

# A Statistical Model for Burst Detection

Lee F. Richardson

November 19, 2018

## 1 Introduction

This document proposes a statistical model to detect “Bursts” and gives a maximum likelihood estimation procedure for this model. By bursts, I mean local (in time) time-series oscillations at a particular frequency. The motivation for developing burst detection methodology comes from Lundqvist et al. (2016), who finds that bursts, *not* sustained oscillations, underly working memory.

## 2 A Statistical Model for Bursts

A graphical example of a theoretical burst is shown in Figure 1. Figure 1 suggests a simple deterministic model for a burst, namely, a cosine function multiplied by a window:

$$g_t = A \cos\left(\frac{2\pi Ft}{N} + \phi\right) \cdot \mathbb{1}_{S, S+L-1}(t) \quad (1)$$

Equation 1 has 5 parameters:

- $A$ : Amplitude.  $A \in [0, \infty]$
- $F$ : Frequency. Number of complete cycles in a length  $N$  signal.  $F \in [0, \frac{1}{2}]$
- $\phi$ : Phase.  $\phi \in [0, 2\pi]$
- $S$ : Start of oscillation.  $S \in [0, 1, \dots, N-2]$
- $L$ : Length of oscillation.  $L \in [1, 2, \dots, N-2]$

We turn Equation 1 into a statistical model by introducing an error term;

$$x_t = g_t + \epsilon_t \quad (2)$$

Where  $\epsilon_t \sim N(0, \sigma^2)$  and  $\epsilon_t$  is i.i.d.

## 3 Estimation

Now that we’ve proposed a Bursts model (Equation 2), given an arbitrary time-series, we want to estimate the parameters. We use maximum likelihood for parameter estimation. Section 3.1 derives analytic solutions for 3 of the 6 parameters, and Section ?? shows how to solve for the remaining 3 parameters numerically, using the Spectrogram.

