RANDOM DIGITIZATION ERRORS AND RELIABILITY OF FOURIER SPECTRA OF STRONG-MOTION ACCELEROGRAMS

Arturo Arias[I] and Horacio Sandoval[II]

SUMMARY

The effects of random digitization errors on Fourier amplitudes of strong-motion accelerograms is studied. A frequency interval in which Fourier spectra are reliable is found to depend on the length of the record and the standard error of digitization. The lower frequency bound is related with uncertainties in the definition of the reference line, while the upper frequency bound arises from uncertainties in measurements of the time coordinate.

INTRODUCTION

Bogert, Healy and Tukey (1963) report an experiment in which log-energy spectra of two versions of the same seismogram are compared; the first version was retrieved directly from an electronically digitized record while the second was obtained by digitizing manually a plot of the original electronic record. The sampling interval was the same in both cases (0.1 sec). The authors do no report the length of the record. Considerable discrepancies between the two versions is found at frequencies above 0.6 f_N; coherency for frequencies over 0.3 f_N is meager (f_N = Nyquist's frequency = 5 Hz, in this case).

Kulhánek and Klīma (1970) have given estimates of a reliable frequency interval for the application of instrumental correction. With this purpose they employed the function

$$f(t) = At^{\alpha} e^{-\beta t} \operatorname{sen} \omega_o t$$

where, A, α, β, ω_o are constants (Berlage pulse) as input of a linear filter that simulates the recording instrument. The output of the filter was plotted by means of a semi-automatic analog-digital converter and compared with the result of an analytical evaluation of the output. The length of the digitized output was 20 sec and the sampling rate 20 points/sec. Significant discrepancies between the two versions are observed at frequencies exceeding 0.25 f_N. In the low frequency range, the digitized version departs significantly from the analytical result for periods larger than 7 sec 0.35 (length of the record).

Linde and Sacks (1971) have studied some of the difficulties involved in Fourier analysis of body waves recorded by standard seismographs. They found that background noise and truncation of the record to eliminate later arrivals introduce large spurious amplitudes at the low frequency end of the spectrum. When corrected for low instrumental response these spurious

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

amplitudes may become dominating in the region of low frequencies of the estimated spectrum. Linde and Sacks remark that simple elimination of the mean or the use of a new straight reference line, to comply with the requirement of zero net area, are not enough to eliminate the low-frequency errors.

Manzoni (1967) has made a theoretical study of the effect of random errors of measurement of the abscissae on the power spectrum of a sinusoidal wave, assuming that the density function of the error is even.

Trifunac (1970), and Trifunac, Udwadia and Brady (1971, 1973) have examined the problem of random digitization errors in strong-motion accelerograms. An important conclusion of their research is that digitization errors of the ordinates are approximately normally distributed, with zero mean and a standard deviation sensibly equal to the limit of resolution of the process of digitization. According to these authors accelerographic data digitized by semi-automatic means are reliable in the frequency range 0.06 - 25 Hz.

Hanks (1975) found that for certain commercially available strong-motion accelerographs, the lower frequency bound has to be raised to 0.125 Hz if reliable estimates of ground displacements obtained by double integration are desired.

The present paper contains a theoretical study of the effects of random digitization errors on the accuracy with which Fourier spectra can be recovered from accelerograms (or seismograms) digitized either manually or by semi-automatic means.

STATEMENT OF THE PROBLEM

Let f(t) be a function that vanishes identically outside the interval $(-\frac{T}{2}, \frac{T}{2})$ and assume that it has been sampled at the abscissae t_k $(k = 0, 1, 2, \ldots, N)$. Due to errors of measurement the point $P(t_k, f_k)$ (where $f_k = f(t_k)$) is replaced by $P'(t_k + \underset{\sim}{\varepsilon}_k, f_k + \underset{\sim}{\eta}_x)$.

When the record is placed on the table of the digitizing machine, its reference line does not coincide with the time-axis of the digitizer, that is to say the position of the record is defined up to a rigid body displacement. Therefore, the (inexact) result of measurements on the coordinates of sampled points can be regarded as the result of exact measurements on the coordinates of sampled points of a function $\tilde{f}(t)$ defined by

$$(1) \qquad \tilde{f}(t) = \underset{\sim}{f}(t) + \underset{\sim}{\eta}(t) + \underset{\sim}{\ell}(t) \quad p(-\frac{T}{2}, \frac{T}{2})$$

where $p(-\frac{T}{2}, \frac{T}{2})$ is the unit pulse in the interval $(-\frac{T}{2}, \frac{T}{2})$ and

$\underset{\sim}{f}(t)$ is the true value of the ordinate at the instant t, to which, as a consequence of digitization errors in the abscissa, an abscissa $t + \underset{\sim}{\varepsilon}$ is assigned.

The term $\underset{\sim}{\eta}(t)$ represents the digitization error on the ordinates.

Finally $\underset{\sim}{\ell}(t) = \underset{\sim}{a}t + \underset{\sim}{b}$, represents the rigid body displacement mentioned above. Here $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are stochastic variables independent of t (or independent of the index k for the discrete case).

It will be assumed that $\varepsilon(t)$ and $\eta(t)$ are stationary stochastic processes; therefore, $\varepsilon_k = \varepsilon(t_k)$ and $\eta_k = \eta(t_k)$ are stochastic variables

Let $\tilde{F}(\omega)$ be the Fourier transform of f(t). To obtain statistics of $\tilde{F}(\omega)$ the following assumptions will be introduced

a) ε_k has gaussian density with zero mean and variance σ^2 indepent of k

(2) $E\left[\varepsilon_k\right] = 0 \quad , \quad E\left[\varepsilon_k^2\right] = \sigma^2$

b) The errors of digitization of the abscissae of any two different sampled points are orthogonal, i. e.,

(3) $E\left[\varepsilon_j \ \varepsilon_k\right] = 0 \qquad j \neq k$

c) η_k has gaussian density with zero mean and variance σ_f^2 indepent of k

(4) $E\left[\eta_k\right] = 0 \quad , \quad E\left[\eta_k^2\right] = \sigma_f^2$

d) η_j, η_k $(j \neq k)$ are orthogonal, i. e.,

(5) $E\left[\eta_j \ \eta_k\right] = 0 \qquad j \neq k$

e) ε_j, η_k are orthogonal

(6) $E\left[\varepsilon_j \ \eta_k\right] = 0$

The effects of the three terms inside the square brackets of Eq. 1 will be analyzed independently.

DIGITIZATION ERRORS ON THE TIME COORDINATE

The ordinate $f(t_k)$ has been assigned an abscissa $t_k + \varepsilon_k$. Therefore, an application of the trapezoidal rule of integration gives

$$\tilde{F}(\omega) = \frac{1}{2}\sum_{k=0}^{N-1}\left[f(t_k)e^{-i\omega(t_k+\varepsilon_k)} + f(t_{k+1})e^{-i\omega(t_{k+1}+\varepsilon_{k+1})}\right](t_{k+1}-t_k+\varepsilon_{k+1}-\varepsilon_k)$$

from which

$$E\left[\tilde{F}(\omega)\right] = E\left[e^{-i\omega\varepsilon}\sum_{k=0}^{N-1}\frac{1}{2}\left(f(t_k)e^{-i\omega t_k} + f(t_{k+1})e^{-i\omega t_{k+1}}\right)(t_{k+1}-t_k)\right]$$

$$= F(\omega)\ E\left[e^{-i\omega\varepsilon}\right]$$

But

$$E\left[e^{-i\omega\varepsilon}\right] = \int_{-\infty}^{\infty}\frac{e^{-\frac{\varepsilon^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\cos\omega\varepsilon\ d\varepsilon = \exp\left(-\frac{\omega^2\sigma^2}{2}\right)$$

So finally

$$(7) \qquad E\left[\tilde{F}(\omega)\right] = F(\omega)\ \exp\left(-\frac{\omega^2\ \sigma^2}{2}\right)$$

Thus, if only the errors in the time coordinate are considered, Fourier amplitudes obtained from digitized data will be in error by defect by a factor $\exp(-\frac{\omega^2\ \sigma^2}{2})$ as compared with true amplitudes. In other words, $E\left[\tilde{F}(\omega)\right]$ is a biased estimator of $F(\omega)$.

