T. A. Ossewaarde

Mathematical analysis of intrinsic mode functions

Master thesis

Thesis advisor:
Dr. O. W. Van Gaans

Date
master exam: 11 Januari 2017

Mathematisch Instituut, Universiteit Leiden

Introduction

Time series analysis is an important specialization in econometrics which studies the nature, behavior and possible future predictions of (often) macroeconomic variables. This can be done either by modeling a single macroeconomic variable as a time-dependent scalar or a group of variables as a (finite dimensional) vector. Available data often shows periodic behavior for a range of frequencies together with a trend and therefor most analysis methods tend to decompose the time series X_t in a number of components. In this context, a filter is any mathematical technique intended to obtain ('filter') one or more of these components, though the aimed decomposition may vary for different filters. Traditionally, most authors [1] assume that X_t is of the form

$$X_t = T_t + C_t + S_t + \epsilon_t \tag{1}$$

or

$$X_t = T_t * C_t * S_t * \epsilon_t \tag{2}$$

where the second form can be transformed into the first by simply taking logarithms. In the following analysis we will therefore only consider the additive variant (1). Each term is assumed to have certain generic properties. We state the properties below.

Firstly, T_t is called the trend component and consists of the 'steady, slow-varying' part of X_t. The trend is not assumed to be periodic within the time frame, i.e. the period of T_t, if it exists, is (much) larger then T (the number of data points).
The second component, C_t, is called the Business cycle component and in this model is assumed to be composed of periodic functions with period in a range from 3 to 5 years. The existence of business cycles for various macroeconomic variables is hardly disputed between economists, though both the mechanisms causing them and the stationarity of their periodicities is.
The third component, S_t is called the seasonal component and behaves very much like C_t. It too is assumed to be composed of periodic functions, only this time with period in a range from 0.5 to 1.5 years. Unlike for the case of the Business cycle, the stationarity of the periods of the seasonal periodicities are not disputed.
The last term, ϵ_t, is the 'error term' and consist of simply $X_t - C_t - S_t$. It is

not estimated but is assumed to consists of high frequencies having (fairly) low amplitude.

There are numerous filtering methods known, both in and outside the field of econometrics, each with its own merits and drawbacks [2]. In this thesis we will examine the most popular and some less frequently used filters. We will also analyze a new method to decompose X_t in an interesting and surprisingly meaningful way. In the remainder of this chapter we will make some remarks on some choices we need to make and give a quick categorization of the different filters we will discuss.

In choosing how to model our macroeconomic variable of interest we have several options. First, we can model the variable either as a function of continuous or discrete time. It seems natural to choose the latter, since the used data (unless modified the input for a filter) will be discrete unless some interpolation has been applied. Mathematical analysis however, is often easier on functions $f : \mathbb{R} \to \mathbb{R}$ and some filtering methods indeed do assume the input variable is of this form: $X_t = f(t)$, $t \in \mathbb{R}$. Note here the difference between input variable and observed variable, the former is an interpolation of the latter. Infiniteness can be troublesome since integration in often required, but simply requiring $f \in l^2(\mathbb{R})$ suffices in most of these cases. Secondly, we can either model the data as being finite or infinite. Like discreteness, boundedness of the domain of the time series (the 'timespan') is a necessary attribute of raw data though modeling the variable as a function on an unbounded domain can have mathematical advantages. However, for the continuous bounded case $X_t : [a, b] \to \mathbb{R}$ the requirement $f \in l^2([a, b])$ becomes trivial since, without loss of generality, the input data X_t can be required to be continuous. In fact, for any $n \in \mathbb{N}$ we can require X_t to be nth times continuously differentiable, since X_t can be obtained from the finite observed data Y_t by means of a polynomial spline interpolation of order $n + 2$. In this thesis we will encounter all four of these models:

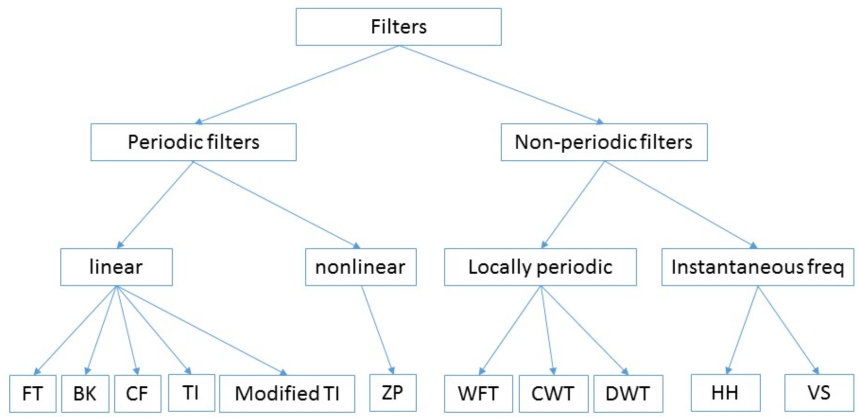
$$X_t = f : \mathbb{Z} \to \mathbb{R} \tag{3}$$
$$X_t = f : \{a, a+1, \cdots, b-1, b\} \to \mathbb{R} \tag{4}$$
$$X_t = f : \mathbb{R} \to \mathbb{R} \tag{5}$$
$$X_t = f : [a, b] \to \mathbb{R} \tag{6}$$
$$\tag{7}$$

And we have loosely argued that these are interchangeable. Without giving full definitions and details we now briefly give a categorization of filters suitable for this thesis.

Prior to our categorization we note that most filters assume the input data has been 'detrended', i.e. has had the trend component already filtered out. Despite this suggestive requirement, for us it is not useful classifying filters based on whether they extract the trend component or not. We will encounter detrending filters later on and, for filters demanding detrended data, we will assume the detrending has been performed. The classification made is based on whether the candidate filter assumes the data is well modeled as a realization of a periodic generating process. In essence this means that the variable can approximately be written as a sum of a small number of sinusoids plus a trend. Slightly abusing terminology, we call this periodic data as opposed to non-periodic data. This is a rather loose criterion but this is typical in time series analysis. For example, there is no clear-cut procedure to determining whether data has a trend (i.e. has to be detrended).

Filters that do require periodic data (abbreviation: periodic filters) will be discussed in chapter two, together with some detrending filters (not necessarily periodic filters themselves). We will review the Fourier Transform (FT) and show how problems such as leakage and the $\omega = 0$-problem arise from this Transform. we will show how these problems are tackled by the Baxter-King (BK) filter, the Christiane-Fitzgerald (CK) filter and the Zero-phase (ZP) filter and how a special class of filters, the linear time-invariant (TI) filters, arise as an interesting modification. It is in this class we find the most widely used detrending filters.

For non-periodic data we need to use filters that take in to account the time-varying nature of the frequency/frequencies. We have divided this category into two subcategories. The first simply assumes that the data is *locally* periodic, i.e. for an interval $[t-\epsilon, t+\epsilon]$ around time t (ϵ not necessarily small) the data is assumed to be periodic and in essence a periodic data analysis is performed on this small interval. This category contains filters like the *Windowed Fourier Transform* (WFT), the *Continuous Wavelet Transform* (CWT) and the *Discrete- Wavelet Transform* (DWT) and will be the content of chapter 3.
The second subcategory relies on the mathematical concept of instantaneous frequency, and consists of the relatively new Hilbert-Huang (HH) filter and a sort of proto-filter described by Vatchev and Sharpley (VS). Instantaneous frequency and the Hilbert-Huang filter will be discussed in chapter 4 and the Vatchev-Sharpley decomposition will be presented in chapter 5. Figure (...) displays the classification in a diagram



Filters
Periodic filters
Non-periodic filters
linear
nonlinear
Locally periodic
Instantaneous freq
FT
BK
CF
TI
Modified TI
ZP
WFT
CWT
DWT
HH
VS

Chapter 1

The Fourier Transform

The best known method to extract components having frequencies lying in a specific bandwidth is by using the Fourier transform. The Fourier transform (FT) maps a signal $x(t)$ to complex-valued function with frequency as independent variable. The name FT (with additional acronyms for different variants) is used both for the operator and for the image $X(\omega)$ obtained by inserting $x(t)$ in the FT (denoting time-dependent functions by small letters and their FT with capital ones is common shorthand). In this thesis we will discuss four versions of the FT who differ in the domain of the to-be-transformed functions and we start by giving a summary of these four versions. The first an simplest will be a transformation on function $x : \{0, 1, \cdots N-1\} \to \mathbb{R}$ called the Discrete Fourier Transform (DFT) together with its inverse, the Inverse Discrete Fourier Transform (IDFT). The second version is the natural extension of the DFT when we let $\{0, 1, \cdots, N-1\} \to \mathbb{Z}$. It is the FT on functions $x \in l^2(\mathbb{Z})$, i.e. $x : \mathbb{Z} \to \mathbb{R}$ s.t. $\sum_{t\in\mathbb{Z}} |x(t)|^2 < \infty$ and is called the Discrete Time Fourier Transform (DTFT) with inverse IDTFT. The third version of the FT will be defined on functions $L^2[0, T]$, i.e. the set of squarely integrable functions on the interval $[0, T]$. In literature this version it is often known as Fourier Series, but we call it the Bounded Interval Fourier Transform (BIFT) and the Inverse Bounded Interval Fourier Transform (IBIFT) for its inverse. The forth and last version will be the natural extension of the BIFT where the domain $[0, T] \to \mathbb{R}$. This version is called the Continuous Fourier Transform (CFT) with inverse ICFT. Both extensions will change a bounded domain to the doubly-unbounded domains $\mathbb{Z}$ and $\mathbb{R}$ respectively. The used derivations will start with alternative version of the DFT and BIFT where the domain has been made symmetrical ($\{0, 1, \cdots, N-1\}$ will be changed to $\{-M+1, \cdots, M-1\}$ for the DFT and $[0, T]$ is changed to $[-T/2, T/2]$ for the BIFT). We will show that these changes in domain do not essentially change the DFT and BIFT: only an expected phase shift occurs. It may seem a bit cumbersome to first define the DFT and BIFT on domains starting at zero, and then giving alternative though equivalent definitions on symmetrical domains. Some authors do prefer the symmetrical domains, but in this thesis domains of signals starting at zero and growing as more data points accumulate in time seems more natural. Furthermore, the adjustments are quite trivial and only needed once (for the derivation of the DTFT and the CFT). Statements made about the FT are understood either to be valid for all given versions or refer to the version the context implies. For completeness we remind that the Fourier transform (FT) is not necessarily defined for real-valued functions only: complex functions are allowed too, though in this thesis we will only consider functions who are real-valued in their time representation. For real-valued functions the FT will, in general, be complex-valued and expressing them as complex exponentials gives a nice interpretation of frequency-dependent amplitude and phase functions.

1.1 The Discrete Fourier Transform (DFT)

We start with the DFT, see for instance [3] transforming $x \in l^2[\{0, \cdots, N-1\}]$, i.e. the set $\{x | x : \{0, \cdots N-1\} \to \mathbb{R}\}$. The set $\{e_i : \{0, \cdots N-1\} \to \mathbb{R}, i = 0, \cdots, N-1\}$ defined by $e_i(t) = \delta_{it}$ (where $\delta_{it} = 1$ if $i = t$ and $\delta_{it} = 0$ otherwise) forms the standard basis for $l^2[\{0, \cdots, N-1\}]$ in time-space with point evaluation forming the coefficients for every function $x(t)$: $x(t) = \sum_{i=0}^{N-1} x(i) e_i(t)$. We could of course have chosen another basis and this is precisely what is done in the Fourier transform. The general goal of the FT is to find a decomposition in terms of sinusoids, and writing these as complex exponentials (giving a clearer analysis) suggests a suitable alternative base would be the set $\{\exp(i\omega t), \omega = 2\pi k/N, k = 0, 1, \cdots N-1\}$. We could have chosen to add a fixed phase in the exponent, but this will not essentially change the resulting theory so for simplicity it is set to zero. This basis of complex exponentials is called the *Fourier basis*. Working out inner products gives

$$\langle \exp(i\omega t), \exp(i\omega t)/\rangle = \sum_{t=0}^{N-1} \exp(i\omega t) \exp(-i\omega t) = N \tag{1.1}$$

and for $\omega_1 \neq \omega_2$

$$\langle \exp(i\omega_1 t), \exp(i\omega_2 t)\rangle = \sum_{t=0}^{N-1} \exp(i(\omega_1 - \omega_2)t) = 0 \tag{1.2}$$

Every element of our set is a non-zero vector (since its norm is non-zero) linearly independent of the others (since orthogonality holds). Since there are N of them, the expressions above show that $\{\exp(i\omega t), \omega = 2\pi k/N, k = 0, 1, \cdots N-1\}$ is indeed an orthogonal (though not normalized) basis. Starting with a signal $x(t)$ (i.e. expressed w.r.t. the time-basis) we obtain the coefficients $X(k)$ (or alternatively but equivalently $X(\omega)$), where k is related to the frequency ω via $\omega = 2\pi k/N$, of this signal with respect to the Fourier basis by computing the inner product. ω and k will be used interchangeably and sometimes we will write c_ω instead of $X(\omega)$ to highlight that it represents a coefficient w.r.t.

a frequency basis. We will try to write k where the Fourier frequencies are discrete, and ω when they are continuous. The Fourier transform $\mathcal{F}x$ is then simply the mapping $k \to X(k)$ so $\mathcal{F}x = (\mathcal{F}x)(k)$ with

$$(\mathcal{F}x)(k) := X(k) = \langle x(t), \exp(2\pi ikt/N)\rangle = \sum_{t=0}^{N-1} x(t)\exp(-2\pi ikt/N) \tag{1.3}$$

Since the DFT is simply a change in basis (so an invertible linear operator on a finite dimensional space) the Inverse Discrete Fourier Transform IDFT can trivially be found. Working out yields

$$x(t) = \mathcal{F}^{-1}X(k) := \frac{1}{N}\sum_{k=0}^{N-1} X(k)\exp(2\pi ikt/N) \tag{1.4}$$

where the r.h.s. of the equation above is understood to be restricted to the original domain of $x(t)$. We can check that this definition is indeed the inverse of the DFT by inserting the expression for $X(k)$ in the equation above. The last step is made by noting that, for $k, s, t \in \{0, 1, \cdots N-1\}$, $\exp(-2\pi ik(s-t)/N$ equals 1 for $s = t$ and is periodic with period N and average 0 for $s \neq t$ so $\sum_{k=0}^{N-1}\exp(-2\pi ik(s-t)/N) = N \cdot \delta_{st}$

$$\begin{aligned}\mathcal{F}^{-1}\mathcal{F}x &= \frac{1}{N}\sum_{k=0}^{N-1} X(k)\exp(2\pi ikt/N)\\ &= \frac{1}{N}\sum_{k=0}^{N-1}\Big(\sum_{s=0}^{N-1} x(s)\exp(-2\pi iks/N)\Big)\exp(2\pi ikt/N)\\ &= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{s=0}^{N-1} x(s)\exp(-2\pi ik(s-t)/N)\\ &= \frac{1}{N}\sum_{s=0}^{N-1} x(s)\sum_{k=0}^{N-1}\exp(-2\pi ik(s-t)/N) = \sum_{s=0}^{N-1} x(s)\delta_{st} = x(t)\end{aligned} \tag{1.5}$$

A few remarks on the DFT are worth making, some of which also hold (analogously or directly) for other versions discussed later. First note that $x_j(t) := \delta_{jt}$ ($j, t \in \{0, 1, \cdots N-1\}$) is the set of functions as 'localised' as possible in the time domain. The DFT for these functions reads

$$(\mathcal{F}x_j)(k) = \sum_{t=0}^{N-1} \delta_{jt}\exp(-2\pi ikt/N) = \exp(-2\pi ikj/N) \tag{1.6}$$

which has modulus 1 for all k. In other words, a function maximally localized is time-space is minimally localized in frequency-space. We will show that this property also holds for the other versions of the FT discussed and in the next chapter we will formalize this localization property.
Secondly, note that the function $\mathcal{F}^{-1}(\mathcal{F}x)$ is a finite sum of complex exponentials

and is hence defined for all $t \in \mathbb{Z}$. Note that for $t' = t+nN$, $t \in \{1, \cdots, N-1\}$, $n \in \mathbb{Z}$ the IDFT of $\mathcal{F}x$ gives

$$\begin{aligned}(\mathcal{F}^{-1}X(k))(t') &:= \frac{1}{N}\sum_{k=0}^{N-1} X(k)\exp(2\pi ikt'/N) \\ = \frac{1}{N}\sum_{k=0}^{N-1} X(k)\exp(2\pi ikt/N)&\exp(2\pi kn) = x(t)\end{aligned} \tag{1.7}$$

So, for any finite signal $x(t)$, performing the composition of the IDFT and the DFT will give a time-dependent signal with domain $\mathbb{Z}$ which coincides with $x(t)$ on $t = 0, 1, \cdots N-1$ and is elsewhere a periodic continuation on $x(t)$. Despite the formal difference we will use the identity $x(t) = \mathcal{F}^{-1}\mathcal{F}x(t)$ which from now on is understood to be valid on the domain of x. Similarly, $\mathcal{F}x$ is also defined for $k' \in \mathbb{Z}\backslash\{0, 1, \cdots, N-1\}$. For $k' = k+nN$, $k \in \{0, 1, \cdots, N-1\}$, $n \in \mathbb{Z}$ we find

$$\begin{aligned}X(k') &= \sum_{t=0}^{N-1} x(t)\exp(-2\pi i\frac{k+nN}{N}t) \\ = \sum_{t=0}^{N-1} x(t)\exp(-2\pi int)&\exp(-2\pi i\frac{kt}{N}) = X(k)\end{aligned} \tag{1.8}$$

