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

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 II of this article will appear in the March issue of THE JOURNAL.

TABLE OF CONTENTS

I. INTRODUCTION

Communications systems and data-processing systems are generally required to handle a large variety of signals in the presence of noise. The design of these systems depends to a large extent upon the statistical properties of both the signals and the noise. In most cases, the noises may be represented, or approximated, as stationary Gaussian random processes with zero averages, so that all of their relevant statistical properties will be contained by the autocovariance function or the power spectrum. In many cases, the signals may also be represented, or approximated, as stationary Gaussian random processes with zero averages.

Noises, signals, or other ensembles of functions (given continuously or at intervals) which are approximately stationary but not Gaussian are often also usefully studied in terms of autocovariance functions or power spectra. Although the average and the spectrum are no longer the *only* relevant statistical properties, they are usually the most useful ones. Thus, we shall do well to keep as much of our treatment generally applicable as possible.

In almost every case, the autocovariance function or power spectrum of either the noise or the signal will be of interest and importance.

To determine the autocovariance function or power spectrum of an (approximately) stationary random process, we are often reduced to the necessity of measurement and computation. Exact determination

would require a perfectly-measured, infinitely-long piece of a random function (or a collection of pieces of infinite total length), and would require infinitely detailed computations. Both of these requirements are, of course, impractical. Approximate determination, on the other hand, raises the questions of how much data of a given accuracy will be required, what computational approach should be used, and how much reliance may be placed upon the results. Practically useful answers to these questions may be found by combining results from transmission theory and the theory of statistical estimation. These answers prove to be relatively simple. The only major difficulty in their practical application is the extensiveness of the data required for highly precise estimates. This requirement is an inherent, irrevocable characteristic of such random processes.

In this account we shall treat only the measurement of spectra of individual noises or signals. The measurement and utilization of the cross-spectra of pairs of series is also important, but is beyond our present scope. Questions of distribution and anticipated variability of cross-spectral estimates, and of certain estimates derived from them, have recently been cleared up by Goodman.[1]

It is natural to feel that the measurement of power spectra is simple, and that no problems deserving extended discussion arise. After all, are there not commercial "wave analyzers" of many sorts; have not Fourier series served for many years to analyze the frequencies of many signals, (musical instruments, human voices, etc.)? Why should there be a serious problem?

There are two reasons why elementary methods fail us rather frequently. On the one hand, the signal may not be available in indefinitely long time stretches. Either the conditions of observation, experimental or otherwise, or the difficulties of careful recording may make it impractical to have so much data that we can afford to analyze carelessly. (The examples of Sections 26 to 28, involving spectra of radar tracking, noise in very short-lived devices, and irregularities in the earth's rotation, respectively, all illustrate this point). Even if observation and recording can be afforded, the cost of computation often forces careful analysis.

On the other hand, the *random* nature of much noise, and some signals, in which the relative amplitudes and phases of different frequencies are not stably related (in contrast to voices and musical notes), introduces much more difficulty with sampling fluctuations and provides much more significant appearing, thus much more misleading statistical artefacts than experience with simpler signals would lead investigators

to expect. (In postwar oceanography, for example, high mechanical ingenuity was expended in the construction of simple and effective wave analysers to produce detailed spectra of ocean waves. The results were quite misleading, because the frequency resolution obtained was too high for the limited length of records used, and almost the entire appearance of the resulting spectra was an illusion due to the particular fluctuations of the particular record. The use of broader filters has since led to meaningful results which could be related to physically satisfying theories.) All too often, the practical study of spectra requires care.

Effective measurement of power spectra requires understanding of a number of considerations and action guided by all of them. Explaining each individual consideration is necessary, but it is equally necessary to explain how they fit together. The general structure of this description of spectral measurement is the following: an introduction to the concepts (Sections 1–3), brief accounts of individual considerations (Sections 4–19), accounts of how these considerations are assembled in analysis (Sections 20–21), and planning for measurement (Sections 22–28, which include discussion of examples), and Sections in Part II giving the details supporting the earlier sections.

We have attempted to provide, somewhere, most of the facts and attitudes that are needed in the practical analysis of (single) power spectra.

Readers interested in either completing their present knowledge or in gaining a brief overview of the subject may wish to proceed next to Sections 20ff, whence they can be referred to specific sections of interest. For some, reading of Sections 1–3 may be a helpful preliminary for Sections 20ff. For others, who want to build more solidly as they go, reading straight through, perhaps with considerable cross-reference to Part II, may be best.

A function of time $X(t)$ generated by a random (or stochastic) process is one of an ensemble of random functions which might be generated by the process. The value of the function at any particular point in time is thus a random variable with a probability distribution induced by the ensemble. Furthermore, the values of the function at any particular set of points, say $t = t_1, t_2, \cdots, t_n$, have an n-dimensional joint probability distribution also induced by the ensemble. Such probability distributions have an important bearing on the design of any communication system or data-processing system which must handle an output from such a random process, be this output "signal" or "noise".

We shall often, but not always, assume that the random process is *Gaussian*. This means that, for every n, $t_1, t_2, \cdots, t_n$, the joint probability distribution of

$$X(t_1), X(t_2), \cdots, X(t_n),$$

is an n-dimensional Gaussian or normal distribution. Each such distribution is completely determined by the ensemble *averages*

$$\bar{X}(t_i) = \text{ave } \{X(t_i)\},$$

and by the *covariances*

$$\begin{aligned} C_{ij} &= \text{cov } \{X(t_i), X(t_j)\} \\ &= \text{ave } \{[X(t_i) - \bar{X}(t_i)]\cdot[X(t_j) - \bar{X}(t_j)]\} . \end{aligned}$$

As a matter of convenience in development we will assume that the averages $\bar{X}(t_i)$ are zero. The covariances then reduce to

$$C_{ij} = \text{ave } \{X(t_i)\cdot X(t_j)\}.$$

Throughout, we will assume that the random process is *stationary* (that is, temporally homogeneous) in the sense that it is unaffected by translations of the origin for time. The covariances C_{ij} now depend only on the time separation $t_i - t_j$ so that

$$C_{ij} = C(t_i - t_j).$$

Thus, the noise is completely specified by a single function of a single variable. In particular, $C(0)$ is the *variance* (for zero average, the average square) of $X(t)$.

If the process were stationary, with zero averages, but were not Gaussian, then knowledge of the covariance as a function of lag, although providing a very large amount of useful information, would not completely specify the process. The results of this paper fall into two categories: (i) those relating to average values of spectral estimates, and (ii) those relating to variability of spectral estimates. The average-value results apply generally under the assumptions of stationarity (and zero averages), and do not depend upon the Gaussian assumption. The variability results are exact under the Gaussian assumption, and are usually rather good approximations otherwise. Thus, our results have practical value for noises and signals which are not closely Gaussian.

Results about variability are naturally used: (i) for planning the approximate extent of measurement effort, (ii) for indicating the presence of changes, during a series of measurements, in the quantities estimated, and (iii) as a means of judging the precision of an over-all estimate. The results given here are mainly for the first planning use. The additional uncertainties in actual variability due to either non-normality of distri-

bution, or to changing of conditions between runs, or to both, are often all too real, but are rarely large enough to affect planning seriously. The same is true of mild nonstationarity. The Gaussian, stationary results can also be applied to the second use, the detection of changes in the true spectrum, but considerable caution is in order. The precision of final over-all values is ordinarily far more wisely judged from the observed consistency of repeated measurements (as by analysis of variance of logarithms of various spectral density estimates at the same nominal frequency) than from any theoretical variability based on a Gaussian assumption.

Communications engineers are more accustomed to work with a single time function of infinite extent than with an ensemble of finite pieces (of such functions). It is perhaps fortunate, therefore, that averages across an ensemble are equivalent (ergodicity) to averages over time along a single function of infinite extent, whenever a process is stationary, Gaussian, has zero averages, and has a continuous power spectrum (no "lines"). (If the process were not stationary the single function approach could not be used in this way.)

Since we seek to make this account as intuitive as possible for communications engineers, we shall define transforms, and make many other computations in terms of averages along a single function (as limits of integrals over centered intervals). In dealing with more specifically statistical issues, however, we shall write "ave" for average value, "var" for variance and "cov" for covariance, and shall do nothing to hinder the interpretation of these operators as acting across the ensemble. (Those who wish can also think of them in single function terms.)

The covariance at lag τ, in single function terms, is given by

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

In ensemble terms, we would write merely

$$C(\tau) = \text{ave}\,\{X(t)\cdot X(t+\tau)\}.$$

The function $C(\tau)$ is frequently called the autocorrelation function, although historical usage in both statistics and the theory of turbulence (Taylor[2]) shows that this name should be applied to the (normalized) ratio $C(\tau)/C(0)$. We shall call $C(\tau)$ the *autocovariance* function. This name is appropriate to our formal definition of $C(\tau)$ because we have assumed that the averages of our process are all zero. Whenever we give up the assumption of zero averages, as we must almost always do when dealing with actual data, we shall use

$$\text{ave } \{X(t)\cdot X(t+\tau)\} = \text{average lagged product,}$$

$$\text{ave } \{[X(t) - \bar{X}]\cdot[X(t+\tau) - \bar{X}]\} = \text{autocovariance function,}$$

where $\bar{X}$ is the common value of ave $\{X(t)\}$ and ave $\{X(t+\tau)\}$, thus preserving accurate usage.

Because of the direct relationship of the joint probability distribution to the autocovariance function, much of the statistical attention given to Gaussian stationary time series (time-sampled random functions) has been expressed in terms of serial-correlation coefficients (corresponding to lag-sampled autocovariance functions).

A stationary Gaussian random process may be regarded (e.g. Rice[3]) as the result of passing *white* Gaussian noise through a fixed linear network with a prescribed transmission function. White Gaussian noise, in turn, may be regarded as the superposition of the outputs of a set of simple harmonic oscillators (continuously infinite in number) with

(a) a continuous distribution in frequency,*

(b) uniform amplitude over the significant frequency range of the transmission system, and

(c) independent and random phases.

This point of view is particularly suited to the techniques employed by communications engineers. It is therefore not surprising that communications engineers have dealt with stationary Gaussian random processes almost entirely in terms of power spectra.

Because the autocovariance function and the power spectrum are Fourier transforms of each other, it would at first appear to be purely a matter of convenience which one is used in any particular situation. Indeed, optimum filter characteristics for the protection of signal against noise in communications systems and in many types of computing devices have, on occasion, been determined by the use of the autocovariance function. In practice, however, the filter designer almost invariably turns to the power spectrum as the final criterion of adequate design and performance.

In practice also, where the autocovariance function or the power spectrum must be determined by measurement and computation, and then interpreted, the choice is now heavily weighted in favor of interpretation of the power spectrum. Although a great deal of theoretical work has been done on the probability distribution of the serial-correlation coefficients for Gaussian stationary time series of finite length, with a view to the estimation of the confidence which may be placed upon practical

* The term "frequency" is used throughout this paper in the communications engineer's sense, viz., cycles per second of a sinusoidal wave. (Exceptional uses in the statistician's sense are explicitly noted.)

results, the criteria which have been developed along this line are so complicated that it is extremely difficult to apply them in practice, where the joint distribution must be considered. On the other hand, the situation with respect to the power spectrum is now very satisfactory for practical purposes. This stems from results obtained by Tukey,[4] and in part independently by Bartlett,[5] about nine years ago, when studies were made of the effects of sampling, of finite length of series, and of choice of computational procedure on the behavior of the estimated power spectrum. Since that time, applications to such diverse fields as ocean waves (Marks and Pierson[6]), aerodynamics (Press and Houbolt[7]), meteorology (Panofsky[8]), and seismology (Wadsworth, Robinson, Bryan, and Hurley[9]), have shown the practical applicability of these results to a wide variety of physical time series.

Shortly after these studies first reached the stage of practical usefulness, the theoretical analysis was reformulated by Blackman, who expressed it from the point of view of transmission theory, for presentation to members of the technical staff of Bell Telephone Laboratories (Out-of-Hours Courses 1950–1951, Communications Development Training Program 1950–1952).

More recent contributions (1950–1957) to the theory of power spectrum estimation have been reviewed by Bartlett and Medhi,[10] by Bartlett,[11] and by Grenander and Rosenblatt.[12]

2. AUTOCOVARIANCE FUNCTIONS AND POWER SPECTRA

First, let us consider the ideal case. The autocovariance function which was defined in the preceding section by

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

may be reduced to the form

$$C(\tau) = \int_{-\infty}^{\infty} P(f)\cdot e^{i2\pi f\tau}\, df$$

where

$$P(f) = \lim_{T\to\infty} \frac{1}{T} \left| \int_{-T/2}^{T/2} X(t)\cdot e^{-i2\pi ft}\, dt \right|^2$$

(cp. Section B.2). The function of frequency $P(f)$ describes the *power spectrum* of the stationary random process considered. More precisely $P(f)\, df$ represents the contribution to the *variance* from frequencies between f and $(f + df)$. If we think of $X(t)$ as a voltage across (or current

through) a pure resistance of one ohm, the long-time average power dissipated in the resistance will be strictly proportional to the variance of $X(t)$. This important special case is the excuse for the adjective "power". The pure statistician might prefer to refer to the covariance spectrum or to the second moment spectrum rather than to the power spectrum. For precision, we shall often refer to $P(f)$ as the *spectral density* or *power spectral density*. When no confusion is likely, we may call $P(f)$ merely the *power spectrum*.

The relation exhibiting the autocovariance function as the Fourier transform of the power spectrum may be inverted to express the power spectrum as the Fourier transform of the autocovariance function. Thus, we have

$$P(f) = \int_{-\infty}^{\infty} C(\tau) \cdot e^{-i2\pi f\tau} \, d\tau \, .$$

The autocovariance function $C(\tau)$ and the power spectrum $P(f)$ are, formally at least, even functions of their respective arguments. Hence, the relation between them may be expressed more simply as two-sided cosine transforms, viz.

$$C(\tau) = \int_{-\infty}^{\infty} P(f) \cdot \cos 2\pi f\tau \cdot df,$$

and

$$P(f) = \int_{-\infty}^{\infty} C(\tau) \cdot \cos 2\pi f\tau \cdot d\tau;$$

or perhaps even more simply, as one-sided cosine transforms, viz.

$$C(\tau) = 2 \int_{0}^{\infty} P(f) \cdot \cos 2\pi f\tau \cdot df$$

and

$$P(f) = 2 \int_{0}^{\infty} C(\tau) \cdot \cos 2\pi f\tau \cdot d\tau.$$

Results are usually more conveniently developed in terms of the two-sided forms than in terms of the one-sided forms. In Sections A.3 and B.4 for example, the use of the two-sided forms with exponential kernels will be found to simplify considerably the expression of the operation of convolution between functions of lag or of frequency. In Section B.6, the use of the two-sided forms with exponential kernels avoids some complicated manipulations of trigonometric identities in the early stages of the development.

It should be particularly noted that

$$\text{ave } \{X(t) \cdot X(t + \tau)\} = \int_0^\infty 2P(f) \cdot \cos 2\pi f\tau \cdot df$$

and that (setting $\tau = 0$)

$$\text{var } \{X(t)\} = \int_0^\infty 2P(f) \cdot df.$$

Thus, it is evident that our definition of the power spectrum differs from the usual one which associates the power spectrum only with positive frequencies. References to the power spectrum in practice are usually in terms of a density $2P(f)$ over $0 \leqq f < \infty$ only.

3. THE PRACTICAL SITUATION

In practice we can obtain only a limited number of pieces of $X(t)$ of finite length. Each piece may be regarded as a sample drawn from a population or universe of pieces of $X(t)$ of the same length. The reduction of the data will therefore yield no more than estimates of the autocovariance function and of the power spectrum — estimates which are subject to sampling variations and to biases in the usual statistical sense. This situation is further complicated in those cases in which we can measure, or desire to use, only values of $X(t)$ at uniformly spaced values of t within each piece of $X(t)$; in other words, those cases in which we are dealing with classical time series (discrete time) rather than with time functions (continuous time).

The theoretical study of sampling variability and bias is much simpler in the case of the estimates of the power spectrum than in the case of the estimates of the autocovariance function (or of serial-correlation coefficients). This reflects the fact that, as we consider longer and longer records, and two narrow frequency bands with an arbitrarily small but fixed separation, we may find estimates of the power in these frequency bands which both (i) become arbitrarily precise, and (ii) become arbitrarily nearly (statistically) independent. The existence of such estimates is another particular consequence of the Gaussian character, as expressible in terms of "random and independent phases", of the random process from which we have one or more samples.

Use of the power spectrum has an additional advantage over use of the autocovariance function. In almost all practical situations, the data analyzed does not represent the actual output of the random process. In such cases the data will have been modified, appreciably if not radically, by the transmission characteristics of the devices employed to se-

cure the data. This modification of the data may in fact be intentional, as we shall see when we come to the discussion of "prewhitening" in Section 15. In any case, the estimates will have to be corrected for the effects of this modification of the data. For estimates of the power spectrum, the correction procedure is a simple division of a frequency function by another frequency function. For estimates of the autocovariance function, however, the correction procedure will require a Fourier transformation, division of the resulting frequency function by another frequency function, and an inverse Fourier transformation. This whole sequence of operations on the autocovariance function is the only practical procedure for the inversion of the convolution (see Appendix A.3) which is the effect to be corrected for. (Details are given at the end of Section B.3.)

As we shall see, the measurements and computational operations may involve the use of either analog or digital computation and handling of either continuous "signals" or discrete data. (Whatever be its relation to some communication or data-handling system, we shall call continuous-time signals or noise which we are analyzing "signals", while discrete-time signals or noise, or discrete-time samples thereof, will be called data.) In actual practice, and for well-defined reasons of instrumentation and computation engineering, only a few of the many possible combinations are used.

Spectrum analysis by analog computation is almost always applied to continuous "signals", and makes use of filtering rather than going through autocovariance or mean lagged products. Digital computation must be carried out on discrete data, perhaps time-sampled from a continuous "signal", and preferably uses an indirect route via mean lagged products rather than trying to isolate individual frequency bands directly. In either case, each data value must enter several computations, and it is rarely economic to carry these computations out directly in real time, especially since there will not usually be enough such analysis on a regular basis to saturate the working capacity of the analog or digital computer used. Consequently, recording, either of "signals" or of data or of both, is almost inevitable.

Thus, five stages will be important in nearly every case:

(1) sensing (pick-up, conversion, etc.)
(2) transmission (to recorder or, possibly, to computer)
(3) recording (including play-back, and, perhaps, time-sampling)
(4) computation (formulas, computing circuit performance, etc.)
(5) interpretation.

In every one of these stages, quality of performance (noise level, distortion, etc.) will be of importance.

The present account concentrates on the computational and interpretational stages, but indicates, from time to time, those considerations in the other stages which are peculiar to power spectrum analysis.

We have been unable to find a wholly satisfactory arrangement for the material we wish to present. In order to facilitate a relatively easy once-over, these introductory sections now continue into a condensed account, from which proofs, some reasons, and many helpful remarks have been postponed to the Appendix and sections in Part II. Readers interested in a survey may find it adequate to read only the condensed account. Others may find it best to skim this condensed account first, to read Appendix A next, and then to study similarly numbered sections of Part II and the condensed account.

The continuous record of finite length will be treated first (Sections 4–11); the modifications required for the discrete equally spaced record are covered next (Sections 12–21), and the opening account concludes with a discussion of the planning and analysis of measurement programs (Sections 22–28).

Appendix A (Sections A.1 to A.6) treats fundamental Fourier techniques, and the transform-pairs most closely associated with diffraction, in both the continuous and equi-spaced cases.

Each section of Part II relates to the similarly numbered section of the main body, and contains details of derivations, further reasons, and additional helpful remarks.

Definitions of the technical terms, arranged alphabetically for reference, are included at the end of Part I. Similar definitions of the notation will be given at the end of Part II.

Continuous Records of Finite Length

4. FUNDAMENTALS

Given a continuous record of finite length, it is clear that we cannot estimate the autocovariance function $C(\tau)$ for arbitrarily long lags. Surely, no estimate can be made for lags longer than the record. Furthermore, as we will find in due course, it is usually not desirable to use lags longer than a moderate fraction (perhaps 5 or 10 per cent) of the length of the record. Thus, in place of

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

for all values of τ, we will have at our disposal

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

only for $| \tau | \leqq T_m < T_n$, where T_n is the length of the record, and T_m is the maximum lag which we desire to use. We will call $C_{00}(\tau)$ the *apparent autocovariance function*, since (on account of ergodicity) its average value is $C(\tau)$ for $| \tau | \leqq T_m$.

The class of estimates for the power spectrum with which we are chiefly concerned will be derived from a *modified apparent autocovariance function* by Fourier transformation. While the modified apparent autocovariance functions, which are obtained by multiplying the apparent autocovariance function by suitable even functions of τ, are often far from being respectable estimates of the true autocovariance function, their transforms are very respectable estimates of *smoothed* values of the true spectral density.

Let $D_i(\tau)$ be a prescribed even function of τ, subject to the restrictions $D_i(0) = 1$, and $D_i(\tau) = 0$ for $| \tau | > T_m$, (where $i = 0, 1, 2, 3, 4$, depending upon the shape of $D_i(\tau)$ for $| \tau | < T_m$), and let the corresponding modified apparent autocovariance function be defined by

$$C_i(\tau) = D_i(\tau) \cdot C_{00}(\tau).$$

We may regard $D_i(\tau)$ as a window of variable transmission which modifies the values of $C_{00}(\tau)$ differently for different lags. It is therefore natural to call $D_i(\tau)$ a *lag window.*

For any lag window which meets the conditions stated above, $C_i(\tau)$ is calculable from the data. Further, it is clear that $C_i(\tau) = 0$ for $| \tau | > T_m$ although $C_{00}(\tau)$ was not defined there. Because $C_i(\tau)$ is defined for all values of τ, it has a perfectly definite Fourier transform $P_i(f)$, which should satisfy the symbolic relation,

$$P_i(f) = Q_i(f) * P_{00}(f)$$

where $Q_i(f)$ is the Fourier transform of $D_i(\tau)$, the asterisk indicates convolution (see Appendix A.3 for discussion), and $P_{00}(f)$ is the Fourier transform of $C_{00}(\tau)$. However, $P_{00}(f)$ is not determinate because $C_{00}(\tau)$ is not specified for $| \tau | > T_m$ (and its definition cannot be directly extended beyond $| \tau | = T_n$). Nevertheless, since

$$\text{ave } \{C_i(\tau)\} = D_i(\tau) \cdot C(\tau)$$

where $C(\tau)$ is the true autocovariance function, it follows that

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

where $P(f)$ is the true power spectrum, that is, the Fourier transform of $C(\tau)$. The average may be thought of as either across the ensemble, or along time. (The latter type of averaging would correspond to replacing

$X(t)$ by $X(t - \lambda)$, thus changing the stretch of $X(t)$ which is observed, and then averaging over λ.) The corresponding explicit relation, viz.

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

exhibits the average value of $P_i(f_1)$ as a smoothing (average-over-frequency) of the true power spectrum density $P(f)$ over frequencies "near" f_1 with weights proportional to $Q_i(f_1 - f)$. In a manner of speaking $P_i(f_1)$ is the collected impression of the true power spectrum $P(f)$ obtained through a window of variable transmission $Q_i(f_1 - f)$. It is therefore natural to call $Q_i(f)$ the *spectral window* corresponding to the *lag window* $D_i(\tau)$.

