TECHNICAL REPORT CR-RD-AS-92-4

BAYESIAN INTERPOLATION AND DECONVOLUTION

G. LARRY BRETTHORST
Washington University
Department of Chemistry
Campus Box 1134
One Brookings Drive
St. Louis, Missouri 63130-4899

Prepared for:

The Advanced Sensor Directorate Research,
Development, and Engineering Center
Contract No. DAAL03-86-D-0001

July 1992

The U. S. Army Missile Command
Redstone Arsenal, Alabama 35898-5000

Approved for public release; distribution is unlimited.

Contents

List of Figures

BAYESIAN INTERPOLATION AND DECONVOLUTION

G. LARRY BRETTHORST
Washington University
Department of Chemistry
Campus Box 1134
One Brookings Drive
St. Louis, Missouri 63130-4899

ABSTRACT. The deconvolution problem is addressed in stages beginning with the interpolation problem when little prior information is available and proceeding to the full deconvolution problem when a great deal of prior information is available. The results of the calculations indicate that good solutions to the deconvolution problem are available even when limited prior information is available and that these results overlap those obtained when a great deal of prior information is available. The difference between them is that the use of uninformative priors causes large uncertainties in the estimated signal, while highly informative priors decreases the uncertainties in the estimated signal.

Introduction

The deconvolution problem is important in many branches of science and engineering. In this problem the "image" or signal is convolved with a smearing function. This function is also called an impulse response function because the ideal noiseless signal that one would obtain in response to an input impulse or delta function is the smearing function for detector. In linear systems the output from an arbitrary input may be written as a convolution or average of the true signal convolved with the impulse response function. Averaging loses information. In addition the signal is contaminated with noise, consequently there is no unique way to deconvolve the signal from the impulse response function; rather one must make inferences about the true signal. In this paper, the deconvolution problem is studied beginning with the simplest "baby" version of this problem and proceeding through stages to more and more complex versions of the problem until, finally, the full deconvolution problem is analyzed. At the end of each stage, numerical examples are supplied to illustrate the calculations.

In the deconvolution problem addressed here, there is a data set D which is postulated to contain a signal $y(t)$ plus additive noise:

$$d(t_i) = y(t_i) + n_i \tag{1}$$

where n_i represents the noise. The data D are a collection of N discrete data samples, $D \equiv \{d(t_1), \ldots, d(t_N)\}$. The signal $y(t)$ is obtained from a "convolution" integral of the form

$$y(t) = \int_{t_1}^{t_N} d\tau r(t-\tau)u(\tau) \tag{2}$$

where $r(t)$ is the impulse response function, and $u(t)$ is the unknown signal. The data D have been written as one dimensional, although the mathematics will take no notice of this and the results may be generalized to higher dimensions by simply relabeling the higher dimensional quantities. The signal that appears in the detector, $y(t)$, will be thought of as a time series, although again the mathematics takes no notice of this, and one could, for example, interpret t as position, as one would in an image. The problem is to make the best inference possible for the unknown signal, $u(t)$, from the data and the prior information.

When the impulse response function $r(\tau)$ is a Dirac delta function

$$r(t-\tau) = \delta(t-\tau), \tag{3}$$

the convolution integral may be evaluated and one obtains

$$d(t_i) = u(t_i) + n_i. \tag{4}$$

The deconvolution problem has reduced to the "data interpolation" problem. *Clearly if one is to understand the deconvolution problem, then one must have a firm understanding of the interpolation problem.* For this reason the data interpolation problem will be studied in the first two sections of this paper.

In the first section, the interpolation problem is addressed, and probability theory will be used to derive the posterior probability for the value of an arbitrary pixel given the data and the prior information. In this baby version of the problem the prior information will be that the signal should be smooth.

In the second section, the analysis of the interpolation problem continues with the use of more informative prior information. This more informative prior information will include information about the functional form of the signal, as well as information about the first and second derivatives. At the end of each sections several numerical examples are given.

In the third section, the full deconvolution problem is addressed using the techniques and procedures developed in the first two sections. Again numerical examples are included at the end of this section. Then in the fourth section the deconvolution is generalized to include more general types of prior information. Additionally, more efficient means of estimating the signal and the uncertainty in the estimate are developed.

1 Data Interpolation – Second Derivative Prior Information

In the data interpolation problem, there is a signal U. This signal is to be estimated at a number of discrete points. These discrete points will be called pixels. These pixels will be labeled $\{u_0, \ldots, u_{\nu+1}\}$ where

$$\nu \equiv \beta(N-1)+1, \tag{5}$$

is the number of the pixel corresponding to the last data value, and pixel u_1 corresponds to the first data value. The pixels labeled $u_1, \ldots, u_\nu$ will be called interior pixels; while u_0, and $u_{\nu+1}$ will be called boundary pixels. These boundary pixels are special because they must be handled differently. The pixel density factor, β, indicates the density of the pixels relative to the data. If $\beta = 1$ there is a one to one correspondence between the pixels and the data (excluding the two boundary pixels). If $\beta = 2$, there are two pixels for every data value, etc. The discrete times t_i correspond to the pixels, not the data. So the sampling times for the data are given by $\{t_1, t_{\beta+1}, t_{2\beta+1}, \ldots, t_\nu\}$, and the data elements will also be labeled to correspond to the pixels: $\{d_1, d_{\beta+1}, d_{2\beta+1}, \ldots, d_\nu\}$. The collection of all of the data will be labeled as D, while the collection of all of the pixels will be labeled as U.

The data D consists of values of the signal U plus noise:

$$d_i = u_i + n_i \qquad i = \{1, \beta+1, 2\beta+1, \ldots, \nu\} \tag{6}$$

where n_i is the value of a randomly varying component that one has no way to predict. The problem is to make the best estimate of any one of the pixels possible. Because we will estimate an arbitrary pixel u_j, we will have estimated all of them by letting j take on any value $\{0 \le j \le \nu+1\}$. From the standpoint of probability theory, all of the information relevant to this inference is contained in a probability density function: $P(u_j|D,I)$, the probability that the signal has value u_j, given the data and the prior information I. This probability is computed using the sum rule

$$P(u_j|D,I) = \int \underbrace{\cdots du_i \cdots}_{i \neq j} P(U|D,I), \tag{7}$$

where $P(U|D,I)$ is the joint probability for all of the pixel values. The integrals are over all pixel values, except u_j.

Bayes theorem [1] may be used to factor $P(U|D,I)$ to obtain

$$P(u_j|D,I) = \int \underbrace{\cdots du_i \cdots}_{i \neq j} \frac{P(U|I)P(D|U,I)}{P(D|I)}, \tag{8}$$

where $P(U|I)$ is the joint prior probability for all the pixel values, $P(D|U,I)$ is the probability for the data given the pixel values, and $P(D|I)$ is a normalization constant.

Making the standard assumptions about the noise, the probability for the data given U is just the likelihood function

$$P(D|\sigma,U,I) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu} (d_i - u_i)^2 \right\}, \tag{9}$$

where the standard deviation of the noise, σ, has been added to the direct probability for the data in a way that indicates its value is known. Later, the rules of probability theory will be applied to remove σ from the problem, if its actual value is unknown. The index i [on the sum in Eq. (9)], means that i starts at 1 and goes to ν in steps of β. Substituting the direct probability into the posterior probability, Eq. (8), and assuming normalization will occur at the end of the calculation, one obtains

$$P(u_j|\sigma,D,I) \propto \int \underbrace{\cdots du_i \cdots}_{i \neq j} P(U|I)\sigma^{-N} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu} (d_i - u_i)^2 \right\}. \tag{10}$$

The problem has been reduced to specifying the prior probability, $P(U|I)$.

If one were to ignore the prior, as would using maximum likelihood, then all of the pixels values associated with the data values are estimated to be equal to the data, $u_j = d_j$, while all of the interpolation pixels are estimated to be zero. This is the maximum likelihood or least squares solution to this problem. But probability theory automatically tells one this is not correct. It is the product of a weighted average of the prior and the likelihood that must be considered. This weighted average will be very different from the maximum likelihood solution. And this difference is maintained even in the limit of very uninformative prior information.

For any given problem there could be a great deal of prior information available. For example, if the data were the output from a continuous wave radar, then the signal will look highly sinusoidal;

yet significant deviations will occur near the beginning and ending of the signal. If the radar were a pulsed radar, the signal would, at least superficially, like the derivative of a Gaussian. Again there could be significant deviations. This information is qualitatively different from that normally associated with a model, where the prior information *insists* that the signal must be of a certain functional form and any deviations from it are to be considered noise. Here the signal should be allowed to make deviations from the functional forms when the data shows evidence for such deviations. This type of prior information will be called "soft" because we do not *insist* that the signal have this functional form.

In addition to this soft prior information about the functional form of the signal, one might know some general characteristics about the signal. For example, the signal might be generated by some analogue electronics. Electronics never generates perfectly sharp signals; it always averages things out. Could smoothness be used as "soft" prior information?

The answers to this question is yes! It is possible to include both types of "soft" prior information in the calculation. Probability theory can be told that the signal is more or less sinusoidal, without *insisting* that it be sinusoidal, just as it will be possible to tell probability theory that the signal should be smooth without *insisting* that the signal must be smooth. To see how to do this, the interpolation problem will be investigated using both of these types of "soft" prior information. We begin by including prior information about the "smoothness' of the signal, and then in the next section proceed to include "soft" information about the functional form of the signal.

1.1 Constraining The Second Derivative

In the traditional interpolation problem, the data is assumed noiseless and one is trying to interpolate between data values. The criteria used in splines is typically minimum arc length, and one seeks the shortest interpolation function. Here noise is allowed into the problem. This noise could be either positive or negative and its effect is to make the data "jitter" around the "signal" in a random way. This jitter should be suppressed as much as possible. Mathematically this jitter corresponds to a rapidly varying second derivative. It can be suppressed if the second derivative of the signal can, in some sense, be made "small."

The data are sampled at discrete times. The first and second derivatives are not defined for discrete functions. However, one can define analogous quantities which reduce to the first and second derivative as the sampling density goes to infinity. The first derivative of a continuous function may be defined as

$$\frac{df(t)}{dt} = \lim_{\Delta \to 0} \frac{f(t+\Delta) - f(t-\Delta)}{2\Delta}. \tag{11}$$

For a discretely sampled function this becomes

$$\frac{df(t_i)}{dt_i} = \frac{f(t_i+\Delta) - f(t_i-\Delta)}{2\Delta} \tag{12}$$

where $f(t_i + \Delta) = f(t_{i+1})$ is the function at the forward sampling time, $f(t_i - \Delta) = f(t_{i-1})$ is the function at the backward sampling time, and

$$\Delta \equiv t_{i+1} - t_i = t_i - t_{i-1} \tag{13}$$

is the sampling time. It is clear from this definition that the discrete first derivative is only an approximation. This approximation is accurate to order Δ. So if delta is 0.01, i.e., if data were collected every 0.01 seconds, then the discrete first derivative will be accurate to ± 0.01.

The second derivative is just a derivative of a derivative and is defined as

$$\frac{d^2 f(t)}{dt^2} = \lim_{\Delta\to 0} \frac{\dfrac{f(t+2\Delta)-f(t)}{2\Delta} - \dfrac{f(t)-f(t-2\Delta)}{2\Delta}}{2\Delta}. \tag{14}$$

This can be rewritten as

$$\frac{d^2 f(t)}{dt^2} = \lim_{\Delta\to 0} \frac{f(t+2\Delta)+f(t-2\Delta)-2f(t)}{4\Delta^2}. \tag{15}$$

The corresponding equation for a discretely sampled signal is given by

$$\frac{d^2 f(t_i)}{dt^2} = \frac{f(t_{i+1})+f(t_{i-1})-2f(t_i)}{\Delta^2}. \tag{16}$$

Note that this approximation is accurate to order Δ^2, so if Δ is small, second derivatives my be evaluated very precisely, provided sufficient machine accuracy is available.

Now that we have a definition of the discrete second derivative, the prior information, that it must be "small" must be translated into a prior probability $P(U|I)$. But the second derivative can be positive or negative. Additionally, the second derivative is defined at every data point, so what is meant by "small"? Here "small' will mean that the mean-square value of the second difference should be small:

$$\sum_{j=1}^{\nu} [u_{j+1} + u_{j-1} - 2u_j]^2 = \delta^2, \tag{17}$$

where δ^2 is the total second difference. This equation will be referred to as a constraint on the second derivative for reasons that will become apparent shortly. The quantity δ, is a measure of the "smallness" of the second derivative. When δ is large, large jitter is allowed and the signal will be estimated to be the data values. When $\delta \to 0$, no jitter is allowed, and the signal will be estimated to be constant. Somewhere between these extreme values is one which will suppress the jitter without suppressing the signal.

Note that this constraint introduces other parameters into the problem. If for example $\beta = 1$, the constraint introduces three new parameters: two "boundary" pixels, u_0 and $u_{\nu+1}$, and a regularization parameter which will be called ϵ and is related to δ^2. If $\beta > 1$, the constraint also introduces the "interpolation" pixels into the problem.

The process of converting Eq. (17) into a prior probability density function is a straightforward application of the principle of maximum entropy and results in the assignment of a Gaussian prior probability:

$$P(u_1, \ldots, u_\nu | u_0, u_{\nu+1}, \epsilon, \sigma, I) \propto \exp\left\{-\frac{\epsilon^2}{2\sigma^2} \sum_{i=1}^{\nu} [u_{i+1} + u_{i-1} - 2u_i]^2\right\}, \tag{18}$$

where ϵ^2/σ^2 is the Lagrange multiplier from the maximum entropy calculation. The fractional variance ϵ^2 will be used to control the amount of smoothing and is related to the mean-square second derivative.

Three additional parameters: u_0 and $u_{\nu+1}$, the boundary pixels, and the fractional variance ϵ have entered the problem. These parameters were added to the prior in a way that indicates that their values are given. Of course in a real problem their values will not be known and inferences must be made about them. All three of these parameters are nuisances in the sense that one would like to formulate the problem independent of their value. This may be done readily for u_0, and $u_{\nu+1}$; but ϵ will prove to be harder to deal with.

What we have derives so far is the prior probability for the interior pixels given the boundary pixels. What is needed is the prior probability for all of the pixels. To compute this the joint prior for all of the pixels is factored using the product rule to obtain:

$$P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma I) = P(u_0,u_{\nu+1}|\epsilon,\sigma,I)P(u_1,\ldots,u_\nu|u_0,u_{\nu+1},\epsilon,\sigma,I) \tag{19}$$

where $P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma I)$ is the joint prior for the interior and boundary pixels; the joint probability for the interior pixels given the boundary pixels, $P(u_1,\ldots,u_\nu|u_0,u_{\nu+1},\epsilon,\sigma,I)$, is given by Eq.(18) and $P(u_0,u_{\nu+1}|\epsilon,\sigma,I)$ is the prior probability for the boundary pixels.

To assign the prior probability for these two boundary pixels, $P(u_0,u_{\nu+1}|\epsilon,\sigma,I)$, a different interpretation of the second derivative will be used. Suppose it is known that adjacent pixels should be approximately equal:

$$u_i \approx \frac{u_{i+1}+u_{i-1}}{2}. \tag{20}$$

This may be rewritten as

$$u_{i+1}+u_{i-1}-2u_i \approx 0. \tag{21}$$

But this is essentially just the statement that the second derivative should be small. So constraining the second derivative to be small is equivalent to asserting that neighboring pixels should be approximately equal. On the boundary this could be interpreted as

$$u_0 \approx u_1 \quad \text{and} \quad u_\nu \approx u_{\nu+1}. \tag{22}$$

Converting this prior information into a prior probability for u_0, one obtains

$$P(u_0|\epsilon,\sigma,I) \propto \exp\left\{-\frac{\epsilon^2}{2\sigma^2}(u_0-u_1)^2\right\}, \tag{23}$$

and similarly for $u_{\nu+1}$

$$P(u_{\nu+1}|\epsilon,\sigma,I) \propto \exp\left\{-\frac{\epsilon^2}{2\sigma^2}(u_{\nu+1}-u_\nu)^2\right\}. \tag{24}$$

To combine these priors, one uses the product rule to factor $P(u_0,u_{\nu+1}|\epsilon,\sigma,I)$, and assuming independence one obtains:

$$P(u_0,u_{\nu+1}|\epsilon,\sigma,I) = P(u_0|\epsilon,\sigma,I)P(u_{\nu+1}|\epsilon,\sigma,I). \tag{25}$$

Substituting for $P(u_0|\epsilon,\sigma,I)$ and $P(u_{\nu+1}|\epsilon,\sigma,I)$, one obtains

$$P(u_0,u_{\nu+1}|\epsilon,\sigma,I) \propto \exp\left\{-\frac{\epsilon^2}{2\sigma^2}[u_0-u_1]^2 - \frac{\epsilon^2}{2\sigma^2}[u_{\nu+1}-u_\nu]^2\right\} \tag{26}$$

as the joint prior probability for the boundary pixels. Substituting the joint prior for the boundary pixels, Eq. (26), and the prior for the interior pixels, Eq. (18), into the prior probability for all of the pixels including the boundary pixels, one obtains

$$\begin{aligned} P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma,I) \quad &\propto \quad \exp\left\{-\frac{\epsilon^2}{2\sigma^2}[u_0-u_1]^2 - \frac{\epsilon^2}{2\sigma^2}[u_{\nu+1}-u_\nu]^2\right\} \\ &\times \quad \exp\left\{-\frac{\epsilon^2}{2\sigma^2}\sum_{i=1}^{\nu}[u_{i+1}+u_{i-1}-2u_i]^2\right\}. \end{aligned} \tag{27}$$

This prior can be rewritten as

$$P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma,I) = [\lambda_0\cdots\lambda_{\nu+1}]^{\frac{1}{2}}\left(\frac{2\pi\sigma^2}{\epsilon^2}\right)^{-\frac{\nu+2}{2}}\exp\left\{-\frac{\epsilon^2}{2\sigma^2}\sum_{k=0}^{\nu+1}\sum_{l=0}^{\nu+1}R_{kl}u_k u_l\right\}, \tag{28}$$

where $\{\lambda_0,\ldots,\lambda_{\nu+1}\}$ are the eigenvalues of the matrix R_{kl} defined as

$$R_{kl} \equiv \begin{pmatrix} 2 & -3 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ -3 & 6 & -4 & 1 & 0 & \ddots & \ddots & \ddots & \vdots \\ 1 & -4 & 6 & -4 & 1 & \ddots & \ddots & \ddots & \vdots \\ 0 & 1 & -4 & 6 & -4 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & -4 & 6 & -4 & 1 & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 & -4 & 6 & -4 & 1 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 1 & -4 & 6 & -3 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & -3 & 2 \end{pmatrix} \qquad (0 \le k,l \le \nu+1). \tag{29}$$

Note that in writing the prior in this form, it has been implicitly assumed that the R_{kl} matrix is not singular. As this prior is written, this is not the case! The R_{kl} matrix has one singular eigenvalue. Apparently one of the two boundary conditions was redundant. This problem must be resolved before any numerical calculations can be done. The condition that the boundary pixels should be approximately equal to the interior pixels at the boundary be maintained. This can be done by making a slight change in the boundary conditions:

$$u_0 \approx 0.999u_1 \qquad \text{and} \qquad u_\nu \approx 1.001u_{\nu+1}. \tag{30}$$

Making this slight change removes the singular eigenvalue and allows the prior to be normalized, without changing the spirit of the boundary condition.

