LIGO-T030157-00-D

SURF Progress Report

Kevin Chan
LIGO Laboratory, California Institute of Technology
Mentor: Alan Weinstein
Co-Mentor: Julien Sylvestre

August 1, 2003

1 Title of Project

The title of my project is **"Evaluation of Techniques to Identify Coincident Bursts in Data from two LIGO Interferometers"**.

2 Work Completed

In the past month I have continued to develop the methods to identify the coincidence of signals from two data streams, and I have begun some preliminary evaluation of the methods. At the beginning of July I had a program to calculate a statistic, the linear correlation coefficient, that gave a measure of how well the data from two streams matched. I improved this program in several ways. I added a parameter to specify the size of the duration over which to calculate the correlation. A larger interval incorporates more data, which is useful if the burst we wish to find is relatively large. A smaller interval focuses on more specific data and is useful if the burst is relatively small. Thus the program has greater ability than before to detect coincidence for a variety of burst durations.

Another addition to the program is the calculation of a series r_k of correlation coefficients as a function of the lag time between two data sets. For actual LIGO data, burst detection in LIGO interferometers can be separated by up to 10 ms, but we do not know this time lag beforehand. The program can now search over the possible time lags for coincidences.

I have also written functions to evaluate the methods. These functions calculate characteristic curves, which are plots of the probability P_d of detecting a signal which is present in both data streams versus the probability P_f of falsely detecting a signal which is not present in the data streams.

I added other related methods for determining coincidence. One method is to Fourier transform the series of linear correlation coefficients in the time domain to a series in the frequency domain. Then this method takes the maximum of the modulus of each sample of this series and compares this maximum to a threshold that we set. If the maximum exceeds the threshold, there is detection; if it is below the threshold, there is no detection.

All the methods work by comparing a statistic computed from the data to a threshold value. The three other methods make use of the cross spectral density of two data streams. These methods fix an interval of data in one stream and compute the cross spectral density between this interval and a corresponding interval in the other stream. The second interval can be shifted by a varying amount of lag. One method finds the maximum power of the cross spectral density for a specific lag time, then finds the maximum of these maxima over all lags within a specified range. The second method averages the power for a specific lag time, then finds the maximimum of these averages over all lags. The third method computes one cross spectrum over a long interval, finds the maximum power and the frequency at which it occurs, and determines the appropriate lag time via the phase of the value of the cross spectrum at that frequency. The maximum in each case is compared to a threshold to determine if there is a detection of coincidence.

Some characteristic curves were plotted for these methods to get a general idea of the performance of each method. Following is the general procedure. I choose a waveform as a signal. I used sine-gaussian and gaussian waveforms. I choose a threshold for detection. For a given method, one threshold will give one value of P_d and one for P_f. For the P_d value, I generate a large number of pairs of sample data streams consisting of the signal injected into random gaussian white noise. A method is run on each of these samples; the number of detections divided by the total samples is the probability of correct detection, or efficiency. Next, the same number of samples is created but with only white noise and without signal. The number of detections divided by the number of total samples is the probability of false detection. By varying the threshold value, we get a set of points that can be plotted to give a characteristic curve.

From these first tests, it appears that the method of taking the Fourier transform of r_k, the series of linear correlation coefficients, works the best. I will be running more tests in the next few weeks.

3 Process and Observations

The process of my work begins with a discussion with my co-mentor, who describes the algorithm and gives a general overview of the theory. From this I attempt to implement the algorithm and understand better the background theory. The programming has been a process of coding, testing, modifying, retesting, finding problems and solving them either on my own or after consultation with my co-mentor. The majority of my work has been in getting the methods to a working stage and improving them or changing them as problems come up or as I have new ideas. The process involves my creating tests for the purpose of observing the correctness of the program before running the programs on real data. After trials and modifications, I have observed the programs work as I expect on test situations.

From my preliminary evaluation of methods, I observed from the plots of characteristic curves that methods working in the fourier domain performed better than those working in the time domain. This corresponds to my idea of how these methods should perform, since for the signals I used, which were mainly periodic, the fourier methods pull out periodic signals from noise better than time domain methods.

4 Problems

One significant problem is our ability to compute the lag time between signals in the two data streams. In the time domain, it is easier because we can specifically compute correlations for each lag. In the fourier domain, I have also used a method that computes the cross spectrum for each lag, but this method requires more computing time than other methods. In methods in which only one cross spectrum is computed, the time lag is related to the phase of the value of the cross spectrum at a maximum. However, since we can only determine the phase modulo a value of 2π, there is some ambiguity. This problem may be inherent in the method.

I found a related problem in my method of computing the cross spectrum. I used a Matlab built-in function to do the computations, but my method computed the wrong lag value. To figure out the problem, I had to understand how the built-in function worked. I realized that it did not use the optimal method for my purposes. I solved the problem by writing out my own method to compute the cross spectrum.

My method of testing functions was the source of another problem. The methods of testing were based on the structure of my original functions, but since my functions constantly changed, I had to change my test methods often as well. This required more effort than I wanted. I have worked to make my test methods more modular and flexible, but as additional changes to my functions are made, I may need to adapt again.

5 Remaining Goals

The overall goal remains to determine the most efficient method I can develop to determine coincidence of signals. Towards this end I plan to run many more tests with more varying waveforms and with more realistic data than I have used so far. I wish to have strong evidence to support my choice of the best method. Finally, not only should this method be the best of the methods that I test, but it should also be proficient for use in real analysis in the future, when real gravitational waves may be detected.

6 Interactions with Mentors

Most of my interactions with my mentors consist of conversations with my co-mentor, Julien Sylvestre. In general, he gives me assignments and suggestions and allows me to work on them somewhat independently. I have met with my official mentor, Alan Weinstein, a couple of times to discuss the progress of my work.

Signature: