Uncertainty estimation and error analysis for linear inversion problems

Kasper van Wijk[†], John A. Scales[†] and William Navidi[††]

[†] Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401 USA.
[††] Department of Mathematics, Colorado School of Mines, Golden, CO 80401 USA.

Abstract

Inverse theory concerns the problem of making inferences about physical systems from measurements. Since the measurements are invariably subject to some uncertainty, to solve an inverse problem it is necessary to have information about the errors in the observations, otherwise it is impossible to say when a feature "fits the data." In practice one seldom has a direct estimate of the data errors, so in the absence of repeated measurements, it is important to be able to distinguish systematic data variations from noise fluctuations.

To estimate the noise we use the L-curve method to construct a reference model of the system. The fluctuations of the data predicted for this reference model about the observations represents an initial estimate of the noise. We then use the resulting noise variance estimate to determine an optimally truncated singular value decomposition (OTSVD) and to construct confidence intervals on the OTSVD model for a VSP problem.

Introduction

The goal of geophysical inversion is to make quantitative inferences about the Earth from a finite number of uncertain observations. It is necessary to know something about the uncertainty in the data in order to define which models are consistent with the observations. There are many contributors to the uncertainty in the data, which can be divided into systematic and random components. If the relation between model parameters and observations is linear, we can write the discretized forward problem as

$$\mathbf{d} = \mathbf{A}(\mathbf{x}) + \mathbf{e} + \mathbf{s}, \tag{1}$$

where the observations are $\mathbf{d} \in R^n$, model parameters $\mathbf{x} \in R^m$ and discrete forward operator $\mathbf{A} \in R^{n \times m}$. The systematic errors $\mathbf{s}$ consist primarily of un-modeled physics and the effects of model discretization. The random errors $\mathbf{e}$ are, by definition, those variations in the data that are not deterministically reproducible. From here on we assume that all physics in our problem is modeled and that the discretization level is fine enough to make the discretization errors insignificantly small [van Wijk et al., 2001].

Estimating the data variance

If there were multiple realizations of the experiment, the distribution of random fluctuations could be directly quantified. Unfortunately, in geophysics there is usually only one realization of the data acquisition. If the data are sufficiently redundant they could be binned so as to approximate the situation of multiple realizations. If the (binned) data are not truly redundant it is necessary to model the systematic variations in the data. A preliminary model could provide synthetic data to be subtracted from the observations. If the preliminary model predicts the systematic variations well, the residuals represent the random fluctuations. We will obtain a preliminary model via Tikhonov optimization, where the L-curve method determined the regularization parameter. The L-curve trades off data prediction vs. model structure [Hansen, 1998], which knows nothing about the errors in the data. The residuals between the data predicted by Tikhonov optimization and the observations can then be used to estimate the noise variance in the data:

$$\tilde{\sigma}^2 = \mathrm{var}\left(\mathbf{A}\tilde{\mathbf{x}} - \mathbf{d}\right). \tag{2}$$

Inversion

If the data variance is known, a traditional approach is to compute the χ^2 function. We use the estimate $\tilde{\sigma}$ to do so:

$$\chi^2(k) = \frac{1}{n}\frac{\|\mathbf{A}\mathbf{x}_k - \mathbf{d}\|^2}{\tilde{\sigma}^2}, \tag{3}$$

where $\mathbf{x}_k$ is the model obtained via truncated SVD, with k singular values:

$$\mathbf{x}_k = \sum_{m=1}^{k} \frac{\mathbf{u}_m^T \mathbf{d}}{\lambda_m} \mathbf{v}_m, \tag{4}$$

where $\mathbf{u}_m$ is the m-th data space singular vector, $\mathbf{v}_m$ is the m-th model space singular vector and λ_m the m-th singular value.

If the estimate of the data variance is correct, a model with $\chi^2 = 1$ fits the data on average to one standard deviation. However, the χ^2 curve can be such that a lot of structure needs to be added to the model to improve the data fit only slightly. We therefore propose an AIC criterion [T. Sakamoto and Kitagawa, 1986] to penalize excessive model structure:

$$\text{AIC(k)} = \chi^2(k) * e^{a \cdot k/n}, \tag{5}$$

where the value of a determines the relative severity of the penalty on the number of singular values used.

