A fast algorithm for extrapolation of discrete-time periodic band-limited signals

Hamid Soltanian-Zadeh and Andrew E. Yagle

Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122, USA

Abstract. We present a fast algorithm for exact extrapolation of a discrete-time and periodic band-limited signal from its values in an interval having the same length as the band-width of the signal. Here band-width is the number of non-zero values in one period of the discrete Fourier transform of the signal. Applications to problems in which the number of given values is unequal to the band-width are also given. The procedure is a simple autoregression on the time-domain values of the signal, and is much simpler than previous algorithms for discrete–discrete extrapolation. which required computation of a large pseudo-inverse. The procedure is highly parallelizable, and the computational savings are especially significant for multidimensional signal extrapolation. Numerical examples for 1-D and 2-D extrapolation demonstrate (1) that the procedure works perfectly in the absence of noise, and (2) that it works well in the presence of band-limited noise. In the presence of wide-band noise the procedure breaks down, due to ill-posedness of the problem; some regularization techniques are proposed.

Zusammenfassung. Wir stellen einen schnellen Algorithmus zur exakten Extrapolation eines zeitdiskreten und periodischen, bandbegrenzten Signals vor, ausgehend von seinen Werten innerhalb eines Intervalls, das die gleiche Länge hat wie die Bandbreite des Signals. Bandbreite bedeutet hier die Zahl der von Null verschiedenen Werte einer Periode der diskreten Fourier-Transformation des Signals. Es werden auch Anwendungen auf Probleme angegeben, bei denen die gegebene Werteanzahl ungleich der Bandbreite ist. Das Verfahren ist eine einfache Autoregression der Zeitbereichswerte des Signals, und ist viel einfacher als vorher angegebene Verfahren für die diskrete–diskrete Extrapolation, die die Berechnung einer großen Pseudoinversen erforderte. Das Verfahren ist hochgradig parallelisierbar, und die rechnerischen Einsparungen sind besonders signifikant bei der multidimensionalen Signalextrapolation. Numerische Beispiele bei 1-D und 2-D Extrapolation zeigen (1) daß das Verfahren perfekt arbeitet bei der Abwesenheit von Rauschen, und (2) daß es gut arbeitet bei der Störung durch bandbegrenztes Rauschen. Bei der Anwesenheit von breitbandigem Rauschen bricht das Verfahren wegen der Schlechtgestelltheit des Problems zusammen; in diesem Fall werden einige Regularisierungsmethoden vorgeschlagen.

Résumé. Nous présentons un algorithme rapide d'extrapolation exact d'un signal périodique à temps discret et à bande limitée à partir de ses valeurs dans un intervalle ayant la même largeur que sa largeur de bande. Par largeur de bande on entend ici le nombre de valeurs non nulles dans une période de la transformée de Fourier discrète du signal. Des applications au problème dans lequel le nombre de valeurs données est différent de la largeur de bande sont également données. La procédure consiste en une simple autorégression sur les valeurs temporelles du signal, et est beaucoup plus simple que les algorithmes pré-existants pour l'interpolation discret–discret, qui nécessitent le calcul d'un pseudo-inverse de grande taille. Cette procédure est hautement parallélisable, et les réductions de calcul sont tout spécialement significatives dans le cas de l'extrapolation de signaux multidimensionnels. Des exemples numériques pour les cas 1-D et 2-D démontrent que (1) la procedure fonctionne parfaitement bien en l'absence de bruit, et (2) qu'elle fonctionne bien en présence de bruit à bande limitée. En présence de bruit à large bande la procédure ne fonctionne pas, du fait du caractère mal posé du problème; des techniques de régularisation sont proposées.

Keywords. Band-limited extrapolation; signal reconstruction; multi-dimensional signal reconstruction; fast algorithms.

Correspondence to: Prof. Andrew E. Yagle, Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122, USA. Fax: 313-763-1503; E-mail: aey@eecs.umich.edu

1. Introduction

1.1. Summary of the problem

We consider the following discrete–discrete band-limited extrapolation problem. Given $2M+1$ values of a discrete-time periodic sequence $x(n)$ with period N whose discrete Fourier transform (DFT) $X(k)$ is known to be zero for $M < |k| \leqslant \frac{1}{2}N$, determine the other values of $x(n)$. The term 'discrete–discrete' is taken from [15], which contains a detailed discussion of the relations between the discrete–discrete problem and the continuous–continuous problem, in which the signal to be extrapolated is continuous and its continuous Fourier transform is known to be zero outside some interval. We do not consider the latter problem in this paper.

Previous solutions to the discrete–discrete problem have used several approaches. One is to formulate the problem as a linear system of equations, which is solved using the singular value decomposition [7, 20], or using a minimum DFT-weighted norm solution [3]. Since the linear system is ill-conditioned [20], some sort of regularization is required. And since the linear system has N unknowns, a large amount of computation is required. Another approach [4, 6] requires solution of a linear system of equations with $2M+1$ unknowns, followed by low-pass filtering. Other approaches have been applied to the continuous–continuous problem, including iterative algorithms [5, 13]; see also [9, 11, 12, 17, 21, 23].

The significance of our approach as compared with those above is that we derive a simple autoregression for extrapolating the unknown values of $x(n)$ *directly* from the known values. This has three major advantages:

1. Since the problem is never formulated as a system of equations, issues of regularization and ill-conditioning do not arise. Our algorithm never requires even as much as a single division.
2. The computation involved in solving an ill-conditioned system of equations is not trivial if N is large, *especially in the multi-dimensional case*. Our Mth

order autoregression is computationally simple and parallelizable in the multidimensional case.

3. Although $x(n)$ is required to be periodic, the period N can be made arbitrarily large with *no attendant increase* in the amount of computation. In contrast, the solutions in [7, 20] require solution of an $N \times N$ linear system, whose conditioning becomes worse as N gets larger [20]. Even the pseudo-inverses in Section 3 below are of size $2M+1$, not of size N. Thus, the continuous Fourier transform of a band-limited and time-limited continuous signal (although no signal can be completely time- and band-limited, many signals of practical interest are effectively zero outside some time and frequency intervals) can be approximated arbitrary closely without more computation.

The ill-posedness of the band-limited signal extrapolation problem is exhibited in the large values of the autoregression coefficients; these are a precise measure of the *sensitivity* of the extrapolation to the given $x(n)$. However, numerical instability in the autoregression can be avoided by computing a desired $x(n)$ directly using the variation of constants formula (see (25) below).

1.2. Applications

Band-limited signal extrapolation (or equivalently, support-limited spectrum extrapolation) is a key problem in signal reconstruction and restoration. A common application is in spectral estimation [1, 6, 13], in which the observed signal is extrapolated in order to improve the spectral resolution. Another application is in image filtering. Ideally, an image filter performs over an infinite-extent image, but in practice we are given only a piece of the image. When the filter is applied to points close to or on the border of the actual image, inaccuracies will result if we assume some arbitrary numbers for the unknown values of the image outside the border, e.g., the image is assumed to be periodic; or a constant number is assumed for the unknown values. Filter per-

formance would improve if unknown values of the image could be extrapolated.

