Spectral Timing Analysis Summary Notes

As a Handbok for the Equations used in the Statistical Package in Python

Ludvig Doeser – doeser@kth.se

January 7, 2022

Important Note: *These summary notes are not meant to be plagiarism-free; multiple copy-pasted sentences from cited articles are included for highest clarity. To align with notational convention, many equations are also copy-pasted; however, not all articles follow the same convention, so where confusion easily could arise, the notation has been altered (when this is the case, it is clearly stated). Lastly, where typos have been found in articles,* ***wrongly*** *has been written prior to the expression.*

Contents

1 Spectral Content of one Light Curve

A light curve is a graph that displays the count rate x, i.e. the number of photons received per time from an astronomical object, versus time t:

$$x_k \equiv x(t = k) \quad \text{for} \quad k = 1, \ldots, N \tag{1}$$

To analyze the count rate variability one often uses the `Discrete Fourier Transformations` (DFT):

$$X_j \equiv \sum_{k=0}^{N-1} x_k \exp(2\pi ijk/N) \quad , \quad j = [-N/2, \ldots, 0, \ldots, N/2] \tag{2}$$

The **power spectrum** $P(f)$, where f is frequency[1], can then be obtained by normalizing the Fourier coefficients according to

$$P\left(f_j\right) = A\left|X\left(f_j\right)\right|^2, \tag{3}$$

where A is a normalization coefficient, see section 1.1. This $P(f)$ is often called the **PSD** (power spectral density) or the **periodogram**, as it is actually only an estimation of the true power spectrum. For a better estimation, an observation is often split into several shorter so-called **time segments** consisting of m number of bins (each of length ΔT), so that $T_{\text{seg}} = m\Delta T$. After having made a power spectrum for each segment one averages[2] over the $K = \texttt{int}(N/m)$ segments to get the final power spectra

$$P(f_j) = \frac{1}{K} \sum_{i}^{K} P_{\text{segment}}(f_j). \tag{4}$$

Note that the segment's power spectrum is normalized (see section 1.1) and noise-subtracted (see section 1.2), using information only from the segment, before being averaged. The error in the power spectrum can be taken as the standard error

$$\delta P(f_j) = \frac{\sigma_P(f_j)}{\sqrt{K}}, \tag{5}$$

where σ_P is the standard deviation over all time segments' power spectra. The error tells us how far away the average $P(f_j)$ (the periodogram) is from the true power spectrum.

If the light curve has a sampling period ΔT, the power spectrum will exist for $m/2$ evenly spaced frequencies $f_j = j/(m\Delta T)$, where $j = 1, 2, \ldots, m/2$, i.e. with frequency resolution $\Delta f = 1/N\Delta T$ and min/max given by

$$f \in [f_1, f_{m/2}] = [1/m\Delta T, 1/2\Delta T = f_{\text{Nyq}}]. \tag{6}$$

Note that negative frequencies (the other $m/2$ of the frequencies) that pop up from the DFT computation are not relevant for us. To cover low frequencies we need to set m high, i.e. partition the original light curve into long segments. This will, however, make it harder to extract the high frequencies, for which we need m low, i.e. short segments. The highest frequency, also called the Nyquist frequency, is given solely by the time resolution.

1.1 Normalizations

There are several possible normalizations A, out of which the *Miyamoto*-normalization is the normalization most often used in analysis of AGN and X-ray binaries because the integrated

[1] Note that this is not the frequency of light in Planck's relation $E = hf = hc/\lambda$, but the frequency of the variability of the light curve. To not get confused, we never talk about the frequency of light itself; instead, we talk about its energy E.

[2] This is known as Bartlett's Method and it minimizes the effect of counting statistics. Note also that we need equally long segments to take the average, so unless N is a multiple of m we will not use all N data points.

periodogram yields the fractional variance[3] of the data. In Revnivtsev et al. (1999, 2000), they claim (**wrongly**[4]) that this normalization is

$$A_{\rm Rev} = \frac{2}{RN_\gamma} = \frac{2}{R^2T}, \tag{7}$$

where N_γ = total number of counts in a time segment of length T, and R = average count rate in the segment. Hence, $R = N_\gamma/T \iff N_\gamma = TR$. The factor 2 is included to get a one-sided RMS normalization, meaning that integration over all positive frequencies yield the full variance. According to Vaughan and Nowak (1997); Uttley et al. (2014), the Miyamoto-normalization looks like

$$A = \frac{2\Delta T}{R^2 N} = \frac{2(\Delta T)^2}{R^2 T} \neq A_{\rm Rev}, \tag{8}$$

where $\Delta T = T/N \iff N = T/\Delta T$, where N is the number of bins in the light curve. **This yeilds correct fractional rms** (see section 1.3), meaning that Revnivtsev et al. (1999) probably just forgot the $(\Delta T)^2$-factor.

Note that in the case that we split the light curve into segments, we use m instead of N in Eq. (8), and the mean count rate in the segment $R_{\rm seg}$ instead of R.

1.1.1 Other Normalizations

The other normalizations are related to the Miyamoto-normalization (also denoted $A_{\rm rms^2}$):

$$R^2 A_{\rm rms^2} = R A_{\rm Leahy} = A_{\rm abs}. \tag{9}$$

Depending on what normalization one chooses, the noise power (see section 1.2) will differ.

