1990ApJS...74..437S

THE ASTROPHYSICAL JOURNAL SUPPLEMENT SERIES, 74:437–461, 1990 October
© 1990. The American Astronomical Society. All rights reserved. Printed in U.S.A.

ANALYSIS OF ZEEMAN EFFECT DATA IN RADIO ASTRONOMY

R. J. SAULT, N. E. B. KILLEEN, J. ZMUIDZINAS, AND R. LOUSHIN
Astronomy Department, University of Illinois
Received 1989 October 9; accepted 1990 March 7

ABSTRACT

We discuss the analysis of Zeeman effect data in radio astronomy and in particular, extend previous techniques to include the case of low signal-to-noise ratios. We consider three statistical techniques for estimating the line-of-sight magnetic field: maximum likelihood, least-squares, and Wiener filters. For high signal-to-noise ratios, all three estimators are essentially unbiased. In the poor to moderate signal-to-noise ratio regime, we conclude that all three estimators are biased; the maximum likelihood technique yields results that are, in general, substantially less biased than least-squares and Wiener filters. However, it is possible to "debias" the least-squares results and obtain estimates that are as good as maximum likelihood under a restricted set of conditions.

Subject headings: magnetic fields — radio sources: general — Zeeman effect

I. INTRODUCTION

The Zeeman effect provides a direct method for measuring the strength of a magnetic field that permeates a gas. The splitting of spectral lines (in emission or absorption), in the weak field case, is proportional to the magnetic field strength. This relatively simple principle has induced many to use the Zeeman effect as a probe of astrophysical magnetic fields (e.g., Verschuur 1969; Troland and Heiles 1986, and references therein).

In this paper, we discuss statistical methods by which the magnetic field strength can be estimated from Zeeman data. However, we will not be concerned in any detail with instrumental effects (see Troland and Heiles 1982, and references therein), which should always be carefully analyzed before Zeeman data is interpreted. Nor will we discuss any inherent ambiguities in the astrophysical interpretation of Zeeman data.

In general, Zeeman experiments have been in the regime of very high signal-to-noise ratio total intensity spectral lines (e.g., Kazes and Crutcher 1986); in this case, the analysis with least squares is fairly straightforward. However, this is the only statistical estimator that has been considered in the literature up to this time. In this paper we investigate analysis techniques for the low signal-to-noise ratio case. We show that the linear least-squares and Wiener filter techniques lead to badly biased results, and that the maximum likelihood approach is superior (although still biased) in this regime. We also find that under certain restricted conditions, it is possible to "debias" the least-squares estimate to produce results that compare favorably with maximum likelihood. In addition we show that for high signal-to-noise ratios, all three of these estimators are unbiased. Finally, we consider the effects of spectral and spatial correlations of the noise and suggest methods with which to deal with them.

This paper is organized as follows. In § II we introduce the basic equations which describe the response of a radio telescope to the Zeeman split lines. In § III we discuss statistical estimation of the Zeeman splitting. These methods are then applied to simulated data in § IV for evaluation, and we summarize our findings in § V. Detailed treatment of the statistical methods summarized in § III is described in the appendices.

II. RADIO OBSERVATIONS OF THE ZEEMAN EFFECT

a) *The Zeeman Effect*

The Zeeman effect is the splitting of spectral lines into multiple components because of the application of an external magnetic field which removes degeneracies in atomic or molecular quantum levels (e.g., Townes and Schawlow 1975). In principle, the Zeeman pattern of lines can become rather complicated. However, in practice, a simple triplet pattern (one unsplit and two split components) is generally all that is necessary to consider when making astrophysical measurements. Although each of the triplet lines may really consist of a group of closely spaced lines (e.g., the 1667 MHz OH line), the spacing is generally substantially below the spectral resolution of the observation for typical astrophysical magnetic field strengths.

Of fundamental importance to Zeeman experiments are the polarization characteristics of the lines. The split lines are circularly and oppositely polarized, the electric vectors rotating in the plane perpendicular to the magnetic field direction. The unsplit line is linearly polarized in the direction of the field. Thus, when observed along the field, the unsplit line is not detected and the split lines are completely circularly polarized. When observed in the direction perpendicular to the magnetic field, all three lines appear linearly polarized; the unsplit line in the direction of the field, the split lines perpendicular to the field. The general case of an arbitrary viewing angle, of course, yields elliptical polarization for the split lines.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

b) Response of the Telescope

To find the response of a radio telescope to these three lines, we assume that the telescope feeds are circularly and oppositely polarized, and we denote their responses by I_+ and I_-. A similar analysis could be carried out for linearly polarized feeds. By considering the power radiated per unit solid angle into a given direction and polarization by each line (e.g., Jackson 1975, p. 396) one can show that

$$I_\pm(\nu) = \tfrac{1}{4}\left[(\cos\theta \pm 1)^2 I_0(\nu + \delta\nu) + (\cos\theta \mp 1)^2 I_0(\nu - \delta\nu) + 2\sin^2\theta I_0(\nu)\right], \quad (2.1)$$

where θ is the angle between the magnetic field and the line of sight, $\delta\nu$ is the frequency shift between the split and unsplit lines, and $I_0(\nu)$ is the Stokes $I = (I_+ + I_-)/2$ (total intensity) spectrum in the case that $\delta\nu = 0$. The first two terms arise from the circularly polarized split lines, and the third component is contributed by the unsplit linearly polarized line. Clearly, the latter component is the same for each feed. It is also evident from equation (2.1) that the response of, say, the I_+ feed to the circularly polarized line represented by the first term, is maximum for $\theta = 0$, and zero for $\theta = 180°$.

i) *Large Splitting*

If the splitting was large enough that all three Zeeman components were well-separated, then the total magnetic field strength could be obtained from the measured separation of the split components. In addition, the angle θ could be obtained from the ratio, γ, of the intensities of either of the split components in each of the circularly polarized feeds. For example, for the range of ν corresponding to the first split component (assuming positive $\delta\nu$),

$$\gamma = \frac{I_+(\nu)}{I_-(\nu)} \approx \frac{(\cos\theta + 1)^2 I_0(\nu + \delta\nu)}{(\cos\theta - 1)^2 I_0(\nu + \delta\nu)} = \frac{(\cos\theta + 1)^2}{(\cos\theta - 1)^2},$$

from which it follows that

$$\cos\theta = \frac{-(1+\gamma) \pm 2(\gamma)^{1/2}}{1-\gamma}.$$

An ideal, noiseless, large splitting case is demonstrated in the upper panel of Figure 1 in which $2\delta\nu$ is 5 times the line FWHM of 40 channels (note we measure all spectra widths or dimensions in channels). In addition, $\theta = 45°$ so that the split lines are elliptically polarized and appear in both feeds. The splitting is obviously recovered correctly by direct measurement and for either the I_+ or I_- profiles, we measure $\gamma = 33.97$ from which $\theta = 45°$ is recovered.

The large splitting case is often applicable to masers. However, in practice, it is rare that all three lines are observed because of depolarizing effects (e.g., Goldreich, Keeley, and Kwan 1973; Lo *et al.* 1975), so that measurement of θ is generally not possible.

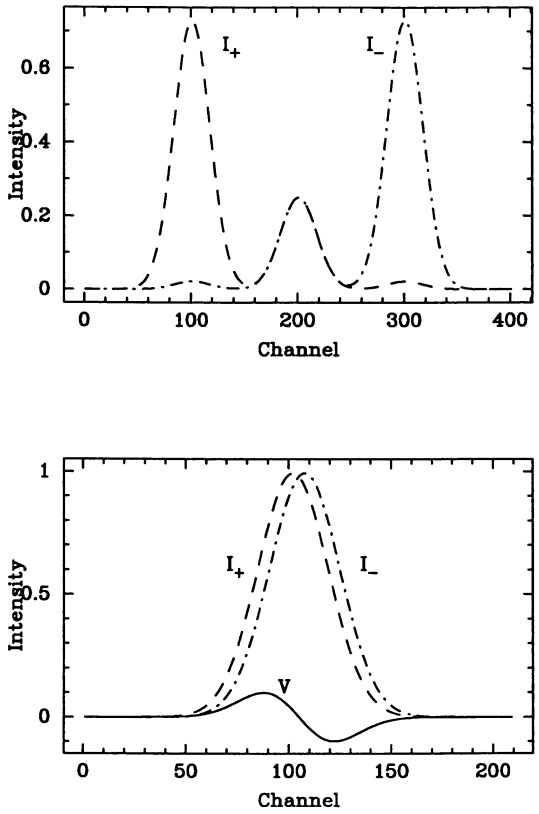

FIG. 1.—Simulated large (*upper panel*) and small (*lower panel*) splitting cases. I_0 is a single Gaussian, and there is no noise. Each panel shows I_+ (*dashed line*) and I_- (*dot-dashed line*) spectra for a magnetic field at 45° to the line of sight. In addition, the resultant V spectrum (*solid line*) is shown for the small splitting case.

ii) *Small Splitting*

Many astrophysical Zeeman experiments are not so well-endowed as to fall into the large splitting case. In the lower panel of Figure 1, we illustrate a case where the three lines of the Zeeman pattern blend together in each feed. Clearly, neither B nor $\cos\theta$ is directly measureable, and we are forced to use a less direct method to extract them. From equation (2.1), it is straightforward to show that the Stokes $V = (I_+ - I_-)/2$ spectrum is

$$V(\nu) = \tfrac{1}{2}\cos\theta\left[I_0(\nu + \delta\nu) - I_0(\nu - \delta\nu)\right]. \quad (2.2)$$

Before proceeding, we add (as is usual) a leakage term to equation (2.2). This term might arise from at least two effects; the leakage of linear polarization into circular polarization, and beam squint (Troland and Heiles 1982). Each effect results in a replica of I appearing in V. For beam squint, the primary beams for the oppositely polarized feeds are offset on the sky by a small fraction of the primary beam width. The effect for a single dish can be rather insidious; a suitably pathological source velocity structure can induce a false Zeeman profile (Troland and Heiles 1982). However,

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

for a synthesis array the effect is far less dangerous. This is because squint is a *primary*, not a *synthesised*, beam effect. In this case, at an instant in time, beam squint causes a gain difference between the oppositely polarized feed responses. It will be a function of distance from the pointing center along the squint pattern direction. If there is no Zeeman splitting, this causes a replica of the I spectrum to appear in the V spectrum, the strength of which increases from the pointing center (until the primary beam tapers off).

We now make the fundamental assumption that the splitting, $\delta\nu$, is small compared to the line width, which allows the following three simplifications. First, the derivative of I_0 is a good approximation to the difference term in equation (2.2); second, the inclusion of leakage in the manner described above is approximately valid; and third, the unmeasured, unsplit spectrum, $I_0(\nu)$, is adequately approximated by the measured spectrum, $I(\nu)$. Therefore, we can write equation (2.2) as

$$V(\nu) = \delta\nu \cos\theta \frac{dI(\nu)}{d\nu} + \beta I(\nu). \tag{2.3}$$

Note that the $\delta\nu$ and $\cos\theta$ terms arise from different sources. The $\delta\nu$ term arises from the difference term in equation (2.2) and is directly related to the Zeeman effect. However, the $\cos\theta$ term is a geometric factor that arises when considering the response of the feeds to polarized radiation. Thus, although $\delta\nu$ and $\cos\theta$ combine mathematically to yield the line-of-sight magnetic field, they arise from completely different considerations.

With the above assumptions, a technique is needed to extract the desired quantities of equation (2.3) from the observed $V(\nu)$ and $I(\nu)$ spectra. In the case of very high signal-to-noise ratios and ignoring leakage, one could divide V by $dI/d\nu$ and find $\delta\nu\cos\theta$ as a function of ν (e.g., Schwarz *et al.* 1986). Alternatively, as is usually done, $\delta\nu\cos\theta$ is extracted by computing $dI/d\nu$ and fitting it to the V spectrum (e.g., Crutcher, Kazes, and Troland 1987).

In the following section, we discuss methods for fitting the numerically computed derivative of the I spectrum to the V spectrum in order to extract $\delta\nu\cos\theta$ and β.

III. ANALYSIS TECHNIQUES

a) *Some Preliminaries*

First, we specify some notation that we use throughout the rest of this paper. A quantity that has a carat is the true value of that quantity. For example, $\hat{I}(\nu)$ is the true Stokes I spectrum that would be measured with a perfect, noiseless receiving system. A quantity with the subscript "est" is a statistical estimate of that quantity. For example, $I_{\rm est}(\nu)$ is an estimate of $\hat{I}(\nu)$, the true spectrum. A quantity with the subscript "obs," is the observed value of that quantity. For example, $I_{\rm obs}(\nu)$ is the observed Stokes I spectrum. If we do not distinguish a quantity by carats or subscripts such as these, then we are referring to it in a generic sense. For example, $I(\nu)$ could mean any of the specific types of the Stokes I spectrum; the context will make which one clear. The quantities that will be subject to this notation in the following sections are $I(\nu)$, $V(\nu)$, α, and β.

We rewrite equation (2.3) as

$$\hat{V} = \hat{\alpha}\frac{d\hat{I}}{d\nu} + \hat{\beta}\hat{I}, \tag{3.1}$$

where $\hat{V}$ and $\hat{I}$ are the true, noiseless spectra, and $\hat{\alpha}$ and $\hat{\beta}$ are the true splitting and leakage terms. In reality, only the noisy measurements $I_{\rm obs}$ and $V_{\rm obs}$ are available. Three techniques for extracting estimates of $\hat{\alpha}$ and $\hat{\beta}$ are considered below: least-squares, maximum likelihood, and Wiener filter. All three techniques minimize a sum of squares, though the techniques differ in the actual form of the sum, and their theoretical justification. The estimates produced by each of these techniques has some expected value and standard deviation. These are denoted by $\bar{\alpha}_{\rm est}, \sigma^2_{\alpha}$, etc. For an unbiased estimator, the expected value will be the true value, and the standard deviation is a measure of the error in the estimate. Unfortunately we will see that all three estimators are biased at sufficiently low signal-to-noise ratios.

We require an estimate of the derivative of the I spectrum. We use either the simple one- or two-sided numerical derivatives given by

$$\frac{dI_i}{d\nu} \approx (I_i - I_{i-1}),$$

and

$$\frac{dI_i}{d\nu} \approx \frac{1}{2}(I_{i+1} - I_{i-1}).$$

Note that when I is noisy, then the noise variance in the two-sided derivative is one-fourth that in the one-sided derivative. Hence we would expect the two-sided derivative to cope better when the signal-to-noise ratio is poor. On the other hand, the two-sided derivative is less accurate at sharp edges or for narrow lines.

Generally it is the signal-to-noise ratio in the derivative of I (rather than I itself) which is of importance. We define $\hat{\eta}$, a measure of signal-to-noise ratio in the derivative spectra, as

$$\hat{\eta} = \frac{1}{\sigma}\frac{\hat{V}_{\rm rms}}{\hat{\alpha}} = \frac{1}{\sigma}\left(\frac{d\hat{I}}{d\nu}\right)_{\rm rms}. \tag{3.2}$$

Here $\hat{V}_{\rm rms}$ and $(d\hat{I}/d\nu)_{\rm rms}$ are the rms values of the noiseless $\hat{V}$ and $d\hat{I}/d\nu$ spectra.

In this paper, the terms "summing" and "averaging" are used with quite distinct meanings. It is important to distinguish between them. All the techniques discussed sum an error measure. This summation can be performed in both the spectral dimension and/or the two spatial dimensions. Note that α and β are assumed constant in this summing region. Provided the region is sufficiently small, this is a good approximation. However, in poor signal-to-noise ratio cases, there is a trade-off between this approximation and summing over a region large enough to gain sufficient signal. Needless to say, in some instances, there is insufficient signal and coherence in α to achieve meaningful results.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

Spatial summing of the error measure is quite distinct from averaging spectra from several spatial locations and then passing the averaged spectrum to the estimation algorithms. Averaging spectra implicitly assumes that the shapes of the spectra are the same, whereas spatial summing does not. Generally, it is preferable to avoid averaging.

