Geostatistically Constrained Seismic Deconvolution

Jonathan Kane, Abdulrahman Al-Moqbel, William Rodi, and M. Nafi Toksöz
Earth Resources Laboratory
Dept. of Earth, Atmospheric, and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

We present a method for combining seismic deconvolution and geostatistical interpolation. Both problems are posed as a single joint inverse problem in the maximum likelihood framework. Joint inversion allows for well data to improve the deconvolution results and, conversely, allows the seismic data to improve the interpolation of well data. Traditional interpolation and trace-by-trace deconvolution are special cases of the joint inverse problem. Inversion is performed on 2-D and 3-D field data sets.

1 Introduction

Geophysical problems often involve estimating some field of petrophysical parameters given multiple data sets related to that field by different operators. These data sets might be of different resolution and provide different, but complementary, information about the parameter field of interest. For example, seismic data potentially provides information about the three-dimensional distribution of velocity and density in the Earth. However, this information is limited by the frequency content of the input signal, dispersion, aperature of the receivers, and other effects. Well data is of very high resolution, approaching actual point samples of the petrophysical parameters, but it is available only locally at limited locations within a three-dimensional volume. Combining these two data sets in a logical manner should result in a superior estimated field than by using just one of them. Simultaneously inverting multiple data sets in a single inverse problem is not neccessarily a straightforward operation. We present a method that solves this problem in a statistically optimal and computational efficient manner.

The particular applications we have in mind, deconvolution and interpolation, are common to many scientific fields, including geophysics. Deconvolution has been used in geophysics over the last half century (Robinson and Trietel, 1980). It removes the blurring in seismic data caused by a seismic wavelet, the Earth, or even previous processing, thus resulting in a sharpened tomographic image. Interpolation has received great attention in the geostatistical community where, given only isolated measurements of a field and some knowledge of its spatial roughness, one attempts to obtain optimal estimates of parameters at other locations.

We begin by first describing the seismic data we use. We then derive the deconvolution and interpolation problems in the maximum likelihood framework, followed by a derivation of the joint problem in the same framework. In looking at the joint problem we see some serious computational considerations to overcome. These problems are solved via an iterative method known as the *conjugate gradients* algorithm (see Appendix A). Joint inversion is then performed on field data sets and results are discussed.

2 Field Data

In this work we will use two seismic data sets, each cut out of a larger volume. All the seismic data is in post-stack form with a sampling interval of 4 ms. Post-stack data ideally means that previous processing (such as migration and stacking) has put the seismic data in a form such that the 1-D convolutional seismic model holds. Thus our observed seismic data is assumed to be the result of a one dimensional linear operation (a convolution) of a seismic wavelet on a reflectivity sequence.

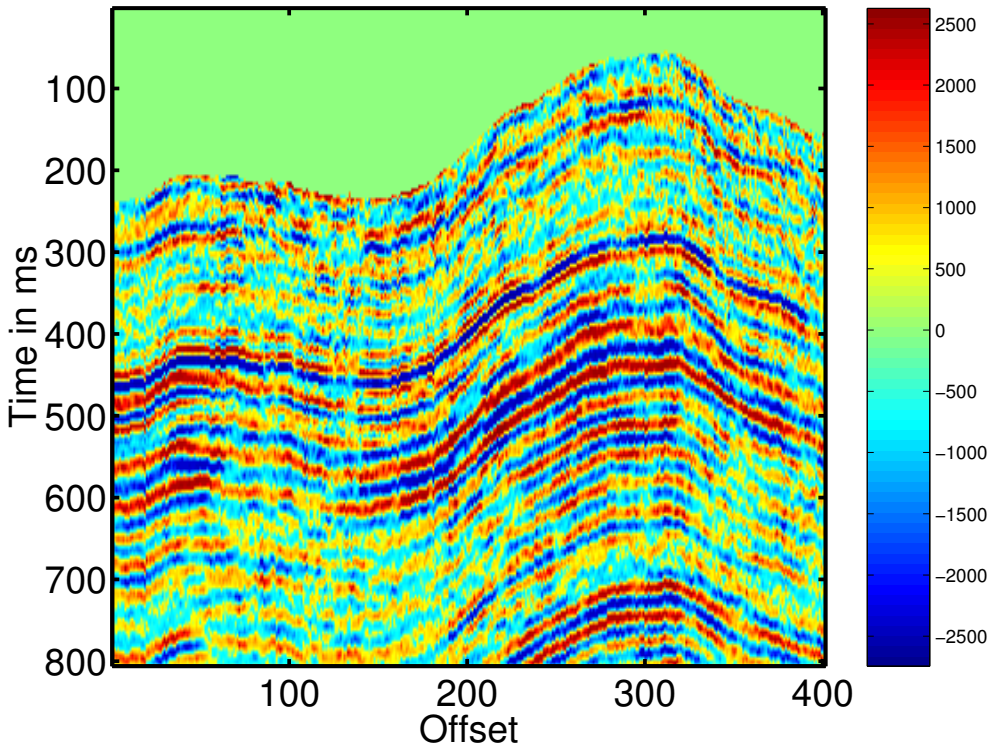

Figure 1: 2-D slice of field seismic data

The first data set is a 2-D slice. Its negative is shown in Figure 1. The negative is plotted because we observe below that positive values of reflectivity in the wells correspond to negative values in the seismic data. Plotting the negative will facilitate easier comparison to reflectivity obtained from inversion. The blank area at the top of the data is due to a later flattening of the seismic data along a horizon. The flattening pushes some of the seismic data out of the cube which is then not used when we perform inversion. Therefore, we do not show it here. The flattening is done along a horizon corresponding to a known reservoir and is necessary for the interpolation of well data.

The other data set used is a 3-D subvolume also cut out of the larger volume. A 3-D slice plot of this subvolume is shown in Figure 2. The negative is again plotted and the top portion is blank because of subsequent flattening. Another alternate 3-D slice plot is shown in Figure 3. This one is plotted for visual comparison of inversion results below.

The 2-D and 3-D flattened seismic data sets are shown in Figures 4 and 5, respectively. After performing inversion the resulting reflectivity field is subsequently unflattened for comparison with the unflattened seismic.

The 2-D section is intersected by one reflectivity well at crossline 211 that goes from 272 ms to 560 ms, as shown in Figure 6. This 2-D data set is sufficient to demonstrate the inversion method of this paper. The subvolume includes 7 reflectivity wells as illustrated in Figure 7, and is sufficient to demostrate the 3-D applicability of the inversion method. The reflectivity well data is obtained by taking existing velocity and density well logs measured in depth and converting them into reflectivity sequences measured in time. It will be used for interpolation purposes in Section 4. The reflectivity well data is correspondingly adjusted in time when the seismic data is flattened.

3 Deconvolution

If the seismic wavelet in the convolutional model is known, a reasonable objective might be to deconvolve it out of the seismic data to obtain the reflectivity, thereby obtaining a better image of subsurface material

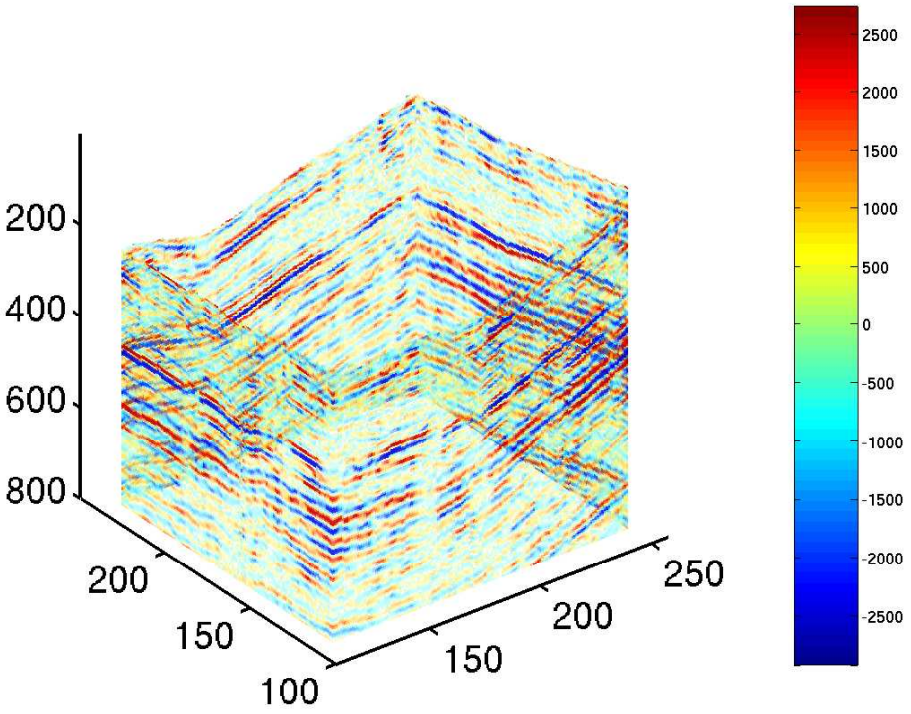

Figure 2: 3-D subvolume of field seismic data

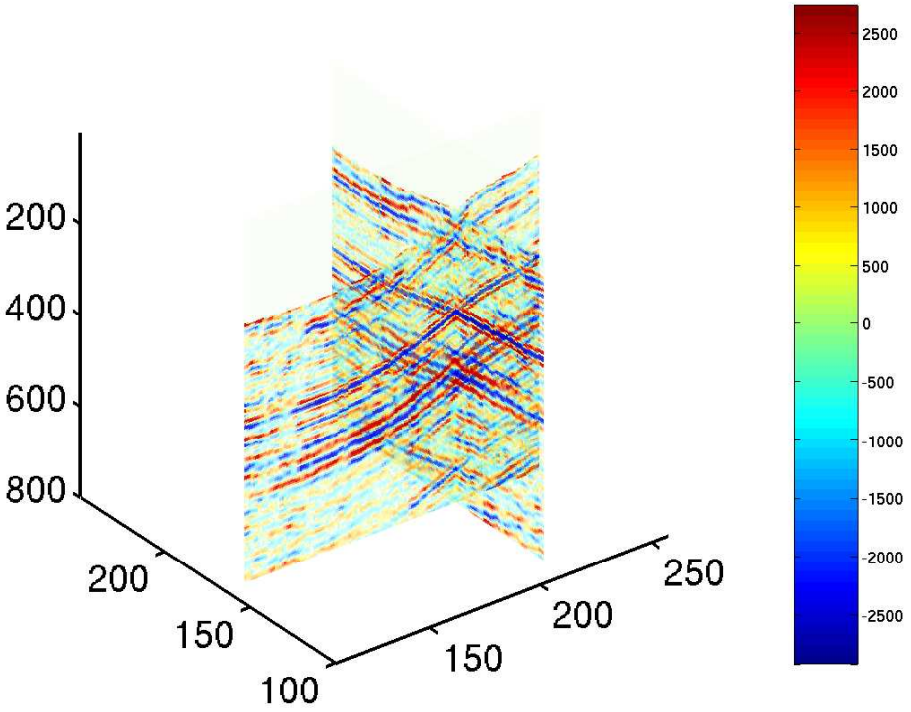

Figure 3: Alternate plot of 3-D seismic subvolume.