1.2 Poisson Noise

The total power $P_j = P(f_j)$ received is partly due to the signal $P_{\rm signal}$ from the object, partly due to noise $P_{\rm noise}$, which is a consequence of observing discrete counts (photons), i.e.

$$P_j = P_{\rm signal} + P_{\rm noise}. \tag{10}$$

If a light curve consists of a binned photon counting signal (and in the absence of other effects such as detector dead-time) the expected Poisson noise 'background' level in the case of a *Miyamoto*-normalization is simply (see Appendix to Vaughan et al. (2003)) given by

$$P_{\rm noise} = \frac{2(R+B)}{R^2}\frac{\Delta T_{\rm samp}}{\Delta T_{\rm bin}} \approx \frac{2}{R}, \tag{11}$$

where R is the mean source count rate, B is the mean background count rate, $\Delta T_{\rm samp}$ is the sampling interval and $\Delta T_{\rm bin}$ is the time bin width. The factor of $\Delta T_{\rm samp}/\Delta T_{\rm bin}$ accounts for aliasing of the Poisson noise level if the original photon counting signal contained gaps. Often $\Delta T_{\rm samp} = \Delta T_{\rm bin}$ and $B \approx 0$ so that the last equality holds. Hence, the signal power is given by

$$P_{\rm signal} = P_j - \frac{2}{R} \tag{12}$$

For other normalizations, we have

$$P_{\rm noise} = \begin{cases} 2 & \text{in the case of } \textit{Leahy}\text{-normalization} \\ 2R & \text{in the case of } \textit{abs}\text{-normalization.} \end{cases} \tag{13}$$

[3] In the *Miyamoto*-normalization, the units for the periodogram is $(\text{rms/mean})^2$ Hz^{-1} (where rms/mean is a dimensionless quantity), or simply Hz^{-1}. See more in section 1.3.

[4] A $(\Delta T)^2$-factor is missing in the nominator.

When I tried to apply the noise correction given by Eq. (11), however, there seemed to be noise power missing. For Cygnus X-1 data between 30/03/96, 19:54:52, until 31/03/96, 01:41:00, a $B \approx 0.1$ was needed, and for Cygnus X-1 data 23/10/96 (this is the same data as in Nowak et al. (1999)), 18:32:32, until 24/10/96, 02:30:28, a $B \approx 0.013$. I then found a review by Uttley et al. (2014), who clamied that if fluxes are not given in count rates (or the data are expressed as a count rate but the statistics are not[5] Poissonian), then the equivalent noise level (in the *Miyamoto*-normalization) can be determined using bars of the light curve and the Nyquist frequency

$$P_{\rm noise} = \frac{\langle \Delta x^2 \rangle}{R^2 f_{\rm Nyq}}, \tag{14}$$

where $\langle \Delta x^2 \rangle$ is the average of the squared error. I am not sure how this is derived, but this noise power is expected in the case of Gaussian statistics. For both aforementioned cases, Eq. (14) give the expected result (as given by HEASOFT's `PowSpec.f`). By comparing other statistics (that make use of the power spectra) with results from articles for the same data, Eq. (14) was also confirmed.

1.3 Fractional Variance Estimations

For a discrete time series the integrated periodogram, if *Miyamoto*-normalized (since $P(f)$ then has the units $(\text{rms/mean})^2\text{Hz}^{-1}$), yields the fractional variance (also called **rms**[6]) $F_{\rm var}$ squared for that particular realisation (see Vaughan et al. (2003); Wilkinson and Uttley (2009)),

$$F_{\rm var}^2 = \sum_{j=1}^{N/2} P\left(f_j\right) \Delta f, \tag{15}$$

where $P(f_j)$ is the power after the noise power has been subtracted at frequency bin j. The rms can also be defined directly by only looking at the light curve,

$$F_{\rm var}^2 = \frac{\sigma_{XS}^2}{R^2}, \tag{16}$$

where σ_{XS}^2 is the so-called excess variance

$$\sigma_{XS}^2 = S^2 - \overline{\sigma_{\rm err}^2}, \tag{17}$$

where, in turn, S^2 is the variance of the light curve, given by

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - R\right)^2, \tag{18}$$

where x_i is the count rate in the ith bin and R is the mean count rate. The expectation of noise variance, given by the average squared-error $\overline{\sigma_{\rm err}^2}$, is subtracted, leaving the 'excess variance', σ_{XS}^2. The excess variance can be computed separately for K segments before being averaged over all segments of the light curve. In the limit of large N these two rms-estimates, Eq. (15) and Eq. (16), are identical.

To estimate the error we can propagate the error in the power spectrum, i.e.,

$$\text{err}(F_{\rm var}) = \frac{1}{2F_{\rm var}} \Delta f \sqrt{\sum_j (\delta P(f_j))^2}, \tag{19}$$

[5] It might be worth mentioning that the Gaussian distribution is the limit of the Poisson distribution in the case of a large number of events N.

[6] The fractional variance also goes under the name fractional root mean square (or even shorter, just **rms**) variability amplitude and has the units [rms/mean].

where $\delta P(f_j)$ is the standard error $\sigma/\sqrt{K}$ over all K segments' $P(f_j)$-values for frequency bin j. There are, however, several other ways to estimate the error, all of which yield the same order of magnitude (it happens, however, that they differ with a factor > 2):

1. Using a Monte Carlo approach: For each frequency bin j, we sample X times from a normal distribution with mean given by the power spectrum average and the standard deviation given by the power spectrum standard error $\delta P(f_j)$. In effect, we then have X new power spectra, for which we compute the rms. The average rms of all these should be the same as the rms computed above and the error is the standard deviation over all these rms.

