Kernel Based Approaches for Change-Point Detection
Midway report

Anirudhan J. Rajagopalan, ajr619

April 3, 2016

Abstract

Electroencephalography (EEG) signals are detected by using electrodes that are applied to the surface of mamals. A sensitive instrument can show continuous fluctuations of the electric potential between the electrodes. These potentials are due to the superposition of the electrical activity of tens of thousands of neuron cells lying on the surface areas of the brain [3]. A typical human EEG signal is produced by 36 electrodes placed on the scalp. This work aims to exploit the spatial relation of the electrodes to identify a Reproducing Kernel Hilbert Space (RKHS) kernel function that will help us find the change points in the EEG signal data.

1 Change point detection

1.1 Stationary process

A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time. Consequently, parameters that define the process such as mean (μ), standard deviation (σ) remain constant

1.2 Change points

A change point in a time series is the times at which the probability distribution of the time series changes. A change point detection problem involves detecting whether or not a change has occured, and identifying the times at which such changes occured.

Change points can be identified when the parameters defining the stationary process, such as mean (μ) and standard deviation (σ) changes. We discuss ways to identify change points when mean of the process changes.

1.3 Offline Change point detection

Offline change point detection or retrospective change point detection methods has the complete data available at the beginning of the program.

1.3.1 Univariate Time series

Lets assume a time series of observations $x_1, x_2, \ldots, x_n$ of independent random variables with parameters $(\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2), \ldots, (\mu_n, \sigma_n^2)$. Also lets assume that each of the observation x_i is normally distributed with mean μ and common variance $\sigma^2 \forall i \in 1, 2, \ldots, n$. When there is no change in mean, the hypothesis of stability (null hypothesis) is defined as

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu \tag{1}$$

Lets suppose that there is a change in the mean in the observations at an unknown point K. This can be define dy

$$H_1 : \mu_1 = \ldots \mu_k \neq \mu_{k+1} \ldots = \mu_n \tag{2}$$

In our experiments we are going to assume that we know μ_1, μ_n and σ are known beforehand (Refer 2.1.1 of [4]).

We can find the change poing by using Maximum Likelihood estimates.

Under H_0,

$$L_0(\mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2}} \tag{3}$$

$$\tag{4}$$

and the Maximum likelihood estimator is given by

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{5}$$

Under H_1,

$$L_1(\mu_1, \mu_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-(\frac{\sum_{i=1}^{k} (x_i-\mu_1)^2 + \sum_{i=k+1}^{n} (x_i-\mu_n)^2}{2})} \tag{6}$$

$$\tag{7}$$

and the Maximum likelihood estimator is given by

$$\hat{\mu_1} = \bar{x_k} = \frac{1}{k} \sum_{i=1}^{k} x_i \tag{8}$$

$$\hat{\mu_n} = \bar{x_{n-k}} = \frac{1}{n-k} \sum_{i=k+1}^{n} x_i \tag{9}$$

We can use the Maximum likelihood estimator directly to find the change points in the given data. But calculating the MLE is computationaly intractable. An alternate set of equations is given in Chapter 2 of [4]. We use the alternate set of equations to find the change points. The experiments using sample data and its results are shown below.

1.3.2 Experiments — Univariate time series

The offline changepoint detection problem, gives a pretty accurate value for changepoint at k = 300. The different plots are as displayed below.

1.3.3 Multivariate Time series

Multivariate model is similar to the above Univariate model except that each and every observation is m-dimensional. Let $x_1, x_2, \ldots, x_n$ be a sequence of independent m-dimensional normal random vectors with parameters $(\mu_1, \Sigma_1), (\mu_2, \Sigma_2), \ldots, (\mu_n, \Sigma_n)$, respectively. Assume $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_n = \Sigma$. The null hypothesis is given by

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu(\text{unknown}) \tag{10}$$