So the periodic continuation is present both in the time and frequency domain. The periodic continuation in time is pretty straightforward to understand, but its counterpart frequency-space is a little harder to understand. A consequence of the periodicity in frequency-space for example, is that one can reconstruct $x(t)$ from the shifted frequency coefficients $c_{k'+nN}$. The math works out since periodicity holds), but intuitively it may seem strange. For $n > 0$ (i.e. frequencies higher than the Fourier frequencies) there is a plot that may prove illuminating. In figure 2.1 the blue dots represent the data points. As can seen in the figure, both the blue and the red sinusoids give an exact fit (the blue dots may in fact not be the data points but the time evaluation of a frequency component) where the red sinusoid has frequency three times higher than the blue one. In other words, the DFT cannot distinguish frequencies $k \in \{0, 1, \cdot N-1\}$ from $k + nN$ which is why they have the same Fourier coefficient. This effect is called aliasing and arises whenever a signal has frequency components higher than the Nyquist frequency . The highest Fourier frequency is called the Nyquist frequency ω_{Ny}, and any frequency component of a signal higher than N_{Ny} will be aliased with frequencies lying in the Fourier bandwidth. This also explains why we can reconstruct $x(t)$ from frequencies higher than the Fourier frequencies. The effect clearly depends on N or, if we consider frequencies to be relative to the time span, the sampling rate. This effect is called aliasing and depending on the signal and sampling rate may have a severe effect. The arithmetically simplest example occurs when a signal consists of harmonics of $\omega_0 = 0$. Consider for example the signal $x(t) = 1 + \cos(2\pi t) + \cos(6\pi t)$ sampled

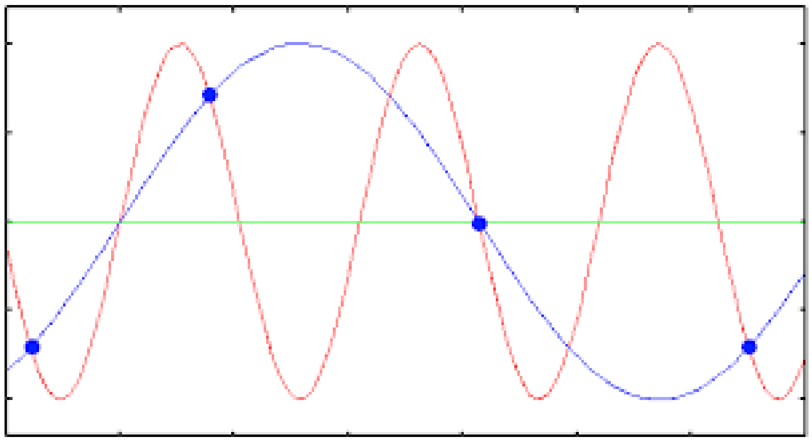


Figure 1.1: Aliasing: for discrete data multiple fitting frequencies can be found

at $t = 0, 1, \cdots N-1$. Since for each $t \in \{0, 1, \cdots, N-1\}$ we have $x_s(t) = 3$, the sampled signal is constant and the only nonzero $X(k)$ will be found at $k = 0$ with $X(0) = 3/\sqrt{N}$. Note that an increase in N only helps identifying the varying components of $x(t)$ if it increases the sampling rate f_s.
The periodic extension of the DFT also includes negative frequencies which have a nice interpretation themselves. For real sinusoids ($\cos(\omega t)$, $\sin(\omega t)$) negative frequencies have little meaning: changing the sign of ω gives $\cos(\omega t)$ and $-\sin(\omega t)$. For complex exponentials there is more to it. Since $\exp(i\omega t) = \cos(\omega t) + i\sin(\omega t)$ a 3d-plot (with axis being t, $\Re(\exp(i\omega t))$, $\Im(\exp(i\omega t))$) shows a spiral. Changing the sign of ω changes the orientation of the spiral. Since we can obtain the original real-valued signal by correlating the Fourier coefficients with these spirals (which includes a time-varying imaginary part whereas $X(\omega)$ is constant in time) its apparent that something is needed to cancel out the imaginary part. This is the role negative frequencies play. Writing z^* as the complex conjugate of z we see

$$\begin{aligned} X(-k) = \sum_{t=0}^{N} x(t)\exp(2\pi ikt/N) = \Big(\sum_{t=0}^{N} x(t)\exp(-2\pi ikt/N)\Big)^* = X(k)^* \\ \Rightarrow \Big(X(k)\exp(2\pi ikt/N) + X(-k)\exp(2\pi i - kt/N)\Big)^* \\ = \Big(X(k)\exp(2\pi ikt/N)\Big)^* + X(k)\exp(2\pi ikt/N) \in \mathbb{R} \end{aligned} \tag{1.9}$$

We stated that the negative frequencies were 'needed' to reconstruct the real signal $x(t)$, but in the IDFT only the (positive) Fourier frequencies are used. The final step in the argument is made by remarking that the negative frequency $-(k + nN)$ ($k \in \{0, 1, \cdots N-1\}$) is aliased with the frequency $N - k$, then trivially giving $X(N-k) = X(k)^*$ (lets call this the conjugate symmetry). This means that the negative frequencies are actually present in the set of Fourier

frequencies. As a help to understanding this, picture an observer measuring the evolution of the top point on the rim of a wheel with radius 1 at times $t = 0, 1, \cdots 9$, so $x(0) = 0, y(0) = 1$. Suppose the wheel rotates with positive (i.e. clockwise) frequency $\omega = 9\pi/5$ (the largest Fourier frequency), so the measured coordinates are $x(t) = \sin(\omega t)$, $y(t) = \cos(\omega t)$. At $t = 1$ we have $x(t) \approx -0.59$ and $y(1) \approx 0.81$ so it appears as though the wheel has rotated counter-clockwise.
The conjugate symmetry shows that, for real signals, we only 'need' the first half of the DFT (that is the values $X(k)$ for $k = 0, 1, \cdots (N-1)/2$ when N is odd and $k = 0, 1, \cdots (N-2)/2$ when N is even. Furthermore, using the conjugate symmetry and the aliasing equation we see that $X(N) = X(0) = X(N)^*$ so $X(0) \in \mathbb{R}$. For N even $X(N/2) = X(N - N/2)^* = X(N/2)^*$ so $X(N/2)$ is real. The conjugate symmetry may seem to contradict the invertible nature of the DFT. For those who like keeping track of (real) parameters however, it may come as a relief. We will give a quick heuristic analysis to show that 'no information is lost'. The DFT maps $\mathbb{R}^N \to \mathbb{C}^N$ linearly, so cant possibly be surjective. In fact, for N even we have shown that the DFT is fully defined by the set $X(0), X(1), \cdots, X(N/2)$ of size $N/2 + 1$ with $X(0), X(N/2) \in \mathbb{R}$. Writing $\mathbb{C} = \mathbb{R} \times i\mathbb{R}$ and ignoring the i we see that the DFT maps $\mathbb{R}^N \to \mathbb{R}^2 \times \mathbb{C}^{N/2+1-2} = \mathbb{R}^2 \times \mathbb{R}^{2(N/2-1)} = \mathbb{R}^N$. Similarly, for N odd the DTF is fully defined by the set $X(0), X(1), \cdots X((N-1)/2)$ where $X(0) \in \mathbb{R}$. For this case the DFT maps $\mathbb{R}^N \to \mathbb{R} \times \mathbb{C}^{(N-1)/2+1-1} = \mathbb{R} \times \mathbb{R}^{2(N-1)/2} = \mathbb{R}^N$. We now show how the $X(\omega)$ relates to the amplitude and phase of sinusoids with frequency ω. Rearranging terms in the IDFT and writing $X(k) =: X_k = |X_k| \cdot \exp(i \arg(X_k))$ to reduce the number of parenthesis we obtain for N odd

$$\begin{aligned}
x(t) &= \frac{1}{N}\Big(X_0 + \sum_{k=1}^{(N-1)/2} X_k \exp(2\pi i k t/N) + X_{N-k} \exp(2\pi i (N-k)/N)\Big) \\
&= \frac{1}{N}\Big(X_0 + \sum_{k=1}^{(N-1)/2} X_k \exp(2\pi i k t/N) + X_k^* \exp(2\pi i k/N)^*\Big) \\
&= \frac{1}{N}\Big(X_0 + \sum_{k=1}^{(N-1)/2} X_k \exp(2\pi i k t/N) + \big(X_k \exp(2\pi i k/N)\big)^*\Big) \\
&= \frac{1}{N}\Big(X_0 + \sum_{k=1}^{(N-1)/2} |X_k|\Big(\exp(i(2\pi \frac{kt}{N} + \arg(X_k))) + \exp(-i(2\pi \frac{kt}{N} + \arg(X_k)))\Big) \\
&= \frac{1}{N}\Big(X_0 + \sum_{k=1}^{(N-1)/2} 2|X_k| \cos(2\pi \frac{k}{N} t + \arg(X_k))\Big)
\end{aligned} \tag{1.10}$$

So we can relate

$$x(t) = \sum_{k=0}^{(N-1)/2} A_k \cos(2\pi \frac{k}{N} t + \phi_k) \tag{1.11}$$

to $X(k)$ via

$$\begin{aligned} A_0 &= X(0)/N \\ A_k &= 2|X(k)|/N \\ \phi_k = \arg(X(k)) &= \arctan \Im(X(k))/\Re(X(k)) \end{aligned} \tag{1.12}$$

In our definition we assumed the time domain to be $0, 1, \cdots N-1$. Though its always possible and common to number all data from 0 to $N-1$, this is no necessary requirement. An broader definition starts with data x of size N on the time domain $N_1, N_1+1, \cdots, N_1+N-1$ ($N_1 \in \mathbb{Z}$), i.e. $x : \{N_1, N_1+1, \cdots, N_1+N-1\} \to \mathbb{R}$. For these x the DFT is defined via

$$X(k) = (\mathcal{F}x)(k) := \sum_{t=N_1}^{N_1+N-1} x(t)\exp(-2\pi i \frac{kt}{N}) \tag{1.13}$$

Note that the two definitions coincide when $N_1 = 0$. What happens to the DFT when we 'translate' our finite data, i.e. renumber them, changing the starting point from N_1 to N_2? Writing $X^{(N_1)}(k), X^{(N_2)}(k)$ for the two DFT's differing in starting point, we derive their relation:

$$\begin{aligned} X^{(N_1)}(k) &= \sum_{t=N_1}^{N_1+N-1} x(t)\exp(-2\pi i \frac{kt}{N}) \\ = \exp(2\pi i \frac{k(N_2-N_1)}{N}) &\sum_{t=N_1}^{N_1+N-1} x(t)\exp(-2\pi i \frac{k(t-(N_1-N2)}{N}) \\ = \exp(2\pi i \frac{k(N_2-N_1)}{N}) &X^{(N_2)}(k) = \phi_{N_1\to N_2}(k))X^{(N_2)}(k) \end{aligned} \tag{1.14}$$

Where $\phi_{N_1\to N_2} := \exp(2\pi i \frac{k(N_2-N_1)}{N})$ is the extra factor picked up when changing the starting point from N_1 to N_2. Note that the two DFT's differ only by a phase modulating factor $\phi(k)$ which does not influence the amplitude of the DFT for any k. Intuitively we do not expect any change in starting point to affect the amplitude of the DFT, since the DFT essentially finds the 'amount' of sinusoids with specific frequencies present in the data. Sinusoids are not localized in space since they are periodic. We would however expect a phase shift to occur when translating data and since the phase is an angular expression (yielding larger translations in time-space for lower frequencies) we would expect the correction factor to be frequency-dependent. The obtained expression therefore should not come as a surprise and is in fact as clean as one could have hoped for. We conclude that changing the starting point does not essentially change the DFT which is why we will treat all finite data as being numbered $0, 1, \cdots N-1$. The translated version of the DFT will be used in the derivation of the inverse Fourier transform on functions $x \in l^2(\mathbb{Z})$.
Technically, the DFT is defined for finite data only. There is one exception to this limitation. For a periodic signal $x(t)$ with period N we can take the DFT

of single period and for this case the identity $x = \mathcal{F}^{-1}\mathcal{F}x$ holds for all $t \in \mathbb{Z}$. One may wonder if the choice of period affects the DFT or what happens if we accidentally take the DFT of a multiple of N. We combine both cases where we renumber our chosen finite subset (containing an integer multiple of periods) to run from 0 to $nN-1$. Let a be the starting point, n the number of periods in our subset and write $k' = nk + j$ ($k' \in \{0, 1, \cdots nN-1\}$, $k \in \{0, 1, \cdots N-1\}$, $j \in \{0, 1, \cdots n-1\}$) and $X_{a+nN}(k)$ for the DFT of our chosen subset (along with $X(k)$ for the DFT of the $\{x(t) | t = 0, 1, \cdots N-1\}$). We work out.

$$
\begin{aligned}
X_{a+nN}(k') &= \sum_{t=0}^{nN-1} x(t+a)\exp(-2\pi i\frac{k't}{nN}) \\
&= \exp(2\pi i\frac{k'a}{nN})\sum_{t=0}^{nN-1} x(t+a)\exp(-2\pi i\frac{k'(t+a)}{nN}) \\
&= \exp(2\pi i\frac{k'a}{nN})\sum_{t=a}^{a+nN-1} x(t)\exp(-2\pi i\frac{k't}{nN}) \\
&= \exp(2\pi i\frac{k'a}{nN})\Big(\sum_{t=a}^{N-1} x(t)\exp(-2\pi i\frac{k't}{nN}) + \sum_{t=N}^{nN-1} x(t)\exp(-2\pi i\frac{k't}{nN}) \\
&\qquad + \sum_{t=nN}^{a+nN-1} x(t)\exp(-2\pi i\frac{k't}{nN}) \\
&= \exp(2\pi i\frac{k'a}{nN})\Big(\sum_{t=a}^{N-1} x(t)\exp(-2\pi i\frac{k't}{nN}) + \sum_{t=N}^{nN-1} x(t)\exp(-2\pi i\frac{k't}{nN}) \\
&\qquad + \sum_{t=0}^{a-1} x(t)\exp(-2\pi i\frac{k't}{nN}) \\
&= \exp(2\pi i\frac{k'a}{nN})\sum_{t=0}^{nN-1} x(t)\exp(-2\pi i\frac{k't}{nN}) \\
&= \exp(2\pi i\frac{k'a}{nN})\sum_{l=0}^{n-1}\sum_{t=0}^{N-1} x(t)\exp(-2\pi i\frac{k'(t+lN)}{nN}) \\
&= \exp(2\pi i\frac{k'a}{nN})\sum_{l=0}^{n-1}\exp(-2\pi i\frac{k'}{n})^{l}\sum_{t=0}^{N-1} x(t)\exp(-2\pi i\frac{k't}{nN})
\end{aligned}
\tag{1.15}
$$

For $k' = nk$ (i.e. $j = 0$) this simplifies to

$$
\begin{aligned}
X_{a+nN}(kn) &= \exp(2\pi i\frac{ka}{N})\sum_{l=0}^{n-1}\exp(-2\pi ik)^{l}\sum_{t=0}^{N-1} x(t)\exp(-2\pi i\frac{kt}{N}) \\
&= \exp(2\pi i\frac{ka}{N})\cdot n\cdot X(k)
\end{aligned}
\tag{1.16}
$$

For the case $j \in \{1, 2, \cdots n-1\}$ we note that $\exp(-2\pi i \frac{(kn+j)t}{n}) = \exp(-2\pi i \frac{jt}{n}) \neq 1$. Since $\exp(-2\pi i \frac{jt}{n})^n = 1 = \exp(-2\pi i \frac{jt}{n})^0$ we see that

$$\exp(-2\pi i \frac{jt}{n}) \sum_{l=0}^{n-1} \exp(-2\pi i \frac{jt}{n})^l = \sum_{l=0}^{n-1} \exp(-2\pi i \frac{jt}{n})^l \tag{1.17}$$

Since $\exp(-2\pi i \frac{jt}{n}) \neq 1$ the sum above must equal zero. Hence

$$X_{a+nN}(kn+j) = \exp(2\pi i \frac{ka}{N}) \sum_{l=0}^{n-1} \exp(-2\pi i \frac{j}{t})^l \sum_{t=0}^{N-1} x(t) \exp(-2\pi i \frac{kt}{N}) = 0 \tag{1.18}$$

We see that the choice of starting point a only induces a phase-modulating factor, and taking the multiple nN as domain linearly translates the DFT values, keeping the rest at zero. The translation for $j = 0$ essentially arises because these frequencies are relative to the the size of the domain (how many periods 'fit' into the domain), which is the same multiple as that taken from the true period N. Intuitively, the other frequencies do not 'fit' into N and since the DFT contains the full information of the signal (it can reconstruct the original signal exactly) we would not expect the corresponding DFT value to be non-zero.

We make a final note on the factor $1/N$ present in IDFT but not in the DFT. Since the DFT and IDFT are linear operations, we can freely multiply the DFT with any $\lambda \neq 0$ and then multiply the IDFT with $1/\lambda$ keeping the identity $x = \mathcal{F}^{-1}\mathcal{F}x$. Therefor it is perfectly legitimate to put the $1/N$ factor in the DFT or put a $1/\sqrt{N}$ factor in both of them, the latter making both the DFT and the IDFT a unitary transformation. However $1/N$ is distributed among the DFT and IDFT, it will only change the modulus of $X(k)$ and hence its relation to the amplitude of the real-valued sinusoids. We have chosen to put the $1/N$ at the IDFT, since this will enable an easy derivation of the DTFT, discussed in the next section.
The $1/N$ obviously refers to the size of the domain of x, which is also the size of the domain on $X(k)$. For the other versions discussed next, a bit more subtlety is needed to interpreted 'size' factor present.

