INFORMATION AND CONTROL 44, 320–338 (1980)

A Robustized Vector Recursive Stabilizer Algorithm for Image Restoration*

IVAN KADAR

IBM T. J. Watson Research Center,† P. O. Box 218, Yorktown Heights, New York 10598

AND

LUDWIK KURZ

Polytechnic Institute of New York, Department of Electrical Engineering, Brooklyn, New York 11201

The ill-posed problem of object reconstruction (or band-limited extrapolation) is reformulated in the framework of the general linear model in new recursive parametric forms. The resultant algorithms are shown to be natural stabilizers of the inherent instabilities of both the iterative and noniterative reconstruction/band-limited extrapolation methods. Both robustized and unrobustized versions of the algorithms are given. The recursive algorithms provide immunity to measurement noise outliers in burst noise of high variance. Unlike procedures suggested previously, these methods eliminate the need for stopping rule constraints and ensure convergence of the algorithms. The recursive formulation of the noniterative method of band-limited extrapolation is also found to be adaptable to multidimensional image restoration. Computer simulations verify the theory and demonstrate the computational efficiency of the method.

INTRODUCTION

Consider the well-known problem of bandlimited extrapolation or spectral estimation. Given a finite segment $g(t)$ of a band-limited signal $f(t)$,

$$g(t) = f(t)\, p_T(t), \qquad p_T(t) = \begin{cases} 1, & |t| < T, \\ 0, & |t| > T, \end{cases} \tag{1}$$

* This work was supported, in part, by the Joint Services Electronics Program under Contract F49620-78-C-0074. This paper is based, in part, on a part of a dissertation submitted by the first author to the Faculty of the Polytechnic Institute of New York in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Electrical Engineering), June 1977.

† Work performed, in part, when Ivan Kadar was with Grumman Aerospace Corporation, Bethpage, N. Y. 11714.

0019-9958/80/030320-19$02.00/0

with $f(t) \leftrightarrow F(\omega) = 0$, $\omega > \sigma$, it is required to extrapolate the unknown portion of $f(t)$ in terms of $g(t)$. By generalizing the time and frequency variables to other spaces (e.g., spatial extent and spatial frequency) and extending the method to two or higher dimensions, the above problem can be restated in terms of object reconstruction or image restoration.

The importance of this problem in signal processing is evident by the renewed interest in extrapolating bandlimited signals and computing their Fourier transform from noisy measurements which can only be taken over a finite segment of the signal (Viano, 1976; Gerchberg, 1974). This problem is of great importance in spectral analysis (Papoulis, 1975; Youla, 1977; Gerchberg, 1974) in optical or antenna systems where one desires to extend the resolution of the instrument beyond the diffraction limit by superresolution (Andrews, 1970; Papoulis, 1968), or in picture processing, where the object (or the entire image) is reconstructed by extrapolating a truncated observed portion (Gerchberg, 1974).

In this paper the iterative and noniterative problems of band-limited extrapolation (or object reconstruction or image restoration) (Gerchberg, 1974; Papoulis, 1975; Viano, 1976; Youla, 1976) are posed in a new recursive framework. The new methods are based upon the adaptation and extensions of the general linear statistical model representation. The compensation of the rapid divergence in the presence of noise of the ill-posed reconstruction methods by robust iterative stabilizing algorithms is the main thrust of this paper.

The first part of the paper is used to set up the framework for both the nonrecursive prolate spheroidal expansion method (Slepian *et al.*, 1961; Viano, 1976) and the iterative methods of Gerchberg (1974), Papoulis (1975), and Youla (1976) to introduce the instabilities associated with the improperly posed problems. Subsequently, two new recursive stabilizing algorithms are introduced providing alternative formulations and computational flexibilities.

In the first algorithm, the iterative method of object reconstruction (or image restoration) (Papoulis, 1975; Youla, 1977) is reformulated in a new way to find parameters rather than functions. With a block of samples taken from an iterative measurement equation of a time-varying signal in additive noise, the block sample amplitudes become a vector of parameters to be estimated in the presence of noise at each iteration step. This leads directly to the use of the multidimensional extension of Gladyshev's Minimum Variance Least-Squares Stochastic Approximation method (SAMVLS) (Kadar and Kurz, 1980a), and subsequently, to the batch preprocessed Mann–Whitney–Wilcoxon Nonparametric Statistic (MWWNS) (Mann and Whitney, 1946) robustized vector SAMVLS (Kadar and Kurz, 1980b).

The robustized SAMVLS algorithm provides immunity to measurement noise outliers in unspecified contaminated noise environments. The method is also both computationally efficient and requires storing of only the last block of batched preprocessed data samples. This represents substantial savings in

storage requirements over the direct application of the noniterative prolate spheroidal expansion method (Viano, 1976). The resultant new robust algorithm is equivalent to a natural stabilizer of the measurement-noise-caused instabilities of the iterative method without the need for stopping rule constraints on the recursion. Computer simulation comparison of the iterative algorithm of Papoulis (1975) with the new SAMVLS stabilizer algorithm in mixture noise demonstrate the advantages of stable reconstruction via SAMVLS.

In the last part of this paper, an alternative recursive stabilizing algorithm is introduced using the Robbins–Monro Stochastic Approximation (RMSA) (Kersten and Kurz, 1976) in the spirit of Kashyap and Blaydon (1968) to estimate the coefficients of general orthogonal expansion methods, e.g., the prolate spheroidal expansion method, or expansions using other complete set of basis functions, such as Walsh functions (Young and Calvert, 1974).

Although not explicitly explored in this paper, the application of Walsh basis functions promises to be a particularly attractive alternative to prolate spheroidal wavefunctions. Walsh basis functions facilitate both recursive generation of eigenfunctions at each iteration step of the recursion and eliminate the storage requirements associated with prolate spheroidal wavefunctions. Furthermore, the extension of the procedure to two- or-higher dimensions via the Walsh–Hadamard relationship (Young and Calvert, 1975) is straightforward.

The Prolate Spheroidal Wavefunction Expansion Method

One possible approach to the theoretical solution of the extrapolation problem has been well known, in the absence of measurement noise, by the use of prolate spheroidal wavefunctions (Viano, 1976; Papoulis, 1975; Slepian *et al.*, 1961).

