Shaping regularization in geophysical estimation problems[a]

[a]Published in Geophysics, 72 , no. 2, R29-R36, (2007)

Sergey Fomel

ABSTRACT

Regularization is a required component of geophysical estimation problems that operate with insufficient data. The goal of regularization is to impose additional constraints on the estimated model. I introduce shaping regularization, a general method for imposing constraints by explicit mapping of the estimated model to the space of admissible models. Shaping regularization is integrated in a conjugate-gradient algorithm for iterative least-squares estimation. It provides the advantage of a better control on the estimated model in comparison with traditional regularization methods and, in some cases, leads to a faster iterative convergence. Simple data interpolation and seismic velocity estimation examples illustrate the concept[a].

[a]Presented at the SEG Annual Meeting in 2005

INTRODUCTION

A great number of geophysical estimation problems are mathematically ill-posed because they operate with insufficient data (Jackson, 1972). Regularization is a technique for making the estimation problems well-posed by adding indirect constraints on the estimated model (Engl et al., 1996; Zhdanov, 2002). Developed originally by Tikhonov (1963) and others, the method of regularization has become an indispensable part of the inverse problem theory and has found many applications in geophysical problems: traveltime tomography (Bube and Langan, 1999; Clapp et al., 2004), migration velocity analysis (Woodward et al., 1998; Zhou et al., 2003), high-resolution Radon transform (Trad et al., 2003), spectral decomposition (Portniaguine and Castagna, 2004), etc.

While the main goal of inversion is to fit the observed data, Tikhonov's regularization adds another goal of fitting the estimated model to a priorly assumed behavior. The contradiction between the two goals often leads to a slow convergence of iterative estimation algorithms (Harlan, 1995). The speed can be improved considerably by an appropriate model reparameterization or preconditioning (Fomel and Claerbout, 2003). However, the difficult situation of trying to satisfy two contradictory goals simultaneously leads sometimes to an undesirable behavior of the solution at the early iterations of an iterative optimization scheme.

In this paper, I introduce *shaping regularization*, a new general method of imposing regularization constraints. A shaping operator provides an explicit mapping of the model to the space of acceptable models. The operator is embedded in an iterative optimization scheme (the conjugate-gradient algorithm) and allows for a better control on the estimation result. Shaping into the space of smooth functions can be accomplished with efficient lowpass filtering. Depending on the desirable result, it is also possible to shape the model into a piecewise-smooth function, a function following geological structure, or a function representable in a predefined basis. I illustrate the shaping concept with simple numerical experiments of data interpolation and seismic velocity estimation.

REVIEW OF TIKHONOV'S REGULARIZATION

If the data are represented by vector $\mathbf{d}$, model parameters by vector $\mathbf{m}$, and their functional relationship is defined by the forward modeling operator $\mathbf{L}$, the least-squares optimization approach amounts to minimizing the least-squares norm of the residual difference $\mathbf{L\,m} - \mathbf{d}$. In Tikhonov's regularization approach, one additionally attempts to minimize the norm of $\mathbf{D\,m}$, where $\mathbf{D}$ is the *regularization* operator. Thus, we are looking for the model $\mathbf{m}$ that minimizes the least-squares norm of the compound vector $\left[\begin{array}{cc} \mathbf{Lm} - \mathbf{d} & \epsilon \mathbf{Dm} \end{array}\right]^T$, where ϵ is a scalar scaling parameter. The formal solution has the well-known form

$$\widehat{\mathbf{m}} = \left(\mathbf{L}^T\,\mathbf{L} + \epsilon^2\,\mathbf{D}^T\,\mathbf{D}\right)^{-1}\,\mathbf{L}^T\,\mathbf{d}\;, \tag{1}$$

where $\widehat{\mathbf{m}}$ denotes the least-squares estimate of $\mathbf{m}$, and $\mathbf{L}^T$ denotes the adjoint operator. One can carry out the optimization iteratively with the help of the conjugate-gradient method (Hestenes and Steifel, 1952) or its analogs. Iterative methods have computational advantages in large-scale problems when forward and adjoint operators are represented by sparse matrices and can be computed efficiently (Saad, 2003; van der Vorst, 2003).

