m \neq n$$

q.e.d.

Dando continuidade as demonstrações dos coeficientes da expansão de Fourier, passemos ao último coeficiente.

$b_n; n \geq 1$:

Efetuando o produto escalar em ambos os lados da expansão com a função $\sin(mx)$, temos:

$$\begin{aligned} f(x) &= \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)] \\ (f(x), \sin(mx)) &= \sum_{n=1}^{\infty} [a_n (\cos(nx), \sin(mx)) + b_n (\sin(nx), \sin(mx))] \end{aligned}$$

Como é sabido, de acordo com a ortogonalidade das funções aqui envolvidas, o produto escalar para cada par de funções iguais na expansão só será diferente de zero para $m = n$. Com isso,

$$(f(x), \sin(mx)) = b_m (\sin(mx), \sin(mx))$$

Consequentemente,

$$\begin{aligned}(f(x), \sin(mx)) &= \int_0^{2\pi} b_m (\sin(mx) \sin(mx)) dx = b_m \int_0^{2\pi} \sin^2(mx) dx \\ &= b_m \int_0^{2\pi} \frac{1}{2} [1 - \cos(2mx)] dx \\ &= b_m \int_0^{2\pi} \frac{1}{2} dx - \underbrace{b_m \int_0^{2\pi} \cos(2mx) dx}_0\end{aligned}$$

Logo,

$$\begin{aligned}\int_0^{2\pi} f(x) \sin(mx) dx &= \frac{1}{2} b_m 2\pi \\ \int_0^{2\pi} f(x) \sin(mx) dx &= b_m \pi\end{aligned}$$

Como $m = n$, temos, portanto:

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx) dx$$

q.e.d.

Justificando a escolha de $m = n$, passemos a demonstração de que para qualquer $n \neq m$, a integral

$$\int_0^{2\pi} \sin(mx) \sin(nx) dx$$

é nula.

Prova:

Efetuada a subtração entre as equações (15) e (16), temos:

$$\begin{aligned}\cos(a + b) - \cos(a - b) &= 2 \sin a \cdot \sin b \\ \sin a \cdot \sin b &= \frac{1}{2} [\cos(a - b) - \cos(a + b)]\end{aligned}$$

Assim,

$$\begin{aligned}
\int_0^{2\pi} \sin(nx) \sin(mx) dx &= \int_0^{2\pi} \frac{1}{2} [\cos(mx - nx) - \cos(mx + nx)] dx \\
&= \int_0^{2\pi} \frac{1}{2} [\cos(x(m - n)) - \cos(x(m + n))] dx \\
&= \frac{1}{2} \int_0^{2\pi} \cos(x(m - n)) dx - \frac{1}{2} \int_0^{2\pi} \cos(x(m + n)) dx \\
&= \frac{1}{2} \underbrace{\left[\frac{1}{m - n} \sin(x(m - n)) \right]_0^{2\pi}}_0 - \underbrace{\left[\frac{1}{m + n} \sin(x(m + n)) \right]_0^{2\pi}}_0
\end{aligned}$$

Logo,

$$\int_0^{2\pi} \sin(nx) \sin(mx) dx = 0; \forall m \neq n$$

q.e.d.

B Dedução da constante complexa C_n

Prova:

No desenvolvimento da expansão em Série de Fourier para funções periódicas de intervalos iguais a 2π definidas em \mathbb{C} vimos que,

$$C_n = \frac{a_n}{2} + \frac{b_n}{2i}$$

Sabe-se que os coeficientes a_n e b_n são dados por:

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx \quad ; \quad b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx) dx$$

Logo,

$$C_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) \cos(nx) dx + \frac{1}{2\pi i} \int_0^{2\pi} f(x) \sin(nx) dx$$

Multiplicando ambos os lados por i^2 , temos:

$$\begin{aligned}
-C_n &= -\frac{1}{2\pi} \int_0^{2\pi} f(x) \cos(nx) dx + \frac{i}{2\pi} \int_0^{2\pi} f(x) \sin(nx) dx \\
-C_n &= \frac{1}{2\pi} \int_0^{2\pi} f(x) [-\cos(nx) + i \sin(nx)] dx
\end{aligned}$$

Multiplicando ambos lados por -1 , sigue:

$$C_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) [\cos(nx) - i \sin(nx)] dx$$

Portanto,

$$C_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx$$

q.e.d.

