An Online Algorithm for Least-Square Spectral Analysis: Applied to Time-Frequency Analysis of Heart Rate

Zhe Zhang, *Student Member, IEEE,* and Philip H.W. Leong, *Senior Member, IEEE*

***Abstract*— We propose a novel online algorithm for computing least-square based periodograms, otherwise known as the *Lomb-Scargle Periodogram*. Our spectral analysis technique has been shown to be superior to traditional discrete Fourier transform (DFT) based methods, and we introduce an algorithm which has $O(N)$ time complexity per input sample. The technique is suitable for real-time embedded implementations and its utility is demonstrated through an application to the high resolution time-frequency domain analysis of heart rate variability (HRV).**

I. Introduction

Heart rate variability (HRV), a measure of the variation in the period of the cardiac cycle, has been shown to reliably indicate physiological factors regulating the heart rate and in turm reflect a number of associated medical conditions [1], [2], [3]. The Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology has published an article [1] where standard measurements and interpretations of HRV were specified in detail. According to this literature, frequency-domain measurements of the HRV mainly involves spectral power analysis of the instantaneous heart rate (HR) signal, or equivalently its reciprocal: the beat-to-beat (RR) interval series.

Because samples of the HR or RR time series are generated at the peak of each R wave, they are inherently irregularly spaced in time. Traditional spectral analysis based on the Discrete Fourier Transform (DFT) requires resampling of the original time series to regular time intervals and such interpolation-based resampling has been shown to introduce systematic errors, leading to inaccuracies [4], [5]. In addition, reference [5] showed such errors significantly increase with the number of nonsinus heart beats due to ectopic depolarization, as these outlying samples must be removed [1].

To circumvent these problems, Moody [4] suggested using the Lomb-Scargle Periodogram (LSP) to estimate power spectra in the frequency analysis of HRV. This was found to be more reliable than autoregressive and DFT-based techniques. Clifford [5] showed that the LSP produces more robust spectral estimates even with a large of number of samples removed due to ectopy. Though shown to be superior, few have studied online algorithms for the LSP even though such an approach has clear advantages in real-time implementations. Most HRV literature has applied the LS method by using the fast Lomb-Scargle algorithm of Press and Rybicki [6]. This algorithm computes an approximation with adjustable accuracy with $O(N \log N)$ time complexity for an input size of N samples. This algorithm significantly reduces computation compared to the $O(N^2)$ direct calculation, but the benefits are mostly relevant to batch computation from a large number of inputs to an output of comparable size (in terms of frequency limits and resolution).

In a typical online, real-time HRV application, input samples are acquired incrementally and irregularly as the QRS waveform is detected and located from the electrocardiogram (ECG). If an updated power estimate has to be computed from beat to beat, the fast Lomb-Scargle algorithm would require $O(N \log N)$ operations per update.

In this paper, a novel online algorithm for the LSP is described. It combines similar reductions as the fast Lomb-Scargle algorithm with a recursive update equation derived from trigonometric identities. The main advantage of this algorithm is its capability to compute the LSP power spectra incrementally where the inclusion of each new sample or the exclusion of an expired sample requires only $O(N)$ computations per update. In addition, the algorithm is an exact computation derived from definitions as opposed to the fast Lomb-Scargle algorithm which is an approximation.

II. Background

In this section, we introduce notation and review the commonly on computing different estimates of power spectra for nonuniform sampling.

The signal to be analysed is a sequence of samples in the time or spatial domain, denoted as $\{y_n\}_{n=1}^N$ or equivalently $\{y(t_n)\}_{n=1}^N$ where $\{t_n\}_{n=1}^N$ are the sampling times. In general, $\{t_n\}_{n=1}^N$ have non-uniform intervals, and it is assumed that t_n is strictly increasing with n. We denote the zero-mean version of this time series as $\hat{y}_n = y_n - \bar{y}$ where $\bar{y} = \frac{1}{N}\sum_{n=1}^N y_n$ is the sample mean.

