The Measurement of Power Spectra from the Point of View of Communications Engineering — Part II

By R. B. BLACKMAN and J. W. TUKEY

(Manuscript received August 28, 1957)

The measurement of power spectra is a problem of steadily increasing importance which appears to some to be primarily a problem in statistical estimation. Others may see it as a problem of instrumentation, recording and analysis which vitally involves the ideas of transmission theory. Actually, ideas and techniques from both fields are needed. When they are combined, they provide a basis for developing the insight necessary (i) to plan both the acquisition of adequate data and sound procedures for its reduction to meaningful estimates and (ii) to interpret these estimates correctly and usefully. This account attempts to provide and relate the necessary ideas and techniques in reasonable detail. Part I of this article appeared in the January, 1958 issue of THE BELL SYSTEM TECHNICAL JOURNAL.

PART I — GENERAL DISCUSSION

CONTINUOUS RECORDS OF FINITE LENGTH

EQUALLY SPACED RECORDS

Details and Derivations

In this part we will reconsider some of the earlier analysis, either in greater detail, or from alternative points of view. We shall assume familiarity with the material on fundamental Fourier techniques presented in Appendix A of Part I.

The sections of this part will be numbered in exact correspondence with the sections of the general account. Thus, for example, Section B.7, below, presents the details and sidelights related to Section 7, of Part I. (Certain sections will be omitted.)

B.1 *Gaussian Processes and Moments*

There are two common modes of description of a random process, intuitively quite different. One uses the idea of an ensemble, the other a function of infinite extent. The first is undoubtedly more flexible, as it can describe processes, even non-stationary (e.g. evolving) ones, which cannot be described by any single function, even one of infinite extent. The first is also, at least in the eyes of the statistician, more fundamental, since uncertainty, which he regards as a central concept, enters directly and explicitly. It is possible to regard the single-function approach as an attempt to minimize recognition of the statistical aspects of the situation. Once, such minimization may have been of some value, but today the essentially statistical nature of communication, be it of symbols, voice, picture or feedback information, is well established. The communication engineer is aware that he must have designed not only for the message which was sent, but also for the one which might have been sent — moreover that his design demanded consideration of the relative probabilities of various messages that might have been sent (and were not).

Such a statistical view of message or noise confronts us with the need, not only of picking out what functions might arise, but also of attaching probabilities to functions (at least to sets of functions). To do this directly and completely requires much careful mathematics. As far as questions associated with observations and data are concerned, there is,

fortunately, no need for such care and complexity. We know that any empirical time function can be adequately represented by some finite number, large or small, of ordinates. Thus, for practical purposes, it suffices to be able to assign probabilities to sets of n ordinates — for n finite but possibly quite large. It is for this reason that we went directly to probability distributions of such n-dimensional sections in the general account.

This replacement of a continuous record by discrete ordinates is related to the sampling theorem of information theory. The relationship is, regrettably, not quite simple. Given a band-limited "signal" defined for all time from $-\infty$ to $+\infty$, and moderately well-behaved otherwise, the sampling theorem (Nyquist[23]), which is also known as the Cardinal Theorem of Interpolation Theory (Whittaker[24]), states that equi-spaced ordinates, if close enough together, extending from $-\infty$ to $+\infty$ will precisely determine the function. Given a band-limited function over a finite interval, the corresponding result is almost practically true. It is not true in a precise impractical sense, since every band-limited function can be obtained from an entire function of exponential type (of a complex variable) by considering only the values taken on along the real axis. Consequently, if we know a band-limited function precisely in an interval, its values are determined everywhere. Theoretically determined, but not practically so, since the kernels expressing this determination behave like hyperdirective antennas. Since the values at equi-spaced points in the interval do not determine the values at equi-spaced points outside the interval (which would be determined by precise knowledge throughout the interval) the latter cannot be obtained from the equi-spaced values in the interval. In practice, however, functions are not quite band-limited, and measurements always involve measurement noise. When these two facts are considered, a sufficiently closely spaced set of equi-spaced ordinates extracts all the practically useful information in the continuous record.

The fact that averages, variances and covariances completely characterize any n-dimensional Gaussian distribution is common statistical knowledge, and follows by inspection of the conventional general form of Gaussian probability density function, in which these moments appear as the only parameters.

B.2 *Autocovariance Functions and Power Spectra for Perfect Information*

If $X(t)$ is a function generated by a stationary Gaussian process, and if

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t)\, dt = 0 \qquad \text{(B-2.1)}$$

then the *autocovariance function* of the process is

$$C(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t) \cdot X(t + \tau)\, dt. \qquad \text{(B-2.2)}$$

In particular, the *variance* of the process is

$$C(0) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [X(t)]^2\, dt. \qquad \text{(B-2.3)}$$

If the function $X(t)$ is passed through a fixed linear network whose impulse response (response to a unit impulse applied at $t = 0$) is $W(t)$, then the output of the network will be

$$X_{\text{out}}(t) = \int_{-\infty}^{\infty} W(\lambda) \cdot X(t - \lambda) \cdot d\lambda\,,$$

and the autocovariance of the output will be

$$C_{\text{out}}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(\lambda_1) \cdot W(\lambda_2) \cdot X(t - \lambda_1) \cdot X(t + \tau - \lambda_2) \cdot d\lambda_1 \cdot d\lambda_2 \cdot dt$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(\lambda_1) \cdot W(\lambda_2) \cdot C(\tau + \lambda_1 - \lambda_2) \cdot d\lambda_1 \cdot d\lambda_2.$$

If we now let

$$C(\tau) = \int_{-\infty}^{\infty} P(f) e^{i\omega\tau}\, df \qquad (\omega = 2\pi f), \quad \text{(B-2.4)}$$

then

$$C_{\text{out}}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(\lambda_1) \cdot W(\lambda_2) \cdot P(f) \cdot e^{i\omega(\tau+\lambda_1-\lambda_2)} \cdot d\lambda_1 \cdot d\lambda_2 \cdot df.$$

But

$$\int_{-\infty}^{\infty} W(\lambda_2) \cdot e^{-i\omega\lambda_2} \cdot d\lambda_2 = Y(f)$$

is the transfer function (ratio of steady-state response to excitation, when the excitation is $e^{i\omega t}$) of the network, and

$$\int_{-\infty}^{\infty} W(\lambda_1) \cdot e^{i\omega\lambda_1} \cdot d\lambda_1 = Y(-f)$$

is the complex-conjugate of $Y(f)$. Hence,

$$C_{\text{out}}(\tau) = \int_{-\infty}^{\infty} |\, Y(f)\,|^2 \cdot P(f) \cdot e^{i\omega\tau} \cdot df\,. \qquad \text{(B-2.5)}$$

In particular, the variance of the output is

$$C_{out}(0) = \int_{-\infty}^{\infty} | Y(f) |^2 \cdot P(f) \cdot df. \qquad \text{(B-2.6)}$$

Since $| Y(f) |^2$ is the power transfer function of the network, it is natural to call $P(f)$ the *power spectrum* of the process. The power spectrum has the dimensions of variance per cycle per second. By equation (B-2.4) we have

$$P(f) = \int_{-\infty}^{\infty} C(\tau) \cdot e^{-i\omega\tau} \, d\tau , \quad (\omega = 2\pi f), \qquad \text{(B-2.7)}$$

and, by equation (B-2.5),

$$P_{out}(f) = | Y(f) |^2 \cdot P(f).$$

Autocovariance functions and power spectra are usually regarded as one-sided functions of lag and frequency, respectively, related by the formulae (Rice[3] p. 285),

$$P(f) = 4 \int_0^{\infty} C(\tau) \cdot \cos \omega\tau \cdot d\tau , \qquad \text{(not used here),}$$

$$C(\tau) = \int_0^{\infty} P(f) \cdot \cos \omega\tau \cdot df , \qquad \text{(not used here).}$$

However, we will find it very convenient for analytical purposes, to continue to regard them as two-sided even functions related by equations (B-2.4) and (B-2.7). This will be evident in Section B.4 where spectral windows will be convolved with power spectra, and in Section B.6 where we would otherwise have to make use of rather complicated trigonometric identities.

In a few places where we contemplate computations, we may write

$$P(f) = 2 \int_0^{\infty} C(\tau) \cdot \cos \omega\tau \cdot d\tau , \qquad (\omega = 2\pi f),$$

instead of equation (B-2.7), but it is important to observe that the power spectrum $P(f)$ thus obtained must still be regarded as a two-sided even function which contains only a half of the total power or variance in the positive frequency range. If we prefer to think ultimately of a one-sided power spectrum (over only positive frequencies), in accordance with engineering practice, then we should take $2P(f)$ as the power density (per cycle per second) for positive frequencies.

The power spectrum $P(f)$ may also be expressed directly in terms of $X(t)$. The formal derivation of this expression on the basis of the formula (B-2.2) for the autocovariance function is complicated by the fact that the integral

$$\int_{-T/2}^{T/2} X(t) \cdot X(t + \tau)\, dt$$

actually depends upon two more-or-less distinct pieces of $X(t)$, one in the range $-T/2 < t < T/2$, the other in the range

$$-\frac{T}{2} - \tau < t < \frac{T}{2} - \tau.$$

We may avoid this complication by using the equivalent formula

$$C(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} G(t) \cdot G(t + \tau) \cdot dt, \qquad \text{(B-2.8)}$$

where

$$G(t) = X(t), \qquad |t| < \frac{T}{2},$$

$$= 0, \qquad |t| > \frac{T}{2}.$$

Let

$$S(f) = \int_{-\infty}^{\infty} G(t) \cdot e^{-i\omega t}\, dt,$$

so that

$$G(t) = \int_{-\infty}^{\infty} S(f) \cdot e^{i\omega t}\, df.$$

Then

$$C(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} G(t) \cdot \int_{-\infty}^{\infty} S(f) \cdot e^{i\omega(t+\tau)}\, df \cdot dt$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} S(f) \cdot e^{i\omega\tau} \int_{-\infty}^{\infty} G(t) \cdot e^{i\omega t}\, dt \cdot df$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} S(f) \cdot S(-f) \cdot e^{i\omega\tau}\, df.$$

Comparing with (B-2.4) we get, at least formally,

$$\begin{aligned} P(f) &= \lim_{T\to\infty} \frac{1}{T} \mid S(f) \mid^2 \\ &= \lim_{T\to\infty} \frac{1}{T} \left| \int_{-\infty}^{\infty} G(t) \cdot e^{-i\omega t}\, dt \right|^2 \\ &= \lim_{T\to\infty} \frac{1}{T} \left| \int_{-T/2}^{T/2} X(t) \cdot e^{-i\omega t}\, dt \right|^2. \end{aligned} \tag{B-2.9}$$

(See Rice,[3] p. 320, and Bennett,[25] p. 621.)

B.3 *The Practical Situation*

In the study of second moments of random processes, the balance between the approach through autocovariances and the approach through power spectra is, in at least one sense, a little closer than Section 3 would seem to imply. As a means for understanding, and as a guide for intelligent design, the power spectrum is without a peer. The autocovariance function is of little use except as a basis for estimating the power spectrum. This is fundamentally because, in most physical systems, power spectra have reasonable shapes, are relatively easily understandable, and often are quite directly influenced by the basic variables of the situation, whatever these may be. The process of using an empirically observed and analyzed power spectrum usually goes through some such chain of steps as this:

(i) Planning and design.

(ii) Observation and recording.

(iii) Analysis and preparation of estimates.

(iv) Comparison of estimates with existing and synthesizable theoretical structures and quantitative information.

(v) Selection of the best working version of a theoretically-guided approximation to the estimates.

(vi) Use of this working version.

Our theoretical understanding of the situation, and of the forces in it, play important roles, which we should never allow ourselves to forget, in steps (i) and (v). (We use this understanding to the utmost, in its proper place, but we do not, and should not, allow it to narrow down steps (ii) and (iii) to the point where we have little or no chance of discovering that it was incomplete or in error. Thus, we estimate a considerable number of smoothed spectral densities, and not merely a few constants of a suggested theoretical curve. After we have compared curve and points, we may then wish to estimate the constants.)

The simplest and most straightforward use we can make of the power spectrum is to predict the output spectrum, or perhaps only the output power, when the process studied provides the input to a linear device with a known power transfer function.

Except for purely descriptive uses, checking on performance for agreement with anticipation, and for predicting the behavior of already designed linear systems, the most elementary use of a power spectrum lies in optimizing the performance of some linear predictor or filter as measured á la least squares. The nature of this situation is not quite that one which most persons imagine.

If we really have no theoretical insight into the situation *at all*, we might as well (nay, perhaps, might better) stay in the time domain. We have autocovariances (say) for some limited range of lags. If the duration of the transient response of our filter or predictor is not going to exceed one-half this time limit, then we can write out the estimated variance of any predictor or filter directly in terms of our estimates of autocovariances and of the time description of the filter or predictor, and could then minimize this directly. With no theoretical insight this should work at least as well as any other way. With no theoretical insight, analysis in the time domain would be relatively good, perhaps even optimal — and probably absolutely poor.

But we do not, and almost always should not, optimize filters or predictors in this way. The reason is simple. Actual power spectra are often simple and understandable. Actual autocovariance functions are hardly, if ever, simple and understandable. The intervention of theoretical insight and human judgment at step (v) is crucial and valuable. This intervention is effective in the frequency domain, but not in the time domain. (Step (v) is likely to stand for some time, as a challenge to the ability of statisticians to wisely and effectively automatize inferential procedures.)

An additional advantage of power spectrum analysis over autocovariance analysis was pointed out in Section 3, namely the ease of compensation for (linear) modification before measurement. When the random function $X(t)$ is passed through a time-invariant linear transmission system whose impulse response is $W(t)$, the output random function, which may be the only function accessible for measurement, is

$$X_{\text{out}}(t) = \int_{-\infty}^{\infty} W(\tau_1) \cdot X(t - \tau_1)\, d\tau_1.$$

The relation between the autocovariance $C_{\text{out}}(\tau)$ of the modified process,

and the autocovariance $C(\tau)$ of the original process may be derived as follows:

$$\begin{aligned} C_{out}(\tau) &= \text{ave}\,\{X_{out}(t)\cdot X_{out}(t+\tau)\} \\ &= \text{ave}\left\{\iint_{-\infty}^{\infty} W(\tau_1)\cdot X(t-\tau_1)\cdot X(t+\tau-\tau_2)\cdot W(\tau_2)\,d\tau_1\,d\tau_2\right\} \\ &= \iint_{-\infty}^{\infty} W(\tau_1)\cdot C(\tau+\tau_1-\tau_2)\cdot W(\tau_2)\,d\tau_1\,d\tau_2. \end{aligned}$$

Putting $\tau_2 = \tau_1 + \lambda$, we get

$$\begin{aligned} C_{out}(\tau) &= \int_{-\infty}^{\infty} C(\tau-\lambda)\cdot\left[\int_{-\infty}^{\infty} W(\tau_1)\cdot W(\tau_1+\lambda)\,d\tau_1\right]d\lambda \\ &= C(\tau) * W(\tau) * W(-\tau). \end{aligned}$$

Measurement of $X_{out}(t)$ can give estimates of $C_{out}(\tau)$ which must subsequently be converted into estimates of $C(\tau)$. The only practical way to make this conversion seems to be through Fourier transformation of the estimates of $C_{out}(\tau)$ into the frequency domain, compensation there, and Fourier retransformation. Such a procedure, in effect, invokes the relation between the power spectrum $P_{out}(f)$ of the modified process and the power spectrum $P(f)$ of the original process. This relation is

$$P_{out}(f) = P(f)\cdot |\,Y(f)\,|^2,$$

where $Y(f)$ is the transfer function corresponding to the impulse response $W(t)$.

Details for Continuous Analysis

B.4 *Power Spectrum Estimation from a Continuous Record of Finite Length*

In the ideal case considered in Section B.2, which assumes that we have an infinite length of $X(t)$, we can calculate the power spectrum $P(f)$ in two ways — either directly from $X(t)$, or indirectly as the Fourier transform of the autocovariance function $C(\tau)$ which is calculable directly from $X(t)$. The basic choice is, leaving limiting problems aside, between squaring a Fourier transform, or Fourier transforming an average of products. In either case multiplication and Fourier transformation must enter.

From the point of view of the ensemble, as opposed to the single time function, we seek to estimate a particular basis for the second moments

of all linear combinations. We may estimate any convenient basis as a start, and then transform.

Clearly, from either point of view, any result obtained in one way can also be obtained in the other. Differences between a time approach and a frequency approach must be differences in (i) ease of understanding, (ii) ease of manipulation of formulas, (iii) ease of calculation with numbers, rather than in anything more essential.

Understanding and a simple description of the procedure which yields reasonably stable estimates, and which we have discussed in general, is more easily obtained by the indirect route, so we shall proceed accordingly, beginning with a general outline of a hypothetical procedure for power spectrum estimation from a continuous record of finite length.

A general outline of a hypothetical procedure for power spectrum estimation from a continuous record of finite length, specifically $X(t)$ for $-T_n/2 \leqq t \leqq T_n/2$, is as follows:

(1) *Calculate the apparent autocovariance function*

$$C_{00}(\tau) = \frac{1}{T_n - |\tau|} \int_{-(T_n-|\tau|)/2}^{(T_n-|\tau|)/2} X\left(t - \frac{\tau}{2}\right) \cdot X\left(t + \frac{\tau}{2}\right) dt \qquad \text{(B-4.1)}$$

for $|\tau| \leqq T_m < T_n$, where T_n is the length of the record, and T_m is the maximum lag to be used. We shall see in Section B.9 that the stability of our power spectrum estimates depends upon how small we take the ratio T_m/T_n.

(For the purpose of the theoretical analysis in this section we assume that the data contain no errors of measurement; in particular, no bias due to a displaced (perhaps drifting) zero. The effects of such errors are considered elsewhere.)

(2) *Calculate the modified apparent autocovariance function*

$$C_i(\tau) = D_i(\tau) \cdot C_{00}(\tau), \qquad \text{(B-4.2)}$$

where $D_i(\tau)$ is a prescribed lag window, an even function such that

$$D_i(0) = 1,$$

and

$$D_i(\tau) = 0 \quad \text{for} \quad |\tau| > T_m.$$

Note that $C_i(\tau) = 0$ for $|\tau| > T_m$ although $C_{00}(\tau)$ is not available for $|\tau| > T_m$.

(3) *Calculate the estimated power spectrum*

$$P_i(f) = 2 \int_0^\infty C_i(\tau) \cdot \cos \omega\tau \cdot d\tau. \qquad \text{(B-4.3)}$$

Our object is now to determine the relation between ave $\{P_i(f)\}$ and $P(f)$, where "ave" denotes the ensemble average, that is, the average over all possible continuous pieces of $X(t)$ of length T_n. (The variability (specifically, the variance) of $P_i(f)$ will be examined in Section B.9.)

Since $C_{00}(\tau)$ is not calculated for $|\tau| > T_m$, it is clear that

$$\text{ave}\,\{C_{00}(\tau)\} = C(\tau), \qquad \text{only for } |\tau| \leq T_m.$$

However, because $D_i(\tau) = 0$ for $|\tau| > T_m$,

$$\text{ave}\,\{C_i(\tau)\} = D_i(\tau)\cdot C(\tau), \qquad \text{for any } \tau.$$

Hence,

$$\text{ave}\{P_i(f)\} = \int_{-\infty}^{\infty} D_i(\tau)\cdot C(\tau)\, e^{-i\omega\tau}\, d\tau.$$

Then, if $Q_i(f)$ is the Fourier transform of $D_i(\tau)$, the relation we seek is, symbolically,

$$\text{ave}\,\{P_i(f)\} = Q_i(f) * P(f),$$

or explicitly,

$$\text{ave}\,\{P_i(f_1)\} = \int_{-\infty}^{\infty} Q_i(f_1 - f)\cdot P(f)\, df. \tag{B-4.4}$$

This relation is in a form, (B-2.6), which is familiar to communications engineers except for the fact that $Q_i(f_1 - f)$ is not an even function of f, when $f_1 \neq 0$, and may be negative in some ranges of f. However, taking advantage of the fact that $P(f)$ is an even function of f, we may write

$$\text{ave}\{P_i(f_1)\} = \int_0^{\infty} H_i(f; f_1)\cdot P(f)\, df, \tag{B-4.5}$$

where

$$H_i(f; f_1) = Q_i(f + f_1) + Q_i(f - f_1). \tag{B-4.6}$$

The function $H_i(f; f_1)$ is an even function of f as well as of f_1. Hence it satisfies one of the necessary conditions for a physically realizable *power transfer function*. However, inasmuch as it may be negative in some ranges of f (actually an advantage as we will see), it may still not be physically realizable. Nevertheless, it is convenient to regard $H_i(f; f_1)$ as the power transfer function of a network, and to regard ave $\{P_i(f_1)\}$ as the long-time-average power output of the network when continuously driven by the random process.

Since $P_i(f_1)$, whether calculated from a single piece of $X(t)$ as outlined

above, or calculated as an average over a finite number of pieces of $X(t)$, is an estimate of ave $\{P_i(f_1)\}$ in the usual statistical sense, it is evident that the calculated power density $P_i(f_1)$ is an estimate of an average-over-frequency of the true power spectrum $P(f)$, and not an estimate of the local power density $P(f_1)$. The calculated power density $P_i(f_1)$ may be regarded as an estimate of the local power density $P(f_1)$ only to the degree to which $Q_i(f)$ approximates $\delta(f)$. However, under the restriction that $D_i(\tau) = 0$ for $|\tau| > T_m$, the degree to which $Q_i(f)$ approximates $\delta(f)$ depends chiefly on how large we take T_m. On the other hand, as we will find in Section B.9, the larger we take T_m/T_n the less stable will the estimates be. Hence, in general, it will be wasteful to demand more frequency resolution than we actually need. In many cases we may even have to take less frequency resolution than we would like to have, in order to secure a reasonable stability of the estimates. Clearly, for any specific value of T_n, (and number of pieces of record), we can increase frequency resolution (or stability) only by sacrificing stability (or frequency resolution).

We have just examined a hypothetical method of power spectrum estimation, in which we compute an apparent autocovariance function, modify it, and take the cosine transform. We will now examine a method, also hypothetical, in which we modify the data, take the sine and cosine transforms, compute the sum of the squares at each frequency, and divide by the length of the record. If the data is $X(t)$ for $0 < t < T_n$, and the weighting function (data window) is $B_i(t)$, the estimated power spectrum is computed essentially according to the formula

$$P_{ei}(f) = \frac{1}{T_n}\left|\int_0^{T_n} B_i(t)\cdot X(t)\cdot e^{-i\omega\tau}\,dt\right|^2. \qquad \text{(B-4.7)}$$

To determine the average it is convenient to assume that $X(t)$ is of unlimited extent, to specify $B_i(t)$ to be identically zero for $t < 0$ and $t > T_n$, and to allow the *data window* $B_i(t)$ to be located anywhere in time by substituting $B_i(t - \lambda)$ for $B_i(t)$. Then

$$P_{ei}(f;\lambda) = \frac{1}{T_n}\left|\int_{-\infty}^{\infty} B_i(t-\lambda)\cdot X(t)\cdot e^{-i\omega\tau}\,dt\right|^2. \qquad \text{(B-4.8)}$$

This is the Fourier transform of

$$C_{ei}(\tau;\lambda) = \frac{1}{T_n}\int_{-\infty}^{\infty} B_i(t-\lambda)\cdot X(t)\cdot B_i(t-\lambda+\tau)\cdot X(t+\tau)\cdot dt. \qquad \text{(B-4.9)}$$

[Compare with (B-2.8) and (B-2.9)]. Now, since the random process is

stationary, the ensemble average is equivalent to the average over λ; that is,

$$\text{ave } \{C_{ei}(\tau)\} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} C_{ei}(\tau;\lambda) \cdot d\lambda.$$

Substituting (B-4.9) and $\lambda = t - \xi$ into the right-hand member, we get

$$\text{ave } \{C_{ei}(\tau)\} = \frac{1}{T_n} \cdot \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} X(t) \cdot X(t + \tau) \cdot \left[\int_{t-(T/2)}^{t+(T/2)} B_i(\xi) \cdot B_i(\xi + \tau) \cdot d\xi \right] \cdot dt.$$

Reversing the order of integration we get

$$\text{ave } \{C_{ei}(\tau)\} = \frac{1}{T_n} \int_{-\infty}^{\infty} B_i(\xi) \cdot B_i(\xi + \tau) \cdot \left[\lim_{T\to\infty} \frac{1}{T} \int_{\xi-(T/2)}^{\xi+(T/2)} X(t) \cdot X(t + \tau) \cdot dt \right] d\xi.$$

The quantity in brackets is the true autocovariance function $C(\tau)$. Hence,

$$\text{ave } \{C_{ei}(\tau)\} = D_{ei}(\tau) \cdot C(\tau) \tag{B-4.10}$$

where

$$D_{ei}(\tau) = \frac{1}{T_n} \int_{-\infty}^{\infty} B_i(\xi) \cdot B_i(\xi + \tau) \cdot d\xi \tag{B-4.11}$$

is the *lag window equivalent* of the data window $B_i(t)$. Therefore,

$$\text{ave } \{P_{ei}(f)\} = Q_{ei}(f) * P(f) \tag{B-4.12}$$

where $Q_{ei}(f)$ is the *spectral window* corresponding to (i.e. the Fourier transform of) the lag window $D_{ei}(\tau)$.