1.2 Eliminating Nuisance Parameters

Now that the prior has been specified, it may be substituted into the posterior probability for pixel u_j, Eq. (10), to obtain

$$\begin{aligned} P(u_j|\epsilon,\sigma,D,I) \;\propto\; & \int \underbrace{\cdots du_i \cdots}_{i\neq j} [\lambda_0\cdots\lambda_{\nu+1}]^{\frac{1}{2}}\,\sigma^{-(N+\nu+2)}\epsilon^{\nu+2} \\ \times\; & \exp\left\{-\frac{\epsilon^2}{2\sigma^2}\sum_{k=0}^{\nu+1}\sum_{l=0}^{\nu+1}R_{kl}u_k u_l\right\} \\ \times\; & \exp\left\{-\frac{1}{2\sigma^2}\sum_{\substack{i=1\\ \text{by }\beta}}^{\nu}[d_i-u_i]^2\right\}. \end{aligned} \tag{31}$$

The integrals are over all pixels, except $i = j$. There are $\nu+1$ integrals to evaluate.

To evaluate these integrals, the exponent in the likelihood is squared to obtain:

$$\begin{aligned} P(u_j|\epsilon,\sigma,D,I) \;\propto\; & \int \underbrace{\cdots du_i \cdots}_{i\neq j} [\lambda_0\cdots\lambda_{\nu+1}]^{\frac{1}{2}}\sigma^{-(\nu+N+2)}\epsilon^{\nu+2} \\ \times\; & \exp\left\{-\frac{1}{2\sigma^2}\left[N\overline{d^2} - 2\sum_{\substack{i=1\\ \text{by }\beta}}^{\nu} d_i u_i + \sum_{k=0}^{\nu+1}\sum_{l=0}^{\nu+1} g_{kl}u_k u_l\right]\right\} \end{aligned} \tag{32}$$

where $\overline{d^2}$ is the mean-square data value, defined as

$$\overline{d^2} \equiv \frac{1}{N} \sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu} d_i^2, \tag{33}$$

and the interaction matrix g_{kl} is defined as

$$g_{kl} \equiv \epsilon^2 R_{kl} + S_{kl} \qquad 0 \leq k, l \leq \nu + 1. \tag{34}$$

The matrix S_{kl} is diagonal and defined as

$$S_{kl} = \begin{cases} 1 & \text{If } k = l \text{ and } \operatorname{mod}(k-1, \beta) = 0 \\ \\ 0 & \text{otherwise,} \end{cases} \tag{35}$$

where "$\operatorname{mod}(k-1, \beta) = 0$" means that $(k-1)$ is evenly divisible by β.

There is no integral over u_j, consequently u_j behaves like a constant. Separating u_j from the integration variables one has

$$\begin{aligned} P(u_j|\epsilon, \sigma, D, I) \quad &\propto \quad \int \underbrace{\cdots du_i \cdots}_{i \neq j} [\lambda_0 \cdots \lambda_{\nu+1}]^{\frac{1}{2}} \sigma^{-(N+\nu+2)} \epsilon^{\nu+2} \\ &\times \quad \exp\left\{ -\frac{N\overline{d^2} - 2d_j u_j z + g_{jj} u_j^2}{2\sigma^2} \right\} \\ &\times \quad \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{\substack{k=0 \\ k \neq j}}^{\nu+1} \sum_{\substack{l=0 \\ l \neq j}}^{\nu+1} g_{kl} u_k u_l - 2 \sum_{\substack{i=1 \\ \text{by } \beta \\ i \neq j}}^{\nu} [d_i - g_{ij} u_j] u_i \right] \right\} \end{aligned} \tag{36}$$

where

$$z \equiv \begin{cases} 1 & \text{if } j = \{1, \beta+1, 2\beta+1, \ldots, \nu\} \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

Now that the dependence on u_j has been separated from the integration variables, the integrals may be done by the following change of variables:

$$A_k = \sqrt{\lambda'_k} \sum_{\substack{i=0 \\ i \neq j}}^{\nu+1} u_i e_{ki} \qquad (k \neq j), \tag{38}$$

where the u_k are given by

$$u_k = \sum_{\substack{i=0 \\ i \neq j}}^{\nu+1} \frac{A_i e_{ik}}{\sqrt{\lambda'_i}} \qquad (k \neq j), \tag{39}$$

and λ'_i is the ith eigenvalue of the jth cofactor of the g_{ik} matrix, Eq. (34), and e_{ik} is the kth component of its ith eigenvector. As a reminder, the jth cofactor of a square matrix of rank $\nu + 2$ is a square matrix of rank $\nu + 1$. The cofactor is formed by deleting the jth row and column from Eq. (34). Note that in defining the cofactor matrix the indices have not been relabeled; they still run from zero to $\nu + 1$; however, the jth item no longer exists and must be skipped in all summation. This will be noted in the equations where applicable. These new integration variables have the property that

$$\sum_{\substack{k=0 \\ k \neq j}}^{\nu+1} g_{lk} e_{ik} = \lambda'_i e_{il} \qquad (i, l \neq j), \tag{40}$$

and

$$\sum_{\substack{k=0 \\ k \neq j}}^{\nu+1} e_{lk} e_{ik} = \delta_{li} \qquad (i, l \neq j) \tag{41}$$

where δ_{li} is the Kronecker delta function. The volume element of the transformation is given by

$$\frac{dA_0 \cdots dA_{j-1} dA_{j+1} \cdots dA_{\nu+1}}{\sqrt{\lambda'_0 \cdots \lambda'_{j-1} \lambda'_{j+1} \cdots \lambda'_{\nu+1}}} = du_0 \cdots du_{j-1} du_{j+1} \cdots du_{\nu+1}. \tag{42}$$

Making the change of variables and introducing a new quantity $h_l(u_j)$:

$$h_l(u_j) \equiv \frac{1}{\sqrt{\lambda'_l}} \sum_{\substack{i=1 \\ \text{by } \beta \\ i \neq j}}^{\nu} [d_i - g_{ij} u_j] e_{li} \qquad (l \neq j) \tag{43}$$

one obtains

$$\begin{aligned} P(u_j|\epsilon, \sigma, D, I) \quad &\propto \quad \sigma^{-(N+\nu+2)} \epsilon^{\nu+2} \exp\left\{ -\frac{N\overline{d^2} - 2d_j u_j z + g_{jj} u_j^2 - h(u_j) \cdot h(u_j)}{2\sigma^2} \right\} \\ &\times \quad \int \underbrace{\cdots dA_i \cdots}_{i \neq j} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{\substack{i=0 \\ i \neq j}}^{\nu+1} [A_i - h_i(u_j)]^2 \right\} \end{aligned} \tag{44}$$

where the square on the quadratic terms was completed, some factors of 2π were dropped, the determinant (which is a constant here) was also dropped. The quantity $h(u_j) \cdot h(u_j)$ is defined as

$$h(u_j) \cdot h(u_j) \equiv \sum_{\substack{i=0 \\ i \neq j}}^{\nu+1} h_i(u_j)^2. \tag{45}$$

Evaluating the $\nu + 1$ integrals gives a factor of $(2\pi\sigma^2)^{(\nu+1)/2}$, and one obtains

$$P(u_j|\sigma, \epsilon, D, I) \propto \sigma^{-(N+1)} \epsilon^{\nu+2} \exp\left\{ -\frac{N\overline{d^2} - 2d_j u_j z + g_{jj} u_j^2 - h(u_j) \cdot h(u_j)}{2\sigma^2} \right\} \tag{46}$$

as the posterior probability for the jth pixel. If as assumed so far, the variance of the noise and the value of the fractional variance ϵ are actually known, then there are a number of additional terms that are constants and these constants will cancel when the distribution is normalized. Dropping these terms, one obtains

$$P(u_j|\sigma, \epsilon, D, I) \propto \exp\left\{ \frac{2d_j u_j z - g_{jj} u_j^2 + h(u_j) \cdot h(u_j)}{2\sigma^2} \right\} \tag{47}$$

as the posterior probability for the jth pixel given the standard deviation of the noise, the fractional variance, the data D, and the prior information I.

1.3 Eliminating σ As A Nuisance Parameter

In most real problems neither σ nor ϵ are known; they are nuisance parameters and should be treated as such. This is easy for σ, but ϵ is more difficult to deal with. To make inferences about u_j independent of σ we apply the sum rule to obtain

$$P(u_j|\epsilon, D, I) = \int d\sigma P(u_j, \sigma|\epsilon, D, I). \tag{48}$$

The right-hand-side of this equation may be factored to obtain

$$\begin{aligned} P(u_j, \sigma|\epsilon, D, I) &= P(u_j, \sigma|\epsilon, I)P(D|u_j, \sigma, \epsilon, I) \\ &= P(u_j|I)P(\sigma|I)P(D|u_j, \sigma, \epsilon, I) \\ &= P(\sigma|I)P(u_j|D, \sigma, \epsilon, I) \end{aligned} \tag{49}$$

where it was assumed that the prior probability, $P(u_j, \sigma|\epsilon, I)$, was independent of ϵ and that $P(u_j, \sigma|I) = P(u_i|I)P(\sigma|I)$. Inserting this result into Eq. (48) one obtains

$$P(u_j|\epsilon, D, I) = \int d\sigma P(\sigma|I)P(u_j|\sigma, \epsilon, D, I) \tag{50}$$

where $P(\sigma|I)$ is the prior probability for the variance, and $P(u_j|\sigma, \epsilon, D, I)$ is proportional to Eq. (46).

The posterior probability for u_j may be computed provided a prior is assigned to the noise standard deviation. Having no specific information about σ, a Jeffreys prior $1/\sigma$ [4] is assigned to obtain:

$$P(u_j|\epsilon, D, I) \propto \int_0^\infty d\sigma \sigma^{-(N+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[N\overline{d^2} - 2d_j u_j z + g_{jj}u_j^2 - h(u_j)\cdot h(u_j)\right]\right\}. \tag{51}$$

Evaluating the integral, one obtains

$$P(u_j|\epsilon, D, I) \propto \left[1 - \frac{h(u_j)\cdot h(u_j) + 2d_j u_j z - g_{jj}u_j^2}{N\overline{d^2}}\right]^{-\frac{N}{2}}. \tag{52}$$

This is a Student t-distribution, and it is this result that is applied in the numerical examples.

Suppose a simple experiment has been run for 100 seconds and a data item was gathered every second, thus obtaining $N = 100$ data samples. Suppose the data gathered in this experiment looked like that shown in Fig. 1 (a constant signal of value 5, plus Gaussian white noise of standard deviation of 1). In the calculation so far, only one pixel may be estimated at a time. But any pixel may be estimated, so all of them may be estimated. In this numerical example, $j = 59$ will be used. At the end of the example, the results will be shown for all of the pixels. To estimate u_{59} one needs only to apply the posterior probability for the pixels. But this probability density function assumes the value of ϵ is known and the estimated pixel value depends on what value of ϵ is chosen. Before the pixel value can be estimated, a procedure must be developed that allows one to estimate or set ϵ to a reasonable value.

1.4 Estimating The Regularizer ϵ

If one follows the rules of probability theory exactly, the way to proceed is to multiply the probability for the pixel given the value of ϵ, by a prior probability for ϵ and integrate. Unfortunately, ϵ appears

Figure 1: Interpolation – The Data

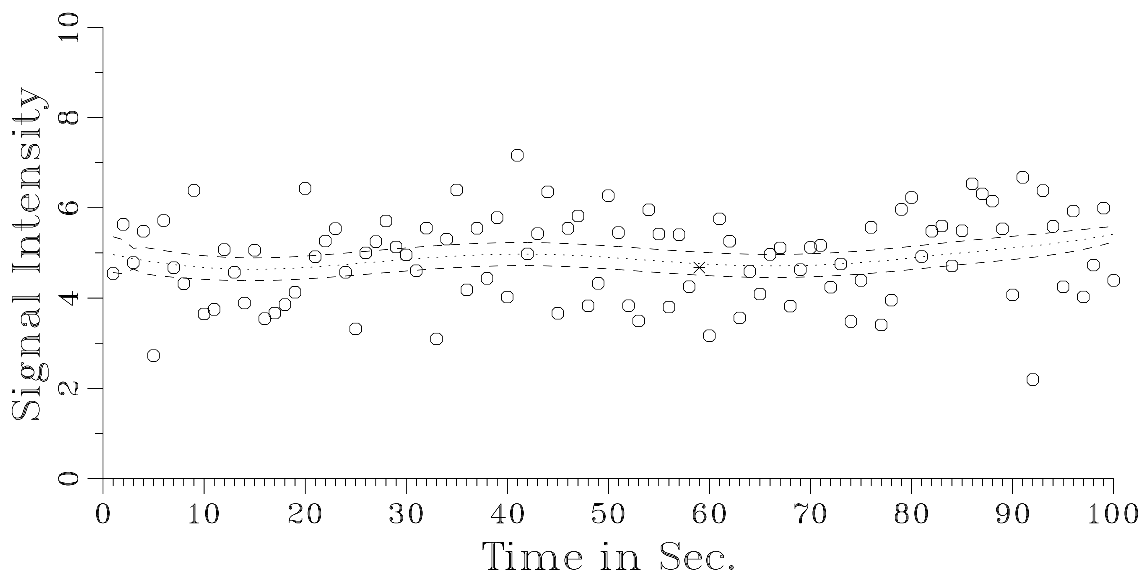

Fig. 1. The data contain a constant signal of value 5, plus noise of standard deviation one. The problem is to make the best estimate of a pixel given only the information that the function must be smooth and these data.

in the problem in a very nonlinear way and evaluating the integral in closed form has not proven possible. However, there are approximations which will allow one to proceed and obtain results that are nearly identical to the exact procedure. If the joint posterior probability for pixel u_{59} and ϵ is sharply peaked, then removing the regularizer by integration, essentially just constrains the regularizer to its value at the maximum of the joint posterior probability. If the value of the regularizer near the maximum can be determined, then ϵ can be constrained to this value in Eq. (52). The results obtained will be nearly identical to what would have been obtained by removing ϵ by integration [7].

To determine a reasonable value of ϵ, the probability density for the regularizer will be computed. From this probability density function one can locate the value of ϵ for which the posterior is maximized. This maximum may be used in Eq. (52) to obtain the posterior probability for the pixels. The estimated pixel value are dependent on the value of ϵ, so it is important that a value near the most probable value be used when estimating the pixels.

To illustrate that a good estimate of ϵ is necessary, consider Fig. 2. Here two different values of ϵ were used: one small and one large. In panel 2(A), $\epsilon = 0.01$. The data values are shown as open circles, and the reconstruction is shown as the solid line. The pixel estimates plotted are the mean or expected values of the pixels. These were computed using the procedures developed in Section 4.1. For now it is enough to know that the values are just the ones given by the maximum of the posterior probability for the pixels, given the value of ϵ, Eq. (52). For small ϵ, the prior information is essentially irrelevant, and the pixels are estimated to be equal to the data values. This effect is seen in panel 2(A), where the reconstruction follows the data almost exactly. The opposite effect occurs when $\epsilon \to \infty$. Here the prior is important and the data are irrelevant, and the pixels are estimated to be a constant, zero. Somewhere between these two extreme values is region which is appropriate for this problem.

Figure 2: The Estimated Pixels As A Function Of ϵ

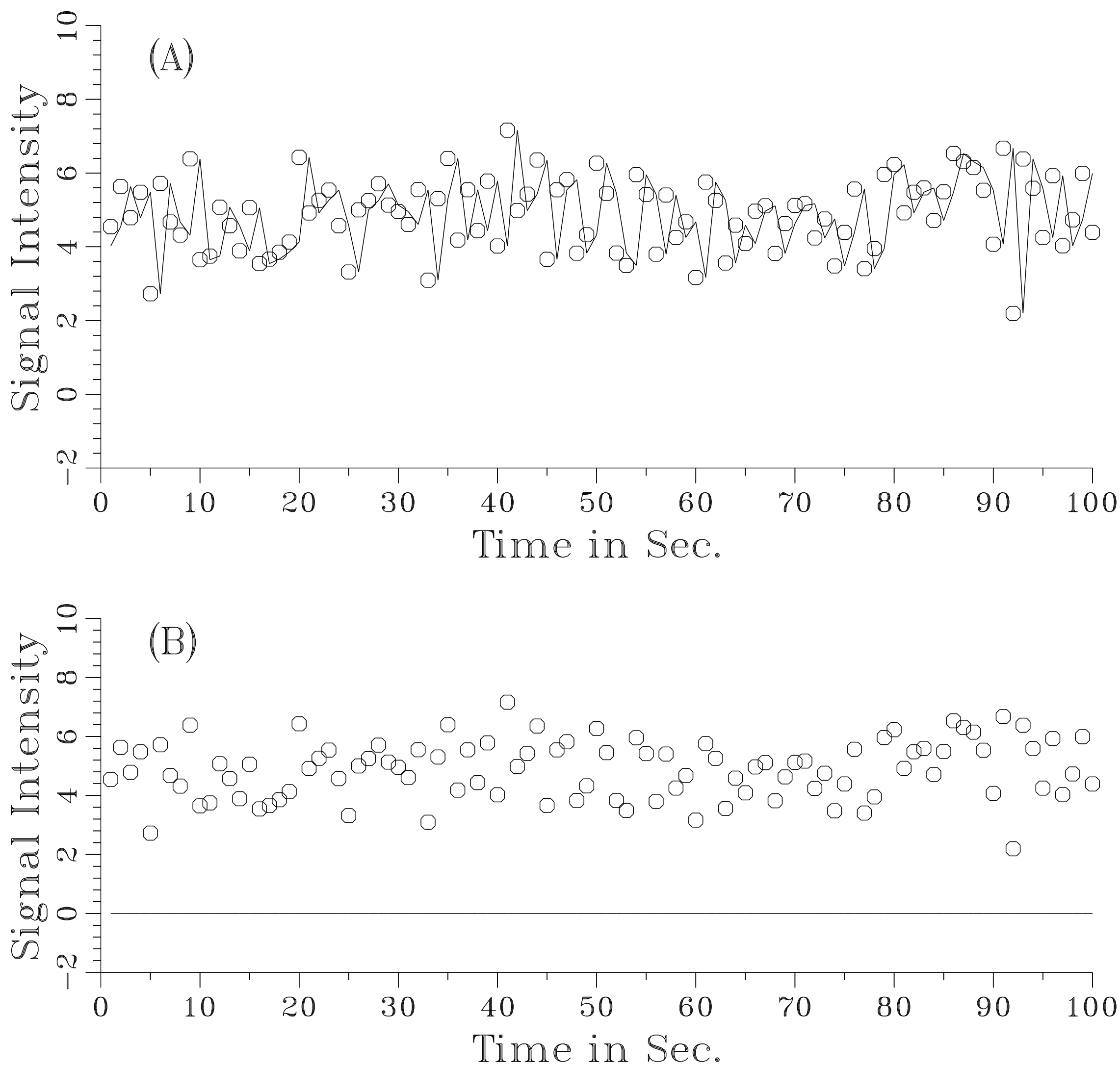

Fig. 2. Panel (A) contains the data (open circles) and the estimated pixel values (solid line) for $\epsilon = 0.01$. Here ϵ is too small and the reconstruction pays too much attention to the data. In panel (B), $\epsilon = 1,000,000$, and is too large; the estimated pixels (solid line) does not pay enough attention to the data (open circles).

