

SAMPLING AND INTERPOLATION OF
STRONG MOTION ACCELEROGRAMS

Arturo Arias^I and Horacio Sandoval^{II}

SUMMARY

An analysis is made of the influence of spacing between sampled points and the interpolation procedure adopted on the accuracy with which Fourier spectra can be recovered from data obtained by mechanical- or optical-recording accelerographs and seismographs.

An interpolation scheme, proposed originally by Maude, is studied in connection with its potential applications to numerical processing of records digitized either manually or by semi-automatic means.

INTRODUCTION

The number of strong-motion accelerographs recording on paper or photographic film now in operation around the world can be estimated to be in the order of two thousand. In spite of the advent of instruments yielding direct digitized records on magnetic tape, it is likely that the number of conventional instruments will not decrease in the near future. The problems that arise when converting "analog" records to their digital version will be with us for some years to come; therefore, it is justified to analyze at some length the consequences of the operations by which such a conversion is performed. The present paper is devoted to the study of some problems related with sampling and interpolation of "analog" accelerographic records. Most of the conclusions are of general validity, and can be applied to other kind of "analog" records.

When digitizing an accelerogram recorded on paper or photographic film, the operator consecutively selects points of the accelerographic trace so that, to his judgment, the polygon having these points as vertexes is an accurate representation of the recorded curve. The result is a time series with values at non-uniformly spaced points of the domain, the spacing being closer where the record departs more significantly from a straight line. Usually, the digitized points include all maximae and minimae and as many intermediate points as it is deemed necessary to get a close reproduction of the trace. For a given number of sampled points, it is plausible to expect that non-uniform sampling will give better accuracy than sampling at equal intervals of time. Conversely, it should be expected that, for a given accuracy, non-uniform sampling will require a smaller number of digitized points.

Some digitizing machines are designed for uniform sampling (Naumovski et al., 1976); however, as the timing marks of the original record are not

(I) Research Engineer, Instituto de Ingeniería, and Prof. of Earthquake Engineering, DEPEFI, UNAM, Mexico

(II) Research Engineer, Instituto de Ingeniería, UNAM

distributed at strictly equal intervals of length, some interpolation procedure will be required to account for this lack of uniformity when preparing the final version of the accelerogram as a discrete series at equal steps in the time domain.

In the numerical processing of accelerograms, it has become current practice to admit that, between consecutive sampled points, the record can be represented by a segment of a straight line. This assumption will be called polygonal approximation. With this assumption as basis, the digitized ("uncorrected") version is subjected to some corrective operations (time or abscissae correction, base line correction, instrument correction, and filtering). A final, "corrected" version of the accelerogram is produced by linear interpolation between the corrected positions of the vertices of the polygonal approximation. This practice introduces spurious components in the high-frequency range which are due to: a) lack of continuity of the first and higher derivatives of the polygonal approximation at each one of the sampled points, b) the discrete character of the representation, and c) the substitution of curved archs by segments of straight lines.

Non-linear interpolation procedures have been employed for processing strong-motion accelerograms, as well as seismograms. Some of these procedures are based on classical interpolation formulae (such as Lagrange's); others make use of the method of splines in some of its several forms. However, for the numerical processing of earthquake records, it has been found that interpolatory schemes not based on the theories of interpolation and approximation of functions can lead to better results. Because of this lack of a strict theoretical justification, these interpolation schemes will be called heuristic. To this class belong, among others, the procedures proposed by Akima (1970), by Wiggins (1976), and by Maude (1973). All of them are local interpolators, in the sense that interpolated values depend only on a few of the neighbouring sampled values. Wiggins's procedure was proposed specifically for its application to earthquake records. Akima's procedure, though originally intended for a quite different purpose, has been applied to seismograms with satisfactory results (Griffiths and Prieto-Díaz, 1977). The procedure proposed by Maude is herein examined regarding its applicability to earthquake records and it is found to yield results that are highly satisfactory.

