Stanford Exploration Project, Report 112, November 11, 2002, pages 161–171

Short Note

Conjugate gradient total least-squares in geophysical optimization problems

Morgan Brown[1]

INTRODUCTION

Total least-squares (TLS) optimization is a methodology to solve least-squares optimization problems when the modeling operator has errors. In standard least-squares optimization, errors are assumed to be concentrated in the data only.

Golub and Loan (1980) presented a numerically-stable TLS algorithm which utilizes the singular value decomposition (SVD). Subsequent refinements to the method predominantly use SVD, and much of the current literature emphasizes stabilization of the inverse and implicit model regularization by SVD truncation (Fierro et al., 1997). Because it is numerically intensive, however, the SVD generally proves unrealistic for use in large-scale problems, which are the rule in exploration geophysics.

The TLS problem can be cast as an extremal eigenvalue/eigenvector estimation problem. Chen et al. (1986) present a conjugate gradient (CG) scheme to compute the minimum eigenvalue/eigenvector of a linear system. Zhu et al. (1997) extend Chen et al.'s algorithm to solve the TLS problem, in the context of optical tomography.

I begin with a short theoretical overview of the TLS problem. I implement the CG method described by Chen et al. (1986), adapted for the TLS problem in a similar fashion as the work of Zhu et al. (1997). I test the algorithm on two familiar geophysical problems: least-squares deconvolution of a 1-D signal, and velocity scan inversion with the hyperbolic Radon transform. Liu and Sacchi (2002) tested an SVD-based, regularized TLS approach on velocity scan inversion using the parabolic Radon transform.

TLS OVERVIEW

Golub and Loan (1980) phrased the TLS problem as follows. Given a forward modeling operator **L** and measured data **d**, assume that both are contaminated with white noise of uniform

[1]**email:** morgan@sep.stanford.edu

variance; matrix $\mathbf{N}$ and vector $\mathbf{n}$, respectively. Then the TLS solution is obtained by minimizing the Frobenius matrix norm of the augmented noise matrix:

$$\min \|[\mathbf{N} \;\; \mathbf{n}]\|_{\mathbf{F}}, \tag{1}$$

subject to the constraint that the solution is in the nullspace of the combined augmented noise and input operators:

$$([\mathbf{L} \;\; \mathbf{d}] + [\mathbf{N} \;\; \mathbf{n}]) \begin{bmatrix} \mathbf{m} \\ -1 \end{bmatrix} = \mathbf{0}. \tag{2}$$

To solve the system of equations (1) and (2), Golub and Loan (1980) introduced a technique based on the Singular Value Decomposition (SVD). Although mathematically elegant, SVD-based approaches are generally unrealistic for the large-scale problems that are the norm in exploration geophysics.

Equivalence with Rayleigh Quotient Minimization

Golub (1973) showed that the constrained minimization problem of equations (1) and (2) is equivalent to minimization of the so-called Rayleigh Quotient. If we define the vector $\mathbf{q} = [\mathbf{m} \;\; -1]^T$ and $\mathbf{A} = [\mathbf{L} \;\; \mathbf{d}]$, the Rayleigh Quotient takes the following form:

$$\min \;\; F(\mathbf{q}) = \left| \frac{\mathbf{q}\mathbf{A}^T\mathbf{A}\mathbf{q}}{\mathbf{q}^T\mathbf{q}} \right|_2 . \tag{3}$$

After the minimization of equation (3), the resultant vector $\mathbf{q}$ is the eigenvector associated with the smallest eigenvalue of $\mathbf{A}^T\mathbf{A}$.

Conjugate Gradient Method for TLS

The Rayleigh Quotient can be minimized by iterative techniques. Zhu et al. (1997) introduced a method based on conjugate gradients (CG) to solve the TLS problem which was adapted from the earlier work of Chen et al. (1986). I implemented this CG-based algorithm and present pseudocode in Appendix A.

