Filtering Seismic Time Series Data in Passive-source Seismic-Processing (PsSp)

Implementation details of the filters provided for isolating seismic signals

Alexander R. Blanchette*

5 July 2023

The goal of this document is to be as useful and accessible as possible. Therefore, it is a *living document*. That means it will change and grow over time in order to ensure the quality of the content. As such, I don't ever truly plan on it being *finished*.

Abstract

Seismic data—whether it is ground motion (displacement), velocity, or acceleration—is recorded as a function of time (time series). Seismic time series recordings contain information from a plethora of sources—natural and anthropogenic. Typically, an analysis task focuses on processing/analyzing signals from a specific target *source*[a]. Generally, seismometers record and are sensitive to a frequency range that is significantly larger than that which any given analytical method/goal requires. Filtering is a critically important tool for isolating signals of interest. In this document I give a brief overview of time series filtering in general and provide the specific details of the filters implemented in Passive-source Seismic-Processing (PsSp).

Plain Language Summary

Filters are an important step in preparing seismic data to be analyzed. In this document I will hint at the basic theory of seismic filtering with most of the focus placed on the specific filters implemented in PsSp. This is intended to serve as a basic guide on the what and how of filtering seismic data without going through any explicit mathematical derivations; instead I hope this will provide a more intuitive understanding to the reader.

[a] *Source* is in the generic sense here; it could refer to the nucleation point of the seismic signals (earthquakes, explosions, etc.) or to structural sources (reflections, refractions, conversions).

*arbCoding@gmail.com

Contents

Figures

Equations

1 Introduction

Filtering of time-dependent recordings (time series data) is a task common among a wide range of different scientific/engineering disciplines. Great advances have been made to analyze as much of the seismic time series as possible—such as the extraction of approximate Green's functions from the ambient seismic noise field. Still, virtually all seismic analysis first requires the seismic time series to be filtered in order to isolate the portion of the recorded data that is most compatible with the intended analysis. In the fairly distant past this was less important due to the limitations of: 1) the available seismographic instruments, and 2) the available computational infrastructure.

At present, relatively inexpensive broadband seismometers can record seismic time series with a flat-sensitivity range starting at a period of 100 seconds (s) (a frequency of 0.01 Hz) up to a high of 100 Hz. Higher quality instruments are sensitive to much longer periods—it is not uncommon to hit 1000 s (0.001 Hz). High frequency nodal seismometers can record time series at frequencies of a few thousand Hz (kHz). Such a large dynamic range of frequencies—containing signals from numerous stationary and non-stationary sources—results in a rather complicated time series. This excess information convolutes the analysis of the signal(s) of interest. Filtering is the solution to this problem; it allows the analyst to isolate the signal(s) of interest by removing the *noise*[1] from the recorded time series. Fortunately—thanks to the many technological advancements over the last few decades—personal computers are sufficiently capable of handling most modern seismic workflows (all except particularly large or high throughput workflows that require more specialized computing infrastructure).

Filtering is a tremendously wide topic of study itself, with a rich history across many otherwise disparate fields of study, well beyond the scope of this simple document. Therefore, I will restrict myself to—at most—provide a brief discussion on the fundamentals of filtering, while leaving many of the details up for the intrepid reader to research on their own. Afterward, I will provide a detailed description of the implementation of filters specific to PsSp. This is not meant to provide a detailed derivation from first principles of the filters employed, but is instead intended to serve as a pragmatic guide on their internal implementation. This is intended as much as a guide to any reader as it is to myself.

2 Background

Filtering of time series data is necessary across numerous disciplines (e.g., seismology, electrical engineering, audio engineering, etc.). Due to its prevalence across a diverse range of disciplines, it has been studied extensively over time—with many of the major theoretical advances discovered more than 100 years ago (Joseph Fourier published on his series expansion in 1822[2]). In fact, the Butterworth filters employed here were first published by Stephan Butterworth in 1930[1]. Of course, there has been significant progress since, both theoretical and practical. While much could be—and has been—written on the history of filtering signals, I think that it is sufficient here to simply say that it is a quite mature field of study. I therefore leave the historical developments for the reader to research on their own, if they so desire.

