Draft version June 18, 2017
Preprint typeset using LaTeX style emulateapj v. 01/23/15

A FAST TEMPLATE PERIODOGRAM FOR DETECTING NON-SINUSOIDAL FIXED-SHAPE SIGNALS IN IRREGULARLY SAMPLED TIME SERIES DRAFT VERSION 1

J. Hoffman[1], J. Vanderplas[2], J. Hartman[1], G. Bakos[1]

Draft version June 18, 2017

ABSTRACT

Astrophysical time series often contain periodic signals. The sheer volume of time series data from photometric surveys demands computationally efficient methods for detecting and characterizing such signals. The most efficient algorithms available for this purpose are those that exploit the $\mathcal{O}(N \log N)$ scaling of the Fast Fourier Transform (FFT). However, these methods are not optimal for non-sinusoidal signal shapes. Template fits (or periodic matched filters) optimize sensitivity for *a priori* known signal shapes but at enormous computational cost. Current implementations of template periodograms scale as $\mathcal{O}(N_f N_{\rm obs})$, where N_f is the number of trial frequencies and $N_{\rm obs}$ is the number of lightcurve observations, and they do not gaurantee the best fit at each trial frequency. In this work, we present a non-linear extension of the Lomb-Scargle periodogram which provides a template-fitting algorithm that is both accurate (the exact optimal solutions are obtained except in rare cases) and computationally efficient (scaling as $\mathcal{O}(N_f \log N_f)$). The non-linear optimization of the template fit at each frequency is recast as a polynomial zero-finding problem, where the coefficients of the polynomial can be computed efficiently with FFTs. We show that our method, which uses truncated Fourier series to approximate templates, is twice as fast as existing algorithms for small problems ($N \lesssim 10$ observations) and 1-2 orders of magnitude faster for long base-line time series with $\mathcal{O}(10^4)$ observations. An open-source implementation of the fast template periodogram is available at github.com/PrincetonUniversity/FastTemplatePeriodogram.

1. INTRODUCTION

Astrophysical time series are challenging to analyze. Unlike time series in other domains such as economics and finance, astrophysical observations are often irregularly sampled in time with heteroskedastic, non-Gaussian, and time-correlated measurement uncertainties.

Irregular sampling thwarts the straightforward application of many well-known time series tools like the discrete Fourier transform (DFT) and the auto-regressive moving average (ARMA) models. The DFT is a particularly unfortunate loss, since the Fast Fourier Transform (Cooley & Tukey 1965) reduces the $\mathcal{O}(N^2)$ DFT to $\mathcal{O}(N \log N)$, and is a powerful tool for finding periodic signals.

The DFT can be extended to irregularly sampled data via what is sometimes referred to as the classical periodogram (Stoica et al. 2009)

$$P_x(\omega) = \frac{1}{N^2}\left|\sum_{n=0}^{N-1} y_n e^{-i\omega t_n}\right|^2. \tag{1}$$

However, as Stoica et al. (2009) point out, this is not an optimal measure of periodicity. A more robust estimate of the power spectrum is given by the Lomb-Scargle periodogram (Lomb 1976; Scargle 1982; Barning 1963; Vaníček 1971).

ARMA models can also be extended to unevenly sampled data with the CARMA model (Kelly et al. 2014; Zinn et al. 2016), but for the purposes of this paper, we focus solely on tools applicable to the detection of periodic signals in astrophysical data.

The Lomb-Scargle periodogram and its extensions can be expressed in terms of least-squares minimization between the data $\{y_n\}_{n=1}^N$ and a model $\hat{y}$. In the original formulation of the Lomb-Scargle periodogram,

$$\hat{y}_{\rm LS}(t|\theta,\omega) = \theta_0 \cos\omega t + \theta_1 \sin\omega t. \tag{2}$$

This is equivalent to performing a DFT if the data is regularly sampled. The Lomb-Scargle periodogram can be obtained from solving the linear system of equations that arise from the condition that the summed squares of residuals between the data and the optimal model,

$$\chi^2(\theta, S) \equiv \sum_i (y_i - \hat{y}(t_i|\theta))^2, \tag{3}$$

must be a local minimum. This means that

$$\left.\frac{\partial \chi^2}{\partial \theta_i}\right|_{\theta=\theta_{\rm best}} = 0 \quad \forall \theta_i \in \theta. \tag{4}$$

jah5@princeton.edu
[1] Department of Astrophysical Sciences, Princeton University, Princeton NJ 08540
[2] eScience Institute, University of Washington, Seattle, WA 98195

The resulting periodogram can be expressed as

$$P_{\rm LS} = \frac{1}{2\sigma^2}\left(\frac{\left[\sum_{n=1}^N (y_n - \bar{y})\cos\omega t_n\right]^2}{\sum_{n=1}^N \cos^2\omega t_i} + \frac{\left[\sum_{n=1}^N (y_n - \bar{y})\sin\omega t_n\right]^2}{\sum_{n=1}^N \sin^2\omega t_i}\right), \tag{5}$$

where $\bar{y} = \mathbb{E}[y_n]$, the mean of the data, and $\sigma = {\rm Var}(y_n)$, the variance of the data.

Heteroskedasticity can be handled by using weighted least squares,

$$\chi^2(\theta, S) \equiv \sum_i \frac{(y_i - \hat{y}(t_i|\theta))^2}{\sigma_i^2}, \tag{6}$$

with weights $w_i = \frac{W}{\sigma_i^2}$, $W \equiv \sum \sigma_i^{-2}$ being a normalization factor to ensure $\sum w_i = 1$, and correlated uncertainties can be accounted for by using the full covariance matrix, $\Sigma_{ij} = {\rm Cov}\left((y_i - \bar{y})(y_j - \bar{y})\right)$.

$$\chi^2(\theta, S) \equiv (y_i - \hat{y}(t_i|\theta))^{\rm T}\Sigma^{-1}(y_i - \hat{y}(t_i|\theta)). \tag{7}$$

If we assume the covariance matrix is diagonal, the Lomb-Scargle periodogram can be evaluated quickly in one of two popular ways. The first, by Press & Rybicki (1989) involves"extirpolating" irregularly sampled data onto a regularly sampled mesh, and then performing FFTs to evaluate the necessary sums. The second, as pointed out in Leroy (2012), is to use the non-equispaced FFT (NFFT) Keiner et al. (2009) to evaluate the sums; this provides an order of magnitude speedup over the Press & Rybicki (1989) algorithm, and both algorithms scale as $\mathcal{O}(N_f \log N_f)$.

There is a growing population of alternative methods for detecting periodic signals in astrophysical data. Some of these methods can reliably outperform the Lomb-Scargle periodogram, especially for non-sinusoidal signal shapes (see Graham et al. (2013) for a recent empirical review of period finding algorithms). However, a key advantage the LS periodogram and its extensions is speed. Virtually all other methods scale as $N_{\rm obs} \times N_f$, where $N_{\rm obs}$ is the number of observations and N_f is the number of trial frequencies, while the Lomb-Scargle periodogram scales as $N_f \log N_f$. The virtual independence of Lomb-Scargle on the number of observations (assuming $N_f \gtrsim N_{\rm obs}$) is especially attractive for lightcurves with $N_{\rm obs} \gg \log N_f \sim 50$.

Algorithmic efficiency will become increasingly important as the volume of data produced by astronomical observatories continues to grow larger. The HATNet survey (Bakos et al. 2004), for example, has already made $\mathcal{O}(10^4)$ observations of $\mathcal{O}(10^6 - 10^7)$ stars. The Gaia telescope (Gaia Collaboration et al. 2016) is set to produce $\mathcal{O}(10 - 100)$ observations of $\mathcal{O}(10^9)$ stars. The Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. (2009)) will make $\mathcal{O}(10^2 - 10^3)$ observations of $\mathcal{O}(10^{10})$ stars during its operation starting in 2023.

This paper develops new extensions to least-squares spectral analysis for arbitrary signal shapes. For non-periodic signals this method is known as matched filter analysis, and can be extended to search for periodic signals by phase folding the data at different trial periods. We refer to the latter technique, the subject of this paper, as template fitting.

Recently, Sesar et al. (2016) found that template fitting significantly improved period and amplitude estimation for RR Lyrae in Pan-STARRS DR1 (Chambers et al. 2016). Since the signal shapes for RR Lyrae in various bandpasses are known *a priori* (see Sesar et al. (2010)), template fitting provides an optimal estimate of amplitude and period, given that the object is indeed an RR Lyrae star well modeled by at least one of the templates. Templates were especially crucial for Pan-STARRS data, since there are typically only 35 observations per source over 5 bands (Hernitschek et al. 2016), not enough to obtain accurate amplitudes empirically by phase-folding. By including domain knowledge (i.e. knowledge of what RR Lyrae lightcurves look like), template fitting allows for accurate inferences of amplitude even for undersampled lightcurves.