A. The Classical Fourier Periodogram

A simple form of spectral analysis is the Classical Fourier Periodogram (CFP) [7] which is defined as

$$P_{CF}(\omega) = \frac{1}{N}\left|\sum_{n=1}^{N} \hat{y}_n e^{-j\omega t_n}\right|^2. \quad (1)$$

where ω is the frequency at which the periodogram is evaluated.

Zhe Zhang zhe.zhang@sydney.edu.au and Philip Leong philip.leong@sydney.edu.au are with the Department of Electrical and Information Engineering, The University of Sydney, NSW 2006, Australia.

This research was partially supported by the Australian Research Council's Linkage Projects funding scheme (project numbers LP110200413 and LP130101034)

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

The CFP is theoretically inappropriate for real-valued nonuniformly sampled data [8], [7]. However, we will show later in this paper that it still provides a useful definition.

B. Least-Squares Periodogram

With the following vector notations (dependencies on ω omitted):

$$\mathbf{c} = \begin{bmatrix} \cos\omega t_1 \\ ... \\ \cos\omega t_N \end{bmatrix}, \mathbf{s} = \begin{bmatrix} \sin\omega t_1 \\ ... \\ \sin\omega t_N \end{bmatrix}, \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ ... \\ \hat{y}_N \end{bmatrix}$$

The least-square periodogram [9] [8] is defined as:

$$P_{LS} = \frac{(\hat{\mathbf{y}}\cdot\mathbf{c})^2(\mathbf{s}\cdot\mathbf{s}) - 2(\hat{\mathbf{y}}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{s})(\hat{\mathbf{y}}\cdot\mathbf{s}) + (\hat{\mathbf{y}}\cdot\mathbf{s})^2(\mathbf{c}\cdot\mathbf{c})}{N((\mathbf{c}\cdot\mathbf{c})(\mathbf{s}\cdot\mathbf{s}) - (\mathbf{c}\cdot\mathbf{s})^2)} \tag{2}$$

where $\hat{\mathbf{y}}\cdot\mathbf{c} = \mathbf{c}\cdot\hat{\mathbf{y}}$ denotes the dot product of $\hat{\mathbf{y}}$ and $\mathbf{c}$. This definition is commonly seen with additional normalizations but for the HRV application, this is not needed because the final measurement is a ratio of power in two frequency bands.

C. The Lomb-Scargle Periodogram

The Lomb-Scargle Periodogram [9], [7] is a modified version of (2) and the most widely known and used variant. Additionally define:

$$\mathbf{c}_\tau = \begin{bmatrix} \cos\omega(t_1-\tau) \\ ... \\ \cos\omega(t_N-\tau) \end{bmatrix}, \qquad \mathbf{s}_\tau = \begin{bmatrix} \sin\omega(t_1-\tau) \\ ... \\ \sin\omega(t_N-\tau) \end{bmatrix}$$

where τ is an ω dependent time reference explicitly chosen to satisfy $\mathbf{c}_\tau\cdot\mathbf{s}_\tau = 0$. The Lomb-Scargle Periodogram is computed as:

$$P_{Lomb} = \frac{1}{N}\left(\frac{(\hat{\mathbf{y}}\cdot\mathbf{c}_\tau)^2}{\mathbf{c}_\tau\cdot\mathbf{c}_\tau} + \frac{(\hat{\mathbf{y}}\cdot\mathbf{s}_\tau)^2}{\mathbf{s}_\tau\cdot\mathbf{s}_\tau}\right). \tag{3}$$

Eqn. (3) (with normalization) is commonly cited in the HRV literature [4], [5], [10]. However, (2) and (3) are numerically identical [9], [8]. Lomb [9] explained that the time-shift modification of Eqn. (3) was solely intended to assist statistical work in analytical forms and for numerical evaluations (2) should be used over (3).

