SIAM J. APPL. MATH.
Vol. 43, No. 2, April 1983

0036-1399/83/4302-0005 $01.25/0

SAMPLING FOR SPLINE RECONSTRUCTION*

ROBERT HUMMEL†

Abstract. The use of a prefilter prior to sampling or resampling can lead to reduced mean square error after reconstruction. This means that the samples should be obtained from the original data by convolution with an optimal local weighting function, whose form depends upon the reconstruction method to be used. We display and discuss these weighting functions for the most common reconstruction methods, namely nearest neighbor, linear and cubic spline interpolation.

1. Introduction. Suppose that a signal $f(x)$ is to be sampled at equispaced points $\{x_i\}_{i=-\infty}^{\infty}$ to obtain a sequence of samples $\{y_i\}_{i=-\infty}^{\infty}$. Eventually, the samples will be used to reconstruct an approximation $\tilde{f}(x)$ to $f(x)$. Both the sampling and reconstruction components of this process may be varied to minimize the error between $f(x)$ and $\tilde{f}(x)$. Typical sampling methods include block averaging:

$$y_i = \int_{x_i-h/2}^{x_i+h/2} f(x)\,dx, \qquad h = x_k - x_{k-1}, \tag{1}$$

or delta function sampling:

$$y_i = \int f(x)\delta(x_i - x)\,dx = f(x_i). \tag{2}$$

Typical reconstruction methods include nearest neighbor, linear or cubic spline interpolation or $(\sin x/x)$ reconstruction, defined by

$$\tilde{f}(x) = \sum_{i=-\infty}^{\infty} y_i \frac{\sin(\pi(x-x_i)/h)}{\pi(x-x_i)/h}. \tag{3}$$

Throughout, h denotes the constant sampling rate, that is, the distance between sampling points.

Each of the reconstruction methods is equivalent to an interpolation scheme, since the reconstructed function satisfies $\tilde{f}(x_i) = y_i$. Sampling, on the other hand, is generally a linear process of the input signal, i.e.,

$$y_i = \int_{-\infty}^{\infty} k_i(x) f(x)\,dx, \tag{4}$$

where $k_i(x) = k(x - x_i)$ is the sampling kernel centered at x_i. Only when $k(x)$ is the delta function does one necessarily have $y_i = f(x_i)$, so that, in general, the reconstructed function $\tilde{f}(x)$ does not even agree with $f(x)$ at the knots $\{x_i\}$.

The Shannon sampling theorem [1] states that if it is known a priori that $f(x)$ is band limited (the Fourier transform of f has bounded support), and if the sample spacing h is sufficiently small, then $\tilde{f}(x) = f(x)$ for all x, provided delta function sampling, (2), and $(\sin x/x)$ reconstruction, (3), are used. If the original signal is not bandlimited, then exact reconstruction is not possible. The theory of optimal prefilters can be used to show that if $(\sin x/x)$ reconstruction is used, then $\tilde{f}$ will be closest to f, in a certain sense, if f is sampled by a low-pass filter [2]. This method of sampling

* Received by the editors March 13, 1981, and in final revised form March 9, 1982.
† Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

is equivalent to weighted sampling, (4), with

$$k(x) = \frac{1}{h} \frac{\sin(\pi x/h)}{\pi x/h}. \tag{5}$$

We will provide a simple independent proof of this result later in this paper.

In this paper, we treat the following question. Suppose the reconstruction method is fixed and specified. Then how should the samples be obtained in order to minimize the $L^2(-\infty, \infty)$ error between f and $\tilde{f}$?

For $(\sin x/x)$ reconstruction, the solution is given by (5). However, in precision processing of digitized images, for example, $(\sin x/x)$ reconstruction is often considered too costly when interpolated values are needed for geometric correction and display magnification [3]. The usual reconstruction schemes include nearest neighbor, linear and cubic spline interpolation, which may be viewed as spline interpolation of degree zero, one and three, respectively [4]. The popularity of spline interpolation methods is presumably related to their ease in implementation and the strongly local dependence of the interpolate values on neighboring samples.

Since it is customary to substitute spline interpolation for $(\sin x/x)$ reconstruction, it is appropriate to replace low-pass filtering and delta function sampling with better sampling methods. Despite the fact that the derivation of these methods is trivial (§ 3), the corresponding sampling kernels (see Fig. 2) are relatively unknown in the fields of signal processing and data compression. The optimal sampling problem is equivalent to finding the spline which is closest to the given function in the L^2 norm. (Note that the values of the spline at the knots are unspecified.) The sampling kernels then represent the local dependence of the solution spline on the data.

