Averaged Lomb Periodograms for Nonuniform Sampling

Thong T [1], McNames J[2], Aboy M[2], Oken B[3]
[1]Dept Biomedical Engineering, OGI/Oregon Health &Science University, [2] Biomedical Signal Processing Laboratory, Dept of Electrical & Computer Engineering, Portland State University, [3] Dept Neurology, Oregon Health & Science University
trant@bme.ogi.edu

1 Introduction

Spectrum analysis techniques have been used extensively to characterize heart rate data [1], [2]. Since the signal is nonuniformly sampled, most analysts use interpolation to resample the signal uniformly and then apply traditional spectrum analysis. Laguna *et al.* [2] reviewed the effects of the interpolation method and recommended the Lomb-Scargle transform [3, 4, 5] for spectral estimation. No resampling is necessary with the Lomb-Scargle transform.

In spite of the fact that a fast Lomb transform [5] has been developed, the evaluation of the Lomb spectrum still consumes considerably more computational cycles than the combined operations of uniform resampling and nonparametric spectral estimation based on the fast Fourier transform (FFT). Like the periodogram, the Lomb periodogram is not a statistically consistent estimator. To obtain a spectral estimate that strikes a reasonable tradeoff between bias and variance, one may want to consider power averaging transforms of several, possibly overlapped, windowed segments of the observed signal. This is similar to the approaches used in averaged Bartlett (adjacent segments) and Welch (overlapped segments) periodograms.

In order to be able to average Lomb periodograms, two problems need to be resolved

1. The Lomb transform, as described in the literature, is variance normalized. Since each record is individually normalized, averaging across records is not meaningful. A de-normalized transform is needed.
2. Since the data are unevenly sampled, how should the long data record be split for averaging purpose?

2 Methods

De-normalization.

The Lomb periodogram of a non-uniformly sampled real-valued data sequence $\{x(t_n)\}$ of length N is defined by [5]

$$P_X(f) = \frac{1}{2\sigma^2}\left\{ \frac{\left[\sum_{n=1}^{N}(x(t_n)-\bar{x})\cos(2\pi f(t_n-\tau))\right]^2}{\sum_{n=1}^{N}\cos^2(2\pi f(t_n-\tau))} + \frac{\left[\sum_{n=1}^{N}(x(t_n)-\bar{x})\sin(2\pi f(t_n-\tau))\right]^2}{\sum_{n=1}^{N}\sin^2(2\pi f(t_n-\tau))} \right\} \tag{1}$$

where $\bar{x}$ and σ^2 are the mean and variance of the series $\{x(t_n)\}$, $\tau(f)$ is a frequency dependent time delay, defined to make the periodogram insensitive to time shift [2], [5] as follows

$$\tan(4\pi f\tau) = \sum_{n=1}^{N}\sin(4\pi f t_n) / \sum_{n=1}^{N}\cos(4\pi f t_n) \tag{2}$$

To derive the de-normalization factor for line spectra, one considers the case of equally sampled data, i.e. $t_n = nT$, and $x(t_n) = A\cos(2\pi nK/N)$, where K is an integer. With this input, we have $\bar{x} = 0$, $\sigma^2 = A^2/2$, $\tau(f) = 0$. The numerator of the first term of Eqn. (1) at f=K/NT is $(AN/2)^2$, the denominator is (N/2). The second term in the bracket is 0 due the orthogonality of the sine and cosine functions. Thus $P_X(K/NT)$ is N/2. Since we expect the peak power to be σ^2 ($=A^2/2$), the line spectra de-normalization factor is $2\sigma^2/N$. Note that in his case $P_X(f \neq K/NT)=0$.

It can be shown that if the input is a white noise of variance σ^2, the expected values of the terms in the {}brackets in (1) is equal to $2\sigma^2$ (for arbitrary t_n), assuming $\tau(f)=0$. Thus $P_x(f) = 1$. The noise power measured in an effective noise bandwidth of $2\pi/NT$ (with T= average (Δt_n), variance (Δt_n) small, $NT=\Sigma\Delta t_n$) is expected to be $2\sigma^2/N$ (power of σ^2 spread over N/2 bins of width 1/NT (Hz) in the bandwidth [0, 1/2T]). Thus, the de-normalization factor for the power spectral density is $2\sigma^2/N$. This is the same as in the case of a sine wave input. Thus, (1) becomes

$$P_X(f) = \frac{1}{N}\left\{ \frac{\left[\sum_{n=1}^{N}(x(t_n)-\bar{x})\cos(2\pi f(t_n-\tau))\right]^2}{\sum_{n=1}^{N}\cos^2(2\pi f(t_n-\tau))} + \frac{\left[\sum_{n=1}^{N}(x(t_n)-\bar{x})\sin(2\pi f(t_n-\tau))\right]^2}{\sum_{n=1}^{N}\sin^2(2\pi f(t_n-\tau))} \right\} \tag{3}$$

Averaging

To split a long record for averaging, we have two choices:

1. Use records with equal number of sample points, or
2. Use records with the same duration.

Due to the non-uniform sampling, in the first case we end up with records of different time duration. Since an implicit rectangular window is used, each output frequency in (3) is effectively convolved with a sinf/f function, the bandwidth of which is proportional to the observation window. Thus, each transform now has a different sinf/f window. When these terms are averaged, they yield inconsistent results due to the different bandwidths.