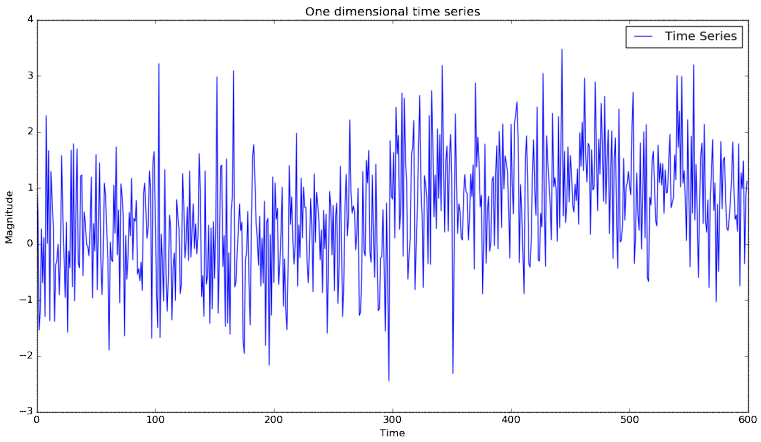

Figure 1: One dimensional time series.

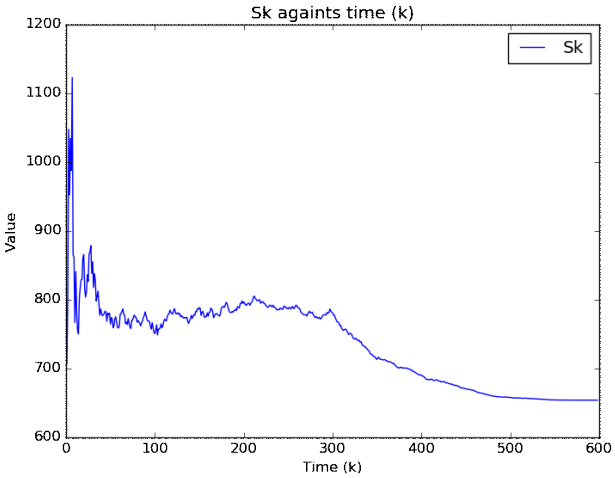

Figure 2: SK values for one dimensional offline detection problem.

If we assume that there is a change at point k in the paramters governing the observation, then the hypothesis (alternate hypothesis) is given by

$$H_1 : \mu_1 = \cdots = \mu_k \neq \mu_{k+1} = \mu_n \tag{11}$$

Where k represents the position of the single change point. (Refer 3.1 of [4]).

This can be solved by following the steps described in section 3.1.1 of [4]. As in the case of univariate model, finding the likelihood directly is computationally

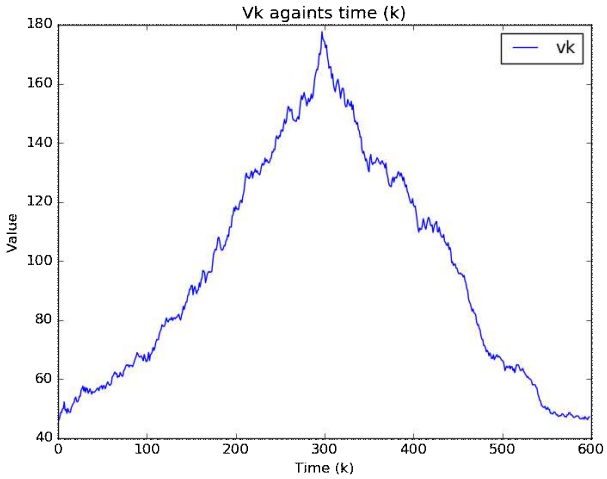

Figure 3: VK values for one dimensional offline detection problem.

intractable as 2π and e has negative powers that can go pretty large and hence the likelihood will always become 1.

1.3.4 Experiments — Multivariate change point detection

We did Several experiments by varying the number of dimension and also the total number of samples. So far, in all variations, we are able to identify the change point pretty accurately using the offline detection method described above. The plots below are for dimension = 6 and number of samples = 600 with change point at 300.

