Reprinted from SPIE Vol. 359—Applications of Digital Image Processing IV

Signal reconstruction from Fourier transform amplitude

Patrick L. Van Hove, Jae S. Lim, Alan V. Oppenheim
Massachusetts Institute of Technology, Research Laboratory of Electronics
Department of Electrical Engineering and Computer Science
Room 36-615, Cambridge, Massachusetts 02139

Abstract

In this paper, we show that a one-dimensional or multi-dimensional sequence is uniquely specified under mild restrictions by its Fourier transform amplitude (magnitude and one bit of phase information). In addition, we develop a numerical algorithm to reconstruct a one-dimensional or multi-dimensional sequence from its Fourier transform amplitude. Reconstruction examples obtained using this algorithm are also provided.

I. Introduction

In a variety of contexts, such as electron microscopy[1], x-ray crystallography[2], optics[3], and Fourier transform signal coding[4], it is desirable to reconstruct a sequence from partial Fourier domain information. As a consequence, considerable attention has been paid to, and some significant results have been developed in this area. For example, it has been previously established[5] that under very mild restrictions a finite extent one-dimensional (1-D) or multi-dimensional (M-D) sequence is uniquely specified to within a scale factor by its Fourier transform (FT) phase, and algorithms for implementing the reconstruction have been developed. It is well known that in contrast, the FT magnitude does not uniquely specify a 1-D sequence. Even for M-D sequences, the FT magnitude specifies a sequence only to within a translation and a central symmetry[5], and reconstruction algorithms developed so far have been successful[6] for only a very restricted class of M-D sequences.

In this paper we consider the reconstruction of 1-D and M-D sequences when the Fourier transform magnitude and one bit of phase information is known. In particular, it is shown that under very mild restrictions, this is sufficient to uniquely specify the sequence.

In Section II of this paper, the basic theory is presented. In Section III an algorithm for implementing the reconstruction is discussed, and Section IV illustrates several examples.

II. Theory

In this section, we discuss the unique specification of a sequence by its FT magnitude and 1 bit of phase. We initially consider the one-dimensional (1-D) case first and then extend the 1-D result to the multi-dimensional (M-D) case. Before we present the theoretical results, we define the notation that will be used throughout the paper.

Let $x(n)$ denote a 1-D sequence which is <u>causal</u> and <u>finite extent</u> so that $x(n)$ is zero outside $0 \le n \le L-1$. Furthermore we restrict $x(n)$ to be real-valued. Let $X(z)$ and $X(\omega)$ represent the z-transform and Fourier transforms of $x(n)$, so that

$$X(z) = \sum_{n=0}^{L-1} x(n) z^{-n} \tag{1}$$

$$X(\omega) = X(z)\Big|_{z=e^{j\omega}} = \sum_{n=0}^{L-1} x(n) e^{-j\omega n} \tag{2}$$

The Fourier transform $X(\omega)$ can be represented in terms of its real part $X_R(\omega)$ and imaginary part $X_I(\omega)$, or in terms of its magnitude $|X(\omega)|$ and phase $\theta_x(\omega)$ as follows:

$$X(\omega) = X_R(\omega) + j\, X_I(\omega) = |X(\omega)| e^{j\,\theta_x(\omega)} \tag{3}$$

To ensure that $\theta_x(\omega)$ is well defined at all ω, we assume that $X(z)$ has no zeros on the unit circle. The phase function $\theta_x(\omega)$ in equation (3) represents the principal value of the phase so that

$$-\pi < \theta_x(\omega) \le \pi \tag{4}$$

The one-bit FT phase information will be represented by the function $S_x^{\alpha}(\omega)$ defined as

$$S_x^{\alpha}(\omega) = \begin{cases} +1 & \alpha-\pi \le \theta_x(\omega) \le \alpha \\ -1 & \text{otherwise} \end{cases} \tag{5}$$

where α is a known constant in the range of $0 < \alpha \le \pi$. Thus, the complex plane is divided into two regions separated by a straight line passing through the origin and at an angle α with the real axis, as shown in Figure 1. For example, for $\alpha = \frac{\pi}{2}$, $S_x^{\pi/2}(\omega)$ represents the algebraic sign of $\mathrm{Re}\{X(e^{j\omega})\}$. More generally, $S_x^{\alpha}(\omega)$ is the algebraic sign of $\mathrm{Re}\{e^{j(\pi/2-\alpha)}X(\omega)\}$. The algebraic sign of zero is assumed to be positive.

The function $G_x^{\alpha}(\omega)$ is defined as

$$G_x^{\alpha}(\omega) = S_x^{\alpha}(\omega)\ |X(\omega)| \tag{6}$$

and will be referred to as the Fourier transform amplitude since it contains both magnitude and sign information. An example of $|X(\omega)|$, $\theta_x(\omega)$, $S_x^{\alpha}(\omega)$ and $G_x^{\alpha}(\omega)$ when $\alpha=\pi/2$ and $X(z)=1+3z^{-1}+5z^{-2}+2z^{-3}$ is shown in Figure 2.

Finally, given a positive integer N, we define a constant P and an interval R as:

$$P = \frac{N-1}{2} \quad \text{and } R = (0,\pi) \text{ for N odd}$$
$$P = \frac{N}{2} \quad \text{and } R = (0,\pi] \text{ for N even} \tag{7}$$