Let us define the relative discrepancy r between

$E\left[\tilde{F}(\omega)\right]$ and $F(\omega)$ by the relation

$$(8) \qquad r = \left| \frac{E\left[\tilde{F}(\omega)\right] - F(\omega)}{F(\omega)} \right|$$

Then, from Eq. 7

$$(9) \qquad r = 1 - \exp\left(-\frac{\omega^2\ \sigma^2}{2}\right)$$

Therefore, the relative discrepancy will be smaller than a given value ρ if ω satisfies the inequality

$$(10) \qquad |\omega| < \omega_\rho = \frac{1}{\sigma}\sqrt{2\ell n\left(\frac{1}{1-\rho}\right)}$$

(Compare Manzoni, 1967, p 252, Eq. 24).

If $\rho << 1$, Eq. 10 can be written as

$$(11) \qquad |\omega| < \omega_\rho \simeq \frac{\sqrt{2\rho}}{\sigma}$$

By a similar procedure it can be shown that

$$(12) \qquad E\ \{(\tilde{F}(\omega) - E\left[\tilde{F}(\omega)\right])(\tilde{F}^*(\omega) - E\left[\tilde{F}^*(\omega)\right])\}$$
$$= |F(\omega)|^2 \left[1 - \exp(-\ \omega^2\ \sigma^2)\right]$$

where asterisks stand for conjugate complex. Therefore, besides a reduction in the expected values of the Fourier amplitude, the variance of Fourier amplitudes increases with increasing $|\omega|$.

If σ is known, the bias of $E\ \tilde{F}(\omega)$ can be corrected defining a new estimator $\tilde{\tilde{F}}(\omega)$ as

$$(13) \qquad \tilde{\tilde{F}}(\omega) = \tilde{F}(\omega)\ \exp(\omega^2\ \sigma^2/2)$$

(14) $$E\left[\tilde{\tilde{F}}(\omega)\right] = F(\omega)$$

and

(15) $$E\left\{\left(\tilde{\tilde{F}}(\omega) - E\left[\tilde{\tilde{F}}(\omega)\right]\right)\left(\tilde{\tilde{F}}^*(\omega) - E\left[\tilde{\tilde{F}}^*(\omega)\right]\right)\right\} = |F(\omega)|^2(\exp(\omega^2\sigma^2) - 1)$$

It is then concluded, that notwithstanding that $F(\omega)$ is estimated without bias by $E\left[\tilde{\tilde{F}}(\omega)\right]$, the variance of this new estimator will grow without bound as $|\omega|$ tends to infinity. The coeficient of variation, (C.V.), of $F(\omega)$ can be expressed as

(16) $$(C.V.)^2 = \exp(\omega^2\sigma^2) - 1$$

Consequently, if it is desired not to exceed a given (C.V.), ω must satisfy the inequality

(17) $$|\omega| \leq \omega_{(C.V.)} = \frac{1}{\sigma}\sqrt{\ell n\left[1 + (C.V.)^2\right]}$$

When (C.V.) $\ll$ 1 this relation takes the approximate form

(18) $$|\omega| \leq \omega_{(C.V.)} \approx \frac{(C.V.)}{\sigma}$$

It is possible to express the above results in terms of a quantity R^2 indistinctly called "coherence" or "coherence squared", which is directly proportional to the square of the coefficient of correlation between two stochastic variables and is equal to the square of the coefficient of correlation between the spected variables (Tick, 1963, p. 200). Herein we shall use the symbol R^2 in the sense given by the following relation

(19) $$R^2 = \frac{\left(E\left[e^{i\omega(\varepsilon' - \varepsilon'')}\right]\right)^2}{E\left[e^{i\omega\varepsilon'}\right] E\left[e^{i\omega\varepsilon''}\right]}$$

where ε' and ε'' are equidistributed, gaussian, orthogonal stochastic variables, with zero mean and variance σ^2.