Theory guarantees that the CG method converges in n steps, where n is the size of the model vector. However, in practical situations with real seismic data, a "useful" model may appear after relatively few ($<< n$) CG iterations. How useful the model and how few the iterations depends on the problem. Nonetheless, in practice, the computational cost and memory requirements are nearly always much less with CG than with SVD.

Relation of TLS to Damped Least-squares (DLS)

The TLS solution is closely related to the classic damped least squares (DLS) solution, where the damping factor, σ^2, is the smallest nonzero singular value of the augmented matrix $[\mathbf{L} \;\; \mathbf{d}]$:

$$\mathbf{m}_{DLS} = \left(\mathbf{L}^T\mathbf{L} + \sigma^2\mathbf{I}\right)^{-1} \mathbf{L}^T\mathbf{d}. \tag{4}$$

The TLS solution can be rewritten (Golub and Loan, 1980; Björck, 1996) as follows.

$$\mathbf{m}_{TLS} = \left(\mathbf{L}^T\mathbf{L} - \sigma^2\mathbf{I}\right)^{-1}\mathbf{L}^T\mathbf{d}. \tag{5}$$

The only difference between equations (4) and (5) is the negative sign on the damping term. Thus the TLS problem is considered a "deregularization" of the standard LS problem, and is guaranteed to be worse conditioned, since $\mathbf{L}^T\mathbf{L}$ is positive-semidefinite at worst (Björck, 1996).

LEAST-SQUARES DECONVOLUTION TESTS

I constructed a simple, yet relevant synthetic test case for the TLS algorithm: deconvolution. The known model is a sequence of spikes of random amplitude and placement. To create data, the known model was convolved with a Ricker wavelet. Gaussian-distributed noise with a variance of 1 was added to the data, and also to the filter used in the deconvolution.

Figures 1-3 compare the standard least-squares (LS), the TLS, and DLS solutions to the problem. The LS solution is undoubtedly poor. In the "quiet" zones of the model, where the known model is zero-valued, the estimated LS model has almost as much energy as where the spikes are. Still, the modeled data appears to fit the input data quite well.

The TLS and DLS solutions appear somewhat similar. Both approaches seem to suppress unwanted noise in the estimated model in the quiet regions. However, the TLS model seems to have better resolution of the true spikes. Also, the TLS method's residual error appears better balanced than the DLS's. Both TLS and DLS have higher residual error energy than the LS solution.

HYPERBOLIC RADON TRANSFORM TESTS

I tested the proposed TLS algorithm on a popular SEP inversion application, the Hyperbolic Radon Transform (HRT) (Nichols, 1994; Lumley et al., 1995; Guitton, 2000b). Figures 4 and 5 compare the results of the TLS, LS, and DLS methods, for 10 and 150 CG iterations, respectively.

The results of the HRT tests are inconclusive. After 10 iterations, the results from the three methods are almost indistinguishable. After 50, the DLS model looks "best," i.e., most interpretable by a human for picking velocities. However, the TLS residual error is the whitest, the best balanced, and contains no correlated energy–the very criteria which Guitton (2000a) uses to define optimality.

CONCLUSIONS AND DISCUSSION

I have introduced total least-squares (TLS) optimization as a possible alternative to "standard" least-squares approaches. TLS approaches incorporate errors in both the data and in the mod-