C Obtenção dos coeficientes da Expansão em Série de Fourier para funções reais de período fundamental T

a_0 ; $n = 0$:

Efetuada o produto escalar em ambos os lados da expansão com a função constante 1, temos:

$$\begin{aligned} (f(x), 1) &= a_0 + \sum_{n=1}^{\infty} a_n \left[\left(\cos \left(\frac{2\pi}{T} nx \right), 1 \right) + \sum_{n=1}^{\infty} b_n \left(\sin \left(\frac{2\pi}{T} nx \right), 1 \right) \right] \\ (f(x), 1) &= \sum_{n=0}^{\infty} \left[a_n \left(\cos \left(\frac{2\pi}{T} nx \right), 1 \right) + \sum_{n=0}^{\infty} b_n \left(\sin \left(\frac{2\pi}{T} nx \right), 1 \right) \right] \\ (f(x), 1) &= a_0 (\cos(0), 1) + \underbrace{b_0 (\sin(0), 1)}_0 + \underbrace{a_1 \cos \left(\frac{2\pi}{T} nx \right)}_0 + \underbrace{b_1 \sin \left(\frac{2\pi}{T} nx \right)}_0 + \dots \end{aligned}$$

Pela definição de produto escalar entre pares distintos de funções de um conjunto ortogonal, percebe-se que o primeiro termo, a_0 , é o único que não será nulo nessa somatória. Desse modo,

$$\begin{aligned} (f(x), 1) &= a_0 (1, 1) \\ \int_0^T f(x) dx &= a_0 \int_0^T dx \\ a_0 T &= \int_0^T f(x) dx \end{aligned}$$

Portanto,

$$a_0 = \frac{1}{T} \int_0^T f(x) dx$$

q.e.d.

$n \geq 1$:

Analogamente, tomemos o produto escalar em ambos os lados da expansão com a função $\cos \left(\frac{2\pi}{T} mx \right)$ para o coeficiente " a_n ", e com a função $\sin \left(\frac{2\pi}{T} mx \right)$ para o coeficiente " b_n ".

a_n :

$$\begin{aligned} \left(f(x), \cos \left(\frac{2\pi}{T} mx \right) \right) &= \sum_{n=1}^{\infty} a_n \left[\cos \left(\frac{2\pi}{T} nx \right), \cos \left(\frac{2\pi}{T} mx \right) \right] + \\ &+ \sum_{n=1}^{\infty} b_n \left[\sin \left(\frac{2\pi}{T} nx \right), \cos \left(\frac{2\pi}{T} mx \right) \right] \end{aligned}$$

De acordo com as condições de ortogonalidade, para $m = n$, segue:

$$\left(f(x), \cos\left(\frac{2\pi}{T}nx\right) \right) = a_n \left[\cos\left(\frac{2\pi}{T}nx\right), \cos\left(\frac{2\pi}{T}nx\right) \right]$$

Conseqüentemente,

$$\begin{aligned} \int_0^T f(x) \cos\left(\frac{2\pi}{T}nx\right) dx &= a_n \int_0^T \cos^2\left(\frac{2\pi}{T}nx\right) dx = \\ &= a_n \int_0^T \frac{1}{2} \left[1 + \cos\left(\frac{4\pi}{T}nx\right) \right] dx \\ &= a_n \frac{T}{2} \end{aligned}$$

Portanto:

$$a_n = \frac{2}{T} \int_0^T f(x) \cos\left(\frac{2\pi}{T}nx\right) dx$$

q.e.d.

b_n :

$$\begin{aligned} \left(f(x), \sin\left(\frac{2\pi}{T}mx\right) \right) &= \sum_{n=1}^{\infty} a_n \left[\cos\left(\frac{2\pi}{T}nx\right), \sin\left(\frac{2\pi}{T}mx\right) \right] + \\ &+ \sum_{n=1}^{\infty} b_n \left[\sin\left(\frac{2\pi}{T}nx\right), \sin\left(\frac{2\pi}{T}mx\right) \right] \end{aligned}$$