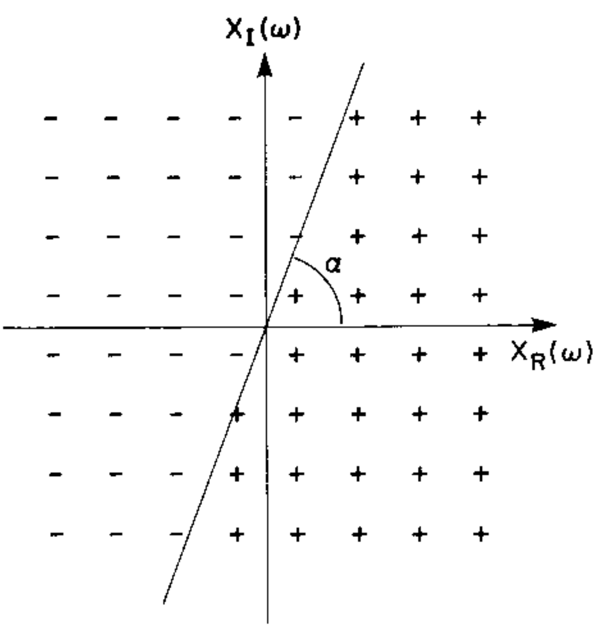

Figure 1: Mapping of the 1-bit phase function

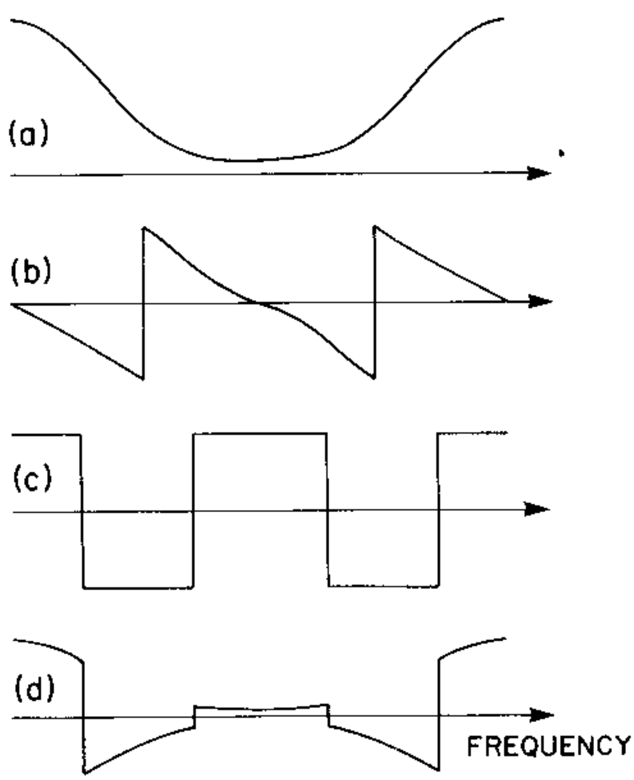

Figure 2: Fourier transform magnitude, phase, 1-bit phase, amplitude of the sequence $X(z)=1+3z^{-1}+5z^{-2}+2z^{-3}$

The uniqueness of a 1-D sequence when the Fourier transform amplitude $G_x^{\alpha}(\omega)$ is specified, is based on the following statements. The proof of these statements is given in the Appendix.

Statement A1

Let x(n) and y(n) be two real, causal, and finite extent sequences. If $|X(\omega)| = |Y(\omega)|$, x(n) and y(n) can always be expressed as

$$x(n) = b(n) * a(n)$$

and

$$y(n) = \varepsilon b(n) * a(N-1-n),$$

where $\varepsilon=+1$ or -1 and a(n) and b(n) are real, causal and finite extent sequences with N corresponding to the length of a(n), i.e. a(n)=0 outside $0 \le n \le N-1$.

Statement A2

Let b(n) be a real, causal, and finite extent sequence. For any positive integer N, the equation

$$\mathrm{Re}\left\{ B(z)\, z^{-\frac{N-1}{2}} \Big|_{z=e^{j\omega}} \right\} = 0$$

is satisfied for at least P distinct values of ω in the interval R, where P and R are as defined in eq. (7).

Statement A3

Let a(n) be a real sequence which is zero outside $0 \le n \le N-1$. If the equation

$$\mathrm{Im}\left\{ A(z)\, z^{\frac{N-1}{2}} \Big|_{z=e^{j\omega}} \right\} = 0$$

is satisfied for at least P distinct values of ω in the interval R, then it is identically equal to zero and a(n)=a(N-1-n).

We use the above three statements, whose proofs are shown in the Appendix, to demonstrate the following statement:

Statement 1

Let x(n) and y(n) be two real, causal, and finite extent sequences with z-transforms which have no zeros on the unit circle. If $G_x^{\pi/2}(\omega)=G_y^{\pi/2}(\omega)$ for all ω, then x(n)=y(n).

To show Statement 1, we note from equations (5) and (6) that the condition $G_x^{\pi/2}(\omega)=G_y^{\pi/2}(\omega)$ is equivalent to

$$\mathrm{sign}\{X_R(\omega)\}\ |X(\omega)| = \mathrm{sign}\ \{Y_R(\omega)\}\ |Y(\omega)| \tag{8}$$

which in turn implies that $|X(\omega)|=|Y(\omega)|$ and therefore that

$$\mathrm{sign}\{X_R(\omega)\} = \mathrm{sign}\{Y_R(\omega)\} \tag{9}$$

From Statement A1, then, x(n) and y(n) can be expressed as

$$\begin{aligned} x(n) &= b(n) * a(n) \\ y(n) &= \varepsilon\ b(n) * a(N-1-n) \end{aligned} \tag{10}$$

where $\varepsilon=\pm 1$. Fourier transforming (10), we obtain:

$$\begin{aligned} X(\omega) &= A(\omega)\ B(\omega) \\ Y(\omega) &= \varepsilon\ e^{-j\omega(N-1)}\ A(-\omega)\ B(\omega) \end{aligned} \tag{11}$$

To show that $\varepsilon=1$ in (11), we evaluate eq. (9) at $\omega=0$ and recognize that $X_R(0)=A(0)B(0)$ and $Y_R(0)=\varepsilon A(0)B(0)$,

$$\mathrm{sign}\{A(0)B(0)\} = \mathrm{sign}\{\varepsilon A(0)B(0)\}. \tag{12}$$

Since $X(\omega)$ is not zero at $\omega=0$, eq. (12) requires that $\varepsilon=+1$.