The techniques for solving this problem are entirely known. The solutions in certain specific cases have even been obtained. For example, Powell [5] has studied the local dependence of the best L^2 approximating *cubic* spline. He obtains the L^2 optimal kernel for cubic spline reconstruction (Fig. 2c in this paper), and displays this kernel along with other kernels obtained by minimizing norm measures other than mean square error. Specifically, he considers the problem of finding a cubic spline $\tilde{f}(x)$ minimizing

$$\int_{-\infty}^{\infty} |f(x) - \tilde{f}(x)|^2 \, dx + c \sum_{i=-\infty}^{\infty} |f(x_i + \theta) - \tilde{f}(x_i + \theta)|^2,$$

where $0 < \theta < h$. Interestingly, some of these alternate norms lead to more localized kernels and are therefore more advantageous for implementation.

Accordingly, the computations of this paper are original only insofar as we emphasize the sampling and reconstruction interpretation of the results. However, we bring together the optimal kernels for linear, cubic spline and $(\sin x/x)$ reconstruction, and show how other reconstruction schemes can be treated similarly. We restrict our attention to distributed L^2 norms, although the results of Powell can be extended to other reconstruction methods treated here. The problem of minimizing the L^P norm could also be treated, but generally leads to nonlinear sampling methods.

A related problem in sampling and reconstruction, which is not treated elsewhere, occurs when previously sampled data must be resampled to achieve data compression. In this case, single samples are taken to represent blocks of values in the original data. The samples can be obtained as a weighted average of the data values in the block and can also involve values in neighboring blocks. Most frequently, linear interpolation is used to recover the data values within each block. Once again, the weights used to obtain the samples should be chosen so as to minimize the sum of square difference between the original data and the reconstructed data. We will treat the case when a sequence of data $\{v_i\}_{i=-\infty}^{\infty}$ is to be resampled to achieve an n-fold compression factor, so that

$$y_i = \sum_{j=-\infty}^{\infty} k_j \cdot v_{ni+j} \tag{6}$$

represents the values in block number i. The weights $\{k_j\}$ (see Fig. 3) play the same role as the kernel $k(x)$ in the continuous signal case.

Viewed as a method for data compression, weighted resampling as expressed by (6) can be compared to more standard encoding schemes, such as Fourier transform or other transform encoding [6], [7] and run length or difference encoding [8]. Higher compression ratios may be possible, but resampling as studied in § 4 is limited to linear encoding methods for which the reconstruction method is given by an interpolation formula. Interpolation methods other than spline reconstruction include Hermite, Lagrange and Chebyshev approximation. Least squares approximation is a familiar and applicable concept for many such interpolation schemes (see, for example, [9]), but is generally restricted to finite sets of sampling data, where the degree of the approximating polynomial increases as the amount of data increases.

Sampling and digitization is increasingly important as high speed digital signal processing technology enters the communications and robotics fields. Pulse code modulation of acoustical signals can be used for transmission and high fidelity recording. In image processing, intensity data sampled on a lattice of picture elements (pixels) can be used for image enhancement, industrial inspection tasks or for subsequent transmission, as from a satellite. In these applications, various sampling kernels can be obtained by suitably manipulating the point spread function of the sensor performing the sampling. If needed, electronic integration methods can be employed.

2. Spline reconstruction. We recall some facts about spline reconstruction of infinitely many samples at equispaced knots. We assume that $x_i - x_{i-1} = h$ for all i and that the set of data $\{y_i\}$ is bounded. For zero order splines (nearest neighbor interpolation) we have

$$s(x) = \sum_{i=-\infty}^{\infty} y_i B_0(x - x_i), \tag{7}$$

where

$$B_0(x) = \begin{cases} 1, & |x| \leq \dfrac{h}{2}, \\ 0, & |x| > \dfrac{h}{2}. \end{cases} \tag{8}$$

First order splines (linear interpolation functions) can be constructed using "chapeau functions"

$$B_1(x) = \begin{cases} 1 - |x|/h, & |x| \leq h, \\ 0, & |x| > h, \end{cases} \tag{9}$$

by the formula

$$s(x) = \sum_{i=-\infty}^{\infty} y_i B_1(x - x_i). \tag{10}$$

Although there is no unique cubic spline interpolating the data $\{(x_i, y_i)\}$, there is always exactly one such *bounded* cubic spline. It can be constructed as follows. Let

(11) $$B_3(x)=\int_{-\infty}^{\infty} B_1(x')B_1(x-x')\,dx', \quad \text{i.e.,} \quad B_3(x)=B_1^{*}B_1(x).$$

The function $B_3(x)$ is the cubic B-spline [4] and satisfies

(12) $$B_3(x)=0 \quad \text{for} \quad |x|>2h, \quad B_3(-h)=B_3(h)=\tfrac{1}{6}, \quad B_3(0)=\tfrac{2}{3}.$$