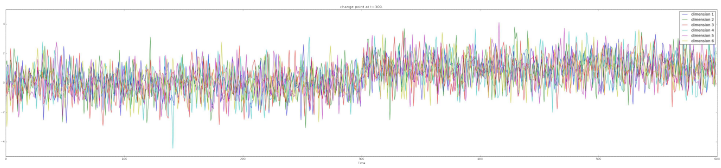

Figure 4: Multi dimensional time series.

1.4 Online Change point detection

Online detection of change point is slightly difficult than offline detection. We have a series of observation defined by $x_1, x_2, \ldots, x_n$. Each observation denoted by x_i is assumed to be made of N samples.

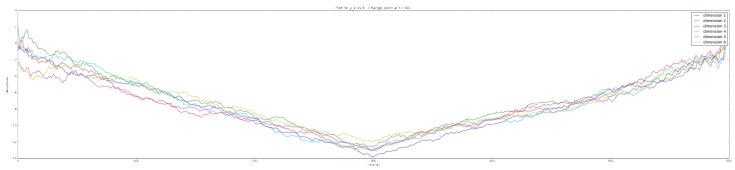

Figure 5: Value of y_k with respect to various k.

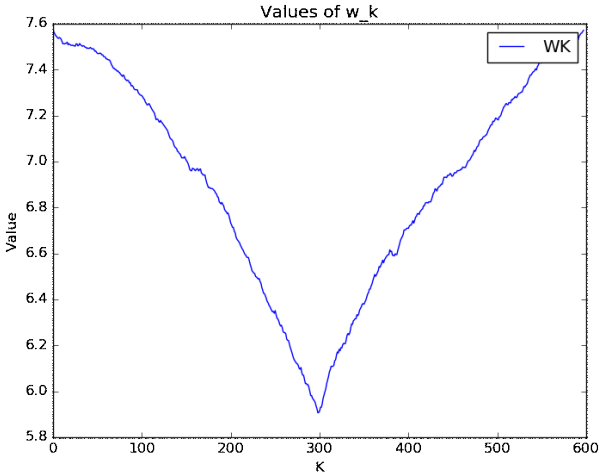

Figure 6: Value of w_k with respect to various k.

We start by defining a sufficient stastic as:

$$s_i = \ln \frac{p\theta_1(y_i)}{p\theta_0(y_i)} \tag{12}$$

We also define a decision rule d given by

$$d = \begin{cases} 0 & if S_1^N < h; H_0 \text{ is chosen} \\ 1 & if S_1^N \geq h; H_1 \text{ is chosen} \end{cases} \tag{13}$$

Where h is a threshold chosen by the user. S_1^N is called as the decision function.

We also have a stopping rule which is defined by

$$t_a = N. \min K : d_K = 1 \tag{14}$$

Where d is the decision taken with the aid of the decision function defined above. This method of finding the stopping rule and alarm time is as given in[1]

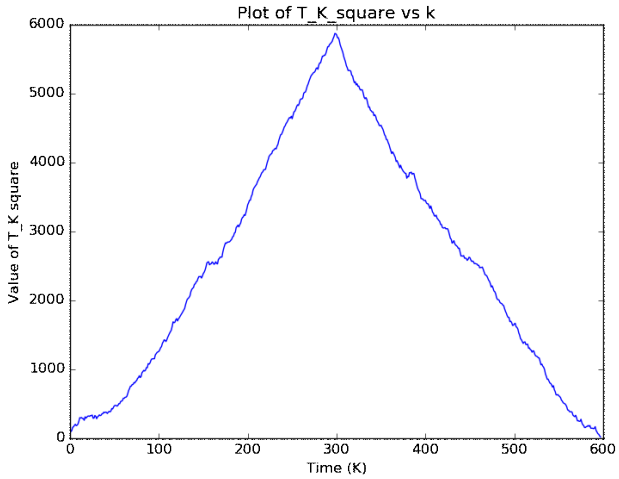

Figure 7: Value of the test stastic T_k^2 with respect to various k.

1.4.1 Experiments — Online change point detection

We run experiments using grid search with various values of h and kappa. In this particular experiment, we used values ranging from 1.5 to 5.5 with step size 0.25 for h and 0.5 to 4 with step size of 0.25 for kappa. We then proceed to find all the values of the stopping rule and also the alarm time for h and kappa combination. The time series, the plots for alarm time and stopping rule are plotted below.