Since $\varepsilon=1$, from (10), showing that $x(n)=y(n)$ is equivalent to showing that $a(n)=a(N-1-n)$. Toward this end, we consider the sum

$$X_R(\omega) + Y_R(\omega)$$

From (11) with $\varepsilon=1$, it can be shown that

$$X_R(\omega) + Y_R(\omega) = 2\ \mathrm{Re}\left[A(\omega)\ e^{j\omega\frac{N-1}{2}}\right]\ \mathrm{Re}\left[B(\omega)\ e^{-j\omega\frac{N-1}{2}}\right] \tag{13}$$

From Statement A2, there are at least P distinct values of ω in the interval R which we denote as ω_i $i=1,2,\ldots,P$ for which

$$\mathrm{Re}\left[B(\omega_i)e^{-j\omega_i\frac{N-1}{2}}\right] = 0\ ,\quad i=1,2,\ldots,P,\ \omega_i\varepsilon R \tag{14}$$

From (13) and (14)

$$X_R(\omega_i) + Y_R(\omega_i) = 0\ ,\qquad i=1,2,\ldots,P,\quad \omega_i\varepsilon R \tag{15}$$

From (9), both terms of the left hand side of (15) have the same sign for all ω. Since a sum of two terms having the same sign can be zero only when both terms are zero, we have

$$X_R(\omega_i) = Y_R(\omega_i) = 0 \quad \text{and therefore also,}$$

$$X_R(\omega_i) - Y_R(\omega_i) = 0,\qquad i=1,2,\ldots,P,\quad \omega_i\varepsilon R \tag{16}$$

From (11) and the fact that $\varepsilon=1$, it can be shown that (16) can be expressed as

$$X_R(\omega_i) - Y_R(\omega_i) = -2\ \mathrm{Im}\left[A(\omega_i)\ e^{j\omega_i\frac{N-1}{2}}\right]\ \mathrm{Im}\left[B(\omega_i)\ e^{-j\omega_i\frac{N-1}{2}}\right] = 0,\ \ i=1,2,\ldots,P,\ \ \omega_i\varepsilon R \tag{17}$$

Since $B(\omega)$ is not zero for any ω, it follows from (14) that the second factor in (17) satisfies the property:

$$\mathrm{Im}\left[B(\omega_i)e^{-j\omega_i\frac{N-1}{2}}\right] \neq 0\ ,\quad i=1,2,\ldots,P,\quad \omega_i\varepsilon R \tag{18}$$

From (17) and (18),

$$\mathrm{Im}\left[A(\omega_i)e^{j\omega_i\frac{N-1}{2}}\right] = 0\ ,\quad i=1,2,\ldots,P,\quad \omega_i\varepsilon R \tag{19}$$

From (19) and Statement A3, $a(n)\equiv a(N-1-n)$ so that $x(n)\equiv y(n)$, thus demonstrating Statement 1.

The result in Statement 1 can be generalized in various ways. Specifically, in Statement 1, we have assumed that $\alpha=\pi/2$, which is a specific representation of the 1 bit phase information. It can be shown that the statement is true for other choices of $0<\alpha<\pi$. When $\alpha=\pi$ so that $S_x^{\pi}(\omega)=\mathrm{sign}[\theta_x(\omega)]$, a sequence is uniquely specified by $G_x^{\pi}(\omega)$ when $x(0)=0$. Statement 1 can also be extended to anti-causal (left-sided) sequences. The proofs of these extensions can be found in[10]. When the above extensions are incorporated in Statement 1, we have the following general statement:

Statement 2

Let $x(n)$ and $y(n)$ be two real, causal (or anti-causal), and finite extent sequences, with z-transforms which have no zeros on the unit circle. If $G_x^{\alpha}(\omega)=G_y^{\alpha}(\omega)$ for all ω and $0<\alpha<\pi$, then $x(n)=y(n)$. When $\alpha=\pi$, if $G_x^{\pi}(\omega)=G_y^{\pi}(\omega)$ and $x(0)=y(0)=0$, then $x(n)=y(n)$.

Statements 1 and 2 explicitly require that the sequences be real-valued and causal (or anti-causal). The necessity of these conditions can be illustrated through counter-examples. Consider first the condition that the sequences be real, and let $y(n)$ equal $e^{j(\alpha-\pi)}x(n)$ where $x(n)$ is real. In this case, it is straightforward to show that $G_x^{\pi}(\omega)=G_y^{\alpha}(\omega)$. Since $G_x^{\pi}(\omega)$ does not uniquely specify $x(n)$, $G_y^{\alpha}(\omega)$ does not uniquely specify $y(n)$. To indicate the necessity of the causality (or anti-causality) condition, consider as one counter-example the two-sided sequences $x(n)$ and $y(n)$ for which the z-transforms are

$$X(z) = -z^2 + 6 - z^{-2} = (z + 2 - z^{-1})(-z + 2 + z^{-1})$$
$$Y(z) = z^2 + 4z + 2 - 4z^{-1} + z^{-2} = (z + 2 - z^{-1})^2 \tag{20}$$

For these two sequences it can be easily shown that $|X(\omega)|=|Y(\omega)|$ and $S_x^{\pi/2}(\omega)=S_y^{\pi/2}(\omega)$. In this case, then, $x(n)$ and $y(n)$ are different sequences but have the same FT amplitude.

In the above discussion, we considered only 1-D sequences. We now extend Statement 2 to M-D sequences. Let $x(\underline{n})$ denote a M-D sequence $x(n_1,n_2,\dots,n_M)$, and let $G_x^{\alpha}(\underline{\omega})$ denote the FT amplitude of $x(\underline{n})$, where $G_x^{\alpha}(\underline{\omega})$ represents $G_x^{\alpha}(\omega_1,\omega_2,\dots,\omega_M)$ and is given by $S_x^{\alpha}(\underline{\omega})|X(\underline{\omega})|$. We define an M-dimensional signal $x(\underline{n})$ to have a one-sided region of support in the M-dimensional space $n_1,n_2,\dots,n_M$ if it only has non-zero values for one polarity of each index n_i. For example, for a two-dimensional sequence there are four possible regions of support which are consistent with the sequence being one-sided, corresponding to the four quadrants. Statement 3, which follows, represents a generalization of Statement 2 to encompass M-D sequences.

