A New method for High-resolution Frequency Measurements

Herbert Weidner [a]

Abstract: Shifting the frequency of a spectral line to a frequency bin of the FFT and omitting the window function improves significantly the frequency resolution. The algorithm works without zero padding and is applicable even with bad SNR.

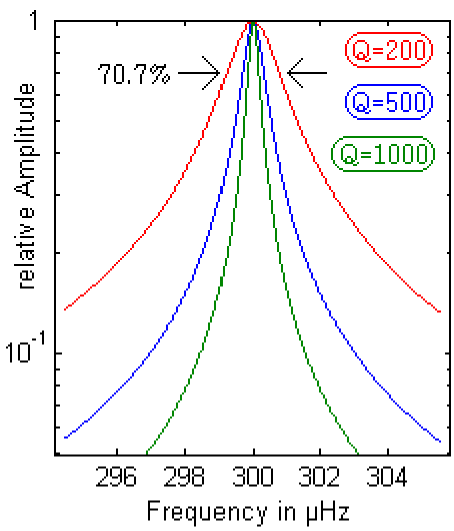
Introduction

The Earth has many different resonances that are excited by earthquakes. The study of the internal structure of the earth requires the precise knowledge of the frequencies. Permanent weak earth-quakes affect the signal-to-noise ratio of gravity data and make it difficult to accurately determine the spectral frequencies. Because the phase insensitive FFT can determine the frequencies only relatively inaccurate, a new method is presented here that is applicable even in bad SNR for its extreme narrow band.

The use of short periods has several advantages: short-term frequency changes can be tracked and measured; Short-duration spikes and amplitude reductions hardly affect the accuracy. By the way, the process acts as a precision rectifier, which can also measure rapid changes in amplitude.

The required Bandwith

In nature there are no undamped oscillations with a constant amplitude and (theoretically) zero bandwidth. Signals have always a finite bandwidth, which is why each measurement may be affected by neighboring frequencies. In particular geo-physical signals are very noisy and must be filtered with narrow bandwidth before any processing (in compliance with the line width).



In most cases, the amplitude of a spectral line can be described by the formula

$$y = A_0 \cdot \exp\frac{-\omega t}{2Q} \sin(\omega t + \varphi)$$

with the initial amplitude A_0 past the excitation of the oscillation and the frequency $\omega = 2\pi f$. No exciting earthquake is synchronized with our clocks, so we need the phase φ to describe this shift. Due to friction inside the earth, the oscillation loses energy and the amplitude descends. The rate is characterized by the quality factor Q.

Changing the amplitude always broadens the line width, as shown in the picture. The total bandwith of an exponentially damped oscillation is infinite and must be limited before analysis. It is hard to define the required bandwith Δf of these filters, because every cutting of sideband frequencies at least changes the envelope of the damped oscillation. An estimate with the formula $\Delta f = f/Q$ leads to a much too small value, because it is based on an amplitude reduction to 70.7 %. At Q = 500, a reduction to the 10% - level increases the bandwith from 0.6 µHz to 6 µHz and decreases the SNR significantly. Gravity data are normally buried in noisy data streams of gravimeters and a reliable detection requires an improvement of the SNR by narrow frequency filters. Another

a) 1. June 2015, email: herbertweidner@gmx.de

concern is that after interference, some filter types cause a strong phase modulation, hindering an accurate frequency determination. In particular, IIR filters should be avoided.

Reducing the Bandwith

Some of these problems can be decreased if the measuring signal is multiplied by an appropriate exponential function that compensates for the natural reduction in the amplitude. Then, the desired signal is (apart from the inevitable faults) a finite-length sinusoid of constant amplitude and may be described by $y = A \cdot \sin(\omega t + \varphi)$. As the frequency is only approximately known, the bandwidth can not be made arbitrarily small, because then there is a danger to analyze the narrow filter ringing, induced by noise. If the in-band signal changes (or ceases) relatively suddenly, the energy stored from previous times still has to be released, and this is the ringing that is observed after the in-band signal has disappeared. However, this ringing may bridge very short data gaps. Experiments with real data have shown that a minimum bandwidth of 0.5 µHz leads to good results.

Of particular importance is the choice of the proper filter type. The subsequent frequency analysis requires narrow-band filters with special properties. IIR filters are not suitable because they produce too much phase modulation. FIR filters have a much lower phase response, but both types tend to distort the measured curve at the beginning of the record strongly. Unsurpassed are Sinc filters having a rectangular passband. The distortions near the end points of the data record are much lower than with other types of filters, an essential prerequisite for the subsequent spectral calculation. The only drawback is the time-consuming process of convolution.

