An analytic approach to minimum phase signals

Michael P. Lamoureux*, and Gary F. Margrave

ABSTRACT

The purpose of this paper is to establish the close connection between minimum phase conditions for signals and outer functions in spaces of analytic functions. The characterization through outer functions is physically motived, more precise mathematically, and opens up results from complex analysis. In particular, we show not all spectra are possible for minimum phase signals, and give alternative formulas for computing minimum phase equivalent signals.

INTRODUCTION

The minimum phase condition for signals is a useful notion in signal processing, including seismic data analysis, where in many situations, certain physical processes produce signals that have the characteristics of "minimum phase." For instance, the blast from a seismic shot (dynamite source), or the impulse from an air gun is often assumed to be minimum phase. A plausible physical argument for this observation is that in such processes, most of the energy occurs near "the beginning" of the signal, a property shared with minimum phase (see Karl (1989) and Oppenheim and Schafer (1998)). Certain data processing algorithms assume the signal under study is of this form, in order to make a more accurate recovery of that signal. Wiener spiking deconvolution, and Gabor deconvolution, are two such instances.

However, this condition is somewhat problematic. The classical definition of minimum phase (minimum phase lag) comes from linear systems theory, and it is not clear that all the properties of such *systems* can be transfered to analogous properties of *signals*. For instance, for a one-dimensional, discrete linear filter, the transfer function of the system is assumed to be in the form of a rational function (a polynomial divided by another polynomial); the minimum phase condition is then a statement about the location of zeros and poles for the rational function. For general signals, however, one cannot assume their spectra are always well-represented by rational functions; in particular, a signal does not easily reveal its zeros and poles, and such poles and zeros can move about rapidly even for small variations of the signal.

An alternate description of the minimum phase condition as "causal, stable signals with causal, stable inverse" is also less than satisfactory, as it is not obvious why a signal should even have a convolutional inverse. For instance, a finite energy signal on the real line can have no convolutional inverse, since the convolution of two finite energy signals is an integrable function, and never equals to the convolution identity, which is the Dirac delta distribution. As another instance, the discrete sampled step function

$$(1, 1, 1, \dots, 1, 0, 0, 0, \dots)$$
 (1)

^{*}Department of Mathematics and Statistics, University of Calgary

fails the "stable causal inverse" property, yet it could be considered minimum phase for physical reasons: it has most of its energy up front, and is the limiting case of a minimum phase signal $(1, r, r^2, \ldots, r^N, 0, 0, 0, \ldots)$, as r increases to 1.

Properties of minimum phase systems (rational functions) do not automatically extend to signals; for instance the phase delay of a signal may be a highly singular function, and its derivative (identified as the group delay) may not even exist. The proofs for the minimum energy delay often assume the system has a rational form, which is not appropriate for signals, and so better proofs are required.

Finally, it is common practice to compute minimum phase signals using the Hilbert transform. This can be difficult as the Hilbert transform is a singular integral typically evaluated as a Cauchy principal value and care must be taken to ensure that this is done correctly, especially in the case of computing from real random data. The "Hilbert transform pair" property of the log amplitude and phase of a minimum phase signal is known for rational functions, but care must be taken in the case of signals. Zeros in amplitude spectrum are a special problem that needs to be addressed.

In this paper, we give an alternative characterization of signals which concentrate their energy near their start. Making a connection with classical complex analysis, such a signal will be identified as precisely one whose spectrum is an outer function in a suitable Hardy space. A key property demonstrated is that the spectrum must not have too many zeros, and so it is impossible to find minimum phase signals with arbitrarily specified spectra. Stable formulas are then presented to compute the minimum phase counterpart of any causal signal.

SAMPLED SIGNALS AND ANALYTIC FUNCTIONS

We restrict our attention to time series of the form $\mathbf{f} = (\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots)$, which can be considered as a data samples of some signal. The case of signals on the real line is considered in Appendix II.