If $J_i(f)$ is the *frequency window* corresponding to (i.e. the Fourier transform of) the data window $B_i(t)$, then

$$Q_{ei}(f) = \frac{1}{T_n} \mid J_i(f) \mid^2. \tag{B-4.13}$$

(It will be noted that (B-4.10) can be obtained more directly from (B-4.9), by taking the ensemble average of the right-hand member of (B-4.9). In this case, λ is superfluous and need not have been introduced at the start.)

These formulas for the "direct" method, where Fourier transforma-

tion (here modified by the data window) precedes multiplication (here squaring) can be used in several ways. Let us consider three possibilities:

(a) we may choose $B_i(t)$ non-zero over as long a range as possible, so that $D_{ei}(\tau)$ will be broad and $Q_{ei}(f)$ narrow,

(b) we may choose $B_i(t)$ non-zero over a much shorter range, making $D_{ei}(\tau)$ not so broad and $Q_{ei}(f)$ not so narrow,

(c) we may choose a number of such short windows, use each for $B_i(t)$ and average the corresponding values of $P_{ei}(f)$ so obtained into a general average.

If we follow choice (a), our result will behave similarly to those obtained from the indirect method when we try to make $Q_i(f)$ like $\delta(f)$. We shall estimate an average over a very short frequency interval, and our estimate will be excessively variable.

If we follow choice (b), we shall estimate an average over a wider frequency interval, but our estimate will remain just as variable.

If we follow choice (c) our estimate will refer to the same sort of smoothing as in (b), but we shall gain increased stability for the estimate. The behavior of the estimate will resemble that of a reasonable estimate by the indirect route.

Finally, there is another way in which we can apply the formulas for the direct route. We may use a long data window, calculate many values of $P_{ei}(f)$, and then average these results over moderately wide frequency intervals. Again our estimates will be estimates of considerably smoothed spectral densities; again our estimates will be moderately stable.

Of all these, the simplest description of an estimate which is moderately stable, and must, consequently, be an estimate of an at least moderately smoothed spectral density, is the indirect route. We shall stick to the indirect route for the present.

B.5 *Particular Pairs of Windows*

In this section we will consider five pairs of windows. They are illustrated in Figs. 1, 14, and 15. We begin with the

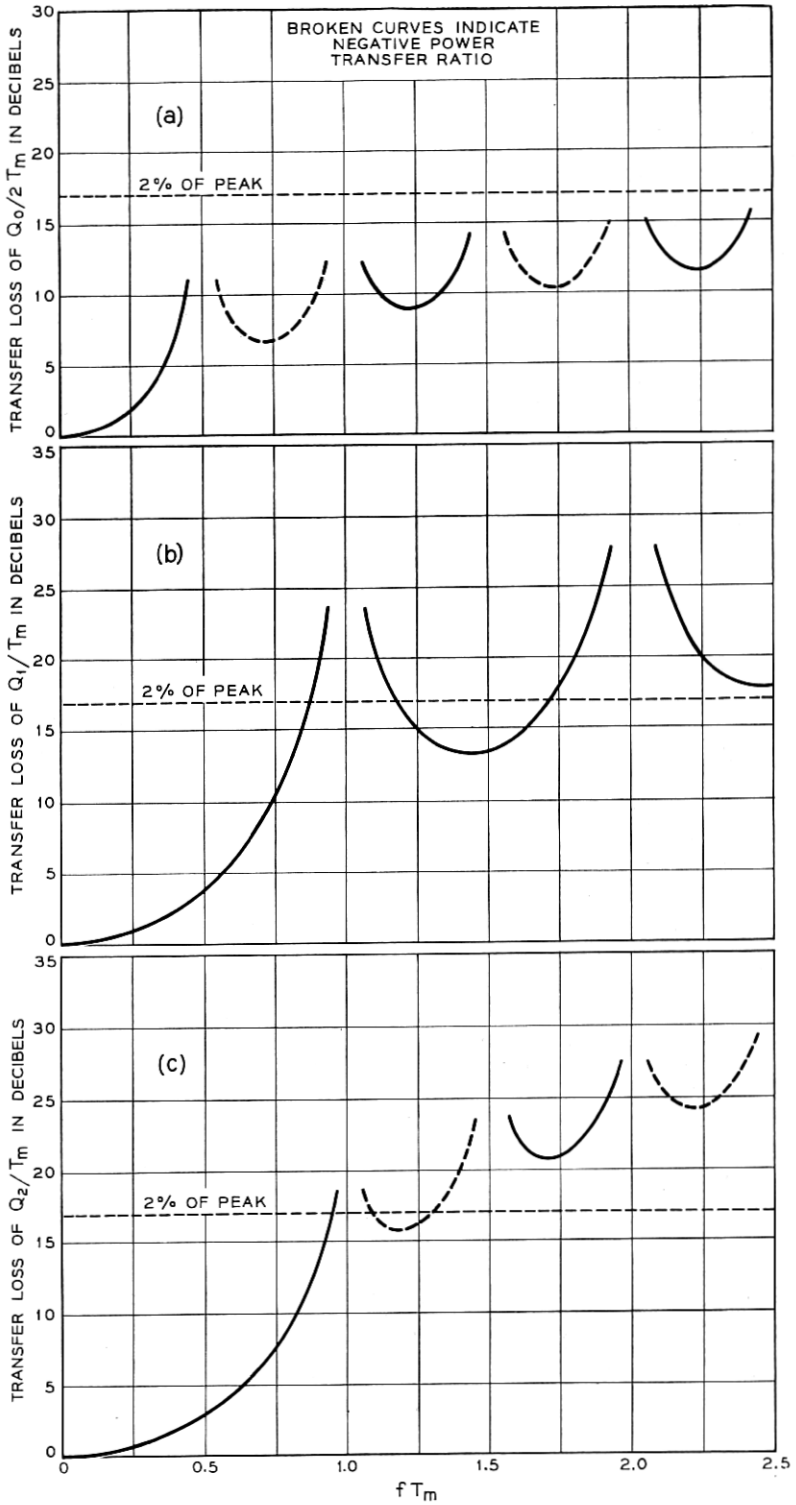
Zeroth Pair

$$\begin{aligned} D_0(\tau) &= 1, && |\tau| < T_m, \\ &= 0, && |\tau| > T_m, \end{aligned}$$

and

$$Q_0(f) = 2T_m \frac{\sin 2\pi f T_m}{2\pi f T_m} = 2T_m \operatorname{dif} 2fT_m.$$

Notice that $C_0(\tau) = D_0(\tau) \cdot C_{00}(\tau)$ coincides with $C_{00}(\tau)$ wherever $C_{00}(\tau)$


BROKEN CURVES INDICATE
NEGATIVE POWER
TRANSFER RATIO
(a)
2% OF PEAK
TRANSFER LOSS OF $Q_0/2\,T_m$ IN DECIBELS
(b)
2% OF PEAK
TRANSFER LOSS OF Q_1/T_m IN DECIBELS
(c)
2% OF PEAK
TRANSFER LOSS OF Q_2/T_m IN DECIBELS
0
5
10
15
20
25
30
35
0
0.5
1.0
1.5
2.0
2.5
$f\,T_m$

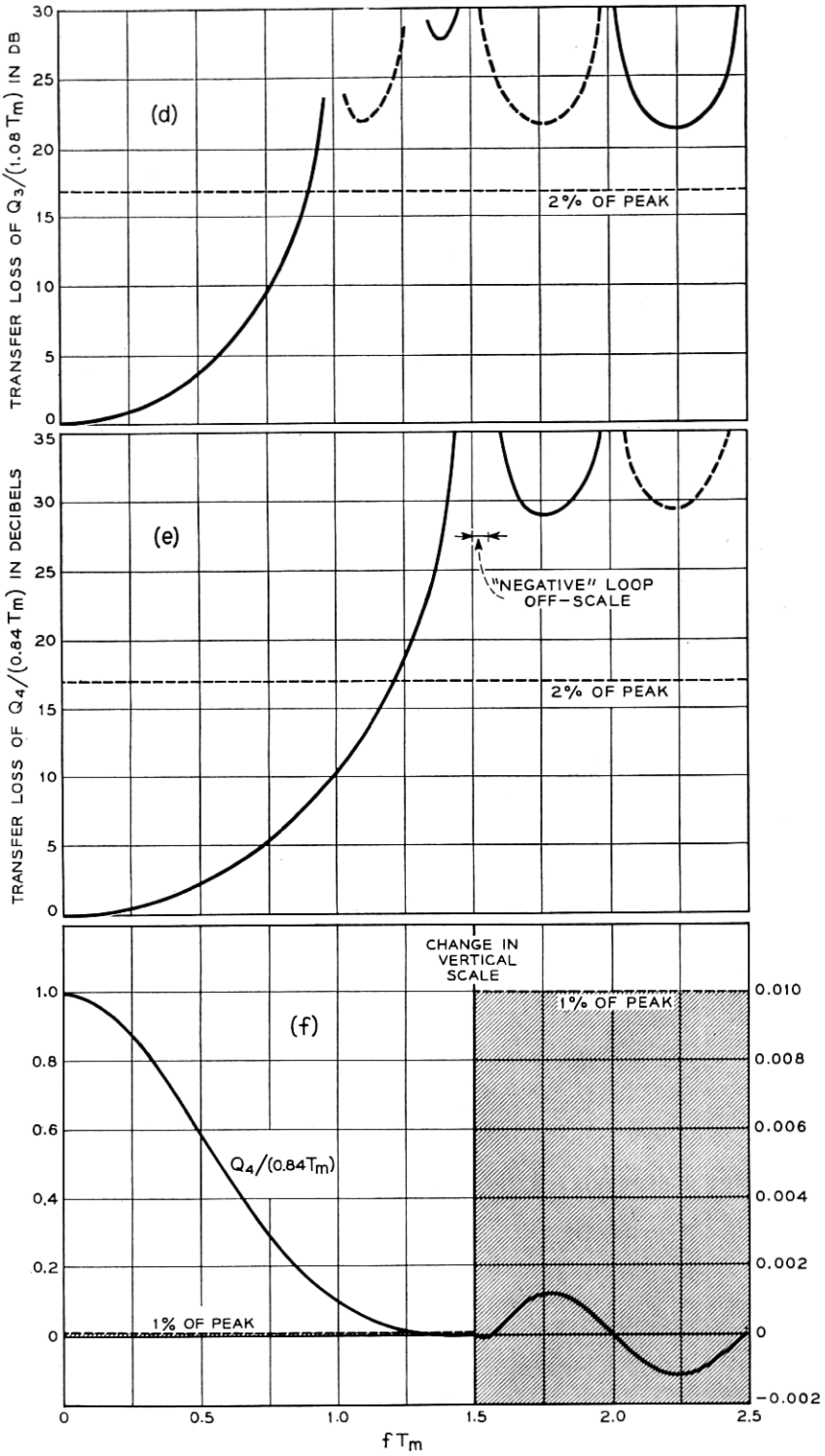


Fig. 15(d)(e)(f) — Transfer loss of spectral windows Q_3, Q_4 in db and power

is defined, and vanishes elsewhere. This is a "do-nothing" pair. However, $Q_0(f)$ is neither simple nor well-behaved, the first side lobe on each side of the main lobe being about $\frac{1}{5}$ the height of the main lobe (and negative).

The specifications of the other four pairs are:

First Pair (Bartlett[5])

$$D_1(\tau) = 1 - \frac{|\tau|}{T_m}, \qquad |\tau| < T_m,$$
$$= 0, \qquad |\tau| > T_m,$$

and

$$Q_1(f) = T_m \left(\frac{\sin \pi f T_m}{\pi f T_m}\right)^2.$$

Second Pair (sometimes called "hanning", after the Austrian meteorologist Julius von Hann)

$$D_2(\tau) = \frac{1}{2}\left(1 + \cos \frac{\pi\tau}{T_m}\right), \qquad |\tau| < T_m,$$
$$= 0, \qquad |\tau| > T_m,$$

and

$$Q_2(f) = \frac{1}{2} Q_0(f) + \frac{1}{4}\left[Q_0\left(f + \frac{1}{2T_m}\right) + Q_0\left(f - \frac{1}{2T_m}\right)\right].$$

Third Pair (sometimes called "hamming", after R. W. Hamming[26])

$$D_3(\tau) = 0.54 + 0.46 \cos \frac{\pi\tau}{T_m}, \qquad |\tau| < T_m,$$
$$= 0, \qquad |\tau| > T_m,$$

and

$$Q_3(f) = 0.54 Q_0(f) + 0.23\left[Q_0\left(f + \frac{1}{2T_m}\right) + Q_0\left(f - \frac{1}{2T_m}\right)\right].$$

Fourth Pair (RBB's not **very** serious proposal)

$$D_4(\tau) = 0.42 + 0.50 \cos \frac{\pi\tau}{T_m} + 0.08 \cos \frac{2\pi\tau}{T_m}, \qquad |\tau| < T_m,$$
$$= 0, \qquad |\tau| > T_m$$

and

$$Q_4(f) = 0.42Q_0(f) + 0.25\left[Q_0\left(f + \frac{1}{2T_m}\right) + Q_0\left(f - \frac{1}{2T_m}\right)\right] + 0.04\left[Q_0\left(f + \frac{1}{T_m}\right) + Q_0\left(f - \frac{1}{T_m}\right)\right].$$

These specifications are all special cases of

$$D_1(\tau) = a_{i0} + 2\sum_{j=1}^{\infty} a_{ij} \cos\frac{j\pi\tau}{T_m}, \qquad |\tau| < T_m,$$
$$= 0, \qquad |\tau| > T_m,$$

with

$$a_{i0} + 2\sum_{j=1}^{\infty} a_{ij} = 1,$$

whence,

$$Q_i(f) = a_{i0}Q_0(f) + \sum_{j=1}^{\infty} a_{ij}\left[Q_0\left(f + \frac{j}{2T_m}\right) + Q_0\left(f - \frac{j}{2T_m}\right)\right].$$

The coefficients in $D_3(\tau)$ may be regarded as convenient approximations to

$$a_{30} = \frac{25}{46}, \qquad a_{31} = \frac{21}{92},$$

which would have produced a zero of $Q_3(f)$ at $|f| = 1.25/T_m$, with other zeros occurring at all integral multiples of $0.5/T_m$ except at 0 and $\pm 0.5/T_m$. (The zero which could have been produced at $|f| = 1.25/T_m$ actually occurs at approximately $|f| = 1.3/T_m$.) The coefficients in $D_3(\tau)$ were actually selected to minimize the height of the highest side lobe.

The coefficients in $D_4(\tau)$ are convenient approximations to

$$a_{40} = \frac{3969}{9304}, \qquad a_{41} = \frac{1155}{4652}, \qquad a_{42} = \frac{715}{18608},$$

which would have produced zeros of $Q_4(f)$ at $|f| = 1.75/T_m$ and $|f| = 2.25/T_m$, with other zeros occurring at all integral multiples of $0.5/T_m$ except at 0, $\pm 0.5/T_m$, and $\pm 1/T_m$.

In view of the fact that

$$D_i(\tau) \cdot D_0(\tau) = D_i(\tau),$$

we have

$$Q_i(f) * Q_0(f) = Q_i(f),$$

relations which may be of interest, as may the fact that using $D_i(\tau) \cdot D_j(\tau)$ for the lag window corresponds in general, to a spectral window $Q_i(f) * Q_j(f)$.

The writers have spent considerable time and effort inquiring into other possible window pairs. One of the most promising approaches was that of Čebyšëv or Chebyshëv polynomials (see Dolph[27]) to obtain side lobes of equal height. Their present conviction is that: (i) special windows *cannot* eliminate the need for prewhitening and rejection filtration and (ii) good prewhitening and rejection filtration *can* eliminate the need for special windows. Accordingly they do not recommend expending extensive effort on special windows.

Readers familiar with physical optics will recognize the close relation between the considerations of this section and diffraction by slits of uniform ($i = 0$) or varying ($i = 1, 2, 3, 4$) width. The literature on apodization (Jaquinot,[28] Boughon, Dossier, Jaquinot[29]) is relevant.

B.6 *Covariability of Power Density Estimates — Basic Result*

To derive a formula for the covariance of two power density estimates $P_i(f_1)$, $P_j(f_2)$ obtained from the same record, we will first derive a formula for the covariance of $M(t_1, \tau_1)$, $M(t_2, \tau_2)$ where

$$M(t, \tau) = X\left(t - \frac{\tau}{2}\right) \cdot X\left(t + \frac{\tau}{2}\right).$$

For this we will use a formula for

$$\text{cov } \{wx, yz\} \equiv \text{ave } \{wxyz\} - \text{ave } \{wx\} \text{ ave } \{yz\}$$

dating back to Isserlis,[30] and used by Hotelling[31] in a similar connection. If w, x, y, z are joint Gaussian variates with zero averages, then

$$\text{cov } \{wx, yz\} = \text{ave } \{wy\} \cdot \text{ave } \{xz\} + \text{ave } \{wz\} \cdot \text{ave } \{xy\}.$$

(This result is easily derived from the "characteristic function".) If we take

$$w = X\left(t_1 - \frac{\tau_1}{2}\right), \qquad x = X\left(t_1 + \frac{\tau_1}{2}\right),$$

$$y = X\left(t_2 - \frac{\tau_2}{2}\right), \qquad z = X\left(t_2 + \frac{\tau_2}{2}\right),$$

and make use of

$$\text{ave}\left\{X\left(t-\frac{\tau}{2}\right)\cdot X\left(t+\frac{\tau}{2}\right)\right\} = C(\tau) = \int_{-\infty}^{\infty} P(f)\cdot e^{i\omega\tau}\,df$$

we get

$$\text{cov}\,\{M(t_1,\tau_1), M(t_2,\tau_2)\} =$$

$$\iint_{-\infty}^{\infty} [e^{i(\omega_1-\omega_2)(\tau_1-\tau_2)/2} + e^{i(\omega_1-\omega_2)(\tau_1+\tau_2)/2}] \qquad \text{(B-6.1)}$$

$$\cdot e^{-i(\omega_1+\omega_2)(t_1-t_2)}\cdot P(f_1)\cdot P(f_2)\cdot df_1\cdot df_2.$$

It may be noted that

$$\text{var}\,\{[X(t)]^2\} = \text{cov}\,\{M(t,0), M(t,0)\} = 2\left[\int_{-\infty}^{\infty} P(f)\cdot df\right]^2,$$

while

$$\text{ave}\,\{[X(t)]^2\} = \int_{-\infty}^{\infty} P(f)\cdot df.$$

Hence,

$$\frac{2\cdot[\text{ave}\,\{[X(t)]^2\}]^2}{\text{var}\,\{[X(t)]^2\}} = 1,$$

in accordance with the fact that $[X(t)]^2$ is a constant multiple of a chi-square variate with one degree of freedom.

Substituting $f_1 = f' + f$, $f_2 = f' - f$, and noting that $df_1\cdot df_2 = 2df\cdot df'$, we get

$$\text{cov}\,\{M(t_1,\tau_1), M(t_2,\tau_2)\} =$$

$$\frac{1}{2}\int_{-\infty}^{\infty} [e^{i\omega(\tau_1-\tau_2)} + e^{i\omega(\tau_1+\tau_2)}]\cdot\Phi(f, t_1-t_2)\cdot df, \qquad \text{(B-6.2)}$$

where

$$\Phi(f,\lambda) = 4\int_{-\infty}^{\infty} P(f'+f)\cdot P(f'-f)\cdot e^{-i2\omega'\lambda}\,df'. \qquad \text{(B-6.3)}$$

Since $\Phi(f,\lambda)$ is an even function of f we may replace $e^{i\omega(\tau_1\pm\tau_2)}$ by $\cos\omega(\tau_1\pm\tau_2)$, and since $P(f'+f)\cdot P(f'-f) = P(f+f')\cdot P(f-f')$ is an

even function of f' we may replace $e^{-i2\omega'\lambda}$ by $\cos 2\omega'\lambda$. Hence, we have

$$\text{cov}\ \{M(t_1, \tau_1), M(t_2, \tau_2)\} =$$

$$\int_{-\infty}^{\infty} \cos \omega\tau_1 \cdot \cos \omega\tau_2 \cdot \Phi(f, t_1 - t_2) \cdot df, \quad \text{(B-6.4)}$$

where

$$\Phi(f, \lambda) = 4 \int_{-\infty}^{\infty} P(f + f') \cdot P(f - f') \cdot \cos 2\omega'\lambda \cdot df'. \quad \text{(B-6.5)}$$

The next step is to determine the covariance between $C_0(\tau_1)$ and $C_0(\tau_2)$. By definition, and for hypothetical computation,

$$C_0(\tau) = \frac{1}{T_n - |\tau|} \int_{-(T_n - |\tau|)/2}^{(T_n - |\tau|)/2} M(t, \tau) \cdot dt, \quad \text{(B-6.6)}$$

but this is inconvenient for present purposes on account of the dependence of both the limits of integration and the divisor upon τ. A more convenient form would be

$$C_0'(\tau) = \frac{1}{T_n'} \int_{-T_n'/2}^{T_n'/2} M(t, \tau) \cdot dt, \quad \text{(B-6.7)}$$

where $T_n' \leqq T_n - T_m$. This form could actually be used for computation, but it would not make the maximum possible use of the data. The range of integration in $C_0'(\tau)$ is less than it is in $C_0(\tau)$ for any τ except possibly $|\tau| = T_m$. A good approximation to the use of $C_0(\tau)$ for computation is to regard this as (approximately) equivalent to the use of a hypothetical, modified $C_0'(\tau)$ with $T_n - T_m < T_n' < T_n$. We will take

$$T_n' = T_n - \alpha_i T_m, \quad \text{(B-6.8)}$$

where α_i depends upon the lag window to be used. Since, for each value of τ, the range of integration in formula (B-6.6) suffers a loss of τ out of T_n, and since the seriousness of this loss depends on the value of $D_i(\tau)$, it seems to be a reasonable approximation to take

$$\alpha_i = \frac{1}{T_m} \frac{\int_0^{T_m} \tau \cdot D_i(\tau) \cdot d\tau}{\int_0^{T_m} D_i(\tau) \cdot d\tau}, \quad \text{(B-6.9)}$$

which yields, for the first four lag windows described in Section B.5, $\alpha_0 = 0.50$, $\alpha_1 = 0.33$, $\alpha_2 = 0.30$, $\alpha_3 = 0.33$.