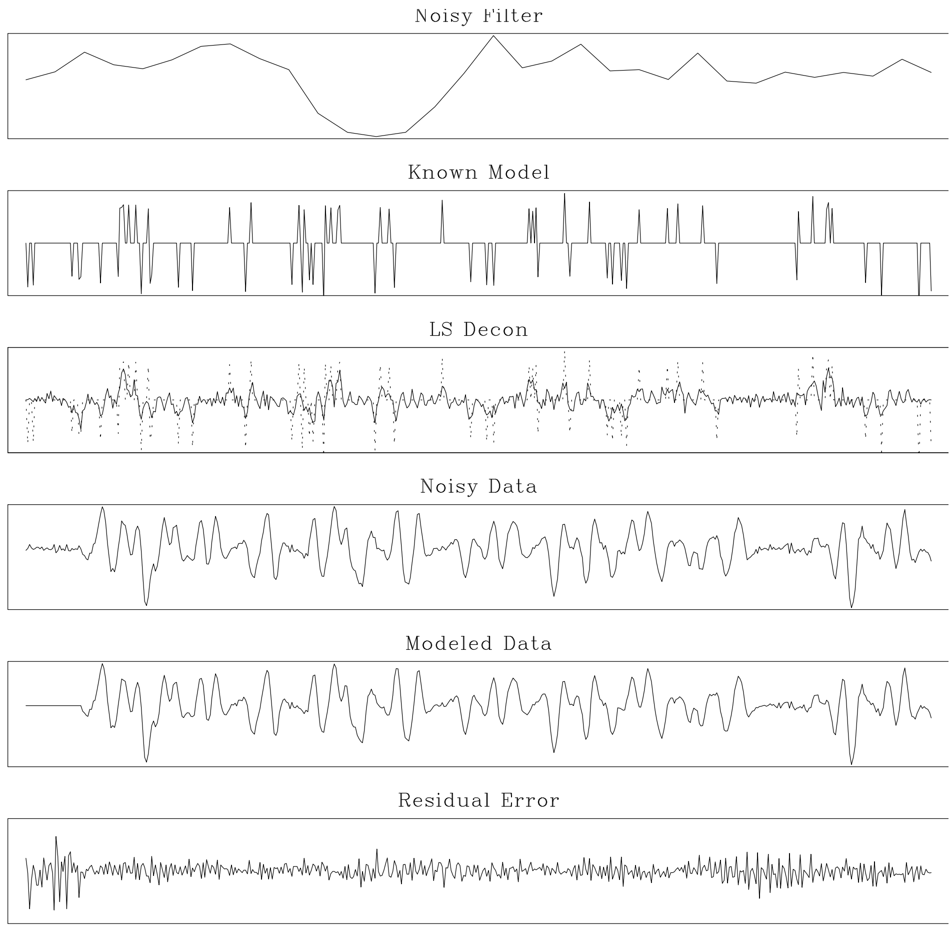

Figure 1: Top to bottom: 1) Known filter plus noise, 2) Known model, 3) Estimated standard least-squares model overlaying known model, 4) Noisy data, 5) Modeled data, 6) Residual error. morgan2-decon.ls.noisy [ER]

eling operator, to produce "more accurate" solutions. I put "more accurate" in quotes because in our real world, frequent appearences of nonempty nullspaces impose some subjectivity on any solution.

I implemented the conjugate gradient TLS solver (TLS-CG) published by Zhu et al. (1997), although in that paper, the authors omit a crucial model normalization step that leads to non-convergence of the algorithm. I present a complete algorithm in Appendix A.

Tests on a synthetic 1-D deconvolution example seem to validate TLS as a tool. In those tests, when ideal noise was added to the filter and data, TLS resolved the true model better than normal least-squares or damped least-squares. Tests using the hyperbolic radon transform were inconclusive; no efforts were made to understand operator error in this case, and in summary, the TLS result looks somewhere in between LS and DLS.

Will TLS be a useful tool in geophysics? My suspicion is that TLS makes only a second order improvement in the quest to account for uncertainty in geophysical inverse problems.