A sampled signal $\mathbf{f} = (\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots)$ is bounded if

$$|f_n| \le M \qquad \text{for all } n; \tag{2}$$

stable if

$$\sum_{n=-\infty}^{\infty} |f_n| < \infty; \tag{3}$$

finite energy if

$$\sum_{n=-\infty}^{\infty} |f_n|^2 < \infty; \tag{4}$$

and *causal* if

 $f_n = 0 \qquad \text{for all } n < 0. \tag{5}$

Given a bounded, causal signal, a complex analytic function F(z) is defined by the power

series

$$F(z) = \sum_{n=0}^{\infty} f_n z^n,$$
(6)

which converges for any complex number z = x + iy in the unit disk

$$\mathbb{D} = \{ z \in \mathbb{C} : |z| < 1 \}.$$
(7)

For stable signals, it is easy to see that the function F(z) is just an extension of the Fourier transform $\hat{\mathbf{f}}$ (or spectrum) of the signal, since its values on the boundary of the disk just gives the usual spectrum for the signal,

$$F(e^{2\pi i\theta}) = \sum_{n=0}^{\infty} f_n e^{2\pi i n\theta} = \widehat{\mathbf{f}}(\theta), \qquad 0 \le \theta \le 1,$$
(8)

where θ is the normalized frequency. The Fourier coefficients of F recover the original signal, as

$$f_n = \int_0^1 F(e^{2\pi i\theta}) e^{-2\pi i n\theta} d\theta.$$
(9)

For finite energy signals, this extension gives a square integrable function on the circle, and the Plancherel theorem indicates that the total energy is given as the integral

$$\int_{0}^{1} |F(e^{2\pi i\theta})|^{2} d\theta = \sum_{n=0}^{\infty} |f_{n}|^{2}.$$
(10)

For other causal signals (not stable, not finite energy), it is not always the case that the corresponding analytic function can be extended to the circle. It is useful to introduce the Hardy space $H^1(\mathbb{D})$ as the set of analytic functions F(z) defined from causal, bounded signals, that have the additional property that the norm

$$||F|| = \lim_{r \to 1} \int_0^1 |F(re^{2\pi i\theta})| d\theta$$
(11)

is finite.[†] These are the functions that can be extended to the circle as a simple limit $F(e^{2\pi i\theta}) = \lim_{r \to 1} F(re^{2\pi i\theta})$. Equivalently, the Hardy space $H^1(\mathbb{D})$ can be defined as the set of integrable functions on the circle, whose negative Fourier coefficients all are identically zero.

The conclusion of this section is that causal signals are very well described by analytic function in a Hardy space, of which there is much known mathematically.

[†]It is possible to define more general Hardy spaces $H^p(\mathbb{D})$ for any number $1 \le p \le \infty$ but we will not need them here.

CAUSAL SIGNALS AND SPECTRUM

In many geophysical applications, we wish to find a causal, minimum phase signal with a given amplitude spectrum. For instance, in deconvolution, the source signature wavelet is found as the minimum phase signal with amplitude spectrum determined by a smoothing of the amplitude spectrum of recorded seismic data.

What is often ignored in practice is that there restrictions on what the amplitude spectrum can be. For a causal signal, there cannot be too many zeros. in practice, it is understood that filling in the zeros with a small constant gives a computable function. But it is important to note that this is more than a computational difficulty: certain amplitude spectra do not come from causal signals. For instance, it is impossible to have a perfectly bandlimited amplitude spectrum for a causal signal, even for a signal of infinite length. That would mean, for instance, that it is impossible to find a causal signal whose spectrum corresponds to a band limited delta spike. The reason is the following theorem:

Theorem 1 Suppose $\mathbf{f} = (f_0, f_1, f_2, ...)$ is a bounded causal signal with corresponding analytic function $F(z) = \sum f_n z^n$ in the Hardy space $H^1(\mathbb{D})$. Then the log of the amplitude spectrum of the signal

$$\log|\widehat{f}(\theta)| = \log|F(e^{2\pi i\theta})| \tag{12}$$

is integrable.

The details are in Hoffman (1962) and Helson (1995). The point is that if the spectrum is zero on, say, an interval, then the log spectrum is minus infinity on an interval, and that cannot be integrated. On the other hand, a few isolated zeroes in the spectrum are not a problem since a logarithmic singularity is integrable.

As an example of the confusion that can arise, we look at computing the minimum phase version of a band limited delta spike. This can be done directly in MATLAB using the "rceps" function, or by other methods involving a numerical Hilbert transform. The problem is that the log amplitude spectrum is used in the calculation, which equals minus infinity at many points, which interferes with the calculation. A common fix is to replace $\log |\hat{f}(\theta)|$ with $\log(|\hat{f}(\theta)| + \epsilon)$, where $\epsilon > 0$ is a small stabilization parameter.