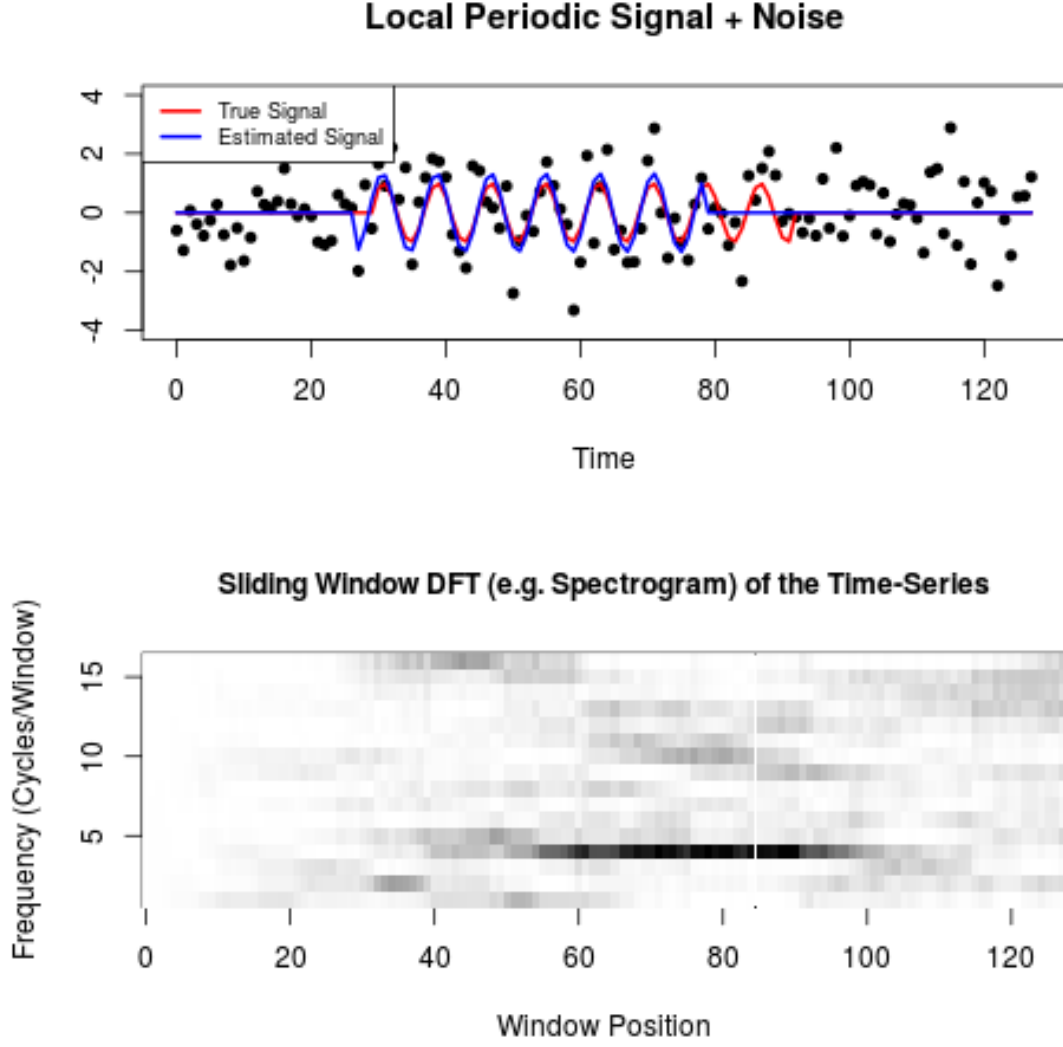


Figure 1: Top: Time series of a local signal (e.g. burst) embedded in noise. The true underlying signal is the red line, and the signal estimated by our procedure is colored in blue. Bottom Sliding Window Discrete Fourier Transform (SWDFT) of the time-series, which indicates where and what frequency the burst occurs. The SWDFT may be considered an estimator of the Spectrogram.

### 3.1 Analytic Solutions for $A, \phi$ , and $\sigma$

Following Equation 2, we know that  $x_t \sim N(g_t, \sigma^2)$ . Denote  $\theta = [A, F, \phi, S, L, \sigma]$  as our vector of unknown parameters, and say that we observe a length  $N$  time-series  $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]$ . The joint pdf of  $\mathbf{x}$  is:

$$\mathbb{P}(x_0, \dots, x_{N-1} | \theta) = \prod_{t=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_t - g_t)^2}{2\sigma^2}\right) \quad (3)$$

Therefore, the likelihood function is:

$$L(\theta) = \prod_{t=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_t - g_t)^2}{2\sigma^2}\right) \quad (4)$$

And the log-likelihood is:

$$\ell(\theta) \propto -N \log(\sigma) - \frac{1}{2\sigma^2} \sum_{t=0}^{N-1} (x_t - g_t)^2 \quad (5)$$

The maximum likelihood estimate is:

$$\hat{\theta}_{ML} = \arg \max_{\theta} \ell(\theta) \quad (6)$$

Since  $S$  and  $L$  are integers, we will solve for them numerically. Similarly, it is hard to get analytic solutions for  $F$ , so we solve for this numerically as well. At the end of this section, we show how we can use the Spectrogram to get starting values numerical solutions. For now, assume that  $S$ ,  $L$ , and  $F$  are known parameters. Then we have

$$\ell(A, \phi, \sigma | S, L, F) \propto -N \log(\sigma) - \frac{1}{2\sigma^2} \sum_{t=0}^{N-1} (x_t - A \cos(\frac{2\pi Ft}{N} + \phi) \cdot \mathbb{1}_{S, S+L-1}(t))^2 \quad (7)$$

Next, notice that for any value of  $\sigma$ , the log-likelihood is *maximized* when the quadratic term inside the sum is minimized (since there is a negative sign in front of the sum). With this, we closely follow the second example of Section 7.10 of Kay (1993), and let  $J(A, \phi)$  be the term inside the sum we want to minimize:

$$J(A, \phi) = \sum_{t=0}^{N-1} (x_t - A \cos(\frac{2\pi Ft}{N} + \phi) \cdot \mathbb{1}_{S, S+L-1}(t))^2 \quad (8)$$

Which looks like least squares. The only problem is that the cosine factor is a non-linear function, but we can linearize it with the following trigonometric identity:

$$A \cos(\frac{2\pi Ft}{N} + \phi) = A \cos(\phi) \cos(\frac{2\pi Ft}{N}) - A \sin(\phi) \sin(\frac{2\pi Ft}{N}) \quad (9)$$