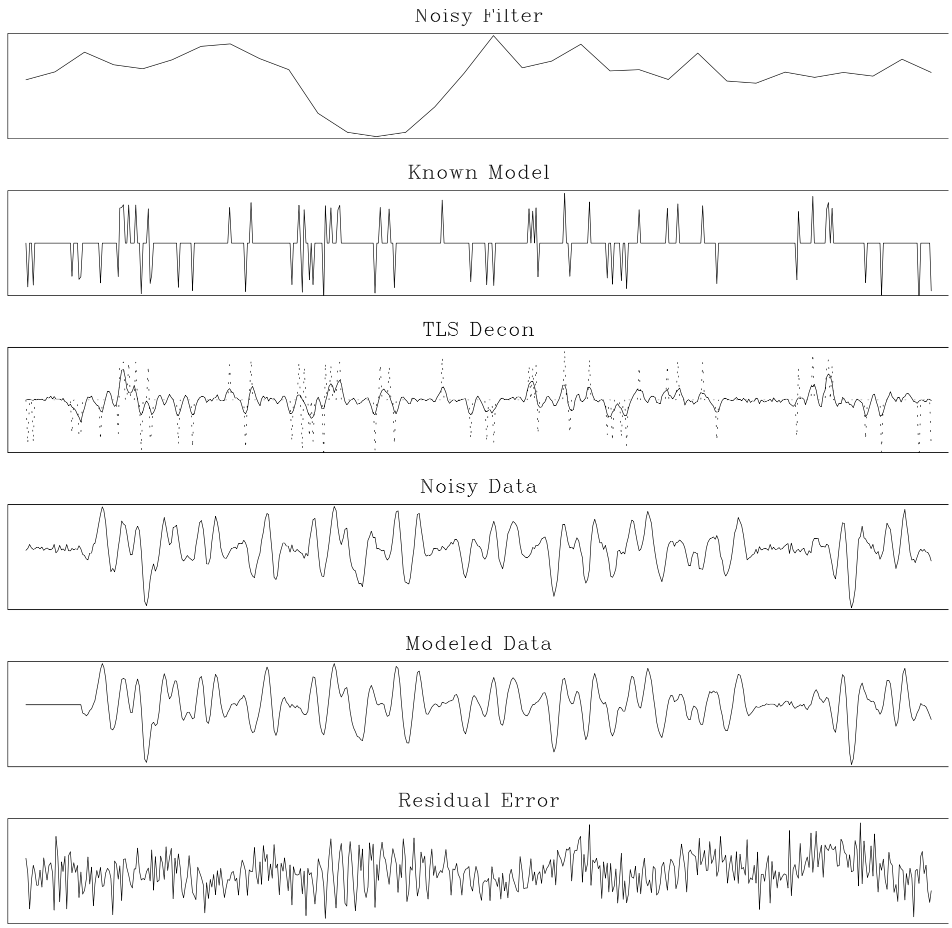

Figure 2: Top to bottom: 1) Known filter plus noise, 2) Known model, 3) Estimated total least-squares model overlaying known model, 4) Noisy data, 5) Modeled data, 6) Residual error. morgan2-decon.tls.noisy [ER]

More interesting are efforts to perturb the nullspace of inverse problems to infer model statistics (Clapp, 2002; Chen and Clapp, 2002).

Discussion: Error Distribution

Recall that in the earlier TLS formulation, the noise which contaminates *both* the operator and data is assumed to be white, with uniform variance. In practice, both the operator and data noise are likely to be correlated, with nonuniform variance. Björck (1996) notes that an appropriate change of variables can restore the validity of the assumptions. He defines a square matrix **D** which is applied, somewhat surprisingly, to the "data matrix" of equation (2).

$$(\mathbf{D}[\mathbf{L}\ \ \mathbf{d}]+[\mathbf{N}\ \ \mathbf{n}])\begin{bmatrix}\mathbf{m}\\-1\end{bmatrix}=\mathbf{0}. \qquad (6)$$

