Scientific Computation Project 4

Lorcan Donnellan - 01818119

April 7, 2021

Question 1

This paper describes three core concepts for time and frequency analysis: Fourier spectra, spectograms and scalograms.

Fourier series and its mathematical background are introduced. We see how the Fourier series is a superposition of waves of frequencies $1/\Omega, 2/\Omega, 3/\Omega, ...$ This explains why it is observed that each frequency in a Fourier spectra is a multiple of the fundamental frequency: $1/\Omega$. Each component wave has an energy or amplitude equal to the absolute value of its corresponding coefficient: $|c_n|^2$. It's this energy that is plotted on the y-axis of the spectra, indicating the relative strength of each frequency, all of which we seen in the notes in this course as well. These wave energies are related to the overall sound signal by Parseval's Equality, for which a continuous version is given in the paper (equation 2.5). This relates to material covered during the term, during which we saw the discrete version of Parseval's theorem where the background energy $|c_0|^2$ is discarded as negligible. The two equations differ in an integral versus a summation sign on the left hand side and arrangement of constants. The summation and integral are obviously analogous between discreet and continuous. Fast Fourier Transforms (FFTs) are also introduced, which approximate Fourier Series coefficients in DFTs (discrete fourier transforms). We know from this course already that this procedure computes all n coefficients in O(n $log_2(n)$). These DFTs can be inverted and preserve energy up to a scale factor. A shortcoming of Fourier spectra becomes clear when analysing signals containing several notes. Each note has a different fundamental frequency (and therefore overtones) and they all appear on the spectrum together, so it becomes difficult to decide which notes correspond to which frequencies as they all appear on top of eachother in the same spectrum. This is where spectograms become useful.

Spectograms are found by multiplying the sound signal $f(t)$ by a superposition of time windows $\{w(t-\tau_m)\}$ that sum to 1. The windows are 1 in the time interval centered at τ_m and smoothly decrease to zero at either side of this interval, so that each window is zero outside of $[\tau_{m-1}+\delta, \tau_{m+1}+\delta]$. This creates a sequence of subwaves $\{f(t)w(t-\tau_m)\}$ which we can discretize and apply FFT to in order to generate Fourier coefficients like before, with the improvement now that only coefficients for the wave in $[\tau_{m-1}+\delta, \tau_{m+1}+\delta]$ show up in this time domain. As the figures in the paper show, the frequency is now on the y axis and time is on the x axis, similar in appearance to bar codes. This gives us a description of the wave frequencies at each time segment in a way which is easy see. We can invert this process, calculating the inverse FFT of each time window's subwave and adding the results together to get back to the discrete original wave values $f(t_k)$. This can be used for compression, where we simply set all values in the spectogram less than a certain threshold to zero before carrying out the inversion process. This results in a version of the original sound wave with small energy frequencies "compressed" out. We discussed inverse fourier transforms during the term, so it was an interesting extension to see how this is used in a real-life application. I was keen to learn more about these methods, so I investigated reference [12] in the paper. The paper shows how we can leverage orthogonal bases to provide sparse representations of certain types of signals with few vectors. These are used for approximation in bases. Essentially, a transform coder decomposes a signal in an orthogonal basis and quantizes the decomposition coefficients. This choice of basis is important, to achieve a high compression rate, the transform code must produce many zero-quantized coefficients, and thus define a sparse signal representation. This is an balance to strike and it was interesting to read of how local cosine bases (corollary 8.1 in book) and lapped orthogonal bases (equation 8.96 in book) could be used for this purpose in audio signals.

Something I learnt that I thought was interesting in the paper was that frequencies of musical notes of consecutive octaves (ie. low C, middle C, high C, etc) form a geometric sequence where each element is equal the previous element multiplied by two. This is described neatly in Music, Physics and Engineering - Olson, Harry F (reference [16] in paper]. Furthermore, ours brain interpret two notes with a ratio equal to any multiple of 2 as similar, or in a sense congruent. This interpretation is also discussed in J.G. Roederer, Introduction to the Physics and Psychophysics of Music ([18] in the paper). In fact it is the ratio between two notes which our brain interprets as their relation, not the difference as we might intuitively expect. As we move up the octaves the distance spanned by an octave actually increases, it doubles each time. Likewise, the difference in frequency between any two notes in one octave is exactly half the difference between the same two notes an octave up. What remains the same is the ratio between the two notes.

The scalogram uses this information by taking the spectogram one step further and incorporating a logarithmic scale of base 2 on the y-axis. In this sense the octave stretching is captured. Scalograms are calculated by a discretization of the continuous wavelet transform (CWT). CWT introduces a dilation in the frequency axis that gives us the logarithmic output. For a wavelet function g, CWT is defined in equation 4.1. We carry out this calculation for a series of values of s. These then correspond to different frequencies on the y axis, and give us a time-frequency distribution of magnitude. The y axis represents the fundamental frequency by 2^i for a range of i in the paper. This fundamental frequency is defined on the parameters of the wavelet function. These are useful where spectograms become crowded in regions of many similar frequencies, as the scalogram "zooms" in. This is seen between figures 2.2 and 4.3 in the paper.