By using records with the same time duration, even though each record has a different number of points, the underlying bandwidths are the same. As discussed in the de-normalization argument leading to (3), if the signal is stationary, consistent powers are obtained. With both consistent power and bandwidth, averaging of the spectra can now be performed.

Thus, the recommended segmentation for averaging purpose is to choose the segments to be of the same time duration D, i.e. such that $\max_m\left\{\{x(t_i)\}_{i=n}^{n+m} : (t_{n+m}-t_n) < D\right\}$.

3 Results

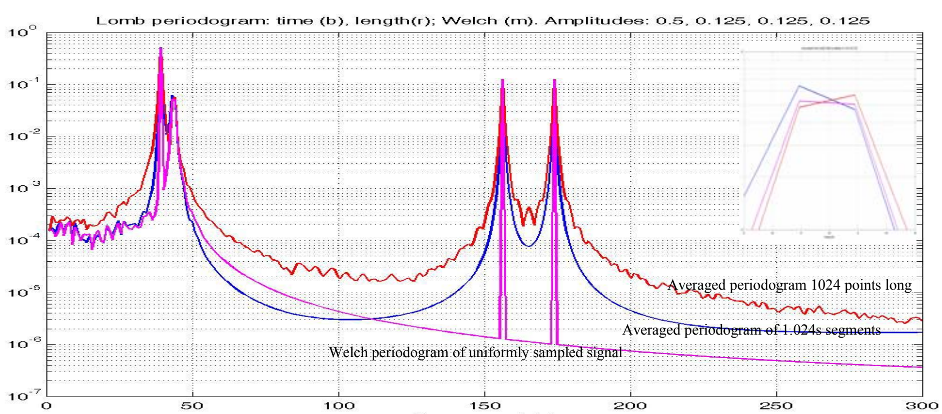

Fig. 1. Lomb (power) periodograms. Input: cosines at f_1= 40/1.024, f_2= 41.5/1.024 (middle of frequency bin), $4f_1$, $4f_2$ with amplitudes 1, and 3x 0.5; lowpass noise (σ^2=1 white noise low-pass filtered to f_c=50Hz). Inset at top right: zoom around f_2. Sampling times: $n+0.5\sin(0.001\pi n)$ ms, n=0… 11000. Frequency resolution of all periodograms: 1/1.024 Hz. Averaging overlap: 50% of record.

4 Discussion

In Fig. 1, the peak powers at f_1, $4f_1$, $4f_2$ are correct, namely ½, $^1/_8$, $^1/_8$. The peak power at f_2, being at the mid-frequency between two bin centers, should come out to be $^1/_{2\pi^2} \sim 0.051$ measured at the 2 adjacent bins. In the inset at the top right of Fig. 1, the two peaks of the uniformly sampled signal are almost equal 0.051. The average of the 2 peaks of the averaged periodogram of 1.024 s segments, appear to be correct. The error is caused by the de-normalization being (incorrectly) set for a signal right at the center of the bin. With the implicit rectangular window used, the power estimate can be off by 20%. Reducing the amount of timing jitter, does reduce the error. A similar effect is also observed with the averaged of 1024 point periodogram.

With regard to noise power, looking at the low frequency components below 50 Hz, the 1.024s averaged periodograms exhibit comparable powers to those for the uniformly sampled periodograms. The averaged 1024 point periodograms, as expected, exhibit slightly higher powers due to inconsistent resolution bandwidths.

Beyond 200 Hz, as the noise power drops below about 2×10^{-6}(~ -57 dB below the peak), the power in the 1.024 s periodograms no longer decreases. Changing the amplitude of the sampling time jitter reduces this level, confirming that the timing jitter is the main contributor of this equivalent noise effect, limiting the noise floor. For biological signals, with a -60 dB noise floor in the case of a very large timing jitter of ±50%, this noise floor is not a limitation for the use of the Lomb periodogram.

5 Conclusions

With the proposed de-normalization factor of $2\sigma^2/N$ and using segments of equal time duration, averaged overlapped Lomb periodograms can be computed efficiently using short transforms.

References

1. Task Force of the Europ Soc of Cardiol & NASPE. Heart rate variability – standards of measurement, physiological interpretation, and clinical use. *Circulation* 1996: 1043-1065.
2. Laguna P, Moody GB, Mark RG. Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals. *IEEE Trans Biomed Eng* 1998:698-715.
3. Lomb NR. Least-squares frequency analysis of unequally spaced data. *Astrophysical and Space Science,* 1976;39: 447-462.
4. Scargle JD. Studies in astronomical time series analysis II. Statistical aspects of spectral analysis of unevenly sampled data.*Astrophysical J,* 1982;263: 835-853.
5. Press WH, Teukolsky SA, Vertterling WT, Flannery BP. *Numerical Recipes in C++*. Cambridge, UK: Cambridge University Press. 2002: 580:589.