HILBERT TRANSFORM OF DISCRETE DATA: A BRIEF REVIEW

Corneliu Rusu\(^1,2\), Pauli Kuosmanen\(^3\) and Jaakko Astola\(^1\)

\(^1\) Tampere International Center for Signal Processing,
Tampere University of Technology,
P.O.Box 553, FIN-33101 Tampere, FINLAND, Jaakko.Astola@tut.fi

\(^2\) Faculty of Electronics and Telecommunications,
Technical University of Cluj-Napoca
Baritiu 26-28, RO-400027 Cluj-Napoca, ROMANIA, Corneliu.Rusu@ieee.org

\(^3\) Institute of Signal Processing,
Tampere University of Technology,
P.O.Box 553, FIN-33101 Tampere, FINLAND, Pauli.Kuosmanen@tut.fi

ABSTRACT

The goal of this work is to review and discuss several methods for the numerical evaluation of certain Cauchy principal value integrals. The problems addressed here appear by discretization of Hilbert transform, Kramers-Kronig relations or Bode relationships. Comments about their implementation are also provided.

1. INTRODUCTION

Hilbert transform has been recognized as very important method in different branches of science and technology, from complex analysis and optics, to circuit theory and control science. Their sampled derivations have been encountered in different applications from applied science and engineering.

In some situations the domain is restricted or other explicit conditions are imposed. A critical issue is related to the singularities involved in the Hilbert transform computation, since we are confronted with an improper integral. If the integral cannot be evaluated in a closed form, as it is the case with discrete input data, numerical implementation is in general complicated.

Hilbert transform has the advantage of not requiring derivatives, but the serious disadvantage that it is not a bounded operator from \(L_\infty\) to \(L_\infty\). To solve the problem, different approaches for gain-phase relationships in logarithmic frequency domain have been proposed. A suitable change of variable can give the bounded operator (10) from \(L_r\) to \(L_\infty\) for any \(r > 1\) [1].

To solve the problem, different approaches for computing Hilbert transform have been proposed. The goal of this paper is to present a brief review of methods used to compute Hilbert transform when the signal is composed of discrete data, sampled at equidistant or arbitrarily instants in time domain.

For this purpose we first recall how Hilbert transform, Kramers-Kronig relations or Bode relationships are known in signal processing, optics, circuit theory and control engineering (Section 2). A general presentation of Hilbert transform from mathematical point of view is addressed in Section 3. Related results about quadrature rules for Cauchy principal value integrals are provided in Section 4. Some comments and remarks about implementation are provided in Section 5.

2. HILBERT TRANSFORM IN FEW FIELDS OF SCIENCE AND ENGINEERING

2.1. Signal processing and communications

Frequently the Hilbert transform is introduced as a convolution between \(f(x)\) and \(-1/(\pi x)\) [2]:

\[
\mathcal{H}f(x) = -\frac{1}{\pi x} * f(x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} f(\tau) \frac{d\tau}{\tau - x}, \tag{1}
\]

We note that the Hilbert transform is occasionally defined with the opposite sign [3] to that given in (1).

If \(x(n)\) is a causal and absolutely summable real sequence with a discrete time Fourier transform \(X(e^{j\omega})\), then Hilbert transform can rewritten in few different forms. In signal processing it is common to designate the Hilbert pair \(X_{re}(e^{j\omega})\) and \(X_{im}(e^{j\omega})\), which are the real and imaginary parts of \(X(e^{j\omega})\). Indeed, \(X_{re}(e^{j\omega})\) and \(X_{im}(e^{j\omega})\) are related by [4]:

\[
X_{im}(e^{j\omega}) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} X_{re}(e^{j\omega}) \coth \left( \frac{\omega - \varphi}{2} \right) d\varphi, \tag{2}
\]

\[
X_{re}(e^{j\omega}) = x(0) + \frac{1}{2\pi} \int_{-\pi}^{\pi} X_{im}(e^{j\omega}) \coth \left( \frac{\omega - \varphi}{2} \right) d\varphi. \tag{3}
\]

The discrete Hilbert transform as introduced by Cizek in [5] is as follows:

\[
g_i = \frac{1}{N} \sum_{k=0}^{N-1} f(k)[1 - (-1)^{k-i}] \coth[(k - i)\pi/N] \tag{4}
\]
and it can be modified for $i$ even or odd in the following way:

- for $i$ odd:
  \[
  g_i = \frac{2}{N} \sum_{k=0,2,4,...} f(k) \coth[(k-i)\pi/N] \quad (5)
  \]

- for $i$ even:
  \[
  g_i = \frac{2}{N} \sum_{k=1,3,5,...} f(k) \coth[(k-i)\pi/N] \quad (6)
  \]

The Hilbert transform has several important applications, which include the following [6]:

1. it can be used to realize phase selectivity in the generation of a special kind of modulation known as single sideband modulation;
2. it provides the mathematical basis for the representation of band-pass signals.