However, as we see in Figure 1, as this stabilization parameter decreases to zero, the resulting signal does not converge to a stable result. In fact, as $\epsilon \to 0$, the major bump in the signal moves off to the right. The problem is not in choosing the right stabilization; the problem is that no causal signal can have that prescribed spectrum with an interval of zeroes.

FRONT LOADED SIGNALS

The key physical property of minimum phase signals that makes them useful in applications is that the energy is concentrated near the front of the signal. Here, we make that our main definition, and show that the usual examples of minimum phase signals are covered



FIG. 1. Four versions of a stabilized min phase signal, using stabilization values .1, .001, .00001, .0000001.

by this case. We also show it is a more general notion that covers more signals, and has an extremely useful characterization in Hardy spaces.

Definition 2 We say a causal signal $\mathbf{f} = (f_0, f_1, f_2, ...)$ is front loaded if its partial energies are maximized, relative to any other casual signal with the same amplitude spectrum.

That is, if $\mathbf{g} = (g_0, g_1, g_2, ...)$ is any other causal signal with amplitude spectrum $|\widehat{\mathbf{g}}(\theta)| = |\widehat{\mathbf{f}}(\theta)|$, then

$$\sum_{n=0}^{N} |g_n|^2 \le \sum_{n=0}^{N} |f_n|^2, \quad \text{for each } N = 0, 1, 2, \dots$$
(13)

As a simple example, consider the causal signals $\mathbf{f} = (2, 1, 0, 0, 0, ...)$ and $\mathbf{g} = (1, 2, 0, 0, 0, ...)$. The signals have the same amplitude spectra, but \mathbf{f} has more energy in the 0-th entry than does \mathbf{g} , so it seems more front loaded. In fact, it is more front loaded than any signal with the same amplitude spectrum.

Our main mathematical result is that these front loaded signals correspond to a wellunderstood class of functions in Hardy space.

Theorem 3 Suppose $\mathbf{f} = (f_0, f_1, f_2, ...)$ is a bounded, causal signal with corresponding analytic function F(z) in the Hardy space $H^1(\mathbb{D})$. Then \mathbf{f} is front loaded if and only if F(z) is an outer function.

That is, F(z) can be written in the form

$$F(z) = \lambda \exp\left(\int_0^1 \frac{e^{2\pi i\theta} + z}{e^{2\pi i\theta} - z} u(e^{2\pi i\theta}) d\theta\right)$$
(14)

where u is a real-valued integrable function on the unit circle, and λ is a complex number of modulus one.

The details are given in Appendix I. The point is, front loaded signals precisely correspond to outer functions, which are given by a specific integral. The function u in the integral is just the log amplitude spectrum of the signal, so we have a precise formula to compute the minimum phase equivalent of any causal signal.

We also note in the following theorem that front loaded generalizes the usual notion of minimum phase.

Theorem 4 Suppose $\mathbf{f} = (f_0, f_1, f_2, ...)$ is a causal stable signal with causal stable inverse. Then \mathbf{f} is front loaded and the analytic function F(z) is outer.

This theorem follows immediately from an exercise in Hoffman (1962): the signal f defines an analytic function F(z) and the inverse signal has analytic function 1/F(z). Since both F(z) and its reciprocal are in the Hardy space, that function must be outer.

For filters, the transfer function is a rational function in z^{-1} and is precisely $F(z^{-1})$. The causal stable filter, with causal stable inverse, is equivalent to having $F(z^{-1})$ with all its zeros and poles of that inside the unit disk.

It is worth noting at least one example that shows the front loaded condition is more general than the "causal stable signal with casual stable inverse." The signal $\mathbf{f} = (1, 1, 0, 0, 0, ...)$ is casual and stable, but its convolutional inverse $\mathbf{f}^{-1} = (1, -1, 1, -1, 1, -1, ...)$ is not stable. So \mathbf{f} is not strictly speaking a minimum phase signal in the classical sense. But it is front loaded, the power series function F(z) = 1 + z is indeed outer, and thus \mathbf{f} can be reconstructed directly from its amplitude spectrum.