2. In Vaughan et al. (2003) as well as in Wilkinson (2011), the error on the rms (Eq. (B2)) is computed as

$$\mathrm{err}(F_{\mathrm{var}}) = \sqrt{\left(\sqrt{\frac{1}{2N}}\frac{\overline{\sigma_{\mathrm{err}}^2}}{\bar{x}^2 F_{\mathrm{var}}}\right)^2 + \left(\sqrt{\frac{\overline{\sigma_{\mathrm{err}}^2}}{N}}\frac{1}{\bar{x}}\right)^2}. \tag{20}$$

2. In Uttley et al. (2014): The errors on the rms spectrum must be calculated based on the errors expected due to Poisson noise (e.g. see Vaughan et al. 2003a), since they are independent between different energy bins in the rms spectrum, whereas the errors in the rms due to intrinsic stochastic variability (i.e. the intrinsic χ^2_2 scatter in the PSD) are correlated between energy bins if the coherence between bins is non-zero. In the limit where the intrinsic coherence is unity, the errors on the (absolute) rms-spectrum are given by:

$$\Delta\sigma_X(\nu_j) = \sqrt{\frac{2\sigma_X^2(\nu_j)\,\sigma_{X,\,\mathrm{noise}}^2 + \left(\sigma_{X,\,\mathrm{noise}}^2\right)^2}{2KM\sigma_X^2(\nu_j)}}, \tag{21}$$

where $\sigma_X(\nu_j)$ is the noise-subtracted rms in the channel-of-interest x and $\sigma_{X,\,\mathrm{noise}}^2$ is the absolute rms-squared value obtained from integrating under the Poisson noise level of the bands, i.e. $\sigma_X^2(\nu_j) = (P_X(\nu_j) - P_{X,\,\mathrm{noise}})\langle x\rangle^2\Delta\nu_j$, $\sigma_{X,\,\mathrm{noise}}^2 = P_{X,\,\mathrm{noise}}\langle x\rangle^2\Delta\nu_j$, where the PSDs are in the fractional rms-squared normalisation. See Vaughan et al. (2003a) for discussion of the rms-spectrum and its error (determined from numerical simulations), and Wilkinson (2011) for a formal derivation of the error.

1.4 RMS–Energy

Here you take a number of light curves in different energy bands. For each light curve you make a power spectrum, and *Miyamoto*-normalize it so that the integral gives rms. You then integrate each power spectrum to obtain rms for the corresponding energy, and plot it in a graph. Sometimes, one takes the square root of the average excess variance plotted against energy, but note that this is just the rms in *absolute*[7] units, i.e. the rms scaled with the mean count rate R (see Eq. (16)).

1.4.1 In a Smaller Frequency Range

Both Eq. (15) and Eq. (16) computes the full rms, i.e. using all frequencies. What if we only want to use a smaller frequency range? In this case, we simply integrate the power spectrum over the desired frequency range.

In case we are able to change the bin size of our data, there is another way to handle a smaller frequency range. By selecting the time bin size and the segment size, we can isolate different time-scales of variability, effectively replicating the Fourier-resolved approach for the two time-scale ranges that we are interested in. For this purpose, we choose two combinations of bin size

[7] In general, fractional/normalized units are often used in preference to absolute units as they are independent of the mean count rate of a specific source. This means that, in principle, normalized amplitudes can be used to compare sources with different fluxes (Vaughan et al., 2003)

and segment size. To look at variations on shorter time-scales we choose 0.1s time bins measured in segments of 4 s (i.e. 40 bins long), i.e. covering the frequency range $0.25 - 5$ Hz. For longer time-scale regions we use 2.7-s bins in segments of 270s, i.e. covering the range $0.0037 - 0.185$ Hz. Note that these two frequency ranges do not overlap and also cover the two parts of both source PSDs which show distinct behaviour in soft and hard bands (Wilkinson and Uttley, 2009; Wilkinson, 2011).

If we want the smallest frequency range possible, we can do this for every frequency-bin, as in Revnivtsev et al. (1999): For every obtained frequency bin the frequency dependent spectra were constructed according to the formula

$$\mathrm{rms}\left(E_i, f_j\right) = R_i \sqrt{P_i\left(f_j\right) \Delta f_j} = \sqrt{\frac{2\left|X_{ij}\right|^2}{T} \Delta f_j}. \tag{22}$$

Note that $P_i\left(f_j\right) \Delta f_j$ is essentially an integration yielding rms in that freqency bin, and that the multiplication with R then yields the excess variance (i.e. the absolute rms). We can then plot rms vs energy.

1.5 Frequency Resolved Spectroscopy (= RMS Spectrum)

This is a continuation of RMS-Energy and involves using the energy spectrum. Having already computed rms for a given energy band and frequency range, we now simply multiply the rms for a given channel[8] by the value of the energy spectra (in counts/s/keV) in that channel. By doing this for multiple energy and frequency bands, we can obtain a frequency resolved spectrum, as discussed in section 3.2 of Wilkinson and Uttley (2009) and shown in Fig 1 of (Revnivtsev et al., 1999). See also Axelsson and Done (2018) for an example how FRS can be used. In summary, we need to do:

1. Calculate rms in % for each energy band over a given frequency range.
2. List which channels each energy band corresponds to.
3. Create a list indicating % for each channel.
4. Take the total energy spectrum and multiply each channel by the respective percentage.

[8] So first we need to convert the energy bands to the corresponding channels.

