Astron. & Astrophys. 13, 169–189 (1971)

The Analysis and Restoration of Astronomical Data via the Fast Fourier Transform

J. W. Brault
Kitt Peak National Observatory*
Tucson, Arizona

O. R. White
High Altitude Observatory
National Center for Atmospheric Research*
Boulder, Colorado

Received January 22, 1971

Since the powerful techniques developed in communications theory have been little used to Fourier analyze astronomical measurements and correct them for smearing, we discuss the application of Fourier transforms and the Fast Fourier Transform algorithm to these problems. Basic sampling theory and the discrete Fourier transform are presented first, and then applied to the analysis of solar time series and to the correction of line spectra for observational smearing. The solution of the empirical restoration problem is based on a filter technique, which suppresses the noise and corrects for smearing in an optimum fashion.

Key words: restoration — power spectra — Fourier transform — sampling

I. The Purpose of this Discussion

Our motivation in writing this paper stems from the realization that fundamental concepts and numerical techniques developed in communications theory are not generally known or fully exploited in observational astronomy today. As a result, a large amount of astronomical data is published in its raw form, which usually means that it is completely uncorrected for either noise or smearing effects. This situation is no longer necessary if the measurements are made and reduced by a mathematically-sound scheme such as one based on discrete Fourier transforms and optimum filter techniques.

Considering the amount of effort devoted to the theoretical analysis of astronomical observations and the often-divergent results from such work, the observers are obliged to present measurements as free of observational distortions as possible. Therefore, our goal is to describe a general method that we find useful in numerically analyzing and improving discrete measurements of astronomical spectra without significantly affecting their true information content.

The correction of astronomical data for smearing effects, i.e., restoration, is a recurrent problem that has not been satisfactorily solved on a routine basis because of practical difficulties in the inversion of the convolution integral in the presence of noise. In the past, Fourier transform techniques were not applied routinely to empirical spectral analysis and restoration because of the long times required for calculation of discrete Fourier transforms of empirical data. The Fast Fourier Transform algorithm (referred to as the FFT throughout the following discussion) now reduces the computing times to very acceptable values, and hence, permits practical application of Fourier techniques to solutions of these problems. Although our experience has been in the analysis of solar time series and the restoration of measurements of the solar Fraunhofer spectrum, the generality of the Fourier transform approach does not limit it to these two particular one-dimensional problems.

In this paper we emphasize the basic concepts through explanations and illustrations, but at the expense of mathematical completeness. However, this subject has been thoroughly investigated in communications theory, and the reader is referred to IEEE Audio Transactions. *Special Issue on the Fast Fourier Transform* (1967), Bracewell (1965),

* Both Kitt Peak National Observatory and the National Center for Atmospheric Research are sponsored by the National Science Foundation.

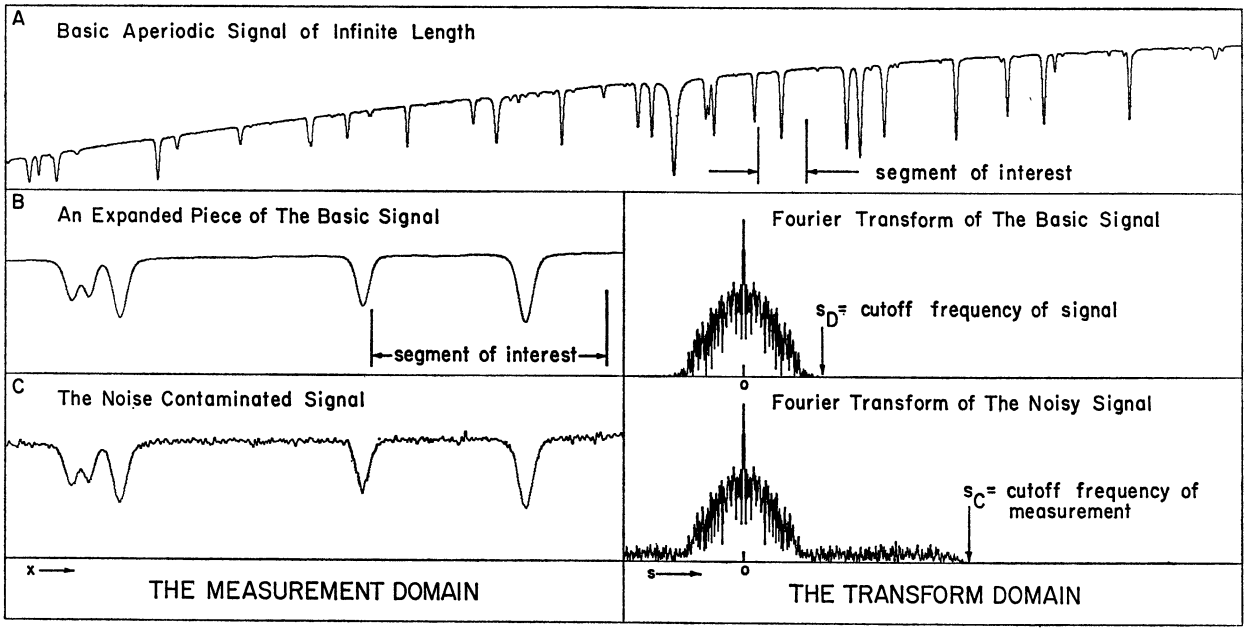

Fig. 1. Illustrations of the basic continuous signal, the noise-contaminated signal, and their transforms

Middleton (1960), Cooley *et al.* (1970a, 1970b), Bergland (1969) and Gold *et al.* (1969) for more complete discussions. The reader is expected to be familiar with the shift, convolution, and power theorems applicable to Fourier transforms (see Bracewell, 1965). The following notation will be used through the text:

$$\text{Complex Fourier Transform of } F(x) = \tilde{F}(s)$$
$$\text{Complex Conjugate of } \tilde{F}(s) = \tilde{F}(s)^* \qquad (1)$$
$$\text{Convolution of } F(x) \text{ with } A(x) = F(x) * A(x)$$

where x is the independent variable in the measurement domain, and s is the corresponding frequency variable in the transform domain.

II. Fundamentals

A. Introduction

In astronomical problems, an infinitely-long, smooth signal is assumed to underlie the measurements. This signal can be either the intensity variation with wavelength in a spectrum, the intensity variation with time in a light curve, or generally any measurable quantity that varies with time, position, or optical wavelength. We measure a portion of such a signal by either recording the photometer output voltage as a function of time in the case of a light curve or by generating a time-varying signal through scanning across a photographic plate or scanning the spectrum directly. The measurement can only cover a finite portion of the basic signal, and it will appear noisy because of fluctuations either inherent in the signal or introduced by the measuring apparatus. In this section we center the discussion on the effects of sampling and finite measurement length on a noise-contaminated signal and its Fourier transform. As we shall see, the sampling theorem will give the conditions under which discrete measurements of such a signal will have a Fourier representation suitable for use in both the restoration and spectral analysis problems. The discrete Fourier transformation and its inverse will also be defined so as to be consistent with the later discussion of the FFT.

B. Definition of the Measurement Process and the Discrete Fourier Transform

1. The Basic Continuous Signal and its Fourier Transform

Eventually we are to be concerned with the measurement of a segment of a long, continuous signal such as the smooth solar Fraunhofer spectrum shown in Figure 1A. However, in order to explain the effect of such a measurement process in the transform domain, we begin with the idealization in which the basic underlying signal is recorded continuously over

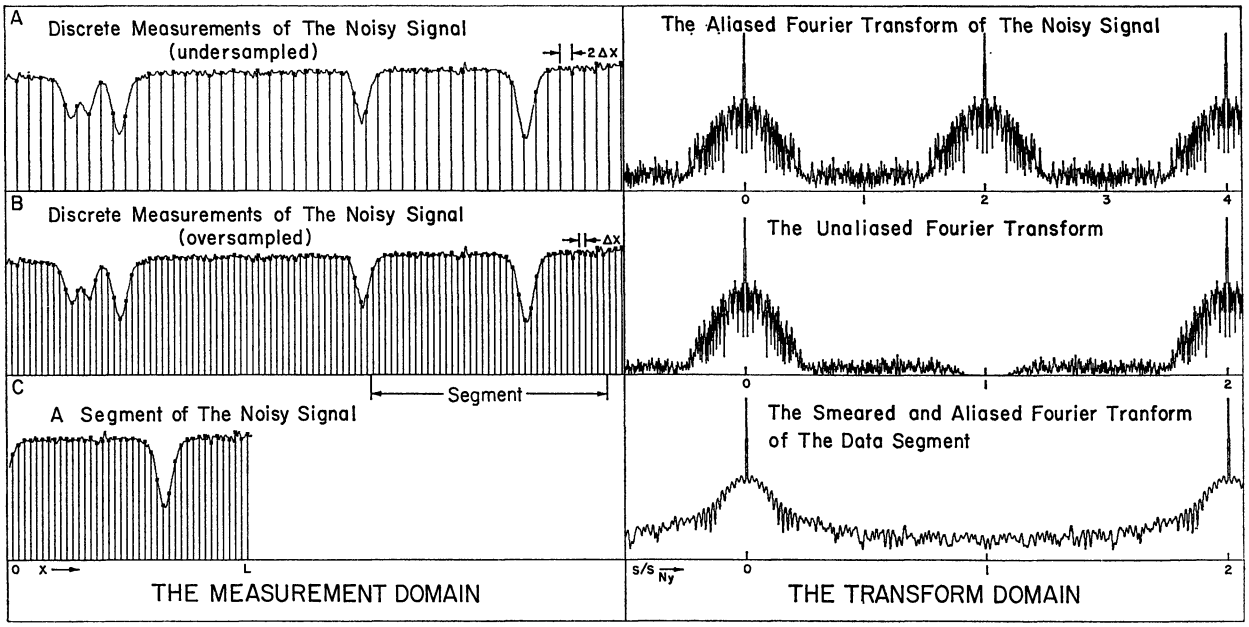

Fig. 2. Plots illustrating the effects of sampling and the finite measurement length in both domains

its entire length. Figures 1, 2 and 4 are illustrations of the signal and its Fourier transform as they appear in the successive steps necessary to go from the ideal continuous measurement to a realistic one composed of a finite number of equally-spaced values. In these three illustrations we plot the *logarithm of the power spectrum* instead of the actual values of the complex transform. This logarithmic technique shows more clearly the subtle power variations at the high frequencies.

Real physical measurements yield data whose Fourier transforms do not extend to arbitrarily high frequencies because of either the finite response time of the electrical and mechanical measuring instrument or the finite spatial resolution of the optics. We say that such data is *band-limited* because its transform is negligible outside a known, finite range or band of frequencies. In addition, the basic signal is effectively band-limited by the characteristics of the physical processes producing it. In the case of optical spectra, the *signal* band-limit is fixed by the characteristic widths of the spectral lines. In a properly designed experiment the measurement band limit should be greater than that inherent to the signal, i.e., in line profile analyses the instrument width must be less than the width of the line profiles under study. On these grounds, we will consider only basic signals that are band-limited in the *low-pass* sense where the non-zero portion of the transform lies in the frequency range from zero to some absolute cutoff frequency $|s_D|$.

Figure 1B shows a *piece* of the basic continuous signal and the logarithm of the modulus of the continuous Fourier transform of the *entire* signal. This latter function displays both the signal band-limit or cutoff at a frequency s_D and the narrow low frequency spike due to the slowly-varying continuum. The noisy appearance of the transform arises from the beating of the sinusoidal components necessary to reproduce the spacing between the spectral lines. The symmetry in the transforms is a natural characteristic of the transform of real data, and it will be discussed later. In Fig. 1C we show the effect of additive noise on the signal and the transform, which now extends to higher frequencies because of the noise contribution. Typically, the noise is *white*, i.e., its transform is essentially a constant out to high frequencies. However, as we depict it here, the measurement process contains a filter that terminates the transform at a frequency s_C beyond the band-limit of the basic signal.

