An Investigation of Wavelet Design using Genetic Algorithms

Yvette Hill, Steven G. O'Keefe, David V. Thiel

Abstract – This paper presents the development of a lightning stroke identification wavelet basis set using genetic algorithms. A basic introduction of the wavelet transform is presented followed by an overview of a genetic algorithm. The frequency content of the best wavelets produced coincided with the spectral content of the spherics examined. As the method is computationally inexpensive, development of the wavelets can be carried out on site. This tailors the identification wavelet to suit spherics at particular localities and allows the wavelet to address changes in frequency due to propagation paths and surface impedance.

I. INTRODUCTION

WAVELETS are mathematical transient functions that when combined form basis sets representing an original time domain signal. Like the Windowed Fourier Transform (WFT), the output coefficients of the Wavelet Transform (WT) indicate the magnitude of the respective wavelet function required to reconstruct the original signal. The sum of the products of each wavelet function in the basis set and its respective coefficient will complete the reconstruction.

Using the original or mother wavelet, the basis set is constructed with equation (1). The scaling factor $\alpha$ controls the dilation or compression of each wavelet and the translation factor $\beta$ determines the shift in time. The original power in the mother wavelet is maintained by the $1/\alpha$ factor.

$$w_{\alpha, \beta}(t) = \frac{1}{\sqrt{\alpha}} w\left(\frac{t-\beta}{\alpha}\right)$$  \hspace{1cm} (1)

$$CWT_p^W(\alpha, \beta) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} p(t) w\left(\frac{t-\beta}{\alpha}\right) dt$$  \hspace{1cm} (2)

Equation (2) is the continuous wavelet transform (CWT), where $p(t)$ is the time domain signal and $w(\alpha \beta)$ is the wavelet as a function of translation and scale. The wavelet transform coefficient gives an indication of the power of the spectral content at a particular dilation and translation. The CWT is the crosscorrelation of two signals, thus for a particular translation and dilation, the higher the output coefficients the closer the wavelet represents the original signal. If wavelets can be produced that will represent a generic form of cloud to ground (CG) strokes and other forms of spheric activity then these can be used in an identification/classification technique.

Our previous work investigated the development of wavelet functions using known transient identification functions such as the Malvar Basis set, and those used by Zhang, Goldak and Paulson. [1] [2] These wavelets accurately identified CG strokes but when robust testing was introduced the wavelets failed to ignore similar signals that contained the same frequency content but differences in the attack, plateau and decay.

The purpose of introducing genetic algorithms to produce wavelets for identification is to improve the performance of the WT in the robust testing phase and to optimise its' performance by using a minimum number of dilations.

II. THEORY

The Genetic Algorithm (GA) is a stochastic search method based on genetic reproduction and natural selection. It is designed to evolve towards an optimal solution for a given problem.[3] The initial population or sample set consists of a number of chromosomes that are permutations of genes of varying magnitudes. To improve the efficiency of the algorithm, real valued chromosomes have been used [4]. Each of the calculated functions forms a wavelet that is then allocated a fitness value according to the selection criteria.

The three basic operators then applied are selection, crossover and mutation which all form part of the reproduction stage. A probabilistic selection has been implemented which is based on a chromosomes' assigned fitness value. The higher the fitness value, the higher the probability of selection. The initial population is then adjusted with chromosomes of a high probability of selection replacing those of a low probability and sets of reproducing parents are randomly selected. Offspring are then produced by the crossover operator.

The simplest crossover function is the single point crossover. Here a randomly selected common point is found in both chromosomes and an exchange of genes from that point is performed.[5] Thus two parents produce two offspring for the next generation. One may choose to completely replace the old population with the new one, known as generational GA's or overlaps can be introduced from one generation to the next, referred to as
steady-state GA's. Offspring may also be produced by the mutation operator.