However, the improved accuracy comes at substantial computational cost: the template fitting procedure took 30 minutes per CPU per object, and Sesar et al. (2016) were forced to limit the number of fitted lightcurves ($\lesssim$ 1000) in order to keep the computational costs to a reasonable level. Several cuts were made before the template fitting step to reduce the more than 1 million Pan-STARRS DR1 objects to a small enough number, and each of these steps removes a small portion of RR Lyrae from the sample. Though this number was reported by Sesar et al. (2016) to be small ($\lesssim 2\%$), it may be possible to further improve the completeness of the final sample by applying template fits to a larger number of objects, which would require either more computational resources, more time, or, ideally, a more efficient template fitting procedure.

The paper is organized as follows. Section 2 poses the problem of template fitting in the language of least squares spectral analysis and derives the fast template periodogram. Section 3 describes a freely available implementation of the new template periodogram. Section 4 summarizes our results, addresses caveats, and discusses possible avenues for improving the efficiency of the current algorithm.

2. DERIVATIONS

We define a template $\mathbf{M}$

$$\mathbf{M} : [0, 1) \to \mathbb{R}, \tag{8}$$

as a mapping between the unit interval and the set of real numbers. We restrict our discussion to sufficiently smooth templates such that $\mathbf{M}$ can be adequately described by a truncated Fourier series

$$\hat{\mathbf{M}}(\omega t|H) = \sum_{n=1}^{H}\left[c_n \cos n\omega t + s_n \sin n\omega t\right] \tag{9}$$

for some $H > 0$. Specifically, we require that

$$\forall t \in (0, 1] : \lim_{H\to\infty}\left|\mathbf{M}(t) - \hat{\mathbf{M}}(t|H)\right| = 0 \tag{10}$$

That the c_n and s_n values are *fixed* (i.e., they define the template) is the crucial difference between the template periodogram and the Multiharmonic Lomb Scargle (or FastChi) (Palmer 2009), where c_n and s_n are *free parameters*.

We now construct a periodogram for this template. The periodogram assumes that an observed time series $S = \{(t_i, y_i, \sigma_i)\}_{i=1}^N$ can be modeled by a scaled, transposed template that repeats with period $2\pi/\omega$, i.e.

$$y_i \approx \hat{y}(\omega t_i | \theta, \mathbf{M}) = \theta_1 \mathbf{M}(\omega t_i - \theta_2) + \theta_3, \tag{11}$$

where $\theta \in \mathbb{R}^3$ is a set of model parameters.

The optimal parameters are the location of a local minimum of the (weighted) sum of squared residuals,

$$\chi^2(\theta, S) \equiv \sum_i w_i (y_i - \hat{y}(\omega t_i|\theta))^2, \tag{12}$$

and thus the following condition must hold for all three model parameters at the optimal solution $\theta = \theta_{\rm opt}$:

$$\left.\frac{\partial \chi^2}{\partial \theta_j}\right|_{\theta=\theta_{\rm opt}} = 0 \quad \forall \theta_j \in \theta. \tag{13}$$

Note that we have implicitly assumed $\chi^2(\theta, S)$ is a C^1 differentiable function of θ, which requires that $\mathbf{M}$ is a C^1 differentiable function. Though this assumption could be violated if we considered a more complete set of templates, (e.g. a box function), our restriction to truncated Fourier series ensures that $\mathbf{M}$ is C^1 differentiable and thus that $\chi^2(\theta, S)$ is C^1 differentiable.

Note that we also implicitly assume $\sigma_i > 0$ for all i and we will later assume that the variance of the observations y_i is non-zero. If there are no measurement errors, i.e. $\sigma_i = 0$ for all i, then uniform weights (setting $\sigma_i = 1$) should be used. If the variance of the observations y is zero, the periodogram is trivial and will be zero for all frequencies. We do not consider the case where $\sigma_i = 0$ for some observations i and $\sigma_j > 0$ for some observations j. In that case, the weighted least squares estimator is undefined.

We can derive a system of equations for $\theta_{\rm opt}$ from the condition given in Equation 13. The explicit condition that must be met for each parameter θ_j is simplified below, using

$$\hat{y}_i = \hat{y}(\omega t_i|\theta) \tag{14}$$

and

$$\partial_j \hat{y}_i = \left.\frac{\partial \hat{y}(\omega t|\theta)}{\partial \theta_j}\right|_{t=t_i} \tag{15}$$

for brevity:

$$\begin{aligned} 0 &= \left.\frac{\partial \chi^2}{\partial \theta_j}\right|_{\theta=\theta_{\rm opt}} \\ &= -2\sum_i w_i\,(y_i - \hat{y}_i)\,(\partial_j \hat{y})_i \end{aligned} \tag{16}$$
$$\sum_i w_i y_i (\partial_j \hat{y})_i = \sum_i w_i \hat{y}_i (\partial_j \hat{y})_i.$$

The above is a general result that extends to all least squares periodograms. To simplify derivations, we adopt the following notation:

$$\langle X \rangle \equiv \sum_i w_i X_i \tag{17}$$
$$\langle XY \rangle \equiv \sum_i w_i X_i Y_i \tag{18}$$
$$\mathrm{Cov}(X, Y) \equiv \langle XY \rangle - \langle X \rangle \langle Y \rangle \tag{19}$$
$$\mathrm{Var}(X) \equiv \mathrm{Cov}(X, X) \tag{20}$$

In addition, we denote the shifted template $\mathbf{M}(x - \theta_2)$ by $\mathbf{M}_{\theta_2}(x)$.

For the amplitude and offset model parameters (θ_1 and θ_3, respectively), we obtain the following relations from Equation 16

$$\langle y\mathbf{M}_{\theta_2}(\omega t)\rangle = \theta_1 \left\langle \mathbf{M}^2_{\theta_2}(\omega t)\right\rangle + \theta_3 \left\langle \mathbf{M}_{\theta_2}(\omega t)\right\rangle \tag{21}$$
$$\theta_3 = \bar{y} - \theta_1 \left\langle \mathbf{M}_{\theta_2}(\omega t)\right\rangle, \tag{22}$$

where $\bar{y} \equiv \langle y \rangle$. Combining these expressions yields

$$\theta_1 = \frac{\langle (y - \bar{y})\mathbf{M}_{\theta_2}(\omega t)\rangle}{\mathrm{Var}(\mathbf{M}_{\theta_2}(\omega t))}. \tag{23}$$

For the offset parameter θ_2,

$$\begin{aligned} \frac{\partial \hat{y}}{\partial \theta_2} &= \theta_1 \frac{\partial \mathbf{M}_{\theta_2}}{\partial \theta_2} \\ &= -\theta_1 \partial \mathbf{M}_{\theta_2}, \end{aligned} \tag{24}$$

where

$$\partial \mathbf{M}_{\theta_2}(x) = \sum_n \left[s_n \cos n(x - \theta_2) - c_n \sin n(x - \theta_2)\right]. \tag{25}$$

From Equations 16, 22, and 24 we obtain

$$\theta_1 = \frac{\langle (y - \bar{y})\partial\mathbf{M}_{\theta_2}(\omega t)\rangle}{\mathrm{Cov}\left(\mathbf{M}_{\theta_2}(\omega t), \partial\mathbf{M}_{\theta_2}(\omega t)\right)}, \tag{26}$$

which, combined with Equation 23, provides a non-linear condition for θ_2:

$$\begin{aligned} &\langle (y - \bar{y})\partial\mathbf{M}_{\theta_2}(\omega t)\rangle\, \mathrm{Var}(\mathbf{M}_{\theta_2}(\omega t)) \\ &= \langle (y - \bar{y})\mathbf{M}_{\theta_2}(\omega t)\rangle\, \mathrm{Cov}(\mathbf{M}_{\theta_2}(\omega t), \partial\mathbf{M}_{\theta_2}(\omega t)). \end{aligned} \tag{27}$$

To obtain an explicit expression for Equation 27, we define several quantities,

$$CC_{nm} \equiv \mathrm{Cov}(\cos n\omega t, \cos m\omega t) \tag{28}$$
$$CS_{nm} \equiv \mathrm{Cov}(\cos n\omega t, \sin m\omega t) \tag{29}$$
$$SS_{nm} \equiv \mathrm{Cov}(\sin n\omega t, \sin m\omega t) \tag{30}$$
$$YC_n \equiv \langle (y - \bar{y}) \cos n\omega t \rangle \tag{31}$$
$$YS_n \equiv \langle (y - \bar{y}) \sin n\omega t \rangle, \tag{32}$$

all of which can be evaluated quickly for every trial frequency ω with the use of NFFTs. We also obtain an expression for the shifted template:

$$\begin{aligned}
\mathbf{M}_{\theta_2}(x) &= \sum_n \left[c_n \cos n(x-\theta_2) + s_n \sin n(x-\theta_2)\right] \\
&= \sum_n \left[(c_n \cos n\theta_2 - s_n \sin n\theta_2)\cos nx \right. \\
&\qquad \left. + (s_n \cos n\theta_2 + c_n \sin n\theta_2)\sin nx\right] \\
&= \sum_n \left[\left(c_n T_n(u) \mp s_n \sqrt{1-u^2}U_{n-1}(u)\right)\cos nx \right. \\
&\qquad \left. + \left(s_n T_n(u) \pm c_n \sqrt{1-u^2}U_{n-1}(u)\right)\sin nx\right] \\
&= \sum_n \left[A_n(u)\cos nx + B_n(u)\sin nx\right]
\end{aligned} \tag{33}$$

where $u \equiv \cos\theta_2$, T_n and U_n are the Chebyshev polynomials of the first and second kind, respectively, and the ± ambiguity arises out of the two possible signs for $\sin\theta_2$.

