# TECHNIQUES, BASED ON EXTREMAL SUBSPACES, FOR IMPROVED RECONSTRUCTION OF SIGNALS FROM SAMPLES*

R.G.Meier and T.W.Parks
E.E.Department, Rice University
Houston, Texas 77001

ABSTRACT

Suppose a signal f drawn from a known class of signals is subject to a finite set of linear functionals, or generalized "samples". This paper studies the problem of estimation of the sampled signal through linear reconstructions based on these measurements. The extent to which a sampling scheme can determine members of a given signal class is measured by the worst-case $\mathcal{L}_2$ error in reconstructing the signals from their samples. This of course depends on the reconstruction technique used. Desire to make this error approach the minimum leads us to seek efficient reconstruction algorithms.

This study makes use of the theory of n-widths and extremal subspaces for function classes originated by A.N.Kolmogoroff. The optimal reconstruction need not, in general, lie in the sampling space. Of itself, projection onto a subspace $\Phi_n$ loses all information about the signal orthogonal to $\Phi_n$; and a best estimate of the original signal is its projection, which of course is restricted to the sampling subspace. A knowledge of the signal-class of which f is a member supplies some information lost by the projection operation. Two reconstruction techniques, based on extremal subspaces, are developed. Error bounds are presented and compared with reconstruction in the sampling subspace.

Finally, to provide a concrete example of this general theory, the results are applied to a much-studied class, the class of time-concentrated, bandlimited signals. The measurement process is here assumed to be the "convenient" one of Nyquist rate time-sampling. For this problem, plots of the error bounds and of several test functions and their reconstruction are presented, both for the proposed algorithms, and for conventional "cardinal sampling theorem" reconstruction.

## I. INTRODUCTION

We consider the estimation of a signal of a known class through linear reconstructions from a finite set of linear measurements. Although Nyquist rate or faster time sampling of bandlimited functions is by far the most common measurement process, we shall generalize our model somewhat, including this as a special case. We will consider as a measurement (or generalized "sample") $b_j$ any bounded linear functional on the signal. By the Reisz Representation theorem such a sample may be considered an inner product of the signal with an appropriate function of $\varphi_j$ of $\mathcal{L}_2$, which we will call a sampling function. Measurement is thus projection onto a subspace, called the sampling subspace, which specifies completely the measurement process. We may therefore consider the samples to be the coordinates of the signal in the sampling subspace,

*This work was supported in part by NASA Grant NGL 44-006-033.

and may without loss of generality restrict the sampling functions to be orthonormal.

Our objective is to use a given finite set of samples to form an estimate of the measured signal, with the greatest possible accuracy. We mean by "greatest possible accuracy" the minimization of the worst possible error of the estimator, with the error measured by an $\mathcal{L}_2$ norm. However, if all that is known about the signal is the sample vector, the estimation error is unbounded for any scheme. For such a problem to make sense it is therefore necessary that the measured signal be known a priori to be a member of a bounded subset of the whole space, some signal class, C. If $\underline{b}$ is the data vector and $\hat{f}(b)$ the estimate of the given signal f in C, the estimator is evaluated by the error $\sup_{f\in C}\|f-\hat{f}(\underline{b})\|$.

We shall restrict $\hat{f}$ to be a linear reconstruction; that is $\hat{f}$ lies in a linear subspace, and the coordinates of $\hat{f}$ in this subspace depend linearly on the data $\underline{b}$. Note however that this reconstruction subspace need not be the same as the sampling subspace. Knowledge of the class of signals to be estimated allows the choice of a reconstruction subspace well suited for representation of the known class. This makes possible a smaller error than the best representation confined to the sampling subspace. This paper presents a study of two schemes for such linear, suboptimal estimators.

## II. EXTREMAL SUBSPACES AND n-WIDTHS OF CLASSES

Before persuing the estimation proglem further, we review some ideas of approximation theory originated by A.N.Kolmogoroff. A good reference is Lorentz's book,* which also contains a good bibliography.