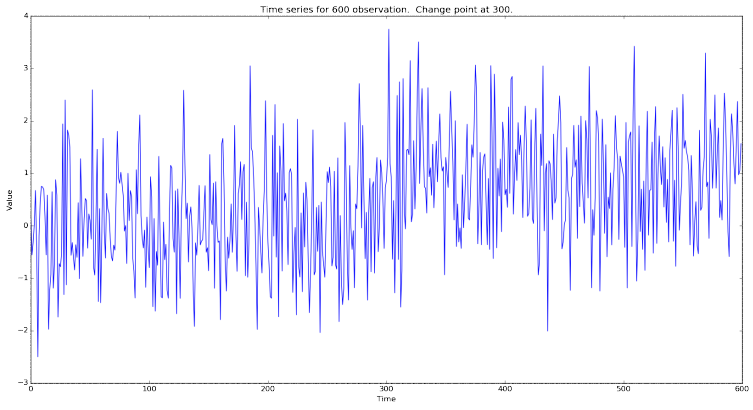

Figure 8: One dimensional time series for online detection.

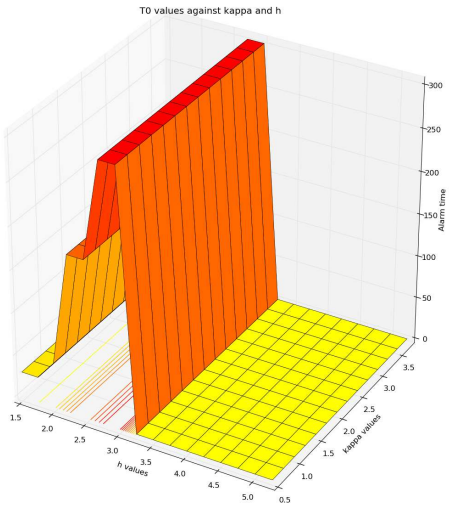

Figure 9: Stopping rule for various values of h and kappa.

2 Spectral Density

Autocovariance function (ACVF) and autocorrelation (ACF), as the name suggests, measures the variance of the time series sample $x_t$with respect to a future time sample x_{t+h}. The formula for finding the autocovariance is given in definition 1.4.4 of [2]. As per the definition

2.1 Auto CoVariance Function

$$\hat{\gamma}(h) \coloneqq n^{-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) \tag{15}$$

Where -n < h < n

2.2 Auto Corelation Function

The autocorelation function is given by the formula:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \tag{16}$$

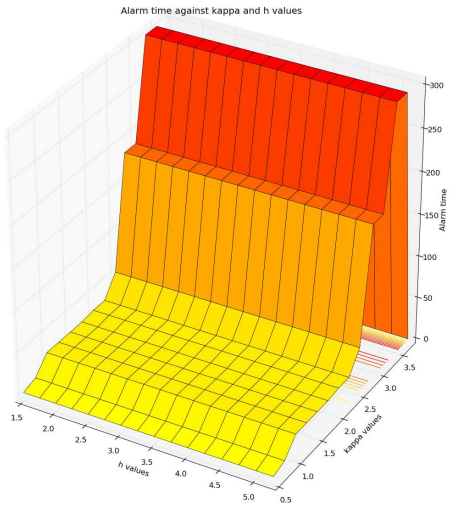

Figure 10: Alarm time for values of h and kappa.