The derivatives of first and second order Chebyshev polynomials are known to be

$$\frac{dT_n}{dx} = nU_{n-1}(x) \tag{34}$$

$$\frac{dU_n}{dx} = \frac{(n+1)T_{n+1}(x) - xU_n(x)}{x^2-1}, \tag{35}$$

and this implies that the first derivative of the shifted template is

$$\begin{aligned}
\partial\mathbf{M}_{\theta_2}(x) &= \sum_n \left[n\left(c_n U_{n-1}(u) \pm s_n \frac{T_n(u)}{\sqrt{1-u^2}}\right)\cos nx \right. \\
&\qquad \left. + n\left(s_n U_{n-1}(u) \mp c_n \frac{T_n(u)}{\sqrt{1-u^2}}\right)\sin nx\right] \\
\partial\mathbf{M}_{\theta_2}(x) &= \sum_n \left[\partial A_n(u)\cos nx + \partial B_n(u)\sin nx\right]
\end{aligned} \tag{36}$$

Using the sums provided in Equations 28 – 32, writing A_n and B_n as as shorthand for $A_n(u)$ and $B_n(u)$, and employing Einstein summation notation, we have that

$$\langle (y-\bar{y})\mathbf{M}_{\theta_2}\rangle = A_n YC^n + B_n YS^n \tag{37}$$

$$\langle (y-\bar{y})\partial\mathbf{M}_{\theta_2}\rangle = \partial A_n YC^n + \partial B_n YS^n \tag{38}$$

$$\begin{aligned}
\mathrm{Var}(\mathbf{M}_{\theta_2}) &= A_n A_m CC^{nm} \\
&\quad + 2A_n B_m CS^{nm} + B_n B_m SS^{nm}
\end{aligned} \tag{39}$$

$$\begin{aligned}
\mathrm{Cov}(\mathbf{M}_{\theta_2}, \partial\mathbf{M}_{\theta_2}) &= A_n \partial A_m CC^{nm} \quad . \\
&\quad + (A_n \partial B_m + B_n \partial A_m) CS^{nm} \\
&\quad + B_n \partial B_m SS^{nm}
\end{aligned} \tag{40}$$

Equation 27 now becomes

$$\begin{aligned}
0 = \quad & A_n A_m \partial A_k \left(YC^n CC^{mk} - YC^k CC^{mn}\right) \\
& + A_n A_m \partial B_k \left(YC^n CS^{mk} - YS^k CC^{mn}\right) \\
& + A_n B_m \partial A_k \left(YC^n CS^{km} + YS^m CC^{kn}\right) \\
& + A_n B_m \partial B_k \left(YC^n SS^{mk} + YS^m CS^{nk}\right) \\
& + B_n B_m \partial A_k \left(YS^n CS^{km} - YC^k SS^{nm}\right) \\
& + B_n B_m \partial B_k \left(YS^n SS^{mk} - YS^k SS^{mn}\right).
\end{aligned} \tag{41}$$

Each $AA\partial A$, $AA\partial B$, ..., $BB\partial B$ can be expressed as

$$AA\partial A,\ AA\partial B,\ ... = p(u) + (1-u^2)^{-1/2}q(u), \tag{42}$$

where both $p(u)$ and $q(u)$ are polynomials in u. Therefore, Equation 41 can be expressed in the same form, and if we assume $|u| < 1$, a polynomial condition can be obtained

$$0 = (1-u^2)\hat{p}^2(u) - \hat{q}^2(u) = \hat{P}(u) \quad |u| < 1 \tag{43}$$

for some polynomials $\hat{p}, \hat{q}$, and $\hat{P}$.

Here and in the implementation presented in Section 3, we do not attempt to obtain an explicit form for the polynomials $\hat{p}$, $\hat{q}$, $\hat{P}$, ... We compute these polynomials as needed with polynomial algebra and the relation given in Equation 41, and precompute the polynomials $AA\partial A$, $AA\partial B$, ... for each template, since they do not depend on the data.

We solve for the zeros of the polynomial condition defined by Equation 43 using the `numpy.polynomial.polyroots` function, which solves for the eigenvalues of the polynomial companion matrix. We also test the special cases where $u = 1$, i.e. $\theta_2 \in \{0, \pi\}$.

The set of real roots of $\hat{P}$, $R \in \mathcal{R}$, provides a set of θ_2 parameters, one of which is the optimal value. We evaluate the periodogram at each $\theta_2 \in R \cup \{-1, 1\}$ to find the optimal value of θ_2.

An explicit form of the polynomial given in Equation 41 may provide some shortcuts that could further improve computational efficiency. As discussed in Section 2.3 and 3, deriving the polynomial $\hat{P}$ and finding its real roots is the limiting computational step for the FTP. The efficiency of the root finding step is aided by the fact that $\hat{P}$ is a polynomial, and thus the roots can be approximated by the eigenvalues of the polynomial companion matrix, or by a variety of other methods **cite rootfinding methods**, which in many cases are faster and more accurate than general non-linear root finding methods **this should be backed up more.**

We have derived an explicit, non-linear system of equations to solve for the parameters θ_1, θ_2, and θ_3. Solving this system of equations requires finding the zeros of a polynomial $\hat{P}(u)$ at the given trial frequency.

2.1. *Negative amplitude solutions*

The model $\hat{y} = \theta_1 \mathbf{M}_{\theta_2} + \theta_0$ allows for $\theta_1 < 0$ solutions. In the original formulation of Lomb-Scargle and in linear extensions involving multlple harmonics, negative amplitudes translate to phase differences, since $-\cos x = \cos(x - \pi)$ and $-\sin x = \sin(x - \pi)$.

However, for non-sinusoidal templates, $\mathbf{M}$, negative amplitudes do not generally correspond to a phase difference. For example, a detached eclipsing binary template $\mathbf{M}_{\rm EB}(x)$ cannot be expressed in terms of a phase-shifted negative eclipsing binary template; i.e. $\mathbf{M}_{\rm EB} \neq -\mathbf{M}_{\rm EB}(x - \phi)$ for any $\phi \in [0, 2\pi)$.

Negative amplitude solutions found by the fast template periodogram are usually undesirable, as they may produce false positives for lightcurves that resemble flipped versions of the desired template, and allowing for $\theta_1 < 0$ solutions increases the number of effective free parameters of the model, which lowers the signal to noise, especially for weak signals.

One possible remedy for this problem is to set $P_{\rm FTP}(\omega) = 0$ if the optimal solution for θ_1 is negative, but this complicates the interpretation of $P_{\rm FTP}$. Another possible remedy is, for frequencies that have a $\theta_1 < 0$ solution, to search for the optimal parameters while enforcing that $\theta_1 > 0$, e.g. via non-linear optimization, but this likely will eliminate the computational advantage of FTP over existing methods.

Thus, we allow for negative amplitude solutions in the model fit and caution the user to check that the best fit θ_1 is positive. Negative amplitude solutions may only be a problem for weak signals, or signals that, when flipped, are similar to other astrophysical signals.

2.2. *Extending to multi-band observations*

2.2.1. *Multi-phase model*

As shown in VanderPlas & Ivezić (2015), the multi-phase periodogram (their $(N_{\rm base}, N_{\rm band}) = (0, 1)$ periodogram), for any model can be expressed as a linear combination of single-band periodograms:

$$P^{(0,1)}(\omega) = \frac{\sum_{k=1}^{K} \chi^2_{0,k} P_k(\omega)}{\sum_{k=1}^{K} \chi^2_{0,k}} \tag{44}$$

where K denotes the number of bands, $\chi^2_{0,k}$ is the weighted sum of squared residuals between the data in the k-th band and its weighted mean $\langle y \rangle$, and $P_k(\omega)$ is the periodogram value of the k-th band at the trial frequency ω.

With Equation 44, the template periodogram is readily applicable to multi-band time series, which is crucial for experiments like LSST, SDSS, Pan-STARRS, and other current and future photometric surveys.

2.2.2. *Shared-phase model*

Extending the template periodogram to multi-band observations while fixing the relative phases of the templates requires additional derivation. A shared-phase model has $2K + 1$ free parameters, with an offset and amplitude for each filter and a shared phase term.