HILBERT TRANSFORM CALCULATION OF FRONT LOADED SIGNALS

We show how to compute a front loaded signal using a the Hilbert transform. The main steps are as follows. Given a causal signal g with amplitude spectrum $|\hat{\mathbf{g}}(\theta)|$, we define two real-valued functions u, v on the circle as

$$u(e^{2\pi i\theta}) = \log |\widehat{\mathbf{g}}(\theta)|$$
$$v(e^{2\pi i\theta}) = \int_0^1 \cot(\pi(\theta - \phi))u(e^{2\pi i\phi})d\phi$$

where we see v is just the convolution of u with the cotangent function; this is Hilbert transform on the unit circle[‡]. The equivalent front end loaded signal **f** has spectrum $\hat{\mathbf{f}}(\theta) = \exp\left(u(e^{2\pi i\theta}) + iv(e^{2\pi i\theta})\right)$, and the signal **f** itself can be recovered using the inverse Fourier transform.

The reason this works is as follows. The front loaded signal f will have the same amplitude spectrum as g and so the real-valued function defining the corresponding outer function F(z) is given by

$$u(e^{2\pi i\theta}) = \log |\widehat{\mathbf{g}}(\theta)|. \tag{15}$$

The outer function is now expressed as

$$F(z) = \exp\left(\int_0^1 \frac{e^{2\pi i\theta} + z}{e^{2\pi i\theta} - z} u(e^{2\pi i\theta}) d\theta\right),\tag{16}$$

and so we let $\log F(z) = u(z) + iv(z)$ with

$$u(z) = \int_0^1 Re\left[\frac{e^{2\pi i\theta} + z}{e^{2\pi i\theta} - z}\right] u(e^{2\pi i\theta}) d\theta,$$

$$v(z) = \int_0^1 Im\left[\frac{e^{2\pi i\theta} + z}{e^{2\pi i\theta} - z}\right] u(e^{2\pi i\theta}) d\theta,$$

[‡]In the Fourier transform domain, the Hilbert transform acts as multiplication by i * sgn(n), the Fourier series coefficients of the cotangent.

where u(z), v(z) are conjugate harmonic functions on the unit disk. Setting $z = re^{2\pi i\phi}$ and letting $r \to 1$ we see in the integral for $v(z) = v(re^{2\pi i\phi})$ that

$$Im\left[\frac{e^{2\pi i\theta} + re^{2\pi i\phi}}{e^{2\pi i\theta} - re^{2\pi i\phi}}\right] = \frac{2r\sin(2\pi(\phi - \theta))}{1 - 2r\cos(2\pi(\phi - \theta)) - r^2} \to \cot(\pi(\phi - \theta))$$
(17)

as $r \to 1$. And that $v(e^{2\pi i\phi})$ is given by the convolution of u with the cotangent, which is the Hilbert transform.

NON-SINGULAR INTEGRAL FORMULA FOR A DISCRETE, MINIMUM PHASE SIGNAL

Given an outer function F(z), the Fourier coefficients can be computed by integrating along an interior circle of the unit disk, so

$$f_n = \frac{1}{r^n} \int_0^1 F(re^{2\pi i\phi}) e^{-2\pi i n\phi} \, d\phi, \qquad \text{any } r < 1 \;. \tag{18}$$

This is not singular, since the function F(z) is continuous on the interior circle. From the previous section, one obtains a formula for F(z) at interior points, and thus

$$f_n = \frac{1}{r^n} \int_0^1 \exp\left(\int_0^1 \frac{e^{2\pi i\theta} + re^{2\pi i\phi}}{e^{2\pi i\theta} - re^{2\pi i\phi}} \log|F(e^{2\pi i\theta})| \, d\theta\right) e^{-2\pi i n\theta} \, d\phi, \tag{19}$$

where $|G(e^{i\theta})|$ is the amplitude spectrum of the desired signal, assumed to be known. Provided F does not vanish on the circle, this double integral is not singular.

In a numerical calculation, the desired amplitude spectrum |F| is either given, or computed from sample data using the FFT. The inner integral is approximated by a sum at the given discrete points;. The exponential is computed at these discrete points, and an inverse FFT used to compute the f_n .

In the case of a zero in the spectrum, it might be best just to factor out that zero and do the reduced calculation.

