Hilbert transform: Mathematical theory and applications to signal processing

Måns Klingspor

November 19, 2015

LiTH - MAT - EX - - 2015/09 - - SE

Hilbert transform: Mathematical theory and applications to signal processing

Måns Klingspor
LiTH - MAT - EX - - 2015/09 - - SE

Examensarbete: **30 hp**

Level: **A**

Supervisor: **Natan Kruglyak**
Matematiska institutionen, Linköpings universitet

Examiner: **Irina Asekritova**
Matematiska institutionen, Linköpings universitet

Linköping: **November 2015**

Contents

Acknowledgements

I would like to thank my supervisor, professor Natan Kruglyak, for his invaluable assistance and guidance. Also, I would like to thank my partner Amanda Karunayake for her support and patience. Last but not least, I would also like to thank my family and friends.

Abstract

The Hilbert transform is a widely used transform in signal processing. In this thesis we explore its use for three different applications: electrocardiography, the Hilbert-Huang transform and modulation. For electrocardiography, we examine how and why the Hilbert transform can be used for QRS complex detection. Also, what are the advantages and limitations of this method? The Hilbert-Huang transform is a very popular method for spectral analysis for nonlinear and/or nonstationary processes. We examine its connection with the Hilbert transform and show limitations of the method. Lastly, the connection between the Hilbert transform and single-sideband modulation is investigated.

URL for electronic version:
http://liu.diva-portal.org/smash/record.jsf?pid=diva2:872439

Introduction

The Hilbert transform is one of the most important operators in the field of signal theory. Given some function $u(t)$, its Hilbert transform, denoted by $\mathcal{H}(u(t))$, is calculated through the integral

$$\mathcal{H}(u(t)) = \lim_{\epsilon \to 0} \frac{1}{\pi} \int_{|s-t|>\epsilon} \frac{u(s)}{t-s} ds.$$

The Hilbert transform is named after David Hilbert (1862-1943). Its first use dates back to 1905 in Hilbert's work concerning analytical functions in connection to the Riemann problem. In 1928 it was proved by Marcel Riesz (1886-1969) that the Hilbert transform is a bounded linear operator on $L^p(\mathbb{R})$ for $1 < p < \infty$. This result was generalized for the Hilbert transform in several dimensions (and singular integral operators in general) by Antoni Zygmund (1900-1992) and Alberto Calderón (1920-1998).

Mainly, the importance of the transform is due to its property to extend real functions into analytic functions. This property certainly induces a vast number of applications, especially in signal theory, and obviously the Hilbert transform is not merely of interest for mathematicians.

This thesis revolves around an aim that is twofold:

(i) To acquire more knowledge about the Hilbert transform and some of its applications to signal processing.

(ii) To better understand why some important applications related to the Hilbert transform, to this day, lack mathematical theory.

The thesis consists of two major parts. In the first part mathematical theory of the Hilbert transform is included. These results are well-known but included to provide a steady ground. In the second part, we consider three different applications of the Hilbert transform. The applications that has been considered are: a) Electrocardiography, b) Hilbert-Huang transform and c) modulation. The first two applications nowadays lacks mathematical theory despite numerous efforts. Thus, in this thesis, computer experiments have been carried out in order to deduce limitations of these methods and also pave the way for future resarch in these areas.

(a) Electrocardiography: The Hilbert transform is a widely used tool in interpreting electrocardiograms (ECGs). In Figure 1 we can see a part of an ECG-signal, $ECG(t)$. A common task when dealing with ECG-signals is to extract the so called *QRS complex* which is the high peak seen in the graph of $ECG(t)$ (see Figure 1). In this thesis we thoroughly explore a method of detecting the QRS complex in a ECG signal, as suggested by [2]. This method is very attractive since we are only required to calculate the Hilbert transform and no numerical derivation is involved.

With the Hilbert transform it is possible to expand a real valued signal into a so called analytic signal.

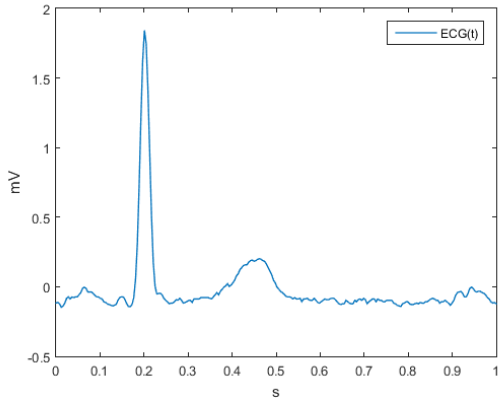

Figure 1: A plot of $ECG(t)$, representing a part of an ECG-signal.

$$z(t) = ECG(t) + i \cdot \mathcal{H}(ECG(t))$$

A parametric plot of $z(t)$, that is, a plot of $ECG(t)$ against $\mathcal{H}(ECG(t))$ reveals interesting things about the $ECG(t)$. Looking at Figure 2, we see that a main loop enclosing the origin is generated. In the thesis, we have shown that if the QRS complex is high enough, it will always produce a closed loop around the origin in the complex plane, distinguishable from the rest of the graph. Also, we have justified this using the fact that the QRS complex resembles a deformed sine wave. By looking at analytic sine waves and deformed sine waves we have established that all type of sine waves, if expanded to analytic signals, form loops enclosing the origin in the complex plane. Thus, the QRS complex, which is a deformed sine wave, also produces enclosed loops in the complex plane.

In our experiments, a limitation was encountered that should be adressed. If the QRS complex to be detected is not high enough, the method does not work. When the QRS complex is too low, the analytic expansion of the QRS complex will not produce a single distinguishable main loop enclosing the origin. It may either not be detected or other peaks in the ECG falsely interpreted as an QRS complex we have found.

(b) The Hilbert-Huang transform: In time series analysis the Fourier transform is the dominating tool. However, this method is not good enough for non-stationary or nonlinear data. For this purpose, the Hilbert-Huang transform (HHT) was proposed in 1996 [10]. This method has gained popularity and is widely used in spectral analysis since it, in contrast to common "Fourier

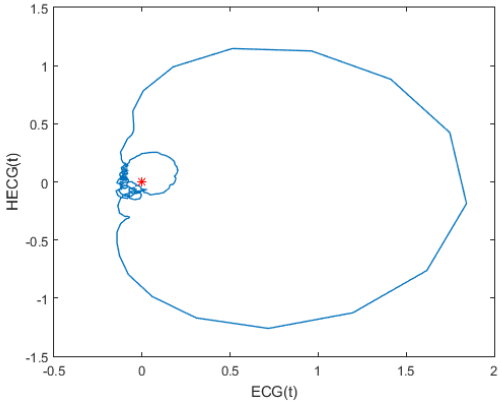

Figure 2: A plot of $ECG(t)$ against $\mathcal{H}(ECG(t))$.

methods", suppose that frequency and amplitude of the harmonics are dependent on time. This is achieved by decomposing the time series into so called intrinsic mode functions (IMFs). However, in spite of considerable efforts, the HHT to this day lacks mathematical framework and the method is entirely empirical. Thus, the investigation of this method is carried out numerically in this thesis to try to understand how the method works and what limitations are inherit.

In the thesis, we state and investigate the following conjecture

Conjecture. Suppose $X(t)$ is a time series and its Hilbert-Huang transform is a decomposition given by

$$X(t) = \sum_{i=1}^{n} c_i(t) + r_n(t).$$

Then the IMFs, $c_k(t)$, are decreasing in complexity as k increases. That is, given $c_k(t)$ and $c_{k+1}(t)$, then $c_k(t)$ has more zero crossings compared to $c_{k+1}(t)$ for $1 \leq k \leq n-1$.

Note that more zero crossing means that $c_k(t)$ has higher frequency content compared to $c_{k+1}(t)$. To check the validity of this conjecture, consider for example a process given by

$$X(t) = \begin{cases} \sin(t), & 0 \leq t \leq 10\pi \\ \sin(t) + \sin(10t), & 10\pi \leq t \leq 100 \end{cases}$$

which also can be seen in Figure 3). In this case we have some background

process given by $\sin(t)$. While the "new" process, given by $\sin(10t)$, is dominating the background process is still active. Obviously, this is a non-stationary situation and therefore the HHT could be very useful in acquiring the harmonics.

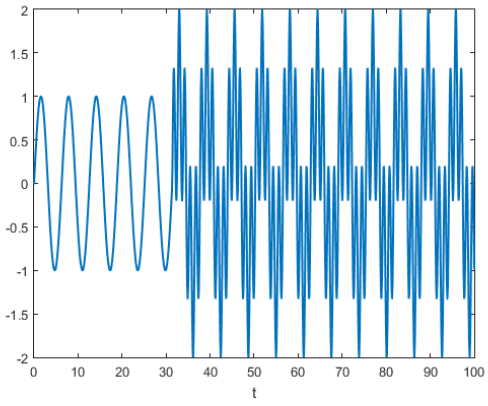

Figure 3: Plot of $X(t)$.

According to the conjecture, we should for $X(t)$ expect the following IMFs.

$$\widetilde{c}_1(t) = \begin{cases} \sin(t), & 0 \leq t \leq 10\pi \\ \sin(10t), & 10\pi \leq t \leq 100 \end{cases}$$

$$\widetilde{c}_2(t) = \begin{cases} 0, & 0 \leq t \leq 10\pi \\ \sin(t), & 10\pi \leq t \leq 100 \end{cases}$$

where the component with the highest frequency in each interval. The actual IMFs, acquired by calculation and denoted by $c_1(t)$ and $c_2(t)$ can be seen in Figure 4.

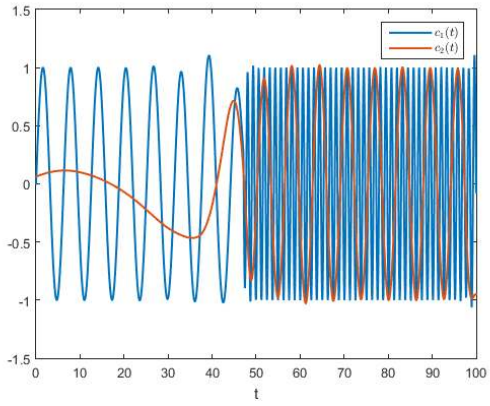

Figure 4: Plot of the two first IMFs, calculated numerically for $X(t)$.

Furthermore, in Figure 45 are the differences of the calculated IMFs and the expected IMFs. Obviously, $c_1(t)$ agrees well with $\widetilde{c}_1(t)$ except for edge effects and in the neighborhood of $t = 10\pi$. Of course, this is because the harmonics shift instantaneously and it is hard for the algorithm to fit this in with IMFs. Despite this, the result of $c_1(t)$ should be considered good since it manages to pick up both harmonics in each interval. For $c_2(t)$, we do not really get what we expected. While it picks up the background harmonics of $\sin(t)$ in $10\pi \leq t \leq 100$, it is a bit off in $0 \leq t \leq 10\pi$. This is because it is hard to fit an IMF to $\widetilde{c}_2(t)$ which has a zero interval and then instantaneous harmonics. To conclude, the HHT is a suitable method of analyzing nonstationary time series. Also, our conjecture seem to hold and may be useful in understanding the HHT.

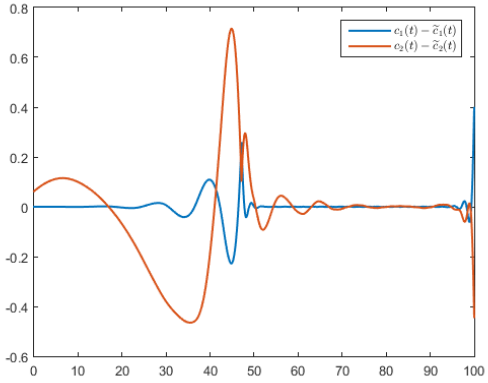

Figure 5: Plot of the difference between the calculated IMFs and the expected IMFs for $X(t)$. That is, $c_1(t) - \widetilde{c}_1(t)$ and $c_2(t) - \widetilde{c}_2(t)$

.

As was seen through the numerical investigations, the method works well as conjectured in analyzing nonstationary and nonlinear data. However, one should be aware of these shortcomings in order to interpret the result correctly. This is especially the case if in a time series there are frequency components that are close to each other (adjacent frequencies). The result is that the algorithm fails to distinguish between these distinct frequencies and this we have investigated thoroughly; both by an example and a more quantative approach. Also, since the method is carried out numerically, finite signals are being considered. The result of this are so called end effects in the IMFs.

(c) Modulation: Transmitting a bandlimited signal, $m(t)$, is usually done in a frequency band centered around some frequency f_c. This is fairly easy accomplished by using modulation

$$m_{AM}(t) = m(t) \cdot \cos(2\pi f_c t).$$

Inherently, this method has a weakness, namely, the frequency content of $M_{AM}(f) = \mathcal{F}(m_{AM}(t))$ is symmetrically doubled around f_c. One way of refining the method is by using the Hilbert transform and properties of analytic signals. This refined modulation is given by

$$m_{SSB}(t) = m(t) \cdot \cos(2\pi f_c t) + \mathcal{H}(m(t)) \cdot \sin(2\pi f_c t).$$

In this section, this refined modulation will be derived in a clear manner.

For the numerical experiments MATLAB R2015a has been used. This version of MATLAB has an inbuilt algorithm `hilbert` that performs the Hilbert transform numerically. An algorithm for the Hilbert-Huang transform, created by Tan, A. in 2008, was downloaded from MATLAB Central [18].

1 The Hilbert transform

1.1 Hilbert transform on the real line

Definition 1.1. (Hilbert transform on $\mathbb{R}$). Let $x(t) \in L^p(\mathbb{R})$ be a function for $1 \leq p < \infty$. Then $\mathcal{H}(x(t))$ is the Hilbert transform of $x(t)$ given by

$$\mathcal{H}(x(t)) = \frac{1}{\pi} \, \mathrm{PV} \int_{-\infty}^{\infty} \frac{x(s)}{t-s} ds$$

Where "PV" is the Cauchy Principal Value of the integral.

For the Hilbert transform the Cauchy Principal Value is necessary in order to handle the singularity at $s = t$. The Cauchy Principal Value is, in this context, utilized in the manner as follows [4].

$$\frac{1}{\pi} \, \mathrm{PV} \int_{-\infty}^{\infty} \frac{x(s)}{t-s} ds = \lim_{\epsilon \to 0^+} \frac{1}{\pi} \int_{|t-s| \geq \epsilon} \frac{x(s)}{t-s} ds$$

It is not obvious that this integral converges and consequently the Hilbert transform is well-defined. This issue will be revisited and adressed in a later section.

Example 1.2. The Hilbert transform for a constant function $x(t) = c$ is easy to calculate using the definition

$$\mathcal{H}(c) = \frac{1}{\pi} \, \mathrm{PV} \int_{-\infty}^{\infty} \frac{c}{t-s} ds = \frac{c}{\pi} \, \mathrm{PV} \int_{-\infty}^{\infty} \frac{1}{t-s} ds = 0$$

The last equality is due to the integrand $1/(t-s)$ being an odd function over a symmetric interval around $s = t$. Hence, $\mathcal{H}(c) = 0$ for any constant c. □

Of course, it is not always this easy and straight forward to calculate the Hilbert transform. For more advanced functions we need to resort to techniques from complex analysis in order to handle the integral. These techniques include contour integrals in the complex plane and the residue theorem. Obviously, it is the singularity in $x = t$ on the real line we need to take care of. For this and other purposes, the following lemmas are very useful.

Lemma 1.3. Let $f(z)$ be analytic in some neighborhood of z_0 where it has a simple pole and C_ϵ is a circular arc defined by $C_\epsilon : z = z_0 + \epsilon \cdot e^{i\varphi}$ with $\varphi : \varphi_1 \to \varphi_2$, then

$$\lim_{\epsilon \to 0^+} \int_{C_\epsilon} f(z) dz = i(\varphi_2 - \varphi_1) \cdot \operatorname*{Res}_{z=z_0} f(z)$$

Proof. Since $f(z)$ has a simple pole at $z = z_0$ we can express its Laurent expansion in some punctured neighborhood around z_0.

$$f(z) = \frac{a_{-1}}{z - z_0} + \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

Let $g(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$. Breaking up the integral into parts we get

$$\int_{C_\epsilon} f(z)dz = a_{-1} \int_{C_\epsilon} \frac{1}{z - z_0} dz + \int_{C_\epsilon} g(z)dz \tag{1}$$

In said neighborhood around z_0 the function $g(z)$ is analytical and bounded so $|g(z)| \leq M$ for some constant M. Using estimations we get

$$\left| \int_{C_\epsilon} g(z)dz \right| \leq M \cdot (\varphi_2 - \varphi_1)\epsilon \to 0 \text{ as } \epsilon \to 0^+$$

For the other integral we get, using the paremetrization stated in the theorem

$$\int_{C_\epsilon} \frac{1}{z - z_0} dz = \int_{\varphi_1}^{\varphi_2} \frac{1}{\epsilon e^{i\varphi}} \epsilon i e^{i\varphi} d\varphi = i \int_{\varphi_1}^{\varphi_2} d\varphi = i(\varphi_2 - \varphi_1)$$

From this and the fact that $a_{-1} = \operatorname*{Res}_{z=z_0} f(z)$ we get in (1) by evaluating limits

$$\lim_{\epsilon \to 0^+} \int_{C_\epsilon} f(z)dz = a_{-1} \cdot i(\varphi_2 - \varphi_1) + 0 = i(\varphi_2 - \varphi_1) \cdot \operatorname*{Res}_{z=z_0} f(z)$$

And the proof is thus complete. □

Remark 1.4. In the case of the Hilbert transform, this simple pole will always be on the real line at the point $z = t$ for a fixed t. To avoid this pole, a small semi-circle in either half-plane with $\varphi_1 = \pi$ and $\varphi_2 = 0$ will suffice.

Lemma 1.5. (Jordan's lemma). If C_R^+ is the semicircle $z = Re^{i\varphi}$, $\varphi : 0 \to \pi$ in the upper halfplane, $a > 0$ and $R > 0$, then

$$\int_{C_R^+} |e^{iaz}||dz| \leq \frac{\pi}{a}$$

Proof. With the parameterization $z = Re^{i\varphi} = R\cos\varphi + iR\sin\varphi$ we get $|e^{iaz}| = |e^{ia(x+iy)}| = |e^{iax-ay}| = e^{-ay} = e^{-aR sin\varphi}$ and $|dz| = |iRe^{i\varphi}d\varphi| = Rd\varphi$.

$$\begin{aligned}
\int_{C_R^+} |e^{iaz}||dz| &= \int_0^{\pi} e^{-aR\sin\varphi} Rd\varphi \overset{(1)}{=} 2\int_0^{\pi/2} e^{-aR\sin\varphi} Rd\varphi \\
&\overset{(2)}{\leq} \int_0^{\pi/2} e^{-aR\cdot 2\varphi/\pi} Rd\varphi = -\frac{\pi}{a}\left[e^{-aR\cdot 2\varphi/\pi}\right]_{\varphi=0}^{\varphi=\pi/2} \\
&= \frac{\pi}{a}(1 - e^{-aR}) \overset{(3)}{\leq} \frac{\pi}{a}
\end{aligned}$$

The equality (1) is because $\sin\varphi$ is symmetric around $\varphi = \pi/2$. Furthermore, since $\sin\varphi$ is concave in the interval $0 \leq \varphi \leq \pi/2$ its graph lies totally above a straight line connecting its endpoints. Therefore, we have that $\sin\varphi \geq 2\varphi/\pi$ when $0 \leq \varphi \leq \pi/2$ and hence inequality (2) holds. Of course, in order for (3) to hold we must have that $a > 0$ [9].

Example 1.6. Let $x(t) = e^{i\omega t}$ where ω is some real parameter. This is an complex exponential function and we would like to calculate its Hilbert transform which is defined by the integral

$$\mathcal{H}(e^{i\omega t})(t) = \frac{1}{\pi}\,\mathrm{PV}\int_{-\infty}^{\infty}\frac{e^{i\omega s}}{t-s}ds \tag{2}$$

Start by looking at the case $\omega > 0$. In order to calculate the integral we will have to use techniques from complex analysis with contour integrals in the complex plane.

Define $f(z) = 1/\pi \cdot e^{i\omega z}/(t-z)$ and the contour $C = C_R^+ + L_{\epsilon,R}^2 + C_\epsilon + L_{\epsilon,R}^1$ which is closed and positively oriented. C_R^+ is a semicircle with radius R in the upper half-plane with the parameterization $z = Re^{i\varphi}, \varphi : 0 \to \pi$. $L_{\epsilon,R}^1$ is a straight line from $z = -R$ to $z = t - \epsilon$ on the real line. C_ϵ is a semicircle with radius ϵ in the upper half-plane with the parameterization $z = \epsilon \cdot e^{i\varphi}, \varphi : \pi \to 0$. $L_{\epsilon,R}^2$ is the straight line from $z = t + \epsilon$ to $z = R$ on the real line. A sketch of the contour C is provided in Figure 6.

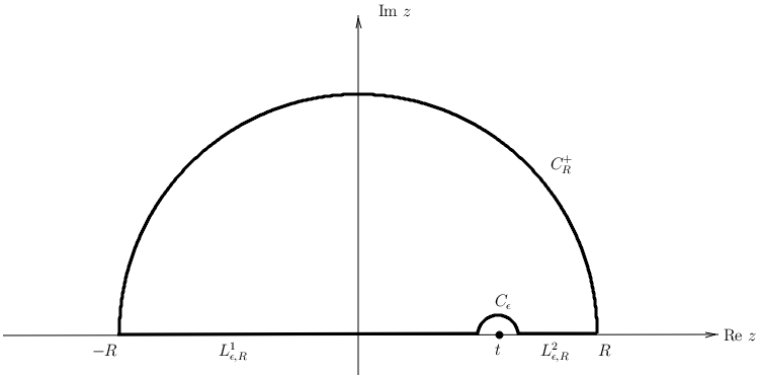

Figure 6: Sketch of the contour C. The contour is positively oriented.