By switching time and frequency, the problem becomes support-limited spectrum extrapolation [14]. One application of this problem is the missing-cone problem in computer tomography, as applied to non-destructive testing [4]. Another application is in seismic data processing. Due to the narrow frequency response of seismic sensors (around 4–40 Hz), the high frequency content of the signal is lost, resulting in loss of resolution of the reconstructed medium. Since seismic signals have short temporal duration (narrow time-domain support), spectrum extrapolation can be used to restore the missing information and resolve finer features of the medium. Finally, there is a recent application in phase retrieval [18].

2. Problem formulation

2.1. The basic problem

A discrete and periodic signal $x(n)$ can be represented as

$$x(n) = \sum_{n=-\infty}^{\infty} x_N(n-mN) , \qquad (1)$$

where $\{x_N(n)\}$ is the sequence $\{x(n)\}$ over one period (N consecutive points), i.e.,

$$x_N(n) = \begin{cases} x(n), & \text{if } -\frac{1}{2}N+1 \leqslant n \leqslant \frac{1}{2}N, \\ 0, & \text{otherwise.} \end{cases} \qquad (2)$$

$x(n)$ is defined as being band-limited to $2M+1$ points if its N-point discrete Fourier transform (DFT)

$$X(k) = \sum_{n=-N/2+1}^{N/2} x(n) e^{-j2\pi kn/N} x(n) e^{-j2\pi kn/N} ,$$

$$k = -\frac{1}{2}N+1, \ldots, \frac{1}{2}N , \qquad (3)$$

has the property

$$X(k) = 0, \quad M < |k| \leqslant \frac{1}{2}N . \qquad (4)$$

For convenience, we use the term 'band-width' to mean $2M+1$, not M. For example, $x(n)$ could be a sampled version of a continuous and periodic band-limited sig-

nal which is sampled at a rate higher than its Nyquist rate.

Signal extrapolation refers to the problem of finding an estimate of a signal outside its observation interval. Specifically, the discrete–discrete band-limited signal extrapolation problem is stated as follows:
Given $x(n)$,

$$n \in [r+1, r+L], \quad L < N, \quad r, L \text{ are integers.}$$

Find

$$y(n), \quad -\frac{1}{2}N+1 \leqslant n \leqslant \frac{1}{2}N, \qquad (5)$$

such that

$$y(n) = x(n), \quad n \in [r+1, r+L],$$

and

$y(n)$ is band-limited to $[-M, M]$

Here and in the sequel N is the size of the DFT, M is the half-band-width of the signal, and L is the length of the observation interval. We shall demonstrate shortly that the problem is underdetermined if $L < 2M+1$, is overdetermined and ill-posed if $L > 2M+1$, and has a unique solution if $L = 2M+1$.

2.2. A previous algorithm

There are two major classes of algorithms for solving the extrapolation problem: iterative procedures and non-iterative two-step procedures. In [4, 16] it is shown that the known iterative procedures are special cases of a general iterative procedure (of Bialy [2]) for solving linear operator equations in Hilbert spaces. Similarly, non-iterative two-step procedures are special cases of some general techniques developed in [22] for solving integral equations. These techniques are general algorithms for solving two optimization problems in Hilbert spaces.

For comparison with our approach. we quickly state a discrete–discrete non-iterative two-step procedure [4, 6]:
1. Solve for $z(m)$ the $L \times (2M+1)$ linear system of equations

$$\sum_{m=-M}^{M} \text{PSINC}_{N,M}(n-m)z(m) = x(n),$$

$$n \in [r+1, r+L]. \tag{6}$$

2. Compute

$$y(n) = \sum_{m=-M}^{M} \mathrm{PSINC}_{N,M}(n-m)z(m),$$

where

$$\mathrm{PSINC}_{N,M}(r) = \frac{1}{N} \sum_{n=-M}^{M} \mathrm{e}^{\mathrm{j}2\pi nr/N} \tag{8}$$

is the impulse response of a discrete–discrete low-pass filter. We relate this method to the method of this paper in Appendix A.

This method finds an intermediate sequence $z(n)$ which has finite support in the time domain identical to the band-width of the original signal in the frequency domain. Its DFT equals the DFT of the original signal inside the band-width, and is different everywhere else. This intermediate sequence is then low-pass filtered to generate the extrapolated signal. Note that if $L < 2M+1$, then (6) corresponds to an underdetermined system of equations, so that the problem does not have a unique solution [17]; one choice is the minimum norm solution, as discussed in [6].

2.3. The new approach

In this paper we develop a new method. Our approach is to exactly extrapolate a strictly band-limited signal from $2M+1$ specified values of it. If $L = 2M+1$ in (5) this results in a unique solution; $L \neq 2M+1$ is discussed later. The extrapolator is linear and makes no assumption about the shape of the signal's spectrum.

The major difference between this work and previous efforts is as follows. Previous methods have explicitly or implicitly written the band-limited constraint in the Fourier domain as (see Appendix A)

$$X(k)S(k) = X(k), \tag{9}$$

where

$$S(k) = \begin{cases} 1, & k \in \text{band-width of } x(n), \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

This corresponds to

$$x(n) * s(n) = x(n) \tag{11}$$

in the time domain. In (11) $s(n)$ is a scaled discrete sinc function and $*$ represents a circular convolution. The continuous version of (11) is the integral equation explicitly or implicitly solved by previous approaches.

Our choice is to replace $S(k)$ in (10) with a different $S(k)$ such that

$$S(k) = \begin{cases} 1, & k \in \text{band-width of } x(n), \\ \neq 1, & \text{otherwise.} \end{cases} \tag{12}$$

Note that this choice of $S(k)$ in conjunction with (9) still implements the band-limited constraint.

The advantage of this choice is that $S(k)$ can be found so that $s(n)$ has finite support, unlike the tails of the discrete sinc function. This $s(n)$ turns the convolution into a regression on $x(n)$. In other words, a $(2M+1)$-point band-limited signal satisfies an autoregression of order $2M+1$; thus there is no need to solve (6). In a sense, (6) and (7) have been replaced by similar equations with a much simpler kernel; the new equations combine to have a simple closed form solution.

3. Derivation of 1-D extrapolation equations

In this section we discuss the one-dimensional (1-D) band-limited signal extrapolation problem. First, the autoregression is derived, and then initialization procedures for when the observation interval is unequal to the band-width (in number of points) are discussed. In Section 4 we extend the method to 2-D and K-D band-limited signal extrapolation. In each case we derive an explicit equation for extrapolating the signal.

3.1 Derivation of 1-D autoregression

We consider a discrete and periodic signal whose N-point DFT (N is even) is band-limited to $2M+1 < N$ points. We prove the following theorem for this class of signals.

THEOREM 1. A 1-D discrete signal $x(n)$ whose DFT $X(k)$ is band-limited to $2M+1$ points satisfies an autoregression of order $2M+1$.

PROOF. Without loss of generality we can assume that the $2M+1$ point band-width consists of $2M+1$ consecutive points centered at the origin. This is the usual case for a band-limited (low-pass) signal. However, the proof can easily be modified for an arbitrary $2M+1$ point band-width if necessary.