Then if we define the following coefficients:

$$\begin{aligned}\beta_1 &= A \cos(\phi) \\ \beta_2 &= -A \sin(\phi)\end{aligned}\tag{10}$$

There is a one-to-one correspondence between  $(\beta_1, \beta_2)$  and  $(A, \phi)$ :

$$\begin{aligned}A &= \sqrt{\beta_1^2 + \beta_2^2} \\ \phi &= \arctan\left(\frac{-\beta_2}{\beta_1}\right)\end{aligned}\tag{11}$$

We can re-write  $J(A, \phi)$  instead as a function of the  $\beta$ 's:

$$J(\beta_1, \beta_2) = \sum_{t=0}^{N-1} (x_t - \beta_1 \cos(\frac{2\pi F t}{N}) \mathbb{1}_{S, S+L-1}(t) - \beta_2 \sin(\frac{2\pi F t}{N}) \mathbb{1}_{S, S+L-1}(t))^2 \tag{12}$$

Equation 12 is just multiple linear regression, which means we can solve for  $\beta$  analytically. We now write  $J(\beta_1, \beta_2)$  in matrix-vector notation. Define the following

$$\begin{aligned}\mathbf{c} &= [\cos(\frac{2\pi F \cdot 0}{N}) \mathbb{1}_{S, S+L-1}(0), \cos(\frac{2\pi F \cdot 1}{N}) \mathbb{1}_{S, S+L-1}(1), \dots, \cos(\frac{2\pi F \cdot (N-1)}{N}) \mathbb{1}_{S, S+L-1}(N-1)]^T \\ \mathbf{s} &= [\sin(\frac{2\pi F \cdot 0}{N}) \mathbb{1}_{S, S+L-1}(0), \sin(\frac{2\pi F \cdot 1}{N}) \mathbb{1}_{S, S+L-1}(1), \dots, \sin(\frac{2\pi F \cdot (N-1)}{N}) \mathbb{1}_{S, S+L-1}(N-1)]^T \\ \mathbf{U} &= [\mathbf{c}, \mathbf{s}] \\ \mathbf{x} &= [x_0, x_1, \dots, x_{N-1}] \\ \beta &= [\beta_1, \beta_2]\end{aligned}\tag{13}$$

$J(\beta_1, \beta_2)$  can now be written as:

$$J(\beta_1, \beta_2) = (\mathbf{x} - \mathbf{U}\beta)^T (\mathbf{x} - \mathbf{U}\beta) \tag{14}$$

And the estimate for  $\hat{\beta}$  is:

$$\hat{\beta} = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{x} \tag{15}$$

Which we transform into  $\hat{A}$  and  $\hat{\phi}$  using Equation 11. For  $\sigma$ , we have:

$$\frac{\partial \ell}{\partial \sigma} = \frac{-N}{\sigma} + \frac{1}{\sigma^3} \sum_{t=S}^{S+L-1} (x_t - g_t)^2 \tag{16}$$

Setting this to 0 and plugging in  $\hat{A}$  and  $\hat{\phi}$  gives:

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{t=0}^{N-1} (x_t - g_t)^2} \quad (17)$$

Similar to standard estimates of  $\sigma$  with normal errors.

So, if we assume  $S$ ,  $L$ , and  $F$  are known, we can estimate  $\hat{A}$ ,  $\hat{\phi}$ , and  $\hat{\sigma}$  analytically. But in practice, we don't know the values of  $S$ ,  $L$ , and  $F$ , so we need to estimate them with numerical methods. This is where we use the Spectrogram.