<u>Statement 3</u>

Let $x(\underline{n})$ and $y(\underline{n})$ be two real, finite extent sequences with one-sided support and with z-transforms which have no zeros at $|\underline{z}|=1$. If $G_x^{\alpha}(\underline{\omega})=G_y^{\alpha}(\underline{\omega})$ for all $\underline{\omega}$ and $0<\alpha<\pi$, then $x(\underline{n})=y(\underline{n})$. When $\alpha=\pi$, if $G_x^{\pi}(\underline{\omega})=G_y^{\pi}(\underline{\omega})$ and $x(\underline{0})=y(\underline{0})=0$, then $x(\underline{n})=y(\underline{n})$.

We demonstrate the validity of Statement 3 for a 2-D sequence which has the first-quadrant support with size $M_1 \times M_2$ so that

$$x(n_1,n_2) = y(n_1,n_2) = 0 \quad \text{outside} \quad 0\le n_1\le M_1-1 \text{ and } 0\le n_2\le M_2-1$$

The proof for a higher dimension and for a different quadrant support is analogous to the 2-D case with the first quadrant support. To demonstrate Statement 3, we map the 2-D sequences $x(n_1,n_2)$ and $y(n_1,n_2)$ into two 1-D sequences $\hat{x}(n)$ and $\hat{y}(n)$ by the following transformation:

$$\hat{x}(n_1 \cdot M_2 + n_2) = x(n_1,n_2)$$
$$\hat{y}(n_1 \cdot M_2 + n_2) = y(n_1,n_2) \tag{21}$$

In essence, the transformation in eq. (21) corresponds to mapping a 2-D sequence to a 1-D sequence by concatenating the columns of the 2-D sequence. Clearly, $\hat{x}(n)$ and $\hat{y}(n)$ given by (21) are real, causal, finite extent sequences. From (21) it is clear that the transformation is invertible. Furthermore, it can be shown that

$$\hat{X}(\omega) = X(\omega_1,\omega_2)\Big|_{\omega_1=\omega\cdot M_2,\ \omega_2=\omega}$$

and

$$\hat{Y}(\omega) = Y(\omega_1,\omega_2)\Big|_{\omega_1=\omega\cdot M_2,\ \omega_2=\omega} \tag{22}$$

From (22), it follows that the FT amplitudes of $\hat{x}(n)$ and $\hat{y}(n)$ are specified by the FT amplitudes of $x(n_1,n_2)$ and $y(n_1,n_2)$. Therefore, if $G_x^{\alpha}(\omega_1,\omega_2)=G_y^{\alpha}(\omega_1,\omega_2)$, then $G_{\hat{x}}^{\alpha}(\omega)=G_{\hat{y}}^{\alpha}(\omega)$. In addition, since $X(z_1,z_2)$ and $Y(z_1,z_2)$ have no zeros at $|z_1|=|z_2|=1$, from (22), $\hat{X}(z)$ and $\hat{Y}(z)$ have no zeros on the unit circle. Since $\hat{x}(n)$ and $\hat{y}(n)$ satisfy all the conditions in Statement 2, it follows from Statement 2 that $\hat{x}(n)=\hat{y}(n)$. Since the transformation (21) is invertible, $x(n_1,n_2)=y(n_1,n_2)$ as required by Statement 3.

The theoretical result in Statement 3 differs from that by Hayes[5] in several respects. In the result by Hayes[5], only samples of the FT magnitude are required, but the sequence is restricted to have a non-factorizable z-transform and the unique specification of the sequence is only to within a sign, a translation, and a central symmetry. In Statement 3, the FT amplitude is required, but the sequence may have a factorizable z-transform and is uniquely specified in the strict sense.

III. Algorithm

In Section II, we showed that under certain conditions, a sequence is uniquely specified by its FT amplitude. In this section, we discuss an algorithm to implement the reconstruction of a sequence $x(\underline{n})$ from its FT amplitude. The sequence $x(\underline{n})$ is assumed to satisfy the conditions of Statement 3. In addition, its FT amplitude $G_x^{\alpha}(\underline{\omega})$ is assumed known.

The algorithm that we have developed is an iterative procedure which is similar in style to other iterative procedures studied by Gerchberg-Saxton[8] and Fienup[6]. In the iterative algorithm, the "time" domain constraint that $x(\underline{n})$ is real and finite extent with a one-sided region of support, and the frequency domain constraint that the FT amplitude of $x(\underline{n})$ is given by $G_x^{\alpha}(\underline{\omega})$, are imposed separately in each iteration. Specifically, let $X_p(\underline{\omega})$ denote the estimate of $X(\underline{\omega})$ at the p^{th} iteration. The estimate $X_p(\underline{\omega})$ is inverse Fourier transformed to the time domain to obtain $x_p'(\underline{n})$

$$x_p'(\underline{n}) = F^{-1}[X_p(\underline{\omega})] \tag{23}$$

From $x_p'(\underline{n})$, we generate an estimate $x_p''(\underline{n})$ which satisfies the time domain constraints

$$x_p''(\underline{n}) = \begin{cases} \mathrm{Re}[x_p'(\underline{n})] & \text{for } \underline{n} \in A \\ 0 & \text{for } \underline{n} \notin A \end{cases} \tag{24}$$

where A represents the known support region of $x(\underline{n})$.