To find this region, one can compute the posterior probability for ϵ. Using the sum rule from probability theory, this is given by

$$P(\epsilon|D,I) = \int du_0 \cdots du_{\nu+1} d\sigma P(\epsilon,\sigma,u_0,\ldots,u_{\nu+1}|D,I). \tag{53}$$

The integrand can be factored using the same steps shown in Eq. (48) to obtain

$$P(\epsilon|D,I) = \int du_0 \cdots du_{\nu+1} d\sigma P(\epsilon,\sigma|I) P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma,D,I) \tag{54}$$

where $P(\epsilon,\sigma|I)$ is the joint prior probability for ϵ and σ. Further, $P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma,D,I)$ can be factored to obtain

$$\begin{aligned} P(\epsilon|D,I) &= \int du_0 \ldots du_{\nu+1} d\sigma P(\epsilon|I) P(\sigma|I) \\ &\times \; P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma,I) P(D|\epsilon,\sigma,u_0,\ldots,u_{\nu+1},I) \end{aligned} \tag{55}$$

where $P(u_0,\ldots,u_{\nu+1}|\epsilon,\sigma,I)$ is the prior probability for all of the pixels given ϵ, σ, and the prior information I; and it is given by Eq. (18), $P(D|\epsilon,\sigma,u_0,\ldots,u_{\nu+1},I)$ is the likelihood for the data and is given by Eq. (9), and $P(\sigma|I)$ is the prior probability for σ and was assumed independent of ϵ. Substituting Eq. (9) for the likelihood, Eq. (18), for the prior probability for the pixels and a Jeffreys prior for both ϵ and σ one obtains:

$$\begin{aligned} P(\epsilon|D,I) &\propto \int du_0 \ldots du_{\nu+1} d\sigma [\lambda_0 \cdots \lambda_{\nu+1}]^{\frac{1}{2}} \sigma^{-(\nu+N+3)} \epsilon^{\nu+1} \\ &\times \; \exp\left\{ -\frac{\epsilon^2}{2\sigma^2} \sum_{k=0}^{\nu+1} \sum_{l=0}^{\nu+1} R_{kl} u_k u_l \right\} \\ &\times \; \exp\left\{ -\frac{1}{2\sigma^2} \sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu} [d_i - u_i]^2 \right\} \end{aligned} \tag{56}$$

where the eigenvalues $\{\lambda_0,\ldots,\lambda_{\nu+1}\}$ must now be kept, because they are functions of ϵ.

To evaluate these $\nu+3$ integrals ($\nu+2$ integrals over the u_i, and one over σ) the quadratic in the likelihood is expanded to obtain something very much like Eq. (32):

$$\begin{aligned} P(\epsilon|D,I) &\propto \int u_1,\ldots,u_\nu \; d\sigma [\lambda_0 \cdots \lambda_{\nu+1}]^{\frac{1}{2}} \sigma^{-(\nu+N+1)} \epsilon^{\nu+1} \\ &\times \; \exp\left\{ -\frac{1}{2\sigma^2} \left[N\overline{d^2} - 2 \sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu} d_i u_i + \sum_{k=0}^{\nu+1} \sum_{l=0}^{\nu+1} g_{kl} u_k u_l \right] \right\}, \end{aligned} \tag{57}$$

where g_{kl} was defined earlier in Eq. (34). Unlike what was done earlier, here there are $\nu+2$ integrals over all of the u_i. Thus no intermediate steps are involved where the cofactor of g_{kl} was defined. All that is necessary is that the g_{kl} matrix be diagonalized.

In the process of doing these calculations, several matrices will have to be diagonalized, and the procedures for doing so are all essentially the same. One introduces a new set of integration variables based on the singular-value decomposition of the interaction matrix, and transforms to the new variables. In these variables all of the Gaussian quadrature integrals separate and may be done trivially. Because all of these integrations are very similar, the details will be omitted and only the results of the calculations given. In this case, after having evaluated the $\nu+2$ integrals the posterior probability for ϵ independent of the pixel values is given by

$$
\begin{aligned}
P(\epsilon|D,I) &= \int d\sigma \left(\frac{\lambda_0\cdots\lambda_{\nu+1}}{\lambda'_0\cdots\lambda'_{\nu+1}}\right)^{\frac{1}{2}} \sigma^{-(N+1)}\epsilon^{\nu+1} \\
&\times \exp\left\{-\frac{N\overline{d^2}-h(\epsilon)\cdot h(\epsilon)}{2\sigma^2}\right\}
\end{aligned}
\tag{58}
$$

where

$$
h_l(\epsilon) \equiv \frac{1}{\sqrt{\lambda'_l}} \sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu} d_i e_{li}, \tag{59}
$$

$$
h(\epsilon)\cdot h(\epsilon) \equiv \sum_{i=0}^{\nu+1} h_i(\epsilon)^2, \tag{60}
$$

$\{\lambda_0,\ldots,\lambda_{\nu+1}\}$ are the eigenvalues of the R_{ik} matrix defined in Eq. (29) and $\{\lambda'_0,\ldots,\lambda'_{\nu+1}\}$ and e_{li} are the eigenvalues and eigenvectors of the g_{ik} matrix defined in Eq. (34).

The remaining integral is very similar to what was done earlier, Eq. (52), when σ was removed and again only the results are given here

$$
P(\epsilon|D,I) \propto \left(\frac{\lambda_0\cdots\lambda_{\nu+1}}{\lambda'_0\cdots\lambda'_{\nu+1}}\right)^{\frac{1}{2}} \epsilon^{\nu+1}\left[1-\frac{h(\epsilon)\cdot h(\epsilon)}{N\overline{d^2}}\right]^{-\frac{N}{2}}. \tag{61}
$$

When $\epsilon\to 0$, there is effectively no prior, and the pixel estimates go to data values. However, when $\epsilon\to\infty$, the prior dominates and forces the second derivative to zero and the reconstruction goes to a constant. As $\epsilon = 0$ the likelihood term [the term in square brackets in Eq. (61)] is going to infinity like ϵ^{-N}. However, the prior term (essentially $\epsilon^{\nu+1}$) is going to zero at exactly the same time. Somewhere between these two extreme values there lies a maximum in the posterior probability that acts as a trade off between the prior and the likelihood.

Figure 1 contains a simple data set with $N = 100$ data values. The “signal” in these data is a constant of value 5, plus additive white noise of standard deviation 1. Using the procedures derived so far, the value of the 59’th pixel is to be estimated. As was mentioned earlier, before the value of pixel u_{59} may be estimated, one must set the value of ϵ. Using the posterior probability for ϵ, this may now be done. This probability density function is plotted in Fig. 3. This probability distribution has a well defined maximum near 70, and a mean value of approximately 93. Note that for values of ϵ smaller than 10 and larger than 270, the probability for ϵ is essentially zero. So whatever value of ϵ is used, it should be somewhere in these bounds.

Figure 4 contains the posterior probability for u_{59} given $\epsilon = 10, 80, 93,$ and 200. Note that for the maximum and mean (Panels B and C), the posterior probabilities are almost identical. However, when ϵ is too small (Panel A), the posterior probability is smeared out and broad; on the other hand, when ϵ is too large (Panel D), the posterior probability is too narrow. It is interesting to note that as $\epsilon\to 0$ the width of the posterior probability becomes large, while the estimated pixel values go to the data values. Estimating the pixels to be equal to the data is the maximum likelihood result. In this limit, there is no prior information about the signal. Probability theory is warning you that there is no way to differentiate between the signal and the noise; the signal could be anything consistent with the total mean-square data value. In the other limit, $\epsilon\to\infty$, deviations from a constant are not allowed. Essentially the results goes to the mean $\pm$ standard deviation estimate of the constant.

Figure 3: The Posterior Probability for ϵ

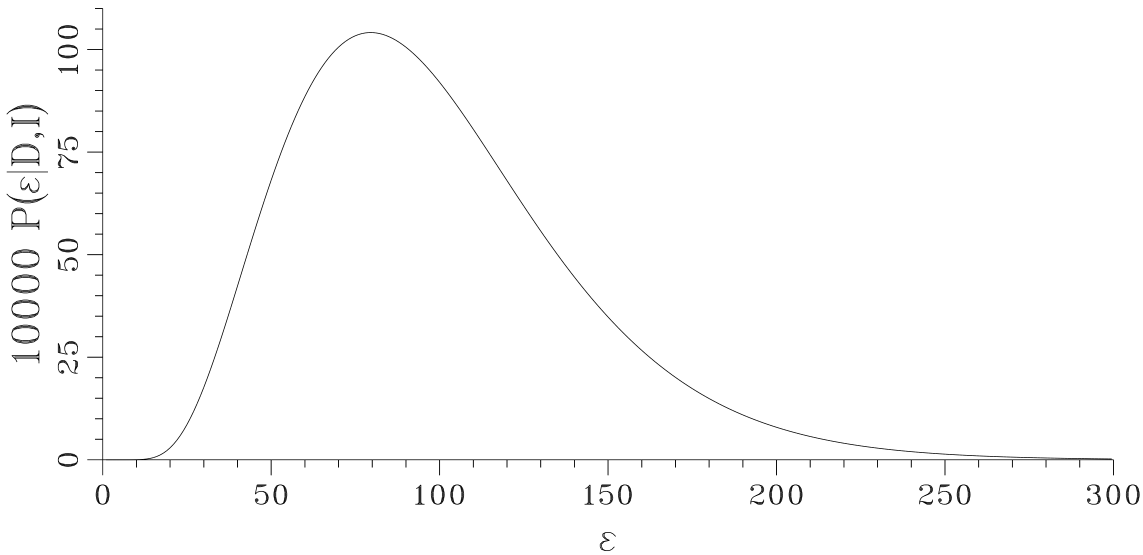

Fig. 3. The posterior probability for ϵ was computed using the constraint on the second derivative. This probability density function has a well defined maximum with a peak near $\epsilon \approx 80$, and a mean value of 93.

2 Data Interpolation – General Prior Information

Before proceeding to the deconvolution problem, the data interpolation problem will be generalized to include other types of prior information. Three types of prior information will be included: information about the functional form of the signal and about its first and second derivatives. As was demonstrated in the previous section, what differentiates the results of a probability theory calculation from a maximum likelihood or least squares calculation is the presence of the prior probability. In the previous section only prior information about the second derivative was used, here three different types of prior information will be used. To utilize all of this information there are two tasks that must be completed: first, each of these three pieces of information must be formulated into a prior probability, and second, these different priors must be combined into a single prior which expresses all three pieces of information.

To see how to convert each of the three types of prior information into a prior probabilities, suppose the signal is known to be sinusoidal. The total difference between the signal and the data is given by

$$\sum_{i=1}^{N} [u_i - A\cos(\omega t_i + \theta)]. \tag{62}$$

What is actually known about this difference? Would one expect this to be zero, positive, or negative? If the signal is known to be more or less sinusoidal, then on average one would expect the difference to be small and its value could be either positive or negative. So the prior information is consistent with a zero mean value: i.e., no information is available that would lead us to expect this difference to be either positive or negative upon repeating the experiment many times. Second, the mean-square difference is expected to be nonzero; i.e., we expect some deviations from the model. Now the principle of maximum entropy can be used to assign a probability density function to this

Figure 4: The Posterior Probability for Pixel u_{59}

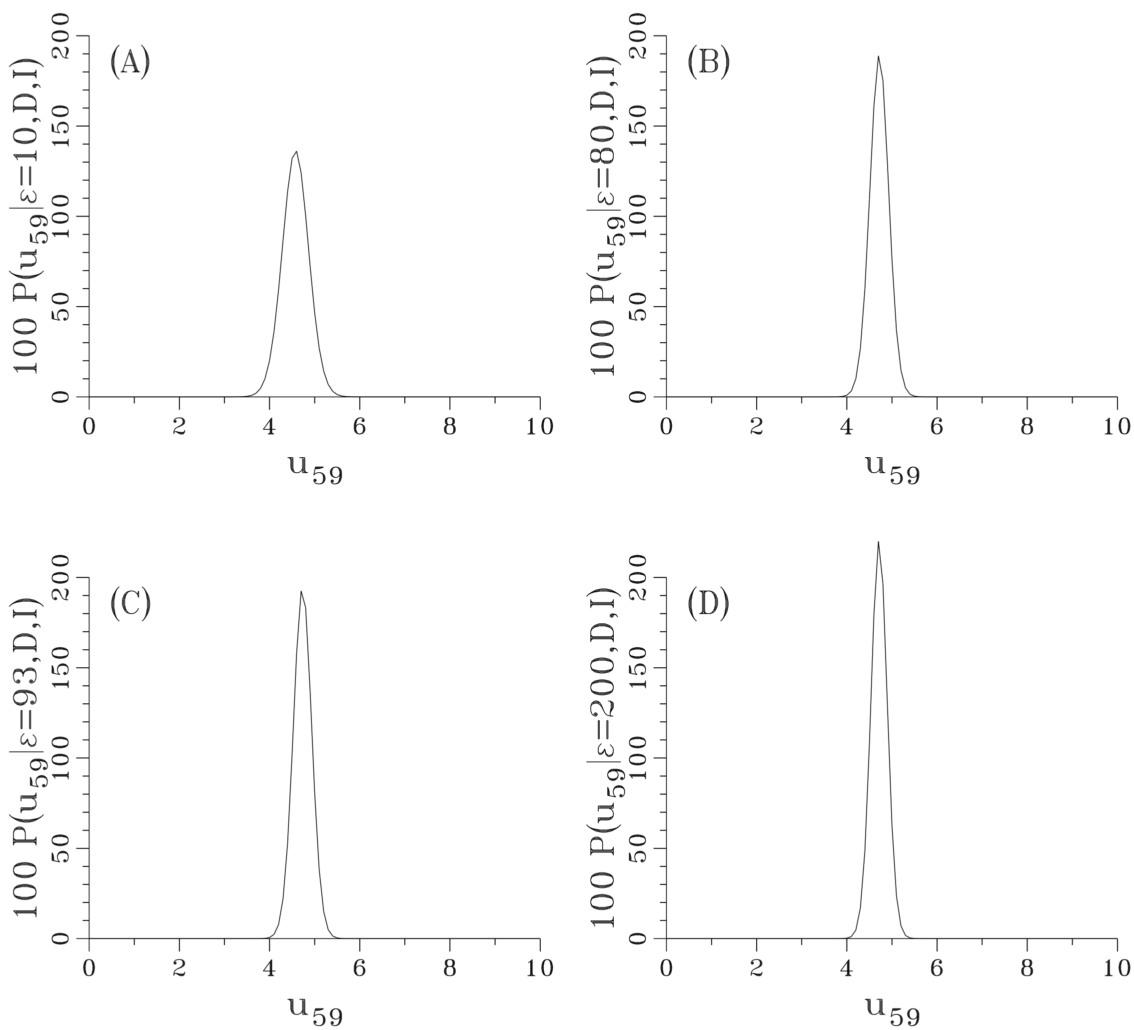

Fig. 4. The posterior probability for u_{59} is shown for $\epsilon = 10$ panel (A), $\epsilon = 80$ panel (B), $\epsilon = 93$ panel (C), and $\epsilon = 200$ panel (D). Panels (B) and (C) correspond to the peak and expected values of ϵ.

difference. But note that this difference is not necessarily noise, it merely reflects our uncertainty in the actual functional form of the signal. When maximum entropy is applied, it will assign a Gaussian prior to this difference. Because for a fixed mean-square the Gaussian has highest entropy, and is therefore the least informative distribution possible. From the Gaussian distribution one can assign a prior probability for the difference between the pixels and the model. For this sinusoidal example, this probability density function is given by

$$P(u_i|A,\omega,\theta,\epsilon,\sigma,I_1) = \sigma^{-N}\epsilon^N \exp\left\{-\frac{\epsilon^2}{2\sigma^2}\sum_{i=1}^{N}\ [u_i - A\cos(\omega t_i + \theta)]^2\right\}, \tag{63}$$

where the parameter ϵ measures the amount of misfit between the pixels and the model. As occurred in the previous example, this prior has introduced a number of additional parameters: A an amplitude, θ a phase, ω a frequency, and a fractional variance ϵ^2. Some of these parameters may be known, but more likely either they will have to be eliminated from the problem, or inferences will have to be made about them. For the time being, no assumptions will be made, and the problem will be formulated in a way that either they may be eliminated as nuisances or inferences may be made about them.

2.1 Formulating The Prior Probability

Three types of prior information will be included in this generalization of the interpolation problem: information on the functional form, and on the first and second derivatives. These will be labeled I_1, I_2 and I_3 respectively. The prior for each of these will be formulated separately and then combined into a single prior for use in the generalized interpolation calculation.

Information I_1 will be addressed first. This information assumes that something is known about the functional form of the signal. The functional form will be written as $Af_1(t_i)$, where A is an amplitude and, for example, $f_1(t_i)$ might be a sinusoid. The total mean-square difference δ_1^2 between the model and the pixels is given by

$$\sum_{i=1}^{\nu} [u_i - Af_1(t_i)]^2 = \delta_1^2. \tag{64}$$

If $\delta_1 = 0$, the model must follow the functional form exactly. If $\delta_1 \to \infty$ then the total squared difference goes to infinity and the reconstruction will follow the data.

Using information I_1 in a maximum entropy calculation results in assigning a Gaussian prior

$$P(u_0,\ldots,u_{\nu+1}|A,\epsilon_1,\sigma,I_1) \propto \exp\left\{-\frac{\epsilon_1^2}{\sigma^2}\sum_{i=0}^{\nu+1}[u_i - Af_1(t_i)]^2\right\} \tag{65}$$

where ϵ_1 is the fractional variance associated with information I_1. As was noted earlier, the prior has introduced two new parameters: A, ϵ_1. Last, note that the prior has not yet been normalized, this will be done after combining the three priors.