Using the approximation described in the preceding paragraph, we find (omitting the prime in C_0')

$$\text{cov } \{C_0(\tau_1), C_0(\tau_2)\} = \int_{-\infty}^{\infty} \cos \omega\tau_1 \cdot \cos \omega\tau_2 \cdot \Gamma(f)\, df \quad \text{(B-6.10)}$$

where

$$\begin{aligned} \Gamma(f) &= \frac{1}{(T_n')^2} \iint_{-T_n'/2}^{T_n'/2} \Phi(f, t_1 - t_2) \cdot dt_1 \cdot dt_2 \\ &= 4 \int_{-\infty}^{\infty} P(f + f') \cdot P(f - f) \cdot \left(\frac{\sin \omega' T_n'}{\omega' T_n'}\right)^2 \cdot df'. \end{aligned} \quad \text{(B-6.11)}$$

The final step is to determine the covariance of $P_i(f_1)$ with $P_j(f_2)$, recalling that, for example,

$$P_i(f) = \int_{-\infty}^{\infty} C_i(\tau) \cdot \cos \omega\tau \cdot d\tau,$$

where $C_i(\tau) = D_i(\tau) \cdot C_0(\tau)$. We get

$$\text{cov } \{P_i(f_1), P_j(f_2)\} = \frac{1}{4} \int_{-\infty}^{\infty} H_i(f; f_1) \cdot H_j(f; f_2) \cdot \Gamma(f) \cdot df, \quad \text{(B-6.12)}$$

where

$$\begin{aligned} H_i(f; f_1) &= 2 \int_{-\infty}^{\infty} D_i(\tau_1) \cdot \cos \omega\tau_1 \cdot \cos \omega_1\tau_1 \cdot d\tau_1 \\ &= Q_i(f + f_1) + Q_i(f - f_1) \end{aligned} \quad \text{(B-6.13)}$$

with a similar formula for $H_j(f; f_2)$. In particular, of course,

$$\text{var } \{P_i(f_1)\} = \frac{1}{4} \int_{-\infty}^{\infty} [H_i(f; f_1)]^2 \cdot \Gamma(f) \cdot df. \quad \text{(B-6.14)}$$

The *power-variance spectrum* is given by (B-6.11), and involves the true power spectrum in an essential way, as we would expect. Together with (B-6.12) and (B-6.14), this is the basic result. It is exact for estimates based on $C_0'(\tau)$, approximate for estimates based on $C_0(\tau)$.

B.7 *Covariability of Estimates — Various Approximations*

We now need to obtain a variety of approximate forms of the basic result, suitable for use under different conditions. We shall, in turn; (i) assume that the typical frequency scale of $H_i(f; f_1)$ and $H_j(f; f_2)$ is much

larger than $1/T_n'$, (ii) assume that $P(f)$ varies slowly enough for the *quadratic* term in its Taylor series to be neglected at distances up to, say, $3/T_n'$, (iii) assume both (i) and an assumption similar to (ii) about $P_{i1}(f)$ and $P_{j2}(f)$, and, finally, (iv) assume only that $P(f)$ is concentrated in a sharp peak, narrow in comparison with $1/T_n'$.

Combining formulas (B-6.11) and (B-6.12) just obtained, we find

$$\operatorname{cov}\{P_i(f_1), P_j(f_2)\} = \iint_{-\infty}^{\infty} H_i(f; f_1)\cdot H_j(f; f_2)\cdot P(f+f')\cdot P(f-f') \left(\frac{\sin \omega' T_n'}{\omega' T_n'}\right)^2 \cdot df\cdot df'. \tag{B-7.1}$$

The last factor in the integrand becomes rapidly negligible for $|f'|$ greater than, say, $1/T_n'$. On the other hand, the typical frequency scale of $H_i(f; f_1)$ and $H_j(f; f_2)$ is $1/T_m$ which is usually much larger than $1/T_n'$. In most cases, then, we will find that the approximation

$$H_i(f; f_1)\cdot H_j(f; f_2) \approx H_i(f+f'; f_1)\cdot H_j(f-f'; f_2)$$

is a good approximation for the values of f' making contributions of any importance to the integral. Under this approximation, and a change of variables of integration to

$$f_1' = f + f', \qquad f_2' = f - f',$$

we have, approximately,

$$\operatorname{cov}\{P_i(f_1), P_j(f_2)\} \approx \frac{1}{2}\int_{-\infty}^{\infty} H_i(f_1'; f_1)\cdot P(f_1')\cdot H_j(f_2'; f_2)\cdot P(f_2') \cdot\left[\frac{\sin \pi(f_1' - f_2')T_n'}{\pi(f_1' - f_2')T_n'}\right]^2 \cdot df_1'\cdot df_2'.$$

If we let

$$P_{i1}(f) = H_i(f; f_1)\cdot P(f), \tag{B-7.2}$$

and

$$P_{j2}(f) = H_j(f; f_2)\cdot P(f), \tag{B-7.3}$$

then, approximately,

$$\text{cov } \{P_i(f_1), P_j(f_2)\}$$

$$\approx 2 \iint_0^\infty P_{i1}(f_1') \cdot P_{j2}(f_2') \cdot \left[\frac{\sin \pi(f_1' - f_2')T_n'}{\pi(f_1' - f_2')T_n'}\right]^2 \cdot df_1' \cdot df_2', \quad \text{(B-7.4)}$$

which is our approximation on assumption (i).

In terms of the same quantities we have, from (B-4.5),

$$\text{ave } \{P_i(f_1)\} = \int_0^\infty P_{i1}(f_1') \cdot df_1', \quad \text{(B-7.5)}$$

and

$$\text{ave } \{P_j(f_2)\} = \int_0^\infty P_{j2}(f_2') \cdot df_2'. \quad \text{(B-7.6)}$$

The following heuristic interpretation of the last three formulas is useful. Frequency f_1' is involved in $P_i(f_1)$ with a net weight of $P_{i1}(f_1')$, while frequency f_2' is involved in $P_j(f_2)$ with a net weight of $P_{j2}(f_2')$. These are net weights, and might represent partial cancellation. The covariance between $P_i(f_1)$ and $P_j(f_2)$ might involve additional sources of variability. To the extent that the approximation to $H_i(f; f_1) \cdot H_j(f; f_2)$ is valid, there are no additional sources of variability — the covariance involves the same net weights. The fact that we have only a record of equivalent length T_n' is represented by a tendency of the frequency f_1' to become entwined with the frequency f_2', measured by the factor

$$\left[\frac{\sin \pi(f_1' - f_2')T_n'}{\pi(f_1' - f_2')T_n'}\right]^2.$$

The fact that there are no additional sources of variability as we pass from first to second moments (to the accuracy of the approximation) is good evidence that we are using the data in a relatively efficient way.

If we begin anew from (B-7.1) by expanding $P(f \pm f')$ in Taylor series around $f' = 0$, we find

$$P(f + f') \cdot P(f - f') = [P(f)]^2 + (f')^2\{P(f) \cdot P''(f) - [P'(f)]^2\} + \cdots\cdots$$

symmetry forcing the odd-order terms, including the first, to vanish. Since the trigonometric factor is very small for $|f'| > 2$ or 3 times $1/T_n'$, we may often replace $P(f + f') \cdot P(f - f')$ by $[P(f)]^2$ to a very good approximation. This yields, after integrating out f',

$$\Gamma(f) \approx \frac{2}{T_n'}[P(f)]^2,$$

whence,

$$\text{cov } \{P_i(f_1), P_j(f_2)\} \approx \frac{1}{T'_n} \int_0^\infty H_i(f; f_1) \cdot H_j(f; f_2)[P(f)]^2 \cdot df$$

$$\approx \frac{1}{T'_n} \int_0^\infty P_{i1}(f) P_{j2}(f)\, df,$$

our useful approximation on assumption (ii).

In carrying out this approximation we only needed smoothness of $P(f)$ for f's which make a non-negligible contribution to the final result. If $H_i(f; f_1)P(f)$ and $H_j(f; f_2)P(f)$ are both smooth, as under assumption (iii), then $P(f)$ must be smooth except where both $H_i(f; f_1)$ and $H_j(f; f_2)$ are small, and the contribution from such regions can be neglected. Thus our useful approximation for the *covariance* under assumptions (iii) is the same as that under assumption (ii). (The approximation for $\Gamma(f)$ need not be so accurate.)

Finally, if $P(f)$ consists only of a very sharp peak (width $\ll 1/T'_n$) at $f = f_0$ (and, of course, at $f = -f_0$), with area A, then

$$P(f) \approx A \cdot [\delta(f + f_0) + \delta(f - f_0)],$$

whence,

$$\Gamma(f) \approx 2A^2 \cdot \left[\delta(f + f_0) + \delta(f - f_0) + 2\left(\frac{\sin \omega_0 T'_n}{\omega_0 T'_n}\right)^2 \cdot \delta(f)\right].$$

Hence,

$$\text{cov } \{P_i(f_1), P_j(f_2)\} \approx A^2 \left\{H_i(f_0\,;\, f_1) \cdot H_j(f_0\,;\, f_2) + H_i(0;\, f_1) \cdot H_j(0;\, f_2) \cdot \left(\frac{\sin \omega_0 T'_n}{\omega_0 T'_n}\right)^2\right\},$$

which is our useful approximation under assumption (iv). In case $f_0 \gg 1/T'_n$, the trigonometric factor may be neglected, and we may write

$$\text{cov } \{P_i(f_1), P_j(f_2)\} \approx \left(\int_0^\infty P_{i1}(f)\, df\right) \cdot \left(\int_0^\infty P_{j2}(f)\, df\right).$$

B.8 *Equivalent Widths*

Under assumption (ii) of Section B.7, i.e. $P(f)$ slowly varying, we have, for the dimensionless variability of $P_i(f_1)$,

$$\frac{\text{var } \{P_i(f_1)\}}{[\text{ave } \{P_i(f_1)\}]^2} = \frac{1}{T'_n W_e}, \tag{B-8.1}$$

where

$$W_e = \frac{\left[\int_0^\infty P_{i1}(f)\cdot df\right]^2}{\int_0^\infty [P_{i1}(f)]^2\cdot df} \tag{B-8.2}$$

is the *equivalent width* of

$$P_{i1}(f) = H_i(f; f_1)\cdot P(f).$$

We will now determine the equivalent width of $P_{i1}(f)$ under the assumption that, for each value of f_1, $P_{i1}(f)$ is essentially a constant times $H_i(f; f_1)$. The value of the constant may depend upon the value of f_1, but it will not have any effect on the value of W_e. It is convenient to express (B-8.2) in terms of the normalized frequency

$$\varphi = 2\pi f T_m, \qquad (\varphi_1 = 2\pi f_1 T_m),$$

so that

$$W_e = \frac{1}{4\pi T_m}\frac{\left[\int_{-\infty}^\infty \tilde{P}_{i1}(\varphi)\cdot d\varphi\right]^2}{\int_{-\infty}^\infty [\tilde{P}_{i1}(\varphi)]^2\cdot d\varphi}, \tag{B-8.3}$$

where

$$\left.\begin{aligned}
\tilde{P}_{i1}(\varphi) &= \tilde{Q}_i(\varphi + \varphi_1) + \tilde{Q}_i(\varphi - \varphi_1),\\
\tilde{Q}_i(\varphi) &= (1 - a_i)\tilde{Q}_0(\varphi) + \frac{a_i}{2}[\tilde{Q}_0(\varphi + \pi) + \tilde{Q}_0(\varphi - \pi)],\\
\tilde{Q}_0(\varphi) &= \frac{\sin\varphi}{\varphi},\\
a_0 &= 0, \qquad a_2 = 0.50, \qquad a_3 = 0.46.
\end{aligned}\right\} \tag{B-8.4}$$

Since

$$\int_{-\infty}^\infty \frac{\sin\varphi}{\varphi}\,d\varphi = \pi,$$

we get

$$\int_{-\infty}^\infty \tilde{P}_{i1}(\varphi)\cdot d\varphi = 2\pi,$$

and, since

$$\int_{-\infty}^\infty \frac{\sin(\varphi + \alpha)}{\varphi + \alpha}\frac{\sin(\varphi + \beta)}{\varphi + \beta}\,d\varphi = \frac{\sin(\alpha - \beta)}{\alpha - \beta}\pi,$$

we get

$$\int_{-\infty}^{\infty} [\tilde{P}_{i1}(\varphi)]^2 \, d\varphi = 2\pi B ,$$

where

$$\begin{aligned} B = {} & \left(1 - 2a_i + \frac{3}{2} a_i^2\right)\left(1 + \frac{\sin 2\varphi_1}{2\varphi_1}\right) \\ & + a_i(1 - a_i)\left[\frac{\sin (2\varphi_1 - \pi)}{2\varphi_1 - \pi} + \frac{\sin (2\varphi_1 + \pi)}{2\varphi_1 + \pi}\right] \\ & + \frac{a_i^2}{4}\left[\frac{\sin 2(\varphi_1 - \pi)}{2(\varphi_1 - \pi)} + \frac{\sin 2(\varphi_1 + \pi)}{2(\varphi_1 + \pi)}\right]. \end{aligned} \tag{B-8.5}$$

Hence, for these windows,

$$W_e = \frac{1}{2T_m B}. \tag{B-8.6}$$

It will be noted that

$$\begin{aligned} B &= 2\left(1 - 2a_i + \frac{3}{2} a_i^2\right), && \text{for } f_1 = 0, \\ &= 1 - a_i + \frac{1}{2} a_i^2, && \text{for } f_1 T_m = \frac{1}{4}, \\ &= 1 - 2a_i + \frac{7}{4} a_i^2, && \text{for } f_1 T_m = \frac{1}{2}, \\ &= 1 - 2a_i + \frac{3}{2} a_i^2, && \text{for } f_1 T_m = \frac{r}{4}, \\ & && (r = 3, 4, 5, \cdots). \end{aligned}$$

To a close approximation,

$$B = 1 - 2a_i + \frac{3}{2} a_i^2, \qquad \text{for } f_1 T_m \geqq 1,$$

whence, for

$i = 0$, the main lobe is $1/T_m$ wide, and the equivalent width of $P_{01}(f)$ is $1/(2T_m)$,

$i = 2$, the main lobe is $2/T_m$ wide, and the equivalent width of $P_{21}(f)$ is $4/(3T_m)$,

$i = 3$, the main lobe is $2/T_m$ wide, and the equivalent width of $P_{31}(f)$ is $1.258/T_m$.

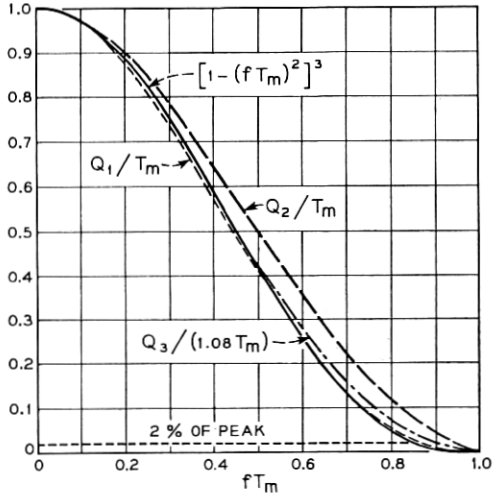


Fig. 16 — Approximation to standard spectral windows.

For $f_1 < 1/T_m$ these equivalent widths are to be reduced. For $f_1 = 0$, they are to be halved.

Assume, next, that

$$Q_i(f) = A\,[1 - (T_m f)^2]^3, \qquad |f| \leqq \frac{1}{T_m}, \tag{B-8.7}$$

$$= 0, \qquad \text{otherwise,}$$

which approximates $Q_1(f)$ quite well for $i = 1$, 2, and 3, as is shown in Fig. 16. Let us further assume that $f_1 \geqq 1/T_m$ and that, at least for $|f - f_1| \leqq 1/T_m$,

$$P(f) = P(f_1)[1 + \beta T_m(f - f_1)]. \tag{B-8.8}$$

Then we can evaluate the equivalent width of

$$P_{i1}(f) = H_i(f; f_1)P(f)$$

by simple integrations, finding

$$W_e = \frac{429}{350 T_m\left(1 + \dfrac{\beta^2}{15}\right)}.$$

Since β^2 cannot exceed 1, we have

$$\frac{1.15}{T_m} < W_e < \frac{1.23}{T_m}, \tag{B-8.9}$$

showing how little effect a linear slope of $P(f)$ across the main lobe of $H_i(f; f_1)$ has on W_e, even when β is large enough for $P(f)$ to vanish at one edge.

If $P(f)$ has exactly the same form as the $H_i(f; f_1)$ corresponding to (B-8.7), then

$$W_e = \frac{0.94}{T_m}, \tag{B-8.10}$$

a rather smaller value.

If, on the other hand, we take

$$\begin{aligned} Q_i(f) &= A[1 - (T_m f)^2]^\ell, \qquad |f| \geqq \frac{1}{T_m}, \\ &= 0, \qquad \text{otherwise}, \end{aligned} \tag{B-8.11}$$

then for $\ell = 2$, and $P(f)$ given by (B-8.8),

$$W_e = \frac{7}{5T_m\left(1 + \frac{\beta^2}{11}\right)},$$

so that

$$\frac{1.28}{T_m} < W_e \leqq \frac{1.40}{T_m}, \tag{B-8.12}$$

while for $\ell = 1$,

$$W_e = \frac{5}{3T_m\left(1 + \frac{\beta^2}{7}\right)},$$

so that

$$\frac{1.46}{T_m} < W_e < \frac{1.67}{T_m}. \tag{B 8.13}$$

In general, we can use

$$W_e \approx \frac{1}{T_m} \tag{B-8.14}$$

as a conservative approximation which provides a factor of safety often near 1.15 or 1.20.

All this was for $f_1 \geqq 1/T_m$. As f_1 is reduced below $1/T_m$ there is overlapping between $Q_i(f - f_1)$ and $Q_i(f + f_1)$ in $H_i(f; f_1)$. As a consequence, the equivalent width decreases in a way which is not worth examining in

great detail. (At $f_1 = 1/(2T_m)$ the equivalent width has decreased about 15 per cent, and at $f_1 = 0$ it has fallen to just half its usual value.)

B.9 *Equivalent Degrees of Freedom*

Let us assume that $P(f)$ is flat — uniformly equal to p_0. Then the power variance spectrum $\Gamma(f)$ will also be flat, but with a value of $2p_0^2/T'_n$. Let us consider $H_i(f; f_1)$ to be the ideal bandpass power transfer function

$$H_i(f; f_1) = A, \qquad f_1 - \frac{W}{2} < |f| < f_1 + \frac{W}{2},$$

$$= 0, \qquad \text{otherwise},$$

where $f_1 > W/2$, although such a transfer function is not even approximately realizable under the requirement that $D_i(\tau) \equiv 0$ for $|\tau| > T_m$. Then

$$\text{ave } \{P_i(f_1)\} = AWp_0,$$

and

$$\text{var } \{P_i(f_1)\} = \frac{A^2Wp_0^2}{T'_n}.$$

If we equate these moments with the corresponding moments of a multiple of a chi-square variate with k degrees of freedom, we get

$$k = \frac{2[\text{ave } \{P_i(f_1)\}]^2}{\text{var } \{P_i(f_1)\}} = 2WT'_n.$$

We get the same result if $H_i(f; f_1)$ is assumed to be the ideal lowpass power transfer function

$$H_i(f; f_1) = A, \qquad |f| < W,$$

$$= 0, \qquad \text{otherwise}.$$

In either case we get only one degree of freedom when $W = 1/(2T'_n)$.

This suggests that the frequency range $f > 0$ be divided into *elementary bands* of width

$$\Delta f = \frac{1}{2T'_n} \tag{B-9.1}$$

with *one degree of freedom in each*. (In the presence of very sharp peaks in the original spectrum it would be somewhat more accurate to divide the frequency range $-\infty < f < \infty$ into bands of width $1/T'_n$ with one

degree of freedom in each, and with one band centered at $f = 0$.) It follows from the results of the preceding section that, if $P(f)$ is reasonably smooth, we may regard the stability of the power density estimate $P_i(f_1)$, for $i = 1$, 2, or 3, as approximately equal to that of a chi-square variate with k degrees of freedom, where

$$k = \frac{2T'_n}{T_m} = 2\left(\frac{T_n}{T_m} - \frac{1}{3}\right) \tag{B-9.2}$$

for one piece (or record).

B.10 *Filtering and Analog Computation*

Power spectrum analysis from continuous data is frequently done by filtering techniques. In this section we will examine some of these techniques. In particular, we will try to express their results in such a form that we are led to express their reliability in equivalent numbers of degrees of freedom.

A common technique is to apply the signal to a narrow-band filter, allow some time for initial transients to become negligible, and then measure the average output power over the remaining time of the record. This corresponds to Mode I as described in Section 10 of Part I. In this technique it is clear that the estimate $P_Y(f_1)$ of the power density $P(f_1)$ in the power spectrum $P(f)$ of the original random process at the filter input, is, substantially,

$$P_Y(f_1) = \int_{-\infty}^{\infty} P_{out}(f; f_1) \cdot df$$

where

$$P_{out}(f; f_1) = |\, Y(f; f_1) \,|^2 \cdot P(f)$$

is the power spectrum of the modified random process at the filter output, and $Y(f; f_1)$ is the transfer function of the filter with a narrow passband around the frequency f_1. The record length is shorter for the modified process than for the original process, by the time allowed for the initial transients to become negligible (at least the reciprocal of the bandwidth in cycles). The effective record length for the modified process determines the width of the elementary bands, and the equivalent number of degrees of freedom is the number of elementary bands in the equivalent width of $P_{out}(f; f_1)$, or the equivalent width of $|\, Y(f; f_1) \,|^2$ if $P(f)$ is reasonably constant in the filter passband.