From Eqs. 16 and 19 it follows that

(20) $$R^2 = \exp(-\omega^2\sigma^2) = \frac{1}{1+(C.V.)^2}$$

Therefore, the coherence will be larger than a given value R provided that ω satisfies the inequality

(21) $$|\omega| \leq \omega_R = \frac{1}{\sigma}\sqrt{\ell n\,\frac{1}{R^2}}$$

DIGITIZATION ERRORS ON THE ORDINATES

The term $\underset{\sim}{\eta}(t)$ in Eq. 1 is essentially "white noise" Therefore the expected energy spectrum of $\underset{\sim}{\eta}(t)$ will be practically a constant in the frequency interval $\frac{2\pi}{T} \leq |\omega| \frac{\pi}{\Delta t}$, where Δt is the sampling interval (assumed for simplicity to by constant). The expected energy spectrum of $\underset{\sim}{\eta}(t)$ can be estimated from Parseval's relation as follows

$$(22) \qquad E|M(\omega)|^2 \approx \sigma_f^2 \, T \, \Delta t = N \, \sigma_f^2 (\Delta t)^2 = \frac{\sigma_f^2 T^2}{N} = \frac{\sigma_f^2 T}{f_s}$$

where N+1 is the number of sampled points $f_s = \frac{1}{\Delta t}$ is the sampling rate, which for simplicity has been considered here to be uniform.

The term $\underset{\sim}{\ell}(t)$ of Eq. 1 arises from the lack of exact coincidence of the reference line with the axis of abscissae of the digitizing machine. Usually the digitized data are employed to correct for this lack of coincidence: a a provisional straight line of reference is determined so that the integral of the squares of the ordinates refered to this straight line is a minimum. There are some variants in this procedure according to whether the record has or has not a "fix" trace. It will be assumed here that both $\underset{\sim}{a}$ and $\underset{\sim}{b}$ have been determined from the original data, e.e., that the record has been translated and rotated as a rigid body to refer it to the provisional base line. In current practice, it is admitted that the parameters of the provisional base line are exactly determined by this operation. However, it cannot be overlooked that really they are stochastic variables dependent on the errors of digitization.

Assuming that the provisional base line has been already determined from the original data as stated above the ordinates still contain an error $\underset{\sim}{\ell}(t) = \underset{\sim}{a} + \underset{\sim}{b}t$. It can be shown that

$$(23) \qquad \begin{array}{l} E[\underset{\sim}{a}] = 0 \;, \quad E[\underset{\sim}{b}] = 0 \;, \quad E[\underset{\sim}{a}\,\underset{\sim}{b}] = 0 \\ E[\underset{\sim}{a}^2] = \dfrac{\sigma_f^2}{N} \;, \quad E[\underset{\sim}{b}^2] = \dfrac{12\,\sigma_f^2}{NT^2} \end{array}$$

where N is the number of sampled points. Therefore the Fourier spectrum $L(\omega)$ of $\underset{\sim}{\ell}(t)$ satisfies the following relations

$$(24) \qquad E[L(\omega)] = 0$$

$$(25) \qquad E\left[|L(\omega)|^2\right] = \frac{\sigma_f^2 T^2}{N} \left\{ \frac{\text{sen}^2(\frac{\omega T}{2})}{\frac{\omega T}{2}} + \frac{3}{(\frac{\omega T}{2})^2} \left(\cos\frac{\omega T}{2} - \frac{\text{sen}\,\frac{\omega T}{2}}{\frac{\omega T}{2}}\right)^2 \right\}$$

The fraction R_o of the total expected energy of $\underset{\sim}{\ell}(t)$ contained in the interval $(-\Omega, \Omega)$

is found to be

$$(25') \qquad R_o = \frac{2}{\pi}\left\{Si(\Omega T) - \frac{2(1-\cos \Omega T) - 2\Omega T \operatorname{sen}\Omega T + (\Omega T)^2 (2-\cos\Omega T)}{(\Omega T)^3}\right\}$$

where $Si(\Omega T)$ is the sine integral funtion.

Employing an asymptotic expansion of this function it is found that

$$(26) \qquad R_o \sim 1 - \frac{4}{\pi\Omega T}$$

It is concluded that the record corrected for provisional base line contains a fraction R_o of the total energy of $\underset{\sim}{\ell}(t)$ in the range of periods larger than

$$(27) \qquad \tau \sim \frac{\pi^2}{2} T(1 - R_o)$$

For example if $R_o = 0.95$, this asymptotic approximation gives

$$\tau \sim \frac{0.05\ \pi^2}{2} T \simeq \frac{T}{4}$$

This conclusion has an important bearing on the design of filters to obtain the final version of the "corrected" accelerogram. For example, if it is desired to eliminate at least 95% of the low-frequency noise arising from digitizing errors, the filter must reject all components whose periods are longer than one quarter of the length of the record.