1.2 The Discrete Time Fourier Transform (DTFT)

The second variant of the FT discussed here will be a transformation on functions $x \in l^2(\mathbb{Z})$, the natural extension of the DFT called the DTFT (Discrete Time Fourier Transform), see for instance [4, Chapter 15],formally defined via

$$(\mathcal{F}x)(\omega) := \sum_{t \in \mathbb{Z}} x(t) \exp(-i\omega t) \tag{1.19}$$

Note that $\sum_{t\in\mathbb{Z}} |x(t)\exp(-i\omega t)|^2 \leq \sum_{t\in\mathbb{Z}} |x(t)|^2 \cdot |\exp(-i\omega t)|^2 = \sum_{t\in\mathbb{Z}} |x(t)|^2 < \infty$ so the operation is well defined for all $\omega \in \mathbb{R}$. The domain in frequency-space (and hence, the expression of the IDTFT) is slightly more complicated. As stated, the domain of the IDFT is $\{k = 0/N, 1/N, \cdots, (N-1)/N\}$. Letting $N \to \infty$ suggests that the IDTFT has the form of an integral instead of a sum with domain $\omega \in [0, 2\pi]$. We will derive the IDTFT as a limiting case of the IDFT.
For every $x \in l^2(\mathbb{Z})$ we can trivially define the truncated function $x_M \in l^2(\mathbb{Z})$ via

$$x_M(t) = \begin{cases} x(t) & \text{if } -M+1 \leq t \leq M-1 \\ 0 & \text{else} \end{cases} \tag{1.20}$$

the truncated function is (apart from two zero-valued tails) of finite length $N := 2M-1$ so the DFT is available for x_M. The DFT of x_M reads

$$X_M(k) = \sum_{t=-M+1}^{M-1} x_M(t)\exp(-2\pi i\frac{kt}{N}) = \sum_{t=-\infty}^{\infty} x_M(t)\exp(-2\pi i\frac{kt}{N}) \tag{1.21}$$

and the IDFT then gives

$$x_M(t) = \begin{cases} \frac{1}{N}\sum_{k=0}^{2M-2} X_M(k)\exp(2\pi i\frac{kt}{N}) & \text{if } -M+1 \leq t \leq M-1 \\ 0 & \text{else} \end{cases} \tag{1.22}$$

Since $\omega = 2\pi k/N$ $(k = 0, 1, \cdots, 2M-2)$ we have $\Delta\omega = \frac{2\pi}{N}$. Inserting this in the equation above and noting that $M \to \infty$ implies $(2M-2)/(2M-1) \to 1$ (so ω has domain $[0, 2\pi]$) is equivalent to $N \to \infty$ which implies $\Delta\omega \to d\omega$ we obtain

$$\begin{aligned} x(t) = \lim_{M\to\infty} x_M(t) &= \lim_{M\to\infty} \frac{1}{2\pi}\sum_{k=0}^{2M-2} X_M(k)\exp(i\omega t)\Delta\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi} X(\omega)\exp(i\omega t)d\omega \end{aligned} \tag{1.23}$$

Hence

$$\mathcal{F}^{-1}X := \frac{1}{2\pi}\int_0^{2\pi} X(\omega)\exp(i\omega t)d\omega \tag{1.24}$$

We briefly check this, noting that $\forall s, t \in \mathbb{Z}$ we have that $\exp(i\omega(s-t))$ has periodic real and imaginary parts completing an integer number of periods on the interval $\omega \in [0, 2\pi)$. Hence $\int_0^{2\pi} \exp(i\omega(s-t))d\omega = 0$.

$$\begin{aligned}\mathcal{F}^{-1}\mathcal{F}x &= \frac{1}{2\pi}\int_0^{2\pi}\Big(\sum_{s=-\infty}^{\infty} x(s)\exp(-i\omega s)\Big)\exp(i\omega t)d\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi}\Big(\sum_{s=-\infty}^{\infty} x(s)\exp(i\omega(s-t))\Big)d\omega \\ &= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} x(s)\int_0^{2\pi}\exp(i\omega(s-t))d\omega \\ &= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} x(s)2\pi\delta_{st} = x(t)\end{aligned} \tag{1.25}$$

A justification of the interchanging of the integral and sum, which is valid both for absolutely summable functions ($\sum_{t=-\infty}^{\infty}|x(t)| < \infty$) and for squarely integrable functions ($\sum_{t=-\infty}^{\infty}|x(t)|^2 < \infty$) can be found at ...

The set $x_j(t) = \delta_{jt}$ $(j, t \in \mathbb{Z})$ contains the functions maximally localized in time-space. Their FT reads:

$$(\mathcal{F}x_j)(\omega) = \sum_{t=-\infty}^{\infty}\delta_{jt}\exp(i\omega t) = \exp(i\omega j) \tag{1.26}$$

Implying that $X_j(\omega)$ is constant so minimally localized in frequency-space. Note the subtle difference between the DFT and DTFT: for the DFT time-localized functions are always either localized w.r.t. a bounded domain (limiting in a way the localization in time) or they are localized w.r.t. the single period of a periodic function (we showed that these two variants are effectively the same). Neither would be considered to be 'perfectly localized'. One can argue that localized functions in $l^2(\mathbb{Z})$ are more localized, since they are localized in an unbounded space. This is reflected in the form of the DTFT for these functions, which gives a uniform interval of Fourier amplitudes (ensuring in no way that the IDTFT is periodic). Unsurprisingly, the most extreme case will be given at the last version of the Ft, the CFT, which transforms $x(t) \in L^2(\mathbb{R})$.

Similar to the DFT the DTFT is defined for all $\omega \in \mathbb{R}$. For $\omega = \omega' + 2\pi n$, $\omega' \in [0, 2\pi)$, $n \in \mathbb{Z}$ we have

$$X(\omega) = \sum_{t\in\mathbb{Z}} x(t)\exp(-i(\omega' + 2\pi n)t) = \sum_{t\in\mathbb{Z}} x(t)\exp(-i\omega' t) \tag{1.27}$$

So the DTFT is periodic with period 2π. Note that this implies that aliasing may occur whenever there are frequency components present with frequencies higher than $2\pi f_s$. Conjugate symmetry also holds (for real-valued signals $x(t)$). Let $\omega' = -\omega$, $\omega \in [0, 2\pi)$ we have

$$\begin{aligned}X(2\pi - \omega) = X(\omega' + 2\pi) = X(\omega') &= \sum_{t\in\mathbb{Z}} x(t)\exp(-i\omega' t) \\ &= \sum_{t\in\mathbb{Z}} x(t)\exp(i\omega t) = X(\omega)^*\end{aligned} \tag{1.28}$$

which again implies $X(0), X(\pi) \in \mathbb{R}$ (even or oddness is obviously irrelevant). Note that, similar ot the DFT case, there is flexibility in the choice of interval in the definition of the IDTFT. Some textbooks, such as [4] prefer the symmetrical form

$$x(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) \exp(2\pi i \omega t) d\omega \tag{1.29}$$

Which is completely equivalent to our definition. We prefer our definition since in clearly implies that the largest frequency that can be 'found' (i.e. is not aliased with a lower positive frequency) is 2π (w.r.t. the sampling frequency f_s).
Technically, the DTFT is defined for functions $x \in l^2(\mathbb{Z})$

1.3 The Bounded Interval Fourier Transform

The third case we consider concerns the space $L^2[0,T]$. Though in literature this version is known as Fourier Series (see [5]), in this thesis we call the variant of the FT on this space the Bounded Interval Fourier Transform (BIFT). In a way, this case is the exact opposite to prior the case, the DTFT. The inner product in $L^2[0,T]$ is defined as

$$\langle f, g\rangle := \int_0^T f(t) g^*(t) dt \tag{1.30}$$

Consider the set $\{\exp(i\omega t)/T | \omega = 2\pi k/T, k \in \mathbb{Z}\}$. Working out inner products, noting that $\exp(2\pi i(k_1 - k_2)t/T)$ completes an integer number of periods on $[0,T]$ for $k_1, k_2 \in \mathbb{Z}$, gives

$$\begin{aligned}
\langle \exp(i\omega t), \exp(i\omega t)\rangle &= \frac{1}{T}\int_0^T \exp(i\omega t)\exp(-i\omega t) dt = 1 \\
\langle \exp(i\omega_1 t), \exp(i\omega_2 t)\rangle &= \frac{1}{T}\int_0^T \exp(i\omega_1 t)\exp(-i\omega_2 t) dt \\
&= \frac{1}{T}\int_0^T \exp(2\pi i(k_1 - k_2)t/T) dt = 0
\end{aligned} \tag{1.31}$$

So the set is orthonormal. One can use Sturm-Liouville theory (noting that the set equals the set of solutions to $\frac{d^2y}{dt^2} = -\omega^2 y$, $y(0) = y(T) = 0$) to show that it lies dense in $L^2[0,T]$. See [6, Theorem 8.27] for a full proof.
Since the Fourier basis is countable and $\langle x, \exp(2\pi ikt/T)\rangle$ is well defined the following definition of the BIFT and IBIFT is rather straightforward.

$$\begin{aligned}
X(k) = (\mathcal{F}x)(k) &:= \frac{1}{T}\int_0^T x(t)\exp(-2\pi ikt/T) dt \\
x(t) = (\mathcal{F}^{-1}X)(t) &:= \sum_{k\in\mathbb{Z}} X(k)\exp(2\pi ikt/T)
\end{aligned} \tag{1.32}$$

where the convergence of the series in the second equation is in the sense of the $L^2[0,T]$ norm. The equality is therefore $x(t) = (\mathcal{F}^{-1}X)(t)$ is understood to valid a.e. and in principle only for the interval $[0,T]$. For a full derivation, including the convergence of the series in the IBIFT, see [5, Theorem 3.1.3]
We now show that in contrast to the two prior versions, $X(k)$ in general will not be periodic. Let $N, n \in \mathbb{Z}$, $k' \in \{0, 1, \cdots N-1\}$. For $k = k' + nN$ we have

$$\begin{aligned} X(k) &= \frac{1}{T}\int_0^T x(t)\exp(-2\pi i(k'+nN)t/T)dt \\ &= \frac{1}{T}\int_0^T x(t)\exp(-2\pi i k' t/T)\exp(-2\pi i n N t/T)dt \end{aligned} \tag{1.33}$$

where the second complex does not vanish since t is not an integer. This implies that aliasing does not occur, which makes sense because $x(t)$ is not sampled but known for all $t \in [0,T]$.
The domain of $(\mathcal{F}^{-1}X)(t)$ can be extended to $\mathbb{R}$ without any technical problematics. At the endpoints 0 and T the relation $x(t) = (\mathcal{F}^{-1}X)(t)$ may not hold, but for $t' \in [0,T)$, $n \in \mathbb{Z}$ we always have

$$\begin{aligned} (\mathcal{F}^{-1}X)(t'+nT) &= \sum_{k\in\mathbb{Z}} X(k)\exp(2\pi i k(t'+nT)/T) \\ &= \sum_{k\in\mathbb{Z}} X(k)\exp(2\pi i k t'/T)\exp(2\pi i n) = (\mathcal{F}^{-1}X)(t') \end{aligned} \tag{1.34}$$

So $\mathcal{F}^{-1}X$ gives a periodic extension of $x(t)$ a.e., similar to what we saw discussing the DFT. When taking the FT of periodic functions on $\mathbb{R}$ it therefor suffices to compute the BIFT of a single period.
Like we did for the previous versions, we calculate the BIFT for a function localized in time. Unfortunately, the simple function $x_s(t) = \delta_{ts}$ (i.e. $x(t) = 1$ for $t = s \in [0,T]$ and 0 otherwise) will not suffice since it is equal to the zero function a.e.. Instead, we will look at the family of functions $\{x_\epsilon^{(s)}(t) | t \in [0,T], s \in (0,T), 0 < \epsilon < s\}$ defined via

$$x_\epsilon^{(s)}(t) = \begin{cases} \frac{1}{2\epsilon} & \text{if } s-\epsilon \le t \le s+\epsilon \\ 0 & \text{otherwise} \end{cases} \tag{1.35}$$

Note that for any s the function $x_\epsilon^{(s)}(t)$ becomes more and more localized in time when we let $\epsilon \to 0$ while $\int_0^T x_\epsilon^{(s)}(t)dt = 1$. Taking the BIFT gives

$$
\begin{aligned}
X(k) &= \frac{1}{T}\int_0^T x_\epsilon^{(s)}(t)\exp(-2\pi ikt/T)dt \\
&= \frac{1}{2\epsilon T}\int_{s-\epsilon}^{s+\epsilon}\exp(-2\pi ikt/T)dt \\
&= -\frac{1}{4\pi\epsilon ik}(\exp(-2\pi ik(s+\epsilon)/T) - \exp(-2\pi ik(s-\epsilon)/T)) \\
&= \frac{\exp(-2\pi iks/T)}{4\pi\epsilon ik}(\exp(2\pi ik\epsilon/T) - \exp(2\pi ik\epsilon/T)) \\
&= \frac{\exp(-2\pi iks/T)}{2\pi\epsilon k}\sin(2\pi k\epsilon/T) \approx \frac{\exp(-2\pi iks/T)}{T}
\end{aligned}
\tag{1.36}
$$

For small ϵ. As $\epsilon \to 0$ we have $|X(k)| \to 1/T$. A similar result is obtained for any family of function becoming increasingly narrow: functions that are well localized in time are minimally localized in frequency.

Similar to what we did for the DFT, we now show what happens when we take a different starting point and/or inadvertently integrate over multiple of the period T.
Suppose we start with a periodic $x : \mathbb{R} \to \mathbb{R}$ with period T and we take the BIFT of $x(t)$ on the domain $[a, a + nT]$. Writing $X_{a+nT}(k)$ for the BIFT taken

over this interval and $k' = nk + j$ we obtain

$$\begin{aligned}
X_{a+nT}(k') = \frac{1}{nT}\int_a^{a+nT} x(t)\exp(-2\pi i\frac{(nk+j)t}{nT})dt \\
= \frac{1}{nT}\int_a^{T} x(t)\exp(-2\pi i\frac{(nk+j)t}{nT})dt \\
+\frac{1}{nT}\sum_{l=1}^{n-1}\int_0^{T} x(t)\exp(-2\pi i\frac{(nk+j)(t+lT)}{nT})dt \\
+\frac{1}{nT}\int_T^{a+T} x(t)\exp(-2\pi i\frac{(nk+j)(t+(n-1)T)}{nT})dt \\
= \frac{1}{nT}\int_a^{T} x(t)\exp(-2\pi i\frac{(nk+j)t}{nT})dt \\
+\frac{1}{nT}\sum_{l=1}^{n-1}\int_0^{T} x(t)\exp(-2\pi i\frac{(nk+j)(t+lT)}{nT})dt \\
+\frac{1}{nT}\int_0^{a} x(t)\exp(-2\pi i\frac{(nk+j)}{nT})dt \\
= \frac{1}{nT}\sum_{l=0}^{n-1}\int_0^{T} x(t)\exp(-2\pi i\frac{(nk+j)(t+lT)}{nT})dt \\
= \frac{1}{nT}\int_0^{T} x(t)\sum_{l=0}^{n-1}\exp(-2\pi i\frac{nkt+nklT+jt+jlT}{nT})dt \\
= \frac{1}{nT}\int_0^{T} x(t)\exp(-2\pi i\frac{nk+j}{nT}t)\sum_{l=0}^{n-1}\Big(\exp(-2\pi i\frac{nk+j}{n})\Big)^l dt \\
= \frac{1}{nT}\int_0^{T} x(t)\exp(-2\pi i\frac{nk+j}{nT}t)\sum_{l=0}^{n-1}\Big(\exp(-2\pi i\frac{j}{n})\Big)^l dt
\end{aligned} \tag{1.37}$$

The summation of exponents above has our interest. Note that for $j = 0$ it reduces to $\exp(-2\pi ik) = 1$ so for $k' = nk$ it holds:

$$X_{a+nT}(nk) = \frac{1}{T}\int_0^T x(t)\exp(-2\pi i\frac{kt}{T})dt = X(k) \tag{1.38}$$

For $j \neq 0$ (i.e. $j = 1, 2, \cdots n-1$) we define m to be the least common multiple (LCM) of j and n. Note that $\exp(-2\pi ij/n)^{m/j} = 0$ and $\{\exp(-2\pi ij/n)^l | l = 0, 1, \cdots m/j - 1\}$ are all distinct. Combining this yields

$$\begin{aligned}
\sum_{l=0}^{n-1}\Big(\exp(-2\pi i\frac{j}{n})\Big)^l &= \frac{m}{j}\sum_{l=0}^{m/j-1}\Big(\exp(-2\pi i\frac{j}{n})\Big)^l \\
&= \frac{m}{j}\frac{1-\exp(-2\pi i\frac{j}{n})^{m/j}}{1-\exp(-2\pi i\frac{j}{n})} = 0
\end{aligned} \tag{1.39}$$

so for $k' = nk + j$, $j \in \{1, 2, \cdots, n-1\}$ we have

$$X_{a+nT}(nk + j) = 0 \tag{1.40}$$

Summarizing, we conclude that the choice of starting point a does not affect the FT and that including multiple periods only induces a stretching of the FT. This stretching was to be expected, since the discretized frequencies correspond to sinusoids 'fitting' into nT (i.e. completing an integer number of cycles), so a frequency k fitting in T will correspond to a frequency nk fitting in nT. Since the IBIFT constructs the periodic extension of $x(t)$ (which gives the same function regardless of n a.e.) we could have expected the proven result $X_{a+nT}(nk+j) = 0$ for $j \neq 0$.
Similar to what we did with the DFT, we will give an alternative definition for the $BIFT$ by changing the starting point. As we did for the DFT we will start by showing that this alternative definition is equivalent up to a phase-modulating factor to the BIFT. The alternative definition will be used in the derivation of the CFT. Let $a \in \mathbb{R}$, $x_a : [a, a+T] \to \mathbb{R}$.

$$(\mathcal{F}x)(k) := \frac{1}{T}\int_a^{a+T} x(t)\exp(-2\pi kt/T)dt \tag{1.41}$$

Defining $x_b : [b, b+T] \to \mathbb{R}$ as the translated x_a via $x_b(t) := x_a(t + a - b)$ we find

$$\begin{aligned}(\mathcal{F}x_b)(k) &= \frac{1}{T}\int_b^{b+T} x(t)\exp(-2\pi ikt/T)dt \\ &= \frac{1}{T}\int_a^{a+T} x(t+a-b)\exp(-2\pi ik(t+a-b)/T)dt \\ &= \exp(-2\pi ik(a-b)/T)\frac{1}{T}\int_a^{a+T} x_a(t)\exp(-2\pi ik(t+a-b)/T)dt \\ &= \phi_{a\to b}(k)\cdot\mathcal{F}x_a\end{aligned} \tag{1.42}$$

Where $\phi_{a\to b}(k) := \exp(-2\pi ik(a-b)/T)$ is a frequency dependent phase modulator.