Mutations occur when a chromosome of a previous generation is duplicated incorrectly and placed in the new generation. Duplication error is normally performed on one of the genes within the chromosome. Mutations are performed with a low probability of occurrence usually within 0.01 to 0.1. The GA will continue to breed new generations until the termination criteria is met. This criteria could be as simple as defining the maximum number of generations or defining the fitness criteria within a certain acceptable range. A more complex form would be meeting a population convergence criteria. The final generation should produce, within its population, the wavelet that will provide the best solution. That is the wavelet which has the maximum fitness value and therefore optimal performance.

III. EXPERIMENT

The GA implemented used the Genetic Algorithms for Optimisation Toolbox (GAOT) in Matlab developed by Houck, Joines and Kay. By implementing a windowing function and introducing normalisation, the design of the wavelet complied with the following properties stipulated by Coifman and Meyer and cited in Hubbard. The decay of the wavelet in both the positive and negative direction is greater than or equal to \(\frac{1}{\sqrt{\omega_0}}\), where \(\omega_0\) is the centre angular frequency of the wavelet's Fourier Transform and the integral of the wavelet over time is equal to zero.

Variations in population size, chromosome length, and windowed trigonometric functions, were investigated in order to determine optimal performance. The original populations in each run varied between 10 to 100 members. It was decided that each chromosome should have nine genes representing frequency, phase, and the amplitude of three Blackman windowed trigonometric functions. The final function selected was the summation of 3 sinusoidal signals. One hundred steady-state generations provided a data set that would converge faster than that of one with a high replacement percentage.

Selection was implemented using Normalised Geometric ranking and proportionate selection, commonly called roulette-wheel selection, with equation (3) determining the probability of selection:

\[
P[\text{selecting the } i\text{th individual}] = q'(1-q)^{r-1}
\]

where:
\(q\) = the probability of the best chromosome being selected;
\(r\) = the chromosomes rank according to its fitness;
\(P\) = the size of the population
\(q' = \frac{q}{1-(1-q)^P}\)

Each of the chromosomes were placed in rank order according to their assigned probability of selection. Chromosomes with higher probabilities replaced lower ranking positions and therefore participate in reproduction more often. A rapid search is provided by a high crossover. Eighty percent of the population was selected to be replaced by the crossover operator and ten percent of the population was mutated.

The fitness function is designed according to the criteria of selection. The function selected for this purpose was the CWT of a known CG signal. The integral of the squared output coefficients of this function would provide a fitness level. Comparisons were drawn between a CWT of one dilation (1D) and one that involved 4 dilations (4D).

IV. RESULTS

A number of test signals were generated in order to examine the ability of the GA to identify the frequencies present. The GA using the 1D CWT continually developed wavelets where the sinusoidal functions contained the spectral content of the test signals. The 4D CWT created a mother wavelet with a frequency content that would maximise the CWT output sum of all the dilations. Hence the mother wavelet frequency content would not focus on the spectral content of the original signal but select a frequency that would allow each of the dilations to address part of the spectral content present.

The test signals were then replaced by a far field CG signal recorded at Brookstead. We found the 1D CWT would focus on the high power spectral content of the original signal. The spectra of the mother wavelet from the 4D CWT focused slightly above the high power spectral content of the CG in order to obtain high coefficients from the dilations. Best results were achieved when each dilation coincided with the high spectral content in the spheric.

A 1D wavelet representation is provided in figure 1 along with the original time domain signal and the Fourier transforms of both signals. Figure 2 shows the same CG signal but a 4D wavelet has been developed and is shown with its Fourier transform.
In this paper we have presented a GA to design CG identification wavelets. Some additional routines were included to maintain the properties required in wavelet design and restrictions established to maintain the integrity of the system. The GA is computationally inexpensive which permits the on site development of tailor made identification wavelets. As the frequency content of CG signals is dependent on the propagation path and the surface impedance of the measurement site, this will be an added benefit. Satisfactory results have been obtained although thorough testing of a large number of transient events still needs to be conducted.

REFERENCES