Thus, for any bounded sequence $\{v_i\}_{i=-\infty}^{\infty}$, the function

(13) $$s(x)=\sum_{i=-\infty}^{\infty} v_i B_3(x-x_i)$$

is a bounded cubic spline interpolating the data $\{(x_i, \frac{1}{6}v_{i-1}+\frac{2}{3}v_i+\frac{1}{6}v_{i+1})\}$. In order to interpolate the data $\{(x_i, y_i)\}_{i=-\infty}^{\infty}$, it suffices to solve the infinite linear system

(14) $$\tfrac{1}{6}v_{i-1}+\tfrac{2}{3}v_i+\tfrac{1}{6}v_{i+1}=y_i \quad \text{for all } i.$$

This system can be solved, as we show in § 5, yielding the linear relation

(15) $$v_i=\sum_{j=-\infty}^{\infty} c_j y_{i+j} \quad \text{for all } i$$

where $\{c_j\}$ is a square summable sequence of fixed constants. Combining (13) and (15), we have

(16) $$s(x)=\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} y_{i+j}c_jB_3(x-x_i)$$

is the unique bounded cubic spline interpolating $\{(x_i, y_i)\}$. This can be written as

(17) $$s(x)=\sum_{k=-\infty}^{\infty} y_k\psi(x-x_k),$$

where

$$\psi(x-x_0)=\sum_{i=-\infty}^{\infty} c_{-i}B_3(x-x_i).$$

Equation (17) is the cubic spline version of (7) and (10). The function $\psi(x-x_0)$ is itself a bounded cubic spline interpolating the data $\{\cdots, (x_{-2}, 0), (x_{-1}, 0), (x_0, 1), (x_1, 0), (x_2, 0), \cdots\}$ and is the cubic spline version of the zero order spline $B_0(x)$ or the linear spline $B_1(x)$ (see Fig. 1). These are the cardinal spline functions of order zero, one and three.

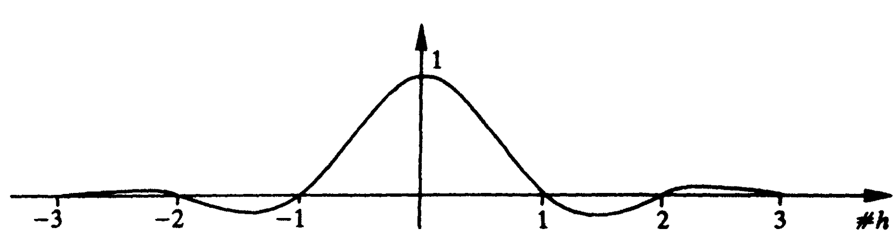

FIG. 1. *Basis cubic spline.*

The cubic B-spline is used more commonly than $\psi(x)$ to represent cubic splines since $B_3(x)$ has compact support. However, $\psi(x)$ decays very rapidly, and thus the sum in (17) is essentially finite for fixed x. Further, $\psi(x)$ reproduces the coefficient data (i.e., $s(x_i)=y_i$ in (17)) so that in practice reconstruction using $\psi(x)$ is probably more appropriate when infinitely many samples (or a very large number of samples) must be interpolated.

3. Derivation of the sampling kernels. In this section we derive the kernels $k(x)$ which should be used for sampling when a fixed reconstruction method is specified. The mathematical derivation given in § 3.1 is standard and included only for completeness. Section 3.1 may be skipped, if desired, since § 3.2 presents a summary of the method for finding the sampling kernels and applies the method to zero, first and third order spline reconstruction. The resulting solution kernels are displayed in Fig. 2.

3.1. Mathematical formulation. Denote by A the linear transformation which takes a sequence of sample values $\{y_i\}_{i=-\infty}^{\infty}$ into the reconstructed function $\tilde{f}(x)$. We write

$$\tilde{f}=A\mathbf{y}$$

regarding the sequence of samples as an infinite vector $\mathbf{y}$. For spline reconstruction, $s=A\mathbf{y}$ is given by equation (7), (10) or (17). We assume a priori that A restricted to square summable sequences ($\mathbf{y}\in l^2$) is a bounded linear transformation into $L^2(\mathbb{R})$.

For a given function $f\in L^2(\mathbb{R})$, the samples $\mathbf{y}$ should be obtained so as to minimize the L^2 norm

$$\|A\mathbf{y}-f\|_{L^2}. \tag{18}$$

A standard argument shows that $\mathbf{y}$ is a solution to the "normal equation" [10]

$$A^*A\mathbf{y}=A^*f,$$

where A^* is the adjoint transformation to A, taking $L^2(\mathbb{R})$ into l^2. Note that A^*A is a bounded linear transformation taking l^2 into l^2. Providing the transformation is nonsingular, we obtain

$$\mathbf{y}=(A^*A)^{-1}A^*f.$$

In particular, the samples y_i depend linearly on the function $f(x)$. That is, y_i is a projection of the linear operator $(A^*A)^{-1}A^*$ operating on f and thus given by a formula