The form just given for ave $\{P_i(f_1)\}$ is natural for our two-sided definition of power spectra, but, in order to view the result from the standpoint of transmission theory for real-valued signals, it is convenient to express the result in a form appropriate to a one-sided definition of power spectra. Taking advantage of the fact that $Q_i(f)$ and $P(f)$ are even functions, we may write

$$\text{ave } \{2P_i(f_1)\} = \int_0^{\infty} H_i(f; f_1) \cdot 2P(f) \cdot df$$

where

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

and where we recall that $2P(f)\ df$ is the amount of power between f and $(f + df)$ in the one-sided true power spectrum. Similarly, $2P_i(f)\ df$ is the amount of power between f and $(f + df)$ in the one-sided estimated power spectrum. The function $H_i(f; f_1)$ has one of the necessary properties of a physically realizable power transfer function inasmuch as it is an even function of f as well as of f_1. In general, however, it does not have the property of being non-negative at all frequencies f. Nevertheless, it is a convenient function to use in the analysis of the variability of the estimated power spectrum. It will be convenient to regard the average value of the smoothed power density estimate ave $\{2P_i(f_1)\}$ as the result of passing the true power spectrum, through a "network" with *power transfer function* $H_i(f; f_1)$.

We see that our procedures will lead us to estimates whose average values are a smoothing (average-over-frequency) of the true power spectral density $P(f)$ over frequencies "near" f_1, and not to estimates of $P(f_1)$ itself. The problem of choosing the shape of the lag window $D_i(\tau)$ so that its Fourier transform $Q_i(f)$ will be concentrated near $f = 0$ is

almost identical to the problem of choosing an intensity distribution along an antenna so that most of the radiation from the antenna will fall in a narrow beam. From this analogy we will use such terms as *main lobe* and *side lobes* for the principal maximum and subsidiary extrema of $Q_i(f)$. (Indeed, any attempt to confine the power transfer function to too narrow a frequency band — too narrow in comparison with the reciprocal of the longest lag used — would be analogous to an attempt to construct a practical hyperdirective antenna.)

It is not surprising that we are led to estimate a smoothed power spectrum. With only a finite length of $X(t)$ available, we should not expect to be able to identify frequencies exactly, and are, indeed, unable to do so. (The presence of neighboring frequencies with random phases will have effects similar to those of noise in preventing such identification.)

5. TWO PARTICULAR WINDOW PAIRS

In order to specify a particular family of estimates within the class of estimates defined in the preceding section, we have only to specify $D_i(\tau)$ or $Q_i(f)$. We would like to concentrate the main lobe of $Q_i(f)$ near $f = 0$, keeping the side lobes as low as feasible. In order to concentrate the main lobe we have to make $D_i(\tau)$ flat and rather blocky. In order to reduce the side lobes, however, we have to make $D_i(\tau)$ smooth and gently changing. Since $D_i(\tau)$ must vanish for $|\tau| > T_m$ we must compromise. So far, cut-and-try inquiry has been more powerful in finding good compromises than has any particular theory.

A simple and convenient compromise is represented by the lag window (whose use is called "hanning")

$$\begin{aligned} D_2(\tau) &= \frac{1}{2}\left(1 + \cos\frac{\pi\tau}{T_m}\right) && \text{for} \quad |\tau| < T_m \\ &= 0 && \text{for} \quad |\tau| > T_m . \end{aligned}$$

(Window pairs 0 and 1 are discussed in Section B.5.) An alternative compromise is represented by the lag window (whose use is called "hamming")

$$\begin{aligned} D_3(\tau) &= 0.54 + 0.46\cos\frac{\pi\tau}{T_m} && \text{for} \quad |\tau| < T_m \\ &= 0 && \text{for} \quad |\tau| > T_m . \end{aligned}$$

These lag windows and the corresponding spectral windows are illustrated in Fig. 1. Notice that the main lobes are four times as wide as the

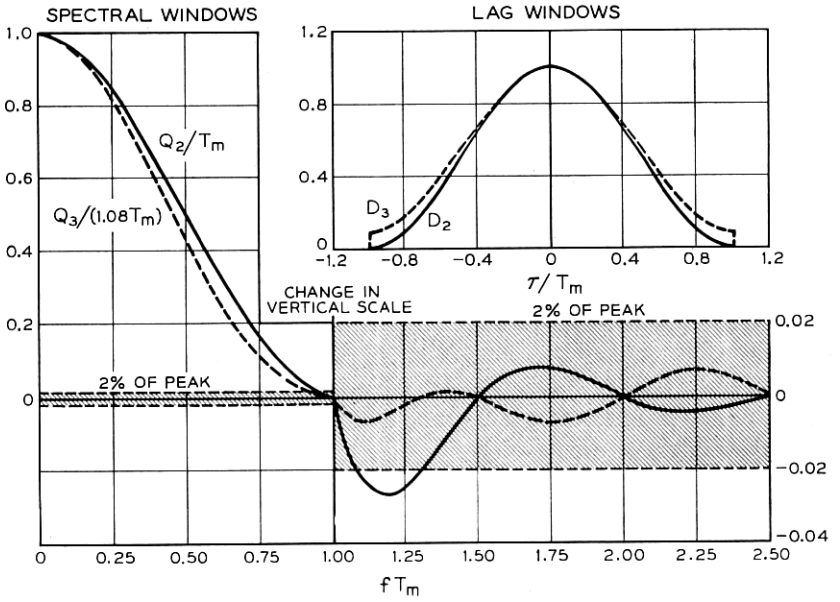

Fig. 1 — Lag windows D_2 and D_3. Spectral windows Q_2 and Q_3.

side lobes (excepting the split side lobes nearest the main lobes), and that the (normal) side lobe width is $1/(2T_m)$.

The general nature of the spectral windows in these two pairs is the same: a main lobe, side lobes at most 1 per cent or 2 per cent of the height of the main lobe. There are differences, which are sometimes relevant, but these may not be obvious. The two most important of these differences are:

(a) The highest side lobe for the "hamming" (spectral) window is about $\frac{1}{3}$ the height of the highest side lobe for the "hanning" window,

(b) The heights of the side lobes for the "hanning" window fall off more rapidly than do those for the "hamming" window.

One difference favors one pair, and one the other.

These and several other window pairs are discussed in Section B.5.

6. COVARIABILITY OF ESTIMATES — BASIC RESULT

It is shown in Section B.6 that, strictly only under Gaussian circumstances, the covariance of any two power density estimates of the sort we have been considering is given to a good degree of approximation by

$$\text{cov } \{2P_i(f_1), 2P_j(f_2)\} \approx \int_0^\infty H_i(f; f_1) \cdot H_j(f; f_2) \cdot 2\Gamma(f) \cdot df$$

where the *power-variance spectrum* $\Gamma(f)$ depends only on the true power spectrum $P(f)$ and the *effective record length* T_n', as described below. Thus, we may regard the covariance of the two power density estimates as the result of passing the power variance spectrum $\Gamma(f)$ through two networks in tandem, one with power transfer function $H_i(f; f_1)$, the other with power transfer function $H_j(f; f_2)$. In other words, we may regard the covariance (of the estimates of the power spectrum) as the power remaining from the *power-variance spectrum* $\Gamma(f)$ after passing through the *two* windows $H_i(f; f_1)$ and $H_j(f; f_2)$ associated with the estimates themselves. *If the windows do not overlap, the estimates do not covary* (at least not in terms of second moments).

In particular, of course,

$$\text{var } \{2P_i(f_1)\} = \text{cov } \{2P_i(f_1), 2P_i(f_1)\}$$

$$\approx \int_0^\infty H_i(f; f_1)^2 \cdot 2\Gamma(f) \cdot df$$

to which we can give a similar interpretation.

These results would become exact if we were to replace $C_{00}(\tau)$ by

$$\tilde{C}_{00}(\tau) = \frac{1}{T_n - T_m} \int_{-(T_n - T_m)/2}^{(T_n - T_m)/2} X\left(t - \frac{\tau}{2}\right) \cdot X\left(t + \frac{\tau}{2}\right) \cdot dt,$$

where $|\tau| \leqq T_m < T_n$. In $C_{00}(\tau)$ we averaged $X(t - (\tau/2)) \cdot X(t + (\tau/2))$ over an interval of t of length $T_n - |\tau|$, varying with τ. In $\tilde{C}_{00}(\tau)$ we would be averaging $X(t - (\tau/2)) \cdot X(t + (\tau/2))$ over an interval of t of length $T_n - T_m$ independent of τ. We could actually do this because $|t \pm (\tau/2)| \leqq T_n/2$ for $|\tau| \leqq T_m$. However, for values of $|\tau|$ less than T_m, $\tilde{C}_{00}(\tau)$ would not make use of some values of $X(t - (\tau/2)) \cdot X(t + (\tau/2))$ which are used in $C_{00}(\tau)$. Thus, $\tilde{C}_{00}(\tau)$ would be wasteful. It seems best, therefore, to use $C_{00}(\tau)$ for computation, but to approximate its variability by the variability corresponding to a $C_{00}'(\tau)$ which could not be calculated from the actual values. This "approximate" hypothetical $C_{00}'(\tau)$ involves a fixed range of integration T_n' part way between $T_n - T_m$ and T_n. The situation is illustrated in Fig. 2, where the ranges of integration are shown for the actually "feasible" $\tilde{C}_{00}(\tau)$, for the $C_{00}(\tau)$ which "wastes not", and for the $C_{00}'(\tau)$ which we use to "approximate" $C_{00}(\tau)$. The shaded areas delineate the products which are actually available.

The best choice of an intermediate value depends somewhat upon the $D_i(\tau)$ and $D_j(\tau)$ involved, and is discussed in Section B.6. In practically useful cases we may take

$$T'_n = T_n - \tfrac{1}{3}T_m .$$

The power-variance spectrum is given approximately and closely by

$$\Gamma(f) = 4\int_{-\infty}^{\infty} P(f+f')\cdot P(f-f')\cdot\left(\frac{\sin \omega' T'_n}{\omega' T'_n}\right)^2 df' \qquad (\omega' = 2\pi f').$$

If we have p pieces of *total length* T_n, and if, in computing our estimate of $C(\tau)$ for each τ, we combine all available lagged products

$$X(t - (\tau/2))\cdot X(t + (\tau/2))$$

without regard to which piece they came from, then we may use this formula for $\Gamma(f)$ with

$$T'_n = T_n - \frac{p}{3} T_m$$

as a satisfactory approximation for the *effective total length.*

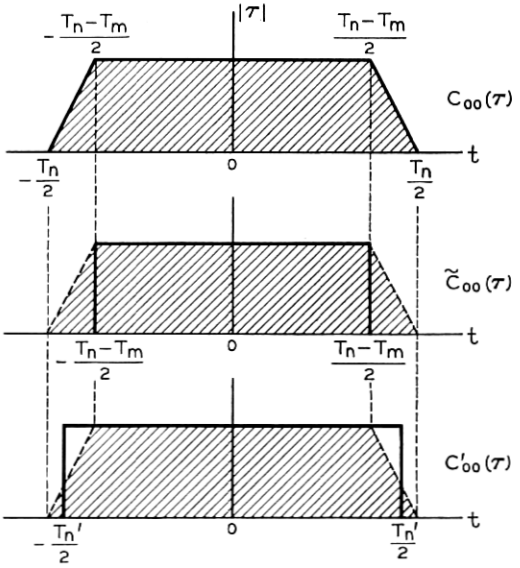

Fig. 2 — Range of integration over t as a function of τ in $\tilde{C}_{00}(\tau)$, $C_{00}(\tau)$, and $C'_{00}(\tau)$.

7. COVARIABILITY OF ESTIMATES — APPROXIMATE FORMS

In assessing the covariability of estimates of the smoothed power spectrum, the relative magnitudes of three distances along the frequency axis are important:

(a) the distance $1/T'_n$, the reciprocal of the effective length of record,

(b) the least distance over which $P(f)$ changes by an important amount for f near f_1, and

(c) the least distance over which $H_i(f; f_1)$ changes by an important amount for f near f_1 (this is of the order of $1/T_m$ and is usually much larger than $1/T'_n$).

If $P(f)$ changes slowly enough to make (b) larger than (a), we may use the approximation

$$\Gamma(f) \approx \frac{2}{T'_n} [P(f)]^2$$

whence, approximately,

$$\text{cov } \{P_i(f_1), P_j(f_2)\} \approx \frac{1}{T'_n} \int_0^\infty P_{i1}(f) \cdot P_{j2}(f) \cdot df$$

where

$$P_{i1}(f) = H_i(f; f_1)P(f)$$
$$P_{j2}(f) = H_j(f; f_2)P(f).$$

In the same terms we have

$$\text{ave } \{P_i(f_1)\} = \int_0^\infty P_{i1}(f) \cdot df$$

and

$$\text{ave } \{P_j(f_2)\} = \int_0^\infty P_{j2}(f) \cdot df.$$

The relation of covariances to averages thus established may be reasonably interpreted as meaning that any cancellations occurring in the average values also occur in the covariances and variances. To the accuracy of this approximation, then, we appear to be using the data rather efficiently.

If, on the other hand, the true spectrum, $P(f)$, consists of a single sharp peak at $f = f_0$, we may use the approximation, derived in Section B.7, namely

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

a result which is not influenced by T'_n (so long as T'_n does not become large enough for $1/T'_n$ to become comparable with the width of the peak).

8. VARIABILITY — EQUIVALENT WIDTHS

If $P(f)$ changes slowly in comparison with $1/T'_n$, then, since

$$\text{var } \{P_i(f_1)\} = \text{cov } \{P_i(f_1), P_i(f_1)\},$$

we may write down the dimensionless variability of $P_i(f_1)$ itself as

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

where

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

is naturally called the *equivalent width* of $P_{i1}(f) = H_i(f; f_1) \cdot P(f)$.

The longer the record, and the wider the equivalent width, the more stable the estimate. (Increasing the width also of course makes the estimate refer to an average power density over a wider frequency interval.)

If, on the other hand, $P(f)$ consists of a sharp peak, then, by the concluding remarks of the preceding section

$$\frac{\text{var } \{P_i(f_1)\}}{[\text{ave } \{P_i(f_1)\}]^2} = 1.$$

The equivalent widths of some simple cases are as follows:

1. If $P_{i1}(f)$ is a rectangle of width W which does not cross $f = 0$, then $W_e = W$.

2. If $P_{i1}(f)$ is a triangle of base W which does not cross $f = 0$, vertex anywhere over the base, then $W_e = 0.75\ W$.

3. If $P_{i1}(f)$ is proportional to

$$\frac{\sin \dfrac{\omega + \omega_1}{W}}{\dfrac{\omega + \omega_1}{W}} + \frac{\sin \dfrac{\omega - \omega_1}{W}}{\dfrac{\omega - \omega_1}{W}}$$

i.e. has the shape of $H_0(f; f_1)$, where $W = W_{\text{main}} = 2W_{\text{side}}$ (these being the widths of main and side lobes, respectively), and if $f_1 \geqq 1/T_m$ then $W_e = 0.5\ W = 0.5\ W_{\text{main}} = W_{\text{side}}$.

4. If $P_{i1}(f)$ has the shape of $H_2(f; f_1)$, i.e. is proportional to a hanning

(0.25, 0.5, 0.25) window, and if $f_1 \geqq 1/T_m$, then $W_e = 0.67\ W_{\text{main}} = 2.67\ W_{\text{side}}$.

5. If $P_{i1}(f)$ has the shape of $H_3(f; f_1)$, i.e. is proportional to a hamming (0.23, 0.54, 0.23) window, and if $f_1 \geqq 1/T_m$, then $W_e = 0.63\ W_{\text{main}} = 2.52\ W_{\text{side}}$.

These cases are illustrated in Fig. 3, a single sketch sufficing for the last two. Note that W_e is close to $\frac{2}{3}W_{\text{main}}$ for practical windows, if $f_1 \geqq 1/T_m$.

For our standard window pairs, hanning or hamming, the width of the normal side lobes is $1/(2T_m)$ and, consequently, $W_e \sim 1.30/T_m$, if $f_1 \geqq 1/T_m$.

These last three equivalent widths decrease somewhat as f_1 becomes small, and the values given should be halved for $f_1 = 0$.

If $P(f)$ varies linearly across $H_i(f; f_1)$, then a calculation discussed in Section B.8 shows that W_e will tend to fall in the range from $1.15/T_m$ to $1.23/T_m$. (A rather peaked case gives $0.94/T_m$.) When we allow for the fact that we are likely to be concerned with processes which are not quite Gaussian, whose variances of estimate are consequently likely to be somewhat larger than for the Gaussian case, a change corresponding to the use of a decreased equivalent width in the formula, the choice

$$W_e \approx \frac{1}{T_m}$$

which introduces a small factor of safety (not more than 1.3) seems de-

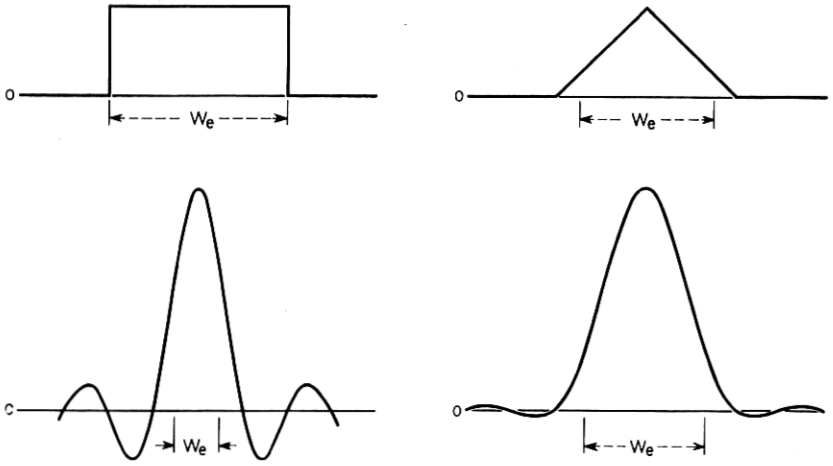

Fig. 3 — Equivalent widths of some spectral windows.

sirable for planning purposes. Consequently, we shall plan according to

$$\frac{\text{var } \{P_i(f_1)\}}{[\text{ave } \{P_i(f_1)\}]^2} \approx \frac{T_m}{T'_n}.$$

If we plan to hold the RMS deviation of each of our estimates below one-third of its average value, we must, accordingly, keep T_m/T'_n below $\frac{1}{9}$. Thus, as noted above, we shall ordinarily keep T_m to a small fraction of T_n.

In making more detailed studies of the variability of spectral estimates, further approximation will be convenient. It is important to note several reasons why we need not be too precise in making such approximations. First, as noted earlier, the variability results depend on the noise being exactly Gaussian. Real noises (and especially real signals) need not be exactly Gaussian. Thus, even exact results in Gaussian theory would be approximations in practice. Second, the chief purposes of studying variability are first to choose, once for all, effective methods of analysis, and then, in each situation, to determine about how much data will be required for the desired or given accuracy. Again, approximate results will be adequate. Third, it would not be safe to use the advance estimates of variability as firm, guaranteed, measures of the stability of the actual computed results in a practical situation, since other sources of variability may well contribute to the deviation of a particular spectral density estimate from its long run value. (Non-constancy of total power level, even with distribution-over-frequency remaining constant, and failures of stationarity are two simple examples.) We must rely on observed changes from trial to trial as basically the safest measure of the lack of stability of our spectral density estimates.

Thus, the purposes of variability theory are well served if its results are approximate — deviations of actual variability from theoretical variability of ± 5 per cent, ± 10 per cent or even ± 20 per cent will be quite satisfactory. Judged by this standard, the variability theory based on (i) the Gaussian assumption and (ii) treating the distribution of the spectral density estimates as if they followed so-called "chi-square" distributions, as we shall do in the next section, will usually be very satisfactory.

9. CHI-SQUARE — EQUIVALENT DEGREES OF FREEDOM

If $y_1, y_2, \cdots, y_k$ are independently distributed according to a standard normal distribution, that is, according to a Gaussian distribution with average zero and unit variance (and, consequently, unit standard

deviation), then

$$\chi_k^2 = y_1^2 + y_2^2 + \cdots + y_k^2,$$

which is obviously positive, follows, by definition, a chi-square distribution with k degrees of freedom. The coefficient of variation of χ_k^2, the ratio of RMS deviation to average value, is $(2/k)^{1/2}$, so that, as k increases, χ_k^2 becomes *relatively* less variable. This statement also applies to any multiple of χ_k^2.

A convenient description of the stability of any positive or nearly-positive estimate is its *equivalent number of degrees of freedom*, the number of degrees of freedom of that χ_k^2 some multiple of which it resembles (in average and variance unless otherwise specified). We can find such a k from

$$k = \frac{2(\text{average})^2}{\text{variance}} = \frac{2}{(\text{coefficient of variation})^2}.$$

Interpretation is aided by Tables I and II. These tables are possible because the distribution of the ratio of any multiple of χ_k^2 to the average value (of that multiple) depends only on k. Thus, if $k = 4$, individual

TABLE I

Distribution of quantities which are distributed as fixed multiple of chi-square. Ratios of individual value to its average value exceeded with given probabilities.

Degrees of freedom	Exceeded by 90% of all values	Exceeded by 50 % of all values	Exceeded by 10% of all values
1	0.016	0.46	2.71
2	0.10	0.70	2.30
3	0.19	0.79	2.08
4	0.26	0.84	1.94
5	0.32	0.87	1.85
10	0.49	0.93	1.60
20	0.62	0.96	1.42
30	0.69	0.98	1.34
40	0.73	0.98	1.30
50	0.75	0.99	1.26
100	0.82	0.99	1.18
200	0.873	1.00	1.139
500	0.920	1.00	1.081
1000	0.943	1.00	1.057

Examples: (1) If the long run average is 10 volts²/cps, then among estimates with 10 degrees of freedom, 10 per cent would fall below 4.9 volts²/cps, and 50 per cent would fall above 9.3 volts²/cps.

(2) If a single observed estimate, with 5 degrees of freedom, is observed to be 2 volts²/cps, then we have 80 per cent confidence that the true long-run value lies between 2/1.85 = 1.08 volts²/cps and 2/0.32 = 6.25 volts²/cps.

TABLE II — BEHAVIOR OF χ_k^2 ON DECIBEL SCALE

Fraction of distribution	Spread* of interval† in db‡	k required for interval† of spread			
		10 db	5 db	2 db	1 db
40%	$6/\sqrt{k-1}$	1	3	11	42
60%	$10/\sqrt{k-1}$	2	5	28	105
80%	$16^*/\sqrt{k-1}$	4	11	63	250
90%	$20/\sqrt{k-1}$	5	18	104	410
96%	$25/\sqrt{k-1}$	8	27	161	640
98%	$29/\sqrt{k-1}$	10	34	207	820

* Accurate to nearest integer in numerator for $k \geqq 4$, except for 80 per cent, where 16 should be replaced by 15 for $k \leqq 11$. Based on Tukey and Winsor.[13] (Spread is the difference between the upper boundary expressed in db, and the lower boundary expressed in db.)

† All intervals are symmetric in the probability sense, half of the non-included probability falling above and half below the interval.

‡ Since we are dealing with measures of variance, analogous to power, 10 db = (factor of 10), and (number of db) = (10 $\log_{10}$ ratio of variances).

values of any particular multiple of χ_4^2 will, in the long run, fall below 0.26 times their average value in 10 per cent of all cases (will be 5.8 db or more below average in 10 per cent of all cases). Similarly, individual values will, in the long run, fall below 0.84 times their average value (be 0.7 db or more below average) in 50 per cent of all cases, and in 90 per cent of all cases will fall below 1.94 times their average value (be 2.9 db or less above average). Thus, in the long run, 80 per cent of all values will fall in an interval of spread (2.9) − (−5.8) = 8.7 db.

Thus, for example, to obtain 4 chances in 5 that a single observed value will lie within ±30 per cent of the true value we require (see Table I) about 40 degrees of freedom, while to obtain 4 chances in 5 that a single observed value will lie in a prescribable interval of length 5 db, we require (see Table II) at least 11 degrees of freedom.