SUMMARY

We have shown a more general definition of minimum phase signals, which we call front loaded signals, referring to the essential physical property that such causal signals maximize the amount of energy at the beginning (front) of the signal. Mathematically, this is equivalent to the property that the Fourier transform (or spectrum) of the signal is an outer function in a Hardy space of analytic functions.

We also note that both causal signals and minimum phase signals have restrictions on their spectrum. In particular, one cannot take an arbitrary amplitude spectrum and hope to find the minimum phase signal corresponding to it. There is an essential restriction on the density of zeros in the amplitude spectrum.

The Hilbert transform can be used to compute the front loaded equivalent of any causal signal. We also present alternative, non-singular integral formulas for computing the front

loaded equivalent, obtained by analytically extending the Fourier transform of a causal signal to the interior of the unit disk in the complex plane.

In summary, we have produced a mathematically rigourous definition of minimum phase signals that is both physically meaningful and computationally useful.

ACKNOWLEDGMENTS

This research is a product of the POTSI group. The authors thank the industrial sponsors of POTSI and CREWES, and the MITACS and NSERC agencies for their support.

REFERENCES

Helson, H., 1995, Harmonic Analysis: Hindustan Book Agency.

Hoffman, K., 1962, Banach Spaces of Analytic Functions: Prentice-Hall, New Jersey.

Karl, J., 1989, An introduction to digital signal processing: Academic Press, San Diego.

Körner, T., 1988, Fourier Analysis: Cambridge University Press.

Oppenheim, A. V., and Schafer, R. W., 1998, Discrete-time Signal Processing: Prentice Hall, New Jersey.

APPENDIX I: PROOFS

Outer functions in the Hardy space $H^1(\mathbb{D})$ were defined in Theorem 3. The following important theorem characterizes some key properties of outer functions.

Theorem 5 Let F be a nonzero function in $H^1(\mathbb{D})$. The following are equivalent:

- *i) F is an outer function;*
- *ii*) $\log |F(0)| = \int_0^1 \log |F(e^{2\pi i\theta})| d\theta$;
- *iii) the functions* $z^n F(z)$, $n \ge 0$, span $H^1(\mathbb{D})$;
- *iv)* F is an extreme point in ball of radius $R = ||F||_1$ in $H^1(\mathbb{D})$.
- v) If G is any other function in $H^1(\mathbb{D})$ such that |G| = |F| almost everywhere on the unit circle, then $|G(z)| \le |F(z)|$ for each point z in the open unit disk;
- *vi) if* G *is any function in* $H^1(\mathbb{D})$ *with* G/F *integrable on the unit circle, then* G/F *is in* $H^1(\mathbb{D})$ *;*

A bounded analytic function G(z) is called *inner* if $|G(e^{i\theta})| = 1$ a.e. on the unit circle.

Every analytic function in $H^1(\mathbb{D})$ can be factored into a product of an inner function times an outer function, unique up to a multiplicative factor of modulus one. The inner function contains the zeros of the function, in the precise sense that it is the product of a Blaschke product, and a singular function. Any analytic function G(z) with the same

These results are stated in Hoffman, or offered as exercises.

We now prove that outer implies front loaded, and conversely.

Theorem 6 Suppose f is a causal signal with F(z) is outer. Then f is front loaded.

Proof: We must show that if g is another signal with the same amplitude spectrum, then for each $N \ge 0$, the partial energies satisfy

$$\sum_{n=0}^{N} |g_n|^2 \le \sum_{n=0}^{N} |f_n|^2, \quad \text{for all } N,$$
(20)

From item vi) of Theorem 5, the function C(z) = G(z)/F(z) is an inner function in $H^1(\mathbb{D})$. Since $|C(e^{i\theta})| = 1$ a.e., the inner function is square-integrable on the circle and corresponds to a square summable causal sequence c. Fix $N \ge 0$ and let $\tilde{\mathbf{f}}$ be the truncated sequence

$$\tilde{\mathbf{f}} = (f_0, f_1, \dots, f_{N-1}, f_N, 0, 0, 0, \dots)$$
 (21)

and set $\tilde{\mathbf{g}} = \mathbf{c} * \tilde{\mathbf{f}}$. Since \mathbf{f} and \mathbf{c} are causal, the full and truncated convolutions agree for the first N + 1 terms, so one has

$$\tilde{g}_n = g_n, \text{ for } 0 \le n \le N.$$
(22)