$$y_i=\int_{-\infty}^{\infty}k_i(x)f(x)\,dx, \tag{19}$$

where $k_i(x)$ is the $L^2(\mathbb{R})$ dual element representing the y_i functional. Finally, if A is covariant with respect to rigid left and right shifts of the samples (i.e., the transformation $y_i\to y_{i+k}$ transforms $s=A\mathbf{y}$ by $s(x)\to s(x+x_k)$), then the kernels $k_i(x)$ are independent of i modulo interval shifts thus yielding a unique kernel $k(x)$,

$$k_i(x)=k(x-x_i).$$

To calculate $k(x)$, it suffices to find $k_0(x)$. Let $\mathbf{v}=(A^*A)^{-1}A^*\delta_t$, where $\delta_t(x)=\delta(x-t)$ is the delta function with unit mass at $x=t$. Then from (19),

$$v_0=\int_{-\infty}^{\infty}k_0(x)\delta(x-t)\,dx, \tag{20}$$

so clearly $v_0=k_0(t)$. We will calculate v_0 and thus $k_0(t)$, when A is given by a formula of the form

$$(21) \qquad (A\mathbf{y})(x) = \sum y_i\phi(x - x_i).$$

For nearest neighbor, linear and cubic spline interpolation, $\phi(x)$ is the cardinal spline $B_0(x)$, $B_1(x)$ and $\psi(x)$ of (17), respectively.

Set $\mathbf{w} = A^*\delta_t = A^*A\mathbf{v}$. If $\mathbf{e}_i \in l^2$ denotes the unit basis vector with a nonzero value in the ith coordinate and (,) denotes the standard inner product in l^2 or in L^2, then

$$(22) \qquad w_i = (\mathbf{w}, \mathbf{e}_i) = (A^*\delta_t, \mathbf{e}_i) = (\delta_t, A\mathbf{e}_i) = \int \delta(x-t)\phi(x - x_i)\,dx = \phi(t - x_i).$$

Further, since $\mathbf{w} = A^*A\mathbf{v}$,

$$w_i = \sum_{j=-\infty}^{\infty} a_{ij}v_j,$$

where

$$a_{ij} = (A^*A\mathbf{e}_i, \mathbf{e}_j) = (A\mathbf{e}_i, A\mathbf{e}_j) = \int_{-\infty}^{\infty} \phi(x - x_i)\phi(x - x_j)\,dx$$
$$= (\phi * \phi)(x_j - x_i) = \tilde{a}_{j-i}.$$

That is, A^*A, which takes l^2 into l^2, can be viewed as an infinite matrix whose entries $a_{ij} = \tilde{a}_{j-i}$ are constant on diagonal bands (i.e., a Toeplitz matrix). In § 5 we show that A^*A can be inverted for the various forms of $\phi(x)$ under consideration, yielding for $\mathbf{w} = A^*A\mathbf{v}$,

$$(23) \qquad v_i = \sum_{j=-\infty}^{\infty} c_{j-i}w_j$$

for a set of constants $\{c_i\} = \mathbf{c}$. Since $k_0(t) = v_0$ and $\mathbf{w}$ is given by (22), we have

$$(24) \qquad k_0(t) = \sum_{j=-\infty}^{\infty} c_j\phi(t - x_j) = A\mathbf{c}.$$

Note that (24) implies that the optimal kernel for a fixed reconstruction method is itself a reconstructed function, using the data $\mathbf{c}$ as sample values. If the reconstruction method uses kth order splines, then the sampling kernel $k(x)$ will be a kth order spline. The data vector $\mathbf{c}$ is obtained from the infinite matrix inverse to the matrix A^*A which in turn is the matrix of L^2 inner products of the basis functions $\phi(x - x_i)$ used for reconstruction.

3.2. The solution kernels. The method for finding the optimal sampling kernel for a reconstruction method of the form (21) can be summarized as follows:

Let $\phi(x)$ denote the basis function which serves to interpolate data $\{y_i\}$ to a function $(A\mathbf{y})(x)$ given by (21). Construct the infinite matrix A^*A whose entries a_{ij} are given by the L^2 inner products

$$a_{ij} = \int_{\infty}^{\infty} \phi(x - x_i)\phi(x - x_j)\,dx = (\phi * \phi)(x_j - x_i),$$

noting that the matrix is symmetric and Toeplitz (i.e., a_{ij} depends only on $|j - i|$). Invert this matrix (see § 5) to obtain a symmetric Toeplitz matrix whose values in diagonal and successive off-diagonal bands are denoted by $c_0, c_1, c_2, \cdots$. The solution kernel is obtained from the reconstruction (using the basis function ϕ) of the data $\{c_i\}$, where $c_{-i} = c_i$.

We supply four applications of this procedure.