Consider the representation of a signal f by a given orthonormal system of m functions $\{\psi_i\}_{i=1}^{m}$. If the approximation $\hat{f} = \sum_{i=1}^{m} a_i\psi_i$, it is well known that the error $\|f-\hat{f}\|_2$ is minimized by $a_i=(f,\psi_i)$. Assuming that all signals f in a bounded class C are represented in this fashion, we define the deviation of the class C from the subspace $\Psi_m$ spanned by the m $\psi_i$:

$$\delta(C,\Psi_m) \triangleq \sup_{f\in C} \inf_{\hat{f}\in\Psi_m} \|f-\hat{f}\| \quad (1)$$

The values $\delta(C,\Psi_m)$ represents the degree of success with which the class C may be represented in $\Psi_m$. If we consider finding the subspace $\tilde{\Psi}_m$ best suited to represent the class C, in the sense of minimizing $\delta(C,\Psi_m)$, we have the idea of the extremal subspace for C, the resulting minimal deviation is called the m-width of C.

$$d_m(C) \triangleq \inf_{\Psi_m} \delta(C,\Psi_m) = \delta(C,\tilde{\Psi}_m) \quad (2)$$

If C is compact, $d_n \to 0$. An important property of the class is the manner in which $d_n$ decreases with increasing m. If most of the decrease occurs around a certain value n, the class might be said to have an essential dimension of n, even though it is technically infinite

*G.G.Lorentz, Approximation of Functions, New York: Holt, Rinehart and Winston, 1966, ch. 9.

dimensional.

The above properties have been studied for numerous classes, however little application of these ideas has been made to practical signal theory problems. One reason for this is the difficulty of implementation of the inner products $(f,\tilde{\psi}_i)$ with extremal basis functions. We now consider the possibility of using a given vector of samples $b_j = (f,\varphi_j)$ $j = 1, \ldots, n$ to estimate the extremal basis coordinates $a_i = (f,\tilde{\psi}_i)$. If this coordinate estimation error can be made sufficiently small, the accuracy characteristic of extremal basis representation of the signal can be achieved using only the given samples. We would thus have a better estimate of the measured signal than a reconstruction in the sampling subspace.

## III. COORDINATE ESTIMATION

To facilitate discussion of the coordinate estimation schemes we propose, we introduce the following notation: $C$ is a class of functions, $\{\tilde{\psi}_i\}_{i=1}^{\infty}$ and $\{\varphi_j\}_{j=1}^{\infty}$ are orthonormal sets of functions which span $\tilde{\Psi}_\infty = \Phi_\infty \supset C$. That is for all $f \in C$, we may expand

$$f = \sum_{j=1}^{\infty} (f,\varphi_j)\varphi_j = \sum_{i=1}^{\infty} (f,\tilde{\psi}_i)\tilde{\psi}_i \tag{3}$$

Denote $M$ the matrix whose elements are $(\tilde{\psi}_i,\varphi_j)$ $i, j=1, \ldots,\infty$. Partition $M$ and $M^{-1} = M^T$ as follows,

$$M = \left[\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right] \tag{4}$$

$$M^{-1} = \left[\begin{array}{c|c} P^T & R^T \\ \hline Q^T & S^T \end{array}\right] \tag{5}$$

where $P$ is $n\times n$, $Q$ and $R$ are semi-infinite and $S$ is infinite. $M$ transforms the coordinates of a vector $f$ in the $\varphi$-basis into the coordinates in the $\tilde{\psi}$-basis. The above partition of $M$ induces a partition of the coordinate vectors $\underline{a}$ and $\underline{b}$, where $a_i = (f,\tilde{\psi}_i)$ and $b_j = (f,\varphi_j)$. Thus $\underline{a} = M\underline{b}$ is partitioned

$$\left[\begin{array}{c} \underline{a}_1 \\ \hline \underline{a}_2 \end{array}\right] = \left[\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right] \left[\begin{array}{c} \underline{b}_1 \\ \hline \underline{b}_2 \end{array}\right] \tag{6}$$

and

$$\left[\begin{array}{c} \underline{b}_1 \\ \hline \underline{b}_2 \end{array}\right] = \left[\begin{array}{c|c} P^T & R^T \\ \hline Q^T & S^T \end{array}\right] \left[\begin{array}{c} \underline{a}_1 \\ \hline \underline{a}_2 \end{array}\right] \tag{7}$$

That is $\underline{a}_1$ and $\underline{b}_1$ are n-vectors, $\underline{b}_2$ and $a_2$ are infinite. The following interpretation will be given these vectors:

$\underline{b}_1$ is a given vector of samples, that is $\Phi_n$ is the sampling subspace

$\underline{a}_1$ is a vector of coordinates in the reconstruction subspace $\tilde{\Psi}_n$. (We shall also consider $\tilde{\Psi}_m \subset \tilde{\Psi}_n$ for $m < n$ as reconstruction subspaces, by using only the first $m$ components of $\underline{a}_1$, $(a_1 \ldots a_m)$).