Information I_2 specifies how the first derivative is to behave. Assuming the functional form of the first derivative is given by $Bf_2(t)$ then

$$\sum_{i=1}^{\nu}\ [u_{i+1} - u_{i-1} - 2Bf_2(t_i)]^2 = \delta_2^2 \tag{66}$$

where B is an amplitude, and δ_2 is the total squared difference. Using this as a constraint in a maximum entropy calculation allows us to assign a prior probability to the difference between the modeled derivative and the pixels:

$$P(u_1,\ldots,u_\nu|B,u_0,u_{\nu+1},\epsilon_2,\sigma,I_2)\propto\exp\left\{-\frac{\epsilon_2^2}{2\sigma^2}\sum_{i=1}^{\nu}\left[u_{i+1}-u_{i-1}-2Bf_2(t_i)\right]^2\right\} \tag{67}$$

where two additional parameters, B the amplitude, and ϵ_2^2 the fractional variance have been introduced.

Information I_3 specifies how the second derivative is to behave. Assuming the functional form of the second derivative is given by $Cf_3(t_i)$, one has

$$\sum_{i=1}^{\nu}\left[u_{i+1}+u_{i-1}-2u_i-Cf_3(t_i)\right]^2=\delta_3^2 \tag{68}$$

where C is an amplitude associated with the second derivative, $f_3(t)$ is its functional form, and δ_3^2 is the total squared difference. Repeating the maximum entropy calculation gives

$$P(u_1,\ldots,u_\nu|C,u_0,u_{\nu+1},\epsilon_3,\sigma,I_3)\propto\exp\left\{-\frac{\epsilon_3^2}{2\sigma^2}\sum_{i=1}^{\nu}\left[u_{i+1}+u_{i-1}-2u_i-Cf_3(t_i)\right]^2\right\} \tag{69}$$

as the prior probability for the pixels given information I_3, where C is the amplitude, and ϵ_3^2 is the associated fractional variance.

Note that three unknown amplitudes A, B, and C, three fractional variances ϵ_1^2, ϵ_2^2 and ϵ_3^2, and two boundary pixels u_0 have entered the problem. The three amplitudes and all of the unknown pixels will be eliminated from the problem as nuisance parameters. In this problem it is critically important to ensure that proper priors are used. A proper prior is one which is normalizable. Improper priors are ones which cannot be normalized. Strictly speaking a function that cannot be normalized is not a probability density function. Two examples of improper priors are the Jeffreys prior and the uniform prior. The Jeffreys prior is improper when the limits on the parameter are taken from zero to infinity. The uniform prior is improper whenever one of the limits is taken to infinity. In spite of this the use of improper or unnormalizable prior probabilities in parameter estimation is often convenient and harmless. However, *in this problem the use of improper priors must be avoided because the normalization factor associated with the prior does not always cancel.* Consequently a normalized prior must be used for A, B and C as well as for all of the pixels. These parameters are location parameters, and the prior which correctly express information about a location parameter is a Gaussian. Consequently, the prior for the three amplitudes A, B, and C will be taken as

$$P(A,B,C|\epsilon_0,I_{\rm old})=(2\pi\sigma^2)^{-3/2}\epsilon_0^3\exp\left\{-\frac{\epsilon_0^2}{2\sigma^2}\left[A^2+B^2+C^2\right]\right\}, \tag{70}$$

where $I\to I_{\rm old}$ was made to differentiate I from I_1, I_2, and I_3. This prior says that the three amplitudes may be either positive or negative and we do not know which it is. If ϵ_0 is small, then this prior does not express a strong opinion about the amplitudes, other than small absolute magnitude is preferred. It will be assumed that ϵ_0 is set from prior information and that its actual value is known. So long as this value is small, the only purpose served by the prior is to prevent any singular mathematics from occurring because. So whatever value is assigned to ϵ_0, it will not change the results, provided it $\epsilon_0\ll\sigma$.

Just as information I_1, I_2 and I_3 can be used to constrain the interior pixels, they may also be used to constrain the boundary pixels. Information I_1 specified a functional form for the signal.

There is no reason that $f_1(t_i)$ cannot be evaluated at the boundary. This will give

$$P(u_0|\epsilon_1, I_1) \propto \exp\left\{-\frac{\epsilon_1^2}{2\sigma^2}[u_0 - Af_1(t_0)]^2\right\} \tag{71}$$

and

$$P(u_{\nu+1}|\epsilon_1, I_1) \propto \exp\left\{-\frac{\epsilon_1^2}{2\sigma^2}[u_{\nu+1} - Af_1(t_{\nu+1})]^2\right\} \tag{72}$$

as the prior probabilities for the lower and upper boundary pixels.

Information I_2 specified prior information about the first derivative. When the first derivative was defined, a symmetric difference equation was written, Eq. (12). This symmetric difference cannot be used on the boundary because it would introduce still more unknown parameters. However, a *forward* and *backward* first derivative can be used. For u_0 one has

$$P(u_0|\epsilon_2, \sigma, I_2) \propto \exp\left\{-\frac{\epsilon_2^2}{2\sigma^2}[u_1 - u_0 - Bf_2(t_0)]^2\right\}. \tag{73}$$

Similarly a backward first derivative may be defined and used to formulate a prior for $u_{\nu+1}$. This prior is given by

$$P(u_{\nu+1}|\epsilon_2, \sigma, I_2) \propto \exp\left\{-\frac{\epsilon_2^2}{2\sigma^2}[u_{\nu+1} - u_\nu - Bf_2(t_{\nu+1})]^2\right\}. \tag{74}$$

But just as occurred earlier, care must be taken here because these boundary conditions are not enough to make the prior associated with I_2 normalizable. So when this prior is actually programmed the constraints will have to be modified just enough to make the matrices associated with them nonsingular.

The last information I_3 may be used to specify a prior on the two boundary pixels. However, now we have a functional form for the second derivative. The second derivative cannot be interpreted as specifying that neighboring pixels be approximately equal. That interpretation was possible because the prior information assumed adjacent pixels were approximately equal. Here we could specify a forward and backward second derivative, but that will not work because the forward second derivative at t_0 is the same as the symmetric second derivative at time t_1, consequently we would be trying to constrain the same quantity to two different values. Without doing something much more complicated, there is no easy way to constrain the boundary pixels using the functional form of the second derivative. This is not a problem, because I_1 and I_2 have already supplied more than enough information to form a normalizable prior for the boundary pixels.

2.2 Combining Different Prior Information

From the previous subsection there are ten probability density functions expressing prior information about the pixels $(u_0, \ldots, u_{\nu+1})$ and the amplitudes. What is needed is a single prior that express the information contained in all ten of these priors. The process of combining these priors is begun by adopting some new notation.

There are three amplitudes, two boundary pixels, and ν interior pixels. There are $\nu + 5$ total parameters (excluding the three fractional variances). All of them but one are to be eliminated as nuisances. To facilitate this, the pixel values $\{u_0, \ldots, u_{\nu+1}\}$ and the three amplitudes A, B and C

will be taken as a collection V. The elements, v_i, are defined as:

$$v_i = \begin{cases} u_i & \text{if } 0 \leq i \leq \nu + 1, \\ A & \text{if } i = \nu + 2, \\ B & \text{if } i = \nu + 3, \\ C & \text{if } i = \nu + 4. \end{cases} \tag{75}$$

In this notation the posterior probability for the jth pixel is given by

$$P(v_j|\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \sigma, D, I) \propto \int \underbrace{\cdots dv_i \cdots}_{i \neq j} P(V|\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \sigma, I)\sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2} \sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu} [d_i - v_i]^2\right\}, \tag{76}$$

where the word "pixel" will be used to refer to all of the V, including the three amplitudes A, B and C. The three regularizes have been added to the probability density function in a way that indicates their value are known. As was done earlier, if there values are not actually known probability theory will be used either estimate them or remove them.

In this problem it will be assumed that the prior information is independent and the prior probability for the pixels given the total prior information is just the product of the probabilities given the individual pieces of information:

$$\begin{aligned} P(V|\epsilon_0, I_1, I_2, I_3, I_{\text{old}}) &= P(v_0, \ldots, v_{\nu+1}|I_1, I_{\text{old}})P(v_0, \ldots, v_{\nu+1}|I_2, I_{\text{old}}) \\ &\times P(v_0, \ldots, v_{\nu+1}|I_3, I_{\text{old}})P(v_{\nu+2}, v_{\nu+3}, v_{\nu+4}|\epsilon_0, I_{\text{old}}), \end{aligned} \tag{77}$$

where $P(v_0, \ldots, v_{\nu+1}|I_1, I_{\text{old}})$ specifies the prior probability for the interior and boundary pixels given information I_1, $P(v_0, \ldots, v_{\nu+1}|I_2, I_{\text{old}})$ specifies the prior given I_2, $P(v_0, \ldots, v_{\nu+1}|I_3, I_{\text{old}})$ specifies the prior given I_3, and $P(v_{\nu+2}, v_{\nu+3}, v_{\nu+4}|\epsilon_0, I_{\text{old}})$ specifies the prior information for the three amplitudes. The independence assumption was used to factor the prior in this particular fashion.

The first three terms are all of the form $P(V|\epsilon_1, I_1, I_{\text{old}})$. These may be factored into a lower boundary prior, and interior prior and an outer boundary prior:

$$P(V|\epsilon_1, I_1, I_{\text{old}}) = P(v_0|\epsilon_1, I_1, I_{\text{old}})P(v_1, \ldots, v_\nu|\epsilon_1, I_1, I_{\text{old}})P(v_{\nu+1}|\epsilon_1, I_1, I_{\text{old}}). \tag{78}$$

These three priors were given in the previous section. This process may be repeated for information I_2 and I_3 with similar results. The remaining term, $P(v_{\nu+2}, v_{\nu+3}, v_{\nu+4}|\epsilon_0, I_{\text{old}})$, was given by Eq. (70).

For information I_1, the prior probability is given by Eq. (65)

$$P(V|\epsilon_0, \epsilon_1, \sigma, I_1, I_{\text{old}}) \propto \exp\left\{-\frac{\epsilon_0^2}{\sigma^2}A^2 - \frac{\epsilon_1^2}{\sigma^2}\sum_{i=0}^{\nu+1} [v_i - Af_1(t_i)]^2\right\}, \tag{79}$$

where the prior information for A, Eq. (70), was included in this equation. This may be rewritten in matrix form as

$$P(V|\epsilon_0, \epsilon_1, \sigma, I_1, I_{\text{old}}) \propto \exp\left\{-\frac{\epsilon_1^2}{2\sigma^2}\sum_{k=0}^{\nu+4}\sum_{l=0}^{\nu+4} W_{kl}v_k v_l\right\} \tag{80}$$

where the matrix W_{kl} is given by

$$W_{kl} \equiv \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & A_0 & 0 & 0 \\ 0 & 1 & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & A_{\nu+1} & \vdots & \vdots \\ A_0 & \cdots & \cdots & \cdots & A_{\nu+1} & A_{\nu+2} & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \tag{81}$$

with $(0 \leq k, l \leq \nu + 4)$ and

$$\begin{aligned} A_i &\equiv -f_1(t_i), \qquad (0 \leq i \leq \nu + 1) \\ A_{\nu+2} &\equiv \frac{\epsilon_0^2}{\epsilon_1^2} + \sum_{i=0}^{\nu+1} f_1(t_i)^2. \end{aligned} \tag{82}$$

Similarly for information I_2 the prior probability becomes

$$\begin{aligned} P(V|\epsilon_0, \epsilon_2, \sigma, I_2, I_{\text{old}}) &\propto \exp\left\{ -\frac{\epsilon_0^2}{\sigma^2} B^2 - \frac{\epsilon_2^2}{2\sigma^2} \sum_{i=1}^{\nu} [v_{i+1} - v_{i-1} - 2Bf_2(t_i)]^2 \right\} \\ &\times \exp\left\{ -\frac{\epsilon_2^2}{2\sigma^2} [v_1 - v_0 - Bf_2(t_0)]^2 \right\} \\ &\times \exp\left\{ -\frac{\epsilon_2^2}{2\sigma^2} [v_{\nu+1} - v_\nu - Bf_2(t_{\nu+1})]^2 \right\}. \end{aligned} \tag{83}$$

And in matrix form this prior is given by

$$P(V|\epsilon_0, \epsilon_2, \sigma, I_2, I_{\text{old}}) \propto \exp\left\{ -\frac{\epsilon_2^2}{2\sigma^2} \sum_{k=0}^{\nu+4} \sum_{l=0}^{N+4} X_{kl} v_k v_l \right\}. \tag{84}$$

The matrix X_{kl} is given by

$$X_{kl} \equiv \begin{pmatrix} 2 & -1 & -1 & 0 & \cdots & \cdots & \cdots & 0 & B_0 & 0 \\ -1 & 2 & 0 & -1 & \ddots & \ddots & \ddots & \vdots & B_1 & \vdots \\ -1 & 0 & 2 & 0 & -1 & \ddots & \ddots & \vdots & B_2 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & -1 & 0 & 2 & 0 & -1 & \vdots & B_{\nu-1} & \vdots \\ \vdots & \ddots & \ddots & -1 & 0 & 2 & -1 & \vdots & B_\nu & \vdots \\ 0 & \cdots & \cdots & 0 & -1 & -1 & 2 & \vdots & B_{\nu+1} & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 & \vdots \\ B_0 & B_1 & B_2 & \cdots & B_{\nu-1} & B_\nu & B_{\nu+1} & 0 & B_{\nu+3} & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \tag{85}$$

where $(0 \leq k, l \leq \nu + 4)$ and

$$\begin{aligned}
B_0 &= f_2(t_0) + f_2(t_1), \\
B_1 &= f_2(t_2) - f_2(t_0), \\
B_i &= f_2(t_{i+1}) - f_2(t_{i-1}), \qquad (2 \leq i \leq \nu - 1) \\
B_\nu &= f_2(t_{\nu+1}) - f_2(t_{\nu-1}), \\
B_{\nu+1} &= -[f_2(t_\nu) + f_2(t_{\nu+1})], \\
B_{\nu+3} &= \frac{\epsilon_0^2}{\epsilon_2^2} + f_2(t_0)^2 + 4\sum_{i=1}^{\nu} f_2(t_i)^2 + f_2(t_{N+1})^2.
\end{aligned} \tag{86}$$

Last, for information I_3, the prior probability for the pixels is given by

$$P(V|\epsilon_0, \epsilon_3, \sigma, I_3, I_{\text{old}}) \propto \exp\left\{-\frac{\epsilon_0^2}{\sigma^2}C^2 - \frac{\epsilon_3^2}{2\sigma^2}\sum_{i=1}^{\nu}\left[v_{i+1} + v_{i-1} - 2v_i - Cf_3(t_i)\right]^2\right\}, \tag{87}$$

which can also be written as

$$P(V|\epsilon_0, \epsilon_3, \sigma, I_3, I_{\text{old}}) \propto \exp\left\{-\frac{\epsilon_3^2}{2\sigma^2}\sum_{k=0}^{\nu+4}\sum_{i=0}^{\nu+4} Y_{kl} v_k v_l\right\} \tag{88}$$

with the matrix Y_{kl} defined as

$$Y_{kl} \equiv \begin{pmatrix}
5 & -4 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 & 0 & C_0 \\
-4 & 6 & -4 & 1 & 0 & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & C_1 \\
1 & -4 & 6 & -4 & 1 & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & C_2 \\
0 & 1 & -4 & 6 & -4 & 1 & \ddots & \ddots & \vdots & \vdots & \vdots & C_3 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \ddots & \ddots & 1 & -4 & 6 & -4 & 1 & 0 & \vdots & \vdots & C_{\nu-2} \\
\vdots & \ddots & \ddots & \ddots & 1 & -4 & 6 & -4 & 1 & \vdots & \vdots & C_{\nu-1} \\
\vdots & \ddots & \ddots & \ddots & \ddots & 1 & -4 & 6 & -4 & \vdots & \vdots & C_\nu \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & -4 & 5 & \vdots & \vdots & C_{\nu+1} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \vdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
C_0 & C_1 & C_2 & C_3 & \cdots & C_{\nu-2} & C_{\nu-1} & C_\nu & C_{\nu+1} & 0 & 0 & C_{\nu+4}
\end{pmatrix} \tag{89}$$

where $(0 \leq k, l \leq \nu + 4)$ and

$$\begin{aligned}
C_0 &= f_3(t_1) \\
C_1 &= f_3(t_2) - 2f_3(t_1) \\
C_i &= 2f_3(t_{i-1}) - 2f_3(t_i) + f_3(t_{i+1}) \qquad (2 \leq i \leq \nu - 1) \\
C_\nu &= f_3(t_{\nu-1}) - 2f_3(t_\nu) \\
C_{\nu+1} &= f_3(t_{\nu+1})] \\
C_{\nu+4} &= \frac{\epsilon_0^2}{\epsilon_3^2} + \sum_{i=1}^{\nu} f_3(t_i)^2.
\end{aligned} \tag{90}$$

Finally, combining these three priors, one obtains the probability for the pixels given $I_{\rm old}$, I_1, I_2, and I_3:

$$P(V|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,I) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=0}^{\nu+4}\sum_{l=0}^{\nu+4}\left(\epsilon_1^2 W_{kl}+\epsilon_2^2 X_{kl}+\epsilon_3^2 Y_{kl}\right)v_k v_l\right\}, \tag{91}$$

where $I_1, I_2, I_3, I_{\rm old} \rightarrow I$.

Before returning to the main problem, this prior must be normalized. To do this the matrix Z_{kl} is defined:

$$Z_{kl} \equiv \epsilon_1^2 W_{kl}+\epsilon_2^2 X_{kl}+\epsilon_3^2 Y_{kl} \tag{92}$$

and the fully normalized prior probability for the pixels is given by

$$P(V|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,I) = (2\pi\sigma^2)^{-\frac{\nu+5}{2}}\sqrt{\lambda_0\cdots\lambda_{\nu+4}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=0}^{\nu+4}\sum_{l=0}^{\nu+4}Z_{kl}v_k v_l\right\} \tag{93}$$

where $\{\lambda_0,\cdots,\lambda_{\nu+4}\}$ is the product of the eigenvalues of the Z_{jk} matrix.

The prior, Eq. (93), may now be inserted into the posterior probability for the jth pixel, Eq. (76), to obtain:

$$\begin{aligned} P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,D,I) &\propto \int\underbrace{\cdots dv_i\cdots}_{i\neq j}\sigma^{-(\nu+5)}\sqrt{\lambda_0\cdots\lambda_{\nu+4}} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=0}^{\nu+4}\sum_{l=0}^{\nu+4}Z_{kl}v_k v_l\right\} \\ &\times \sigma^{-N}\exp\left\{-\frac{1}{2\sigma^2}\sum_{\substack{i=1\\ \text{by }\beta}}^{\nu}(d_i-v_i)^2\right\} \end{aligned} \tag{94}$$

where a factor of $(2\pi)^{-\frac{\nu+5}{2}}$ was dropped.