We accept the level of truncation to be the smaller of the k's associated with the global minimum of the AIC and the $\chi^2 = 1$. The resulting smooth truncated SVD model is assumed to fit the signal in the data, so that the residuals can be used to update the error estimate. The algorithm now looks as follows.

Algorithm 1 OTSVD: Optimally Truncated Singular Value Decomposition

Let λ_L be the optimal Tikhonov regularization parameter as determined by the L curve method. Let $\tilde{\sigma}^2$ be the variance of the data residuals associated with the Tikhonov regularized using $\lambda = \lambda_L$.

1. *k_{aic} is the singular value associated with the minimum of the AIC.*
2. *Compute the TSVD model $\mathbf{x}_k$ and $\chi^2(\mathbf{x}_k)$ for values of k up to some upper limit $k_{\max}$.*
3. *k_χ is the smallest k such that $\chi^2(\mathbf{x}_k) < 1$.*
4. *The optimal number of singular values is $k_o = min(k_{\text{aic}}, k_\chi)$.*
5. *Compute the TSVD model with $k = k_o$: $\mathbf{x}_o = \mathbf{A}_o^\dagger \mathbf{d}$.*
6. *Update the noise variance estimate: $\tilde{\sigma}^2 = \|\mathbf{A}\mathbf{x}_o - \mathbf{d}\|^2$.*

We solve the inversion problem by estimating confidence intervals on the model parameters. A cartoon of the sequence of steps is depicted in figure 1.

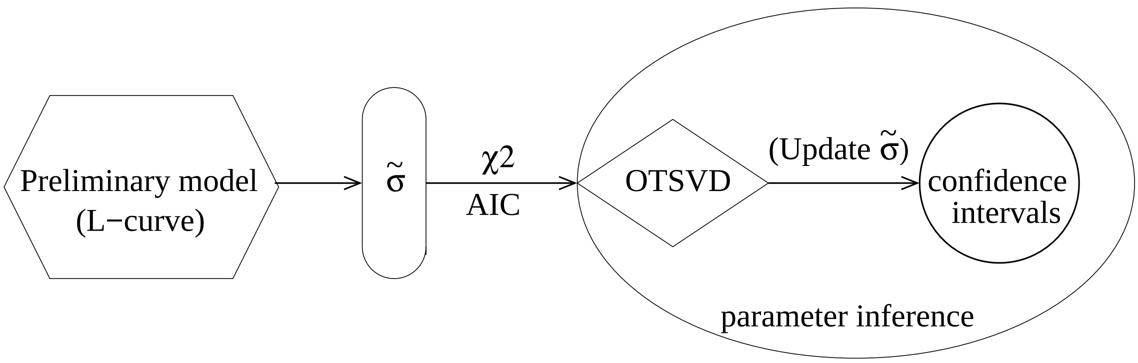

Figure 1: Cartoon of OTSVD algorithm with confidence estimation

Vertical Seismic Profile

In travel time tomography rays are traced through the sub-surface. The travel time t of a single ray is the integral of the local *slowness*, i.e. the reciprocal of velocity v, along the ray:

$$t = \int_{ray} \frac{1}{v(l)} dl. \tag{6}$$

We will investigate the discretized one-dimensional tomography problem of Vertical Seismic Profiling (VSP). The geometry is shown in Figure 2. A single source is on the surface directly above a number of receivers that are lowered into a bore-hole. The source emits a pulse of energy and the arrival times to n receivers form the data vector $\mathbf{d}$. A model vector $\mathbf{x}$ consists of m layers of constant thickness and slowness. In practice, we over-parameterize the model space [van Wijk et al., 2001]. The discrete forward operator is a matrix $\mathbf{A}$ of dimensions $n \times m$, whose element A_{ij} is the length of the i-th ray in the j-th layer. The The data are contaminated with uncorrelated noise drawn from a pseudo-random Gaussian distribution with $\mu = 0$ and $\sigma = 1$ ms, so that linear system is that as defined in equation (1).