2.1 Mathematical Background

While I do intend for this document to focus more on the development of the reader's intuitive understanding of the filtering process, it would be a significant disservice to totally forgo all mathematical formalism. In fact, the equations themselves provide a great deal of insight into how filters work. In order to provide a sufficiently comprehensive introduction to the application of filters, I will need to first introduce a few mathematical concepts. In particular, I need to discuss how it is that we take data that is measured over time and *transform*[2] it to the frequency (spectral) domain (via the *Fourier Transform*) and then back to the time domain (via the *Inverse Fourier Transform*). I must also discuss how we take a filter that is defined in the Laplace (Complex-Frequency) domain and apply it to a signal in the (non-complex) frequency domain.

2.1.1 The Temporal (Time) Domain

Observations are always a function of time. Whether the observed process is constant or dynamic, the very nature of measurement is time-dependent itself. We can, at best, make the rather artificial assumption of constancy and pretend that time doesn't exist—but we can never truly remove the temporal component from

[1] *Noise* is a bit of a misunderstood term in seismology, especially with the advances in the last few decades in ambient noise seismology. The word 'noise' is used here to mean 'any signal that is not of interest for a specific analysis' and does not provide any suggestion as to the source of the noise.

[2] *Transform* is commonly used in mathematics to refer to a linear change of basis. That is, to change the meaning of the dependent-variable in a well-behaved and reversible manner.

our measurements. Seismographic stations record ground motion as a function of time. The distinction between measurements of displacement, velocity, and acceleration is simply a matter of differentiation ($x \to v \to a$) or integration ($a \to v \to x$). The distinction between ground motion and the actual instrumental measurement is related to the internal *instrument response*[3].

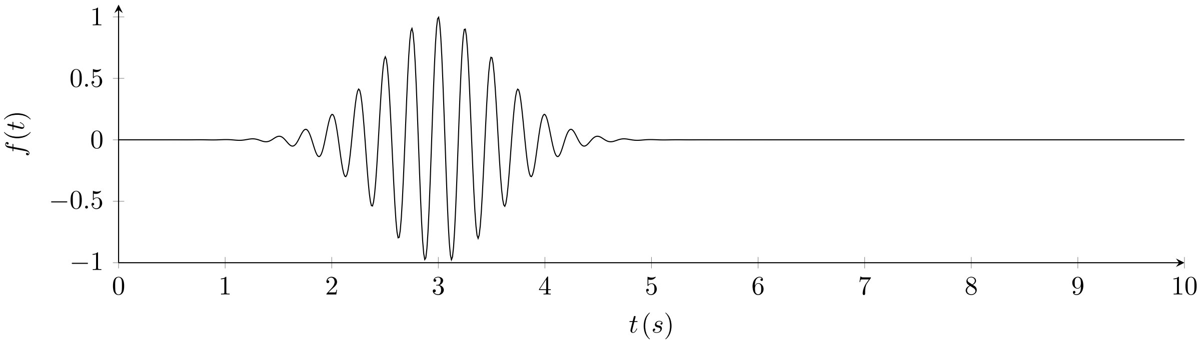

Figure 1: A simple schematic time series.

The time series recorded at a seismographic station is a seismogram. A rather simple example time series can be seen in Fig. 1. This time series shows a wave-packet with an *arrival time*[4] of approximately one second[5], and it is centered at three seconds. For the sake of completeness: the waveform plotted in Fig. 1 was generated using Eq. (1). It is simply a time-shifted cosine function multiplied by a non-normalized *Gaussian*[6].

$$f(t) = \cos(8\pi(t-3))e^{-\left(\frac{2\pi(t-3)^2}{9}\right)} \tag{1}$$

Schematic Time Series

In this case we can determine at least one frequency purely by visual inspection. The cosine function is 2π-periodic, meaning that $cos(t)$ has a frequency of $\frac{1}{2\pi}$ Hz. In this particular case I've chosen to have a frequency greater than one (it is exactly four Hz). There are signals of other frequencies contained in Eq. (1), in fact it contains a range of frequencies—an observation that is not immediately obvious when written as a function of time. I'll return to the frequencies hidden in this example time series once we've been introduced to the Fourier Transform of a function.