2 Spectral Content of Two Simultaneous Light Curves

Another class of spectral analysis techniques measures the correlated variability between two simultaneous light curves. The four main (and related) statistics that we will discuss here are:

- Cross Spectrum
- Covariance
- Coherence
- Time-Lag

2.1 Cross-spectrum

The power spectrum can be called the autospectrum since it is the square of the Fourier transform of the count rate (see Eq. (3)). The **cross-spectrum**, on the other hand, compares two simultaneous light curves, e.g. from different energy bands, say $x(t)$ and $y(t)$. The product of the Fourier transforms of the two light curves gives the cross-spectrum:

$$C(f) = X^*(f)Y(f) = |X(f)||Y(f)|\mathrm{e}^{\mathrm{i}[\phi_y(f)-\phi_x(f)]} \equiv S_1^*(f)S_2(f), \tag{23}$$

where $*$ here stands for the complex conjugate. In the last step we let S_i denote the Fourier transformations as this is common notation for the full signal. Note that no noise subtraction or normalization is used.

2.2 Coherence

Another tool for measuring variability is **coherence**, $\gamma^2(f)$, which is a Fourier-frequency–dependent measure of the degree of linear correlation between two concurrent time series. Specifically, it gives the fraction of the mean-squared variability at f of one time series that can be attributed to, or equivalently predicted from, the other (Nowak et al., 1999). The coherence is a real-valued function that is computed as the squared cross-spectrum normalized by the autospectra (PSDs) of the two light curves,

$$\gamma^2(f) = \frac{|\langle C(f)\rangle|^2}{\langle |S_1(f)|^2\rangle \langle |S_2(f)|^2\rangle}. \tag{24}$$

The angled brackets $\langle\rangle$ here represent an averaging over light-curve segments, i.e. using $m \ll N$, and/or consecutive frequencies. Note the different orders on averaging and taking absolute value: $|\langle\ \rangle|$ vs $\langle|\ |\rangle$. Note also that the full signals are used, that is without subtracting any noise power; this estimation of the coherence is, thus, most relevant for noiseless signals. As presented in Nowak et al. (1999), the uncertainty on the coherence is in the case of Gaussian statistics and noiseless signals given by

$$\delta\gamma^2(f) = \frac{\sqrt{2}\left[1-\gamma^2(f)\right]}{|\gamma(f)|\sqrt{K}}. \tag{25}$$

Note that Vaughan and Nowak (1997) **wrongly** have an extra factor $\gamma^2(f)$ in the nominator.

2.3 Intrinsic Coherence for Noisy Signals

A strong noise component in the light curves may lead to a loss of coherence (as random noise signals are not correlated). To further investigate this possibility, we can turn to the intrinsic coherence (Eq. (8) from Vaughan and Nowak (1997)), which takes into account noisy signals by applying a correction term to the measured coherence.

Since the notation vary from article to article, which can be confusing, we will look at different presentations[9] of the intrinsic coherence. We will focus on the *High powers, high measured coherence*–case presented in Vaughan and Nowak (1997), in which we need that the noise-subtracted power (e.g. $|S|^2$ in Vaughan and Nowak (1997)) is greater than a few times the noise power divided by the number of segments[10] ($|N|^2/\sqrt{K}$) in each channel (light curve), and that the measured coherence satisfies γ^2 greater than a few times $n^2/(P_1P_2)$, where P_i is the full signal power and n^2 is the noise-correction term to be introduced. Let's first note that when we calculate the coherence we choose $A = 1$ in Eq. (3), i.e. no normalization. Effectively, we therefore need to alter the noise power with the same factor. In the case of Poisson noise[11], we have

$$A = \frac{2\Delta T}{R^2 N} \mapsto 1 \iff P_{\text{noise}} = \frac{2}{R} \mapsto \frac{RN}{\Delta T} \tag{26}$$

In the case of Gaussian noise, we have

$$P_{\text{noise}} = \frac{\langle \Delta x^2 \rangle}{R^2 f_{\text{Nyq}}} = \frac{2\Delta T \langle \Delta x^2 \rangle}{R^2} \mapsto N \langle \Delta x^2 \rangle . \tag{27}$$

2.3.1 Vaughan and Nowak (1997)

We now have a look at the intrinsic coherence as presented by Vaughan and Nowak (1997). With noise present, the noiseless coherence Eq. (24) needs to be replaced by

$$\gamma^2_{int} = \frac{|\langle C \rangle|^2 - n^2}{|S_1|^2 |S_2|^2}, \tag{28}$$

where subscript I denotes the intrinsic coherence, and the noise correction term n^2 is (**wrongly**[12]) given by

$$n^2 \equiv \frac{1}{K} \left[|S_1|^2 |N_2|^2 + |N_1|^2 |S_2|^2 + |N_1|^2 |N_2|^2 \right], \tag{29}$$

where, as already mentioned, m is used instead of K in the article. The error can be estimated as

$$\delta\gamma^2_{int} = \frac{\gamma^2_{int}}{\sqrt{K}} \left[\frac{2n^4 K}{\left(|\langle C \rangle|^2 - n^2 \right)^2} + \frac{|N_1|^4}{|S_1|^4} + \frac{|N_2|^4}{|S_2|^4} + \frac{K \delta\gamma^2}{\gamma_I^4} \right]^{1/2} . \tag{30}$$

It is, however, not clear how C ought to be computed. It is discussed in Vaughan and Nowak (1997) that since the full signal power is given by the sum of the noise-free signal power and noise power,

$$P_i = |S_i|^2 + |N_i|^2, \tag{31}$$

C is now given by

$$|\langle C \rangle|^2 = |\langle S_1^* S_2 \rangle + \langle S_1^* N_2 \rangle + \langle N_1^* S_2 \rangle + \langle N_1^* N_2 \rangle|^2 . \tag{32}$$

They then claim that the three last terms average to a mean of zero with a variance term that is equal to $n^2/2$. My interpretation of introducing Eq. (32) is thus to motivate the noise correction term, all the while still computing $|\langle C \rangle|^2$ in the same way as before (using the full signal).

