APPROXIMATE RECONSTRUCTION OF RANDOMLY SAMPLED SIGNALS

J. JIMÉNEZ

IBM Scientific Centre, Paseo Castellana 4, 28046 Madrid, Spain

J.C. AGÜÍ†*

School of Aeronautics, Universidad Politécnica, 28040 Madrid, Spain

Abstract. We study the use of polynomial interpolation to approximate a function specified by samples taken at random moments satisfying a Poisson distribution with uniform mean sampling rate. Two different selection schemes are considered to determine which samples should be used in the construction of the polynomials, and detailed error estimates are derived for each case. The results are compared with the classical interpolation methods of convolution with a smoothing window. It is concluded that only low order polynomials are useful for interpolation in the presence of noise, but that they are comparable or superior to nonadaptive convolution in most cases, as well as computationally more efficient. Some simulation experiments are presented to support the theoretical estimates.

Zusammenfassung. Wir untersuchen die Anwendung der Polynominterpolation zur Annäherung einer Funktion, die definiert ist durch einzelne Abtastwerte an zufällig ausgewählten Stellen, wobei die Abtastintervalle poissonverteilt sind und einen konstanten Mittelwert aufweisen. Betrachtet werden zwei verschiedene Auswahlschemen zur Entscheidung darüber, welche Abtastwerte bei der Konstruktion der Polynome verwendet werden; hierbei werden für jeden betrachteten Fall die Schätzfehler eingehend analysiert. Die Erbegnisse werden verglichen mit der klassischen Interpolationsmethode der Faltung mit Hilfe eines Glättungsfilters. Es zeigt sich, daß für die Interpolation verrauschter Signale nur Polynome niedrigen Grades nützlich sind; diese jedoch ergeben in den meisten Fällen bessere Resultate als die Faltungsmethode und sind darüber hinaus weniger rechenaufwendig. Einige Ergebnisse von Simulationsexperimenten werden beschrieben; sie unterstützen die theoretisch erhaltenen Resultate.

Résumé. Nous étudions l'utilisation de l'interpolation polynomiale pour approcher une fonction spécifiée par des échantillons prélevés à des instants aléatoires, satisfaisant une distribution de Poisson avec une cadance d'échantillonnage moyenne uniforme. Deux méthodes différentes de sélection ont été considérées pour déterminer quels échantillons doivent utilisés dans la construction des polynômes, et des estimations d'erreurs detaillées sont dérivées pour chaque cas. Les résultats sont comparés avec les méthodes d'interpolation classiques de convolution avec une fenêtre d'adoucissement. On conclue que seuls les polynômes d'ordre faible sont utiles pour l'interpolation en présence de bruit, mais qu'ils sont comparables ou supérieurs à la convolution non adaptative dans la plupart des cas tout en étant plus efficace du point de vue de charge de calcul. Certains expériences de simulation sont présentées pour renforcer les estimations théoriques.

Keywords. Interpolation, random sampling, polynomial, approximation theory, experimental fluid mechanics.

1. Introduction

The basic result in the theory of signal recon- struction from discrete samples is that a function whose Fourier transform is limited to frequencies smaller than ν can be reconstructed exactly from samples taken at a uniform rate higher than the Nyquist limit 2ν [17]. This theorem can be inter- preted as stating that the number of degrees of freedom in a bandlimited function defined in an interval $0 \leqslant x \leqslant L$ is approximately $2L\nu$, and it has long been recognised that those degrees of freedom can be satisfied in many different ways, including mixed values of the function and its derivatives [6, 10, 17], and a variety of nonuniform sampling schemes [2, 13, 14, 15, 17, 18, 19]. In this paper we are interested in the reconstruction of a signal from samples obtained at random times satisfying a Poisson distribution with an average sampling rate λ.

Nonuniform sampling schemes appear in several applications, including missing-data recovery [3, 12, 13, 19], signal coding [14, 15], and the correction of jitter error [13, 18]. Completely random coding seems to have been mainly of theo- retical interest, although Poisson sampling occurs naturally in some experimental applications [3], in particular when a measurement has been done in conjunction with particular events which are outside the control of the experimenter. Think, for example, of trying to monitor a particular point on Earth from a large swarm of satellites which pass above Earth at effectively random times, or think of trying to gauge the standard of living in a foreign city from occasional visits of friends who live there. We came across it, in a two-dimensional context, while working on the determination of flow velocities from tracking particles dispersed in turbulent flows [5, 8, 11].

Quite a lot is known about Poisson sampling. It was shown in [3, 12] that Poisson samples of any infinite signal, not necessarily bandlimited, are enough for the alias-free determination of its power spectrum. A practical method for computing the spectrum is given in [7]. The conditions for

reconstruction of a signal are generally more demanding than those for spectral determination, but it was shown in [2] that a bandlimited infinite signal can exactly be reconstructed from its Poisson samples iff the average sampling rate is at least the Nyquist rate.

Practical reconstruction algorithms are less well developed. For uniform sampling rates, the exact reconstruction is obtained by convolution of the discrete samples with a smoothing window of a particular form [4, 17]. For some nonuniform sampling schemes, the exact reconstruction kernel is also known [19], but the expressions grow increasingly complicated with the amount of non- uniformity, and the numerical conditioning of the reconstruction is also known to deteriorate rapidly. For almost uniform sampling, it is possible to get a good approximate reconstruction using the same convolution kernel as for the uniform case, con- trolling the resulting 'jitter' error [13] with appropriate techniques [14, 15]. None of these methods works very well for Poisson sampling, which is highly nonuniform.

The result quoted above, namely that an infinitely large set of Poisson samples is enough to completely reconstruct the spectrum of a signal, independently of bandwidth or mean sampling rate, has interesting consequences in reconstruc- tion. Intuitively, it is due to the fact that, for any arbitrarily small threshold, there is a finite proba- bility of finding two samples which are closer together than that threshold. Unfortunately, the same is true for the probability of finding consecu- tive samples which are farther apart than any given distance. That means that some stretches of the signals are highly oversampled with respect to the mean sampling rate, while others are badly under- sampled, and that it is impossible to treat the sampled signal as bandlimited. As a consequence, a lot of the standard interpolation theory is not applicable.