UNIFORM VERSUS NON-UNIFORM SAMPLING

Let a large number of cycles of a sinusoidal curve be sampled at equal intervals of the abscissa. The corresponding polygonal approximation can be expanded as a (complex) Fourier series, provided that the number of sampled points per cycle is an integer (it suffices that the wave length and the sampling step have a common multiple). The amplitude of the lowest frequency term of this expansion compared by quotient with the amplitude of the original sinusoid is, by definition, the modulus of the transfer function of the polygonal approximation, for the frequency of the original wave and the selected sampling rate. It depends, of course, on the relative position or phase difference of the sinusoid and the "net" of sampling abscissae. If the number of sampled points per cycle is $2n$, the modulus of the transfer function will be maximum either when one of the sampling points coincides with a peak of the sinusoid, if n is odd, or when a peak of the sinusoid is

located at the mid-point of the interval between two consecutive sampling points, if n is even. These two cases will be called in-phase uniform sampling. The transfer function for in-phase uniform sampling has been evaluated by Trifunac et al. (1971, 1973) and is represented by the curve of Fig. 1 which is a freehand trace joining the points obtained for 2, 4, 6, ... samples per cycle. This curve represents the optimum situation that can be reached by uniformly sampling a sinusoid with an even number of sampled points per cycle.

For $n=1$ (two points per cycle), the modulus of the transfer function for in-phase uniform sampling is $8/\pi^2 = 0.81056\dots$. In this case the sinusoid is approximated by a triangular wave with turning points at the peaks and valleys of the original curve. It may then be concluded that, at the Nyquist frequency, the polygonal approximation allows recovery of the Fourier amplitude spectrum with a relative error by defect not smaller than $100(1 - 8/\pi^2) \approx 19\%$. If sampling is uniform but not in-phase, the relative error will be larger and always by defect; for $n=1$ (two points per cycle) the modulus of the transfer function varies with phase difference from zero to $8/\pi^2$. If the phase difference is assumed to be randomly distributed in the interval $(0, 2\pi)$, the expected value of the modulus of the transfer function is $16/\pi^3 = 0.516\dots$, i.e., one should expect that at the Nyquist frequency the Fourier amplitude spectrum is recovered from the polygonal approximation with an error by defect which on the average is $100(1 - 16/\pi^3) \approx 48\%$.

Some improvement can be obtained within the frame of polygonal approximation if uniform sampling is replaced by a non-uniform sampling strategy that selects the sampled points so as to follow the variations of curvature of the sinusoid. For $n=1$ no improvement is possible. For $n=2$, (four points per cycle) the best choice of the sampled points gives 0.9216 for the modulus of the transfer function at the frequency of the original wave; the corresponding result for $n=3$ (six points per cycle) is 0.9572. These results are represented in Fig. 1 by small circles. The improvement over in-phase uniform sampling is very slight.

Trifunac et al. report an experiment in which several cycles of a sinusoidal curve were digitized by optical means at different non-uniform sampling rates. The corresponding moduli of the transfer function of the polygonal approximation are shown in Fig. 1 by triangular dots. Contrary to what it could be expected, these moduli are consistently smaller than those obtained for the case of uniform in-phase sampling. Therefore, unless the sampling strategy is carefully chosen, the polygonal approximation can lead to significant errors by defect in the Fourier amplitudes at the high-frequency range. If the average sampling rate \bar{f}_s is fixed, and a desired accuracy in the Fourier ordinates is chosen, then there exists an upper frequency limit, f_{\max} , below which the required accuracy can be attained, and above which it cannot be reached with the polygonal approximation. For example, let $\bar{f}_s = 50$ points/sec, and let the maximum allowable error in the Fourier ordinates be 5%; from Fig. 1 it follows that this accuracy can be attained if the number of points per cycle is at least equal to 6. Therefore, in this case, $f_{\max} = \bar{f}_s / 6 = 8.33$ Hz.