III. An Online Algorithm for the Periodograms

A. The Recursive Update of Trigonometric Sums

First we focus on recursive evaluation of trigonometric series in the form of

$$Y_G(N,\omega) = \sum_{n=1}^{N} y_n e^{-j\omega t_n}. \tag{4}$$

1) Simple Growing Window: If new samples are iteratively added to the summation, a recursive update for (4) is simply:

$$Y_G(N,\omega) = Y_G(N-1,\omega) + y_N\, e^{-j\omega t_N}.$$

2) Growing Window with Moving Time Reference: With (4) the absolute time reference $t = 0$ was used. If we change the time reference to be the latest sample acquired, i.e., t_N, (4) becomes

$$Y_{MT}(N,\omega) = e^{j\omega t_N}\sum_{n=1}^{N} y_n e^{-j\omega t_n}. \tag{5}$$

With a similar expansion of the summation, the recursive relation is now

$$Y_{MT}(N,\omega) = Y_{MT}(N-1,\omega)\, e^{j\omega(t_N - t_{N-1})} + y_N. \tag{6}$$

Remarkably, the only reference to time is the time interval between the two most recent samples $\Delta t = t_N - t_{N-1}$.

3) Sliding Window: For a sliding window with fixed width, (6) is only half a solution. This ingress update is performed whenever a sample becomes available and the window slides forward in time to include the sample. On the other end, older samples will eventually reach the trailing edge of the window and this requires egress updates.

We consider a sliding rectangular window which only serves the purpose of localizing a small section of the full or infinite time series in our formalization. Spectral smoothing window functions, e.g. Hamming and Hann are rarely used with irregular sampling due their interaction with the spectral window [7], [5].

Let the rectangular window function with width T be zero outside $-T < t \le 0$. An egress update is required for each sample, T unit-time after its ingress event, i.e., at $t_n + T$. Buffering of the windowed samples and scheduling of the egress updates must be handled by the implementation.

Having established the leading window edge as the moving time reference, a ingress sample denoted as y_{in}, always has relative time $t_{in} = 0$ and similarly an egress sample y_{eg} will have $t_{eg} = -T$. In addition, let $Y(\omega)$ be similarly defined as $Y_{MT}(\omega)$ from (5) but with windowed samples only. Let $Y'(\omega)$ be the previous value of $Y(\omega)$ and Δt be the time lapsed from the previous update, whether it was ingress or egress. With the new notation, (6) is re-written as the ingress sample update

$$Y(\omega) = Y'(\omega)e^{j\omega\Delta t} + y_{in}. \tag{7}$$

An egress sample update can be similarly derived as

$$Y(\omega) = Y'(\omega)e^{j\omega\Delta t} - y_{eg}\, e^{j\omega T}. \tag{8}$$

The rotation coefficient $e^{j\omega T}$ reduces to one for $\omega = k\frac{2\pi}{T}$, i.e., harmonics of the fundamental frequency.

When ingress and egress events occur simultaneously, both updates are required but with the later computed one having $\Delta t = 0$.

B. Online Computation of the Periodograms

We now explain how to compute the CFP and the LSP online using the recursive updates (7) and (8). To avoid changing from the notation of Section II, it is assumed that the index range $n = 1, \ldots, N$ corresponds to the windowed samples. In addition, we implicitly apply a moving time

reference t_N and use $\{t_n - t_N\}_{n=1}^{N}$ consistent with the sliding window. This change of time reference has no effect on the final values of the periodograms (1), (2) and (3) because of their time-translation invariance [7].

1) Classical Fourier Periodogram: The CFP (1) can be expressed as:

$$P_{CF}(\omega) = \frac{1}{N}\left|Y(\omega) - \bar{y}S(\omega)\right|^2.$$

where $S(\omega) = \sum_{n=1}^{N} e^{-j\omega t_n}$ is known as the spectral window.

$Y(\omega)$ and $S(\omega)$ can be recursively computed as two individual instances of (5). N and $\bar{y}$ are updated correspondingly for each ingress and egress event. Since recursive updates of all four quantities N, $\bar{y}$, $Y(\omega)$, $S(\omega)$ require constant numbers of operations, the time complexity for an update remains $O(1)$ per frequency ω.