Given a finite segment of a signal, $f(t)$,

$$g(t) = f(t)\, p_T(t), \qquad p_T(t) = \begin{cases} 1, & |t| < T, \\ 0, & |t| > T, \end{cases} \tag{1}$$

with $f(t) \leftrightarrow F(\omega) = 0$, $|\omega| > \sigma$, and it is further assumed that $g(t) = 0$, $|t| > T$. The resultant extrapolated signal is then obtained in the form of an expansion (Viano, 1976; Papoulis, 1975) as

$$f(t) = \sum_{k=1}^{\infty} g_k (\lambda_k)^{1/2}\, \phi_k(t) \tag{2}$$

and

$$\lim_{N \to \infty} \int_{-T}^{T} \left[g(t) - \sum_{k=1}^{N} g_k \frac{\phi_k(t)}{\lambda_k^{1/2}} \right]^2 dt \to 0, \tag{3}$$

where

$$g_k = (1/\lambda_k) \int_{-\infty}^{\infty} g(t)\, \phi_k(t)\, dt.$$

The ϕ_k are the eigenfunctions and the λ_k are the eigenvalues of the associated integral equation (Slepian *et al.*, 1961). The eigenvalues λ_k are such that (Slepian *et al.*, 1961) $1 > \lambda_0 > \cdots \lambda_k > \cdots > 0$ and $\lambda_k \to 0$ as $k \to \infty$. Hence, if $g(t)$ is corrupted by noise, i.e.,

$$y(t) \triangleq g(t) + v(t) \tag{4}$$

for k large λ_k is small and an arbitrary small noise on the measurements can create large errors in the extrapolated signal due to the smoothing action of the kernel of the associated integral equation (Viano, 1976). As a matter of fact, as $k \to \infty$, $\lambda_k \to 0$ and the problem is improperly posed (Viano, 1976 and Youla, 1976), (and becomes unstable as a limit point) even in the absence of noise since the solution depends uniquely, but not continuously, on the data. This condition in the presence of noise requires stabilizing constraints in the form of stopping rules on the number of coefficients used in the expansions, some of which are given by Viano (1976).

The instabilities associated with the noniterative procedure also occur in the iterative algorithms. This is further amplified in the next section.

The Iterative Algorithm

Papoulis (1975) has recently developed an iterative method for computing the Fourier transform of a bandlimited function $f(t) \leftrightarrow F(\omega) = 0$, $|\omega| > \sigma$ from a given time-limited segment $g(t) = f(t)\, p_T(t)$, and for extrapolating the bandlimited function (object reconstruction). An algorithm similar to the one proposed by Papoulis (1975) has been applied by Gerchberg (1974) in picture processing with real data corrupted by uniform (white) Gaussian noise.

Starting with $G(\omega) = G_0(\omega) \leftrightarrow g(t) = g_0(t)$ at the nth iteration step, one forms

$$F_n(\omega) = G_{n-1}(\omega)\, p_\sigma(\omega), \qquad p_\sigma(\omega) = \begin{cases} 1, & |\omega| < \sigma \\ 0, & |\omega| > \sigma \end{cases}$$

by truncating $G_{n-1}(\omega)$ within $(-\sigma, \sigma)$. Then, compute the inverse transform

$$f_n(t) \leftrightarrow F_n(\omega).$$

Next form the function

$$g_n(t) = f_n(t) + [f(t) - f_n(t)]p_T(t) = \begin{cases} g(t), & |t| < T \\ f_n(t), & |t| > T \end{cases}$$

by replacing the segment of $f_n(t)$ in the interval $(-T, T)$ by the known segment $g(t)$ of $f(t)$. The nth step ends by computing $G_n(\omega) \leftrightarrow g_n(t)$ and the function $g_n(t)$ so formed.

The proof of the convergence is based upon the properties of prolate spheroidal functions, and it can be shown (Papoulis, 1975) that the iterative method converges in the mean-square sense. Therefore, as one would expect, the effects of measurement noise, roundoff error, and aliasing due to the finite sample representation can cause instabilities in the method, which is handled by methods equivalent to stopping rule constraints (Viano, 1976) and the limit point is never reached. However, the important point is that the problem becomes unstable only as a limit point in the absence of noise. The presence of noise, however, introduces instabilities due to the monotonically decreasing eigenvalues in the expansion methods or equivalently due to the smoothing effect of the convolution term in the recursive method. This gives rise to large errors due to the noise perturbations.

In the presence of burst noise of large variance, the recursion might completely diverge. This method has no immunity to outliers. The procedure, by itself, cannot be robustized (i.e., made insensitive to the effects of changes of the noise distributions); and even in the absence of outliers, the instability becomes more pronounced as the recursion progresses by having its own instability problems.

The instability becomes more pronounced as the recursion progresses. However, by combining the recursive procedure with a form of stochastic approximation (SA) in a regression framework, i.e., SAMVLS (Kadar and Kurz, 1980a), one recursively estimates parameters (i.e., the amplitude of a time-varying signal) and the noisiness of the parameter estimates reduce with each step. In the limit the recursive parameter estimator becomes asymptotically Gaussian with variance approaching zero. Therefore, the SA procedure ideally compensates for the instability problem since the parameter estimates improve as the recursion progresses, which adaptively corrects for the actual instabilities caused by noise which would otherwise become more pronounced as the recursion progresses. Hence, SA in this case provides a continuous compensation for instabilities and the stopping rules need only be imposed from practical considerations on the dimensionality of the block sample representation of the restored image. As a matter of fact, in practical problems using the SA method, the limit point is never reached because the recursion is stopped long before. An additional advantage of the SA method is that no assumptions need to be made about the underlying noise distributions other than the samples i.i.d. at each iteration step. Furthermore, one can robustize the procedure which guarantees near optimum convergence rate and insensitivity to changes in the noise distribution.

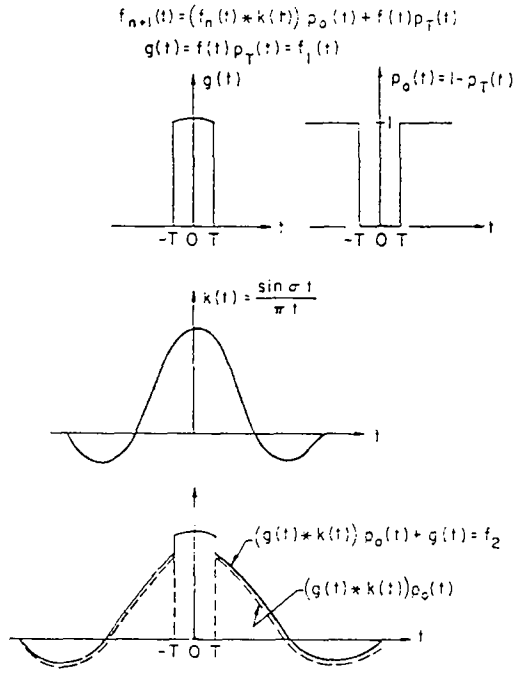

FIG. 1. Signals at first iteration.

THE PARAMETRIZED ITERATIVE OBJECT RECONSTRUCTION ALGORITHM

Consider the object reconstruction problem from a truncated measured (known) portion, $g(t) = f(t)\, p_T(t)$ for which the iterative algorithm described in (Papoulis, 1975) expressed in a convolution form becomes (refer to Fig. 1):

$$g_{n-1}(t) = [g_n(t) * k(t)]\, p_0(t) + f(t)\, p_T(t), \qquad n = 1, 2, \ldots, \tag{5}$$

where

$$g_1(t) = f(t)\, p_T(t); \qquad p_0(t) \triangleq 1 - p_T(t)$$

and

$$k(t) = (\sin \sigma t)/\pi t.$$

It is clear that

$$g_{n+1}(t) = g(t), \qquad t < T,$$
$$= f_{n+1}(t), \qquad t > T.$$

In this case the kernel $k(t)$ represents the band-limitedness of $f(t)$, i.e., $f_n(t) = g_{n-1}(t) * k(t)$.