The χ^2 corresponding to the shared-phase multiband model is

$$\chi^2 = \sum_i^K \sum_j^{N_i} w_j^{(i)} (y_j^{(i)} - \hat{y}_j^{(i)})^2, \tag{45}$$

where $\hat{y}_j^{(i)} = \theta_1^{(i)} \mathbf{M}_{\theta_2}^{(i)} + \theta_3^{(i)}$ represents the model corresponding to a given filter i. The system of equations derived by setting $\partial_\theta \chi^2 = 0$ for each parameter θ reduces to the familiar expressions for the amplitude and offset:

$$\theta_1^{(i)} = \frac{\left\langle (y_j^{(i)} - \bar{y}^{(i)}) \mathbf{M}_{\theta_2}^{(i)} \right\rangle}{\mathrm{Var}(\mathbf{M}_{\theta_2}^{(i)})} \tag{46}$$

$$\theta_3^{(i)} = \bar{y}^{(i)} - \theta_1^{(i)} \mathrm{Var}(\mathbf{M}_{\theta_2}^{(i)}), \tag{47}$$

as well as a non-linear relation that must be satisfied for the shared-phase parameter:

$$\sum_i^M \left\langle (y_j^{(i)} - \bar{y}^{(i)}) \mathbf{M}_{\theta_2}^{(i)} \right\rangle \left(\prod_{k \neq i} \left(\mathrm{Var}\left(\mathbf{M}_{\theta_2}^{(k)} \right) \right)^2 \right) \tag{48}$$

$$\times \left(\left\langle (y_j^{(i)} - \bar{y}^{(i)}) \partial \mathbf{M}_{\theta_2}^{(i)} \right\rangle \mathrm{Var}\left(\mathbf{M}_{\theta_2}^{(i)} \right) - \left\langle (y_j^{(i)} - \bar{y}^{(i)}) \mathbf{M}_{\theta_2}^{(i)} \right\rangle \mathrm{Cov}\left(\mathbf{M}_{\theta_2}^{(i)}, \partial \mathbf{M}_{\theta_2}^{(i)} \right) \right) = 0. \tag{49}$$

2.2.3. *Other multi-band extensions*

Other multi-band template periodograms are derivable; for instance, Sesar et al. (2016) used what amounted to a shared phase, amplitude, and offset model. Their model can be translated to

$$\hat{y}_j^{(i)} = \theta_1 \left(\underbrace{\alpha^{(i)} \mathbf{M}_{\theta_2}^{(i)} + \beta^{(i)}}_{\widetilde{M}_{\theta_2}^{(i)}} \right) + \theta_3 \tag{50}$$

where the $\alpha^{(i)}$ and $\beta^{(i)}$ parameters are fixed. This model has only 3 free parameters, which allowed Sesar et al. (2016) to achieve high signal-to-noise despite having less than 10 epochs in each band.

In this case, the requirement that $\partial_{\theta_1}\chi^2 = 0$ gives

$$0 = \sum_i^K \left\langle y_j^{(i)}\widetilde{M}^{(i)}_{\theta_2}\right\rangle - \theta_1 \sum_i^K \left\langle \left(\widetilde{M}^{(i)}_{\theta_2}\right)^2\right\rangle \tag{51}$$

$$- \theta_3 \sum_i^K \left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle . \tag{52}$$

Requiring $\partial_{\theta_3}\chi^2 = 0$ gives another relation between the offset and amplitude parameters

$$0 = \sum_i^K \overline{y}^{(i)} - \theta_1 \sum_i^K \left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle - K\theta_3 \tag{53}$$

$$\implies \theta_3 = \overline{y^{(i)}} - \theta_1 \overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle}. \tag{54}$$

which implies that

$$0 = \sum_i^K \left\langle y_j^{(i)}\widetilde{M}^{(i)}_{\theta_2}\right\rangle - \theta_1 \sum_i^K \left\langle \left(\widetilde{M}^{(i)}_{\theta_2}\right)^2\right\rangle \tag{55}$$

$$- K\overline{y^{(i)}}\,\overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle} + K\theta_1 \overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle}^2 \tag{56}$$

$$= \overline{\left\langle (y_j^{(i)} - \overline{y^{(i)}})\widetilde{M}^{(i)}_{\theta_2}\right\rangle} - \theta_1 \left(\overline{\left\langle \left(\widetilde{M}^{(i)}_{\theta_2}\right)^2\right\rangle} \right. \tag{57}$$

$$\left. - \overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle}^2 \right) \tag{58}$$

$$\theta_1 = \frac{\overline{\left\langle (y_j^{(i)} - \overline{y^{(i)}})\widetilde{M}^{(i)}_{\theta_2}\right\rangle}}{\overline{\left\langle \left(\widetilde{M}^{(i)}_{\theta_2}\right)^2\right\rangle} - \overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle}^2} \tag{59}$$

Where here, $\bar{\cdot}$ denotes a quantity averaged over all bandpasses. Combining these relations with the requirement that $\partial_{\theta_2}\chi^2 = 0$, i.e.

$$0 = \theta_1 \sum_i^M \left\langle y_j^{(i)}\partial\widetilde{M}^{(i)}_{\theta_2}\right\rangle - \theta_1 \left\langle \widetilde{M}^{(i)}_{\theta_2}\partial\widetilde{M}^{(i)}_{\theta_2}\right\rangle \tag{60}$$

$$- \theta_3 \left\langle \partial\widetilde{M}^{(i)}_{\theta_2}\right\rangle , \tag{61}$$

produces the condition that must be satisfied by the optimal phase parameter θ_2,

$$\overline{\left\langle (y_j^{(i)} - \overline{y^{(i)}})\partial\widetilde{M}^{(i)}_{\theta_2}\right\rangle} \left(\overline{\left\langle \left(\widetilde{M}^{(i)}_{\theta_2}\right)^2\right\rangle} - \overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle}^2 \right) - \overline{\left\langle (y_j^{(i)} - \overline{y^{(i)}})\widetilde{M}^{(i)}_{\theta_2}\right\rangle} \left(\overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\partial\widetilde{M}^{(i)}_{\theta_2}\right\rangle} - \overline{\left\langle \widetilde{M}^{(i)}_{\theta_2}\right\rangle\left\langle \partial\widetilde{M}^{(i)}_{\theta_2}\right\rangle} \right). \tag{62}$$

2.3. *Computational requirements*

For a given number of harmonics H, the task of deriving $\hat{P}$ requires a triple sum over H terms, with each sum requiring $\mathcal{O}(n_{\hat{P}})$ operations, where $n_{\hat{P}}$ is the order of $\hat{P}$. The order of $\hat{P}$ can be shown to be

$$6H - \deg\left(\text{gcf}\left((1-u^2)\star\hat{p}^2, \hat{q}^2\right)\right) \propto H, \tag{63}$$

where gcf(p_1, p_2) denotes the greatest common polynomial factor between polynomials p_1 and p_2, and deg(p) denotes the degree of polynomial p. Computing the coefficients of $\hat{P}$ therefore scales as $\mathcal{O}(H^4)$ at each trial frequency.

The computational complexity of polynomial root finding is algorithm dependent. If we choose to perform singular value decomposition of the polynomial companion matrix[3], the root finding step scales as $\mathcal{O}(n_{\hat{P}}^3) = \mathcal{O}(H^3)$. The polynomial root-finding step should be asymptotically faster (for large H) than the computation of the polynomial coefficients.

When considering N_f trial frequencies, the polynomial computation and root-finding step scales as $\mathcal{O}(H^4 N_f)$. The computation of the sums (Equations 28 – 32) scales as $\mathcal{O}(HN_f \log HN_f)$. Therefore, the entire template periodogram scales as

$$\mathcal{O}(HN_f \log HN_f + H^4 N_f). \tag{64}$$

For a fixed number of harmonics H, the template periodogram scales as $\mathcal{O}(N_f \log N_f)$. However, for a constant number of trial frequencies N_f, the template algorithm scales as $\mathcal{O}(H^4)$, and computational resources alone limit H to reasonably small numbers $H \lesssim 15$ (see Figure 1).

2.3.1. *Computational considerations of multi-band extensions*

Multi-band extensions generally require more computational time than the single band case. For the shared-phase multi-band periodogram, the scaling becomes $\mathcal{O}KHN_f \log HN_f + K^4H^4N_f$. The final expression given in Equation 48 produces a root-finding polynomial that is of order $2H(5K-1)$ for $K > 1$, compared with the order $\sim 6H$ polynomial obtained for the $K = 1$ case.

For the shared phase, amplitude, and offset case, however, the polynomial order remains the same as for the single band case, though the time needed to compute the polynomial coefficients increases by a factor of K.

[3] The `numpy.polynomial.polyroots` function uses this method.

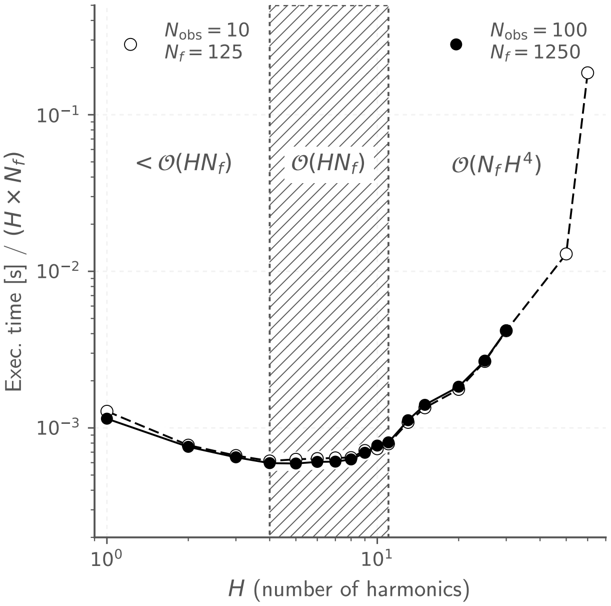

FIG. 1.— Computation time of FTP scaled by NH for different numbers of harmonics. For $H \lesssim 3$, FTP scales sublinearly in H (possibly due to a constant overhead per trial frequency, independent of H). When $3 \lesssim H \lesssim 11$, FTP scales approximately linearly in H, and when $H \gtrsim 11$ FTP approaches the $\mathcal{O}(H^4)$ scaling limit.