The summations in this paper should strictly be represented by triple sums (the spectral dimension and the two spatial dimensions). However, for brevity, we use only one summation symbol, with the understanding that extension to the more general case is moderately straightforward. We usually assume we are summing over N data points.

In developing analysis techniques for cases of poor signal-to-noise ratio, attention must be paid to the properties of the noise processes present. We assume that the characteristics of the noise processes in the I_+ and I_- spectra are identical, and in particular, that they are zero mean Gaussian noise. Although we assume the noise is uncorrelated between the I_+ and I_- spectra, we allow the noise to be correlated between channels and pixels (i.e., along the spectral and spatial dimensions). Thus, as with the summations, the autocorrelation function of the noise process varies over three dimensions. We express it as a function of a single index, with the understanding that extension is possible. Let the autocorrelation function for the $I_\pm$ spectra be $2\sigma^2 r_n(i)$, where the noise variance is $2\sigma^2$. That is, $r_n(i)$ is the correlation coefficient for two points i elements apart. Under these assumptions, the noise processes in the I and V spectra are also zero mean Gaussian noise. The noise between I and V is uncorrelated, though there is a correlation between channels and pixels, with autocorrelation function $\sigma^2 r_n(i)$.

All these assumptions are accurate characterizations of the noise processes that occur in radio observations. If thermal noise in the receiver is the dominant noise source, we expect Gaussian noise. Because the I_+ and I_- receivers are independent (or the same receiver but switched in time), the noise is uncorrelated. The assumption that the noise processes in the two receivers have identical variances is generally a good approximation. This is especially true in the case of synthesis telescopes, since the $I_\pm$ spectra at a given spatial location contain contributions from many independent receivers. The correlation between frequency channels is determined by the telescope spectrometer and the on-line data gathering software. For example, the spectra may be Hanning smoothed on-line, which introduces correlations between otherwise independent frequency channels in a predictable fashion. Consequently, we can often derive the channel-to-channel autocorrelation function purely by considering the telescope characteristics and observing mode. For a synthesis array, the correlation along the spatial dimensions is proportional to the autocorrelation of the synthesized beam, which in turn is determined by the array geometry and the imaging software. Generally the autocorrelation function in the spectral dimension is separable from the two spatial dimensions. Thus, the spectral and spatial autocorrelation functions are usually computable from the characteristics of the observation. If this is not possible, the autocorrelation can be determined by finding the numerical autocorrelation in signal-free regions. At the least, this is a good check.

b) Summary of Definitions

To summarize, we observe Stokes I and V spectra, $I_{\rm obs}$ and $V_{\rm obs}$, which are noisy versions of the true spectra $\hat{I}$ and $\hat{V}$. The spectra are of length N channels and the variance of the noise is σ^2. The true splitting parameter is $\hat{\alpha}$ channels, and our aim is to find an estimate of this, $\alpha_{\rm est}$. We allow $V_{\rm obs}$ to be corrupted by a factor β of $\hat{I}$. A parameter giving the signal-to-noise ratio of the rms value of the $\hat{I}$ derivative spectrum is defined as $\hat{\eta} = \sigma^{-1}(d\hat{I}/{\rm d}\nu)_{\rm rms}$.

c) Statistical Estimators

i) *Least Squares*

Least squares is a simple technique which finds those values of α and β which minimize

$$\varepsilon^2 = \sum_i \left(V_{{\rm obs},i} - \alpha \frac{dI_{{\rm obs},i}}{d\nu} - \beta I_{{\rm obs},i} \right)^2 .$$

It deals neither with any correlations nor with the noise processes themselves so that the justification for doing this is intuitive rather than theoretical. The primary advantage of this technique (over the others discussed) is that it is computationally cheap and easy, requiring calculation of some inner products (see Appendix A for details). When the signal-to-noise ratio is high, the variance in the uncertainty of the least-squares estimate is given by (see Appendix A)

$$\sigma_\alpha^2 = \sigma^2 \Big/ \sum_i \left(\frac{d\hat{I}_i}{d\nu} \right)^2 = \frac{\hat{\alpha}^2 \sigma^2}{\sum_i \hat{V}_i^2} . \tag{3.3}$$

This equation requires the values of $\hat{I}$ or $\hat{\alpha}$ and $\hat{V}$. In practice, we use $I_{\rm obs}$. A serious disadvantage of least squares is that the estimate of $\hat{\alpha}$ is biased. By this we mean that the expected value of $\alpha_{\rm est}$ (that is, the average of an infinite number of trials of the same experiment) is not $\hat{\alpha}$. More specifically (see Appendix A), the expected value is

$$\bar{\alpha}_{\rm est} = \frac{\hat{\alpha}}{1 + a\hat{\eta}^{-2}} , \tag{3.4}$$

where $a = 2$ or $1/2$ for the one- and two-sided derivatives, respectively. This equation assumes that N is large, and that the data are spectrally uncorrelated (this assumption is relaxed in Appendix A). The quantity $\hat{\eta}$ is a measure of the signal-to-noise ratio in the derivative spectra. So as the signal-to-noise in the derivative spectra decreases, the bias increases. If the error variance given in equation (3.3) is calculated using $I_{\rm obs}$, then it too will be biased by exactly the same factor.

If we are given σ^2, then we can partially remove the bias from the equations used to estimate $\hat{\alpha}$. That is, we can "debias" the least-squares estimate. This is discussed further in § IV*e* and Appendix A.

ii) *Maximum Likelihood*

The maximum likelihood principle (Jeffreys 1939; Parratt 1961) is a general guiding principle for developing algorithms

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

to extract parameters from noisy or random data. For maximum likelihood techniques, a probability density function is derived in terms of the experimentally measured quantities and the unknown parameters. This density function expresses the probability of obtaining the experimental results for given values of the unknown parameters. The maximum likelihood principle states that we choose, as our solution, those parameters which maximize the probability density function.

One should not confuse maximum likelihood techniques with least-squares techniques. The maximum likelihood principle is far more general than least squares. However, in special cases they are identical. One such case occurs when the independent variables are noiseless and the dependent variable is corrupted by uncorrelated Gaussian noise. This is not the case with the Zeeman fitting problem, as the "independent variables" (the I spectra) are noisy. Thus the maximum likelihood and least-squares solutions are distinct in our case.

For clarity of exposition, we give only an outline here. In addition, we assume that the noise processes are totally uncorrelated. More detailed analysis (including the correlated noise case) is given in Appendix B. The desired probability density function is proportional to

$$\exp\left(\frac{-\chi^2}{2}\right), \tag{3.5}$$

where

$$\chi^2 = \frac{1}{\sigma^2}\left[\sum_i n_{I_i}^2 + \sum_i n_{V_i}^2\right]. \tag{3.6}$$

With the use of equation (3.1), the noise processes, n_I and n_V, can be given in terms of the measured quantities and the unknowns $\hat{\alpha}$, $\hat{\beta}$, and $\hat{I}$:

$$n_I = I_{\rm obs} - \hat{I},$$

$$n_V = V_{\rm obs} - \hat{V} = V_{\rm obs} - \hat{\alpha}\frac{d\hat{I}}{d\nu} - \hat{\beta}\hat{I}. \tag{3.7}$$

Thus, combining equations (3.6) and (3.7) yields

$$\chi^2(\hat{\alpha},\hat{\beta},\hat{I}) = \frac{1}{\sigma^2}\left[\sum_i \left(I_{{\rm obs},i} - \hat{I}_i\right)^2 + \sum_i \left(V_{{\rm obs},i} - \hat{\alpha}\frac{d\hat{I}_i}{d\nu} - \hat{\beta}\hat{I}_i\right)^2\right]. \tag{3.8}$$

The maximum likelihood estimates for $\hat{\alpha}$, $\hat{\beta}$, and $\hat{I}$ are those that minimize χ^2, and are denoted by $\alpha_{\rm est}$, $\beta_{\rm est}$, and $I_{\rm est}$. The minimization is a nonlinear problem, and Appendix B gives details of an algorithm to solve this. This minimization must be treated with some caution. Multiple minima are possible and, indeed, do occur when the signal-to-noise ratio is poor. In these cases there is often a spurious minimum close to $\alpha = 0$.

It can be shown that if either $\hat{\alpha}$ and $\hat{\beta}$ or $\hat{I}$ are known, the maximum likelihood technique leads to an unbiased estimate of the remaining unknowns. However, when none of these are known, the maximum likelihood technique proves to be biased if the signal-to-noise ratio is poor. Because of the nonlinear nature of the problem, an expression for the bias has not been derived; we have determined it for certain cases through numerical simulation (see § IV).

A measure of σ_α, the uncertainty of $\alpha_{\rm est}$, is as important as the estimation itself. We can derive this quantity by observing how quickly χ^2 varies with α near the minimum (see Press *et al.* 1986, and references therein). This technique is valid, provided that the signal-to-noise ratio is high enough so that χ^2 is essentially parabolic within the range $\alpha_{\rm est} \pm \sigma_\alpha$. The expression for σ_α is given by (see Appendix B)

$$\sigma_\alpha^2 = \left(\frac{1}{2}\frac{d^2}{d\alpha^2}\left[\chi^2(\alpha,\beta_\alpha,I_\alpha)\right]\right)^{-1}, \tag{3.9}$$

where β_α and I_α are those values of β and I which minimize χ^2 (as defined in eq. [3.8]) for a given value of α. The second derivative in this expression is evaluated at $\alpha_{\rm est}$. Some arguments indicate that, as the signal-to-noise ratio increases, the second derivative in equation (3.9) approaches the second partial derivative of χ^2 with respect to α. This leads to an explicit expression of σ_α for high signal-to-noise ratios:

$$\sigma_\alpha^2 \approx \sigma^2 \Big/ \sum_i \left(\frac{dI_{{\rm obs},i}}{d\nu}\right)^2$$

This is the same error variance as the least-squares algorithm (eq. [3.3]). When the signal-to-noise is moderate to poor, it is better to determine the second derivative by fitting a parabola to $\chi^2(\alpha,\beta_\alpha,I_\alpha)$ (Appendix B). When the signal-to-noise ratio is quite poor, these methods underestimate σ_α.

We now consider briefly the case where the noise is correlated. As discussed in Appendix B, the estimated values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{I}$ vary little if we do a full treatment of the correlation, or if we simply ignore it. However, the value of σ_α is significantly affected. A full treatment is unattractive for a number of reasons, so two alternate schemes will be considered. The first alternative is to perform several simulations of the estimation of $\hat{\alpha}$, using different trials of the noise. The variation of α in the simulations is then used as an estimate of the formal error. This approach was used by Tan and Gull (1985) for a similar problem. A second approach is to ignore correlation when solving for α and β, but to include the effects of correlation when estimating the formal errors. These approaches are considered further in § IV.

iii) *Wiener Filters*

Wiener filters are a general technique for obtaining conservative estimates of a signal, given a distorted and noisy version of it (e.g., Davenport and Root 1958). For simplicity we ignore leakage and sum along the spectral dimension only. The Wiener filter approach assumes the signals, as well as the noise, can be modeled as stochastic processes. Let the

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

autocorrelation function of the $\hat{I}$ spectrum be $r_I(i)$. We formulate our problem as that of finding linear operators (filters), which, given the noisy $I_{\rm obs}$ and $V_{\rm obs}$ data, form an estimate of the noiseless $\hat{I}$. These filters are chosen so that the expected value of the error between the true and estimated value of $\hat{I}$ is minimized. Specifically, we find filters h_I and h_V which minimize

$$\mathrm{E}\left[\left(\hat{I}-(h_I * I_{\rm obs}+h_V * V_{\rm obs})\right)^2\right],$$

where "$*$" represents convolution. Once this is achieved, we can use a least-squares algorithm to determine an estimate of $\hat{\alpha}$ from the estimate for $\hat{I}$ and $V_{\rm obs}$. This leads to an iterative algorithm, because we must know α to determine the filters. It can be shown that the Fourier transforms of the optimum Wiener filters are given by

$$H_I=\frac{1}{1+\alpha^2 D^* D+\sigma^2 R_n/R_I},$$

$$H_V=\frac{\alpha D^*}{1+\alpha^2 D^* D+\sigma^2 R_n/R_I},$$

where R_n and R_I are the power spectra corresponding to autocorrelation functions r_n and r_I, and D is the Fourier transform of the derivative operator.

The advantage of this algorithm is that it allows simpler treatment of correlation of the noise processes. In addition, it adds information about the autocorrelation of the signal. Even though we do not know the autocorrelation of I, we generally believe I is a "smooth" function. We can add this *a priori* information by using, for the I spectra autocorrelation function, a Gaussian whose width is typical of widths of features in the I spectra.

Wiener filters are intended to give conservative, rather than unbiased, estimates. Indeed, as the signal-to-noise ratio approaches zero, the Wiener filter estimate of $\hat{I}$ will also approach zero. This leads to the estimate of $\hat{\alpha}$ being biased on the high side.

IV. EVALUATION OF TECHNIQUES

In this section, we use numerical simulation to examine the performance of the estimators discussed in § III and the appendices. Since we cannot duplicate the infinite variety of line profiles encountered in astronomy, we content ourselves with the Gaussian. Therefore, all our tests were conducted with a single line of Gaussian shape and should yield indicative results. Independent noise in the $I_{\pm}$ spectra was generated by summing 12 samples from a uniform random number generator (Park and Miller 1988). According to the central limit theorem, this sum approximates a Gaussian random variable. Unless explicitly stated, our tests did not involve a leakage term nor any correlation of the noise. Finally, all quantities which express a spectral width (e.g., the splitting α and line FWHM) are given in units of channels.

The large computations were carried out on the Cray-2 running UNICOS at the National Center for Supercomputing Applications in Champaign-Urbana, Illinois. Smaller trials were undertaken on the Astronomy Department microVAX running VMS at the University of Illinois.

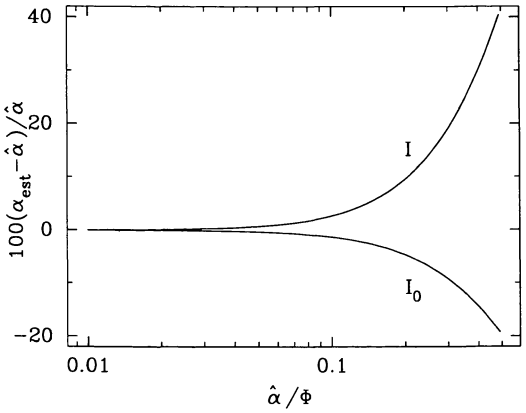

FIG. 2.—Percentage error in $\alpha_{\rm est}$ vs. the ratio of the splitting to the line FWHM for the maximum likelihood method and a noiseless, well-sampled line. Both the "observed" I and the "unobserved," unsplit I_0 spectra are shown.

a) *Breakdown of the Small $\delta\nu$ Approximation*

We first establish the regime in which the fundamental small splitting assumption is valid. Figure 2 shows the percentage error in $\alpha_{\rm est}$ versus $\hat{\alpha}/\Phi$ (Φ is the line FWHM). The noiseless line is well-sampled with $\Phi = 40$ channels. We show the results from both the unsplit I_0 and split I spectra, although in practice, only the I spectrum is actually measured. From Figure 2 we see the small $\delta\nu$ assumption fail in two ways. First, even if the I_0 spectrum is used, the result begins to deteriorate significantly at $\hat{\alpha}/\Phi \approx 0.1$ because the assumption that $I_0(\nu+\delta\nu)-I_0(\nu-\delta\nu)\approx 2\delta\nu(dI_0/d\nu)$ fails (see § II). Second, additional error is incurred because the assumption that $I = I_0$ also becomes poor.

This test was conducted with the maximum likelihood technique and a one-sided derivative. However, the position of the breakdown point does not depend significantly on either of these choices. This is because the line is noiseless (see § IV*c*) and well-sampled (see § IV*b*).