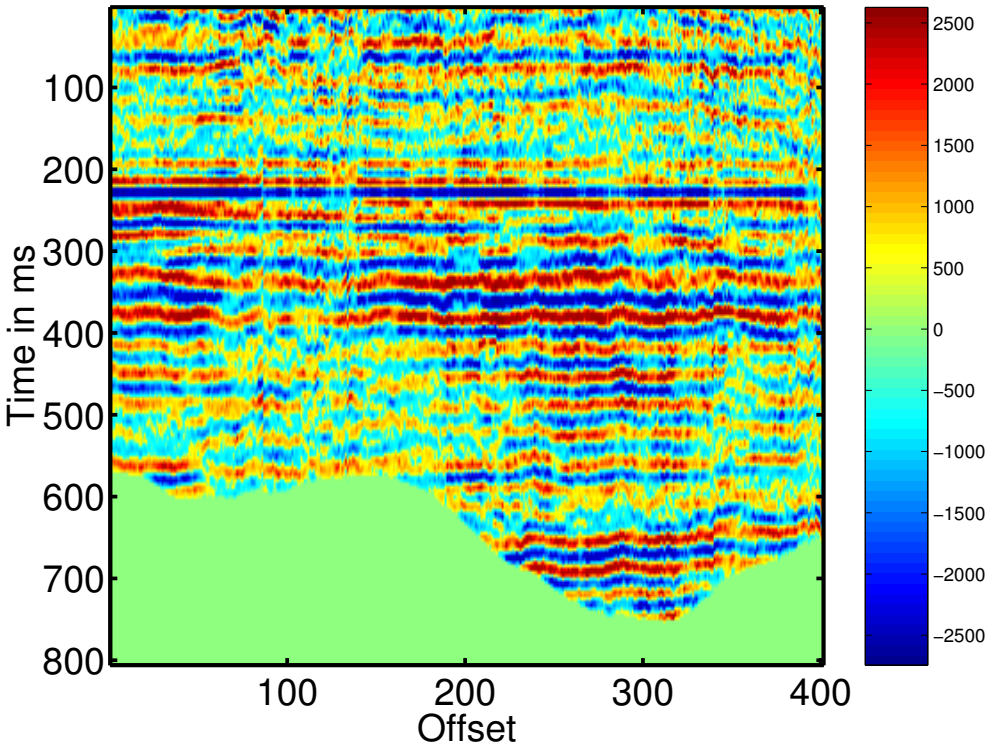

Figure 4: Flattened 2-D seismic data.

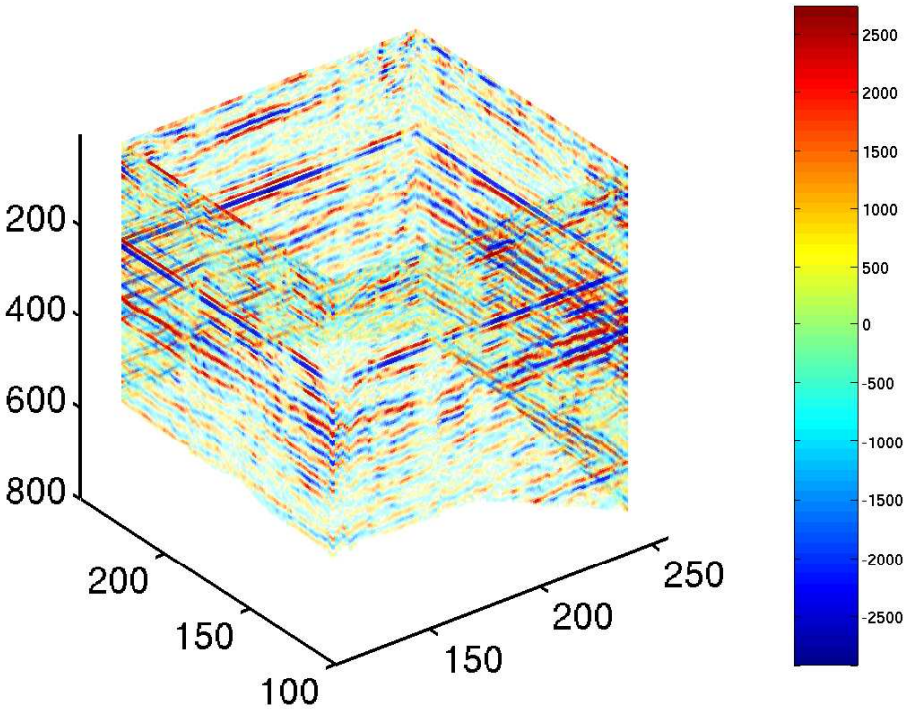

Figure 5: Flattened 3-D seismic data.

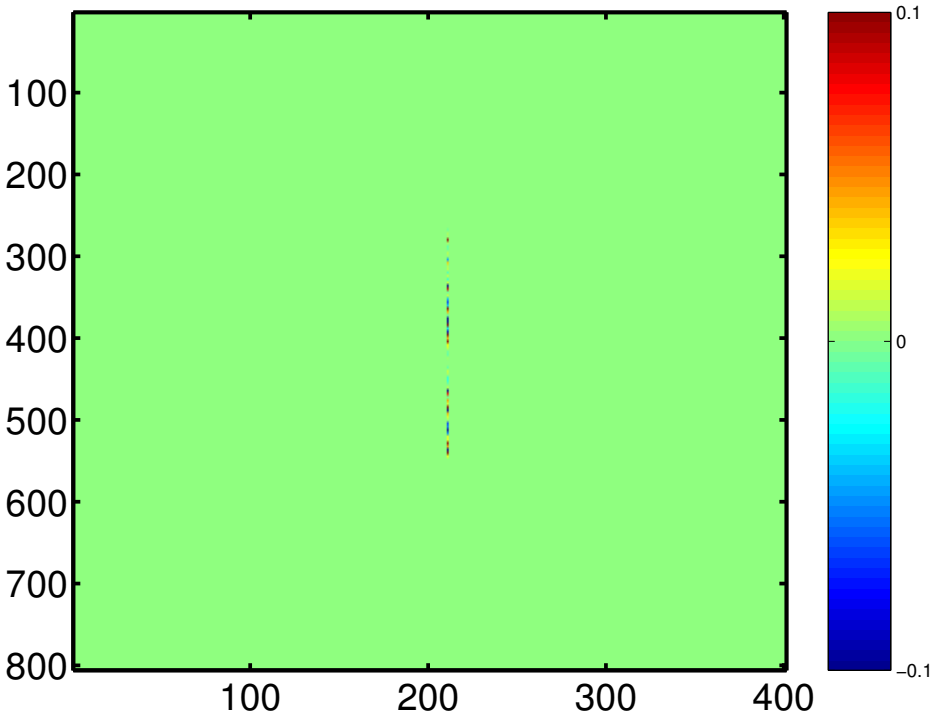

Figure 6: Reflectivity well data in unflattened 2-D section.

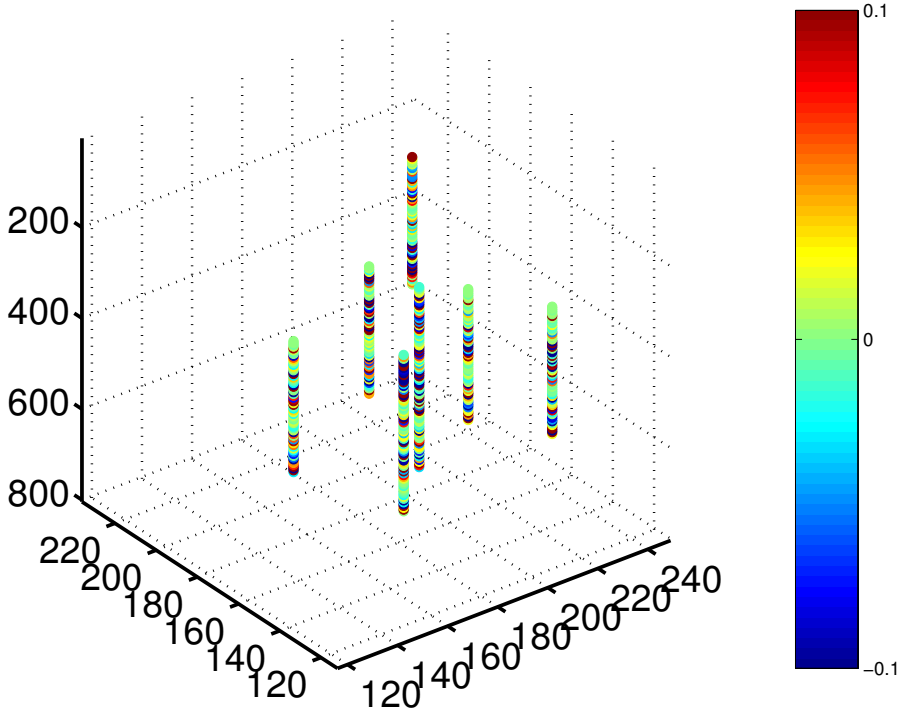

Figure 7: Reflectivity well data in unflattened 3-D subvolume.

discontinuities.

There exists at least 50 years of literature on deconvolution methods (Wiener, 1949; Robinson and Trietel, 1980; Yilmaz, 1987; Saggaf and Robinson, 2000). The convolution of two sequences $\check{h}[k]$ and $\check{f}[k]$ to produce a third sequence $\check{s}[k]$ is written

$$\begin{aligned} \check{s}[k] &= \check{h}[k] * \check{f}[k] && (1) \\ &= \sum_{k'=-\infty}^{\infty} \check{h}[k-k']\check{f}[k']. && (2) \end{aligned}$$

Since convolution is a linear forward problem, we will represent equation 1 as a matrix/vector product:

$$\check{\mathbf{s}}_l = \check{\mathbf{H}}\check{\mathbf{f}}_l. \tag{3}$$

$\check{\mathbf{f}}_l, l \in \{1, 2, \ldots, N\}$, represents each 1-D vertical trace of reflectivity within either a 2-D cross section or 3-D subvolume of reflectivity, where N is the number of traces. $\check{\mathbf{s}}_l, l \in \{1, 2, \ldots, N\}$, represents each 1-D vertical seismic trace in one of the seismic data sets.