In the technique described as Mode II in Section 10 we integrate *all* of the power output of the filter, and divide by the length T of the original record to obtain the estimate $P_Y(f_1)$. The result may clearly be ex-

pressed in the form

$$P_Y(f_1) = \frac{1}{T}\int_{-\infty}^{\infty} \{W(t; f_1) * [B(t)\cdot X(t)]\}^2\, dt,$$

where $W(t; f_1)$ is the impulse response of the filter, and $B(t) = 0$ for $t < 0$ and $t > T$, but is otherwise arbitrary. To reduce this to a familiar form we first write it out in detail as

$$P_Y(f_1) = \frac{1}{T}\iiint_{-\infty}^{\infty} W(\tau_1; f_1)\cdot W(\tau_2; f_1)\cdot B(t-\tau_1)\cdot B(t-\tau_2)\cdot X(t-\tau_1)\cdot X(t-\tau_2)\cdot d\tau_1\, d\tau_2\, dt,$$

so that

$$\text{ave}\,\{P_Y(f_1)\} = \frac{1}{T}\iiint_{-\infty}^{\infty} W(\tau_1; f_1)\cdot W(\tau_2; f_1)\cdot B(t-\tau_1)\cdot B(t-\tau_2)\cdot C(\tau_1-\tau_2)\cdot d\tau_1\, d\tau_2\, dt.$$

Now,

$$C(\tau_1 - \tau_2) = \int_{-\infty}^{\infty} P(f)\cdot e^{-i\omega(\tau_1-\tau_2)}\, df,$$

while, if $J(f)$ is the Fourier transform of $B(t)$,

$$\begin{aligned}\int_{-\infty}^{\infty} B(t-\tau_1)\cdot B(t-\tau_2)\cdot dt &= \iiint_{-\infty}^{\infty} J(f')\cdot J(f'')\cdot e^{i\omega'(t-\tau_1)+i\omega''(t-\tau_2)}\, df'\, df''\, dt\\ &= \iint_{-\infty}^{\infty} J(f')\cdot J(f'')\cdot\delta(f'+f'')\cdot e^{-i(\omega'\tau_1+\omega''\tau_2)}\, df'\, df''\\ &= \int_{-\infty}^{\infty} |J(f')|^2\cdot e^{-i\omega'(\tau_1-\tau_2)}\, df'.\end{aligned}$$

Hence,

$$\text{ave}\,\{P_Y(f_1)\} = \frac{1}{T}\iiiint_{-\infty}^{\infty} |J(f')|^2\cdot W(\tau_1; f_1)\cdot W(\tau_2; f_1)\cdot P(f)\cdot e^{i(\omega-\omega')(\tau_1-\tau_2)}\, d\tau_1\, d\tau_2\, df'\, df.$$

Further, since $Y(f; f_1)$ is the Fourier transform of $W(t; f_1)$,

$$\text{ave}\{P_Y(f_1)\} = \frac{1}{T}\int\int_{-\infty}^{\infty} |J(f')|^2 \cdot |Y(f - f'; f_1)|^2 \cdot P(f) \cdot df'\, df.$$

Finally, therefore,

$$\text{ave}\{P_Y(f_1)\} = \int_0^\infty H_Y(f; f_1) \cdot P(f) \cdot df, \tag{B-10.1}$$

where

$$H_Y(f; f_1) = \frac{2}{T} |J(f)|^2 * |Y(f; f_1)|^2. \tag{B-10.2}$$

Since (B-10.1) is in the same form as (B-4.5), we may now apply the results of Sections B.8 and B.9. In particular, if $P(f)$ is reasonably smooth, we may regard the stability of the estimate $P_Y(f_1)$ as approximately equal to that of a chi-square variate with k degrees of freedom, where

$$k = 2T \frac{\left[\int_0^\infty H_Y(f; f_1) \cdot df\right]^2}{\int_0^\infty [H_Y(f; f_1)]^2 \cdot df}. \tag{B-10.3}$$

From the fact that $H_Y(f; f_1)$ is the convolution of the power transfer function of the filter with $(2/T)|J(f)|^2$ it is clear that the passband of $H_Y(f; f_1)$ is at least as wide as that of the wider of the filter and $(2/T)|J(f)|^2$. Hence, the resolving power of the filter method of power spectrum analysis is limited by the length T of available data, just as is that of any other method. If the filter passband is made narrower than $1/T$, say, not only do we gain very little in resolving power, but the stability of our power density estimates is then, at best, approximately that of a chi-square variate with only one or two degrees of freedom. To obtain a reasonable degree of stability the filter passband should be several to many times $1/T$ wide. The resolving power then depends largely on this width. Under these circumstances it should not make much difference which of the four modes described in Section 10 is used.

In Mode III, the output power may clearly be expressed in the form

$$P_Y(f_1) = \frac{1}{T}\int_{-\infty}^{\infty} [\tilde{B}(t) \cdot \{W(t; f_1) * [B(t) \cdot X(t)]\}]^2\, dt,$$

where $\tilde{B}(t) = 0$ for $t < 0$ and $t > T$, but is arbitrary otherwise. The re-

duction of this to a familiar form follows closely that of the preceding case. The final result is that

$$\text{ave } \{P_Y(f_1)\} = \int_0^\infty H_Y(f; f_1) \cdot P(f) \cdot df, \tag{B-10.4}$$

where

$$H_Y(f; f_1) = \frac{2}{T} \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} J(f'') \cdot Y(f'' - f; f_1) \cdot \tilde{J}(f' - f'') \cdot df'' \right|^2 \cdot df'. \tag{B-10.5}$$

The analysis of Mode IV differs from that of Mode III only in the specifications of $\tilde{J}(f)$. Indeed, the results for Modes I and II are also special cases of the result for Mode III. Mode II corresponds to $\tilde{J}(f) = \delta(f)$, while Mode I corresponds to $J(f) = \delta(f)$ as well as $\tilde{J}(f) = \delta(f)$.

It was noted in Section 10 that the zero of the input noise may not be quite at ground potential in Fig. 5. The discrepancy may be considered to produce a spurious *line* or *delta* component at zero frequency in the spectrum of the input noise. It will have no effect on the average output power in Mode I if $f_1 \neq 0$, because the spectral window, which is simply $| Y(f; f_1) |^2$ is opaque at zero frequency. In Mode II, however, the spectral window specified by (B-10.2) may not be sufficiently opaque at zero frequency. The effect on the average output power depends upon the value of $H_Y(0; f_1)$. If the spectral window $H_Y(f; f_1)$ were ideal in the sense that it has no side lobes, there would be no effect on the average output power for values of f_1 at least half of the width of the spectral window. This indicates the desirability of using a graded data window $B(t)$ in order to reduce the side lobes in the spectral window $H_Y(f; f_1)$, that is, essentially the side lobes in $(2/T) |J(f)|^2$. Since the latter window is necessarily positive for all values of f, reduction of side lobes here is not quite as easy as it is in the indirect computation technique described in Section B.4, where selection can be exercised directly on the lag window, and the spectral window can be negative at some frequencies.

Another filter technique frequently used in power spectrum analysis is to apply the available record to the filter as a periodic function with a period equal to the length of the record. If a data window is used, so that the periodic function applied to the filter is of the form $B(t) \cdot X(t)$ in the interval $0 < t < T$, then the power spectrum of the input function is a *line spectrum* with power concentrated at integral multiples of $1/T$

cps. The total power in this spectrum, expressed in a form which displays the distribution in frequency, is

$$\text{input power} = \frac{1}{T} \sum_{q=-\infty}^{q=\infty} P_{\text{in}}\left(\frac{q}{T}\right),$$

where, essentially,

$$P_{\text{in}}(f) = \frac{1}{T}\left| \int_{-\infty}^{\infty} B(t) \cdot X(t) \cdot e^{-i\omega t} \cdot dt \right|^2.$$

Hence, the output power of the filter, is

$$P_Y(f_1) = \frac{1}{T} \sum_{q=-\infty}^{q=\infty} \left| Y\left(\frac{q}{T}; f_1\right) \right|^2 \cdot P_{\text{in}}\left(\frac{q}{T}\right).$$

Aside from the fact that we are now dealing with sums instead of integrals, this analysis parallels very closely our analysis of Mode II. It should be noted, however, that while our summands are even functions of q we may not run our sums from $q = 0$ to $q = \infty$, and then double the result, unless we take only half of the first ($q = 0$) term. Hence, with this technique the equivalent number of degrees of freedom should be taken as

$$k = \frac{\left[\sum_{q=-\infty}^{q=\infty} \left| Y\left(\frac{q}{T}; f_1\right) \right|^2 \right]^2}{\sum_{q=-\infty}^{q=\infty} \left| Y\left(\frac{q}{T}; f_1\right) \right|^4}$$

provided that this turns out to give $k > 3$ or 4, say. It is easily seen that this formula discourages the use of this technique with a filter whose passband is only a few times $1/T$ wide.

B.11 *Prewhitening*

The techniques required here are standard in communication engineering and do not require specific treatment. As these techniques are ordinarily used, the original stationary random process is in effect continuously acting on the input end of the prewhitening filter, so that the output of the filter may be regarded as another stationary random process for purposes of spectral analysis. If these techniques are used where it is practical to obtain only a finite length of the original process to apply to the input of the filter, we must discard an initial portion of the output, corresponding to the time required for the filter transients to die out, as well as all of the output after the input has ceased. These

considerations are taken up in greater detail in Sections B.15 and 16 in connection with the analysis of equi-spaced discrete data.

Details for Equi-Spaced Analysis

B.12 *Aliasing*

Let us consider a stationary random process whose autocovariance function $\tilde{C}(\tau)$, and power spectrum $\tilde{P}(f)$ are known exactly. If we now take only the values of $\tilde{C}(\tau)$ at uniformly spaced values of τ, viz.

$$\tau = 0, \pm\Delta\tau, \pm 2\Delta\tau, \cdots$$

we can, in principle, calculate a corresponding (aliased) power spectrum $\tilde{P}_a(f)$, by using the formula

$$\tilde{P}_a(f) = \int_{-\infty}^{\infty} [\nabla(\tau; \Delta\tau) \cdot \tilde{C}(\tau)] \cdot e^{-i\omega\tau}\, d\tau,$$

where $\nabla(\tau; \Delta\tau)$ is an infinite Dirac comb as defined in Section A.2. Making use of the results of Sections A.2 and A.3 we have

$$\tilde{P}_a(f) = A\left(f; \frac{1}{\Delta\tau}\right) * \tilde{P}(f),$$

or, explicitly,

$$\tilde{P}_a(f) = \sum_{q=-\infty}^{q=\infty} \tilde{P}\left(f - \frac{q}{\Delta\tau}\right).$$

Thus, if it happens that $\tilde{P}(f)$ is zero for $|f| > 1/(2\,\Delta\tau)$, as illustrated in Fig. 17(a), there will be no overlapping of the individual terms in $\tilde{P}_a(f)$, and the result of the summation will be as illustrated in Fig. 17(b). In this case, we can restore the original spectrum by multiplying $\tilde{P}_a(f)$ by a rectangular function according to the formula

$$\tilde{P}(f) = \tilde{P}_A(f) = \tilde{S}(f) \cdot \tilde{P}_a(f),$$

where

$$\tilde{S}(f) = 1, \qquad |f| < \frac{1}{2\Delta\tau},$$

$$= 0, \qquad |f| > \frac{1}{2\Delta\tau},$$

for, in the absence of such overlapping, $\tilde{P}_a(f) = \tilde{P}(f)$ for $|f| < 1/(2\Delta\tau)$. Hence, we will recover the original autocovariance function $\tilde{C}(\tau)$ from

that for the discrete series by a convolution according to the formula

$$\tilde{C}(\tau) = \tilde{G}(\tau) * [\nabla(\tau; \Delta\tau) \cdot \tilde{C}(\tau)]$$
$$= \Delta\tau \cdot \sum_{q=-\infty}^{q=\infty} \tilde{G}(\tau - q\Delta\tau) \cdot \tilde{C}(q\Delta\tau),$$

where

$$\tilde{G}(\tau) = \frac{\sin \frac{\pi\tau}{\Delta\tau}}{\pi\tau} = \frac{1}{\Delta\tau} \cdot \text{dif} \frac{\tau}{\Delta\tau}.$$

Now let us consider a second stationary random process whose auto-covariance function $C(\tau)$ happens to be related to $\tilde{C}(\tau)$ by

$$C(\tau) = G(\tau) \cdot \tilde{C}(\tau),$$

where $G(\tau)$ is unity at $\tau = 0, \pm\Delta\tau, \pm 2\Delta\tau$, etc. Since

$$\nabla(\tau; \Delta\tau) \cdot C(\tau) = \nabla(\tau; \Delta\tau) \cdot \tilde{C}(\tau),$$

it is clear that $P_a(f)$ will be quantitatively, and in the utmost detail

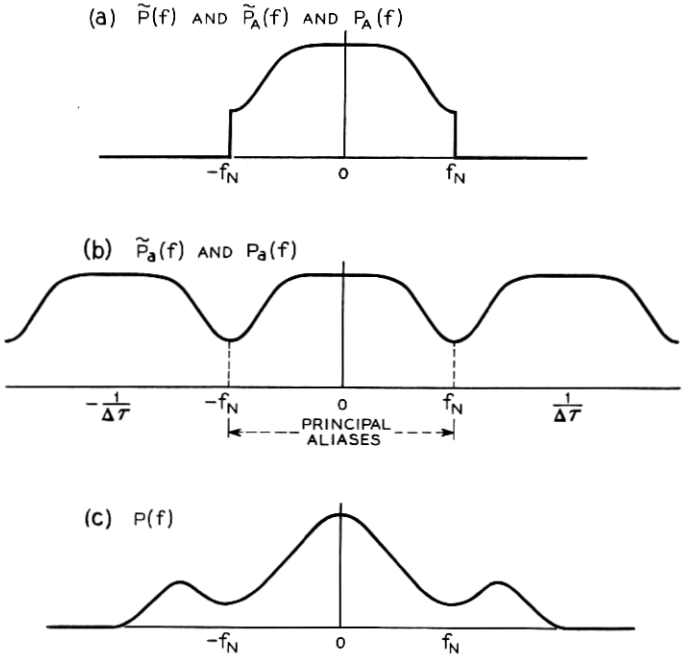


Fig. 17 — Effect of sampling on power spectra.

identical with $\tilde{P}_a(f)$. Hence, the *principal part* $\tilde{P}_A(f)$ of the aliased spectrum $\tilde{P}_a(f)$, that is, the part in the range $|f| < 1/(2\Delta\tau)$, will not in general be the power spectrum of the second process. In fact, if $S(f)$ is the Fourier transform of $G(\tau)$, the power spectrum of the second process is related to the power spectrum of the first process by

$$P(f) = S(f) * \tilde{P}(f).$$

In general, this spectrum covers an infinite range of frequencies, so that the aliased spectrum

$$P_a(f) = A\left(f; \frac{1}{\Delta\tau}\right) * P(f) = \sum_{q=-\infty}^{\infty} P\left(f - \frac{q}{\Delta\tau}\right)$$

will involve overlapping of the individual terms. This overlapping will account for the quantitative identity of $P_a(f)$ with $\tilde{P}_a(f)$, and the failure of its principal part $P_A(f)$ to represent $P(f)$ even in the range $|f| < 1/(2\ \Delta\tau)$. Fig. 17(c) illustrates such a $P(f)$.

It may well have already occurred to readers familiar with amplitude modulation that using only uniformly spaced values of $C(\tau)$, viz., $C(r\Delta\tau)$ where $r = 0, \pm 1, \pm 2, \cdots$, has the same effect on the power spectrum as the simultaneous amplitude modulation of carrier waves with frequencies $q/\Delta\tau$ where $q = 0, \pm 1, \pm 2, \cdots$. If the two-sided power spectrum $P(f)$ corresponding to $C(\tau)$ is visualized as side-bands on a zero-frequency carrier, then the aliased spectrum $P_a(f)$ corresponding to $C(r\Delta\tau)$ will be naturally visualized as the same sidebands on carrier frequencies $q/\Delta\tau$, where $q = 0, \pm 1, \pm 2, \cdots$. If each sideband of $P(f)$ does not extend beyond the frequency $1/(2\Delta\tau)$, then there will be no overlapping of side-bands in $P_a(f)$, and the principal part $P_A(f)$ of $P_a(f)$ will be identical to $P(f)$. Contrariwise, if each sideband of $P(f)$ extends beyond the frequency $1/(2\Delta\tau)$,then there will be overlapping of side-bands in $P_a(f)$, and the principal part $P_A(f)$ of $P_a(f)$ will not be identical to $P(f)$. In any case, however, it is important to note that

$$\int_{-1/(2\Delta\tau)}^{1/(2\Delta\tau)} P_a(f)\, df = \int_{-1/(2\Delta\tau)}^{1/(2\Delta\tau)} P_A(f)\, df = \int_{-\infty}^{\infty} P_A(f)\, df = \int_{-\infty}^{\infty} P(f)\, df.$$

If we examine the relation of $P_A(f')$ to $P(f)$, where

$$0 \leqq f' \leqq \frac{1}{2\Delta\tau}, \quad \text{and} \quad f \geqq 0,$$

we find that

$$P_A(f') = P(f') + \sum_{q=1}^{\infty} \left[P\left(\frac{q}{\Delta\tau} - f'\right) + P\left(\frac{q}{\Delta\tau} + f'\right)\right],$$

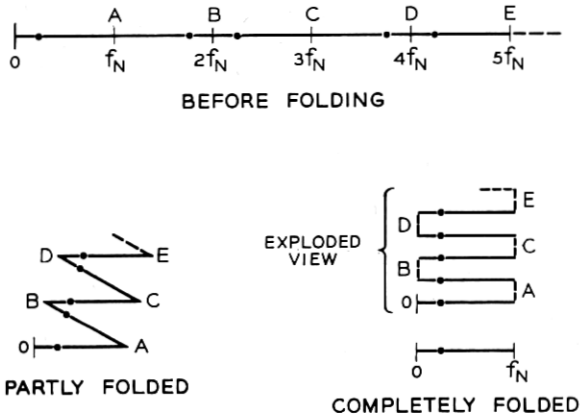


Fig. 18 — Spectrum folding.

(the right-hand side representing $P_a(f')$ for all f'). We say, therefore, that the power density at $f = (q/\Delta\tau) - f'$ and at $f = (q/\Delta\tau) + f'$ in the true spectrum $P(f)$, where $q = 1, 2, \cdots, \infty$, are *aliased* at f' in the principal part $P_A(f')$ of the aliased spectrum. Clearly, the presence of aliases in power density estimates from time series is a matter of some concern.

Aliasing is sometimes called *spectrum folding*, because the pattern by which various frequencies are aliased with one another corresponds to the result of folding up the frequency axis, as illustrated in Fig. 18. The frequency of the first fold is called the *folding* or *Nyquist frequency*. In the discussion above this was $f_N^* = 1/(2\Delta\tau)$, which usually coincides with $f_N = 1/(2\Delta t)$.

B.13 *Transformation and Windows*

A general outline of a hypothetical procedure for power spectrum estimation from a uniformly spaced discrete time series of finite length, by the indirect route, is as follows:

(1) Let $X_0, X_1, \cdots, X_n$ be the time series and let Δt be the time interval between adjacent values. Compute *mean lagged products*, with lag interval $\Delta\tau = h\Delta t$, according to the formula

$$C_r = \frac{1}{n - hr} \sum_{q=0}^{q=n-hr} X_q \cdot X_{q+hr}$$

$$\left(r = 0, 1, \cdots, m \quad \text{where} \quad m \leqq \frac{n}{h}\right).$$

(In a practical procedure, this formula will have to be modified to avoid difficulties with spurious low-frequency components.)

(2) Compute *raw spectral density estimates* according to the formula

$$V_r = \Delta\tau \cdot \left[C_0 + 2 \sum_{q=1}^{q=m-1} C_q \cdot \cos \frac{qr\pi}{m} + C_m \cdot \cos r\pi \right].$$

Since V_r is symmetric in r about every integral multiple of m, it is necessary to compute it only for $r = 0, 1, \cdots, m$. The frequency corresponding to r is $r/(2m\Delta\tau)$, as shown below.

(3) Compute *refined spectral density estimates* according to the formula

$$U_r = a_{i0} V_r + \sum_{j=1}^{\infty} a_{ij} [V_{r+j} + V_{r-j}] ,$$

where the a's are the same as in Section B.5. In particular, for the third pair of lag and spectral windows described in Section B.5, we have $a_{30} = 0.54$ and $a_{31} = 0.23$, all others being zero, so that

$$U_r = 0.23\ V_{r-1} + 0.54\ V_r + 0.23\ V_{r+1} .$$

(These power density estimates should of course be doubled if they are to be referred to positive frequencies only. This doubling may in fact be introduced through the mean lagged products.)

Comparing this outline with the one for the continuous case (Section B.4), it will be noted that the window is introduced by different methods. In the continuous case the window is introduced as a lag window before cosine transformation, in order to avoid convolution after transformation. In the discrete series case, since convolution is not difficult, indeed is very simple, convolution after transformation is convenient, and the lag window is shaped after the cosine transformation.

To relate the outlined procedure for the discrete series case to that for the continuous case, we note that

$$\text{ave } \{C_r\} = C(r\Delta\tau),$$

where $C(\tau)$ is the true autocovariance function. Hence,

$$\text{ave } \{V_r\} = \Delta\tau \cdot \left[C(0) + 2 \sum_{q=1}^{q=m-1} C(q\Delta\tau) \cos \frac{qr\pi}{m} + C(m\Delta\tau) \cos r\pi \right].$$

This may be expressed in the form of a Fourier transform, viz.,

$$\text{ave } \{V_r\} = \int_{-\infty}^{\infty} [\nabla_m(\tau; \Delta\tau) \cdot C(\tau)] \cdot e^{-i\omega\tau}\, d\tau ,$$

where $\nabla_m(\tau;\ \Delta\tau)$ is the finite Dirac comb defined in Section A.2, and

$$f = \frac{\omega}{2\pi} = \frac{r}{2m\Delta\tau}.$$

Therefore,

$$\text{ave } \{V_r\} = \left[Q_0(f;\ \Delta\tau) * P(f) \right]_{f=r/(2m\cdot\Delta\tau)},$$

where $P(f)$ is the true power spectrum. Hence, V_r may be regarded as an estimate of an average-over-frequency of $P(f)$ in the *aliased* neighborhood of $f = r/(2m\Delta\tau)$, with the spectral window $Q_{0A}(f) = Q_0(f;\ \Delta\tau)$ illustrated in Fig. 9.

Three other views of the relation of V_r to $P(f)$ may be developed from the fact that (from the end of Section A.2)

$$\nabla_m(\tau;\ \Delta\tau) = D_0(\tau)\cdot\nabla(\tau;\ \Delta\tau),$$

$$Q_0(f;\ \Delta\tau) = Q_0(f) * A\left(f;\ \frac{1}{\Delta\tau}\right),$$

whence,

$$\text{ave } \{V_r\} = \left[Q_0(f) * A\left(f;\ \frac{1}{\Delta\tau}\right) * P(f) \right]_{f=r/(2m\cdot\Delta\tau)}.$$

The view taken in the preceding paragraph corresponds to the substitution

$$Q_0(f) * A\left(f;\ \frac{1}{\Delta\tau}\right) = Q_{0A}(f).$$

A second view corresponds to the substitution

$$A\left(f;\ \frac{1}{\Delta\tau}\right) * P(f) = P_a(f),$$

and a third to the identity

$$A\left(f;\ \frac{1}{\Delta\tau}\right) * P(f) = A\left(f;\ \frac{1}{\Delta\tau}\right) * P_A(f).$$

The latter two correspond to the results

$$\text{ave } \{V_r\} = \left[Q_0(f) * P_a(f) \right]_{f=r/(2m\cdot\Delta\tau)},$$

$$\text{ave } \{V_r\} = \left[Q_{0A}(f) * P_A(f) \right]_{f=r/(2m\cdot\Delta\tau)}.$$

Hence, V_r may also be regarded as an estimate of an average-over-frequency of the entire *aliased power spectrum* $P_a(f)$ in the neighborhood of $f = r/(2m\Delta\tau)$, with the spectral window $Q_0(f)$ illustrated in Fig. 14, or, and usually most usefully, as an estimate of an average-over-frequency of the *principal part* $P_A(f)$ *of the aliased spectrum*, in the neighborhood of $f = r/(2m\Delta\tau)$, with the aliased window $Q_{0A}(f)$, which is illustrated in Fig. 9.

Finally, a fourth view corresponds to the substitution (from Section B.4)

$$Q_0(f) * P(f) = \text{ave } \{P_0(f)\}.$$

Hence, V_r may also be regarded as an aliased version of the estimated power spectrum $P_0(f)$ which would have been obtained from the continuous data in accordance with the hypothetical procedure outlined in Section B.4.

Regarded from any one of the four points of view developed above, we see that the raw spectral density estimates V_r, (assuming $\Delta\tau$ sufficiently small, so that aliasing is negligible) are in the same position as the estimated power spectrum $P_0(f)$ in the continuous case; the spectral window $Q_{0A}(f)$ is essentially $Q_0(f)$ with the same undesirable side lobes. Hence, the raw power density estimates must be refined. As in the continuous case, the refinement can be done by using a graded lag window before transformation. In the discrete series case, it may also be done by convolving the raw estimates with the a_{ij}'s of Section B.5 to obtain the refined estimates, as described under (3) in the outline.