In summary, the small splitting assumption is valid only when the condition $\hat{\alpha}/\Phi < 0.1$ is satisfied. This result is subject to some variation depending on the actual line shape, but should be indicative.

b) *The Effect of Sampling*

We investigate the effect of sampling on a noiseless line with the one- and two-sided derivatives. We choose $\hat{\alpha}/\Phi = 0.025$ so that the small splitting assumption is satisfied. Figure 3 displays the percentage error in $\alpha_{\rm est}$ versus Φ for the maximum likelihood technique. Examination of Figure 3 suggests that a minimum of $\Phi \approx 10$ channels is necessary to obtain a representative derivative of the line and that the two-sided derivative makes a larger error (as expected) when the sampling is poor. This result is also subject to some variation depending on the actual line shape.

c) *The Bias in the Estimators*

Having established the conditions that satisfy the small splitting and good sampling criteria, we now compare each statistical method to evaluate their relative biases. Appen-

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

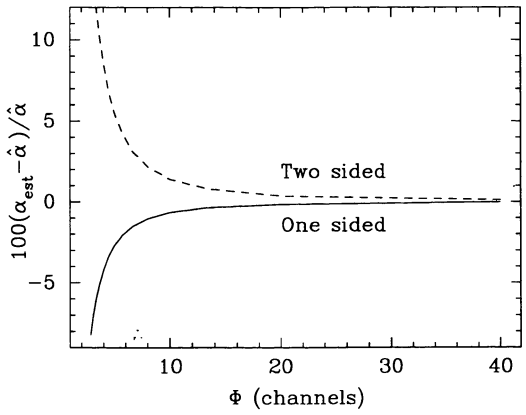

FIG. 3.—Percentage error in α_{est} vs. the line FWHM, for the maximum likelihood method, and a noiseless line (small splitting criterion satisfied). Both the one- and two-sided derivatives are shown.

dices A and B give more detailed discussion on the source of such biases.

For the least-squares technique, $\bar{\alpha}_{est}$ (the expected value of the estimate of $\hat{\alpha}$) is given by a simple formula (eq. [3.4]). We define the bias by

$$\frac{\bar{\alpha}_{est} - \hat{\alpha}}{\hat{\alpha}} = \frac{-a}{\hat{\eta}^2 + a}, \tag{4.1}$$

where, as before, a is 2 or 1/2 for the one- and two-sided derivatives, respectively, and $\hat{\eta}$ is the rms signal-to-noise ratio (see eq. [3.2]) in the $d\hat{I}/d\nu$ spectrum. Unfortunately, an expression such as equation (4.1) cannot be derived for the maximum likelihood or Wiener filter methods (see Appendix B). Therefore we start with a specific case and compare all three estimators. We choose $\hat{\alpha}/\Phi = 0.025$ (small splitting satisfied), a well-sampled line ($\Phi = 30$), and the number of channels $N = 128$, so that all channels effectively contain signal.

The solid dots on the upper panel of Figure 4 show the percentage errors in $\bar{\alpha}_{est}$ as a function of $\hat{\eta}$ (as the least-squares bias error is a function of $\hat{\eta}$) for each of our estimators. For the maximum likelihood and least-squares techniques, the results obtained with both the one- and two-sided derivatives are shown (labeled "1" and "2"). Each point on the plot is the average of 2000 different trials (i.e., the $I_{\pm}$ spectra were computed 2000 times with the same noise variance, but different and independent noise samples). For least squares, the solid lines generated from equation (4.1) verify that experiment and theory are in excellent agreement for the least-squares technique. The dash and dot-dash lines just join up the points for the maximum likelihood and Wiener filter methods. It is evident that all the estimators are unbiased for large $\hat{\eta}$ but that they all become progressively more biased as $\hat{\eta}$ decreases. Clearly the maximum likelihood estimate is the least biased, with the least squares and Wiener filter approaches about equally and oppositely biased for small $\hat{\eta}$.

The upper panel of Figure 4 also shows that one-sided derivative provides a more biased estimate than the two-sided derivative for noisy data (when the small splitting and good sampling criteria are met). This is because the two-sided derivative provides a less noisy estimate of $d\hat{I}/d\nu$.

The errors for the mean estimates are small (a few percent at the most) compared with the bias, so that it is a true bias, not just a statistical fluctuation in the mean. This is true of all the simulations that we present below so that errors in the mean estimates will not be discussed further.

In the upper panel of Figure 4, we plotted the abcissa as $\hat{\eta}$. However, this is not a directly observable quantity because it is noiseless. It can be shown that $\hat{\eta}$ is related to its noisy equivalent by

$$\eta^2 = E\left[\frac{1}{N\sigma^2}\sum_i \left(\frac{dI_{obs,i}}{d\nu}\right)^2\right] = \hat{\eta}^2 + a. \tag{4.2}$$

Thus, the lower panel of Figure 4 redisplays the bias in $\bar{\alpha}_{est}$ as a function of η (determined by eq. [4.2]) for the two-sided derivative results. In an actual observation, one would estimate η from the I_{obs} spectrum (eq. [4.2] would not hold exactly because there would be only one trial of the noise). Note that $\eta = \sqrt{a}$ for the case of pure noise so that the points bunch up at the low signal-to-noise end of the plot (we have also excluded the highest signal-to-noise ratio point of the upper panel in the lower panel). Although Figure 4 illustrates a special case, the maximum likelihood estimate should always be less biased than that of least-squares (see Appendix B).

Further least-squares simulations, varying N from 40 to 200, and $\hat{\alpha}$ from 0.1 to 2, show that the bias is in good agreement with equation (3.4). Because we have no expression for the maximum likelihood function bias, we determine its dependence on $\hat{\alpha}$ and N numerically. Hence, Figure 5 illustrates the percentage error in $\bar{\alpha}_{est}$ for the two-sided derivative maximum likelihood technique for various values of $\hat{\alpha}$. For reference, the least-squares bias curve is also shown. Again, the abcissa is shown as either $\hat{\eta}$ (*upper panel*) or η (*lower panel*). We have varied the number of noise trials in order to keep the relative error in $\bar{\alpha}_{est}$ small. Hence, for $\hat{\alpha} = 2$, 1.25, and 0.75 we used 500, 1300, and 3500 trials of the noise, respectively. It is computationally prohibitive to investigate thoroughly significantly smaller values of $\hat{\alpha}$ in this manner. We would need some 2×10^5 trials of the $I_{\pm}$ spectra for $\hat{\alpha} = 0.1$. Even with the Cray-2, this is a huge computational load. Thus, we have computed just four points for $\hat{\alpha} = 0.1$, but with only 10^4 trials. The curve is therefore noisier, but should still be representative. We used $N = 160$ and $\Phi = 40$, so that the small splitting and good sampling criteria are met.

From Figure 5 it is clear that the bias decreases with increasing $\hat{\alpha}$. This trend should continue until the small splitting criterion fails. We believe that as $\hat{\alpha}$ approaches zero, the maximum likelihood bias converges to the least-squares bias curve. However, we do not have the computational resources at this time to investigate values of $\hat{\alpha}$ that are orders of magnitude lower. Thus, we must be content that the trend from the trials we do have, is in the correct sense for this convergence.

The variation of the percentage error in $\bar{\alpha}_{est}$ with the number of channels for the two-sided derivative maximum likelihood method is displayed in Figure 6*a*. Again, we plot

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

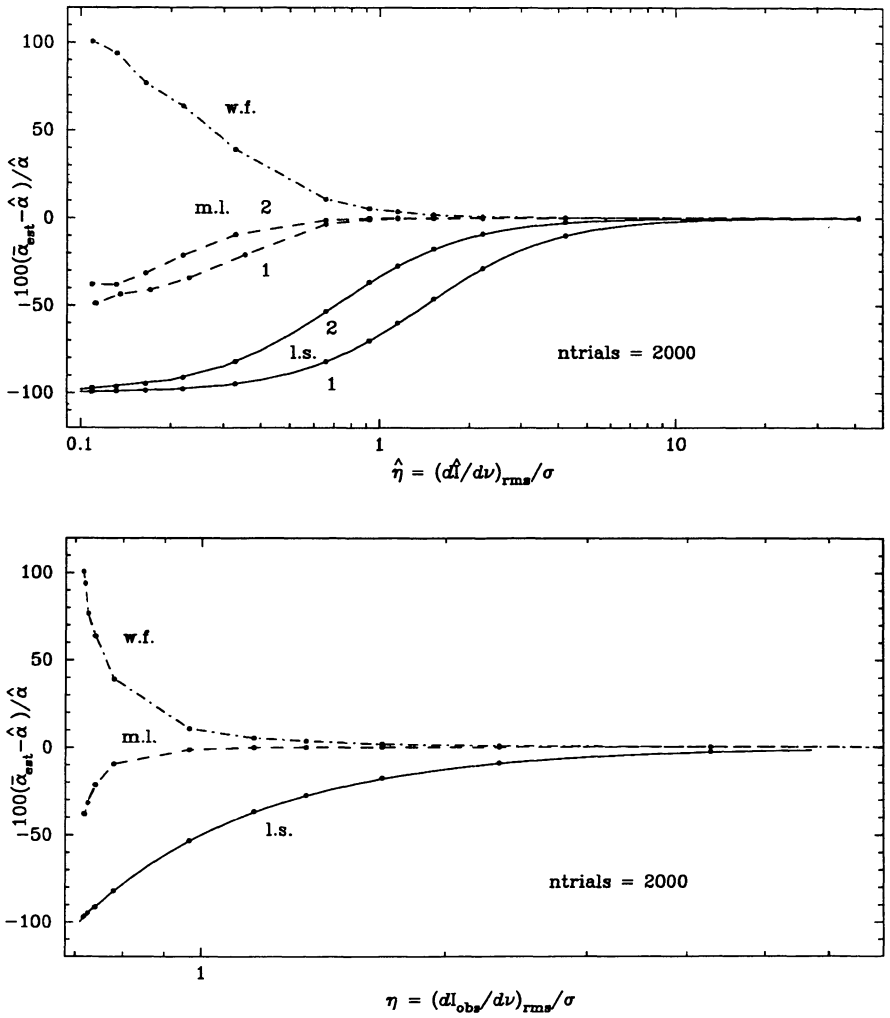

FIG. 4.—Percentage error in $\bar{\alpha}_{est}$ (from 2000 trials) vs. the noiseless (*upper panel*) and noisy (*lower panel*) rms signal-to-noise ratio in the I derivative spectrum of a line for which $\Phi = 30$, $\hat{\alpha} = 0.75$, and $N = 128$. The least-squares, maximum likelihood, and Wiener filter results are all shown. For the upper panel, both two- and one-sided derivative results are shown; the solid lines are the theoretical least-squares curves (eq. [4.1]), and the dash and dot-dash lines just join up the points. For the lower panel, just two-sided derivative results are shown.

the abcissa as $\hat{\eta}$ (*upper panel*) and η (*lower panel*), and the four curves correspond to $N = 160$, 120, 80, and 40. In addition, for this test, we use $\hat{\alpha}/\Phi = 0.025$ (small splitting satisfied), $\Phi = 20$ channels (good sampling satisfied), and 5000 trials. For $N = 80$, essentially all of each spectrum contains signal. For $N = 40$, not all of the line is included in the spectra, and for $N = 160$, the first and last 40 channels contain essentially noise. Fairly clearly, there is only a weak dependence of the maximum likelihood bias on the number of channels, within the range we have examined.

If one considers a single spectrum and just adds more and more noise channels, then η will decrease, and one moves obliquely from curve to curve on Figure 6*a*. This is presented slightly differently in Figure 6*b*, where we show the two-sided derivative maximum likelihood and least-squares bias versus the number of channels for the particular case of $\Phi = 20$, $\hat{\alpha} = 0.5$, and a constant noise variance. When $N > 80$, the additional channels contain essentially only noise because $\Phi = 20$. In addition, we label each point on the plot with its value of $\hat{\eta}$. This plot emphasizes the weak dependence of maximum likelihood on the number of channels and the need to exclude all noise channels when applying the least-squares technique.

For line shapes other than Gaussians, we find that the least-squares bias is still well-represented by equation (3.4). However, the maximum likelihood bias appears to be a function of line shape. We have found instances of the maximum likelihood technique being biased high. The magnitude of the bias, however, is of the same order, and the break point in $\hat{\eta}$, below which the estimate becomes significantly biased, remains approximately constant.

We now include a small leakage term ($\hat{\beta} = 0.05$) and examine the percentage error in $\bar{\beta}_{est}$, as well as $\bar{\alpha}_{est}$. We show only the maximum likelihood results in Figure 7 for $\hat{\alpha}/\Phi = 0.025$, $\Phi = 40$, $N = 160$, and 1000 trials of the noise. The leakage was included in a somewhat *ad hoc* manner by simply adding a fraction of the noiseless I spectrum into the noisy V, rather than computing the relevant quantities in the

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

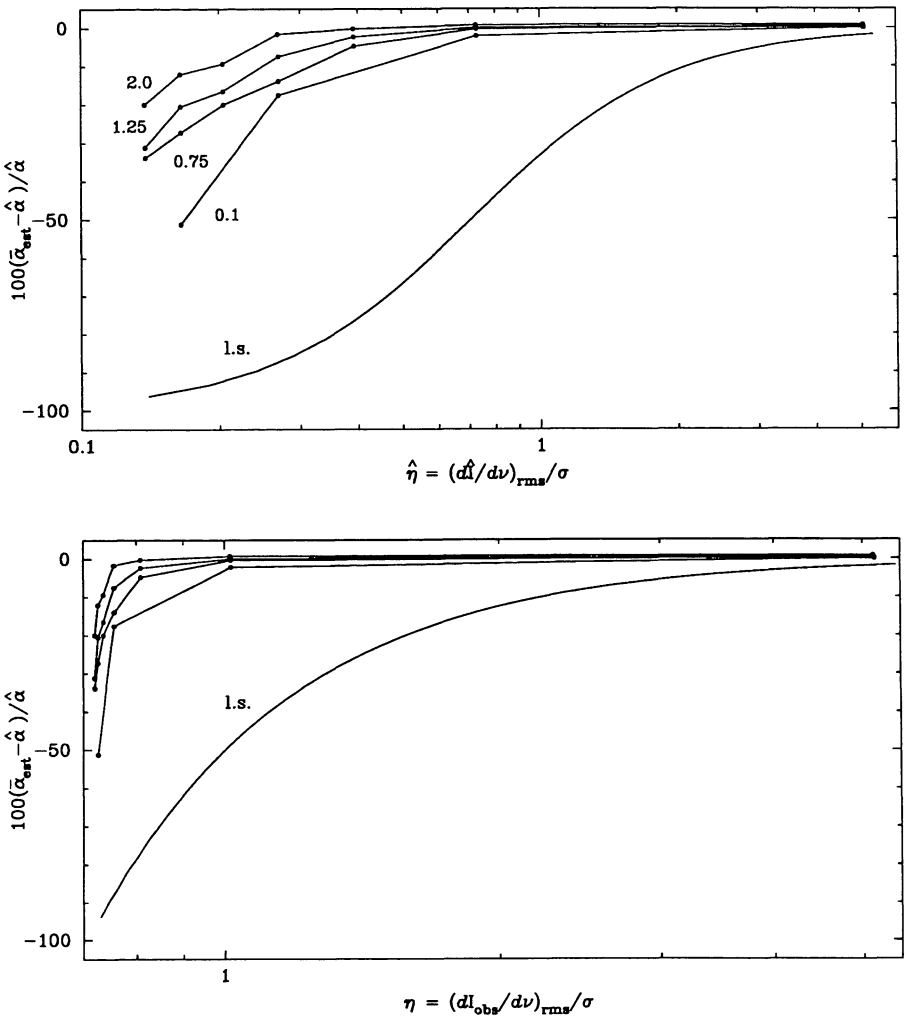

FIG. 5.—Percentage error in $\bar{\alpha}_{est}$ from the maximum likelihood method (two-sided derivative) vs. the noiseless (*upper panel*) and noisy (*lower panel*) rms signal-to-noise ratio in the I derivative spectrum of a line for which $\Phi = 40$, $N = 160$, and the splitting is varied. For $\hat{\alpha} = 2.0$, 1.25, 0.75, and 0.1 the number of noise trials was 500, 1300, 3500, and 10000 in order to keep the relative error roughly constant from curve to curve. The least-squares curve is shown for reference.