2) Least-Squares Periodogram: To compute (2), use the following trigonometric formulas from [6]:

$$\begin{aligned}
&\mathbf{c}\cdot\mathbf{c} = (N + \Re(S(2\omega)))/2, \quad &&\hat{\mathbf{y}}\cdot\mathbf{c} = \Re(Y(\omega)) - \Re(S(\omega))\bar{y} \\
&\mathbf{s}\cdot\mathbf{s} = (N - \Re(S(2\omega)))/2, \quad &&\hat{\mathbf{y}}\cdot\mathbf{s} = \Im(S(\omega))\bar{y} - \Im(Y(\omega)) \\
&\mathbf{c}\cdot\mathbf{s} = -\Im(S(2\omega))/2. &&
\end{aligned} \tag{9}$$

Similar to the CFP, in addition to running updates of N and $\bar{y}$, trigonometric sums $Y(\omega)$, $S(\omega)$ and $S(2\omega)$ can be recursively computed as instances of (5). With these quantities updated, $\hat{\mathbf{y}}\cdot\mathbf{c}$, $\hat{\mathbf{y}}\cdot\mathbf{s}$, $\mathbf{c}\cdot\mathbf{c}$, $\mathbf{s}\cdot\mathbf{s}$, $\mathbf{c}\cdot\mathbf{s}$ can be evaluated using (9) and subsequently used to compute (2). As before, all computations here require constant numbers of operations per update for a given ω.

3) Lomb-Scargle Periodogram: Evaluation based on (3) for the Lomb-Scargle Periodogram is not recommended because (3) and (2) are numerically identical while (3) requires extra computations for τ and subsequent time-shifting of the trigonometric sums [6]. The arctangent function is also expensive to compute compared to multiplication and addition. Thus in agreement with Lomb [9] and Stoica et al. [8], (2) should be used over (3) for all computation work.

IV. Discussion

A. Online LSP versus the Fast Lomb-Scargle in HRV

The HRV standard [1] defines the minimum period of time over which short-term HRV can be meaningfully measured and interpreted to be five minutes. This constrains the width of the sliding window to be $T = 300\ s$ with a fundamental frequency of $f_1 = 1/300\ Hz$. The primary HRV metrics associated with frequency domain methods are LF, HF and LF/HF ratio, where LF and HF are defined as the total spectral power in the frequency band of 0.04 Hz to 0.15 Hz and from 0.15 Hz to 0.4 Hz respectively. Therefore, computing the LSP for all harmonics of the fundamental frequency within the range of 0.04 Hz (12th harmonic) to 0.4 Hz (120th harmonic), a total of 109 frequencies, is sufficient to evaluate these metrics. Since this frequency resolution is a function of T which is approximately proportional to the number of samples (heart beats) N in the windowed input, the computational complexity is also expressed in terms of N.

In time-frequency analysis terms, increasing time-resolution without sacrificing frequency-resolution requires window instances with overlaps. Therefore whether the online LSP or the fast Lomb-Scargle is more computationally efficient is determined by the amount of overlap between the windows for successive outputs. In theory, an overlapping percentage of $1 - \frac{\log_2 N}{N}$ or more would yield in favour of the online LSP. For a typical real-world value of $N = 400$ or an average of 80 BPM over 5 minutes, this amount of overlap yields a rate of output update every 6.4 seconds.

In practice, however, the online LSP may still be preferable over the fast Lomb-Scargle at an overlapping percentage that is much less than the above mentioned. This is because the fast Lomb-Scargle algorithm involves a pre-processing step which has complexity $O(N)$ but the constant in front of this order is so large that it is typically much slower to compute than the $O(N \log N)$ FFT part of the algorithm [11].

From an implementation perspective, the fast Lomb-Scargle algorithm [6] is substantially more complicated in structure than the proposed online algorithm and may lead to high area costs in custom hardware implementations.