In any event, we may write, as we did at the end of Section 4, of Part I,

$$\text{ave } \{U_r\} = \int_0^\infty H_i\left(f; \frac{r}{2m\Delta\tau}\right) \cdot P_a(f) \cdot df,$$

where

$$H_i(f; f_1) = Q_i(f + f_1) + Q_i(f - f_1).$$

Alternatively, we may write

$$\text{ave } \{U_r\} = \int_0^\infty H_{iA}\left(f; \frac{r}{2m\Delta\tau}\right) \cdot P(f) \cdot df,$$

or, more usually,

$$\begin{aligned} \text{ave } \{U_r\} &= \int_0^{1/(2\cdot\Delta\tau)} H_{iA}\left(f; \frac{r}{2m\Delta\tau}\right) \cdot P_a(f) \cdot df \\ &= \int_0^\infty H_{iA}\left(f; \frac{r}{2m\Delta\tau}\right) \cdot P_A(f)\ df, \end{aligned}$$

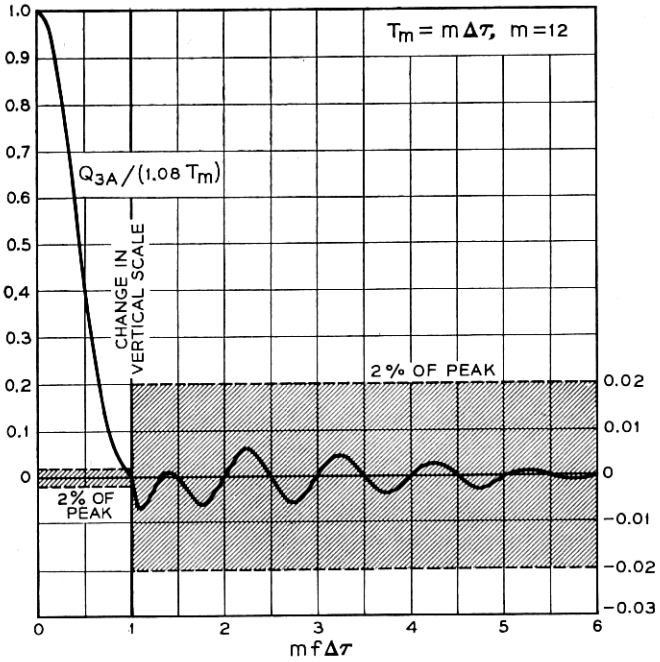


Fig. 19 — Aliased spectral window Q_{3A} for $m = 12$.

where

$$H_{iA}(f; f_1) = Q_{iA}(f + f_1) + Q_{iA}(f - f_1),$$

and

$$Q_{iA}(f) = Q_i(f) * A\left(f; \frac{1}{\Delta\tau}\right) = \sum_{q=-\infty}^{+\infty} Q_i\left(f - \frac{q}{\Delta\tau}\right).$$

In particular, $Q_{2A}(f)$ and $Q_{3A}(f)$ are illustrated in Figs. 8 and 19.

B.14 *Variability and Covariability*

The analysis of variability of power density estimates from discrete time series is a repetition of Section B.6 up to and including (B-6.4) for cov $\{M(t_1, \tau_1), M(t_2, \tau_2)\}$. Beyond this point we now have to deal with summations rather than integrations with respect to t_1 and t_2.

If the range of summation in the formula for the mean lagged products, viz., *n-hr*, is replaced by an average range n', which may be regarded as

the effective length of the series, and which may usually be taken as

$$n' = \text{largest integer less than } \left(n - \frac{hm}{3}\right)$$

then

$$\text{cov } \{C_r, C_s\} = \int_{-\infty}^{\infty} \cos \omega r\Delta\tau \cdot \cos \omega s\Delta\tau \cdot \Gamma_{\Delta t}(f) \cdot df,$$

where

$$\Gamma_{\Delta t}(f) = 4 \int_{-\infty}^{\infty} P(f + f') \cdot P(f - f') \cdot \left(\frac{\sin \omega' n' \Delta t}{n' \sin \omega' \Delta t}\right)^2 \cdot df'$$

(reducing to (B-6.10) and (B-6.11) as $\Delta t \to 0$ with $n'\Delta t = T'_n$). Finally, if U_r and U_s are refined power density estimates based on applying the spectral windows $Q_i(f)$ and $Q_j(f)$, respectively, to the aliased power spectrum $P_a(f)$, then

$$\text{cov } \{U_r, U_s\} = \frac{1}{4} \int_{-\infty}^{\infty} H_{iA}\left(f; \frac{r}{2m\Delta\tau}\right) \cdot H_{jA}\left(f; \frac{s}{2m\Delta\tau}\right) \cdot \Gamma_{\Delta t}(f) \cdot df.$$

In particular, of course

$$\text{var } \{U_r\} = \frac{1}{4} \int_{-\infty}^{\infty} \left[H_{iA}\left(f; \frac{r}{2m\Delta\tau}\right)\right]^2 \cdot \Gamma_{\Delta t}(f) \cdot df.$$

We see now that there is essentially no difference between power density estimation from uniformly spaced time series and power density estimation from continuous data, except for aliasing and its secondary consequences. In particular, if $P(f)$ is reasonably smooth, and if aliasing is negligible, then we may judge the stability of the power density estimates U_r for the uniformly spaced case by analogy with a chi-square variate with k degrees of freedom, where

$$k = 2\left(\frac{n\Delta t}{m\Delta\tau} - \frac{1}{3}\right), \quad r = 1, 2, \cdots, (m - 1),$$

$$= \text{half as much for } r = 0, m.$$

B.15 AND 16 *Transversal Filtering**

If the Z's are moving linear combinations of the X's, for example

$$Z_q = c_0 X_q + c_1 X_{q-1} + \cdots + c_k X_{q-k},$$

* See Kallmann[32] for the origin of this term.

and we take $\Delta t = 1$ for convenience, then the spectra are related by

$$\begin{aligned} P_Z(f) &= P_X(f) \mid c_0 + c_1 e^{-i\omega} + c_2 e^{-2i\omega} + \cdots + c_k e^{-ki\omega} \mid^2 \\ &= P_X(f)[(c_0^2 + c_1^2 + c_2^2 + \cdots + c_k^2) \\ &\quad + 2(c_0c_1 + c_1c_2 + \cdots + c_{k-1}c_k) \cos \omega \\ &\quad + 2(c_0c_2 + c_1c_2 + \cdots + c_{k-2}c_k) \cos 2\omega + \cdots \\ &\quad + 2(c_0c_k) \cos k\omega], \end{aligned}$$

where $\omega = 2\pi f$. The first equality arises by considering $X_q \equiv e^{iq\omega}$, and the second involves the sequence of coefficients

$$\begin{aligned} b_{-k} &= c_0c_k, \\ b_{-k+1} &= c_0c_{k-1} + c_1c_k, \\ &\vdots \\ b_{-1} &= c_0c_1 + c_1c_2 + \cdots + c_{k-1}c_k, \\ b_0 &= c_0^2 + c_1^2 + \cdots + c_k^2, \\ b_1 &= c_1c_0 + c_2c_1 + \cdots + c_kc_{k-1}, \\ b_2 &= c_2c_0 + c_3c_1 + \cdots + c_kc_{k-2}, \\ &\vdots \\ b_{k-1} &= c_{k-1}c_0 + c_kc_1, \\ b_k &= c_kc_0, \end{aligned}$$

which represent the convolution of the sequence $c_0, c_1, \cdots, c_k$ with itself.

As a filter, the moving linear combination is characterized by

$$\mid Y(f)\mid^2 = b_0 + 2\, b_1 \cos \omega + 2\, b_2 \cos 2\omega + \cdots + 2\, b_k \cos k\omega,$$

which is never negative (as the square of an absolute value). Since we can write $\cos j\omega$ as a polynomial in $\cos \omega$ of degree j, we know that:

(1) $\mid Y(f) \mid^2$ can be written as a polynomial of degree k in $\cos \omega$,

(2) for $-1 \leqq \cos \omega \leqq 1$, it is not negative.

Any such polynomial can be realized as a moving linear combination (in several ways, see Wold[33]). The simplest way to see this is to factor the given polynomial into linear and quadratic factors. By appropriate

choice of signs these factors will satisfy (2), and, if each corresponds to a real moving linear combination, the moving linear combination obtained by applying them successively will correspond to the given polynomial.

Any such linear factor takes the form

$$| Y_i(f) |^2 = A^2(1 + a \cos \omega)$$

with $| a | \leqq 1$, which may be realized by $k = 1$ and

$$c_0 , c_1 = \tfrac{1}{2}A(\sqrt{1 + a} \pm \sqrt{1 - a}).$$

Any such quadratic factor takes the form

$$| Y_j(f) |^2 = A^2(1 + a_1 \cos \omega + a_2 \cos^2 \omega).$$

The condition for this to be non-negative for $| \cos \omega | \leqq 1$ is

$$| a_1 | \leqq 1 + a_2,$$

and if $a_2 \geqq 1$, so that the first condition forces an internal extremum on $[-1, +1]$, we must also have $a_1^2 \leqq 4a_2$. We may write

$$| Y_j (f) |^2 = A^2 \left(1 + \frac{a_2}{2} + a_1 \cos \omega + \frac{a_2}{2} \cos 2\omega\right),$$

which is realized by a three-point moving linear combination with

$$c_0 , c_2 = \frac{A}{2\sqrt{2}} (\sqrt{1 + a_2 + \sqrt{\ }} \pm \sqrt{1 - a_2 + \sqrt{\ }}),$$

$$\sqrt{\ } = \sqrt{(1 + a_2)^2 - a_1^2},$$

$$c_1 = \frac{A}{2} (\sqrt{1 + a_2 + a_1} - \sqrt{1 + a_2 - a_1}),$$

the conditions stated above ensuring that all radicals are real.

These two cases — linear and quadratic factors — not only prove that every polynomial non-negative on $(-1, +1)$ can be obtained, but provide at least one way to find a moving linear combination with an assigned polynomial.

One special case deserves record. If we require a simple moving linear combination with $Y(f_0) = 0$, we may use

$$c_0 = \frac{1}{2 \cos \omega_0},$$

$$c_1 = -1,$$

$$c_2 = \frac{1}{2 \cos \omega_0},$$

which has $| Y(f) |^2 = [1 - (\cos \omega / \cos \omega_0)]^2$.

Now that methods are available for finding moving linear combinations whose spectral windows are prescribed polynomials in cos ω, some attention should be given to approximating an arbitrary desired response by polynomials. If a roughly "equal ripple" approximation (where the local maxima deviations of approximation from desired are roughly equal) is desired, then the techniques described in the next paragraph will be quite effective. If, as seems likely, however, we desire the *fractional* error

$$\frac{\text{approximation} - \text{desired}}{\text{desired}}$$

to have roughly equal ripples, then no specific method seems to be available. All we can suggest is the following procedure: (i) find a roughly equal ripple approximation, (ii) find the zeros of its error, (iii) squeeze these zeros together where smaller ripples are desired, and open them out where larger ripples can be permitted, keeping as much of the same general pattern of distances between zeros as possible, (iv) construct a new polynomial with these points for the zeros of its error, (v) adjust the result slightly, if necessary, to make it non-negative. (This procedure sounds quite plausible, but the reader should be warned that we do not know how to be more specific about "keeping as much of the same general pattern of distances as possible". However, we expect the procedure to work in many hands.)

The construction of the roughly "equal ripple" approximation can proceed in many ways. In almost every case, one should begin by calculating values of the desired response at values of cos ω equally spaced from cos $\omega = +1$ to cos $\omega = -1$. The semi-classical approach (DeLury,[34] Fisher-Yates,[35] or Milne[36]) would be to fit orthogonal polynomials to these equi-spaced points. The results would be least-square fits, but might be far from equal-ripple. The process of "economization" (Lanczos,[37] Lanczos,[38] streamlined by Minnick[39]) will allow us to take an over-fitted least-square fit and back up to an "equal-ripple" polynomial of lower order. However, the direct attack, based on central differences of the desired response (at equi-spaced values of cos ω) as proposed by Miller[40] seems likely to be shorter, even allowing for the expansion of the Čebyšëv or Chebyshëv polynomial series into single polynomials.

One further set of considerations remains which is sometimes important. These relate to the starting up of a moving linear combination. If Z_q involves X_q back to X_{q-k}, then there will be k less Z's than X's, and nothing can be done about this. Any transversal filtration causes the loss of some data, and if the filter characteristic is complicated (as a

polynomial in $\cos \omega$) the loss will have to be correspondingly great. This is usually unimportant, but, with very short pieces of record, might become crucial.

In Section 15 we remarked that an autoregressive relation, e.g.,

$$X_q = c_0 Z_q + c_1 Z_{q-1} + \cdots + c_k Z_{q-k},$$

between X's and Z's enabled us to obtain the reciprocal of any suitably non-negative cosine polynomial as the ratio of the spectrum of Z to the spectrum of X. There are different ways of looking at the situation which make this statement true, not true, or partly true. If we have all the X's back to $q = -\infty$, we can calculate the corresponding Z's and it is true. If we have only a finite number of X's, as always in practice, then we have to start the calculation up in some other way, perhaps like this

$$\begin{aligned} X_0 &= c_0 Z_0, \\ X_1 &= c_0 Z_1 + c_1 Z_0, \\ &\vdots \\ X_{k-1} &= c_0 Z_{k-1} + c_1 Z_{k-2} + \cdots + c_{k-1} Z_0, \\ X_k &= c_0 Z_k + c_1 Z_{k-1} + \cdots + c_k Z_0. \end{aligned}$$

As a consequence, we will have introduced an initial transient into the form of the autoregressive transformation so that our Z's are never related to the X's in the way we supposed. In this sense the statement is untrue. In many cases, however, this initial transient dies out quite quickly, and if we discard enough initial Z's, perhaps $2k$ to $4k$ of them, we can regard the reciprocal of the cosine polynomial as a satisfactory approximation. In this sense the statement is partly true.

In theory, an autoregressive scheme corresponds to an infinitely long moving linear combination. In practice it corresponds to a sequence of changing moving linear combinations of finite but increasing length which approximate the infinitely long one. Sometimes the approximation is quite good enough. (Perhaps the main advantage to the autoregressive scheme is its likely reduction in arithmetic — a few autoregressive coefficients corresponding to a long moving linear combination.)

B.17 *Smoothing and Decimation Procedures*

We now study the effects of applying, successively and in any order, simple smoothing and decimation. The basic operations are taking

equally weighted means of ℓ consecutive values, and discarding all but every jth value. Simple sums are usually more convenient than means, and, since the results differ only by a fixed constant factor, lead to the same spectra except for a constant. We define, then, as our basic operations, S_ℓ and F_j, where (for definiteness)

$$Y = S_\ell X$$

means

$$Y_i = X_i + X_{i+1} + \cdots + X_{i+\ell-1}, \qquad (\ell \text{ terms}),$$

while

$$Z = F_j X$$

means

$$Z_i = X_{1+(i-1)j}.$$

It will also be convenient in dealing with the algebra of these operations, to use an operator for simple summing at wider spacings. We therefore define $S_\ell^{(h)}$ by taking

$$W = S_\ell^{(h)} X$$

to mean

$$W_i = X_i + X_{i+h} + \cdots + X_{i+(\ell-1)h}, \qquad (\ell \text{ terms}).$$

It is immediately clear that

$$S_\ell^{(h)} S_h = S_{\ell h} \qquad \text{(B-17.1)}$$

both sides corresponding to forming sums of ℓh consecutive X_t's.

Now, consider the equation

$$Z = S_\ell F_j X,$$

which means that if

$$Y = F_j X,$$

then

$$Z = S_\ell Y.$$

These, in turn, mean that

$$Y_i = X_{1+(i-1)j},$$

and

$$Z_i = Y_i + Y_{i+1} + \cdots + Y_{i+\ell-1}.$$

Hence,

$$Z_i = X_g + X_{g+j} + \cdots + X_{g+(\ell-1)j},$$

where $g = 1 + (i - 1)j$. Now, since

$$W = S_\ell^{(j)} X$$

means

$$W_g = X_g + X_{g+j} + \cdots + X_{g+(\ell-1)j},$$

we have

$$Z_i = W_{1+(i-1)j},$$

which corresponds to

$$Z = F_j W.$$

Thus, we have shown that

$$S_\ell F_j = F_j S_\ell^{(j)}. \tag{B-17.2}$$

Similarly, we find that

$$S_\ell^{(h)} F_j = F_j S_\ell^{(jh)}. \tag{B-17.3}$$

The order in which F_j and S_ℓ or $S_\ell^{(h)}$ are applied is thus important. On the other hand, it is easy to show that

$$S_\ell S_h = S_h S_\ell \neq S_{\ell h}, \tag{B-17.4}$$

and

$$F_j F_h = F_h F_j = F_{jh}. \tag{B-17.5}$$

Thus, sample reductions are

$$S_2 F_4 S_4 = F_4 S_2^{(4)} S_4 = F_4 S_8,$$

and

$$F_2 S_4 F_2 S_2 = F_2 F_2 S_4^{(2)} S_2 = F_4 S_8.$$

It may be noted also that $F_\ell S_\ell$ corresponds to "summing in (barely) non-overlapping blocks of ℓ terms".

A reason for these differing relations is easily found. The changes in the spectrum due to S_ℓ and F_j are quite different in character. S_ℓ multiplies spectra by the power transfer function

$$|\, 1 + e^{-i\omega\Delta t} + \cdots + e^{-i(\ell-1)\omega\Delta t} \,|^2 = \left(\frac{\sin \dfrac{\ell\omega\Delta t}{2}}{\sin \dfrac{\omega\Delta t}{2}} \right)^2,$$

while $S_\ell^{(h)}$ multiplies spectra by

$$\left(\frac{\sin\frac{\ell h\omega\Delta t}{2}}{\sin\frac{h\omega\Delta t}{2}}\right)^2.$$

On the other hand, F_j changes the folding frequency, f_N, to a new value $(1/j)^{\text{th}}$ as large as before, namely

$$f'_N = f_N/j,$$

and aliases together the old principal aliases in sets of j. The j old principal aliases which have f as their new principal alias are

$$f,\ 2f'_N - f,\ 2f'_N + f,\ 4f'_N - f,\ 4f'_N + f,\ \cdots\ (j \text{ terms}).$$

Our combined operations will, in general, change both amplitudes and principal aliases. If

$$P_{+A}(f) = T(f)P_A(f) + T(2f'_N - f)P_A(2f'_N - f) + T(2f'_N + f)P_A(2f'_N + f) + \cdots$$

where $P_{+A}(f)$ is the new aliased spectrum, $P_A(f)$ is the old aliased spectrum, f'_N is the new folding (Nyquist) frequency, and the summation continues as long as the arguments lie below the old folding frequency, then we call $T(f)$, for any f up to the old folding frequency, the *transmission*. (If there is no new aliasing, the transmission is merely the power transfer function.) The ratio of transmission for a desired frequency to the transmission for an undesired alias of that frequency, such as $T(f)/T(2f'_N - f)$ for example, will be called the *protection ratio*.

Fig. 20 illustrates some simple cases. Curve (b) is the transmission of S_3, F_2S_3, or, indeed any F_jS_3. All of these are given by

$$\left(\frac{\sin\frac{3\omega\Delta t}{2}}{\sin\frac{\omega\Delta t}{2}}\right)^2.$$

Scale (c) illustrates the reduced range of the principal aliases for F_2, F_2S_3, or, indeed any $S_jF_2S_k$. Curve (d) shows the power transfer function of S_3, regarded as following F_2. In terms of the respective folding frequencies, this curve duplicates curve (b). On an absolute frequency scale it is different. Curve (e) presents the transmission corresponding to $S_3F_2 = F_2S_3^{(2)}$ which aliases into curve (d).

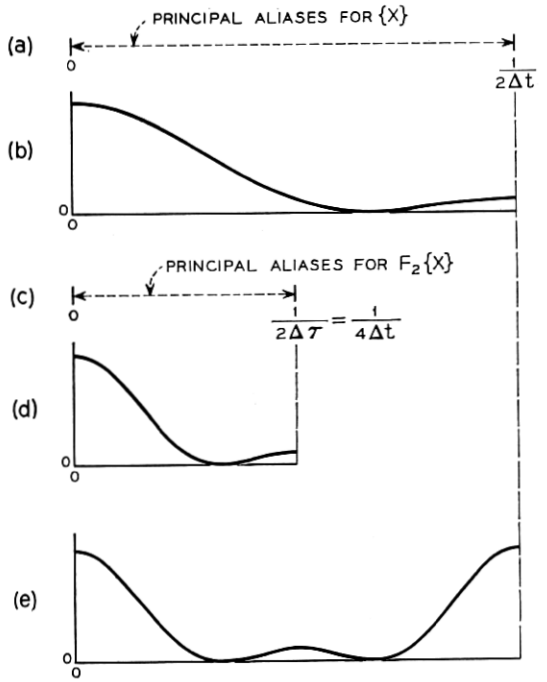


Fig. 20 — Transmission curves for a simple example of smoothing and decimation.

The transmission (= power transfer function) of S_8 is shown in Fig. 21.

The smoothing operation described in Section 17 was, in our present notation, F_jS_k where $j = k/4$, which equals $S_4F_jS_j$. In the case $j = 2$, we have $F_2S_8 = S_4F_2S_2$, and the highlights of transmission and folding pattern are as follows, where the frequency scale has been chosen to make the original folding frequency = 8.

Original freq.	0	1	2	3	4	5	6	7	8
Aliased freq.	0	1	2	3	4	3	2	1	0
Transmission	64	26	zero	3.2	zero	1.4	zero	1.0	zero

The protection provided for frequency 1 against aliasing from frequency 7 is only in the ratio of 26 to 1. If the spectrum falls toward higher frequencies, this may be enough, but adequacy is far from certain.

Double use of S_8 before selecting every second observation would square this protection ratio, giving a ratio near 700. If we are to sum by groups twice, however, we can do better than to use the same length

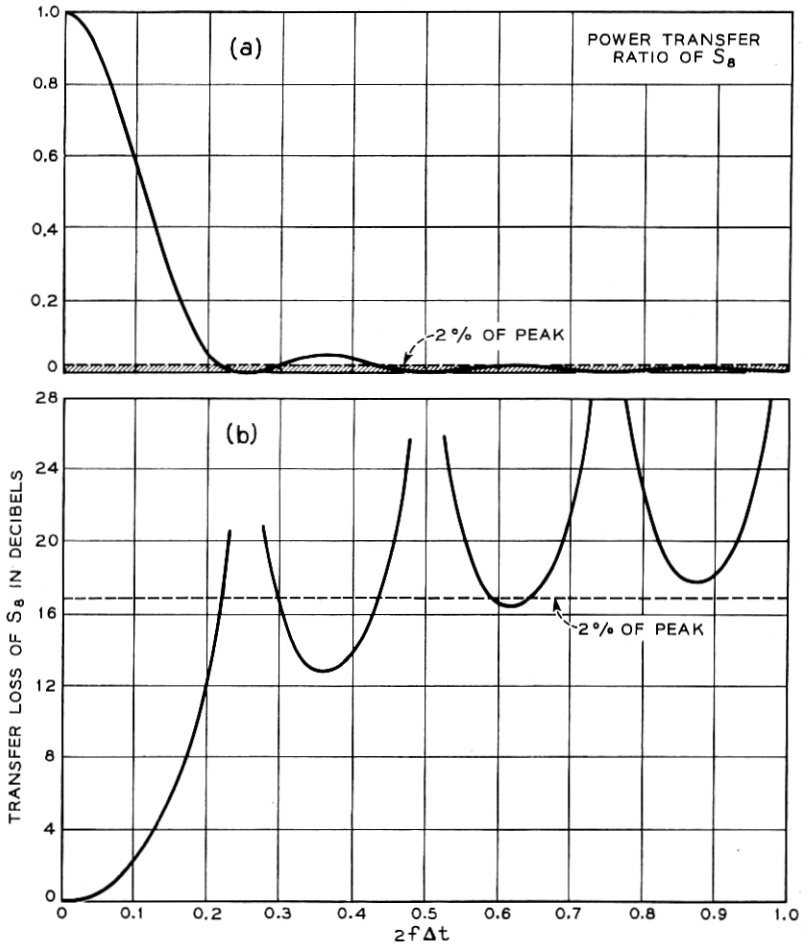


Fig. 21 — Transmission curves for S_8.

of group twice. By using slightly different lengths, we can spread out the zeros of the transmission curve, and tend to hold the right-hand end of the transmission curve nearer the origin. Table V shows the transmission and folding behavior of $F_3S_7S_8$, with the original folding frequency taken as 18. Over the lower half of the folded spectrum, this choice yields protection (against aliasing) by a ratio of about 3,000, which should suffice under even moderately extreme circumstances.