3. IMPLEMENTATION

An open-source implementation of the template periodogram in Python is available.[4] Computing $\hat{P}(u)$ is done using the `numpy.polynomial` module (Jones et al. 2001–). The `nfft` Python module, [5] which provides a Python implementation of the non-equispaced fast Fourier transform (Keiner et al. 2009), is used to compute the necessary sums for a particular time series.

No explicit parallelism is used anywhere in the current implementation, however certain linear algebra operations in `Scipy` use OpenMP via calls to BLAS libraries that have OpenMP enabled.

All timing tests were run on a quad-core 2.6 GHz Intel Core i7 MacBook Pro laptop (mid-2012 model) with 8GB of 1600 MHz DDR3 memory. The `Scipy` stack (version 0.18.1) was compiled with multi-threaded MKL libraries. However, the slowest portion of the algorithm, computing the polynomial coefficients, uses the `numpy.einsum` function which is not multi-threaded.

3.1. *Comparison with non-linear optimization*

Within the Python implementation of the FTP, we have provided access to slower methods that employ non-linear optimizationto compare the accuracy and speed of the template periodogram described in this paper.

Periodograms computed in Figures 2, 3, and 4 used simulated data. The simulated data has uniformly random observation times, with gaussian-random, homoskedastic, uncorrelated uncertainties. An eclipsing binary template, generated by fitting a well-sampled, high signal-to-noise eclipsing binary in the HATNet dataset (BD+56 603) with a 10-harmonic truncated Fourier series.

3.1.1. *Accuracy*

For weak signals or signals folded at the incorrect trial period, there may be a large number of local χ^2 minima in the parameter space, and thus non-linear optimization algorithms may have trouble finding the global minimum. The FTP, on the other hand, solves for the optimal parameters directly, and thus is able to recover optimal solutions even when the signal is weak or not present.

Figure 3 illustrates the accuracy improvement with FTP. Many solutions found via non-linear optimization are significantly suboptimal compared to the solutions found by the FTP.

Figure 4 compares FTP results obtained using the full template ($H = 10$) with those obtained using smaller numbers of harmonics. The left-most plot compares the $H = 1$ case (weighted Lomb-Scargle), which, as also demonstrated in Figure 2, illustrates the advantage of the template periodogram for known, non-sinusoidal signal shapes.

3.1.2. *Computation time*

FTP scales asymptotically as $\mathcal{O}(N_f H \log N_f H)$ with respect to the number of trial frequencies, N_f and as $\mathcal{O}(N_f H^4)$ with respect to the number of harmonics in which the template is expanded, H. However, for reasonable cases ($N_f \lesssim 10^{120}$ when $H = 5$) the computation time is dominated by computing polynomial coefficients and root finding, both of which scale linearly in N_f.

The number of trial frequencies needed for finding astrophysical signals in a typical photometric time series is

$$N_f = 1.75 \times 10^6 \left(\frac{\text{baseline}}{10 \text{ yrs}}\right)\left(\frac{\alpha}{5}\right)\left(\frac{15 \text{ mins}}{P_{\text{min}}}\right) \quad (65)$$

where α represents the "oversampling factor," $\Delta f_{\text{peak}}/\Delta f$, where $\Delta f_{\text{peak}} \sim 1/\text{baseline}$ is the typical width of a peak in the periodogram and Δf is the frequency spacing of the periodogram.

Extrapolating from the timing of a test case (500 observations, 5 harmonics, 15,000 trial frequencies), the summations account for approximately 5% of the computation time when $N_f \sim 10^6$. If polynomial computations and root-finding can be improved to the point where they no longer dominate the computation time, this would provide an order of magnitude speedup over the current implementation.

Figure 5 compares the timing of the FTP with that of previous methods that employ non-linear optimization. For the case when $N_f \propto N_{\text{obs}}$, FTP achieves a factor of 3 speedup for even the smallest test case (15 datapoints), while for larger cases ($N \sim 10^4$) FTP offers 2-3 orders of magnitude speed improvement. For the constant baseline case, FTP is a factor of ~ 2 faster for the smallest test case and a factor of ~ 20 faster for $N_{\text{obs}} \sim 10^4$. Future improvements to the FTP implementation could further improve speedups by 1-2 orders of magnitude over non-linear optimization.

[4] https://github.com/PrincetonUniversity/FastTemplatePeriodogram

[5] https://github.com/jakevdp/nfft

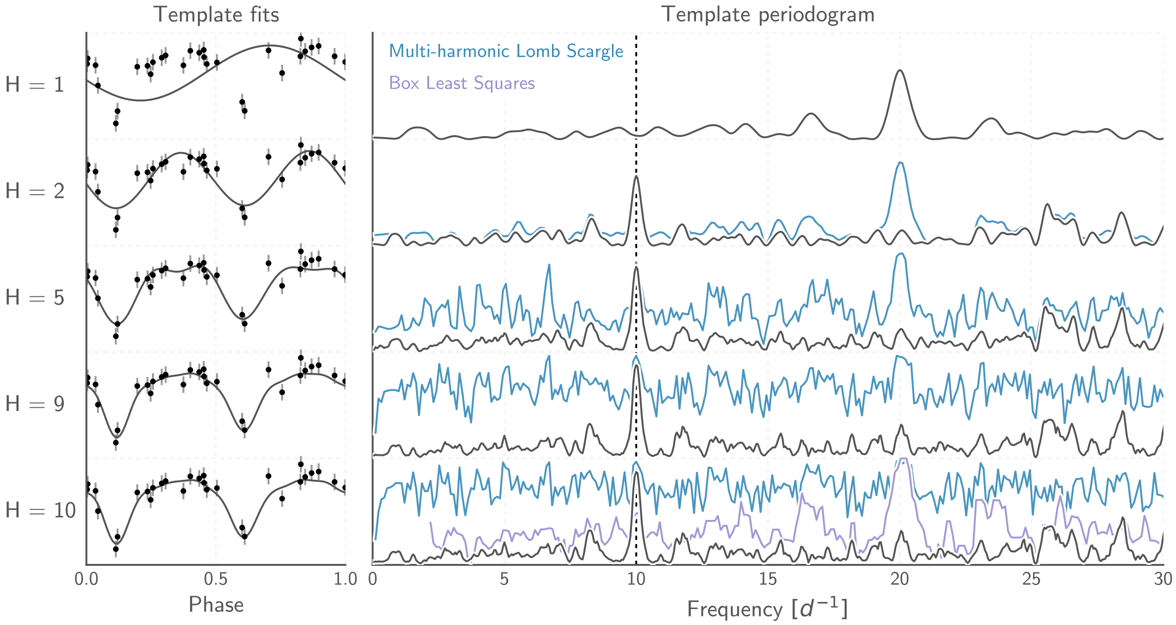

FIG. 2.— Template periodograms performed on a simulated eclipsing binary lightcurve (shown phase-folded in the left-hand plots). The top-most plot uses only one harmonic, equivalent to a Lomb-Scargle periodogram. Subsequent plots use an increasing number of harmonics, which produces a narrower and higher peak height around the correct frequency. For comparison, the multi-harmonic extension to Lomb-Scargle is plotted in blue, using the same number of harmonics as the FTP. The Box Least-Squares (Kovács et al. 2002) periodogram is shown in the final plot.

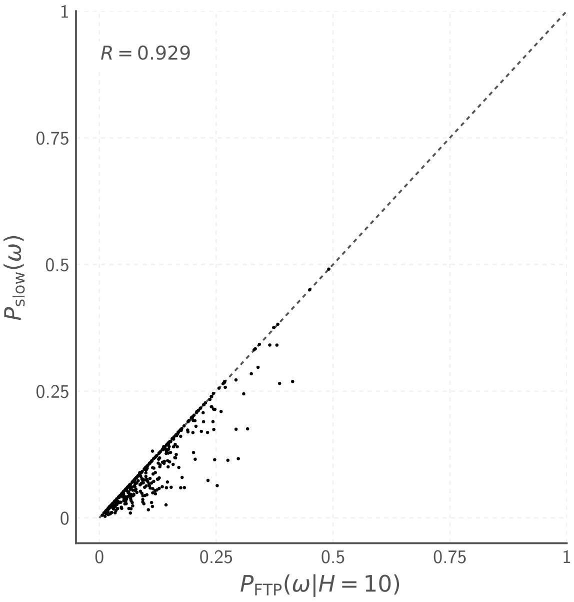

FIG. 3.— Comparing accuracy between previous methods that rely on non-linear optimization at each trial frequency with the fast template periodogram described in this paper. Both methods are applied to the same simulated data as shown in Figure 2. The FTP consistently finds more optimal template fits than those found with non-linear optimization, which do not gaurantee convergence to a globally optimal solution. The FTP solves for the optimal fit parameters directly, and therefore is able to achieve greater accuracy than template fits done via non-linear optimization.