A particularly bothersome problem is the effect of high frequency noise. In dealing with uniformly sampled signals, it is permissible to assume that the underlying continuous process contains no

frequencies above the Nyquist limit. In essence, they have been filtered out by the sampling process. In the case of Poisson sampling, this is not so. Some intervals of the signal, those which are over-sampled by the Poisson process, will contain high frequencies which, according to the Nyquist criterion, cannot be adequately reconstructed in a global sense, but which have an effect in any practical filter, and have to be dealt with. Moreover, this effect is not just global, but occurs microscopically at individual points. Consider the problem of interpolating at a point which has two closely spaced samples to one side, and just one, very distant sample, to the other. It will be seen below that the effect of these spurious frequencies is to induce very large errors in some interpolators, which should be avoided. On the other hand, even if these high frequencies cannot be reconstructed globally, they can be reconstructed locally, at those intervals which happen to be densely sampled, and it would be wrong to throw them away everywhere by using an interpolation filter matched only to the mean sampling rate. In particular, it should be possible to interpolate the signal at a rate above the Nyquist limit in such a way that the fast variations are retained locally wherever possible, without introducing spurious results in the poorly sampled intervals. It even makes sense to try to interpolate the signal to a *continuous* waveform. What is needed is a filtering method which adapts itself to the local structure of the samples, while minimising the unwanted effects of the mixed sampling rate.

Here, we explore the use of direct polynomial interpolation. It was shown in [10] that the exact convolution windows are the limit of the Lagrange interpolating polynomials when the order of interpolation tends to infinity. Only for some particular sample distributions does this limit tend to a function that can be expressed in closed form and, in any case, it can be shown that exact reconstruction implies windows that are infinite in extent. Since this is not practical, windows are usually approximated by finite versions that induce an error even in the cases in which the exact kernel is known.

In the case of random sampling, this suggests the use of relatively low order interpolation polynomials based on the samples closest to the point being approximated. These polynomials are computationally efficient and have the advantage of using information from the known positions of the sample points, without incorporating unnecessary data from samples which are far away and which would be neglected in any case by the approximation of the window to a finite interval.

In the next three sections we consider the problem of how to choose the samples in which to base an interpolating polynomial of a given order so as to minimise the error, and we study the errors induced as a function of the frequency of the original signal. We consider two different choice criteria and give both theoretical and numerical simulation results for the approximation errors. We consider next which are the implications of these results for the interpolation of noisy signals and, finally, compare the errors obtained with those resulting from the convolution with smoothing windows, as well as the computational complexity of both procedures.

2. Polynomial approximations for smooth functions

The theory of polynomial interpolation of a function from discrete samples is well developed. Assume, without loss of generality, that we are interested in approximating the function at $x = 0$, and that we are given values f_i at $n+1$ points x_i, $i = 0, 1, \ldots, n$. The unique interpolating polynomial of order at most n is defined by the Lagrange formula [9]

$$z(x) = \sum_i \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)} f_i,$$

(1)

and, if the $(n+1)$st derivative of the original function exists, the pointwise error at $x = 0$ can be shown to be

$$z(0) - f(0)$$
$$= (-1)^{n+1} x_0 x_1 \cdots x_n f^{(n+1)}(\zeta)/(n+1)!, \quad (2)$$

where ζ is a point in the interval defined by $x = 0$ and by all the sample points x_i. If the function is smooth enough, we can approximate the derivative in equation (2) by its value at zero, and the error becomes simply proportional to the product of the abscissae at the sampling points,

$$Q = x_0 x_1 \cdots x_n. \tag{3}$$

To minimise this product, and the interpolation error, the simplest choice is to use for the interpolation the $n + 1$ points closest to $x = 0$. For reasons that will become clear later, we shall refer to this choice as 'unbalanced'.

Assume that we label those points in order of increasing distance from $x = 0$. The closest point x_0 has an equal probability of lying to either side of the origin. The second point x_1 has again a uniform probability of lying anywhere outside the segment $(-x_0, x_0)$ and, in particular, also has an equal probability of being positive or negative. The same is true for any of the $n + 1$ sampling points and, as a consequence, for the product Q. The average error is then zero and the estimation based in this set of points is unbiased. The expected absolute value of the error depends on the higher moments of distribution of Q, which can be computed exactly and which are given in Appendix A. As expected from the consideration that λ is the only dimensional quantity with which to scale x, the qth moment is proportional to $\lambda^{-q(n+1)}$ and, from equation (2), the root mean square error of the approximation of f by an nth-order polynomial is proportional to $f^{(n+1)}(0)/\lambda^{n+1}$. When the function being approximated is smooth or, in other words, when the frequencies being considered are much lower than the average sampling frequency λ, this error is small and decreases exponentially with increasing order of the interpolating polynomial.

3. The high frequency limit

The estimates in the previous section do not hold when the frequencies involved are comparable or

larger than the sampling frequency, in which case the substitution of ζ by zero in equation (2) is not justified and the analysis becomes more complicated. As argued in Section 1, even if those frequencies are above the *mean* Nyquist limit, they are present in the original set of samples, and their effect has to be taken into account.

The interesting case here is the limit in which the signal frequency is much higher than the mean sampling rate. Consider, to fix ideas, that the function being interpolated is a sine wave of unit amplitude,

$$f(x) = \sin(2\pi\lambda x/\alpha), \tag{4}$$

where the parameter α is the average number of sampling points per period, and the high frequency limit is $\alpha \to 0$. Since all the interpolators considered here are linear, the behaviour of any signal can be estimated from the superposition of components of the form (4).

The simplest case to analyse is the zeroth order interpolation, in which the function is approximated by its value at the closest sampling point. In the high frequency limit, this value is essentially uncorrelated to the value at the point of interpolation, and the variance of the error is of the same order of magnitude as the amplitude of the band of variation of the function. In the particular case of equation (4),

$$\sigma_0^2 = \langle (f(x) - f(x_0))^2 \rangle$$
$$= \langle f^2(x) \rangle + \langle f^2(x_0) \rangle = 1. \tag{5}$$

The next simplest case is the linear approximation ($n = 1$). The error is then

$$z(0) - f(0) = \frac{x_1 f_0 - x_0 f_1}{x_1 - x_0}, \tag{6}$$

where we have assumed, again without loss of generality, not only that $x = 0$ but that $f(0) = 0$ at the point of interpolation.