If we do not wish to fold quite so far (or in multiples of 2) then $F_2S_5S_6$ gives a protection ratio of 1,500 or better over the lower half of the folded spectrum.

TABLE V — TRANSMISSION AND FOLDING BEHAVIOR OF $F_3S_7S_8$ OR $F_3S_8S_7$ (ORIGINAL FOLDING FREQUENCY = 18)

Original Frequency	Folded Frequency	Relative Transmission
0	0	3136
1	1	2356
2	2	946
3	3	156
4	4	3.5
5	5	0.03
6	6	3.0
7	5	7.4
8	4	2.4
9	3	zero
10	2	0.04
11	1	0.39
12	0	1.0
13	1	0.17
14	2	0.09
15	3	0.06
16	4	0.12
17	5	0.28
18	6	zero

So long as we are satisfied with such, only moderately large, protection ratios, and with a substantial fall-off of transmission over the useful part of the spectrum, such repeated *unweighted* summing-or-averagings by groups are likely to be most desirable from the point of view of machine computation. If requirements on the smoothing-and-decimation operation are more stringent, smoothing with a suitably chosen set of graded weights is likely to be required.

B.18 *Modified Pilot Estimation, Cascade Estimation.*

The addition-and-subtraction calculation discussed in Section 18 (i) yields only one estimate per octave, (ii) is unduly sensitive to trends, (iii) involves windows which are broad even for a one-per-octave spacing, and (iv) is deliberately wasteful of data in the interest of computational simplicity. There may well be a desire to correct any or all of these.

The δ operator, in a notation extending that of the last section, takes the form $\delta = F_2D$, where F_2 is the operation of dropping every alternate value and D is the operation of differencing adjoining values. We may replace δ by D in the computing routine, keeping σ the same, with the following effects: (i) loss of symmetry in the procedure, (ii) introduction of numerical factors like 8/15 (to be applied to the sums of squared differences), (iii) near doubling of number of differences available for

squaring and summing. If not enough data is available to give adequate stability to the results of the standard pilot method, then this modification may be worth while.

If we denote the operation of omitting the *other* half of the values by F_2' and introduce $\sigma' = F_2'S_2$ as well as $\sigma = F_2S_2$, then, by using D, σ and σ' successively, we may obtain even more differences for squaring and summing. Table VI shows a convenient pattern of computation, in which sums and differences are *not* located in lines near the lines from which they are derived. Table IX, below, (in Section B.28) provides a numerical example.

TABLE VI — COMPACT CALCULATION ARRANGEMENT FOR COMPLETE VERSION OF PILOT ESTIMATION PROCEDURE

X	DX	$\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}X$	$D\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}X$	$\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}X$	$D\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}X$	$\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \sigma'\end{matrix}\right\}X$
X_1		SX_1		SSX_1		$SSSX_1$
	DX_1		DSX_1		$DSSX_1$	
X_2		SX_3		SSX_5		
	DX_2		DSX_3			
X_3		SX_5				
	DX_3		DSX_5			
X_4		SX_7		SSX_3		
	DX_4					
X_5						
	DX_5			SSX_2		
X_6		SX_2				
	DX_6		DSX_2			
X_7		SX_4				
	DX_7		DSX_4			
X_8		SX_6		SSX_4		

$$DX_q = X_{q+1} - X_q\,, \qquad SSX_q = (SX_{q+2}) + (SX_q)$$
$$SX_q = X_{q+1} + X_q\,, \qquad DSSX_q = (SSX_{q+4}) - (SSX_q)$$
$$DSX_q = (SX_{q+2}) - (SX_q) \qquad SSSX_q = (SSX_{q+4}) + (SSX_q)$$

Two methods are available to cope with trend difficulties. The original series may be differenced, and the differenced series subjected to pilot estimation. (The main disadvantage being some loss in accuracy, etc., at very low frequencies.) Or each column of differences may be "corrected for its mean" adjusting the corresponding divisor from k to $k - 1$. (This is only recommended when the first modification is in use and the column contains all D's, not merely the corresponding δ's. Instead of

$$\frac{k+1}{2k}\sum_1^k (D_q)^2$$

then, the corrected sum, corresponding to $(k + 1)/2$ squares, becomes

$$\frac{k+1}{2(k-1)}\left[\sum_1^k (D_q)^2 - \frac{1}{k}\left(\sum_1^k D_q\right)^2\right].$$

In view of the fact that $\sum D_q$ should equal the last value in the preceding column minus the first value there, a convenient check on the D_q exists.

If, as may not be too unlikely, none of these modifications suffices, as will surely be the case if more than one estimate per octave is required, something better than the modified pilot estimation method, without going all the way to the detailed method's equi-spacing in frequency, may be desired. Such an intermediate method should give roughly constant spacing on a logarithmic scale, and provide reasonably clean windows, with about a 100-to-1 ratio between major lobes and minor lobes. It would be useful for high quality pilot estimation, and might sometimes suffice for the complete analysis.

Easy calculation and a roughly logarithmic scale both favor a cascade process, in each of whose cycles some computation is carried out on given values, and then half as many values are computed ready for the next cycle. We may expect, then, that it will be possible to think of each cycle as having three phases:

(a) computation of estimates
(b) smoothing of given values
(c) deletion of alternate smoothed values.

Our main attention needs to be given to the last two phases, since the first is a branch which can be changed rather freely.

We are concerned, therefore, with smoothing procedures to precede halving of the folding frequency. We have, then, to choose two frequencies averaging to the new folding frequency such that we plan to make estimates based on the smoothed (and halved) values up to the lower of these, while effectively eliminating frequencies above the higher. If the folding frequency at the start of the present cycle is f_0, these two frequencies can be written as αf_0 and $(1 - \alpha)f_0$. In the next cycle, then, we anticipate estimation up to αf_0. In the present cycle, we anticipate estimation up to $2\alpha f_0$ and must consequently cover the octave from αf_0 to $2\alpha f_0$.

In the choice of α we must balance two considerations of computing effort. For a given number of lags (a given number of multiplications in forming mean lagged products per cycle) our estimates are spaced a fixed fraction of f_0. The larger α and $2\alpha f_0$, the more of these we may use. Contrariwise, the smaller α, the easier it is, in terms of computational effort, to provide smoothing which suppresses the whole interval from $2\alpha f_0$ to f_0 by a factor of about 100 in comparison with what it does to any frequency between 0 and αf_0.

Trial suggests that a value of α near $\frac{1}{3}$ is reasonable. We wish, therefore, a smoothing which suppresses frequencies between $(\frac{2}{3})f_0$ and f_0.

Two elementary smoothings with zeros in this range are running means-or-sums of 2 (with a zero at f_0) and running means-or-sums of 3 (with a zero at $(\frac{2}{3})f_0$). If, for convenience, we use sums, these operations multiply the spectrum by, respectively,

$$2 + 2 \cos \pi f/f_0 ,$$

and

$$3 + 4 \cos \pi f/f_0 + 2 \cos 2\pi f/f_0 .$$

If both are applied, the factor takes on the following values:

f/f_0 =	0.0	0.3	$\frac{1}{3}$	0.5	0.6	0.62	$\frac{2}{3}$	0.8	0.82	$\frac{5}{6}$	1
factor =	36	15	12	2	0.23	0.08	0	0.146	0.147	0.143	0

The result is a protective factor (against the effects of aliasing on halving) of just over 100 for $\alpha = 0.3$ and of about 80 for $\alpha = \frac{1}{3}$. This should be entirely satisfactory for most pilot purposes. (If further protection is needed, $Z_q = 0.6X_{q-1} + X_q + 0.6X_{q+1}$, which has zero transmission at $f/f_0 = 0.815$, could also be applied.)

If we wish to reduce the number of additions, we might smooth by threes, and then combine smoothing by twos and halving, e.g.,

$$X_q^* = X_{q-1} + X_q + X_{q+1} ,$$
$$Z_q = X_{2q}^* + X_{2q+1}^* ,$$

which requires 2.5 additions per X-value. If we rearrange, however, to

$$\tilde{X}_{2q} = X_{2q} + X_{2q+1} ,$$
$$2\tilde{X}_{2q} = \tilde{X}_{2q} + X_{2q} ,$$
$$Z_q = X_{2q-1} + 2\tilde{X}_{2q} + X_{2q+2} ,$$

we find only 2 additions per X-value. Thus this type of smoothing and halving is computationally quite simple.

If we use this smoothing-and-halving process cycle after cycle we must, after obtaining our spectral estimates for the series actually processed in a given cycle, adjust them for the effects of all the smoothings in preceeding cycles. Near zero frequency this factor is $(36)^{-d}$ for d previous smoothings. At a few selected frequencies the factors are as follows:

f/f_0 =	0	0.075	0.15	0.225	0.30
$j = 1$	$(36)^{-1}$	$1.05(36)^{-1}$	$1.23(36)^{-1}$	$1.60(36)^{-1}$	$2.4(36)^{-1}$
$j = 2$	$(36)^{-2}$	$1.06(36)^{-2}$	$1.29(36)^{-2}$	$1.81(36)^{-2}$	$2.95(36)^{-2}$
$j = 3$	$(36)^{-3}$	$1.07(36)^{-3}$	$1.31(36)^{-3}$	$1.86(36)^{-3}$	$3.10(36)^{-3}$
$j = 4$	$(36)^{-4}$	$1.07(36)^{-4}$	$1.31(36)^{-4}$	$1.87(36)^{-4}$	$3.16(36)^{-4}$

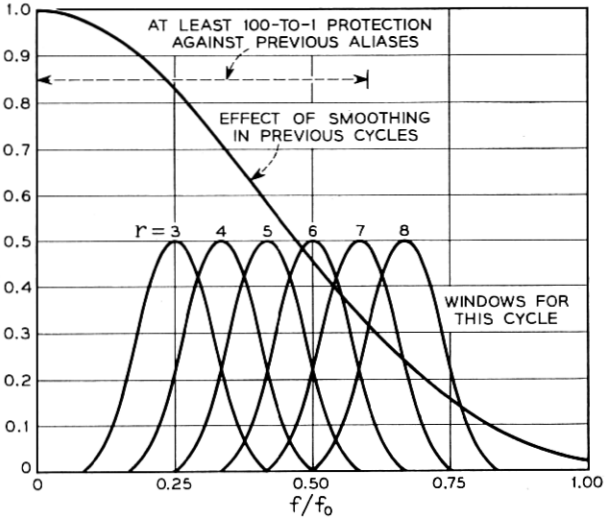


Fig. 22 — Windows for $m = 12$, and smoothing from previous cycles.

When an ordinary calculation with $m = 12$, for example, and hanning is used at each cycle, the spectral windows of the more relevant estimates and the effect of smoothing in previous cycles appear as in Fig. 22. Clearly the estimates for $r = 3, 4, 5$ and probably 6 will be quite usable. The estimate for $r = 7$ may well be usable, but its window extends up to a point where protection against aliasing is beginning to be much reduced. Computation for all r from 0 to 12 is probably worthwhile, lower values of r providing rough checks for later cycles and higher values indicating the extent of the danger from aliasing (to estimates of the next cycle) during the smooth-and-halve phase of the present cycle.

The resulting spectral windows are shown for parts of 4 cycles in Fig. 23. The windows 6, 5, 4, 3 $\sim$ 6′, 5′, 4′, 3′ $\sim$ 6″, 5″, 4″, 3″ $\sim$ 6‴, 5‴, 4‴, 3‴, and so forth, will give a fairly effective set of coarsely spaced spectral estimates.

B.19 *Rejection Near Zero Frequency*

We come now to the details of compensation, on the average, for nonzero but constant averages or for averages changing linearly with time. (As a convenient shorthand we will refer to these as "constant" and "linear" trends.) The X_h, on a sample of which our calculations are to be based, can each usefully be regarded as a sum of two terms: (i) a fixed (but unknown) trend and (ii) a sample from a stationary ensemble with

average zero. These terms are added together and are statistically independent (since one is fixed).

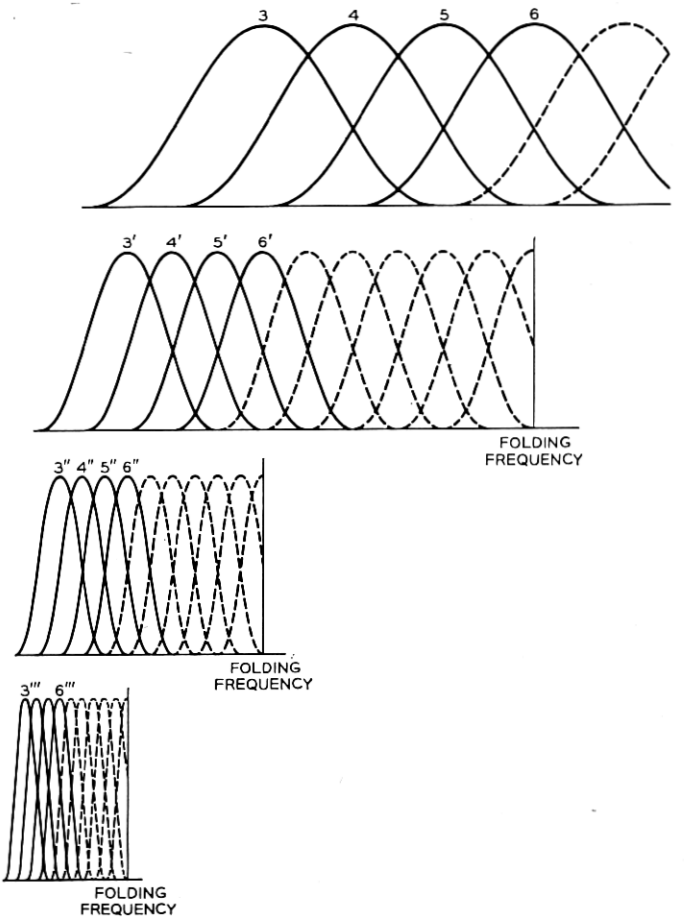


Fig. 23 — Spectral windows in successive cycles (smoothing corrected for).

Consider any quadratic expression in the X_h. If we introduce the two terms for each X_h, we may write the quadratic expression as a sum of three parts: (i) a quadratic in the trend (fixed, of course), (ii) an expression linear both in the trend and in the stationary fluctuations (and hence of average value zero), and (iii) a quadratic in the stationary fluctuations. The first and last of these three parts are, of course, just what would have arisen had the trend alone or, respectively, the fluctuations alone been present. Hence, since the average value of the middle

part vanishes, the average value of the whole quadratic is the sum of the average value for the trend alone and the average value for the fluctuations alone. To study the efficacy with which a quadratic expression rejects a trend in the presence of fluctuation, *we have only to study its behavior in the presence of trend alone.* After this, we shall want to study the behavior for fluctuations alone of those expressions which satisfactorily reject trend.

(The variance of the quadratic expression will be determined from the middle and last parts. If third moments of the fluctuations vanish, as will be the case for Gaussian ensembles, the contributions will come from these parts separately, the covariance between parts vanishing.)

A few formulas will be useful in discussing possible rejection techniques. These are conveniently derived on the basis of $n = 2k + 1$ available equi-spaced values extending from $-k$ through 0 to $+k$. (The indexing of the values is only a convenience, and cannot affect any essential result. The limitation to an odd number of points is essentially a convenience also — we shall replace $2k + 1$ by an unrestricted n rather freely.)

The first formulas relate to constants and linear trends, and are as follows:

$$\text{ave } \{C_r \mid X_t \equiv 1\} = \frac{1}{2k + 1 - r} \sum_{-k}^{k-r} (1)(1) = 1,$$

$$\text{ave } \{C_r \mid X_t \equiv t\} = \frac{1}{2k + 1 - r} \sum_{-k}^{k-r} h(h + r)$$

$$= \frac{1}{6} (2k(k + 1) - r(2k + 1) - r^2),$$

$$= \frac{1}{12} n^2 K_r,$$

where

$$K_r = 1 - \frac{1}{n^2} - 2\frac{r}{n} - 2\frac{r^2}{n^2} \approx 1,$$

whence,

$$\text{ave } \{C_r \mid X_t \equiv \alpha + \beta t\} = \frac{1}{2k + 1 - r} \sum_{-k}^{k-r} (\alpha + \beta h)(\alpha + \beta h + \beta r)$$

$$= \alpha^2 + \frac{\beta^2}{12} n^2 K_r,$$

while

$$\frac{1}{c+d+1}\sum_{-c}^{d}(\alpha+\beta h)=\alpha+\beta\frac{d-c}{2},$$

$$\frac{1}{a+b+1}\sum_{-a}^{b}\frac{1}{c+d+1}\sum_{-c}^{d}(\alpha+\beta h)(\alpha+\beta g)$$
$$=\alpha^2+\alpha\beta\left(\frac{d-c+b-a}{2}\right)+\beta^2\frac{(d-c)(b-a)}{4}.$$

From the first two formulas, as combined in the fourth, we learn what dependence on α and β is required if we are to reject both constants and linear trends for any fixed value of r. The fifth formula, as developed into the last, shows that this cannot be done by subtracting the product of any two simple, equally weighted means of the X's.

Suppose now that $k=3j-2$, so that the means of the lower one-third of all, of the upper one third of all, and of all X's are, respectively

$$\overline{X^-}=\frac{1}{2j-1}\sum_{-k}^{-j}X_h,$$

$$\overline{X^+}=\frac{1}{2j-1}\sum_{j}^{k}X_h,$$

$$\bar{X}=\frac{1}{6j-3}\sum_{-k}^{k}X_h,$$

and so that, when $X_t=\alpha+\beta t$, we have

$$\overline{X^-}=\alpha-\beta\left(\frac{k+j}{2}\right)=\alpha-(2j-1)\beta=\alpha-\frac{n}{3}\beta,$$

$$\overline{X^+}=\alpha+\beta\left(\frac{k+j}{2}\right)=\alpha+(2j-1)\beta=\alpha+\frac{n}{3}\beta,$$

$$\bar{X}=\alpha.$$

Thus if we wish to reject a constant we may choose

$$E_{0r}=(\bar{X})^2 \text{ (independent of } r)$$

and have

$$\text{ave}\,\{C_r\mid X_t\equiv\alpha\}\equiv\text{ave}\,\{E_{0r}\mid X_t\equiv\alpha\}.$$

This choice, however, produces no rejection of $X_t\equiv\beta t$ at all.

If we wish to reject either a constant or a linear trend or both in a

simple way, we may choose

$$E_{1r} = \bar{X}^2 + \frac{3}{16} K_r \cdot (\overline{X^+} - \overline{X^-})^2$$

for which

$$\text{ave } \{C_r \mid X_t \equiv \alpha + \beta t\} \equiv \text{ave } \{E_{1r} \mid X_t \equiv \alpha + \beta t\}.$$

An alternate calculation, nearly equivalent to the use of E_{0r} is illustrated by

$$\begin{aligned}
&\frac{1}{2k+1-r} \sum_{-k}^{k-r} (X_h - \bar{X})(X_{h+r} - \bar{X}) \\
&\qquad = \frac{1}{2k+1-r} \sum_{-k}^{k-r} X_h X_{h+r} - \bar{X}^2 \\
&\qquad\quad - \bar{X} \left(\frac{r\bar{X} - \sum_{k-r+1}^{k} X_h}{2k+1-r} + \frac{r\bar{X} - \sum_{-k}^{-k+r-1} X_h}{2k+1-r} \right) \\
&\qquad = \frac{1}{2k+1-r} \sum_{-k}^{k-r} X_h X_{h+r} - \bar{X}^2 - \bar{X} Q_r ,
\end{aligned}$$

where

$$Q_r = \frac{1}{2k+1-r} \left[2r\bar{X} - \sum_{k-r+1}^{k} X_h - \sum_{-k}^{-k+r-1} X_h \right].$$

Thus subtracting the mean of all the observations from each observation before forming the C_r's is equivalent to using

$$E_{2r} = \bar{X}^2 + \bar{X} Q_r .$$

Similarly, fitting and subtracting the elementary-least-squares straight line corresponds to

$$\hat{\beta} = 3 \frac{\sum_{-k}^{k} hX_h}{k(k+1)(2k+1)} = \frac{12}{n(n^2-1)} \sum_{-k}^{k} hX_h ,$$

and

$$\begin{aligned}
&\frac{1}{2k+1-r} \sum_{-k}^{k-r} (X_h - \bar{X} - \hat{\beta}h)(X_{h+r} - \bar{X} - \hat{\beta}(h+r)) \\
&\qquad = \frac{1}{2k+1-r} \sum_{-k}^{k-r} X_h X_{h+r} - \bar{X}^2 - \bar{X} Q_r \\
&\qquad + \frac{\hat{\beta}}{n-r} \sum_{k-r+1}^{k} (h+r) \cdot (X_h - X_{-h}) - \frac{n^2 \hat{\beta}^2}{12} \left[\frac{2(n^2-1)}{n(n-r)} - K_r \right].
\end{aligned}$$

Thus linear fitting corresponds to the use of

$$E_{3r} = \bar{X}^2 + \bar{X}Q_r - \frac{\hat{\beta}}{n-r}\sum_{k-r+1}^{k}(h+r)\cdot(X_h - X_{-h}) + \frac{n^2\hat{\beta}^2}{12}\left[\frac{2(n^2-1)}{n(n-r)} - K_r\right].$$

and if we wish to simplify computation by neglecting "end corrections", consider

$$E^*_{3r} = \bar{X}^2 + \bar{X}Q_r + \frac{n^2\hat{\beta}^2}{12}K_r.$$

Another version can be obtained by recalling the usual formulation of a sample estimate of a covariance between two unlagged variates. If we write

$$\overline{X^-}_{(r)} = \frac{1}{2k+1-r}\sum_{-k}^{k-r} X_h$$

$$\overline{X^+}_{(\cdot)} = \frac{1}{2k+1-r}\sum_{-k}^{k-r} X_{h+r}$$

then we are led to use the expression

$$\frac{1}{2k+1-r}\sum_{-k}^{k-r}(X_h - \overline{X^-}_{(r)})(X_{h+r} - \overline{X^+}_{(r)}) = \frac{1}{2k+1-r}\sum_{-k}^{k-r} X_hX_{h+r} - \overline{X^-}_{(r)}\overline{X^+}_{(r)}$$

which corresponds to the use of

$$E_{4r} = \overline{X^-}_{(r)}\overline{X^+}_{(r)}$$

or, in the case of a linear trend, perhaps, to

$$E_{5r} = \overline{X^-}_{(r)}\overline{X^+}_{(r)} + \frac{n^2\hat{\beta}^2}{12}\left[K_r + \frac{3r^2}{n^2}\right].$$

(The E's corresponding to fitting a separate straight line to $X_{-k}, \cdots, X_{k-r}$ and to $X_{-k+r}, \cdots, X_k$ do not seem worth writing down.)

Having now obtained a collection of quadratic expressions which, when subtracted from the corresponding C_r, remove the average effect of trend, *we have now to learn what effect these subtractions will have on the contribution of the fluctuations to the average values of the corresponding quantities.* As noted at the beginning of this section, it suffices to consider pure fluctuations, so that we need only study the

spectral windows corresponding to the E_{kr}, and to their finite cosine transforms.