The problem with which we shall be concerned is to estimate $\underline{a}_1$ given $\underline{b}_1$, and to bound the error which results when the approximate coordinates $\hat{\underline{a}}_1$ are used to reconstruct $f$.

Perhaps the most obvious thing to do is use the truncated expansion

$\hat{\underline{a}}_1 = P\ \underline{b}_1$, with an error $Q\ \underline{b}_2$. We shall refer to this as Method 1.

With reference to the matrix formulation, Method 1 amounted to assuming the remaining samples $\underline{b}_2$ were zero, producing an error $Q\underline{b}_2$. If we assume instead that the original signal lay exactly in the reconstruction subspace, that is that $\underline{a}_2$ is zero, solving the resulting equation gives the estimate $\hat{\underline{a}}_1 = (P^T)^{-1}\ \underline{b}_1$, and an error, $(P^T)^{-1}\ R^T\underline{a}_2$ (assuming P is nonsingular). Since the $\underline{a}_2$ coordinates are small for all $f \epsilon C$, and the same does not hold for $\underline{b}_2$, it would seem that this method, Method 2, is always superior to Method 1. This is not true because of a greater sensitivity of Method 2 to the relative alignment of the sampling and reconstruction subspaces.

Considerable insight into the behavior of these two schemes and the behavior of their error bounds may be gained by considering a simple example in a two dimensional space. Figure 1 shows a two dimensional, ellipsoidal class C, the one-dimensional sampling subspace $\Phi_1$ and extremal reconstruction subspace $\widetilde{\Psi}_1$. The 1-width $d_1(C) = \delta(C,\widetilde{\Psi}_1)$ and the deviation $\delta(C,\Phi_1)$ of the class from the sampling subspace are also indicated.

We assume that the only data regarding $f \epsilon C$ is the sample $(f,\varphi_1)=b_1$ given by the distance $D(0,b)$. We wish to estimate $a_1 = (f,\widetilde{\psi}_1) = D(0,c)$. The matrix P is simply $\cos\theta$. Method 1 uses $\hat{a}_1 = Pb_1 = b_1 \cos\theta$; the Method 1 approximation to point c is point d. Method 2 uses $\hat{a}_1 = (P^T)^{-1} b_1$ or $b_1/\cos\theta$, which gives point e. That is, Method 1 merely projects onto $\widetilde{\Psi}_1$ the projection of f on $\Phi_1$. Method 2 finds the member of $\widetilde{\Psi}_1$ whose projection on $\Phi_1$ is the same as $b_1 = (f,\varphi_1)$, on the assumption that the actual f is known to be close to $\widetilde{\Psi}_1$, by virtue of its class membership. If we call the distance $D(c,d) \triangleq \epsilon_1$, $D(c,e) \triangleq \epsilon_2$, $D(a,b) \triangleq h$, and $D(a,c) \triangleq s$, we find by direct application of a little trigonometry, the coordinate errors

$$\epsilon_1 = h \sin\theta \leq \delta(C,\Phi_1) \sin\theta \tag{8}$$

$$\epsilon_2 = s \tan\theta \leq d_1(C) \tan\theta \tag{9}$$

Noting that the coordinate and truncation errors are orthogonal we have the overall estimation errors for both methods:

$$\text{Method 1:} \quad \sup_{f \epsilon C} \|f-\hat{f}_1\|^2 \leq d_m^2(C) + \delta^2(C,\Phi_1) \sin^2\theta \tag{10}$$

$$\text{Method 2:} \quad \sup_{f \epsilon C} \|f-\hat{f}_2\| \leq d_n^2(C) + d_n^2(C) \tan^2\theta \tag{11}$$

We have found bounds of this type for the general problem. The generalization of the term $\sin^2\theta$ is $\min\{1, \sum_{i=1}^{m} [1 - \sum_{j=1}^{n} (\varphi_j,\psi_i)^2]\}$, and that of $\tan^2\theta$ is $\lambda(m,n) \min\{1, \sum_{j=1}^{n} [1 - \sum_{i=1}^{n} (\varphi_j,\psi_i)^2]\}$, where $\lambda(m,n)$ is the norm of the coordinate transformation matrix. These terms reduce to $\sin^2\theta$ and $\tan^2\theta$ for our two dimensional case.