The results of the preceding section indicate that, for an estimate of smoothed spectral density, when $P(f)$ is smooth, the number of degrees of freedom is given by

$$k = 2T_n' W_e = \frac{W_e}{\Delta f},$$

where the latter form expresses the number of degrees of freedom as the number of *elementary frequency bands*, each of width

$$\Delta f = \frac{1}{2T_n'},$$

contained in the equivalent width W_e.

For design purposes, the relation of the last section (including the small safety factor) indicates that

$$k = \frac{2}{[\text{var } \{P_i(f_1)\}]/[\text{ave } \{P_i(f_1)\}]^2} \approx \frac{2T_n'}{T_m}$$

when $P(f)$ varies slowly. (This will usually be the case if (*i*) $k > 3$ or 4, say, and (*ii*) $P_{il}(f)$ is a moderately smooth single hump. For, under these circumstances, $P_{il}(f)$ will not change rapidly in a frequency interval $1/T_n'$ and the same property can then be inferred for $P(f)$ itself.)

When, on the other hand, $P(f)$ consists of a single sharp peak, we find, using the last result of Section 7, that $k \approx 2$, so long as $1/T_n'$ is not small enough to be comparable with the width of the peak. At first glance, this result may appear a little surprising, but when we notice that a single spectral line corresponds either (*a*) to frequency $+f_0$ *and* to frequency $-f_0$, or (b) to $\cos \omega_0 t$ *and* to $\sin \omega_0 t$, or (c) to amplitude *and* to phase, it appears quite natural that a sharp line carries *two* degrees of freedom and not merely one.

We may summarize the semi-quantitative study of the stability of estimates of the smoothed power spectrum as follows:

(a) It is not necessary to judge stability with very high accuracy.

(b) It is convenient to measure stability by analogy with the number of degrees of freedom associated with a multiple of a chi-square variate.

(c) The equivalent number of degrees of freedom can be regarded as the number of elementary bands of width Δf in the equivalent width W_e of the *filtered* spectrum

$$2P_{il}(f) = H_i(f; f_1) \cdot 2P(f) \qquad (f \geqq 0)$$

if the result is not too small (say > 3 or 4) and $P_{il}(f)$ is moderately smooth.

(*d*) If the filtered spectrum approaches a single sharp peak, the equivalent number of degrees of freedom for the corresponding estimate approaches two.

In interpreting the concept of equivalent number of degrees of freedom, it may be helpful to imagine the continuous density of the *filtered* spectrum replaced by a discrete set of ordinates, one per elementary frequency band. If these ordinates are p_0, p_1, p_2, $\cdots$, the natural approximation to the number of degrees of freedom is

$$k = \frac{(p_0 + p_1 + p_2 + \cdots)^2}{p_0^2 + p_1^2 + p_2^2 + \cdots}$$

as illustrated in Fig. 4. This approximation will usually be satisfactory

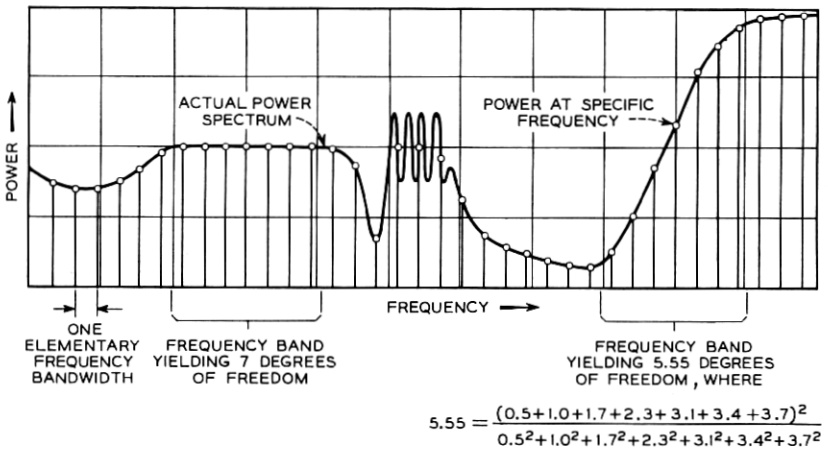

Fig. 4 — Equivalent degrees of freedom.

as long as the effect on k of moving each ordinate around within its elementary frequency band can be neglected. (In more extreme cases, an approximation based on two ordinates per pair of elementary frequency bands is more precise.)

10. DIRECT ANALOG COMPUTATION — GRADED DATA WINDOWS

We have been dealing thus far with continuous time, and the communications engineer will naturally ask, "Why introduce autocovariance functions and all that, why not measure the spectrum by filtering, rectifying, and smoothing?". The only fair answer is "By all means, do so if you can obtain, and maintain, the necessary accuracy economically!" Let us apply our results to such a measurement technique.

Let $X(t)$ be the noise or signal whose power spectrum $P(f)$ we wish to study. Let us pass it through a filter of transfer function $Y(f)$, and designate the result by $X_{\text{out}}(t)$. Its power spectrum, $P_{\text{out}}(f)$, will be given by

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

and if a section of $X_{\text{out}}(t)$ of length T_n is applied to an ideal quadratic rectifier and smoothed by a smoothing circut of infinite time constant, the result will be

$$\int_0^{T_n} [X_{\text{out}}(t)]^2 \, dt.$$

The average value of this result divided by T_n is

$$\int_0^\infty 2P_{out}(f)\, df,$$

and the number of equivalent degrees of freedom is the number of elementary frequency bands, of bandwidth $1/(2T_n)$, contained by the equivalent width of $|\, Y(f)\,|^2 \cdot P(f)$. This last function is of the form

(power transmission function)(original power spectrum)

just as before. We see that the ideal process of filtering, rectifying, and smoothing the actual input has produced the same accuracy as the ideal process of calculating, modifying, and transforming the apparent auto-covariance, provided that $|\, Y(f)\,|^2 = H_i(f; f_1)$ for a suitable choice of T_m, f, and f_1. This is what we ought to have expected, since we believe that either method extracts nearly all the information about the spectrum which the data provides.

A few practical considerations deserve mention. They center around the actual switching sitations which can arise, especially when we have only a finite sample of the original noise. In Fig. 5, the watt-second meter includes quadratic rectification and integration functions which we think of as ideal. (It may be very important to allow for the fact that the "ground" position of switch A is not quite at the same potential as the zero of the input noise, but we shall neglect this effect for the moment.)

Some four sorts of operation can arise according to the times at which switch B is operated. The watt-second meter may be connected either at the beginning of the running period T or after some interval of time (to allow initial transients to become negligible), and may be disconnected either at the end of the running period T or after some interval of time (to allow the meter to reach a maximum). These four modes of operation are illustrated in Fig. 6.

In Mode I, providing the initial waiting period is long enough to allow transients to become negligible, the filter output is essentially stationary, and the earlier discussion in this section applies.

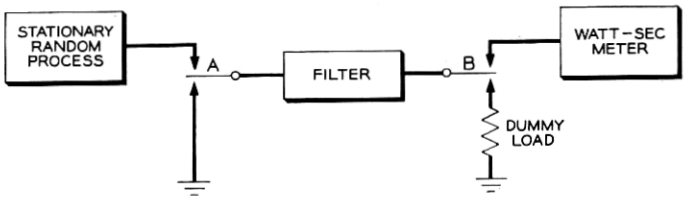

Fig. 5 — Schematic analog circuit.

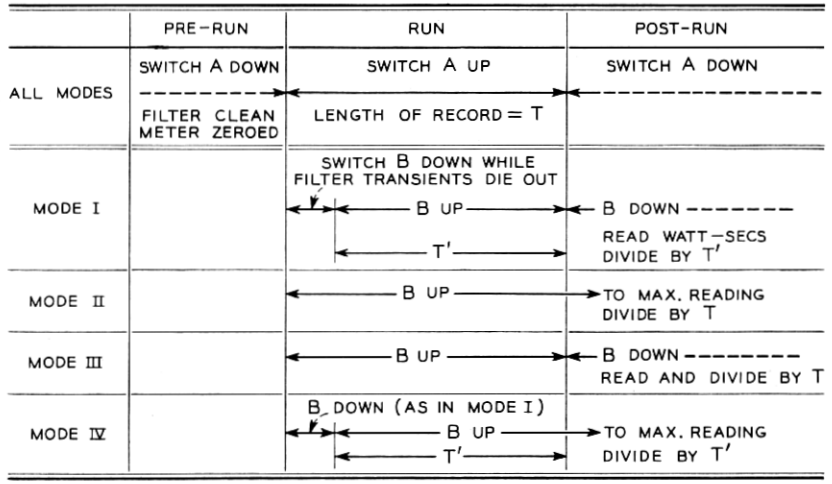

Fig. 6 — Time histories of operation for different modes.

In Mode II, all of the energy output is recorded on the meter, but the reading is divided only by the length of the input data. This mode is amenable to exact and complete analysis which is given in some detail in Section B.10. The results differ from those of Mode I in that the transform of the boxcar function of length T (running period) is convolved twice into the spectral window. (Convolution is defined and discussed in Appendix A.3.) If T is not large, the effects may be somewhat uncomfortable in that the spectral window becomes wider and more ragged.

Mode III, discussed briefly in Section B.10, differs from Mode II by an additional convolution whose effect again disappears as $T \to \infty$.

Mode IV resembles Mode I in that the noise input is passed through the filter until transient effects have become negligible, when the meter is switched on at the filter output. It differs from Mode I in that the meter is read *after* a final waiting period. This seems to offer no advantages over Mode I, and will not be discussed further.

The contrast between Mode I and Mode II is another example of what should now be becoming familiar. Mode I has no additional convolution in the spectral window. Mode II provides data economy by making it possible to integrate over the whole length of the available record. We should really like both advantages.

We can, indeed, obtain most of both advantages, but only by replacing the sharp edges of the switched data window by the smoothed outlines of a graded data window. In other words, we need to introduce

$$X_{in}(t) = B(t) \cdot X(t)$$

at the input of the filter, where $B(t)$ vanishes except for $0 < t < T$, and is smooth enough to have its Fourier transform $J(f)$ concentrated near $f = 0$. Details are discussed in Section B.10.

Difficulties arising from the fact that the zero of the $X(t)$ input might not be at ground are shown in Section B.10 to behave similarly to those arising from switching transients, namely, no effect in Mode I, possibly uncomfortable in Modes II and III, usually negligible when a well-chosen graded data window is used.

Another device is sometimes used to make maximum use of a finite noise record. The record is merely closed into a continuous loop, and the rectifier-smoother output averaged. It is shown in Section B.10 that here, too, we must use a graded data window $B(t)$.

11. DISTORTION, NOISE, HETERODYNE FILTERING AND PREWHITENING

Another group of very important practical considerations center around the spectrum of the "signal" as it is handled (either instantaneously, or in recorded form). We have spoken of "filtering, rectifying and smoothing" and have treated all these steps as ideal. No attention has been given to the equally vital "gathering" and "transmission and recording" steps. Tacitly, they too have been treated as ideal. Realistically, we must expect a certain amount of distortion (nonlinearity, intermodulation, etc.) and the addition of a certain amount of background noise in all three of the first steps: gathering, transmission and recording, filtration. It often proves to be most important to lessen the ill effects of such distortion and noise addition.

In a perfect system, and with a fixed spectral window, the fluctuations of an estimate are proportional to its average value. If we have a fixed uniform noise level, it will do the least additional damage if all the average values of the estimates are of about the same size, for then no low estimate can "disappear" into the noise.

Intermodulation distortion will have the greatest effect on the signal being transmitted when two strong frequencies combine to produce a modulation product whose frequency falls in a very weak region of the spectrum, for it is in such situations that the fractional distortion of the spectrum reaches its maximum. To minimize possible effects of intermodulation distortion it is again desirable to transmit, record and generally handle signals with a roughly flat spectrum.

To these noise and intermodulation considerations another sort of consideration may be added. Many frequency analyzers use a hetero-

dyne system, bringing the frequency band to be studied to a fixed filter, rather than tuning a filter across a wide frequency band. The power transfer function of the combination of heterodyne modulator and fixed filter, referred to input frequency, will depend only on Δf, the deviation of $|f|$ from $|f_0|$, where f_0 is the nominal frequency of the fixed filter, and will be denoted by $Q_i(\Delta f)$. If demands at different frequencies differ, the shape of $Q_i(\Delta f)$ may have to be a compromise. One sort of demand arises when $P(f)$ varies very rapidly. The net contribution near frequency f_1 to the average value of the spectral density estimate is measured by $H_i(f; f_1) \cdot P(f)$, where, as elsewhere, $H_i(f; f_1) = Q_i(f + f_1) + Q_i(f - f_1)$. If our estimate is to be useful, only f's near f_1 should have a substantial net contribution. If $P(f)$ rises steeply as f leaves f_1, we may have to require a very rapid fall-off in $H_i(f; f_1)$, here practically equal to $Q_i(f - f_1)$, in order to attain this as f leaves f_1. We may thus be forced to compromise properties of $Q_1(\Delta f)$ useful near other frequencies. The simplest way to avoid such problems is to arrange for the $P(f)$ of the "signal" analyzed to be fairly constant, or at most slowly varying.

Thus, for a variety of reasons, we can often gain by introducing "compensation" or "preemphasis" to make more nearly constant the spectrum of the "signal" actually transmitted or recorded, and analyzed. Since the ideal would be to bring the spectrum close to that of white noise, it is natural to refer to this process as *prewhitening*. Such flattening of the spectrum need not be precise, or even closely approximate. We need only to make the rate of change of $P(f)$ with frequency relatively small.

Because of advantages related to the noise and intermodulation distortion introduced in various steps of the sequence, it will be best, other considerations aside, to carry out such prewhitening at as early a point in the measurement-analysis sequence as possible. Sometimes this can even be done in the pick-up or sensing element.

This whole philosophy of prewhitening, which appears quite natural to the communication engineer familiar with preemphasis and other techniques for increased information transfer within a given frequency interval, comes as a great change to the instrumentation engineer, whose clients ordinarily require "faithful" reproduction of an input at the output, by which they mean phase shifts nearly linear with frequency, and a nearly constant amplitude response up to some high frequency. It will be rare indeed, in practical spectrum analysis, that the ideal response for the initial transducer and amplifier will be flat. Instead it should have a characteristic contributing to prewhitening. This characteristic will, of course, have to be measured separately and the corresponding

adjustments to the estimates of the spectral density will have to be made so that these estimates, instead of applying to the "signal" actually analyzed, apply to the original "input signal", but such labor will often be many times repaid.

One further consideration about frequency responses in measurement now enters naturally. In almost every power spectrum problem there is an upper frequency beyond which there is no appreciable interest. In most components used in measurement, transmission, recording, etc., the noise level, and often the level of intermodulation distortion, is roughly a fixed fraction of the peak useful level. If substantial power is present at frequencies so high as to be uninteresting, then the need to keep *total* power below the peak useful level forces us to handle the interesting frequencies at a power level below that which could otherwise be used. The ratio of noise and intermodulation distortion to interesting signal is thus raised — the quality of the analysis and its results degraded. The appropriate remedy is to filter out the uninteresting high frequencies at as early a stage as possible. This is a further reason why a carefully tailored frequency response is an important part of a power spectrum measuring process.

Together with the need for *adequately wide filters* (we can of course use narrower filters when we are prepared to average over homogeneous records of sufficiently long total duration) to provide enough equivalent degrees of freedom, and hence enough stability for the estimates, this tailoring of frequency response is often the crucial part of a power spectrum measuring program. Indeed, there may sometimes be no reasonable way to measure power spectra with an ill-tailored frequency response, even if this response be "flat".

Equally Spaced Records

We come now to treat a modified situation of great practical importance, where the observations are used for analysis only at equally spaced intervals of time — not as a continuous time record. Two new and important features enter: there is aliasing of frequencies, and practical analysis will involve digital rather than analog computation. In general, however, the situation is surprisingly similar to the case of a continuous record, with limitations on data-gathering effort still forcing us to compromise resolution and stability. Advantages of convenient calculation and noise reduction still lead us to prewhitening. Filtering of equi-spaced data must involve transversal filters (see Glossary of Terms for definition) whose transmission properties (in frequency) exhibit a periodic symmetry. This exerts additional pressure toward prewhitening.

Questions regarding computational techniques arise anew because of the nature of digital computation. These include means for reducing the effects of a displaced (perhaps drifting) zero, smoothing by groups to economize arithmetical operations on the whole, and preliminary rough estimation as an aid to planning.

12. ALIASING

We now suppose that $X(t)$ is available, or is to be used, only for uniformly spaced values of t, which we may as well suppose to be

$$t = 0, \Delta t, 2\Delta t, 3\Delta t, \cdots, n\Delta t,$$

so that $C(\tau)$ can only be estimated for

$$|\tau| = 0, \Delta t, 2\Delta t, \cdots, n\Delta t.$$

Now, the equations

$$C(\tau) = \int_0^\infty 2P_A(f) \cdot \cos 2\pi f\tau \cdot df,$$

$$|\tau| = q\Delta t, \qquad q = 0, 1, \cdots, n,$$

if soluble at all, can always be satisfied by a $P_A(f)$ which vanishes for $f > f_N = 1/(2\Delta t)$, although the power spectrum $P(f)$ of the original process (for which the $C(\tau)$ was defined) might actually cover a much wider frequency range. (We shall reserve the notation $P_A(f)$ for such a function, vanishing for $|f| > f_N$.) While frequencies between $f = 0$ and $f = f_N$ are clearly distinct from one another, we face a problem of aliasing, since frequencies above f_N usually contribute some power. Each frequency, no matter how high, is indistinguishable from one in the band from 0 to f_N.

The essential, unavoidable nature of this problem is made clear by Fig. 7 which illustrates how equally spaced time samples from any

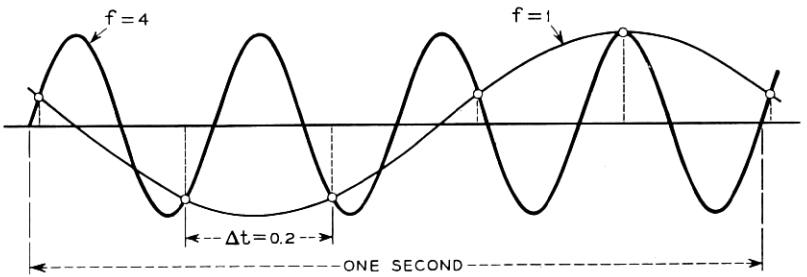

Fig. 7 — Sampling of sinusoidal waves.

cosine wave *could* have come from each of many other cosine waves. (The familiar stroboscope uses a particular expression of this fact in apparently "slowing down" rapidly rotating or oscillating machinery.)

The logical position about $P_A(f)$ depends very much on whether $X(t)$ is thought of as having any real existence for $|t| \neq q\Delta t$.

If $X(t)$ really exists for continuous t, although we have (*i*) failed to observe or record it, or (*ii*) failed to "read" the record, or (*iii*) decided to neglect the available values, then there is a well-defined $P(f)$ corresponding to the process from which each $X(t)$ is a sample, and we must be very careful about the relation between $P(f)$, which is our true concern, and $P_A(f)$, which is clearly all we can strive to estimate directly from the data. It can be shown (see Section B.12) that, in the form appropriate for a one-sided spectrum, if we set

$$2P_a(f) = 2P(f) + 2P(2f_N - f) + 2P(2f_N + f) + 2P(4f_N - f) + 2P(4f_N + f) + \cdots$$

then we may take

$$P_A(f) = \begin{cases} P_a(f), & 0 \leqq |f| \leqq f_N, \\ 0, & \text{otherwise} \end{cases}$$

where $f_N = 1/(2\Delta t)$ is the *folding* (or Nyquist) frequency. We naturally call the frequencies f, $2f_N - f$, $2f_N + f$, $4f_N - f$, $4f_N + f$, and so on, *aliases* of one another, f being the *principal alias*. The *aliased spectrum* $P_a(f)$ is the result of *aliasing* $P(f)$. The *principal part of the aliased spectrum* $P_A(f)$ is the part of $P_a(f)$ which corresponds to principal aliases, positive and negative.

(If $X(t)$ has no natural existence for t's which are not integral multiples of Δt, then $P(f)$ is not uniquely defined, and we are at liberty to choose any normalization we desire. In particular, we may decide to limit $P(f)$ to the interval $|f| \leqq 1/(2\Delta t)$, in which case we will be enforcing $P(f) = P_A(f)$ without any trace of aliasing. We mention this case for logical completeness, but remark that it seems to occur infrequently in practice, whatever the field.)

If the Gaussian noise we are considering has a power spectrum $P(f)$ which extends outside $|f| \leqq 1/(2\Delta t)$, then the Gaussian noise with spectrum $P_A(f)$ is not the same for continuous time. However, if we consider these two noises only for equi-spaced times

$$t = 0, \Delta t, 2\Delta t, \cdots\cdots$$

they are identical. For all first moments vanish and all second moments

coincide, which implies coincidence of the joint distributions of any finite set from $\cdots, X_{-q}, \cdots, X_{-1}, X_0, X_1, \cdots, X_q, \cdots$, and this is our definition of the coincidence of two noises. (If a result concerning such equally spaced values can be established for a Gaussian noise restricted to have $P(f)$ vanish outside $|f| \leqq 1/(2\Delta t)$, it must trivially hold, under the same restriction, when all occurrences of $P(f)$ are changed to $P_A(f)$. It is a consequence of the identification just established that the result, when expressed in terms of $P_A(f)$, must also hold for any Gaussian noise whatever.)

The frequency interval from 0 to f_N contains a certain number of elementary frequency bands in the sense of our treatment of variability. The total length of record is $T_n = n\Delta t$, and if we write $T'_n = n'\Delta t$ for the effective length, then, since

$$\frac{f_N}{\text{elementary frequency bandwidth}} = \frac{\frac{1}{2\Delta t}}{\frac{1}{2T'_n}} = n'$$

there are n' elementary frequency bands between 0 and f_N. As a statistician would have anticipated, we gain one elementary frequency band — one degree of freedom — for each added observation.

It is perhaps natural to base a hope on this result — a hope that taking data more frequently over the same time interval would gain us many degrees of freedom and reduce our difficulties with variability. However, this is not the case (as the expression for the width of an elementary frequency band $1/(2T'_n)$ should have warned us). Taking observations twice as frequently yields twice as many elementary frequency bands, but also doubles the folding frequency f_N and, thus, doubles the frequency interval occupied by principal aliases. The density of elementary frequency bands is not increased one iota. (Clearly, iota was the Greek word for bit!).

It is usual for aliasing to be present and to be of actual or potential importance. This is an inescapable consequence of data taken or read at uniform intervals. (It is not infrequently suggested that there should be a workable scheme of taking discrete data in some definite, but not uniformly spaced pattern, and estimating the power spectrum without aliasing. No such scheme seems so far to have been developed).

13. TRANSFORMATION AND WINDOWS.

Given uniformly spaced values of $X(t)$ — values which we shall now designate $X_0, X_1, \cdots, X_n$, as well as $X(0), X(\Delta t), \cdots, X(n\Delta t)$ — we

expect to calculate "sample autocovariances", modify them, and then Fourier transform the results. There is no possibility of calculating autocovariances for lags other than 0, Δt, $\cdots$, $n\Delta t$, and so we may as well write C_0, C_1, $\cdots$, C_m in place of $C_i(0)$, $C_i(\Delta t)$, $\cdots$, $C_i(m\Delta t)$. If we Fourier transform these $m + 1$ numbers, as obtained or modified, we might obtain a smoothed spectral estimate for any frequency between 0 and $f_N = 1/(2\Delta t)$ that we may wish. It is not surprising, however, that we lose no information (and little explication) if we calculate only $m + 1$ such estimates (one for each C_r). Nor is it surprising that we regularly take these estimates equally spaced over $0 \leqq f \leqq f_N$, and hence at intervals of $f_N/m = 1/(2m\Delta t)$. As a consequence we have to deal with *finite* Fourier (cosine) *series* transformation (classical harmonic analysis) rather than with *infinite* Fourier *integral* transformation, but the correspondence between multiplication and convolution persists.

