A toolbox for analysing harmonic signals

David Lawunmi¹ and Soodamani Ramalingam²

1 QMUL (London), Room W209 Queens' Building Mile End Rd, London E1 4NS, UK

2 Centre for Engineering Research, School of Engineering and Technology, University of Hertfordshire (UH), Hatfield, Hertfordshire, United Kingdom, AL10 9AB

Abstract

We develop a toolbox approach to analysing harmonic signals. The methodology that is pursued in this discussion is based on interpolating signals with Chebyshev polynomials. This approach provides the basis for transforming the initial input signal into another signal based on the input by using the analytic properties of harmonic functions and Chebyshev polynomials. This methodology has significant benefits when attempting to analyse difficult problems such as characterising harmonic signals where data has been collected over a fraction of a cycle, and it has significant ramifications for the analysis of noise and the measurement uncertainties that are associated with a signal.

Introduction

A toolbox approach to analysing exponential and harmonic signals can be developed by interpolating a signal of interest with Chebyshev polynomials. This approach provides a mechanism for transforming the input signal into a form that has properties that are amenable to analysis that allow the data to be explored for properties that are consistent with a signal of interest. In addition this 'toolbox approach' provides a means for extracting numerical estimates for the parameters of interest.

We apply this approach to a technique that attempts to characterise a 'harmonic' signal from data that spans a fraction of a cycle. In this study we exploit the properties of Chebyshev polynomials to develop an analytic expression for a harmonic signal and we develop tests to ascertain that the parameters that are obtained for the harmonic signal are consistent with the data.

Another factor that is important when analysing signals is the impact of noise and uncertainty on the parameter estimates. In this study we adopt a simple ‘Monte Carlo’ approach and we investigate the impact of noise on the signal by looking at the impact of normally distributed noise on harmonic signals. This approach can be extended to other noise distributions and as it is analytic, it can readily be integrated into more sophisticated statistical approaches such as the Bayesian approach. The starting point is a technique for extracting information about a harmonic signal from noisy environment when the signal spans a fraction of a cycle, Lawunmi 1995[1].

The technique exploits the properties of harmonic functions, consider a signal of the form $f(t) = A \cos \omega t + n(t) + C$, where $n(t)$ is a stochastic function that represents noise, C and A are constants. Multiply this signal by $\cos \omega' t$. Where the angular frequency ω' is treated as a variable parameter.

The noise free case where $n(t) = 0$, and $f(t)$ is integrated between the upper limit x_{UL} and the lower limit x_{LL} , where the lower limits and upper limits are symmetrically distributed about a point of odd symmetry of $\cos \omega t$. This can be expressed as

$$\text{Integral} = \int_{x_{LL}}^{x_{UL}} dt (\cos \omega' t (A \cos \omega t + n(t) + C)) \quad (1)$$

When this is divided by $\frac{1}{2}(x_{UL} - x_{LL}) + \frac{1}{4}[\sin(2x_{UL}\omega) - \sin(2x_{LL}\omega)]$
In general the resulting expression

$$\begin{aligned} I &= \frac{\int_{x_{LL}}^{x_{UL}} dt (A \cos \omega t + C) \cos \omega' t}{\frac{1}{2}(x_{UL} - x_{LL}) + \frac{1}{4\omega}[\sin(2x_{UL}\omega) - \sin(2x_{LL}\omega)]} \\ &= \frac{\int_{x_{LL}}^{x_{UL}} dt A \cos \omega t \cos \omega' t}{\frac{1}{2}(x_{UL} - x_{LL}) + \frac{1}{4\omega}[\sin(2x_{UL}\omega) - \sin(2x_{LL}\omega)]} \end{aligned} \quad (2)$$

ω' can be treated as a parameter, and the integral can be evaluated for a set of the lower limit of the integral and the upper limit of the integral,

$\{x_{UL}, x_{LL}\}$. The integral will tend to fluctuate over a wide range, when the integral is evaluated over the set of upper limits and lower limits, e.g. $\{x_{UL}, x_{LL}\}$ that are symmetrically distributed around a point of odd symmetry of $\cos \omega t$. However when $\omega = \omega'$, the integral will be constant (or ‘close’ to constant) over a set of values of upper and lower limits that are symmetrically distributed around a point of odd symmetry of $\cos \omega t$. According to the analysis in Lawunmi (1995)[1], when the effects of noise are included, equation 2 can be fairly robust, and if $\omega \sim \omega'$ the value of expression (2) is ‘approximately constant’ when it is evaluated over a set of values of upper and lower limits that are symmetrically distributed around a point of odd symmetry of $\cos \omega t$. This can be a useful way of identifying ω even when the signal to noise ratio is fairly low. The corresponding expression for $\sin \omega t$ that are distributed symmetrically about a point of odd symmetry of $\sin \omega t$ is given by