The Hilbert transform applies to energy signals as well as power signals:

- an energy signal and its Hilbert transform are orthogonal over the entire interval $(-\infty, \infty)$;

- similarly, a power signal and its Hilbert transform are orthogonal over one period.

### 2.2. Kramers-Kronig relations

Hilbert transforms arise in many applications from optics, and they are often called there by different names such as dispersion relations, Kramers-Kronig transforms and Cauchy principal value integrals [7]. According to Kramers-Kronig theory, the attenuation (or amplification) of light is always connected with a phase shift.

The interaction between a light wave and matter can be described with a complex susceptibility of the material, whose real and imaginary parts are connected with the phase shift and the amplitude variation of the wave, respectively. Since the susceptibility must be an analytical function of the excitation frequencies (e.g., electronic or vibrational) of the material. The Kramers-Kronig relations are more general, however, since their origin is purely mathematical and is not related to a specific effect in physics. Therefore, they can be used with any variable as long as the susceptibility as a function of this variable fulfills the well-known mathematical conditions.

Let $F$ be a suitably regular function on $\mathbb{R}$ that vanishes on the half line $s \leq 0$ [8], and let

\[
f(x) = \int_0^\infty F(s)e^{i2\pi sx} ds.
\]

Since $\text{sgn} \cdot F = F$, we can write:

\[
\mathcal{H}f = -j F^{-1} \text{sgn} F f = -j F^{-1} (\text{sgn} F) = -j f
\]
or equivalently,

\[
\mathcal{H}\{f_R + jf_I\} = \{f_R + jf_I\}
\]

where $f_R, f_I$ are the real and imaginary parts of $f$.

Since $\mathcal{H}$ commutes with complex conjugate operator, then $\mathcal{H}f_R, \mathcal{H}f_I$ are real, so we can equate the real and imaginary parts of this identity to obtain the Kramers-Kronig relations:

\[
\mathcal{H}f_R = f_I, \quad \mathcal{H}f_I = -f_R,
\]

when $F(s) = 0$ for $s \leq 0$.

There is an extensive work devoted to the application of Kramers-Kronig transforms to experimental data. Some strategies have focused on avoiding the principal value integrals altogether. A primary area of application is the analysis of optical data, that is, determination of the principal value integrals altogether. A primary area of application is the analysis of optical data, that is, determination of the dispersive mode from a measurement of the absorptive mode and viceversa.

There are two principal issues involved in optical data analysis:

1. fitting measurements to some particular functional form, including a resolution of the extrapolation problem to regions outside which spectral measurements have been made;

2. solving the Kramers-Kronig inversion either analytically or numerically.

### 2.3. Bode relationships

In circuit theory and control engineering, there has been a constant interest in developing techniques using logarithmic frequency axis. Let us consider $H(j\omega)$ the Fourier transform of a causal function $h(t)$:

\[
H(j\omega) = \int_0^\infty h(t)e^{-j\omega t} dt = R(\omega) + jI(\omega).
\]

Proofs based on Cauchy’s residue theorem [9] or convolution [10] established that we have

\[
R(\omega) = R(\infty) - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{I(y)}{y - \omega} dy
\]

\[
= R(\infty) - \frac{2}{\pi} \int_0^{\infty} \frac{yI(y) - \omega I(\omega)}{y^2 - \omega^2} dy,
\]

and

\[
I(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{R(y)}{y - \omega} dy = \frac{2\omega}{\pi} \int_0^{\infty} \frac{R(y) - R(\omega)}{y^2 - \omega^2} dy,
\]

which establish the Hilbert pair of $R(\omega)$ and $I(\omega)$. 
One can easily obtain the gain-phase relations (or Bode relationships) from (7) and (8) directly by taking logarithms [10], after fulfilling the requirements needed to satisfy the right half plane analyticity conditions of the Hilbert transform, i.e., the stable and minimum phase conditions. Under the assumption that \( H(s) \) is not only analytic, but has no zeros for \( \text{Re}(s) \geq 0 \), then:

\[
\ln(H(j\omega)) = \alpha(\omega) + j\beta(\omega)
\]

will also be analytic in the right-hand plane. Thus the phase \( \beta(\omega) \) will be uniquely determined from the gain (in nepers) \( \alpha(\omega) \) by using (8).