Question 2

The spectogram and the scalogram appear to be quite similar for the test signal, and this is because they are. The spectogram in general displays frequencies plotted against time where the y axis is linear. The scalogram does the same with the key difference that the y axis is logarithmic, with values equalling $v_0 s^{-1}$ in the notation of the paper. They are calculated differently also. The logarithmic scale is seen to be useful when analysing signals with several closely located frequencies on the linear scale, as they can be very difficult to read of a spectogram. This is seen in figure 4.1 (a) of the paper. A scalogram "zooms in" to these, by use of the logarithmic scale, as is seen in figure 4.1 (c) in the paper. Also the scalogram is more musically suitable, as the difference factor in frequency between two notes an octave apart is 2. Therefore in this sense, the scalogram can represent each octave with equal space in the graph.

In this test signal example, even in the spectogram the signals are adequately spaced out, so this "zooming in" advantage is rendered obsolete. In fact, the two figures are generally quite similar. This is evident in our figures. We see strong black bars in the spectogram at the same times we see high test signal amplitude in the signal. We plot the signal itself also to confirm this. These roughly higher energy times are centered around t=0.2, 0.5, 0.8, with the reason for these locations taken directly from the equation 4.7 in the paper. There are gaps between these bursts in the every figure. The two graphs agree with respect to the zero areas as well, the same gaps appear with respect to time and frequency. Also, the frequencies we see higher energies for are 160 Hz, 320 Hz, and 640 Hz in both the spectogram and scalogram graphs, as expected based on the values of v1, v2 and v3 in the original equation. So the peaks of the spectogram and scalogram agree with respect to time and frequency. This is the key similarity between the spectogram and the scalogram.

Another similarity between the graphs is that the 160Hz and 320Hz frequency peaks are lower than the peaks seen for the 640Hz frequency. This can be observed by the color intensity, according to the color bar on each graph. We have plotted both with identical color schemes for clarity, with a color bar to indicate darker means greater. Clearly the two peaks located roughly at t=0.5 and t=0.8 for the frequency 640Hz are darker compared to the other frequencies. This is due to the coefficients in the equation 4.7, where the waves with frequency 640 have higher magnitude coefficients.

An observable difference between the two graphs is that the lines in the spectogram are slightly thicker. This is related to the method more so than the actual signal. There are only 26 different frequencies in total in the grid for the spectogram. So there are 5 frequencies between each grid line on the y axis, where moving from one grid line to the next is an increase of 200. By comparison, between every $v_0 2^i$ and $v_0 2^{i+1}$ there are 10 equally spaced frequencies. In these inbetween frequencies do not have any magnitude in both graphs, because the test signal has clearly defined frequencies. Therefore the scalogram is more tightly squeezed from both sides, making the lines slightly thinner.

So all in all the two graphs are very similar, with the main differences being in how they are calculated, as is outlined in the paper, and the scaling of the y axis. The two are quite similar in this case, but the differences in usefulness can be seen more evidently for other cases.

Figures for Question 2

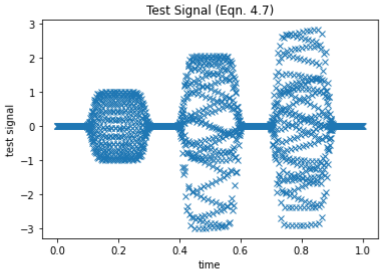

Figure 1: Test Signal (Eqn 4.7)