[9] Best articles: 1)Nowak et al. (1999): some errors, e.g. a + instead of − in the last term in the expression for n^2 and not so clear how to compute $C(f)$. 2) Vaughan and Nowak (1997): brief discussion, but with nice examples. 3) Epitropakis and Papadakis (2017): correct equations, and here it seems like $C(f)$ should be calculated without noise subtraction. 4) Uttley et al. (2014): perform noise subtraction in strange order.

[10] In both articles, they use m instead of K for the number of independent samples, i.e. the number of segments. Different notations everywhere, so be aware...

[11] We start from the expressions in the *Miyamoto*-normalization, but this is an arbitrary choice since we scale both A and P_{noise}, making the normalization irrelevant.

[12] The last sign in the nominator should be a minus.

2.3.2 Epitropakis (2017)

In (Epitropakis and Papadakis, 2017), Eq. (24) is presented as

$$\hat{\gamma}^2_{xy}\left(v_p\right) \equiv \frac{\left|\hat{C}_{xy}\left(v_p\right)\right|^2}{\hat{P}_x\left(v_p\right)\hat{P}_y\left(v_p\right)}, \tag{33}$$

where $\hat{C}_{xy}$ was obtained by averaging the K individual cross-periodograms (once again, m is used instead of K in the article) at each frequency. With noise present, this changes to

$$\hat{\gamma}^2_{\mathrm{int},xy}\left(v_p\right) = \frac{\left|\hat{C}_{xy}\left(v_p\right)\right|^2 - \left|\hat{\zeta}\left(v_p\right)\right|^2}{\left[\hat{P}_x\left(v_p\right) - P_{\epsilon_x}\right]\left[\hat{P}_y\left(v_p\right) - P_{\epsilon_y}\right]} \tag{34}$$

where v_p here is the frequency, $\hat{}$ denotes the average over K segments, P_x (P_y) is the full signal power, P_{ϵ_x} (P_{ϵ_y}) is the noise power, and

$$\left|\hat{\zeta}\left(v_p\right)\right|^2 = \frac{1}{K}\left[\hat{P}_x\left(v_p\right)P_{\epsilon_y} + \hat{P}_y\left(v_p\right)P_{\epsilon_x} - P_{\epsilon_x}P_{\epsilon_y}\right]. \tag{35}$$

Here it is clearer that we ought to compute $\left|\hat{C}_{xy}\left(v_p\right)\right|^2$ as in the noiseless case. It is also clearer that we ought to use the noise-subtracted powers in the denominator (which can be confusing when comparing Eq. (24) (where $|S_i|^2 = P_i$) and Eq. (28) (where $|S_i|^2 = P_i - |N_i|^2$)). Note also the minus sign between the last two terms in Eq. (35), which wrongly is a plus in Eq. (29).

The error is given by

$$\hat{\sigma}_{\hat{\gamma}^2_{\mathrm{int},xy}}\left(v_p\right) = \frac{\hat{\gamma}^2_{\mathrm{int},xy}\left(v_p\right)}{\sqrt{K}}\left\{\frac{2\left|\hat{\zeta}\left(v_p\right)\right|^4 K}{\left[\left|\hat{C}_{xy}\left(v_p\right)\right|^2 - \left|\hat{\zeta}\left(v_p\right)\right|^2\right]^2} + \left[\frac{P_{\epsilon_x}}{\hat{P}_x\left(v_p\right) - P_{\epsilon_x}}\right]^2 + \left[\frac{P_{\epsilon_y}}{\hat{P}_y\left(v_p\right) - P_{\epsilon_y}}\right]^2 + K\left[\frac{\sigma_{\hat{\gamma}^2}\left(v_p\right)}{\hat{\gamma}^2_{xy}\left(v_p\right)}\right]^2\right\}^{1/2}. \tag{36}$$

This expression is identical with Eq. (30) up to the last term, where it seems that the coherence $\hat{\gamma}^2_{xy}$ (not intrinsic coherence) should be used, while Eq. (30) uses the intrinsic coherence. I haven't derived the error-expression, but it feels more likely that the intrinsic coherence would be used here. I haven't investigated the difference, but I have used $\hat{\gamma}^2_{\mathrm{int},xy}$.

2.3.3 Uttley (2014)

In terms of inconsistencies, I found one more. In Uttley et al. (2014) it is stated that

$$\gamma^2_I\left(\nu_j\right) = \frac{\left|\bar{C}_{XY}\left(\nu_j\right)\right|^2 - n^2}{\bar{P}_X\left(\nu_j\right)\bar{P}_Y\left(\nu_j\right)}, \tag{37}$$

and

$$n^2 = \left[\left(\bar{P}_X\left(\nu_j\right) - P_{X,\mathrm{noise}}\right)P_{Y,\mathrm{noise}} + \left(\bar{P}_Y\left(\nu_j\right) - P_{Y,\mathrm{noise}}\right)P_{X,\mathrm{noise}} + P_{X,\mathrm{noise}}P_{Y,\mathrm{noise}}\right]/KM, \tag{38}$$

where $K = N//2, M =$ number of bins per segment. That is, Uttley et al. (2014) subtract the noise power in differently (compare with Eq. (34) and (35)). They do not subtract the noise power from the power spectra in the denominator of γ^2_I but do subtract the noise power from the power spectra in the two first terms of n^2. I have not used these expression nor have I managed to show if these changes yields the same expression in the end.