Our seismic model may not be precise. We introduce the vectors $\check{\mathbf{n}}_{\mathbf{H},l}, l \in \{1, 2, \ldots, N\}$, to denote the amount of misfit between our theoretically predicted data and the real data for each trace $\check{\mathbf{s}}_l$. This changes equation 3 into

$$\check{\mathbf{s}}_l = \check{\mathbf{H}}\check{\mathbf{f}}_l + \check{\mathbf{n}}_{\mathbf{H},l}. \tag{4}$$

The operator $\check{\mathbf{H}}$ must be operated on each vertical reflectivity trace, so we put $\check{\mathbf{H}}$ into a block diagonal matrix and arrange the vectors $\check{\mathbf{f}}_l$, $\check{\mathbf{s}}_l$, and $\check{\mathbf{n}}_{\mathbf{H},l}$ head to tail, resulting in the following matrix/vector system of equations:

$$\underbrace{\begin{bmatrix} \check{\mathbf{s}}_1 \\ \check{\mathbf{s}}_2 \\ \check{\mathbf{s}}_3 \\ \vdots \\ \check{\mathbf{s}}_N \end{bmatrix}}_{\mathbf{s}} = \underbrace{\begin{bmatrix} \check{\mathbf{H}} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{H}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \check{\mathbf{H}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \check{\mathbf{H}} \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \check{\mathbf{f}}_1 \\ \check{\mathbf{f}}_2 \\ \check{\mathbf{f}}_3 \\ \vdots \\ \check{\mathbf{f}}_N \end{bmatrix}}_{\mathbf{f}} + \underbrace{\begin{bmatrix} \check{\mathbf{n}}_{\mathbf{H},1} \\ \check{\mathbf{n}}_{\mathbf{H},2} \\ \check{\mathbf{n}}_{\mathbf{H},3} \\ \vdots \\ \check{\mathbf{n}}_{\mathbf{H},N} \end{bmatrix}}_{\mathbf{n_H}}. \tag{5}$$

This equation represents the operation of convolving each reflectivity trace with the same seismic wavelet and adding noise to produce the seismic data. We represent equation 5 more concisely as

$$\mathbf{s} = \mathbf{H}\mathbf{f} + \mathbf{n_H}. \tag{6}$$

If $\check{\mathbf{H}}$ is invertible, then the block diagonal matrix $\mathbf{H}$ has a block diagonal inverse, $\mathbf{H}^{-1}$. If $\check{\mathbf{H}}$ is non-invertible, or if the noise $\mathbf{n_H}$ is too large, the inverse problem becomes *ill-posed* (Rodi, 1989) and some regularization is required. The block diagonal form of $\mathbf{H}$ allows for fast forward and inverse operations, since convolution or deconvolution can be done in a trace by trace manner. The whole matrix never need be constructed. If a regularization operator is used that enforces horizontal correlation between separate reflectivity traces, the generalized inverse matrix (Rodi, 1989) will no longer be block diagonal and we can't perform the inversion in a trace by trace manner. There are still ways of solving equation 6 for $\mathbf{f}$ without explicitely forming and inverting $\mathbf{H}$. These methods are detailied in Section 7.

$\mathbf{n_H}$ is a random vector and can be thought of as the combination of noise in the seismic data gathering process, inaccuracies in the convolution model, and errors in previous processing steps. There may exist statistical covariance in this noise but it is all but impossible to estimate. Therefore we idealize it to be white noise with equal variance at all locations. We further constrain it and all other random vectors in this chapter to be Gaussian. This is done mainly because of the mathematical tractability of Gaussian random vectors. We denote the variance of $\mathbf{n_H}$ as $\sigma^2_{\mathbf{H}}$.

3.1 Estimating a Seismic Wavelet

It is useful to extract a seismic trace at a location where there also exists reflectivity well data for a visual comparison. This is done for the well data at crossline 211 in the 2-D section, shown with its corresponding

negative seismic trace in Figure 8. The scales of the reflectivity and seismic data differ by several orders of magnitude so, to visually compare them, the amplitude of the reflectivity data is multiplied by $12,000$. Plotting the negative of the seismic versus the reflectivity in Figure 8 allows us to see the close match between major events in the log and the seismic. This match is not exact though. In addition to the polarity reversal and scale difference of the two data sets, we also see that they appear to have different frequency content (the seismic is blurred compared to the reflectivity). Assuming convolution with a wavelet as the operator that changes reflectivity into seismic data, we must estimate the wavelet. To do this it is useful to look at both data in the Fourier domain.

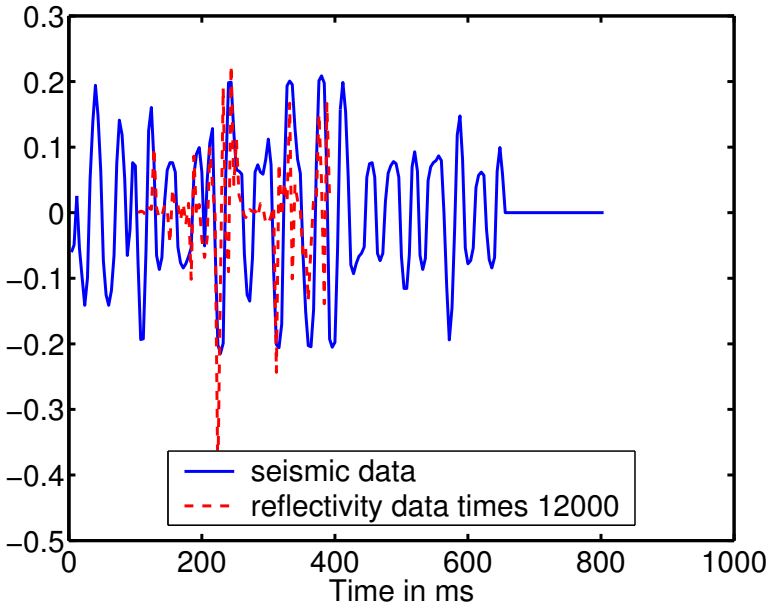

Figure 8: Seismic trace and reflectivity well.

In Figure 9 we plot the amplitude spectra of the seismic data from Figure 8 in a blue solid line. Noticing that its spectra looks similar to that of a Ricker wavelet, we take the amplitude spectra of Ricker wavelets of varying central frequency and multiply them by the amplitude spectra of the reflectivity until a good is fit is found to the seismic data. The best fitting result is shown in Figure 9 as the dashed red line for a Ricker wavelet with center frequency of $28Hz$. The amplitude spectrum of the best fitting Ricker wavelet is shown along with the reflectivity amplitude spectrum in Figure 10. Inverse transforming the spectrum of the Ricker wavelet gives the zero phase wavelet itself, shown in Figure 11. It has negative polarity because of the polarity relationship between the seismic and reflectivity data.

If we reconvolve our chosen wavelet from Figure 11 with the reflectivity shown in Figure 8, we don't quite reproduce the observed seismic. This is shown in Figure 12. This shows the inherent inaccuracy in our seismic convolutional model expressed by the $\mathbf{n_H}$ vector.

In this work we have restricted our seismic wavelet to be a Ricker wavelet. Alternatively one could estimate a seismic wavelet from locations in the field where well data and seismic overlap. This was tried for this field data set and the results were sub-optimal. The estimated wavelet at one well location was different from another and using a single best fitting Ricker gave superior results.

3.2 Estimating $\sigma^2_{\mathbf{H}}$

In order to estimate the noise variance, $\sigma^2_{\mathbf{H}}$, in the seismic data, we observe the energy in the amplitude spectrum of the seismic trace in Figure 9 at high frequencies. This deviates from the product of the wavelet and reflectivity amplitude spectra shown in the same plot. We will assume that this energy in the seismic data at high frequencies is the noise in the system. A green dash-dot line is fit to this noise. We consider

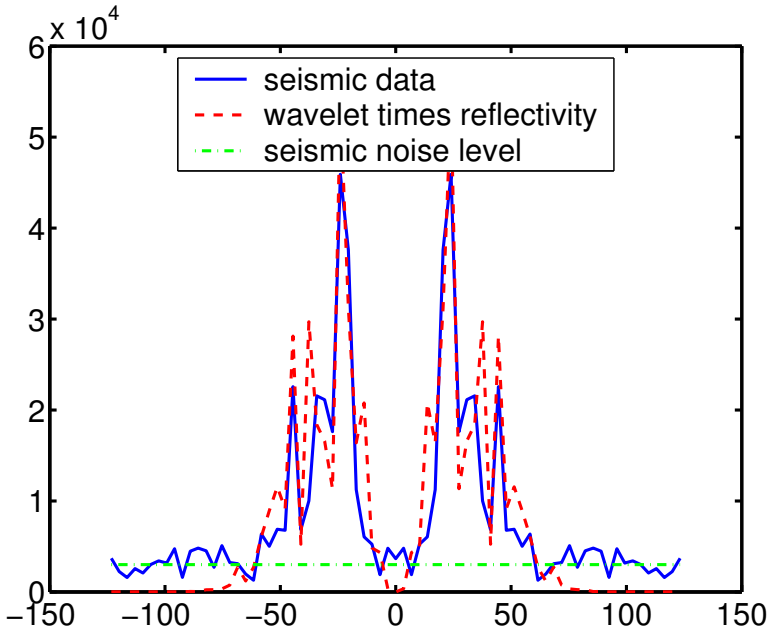

Figure 9: Amplitude spectra of seismic trace, best fitting wavelet convolved with reflectivity in well, and noise.

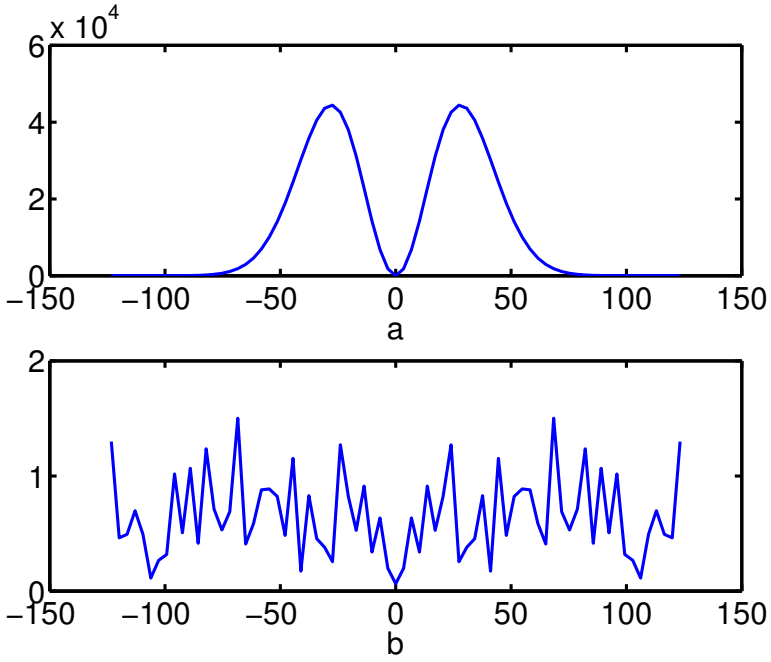

Figure 10: Amplitude spectra of a) best fitting Ricker wavelet, b) reflectivity in well.