The question of modification also requires discussion. In the continuous case we Fourier transformed

$$C_i(\tau) = D_i(\tau) \cdot C_{00}(\tau) = D_i(\tau) \cdot C_0(\tau)$$

where $C_0(\tau)$ coincided with $C_{00}(\tau)$ wherever the latter was defined, and is zero otherwise (cp. Section B.5). The result was, consequently (e.g. see Appendix A.3), the convolution of the Fourier transforms of $D_i(\tau)$ and $C_0(\tau)$. So long as time was continuous and computation was presumably by analog devices, there was a real advantage to modification before transformation. Now that time is discrete and computation presumably digital, the advantage is transferred to first transforming and then convolving. Indeed, because the $D_i(\tau)$, for $i > 1$, are finite sums of cosines, so that their transforms are simply sums of spikes (Dirac delta-functions) at the appropriate spacing, convolution means only smoothing with weights

$$0.25,\ 0.5,\ 0.25 \qquad (i = 2, \text{ hanning})$$

$$0.23,\ 0.54,\ 0.23 \qquad (i = 3, \text{ hamming})$$

and is very simply carried out.

In discussing this program, we gain some generality by using $m + 1$ lags separated by $\Delta\tau = h\Delta t$ for an integer $h > 0$, while our results are no more complicated than if we were to confine ourselves to $h \equiv 1$, which is the practical case. Thus, we first compute the *mean lagged products*

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

for $r = 0, 1, 2, \cdots, m$, where $mh < n$. Note that C_r is heuristically as close as we can come to the apparent autocovariance $C_{00}(\pm r\Delta\tau)$ with the available (equi-spaced) data. Note further that, so far as functions of the C_r are concerned, our effective folding frequency is

$$f_N^* = \frac{1}{2\Delta\tau} = \frac{1}{h} f_N .$$

We will usually need to adjust the C_r somewhat to improve very-low-frequency performance, as discussed in Section 19, but this need not concern us for the moment.

Applying a *discrete finite* cosine *series* transform to the sequence C_0, C_1, $\cdots$, C_m, we find

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

(We may regard this as arising from replacing $C_0(\tau)$ in the expression for $P_0(f)$ as its Fourier integral transform by a finite sequence of spikes (Dirac delta functions) of intensities (areas) proportional to the corresponding values of $C_0(\tau)$.) If we put

$$P_{0A}\left(\frac{r}{2m \cdot \Delta\tau}\right) = V_r$$

then it is shown in Section B.13 that

$$\text{ave } \{P_{0A}(f)\} = \int_{-\infty}^{\infty} Q_0(f - f'; \Delta\tau) \cdot P(f') \cdot df'$$

where

$$Q_0(f; \Delta\tau) = \Delta\tau \cdot \cot \frac{\omega\Delta\tau}{2} \cdot \sin m\omega\Delta\tau.$$

In terms of $Q_0(f)$, which is treated in Section B.5, we have

$$Q_0(f; \Delta\tau) = \sum_{q=-\infty}^{\infty} Q_0\left(f - \frac{q}{\Delta\tau}\right) = Q_{0A}(f).$$

Just as the average value of $P_0(f)$ in the continuous case is the corresponding value of $Q_0(f) * P(f)$, so here the average value of $P_{0A}(f)$ is the corresponding value of $Q_0(f; \Delta t) * P(f)$. Thus, we may consider $P_{0A}(f)$ as estimating the result of "smoothing" $P(f)$ with a window $Q_0(f; \Delta\tau)$ which has repeated major (and concomitant minor) lobes at intervals of $2f_N^* = (\Delta\tau)^{-1}$. This is not the most convenient way to consider matters, and in Section B.13 it is shown that there are two equivalent forms for

ave $\{P_{0A}(f)\}$ and, correspondingly, two other, equally appropriate, ways to consider the situation.

These arise from the three-fold identity

$$Q_{0A}(f) * P(f) \equiv Q_0(f) * P_a(f) \equiv Q_{0A}(f) * P_A(f),$$

any member of which represents the average value of $P_{0A}(f)$. Thus, we can also consider $P_{0A}(f)$: (i) as estimating the result of smoothing the infinite, periodic aliased spectrum $P_a(f)$ with the same window as for the continuous case, or (ii) as estimating the result of smoothing the principal part of the aliased spectrum $P_A(f)$ with the aliased window $Q_{0A}(f)$. The latter choice is usually the most helpful of the three possibilities, and is the one we shall adopt.

All this has been discussed for the immediate results of transforming unmodified C_r's. This is only the case $i = 0$ of the identity

$$Q_{iA}(f) * P(f) \equiv Q_i(f) * P_a(f) \equiv Q_{iA}(f) * P_A(f)$$

which holds in general. We should thus usually be concerned with $Q_{iA}(f)$ and with $P_A(f)$.

The case $i = 2$ (hanning) corresponds to the following smoothing after transformation:

$$\begin{aligned} U_0 &= 0.5\ V_0 + 0.5\ V_1\,, \\ U_r &= 0.25\ V_{r-1} + 0.5\ V_r + 0.25\ V_{r+1}\,, \qquad 1 \leqq r \leqq m-1, \\ U_m &= 0.5\ V_{m-1} + 0.5\ V_m\,, \end{aligned}$$

for which $Q_{2A}(f)$ has the form shown in Fig. 8. The curve is for $m = 12$, and the circles are for $m = \infty$, which corresponds exactly to the continuous case. Clearly, for usual values of m, the modification in the lobes due to aliasing is almost surely unimportant.

The frequency separation between adjacent estimates is

$$\frac{1}{2T_m} = \frac{1}{2m\Delta\tau},$$

but the equivalent width of the windows (for $1 \leqq r \leqq m-1$) is about

$$\frac{1.30}{T_m} = \frac{1.30}{m\Delta\tau},$$

just as for the continuous case (see Section 8). For most purposes we may again take the bandwidth corresponding to each estimate as $1/T_m$,

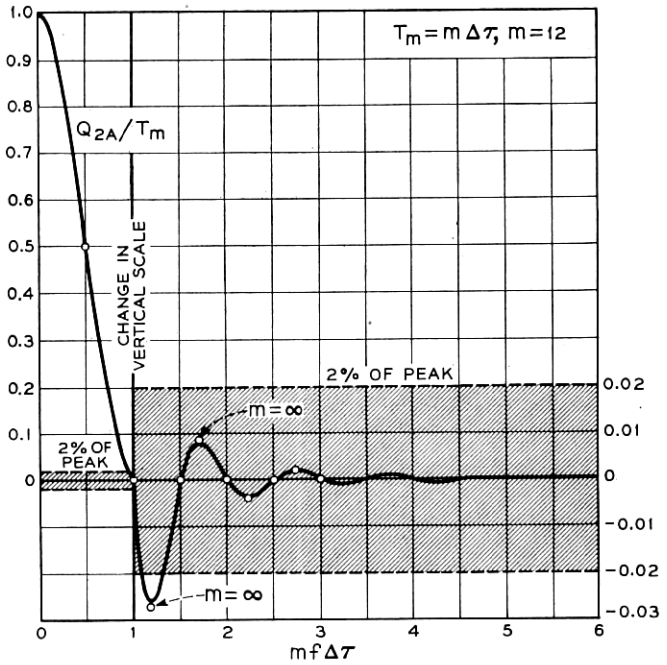

Fig. 8 — Aliased spectral window Q_{2A} for $m = 12$.

so that m satisfies

$$\frac{1}{m} = (\text{bandwidth of estimates}) \cdot (\Delta\tau).$$

If we had neither modified before Fourier transformation, nor smoothed after transformation, we should have faced the uncomfortable minor lobes of $Q_{0A}(f)$ shown in Fig. 9 for $m = 12$ (with circles for $m = \infty$). Generally speaking, all we learned about desirable lag windows for the continuous case carries over with minor modifications, at most. The only serious effect of going to uniformly spaced values is the aliasing (and this may be very serious indeed).

It is well worth noting that the possible spectral windows $Q_{iA}(f)$ are now restricted to be finite Fourier series in $\cos \omega\Delta\tau$, $\cos 2\omega\Delta\tau$, $\cdots$, $\cos m\omega\Delta\tau$, or equivalently, to be polynomials in $\cos \omega\Delta\tau$ of degree m at most.

14. VARIABILITY AND COVARIABILITY

We now extend all our other notation: $H_i(f; f_1)$, $P_i(f_1)$, etc. to corresponding $H_{iA}(f; f_1)$, $P_{iA}(f_1)$, etc. for the uniformly spaced case as

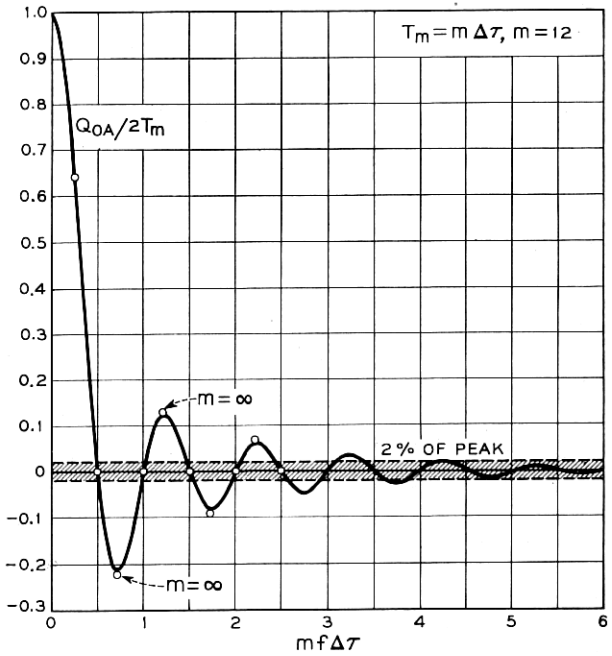

Fig. 9 — Aliased spectral window Q_{0A} for $m = 12$.

specified in Sections B.13 and B.14. It is shown in the latter section that we now have

$$\text{cov } \{2P_{iA}(f_1), 2P_{jA}(f_2)\} = \int_0^\infty H_{iA}(f; f_1) \cdot H_{jA}(f; f_2) \cdot 2\Gamma_{\Delta t}(f) \cdot df$$

where

$$\Gamma_{\Delta t}(f) \approx 4 \int_{-\infty}^{\infty} P_A(f' + f) \cdot P_A(f' - f) \cdot \left(\frac{\sin \omega' T_n'}{\omega' T_n'}\right)^2 \cdot \left(\frac{\sin \omega' \Delta t}{\omega' \Delta t}\right)^{-2} df',$$

($\omega' = 2\pi f'$), with a very slightly different determination of T_n' than before. The only essential change has been the introduction of a new factor, corresponding to aliasing,

$$\left(\frac{\sin \omega' \Delta t}{\omega' \Delta t}\right)^{-2}$$

into the integrand of the power-variance spectrum $\Gamma_{\Delta t}(f)$. For usable values of n, this factor will vary much more slowly than

$$\left(\frac{\sin \omega' T_n'}{\omega' T_n'}\right)^2$$

and can usually be treated as sensibly equal to unity. All the approximate analysis of covariability and variability given for the continuous case now goes through without essential change.

15. PREWHITENING

If the equally spaced data is sampled from a continuously transmitted "signal" or "read" from a continuous recording, then all the points made in Section 11 in favor of early prewhitening are still applicable. If the equally spaced data arises more directly, as by photographing a physical situation, we may not be able to apply prewhitening early. In either case it may still be desirable to prewhiten after the data is obtained at equal intervals, either as a supplement to, or as a partial replacement for, early prewhitening.

The average value of a power density estimate $P_{iA}(f_1)$ is

$$\text{ave } \{P_{iA}(f_1)\} = \int_0^\infty P_{iA1}(f) \cdot df,$$

where

$$P_{iA1}(f) = H_{iA}(f; f_1) \cdot P_A(f).$$

We want this quantity to tell us about the values of $P(f)$ for f near f_1. To do this we must: (i) reduce variability, (ii) ensure that $P_A(f)$ resembles $P(f)$ sufficiently, and (iii) concentrate $P_{iA1}(f)$ near $f = f_1$. We must be concerned with: (i) adequately broad windows, (ii) sufficiently weak aliasing, and (iii) enough sharpness in the effective filter. This sharpness can be obtained in a combination of ways.

Note that we asked for $P_{iA1}(f)$, which measures the net contribution to the average value, to be localized. We did not merely ask that $H_{iA}(f; f_1)$ should be localized. For, if

$$P_A(f_2) >>> P_A(f_1),$$

although

$$H_{iA}(f_2 ; f_1) << H_{iA}(f_1 ; f_1),$$

it is still possible for

$$P_{iA1}(f_2) = H_{iA}(f_2 ; f_1) \cdot P_A(f_2)$$

to outweigh

$$P_{iA1}(f_1) = H_{iA}(f_1 ; f_1) \cdot P_A(f_1),$$

so that our estimate tells us more about $P(f)$ near $f = f_2$ than it does about $P(f)$ near $f = f_1$. To avoid such unfortunate situations either we

must choose our window pair in a very particular manner (so as to make $H_{iA}(f_2\,;\,f_1)$ exceptionally small) or we must avoid $P_A(f_2) >>> P_A(f_1)$. Both courses are possible and sometimes necessary. Usually, the second course is simpler.

Following the second course is simple in principle. Given actual values X_q, we apply a selected linear procedure to obtain new values $\tilde{X}_q$ and analyze these. The aliased spectrum $\tilde{P}_A(f)$ of the $\tilde{X}_q$ differs from the aliased spectrum $P_A(f)$ of the X_q by a known multiplicative function of frequency. (See Section B.15 for details.) Thus, (i) we may convert estimates of $\tilde{P}_A(f)$ into estimates of $P_A(f)$, and (ii) we may choose the linear procedure to make the aliased spectrum $\tilde{P}_A(f)$ of the $\tilde{X}_q$ reasonably flat.

The simplest linear procedures are probably the formation of moving linear combinations and the construction of autoregressive series. A simple example of a moving linear combination is

$$\tilde{X}_q = X_q - \alpha X_{q-1} - \beta X_{q-2} - \gamma X_{q-3}$$

for which the relation between the spectra is

$$\frac{\tilde{P}_A(f)}{P_A(f)} = \frac{\tilde{P}(f)}{P(f)} = |\,1 - \alpha e^{-i\omega\Delta t} - \beta e^{-i2\omega\Delta t} - \gamma e^{-i3\omega\Delta t}\,|^2$$

$$= \text{a cubic in } \cos\,\omega\Delta t.$$

A suitable moving linear combination will generate any desired non-negative polynomial in $\cos\,\omega\Delta t$.

A simple example of an autoregressive combination is

$$\tilde{X}_q = X_q + \lambda\tilde{X}_{q-1} + \mu\tilde{X}_{q-2} + \nu\tilde{X}_{q-3}$$

for which the relation (reciprocal to that just considered) between the spectra is

$$\frac{\tilde{P}_A(f)}{P_A(f)} = \frac{\tilde{P}(f)}{P(f)} = \{|\,1 - \lambda e^{-i\omega\Delta t} - \mu e^{-i2\omega\Delta t} - \nu e^{-i3\omega\Delta t}\,|^2\}^{-1}$$

$$= (\text{a cubic in } \cos\,\omega\Delta t)^{-1}.$$

A suitable autoregressive combination will, when indefinitely continued, generate the reciprocal of any desired non-negative polynomial in $\cos\,\omega\Delta t$.

By combining a suitable moving linear combination with suitable autoregression, as for instance in

$$\tilde{X}_q = X_q - \alpha X_{q-1} + \lambda\tilde{X}_{q-1}\,,$$

which may also be written

$$\tilde{X}_q - \lambda \tilde{X}_{q-1} = X_q - \alpha X_{q-1} ,$$

for which

$$\frac{\tilde{P}_A(f)}{P_A(f)} = \frac{\tilde{P}(f)}{P(f)} = \left| \frac{1 - \alpha e^{-i\omega\Delta t}}{1 - \lambda e^{-i\omega\Delta t}} \right|^2$$

$$= \frac{1 + \alpha^2 - 2\alpha \cos \omega\Delta t}{1 + \lambda^2 - 2\lambda \cos \omega\Delta t}$$

$$= \text{a rational function of } \cos \omega\Delta t,$$

we can modify $P_A(f)$ by multiplication by an arbitrary non-negative rational function of $\cos \omega\Delta t$.

Freedom to multiply by any (simple) non-negative rational function of $\cos \omega\Delta t$ is very substantial freedom. If we have a rough idea (see Section 18) of the behavior of $P_A(f)$, and if this behaviour is moderately smooth, though perhaps quite steep in places, we can usually do a very good job of flattening the spectrum by prewhitening after obtaining discrete (digital) values. Unless still bothered with steep slopes, we will usually then find that hanning, with its (0.25, 0.50, 0.25) weights and lower outer lobes is slightly preferable to hamming, with its (0.23, 0.54, 0.23) weights and reduced first minor lobes.

The main purpose of prewhitening *after* data has been obtained in digital form at equally spaced intervals is to avoid difficulty with the minor lobes of our spectral windows. We may regard the whole process of prewhitening, analysis with standard spectral windows, and, finally, compensation of estimate, as a means of constructing a set of specially shaped spectral windows, one for each center frequency, specially adapted to the data we are processing. This point of view is illustrated in Fig. 10. The uppermost curve shows the power transfer function of a hypothetical prewhitening filter, one which enhances mid-frequencies in comparison with those lower and higher. The next line shows two standard spectral windows, with symmetrical side lobes. The third line shows the effective spectral windows when prewhitening is followed by standard analysis, as given by the product of prewhitening power transfer function and spectral window. In either case, the side lobe toward mid-frequencies is higher than the corresponding side lobe on the opposite side, which is lower than for the standard. The lowest curve shows alternative spectra for time series which might reasonably be processed by the combination of prewhitening and standard analysis shown (since the prewhitened spectra would change only slowly). In every case, the side lobes of the

special spectral windows are automatically so related to these spectra, as to balance and reduce the amount of leakage through them, as given by the product of special spectral window side lobe and original spectral density.

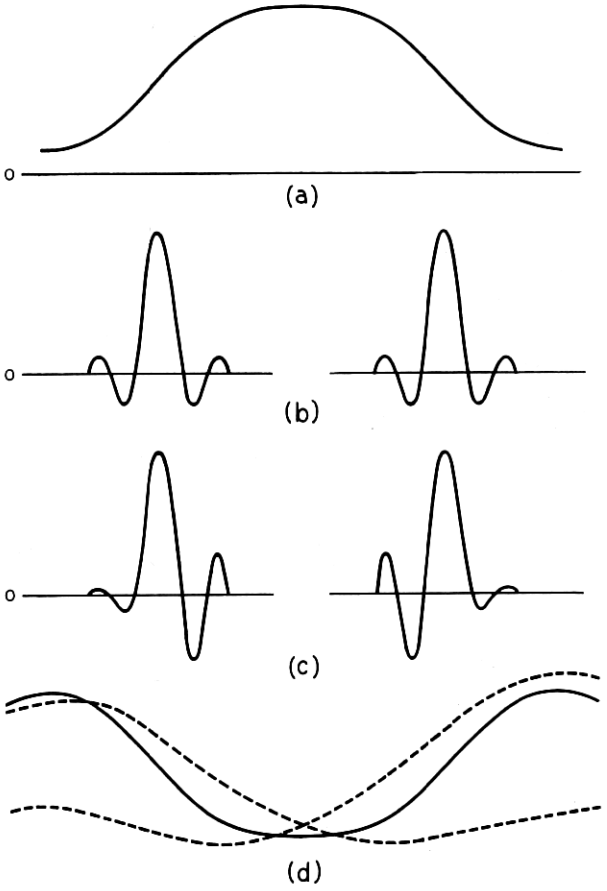

Fig. 10 — Illustration of prewhitening; (a) prewhitening power transfer function, (b) standard spectral windows, (c) effective spectral windows, and (d) typical input spectra to which (a) might be applied.

Easing of requirements for accuracy (number of significant figures, etc.) during computation are ordinarily quite secondary, though pleasant, advantages of prewhitening during digital calculation.

16. REJECTION FILTERING AND SEPARATION

If the difficulties in handling $P(f)$ are due, wholly or in part, to one or more quite narrow and very high peaks ("lines" or "narrow bands")

then we cannot expect either to afford, or to be able to estimate, the great number of accurately chosen constants which would be required to obtain a rational function whose reciprocal has a shape very close to the given narrow peak. We must adopt a slightly different approach, and plan to make at least two analyses of the data — one to estimate the behavior at the peak, and another to estimate the behavior away from the peak.

In order to separate the bulk of the information in the data from the variation associated with the sharp peak which may be troubling us, we may apply to the data a moving linear combination (possibly combined with autoregression) whose power transfer function (the factor by which the spectrum is altered) has one or more zeroes near the peak. The resulting sequence will be largely free of contribution from the peak and hence will be suitable for further prewhitening (if required) and analysis. (This operation can often, of course, be combined with further prewhitening so far as actual calculation goes. It will of course be necessary to compensate for the effects of this transformation at frequencies away from the peak, when preparing the final spectrum estimates for interpretation.)

There remains the estimation of the power in the peak, and possibly some inquiry into its width. A number of approaches are possible:

(1) We may analyze the original data as well as the data with the peak rejected, obtaining an estimate at the peak and possibly confirmatory estimates far from the peak.

(2) We may subtract a suitable multiple of the modified data from the original data so as to retain the peak and partially reduce other frequencies; and then analyze the difference.

(3) We may apply a band-pass filter to isolate frequencies at and near the peak, and then analyze the result.

Any of these techniques may be applicable in suitable circumstances.

Other related procedures are sometimes more natural than the use of moving linear combinations. Rejection of zero frequency, for example, is more naturally, and computationally more easily, accomplished by subtraction of the mean of all the data from each X_q than by the subtraction of a *moving* linear combination from each.

Rejection filtration has been applied in oceanography by Groves,[14] Seiwell,[15] Seiwell and Wadsworth,[16] to the elimination of various well-defined tides from records. It almost always has to be used to eliminate possible peaks at zero frequency (see Section 19 below).

In electronic measurements we may also anticipate its possible use in measurements: (i) close to a substantial harmonic of 60 cycles per second (such as 120 cps or 1380 cps), or (ii) near some strong "carrier".

17. SMOOTHING BY GROUPS

The cost of digital power spectrum analysis, once initial investments in programming, etc. have been made, and assuming records to have already been made and "read", is likely to be associated with the number of multiplications involved in computing the mean lagged products (in original or modified form). If there are n observations, and m lags are used, then there will be roughly nm multiplications.

Ways of reducing this number substantially are naturally of interest. Most of these must depend for their efficacy on our interest in something less than the whole spectrum. We have already discussed (in passing) a situation which would naturally arise only when we are interested only in the lower part of the aliased spectrum. This is the use of lags which are multiples of $\Delta\tau = h\Delta t$ with $h > 1$. The use of lags up to $m\Delta\tau = hm\Delta t$ allows us to explore the spectrum down to frequencies almost of the order $1/hm\Delta t$, which, had we used all multiples of Δt up to hm, would have required $hm + 1$ values of C_τ (or of its modifications) instead of $m + 1$. The price of doing this is the aliasing of the spectrum with folding frequency $1/(2\Delta\tau) = (1/h)\,(1/(2\Delta t))$, which is h times as much aliasing as if all multiples of Δt up to hm had been used, yielding a folding frequency of $1/(2\Delta t)$.