If the components of $\underset{\sim}{\ell}(t)$ due to uncertainties in $\underset{\sim}{a}$ (translation of the record) and $\underset{\sim}{b}$ (rotation of the record) are examined separately, it is found that the fractions of the expected energy $\underset{\sim}{\ell}(t)$ contained in the interval $(-\Omega,\Omega)$ are given (asymptoticaly) by

$$(29) \qquad R_1 \sim 1 - \frac{2}{\pi\Omega T}$$

for translation, and by

$$(29) \qquad R_2 \sim 1 - \frac{6}{\pi\Omega T}$$

for rotatiom. (Incidentally, $R_o = \frac{1}{2}(R_1 + R_2)$.)

Therefore, the uncertainties in the low frequency range due to uncertainties in the rotation of the record with respect to the true base line are more critical than those due to translation.

It may then be concluded that in the calculation of the provisional base line it is sound practice to include both parameters a and b, ever in the case when the record has a "fix" trace.

CONCLUSIONS

1. Digitization errors affecting the time coordinate have, on the average, a depressing affect on the Fourier amplitudes at high frequencies without affecting the Fourier phase.

2. If the variance of the error of digitization in the time coordinate is known, the bias on the estimates of the Fourier spectrum can be accounted for defining an unbiased estimator as in Eq. 13

3. If the correction mentioned in the preceding conclusion is not applied, then Eq. 10 gives a high-frequency limit for a given desired accuracy of the Fourier ordinates. If the correction is applied, the limiting frequency is given by Eq. 17.

4. Digitization errors affecting the ordinates of sampled points give rise to two kinds of uncertainties in the Fourier ordinates. One of them is essentially white-noise; the other one is due to uncertainties in the parameters defining the provisional reference line of the accelerogram. In what concerns the first of these components there is nothing much to be done, except careful digitization and the selection of a relatively high sampling rate.

5. The second component mentioned in the preceding conclusion specially affects the quality of the estimated spectrum in the low frequency range. This source of error is specially annoying for periods larger than about one third or one quarter of teh length of the record.

6. Uncertainties in the rotation component of the position of the record, as referred to a "true" base line, are more critical than uncertainties in the translation component. Therefore both parameters of position should be included in the calculation of the provisional base line, even in the case that the original record has a "fix"trace.

Additional results and applications as well as detailed derivations can be found in Arias and Sandoval (1980).

REFERENCES

Arias, A., and H. Sandoval (1980) "Intervalo confiable en el espectro de Fourier obtenido de acelerogramas digitizados manualmente o con medios semiautomáticos" Instituto de Ingeniería, UNAM Mexico (in press)

Bogert, B. P., Healy, M.J.R., and J. W. Tukey (1963) Proc. Symp. on Time Series Analysis, Brown University, June 11-14, 1962. Chap. 15, M. Rosenblatt (ed.) Wiley, N. York and London.

Hanks, T. C. (1975) Bull. Seism Soc. Am. 65 (1), p. 193-225

Kulhánek, O., and K. Klíma (1970) Geophys, Jour. Roy. Astr. Soc. 21, p. 235-242

Linde, A. T., and I.S. Sacks (1971) J. Geophys. Res., 76 (14) O. 3326-3336.

Manzoni, G. (1967) Bull. Geofis. Teor. ed Appl., Vol IX, No. 35.

Tick, L. J. (1963) Proc. Symp. on time Series analysis, Brown University, June 11-14, 1962. Chap. 13, Wiley, N. York and London.

Trifunac, M. D. (1970) EERL 70-07, Caltech, Pasadena

Trifunac, M. D., Udwadia, F. E., and A. G. Brady (1971) EERL 71-05 Caltech. Pasadena

——— (1973) Bull. Seism. Soc. Am., 63 (1), p. 157-187.