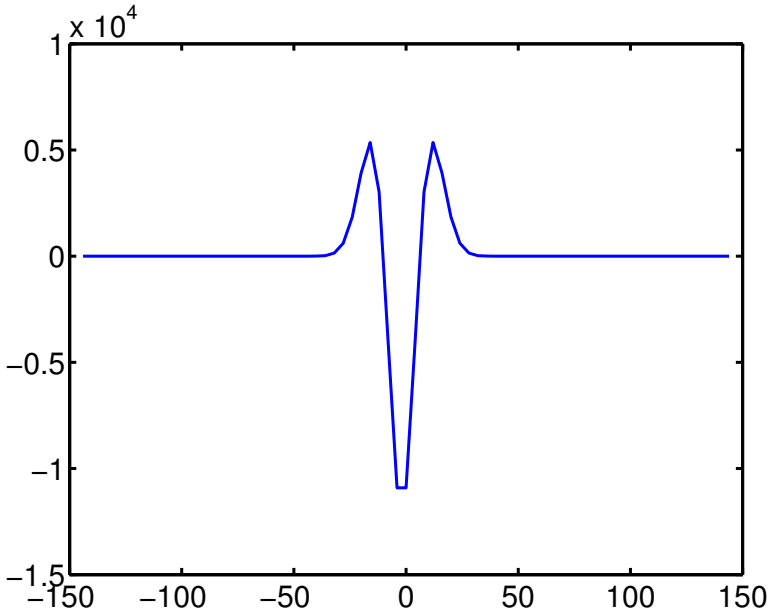

Figure 11: Ricker wavelet in time.

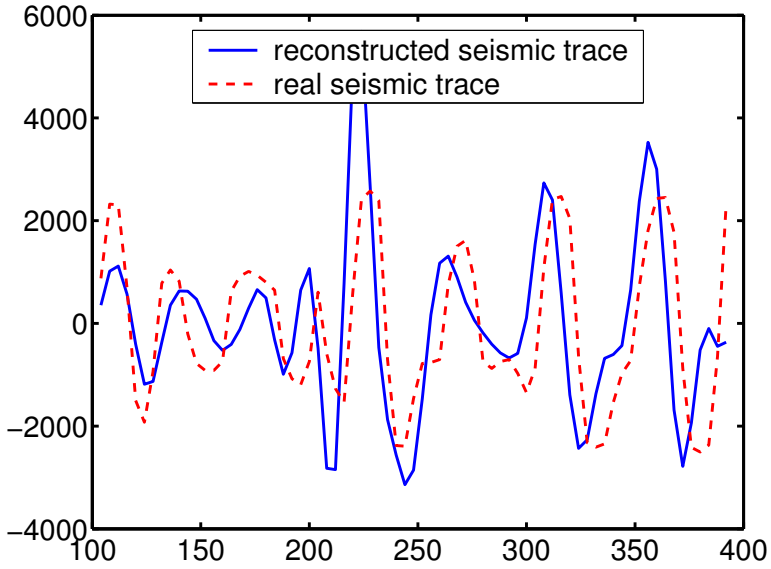

Figure 12: Reconvolved wavelet with reflectivity and true seismic trace.

this line to be the amplitude spectrum of the noise. By squaring the amplitude spectrum we get the power spectrum. Inverse Fourier transforming the power spectrum gives the covariance function. The value of the covariance function at zero lag is $\sigma^2_{\mathbf{H}}$. By this method we obtain the value $\sigma^2_{\mathbf{H}} = 1.05 \times 10^6$ to use in inversion in Section 8 below.

4 Interpolation

Interpolation of geophysical data is usually derived in a minimum variance framework and is called *kriging* in the geostatistical community (Deutsch and Journel, 1998). We will instead define the interpolation problem in the maximum likelihood framework (Rodi, 1989), as we did for deconvolution, because it allows for the solution of joint problems. Since we are not using the minimum variance derivation, we will refer to the interpolation problem as *interpolation* rather than *kriging* for the remainder of this document.

The minimum variance framework leads to the same solution as the maximum likelihood framework, but it leads to a different system of equations to solve (Fomel, 2001). In fact, the system of equations it produces is not sparse, which disallows efficient solution unless many ad-hoc approximations are made. This is especially important in the joint inverse problem presented below. Iterative methods are the only practical method of solution and are only possible in the maximum likelihood framework.

Posing interpolation as a discrete linear inverse problem in the maximum likelihood framework requires 1) a "picking" operator that picks isolated observations out of a field of values, and, 2) a regularization constraint that forces continuity between adjacent estimated values. We examine the first of these requirements in this section and address regularization in the next.

As in Section 3, we let $\check{\mathbf{f}}_l, l \in \{1, 2, \ldots, N\}$, again be a reflectivity trace. $\check{\mathbf{d}}_m, m \in \{1, 2, \ldots, M\}$, will represent an observed reflectivity well picked out the reflectivity data with the picking operator, $\check{\mathbf{P}}$, at M trace locations. $\check{\mathbf{d}}_m$ is related to $\check{\mathbf{f}}_l$ via the following equation:

$$\underbrace{\begin{bmatrix} \check{\mathbf{d}}_1 \\ \check{\mathbf{d}}_2 \\ \vdots \\ \check{\mathbf{d}}_M \end{bmatrix}}_{\mathbf{d}} = \underbrace{\begin{bmatrix} \check{\mathbf{P}}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots\cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \check{\mathbf{P}}_2 & \cdots\cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots\cdots & \check{\mathbf{P}}_M \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \check{\mathbf{f}}_1 \\ \check{\mathbf{f}}_2 \\ \check{\mathbf{f}}_3 \\ \check{\mathbf{f}}_4 \\ \vdots \\ \vdots \\ \check{\mathbf{f}}_N \end{bmatrix}}_{\mathbf{f}} + \underbrace{\begin{bmatrix} \check{\mathbf{n}}_{\mathbf{P},1} \\ \check{\mathbf{n}}_{\mathbf{P},2} \\ \vdots \\ \check{\mathbf{n}}_{\mathbf{P},M} \end{bmatrix}}_{\mathbf{n_P}}, \tag{7}$$

or, more concisely,

$$\mathbf{d} = \mathbf{P}\mathbf{f} + \mathbf{n_P}. \tag{8}$$

$\check{\mathbf{P}}_m$ is a 201×201 matrix of zeros with a 1 value on the diagonal at locations corresponding to locations in the corresponding $\check{\mathbf{f}}_l$ where well data is observed. Since well data is only observed in a few of the reflectivity traces out of the whole vector $\mathbf{f}$, the $\mathbf{P}$ matrix is rectangular with a $\check{\mathbf{P}}$ entry only for each well log observed. It can be seen that $\mathbf{P}$ itself is a rectangular matrix made up of a subset of the rows of the identity matrix. It "picks" the observed data out of the reflectivity field $\mathbf{f}$ and adds some noise, $\mathbf{n_P}$, to it to make the well data. $\mathbf{n_P}$ is another Gaussian white noise random vector that we assume to be uncorrelated with $\mathbf{f}$ and with constant variance $\sigma^2_{\mathbf{P}}$ at all points.

The matrix $\mathbf{P}$ is wider than it is long, thus it does not have a unique inverse. The well data cannot be inverted to give interwell data unless we add some kind of regularization constraint to the problem. Also, the interpolation method we are presenting can only interpolate linear trends. Any complicated structure between points to be interpolated must be accounted for before the interpolation is performed. This was done when the seismic data was flattened and the well data adjusted accordingly. The interwell nonlinear trends in the reflectivity field were effectively removed due to this data transformation.

4.1 Estimating $\sigma_{\mathbf{P}}^2$

It can be safely assumed that the well data is of much higher accuracy than the seismic data. Therefore $\sigma_{\mathbf{P}}^2$ should be much smaller than $\sigma_{\mathbf{H}}^2$. As $\sigma_{\mathbf{P}}^2 \mapsto 0$ we have noise free well data. In this work we set $\sigma_{\mathbf{P}}^2 = 7.65 \times 10^{-7}$ which, for all practical purposes, says that the well data should be exactly reproduced in any solution to the joint inverse problem.

5 Regularization

In the deconvolution problem, some regularization may be needed to make the matrix $\mathbf{H}$ invertible, or to account for excessive noise in the system. In the interpolation problem, regularization is absolutely required if any interpolation is to be done. Besides making inverse problems solvable, a regularization term tells us something about the continuity of the estimated field.

The covariance[1] of a certain class of functions can be represented equivalently by a differential operator. By attempting to minimize the result of a differential[2] operator, L, acting on the reflectivity field, f, we effectively smooth the reflectivity. Put another way, we believe that for a function f with α bounded derivatives there exists some operator L such that the new function $n_L = Lf$ has exactly 0 bounded derivatives. This allows us to use an operator to describe exactly how smooth f is. f in our case is multidimensional, so derivatives can be taken in different directions. In this thesis we will define an L only in horizontal directions: L_X and L_Y. Reflectivity is actually *anti-correlated* vertically. Thus an operator L_Z would be some kind of integral operator instead of a differential operator in that direction, and might destroy the positive-definiteness of the $\mathbf{L}^T\mathbf{L}$ matrix below. Therefore we impose no constraint on the vertical correlation structure of the reflectivity.

Since we are dealing with discrete vectors of coefficients, $\mathbf{f}$, rather than continuous functions, we represent the differential operators L_X and L_Y by the matrices $\mathbf{L}_X$ and $\mathbf{L}_Y$. We will assume that all functions f have the same order of differentiability at all points. This will make $\mathbf{L}_X$ and $\mathbf{L}_Y$ Toeplitz matrices. We then have the equation:

$$\underbrace{\begin{bmatrix}\mathbf{L}_X\\ \mathbf{L}_Y\end{bmatrix}}_{\mathbf{L}}\mathbf{f} = \underbrace{\begin{bmatrix}\mathbf{n}_{\mathbf{L}_X}\\ \mathbf{n}_{\mathbf{L}_Y}\end{bmatrix}}_{\mathbf{n}_\mathbf{L}}, \tag{9}$$

where $\mathbf{n}_\mathbf{L}$ is a discrete vector of uncorrelated coefficients.

