Lomb periodogram

Pedro Pons-Villalonga, 2021

1 Ordinary least squares

Consider a set of observations $\{y_i\}$ at points $\{x_i\}$, that we model as a sum of functions:

$$z_i = \sum_j f_j(x_i)\beta_j \tag{1}$$

Where the β_j are coefficients. In matrix form:

$$\vec{z} = \mathrm{F}\vec{\beta}$$

Where:

$$\mathrm{F} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_1) & \cdots & \\ \vdots & & \ddots & \\ f_1(x_N) & \cdots & & f_p(x_N) \end{pmatrix}$$

We want to minimize the sum of the squares of the differences between the observations and the model:

$$S \equiv \|\vec{y} - \vec{z}\|^2 = \left\|\vec{y} - \mathrm{F}\vec{\beta}\right\|^2 = \sum_i \left(y_i - \sum_j f_j\left(x_i\right)\right)^2 \tag{2}$$

The values of the β_p parameters can be found solving the normal equations (theorem):

$$\left(\mathrm{F}^T\mathrm{F}\right)\vec{\beta} = \mathrm{F}^T\vec{y} \tag{3}$$

Then, we can find $\mathrm{F}\vec{\beta}$ from this equation:

$$\mathrm{F}\vec{\beta} = \mathrm{F}\left(\mathrm{F}^T\mathrm{F}\right)^{-1}\left(\mathrm{F}^T\mathrm{F}\right)\vec{\beta} = \mathrm{F}\left(\mathrm{F}^T\mathrm{F}\right)^{-1}\mathrm{F}^T\vec{y} \tag{4}$$

We can rename:

$$\mathrm{P} \equiv \mathrm{F}\left(\mathrm{F}^T\mathrm{F}\right)^{-1}\mathrm{F}^T \qquad\qquad \mathrm{M} \equiv \mathbb{1} - \mathrm{P} \tag{5}$$

Where P and M are symmetric, idempotent matrices:

$$\mathrm{P}^T = \mathrm{P} \qquad \mathrm{P}^2 = \mathrm{P} \qquad \text{and} \qquad \mathrm{M}^T = \mathrm{M} \qquad \mathrm{M}^2 = \mathrm{M}$$

Then:

$$\mathrm{F}\vec{\beta} = \mathrm{P}\vec{y}$$

So, using the definitions and properties of M and P:

$$S = \|\vec{y} - \mathrm{P}\vec{y}\| = \|\mathrm{M}\vec{y}\| = \vec{y}^T\mathrm{M}^T\mathrm{M}\vec{y} = \vec{y}^T\vec{y} - \vec{y}^T\mathrm{P}\vec{y} \tag{6}$$

The last term is what is called by Lomb the "reduction in the sum of squares".

2 One-dimensional Lomb periodogram

Lomb [2] and Scargle [3] studied the statistics of the least-squares method for frequency analysis, first introduced and developed in [1, 4].

Let $\{y_i\}$ be a set of N observations with zero mean, obtained at times $\{t_i\}$, that can be evenly spaced or not. We can fit this observation data to the model:

$$y_k + \epsilon_k = a\cos\omega t_k + b\sin\omega t_k \tag{7}$$

Where ϵ_k are the errors, supposedly independent, with zero mean and variance σ^2.

We can fit this model using least squares as described above. For that, our matrix F will be:

$$\mathrm{F} = \begin{pmatrix} \cos\omega t_1 & \sin\omega t_1 \\ \cos\omega t_2 & \vdots \\ \vdots & \sin\omega t_N \end{pmatrix} \tag{8}$$

Then, $\mathrm{F}^T\mathrm{F}$ will be:

$$\mathrm{F}^T F = \begin{pmatrix} \mathrm{CC} & \mathrm{CS} \\ \mathrm{CS} & \mathrm{SS} \end{pmatrix}$$

And the normal equations will become:

$$\left(\mathrm{F}^T\mathrm{F}\right)\vec{\beta} = \begin{pmatrix} \mathrm{CC} & \mathrm{CS} \\ \mathrm{CS} & \mathrm{SS} \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \mathrm{F}^T\vec{y} = \begin{pmatrix} \mathrm{CY} \\ \mathrm{SY} \end{pmatrix} \tag{9}$$

Where we have called:

$$\mathrm{SY} = \sum_i y_i \sin\omega t_i \qquad \mathrm{CC} = \sum_i \cos^2\omega t_i \qquad \mathrm{CS} = \sum_i \cos\omega t_i \sin\omega t_i$$
$$\mathrm{CY} = \sum_i y_i \cos\omega t_i \qquad \mathrm{SS} = \sum_i \sin^2\omega t_i \qquad \mathrm{YY} = \sum_i y_i^2$$