The expected magnitude of the error depends on whether x_0 and x_1 have the same or opposite signs. Both cases are equally probable. In the second case, it is easy to see that the interpolated value is bounded by f_0 and f_1 (see Fig. 1(a)), and
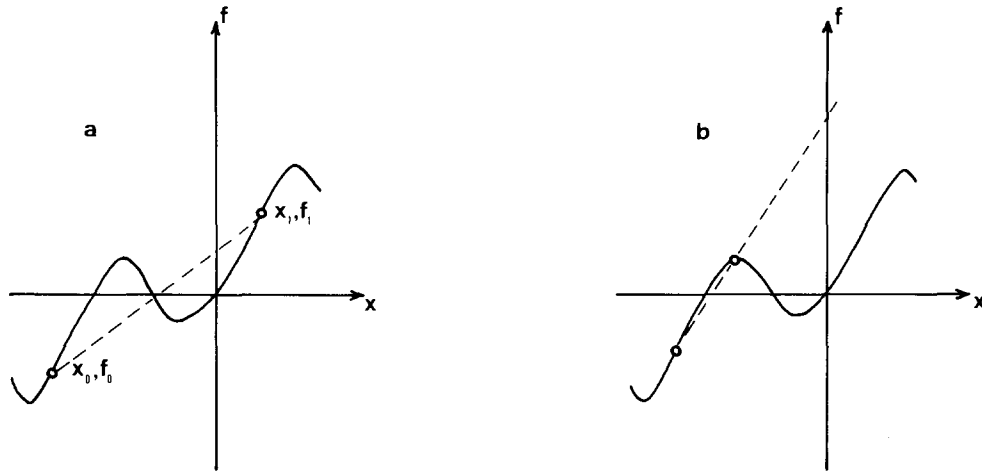
Fig. 1. Errors in linear interpolation are bounded by the range of the signal when the samples bracket the origin (a), but may become much larger otherwise (b).

the error is again of the same order of magnitude as the band of variation of the function. In the first case, the errors can be larger, in essence because we are using an extrapolation rather than an interpolation formula.

It follows that the probability distributions obtained in Appendix A that the variance of the linear interpolation error can be written as the integral

$$\sigma_1^2 = 2\lambda^2 \int_0^\infty \mathrm{d}x_1 \, \mathrm{e}^{-2\lambda x_1} \int_0^{x_1} \mathrm{d}x_0 \, [z(0) - f(0)]^2$$

$$= 2\lambda^2 \int_0^\infty \mathrm{d}x_1 \, \mathrm{e}^{-2\lambda x_1}$$

$$\times \int_0^{x_1} \mathrm{d}x_0 \left[f_1 - x_1 \frac{f_1 - f_0}{x_1 - x_0} \right]^2. \tag{7}$$

If we assume that, in the high frequency limit, f_1 and f_0 are uncorrelated, this integral is divergent, due to the behaviour of the inner bracket as $x_1 \to x_0$. The divergence is avoided in practice, at least for signals with finite bandwidth, because samples are not really uncorrelated when their separation tends to zero.

Assume now that f can be written as a smooth function in terms of a 'fast' variable

$$\xi = 2\pi\lambda x / \alpha.$$

In the limit $\alpha \to 0$, the integral (7) is dominated by the neighbourhood of the 'near singularity' at $x_1 = x_0$, and can be approximated by

$$\sigma_1^2 \sim \frac{4\pi\lambda^3}{\alpha} \int_0^\infty \mathrm{d}x_1 \, x_1^2 \, \mathrm{e}^{-2\lambda x_1}$$

$$\times \int_0^\infty \mathrm{d}\zeta \frac{[f(\xi_1) - f(\xi_1 - \zeta)]^2}{\zeta^2}$$

$$\sim \frac{2\pi}{\alpha} \int_0^\infty \mathrm{d}\zeta \frac{[f(\xi_1) - f(\xi_1 - \zeta)]^2}{\zeta^2}, \tag{8}$$

where the inner integral is now $O(1)$ since f is smooth in the new variables. For the particular case of equation (4), a representative value for that integral is, at $x_1 = 0$,

$$\int_0^\infty \mathrm{d}\zeta \sin^2 \zeta / \zeta^2 = \tfrac{1}{2}\pi. \tag{9}$$

A more rigorous estimate, taking into account all the possible values of x_1, gives the same result. This estimate only reflects the contribution to σ_1 from those situations in which both samples are located to the same side of the interpolation point and separated by a distance comparable or smaller than the signal wavelength (Fig. 1(b)). This contribution turns out to be $O(1/\alpha) \gg 1$, and dominates

the total error, giving the final result

$$\sigma_1 = \pi/\sqrt{2\alpha}, \tag{10}$$

which is valid for equation (4) in the high frequency limit. Note that the particular form of the signal determines the coefficient in equation (10), but that the dependence with frequency is universal.

In the same way in which the errors of the linear interpolation are dominated by compact point pairs, it is easy to convince oneself that the dominant errors for nth-order approximations are caused by compact groups of $n+1$ sampling points, clustered far away from the interpolation point. An analysis similar to the previous one would give in those cases

$$\sigma_n \sim \alpha^{-n/2}, \tag{11}$$

which is again a high frequency estimate. A calculation of the coefficients for these estimates is, in principle, straightforward but laborious. Its practical utility is dubious since the general message is that the errors from high-order interpolations are large in the high frequency limit and, therefore, that high-order polynomials should be avoided when high frequencies are present, as in the cases in which the signal bandwidth is large with respect to the mean sampling rate, or when the signal is contaminated by high frequency noise.

To check all these estimates we have performed a set of simulation experiments, interpolating equation (4) from samples taken at 1000 points, distributed randomly in the interval (0, 1). For each set of samples, the function was interpolated at 100 different points and the whole simulation was repeated between 100 and 1000 times for each value of n and α. The mean and standard deviation of the interpolation error were computed in each case and compared to their theoretical values. It was found that the estimation was indeed unbiased, not only 'globally' when the average was computed over many interpolation points, but 'locally' when a single point was interpolated many times from different samples. The results for the r.m.s. error are presented in Fig. 2. In that figure,
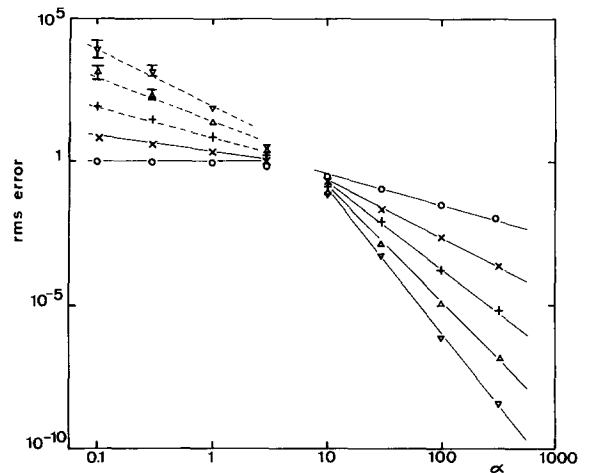

Fig. 2. Root mean square approximation error in polynomial 'unbalanced' interpolations of equation (4). Solid lines are estimates derived from equations (5) and (10), and from (A.6) in the appendix; dashed lines are from equation (11). Symbols are experimental results from simulation. $\bigcirc$: $n=0$; $\times$: $n=1$; $+$: $n=2$; $\triangle$: $n=3$; ∇: $n=4$. Error bars in some points are r.m.s. scatter of simulation results.

the points are experimental values while the different solid lines are the theoretical estimates from Appendix A and from equations (5) and (10). The dashed lines represent the slopes given by equation (11). The agreement is generally good, not only in the limit of smooth functions, in which the theory is well developed, but also in the high frequency range where the arguments are more qualitative.