In an alternative approach, one obtains the regularized estimate by minimizing the least-squares norm of the compound vector $\left[\begin{array}{cc} \mathbf{p} & \mathbf{r} \end{array}\right]^T$ under the constraint

$$\epsilon\mathbf{r} = \mathbf{d} - \mathbf{Lm} = \mathbf{d} - \mathbf{LPp}\;. \tag{2}$$

Here $\mathbf{P}$ is the *model reparameterization* operator that translates vector $\mathbf{p}$ into the model vector $\mathbf{m}$, $\mathbf{r}$ is the scaled residual vector, and ϵ has the same meaning as before. The formal solution of the preconditioned problem is given by

$$\widehat{\mathbf{m}} = \mathbf{P}\,\widehat{\mathbf{p}} = \mathbf{P}\,\mathbf{P}^T\,\mathbf{L}^T\,\left(\mathbf{L}\,\mathbf{P}\,\mathbf{P}^T\,\mathbf{L}^T + \epsilon^2\,\mathbf{I}\right)^{-1}\,\mathbf{d}\;, \tag{3}$$

where $\mathbf{I}$ is the identity operator in the data space. Estimate 3 is mathematically equivalent to estimate 1 if $\mathbf{D}^T\,\mathbf{D}$ is invertible and

$$\left(\mathbf{D}^T\,\mathbf{D}\right)^{-1} = \mathbf{P}\,\mathbf{P}^T = \mathbf{C}\;. \tag{4}$$

Statistical theory of least-squares estimation connects $\mathbf{C}$ with the model covariance operator (Tarantola, 2004). In a more general case of reparameterization, the size of $\mathbf{p}$ may be different from the size of $\mathbf{m}$, and $\mathbf{C}$ may not have the full rank. In iterative methods, the preconditioned formulation often leads to faster convergence. Fomel and Claerbout (2003) suggest constructing preconditioning operators in multi-dimensional problems by recursive helical filtering.

SMOOTHING BY REGULARIZATION

Let us consider an application of Tikhonov's regularization to one of the simplest possible estimation problems: smoothing. The task of smoothing is to find a model $\mathbf{m}$ that fits the observed data $\mathbf{d}$ but is in a certain sense smoother. In this case, the forward operator $\mathbf{L}$ is simply the identity operator, and the formal solutions 1 and 3 take the form

$$\widehat{\mathbf{m}} = \left(\mathbf{I} + \epsilon^2\,\mathbf{D}^T\,\mathbf{D}\right)^{-1}\,\mathbf{d} = \mathbf{C}\,\left(\mathbf{C} + \epsilon^2\,\mathbf{I}\right)^{-1}\,\mathbf{d}\;. \tag{5}$$

Smoothness is controlled by the choice of the regularization operator $\mathbf{D}$ and the scaling parameter ϵ.

Figure 1 shows the impulse response of the regularized smoothing operator in the 1-D case when $\mathbf{D}$ is the first difference operator. The impulse response has exponentially decaying tails. Repeated application of smoothing in this case is equivalent to applying an implicit Euler finite-difference scheme to the solution of the diffusion equation

$$\frac{\partial \mathbf{m}}{\partial t} = -\mathbf{D}^T\,\mathbf{D}\,\mathbf{m} \tag{6}$$

The impulse response converges to a Gaussian bell-shape curve in the physical domain, while its spectrum converges to a Gaussian in the frequency domain.

As far as the smoothing problem is concerned, there are better ways to smooth signals than applying equation 5. One example is triangle smoothing (Claerbout, 1992). To define triangle smoothing of one-dimensional signals, start with box smoothing, which, in the Z-transform notation, is a convolution with the filter

$$B_k(Z) = \frac{1}{k}\,\left(1 + Z + Z^2 + \cdots + Z^k\right) = \frac{1}{k}\,\frac{1-Z^{k+1}}{1-Z}\;, \tag{7}$$

where k is the filter length. Form a triangle smoother by correlation of two boxes

$$T_k(Z) = B_k(1/Z)\,B_k(Z) \tag{8}$$

Triangle smoothing is more efficient than regularized smoothing, because it requires twice less floating point multiplications. It also provides smoother results while having