And the sum of squared differences will be:

$$S = \vec{y}^T\vec{y} - \vec{y}^T\mathrm{F}\left(\mathrm{F}^T\mathrm{F}\right)^{-1}\mathrm{F}^T\vec{y} \tag{10}$$

Explicitly:

$$S = \mathrm{YY} - \begin{pmatrix} \mathrm{CY} & \mathrm{SY} \end{pmatrix}\frac{1}{\mathrm{CC}\cdot\mathrm{SS} - \mathrm{CS}^2}\begin{pmatrix} \mathrm{SS} & -\mathrm{CS} \\ -\mathrm{CS} & \mathrm{CC} \end{pmatrix}\begin{pmatrix} \mathrm{CY} \\ \mathrm{SY} \end{pmatrix} \tag{11}$$

It would be advantageous if $\mathrm{CS} = 0$. For that, we can introduce a time delay τ in our model, so that:

$$y_k + \epsilon_k = a\cos\omega(t-\tau) + b\sin\omega(t-\tau)$$

We must update our definitions CC, etc. to include this parameter. We can find τ using the condition $\mathrm{CS} = 0$ and some trigonometric identities:

$$\mathrm{CS} = 0 = \sum_i \cos\omega(t-\tau)\sin\omega(t-\tau)$$

$$0 = \sum_i \left[\left(\cos\omega t_i \cos\omega\tau + \sin\omega t_i \sin\omega\tau\right) \left(\sin\omega t_i \cos\omega\tau - \cos\omega t_i \sin\omega\tau\right) \right]$$

$$0 = \sum_i \left[\cos\omega t_i \sin\omega t_i \cos^2\omega\tau - \cos^2\omega t_i \cos\omega\tau \sin\omega\tau + \sin^2\omega t_i \sin\omega\tau \cos\omega\tau - \sin\omega t_i \cos\omega t_i \sin^2\omega\tau \right]$$

Separating the terms with τ and the ones with ωt we get:

$$\left(\cos^2\omega\tau - \sin^2\omega\tau\right) \sum_i \cos\omega t_i \sin\omega t_i = \cos\omega\tau \sin\omega\tau \sum_i \left(\cos^2\omega t_i - \sin^2\omega t_i\right)$$

Using the identities ($\cos 2x = \cos^2 x - \sin^2 x$) and ($\sin x \cos x = \frac{1}{2}\sin 2x$), we have, finally:

$$\boxed{\tan 2\omega\tau = \frac{\sum_i \sin 2\omega t_i}{\sum_i \cos 2\omega t_i}} \tag{12}$$

Where τ depends on the frequency. Now that we can find τ, let's get back to the sum of squared differences. As $\mathrm{CS} = 0$, now:

$$S = \mathrm{YY} - \begin{pmatrix} \mathrm{CY} & \mathrm{SY} \end{pmatrix} \begin{pmatrix} \frac{1}{\mathrm{CC}} & 0 \\ 0 & \frac{1}{\mathrm{SS}} \end{pmatrix} \begin{pmatrix} \mathrm{CY} \\ \mathrm{SY} \end{pmatrix}$$

Then:

$$S = \mathrm{YY} - \left[\frac{\mathrm{CY}^2}{\mathrm{CC}} + \frac{\mathrm{SY}^2}{SS} \right] \tag{13}$$

We can define a normalized spectral function using the last term in the above equation:

$$P(\omega) = \frac{1}{\mathrm{YY}} \left(\frac{\mathrm{CY}^2}{\mathrm{CC}} + \frac{\mathrm{SY}^2}{SS} \right) \tag{14}$$

In this case, $0 \le P(\omega) \le 1$.

3 Three-dimensional periodogram

We are digitizing signals at different points of space, however. The signals can be modeled as [5]:

$$A(\rho,\theta,\phi,t) = \sum_{n,m} A_{nm}(\rho) \exp\left[-i\left(m\theta - n\phi - \omega t\right)\right] \tag{15}$$

Where the angles θ,ϕ are in the boozer coordinate system, and the mode numbers $n,m \in \mathbb{Z}$. The assumption of zero mean still holds.