Define

$$S(k) = 1 + \prod_{i=-M}^{M} (Z_k - Z_i),$$

$$k = -\tfrac{1}{2}N+1, \ldots, \tfrac{1}{2}N, \tag{13}$$

where

$$Z_i = e^{-j2\pi l/N}, \quad l = -\tfrac{1}{2}N+1, \ldots, \tfrac{1}{2}N, \tag{14}$$

are the N Nth roots of unity. Then $S(k)$ satisfies

$$S(k) = \begin{cases} 1, & \text{if } -M \leqslant k \leqslant M, \\ \neq 1, & \text{otherwise,} \end{cases} \tag{15}$$

and the fact that $x(n)$ is band-limited can be written as

$$X(k) = X(k)S(k), \tag{16}$$

which corresponds to

$$x(n) = x(n) * s(n) \tag{17}$$

in the time domain.

Simplifying (13) by multiplying different terms and collecting the coefficients $c(n)$ for identical powers of Z_k, and noting that $\prod_{i=-M}^{M}(-Z_i) = -1$, we obtain

$$S(k) = \sum_{n=1}^{2M+1} c(n) Z_k^n$$

$$= \sum_{n=1}^{2M+1} c(n) \, e^{-j2\pi kn/N}. \tag{18}$$

Equations (13) and (18) imply that

$$s(n) = \mathscr{F}^{-1} \left\{ 1 + \prod_{i=-M}^{M} (Z_k - Z_i) \right\}$$

$$= \begin{cases} c(n), & \text{if } 1 \leqslant n \leqslant 2M+1, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Substituting (19) in (17) results in

$$x(n) = \sum_{i=1}^{2M+1} c(i)x(n-1), \tag{20}$$

which is an autoregression of order $2M+1$ over the band-limited signal $x(n)$. $\square$

COROLLARY 1.1. A band-limited signal also satisfies the higher order autoregressions obtained for band-limited signals with a larger band-width which includes the original signal's band-width.

COROLLARY 1.2. A necessary and sufficient condition for a signal to have a $(2M+1)$-point band-width centered at the origin is that it satisfy the autoregression (20).

3.2. Derivation of explicit equation for 1-D extrapolation

When a band-limited signal is observed over an interval with length L (in number of points) equal to its band-width $2M+1$, (20) implements an exact extrapolation algorithm. Indeed, we now derive an explicit expression for extrapolating to any desired point. Define a state vector x_n as

$$x_n = [x(n), x(n-1), \ldots, x(n-2M)]^T. \tag{21}$$

Then the autoregression (20) can be written as the state equation

$$x_{n+1} = \begin{bmatrix} c(1) & \cdots & c(2M) & 2(2M+1) \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} x_n. \tag{22}$$

Let A be the square matrix of size $2M+1$ in (22). Note that $c(2M+1) = 1$, since $\prod_{i=-M}^{M}(Z_k) = Z_k^{2M+1}$. A is in controllable (phase-variable) canonical form; from (19) its eigenvector decomposition is as follows [8]:

$$A = U\Lambda U^{-1}, \tag{23}$$

where

$$U = \begin{bmatrix} 1 & \cdots & 1 & \cdots & 1 \\ Z_{-M} & \cdots & Z_0 & \cdots & Z_M \\ Z_{-M}^2 & \cdots & Z_0^2 & \cdots & Z_M^2 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ Z_{-M}^{2M} & \cdots & Z_0^{2M} & \cdots & Z_M^{2M} \end{bmatrix}, \quad (24)$$

$$\Lambda = \mathrm{DIAG}[Z_M \cdots Z_{-M}],$$

and Z_l was defined in (14). From (22) and (23) we have

$$x_m = A^{m-n} x_n = U \Lambda^{m-n} U^{-1} x_n, \quad (25)$$

which allows any $x(m)$ (the first element of x_m) to be extrapolated from any $2M+1$ consecutive values of $x(n)$, i.e., $\{x(i), n-2M \leqslant i \leqslant n\}$.

Three comments are in order here:

1. From (13) and (14), it is clear that $S^*(k) = S(-k)$. Hence the autoregression coefficients $c(n)$ are all real.

2. If the signal $x(n)$ is complex, the autoregression (20) applies separately to its real and imaginary parts.

3. Since the eigenvalues of A are all Nth roots of unity [8], the extrapolated $x(n)$ will correctly be periodic with period N (note that $x_n = x_0$).

The system in (22) is reminiscent of state-space methods for harmonic retrieval [7], but there are some differences. Here we are concerned with band-limited signal extrapolation, while previous work dealt with the identification of sinusoidal signals at unknown frequencies. Also, this procedure applies to the general case of any band-limited signal, while previous work considered only sinusoidal signals.

3.3. Initialization of autoregression: $L = 2M + 1$

For notational convenience, let the given segment of signal to be extrapolated be $\{x(-L+1), x(-L+2), \ldots, x(0)\}$, i.e., $l = -L$ in (5). If $L = 2M+1$, then the given segment is precisely the elements of x_0, and the entire signal may be extrapolated using

$$x_{(2M+1)i} = U \Lambda^{(2M+1)i} U^{-1} x_0,$$

$$1 \leqslant i \leqslant p, \quad p = [N/2M+1)], \quad (26)$$

and the fact that $x(n)$ is periodic with period N (the

brackets in (26) denote greatest integer function). Each extrapolated value of $x(n)$ will be a component of one of the $x_{(2M+1)i}$.

3.4. Initialization of autoregression: $L > 2M + 1$

If $L > 2M + 1$, then the problem is *overdetermined*, since any $2M+1 < L$ values of $x(n)$ uniquely determine the other values. Indeed, unless the extra $L - 2M - 1$ values of $x(n)$ are entirely consistent with the first $2M+1$ values, the extrapolation problem has no solution. Due to observation noise, this consistency is extremely unlikely; hence the problem is ill-posed.

In this case we may estimate a segment $\{\hat{x}(-2M), \ldots, \hat{x}(0)\}$ from $\{x(-L+1), \ldots, x(0)\}$, and use $\{\hat{x}(-2M), \ldots, \hat{x}(0)\}$ in (26) to extrapolate a properly band-limited signal. The least-squares solution minimizing $\sum_{i=-L+1}^{0} (x(i) - \tilde{x}(i))^2$, where the additional $\hat{x}(i)$ have been extrapolated using (26), can be found as follows. Augmenting (25) and using (21) yields

$$\begin{bmatrix} x(0) \\ \vdots \\ x(-L+2M+2) \\ x(-L+2M+1) \\ \vdots \\ x(-L+1) \end{bmatrix}$$

$$= \begin{bmatrix} I_{L-(2M+1)} & 0 \\ A^{2M+1-L} & \end{bmatrix} \begin{bmatrix} \hat{x}(0) \\ \vdots \\ \hat{x}(-L+2M+2) \\ \hat{x}(-L+2M+1) \\ \vdots \\ \hat{x}(-2M) \end{bmatrix}$$

$$= C \begin{bmatrix} \hat{x}(0) \\ \vdots \\ \hat{x}(-L+2M+2) \\ \hat{x}(-L+2M+1) \\ \vdots \\ \hat{x}(-2M) \end{bmatrix}, \quad (27)$$

provided that $2M+1 < L \leqslant 4M+2$. If $L > 4M+2$ the matrix C will also have an A^{4M+2-L} element, i.e.,

$$C = \begin{bmatrix} I_{L-(4M+2)} & 0 \\ A^{4M+2-L} & \\ A^{2M+1-L} & \end{bmatrix}.$$