A word should be said here about the nature of the large errors introduced by the high-order polynomials. It follows from both the simulation experiments and the theoretical estimates that most of the variance is due to a very small percentage of cases in which the errors are huge. Both in the high and in the low frequency limits, it can be shown that the probability of finding an error outside a few standard deviations is small and decreases with increasing order of interpolation. On the other hand, for high-order polynomials, the few outliers are liable to have values several orders of magnitude larger than the amplitude of the original signal. This is consistent with the well-

known property of high-order interpolating polynomials oscillating wildly and unpredictably, which is aggravated here by the appearance of close clusters and the possibility of extrapolation. Some theoretical and experimental probability functions are given in Appendix A, for the case of smooth functions. The general result is that the probability density function for the magnitude Q, proportional to the error of an nth-order polynomial, behaves as $\exp\{-2\lambda Q^{1/(n+1)}\}$. For large n, this function decreases very fast for small errors but has a long 'tail' in which large errors remain relatively probable.

While it might be possible, in practical cases, to try to detect these spurious values by some sort of a posteriori filtering procedure, this process is complicated by the sheer size of the errors. Thus, a relatively small contamination of a signal by high frequency noise, or even numerical error, can produce spikes in the interpolated signal which are of order unity and difficult to distinguish from real features in the data. It would appear, therefore, that only zeroth- or first-order interpolation are useful in the presence of noise.

On the other hand, the good approximating properties of higher-order polynomials for smooth functions make their use attractive, as long as their high frequency behaviour can be improved. A better choice of the sampling points used in the interpolation, which essentially seeks to control the possibility of extrapolation, is presented in the next section.

4. Balanced interpolation

We have seen that the worst problems in the interpolation of high frequency signals are due to those situations in which all the samples used to compute a polynomial are clustered to one side of the point being approximated, in such a way that the polynomial is used as an extrapolator. An obvious remedy is to avoid this situation as much as possible by choosing the samples in an appropriate way. Consider the following 'balanced' pro-

cedure. We have to choose $n+1$ samples to construct an nth-order interpolating polynomial and again assume that we are trying to approximate the function at $x = 0$. For the first point, x_0, we choose the one closest to the origin, whether it is positive or negative. For the second one we choose the next closest point, but with a sign opposite to that of x_0. The third one is again the closest remaining point, independent of sign. The fourth is forced to have a sign different from the third, and so on. It is clear that this procedure generates a set of points which is as close to the origin as possible but which is guaranteed to bracket it.

The effect of this new procedure on smooth functions is negligible and the errors are only slightly larger than in the previous case. The worst effect is that the estimation is not always unbiased. For an odd number of points, the product Q still has the same probability of being positive as negative and the expected bias is zero but, for an even number, the product always has the same sign, and the sign of the error follows that of the corresponding derivative. As an example, the error due to a balanced linear interpolation always has a sign opposite to the local second derivative, and has the effect of lowering the maxima and raising the minima, leading to an underestimation of the range of variation of the signal. Theoretical estimates for the bias and standard deviation in the case of smooth functions are given in Appendix A, and Fig. 3 shows r.m.s. values for the bias, both theoretical and experimental, and the results of the simulation. In the worst case, which is the linear interpolation, the bias is half as large as the standard deviation.

In the high frequency limit, the interpolated values are essentially uncorrelated to the function. In this case, the average value of the interpolation is always equal to the local mean value of the function (zero in the case of equation (4)), and the bias is equal to the difference between the value of the signal at the particular point being considered and the local mean. This is true both for 'balanced' and 'unbalanced' interpolation schemes.
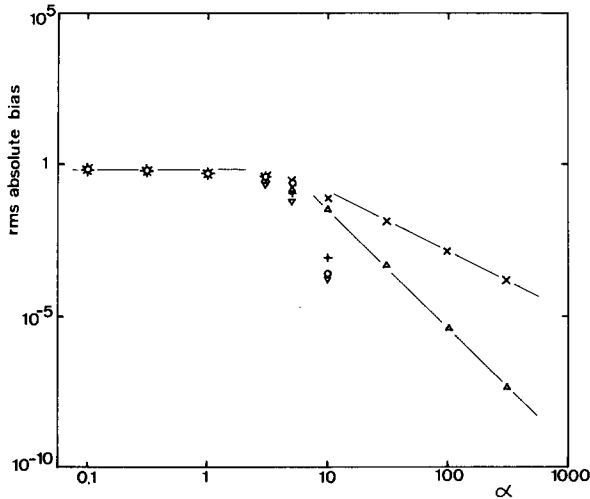
Fig. 3. Root mean square bias for 'balanced' polynomial interpolation of equation (4). Theoretical lines are from equation (A.11). Symbols are as in Fig. 2. Experimental values for even orders and $\alpha > 10$ are consistent with zero.
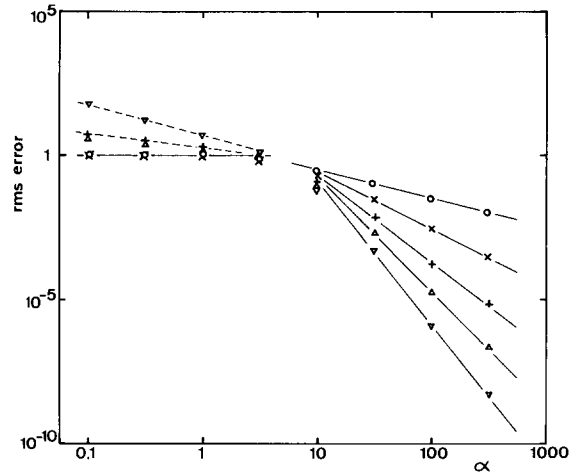

Fig. 4. Root mean square error for 'balanced' polynomial interpolation of equation (4). Theoretical lines are from equations (12), (A.12), and (A.14). Symbols are as in Fig. 2.

The effect of the new interpolation strategy on the standard deviation for high frequency signals is stronger. A priori, we would expect that there will never be extrapolation with one or two points, and that the standard deviation will be at most of order unity. For 2nd- and 3rd-order polynomials (3 or 4 points) there is the possibility of close pairs in either side of the origin, but of no triplets. In this case we would expect deviations similar to equation (10). In general, an nth-order polynomial will be based on $n + 1$ points, with a largest possible compact group of $\lfloor \frac{1}{2}n \rfloor + 1$ to one side of the origin. Using arguments similar to the ones in the previous section, we would expect the standard deviation to behave as

$$\sigma_n \sim \alpha^{-(\lfloor n/2 \rfloor)/2}, \tag{12}$$

which is asymptotically substantially smaller than equation (11).