2.3.4 Condtions to be met for Useful Estimation

The intrinsic coherence can be usefully estimated (see Vaughan and Nowak (1997)) when:

$$\begin{aligned} \left|\hat{C}_{xy}\right| &\geq n, \\ |S_1|^2/|N_1|^2 &\geq 1/\sqrt{K}, \\ |S_2|^2/|N_2|^2 &\geq 1/\sqrt{K}. \end{aligned} \tag{39}$$

Note that in order to display a coherence-figure as Fig. 5 of Nowak et al. (1999), we cannot just compute $\gamma^2_{\rm int}$ once. We normally need to use m high (i.e. many segments $= K$ low) to cover low frequencies and m low (K high) to cover high frequencies. The m-value determines not only the frequency resolution $1/(m\Delta T)$, but also highly influences the frequency range for which the coherence can be usefully estimated through Eq. (39).

Due to these conditions, each m yields an upper frequency limit $f_{\rm upper}$, which is the lowest frequency at which at least one condition is broken. I have therefore implemented a way to automatically change m, starting with a high m-value, which is then lowered each iteration. As a criteria I chose to demand that the next iteration's frequency resolution Δf is increased until it satisfies

$$q_1 \cdot \Delta f > f_{\rm upper}, \tag{40}$$

where q_1 is an arbitrary multiplicative factor (now set to 10) for conservative purposes. Since $\Delta f = 1/(m\Delta t)$ we require that

$$m = 2^k < \frac{q_1}{\Delta t \cdot f_{\rm upper}}, \tag{41}$$

where we want to find the largest k that satisfies the inequality each iteration. The lowest m that is acceptable is now set to 2^8. Yet another multiplicative factor, q_2, is used to combine these coherence-computations for different m; the used frequency range for each m will be $[\Delta f, q_2 \cdot f_{\rm upper}]$.

2.4 Time Lag

The phase of the cross-spectrum Eq. (23) (i.e. the argument of the complex cross-spectrum vector) $\phi(\nu_j) \equiv \arg[C(f)]$, gives the average phase lag between the two light curves at frequency f. The time lag is thus given by

$$\tau\left(f\right) = \frac{\phi}{2\pi f}. \tag{42}$$

Here, it is important to note that the standard domain for $\arg[C]$ is $(-\pi, \pi]$, so use e.g. `np.angle()`. Using the noiseless coherence, the error is simply given by[13] Epitropakis et al. (2016),

$$\Delta\phi\left(f\right) = \sqrt{\frac{1-\gamma^2\left(f\right)}{2\gamma^2\left(f\right)K}}, \tag{43}$$

where K is the number of segments and the time-lag error is simply $\Delta\tau = \Delta\phi/\left(2\pi f\right)$. Note that this is the same equation as Eq. (16) from Nowak et al. (1999).

There are two important notes:

1. The time lag is, like the coherence, computed using several m-values, as this is required to cover very low and very high frequencies.

[13] Note that Uttley et al. (2014) uses M as the number of segments and K as the number of frequencies within a frequency bin. In that article's Eq. (9) there is a factor $1/\sqrt{K}$, which comes from also performing an average over frequency bins. As in Epitropakis and Papadakis (2017), we don't average C over neighbouring frequencies as this can introduce a bias at low frequencies (Epitropakis et al., 2016).

2. Positive time lags do by convention usually mean that soft leads (hard lags). When computing $\tau = \texttt{np.angle}(C)/(2\cdot \texttt{np.pi}\cdot f)$ we get the opposite, so we need to use a minus sign later. This is what the `time_lag`-function returns. However, the `plot_time_lag`-function works as it should.

2.5 Covariance Spectrum

The covariance spectrum is the cross-spectral counterpart of the rms-spectrum (in the same way as the cross spectrum is the cross-spectral counterpart of the power spectrum). Remember that the rms spectrum displays the rms amplitude of variability as a detailed function of energy. In the same way, we can use the cross-spectrum to obtain a Fourier-frequency resolved covariance spectrum, which shows the spectral shape of the components which are correlated with the reference band. Not that multiplication with the energy spectra turns the **covariance** into a **covariance** spectrum.

This technique overcomes the problems of low signal-to-noise ratio and bias associated with the rms spectrum and has smaller statistical errors (Wilkinson and Uttley, 2009). The problem with the rms spectrum is when the signal-to-noise ratio is low, e.g. at higher energies. It is possible for the expectation value of the Poisson variance term to be larger than the measured average variance term, producing negative average excess variances (see Eq. (17)). If this is the case, it is not possible to calculate the rms at these energies and this introduces a bias towards the statistically higher-than-average realizations of rms values, which can still be recorded.

2.5.1 Wilkinson and Uttley (2009)

In order to overcome the problems related to the rms, Wilkinson and Uttley (2009) have developed a technique called the 'covariance spectrum'; the covariance is calculated as

$$\sigma^2_{\rm cov} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right), \tag{44}$$

where Y_i now refers to the light curve for a 'reference band' running over some energy range where the variability signal-to-noise ratio is large (i.e. usually a low-energy band).