2. Discrete Measurements and the Effects of Sampling

Thus far, only long, continuous signals and their transforms have been described, but now consider the case shown in Fig. 2A where the signal is sampled

at equal intervals in the x coordinate. The *continuous* Fourier transform of this set of data values is computed from the sum

$$\tilde{F}(s) = N^{-1} \sum_{j=1}^{N} F(x_j)\, e^{-i2\pi x_j s}, \qquad (2)$$

where $x_j = (j-1)\,\Delta x$, $\Delta x =$ sampling interval, $N =$ no. of data points (very large but finite), $s =$ generalized frequency variable, and $F(x) =$ signal. Our particular choice of normalization will be discussed later in Section II. *C*.

The sampling causes a replication of the basic transform of the original continuous data at intervals of $\Delta s = \Delta x^{-1}$ in the transform domain (see Bracewell, 1965, p. 192 for a rigorous proof). When the contributions from the successive replicas are added together, the result is a periodic transform as shown in Figs. 2A and 2B. In order to appreciate the cause of this fundamental periodicity, we may consider the sampled data values to be the coefficients of a Fourier series expansion of the continuous transform $\tilde{F}(s)$. Equation (2) is the required expansion, but such a series representation is possible only if the continuous function is periodic. Thus, the Fourier transform of a sequence of sample values must be periodic as illustrated in Figs. 2A and 2B.

This replication of the basic transform introduced by the sampling can lead to unreliable transforms if the replicas overlap. For example, Fig. 2A shows undersampling where the repeated basic transforms overlap to produce a distorted empirical transform. In Fig. 2B the sampling interval has been halved to show how the repeated transforms now move further apart, and become completely separated. These examples illustrate the importance of the *folding* frequency midway between the centers of the repeated transforms; this characteristic frequency due to sampling is called the Nyquist or critical frequency, and is equal to one half of the sampling rate, i.e., $s_{\mathrm{Ny}} = 0.5/\Delta x$. If the basic transform extends to frequencies greater than the Nyquist frequency, the transform computed from measurements will contain contributions due to the overlap or folding of the high-frequency amplitudes into the low-frequency band from $s = 0$ to s_{Ny}. When contributions from undersampled high-frequency components cause the computed amplitudes to differ from those of the basic transform, the Fourier transform is said to be aliased.

However, if the signal being measured is band-limited so that its transform vanishes above some finite frequency, the proper choice of the sampling interval will separate the replicas enough to prevent their mutual distortion. This implies the existence of a fundamental relation between the sampling interval and the measurement band-width in the transform domain. This relationship is known as the sampling theorem, which (following Bracewell, 1965, p. 189) states: *If a function has a Fourier transform that is zero at frequencies greater than or equal to some finite frequency s_C, this function is fully specified by values spaced at equal intervals not exceeding $0.5/s_C$.* As long as the band-width of the measurement is less than the Nyquist frequency, the sampling theorem is satisfied. If the theorem is not violated in either the sampling process or later manipulations of the data, the discrete samples will contain all of the information present in the original continuous signal, and the continuous Fourier representation of the discrete samples (computed from Eq. 2) will be identical to that of the original signal at frequencies below s_{Ny}. Noise is broadband, however, and a band-limiting analog filter must be employed prior to sampling to terminate the noise spectrum in practical measurements if aliasing by the transform of the random noise signal is to be avoided.

Thus far in our discussion, the entire signal has been sampled and recorded; however, only a short segment of the total signal can be considered in a real measurement. This practical restriction is conceptually equivalent to the extraction of the measured values from the infinitely long, sampled signal shown in Fig. 2B; and, analytically, it is accomplished by multiplying the infinitely-long signal by a rectangular window function of unit height and width equal to the length of the measurement (the rectangle function $\Pi(x)$ given by Bracewell, 1965, p. 52).

One result of this extraction process is the occurrence of the discontinuities at the beginning and end of the measurement as shown in Fig. 2C. Such discontinuities are inconsistent with the original measuring process because the measurement band-width is usually too small to allow such rapid signal changes. Unfortunately, the Fourier representation is that of the entire segment, including the discontinuities, in which case the transform of the data can extend beyond the band-limit of the measurement, i.e., the effective signal is undersampled. This is an example of a necessary step in the data-gathering process that can introduce aliasing in the computed transforms even though the sampling theorem is not violated directly in the measurement.

Because we must use measurements of finite length, the empirical transform cannot be identical to the transform of the long, basic signal, even in the

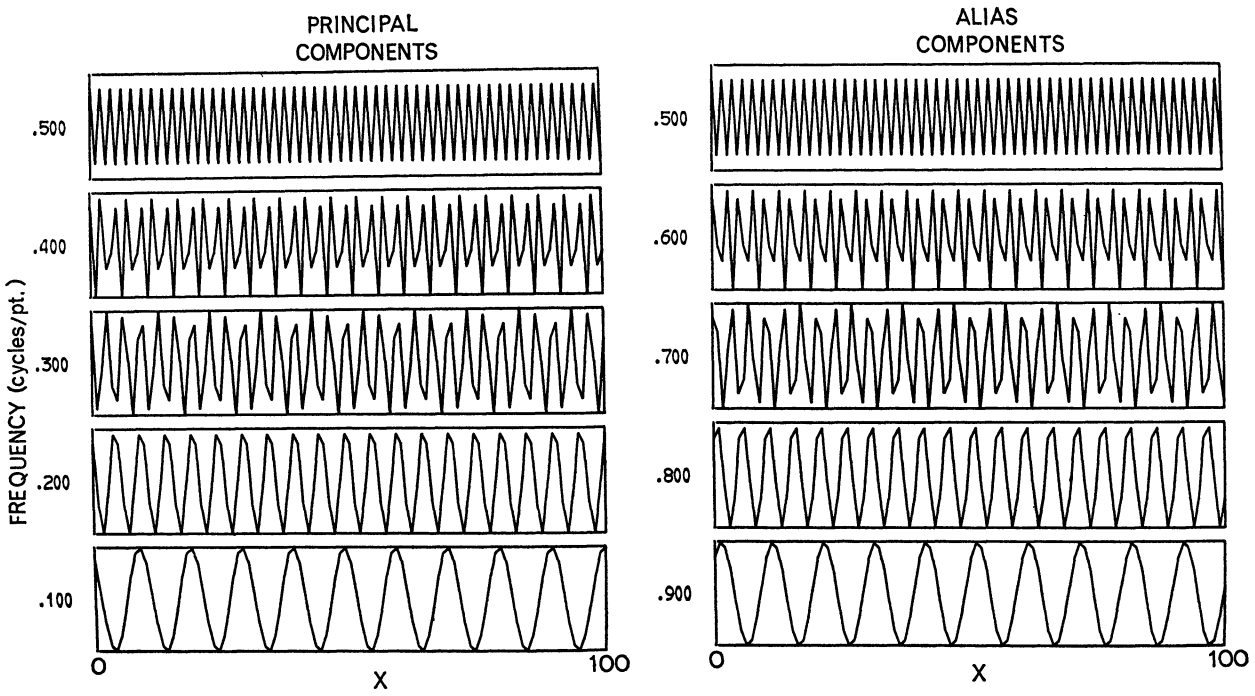

Fig. 3. An illustration of aliasing due to undersampling. The nine sinusoidal components of the artificial signal are plotted with straight lines connecting the sample values. The apparent equality between the frequency of components above and below $s_{Ny} = 0.5$ shows how undersampling makes the amplitudes of the components above s_{Ny} indistinguishable from lower frequency terms

limit of continuous data. The computed continuous transform is the result of the convolution of the basic transform with the resolution function $\sin(\pi L s)/\pi s$, where L is the measurement length. This convolution, inherent to all practical measurements, redistributes the Fourier amplitudes or, equivalently, mixes frequencies in the computed transform. This smearing effect is commonly referred to as *leakage*. The redistribution of the low-frequency spike can significantly distort the computed transform as shown in Fig. 2C where the distortion is quite evident in the frequency range from the measurement cutoff to the Nyquist frequency. Because the corruption of the transform by aliasing and leakage can affect the results of data manipulation via the Fourier transform, later we will describe procedures to minimize these effects by smoothing the sharp ends of the measured segment and by removing low-frequency components that produce the spike.

We often find that the meaning and implications of aliasing are not appreciated by researchers who want to use Fourier techniques as a tool. So, at the risk of being even more tedious, we consider the meaning of aliasing a bit further. Aliasing means that the transform value computed at a frequency s is not only composed of the true amplitude at that frequency but also of amplitudes at frequencies higher than the Nyquist frequency s_{Ny}. The Fourier representation of undersampled data is ambiguous because the true frequency content of the signal cannot be determined. In order to demonstrate this ambiguity, in Fig. 3 we show a *visual* example of how aliasing occurs through undersampling. This figure shows nine sinusoids with frequencies 0.1, 0.2, 0.3, . . . , 0.9 (cycles/unit of x) sampled at an interval of $\Delta x = 1$ and plotted with straight lines connecting the sample values. Note that in this case the Nyquist fequency is 0.5. A signal constructed by adding these nine sampled sinusoids together will be aliased because those with frequencies greater than the Nyquist frequency have an *apparent* frequency that is the *same* as one of the low-frequency sinusoids, i.e., higher-frequency components ($s > 0.5$) cannot be distinguished from true components ($s < 0.5$). The Fourier transformation of the composite signal will yield amplitudes that are too large at frequencies less than 0.5. The difficulty cannot be removed by computing the transform at frequencies greater than s_{Ny} because of the periodicity in the

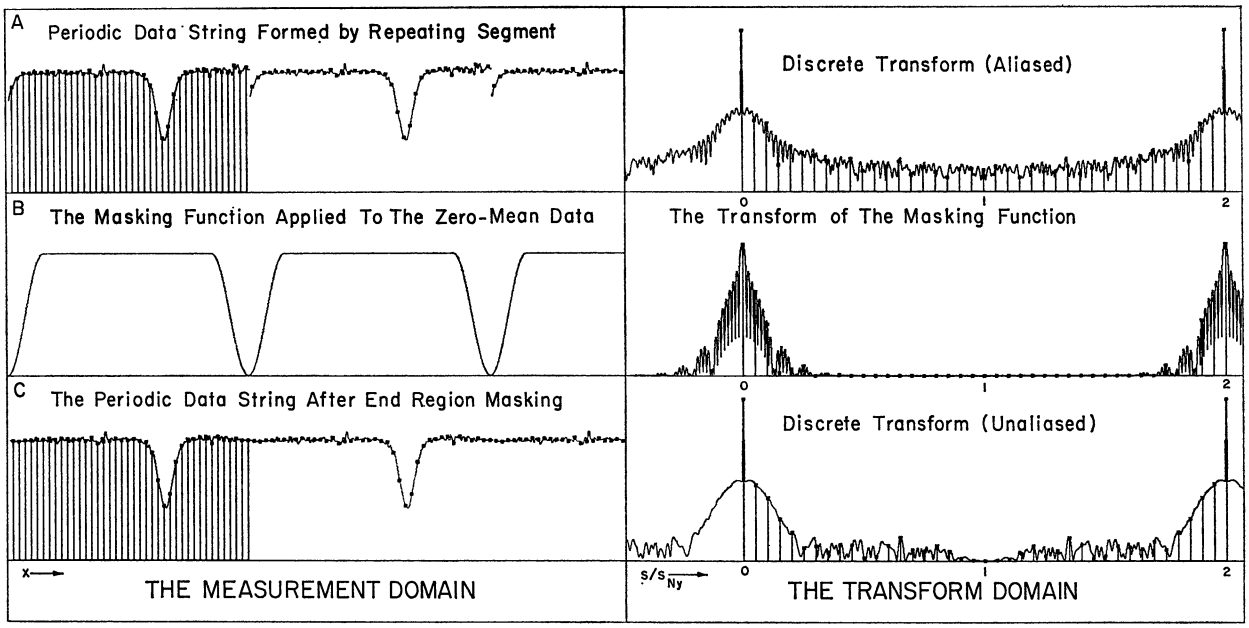

Fig. 4. Illustrations of (A) the periodicity in the effective signal required by discreteness of the transform, (B) the 15 percent data window used to mask the zero-mean data, and (C) the effects due to end-region masking of the data segment

transform. Therefore, no new information on the frequency content of a discretely-sampled signal can be obtained by computing the Fourier transform at frequencies greater than the Nyquist frequency. It is essential that the part of the transform originating in the basic, noise-free signal not be aliased; otherwise, application of the Fourier convolution, shift, and power theorems will yield neither correctly restored and shifted data nor correct power spectra.