If such intensive aliasing is bearable, this procedure with $\Delta\tau > \Delta t$ is simple, even though it is not necessarily economical. Indeed, if so much aliasing were permissible, we need only have "read" every hth data value. In many situations, however, especially where Δt has been taken as large as aliasing will permit, such further aliasing is unbearable. If we are to look at the low frequency part of the aliased spectrum $P_A(f)$with computational economy, another course will have to be found.

Our use of linear schemes in prewhitening shows us a possible course. Let us begin by applying a linear scheme to the given values X_q, which attenuates all high frequencies. Then we can face further aliasing, and proceed apace.

If simplicity is controlling, then we take

$$\tilde{X}_q = X_q + X_{q-1} + \cdots + X_{q-k+1} \qquad (k \text{ terms})$$

for which the relation between the spectra (the power transfer function of the smoothing) is

$$\frac{\tilde{P}(f)}{P(f)} = |\, 1 + e^{-i\omega\Delta t} + \cdots + e^{-i(k-1)\omega\Delta t}\,|^2 = \left[\frac{\sin\dfrac{k\omega\Delta t}{2}}{\sin\dfrac{\omega\Delta t}{2}}\right]^2.$$

This will give us zeroes at frequencies which are multiples of $1/k\Delta t$, and we can avoid folding the first two side lobes of this function onto the main lobe and still take a folding frequency as small as $2/k\Delta t$. Such a choice will fold the second, fifth, sixth, etc. side lobes onto the first side lobe, and it will fold the third, fourth, seventh, eighth, etc. side lobes onto the main lobe. We obtain such a folding frequency by retaining only one in every $k/4$ of the $\tilde{X}_q$'s. These *decimated** $\tilde{X}_q$'s may, in particular, be obtained by summing the $\tilde{X}_q$'s in non-overlapping blocks of $k/4$, and then summing these block sums in all possible (overlapping) sets of four successive blocks. (This requires $(k + 8)/k$ additions per original value.) The estimated spectrum below $1/k\Delta t$ has to be multiplied by

$$\left[\frac{\sin\dfrac{\omega\Delta t}{2}}{\sin\dfrac{k\omega\Delta t}{2}}\right]^2,$$

and only aliases which are usually negligible will have been superposed on the principal aliases. About one kth of the original principal spectrum will be available for analysis.

The stability obtained by this process can be easily compared with that obtained by using all X_q and taking $\Delta\tau = k\Delta t/4$. In each case, the width of the elementary frequency bands is approximately $1/2T'_n$ where T'_n has slightly different, but not substantially different values. The process just described yields nearly the same stability as $\Delta\tau = k\Delta t/4$, and usually involves much less computation, besides avoiding serious aliasing. It will almost always be preferred to using $\Delta\tau = h\Delta t$ with $h > 1$.

Other schemes of smoothing by groups are discussed in Section B.17.

18. PILOT ESTIMATION

The prewhitening procedure demands a rough knowledge of the spectrum for its effective use. Sometimes this rough knowledge can be obtained from theoretical considerations, or from past experience, but in many cases it must be obtained from a preliminary (pilot) analysis of the data. Such pilot analyses should be as simple and cheap as possible. We now discuss a pilot analysis giving very rough results quite easily.

Table III exemplifies a form of calculation which is easily carried out either entirely by hand, or with a desk calculator. The symbols "δ" and "σ" refer to differences and sums of consecutive numbers in non-overlapping pairs. Taking the numbers in non-overlapping pairs is not neces-

* Although this word should refer strictly to the deletion of only every 10th item, we shall apply it to the retention of only every jth item, for whatever j may be relevant.

TABLE III — COMPUTATION OF PILOT ESTIMATES

q	X_q	δX_q	$(\delta X_q)^2$	σX_q	$\delta\sigma X_q$	$(\delta\sigma X_q)^2$	$\sigma^2 X_q$	$\delta\sigma^2 X_q$	$(\delta\sigma^2 X_q)^2$
1	3								
2	4	1	1	7					
3	−1								
4	−2	−1	1	−3	−10	100	4		
5	2								
6	7	5	25	9					
7	5								
8	−1	−6	36	4	−5	25	13	9	81
9	−3								
10	2	5	25	−1					
11	5								
12	4	−1	1	9	10	100	8		
13	7								
14	3	−4	16	10					
15	4								
16	−1	−5	25	3	−7	49	13	5	25
17	−4								
18	2	6	36	−2					
19	4								
20	0	−4	16	4	6	36	2		
21	1								
22	−1	−2	4	0					
23	1								
24	2	1	1	3	3	9	3	1	1
25	4								
26	3	−1	1	7					
27	0								
28	−4	−4	16	−4	−11	121	3		
29	−1								
30	−2	−1	1	−3					
31	−2								
32*	−2	0	0	−4	−1	1	−7	−10	100
Totals			205			441			207

CONTINUATION OF TABLE III TO THE RIGHT (COMPRESSED)

q	$\sigma^3 X_q$	$\delta\sigma^3 X_q$	$(\delta\sigma^3 X_q)^2$	$\sigma^4 X_q$	$\delta\sigma^4 X_q$	$(\delta\sigma^4 X_q)^2$	$\sigma^5 X_q$	$(\sigma^5 X_q)^2$
8	17							
16	21	4	16	38				
24	5							
32	−4	−9	81	1	−37	1369	39	1521
			97			1369		1521

(* Note: $32 = 2^5$.)

sary, but saves much calculation at little cost in accuracy. (In this table sums and differences are entered in the lower of the two lines to which they correspond.)

The final sums of squares are roughly proportional to the power in successive octaves coming down from the folding frequency. They differ by only a constant factor, equal to the number 2^ℓ of values X_q used,

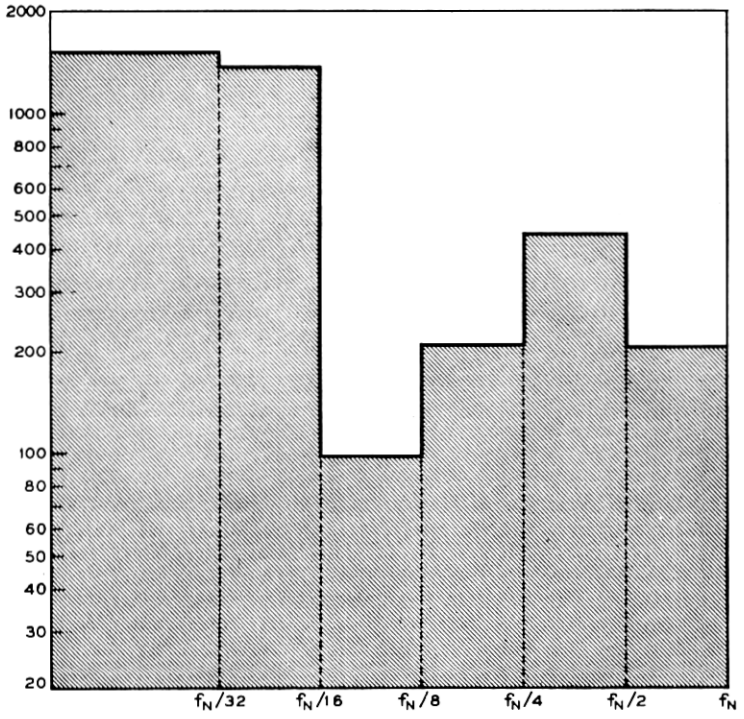

Fig. 11 — Pilot-estimated power spectrum.

from the mean squares of a nested analysis of variance. For many purposes they can be used as they come.

For the example of Table III we obtain sums of squares of 205, 441, 207, 97, 1369, and 1521. These are plotted in Fig. 11 for the successive octaves f_N to $f_N/2$, $f_N/2$ to $f_N/4$, $f_N/4$ to $f_N/8$, $f_N/8$ to $f_N/16$, $f_N/16$ to $f_N/32$, and the remaining range $f_N/32$ to 0. We see that the spectrum is roughly flat.

When medium or large stored-program digital computers are available, and the data is already available in machine-processable form (so-called diamond copy), it will often pay to use less elementary pilot calculations. Possible alternatives are discussed in Section B.18.

19. VERY LOW FREQUENCIES

The change from continuous "signals" processed in analog equipment to equally spaced "data" processed digitally has another important practical effect. Analog equipment, unless special care is taken, does not respond all the way down to zero frequency, and this automatically filters

out the very lowest frequencies. This fact allowed us, in dealing with continuous records, to treat the "signals" being processed as if they had zero means. In dealing digitally with equally spaced data, all frequencies down to zero are transmitted, *unless* we take special precautions. Consequently, we must give serious attention to the very lowest frequencies.

(We must now distinguish between power (in the sense of a line) at zero frequency and power density at zero frequency. The power spectrum of a stationary random process with zero means may have finite power density at zero frequency without having finite power there. However, finite power at zero frequency may be introduced into the data in measurement. It would then be desirable to filter out the power at (exactly) zero frequency without affecting the power density at and near zero frequency due to the stationary random process, but this cannot be done perfectly.)

The need for such attention becomes clear when we consider the effect of "small" displacements of the average. Suppose that most of the observations (say about 999 in 1000) lie between -100 to $+100$, with a few falling outside one limit or the other. This would be the case when the standard deviation is about 30, the variance about 900. If the average of the observations were 5 or even 10, we might or might not detect at a glance its failure to be zero.

The total power is the square of the average (dc power) plus the variance. Numerically, perhaps $25 + 900 = 925$ or $100 + 900 = 1000$. All the dc power belongs to the very lowest frequency band, whose width is

$$\Delta f = \frac{1}{2T'_n} .$$

If we have data at one second intervals for a period of 15 minutes, a total of 900 points, we will have a folding frequency of one-half cycle per second, and 900 elementary frequency bands before we reach the folding frequency. Thus up to one tenth of all the power may be concentrated in one 900th of the spectrum, so that the lowest frequency band has a power density up to 90 times that of the average of the 899 others. It is not surprising that precautions need to be taken to deal with such possibilities. (After all, our standard spectral windows have side lobes more than 1 per cent the height of the main lobe.)

Slow trends, which may reasonably be regarded as zero-frequency sine waves, just as constant displacements are regarded as zero frequency cosine waves, are not nearly so likely to involve quite so substantial excesses of power density, but instances of this may and do arise.

Any way of dealing with these effects must essentially remove the

lowest elementary frequency band, or both this band and the next to lowest one. In the process it will also have to eliminate some parts of the next higher elementary bands as well, since we cannot design a filtering procedure entirely free of side lobes. Two classes of ways of doing this are important. Either the X_i's can be linearly altered, as by subtracting the mean of them all from each of them, before the mean lagged products are calculated — calculated from modified data as if they were original data — or additional computations may be made and combined with either the mean lagged products or their cosine series transforms. Thus, for example, the mean of all data may be calculated and the square of this mean subtracted from each and every mean lagged product. The effect of all of these modifications can, however, be summarized as applying the finite cosine series transform to

$$C_r - E_{kr}$$

where k identifies a specific method of modification, rather than to the C_r alone.

In place of

$$\text{ave } \{P_{iA}(f)\} = Q_{iA}(f) * P_A(f),$$

we shall now have

$$\text{ave } \{P_{iAk}(f)\} = Q_{iAk}(f) * P_A(f) = [Q_{iA}(f) - R_{ik}(f)] * P_A(f)$$

where $R_{ik}(f)$ is related to the E_{kr} in the same way that $Q_i(f)$ is related to the C_r.

Details for certain special choices for E_{kr} are given in Section B.19. It is there concluded that, among others, satisfactory choices for practical calculation appear, for the present, to be, for removing possible constants,

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

and, for removing the effects of both possible constants and possible linear trends,

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

where $\overline{X^+}$ and $\overline{X^-}$ are the means of the right- and left-hand thirds of the X values.

WARNING: It will *almost never be wise* to fail to use some E_{kr} in a digital computation.

Analysis in Practice

The two sections which follow discuss the questioning and planning required whenever a digital analysis of equally spaced data is to be made, and exhibit a sample sequence of calculation formulas which might result from such planning. They are intended to summarize the previous material in its application to analysis. (Application to planning for measurement is treated next after this.)

20. Practical Analysis of an Equally Spaced Record

We may logically and usefully separate the analysis of an equally spaced record into four stages — each stage characterized by a question:

(a) Can the available data provide a meaningful estimated spectrum?

(b) Can the desires of the engineer for resolution and precision be harmonized with what the data can furnish?

(c) What modifications of the data are desirable or required before routine processing?

(d) How should modification and routine processing be carried out?

Failure to adequately consider any one question properly, or failure to apply any one answer, can make the entire analysis worthless.

The data presented will have come about by measuring some physical phenomenon at regular intervals. Thus,

1. the spectrum of the phenomenon
2. the frequency response of the instruments used to make the measurements
3. the probable magnitudes of measuring, and recording or reading errors, and
4. the time separation between adjacent values

are all relevant.

The first stage of consideration is to inquire generally about these quantities, and to determine whether either aliasing (see Section 12) or background noise is so heavy as to make the values almost wholly useless. Thus, if the spectrum is believed to extend up to 10 megacycles with substantial intensity, if the measuring equipment is flat to 1.2 kilocycles and is 60 db down at 5 kilocycles, and if the values are measured every $\frac{1}{200}$ of a second, we may as well stop here and go no further, since the whole available spectrum (up to 100 cycles) will be aliased more than a dozen times over. (The 1.2 kilocycle measurement bandwidth, which will be aliased 12 layers deep, will control rather than the 10 megacycle phenomenon bandwidth.)

If, on the other hand, the equipment was flat to 10 cycles, down about

6 db at 20 cycles, 15 db at 30 cycles, and 60 db at 50 cycles, we would not expect any irremovable aliasing difficulties, and would expect to be able to estimate the spectrum up to some moderate frequency — up to, say, 20 cycles, 30 cycles, or 40 cycles, depending upon how much background noise was present. (The energy above 100 cycles would not be recorded.)

In the next stage we should inquire into

1. the frequency resolution required
2. the fractional accuracy of estimation required, and
3. the total duration of data available, and the number of pieces into which it falls.

Items 1 and 3 can be combined and converted into the approximate number of elementary frequency bands (number of degrees of freedom — see Section 9 which is based on Sections 6 to 8) possibly available for each of the proposed estimates. This number can then be compared with the number of degrees of freedom required (also see Section 9) to give the desired fractional accuracy. If these are consistent, or if the desired accuracy, or the desired resolution, or both can be modified to make them consistent, then there is a good chance that the data can be persuaded to yield the desired results, and further inquiry is indicated. If not, we should stop here.

Explicit relations among duration, resolution, and fractional accuracy, the latter expressed in terms of 90 per cent interval (cp. Tables I and II), are given in Section B.23. These lead to an approximate 90 per cent spread, expressed in db (decibels), of

$$\frac{14}{\sqrt{(\text{total duration in secs})(\text{resolution in cps}) - \frac{1}{2} - \frac{1}{3}\,(\text{number of pieces})}}$$

a result which may often be conveniently used in such an inquiry.

At the beginning of the third stage, information should be sought as to

1. over what range of frequencies the spectrum is desired, and
2. whether any lines or high and narrow peaks are to be expected, and at what frequencies.

Guided by this information, it should be possible to decide whether either

a. smoothing by groups (as in Section 17) to reduce computation without loss of low-frequency information, or

b. rejection filtration (as in Section 16) to suppress well-established lines or high and narrow peaks,

or both, are desirable. If desirable, they are then carried out before or during the next step.

Unless advance information about the spectrum is exceedingly good,

a pilot analysis (see Section 18) to establish the rough form of the spectrum will now be very much worthwhile. The result (or the very good advance information, if available) will now make it possible to choose a reasonable prewhitening procedure (or, possibly, to choose not to prewhiten). Once suitable prewhitening (see Sections 11 and 15) has been chosen, and either carried out or planned for, the third stage is complete.

Finally, the information on resolution and accuracy combine to specify the width of spectral window desired, and hence (see Section 13) the number of lags for which mean lagged products should be calculated. When these are in hand, they are modified and transformed (or, perhaps more simply, transformed and convolved — see Section 13), adjusted to screen out very low frequencies, and the resulting power density estimates are corrected for the prewhitening, and for grouping and/or rejection filtration (if any) used. The final estimates are best plotted on a logarithmic power scale, since their accuracy will be roughly constant on this scale. Crude confidence limits can then be calculated from the number of degrees of freedom (see Section 9) which would be present in the individual estimates if: (i) the process were Gaussian, and (ii) the prewhitened spectrum were flat. (The factor of safety of Section 8 will ordinarily be adequate.)

21. SAMPLE COMPUTING FORMULAS

We cannot prescribe one set of computing formulas for general use, since there are rational reasons for different choices. All we can do is illustrate a procedure which may work fairly well in many cases. (And our example is not likely to be the only one with such properties. If the reader understands, by comparison with adjacent sections, just why we do what we do, he can compare other procedures with this example in a meaningful way. He will have to understand much of what is said in order to do this.)

If X_t, $t = 0, 1, \cdots, n$ are the given observations, which we will treat as if at unit spacing, it is likely that $P_A(f)$ decreases substantially as f goes from 0.0 to $0.5 = f_N$. (If it does not, then aliasing is likely to have been serious, and satisfactory analysis at this spacing may be impossible.) Prewhitening by

$$\tilde{X}_t = X_t - 0.6\, X_{t-1}$$

which multiplies $P_A(f)$ by

$$1.36 - 1.20 \cos 2\pi f,$$

a factor increasing from 0.16 to 2.56, *may* be a wise prewhitening. (The index t will now start at 1, and not at zero.)

We calculate next

$$C_r' = \frac{1}{n-r}\sum_{1}^{n-r} \tilde{X}_t \tilde{X}_{t+r} - \left(\frac{1}{n}\sum_{1}^{n} \tilde{X}_t\right)^2,$$

namely mean lagged products with an adjustment for the mean. (Further adjustment for a linear trend might have been necessary. See Section 19.) Let us suppose that we do this for $r = 0, 1, 2, \cdots, 24 = m$. (Some other choice may have been appropriate.)

Next we calculate the finite cosine series transform

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

and the results of hanning (see Sections 5 and 13)

$$\begin{aligned} U_0 &= \tfrac{1}{2}(V_0 + V_1) \\ U_r &= \tfrac{1}{4}V_{r-1} + \tfrac{1}{2}V_r + \tfrac{1}{4}V_{r+1}, \qquad 1 \leqq r \leqq m-1, \\ U_m &= \tfrac{1}{2}V_{m-1} + \tfrac{1}{2}V_m. \end{aligned}$$

These can then be corrected for both prewhitening and the correction for the mean by forming (see Section B.21)

$$\frac{n}{n-m}\,\frac{1}{1.36 - 1.20\cos\dfrac{2\pi}{6m}}\,U_0,$$

$$\frac{1}{1.36 - 1.20\cos\dfrac{2r\pi}{2m}}\,U_r, \qquad 1 \leqq r \leqq m-1,$$

$$\frac{1}{1.36 - 1.20\cos\left(1 - \dfrac{1}{6m}\right)2\pi}\,U_m,$$

as smoothed estimates of the power density. Estimates with subscript 0 will apply in the range just above zero frequency, those with subscript r near a frequency of $r/(2m)$ cycle per observation, and those with subscript m in the range just below a frequency of 0.5 cycle per observation.

In interpreting these estimates four cautions are important:

(a) aliasing of frequencies (see Section 12) may have taken place,

(b) the estimates are smoothed with a crudely isosceles triangular weighting function (see Sections 5 and 13) of full width $4/(2m)$,

(c) no estimate will be more stable than chi-square on $(2n)/m$ degrees of freedom and, wherever the spectrum is not smooth, the stability of the estimates will be appreciably less (see Section 9),

(d) adjacent estimates will not have independent sampling errors, though those not adjacent are at least very close to being uncorrelated.

The units involved are such that the smoothed one-sided, aliased power density on $0.0 \leqq f \leqq 0.5$ is approximated by twice the estimates. The pieces into which the variance would be divided, each coming from a frequency band of width $1/(2m)$ cycles per observation, are estimated by $1/(2m)$ times the corrected estimates.

Planning for Measurement

Up to this point, with the exception of part of Section 11, our discussion has been concerned (i) with what happens when certain operations are performed, and hence (ii) with how we should make the best of what we already have.

The third aspect — planning the measurements or observations to meet requirements — has not been adequately treated. (Both statisticians and engineers concerned with measurement will agree that this is the most vital aspect of all, but will, unfortunately, also have to admit that, all too often, "salvage" work will be required because this third aspect was omitted, and the observations made unwisely.)

In discussing "What data shall we take?", "How shall we measure it?", the same considerations will recur as in discussing "How shall we analyze it?", but (i) they will be looked at from quite different aspects and (ii) they will be even more important. Now, by planning in advance of data-gathering, we may be able either to replace useless or difficult-to-analyze measurements by usable ones, or to avoid making measurements which could never provide the desired information.

The first basic decision has to do with the type of recording and analysis to be used. Three types are in use today:

(1) *Spaced:* Analog use of intermittent recorders (photography of situations or of dials, etc.) or digital recording at equally spaced intervals (electronic reading of dials, photography of counters, etc.).

(2) *Mixed:* Continuous recording (on film, calibrated paper rolls, etc.) with the intention of analyzing equally spaced values to be "read" from these continuous records.

(3) *Continuous:* Continuous recording (FM recording on magnetic tape, etc.) with the intention of making an analog analysis.

The choice among these types will depend on their particular advantages and disadvantages, and on the availability of equipment, both for recording and analysis. In almost every case, however, the detailed problems will be surprisingly similar.

22. CHOICE OF FREQUENCY RESPONSE

In each instance there will be a problem of the response of the observing and transmitting or recording elements to high frequencies. When less quantitative studies are made, it is usual to worry whether the high-frequency response is large enough to "follow" the phenomena precisely. To be sure, if recording is only at intervals, and the needle is so blurred as not to be read, the high-frequency response may indeed be reduced by filtering. Such filtering is too likely to be regarded as unfortunate rather than helpful. Effort tends always to be applied for "faithful" recording. This is appropriate for recording specific *individual* time histories for *visual* study, but is often *most inappropriate* for recording *sample* time histories for *statistical* study with the aid of sensitive *filters* (analog or digital). (When the recording is continuous, be it on film, oscillograph paper, or magnetic tape, the "writing" means has a limited frequency response, and this will usually help to keep the record from blurring.)

When the analysis is to be made on equally spaced data, whether the recording be continuous or equi-spaced, there is a real problem of aliasing. And there is need for a basic choice of a frequency cutoff, usually in terms of two frequencies such that (i) the experiment is only concerned with frequencies up to the lower one, and (ii) frequencies beyond the upper one will not be recorded. The need for such a choice in a continuous system may not appear to be so acute, since only problems of noise or non-linear distortion are involved (see Section 11). Yet in practice, it will almost always be made — indirectly — by the choice of a writing speed (which implies a frequency cutoff for a continuous recorder). Economic pressures to reduce both the volume of record, and the extent of measurement and computation, act to lower the frequency cutoff, while desires to follow the spectrum to higher frequencies act to raise it. The proper choice comes from balancing these pressures.

Sometimes in mixed systems, when continuously recorded data is to be subjected to equi-spaced analysis, an attempt is made to compromise matters by recording with a high cutoff, and then asking that the measurements of this record be "eye averages" over periods long enough for the record to show considerable variation. Such compromises do not seem to work nearly as well in practice as their proponents suppose. Re-

placing the "averages" by the results of "reading to the line" at equi-spaced points often seems to give better results, even though a smaller, but unknown amount of aliasing is thus replaced by a larger, known amount. Putting the filtering into the observing and writing equipment, rather than into the (human) measurer and transcriber, will usually do even better — better by a large margin.

If one can be confident of the upper limit, beyond which the power spectrum will not be needed, it is usually best to record with a related frequency cutoff, thus reducing noise complications, aliasing difficulties, and the necessary bulk of the record.