If $g(t)$ is corrupted by noise, then the known signal portion $g(t) = f(t)\, p_T(t)$ in (5) is replaced with

$$y_n(t) \triangleq g(t) + V_n(t), \tag{6}$$

where $V_n(t)$ is an additive noise term, i.i.d. for every step n of the iteration. At this point no further assumptions need to be made about the statistics of the noise. At each step of the iteration starting with $n = 1$, a block of samples is taken of all terms in (5). If the functions under consideration are band limited, then the number of samples needed to represent the functions is fixed. Otherwise, one has to assess the significant frequencies of interest and select the sample size to make the aliasing error vanishingly small. Some of these considerations are mentioned in (Papoulis, 1975; Slepian *et al.*, 1961). It will be assumed that even if the sampling rate is greater than the Nyquist rate, the cross-correlation between the block of signal and noise samples is small enough to be negligible and the samples are i.i.d.

Now let $n = 1, 2, 3, \ldots$ and by taking block samples of each time function, one can represent (1) in terms of "signal," S_n and "noise" $\mathbf{W}_n$ terms as a measurement equation in a general linear model framework

$$\mathbf{g}_{n-1} = S_n \boldsymbol{\alpha} + \mathbf{W}_n, \tag{7}$$

where, say, at some step $n = k \geqslant 1$, $\mathbf{g}$, $\boldsymbol{\alpha}$, and $\mathbf{W}$ are m-vectors corresponding to m samples per block where $\boldsymbol{\alpha}$ is a vector of parameters representing the amplitude of the time-varying signal S_n which is represented by a diagonal $(m \times m)$ matrix whose m diagonal elements are the block samples of S_n,

$$S_n = \operatorname{diag}(S_{11}, S_{22}, \ldots, S_{mm}),$$

where the diagonal entries of S_n are obtained from the block samples of

$$\mathbf{X}_n = \left\{ [I]_{(m \times m)} \sum_{j=0}^{i} g_n(j)\, K(i-j) \right\} \mathbf{p}_0 + \mathbf{g} \tag{8}$$

with

$$X_{1n} = S_{11}, \qquad X_{2n} = S_{22}, \ldots, \qquad X_{mn} = S_{mm}.$$

The sum within the brackets is the discrete convolution of, say, r_g samples of $g_n(t)$ with r_k samples of the kernel $K_n(t)$ such that the total number of samples $m = r_g + r_k - 1$; $\mathbf{p}_0$ and $\mathbf{g}$ are m-dimensional vectors.

The noise term $\mathbf{W}_n$ is given by

$$\mathbf{W}_n = \left\{ [I]_{(m \times m)} \sum_{j=0}^{i} V_n(j)\, K(i-j) \right\} \mathbf{p}_0 + \mathbf{V}_n \tag{9}$$

and it is clear from (5) and (9) be the discrete convolution operation that any noise perturbation is spread into the reconstructed initially unknown signal portion. It is important to note that the dimension, m, grows with each iteration step n due to the convolution operation which is performed before the recursion for g_{n+1} is iterated. This is not unexpected since the algorithm is extrapolating the known signal segment[1].

It should be noted at this point that the convolution form of the iterative (restoration or extrapolation) algorithm (5) is used rather than the FFT implementation suggested by Papoulis (1975) and Youla (1976). The convolution form can be adapted directly to the stochastic approximation framework with its similarity to batch preprocessing (Kersten and Kurz, 1976) used in SAMVLS (Kadar and Kurz, 1980a), albeit in this case the convolution is among the block (vector) samples and does not require an initial delay. Furthermore, from a practical point of view, charge coupled devices (CCDs) used as analog shift registers are being developed in the 1977–1978 time frame for real-time high speed convolution applications for imaging sensor spaceborne signal processing (Howle, 1977). The convolution operation is performed with CCDs, either by direct storage of the samples and shifting operations, or by using CCDs to implement the FFT algorithm and form the convolution in a two-step operation. There is no published information available at this time on the relative performance and complexity of the two methods. Both theoretical and experimental work is in progress (Howle, 1977).

The SAMVLS Stabilizer Algorithm

The parametric form of the iterative restoration or extrapolation algorithm $\mathbf{g}_{n+1} = S_n\boldsymbol{\alpha} + \mathbf{W}_n$ (7) is of the form which can be viewed as a measurement equation, where the noise term, $\mathbf{W}_n$, can be contaminated by outliers (Kadar and Kurz, 1980b), directly suggests the application of the vector extension of Gladyshev's theorem to Minimum Variance Least-Squares (SAMVLS) (Kadar and Kurz, 1980a) to estimate the time-varying signal parameter, $\boldsymbol{\alpha}$ at each step, n.

However, before we apply the vector SAMVLS algorithm to the iterative reconstruction algorithm we review some important properties needed in the sequel!

In essence, the vector SAMVLS algorithm is a recursive procedure for estimating a multiplicative parameter vector in a linear regression model (Kadar and Kurz, 1980a). The SAMVLS procedure is defined by the recursion equation

$$\hat{\boldsymbol{\alpha}}_{k+1} = \hat{\boldsymbol{\alpha}}_k + \frac{1}{k} A_k(\hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_k)\, \mathbf{Y}(\hat{\boldsymbol{\alpha}}_k, \boldsymbol{\alpha}), \tag{10}$$

[1] If the significant signal portion to be reconstructed is assumed to be of finite support, then, in practical applications, the dimension, m, can be held constant.

where $\hat{\alpha}_k$ (the current estimate of α), $Y(\hat{\alpha}_k, \alpha)$, and α are all m-dimensional vectors, $A_k(\cdot)$ is a diagonal $(m \times m)$ adaptive gain matrix, and $E\mathbf{Y}(\hat{\alpha}_k, \alpha) = M(\hat{\alpha}_k, \alpha)$ is linear and has a unique root $\mathbf{0}$ at $\hat{\alpha} = \alpha$. The adaptive gain matrices are assumed to be positive definite with eigenvalues $a_1^{(k)}(\cdot) > \cdots > a_m^{(k)}(\cdot) > 0$, $\sum a_m^{(k)}(\cdot)/k = \infty$ wpl. These assumptions are the multidimensional extension of the scalar SA case of (Robbins and Monro, 1951), generalized and extended by (Kersten and Kurz, 1976).

The following conditions establish the relationship of the i.i.d. random noise vectors $Z_k(\hat{\alpha})$ to the linear regression function, namely, the noise in the measurement equation describing the parametrized linear model is additive and zero mean.