2.3 Eliminating Nuisance Parameters

To obtain the posterior probability for v_j, all of the parameters except v_j must be removed by integration. There are $\nu+4$ integrals that must be evaluated. These integrals are very similar to those evaluated in the previous section and few of the details will be given. To evaluate these integrals, the exponent in the likelihood, Eq. (94), is squared to obtain:

$$\begin{aligned} P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,D,I) &\propto \int\underbrace{\cdots dv_i\cdots}_{i\neq j}[\lambda_0\cdots\lambda_{\nu+4}]^{\frac{1}{2}}\sigma^{-(N+\nu+5)} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\left[N\overline{d^2}-2\sum_{\substack{i=1\\ \text{by }\beta}}^{\nu}d_i v_i+\sum_{k=0}^{\nu+4}\sum_{l=0}^{\nu+4}g_{kl}v_k v_l\right]\right\} \end{aligned} \tag{95}$$

where $\overline{d^2}$ is the mean-square of the data, Eq. (33), and the interaction matrix, g_{kl}, is given by

$$g_{kl} \equiv Z_{kl}+S_{kl} \qquad 0\le k,l\le \nu+4 \tag{96}$$

where S_{kl} was defined earlier, Eq. (35).

Note that the g_{kl} matrix has been redefined. In fact that is not quite true, it has been generalized. As we proceed though this calculation, each section will generalize the results from the previous sections. Whenever possible, these generalized equations will use the same notation to represent the generalized quantity.

The integrals are over all of the V, except v_j. Pixel v_j behaves like a constant and must be handled in a special manner. Separating v_j from the integration variables one has

$$\begin{aligned} P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,D,I) &\propto \int \underbrace{\cdots dv_i \cdots}_{i\neq j} [\lambda_0 \cdots \lambda_{\nu+4}]^{\frac{1}{2}} \sigma^{-(N+\nu+5)} \\ &\times \exp\left\{-\frac{N\overline{d^2} - 2d_j v_j z + g_{jj} v_j^2}{2\sigma^2}\right\} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{\substack{k=0 \\ k\neq j}}^{\nu+4} \sum_{\substack{l=0 \\ l\neq j}}^{\nu+4} g_{kl} v_k v_l - 2\sum_{\substack{i=0 \\ i\neq j}}^{\nu} (d_i - g_{ij} v_j) v_i\right]\right\} \end{aligned} \tag{97}$$

where z was defined earlier, Eq. (37).

Evaluating the $\nu + 4$ integrals gives

$$P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,D,I) \propto \exp\left\{-\frac{N\overline{d^2} - 2d_j v_j z + g_{jj} v_j^2 - h(v_j)\cdot h(v_j)}{2\sigma^2}\right\} \tag{98}$$

as the posterior probability for the jth pixel with

$$h_l(v_j) \equiv \frac{1}{\sqrt{\lambda_l'}} \sum_{\substack{i=1 \\ \text{by } \beta \\ i\neq j}}^{\nu} [d_i - g_{ji} v_j] e_{li} \qquad (l \neq j) \tag{99}$$

and

$$h(v_j)\cdot h(v_j) \equiv \sum_{i=0}^{\nu+4} h_i(v_j)^2 \qquad (i \neq j) \tag{100}$$

where e_{li} is the ith component of the lth eigenvector of the jth cofactor of Eq. (96) and λ_i' is the ith eigenvalue of this matrix.

If the variance of the noise and the regularizes are actually known then the problem is completed and Eq. (98) represents the best estimate of the jth pixel one can make given the three types of prior information. However, in general σ and $\epsilon_1, \ldots, \epsilon_3$, are not known and must be determined from the data.

2.4 Eliminating σ As A Nuisance Parameter

The posterior probability for v_j independent of σ is computed in a way analogous to what was done in subsection 1.3. The details of the calculation will not be repeated here. However, as a reminder, one must assign a prior probability to the standard deviation (here this is a Jeffreys prior), and integrate with respect to σ over its valid range of values. Note that we cautioned against using improper priors in this calculation and this is essential for location parameters. However, for the scale parameters (the fractional variances, and σ) use of improper priors is harmless. This distribution is given by

$$P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,D,I) \propto \left[1 - \frac{h(v_j)\cdot h(v_j) + 2d_j v_j z - g_{jj} v_j^2}{N\overline{d^2}}\right]^{-\frac{1+N}{2}}. \tag{101}$$

This is a Student t-distribution and this result will be applied in a numerical example, but before it can be used, ϵ_0, ϵ_1, ϵ_2, and ϵ_3 must either be known or be estimated.

2.5 Estimating The Regularizes

The joint posterior probability for ϵ_1, ϵ_2, and ϵ_3 will be computed and used to set the regularizes. In this calculation ϵ_0 will be assumed known. This parameter relates to the prior uncertainty about the amplitudes A, B, and C. It will be assumed that the experiment was designed in such a way that is adequate to actually capture the data in question. This implies that one knows the strength of the signal, at least to order of magnitude and this information was used in setting ϵ_0. However, the other three parameters, ϵ_1, ϵ_2, and ϵ_3 relate to how important the prior information is compared to the data and this is probably not known before actually taking the data. Inferences will have to be made about these parameters.

To make inferences about these parameters, one uses the rules of probability theory to eliminate the nuisance parameters from the problem. Here the standard deviation, σ, will be removed from the posterior probability, and then the rules of probability theory will be used to make inferences about the three regularizer. This calculation is again essentially identical to what was done in subsection 1.4 and the details of the calculation will not be given. To proceed a prior for σ, ϵ_1, ϵ_2 and ϵ_3 must be assigned. Here a Jeffreys priors will be used for the prior probability for all of the regularization parameters; one obtains

$$P(\epsilon_1,\epsilon_2,\epsilon_3|\epsilon_0,D,I) \propto (\epsilon_1\epsilon_2\epsilon_3)^{-1}\left(\frac{\lambda_0\cdots\lambda_{\nu+4}}{\lambda'_0\cdots\lambda'_{\nu+4}}\right)^{\frac{1}{2}}\left[1-\frac{h(\epsilon_1,\epsilon_2,\epsilon_3)\cdot h(\epsilon_1,\epsilon_2,\epsilon_3)}{N\overline{d^2}}\right]^{-\frac{N}{2}} \tag{102}$$

as the joint posterior probability for ϵ_1, ϵ_2, and ϵ_3, where

$$h_l(\epsilon) \equiv \frac{1}{\sqrt{\lambda'_l}}\sum_{\substack{i=1\\ \text{by }\beta}}^{\nu} d_i e_{li}, \tag{103}$$

and

$$h(\epsilon_1,\epsilon_2,\epsilon_3)\cdot h(\epsilon_1,\epsilon_2,\epsilon_3) \equiv \sum_{i=0}^{\nu+4} h_i(\epsilon_1,\epsilon_2,\epsilon_3)^2 \tag{104}$$

where $\{\lambda'_0,\ldots,\lambda'_{\nu+4}\}$ and e_{li} are the eigenvalues and eigenvectors of Eq. (96).

To illustrate the use of the joint posterior probability, the example begun in the previous subsection will be continued. For simplicity only prior information about the functional form of the signal will be used in this example. The data in this example are the same data used in Fig. 1. These data has been repeated in Fig. 5. The solid line in Fig. 5 is the estimate of all of the pixel values when the maximum of the posterior probability is used as the estimate. The dashed lines are the estimated uncertainty in the pixel values in the (mean $\pm$ standard deviation) sense. These estimates assumed the value of the regularizer was known.

To set the regularizer, the posterior probability for ϵ_1 was computed. This is given by Fig. 6(A). Note that this posterior probability density function has a well defined maximum near 3. If one computes the mean value of ϵ_1, one finds $\langle\epsilon_1\rangle = 7.21$. It is this mean value for ϵ_1 that was used to compute the estimates shown in Fig. 5 as the solid line. Note that the estimated signal is flat and only very small deviations are observed from a constant value. Also note that the estimates overlap the true value of the constant, 5.

Next the posterior probability for u_{59} was computed given that $\epsilon_1 = 1$, see Fig. 6(B). This value is relatively far from the value indicated by probability theory. Note that the probability for the pixel is broad and smeared out, indicating that u_{59} has not been well estimated. But also note that true value of the pixel is covered by this posterior! Panel (C) of Fig. 6 is the posterior probability for u_{59} given that $\epsilon_1 = 7.21$. Here, the posterior is much sharper, and the pixel is better resolved.

Figure 5: Interpolation – Functional Form Prior Information

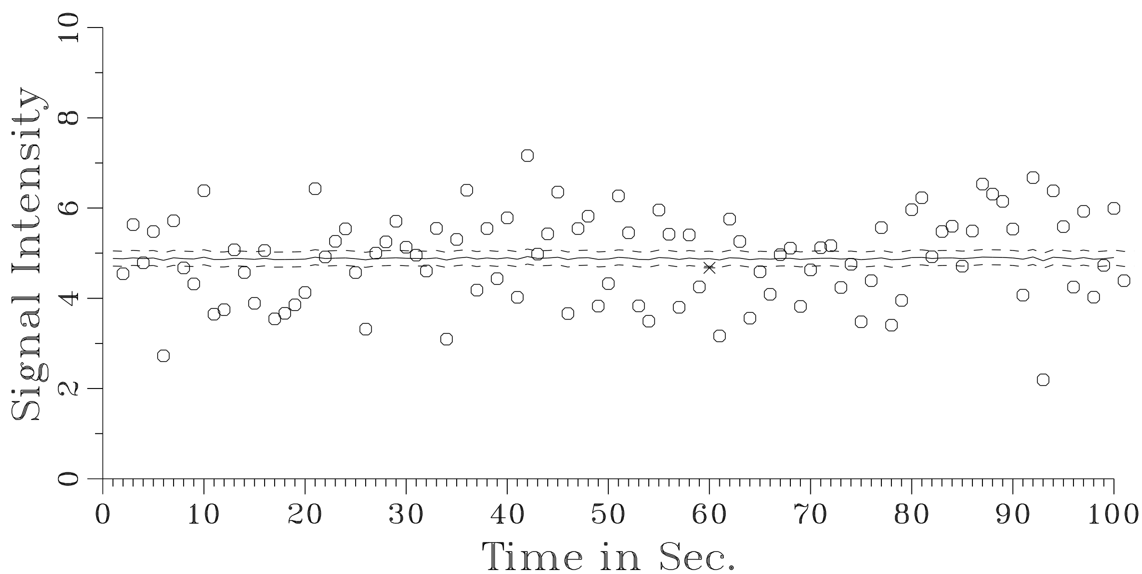

Fig. 5. The functional form of the signal was used in the prior probability. The maximum of the posterior probability for the pixels with $\epsilon = 7.21$ is given by the solid line. The one-standard-deviation width of the posterior is shown as the dotted lines. The data (open circles) are shown for reference.

It is possible to estimate the variance of the noise when $\epsilon_1 = 1$ and $\epsilon_1 = 7.21$. When $\epsilon_1 = 1$ the variance of the noise is estimated to be small: $\langle\sigma\rangle = 0.77$. When $\epsilon_1 = 7.21$ it is estimated to be $\langle\sigma\rangle = 0.99$. When the posterior probability for the pixels is computed, one finds that the estimate with the largest estimated noise level has a better determination of the pixels.

Last, note that the one-standard-deviation error bars shown in Fig. 5 are much narrower than those shown in Fig. 1, indicating that the constraint on the functional form was much more informative than the constraint on the second derivative. But in both cases the estimates easily overlap the true answers at one standard deviation.

3 Deconvolution

Now that the data interpolation problem has been thoroughly addressed, we are in a position to proceed to the full deconvolution problem. Fortunately, the preceding sections have essentially solved the deconvolution problem. As a reminder, in the deconvolution problem there is a data set D that is composed of a signal plus additive noise:

$$d_i = \int_{t_1}^{t_N} d\tau r(t_i - \tau)u(\tau) + n_i \qquad i = \{1, \beta+1, 2\beta+1, \ldots, \nu\} \tag{105}$$

where $r(t_i - \tau)$ is the impulse response function. On a discrete grid, τ takes on values only at the discrete times τ_i and this equation is written

$$d_i = \sum_{j=0}^{\nu+1} r_{ij}u_j + n_i \qquad i = \{1, \beta+1, 2\beta+1, \ldots, \nu\} \tag{106}$$

Figure 6: The Posterior Probability for ϵ_1, and u_{59}

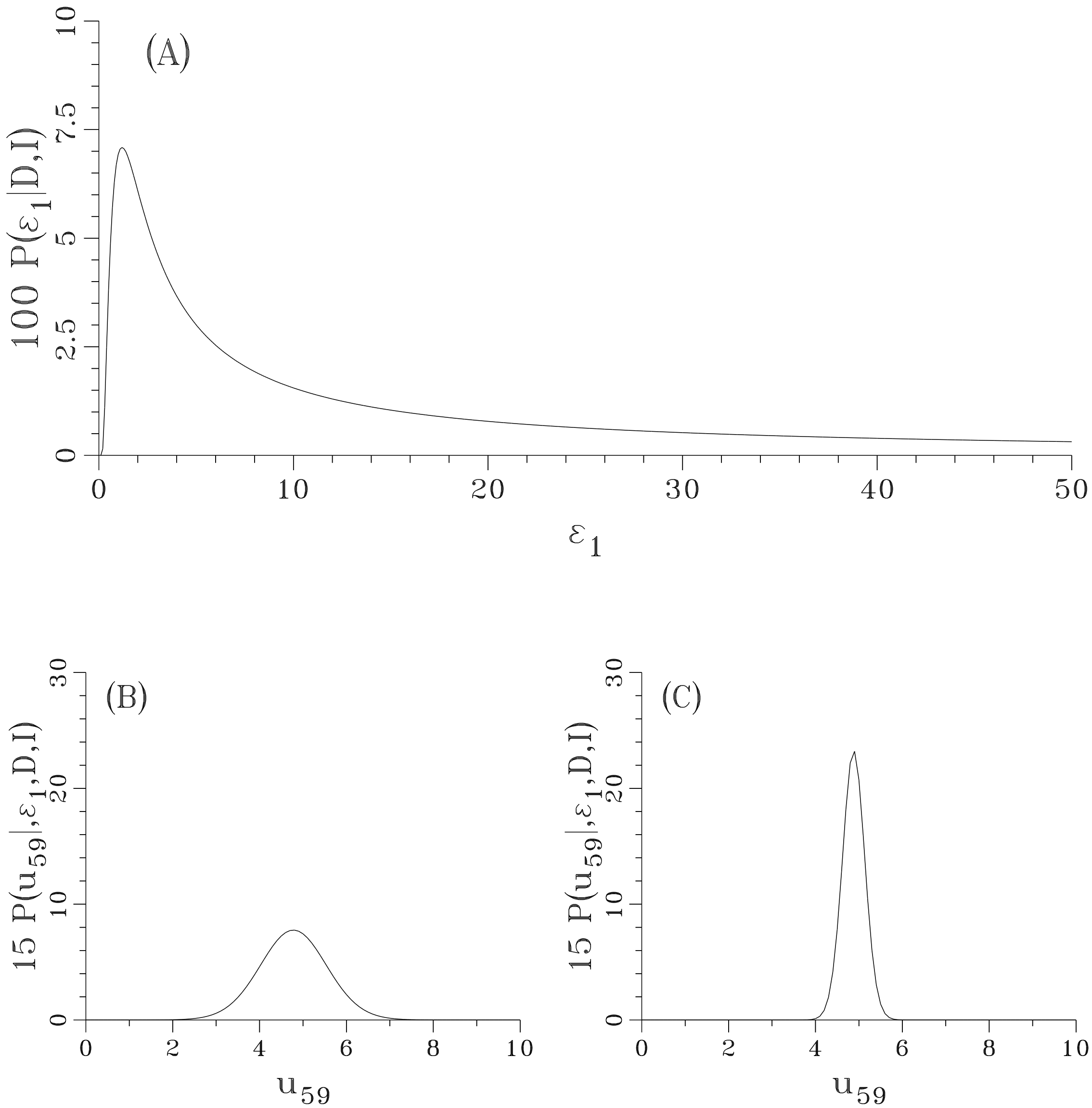

Fig. 6. The posterior probability for ϵ_1 is shown in panel (A). Note that it rises very sharply and then falls off very slowly. The sharp rise indicates that the likelihood and the prior jointly determine a minimum value for ϵ_1 well. But the likelihood is uninformative about large values of ϵ_1. The slow drop off is just the $1/\epsilon_1$ behavior in the prior. Panel (B) is the posterior probability for u_{59} given that $\epsilon_1 = 1$. Note that the probability is broad and smeared out, indicating that u_{59} has not been well estimated. Panel (C) is the posterior probability for u_{59} given that $\epsilon_1 = 7.21$. Here the posterior is much sharper, and the pixel is better resolved.

where

$$r_{ij} \equiv r(t_i - \tau_j)\Delta\tau \tag{107}$$

and $\Delta\tau$ is the time interval between pixel values.

The calculation for the posterior probability for the pixel values proceeds just as in the previous sections. The posterior probability for pixel v_j is given by

$$P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,D,I) \propto \int \underbrace{\cdots dv_i \cdots}_{i\neq j} P(V|I)\sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{\substack{i=1\\ \text{by } \beta}}^{\nu}\left[d_i - \sum_{k=0}^{\nu+1} r_{ik}v_k\right]^2\right\}, \tag{108}$$

where the prior $P(V|I)$ will be taken as Eq. (93). Introduction of the convolution operation only complicates the direct probability or likelihood, not the prior.