CRITERIA FOR A SAMPLING STRATEGY

According to Båth (1974, pp. 131-134), the maximum absolute deviation, $|\delta_y|_{\max}$, of the polygonal approximation from the true trace $y = f(x)$ is given by

$$(1) \quad |\delta_y|_{\max} \approx \frac{(\Delta x)^2}{3} |f''(y)|_{\max}$$

where Δx is the length of the interval between consecutive sampled points. It would be inferred that in order to put a uniform bound on $|\delta_y|$, the length Δx of each individual interval should be chosen to vary as $(|f''|_{\max})^{-1/2}$. However, Båth's derivation of Eq. 1 tacitly assumes that f'' does not vanish in the interval of interest and that the coordinates of the sampled points are known without error. These two issues have been examined with some detail by Arias and Sandoval (1979). The main results are as follows:

1. If the uncertainties in the abscissae are much smaller than Δx , then

$$(2) \quad |E[\delta_y]| = \frac{\bar{F}'' \cdot (\Delta x)^2}{4}$$

where \bar{F}'' is the average curvature of the true trace in the interval considered. For a segment of the trace having constant curvature, this is twice the value given by Båth's result.

2. If

$$(3) \quad \Delta x \geq (2\sigma_x)^{2/3} |\bar{F}'|^{1/3} |\bar{F}''|^{-1/3}$$

there exists an upper bound K for the expected value of the deviation δ_y , given by

$$(4) \quad |E[\delta_y]| < \frac{|\bar{F}''|}{8} \left[(\Delta x)^2 + \left(\frac{4\sigma_x^2 \bar{F}'}{\bar{F}''(\Delta x)^2} \right)^2 \right] = K$$

where σ_x^2 is the variance of the error of measurement of the abscissae, and \bar{F}' is the average slope of the trace in the interval considered.

3. The bound K reaches a minimum if the length of the interval is chosen to be

$$(5) \quad \Delta x = 2^{5/6} \left[\sigma_x^2 \frac{|\bar{F}'|}{|\bar{F}''|} \right]^{1/3}$$

With this choice

$$(6) \quad K_{\min} = \frac{3}{8} (4 \sigma_x^4 \bar{F}'^2 |\bar{F}''|)^{1/3}$$

From these results the following conclusions regarding sampling strategies adequate for polygonal approximation can be derived.

- a) To obtain a uniform bound on $|E[\delta_y]|$ sampling must be non-uniform. The sampling rate for segments of non-vanishing curvature should vary as $|\bar{f}''|^{1/3} |\bar{f}'|^{-1/3}$.
- b) Errors of measurement affecting the abscissae are specially undesirable in those segments of the trace where $(f'/f'')^2$ is large.
- c) The optimal sampling rate is proportional to $\sigma_x^2/3$; therefore, there is no advantage in using large sampling rates if the accuracy of the digitizing process is low.
- d) Δx should not be selected too small for those segments of the trace having large slope and small curvature. This conclusion is contrary to the usual recommendation requiring the sampling of all inflexion points. These should be sampled only when the tangent at the inflexion point is horizontal or nerly so.

AVERAGE SAMPLING RATE FOR A GIVEN ACCURACY

Let the overall resolution of the digitizing process be δ (cm); then the smallest wavelength that can be recovered from the digitized data is $\lambda_{\min} = 2\delta$ (cm). It is clear that the existence of a limit of resolution implies that measurement of the coordinates of sampled points is affected by error. Arias and Sandoval (1979) have been able to show that the coefficient of variation of λ_{\min} is

$$(7) \quad \text{C.V.}\{\lambda_{\min}\} = \frac{\sigma_x}{\delta} \sqrt{\frac{11}{6}}$$

From the results of the digitizing experiments reported by Trifunac et al., $\sigma_x = 1.41$ points of the scale of the digitizing machine. Therefore, if $\delta = 1$ point, as stated by those authors, then $\text{C.V.}\{\lambda_{\min}\} = 1.91$. This is, by any standard, a large value of the coefficient of variation. It follows that if it were possible to evaluate exactly the Fourier amplitude for frequencies equal to or nearly equal to f_{\max} , it would be impossible to know, with any fairly acceptable precision, the value of the frequency for which the spectral ordinate was evaluated.