(i) *Nearest neighbor reconstruction.* When $\phi(x)$ is $B_0(x)$, as defined in (8), we have

$$a_{ii} = \int [B_0(x - x_i)]^2 \, dx = h,$$

$$a_{ij} = \int B_0(x - x_i) B_0(x - x_j) \, dx = 0, \qquad i \neq j$$

so $A^*A = hI$, where I is the infinite identity matrix. Thus, $(A^*A)^{-1} = 1/hI$ and $\mathbf{c} = (\cdots 0, 0, 1/h, 0, 0 \cdots)$, so $k(x) = B_0(x)/h$ (see Fig. 2a).

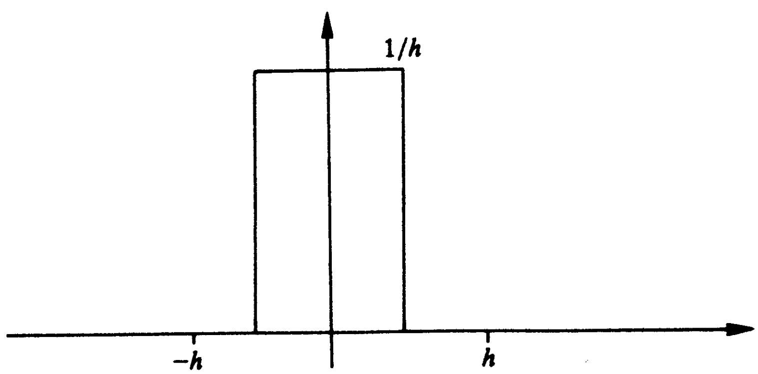

FIG. 2a. *Sampling kernel for nearest neighbor reconstruction.*

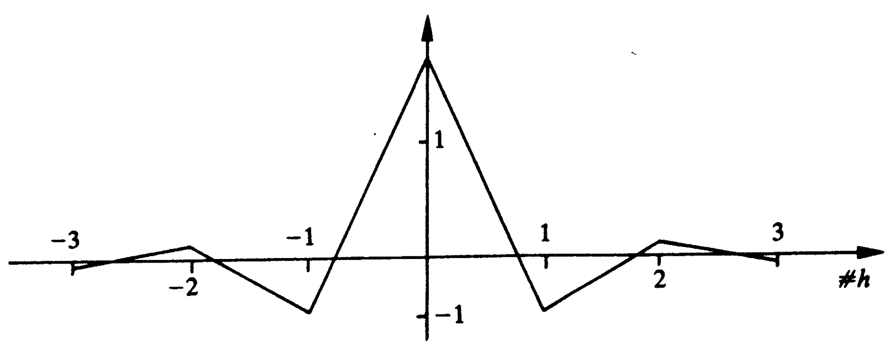

FIG. 2b. *Sampling kernel for linear reconstruction.*

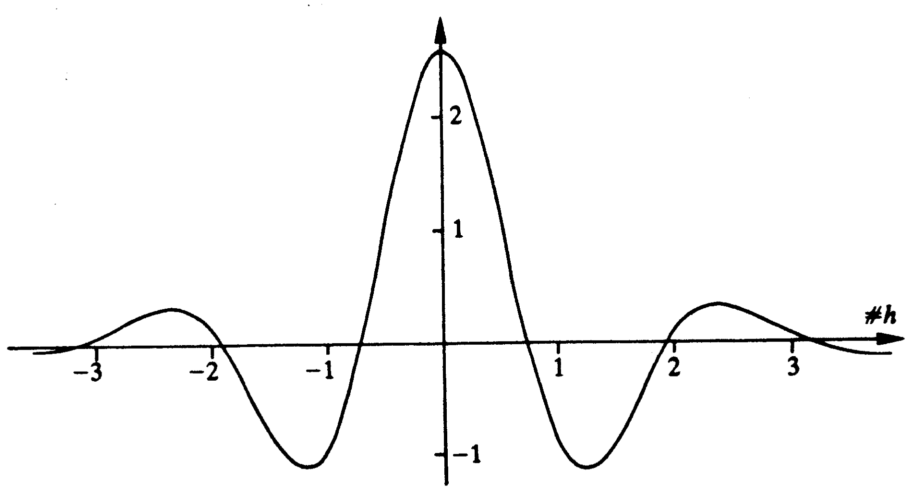

FIG. 2c. *Sampling kernel for cubic spline reconstruction.*

Note that the sampling kernel corresponds to a sampling method based on unweighted block averaging. That is, if nearest neighbor interpolation is to be used as the reconstruction method, then the samples should be obtained by computing the mean value of the input signal in intervals centered at each sampling point, each interval having the same length as the sampling rate.