$$\begin{aligned}
J &= \frac{\int_{x_{LL}}^{x_{UL}} dt (A \sin \omega t + C) \sin \omega' t}{\frac{1}{2}(x_{UL} - x_{LL}) + \frac{1}{4\omega} [\sin(2x_{LL}\omega) - \sin(2x_{UL}\omega)]} \\
&= \frac{\int_{x_{LL}}^{x_{UL}} dt A \sin \omega t \sin \omega' t}{\frac{1}{2}(x_{UL} - x_{LL}) + \frac{1}{4\omega} [\sin(2x_{LL}\omega) - \sin(2x_{UL}\omega)]}
\end{aligned} \tag{3}$$

Interpolation

The data is interpolated using Chebyshev polynomials. The underlying ideas are similar to those used in the NAG Fortran subroutine **E02ADF**. We implemented the Chebyshev interpolation code in the Java programming language with the algorithm presented in Mason (1984)[2], and Powell (1981)[3].

It is possible to obtain a set of basis functions with the property that satisfies the orthogonality relationship

$$\sum_{i=1}^m g_j(x_i) g_k(x_i) = \begin{cases} 1 & j=k \\ 0 & j \neq k \end{cases} \tag{4}$$

for a set of data points $\{x_i\}$. The polynomials can be obtained by making use of the following relationships, Mason(1984)[2], Powell (1981)[3]

$$\begin{aligned} g_1(x) &= 1 \\ g_{j+1}(x) &= (x - B_j)g_j(x) + C_j \\ g_{j+1}(x) &= (x - B_j)g_j(x) + C_j g_{j-1}(x) (j = 1, 2, \dots) \end{aligned} \quad (5)$$

where C_j is a constant.

Interpolation of data points is a useful and robust way of obtaining a good approximation of a data set that correspond to a discrete set of points in terms of a continuous function. It provides us with a way of dealing with a data set corresponding to a set of measurements corresponding to a signal representing a harmonic signal in terms of a continuous function.

Clenshaw [4] developed a modification of the basic approach illustrated above in terms of the generalised orthogonal polynomials. This involved expressing the polynomials in terms of Chebyshev polynomials. This is significant as it has benefits in terms of storage efficiency and the favourable stability properties that Chebyshev polynomials have over the general polynomials $g_i(x)$, when used as a basis set to approximate a function. Chebyshev series have some useful properties that are helpful when evaluating the product of two Chebyshev series, or the product of an interpolated data set and a series that is used to approximate an analytic function. It's worth noting that the product of two Chebyshev series can be expressed as a Chebyshev series, the following expression can be used to generate the terms of a Chebyshev series that arises when two Chebyshev series are multiplied

$$2T_m(x)T_n(x) = T_{m+n}(x) + T_{|m-n|}(x) \quad (6)$$

Gil et al (2008),[5].

Analysing a harmonic signal using data from a fraction of a cycle

Some tables are generated to illustrate the ideas behind extracting parameters that characterise a sinusoidal signal from a series measurements. The general form of the data is

$$y = A \sin \omega t + C + n(t) \quad (7)$$

where

- A is the amplitude of a harmonic signal
- ω represents the frequency
- $n(t)$ is a function that represents random noise
- C is a constant

In this analysis the noise is Gaussian (normally distributed).

The data points were measured at times that were obtained from a randomised grid, this is more general than equally spaced data and it helps to illustrate the benefits of interpolation as having a Chebyshev expansion for the integral permits evaluating the integral at points that are symmetrically distributed about a point of odd symmetry, even if the data isn't collected at pairs of points that match up with respect to a reference point of odd symmetry.

The algorithm for randomisation involved creating a series of equispaced points between a starting point and an end point. Other than the starting point and the end point, the remaining points are determined using the uniform distribution. The first point and the last point are retained. The intervening points are obtained from the set of equidistant points, starting with the first point x_1 and the second point x_2 in the equispaced set of points, they are used as the lower limit and the upper limit respectively for selecting a random number using the uniform distribution. The uniform distribution used in this analysis is the Java implementation in the 'java.util.Random' class. This process was used to generate the $[x_{LL}, x_{UL})$ that appeared in expression (3) and (2)

As the method for the uniform distribution in the class `java.util.Random` selects a number from a 'half open interval' $[x_{LL}, x_{UL}]$ when generating a random variable for the uniform distribution, we prevented the very small possibility of having two identical points for x_{LL} and x_{UL} by including a checking condition in the code that added a small number corresponding to machine precision if the random number generator returned x_{LL} . This process for selecting points generates $n + 1$ points, including the two endpoints of the initial set of equispaced points.