B. Holland and Aboy's RLSFT

Starting from the inverse DFT, Holland and Aboy [10] derived a recursive solution for the case of adding a new sample using the well-known recursive least-squares (RLS) formulation to produce the "Recursive Least Squares Fourier Transform" (RLSFT). The RLSFT has some similarities to the CFP but is less general as some restrictive DFT conventions were enforced e.g. standard frequency bins and quantization of sampling times. Our efforts to reproduce their work resulted in rather erratic spectra where significant side-lobe leakage is observed (see Fig. 1). In addition, inaccurate heights of spectral peaks compared to an FFT reference were also apparent.

C. CFP versus LSP for HRV measureaments

We included the CFP in our discussion for two reasons. Firstly, Holland and Aboy's work [10] is very similar to the CFP in formalization. More importantly, Scargle suggested [7] that "the actual values are typically not changed much" from the CFP to the LSP.

To compare CFP and LSP accuracy in a HRV application, we duplicated the test bench set up used by Clifford and Tarassenko [5]. A synthetic RR signal $RR(t)$ with an LF and an HF sinusoid components with a theoretical LF/HF power ratio of 0.64 is generated at a high sampling rate (1000 Hz) for a total period of $T = 300\ s$ or five minutes. The time series is then searched and those samples occurring at time intervals consistent with their values (RR intervals) are collected, forming the synthetic RR tachogram $\{RR_n\}_{n=1}^{N}$. Because the mean heart rate was chosen as 60 BPM or one beat per second, the length of the generated RR tachogram is $N = 300$ beats. The RR tachogram is then processed with randomly occuring ectopic beats, where the occurrence of

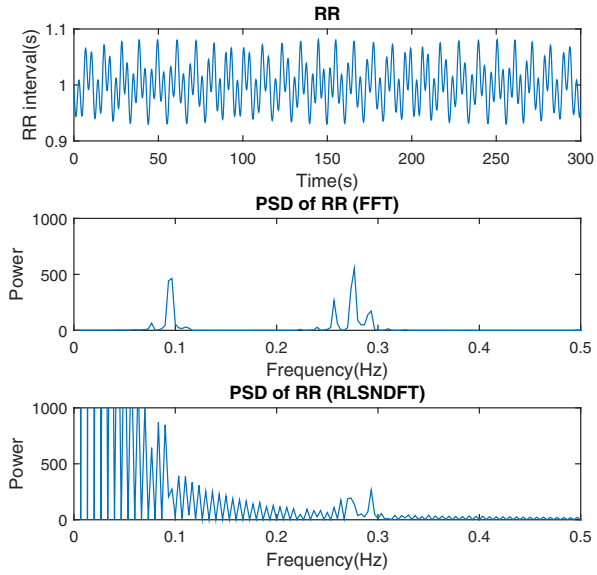

Fig. 1. The RLSFT power spectrum of a synthetic RR signal compared to an FFT reference. The RLSFT seems to show severe local leakage from the DC bin for this particular signal.

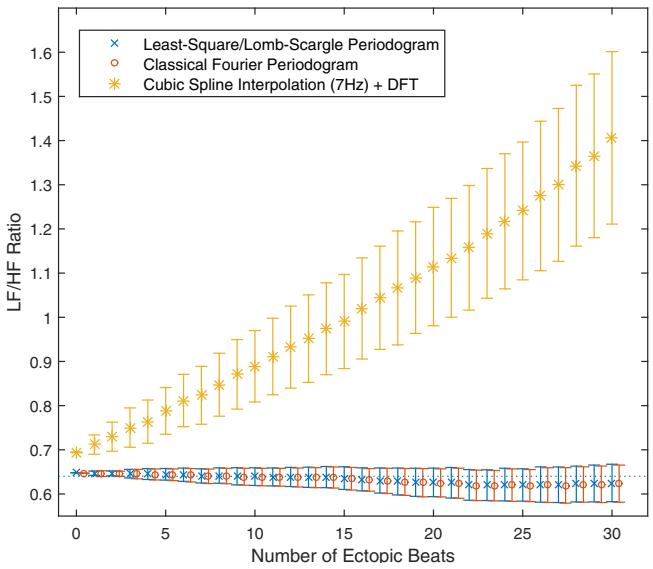

Fig. 2. Distribution of LF/HF ratio estimates for the LSP and the CFP. The dotted line indicates the theoretical standard of LF/HF = 0.64 and errorbars indicating one standard deviation. The CFP's series are ploted with a slight offset for graphical clarity. The result of a baseline method (interpolation followed by DFT) is also included.