Conversely, however, points must be recorded or measured frequently enough (or a high-enough writing speed used) so that aliasing (or loss of high-frequency response) is not serious. (For a given maximum usable frequency, the sharper the cutoff, the less stringent this requirement.)

To summarize, the problems surrounding aliasing should lead to the choice of a frequency cutoff which is usefully described by two frequencies (which may reasonably be in the ratio of 1 to 2):

(a) a lower frequency, which is the highest at which important power spectrum estimates will be made, and

(b) a higher frequency, at and above which no serious amount of recording is done.

Both of these need to be chosen before settling finally on observing and recording equipment. If equi-spaced data is produced, the folding frequency may be as low as half-way between these two frequencies.

A prime essential to keep in mind is that all measurement, transmission, *and analysis* systems are essentially band-limited. It is always inadvisable to try to cover too many octaves of log frequency while using exactly the same techniques.

23. DURATION OF DATA REQUIRED

Instead of trying to compromise resolution and stability within the limitations of available data, we may now consider the costs and advantages of getting still more data, or, perhaps, somewhat less data. We face a three-way compromise among effort, resolution, and stability (precision) of estimate.

Effort has to be measured in various ways, but the duration of initial record will almost certainly have to be considered as one measure. It is shown in Section B.23, where both precise definitions of the quantities, and a corresponding formula for the necessary numbers of pieces of a given length will also be found, that

$$(\text{total duration in seconds}) = \frac{\frac{1}{2} + \frac{200}{(90\%\text{ range in db})^2} + \frac{(\text{pieces})}{3}}{(\text{resolution in cps})}.$$

If, for example, a resolution of 0.1 cps is to be obtained from 6 pieces of record and is to furnish stability of ± 2 db for (on the average) $\frac{9}{10}$ of the individual estimates, then the necessary duration will be

$$\frac{\frac{1}{2} + \frac{200}{4^2} + \frac{6}{3}}{0.1} = 150 \text{ seconds.}$$

This applies equally to analog processing of continuous records or to digital processing of spaced records, so long as we apply the best methods which we know to a shape of spectrum which is not exceptionally difficult to handle.

24. AMOUNT OF DIGITAL DATA-HANDLING REQUIRED

If spaced data are to be digitally processed, both the number of data points to be used and the number of multiplications involved are of interest.

If we can easily build in the desirable frequency cutoff, and have to resolve a number of equally spaced bands spaced evenly from zero frequency to some maximum frequency, then we will require about

$$\left[\frac{3}{2} + \frac{600}{(90\%\text{ range in db})^2} + (\text{pieces})\right] (\text{number of bands resolved})$$

data points and, roughly about

$$\left(\frac{9}{2} + \frac{1800}{(90\%\text{ range in db})^2} + 3\ (\text{pieces})\right) (\text{number of bands resolved})^2$$

multiplications.

These last two results often give only preliminary indications. Aliasing difficulties will increase these numbers. The possibility of smoothing by groups will decrease them. Details and possible modifications of the proposed system of data gathering and analysis need to be studied carefully before final estimates of the number of data points and the rough number of multiplications are finally settled upon.

25. QUALITY OF MEASUREMENT AND HANDLING

In every case, careful consideration should be given to the quality of measurement and data handling required (in terms of the dangers of,

e.g.: time-varying frequency response, introduced noise, intermodulation distortion, etc.). An extensive catalog would be out of place here, since the problems are basically those of instrumentation engineering. But a few reminders may indicate the diversity of problems which might arise.

A camera may be "clamped" to some object to record the relative orientation of that object and something visible to the camera. The mounting of the camera is never perfectly rigid, and vibrations will occur ordinarily at frequencies far above the data-taking rate. Whatever the frequency, these vibrations will introduce "noise" into the record. At least an order-of-magnitude calculation of the effects of likely vibration is needed.

Storage of a signal on magnetic tape will be a part of many measurement-analysis systems. Because only rough spectra are wanted, AM (amplitude modulation) recording may be planned. If the fact that AM recording and playback is subject to considerable fluctuation in over-all gain (db's, not tenths db) is neglected, measurement planning may be quite misleading.

In a complex analysis, where several spectra and cross-spectra (whose analysis we have not specifically discussed) are involved, it might be planned to plot the estimates of each spectrum and cross-spectrum against frequency, draw smooth curves, and compute derived quantities from values read from these curves. Such a process has led to great difficulties in certain actual situations, because of the "noise" introduced by such visual smoothing which appears to have distinctive but unknown properties. Such a graphical step may appear to be good engineering, but it cannot be high quality data handling. Its use may nullify the careful selection of other data processes, some of which are delicately balanced.

Graphical analysis should ordinarily be reserved for:

(a) display of whatever spectrum or function of spectra is really a final output,

(b) description of the actual effects of computational procedures, and

(c) trouble-shooting.

26. EXAMPLE A

Suppose first that the spectrum of some aspect of the angular tracking performance of a new radar is to be obtained; that angular tracking can only be studied by photographing the target with a camera clamped to the antenna; that frequencies near 0.27 cps are of special interest; that the spectrum of tracking performance at higher frequencies is relatively flat up to 10 cps and then falls rapidly enough to be negligible beyond 40

cps; that estimates at all frequencies up to 25 cps are desired; and that stability to ± 1 db is derived. What are the requirements?

The total amount of tracking required is fixed by the resolution requirement near 0.27 cps, which we may suppose to be either 0.05 cps or 0.02 cps. These lead, respectively to durations of

$$\left(\frac{1}{2} + 50 + \frac{p}{3}\right)\frac{1}{0.05} > 1000 \text{ seconds}$$

and

$$\left(\frac{1}{2} + 50 + \frac{p}{3}\right)\frac{1}{0.02} > 2500 \text{ seconds.}$$

Single stretches of either 16 or 40 minutes continuous tracking are almost certain to be out of the question. The length of piece available would depend on the aspect of tracking performance studied, but a fair figure for this illustration might be 200 seconds. Going to Section B.23 for the necessary formula, we find

$$\text{(number of pieces)} = \frac{\frac{1}{2} + 50}{(200)(0.05) - \frac{1}{3}} = \frac{50.5}{9.67} = 5.2$$

or

$$\text{(number of pieces)} = \frac{\frac{1}{2} + 50}{(200)(0.02) - \frac{1}{3}} = \frac{50.5}{3.67} = 13.7.$$

From a purely experimental point of view, these amounts of data are moderately hard to substantially hard to obtain, but we may suppose them available as far as radar and target availability are concerned.

We come next to data taking and availability problems. We must study the spectrum up to 25 cps. Since the spectrum is negligible only above 40 cps, our folding frequency must be at least 32.5 cps, which would fold 40 cps exactly back to 25 cps. Hence we need at least 65 frames a second. Consideration of available frame rates bring us to 64 frames a second as probably reasonable. This is 12,800 frames in each 200 second piece, a total film reading load of between 50 and 150 thousand frames. This will require some hundreds of man-days of film reading, but may perhaps be faced.

To calculate *directly* the rough number of multiplications involved, we

may begin by assuming that we are going to require the 0.05 or 0.02 cps resolution all the way from 0 to 25 cps. Were this the case, then we would require to resolve from

$$\frac{25}{0.05} = 500$$

to

$$\frac{25}{0.02} = 1{,}250$$

frequency bands. The corresponding numbers of multiplications range from

$$[4.5 + 450 + 3 \text{ (pieces)}]\ (500)^2 \approx 120 \text{ million}$$

to

$$[4.5 + 450 + 3 \text{ (pieces)}]\ (1{,}250)^2 \approx 750 \text{ million.}$$

The running time of an IBM 650 calculator on such a problem is about 10 hours per million multiplications, so that between

$$1{,}200 \text{ hours} = 30 \text{ shift-weeks}$$

and

$$7{,}500 \text{ hours} = 188 \text{ shift-weeks}$$

would be required. Clearly these machine times are out of line, and attention should be given to ways of reducing this aspect of effort.

An application of smoothing by groups seems most likely to be effective, especially since the high resolution is only wanted near the low frequency of 0.27 cps. Let us suppose that, in view of the supposed rather flat spectrum out to 10 cps, the engineers concerned will be content with two spectrum analyses, one with 0.5 cps resolution extending all the way to 25 cps, and the other with 0.02 cps resolution extending only to 1 cps. What effect will this have on the computational load?

Notice first that it will have no effect on the radar-and-target operating and film-reading loads. These were fixed by the resolution-precision requirements, and by the combination of this with the upper limit of the actual spectrum affecting the camera. Replanning details of the analysis will save nothing on either of these.

The broad-frequency low-resolution analysis will resolve about

$$\frac{25}{0.5} = 50 \text{ bands}$$

and require roughly

$$[4.5 + 450 + 3(14)] \cdot (50)^2 = 1.24 \text{ million multiplications}$$

(since we shall need 14 pieces to obtain the required precision at a resolution of 0.02 cps). This would require about 12.4 hours machine time, a quite reasonable amount.

The preparation of data for the low-frequency high-resolution analysis — if we follow the suggestion of Section 17, requires less than 1.5 additions per original frame, since each datum contributes to four means. This is at most 0.2 million additions and can probably be combined with the next step so as not to involve substantial machine time.

The conduct of the low-frequency high-resolution analysis will resolve about

$$\frac{1.0}{0.02} = 50 \text{ bands}$$

and will require about another 12.4 hours of machine time.

Thus we have reduced machine time to about 25–30 hours, in pleasant contrast with the remaining requirements of some hundreds of hours of film reading and 14 test runs of 200 seconds each. The balance is approximately restored.

Our apparently blind use of the multiplications-required formula has concealed one important point. Our calculation of the time required for the high-frequency low-resolution analysis tacitly assumed that we have processed no more of the data than is required to meet the actual resolution-precision requirement.

The loosening of resolution from 0.02 cps to 0.5 cps in this part of the analysis has reduced by a factor 25 the amount of data which must be processed to meet the ± 1 db (90 per cent) requirement. Hence the two hours machine time is predicated on processing only $\frac{1}{25}$th of the available data. If only about $\frac{1}{25}$ of the data is to be processed for the high frequency analysis, then it will be desirable to take the most typical 8 or 10 seconds from each piece. The losses due to end effects will be somewhat greater, it is true, but the advantages of increased coverage of the effects of unplanned variation, consequent on using parts of all 14 runs, far outweigh such considerations.

It would be possible to use only one run for the high-frequency analysis, a possibility which emphasizes the fact that $\frac{3}{4}$ths of the film reading is done to obtain the raw material for averaging, for filtering out high frequencies. If the hundreds of man-days of film reading look out of line, *and if* the line from the radar to the target is known not to change

rapidly (with respect to an inertial frame of reference), then we are driven to consider whether the "clamping" of the camera to the antenna could modified in such a way as to provide a frequency cutoff between antenna position and camera position. What would be desired would be a reliable mechanical filter with a cutoff at 1 or 2 cps, and substantial, reproducible transmission up to, say, 0.5 cps. If such a mount could be taken down from the shelf, then it would suffice to make (a) one 200-second run with a stiff mount and 64 frames per second, and, say, (b) thirteen 200-second runs with a mount of such designed softness, and, say 4 frames per second. The total number of frames for reading would now be 12,800 for run (a) and 800 for each run (b), a total of about 23,000 frames. This might require about a man-month to read, a saving of several man-months. Unfortunately, such a sharply-tuned low-pass mount would not be likely to be on the shelf.

27. EXAMPLE B

As a second example, suppose a new solid-state device develops a noise voltage with a power spectrum roughly proportional to $1/f^2$ when under test under most extreme circumstances — circumstances so extreme that its average life is 30 to 50 milliseconds, and suppose that the detailed behaviour of this spectrum is believed likely to provide a clue to the proper theoretical treatment of some of the properties of this device. Suppose further that, while it was believed that the shape of the spectrum of the noise from different examples of this device was the same, the voltage levels of different devices were quite different. It might be reasonable to ask for spectral measurements to ± 0.25 db resolving 1 cps and covering from 1 cps to 500 cps. Direct measurements are likely to be most difficult, for the power between 499 and 500 cps is about $\frac{1}{120,000}$th the power between 1 and 2 cps, a difference of 51 db in level. Our recording and processing equipment is not likely to have the dynamic range required for direct analysis.

Clearly we should prewhiten our noise as early in the measurement and analysis system as we reasonably can. Fortunately, prewhitening here is

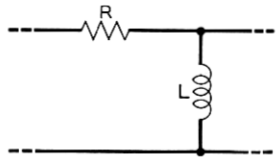

Fig. 12 — RL voltage divider.

operationally simple. A RL voltage divider, as indicated in Fig. 12 will introduce an attenuation of voltage, if the load impedance is high, amounting to

$$\left| \frac{R + j\omega L}{j\omega L} \right|^2 = 1 + \frac{R^2}{\omega^2 L^2}.$$

If the original spectrum were

$$P(f) = \frac{A}{f^2} = \frac{4\pi^2 A}{\omega^2},$$

then the prewhitened spectrum would be

$$\frac{\frac{4\pi^2 A}{\omega^2}}{1 + \frac{R^2}{\omega^2 L^2}} = \frac{4\pi^2 A L^2}{R^2 + \omega^2 L^2}$$

which will be initially constant, and then decrease 6 db per octave, with a corner at $\omega_c = R/L, f_c = R/2\pi L$. As a first step in measuring a spectrum out, to say, $f = 2R/2\pi L$, at which frequency the prewhitened spectrum would be down about 7 db, such a change would be useful. The range of frequencies which could be usefully studied would not be appreciably reduced by such a change, even though the low frequency power level would be greatly reduced by the prewhitening network, since the low-frequency power level would not be seriously reduced below the former power level at the corner frequency. If one could have been studied, the other can be studied.

28. EXAMPLE C

The irregularities in the earth's rotation have been studied by Brouwer,[17] who reduced the available observations (times of occultation and meridian passage) by averaging over individual years. He states "occultations so reduced in recent years have been demonstrated to yield annual means essentially free from systematic errors if the observations are well distributed over the year. . . . The δ's may themselves be the accumulations of numerous smaller random changes with average intervals much smaller than a year. The astronomical evidence throws no further light on this, though perhaps something may be gained by an analysis of residuals in the moon's mean longitude taken by lunations."[18] These comments suggest that astronomical data can supply values once a year, possibly no more frequently, and may be able to supply values

about 13 times a year (once per lunation), certainly no more frequently. Let us accept the first possibility as a basis for an example. (This is the best example we know of a situation where equally spaced data cannot, in principle, be had at a finer spacing.)

The information most nearly directly supplied by the astronomical observations is Δt, the difference between ephemeris time and mean solar time. Brouwer discusses two statistical models for its structure, both of which are most easily described in terms of the behavior of the second differences of the observations. In the first, the true second differences are constant over periods of varying length. In the second model, the true second differences are independently and randomly distributed. In either case, observational errors, independent from observation-period to observation-period also contribute to the observed Δt's.

If we were to plan an observational program to decide between these hypotheses by spectral analysis we need first to specify the alternative spectra. The first model seems never to have been made as precise statistically as the second. Brouwer's fitted curves correspond to constancy over periods of from 4 to 15 years. We should like to get a general idea of the possible spectra corresponding to this model without making the model too specific. Consider first a situation in which, except for the effects of second differences of experimental errors, the observations are constant in blocks of five, and where the values assigned to different blocks are independent. The successive average lagged products (starting with lag zero) are proportional to 5, 4, 3, 2, 1, 0, 0, 0, . . . and it follows that the power density is proportional to

$$1 + \frac{8}{5}\cos \pi f/f_N + \frac{6}{5}\cos 2\pi f/f_N + \frac{4}{5}\cos 3\pi f/f_N + \frac{2}{5}\cos 4\pi f/f_N .$$

Calculation shows that this is high near zero frequency, falling rapidly until, beyond about $f/f_N = 0.3$, it consists of ripples with an average height of less than $\frac{1}{25}$th the low frequency peak. If, instead of "constant by fives", the specific model were "constant by eights" or "constant by tens", still with independence between blocks, this peaking would be more pronounced and confined to still lower frequencies. If the lengths of the blocks were to vary at random, according to some distribution, still with independence of value, the spectrum would be the corresponding average of such spectra for fixed block lengths. The spectrum to be expected *for second differences* of annual average observations then should consist of a sum of two components:

(1) a "true" component peaked at low frequencies, falling rapidly by, say, $f/f_N = 0.2$ or 0.25, and continuing to $f/f_N = 1.00$ with an average height perhaps 1 per cent or 2 per cent of the low-frequency value,

(2) an "observational error" component, corresponding to independent errors in the annual averages, and hence proportional to $(1 - \cos(\pi f/f_N))^2$.

In case the second model should apply, the first component would be replaced by one with a flat density.

Fig. 13 shows the shapes of the three possible components. The natural way to try to distinguish between the two models by spectral analysis is to compare the spectral density in the middle range, say $f/f_N = 0.25$ to 0.5 with that in a lower range, say below $f/f_N = 0.25$. According to Model I, the low-range density should be substantially higher than the middle-range density, the latter consisting of the effects of observational error (whose strength can be well estimated at the upper end of the spectrum). According to Model II, the middle-range density should be slightly to somewhat greater than the low-range density, the increment representing effects of observational error.

Without more detailed estimates of the relative sizes of the components, it would be difficult to specify exactly how many observations would be required to separate Model I from Model II, but 10 to 20 degrees of freedom in each of the ranges discussed should be quite helpful. This suggests 100 values of annual second differences, corresponding to 102 years of careful astronomy, as likely to be helpful. Since Brouwer gives annual values for 131 years, some 129 annual second differences are

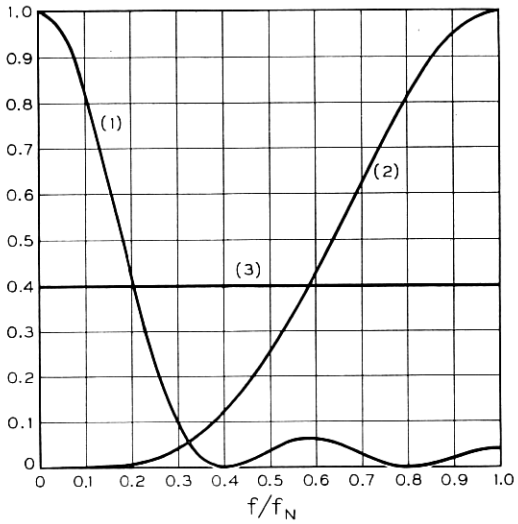

Fig. 13 — Components for two models of earth-rotation irregularities: (1) "true irregularity" component for first model, (2) "observational error" component for either model, (3) "true irregularity" component for second model.

available for trial, and it may be possible to answer the question without waiting for many more years to pass.

It might well suffice to estimate smoothed densities over octaves such as $0.0625 \leqq f/f_N \leqq 0.125$, $0.125 \leqq f/f_N \leqq 0.25$, $0.25 \leqq f/f_N \leqq 0.5$ and $0.5 \leqq f/f_N \leqq 1$. Thus we might consider using the add-and-subtract pilot estimation method for initial exploration. The actual analysis of Brouwer's data is considered further in Section B.28.

Appendix A

Fundamental Fourier Techniques

In this appendix we review briefly certain aspects of Fourier transformation. These aspects may be regarded as dealing mainly with diffraction by slits, rectangular or graded, and by analogs made up of discrete "lines". Convolution and the so-called Dirac functions are specially important as convenient tools. Some parts of the discussion will have no direct bearing on the analysis of procedures for power spectrum estimation, but are intended to familiarize the reader with analytical tools which are used frequently throughout the remainder of this paper, and which may be used to advantage in many other analyses of a similar nature.

A.1 *Fourier Transformation*

There are several formulations of Fourier transformation which differ according to custom, convenience, or taste. The formulation which we will adopt here is the one used by Campbell and Foster.[19] Given a function of time, $G(t)$, its Fourier transform is a function of frequency, and is given by the formula

$$S(f) = \int_{-\infty}^{\infty} G(t) \cdot e^{-i\omega t}\, dt \qquad (\omega = 2\pi f).$$

Conversely, given a function of frequency, $S(f)$, its Fourier transform is a function of time, and is given by the formula

$$G(t) = \int_{-\infty}^{\infty} S(f) \cdot e^{i\omega t}\, df \qquad (\omega = 2\pi f).$$

The term "frequency" is used here, not in the probability or statistical sense, but in the sense of sinusoidal or cisoidal functions of time ($\cos \omega t$, $\sin \omega t$, $e^{i\omega t}$).

Our preference for the Campbell-Foster formulation is based on the

following points, arranged approximately in the order of increasing weight.

1. Frequencies are expressed in cycles per second more naturally and much more frequently than in radians per second. (In our analysis we use ω only as an abbreviation of $2\pi f$, and only if it is typographically convenient.)

2. Except for the sign of the exponent in the kernels, the transformation formulae are symmetrical. The assignment of the signs here is the conventional one in transmission theory.

3. In most of the applications to communications problems, the frequency functions are rational functions of $p = i\omega$, with real coefficients. Hence, the reformulation of the transformation of $S(f)$ to $G(t)$ as

$$G(t) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} S\left(\frac{p}{2\pi i}\right) \cdot e^{pt}\, dp$$

is a natural and convenient step in the calculation of the integral by the method of residues.

4. The transformation formulae correspond to the conventional relations between the *impulse response* (response due to a unit impulse applied at $t = 0$) and the *transfer function* (ratio of steady-state response to excitation, for the complex excitation $e^{i\omega t}$) of a fixed linear transmission network. These network functional relations are commonly regarded as Laplace transformations rather than Fourier transformations. As a matter of fact, however, the circumstances in almost all practical applications are such that there is no essential difference between Laplace transformations and Fourier transformations. Impulse responses are zero for $t < 0$ and vanish exponentially as $t \to \infty$, and transfer functions are analytic on and to the right of the imaginary axis (including the point at infinity) in the complex p-plane. On the very rare occasions when a communications engineer might be interested in the behavior of a network under energetic initial conditions, he has ways of introducing the initial conditions without using Laplace transforms (Guillemin[20]).

It should be noted that, since $G(t)$ must be a real function, the real part of $S(f)$ must be an even function, and the imaginary part of $S(f)$ must be an odd function. The even part of $G(t)$ and the even (real) part of $S(f)$ are cosine-transforms of each other. The odd part of $G(t)$ and the odd (imaginary) part of $S(f)$ are negative sine-transforms of each other. It should be noted also that if $G(t)$ and $S(f)$ constitute a transform-pair, then $G(-t)$ and $S(-f)$ also constitute a transform-pair. Further, $S(-f)$ is equal to $S^*(f)$, the complex conjugate of $S(f)$.

A.2 *Some Transform-Pairs*

We will now turn our attention to some transform-pairs which we will require directly or indirectly in the analysis of procedures for power spectrum estimation. We will use special symbols for some of these transform-pairs. For later reference, these transform-pairs will be collected in Table IV.