If we assume that \( H(s) \) is not only analytic, but has no zeros for \( \text{Re}(s) \geq 0 \), then \( \ln(H(j\omega)) = \alpha(\omega) + j\beta(\omega) \) will also be analytic in the right-hand plane, and the phase \( \beta(\omega) \) will be uniquely determined from the gain (in nepers) \( \alpha(\omega) \):

\[
\beta(\omega) = \frac{2\omega}{\pi} \int_0^\infty \frac{\alpha(y) - \alpha(\omega)}{y^2 - \omega^2} dy. \tag{9}
\]

A change of variable \( u = \ln(y/\omega_c) \) where \( \omega_c \) is a normalizing frequency, is usually introduced. The results are [9]:

\[
\beta(\omega_c) = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\alpha(\omega_c e^u) - \alpha(\omega_c e^{-u})}{e^u - e^{-u}} du
\]

\[
= \frac{2}{\pi} \int_0^{\infty} \frac{\alpha(\omega_c e^u) - \alpha(\omega_c e^{-u})}{e^u - e^{-u}} du
= \frac{1}{\pi} \int_{-\infty}^{\infty} \left( \frac{d}{du} \alpha(\omega_c e^u) \right) \ln \left( \coth \frac{|u|}{2} \right) du. \tag{10}
\]

Equation (10) shows mostly that the phase characteristic is proportional to the derivative of the gain characteristic on a logarithmic frequency scale, weighted by an even function of frequency.

This result is the basic of a method often used for a long time in analogue electronics in order to draw the phase characteristics as concatenation of straight lines.

Bode straight-line approximations are an extremely useful tool also in the study of system frequency response. These approximations give good insight into the frequency variation of the amplitude and the phase of a system response without the use of computer simulation or complex calculations.

### 3. HILBERT TRANSFORM

To define the Hilbert transform for all type of signals, from mathematical point of view, we need few preliminaries.

We denote by \( \mathbb{R} \) and \( \mathbb{Z} \) the sets of real numbers and of integers, respectively. For \( p > 0 \), we consider the functions which are \( p \)-periodic, i.e., on the circle \( \mathbb{T}_p \). For \( N \) a positive integer, \( \mathbb{P}_N \) consists of \( N \) uniformly spaced points on the circle \( \mathbb{T}_N \) [8].

We define the odd signum function on \( \mathbb{R}, \mathbb{T}_p, \mathbb{Z} \) and \( \mathbb{P}_N \) [8], by writing:

\[
\text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases} \tag{11}
\]

\[
\text{sgn}(x) = \begin{cases} 1 & \text{if } 0 < x < p/2 \\ 0 & \text{if } x = 0, p/2 \\ -1 & \text{if } p/2 < x < p \end{cases} \tag{12}
\]

\[
\text{sgn}(n) = \begin{cases} 0 & \text{if } n = 0 \\ 1 & \text{if } n = 1, 2, \ldots \\ -1 & \text{if } n = -1, -2, \ldots, N - [N/2] \\ \end{cases} \tag{13}
\]

respectively.

We then form the Hilbert transform operator

\[
\mathcal{H} \equiv -j \mathcal{F}^{-1} \mathcal{M}_{\text{sgn}} \mathcal{F}, \tag{15}
\]

where \( \mathcal{F} \) is Fourier transform operator and \( \mathcal{M}_g f = g \cdot s \) is the multiplication operator.

Let denote by \( f^\# \) to show that \( \mathcal{H} \) has been applied:

\[
f^\# \equiv \mathcal{H} f. \]

Thus when \( f \) is a suitably regular function on \( \mathbb{R} \), \( \mathbb{T}_p, \mathbb{Z} \) and \( \mathbb{P}_N \) we have:

\[
f^\# = -j \int_{-\infty}^{\infty} \text{sgn}(s) \mathcal{F}(s) e^{j2\pi xs} ds, \tag{16}
\]

\[
f^\# = -j \sum_{k=-\infty}^{\infty} \text{sgn}(k) \mathcal{F}(k) e^{j2\pi kx/p} ds, \tag{17}
\]

\[
f^\# = -j \int_0^p \text{sgn}(s) \mathcal{F}(s) e^{j2\pi sn/p} ds, \tag{18}
\]

\[
f^\# = -j \sum_{k=0}^{N-1} \text{sgn}(k) \mathcal{F}(k) e^{j2\pi kn/N} ds, \tag{19}
\]

respectively.