I've enclosed the key parameters that are associated with a harmonic signal of the form $A \sin \omega t + C + n(t)$, as given in expression (7), in the table below

Key parameters

Parameter	Value
Upper Limit	1.0 unit
Lower Limit	-1.0 unit
Number of points	45
C see equation 7	100.0 unit
ω see equation 7	1.0 unit
A see equation 7	100.0 unit

In the table above the terms C , A , and ω are defined in equation 7. As this analysis is based on a simulation I've kept the unit of measurement general and used the generic term 'unit' to encompass the various appropriate units for each context where a unit of measurement is required.

The first data analysis study is generated with a negligible level of noise (close to zero noise). The ratio defined by expression (3) is constant whenever the integral is evaluated with an upper limit and lower limit that are symmetrically distributed about the point of '*odd symmetry*' $x = 0$, as expected.

Table: Close to zero noise

Upper limit (radians)	Lower limit (radians)	ratio
0.0490	-0.0490	1.0000
0.0867	-0.0867	1.0000
0.1379	-0.1379	1.0000
0.1723	-0.1723	1.0000
0.2472	-0.2472	1.0000
0.2993	-0.2993	1.0000
0.3220	-0.3220	1.0000
0.3661	-0.3661	1.0000
0.4290	-0.4290	1.0000
0.4850	-0.4850	1.0000
0.5231	-0.5231	1.0000
0.5685	-0.5685	1.0000
0.6059	-0.6059	1.0000
0.6410	-0.6410	1.0000
0.6949	-0.6949	1.0000
0.7425	-0.7425	1.0000
0.7896	-0.7896	1.0000
0.8495	-0.8495	1.0000
0.8933	-0.8933	1.0000
0.9203	-0.9203	1.0000
0.9778	-0.9778	1.0000
1.0000	-1.0000	1.0000

The standard deviation of the noise is 1×10^{-10} units

Ramping up the noise a little to a ‘low’ level of noise $\sigma = 1 \times 10^{-2}$ units has an impact on the ratio, but it’s still very close to 1.0000 for all of the pairs of points for which measurements were taken, (including pairs of points that are relatively close to the point of odd symmetry).

Table: Low level of noise

Upper limit (radians)	Lower limit (radians)	ratio
0.0067	-0.0067	1.0009
0.0689	-0.0689	1.0005
0.1100	-0.1100	1.0000
0.1397	-0.1397	0.9996
0.1844	-0.1844	0.9993
0.2433	-0.2433	0.9996
0.2578	-0.2578	0.9999
0.3070	-0.3070	1.0013
0.3941	-0.3941	1.0049
0.4352	-0.4352	1.0063
0.4862	-0.4862	1.0073
0.5045	-0.5045	1.0073
0.5761	-0.5761	1.0058
0.5931	-0.5931	1.0051
0.6424	-0.6424	1.0029
0.7066	-0.7066	1.0006
0.7432	-0.7432	1.0002
0.8073	-0.8073	1.0011
0.8463	-0.8463	1.0023
0.8936	-0.8936	1.0036
0.9354	-0.9354	1.0043
0.9767	-0.9767	1.0045
1.0000	-1.0000	1.0047

The standard deviation of the noise is 1×10^{-2} units

Ramping up the noise further by increasing the standard deviation to $\sigma = 1 \times 10^{-1}$ units, lets describe this as a ‘moderate’ level of noise, has an impact on the ratio, however it is still possible to see a strong pattern for points that are around 0.3759 units or more from the point of odd symmetry.

Table: Moderate level of noise

Upper limit (radians)	Lower limit (radians)	ratio
0.0102	-0.0102	1.8524
0.0393	-0.0393	1.8378
0.1068	-0.1068	1.7442
0.1166	-0.1166	1.7249
0.1856	-0.1856	1.5663
0.2424	-0.2424	1.4296
0.3009	-0.3009	1.3092
0.3321	-0.3321	1.2592
0.3619	-0.3619	1.2214
0.4092	-0.4092	1.1803
0.4822	-0.4822	1.1495
0.4955	-0.4955	1.1464
0.5391	-0.5391	1.1380
0.5979	-0.5979	1.1267
0.6507	-0.6507	1.1153
0.6857	-0.6857	1.1083
0.7527	-0.7527	1.0982
0.8024	-0.8024	1.0915
0.8302	-0.8302	1.0864
0.8656	-0.8656	1.0779
0.9125	-0.9125	1.0657
0.9667	-0.9667	1.0612
1.0000	-1.0000	1.0586

The standard deviation of the noise is 1×10^{-1} units

. The noise has a bigger impact here especially for points that are close to the central point of odd symmetry. Adopting a cut-off point for points that are close to the origin, is a way of looking for promising matches.

