

## A fast template periodogram

John Hoffman<sup>1,\*</sup>, Jake VanderPlas<sup>2,\*\*</sup>, Joel Hartman<sup>1,\*\*\*</sup>, and Gáspár Bakos<sup>1,\*\*\*\*</sup>

<sup>1</sup>*Department of Astrophysical Sciences, Princeton University, Princeton NJ 08540*

<sup>2</sup>*eScience Institute, University of Washington, Seattle, WA 98195*

**Abstract.** This proceedings contribution presents a novel, non-linear extension to the Lomb-Scargle periodogram that allows periodograms to be generated for arbitrary signal shapes. Such periodograms are already known as “template periodograms” or “periodic matched filters,” but current implementations are computationally inefficient. The “fast template periodogram” presented here improves existing techniques by a factor of  $\sim$ a few for small test cases ( $O(10)$  observations), and over three orders of magnitude for lightcurves containing  $O(10^4)$  observations. The fast template periodogram scales asymptotically as  $O(HN_f \log HN_f + H^4 N_f)$ , where  $H$  denotes the number of harmonics required to adequately approximate the template and  $N_f$  is the number of trial frequencies. Existing implementations scale as  $O(N_{\text{obs}} N_f)$ , where  $N_{\text{obs}}$  is the number of observations in the lightcurve. An open source Python implementation is available on GitHub.

### 1 Introduction

Template fitting, or periodic matched filtering, is a powerful technique for identifying, classifying, and characterizing periodic variability in noisy time series data. However, template fitting requires several orders of magnitude more computational resources than running a Lomb Scargle periodogram ([1–4]) on the same lightcurve, and the computational requirements scale as  $N_{\text{obs}} N_f$ , where  $N_{\text{obs}}$  is the number of observations in the lightcurve and  $N_f$  is the number of trial frequencies.

For example, [5] used template fitting to model RR Lyrae in the Pan-STARRS DR1 ([6]) and found that, for a given completeness, template fitting produced purer samples of RR Lyrae than Lomb-Scargle or multi-term extensions ([7]). However, [5] found that the computational resources needed to produce template fits (approximately 30 minutes per source per CPU) limited the number of lightcurves for which a full template fit could be performed to  $O(10^3)$  out of  $O(10^6)$  Pan-STARRS DR1 lightcurves.

Pan-STARRS lightcurves typically have less than 35 observations in 5 photometric filters ( $\sim 7$  per filter), but many photometric surveys have orders of magnitude more observations – CoRoT ([8]), SuperWASP ([9]), HATNet ([10]), and *Kepler* ([11]), for example, have  $O(10^4 - 10^5)$  observations per lightcurve. Since template fitting scales as  $O(N_{\text{obs}}^2)$ ,<sup>1</sup> running template periodograms on these surveys currently requires a prohibitively large amount of computational resources.

\*jah5@princeton.edu

\*\*jakevdp@cs.washington.edu

\*\*\*jhartman@astro.princeton.edu

\*\*\*\*gbakos@astro.princeton.edu

<sup>1</sup>The actual scaling is technically  $O(N_{\text{obs}} N_f)$ , however the number of trial frequencies  $N_f$  scales linearly with  $N_{\text{obs}}$  ([12]).

However, if template periodograms were computationally efficient enough to run on entire surveys with  $\mathcal{O}(10^3 - 10^5)$  observations per lightcurve, they would have the potential to be a powerful tool for automated classification. For example, by selecting a group of templates that capture the variety of periodic signals present in astrophysical data, a template periodogram could provide *simultaneous* detection and (pre-)classification of all lightcurves with higher signal to noise ratio than multi-harmonic periodograms, which are also designed to detect non-sinusoidal signals.

The advantage of templates is that there are only three free parameters (amplitude, phase, and a constant offset), compared to the multi-harmonic periodogram, which has  $2H + 1$  free parameters, where  $H$  is the number of harmonics. Using a collection of templates encodes “domain knowledge” (see, e.g., [13]) into the periodogram, which helps to reduce the parameter space to a set of physically plausible lightcurve shapes.

## 2 Extending Lomb-Scargle to arbitrary shapes

The Lomb-Scargle periodogram ([1–4]) is also known as least squares spectral analysis. The formulation of the periodogram given in [1] and [2] fits a linear model

$$\hat{y}(t|\theta, \omega) = \theta_1 \cos \omega t + \theta_2 \sin \omega t \quad (1)$$

to time series data  $D = \{t_n, y_n, \sigma_n | n < N\}$  at a number of trial frequencies  $\omega$ . The Lomb-Scargle periodogram,

$$P_{\text{LS}}(\omega) = \frac{1}{2\sigma^2} \left( \frac{\left[ \sum_{n=1}^N (y_n - \bar{y}) \cos \omega t_n \right]^2}{\sum_{n=1}^N \cos^2 \omega t_i} + \frac{\left[ \sum_{n=1}^N (y_n - \bar{y}) \sin \omega t_n \right]^2}{\sum_{n=1}^N \sin^2 \omega t_i} \right), \quad (2)$$

is similar to a Fourier intensity for unevenly spaced data. As shown in [2], the periodogram can be derived by minimizing the squared residuals and setting

$$P(\omega) = \frac{\chi_0^2 - \chi^2(\omega)}{\chi_0^2}, \quad (3)$$

where  $\chi^2(\omega)$  is the minimum (weighted) sum of squared residuals for a model fit at frequency  $\omega$ , and  $\chi_0^2$  is the weighted sum of squared residuals for a constant fit.