$I_{\pm}$ spectra (this would also affect I). However, for small leakages, these factors are fairly unimportant. This plot is interesting because $\bar{\beta}_{est}$ remains unbiased even when $\bar{\alpha}_{est}$ is significantly wrong. Although we have not shown it, least squares also extracts an essentially unbiased value of $\bar{\beta}_{est}$ for small leakage terms.

As a special case, we examine the maximum likelihood and least-squares results for $\hat{\alpha} = 0$. According to equation (3.4) and Appendices A and B, the expected value of α_{est} should be zero for both techniques. This is verified by Figure 8, which shows the average results for 500 trials of the noise.

The main conclusion that can be drawn from this section is that one can obtain an essentially unbiased estimate of $\hat{\alpha}$ and $\hat{\beta}$ with the maximum likelihood method, for η as low as about unity if $\hat{\alpha} > 0.1$. For smaller values of $\hat{\alpha}$, this lower limit for η will increase, but we cannot say how rapidly. For the least-squares method, it is necessary that $\eta > 10$ in order that the estimate be essentially unbiased; there is no dependence on $\hat{\alpha}$ in this case. We have not thoroughly investigated the Wiener filter method. However, initial results indicate that for small values of η, it is as biased (but in the opposite sense) as least squares.

d) The Effect of Noise on the Shape of χ^2

We have shown that the maximum likelihood estimate becomes biased as the signal-to-noise ratio decreases. Although our algorithm (see Appendix B) always finds the global minimum of χ^2, that minimum is not always at $\alpha_{est} = \hat{\alpha}$, owing to the presence of error in its estimate of $\hat{I}$. To give a qualitative feel for this, Figure 9 shows a normalized χ^2 as a function of α for six different values of $\hat{\alpha}/\sigma_{\alpha}$. As usual, the line is well-sampled ($\Phi = 40$ and $N = 160$) and satisfies the small splitting criterion with $\hat{\alpha} = 1.0$. Clearly, in the top three cases, the global minimum is at the correct value, $\alpha = \hat{\alpha}$. However, in the bottom three cases, the wrong solution is obtained. It should be noted that the same samples of the random noise were used in generating each of these

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

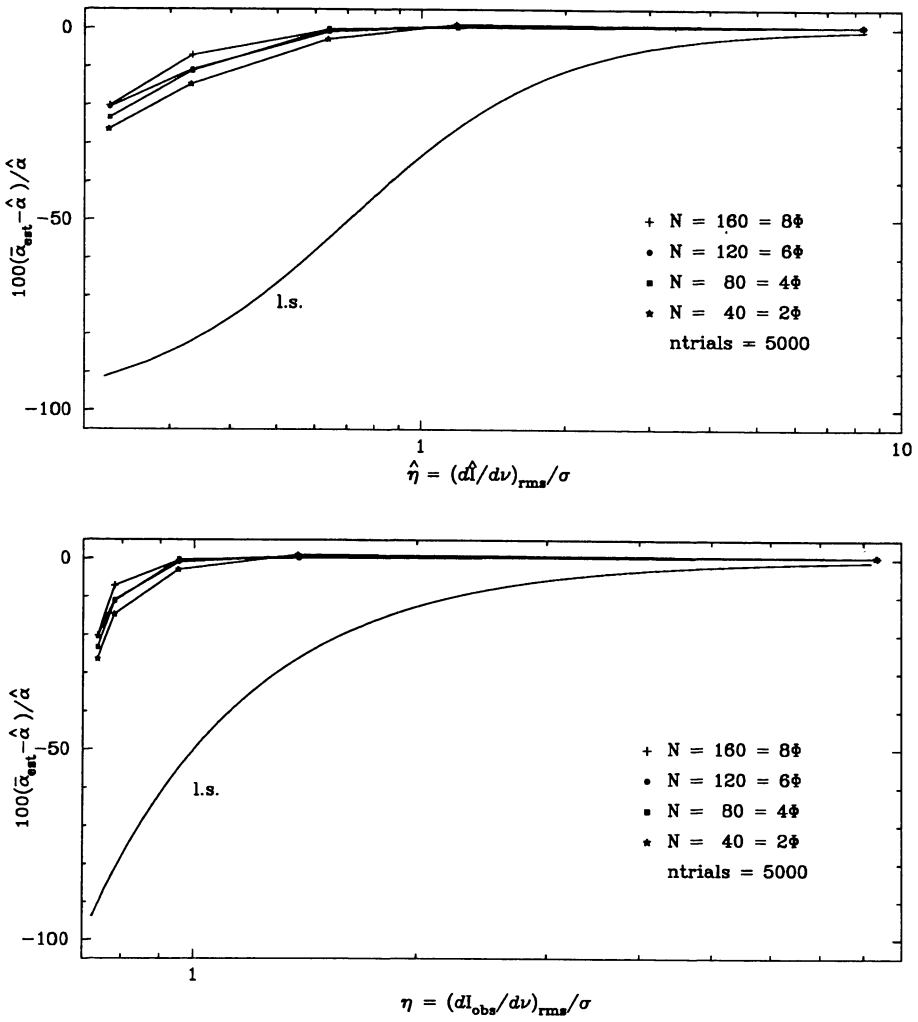

FIG. 6*a*

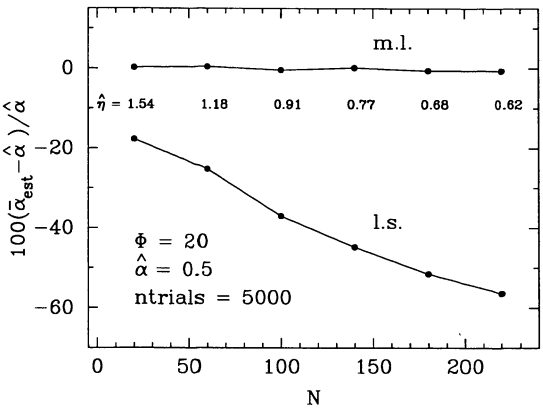

FIG. 6*b*

FIG. 6.—(*a*) Percentage error in $\bar{\alpha}_{est}$ from the maximum likelihood method (two-sided derivative) vs. the noiseless (*upper panel*) and noisy (*lower panel*) rms signal-to-noise ratio in the I derivative spectrum of a line for which $\Phi = 20$, $\hat{\alpha} = 0.5$ and $N = 160$, 120, 80, and 40. There were 5000 trials of the noise, and the least-squares curve is shown for reference. (*b*) Percentage error in $\bar{\alpha}_{est}$ from the two-sided derivative maximum likelihood and least-squares methods vs. the number of channels for a line with $\Phi = 20$, $\hat{\alpha} = 0.5$, and a constant noise variance. Each point is labeled with the corresponding value of $\hat{\eta}$, and there were 5000 trials of the noise. For $N > 80$, the extra channels are essentially noise.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

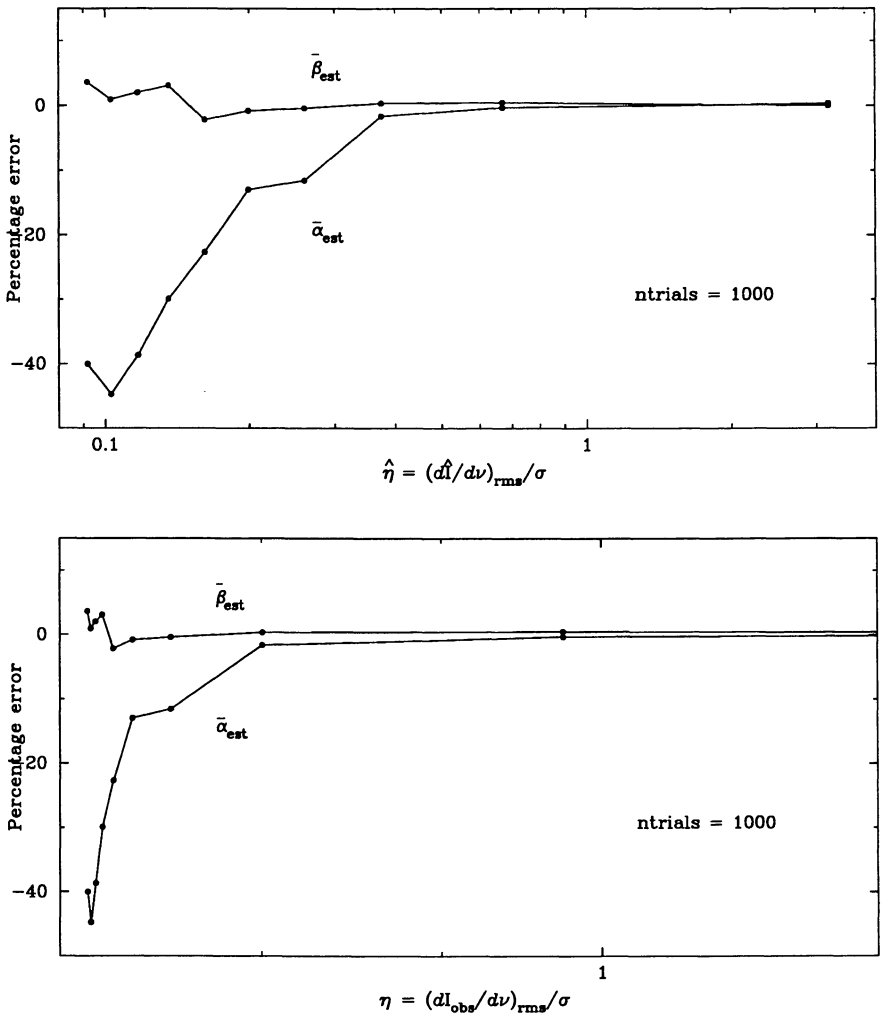

FIG. 7.—Percentage error in $\bar{\alpha}_{est}$ and $\bar{\beta}_{est}$ from the maximum likelihood method (two-sided derivative) vs. the noiseless (*upper panel*) and noisy (*lower panel*) rms signal-to-noise ratio in the I derivative spectrum of a line for which $\Phi = 40$, $\hat{\alpha} = 1.0$, $\hat{\beta} = 0.05$, and $N = 160$. There were 1000 trials of the noise.

plots, so that as the signal-to-noise ratio decreases, the same spurious feature grows until it dominates the location of the minimum.

It is also instructive to examine a few trials of the noise with the same value of $\hat{\alpha}/\sigma_{\alpha}$. Figure 10 shows nine trials for the case of $\hat{\alpha}/\sigma_{\alpha} = 4.3$ and the other parameters as before. We see that the global minimum is in roughly the correct place for six of the cases. For the remaining three cases (*top center*, *bottom left*, *bottom right*) the answer is clearly wrong and the bias is such that there are more wrong results on the low side than the high.

e) Debiasing Least Squares

In § IV*c* above we showed that maximum likelihood makes superior estimates of $\hat{\alpha}$ compared to least squares. However, it is possible to partially debias the least-squares estimate (see Appendix A). This is not true of maximum likelihood. Thus, before we dismiss least squares as inferior to maximum likelihood, we must compare the debiased least-squares results with those of maximum likelihood. Debiasing should be used with caution, as any error in the value of σ^2 used in the debiasing will cause the bias to be incompletely removed. In the simulations that are presented in this section, we have "cheated," in that we know the exact value of σ^2.

First we reexamine the cases demonstrated in Figure 4, for the two-sided derivative. The mean debiased least-squares estimate from 2000 trials is shown in Figure 11*a*, together with the maximum likelihood and biased least-squares results. The lower panel excludes the five lowest $\hat{\eta}$ points from the upper panel for greater clarity. Once again recall that, while the least-squares bias is independent of $\hat{\alpha}$, this is not so with maximum likelihood. The least-squares curves are universal, but the maximum likelihood curve is dependent on $\hat{\alpha}$ as well as $\hat{\eta}$. This plot is an indication that the mean debiased least-squares estimate is essentially as good as the maximum likelihood result when $\hat{\eta} > 1$. Below this point, the debiased least-squares result rapidly deteriorates. Note that debiasing extends the range in which an unbiased least-

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

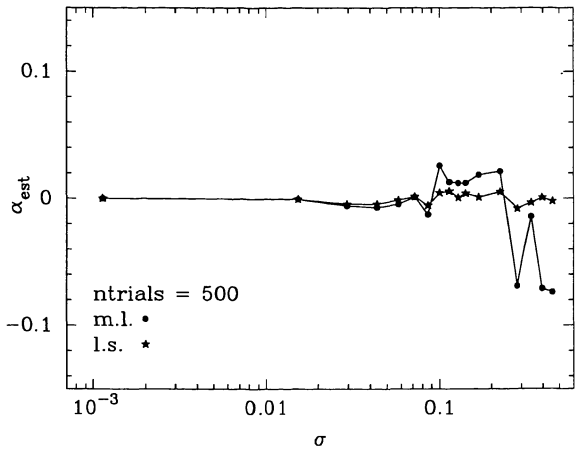

FIG. 8.—Two-sided derivative maximum likelihood and least-squares estimates of $\hat{\alpha}$ vs. the noise in I (and V) for a line with $\hat{\alpha} = 0.0$, $\Phi = 40$, $N = 160$, and a peak $\hat{I}$ intensity of 1.0. There were 500 trials of the noise.

squares estimated of $\hat{\alpha}$ can be obtained by approximately an order of magnitude.

Further simulations (varying $\hat{\alpha}$ from 0.1 to 2, N from 40 to 200, and line shape) show that $\bar{\alpha}_{est}$ is essentially independent of $\hat{\alpha}$ and N, for $\hat{\eta} > 1$. This is not true for $\hat{\eta} < 1$.

We conclude our discussion of debiased least squares by examining some distributions of α_{est} and comparing them to those of maximum likelihood. It is pointless to debias a least-squares estimate if the error variance of the estimate would be much larger than that of maximum likelihood. The results of the appendices indicate that the debiased least-squares and maximum likelihood techniques should have approximately the same variance. Let us examine a case where $\hat{\eta} = 1.0$; this is about the limit below which both maximum likelihood and debiased least squares fail. We examine distributions for $\hat{\alpha} = 1.0$ and 0.1, choose $\Phi = 40$, $N = 160$, and use the two-sided derivative. Two thousand trials of the noise were used.