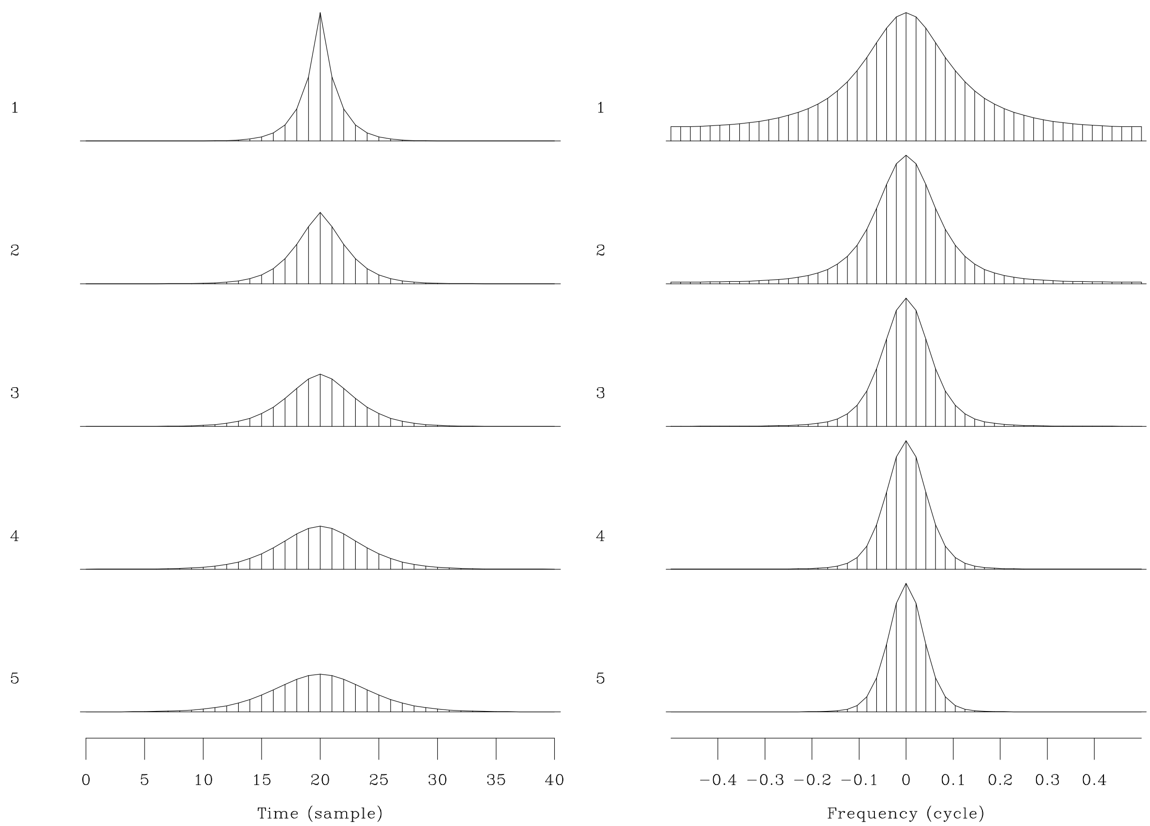

Figure 1: Left: impulse response of regularized smoothing. Repeated smoothing converges to a Gaussian bell shape. Right: frequency spectrum of regularized smoothing. The spectrum also converges to a Gaussian.

a compactly supported impulse response (Figure 2). Repeated application of triangle smoothing also makes the impulse response converge to a Gaussian shape but at a significantly faster rate. One can also implement smoothing by Gaussian filtering in the frequency domain or by applying other types of bandpass filters.