Thus, the partial energies satisfy

$$\sum_{n=0}^{N} |g_n|^2 = \sum_{n=0}^{N} |\tilde{g}_n|^2 \le \sum_{n=0}^{\infty} |\tilde{g}_n|^2 = \sum_{n=0}^{\infty} |\tilde{f}_n|^2 = \sum_{n=0}^{N} |f_n|^2$$
(23)

where the equality in the middle follows from the Plancheral theorem on $L^2(\mathbb{T})$, noting that $|\tilde{G}| = |C\tilde{F}| = |\tilde{F}|$ a.e. on the unit circle. QED

The converse is quite easy:

Theorem 7 Suppose f is a front loaded signal. Then F(z) is outer.

Proof: Let G(z) be an outer function in $H^1(\mathbb{D})$ with |G| = |F| on the unit circle, corresponding to a casual sequence g. Since f is front loaded, its partial energy at zero is bigger than that for g, so

$$|F(0)| = |f_0| \ge |g_0| = |G(0)|.$$
(24)

On the other hand, by v) of Theorem 5, the outer function G(z) satisfies

$$|G(0) \ge |F(0)|. \tag{25}$$

Combining the two inequalities shows |F(0)| = |G(0)| and thus by ii) of Theorem 5, we have

$$\log |F(0)| = \log |G(0)| = \int_0^1 \log |F(e^{2\pi i\theta})| d\theta.$$
(26)

Hence F(z) is outer. QED

APPENDIX II: CONTINUOUS TIME SIGNALS

This work generalizes to continuous signals on the real line. The key ideas are that the Fourier transform of a causal signal $f : \mathbb{R} \to \mathbb{C}$ extends to an analytic function on half the complex plane by

$$F(z) = \int_{0}^{\infty} f(t)e^{2\pi i t z} dt, \qquad Im(z) > 0,$$
(27)

since the exponential decay cause by the choice of Im(z) > 0 makes the integral converge nicely. The Hardy space $H^2(\mathbb{R})$ is defined as the image of the square integrable causal functions under this transform. Again we have inner and outer functions, and the various factorization theorems.

We show here that a causal signal whose Fourier transform is outer, is necessarily front loaded.

Theorem 8 Suppose f is a causal signal in $L^2(\mathbb{R})$ with Fourier transform an outer function in $H^2(\mathbb{R})$), and g is another causal signal with the same amplitude spectrum. Then for each $T \ge 0$, the partial energies satisfy

$$\int_{0}^{T} |g(t)|^{2} dt \leq \int_{0}^{T} |f(t)|^{2} dt.$$
(28)

Proof: Since g the same amplitude spectrum as f, it is also in $L^2(\mathbb{R}^+)$. By the factorization theorem, the corresponding Fourier transform functions F(z), G(z) in the Hardy space are related by an inner function C(z), with G(z) = C(z)F(z). Since |C(x)| = 1 a.e. on the real line, the inner function is in H^{∞} and corresponds to a distribution c(t) on the positive real line. Fix $T \ge 0$ and let \tilde{f} be the truncated signal

$$\tilde{f}(t) = f(t) \quad \text{for } 0 \le t \le T,$$
(29)

$$= 0$$
 otherwise, (30)

and set $\tilde{g} = c * \tilde{f}$. Since f and c are causal, the full and truncated convolutions agree up to T, so we have

$$\tilde{g}(t) = g(t), \text{ for } 0 \le t \le T$$
 . (31)

Thus, the partial energies satisfy

$$\int_{0}^{T} |g(t)|^{2} dt = \int_{0}^{T} |\tilde{g}(t)|^{2} dt \le \int_{0}^{\infty} |\tilde{g}(t)|^{2} dt = \int_{0}^{\infty} |\tilde{f}(t)|^{2} dt = \int_{0}^{T} |f(t)|^{2} dt \quad (32)$$

where the equality in the middle follows from the Plancheral theorem for $H^2(\mathbb{R})$ functions, noting that $|\tilde{G}| = |C\tilde{F}| = |\tilde{F}|$ a.e. on the real line. QED

At this point, we are not able to show that front loaded implies outer.