(i) $E[\mathbf{Y}(\hat{\alpha}_k, \alpha)/\hat{\alpha}_k, \ldots, \hat{\alpha}_1] = E[\mathbf{Y}(\hat{\alpha}_k, \alpha)/\hat{\alpha}_k] = M(\hat{\alpha}_k, \alpha) \triangleq M_k$,

(ii) $\mathbf{Y}(\hat{\alpha}_k, \alpha) = M(\hat{\alpha}_k, \alpha) + \mathbf{Z}(\hat{\alpha}_k) = M(\hat{\alpha}_k) + \mathbf{Z}(\hat{\alpha}_k) \triangleq M_k + Z_k$, where $EZ_k = EZ_k(\hat{\alpha}_k) = E[Z_k/\hat{\alpha}_k] = E[\mathbf{Z}_k/\hat{\alpha}_k, \ldots, \hat{\alpha}_1] = 0$ wpl,

(iii) $M(\hat{\alpha}_k, \alpha) = B_k(\hat{\alpha}_k - \alpha)$ has a unique root $\mathbf{0}$ at $\hat{\alpha} = \alpha$.

$\mathbf{Y}(\hat{\alpha}_k, \alpha)$ is obtained directly from the measurement equation describing the parametrized model by cross-correlating the measurements $\xi_k = H_k\alpha_k + Z_k$ with the known signal or measurement matrix H_k, i.e.,

$$\mathbf{Y}(\hat{\alpha}_k, \alpha) = H_k^T[H_k(\hat{\alpha}_k - \alpha) + Z_k] = M_k + Z_k^* = B_k(\hat{\alpha}_k - \alpha) + Z_k^*, \quad (11)$$

from which (iii) follows directly.

To show convergence, the class of regression functions is usually restricted (conditions (iv–vi) of (Kersten and Kurz, 1976)), to be linear in the neighborhood of the root (i.e., at $\hat{\alpha} = \alpha$). These conditions are obviously satisfied by the assumption of a linear regression function and are not stated here. However, to decouple the estimates and to diagonalize the covariance matrix of the estimate an orthogonal transformation P is assumed s.t. $PP^T = I$, $P^TB_kP =$ diagonal, and $P^TA_kP =$ diagonal matrix for each k.

The additive noise vector $\mathbf{Z}(\hat{\alpha})$ should have a uniformly bounded variance and have a well-defined covariance matrix as $\hat{\alpha} \to \alpha$ a.s.

In addition $E\|\xi\|^4 < \infty$ and $E\|B_k\|^4 < \infty$ are needed to guarantee minimum mean-square error estimates. Furthermore, the adaptive gain matrix needs to be well behaved and a consistent "mean-square" estimator of a constant matrix. These requirements (vii–ix) are stated in (Kersten and Kurz, 1976) and are summarized here for convenience:

(viii) (a) $\sup_{\hat{\alpha}} E\|\mathbf{Z}(\hat{\alpha})\|^{2+2\epsilon} < \infty$ for some $\epsilon > 0$.

(b) $\lim_{\hat{\alpha}\to\alpha} E[\mathbf{Z}(\hat{\alpha})\,\mathbf{z}(\hat{\alpha})] = \pi$ where π is a nonnegative definite matrix and where the limit is in the sense of the norm.

(viii) $a_1^{(k)} \geq a_2^{(k)} \geq \cdots \geq a_m^{(k)} > 0$ be the eigenvalues of A_k and $b_1^{(k)} \geq b_2^{(k)} \geq \cdots \geq b_m^{(k)} > 0$ be the eigenvalues of B_k and $a_1 \geq a_2 \geq \cdots \geq a_m > 0$ be the eigenvalues of A.

(a) $0 < a_1' \leq \inf_{\hat{\alpha}} | A_k | \leq \sup_{\hat{\alpha}} | A_k | \leq a'' < \infty$ wpl for k large.

(b) $\underline{\lim}_k a_m^{(k)}(\cdot) \geq a_1' > 0$ wpl where A is a constant matrix s.t. $a' \leq | A | \leq a''$.

(ix) $a_1' b_m^{(k)} - \epsilon > \frac{1}{2}$ and $a_m b_m^{(k)} - \epsilon > \frac{1}{2}$.

Under the above set of assumptions the sequence $\hat{\alpha}_k$ converges wpl to α which minimizes $E | H\alpha - \xi |^2$. The associated theorem establishing convergence in distribution is stated in the application of vector SAMVLS to the parametrized reconstruction algorithm.

To apply SAMVLS to the vector parameter estimation problem, one considers each component of the measurement separately and forms, initially, an m-dimensional SAMVLS (since the dimension, m, grows with each iteration step) which can be considered as m scalar SAMVLS algorithms operating in parallel. Specifically, with $\hat{\alpha}_k$ the current estimate of α, the vector SAMVLS for (5) becomes

$$\hat{\alpha}_{k+1} = \hat{\alpha}_k - \frac{1}{k} A_k(\hat{\alpha}_1, \ldots, \hat{\alpha}_k)[\mathbf{Y}(\hat{\alpha}_k, \alpha)], \tag{12}$$

where, initially, $\hat{\alpha}_k$ is an m-vector, $A_k(\cdot)$ is a diagonal $(m \times m)$ adaptive gain matrix. Since $\hat{\alpha}_1$ is arbitrary in SAMVLS, the optimum choice in this case is to let it equal the amplitude samples of the known signal $g(t)$ for $| t | < T$ and zero for sample values $| t | > T$, making up an $(m \times 1)$ block sample vector. Actually, the iterative measurement equation (algorithm within the SAMVLS algorithm) is always one step behind as the parameter vector is estimated at step k and the inner algorithm updated. $\mathbf{Y}(\hat{\alpha}_k, \alpha)$, $(m \times 1)$, is obtained by correlating the known signal at step $k = n$ with $\mathbf{g}_{n+1}$,

$$S_k^T \mathbf{g}_{k+1} = S_k^T(S_k \alpha + \mathbf{W}_k) \qquad \text{and} \qquad \mathbf{Y}(\hat{\alpha}_k, \alpha) \triangleq M(\hat{\alpha}_k, \alpha) + \mathbf{Z}(\hat{\alpha}_k),$$

which becomes $\mathbf{Y}(\hat{\alpha}_k, \alpha) = S_k^T \mathbf{g}_{k+1} - S_k^T S_k \hat{\alpha}_k$, where $\hat{\alpha}_k$ denotes the current estimate of α. Substituting in (10) the expression for $Y(\hat{\alpha}_k, \alpha)$,

$$\mathbf{Y}(\hat{\alpha}_k, \alpha) = S_k^T[(\alpha - \hat{\alpha}_k) S_k + \mathbf{W}_k]$$

and the regression function is given by

$$E\mathbf{Y}(\hat{\alpha}_k, \alpha) = \mathbf{M}(\hat{\alpha}_k, \alpha) = S_k^T S_k(\alpha - \hat{\alpha}_k),$$

which is linear and has a unique root $\mathbf{0}$ at $\alpha = \hat{\alpha}$.