Figure 11*b* shows histograms of the distributions of the percentage error in α_{est}. The top panels are for $\hat{\alpha} = 1.0$, and

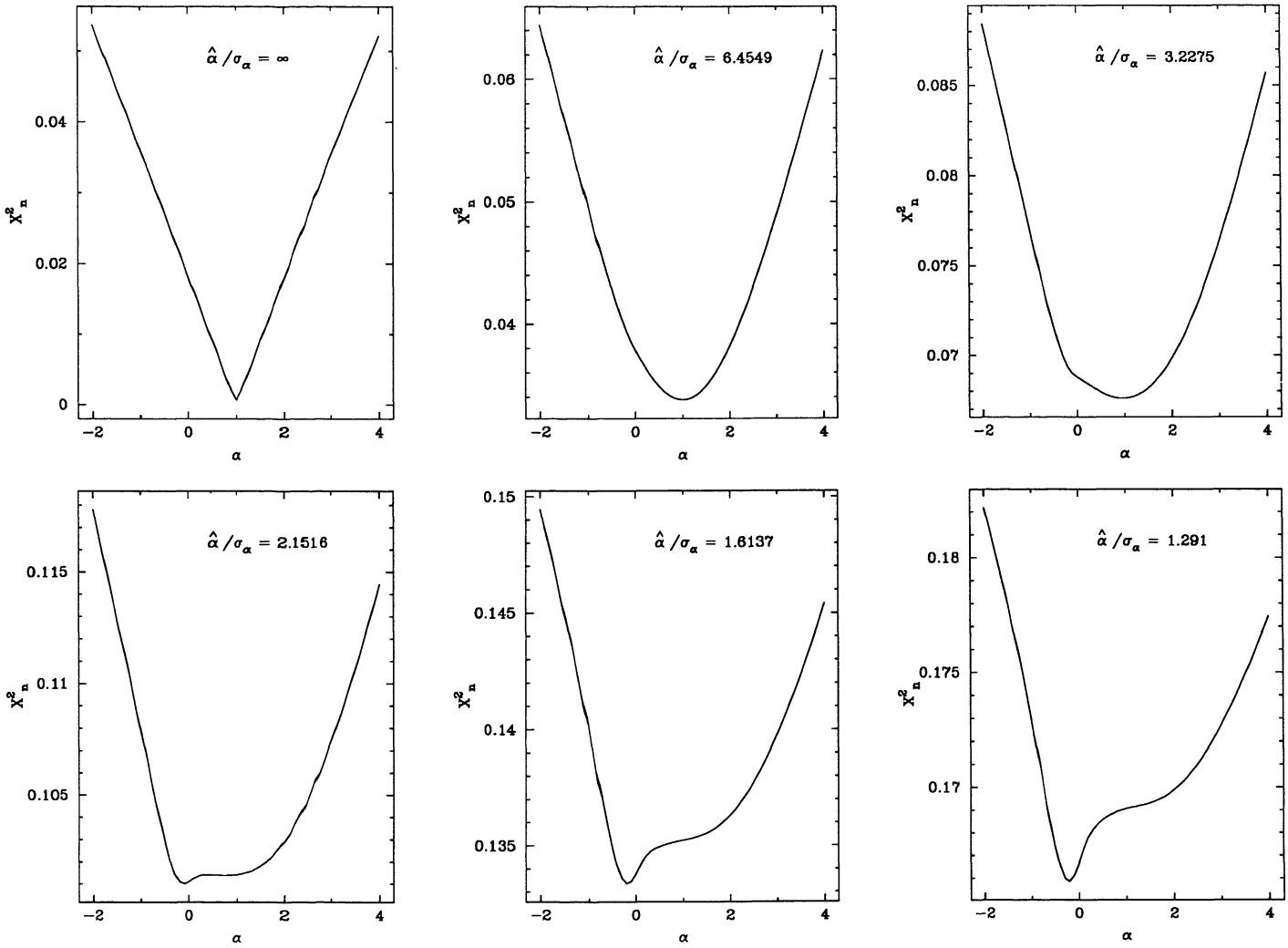

FIG. 9.—Normalized χ^2 for the maximum likelihood method (two-sided derivative) vs. α as the signal-to-noise ratio ($\hat{\alpha}/\sigma_\alpha$) varies from infinity to unity. The line has $\hat{\alpha} = 1.0$, $\Phi = 40$, and $N = 160$. The same samples of the random noise were used for each plot, so that as $\hat{\alpha}/\sigma_\alpha$ decreases, the same spurious feature grows until it dominates the location of the minimum.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

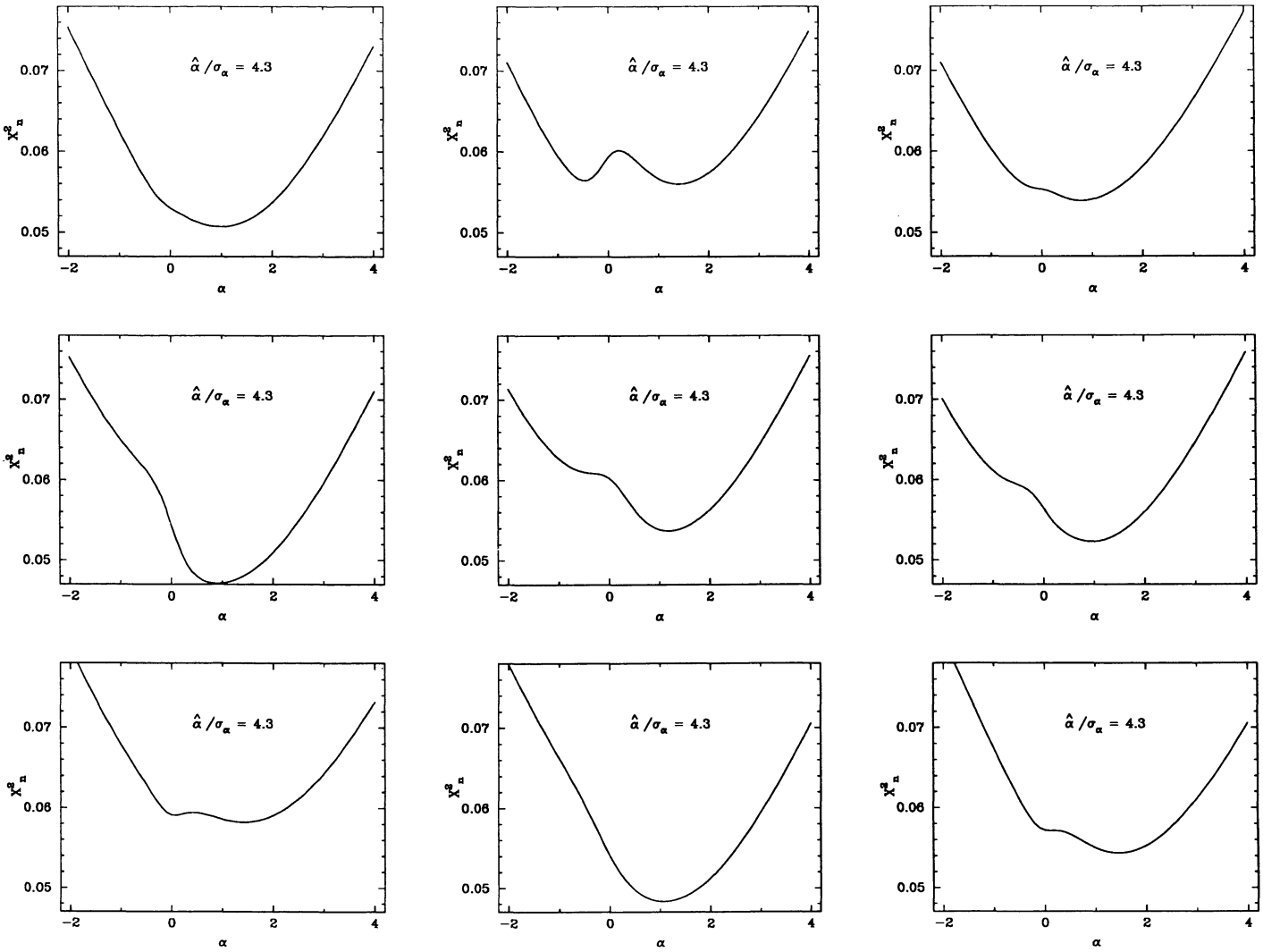

FIG. 10.—Normalized χ^2 from the maximum likelihood method (two-sided derivative) vs. α for 6 different trials of the noise in a line with $\hat{\alpha}/\sigma_\alpha = 4.3$, $\hat{\alpha} = 1.0$, $\Phi = 40$, and $N = 160$.

the bottom panels are for $\hat{\alpha} = 0.1$. Each panel shows the results for least-squares, debiased least-squares and maximum likelihood. For the case of $\hat{\alpha} = 1.0$, Figure 11*b* shows that not only does least-squares yield a wrong result (mean is -32%), there is little probability of obtaining the correct result. The debiased least-squares shows a distribution which is clearly no longer biased, and the standard deviation is small, so reliable estimation is possible. The maximum likelihood distribution is only modestly better than the debiased least squares. There is little to choose between the maximum likelihood and debiased least-squares distributions.

We now turn to the lower panels of Figure 11*b*, where $\hat{\alpha} = 0.1$. The biases show negligible change, but the widths of the percentage error distributions have increased. The debiased least-squares and maximum likelihood distributions are essentially indistinguishable.

In summary, debiased least-squares and maximum likelihood yield comparable results until $\hat{\eta}$ is lower than about unity. Below this value, one obtains more reliable (but still biased) estimates with maximum likelihood.

f) *The Reliability of the Error Estimates*

i) *Uncertainty in* α

We have demonstrated so far that in a fairly well-defined part of parameter space, we can extract unbiased estimates of $\hat{\alpha}$ and $\hat{\beta}$. Unfortunately, in practice, we only get one chance to estimate these parameters, so we *must* have a good estimate of the uncertainties or our results are useless. We investigate the reliability of our error estimates for the maximum likelihood method with two representative cases in Figure 12. The upper and lower rows of panels are for a moderately good and a poor signal-to-noise ratio case. As usual, we make sure the small splitting and good sampling criteria are well met ($\hat{\alpha} = 1.0$ and $\Phi = 40$). In addition, we use $N = 160$ and the two-sided derivative.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

Consider first the good signal-to-noise ratio case shown in the upper row. The leftmost panel shows typical I_{obs} and $dI_{obs}/d\nu$ spectra, and we give (in the top left corner) the signal-to-noise ratio parameter $\hat{\eta} = 2.6$ (see eq. [3.2]). Referring to Figures 4 and 5, we see that in this case the estimate of $\hat{\alpha}$ should be virtually unbiased. The second and third panels show histograms of the percentage error of α_{est} and the estimated standard deviation of α_{est} (as a percentage of $\hat{\alpha}$) from 1000 trials of the noise. Both the standard deviation of the distribution for α_{est} and the mean of the distribution for σ_α are 3.1%.

We now turn to the lower row of panels in Figure 12, a case of greater noise corruption such that $\hat{\eta} = 0.5$. The experimental standard deviation of the distribution for α_{est} is 21.2%. The mean of the distribution of the estimated percentage standard deviations is 15.3%. In general, it appears that for poor signal-to-noise ratios, the error estimate seems to be biased low.

ii) *Uncertainty in* β

We must also examine our error estimates for the leakage parameter β. Thus, Figure 13, in the manner of Figure 12, shows representative spectra (I_{obs} and V_{obs} in this case) and histograms of the β_{est} and estimated σ_β distributions from 1000 trials of the noise. The actual leakage term is 5%. We see that the standard deviation of the distribution for β_{est} is in good agreement with the mean of the distribution for σ_β in both the moderate and poor signal-to-noise ratio cases, so that we are confident in our ability to supply meaningful error estimates for β.

In summary, we can supply good estimates of the standard error in α_{est} when the estimate of $\hat{\alpha}$ is not biased. As α_{est} becomes more biased, so does the error estimate, and in the same sense. We have not explored large leakage terms, but for small leakages, the estimated standard error in β_{est} is reliable.

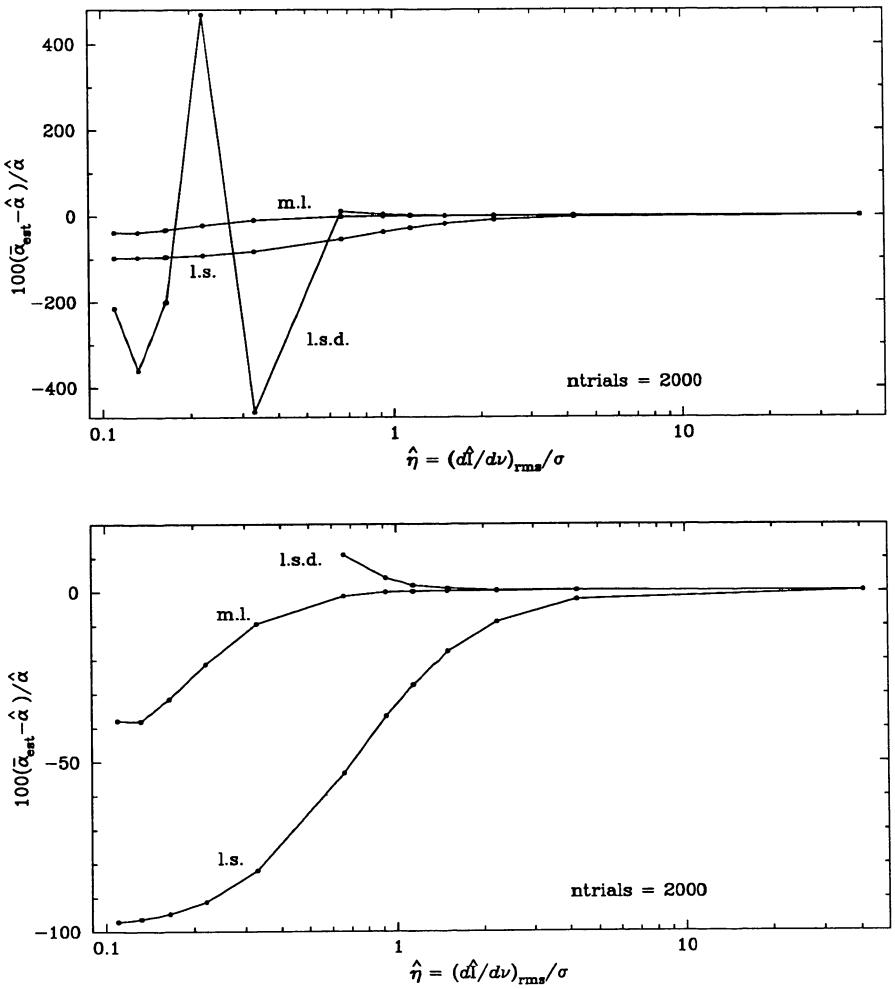

FIG. 11*a*

FIG. 11.—(*a*) Percentage error in $\bar{\alpha}_{est}$ (from 2000 trials) vs. the noiseless rms signal-to-noise ratio in the I derivative spectrum of a line for which $\Phi = 30$, $\hat{\alpha} = 0.75$, and $N = 128$. The two-sided derivative least-squares, debiased least-squares, and maximum likelihood results are shown. The lower panel excludes the five noisiest points for debiased least-squares. (*b*) Comparison between least-squares, debiased least-squares, and maximum likelihood distributions of the percentage error in α_{est} from 2000 trials of the noise for a line in which $\Phi = 40$, $N = 160$, and $\hat{\eta} = 1.0$. In the upper panels, $\hat{\alpha} = 1.0$, while in the lower panels $\hat{\alpha} = 0.1$. From left to right, the panels are for the least-squares, debiased least-squares, and maximum likelihood techniques, respectively.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