3. The Discrete Forward and Inverse Fourier Transformations

In the previous section only a discretely-sampled signal and its continuous Fourier transform were considered; but if the digital computing technique is to be exploited fully, both the data and its Fourier transform must be discrete, i.e., we must sample in both domains. In order to suitably define the discrete forward transformation, let us sample the continuous transform given by Eq. (2) at N discrete frequencies s_k spaced at equal intervals of $\Delta s = L^{-1} = (N\Delta x)^{-1}$ in the range from $s = 0$ to $(N-1)/N\Delta x$. In this case the k^{th} frequency is given by

$$s_k = \frac{k-1}{N\Delta x}, \; k = 1, 2, 3, \ldots N . \qquad (3)$$

Equation (2) evaluated at these particular frequencies defines the discrete forward Fourier transformation of N data values. This choice of frequency spacing is optimum in the sense that the transform values are spaced one resolution width ($(N\Delta x)^{-1}$) apart.

The discrete inverse Fourier transformation can now be defined as

$$F(x_j) = \sum_{k=1}^{N} \tilde{F}(s_k)\, e^{+i2\pi x_j s_k} \qquad (4)$$

where the normalization is chosen such that a forward transform (Eq. 2) followed by an inverse transform (Eq. 4) returns the original function unchanged.

In analogy with the previous case of periodic transforms produced by sampled data, the sampling of the continuous transform requires that the observed signal segment must be thought of as one period of a periodic function, where the segment is repeated indefinitely along the measurement axis. Because of this periodicity as illustrated in Fig. 4A, the concepts of aliasing and leakage are also applicable in the measurement domain. Undersampling in frequency can produce aliased or overlapped data segments, but this aliasing in the measurement domain (often called *aliasing in time*) is usually not a problem in our applications since the sampling interval in frequency is fixed by the number of data points. Much more significant is the leakage or *wrap around* from one segment to the other caused by convolution in the measurement domain (e.g. as in filtering or inter-

polating). The wrap-around problem can be reduced by extending the data sequence with values of the mean thereby increasing the separation between the repeated segments.

Since the result of the discrete forward transform of N real data values is a set of N complex numbers, there must be a two-fold redundancy in the computed set of transform values. The transform of real data has conjugate symmetry about both the origin and the Nyquist frequency, i.e., $\tilde{F}(s) = \tilde{F}(-s)*$ and $\tilde{F}(s_{\mathrm{Ny}} + \Delta s) = \tilde{F}(s_{\mathrm{Ny}} - \Delta s)*$. Figure 5 illustrates this symmetry by showing the amplitudes of the real and imaginary parts of a real-data transform plotted from 0 to $2\,s_{\mathrm{Ny}}$ in frequency. Clearly, no new information is contained in the values above the Nyquist frequency. Since we compute a *one-sided* transform in the frequency range from 0 to $2\,s_{\mathrm{Ny}}$, which is just one full cycle of the total periodic transform, the values above the Nyquist frequency are the coefficients of the negative frequency terms in the normal *two-sided* case. As a further result of the conjugate symmetry, the imaginary parts of the transform at the origin and Nyquist frequency are identically zero. Thus, the discrete forward transform yields N non-zero, independent Fourier coefficients of which $N/2 + 1$ apply to the real terms and $N/2 - 1$ apply to the imaginary ones.

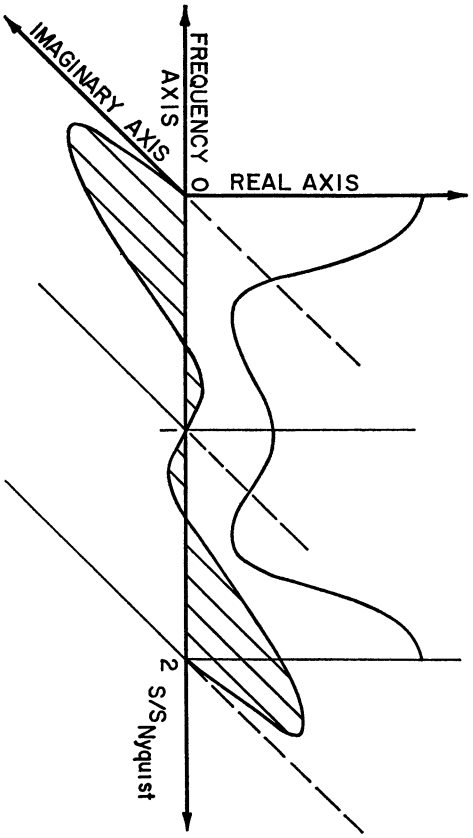

Fig. 5. A sketch of the real and imaginary components that illustrates the conjugate symmetry of the one-sided Fourier transform of real, asymmetric data

4. Low Frequency Problems

Although we have shown the data on its true intensity scale in Fig. 4, in practice, the mean of the measured values is subtracted prior to any Fourier transformations and modifications of the data. The mean is essentially an invariant under the types of filtering we will consider; therefore, the true intensity scale is recovered by simply adding the mean at the end of the computation. The consistent use of data with a zero mean simplifies the end-point problems described in the next section.

Removal of the sample mean from the data only affects the one transform value at zero frequency, but other very low-frequency effects such as slow signal drifts and real continuum variations in a spectrum can produce a low-frequency spike in the transform of the data. The removal of such spurious *trends* is necessary to minimize the effect of the redistribution of the spike by the resolution function. Often, low-order polynomials obtained from least squares fitting of the data are used to remove trends from the data prior to Fourier analysis.

5. The Special Problem at the Ends of Data Strings

As mentioned earlier, the discontinuities at the ends of the measured segment introduce a rapid signal change that is not consistent with the bandwidth of the measurement. The difficulties at the end of the data string become apparent as oscillations or Gibbs phenomena if attempts are made to interpolate or filter the signal. In order to lessen the effects of these discontinuities, we force the end regions to make a smooth transition to the mean of the measured values by multiplying the sequence of data values with a set of weights called a *data window*. A data window should only affect the ends of the segment where the weights effectively mask the actual data values. The *application of a data window*, and *end region masking* are different names for the same operation, and they will be used interchangeably throughout the remainder of the paper. In Fig. 4B we show several periods of a data window that modifies only the first and last 15 percent of the data

segment. Figure 4C shows the smoothing that results from its application to the periodic data string in Fig. 4A. Such a set of weights apodizes the resolution function in the frequency domain, and thus reduces the leakage due to the wings of the function. Note that a data window should only be applied to data with a zero mean; otherwise, the smoothed step at the ends of the segment will be larger than is either necessary or desirable.

C. Summary

We have described a linear transformation, the discrete Fourier transform, that converts N discrete measurements into N complex numbers by means of the defining equation

$$\tilde{F}(s_k) \equiv N^{-1} \sum_{j=1}^{N} F(x_j)\, e^{-i2\pi x_j s_k} \qquad (5)$$

where $x_j = (j-1)\,\Delta x$ for $j = 1, 2, \ldots N$ and $s_k = (k-1)/N\Delta x$ for $k = 1, 2, \ldots N$. The discrete forward Fourier transform for real data has conjugate symmetry about the Nyquist frequency, so that only the first $N/2 + 1$ complex values need to be considered. The normalization of the forward transform is such that the computed amplitudes are on the same scale as the original measurements, except that one must include the transform values at negative frequencies to determine the total contribution to the transform at some absolute frequency $|s_k|$. The corresponding inverse transform is

$$F(x_j) \equiv \sum_{k=1}^{N} \tilde{F}(s_k)\, e^{+i2\pi x_j s_k}. \qquad (6)$$

The inverse transform is normalized in a manner to make the transform pair cyclical. These definitions of the discrete Fourier transform pair are general enough to allow the input sequence $F(x_j)$ to be complex.

If the sampling theorem is not violated, the intermediate values of both the continuous data function and the continuous complex transform can be recovered by convoluting the discrete values with the proper interpolation function. The interpolation formulae are as follows.

Measurement Domain Interpolation:

$$F(x) = N^{-1} \sum_{J=1}^{N} F(x_j) \frac{\sin[\pi(x - x_j)/\Delta x]}{\tan[\pi(x - x_j)/N\Delta x]} \qquad (7)$$

Transform Domain Interpolation:

$$\tilde{F}(s) = N^{-1} \sum_{J=1}^{N} \tilde{F}(s_j) \frac{\sin[\pi(s - s_j) N\Delta x]}{\tan[\pi(s - s_j)\Delta x]} \qquad (8)$$

The following procedure will help the user avoid common problems in computing reliable transforms. First, the sampling rate must be such that the basic signal is oversampled to prevent aliasing by the replicas of the basic transform ($1/\Delta x \gtrless 2\, s_c$). Second, the sampling instrument must contain a real-time filter with a cutoff frequency no greater than one half of the sample rate if aliasing by noise is to be avoided. Third, the low frequency spike should be minimized by subtracting the sample mean and, if necessary, by removing trends from the data. Fourth, the data values must be multiplied by a data window which will reduce leakage effects due to the finite measurement length. Finally, the modified data sequence can be artificially extended with zeroes to reduce leakage between the repeated data segments.

III. Calculation of Fourier Transform

A. The FFT Algorithm

The computational methods now used to evaluate the discrete transform given by Eqs. (5) and (6) are described by Cochran *et al.* (1967) and Cooley *et al.* (1970a). In earlier papers Cooley and Tukey (1965), and Gentleman and Sande (1966) describe the basic algorithm. White and Cha (1967) give a detailed description of one version of the algorithm as used by Forman (1966). The literature on the FFT is now extensive but thoroughly listed in IEEE Audio Transactions, Special Issue on the Fast Fourier Transform (1967), Bergland (1969) and Cooley *et al.* (1970a). There is no need to give the algorithm in algebraic detail since this task is already well done in the existing literature, but the general ideas underlying it are presented below.

The basic algorithm applies when the number of points in the sample is an integral power of 2; however, it has been generalized to the case where N is a composite integer. The algorithm converts the serial product in Eq. (5) into a set of nested sums whose evaluation requires fewer additions and multiplications than the straightforward evaluation of the serial product. This reduction in addition and multiplication is the principal mechanism by which the computation time is decreased. For example, the direct method requires on the order of N^2 multiplications in contrast to $2N \log_2 N$ in the FFT. In addition the FFT results are more accurate because the reduction in the number of arithmetic operations also reduces the cumulative round-off errors. The FFT programs eliminate redundancy in calculation of sines and cosines, i.e., only $N/4 + 1$ sine and cosine

values are required. All of these time-saving devices combine to give speed increases of 50 to 100 times when the FFT is applied to typical solar problems.

B. Comments on Available FFT Routines

We use FFT routines written by M. L. Forman of AFCRL (Forman, 1966), Norman Brenner of MIT (Brenner, 1967, 1968), G. D. Bergland of Bell Labs (Bergland, 1968) and one of us (JWB). The Bell Lab subroutines are designed for one-dimensional, real data; whereas, the more-general MIT programs can handle multidimensional, complex input data. Using the NCAR CDC 6600 computer the MIT FOUR2 routine requires 0.1 seconds for calculation of the transform of 1024 points as compared to 0.07 seconds for Bergland's FFA program. Many new FFT routines are now available such as those described by Singleton (1968, 1969).