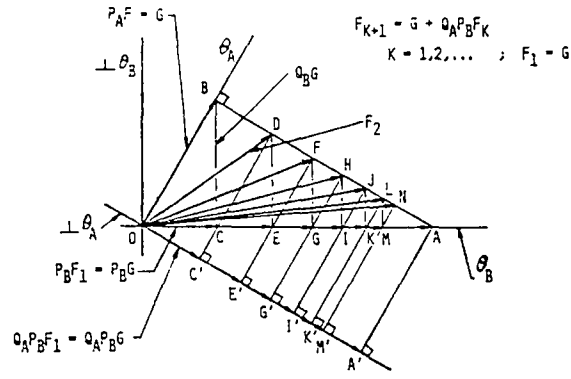

FIG. 2. Geometry of reconstruction in Hilbert space.

It should be noted that the regression function $\mathbf{M}(\hat{\boldsymbol{\alpha}}_k, \boldsymbol{\alpha}) = E\{S_k^T[\mathbf{g}_{k+1} - S_k\hat{\boldsymbol{\alpha}}_k]\}$ is actually a statement of the multidimensional orthogonality principle, with $[\mathbf{g}_{k+1} - S_k\hat{\boldsymbol{\alpha}}_k]$ the "error term" and S_k^T the data. It is well known that the error is orthogonal to the data (Papoulis, 1965) and $E\{S^T[\mathbf{g} - S\boldsymbol{\alpha}]\} = \mathbf{0}$ and $\boldsymbol{\alpha}$ minimizes the mean-square error $E\|\mathbf{g} - S\hat{\boldsymbol{\alpha}}\|^2$. This can be interpreted geometrically in terms of orthogonal projections, very much alike to the method of Youla (1976). In reference to Fig. 2, consider the Hilbert space setting of Youla (1976) (refer to Appendix A) and the orthogonality principle is represented by $(\overline{DE}) \perp (\overline{OE})$ since $(\overline{OD})$ is equivalent to $\mathbf{g}_{k+1}$.

Reflecting at this point to the conditions required by Theorem 1 of (Kersten and Kurz, 1976), the linear regression function satisfies conditions (iv–vi) with $M(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}) = B_k(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$, where $B_k = S_k^T S_k$ is a positive definite diagonal $m \times m$ matrix for each k, s.t. $\|B_k\| < \infty$, since one can reasonably assume that the recursion is terminated after a finite number of steps (with both B_k and A_k diagonal for each k, $P = I$). The additive noise term $\mathbf{Z}(\hat{\boldsymbol{\alpha}}) = S_k^T\mathbf{W}_k$ with $E\mathbf{Z}_k(\hat{\boldsymbol{\alpha}}) = \mathbf{0}$, must have a uniformly bounded variance and have a well-defined covariance matrix as $\hat{\boldsymbol{\alpha}}_k \to \boldsymbol{\alpha}$ a.s. This is reflected in condition (vii, a), $\sup_{\hat{\boldsymbol{\alpha}}} E\|\mathbf{Z}(\hat{\boldsymbol{\alpha}})\|^{2+2\epsilon} < \infty$ for some $\epsilon > 0$ which is satisfied with the Euclidean norm, $\sup_{\hat{\boldsymbol{\alpha}}}\{\mathrm{tr}[S_k^T E(\mathbf{W}\mathbf{W}^T)S_k]^{1+\epsilon} < \infty$ for each k, with $E(\mathbf{W}\mathbf{W}^T)$ nonnegative definite. Condition (vii, b) $\lim_{\hat{\boldsymbol{\alpha}}\to\boldsymbol{\alpha}} E[\mathbf{Z}(\hat{\boldsymbol{\alpha}})\mathbf{Z}^T(\hat{\boldsymbol{\alpha}})] = \pi$, where π is a nonnegative definite matrix and where the limit is in the sense of the norm, is satisfied with $\lim_{\hat{\boldsymbol{\alpha}}\to\boldsymbol{\alpha}}[S_k^T E(\mathbf{W}\mathbf{W}^T)S_k] = \pi$. S_k is diagonal for each k and $E(\mathbf{W}\mathbf{W}^T)$ is nonnegative definite and diagonal by the i.i.d. assumption of the problem.

The adaptive gain matrix $A_k(\hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_k)$ needs to be well behaved and a consistent "mean-square" estimator of a constant matrix. With $A_k(\cdot)$ diagonal by the assumption of the problem, the eigenvalues of A_k are $a_1^{(k)} \geqslant a_2^{(k)} \geqslant \cdots \geqslant a_m^{(k)} > 0$ and the eigenvalues of $B_k = S_k^T S_k$, $B_k = \mathrm{diag}(S_{11}^{2(k)}, \ldots, S_{m,m}^{2(k)})$, a

diagonal matrix, are $b_1^{(k)} \geqslant b_2^{(k)} \geqslant \cdots \geqslant b_m^{(k)} > 0$. With $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_m > 0$ the eigenvalues of A, condition (viii, a) is satisfied

$$0 < a_1{}^1 \leqslant \inf_{\hat{\alpha}} \left\{ \sum_{i=1}^{m} [a_i^{(k)}]^2 \right\}^{1/2} \leqslant \sup_{\hat{\alpha}} \left\{ \sum_{i=1}^{m} [a_i^{(k)}]^2 \right\}^{1/2} \leqslant a'' < \infty$$

wpl for k large; and $\underline{\lim}_k a_m^{(k)}(\cdot) \geqslant a_1{}^1 > 0$ wpl where A is a constant matrix s.t. $a^1 \leqslant \sum_{i=1}^{m} [a_i{}^2]^{1/2} \leqslant a''$. Condition (ix) $a_1{}^1 b_m^{(k)} - \epsilon > 1/2$ and $a_m n_m^{(k)} - \epsilon > 1/2$ for each k, which is required for convergence within the proof.

THEOREM 1. *Under assumptions* (i–ix) *which were shown to be satisfied above, let* $r_1^{(k)} > r_2^{(k)} > \cdots > r_m^{(k)} > 0$ *be eigenvalues of* $AB, B_k = S_k{}^T S_k$. *Then* $k^{1/2}(\hat{\alpha}_k - \alpha)$ *is asymptotically normal with mean zero and covariance matrix* Q, *where* Q *is a diagonal matrix whose elements are* $a_{ii}^2 \pi_{ii} \cdot [2a_{ii} S_{ii}^2 - 1]^{-1}$, *where* π_{ii} *are the elements of* $\pi = S^T E(\mathbf{W}\mathbf{W}^T) S$, *a diagonal matrix,* $i = 1, 2, \ldots, m(k)$. *For proof see* (*Kersten and Kurz*, 1976).

Comment. It is clear from the above theorem and from the form of Q, that the asymptotic variance is a function of the power in the signal samples (which are the eigenvalues of B) and the covariance of i.i.d. noise vector samples, with the dimensionality $m = m(k)$ of Q increasing with each iteration step. The optimum gain coefficient is given by $a_{ii}^{(k)} = 1/(S_{ii}^2)_k$ which minimizes the components of the variance and assures the rate of convergence to be optimal. One has to be careful, however, that $a_{ii}^{(k)} \neq 1/2(S_{ii}^2)_k$ since the variance becomes infinite and the recursion diverges. To avoid the instability with $a_{ii}^{(k)} \to 1/2(S_{ii}^2)_k$, one could use the average power in the block signal samples which would rapidly become independent of k and would still guarantee near optimum convergence.