The least-squares solution to (27) is

$$
\begin{bmatrix} \hat{x}(0) \\ \vdots \\ \hat{x}(-2M) \end{bmatrix} = (C^{\mathrm{T}}C)^{-1}C^{\mathrm{T}} \begin{bmatrix} x(0) \\ \vdots \\ x(-L+1) \end{bmatrix}. \qquad (28)
$$

3.5. Initialization of autoregression: $L < 2M + 1$

If $L < 2M + 1$, then the problem is *underdetermined*, since the unspecified $2M + 1 - L$ values of $x(n)$ may be assigned arbitrary values, and a properly band-limited signal can still be extrapolated. Indeed, note that *all* solutions to the band-width extrapolation problem are easily characterized, with the unspecified $2M + 1 - L$ values of $x(n)$ as free parameters.

The simplest way to choose a solution is to tighten the band-width constraint from $2M + 1$ to L (assuming L is odd). This specifies the signal that matches the given L values, and is band-limited to $-M < -\frac{1}{2}(L-1) \leqslant k \leqslant \frac{1}{2}(L-1) < M$.

The *minimum norm* solution is also of interest; however, it requires more work. We now wish to determine the signal that matches the given L values and has band-width $2M + 1$, such that

$$
E = \sum_{n=0}^{N-1} \| x(n) \|^2 = \frac{1}{N} \sum_{k=0}^{N-1} \| X(k) \|^2 \qquad (29)
$$

is minimized. Using (25) and the periodicity of $x(n)$, this can be rewritten as

$$
E = \sum_{n=0}^{N-1} \| x(n) \|^2 = \sum_{i=1}^{p} \boldsymbol{x}_{(2M+1)i}^{\mathrm{H}} \boldsymbol{x}_{(2M+1)i}
$$

$$
+ \boldsymbol{x}_0^{\mathrm{H}} \begin{bmatrix} I_{N-(2M+1)p} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{x}_0 = \boldsymbol{x}_0^{\mathrm{H}} B \boldsymbol{x}_0, \qquad (30)
$$

where $p = [N/(2M+1)]$ as before and B is defined as

$$
B = \sum_{i=1}^{p} (A^{(2M+1)i})^{\mathrm{T}} A^{(2M+1)i}
$$

$$
+ \begin{bmatrix} I_{N-(2M+1)p} & 0 \\ 0 & 0 \end{bmatrix}
$$

$$
= U^{-\mathrm{T}} \left(\sum_{i=1}^{p} \Lambda^{(2M+1)i} (U^{\mathrm{T}}U) \Lambda^{(2M+1)i} \right) U^{-1}
$$

$$
+ \begin{bmatrix} I_{N-(2M+1)p} & 0 \\ 0 & 0 \end{bmatrix}
$$

$$
= \begin{bmatrix} B_1 & B_3 \\ B_3^{\mathrm{T}} & B_2 \end{bmatrix}. \qquad (31)
$$

Note that B is symmetric and positive definite, and is partitioned so that B_1 is $L \times L$ and B_2 is $(2M+1-L) \times (2M+1-L)$.

Now partition $\boldsymbol{x}_0$ into

$$
\boldsymbol{x}_0 = [x(0), \ldots, x(-L+1), \; x(-L), \ldots,
$$

$$
x(-2M)]^{\mathrm{T}}
$$

$$
= [\boldsymbol{d}^{\mathrm{T}} \vdots \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}, \qquad (32)
$$

where $\boldsymbol{d}$ is the L-vector of given signal values and $\boldsymbol{z}$ is the $(2M+1-L)$-vector of undetermined signal values. We wish to determine $\boldsymbol{z}$ such that E is minimized; then (26) can be used to compute the band-limited extrapolation from $\boldsymbol{x}_0 = [\boldsymbol{d}^{\mathrm{T}} \vdots \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}$. Inserting (31) and (32) into (30) gives

$$
E = [\boldsymbol{d}^{\mathrm{H}} \vdots \boldsymbol{z}^{\mathrm{H}}] \begin{bmatrix} B_1 & B_3 \\ B_3^{\mathrm{T}} & B_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{z} \end{bmatrix}
$$

$$
= \boldsymbol{d}^{\mathrm{H}} B_1 \boldsymbol{d} + \boldsymbol{z}^{\mathrm{H}} B_3^{\mathrm{T}} \boldsymbol{d} + \boldsymbol{d}^{\mathrm{H}} B_3 \boldsymbol{z} + \boldsymbol{z}^{\mathrm{H}} B_2 \boldsymbol{z}, \qquad (33)
$$

which is minimized when

$$
\boldsymbol{z} = -B_2^{-1} B_3^{\mathrm{T}} \boldsymbol{d}. \qquad (34)
$$

Hence the minimum norm solution is determined as follows:

1. Compute B from (31) and partition it.
2. Compute $\boldsymbol{z}$ from the given signal values $\boldsymbol{d}$ using (34).
3. Extrapolate the other signal values using (26) and (32).

Note that even if the given L values are *not contiguous*, this procedure may still be used by simply reordering the rows and columns of B. Also note that the only extra computation required is the computation of B and the solution of a linear system of equations (34) of order $2M + 1 - L$.

4. Derivation of 2-D and K-D extrapolation equations

4.1. Derivation of 2-D autoregression

We now consider a 2-D discrete and periodic signal

whose $(N \times N)$-point 2-D DFT is band-limited to $(2M+1) \times (2M+1)$ points. We prove the following theorem for this class of signals.

THEOREM 2. A 2-D discrete signal $x(n_1, n_2)$ whose $(N \times N)$-point 2-D DFT $X(k_1, k_2)$ is band-limited to a square region of $(2M+1) \times (2M+1)$ points $(M < \frac{1}{2}N)$ satisfies an autoregression of order $2M+1$ over each row and each column.

PROOF. Without loss of generality we can assume that the $(2M+1) \times (2M+1)$-point band-width consists of the square region containing $(2M+1) \times (2M+1)$ points centered at the origin. This is the usual case for a 2-D band-limited (low-pass) signal. However, the proof can be modified for an arbitrary square support of $(2M+1) \times (2M+1)$ points if necessary.