Fig. 4 gives simulation results for the standard deviations of balanced interpolations for different orders, computed with parameters identical to those of Fig. 2. The results confirm the estimates and the fact that the errors produced in this method by the high frequency components are generally smaller than the ones found in the 'balanced'

approximation. Note that the largest error of fourth-order polynomials in the left part of Fig. 4 is almost three orders of magnitude below the corresponding value in Fig. 2.

5. The effect of noise

We have seen that all polynomial interpolations fail in the high frequency limit. High-order polynomials do so catastrophically, resulting in errors which are orders of magnitude above the original signals, while low-order ones, especially those based on balanced point sets, just fail to predict anything. This latter error is not particularly harmful and is just a consequence of the sampling theorem; its magnitude is of the same order as that of the high frequency components. The former kind, on the other hand, can have important negative effects in the presence of even small amounts of noise.

In uniform sampling schemes, there is no noise component with a frequency higher than the Nyquist limit associated to the (constant) sampling rate. Any high frequency noise originally present in the signal is folded by the sampling process into

frequencies lower than that limit. In the random case, however, the high frequency components present in the signal may still appear in the sampled version through the locally denser sampling rates at some intervals. Therefore, if an interpolation scheme amplifies these frequencies by a large factor, the interpolated values become swamped by the resulting error even for moderately small contamination by noise.

It is clear from the derivation that led to equation (10) that, if a given signal is Fourier analysed, the errors due to the different frequency components should add approximately as independent random variables, at least in the high frequency range in which the different components can be considered uncorrelated to each other. Therefore, in this limit, the squares of the response curves in Figs. 2 and 4 can be considered approximately as transfer functions for the power spectrum of the original signal. Since, for the high-order polynomials, these transfer functions are clearly nonintegrable in the high frequency limit, any amount of high frequency contamination can induce large errors even for comparatively smooth signals.

Consider the case of white noise, bandlimited to frequencies below ν. Using equation (11) as the transfer function, the error induced by contamination from noise with an r.m.s. amplitude ε would be

$$\sigma_U^2 = \int^{\nu} \varepsilon^2 \omega^n \, d\omega / \nu \sim \varepsilon^2 \nu^n, \tag{13}$$

which is valid for the unbalanced polynomials, or

$$\sigma_B^2 \sim \varepsilon^2 \nu^{\lfloor n/2 \rfloor}, \tag{14}$$

which is based on equation (12) and applies to the balanced ones. Some simulation results are given in Fig. 5, which presents r.m.s. interpolation errors for a smooth sine wave ($\alpha = 30$), both in the absence of noise, and with the addition of uniformly distributed white noise with a peak-to-peak amplitude of 1% of that of the base signal. For the noise-free data, the interpolation error decreases exponentially with the order of interpo-
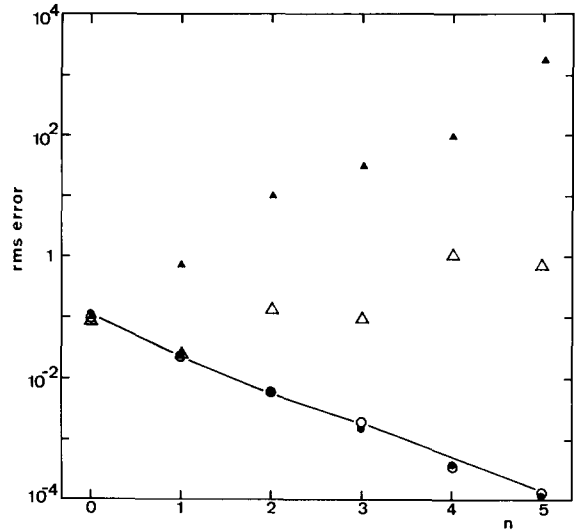

Fig. 5. Root mean square interpolation error for a smooth sine wave ($\alpha = 30$) vs. the order of interpolation. Circles are the results in the absence of noise. Triangles are the results when 1% white noise is added. Open symbols stand for balanced interpolation, closed for unbalanced interpolation. The straight line is the result of equation (A.6).

lation, as expected from the results in Appendix A. In the noisy case, however, and even with this relatively small amount of noise, the only polynomial behaving better than the simple nearest neighbour interpolation is the linear balanced one. In agreement with equations (13) and (14), the slope of the errors for unbalanced polynomials is roughly twice that for the balanced ones.

The previous discussion only gives a general estimate for the behaviour of the error with n. The prediction of its absolute magnitude requires an analysis similar to the one given in the previous section for the high frequency signals. For unbalanced interpolation and $n = 1$, this is equivalent to the evaluation of the integral (7) for white noise. As for continuous functions, some sort of high frequency cut-off is needed to get a finite result. For analog signals, this is provided by the frequency response of the instrument while, for computer processed data, it is bounded, in any case, by the finite resolution of the time discretisation. In general, we have to evaluate the integral for $|x_1 - x_0| > 1/\nu$, where ν is the highest frequency

in the noise. To leading order, if $\nu \gg \lambda$,

$$\sigma_1^2 = 2\lambda^2 \int_0^\infty dx_1 \, x_1^2 \, e^{-2\lambda x_1}$$

$$\times \int_{1/\nu}^{x_1} d\xi \, \langle (f_1 - f_0)^2 \rangle / \xi^2$$

$$\sim \varepsilon^2 \nu / 2\lambda, \tag{15}$$

which applies to unbalanced linear interpolation. In the simulations in Fig. 5, the finite precision of the computer fixes $\nu \sim 10^8$, which would result in $\sigma_1 \sim 0.6$, in fair agreement with the experimental result.

6. Convolution windows

From the discussion in the previous sections, it is clear that low-order balanced polynomials are an attractive possibility for the interpolation of randomly sampled functions. As we saw in Section 1, the classical alternative is the use of convolution windows, in which the value at x is approximated by

$$z(x) = \frac{\sum w(x_i - x) f_i}{\sum w(x_i - x)}. \tag{16}$$

This scheme is easy to implement and robust, since the high frequency components are essentially averaged out by the smoothing window and, as a consequence, the errors introduced by noise are generally small. Its behaviour depends on the shape and width of the window. Wide windows have good noise rejection properties, but they tend to filter out some of the interesting frequencies of the signal. Narrow ones, on the other hand, tend to select just those samples which are nearest to the point of interpolation and, in that respect, they are similar to the interpolation methods. In fact, as the width of the window approaches $1/\lambda$, the only point used by equation (16) is the nearest one, and the result approaches that of the $n = 0$ polynomials.