The covariance spectrum does not suffer from the same problems as the rms spectrum, as no Poisson error term has to be subtracted, since uncorrelated noise tends to cancel out and any negative residuals do not affect the calculation. To remove the reference band component of the covariance, and produce a spectrum in count-rate units, we obtain the normalized covariance for each channel using

$$\sigma_{\rm cov,norm} = \frac{\sigma^2_{\rm cov}}{\sqrt{\sigma^2_{\rm xs,y}}}, \tag{45}$$

where $\sigma^2_{\rm xs,y}$ is the excess variance of the reference band[14]. **In short,** one can think of the covariance technique as applying a matched filter to the data, where the variations in the good signal-to-noise ratio reference band pick out much weaker correlated variations in the energy channel of interest that are buried in noise.

The error is computed as

$$\mathrm{Err}\left[\sigma_{\rm cov,norm}\right] = \sqrt{\frac{\sigma^2_{\rm xs,x}\overline{\sigma^2_{\rm err,\,y}} + \sigma^2_{\rm xs,y}\overline{\sigma^2_{\rm err,x}} + \overline{\sigma^2_{\rm err,x}\sigma^2_{\rm err,y}}}{NK\sigma^2_{\rm xs,y}}}, \tag{46}$$

[14]Therefore, the only requirement for there being a valid, unbiased value of covariance at a given energy is that the reference excess variance is not negative. This is usually the case, since the reference band is chosen to include those energies with the largest absolute variability.

where N is the total number of data points, K (M in article) is the number of segments, and subscripts x and y identify excess variances and Poisson variance terms for the channel of interest and reference band, respectively. The errors on the covariance are smaller than corresponding errors on the rms values.

When the raw counts rms and covariance spectra are overlaid, as in Fig. 2 (Wilkinson and Uttley, 2009), they match closely, indicating that the reference band is well correlated with all other energies (the spectral 'coherence' is high; e.g. see Vaughan & Nowak 1997).

2.5.2 Uttley (2014)

In (Uttley et al., 2014), the expression for the covariance in absolute units is

$$Cv\left(\nu_j\right) = R_x \sqrt{\frac{\Delta\nu_j \left(\left|\bar{C}_{XY}\left(\nu_j\right)\right|^2 - n^2 \right)}{\bar{P}_Y\left(\nu_j\right) - P_{Y,\,\mathrm{noise}}}}, \tag{47}$$

where ν_j is the j:th frequency bin. This means that the covariance spectrum can be computed directly from the coherence, since the above is equivalent to

$$\sigma^2_{\mathrm{cov}} \equiv Cv(\nu_j) = R_x \sqrt{\gamma^2_{\mathrm{int}}\left(\nu_j\right)\left(\bar{P}_X\left(\nu_j\right) - P_{X,\,\mathrm{noise}}\right)\Delta\nu_j}, \tag{48}$$

where we here should note that we in order to get the covariance in the frequency interval of interest need to see the bar in $\bar{P}_X\left(\nu_j\right)$ as the average of the normalized power spectra over the frequency interval of interest. We can, furthermore, note the resemblance with the rms. Suppose the intrinsic coherence is $= 1\ \forall\nu$, in which case we can use Eq. (15) and the *first mean value theorem for definite integrals*[15] to obtain the absolute rms,

$$\sigma^2_{\mathrm{cov}}(\gamma^2_{\mathrm{int}} = 1) = R_x \sqrt{\left(\bar{P}_X\left(\nu_j\right) - P_{X,\,\mathrm{noise}}\right)\Delta\nu_j} = R_x F_{\mathrm{var},x}. \tag{49}$$

So we see that the covariance spectrum is closely related to the coherence, and in the limit of unity coherence, the covariance spectrum should have the same shape as the rms spectrum. However, even in this case, the signal-to-noise of the covariance spectrum is substantially better than that of the rms spectrum, since the reference band light curve is effectively used as a 'matched filter' to pick out the correlated variations in each channel- of-interest (Uttley et al., 2014).

In Uttley et al. (2014), the error is given by

$$\Delta Cv\left(\nu_j\right) = \sqrt{\frac{\left[Cv\left(\nu_j\right)\right]^2 \sigma^2_{Y,\,\mathrm{noise}} + \sigma^2_Y\left(\nu_j\right)\sigma^2_{X,\,\mathrm{noise}} + \sigma^2_{X,\,\mathrm{noise}}\,\sigma^2_{Y,\,\mathrm{noise}}}{2KM\sigma^2_Y\left(\nu_j\right)}} \tag{50}$$

where $\sigma^2_Y\left(\nu_j\right)$ is the noise-subtracted absolute rms-squared of the reference band and $\sigma^2_{Y,\,\mathrm{noise}} = P_{Y,\,\mathrm{noise}}\,\langle y\rangle^2\Delta\nu_j$, where the PSDs are again in the fractional rms squared *Miyamoto*-normalisation. Note that $\left[Cv\left(\nu_j\right)\right]^2$ is used instead of the absolute rms-squared of the $\sigma_X\left(\nu_j\right)$ in this formula, because we assume unity intrinsic coherence, in which case $\left[Cv\left(\nu_j\right)\right]^2$ gives a significantly more accurate measure than the equivalent value obtained from the rms-spectrum. Note also that this equation is otherwise consistent with Eq. (46); the denominator $2KM$ in Eq. (50), where $K = N/2$ in Uttley et al. (2014) and M is number of segments, is equal to NK in Eq. (46).

2.5.3 In a Smaller Frequency Band

As for the rms, we can compute the covariance in a smaller frequency interval. Since the coherence usually only can be estimated up to an upper limit, we can use the coherence-approach, that is

[15] Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then there exists c in $[a, b]$ such that $\int_a^b f(x)dx = f(c)(b - a)$.