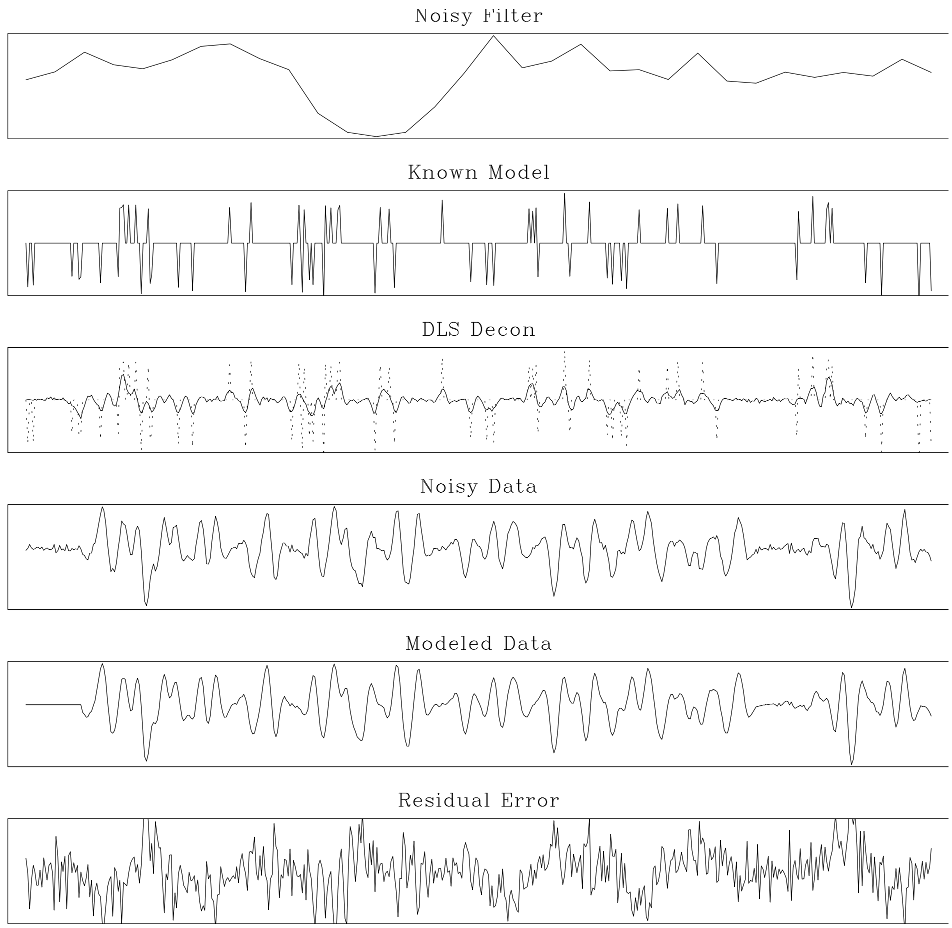

Figure 3: Top to bottom: 1) Known filter plus noise, 2) Known model, 3) Estimated damped least-squares model overlaying known model, 4) Noisy data, 5) Modeled data, 6) Residual error. morgan2-decon.dls.noisy [ER]

While it may seem intuitive to scale the noise, rather than the data, if the operator is diagonal (as it is in the fairytale world of uncorrelated noise), the inverse is trivial. Even if the noise is correlated, at SEP, we have considerable experience with the design of invertible decorrelation and balancing operators.

Are the restrictions (white, balanced) on the noise crippling? Zhu et al. (1997) claim that in scattering tomography experiments, correlated noise does not unduly harm the TLS result, and also that the TLS result in this case is still better than the normal LS result.

ACKNOWLEDGEMENT

At the 2002 SEP sponsor meeting, Peter Harris of CGG suggested I look into TLS.