If we consider white noise (of unit variance) for a moment, the X_h become independent with average zero and unit variance and we can calculate the average values of the various E_{kr} in an elementary way. This is of interest, because these average values for white noise are exactly proportional to the integrals of the corresponding spectral windows. We find the following results:

Quadratic expression	Average value for unit white noise	Approximation to this
E_{0r}	$\frac{1}{n}$	$\frac{1}{n}$
E_{1r}	$\frac{1}{n}\left(1 + \frac{9K_r}{8}\right)$	$\sim \frac{2}{n}$
E_{2r}	$\frac{1}{n}$	$\frac{1}{n}$
E^*_{3r}	$\frac{1}{n}\left(1 + \frac{n^2}{n^2 - 1} K_r\right)$	$\sim \frac{2}{n}$
E_{4r}	$\frac{n - 2r}{(n - r)^2}$	$\sim \frac{1}{n}$
E_{5r}	$\frac{n - 2r}{(n - r)^2} + \frac{n}{n^2 - 1}\left(K_r + \frac{3r^2}{n^2}\right)$	$\sim \frac{2}{n}$

The quadratic expressions which eliminate the average effect of a constant trend (one constant) have spectral windows integrating approximately proportional to $1/n$, while those which compensate for general linear trends (two constants) integrate approximately proportional to $2/n$. This is simple, and seems straightforward.

It is possible, however, to eliminate the average effect of a centered linear trend (one with $\alpha = 0$) by subtracting a windowless quadratic expression, one whose average vanishes for all *stationary* ensembles. We need only consider

$$X_{-2}X_0 - 2X_{-1}X_1 + X_0X_2$$

whose average value clearly vanishes under stationarity, but whose value when $X_t \equiv \alpha + \beta t$ is $2\beta^2$, to see that this is so. (Cursory inquiry into variability suggests that such use of windowless quadratics may

increase the variability of the finally resulting spectral density estimates.)

Leaving such possibilities aside, let us compare the E_{kr} for $k = 0, 1, 2, 3, 4, 5$. The corresponding spectral windows can be written down with the aid of the formulas of Section A.6, but as they are rather complicated, we shall avoid doing so. The three choices eliminating constant trends, E_{0r}, E_{2r} and E_{4r}, all have integrated spectral windows approximating $1/n$. The dependence on r is in any case weak, and where present increases with r. For E_{0r} there is no dependence on r.

We are, of course, more concerned with the spectral windows associated with the modified V_r's rather than with those associated with the modifying E_{kr}. If we use E_{0r}, then the change in spectral window is the same for each lag. Consequently, only the $R_{00}(f)$ window is affected, the $R_{0r}(f)$ windows corresponding to the other V_r vanish. The situation is more complicated for the other cases, and no clear advantage over the use of E_{0r} appears.

The spectral window corresponding to $\bar{X}^2$ is, of course, (dif nf/dif f)^2 and is both always positive and well concentrated near zero. (If we replaced $\bar{X}$ by the average over a graded data window we could decrease the corresponding spectral window for f beyond the first side lobe of dif nf, but the concomitant broadening of the main lobe would make the result much less useful.) Clearly the use of E_{0r} will be quite satisfactory.

Comparing E_{1r}, E_{3r} and E_{5r} is not so simple. E^*_{3r} has the smaller integrated spectral window for r/n small, but the simplicity of calculation of E_{1r} will often outweigh this fact. (If economizing on the spectral window is important, we could use a windowless quadratic to eliminate the average effect of β.) Accordingly, the use of E_{1r} is recommended, unless it is simpler to subtract a fitted linear function from all X_t.

(There is no Section B.20.)

B.21 *Sample Computing Formulas*

Only the formulas for correction for prewhitening and for correction for the mean require discussion. The factor for $1 \leqq r \leqq m - 1$ is exactly the reciprocal of the prewhitening transfer function, calculated at the nominal frequency of the estimate. The factor for $r = m$, and the main portion of that for $r = 0$, differ only in the selection of the frequency at which the prewhitening transfer function is evaluated. Since, in each case, the nominal frequency is at one edge of the band of frequencies covered, the frequency of evaluation was displaced from the nominal frequency toward the center of the corresponding band. The choice of a point one-third of the way across the band was somewhat arbitrary.

Finally, there is the question of compensation for the correction for the mean. The raw estimate for $r = 0$ would naturally be thought of as corresponding to the interval from $-1/(4m)$ to $+1/(4m)$ cycles per observation, while the hanned estimate would cover from $-1/(2m)$ to $+1/(2m)$ cycles per observation. Since n degrees of freedom are associated with the entire interval from $-\frac{1}{2}$ to $+\frac{1}{2}$ cycle per observation, the hanned estimate for $r = 0$ is associated with n/m degrees of freedom. One of these has been eliminated by correction for the mean, as would also have been the case had we used E_{2r} or E_{4r}, so that we need to compensate for a reduction in the ratio

$$\frac{\frac{n}{m} - 1}{\frac{n}{m}} = \frac{n - m}{n}$$

whose reciprocal is the first compensating factor for $r = 0$. (Had we used E_{1r}, E_{3r}, or E_{5r}, we would have had to compensate for the loss of two degrees of freedom by a factor $n/(n - 2m)$.)

(There is no Section B.22.)

Details for Planning

B.23 *Duration Requirement Formulas*

We are now in a position to assemble and modify formulas from a number of sections as a basis for formulas expressing explicit requirements. In the process we shall have to give explicit definitions for certain concepts. The first of these is *resolution*. If we hann or hamm, we obtain estimates every $1/(2T_m)$ cps. Adjacent estimates have very considerably overlapping windows, and consequently the estimates have substantially related sampling fluctuations and refer to overlapping frequency regions. It would be a clear mistake to consider these estimates as completely resolved. When we come to next-adjacent estimates, however, the situation is quite different. The overlap is small, the covariance being about 5 per cent of either variance for a moderately flat spectrum. We shall consequently treat such pairs of estimates as completely resolved, and place

$$(\text{resolution in cps}) = 2\,\frac{1}{2T_m} = \frac{1}{T_m \text{ in seconds}}.$$

We can express the stability, so far most often expressed as the number of elementary frequency bands or equivalent degrees of freedom

associated with each estimate, in terms of the spread in db of an interval containing, with prescribed probability, the ratio of true smoothed power to estimated smoothed power. Reference to Table II in Section 9 shows the inter-relations to be, approximately

$$k = 1 + \frac{250}{(80 \text{ per cent range in db})^2},$$

$$k = 1 + \frac{400}{(90 \text{ per cent range in db})^2},$$

$$k = 1 + \frac{625}{(96 \text{ per cent range in db})^2},$$

$$k = 1 + \frac{840}{(98 \text{ per cent range in db})^2}.$$

We shall write our combined formulas in terms of the 90 per cent range. Formulas for other per cent ranges are easily obtained by replacing 400 by the appropriate constant. It must be emphasized that when we use a 90 per cent range we only have 9 chances in 10 of finding *each* individual estimate correspondingly close to its average value and that if we have, say, 30 estimates, we are quite sure that at least one will be more discrepant than this.

We recall that we adopted (see end of Section 6 of Part I)

$$T'_n = (\text{total length of record}) - \frac{p}{3} T_m,$$

where p was the number of pieces, and, for design purposes.

$$k \approx \frac{2T'_n}{T_m}.$$

Hence,

$$(\text{duration in seconds}) = T_n = \left(\frac{T'_n}{T_m} + \frac{p}{3}\right) T_m$$

$$= \left(\frac{k}{2} + \frac{p}{3}\right) \frac{1}{(\text{resolution in cps})}$$

$$= \left(\frac{1}{2} + \frac{200}{(90 \text{ per cent range in db})^2} + \frac{(\text{pieces})}{3}\right) \Big/ (\text{resolution in cps}).$$

Writing (length of each piece) · (number of pieces) for (duration) and

solving for the number of pieces yields

$$(\text{number of pieces}) = \frac{\frac{1}{2} + \frac{200}{(\text{90 per cent range in db})^2}}{(\text{length of each piece})(\text{resolution in cps}) - \frac{1}{3}}.$$

These relations are general, and apply equally to analog processed continuous records or digitally processed equi-spaced records, provided the spectrum is, or can be prewhitened to be, reasonably flat.

B.24 *Digital Requirement Formulas*

If we are to use equi-spaced digital analysis, if we can provide frequency cutoff easily, and if we need to cover frequencies up to some $f_{\max}$, then we can probably take our folding frequency at about $\frac{3}{2} f_{\max}$. The necessary number of lags m then follow from

$$\Delta t = \frac{1}{3 f_{\max}} = \frac{1}{2 f_N},$$

$$m = \frac{T_m}{\Delta t} = 3 T_m f_{\max}.$$

The necessary number of data points follows from

$$\begin{aligned} n = \frac{T_n}{\Delta t} &= \left(T_n' + \frac{p}{3} T_m\right) 3 f_{\max} \\ &= 3 T_n' f_{\max} + p T_m f_{\max} \\ &= \left(3 \frac{T_n'}{T_m} + p\right) (T_m f_{\max}), \end{aligned}$$

and the rough number of multiplications is

$$mn = 9 T_m T_n' (f_{\max})^2 + 3p (T_m f_{\max})^2.$$

The quantity $T_m f_{\max}$ can be written as

$$\frac{f_{\max}}{1/T_m} = \frac{\text{maximum frequency}}{\text{resolution}} = (\text{number of resolved bands}),$$

so that the number of data points becomes

$$\begin{aligned} n &= \left(3 \frac{T_n'}{T_m} + p\right) (\text{number of resolved bands}) \\ &= (1.5k + p) (\text{number of resolved bands}), \end{aligned}$$

Table VII

Second differences of Brouwer's data and the add-and-subtract pilot estimation process as applied to them. (Block 2 only.)

Date	10F*	diff.	2nd diff. $= X_q$	δ	σ	$\delta\sigma$	$\sigma\sigma$	$\delta\sigma\sigma$	$\sigma\sigma\sigma$	DX_q
1853.5	56									
		0								
1854.5	56		−4							
		−4		4	−4					4
1855.5	52		0							
		−4				7	−1			−2
	48		−2							
		−6		7	3					7
	42		+5							
		−1						−2	−4	−9
	41		−4							
		−5		11	3					11
	36		+7							
		+2				−9	−3			−30
1860.5	38		−23							
		−21		40	−6					40
	17		+17							
		−4								−19
	13		−2							
		−6		−4	−8					−4
	7		−6							
		−12				18	2			+7
	−5		+1							
		−11		+8	10					+8
1865.5	−16		+9							
		−2						−23	−19	−20
	−18		−11							
		−13		+13	−9					+13
	−31		+2							
		−11				−3	−21			−1
	−42		+1							
		−10		−14	−12					−14
	−52		−13							
		−23								+15
1870.5	−75		+2							
		−21		+7	11					+7
	−96		+9							
		−12				−17	5			−20
	−108		−11							
		−23		+16	−6					+16
	−131		+5							
		−18						4	14	+4
	−149		+9							
		−9		−9	9					−9
1875.5	−158		0							
		−9				−9	9			+4
	−167		+4							
		−5		−8	0					−8
	−172		−4							
		−9								+8
	−181		+4							
		−5		−10	−2					−10
	−186		−6							
		−11				6	2			+15

TABLE VII, CONTINUED

Date	10F*	diff.	2nd diff. $= X_q$	δ	σ	$\delta\sigma$	$\sigma\sigma$	$\delta\sigma\sigma$	$\sigma\sigma\sigma$	LX_q
1880.5	−197		+9							
		−2		−14	4					−14
	−199		−5							
		−7						1	5	−6
	−206		−11							
		−18		+33	11					+33
	−224		+22							
		+4				−19	3			−30
	−220		−8							
		−4		+8	−8					+8
1885.5	−224		0							
		−4								
1886.5	−228									

* (Brouwer's notation) is the fluctuation in the earth's rotation. Here 10F is the fluctuation expressed in tenths of seconds of time.

and the rough number of multiplications becomes

$$nm = (4.5k + 3p)(\text{number of resolved bands})^2$$

$$= \left[\frac{9}{2} + \frac{1800}{(\text{90 per cent range in db})^2} + 3(\text{pieces})\right] \cdot (\text{number of resolved bands})^2.$$

These last formulas assume satisfactory shaping of frequency cutoff, and constant resolution up to the maximum frequency of interest. Particular situations may deviate from this in either direction.

(There are no Sections B.25, B.26, B.27.)

B.28. *Analysis of Example C*

It is most desirable that an account of this sort include enough details of a numerical example to allow those readers who wish to do so, to follow through and check. This is impractical when even a few tenths (or even a few hundredths) of a million multiplications are involved. Example C, however, offers us an opportunity to present such details for one example, even if it is quite atypical.

We remarked in Section 28 that the add-and-subtract pilot estimation procedure of Section 18 *might* be applied to the second differences of Brouwer's values (themselves the differences ephemeris time *minus* mean solar time). During the 131 years from 1820 to 1950 astronomical techniques have improved, and observational errors seem, according to Brouwer's own analysis, to have somewhat decreased. In

TABLE VIII — SUMS OF SQUARED DIFFERENCES

	δ column	$\delta\sigma$ column	$\delta\sigma\sigma$ column	(*)	(†)
Block 1	8615	2083	321	−502	−971
Block 2	4130	1230	550	−9	−70
Block 3	3447	2591	273	1557	−244
Block 4	1553	304	516	−162	283
Noise	5/2	3/4	3/8	0	0
Flat	1	1	1	0.7	0.85

(*) Sum for $\delta\sigma$ column less 0.3 times sum for δ column.
(†) Sum for $\delta\sigma\sigma$ column less 0.15 times sum for δ column.

order to reflect this fact, and to provide some external estimate of error we shall study the second differences in 4 separate blocks, 4 separate time intervals of 32 consecutive years each. In view of the fact that Brouwer estimates the mean observational error to have decreased from 0.38 to 0.17 seconds of time over the period, it will clearly suffice to work in units of 0.1 second of time. Table VII shows, for the second time block, calendar dates, Brouwer's values, their first and second differences (the latter being the series we consider as the X_q) and the results of the add-and-subtract pilot estimation model. In this table, sums and differences are shown on lines half-way between the lines containing the entries from which they are formed. The last column shows the result of completing the δX_q column to a DX_q column. The resulting sums of squared differences are shown in Table VIII for each of the four blocks, together with comparative values.

These comparative values show the anticipated relative sizes of such sums of squared differences in case the spectral density were

(1) proportional to $(1 - \cos \pi f/f_N)^2 = (1 - u)^2$, the shape assumed for the "noise" component,

(2) flat, the shape assumed for the other component according to the second model,

where we have introduced u as an abbreviation for $\cos \pi f/f_N$. The successive columns have average sums of squared differences which are easily seen to be obtained by multiplication of the spectrum by (when we start with 32 X's):

$$16\,(2 - 2\cos \pi f/f_N) = 32\,(1 - u),$$

$$8\,(2 + 2\cos \pi f/f_N)\,(2 - 2\cos 2\pi f/f_N) = 64(1 + u - u^2 - u^3)$$
$$= 64\,(1 + u)\,(1 - u^2),$$

$$4\,(2 + 2u)\,(4u^2)\,(16\,u^2 - 16\,u^4) = 512(u^4 + u^5 - u^6 - u^7),$$

and integration.

When we recall that

$$\int_0^{f_N} u^k \, df$$

vanishes for odd k, and is equal to f_N, $f_N/2$, $3f_N/8$, $5f_N/16$ and $35f_N/128$ respectively, for $k = 0, 2, 4, 6$ and 8, we easily obtain the values given for comparison in Table VIII (after removing $32f_N$ as a common factor).

The last two columns of this table represent attempts to combine sums of squared differences so as to estimate the first component free of the "noise" component whose spectrum is proportional to $(1 - u)^2$. The attempts appear quite useless.

One reason for the lack of success is easy to find. There are only 4 values of $\delta\sigma\sigma$ to square and sum for each 32-year block. This means no more than 4 degrees of freedom, and, consequently very poor stability. We can partially correct this difficulty by going over to the modified add-and-subtract method in which all possible differences of a given sort are calculated.

Table IX presents the compact calculation for block 1, arranged as in Table VI. Here (σ) stands for first σ and then σ'. The mean squares for each column of differences lead to the results summarized in Table X for all four blocks. The difference estimates are now seen to be more stable and to increase somewhat with decreasing frequency, although not as much as would be expected even for a flat component.

The analysis is doing better, but is not yet satisfactory. One likely reason for this appears when we inquire what sort of spectral windows go with (*), (†) and (‡). The windows corresponding to $D(\sigma)$ and to (*) are shown in Fig. 24. The amount of negative area near $f/f_N = 1$ required to compensate for the rather large pickup of the "noise" component by the $D(\sigma)$ column is quite substantial, suggesting probable increased variability.

We can reduce our difficulties from such causes by using only slightly more complex processes. We can probably use the D column satisfactorily as an indication of the noise component.

We need to obtain two other composite measures of the spectrum, both of which avoid large values of f/f_N, one of which is concentrated near $f/f_N = 0$ and the other of which avoids $f/f_N = 0$. If we can obtain a smoothed and decimated sequence which avoids the upper part of the spectrum, then sums and differences of such values will have mean squares with the appropriate properties. The results of Section B.18 suggest trying $F_2S_2S_3$ as the operation generating the modified sequence, to which D (differencing) and S_2 are then to be applied.

TABLE IX

Compact calculation applied to second differences (in tenths of seconds of time) for block 1 (values in Brouwer's Table VIII(c) used where appropriate). (Entries arranged as in Table VI.)

X_q	D	(σ)	$D(\sigma)$	$(\sigma)^2$	$D(\sigma)^2$	$(\sigma)^3$	$D(\sigma)^3$	$(\sigma)^4$	$D(\sigma^4)$
−27		−6		−10		−10		−9	
	+48		2		10		11		13
+21		−4		0		+1		4	
	−27		4		−6		−5		
−6		0		−6		−4			
	+8		0		13		12		
+2		0		7		8		−3	
	−8		−1		−12				
−6		−1		−5					
	+12		−4		6				
+6		−5		1		−6		−4	
	+2		6		1		8		
+8		1		2		2			
	−16		5		4		1		
−8		6		6		3		5	
	+1		−10						
−7		−4							
	+13		3						
+6		−1		−4		−5		−7	
	0		3		3		3		
+6		2		−1		−2			
	−17		−3		−3		24		
−11		−1		−4		22		20	
	+16		23		6				
+5		22		2					
	−9		−42		−1				
−4		−20		1		−5		−2	
	+8		15		20		8		
+4		−5		21		3			
	−2		16		−46		−7		
+2		11		−25		−4		−1	
	+6								
+8									
	−20								
−12		15		11		10		26	
	+23		−19		−12		6		
+11		−4		−1		16			
	−23		18		7		−23		
−12		14		6		−7		9	
	+31		−29		4				
+19		−15		10					
	−36		27		−4				
−17		12		6		5		21	
	+27		−18		−19		11		
+10		−6		−13		16			
	−21		6		24		−18		
−11		0		11		−2		14	
	+16		10						
+5		10							
	+12		−11						

TABLE IX, CONTINUED

X_q	D	(σ)	$D(\sigma)$	$(\sigma)^2$	$D(\sigma)^2$	$(\sigma)^3$	$D(\sigma)^3$	$(\sigma)^4$	$D(\sigma^4)$
+17		−1		10		7		10	
	−45		8		−13		−4		
−28		7		−3		3			
	+36		−14		−3		−20		
+8		−7		−6		−17		−14	
	+6		1		15				
+14		−6		9					
	−33		−5		−9				
−19		−11		0		−9		0	
	+28		33		−17		18		
+9		22		−17		9			
	−7		−32		29		−14		
+2		−10		12		−5		4	

TABLE X — MEAN SQUARES FOR DIFFERENCE COLUMNS AND SUITABLE LINEAR COMBINATIONS THEREOF FOR THE ANALYSIS ILLUSTRATED BY TABLE IX

Block	D	$D(\sigma)$	$D(\sigma)^2$	$D(\sigma)^3$	(*)	(†)	(‡)
1	485	283	234	176	−8	−57	−115
2	252	207	206	317	56	55	166
3	207	258	145	148	134	21	24
4	76	41	80	106	−5	34	60
Ave. of 2, 3 and 4					51	38	78
Noise	5/2	3/2	3/2	3/2	0	0	0
Flat	1	2	4	8	1.4	3.4	7.4

(*) Mean square for $D(\sigma)$ less 0.6 times that for D.
(†) Mean square for $D(\sigma)^2$ less 0.6 times that for D.
(‡) Mean square for $D(\sigma)^3$ less 0.6 times that for D.

The corresponding windows are, for the square of a single value of $S_2(F_2)S_2S_3X_q$,

$$4u^2(2 + 2u)\,(1 + 4u + 4u^2) = 8(u^2 + 5u^3 + 8u^4 + 4u^5),$$

and, for the square of a single value of $D(F_2)S_2S_3X_q$,

$$(4 - 4u^2)(2 + 2u)\,(1 + 4u + 4u^2) = 8(1 + 5u + 7u^2 - u^3 - 8u^4 - 4u^5),$$

to which should be compared that for the square of a single value of DX_q

$$(2 - 2u) = 2(1 - u).$$

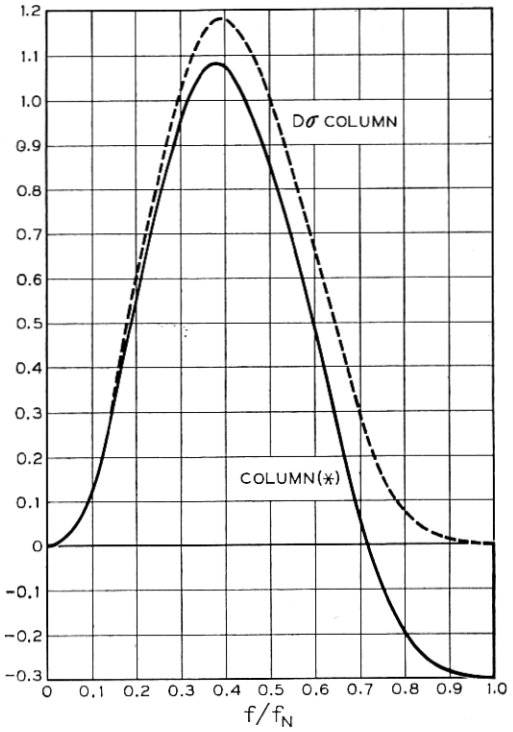


Fig. 24 — Windows corresponding to $D(\sigma)$ and to (*).

If now we form all values of $S_2(F_2)S_2S_3X_q$, $D(F_2)S_2S_3X_q$ and DX_q (Table XI illustrates the parallel calculation of F_2 and F_2' quantities), and form the corresponding mean squares, we obtain the results shown in Table XII.

If we compare the S-0.2δ column with the last column, we find a definite tendency for the last to be smaller. As indicated by the last two lines of the table, neither of these columns reflects, on the average, a "noise" component, while, on the average, both reflect a "flat" component equally. The natural conclusion is that the observed spectrum is more peaked toward $f/f_N = 0$ than would be expected from a mixture of "flat" and "noise" components. It might seem natural to conclude that the first model is to be preferred.

However, more careful examination shows that not only do the δ mean squares appear to decrease with improvements in astronomical technique, but (leaving aside block 1) the values in the S-0.2δ column are decreasing and those in the D-0.6δ column may well be doing the

TABLE XI

Calculation of $S_2(F_2)S_2S_3$ and $D(F_2)S_2S_3$ for the second differences of Brouwer's fluctuations in the earth's rotation. (Portion of block 2.)