Eq. (48), to find the covariance in the frequency range [$f_{\rm min}$,`freq_high`], where `freq_high` can be chosen arbitrarily.

Another approach is presented in Gardner and Done (2014): The light curves are then Fourier transformed and the power set to zero for all frequencies other than the range of interest. An inverse Fourier transform is then applied to transform the filtered periodogram back into a light curve, now containing variability only in a narrow frequency range. The covariance between the filtered energy bin and reference band lightcurves is then calculated as in Wilkinson and Uttley (2009).

A Dead Time

A PSD constructed directly from detector counts consist of two components: the intrinsic noise from the source which is being studied, and the noise from counting statistics, which, in the presence of detector dead time, no longer obeys Poisson statistics. While effects of detector dead time on counting statistics had been recognized and extensively studied in the field of nuclear physics (see Evans 1955, p. 791 , for a historical summary, and see Müller 1973,1974 for a relatively recent update), it was not until the timing analyses of the EXOSAT data that their consequences were fully appreciated in the context of X-ray astrophysics (Zhang et al., 1995).

- Paralyzable: every incident event causes dead time of t_d, even if it is not detected. The dead time is cumulative and a contigous dead time segment can thus have length $t \in [t_d, \infty)$.
- Non-paralyzable: only a detected event causes the detector to be dead for a period of time t_d.

Paralyzable

The incident event rate, $r_{\rm in}$, is related to the detected event rate, r_0, through

$$r_0 = r_{\rm in} \exp(-r_{\rm in} t_d) = \frac{1}{\tau} \exp\left(-\frac{t_d}{\tau}\right), \tag{51}$$

with $\tau = 1/r_{\rm in}$ is the average time interval between two consecutive incident events.

Carrying out the cosine transform, we have the following power spectrum:

$$\langle P_{\rm noise}(f) \rangle = 2 - 4 \exp\left(-\frac{t_d}{\tau}\right) \frac{\sin(2\pi f t_d)}{2\pi f \tau}, \tag{52}$$

where we have multiplied the right-hand side by a factor of $2/r_o$ so as to adopt the Leahy normalization (Zhang et al., 1995). In the case of *Miyamoto*-normalization, we should have Eq.(52)$/R$.

B Error Propagation

Either when you need to subtract light curves or subtract spectra. Let f be the full light curve/spectra at hand being the sum of individual components:

$$f = A + B \rightarrow \sigma_f = \sqrt{\sigma_A^2 + \sigma_B^2}. \tag{53}$$

What we want to do is to subtract A to find σ_B:

$$f_2 = f - A \rightarrow \sigma_{f_2} = \sigma_B \tag{54}$$

To get it right, we just need to do:

$$\sigma_{f_2} = \sqrt{\sigma_B} = \sqrt{\sigma_f^2 - \sigma_A^2} \tag{55}$$

Bibliography

Axelsson, M. and Done, C. (2018). Breaking the spectral degeneracies in black hole binaries with fast timing data: The hard state of Cygnus X-1. *Monthly Notices of the Royal Astronomical Society*, 480(1):752–759.

Epitropakis, A. and Papadakis, I. E. (2017). The X-ray continuum time-lags and intrinsic coherence in AGN. *Monthly Notices of the Royal Astronomical Society*, 468(3):3568–3601.

Epitropakis, A., Papadakis, I. E., Dovčiak, M., Pecháček, T., Emmanoulopoulos, D., Karas, V., and McHardy, I. M. (2016). Theoretical modelling of the AGN iron line vs. Continuum time-lags in the lamp-post geometry. *Astronomy and Astrophysics*, 594:71.

Gardner, E. and Done, C. (2014). A physical model for the X-ray time lags of narrow-line Seyfert type 1 active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, 442(3):2456–2473.

Nowak, M. A., Vaughan, B. A., Wilms, J., Dove, J. B., and Begelman, M. C. (1999). Rossi X-Ray Timing Explorer Observation of Cygnus X-1. II. Timing Analysis. *The Astrophysical Journal*, 510(2):874–891.

Revnivtsev, M., Gilfanov, M., and Churazov, E. (1999). The frequency resolved spectroscopy of Cyg X-1: Fast variability of the Fe Kα line.

Revnivtsev, M., Gilfanov, M., and Churazov, E. (2000). High frequencies in the power spectrum of Cyg X-1 in the hard and soft spectral states. *Astronomy and Astrophysics*, 363(3):1013–1018.

Uttley, P., Cackett, E. M., Fabian, A. C., Kara, E., and Wilkins, D. R. (2014). X-ray reverberation around accreting black holes.

Vaughan, B. A. and Nowak, M. A. (1997). X-Ray Variability Coherence: How to Compute It, What It Means, and How It Constrains Models of GX 3394 and Cygnus X-1. *The Astrophysical Journal*, 474(1):L43–L46.

Vaughan, S., Edelson, R., Warwick, R. S., and Uttley, P. (2003). On characterizing the variability properties of X-ray light curves from active galaxies.

Wilkinson, T. (2011). Using novel spectral-timing techniques to probe the rapid variability of X-Ray Binaries.

Wilkinson, T. and Uttley, P. (2009). Accretion disc variability in the hard state of black hole X-ray binaries. *Monthly Notices of the Royal Astronomical Society*, 397(2):666–676.

Zhang, W., Jahoda, K., Swank, J. H., Morgan, E. H., and Giles, A. (1995). Dead-Time Modifications to Fast Fourier Transform Power Spectra. *The Astrophysical Journal*, 449:930–935.