1.4 The Continuous Fourier Transfrom (CFT)

The last version of the Fourier transform presented here concerns the function space $L^2(\mathbb{R})$ which we call the Continuous Fourier Transform (CFT), also described in [5]. We derive it using the alternative definition for the BIFT setting

$a = -T/2$ and letting $T \to \infty$. Since $\omega = 2\pi k/T$ we have $\Delta\omega = 2\pi/T$

$$\begin{aligned}(\mathcal{F}x)(k) &= \int_{-T/2}^{T/2} x(t)\exp(-2\pi ikt/T)dt \\ (\mathcal{F}X)(t) &= \frac{1}{T}\sum_{k=-\infty}^{\infty} X(k)\exp(2\pi ikt/T) \\ &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} X(k)\exp(i\omega t)\Delta\omega\end{aligned} \tag{1.43}$$

Letting $T \to \infty$ we obtain

$$\begin{aligned}(\mathcal{F}x)(\omega) &= \int_{-\infty}^{\infty} x(t)\exp(-i\omega t)dt \\ (\mathcal{F}^{-1}X)(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\exp(i\omega t)d\omega\end{aligned} \tag{1.44}$$

As a worked out example of the CFT, consider the family of functions $\{x_\epsilon^{(s)}(t) : \epsilon > 0\}$ with $x_\epsilon^{(s)}(t)$ defined via

$$x_\epsilon^{(s)}(t) := \begin{cases} \frac{1}{2\epsilon} & \text{if } s-\epsilon \le t \le s+\epsilon \\ 0 & \text{otherwise}\end{cases} \tag{1.45}$$

The CFT of $x_\epsilon^{(s)}$ reads

$$\begin{aligned}\mathcal{F}x_\epsilon^{(s)} &= \int_{-\infty}^{\infty} x_\epsilon^{(s)}\exp(-i\omega t)dt = \frac{1}{2\epsilon}\int_{s-\epsilon}^{s+\epsilon}\exp(-i\omega t)dt \\ &= -\frac{1}{2i\omega}\Big(\exp(-i\omega(s+\epsilon)) - \exp(-i\omega(s-\epsilon))\Big) \\ &= \frac{\exp(-i\omega s)}{\omega}\sin(\omega\epsilon) \approx \exp(-i\omega s)\epsilon\end{aligned} \tag{1.46}$$

for $\epsilon \ll 1$. Note that as $\epsilon \to 0$ the function $x_\epsilon^{(s)}$ becomes more and more localized in space, whereas its FT (more precisely the amplitude of the FT) becomes more and more uniformly distributed over $\mathbb{R}$. Also note that the limit $\lim_{\epsilon\to 0} x_\epsilon^{(s)}$ is not defined since its infinite at $t = s$ so the FT is only defined for $\epsilon > 0$.
As another example of a family of functions becoming increasingly localized in time, consider the family of normalized Gaussians $\{y_\epsilon^{(s)}(t) : \epsilon > 0\}$ defined via

$$y_\epsilon^{(s)}(t) := \frac{1}{\sqrt{2\pi\epsilon}}\exp\left(-\frac{(t-s)^2}{2\epsilon}\right) \tag{1.47}$$

Writing $u = (t-s)/\sqrt{2\epsilon}$ and $z = u + i\omega\sqrt{\epsilon/2}$ the FT becomes

$$\begin{aligned}
\mathcal{F}y_\epsilon^{(s)} &= \frac{1}{\sqrt{2\pi\epsilon}}\int_{-\infty}^{\infty}\exp(-\frac{(t-s)^2}{2\epsilon})\exp(-i\omega t)dt \\
&= \frac{\exp(-i\omega s)}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp(-u^2)\exp(-i\sqrt{2\epsilon}\omega u)du \\
&= \frac{\exp(-i\omega s)}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp(-(u^2+i\sqrt{2\epsilon}\omega u))du \\
&= \frac{\exp(-i\omega s)}{\sqrt{\pi}}\exp(-\frac{\epsilon\omega^2}{2})\int_\gamma \exp(-(u+i\omega\sqrt{\epsilon/2})^2)dz \\
&= \frac{\exp(-i\omega s)}{\sqrt{\pi}}\exp(-\frac{\epsilon\omega^2}{2})\int_\gamma \exp(-z^2)dz
\end{aligned} \tag{1.48}$$

Where γ is the line in $\mathbb{C}$ defined by $\gamma := \{z = a + i\omega\sqrt{\epsilon/2} : a \in \mathbb{R}\}$. The integral above can be solved using contour integration and Cauchy's Residue Theorem, as can be found in [7]. Let γ_R be the positive (counter clockwise) oriented rectangle with the 4 corners being $\{-R, R, R+i\omega\sqrt{\epsilon/2}, -R+i\omega\sqrt{\epsilon/2}\}$. The complex function $\exp(-z^2)$ has no singularities in $\mathbb{C}$ so $\int_{\gamma_R}\exp(-z^2)dz = 0$.

$$\begin{aligned}
0 &= \lim_{R\to\infty}\int_{\gamma_R}\exp(-z^2)dz \\
&= \lim_{R\to\infty}\int_{-R}^{R}\exp(-v^2)dv + \lim_{R\to\infty}\int_0^{\omega\sqrt{\epsilon/2}}\exp(-(R+iv)^2)dv \\
&- \lim_{R\to\infty}\int_{-R}^{R}\exp(-(v+i\omega\sqrt{\epsilon/2})^2)dv - \lim_{R\to\infty}\int_0^{\omega\sqrt{\epsilon/2}}\exp(-(-R+iv)^2)dv
\end{aligned} \tag{1.49}$$

The first integral equals $\sqrt{\pi}$, the third is the integral to be calculated and the second and forth integral converge to zero uniformly:

$$\begin{aligned}
\lim_{R\to\infty}\Big|\int_0^{\omega\sqrt{\epsilon/2}}\exp(-(R+iv))\Big|dv &\le \lim_{R\to\infty}\int_0^{\omega\sqrt{\epsilon/2}}\exp(-R^2+v^2)dv = 0 \\
\lim_{R\to\infty}\Big|\int_0^{\omega\sqrt{\epsilon/2}}\exp(-(-R+iv))\Big|dv &\le \lim_{R\to\infty}\int_0^{\omega\sqrt{\epsilon/2}}\exp(-R^2+v^2)dv = 0
\end{aligned} \tag{1.50}$$

This implies that

$$\int_{-\infty}^{\infty}\exp(-(v+i\omega\sqrt{\frac{\epsilon}{2}})^2)dv = \sqrt{\pi} \tag{1.51}$$

Hence

$$\mathcal{F}y_\epsilon^{(s)} = \exp(-i\omega s)\exp(-\frac{\epsilon\omega^2}{2}) \tag{1.52}$$

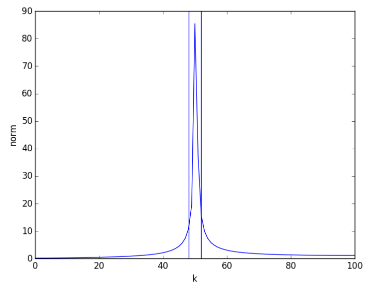
From the obtained expression we see that the FT of a Gaussian is again a (modulated for $s \neq 0$) Gaussian. Furthermore, the variance is inverted in the FT, so narrow functions in time (small ϵ) will have spread out FT.
We conclude this section by mentioning that a more general form of the FT, unifying the four versions discussed here, can be found in the theory of distributions. For a comprehensive overview, see [8]

1.5 Filtering with the FT

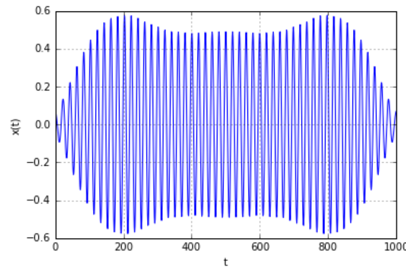
Having discussed the four versions of the FT, we are now ready to discuss filtering for each of these. Much of the analysis is the next two paragraphs can be found in [2] Often, the goal of filtering is to subtract the frequency components lying within a specific bandwidth $[\omega_1, \omega_2]$. The different versions lead to different problems and limitation, which will be the topic of this section.
We start with the DFT, so suppose we start with a signal $x : \{0, 1, \cdots, N-1\} \to \mathbb{R}$. In principle, filtering with the DFT is a trivial procedure. We simply take the DFT of x, multiply it with the indicator function $\mathbb{1}_{[\omega_1,\omega_2]}$ and finally take the IDFT. Mathematically this results in

$$x_{[\omega_1,\omega_2]} = \mathcal{F}^{-1}(X \cdot \mathbb{1}_{[\omega_1,\omega_2]}) = \frac{1}{N} \sum_{\frac{\omega_1}{2\pi} \leq k \leq \frac{\omega_2}{2\pi}} \Big(\sum_{t=0}^{N-1} x(t) \exp(-2\pi ikt) \Big) \exp(2\pi ikt) \tag{1.53}$$

The devil is in the details. For successive Fourier frequencies k_{11} and k_{12} the choice of ω_1 satisfying $2\pi k_{11} < \omega_1 < 2\pi k_{12}$ will not result in a different filter and analogously there is room picking $2\pi k_{21} < \omega_2 < 2\pi k_{22}$ without changing the result. In other words, we may want to filter a bandwidth having endpoints unequal to the Fourier frequencies but the DFT essentially forces us to pick the endpoints lying within the set of Fourier frequencies. Another problem arises when the data function x_s is sampled from a function x containing frequencies unequal to the Fourier frequencies. Before giving a concrete example, we point out that the periodic continuation of the IDFT of X_s will not match x on the discrete domain outside $\{0, 1, \cdots, N-1\}$ due to this (and maybe other non-Fourier frequency) component. As a concrete example, consider the function $x(t) = \sin(2\pi kt/N)$ with $k = 50.5$ and $N = 1000$, sampled at $t = 0, 1, \cdots, N-1$. In figure ... We have plotted the norm of $X(k)$, called the spectrogram of x, for the low values of k together with the 'cut off' lines $k_1 = 48$ and $k_2 = 52$. Though $x(t)$ mono-frequent, its frequency is not a Fourier frequency so instead of a sharp peak the DFT is more spread out around $k = 50$ and $k = 51$. The cut off shows some part of the signal is lost after filtering, which can clearly be seen in figure (b). This effect, where non-Fourier frequency components are spread out in the DFT is called leakage. Leakage effects decreases when N increases and when the spectrogram has lower peaks (the latter is used in the to-be-discussed Zhero-phase filter). Also note that non-Fourier frequencies near



(a) Spectrogram and cut off



(b) Original and filtered signal

Figure 1.2: Leakage

the borders k_1 and k_2 induce particularly strong leakage since they produce significant tails outside the interval $[k_1 \leq k \leq k_2]$.

1.6 Linear Translation Invariant Filters

Before we discuss problems concerning the DTFT, we first introduce a widely used class of filters working on the set of functions $\mathbb{R}^{\mathbb{Z}} := \{x : \mathbb{Z} \to \mathbb{R}\}$, which includes a 'moving' version of the DTFT. Modeling data as a function $x \in \mathbb{R}^{\mathbb{Z}}$ is a popular approach in econometric time series analysis, since a specific yet quite broad class of filters can be defined on these $x \in \mathbb{R}^{\mathbb{Z}}$ that gives rise to a simple but rich analysis. Economic data is of course finite, but does grow periodically (monthly, quarterly or yearly for most variables). Furthermore, the oldest data points are the least significant (in terms of extracting frequency components, predicting future variable values and other matters time series analysis is concerned with) so modeling the data as being doubly infinite instead of an ever growing finite set is defensible.

In general, a filter should be a mapping $F : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ and our class of filters will have the two additional properties of being linear and mappings of translated signals equal the translated mappings of the original signals. Using the acronym, we write LTIF for the set of linear translation invariant filters. To reduce the number of parentheses, we write x_t and Fx_t instead of $x(t)$ and $(Fx)(t)$ respectively.
Linearity of F is trivially defined by requiring $\forall F \in \text{LTIF}$, $x, y \in \mathbb{R}^{\mathbb{Z}}$, $\lambda, \mu \in \mathbb{R}$, $t \in \mathbb{Z}$: $F(\lambda x + \mu y)_t = (\lambda F x + \mu F y)_t := \lambda F x_t + \mu F y_t$ (note that the latter equality is not a requirement but a definition). To make the translation property mathematically precise, we define for $a \in \mathbb{Z}$ the lag operator $L^a : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ via $L^a x_t = x_{t-a}$ and ask of any linear filter F and $x \in \mathbb{R}^{\mathbb{Z}}$ to satisfy $FL^a x = L^a F x$. Or equivalently, using commutator notation: $\forall a \in \mathbb{Z} : [F, L^a] := FL^a - l^a F = 0$.

Summarizing, we obtain the following definition

$$\text{LTIF} := \{F : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}} | \forall x, y \in \mathbb{R}^{\mathbb{Z}},\ t, a \in \mathbb{Z} : F(x+y)_t = Fx_t + Fy_t, [F, L^a] = 0\} \tag{1.54}$$

Note that $\forall a \in \mathbb{Z} : L^a \in \text{LTIF}$. An example of a simple parametric family widely used in econometric time series analysis is the *moving average* filter of order n, notation: $F_n^{(ma)} \in \text{LTIF}$. We give two equivalent definitions, one pointwise and one as a sum of lag operators

$$\begin{aligned} F_n^{(ma)} x_t &:= \frac{1}{2n+1} \sum_{k=-n}^{n} x_{t+k} \\ F_n^{(ma)} &:= \frac{1}{2n+1} \sum_{k=-n}^{n} L^k \end{aligned} \tag{1.55}$$

The following lemma states that any $F \in \text{LTIF}$ is in fact a weighted sum of lag operators. We start by showing that Fx_t is a linear combination of $\{x_s | s \in \mathbb{Z}\}$. We can $x = \sum_{s \in \mathbb{Z}} x_s e^{(s)}$ where $e^{(s)} \in \mathbb{R}^{\mathbb{Z}}$ is defined via $e_t^{(s)} = \delta_{st} = 1$ for $s = t$ and zero elsewhere. Defining $\lambda^{(st)} := (Fe^{(s)})_t$ we see that $\lambda^{(st)}$ is independent of x and since F is translation invariant we can reduce the number of parameters in $\lambda^{(st)}$ since $\lambda^{(s(t-a))} = (Fe^{(s)})_{t-a} = (FL^a e^{(s)})_t = (Fe^{(s-a)})_t = \lambda^{((s-a)t)}$

Using the linear property and noting that $\lambda^{(st)} := (Fe^{(s)})_t$ is independent of x we see that

$$Fx_t = F\big(\sum_{s \in \mathbb{Z}} x_s e^{(s)}\big)_t = \sum_{s \in \mathbb{Z}} x_s \big(Fe^{(s)}\big)_t = \sum_{s \in \mathbb{Z}} \lambda^{(st)} x_s \tag{1.56}$$

For $F \in \text{LTIF}$ there are $a, b \in \mathbb{Z}$ and $\{f_k \in \mathbb{R} : k = a, a+1, \cdots, b\}$ s.t.

$$F = \sum_{k=a}^{b} f_k L^k \tag{1.57}$$

Every $x \in \mathbb{R}^{\mathbb{Z}}$ can be written as $x = \sum_{t \in \mathbb{Z}} x_t e^{(t)}$ with $e^{(t)} \in \mathbb{R}^{\mathbb{Z}}$ defined as $e_s^{(t)} = \delta_{st}$. We start by noting that, for any $s \in \mathbb{Z}$, the evaluation Fx_s is a function of x_t, $t \in \mathbb{Z}$. the filter F is fully determined by Fx_s, since for any $t \in \mathbb{Z}$ we then have that $Fx_t =$

The moving average filter takes the average around every point, which is why it used to estimate a trend, a feature we will analyze in a moment. We now prove the following claim. Every $F \in \text{LTIF}$ can be written as a finite weighted sum of lag operators. We first show that any linear $F : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ implies that Fx_t only depend on a finite set of data points. For any $x \in \mathbb{R}^{\mathbb{Z}}$ we have that $Fx_t = f(\{x_t\}_{t \in \mathbb{Z}})$. Since F is linear $f(\{x_t + y_t\}_{t \in \mathbb{Z}}) = F(x_t + y_t) = Fx_t + Fy_t = f(\{x_t\}_{t \in \mathbb{Z}}) + f(\{y_t\}_{t \in \mathbb{Z}})$

$\{x_t, t = a, a+1, \cdots, b\}$. Suppose it does not. Linearity implies F is of the form $Fx_t = \sum_{k=-\infty}^{\infty} a_k x_{t+k}$, where an infinite number of a_k's are nonzero. For $x \in \mathbb{R}^{\mathbb{Z}}$ defined via

$$\tag{1.58}$$

A filter F is called linear if $\forall x, y \in l$, $t \in \mathbb{Z}$ we have $(F[x+y])_t = (Fx+Fy)_t = Fx_t + Fy_t$ where the last equation is not a property of linearity but simply the definition of addition in l. It can be shown that a linear filter is translation-invariant if and only if it is of the form

$$Fx_t = \sum_{l=a}^{b} g_l x_{t-l} = \Big(\sum_{l=a}^{b} g_l L^l\Big) x_t =: G(L) x_t$$

One can wonder whether the summation above may extend to infinite sums. We will soon only consider functions on finite data so this curiosity is hardly relevant, but as an exercise we will show that this is not the case. Note that there are no restriction placed on the decay or even the boundedness of $x \in l$ so for any F of the form $\sum_{l=-\infty}^{\infty} g_l L^l$ we can choose $x \in l$ given by $x_t = 1/g_{-t}$. Then

$$y_0 = Fx_0 = \sum_{l=-\infty}^{\infty} g_l x_{-l} = \sum_{l=-\infty}^{\infty} 1 = \infty$$

So any linear, translation-invariant filter is a finite linear combination of lag operators. The filters in the examples above are all linear. As a realistic example of a non-linear translation-invariant filter (again to estimate the trend) consider

$$F_3 x_t = \text{median}(x_{t-1}, x_t, x_{t+1})$$

where median$(\cdot)$ is the median of its argument. To show that F_3 is not linear choose $x, y \in l$ as

$$x_t = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t \neq 0 \end{cases}$$

$$y_t = \begin{cases} 1 & \text{if } t = 1 \\ 0 & \text{if } t \neq 1 \end{cases}$$

we see that $F_3 x, F_3 y \equiv 0$ but

$$F_3(x+y)_t = (x+y)_t = \begin{cases} 1 & \text{if } t = 0 \\ 1 & \text{if } t = 1 \\ 0 & \text{elsewhere} \end{cases}$$

Unsurprisingly, the subset within the class of translation-invariant filters that is easiest to analyze is the set of *linear* translation-invariant filters. Their appealing properties are the subject of the next paragraph.