### 3.2 Analytic Justification for Spectrogram Based Starting Values

One might wonder if it's *also* possible to analytically solve for the parameter  $F$ . This section considers what that would look like. If we extend our minimization to include  $F$ , we have:

$$J(\hat{\beta}_1, \hat{\beta}_2, F) = (\mathbf{x} - \mathbf{U}\hat{\beta})^T (\mathbf{x} - \mathbf{U}\hat{\beta}) \quad (18)$$

Plugging in Equation 15 gives (Kay (1993) Equation 7.65):

$$J(\hat{\beta}_1, \hat{\beta}_2, F) = \mathbf{x}^T (\mathbf{I} - \mathbf{U}(\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T) \mathbf{x} \quad (19)$$

Which is equivalent to finding the value of  $F$  that maximizes:

$$\arg \max_F \quad \mathbf{x}^T \mathbf{U} (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{x} \quad (20)$$

We can rewrite this maximization in terms of  $\mathbf{c}$ ,  $\mathbf{s}$ , and  $\mathbf{x}$ :

$$\begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}^T \begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}$$

We can approximate the two diagonal elements of the middle term with:

$$\begin{aligned} \mathbf{c}^T \mathbf{c} &\approx \frac{L}{2} \\ \mathbf{s}^T \mathbf{s} &\approx \frac{L}{2} \end{aligned} \quad (21)$$

For the off diagonal terms, we have:

$$\begin{aligned} \mathbf{c}^T \mathbf{s} &= \mathbf{s}^T \mathbf{c} \\ &= \sum_{t=S}^{S+L-1} \cos\left(\frac{2\pi Ft}{N}\right) \sin\left(\frac{2\pi Ft}{N}\right) \\ &= 2 \sum_{t=S}^{S+L-1} \sin\left(\frac{4\pi Ft}{N}\right) \end{aligned} \quad (22)$$

When the interval  $(S, S + L - 1)$  forms a complete cycle, we have that  $\mathbf{c}^T \mathbf{s} \approx 0$ . While this isn't always true, we assume it for now to keep the derivations simple. Then we have

$$\approx \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}^T \begin{bmatrix} \frac{L}{2} & 0 \\ 0 & \frac{L}{2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}$$

Simplifying this gives:

$$\frac{2}{L} \left[ \left( \sum_{t=S}^{S+L-1} x_t \cos\left(\frac{2\pi F t}{N}\right) \right)^2 + \left( \sum_{t=S}^{S+L-1} x_t \sin\left(\frac{2\pi F t}{N}\right) \right)^2 \right] = \frac{2}{L} \left| \sum_{t=S}^{S+L-1} x_t \omega_N^{-tF} \right|^2$$

And if we let  $\frac{f}{L} = \frac{F}{N}$ , this simplifies to

$$\frac{2}{L} \left| \sum_{t=S}^{S+L-1} x_t \omega_L^{-t f} \right|^2 \quad (23)$$

Which looks very similar to the formula for the Periodogram. Specifically, recall that the Sliding Window Discrete Fourier Transform (SWDFT) for length  $n$  windows is:

$$a_{k,p} = \sum_{j=0}^{n-1} x_{p-n+1+j} \omega_n^{-jk} \quad (24)$$

Which shows that Equation 23 is the maximum of the SWDFT with length  $L$  windows.

So, by assuming that the local signal interval  $(S, S + L - 1)$  has an integer number of complete cycles, the value of  $F$  that maximizes the likelihood is the maximum of the SWDFT with length  $L$  windows. Nicely, this also tells us that the values of  $S$  and  $L$  that maximize the likelihood are the values of  $S$  and  $L$  that correspond to the maximum of the Spectrogram. We use these insights to formally specify the parameters in the numerical search.

### 3.3 Numerical Solutions for $S$ , $L$ , and $F$

Now that we have justified using the Spectrogram to determine starting values for estimation of  $S$ ,  $L$ , and  $F$ , this section more formally specifies which values we use. To estimate  $F$ , we simply use the maximum of the Spectrogram. To keep things simple, we estimate the Spectrogram using the Sliding Window DFT (Equation 24). Then the maximum of is:

$$(\hat{k}, \hat{p}) = \arg \max_{k,p} |a_{k,p}|^2 \quad (25)$$

And since the maximum of the SWDFT corresponds to the frequency closest to the true frequency (see Section 3.3. of Richardson and Eddy (2018)), we can fix the range of  $F$  to search as:

$$F_{min} = (\hat{k} - \frac{1}{2}) \cdot \frac{N}{n}$$

$$F_{max} = (\hat{k} + \frac{1}{2}) \cdot \frac{N}{n}$$

To understand how we specified the range for  $S$  and  $L$ , consider Figure 2. Figure 2 shows the time-series corresponding to frequency  $\hat{k}$  of the Spectrogram. Now, we know that the local periodic signal occurs when this time-series is large, and in particular, where there are consecutive large values of the periodogram. For this reason, we first determine the range of large values surrounding the maximum (e.g.  $a_{\hat{k}, \hat{p}}$ ). To do this, we first determine a threshold for “large” values, which may be determined through either simulation, asymptotic, or an empirical estimation if possible. Then we extract the values:

$$p_{min} = \text{Largest value to the left of } \hat{p} \text{ above a threshold}$$

$$p_{max} = \text{Largest value to the right of } \hat{p} \text{ above a threshold}$$