The symmetry-preserving properties show that even functions have odd Hilbert transforms and vice versa. Also real functions have real Hilbert transforms and hermitian functions have antihemitian transforms and vice versa. Moreover, complex conjugate operator commutes with Hilbert transform.

The convolution rule for Hilbert transforms is as follows:

\[
(f * g)^\# = f^\# * g = f * g^\#. \tag{20}
\]
Now we shall introduce the analytic function for the signal. Let $f$ be a suitably regular function on $\mathbb{T}_p$ with the Fourier representation
\[
f(x) = \sum_{k=-\infty}^{\infty} a_k e^{i2k\pi x/p}.
\] (21)
We consider also its Hilbert transform or conjugate function:
\[
f^\#(x) = -\frac{j}{\pi} \int_{-\infty}^{\infty} \frac{f(t)}{x-t} dt.
\]
and we combine both to form
\[
f(x) +jf^\#(x) = \sum_{k=-\infty}^{\infty} a_k [1 + \text{sgn}(k)] e^{i2k\pi x/p} =
\]
a_0 + 2 \sum_{k=1}^{\infty} a_k e^{i2k\pi x/p} = A(e^{i2\pi x/p}).
\] (23)
The function
\[
A(z) = a_0 + 2a_1 + 2a_2 + 2a_3 + \cdots
\] (24)
is called the analytic function for $f$. Since $a_{-1}, a_{-2}, \ldots$ do not appear in (24), we cannot recover $f$ from $A$.

On the other hand, we can generate a Hilbert transform pair [8]:
\[
g(x) = \frac{1}{2} [A(e^{i2\pi x/p}) + A(e^{-i2\pi x/p})]
g(x) = \frac{1}{2j} [A(e^{i2\pi x/p}) - A(e^{-i2\pi x/p})]
\] (25)
for any power series $A$ that converges on the unit circle.

4. CAUCHY PRINCIPAL VALUE INTEGRALS

Considerable effort over a long period of time has been devoted to the numerical evaluation of Cauchy principal value integrals, and this was a consequence of the important applications of Hilbert transforms. In general, numerical evaluation of weighted Cauchy principal value integrals of the form:
\[
\int_{-1}^{1} w(x) \frac{f(x)}{x-\lambda} dx = \lim_{\epsilon \to 0^+} \left( \int_{-1}^{\lambda-\epsilon} \frac{f(x)}{x-\lambda} dx + \int_{\lambda+\epsilon}^{1} \frac{f(x)}{x-\lambda} dx \right) = I_w[f;\lambda]
\]
with $\lambda \in (-1, 1)$ and $f \in C^1[-1,1]$ is a problem frequently encountered in aerodynamics, fluid and fracture mechanics and many other fields of physics and the engineering sciences. For $w = 1$ we retrieve something is called the finite Hilbert transform of $f$.

Since, in many applications, the integral has to be evaluated for a number of different values of $\lambda$, there are two basic requirements to be fulfilled by a quadrature formula:

- uniform convergence for $\lambda \in (-1, 1)$; this is especially important for the solution of singular integral equations.

Here, we are going to discuss this problem in some classes of $k$-times differentiable functions. For instance Gaussian quadrature approach is a very popular way to numerical evaluate integrals. The integrals takes the form:
\[
\int_{a}^{b} f(x) dx \approx \sum_{i=1}^{M} \Omega_i f(x_i),
\]
where $M$ is the number of sample points in the interval, which can be open or closed, $x_i$ denotes the points at which the integrand is sampled, and the $\Omega_i$ represent the weighting coefficients at the sampling points.

The simplest example of this approach are the trapezoidal rule and Simpson’s rule, but the selection of abscissa values is in an equally spaced fashion. In Gaussian quadrature schemes, the restriction of equally spaced evaluation points is dropped. This has the immediate effect of doubling the number of variables that can be used to optimize the calculation of the integral.

Usually such quadrature formulae have been developed taking into account that the integrand is a polynomial or spline function [11, 12], but this is not always the case in practical situations.

Another quadrature formulae considered usually for this problem are based on the method of substraction of singularity. One writes:
\[
I_w[f;\lambda] = \int_{-1}^{1} w(x) \frac{f(x) - f(\lambda)}{x-\lambda} dx + f(\lambda) \int_{-1}^{1} w(x) \frac{dx}{x-\lambda}.
\] (26)
and sees that the integrand on the right side of (26) is continuous because $f$ is differentiable. Hence we approximate this integral using a classical quadrature formula with respect to the weight function $w$. This method has frequently been considered by numerous authors and the resulting formulae are usually called modified quadrature formulae. Although the so-called substraction of singularity approach offers very nice theoretical convergence results under some smoothness assumptions on $f$, it is severely unstable from the numerical point of view, if $\lambda$ happens to be too close to one of the nodes used in the approximation process, and furthermore, it may be even divergent if the smoothness of $f$ is very weak [13].