In the previous paragraph we already encountered the definition of $G(L)$ which uniquely characterizes any linear translation-invariant filter (all future filters will be, unless stated otherwise, translation-invariant so we will drop this description and simply call them (non-)linear filters or even just filters). Since we want to filter on certain frequencies, even when detrending, it is natural to consider the effect of a linear filter on a sequence $x_t = \cos(\omega t + \phi)$. Since the filters we consider are linear we can add any imaginary part to it without changing the results (well, the real part of the result anyway). adding $i\sin(\omega t+\phi)$ will give us a cleaner, more suggestive notation.

$$x_t^* := x_t + i\sin(\omega t+\phi) = \cos(\omega t+\phi) + i\sin(\omega t+\phi) = \exp(i(\omega t+\phi)) \quad (1.59)$$

$$y_t^* = G(L)x_t^* = \sum_{l=a}^{b} g_l x_{t-l}^* = \sum_{l=a}^{b} g_l \exp(i(\omega(t-l)+\phi)) \quad (1.60)$$

$$= \exp(i(\omega t+\phi))\sum_{l=a}^{b} g_l \exp(-i\omega l) = x_t^* G(\exp(-i\omega)) \quad (1.61)$$

Note that the $G(\exp(-i\omega))$ is time independent making the complex exponential an eigenfunction of any linear filter. The corresponding complex eigenvalue $G(\exp(-i\omega))$ has an amplitude and phase which determine y_t via:

$$|y_t^*| = |x_t^*| \cdot |G(\exp(-i\omega))| \quad (1.62)$$

$$\arg(y_t^*) = \arg(x_t^*) + \arg(G(\exp(-i\omega))) \quad (1.63)$$

Since $G(\exp(-i\omega))$ determines, for every frequency ω, how sinusoids of this frequency are amplified and shifted it is called the Frequency Transfer Function (notation: FTF). It is complex-valued and in exponential notation its components are called:

$$\mathrm{Gain}(\omega) = |G(\exp(-i\omega))| \quad (1.64)$$

$$\mathrm{Phase}(\omega) = \frac{1}{2\pi}\arg(G(\exp(-i\omega))) \quad (1.65)$$

where we have divided by 2π in the last equation to express the phase shift as a fraction of the wavelength. Note that if F is a linear we have that $1-F$ defined by $(1-F)x_t = x_t - Fx_t$ is also a linear filter. Its FTF (inlag form) is trivially given by $1-G(L)$. As a second note, consider two filters F_1, F_2 and $F := F_2 \circ F_1$. Then $\mathrm{Gain}_F(\omega) = \mathrm{Gain}_{F_1}(\omega)\cdot\mathrm{Gain}_{F_2}(\omega)$ and $\mathrm{Phase}_F(\omega) = \mathrm{Phase}_{F_1}(\omega) + \mathrm{Phase}_{F_2}(\omega)$.

We will now derive an easy condition for linear filters F to have $\mathrm{Phase}_F(\omega) = 0$.

$$\mathrm{Phase}_F(\omega) = 0 \Leftrightarrow i\sum_{l=a}^{b} g_l \sin(\omega l) = 0 \Leftrightarrow a = -b \text{ and } g_l = g_{-l}$$

So a linear filter is a zero-phase filter if and only if it is symmetrical.

Let us take a moment to appreciate the FTF. Each sequence x can be represented in frequency-space by the Fourier transform of x i.e. we can write x as a linear combination of complex exponentials with different frequencies and phases. The linearity of the filter then implies that each of these complex exponentials is multiplied by the FTF (whose argument only depends on the frequency, not on the phase) to obtain the filtered series. To analyze the effect

Figure 1.3

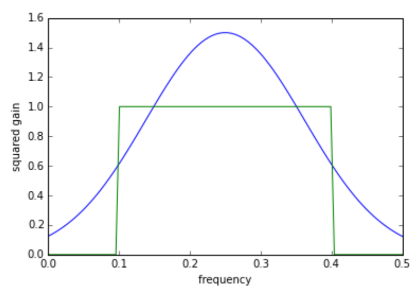
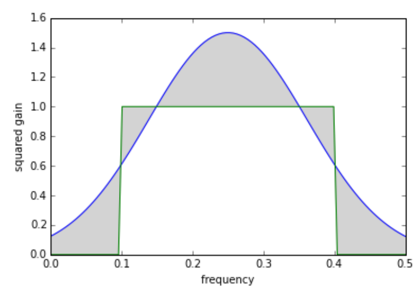


Figure 1.4



of the filter on any series it therefore suffices to consider the Gain and the Phase of the FTF. Since our goal will be to decompose the series based on a partitioning in frequency-space, it is easy to set criteria for linear filters in terms of the FTF. For the filters we will consider (used to make an estimation of the trend, business cycle or seasonal fluctuations), we ideally want the filtered series to retain frequencies within a certain band of the original without shifting them, i.e.

$$\text{Gain}^2_{\text{Ideal}}(\omega) = \begin{cases} 1 & \text{if } \omega \in [\omega_1, \omega_2] \\ 0 & \text{if } \omega \notin [\omega_1, \omega_2] \end{cases}$$

$$\text{Phase}_{\text{Ideal}}(\omega) = 0 \quad (\text{at least for } \omega \in [\omega_1, \omega_2])$$

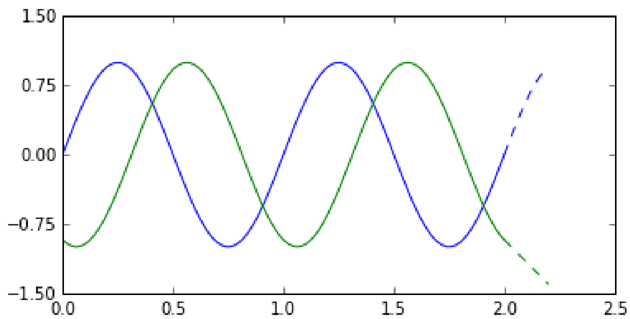
Where we have use the squared gain for reasons that will become apparent later. We will see that it is impossible to obtain a filter with an ideal FTF as stated above, so filters typically deviate from this ideal filter. In figures 2.3 and 2.4 the squared gain of a filter (blue) is compared to the squared gain of the ideal filter (green). The shaded area in figure 2.4 represents the total deviation from the ideal filter. Note that the highest frequency is $\frac{1}{2}$ as we have explained, but the domain is shown as a continuum. This is because the discretization is dependent on the sample size (approaching an continuum when $N \to \infty$).

Now let $c > 0$, G a linear filter (determined by its coefficients $g_a, g_{a+1}, \cdot, g_{b-1}, g_b$) and cG be the linear filter defined by $cg_a, cg_{a+1}, \cdot, cg_{b-1}, cg_b$. Note that

$$\text{Gain}_{cG}(\omega) = |G_{cG}(e^{-i\omega})| = \Big|\sum_{l=a}^{b} cg_l e^{-i\omega l}\Big| = |c|\Big|\sum_{l=a}^{b} g_l e^{-i\omega l}\Big| = |c|\text{Gain}_G(\omega) \tag{1.66}$$

So we can easily scale the gain of a filter by scaling the coefficients (note that the scaling is done in the output domain rather than in the time domain). Furthermore, this scaling does not influence the quality of the filter. Bearing this in mind we take another look at figure 2.4. In order to make a fair comparison between the squared gain of two filters we must demand that the squared gain is normalized. Only then can we say that the shaded area of figure 2.4 is an

Figure 1.5



appropriate measure of error compared to the ideal filter. Finally, we wish that a 'large error over a small interval' is worse than 'a small error over a large interval'. This is expressed by integrating over the squared difference. Let $c := \int_0^{\frac{1}{2}} |G(\omega)|^2 d\omega$. Then

$$d(G_{\text{Gain}}, I_{\text{Gain}}) := \int_0^{\frac{1}{2}} \left| \frac{|G(\omega)|^2}{c} - 1_{[\omega_1,\omega_2]} \right|^2 d\omega = \int_{\omega_1}^{\omega_2} \left| \frac{|G(\omega)|^2}{c} - 1 \right|^2 d\omega \quad (1.67)$$

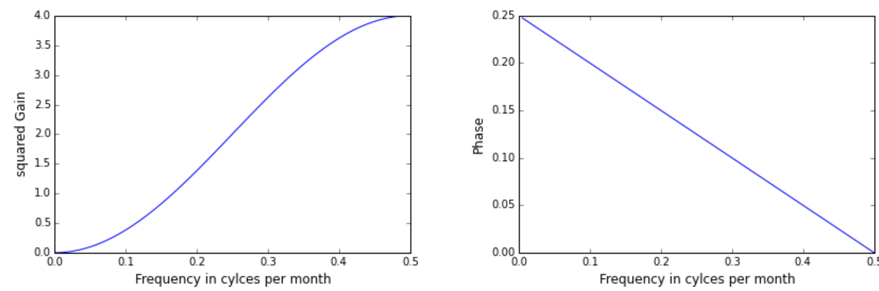
So how about the phase shift of a linear filter? Ideally we would have no phase shift for any frequency. To illustrate why figure 2.5 shows the effect of a filter on a pure sinusoid whose frequency induces a positive phase shift (and a positive gain) in the FTF. At the end of the sample we have extrapolated the filtered (green) and the original series to obtain a short term prediction. Clearly visible is that the prediction resulting from the filtered series greatly deviates from the realization of the original series. This is the mayor disadvantage of a non zero phase shift in the FTF.

It is clear that a filter with no phase shift for all frequencies would be ideal (from the phase shift point of view) but is the converse true?. Should we simply integrate then over the entire domain $[0, \frac{1}{2}]$? Or do phase shifts for different frequencies have different impact? Well, if for a frequency ω the gain is zero, then the phase shift has no influence. For a frequency with a low gain, the phase is correspondingly less important. Hence we introduce the weighted phase shift error.

$$d(G_{\text{Phase}}, I_{\text{Phase}}) = \int_0^{\frac{1}{2}} \text{Phase}(\omega)\text{Gain}(\omega) d\omega \quad (1.68)$$

Some Linear Filters The simplest and perhaps most often used linear translation-invariant filter (jsut called filter in this chapter) is the first order difference fiolter defined by

$$y_t = x_t - x_{t-1} = (1 - L)x_t \quad (1.69)$$



Where the on the far right we have written tyhe expression inb terms of the lag operator. We can easily compute the FTF:

$$G(\exp(-i\omega)) = 1 - \exp(-i\omega) = 1 - \cos(\omega) + i\sin(\omega) \tag{1.70}$$

$$\text{Gain}(\omega) = |G(\exp(-i\omega))| = \sqrt{2 - 2\cos(\omega)} \tag{1.71}$$

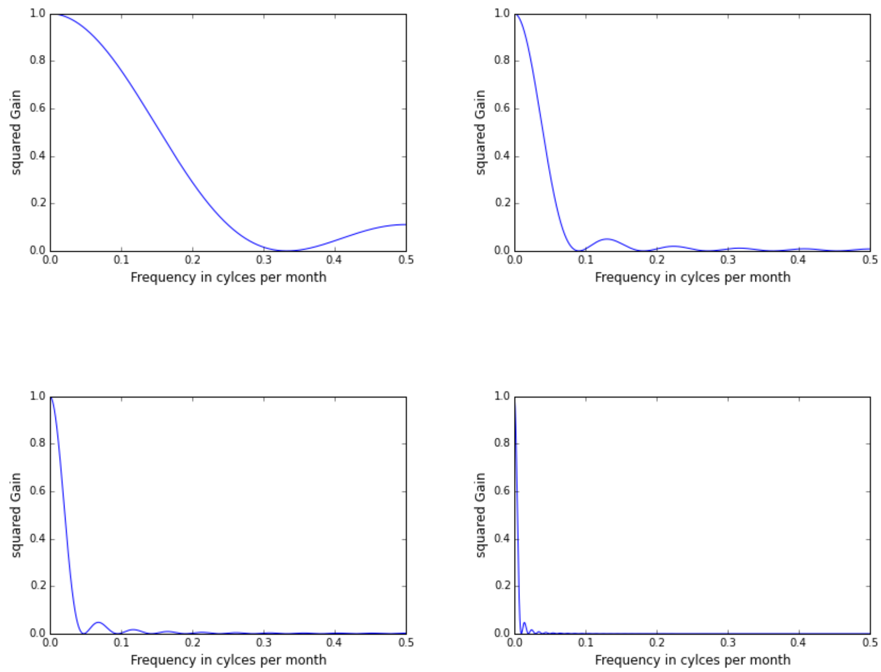
$$\text{Phase}(\omega) = \frac{\arg(G(\exp(-i\omega)))}{2\pi} = \frac{1}{2\pi}\arctan\left(\frac{\sin(\omega)}{1-\cos(\omega)}\right) \tag{1.72}$$

Below we have plotted the squared Gain (henceforth: Power Transfer Function or PTF) and the phase of this FTF. We can see that frequencies below 1/6 cycles per month i.e. having a period longer than half a year are being damped with this filter whereas higher frequencies are being amplified. For this reason the FOD filter is used to detrend data. It is easy to show that data containing a linear trend are perfectly detrended by the FOD filter, but for data having a strongly non-linear trend (for example strong exponential or quadratic) this is not the case. This can be solved by applying th FOD filter consecutive times until the data is stationary (i.e. has no trend). Often once or twice is sufficient.

Moving Averages Another family of filters popular for their simplicity if the family of moving averages. They are defined for each $n \leq \frac{1}{2}(\text{size(data)}-1)$

$$y_t = \frac{1}{2n+1}\sum_{l=-n}^{n} x_{t+l} \tag{1.73}$$

A moving average filter (MA) takes the average of a point and its closest n neighbors to each side. Intuitively we would expect rapid changing behavior (i.e. high frequencies) to cancel out or at least be suppressed. Very slow varying behavior is expected not to be affected by an MA since consecutive points are approximately equal and the average around x_t is approximately x_t. To visualize this we have plotted the PTF for $n = 1$, $n = 5$, $n = 10$ and $n = 50$. Note that all MA's are symmetric so no phase shift is induced.
It is clear that MA are used as detrending filters. Depending on the volatility



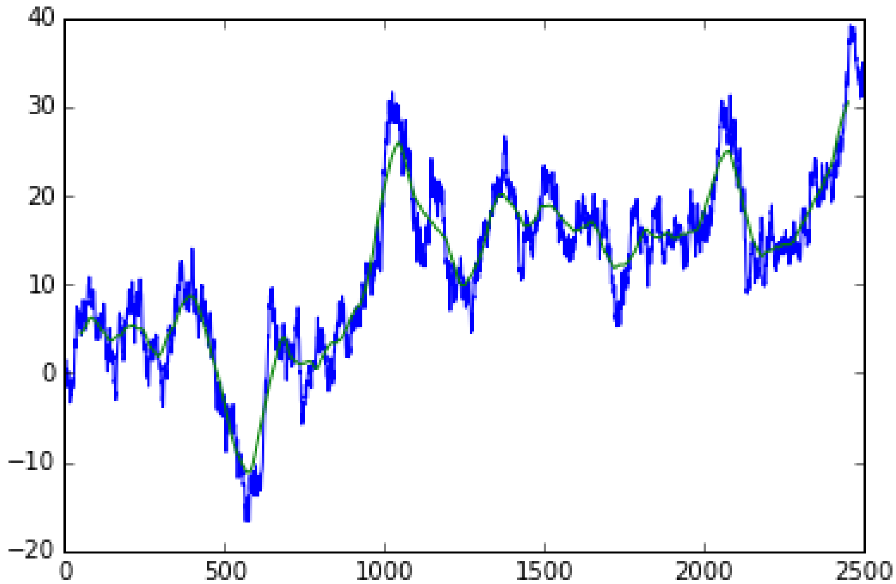
of the data a high value for n is needed to extract a sufficiently smooth trend component. This presents a mayor drawback of MA's: the last (and the first) n points have too few neighbors to obtain a value in the filtered series. There are ways to obtain a similar filtered value at these points (like decreasing the window size as the points get nearer to the end of the sample) but all these methods make the filter time-variant and may also introduce a phase shift.