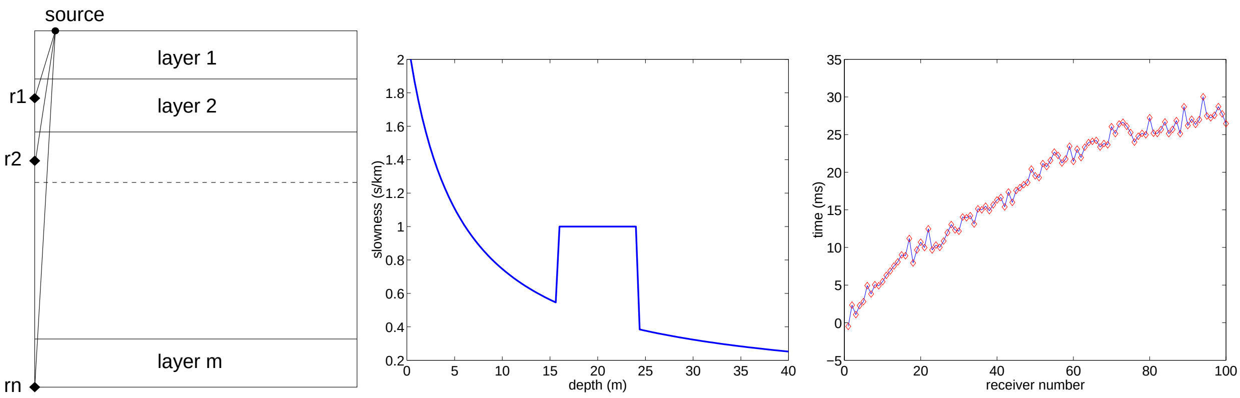

Figure 2: The synthetic VSP configuration, the true slowness model and the synthetic data.

In Figure 3 we see a typical L-curve for this example. The optimal value of the Tikhonov regularization parameter is defined to be the point of maximum curvature. With this value of λ, the Tikhonov optimization leads to a preliminary model $\tilde{\mathbf{x}}$, which will then serve to determine an estimate of the data variance $\tilde{\sigma}^2$ as defined in (2). The data uncertainty estimate is within 4 percent of the true error, in this case. Note that this success is solely the result of a difference in frequency between noise and signal. We will not accept the Tikhonov model, because it has more structure than needed to fit the data.

With this estimate $\tilde{\sigma}$, OTSVD (algorithm 1) is initiated. The AIC curve is plotted in the middle of figure 3. On the right side of this figure, you can see the data predicted by the OTSVD model and the Tikhonov model, regularized by the L-curve method. The data predictions are in agreement, but the OTSVD data is obtained with a model with much less structure (see the left side of Figure 4).

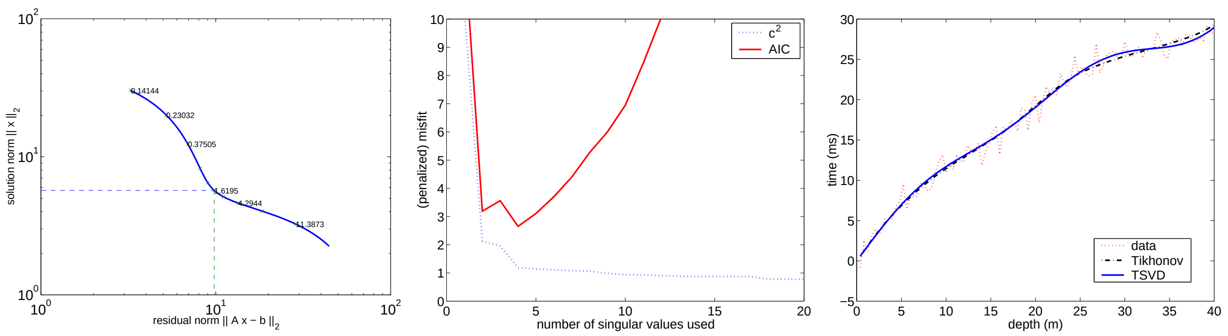

Figure 3: Typical L-curve, the AIC curve penalizing the number of singular values used (here $a = 20$ in the exponent), and the predicted and observed travel times.