(ii) *Linear interpolation.* When linear interpolation is used for reconstruction, $\phi(x)=B_1(x)$, (10), $(\phi^*\phi)(x)=B_3(x)$ and for $a_{ij}=\tilde{a}_{j-i}$,

$$\tilde{a}_{-1}=B_3(-h)=\tfrac{1}{6}, \qquad \tilde{a}_0=B_3(0)=\tfrac{2}{3},$$

$$\tilde{a}_1=B_3(h)=\tfrac{1}{6}, \qquad \tilde{a}_n=B_3(nh)=0, \qquad |n|\geqq 2.$$

Thus, $A^*A\mathbf{v}=\mathbf{y}$ is the transformation

$$y_i=\tfrac{1}{6}v_{i-1}+\tfrac{2}{3}v_i+\tfrac{1}{6}v_{i+1} \quad \text{for all } i.$$

In § 5 we show that this transformation can be inverted and gives rise to the vector $\mathbf{c}$ satisfying $c_i=c_{-i}$ and approximately $(c_0, c_1, \cdots)=(1.732, -.464, .124, -.033, .009, -.002, \cdots)$. More precisely,

$$c_0=\sqrt{3}, \qquad c_{i+1}=(\sqrt{3}-2)c_i \quad \text{for } i>0. \tag{25}$$

For linear interpolation the optimal sampling kernel $k(x)$ is the linear spline interpolating this data, as displayed in Fig. 2b.

Accordingly, when linear interpolation will be used to reconstruct functions from sampled data, the optimal samples are obtained by convolving the original function against the piecewise linear kernel displayed in Fig. 2(b). Note that this kernel decays geometrically so that the convolution has essentially finite support, with a radius of three or four intervals for high accuracy. The kernel can be generated easily using (25) and can be implemented quite easily in special purpose electronic hardware embedded in the sampling sensor.

(iii) *Cubic spline interpolation.* For cubic spline interpolation, it is somewhat simpler to use the B-spline $B_3(x)$ (11) and to define the transformation $s=A\mathbf{v}$ by (13) to determine the kernel $k(x)$. The vector $\mathbf{v}$ which minimizes $\|s-f\|=\|A\mathbf{v}-f\|$ satisfies, as before,

$$\mathbf{v}=(A^*A)^{-1}A^*f$$

and yields the optimal samples $\mathbf{y}$ given by (14). In this case, A^*A is an infinite matrix with seven nonzero bands containing the constant values $B_7(ih)$, $i=-3,-2,-1,0,1,2,3$, where $B_7(x)=(B_3^*B_3)(x)$ is the B-spline of degree 7 centered at zero. The inversion of this matrix can be done numerically. The reconstructed function gives a kernel for determining the components of $\mathbf{v}$, which in turn determines by means of (14) a kernel $k(x)$ for sampling the values of $\mathbf{y}$.

In fact, the inversion of the matrix A^*A yields the values $\{c_i\}$ given approximately by $(c_0, c_1, \cdots)=(4.96, -3.09, 1.71, -.92, .49, -.26, .14, \cdots)$. Then $k_0(x)=\sum_{i=-\infty}^{\infty} c_iB_3(x-x_i)$, and the resulting kernel $k(x)$ (which is a cubic spline) is shown in Fig. 2c.

The kernels for both linear and cubic spline interpolation (Fig. 2b and 2c) exhibit oscillatory behavior reminiscent of a $\sin x/x$ kernel. The functions are *not* spline approximations to $\sin x/x$, however, since they decay exponentially. The oscillatory behavior and exponential decay result from inverting the positive matrix A^*A.

For the sampling kernel in Fig. 2c, a truncated version with a support radius of four intervals will yield very nearly the same samples.

(iv) ($\sin x/x$) *reconstruction*. In any case the observation that the kernels in Fig. 2b and 2c behave qualitatively like ($\sin x/x$) is consistent with the result stated earlier that ($\sin x/x$) reconstruction calls for low-pass prefiltering. We can use the methods of § 3.1 to verify this result. Specifically, we use the basis function

$$\phi(x)=\frac{\sin(\pi x/h)}{\pi x/h}\equiv \operatorname{sinc}\left(\frac{\pi x}{h}\right).$$

The coefficients of the A^*A matrix are calculated from the function given by the convolution of $\phi(x)$ with itself. However, the Fourier transform of $\operatorname{sinc}(x)$ is $\pi\cdot\operatorname{rect}(x)$. (NB: $\operatorname{rect}(x)$ is the characteristic function of the interval $[-1, 1]$.) Thus, $(\operatorname{sinc}^*\operatorname{sinc})(x)$ is a function whose Fourier transform is $\pi^2\cdot\operatorname{rect}^2(x)=\pi^2\cdot\operatorname{rect}(x)$, namely, $\pi\cdot\operatorname{sinc}(x)$. After a scaling substitution, we obtain $\phi^*\phi=h\phi$ and $(A^*A)_{i,j}=h\phi(x_i-x_j)=h\delta_{i,j}$. That is $A^*A=hI$, and from (24), it follows that the sampling kernel is $(1/h)\phi$, as stated in (5), which implies a low-pass filter.