However, even if one could find an estimator for $a_{ii} = 1/S_{ii}^2$ at every step by somehow measuring the energy in the signal samples in the absence of noise, the variance is a function of the covariance of the noise $E(\mathbf{W}_k \mathbf{W}_k{}^T)$, and the variance of the recursion, both asymptotically and in the small sample case, is influenced by the variations in $E(\mathbf{W}_k \mathbf{W}_k{}^T)$. One should recall at this point that the elements of S_k are $\mathbf{X}_k = \{[I]_{(m \times m)} \sum_{j=0}^{i} g_k(j) K(i-j)\} \mathbf{p}_0 + \mathbf{g}$ and the additive noise term is given by

$$\mathbf{W}_k = \left\{ [I]_{(m \times m)} \sum_{j=0}^{i} V_k(j) K(i-j) \right\} \mathbf{p}_0 + \mathbf{V}_k .$$

To alleviate this dependence, one needs to robustize the SAMVLS algorithm. However, it should be noted that even in the above unrobustized case, the SAMVLS algorithm reduces the mean-squared error $E \| \hat{\alpha}_k \|^2 = O(1/k)$ due to the noise in the data, while the iterative measurement equation (algorithm

within the SAMVLS algorithm) converges in mean-square as shown by Papoulis (1975). Thus, the reduction of the noise contribution at each iteration step eliminates the instability associated with the iterative measurement algorithm, as long as the rate of convergence of the SAMVLS compensates for the reduction of eigenvalues of the prolate spheroidal functions which reduce at a rate depending on the time-bandwidth product, $T\sigma$, with each iteration step.

Batch-Nonlinear-Integer Rank Transformation Robustized SAMVLS Image Restoration Algorithm

To robustize (10) one introduces batch preprocessing and a nonparametric rank statistic of the form (Kadar and Kurz, 1980b)

$$W_k^q(\hat{\alpha}_k, \alpha) = \frac{1}{q^2} \sum_{i=1}^{q} \sum_{j=1}^{q} \operatorname{sgn}\{S^T_{[i+q(k+1)]}\mathbf{g}_{[i+qk]} - [S^T S\alpha]_{[j+q(k-1)]}\}, \tag{13}$$

where $W_k^q(\cdot)$ is an $(m(k) \times 1)$ vector operator applied component by component. $W_k^q(\cdot)$ is a symmetric version of the Mann–Whitney–Wilcoxon Nonparametric Statistics (MWWNS) with properties summarized here for convenience

$$EW^q = 0,$$
$$\operatorname{Var} W^q(\cdot) = (2q + 1)/3q^2,$$
$$\sup_{F,G} W^q = -\inf_{F,G} W^q = 1,$$

and under the hypothesis, H, and alternative, K,

$$\lim_{N \to \infty} P\left[\frac{W^q - EW^q}{(\operatorname{Var} W^q)^{1/2}} \leqslant t\right] = \Phi(t) \qquad \text{(unit normal CDF)},$$

with asymptotic normality reached with as few as $q = 8$ samples. Furthermore, the above properties do not require symmetry of the CDF, either under H or K.

The robust vector SAMVLS in this case is of the form

$$\hat{\alpha}_{k+1} = \hat{\alpha}_k - (1/k)A_k(\hat{\alpha}_1, \ldots, \hat{\alpha}_k)\, W^q \mathbf{Y}(\hat{\alpha}_k, \alpha), \tag{14}$$

which has to satisfy the conditions of Theorem 1. It should be noted here that the batch preprocessing requires an initial delay of q samples during which the algorithm is iterated and the samples are stored. This means that during this period the SAMVLS operates essentially as an unrobustized algorithm and no protection is provided against measurement noise outliers, causing possible instabilities in the "inner" measurement algorithm. Proceeding in a manner similar the scalar case (Evans, Kersten, and Kurz, 1976), it can be shown that for each $m(k)$ component, say, $l = 1, 2, \ldots, m(k)$ with $\mathbf{M}(\hat{\alpha}, \alpha) = EW^q[\mathbf{Y}(\hat{\alpha}, \alpha)] =$

$B(\hat{\alpha} - \alpha)$ the diagonal elements of the B-matrix (the slope of the regression function) $B = \text{diag}(b_{k1},..., b_{k,m(k)})$ in this case are $b_k = 2f_{u_{kl}-v_{kl}}(0)$, $l = 1, 2,..., m(k)$, where u_k and v_k are the batched components of the first and second terms in $\mathbf{Y}(\hat{\alpha}, \alpha)$, (13), respectively, and $f_{u_{kl}-v_{kl}}(0)$ is a one-dimensional density corresponding to the lth component of $\mathbf{Y}(\hat{\alpha}, \alpha)$. It is clear from the previous definition of the terms above that $f_{u-v}(0)$ is a function of the block signal samples which are time-varying from block to block. This means that the optimum gain coefficient, $A_k = \text{diag}(a_{1k}, a_{2k},..., a_{m(k),k})$ with $a_{i_k} = 1/4[f_{u_{ki}-v_{ki}}(0)]$ is also time varying. By assuming that the block signal sample amplitudes can be approximated by an averaged signal level between adjacent blocks, $S_k = S = \text{diag}(\text{avg} \mid S_1 \mid,..., \text{avg} \mid S_{m(k)} \mid)$, $f_{u-v}(0)$ becomes a constant for each k. Now a simple approximation can be derived by representing $f_{u-v}(0)$, $i = 1, 2,..., m(k)$, as a generalized Gaussian noise pdf (Kendall and Stuart, 1963) for a wide class of both thin and heavy tailed pdf's $f_{u_i-v_i}(0)$, only varies in the range of 2 to 1. This means that one does not need a very precise estimator of a_{i_k} at each step, k, since the efficiency of the SAMVLS is not sensitive to changes in a_{i_k} (Evans *et al.*, 1976). However, the estimator of a_{i_k} should be robust if one desires high efficiency independent of the CDF of the measurement noise. Such an estimator is given in (Kersten and Kurz, 1976), which for each component of $A_k(\cdot)$

$$\hat{a}_{i_k} = \frac{q-1}{4(k-1)} \sum_{j=1}^{k-1} [Z_{j,[(q-2)+1]} - Z_{j,[q/2]}],$$

where $Z_{j,[q/2+1]} = q/2 +$ 1th-order statistic from the batched component u_{kl} of $\mathbf{Y}(\hat{\alpha}, \alpha)$ and $Z_{j,[q/2]} = q/2$th-order statistic from the batched component u_{kl}, $l = 1, 2,..., m(k)$, and $[\epsilon]$ defined to be the greatest integer less than or equal to ϵ. It has been shown in (Kersten and Kurz, 1977) that the above robust adaptive estimator of the optimum gain coefficient satisfies the conditions of Theorem 1.