one ectopic beat leads to the removal of a pair of affected RR interval samples being removed from the series. For each number of ectopic beats simulated, starting from 1 upto a maximum of 30 beats (or 60 samples being removed out of 300), 1000 randomly seeded trials were tested. For each trial the resulting RR series with beats removed are spectrally analysed using the CFP and the LSP and the LF/HF ratio computed.

Fig. 2 shows the mean and standard deviation of the LF/HF ratio for 1000 simulations and how they scale with the increasing number of samples removed due to ectopy. For the LSP, our test results have accurately reproduced those shown in Fig. 5. of [5], on which we modelled this experiment. In addition, the CFP gave results which were nearly indistinguishable from the LSP, consistent with Scargle's suggestion regarding the numerical differences [7]. Recalling from section III-B, the CFP is cheaper to compute than the LSP, this leads us to the conclusion that the CFP may be a viable approximation to the LSP when applied to HRV analysis.

V. Conclusions

Online algorithms for the Classical Fourier Periodogram (CFP) and Lomb-Scargle Periodogram (LSP) were presented. For each new sample, this technique computes an update based on previous values, avoiding redundant computations and interpolation. The technique is eminently suitable in embedded implementations of high resolution time-frequency analysis of Heart Rate Variability. We further evaluated the performance of the CFP compared with the LSP in HRV analysis, our results showing that the two are similar in accuracy and the CFP may offer computational advantages due to its simpler form.

References

[1] T. F. of the European Society of Cardiology, T. F. of the European Society of Cardiology *et al.*, "the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[2] R. E. Kleiger, J. P. Miller, J. T. Bigger, and A. J. Moss, "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction," *The American journal of cardiology*, vol. 59, no. 4, pp. 256–262, 1987.

[3] M. T. La Rovere, J. T. Bigger, F. I. Marcus, A. Mortara, P. J. Schwartz, A. A. Tone, R. A. M. I. Investigators *et al.*, "Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction," *The Lancet*, vol. 351, no. 9101, pp. 478–484, 1998.

[4] G. B. Moody, "Spectral analysis of heart rate without resampling," in *Computers in Cardiology 1993, Proceedings.* IEEE, 1993, pp. 715–718.

[5] G. Clifford and L. Tarassenko, "Quantifying errors in spectral estimates of HRV due to beat replacement and resampling," *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 4, pp. 630–638, April 2005.

[6] W. H. Press and G. B. Rybicki, "Fast algorithm for spectral analysis of unevenly sampled data," *The Astrophysical Journal*, vol. 338, pp. 277–280, 1989.

[7] J. D. Scargle, "Studies in astronomical time series analysis. II-statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, pp. 835–853, 1982.

[8] P. Stoica, J. Li, and H. He, "Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram," *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 843–858, March 2009.

[9] N. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and Space Science*, vol. 39, no. 2, pp. 447–462, 1976. [Online]. Available: http://dx.doi.org/10.1007/BF00648343

[10] A. Holland and M. Aboy, "A novel recursive Fourier transform for nonuniform sampled signals: application to heart rate variability spectrum estimation," *Medical & biological engineering & computing*, vol. 47, no. 7, pp. 697–707, 2009.

[11] S. Kestur, S. Park, K. Irick, and V. Narayanan, "Accelerating the nonuniform fast Fourier transform using FPGAs," in *Field-Programmable Custom Computing Machines (FCCM), 2010 18th IEEE Annual International Symposium on*, May 2010, pp. 19–26.