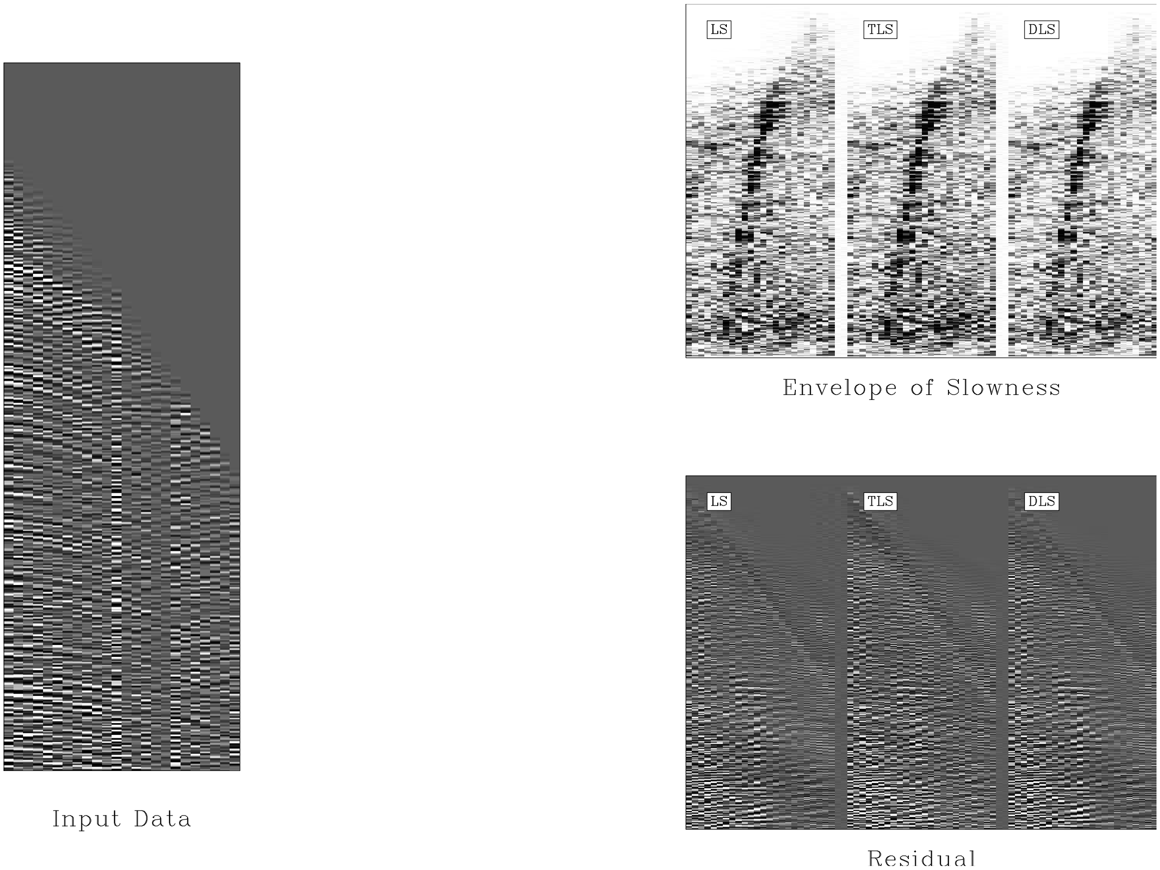

Figure 4: Left panel: Input data. Right-top: Envelope of estimated slowness model for LS, TLS, and DLS methods after 10 iterations. Right-bottom: Residual error for LS, TLS, and DLS solutions. morgan2-hrtcomp.10 [ER]

REFERENCES

Björck, A., 1996, Numerical methods for least squares problems: Society for Industrial and Applied Mathematics.

Chen, W., and Clapp, R., 2002, Exploring the relationship between uncertainty of ava attributes and rock information: SEP–**112**, 259–268.

Chen, H., Sarkar, T., Dianat, S., and Brulé, J., 1986, Adaptive spectral estimation by the conjugate gradient method: IEEE Transactions on Acoustics, Speech, and Signal Processing, **ASSP-34**, no. 2, 272–284.

Clapp, R. G., 2002, Effect of migration velocity uncertainty on amplitude information: 72nd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, submitted.

Fierro, R., Golub, G., Hansen, P., and O'Leary, D., 1997, Regularization by truncated total least squares: SIAM J. Sci. Comput., **18**, no. 4, 1223–1241.