This is all well and good, but there are some outstanding questions, such as what impact does the noise have on the ratio defined by equation (3)? I'll illustrate this by considering the spread in the value of the ratio defined by equation ((3)) when the signal consists of noise.

In the figures the data bins have a width of 0.1 units, they measure the value of the ratio defined by expression (3). A simple calculation was

undertaken to look at the size of the ratio that is generated when the signal consists of normally distributed noise, the experiment was repeated several times in order to get an estimate of the probability that noise with specific standard deviation (the mean was set to zero as the integral of a constant signal using expression (3) is zero).

For the first figure the noise level is very low, the standard deviation is 10^{-2} units, the ratio obtained for 100 simulated calculations lies in two bins centred at $1 = 0^{-2}$ units, 45 are in the range $-0.1 \leq \text{ratio} < 0$ and 55 are in the range $0.0 \leq \text{ratio} < 0.1$

Increasing the noise results in a bigger spread of values, the results for a simulation when the standard deviation is 10^{-1} units are placed in the table below.

Table: Moderate level of noise

Upper limit (radians)	Lower limit (radians)	ratio
lower limit	upper limit	number of values
-0.6	-0.5	1
-0.5	-0.4	4
-0.4	-0.3	4
-0.3	-0.2	5
-0.2	-0.1	15
-0.1	0.0	22
0.0	0.1	27
0.1	0.2	10
0.2	0.3	7
0.3	0.4	4
0.4	0.5	1

Increasing the standard deviation has resulted in a bigger spread in the values of the ratio defined by expression (3), the larger values could cause a problem for a signal with an amplitude of the order of 1 unit, e.g. a misleading estimate of an amplitude or a big spread in the ratio at the points at which it has been estimated and this could lead to the correct value for the frequency being rejected in favour of some other value.

Analysing 23rd point, Normal distribution $\sigma=0.1$, $\mu=0.0$

One question that arises with this approach is what happens to the ratio defined in equation (3) when the angular frequencies ω and ω' are different?

In Lawunmi (1995) [1]), a scheme was developed to determine the frequency of a harmonic from data that spanned a fraction of a cycle. This approach used simple numerical integration to estimate the ratio given by equation (3). In this study we expect greater accuracy due to our use of Chebyshev polynomials and the formula for integrating Chebyshev polynomials to generate the Chebyshev series of the integral

$$\int T_n dx = \frac{1}{2} \left(\frac{T_{n+1}}{n+1} - \frac{T_{n-1}}{n-1} \right) \quad (8)$$

In practice based on the analytic properties of the ratio, if the noise level is very low or zero there should be a noticeable difference, with the caveat that this difference should become very small as $\omega - \omega' \rightarrow 0$.

I have illustrated this in the table where ω is ‘fairly close’ to ω . In the table the fixed frequency $\omega = 1.0$ unit and the varying test frequency $\omega' = 0.9$ units.

Table: Close to zero noise

Upper limit (radians)	Lower limit (radians)	ratio
0.0225	-0.0225	0.9000
0.0302	-0.0302	0.9000
0.0806	-0.0806	0.9001
0.1532	-0.1532	0.9004
0.1696	-0.1696	0.9005
0.2213	-0.2213	0.9008
0.2768	-0.2768	0.9013
0.3220	-0.3220	0.9018
0.3951	-0.3951	0.9027
0.4198	-0.4198	0.9030
0.4808	-0.4808	0.9040
0.5158	-0.5158	0.9046
0.5664	-0.5664	0.9055
0.6089	-0.6089	0.9064
0.6619	-0.6619	0.9075
0.6953	-0.6953	0.9083
0.7630	-0.7630	0.9100
0.7813	-0.7813	0.9105
0.8318	-0.8318	0.9120
0.8904	-0.8904	0.9137
0.9312	-0.9312	0.9150
0.9742	-0.9742	0.9165
1.0000	-1.0000	0.9174

The standard deviation of the noise is 1×10^{-10} units, $\omega = 1.0$ unit and the varying test frequency $\omega' = 0.9$ units.

Predictions, properties and testing

The fitting process has mathematical consequences. The mathematical properties of harmonic functions and Chebyshev polynomials can be used to test the parameters that are obtained from the ratio analysis.