Since $f(z)$ is analytical both on and inside the contour C we have

$$\int_C f(z)dz = 0$$

because of Cauchy's integral theorem. Breaking up the integral thus yields

$$\int_{L_{\epsilon,R}^1 + L_{\epsilon,R}^2} f(z)dz = -\int_{C_R^+} f(z)dz - \int_{C_\epsilon} f(z)dz \tag{3}$$

Because $\omega > 0$ and C_R^+ is in the upper half-plane Jordan's lemma gives us

$$\left|\int_{C_R^+} f(z)dz\right| \leq \int_{C_R^+} |f(z)||dz| = \int_{C_R^+} \frac{|e^{i\omega z}|}{|\pi(t-z)|}|dz| \leq \frac{1}{\pi}\cdot\frac{1}{|R-t|}\int_{C_R^+}|e^{i\omega z}||dz|$$
$$\leq \frac{1}{\pi}\cdot\frac{1}{|R-t|}\cdot\frac{\pi}{2} = \frac{1}{2}\cdot\frac{1}{|R-t|}$$

Furthermore, for the integral over the contour C_ϵ we get by Lemma 1.3

$$\int_{C_\epsilon} f(z)dz = i(0-\pi)\cdot\operatorname*{Res}_{z=t} f(z) = -i\pi\cdot\left(\frac{e^{i\omega t}}{-\pi}\right) = ie^{i\omega t}$$

Of course, the integral over the contour $L^1_{\epsilon,R} + L^2_{\epsilon,R}$ in (2) coincide with the integral in (1) when $R \to \infty$ and $\epsilon \to 0^+$. For a fixed t, applying said limits and for $\omega > 0$ in (2) yield

$$\frac{1}{\pi}\,\mathrm{PV}\int_{-\infty}^{\infty}\frac{e^{i\omega t}}{t-s}ds = -0 - ie^{i\omega t} = -ie^{i\omega t},\ \ \omega > 0$$

For the case $\omega < 0$ one may use the same technique as above with the exeception that the contours C_R^+ and C_ϵ are instead semicircles in the lower half-plane. This would for $\omega < 0$ simply give us

$$\frac{1}{\pi}\,\mathrm{PV}\int_{-\infty}^{\infty}\frac{e^{i\omega t}}{t-s}ds = ie^{i\omega t},\ \ \omega < 0$$

For $\omega = 0$ the complex exponential function is simply a constant and thus its Hilbert transform becomes 0 (as shown in Example 5.1) . Consequently, we may conclude that for all $\omega \in \mathbb{R}$

$$\mathcal{H}(e^{i\omega t}) = -i\,\mathrm{sgn}\,(\omega)e^{i\omega t}$$

This important result concludes this example. □

1.2 Basic properties of the Hilbert transform

Theorem 1.7. Let $y(t) = \mathcal{H}(x(t))$, $y_1(t) = \mathcal{H}(x_1(t))$, $y_2(t) = \mathcal{H}(x_2(t))$ and let a, a_1, a_2 be some arbitrary constants. Then the Hilbert transform satisifies the following basic properties:

(i) **Linearity:** $\mathcal{H}(a_1x_1(t) + a_2x_2(t)) = a_1\mathcal{H}(x_1(t)) + a_2\mathcal{H}(x_2(t))$

(ii) **Time shift:** $\mathcal{H}(x(t-a)) = y(t-a)$

(iii) **Scaling:** $\mathcal{H}(x(at)) = y(at), \quad a > 0$

(iv) **Time reversal:** $\mathcal{H}(x(-at)) = -y(-at), \quad a > 0$

(v) **Derivative:** $\mathcal{H}(x'(t)) = y'(t)$

Proof: Property (i) follows directly from the fact that the integral itself is an linear operation. Properties (ii)-(iv) can all be shown by a trivial change of variables. Property (v), which requires differentiability, will be shown later.

Example 1.8. With the properties above we may calculate the Hilbert transforms of the basic trigonometric functions. Suppose $\omega > 0$.

$$\begin{aligned}\mathcal{H}(\cos(\omega t)) &= \mathcal{H}\left(\frac{e^{i\omega t}+e^{-i\omega t}}{2}\right) = \frac{\mathcal{H}(e^{i\omega t})+\mathcal{H}(e^{i(-\omega)t})}{2} \\ &= \frac{-i\,\mathrm{sgn}\,(\omega)e^{i\omega t} - i\,\mathrm{sgn}\,(-\omega)e^{i(-\omega)t}}{2} \\ &= \frac{-ie^{i\omega t}-ie^{-i\omega t}}{2} = \frac{e^{i\omega t}-e^{-i\omega t}}{2i} = \sin(\omega t) \\ \mathcal{H}(\sin(\omega t)) &= \mathcal{H}\left(\cos\left(\omega t - \frac{\pi}{2}\right)\right) = \sin\left(\omega t - \frac{\pi}{2}\right) = -\cos(\omega t)\end{aligned}$$

In conclusion, if $\omega > 0$ (which usually is the case) then $\mathcal{H}(\cos(\omega t)) = \sin(\omega t)$ and $\mathcal{H}(\sin(\omega t)) = -\cos(\omega t)$. By the time reversal property, if $\omega < 0$ then $\mathcal{H}(\cos(\omega t)) = -\sin(\omega t)$ and $\mathcal{H}(\sin(\omega t)) = -(-\cos(\omega t)) = \cos(\omega t)$. Written in a more compact form, for all $\omega \in \mathbb{R}$ we have

$$\begin{aligned}\mathcal{H}(\cos(\omega t)) &= \mathrm{sgn}\,(\omega)\sin(\omega t) \\ \mathcal{H}(\sin(\omega t)) &= -\,\mathrm{sgn}\,(\omega)\cos(\omega t)\end{aligned}$$

which are the Hilbert transforms for the basic cosine and sine-functions. □

What properties does the Hilbert transform exhibit for even and odd functions? The following theorem gives us the answer.

Theorem 1.9. (Even and odd functions). Suppose $f(t)$ has a well-defined Hilbert transform and that $f(t)$ is either even or odd. Then, the following holds

(i) If $f(t)$ is even, its Hilbert transform is an odd function

(ii) If $f(t)$ is odd, its Hilbert transform is an even function

Proof. From the definition of the Hilbert transform it follows that

$$\mathcal{H}(f(t)) = \frac{1}{\pi}\,\mathrm{PV}\int_{-\infty}^{\infty}\frac{f(s)}{t-s}ds = \frac{1}{\pi}\,\mathrm{PV}\int_{0}^{\infty}\left(\frac{f(s)}{t-s}+\frac{f(-s)}{t+s}\right)ds$$

If $f(t)$ is an even function, that is, $f(-t) = f(t)$ we get

$$\begin{aligned}\mathcal{H}(f(t)) &= \frac{1}{\pi}\,\mathrm{PV}\int_0^\infty \left(\frac{f(s)}{t-s} + \frac{f(s)}{t+s}\right) ds \\ &= \frac{1}{\pi}\,\mathrm{PV}\int_0^\infty \left(\frac{(t+s)f(s) + (t-s)f(s)}{t^2-s^2}\right) ds \\ &= \frac{2t}{\pi}\,\mathrm{PV}\int_0^\infty \frac{f(s)}{t^2-s^2} ds\end{aligned}$$

and if $f(t)$ is an odd function, that is, $f(-t) = -f(t)$ we get

$$\begin{aligned}\mathcal{H}(f(t)) &= \frac{1}{\pi}\,\mathrm{PV}\int_0^\infty \left(\frac{f(s)}{t-s} - \frac{f(s)}{t+s}\right) ds \\ &= \frac{1}{\pi}\,\mathrm{PV}\int_0^\infty \left(\frac{(t+s)f(s) - (t-s)f(s)}{t^2-s^2}\right) ds \\ &= \frac{2}{\pi}\,\mathrm{PV}\int_0^\infty \frac{sf(s)}{t^2-s^2} ds.\end{aligned}$$

From these two expressions the theorem follows and the proof is complete. □

1.3 Hilbert transform and the Fourier transform

It is easy to see that the Hilbert transform $y(t) = \mathcal{H}(x(t))$ actually can be interpreted as an convolution between $x(t)$ and $1/(\pi t)$. This requires, however, a rigorous proof. For this purpose, we need a couple of important theorems from functional analysis [11].

Theorem 1.10. (Hölder's inequality). Suppose $f(t) \in L^p(\mathbb{R})$, $g(t) \in L^q(\mathbb{R})$ where $1/p + 1/q = 1$ for $1 \leq p, q \leq \infty$. Then,

$$||fg|| \leq ||f||_p ||g||_q$$

Theorem 1.11. If $f(t) \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$ and $g(t) \in L^1(\mathbb{R})$, then the convolution $(f * g)(t)$ is in $L^p(\mathbb{R})$ and

$$\|f * g\|_p \leq \|f\|_p \|g\|_1$$

Proof. For the proof we will consider three different cases, namely $p = 1$, $p = \infty$ and $1 < p < \infty$.

(i) $p = 1$: This case is trivial since

$$\begin{aligned}\|f * g\|_1 &= \left\| \int_{\mathbb{R}} f(s)g(t-s)ds \right\| = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(s)g(t-s)ds \right| dt \\ &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |f(s)||g(t-s)|dsdt = \int_{\mathbb{R}} |f(s)| \int_{\mathbb{R}} |g(t-s)|dtds \\ &= \int_{\mathbb{R}} |f(s)| \cdot \|g\|_1 ds = \|g\|_1 \int_{\mathbb{R}} |f(s)|ds = \|g\|_1 \|f\|_1 = \|f\|_1 \|g\|_1\end{aligned}$$

(ii) $p = \infty$: As a reminder, $\|f\|_\infty = \sup_t |f(t)| < \infty$ if $f(t) \in L^\infty(\mathbb{R})$.

$$\begin{aligned}||f * g||_\infty &= \sup_t \left| \int_{\mathbb{R}} f(s)g(t-s)ds \right| = \sup_t \left| \int_{\mathbb{R}} f(t-s)g(s)ds \right| \\ &\leq \sup_t \int_{\mathbb{R}} |f(t-s)||g(s)|ds = \int_{\mathbb{R}} \sup_t |f(t-s)||g(s)|ds \\ &= \int_{\mathbb{R}} \|f\|_\infty |g(s)|ds = \|f\|_\infty \int_{\mathbb{R}} |g(s)|ds = \|f\|_\infty \|g\|_1\end{aligned}$$

(iii) $1 < p < \infty$: Let q satisfy $\frac{1}{p} + \frac{1}{q} = 1$.

$$\begin{aligned}|(f * g)(t)| &= \left| \int_{\mathbb{R}} f(s)g(t-s)ds \right| \leq \int_{\mathbb{R}} |f(s)||g(t-s)|ds \\ &= \int_{\mathbb{R}} |f(s)||g(t-s)|^{1/p} |g(t-s)|^{1/q} ds \\ &\overset{(*)}{\leq} \left(\int_{\mathbb{R}} |f(s)|^p |g(t-s)|ds \right)^{1/p} \left(\int_{\mathbb{R}} |g(t-s)|ds \right)^{1/q} \\ &= \|g\|_1^{1/q} \left(\int_{\mathbb{R}} |f(s)|^p |g(t-s)|ds \right)^{1/p}\end{aligned}$$

Where $(*)$ is because of Hölder's inequality. Taking the $L^p(\mathbb{R})$-norm of both sides yields

$$\begin{aligned}\|f * g\|_p &\leq \|g\|_1^{1/q} \left(\int_{\mathbb{R}} \int_{\mathbb{R}} |f(s)|^p |g(t-s)|dsdt \right)^{1/p} \\ &= \|g\|_1^{1/q} \left(\int_{\mathbb{R}} |f(s)|^p \int_{\mathbb{R}} |g(t-s)|dtds \right)^{1/p} \\ &= \|g\|_1^{1/q} \left(\int_{\mathbb{R}} |f(s)|^p \|g\|_1 ds \right)^{1/p} = \|g\|_1^{1/q} \|g\|_1^{1/p} \left(\int_{\mathbb{R}} |f(s)|^p ds \right)^{1/p} \\ &= \|g\|_1^{1/p+1/q} \|f\|_p = \|f\|_p \|g\|_1\end{aligned}$$

With cases (i), (ii) and (iii) proved the proof is thus complete.

Theorem 1.12. Suppose $x(t) \in L^p(\mathbb{R})$, $1 < p \leq 2$ and $\mathcal{F}(x(t))$ is the Fourier transform of $x(t)$. Then the Fourier transform of $\mathcal{H}(x(t))$ is given by $\mathcal{F}(\mathcal{H}(x(t))) = (-i \operatorname{sgn}(f)) \cdot \mathcal{F}(x(t))$.

Proof. Define the function

$$u_{\epsilon,R}(t) = \begin{cases} \frac{1}{\pi t}, & 0 < \epsilon < |t| < R < \infty \\ 0 & \text{otherwise} \end{cases}$$

and let $\mathcal{H}_{\epsilon,R}$ be a truncated Hilbert transform. Clearly

$$\mathcal{H}_{\epsilon,R}(x(t)) = \frac{1}{\pi} \int_{\epsilon<|t|<R} \frac{x(s)}{t-s} ds = \int_{-\infty}^{\infty} x(s) u_{\epsilon,R}(t-s) ds = (x * u_{\epsilon,R})(t)$$

is a convolution that makes sense since $u_{\epsilon,R} \in L^1(\mathbb{R})$ and $x(t) \in L^p(\mathbb{R})$, $1 < p \leq 2$. Then, according to Theorem 5.10, $\|x * u_{\epsilon,R}\|_1 \leq \|x\|_p \|u_{\epsilon,R}\|_1$. This also means that $\mathcal{H}_{\epsilon,R}(x(t)) \in L^p(\mathbb{R})$, $1 < p \leq 2$. The Fourier transform of $\mathcal{H}_{\epsilon,R}(x(t))$ is given by

$$\mathcal{F}(\mathcal{H}_{\epsilon,R}(x(t))) = \mathcal{F}\left((x * u_{\epsilon,R})(t)\right) = \mathcal{F}(x(t)) \cdot \mathcal{F}(u_{\epsilon,R}(t))$$

and we need to calculate $\mathcal{F}(u_{\epsilon,R}(t))$.

$$\begin{aligned} \mathcal{F}(u_{\epsilon,R}(t)) &= \int_{\epsilon<|t|<R} \frac{e^{-2\pi i f t}}{\pi t} dt = \int_{-R}^{-\epsilon} \frac{e^{-2\pi i f t}}{\pi t} dt + \int_{\epsilon}^{R} \frac{e^{-2\pi i f t}}{\pi t} dt \\ &= -\int_{\epsilon}^{R} \frac{e^{2\pi i f t}}{\pi t} dt + \int_{\epsilon}^{R} \frac{e^{-2\pi i f t}}{\pi t} dt = -\frac{1}{\pi} \int_{\epsilon}^{R} \frac{e^{2\pi i f t} - e^{-2\pi i f t}}{t} dt \\ &= -\frac{2i}{\pi} \int_{\epsilon}^{R} \frac{\sin 2\pi f t}{t} dt = -\frac{2i \operatorname{sgn}(2\pi f)}{\pi} \int_{2\pi|f|\epsilon}^{2\pi|f|R} \frac{\sin t}{t} dt \\ &= -\frac{2i \operatorname{sgn}(f)}{\pi} \int_{2\pi|f|\epsilon}^{2\pi|f|R} \frac{\sin t}{t} d \end{aligned}$$

The integral at the end has the limit $\pi/2$ as $\epsilon \to 0^+$, $R \to \infty$. This can be shown quite easily just using earlier employed techniques with contour integrals in the complex plane. Obviosuly, we have that $\mathcal{F}(u_{\epsilon,R}(t)) \to -i \operatorname{sgn}(f)$ as $\epsilon \to 0^+$, $R \to \infty$ for every real value of f.

Since $\mathcal{F}(u_{\epsilon,R}(t)) \to -i \operatorname{sgn}(f)$ as $\epsilon \to 0^+$, $R \to \infty$ we have that $|\mathcal{F}(u_{\epsilon,R}(t))| \leq C$ for some C depending on the values of ϵ and R. We can now show that

$$\mathcal{H}(x(t)) = \lim_{\epsilon \to 0^+} \lim_{R \to \infty} \mathcal{H}_{\epsilon,R}(x(t)) = \mathcal{F}^{-1}\left(-i \operatorname{sgn}(f) \cdot \mathcal{F}(x(t))\right)$$

by making use of Lebesgue's theorem of dominated convergence. For $1 < p \leq 2$

$$\begin{aligned}
&\lim_{\epsilon\to 0^+}\lim_{R\to\infty} \|\mathcal{H}_{\epsilon,R}(x(t)) - \mathcal{F}^{-1}\left(-i\,\mathrm{sgn}\,(f)\cdot\mathcal{F}(x(t))\right)\|_p \\
&\quad \overset{(*)}{=} \lim_{\epsilon\to 0^+}\lim_{R\to\infty} \|\mathcal{F}(\mathcal{H}_{\epsilon,R}(x(t))) - (-i\,\mathrm{sgn}\,(f)\cdot\mathcal{F}(x(t)))\|_p \\
&\quad = \lim_{\epsilon\to 0^+}\lim_{R\to\infty} \|\mathcal{F}(x(t))\cdot\mathcal{F}(u_{\epsilon,R}(t)) - (-i\,\mathrm{sgn}\,(f)\cdot\mathcal{F}(x(t)))\|_p \\
&\quad = \lim_{\epsilon\to 0^+}\lim_{R\to\infty} \|(\mathcal{F}(u_{\epsilon,R}(t)) - (-i\,\mathrm{sgn}\,(f)))\cdot\mathcal{F}(x(t))\|_p = 0
\end{aligned}$$

where $(*)$ is because of Parseval's identity. Thus we have established the limit above (in the sense of the L^p norm). Now, by just taking the Fourier transform of both sides we get because of the Fourier inversion theorem that

$$\mathcal{F}(\mathcal{H}(x(t))) = (-i\,\mathrm{sgn}\,(f))\cdot\mathcal{F}(x(t))$$

and the proof is complete. □

Remark 1.13. Why does the theorem not hold for the case $p = 1$? Because $x(t) \in L^1(\mathbb{R})$ does not necessarily imply that $\mathcal{H}(x(t)) \in L^1(\mathbb{R})$. For example, let $x(t) = \chi_{[0,1]}(t) \in L^1(\mathbb{R})$. Then its Hilbert transform is given by

$$\mathcal{H}(x(t)) = \frac{1}{\pi}\,\mathrm{PV}\int_0^1 \frac{1}{t-s}ds = \frac{1}{\pi}\ln\left|\frac{t}{t-1}\right|$$

and clearly $\mathcal{H}(x(t)) \notin L^1(\mathbb{R}$. Therefore, the Fourier transform of $\mathcal{H}(x(t))$ does not exist in the usual sense and the theorem does not hold. However, if both $x(t)$, $\mathcal{H}(x(t)) \in L^1(\mathbb{R})$ then all steps in the proof are valid and the theorem holds even for $p = 1$.

With this relationship between the Hilbert transform and the Fourier transform we may show various different properties of the Hilbert transform. We begin with the differentiation property.

Theorem 1.14. (Differentiation). If $x(t) \in L^p(\mathbb{R})$, $1 < p \leq 2$ is differentiable, then it holds that $\mathcal{H}(x'(t)) = \frac{d}{dt}\mathcal{H}(x(t))$.

Proof. By the differentation property of the Fourier transform and Theorem 1.11 we get by some simple algebra

$$\begin{aligned}
\mathcal{F}(\mathcal{H}(x'(t))) &= (-i\,\mathrm{sgn}\,(f))\cdot\mathcal{F}(x'(t)) = (-i\,\mathrm{sgn}\,(f))\cdot(2\pi i f\cdot\mathcal{F}(x(t)) \\
&= 2\pi i f\cdot(-i\,\mathrm{sgn}\,(f))\cdot\mathcal{F}(x(t)) = 2\pi i f\cdot\mathcal{F}(\mathcal{H}(x(t))) \\
&= \mathcal{F}\left(\frac{d}{dt}\mathcal{H}(x(t))\right)
\end{aligned}$$

Through this, since they share the same Fourier transform, it is established that $\mathcal{H}(x'(t)) = \frac{d}{dt}\mathcal{H}(x(t))$, thus the differentiation property of the Hilbert transform.

Another interesting property we can show is the inversion property.

Theorem 1.15. (Inversion). Suppose $x(t) \in L^p(\mathbb{R})$, $1 < p \leq 2$. Then it holds that $\mathcal{H}(\mathcal{H}(x(t)) = -x(t)$.