Então, para $m = n$:

$$\left(f(x), \sin\left(\frac{2\pi}{T}nx\right) \right) = b_n \left[\sin\left(\frac{2\pi}{T}nx\right), \sin\left(\frac{2\pi}{T}nx\right) \right]$$

Que implica em,

$$\begin{aligned} \int_0^T f(x) \sin\left(\frac{2\pi}{T}nx\right) dx &= b_n \int_0^T \sin^2\left(\frac{2\pi}{T}nx\right) dx = \\ &= a_n \int_0^T \frac{1}{2} \left[1 - \cos\left(\frac{4\pi}{T}nx\right) \right] dx \\ &= b_n \frac{T}{2} \end{aligned}$$

Portanto:

$$b_n = \frac{2}{T} \int_0^T f(x) \sin\left(\frac{2\pi}{T}nx\right) dx$$

q.e.d.

Agradecimentos

Esta monografia marca o final de um ciclo. Ao longo dos últimos cinco anos, obtive além do conhecimento teórico para exercer a profissão de Físico Licenciado, a amizade e estima de alguns professores que foram essenciais para a minha formação como profissional e como ser humano:

Prof. Dr. André Maurício Conceição de Souza (DFI - UFS);
Prof. Dr. Bruno Luis de Andrade Santos (DMA - UFS);
Prof. Dr. Edvaldo Alves de Souza Júnior (DFI - UFS);
Prof. Dr. Frederico Guilherme de Carvalho Cunha (DFI - UFS);
Prof. Dr. Jhon Freddy Martinez Ávila (DFI - UFS);
Prof. Dr. José Carlos Leite dos Santos (DMA - UFS);
Prof. Dr. Marcos Antônio Couto dos Santos (DFI - UFS);
Prof. Dr. Naldisson dos Santos (DMA - UFS);
Prof. Dr. Osmar de Souza e Silva Júnior (DFI - UFS);
Prof. Dr. Raimundo Lopes de Oliveira Filho (DFI - UFS),
Prof. Dr. Stoian Ivanov Zlatev (DFI - UFS).

A construção do protótipo, tal como apresentado, foi possível graças ao apoio fundamental da seguinte equipe:

André Mansfield Andrade de Araújo (Técnico de Laboratório - DFI);
Antelmo Santos Chaves (Técnico de Laboratório - DFI);
Clifsson Rolemberg Andrade (Técnico de Laboratório - DFI);
Jaime José dos Santos (Carpinteiro e Maceneiro- UFS);
José Cristiano dos Santos (Serralheiro - UFS);
Luiz Macedo Fiaes Júnior (Técnico de Laboratório - DFI),
Vilton Ricardo dos Santos (Técnico de Laboratório - DFI)

A eles e aos demais professores e colegas do Departamento de Física, estimo meus sinceros votos de gratidão.

Referências

- CHURCHILL, R. V. (1978). Séries de Fourier e Problemas de Valores de Contorno. Guanabara Dois.
- IONIDES, E. (2016). Smoothing in the time and frequency domains. <https://ionides.github.io/531w16/notes08/notes8.html>.
- PUPIN, J. R. (2011). Introdução às séries e transformadas de fourier e aplicações no processamento de sinais e imagens. Departamento de Matemática, Universidade Federal de São Carlos.
- SAGAN, C. (1998). Bilhões e Bilhões. Companhia das Letras.
- SANTOS, F. J. (2004). Introdução às séries de fourier. <https://pt.scribd.com/document/409740787/sf-pdf>.
- SCARGLE, J. D. (1982). Studies in astronomical time series analysis.ii. statistical aspects of spectral analysis of unevenly spaced data. The Astrophysical Journal.
- SILVA JR., O. S. (2009). Métodos de Física Teórica I e II. Universidade Federal de Sergipe, CESAD. São Cristóvão.
- VANDERPIAS, J. T. (2018). Understanding the lomb-scargle periodogram. The Astrophysical Journal.