There are many extensions to the Lomb-Scargle periodogram that are useful for time domain astronomy. [14] used *weighted* squared residuals and added a constant offset to  $\hat{y}$ , [7, 15] further extended the Lomb-Scargle periodogram for models with more than one harmonic,

$$\hat{y}(t|\theta, \omega) = \theta_0 + \sum_{h=1}^H \theta_{2h} \cos h\omega t + \theta_{2h-1} \sin h\omega t, \quad (4)$$

and recently [12] provided a formalism for applying periodograms to multi-band time series.

The multiharmonic periodogram is designed to find non-sinusoidal periodic signals in irregularly spaced time series data. However, as shown in [12], increased flexibility comes at a cost – the noise level of the periodogram relative to the maximum peak height tends to increase as the number of free parameters in the model  $\hat{y}$  increases, and in some cases the best frequency can tend to a harmonic or a subharmonic of the true frequency.

If the signal shape of interest is known *a priori* and can be expressed as a truncated Fourier series with  $H$  harmonics,

$$M(\omega t) = \sum_{h=1}^H \psi_{2h} \cos h\omega t + \psi_{2h-1} \sin h\omega t, \quad (5)$$

then a model of the form

$$\hat{y}_M(t|\theta, \omega) = \theta_0 + \theta_1 M(\omega(t - \theta_2)) \quad (6)$$

should provide improved signal to noise over the multiharmonic periodogram, since there are only three free parameters at each trial frequency, regardless of the number of harmonics used to express the template  $M$ . The difference between the template periodogram (i.e., the periodogram derived from  $\hat{y}_M$ ) and the multiharmonic periodogram of [7, 15] is that the harmonic amplitudes  $\psi_i$  are *not free parameters* in the template periodogram, while they are free parameters for the multiharmonic periodogram.

Unlike the Lomb-Scargle extensions previously mentioned, the template periodogram is *non-linear*. The non-linearity arises from the phase shift parameter  $\theta_2$ . For Lomb-Scargle and multiharmonic periodograms, the phase shift is equivalent to changing the relative amplitudes of the cosine and sine terms. This is not true for the template periodogram, since  $\psi_i$  terms are deliberately fixed. Instead, by working with  $x \equiv \cos \omega\theta_2$ , the template can be expressed as:

$$M(\omega(t - \theta_2)) = \sum_{h=1}^H \left[ \left( \psi_{2h} T_h(x) \mp \psi_{2h-1} \sqrt{1-x^2} U_{h-1}(x) \right) \cos h\omega t + \left( \psi_{2h-1} T_h(x) \pm \psi_{2h} \sqrt{1-x^2} U_{h-1}(x) \right) \sin h\omega t \right], \quad (7)$$

where  $T_n$  and  $U_n$  are the Chebyshev polynomials of the first and second kind, respectively, and the  $\pm$  sign ambiguity reflects the two possible signs of  $\sin \omega\theta_2$ .

Detailed derivations of the template periodogram will be presented in a later paper, but the non-linearity of the template periodogram can be reduced to finding the real roots of a polynomial of order  $\sim 6H$  at each trial frequency. The root finding step scales as  $O(H^3 - H^4)$  at each trial frequency, and the polynomial coefficients can be computed with a single non-equispaced fast Fourier transform (NFFT, [16]) of size  $N_f H$ . Thus, the template periodogram scales as

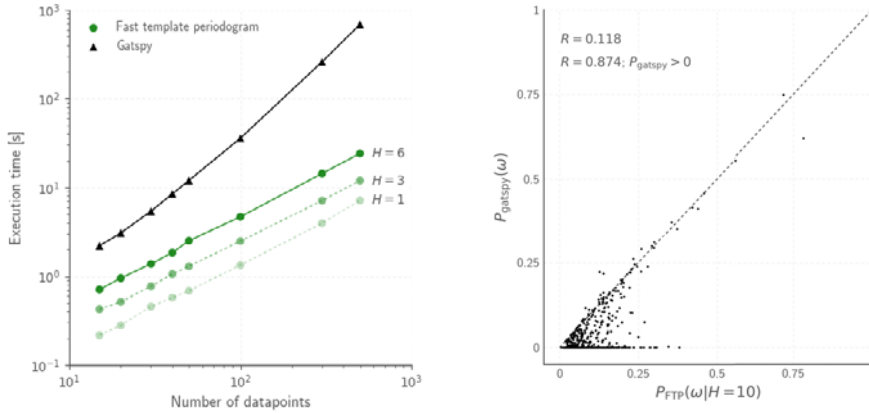
$$O(N_f H^4 + N_f H \log N_f H). \quad (8)$$

### 3 Implementation

A Python implementation of the fast template periodogram is available at

<https://github.com/PrincetonUniversity/FastTemplatePeriodogram>.