Frequency determination by FFT

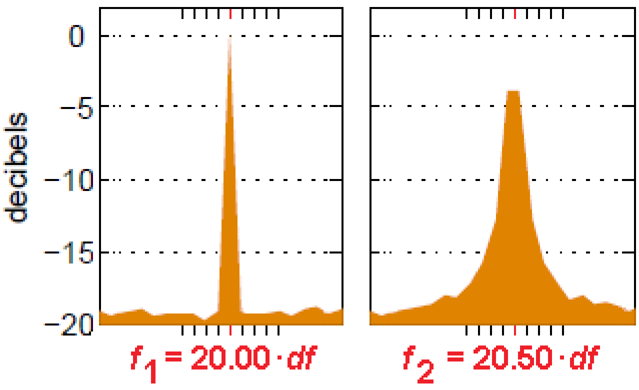
The Fourier transform (FFT) is very well suited to gain a quick overview of the frequency distribution in a data record containing *N* readings. For many reasons FFT is an unsuitable tool for the precise determination of frequencies. It was never designed for this purpose. There are various causes: Each Discrete Fourier transform (DFT) calculates always a limited number of frequencies, depending on the sampling frequency f_S and the data length *N*. The FFT calculates exactly *N*/2 bins between zero and the Nyquist frequency, and therefore the fundamental frequency $df = f_s / N$ limits the accuracy. By zero-padding, intermediate values can be generated to allow a more accurate guess of position of the amplitude maximum. But the accuracy can not be improved, because the extra zeros provide no additional information. Therefore, the method is similar to an interpolation, assuming a symmetrical bell curve.

Additional inaccuracy is caused by another property of the DFT, which users are not always aware. The DFT transforms *not* the isolated input data block, extended by zeros on both sides, but an infinite repetition of this data block. The last sample is followed by the first one, as if all samples would lie on a circle. The data block is repeated endlessly. At the junction of two blocks, almost always a discontinuity arises with devastating effects on the spectrum. This is not based on erroneous data, it is a mathematical consequence of the algorithm of DFT. Details are described here[1] worth reading.

Each discontinuity in the data flow generates additional spectral components. Usually, the most harmful consequences of these discontinuities are reduced by multiplication of the data block with a bell-shaped window function. This corresponds to a modulation of the measured data and generates additional spectral lines in the environment of strong lines, whereby the FWHM is broadened remarkably. All window functions broaden the linewidth, which contradicts the aim of a precise frequency determination.

Window functions prefer the samples near the center of the data record. The first and last samples are more or less neglected, even if they have better SNR than samples in the central region of the record. Weighing the positive against the negative effects of "windowing" is not an easy job and led to the invention of about twenty different windows. Therefore, an algorithm is presented below that avoids a window function.

What happens if one does *not* use a Window (some people call it a rectangular window)? If the actual frequency of the sinusoid happens to coincide with an integer multiple of the fundamental frequency *df*, the maximum value of the spectrum is accurately measured by that bin. Even more important: the neighbor bins are zero, if the signal is undisturbed. No other window function can achieve this selectivity. If the actual frequency of the sinusoid does not match the frequency comb, the results deteriorate.



In the left picture, the frequency of a noisy signal is exactly 20 times the fundamental frequency *df* of the FFT analysis. At the junction of successive data sequences, there is no phase discontinuity. The peak is exactly one bin wide without spectral leakage to neighboring frequencies. A dream result! A Dirac delta function and not a wide bell curve with blurred peak.

In the right picture, the signal frequency is in the middle between two frequency bins, resulting in a maximum phase shift of the signal at the junction of successive sequences. While the amplitude reduction is not serious, the spectral broadening is very disturbing and can not be eliminated by zero padding or other tricks.

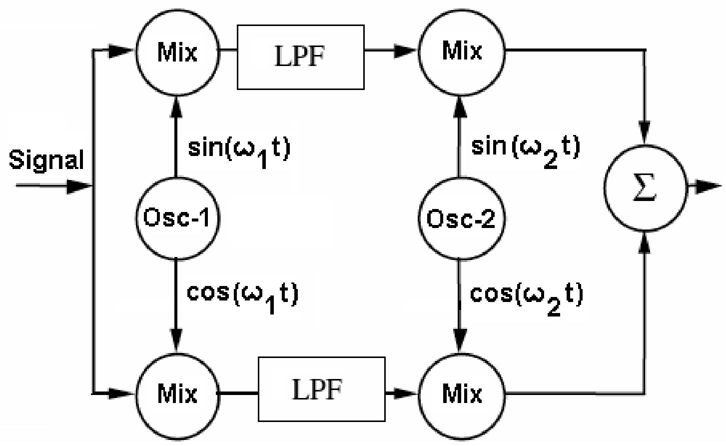
All window functions equalize the two pictures: The left, perfect result (a Dirac delta function) is broadened, the right bell curve becomes narrower. Regardless of their frequency, all the spectral lines are represented by a uniformly shaped bell curve (unless the actual FWHM is even wider).