Proceeding similarly to the 1-D problem, define

$$S_1(k_1, k_2) = 1 + \prod_{i=-M}^{M} (Z_{k_1} - Z_i),$$

$$k_1, k_2 = -\tfrac{1}{2}N+1, \ldots, \tfrac{1}{2}N, \qquad (35)$$

where Z_i was previously defined in (14) (note that $S_1(k_1, k_2)$ is actually independent of k_2). Then

$$S_1(k_1, k_2) = \begin{cases} 1, & \text{if } -M \leqslant k_1 \leqslant M \\ & \quad \text{and } -\tfrac{1}{2}N+1 \leqslant k_2 \leqslant \tfrac{1}{2}N, \\ \neq 1, & \text{otherwise.} \end{cases}$$

$$(36)$$

The fact that $x(n_1, n_2)$ is band-limited in the n_1-direction can be written as

$$X(k_1, k_2) = X(k_1, k_2) S_1(k_1, k_2), \qquad (37)$$

which in the spatial domain becomes

$$x(n_1, n_2) = x(n_1, n_2) * s_1(n_1, n_2). \qquad (38)$$

Simplifying (35) by multiplying different terms and collecting the coefficients $c(n_1)$ for identical powers of Z_{k_1}, and noting that $\prod_{i=-M}^{M}(-Z_i) = -1$, we obtain

$$S_1(k_1, k_2) = \sum_{n_1=1}^{2M+1} c(n_1) Z_{k_1}^{n_1}$$

$$= \sum_{n_1=1}^{2M+1} c(n_1) e^{-j2\pi k_1 n_1/N}. \qquad (39)$$

Equations (35) and (39) imply that

$$s_1(n_1, n_2) = \mathscr{F}^{-1}\left\{ 1 + \prod_{i=-M}^{M} (Z_{k_1} - Z_i) \right\}$$

$$= \begin{cases} c(n_1), & \text{if } 1 \leqslant n_1 \leqslant 2M+1 \\ & \quad \text{and } n_2 = 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (40)$$

Substituting (40) in (38) results in the autoregression

$$x(n_1, n_2) = \sum_{i=1}^{2M+1} c(i) x(n_1 - i, n_2), \qquad (41)$$

which is an autoregression of order $2M+1$ over each row of the signal $x(n_1, n_2)$.

By interchanging the roles of k_1 and k_2 in the proof. we obtain the same autoregression over the columns of $x(n_1, n_2)$. $\quad\square$

Note that the autoregression coefficients are identical to those we found for 1-D band-limited signals in Theorem 1.

COROLLARY 2.1. If the signal band-width lies inside a rectangle of size $(2M_1+1) \times (2M_2+1)$, then the signal rows satisfy an autoregression of order $2M_1+1$ while the signal columns satisfy an autoregression of order $2M_2+1$.

COROLLARY 2.2. A $(2M+1) \times (2M+1)$ band-limited signal satisfies the higher order auto-regressions obtained for $(2(M+a)+1) \times 2(M+b)+1)$-point $(a, b > 0)$ band-limited signals, if the larger band-width includes the smaller band-width.

COROLLARY 2.3. A necessary and sufficient condition for a 2-D signal to have a $(2M+1) \times (2M+1)$-point band-width centered at the origin is that all of its rows and columns satisfy the autoregression (20).

When a two-dimensional band-limited signal is observed over a rectangular region with sides equal to those of the signal band-width, the autoregression in (20) implements an exact extrapolation algorithm. To perform a 2-D extrapolation, 1-D extrapolations can be

done over rows and columns in parallel. The separation of the 2-D extrapolation into decoupled 1-D extrapolations, and the resulting parallelization of the procedure, makes this extrapolation procedure fast, and therefore attractive.

4.2. Derivation of K-D algorithm

Consider a K-D discrete signal whose K-D DFT has a band-width of $2M + 1$ points along all directions of the Cartesian coordinate axes. The following theorem can be proved for this class of signals.

THEOREM 3. A K-D discrete signal $x(n_1, n_2, ..., n_k)$ whose K-D DFT $X(k_1, k_2, ..., k_k)$ is band-limited to $2M + 1 < N$ points in all directions satisfies an auto-regression of order $2M + 1$ over each direction.

PROOF. The same method used in the proof of Theorem 2 can be applied. Thus the proof is omitted.

COROLLARY 3.1. If the signal band-width is different in different directions, then the signal satisfies auto-regressions of corresponding orders in different directions.

COROLLARY 3.2. A K-D band-limited signal with a certain band-width in a specific direction satisfies the higher order autoregressions obtained for K-D band-limited signals with a larger band-width in that particular direction, if the larger band-width includes the original signal's band-width in that direction.

COROLLARY 3.3. A necessary and sufficient condition for a signal to have a $2M + 1$ point band-width in all directions is that it satisfies the autoregression (20) in all directions.

When a multi-dimensional band-limited signal is observed over a region with sides equal to those of the signal band-width, the autoregression (20) implements an exact extrapolation algorithm. To perform a K-D extrapolation, 1-D extrapolations can be done over different directions, in parallel. This makes the extrapolation procedure fast, and therefore attractive.

5. Numerical results

In this section some numerical examples are given, in a variety of noise situations. Some possibilities for regularization in the presence of wideband noise, which makes the problem ill-posed, are also briefly discussed. Three noise situations are examined:
1. noiseless data;
2. additive band-limited noise;
3. additive wide-band noise.
Since 2-D and K-D band-limited signals satisfy an auto-regression identical to that of 1-D band-limited signals in all directions, it suffices to investigate the performance of the extrapolator for 1-D band-limited signals. However, for illustration a 2-D example is also given.

5.1. Noiseless data

In Fig. 1 a band-limited signal whose 64-point DFT has a 9-point support is shown. The magnitude of its DFT is shown in Fig. 2. The extrapolator coefficients for this situation are $[c(1), ..., c(9)] = [8.7136, -34.0200, 78.1091, -116.2225, 116.2225, -78.1091, 34.0200, -8.7136, 1.0000]$. Figure 3 shows the nine points from which the signal is to be extrapolated. The extrapolated signal is shown in Fig. 4; it can be seen that the signal is extrapolated perfectly. A 2-D example is shown in Figs. 5–8.
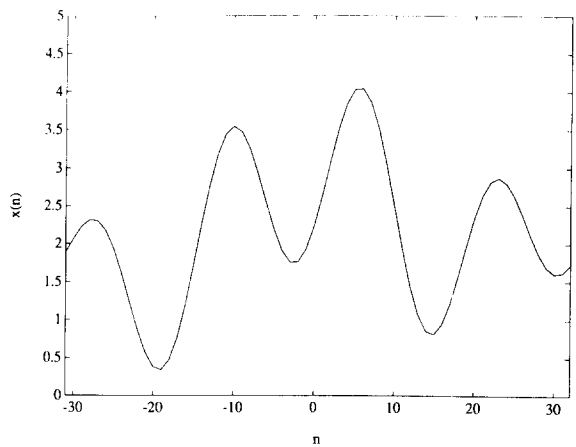

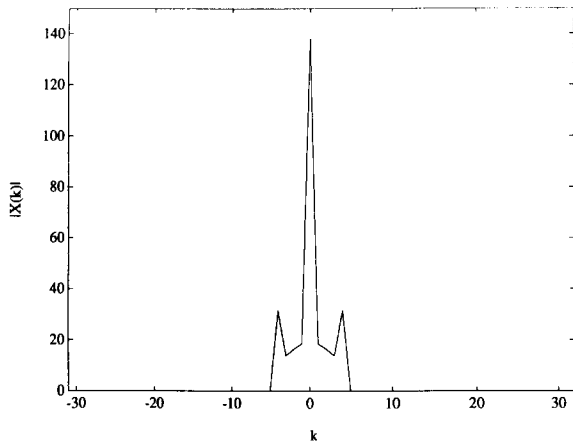Fig. 1. A band-limited signal whose 64-point DFT has a 9-point band-width.