2.1.2 The Spectral (Frequency) Domain

To set the stage for us to discuss spectral analysis/processing I must first introduce some basic terminology.

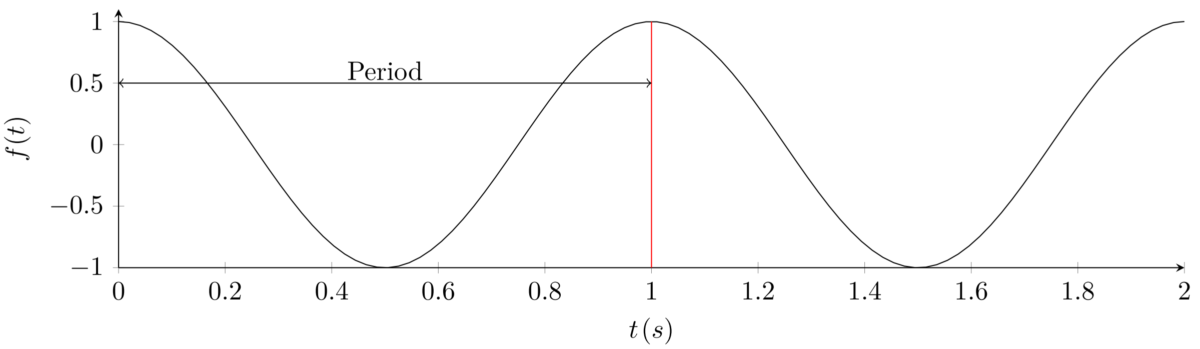

Figure 2: Plot of a cosine function that shows its period.

[3] *Instrument response* removal can be thought of as a form of filtering, but it will be a topic of a separate document.

[4] *Arrival time* is an ambiguous term with multiple potential meanings. In this case I'm referring to the first deviation from zero amplitude (often referred to as the *first break*).

[5] While it is *technically* true in this case that the signal has infinite duration and therefore the first deviation from zero is *technically* at $-\infty$, real data always has some background noise that renders this technicality purely *academic/pedantic*.

[6] *Gaussian* is one of multiple names for this type of function. The name used is typically dependent upon the field of study. Other names you may have heard it called are a *Bell curve* and a *Normal distribution*.

In Fig. 2 we can see a cosine function with a period of one second. It is the time-analogue to the spatial wavelength.

Fourier Series

Fourier Transform

Linearity of the Fourier Transform

Real/Imaginary vs Amplitude/Phase

Convolution

Return to the Example Time Series

2.1.3 The Laplace (Complex-Frequency) Domain

2.1.4 Analysis of Discrete Time Series Data

The Discrete Fourier Transform

Windowing Functions

Discrete Z-Transform

3 The Butterworth Filter

Ideal Filter

Unrealizable

3.1 Low-pass

Test equation reference for Eq. (2).

$$H(s) = \frac{G_0}{B_n(a)}; a = \frac{s}{w_c} \tag{2}$$

Butterworth Low-pass Filter

$$B_n(s) = \sum_{k=0}^{n} a_k s^k \tag{3}$$

Normalized Butterworth Polynomial

$$a_{k+1} = a_k \frac{cos(k\gamma)}{sin((k+1)\gamma)}; a_0 = 1; \gamma = \frac{\pi}{2n}; a_k = a_{n-k} \tag{4}$$

Normalized Butterworth Polynomial Coefficients

3.2 High-pass

3.3 Band-pass

3.4 Band-reject

Glossary

frequency The number of cycles completed in one second. 1, 3, 4

period The amount of time it takes for a process to complete a full cycle. 3–5

wavelength The distance over which a wave completes a full cycle. 5

Acronyms

PsSp Passive-source Seismic-Processing 1

References

[1] Stephen Butterworth et al. "On the theory of filter amplifiers." In: *Wireless Engineer* 7.6 (1930), pp. 536–541 (cit. on p. 3).

[2] JBJ Fourier. "Théorie analytique de la chaleur: Paris." In: *Académie des Sciences* 3 (1822) (cit. on p. 3).