The first transform-pair, which is easily worked out, involves a symmetrical rectangular time function (box car of length $2T_m$), viz.

$$\begin{aligned} D_0(t) &= 1, && |t| < T_m, \\ &= \tfrac{1}{2}, && |t| = T_m, \\ &= 0, && |t| > T_m. \end{aligned}$$

TABLE IV

1.	$D_0(t) = 1, \quad \lvert t \rvert < T_m$ $= \frac{1}{2}, \quad \lvert t \rvert = T_m$ $= 0, \quad \lvert t \rvert > T_m$	$Q_0(f) = 2T_m \frac{\sin \omega T_m}{\omega T_m}$ $= 2T_m \operatorname{dif} 2fT_m$
2.	$D_1(t) = 1 - \frac{\lvert t \rvert}{T_m}, \quad \lvert t \rvert \leqq T_m$ $= 0, \quad \lvert t \rvert \geqq T_m$	$Q_1(f) = T_m \left(\frac{\sin \pi f T_m}{\pi f T_m}\right)^2$ $= T_m (\operatorname{dif} fT_m)^2$
3.	$\delta(t - t_0)$	$e^{-i\omega t_0}$
4.	$\cos \omega_0 t$	$\frac{1}{2}[\delta(f + f_0) + \delta(f - f_0)]$
5.	$\nabla_m(t; \Delta t) = \frac{\Delta t}{2}\delta(t + m\Delta t) + \Delta t \cdot \sum_{q=-m+1}^{q=m-1} \delta(t - q\Delta t) + \frac{\Delta t}{2}\delta(t - m\Delta t)$	$Q_0(f; \Delta t) = \Delta t \cdot \cot \frac{\omega \Delta t}{2} \cdot \sin m\omega\Delta t$ $= 2(m \cdot \Delta t) \cos (\pi f \cdot \Delta t) \frac{\operatorname{dif} 2f(m \cdot \Delta t)}{\operatorname{dif} f \cdot \Delta t}$
6.	$\nabla(t; \Delta t) = \Delta t \cdot \sum_{q=-\infty}^{q=\infty} \delta(t - q\Delta t)$	$A\left(f; \frac{1}{\Delta t}\right) = \sum_{q=-\infty}^{q=\infty} \delta\left(f - \frac{q}{\Delta t}\right)$
7.	$A(t; \Delta t)$	$\nabla\left(f; \frac{1}{\Delta t}\right)$

The corresponding frequency function is

$$Q_0(f) = 2T_m \frac{\sin \omega T_m}{\omega T_m} = 2T_m \cdot \mathrm{dif}\, 2fT_m.$$

(The values assigned to $D_0(t)$ at the end points $|t| = T_m$ are those resulting from the transformation of $Q_0(f)$ to $D_0(t)$. Of course the values assigned at these two points do not influence the result of the transformation of $D_0(t)$ to $Q_0(f)$). Except for scale factors, this frequency function is the function $\mathrm{dif}\, u = \sin \pi u / \pi u$ which recurs constantly in this subject. It is often convenient to regard it as the diffraction pattern (in frequency) due to passage through a rectangular slot (in time). The behaviour of $\mathrm{dif}\, 2fT_m$ is shown in Fig. 14.

The second transform-pair, which is almost as readily worked out as the first, involves a symmetrical triangular time function, viz.

$$D_1(t) = 1 - \frac{|t|}{T_m}, \qquad |t| \leqq T_m,$$

$$= 0, \qquad |t| \geqq T_m.$$

The corresponding frequency function is

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

Except for scale factors, this frequency function behaves as shown in Fig. 14.

The third transform-pair involves a so-called Dirac *function* as the time function. The Dirac function is not a function in the strict mathematical sense. It is called a "measure" by L. Schwartz.[21] For our purposes, it will only be necessary to identify $\delta(t - t_0) \cdot dt$ formally with $dh(t - t_0)$ where $h(t - t_0)$ is Heaviside's unit-step function, viz.,

$$h(t - t_0) = 0, \qquad t < t_0$$

$$= 1, \qquad t > t_0$$

and to interpret all integrals as Stieltjes integrals. Hence if the time function (to use the term loosely) is

$$G(t) = \delta(t - t_0)$$

then, the corresponding frequency function is

$$S(f) = e^{-i\omega t_0}.$$

It should be noted that while $\delta(t - t_0)$ is easily formally transformed

into a frequency function, the latter is not so readily transformed into the original time function.

The fourth transform-pair involves a symmetrical pair of Dirac functions as the frequency function. Thus, the time function

$$G(t) = \cos \omega_0 t \qquad (\omega_0 = 2\pi f_0)$$

corresponds to the frequency function

$$S(f) = \tfrac{1}{2}[\delta(f + f_0) + \delta(f - f_0)].$$

If the reader is disturbed over the fact that we are evidently going to base our analysis, at least initially, on the use of Dirac functions, he should note that Dirac functions are always paired with functions which are used widely and freely in transmission theory although they are not realistic in a physical sense. Functions of time, such as $\cos \omega_0 t$, which represent an infinitely long past and future history of activity, are not a bit more realistic in a physical sense than are "infinitely sharp" lines in the frequency spectrum. Similarly, functions of frequency, such as $\exp(-i\omega t_0)$, whose absolute values do not vanish as $f \to \infty$, are not a bit more realistic than impulsive "functions" of time. Nevertheless, as we

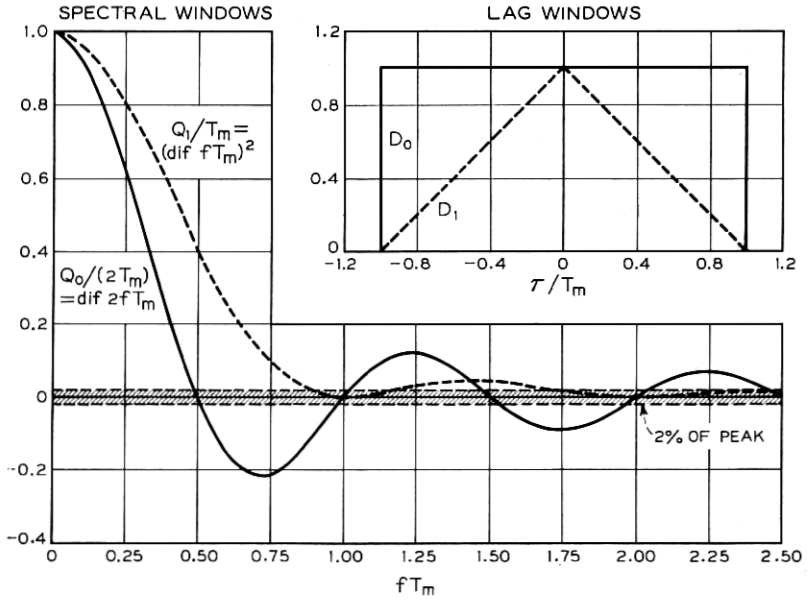

Fig. 14 — Lag windows D_0 and D_1. Spectral windows Q_0 and Q_1.

will see later on, these unrealistic pairs may be used as convenient bases for a wide variety of realistic pairs. They thus serve a very useful purpose.

The fifth transform-pair involves a *finite Dirac comb* as the time function, viz.

$$\nabla_m(t;\,\Delta t) \;=\; \frac{\Delta t}{2}\,\delta(t + m\Delta t) + \Delta t\cdot \sum_{q=m+1}^{q=m-1} \delta(t - q\Delta t) + \frac{\Delta t}{2}\,\delta(t - m\Delta t).$$

This is clearly a discrete approximation to $D_0(t)$ for $T_m = m\cdot\Delta t$. The corresponding frequency function, which is easily worked out with the help of the third transform-pair (summing the exponential terms before introducing trigonometric equivalents), is

$$Q_0(f\,;\,\Delta t) \;=\; \Delta t\,\cot\frac{\omega\Delta t}{2}\cdot\sin m\omega\Delta t \;=\; 2(m\cdot\Delta t)\,\cos\,(\pi f\cdot\Delta t)\,\frac{\mathrm{dif}\,2f(m\cdot\Delta t)}{\mathrm{dif}\,f\cdot\Delta t}\,.$$

Except for a scale factor, the initial behaviour of this frequency function is illustrated in Fig. 9. Clearly, since $\cos 0 = \mathrm{dif}\ 0 = 1$, the limit of $Q_0(f;\,\Delta t)$, when $\Delta t \to 0$ with $m\cdot\Delta t = T_m$ held constant, is $Q_0(f)$. This corresponds to the formal convergence of $\nabla_m(t;\,\Delta t)$ to $D_0(t)$.

We have defined this finite Dirac comb with a half-sized Dirac function at each end because the corresponding frequency function has smaller side lobes, relative to the main lobe, than for the finite Dirac comb with a whole Dirac function at each end. This is easily seen from the fact that the effect of adding a further half-sized Dirac function at each end of $\nabla_m(t;\,\Delta t)$ is to add $\Delta t\cdot\cos m\omega\Delta t$ to $Q_0(f;\,\Delta t)$.

The frequency function $Q_0(f;\,\Delta t)$ is periodic, with a period of $1/\Delta t$ cps. It is symmetrical about every integral multiple of $1/(2\Delta t)$ cps. Thus, it has an absolutely maximum value of $2m\cdot\Delta t$ at the integral multiples of $1/\Delta t$ cps. It is zero at the integral multiples of $1/(2m\Delta t)$ cps which are not integral multiples of $1/\Delta t$ cps. For large values of m and small values of $\omega\Delta t$, it behaves approximately like $Q_0(f)$.

The sixth transform-pair involves an *infinite Dirac comb* in time, and, as it turns out, also an infinite Dirac comb in frequency. The time function is the formal limit of $\nabla_m(t;\,\Delta t)$ as $m \to \infty$, namely,

$$\nabla(t;\,\Delta t) \;=\; \Delta t\cdot \sum_{q=-\infty}^{q=\infty} \delta(t - q\Delta t).$$

The corresponding frequency function is

$$A\left(f;\,\frac{1}{\Delta t}\right) = \sum_{q=-\infty}^{q=\infty} \delta\left(f - \frac{q}{\Delta t}\right) = \Delta t\cdot\nabla\left(f;\,\frac{1}{\Delta t}\right).$$

This may be surmised from the fact that

$$\int_{-1/(2\Delta t)}^{1/(2\Delta t)} Q_0(f;\Delta t)\,df = 1 \qquad \text{for any } m$$

while

$$\lim_{m\to\infty}\int_{-\epsilon}^{\epsilon} Q_0(f;dt)\,df = \frac{2}{\pi}\lim_{m\to\infty} Si(2\pi m\epsilon\Delta t), \quad \left(\text{where } Si(x) \equiv \int_0^x \frac{\sin y}{y}\,dy\right)$$

$$= 1 \text{ for any } \epsilon \text{ in } 0 < \epsilon < \frac{1}{2\Delta t}.$$

The result may indeed be obtained by applying the fourth transform-pair with $T_m = m\cdot\Delta t$ to the formal Fourier series representation of the infinite comb

$$\nabla(t;\Delta t) = 1 + 2\sum_{q=1}^{q=\infty}\cos\frac{2\pi qt}{\Delta t}.$$

Since

$$\nabla_m(t;\Delta t) = D_0(t)\cdot\nabla(t;\Delta t)$$

we also have, as we shall see in the next section,

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

$$= \sum_{q=-\infty}^{\infty} Q_0\left(f - \frac{q}{\Delta t}\right).$$

The seventh transform-pair arises from the sixth by dividing by Δt on both sides.

A.3 *Convolution*

If $G(t) = G_1(t)\cdot G_2(t)$, then the Fourier transform of $G(t)$ may be expressed in terms of those of $G_1(t)$ and $G_2(t)$ as follows.

$$S(f) = \int_{-\infty}^{\infty} G_1(t)\cdot G_2(t)\cdot e^{-i\omega t}\,dt,$$

$$= \int_{-\infty}^{\infty} G_1(t)\cdot\left[\int_{-\infty}^{\infty} S_2(\xi)\cdot e^{i2\pi\xi t}\,d\xi\right]\cdot e^{-i\omega t}\,dt,$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} G_1(t)\cdot e^{-i2\pi(f-\xi)t}\,dt\right]\cdot S_2(\xi)\,d\xi,$$

$$= \int_{-\infty}^{\infty} S_1(f-\xi)\cdot S_2(\xi)\,d\xi.$$

This relation, in which $S_1(f)$ and $S_2(f)$ are interchangeable, is commonly expressed in the symbolic form

$$S(f) = S_1(f) * S_2(f).$$

The implied operation on $S_1(f)$ and $S_2(f)$ is called a *convolution*. In particular, $S(f)$ is said to be the convolution of $S_1(f)$ with $S_2(f)$.

Similarly, if $S(f) = S_1(f) \cdot S_2(f)$, then

$$G(t) = \int_{-\infty}^{\infty} G_1(t - \lambda) \cdot G_2(\lambda)\, d\lambda$$

$$= G_1(t) * G_2(t).$$

Thus, multiplication and convolution constitute an *operational transform-pair*.

(Convolution is often called by a variety of names such as Superposition theorem, Faltungsintegral, Green's theorem, Duhamel's theorem, Borel's theorem, and Boltzmann-Hopkinson theorem.)

It may be noted in the detailed derivation above (putting $f = 0$), that

$$\int_{-\infty}^{\infty} G_1(t) \cdot G_2(t)\, dt = \int_{-\infty}^{\infty} S_1^*(f) \cdot S_2(f) \cdot df$$

where $S_1^*(f)$ is the complex conjugate of $S_1(f)$. This is *Parseval's theorem* of which a very useful special case is

$$\int_{-\infty}^{\infty} [G(t)]^2\, dt = \int_{-\infty}^{\infty} |\, S(f)\, |^2\, df.$$

An example of convolution is supplied by the symmetrical triangular time function in the second transform-pair. This time function is the convolution of two symmetrical rectangular time functions from the first transform-pair, with appropriate scalar adjustments. Another example is the infinite Dirac comb $\nabla(t;\, \Delta t)$, which may be regarded as the convolution of the finite Dirac comb $\nabla_m(t;\, \Delta t)$ with the infinite Dirac comb $A(t;\, 2m\Delta t)$, that is

$$\nabla(t;\, \Delta t) = \nabla_m(t;\, \Delta t) * A(t;\, 2m\Delta t).$$

As the reader may easily verify, this corresponds to

$$A\left(f;\, \frac{1}{\Delta t}\right) = Q_0(f;\, \Delta t) \cdot \nabla\left(f;\, \frac{1}{2m\Delta t}\right).$$

Convolution of time functions occurs in communications systems whenever a signal is transmitted through a fixed linear network. If the input

signal is $G_1(t)$, and if the impulse response of the network is $W(t)$, then the output signal is*

$$G(t) = \int_{-\infty}^{\infty} W(t - \lambda) \cdot G_1(\lambda)\, d\lambda$$

$$= W(t) * G_1(t).$$

The so-called *linear distortion* of the signal due to transmission through the network can be (and occasionally is) examined in terms of the effects of convolution, but the common practice among circuit engineers is to conduct the examination in terms of the corresponding frequency functions. There are good reasons for this common practice. The most important of these reasons are:

1. The relation between the frequency functions is simpler, viz.

$$S(f) = Y(f) \cdot S_1(f)$$

where $Y(f)$ is the transfer function of the network.

2. The effects of *amplitude distortion* of the signal and of *phase distortion* (of the unmodulated signal) may be examined independently. While phase distortion is critical in the transmission of pictures (facsimile), it is relatively unimportant in the transmission of speech or music.

3. The transmission characteristics of fixed linear networks are most easily calculated or measured accurately in terms of frequency rather than time.

4. Fixed linear network design techniques based on frequency functions are today much further developed (simpler, more powerful, and more versatile) than those based on time functions.

Convolution of frequency functions occurs in communications systems whenever a carrier wave is amplitude-modulated by a signal. If the input signal is $G_1(t)$, and if the carrier wave is $\cos \omega_0 t$, then the output signal, with *suppressed carrier*, is

$$G(t) = G_1(t) \cdot \cos \omega_0 t$$

* It may be of some help here to think of λ as "excitation time", and of t as "response time". In the equivalent formulation

$$G(t) = \int_{-\infty}^{\infty} W(\tau) \cdot G_1(t - \tau)\, d\tau$$

we may think of $\tau = t - \lambda$ as the "age" of input data at response time.

At this point attention is called to a device which will be used many times to simplify analysis, which is to use $-\infty$ and $+\infty$ as limits of integration, letting the integrand take care of the effective range of integration. In this case, if $G_1(\lambda) \equiv 0$ for $\lambda < t_0$, and $W(\tau) \equiv 0$ for $\tau < 0$, the effective range of integration would be $t_0 < \lambda < t$ or $0 < \tau < t - t_0$.

and the relation among the corresponding frequency functions is

$$S(f) = S_1(f) * \tfrac{1}{2}\,[\delta(f + f_0) + \delta(f - f_0)]$$
$$= \tfrac{1}{2}\,S_1(f + f_0) + \tfrac{1}{2}\,S_1(f - f_0).$$

The convolution of frequency functions corresponding to the amplitude-modulation of a carrier wave is so naturally visualized simply as shifting the signal spectrum (frequency function) that it is almost never visualized in any other way. It should be observed, however, that this point of view depends critically upon the two-sided specification of the signal spectrum, in amplitude and phase, to give the correct picture of the sidebands, whether the amplitude-modulation scheme under consideration be double-sideband, single sideband, vestigial sideband, or two-phase (as in TV chrominance signals). Further, the two-sided specification of the modulated-carrier spectrum is essential for a correct picture of the demodulation process used to recover the signal.

For present purposes we will be interested in convolution not only as a tool for the synthesis of new transform-pairs but also as an analytical tool. For example, by regarding a time function $G(t)$ as the product of two other time functions $G_1(t)$ and $G_2(t)$ we can make use of the relation $S(f) = S_1(f) * S_2(f)$ to reach insights about $S(f)$ which do not come easily from the explicit form of $S(f)$.

To make convolution a useful analytical tool, we have to visualize it in some convenient way. This may be done in three ways. The relative merits of these three points of view depend upon the circumstances in any particular case.

In the first place, convolution may be visualized as a stretching process. For example, in the equation

$$G(t) = \int_{-\infty}^{\infty} G_1(t - \lambda)\cdot G_2(\lambda)\, d\lambda$$

we visualize $G_2(\lambda)\cdot d\lambda$ as a rectangular element of $G_2(t)$, originally concentrated at $t = \lambda$. This rectangular element is then stretched into the area under the elementary curve $G_1(t - \lambda)\cdot G_2(\lambda)\cdot d\lambda$ regarded as a function of t. This elementary curve has the shape of $G_1(t)$ with origin shifted to $t = \lambda$. The total effect at any particular value of t is then obtained by integration over λ. In this example, we have regarded $G_1(t)$ as the "stretcher" operating on each element of $G_2(t)$. Of course, since convolution is commutative, we may interchange the roles of the two functions.

In the second place, if one of the functions in the convolution consists exclusively of Dirac functions, each Dirac function may be regarded as a "shifter" operating on the other function in the convolution. For ex-

ample,

$$\delta(t - a) * G(t) = \int_{-\infty}^{\infty} \delta(t - a - \lambda) \cdot G(\lambda)\, d\lambda = G(t - a).$$

In the third place, convolution may be visualized as a weighted integration with a moving weight function. For example, in the equation

$$G(t) = \int_{-\infty}^{\infty} G_1(t - \lambda) \cdot G_2(\lambda)\, d\lambda$$

we regard $G(t)$ as the integral of $G_2(\lambda)$ with weight function $G_1(t - \lambda)$. The position of the weight function with respect to the λ scale depends upon the value of t. In the event that the weight function has unit area, $G(t)$ may be regarded as the moving weighted average of $G_2(\lambda)$. (As previously noted, the roles of the two functions may be interchanged.)

As an example of the use of the ideas described above, let us assume that we have a function $G_0(t)$ which is zero outside of the interval $0 < t < T$, and for which the frequency function is $S_0(f)$. Let us generate a periodic function $G(t)$ by convolving $G_0(t)$ with $A(t; T)$. Then, since

$$G(t) = G_0(t) * A(t; T)$$

the frequency function corresponding to $G(t)$ is, from Item 7 of Table IV,

$$S(f) = S_0(f) \cdot \nabla\left(f; \frac{1}{T}\right).$$

As we expect, $S(f)$ consists of "lines" (of infinite height but finite area) at uniform intervals of $1/T$ cps. The complex intensities (areas) of these lines represent the amplitudes and relative phases of the terms in the conventional Fourier series representation of $G(t)$. Thus,

$$\begin{aligned} G(t) &= \int_{-\infty}^{\infty} S(f) \cdot e^{i\omega t}\, df \\ &= \frac{1}{T} \sum_{q=-\infty}^{q=\infty} S_0\left(\frac{q}{T}\right) \cdot e^{(i2\pi q t)/T}. \end{aligned}$$

As a second example, which is in a sense the dual of the first, let us assume that we have a function $G_0(t)$ for which the frequency function $S_0(f)$ is zero outside of the band $-f_0 < f < f_0$. Let us generate a discrete time series $G(t)$ by sampling $G_0(t)$ at uniform intervals of $1/(2f_0)$ seconds. If we regard sampling as a multiplication by (or as amplitude-modula-

tion of) an infinite Dirac comb, then

$$G(t) = G_0(t) \cdot A\left(t; \frac{1}{2f_0}\right).$$

Hence, the frequency function corresponding to $G(t)$ is

$$S(f) = S_0(f) * \nabla(f; 2f_0),$$

or, explicitly,

$$S(f) = 2f_0 \cdot \sum_{q=-\infty}^{q=\infty} S_0(f - 2qf_0).$$

If this frequency function is multiplied by a frequency function $S_1(f)$, where

$$S_1(f) = \frac{1}{2f_0}, \quad |f| < f_0$$
$$= 0, \quad |f| > f_0$$

it will revert to $S_0(f)$. Thus,

$$S_1(f) \cdot S(f) = S_0(f).$$

Hence, if $G_1(t)$ is the time function corresponding to $S_1(f)$, namely,

$$G_1(t) = \frac{\sin \omega_0 t}{\omega_0 t},$$

then

$$G_1(t) * G(t) = G_0(t).$$

Thus, sampling $G_0(t)$ to get the discrete time series $G(t)$, and convolving $G(t)$ with $G_1(t)$, restores $G_0(t)$ exactly. This result reflects the well-known *sampling theorem* in information theory. The effect of sampling $G_0(t)$ at uniform intervals of other than $1/(2f_0)$ seconds is readily visualized.

A.4 *Windows*

If a time function is even (and of course real), the corresponding frequency function is real (and of course even), and conversely. These circumstances will prevail when we deal with autocovariance functions, power spectra, and appropriate weight functions. Under these circumstances, the weight functions will be called *windows*. Such windows will be considered in transform-pairs, and the members of any pair will be distinguished as the *lag window*, and the *spectral window*.

Time windows convolved with periodic functions of time have been used by Guillemin,[22] under the name "scanning functions", to examine the behavior of weighted partial sums of Fourier series. We use them in Sections B.4 and B.10 where we call them *data windows*, and their Fourier transforms (which may be complex) *frequency windows*.

A.5 *Realistic Pairs from Unrealistic Pairs*

Transform-pairs which involve Dirac functions are very easily converted into a wide variety of realistic pairs. As an example, let us consider the sixth pair (infinite Dirac combs) which requires two convolutions for conversion to a realistic pair. If we convolve the time functions of the first and sixth pairs, taking $T_m \ll \Delta t$, we get a time function which represents an infinite train of narrow rectangular pulses of unit height. The corresponding frequency function still consists of Dirac functions but these now do not have a uniform intensity. If we next multiply the time function of this pair by the time function of the first pair, taking $T_m \gg \Delta t$, we get a time function which represents a long but finite train of narrow rectangular pulses. The corresponding frequency function is continuous and consists chiefly of very narrow peaks of finite height approaching zero as $f \to \infty$.

A sinusoidal carrier wave of finite though great length may be represented as the product of the time functions of the first and fourth pairs with $T_m \gg 1/f_0$. The corresponding frequency function is continuous and consists of very narrow peaks at $\pm f_0$, with much lower subsidiary peaks of height approaching zero as $f \to \infty$.

If the time function of the third pair is convolved with the time function

$$\begin{aligned} G(t) &= 0 && t < 0 \\ &= \frac{1}{T} e^{-t/T} && t > 0, \end{aligned}$$

the resultant frequency function is

$$S(f) = \frac{1}{1 + i\omega T} e^{-i\omega t_0}$$

of which the absolute value falls off asymptotically like $1/f$ as $f \to \infty$, however small $T(>0)$ might be.