Proof. By using Theorem 1.11 twice we get

$$\begin{aligned}\mathcal{F}(\mathcal{H}(\mathcal{H}(x(t)))) &= (-i\,\mathrm{sgn}\,(f)) \cdot \mathcal{F}(\mathcal{H}(x(t))) \\ &= (-i\,\mathrm{sgn}\,(f)) \cdot (-i\,\mathrm{sgn}\,(f)) \cdot \mathcal{F}(x(t)) \\ &= i^2 \cdot (\mathrm{sgn}\,(f))^2 \cdot \mathcal{F}(x(t)) = -\mathcal{F}(x(t)) = \mathcal{F}(-x(t))\end{aligned}$$

Since $\mathcal{H}(\mathcal{H}(x(t)))$ and $-x(t)$ have the same Fourier transform almost everywhere we can conclude that $\mathcal{H}(\mathcal{H}(x(t))) = -x(t)$ and the proof is complete.

In words, applying the Hilbert transform twice on a function $x(t)$ it simply yields $-x(t)$ - the very same function except for a minus sign. Thus, if $\mathcal{H}$ is an Hilbert operator the inverse is $\mathcal{H}^{-1} = -\mathcal{H}$.

Theorem 1.16. (Orthogonality). Suppose $x(t) \in L^2(\mathbb{R})$ is a purely real function with the Fourier transform, $X(f) = \mathcal{F}(x(t))$. Then $x(t)$ and $\mathcal{H}(x(t))$ are orthogonal functions, that is

$$\int_{-\infty}^{\infty} x(t) \cdot \mathcal{H}(x(t))dt = 0$$

Proof. With Parseval's identity and Theorem 1.11 we get

$$\begin{aligned}\int_{-\infty}^{\infty} x(t) \cdot \mathcal{H}(x(t))dt &= \int_{-\infty}^{\infty} X(f) \cdot \overline{(-i\,\mathrm{sgn}\,(f) \cdot X(f))}df \\ &= \int_{-\infty}^{\infty} X(f) \cdot i\,\mathrm{sgn}\,(f) \cdot \overline{X(f)}df \\ &= i\int_{-\infty}^{\infty} \mathrm{sgn}\,(f)|X(f)|^2 df.\end{aligned}$$

Since $x(t)$ is real it means that $|X(f)| = |X(-f)|$ (more about this in section 1.5). Therefore, $\mathrm{sgn}\,(f)|X(f)|^2$ is an odd function since $\mathrm{sgn}\,(f)$ is odd and

$|X(f)|^2$ even. With the symmetric interval of integration, the integral is zero and we have the desired result,

$$\int_{-\infty}^{\infty} x(t) \cdot \mathcal{H}(x(t))dt = 0$$

which means that $x(t)$ and $\mathcal{H}(x(t))$ are orthogonal functions. Proof is complete.

Remark 1.17. The differentiation, inversion and orthogonality properties of the Hilbert transform can be generalized for a broader class of functions.

For the general case, the Hilbert transform offer no shortcuts for calculating the Hilbert transform of products of functions. There is however a special case covered by the so called *Bedrosian's theorem* [4].

Theorem 1.18. (Bedrosian's theorem). Suppose $f(t)$ and $g(t)$ have Fourier transforms $F(f)$ and $G(f)$, respectively, where $F(f) = 0$ for $|f| > a$ with $a > 0$ and $G(f) = 0$ for $|f| < a$. Then

$$\mathcal{H}(f(t)g(t)) = f(t)\mathcal{H}(g(t))$$

Proof. Using the fact that $\mathcal{H}(e^{ist}) = -i\,\mathrm{sgn}\,(s)e^{ist}$ for $s \in \mathbb{R}$ we have

$$\begin{aligned}
\mathcal{H}(f(t)g(t)) =& \mathcal{H}\left(\int_{-\infty}^{\infty} F(u)e^{2\pi iut}du \int_{-\infty}^{\infty} G(v)e^{2\pi ivt}dv\right) \\
=& \mathcal{H}\left(\int_{-\infty}^{\infty} F(u) \int_{-\infty}^{\infty} G(v)e^{2\pi i(u+v)t}dvdu\right) \\
=& \int_{-\infty}^{\infty} F(u)du \int_{-\infty}^{\infty} G(v)\mathcal{H}(e^{2\pi i(u+v)t})dv \\
=& \int_{-\infty}^{\infty} F(u)du \int_{-\infty}^{\infty} G(v)(-i\,\mathrm{sgn}\,(u+v)e^{2\pi i(u+v)t})dv \\
=& -i\int_{-\infty}^{\infty} F(u)e^{2\pi iut}du \left(\int_{-\infty}^{-a} G(v)e^{2\pi ivt}\,\mathrm{sgn}\,(u+v)dv \right. \\
&\left. + \int_{a}^{\infty} G(v)e^{2\pi ivt}\,\mathrm{sgn}\,(u+v)dv\right)
\end{aligned}$$

Make a change of variables in the inner integral, $w = u + v$

$$\begin{aligned}\mathcal{H}(f(t)g(t)) =& -i\int_{-a}^{a} F(u)e^{2\pi iut}du\left(\int_{-\infty}^{-a+u} G(w-u)e^{2\pi i(w-u)t}\operatorname{sgn}(w)dw\right.\\ &\left.+\int_{a+u}^{\infty} G(w-u)e^{2\pi i(w-u)t}\operatorname{sgn}(w)dw\right)\\ =& -i\int_{-a}^{a} F(u)du\left(\int_{-\infty}^{-a+u} G(w-u)e^{2\pi iwt}\operatorname{sgn}(w)dw\right.\\ &\left.+\int_{a+u}^{\infty} G(w-u)e^{2\pi iwt}\operatorname{sgn}(w)dw\right)\end{aligned}$$

For the first integral the integrand is non-zero when $w - u < -a \iff w < -a + u$ and for the second integral the integrand is non-zero when $w - u > a \iff w > a + u$. Since $-a < u < a$ the sign-function in each integral takes only one value and we may simplify

$$\begin{aligned}\mathcal{H}(f(t)g(t)) =& -i\int_{-a}^{a} F(u)du\left(-\int_{-\infty}^{-a+u} G(w-u)e^{2\pi iwt}dw\right.\\ &\left.+\int_{a+u}^{\infty} G(w-u)e^{2\pi iwt}dw\right)\end{aligned}$$

Make another change of variables, let $y = w - u$ which gives

$$\begin{aligned}\mathcal{H}(f(t)g(t)) =& -i\int_{-a}^{a} F(u)du\left(-\int_{-\infty}^{-a} G(y)e^{2\pi i(y+u)t}dy\right.\\ &\left.+\int_{a}^{\infty} G(y)e^{2\pi i(y+u)t}dy\right)\\ =& -i\int_{-a}^{a} F(u)e^{2\pi iut}du\left(-\int_{-\infty}^{-a} G(y)e^{2\pi iyt}dy\right.\\ &\left.+\int_{a}^{\infty} G(y)e^{2\pi iyt}dy\right)\\ =&\int_{-a}^{a} F(u)e^{2\pi iut}du\left(\int_{-\infty}^{-a}(-i\operatorname{sgn}(y))G(y)e^{2\pi iyt}dy\right.\\ &\left.+\int_{a}^{\infty} G(y)(-i\operatorname{sgn}(y))e^{2\pi iyt}dy\right)\\ =&f(t)\left(\int_{-\infty}^{-a} G(y)\mathcal{H}(e^{2\pi iyt})dy+\int_{a}^{\infty} G(y)\mathcal{H}(e^{2\pi iyt})dy\right)\\ =&f(t)\mathcal{H}\left(\int_{-\infty}^{-a} G(y)\mathcal{H}(e^{2\pi iyt})dy+\int_{a}^{\infty} G(y)\mathcal{H}(e^{2\pi iyt})dy\right)\\ =&f(t)\mathcal{H}(g(t))\end{aligned}$$

And the proof is thus complete. □

1.4 Hilbert transform of periodic functions

One can also define the Hilbert transform for periodic functions. Suppose $u(t)$ is a function with period $2T$ - then it can be expressed as a Fourier series

$$u(t) = \sum_{n=-\infty}^{\infty} c_n e^{\pi i n t/T}$$

where each coeffiecent c_n for a fixed n is given by

$$c_n = \frac{1}{2T}\int_{-T}^{T} u(s)e^{-\pi i n s/T} ds$$

With the facts that $\mathcal{H}(c) = 0$ for any constant c and $\mathcal{H}(e^{ist}) = -i\,\mathrm{sgn}\,(s)e^{ist}$ for any real number s we get

$$\begin{aligned}
\mathcal{H}(u(t)) &= \mathcal{H}\left(c_0\right) + \mathcal{H}\left(\sum_{n=1}^{\infty} c_n e^{\pi i n t/T}\right) + \mathcal{H}\left(\sum_{n=-\infty}^{1} c_n e^{\pi i n t/T}\right) \\
&= 0 + \mathcal{H}\left(\sum_{n=1}^{\infty} c_n e^{\pi i n t/T}\right) + \mathcal{H}\left(\sum_{n=1}^{\infty} c_n e^{-\pi i n t/T}\right) \\
&= \sum_{n=1}^{\infty} c_n \mathcal{H}(e^{\pi i n t/T}) + \sum_{n=1}^{\infty} c_n \mathcal{H}(e^{-\pi i n t/T}) \\
&= -i\sum_{n=1}^{\infty} c_n \left(e^{\pi i n t/T} - e^{-\pi i n t/T}\right) \\
&= -\frac{i}{2T}\sum_{n=1}^{\infty}\left(\int_0^T u(s)e^{-\pi i n s/T} ds\right)\left(e^{\pi i n t/T} - e^{-2\pi i n t/T}\right) \\
&= \frac{-i}{2T}\int_{-T}^{T} u(s)\left(\sum_{n=1}^{\infty} e^{\pi i n(t-s)/T} - e^{-\pi i n(t-s)/T}\right) ds \\
&= \frac{1}{2T}\int_{-T}^{T} u(s)\sum_{n=1}^{\infty} 2\sin\left(\frac{\pi n(t-s)}{T}\right) ds \\
&= \frac{1}{2T}\,\mathrm{PV}\int_{-T}^{T} u(s)\cot\left(\pi\frac{t-s}{2T}\right) ds
\end{aligned}$$

The last equality is because of the identity

$$2\sum_{k=1}^{\infty}\sin kx = \cot\left(\frac{x}{2}\right)$$

which is valid in the sense of distributions [7]. Now, the following definition of a periodic Hilbert transform makes sense.

Definition 1.19. (Hilbert Transform of a periodic function) Let $u(t)$ be a periodic function with periodicity $2T$. Then $\mathcal{H}_T(u(t))$ is the periodic Hilbert transform of $u(t)$ given by

$$\mathcal{H}_T(u(t)) = \frac{1}{2T}\,\mathrm{PV}\int_{-T}^{T} u(s)\cot\left(\pi\frac{t-s}{2T}\right)ds$$

Note: If $T \to \infty$ then $u(t)$ has infinite periodicity and is then, obviously, an non-periodic function. What happens to $\mathcal{H}_T$ when $T \to \infty$? Evaluate the limit of the integrand when $T \to \infty$.

$$\begin{aligned}\lim_{T\to\infty}\frac{\cot\left(\pi\frac{t-s}{2T}\right)}{2T} &= \lim_{T\to\infty}\frac{\cos\left(\pi\frac{t-s}{2T}\right)}{\sin\left(\pi\frac{t-s}{2T}\right)}\cdot\frac{1}{2T} = \lim_{y\to 0^+}\frac{\cos y}{\sin y}\cdot\frac{y}{\pi(t-s)}\\ &= \lim_{y\to 0^+}\frac{\cos y}{\pi(t-s)}\cdot\frac{y}{\sin y} = \frac{1}{\pi(t-s)}\cdot 1 = \frac{1}{\pi(t-s)}\end{aligned}$$

Above we made a change of variables, $y = \pi\frac{t-s}{2T}$ and as $T \to \infty$ we get $y \to 0^+$.

$$\begin{aligned}\lim_{T\to\infty}\mathcal{H}_T(u(t)) &= \lim_{T\to\infty}\frac{1}{2T}\,\mathrm{PV}\int_{-T}^{T} u(s)\cot\left(\pi\frac{t-s}{2T}\right)ds\\ &= \mathrm{PV}\int_{-\infty}^{\infty}\frac{1}{\pi(t-s)}u(s)ds\\ &= \frac{1}{\pi}\,\mathrm{PV}\int_{-\infty}^{\infty}\frac{u(s)}{t-s}ds\end{aligned}$$

We recognize this last expression as the Hilbert transform on the real line. Obviously, $\mathcal{H}_T \to \mathcal{H}$ as $T \to \infty$ as should be expected [4].

1.5 Analytic representation of a signal

The Hilbert Transform is widely used in signal processing. The main reason for that is that the Hilbert transform can help creating an analytic representation of real signals [12].

Definition 1.20. (Analytic signal). Let $f(z)$ be an analytic function in the upper half-plane. If $f(z)$ on the real line can be written as

$$f(t) = f(t + 0i) = g(t) + ih(t)$$

where $g(t)$ and $h(t)$ are real-valued functions and a Hilbert transform pair, then $f(t)$ is said to be an analytic signal.

Note that this is one of various ways to define an analytic signal. Not all analytic functions are analytic signals on the real line. One sufficient condition is that $|f(z)| \to 0$ as $|z| \to \infty$, $\operatorname{Im} z > 0$ which is explored in the next theorem.

Theorem 1.21. Let $f(z)$ be an analytic function for $\operatorname{Im} z \geq 0$ that vanishes in the upper half-plane ($|f(z)| \to 0$ as $|z| \to \infty$). Then, on the real line $f(z)$ is an analytic signal and can be written as

$$f(t) = g(t) + i\mathcal{H}(g(t))$$

where $g(t) = \operatorname{Re} f(t + 0i) = u(t, 0)$ and $\mathcal{H}(g(t)) = \operatorname{Im} f(t + 0i) = v(t, 0)$ are real valued functions.

Proof. Define the contour C like in Example 1.6. Then, since $f(z)$ is analytic in the upper half-plane we have according to Cauchy's integral theorem

$$\int_C \frac{f(z)}{t - z} dz = 0$$

since the integrand $f(z)/(t - z)$ is analytical inside the contour. Breaking up the integral in parts we get

$$\int_{L^1_{\epsilon,R} + L^2_{\epsilon,R}} \frac{f(z)}{t - z} dz = -\int_{C^+_R} \frac{f(z)}{t - z} dz - \int_{C_\epsilon} \frac{f(z)}{t - z} dz$$

For the integral over the contour C^+_R we have that

$$\left| \int_{C^+_R} \frac{f(z)}{t - z} dz \right| \leq \frac{\max\limits_{z \in C^+_R} |f(z)|}{|t - R|} \cdot \pi R = \frac{\pi}{|t/R - 1|} \cdot \max_{0 < \varphi < \pi} |f(Re^{i\varphi})|$$

and since $|f(z)| \to 0$ as $|z| \to \infty$, $\operatorname{Im} z > 0$ by assumption it therefore holds that $|f(Re^{i\varphi})| \to 0$ as $R \to \infty$ for $0 < \varphi < \pi$. Thus, as $R \to \infty$

$$\left| \int_{C^+_R} \frac{f(z)}{t - z} dz \right| \to \pi \cdot 0 = 0$$

Looking on the integral over the the contour C_ϵ, from Lemma 1.3, we have that

$$\int_{C_\epsilon} \frac{f(z)}{t - z} dz = i(0 - \pi) \cdot \operatorname*{Res}_{z=t} \frac{f(z)}{t - z} = -i\pi \cdot (-f(t)) = i\pi f(t)$$

Letting $\epsilon \to 0^+$, $R \to \infty$, dividing by π and parameterizing we get

$$\frac{1}{\pi} \operatorname{PV} \int_{-\infty}^{\infty} \frac{f(s)}{t - s} ds = -0 - if(t) = -if(t)$$

We have that $\mathcal{H}(f(t)) = -if(t)$. Since $f(z)$ is a complex and analytic function we can write $f(z) = f(x+iy) = u(x,y) + iv(x,y)$ where $u(x,y)$ and $v(x,y)$ are real valued functions. On the real line ($x = t$, $y = 0$) we therefore get

$$\begin{aligned}\mathcal{H}(f(t)) &= \mathcal{H}(u(t,0) + iv(t,0)) = \mathcal{H}(u(t,0)) + i\mathcal{H}(v(t,0)) = -if(t) \\ &= -i(u(t,0) + iv(t,0)) = v(t,0) - iu(t,0)\end{aligned}$$

Identifying real and imaginary parts we have that $\mathcal{H}(u(t,0)) = v(t,0)$ and $\mathcal{H}(v(t,0)) = -u(t,0)$. Thus, if we let $g(t) = u(t,0)$ we simply have $f(t) = g(t) + i\mathcal{H}(g(t))$ and $f(t)$ is an analytic signal.

Example 1.22. Let $f(z) = e^{i\omega z}$ for $\omega > 0$. Clearly $f(z)$ is analytic for $\operatorname{Im} z \geq 0$ and since $|f(z)| = |e^{i\omega z}| = |e^{i\omega(x+iy)}| = |e^{-\omega y + i\omega x}| = e^{-\omega y} \to 0$ as $|z| \to \infty$, $\operatorname{Im} z > 0$ we have that $f(z)$ vanishes in the upper half-plane. Hence, on the real line $f(z)$ is an analytic signal.

$$f(t) = f(t + i0) = e^{i\omega(t+0i)} = e^{i\omega t} = \cos\omega t + i\sin\omega t$$

In this case we have $g(t) = \cos\omega t$ and as should be expected $\mathcal{H}(g(t)) = \sin\omega t$.

Example 1.23. Let $f(z) = \frac{1}{a-iz}$ for $a > 0$. The function is singular in the point $z = -ia$ and analytic everywhere else. Thus, in the upper half-plane $f(z)$ is analytic and obviously decreases with the same rate as $1/|z|$. Therefore, $f(z)$ represents an analytic signal on the real line and we have that

$$f(t) = f(t + 0i) = \frac{1}{a - it} = \frac{a + it}{a^2 + t^2} = \frac{a}{a^2 + t^2} + i\frac{t}{a^2 + t^2}$$

and we have shown that $\mathcal{H}\left(\frac{a}{a^2+t^2}\right) = \frac{t}{a^2+t^2}$ and $\mathcal{H}\left(\frac{t}{a^2+t^2}\right) = -\frac{a}{a^2+t^2}$ for $a > 0$.

The following theorem we have esentially shown already but because of its usefulness it is worthful to state in a clear manner.

Theorem 1.24. Suppose $f(t)$ is an analytic signal. Then its Hilbert transform is given by $\mathcal{H}(f(t)) = -if(t)$.

Proof. See the proof of Theorem 1.21.

Theorem 1.25. Suppose $f_1(t)$ and $f_2(t)$ are analytic signals. Then it holds that $\mathcal{H}(f_1(t)) \cdot f_2(t) = f_1(t) \cdot \mathcal{H}(f_2(t))$.

Proof. The proof is just a matter of rearranging and using Theorem 1.13.

$$\mathcal{H}(f_1(t)) \cdot f_2(t) = -if_1(t) \cdot f_2(t) = f_1(t) \cdot (-if_2(t)) = f_1(t) \cdot \mathcal{H}(f_2(t))$$

Theorem 1.26. Let $f(t) = g(t) + i\mathcal{H}(g(t))$ be an analytic signal. Then its Fourier transform, $F(f)$, is given by

$$F(f) = (1 + \operatorname{sgn}(f)) \cdot G(f) = \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases}$$

Proof. By using Theorem 1.8 we readily get

$$\begin{aligned} F(f) &= \mathcal{F}(f(t)) = \mathcal{F}(g(t) + i\mathcal{H}(g(t))) = \mathcal{F}(g(t)) + i \cdot \mathcal{F}(\mathcal{H}(g(t))) \\ &= G(f) + i \cdot (-i \operatorname{sgn}(f)) \cdot G(f)) = (1 + \operatorname{sgn}(f)) \cdot G(f) \end{aligned}$$

and the proof is complete. Note that for the sign function we have $\operatorname{sgn}(0) = 0$. □

This theorem is very interesting since the function $g(t)$ is real. Because $g(t)$ is purely real and that conjugation is a linear operation we have that

$$\begin{aligned} \overline{G(f)} &= \overline{\int_{-\infty}^{\infty} g(t) e^{-2\pi i f t} dt} = \int_{-\infty}^{\infty} \overline{g(t) e^{-2\pi i f t}} dt = \int_{-\infty}^{\infty} g(t) \overline{e^{-2\pi i f t}} dt \\ &= \int_{-\infty}^{\infty} g(t) e^{2\pi i f t} dt = \int_{-\infty}^{\infty} g(t) e^{-2\pi i (-f) t} dt = G(-f) \end{aligned}$$

and thus $G(f)$ has Hermitian symmetry since $G(-f) = \overline{G(f)}$. In particular it holds that $|G(f)| = |G(-f)|$ since $|\overline{G(f)}| = |G(f)|$. Consequently, when we consider the amplitude spectrum $|G(f)|$, it is an even function with symmetry around $f = 0$. Because of these symmetries, it is clearly sufficient to consider $G(f)$ for $f \geq 0$. This is one of many advantages with analytic signals - they carry only positive frequency components and discard the negative frequencies since they are superfluous for real signals [12].