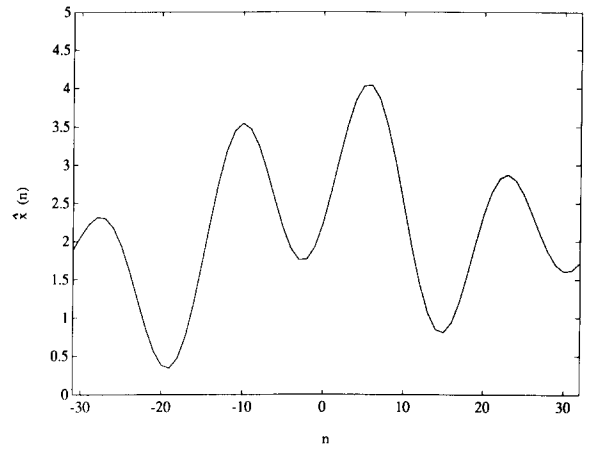Fig. 2. DFT magnitude of the band-limited signal shown in Fig. 1.

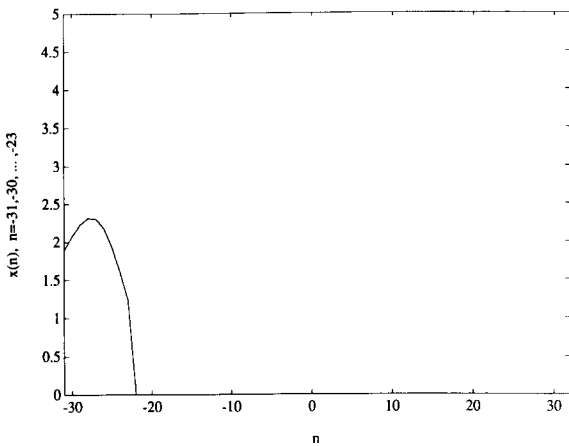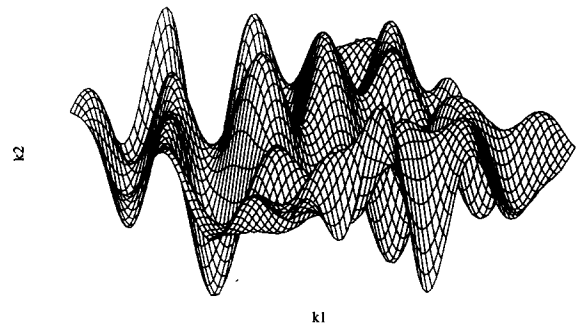Fig. 4. Extrapolated version of the signal shown in Fig. 1.

Fig. 3. Known values of the signal in Fig. 1 which are used for extrapolation.

Fig. 5. A 2-D band-limited signal whose (64×64)-point DFT has a (9×9)-point band-width.

Fig. 6. DFT magnitude of the band-limited signal shown in Fig. 5.

5.2. Additive band-limited noise

Suppose the given data $\{x(n), r+1 \leqslant n \leqslant r+L\}$ are contaminated with additive *band-limited* noise, where the band-width of the noise is equal to or less than the band-width of $x(n)$. One way this situation can arise is for the band-limitation to occur *after* the signal is contaminated with noise. For example, seismic data are band-limited due to the frequency response of the measuring equipment (seismometers); any noise in the data will be similarly band-limited.

Since the signal plus noise is still band-limited, the given noisy data may still be extrapolated using (20). Although it might seem that extrapolation should mag-

nify the noise, the signal-to-noise ratio (SNR) of the extrapolated signal will be the same as the actual SNR, since the *signal-plus-noise* is reconstructed perfectly. Indeed, there is no direct procedure for distinguishing a band-limited signal from similarly band-limited noise; other a priori information must be used.
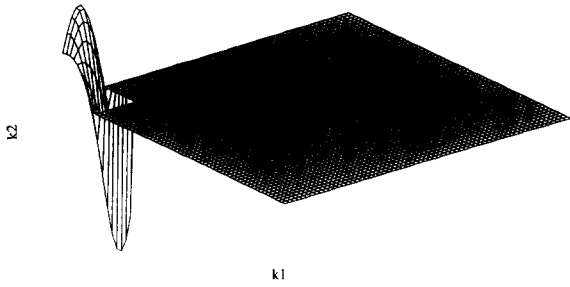
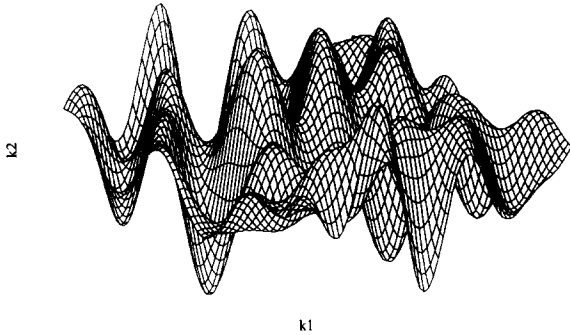Fig. 7. Known values of the signal in Fig. 5 which are used for extrapolation.

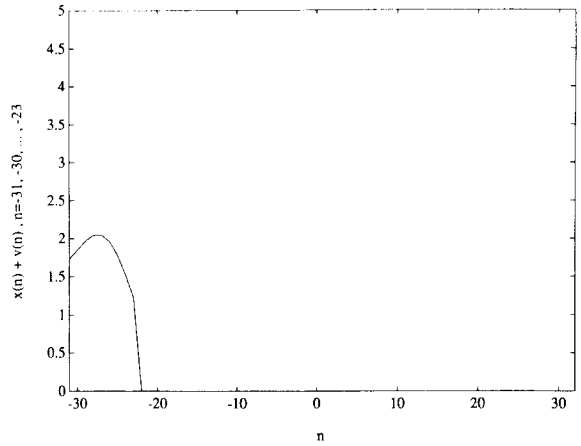Fig. 8. Extrapolated version of the signal shown in Fig. 5.

Fig. 9. Noisy observations (with band-limited noise) of the signal in Fig. 1.

Fig. 10. Extrapolated signal using the noisy observations shown in Fig. 9; the original (noise-fee) signal is drawn in dotted line.

Once the signal-plus-noise has been completely extrapolated, standard noise-reducing signal processing techniques may be employed. For example, if the signal and noise are both realizations of random processes, a Wiener filter may be used to compute the linear least-squares estimate of the signal.