Consider a data set of the form $\sin(\Omega t)$. If this data set is multiplied a time series $2 \cos(\phi t)$, e.g. by multiplying their respective Chebyshev series, e.g.

by using expression (6). This is a way of generating a time series that spans a larger fraction of the period of the signal than the initial input time series. From the perspective of the ratio analysis technique, this process could be the basis of testing the output as it enables a prediction to be made about a time series that is generated by multiplying the initial data set by a harmonic time series with a known frequency ϕ . This process generates a signal $s(t)$ of the form

$$s(t) = 2 \sin(\Omega t) \cos(\phi t) = \sin([\Omega + \phi]t) + \sin([\Omega - \phi]t) \quad (9)$$

If $\phi = \Omega$ the resulting time series will have the form $\sin(2\Omega t)$. In practice it is worth noting that the estimated frequency is more likely to be close to the correct frequency as opposed to being exactly the same as the frequency of the input signal, the attempt at doubling the signal frequency can be viewed as an approximation, e.g. if Ω is the correct frequency, and $\Omega + \Delta$ is the estimated frequency, multiplying $\sin \Omega t$ by $2 \cos(\Omega + \Delta)t$ yields a signal $s'(t)$ of the form

$$s'(t) = \sin(2\Omega + \Delta)t - \sin \Delta t \quad (10)$$

In principle this can be repeated, if Δ isn't too big, and noise isn't too big a factor, a single harmonic should be a good approximation for the new signal $s'(t)$. However if Δ is 'large', the process may result in two new harmonics and the approximation by a single harmonic will break down. Another angle on prediction arises from the property of the Chebyshev expansion of a harmonic function. They can be expressed analytically, e.g.

$$\cos(at) = J_0(a)T_0(t) + 2 \sum_{n=1}^{\infty} (-1)^n J_{2n}(a)T_{2n}(t) \quad (11)$$

and

$$\sin(at) = 2 \sum_{n=0}^{\infty} (-1)^n J_{2n+1}(a)T_{2n+1}(t) \quad (12)$$

Snyder (1966)[6]. These expressions are potentially useful, as they can be used for a direct comparisons with one or more harmonic signals with (known or estimated) amplitudes and signals. This could involve comparing the value of the coefficient of a Chebyshev polynomial of some order that has been generated from a signal with the theoretical value corresponding

to a harmonic signal with a known or estimated angular frequency and amplitude. It's also possible to use the properties of Bessel functions such as zeros when making comparisons. Low order coefficients of the Chebyshev expansion are likely be the best for comparisons as they will tend to be least effected by noise.

Inspection of the coefficients of Chebyshev coefficients can provide useful predictive information when it is combined with the angle addition formulae, for example using expression (9) for the sum of two sine functions

$$\begin{aligned} & \sin([\Omega + \phi]t) + \sin([\Omega - \phi]t) \\ &= 2 \sum_{n=0}^{\infty} (-1)^n (J_{2n+1}([\Omega + \phi]) + J_{2n+1}([\Omega - \phi])) T_{2n+1}(t) \quad -1 \leq t \leq 1 \end{aligned} \tag{13}$$

The Bessel functions have zeros, that is values for which $J_n(a) = 0$, thus if either $\Omega + \phi$ or $\Omega - \phi$ is chosen so that it corresponds to a zero of T_{2n+1} for a certain value of n , the n th order Chebyshev coefficient should be determined (approximately for low levels of noise) by a single evaluation of a Bessel function, instead of the usual two evaluations. Thus a zero for $\Omega + \phi$ corresponds to a coefficient determined by evaluating $J_{2n+1}(a)$ for $a = \Omega - \phi$.

Key questions that arise are frequency resolution and the determination of the signal amplitude. The signal amplitude and frequency will be affected by factors such the length of the time interval over which the signal has been measure; the nature and the magnitude of the noise; measurement uncertainty; the number of data points; the spacing of the data points; the presence of outliers points due to factors such as noise and measurement uncertainty.

The ratio in equations (2) and (3) is a constant or approximately constant when ω' the variable angular frequency is equal to or close to the true frequency ω . The spread in the ratio will tend to diverge as the difference between ω and ω' increases.

Generally measuring a signal over a large time span which covers a significant fraction of a cycle (or more than a full cycle), and taking these measurement regularly (i.e. a large number of data points per unit measurement) will tend to help with the frequency resolution. In essence from an information theoretical perspective, a measuring process that captures more information about a signal will tend to provide better

information about the properties of the signal.