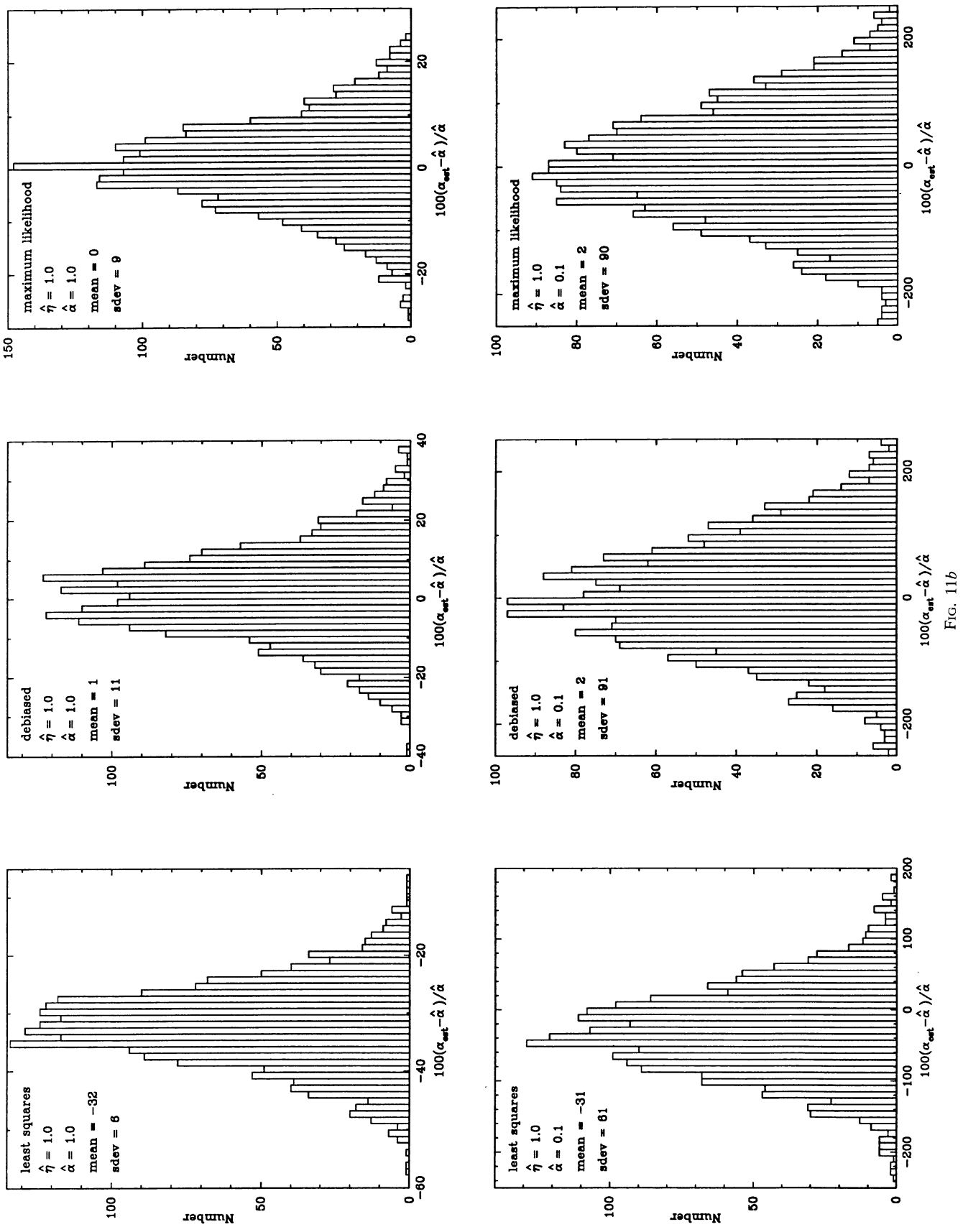

FIG. 11*b*

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

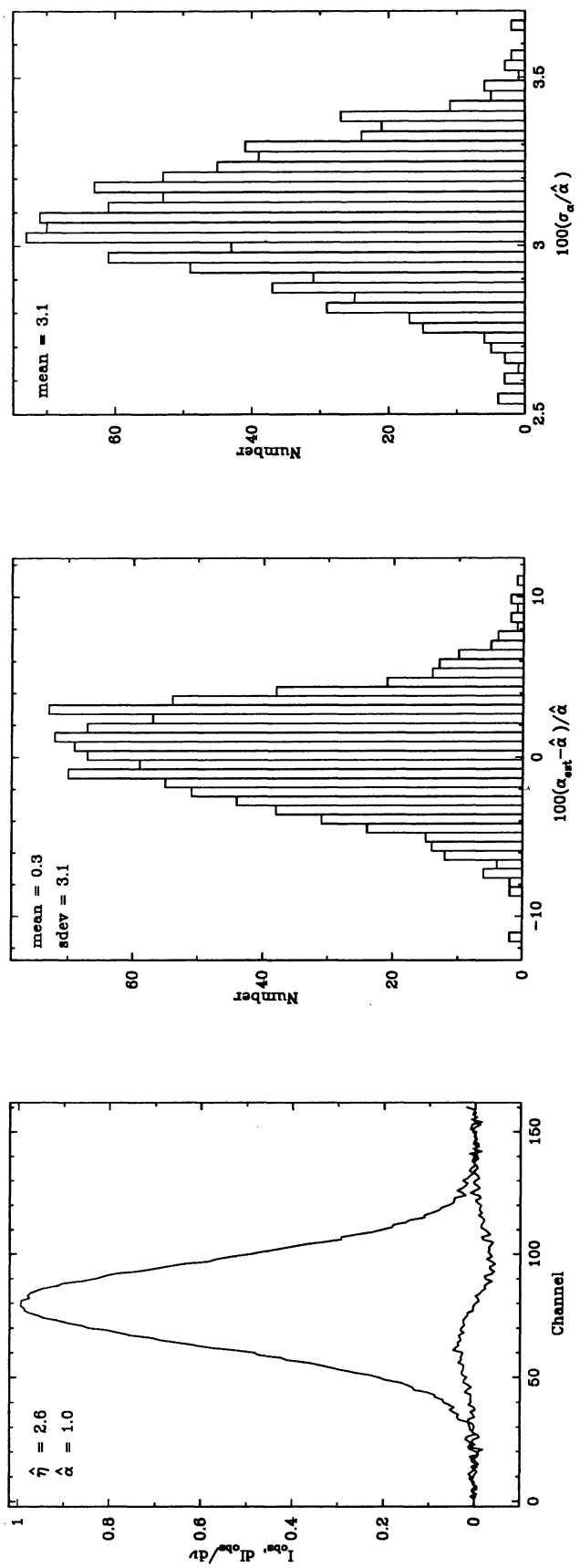

FIG. 12.—Reliability of the maximum likelihood (two-sided derivative) error estimates for $\hat{\alpha}$ demonstrated for two cases: a moderately good ($\hat{\eta} = 2.6$, *upper panels*) and a poor ($\hat{\eta} = 0.5$, *lower panels*) signal-to-noise ratio. The lines have $\hat{\alpha} = 1.0$, $\Phi = 40$, and $N = 160$. The leftmost panels show representative I and $dI/d\nu$ spectra, the middle panels show histograms of the percentage error in $\alpha_{\rm est}$, and the rightmost panels show histograms of the estimated standard deviations (as a percentage of $\hat{\alpha}$) in α from 1000 trials of the noise.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

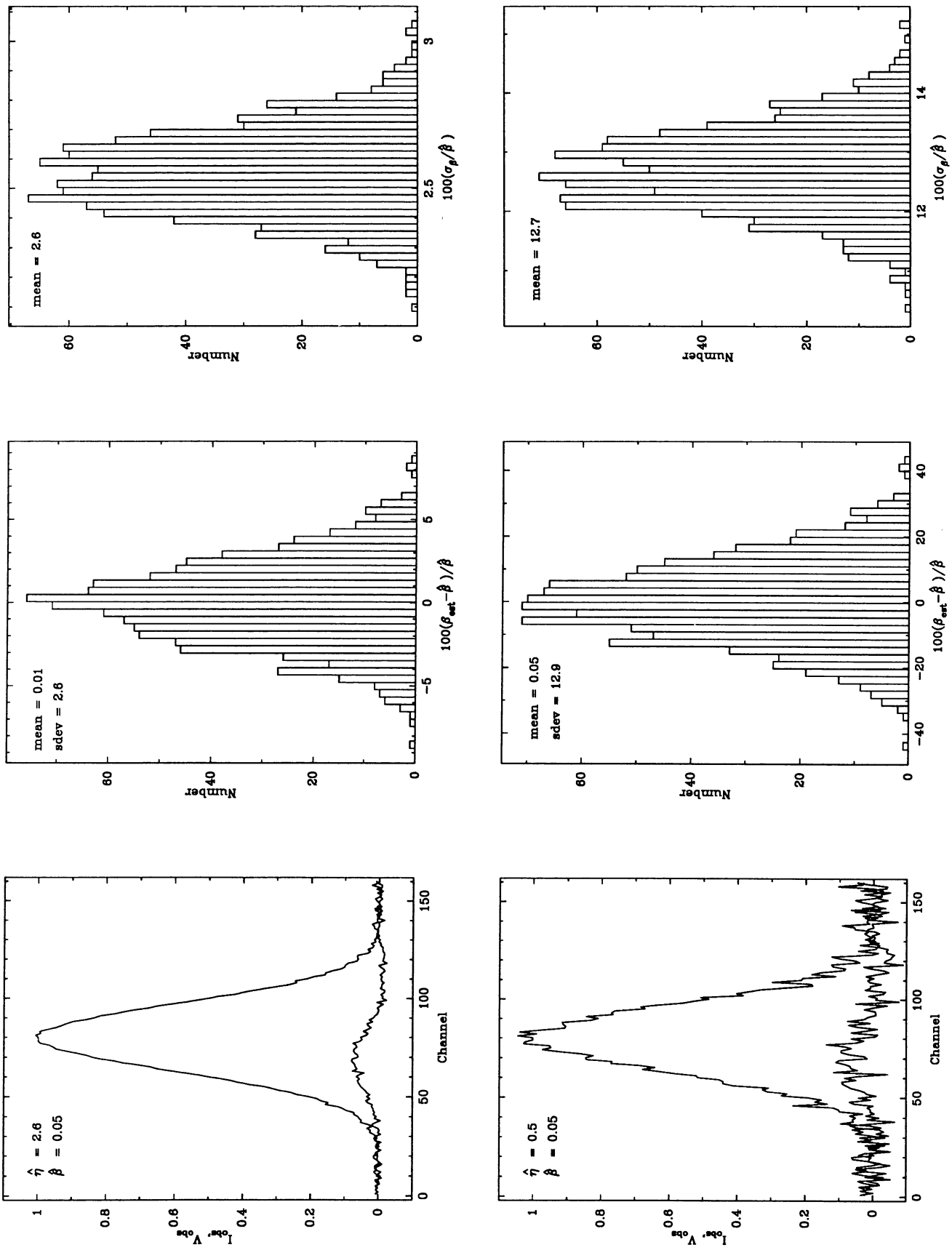

FIG. 13.—Reliability of maximum likelihood (two-sided derivative) error estimates for $\hat{\beta}$ demonstrated for two cases: a moderately good ($\hat{\eta} = 2.6$, *upper panels*) and a poor ($\hat{\eta} = 0.5$, *lower panels*) signal-to-noise ratio. The lines have $\hat{\alpha} = 1.0$, $\hat{\beta} = 0.05$, $\Phi = 40$, and $N = 160$. The leftmost panels show representative I and V spectra, the middle panels show histograms of the percentage error in β_{est}, and the rightmost panels show histograms of the estimated standard deviations (as a percentage of $\hat{\beta}$) in β from 1000 trials of the noise.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

g) The Effect of Spatial and Spectral Correlations

This section considers the effects of correlated noise on the maximum likelihood algorithm; the theory is discussed in Appendix B. Basically, noise correlation has little effect on the estimates of $\hat{\alpha}$ and $\hat{\beta}$, but has a significant effect when determining the uncertainty associated with these parameters. Noise correlation effects depend strongly on the autocorrelation functions, and so general results cannot be obtained. Rather we present results of some simulations which should be indicative. Two sets of simulations are discussed, one with a high and one with a moderate signal-to-noise ratio. All tests summed over 128 channels by 20 pixels, and used 100 trials of the noise.

Four cases of noise correlation are considered: no correlation, spectral correlation only, spatial correlation only, and both spectral and spatial correlation. The spectral correlation is such that there is only correlation between adjacent channels, for which the correlation coefficient is 1/6. This is the form of correlation which occurs if the spectral dimension is 3-point Hanning smoothed, and then every second channel is discarded. (This choice was motivated by such a correlation being present in some of our data.) The spatial correlation is that caused by a sinc-shaped beam, with FWHM of 4 pixels. To generate noise with the appropriate correlations, we convolved Gaussian independent noise. In doing so, we padded with extra noise samples (the number determined by the width of the convolving function) to avoid edge effects.

Three techniques for dealing with correlation are considered: full noise correlation treatment, partial treatment, and a simulation approach. By full treatment, we mean that the covariance matrix was included at all stages of the estimation algorithm (see Appendix B). We have not developed code to perform this full treatment when there is spatial correlation, and so we do not present results in these cases. The partial treatment approach ignores correlation in determining estimates of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{I}$, but correctly includes it when determining σ_α. The third approach, simulation, simply measures the standard deviation in $\alpha_{\rm est}$ when the experiment is simulated many times, with different trials of the noise.

The first result of the simulations is that the estimated values of $\hat{\alpha}$ for the full and partial treatments agree closely. Typically they do not differ by more than $\sigma_\alpha/10$. Thus, to good approximation, we can ignore the effects of noise in estimating $\hat{\alpha}$.

However, we cannot ignore correlation when estimating the standard error of the estimate. Tables 1 (high signal-to-noise ratio) and 2 (moderate signal-to-noise ratio) give the estimate of σ_α (expressed as a percentage of $\hat{\alpha}$) for the three techniques and for different forms of noise correlation. In the case of a high signal-to-noise ratio, the agreement between the estimated error, for the full and partial treatments, agrees well with the actual variation of $\alpha_{\rm est}$ seen in the simulations. However, the agreement is poorer when the signal-to-noise ratio is moderate. In particular, the agreement when there is spatial correlation, is quite poor. Here the estimated σ_α differs from that observed in the simulations by almost a factor of 2. Standard error estimates are consistently lower than the results of simulation. The bias seems to be more pronounced when correlations are present, particularly spatial correlations. We believe this is caused by the error in $I_{\rm est}$. In calculating error estimates for the correlated noise, this error can be amplified by the matrix operations involved. This is particularly true of spatial correlation, because the matrices involved are inevitably ill-conditioned (see Appendix B for more details). Indeed, we find error estimates that are grossly incorrect when the following three conditions are all satisfied: the signal-to-noise ratio is poor, spatial summing is being performed over a large region (5×5 pixels or greater), and the effect of spatial correlation is included in the error analysis.

When summing over large spatial regions, the true σ_α is probably no more than a factor of 2–6 times larger than the value of σ_α that would be calculated assuming no spatial correlation. This factor will be a function of the source structure and beam size and shape. When a better estimate of σ_α is needed, simulations of the estimation process can provide better bounds.

V. SUMMARY

We have evaluated the performance of three statistical methods in the analysis of Zeeman effect radio astronomical data: maximum likelihood, least-squares, and Wiener filters. We now summarize our findings, which are based upon

TABLE 1

EFFECTS OF NOISE CORRELATION ON σ_α: HIGH SIGNAL-TO-NOISE CASE

Correlation	Full Treatment	Corrected σ_α	Simulation
No correlation	0.36%	0.36%	0.36%
Spectral correlation	0.39%	0.39%	0.41%
Spatial correlation	...	0.47%	0.50%
Spectral and spatial correlation	...	0.54%	0.60%

TABLE 2

EFFECTS OF NOISE CORRELATION ON σ_α: MODERATE SIGNAL-TO-NOISE CASE

Correlation	Full Treatment	Corrected σ_α	Simulation
No correlation	2.6%	2.6%	3.2%
Spectral correlation	2.8%	2.8%	3.4%
Spatial correlation	...	3.2%	5.2%
Spectral and Spatial Correlation	...	3.3%	6.1%

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

numerical simulations with a single Gaussian line. These conclusions, although indicative, are of course subject to some fluctuation depending on the details of the line shape. We then suggest a recipe which should lead to a good estimate, if it can be made, of the splitting. The reader may wish to review the definitions of symbols, which are summarized in § III*b*.

a) Conclusions

We assume that the first two items in the following list are satisfied for all succeeding items. We find the following:

1. The fundamental Zeeman analysis equation (eq. [3.1]) rests upon the assumption that the splitting is small. The latter assumption requires that $\hat{\alpha}/\Phi < 0.1$ be approximately satisfied (recall that Φ is the line FWHM, measured in channels). Unfortunately, the observer has no control over this ratio.

2. To obtain a good numerical derivative, the line must be well-sampled. This requires that the condition $\Phi > 10$ be approximately satisfied.

3. The estimators we have investigated are biased when the signal-to-noise ratio is poor ($\hat{\eta} < 1$ for maximum likelihood, $\hat{\eta} < 10$ for least squares). The maximum likelihood method is the least biased, and therefore preferable. The only disadvantage of the maximum likelihood technique is that it is computationally more expensive and requires more coding effort.

Least-squares is the simplest technique in all regards, and it is possible to write down an expression for the bias in the least-squares estimate of the splitting. Thus, debiasing least squares is also a computationally attractive alternative. However, you *must* have a good estimate of the noise variance, σ^2, in order to do this successfully. Accurately debiased least squares provides unbiased results for $\hat{\eta} > 1$, the same range as for maximum likelihood.

4. If the splitting is constant (or approximately constant) over a region (both spectral and spatial), then the data for this entire region should be input to the estimation process to form a single set of estimates of the parameters. This leads to estimates with lower variances and is distinct, and preferable, to averaging the data before the estimation process. However, if the signal-to-noise ratio is poor (i.e., bias is a concern), then averaging might be required as a last resort.