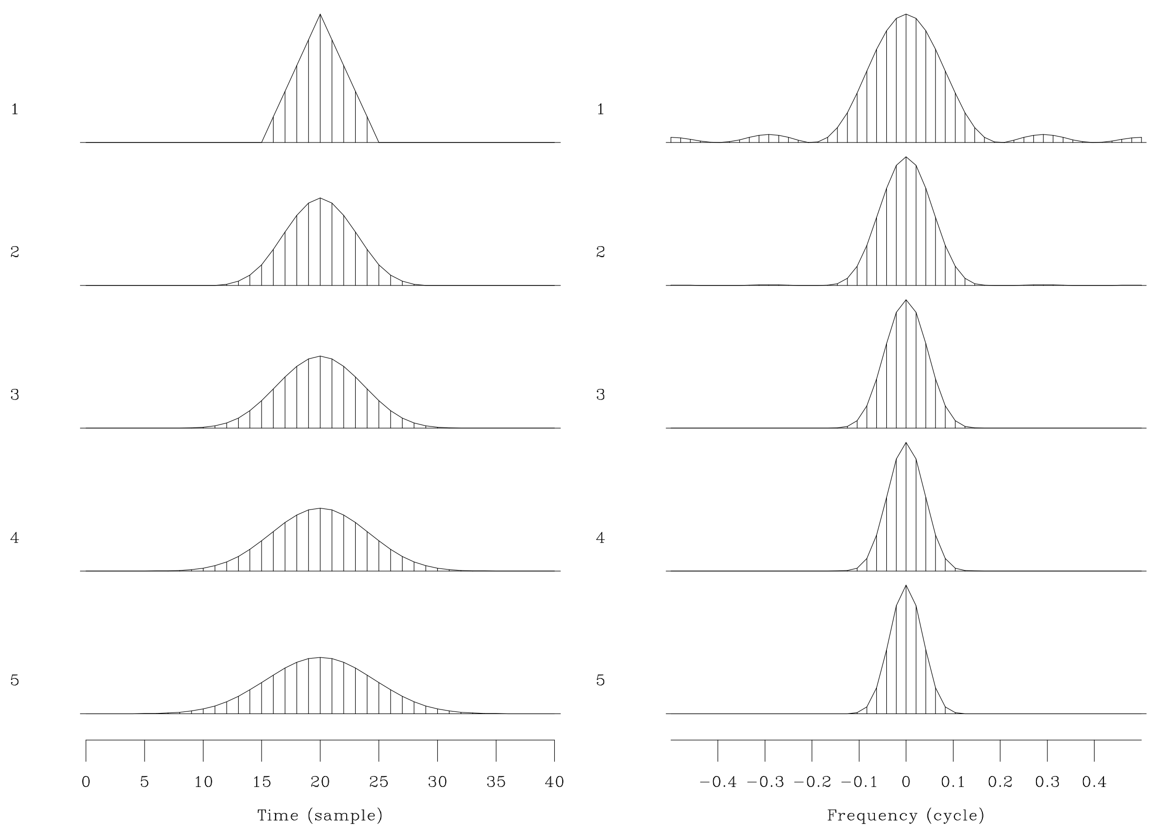

Figure 2: Left: impulse response of triangle smoothing. Repeated smoothing converges to a Gaussian bell shape. Right: frequency spectrum of triangle smoothing. Convergence to a Gaussian is faster than in the case of regularized smoothing. Compare to Figure 1.

SHAPING REGULARIZATION IN THEORY

The idea of shaping regularization starts with recognizing smoothing as a fundamental operation. In a more general sense, smoothing implies mapping of the input model to the space of admissible functions. I call the mapping operator *shaping*. Shaping operators do not necessarily smooth the input but they translate it into an acceptable model.

Taking equation 5 and using it as the definition of the regularization operator $\mathbf{D}$, we can write

$$\mathbf{S} = \left(\mathbf{I} + \epsilon^2\,\mathbf{D}^T\,\mathbf{D}\right)^{-1} \tag{9}$$

or

$$\epsilon^2\,\mathbf{D}^T\,\mathbf{D} = \mathbf{S}^{-1} - \mathbf{I}\;. \tag{10}$$

Substituting equation 10 into 1 yields a formal solution of the estimation problem regularized by shaping:

$$\widehat{\mathbf{m}} = \left(\mathbf{L}^T\,\mathbf{L} + \mathbf{S}^{-1} - \mathbf{I}\right)^{-1}\,\mathbf{L}^T\,\mathbf{d} = \left[\mathbf{I} + \mathbf{S}\,\left(\mathbf{L}^T\,\mathbf{L} - \mathbf{I}\right)\right]^{-1}\,\mathbf{S}\,\mathbf{L}^T\,\mathbf{d}\;. \tag{11}$$