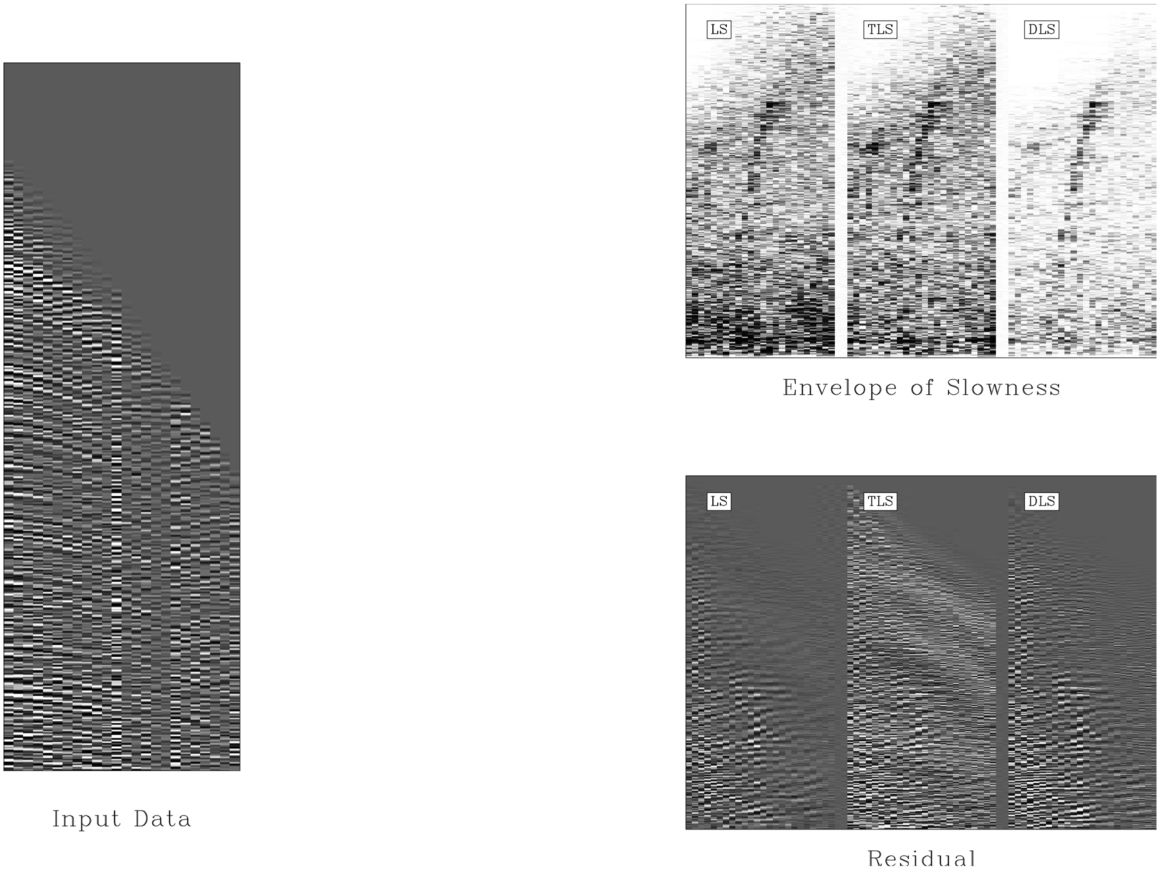

Figure 5: Left panel: Input data. Right-top: Envelope of estimated slowness model for LS, TLS, and DLS methods after 50 iterations. Right-bottom: Residual error for LS, TLS, and DLS solutions. morgan2-hrtcomp.50 [ER]

Golub, G., and Loan, C. V., 1980, An analysis of the total least squares problem: SIAM J. Numer. Anal., **15**, no. 17, 883–893.

Golub, G., 1973, Some modified matrix eigenvalue problems: SIAM (Soc. Ind. Appl. Math.) Rev., **15**, no. 15, 318–334.

Guitton, A., 2000a, Coherent noise attenuation using Inverse Problems and Prediction Error Filters: SEP–**105**, 27–48.

Guitton, A., 2000b, Prestack multiple attenuation using the hyperbolic Radon transform: SEP–**103**, 181–201.

Liu, Y., and Sacchi, M., 2002, Regularization of inverse problems using total least squares: Ann. Mtg, Can. Soc. Expl. Geophys., Expanded Abstracts, pages 1–3.

Lumley, D. E., Nichols, D., Ecker, C., Rekdal, T., and Berlioux, A., 1995, Amplitude-preserved processing and analysis of the Mobil AVO data set: SEP–**84**, 125–152.

Nichols, D., 1994, Velocity-stack inversion using $\mathbf{L_p}$ norms: SEP–**82**, 1–16.