The sequence $x_p''(\underline{n})$ is then Fourier transformed back to the frequency domain to obtain $X_p''(\underline{\omega})$

$$X_p''(\underline{\omega}) = F[x_p''(n)] \tag{25}$$

The new frequency domain estimate $X_{p+1}(\underline{\omega})$ is then obtained by enforcing the constraint that $G_{x_{p+1}}^{\alpha}(\underline{\omega}) = G_x^{\alpha}(\underline{\omega})$ as follows:

$$X_{p+1}(\underline{\omega}) = \begin{cases} |X(\underline{\omega})|\, e^{j\theta_{x_p''}(\underline{\omega})} & \text{if } S_{x_p''}^{\alpha}(\underline{\omega}) = S_x^{\alpha}(\underline{\omega}) \\ |X(\underline{\omega})|\, e^{j(2\alpha - \theta_{x_p''}(\underline{\omega}))} & \text{if } S_{x_p''}^{\alpha}(\underline{\omega}) = -S_x^{\alpha}(\underline{\omega}) \end{cases} \tag{26}$$

Specifically, the correct magnitude is substituted for the estimated magnitude. If $S_{x_p''}^{\alpha}(\underline{\omega}) = S_x^{\alpha}(\underline{\omega})$, then the phase of the estimate is retained. Otherwise, the estimate is reflected about a line through the origin and with slope α to correct the sign of $S_{x_p''}^{\alpha}(\underline{\omega})$. This completes one iteration. The initial estimate $X_0(\underline{\omega})$ we have used is given by

$$X_0(\underline{\omega}) = |X(\underline{\omega})|\, e^{j\theta_{x_0}(\underline{\omega})} \tag{27}$$

where $\theta_{x_0}(\underline{\omega})$ is given by

$$\theta_{x_0}(\underline{\omega}) = \begin{cases} \alpha - \frac{\pi}{2} & \text{for } S_x^{\alpha}(\underline{\omega}) = +1 \\ \alpha + \frac{\pi}{2} & \text{for } S_x^{\alpha}(\underline{\omega}) = -1 \end{cases} \tag{28}$$

The iterative algorithm discussed above is illustrated in Figure 3.

The asymptotic behavior of the algorithm in Figure 3 has not yet been studied theoretically. We have observed experimentally that a stable estimate of the sequence to be retrieved is always attained after a large number of iterations.

To implement the algorithm in Figure 3, the Fourier and inverse Fourier transform operations are approximated by discrete Fourier transform (DFT) and inverse DFT (IDFT) operations. Although the uniqueness is not guaranteed in terms of the FT amplitude samples, we have empirically observed that the algorithm reconstructs the desired sequence provided that the FT amplitude is densely sampled in the frequency domain, so that the FT magnitude is completely specified and the discontinuities of $S_x^{\alpha}(\underline{\omega})$ are individually resolved by the

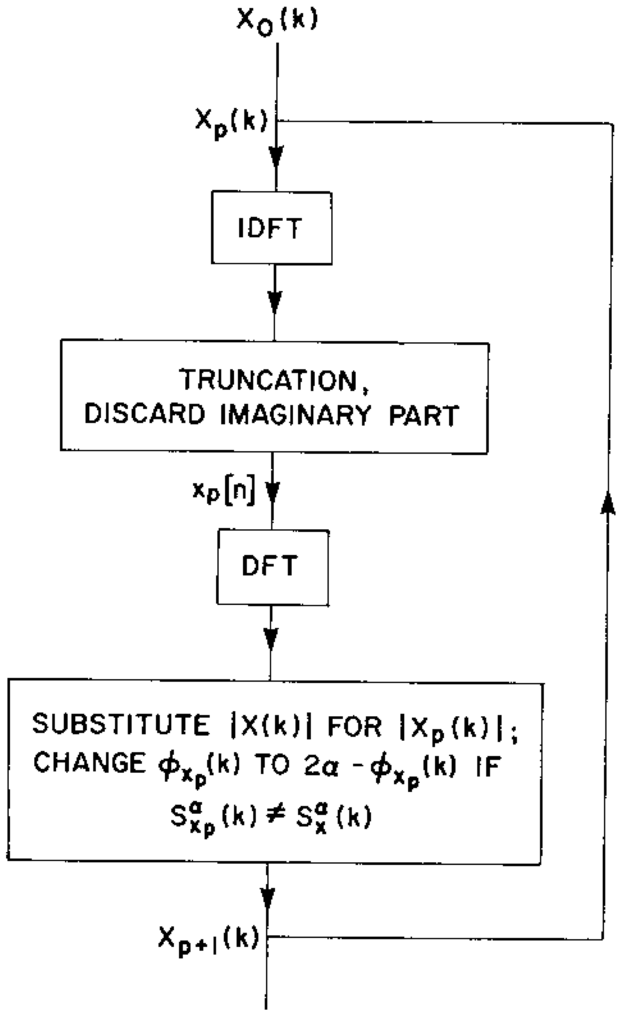

Figure 3: Block diagram of the iterative algorithm

samples of $S_x^{\alpha}(\underline{\omega})$. The FT magnitude $|X(\underline{\omega})|$ is completely specified by samples of $|X(\underline{\omega})|$ when the DFT size is twice the size of the known support of $x(\underline{n})$ in each dimension.

IV. Examples

The algorithm discussed in Section III has been used to reconstruct a variety of different 1-D and 2-D sequences from their FT amplitudes. In this section, we present some of these examples.

Figure 4 illustrates one example in which a 1-D sequence is reconstructed from its FT amplitude. In Figure 4(a) is shown a 47-point sequence obtained by sampling female speech at a 10 kHz rate. In Figure

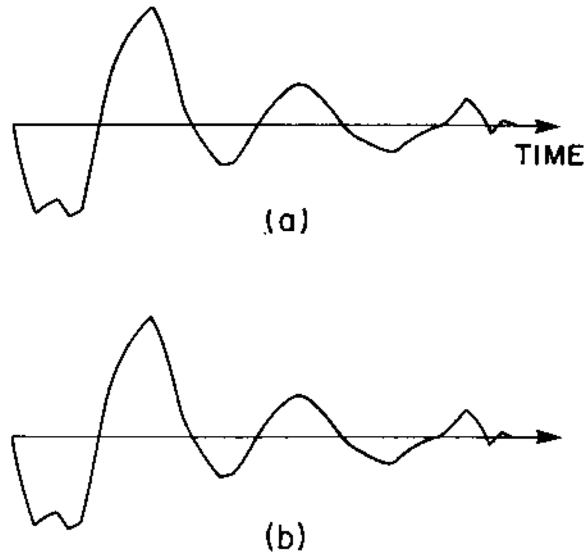

Figure 4: Speech segment sampled at 47 points a) original sequence; b) reconstructed sequence after 50 iterations

4(b) is shown the sequence reconstructed by using the iterative algorithm with the DFT size of 1024 after 50 iterations. In addition to the above example, a number of other examples have been considered. In all cases, we observed that the algorithm reconstructs the desired sequence.