A full analysis of the statistics of the errors of expression (16) is difficult because of the nonlinear dependence in $w(x_i)$ and, to check its performance

with respect to the polynomials approximations, we run a few numerical experiments in which (16) was tested against the nearest neighbour and linear balanced interpolators. The conditions of the experiments were identical to the one in the previous sections, and the results are given in Fig. 6. In each case, we tested the several window shapes and varying widths.

Wide windows just averaged the signal and gave zero as the interpolated value. The resulting r.m.s. error was equal to the r.m.s. amplitude of the signal, and corresponds in Fig. 6 to the horizontal boundary of the shaded area. The best approximations were obtained with relatively narrow windows, which are presented in this figure by the curved boundaries of the shaded area. The region between these two limits could be reached by intermediate choices of the convolution window.

Note that, for smooth functions, estimator (16) approaches the performance of the nearest neighbour approximation, which is given in the figure by the dotted and dashed line. At the high frequency limit, the narrow windows, which are good for smooth functions, have slightly worse noise rejection properties than the wide ones, but only by approximately 10%. The cross-over point between the two regimes (best narrow and wide
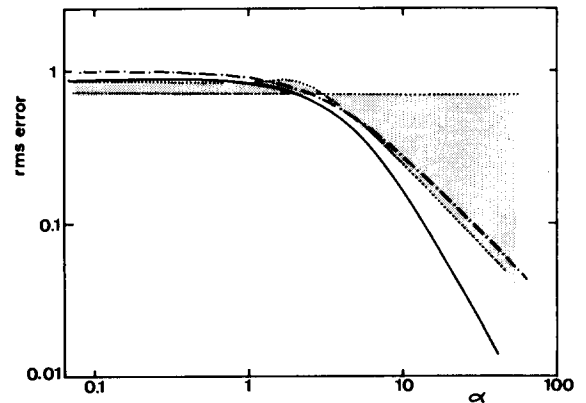

Fig. 6. Root mean square error of convolution windows on equation (4). The shaded area is the error attainable with the convolution estimator (16). The solid line is linear balanced polynomial interpolation; the dotted and dashed line is nearest neighbour interpolation.

windows) happens to be very near to the Nyquist limit, $\alpha = 2$.

The solid line in Fig. 6 represents the error from the linear balanced interpolation. Note that it outperforms all the other estimates in all ranges, except at the high frequency limit, in which its behaviour is similar to all the other estimates. In fact, in practical applications, it seems to be the most convenient estimator.

It might be possible to design an optimum convolution window with better approximation properties than the polynomial interpolators. The errors given in Fig. 6, however, seem to be fairly independent of the detailed shape of the window used in the convolution. The most obvious improvement, which was not applied in the experiments for that figure, would be to use windows whose shape and width adapt to the *local* sampling frequency. All the windows used in Fig. 6 were constant, independent of position, and could be considered as being too wide for the sampling rate at some locations and too narrow for others. The interpolation, on the other hand, is intrinsically adaptive to the local distribution of sampling points, which probably accounts for its better overall performance.

It is interesting to consider here the computational complexity associated to both convolution and interpolation. Convolution windows are theoretically of infinite extent but, in practice, are limited to some range around their centre. This limit depends on the window but, in any case, it should be wide enough to make sure that at least one point is always included inside it. For Poisson sampling, a good rule is that the window should span at least 10 to 15 average sampling intervals. The problem of identifying which samples fall inside the window is roughly equivalent to that of finding the closest points to be used of interpolation, and will not be considered here. The convolution itself needs $O(N)$ multiplications and additions, where N is the number of samples inside the window, plus N evaluations of the weighting function. For most windows, the last step is the most expensive part of the computation.

On the other hand, interpolation by an nth-order polynomial needs only $O(n^2)$ operations, which is a very small number in the preferred cases of $n = 0$ or $n = 1$. No window evaluation is needed in this case.

7. Conclusions

Summarising the discussion up to now, we have shown that low-order 'balanced' interpolation polynomials provide a very convenient method for the approximate reconstruction of signals from their Poisson samples. The choice of the order of interpolation depends on the amount of noise expected. Randomly sampled signals cannot be considered a priori as bandlimited, and the behaviour of any interpolation method with regard to signal frequencies above the Nyquist limit has to be considered explicitly.

Constant $(n = 0)$ or linear $(n = 1)$ polynomials are fairly insensitive to those frequencies and can be used in all cases. They are computationally efficient and their approximation properties match, and in many cases outperform, those of classical (nonadaptive) convolution windows, especially in the linear case. The errors introduced are tolerable for frequencies three or four times lower than the Nyquist limit corresponding to the average sampling rate.

Quadratic or cubic balanced polynomials have moderate noise rejection properties and give substantially better approximations at low frequencies. They should probably be used on signals known to be very clean or very densely sampled. Higher-order polynomials are useless in most cases due to their extreme sensitivity to noise.

Finally, a word should be said about the two-dimensional problem which was the original motivation for this work. The two-dimensional analog for a segment bracketing the interpolation point is a triangle containing it, and the equivalent of balanced linear interpolation is the approximation of the function by a set of planar triangular tiles. This

idea has been used for a long time in the production of two-dimensional graphics and in finite element analysis, and has been extended to higher-order methods using two-dimensional splines based on the triangular regions. There are also efficient ways of producing optimal triangular tesselations of the plane, using arbitrary point sets. Two published algorithms that generate both triangulations and interpolated values can be found in [1, 16]. A review of some of the problems associated with this approach in the particular field of experimental fluid mechanics can be found in [11].

Appendix A. Approximation of smooth functions

The problem is to estimate the statistical properties of the errors committed in approximating, at $x = 0$, a smooth function by an nth-order Lagrange polynomial based on $n + 1$ points extracted from a Poisson population with an average distance $1/\lambda$ between consecutive points. Following the arguments in the body of this paper, this is equivalent to estimating the statistics of the product

$$Q_n = x_0 x_1 \cdots x_n \tag{A.1}$$

of the abscissae of the $n + 1$ points used to compute the polynomial. We shall examine separately the methods of point selection which are referred to as 'balanced' and 'unbalanced' in the body of this paper.