The meaning of equation 11 is easy to interpret in some special cases:

- If $\mathbf{S} = \mathbf{I}$ (no shaping applied), we obtain the solution of unregularized problem.
- If $\mathbf{L}^T\,\mathbf{L} = \mathbf{I}$ ($\mathbf{L}$ is a unitary operator), the solution is simply $\mathbf{S}\,\mathbf{L}^T\,\mathbf{d}$ and does not require any inversion.
- If $\mathbf{S} = \lambda\,\mathbf{I}$ (shaping by scaling), the solution approaches $\lambda\,\mathbf{L}^T\,\mathbf{d}$ as λ goes to zero.

The operator $\mathbf{L}$ may have physical units that require scaling. Introducing scaling of $\mathbf{L}$ by $1/\lambda$ in equation 11, we can rewrite it as

$$\widehat{\mathbf{m}} = \left[\lambda^2\,\mathbf{I} + \mathbf{S}\,\left(\mathbf{L}^T\,\mathbf{L} - \lambda^2\,\mathbf{I}\right)\right]^{-1}\,\mathbf{S}\,\mathbf{L}^T\,\mathbf{d}\;. \tag{12}$$

The λ scaling in equation 12 controls the relative scaling of the forward operator $\mathbf{L}$ but not the shape of the estimated model, which is controlled by the shaping operator $\mathbf{S}$.

Iterative inversion with the conjugate-gradient algorithm requires symmetric positive definite operators (Hestenes and Steifel, 1952). The inverse operator in equation 12 can be symmetrized when the shaping operator is symmetric and representable in the form $\mathbf{S} = \mathbf{H}\mathbf{H}^T$ with a square and invertible $\mathbf{H}$. The symmetric form of equation 12 is

$$\widehat{\mathbf{m}} = \mathbf{H}\,\left[\lambda^2\,\mathbf{I} + \mathbf{H}^T\,\left(\mathbf{L}^T\,\mathbf{L} - \lambda^2\,\mathbf{I}\right)\,\mathbf{H}\right]^{-1}\,\mathbf{H}^T\,\mathbf{L}^T\,\mathbf{d}\;. \tag{13}$$

When the inverted matrix is positive definite, equation 13 is suitable for an iterative inversion with the conjugate-gradient algorithm. Appendix A contains a complete algorithm description.

FROM TRIANGLE SMOOTHING TO TRIANGLE SHAPING

The idea of triangle smoothing can be generalized to produce different shaping operators for different applications. Let us assume that the estimated model is organized in a sequence of records, as follows: $\mathbf{m} = \left[\begin{array}{cccc}\mathbf{m}_1 & \mathbf{m}_2 & \vdots & \mathbf{m}_n\end{array}\right]^T$. Depending on the application, the records can be samples, traces, shot profiles, etc. Let us further assume that, for each pair of neighboring records, we can design a prediction operator

$\mathbf{Z}_{k\to k+1}$, which predicts record $k+1$ from record k. A global prediction operator is then

$$\mathbf{Z} = \left[\begin{array}{cccccc} 0 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{Z}_{1\to 2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{Z}_{2\to 3} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \mathbf{Z}_{3\to 4} & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \mathbf{Z}_{n-1\to n} & 0 \end{array}\right] . \tag{14}$$

The operator $\mathbf{Z}$ effectively shifts each record to the next one. When local prediction is done with identity operators, this operation is completely analogous to the Z operator used in the theory of digital signal processing. The $\mathbf{Z}$ operator can be squared, as follows:

$$\mathbf{Z}^2 = \left[\begin{array}{cccccc} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \mathbf{Z}_{2\to 3}\,\mathbf{Z}_{1\to 2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \mathbf{Z}_{3\to 4}\,\mathbf{Z}_{2\to 3} & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{Z}_{n-1\to n}\,\mathbf{Z}_{n-2\to n-1} & 0 & 0 \end{array}\right] . \tag{15}$$