5. The choice of derivative is dictated by conflicting requirements. If you are close to the limit of good sampling ($\Phi \approx 10$), then the one-sided derivative is preferable. If the signal-to-noise ratio is poor and the line is reasonably well sampled, then the two-sided derivative is less biased.

6. Small leakage terms are extracted essentially unbiased by both least squares and maximum likelihood.

7. For both maximum likelihood and least squares, we can produce reliable standard error estimates for α when $\alpha_{\rm est}$ is unbiased. As $\hat{\eta}$ decreases and $\alpha_{\rm est}$ becomes progressively more biased, so do the estimates of the standard error in $\alpha_{\rm est}$. Both least squares and maximum likelihood make reliable estimates of the standard error in β for small leakage terms.

8. The effects of noise correlation must be considered when determining the error in the estimate of $\hat{\alpha}$; the estimate of $\hat{\alpha}$ itself is only weakly dependent on such correlations. Although correlations should be calculable from the characteristics of the observation, they can also be tested for by computing the autocorrelation function of a signal-free region. The maximum likelihood algorithm can handle spectral correlations, or spatial correlations in cases of good signal-to-noise ratios. In other instances, or for the least-squares technique, simulation is the best approach in determining error estimates.

b) Recommended Recipe for Success

Finally, we suggest a procedure which should lead to a good estimate of $\hat{\alpha}$ if there is sufficient signal to do so:

1. Ensure that the good sampling criterion is met.

2. Decide upon choice of derivative, remembering the trade-off between good sampling (use one-sided if close to the limit) and noisy data (use two-sided if well-sampled).

3. Select the region (spatial and spectral) of the data that contains the signal of interest, and over which you are assuming the splitting is constant. Then estimate $\hat{\eta}$ (using eq. [4.2]):

$$\eta \approx \frac{1}{\sigma}\sqrt{N^{-1}\sum_i \left(\frac{dI_{{\rm obs},i}}{d\nu}\right)^2},$$

and

$$\hat{\eta}^2 = \eta^2 - a.$$

Here $a = 2$ or $1/2$ for the one- and two-sided derivatives, respectively (assuming uncorrelated noise). If you are summing in the spatial as well as in the spectral dimension, then the sum is extended to the spatial dimensions as well (and N is the number of channels in each spectrum times the number of spectra).

4. Having computed $\hat{\eta}$, consider three regimes. (*a*) If $\hat{\eta} > 10$, then all estimators are unbiased. The choice of maximum likelihood or least squares is largely a matter of convenience in this case; least squares is much easier to implement and less computationally intensive. (*b*) If $1 < \hat{\eta} < 10$, then maximum likelihood yields essentially unbiased results (with the caveat that we have not explored very small values of $\hat{\alpha}$). Alternatively, one could use debiased least squares. (*c*) If $\hat{\eta} < 1$, then the estimate is biased no matter what estimator you use. There are two solutions. Either try to improve the signal-to-noise ratio by averaging spectra (this means that the spectral structure must be similar in the averaged spectra), or, better, get more data.

5. Both spatial and spectral correlations can be ignored when estimating $\hat{\alpha}$, but the standard error estimates depend significantly on the correlations. The maximum likelihood algorithm can provide good error estimates for high signal-to-noise ratios, but, for noisy data, the recommended technique for evaluating the effect of spatial correlations on the errors is by simulation. One can determine if spectral correlations exist by empirically computing the autocorrelation function of a signal-free spectrum or by knowledge of the characteristics of the observation.

6. Check the results for self-consistency by making sure that $\alpha_{\rm est}/\Phi < 0.1$. This is the small splitting assumption.

The facilities of the Cray-2 at the National Center for Supercomputing Applications were a vital part of the numerical simulations presented, and we acknowledge the support of that program. In addition, we are grateful for computational support from the Laboratory for Astronomical Imaging in the Astronomy Department, at the University of Illinois.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

APPENDIX A

THE LEAST-SQUARES TECHNIQUE

For the least-squares technique, we wish to minimize

$$\varepsilon^2 = \sum_i \left(V_{\text{obs},i} - \alpha \frac{dI_{\text{obs},i}}{d\nu} - \beta I_{\text{obs},i} \right)^2 . \tag{A1}$$

a) Matrix Notation

In this and the following appendix, it is convenient to formulate the problem in terms of matrix operations. Initially, we consider summation over the spectral dimension only and consider I and V to be vectors. If we have N channels, then vectors I_{obs} and $\hat{I}$ will be of length N. In estimating the derivative, we need one or two more channels in the I spectrum than the V spectrum so that V will be of length $N-1$ or $N-2$ for the one- and two-sided derivatives, respectively.

Let **D** be the matrix representing the derivative operator, and let **E** be an identity matrix padded on the left, and possibly the right, by a column of zeros. This is used to discard the first, and possibly the last, I channels (depending on the derivative used). Let **B** be a matrix combining α, the derivative operator, and the leakage term according to

$$\mathbf{B}I = \alpha \mathbf{D}I + \beta \mathbf{E}I .$$

The matrices **B**, **D**, and **E** will be of size $(N-1)\times N$ or $(N-2)\times N$. For the one-sided case,

$$\mathbf{B} = \begin{pmatrix} -\alpha & \alpha+\beta & & \\ & -\alpha & \alpha+\beta & \\ & \cdot & \cdot & \\ & & -\alpha & \alpha+\beta \end{pmatrix},$$

whereas for the two-sided case,

$$\mathbf{B} = \begin{pmatrix} -\alpha/2 & \beta & \alpha/2 & & \\ & -\alpha/2 & \beta & \alpha/2 & \\ & & \cdot & \cdot & \\ & & -\alpha/2 & \beta & \alpha/2 \end{pmatrix}.$$

b) Solution of the Least-Squares Technique

We now rewrite equation (A1) as

$$\varepsilon^2 = (V_{\text{obs}} - \mathbf{B}I_{\text{obs}})^T (V_{\text{obs}} - \mathbf{B}I_{\text{obs}}) .$$

Differentiating with respect to α and β, equating to zero, and solving the resultant pair of linear equations, results in the solution

$$\alpha_{\text{est}} = \frac{S_{EE}S_{VD} - S_{DE}S_{VE}}{S_{DD}S_{EE} - S_{DE}^2}, \qquad \beta_{\text{est}} = \frac{S_{DD}S_{VE} - S_{DE}S_{VD}}{S_{DD}S_{EE} - S_{DE}^2}, \tag{A2}$$

where

$$S_{DD} = I_{\text{obs}}^T \mathbf{D}^T \mathbf{D} I_{\text{obs}}, \qquad S_{DE} = I_{\text{obs}}^T \mathbf{D}^T \mathbf{E} I_{\text{obs}}, \qquad S_{EE} = I_{\text{obs}}^T \mathbf{E}^T \mathbf{E} I_{\text{obs}},$$

$$S_{VD} = I_{\text{obs}}^T \mathbf{D}^T V_{\text{obs}}, \qquad S_{VE} = I_{\text{obs}}^T \mathbf{E}^T V_{\text{obs}} .$$

These S terms are simple inner products between I_{obs}, V_{obs}, and derivative spectrum.

In the absence of a leakage term, the above equations simplify to

$$\alpha_{\text{est}} = \frac{S_{VD}}{S_{DD}} = \frac{I_{\text{obs}}^T \mathbf{D}^T V_{\text{obs}}}{I_{\text{obs}}^T \mathbf{D}^T \mathbf{D} I_{\text{obs}}} = \sum_i V_{\text{obs},i} \frac{dI_{\text{obs},i}}{d\nu} \Bigg/ \sum_i \left(\frac{dI_{\text{obs},i}}{d\nu} \right)^2 \tag{A3}$$

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

c) Bias and Error Analysis

We now analyze the bias and error of the estimate. Here we assume there is no leakage. In addition, we initially assume there is no noise correlation. The estimate of $\hat{\alpha}$ will have some fluctuation caused by the noise. The numerator and denominator of equation (A3) can be treated as random variables with a mean value, and some variation about this mean. Let

$$\alpha_{\rm est} = \frac{\mu_{VD} + \mathbf{X}_{VD}}{\mu_{DD} + \mathbf{X}_{DD}} \tag{A4}$$

where μ_{VD} and μ_{DD} are the mean values of S_{VD} and S_{DD}, and the $\mathbf{X}$ quantities are the variation about this mean. In both the numerator and denominator, the value of μ and the variance of $\mathbf{X}$ are proportional to N. Hence as N increases, the $\mathbf{X}/\mu$ ratios are proportional to $N^{-1/2}$. So for sufficiently large N, $\mathbf{X}$ is much smaller than μ, and we can replace equation (A4) with its first-order Taylor expansion:

$$\alpha_{\rm est} \approx \frac{\mu_{VD}}{\mu_{DD}} + \frac{\bar{1}}{\mu_{DD}}\mathbf{X}_{VD} - \frac{\mu_{VD}}{\mu_{DD}^2}\mathbf{X}_{DD}.$$

We thus have

$$\bar{\alpha}_{\rm est} = \mathrm{E}[\alpha_{\rm est}] = \frac{\mu_{VD}}{\mu_{DD}},$$

and

$$\sigma_\alpha^2 = \mathrm{var}[\alpha_{\rm est}] = \frac{1}{\mu_{DD}^2}\mathrm{var}[\mathbf{X}_{VD}] + \frac{\mu_{VD}^2}{\mu_{DD}^4}\,\mathrm{var}\,[\mathbf{X}_{DD}] - 2\frac{\mu_{VD}}{\mu_{DD}^3}\mathrm{E}[\mathbf{X}_{VD}\mathbf{X}_{DD}]. \tag{A5}$$

If we assume uncorrelated noise, some analysis shows

$$\begin{aligned} &\mu_{VD} = \hat{\alpha}\hat{I}^T\mathbf{D}^T\mathbf{D}\hat{I}, \qquad \mu_{DD} = \hat{I}^T\mathbf{D}^T\mathbf{D}\hat{I} + aN\sigma^2, \\ &\mathrm{var}[\mathbf{X}_{VD}] = \sigma^2\hat{I}^T\mathbf{D}^T\mathbf{D}\hat{I} + \sigma^2\hat{\alpha}^2\hat{I}^T\mathbf{D}^T\mathbf{D}\mathbf{D}^T\mathbf{D}\hat{I} + aN\sigma^4, \qquad \mathrm{var}[\mathbf{X}_{DD}] = 4\sigma^2\hat{I}^T\mathbf{D}^T\mathbf{D}\mathbf{D}^T\mathbf{D}\hat{I} + 2bN\sigma^4, \\ &\mathrm{E}[\mathbf{X}_{VD}\mathbf{X}_{DD}] = 2\hat{\alpha}\sigma^2\hat{I}^T\mathbf{D}^T\mathbf{D}\mathbf{D}^T\mathbf{D}\hat{I}. \end{aligned} \tag{A6}$$

Here $a = 2$ or $1/2$, and $b = 6$ or $3/8$ for the one- and two-sided derivative, respectively.

In simplifying, we ignore $I^T\mathbf{D}^T\mathbf{D}\mathbf{D}^T\mathbf{D}\hat{I}$, as this is small compared with the other terms for typical well-sampled lines. For example, the ratio of this term to $\hat{I}^T\mathbf{D}^T\mathbf{D}\hat{I}$ is approximately $(12 \log_e 2)/\Phi^2$, for a Gaussian line with FWHM of Φ channels. We define

$$\hat{\eta} = \frac{1}{\sigma}\left(\frac{d\hat{I}}{d\nu}\right)_{\rm rms}. \tag{A7}$$

This is a measure of the signal-to-noise in the derivative spectrum. This gives

$$\bar{\alpha}_{\rm est} = \frac{\hat{\alpha}}{1 + a\hat{\eta}^{-2}},$$

and

$$\sigma_\alpha^2 = \frac{1}{\hat{\eta}^2 N}\left[\frac{\bar{\alpha}_{\rm est}}{\hat{\alpha}} + \frac{2b\hat{\alpha}^2}{\hat{\eta}^2}\left(\frac{\bar{\alpha}_{\rm est}}{\hat{\alpha}}\right)^4\right]. \tag{A8}$$

The expression for $\bar{\alpha}_{\rm est}$ shows that the estimate is essentially unbiased for large values of $\hat{\eta}$ (i.e., high signal-to-noise ratios) but becomes progressively more biased as $\hat{\eta}$ decreases.

The second term in the variance (the term containing b) is not a significant term except when $\hat{\eta} \approx 1$, and it can be ignored when the signal-to-noise ratio is high (or very low). In addition, for large signal-to-noise ratios, there is essentially no bias, so $\bar{\alpha}_{\rm est}/\hat{\alpha} \approx 1$. So for high signal-to-noise ratios,

$$\sigma_\alpha^2 \approx \frac{1}{\hat{\eta}^2 N} = \sigma^2 \Big/ \sum_i \left(\frac{d\hat{I}_i}{d\nu}\right)^2.$$

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

The above analysis also holds if we consider spatial summing. The summations contained in the above equations are now performed along the spatial as well as spectral dimensions. If the noise is correlated, however, the analysis changes. The equation for $\bar{\alpha}_{\rm est}$ remains the same, except that the value of a becomes $2[1-r_n(1)]$ or $[1-r_n(2)]/2$ for the one- or two-sided derivative, respectively [recall that $r_n(i)$ is the correlation coefficient for the noise samples separated by i channels]. The error variance changes significantly and will not be presented here.

d) Debiasing

The bias is in the least-squares estimate for $\hat{\alpha}$ arises simply from the fact that it ignores the noise in the $I_{\rm obs}$ spectrum. When the signal-to-noise ratio in the derivative of I spectrum becomes small (less than about 10), this assumption becomes inappropriate and manifests as a bias in the expected value of S_{DD}. The error variance (eqs. (3.3) and [A8]) is biased by an identical factor.

The bias caused by the term $aN\sigma^2$ in μ_{DD}. If we know the rms noise, σ, we may subtract $aN\sigma^2$ from the calculated value of S_{DD}, and so (to first order) we can "debias" the least-squares algorithm. Similarly we can eliminate the bias from the error variance, σ_α. One must use some caution in performing debiasing, as error in σ will incompletely remove the bias. However, provided any uncertainty in σ^2 contributes negligibly to the overall uncertainty, equations (3.3) and (A8) remain valid equations for the error variance of the debiased least-squares estimate of $\hat{\alpha}$.

APPENDIX B

THE MAXIMUM LIKELIHOOD TECHNIQUE

a) Solution of the Maximum Likelihood Technique

In this appendix, we continue to use the matrix notation introduced in Appendix A. Again we initially consider the case of handling one spectrum only. However, we include noise correlation in our analysis. Let $\sigma^2\mathbf{R}$ be the covariance matrix of the noise process. $\mathbf{R}$ is the square Toeplitz matrix corresponding to the noise autocorrelation function $r_n(i)$. This matrix will be of size $N\times N$ when it is associated with I, and $(N-1)\times(N-1)$ or $(N-2)\times(N-2)$ when associated with V. We distinguish these as $\mathbf{R}_{N_I}$ and $\mathbf{R}_{N_V}$. If the noise is uncorrelated, $\mathbf{R}$ becomes an identity matrix.