Example 1.27. Consider $g(t) = \cos \omega t$ for $\omega > 0$. As has been seen already we can expand $g(t)$ into an analytic signal.

$$f(t) = \cos \omega t + i\mathcal{H}(\cos \omega t) = \cos \omega t + i \sin \omega t = e^{i\omega t}$$

As is known, $G(f) = \mathcal{F}(\cos(\omega t)) = \frac{1}{2}\left(\delta\left(f - \frac{\omega}{2\pi}\right) + \delta\left(f + \frac{\omega}{2\pi}\right)\right)$. As expected, because $g(t)$ is purely real $G(f)$ has Hermitian symmetry. Also, because $G(f)$ is purely real as well, $\overline{G(f)} = G(f)$ and so $G(-f) = G(f)$. For the analytic signal we consequently get the Fourier transform

$$F(f) = \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases} = \begin{cases} \delta\left(f - \frac{\omega}{2\pi}\right) + \delta\left(f + \frac{\omega}{2\pi}\right), & f > 0 \\ 0, & f = 0 \\ 0, & f < 0 \end{cases}$$

Of course, $\delta\left(f-\frac{\omega}{2\pi}\right)+\delta\left(f+\frac{\omega}{2\pi}\right)=\delta\left(f-\frac{\omega}{2\pi}\right)$ for $f>0$. To conclude, we have that $F(f)=\delta\left(f-\frac{\omega}{2\pi}\right)$ and as expected $F(f)=0$ for $f<0$. The term $\delta\left(f+\frac{\omega}{2\pi}\right)$ is discarded but no information is lost since the same information is carried in the term $\delta\left(f-\frac{\omega}{2\pi}\right)$.

To transmit the signal $g(t)$ completely one needs to consider the frequency components in $-\frac{\omega}{2\pi}\leq f\leq\frac{\omega}{2\pi}$, resulting in $\frac{\omega}{2\pi}-\left(-\frac{\omega}{2\pi}\right)=\frac{\omega}{\pi}$ as the smallest possible bandwidth. In contrast, for $f(t)$ is is enough to consider frequency components for $0\leq f\leq\frac{\omega}{2\pi}$ and because $\frac{\omega}{2\pi}-0=\frac{\omega}{2\pi}=\frac{1}{2}\cdot\frac{\omega}{\pi}$ the bandwidth required for complete transmission is halved.

1.6 Boundedness of the Hilbert transform

The Hilbert transform is a bounded linear operator on $L^p(\mathbb{R}), 1<p<\infty$. The following inequality was shown by Riesz (1928) and is known as the "Riesz inequality" [4].

Theorem 1.28. Suppose $f\in L^p(\mathbb{R})$, $1<p<\infty$. Then $\|\mathcal{H}f\|_p\leq C_p\|f\|_p$.

Proof. We will provide a proof of this inequality for $p=2$. From the steps acquired in the proof of Theorem 1.12 we have that

$$\mathcal{H}(x(t))=\mathcal{F}^{-1}\left(-i\operatorname{sgn}(f)\cdot\mathcal{F}(x(t))\right)$$

Taking the L^2-norm and using Parseval's identity twice yields

$$\begin{aligned}\|\mathcal{H}(x(t))\|_2&=\|\mathcal{F}^{-1}\left(-i\operatorname{sgn}(f)\cdot\mathcal{F}(x(t))\right)\|_2=\|-i\operatorname{sgn}(f)\cdot\mathcal{F}(x(t))\|_2\\&=\|\mathcal{F}(x(t))\|_2=\|x(t)\|_2\end{aligned}$$

and therefore $\|\mathcal{H}f\|_2=\|f\|_2$ and thus $\|\mathcal{H}f\|_2\leq C_2\cdot\|f\|_2$ and $C_2=1$.

Remark 1.29. The best constant C_p was found by Pichorides (1972) and is

$$C_p=\begin{cases}\tan\frac{\pi}{2p} & 1<p\leq 2\\ \cot\frac{\pi}{2p} & 2\leq p<\infty.\end{cases}$$

This theorem does not hold for the case $p=1$. A counterexample was given in Remark 1.13 where $x(t)=\chi_{[0,1]}(t)\in L^1(\mathbb{R})$ and $\mathcal{H}(x(t))=\frac{1}{\pi}\ln\left|\frac{t}{t-1}\right|\notin L^1(\mathbb{R})$. However, since it is possible to prove that, given $f(t)\in L^1(\mathbb{R})$, it holds that

$$|t\in\mathbb{R}:|\mathcal{H}(f)(t)|>\lambda|\leq\frac{K\|f\|_1}{\lambda}=\frac{K\|f\|_1}{\lambda^1}$$

for some $K > 0$, the Hilbert operator $\mathcal{H}$ is said to be of weak type of $(1, 1)$ This space is sometimes written as $L^{1,w}(\mathbb{R})$ equipped with the following norm

$$\|f\|_{L^{1,w}} = \sup_{\lambda>0} \lambda \cdot |t \in \mathbb{R} : |f(t)| > \lambda|.$$

The following theorem summarizes the result [4].

Theorem 1.30. Suppose $f(t) \in L^1(\mathbb{R})$. Then $\mathcal{H}(f(t)) \in L^{1,w}(\mathbb{R})$.

For completeness, let us take a look at the multidimensional Hilbert transform and an inequality due to Alberto Calderón and Antoni Zygmund that implies boundedness. However, first a definition of the n-dimensional Hilbert transform is needed. It is very straightforward [5].

Definition 1.31. The n-dimensional Hilbert transform of $f(t) = f(t_1, ..., t_n)$ is given by

$$\mathcal{H}_n(f(t)) = \lim_{\epsilon_1 \to 0^+} ... \lim_{\epsilon_n \to 0^+} \int_{D_1} ... \int_{D_n} f(s_1, ..., s_n) \prod_{j=1}^{n} \frac{1}{(x_j - s_j)} ds_1 ... ds_n$$

where D_k is given by $|x_k - s_k| > \epsilon_k$.

Remark 1.32. Note that the Cauchy Principal Value is applied on each and everyone of the integrals. For the case $n = 1$, we acquire the "normal", one-dimensional Hilbert transform.

For the n-dimensional Hilbert transform a result like Riesz inequality is valid. This inequality is known as the Calderón-Zygmund inequality.

Theorem 1.33. (Calderón-Zygmund inequality). Let $f(t) \in L^p(\mathbb{E}^n)$ with $1 < p < \infty$. Then

$$\|\mathcal{H}_n f\|_p \leq C_{p,n} \|f\|_p$$

where the constant $C_{p,n}$ is only dependent on n and p.

By this theorem it follows that $\mathcal{H}_n$ is a bounded linear operator on $L^p(\mathbb{R}^n)$ for $1 < p < \infty$.

1.7 Harmonic analysis

The Hilbert transform is a particular case of the so called singular integral operators studied by Calderón and Zygmund. This is closely related to the Dirichlet problem. This problem will be briefly discussed in this section.

Definition 1.34. (Dirichlet problem). Suppose we are looking for a harmonic function ϕ that is continuous and harmonic on a domain Ω and $\phi = f$ on $\partial\Omega$ where f is continuous. This problem is called Dirichlet problem.

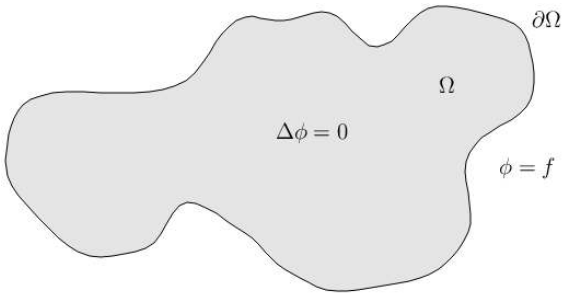

Figure 7: A sketch of the Dirichlet problem for some region Ω.

If Ω is the upper half-plane and $f(t)$ is the function on the real line, we have seen that the Hilbert transform is a helpful tool to expand $f(t)$ to an analytic function $F(t)$ given by

$$F(t) = f(t) + i\mathcal{H}(f(t))$$

where $F(t)$ is an analytic function on the real line, that is, $F(z)$ is an analytic function and $F(t) = F(t + 0i)$. Since $F(z)$ is analytic, it can be written as

$$F(z) = F(t + iy) = \phi(t, y) + i\psi(t, y)$$

where ϕ, ψ are harmonic functions. If $f(t) \in L^2(\mathbb{R})$ as $y \to 0$, $\phi(t, y) \to f(t)$ almost everywhere. The question is now, how can an harmonic function $\phi(x, y)$ be derived from $f(t)$ if Ω is the upper half-plane?

Suppose Ω is the upper half-plane and $f = \phi + i\psi$ is analytic in Ω. Also, suppose $C = C_R^+ + L_R$ is a positively contour in the upper half-plane (given in Figure 8). Then, if z is a point in the upper half-plane, we have according to Cauchy's integral formula

$$f(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{\zeta - z} d\zeta.$$

Furthermore, because $\overline{z}$ is outside the contour, it holds that

$$\frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta-\overline{z}}d\zeta = 0$$

thus yielding

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta - z}d\zeta - \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta-\overline{z}}d\zeta = \frac{1}{2\pi i}\int_C \frac{f(\zeta)(z-\overline{z})}{(\zeta-z)(\zeta-\overline{z})}d\zeta.$$

Note that $z - \overline{z} = 2i\,\mathrm{Im}\, z$. Now, suppose $|f(z)| \leq M$, then on C_R^+ we have

$$\begin{aligned}\left|\frac{1}{2\pi i}\int_{C_R^+} \frac{f(\zeta)(z-\overline{z})}{(\zeta-z)(\zeta-\overline{z})}d\zeta\right| &= \frac{\mathrm{Im}\, z}{\pi}\cdot\left|\int_{C_R^+}\frac{f(\zeta)}{(\zeta-z)(\zeta-\overline{z})}d\zeta\right| \\ &\leq \frac{\mathrm{Im}\, z}{\pi}\frac{M}{(R-|z|)^2}\cdot \pi R\end{aligned}$$

Thus, letting $R\to\infty$ and $z = x+iy$ yields

$$f(z) = f(x+iy) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(t)\cdot 2i\,\mathrm{Im}\, z}{(t-z)(t-\overline{z})}dt = \frac{y}{\pi}\int_{-\infty}^{\infty}\frac{f(t)}{(t-x)^2+y^2}dt$$

since the integral vanishes over C_R^+. Furthermore

$$f(x+iy) = \phi(x,y) + i\psi(x,y) = \frac{y}{\pi}\int_{-\infty}^{\infty}\frac{\phi(t,0)+i\psi(t,0)}{(t-x)^2+y^2}dt.$$

By taking real parts of both sides we have that

$$\phi(x,y) = \frac{y}{\pi}\int_{-\infty}^{\infty}\frac{\phi(t,0)}{(t-x)^2+y^2}dt$$

and in the integral the so called Poisson kernel is present. This expression is known as the Poisson integral formula and it may be further generalized. This generalization is found in Theorem 1.35 and is the solution of the Dirichlet problem in the upper half-plane.

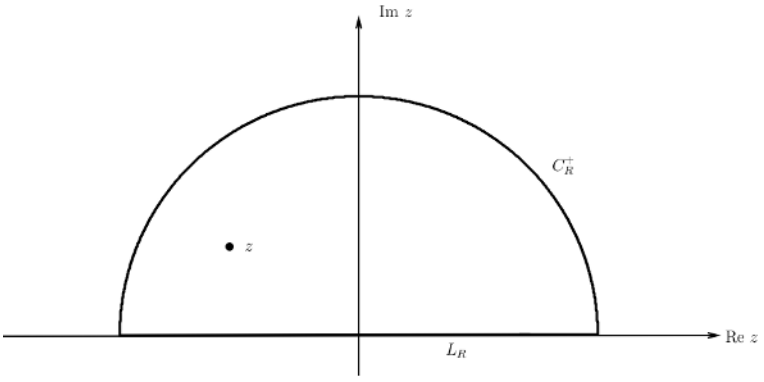

Figure 8: The contour $C = C_R^+ + L_R$.

Theorem 1.35. The solution of a Dirichlet problem in the upper half-plane where $\phi(x, y)$ attains the values $f(x)$ on the real line is given by

$$\phi(x, y) = \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{f(s)}{(t-s)^2 + y^2} ds$$

2 Hilbert transform and electrocardiogram

In the field of medicine, one of the most important tool is a so called *electrocardiaogram*, abbreviated ECG (and sometimes EKG which is an abbreviation of the German word *Elektrokardiogramm*, so ECG is a loanword). An ECG is simply a recording over the electrical activity in a person's heart over time. Naturally, because the heart beats in a periodic manner, this recording will have a clear waveform. The ECG is a very helpful tool in diagnosing of heart diseases and irregularites like arrhythmia. While a skillful physician is able to single-handedly interpret an ECG and find various possible heart defects, computers are an invaluable resource in helping to interpret ECG:s [1]. This is especially the case when the electrocardiography is carried out in real time and serious defects and irregularities must be detected and dealt with swiftly. Of course, algorithms are needed to process and interpret the ECG. For this purpose, the Hilbert transform and its close relationship with analytic signals is a very useful tool. In this section we will examine how the Hilbert transform may be used in this very respect.

2.1 Basics of electrocardiography

This section will focus on providing a basic understanding of electrocardiography. That is, what an ECG actually looks like and how it should be interpreted. Note that this is merely an overview of the medical aspects.

The general ECG-signal, recorded from a heart with normal functionality, will exhibit periodicity. Each cycle (representing a heartbeat) can be divided into unique and distinguishable segments. In Figure 9 we can see a simplified sketch of these segments that together represent a single heartbeat.

This part we will mainly focus on is the so called "QRS complex"-part of the wave. It is corresponding to the depolarization of the right and left ventricles of the heart. Various different heart defects can be detected by looking at the QRS complex. For example, if the QRS complex is too wide (duration is too long) it can suggest problems with the heart's conduction system. Also, too low amplitude of the QRS complex may suggest pericardial effusion or infiltrative myocardial disease while too high amplitude could mean left ventricular hypertrophy. Thus, locating and extractring all QRS complex-parts of an ECG-wave is an essential problem in electrocardiography [2], [15], [13].

2.2 Hilbert transform of the QRS complex

Consider the main part of the QRS complex as clearly illustrated in Figure 9. Looking at its two turning points points, it actually resembles a "slightly" deformed sine wave [2]. This can be very well used in our analysis of the QRS complex along with the Hilbert transform and analytic signals.

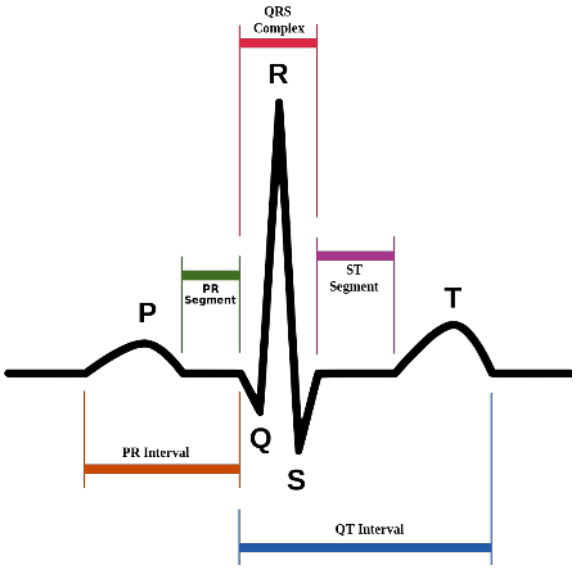

Figure 9: Segments of a single heartbeat cycle from a sketch of an ECG-wave recorded of a normally functioning heart. The QRS complex is the most distinguishable segment and will be our main focus in this section.

Start by making the definition $x(t) = \sin(t)$. As has been seen earlier we know that $\mathcal{H}(\sin(t)) = -\cos(t)$. By this, we may construct an analytic signal.

$$z(t) = x(t) + i\mathcal{H}(x(t)) = \sin(t) + i\mathcal{H}(\sin(t)) = \sin(t) - i\cos(t) = -ie^{it}$$

Clearly, $z(t)$ is an analytic signal and it is also a parameterization of the unit circle in the complex plane. As can be seen in Figure 10 and Figure 11 this is also the case when we consider the situation numerically.

Now, to better imitate the QRS complex we can modify the sine wave slightly by deforming it. Calculations are the same as before but now $u(t)$ and $\mathcal{H}(u(t))$ are as in Figure 12. Also, the analytical signal $w(t) = u(t) + i\mathcal{H}(u(t))$ is constructed. Clearly, $u(t)$ in this plot resembles the QRS complex given in Figure 9 more than a pure sine wave provided by $x(t)$. From the the corresponding plot of $u(t)$ against $\mathcal{H}(u(t))$ seeen in Figure 13 some conclusions can be drawn. As one would expect, since $u(t)$ is no longer a pure sine wave the corresponding plot is not a circle, but rather a deformed circle. Note that the graph of this plot is still closed and encloses the origin.

To conclude, if $z(t) = x(t) + i\mathcal{H}(x(t))$ is an analytical expansion of $x(t)$ where $x(t)$ is a QRS complex graph - then the plot of $(x(t), \mathcal{H}(x(t)))$ will be an enclosed loop around the origin. Of course, this requires that $x(t)$ is symmetric around the x-axis and for an actual ECG to have mean 0 [V]. If this is not the case, it may be fixed by removing the non-zero mean of the ECG-wave and

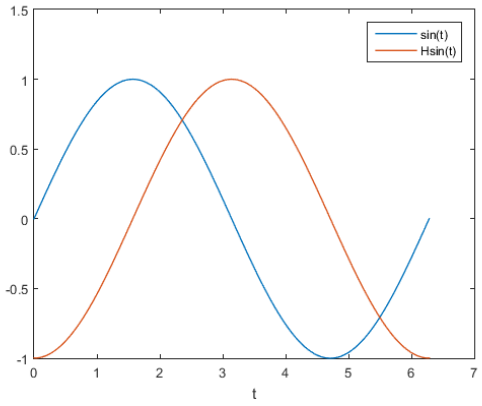

Figure 10: Plot of $x(t) = \sin(t)$ *and its Hilbert transform,* $\mathcal{H}(x(t)) = -\cos(t)$ *in the interval* $0 \leq t \leq 2\pi$*. The Hilbert transform of* $x(t) = \sin(t)$ *has been calculated numerically.*

consequently achieving desired symmetry.

Before examining some datasets from real ECG:s it is interesting to examine what orientation the closed loop of an QRS complex expanded into an analytic signal has. Of course, we are working under the assumption that the QRS complex wave is some deformed sine wave. Plainly deforming the wave will not change the orientation of the graph. Thus, it is perfectly sufficient to examine the simple sine wave that we looked at initially.

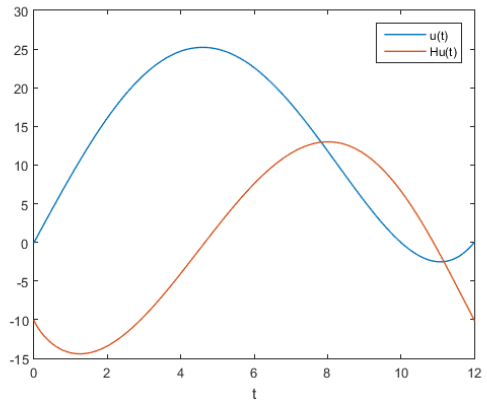

Figure 12: Plot of $u(t)$ *and* $\mathcal{H}(u(t))$*. Here* $u(t)$ *is a slightly deformed sine wave generated numerically in the interval* $0 \leq t \leq 12$*. Its Hilbert transform,* $\mathcal{H}(u(t))$*, was also calculated numerically.*

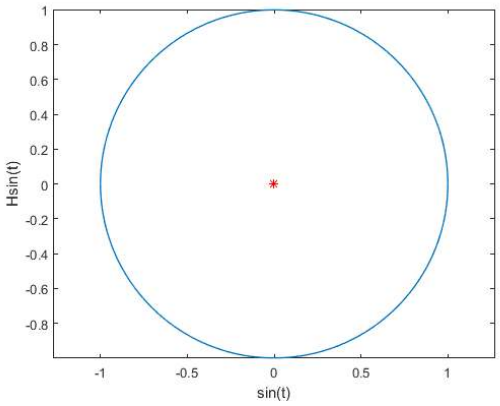

Figure 11: Plot of $x(t) = \sin(t)$ *against* $\mathcal{H}(x(t)) = -\cos(t)$ *with the parameter interval* $0 \leq t \leq 2\pi$. *As expected from theory, when* $x(t) = \sin(t)$ *is plotted against* $\mathcal{H}(x(t)) = -\cos(t)$ *the result is the unit circle. The Hilbert transform of* $x(t) = \sin(t)$ *has been calculated numerically.*

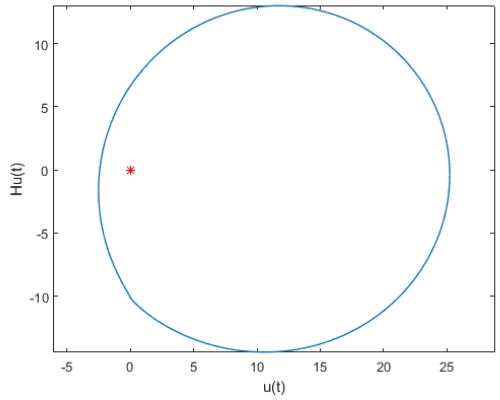

Figure 13: Plot of $u(t)$ *against* $\mathcal{H}(u(t))$ *with the parameter interval* $0 \leq t \leq 12$. *They are both numerically produced and as can be seen in the plot they now represent a somewhat deformed circle.*

$$z(t) = \sin(t) - i\cos(t) = -ie^{it}$$

As before, the parameter interval is $0 \leq t \leq 2\pi$ and by simple calculation we get $z(0) = -i, z(\pi/2) = 1, z(\pi) = i, z(3\pi/2) = -1$ and $z(2\pi) = -i$. Clearly the orentation of the closed curve is counter-clockwise which is also illustrated in Figure 14 with juxtaposed plots.