In a shorter notation, we can denote prediction of record j from record i by $\mathbf{Z}_{i\to j}$ and write

$$\mathbf{Z}^2 = \left[\begin{array}{cccccc} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \mathbf{Z}_{1\to 3} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \mathbf{Z}_{2\to 4} & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{Z}_{n-2\to n} & 0 & 0 \end{array}\right] . \tag{16}$$

Subsequently, the prediction operator $\mathbf{Z}$ can be taken to higher powers. This leads immediately to an idea on how to generalize box smoothing: predict each record from the record immediately preceding it, the record two steps away, etc. and average all those predictions and the actual records. In mathematical notation, a box shaper of length k is then simply

$$\mathbf{B}_k = \frac{1}{k}\,\left(\mathbf{I} + \mathbf{Z} + \mathbf{Z}^2 + \cdots + \mathbf{Z}^k\right)\;, \tag{17}$$

which is completely analogous to equation 7. Implementing equation 17 directly requires many computational operations. Noting that

$$\left(\mathbf{I} - \mathbf{Z}\right)\,\mathbf{B}_k = \frac{1}{k}\,\left(\mathbf{I} - \mathbf{Z}^{k+1}\right)\;, \tag{18}$$

we can rewrite equation 17 in the compact form

$$\mathbf{B}_k = \frac{1}{k}\,\left(\mathbf{I} - \mathbf{Z}\right)^{-1}\,\left(\mathbf{I} - \mathbf{Z}^{k+1}\right)\;, \tag{19}$$

which can be implemented economically using recursive inversion of the lower triangular operator $\mathbf{I} - \mathbf{Z}$. Finally, combining two generalized box smoothers creates a symmetric generalized triangle shaper

$$\mathbf{T}_k = \mathbf{B}_k^T \, \mathbf{B}_k \; , \tag{20}$$

which is analogous to equation 8. A triangle shaper uses local predictions from both the left and the right neighbors of a record and averages them using triangle weights.