4. DISCUSSION

Template fitting is a powerful technique for accurately recovering the period and amplitude of objects with *a priori* known lightcurve shapes. It has been used in the literature by, e.g. Sesar et al. (2016, 2010), to analyze RR Lyrae in the SDSS and PS1 datasets, where it has been shown to produce purer samples of RR Lyrae at a given completeness. The computational cost of current template fitting algorithms, however, limits their application to larger datasets or with a larger number of templates.

We have presented a novel template fitting algorithm that extends the Lomb-Scargle periodogram (Lomb 1976; Scargle 1982; Barning 1963; Vaníček 1971) to handle non-sinusoidal signals that can be expressed in terms of a truncated Fourier series with a reasonably small number of harmonics ($H \lesssim 10$).

The fast template periodogram (FTP) asymptotically scales as $\mathcal{O}(N_f H \log N_f H + N_f H^4)$, while previous template fitting algorithms such as the one used in the `gatspy` library (VanderPlas 2016), scale as $\mathcal{O}(N_f N_{\rm obs})$. However, the FTP effectively scales as $\mathcal{O}(N_f H^4)$, since the time needed to compute polynomial coefficients and perform zero-finding dominates the computational time for all practical cases ($N_f \lesssim 10^{120}$). The H^4 scaling effectively restricts templates to those that are sufficiently smooth to be explained by a small number of Fourier terms.

FTP also improves the accuracy of previous template fitting algorithms, which rely on non-linear optimization at each trial frequency to minimize the χ^2 of the template fit. The FTP routinely finds superior fits over non-linear optimization methods.

An open-source Python implementation of the FTP is

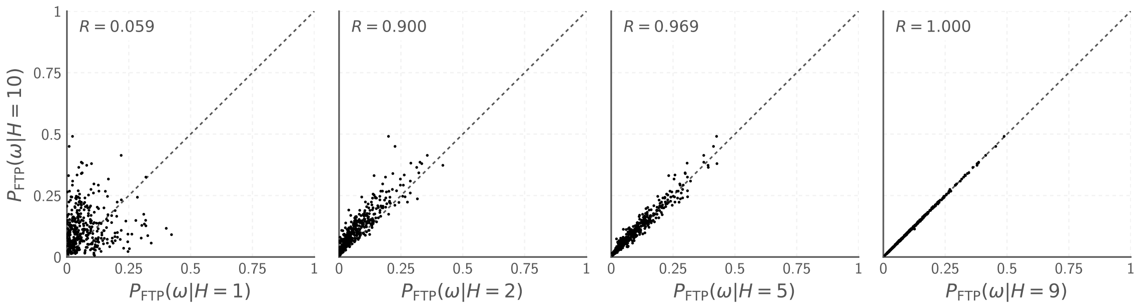

FIG. 4.— Comparing the template periodogram calculated with $H = 10$ harmonics to the template periodogram using a smaller number of harmonics $H < 10$. The template and data used to perform the periodogram calculations are the same as those shown in Figure 2.

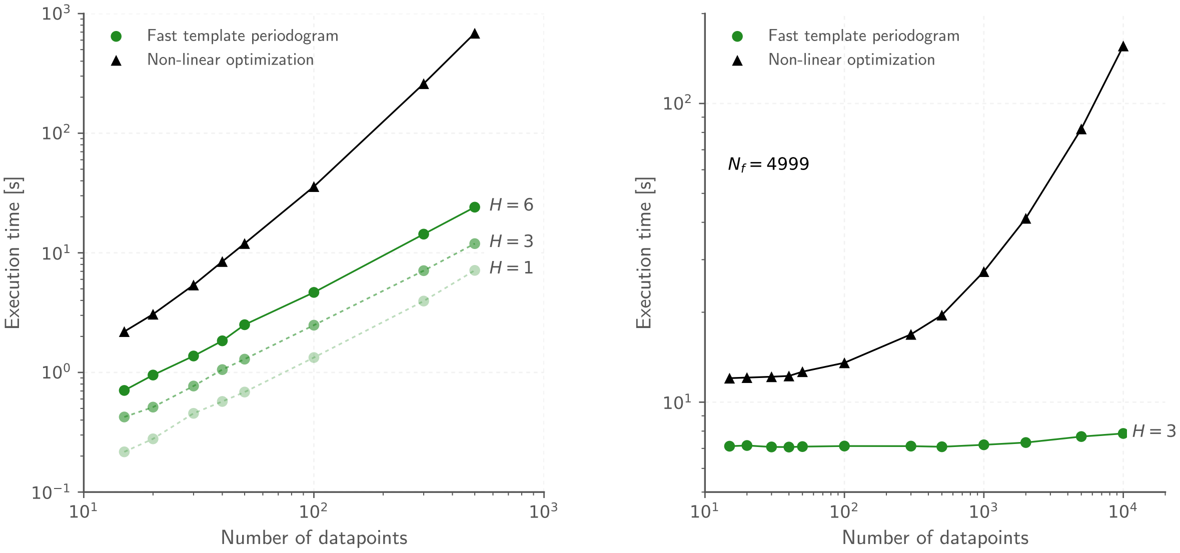

FIG. 5.— Computation time of FTP compared with alternative techniques that use non-linear optimization at each trial frequency. *Left*: timing for the case when $N_f = 12N_{\rm obs}$, i.e. the cadence of the observations is constant. *Right*: timing for the case when N_f is fixed, i.e. the baseline of the observations are constant. Non-linear optimization techniques scale as $\mathcal{O}(N_f N_{\rm obs})$ while the FTP scales as $\mathcal{O}(HN_f \log HN_f + N_f H^4)$, where H is the number of harmonics needed to approximate the template.

available at GitHub.[6] The current implementation could likely be improved by:

1. Improving the speed of the polynomial coefficient calculations and the zero-finding steps. This could potentially yield a speedup of $\sim 1 - 2$ orders of magnitude over the current implementation.

2. Exploiting the embarassingly parallel nature of the FTP using GPU's.

For a constant baseline, the current implementation improves existing methods by factors of a ∼few for lightcurves with $\mathcal{O}(100)$ observations, and by an order of magnitude or more for objects with more than 1,000 observations. These improvements, taken at face value, are not enough to make template fitting feasible on LSST-sized datasets. However, optimizing the polynomial computations could yield a factor of $\sim 25 - 100$ speedup over the current implementation, which would make the FTP 1-3 orders of magnitude faster than alternative techniques.

(acknowledge GRANTS.)

APPENDIX

EXPLICIT EXPRESSION OF POLYNOMIAL CONDITION

We introduce the following definitions:

[6] https://github.com/PrincetonUniversity/ FastTemplatePeriodogram

$$
\begin{aligned}
2uu_{nm} &= c_n c_m + s_n s_m \tag{A1}\\
2uv_{nm} &= c_n s_m + s_n c_m \tag{A2}\\
2vu_{nm} &= c_n s_m - s_n c_m \tag{A3}\\
2vv_{nm} &= c_n c_m - s_n s_m \tag{A4}\\
2UU_{nm} &= CC_{nm} + SS_{nm} \tag{A5}\\
2VV_{nm} &= CC_{nm} - SS_{nm} \tag{A6}\\
\widehat{YM} &= \langle y\mathbf{M}_{\theta_2}\rangle \tag{A7}\\
\widehat{MM} &= \left\langle \mathbf{M}_{\theta_2}^2 \right\rangle \tag{A8}\\
\overline{M} &= \langle \mathbf{M}_{\theta_2}\rangle \tag{A9}\\
MM &= \widehat{MM} - \overline{M}^2 \tag{A10}\\
YM &= \widehat{YM} - \bar{y}\overline{M} \tag{A11}
\end{aligned}
$$

For a given phase shift θ_2, the optimal amplitude and offset are obtained from requiring the partial derivatives of the sum of squared residuals, χ^2, to be zero.

Namely, we obtain that

$$
\begin{aligned}
0 = \frac{\partial\chi^2}{\partial\theta_1} &= 2\sum_i w_i(y_i - \hat{y}_i)\left(-\frac{\partial\hat{y}}{\partial\theta_1}\right)_i \tag{A12}\\
&= \sum_i w_i(y_i - \theta_1\mathbf{M}_{\theta_2} - \theta_3)\mathbf{M}_{\theta_2} \tag{A13}\\
&= YM - \theta_1\widehat{MM} - \theta_3\overline{M} \tag{A14}
\end{aligned}
$$

and

$$
\begin{aligned}
0 = \frac{\partial\chi^2}{\partial\theta_3} &= 2\sum_i w_i(y_i - \hat{y}_i)\left(-\frac{\partial\hat{y}}{\partial\theta_3}\right)_i \tag{A15}\\
&= \sum_i w_i(y_i - \theta_1\mathbf{M}_{\theta_2} - \theta_3) \tag{A16}\\
&= \bar{y} - \theta_1\overline{M} - \theta_3 \tag{A17}
\end{aligned}
$$