Timing comparisons with an existing Python template fitting implementation (`gatspy`, [17]) shown in Figure 1 demonstrate the improved computational efficiency of the fast template periodogram. In terms of accuracy, Figure 1 also shows that the fast template periodogram, which solves for the optimal parameters directly, is usually able to find better fits than the `gatspy` implementation, which uses non-linear optimization tools to find the optimal parameters.



**Figure 1.** Timing (*left*) and accuracy (*right*) comparisons between the *gatspy* ([17]) template fitting algorithm and the fast template periodogram. The fast template periodogram improves computational efficiency by a factor of  $\sim$ a few for small test cases and this improvement factor grows linearly in  $O(N_f \sim N_{\text{obs}})$ . Though the fast template periodogram has  $O(N_f \log N_f)$  asymptotic scaling for a given value of  $H$ , the root-finding utilities (which scale as  $O(N_f)$ ) dominate the computational time for all test cases shown here. In most cases, the fast template periodogram finds better template fits than *gatspy* (points that fall below the dashed line on the *right*-most figure).

## 4 Conclusion

Finding non-sinusoidal signals in irregularly spaced astrophysical data is a challenging problem that requires balancing model flexibility against the need to use as few model parameters as possible to minimize the relative noise of the periodogram. Template periodograms, or periodic matched filters, allow for the detection of periodic signals with fixed, arbitrary non-sinusoidal shapes without the increase of periodogram noise inherent in the multiharmonic periodogram presented in [7, 15]. However, template periodogram implementations to date ([5, 17]) have relied on non-linear optimization at each trial frequency to obtain the optimal amplitude, phase, and offset of the template. Thus, applying template periodograms to large photometric surveys like Pan-STARRS ([6]) requires prohibitive amounts of computational resources and scale as  $O(N_f N_{\text{obs}} \sim N_{\text{obs}}^2)$ .

However, template periodograms can be made orders of magnitude more computationally efficient. By using the algorithmic shortcuts described here, template periodograms could potentially prove computationally feasible on large photometric surveys with many ( $\gtrsim 10^2$ ) observations per lightcurve, like CoRoT ([8]), HATNet ([10]), and others. An open source implementation of the fast template periodogram is available on GitHub, and improvements to the existing implementation should continue to improve computational efficiency.

## References

- [1] Lomb, N. R., *Ap&SS*, **39**, 447 (1976)
- [2] Scargle, J. D., *ApJ*, **263**, 835 (1982)
- [3] Barning, F. J. M., *BAIN*, **17**, 22 (1963)
- [4] Vaníček, P., *Ap&SS*, **12**, 10 (1971)

- [5] Sesar, B., Hernitschek, N., Mitrović, S., et al., *AJ*, **153**, 204 (2017)
- [6] Chambers, K. C., Magnier, E. A., Metcalfe, N., et al., preprint (arXiv:1612.05560) (2016)
- [7] Palmer, D. M., *ApJ*, **695**, 496 (2009)
- [8] CoRoT Team, *The CoRoT Legacy Book: The Adventure of the Ultra High Precision Photometry from Space* (2016)
- [9] Smith, A. M. S., & WASP Consortium, *Contr. Astron. Obs. Skalnaté Pleso*, **43**, 500 (2014)
- [10] Bakos, G., Noyes, R. W., Kovács, G., Stanek, K. Z., Sasselov, D. D., & Domsa, I., *PASP*, **116**, 266
- [11] Koch, D. G., Borucki, W. J., Basri, G., et al., *ApJL*, **713**, L79 (2010)
- [12] VanderPlas, J. T., & Ivezić, Ž., *ApJ*, **812**, 18 (2015)
- [13] Yu, T., Simoff, S., & Jan, T., *Neurocomputing*, **73**, 2614 (2010)
- [14] Zechmeister, M., & Kürster, M., *A&A*, **496**, 577
- [15] Schwarzenberg-Czerny, A., *ApJL*, **460**, L107 (1996)
- [16] Keiner, J., Kunis, S., & Potts, D., *ACM Trans. Math. Softw.*, **36**, 19 (2009)
- [17] VanderPlas, J., *gatspy: General tools for Astronomical Time Series in Python*, *Astrophysics Source Code Library* (<http://ascl.net/1610.007>) (2016)