Figure 5 illustrates an example in which a 2-D sequence is reconstructed from its FT amplitude. In Figure 5(a) is shown an image of size 256×256 pixels. In Figure 5(b) is shown the image reconstructed by using the iterative algorithm using the DFT size of 512×512 after 10 iterations.

In addition to the examples shown in this section, we have studied a number of other examples. From these examples, we have made the following observation about the iterative algorithm. First, for sequences satisfying the uniqueness constraints, if a DFT size below some threshold value is used, the algorithm does not lead to the desired sequence. Second, the convergence rate of the iterative algorithm is rapid initially and becomes slow as the number of iterations is increased. Third, the threshold DFT length is approximately the same for different choices of α, as long as α is not too close to 0 or π. As α approaches 0 or π, the threshold length is significantly increased. The choice of $\alpha=\pi/2$ permits the use of FFT routines specific to real sequences and uses, therefore, less computation time and less storage space. Fourth, we have observed that the mean square error between the original and reconstructed sequences decreases monotonically as the number of iterations increases. Fifth, the convergence rate of the algorithm can be significantly improved by using an acceleration procedure similar to that used by Oppenheim, et.al.[9]. Further details on the behavior of the iterative algorithm can be found in Van Hove[10].

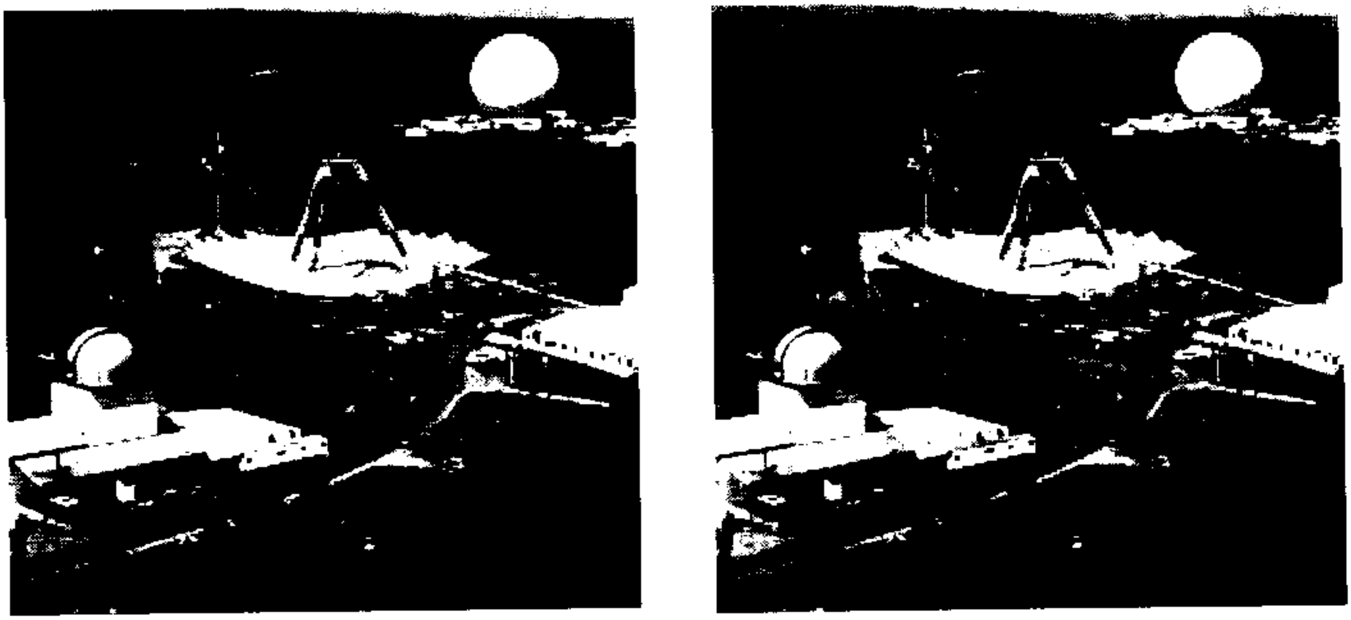

Figure 5: Image of size 256x256 pixels

a) original image b) reconstructed image after 10 iterations

V. Conclusions

In this paper, we have shown that a 1-D or M-D sequence is uniquely specified under mild restrictions by its FT amplitude. In addition, we have developed an iterative algorithm to reconstruct a 1-D or M-D sequence from its FT amplitude. When this result is combined with the previous result[5] on the problem of reconstructing a 1-D or M-D sequence from its FT phase, we obtain a very general result that a 1-D or M-D sequence is uniquely specified by its FT phase or its FT amplitude. In addition, under mild restrictions, an iterative algorithm which is similar in style can be used to reconstruct a 1-D or M-D sequence from its FT phase or amplitude.

Several important questions remain unanswered and remain as topics for future research. In this paper, the uniqueness question has been studied under the assumption that the FT amplitude is given for all $\underline{\omega}$. The unique specification of a sequence by samples of its FT amplitude remains as a topic for future research. The algorithm that we have developed has been empirically observed to converge to the desired sequence. It is not known theoretically, however, if the algorithm always converges.