2.3 Analysis of real ECG-signals

Now, we are provided a real ECG-signal from *PhysioNet* [3]. It is a renowned database which holds a large collection of recorded physiologic signals.

Make the definition $ECG(t)$ as our real ECG-signal. In Figure 15 a plot of $ECG(t)$ is readily provided. We see that $ECG(t)$ has total duration of 5 seconds ($0 \leq t \leq 5$, t [s]). Comparing the sketch of a heartbeat cycle given in Figure 9 with the graph of $ECG(t)$, it is very easy to distinguish the QRS complex-segments in $ECG(t)$. Visibly, there are exactly 5 relatively high and distinguishable peaks in $ECG(t)$. Each one of said peaks represent a QRS complex. Notably, after each QRS complex it is also easy to detect the T-waves which are represented by fairly high peaks. The P-waves are a bit more difficult to detect but still possible if one looks closely before each QRS complex.

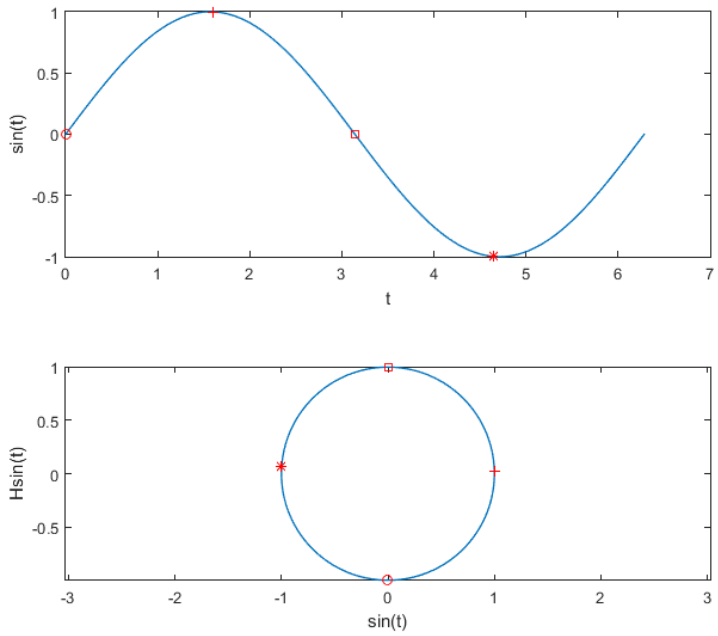

Figure 14: Plot of $x(t) = \sin(t)$ *and a plot of* $\sin(t)$ *against* $\mathcal{H}(\sin(t))$ *in the parameter interval* $0 \leq t \leq 2\pi$. *These two plots are juxtaposed in order to confirm the orentation of the closed (circular) loop. This very figure confirms that the orentation is indeed counter-clockwise which was analytically derived.*

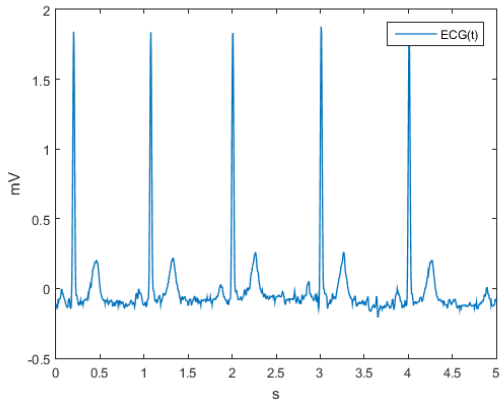

Figure 15: Plot of $ECG(t)$ in the interval $0 \leq t \leq 5$ where t is in seconds. There are 5 sets of QRS complex visible in the plot, represented by the highest peaks. The second highest peak which appear after a QRS complex is a T-wave. The third highest peak which appear before a QRS complex is a P-wave.

Numerically it is possible to calculate the Hilbert transform of $ECG(t)$. In Figure 16 the graph of $\mathcal{H}(ECG(t))$ is plotted in the time domain together with $ECG(t)$. Interestingly, close to the large peaks of $ECG(t)$, the Hilbert transform $\mathcal{H}(ECG(t))$ seems to behave like $1/t$. This is because the large peaks in $ECG(t)$ do resemble the Dirac delta function. While these peaks are of course not true Dirac delta functions (remember, the Dirac delta function is a distribution), they can still be approximated by $f_n(t) = \frac{1}{2}\sqrt{\frac{n}{\pi}} \cdot e^{-nt^2/4}$ which is a sequence of functions. As $n \to \infty$, $f_n(t) \to \delta(t)$ in the sense of distributions. Numerically, as can be seen in Figure 17, as n is increasing $f_n(t)$ is starting to resemble a true pulse. Furthermore, in Figure 18 there is a plot of $\mathcal{H}(f_n(t))$ for the same values for n. Likewise, as n is increasing, $\mathcal{H}(f_n(t))$ gradually tends to something in the likes of $1/t$. This is no accident because in the sense of distributions we according to [4]

$$\mathcal{H}(\delta(t)) = \frac{1}{\pi}\,\mathrm{PV}\int_{-\infty}^{\infty}\frac{\delta(s)}{t-s}ds = \frac{1}{\pi}\cdot\frac{1}{t} = \frac{1}{\pi t}$$

which explains the apperance of $\mathcal{H}(ECG(t))$ around the high peaks.

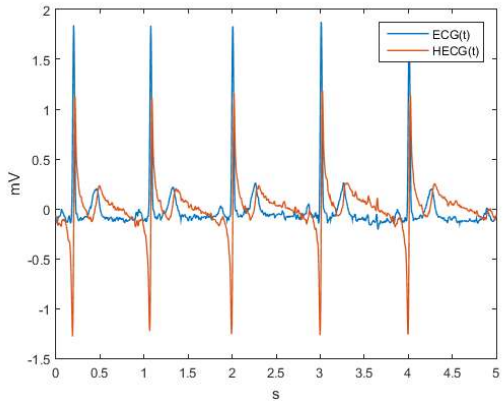

Figure 16: Plot of $ECG(t)$ and $\mathcal{H}(ECG(t))$ in the interval $0 \leq t \leq 5$. Close to the peaks, $\mathcal{H}(ECG(t))$ to some degree resembles the function $1/t$.

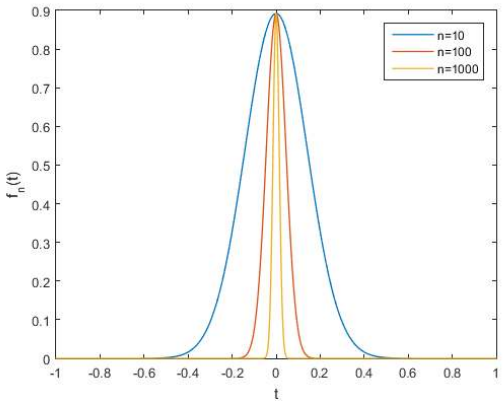

Figure 17: Plot of the sequence of functions defined by $f_n(t) = \frac{1}{2}\sqrt{\frac{n}{\pi}} \cdot e^{-nt^2/4}$ in the interval $-1 \leq t \leq 1$ for $n = 10, 100, 1000$. As n is increasing, $f_n(t)$ is starting to appear like a pulse. Analytically, $f_n(t) \to \delta(t)$ in the sense of distributions as $n \to \infty$.

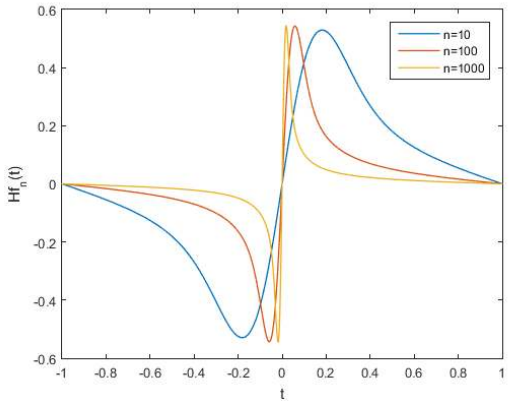

Figure 18: Plot of $\mathcal{H}(f_n(t)) = \mathcal{H}\left(\frac{1}{2}\sqrt{\frac{n}{\pi}} \cdot e^{-nt^2/4}\right)$ *in the interval* $-1 \leq t \leq 1$ *for* $n = 10, 100, 1000$. *The Hilbert transform for each* $f_n(t)$ *has been calculated numerically. As* n *increases,* $\mathcal{H}(f_n(t))$ *begins to resemble* $1/t$.

To continue, as we established in the preceding section, the QRS complex imitates a deformed sine wave. Can this be used for the whole $ECG(t)$ for dection of peaks that are QRS complex? Given $ECG(t)$ and $\mathcal{H}(ECG(t))$ we can expand the ECG signal into an analytic signal.

$$z(t) = ECG(t) + i\mathcal{H}(ECG(t))$$

A plot of $z(t)$ against $\mathcal{H}(ECG(t))$ is provided in Figure 19. By close inspection of this plot it is possible to detect 5 main loops enclosing the origin. By our theory and because the Hilbert transform is a linear operator, each main loop should correspond to exactly one QRS complex. If we can show that this is the case, we have a valuable tool of detecting a QRS complex.

Let us look at a single QRS complex. To do this, focus on the first QRS complex of the ECG, that is, $ECG(t)$ for $0 \leq t \leq 1$. This piece of the ECG can be seen in Figure 20 and its corresponding plot with $ECG(t)$ against $\mathcal{H}(ECG(t))$ for the parameter interval $0 \leq t \leq 1$ is provided in Figure 20. As can be seen in this plot, we now only have one main loop, agreeing well with the fact that inside said interval $ECG(t)$ only carry one QRS complex. Still, we have some "garbage" and a smaller loop close to the origin. We need to show that this does not belong to the QRS complex.

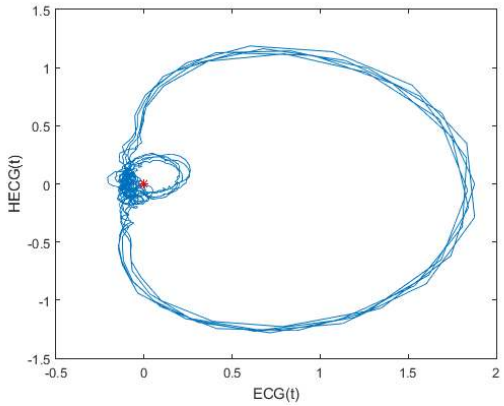

Figure 19: Plot of $ECG(t)$ against $\mathcal{H}(ECG(t))$ in the parameter interval $0 \leq t \leq 5$. Looking very closely at the plot, one can detect 5 main loops.

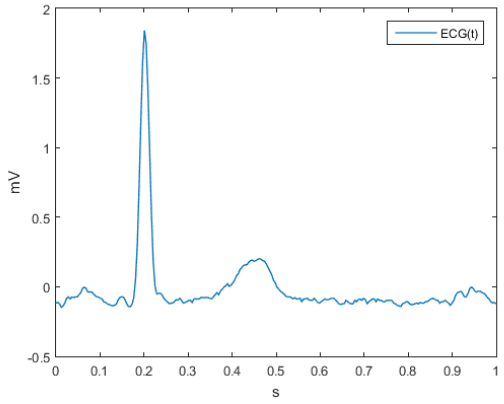

Figure 20: Plot of $ECG(t)$ in the interval $0 \leq t \leq 1$. Exactly one peak representing a QRS complex is present inside this interval.

In order to show this, we define the following for $0 \leq t \leq 1$

$$QRS(s) = ECG(t) \cdot \chi_{[0.18,0.38]}(t)$$

then we cut out the QRS complex from the graph in Figure 20 and the result can be seen in Figure 22 which is the graph of $QRS(t)$. We see clearly that $QRS(t)$ indeed resembles a deformed sine wave, here with some minor numerical disturbances. Now, expand $QRS(t)$ into an anlytic signal

$$z_{QRS}(t) = QRS(t) + i\mathcal{H}(QRS(t))$$

and also make a definition of a residual signal $r(t)$, that is, everything of $ECG(t)$ in the interval $0 \leq t \leq 1$ that is not the QRS complex.

$$r(t) = ECG(t) \cdot \chi_{[0,1]} - QRS(t)$$

Firstly, consider the analytic signal $z_{QRS}(t)$. A plot of $QRS(t)$ against $\mathcal{H}(QRS(t))$, which can be seen in Figure 23 effectively shows that $z_{QRS}(t)$ by itself generates a large main loop around the origin. However, we still need to show that the numerical Hilbert transform we are using "exhibits" linear behavior. If it is linear, then we should have that

$$\begin{aligned} 0 &= |\mathcal{H}(QRS(t)) - \mathcal{H}(QRS(t))| = \left|\mathcal{H}(QRS(t)) - \mathcal{H}(ECG(t) \cdot \chi_{[0,1]} - r(t))\right| \\ &= \left|\mathcal{H}\left(QRS(t)\right) + \mathcal{H}\left(r(t)\right) - \mathcal{H}\left(ECG(t) \cdot \chi_{[0,1]}\right)\right| \end{aligned}$$

In Figure 24 we see that this error is very small and close to zero throughout the whole interval. It is safe to say that our numerical Hilbert transform is linear and therefore the theory holds.

Consequently, given an ECG-signal, $ECG(t)$, and the objective to detect one or several QRS complex-waves the following method is valid.

1. If necessary, filter the ECG-signal to remove noise and to get a smoother curve. For some ECG-signals this method works better if the curve is a bit smooth.

2. Remove any offset from $ECG(t)$ to make sure that the ECG-wave has mean zero. If this is not the case, this can be done numerically by subtracting the non-zero mean from $ECG(t)$ (sometimes known as "detrending").

3. Expand the detrended version of $ECG(t)$ into an analytic signal using the Hilbert transform. First calculate $\mathcal{H}(ECG(t))$ and then make the definition $z(t) = ECG(t) + i\mathcal{H}(ECG(t))$.

4. Analyze $z(t)$ along the parametric interval. As shown, a QRS complex wave in the time domain creates a large loop enclosing the origin in a plot of $ECG(t)$ against $\mathcal{H}(ECG(t))$ with counter-clockwise orentation as t increases. It is also possible to use the zero crossings of the real- and imaginary axis since $z_{QRS}(t)$ completely encloses the origin and does not go through it.

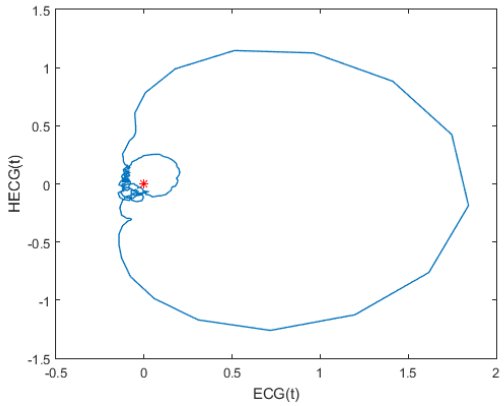

Figure 21: Plot of $ECG(t)$ against $\mathcal{H}(ECG(t))$ in the parameter interval $0 \leq t \leq 1$. There is only one main loop in the plot, still enclosing the origin.

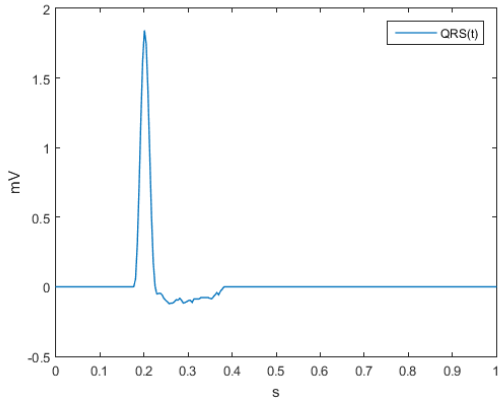

Figure 22: Plot of $QRS(t) = ECG(t) \cdot \chi_{[0.18,0.38]}(t)$ in the interval $0 \leq t \leq 1$ where the QPRS complex has been cut out and the rest of the $ECG(t)$ is set to zero. As can be seen in the plot, the QRS-function has the apperance of a derfomed sine wave with some minor numerical disturbances.

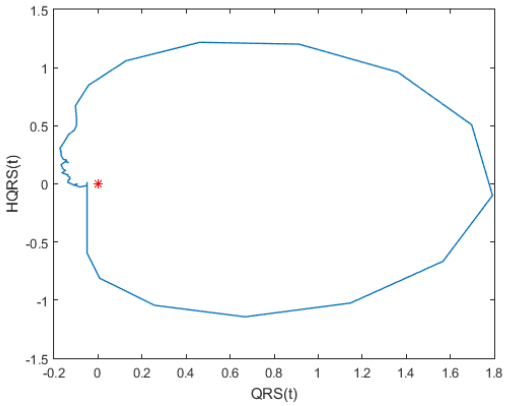

Figure 23: Plot of $QRS(t)$ against $\mathcal{H}(QRS(t))$ in the parameter interval $0 \leq t \leq 1$. The only thing visible in the plot is a large main loop. This means that the analytic expansion of $QRS(t)$ itself corresponds to a main loop enclosing the origin.

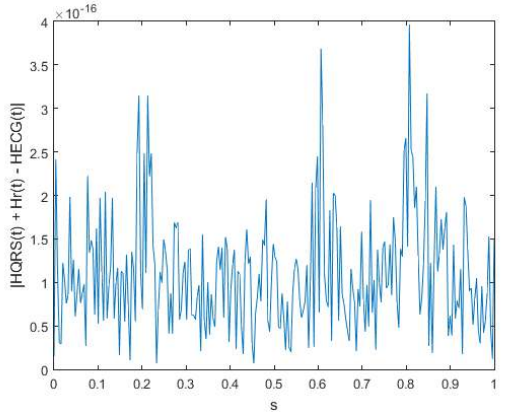

Figure 24: A plot over the error $\left|\mathcal{H}\left(QRS(t)\right) + \mathcal{H}\left(r(t)\right) - \mathcal{H}\left(ECG(t) \cdot \chi_{[0,1]}\right)\right|$. Since the error is very small, the numerical Hilbert transform exhibits linearity.

2.4 Limitation of the method

When does this method not work for detection of the QRS complex? Consider some other ECG-signal, $ECG(t)$, as in Figure 25. The difference here is that the altitude of the QRS complex is considerably lower. As before,we expand $ECG(t)$ into an analyhtic signal

$$z(t) = ECG(t) + i\mathcal{H}(ECG(t))$$

and in Figure 26 there is a plot of $ECG(t)$ against $\mathcal{H}(ECG(t))$. In this plot we see that there is not *one* distinguishable main loop anymore. Therefore, the algorithm will possibly fail to detect the QRS complex or the T-wave will be interpreted as an QRS complex.

To conclude, for the method to work the height QRS complex needs to be distinguishably higher than the rest of the ECG-signal.

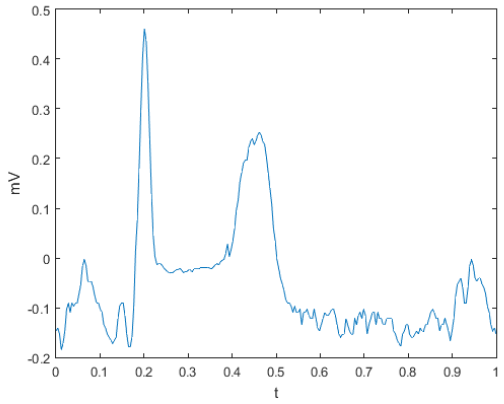

Figure 25: A slightly modified ECG-signal with a considerably lower QRS complex.

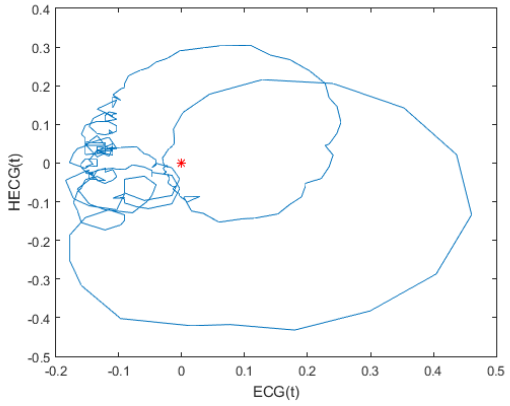

Figure 26: A slightly modified ECG-signal with a considerably lower QRS complex.

3 The Hilbert-Huang transform

A time series is a collection of data points for a fluctuating variable that has been sampled sequentially in time. We will denote a time series by $X(t)$ which simply denotes the value of the time series at a time t. The time series that will be considered here will be continuous ones, meaning that they have been sampled continuously in time [5] .

There are several types of tools available to analyze a time series, most famously is perhaps the Fourier transform. However, if the data in question comes from a nonstationary or nonlinear process there are limitations and the Fourier transform may not be a suitable tool. For these time series the *Hilbert-Huang transform* (HHT) is useful. The method is developed entirely in an emperical sense and is lacking an underlying mathematical framework.

Before presenting the method and an algorithm to carry out said method we need, however, a definition and the introduction of some new concepts [5].

Definition 3.1. (IMF). An Intrinsic Mode Function (IMF) is a continous, real-valued function that satisfies the following two conditions

(i) The number of local extrema and zero crossings must be equal or differ by most one

(ii) The mean value of the two envelopes curves formed by the extremas (local minima and maxima, respectively) should be zero at any time

Remark 3.2. Certainly, it begs the question, what is the "envelope formed by the extremas"? This is simply some smooth curve connecting the local maxima or local minima respectively. An envelope curve for the maxima is called $e_u(t)$ and for the minima $e_l(t)$. Usually, cubic spline interpolation is used to construct these two envelope curves.