Two examples of extrapolation from signals with additive band-limited noise are illustrated in Figs. 9–12. Figures 9 and 11 show the noisy data; Figs. 10 and 12 the extrapolations. It can be seen that the noise is not amplified unduly in the reconstructed signals.

5.3. Additive wide-band noise

Now suppose the given data $\{x(n), r+1 \leqslant n \leqslant r+L\}$ are contaminated with additive *wide-band* noise having a band-width greater than the band-width of $x(n)$. This is a well-known ill-posed problem [5, 9–12, 19–21, 23]. Now extrapolation using (20) will give poor results, since (20) was derived under the assumption that the data is a segment of a band-limited signal, which is no longer true. (In fact, if white noise is added, the extrapolated signal tends to resemble

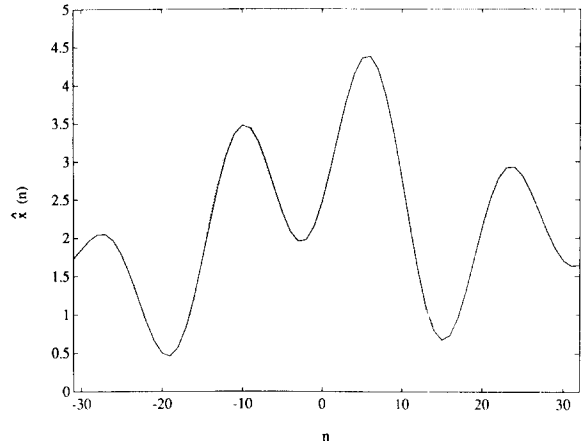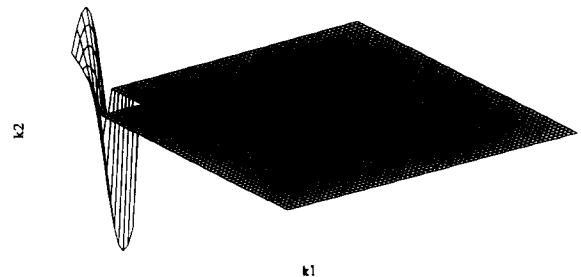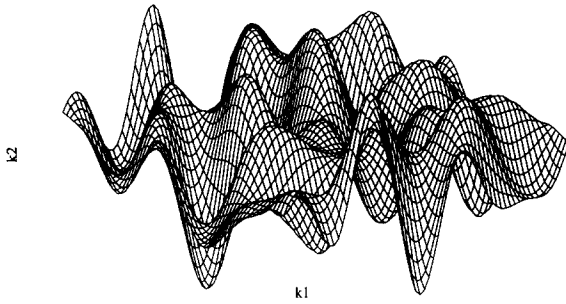Fig. 11. Noisy observations (with band-limited noise) of the signal in Fig. 5.

Fig. 12. Extrapolated signal using the noisy observations shown in Fig. 11.

the impulse response of (20).) Usual noise-reducing signal processing techniques cannot be used, since only a segment of the noisy signal is available. What can be done depends on the relative sizes of the bandwidth $2M + 1$ and data length L.

If $L = 2M + 1$, the given data can be low-pass filtered by convolving it with a discrete sinc function, windowing to zero data values outside the interval $[r + 1, r + L]$ [6]. This will introduce edge artifacts into the spectrum, since the window will convolve the spectrum with a discrete sinc function. However, if the low-pass filter cut-off frequency is less than the maximum signal frequency, the resulting filtered data will be approximately band-limited. Then the filtered data can be extrapolated using (20), with reduced error. Another possibility is to perform a wavelet decomposition of the data, and discard the high-frequency components. However, 'high-frequency' will be defined in terms of the wavelet basis functions, not the complex exponentials, so this will be approximate at best.

If $L > 2M + 1$, and the noise is band-limited to $[-\frac{1}{2}(L-1), \frac{1}{2}(L-1)]$, we can do much better by using an L-point extrapolator, rather than a $(2M + 1)$-point extrapolator, to extrapolate exactly the signal-plus-noise, and then low-pass filter the result to be band-limited to $[-M, M]$. Wiener filtering could then be used to reduce the noise present in the frequency band $[-M, M]$.

Still another possibility is to use a procedure similar to one suggested in [4]. The error in performing the extrapolation using the $(2M + 1)$-point extrapolator on noisy data $y(n) = x(n) + v(n)$ is $s(n) * v(n)$. Here '$s(n) *$' symbolizes the extrapolation operation. If

$v(n)$ can be simulated, e.g., by using a shaping filter on white noise, the extrapolation error can be approximately compensated.

6. Conclusion

A new procedure for extrapolation of discrete band-limited signals has been presented. Unlike previous methods requiring solution of a system of equations, or repeated iterations, the new procedure is a simple auto-regression. A matrix form of the algorithm, which is also simple and may be more numerically stable, was also presented. The method extends to two- and higher-dimensional signals; in these cases the algorithm is highly parallelizable. Numerical examples for both 1-D and 2-D signals demonstrated the procedure.

If the data length and band-width are unequal, or if noise is added to the data, some sort of regularization is required. Regularization procedures were proposed for (1) data length less than band-width, (2) data length greater than band-width, (3) additive band-limited noise, and (4) additive wide-band noise. We believe that the simplicity of the algorithm sheds new light on the band-width extrapolation problem.

Acknowledgments

Appendix A

In this appendix we compare our procedure with the procedure of [6] (specified by (6) and (7)), and explain why our procedure, unlike that of [6], avoids solving a linear system of equations.

A.1. Matrix formulation of the problem

A matrix formulation will be used throughout. For

either procedure, the band-limited constraint can be written in the frequency domain as $X(k)S(k) = X(k)$, and in the time domain as $x(n) * s(n) = x(n)$. Defining the matrix S as having (i, j)th element $S(i - j)$ and the vector $x = [x(-\frac{1}{2}N+1), \ldots, x(\frac{1}{2}N)]^T$, $x(n) * s(n) = x(n)$ can be written in matrix form as $Sx = x$.

Now suppose $x(n)$ is band-limited in frequency to the interval $[-M, M]$, and values of $x(n)$ are given in time in the interval $[-M, M]$. Let $x_0 = [x(-M), \ldots, x(M)]^T$ be the vector of given values of $x(n)$. Partitioning $Sx = x$ appropriately results in

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \begin{bmatrix} a \\ x_0 \\ b \end{bmatrix} = \begin{bmatrix} a \\ x_0 \\ b \end{bmatrix}, \tag{42}$$

where $a = [x(-\frac{1}{2}N+1), \ldots, x(-M-1)]^T$ and $b = [x(M+1), \ldots, x(\frac{1}{2}N)]^T$ are vectors of the unknown values of $x(n)$. The band-width extrapolation problem is to solve (42) for a and b.

A.2. The procedure of Jain et al.