Appendix

Statement A1

Let x(n) and y(n) be two real, causal and finite extent sequences. If $|X(\omega)|=|Y(\omega)|$, x(n) and y(n) can always be expressed as

$$x(n) = b(n) * a(n)$$

$$y(n) = \varepsilon b(n) * a(N-1-n)$$

where $\varepsilon=+1$ or -1 and a(n) and b(n) are real, causal and finite extent with N corresponding to the length of a(n), i.e. a(n)=0 outside $0 \le n \le N-1$.

Proof

A general expression of the z-transform X(z) of a sequence x(n) which is causal and has a finite support is given by

$$X(z) = z^{-n_1} x_0 \prod_{i=1}^{Q} (1 - z_i z^{-1}) \qquad \text{(A1.1)}$$

where z_i, $i=1,2,\ldots,Q$, are the zeros of $X(z)$, x_0 is the first non-zero sample, and n_1 is the positive initial delay in $x(n)$. It is well known that the FT magnitude of a finite extent 1-D sequence remains unchanged only when the sequence is subject to linear shifts, sign inversions and/or zero "flipping". The z-transform $Y(z)$ may therefore be written as

$$Y(z) = \pm z^{-n_2} x_0 \prod_{i\varepsilon\{u\}} (1 - z_i z^{-1}) \prod_{i\varepsilon\{r\}} (-z_i + z^{-1}) \tag{A1.2}$$

where n_2 is the positive initial delay in $y(n)$, $\{r\}$ is the set of indexes of the R zeros of $Y(z)$ which are zeros of $X(z)$ reflected across the unit circle and $\{u\}$ is the set of indexes of zeros which are unchanged from $X(z)$ to $Y(z)$. We may also write (A1.1) and (A1.2) as

$$X(z) = A(z)\cdot B(z)$$

$$Y(z) = \pm C(z)\cdot B(z)$$

or

$$x(n) = a(n) * b(n)$$

$$y(n) = \pm c(n) * b(n) \tag{A1.3}$$

where

$$A(z) = z^{-(n_1-n_2)} \prod_{i\varepsilon\{r\}} (1 - z_i z^{-1})$$

$$B(z) = z^{-n_2} x_0 \prod_{i\varepsilon\{u\}} (1 - z_i z^{-1})$$

$$C(z) = \prod_{i\varepsilon\{r\}} (-z_i + z^{-1}) \tag{A1.4}$$

We now show that $c(n)$ is $a(n)$ time reversed, represented by $a'(n)$. The length of the sequence $a'(n)$ is $N=n_1-n_2+R+1$, if we include the leading zeros. Therefore,

$$a'(n) = a(N-1-n)$$

$$A'(z) = A(z^{-1})z^{-(N-1)} = z^{-R} \prod_{i\varepsilon\{r\}} (1 - z_i z^{-1}) = C(z)$$

so that $c(n)=a(N-1-n)$. From (A1.3), the sequences $x(n)$ and $y(n)$ are expressed in the adequate form. To characterize $a(n)$ and $b(n)$, we examine their z-transforms. Since $B(z)$ contains only a finite number of negative powers of z, the sequence $b(n)$ has a finite causal support. Since $A(z)$ and $A'(z)=C(z)$ contain only negative powers of z, it follows that $a(n)$ and $a(N-1-n)$ are causal so that $a(n)$ is zero outside $0<n<N-1$. If the z-transform $X(z)$ contains a pair of complex conjugate zeros, then they must belong both to $\{u\}$ or both the $\{r\}$ for $y(n)$ to be real-valued. The z-transforms $A(z)$ and $B(z)$ may therefore contain complex zeros only in conjugate pairs so that $a(n)$ and $b(n)$ are real. In the case $n_2>n_1$, we simply exchange the roles of $x(n)$ and $y(n)$. This completes the proof of Statement A1.

Statement A2

Let $b(n)$ be a real, causal, and finite extent sequence. For any positive integer N, the equation

$$\mathrm{Re}\left[B(z)\, z^{-\frac{N-1}{2}} \Big|_{z=e^{j\omega}} \right] = 0$$

is satisfied for at least P distinct values of ω in the interval R, where P and R are as defined in eq. (7) of the text.

To prove this statement, we introduce the notion of unwrapped phase. Given a Fourier transform $M(\omega)$ which has no zeros, we define its unwrapped phase $\phi_M(\omega)$ as the unique continuous function of ω which satisfies

$$M(\omega) = |M(\omega)| e^{j\phi_M(\omega)} \tag{A2.1}$$

for all ω and which takes the value of 0 or $-\pi$ at $\omega=0$. The unwrapped phase has the following properties. If we define the function $F(\omega)$ as

$$F(\omega) = D(\omega)\ B(\omega) \tag{A2.2}$$

then it follows that

$$\phi_F(\omega) = \phi_D(\omega) + \phi_B(\omega) + 2\alpha\pi$$

where $\alpha = 1$ if $\phi_D(0) = \phi_B(0) = -\pi$ $\alpha = 0$ otherwise (A2.3)

The unwrapped FT phase $\phi_B(\omega)$ of a causal sequence $b(n)$ satisfies

$$\phi_B(0) \geq \phi_B(\pi) \tag{A2.4}$$

The unwrapped phase of the function

$$D(\omega) = e^{-j\omega \frac{N-1}{2}} \tag{A2.5}$$

is

$$\phi_D(\omega) = -\omega\ \frac{N-1}{2} \tag{A2.6}$$

We now proceed to the proof of statement A2. We consider the unwrapped phase $\phi_F(\omega)$ of the function

$$F(\omega) = B(\omega) e^{-j\omega \frac{N-1}{2}}$$

The equation $Re(F(\omega))=0$ has the same roots as the equation

$$\phi_F(\omega) = \frac{\pi}{2} + k\pi \quad , \text{ with k an integer,}$$

since $F(\omega)$ has no zeros. From our previous discussion, we have

$$\phi_F(\pi) - \phi_F(0) = \phi_B(\pi) - \phi_B(0) + \phi_D(\pi) - \phi_D(0) \leq -\left(\frac{N-1}{2}\right)\pi$$

Since the continuous function $\phi_F(\omega)$ decreases at least by $(N-1)/2\ \pi$ on the interval R, it follows that the graph of $\phi_F(\omega)$ crosses at least N/2 lines of phase $\pi/2 + k\pi$ in $(0,\pi]$ if N is even. Figure 6 shows $\phi_F(\omega)$ when $b(n)=\delta(n)$, for the cases N=4 and N=5.