The Hodrick-Prescott Filter Another widely used filter is the Hodrick-Prescott filter (hpf), as described in [9]. This filter fits a line through the data where an extra penalty λ is placed on the square of the second order derivative of this fit. For a finite sample of size T we find y_t by performing the following minimization:

$$\min_{y_0,\cdots,y_{T-1}} \sum_{t=1}^{T} (x_t - y_t)^2 + \lambda\big[(y_{t+1} - y_t) - (y_t - y_{t-1})\big]^2 \tag{1.74}$$

Solving this equation gives

$$y_t(I + \lambda A'A)^{-1} x_t \tag{1.75}$$



where A is the $(T-2)\times T$ matrix:

$$A = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix} \tag{1.76}$$

Approximating ideal filter Ideally we would have

$$1_{[\omega_1,\omega_2]} = G(\omega) = \sum_{l=a}^{b} g_l \exp(-i\omega l) \tag{1.77}$$

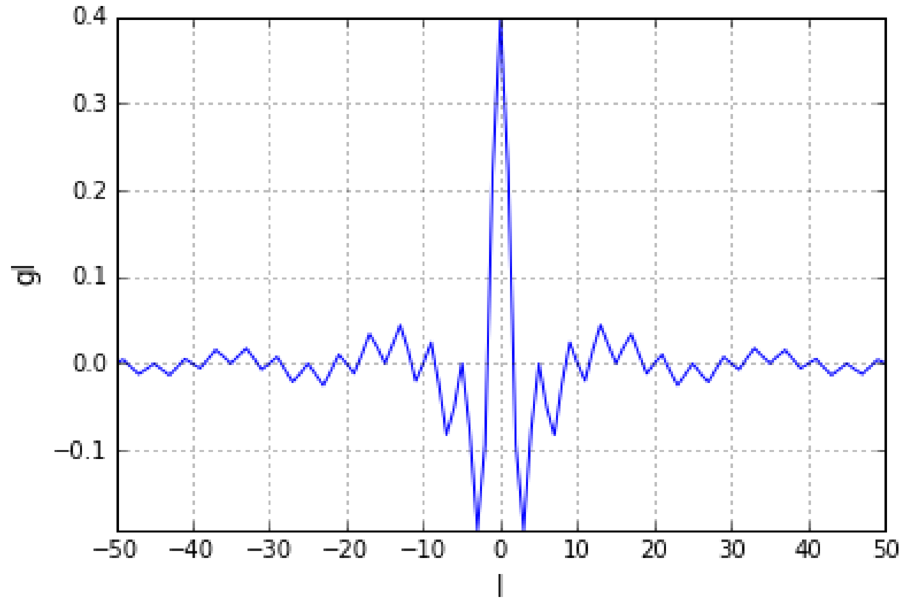
Since the right hand side of the equation above is the DCFT of the sequence$\{g_l\}$ we have

$$1_{[\omega_1,\omega_2]} = G(\omega) = \sum_{l=a}^{b} g_l \exp(-i\omega l) \tag{1.78}$$

$$\Leftrightarrow g_l = \frac{1}{2\pi}\int_0^{2\pi} G(\omega)\exp(-i\omega l)d\omega \quad l = a,\cdots,b \tag{1.79}$$

If we demand that out filter is symmetrical (i.e. zero phase) we obtain the extra equaltiy:

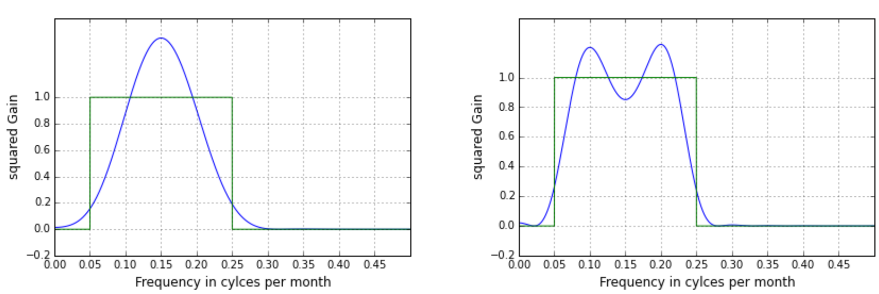
$$g_l = \frac{1}{2}(g_l + g_{-l}) \tag{1.80}$$

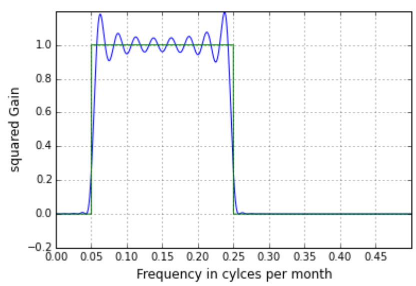


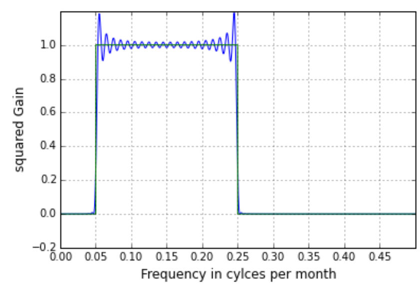
Combining these results and substitute $G(\omega) = 1_{[\omega_1,\omega_2]}$ we find

$$g_l = \frac{1}{\pi}\int_{\omega_1}^{\omega_2}\cos(\omega l)d\omega = \begin{cases} \frac{\sin(\omega_2 l)-\sin(\omega_1 l)}{\pi l} & \text{for } l \neq 0 \\ \frac{\omega_2-\omega_1}{\pi} & \text{if } l = 0 \end{cases} \tag{1.81}$$

Since the support of $\{g_l\}$ is not bounded and we only have a finite data sample we always need to truncate the obtained series. One can prove that at n neighbors is gives the best approximation of the ideal filter for n neighbors. The consistent overshoot near ω_1 and ω_2 is due to Gibbs phenomenon is roughly 9% is magnitude. This overshoot can be decreased by using a more smoother $G(\omega)$ in the computation of $\{g_l\}$. The obvious drawback is that the PTF resembles







less the PTF of the perfect filter. The approximated ideal filter has another drawback as well. The value $G(0)$ is non-zero around zero so an and part of a linear trend (with the offset if present) is still present after filtering. The Baxter-King filter is designed to solve this problem as much as possible.

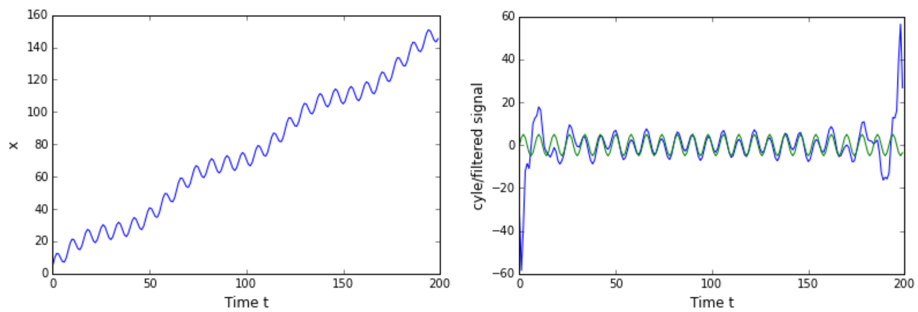
Filtering in Frequency Space There is an obvious method to filter frequencies between ω_1 and ω_2 that we have not yet discussed. We can calculate the DDTF of the data, truncate outside $[\omega_1, \omega_2]$ and then transform back to the time domain with IDDFT. In this way we are actually filtering in the frequency space. Unfortunately, this apparent easy solution comes has a very unpleasant side effect. In the example below we have created a series consisting of the following elements:

$$x_t = \text{trend}_t + \text{long}_t + \text{cycle}_t \tag{1.82}$$

where

$$\begin{aligned}
\text{trend}_t &= 5 + 0.7t \\
\text{long}_t &= 5\sin(2\pi t/60) \\
\text{cycle}_t &= 5\sin(2\pi t/8)
\end{aligned}$$

We have filtered the signal using a smoothed version of the bloc function $1_{[\omega_1, \omega 2]}$ (where $\omega_1 = 2\pi/20$ and $\omega_2 = 2\pi/4$)to achieve better results. We see that the filtered series follows the cycle component quite well in the middle of the sample, but deviates strongly at the end of the sample. This phenomenon is called leakage and occurs whenever the signal contains frequencies which are not Fourier frequencies, i.e. frequencies which are not of the form $n\pi/T$ where T is the length of the signal. In our example we have $T = 200$ and a filtered frequency of $1/60$ which is clearly not a Fourier frequency. It can be shown that leakage can be reduced by either increase the number of data points or by decreasing the maximum value(s) in the spectrogram of the signal. While an increase in data points is not really feasible for historic economic data, the second method does present a realistic opportunity for improvement. This method is the principle idea behind the Zero-Phase filter. It recognizes that filtering in the frequency domain is in principle the optimal way to approximate the ideal bandpass filter



with the optimality obtained by reducing leakage as much as possible. This is done by fitting a number of best correlating sinusoids (fitting frequency, amplitude *and* phase), subtracting these from the signal (thereby reducing the maximum in the spectrogram) and finally filtering both the subtracted and residue signals in frequency space.

It can be shown that the set $\{\exp(i\omega t) | \omega = 2\pi k/T, k = 0, 1, \cdots\}$ forms a dense orthogonal basis (the Fourier basis) in $l^2[0, T]$. Again we will use ω and k interchangeably. The Fourier transform $\mathcal{F}x$ of $x \in l^2[0, T]$ is again defined point-wise by computing the inner product of x w.r.t. the Fourier basis:

$$(\mathcal{F}x)(\omega) = X(k) := \langle x(t), \exp(i\omega t)\rangle = \int_0^T x(t)\exp(i\omega t)dt \tag{1.83}$$

Note that the domain of x is $[0, T]$ whereas the domain of $\mathcal{F}x$ is $\frac{2\pi}{T}\mathbb{Z}_{\geq 0}$. The following expression should feel like a suitable candidate for the inverse Fourier transform:

$$\mathcal{F}^{-1}X(\omega) := \frac{1}{T}\sum_{k\in\mathbb{Z}_{\geq 0}} X(k)\exp(i\omega t) \tag{1.84}$$

To test this we insert the expression above into the FT and simplify:

$$\begin{aligned} X(k) &= \frac{1}{T}\int_0^T \exp(-2\pi i\frac{k}{T}t)\sum_{l\in\mathbb{Z}_{\geq 0}} X(l)\exp(2\pi i\frac{l}{T}t)dt \\ &= \frac{1}{T}\sum_{l\in\mathbb{Z}_{\geq 0}} X(l)\int_0^T \exp(2\pi i\frac{l-k}{T}t)dt \\ &= \frac{1}{T}\sum_{l\in\mathbb{Z}_{\geq 0}} X(l)\cdot T\cdot\langle\exp(2\pi i\frac{l}{T}), \exp(2\pi i\frac{k}{T})\rangle = X(k) \end{aligned} \tag{1.85}$$

since $\langle\exp(-2\pi i\frac{k}{T}), \exp(-2\pi i\frac{l}{T})\rangle = T$ if $k = l$ and zero otherwise. For a full overview of the derivation including the legitimacy of interchanging summation and integration in the analysis see [8].

Chapter 2

The Hilbert-Huang Transform

The subject of this chapter will be the Hilbert Huang Transform (HHT), first proposed by Huang in 1996 [10] as an alternative method to analyze non-stationary signals. In contrast to the WFT and the various wavelet transforms, it does not start out with an a priori basis with which the signal is correlated with. Instead, the HHT provides a procedure which extracts components called *intrinsic mode functions* (IMF) until a stopping criterion is met. As such, the HHT is not a conventional transform since it maps real-valued signals to a sum of IMF's plus a remainder. The usefulness of the HHT lies in a crucial property of IMF's: these, soon to be defined, functions allow for a meaningful definition of instantaneous frequency. Before defining IMF's and presenting the HHT procedure, we start this chapter with a section on instantaneous frequency.

2.1 Instantaneous frequency

Suppose we start with a signal $x(t)$ of the form $x(t) = r(t)\cos(\phi(t))$. One would naturally define the *instantaneous amplitude* as $r(t)$ and the *instantaneous phase* as $\phi(t)$. The *instantaneous angular frequency* would consequently be defined as the derivative of the instantaneous phase ($\omega(t) := \frac{d}{dt}\phi(t)$) and the *instantaneous frequency* is then simply $f(t) = \omega(t)/2\pi$. The problem central to this analysis is the ambiguity arising naturally from deriving the two functions $r(t)$ and $\phi(t)$ from the single function $x(t)$. As an example, consider the signal $x(t) = \sin(2t)\sin(3t)$. As a possible representation one could choose $r(t) = \sin(2t)$ and $\phi(t) = 3t - \pi/2$, or $r(t) = \sin(3t)$ and $\phi(t) = 2t - \pi/2$ or a completely other representation satisfying $x(t) = r(t)\cos(\phi(t))$ (writing $\sin(2t)\sin(3t) = 2\sin(t)\cos(t)\sin(3t)$ shows a third possible representation: $r(t) = 2\sin(t)\sin(3t)$ and $\phi(t) = t$). Each of these representations yield completely different functions for the instantaneous amplitude and instantaneous frequency. Obviously, either the concept of instantaneous frequency is useless or we need to impose restrictions on the signal $x(t)$. Thankfully, the latter option is available. Without giving a full derivation we state that as long as $x(t)$ has a representation $x(t) = r(t)\cos(\phi(t))$ with a $r(t)$ and $\phi(t)$ slowly varying, the representation 'makes sense', i.e. alternative representations with slowly varying $r'(t)$ and $\phi'(t)$ will imply $r(t) \approx r'(t)$ and $\phi(t) \approx \phi'(t)$. Some ambiguity in both the amplitude and phase is unavoidable, giving rise to the definition of *instantaneous bandwidth*: the 'spread' in frequency at any time t resulting from this ambiguity. The soon to be defined IMF's are a class of functions for which the instantaneous frequency makes sense and the goal of the HHT is to decomposes a signal $x(t)$ as a sum of IMF's (plus a low energy remainder). Before defining IMF's we show how to obtain a representation $x(t) = r(t)\cos(\phi(t))$ such that $\phi(t)$ is non-negative.

Using the Euler representation we can always write any complex-valued signal $z(t)$ as $z(t) = A_z(t)\exp(i\phi_z(t))$. The real part of $z(t)$ then equals the expression $A_z(t)\cos(\phi_z(t))$. For real-valued signals the easiest way to obtain the desired product $x(t) = r(t)\cos(\phi(t))$ is to add an imaginary component to x and then use the Euler representation to find $A_x(t)$ and $\phi_x(t)$. We have already seen that negative frequencies have little meaning for real-valued functions, a feature that is reflected in the conjugate symmetry $X(-\omega) = X(\omega)^*$. The goal when creating the complex extension of x is to prevent negative frequencies from emerging. This is achieved by using the Hilbert transform. The Hilbert transform (HT) of a function $x : \mathbb{R} \to \mathbb{R}$ is defined via

$$(\mathcal{H}x)(t) := \frac{1}{\pi}\text{p.v.}\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{2.1}$$

Where p.v. denotes the *Cauchy principle value* of the integral. The expression can be written explicitly as

$$(Hx)(t) = -\frac{1}{\pi}\lim_{\epsilon\to 0}\int_{\epsilon}^{\infty} \frac{x(t+\tau) - x(t-\tau)}{\tau} d\tau \tag{2.2}$$

We state without proof that

$$(\mathcal{F}(\mathcal{H}x))(\omega) = -i\text{sign}(\omega)(\mathcal{F}x)(\omega) \tag{2.3}$$

and hence

$$(\mathcal{F}(x + i\mathcal{H}x))(\omega) = (\mathcal{F}x)(\omega) + \text{sign}(\omega)(\mathcal{F}x)(\omega) = 2(\mathcal{F}x)(\omega) \cdot \mathbb{1}_{\omega\geq 0} \tag{2.4}$$

So the complex function $z(t) := x(t) + i(\mathcal{H}x)(t)$ has no negative frequencies and $\Re(z) = x(t)$, exactly as desired.

2.2 The HHT procedure

The idea of IMF's is that they represent oscillations with a slowly varying phase and amplitude. The definition contains requires two conditions to be met. A

signal $x : [a, b] \rightarrow \mathbb{R}$, $a, b \in \mathbb{R}$ is called an *intrinsic mode function* (IMF) whenever the following two conditions are met:
(1) In the domain $[a.b]$ the number of extrema and zero-crossings must equal or differ by at most one.
(2) At any $t \in [a, b]$ the mean value of the envelope defined by the local maxima and the envelope defined by the local minima must equal zero

The envelope defined by the local maxima is defined to be the qubic spline interpolating the local maxima of x in $[a, b]$, with an analogous definition for the minima envelope. Note that the first requirement is a formal way to define simple (as opposed to composite) oscillations and the second is a way to prevent trend components to be present in IMF's (IMF's must oscillate around zero). Also note that there is some arbitrariness in the cubic spline requirement: some other method of ensuring the 'locally mean zero' property could also be used, possibly yielding very different results. As a final note, we trivially state that harmonics (simple sinusoids) satisfy the definition so IMF's is a superset containing harmonic functions.

The HHT procedure works as follows.

2.2.1 Step 1

For a signal $x : [a, b] \rightarrow \mathbb{R}$ determine the extrema (not counting end points unless they are one-sided extrema) and construct the two functions $u : [a, b] \rightarrow \mathbb{R}$ (upper envelope) and $l : [a, b] \rightarrow \mathbb{R}$ (lower envelope) as the cubic splines connecting the maxima of x for u and the minima for l. Note that each piecewise cubic function has four parameters. For the first interval $[a, \max_1]$ (assuming $a \neq \max_1$), where a is the starting point of x and $\max_1$ the first maximum on $[a, b]$), the only requirement to be satisfied is $u(\max_1) = x(\max_1)$. This leaves three free parameters. For the consequent intervals $[\max_i, \max_{i+1}]$ we can use all four parameters by requiring $u(\max_i) = x(\max_i)$, $u(\max_{i+1}) = x(\max_{i+1})$, $u'(\max_i) = x'(\max_i)$, $u"(\max_i) = x"(\max_i)$. The remaining free parameters can be reduced by imposing certain periodic conditions on the spline, as discussed in......
After the two splines u and l are generated, their mean $m_1 := \frac{1}{2}(u + l)$ is subtracted from x to obtain the first *proto-IMF*: h_1.