Squaring the likelihood and substituting Eq. (93) for the prior, one obtains

$$\begin{aligned} P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,D,I) \quad &\propto \quad \int \underbrace{\cdots dv_i \cdots}_{i\neq j} \sigma^{-(N+\nu+5)}\sqrt{\lambda_0\ldots\lambda_{\nu+4}} \\ &\times \quad \exp\left\{-\frac{1}{2\sigma^2}\left[N\overline{d^2} - 2\sum_{k=0}^{\nu+4} v_k D_k + \sum_{k=0}^{\nu+4}\sum_{l=0}^{\nu+4} g_{kl}v_k v_l\right]\right\} \end{aligned} \tag{109}$$

where D_k is a kind of weighted averaged over the data, and is defined as

$$D_k \equiv \begin{cases} \displaystyle\sum_{\substack{i=1\\ \text{by } \beta}}^{\nu} r_{ik}d_i & \text{if } 0 \leq k \leq \nu+1 \\ 0 & \text{otherwise,} \end{cases} \tag{110}$$

the g_{kl} matrix generalizes to

$$g_{kl} = Z_{kl} + S_{kl}, \tag{111}$$

and S_{kl} defined as

$$S_{kl} \equiv \begin{cases} \displaystyle\sum_{\substack{i=1\\ \text{by } \beta}}^{\nu} r_{ik}r_{il} & \text{if } 0 < k,l \leq \nu+1 \\ 0 & \text{otherwise.} \end{cases} \tag{112}$$

3.1 Eliminating Nuisance Parameters

As observed in subsection 1.3, the pixel being estimated, v_j, behaves as if it were a constant in the integrals and must be treated specially. This is done by separating v_j from the integration variables to obtain:

$$\begin{aligned} P(v_j|\epsilon_0,\epsilon_1,\epsilon_2,\epsilon_3,\sigma,D,I) \quad &\propto \quad \int \underbrace{\cdots dv_i \cdots}_{i\neq j} \sigma^{-(N+\nu+5)} \exp\left\{-\frac{N\overline{d^2} - 2D_j v_j + g_{jj}v_j^2}{2\sigma^2}\right\} \\ &\times \quad \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{\substack{k=0\\ k\neq j}}^{\nu+4}\sum_{\substack{l=0\\ l\neq j}}^{\nu+4} g_{kl}v_k v_l - 2\sum_{\substack{i=0\\ i\neq j}}^{\nu+4}(D_i - g_{ji})v_i\right]\right\}. \end{aligned} \tag{113}$$

Evaluating the $\nu + 4$ integrals gives

$$P(v_j|\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \sigma, D, I) \propto \exp\left\{-\frac{N\overline{d^2} - 2d_j v_j + g_{jj}v_j^2 - h(v_j)\cdot h(v_j)}{2\sigma^2}\right\} \tag{114}$$

as the posterior probability for the jth pixel, where

$$h_l(v_j) \equiv \frac{1}{\sqrt{\lambda_l'}} \sum_{\substack{i=0 \\ i\neq j}}^{\nu+4} [D_i - g_{ji}v_j]e_{li} \qquad (l \neq j), \tag{115}$$

$$h(v_j)\cdot h(v_j) \equiv \sum_{\substack{i=0 \\ i\neq j}}^{\nu+4} h_i(v_j)^2, \tag{116}$$

e_{li} is the ith component of the lth eigenvector of the jth cofactor of Eq. (111), and λ_i' is its ith eigenvalue.

3.2 Eliminating σ As A Nuisance Parameter

Computing the posterior probability for u_j independent of σ is essentially identical to what was done in subsection 1.4 and the details of this calculation will not be given here. The posterior probability for jth pixel value is given by

$$P(v_j|\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, D, I) \propto \left[1 - \frac{h(v_j)\cdot h(v_j) + 2D_j v_j - g_{jj}v_j^2}{N\overline{d^2}}\right]^{-\frac{1+N}{2}}. \tag{117}$$

This is a Student t-distribution and it is this result that is applied in the numerical examples. But before any numerical calculation may be done ϵ_0 ϵ_1, ϵ_2, and ϵ_3 must either be known or estimated.

3.3 Estimating The Regularizes

As was done previously, ϵ_0 will be assumed known and inferences about ϵ_1, ϵ_2, and ϵ_3 will be made. To proceed, a prior for σ, ϵ_1, ϵ_2 and ϵ_3 must be assigned. Here Jeffreys priors will be used for all of the parameters. The prior probability for the pixels was already assigned, Eq. (93), and the probability for the data is given by Eq. (9). Using these, one obtains the joint posterior probability for ϵ_1, ϵ_2 and ϵ_3:

$$\begin{aligned} P(\epsilon_1, \epsilon_2, \epsilon_3|\epsilon_0, D, I) \;&\propto\; \int \frac{dv_0 \ldots dv_{\nu+4}d\sigma}{\sigma\epsilon_1\epsilon_2\epsilon_3}[\lambda_0\cdots\lambda_{\nu+4}]^{\frac{1}{2}}\sigma^{-(N+\nu+5)} \\ &\times\; \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=0}^{\nu+4}\sum_{j=0}^{\nu+4} Z_{ij}v_iv_j\right\} \\ &\times\; \exp\left\{-\frac{1}{2\sigma^2}\sum_{\substack{i=1 \\ \text{by } \beta}}^{\nu}[d_i - \sum_{k=0}^{\nu+1} r_{ik}v_k]^2\right\}, \end{aligned} \tag{118}$$

where the eigenvalues, $\{\lambda_0, \ldots, \lambda_{\nu+5}\}$, are the eigenvalues of Eq. (92), and u_i [from Eq. (9)] was replaced by v_i to conform to the current notation. Evaluating all of the integrals and dropping some irrelevant constants, one obtains:

$$P(\epsilon_1,\epsilon_2,\epsilon_3|\epsilon_0,D,I)\propto(\epsilon_1\epsilon_2\epsilon_3)^{-1}\left(\frac{\lambda_0\cdots\lambda_{\nu+4}}{\lambda'_0\cdots\lambda'_{\nu+4}}\right)^{\frac{1}{2}}\left[1-\frac{h(\epsilon_1,\epsilon_2,\epsilon_3)\cdot h(\epsilon_1,\epsilon_2,\epsilon_3)}{N\overline{d^2}}\right]^{-\frac{N}{2}} \tag{119}$$

as the joint posterior probability for the three fractional variances, where

$$h_l(\epsilon_1,\epsilon_2,\epsilon_3)\equiv\frac{1}{\sqrt{\lambda'_l}}\sum_{i=0}^{\nu+4}D_ie_{li}, \tag{120}$$

$$h(\epsilon_1,\epsilon_2,\epsilon_3)\cdot h(\epsilon_1,\epsilon_2,\epsilon_3)\equiv\sum_{i=0}^{\nu+4}h_i(\epsilon_1,\epsilon_2,\epsilon_3)^2, \tag{121}$$

$\{\lambda'_0,\ldots,\lambda'_{\nu+4}\}$ are the eigenvalues of the g_{kl} matrix, Eq. (111), and e_{li} is the ith component of the lth eigenvector of this matrix.

3.4 Examples – Deconvolution

To illustrate this calculation, several deconvolution examples will be given which incorporate different types of prior information. In the first example, very little prior information will be available; all that will be used is a constraint on the smoothness of the function. In the second example, more prior information will be available, and the functional form of the signal will be used to constrain the deconvolution. In the third example, both sets of prior information will be used to constrain the deconvolution. The data will be simulated sinusoidal data that have been low-pass filtered. This problem is important in radar target identification, because it is the free space signal that must be known in the target identification problem.

The signal function will be taken to be a pure sinusoid of known frequency and phase:

$$u_i=10\cos(0.3t_j). \tag{122}$$

However, this signal has been filtered using a low-pass filter:

$$r(t_i)=\frac{1}{c}\mathrm{e}^{-0.25t_i} \tag{123}$$

where the constant c is given by

$$c=\sum_{i=1}^{N}\mathrm{e}^{-0.25t_i} \tag{124}$$

and the times t_i were taken to be $0,1,\ldots,N-1$. Note that the smearing function is defined to be zero for times less than t_1 or greater than t_N. The data are a convolution between the signal function, $u(t)$, and the impulse response function $r(t)$:

$$d_i=\sum_{j=1}^{N}10\cos(0.3t_j)\mathrm{e}^{-0.25[t_i-t_j]}\theta(t_i-t_j)/c+n_i, \tag{125}$$

where n_i represents noise of unit standard deviation, and $\theta(t_i-t_j)$ is the unit step function.

The filter changes the amplitude of the response. Consequently, the time domain signal-to-noise ratio is not 10; rather it is more like 5. A plot of the impulse response function for t_{100}, is shown in Fig. 7(A). Data item 100 is a weighted average of all of the preceding signal values. As you go

Figure 7: Deconvolution – The Data

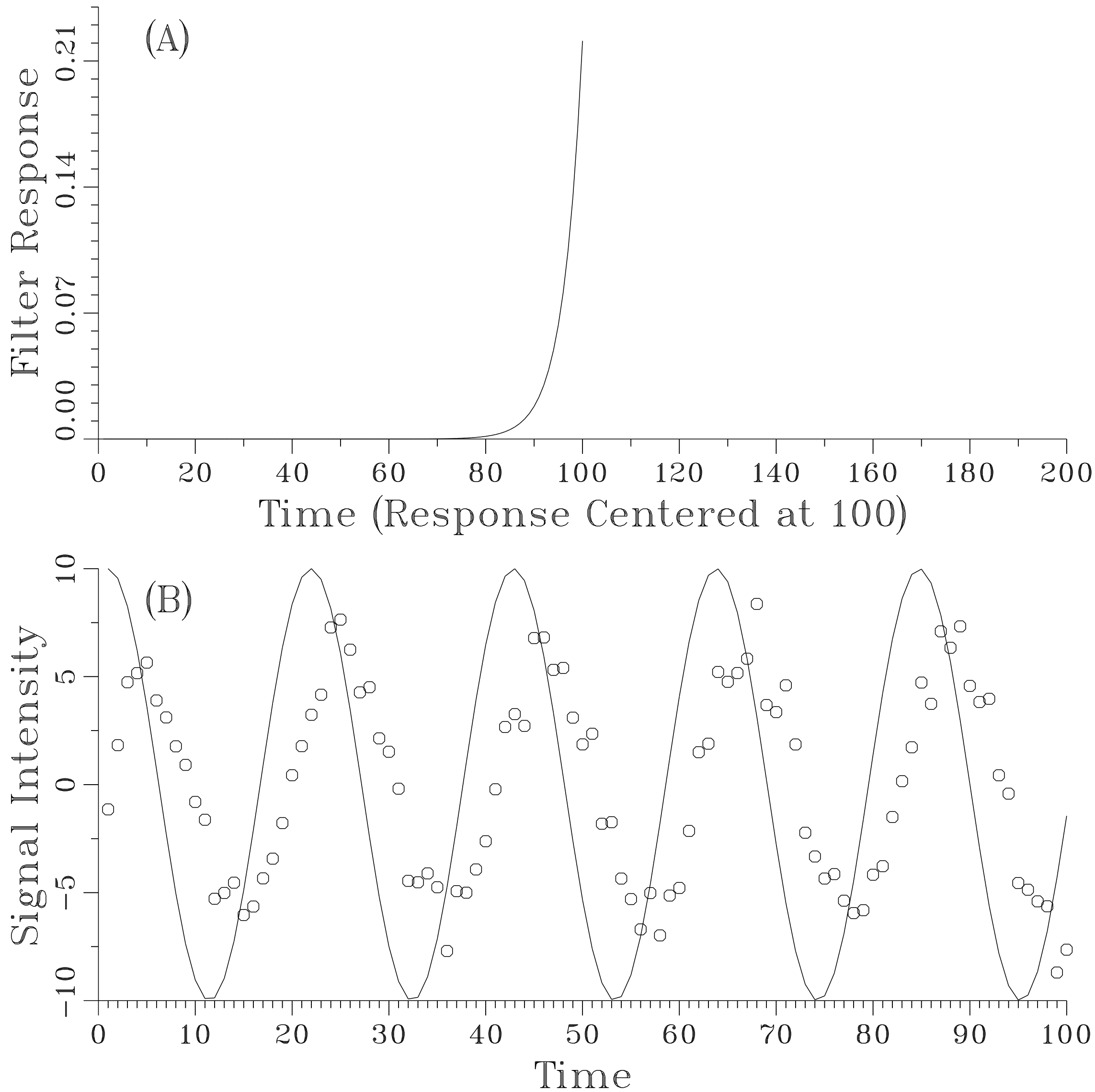

Fig. 7. The filter response function, panel (A), mixes the components of the sinusoid. The mixing takes place as a weighted average. Here all values from $t = 100$ backward in time are averaged to produce the output from the filter. The data, open circles panel (B), remain a sinusoid but shifted in time and with decreased amplitude. The problem is to recover the original signal, solid line in panel (B).

back in time, the signal values become less and less important in this average, finally dropping to essentially zero after 20 sampling intervals. The data values (convolved signal + noise) are shown in Fig. 7(B) as the open circles. The true signal is shown as the solid line in panel 7(B). The convolution introduces an effective amplitude change and phase shift, while the "noise" introduces uncertainty about the "true" convolved signal.

To remove the effect of the convolution, a constraint on the second derivative will be used. To apply the posterior probability for the pixels, one must first set the value of the regularizer. This is done by computing the posterior probability for the regularizer given the data and the prior information, Eq. (119). This is shown in Fig. 8(A). As in the previous examples, this probability density function has a well defined maximum near $\epsilon_3 \approx 0.8$. This maximum value was used in computing the posterior probability for the pixel values, Eq. (117). The maximum of the posterior probability for the pixels (solid line) plus or minus one standard deviation (dashed line) is shown in 8(B). The true signal values are shown as the plus signs. Notice the true signal is covered almost everywhere at one standard deviation. Also note that there is a systematic misfit in the peaks and valleys. That is because the prior information tries to make the second derivative as small as possible. At these turning points the second derivative is at its maximum, so of course the reconstruction will undershoot the mark here. Last, note that the reconstruction is bad near time $t = 100$. But probability theory knows this and has widened the error bars, so that the true value is still overlapped at two standard deviations.

In the second part of this example, use of the correct functional form of the signal will be investigated. Here it will be assumed that the signal must be a cosine with the known correct frequency. The posterior probability for the regularizer, Eq. (119), is shown in Fig. 9(A). Again there is a peak near $\epsilon_1 \approx 0.25$. This value of ϵ_1 was then used to compute the posterior probability for each of the pixels, Eq. (117). The maximum of the posterior probability for each pixel is shown in Fig 9(B) as the solid line. The one standard deviation error bars are shown as the dashed lines. The true signal is shown as the plus signs. Note that the reconstruction follows the signal much more closely: The true signal is easily covered by the one-standard-deviation error bars. However, unlike the previous example this reconstruction does not know about the "smoothness" of the function so the reconstruction is jagged, even though it actually fits the data better. This suggest that these two pieces of prior information could be combined, and this reconstruction would be better than either separately.

Repeating this example using both the second derivative constraint and the functional form of the signal is more difficult because now there are two regularizes: ϵ_1 the regularizer associated with the functional form, and ϵ_3 associated with the second derivative constraint. As in the other examples, to compute the posterior probability for a pixel, we must set these regularizes. This is done by computing the joint posterior probability for the regularizes, Eq. (119), and then marginalizing over either ϵ_1 or ϵ_3. In Fig. 10 the joint posterior probability for these two regularizes has been plotted. The dashed contours are the base 10 logarithm of $P(\epsilon_1, \epsilon_3|\epsilon_0, D, I)$. The region enclosed by the contour labeled 9 contains 99% of the total probability. The region enclosed by the contour labeled 8 contains 99.9% of the total probability, etc. The solid lines inside of contour 9 is the fully normalized joint posterior probability.

From this joint posterior probability for ϵ_1 and ϵ_3, it is possible to compute the posterior probability for either ϵ_1 or ϵ_3 by using the sum rule from probability theory. The posterior probability for ϵ_1 is given by

$$P(\epsilon_1|\epsilon_0, D, I) = \int d\epsilon_3 P(\epsilon_1, \epsilon_3|\epsilon_0, D, I) \tag{126}$$

Figure 8: Deconvolution – Second Derivative Constraint

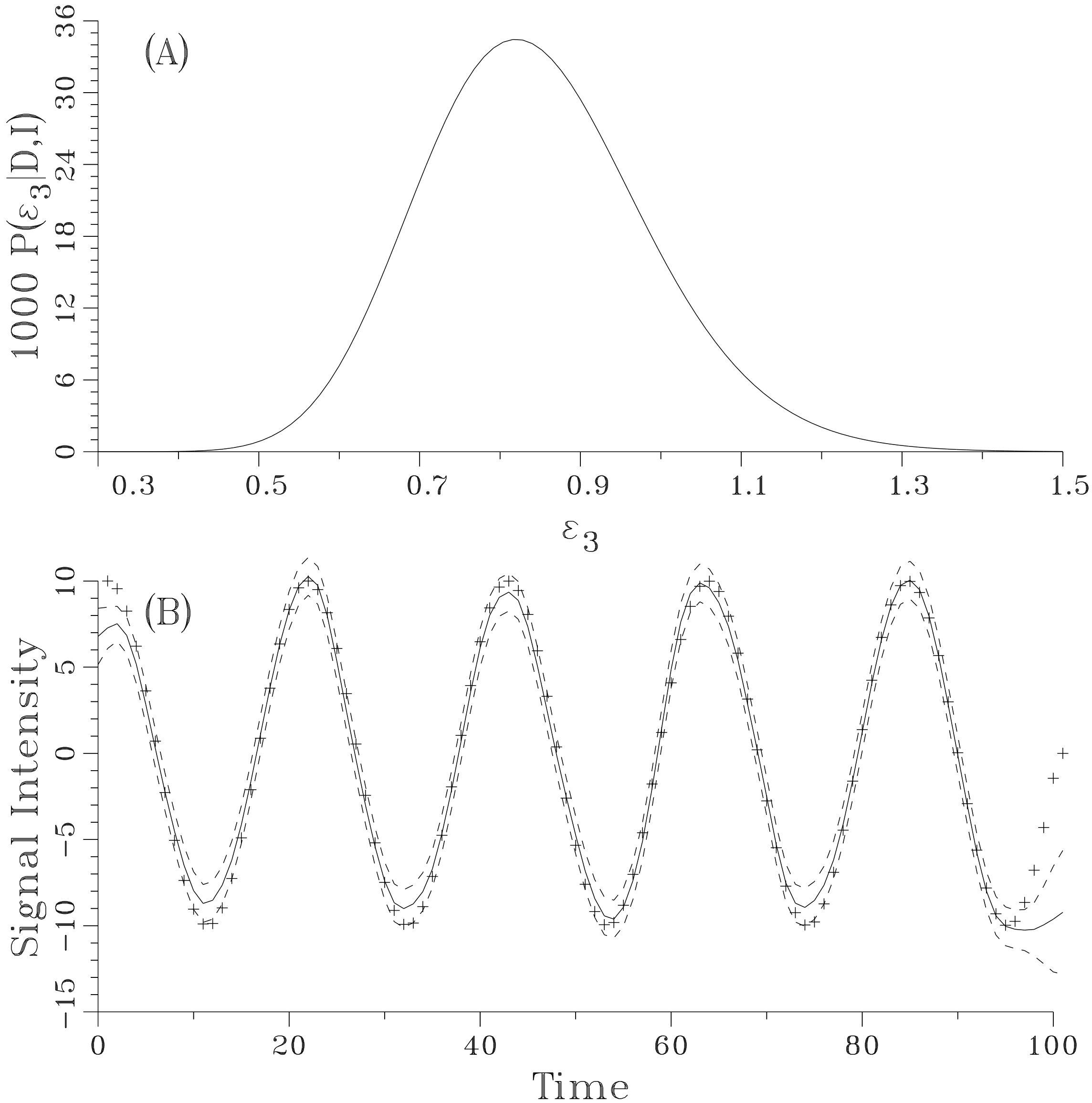

Fig. 8. In panel (A) the posterior probability for the regularizer ϵ_3 is shown. As in previous examples any value of ϵ_3 close to the maximum yields essentially identical deconvolutions. Panel (B) shows the peak value of each estimated pixel value (solid line) plus or minus one standard deviation (dashed line). The true values are shown as the plus signs.