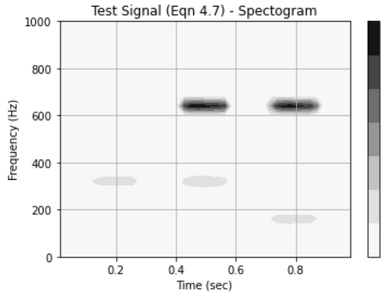

Figure 2: Test Signal: Spectogram

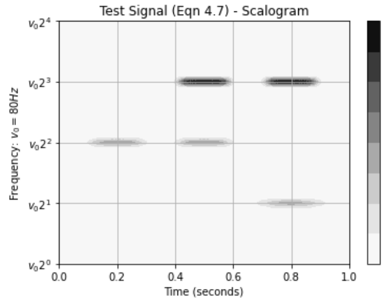

Figure 3: Test Signal: Scalogram

Question 3

In this section we will analyse a musical excerpt with respect to corresponding scalograms. The excerpt is 15 seconds long and consists of three different tunes. Each part is played at a different tempo, with different instruments, at different levels of volume. For this reason in the figures below we provide a scalogram for the entire excerpt, and for each individual part so as examine holistically and to zoom into the details of each respectively. We plot these scalograms for audible frequencies only. Humans can hear frequencies between 20 Hz and 20000Hz. Beyond this threshold waves become ultra sonic. We will not comment on frequencies beyond this threshold in this part of the analysis. We found that the most audible frequencies lay between 2^6 and 2^{14} for all parts of the excerpt, so we set our y limits appropriately. As in the scalograms in the paper we chose a logarithmic scale on the y axis with base 2. This is in tune with (pardon the pun) how octaves in music relate to one another, where a frequency produced by a single note doubled is the same note but in the next octave. We use the morlet2 function from python. The key parameter is set at w = 10 for the following analysis as important features can be seen for this value. It gives frequencies adequately accurately located along the frequency and time axis. The scalogram for w=5 is also provided below next to the scalogram for w=10. We see the areas of high value seem to be larger. This is good in one sense, but it dilutes the accuracy of the frequency estimation, so we use w=10 for this analysis. Going lower than w=5 this trend continues. Bright spots become larger. The trade off for increased time accuracy is decreased frequency accuracy in this sense. We also use a different color map for this exercise because I think it is visually easier to read.

Comparing the scalogram for the three parts, it is clear they are each different. One factor is that each contains music played by different instruments. This is a key difference between the anslysed pieces in the paper and here. Here we ave more than one instrument playing more than one note, whereas most of the examples in the paper have one instrument playing certain notes. Therefore we can't make some of the conclusions made in the paper. Two different instruments playing the same note may share a fundamental frequency, but even so the different instruments produce different overtones. This is discussed in section 2.1 of the paper and illustrated in figure 2.2 in the paper. Furthermore, the range of different instruments includes different octaves, for example a bass is an octave lower than a guitar. So the fundamental frequencies of different instruments can lie in different octaves to begin with. As opposed to some single note examples seen in the paper, the instruments in the excerpt play a string of notes making up a tune. Therefore we don't expect the neatly spaced out frequencies seen in the paper in figure 4.1 (c). The fundamental frequency corresponding to the note being played will move along with its overtones as the tune progresses. The fact that there are more than one of these leads to a scalogram like the one we see for the excerpt. We see nothing in the scalogram in areas between the parts, which is expected and a similarlity between the parts.

I believe the first part features a viola and either a cello or double bass. The two instruments follw an identical melody in two different octaves and are the only two instruments in the part. In other terms, these are middle range and low range instruments respectively. With a little online research we can test whether this is even plausible based on the scalogram we see. It is found that these instruments have frequency ranges of 125 Hz–1KHz (Viola) and 63 Hz–630 Hz (Cello). These frequency ranges fit what we see in the scalogram for part 1. We seem to observe two strong frequencies moving between 2^7 and just over 2^9. These are indicated by the yellow/green colours. The melody consists of a run from a high to a low and back to a high. This is pretty clear to see for the middle range instrument, but seems to get slightly lost in translation for the lower range one. This could be to do with that lower ranges are harder to pick up on, less energy. We also observe dark green areas between 2^{13} and 2^{14} while the music is playing. These could correspond to overtones from the instruments. Although they are very high, right at the top of the audible spectrum, and are more likely scratches and other unavoidable sounds made from playing a wooden instrument. Sibilants like the S's of words and sizzles of snare drums are typical of what tends to be found in this range. Their darker colour suggests we just may not notice them clearly in the mix.

The second part of the excerpt is a different tune and is played by different instruments to part 1. I would guess the instruments playing are a viola and a violin, which are middle and high range string instruments respectively. The frequency range of a viola is 125 Hz–1KHz, while the range of the violin is 200 Hz–3.5KHz. This once again fits into the range we see on the scalogram. We see the strongest frequencies between just below 2^8 and just over 2^{10}, in keeping with this change in instrumentation. Furthermore, I would argue that the stop start nature of the melody in part 2 is also recognisable. From listening to the excerpt, it's obvious the part 2 melody does not follow the same melody as the first. The melody in the first was a smooth run, from high to low then back to high. The melody in part 2 is instead more stop start, both instruments go from a lower note to a higher note and back with increasing urgency as the part goes on. This corresponds to the sparsity of the zoomed in scalogram for part two. The leaps or jumps from one note to the other can be seen, where as in part 1 the scalogram almost intuits one long signal changing frequency as it goes. Once again, we see the dark green areas between 2^{13} and 2^{14} while part 2 is playing. We commented on this for the previous part and the same applies here.