A.1. Unbalanced approximation

In this method, the ith point is just the ith closest point to the origin. As discussed in this paper, each point has an equal probability of having either sign. Therefore, we shall treat the abscissae in equation (A.1) as positive numbers, remembering in any case that they represent the absolute values of signed quantities. The probability density that x_0 lies between ξ and $\xi + d\xi$ is the probability that no point lies in the interval $(-\xi, \xi)$ multiplied by the probability that one point lies in $(\xi, \xi + d\xi)$,

$$p_1(\xi) = 2\lambda \, e^{-2\lambda\xi}. \tag{A.2}$$

Note that a factor of two has been added to account for the possibility of both signs. In the same way, the probability density function for x_i, conditional to x_{i-1}, is

$$p_i(\xi \mid x_{i-1}) = \begin{cases} 0 & \text{when } \xi < x_{i-1}, \\ 2\lambda \, e^{-2\lambda(\xi - x_{i-1})} & \text{otherwise.} \end{cases} \tag{A.3}$$

The compound probability that a given point set lies in the volume $d\xi_0 \, d\xi_1 \cdots d\xi_n$ is then

$$(2\lambda)^{n+1} e^{-2\lambda\xi_n} d\xi_0 \, d\xi_1 \cdots d\xi_n,$$

as long as $\xi_0 \leqslant \xi_1 \leqslant \cdots \leqslant \xi_n$, and zero otherwise. The qth moment of the absolute value of the product Q_n can be expressed as the integral

$$U_{nq} \equiv \langle Q_n^q \rangle = (2\lambda)^{n+1} \int_0^\infty \xi_n^q \, e^{-2\lambda\xi_n} \, d\xi_n \int_0^{\xi_n} \xi_{n-1}^q \, d\xi_{n-1} \cdots \int_0^{\xi_1} \xi_0^q \, d\xi_0, \tag{A.4}$$

which can be evaluated from right to left with the result

$$U_{nq} = (nq + n + q)! / [n!(q + 1)^n (2\lambda)^{(n+1)q}]. \tag{A.5}$$

In particular, the variance of the error is given by

$$\sigma_n^2 = (3n + 2)! f^{(n+1)}(0)^2 / [n!(n + 1)!^2 3^n (2\lambda)^{2(n+1)}]. \tag{A.6}$$

We have not found an easy way of computing the exact probability density function for the (absolute value) of the product Q, but an estimate of its behaviour for large error amplitudes can be deduced from the asymptotic properties of its high-order moments. Using Stirling's formula to approximate equation (A.5) for $q \gg 1$ we get, after some manipulation,

$$U_{nq} \sim \frac{(n+1)^n (2\lambda)^{-(n+1)q}}{n!} [\Gamma(nq+q+1) - \tfrac{1}{2}n(n+1)\Gamma(nq+q) + \cdots]. \qquad (A.7)$$

This value can be related to the probability density function $p_n(Q)$ for the absolute value of Q, using directly the definition of the moments,

$$U_{nq} = \int_0^\infty Q^q p_n(Q) \, dQ. \qquad (A.8)$$

When $q \gg 1$, the dominant part in the integrand is the range $Q \gg \sigma_n$, and we can estimate its behaviour by equating (A.7) to (A.8) and using the definition of the Gamma function. The result is

$$p_n(Q) \sim \frac{(2\lambda)^n (n+1)^{n-1}}{n!} Q^{-n/(n+1)} e^{-2\lambda Q^{1/(n+1)}} \left[1 - \frac{n(n+1)}{4\lambda} Q^{-1/(n+1)} + \cdots \right]. \qquad (A.9)$$

This approximation is only valid for large values of Q, but it can be used to estimate the probability of committing an error above a certain threshold. If we express the threshold as a multiple of the standard deviation, we get the results in Fig. A.1. This figure shows that, for high-degree polynomials, most of the mass of the histogram is concentrated below one standard deviation, but that very large errors become more and more probable. Table A.1 is the inverse of Fig. A.1, and shows the multiple of σ_n that has to be included to be sure that the probability of error is smaller than a certain fraction. This table confirms
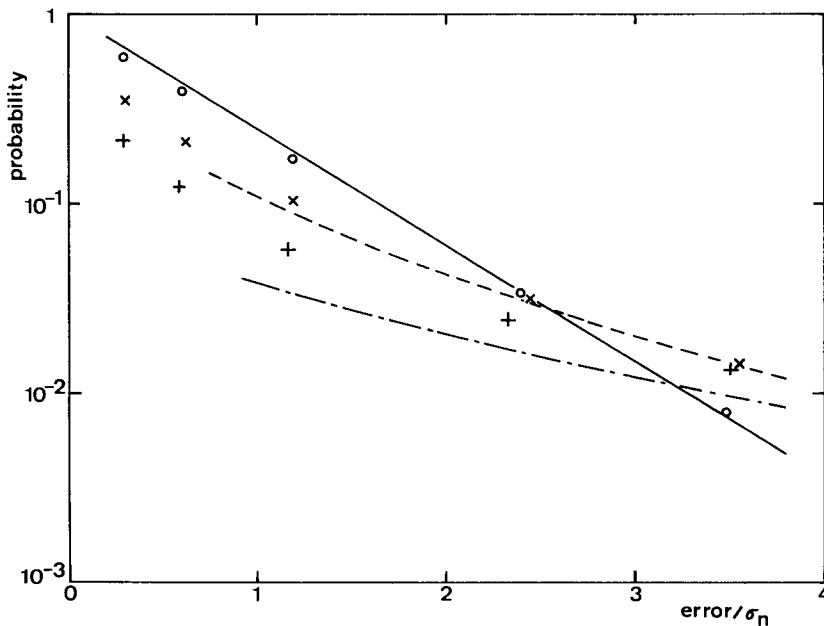

Fig. A.1. Probability that actual interpolation error is larger than a given fraction of its root mean square value. Unbalanced interpolation. Lines are derived from equation (A.9), and symbols from the simulation experiments. $\bigcirc$ and ——— : $n = 0$; $\times$ and — — : $n = 1$; $+$ and — $\cdot$ — : $n = 2$.

Table A.1

Multiple of the root mean square error that has to be included to be sure that the probability of error is smaller than a certain fraction; empty entries are the range in which the theoretical estimate is not reliable

Error probability	Order of interpolation			
	0	1	2	3
10^{-1}	1.62	1.11	–	–
10^{-2}	3.25	4.13	3.45	–
10^{-3}	4.88	8.83	10.82	9.107
10^{-4}	6.51	15.20	23.81	28.08
10^{-5}	8.14	23.25	43.99	63.69

the results in the figure; for relatively large error probabilities (~ 0.1) the high-order polynomials are better, but as soon as the error conditions are made more restrictive, they loose their advantage to the coarser but more robust low-order interpolations. It should be remembered, however, that these results are normalised with the standard deviation, σ_n, and that, for very smooth functions, the value of σ_n decreases strongly with n, so that the absolute behaviour of the high-order polynomials is always better in that limit.