Depending on whether we are doing stochastic or Bayesian inversion the noise term $\mathbf{n}_\mathbf{L}$ will have either a deterministic or probabilistic interpretation. If we are Bayesian, $\mathbf{n}_\mathbf{L}$ is a white Gaussian random vector with equal variance $\sigma_\mathbf{L}^2$ at all points. We can then write that $\mathbf{n}_\mathbf{L} = \sigma_\mathbf{L}\mathbf{n}'_\mathbf{L}$, where $\mathbf{n}'_\mathbf{L}$ is a unit variance white noise Gaussian random vector. We then have

$$\mathbf{Lf} = \sigma_\mathbf{L}\mathbf{n}'_\mathbf{L}. \tag{10}$$

Assuming that $\mathbf{L}$ is invertible, we can rearrange to have

$$\mathbf{f} = \sigma_\mathbf{L}\mathbf{L}^{-1}\mathbf{n}'_\mathbf{L}. \tag{11}$$

This equation implies that $\mathbf{f}$ is also a random vector. From this we can derive the covariance matrix, $\mathbf{C}$, of the random vector $\mathbf{f}$. Noting that reflectivity is theoretically a zero mean random variable, this is done by

[1] To be speaking of *covariance* requires that the function in question to be stochastic in nature (Papoulis, 1991). For deterministic functions there is a corresponding structure called the *reproducing kernel* (Wahba, 1990) that describes the inter-relationship between points in the function. In this paper we will use the term *covariance* to describe either the stochastic or deterministic situation.

[2] The "differential" operator need not be an exact derivative operator, but rather any operator that tends to boost high frequencies relative to low ones in the Fourier domain. In fact, since a true derivative operator has a zero eigenvalue at zero frequency, it is not uniquely invertible and therefore it is not possible to invert it for a covariance function.

expanding $\mathbf{f}\mathbf{f}^T$ and taking the expected value as follows

$$\begin{aligned}
\mathbf{C} &= E\left[\mathbf{f}\mathbf{f}^T\right] && (12)\\
&= \sigma_{\mathbf{L}}^2 E\left[\mathbf{L}^{-1}\mathbf{n}'_{\mathbf{L}}(\mathbf{n}'_{\mathbf{L}})^T\mathbf{L}^{-T}\right] && (13)\\
&= \sigma_{\mathbf{L}}^2 \mathbf{L}^{-1} \underbrace{E\left[\mathbf{n}'_{\mathbf{L}}(\mathbf{n}'_{\mathbf{L}})^T\right]}_{\mathbf{I}} \mathbf{L}^{-T} && (14)\\
&= \sigma_{\mathbf{L}}^2 \mathbf{L}^{-1}\mathbf{L}^{-T} && (15)\\
&= \sigma_{\mathbf{L}}^2 (\mathbf{L}^T\mathbf{L})^{-1} && (16)\\
&= \sigma_{\mathbf{L}}^2 \mathbf{R}. && (17)
\end{aligned}$$

$\mathbf{R}$ has the value 1 along the diagonal and decreasing values away from the diagonal.

If we are performing stochastic inversion then $\sigma_{\mathbf{L}}$ must be regarded as a deterministic parameter and $\mathbf{f}$ loses its probabilistic character. In this case $\sigma_{\mathbf{L}}^2$ becomes simply an adjustable constant, and $\mathbf{R}$ is a matrix that affects the smoothness of the inversion result.

It is important to review the difference between the stochastic inversion and Bayesian inversion lines of thought. Stochastic inversion doesn't allow the idea of a probablility distribution to enter on the vector $\mathbf{f}$ to be estimated. It simply uses $\sigma_{\mathbf{L}}^2\mathbf{R}$ as a stabilizing function that helps make an unsolvable inverse problem solvable. The optimal $\sigma_{\mathbf{L}}^2$ is to be determined such that it produces a solution $\mathbf{f}_{est}$ as close to the true $\mathbf{f}$ as possible. The Bayesian method believes that $\mathbf{f}$ has a probability distribution that must be honored in finding the optimal $\mathbf{f}_{est}$. In short, stochastic inversion frees the inverse theorist to improve the solution by varying $\sigma_{\mathbf{L}}^2$ while Bayesian inversion fixes the solution (perhaps sub-optimally) based on prior knowledge of $\sigma_{\mathbf{L}}^2$.

5.1 Estimating $\sigma_{\mathbf{L}}^2$

As mentioned above, in the stochastic inversion framework, $\sigma_{\mathbf{L}}^2$ is an adjustable parameter and no estimation needs to be done. We choose it by varying its value until a result is reached that fits our conception of a realistic solution.

For Bayesian inversion a sample data set is required. From a single sample data set both the operator $\mathbf{L}$ and the noise level $\sigma_{\mathbf{L}}^2$ could be estimated. Estimating $\mathbf{L}$ could be done by calculating a multidimensional power spectrum of the sample $\mathbf{f}$, taking the square root of its inverse, and then inverse Fourier transforming. The $\mathbf{L}$ obtained this way might not be very sparse, which would slow its use in practice. Sparse approximate models could be fit to the true one to circumvent this problem. $\sigma_{\mathbf{L}}^2$ would simply be the estimated variance of the sample $\mathbf{f}$.

5.2 Choosing a differential operator L

So far we have said nothing about the exact structure of L. It was mentioned above that in order to define a covariance in terms of a differential operator, the differential operator needs to be invertible. For the puposes of this paper we will use an easily defined class of differential operators, L, with a corresponding covariance function that has an adjustable parameter, a, controlling the macroscopic correlation length. We define such an operator in each horizontal direction as

$$L_X = \frac{1}{\sqrt{a}}\left(\frac{d}{dx} - a\delta(x)\right), \quad (18)$$

$$L_Y = \frac{1}{\sqrt{a}}\left(\frac{d}{dy} - a\delta(y)\right). \quad (19)$$

The covariance operator corresponding to either of these L operators is the familiar exponential covariance function used in geostatistics (Tarantola, 1987, p. 579):

$$C(x, x') = e^{\frac{|x-x'|}{a}} \quad (20)$$

Discretizing these L operators will result in matrices

$$\mathbf{L}_X = \frac{1}{\sqrt{2a}}(\mathbf{D}_X - a\mathbf{I}) \tag{21}$$

and

$$\mathbf{L}_Y = \frac{1}{\sqrt{2a}}(\mathbf{D}_Y - a\mathbf{I}), \tag{22}$$

where $\mathbf{D}_X$ and $\mathbf{D}_Y$ are first differencing matrices and $\mathbf{I}$ is the identity matrix. These matrices correspond to the continuous operators $\frac{d}{dx}$ and $\delta(x)$ in equations 18 and 19. In this case $\frac{d}{dx}$ becomes a first differencing matrix and $\delta(x)$ becomes the identity matrix. Applying these matrices to a vector then amounts to explicit convolution with a 2 point filter - a very fast $O(N)$ operation. Writing out a few rows of the $\mathbf{L} = \begin{bmatrix} \mathbf{L}_X \\ \mathbf{L}_Y \end{bmatrix}$ matrix explicitely, we have

$$\underbrace{\begin{bmatrix} \mathbf{n_L}_{,1} \\ \mathbf{n_L}_{,2} \\ \mathbf{n_L}_{,3} \\ \mathbf{n_L}_{,4} \\ \mathbf{n_L}_{,5} \\ \mathbf{n_L}_{,6} \\ \mathbf{n_L}_{,7} \\ \vdots \\ \mathbf{n_L}_{,N} \end{bmatrix}}_{\mathbf{n_L}} = \frac{1}{\sqrt{2a}} \underbrace{\left[\begin{array}{ccccccccc} -(1+a) & a & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & -(1+a) & a & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -(1+a) & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \multicolumn{9}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ a & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -(1+a) \\ \hline -(1+a) & 0 & 0 & \cdots & a & 0 & 0 & \cdots & 0 \\ 0 & -(1+a) & 0 & \cdots & 0 & a & 0 & \cdots & 0 \\ 0 & 0 & -(1+a) & \cdots & 0 & 0 & a & \cdots & 0 \\ \multicolumn{9}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -(1+a) \end{array}\right]}_{\mathbf{L}} \underbrace{\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \\ \mathbf{f}_5 \\ \mathbf{f}_6 \\ \mathbf{f}_7 \\ \vdots \\ \mathbf{f}_N \end{bmatrix}}_{\mathbf{f}}. \tag{23}$$

We again notice the extreme sparsity of this matrix.

6 Joint Problem

Since $\mathbf{n_H}$ and $\mathbf{n_P}$ are multi-Gaussian random vectors, their probability distributions have particularly tractable and convenient forms:

$$Pr(\mathbf{n_H}) = \frac{1}{(2\pi^{\frac{N_\mathbf{s}}{2}}\ det(\mathbf{C_H}))}\ exp[-\frac{1}{2}(\mathbf{s}-\mathbf{Hf})^T\mathbf{C_H}^{-1}(\mathbf{s}-\mathbf{Hf})] \tag{24}$$

and

$$Pr(\mathbf{n_P}) = \frac{1}{(2\pi^{\frac{N_\mathbf{d}}{2}}\ det(\mathbf{C_P}))}\ exp[-\frac{1}{2}(\mathbf{d}-\mathbf{Pf})^T\mathbf{C_P}^{-1}(\mathbf{d}-\mathbf{Pf})]. \tag{25}$$

As mentioned above, we assume $\mathbf{n_H}$ and $\mathbf{n_P}$ to be identically distributed white noise. Thus $\mathbf{C_H} = \sigma_\mathbf{H}^2\mathbf{I}$ and $\mathbf{C_P} = \sigma_\mathbf{P}^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. $N_\mathbf{s}$ and $N_\mathbf{d}$ are the number of data points in $\mathbf{s}$ and in $\mathbf{d}$, respectively.

Our decision to describe the noise in equations 24 and 25 as multivariate Gaussian should not be ad-hoc, but rather based on prior observations of the noise in the system. If the true noise departs significantly from the multi-Gaussian distribution, the results of our estimation will suffer by making the multi-Gaussian assumption. It can be shown via the Law of Large Numbers (Papoulis, 1991) that linear combinations of a large number of independent random variables, regardless of their distribution, converge to a Gaussian distribution (assuming they have finite variances). Thus, in addition to being tractable, the Gaussian noise assumption is a plausible one in the absence of other information. In this work we did not have prior knowledge of the distribution of the noise vectors and thus fell back to the multi-Gaussian white noise hypothesis.