In line with this discussion, it should be noted that a realistic "white noise" spectrum must be effectively band-limited by an asymptotic fall-off at least as fast as $1/f^2$. Under certain circumstances, however, we may assume that the spectrum is flat to any frequency. Let us suppose that

the spectrum is in fact

$$P(f) = \frac{\frac{\sigma^2}{\pi f_c}}{1 + \left(\frac{f}{f_c}\right)^2}, \qquad \int_{-\infty}^{\infty} P(f)\, df = \sigma^2$$

where σ^2 is the variance. The autocovariance is

$$C(\tau) = \sigma^2 \cdot e^{-\omega_c |\tau|}.$$

If we transmit this noise through a network with an effective cutoff frequency well below f_c, we may assume for an approximation that

$$P(f) \approx \frac{\sigma^2}{\pi f_c}$$

and, therefore, that

$$C(\tau) \approx \frac{\sigma^2}{\pi f_c}\, \delta(\tau)$$

although such an assumption is unrealistic if carried to indefinitely high frequencies (the input noise would have infinite variance). Hence, if the impulse response of the network is $W(t)$, the autocovariance of the output noise is

$$\begin{aligned} C_{\text{out}}(t_i - t_j) &= \text{ave}\left\{\int_{-\infty}^{\infty} W(\tau_1) X(t_i - \tau_1)\, d\tau_1 \right. \\ &\qquad \left. \cdot \int_{-\infty}^{\infty} W(\tau_2) \cdot X(t_j - \tau_2)\, d\tau_2 \right\} \\ &= \iint_{-\infty}^{\infty} W(\tau_1) W(\tau_2) \cdot C(t_i - t_j - \tau_1 + \tau_2) d\tau_1\, d\tau_2 \\ &\approx \frac{\sigma^2}{\pi f_c} \int_{-\infty}^{\infty} W(\tau_1) \cdot W(\tau_1 - t_i + t_j)\, d\tau_1 . \end{aligned}$$

In particular, the variance of the output noise is

$$C_{\text{out}}(0) \approx \frac{\sigma^2}{\pi f_c} \int_{-\infty}^{\infty} [W(\tau_1)]^2\, d\tau_1$$

which by Parseval's theorem is equivalent to

$$C_{\text{out}}(0) \approx \frac{\sigma^2}{\pi f_c} \int_{-\infty}^{\infty} |\, Y(f)\, |^2\, df$$

where $Y(f)$ is the transfer function of the network. These results are

realistic. (The variance of the output noise is finite and approximately correct).

A.6 *Some Trigonometric Identities*

In this section we develop some trigonometric identities which will be needed later on. We start with the equation

$$\sum_{-a}^{b} \cos(\psi + 2hu) = \frac{\sin(a + b + 1)u}{\sin u} \cos[\psi + (b - a)u]$$

which is easily obtained by substituting

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}$$

in the left-hand member, summing the exponential terms and making some elementary trigonometric substitutions. By substituting $\psi \pm \pi/2$ for ψ we then get

$$\sum_{-a}^{b} \sin(\psi + 2hu) = \frac{\sin(a + b + 1)u}{\sin u} \sin[\psi + (b - a)u].$$

Now, setting $u = \pi f$, and using the function introduced in Section A.2,

$$\frac{\sin pu}{p \sin u} = \frac{(\sin p\pi f)/p\pi f}{(\sin \pi f)/\pi f} = \frac{\text{dif } pf}{\text{dif } f},$$

which, on differentiation, yields

$$\frac{d}{df}\frac{(\text{dif } pf)}{(\text{dif } f)} = \left(\frac{\text{dif } pf}{\text{dif } f}\right)\left(p\frac{\text{dif}' pf}{\text{dif } pf} - \frac{\text{dif}' f}{\text{dif } f}\right).$$

Before we rewrite our summation formulas in terms of such ratios of "dif" functions, we need to appreciate their behavior. For p not very small, (dif pf)/(dif f) behaves much like the numerator for pf small and moderate. The effect of the denominator is to force symmetry around integer multiples of $\frac{1}{2}$, so that the peak at $f = 0$ is repeated at $f = 1, 2, 3, \cdots$, thus making its behavior consistent with aliasing. For $0 \leqq f \leqq \frac{1}{2}$ its other effects are minor, since in this range $(2/\pi) \leqq \text{dif } f \leqq 1$, while the extrema of dif pf have shrunk from $+1$ to $\pm 2/(p\pi)$. For most considerations, therefore, we can approximate this ratio by the numerator.

We now rewrite our summations as means, introducing (dif pf)/(dif f), finding

$$\frac{1}{a + b + 1}\sum_{-a}^{b} \cos(\psi + 2h\pi f) = \frac{\text{dif } (a + b + 1)f}{\text{dif } f} \cos[\psi + (b - a)\pi f]$$

$$\frac{1}{a + b + 1}\sum_{-a}^{b} \sin(\psi + 2h\pi f) = \frac{\text{dif } (a + b + 1)f}{\text{dif } f} \sin[\psi + (b - a)\pi f].$$

Differentiating with respect to f, and multiplying through by

$$-(a + b + 1)/(2\pi),$$

we get

$$\sum_{-a}^{b} h \sin(\psi + 2h\pi f) = \left(\frac{\operatorname{dif}(a + b + 1)f}{\operatorname{dif} f}\right) \cdot \left[\frac{b - a}{2}(a + b + 1) \sin[\psi + (b - a)\pi f] - \left(\frac{(a + b + 1)^2}{2\pi} \frac{\operatorname{dif}'(a + b + 1)f}{\operatorname{dif}(a + b + 1)f} - \frac{a + b + 1}{2\pi} \frac{\operatorname{dif}' f}{\operatorname{dif} f}\right) \cos[\psi + (b - a)\pi f]\right]$$

with a similar formula for

$$\sum_{-a}^{b} h \cos(\psi + 2h\pi f).$$

We shall now use these formulas to obtain results about the average values of certain quadratic functions of chance variables $X_0, X_1, \cdots, X_n$. The average value of any such quadratic function can be represented in terms of a corresponding spectral window $Q(f)$ in the form

$$\int_0^\infty Q(f) \cdot 2P(f)\, df$$

whenever

$$\text{ave}\{X_t X_{t+q}\} = \int_0^\infty \cos 2\pi q f \cdot 2P(f)\, df$$

for all suitable integers t and q, since the quadratic function can be expressed as a sum of multiples of terms of the form $X_t X_{t+q}$. To determine the height, $Q(f_0)$, of the spectral window corresponding to a specific quadratic function, it suffices to consider the special case $2P(f) = \delta(f - f_0)$, for which ave $\{X_t X_{t+q}\} = \cos 2\pi q f_0$, when the average value of the quadratic function for such a special set of X_t is exactly $Q(f_0)$.

If ave $\{X_t X_{t+q}\} = \cos 2\pi q f$, we easily find that

$$\text{ave}\left\{\left(\frac{1}{a + b + 1} \sum_{-a}^{b} X_h\right)\left(\frac{1}{c + d + 1} \sum_{-c}^{d} X_g\right)\right\} = \frac{1}{a + b + 1} \sum_{-a}^{b} \frac{1}{c + d + 1} \sum_{-c}^{d} \cos 2\pi f(g - h)$$

$$= \frac{1}{a+b+1} \sum_{-a}^{b} \frac{\operatorname{dif}(c+d+1)f}{\operatorname{dif} f} \cos(-2\pi fh + (d-c)\pi f)$$

$$= \frac{\operatorname{dif}(a+b+1)f}{\operatorname{dif} f} \frac{\operatorname{dif}(c+d+1)f}{\operatorname{dif} f} \cos(d-c-a+b)\pi f$$

$$\approx \operatorname{dif}(a+b+1)f \operatorname{dif}(c+d+1)f \cos(d-c-a+b)\pi f$$

any of these expressions being the spectral window corresponding to

$$\left(\frac{1}{a+b+1} \sum_{-a}^{b} X_h\right)\left(\frac{1}{c+d+1} \sum_{-c}^{d} X_g\right).$$

Making the same assumption, we find that (where $n = 2\ell + 1$)

$$\operatorname{ave}\left\{\left(\sum_{-\ell}^{\ell} hX_h\right)\left(\sum_{-\ell}^{\ell} gX_g\right)\right\}$$

$$= \sum_{-\ell}^{\ell} \sum_{-\ell}^{\ell} gh \cos 2\pi f(g-h)$$

$$= \sum_{-\ell}^{\ell} g \left(\frac{\operatorname{dif} nf}{\operatorname{dif} f}\right)\left(\frac{n^2}{2\pi} \frac{\operatorname{dif}' nf}{\operatorname{dif} nf} - \frac{n}{2\pi} \frac{\operatorname{dif}' f}{\operatorname{dif} f}\right) \sin 2\pi fg$$

$$= \frac{n^4}{4}\left(\frac{\operatorname{dif} nf}{\operatorname{dif} f}\right)^2 \left(\frac{1}{\pi} \frac{\operatorname{dif}' nf}{\operatorname{dif} nf} - \frac{1}{n\pi} \frac{\operatorname{dif}' f}{\operatorname{dif} f}\right)^2.$$

These expressions therefore represent the spectral windows corresponding to

$$\left(\sum_{-\ell}^{\ell} hX_h\right)^2.$$

Glossary of Terms

Add-and-subtract method

A method of roughly estimating spectra based on successive additions by non-overlapping two's followed by a differencing. (18, B.18 and B.28.)

Alias

In equally spaced data, two frequencies are aliases of one another if sinusoids of the corresponding frequencies cannot be distinguished by their equally spaced values (this occurs when $f_1 = 2kf_N \pm f_2$ for integer k); the *principal aliases* lie in the interval $-f_N \leqq f \leqq f_N$. (See also 14.) (Also *aliased*, *aliasing*, etc.)

Aliased spectrum

See Spectrum, aliased.

Analysis, pilot

Any of a number of methods of obtaining a rough spectrum, including the add-and-subtract method (18, etc.) the cascade method (B.18), the complete add-and-subtract method (B.18).

Autocorrelation function

The normalized autocovariance function (normalized so that its value for lag zero is unity).

Autocovariance function

The covariance between $X(t)$ and $X(t + \tau)$ as a function of the lag τ. If averages of $X(t)$ and $X(t + \tau)$ are zero, it is equal to the average value of $X(t) \cdot X(t + \tau)$. It can be defined for a whole ensemble, a whole function stretching from $-\infty$ to $+\infty$, or for a finite piece of a function; in the latter case it is called the *apparent autocovariance function* (see 4). Certain related functions are called modified apparent autocovariance functions (also see 4).

Autoregressive series

A time series generated from another time series as the solution of a

linear difference equation. (Usually where previous values of output series enter into determination of current value.)

Average

The arithmetic mean, usually over an ensemble, a population, or some reasonable facsimile thereof.

Band-limited function

Strictly, a function whose Fourier transform vanishes outside some finite interval (and hence is an entire function of exponential type); practically, a function whose Fourier transform is very small outside some finite interval.

Box-car function

A function zero except over a finite interval, in the interior of which it takes a constant value (often $+1$).

Cardinal theorem (of interpolation theory)

A precise statement of the conditions under which values given at a doubly infinite set of equally spaced points can be interpolated (with the aid of the function $(\sin (x - x_i))/(x - x_i)$ to yield a continuous band-limited function. (See B.1.)

Cascade process (of spectral estimation)

A process of spectral estimation in which a single step is repeated again and again, each step yielding both certain estimates and a condensed set of data (ready for input to the next step). (See B.18.)

Chi-square

A quantity distributed (strictly exactly, but practically approximately) as $x_1^2 + x_2^2 + \cdots + x_k^2$ where $x_1, x_2, \cdots, x_k$ are independent and Gaussian, and have average zero and variance unity.

Continuous power spectrum

A power spectrum representable by the indefinite integral of a suitable (spectral density) function. (All power spectra of physical systems are continuous.)

Convolution

The operation on one side of a Fourier transformation corresponding to multiplication on the other side. (See A.3 for detailed discussion.)

Cosine transform

A series (see 13) or integral (see 2) transform in which a cosine of the product of the variables is the kernel.

Covariance

A measure of (linear) common variation between two quantities, equal to the average product of deviations from averages. (See 1.)

Cross-spectrum

The expression of the mutual frequency properties of two series analogous to the spectrum of a single series. (Because mutual relations at a single frequency can be in phase, in quadrature, or in any mixture of these, either a single complex-valued cross-spectrum or a pair of real-valued *cross-spectra* are required.) (Also *cross-spectral.*)

Data

As specifically used in this paper, values given at equally spaced intervals of time (often called time series).

Data window

A time function which vanishes outside a given interval and which is regarded as multiplying data or signals defined for a more extended period. (Data windows are usually smooth (graded) to improve the quality of later frequency analysis.)

Degrees of freedom

As applied to chi-square distributions arising from quadratic forms in Gaussian (normal) variables, the number of linearly independent squared terms of equal size into which the form can be divided. In general, a measure of stability equal to twice the square of the average divided by the variance.

Delta-component

A finite contribution to the spectrum at one frequency (B.10 only).

Diffraction function

$$\text{dif } x \equiv \frac{\sin \pi x}{\pi x}.$$

Dirac comb

An array of equally spaced Dirac functions, usually most of which are of equal height.

Dirac function

The limit of functions of unit integral concentrated in smaller and smaller intervals near zero. (See A.2 for fuller discussion.)

Distortion

Failure of output to match input. (Often specified as to kind of failure as *linear*, *amplitude*, *phase*, *non-linear*, etc., cp. A.3.)

Effective record length

Actual length of record available reduced to allow for end effects. (See 6.)

Elementary frequency band

An interval of frequency conveniently thought of as containing "a single degree of freedom", equal to the reciprocal of twice the duration of observation or record. (Since both sines and cosines may occur, it requires *two* elementary frequency bands to contain "an independently observable frequency.")

Ensemble

A family of functions (here functions of either continuous or equi-spaced time) with probabilities assigned to relevant sub-families.

Equivalent number (of degrees of freedom)

See second sentence under *degrees of freedom.*

Equivalent width

The extent of a function regarded as a window as expressed by the ratio of the square of its integral to the integral of its square. (See 8.)

Filtered spectrum

Spectrum of the output from any process which can be regarded as a filter.

Folding frequency

The lowest frequency which "is its own alias", that is, is the limit of both a sequence of frequencies and of the sequence of their aliases, given by the reciprocal of twice the time-spacing between values, also called *Nyquist* frequency.

Fourier transform

Operations making functions out of functions by integration against a kernel of the form exponential function of $\sqrt{-1}$ times frequency times time. Often, including here, defined differently for transforming time functions into frequency functions than for transforming frequency functions into time functions. (See A.1 for details.)

Frequency

A measure of rate of repetition; unless otherwise specified, the number of cycles per second. The *angular frequency* is measured in radians per second, and is, consequently, larger by a factor of 2π.

Frequency window

The Fourier transform of a data window.

Gaussian

A single quantity, or a finite number of quantities distributed according to a probability density representable as e to the power minus a quadratic form. (Also called *normal*, *Maxwellian*, etc.) Also, a function or ensemble, distributed in such a way that all finite sections are Gaussian. (See 1.)

Hamming

The operation of smoothing with weights 0.23, 0.54 and 0.23. (After R. W. Hamming.)

Hanning

The operation of smoothing with weights 0.25, 0.50 and 0.25. (After Julius von Hann.)

Hyperdirective antenna

An antenna or antenna system so energized as to have a more compact directional pattern than naturally corresponds to its extent (as measured in wavelengths).

Impulse response

The time function describing a linear system in terms of the output resulting from an input described by a Dirac function.

Independence (statistical, of estimates)

In general, two quantities are statistically independent if they possess a joint distribution such that (incomplete or complete) knowledge of one does not alter the distribution of the other. Estimates are statistically independent if this property holds for each fixed true situation.

Independent phases

An ensemble has independent phases when it can be approximated by ensembles consisting of finite sums of (phased) cosines (of fixed frequencies) whose phases are mutually independent. Continuous spectrum and independent phases imply Gaussian character. Every Gaussian ensemble has independent phases.

Intermodulation distortion

Non-linear distortion, especially as recognized in the output of a system when two or more frequencies enter the input simultaneously.

Joint probability distribution

Expression of the probability of simultaneous occurrence of values of two or more quantities.

Lag

A difference in time (epoch) of two events or values considered together.

Lag window

A function of lag, vanishing outside a finite interval, and either multiplying or regarded as multiplying the quantities of a family of quantities with differing lags.

Lagged product

The product of two values corresponding to different times. (In a *mean lagged product* the lags are usually all the same.)

Lead

The negative of lag.

Line (in a power spectrum)

Theoretically, and as used in this paper, a finite contribution associated with a single frequency. Physically, not used here, a finite contribution associated with a very narrow spectral region.

Lobe

A bulge, positive or negative, especially in a spectral window. (In most spectral windows, a large central *main lobe* is surrounded on both sides by smaller *side lobes*.)

Mean lagged product

The (arithmetic) mean of products of equally lagged quantities.

Moving linear combination

A transformation expressing the values of an output time series as linear combinations of values of the input series in specified relations of lag (or lead).

Negative frequencies

When sines and cosines are jointly represented by two imaginary exponentials, one has a positive frequency and the other a negative frequency. (Not specifiable for a single time function in real terms.)

Network (linear)

In this account, an otherwise unspecified physical device which converts an input function (of continuous time) *linearly* into an output function (of continuous time).

Noise

In general, an undesired time-function, or component of a function.

Non-normality

Failure to follow a normal or Gaussian distribution.

Normality

The property of following a normal or Gaussian distribution.

Nyquist frequency

The lowest frequency coinciding with one of its own aliases, the reciprocal of twice the time interval between values (same as *folding frequency*).

Octave

An interval of frequencies, the highest of which is double the lowest.

Pilot (*analysis* or *estimation*)

A process yielding rough estimates of spectral density intended mainly as a basis for planning more complete and precise analyses.

Population

A collection of objects (in particular, of numbers or of functions), with probabilities attached to relevant subcollections.

Power transfer function

The function expressing the ratio of output power near a given frequency to the input power near that frequency.

Power-variance spectrum

A function of frequency, in terms of which the variances and covariances of a family of spectral estimates can be expressed in standard form. (See 6 and 14 for details in the continuous and equi-spaced cases, respectively.)

Preemphasis

Emphasis of certain frequencies (in comparison with others), before processing, as an aid to the quality of result.

Prewhitening

Preemphasis designed to make the spectral density more nearly constant (the spectrum more nearly flat).

Principal alias

An alias falling between zero and plus or minus the folding or Nyquist frequency.

Process (*random* or *stochastic*)

An ensemble of functions. (Often composed of functions of time regarded as unfolding or developing.)

Protection ratio

The ratio of transmission at a desired frequency to the transmission at an undesired alias of that frequency.

Recording

Is *spaced* when originally taken at equal intervals, *mixed* when taken continuously and processed at equal intervals, *continuous* when taken and processed on a continuous basis.

Resolution

A measure of the concentration of a spectral estimate expressed in frequency units, here taken (for the important cases) as equal to the width of the major lobe. (See B.23.)

Resolved bands (number of)

The ratio of the Nyquist or folding frequency to the resolution.

Sampling theorem (of information theory)

Nyquist's result that equi-spaced data, with two or more points per cycle of highest frequency, allows reconstruction of band-limited functions. (See *Cardinal theorem.*)

Serial correlation coefficients

Ratios of the autocovariances to the variance of a process, ensemble, etc.

Signal

A time function desired as (potentially) carrying intelligence.

"Signal"

A function of continuous time, which may be either a signal, a noise, or a combination of both. (Contrasted with *data*, a function of discrete time.)

Single function approach

A mode of representing certain ensembles by the translations of a single time function (in single function terms).

Smoothed function

The result of weighted averaging of nearby values of the original function.

Smoothing

In the narrow sense, forming (continuous or discrete) moving linear combinations with unit total weight.

Smoothing and decimation procedure

A procedure which may be regarded as the formation of discrete moving linear combinations, followed by the omission of all but every kth such. (See 17 and B.17.)

Spectrum (*also power spectrum*)

An expression of the second moments of an ensemble, process, etc. (i) in terms of frequencies, (ii) in such a form as to diagonalize the effects on second moments of time-invariant linear transformations applied to the ensemble or process. (adjective: *spectral*).

Spectrum, aliased

For equally spaced data, the *principal part* of the *aliased spectrum* expresses contributions to the variance in terms of frequencies between zero and the Nyquist or folding frequency, all contributions from frequencies having the same principal alias and sign having been combined by addition. (The *aliased spectrum* repeats the principal part periodically with period $2f_N$. See 14.)

Spectral density

A value of a function (or the entire function) whose integral over any frequency interval represents the contribution to the variance from that frequency interval.

Spectral density estimates

Estimates of spectral density, termed *raw* when obtained from equispaced mean lagged products by cosine series transformation, *refined* when hanned or hammed from raw estimates or obtained by an equivalent process. (See B.13.)

Spectral window

A function of frequency expressing the contribution of the spectral density at each frequency to the average value of an estimate of (smoothed) spectral density.

Stationary (*ensemble* or *random process*)

An ensemble of time functions (or random process) is stationary if any translation of the time origin leaves its statistical properties unaffected.

Superposition theorem

A statement that the output of a linear device is the convolution of its input with its impulse response. (See B.2.)

Temporally homogeneous

Sometimes used in place of *stationary*, especially when speaking of stochastic processes.

Transfer function

The *transfer function* of a network or other linear device is a complex-valued function expressing the amplitude and phase changes suffered by cosinusoidal inputs in becoming outputs. (See A.5.) The square of the absolute value of the transfer function is the *power transfer function*, which expresses the factors by which spectral densities are changed as inputs become outputs. (See 4.)

Transmission

The coefficient with which power at a given frequency contributes to power at the (new) principal alias as a result of the application of a smoothing and decimation procedure.

Transversal filtering

Time domain filtering by forming linear combinations of lagged values, use of moving linear combinations for filtering. (See Kallmann[32] for the origin of this term.)

Trend

A systematic, smooth component of a time function (time series), as, for example, a linear function of time (a *linear trend*).

True

Often used to refer to average values over the ensemble, as contrasted with mean values over the observations.

Universe

A collection of objects (numbers, functions, etc.) with probabilities attached to relevant subcollections.

Variance

A quadratic measure of variability, the average squared deviation from the average.

White noise

An ensemble whose spectral density is sensibly constant from zero frequency through the frequencies of interest (in equi-spaced situations, up to the folding or Nyquist frequency). (The values of equi-spaced white noise at different times are independent.)

Window

A function expressing, as a multiplicative factor, the tendency or possibility of the various values of some function to enter into some calculation or contribute to the average value of some quantity. (See *data*, *lag*, *spectral*, etc. for specific instances.)

Windowless quadratic

A quadratic expression is windowless if its average value vanishes for every stationary ensemble of finite variance (See B.19).

Window pair

Two windows related by a Fourier transformation, as lag and spectral windows or data and frequency windows. (See A.4 and 4.)

Zero-frequency waves (cosine and sine)

The limiting forms of very-low-frequency cosinusoids, namely constants and linear trends. (See 19.)

ACKNOWLEDGEMENT

Many persons have contributed by discussion, questioning and suggestion to the development of these ideas and techniques. An attempt to list all those who have helped during the eight years or so of growth and development could not succeed. Thus, we must express our thanks, generally, to many colleagues interested in transmission theory, statistics or computation, here in Bell Telephone Laboratories, and at Princeton and other universities, and to many scientists and engineers who have used the techniques to solve their own problems, and by real examples have shown us what had to be considered. Special thanks must go to Miss Dorothy T. Angell for patient and careful computations during numerical exploration, and to R. W. Hamming for continuing active discussion, especially of computational and expository problems. A careful reading of the final draft by S. O. Rice led to the elimination of many slips and ambiguities (of course those which remain are the authors' sole responsibility).

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*, Vols. 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 Ed. 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 periodes, 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.)