## IV. APPLICATION

An example problem to which we have applied this theory is diagramed in Fig. 2. Given, finite energy signals are bandlimited, forming the class P. Members of P are then time-sampled at the Nyquist rate for n samples. The interpolator is to perform an accurate linear reconstruction of the sampled signal.

Error bounds for Methods 1 and 2 were evaluated for various numbers n of time samples and m of coordinates estimated. The results are presented in Figures 3 and 4, which plot the value of the error bound versus m, with n as a parameter. Of course for both sets of curves, for a given m, the bound is strictly decreasing with increasing n. This decline is asymptotically limited by the m-width of the class, however, since as the coordinate estimation becomes perfect, the only error is due to the limited number of coordinates estimated. The curves indicate that when such saturation occurs, estimation of a larger number of coordinates is warranted by the number of time samples available.

Using the data of Figures 3 and 4, both methods were optimized with respect to m for each n by picking the value yielding the minimum point on the curves. With the understanding then that the methods estimate the optimal number of coordinates for the number n of samples given, error bounds were plotted versus n for both methods (see Fig. 5). It is clear that for the problem being considered, the situation is as suggested by the geometric interpretation shown in Fig. 1. That is, the sampling and reconstruction subspaces are sufficiently close that Method 2 significantly outperforms Method 1. In fact the Method 2 bound behaves much like the m-width of the class, while the Method 1 bound looks like the deviation from the sampling subspace just shifted by a scale factor less than 1.

The final phase of this application consisted of actually generating some typical members of P, and reconstructing them from their time samples by 1) the cardinal sampling theorem, 2) Method 1, 3) Method 2.

Figures 6 and 7 show two such functions and their reconstructions derived from five time samples. Reconstruction error was computed for Methods 1 and 2 from a knowledge of the PSWF coordinates, exact and estimated. The error for the sampling theorem reconstruction could not be calculated other than by numerical integration over $(-\infty,\infty)$, and this was not done. The errors for reconstructions from various numbers of samples are summarized below. From inspection of the corresponding plots, it was considered obvious that the sampling theorem reconstruction of these functions never performed any better than Method 1. (Just as this is clearly the case in Figs. 6 and 7.)

| | # of Samples | Method 1 Error | Method 2 Error |
|---|---|---|---|
| Function A | 3 | $0.191 \times 10^{-0}$ | $0.127 \times 10^{-0}$ |
| | 4 | $0.104 \times 10^{-1}$ | $0.457 \times 10^{-1}$ |
| | 5 | $0.952 \times 10^{-2}$ | $0.150 \times 10^{-2}$ |
| | 6 | $0.575 \times 10^{-2}$ | $0.175 \times 10^{-3}$ |
| | 7 | $0.460 \times 10^{-2}$ | $0.283 \times 10^{-4}$ |
| Function B | 5 | $0.453 \times 10^{-1}$ | $0.236 \times 10^{-3}$ |

These values are consistent with the error bounds of Figure 5. As we expected, Method 2 seems distinctly superior to Method 1 in this case.