Zhu, W., Wang, Y., Yao, Y., Chang, J., Graber, H., and Barbour, R., 1997, Iterative total least-squares image reconstruction algorithm for optical tomography by the conjugate gradient method: J. Opt. Soc. Am. A, **14**, no. 4, 799–807.

APPENDIX A: CONJUGATE GRADIENT MINIMIZATION OF THE RAYLEIGH QUOTIENT

Recall that $\mathbf{q} = [\mathbf{m} \quad -1]^T$, where $\mathbf{m}$ is the "usual" model (i.e., **Lm=d**). $\mathbf{q}_i$ is the estimated model vector at iteration `i`.

$$
\begin{aligned}
\mathbf{q}_0 &= \frac{\mathbf{q}_0}{\sqrt{\mathbf{q}_0^T\mathbf{q}_0}} \quad \Leftarrow \text{Model} && \text{(A-1)}\\
\lambda_0 &= \mathbf{q}_0^T\mathbf{A}^T\mathbf{A}\mathbf{q}_0 \quad \Leftarrow \text{Estimated smallest eigenvalue} && \text{(A-2)}\\
\mathbf{r}_0 &= \lambda_0\mathbf{q}_0 - \mathbf{A}^T\mathbf{A}\mathbf{q}_0 \quad \Leftarrow \text{Residual} && \text{(A-3)}\\
\mathbf{s}_0 &= \mathbf{r}_0 \quad \Leftarrow \text{Solution Step} && \text{(A-4)}\\
\text{iterate} &\;\{ \quad (k=0, n_{iter}) && \text{(A-5)}\\
P_{a,k} &= \mathbf{q}_k^T\mathbf{A}^T\mathbf{A}\mathbf{s}_k && \text{(A-6)}\\
P_{b,k} &= \mathbf{s}_k^T\mathbf{A}^T\mathbf{A}\mathbf{s}_k && \text{(A-7)}\\
P_{c,k} &= \mathbf{s}_k^T\mathbf{q}_k && \text{(A-8)}\\
P_{d,k} &= \mathbf{s}_k^T\mathbf{s}_k && \text{(A-9)}\\
b &= P_{b,k} - \lambda_k P_{d,k} && \text{(A-10)}\\
c &= P_{a,k} - \lambda_k P_{c,k} && \text{(A-11)}\\
d &= P_{b,k}P_{c,k} - P_{a,k}P_{d,k} && \text{(A-12)}\\
\alpha_k &= \frac{-b+\sqrt{b^2-4dc}}{2d} && \text{(A-13)}\\
\mathbf{q}_{k+1} &= \mathbf{q}_k + \alpha_k\mathbf{s}_k && \text{(A-14)}\\
\mathbf{q}_{k+1} &= \frac{\mathbf{q}_{k+1}}{\sqrt{\mathbf{q}_{k+1}^T\mathbf{q}_{k+1}}} && \text{(A-15)}\\
\lambda_{k+1} &= \mathbf{q}_{k+1}^T\mathbf{A}^T\mathbf{A}\mathbf{q}_{k+1} && \text{(A-16)}\\
\mathbf{r}_{k+1} &= \lambda_{k+1}\mathbf{q}_{k+1} - \mathbf{A}^T\mathbf{A}\mathbf{q}_{k+1} && \text{(A-17)}\\
\beta_k &= -\frac{\mathbf{s}_k\mathbf{A}^T\mathbf{A}\mathbf{r}_{k+1}}{\mathbf{s}_k\mathbf{A}^T\mathbf{A}\mathbf{s}_k} && \text{(A-18)}\\
\mathbf{s}_{k+1} &= \mathbf{r}_{k+1} + \beta_k\mathbf{s}_k && \text{(A-19)}\\
\}& && \text{(A-20)}\\
\mathbf{q}_{n_{iter}} &= \mathbf{q}_{n_{iter}} / \left(-\mathbf{q}_{n_{iter}}[m+1]\right) && \text{(A-21)}
\end{aligned}
$$