The approach and idea of HHT is based on *Emperical Mode Decomposition* (EMD) where the time series is decomposed into IMFs. The extraction of an IMF from a time series is refered to as *sifting*. Each IMF will represent an intrinsic oscillation that is present in said time series [5], [17], [16].

3.1 An algorithm for decomposition

Consider a continuous time series $X(t)$. An algorithm to carry out an EMD for the time series is as follows. Let $h_{10} = X(t)$.

(i) Locate each local maximum of $X(t)$ and construct an envelope curve connecting all maxima using cubic spline interpolation.

(ii) Locate each local minimum of $X(t)$ and construct an envelope curve connecting all minimas using cubic spline interpolation.

(iii) Construct a mean curve of these two envelope curves and call it $m_{11}(t)$. The first index refers to the particular IMF under construction and the second index tells us which iteration (extracting one IMF may require several iterations) we are in for said IMF.

(iv) Calculate $h_{11}(t) = h_{10}(t) - m_{11}(t) = X(t) - m_{11}(t)$. This should be close to an IMF but it may need some refinement.

(v) Treat $h_{11}(t)$ as input data and repeat steps (i), (ii) and (iii) with $h_{12}(t) = h_{11}(t) - m_{12}(t)$.

(vi) Repeat k times with an appropriate stopping criterion finally ending up with $h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t)$.

(vii) Finally $h_{1k}(t)$ is subtracted from $X(t)$ to give $h_{20}(t)$. Process then restarts with the dataset $h_{20}(t)$.

(viii) Set $c_i(t) = h_{ik}(t)$.

For step (vi) we need a stopping criterion. The most commonly used criterion is by looking at the sum of the difference, SD, derived from two consecutive iterations. That is, $h_{n(k-1)}(t)$ and $h_{nk}(t)$ [10]. Also, since the algorithm is carried out numerically, suppose the time series is in a finite time interval, $0 \leq t \leq T$.

$$\mathrm{SD}_n = \sum_{t=0}^{T} \left(\frac{|h_{n(k-1)}(t) - h_{nk}(t)|^2}{h_{n(k-1)}^2(t)} \right)$$

Empirically derived, an appropriate stopping criterion is then given by $\mathrm{SD}_n < \epsilon$ where ϵ is some number between 0.2 and 0.3. But when should the overall sifting process stop? We may calculate

$$r_n(t) = X(t) - \sum_{i=1}^{n} c_i(t)$$

where $r_n(t)$ is a residual term. Naturally, when $r_n(t)$ has become monotonic or constant, no more IMFs may be extracted from $X(t)$ and the sifting should be stopped. Therefore, when $r_n(t)$ is either monotonic or constant the sifting process is complete. The final decomposition of $X(t)$ has thus been obtained and is obviously given by the expression

$$X(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)$$

where $r_n(t)$ is, again, a residual term. Since $r_n(t)$ is either monotonic or constant [10] one of the useful properties of the HHT is that it seperates trends or means from the harmonics of a time series. Thus, when examining a time series there is no inherent requirement of removing linear trends or unwanted offsets.

Example 3.3. We will consider one step of this algorithm where the first IMF, $c_1(t)$, is extracted from a function $X(t)$ which can be seen in Figure 27. Put $h_{10}(t) = X(t)$. In Figure 28 we see that $m_{11}(t)$, which is is the mean curve of the two envelope curves is not zero. Therefore, $h_{10}(t)$ is not an IMF itself and $h_{11}(t) = h_{10}(t) - m_{11}(t)$ is calculated.

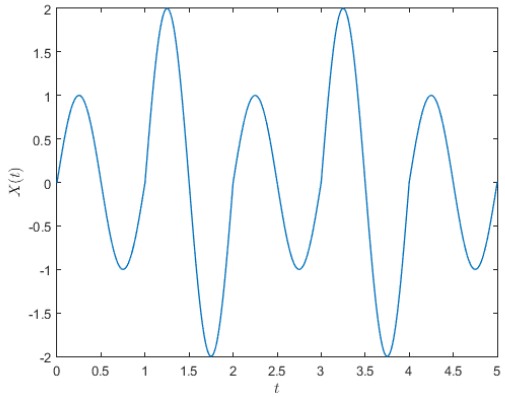

Figure 27: Plot of function $X(t)$ from which we will extract $c_1(t)$.

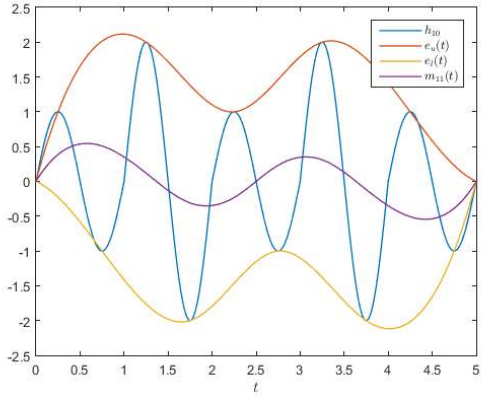

Figure 28: Plot of $h_{10}(t)$ with its envelope curves and mean curve.

Visibly, in Figure 29 we can see that $h_{11}(t)$ is much closer to an IMF compared to $h_{10}(t)$. In Figure 30 - 31 the above steps are repated in order to refine $h_{11}(t)$ and make it look like a true IMF.

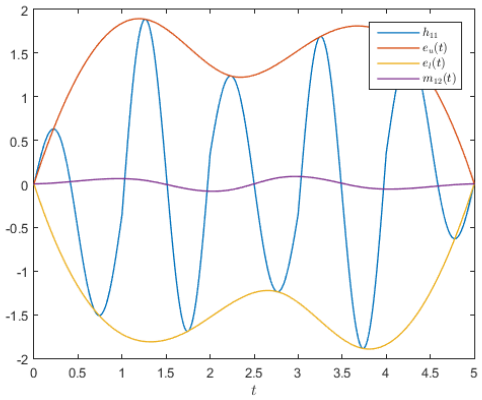

Figure 29: Plot of $h_{11}(t)$ with its envelope curves and mean curve.

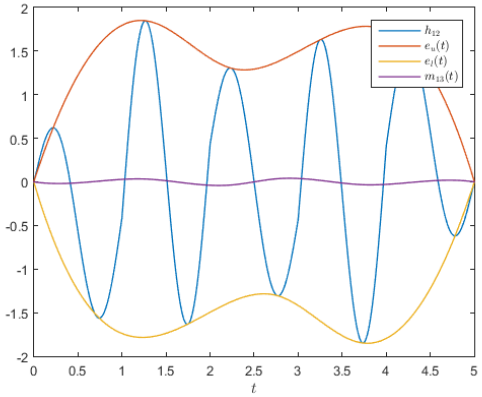

Figure 30: Plot of $h_{12}(t)$ with its envelope curves and mean curve.

In Figure 31 we se that $m_14(t)$ is very close to zero and for these reasons we may put $c_1(t) = h_{13}(t)$ (of course, further refinement is possible). Lastly, in Figure 32 we have the result, $c_1(t)$, juxtaposed with $X(t)$. Also in this figure, there is $X(t) - c_1(t)$. From $X(t) - c_1(t)$ the next IMF, $c_2(t)$ is to be extracted.

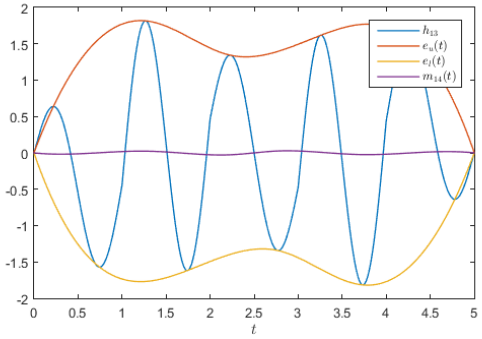

Figure 31: Plot of $h_{13}(t)$ with its envelope curves and mean curve.

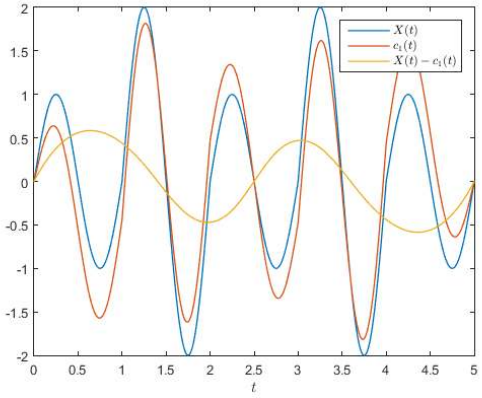

Figure 32: Plot of $X(t)$, $c_1(t)$ and $X(t) - c_1(t)$. Extraction of the next IMF, $c_2(t)$, will continue from $X(t) - c_1(t)$ as the new "input".

3.2 Interpretation of the decomposition

Let $x(t)$ be a real-valued, continuous signal. If it is fairly well-behaved, we may expand $x(t)$ into an analytic signal.

$$z(t) = x(t) + i\mathcal{H}(x(t))$$

Since $z(t)$ is a complex number it may be written in, instead of rectangular form, polar form $z(t) = a(t)e^{i\varphi(t)}$ where $a(t) = |z(t)|$ and $\varphi(t) = \arg(z(t))$. Now, assume $X(t)$ is a continuous time series that has been decomposed using the HHT. Expand $X(t)$ into an analytical signal using the Hilbert transform.

$$Z(t) = X(t) + i\mathcal{H}(X(t)).$$

Using the decomposition, we disregard the residual term, $r_n(t)$. Either $r_n(t)$ is a constant and then $\mathcal{H}(r_n(t)) = \mathcal{H}(r) = 0$, or it is monotonic. If $r_n(t)$ is monotonic then it can potentially overpower the harmonics and should therefore be left out.

Define $\omega(t) = \frac{d}{dt}\varphi(t)$ where $\omega(t)$ is angular frequency. From this definition it also follows that $\omega(t) = 2\pi f(t)$ where $f(t)$ is called frequency. Now, from the analytical expansion of $X(t)$ along with the decomposition of $X(t)$ through HHT we get

$$\begin{aligned} Z(t) &= \sum_{j=1}^{n} c_j(t) + i\mathcal{H}\left(\sum_{j=1}^{n} c_j(t)\right) = \sum_{j=1}^{n}(c_j(t) + i\mathcal{H}c_j(t)) \\ &= \sum_{j=1}^{n} a_j(t)e^{i\varphi_j(t)} = \sum_{j=1}^{n} a_j(t)e^{i(\varphi(0)+\int_0^t \omega_j(s)ds)} \\ &= \sum_{j=1}^{n} a_j(t)e^{i\varphi(0)+2\pi i\int_0^t f_j(s)ds}. \end{aligned}$$

Thus, another way of representing the time series $X(t)$ is given by

$$X(t) = \mathrm{Re}\,(Z(t)) = \mathrm{Re}\left(\sum_{j=1}^{n} a_j(t)e^{i\varphi(0)+2\pi i\int_0^t f_j(s)ds}\right). \tag{1}$$

Now, suppose that $X(t)$ were to be expanded into a Fourier representation. Given in complex form, its Fourier representation would be

$$X(t) = \sum_{j=-\infty}^{\infty} a_j e^{2\pi i f_j t}. \tag{2}$$

Obviously, these expressions are very similar but they differ in an very important aspect. In (2), the amplitudes, a_j, and frequencies, f_j, are constant while the corresponding terms are time dependent in (1). Thus, the HHT with its decomposition can be viewed as a generalized Fourier series expansion. The time varying amplitudes and frequencies allows for a better representation of a nonstationary time series [5].

This new expression of a time series allows us to represent the amplitude and frequncy as functions of time in a 3-D plot. Having a plane with a f- and t-axis the amplitude may be contoured. This f-t-distibution of the amplitude is called the *Hilbert spectrum* and usually denoted as $H(f,t)$. Also, with $H(f,t)$ defined it makes sense to also define the marginal spectrum, $h(f)$, which is given by

$$h(f) = \int_0^T H(f,t)dt.$$

The marginal spectrum can, in some sense, be compared with the amplitude spectrum of a Fourier representation of $X(t)$ given by $|X(f)|$ in the frequency domain. However, there is one important distinction that one has to keep in mind. For the Fourier representation, if $|X(f)| \neq 0$ for a certain f, this means that a harmonic with the same frequency persisted throughout the whole time series. If $h(f) \neq 0$ for the very same frequency f, it merely means that there is a higher likelihood for such a harmonic to have existed locally somehwere in the distribution. The exact time of occurence of this harmonic is provided by the complete Hilbert spectrum [10], [15].

Example 3.4. Consider the function

$$X(t) = \begin{cases} \sin(t^2) + t, & 0 \leq t \leq 25 \\ \sin(20\pi t) + t, & 25 \leq t \leq 100 \end{cases}.$$

In this function there is a chirp where the frequency increases linearly in time. That is, $f(t) = f_0 + kt$ and the chirp signal is $\sin(f(t) \cdot t)$ (in our case, $f_0 = 0$ and $k = 1$). Also, there is also a linear trend present in the function. Clearly, the process represented by the given function is non-linear and nonstationary. Using methods from Fourier analysis to analyze this function is therefore not a good idea since the linear trend will dominate. Using `detrend` in MATLAB removes the linear trend and in Figure Figure 33 we have the amplitude spectrum, $|X(f)|$, of $X(t)$. While the frequency components we expect are present, the Fourier transform does not tell us *when* the two different harmonics occur. We may, however, use the HHT to demcompose $X(t)$ into IMFs using the HHT. With the decomposition, a Hilbert spectrum may be plotted. In Figure 34 a Hilbert spectrum for $X(t)$ can be seen.

From this spectrum it is easy to see that in the interval $0 \leq t \leq 25$ the frequency of the harmonics are given by $\frac{1}{2\pi} \cdot f(t) = \frac{t}{2\pi}$. Furthermore, in the interval $25 \leq t \leq 100$ the frequency of the harmonics are given by $f = 20\pi/(2\pi) = 10$ and this also clearly reflected by the Hilbert spectrum.

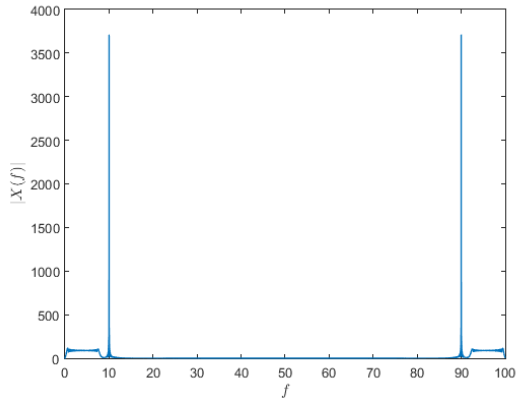

Figure 33: Amplitude spectrum of a "detrended" version of $X(t)$. This spectrum conceived by Fourier methods does not reflect that we have a chirp signal and that $X(t)$ is nonstationary.

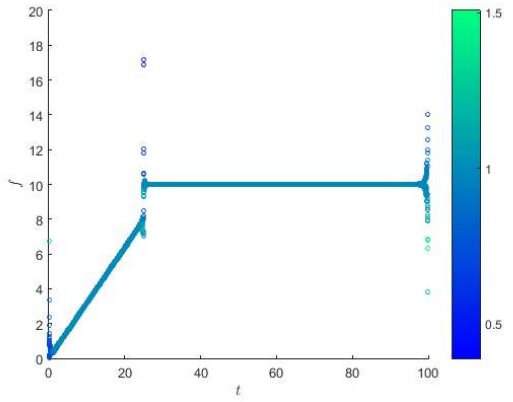

Figure 34: Hilbert spectrum of $X(t)$.

In this section some various numerical investigations of the HHT will be carried out. This is an attempt to try to deduce some properties of the transform in an empirical manner. Also, what weaknesses are inherit in the method?

By looking at the HHT algorithm (or empirical investigation) it is clear that the IMFs given by $c_k(t)$, $1 \leq k \leq n$ decreases in complexity as k increases. The higher complexity of an IMF (more zero crossings and extremas) the higher is its frequency content. Therefore, the following conjecture makes sense.

Conjecture 3.5. Suppose $X(t)$ is a time series and its Hilbert-Huang transform is a decomposition given by

$$X(t) = \sum_{i=1}^{n} c_i(t) + r_n(t).$$

Then the IMFs, $c_k(t)$, are decreasing in complexity as k increases. That is, given $c_k(t)$ and $c_{k+1}(t)$, then $c_k(t)$ has more zero crossings compared to $c_{k+1}(t)$ for $1 \leq k \leq n-1$.

The validity of this conjecture will be investigated throughout the thesis. Note that more zero crossing means that $c_k(t)$ has higher frequency content compared to $c_{k+1}(t)$.

3.3 Trigonometric polynomials and the HHT

Definition 3.6. A function $f(t)$ is said to be a trigonometric polynomial of degree N if it can be written in the form

$$X(t) = a_0 + \sum_{n=1}^{N} a_n \cos(nt) + \sum_{n=1}^{N} b_n \sin(nt)$$

where a_n, b_k for $1 \leq n \leq N$ are any complex numbers.

It is easy to see that $a_n \cos(nt)$ and $b_n \sin(nt)$ for any fix n will satisfy both conditions (i) and (ii) from Definition 3.1. Therefore, an interesting observation is that a trigonometric polynomial is in fact a sum of several functions that each one is an IMF, plus some constant a_0. If given a function $X(t)$ known to be a trigonometric polynomial, is it possible to decompose it using the HHT?

Example 3.7. Consider the function

$$X_1(t) = 4\cos(10t) + 2\cos(t)$$

which is a very simple trigonometric polynomial consisting of only two components. As can be seen in a plot of the function in Figure 35 it is, despite $X_1(t)$ being a fairly simple function, hard to determine what components $X_1(t)$ actually contains. To extract the different components from $X_1(t)$, the Hilbert-Huang Transform should prove useful.

When performing HHT on $X_1(t)$ we should, according to Conjecture 3.5,

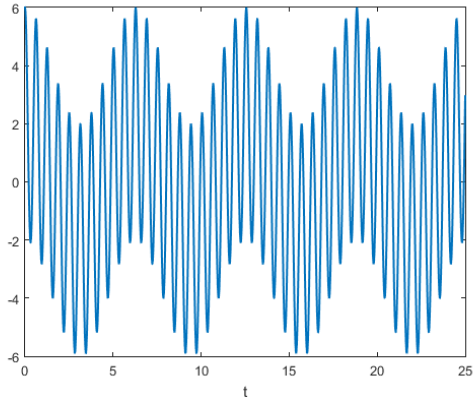

Figure 35: Plot of $X_1(t) = 4\cos(10t) + 2\cos(t)$.

for the first IMF receive $c_1(t) = 4\cos(10t)$ since this is the component with the highest frequency content in $X_1(t)$. Naturally, for the second IMF we expect to recieve $c_2(t) = 2\cos(t)$ which of course is the component in $X_2(t)$ with the second highest frequency content.

The results of performing the HHT numerically on $X(t)$ can be seen in Figure 36. The results seem to be in line with our predictions and it looks very much like $c_1(t) = 4\cos(10t)$ and $c_2(t) = 2\cos(t)$ as expected. This can be seen more clearly in Figure 37 where the differences $c_1(t) - 4\cos(10t)$ and $c_2(t) - 2\cos(t)$ are plotted. Except for some edge effects (which will be adressed later) it is clear that the differences are virtually zero throughout the interval. To summarize,

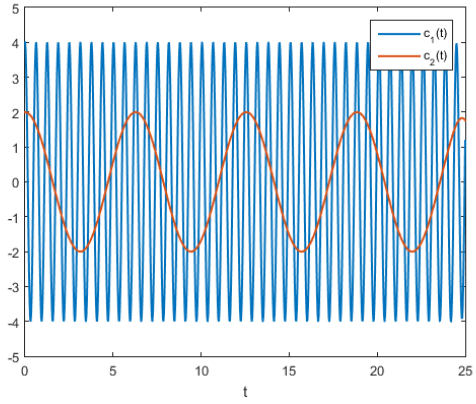

Figure 36: Plot of $c_1(t)$ *and* $c_2(t)$ *calculated numerically through the HHT algorithm for* $X_1(t)$ *for* $0 \leq t \leq 25$.

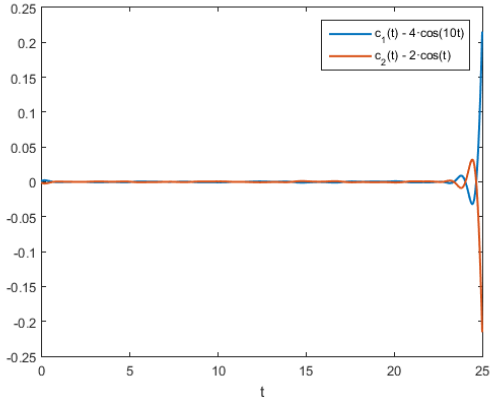

Figure 37: Plot of the difference between expected IMFs and the actual IMFs calculated through the HHT for $0 \leq t \leq 25$.

the HHT worked as exepected on $X_1(t)$, both with regards to Conjecture 3.5 and as a suitable tool for extracting the compononents of a trigonometric polynomial. Note that here is the residual term, $r_2(t)$, equal to zero because $X_1(t)$ consists completely of IMFs.