The faster routines usually require that the number of data values be an integral power of two ($N = 2^n$). In cases where the number of measured points is not the desired power of two, we can use the artifice of extending the data sequence with values of the mean. The immediate effect of such extensions is to decrease the frequency sampling interval to less than the optimum spacing of $\Delta s = (N\Delta x)^{-1}$, i.e., the transform of extended data will be computed at a higher frequency resolution than is justified by the length of the measurement. The extension of a data sequence also effects the magnitude of the transform since the normalizing constant N in the forward transform (Eq. 5) is the total number of points used in the calculation and not only the number of measurements. In order to preserve the scaling for different extensions of data with a zero mean, the forward transform must be multiplied by N/M, where M is the number of measurements, and $N-M$ is the number of appended zeroes.

Users of the available subroutines must know the normalization and the coordinate system of the computed transforms. For example, the MIT and Bell Lab programs contain no normalization; therefore, recovery of the data after a forward → inverse transform sequence requires multiplication by a scaling factor (N^{-1}). The coordinate system for these routines is such that both the data and the transform are one sided, i.e., the data extends from $x = 0$ to $N\Delta x$, and the transform from $s = 0$ to $2\,s_{\mathrm{Ny}}$. This lack of symmetry can cause confusion when one is first checking the FFT routines by computing the transform of a simple analytic function for which the transform values are known. For example, consider a simple gaussian exponential with the center of symmetry placed midway in the data string, $x_0 = N\Delta x/2$. The Fourier transform of a gaussian exponential is another gaussian, in which case we may naively expect the computed transform to show the exponential in the real part of the transform and zeroes for all of the imaginary amplitudes. Instead, the real amplitudes alternate in algebraic sign as we go from point to point along the frequency axis. This wildly fluctuating function is simply the desired gaussian transform multiplied by the shift theorem factor $\exp(-i2\pi x_0 s_j)$, which in this case is simply $(-1)^{j-1}$ since $s_j = (j-1)/N\Delta x$. Such fluctuations due to the shift of the input data are easily removed by applying the opposite shift factor $\exp(i2\pi x_0 s_j)$ to the complex data transform. In applying the shift and derivative theorems in data manipulation, the necessary complex operators (such as $\exp(-i2\pi x_0 s_j)$) must have conjugate symmetry about the Nyquist frequency, i.e., the values of the operators in the upper frequency range $s_{\mathrm{Ny}} < s_j < 2s_{\mathrm{Ny}}$ must be computed as though they apply at the negative frequencies $s_j' = s_j - 2s_{\mathrm{Ny}}$.

IV. Applications of the FFT to the Spectral Analysis and Restoration Problems

A. Classes of Measurement

Astronomical measurements amenable to analysis by Fourier techniques fall into two categories: 1) data that display both correlation and randomness such as the solar oscillations displayed by the Doppler shifts of spectral lines, intensity fluctuations with time, and intensity variations due to changes in fine structure across the solar disk and 2) deterministic data that appear to be repeatable such as scans of astronomical spectra, interferograms, single photographs of solar fine structure, and scans of the solar limb. In category 1 we wish to estimate the power spectra of the solar process that causes the intensity and velocity oscillations; whereas, in category 2 the measurements require correction for linear, *time* invariant effects that can be represented by a convolution integral.

In the following two sections, FFT techniques for computing power spectra and for restoring measurements will be discussed. The utility of the method will be stressed in order to convince observers that they can practically, and with some confidence, reduce instrument effects and noise in their measurements in an optimum way and thus recover as much

of the stellar information as possible. The explanatory figures to be shown are examples of solar measurements made by the authors at Kitt Peak National Observatory and Sacramento Peak Observatory.

B. Calculation of Power Spectra

1. The Unmodified Power Spectrum and Its Use

As we shall demonstrate, the raw power spectrum of a measurement is essential for experiment diagnostics and the design of noise-suppression filters. The raw power spectrum is immediately computed from the discrete Fourier transform of the measurement by the usual relation

$$P(s_k) = \tilde{F}(s_k)\,\tilde{F}(s_k)\star = \tilde{F}(s_k)^2_{\text{Real}} + \tilde{F}(s_k)^2_{\text{Imag}} \,. \quad (9)$$

This direct calculation is the alternative to the common autocovariance approach, but the two methods are equivalent as can be shown from the power theorem (Parseval's theorem). However, the direct calculation using the FFT and Eq. (9) can be performed much faster for long data sequences.

Besides the usual self-convolution method, the *mean lagged product*

$$C(\chi_n) = N^{-1} \sum_{k=0}^{N-|n|} F(x_k)\,F(x_k + |\chi_n|)\,, \quad (10)$$

where

$$\chi_n = n\,\Delta x \text{ for } n = 0, 1, 2, \ldots N-1\,,$$

is easily computed by taking the inverse transform of the raw power spectrum $P(s)$, provided certain precautions are taken. One must remember that the entire mean lagged product for an N point measurement is symmetric about $\chi_0 = 0$, and is composed of $2N-1$ values. In contrast, the symmetric power spectrum computed by applying the FFT to the same data will have only N values. Before an unaliased $C(\chi_n)$ can be computed from the N point power spectrum, the length of the initial data sequence must be at least doubled by extending it with values of the sample mean. It is best to anticipate the need for a correlation function and allow for its calculation by routinely extending the set of sample values at the beginning of the reduction. Figure (6) illustrates the two alternate routes for calculation of the raw power spectrum and the corresponding mean lagged product.

The two empirical functions, $P(s)$ and $C(\chi)$, described above are related to the periodogram, autocorrelation, and autocovariance function as follows.

Periodogram $= N\,P(s_k)$

Autocorrelation $= C(\chi_n)/C(o)$

Autocovariance $= C(\chi_n)$ when sample mean is zero.

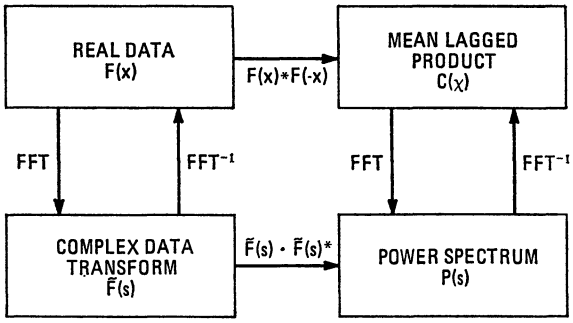

Fig. 6. A flow diagram showing the two routes for computing power spectra and mean lagged products

Note that the calculation yields the *biased* form of the correlation function. If the *unbiased* correlation,

$$C(\chi_n) = \frac{1}{N-|k|} \sum_{k=0}^{N-|n|} F(x_k)\,F(x_k - |\chi_n|)\,, \quad (11)$$

is required; the weighting by the factor $(N-|k|)^{-1}$ must be performed separately. This weighting is a correction for the finite length of the data sequence, but it does increase the contribution of the most unreliable values of the lagged product, i.e., those for large lags. For comments on the use of biased and unbiased autocovariance functions see Jenkins and Watts (1968, p. 174).

The raw power spectrum is a very useful diagnostic tool because it reveals defects in the data that can be difficult to detect and evaluate by direct examination of the measurements. For example, the power spectrum of a measurement will show any spurious periodic interference such as the 30 Hz frame rate of TV monitors located near photometers, image motion due to motor and compressor vibrations, etc. Low frequency problems due to electronic drift and sky transparency changes will appear as an increased "1/*f*" noise spectrum. Infrequent random spikes or data "glitches" will raise the high frequency end of the empirical spectrum. Even more important, such preliminary spectral analysis immediately shows the empirical relationship between the signal power and noise power, and thus, allows the experimenter to locate the frequency band in which significant signal information is carried. The knowledge of the usable signal spectrum then permits specification of a numerical filter to preferentially suppress the noise. In practice, use of a data window will give better power spectra for diagnostic purposes. This technique will be discussed in the next section under *modified* power spectra.

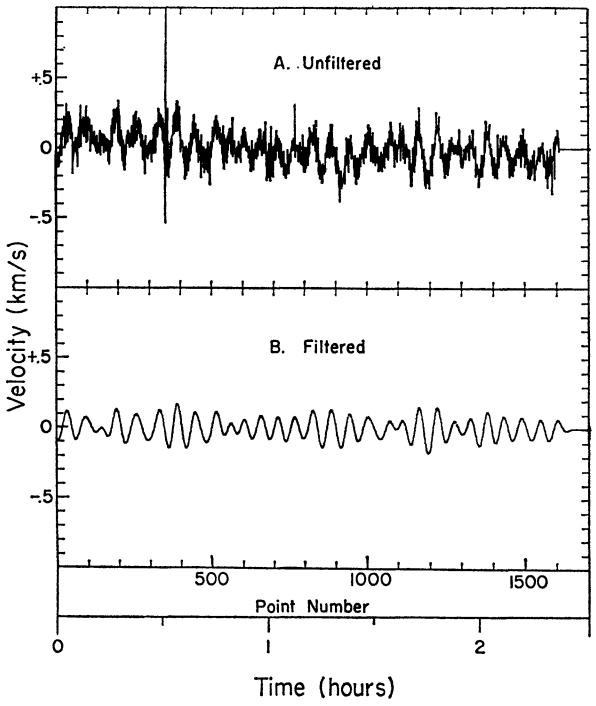

Fig. 7. A measurement of the photospheric velocity oscillations displayed by Doppler shifts of the solar Fe I line at 6290.974 Å: (A) raw data and (B) filtered data. The observations were made with a fast spectrum scanner at Sacramento Observatory on April 29, 1969

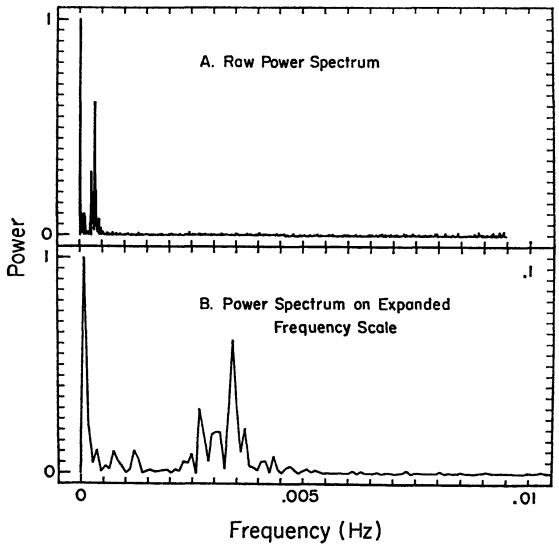

Fig. 8. The raw power spectrum of the unfiltered, 2.5 hour sample of solar photospheric oscillations shown in Fig. 7A: (A) the power spectrum from 0 to 0.1 Hz, and (B) the low frequency end of the same power spectrum from 0 to 0.01 Hz

Consider the measurements of the quasi-periodic solar signal shown in Fig. 7A. The raw power spectrum of this measurement is shown in Fig. 8. From this empirical spectrum we see that the predominant oscillations, those well removed from the origin, lie in the frequency range from 0.0015 to 0.0045 Hz. A band-pass filter with a cosine-bell shape in the transform domain was chosen to preserve this band but to totally eliminate all amplitudes at frequencies above 0.0066 Hz and to strongly attenuate those below 0.001 Hz. The result of this filtering is shown in Fig. 7B where the modulated or wave-packet character off the 300 second solar oscillations now shows very clearly because the high frequency noise and the low frequency drifts (or perhaps low frequency solar oscillations) have been attenuated by the numerical filtering.

Such filters are conceptually easier to specify and apply in the transform domain than in the measurement domain where they can be rather complicated functions to be convolved with the data. Once the desired frequency response is determined from the empirical power spectrum, the transform of the measurement is multiplied by this response function; and the filtered data is obtained immediately from the inverse transformation of the data transform as modified by the filter response function.

Band-pass filters such as the one used to obtain the results in Fig. 7B must be used carefully, because if the pass band is too narrow, the smoothed data may not represent a real oscillation originating in the star under study. To see how this problem may arise, consider the application of a very narrow filter to a pure noise signal. The filter will certainly smooth the data and leave an oscillatory signal, but this smooth wave-form is only one Fourier component of a broad-band signal that has no principal resonance.