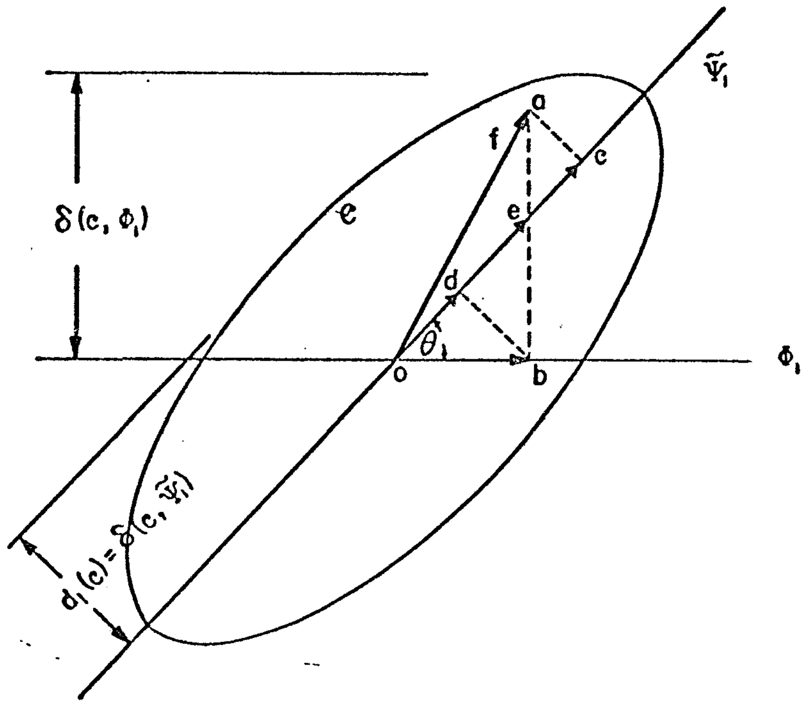


Figure 1: Geometric interpretation of Methods 1 and 2. A signal f of the class C is projected onto $\Phi_1$ by the sampling operation. The Method 1 reconstruction is point d; Method 2 gives point e. The best representation of f in $\tilde{\Psi}_1$ is point c.

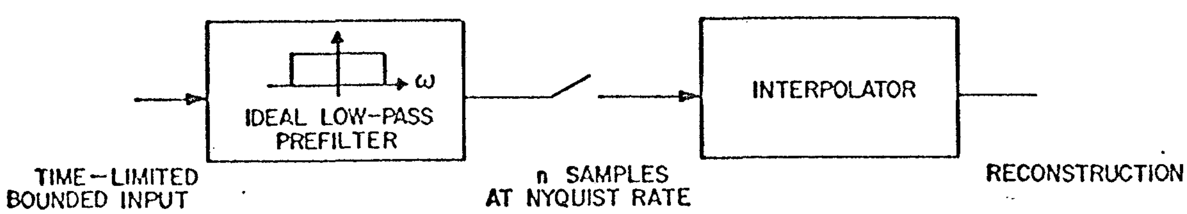


Figure 2: Block diagram for example problem. Given, finite-energy, time-limited input signals are bandlimited, forming the class P. Members of P are then sampled at the Nyquist rate. Using these n samples, the interpolator can be designed to do better than the cardinal reconstruction using sinc functions.