2.2.2 Step 2

Ideally, h_1 satisfies the IMF criteria, but generally it will contain extrema not present in x. To visualize this, imagine a point $x(t)$ s.t. $|x"(t)| \ll 1$. Subtracting m_1 may produce a new extrema at t. One can easily check that a negative maximum (i.e. $h_1(t) \leq 0$, $h_1'(t) = 0$, $h_1"(t) < 0$) or positive minimum (i.e. $h_1(t) \geq 0$, $h_1'(t) = 0$, $h_1"(t) > 0$) implies that h_1 is not an IMF, since the first criterion will not be met. The second criterion is also not met due to these new extrema since positive minima and negative maxima implies the mean of the envelopes is non-zero. The remove these unwanted extrema we treat h_1 as data. We construct two new upper and lower splines, call their mean m_{11} and define $h_{11} := h_1 - m_{11}$. We repeat this procedure k times to obtain h_{1k}. The nature of the procedure ensures the mean of consecutive pairs of upper and lower envelopes will approach zero. Furthermore, the number of unwanted extrema will in general be reduced by for an increased number of iterations. Ideally, after k iterations the resulting function h_{1k} is an IMF, but in general there is no guarantee that this can be achieved after a finite number of steps. Therefore, a stoppage criterion is introduced such that h_{1k} satisfies the IMF criteria 'sufficiently'. The most commonly used stoppage criterion is a Cauchy convergence test, requiring the normalised squared difference between two consecutive $h_{1(k-1)}$ and h_{1k} to be less than some predetermined value ϵ. Since x (and hence h_1, m_1, h_{1k}, etc) is actually finite data (we will not considering modeling the HHT for other spaces) we obtain the following expression (assuming the data ranges from $0, 1, \cdots, T$).

$$SD_{1k} := \frac{\sum_{t=0}^{T} |h_{1(k-1)}(t) - h_{1k}(t)|^2}{\sum_{t=0}^{T} h_{k-1}(t)^2} < \epsilon \tag{2.5}$$

Note that this stoppage criterion completely ignores the second criterion for IMF's. An alternative stoppage criterion, proposed by Huang, uses a predefined number S such that the iterations are stopped and a 'sufficiently close' IMF is obtained whenever S consecutive iterations have a number of zero crossings and extrema differing by at most one. Any choice of S is ad hoc, but Huang established an empirical guide and in most cases S will be between 4 and 6. For more details see.....

Assuming the stoppage criterion is met, we have obtained the first IMF c_1.

2.2.3 Step 3

Subtract c_1 from the data x to obtain the first residual $r_1(t)$:

$$r_1(t) := x(t) - c_1(t) \tag{2.6}$$

The procedure is repeated from step one, treating each $r_j(t)$ as data. The procedure is stopped whenever $r_n(t)$ is small enough (using the $L^2[0, T]$ norm) or whenever $r_n(t)$ is monotone, essentially representing the trend of x. Note that any monotone function cannot contain an IMF since it will not contain any extrema and hence no upper and lower envelope can be constructed.

The procedure will result in a set $\{c_j(t)|j = 1, 2, \cdots, n\} \cup \{r_n(t)\}$ such that $x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t)$. Each c_j is an IMF from which we can meaningfully extract the instantaneous amplitude, phase and frequency via

$$\begin{aligned} A_{c_j}(t) &= |c_j(t) + i(\mathcal{H}c_j)(t)| \\ \phi_{c_j}(t) &= \arctan(\mathcal{H}c_j)(t)/c_j(t)) \\ \omega_{c_j}(t) &= \frac{1}{2\pi}\frac{d}{dt}\phi_{c_j}(t) \end{aligned} \tag{2.7}$$

We have already shown that the CFT of a real-valued signal $x(t)$ is fully determined by the sub-domain containing the non-negative frequencies. This holds due to the conjugate symmetry $X(-\omega) = X(\omega)^*$.

Chapter 3

The Vatchev-Sharpley Decomposition

The Hilbert-Huang transform discussed in the previous chapter has the mayor drawback of being a rather ad hoc approach to analyze frequency components in signals. As a consequence, a deep (or even rather shallow) mathematical analysis is not available for the HHT. In this chapter we will discuss a theorem presented by Robert Sharpley and Vesselin Vatchev in a 2006 paper [**?**], which takes a more formal approach of Intrinsic mode functions. They introduce the concept of *weak* IMF, which eseentially is an IMF satisfying only the first criterion presented in the previous chapter, though they use a slightly different (though equivalent) definition. In this chapter, we freely use the definition, lemma's theorems and their derivations presented in Vatchevs and Sharpley's paper.

Definition: A function $\psi : \mathbb{R} \to \mathbb{R}$ is called an Intrinsic mode function (IMF) if the following two criteria are met:
(a) ψ has exactly one zero between any two consecutive extrema.
(b) ψ has zero 'local mean'.

A function ψ only satisfying the first criterion is called a *weak* IMF, and the set of weak IMFs is denoted as $\mathcal{W}$. Note that Vatchev and Sharpley are purposefully ambiguous in their second criterion, though they immediately state that upper and lower envelope defined by the cubic splines is the most widely used method. Their paper however focuses primarily on weak IMFs so the second criterion is less relevant.
The importance of weak IMF's is illustrated in a 2006 paper by Sharpley & Vatchev where they characterize weak IMFs as solutions of self-adjoint differential equations. The following theorem from their paper captures the importance of weak IMFs.

Theorem. Let $\psi \in C^2(I)$ be a weak IMF with simple zeros and extrema, then there exist positive and continuously differentiable functions P and Q such that ψ is the solution of the initial-value problem

$$(Pf')' + Qf = 0, \quad f(\tau) = \psi(\tau), \quad f'(\tau) = \psi(\tau) \tag{3.1}$$

For some $\tau \in I$.
Using this theorem we can define the instantaneous frequency and bandwidth pair $(\theta', r'/r)$ implicitly by using the Prufer substitution defined below:

$$Pf'(t) := r(t)\cos(\theta(t)), \quad f(t) = r(t)\sin(\theta(t)) \tag{3.2}$$

When Q is positive we can define the general Sturm-Liouville operator (SLO) via

$$Lf := -\frac{1}{Q}(Pf')' \tag{3.3}$$

So weak IMFs are eigenfunctions of SLOs, which have the appealing property of having a simple interpretation of time-dependent mass and spring constant (P and Q respectively) in addition to the theoretical theory available for Sturn-Liouville operators, a strong benefit over the HHT. Sharpley and Vatchev claim that any sufficiently smooth function on the finite interval I can be written as the linear sum $a_1\psi_1 + a_2\psi_2$, where ψ_1 and ψ_2 are solutions to the same SLO with different eigenvalues. Unfortunately, in their derivation and construction of P and Q an error was made, and in the general case one cannot construct P and Q such that they are continuously differentiable and positive. The construction of P, Q and the help function x are constructed piecewise on the interval I, which is partitioned into mutually disjoint sets I_i and J_j such that $I = (\cup_{i=0}^{N_I} I_i) \cup (\cup_{j=0}^{N_J} J_j)$. The intervals I_i contain simple inflection points whereas the intervals J_j do not. This partition is needed to find solutions of the differential equations that emerge from the derivation in the paper using two different lemmas presented by Vatchev & Sharpley. We start by presenting these lemmas together with their proof, since we will need both in understanding the construction. We freely quote from Vatchev & Sharpley

We denote the finite collection of inflection points of f by

$$\Xi = \{\xi : f''(\xi) = 0, \xi \in I\} \tag{3.4}$$

In the next result, it is assumed that f, p, q, k_1, k_2 are given, and in order to simplify notation we set

$$G(t, x, z) = k_1 qf - (k_1 + k_2)qx + (k_1 q\frac{f}{f'} + \frac{f''}{f'})z - \frac{k_1 q}{f'}xz \tag{3.5}$$

Lemma A.1 Suppose k_1 and k_2 are positive real numbers and q is continuous on $I_0 \subset I$ with $||q||_\infty < 2||f||_\infty$. Suppose further that ξ is an inflection point of the twice continuously differentiable function f, such that $f'(\xi) \neq 0$, then there exists a subinterval $I_0' = [a,b] \subset I_0$ containing ξ in its interior, such that the initial value problem (IVP)

$$\begin{aligned} x" &= G(t, x(t), x"(t)) \\ x(a) = y_0, &\quad x'(a) = y_1 \end{aligned} \tag{3.6}$$

has a solution on I_0' for any real y_0, y_1.

Proof. Without loss of generality, we may assume from the hypothesis that f' does not vanish on I_0 and by compactness that $|f'(t)| \geq \sigma > 0$ for all $t \in I_0$. Denote, respectively, the constants

$$\begin{aligned} m_0 &:= 2k_1||f"||_\infty ||f||_\infty \\ m_1 &:= 2(k_1 + k_2)||f"||_\infty \\ m_2 &:= ||2k_1|f| + 1||_\infty \frac{||f"||_\infty}{\sigma} \\ m_3 &:= 2k_1 \frac{||f"||_\infty}{\sigma} \end{aligned} \tag{3.7}$$

Further, select positive c_m such that

$$c_m^3 > \max\left(1, \left|m_0 - 3\frac{m_1 m_2}{m_3}\right|, m_3 c_m^2 + (m_1 + m_2)c_m + m_1\right) \tag{3.8}$$

I_0' is then set to the interval centered at ξ with length $\delta := (1/c_m^2)$. If necessary, c_m can be can be increased to ensure $\delta < 1$ and $I_0' \subset I_0$. We show that I_0' is the interval whose existence is stated in the lemma.
Write the second-order equation as its equivalent first-order system in integral form,

$$\begin{aligned} x(t) &= y_0 + \int_a^t z(\tau) d\tau \\ z(t) &= y_1 + \int_a^t G(\tau, x(\tau), z(\tau)) d\tau \end{aligned} \tag{3.9}$$

A standard technique to show the existence of solutions of differential equations is the method of successive approximations. In particular, we may, without loss of generality, assume that $y_0 = y_1 = 0$ and let $x_0 = z_0 = 0$, the inductively construct the sequence pairs

$$\begin{aligned} x_{n+1}(t) &= y_0 + \int_a^t z_n(\tau) d\tau \\ z_n(t) &= y_1 + \int_a^t G(\tau, x_n(\tau), z_n(\tau)) d\tau \end{aligned} \tag{3.10}$$

If we show $|x_n| \leq c_m$ and $|z_n| \leq c_m$ for $n = 0, 1, 2, \cdots$ we can use s standard argument to then establish that the limit function x and z exist and the convergence is uniform. Using the inductive $|x_n| \leq c_m$ and $|z_n| \leq c_m$, we show that $|x_{n+1}| \leq c_m$ and $|z_{n+1}| \leq c_m$. For x_{n+1} we have $|x_{n+1}| \leq |z_n|\delta \leq |z_n| \leq c_m$, since $\delta < 1$. For z_{n+1}, using the bound $q \leq 2||f"||_\infty$, we have $|z_{n+1}| \leq \delta(m_0 + m_1|x_n| + m_2|z_n| + m_3|x_n||z_n|)$. The polynomial function $b(x,z) = m_0 + m_1 x + m_2 z + m_3 xz$ has only one local extremum and and the value at the extremum is $m = |m_0 - 3(m_1 m_2/m_3)| < c_m^3$, from the choice of c_m. The global maxima of $b(x,z)$ on the square $|x| \leq c_m$, $|z| \leq c_m$ are either the value at the extremum or a value from the boundary. It is easy to see that the maximum value on the boundary is $b(c_m, c_m) = m_0 + (m_1 + m_2)c_m + m_3 c_m^2 < c_m^3$, from the choice of c_m. Hence $|z_{n+1}| \leq \delta \max|b(|x|,|z|)| \leq (1/c_m^2)c_m^3 = c_m$ and the limit functions $x = \lim_{n\to\infty} x_n$, and $z = \lim_{n\to\infty} z_n$ is the solution pair of the first-order system defined above. Hence x is the solution of the IVP on the interval $[a, a+\delta]$ that contains the point ξ.

The second lemma addresses intervals that have no reflection points, and in the ultimate contruction are formed by the 'gaps' between the intervals containing inflection points who have reduced in size using the previous lemma.

Lemma A.2. Let $f \in C^2(J_0)$ with simple zeros and extrema. Let $\tilde{t} \in J_0$. Then for any non-zero λ_1 and λ_2 and any u, there exists a positive real number T_0, depending only on J_0 and f, such that for each T, $|T| \geq T_0$, the integral equation,

$$\phi(t) = T + (t-\tilde{t})u + \frac{\lambda_1}{\lambda_2}(f(\tilde{t}) + f'(\tilde{t})(t-\tilde{t}) - f(t)) - \frac{1}{\lambda_2}\int_{\tilde{t}}^{t}\int_{\tilde{t}}^{\tau}\frac{f(v)f"(v)}{\phi(v)}dv \dagger \tau \tag{3.11}$$

has a solution Φ, such that $|\Phi| > |T|/2$ on J_0, with $\phi(\tilde{t}) = T$, $\phi'(\tilde{t}) = u$.

Proof. We prove the lemma in the case $\tilde{t} = a$, where $J_0 = [a,b]$. The case $\tilde{t} \in (a,b]$ is similar. Let $c_1 = |(b-a)u| + (|\lambda_1|/|\lambda_2|)(|f'(a)(b-a)| + |f(a)| + ||f||_\infty)$ and $c_2(|J_0|^2/|\lambda_2|)||ff"||_\infty$. Since $f, f', f"$ are bounded on J_0, it follows that c_1 and c_2 are finite constants. Then pick T_0 positive such that $|T| \geq T_0$ implies

$$|T| > c_1 + \sqrt{c_1^2 + 4c_2} \tag{3.12}$$

Substituting $\phi + T$ for ϕ in the integral equation and applying the method of successive approximations, we consider a sequence of functions ϕ_n, that satisfies the initial conditions $\phi_n(a) = 0$, $\phi_n'(a) = u$. Initialize $\phi_0 = 0$ and inductively define

$$\phi_{n+1}(t) = u(t-a) + \frac{\lambda_1}{\lambda_2}(f(a) + f'(a)(t-a) - f(t)) - \frac{1}{\lambda_2}\int_a^t\int_a^\tau \frac{f(v)f"(v)}{\phi_n(v)+T}dvd\tau \tag{3.13}$$

By construction $|\phi_0| = 0 < |T|/2$. Assume $|\phi_n| < |T|/2$, then $|phi_n + T| \geq |T|/2$. Estimating the r.h.s. of the equation above, it follows from the selection of T that

$$|phi_{n+1}| \leq c_1 + \frac{2c_2}{|T|} < |T|/2 \tag{3.14}$$

Hence the limit function $\Phi = T + \lim_{n\to\infty} \phi_n$ satisfies the integral equation with $\phi(\tilde{t}) = T$, $\phi'(\tilde{t}) = u$ and $|\Phi| \geq |T|/2$ on J_0.

The following theorem is presented in Vatchev & Sharpley, which serves as preliminary result for their further theorems, where they modify the theorem in order to make P and Q differentiable. Unfortunately, this is already the place where a critical error is made. In its proof it uses the lemmas and their proofs presented above.

Theorem 2.1 (Vatchev & Sharpley). For a function $f \in C^2(I)$ with simple extrema and for any positive numbers k_1 and k_2, there exists continuous and positive functions P and Q, depending upon f,and twice differentiable function x such that f and x satisfy the following system of differential equations:

$$\begin{aligned} (Px')' &= -k_2 Qx + k_1 Q(f - x) \\ (Pf')' &= -k_1 Q(f - x) \end{aligned} \tag{3.15}$$

We follow the proof and in an attempt to clarify the mistake made (hard to see, with numerous paramaters to keep track of) we present the simplest of cases where to theorem does not hold: by following the algorithm for the function $f(t) = \cos(t)$. In essence, the algorithm first determines the inflection points of f, then carefully finds intervals I_i around each inflection point such that lemma A.1 can be applied to each of these intervals to construct P, Q and x on each I_i. The gaps between consecutive I_i (plus possible intervals at the end points) form by definition the intervals J_j who, by their very construction, do not contain inflection points. In the proof the author describe the procedure for the case when the interval partition starts with J_0 (i.e. the interval containing the first inflection point, I_0, does not contain the starting point a). We now fully present the proof and then work out the example for $f(t) = \cos(t)$. *Proof*: For each inflection point $\xi \in \Xi$, suppose $I_{j-1} = [b_{j-1}, a_j]$ contains ξ_j and the initial value problem (IVP) considered in lemma A.1 has one solution. Let $a_0 = a$, $b_M = b$ and $J_j = [a_j, b_j]$. First assume $a_0 \in J_0$. The existence of P and Q is shown by inductively constructing them on J_0 and I_0 and repeating the same procedure on J_j and I_j for $j > 0$.
For fixed k_1, k_2 positive, set

$$\begin{aligned} \lambda_1 : &= -\Big(\frac{1}{k_1} + \frac{2}{k_2}\Big) \\ \lambda_2 &= \frac{1}{k_1 k_2}. \end{aligned} \tag{3.16}$$

Since f, f' and f" are bounded on I, then for any real number u, it follows that $C_1 = |(b-a)u| + (|\lambda_1|/|\lambda_2|)(|f'(a)(b-a)| + |f(a)| + ||f||_\infty)$ and $C_2 = ((b-a)^2/|\lambda_2|)||ff"||_\infty$ are finite constants. Pick $\tilde{T} > \max\{1, C_1 + \sqrt{C_1^2 + 4C_2}\}$. (We note that these are the primary constants and parameters in lemma A.2. Although somewhat redundant, introducing these here motives the proof of alternating the successive application of the two lemmas.)