6.1 Stochastic inversion approach

The probability density functions of the noise terms in equations 24 and 25 can be equally well represented as likelihood functions of the data sets conditioned on $\mathbf{f}$:

$$Pr(\mathbf{n_H}) = Lk(\mathbf{s}|\mathbf{f}) \quad (26)$$

$$Pr(\mathbf{n_P}) = Lk(\mathbf{d}|\mathbf{f}). \quad (27)$$

These likelihood functions are only probability densities for a fixed vector $\mathbf{f}$ (Tarantola, 1987, p. 34); otherwise they are not normalizable. For an observed $\mathbf{s}$ and $\mathbf{d}$, the optimal reflectivity field $\mathbf{f}$ is found by simultaneously maximizing both likelihood functions. Maximizing the multi-Gaussian likelihood functions corresponds to minimizing the negative exponents of equations 24 and 25. Putting these exponents together we have

$$\epsilon(\mathbf{f}|\mathbf{s},\mathbf{d}) = \frac{1}{2}\left(\frac{1}{\sigma_{\mathbf{H}}^2}(\mathbf{s}-\mathbf{Hf})^T(\mathbf{s}-\mathbf{Hf}) + \frac{1}{\sigma_{\mathbf{P}}^2}(\mathbf{d}-\mathbf{Pf})^T(\mathbf{d}-\mathbf{Pf})\right). \quad (28)$$

Minimizing this equation over $\mathbf{f}$ will perform no interpolation and may not lead to a unique inverse unless the regularization constraint is imposed. We do this by additionally minimizing $\|\frac{1}{\sigma_{\mathbf{L}}}\mathbf{Lf}\|$. This leads to the augmented equation

$$\epsilon(\mathbf{f}|\mathbf{s},\mathbf{d}) = \frac{1}{2}\left(\frac{1}{\sigma_{\mathbf{H}}^2}(\mathbf{s}-\mathbf{Hf})^T(\mathbf{s}-\mathbf{Hf}) + \frac{1}{\sigma_{\mathbf{P}}^2}(\mathbf{d}-\mathbf{Pf})^T(\mathbf{d}-\mathbf{Pf}) + \frac{1}{\sigma_{\mathbf{L}}^2}\mathbf{f}^T\mathbf{L}^T\mathbf{Lf}\right). \quad (29)$$

To find the minimum we take the derivative of equation 29 with respect to $\mathbf{f}$ and set the result equal to zero:

$$\frac{\partial\epsilon(\mathbf{f}|\mathbf{s},\mathbf{d})}{\partial\mathbf{f}} = (\frac{1}{\sigma_{\mathbf{H}}^2}\mathbf{H}^T\mathbf{H} + \frac{1}{\sigma_{\mathbf{P}}^2}\mathbf{P}^T\mathbf{P} + \frac{1}{\sigma_{\mathbf{L}}^2}\mathbf{L}^T\mathbf{L})\mathbf{f} - \frac{1}{\sigma_{\mathbf{H}}^2}\mathbf{H}^T\mathbf{s} - \frac{1}{\sigma_{\mathbf{P}}^2}\mathbf{P}^T\mathbf{d} = 0. \quad (30)$$

Rearranging, we have

$$(\frac{1}{\sigma_{\mathbf{H}}^2}\mathbf{H}^T\mathbf{H} + \frac{1}{\sigma_{\mathbf{P}}^2}\mathbf{P}^T\mathbf{P} + \frac{1}{\sigma_{\mathbf{L}}^2}\mathbf{L}^T\mathbf{L})\mathbf{f} = \frac{1}{\sigma_{\mathbf{H}}^2}\mathbf{H}^T\mathbf{s} + \frac{1}{\sigma_{\mathbf{P}}^2}\mathbf{P}^T\mathbf{d}. \quad (31)$$

$\sigma_{\mathbf{H}}^2$ expresses the degree of uncertainty in the seismic data and $\sigma_{\mathbf{P}}^2$ expresses the degree of uncertainty in the observed well data. $\sigma_{\mathbf{L}}^2$ can be interpreted as saying how much we believe the wells should tie into the rest of the reflectivity field. Put another way, it stresses how much horizontal correlation (smoothing) the resulting reflectivity field should exhibit. Different combinations of these parameter values give drastically different results to the joint inverse problem. Among the possible solutions, traditional interpolation and damped trace-by-trace deconvolution are special cases.

6.2 Bayesian approach

If we define a prior probability distribution on $\mathbf{f}$ the solution of an inverse problem involves maximizing the *posterior* probability density function (Tarantola, 1987):

$$Pr(\mathbf{f}|\mathbf{s},\mathbf{d}) = k\ Lk(\mathbf{s}|\mathbf{f})Lk(\mathbf{d}|\mathbf{f})Pr(\mathbf{f}) \quad (32)$$

where k is a normalizing constant.

If we further assign $\mathbf{f}$ the multi-Gaussian distribution,

$$Pr(\mathbf{f}) = \frac{1}{(2\pi^{\frac{N_\mathbf{f}}{2}}\ det((\frac{1}{\sigma_{\mathbf{L}}^2}\mathbf{L}^T\mathbf{L})^{-1}))}\ exp[-\frac{1}{2\sigma_{\mathbf{L}}^2}(\mathbf{f}^T\mathbf{L}^T\mathbf{Lf})], \quad (33)$$

the posterior takes the form

$$Pr(\mathbf{f}|\mathbf{s},\mathbf{d}) = k\ exp\left[-\epsilon(\mathbf{f}|\mathbf{s},\mathbf{d})\right]. \quad (34)$$

The negative exponent is equation 29. Maximizing this posterior distribution corresponds to minimizing the negative exponent and we arrive again at equation 31. The equations we solve in Bayesian inversion with a Gaussian prior are the same as those for stochastic inversion; the difference is mainly philosophical. If $\sigma_{\mathbf{L}}^2$ is an adjustable parameter that can be changed to improve the results of the inversion, we are in the stochastic inversion framework. If it is decided by our prior knowledge, we are in the Bayesian inversion framework.

7 Computational Issues and the Conjugate Gradients Algorithm

Let us pose the estimation of $\mathbf{f}$ as a matrix inversion by setting $\mathbf{Af} = \mathbf{b}$, where $\mathbf{A} = (\frac{1}{\sigma_{\mathbf{H}}^2}\mathbf{H}^T\mathbf{H} + \frac{1}{\sigma_{\mathbf{P}}^2}\mathbf{P}^T\mathbf{P} + \frac{1}{\sigma_{\mathbf{L}}^2}\mathbf{L}^T\mathbf{L})$ and $\mathbf{b} = \frac{1}{\sigma_{\mathbf{H}}^2}\mathbf{H}^T\mathbf{s} + \frac{1}{\sigma_{\mathbf{P}}^2}\mathbf{P}^T\mathbf{d}$. Then we must numerically perform the equivalent of $\mathbf{f} = \mathbf{A}^{-1}\mathbf{b}$ to obtain our solution. There are many ways to perform this solution (Strang, 1986; Trefethen and Bau, 1997; Press et al., 1995). However, the sheer size of $\mathbf{A}$ for a problem of realistic dimensions rules out all but a few methods.

To demonstrate the computational cost we examine our 3-D data set. $\mathbf{f}$ represents a 3-D cube of reflectivity traces with dimensions $201 \times 134 \times 125$. $\mathbf{f}$ is stored as a 1-D array in a trace by trace fashion. Thus we put the $134 \times 125 = 16,750$ reflectivity traces tip-to-tail into $\mathbf{f}$ to make a $3,366,750$ entry long vector that takes $26,934,000$ bytes to store in double precision. The 3-D cube of seismic data is of the same dimensions as the reflectivity. The $\mathbf{A}$ matrix for a problem of this size would then have $3,366,750 \times 3,366,750 = 1.13350055625 \times 10^{13}$ entries and, at 8 bytes per entry, we would require $9.06800445 \times 10^{13}$ bytes of computer RAM just to store the matrix! Most modern computers don't have 90 Terabytes of RAM so direct creation and inversion of this hypothetical matrix is impossible. There is, however, an excellent property of this matrix that we should and do exploit. $\mathbf{A}$ is exceedingly sparse. Most of its entries are 0. This prompts us to turn to iterative methods of solving a system of linear equations. The method we use is conjugate gradients (Strang, 1986; Trefethen and Bau, 1997; Press et al., 1995).

Iterative methods require the repeated application of the $\mathbf{A}$ matrix. We show how this can be quickly executed by looking individually at each component of $\mathbf{A}$. The first component of $\mathbf{A}$ is $\mathbf{H}^T\mathbf{H}$. This has the effect of going through each reflectivity trace and convolving it with the autocorrelation function of the seismic wavelet. In looking at $\mathbf{H}^T\mathbf{H}$ as a matrix, it is a block diagonal matrix with each diagonal block an identical symmetric Toeplitz matrix. Since each diagonal block is identical, we need to store only one of them, $\check{\mathbf{H}}^T\check{\mathbf{H}}$. However, since convolution is done more efficiently in the Fourier domain for long seismic wavelets, we simply store the Fourier transform of the autocorrelation function (i.e. the power spectrum) in memory. We have thus compressed a matrix with $1.13350055625 \times 10^{13}$ entries into a single vector with only 201 entries! When operating $\mathbf{H}^T\mathbf{H}$ on $\mathbf{f}$ we need only cycle through $\mathbf{f}$ each 201 entries at a time, transform those 201 values to the Fourier domain, multiply by the power spectrum of $\check{\mathbf{H}}^T\check{\mathbf{H}}$, and then transform back. That becomes approximately $O(16,750 \times 201 \times \log_2 201)$ operations for one application of $\mathbf{H}^T\mathbf{H}$.

As mentioned above, we model $\mathbf{L}$ as a multi-dimensional differencing operator. Storage of $\mathbf{L} = \begin{bmatrix} \mathbf{L}_X \\ \mathbf{L}_Y \end{bmatrix}$ on a computer amounts to storing the indices and values of the differencing operator (i.e. the non-zero elements of one row of both $\mathbf{L}_X$ and $\mathbf{L}_Y$). For low order differencing operators, this amounts to just a few entries in two vectors. The action of $\mathbf{L}_X\mathbf{f}$ and $\mathbf{L}_Y\mathbf{f}$ then amounts to two explicit convolutions. The $\mathbf{L}^T$ operator is then applied by reversing the order of the values in the vector of the $\mathbf{L}_X$ and $\mathbf{L}_Y$ coefficients and applying another convolution to the output of $\mathbf{Lf}$. For our purposes $\mathbf{L}^T\mathbf{Lf}$ amounts to $O(2 \times 2 \times \nu \times 3,366,750)$ operations, where ν is the number of elements in each differencing operator (in our case $\nu = 2$).

Finally we come to $\mathbf{P}^T\mathbf{P}$, the simplest of the operators to apply. This operator has the effect of extracting reflectivity profiles at a few locations and adding them to the output of $\mathbf{H}^T\mathbf{H}$ and $\mathbf{L}^T\mathbf{L}$ at those locations. This can be done explicitly and requires no computation.

The conjugate gradients algorithm proceeds by generating a sequence of vectors $\mathbf{f}_k$ that converges to the true solution of $\mathbf{f} = \mathbf{A}^{-1}\mathbf{b}$. The matrix $\mathbf{A}^{-1}$ is never constructed, neither is Gaussian elimination or any other standard matrix inversion technique employed (Press et al., 1995; Strang, 1986; Golub and Loan, 1989; Trefethen and Bau, 1997). The algorithm instead proceeds by operating the matrix iteratively and therefore requires the fast forward operation of the matrix $\mathbf{A}$. The fast operation of $\mathbf{H}^T\mathbf{H}$, $\mathbf{L}^T\mathbf{L}$, and $\mathbf{P}^T\mathbf{P}$, as described above, show that the conjugate gradients algorithm is ideal for this particular $\mathbf{A}$ matrix.