Using these two formulas we can find the ACVF and ACF for a time series. The plot of the time series and the power spectral density obtained by np.fft is given below

2.2.1 Experiments — Spectral density for univariate time series

2.2.2 Experiments — Spectral density for linear combination of sinusoidals

Suppose we have a time series defined as

$$X_t = \sum_{j=1}^{k} (A_j \cos(\omega_j t) + B_j \sin(\omega_j t)), \qquad 0 < \omega_1 < \cdots < \omega_k < \pi \tag{17}$$

The time series is given in Figure 14 The timeseries is generated with values

$$\begin{aligned} K =& 2 \\ \omega =& [1.57079633, 0.78539816], \\ A =& [-0.31201389 - 1.04898091] \quad \text{and} \\ B =& [-0.33909457 - 0.1755208]. \end{aligned}$$

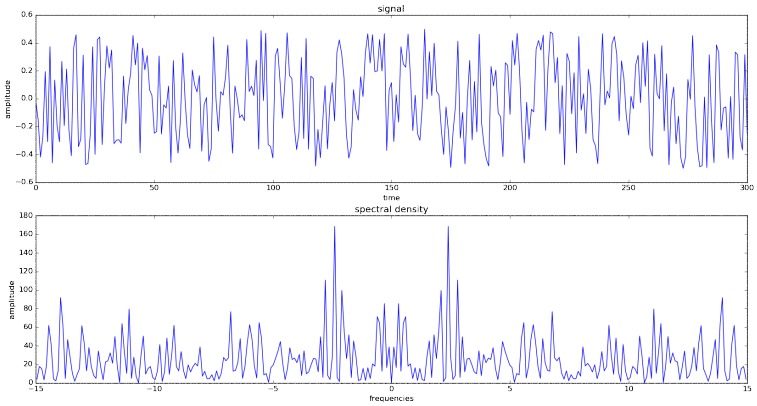

Figure 11: Plot of power spectral density (bottom) obtained after using fourier transform numpy module on the timeseries given in the topplot.

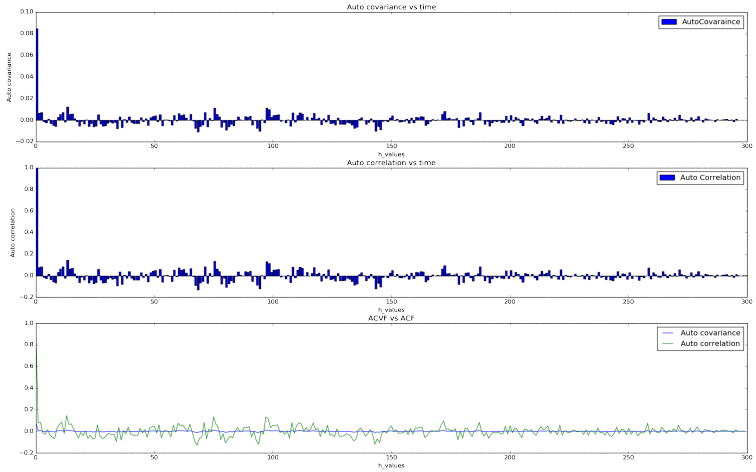

Figure 12: ACVF and ACF of the time series.

The Auto Covariance Function (ACVF) of this time series is given by

$$\gamma(h) = \sum_{j=1}^{k} \sigma^2 \cos(\omega_j h) \tag{18}$$

as defined in [2].
The ACVF as a function of h is shown in Figure 15

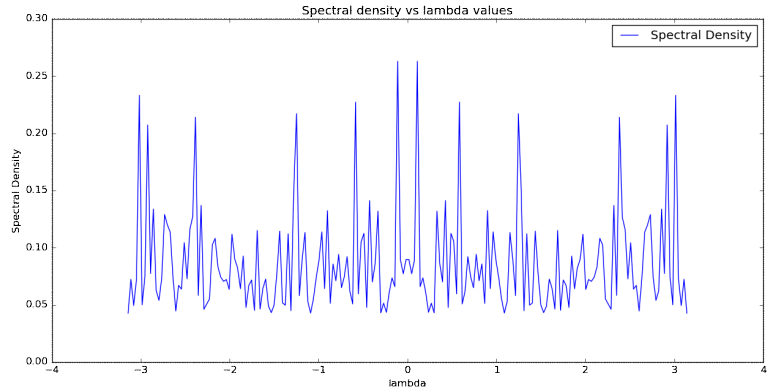

Figure 13: Spectral density for lambdas ranging from -3.14 to 3.14.