The covariance matrix of the asymptotically normal robustized estimator $k^{1/2}(\hat{\alpha} - \alpha)$ with mean zero is given by $a_{ii}^2 \pi_{ii}[2a_{ii}b_{ii} - 1]^{-1}$, where $\pi_{ii} = \sigma_{w_{ii}}^2[I]S^TS$, $i = 1, 2,..., m(k)$, and $\sigma_{w_{ii}}^2 = (2q + 1)/3q^2$, which is independent of the measurement noise statistics (compare this result with the covariance of the unrobustized SAMVLS) and the a_{ii}^2 are given by the batched order statistic estimator, defined previously. In this case, the robustness is reached after a small number of iterations and the instability of the recursion is only reached as a limit point as $k \to \infty$.

Simulation Results

The SAMVLS method was applied to a signal $f(t) = \sin \sigma t/\pi t$ choosing for T the value $\pi/5\sigma$ in the presence of noise contamination described by the mixture distribution model $f(v) = 0.9n(0, 1) + 0.1n(0, 8)$. The above signal is the same

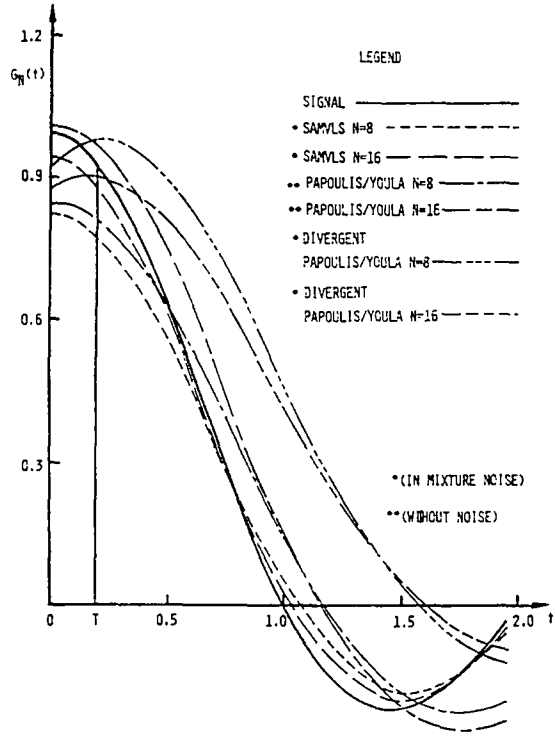

FIG. 3. Computer simulation comparison of reconstruction algorithms.

as the one used in (Papoulis, 1975) which is used to illustrate the performance of the "inner" algorithm (5), in the absence of noise. This allowed comparing and checking the results with the one in (Papoulis, 1975) for the noise-free case using the convolution approach. The result of the simulation of the noise-free "inner" algorithm by the convolution approach is shown in Fig. 3. The effect of noise contamination on the convolution implementation of the "inner" algorithm is shown in the same figure indicating divergence. The stable performance of a SAMVLS algorithm in the presence of noise contamination is also shown in Fig. 3.

AN ALTERNATIVE METHOD—THE ROBBINS–MONRO APPROACH

The foregoing recursive regression framework representation is not limited to the iterative algorithm of bandlimited extrapolation. It is also possible to apply the regression framework to the noniterative approach using prolate spheroidal wave functions (Viano, 1976) or, in general, to methods using any other complete set of basis functions.

More specifically, one can apply a vector Robbins–Monro Stochastic Approximation (RMSA) (Robbins and Monro, 1951; Kersten and Kurz, 1976) algorithm directly to estimate the coefficients of the orthogonal expansion methods (e.g., the prolate spheroidal expansion method introduced earlier) in a manner similar to the one used for estimation of density functions in (Kashyap and Blaydon, 1968). The method involves expanding the given function in a finite linear combination of orthogonal functions, i.e., $h(x) = \sum_{i=1}^{N} a_i \phi_i(x)$ and minimizes the integral-square error (ISE). ISE $= \int [h(x) - \sum_{i=1}^{N} a_i \phi_i(x)]^2 \, dx$ with respect to a_i, $i = 1, 2, \ldots, N$.

One can consider the problem as if $g(t)$ is unknown (or cannot be directly observed) and we estimate an approximation to $y(t)$ (4) using a sequence of independent samples t_k, $k = 1, 2, \ldots$ A block sample of $y(t)$ is approximated as

$$\mathbf{y}(t) = \sum_{k=1}^{n} a_k^* \psi_k(t) = \mathbf{a}^{*T} \boldsymbol{\psi}(t),$$

where $\psi_k(t) = \phi_k(t)/(\lambda_k)^{1/2}$, $\mathbf{a}^*$ is an n-vector of unknown coefficients, $\{\psi_k(t), k = 1, \ldots, n\}$ is a set defined by the eigenfunctions of the integral equation (Slepian *et al.*, 1961). The problem is to find $\mathbf{a}^*$ which minimizes the ISE. The integral-square-error criterion is given by

$$I(\mathbf{a}) = \int_{\theta} (\mathbf{y}(t) - \mathbf{a}^T \boldsymbol{\psi}(t))^2 \, dt,$$

where

$$dt = \prod_{k=1}^{m} dt_k, \qquad \theta = (-T, T).$$

The value of $\mathbf{a} = \mathbf{a}^*$ which minimizes this criterion is (Kashyap and Blaydon, 1968)

$$\mathbf{a}^* = \left[\int \boldsymbol{\psi}(t) \, \boldsymbol{\psi}^T(t) \, dt \right]^{-1} \int_{\theta} \boldsymbol{\psi}(t) \, \mathbf{y}(t) \, dt.$$

To evaluate $\mathbf{a}^*$ numerically, using the RMSA, one forms

$$\mathbf{a}_{k+1} = \mathbf{a}_k - (A/k) \, \mathbf{Y}(\mathbf{a}_k, t_k),$$

where A is a diagonal gain matrix, the function $\mathbf{Y}(\mathbf{a}, t)$ is chosen so that

$$E[\mathbf{Y}(\mathbf{a}, t) | \mathbf{a}] = \frac{\partial I}{\partial \mathbf{a}} = -\int_{\theta} \boldsymbol{\psi}(\eta) \, \mathbf{g}(\eta) \, d\eta + \left[\int_{\theta} \boldsymbol{\psi}(\eta) \, \boldsymbol{\psi}^T(\eta) \, d\eta \right] \mathbf{a},$$

$$\eta = \text{col}(\eta_1, \ldots, \eta_m), \qquad d\eta = \prod_{i=1}^{m} d\eta_i.$$

The function $\mathbf{Y}(\mathbf{a}, t)$ satisfying the above condition is given below

$$\mathbf{Y}(\mathbf{a}, t) = -\beta(t) + K\mathbf{a},$$

where

$$\begin{aligned} \beta(t) &= \psi(t) \quad \text{if} \quad t \in \theta \\ &= 0 \quad \text{if} \quad t \notin \theta \end{aligned}$$

and

$$K = \int_{\theta} \psi(t)\, \psi^T(t)\, dt,$$

which can be evaluated directly from knowledge of the eigenfunctions. The vector RMSA algorithm can be written as

$$\mathbf{a}_{k+1} = \mathbf{a}_k + (A/k)(\beta(t_k) - K\mathbf{a}_k).$$

The above algorithm has been shown to converge in the mean-square sense and wpl (Kashyap and Blaydon, 1968). The conditions for convergence are given in Theorem 1 and in (Kersten and Kurz, 1976).