Measurement uncertainty reflects the limitations of the measuring processes that are used to determine the signal, e.g. input uncertainties could reflect the limitations on the accuracy of the time of measurement, output uncertainties could reflect the limiting accuracy of the the processes that are used to measure the signal. In essence the difference between error (noise) and uncertainty, are that error is a ‘point estimate of the difference between the estimated and the ‘true’ value of the measured quantity, whilst uncertainty reflects a range of values in which there is strong likelihood of finding the true measurement, Lira (2002)[7]. Ultimately both noise and measurement uncertainty will have a deleterious impact on the measurement of the true value of the signal, and moderate values of these quantities have the potential to make a direct interpretation of an analysis with the ratio approach difficult without resorting to more sophisticated statistical analysis techniques that can model noise and uncertainty such as the use of Bayesian approaches Givens and Hoeting (2013) [8] and D’Agostini (2003), [9].

A key factor with harmonic signals is that they have properties that allow the estimates of their properties to be tested, in addition if the signal frequency isn’t too long they also allow predictions to be made about future measurements of the signal.

I’ve performed a simple comparison of the analytic result of the coefficients of a Chebyshev expansion of a sinusoid with a Chebyshev expansion obtained from computer program that interpolates $\sin ax$ for $a = 1$ over the interval $-1 \leq x \leq 1$. It’s worth noting that the Chebyshev expansion of $\sin x$ is fairly ‘compact’ in the sense that it converges fairly quick, i.e. a small number of terms are required for a ‘fairly’ accurate representation of a harmonic function.

Comparison of the analytic result to the results for $\sin x$

Order	Analytic result	Computer result
1	0.88010117	0.88010117
3	-0.03912671	-0.03912671
5	0.00049952	0.00049952
7	-0.00000300	-0.00000300
9	0.00000001	0.00000001
11	0.00000000	0.00000000
13	0.00000000	0.00000000
15	0.00000000	0.00000000

In the table below I've compared the analytic result for $\sin 2x$ obtained by multiplying the Chebyshev expansion for $\sin x$ by the Chebyshev expansion for $2 \cos x$ using expression (6) to obtain an expression for the product of two Chebyshev series.

Comparison of the analytic result to the results $\sin 2x$

Order	Analytic result	Computer result
1	1.15344962	1.15344962
3	-0.25788650	-0.25788650
5	0.01407926	0.01407926
7	-0.00034989	-0.00034989
9	0.00000498	0.00000498
11	-0.00000005	-0.00000005
13	0.00000000	0.00000000
15	0.00000000	0.00000000

In addition to comparing the analytic result to a noise free interpolation it's useful to look at the impact of noise on the Chebyshev series of the interpolated sinusoid. I've enclosed some Chebyshev expansion coefficients for the function $\sin x + n(x)$ below, where $n(x)$ is normally distributed with a mean of 0.0 units and a standard deviation of 1.0×10^{-2} units. The first column is the analytic result the second column represents the expansion coefficients that were obtained from interpolating $\sin x + n(x)$.

Comparison of the analytic result $\sin x$ to the results $\sin x + n(x)$

Order	Analytic result	Computer result
1	1.15344962	1.15895812
3	-0.25788650	-0.25633695
5	0.01407926	0.01266066
7	-0.00034989	0.00250995
9	0.00000498	-0.00899703
11	-0.00000005	-0.00650552
13	0.00000000	0.00100265
15	0.00000000	-0.00311431

In the table below I've compared the analytic result for $\sin 2x + n'(x)$ obtained by multiplying the Chebyshev expansion for $\sin x + n(x)$ by the Chebyshev expansion for $2 \cos x$ using expression (6) to obtain an expression for the product of two Chebyshev series. The stochastic function $n'(x)$ is the product of the stochastic noise function $n(x)$ with $2 \cos x$.

Comparison of the analytic result to the results $\sin 2x + n(x)$

Order	Analytic result	Computer result
1	0.88010117	0.88461336
3	-0.03912671	-0.03753806
5	0.00049952	-0.00006023
7	-0.00000300	0.00080988
9	0.00000001	-0.00655205
11	0.00000000	-0.00529542
13	0.00000000	-0.00043154
15	0.00000000	-0.00208281

I've used an approach based on a comparison of Chebyshev expansion coefficients from an interpolated data set with the analytic result for illustrative purposes. In the low noise and measurement uncertainty regime it may be appropriate to use the lower order coefficients and to compare them directly with the analytic result. This might be more challenging for data where the noise and measurement uncertainty has a large impact on the signal.

In practice as with any technique that's applied to noisy data a sensible

approach will involve doing some analysis to ascertain that the comparison of the noisy coefficients with the analytic result is appropriate. This could involve formulating the problem robustly in a way that takes account of noise and uncertainty, e.g. using approaches such as the singular value decomposition (SVD), the wavelet transform Patil (2015) [10], and Bayesian analysis D’Agostini (2003) [9], Lira (2002) [7].