Most of the classical methods are based on replacing the integrand function $f$ by some approximation $f_n$, based on $n$ function values of $f$. Such values come from an interpolating or approximating polynomial or spline. Then $I_w[f_n;\lambda]$ is used as an approximation for $I_w[f;\lambda]$. Here the choice of nodes may or may not depend on $\lambda$. Consequently we can get several computational disadvantages:

1. by choosing all the nodes independent of the location $\lambda$ in combination with spline (or, more gen-
Figure 1. The linear (1) and the logarithmic (2) domain.

eral, piecewise polynomial) approximation leads to methods with rather slow convergence;

2. for usual adaptive approach, the subdivision scheme necessarily implies that a change of $\lambda$ leads to a complete change of set of nodes, so all or almost all the functions evaluations of $f$ must be repeated. This is highly undesirable from the point of view of computing time.

5. IMPLEMENTATION COMMENTS AND REMARKS

This section is concerned with recalling several approaches to the numerical evaluation of Hilbert transforms, Kramers-Kronig transforms and Bode relationships.

In general, a large number of data may provide more accurate approximation [14]. However in situations with a small number of samples we need a certain approximation formula, otherwise the error is too large [15].

It is also convenient to know whether the samples are uniform spread in linear [16] or logarithmic domain [17] (Figure 4), as for different situations, different approaches are recommended [18]. According to [14], the best achievements can be obtained using the linear frequency domain approximation, unless when the number of gain samples is low, where the logarithmic sampling of gain is more attractive and provides better results.

When the data is noisy, it is also important if the filtering of data is used [19] or interpolation or extrapolation is needed [20]. The most popular method consists in using polynomials, though polynomials may not have any physical meaning in certain applications (for instance magnitude response in circuits with concentrated parameters).

It is also possible to replace the polynomial interpolation by other methods such as, e.g. spline interpolation, if this is desired. This may be useful if it is known that the function is very smooth in one part of the interval of integration and behaves very irregular somewhere else. Then, one may locate more nodes in the irregular part to improve the modelling of the function.

There are methods which give approximation of phase for all values of frequency, unlike other approaches which provide approximations of phase only at the frequency sampling points, and other phase values should be obtained later on by interpolation or extrapolation.

It should be noted that in the case of non minimum-phase functions, only gain information is not enough to reconstruct the signal, additional information may be necessary [21]. The supplementary information, necessary to derive a unique solution, can consist also of some time samples of the signal.

Certain methods request for special requirements. For instance, in [16], the procedure cannot be employed when the gain characteristic has slopes different from zero at zero and high frequencies. The phase characteristic of a factor in transfer function whose gain characteristic does not have the above mentioned property can be calculated alone. However, this may be a sensitive problem when only a small number of gain samples are available.

Moreover, since the data are collected over a finite frequency range, a reasonable question is whether the experimentalist can be sure that, outside the measured frequency interval he will get a reasonable representation.

Testing the results may be done with examples evaluated in closed form. The analytic solution therefore serves as valuable comparison point for numerical quadrature approach. When this is not possible for reasons as mentioned above, we have to change them. Thus the Bode functions [9], have been modified accordingly [14].

Some methods do not take into account any realizability condition. Indeed, methods where the gain is with compact support, the outcomes proposed have no mirror in real circuits, as they do not respect the Paley-Wiener Theorem. However, by employing frequency scaling, subtraction from a constant asymptote or using many similar devices we can achieve a good approximation of the gain
corresponding to a real circuit.

Difficulties may appear when the gain function has finite poles on the imaginary axis. As we usually get only finite gain samples from practical measurements, this could be avoided.

6. CONCLUSIONS

In this paper we have presented a brief review of methods used to compute Hilbert transform when the signal is composed of discrete data, sampled at equidistant or arbitrarily instants in time domain.

We have recalled how Hilbert transform, Kramers-Kronig relations or Bode relationships are known in signal processing, optics, circuit theory and control engineering. Then from mathematical point of view, a general presentation of Hilbert transform and Cauchy principal value integrals has been addressed.

We pointed out for every approach the cases when it is recommended by taking into account the situations when many samples are available or only a small number of gain samples are provided.

7. REFERENCES