Posterior uncertainty analysis

In Figure 4, the smooth OTSVD model centers error bars consisting of the estimated standard deviation. These error bars tell us how much the model parameters vary as a result of the random fluctuations in the data. By definition, σ is the square root of the diagonal elements of the covariance matrix:

$$\mathrm{cov}(\mathbf{x}_o) = E[\mathbf{x}_o(\mathbf{x}_o)^T] = E[\mathbf{A}_o^\dagger \mathbf{d}(\mathbf{A}_o^\dagger \mathbf{d})^T] = \mathbf{A}_o^\dagger \mathrm{cov}(\mathbf{d}) \mathbf{A}_o^{\dagger T}. \tag{7}$$

If the OTSVD solution was un-biased, these bars would be the final uncertainty estimate. It is usually impossible to compute the bias of the model parameters since it involves knowing the true model:

$$\text{Bias}(\mathbf{x}_o) = E[\mathbf{x}_o - \mathbf{x}_{\text{true}}] = E[(\mathbf{A}_o^\dagger \mathbf{A} - \mathbf{I})\mathbf{x}_{\text{true}} + \mathbf{A}_o^\dagger \mathbf{e}] = (\mathbf{A}_o^\dagger \mathbf{A} - \mathbf{I})\mathbf{x}_{\text{true}} = \mathbf{B}\mathbf{x}_{\text{true}}. \tag{8}$$

A conservative alternative would be to compute bounds on the bias with prior bounds on the solution. However, since we over-parameterized, posterior bounds on the bias will be no smaller than the prior ones. Extra conditions like smoothness constraints could possibly lead to smaller confidence intervals.

To show the implications of - and relations between - covariance and bias, we have calculated the bias in this synthetic example, plotted in Figure 4. From equations (7) and (8) it is clear that there is a trade off between the covariance and the bias. Increasing the level of truncation results in smooth solutions that vary less with random fluctuations in the data, causing artificially small variance estimates, while the bias increases.

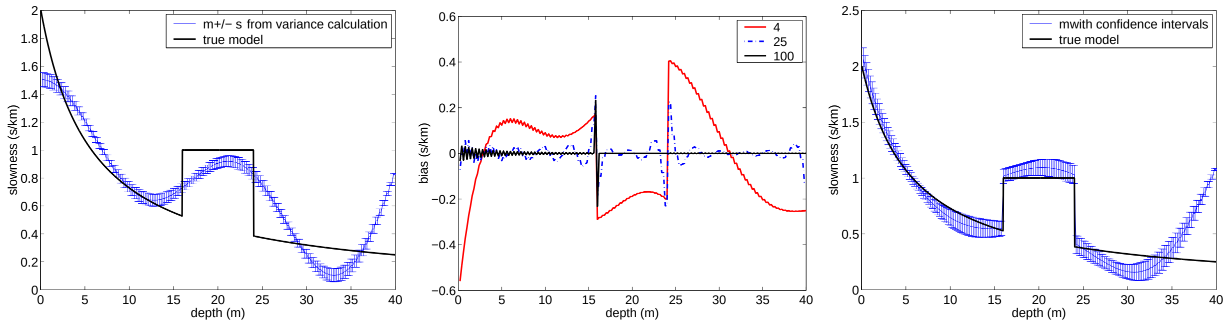

Figure 4: The mean result of the OTSVD model $\pm\sigma$, the bias in the models for different levels of truncation, and the un-biased OTSVD solution.

Conclusions

When provided with nothing but a single realization of noisy data, it is impossible to separate signal from noise: at some point choices based on prior information need to be made. In the OTSVD algorithm we propose, the initial estimate of the data variance came from a method (L-curve) that knows nothing about the noise level in the data. Therefore, there are no guarantees that the L-curve predicts the data variance well. In this example it was very successful, but tests on 2-d tomography led to failure of the L-curve method [van Wijk et al., 2001]. In addition, the size of the exponent in AIC is a subjective choice, that will vary from case to case.

Simple interval constraints on parameters are insufficient to make the posterior uncertainty estimates smaller than the prior, because of the nature of the null space vectors. It is misleading to look at the smooth models produced by the OTSVD or Tikhonov regularization and see how closely they resemble the true model; the true model is smooth and we are using methods that produce smooth estimated models. However, the fact is that with only interval constraints on the slowness parameters, quite pathological models can be constructed that fit the data. This is reflected in posterior uncertainties that are the same size as the prior uncertainties, even though the estimated model looks quite reasonable.

References

[Hansen, 1998] Hansen, P. C. (1998). *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM monographs on mathematical modeling and computation.

[T. Sakamoto and Kitagawa, 1986] T. Sakamoto, M. I. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. D. Reidel, Holland.

[van Wijk et al., 2001] van Wijk, K., Scales, J. A., and Navidi, W. (2001). Optimally truncated singular value decomposition for linear inverse problems. pre-print.