Let $F(\omega)$ be the Fourier transform of the trace function $f(t)$, where $t = x/MV$ is the time coordinate ($MV =$ photographic magnification multiplied by instrument recording velocity). It can be shown (Arias and Sandoval, 1980) that, the Fourier transform $\tilde{F}(\omega)$ recovered from digitized data affected by error in the measurement of the abscissae of sampled points satisfies the relation

$$(8) \quad E[\tilde{F}(\omega)] = F(\omega) \exp\left(-\frac{\omega^2 \sigma_x^2}{2(MV)^2}\right)$$

The relative discrepancy r between $E[\tilde{F}(\omega)]$ and $F(\omega)$ is

$$(9) \quad r = \frac{F(\omega) - E[\tilde{F}(\omega)]}{F(\omega)} = 1 - \exp\left(-\frac{\omega^2 \sigma_x^2}{2(MV)^2}\right)$$

Therefore, if the allowable relative discrepancy is r_a , the frequency cannot exceed the value $f_{\max|r_a}$ given by

$$(10) \quad f_{\max|r_a} = \frac{MV}{2\pi\sigma_x} \sqrt{2\log_e\left(\frac{1}{1-r_a}\right)}$$

Now, for uniform sampling, the digitized version of the accelerogram contains information up to Nyquist's frequency, $f_N = f_s/2 = MV/2\Delta x$ where f_s is the sampling frequency (in Hz) and Δx is the sampling step (in cm). Therefore the sampling rate will be compatible with a given desired accuracy if $f_{\max|r} = f_N$, i.e., if

$$(11) \quad \Delta x = \frac{\pi\sigma_x}{\sqrt{2\log_e\left(\frac{1}{1-r_a}\right)}}$$

If a larger sampling step is used, more information will be lost. If the sampling step is chosen to be smaller the value given by Eq. 11, then the accuracy desired will not be attained in the range of frequency near to the Nyquist frequency.

For example, assuming that $MV = 2$ cm/sec and the required accuracy is 5%, the Fourier spectrum will be reliable at this level of accuracy up to a frequency of $2 \times 11.78 \approx 24$ Hz, and the sampling rate required to achieve the accuracy prescribed is $2 \times (2 \times 11.78) \approx 47$ points/sec, for uniform sampling.

NON-LINEAR INTERPOLATION

As stated before, non uniform sampling offers some possibilities of obtaining better results than uniform sampling for the same average sampling rate. However, these possibilities are not exploited to their full extent if one persists in adhering to linear interpolation. Much better results can be obtained if non uniform sampling is combined with some non-linear interpolation scheme.

Maude's interpolation scheme. The interpolation procedure proposed by Maude (1973) is based on the idea of replacing the original function in each interval between sampled points by a weighted mean of two polynomials, where the weight is a function of the abscissa whose derivative vanishes at the two end points of the interval. One of the simplest choices is to use two quadratic polynomials to interpolate the function $f(x)$ in the interval (x_n, x_{n+1}) ; the first one, F_n , adjusted to $f(x)$ at the points x_{n-1}, x_n, x_{n+1} and the second one, F_{n+1} , adjusted to $f(x)$ at the points x_n, x_{n+1}, x_{n+2} . The weighted mean or interpolating function $F(x)$ is then defined as

$$(12) \quad F(x) = w(x) F_n(x) + [1 - w(x)] F_{n+1}(x)$$

where $w(x)$ is a monotonic twice differentiable decreasing function satisfying the restrictions

$$(13) \quad \begin{aligned} 0 &\leq w(x) \leq 1 && \text{in } x_n \leq x \leq x_{n+1} \\ w(x_n) &= 1 \\ w(x_{n+1}) &= 0 \\ w'(x_n) &= w'(x_{n+1}) = 0 \end{aligned}$$

There are several choices possible for $w(x)$. One of the simplest is to employ a polynomial. The polynomial of lowest degree satisfying the above conditions is

$$(14) \quad w(X) = 1 - 3X^2 + 2X^3 = (1-X)^2 (1 + 2X)$$

where

$$(15) \quad X = \frac{x - x_n}{x_{n+1} - x_n}$$

This is the weighting function proposed by Maude, who remarks that with this choice, $w(X)$ has the following symmetry property

$$(16) \quad w(1-X) = 1-w(X)$$

Most of the properties of Maude's interpolation scheme are consequences of Eqs. 12, 13 and 15 and do not depend on the particular form of Eq. 14.