4. Resampling. We will consider a particular example. More general cases can be treated analogously.

Suppose we are given a sequence of data $\{v_i\}_{i=-\infty}^{\infty}$, and wish to resample to achieve a threefold compression factor. That is, we wish to find samples $\{y_i\}_{i=-\infty}^{\infty}$ so that y_i represents the block of data containing v_{3i-1}, v_{3i} and v_{3i+1} (see (6)). The sample $\{y_i\}$ should be determined so that the vector $\mathbf{v}'$ reconstructed from $\mathbf{y}$ by, say, linear interpolation should be the closest l^2 approximate to $\mathbf{v}$. We write $\mathbf{v}'=A\mathbf{y}$ and note that

$$\begin{aligned} v_0' &= y_0,\\ v_1' &= \tfrac{2}{3}y_0+\tfrac{1}{3}y_1,\\ v_2' &= \tfrac{1}{3}y_0+\tfrac{2}{3}y_1,\\ v_3' &= y_1, \quad \text{etc.} \end{aligned}$$

The operator A is represented by an infinite matrix, whose columns contain the values $(\cdots 0, \frac{1}{3}, \frac{2}{3}, 1, \frac{2}{3}, \frac{1}{3}, 0, \cdots)$. These values are shifted downward three rows in each successive column. As before, the optimal samples are given by

$$\mathbf{y}=(A^*A)^{-1}A^*\mathbf{v},$$

where A^* is the adjoint, or transpose, matrix to A. The sampling weights $\{k_j\}$ (see (6)) will arise as the "middle" row of values in the matrix $(A^*A)^{-1}A^*$.

To calculate the weights $\{k_j\}$, we first observe that A^*A is given by a matrix $\{a_{ij}\}$, with constant bands $a_{ij}=\tilde{a}_{j-i}$, and

$$\tilde{a}_{-1}=\tfrac{4}{9},\quad \tilde{a}_0=\tfrac{19}{9},\quad \tilde{a}_1=\tfrac{4}{9},\quad \tilde{a}_k=0,\quad |k|\geqq 2.$$

When this matrix is inverted (§ 5), one obtains a symmetric matrix with constant bands containing the values (approximately)

(26) $\quad c_0=.5222,\quad c_1=-.1153,\quad c_3=.0255,\quad c_4=-.0056,\quad$ etc.

The resulting weights $\{k_j\}$ are obtained by linearly interpolating two values between each component of $\mathbf{c}$, yielding the weights plotted in Fig. 3.

Once again, we observe an oscillatory behavior of the weights and a geometric decay.

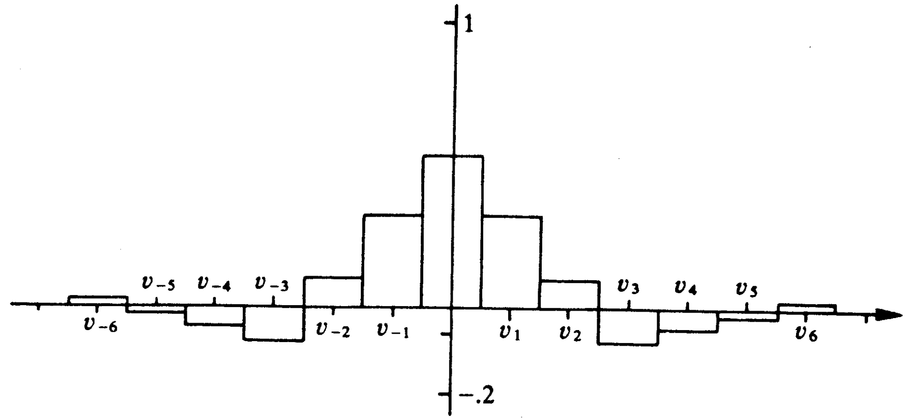

FIG. 3. *Sampling weights for 3 to 1 data compression, assuming linear reconstruction.*

5. Inverting infinite matrices. Consider a transformation T taking l^2 into l^2 of the form

$$w_i = (T\mathbf{v})_i = \sum_{j=-l}^{l} a_j v_{i+j} \quad \text{for all } i.$$

When $l = 1$ there are three coefficients a_{-1}, a_0 and a_1 and so the infinite matrix corresponding to T is tridiagonal. When $a_j = a_{-j}$, the transformation and its matrix are symmetric.