On the interval J_0 the function f does not have an inflection, and hence $|f"| > \eta > 0$. By applying lemma A.2 with $T = -\text{sign}(f"(a_0))\tilde{T}$ and $\phi'(a_0) = u$, we obtain a twice differentiable solution Φ that does not change sign on J_0. Furthermore, since $f"$ does not change its sign on $J_0$and owing to the choice of the initial condition for $\Phi(a_0)$, it follows that the function $Q_0 := -(P_0 f"/\Phi)$ is positive and continuous on J_0 for any fixed function $P_0 > 0$. Fix P_0 a positive constant and define $x = f - (1/k_1)\Phi$, then it follows that

$$P_0 f" = -k_1 Q_0 (f - x) \tag{3.17}$$

Twice differentiating the integral equation in lemma A.2, we get $\Phi" = -(\lambda_1/\lambda_2)f" - (1/\lambda_2)(ff"/\Phi)$ and after substituting in $x" = f" - (1/k_1)\Phi"$, we obtain the following differential equation for x:

$$P_0 x" = P_0 f"\Big(1 + \frac{\lambda_1}{\lambda_2 k_1}\Big) + \frac{P_0}{k_1\lambda_2} f\Big(-\frac{Q_0}{P_0}\Big) = Q_0(-k_2 x + k_1(f - x)) \tag{3.18}$$

By combining the two equations above we have constructed the desired system stated in the theorem for J_0.
Next we construct the functions P_0 and Q_0 on the interval I_0, but here the function f has an inflection point and lemma A.2 cannot be applied. Instead, on this interval this will be the role of lemma A.1. Note that the system in the theorem is equivalent to the system

$$\begin{aligned} x" + px' &= -k_2 q x + k_1 q (f - x) \\ f" + pf' &= -k_1 q (f - x) \end{aligned} \tag{3.19}$$

where $p := (P'/P)$ and $q := (Q/P)$. Let q be the linear function connecting the points $(b_0, (Q_0(b_0)/P_0))$ and $(a_1, (|f"(a_1)|/\tilde{T}))$. Since $f" \geq \eta > 0$ on I_0, then the second equation can be solved for p

$$p = -\frac{f"}{f'} - k_1 q\Big(\frac{f}{f'} - \frac{x}{f'}\Big) \tag{3.20}$$

and upon substituting in the first equation we obtain the differential equation in lemma A.2 for x. The initial conditions naturally come from the already determined values $y_0 = x(b_0)$, $y_1 = x'(b_0)$. On J_0 we have that $\Phi = -(1/q)f"$ and $|\Phi| \geq |\tilde{T}/2|$ including the right endpoint b_0, and hence the estimate $|q_0(b_0)| = (|f"(b_0)|/|\Phi(b_0)|) \leq (2|f"(b_0)|/\tilde{T}) \leq 2||f"||_\infty$ holds true. From the definition of q it follows that $|q(a_1)| \leq (2||f"||_\infty/\tilde{T})$ and since$q$ is linear on I_0 we have that the last estimate holds on over all of I_0. All the conditions of lemma A.1 are met , and hence there exists a function x that is a solution to the differential equation in the lemma. Substituting x in the equation above, we get a continuous function p on $J_0 \cup I_0$. The functions P_1 and Q_1 are determined from the system

$$\begin{aligned} \frac{P_1'}{P_1} &= p \\ \frac{Q_1}{P_1} &= q \end{aligned} \tag{3.21}$$

and have solutions $P_1(t) = P_0 \exp(\int_{b_0}^t p(v)dv)$, $Q_1(t) = P_0 q(t) \exp(\int_{b_0}^t p(v)dv)$, where $P_0 = P_1(b_0)$ and $Q_0(b_0) = Q_1(b_0)$. We have now constructed continuous positive functions P, Q and x, which satisfy the system of the theorem. Repeating the same procedure on J_j and I_j, $j \geq 1$, we construct continuous positive functions P, Q and x, which satisfy the system of the theorem. This completes the proof if the partition of $I = [a, b]$ begins with J_0.

We now work the easiest non-trivial example containing a single inflection point (resulting in the partition $I = J_0 \cup I_0 \cup J_1$). Unfortunately, the differential equations to be solved on these piecewise intervals are so complicated that even for our simple example no closed form solution can be found. It will however, help understand what goes wrong in the theorem.
Consider the cubic function $f(t) = t^3 - 3t^2 + 2$ on the interval $I = [0, 2]$. Note that $f \in C^2(I)$ and f has extrema at the endpoints $t = 0$ and $t = 2$ and a single inflection point at $t = 1$. It is easily checked that both extrema and the zero (at $t = 1$) are simple. We start by determining a suitable interval I_0 around the inflection point $t = 1$. We use lemma A_1 to find it, where we have chosen the positive real numbers (free to choose according to theorem) $k_1 = k_2 = 1$. Using lemma A.1 we deduce that the starting interval around the inflection point $\xi = 1$ can be chosen freely, a long as $|f'(t)| > 0$ on this interval. Following lemma A.1 the starting interval may become smaller to satisfy the needed requirements. We start by setting $I_0 = [1/2, 3/2]$. The minimum of $|f'(t)|$ on I_0 is defined as σ: $\sigma := \min_{t \in I_0} |f'(t)| = 9/4$, and other useful quantities include

$$\begin{aligned}
\|f\|_\infty &= 11/8 \\
\|f"\|_\infty &= 3 \\
m_0 &= 2k_1 \|f"\|_\infty \|f\|_\infty = 33/4 \\
m_1 &= 2(k_1 + k_2)\|f"\|_\infty = 12 \\
m_2 &= \|2k_1|f| + 1\|_\infty \frac{\|f"\|_\infty}{\sigma} = 5 \\
m_3 &= 2k_1 \frac{\|f"\|_\infty}{\sigma} = 8/3
\end{aligned} \tag{3.22}$$

Having defined these constants we now search for a suitable value for c_m, which will determine the final size of the interval I_0. The implicit inequality concerning c_m implies that

$$c_m^3 > m_3 c_m^2 + (m_1 + m_2)c_m + m_1 \tag{3.23}$$

Since c_m needs to be positive and larger than two other arguments in the inequality, we let c_m be a slightly larger value than the highest root of the cubic equation $c_m^3 = m_3 c_m^2 + (m_1 + m_2)c_m + m_1$. By continuity, this implies that the inequality above is satisfied.
The largest root of $c_m^3 = m_3 c_m^2 + (m_1 + m_2)c_m + m_1$ is approximately $c_m \approx 5.8955$ so for convenience we choose $c_m = 6$. Since $|m_0 - 3m_1 m_2/m_3| = 237/4$ we easily check that $c_m^3 = 216 > 237/4 = |m_0 - 3m_1 m_2/m_3|$ so $c_m = 6$ is a suitable candidate. Note that this implies that $\delta := 1/c_m^2 = 1/36$ so $I_0 = [35/36, 37/36]$.

This leaves $J_0 = [0, 35/36]$ and $J_1 = [37/36, 2]$. For this example the interval partition starts with J_0 (rather than I_0) just as the case described in the proof of the theorem.
Having found the partition we follow the rest of the theorem. This yields

$$
\begin{aligned}
&a = 0, \quad b = 2, \quad f'(a) = 0, \quad f(a) = 2, \quad u = 0 \\
&\|f\|_\infty = 2 \quad \|ff''\|_\infty = 12 \\
&\lambda_1 := -\Big(\frac{1}{k_1} + \frac{2}{k_2}\Big) = -3 \\
&\lambda_2 := \frac{1}{k_1 k_2} = 1 \\
&C_1 := |(b-a)u| + (|\lambda_1|/|\lambda_2|)(|f'(a)(b-a)| + |f(a)| + \|f\|_\infty) = 12 \\
&C_2 := ((b-a)^2/|\lambda_2|)\|ff''\|_\infty = 48 \\
&\max\{1, C_1 + \sqrt{C_1^2 + 4C_2}\} = 12 + 4\sqrt{21} \approx 30.33
\end{aligned}
\tag{3.24}
$$

We need to pick $\tilde{T} > \max\{1, C_1 + \sqrt{C_1^2 + 4C_2}\}$ so let $\tilde{T} = 31$. We are now ready to construct the function ϕ on the interval J_0 which we need to find Q, P and the help function x. The initial values for the ODE governing ϕ are $T = -\text{sign}(f''(a_0))\tilde{T} = \tilde{T} = 31$ and $\phi'(a_0) = u = 0$. Writing the integral equation of appendix A.2 as an ODE yields

$$
\begin{aligned}
\phi''(t) &= -\frac{\lambda_1}{\lambda_2} f''(t) - \frac{1}{\lambda_2}\frac{f(t)f''(t)}{\phi(t)} \\
&= 3(6t - 6) - \frac{12(t-1)(2t^2 - 4t - 1)}{\phi(t)} \\
\phi(0) &= T = 31 \\
\phi'(0) &= u = 0
\end{aligned}
\tag{3.25}
$$

This ODE needs to solved numerically for the interval $J_0 = [0, 35/36]$. Picking $P_0 = 1$ than trivially gives Q_0: $Q_0 = -(P_0 f'')/\phi$ Below we plotted ϕ and Q_0 (for completeness we have added P_0 to the plot of Q_0). Next, we need to construct our variables of interest on the interval I_0. To do this the thesis defines the help function q on I_0 as the linear function connecting $(b_0, Q_0(b_0)/P_0)$ and $(a_1, |f''(a_1)|/\tilde{T})$. Evaluating Q_0 at the point $b_0 = 35/36$ and inserting the other parameters gives (rounded) $q(t) = |f''(a_1)|/(\tilde{T}(a_1 - b_0))(t - b_0) + Q_0(b_0)/(P_0(b_0 - a_1))(t - a_1) \approx -0.02312659385t + 0.02914534332 := ct + d$. The differential equation to be solved on the interval I_0 is obtained from lemma A.1, with initial values obtained via $x(b_0) = f(b_0) - (1/k_1)\phi(b_0) \approx -24.9373744949638$ and $x'(b_0) = f'(b_0) - (1/k_1)\phi'(b_0) \approx -2.76038659024925$

$$
\begin{aligned}
&x''(t) = G(t, x(t), x'(t)) \\
&G(t, x(t), x'(t)) = k_1 q f - (k_1 + k_2) q x + \Big(k_1 q \frac{f}{f'} + \frac{f''}{f'}\Big) x' - \frac{k_1 q}{f'} x x' \\
&x(b_0) = -24.9373744949638 \\
&x'(b_0) = -2.76038659024925
\end{aligned}
\tag{3.26}
$$

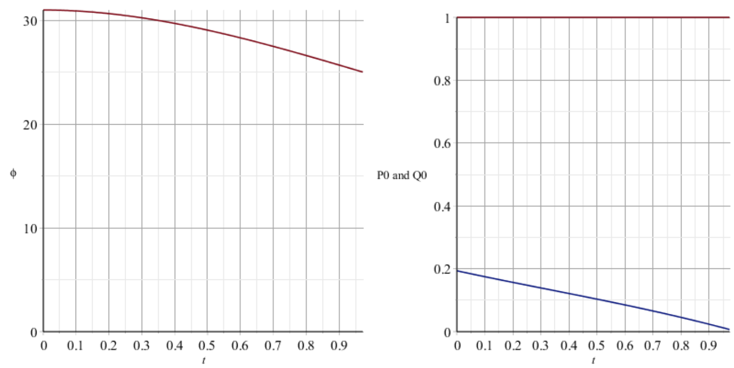


Figure 3.1: ϕ, Q_0 and P_0

Inserting the parameters and writing $q(t) = ct + d$ to avoid the long decimal expansions, we obtain

$$\begin{aligned} x'' &= (ct+d)(t^3 - 3t^2 + 2) - 2(ct+d)x \\ &+ \Big((ct+d)\frac{t^3 - 3t^2 + 2}{3t^2 - 6t} + \frac{3t^2 - 6t}{6t - 6}\Big)x' - \frac{ct+d}{3t^2 - 6t}xx' \end{aligned} \tag{3.27}$$

There is no method known to us that can solve this ODE analytically, but for our purpose a numerical will suffice. Having obtained a numerical approximation for the function x on I_0, the theorem introduces the help function p, dependent on x, through which we can calculate P and Q. As a reminder we restate these relations.

$$\begin{aligned} p(t) &= -\frac{f''}{f'} - k_1 q\Big(\frac{f}{f'} - \frac{x}{f'}\Big) \\ P_1(t) &= P_0 \exp\big(\int_{b_0}^{t} p(v)dv\big) \\ Q_1(t) &= P_0 q(t) \exp\big(\int_{b_0}^{t} p(v)dv\big) \end{aligned} \tag{3.28}$$

We insert the rhs of the first equation in the second and third to obtain an integral containing a numerical expression (the function x) in its integrand. Hence, the integrals themselves need to be calculated numerically. Below we have plotted x, P_1 and Q_1 on I_0. The final step in our worked out example consists of finding the graphs of ϕ, P and Q on the last interval J_1. According to the proof of the theorem, on this last interval we need the same constants λ_1 and λ_2 and initial conditions for ϕ on J_1 set via

$$\begin{aligned} \phi_{J_1}(a_1) &= -\text{sign}(f''(a_1))q(a_1) \\ \phi'_{J_1}(a_1) &= u \end{aligned} \tag{3.29}$$

We would also like to stress that ϕ on J_1 is a help function and is not directly related to ϕ on J_1, which can made apparent by noting that ϕ is not defined on

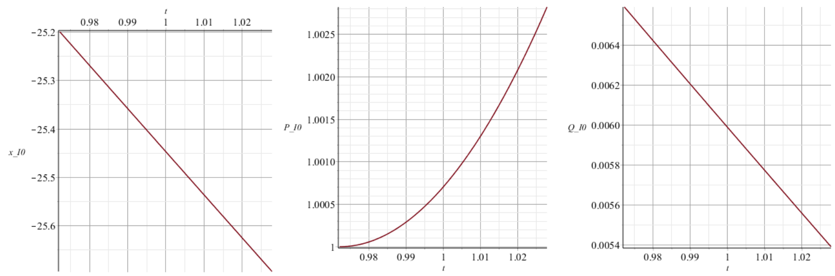


Figure 3.2: x, P_{I0} and Q_{I0}

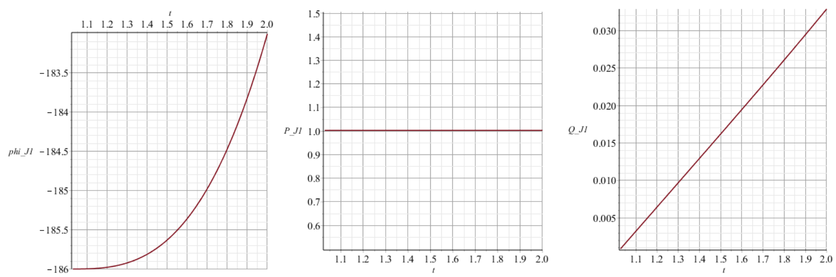


Figure 3.3: x, P_{J1} and Q_{J1}

the middle interval I_0. Finally, the theorem states that u is again free to choose and as before we have set it to zero. The theorem does not mention what P should look like on the interval J_1, but it states that the procedure creating partial solutions for P and Q on J_0 and I_0 should be repeated until the end of the interval $I = [a, b]$ (so for our example we only need to work out the solutions on J_1 to finish, and since the problem first arises in the construction of J_1 our example is in sense of 'minimal complexity').
Apart from requiring a repetition of steps, no further requirements are made. This repetition would suggest that P_{J_1} is constant, and since the theorem claims that P (and Q) are continuous, the only available choice for a constant P_{J_1} is $P_{J_1}(t) \equiv P_{I_0}(a_1)$. In order to calculate Q_{J_1} we need ϕ_{J_1}, which we have plotted below. We can simply show that is not continuous by comparing $Q_{I_0}(a_1) = 0.00539152747311828$ to $Q_{J_1}(a_1) = 0.000898587912366098$. It appears that the requirements on Q on the interval J_1 do not leave sufficient freedom to ensure continuity at the left endpoint a_1. The example demonstrates that continuity of Q is not guaranteed on the right-end boundary points of the intervals I_i. As a final remark, efforts have been made to repair the continuity of Q but it turns out this destroys the continuity of P. Unfortunately, continuity of Q is esential for a meaningful decomposition of the original signal into intrinsic mode functions.

Bibliography

[1] Tinbergen, J., *Economische Bewegingsleer*, North-Holland Publishing Company, Amsterdam, 1946.

[2] Steehouwer, H., *Macroeconomic Scenarios and Reality*, Hens Steehouwer, 2005.

[3] Smith, J., *Mathematics with the Discrete Fourier Transform (DFT): with Audio Applications*, 2nd edition, W3K Publishing, 2007.

[4] Poularikas, A.D.,*The Transform and Applications*, 2nd edition, CRC Press, 2000

[5] Stein, E. M. & Sharkarchi, R., *Fourier Analysis*, Princeton University Press, 2003.

[6] Rynne, B.P. & Youngson, M. A., *Linear Functional Analysis*, 2nd edition, Springer, 2008.

[7] Priestley, H. A., *Complex Analysis*, 2nd edition, Oxford University Press, 2003.

[8] Rudin, W., *Functional Analysis*, 2nd edition, McGraw–Hill, 1991.

[9] Hodrick, R. T. & Prescott, E. C., Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking*, Vol. 29, No. 1 (Feb., 1997), pp. 1-16, Blackwell Publishing, 1997

[10] Huang, N. E. & Long Zheng Shen, S. R., The Mechanism for Frequency Downshift in Nonlinear Wave Evolution, *Advances in Applied Mechanics* Volume 32, 1996, Pages 59–117, 117A, 117B, 117C, Elsevier, 1996.

[11] Vatchev, V. & Sharpley, R., Decomposition of functions into pairs of intrinsic mode functions, *Proc. R. Soc. A*, 464, pp. 2265-2280, 2008.