This system of equations can then be rewritten as

$$
\begin{pmatrix} \widehat{MM} & \overline{M} \\ \overline{M} & 1 \end{pmatrix}\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \widehat{YM} \\ \bar{y} \end{pmatrix} \tag{A18}
$$

which reduces to

$$
\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \frac{1}{\widehat{MM} - \overline{M}^2}\begin{pmatrix} 1 & -\overline{M} \\ -\overline{M} & \widehat{MM} \end{pmatrix}\begin{pmatrix} \widehat{YM} \\ \bar{y} \end{pmatrix} = \frac{1}{\widehat{MM} - \overline{M}^2}\begin{pmatrix} \widehat{YM} - \bar{y}\overline{M} \\ \widehat{MM}\bar{y} - \widehat{YM}\overline{M} \end{pmatrix} \tag{A19}
$$

Letting $MM = \widehat{MM} - \overline{M}^2$ and $YM = \widehat{YM} - \bar{y}\overline{M}$, we have

$$
\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} YM/MM \\ 1 - \overline{M}(YM/MM) \end{pmatrix} \tag{A20}
$$

This means we can rewrite the model $\hat{y} = \theta_1\mathbf{M}_{\theta_2} + \theta_3$ as

$$
\hat{y}_i = \bar{y} + \left(\frac{YM}{MM}\right)(M_i - \overline{M}) \tag{A21}
$$

To obtain an expression for the periodogram, $P = 1 - \chi^2/\chi_0^2$, we first compute χ^2

$$\chi^2 = \sum_i w_i (y_i - \hat{y}_i)^2 \tag{A22}$$

$$= \sum_i w_i (y_i^2 - 2y_i\hat{y}_i + \hat{y}_i^2) \tag{A23}$$

$$= \sum_i w_i ((y_i - \bar{y})^2 + 2y_i(\bar{y} - \bar{y} - \left(\frac{YM}{MM}\right)(M_i - \overline{M})) - \bar{y}^2 + \bar{y}^2 \tag{A24}$$

$$+ 2\bar{y}\left(\frac{YM}{MM}\right)(M_i - \overline{M}) + \left(\frac{YM}{MM}\right)^2 (M_i^2 - 2M_i\overline{M} + \overline{M}^2) \tag{A25}$$

$$= YY - 2\frac{(YM)^2}{MM} + \frac{(YM)^2}{MM} \tag{A26}$$

$$= YY - \frac{(YM)^2}{MM} \tag{A27}$$

Since, $\chi_0^2 = YY$, we have

$$P(\omega) = \frac{(YM)^2}{YY \cdot MM} \tag{A28}$$

We wish to maximize $P(\omega)$ with respect to the phase shift parameter θ_2,

$$\partial_{\theta_2} P = 0 = \frac{YM}{YY \cdot MM}\left(2\partial_{\theta_2}(YM) - \frac{YM}{MM}\partial_{\theta_2}(MM)\right) \tag{A29}$$

$$= 2MM\partial_{\theta_2}(YM) - YM\partial_{\theta_2}(MM). \tag{A30}$$

The final expression is the non-linear condition that must me met by the optimal phase shift parameter θ_2. However, satisfying Equation A29 is not *sufficient* to gaurantee that θ_2 is optimal. The value of the periodogram at each θ_2 satisfying Equation A29 must be computed, and the globally optimal solution chosen from this set.

We derive a more explicit form for Equation A29. To do so, we must first derive expressions for MM and YM.

$$MM \equiv \sum_i w_i M_i^2 \tag{A31}$$

$$= \sum_i w_i \left(\sum_n A_n \cos\omega n t_i + B_n \sin\omega n t_i\right)^2 \tag{A32}$$

$$= \sum_n \sum_m A_n A_m CC_{nm} + (A_n B_m CS_{nm} + B_n A_m (CS^T)_{nm}) + B_n B_m SS_{nm} \tag{A33}$$

$$= \sum_n \sum_m A_n A_m CC_{nm} + 2A_n B_m CS_{nm} + B_n B_m SS_{nm} \tag{A34}$$

$$\tag{A35}$$

Recall that

$$A_n(x) = c_n T_n(x) - s_n q\sqrt{1 - x^2}U_{n-1}(x) \tag{A36}$$

$$B_n(x) = s_n T_n(x) + c_n q\sqrt{1 - x^2}U_{n-1}(x) \tag{A37}$$

$$\tag{A38}$$

From this, we can derive

$$A_nA_m = \left(c_nT_n - qs_n\sqrt{1-x^2}U_{n-1}\right)\left(c_mT_m - qs_m\sqrt{1-x^2}U_{m-1}\right) \tag{A39}$$

$$= c_nc_mT_nT_m + s_ns_m(1-x^2)U_{n-1}U_{m-1} - q\sqrt{1-x^2}\left(c_ns_mT_nU_{m-1} + s_nc_mU_{n-1}T_m\right) \tag{A40}$$

$$= \left(\frac{c_nc_m + s_ns_m}{2}\right)\left(T_nT_m + (1-x^2)U_{n-1}U_{m-1}\right) + \left(\frac{c_nc_m - s_ns_m}{2}\right)\left(T_nT_m - (1-x^2)U_{n-1}U_{m-1}\right) \tag{A41}$$

$$- q\sqrt{1-x^2}\left(\left(\frac{c_ns_m + s_nc_m}{2}\right)\left(T_nU_{m-1} + U_{n-1}T_m\right) + \left(\frac{c_ns_m - s_nc_m}{2}\right)\left(T_nU_{m-1} - U_{n-1}T_m\right)\right) \tag{A42}$$

$$= uu_{nm}T_{n-m} + vv_{nm}T_{n+m} - q\sqrt{1-x^2}\left(vu_{nm}U_{n-m-1} + uv_{nm}U_{n+m-1}\right) \tag{A43}$$

From this, we can quickly obtain A_nB_m by letting $c_m \to s_m$ and $s_m \to -c_m$, which implies $uv \to -vv$, $uu \to vu$, $vu \to -uu$, and $vv \to uv$, and B_nB_m via $vv \to -vv$ and $uv \to -uv$:

$$A_nB_m = vu_{nm}T_{n-m} + uv_{nm}T_{n+m} + q\sqrt{1-x^2}\left(uu_{nm}U_{n-m-1} + vv_{nm}U_{n+m-1}\right) \tag{A44}$$

$$B_nB_m = uu_{nm}T_{n-m} - vv_{nm}T_{n+m} - q\sqrt{1-x^2}\left(vu_{nm}U_{n-m-1} - uv_{nm}U_{n+m-1}\right) \tag{A45}$$

From this, we can derive MM:

$$MM_{nm} = \left(uu_{nm}\left(CC_{nm} + SS_{nm}\right) + 2vu_{nm}CS_{nm}\right)T_{n-m} \tag{A46}$$

$$+ \left(vv_{nm}\left(CC_{nm} - SS_{nm}\right) + 2uv_{nm}CS_{nm}\right)T_{n+m} \tag{A47}$$

$$- q\sqrt{1-x^2}\Big\{\left(vu_{nm}\left(CC_{nm} + SS_{nm}\right) - 2uu_{nm}CS_{nm}\right)U_{n-m-1} \tag{A48}$$

$$+ \left(uv_{nm}\left(CC_{nm} - SS_{nm}\right) - 2vv_{nm}CS_{nm}\right)U_{n+m-1}\Big\} \tag{A49}$$

$$= 2\left(uu_{nm}UU_{nm} + vu_{nm}CS_{nm}\right)T_{n-m} \tag{A50}$$

$$+ 2\left(vv_{nm}VV_{nm} + uv_{nm}CS_{nm}\right)T_{n+m} \tag{A51}$$

$$- 2q\sqrt{1-x^2}\Big\{\left(vu_{nm}UU_{nm} - uu_{nm}CS_{nm}\right)U_{n-m-1} \tag{A52}$$

$$+ \left(uv_{nm}VV_{nm} - vv_{nm}CS_{nm}\right)U_{n+m-1}\Big\} \tag{A53}$$

The expression for YM is much simpler:

EVEN SHORTER SHORTCUT

The expression for $A_n(x)$ and $B_n(x)$ can be rewritten as

$$A_n(x) = c_nT_n(x) - qs_n\sqrt{1-x^2}U_{n-1}(x) \tag{B1}$$

$$= \left(\frac{c_n - iqs_n}{2}\right)\left(T_n(x) + \sqrt{x^2-1}U_{n-1}(x)\right) + \tag{B2}$$

$$\left(\frac{c_n + iqs_n}{2}\right)\left(T_n(x) - \sqrt{x^2-1}U_{n-1}(x)\right) \tag{B3}$$

$$B_n(x) = \left(\frac{s_n + iqc_n}{2}\right)\left(T_n(x) + \sqrt{x^2-1}U_{n-1}(x)\right) + \tag{B4}$$

$$\left(\frac{s_n - iqc_n}{2}\right)\left(T_n(x) - \sqrt{x^2-1}U_{n-1}(x)\right) \tag{B5}$$