The above theory of recursive estimation of the coefficients of orthogonal expansion methods via RMSA is completely general, and is not dependent on the choice of the basis functions. Hence, the method is applicable to prolate spheroidal wave functions or to any other convenient orthogonal set. But in the case of prolate spheroidal wave functions, unless one can describe $\psi(t)$ in functional form, the method would require storage or generation of eigenfunctions and would not possess the economy of robustized SAMVLS.

However, using Walsh basis functions (Young and Calvert, 1974) the eigenfunctions of the resultant expansions can be recursively generated at each iteration step of RMSA. Hence the procedure becomes completely recursive and can be robustized in a manner similar to the methods used in SAMVLS. Furthermore, the extension of this procedure to two or multidimensional signals via the Walsh–Hadamard relationship (Young and Calvert, 1974) promises to generate an interesting and straightforward approach to image reconstruction.

Conclusions

A new application of the vector extension of Gladyshev's theorem to Minimum Variance Least Squares (SAMVLS) was introduced to the image extrapolation or the object reconstruction problem. The new algorithm which uses the iterative vector measurement equation in parametric form, depicting a band-limited extrapolation process, which possesses inherent stabilizing properties which eliminate the instabilities associated with the improperly posed problem.

The new algorithm is robustized via the Batch-Nonlinear-Linear (B-N-L) approach (Kadar and Kurz, 1980b), a procedure which ensures "small sample" asymptotic robustness by a "prewhitening-like" operation using the Mann-Whitney-Wilcoxon Nonparametric Statistics (MWWNS) and adaptive gain coefficients. The resultant diagonal covariance matrix of the asymptotically normal robustized estimator of the parameters thus becomes independent of the measurement noise process and completely insensitive to noise outliers after a small number of iteration steps. The use of the adaptive gain coefficients guarantees near optimum convergence rates. Computer simulation verified the stable performance of the method in Gaussian mixture noise.

The direct application of the RMSA method to estimating the coefficients of orthogonal expansion methods was also considered. The resultant recursive method was shown to be applicable to any convenient orthogonal set of basis functions (e.g., Walsh functions) and promises to generate an interesting approach to multidimensional image reconstruction.

Appendix A

In a Hilbert space setting (Youla, 1977) the complete image, f, is represented as a vector known a priori to belong to a linear subspace θ_b but all that is available is its projection $P_a f$ onto a known linear subspace θ_a. By defining P_a, Q_a, P_b, and Q_b as projection operators projecting onto θ_a, $\perp\theta_a$, θ_b, and $\perp\theta_b$, respectively, (e.g., $f \in \theta_b \rightarrow f = P_b f$) a recursive method is evolved to determine f uniquely from $P_a f$ (and with stability in the face of noise) given by

$$f_{k+1} = g + Q_a P_b f_k, \qquad k = 1 \rightarrow \infty; f_1 = g, \tag{A1}$$

and it is shown to converge to f in norm, $\lim_{k\to\infty} \| f_k - f \| = 0$ (Youla, 1977). The geometric representation of the steps involved is illustrated in Fig. 2. The above algorithm becomes improperly posed and can become unstable in the presence of noise, if the angle $\psi(\theta_b, \perp\theta_a) = 0$, which is clear from Fig. 2, where $\psi(\theta_b, \perp\theta_a)$ is the angle at 0 of the $OA'A$ triangle. The instability, due to aliasing, truncation error and measurement noise, is handled as before by terminating the recursion after certain "optimal" number of steps, albeit the theory for the optimality is not established. Papoulis' (1975) algorithm is a special case of (A1).

Received: November 3, 1979; Revised: December 5, 1979

References

Andrews, H. C. (1970), "Computer Techniques in Image Processing," Academic Press, New York.

EVANS, J., KERSTEN, P., AND KURZ, L. (1976), Robustized recursive estimation with applications, *Inform. Sci.* **11**, No. 1, 69–92.

GERCHBERG, R. W. (1974), Super-resolution through error energy reduction, *Optica Acta* **21**, No. 9, 709–720.

HOWLE, W. (1977), Private communication.

KADAR, I., AND KURZ, L. (1980a), Robustized vector form of Gladyshev's theorem, to appear.

KADAR, I., AND KURZ, L. (1980b), A robust Kalman filter, to appear.

KASHYAP, K. L., AND BLAYDON, C. C. (1968), Estimation of probability density and distribution functions, *IEEE Trans. Information Theory* **IT-14**, No. 4, 549–556.

KENDALL, M. G., AND STUART, A. (1963), "The Advanced Theory of Statistics," Vol. 1, 80–83, Hafner, New York.

KERSTEN, P., AND KURZ, L. (1976), Robustized vector Robbins-Monro algorithm with applications to M-interval detection, *Inform. Sci.* **11**, No. 2, 121–140.

KERSTEN, P., AND KURZ, L. (1977), Bivariate M-interval classifiers with applications to edge detection, *Inform. Contr.* **34**, 152–168.

MANN, H. B., AND WHITNEY, D. R. (1946), On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Statist.* **18**, 50–60.

PAPOULIS, A. (1965), "Probability, Random Variables and Stochastic Processes," McGraw–Hill, New York.

PAPOULIS, A. (1968), "Systems and Transforms with Application in Optics," McGraw–Hill, New York.

PAPOULIS, A. (1975), A new algorithm in spectral analysis and bandlimited extrapolation, *IEEE Trans. Circuits and Systems*, **CAS-22**, No. 9, 735–742.

ROBBINS, H., AND MONRO, S. (1951), A stochastic approximation method, *Ann.Math. Statist.* **22**, 400–407.

SLEPIAN, D., POLLACK, H. O., AND LANDAU, H. J. (1961), Prolate spheroidal wave functions, *Bell Syst. Tech. J.* **40**, 43–84.

VIANO, G. A. (1976), On the extrapolation of optical image data, *J. Math. Phys.* **17**, No. 7, 1160–1165.

YOULA, D. C. (1976), Generalized image restoration by the method of alternating orthogonal projections, *in* "Proceedings, Image Science Mathematics Symposium, November 10–12, Monterey, California," pp. 176–185, Western Periodicals, North Hollywood, Calif.

YOUNG, T. Y., AND CALVERT, T. W. (1974), "Classification, Estimation and Pattern Recognition," American Elsevier, New York.