Date	X_q	S_3X_q	$F_2S_2S_3X_q$	$DF_2S_2S_3X_q$	$S_2F_2S_2S_3X_q$	$F_2'S_2S_3X_q$	$DF_2'S_2S_3X_q$	$S_2F_2'S_2S_3X_q$
1854.5	−4							
1855.5	0	−6						
			−3					
	−2	+3						
				+10	+4	2		
	+5	−1						
			+7				−14	−10
	−4	+8						
				−26	−12	−12		
	+7	−20						
			−19				+5	−19
1860.5	−23	+1						
				+20	−18	−7		
	+17	−8						
			+1				+9	−5
	−2	+9						
				−4	−2	2		
	−6	−7						
			−3				+1	5
	+1	+4						
				+2	−4	3		
1865.5	+9	−1						
			−1				−11	−5
	−11	0						
				−17	−19	−8		
	+2	−8						
			−18				−12	−28
	+1	−10						
				+6	−30	−20		
	−13	−10						
			−12				18	−22
1870.5	+2	−2						
				15	−9	−2		
	+9	0						
			3				8	4
	−11	+3						
				14	20	6		
	+5	+3						
			17				21	33
	+9	+14						
				−4	30	27		
1875.5	0	+13						
			13				−23	31
	+4	0						
				−15	11	4		
	−4	+4						
			−2				−3	5
	+4	−6						
						1		
	−6	+7						
1880.5	9							

TABLE XII — MEAN SQUARES OF COLUMNS FOR SECOND DIFFERENCES OF BROUWER'S DATA

For	S	D	δ	S-0.2δ	D-0.6δ	2.5 (D-0.6δ)
Block 1	83	228	484	−14	−62	−155
Block 2	290	178	252	240	27	68
Block 3	153	168	207	112	44	110
Block 4	62	47	75	47	2	5
Average (2, 3, 4)	168	131	178	133	24	61
Noise	1	3	5	0	0	0
Flat	28	12	2	27.6	10.8	27.6

S = mean square value of $S_2(F_2)S_2S_3X_q$.
D = mean square value of $D(F_2)S_2S_3X_q$.
δ = mean square value of DX_q .

same. The suggestion is — which, after we have seen another block or two or three (of 32 years each), we may be able to confirm or deny — that the lower frequency component, as well as the noise component, is decreasing as astronomical technique improves. If this be true, then most of the low-frequency power in blocks 2 and 3 may represent observational and reductional sources of variation rather than changes in the rotation of the earth.

While more refined analyses might show something more, 131 values are not a great number, especially since technique has changed during their measurement, and it is likely that we shall have to wait a while for a more definite answer to this question.

This example was included, not because of the peculiar importance of the irregularities in the rotation of the earth, but rather to illustrate certain general points, particularly these:

(1) While careful spectral analysis requires a very considerable amount of arithmetic, there are situations where simple analysis will yield useful results. (All the values in this section were obtained with pen or pencil and an occasional use of a slide rule. The use of simple differencing and summing techniques to produce moderately sharp windows has been studied by the Labroustes.[41, 42, 43] Although they felt themselves interested in line spectra, their methods, properly reinterpreted, are easily and directly applicable to continuous spectra.)

(2) The windows corresponding to the add-and-subtract pilot analysis are broad and the estimation far from exhaustive. (The procedure is intended for use in planning prewhitening, not as a tool for answering questions of even moderate difficulty.)

(3) Subdivision of the data before analysis is quite often helpful.

(If blocks 2, 3, and 4 had been analyzed as one unit, it would have been easy to jump to a conclusion which now seems dangerous.)

(4) Spectral analysis can lead to results unsuspected before the calculations were made. (The decrease in amplitude of the low-frequency component which Brouwer's data now suggests was not at all suspected until the values given in Table XII were pulled together.)

Index of Notations

In case a notation is not widely used, the Sections in which it is used are specified in parentheses:

a = an integer (A.6 and B.19),

a_0, a_1, a_2 = real constants (B-8.4),

a_{i0}, a_{ij} = assorted constants (13 and B.5),

A = various constants (27, B.9, B.15),

$A\left(f; \frac{1}{\Delta t}\right)$ = infinite Dirac comb made up of unit δ-functions spaced $1/\Delta t$ in frequency (see A.2 or Table IV for formula),

b = an integer (A.6 and B.19),

B = auxiliary quantity defined in B-8.5 (B.8),

$B(t)$, $B_i(t)$, $\tilde{B}(t)$ = a data window (graded) (10 and B.10),

c = an integer (A.6 and B.19),

c_0, c_1, $\cdots$, c_k = real constants (B.15),

C_{ij} = autocovariance corresponding to t_i and t_j (1),

$C(\tau)$ = autocovariance at lag τ,

$C_{00}(\tau)$ = autocovariance estimated from record of finite length,

$C'_{00}(\tau)$, $\tilde{C}_{00}(\tau)$, etc. = hypothetical or actual analogs of $C_{00}(\tau)$,

$C_i(\tau) = D_i(\tau)C_{00}(\tau)$ = modified autocovariance estimate (defined in 4),

C_r = sample autocovariance at lag $r\Delta\tau$ (usually = $r\Delta t$ in practice),

d = an integer (A.6, B.18 and B.19),

D = the operation of differencing adjoining values,

$D_i(\tau)$ = a prescribed even function of time shift, often a lag window (cp. 4, 5, B.4, B.5),

$D_{ei}(\tau)$ = lag window equivalent of $B_i(\tau)$ (see B-4.11 for representation),

e = 2.71828182845 $\cdots$,

E_{kr} = k^{th} alternative correction at lag r (19, B.19),

f = frequency (in cycles per second where not otherwise specified), in equi-spaced discrete situations $f = r/(2m \cdot \Delta t) = (r/m)f_N$,

f_0 = a particular frequency (in Section B.18 the folding frequency of the cycle in question),

f_1 = a frequency, often the nominal frequency of a smoothed estimate,

f_N = Nyquist or folding frequency = $1/(2 \cdot \Delta t)$,

f_N* = effective Nyquist or folding frequency = $1/(2\cdot\Delta\tau)$, (13),
F_j = the operation of retaining only every j^{th} value ($j = 2, 3, \cdots$) (B.19, B.28),
F_2' = the operation of retaining the alternate values dropped by F_2 (B.19, B.28),
g = an index running from $-c$ to $+d$ (A.6 and B.19),
$G(t)$ = a function of the time, Fourier transform of $S(f)$ (correspondence also required with subscripts 0, 1, 2),
$\tilde{G}(t)$ = the Fourier transform of a particular box-car function (B.12),
h = an integer, generally satisfying $\Delta\tau = h\cdot\Delta t$, in A.6 and B.19 an index running through indicated ranges,
$H_i(f; f_1) = Q_i(f + f_1) + Q_1(f - f_1)$ = power transfer function (from power at absolute frequency f to estimate at nominal frequency f_1),
$H_i(f - f_1)$ = special form of $H_i(f; f_1)$,
$H_Y(f; f_1)$ = power transfer function defined in B-10.2,
$i = \sqrt{-1}$,
Subscript i, values often 0, 1, 2, 3, 4, usually identifies quantities or functions associated with i^{th} window pair,
j = an integer, often such that $3j - 2 = k$ or ℓ (B.19),
Subscript j, values often 0, 1, 2, 3, 4, alternative to subscript i,
k = usually number of equivalent degrees of freedom or number of elementary frequency bands; in Section 17, length of group averaged,
ℓ = an integer; in 18 an exponent of 2, in A.6 satisfying $n = 2\ell + 1$,
L = inductance (in example of 27),
m = integer, number of longest lag (longest lag = $m\cdot\Delta\tau$, usually = $m\Delta t$),
$M(t, \tau)$ = bilinear monomial in the X's
$= X(t + \tau/2)\cdot X(t - \tau/2)$,
n = usually one less than the number of discrete data points, in A.6 and B.19 an integer = $2\ell + 1$,
n' = effective length of record ($n'\cdot\Delta t = T_n'$),
p = usually the number of pieces of record; also = $i\omega$ (A.1 only); a real number (usually integral, A.6),
p_0 = a constant power density (B.9),
$p_0, p_1, p_2, \cdots$ = ordinates in general example (9),
$P(f)$ = density of power spectrum (normalized so that variance = $\int_0^\infty 2P(f)\,df$), also with various subscripts,
$P_a(f)$ = aliased density of power spectrum (periodic in frequency, cp. 12),
$P_A(f)$ = principal part of aliased density of power spectrum (confined to $|f| \leqq 1/(2\cdot\Delta\tau)$, cp. 12),
$P_Y(f_1)$ = estimate of spectral density based on filtered signal (B.10),

$P_{ei}(f)$ = spectral density estimated with the data window $B_i(t)$,
$P_{00}(f)$ = symbolic Fourier transform of $C_{00}(\tau)$,
$P_{0A}(f)$ = aliased estimate of smoothed $P(f)$,
$P_{i1}(f) = H_i(f; f_1)P(f)$, roughly a filtered version of $P(f)$ (subscript i interchangeable with j, also 1 with 2),
$P_{iA}(f_1)$ = aliased estimate of smoothing of $P(f)$ by $H_i(f; f_1)$,
$P_{iA1}(f) = H_i(f; f_1) \cdot P_A(f)$ = aliased filtered spectral density,
$P_{iAk}(f)$ = corrected aliased estimate of $P(f)$ smoothed by $Q_i(f)$,
$P_{\text{out}}(f; f_1) = |\, Y(f; f_1)\,|^2 \cdot P(f)$ = spectral density of output,
$\tilde{P}(f)$ = a perfectly known power spectrum (B.12),
$\tilde{P}_A(f)$ = either the aliased spectrum corresponding to $\tilde{P}(f)$ (B.12) *or* the aliased spectrum of the $\tilde{X}_q$ series (15, 17),
$\tilde{P}_{i1}(f)$ = an auxiliary quantity defined in B-8.4,
q = index running as indicated,
Q_r = an auxiliary quantity (B.19),
$Q_0(f)$ = Fourier transform of centered box-car function of length $2T_m$ (See also $Q_i(f)$ for i),
$Q_i(f)$ = Fourier transform of $\mathrm{R}_i(\tau)$ (see B.5 for $i = 0, 1, 2, 3, 4$ and B-8.7 for another choice),
$Q_0(f;\, \Delta\tau)$ = spectral window corresponding to (Fourier transform of) a discrete box-car function,
$= \Delta\tau \cdot \cos \tfrac{1}{2}(w \cdot \Delta\tau) \sin (mw \cdot \Delta\tau)$,
$Q_{ei}(f)$ = spectral window corresponding to (Fourier transform of) $D_{ei}(\tau)$,
$Q_{0A}(f)$ = aliased form of $Q_0(f) = Q_0(f;\Delta\tau)$,
$Q_{iA}(f)$ = aliased form of $Q_i(f)$,
$\tilde{Q}_0(f)$, $\check{Q}_i(f)$ = auxiliary quantities (B.8),
r = integer index almost always running from 0 or 1 to m or $m - 1$,
R = resistance (in example of Section 27),
$R_{ik}(f)$ = spectral window associated with $Q_i(f)$ and the sequence E_{k0}, E_{k1}, $\cdots$, E_{km},
S_j = operation of summing by (overlapping) sets of j each (B.17, B.28),
$S(f)$ = function of frequency, Fourier transform of $G(t)$ (also with subscripts 1, 2, $\cdots$),
$\tilde{S}(f)$ = a particular box-car function (B.12),
t = time in the sense of epoch (also with various subscripts),
T = a time, usually positive,
T_m = half-length of box-car function of time, *or* greatest lag used or considered,
T_n = length of record,
T_n' = effective length of record, approximately $T_n - \dfrac{p}{3} T_m$,

u = general real variable,
U_r = corrected estimate of smoothed power density (at nominal frequency $r/(2m \cdot \Delta t)$,
V_r = raw estimate of smoothed power density at nominal frequency $r/(2m \cdot \Delta t)$,
w = a chance quantity (B.6),
W = a frequency band width (B.9),
W_e = equivalent width, often of $P_{il}(f)$ defined in Section 8,
W_{main} = width of main lobe,
W_{side} = width of side (unsplit) lobe,
$W(t)$ = impulse response of linear transmission system,
$W(t; f_1)$ = in B.10 the impulse response of filter with transfer function $Y(f; f_1)$,
x = usually a real variable, in B.6 a chance quantity (random variable),
X_q = qth value of discrete, equi-spaced time series,
$X(t)$ = value of time function,
$\bar{X}$ = in Section 1 an average value (along an infinite function or across an ensemble), in 19 ff. the mean of the observed X_q's,
$\overline{X^+}$, $\overline{X^-}$ = means of end thirds of observed values,
$\tilde{X}_q$, $X_q{}^*$ = qth values of linearly transformed time series,
y = in Section B.6 a chance quantity (random variable),
$Y(f)$ = steady-state transfer function corresponding to linear transformation or linear transmission system,
z = in Section B.6 a chance quantity (random variable),
α = a real constant (15) *or* an indeterminate (B.8), *or* a certain fraction (B.18) *or* an unknown constant (B.19),
α_i = factor indicating extent of end effect losses (8, B.6) defined by B-6.9,
β = real constant (15) *or* an indeterminate (B.8) *or* a constant defined by B-8.8, *or* an unknown (slope) constant (B.19),
$\hat{\beta}$ = estimate of β(B.19),
γ = real constant (15),
$\Gamma(f)$ = power-variance spectral density,
$\Gamma_{\Delta t}(f)$ = power-variance spectral density in the equi-spaced discrete case,
δ = operation of forming alternate differences,
δ' = operation of forming alternate differences complementary to δ,
Δf = a change in frequency, sometimes the width $1/(2T_n')$ of an elementary frequency band,
Δt = time interval, usually that between data values,
$\Delta\tau$ = time interval, usually that between lags used ($=h \cdot \Delta t$),

ϵ = a small real number (A.2),
ξ = real, time-like variable of integration (A.3),
λ = real, time-like variable of integration (A.3) *or* a real constant (15),
μ = real constant (15),
ν = real constant (15),
ω = angular frequency (in radians per second unless otherwise specified) always = $2\pi f$ [with any sub- or superscripts],
$\varphi = 2\pi f T_m$, a normalized frequency (B.8),
$\Phi(f, \lambda)$ = an auxiliary function (B.6) (defined in B-6.3),
ψ = a phase angle (A.6),
χ_k^2 = a quantity distributed as chi-square on k degrees of freedom (9),
σ, σ' = complementary operations of summing by adjacent parts and then omitting every alternate sum,
τ = time difference or lag,
$*$ = sign of convolution,
superscript * = sign of complex conjugate (A.3),
$\nabla(t; \Delta t)$ = infinite Dirac comb approximating the constant unity [formula in A.2 or Table IV],
$\nabla_m(t; \Delta t)$ = finite Dirac comb approximating a unit-height, centered, box-car function [formula in A.2 or Table IV].

REFERENCES (PARTS I AND II)

1. N. R. Goodman, *On the Joint Estimation of the Spectra, Cospectrum and Quadrature Spectrum of a Two-Dimensional Stationary Gaussian Process*, Scientific Paper No. 10, Engineering Statistics Laboratory, New York University, 1957. (Ph.D. Thesis, Princeton University).
2. G. I. Taylor, Statistical Theory of Turbulence, Proc. Roy. Soc. (London), **A151,** pp. 421–478, 1935.
3. S. O. Rice, Mathematical Analysis of Random Noise, B.S.T.J., **23,** pp. 282–332, July, 1944; **24,** pp. 46–156, Jan. 1956. (This article also appears in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax, Dover Publ., New York, 1954, pp. 133–294.)
4. J. W. Tukey, The Sampling Theory of Power Spectrum Estimates, in *Symposium on Applications of Autocorrelation Analysis to Physical Problems*, Woods Hole, June 13, 1949. NAVEXOS-P-735, Office of Naval Research.
5. M. S. Bartlett, Periodogram Analysis and Continuous Spectra, Biometrika, **37,** pp. 1–16, 1950.
6. W. Marks and W. J. Pierson, The Power Spectrum Analysis of Ocean-Wave Records, Trans. Am. Geophysical Union, **33,** pp. 834–844, 1952.
7. H. Press and J. C. Houbolt, Some Applications of Generalized Harmonic Analysis to Gust Loads on Airplanes, J. Aero. Sciences, **22,** pp. 17–26, 1955.
8. H. A. Panofsky, Meteorological Applications of Power Spectrum Analysis, Bull. Am. Meteorological Soc., **36,** pp. 163–166, 1955.
9. G. P. Wadsworth, E. A. Robinson, J. G. Bryan, and P. M. Hurley, Detection of Reflections on Seismic Records by Linear Operators, Geophysics, **18,** pp. 539–586, 1953. (Also see later reports, Geophysical Analysis Group, M.I.T.)
10. M. S. Bartlett and J. Mehdi, On the Efficiency of Procedures for Smoothing Periodograms from Time Series with Continuous Spectra, Biometrika, **42,** pp. 143–150, 1955.

11. M. S. Bartlett, *An Introduction to Stochastic Processes*, Cambridge Univ. Press, Cambridge, England, 1955.
12. U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, Wiley, New York, 1957.
13. J. W. Tukey and C. P. Winsor, *Note on Some Chi-Square Normalizations*, Memorandum Report 29, Statistical Research Group, Princeton, 1949.
14. G. Groves, Numerical Filters for Discrimination Against Tidal Periodicities, Trans. Am. Geophysical Union, **36,** pp. 1073–1084, 1955.
15. H. R. Seiwell, The Principles of Time Series Analysis Applied to Ocean Wave Data, Proc. Nat. Acad. Sciences, **35,** pp. 518–528, Sept., 1949.
16. H. R. Seiwell and G. P. Wadsworth, A New Development in Ocean Wave Research, Science, **109,** pp. 271–274, 1949.
17. D. Brouwer, A Study of Changes in the Rate of Rotation of the Earth, Astronomical J., **57,** pp. 126–146, 1952.
18. Brouwer, pp. 135, 138.
19. G. A. Campbell and R. M. Foster, Fourier Integrals for Practical Applications, B.S.T.J., **7,** pp. 639–707, 1928; also, Van Nostrand, New York, 1942.
20. E. A. Guillemin, *Introductory Circuit Theory*, Wiley, New York, 1953, p. 257.
21. L. Schwartz, *Theorie des Distributions*, Vol. I and II, Hermann et Cie, Paris, 1950. (See also Halperin; Friedman; and Clavier, noted in Bibliography.)
22. E. A. Guillemin, *The Mathematics of Circuit Analysis*, Wiley, 1949, pp. 485–501.
23. H. Nyquist, Certain Topics in Telegraph Transmission Theory, Trans. A.I.E.E., pp. 617–644, April, 1928.
24. J. M. Whittaker, *Interpolatory Function Theory*, Cambridge Univ. Press, Cambridge, England, 1935, Chapter IV.
25. W. R. Bennett, Methods of Solving Noise Problems, Proc. I.R.E., **44,** 1956.
26. R. W. Hamming and J. W. Tukey, Measuring Noise Color, unpublished memorandum.
27. C. L. Dolph, A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beam Width and Side-Lobe Level, Proc. I.R.E., **34,** pp. 335–348, 1946.
28. P. Jaquinot, Quelques recherches sur les raies faible dans les spectres optiques, Proc. Phys. Soc. (London), **63B,** pp. 969–979, 1950.
29. P. Boughon, B. Dossier, and P. Jaquinot, Determination des fonctions pour l'amelioration des figures de diffraction dans le spectroscope, C. R. Acad. Sci. Paris, **223,** pp. 661–663, 1946.
30. L. Isserlis, On a Formula for the Product-Moment Coefficient of Any Order of a Normal Frequency Distribution in Any Number of Variables, Biometrika, **12,** pp. 134–139, 1918.
31. H. Hotelling, Relations Between Two Sets of Variates, Biometrika, **28,** pp. 321–377, 1936.
32. H. E. Kallmann, Transversal Filters, Proc. I.R.E., **28,** pp. 302–310, 1940.
33. H. Wold, *A Study in the Analysis of Stationary Time Series*, 2nd Edition, Almquist and Wiksell, Stockholm, 1954.
34. Daniel B. DeLury, *Values and Integrals of the Orthogonal Polynomials up to* $n = 26$, Toronto, University Press, 1950.
35. Ronald A. Fisher and Frank Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Edinburgh, Oliver and Boyd, 4th Ld. 1953, (especially Table XXII).
36. William E. Milne, *Numerical Calculus*, Princeton University Press, 1949.
37. Cornelius Lanczos *Tables of Chebyshëv Polynomials* $S_n(x)$ and $C_n(x)$, Nat'l. Bur. Stds. App'd. Math. Series 9, Washington, Gov't Print. Off., 1952, (especially pp. xiv to xvi).
38. Cornelius Lanczos, Trigonometric Interpolation of Empirical and Analytical Functions, J. Math. Phys., **17,** pp. 123–199, 1938.
39. Robert C. Minnick, Tshebycheff Approximations for Power Series, J. Assoc. Comp. Machinery, **4,** pp. 487–504, 1957.
40. J. C. P. Miller, Two Numerical Applications of Chebyshëv Polynomials, Proc. Roy. Soc. Edinb., **62,** pp. 204–210, 1946.
41. H. Labrouste and Y. Labrouste, Harmonic Analysis by Means of Linear Combinations of Ordinates, J. Terr. Magnetism, **41,** 1936.

42. H. Labrouste and Y. Labrouste, Analyse des graphiques resultant de la superposition de sinusoides, Mem. Acad. Sci. Inst. France, (2) **64,** No. 5, 1941. (See also Ann. de l'Inst. de Phys. du Globe de Univ. de Paris, **7,** pp. 190–207; **9,** pp. 99–101; **11,** pp. 93–101; **14,** pp. 77–91, for much of this material.)
43. H. Labrouste and Y. Labrouste, *Tables Numeriques precedes d'un expose de la methode d'analyse par combinaison d'ordonnees*, Presses Univ., Paris, 1943.

BIBLIOGRAPHY

In addition to the articles cited in the text, the reader's attention is called to the additional titles below.

P. Boughon, B. Dossier, and P. Jaquinot, Apodisation des raies spectrales au moyen d'ecrans absorbents, J. Recherches C.N.R.S., No. **11,** pp. 49–69, 1950.
P. A. Clavier, Some Applications of the Laurent Schwarz Distribution Theory to Network Problems, *Proceedings of the Symposium on Modern Network Synthesis, II,* Polytechnic Institute of Brooklyn, 1955.
L. Couffignal, La Methode de H. Labrouste pour la recherche des periods, Ciel et Terre, **66,** pp. 78–86, 1950.
B. Friedman, *Principles and Techniques of Applied Mathematics*, Wiley, New York, 1956, Chapter 3.
I. Halperin, *Introduction to the Theory of Distributions*, University of Toronto Press, Toronto, 1952.
B. Mazelsky, Extension of the Power Spectrum Methods of Generalized Harmonic Analysis to Determine Non-Gaussian Probability Functions of Random Input Disturbances and Output Responses of Linear Systems, J. Aeronautical Sciences, **21,** pp. 145–153, March, 1954.
H. A. Panofsky and I. van der Horen, Spectra and Cross-Spectra of Velocity Components in the Mesometeorological Range, Quart. J. Roy. Meteorol. Soc., **81,** pp. 603–605, 1955.
H. A. Panofsky and R. A. McCormack, The Vertical Momentum Flux at Brookhaven at 109 Meters, Geophys. Res. Papers (USAF, Cambridge Research Center), No. **19,** pp. 219–230, 1952.
H. Press and J. W. Tukey, Power Spectral Methods of Analysis and Their Application to Problems in Airplane Dynamics, *Flight Test Manual, NATO, Advisory Group for Aeronautical Research and Development,* IV-C, pp. 1–41, June, 1956. (Reprinted as Bell System Monograph No. 2606.)