The Chebyshev polynomials satisfy the following property:

$$T_n(x) + \sqrt{x^2-1}U_{n-1}(x) = (x + \sqrt{x^2-1})^n \tag{B6}$$

From which we can also derive that

$$T_n(x) - \sqrt{x^2-1}U_{n-1}(x) = (x + \sqrt{x^2-1})^{-n}. \tag{B7}$$

This provides a convenient change of variables

$$y = x + \sqrt{x^2-1} \tag{B8}$$

with which to rewrite A_n and B_n

$$\begin{aligned}
A_n(y) &= \left(\frac{c_n - iqs_n}{2}\right) y^n + \left(\frac{c_n + iqs_n}{2}\right) y^{-n} && \text{(B9)}\\
&= \alpha_n y^n + \alpha_n^* y^{-n} && \text{(B10)}\\
B_n(y) &= \left(\frac{s_n + iqc_n}{2}\right) y^n + \left(\frac{s_n - iqc_n}{2}\right) y^{-n} && \text{(B11)}\\
&= (iq)\left(\alpha_n y^n - \alpha_n^* y^{-n}\right) && \text{(B12)}\\
A_n A_m &= \left(\alpha_n y^n + \alpha_n^* y^{-n}\right)\left(\alpha_m y^m + \alpha_m^* y^{-m}\right) && \text{(B13)}\\
&= \alpha_n \alpha_m y^{n+m} + \alpha_n^* \alpha_m y^{m-n} + \alpha_n \alpha_m^* y^{n-m} + \alpha_n^* \alpha_m^* y^{-(n+m)} && \text{(B14)}\\
A_n B_m &= (iq)\left(\alpha_n y^n + \alpha_n^* y^{-n}\right)\left(\alpha_m y^m - \alpha_m^* y^{-m}\right) && \text{(B15)}\\
&= (iq)\left\{\alpha_n \alpha_m y^{n+m} + \alpha_n^* \alpha_m y^{m-n} - \alpha_n \alpha_m^* y^{n-m} - \alpha_n^* \alpha_m^* y^{-(n+m)}\right\} && \text{(B16)}\\
B_n B_m &= -\left(\alpha_n y^n - \alpha_n^* y^{-n}\right)\left(\alpha_m y^m - \alpha_m^* y^{-m}\right) && \text{(B17)}\\
&= -\left\{\alpha_n \alpha_m y^{n+m} - \alpha_n^* \alpha_m y^{m-n} - \alpha_n \alpha_m^* y^{n-m} + \alpha_n^* \alpha_m^* y^{-(n+m)}\right\} && \text{(B18)}
\end{aligned}$$

Now we have that

$$\begin{aligned}
MM_{nm} &= A_n A_m CC_{nm} + 2A_n B_m CS_{nm} + B_n B_m SS_{nm} && \text{(B19)}\\
&= \alpha_n \alpha_m \left(CC_{nm} + 2(iq)CS_{nm} - SS_{nm}\right) y^{n+m} && \text{(B20)}\\
&\quad + 2\alpha_n \alpha_m^* \left(CC_{nm} - \frac{iq}{2}\left(CS_{nm} - CS^T_{nm}\right) - SS_{nm}\right) y^{n-m} && \text{(B21)}\\
&\quad + (\alpha_n \alpha_m)^* \left(CC_{nm} - 2(iq)CS_{nm} + SS_{nm}\right) y^{-(n+m)} && \text{(B22)}\\
&= \alpha_n \alpha_m \widetilde{CC}_{nm} y^{n+m} + 2\alpha_n \alpha_m^* \widetilde{CS}_{nm} y^{n-m} + \alpha_n^* \alpha_m^* \widetilde{SS}_{nm} y^{-(n+m)} && \text{(B23)}
\end{aligned}$$

For YM, we can show that

$$\begin{aligned}
YM_k &= A_k YC_k + B_k YS_k && \text{(B24)}\\
&= \alpha_k YC_k y^k + \alpha_k^* YC_k y^{-k} + (iq)\left(\alpha_k YS_k y^k - \alpha_k^* YS_k y^{-k}\right) && \text{(B25)}\\
&= (YC_k + iqYS_k)\alpha_k y^k + (YC_k - iqYS_k)\alpha_k^* y^{-k} && \text{(B26)}\\
&= \alpha_k \widetilde{YC}_k y^k + \alpha_k^* \widetilde{YC}^*_k y^{-k} && \text{(B27)}
\end{aligned}$$

since $n, m, k > 0$, we also have that

$$\begin{aligned}
\partial YM_k &= k\left(\alpha_k \widetilde{YC}_k y^{k-1} - \alpha_k^* \widetilde{YC}^*_k y^{-(k+1)}\right) && \text{(B28)}\\
\partial MM_{nm} &= (n+m)\alpha_n \alpha_m \widetilde{CC}_{nm} y^{n+m-1} + 2(n-m)\alpha_n \alpha_m^* \widetilde{CS}_{nm} y^{n-m-1} && \text{(B29)}\\
&\quad - (n+m)\alpha_n^* \alpha_m^* \widetilde{SS}_{nm} y^{-(n+m+1)} && \text{(B30)}
\end{aligned}$$

and thus the final polynomial expression for y is

$$\begin{aligned}
&2k\left(\alpha_k \widetilde{YC}_k y^{k-1} - \alpha_k^* \widetilde{YC}^*_k y^{-(k+1)}\right)\Big(\alpha_n \alpha_m \widetilde{CC}_{nm} y^{n+m} && \text{(B31)}\\
&\qquad +2\alpha_n \alpha_m^* \widetilde{CS}_{nm} y^{n-m} + \alpha_n^* \alpha_m^* \widetilde{SS}_{nm} y^{-(n+m)}\Big) && \text{(B32)}\\
&- \left(\alpha_k \widetilde{YC}_k y^k + \alpha_k^* \widetilde{YC}^*_k y^{-k}\right)\Big((n+m)\alpha_n \alpha_m \widetilde{CC}_{nm} y^{n+m-1} && \text{(B33)}\\
&\qquad +2(n-m)\alpha_n \alpha_m^* \widetilde{CS}_{nm} y^{n-m-1} - (n+m)\alpha_n^* \alpha_m^* \widetilde{SS}_{nm} y^{-(n+m+1)}\Big) = 0 && \text{(B34)}
\end{aligned}$$

$$2k\Big(\alpha_k\widetilde{YC}_k y^{H+k} - \alpha_k^*\widetilde{YC}_k^* y^{H-k}\Big)\Big(\alpha_n\alpha_m\widetilde{CC}_{nm}y^{2H+n+m} \tag{B35}$$

$$+2\alpha_n\alpha_m^*\widetilde{CS}_{nm}y^{2H+n-m} + \alpha_n^*\alpha_m^*\widetilde{SS}_{nm}y^{2H-n-m)}\Big) \tag{B36}$$

$$-\Big(\alpha_k\widetilde{YC}_k y^{H+k} + \alpha_k^*\widetilde{YC}_k^* y^{H-k}\Big)\Big((n+m)\alpha_n\alpha_m\widetilde{CC}_{nm}y^{2H+n+m} \tag{B37}$$

$$+2(n-m)\alpha_n\alpha_m^*\widetilde{CS}_{nm}y^{2H+n-m} - (n+m)\alpha_n^*\alpha_m^*\widetilde{SS}_{nm}y^{2H-n-m}\Big) = 0 \tag{B38}$$

REFERENCES

Bakos, G., Noyes, R. W., Kovács, G., et al. 2004, PASP, 116, 266 1

Barning, F. J. M. 1963, Bull. Astron. Inst. Netherlands, 17, 22 1, 4

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints, arXiv:1612.05560 1

Cooley, J. W., & Tukey, J. W. 1965, Math. Comput., 19, 297 1

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1 1

Graham, M. J., Drake, A. J., Djorgovski, S. G., et al. 2013, MNRAS, 434, 3423 1

Hernitschek, N., Schlafly, E. F., Sesar, B., et al. 2016, ApJ, 817, 73 1

Jones, E., Oliphant, T., Peterson, P., et al. 2001–, SciPy: Open source scientific tools for Python, [Online; accessed 2017-01-19] 3

Keiner, J., Kunis, S., & Potts, D. 2009, ACM Trans. Math. Softw., 36, 19:1 1, 3

Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, ApJ, 788, 33 1

Kovács, G., Zucker, S., & Mazeh, T. 2002, A&A, 391, 369 2

Leroy, B. 2012, A&A, 545, A50 1

Lomb, N. R. 1976, Ap&SS, 39, 447 1, 4

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints, arXiv:0912.0201 1

Palmer, D. M. 2009, ApJ, 695, 496 2

Press, W. H., & Rybicki, G. B. 1989, ApJ, 338, 277 1

Scargle, J. D. 1982, ApJ, 263, 835 1, 4

Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, ApJ, 708, 717 1, 4

Sesar, B., Hernitschek, N., Mitrović, S., et al. 2016, ArXiv e-prints, arXiv:1611.08596 1, 2.2.3, 2.2.3, 4

Stoica, P., Li, J., & He, H. 2009, IEEE Transactions on Signal Processing, 57, 843 1, 1

VanderPlas, J. 2016, gatspy: General tools for Astronomical Time Series in Python, Astrophysics Source Code Library, ascl:1610.007 4

VanderPlas, J. T., & Ivezić, Ž. 2015, ApJ, 812, 18 2.2.1

Vaníček, P. 1971, Ap&SS, 12, 10 1, 4

Zinn, J. C., Kochanek, C. S., Kozłowski, S., et al. 2016, ArXiv e-prints, arXiv:1612.04834 1