For such infinite symmetric tridiagonal matrices, it can be shown that the matrix is invertible if and only if it is diagonally dominant, in which case the inverse is given by the formulas

$$v_i = \sum_{j=-\infty}^{\infty} c_j w_{i+j}, \tag{27}$$

where

$$c_j = c_{-j}, \quad c_0 = \frac{1}{a_1\sqrt{(a_0/a_1)^2 - 4}}, \quad c_{k+1} = \alpha c_k.$$

Here α is a root of the polynomial $a_1x^2 + a_0x + a_1$ satisfying $|\alpha| < 1$. The values in (25) and (26) come from applying these formulas.

When there are more than three nonzero constant bands, such as in the calculation of the kernel for cubic spline interpolation in § 3.2(iii), formulas for a bounded inverse of the form of (27) depend on the roots of the polynomial $\sum a_i z^{i+l}$. It can be shown that an inverse exists if none of the roots in the complex plane lie on the unit circle. In that case the coefficients c_j in the inverse transformation, (27), will decay rapidly but not necessarily in a precise geometric fashion as in the tridiagonal case. For symmetric infinite matrices with five, seven and nine nonzero bands, explicit formulas using roots to quadratic, cubic and quartic polynomials are at least theoretically available. In practice, however, if an inverse exists, the middle row of the inverses of finite versions of the matrix will converge rapidly to the sequence of values c_i in the inverse. This process is frequently simpler to apply than the explicit formulas. In any case our viewpoint in terms of the applications of these results to sampling is that the inverse should be obtained once and for all for the purpose of finding the sampling kernel. The kernel will generally be implemented directly, so that the linear system will never require resolving. Since the values c_j in the bands of the inverse Toeplitz matrix asymptotically decay exponentially, only a few significant nonzero values are involved in the construction of the kernels.

6. Summary. The substitution of the appropriate sampling kernel of Fig. 2a, b, c for the more standard delta function sampling method, whenever practical, may lead to reduced mean square error after reconstruction. The expected decrease in mean square error depends on the autocorrelation function of the initial function but in some instances can be substantial. For example, if the autocorrelation function decays exponentially and if the sampling is very fine, a numerical calculation shows that a 35% reduction in total expected mean square error can be expected by using the optimal sampling kernel prior to linear reconstruction as opposed to normal delta function sampling. The improvement compared to block averaging is a less dramatic 17% of the mean square error.

Alternatively, one can view optimal sampling as permitting wider sampling distances for a fixed mean square error. For example, suppose that the autocorrelation function decays exponentially and that a sampling rate is desired so that the expected mean square error will be no larger than 10% of the signal variance when linear interpolation is used for reconstruction. A numerical analysis shows that relative sampling rates of .3, .4 and .475 are required for delta sampling, block averaging and sampling using the optimal kernel for linear reconstruction respectively. Thus fewer samples are needed if optimal sampling is used. Of course, for different autocorrelation functions, these figures will change, but since the optimal kernel is derived by minimizing expected mean square error, some improvement in expected L^2 error is guaranteed.

Extension to sampling in two or more dimensions is straightforward, especially if the reconstruction basis functions are separable in each variable. Extending these results to higher order splines or other linear reconstruction methods is straightforward. It is also possible to find the optimal kernel for a fixed spline reconstruction scheme, subject to the constraint that the kernel vanish identically outside a fixed neighborhood of the origin. In that case, the best kernel depends on the autocorrelation function of the original data and can be determined from the solution to a variational problem. In all realistic cases, the resulting kernel is closer to a truncated version of the optimal kernel than it is to a delta function or to the block averaging kernel.

Acknowledgment. The author thanks Professor Azriel Rosenfeld, who suggested this study in the context of sampling and data compression. Our initial interest was to determine weights for resampling digitized image data to reduce tape storage requirements.

REFERENCES

[1] C. E. SHANNON, *Communications in the presence of noise*, Proc. IRE, 37 (1949), pp. 10–21.
[2] W. M. BROWN, *Optimal prefiltering of sampled data*, IRE Trans. Inform. Theory, 7 (1961), pp. 269–270.
[3] R. BERNSTEIN, *Precision processing of earth image data*, American Scientist, 64 (1976), pp. 500–508.
[4] J. H. AHLBERG, E. N. NILSON AND J. L. WALSH, *The Theory of Splines and Their Applications*, Academic Press, New York, 1967.
[5] M. J. D. POWELL, *The local dependence of least squares cubic splines*, SIAM J. Numer. Anal., 6 (1969), pp. 398–413.
[6] E. L. HALL, *Computer Image Processing and Recognition*, Academic Press, New York, 1979.
[7] W. K. PRATT, *A comparison of digital image transforms*, Data Compression, L. Davisson and R. Gray, eds., Dowden, Hutchinson and Ross, 1976.
[8] P. ELIAS, *Predictive coding*, IRE Trans. Inform. Theory, 1, (1955), pp. 16–23, 30–33.
[9] A. RALSTON, *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1965.
[10] G. W. STEWART, *Introduction to Matrix Computations*, Academic Press, New York, 1973.