Much past effort has been devoted to *physically realizable* filters which can be duplicated electronically. The time or impulse response of such filters must be asymmetric since future values of the signal are not known. However, numerical filtering of recorded data is not restricted to the *physically realizable* case, and the impulse response can be symmetric. By specifying that the filter's amplitude response function be purely real in the transform domain, we insure that the smearing or impulse response function will be symmetric in the measurement domain. The use of such symmetric filters will not skew the measurements.

2. Smoothed and Modified Power Spectra

The raw power spectrum is indeed the correct spectrum for a particular measurement, but it may be a very poor estimate of the average power spectrum underlying many such measurements, particularly when the basic signal displays random amplitude and phase changes with time. This statistical uncertainty in the power values computed from a strictly random signal is quantitatively expressed by the relation

$$\varepsilon^2 \approx (L\, W_s)^{-1}\,, \tag{12}$$

where ε is the fractional variance of the power, L is the length of the measured signal, and W_s is the approximate width of the spectral resolution function (see Blackman and Tukey, 1958, p. 21, Richards, 1967, p. 87). This approximate relation states that the RMS uncertainty in the power values computed from Eq. (9) is equal to the mean value of the power, i.e., for an unmodified power spectrum $\varepsilon^2 = 1$; whenever, $W_s = L^{-1}$. If a more precise power estimate is to be made from a given measurement, the resolution width W_s must increase. One must realize, however, that the discouraging statistical uncertainty given by Eq. (12) is derived for purely random signals described by gaussian statistics and a flat, broadband spectrum.

A related problem is that of spectral purity in the computed power spectrum. The basic resolution function has wings that decay as s^{-2}; and since the power spectrum is the result of convolution of this slowly-decaying function with the true power spectrum, there is frequency mixing. Recent work with the FFT emphasizes the data window method for reducing this effect through apodization of the resolution function prior to spectral smoothing. In the following paragraphs we shall discuss these two aspects of power spectra calculation: 1) the use of data windows and 2) three methods of spectral smoothing. According to current usage, a *modified* power spectrum is one computed directly from a data sequence that has been multiplied with a data window function, and a *smoothed* power spectrum is obtained by filtering the spectrum itself to improve the *statistical stability* of the power estimate.

(a) Data Windows and Modified Power Spectra.

In Section II.*B*.5 we discussed the inconsistency that arises because the effective signal has discontinuities at the ends of the measurement. The remedy was to apply a set of weights to the data sequence in order to smooth the discontinuities before taking the transform. The use of such a data window is to be considered as a necessary part of any reduction scheme that requires reliable transforms and power spectra.

The choice of the window function is somewhat arbitrary, but Bingham *et al.* (1967) suggest use of a cosine bell over the first and last 10 percent of the data string. In the more general case where the cosine bell is applied over short sections fL units wide at both ends of the measurement, the weights for the cosine bell are computed from the function

$$W(x) = \begin{cases} \frac{1}{2}\,[1 - \cos(\pi x/fL)] & 0 \leqq x \leqq fL \\ 1 & fL \leqq x \leqq L(1-f) \\ \frac{1}{2}\,[1 - \cos(\pi(L-x)/fL)] & \\ & L(1-f) \leqq x < L\,, \end{cases} \tag{13}$$

where L is the length of the data string and f is the fraction of the length over which the data is masked by the cosine bell.

The effect of such data modification on the transform can be seen from the transform of this continuous window function, which is

$$\tilde{W}(s) = \frac{\sin(\pi L s(1-f))}{\pi L s}\,\frac{\cos(\pi f L s)}{1-(2fLs)^2}\,. \tag{14}$$

The sine term in this transform is very similar to the basic resolution function, and it fixes the width of the central lobe; whereas, the cosine factor apodizes the wings of the sine term. For example, notice that for $2fLs \gg 1$, the transform decays as s^{-3} compared to s^{-1} for an unmodified data set. As a rule of thumb, the width of the central lobe will be $(1-f)^{-1}$ in units of the dimensionless frequency Ls. Similarly for the cosine term, the frequency at which apodization becomes effective is f^{-1} or, conservatively, $2f^{-1}$. Using these approximate rules, we see that a data window designed to affect only the first and the last 10 percent of the data sequence produces a spectral resolution function with a central lobe 1.1 resolution units wide and an apodization starting 10 to 20 resolution units from the central maximum.

If we apply such a data window to a measurement, and then compute the modified power spectrum; the result is a high resolution spectrum with less leakage than would be obtained with an unmodified data set. As Bingham *et al.* (1967) point out, the power spectra of *deterministic* signals relatively free of random effects may be adequately estimated by modified power spectra alone; but when the measurement is *noise-like*, and the uncertainty in the modified power estimates is intolerable, we must also smooth the modified power spectrum.

(b) Smoothed Power Spectra.

The three methods for spectral smoothing are the lag window technique, smoothing by direct convolution in the frequency domain, and segmental averaging. Since Cooley *et al.* (1970b) discuss all three methods and show examples, we will not make an exhaustive analysis but only summarize the important characteristics of the methods.

(1) The Lag Window Method

The standard lag window method described by Blackman and Tukey (1958) was the most efficient way to compute power spectra prior to the advent of the FFT. In this approach, the empirical autocovariance function is computed directly from the data sequence and then smoothly truncated by multiplying it with a set of weights called a lag window. The power spectrum is then obtained by taking the cosine transform of this modified autocovariance function. Since this multiplicative process in the measurement domain is equivalent to a convolution of the power spectrum with the transform of the lag window, it results in a smoothing of the power spectrum. The degree of smoothing varies inversely with the maximum lag or truncation point of the autocovariance: $L_{\max} = W_s^{-1}$. The Blackman-Tukey recipe (Blackman and Tukey, 1958, p. 21) calls for maximum lags of ~ 10 percent of the total measurement length. This 10 percent lag specification comes from their requirement that the RMS uncertainty ε in the power computed for a random signal be $\sim 1/3$ of the average power, i.e., $\varepsilon = 1/3$ and $W_s = L_{\max}^{-1}$ in Eq. (12). The smoothed power spectrum computed from such a recipe will have a resolution width 10 times that of the raw periodogram, and the statistical stability of the smoothed power estimate will be increased 3 times for a pure *white noise* signal.

In the absence of the FFT, the computational advantage of the standard lag window method for small maximum lags is clear: the number of values involved in the cosine transform of the autocovariance function is much less than the number of data values. However, the method suffers from the drawback that the spectral resolution functions corresponding to convenient lag windows have slowly decaying wings. In some cases, the resolution function can have appreciable side lobes and may take on negative values. Thus, the use of simple lag windows can lead to spurious peaks in the spectrum and even negative power values because of leakage effects in the frequency domain. Figure 9 illustrates the steps in the application of the FFT to the lag window method.

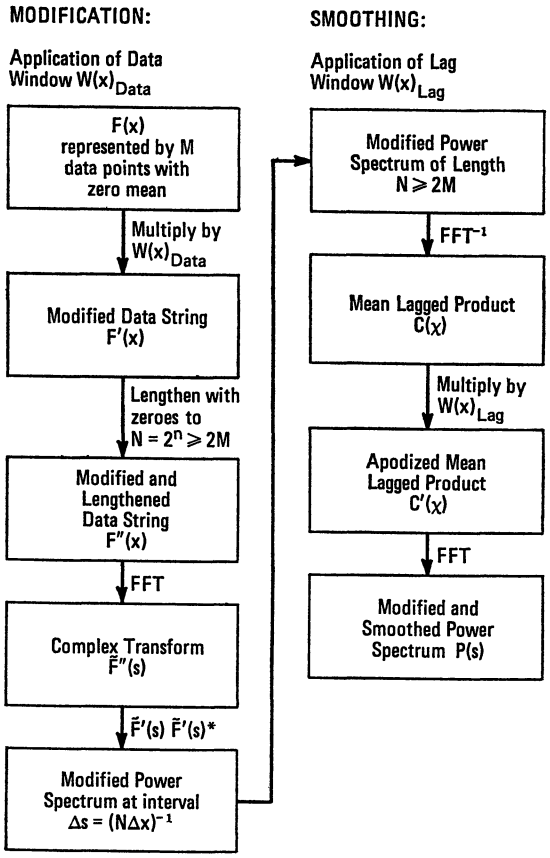

Fig. 9. A flow diagram of the combined modification and lag window smoothing process using the FFT

(2) Smoothing by Direct Convolution

Since the FFT allows us to economically compute the transform and the raw power spectrum for long data sequences, we can bypass the calculation of autocovariance functions and consider direct convolution of the modified periodogram as the smoothing device. In this approach we can specify the smoothing function to suit our needs, and thus avoid leakage inherent to the lag window method. The smoothing function is represented by a set of weights whose sum must be unity in order to preserve the area under the power spectrum being smoothed.

As an example of the convolution smoothing, we can take the simple running mean over k points in the modified power spectrum. Such an average over 11 points in the spectrum is statistically equivalent to use of a 10 percent lag window although the leakage

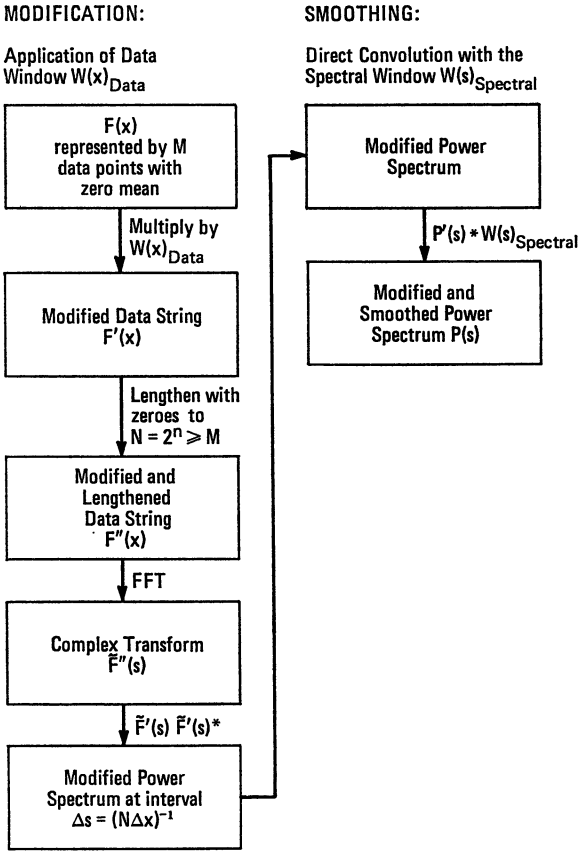

Fig. 10. A flow diagram of the combined modification and direct smoothing method

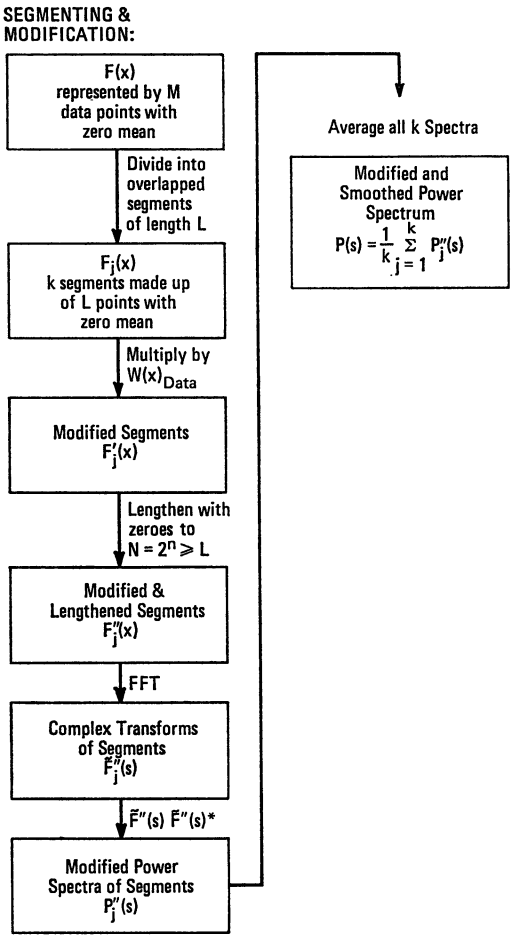

Fig. 11. A flow diagram of the segmental averaging technique for smoothing power spectra

characteristics are quite different. The lag window and direct smoothing techniques will give identical results if the weights for the lag window are the discrete inverse transform of the spectral weights that define the smoothing function in the frequency domain. Figure 10 illustrates the sequence of operations in the direct smoothing method.