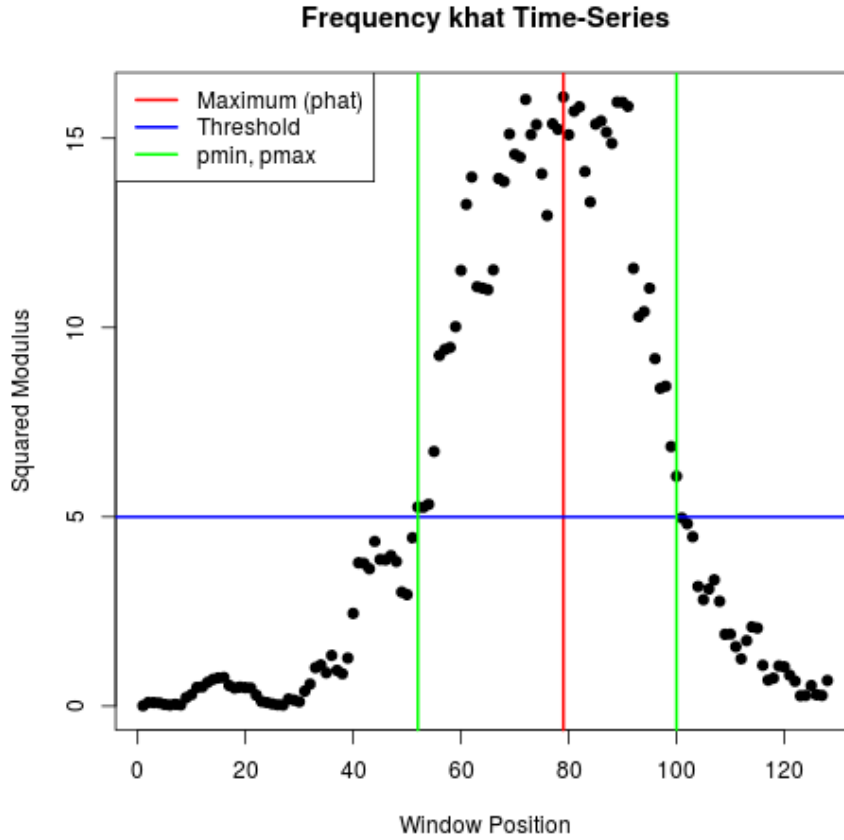


Figure 2: Frequency  $k$  time-series of the SWDFT (e.g. the Spectrogram).

Now we use  $p_{min}$  and  $p_{max}$  to specify the range of  $S$  and  $L$  to numerically search (to be safe, we could search all combinations, but this would be very expensive computationally). For  $S$ , we know that the signal starts when the time-series starts getting large. We also know that the time-series starts getting large when more of the oscillations enter the window, so we can set:

$$\begin{aligned} S_{min} &= p_{min} - n \\ S_{max} &= p_{min} \end{aligned} \tag{26}$$

Finally, we know that the maximum occurs when the maximum of the oscillating part of the local signal is in the window. So roughly speaking, we can set:

$$\begin{aligned} L_{min} &= p_{max} - p_{min} - n \\ L_{max} &= p_{max} - p_{min} \end{aligned}$$

These are most heuristic for now, but can be more properly specified in future work.

## References

- Kay, S. M. (1993). Fundamentals of statistical signal processing, volume i: Estimation theory (v. 1). *PTR Prentice-Hall, Englewood Cliffs*.
- Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., and Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron*, 90(1):152–164.
- Richardson, L. F. and Eddy, W. F. (Submitted. 2018). The sliding window discrete fourier transform. *arXiv preprint arXiv:1807.07797*.