Remark 3.8. The interval $0 \leq t \leq 25$ is arbitrary. One should note that the HHT performs better if the interval is larger. This is of course due to the numerical algorithm having more data points to consider. The interval in question gives good result and thus will be used.

Example 3.9. Consider the function

$$X_2(t) = \cos(5t) + \cos(t) + 3$$

which can be seen in Figure 38. What happens when $a_0 \neq 0$ for a trigonometric polynomial in regards to the HHT? Obviously the constant a_0 is not an IMF function since it does not satisfy condition (ii) in Definition 5.1. Therefore, the constant will end up in the residual term. In this case, we should get $c_1(t) = \cos(5t)$, $c_2(t) = \cos(t)$ and $r_2(t) = 3$. Like in the first example, this is empirically proven by calculating the HHT for $X_2(t)$ and the results can be seen in Figure 39 and Figure 40.

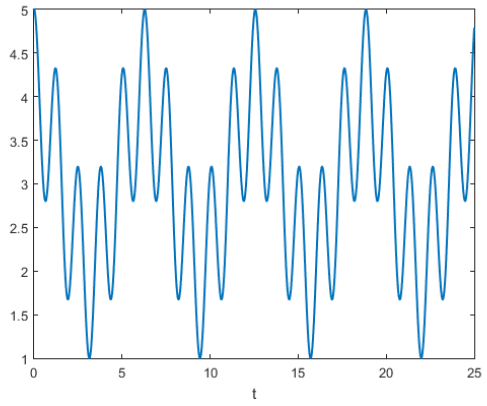

Figure 38: Plot of $X_2(t) = \cos(5t) + \cos(t) + 3$.

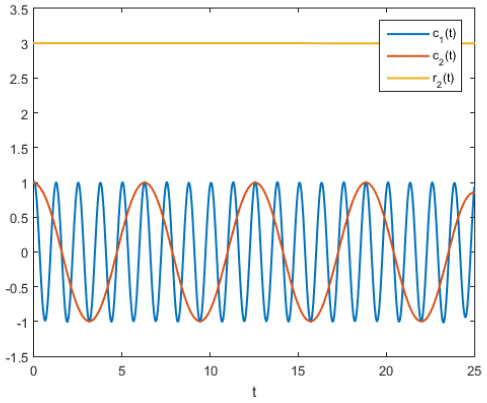

Figure 39: Plot of $c_1(t)$, $c_2(t)$ *and* $r_2(t)$ *calculated numerically through the HHT algorithm for* $X_2(t)$.

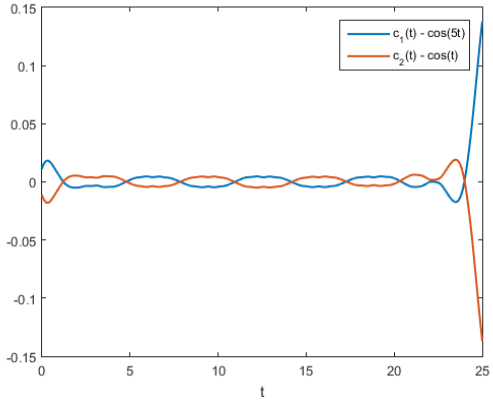

Figure 40: Plot of the difference between expected IMF:s and the actual IMF:s calculated through the HHT.

3.4 Nonstationary processes and the HHT

One of the benefits of the HHT is that it can handle non-stationary time series.

Example 3.10. Consider the function

$$X_3(t) = \begin{cases} \sin(t), & 0 \leq t \leq 4\pi \\ \sin(3t), & 4\pi \leq t \leq 20 \end{cases}$$

which can be seen in Figure 41. Clearly, $X_3(t)$ is continuous everywhere and both of the two seperate functions in each interval is an IMF and thus $X_3(t)$ itself is an IMF. Ideally, the Hilbert-Huang transform of $X_3(t)$ should simply return $X_3(t)$ itself. This is also the case as can be seen in Figure 42 where $c_1(t)$ appears to be identical to $X_3(t)$. This is confirmed to be the case in the very same figure. We see that the difference of $c_1(t)$ and $X(t)$ is zero everywhere. Also, $r_1(t)$ is zero everywhere as should be expected.

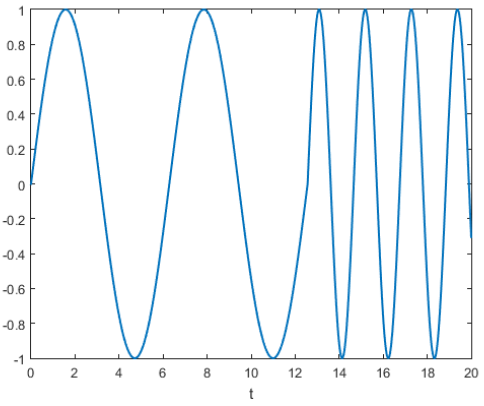

Figure 41: Plot of $X_3(t)$.

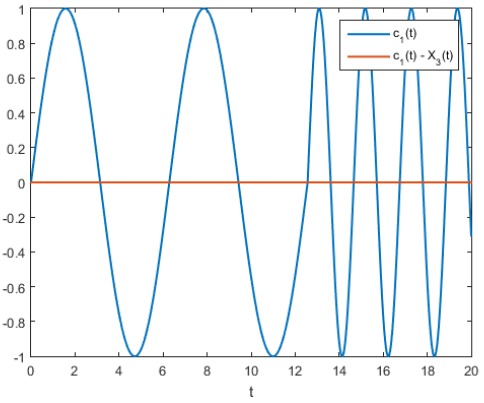

Figure 42: Plot of $c_1(t)$ and the difference of $c_1(t)$ and $X_3(t)$.

Example 3.11. Consider the function

$$X_4(t) = \begin{cases} \sin(t), & 0 \leq t \leq 10\pi \\ \sin(t) + \sin(10t), & 10\pi \leq t \leq 100 \end{cases}$$

which can be seen in Figure 43. Calculating the HHT for a function like this and succesfully extracting and separating the harmonics could be very useful. Imagine we have some background process that is suddenly interrupted by another process (like an earthquake). While the "new" process is dominating the background process is still active. Obviously, this is a nonstationary situation and therefore the HHT could be very useful in acquiring the harmonics.

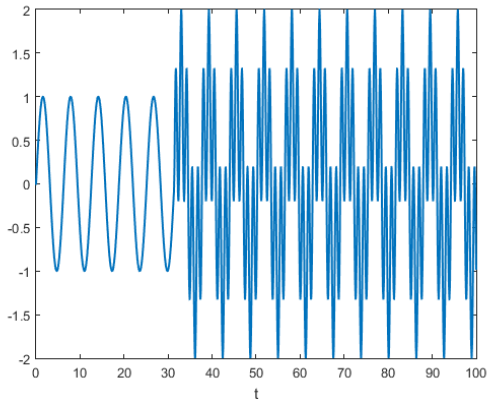

Figure 43: Plot of $X_4(t)$.

According to Conjecture 3.5, the IMFs we should expect are

$$\widetilde{c}_1(t) = \begin{cases} \sin(t), & 0 \leq t \leq 10\pi \\ \sin(10t), & 10\pi \leq t \leq 100 \end{cases}$$

$$\widetilde{c}_2(t) = \begin{cases} 0, & 0 \leq t \leq 10\pi \\ \sin(t), & 10\pi \leq t \leq 100 \end{cases}$$

where the component with the highest frequency in each interval. The actual IMFs, acquired by calculation and denoted by $c_1(t)$ and $c_2(t)$ can be seen in Figure 44. Furthermore, in Figure 45 are the differences of the calculated IMFs and the expected IMFs.

What conclusions can be drawn? Obviously, $c_1(t)$ agrees well with $\widetilde{c}_1(t)$ except for edge effects and in the neighborhood of $t = 10\pi$. Of course, this is because the harmonics shift instantaneously and it is hard for the algorithm to fit this in with IMFs. Despite this, the result of $c_1(t)$ should be considered good since it manages to pick up both harmonics in each interval. For $c_2(t)$, we do not really get what we expected. While it picks up the background harmonics of $\sin(t)$ in $10\pi \leq t \leq 100$, it is a bit off in $0 \leq t \leq 10\pi$. This is because it is hard to fit an IMF to $\widetilde{c}_2(t)$ which has a zero interval and then instantaneous harmonics.

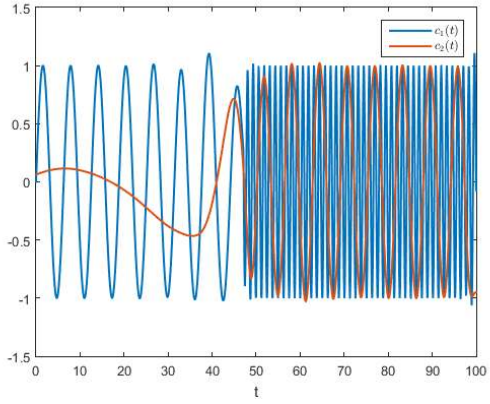

Figure 44: Plot of the two first IMFs, calculated numerically for $X_4(t)$.

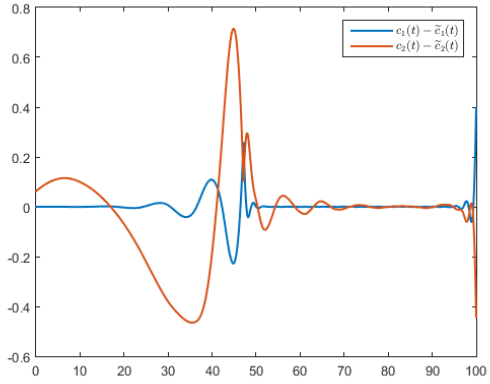

Figure 45: Plot of the difference between the calculated IMFs and the expected IMFs for $X_4(t)$. That is, $c_1(t) - \widetilde{c}_1(t)$ and $c_2(t) - \widetilde{c}_2(t)$

.

3.5 Limitations of HHT

As have been seen in the examples, there are some limitations inherent in the HHT. These limitations are due to the algorithm being carried out numerically and because the data sets are indeed finite in the number of elements.

3.5.1 End effects

Comparing the difference between the expected IMF and the IMF calculated numerically it has been clear that the difference can relatively large at the end points. This effect may be seen very clearly in Figure 37 for example. Expected

IMF and the IMF yielded by the algorithm correspond very well everywhere in the interval except at the end point.

This unwanted effect is due to the datasets being of finite numbers of elements. Therefore, both end points of the data set only have one (instead of two) neighboring point to be considered. Unfortunately, these end effects will become amplified as for every IMF being calculated these effects will spread down to the next and less complex IMF to be considered.

3.5.2 Adjacent frequencies

To discuss this limitation of the HHT, we will look at a concrete example.

Example 3.12. Consider the functions

$$X(t) = \cos(10t) + \cos(t)$$
$$Y(t) = \cos(2t) + \cos(t)$$

which can be seen in Figure 46. Both functions are simple trigonometric polynomials. Using the HHT to extract the components from each function should therefore be fairly straightforward. In Figure 47 the two first IMFs for $X(t)$ and $Y(t)$ can be seen. For $X(t)$ we have $c_1^X(t)$, $c_2^X(t)$ and for $Y(t)$, in the same manner, $c_1^Y(t)$, $c_2^Y(t)$ as the numerically calculated IMFs through the HHT.

Apart from the edge effects it is clear that the HHT algorithm provides good results for decomposing $X(t)$. For $X(t)$, the numerically calculated IMFs are close to the expected IMFs. That is, $c_1^X(t) - \cos(10t)$ and $c_1^X(t) - \cos(t)$ are zero or close to zero throughout the whole interval which can be seen in Figure 49. To conclude, the HHT works very well for $X(t)$ and it manages to decompose the harmonic components of this function.

Now, despite $X(t)$ and $Y(t)$ being very similar functions, the performance of the HHT is very different for $Y(t)$. In Figure 50 it is clear that that the numerically calculated IMFs are not close to the expected IMFs. That is, $c_1^Y(t) - \cos(2t)$ as well as $c_2^Y(t) - \cos(t)$ are not close to zero throughout the interval. Quite the contrary, both differences osciallate between fairly large numbers. To conclude, the HHT actually seems to perform quite badly for $Y(t)$.

Question is, why is there such a difference in the performance of the HHT between $X(t)$ and $Y(t)$? The answer is obtained by looking at and comparing the frequency components of $X(t)$ and $Y(t)$, respectively. Considering the Fourier transform of $X(t)$ we see that it contains $f_X = \frac{10}{2\pi}, \frac{1}{2\pi}$ (note that these are separated by a factor 10). The same reasoning for $Y(t)$ provides the frequency components $f_Y = \frac{2}{2\pi}, \frac{1}{2\pi}$. Note that for $Y(t)$ the different frequency components are closer to each other, almost ajdacent. This becomes even clearer by looking

at the actual Fourier transforms of $X(t)$ and $Y(t)$ [6].

$$X(f) = \frac{1}{2}\left(\delta\left(f - \frac{10}{2\pi}\right) + \delta\left(f + \frac{10}{2\pi}\right)\right) + \frac{1}{2}\left(\delta\left(f - \frac{1}{2\pi}\right) + \delta\left(f + \frac{1}{2\pi}\right)\right)$$
$$Y(f) = \frac{1}{2}\left(\delta\left(f - \frac{2}{2\pi}\right) + \delta\left(f + \frac{2}{2\pi}\right)\right) + \frac{1}{2}\left(\delta\left(f - \frac{1}{2\pi}\right) + \delta\left(f + \frac{1}{2\pi}\right)\right)$$

Analytically, the frequency components are well-separeted by the distinct Dirac delta functions. However, in a numerical representation these Dirac deltas functions will look like "hills" in the frequency domain. This can be seen in Figure 50 where $|X(f)| = |\mathcal{F}(X(t))|$ and $|Y(f)| = |\mathcal{F}(Y(t))|$ have been calculated numerically and plotted in the frequency domain. If these hills are too close to each other, the algorithm will have trouble to differ between the two different frequency components. As a result, we will not receieve as neat and clear-cut IMFs for $Y(t)$ compared to $X(t)$ and the numerical errors will increase if frequencies are adjacent and too alike.

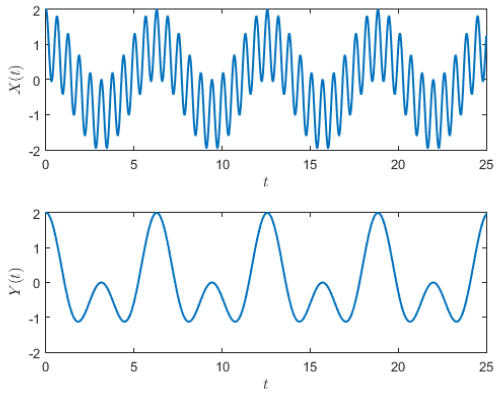

Figure 46: Plots of $X(t) = \cos(10t) + \cos(t)$ *and* $Y(t) = \cos(2t) + \cos(t)$.

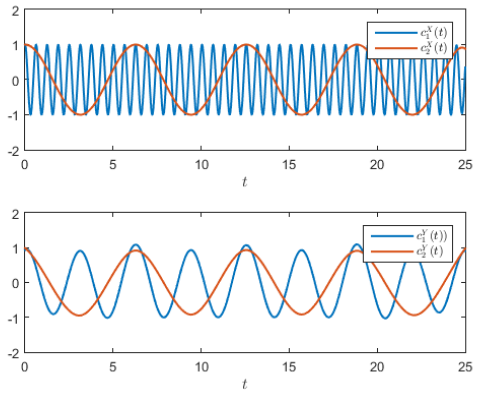

Figure 47: Plots of the IMFs for $X(t)$ and $Y(t)$. The first plots shows the IMFs for $X(t)$ which are $c_1^X(t)$ and $c_2^X(t)$. The second plot shows the IMFs for $Y(t)$ which are $c_1^Y(t)$ and $c_2^Y(t)$.

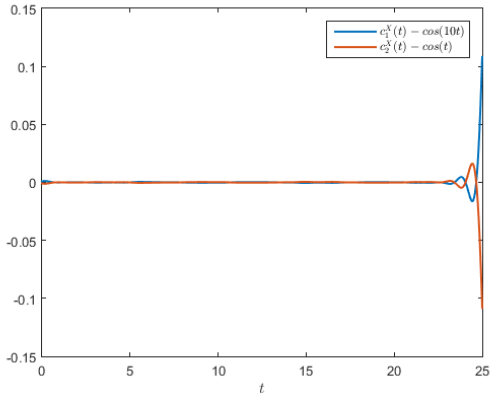

Figure 48: Plot of $c_1^X(t) - \cos(10t)$ and $c_2^X(t) - \cos(t)$. As can be seen in the plot, the expected IMFs are close to the calculated IMFs.

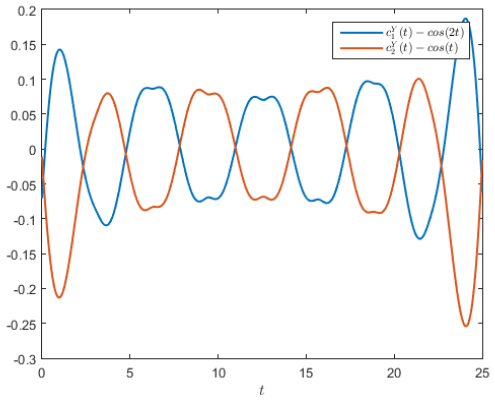

Figure 49: Plot of $c_1^Y(t) - \cos(2t)$ *and* $c_2^Y(t) - \cos(t)$*. Visibly, the expected IMFs are not close to the calculated IMFs.*

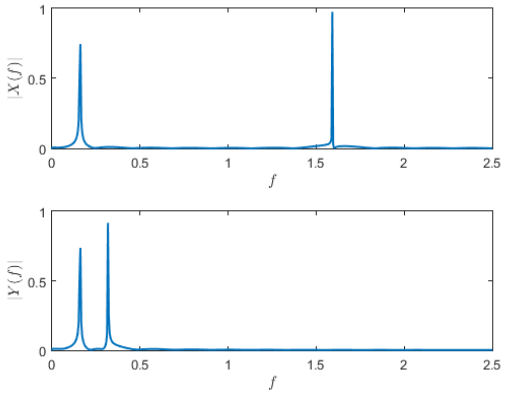

Figure 50: Plot of the power spectrums, $|X(f)|$ *and* $|Y(f)|$ *for* $f \geq 0$ *(note that the power spectrums are symmetric, since the signals in the time domain are real). For* $|X(f)|$ *we expect peaks at* $f = \frac{1}{2\pi} \approx 0.16$ *and* $f = \frac{10}{2\pi} \approx 1.60$*. For* $Y(f)$ *we expect peaks at* $f = \frac{1}{2\pi} \approx 0.16$ *and* $f = \frac{2}{2\pi} \approx 0.32$*. Looking at the power spectrums, this is also the case.*

Example 3.13. For a more quantative approach to the limitation of ajdacent frequencies, consider the function

$$X(t, \epsilon) = \cos((1+\epsilon)t) + \cos(t)$$

for some $\epsilon > 0$. How well does the HHT perform for different values of ϵ? Ideally, when $X(t)$ is subject to HHT, the IMFs we expect are $\widetilde{c}_1(t, \epsilon) = \cos((1+\epsilon)t)$ and $\widetilde{c}_2(t) = \cos(t)$. To determine how well HHT performs for different values

of ϵ, let us calculate $\|c_1(t,\epsilon) - \widetilde{c}_1(t,\epsilon)\|_2$ and $\|c_2(t,\epsilon) - \widetilde{c}_2(t)\|_2$ where $c_1(t,\epsilon)$ and $c_2(t,\epsilon)$ are the numerically calcuated IMFs for $X(t,\epsilon)$. The result can be seen in Figure 51 where $0.1 \leq \epsilon \leq 30$ with a step length of $\Delta\epsilon = 0.1$. The frequency components of $X(t)$ are $f = \frac{1+\epsilon}{2\pi} = \frac{1}{2\pi} + \frac{\epsilon}{2\pi}$ and $f = \frac{1}{2\pi}$. Obviously, as ϵ increases, the frequencies moves farther away from one another. Thus, the HHT experiences less problems of differentiating the two frequencies and the results are better. In contrast, when ϵ is small, the frequencies are adjacaent and the results are bad as can be seen in said figure.

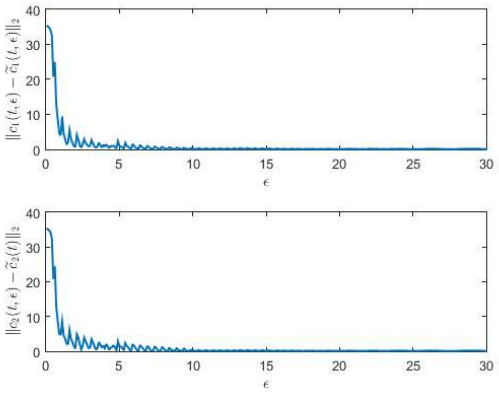

Figure 51: Plot of $\|c_1(t,\epsilon) - \widetilde{c}_1(t,\epsilon)\|_2$ *and* $\|c_2(t,\epsilon) - \widetilde{c}_2(t)\|_2$ *for* $0.1 \leq \epsilon \leq 30$.

For reasonable results with the HHT, frequencies should be seperated by a factor 5 to avoid problems with adjacency. For good results, a factor 10 is desirable.