(3) Segmental Averaging

Spectral smoothing can also be performed by dividing a long time series into several segments of equal length, computing the power spectra for each segment, and then averaging the set of spectra. This technique is equivalent to use of an appropriate lag window (see Richards, 1967, p. 86 and Welch, 1967, p. 70). Spectral stability has been gained at the expense of spectral resolution because the effective spectral resolution width is now fixed by the short segment lengths and not the longer duration of the original time series.

Cooley *et al.* (1970b) recommend that the segments of the data sequence overlap by one-half of the desired segment length. Such overlapping introduces some correlation between the segments, but there is a net gain in the statistical stability because the number of spectra averaged together is doubled relative to that without overlap. The spectral resolution function for this method has the same general shape as in the lag window case: a strong central lobe whose width is determined by the segment lengths and long wings whose rate of decay is determined by the shape of the data window applied to each segment. These wings tend to have a small amplitude lobe structure, but they cannot take on negative values as is possible for certain lag windows. Figure 11 shows the various steps in computing the smoothed power spectrum by the segmental averaging method.

(4) Comparison of the Three Methods for Spectral Smoothing

Cooley *et al.* (1970b) show that for a *gaussian white noise* signal the three smoothing techniques are statistically equivalent when their smearing widths are the same in the frequency domain. As guides for this equivalence, the approximate smearing widths to be compared for the lag window, direct smoothing, and segmental averaging methods are the reciprocal of the maximum lag ($W_s \approx (L_{\max})^{-1}$), the width of the smoothing function at the half power points, and the reciprocal of the segment length ($W_s \approx (L_{\text{seg}})^{-1}$), respectively.

More significant differences lie in the shapes of their spectral resolution functions, since the shape determines the degree of leakage and frequency range over which it occurs. In the direct smoothing case, the smoothing function can be chosen to give a more uniform frequency mixing over a narrower frequency band than for the other two methods, but the shape of any narrow *lines* in the power spectrum will also be distorted more by concentrated, flat-topped functions such as the running mean.

The more obvious differences are practical ones. Direct smoothing and segmental averaging are computationally more efficient for long data sequences because the FFT can be used. Segmental averaging has the added advantage that the spectra of successive segments can be compared to give an indication of the stationarity of the measurement with time, i.e., is the underlying power spectrum obviously changing during the measurement process?

The degree of smoothing required in any particular case still remains unspecified because it depends so strongly on the statistical character of the measured signal. This uncertainty leaves the theory of empirical spectral analysis in a rather unsatisfactory state, but we are not prepared to discuss the problem of signal classification. However, the use of a data window to reduce leakage plus the option of using either the lag window, direct smoothing, or segmental averaging method for spectral smoothing gives a set of flexible, defensible tools for computing power spectra.

C. Restoration of Smeared Data

In the following discussion, we consider in some algebraic detail the optimum restoration process. Since optimum filtering and restoration are not as well known nor widely-discussed as Fourier analysis and power spectra calculation, we feel that a more mathematical discussion is necessary for the reader to appreciate the basis for this technique.

1. Optimum Restoration

Many physical observations are degraded by smearing effects that can be represented by a convolution integral

$$S(x) = \int_{-\infty}^{\infty} T(y)\, A(x-y)\, dy \equiv T(x) * A(x)\,, \quad (15)$$

where $T(y)$ is the true signal, $S(x)$ the smeared signal as observed in the absence of noise, and $A(x)$ is the smearing function or apparatus response function. The smearing of spectral lines by instruments with finite resolution and the two-dimensional smearing of solar fine structure by atmospheric turbulence and telescopes of finite aperture are perhaps the most common examples of the problem. What we now seek is the best possible means of correcting for such effects.

We begin by noting that in the transform domain, the convolution becomes a simple product:

$$\tilde{S}(s) = \tilde{T}(s) \cdot \tilde{A}(s)\,. \quad (16)$$

These relations are illustrated in Fig. 12A, which shows the basic convolution process as seen in both the measurement and transform domains. The example is that of a single noise-free spectral line smeared by a typical apparatus function. In the transform domain, the restoration is trivial and exact as long as $\tilde{A}(s)^{-1}$ exists for all frequencies:

$$\tilde{R}(s) = \tilde{S}(s) \cdot \tilde{A}(s)^{-1} = \tilde{T}(s)\,. \quad (17)$$

This simple restoration process is shown in Fig. 12B.

In the presence of the inevitable noise associated with all measurements, however, the simple solution indicated by Eq. (17) proves to be disastrous, as demonstrated in the next figure. Figure 13A shows the addition of noise to the smeared profile and its effect on both the observed profile and its transform.

In this case the observed distribution is

$$O(x) = S(x) + N(x)\,, \quad (18)$$

where $N(s)$ is additive random noise. The result of restoration in the transform domain is then

$$\tilde{R}_0(s) = \tilde{O}(s)/\tilde{A}(s) = \tilde{T}(s) + \tilde{N}(s)/\tilde{A}(s)\,, \quad (19)$$

which produces a very noisy restored profile $R_0(s)$. The poor results obtained by this naive approach are thus traced to the strong amplification of high frequency noise components during this inverse operation as illustrated in Fig. 13B.

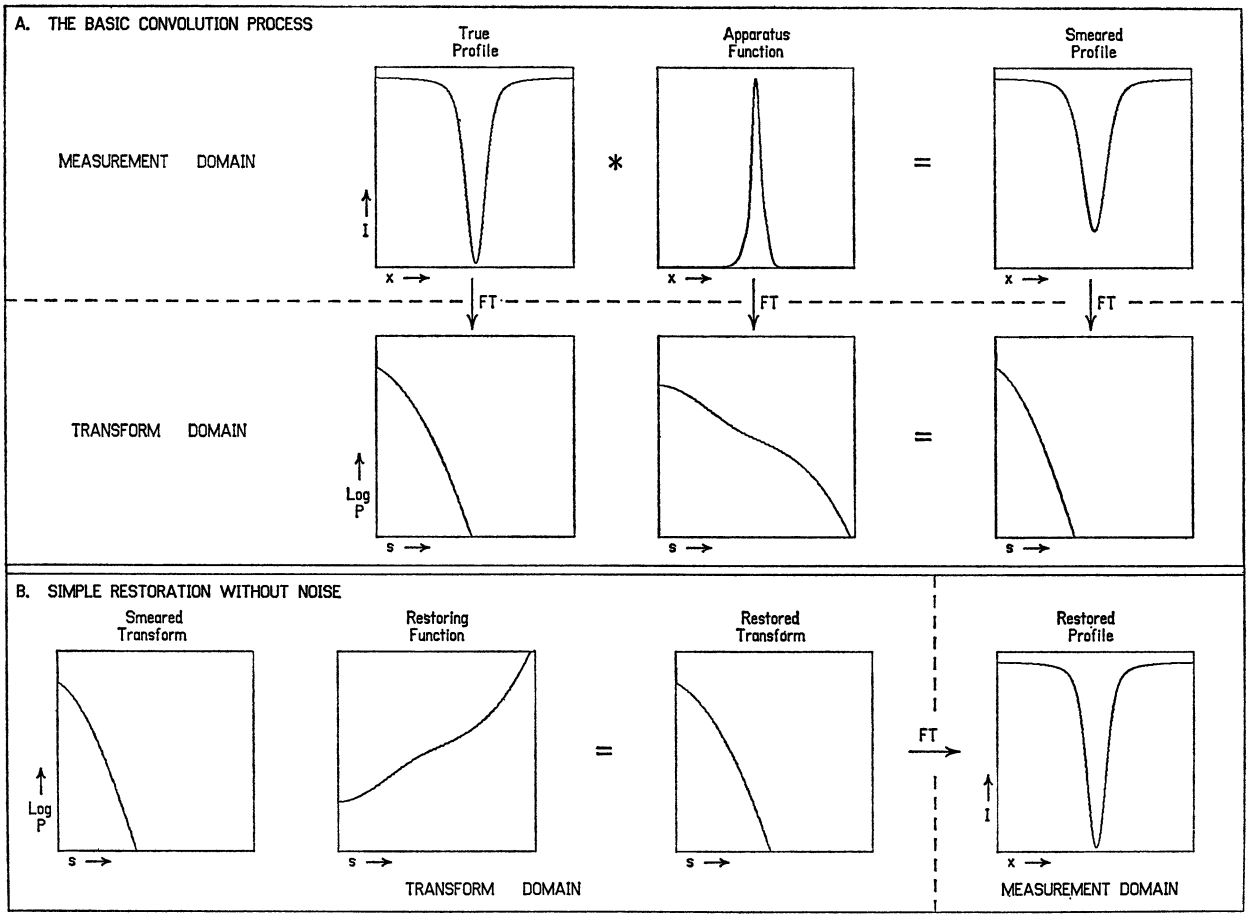

Fig. 12. Illustrations of (A) the convolution process as seen in both domains and (B) the simple Fourier restoration in the absence of noise

In fact, a close examination of the power spectrum of conservatively-sampled data invariably reveals that the highest frequency components contain virtually no information, and are essentially pure noise. It seems intuitively obvious that if the restored data is to be as close as possible to the true distribution, such high frequency components due to noise must be suppressed rather than enhanced, while those components that contain much more signal than noise should be rather completely restored. Subject to some very reasonable assumptions, these ideas lead directly to the optimum filter. The following development is intended to be illustrative rather than rigorous; more formal treatments may be found Wiener (1949), Middleton (1960, p. 697) and Bracewell (1958).

Essentially, three assumptions are necessary:

(i) The correction for the presence of noise is to take the form of a *filter*; that is, the various frequency components will be weighted in some optimum way:

$$\tilde{R}(s) = \frac{\tilde{O}(s)}{\tilde{A}(s)} \tilde{\Phi}(s) , \qquad (20)$$

where $\tilde{\Phi}(s)$ is the optimum filter that gives the „best" restored profile. Our approach will yield filter functions that are real in both domains; hence, they are not physically realizable as real-time electronic devices. Middleton (1960), p. 697, gives a more general discussion of the optimum filter with and without the constraint of physical realizability. Also implicit in our approach is the belief that the signal and the noise have fundamentally different frequency distributions.