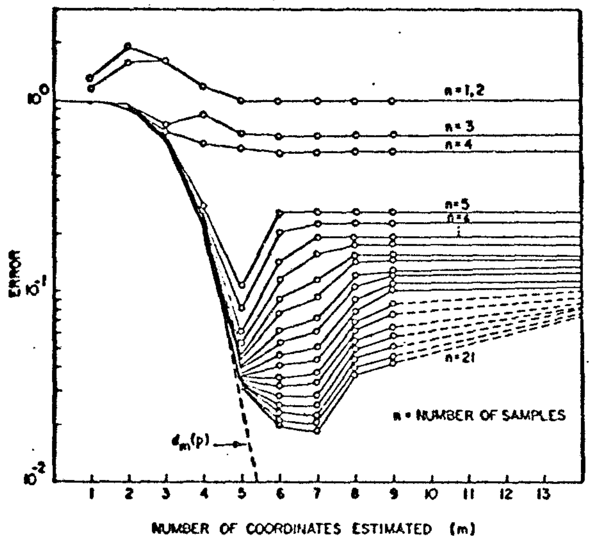


Figure 3

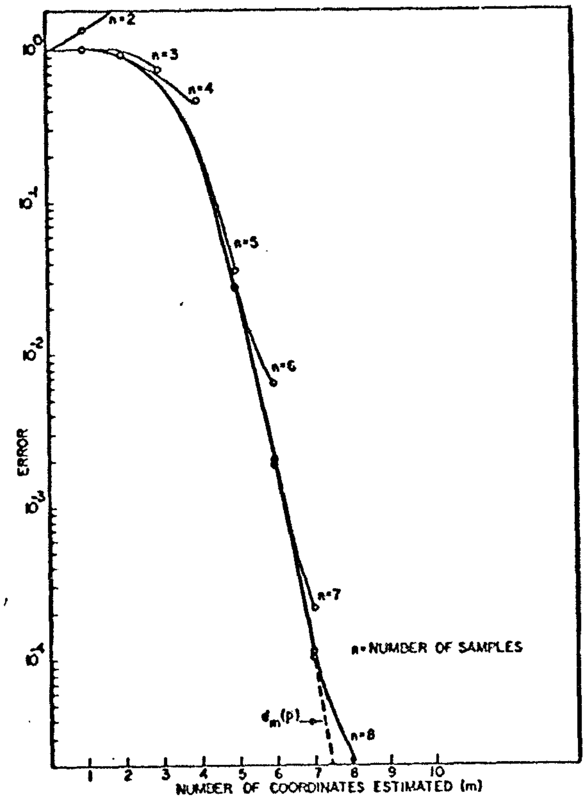


Figure 4

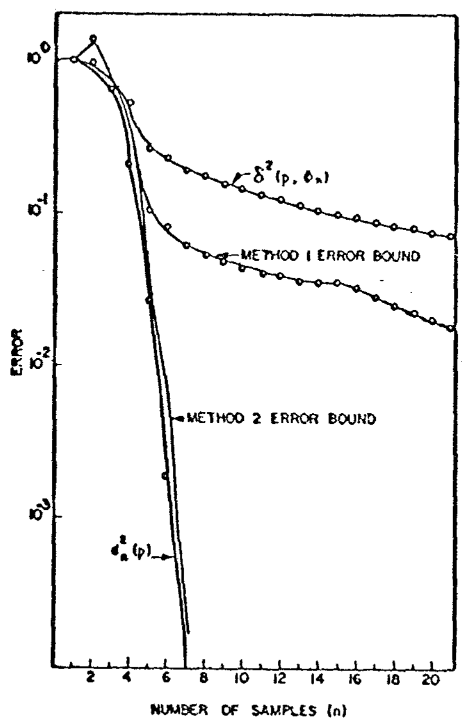


Figure 5

Figure 3: Error bounds for Method 1 for the class P.

Figure 4: Error bounds for Method 2 for the class P.

Figure 5: Error bounds for Methods 1 and 2 assuming the optimal number of coordinates are estimated for the given number of samples available. The deviation of the class P from the sampling subspace, and from the extremal (PSWF) subspaces is also shown.