The appropriate probability density function is a multivariate Gaussian process, where the noise samples are not independent. This probability density function (see Davenport and Root 1958) is proportional to

$$\exp\left(\frac{-\chi^2}{2}\right),$$

where

$$\sigma^2\chi^2 = n_I^T\mathbf{R}_{N_I}^{-1}n_I + n_V^T\mathbf{R}_{N_V}^{-1}n_V = \left(I_{\rm obs}-\hat{I}\right)^T\mathbf{R}_{N_I}^{-1}\left(I_{\rm obs}-\hat{I}\right)+\left(V_{\rm obs}-\mathbf{B}\hat{I}\right)^T\mathbf{R}_{N_V}^{-1}\left(V_{\rm obs}-\mathbf{B}\hat{I}\right).$$

The maximum likelihood solution for the unknowns $\alpha_{\rm est}$, $\beta_{\rm est}$, and $I_{\rm est}$ (an estimate of $\hat{I}$), are those values which maximize the probability density function (or, equivalently, minimize χ^2). By differentiating with respect to the unknowns, it can be shown that at a local minimum of χ^2

$$I_{\rm est} = \left(\mathbf{R}_{N_I}^{-1}+\mathbf{B}^T\mathbf{R}_{N_V}^{-1}\mathbf{B}\right)^{-1}\left(\mathbf{R}_{N_I}^{-1}I_{\rm obs}+\mathbf{B}^T\mathbf{R}_{N_V}^{-1}V_{\rm obs}\right),$$

$$\alpha_{\rm est} = \frac{S_{EE}S_{VD}-S_{DE}S_{VE}}{S_{DD}S_{EE}-S_{DE}^2}, \qquad \beta_{\rm est} = \frac{S_{DD}S_{VE}-S_{DE}S_{VD}}{S_{DD}S_{EE}-S_{DE}^2},$$

where

$$S_{DD} = I_{\rm est}^T\mathbf{D}^T\mathbf{R}_{N_V}^{-1}\mathbf{D}I_{\rm est}, \qquad S_{DE} = I_{\rm est}^T\mathbf{D}^T\mathbf{R}_{N_V}^{-1}\mathbf{E}I_{\rm est}, \qquad S_{EE} = I_{\rm est}^T\mathbf{E}^T\mathbf{R}_{N_V}^{-1}\mathbf{E}I_{\rm est}$$

$$S_{VD} = I_{\rm est}^T\mathbf{D}^T\mathbf{R}_{N_V}^{-1}V_{\rm obs}, \qquad S_{VE} = I_{\rm est}^T\mathbf{E}^T\mathbf{R}_{N_V}^{-1}V_{\rm obs}. \tag{B1}$$

The S terms are quite similar to those in the least-squares case, except they are inner products involving $I_{\rm est}$ (rather than $I_{\rm obs}$), and that $\mathbf{R}_{N_V}^{-1}$ is used as the metric. If $\mathbf{R}$ is an identity matrix (i.e., the uncorrelated noise case), then the similarity becomes greater. Note that the equations of equation (B1) are not independent of each other.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

Since $N+2$ parameters (α, β, and I) are being fitted to $2N$ data points (N is the number of spatial and spectral points included in the summation), the value of χ^2 at the minimum follows a χ^2 distribution with $2N-(N+2)=N-2$ degrees of freedom, provided that the signal-to-noise ratio is sufficiently large (see next paragraph). Thus, χ^2 ($\alpha_{\rm est}, \beta_{\rm est}, I_{\rm est}$) has a mean value of $N-2$ and an rms deviation of $\sqrt{2(N-2)}$. This fact can be used to check the goodness of fit or the accuracy of the estimate of σ^2.

The derivation of the formal errors of the fitted parameters $\alpha_{\rm est}$ and $\beta_{\rm est}$ follows the standard treatment (see Press *et al.* 1986, and references therein). This method is valid provided that the signal-to-noise ratio is high enough so that χ^2 may be approximated by a parabola within the range of the uncertainties of the fitted parameters. In such circumstances, the statements in the previous paragraph regarding the distribution of $\chi^2(\alpha_{\rm est}, \beta_{\rm est}, I_{\rm est})$ are valid. We assume this to be the case in the following discussion. We will limit our discussion to σ_α, the error estimate for $\alpha_{\rm est}$, but the extension of the method to the other fitted parameters is obvious. We begin by defining

$$\Delta\chi^2(\alpha) = \chi^2(\alpha, \beta_\alpha, I_\alpha) - \chi^2(\alpha_{\rm est}, \beta_{\rm est}, I_{\rm est}), \tag{B2}$$

where β_α and I_α are chosen to minimize χ^2 at a given value of α. The key point is that $\Delta\chi^2$ as defined above *and evaluated at* $\alpha=\hat{\alpha}$, *where* $\hat{\alpha}$ *is the true value*, has a χ^2 distribution with one degree of freedom. Thus, the 68% confidence interval for $\hat{\alpha}$ is given by $\alpha_{\rm est} \pm \sigma_\alpha$, where σ_α is determined from

$$\Delta\chi^2(\alpha_{\rm est} \pm \sigma_\alpha) = 1. \tag{B3}$$

To restate this in simpler terms: the interval $(\alpha_{\rm est}-\sigma_\alpha, \alpha_{\rm est}+\sigma_\alpha)$ has a probability of 0.68 of enclosing $\hat{\alpha}$. We may derive an explicit expression for σ_α by expanding $\Delta\chi^2$ around its minimum:

$$\Delta\chi^2(\alpha) = \Delta\chi^2(\alpha_{\rm est}) + \frac{1}{2}\frac{d^2\Delta\chi^2}{d\alpha^2}(\alpha-\alpha_{\rm est})^2 = \frac{1}{2}\frac{d^2}{d\alpha^2}\left[\chi^2(\alpha,\beta_\alpha,I_\alpha)\right](\alpha-\alpha_{\rm est})^2. \tag{B4}$$

Combining equations (B3) and (B4) leads to the result:

$$\frac{1}{\sigma_\alpha^2} = \frac{1}{2}\frac{d^2}{d\alpha^2}\left[\chi^2(\alpha,\beta_\alpha,I_\alpha)\right], \tag{B5}$$

where the second derivative is evaluated at $\alpha_{\rm est}$. Some appropriate analysis shows that, when the signal-to-noise ratio is large, the second (complete) derivative of χ^2 can be replaced with the second partial derivative of χ^2 with respect to α. Empirical studies also have shown this approximation to be better than 1% in the cases tested. This gives

$$\frac{1}{\sigma_\alpha^2} \approx \frac{1}{\sigma^2}\sum_i \left(\frac{d\hat{I}}{d\nu}\right)^2.$$

When the signal-to-noise ratio is poor, χ^2 is far from parabolic. In this case, it is more appropriate to fit a parabola to the region around the minimum of χ^2 and use the σ_α that this parabola implies. This parabola is not meant to approximate χ^2, but rather to measure the broadness of the minimum. In practice, we fit a parabola to χ^2 at three points, viz., at the minimum and at the two points where $\Delta\chi^2=16$ (i.e., $\pm 4\ \sigma_\alpha$). When the signal-to-noise ratio is moderate, this approach leads to an estimate of σ_α which is only slightly larger than implied by the derivative. In poor signal-to-noise cases, it can lead to values a factor of 1.5 to 2 larger. However, it must be realized that the only completely rigorous method for estimating σ_α in the poor signal-to-noise cases is to perform a Monte Carlo simulation.

b) General Algorithmic Concerns

A two-step iterative procedure was developed to find minima of χ^2. The first step is to calculate $I_{\rm est}$, assuming we know α and β. This involves the solution of the system of linear equations (first line of eq. [B1]). If $\mathbf{R}$ is an identity matrix, then the linear system reduces to solving a system involving the matrix $\mathbf{1}+\mathbf{B}^T\mathbf{B}$ ($\mathbf{1}$ is the identity matrix). This is a tridiagonal or pentadiagonal matrix for the one- and two-sided derivatives, respectively, so the calculation of $I_{\rm est}$ is reasonably efficient. During the second step, we assume we know $I_{\rm est}$ and then calculate the inner products needed to determine $\alpha_{\rm est}$ and $\beta_{\rm est}$. We iterate in this fashion until the unknowns converge. We start the algorithm by giving initial estimates of α and β. In cases where the signal-to-noise ratio is poor, χ^2 has multiple minima. Often there is a spurious local minimum near $\alpha=0$. To attempt to find the global minimum, our algorithm tries three initial estimates of α. For the first run, the least-squares solution for α is used. The second and third runs use $\alpha_1 \pm 15\sigma_{\alpha 1}$, where α_1 and $\sigma_{\alpha 1}$ are the values of α and σ_α that the first run converged to. The variable β is always initially set to 0. The best of these three minima is then chosen (often the three minima are identical). In practice, we have never noted this scheme failing to converge to the global minimum.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

c) *Spatial Summing Extension*

Equations (B1) and (B2) apply equally well to the case when we are performing spatial summing; it is just the details of the matrices which change. Though it is possible to formally define the forms of these matrices, we will give only an intuitive approach here. See, for example, Andrews and Hunt (1977) for matrix approaches in similar situations.

We temporarily assume the noise is uncorrelated; that is, the covariance matrix is an identity matrix. If we are given α and β, we note that the calculation of $I_{\rm est}$ is dependent on values of I and V in the same spectrum but is independent of I and V at different spatial locations. Hence we deal with only one spectrum at a time when calculating $I_{\rm est}$. The spatial extension of calculating the inner products (the S terms in eq. B1), needed to estimate α and β, is straightforward. These simply involve summing the products over the spatial dimensions, as well as the spectral dimension. Thus we again have a two-step algorithm, where we iterate between calculating $I_{\rm est}$ and $\alpha_{\rm est}$ and $\beta_{\rm est}$.

d) *Correlated Noise Case*

The presence of $\mathbf{R}^{-1}$ in these equations adds significant extra complexity to the algorithm. This is particularly true when there is spatial summing, as this would couple the solutions of $I_{\rm est}$ for one spectrum with the solutions of other spectra. The matrices involved would quickly become unmanageably large. We now discuss two approaches to simplify the case where there is correlation.

The first approach is to ignore the correlation, as the solution does not depend strongly on $\mathbf{R}$. In particular, the equation which gives $I_{\rm est}$ is quite weakly dependent on $\mathbf{R}$. To see this, we note that all the matrices involved are Toeplitz in form. If the spatial summing region or the number of channels is large (which is invariably true), then the Toeplitz matrices are approximately circulants (see Bellman 1960; or Andrews and Hunt 1977). Circulant matrices are commutative and the $\mathbf{R}$ factors cancel out in the equation for $I_{\rm est}$. In this case, solving for $I_{\rm est}$ involves inverting the matrix $\mathbf{1}+\mathbf{B}^T\mathbf{B}$.

Simulation has also shown that the solutions of $I_{\rm est}$, $\alpha_{\rm est}$, and $\beta_{\rm est}$ are all weakly depending on the covariance matrix. However, this is not true of the formal error, σ_α. If we completely ignore the covariance matrix, then the resulting formal error will be too low. This is because we have fewer truly independent samples than the algorithm has assumed. Even if we ignore correlation when solving for α and β, we can still include it when estimating the formal errors. Another way of estimating the formal error would be by simulation. These possibilities are examined in § IV.

A second approach for dealing with $\mathbf{R}^{-1}$ is to look for matrix approximations which help simplify the calculations. As noted, the matrices are approximately circulants, and so Fourier transform techniques could be used to perform the matrix operations. Another approximation, which might be applicable if the covariance matrix differs little from the identity matrix, is to use a Taylor series expansion. Specifically, if $\mathbf{R}=\mathbf{1}+\Delta$, where Δ is small in some sense, then $\mathbf{R}^{-1}=(\mathbf{1}+\Delta)^{-1}\approx\mathbf{1}-\Delta$. This approach is more applicable for the spectral dimension, where the correlation lengths are typically fairly short.

Although we have developed code to perform a full treatment of noise correlation in the spectral dimension, we have not done so for the spatial dimensions. Rather we assume that the correlation does not affect the solution greatly. However, we have developed code to perform a proper treatment of correlation when calculating σ_α. Because the noise autocorrelation function is separable into a spatial function and a spectral function, the metric used in these inner products is also separable. Provided the spatial region is small, then the matrix operations involved are manageable. For regions of even moderate size, we must be cautious of the spatial covariance matrix being ill-conditioned or singular. This is an inevitable problem with radio synthesis observations, being caused by incomplete coverage of the $(u-v)$ plane. This singularity or near-singularity can result in the errors in $I_{\rm est}$ being wildly amplified and leading to grossly incorrect values of σ_α. For the dirty beams from our observations, we find that 3×3 is the largest spatial summing region we can handle, before ill conditioning in the spatial covariance matrix becomes a problem.

This near-singularity is not a warning sign that our problem, or the maximum likelihood technique, is fundamentally ill-conditioned. Rather it is a sign that the standard approach, of using the inverse of the covariance matrix, is not appropriate.

e) *Maximum Likelihood Bias*

Bias conclusions for the maximum likelihood case are more involved than the least-squares case. The problem is again caused by biases in the expected values of S_{VD} and S_{DD}. In calculating these for the maximum likelihood technique, we use $I_{\rm est}$ rather than $I_{\rm obs}$, and it is the error in $I_{\rm est}$ which causes the bias. Nevertheless, $I_{\rm est}$ is a better estimate of $\hat{I}$ than $I_{\rm obs}$. This is because maximum likelihood uses information in both $I_{\rm obs}$ and $V_{\rm obs}$ when determining $I_{\rm est}$. Therefore, the maximum likelihood technique should be less biased than the least-squares technique.

The error in the maximum likelihood estimate of $\hat{I}$ is a function of $\hat{\alpha}$ because the estimate of $\hat{I}$ is a function of $\alpha_{\rm est}$. Indeed, as $\hat{\alpha}$ approaches zero, $I_{\rm est}$ approaches $I_{\rm obs}$, and consequently the bias in the maximum likelihood technique should approach that of the least-squares technique. On the other hand, for given signal and noise levels in $I_{\rm obs}$, as the magnitude of $\hat{\alpha}$ increases, the error in $I_{\rm est}$ decreases, and the bias disappears.

As we do not have an expression for the bias of the maximum likelihood technique, and because the bias is also a function of $\hat{\alpha}$, attempting to debias is much less attractive than it is for least squares.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1990ApJS...74..437S

REFERENCES

Andrews, H. C., and Hunt, B. R. 1977, *Digital Image Restoration* (Englewood Cliffs: Prentice-Hall).
Bellman, R. E. 1960, *Introduction to Matrix Analysis* (New York: McGraw-Hill).
Crutcher, R. M., Kazes, I., and Troland, T. H. 1987, *Astr. Ap.*, **181**, 119.
Davenport, W. B., and Root, W. L. 1958, *An Introduction to the Theory of Random Signals and Noise* (New York: IEEE Press).
Goldreich, P., Keeley, D. A., and Kwan, J. Y. 1973, *Ap. J.*, **182**, 55.
Jackson, J. D. 1975, *Classical Electrodynamics* (New York: Wiley).
Jeffreys, H. 1939, *Theory of Probability* (3d ed.; London: Oxford University Press).
Kazes, I., and Crutcher, R. M. 1986, *Ap. J.*, **164**, 328.
Lo, K. Y., Walker, R. C., Burke, B. F., Moran, J. M., Johnston, K. J., and Ewing, M. S. 1975, *Ap. J.*, **202**, 650.
Park, S. K., and Miller, K. W. 1988, *Comm. A.C.M.*, **31**, No. 10.
Parratt, L. G. 1961, *Probability and Experimental Errors in Science* (New York: Wiley).
Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. 1986, *Numerical Recipes* (New York: Cambridge University Press).
Schwarz, U. J., Troland, T. H., Albinson, J. S., Bregman, J. D., Goss, W. M., and Heiles, C. 1986, *Ap. J.*, **301**, 320.
Tan, S. M., and Gull, S. F. 1985, *M.N.R.A.S.*, **216**, 949.
Townes, C. W., and Schawlow, A. L. 1975, *Microwave Spectroscopy* (New York: Dover).
Troland, T. H., and Heiles, C. 1982, *Ap. J.*, **252**, 179.
_____. 1986 *Ap. J.*, **301**, 339.
Verschuur, G. L. 1969, *Phys. Rev. Letters*, **21**, 775.

N. E. B. Killeen and R. J. Sault: Australia Telescope National Facility, CSIRO, P.O. Box 76, Epping, NSW, 2121, Australia.

R. Loushin: Astronomy Department, University of Illinois, 1011 West Springfield Avenue, Urbana, IL 61801

J. Zmuidzinas: Caltech, Downs Laboratory, 320-47, Pasadena, CA 91125

© American Astronomical Society • Provided by the NASA Astrophysics Data System