Figure 9: Deconvolution – Functional Form Constraint

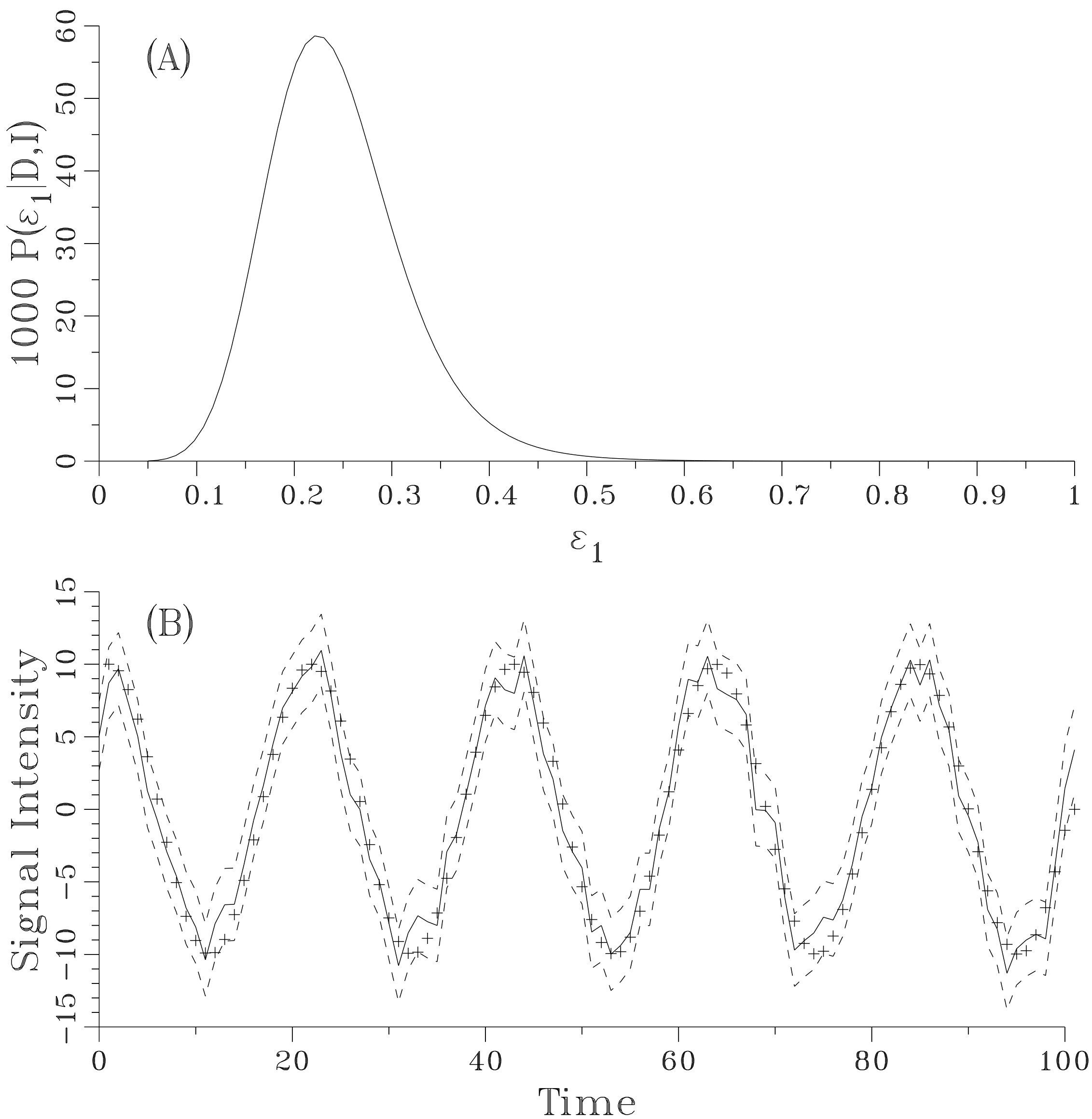

Fig. 9. In panel (A) the posterior probability for the regularizer ϵ_1 is shown. As in previous examples, any value of ϵ_1 close to the maximum yields essentially identical deconvolutions. Panel (B) shows the peak value of each estimated pixel value (solid line) plus or minus one standard deviation (dashed line). The true values are shown as the plus signs.

Figure 10: The Joint Probability for ϵ_1 and ϵ_3

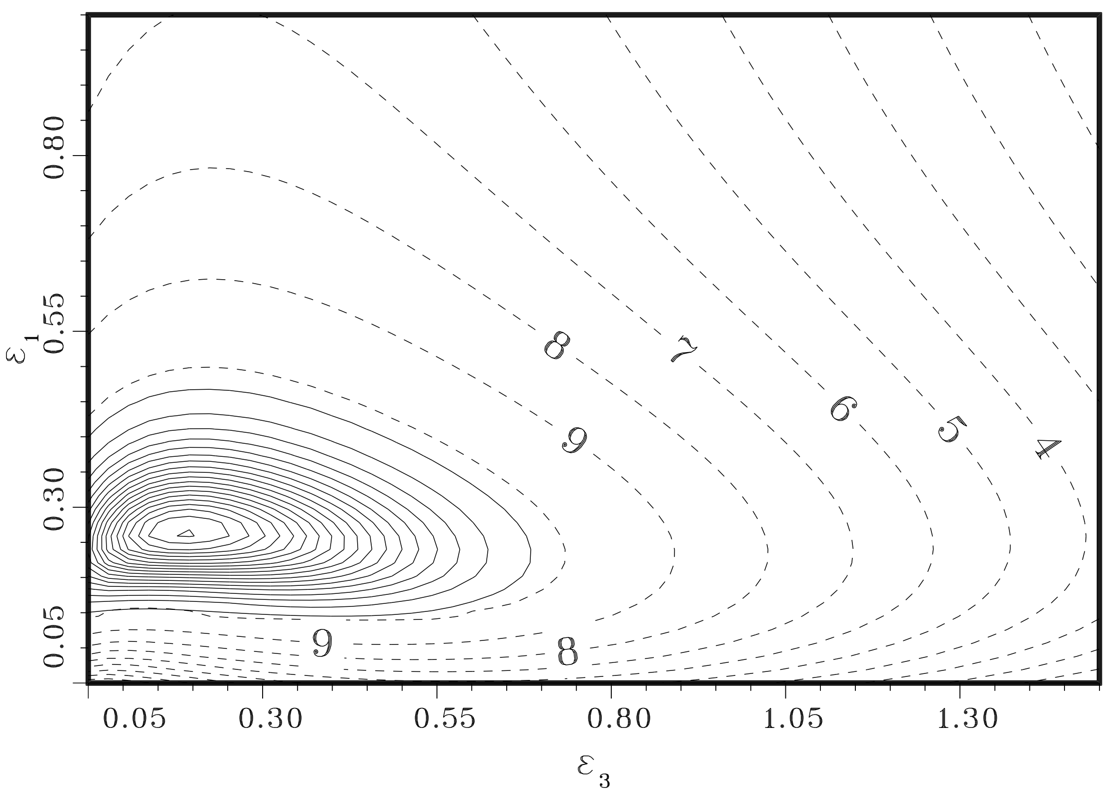

Fig. 10. When both constraints are used (second derivative, and functional form) the joint posterior probability for the regularizer is a function of both ϵ_1 and ϵ_3. The dashed lines are the base 10 logarithm of $P(\epsilon_1, \epsilon_3|\epsilon_0, D, I)$. A change of one from the maximum corresponds to including 90% of the total probability. So effectively everything outside of the contour labeled 9 is irrelevant. The closely spaced solid contours are the fully normalized posterior probability. The region covered by these contours covers 99% of the total probability.

Figure 11: Deconvolution – The Estimated Parameters

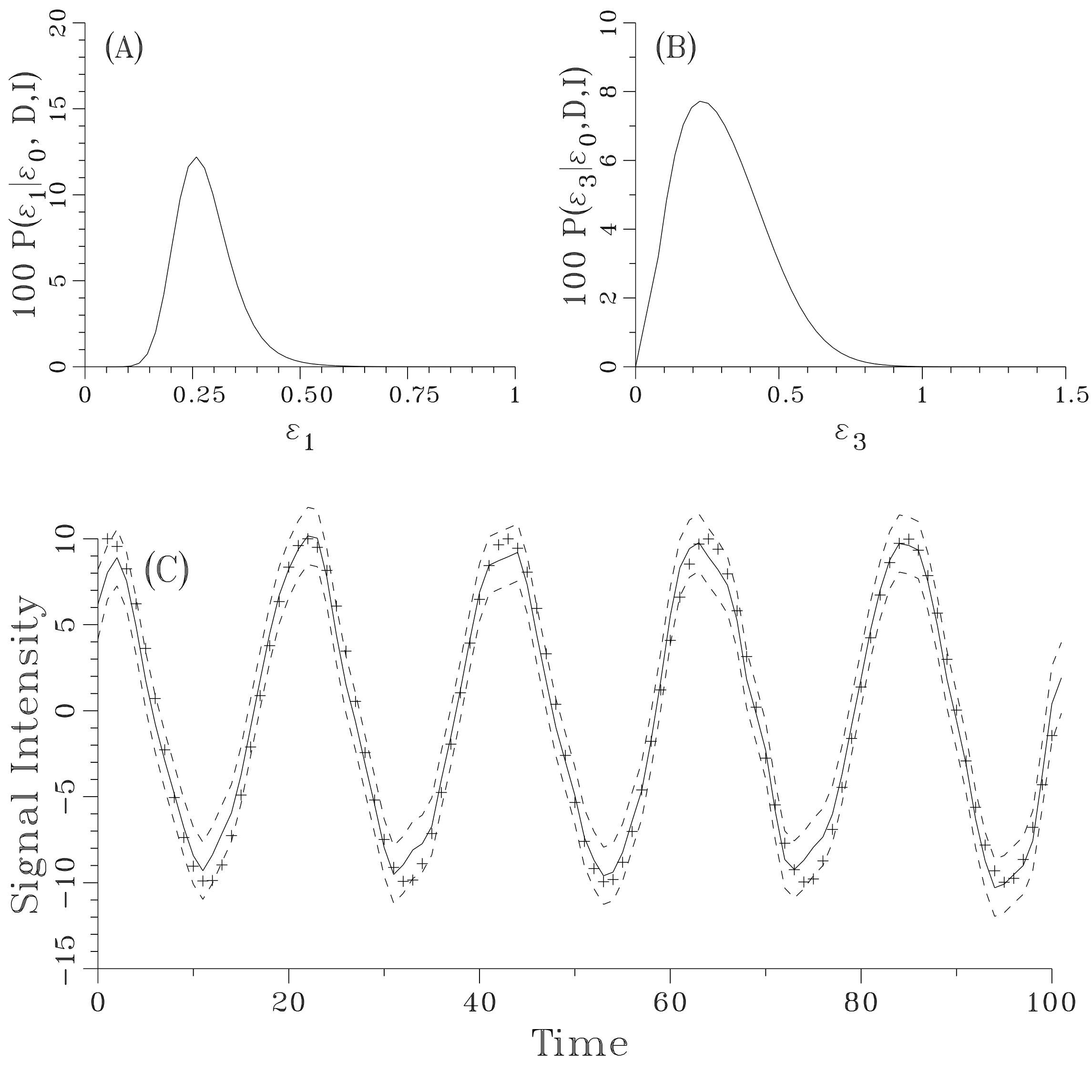

Fig. 11. From the joint posterior probability for ϵ_1 and ϵ_3 (Fig. 10), one can easily compute the posterior probability for each ϵ_1, panel (A), and ϵ_3, panel (B). Using the maxima from these marginal distributions, a $\pm$ standard deviation estimate for the pixel values were computed. The maxima are shown in (C) as the solid line, the one-standard-deviation error bars are shown as the dashed lines, and the true signal values are given by the plus signs.

and the posterior probability for ϵ_3 is given by

$$P(\epsilon_3|\epsilon_0, D, I) = \int d\epsilon_1 P(\epsilon_1, \epsilon_3|\epsilon_0, D, I). \tag{127}$$

These two probability density functions have been plotted in Fig 11(A) and (B) respectively. The peak value for ϵ_1 is approximately 0.35 and for ϵ_3 it is approximately 0.25. These values were used to compute the posterior probability for the pixels. The peak values are shown in (C) as the solid line, the one-standard-deviation error bars are shown as the dashed lines, and the true signal values are given by the plus signs. Note that this reconstruction has combined the best features of the two previous examples: The use of information about the functional form causes the reconstruction to follow the true signal much more closely, while use of the smoothing constraint has suppressed much of the random fluctuation.

4 Deconvolution – Generalizations

The results of the preceding calculations can be generalized in a number of ways by allowing more general types of prior information. When the priors were established for the deconvolution problem, only one function per type of prior information was allowed. There is no reason why more functions cannot be allowed, and in many cases the need for them is obvious. For example, suppose f_1 (the functional form) were a cosine, then a second function, a sine, is needed to properly express the phase of the sinusoid. Additionally, only three pieces of prior information were used: one on the functional form of the signal, one on its first derivative, and one on its second derivative. There is no reason why one could not have more then three pieces of prior information, and these could constrain more complicated functions of the pixels than just the first and second derivatives.

In this section, the deconvolution results presented in the previous sections will be generalized to allow for any number of pieces of prior information. This information can specify functional forms containing any number of amplitudes and functions, and these functions will be allowed to constrain an arbitrary linear combination of the pixels. The total number of pieces of prior information will be designated as r. Each piece of prior information will be designated as $I_1, \ldots, I_r$. For information I_μ, the constraint will be written

$$\sum_{i=0}^{\nu+1} \left[\sum_{j=0}^{\nu+1} a_{ij}^\mu u_j - \sum_{k=1}^{m_\mu} A_k^\mu f_k^\mu(t_i) \right]^2 = \delta_\mu^2 \tag{128}$$

where a_{ij}^μ is a known matrix of coefficients that describe how the pixels interact. For example, it could describe the second-derivative constraint used earlier. The coefficients A_k^μ are the amplitudes or intensities of the signal functions, and they will be considered as unknown, nuisance parameters. The constraint functions f_k^μ are the analogue of the functions (f_1, f_2, and f_3) used earlier. However, there are m_μ of these functions for each of the r constraints. There are a total of $\sum m_\mu$ functions and amplitudes. Each constraint will have a fractional variance or regularizer associated with it. These regularizes will be designated as $\epsilon_1, \ldots, \epsilon_r$. Last, note that the sum over discrete times (the i index) runs from $0 \leq i \leq \nu + 1$. So the above constraints are written *implicitly* to include the boundary conditions.

Converting the μth constraint into a probability density function for the pixels is straightforward and results in

$$P(V|\epsilon_0, \epsilon_\mu, I_\mu, I) \propto \exp\left\{ -\frac{\epsilon_0^2}{2\sigma^2} \sum_{l=1}^{m_\mu} (A_l^\mu)^2 - \frac{\epsilon_\mu^2}{2\sigma^2} \sum_{i=0}^{\nu+1} \left[\sum_{j=0}^{\nu+1} a_{ij}^\mu u_j - \sum_{k=1}^{m_\mu} A_k^\mu f_k^\mu(t_i) \right]^2 \right\}, \tag{129}$$

where the first term expresses the prior information about the amplitudes and the second expresses the prior information about the pixels. As in the previous examples, this prior may be converted into a prior with a double sum; this gives

$$P(V|\epsilon_0, I_\mu, I) \propto \exp\left\{-\frac{\epsilon_\mu^2}{2\sigma^2}\sum_{k=0}^{\eta}\sum_{l=0}^{\eta} W_{kl}^\mu v_k v_l\right\}, \tag{130}$$

where

$$\eta = \nu + 1 + \sum_{k=1}^{r} m_k \tag{131}$$

and $(\eta + 1)$ is the total number of unknown generalized pixels v_i. Following what was done earlier, these generalized pixels are defined as:

$$v_i \equiv \begin{cases} u_i & \text{if } 0 \leq i \leq \nu + 1 \\ A^1_{i-\nu-1} & \text{if } \nu + 1 < i \leq \nu + 1 + m_1 \\ A^2_{i-\nu-1-m_1} & \text{if } \nu + 1 + m_1 < i \leq \nu + 1 + m_1 + m_2 \\ \vdots & \\ A^r_{i-\nu-m_1-\cdots-m_r} & \text{if } \nu + 1 + m_1 + \cdots + m_{r-1} < i \leq \eta. \end{cases} \tag{132}$$

Last the matrix W_{kl}^μ is defined as

$$W_{kl}^\mu = b_{kl}^\mu - c_{kl}^\mu + d_{kl}^\mu, \tag{133}$$

where b_{kl}^μ, c_{kl}^μ and d_{kl}^μ correspond to the coefficients of the terms obtained by squaring the exponent, combining all of the terms and carrying out the sum over i. The matrix b_{kl}^μ is just the coefficient of the first term of the square in Eq. (129), and is given by

$$b_{kl}^\mu \equiv \begin{cases} \displaystyle\sum_{i=0}^{\nu+1} a_{ik}^\mu a_{il}^\mu & \text{if } 0 \leq k, l \leq \nu + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{134}$$

Note that in setting up the general W_{kl}^μ matrix, the indices are allowed to take on values $0 \leq k, l \leq \eta$, so in the definition of b_{kl}^μ it was necessary to state explicitly that this term is zero when either k or l was greater than $\nu + 1$. The matrix c_{kl}^μ corresponds to the coefficient of the cross term, and is given by

$$c_{kl}^\mu \equiv \begin{cases} \displaystyle\sum_{i=1}^{\nu+1} a_{ik}^\mu f_n^\mu(t_i) & \left\{\begin{array}{l} \text{If } k \leq \nu + 1 \text{ and } \nu + 1 + m_1 + \cdots + m_{\mu-1} < l \\ \text{and } l \leq \nu + 1 + m_1 + \cdots + m_\mu \\ \text{where } n = l - \nu - 1 - m_1 - \cdots - m_{\mu-1} \\ \text{OR} \\ \text{If } l \leq \nu + 1 \text{ and } \nu + 1 + m_1 + \cdots + m_{\mu-1} < k \\ \text{and } k \leq \nu + 1 + m_1 + \cdots + m_\mu \\ \text{where } n = k - \nu - 1 - m_1 - \cdots - m_{\mu-1} \end{array}\right. \\ 0 & \text{otherwise.} \end{cases} \tag{135}$$

The third term is the square plus first the prior probability for the amplitudes, and is defined as

$$d^{\mu}_{kl} \equiv \begin{cases} \dfrac{\epsilon_0^2}{\epsilon_\mu^2}\delta_{kl} + \displaystyle\sum_{i=0}^{\nu+1} f^{\mu}_{n_1}(t_i) f^{\mu}_{n_2}(t_i) & \begin{cases} \text{if } \nu+1+m_1+\cdots+m_{\mu-1} < k,l \\ \text{and } k,l \leq \nu+1+m_1+\cdots+m_\mu \\ \text{where } n_1 = k-\nu-1-m_1-\cdots m_{\mu-1} \\ \text{and } n_2 = l-\nu-1-m_1-\cdots m_{\mu-1}. \end{cases} \\ 0 & \text{otherwise} \end{cases} \tag{136}$$

where δ_{kl} is the Kronecker delta function. As was done previously, the individual priors may be combined to obtain a single prior which expresses all of the prior information. This prior is given by

$$P(V|\epsilon_0,\ldots,\epsilon_r,I) = \sqrt{\lambda_0\cdots\lambda_\eta}\left(2\pi\sigma^2\right)^{\eta+1}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=0}^{\eta}\sum_{l=0}^{\eta} Z_{kl}v_k v_l\right\} \tag{137}$$

where

$$Z_{kl} = \sum_{\mu=1}^{r}\epsilon_\mu^2 W^{\mu}_{kl} \tag{138}$$

and $\{\lambda_0\cdots\lambda_\eta\}$ are the eigenvalues of the Z_{kl} matrix.