In the absence of numerical round-off error, the conjugate gradients algorithm will reach the minimizing solution in a finite number of iterations. It will still give a good solution if stopped short of this final number. In fact, as the algorithm runs, smaller and smaller details are added to the solution. This has the effect that, if stopped too early, the solution will be effectively "regularized". The effect of this sort of regularization and that produced by the regularizing matrix, $\mathbf{L}^T\mathbf{L}$, might be hard to distinguish. Therefore, after obtaining a satisfactory solution one should run the algorithm again with more iterations to make sure that the solution does not change significantly, and that the smoothness in the inversion result is not just an artifact of the

algorithm. For more details on the algorithm see Appendix A.

8 Inversion of Field Data and Discussion

We now examine the results of the joint inversion on the field data. Inversion is done for 3 sets of σ^2 parameters. The combination of parameters are shown in table 1 and illustrate 3 important special cases of the inversion: interpolation (with no use of seismic data), joint inversion (with optimal use of all data), and damped trace-by-trace deconvolution (which makes little use of well data). We see that a different $\sigma^2_{\mathbf{L}}$ is used

Inversion Type	$\sigma^2_{\mathbf{H}}$	$\sigma^2_{\mathbf{P}}$	$\sigma^2_{\mathbf{L}}$ in 2-D	$\sigma^2_{\mathbf{L}}$ in 3-D
Interpolation Only	∞	7.65×10^{-7}	0.005	0.005
Joint Inversion	1.05×10^{6}	7.65×10^{-7}	0.005	0.15
Damped Trace-By-Trace Decon	1.05×10^{6}	10000	0.1	100

Table 1: Parameter combinations for 3 different inversion schemes.

in 2-D and 3-D. The nature of the problem changes from 2-D to 3-D: in 2-D we only have L_X operating, whereas in 3-D both L_X and L_Y operate. One is essentially solving a different inverse problem in 3-D than in 2-D.

Using the well data shown in Figures 6 and 7, we perform interpolation with the parameters from table 1. We use a correlation length of $a = 250$ in both 2-D and 3-D. This correlation length is used in all results except for the trace-by-trace deconvolution below. The interpolation results are shown in Figure 13 and 14 for 2-D and 3-D, respectively. From these figures we can see that the well data has been extrapolated a large distance because of the correlation length input with the $\mathbf{L}$ operator. Interpolation can tell us nothing, however, about reflectivity at locations above and below the wells. It also cannot account for any variations in reflectivity at inter-well locations.

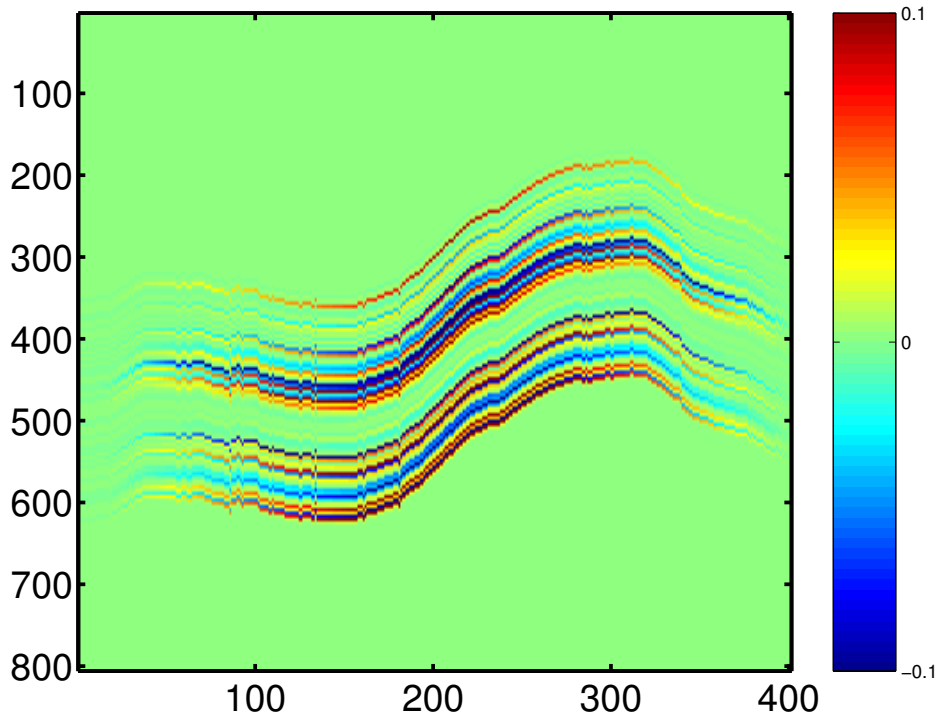

Figure 13: 2-D interpolated reflectivity field.

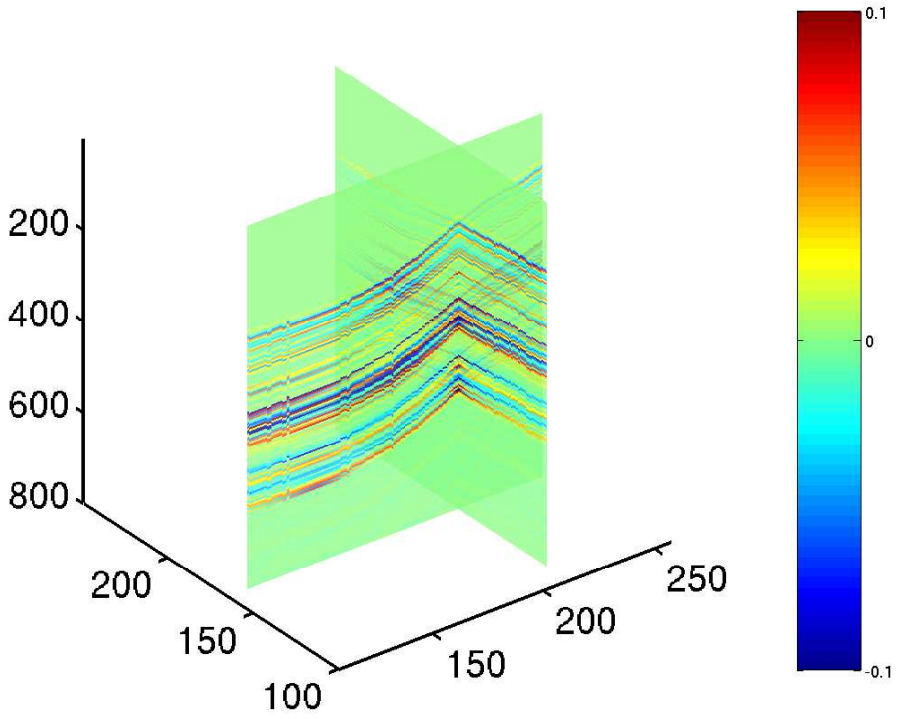

Figure 14: 3-D interpolated reflectivity field.

Trace-by-trace deconvolution makes little use of the well data. It does not force the result of the inversion to fit well data at well locations because $\sigma_{\mathbf{P}}^2$ is very large. Nor does it enforce much horizontal correlation because the correlation length is set to ≈ 1. Thus it essentially only performs a damped deconvolution at each seismic trace location. The results of such an inversion are shown in Figure 15 and 16 for 2-D and 3-D, respectively. The lateral continuity is as bad as the original seismic but vertically the resolution has been improved because of deconvolution in that direction. We next combine the global resolution improvement of trace-by-trace damped deconvolution with the well tying and horizontal correlation of the interpolation by performing an optimal joint inversion.

In Figures 17 and 18 we see the results of the joint inversion in 2-D and 3-D, respectively. Away from well locations we see that the inversion produces a laterally consistent deconvolution; it sharpens the image and shows greater lateral continuity than the original seismic. Near the wells the inversion makes use of the high resolution well logs and produces a high resolution image of reflectivity not attainable with damped deconvolution by itself. These results can greatly improve an interpreter's job of tying well data to the seismic data and tracing lateral events.

9 Conclusions

We have presented a method for simultaneously inverting seismic and well data for a single model of subsurface reflectivity. The maximum likelihood inverse formulation generated a system of equations to solve but the size of the problem precluded all but a few methods of solution. An iterative solution algorithm called the conjugate gradients method solved the problem.

By posing deconvolution and interpolation as a single joint inversion we can still arrive at either interpolation or trace-by-trace deconvolution as special cases, but incorporating both of the data sets gives a superior solution. The well data increases the resolution of the deconvolution near the wells, the regularization operator ensures lateral continuity, and the seismic data helps the interpolation away from the wells.

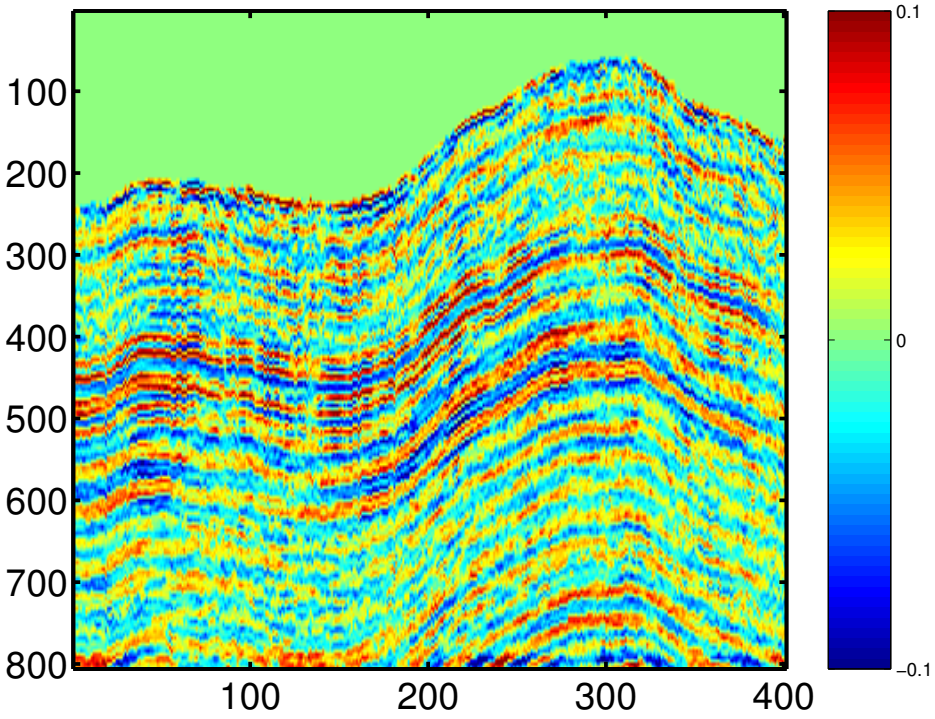

Figure 15: Trace-by-trace damped deconvolution of 2-D seismic data.