Some of these properties are listed below.

1. From Eqs. 12 and 13 it follows that the interpolating function $F(x)$ and its first and second derivatives are continuous at each intermediate sampled point.
2. From the preceding statement it follows that the Fourier transform of $F(x)$ tends to zero when $|\omega| \rightarrow \infty$ at least as fast as $|\omega|^{-2}$, if no special care is taken with continuity conditions at the first and last end points, and as fast as $|\omega|^{-3}$ if by some artifice continuity of the first and second derivatives at the first and last end point is imposed.
3. The average value of $F''(X)$ in $0 \leq X \leq 1$ is equal to the arithmetic mean of $F''(X)$ and $F''_{n+1}(X)$. This result does not depend on the particular form of the weighting function w as long as it satisfies the symmetry property of Eq. 16.
4. The graph of the interpolating function $F(X)$ intersects the straight line joining the n th and $(n+1)$ th sampled point at most once in the open interval $x_n < x < x_{n+1}$. This property insures that the interpolator does not introduce high-frequency components of significant amplitude.
5. If the original function has an extremum (or an inflexion point) in (x_n, x_{n+1}) then the interpolating function will also have an extremum (or an inflexion point) inside the same interval.

Several tests with Maude's interpolator are reported in Arias and Sandoval (1979).

REFERENCES

- Akima, H. (1970). "A new method of interpolation and smooth curve fitting based in local procedures" Jour. Ass. Comp. Mach., 17, (4), 589-602
- Arias, A., and H. Sandoval (1979) "Muestreo e interpolación de acelerogramas obtenidos en instrumentos de registro óptico o mecánico, Instituto de Ingeniería, Publ. No 416 UNAM, México (52 pp)
- Arias, A., and H. Sandoval (1980) "Intervalo confiable en el espectro de Fourier obtenido de acelerogramas digitizados manualmente o con medios semi automáticos", Instituto de Ingeniería, UNAM (in press)
- Bath, M. (1974) "Spectral Analysis in Geophysics", Elsevier, Amsterdam.
- Griffiths, L.J., and R. Prieto-Díaz (1977) "Spectral analysis of natural events using autoregressive technique", IEEE Trans. Geoscience Electronics, Vol GE-15(1).
- Maude, D. D. (1973) "Interpolation—Mainly for graph plotters" Comp. Jour., 16 (1)
- Naumowski, N., et al. (1976) "Strong-motion earthquake accelerograms. Digitized and plotted data" Inst. Earthq. Engng. and Engng. Seism. Univ. Skopje, Yugoslavia.
- Trifunac, M.D., Udawadia, F.E, and A. G. Brady (1971) "High frequency errors and instrument corrections of strong-motion accelerograms" EERL 71-05, Caltech, Pasadena, Cal.
- Trifunac, M. D. and V. Lee "Routine computer processing of strong motion accelerograms", EERL 73-03, Caltech, Pasadena, Cal.
- Wiggins, R. A. (1976) "Interpolation of digitized curves" Bull. Seism. Soc. Am., 66 (6).

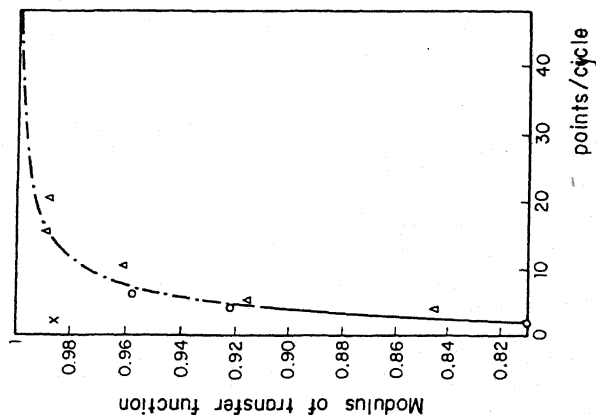


Fig 1