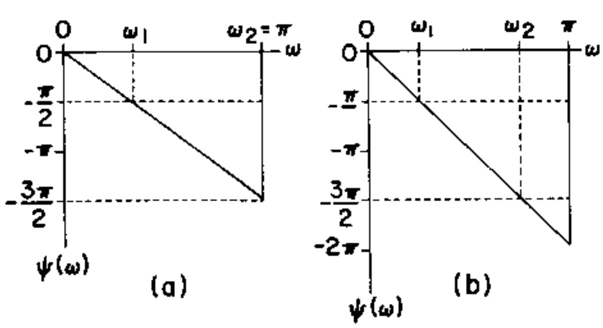

Figure 6: Unwrapped phase of the function $F(\omega)$ for $b(n)=\delta(n)$.

a) N=4 b) N=5

Statement A3

Let $a(n)$ be a real valued sequence which is zero outside $0 \le n \le N-1$. If the equation

$$\mathrm{Im}\left\{ A(z)\, z^{\frac{N-1}{2}} \Big|_{z=e^{j\omega}} \right\} = 0$$

is satisfied for at least P distinct values of ω in the interval R, then it is identically equal to zero and $a(n)=a(N-1-n)$. P and R are defined as in eq. (7) in the text

$$P = \frac{N-1}{2} \quad \text{and} \quad R = (0,\pi) \quad \text{for N odd}$$

$$P = \frac{N}{2} \quad \text{and} \quad R = (0,\pi] \quad \text{for N even}$$

Proof for N odd

With the use of trigonometric formulas, we obtain

$$G(\omega) = \mathrm{Im}\left\{ A(\omega) e^{j\omega \frac{N-1}{2}} \right\} = \sum_{n=0}^{N-1} a[n] \sin\left(\frac{N-1}{2} - n\right)\omega \tag{A3.1}$$

$$G(\omega) = \sum_{n=1}^{\frac{N-1}{2}} \left\{ a\left(\frac{N-1}{2} - n\right) - a\left(\frac{N-1}{2} + n\right) \right\} \sin n\omega \tag{A3.2}$$

Since the set of the $(N-1)/2$ functions $\sin\omega, \sin 2\omega, \ldots, \sin (N-1)\omega/2$ is a Chebychev set on the interval $(0,\pi)$ as is shown in[10], and since $G(\omega)$ has at least $(N-1)/2$ distinct roots in the interval $(0,\pi]$, it follows that the coefficients of the expansion in the right-hand side of (A3.2) must vanish

$$a\left(\frac{N-1}{2} - n\right) = a\left(\frac{N-1}{2} + n\right) = 0; \qquad n=1,2,\ldots,\frac{N-1}{2}$$

or

$$a(n) = a(N-1-n); \quad n=0,1,\ldots,N-1$$

When N is even, the expansion of $G(\omega)$ is

$$G(\omega) = \sum_{n=0}^{\frac{N}{2}-1} \left\{ a\left(\frac{N}{2} - 1 - n\right) - a\left(\frac{N}{2} + n\right) \right\} \sin\left(n + \frac{1}{2}\right)\omega$$

Since the functions $\sin \omega/2, \sin 3\omega/2, \ldots, \sin N-1/2\, \omega$ form a Chebychev set on the interval $(0,\pi]$ as is shown in[10], it follows that

$$a\left(\frac{N}{2} - 1 - n\right) - a\left(\frac{N}{2} + n\right) = 0; \qquad n=0,1,\ldots,\frac{N}{2} - 1$$

or

$$a(n) = a(N-1-n)\ ; \qquad n=0,1,\ldots,N-1$$

This completes the proof of Statement A3.

Acknowledgements

This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102.

References

1. W. O. Saxton, Computer Techniques for Image Processing in Electron Microscopy, Academic Press, New York, 1978.

2. G. N. Ramachandran and R. Srinivasan, Fourier Methods in Crystallography, Wiley-Interscience, New York, 1970.

3. D. L. Misell, "An Examination of an Iterative Method for the Solution of the Phase Problem in Optics and Electron Optics: I and II," J. Phys. D, Appl. Phys., vol. 6, pp. 2200-2225, 1973.

4. W. K. Pratt, Digital Image Processing, Wiley, New York, 1978.

5. M. H. Hayes, "The Reconstruction of a Multidimensional Sequence from the Phase or Magnitude of its Fourier Transform," IEEE Trans. on Acoustics, Speech and Signal Proc., vol. ASSP-30, no. 2, pp. 140-154, April 1982.

6. J. R. Fienup, "Space Object Imaging Through the Turbulent Atmosphere," Optical Engineering, vol. 18, pp. 529-534, Sept. 1979.

7. M. H. Hayes, "Signal Reconstruction from Phase or Magnitude," Sc.D. Thesis, Mass. Institute of Tech., Dept. of Elec. Eng. and Comp. Sci., June 1981.

8. R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," Optik, vol. 35, pp. 237-246, 1972.

9. A. V. Oppenheim, M. H. Hayes and J. S. Lim, "Iterative Procedure for Signal Reconstruction from Fourier Transform Phase," Optical Eng., vol. 21, pp. 122-127, Jan./Feb. 1982.

10. P. L. Van Hove, "Signal Reconstruction from Fourier Transform Amplitude," S.M. Thesis, Mass. Institute of Tech., Dept. of Elec. Eng. and Comp. Sci., September 1982.