In [6] $S(k)$ is chosen to be (10), so that $s(n)$ is a discrete sinc function (8). Since the elements of S are almost all nonzero, the solution of (42) requires some work. In the present formulation, the procedure of [6] can be summarized as follows. First, solve the system of equations $S_{22}z = x_0$ for the $(2M + 1)$-vector z (note that this is the system (6)). Using the middle equation of (42),

$$z = S_{22}^{-1}x_0 = S_{22}^{-1}S_{21}a + S_{22}^{-1}S_{23}b + x_0. \tag{43}$$

Second, compute the unknown a and b from z using (note that this implements the low-pass filtering operation (7))

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \begin{bmatrix} 0 \\ z \\ 0 \end{bmatrix} = \begin{bmatrix} S_{12}z \\ x_0 \\ S_{32}z \end{bmatrix} = \begin{bmatrix} a \\ x_0 \\ b \end{bmatrix}. \tag{44}$$

The second equality in (44) follows from the first and third equations of (42) and the verifiable relations (valid for $s(n)$ being the discrete sinc function)

$$\begin{aligned} S_{11} &= S_{12}S_{22}^{-1}S_{21}, & S_{13} &= S_{12}S_{22}^{-1}S_{23}, \\ S_{31} &= S_{32}S_{22}^{-1}S_{21}, & S_{33} &= S_{32}S_{22}^{-1}S_{23}. \end{aligned} \tag{45}$$

Note that the third relation of (45) is the transpose of the second, since S is symmetric, and that various relations state that various Schur complements of various submatrices of S are identically zero.

A.3. Our procedure

In this paper we have chosen $S(k)$ to be (12), not (10), so that $s(n) = 0$ unless $1 \le n \le 2M + 1$. Hence the matrix S is *banded*. Equation (42) for this choice of $S(k)$ has the form (note that S is also circulant)

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & * & * & \cdots & * & * & * \\ * & 0 & \cdots & 0 & 0 & * & \cdots & * & * & * \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & * & * & \cdots & * & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & * & \cdots & * & * & 0 \end{bmatrix} \begin{bmatrix} a \\ x_0 \\ b \end{bmatrix}$$

$$= \begin{bmatrix} a \\ x_0 \\ b \end{bmatrix}. \tag{46}$$

where each $*$ represents a non-zero value. Because of this banded structure, the elements of a and b can be recursively computed from x_0, without any of the above machinery. In particular, the $(\frac{1}{2}N+M+1)$th row of (46) computes $x(M+1)$ from $\{x(-M), \ldots, x(M)\}$; succeeding rows compute other values of $x(n)$. The simplicity of our procedure thus stems from the use of a *banded* matrix S, instead of the non-banded S matrix of discrete sinc functions.

References

[1] K.S. Arun and L.C. Potter, "Existence and uniqueness of band-limited positive semidefinite extrapolation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-38, No. 3, March 1990, pp. 547–549.

[2] H. Bialy, "Iterative Behandlung linearen Funktional-gleichungen", *Arch. Rational Mech. Anal.*, Vol. 4, 1959, pp. 166–176.

[3] S.D. Cabrera and T.W. Parks, "Extrapolation and spectral estimation with iterative weighted norm modification", *IEEE Trans. Signal Process.* Vol. 39, No. 4, April 1991, pp. 842–851.

[4] T.S. Huang, *Advances in Computer Vision and Image Processing*, JAI Press, London, 1984, Chapter 1.

[5] T.S. Jaffe, "Band-limited extrapolation through iterative deconvolution". *SPIE Inverse Optics II*, Vol. 558, 1985, pp. 73–79.

[6] A.K. Jain and S. Ranganath, "Extrapolation algorithms for discrete signals with application in spectral estimation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-29, No. 5, October 1981, pp. 830–845.

[7] S.Y. Kung, K.S. Arun and D.V. Bhaskar Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem", *J. Opt. Soc. Amer.*, Vol. 73, No. 12, December 1983, pp. 1799–1811.

[8] B.C. Kuo, *Digital Control Systems*. Holt, Rinehart & Winston, New York, 1980, pp. 195–196.

[9] H. Lee and T.S. Huang, "On discrete band-limited signal extrapolation", *Proc. Internat. Conf. Acoust. Speech Signal Process. '85*, 1985, pp. 465–468.

[10] H. Lee, Z.-C. Lin and T.S. Huang. "Performance and limitation of discrete band-limited signal extrapolation", *Proc. Internat. Conf. Acoust. Speech Signal Process. '86*, 1986, pp. 1645–1648.

[11] H. Lee. D.P. Sullivan and T.S. Huang, "Improvement of discrete band-limited signal extrapolation by iterative subspace modification". *Proc. Internat. Conf. Acoust. Speech Signal Process. '87*, 1987, pp. 1569–1572.

[12] K. Nakayama. "A band-limited signal extrapolation algorithm using pseudo inverse filtering heuristic optimization", *Proc. Internat. Conf. Acoust. Speech Signal Process. '88*, 1988, pp. 1754–1757.

[13] A. Papoulis. "A new algorithm in spectral analysis and band-limited extrapolation", *IEEE Trans. Circuits and Systems*, Vol. CAS-22, No. 9, September 1975, pp. 735–742.

[14] C.K. Rushfort, A.E. Crawford and Y. Zhou, "Least squares reconstruction of objects with missing high frequency components", *J. Opt. Soc. Amer.*, Vol. 72, No. 2, February 1982, pp. 204–211.

[15] J.L.C. Sanz and T.S. Huang, "Some aspects of band-limited signal extrapolation: Models, discrete approximations, and noise", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-31, No. 6, December 1983, pp. 1492–1501.

[16] J.L.C. Sanz and T.S. Huang, "Band-limited signal extrapolation in the presence ot noise", *Proc. Internat. Conf. Acoust. Speech Signal Process. '84*, 1984, pp. 31.6.1–31.6.4.

[17] J.L.C. Sanz and T.S. Huang, "Discrete and continuous band-limited signal extrapolation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-31, No. 5, 1983, pp. 1276–1285.

[18] H. Soltanian-Zadeh and A.E. Yagle, "A wavenumber-domain Kalman filtering approach to one-dimensional phase retrieval", *Proc. Internat. Conf. Acoust. Speech Signal Process. '90*, Vol. 3, April 1990, pp. 1555–1558.

[19] B.J. Sullivan, "A comparison of three methods of discrete-time signal extrapolation", *Proc. Internat. Conf. Acoust. Speech Signal Process. '85*, 1985, pp. 469–472.

[20] B.J. Sullivan and B. Liu, "On the use of singular value decomposition and decimation in discrete-time band-limited signal extrapolation". *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-32, No. 6, December 1984, pp. 1201–1212.

[21] M.J. Tsai and D.A. O'Connor, "Experiments with extrapolation or band-limited signal", *Proc. Internat. Conf. Acoust. Speech Signal Process. '84*, 1984, pp. 31.9.1–31.9.4.

[22] G. Wahba, On the numerical solution of Fredholm integral equations of the first kind, Technical Report No. 217, Department of Statistics, University of Wisconsin, 1969.

[23] C. Yan-Yi and K. Nakayama, "A band-limited signal extrapolation algorithm using waveform synthesis digital filter", *IEEE/IECE Global Telecommumications Conf. Recording*, Vol. 1, 1987, pp. 234–238.