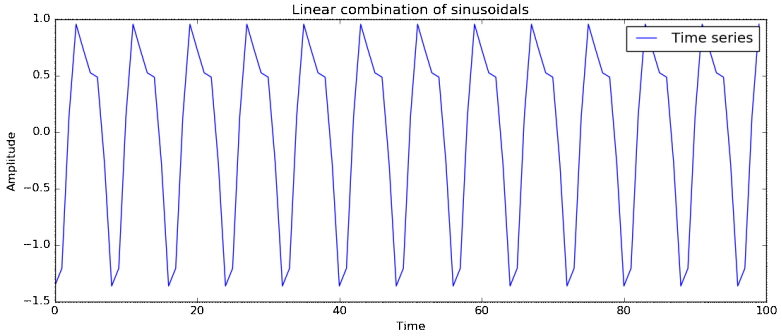

Figure 14: Timeseries generated by Linear combination of sinusoidals.

Also the spectral density is given by:

$$F_j(\lambda) = \begin{cases} 0.0 & \text{if} \quad \lambda < -\omega_j, \\ 0.5 & \text{if} \quad -\omega_j \leq \lambda < \omega_j, \\ 1.0 & \text{if} \quad \lambda \geq \omega_j. \end{cases} \tag{19}$$

The spectral density of the function for $h \in (-\pi, \pi)$ is shown in Figure 16.

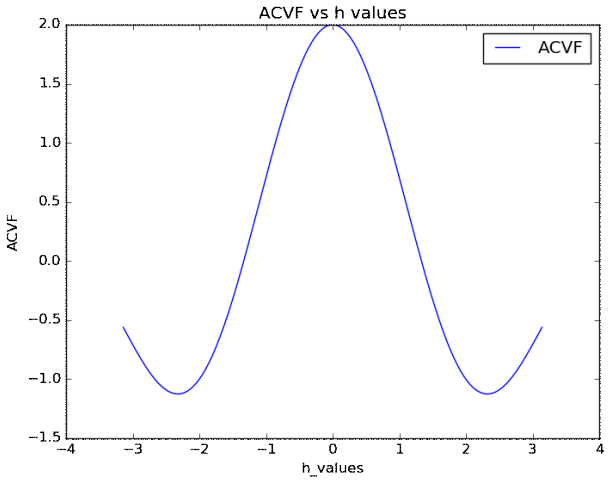

Figure 15: ACVF for the timeseries defined in Figure 14.

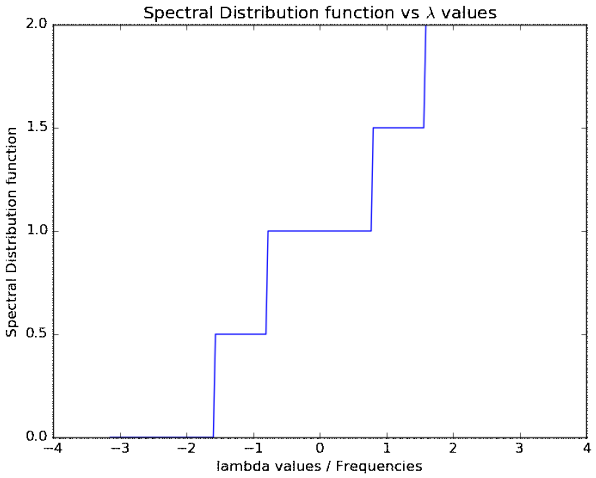

Figure 16: Spectral density for the timeseries defined in Figure 14.

2.3 Peridogram

2.4 Change points using spectral density

References

[1] Michele Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. PTR Prentice-Hall, Inc., 1993.

[2] Peter J Brockwell. *Introduction to Time Series and Forecasting, Second Edition*. New York, Springe, 2002.

[3] B.E. Brodsky and B.S. Darkhovsky. *Non-Parametric Stastical Diagnosis — Problems and methods*. Springer — Science + Business media, B.V., 2000.

[4] Jie Chen and Arjun K. Gupta. *Parametric Stastical Change Point Analysis - With applications to Genetics, Medicine and Finance*. Birkhauser, 2012.