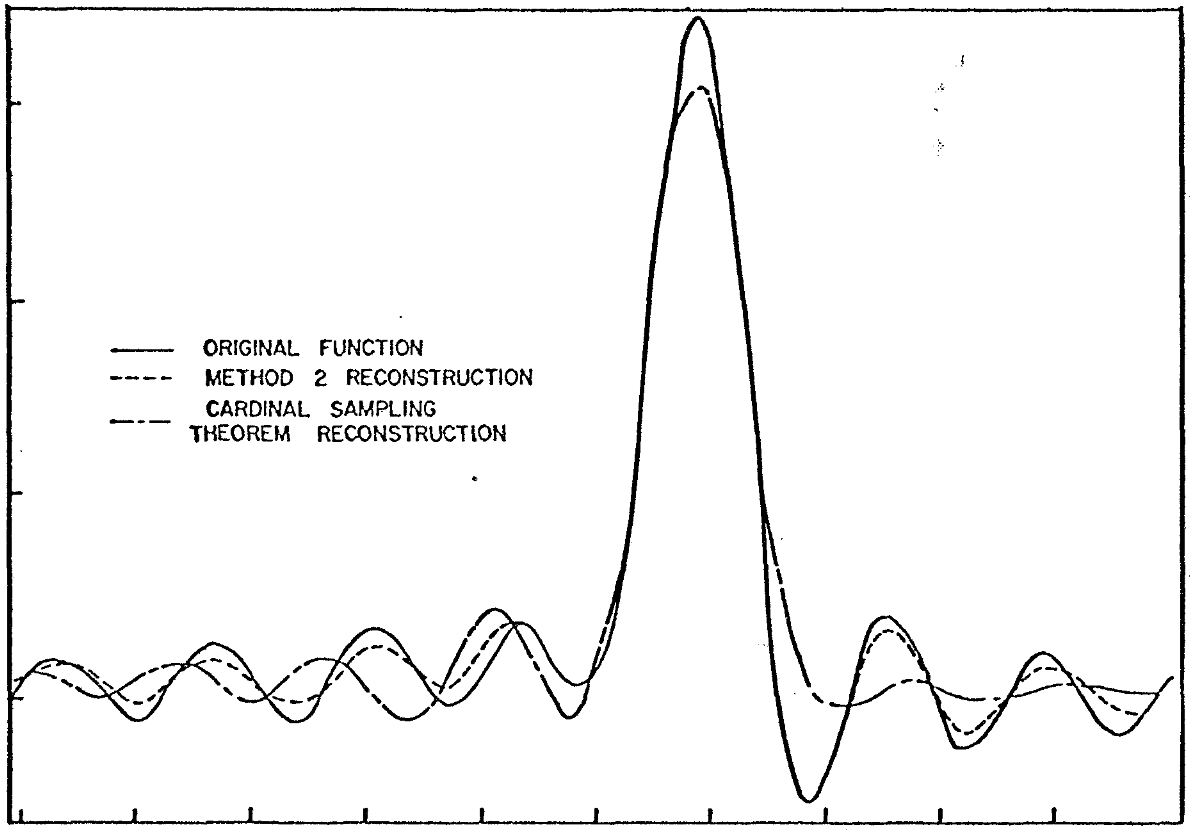


Figure 6: A member of class P (function A) and its reconstruction by Method 2 and by the cardinal sampling theorem. The Method 1 estimate is not shown, for this case it is quite close to Method 2.

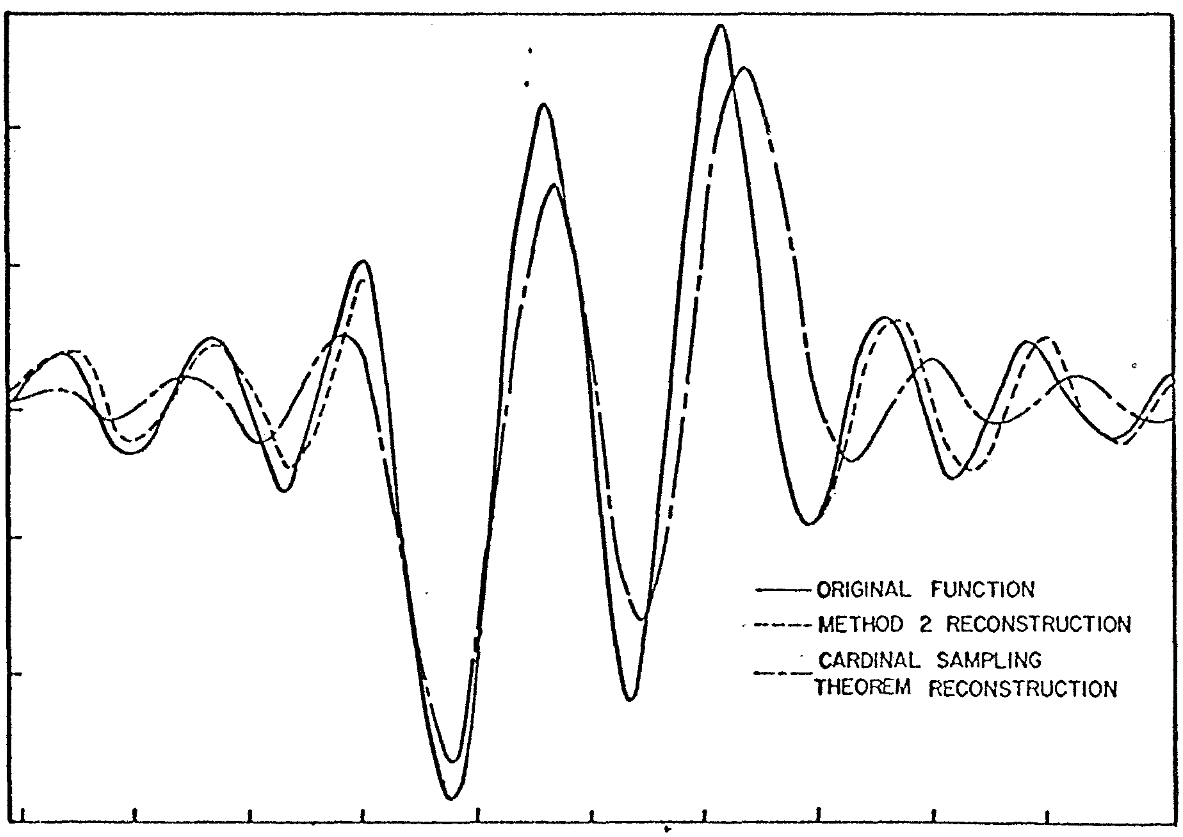


Figure 7: Another member of class P (function B) and reconstructions.