The third part is a different tune to the second part, and it sound like it is being played on a viola accompanied by a violin. The lower instrument actually plays the same tune as in the first part, however the accomapanyment is different, with the higher range string playing elongated notes. This trails off towards the end. All in all we seem to see less strong frequencies for this part. This could be due in part to the energy being lower in general for this part. This is clear from the first figure where just plot the signal. The amplitudes are much lower. As we seen earlier, there is some cross over in the frequency ranges of the instruments, so it is still plausible that our spectograms has captured that aspect of the music here too. We also include figure 10, which is a sclogram like the others with a different color mapping, black and white. Here we see some registered frequencies for over 2^{14}. These could potentially relate to Idiosyncrasies of the recording equipment, producing higher range frequencies than we can typically hear. Alternatively they could an artifact of computing the transform at frequencies that are too high.

Thus we have analyzed the musical excerpt with respect to its scalograms. There are some definite similarities and differences to be heard in the tunes which coincide with what the scalogram tells us with respect to frequency strengths and timing.

Figures for Question 3

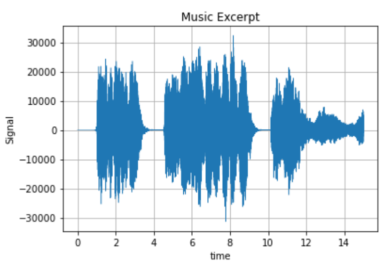

Figure 4: Musical Excerpt: Signal

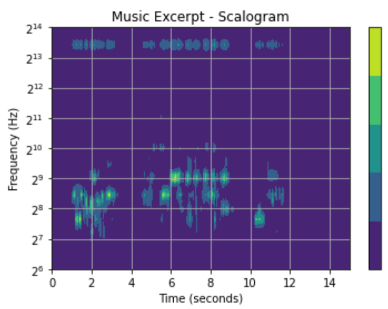

Figure 5: Entire Excerpt Scalogram: w=10

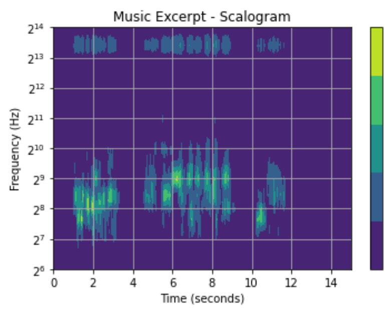

Figure 6: Entire Excerpt Scalogram: w=5

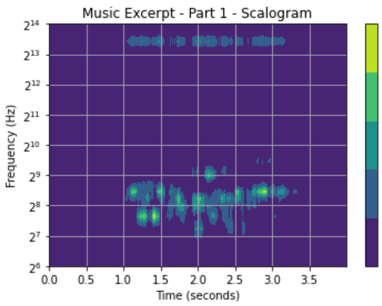

Figure 7: Part 1 Scalogram

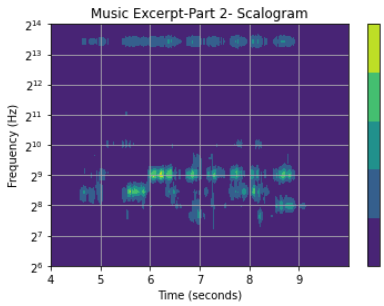

Figure 8: Part 2 Scalogram

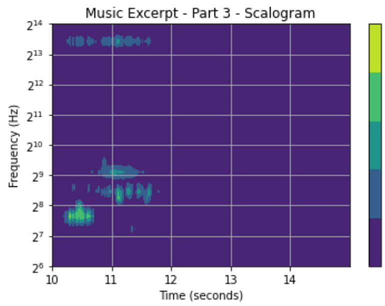

Figure 9: Part 3 Scalogram

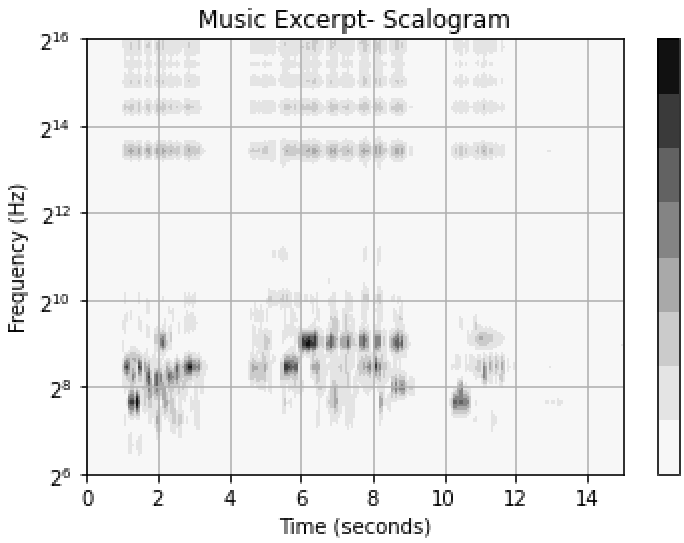

Figure 10: Music Excerpt - Scalogram (including ultrasonic frequencies)