An outline of an approach for looking at signals containing more than one harmonic

The Chebyshev expansion is a mathematically versatile way of expressing a harmonic function as it provides a mathematical foundation for using the properties of harmonic functions to transform a harmonic function into a form that is amenable to analysis, (noise and measurement uncertainty permitting). For example, transforming a harmonic series from $\sin x$ to $\sin 2x$. The analytic properties of harmonic functions provide a wide range of useful tools that can potentially be used to transform a signal into a form that is amenable for analysis. For example if a harmonic signal has more than one component, e.g. a signal of the form

$$\sum_{i=1}^n a_i \sin \omega_i t + n(t) \quad (14)$$

where $n(t)$ is a stochastic function that represents the noise
 If it is multiplied by $\sin \omega_0 t$
 and integrated with respect to t we have

$$\sum_{i=1}^n a_i \left(\frac{\sin(\omega_i - \omega_0)t}{2(\omega_i - \omega_0)} - \frac{\sin(\omega_i + \omega_0)t}{2(\omega_i + \omega_0)} \right) + n'(t) + \xi \quad (15)$$

where $n'(t)$ is a stochastic function representing the noise, and ξ is a constant and can be set equal to zero, (as we are dealing with sine functions which are asymmetric about the origin). This process can be used to try to enhance the contribution to the overall signal from one harmonic whilst diminishing the others in order to get a signal that is dominated by a single frequency, noise and measurement uncertainty permitting.

Discussion

A study was undertaken to develop a method for analysing a data set to ascertain whether a harmonic signal is present when the data is collected over a time interval that is less than the time it takes for the signal to complete a full cycle. The approach is based on interpolation with Chebyshev polynomials together with some basic noise and false alarm analysis. There are many benefits to using interpolation based on a expansion in terms of Chebyshev polynomials. This approach makes it straightforward to separate odd components of the harmonic signal (sine), which correspond to the odd order terms in the Chebyshev expansion from the even components (cosine) which correspond to even order terms in the Chebyshev expansion. From looking at the tables of the Chebyshev expansion of $\sin ax$, the expansion is fairly ‘compact’ in the sense that a small number of terms can approximate $\sin ax$ fairly accurately.

This observation can be explained in terms of the mathematical properties of Bessel functions. According to Cooke (1937), [11]

“the graph of the Bessel function $J_\nu(x)$ ($\nu > -1, x > 0$) consists of waves alternately above and below the axis of x , whose areas $A_1, A_2, A_3, \dots$ (in order) form a steadily decreasing sequence”.

As Bessel functions are oscillatory, there isn’t a straightforward monotonic relationship that describes the dependency of magnitude of the value of $J_\nu(x)$ on the order of the Bessel function. In addition for a specific argument x' and order ν' $J_{\nu'}(x')$ might be zero, for some values of x' . However, there are some mathematical expressions that place some bounds on $J_\nu(x)$. For example according to Cerone (2007), [12]

$$|J_\nu(x)| \leq \left| \frac{z}{2} \right|^\nu \frac{1}{\Gamma(\nu + 1)} \quad (16)$$

Noting that when ν is an integer, the gamma function can be expressed in terms of a factorial $\Gamma(\nu + 1) = \nu!$ and

$$\lim_{\nu \rightarrow +\infty} \left(\frac{x^\nu}{\nu!} \right) = 0 \quad (17)$$

which is proved in Binmore (1977), exercise §4.20 (4), [13].

A similar somewhat less strict bound was given in Watson (1966) p. 16, [14].

$$|J_\nu(x)| \leq \left|\frac{z}{2}\right|^\nu \frac{1}{\nu!} \exp\left(\frac{1}{4}|z^2|\right) \quad (18)$$

It follows that

$$\frac{|J_{\nu+m}(x)|}{|J_\nu(x)|} \leq \left|\frac{z}{2}\right|^m \frac{\nu!}{(\nu+m)!} \quad (J_\nu(x) \neq 0) \quad (19)$$

The bounds on the Bessel functions suggest for some order ν , and a constant K that if $|J_\nu(x)| > |K|$, there is an integer $m > 0$

$$|J_\nu(x)| > |K| > |J_{\nu'}(x)| \quad \forall \nu' \geq \nu + m \quad (20)$$

The nature of the convergence based on expression (19) will depend on factors such as the choice of order, ν selected so that for ν , $|J_\nu(x)| > |K|$, the value of x . Additionally the convergence of the expansion coefficients of the harmonic functions $\sin x$ and $\cos x$ in expressions (12) and (11) increase in steps of 2.