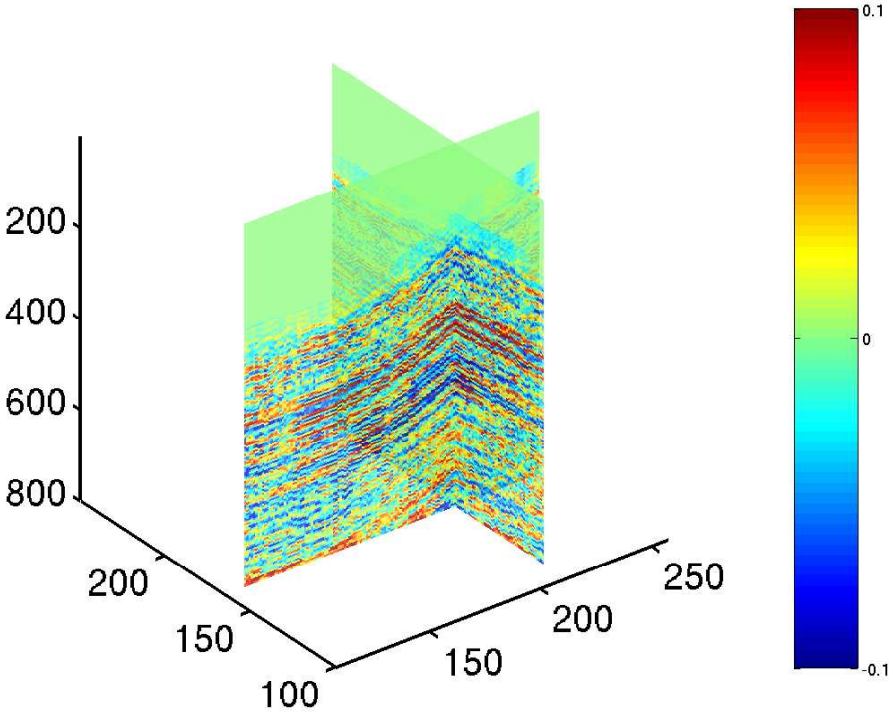

Figure 16: Trace-by-trace damped deconvolution of 3-D seismic data.

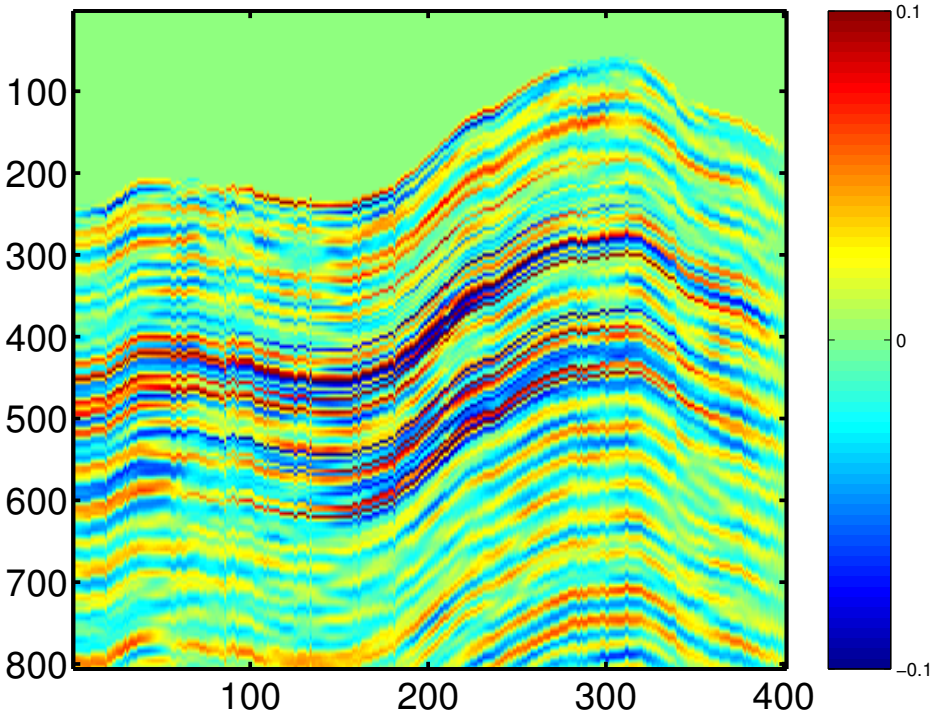

Figure 17: 2-D joint inversion of well and seismic data.

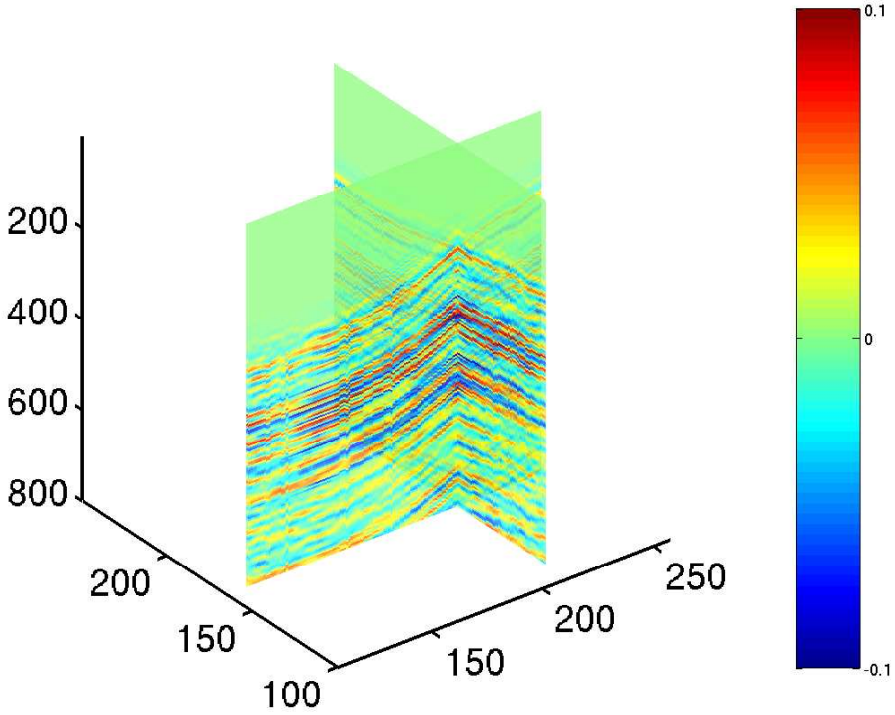

Figure 18: 3-D joint inversion of well and seismic data.

10 Acknowledgments

We would like to thank the ERL Founding Members for financial sponsorhip of this work. We especially thank Chevron Texaco for their fellowship supporting Jonathan Kane.

A The Conjugate Gradients Algorithm

The conjugate gradients method, like many iterative methods, constructs what is known as a *Krylov subspace* of vectors with the property such that at the k^{th} iteration the quantity $(\hat{\mathbf{f}}-\mathbf{f}_k)^T\mathbf{A}(\hat{\mathbf{f}}-\mathbf{f}_k)$ is minimum, where $\hat{\mathbf{f}}$ is the true solution to the equation $\mathbf{Af}=\mathbf{b}$.

Furthermore, the residual $\mathbf{r}_k=\mathbf{b}-\mathbf{Af}_k$ is orthogonal to all previous residuals: $\mathbf{r}_j^T\mathbf{r}_k^T=0$, $j<k$. Also, the search direction $\mathbf{d}_k$ is $\mathbf{A}$ orthogonal to all previous search directions: $\mathbf{d}_j^T\mathbf{Ad}_k^T=0$, $j<k$.

When applied to a positive definite matrix, the conjugate gradients algorithm can be viewed as minimizing the objective function $\Phi(\mathbf{f})=\frac{1}{2}\mathbf{f}^T\mathbf{Af}-\mathbf{f}^T\mathbf{b}$. Each iteration lowers the objective function and brings the vector $\mathbf{f}$ closer to the minimizing solution. This is one of the attractions of the method. It can be stopped any time before the final solution is obtained and still give a useful answer. This is not the case with other standard matrix inversion algorithms, such as Gaussian elimination. Another attraction has already been mentioned: the ability to invert a huge matrix (assuming it is sparse).

The algorithm is as follows (Trefethen and Bau, 1997):

1. Set $\mathbf{f}_0=$ arbitrary, and $\mathbf{r}_0=\mathbf{b}-\mathbf{Af}_0$
2. $\beta_j=\mathbf{r}_{j-1}^T\mathbf{r}_{j-1}/\mathbf{r}_{j-2}^T\mathbf{r}_{j-2}$ (except $\beta_1=0$)
3. $\mathbf{d}_j=\mathbf{r}_{j-1}+\beta_j\mathbf{d}_{j-1}$ (except $\mathbf{d}_1=\mathbf{r}_0$)
4. $\alpha_j=\mathbf{r}_{j-1}^T\mathbf{r}_{j-1}/\mathbf{d}_j^T\mathbf{Ad}_j$
5. $\mathbf{f}_j=\mathbf{f}_{j-1}-\alpha_j\mathbf{d}_j$
6. $\mathbf{r}_j=\mathbf{r}_{j-1}-\alpha_j\mathbf{Ad}_j$
7. Repeat step 2.

References

Deutsch, C. and Journel, A. (1998). *GSLIB: Geostatistical Software Library and User's Guide, 2nd ed.* Oxford University Press, Inc.

Fomel, S. (2001). *Three-dimensional Seismic Data Regularization*. PhD thesis, Stanford University.

Golub, G. H. and Loan, C. F. V. (1989). *Matrix Computations*. The Johns Hopkins University Press.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1995). *Numerical Recipes in C*. Cambridge University Press.

Robinson, E. A. and Trietel, S. (1980). *Geophysical Signal Analysis*. Prentice-Hall, Inc.

Rodi, W. (1989). *Regularization and Backus-Gilbert Estimation in Non-linear Inverse Problems: Application to Magnetotellurics and Surface Waves*. PhD thesis, The Pennsylvania State University.

Saggaf, M. M. and Robinson, E. A. (2000). A unified framework for the deconvolution of traces of non-white reflectivity. *Geophysics*.

Strang, G. (1986). *Introduction to Applied Mathematics*. Wellesley-Cambridge Press.

Tarantola, A. (1987). *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier.

Trefethen, L. N. and Bau, D. (1997). *Numerical Linear Algebra*. Society of Industrial and Applied Mathematics.

Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.

Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Wiley.

Yilmaz, Özdoğan. (1987). *Seismic Data Processing*. Society of Exploration Geophysicists.