(ii) The optimum filter must result in the closest smooth approach of $R(x)$ to $T(x)$ in the root-mean-square error sense:

$$\int [T(x) - R(x)]^2 \, dx = \text{a minimum} . \qquad (21)$$

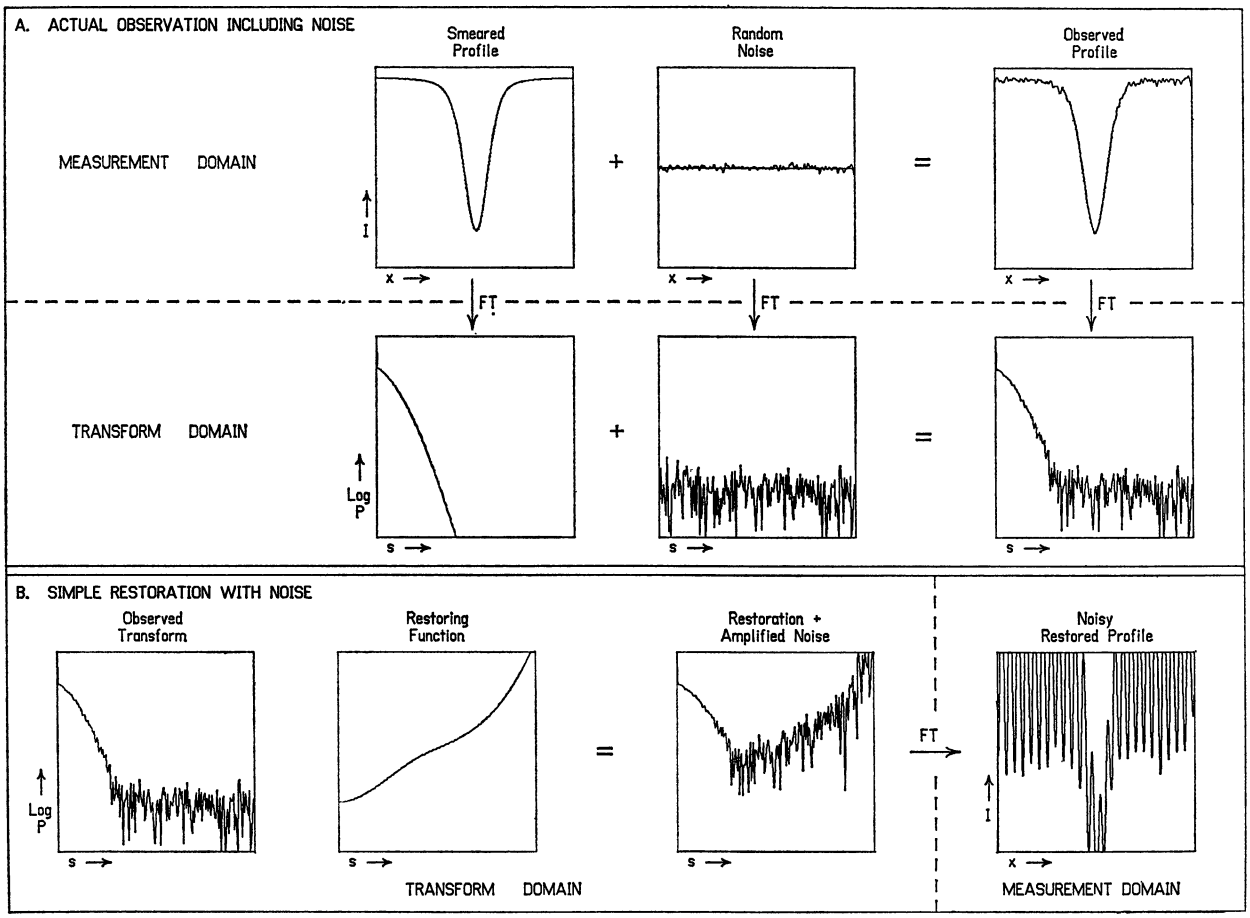

Fig. 13. Illustrations of (A) the effect of noise on the observed profile and its transform, and (B) the disastrous result of noise amplification in a simple restoration

(iii) The noise is random and uncorrelated with the signal. This assumption, though not essential, is usually valid and simplifies the problem.

The power theorem tells us that we may treat the RMS error in either domain:

$$\int [T(x) - R(x)]^2\, dx = \int |\tilde{T}(s) - \tilde{R}(s)|^2\, ds\ . \quad (22)$$

Hence the problem is reduced to finding the filter $\tilde{\Phi}(s)$ that makes

$$\int |\tilde{T}(s) - \tilde{R}(s)|^2\, ds = \int \left| \tilde{T}(s) - \frac{\tilde{O}(s)}{\tilde{A}(s)}\tilde{\Phi}(s) \right|^2 ds \quad (23)$$

a minimum.

Using Eqs. (15) and (18), we find that the residual error ε is

$$\varepsilon = \tilde{T}(s) - \tilde{R}(s) = \tilde{T}(s) - [\tilde{T}(s)\,\tilde{A}(s) + \tilde{N}(s)]\frac{\tilde{\Phi}(s)}{\tilde{A}(s)}$$

$$= \tilde{T}(s)\,[1 - \tilde{\Phi}(s)] - \frac{\tilde{N}(s)}{\tilde{A}(s)}\tilde{\Phi}(s)\ . \quad (24)$$

The error is now seen to be composed of two parts, the first due to incomplete restoration because of the presence of the filter, and the second the usual enhanced noise term now *reduced* by the filter. If we accept the third assumption above, the two terms are uncorrelated; and hence, on the average, their cross product will contribute nothing to the sum of squares. We are then left with finding $\tilde{\Phi}(s)$ such that

$$\int \varepsilon^2\, ds = \int \left\{ |\tilde{T}(s)|^2 (1 - \tilde{\Phi}(s))^2 + \left|\frac{\tilde{N}(s)}{\tilde{A}(s)}\right|^2 \tilde{\Phi}^2(s) \right\} ds$$

$$= \text{a minimum.} \quad (25)$$

This is a standard calculus of variations problem whose solution is given by

$$\frac{\partial \varepsilon^2}{\partial \tilde{\Phi}(s)} = 0 = -2\,[1 - \tilde{\Phi}(s)]\,|\tilde{T}(s)|^2 + 2\,\tilde{\Phi}(s) \left|\frac{\tilde{N}(s)}{\tilde{A}(s)}\right|^2 \quad (26)$$

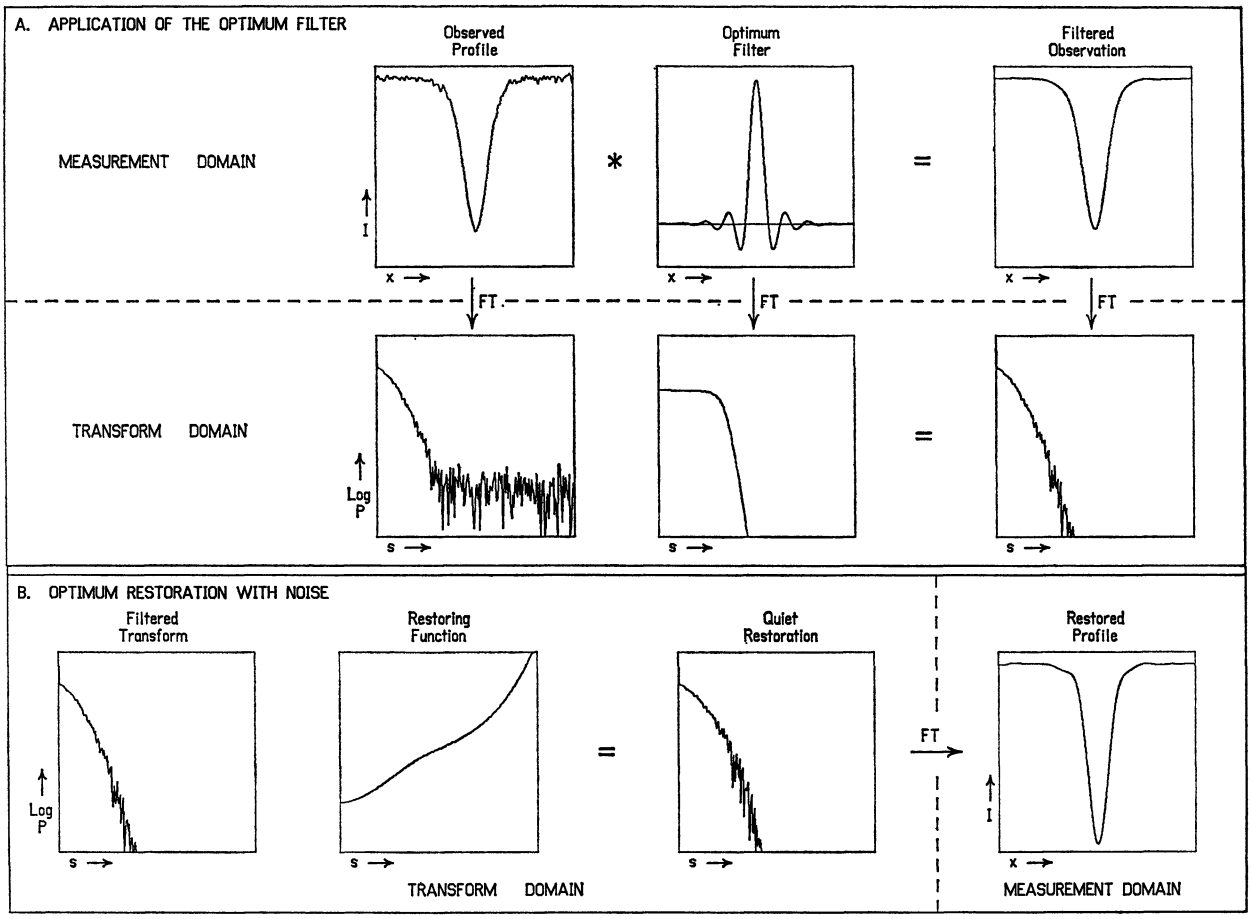

Fig. 14. Illustrations of (A) the noise reduction provided by the optimum filter, and (B) the resulting quiet restoration

or

$$\tilde{\Phi}(s) = \frac{|\tilde{T}(s)\,\tilde{A}(s)|^2}{|\tilde{T}(s)\,\tilde{A}(s)|^2 + |\tilde{N}(s)|^2} \equiv \frac{P_S(s)}{P_S(s) + P_N(s)} \qquad (27)$$

where $P_S(s)$ and $P_N(s)$ are the power spectra of the smeared signal and the noise. The optimum filter is thus seen to be determined only by the power of the smeared signal (i.e., that which would be observed in the absence of noise) and the actual noise power; there is no explicit dependence on the apparatus function. Also note that this optimum filter function is real and even in both domains.

The derivation of the filter function does not actually need to be connected to the restoration problem at all. The same answer is obtained by asking which filter, when applied to the observations $O(x)$, gives the best RMS approximation to the noise-free signal $S(x)$. Optimum restoration is thus nothing more than optimum filtering followed by straightforward Fourier restoration (see Fig. 14).

The optimum restoration can also be formulated in terms of a linear operator to be convolved directly with the observed data. In Fig. 14A we show this convolution process for optimum filtering alone, but by combining the filter and restoration functions the complete optimum restoration operator can be found for the convolution approach. Rybicki and Harrison (Harrison, 1968) are the principal proponents of the direct linear-operator technique, and they have successfully applied it to the restoration of both stellar and solar spectra.

The complete dependence of the optimum filter on observed quantities also indicates why the traditional Burger-van Cittert iterative restoration technique (see Burger *et al.*, 1932, 1933) should be avoided. Using this technique, the resulting transform of the restoration operator at the end of n iterations is

$$\tilde{R}(s) = \tilde{O}(s) \sum_{J=0}^{n} [1 - \tilde{A}(s)]^J . \qquad (28)$$

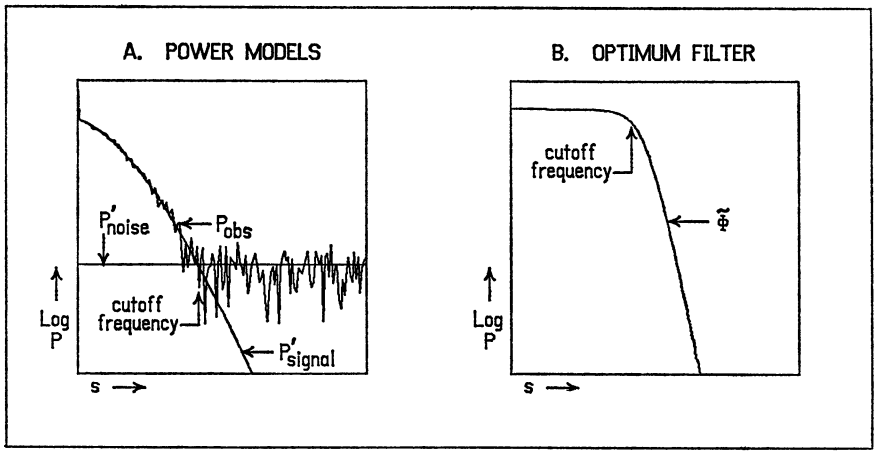

Fig. 15. Illustration of (A) how the smooth power models of the signal and noise are determined by the least squares fitting of the computed power spectrum of the data, and then used to construct the optimum filter as shown in (B)

The bracketed quantity is seen to be the first $(n+1)$ terms of a geometric progression, so we may immediately write

$$\tilde{R}(s) = \frac{\tilde{O}(s)}{\tilde{A}(s)}\{1 - [1 - \tilde{A}(s)]^{n+1}\}. \qquad (29)$$

Comparing this result with Eq. (20), we see that this process is formally equivalent to the use of a filter of the form

$$\tilde{\Phi}(s) = 1 - [1 - \tilde{A}(s)]^{n+1}. \qquad (30)$$

As a filter, this leaves much to be desired. As n becomes large, the term in brackets tends to vanish at all frequencies where $\tilde{A}(s)$ is significantly greater than zero. In a properly designed system, this means all frequencies up to the cut-off frequency. The net result is that $\tilde{\Phi}(s)$ approaches unity at all frequencies; i.e., no filtering at all. We are thus back to the unhappy situation of Fig. 13B. This is obviously unsatisfactory, and the usual remedy is to stop the process at some relatively small value of n, thus achieving some filtering action. Unlike the optimum filter, however, this filter is independent of observed quantities and depends only on the apparatus function. The all-important properties of the observed distribution thus enter only subjectively and indirectly through the choice of the number of iterations n.