However irrespective of the specific set of circumstances, expressions (19) and (20) provide a mathematical description of the dependence of the convergence of the Chebyshev coefficients on the order ν' .

Overall from a practical standpoint, it's desirable that the Chebyshev expansion is 'compact', as an expansion that depends on several small high order terms is likely to be prone to inaccuracy as the small higher order coefficients are more likely to be dominated by noise than a small set of low order expansion coefficients.

This analysis involved developing an approach to analysing a noisy signal to ascertain whether a harmonic signal is present over time spans that could less than a full cycle based on interpolation with Chebyshev polynomials. It involved comparing the analytic result with an integral obtained from the signal data. This approach to harmonic analysis applies to cosine functions, expression (2) and to sine functions expression (3).

The effectiveness of this approach to determine the frequency and the amplitude of a harmonic is affected by factors such the length of the time interval over which the signal has been measured; the nature and the magnitude of the noise; measurement uncertainty; the number of data points; the spacing of the data points; the presence of outliers points.

Generally measuring a signal over a large time span which covers a significant fraction of a cycle (or more than a full cycle), and taking these measurement regularly (i.e. a large number of data points per unit measurement) will tend to help with the frequency resolution. However one of the benefits of this approach is that it can be applied to analyse data sets where the measurements span an interval that is less than the full cycle of the underlying harmonic signal. It can be used as the basis for a direct interpretation and estimate of parameters for small to moderate levels of noise, when supported with an appropriate false alarm analysis. For larger levels of noise an approach that models the noise and the measurement uncertainty such as Bayesian statistical techniques could be a way of extending this approach to extract quantitative information from a signal of interest.

In addition to making estimates of the key parameters that characterise a harmonic signal, this approach provides a variety of ways to test the estimates of the parameters of the signal. There are some options for constructing tests to validate the results of an analysis. Tests can make direct use of the properties of harmonic functions and the analytic expressions connected with the Chebyshev expansions of harmonic functions such as expression (11) and expression (12).

Conclusion

We have developed an approach for analysing data to ascertain whether a harmonic signal is present and applied it to look at noisy data sets where the period of the signal is greater than the time interval over which the signal data was collected. This approach involved interpolating the signal with Chebyshev polynomials. It can be thought of as a *‘toolbox approach’* for analysing harmonic signals, as it provides the basis for a variety of techniques for combining the original data set with harmonic functions to generate another signal that has properties that are more conducive for extracting the parameters of interest. Another benefit of this approach is that the mathematical properties of harmonic functions and Chebyshev polynomials can be used to test the values of the parameters of the harmonic function that are obtained from a fraction of a cycle. In effect the estimates of the parameters should be consistent with the underlying

interpolation and the fundamental properties of harmonic functions.

References

- [1] Lawunmi D. Extracting keplerian signals from noisy data. *Journal of Physics A: Mathematical and General*, 28:3269– 3278, 1995.
- [2] Mason J. C. *BASIC matrix methods: including applications in approximation and data fitting*. Butterworths BASIC series, 1984.
- [3] Powell M. J. D. *Approximation Theory and Methods*. Cambridge University Press, 1981.
- [4] Clenshaw C. W. Curve fitting with a digital computer. *Comput. J.*, 2:170–173, 1960.
- [5] Gil A., Segura J., and Temme N. *Numerical Methods for Special Functions*. Society for Industrial and Applied Mathematics, 2008.
- [6] Snyder M. A. *Chebyshev methods in numerical approximation*. Prentice-Hall, 1966.
- [7] Lira I. *Evaluating the Measurement Uncertainty: Fundamentals and Practical Guidance*. CRC Press, 2002.
- [8] Givens G. H. and Hoeting J. A. *Computational Statistics, Second Edition*. John Wiley & Sons, 2013.
- [9] D’Agostini G. *Bayesian Reasoning in Data Analysis: A Critical Introduction*. World Scientific Publishing, 2003.
- [10] Patil R. Noise reduction using wavelet transform and singular vector decomposition. *Procedia Computer Science*, 54:849 – 853, 2015.
- [11] Cooke R. G. A monotonic property of bessel functions. *Journal of the London Mathematical Society*, s1-12:180– 185, 1937.
- [12] Cerone P. Special functions: Approximations and bounds. *Applicable Analysis and Discrete Mathematics*, 1:72–91, 2007.

- [13] Binmore K. G. *Mathematical Analysis (A straightforward approach)*. Cambridge University Press, 1977.
- [14] Watson G. N. *A Treatise on the Theory of Bessel Functions, Second Edition*. Cambridge Cambridge University Press, 1966.