4 Hilbert transform and modulation

4.1 Introduction to Amplitude modulation

Consider a purely real and continous signal $m(t)$ that we want to transmit. Suppose its Fourier transform, $M(f)$, is bandlimited ($M(f) = 0$ if $|f| > W$ for some W) and also $M(0) = 0$. Usually when transmitting signals it is desired that they are within a given frequency band to avoid interference from other transmissions. To achieve this property *amplitude modulation* may be used. A standard amplitude modulated signal is given by

$$m_{AM}(t) = m(t)\cos(2\pi f_c t)$$

where f_c is called the *carrier frequency*. The carrier frequency is much larger than the bandwidth of $m(t)$, that is, $f_c \gg W$ [8]. Given the Fourier transform

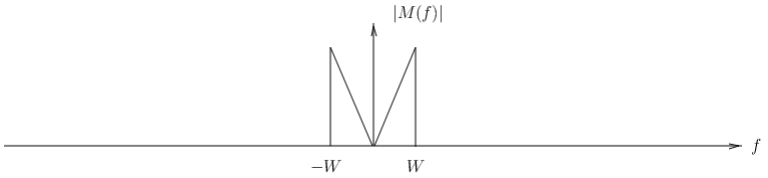

Figure 52: Plot of $|M(f)|$, the amplitude spectrum of $M(f)$. The symmetry is because $m(t)$ is a real signal.

$M(f)$ of $m(t)$, the Fourier transform of $m_{AM}(t)$ is then given by

$$\begin{aligned} M_{AM}(f) &= \mathcal{F}\left(m_{AM}(t)\right) = \mathcal{F}\left(m(t)\cos(2\pi f_c t)\right) = \mathcal{F}(m(t)) * \mathcal{F}(\cos(2\pi f_c t) \\ &= M(f) * \frac{1}{2}\left(\delta(f - f_c) + \delta(f + f_c)\right) = \frac{1}{2}\left(M(f - f_c) + M(f + f_c)\right) \end{aligned}$$

As can be seen in the expression above, the effect $\cos(2\pi f_c t)$ has on the message signal $m(t)$ is that it moves the frequency content of $M(f)$ to f_c and $-f_c$. Thus, $M_{AM}(f)$ has ended up within a more desirable frequency which is around the carrier frequency. One should note that $M(f)$ is symetrically doubled around the origin and each copy of $M(f)$ is centered around f_c and $-f_c$ rescpectively in $M_{AM}(f)$. Thus, the bandwidth used in $M_{AM}(f)$ is doubled compared to the original message. One way to refine this method and to avoid the doubling in bandwidth is to use the method of *Single-Sideband modulation* (SSB). This method makes use of the Hilbert transform and analytic signals and will be investigated further later.

Suppose that $m_{AM}(t)$ has been transmitted and received. How is the orignal message signal, $m(t)$, extracted by the receiver? By simply multiplying with

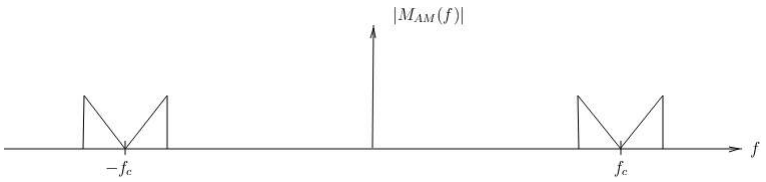

Figure 53: Plot of $|M_{AM}(f)|$, the amplitude spectrum of $M_{AM}(f)$.

$\cos(2\pi f_c t)$ once again we receive

$$m_{AM}(f)\cos(2\pi f_c t) = m(t)\cos^2(2\pi f_c t) = m(t)\cdot\frac{1}{2}(\cos(4\pi f_c t)+1)$$
$$= \frac{1}{2}\cdot m(t) + \frac{1}{2}\cdot m(t)\cos(4\pi f_c t)$$

and extracting $m(t)$ is just a matter of applying a suitable low-pass filter and multiplying with a factor 2. The term $\frac{1}{2}\cdot m(t)\cos(4\pi f_c t)$ is clearly high-frequent compared to $\frac{1}{2}\cdot m(t)$ and will therefore disappear when said low-pass filter is applied onto the signal $m_{AM}(f)\cos(2\pi f_c t)$. The act of extracting the original signal from a modulated version of said signal is called *demodulation*.

4.2 Single-sideband modulation

Obviously, when using standard techniques of amplitude modulation the frequency content is doubled and more bandwidth is required for transmission. This is avoidable and modulation can be performed more efficiently using Single-sideband modulation (SSB) [8]. This technique employs the connection between the Hilbert transform and analytic signals.

Suppose as $m(t)$ is as in the previous section. Begin by expanding $m(t)$ into an analytic signal using the Hilbert transform.

$$z(t) = m(t) + i\mathcal{H}(m(t))$$

Here $z(t)$ is an analytic signal and because $m(t)$ is purely real, Theorem 1.26 readily gives us the Fourier transform of $z(t)$.

$$Z(f) = \begin{cases} 2M(f), & f>0 \\ M(0), & f=0 \\ 0, & f<0 \end{cases} = \begin{cases} 2M(f), & f>0 \\ 0, & f\leq 0 \end{cases}$$

Of course, since $m(t)$ is purely real, no information is lost in $Z(f)$ due to $M(f)$ exhibiting Hermitian symmetry. That is, $M(-f) = \overline{M(f)}$ and consequently $|M(-f)| = |M(f)|$ as discussed in Example 1.27. Like before, it is desirable to use amplitude modulation on $z(t)$ to move the frequency content into a given frequency band. Construct a new signal by multiplying $z(t)$ with $e^{2\pi i f_c t}$.

$$z_a(t) = z(t)\cdot e^{2\pi i f_c t}$$

The Fourier transform of $z_a(t)$ is easily calculated using convolution.

$$\begin{aligned} Z_a(f) &= \mathcal{F}(z_a(t)) = \mathcal{F}\left(z(t) \cdot e^{2\pi i f_c t}\right) = \mathcal{F}(z(t)) * \mathcal{F}\left(e^{2\pi i f_c t}\right) \\ &= Z(f) * \delta(f - f_c) = Z(f - f_c) \end{aligned}$$

Using the derived expression for $Z(f)$ from above we get the expression

$$Z_a(f) = Z(f - f_c) = \begin{cases} 2M(f - f_c), & f > f_c \\ 0, & f \leq f_c \end{cases}$$

and accordingly it has both only positive frequency components and due to amplitude modulation, said components lie about $f = f_c$. Since all analytic signals have only positive frequency components, $z_a(t)$ is an analytic represenation of an SSB-signal that we define as $m_{SSB}(t)$.

$$z_a(t) = m_{SSB}(t) + i\mathcal{H}(m_{SSB}(t))$$

But we also have that

$$\begin{aligned} z_a(t) =& z(t) \cdot e^{2\pi i f_c t} = (m(t) + i\mathcal{H}(m(t))) \cdot (\cos(2\pi f_c t) + i \sin(2\pi f_c t)) \\ =& m(t) \cdot \cos(2\pi f_c t) - \mathcal{H}(m(t)) \cdot \sin(2\pi f_c t) + i \cdot [m(t) \cdot \sin(2\pi f_c t) + \\ &+ \mathcal{H}(m(t)) \cdot \cos(2\pi f_c t)] \end{aligned}$$

Since $m_{SSB}(t) = \operatorname{Re} z_a(t)$ we get by simply identifying real parts

$$m_{SSB}(t) = m(t) \cdot \cos(2\pi f_c t) - \mathcal{H}(m(t)) \cdot \sin(2\pi f_c t)$$

By using Theorem 1.12 we can acquire the Fourier transform of $m_{SSB}(t)$.

$$\begin{aligned} M_{SSB}(f) &= \mathcal{F}(m_{SSB}(t)) = \mathcal{F}(m(t) \cdot \cos(2\pi f_c t) - \mathcal{H}(m(t)) \cdot \sin(2\pi f_c t)) \\ &= \mathcal{F}(m(t)) * \mathcal{F}(\cos(2\pi f_c t)) - \mathcal{F}(\mathcal{H}(m(t))) * \mathcal{F}(\sin(2\pi f_c t)) \\ &= M(f) * \frac{1}{2}(\delta(f - f_c) + \delta(f + f_c)) - \\ &- (-i \operatorname{sgn}(f)) M(f) * \frac{1}{2i}(\delta(f - f_c) - \delta(f + f_c)) \\ &= \frac{1}{2}(1 + \operatorname{sgn}(f - f_c)) M(f - f_c) + \frac{1}{2}(1 - \operatorname{sgn}(f + f_c)) M(f + f_c) \\ &= \begin{cases} M(f - f_c), & f > f_c \\ 0, & -f_c \leq f \leq f_c \\ M(f + f_c), & f < -f_c \end{cases} \end{aligned}$$

From this we can draw the conclusion that the frequency content for the SSB has been halved compared to the amplitude modulated signal. Demodulation, recovering $m(t)$ from $m_{SSB}(t)$ that is, works in a similar fashion as before -

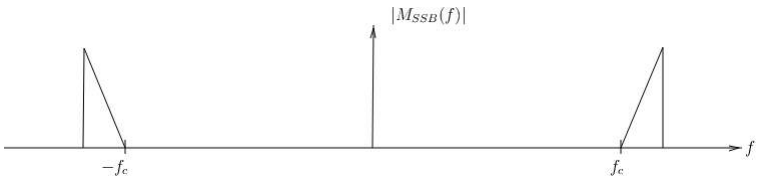

Figure 54: Plot of $|M_{SSB}(f)|$, the amplitude spectrum of $M_{SSB}(f)$.

begin by multiplying the signal with $\cos(2\pi f_c t)$.

$$\begin{aligned} m_{SSB}(t)\cos(2\pi f_c t) &= (m(t)\cdot\cos(2\pi f_c t) - \mathcal{H}(m(t))\cdot\sin(2\pi f_c t))\cdot\cos(2\pi f_c t) \\ &= m(t)\cdot\cos^2(2\pi f_c t) - \mathcal{H}(m(t))\cdot\cos(2\pi f_c t)\sin(2\pi f_c t)) \\ &= m(t)\cdot\frac{1}{2}(\cos(4\pi f_c t)+1) - \mathcal{H}(m(t))\cdot\frac{1}{2}\sin(4\pi f_c t) \\ &= \frac{1}{2}\cdot m(t) + \frac{1}{2}\cdot m(t)\cos(4\pi f_c t) - \frac{1}{2}\cdot m(t)\sin(4\pi f_c t) \end{aligned}$$

In exact manner as before, a low-pass filter and multiplication with a factor 2 help us to recover $m(t)$. A suitable low-pass filter will accordingly remove the terms $\frac{1}{2}\cdot m(t)\cos(4\pi f_c t)$ and $-\frac{1}{2}\cdot m(t)\sin(4\pi f_c t)$ which are high-frequent.

In this application of the SSB we get the upper sidebands. An alternative approach is to instead get the lower sidebands [8]. Start by considering

$$w(t) = \overline{z(t)} = m(t) - i\mathcal{H}(m(t))$$

The Fourier transform of this conjugated function is given by

$$\begin{aligned} W(f) = \mathcal{F}(w(t)) &= \mathcal{F}\left(\overline{z(t)}\right) = \mathcal{F}(m(t) - i\mathcal{H}(m(t))) \\ &= \mathcal{F}(m(t)) - i\cdot\mathcal{F}(\mathcal{H}(m(t)) = M(f) - i\cdot(-i\,\mathrm{sgn}(f)\cdot M(f)) \\ &= M(f) - \mathrm{sgn}(f)\cdot M(f) = (1-\mathrm{sgn}(f))\cdot M(f) \end{aligned}$$

This expression may be simplified by looking at subcases.

$$W(f) = \begin{cases} 0, & f>0 \\ M(0), & f=0 \\ 2M(f), & f<0 \end{cases} = \begin{cases} 0, & f\geq 0 \\ 2M(f), & f<0 \end{cases}$$

Obviously, $w(t)$ is not an analytic signal since its Fourier transform $W(f)$ is non-zero for $f<0$ and zero for $f>0$. However, we multiply by the factor $e^{2\pi i f_c t}$ and thus modulate $w(t)$.

$$w_a(t) = w(t)\cdot e^{2\pi i f_c t}$$

By earlier examples, its Fourier transform is given by

$$W_a(f) = W(f-f_c) = \begin{cases} 0, & f\geq f_c \\ 2M(f-f_c), & f<f_c \end{cases}$$

and if f_c is large enough it is possible that $W_a(f) = 0$ for all $f \leq 0$. How large does f_c have to be? For $f = 0$ we have that $W_a(0) = 2M(-f_c)$ and because it is necessary that $W_a(f) = 0$ for all $f \leq 0$ we must have that $-f_c < -W$ which means that $f_c > W$. If this is the case, $w_a(t)$ is an analytic signal since $W_a(f) = 0$ for all $f \leq 0$. Also, note that it is the lower sideband of $M(f)$ in the expression belonging to $W_a(f)$. Consequently, it makes sense to put

$$w_a(t) = m_{LSB}(t) + i\mathcal{H}(m_{LSB}(t))$$

but as before we also have that

$$\begin{aligned} w_a(t) =& w(t) \cdot e^{2\pi i f_c t} = (m(t) - i\mathcal{H}(m(t))) \cdot (\cos(2\pi f_c t) + i \sin(2\pi f_c t)) \\ =& m(t) \cdot \cos(2\pi f_c t) + \mathcal{H}(m(t)) \cdot \sin(2\pi f_c t) + i \cdot [m(t) \cdot \sin(2\pi f_c t) + \\ & - \mathcal{H}(m(t)) \cdot \cos(2\pi f_c t)] \end{aligned}$$

Since $m_{LSB}(t) = \operatorname{Re} w_a(t)$ we get, once again, by simply identifying real parts

$$m_{LSB}(t) = m(t) \cdot \cos(2\pi f_c t) + \mathcal{H}(m(t)) \cdot \sin(2\pi f_c t)$$

and the Fourier transform of $m_{LSB}(t)$ is given by

$$\begin{aligned} M_{LSB}(f) &= \mathcal{F}(m_{LSB}(t)) = \mathcal{F}(m(t) \cdot \cos(2\pi f_c t) + \mathcal{H}(m(t)) \cdot \sin(2\pi f_c t)) \\ &= \mathcal{F}(m(t)) * \mathcal{F}(\cos(2\pi f_c t)) + \mathcal{F}(\mathcal{H}(m(t))) * \mathcal{F}(\sin(2\pi f_c t)) \\ &= M(f) * \frac{1}{2}(\delta(f - f_c) + \delta(f + f_c)) + \\ &+ (-i \operatorname{sgn}(f)) M(f) * \frac{1}{2i}(\delta(f - f_c) - \delta(f + f_c)) \\ &= \frac{1}{2}(1 - \operatorname{sgn}(f - f_c)) M(f - f_c) + \frac{1}{2}(1 + \operatorname{sgn}(f + f_c)) M(f + f_c) \\ &= \begin{cases} M(f + f_c), & f > f_c \\ M(f - f_c) + M(f + f_c), & -f_c \leq f \leq f_c \\ M(f - f_c), & f < -f_c \end{cases} \end{aligned}$$

however, since $M(f + f_c) = 0$ for $f > f_c$ and $M(f - f_c) = 0$ for $f < -f_c$ the expression above may be further simplified.

$$M_{LSB}(f) = \begin{cases} 0, & f > f_c \\ M(f - f_c) + M(f + f_c), & -f_c \leq f \leq f_c \\ 0, & f < -f_c \end{cases}$$

Clearly, in $M_{LSB}(f)$ only the lower sidebands are present as opposed to the expression of $M_{SSB}(f)$ where we had the upper sidebands. Both methods have the advantage of halving the required bandwidth for transmission and receiving their respective signals in the time domain.

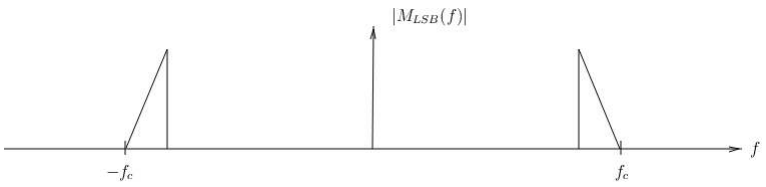

Figure 55: Plot of $|M_{LSB}(f)|$, the amplitude spectrum of $M_{LSB}(f)$.

5 Discussion

From the results presented in this thesis some conclusions may be drawn. Of course, the discussion is focused around the Electrocardiagram and the Hilbert-Huang transform for which original research has been carried out.

5.1 Electrocardiogram

Provided an ECG-signal, $ECG(t)$, and expanding this real valued signal into an analytic signal given by

$$z(t) = ECG(t) + i\mathcal{H}(ECG(t))$$

has proven to be an effective way of detecting the QRS complex. Since the QRS resembles a deformed sine wave, the parametric plot of $z(t)$ will produce a distinguishable main loop in the complex plan if there is a QRS complex within the parametric interval. This main loop encloses the origin and therefore crosses all the axes.

The main reason that the method suggested to extract QRS complex in an ECG signal lacks mathematical theory is because it does not work for all possible ECG-signals. As was seen in the example in section 2.4, if the QRS complex is not high enough, the method does not work. This is because the lower QRS complex does not produce any distinguishably large main loop in the parametric plot of $z(t)$. Depending on how the algorithm is constructed, either it will fail to detect the QRS complex or a T-wave may be misinterpreted as in the parametric plot and thus the detection of the QRS complex fails or a T-wave is incorrectly interpreted as a QRS complex.

The possibilities to expand this research is to prove the method mathematically as opposed to numerical justifications (given one consider the limitations mentioned above). Also, find ways for QRS complex detection to work beyond the mentioned limitation.

5.2 Hilbert-Huang transform

The Hilbert-Huang transform (HHT) is a tool for analyzing time series that has proven very effective. However, even though the method is gaining popularity it is still questionable why "Fourier methods" still dominate spectral analysis. The HHT is a more versatile tool and has the ability to treat frequencies and amplitude as dependent on time. Thus, the HHT is very appropriate in analyzing nonstationary and nonlinear time series in contrast to, for example, the Fast Fourier Transform (FFT). One could easily argue that the HHT should be the standard method for analyzing time series since stationarity and linearity is quite rare in real examples [10].

The HHT lacks mathematical theory because the empirical conceived method

does not work for all cases either. As seen in example 3.11, the algorithm fails when frequencies are too cloose to each other (ajdacent frequencies).

The limitations of the HHT discovered during the experiments must be considered. One possibility of further research in this area is how to further reduce these limitations - especially in the case of "adjacent frequencies". Possibly the algorithm of decomposing could be refined to better deal with adjacent frequencies. Also, reducing, or even removing, the apparent end effects would of course be desirable.

The biggest advantage of the HHT is that it handles nonstationary processes so well. For example, considering a scenario where we are measuring seismic waves in connection with earthquake activity. Because of the nonstationary nature of these situations, as the frequency of the seismic waves fluctuate with time, the HHT is a very appropriate tool in analyzing these waves. In these types of situations, a Fourier spectral analysis oftentimes not suitable.

References

[1] Werner, J., Krantz, M., *EKG*, Bokakademin, Linköping, 2003.

[2] Abeysekera, R.M.S.S., Bolton, R.J., Westphal, L.C. and Boashash, B., Patterns in Hilbert Transforms and Wigner-Wille Distributions of Electrocardiogram Data, *IEEE International Conference on ICASSP*, **11** (1986), 1973-1978.

[3] *European ST-T Database*, 2000.
Available online at `http://physionet.org/physiobank/database/edb/.`

[4] King, F.W., *Hilbert Transforms Volume 1*, Cambridge University Press, Cambridge, 2009.

[5] King, F.W., *Hilbert Transforms Volume 2*, Cambridge University Press, Cambridge, 2009.

[6] Hahn, S., *Hilbert transforms in signal processing*, Artech House, Inc., Norwood, 1996.

[7] Aigner, M., *Fourieranalys*, Bokakademin, Linköping, 2015.

[8] Olofsson, M., *Signal theory*, Studentlitteratur, Lund, 2011.

[9] Alexandersson, L., *TATA45 Komplex analys*, Bokakademin, Linköping, 2014.

[10] Huang, N.E., et al, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *The Royal Society*, **454** (1998), 903-995.

[11] Kreyszig, E., *Introductory Functional Analysis with Applications*, John Wiley & Sons Inc, Windsor, 1978.

[12] Alavi-Sereshki, M.M., *Analytical signals and Hilbert transforms*, Texas Tech University, Lubbock, 1972.

[13] Benitez, D., et al., The use of the Hilbert transform in ECG signal analysis, *Computers in Biology and Medicine*, **31** (2001), 399–406.

[14] Linderhed, A., *Adaptive image compression with wavelet packets and empirical mode decomposition*, Linköping University Post Print, Linköping, 2004.

[15] Sahoo, J.P., *Analysis of ECG signal for Detection of Cardiac Arrhythmias*, National Institute Of Technology Rourkela, Orissa, 2011.

[16] Tolwinski, S., *The Hilbert Transform and Empirical Mode Decomposition as Tools for Data Analysis*, University of Arizona, Tuscon, 2007.

[17] Gong, K., *Speech processing using the empirical mode decomposition and the Hilbert transform*, Concordia University, Montreal, 2004.

[18] Tan, A., Hilbert-Huang Transform algorithm for MATLAB, 2008. Available online at `http://goo.gl/Fejjvy`.

[19] Huang, N.E., *Hilbert-Huang Transform And Its Applications*, World Scientific Publishing Co Pte Ltd, Singapore, 2005.

LINKÖPINGS UNIVERSITET