In this case, we can apply the periodogram technique, as the basis functions are the same. However, the phase argument changes, and has a spatial dependency. This forces us to change the parameter τ, as it will have a dependency in θ,ϕ. The easiest way to go about this is to simply decouple it from the frequency and subtract it as a global phase:

$$F = \begin{pmatrix} \cos\left(\omega t_1 - m\theta_1 + n\phi_1 - \tau\right) & \sin\left(\omega t_1 - m\theta_1 + n\phi_1 - \tau\right) \\ \cos\left(\omega t_2 - m\theta_1 + n\phi_1 - \tau\right) & \sin\left(\omega t_2 - m\theta_1 + n\phi_1 - \tau\right) \\ \vdots & \vdots \\ \cos\left(\omega t_1 - m\theta_2 + n\phi_2 - \tau\right) & \sin\left(\omega t_1 - m\theta_2 + n\phi_2 - \tau\right) \\ \cos\left(\omega t_2 - m\theta_2 + n\phi_2 - \tau\right) & \sin\left(\omega t_2 - m\theta_2 + n\phi_2 - \tau\right) \\ \vdots & \vdots \end{pmatrix}$$

Then,

$$\mathrm{CS} = 0 = \sum_{ij} \cos\left(\omega t_i - m\theta_j + n\phi_j - \tau\right) \cos\left(\omega t_i - m\theta_j + n\phi_j - \tau\right)$$

Where i is the time label and j is the coil label. We can rename:

$$\boxed{\alpha_{ij} \equiv \alpha_{\omega,m,n}(t_i, \theta_j, \phi_j) = \omega t_i - m\theta_j + n\phi_j} \tag{16}$$

Then, we have:

$$\mathrm{CS} = 0 = \sum_{ij} \cos\left(\alpha_{ij} - \tau\right) \sin\left(\alpha_{ij} - \tau\right)$$

And, following the same steps as in the one-dimensional case, we can find τ:

$$\tan 2\tau = \frac{\sum_{ij} \sin 2\alpha_{ij}}{\sum_{ij} \cos 2\alpha_{ij}} \tag{17}$$

Again, this τ depends on the frequencies and mode numbers: $\tau = \tau(\omega, m, n)$.

Then, the periodogram will have the same expression as in the one dimensional case, with the phase ωt substituted by α. Also, each data point is now labeled by the time index i and the coil index j.

$$P(\omega, n, m) = \frac{1}{YY} \left(\frac{\left[\sum_{ij} y_{ij} \cos\left(\alpha_{ij} - \tau\right)\right]^2}{\sum_{ij} \cos^2(\alpha_{ij} - \tau)} + \frac{\left[\sum_{ij} y_{ij} \sin\left(\alpha_{ij} - \tau\right)\right]^2}{\sum_{ij} \sin^2(\alpha_{ij} - \tau)} \right) \tag{18}$$

Where YY is the same as before.

3.1 Using the periodogram

The computation of the periodogram is relatively expensive, because we must sweep all the mode numbers. However, it also is moderately robust to frequency displacements, because for the typical signal length the peaks it produces are fairly broad (the width of the peaks should decrease with signal length, but I have to make sure of it). Because of that, the best approach, in my opinion, is to find the frequency (frequencies) of the modes via a Fourier transform and then calculate the periodogram for this frequency and all mode numbers.

4 Frequency-varying signals

What if the frequency of our signal is rapidly-varying on time? In this case, the periodogram loses its utility, as the advantage it presented, namely the phase τ that simplified the calculations, cannot be applied anymore. However, the least-squares fit remains valid, if now for a time-dependent frequency, that must be provided. This time-dependence is relatively easy to provide for most signals: we need just to trace the spectrogram and get a vector $\omega(t)$ the same size as t.

Then, from equations 5 and 6, we have:

$$S = y^T y - y^T F \left(F^T F\right)^{-1} F^T y \tag{19}$$

TODO

References

[1] Fredericus JM Barning. “The numerical analysis of the light-curve of 12 Lacertae”. In: *Bulletin of the Astronomical Institutes of the Netherlands* 17 (1963), p. 22.

[2] N. R. Lomb. “Least-squares frequency analysis of unequally spaced data”. In: *Astrophysics and Space Science* 39.2 (1976), pp. 447–462. DOI: 10.1007/BF00648343.

[3] J. D. Scargle. “Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data.” In: *The Astrophysical Journal* 263 (1982), pp. 835–853. DOI: 10.1086/160554.

[4] Petr Vaníček. “Further development and properties of the spectral analysis by least-squares”. In: *Astrophysics and Space Science* 12.1 (1971), pp. 10–33. DOI: 10.1007/BF00656134.

[5] Stefan Zegenhagen et al. “Analysis of Alfvén eigenmodes in stellarators using non-evenly spaced probes”. In: *Plasma Physics and Controlled Fusion* 48.9 (2006), pp. 1333–1346. DOI: 10.1088/0741-3335/48/9/005.