2. Practical Realization of the Optimum Filter

The realization of the filter represented by Eq. (27) obviously presents a problem, since it involves both the quantity $\tilde{S}(s)$ we are attempting to measure and the noise $\tilde{N}(s)$ that is interferring with that measurement, What saves us is the fact that this is an *optimum* filter; therefore, small deviations from the true filter shape should result only in second order increases in the error. Consequently, it is usually adequate to replace the actual complicated signal and noise power spectra by smooth, simple models involving only a few parameters. If $P'_S(s; a_1, a_2, \ldots)$ represents a model of the signal power with parameters a_1, a_2, etc., and $P'_N(s; b_1, b_2 \ldots)$ represents a model of the noise power with parameters b_1, b_2, etc., we use a fitting procedure to adjust the a's and b's until

$$P'_{\text{total}}(s) = P'_S(s) + P'_N(s) \qquad (31)$$

is a good smooth approximation to P_{obs}. Having thus modelled both the signal and noise powers, we take the filter to be

$$\tilde{\Phi}(s) = \frac{P'_S(s)}{P'_S(s) + P'_N(s)} \qquad (32)$$

as shown in Fig. 15.

The choice of a power model depends somewhat on the type of data involved, and one should have considerable familiarity with empirical power spectra typically produced by the source and instrument. Every available piece of physical information should be used in setting up the model. Nonetheless, there are a few helpful generalizations.

(i) In most cases, a constant power is an adequate model of the noise power. Such "white" noise is typical of the statistical fluctuations in photon flux that are the dominant noise source in many measurements, and it is also characteristic of amplifiers

operated at reasonably high frequencies. For systems operated at very low frequencies, a noise power proportional to 1/s may be more appropriate.

(ii) The simplest multi-line spectra, such as those produced by typical laboratory emission sources, consist of a number of lines with essentially identical width and shape, and the shape is usually well represented by a Voigt function (van de Hulst *et al.*, 1947; Elste, 1953). The smoothed power spectra of such sources can be represented very well by a Voigt transform of the form

$$P'_s = P_0 \exp[-2\sqrt{2}\,\alpha\, ks - k^2 s^2]. \qquad (33)$$

The parameters α and k that correspond to the damping ratio and the Doppler width may often be specified from the known properties of the source, which leaves only P_0 to be determined from the observations.

(iii) The most complex case of all is typified by stellar absorption spectra, which contain randomly-spaced lines of widely differing width, damping and saturation. Such spectra are especially troublesome when narrow telluric lines are also involved. Even for such cases, however, a power model consisting of two Voigt transforms of different widths will usually suffice.

A smooth model of the power spectrum applicable to case (iii) is in regular use at Kitt Peak, but the number of the free parameters is reduced to 3 by arbitrarily fixing the damping ratios and by making the second transform twice as wide as the first. Using a constant for the noise term, we then have a 4-parameter model to be fitted to the observed power spectrum:

$$P'_{\text{total}}(s; P_0, P_1, k, P_n) = P_0 \exp(-2\sqrt{2}\alpha\, ks - k^2 s^2) + P_1 \exp[-2\sqrt{2}\,\alpha(k/2)\, s - (k/2)^2 s^2] + P_N. \qquad (34)$$

In order to emphasize the important region where the signal power equals the noise power, we determine the parameters by minimizing the error in the ratio $P_{\text{obs}}(s)/P'_{\text{total}}(s)$:

$$\Sigma\left(\frac{P^{\text{obs}}(s)}{P'_{\text{total}}(s)} - 1\right)^2 = \text{minimum} \qquad (35)$$

3. Treatment of the Data

Before any of the operations involved in modelling the power spectrum or restoring the data are performed, it is absolutely essential that an appropriate data window be applied to the observations as described in Section II.*B*.5. If this is *not* done, the true shape of the power spectrum in the middle and high frequencies will be hidden by leakage from other parts of the spectrum. As a practical rule the end-region masking should be applied over 2 to 4 characteristic line-widths at both ends of the spectral scan.

The cosine bell (see Section IV.*B*.2, Eq. 13) is a convenient functional form for the data window. The effect of masking cannot be completely removed from the restored data because of the band-width restriction; therefore, the data within the masked end-regions should not be used in further analysis.

4. Treatment of the Apparatus Function

Two important properties of a convolution must be recognized in the handling of the apparatus function (see Bracewell, 1965, p. 111):

(i) The area under a convolution is equal to the product of the areas of the convolved functions.

(ii) The centroid of the convolution is equal to the sum of the centroids of the functions being convolved.

The restoration process, of course, produces the exact opposite effects. To avoid wavelength shifts and scale changes, the apparatus function should be normalized to unit area and shifted until its centroid coincides with the coordinate origin. The procedure we commonly use is to find the centroid (and hence the needed shift) in the measurement domain, take the transform, and apply the shift theorem in the transform domain. Dividing all values of the transform by the zero-frequency term then completes the area normalization, and the transform of the apparatus function is ready for use in restoration.

Occasionally the apparatus function transform will be zero at some frequency, usually due to the filtering action of an optical aperture. Since this transform appears in the denominator of the restoration Eq. (20), the noise at the nodal frequency will be enormously amplified. This practical problem may be avoided by setting an upper bound to the allowed amplification, or by limiting the range of the restoration to frequencies below the nodal frequency.

5. Applicability to Photon-Counting Photometers

At first glance, a photon-counting system appears to be fundamentally different from an analog photometer because it cannot contain an analog filter to limit the band-width of the measurement. In a photon counter, the noise suppression is achieved instead by integrating over a time interval τ. Such integration is equivalent to a filter function of the form $\sin(\pi\tau s)/\pi s$ in the frequency domain. Because of the slow decay

of this function, many aliases contribute to the noise power at a given frequency. Although such aliasing appears to be bad, the resultant sum gives a flat noise spectrum that is equivalent to that obtained by filtering the noise signal with an ideal filter which has its cutoff at the Nyquist frequency. Consequently, as long as the signal is over-sampled, the optimum restoration technique described previously is directly applicable to photon-counting measurements.

V. Conclusions

If a truly general conclusion is to be drawn from our experience with Fourier analysis and restoration of measurements, it is that such processes are strongly dependent on the characteristics of the measured signal. The blind application of recipes, which by their very nature tend to be conservative, can then give compromised or even poor results. One must examine all of the available physical properties of the measurement, i.e., the data itself, the modified power spectrum, and the known properties of the instrument and source, before any technique can be applied intelligently. The Fast Fourier Transform proves to be a useful tool for efficiently making these examinations of the data as well as for performing the final analysis and restoration.

Acknowledgements. We thank Drs. R. M. Bracewell and G. Rybicki for their contributions to our work toward practical solutions of the convolution problem for empirical data. Their work introduced us to the optimum filter, and we continue to rely on them for elegant, rigorous mathematical descriptions of the background theory.

Drs. K. Pierce, G. Rybicki, R. Canfield, R. Altrock, M. Cha, R. Miller, A. Code, D. Hall and Mr. C. Slaughter served patiently as manuscript readers, and we are indebted to them for their suggestions, many of which have been incorporated into the text. Likewise, two unknown referees contributed to this work by pointing out the shortcomings in notation and presentation and thus substantially improved the final result. We thank Miss C. J. Webb of the University of Texas who directed us to the recent work of Dr. J. W. Cooley and his collaborators. Credit is certainly due Ruby Fulk of HAO who carefully prepared the many preliminary manuscripts.

References

Bergland, G.D. 1968, *Comm. ACM* **11**, 703.
Bergland, G.D. 1969, *IEEE Spectrum* **6**, 7, 41.
Bingham, C., Godfrey, M.D., Tukey, J.W. 1967, *IEEE Trans.* AU-15 **2**, 56.
Blackman, R.B., Tukey, J.B. 1958, The Measurement of Power Spectra, Dover Publ. New York.
Bracewell, R.N. 1958, *Proc. IRE* **46**, 106.
Bracewell, R.N. 1965, The Fourier Transform and Its Applications, McGraw-Hill, New York.
Brenner, N.M. 1967, *Lincoln Lab. Techn. Note* 1967-2, MIT.
Brenner, N.M. 1968 (private communication.)
Burger, H.C. van Cittert, P.H. 1932, *Z. Phys.* **79**, 722.
Burger, H.C., van Cittert, P.H. 1933, *Z. Phys.* **81**, 428.
Cochran, W.T., Cooley, J.W. Favin, D.L., Helms, H.D., Kaenel, R.A., Lang, W.W., Maling, G.C., Nelson, D.E., Rader, C.M., Welch, P.D. 1967, *IEEE Trans. Audio Electroacoustics*, Special Issue on Fast Transform and Its Application to Digital Filtering and Spectral Analysis, AU-15, **2**, 45.
Cooley, J.W., Tukey, J.W. 1965, *Math. Comput.* **19**, 297.
Cooley, J.W., Lewis, P.A.W., Welch, P.D. 1970a, *J. Sound Vib.* **12** (3), 315.
Cooley, J.W., Lewis, P.A.W., Welch, P.D. 1970b, *J. Sound Vib.* **12** (3), 339.
Elste, G. 1953, *Z. Astrophys.* **33**, 39.
Forman, M.L. 1966, *J. O. S. A.* **56**, 978.
Gentleman, W.M., Sande, G. 1966, American Federation of Information Processing Societies Proc. Vol. 29, Spartan Books, Washington.
Gold, B., Rader, C.M. 1969, Digital Processing of Signals, Lincoln Laboratory Publications, Cambridge, Massachusetts.
Harrison, P. 1968, Thesis: Restoration of Astronomical Spectra, Harvard.
Hulst, van de, H.C., Reesnick, J.J. 1947, *Ap. J.* **106**, 121.
Jenkins, G.M., Watts, D.G. 1968, Spectral Analysis and Its Applications, Holden-Day, San Francisco.
Middleton, D. 1960, An Introduction to Statistical Communications Theory, McGraw-Hill, New York.
Richards, P.I. 1967, *IEEE Spectrum*, Computing Reliable Power Spectra, p. 84.
Singleton, R.C. 1968, *Comm. ACM* **11**, 773.
Singleton, R.C. 1969, *IEEE Trans.* AU-17, p. 93.
Welch, P.D. 1967, *IEEE Trans.* AU-15, **2**, 70.
White, O.R., Cha, M. 1967, Application of the Cooley-Tukey Algorithm in Computing Fourier Transform, Internal Report No. 3, Astrophysics Division, Dept. of Physics and Astronomy, University of Hawaii.
Wiener, N. 1949, Extrapolation, and Smoothing of Stationary Time Series, Wiley, New York.

J. W. Brault
Kitt Peak National Observatory
Tucson, Arizona 85717, USA

O. R. White
High Altitude Observatory
National Center for
Atmospheric Research
Boulder, Colorado 80302, USA