The mathematics from the three previous sections may now be repeated to obtain a generalized result. First, the posterior probability for the jth generalized pixel is given by

$$P(v_j|\epsilon_0,\ldots,\epsilon_r,\sigma,D,I) \propto \exp\left\{-\frac{N\overline{d^2} - 2D_j v_j + g_{jj}v_j^2 - h(v_j)\cdot h(v_j)}{2\sigma^2}\right\} \tag{139}$$

where D_j was defined earlier, Eq. (110),

$$h_l(v_j) \equiv \frac{1}{\sqrt{\lambda'_l}}\sum_{\substack{i=0\\ i\neq j}}^{\eta}[D_i - g_{ji}v_j]e_{li} \qquad (l\neq j) \tag{140}$$

$$h(v_j)\cdot h(v_j) \equiv \sum_{\substack{i=0\\ i\neq j}}^{\eta} h_i(v_j) \tag{141}$$

and $\lambda'_0\cdots\lambda'_\eta$ are the eigenvalues of the jth cofactor of the g_{kl} matrix. The g_{kl} matrix is defined as

$$g_{kl} \equiv Z_{kl} + S_{kl} \tag{142}$$

where S_{kl} was defined in Eq. (112).

Next, the posterior probability for the jth pixel value, independent of the variance of the noise, is given by

$$P(v_j|\epsilon_0,\ldots,\epsilon_r,D,I) \propto \left[1 - \frac{h(v_j)\cdot h(v_j) + 2D_j v_j - g_{jj}v_j^2}{N\overline{d^2}}\right]^{-\frac{1+N}{2}}. \tag{143}$$

Last, the joint marginal posterior probability for the regularizes is given by

$$P(\epsilon_1,\ldots,\epsilon_r|\epsilon_0,D,I) \propto (\epsilon_1\cdots\epsilon_r)^{-1}\left(\frac{\lambda_0\cdots\lambda_{\nu+1}}{\lambda'_0\cdots\lambda'_\eta}\right)^{\frac{1}{2}}\left[1 - \frac{h(\epsilon_1,\ldots,\epsilon_r)\cdot h(\epsilon_1,\ldots,\epsilon_r)}{N\overline{d^2}}\right]^{-\frac{N}{2}} \tag{144}$$

where $\lambda'_0 \cdots \lambda'_\eta$ are the eigenvalues of the g_{jk} matrix, Eq. (142),

$$h(\epsilon_1, \ldots, \epsilon_r) \cdot h(\epsilon_1, \ldots, \epsilon_r) \equiv \sum_{i=0}^{\eta} h_i(\epsilon_1, \ldots, \epsilon_r)^2 \tag{145}$$

and

$$h_i(\epsilon_1, \ldots, \epsilon_r) = \frac{1}{\sqrt{\lambda'_i}} \sum_{j=0}^{\eta} D_j e_{ij} \tag{146}$$

where in Eqs. (144–146), the eigenvalues $\{\lambda'_0, \ldots, \lambda'_\eta\}$ are the eigenvalues of the g_{jk} matrix, Eq. (142), and e_{ij} are its eigenvectors.

Note that care must be taken when interpreting the results of these calculations, because the notation for the eigenvalues and eigenvectors has not been changed when different matrices were used. The meaning should remain clear because when each formula is given the matrices being diagonalized are clearly stated. But just to be clear on this point, when the posterior probability for the jth pixel is being computed the eigenvalues $\lambda'_0, \ldots, \lambda_n u'$ and eigenvectors e_{ij} refer to the jth cofactor of the g_{jk} matrix. However, when the posterior probability for the regularizes is computed $\lambda'_0 \ldots \lambda'_\nu$ refer to the eigenvalues of the g_{jk} matrix (not the$j$th cofactor) and e_{jk} refer to its eigenvectors of g_{jk}.

4.1 Estimating The Pixel Values

It is one thing to formally derive a result and quite another for it to be useful. The posterior probability for the individual pixels given all of the prior information, Eq. (143), is one of these types of results. While this result will prove useful in examining individual, important, pixels it is not the way to estimate *all* of the pixels. Even if one were to compute this posterior probability density for all of the pixels, it still would not give one an estimated signal; rather it would tell one what is actually known about the signal values and the uncertainty in those values. What is actually needed is an estimate of the pixels and the uncertainty in the estimate.

There are many ways to estimate a parameter using probability theory and the estimate of choice will depend on what one stands to lose if one is wrong. Two different types of estimates are the maximum of the posterior probability, and mean or expected value of a parameter. In this calculation, the expected value and peak values are the same, so the pixel estimates will be given as the mean plus or minus the standard deviation.

The expected value of the jth generalized pixel is given by

$$\langle v_j | \epsilon_0, \ldots, \epsilon_r, \sigma, I \rangle = \int dv_0 \ldots dv_\eta v_j P(v_0, \ldots, v_\eta | \epsilon_0, \ldots, \epsilon_r, \sigma, D, I) \tag{147}$$

where the notation $\langle v_j | \epsilon_0, \ldots, \epsilon_r, \sigma, I \rangle$ means the expected value of pixel v_j given that $\epsilon_0, \ldots, \epsilon_r$, and σ are known. But note that it is the fully normalized joint probability density function that is to be used. Consequently, when this calculation is performed the probability density function will have to be normalized:

$$\langle v_j | \epsilon_0, \ldots, \epsilon_r, \sigma \rangle = \int \frac{dv_0 \ldots dv_\eta v_j}{\text{Normalization}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{k=0}^{\eta} \sum_{l=0}^{\eta} g_{kl} v_k v_l - 2 \sum_{k=0}^{\eta} D_k v_k \right\} \tag{148}$$

where

$$\text{Normalization} = \int dv_0 \ldots dv_\eta \exp\left\{ -\frac{1}{2\sigma^2} \sum_{k=0}^{\eta} \sum_{l=0}^{\eta} g_{kl} v_k v_l - 2 \sum_{k=0}^{\eta} D_k v_k \right\}. \tag{149}$$

Evaluating the integrals, one obtains

$$\langle v_j|\epsilon_0,\ldots,\epsilon_r,\sigma\rangle = \sum_{i=0}^{\eta} \frac{e_{ij}h_i}{\sqrt{\lambda_i'}}. \tag{150}$$

Similarly the expected mean-square value of the pixels is given by

$$\langle v_j v_k|\epsilon_0,\ldots,\epsilon_r,\sigma\rangle = \int dv_0\ldots dv_\eta v_j v_k P(v_0,\ldots,v_\eta|\epsilon_0,\ldots,\epsilon_r,\sigma,D,I), \tag{151}$$

and one finds

$$\langle v_j v_k|\epsilon_0,\ldots,\epsilon_r,\sigma\rangle = \langle\sigma^2\rangle \sum_{i=0}^{\eta} \frac{e_{ij}e_{ik}}{\sqrt{\lambda_i'}}. \tag{152}$$

From which one would make a (mean $\pm$ standard deviation) estimate of

$$(v_j)_{est} = \sum_{i=0}^{\eta} \frac{e_{ij}h_i}{\sqrt{\lambda_i'}} \pm \sqrt{\langle\sigma^2\rangle \sum_{i=0}^{\eta} \frac{e_{ij}e_{ij}}{\sqrt{\lambda_i'}}}. \tag{153}$$

Note that while the individual probability distributions would require one to invert a matrix for each value of v_j; the (mean $\pm$ standard deviation) estimate may be done for all of the pixels with a single matrix inversion.

4.2 Estimating The Noise Level

Before the above result can be used, $\langle\sigma^2\rangle$ must be computed. To compute $\langle\sigma^2\rangle$, one must evaluate

$$\langle\sigma^2\rangle = \int_0^\infty d\sigma\sigma^2 P(\sigma|\epsilon_0,\cdots,\epsilon_r,D,I)d\sigma \tag{154}$$

where $P(\sigma|\epsilon_0,\cdots,\epsilon_r,D,I)$ is the fully normalized posterior probability for σ given the regularizes and the data. But using the rules of probability theory, this is just the prior probability for σ times the probability for the regularizes given σ. So the expectation value may be written as:

$$\langle\sigma^2\rangle = \int_0^\infty d\sigma\sigma^2 d\sigma P(\sigma|I)P(\epsilon_1,\cdots,\epsilon_r|\epsilon_0,\sigma,D,I). \tag{155}$$

where

$$P(\epsilon_1,\cdots,\epsilon_r|\epsilon_0,\sigma,D,I) \propto \sigma^{m_1+\cdots+m_r-N} \exp\left\{\frac{N\overline{d^2} - h(\epsilon_1,\ldots,\epsilon_r)\cdot h(\epsilon_1,\ldots,\epsilon_r)}{2\sigma^2}\right\}, \tag{156}$$

$$P(\sigma|I) \propto \frac{1}{\sigma} \tag{157}$$

and $h(\epsilon_1,\ldots,\epsilon_r)\cdot h(\epsilon_1,\ldots,\epsilon_r)$ is given by Eq. (145). The normalization constant needed to ensure that the total probability is one is given by

$$\text{normalization} = \int_0^\infty d\sigma P(\sigma|I)P(\epsilon_1,\cdots,\epsilon_r|\epsilon_0,\sigma,D,I). \tag{158}$$

Making the appropriate substitutions and evaluating the integrals gives

$$\langle\sigma^2\rangle = \frac{1}{N-m_1-\cdots-m_r}\left[N\overline{d^2} - h(\epsilon_1,\ldots,\epsilon_r)\cdot h(\epsilon_1,\ldots,\epsilon_r)\right] \tag{159}$$

as the estimated standard deviation for the noise.

At this point in the calculation it would appear that another numerical example is needed to illustrate these new additional calculations and generalizations. However that is not the case, because all of the examples given in the text were computed by using these final results. That is to say, all of the computer programs used in the numerical calculations implemented this generalized calculation. To produce any specific example the model functions and the pixel smearing matrices were changed to produce the desired calculation.

5 Summary And Conclusions

Proceeding through stages, this paper has explored the deconvolution problem in varying degrees of complexity. In the first two sections, the deconvolution problem was simplified to the interpolation problem. This problem was then explored to see how varying the prior information affects the results of the calculation. These calculations illustrate that the interpolation problem is easily solved by incorporating prior information into the problem. The more cogent the prior information the better the reconstructions. However, even with very crude prior information probability theory does not lie. The interpolations always covered the correct signal at one and sometimes two standard deviations.

After obtaining an understanding of the interpolation problem, the calculation was then generalized to include the convolution. Including the convolution did not actually change the results from the first two sections, it only generalized them. The effect of prior information was then explored again to show how including different types of prior information affects the results. Again the results were essentially identical to what was found in the first two sections: Including more cogent prior information helps the deconvolution problem; but again when only limited prior information is available, the results obtained overlapped the correct result at one and sometimes two standard deviations.

Last, the entire formalism was generalized to include much more arbitrary types of prior information. This formalism, given in the preceding section, is the only version of the calculation programmed on the computer. Every example given in this work was essentially an example of the power of the general calculation presented in the previous section.

This work represents at best, a first initial exploration of the deconvolution problem. Much remains to be done. For example this work did not address the use of priors outside of the class of general Gaussian priors. While this class is wide, it does not include such priors as the entropy prior. An interesting problem would be to try to combine the best aspects of both the entropy prior and the Gaussian priors used in this calculation. Indeed there is some evidence based on work in other fields that this could be very productive, [28].

Last, this work suggests how to use probability theory to solve other types of outstanding problems. In particular relatively straightforward modification to this calculation will allow inhomogeneous linear differential equations with either boundary value, initial value, or any other type of asymptotic condition to be solved. Additionally, using the techniques developed in this paper, the moment problem, i.e., inferring a function from a finite number of its moments, should now be a solvable problem. The only change in this calculation is that the limit as the noise variance goes to zero is needed to solve this problem.

If there is a single major accomplishment for this paper, it was to demonstrate that the results one obtains depends critically on the prior information put into the problem. To put it bluntly, there is no such thing as a single best deconvolution. Every result from a Bayesian calculation is only as good as the prior information that goes into it. However, every Bayesian calculation carries with it a measure of the uncertainty in the calculation. While some priors will give poor reconstructions, probability theory warns one of this by making the uncertainty in the estimates large (large enough to cover the correct value of the signal). So even the results from very uninformative priors still give meaningful reconstructions.

Acknowledgments

The comments of Dr. C. R. Smith, and the conversations with Professor E. T. Jaynes are greatly appreciated. This work was supported by the U. S. Army through the Scientific Services Program.

References

[1] Bayes, Rev. T., "An Essay Toward Solving a Problem in the Doctrine of Chances," *Philos. Trans. R. Soc. London* **53**, pp. 370-418 (1763); reprinted in *Biometrika* **45**, pp. 293-315 (1958), and *Facsimiles of Two Papers by Bayes,* with commentary by W. Edwards Deming, New York, Hafner, 1963.

[2] Bretthorst, G. Larry, "An Introduction To Model Selection, Using Bayesian Probability Theory"

[3] Bretthorst, G. Larry, "Including The Unknown Target In The Radar Target Identification Problem"

[4] Jeffreys, H., *Theory of Probability,* Oxford University Press, London, 1939; Later editions, 1948, 1961.

[5] Jaynes, E. T., Journal of the American Statistical Association, Sept. 1979, p. 740, review of "Inference, Methods, and Decision: Towards a Bayesian Philosophy of Science." by R. D. Rosenkrantz, D. Reidel Publishing Co., Boston.

[6] Gull, S. F., "Bayesian Inductive Inference and Maximum Entropy," in *Maximum Entropy and Bayesian Methods in Science and Engineering"* **1**, pp. 53-75, G. J. Erickson and C. R. Smith *Eds.,* Kluwer Academic Publishers, Dordrecht the Netherlands.

[7] Bretthorst, G. Larry, "Bayesian Spectrum Analysis and Parameter Estimation," in *Lecture Notes in Statistics* **48**, Springer-Verlag, New York, New York, 1988.

[8] Bretthorst, G. Larry, "Bayesian Analysis I: Parameter Estimation Using Quadrature NMR Models," *J. Magn. Reson.,* in preparation.

[9] Bretthorst, G. Larry, "Bayesian Analysis II: Model Selection," *J. Magn. Reson.,* in preparation.

[10] Bretthorst, G. Larry, "Bayesian Analysis III: Spectral Analysis,"

[11] Tribus, M., *Rational Descriptions, Decisions and Designs,* Pergamon Press, Oxford, 1969.

[12] Zellner, A., *An Introduction to Bayesian Inference in Econometrics,* John Wiley and Sons, New York, 1971; Second edition 1987.

[13] Jaynes, E. T., "How Does the Brain do Plausible Reasoning?" unpublished Stanford University Microwave Laboratory Report No. 421 (1957); reprinted in *Maximum-Entropy and Bayesian Methods in Science and Engineering* **1**, pp. 1-24, G. J. Erickson and C. R. Smith, *eds.,* 1988.

[14] Jaynes, E. T., "Probability Theory – The Logic of Science," in preparation. Copies of this manuscript are available from E. T. Jaynes, Washington University, Dept. of Physics, St. Louis, MO 63011.

[15] Bretthorst, G. Larry, "An Introduction to Parameter Estimation Using Bayesian Probability Theory," in *Maximum Entropy and Bayesian Methods,* Dartmouth College 1989, P. Fougere, *ed.,* Kluwer Academic Publishers, Dordrecht the Netherlands, 1990.

[16] Jaynes, E. T., "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics,* SSC-4, pp. 227-241 (1968); reprinted in [20].

[17] Jaynes, E. T., "Marginalization and Prior Probabilities," in *Bayesian Analysis in Econometrics and Statistics,* A. Zellner, *ed.,* North-Holland Publishing Company, Amsterdam, 1980; reprinted in [20].

[18] Shore J. E., and R. W. Johnson, *IEEE Trans. on Information Theory,* **IT-26,** No. 1, pp. 26-37, 1981.

[19] Shore J. E., and R. W. Johnson, *IEEE Trans. on Information Theory,* **IT-27,** No. 4, pp. 472-482, 1980.

[20] Jaynes, E. T., *Papers on Probability, Statistics and Statistical Physics,* a reprint collection, D. Reidel, Dordrecht the Netherlands, 1983; second edition Kluwer Academic Publishers, Dordrecht the Netherlands, 1989.

[21] Shannon, C. E., "A Mathematical Theory of Communication," *Bell Syst. Tech. J.* **27,** pp. 379-423 (1948).

[22] Laplace, P. S., *A Philosophical Essay on Probabilities,* unabridged and unaltered reprint of Truscott and Emory translation, Dover Publications, Inc., New York, 1951, original publication data 1814.

[23] Abel, N. H., *Crelle's Jour.,* Bd. 1 (1826).

[24] Cox, R. T., "Probability, Frequency, and Reasonable Expectations," *Amer. J. Phys.* **14,** pp. 1-13 (1946).

[25] Jaynes, E. T., "Where Do We Stand On Maximum Entropy?" in *The Maximum Entropy Formalism,* R. D. Levine and M. Tribus *Eds.,* pp. 15-118, Cambridge: MIT Press, 1978; Reprinted in [20].

[26] Bretthorst, G. Larry, "Bayesian Model Selection: Examples Relevant to NMR," *Maximum Entropy and Bayesian Methods,* J. Skilling *ed.,* pp. 377-388, Kluwer Academic Publishers, Dordrecht the Netherlands, 1989. *J. Magn. Reson.,* in preparation.

[27] Jaynes, E. T,'., 1989, "The Theory of Radar Target Discrimination"

[28] Sivia, D. S., and C. J. Carlie, 1992, "Molecular spectroscopy and Bayesian spectral analysis – how many lines are there?" *J. Chem. Phys.,* **96**(1) pp. 170-178.