A problem arises, if two spectral lines have approximately the same frequency but different amplitude. Then the bell-shaped curves add up and the spacing of the peaks decreases. This prevents an accurate frequency determination by FFT.

Tuning the frequency

It is highly unlikely that the frequency of the sinusoid coincides with one of the frequency bins of the FFT. Therefore, the frequency must be shifted by a programmable amount. Geophysical signals are always severely disturbed and require a narrow-band filtering to detect weak signals in noise. Fortunately, a method[2] was developed many years ago by Weaver (for an entirely different purpose), which satisfies both requirements in a perfect manner. It is a *frequency-shifting bandpass filter*. In analog technology, the implementation was difficult, but in the era of digital signal processing this filter is programmed with a few and simple lines of code.

The oscillator-1 determines the central frequency f_1 of the filter and should match the value of the expected spectral frequency. The first two mixers produce the sum and difference of both frequencies. The following identical low-pass filters let pass through only signals within a narrow range around f_1 . The second pair of mixers shift the Signal to the arbitrary frequency f_2 . For $f_1 = f_2$, the circuit is a complex bandpass filter. More valuable is the case $f_1 \neq f_2$, because an arbitrary



spectral frequency may be shifted to any predetermined value. Here, the signal frequency will be shifted so that it exactly matches a frequency bin of the downstream FFT. In radio-frequency technology, the corresponding principle is a superheterodyne receiver, where the signal frequency is shifted to a better processable intermediate frequency without changing the message content.

The precise determination of the frequency

It is highly unlikely that the frequency of a spectral line coincides with a multiple of the fundamental frequency df of the FFT, called frequency bin. Shifting the frequency of the sine wave (with the above-described band-pass filter) to exactly the point of highest resolution makes the window function redundant with all its disadvantages. The correct displacement is easy to check: it corresponds to a maximum of the calculated amplitude of the spectral line.

The process can probably be best explained with an example based on real data. A seismometer in Membach supplied a data record of 256 samples, which were taken at intervals of 360 seconds. In the data, a signal near 318. 4 µHz could be hidden whose frequency is to be determined precisely. The 256-point-FFT calculates the fundamental frequency 10.850694 µHz and multiples thereof. A rather coarse resolution. The nearest neighbor in the frequency comb is bin number 30 with the value 314.6701 µHz. Other bins are just as well suited.

The offset -3.73 µHz is the starting value of an iteration, which uses the *frequency-shifting band-pass filter* described above. The three target frequencies f_{2A} = 314.57 µHz, f_{2B} = 314.67 µHz and f_{2C} = 314.77 µHz yield three amplitudes Y_A, Y_B and Y_C of bin number 30. These three pairs of numbers define a parabola whose vertex is located at offset -3.7913 µHz. A repetition of the process results in the final offset -3.791215 µHz. Now it is an easy job to calculate the signal frequency. Y_B is the average signal amplitude during the sampling period.

If the iteration does not converge satisfactorily after the fourth iteration, it should be terminated due to insufficient SNR.

Summary

The method is based on an infinitely long chain of measurements. The FFT algorithm creates this chain automatically, although the underlying record contains only 256 samples. The chain must not have any discontinuities, which is why only certain frequencies are allowed and the record must include exactly $N=2^m$ samples. With the sampling frequency f_S , the allowable frequencies are integer multiples of the fundamental frequency $df = f_S / N$.

Prior to FFT, the signal *must* be band-pass filtered. After filtering the record, the sampled data correspond essentially to a single frequency of constant amplitude. Since only the amplitude of this single frequency is required, the FFT process may be replaced by the faster Goertzel algorithm. The use of window functions and zero padding destroys the foundations of this method.

This method has all the qualities of a synchronous rectifier and measures very accurately the average amplitude during the measurement period.

In the course of an iteration, the frequency f_1 of the bandpass filter can track the currently calculated frequency. It may also be useful to reduce the bandwidth gradually to improve the SNR (adaptive filter).

[1] Steven W. Smith, Ph.D., The Scientist and Engineer's Guide to Digital Signal Processing, http://www.dspguide.com/ch10.htm
[2] D K Weaver Jr., A Third Method of Generation and Detection of Single-Sideband Signals, Proc. IRE, Dec. 1956