A.2. Balanced approximation

It is convenient, in this case, to treat the points to the right of $x = 0$ differently from those to the left. Call the abscissa of the ith point to the right a_i, and to the left $-b_i$. It is clear, from the same arguments used in the previous section, that the probability density function for a_i is the same as the one in (A.3), but with 2λ substituted by λ to take into account that only one sign is now possible. Moreover, when the number of points in the product Q is even, half of the points are positive and half are negative, and their contributions to the qth moment are equal (except for the sign), resulting in a formula equivalent to (A.5):

$$U_{2n+1,q} = (-1)^{(n+1)q} \left[\frac{(nq+n+q)!}{n!(q+1)^n \lambda^{(n+1)q}} \right]^2. \tag{A.10}$$

In particular, the bias is

$$B_{2n-1} = (-1)^n (2n-1)! f^{(2n)}(0) / [n!(n-1)! 2^{2n-1} \lambda^{2n}], \tag{A.11}$$

and the root mean square error is

$$\sigma_{2n-1} = (3n-1)! |f^{(2n)}(0)| / [(2n)!(n-1)! 3^{2n-1} \lambda^{2n}]. \tag{A.12}$$

The ratio of these two numbers for the first few values of n is given in Table A.2. It is clear that, for low-order polynomials with an even number of points, the magnitude of the bias can be of the same order as the expected standard deviation, but that it decreases faster with increasing order.

The case with an odd number of points is slightly more complicated, since we need to take into account the two different possibilities that the next point is chosen positive or negative. The probability distribution of the $(2i+1)$st point depends on whether a_i is larger or smaller than b_i but, since both cases are equally probable, we can assume any one of them, say $a_i \geqslant b_i$, and multiply the final moments by a factor of two. With this assumption, it is easy to see that the probability density function for the absolute value of x_{2i+1},

Table A.2

Dimensionless bias due to balanced polynomial
interpolations of different odd orders

n	$2n-1$	$|\text{Bias}/\sigma|_{2n-1}$
1	1	0.500
2	3	0.255
3	5	0.100
4	7	0.045

conditional to a_i and b_i, is

$$p_{2i+1}(\xi \,|\, a_i, b_i) = \begin{cases} 0 & \text{when } \xi < b_i, \\ \lambda\, e^{-\lambda(\xi-b_i)} & \text{when } \xi \leqslant a_i, \\ 2\lambda\, e^{-\lambda(2\xi-a_i-b_i)} & \text{otherwise.} \end{cases} \tag{A.13}$$

The qth moment can then be evaluated to the rather complicated expression

$$U_{2n,q} = \frac{(nq+n)!}{n!^2 (q+1)^{2n} 2^{nq+n+q} \lambda^{(2n+1)q}} \sum_{j=0}^{nq+n} \frac{(nq+n+q+j)!}{2^j j!}. \tag{A.14}$$

The standard deviation derived from this formula is reflected in the theoretical estimates used in Fig. 4. From reasons of symmetry, the interpolation based on an odd number of points should be unbiased, which is confirmed by the simulation experiments.

An argument similar to the one used for the unbalanced interpolation can be used to estimate the behaviour of the error probability for large error magnitudes. The final formulas are more complicated in this case, but the general conclusions are the same; the probability of finding errors larger than Q falls like $\exp\{-2\lambda Q^{1/(n+1)}\}$ for an nth-order interpolating polynomial. Therefore, also in this case, the low-order polynomials are much more robust than the high-order ones.

References

[1] H. Akima, "A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points", *ACM Trans. Math. Softw.*, Vol. 4, 1978, pp. 148-159; also: "Algorithm 526", *ibid.*, pp. 160-164.

[2] F.J. Beutler, "Error-free recovery of signals from irregularly spaced samples", *SIAM Rev.*, Vol. 8, 1966, pp. 328-335.

[3] F.J. Beutler, "Alias-free randomly timed sampling of stochastic processes", *IEEE Trans. Inform. Theory*, Vol. IT-16, 1970, pp. 147-152.

[4] R.E. Crochiere and L.R. Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[5] P.E. Dimotakis, F.D. Debussy and M.N. Koochesfahani, "Particle streak velocity field measurements in a two-dimensional mixing layer", *Phys. Fluids*, Vol. 24, 1981, pp. 995-999.

[6] L.J. Fogel, "A note on the sampling theorem", *IRE Trans. Inform. Theory*, Vol. IT-1, 1955, pp. 47-48.

[7] M. Gaster and J.B. Roberts, "Spectral analysis of randomly sampled signals", *J. Inst. Math. Appl.*, Vol. 15, 1975, pp. 195-216.

[8] K. Imaichi and K. Ohmi, "Numerical processing of flow-visualisation pictures: Measurement of two-dimensional vortex flow", *J. Fluid Mech.*, Vol. 129, 1983, pp. 283-311.

[9] E. Isaacson and H.B. Keller, *Analysis of Numerical Methods*, Wiley, New York, 1966, pp. 187-192.

[10] D.L. Jagerman and L.J. Fogel, "Some general aspects of the sampling theorem", *IRE Trans. Inform. Theory*, Vol. IT-2, 1956, pp. 139-146.

[11] J. Jimenez, "Some computational problems in computer-assisted flow visualisation", *Proc. Internat. Symp. on Computational Fluid Dynamics*, Tokyo, September 1985, pp. 145-156.

[12] E. Masry, "Random sampling and reconstruction of spectra", *Inform. and Control*, Vol. 19, 1971, pp. 275-288.

[13] A. Papoulis, "Error analysis in sampling theory", *Proc. IEEE*, Vol. 54, 1966, pp. 947-955.

[14] E.I. Plotkin, L.M. Roytman and M.N.S. Swamy, "Nonuniform sampling of bandlimited modulated signals", *Signal Processing*, Vol. 4, 1982, pp. 295-303.

[15] E.I. Plotkin, L.M. Roytman and M.N.S. Swamy, "Reconstruction of nonuniformly sampled bandlimited signals and jitter error reduction", *Signal Processing*, Vol. 7, 1984, pp. 151-160.

[16] R.J. Renka, "Triangulation and interpolation at arbitrarily distributed points in the plane", *ACM Trans. Math. Softw.* (Algorithm 624), Vol. 10, 1984, pp. 440-442.

[17] C.E. Shannon, "Communication in the presence of noise", *Proc. IRE*, Vol. 37, 1949, pp. 10-21.

[18] R.G. Wiley, "Recovery of bandlimited signals from unequally spaced samples", *IEEE Trans. Commun.*, Vol. COM-26, 1978, pp. 135-137.

[19] J.L. Yen, "On nonuniform sampling of bandwidth-limited signals", *IRE Trans. Circuit Theory*, Vol. CT-3, 1956, pp. 251-256.