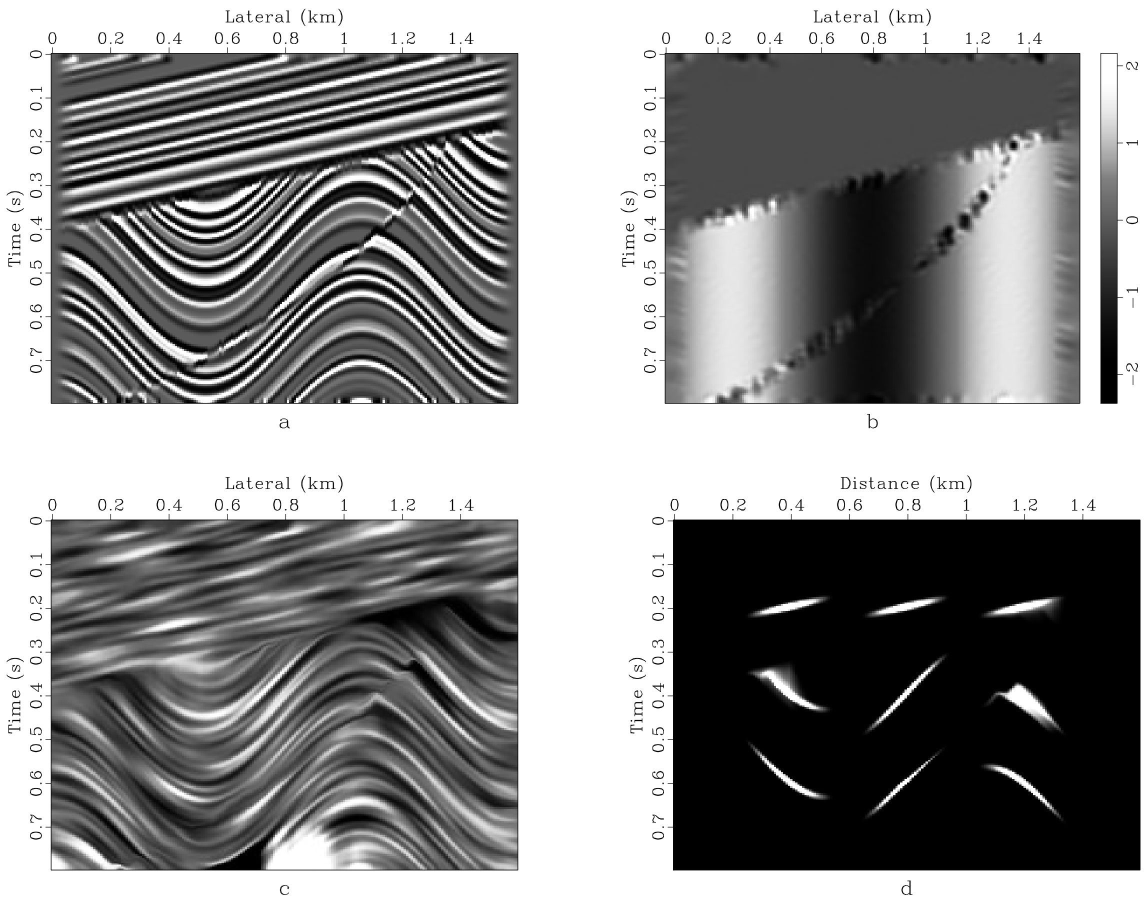

Figure 3: Shaping by smoothing along local dip directions according to operator $\mathbf{T}_k$ from equation 20. a: an example image, b: local dip estimation, c: smoothing random numbers along local dips, d: impulse responses of oriented smoothing for nine different locations in the image space.

Figure 3 illustrates generalized triangle shaping by constructing a non-stationary smoothing operator that follows local structural dips. Figure 3a shows a synthetic image from Claerbout (2006). Figure 3b is a local dip estimate obtained with plane-wave destruction (Fomel, 2002). Figure 3c is the result of applying triangle smoothing oriented along local dip directions to a field of random numbers. Oriented smoothing generates a pattern reflecting the structural composition of the original image. This construction resembles the method of Claerbout and Brown (1999). Figure 3d shows the impulse responses of oriented smoothing for several distinct locations in the image space. As illustrated later in this paper, oriented smoothing can be applied for generating geophysical Earth models that are compliant with the local geological structure (Sinoquet, 1993; Versteeg and Symes, 1993; Clapp et al., 2004).

Appendix B describes general rules for combining elementary shaping operators.

EXAMPLES

Two simple examples in data regularization and seismic velocity estimation illustrate the method of shaping.

1-D inverse data interpolation

I start with a simple 1-D example: a benchmark data regularization test used previously by Fomel and Claerbout (2003).

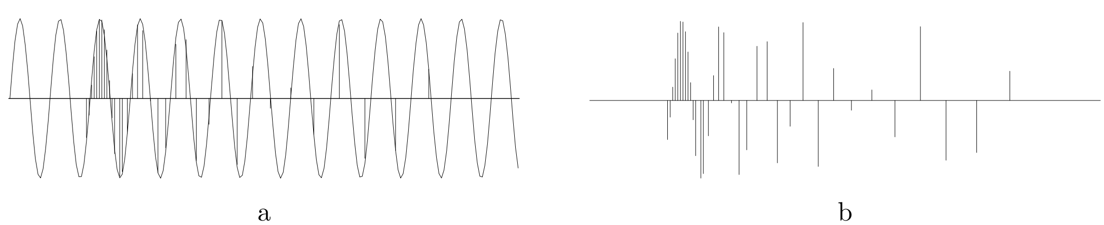

Figure 4: The input data (b) are irregular samples from a sinusoid (a).

The input synthetic data are irregular samples from a sinusoidal signal (Figure 4). The task of data regularization is to reconstruct the data on a regular grid. The forward operator $\mathbf{L}$ in this case is forward interpolation from a regular grid using linear (two-point) interpolation.

Figure 5 shows some of the first iterations and the final results of inverse interpolation with Tikhonov's regularization using equation 1 and with model preconditioning using equation 3. The regularization operator $\mathbf{D}$ in equation 1 is the first finite difference, and the preconditioning operator $\mathbf{P}$ in 3 is the inverse of $\mathbf{D}$ or causal integration. The preconditioned iteration converges faster but its very first iterations produce unreasonable results. This type of behavior can be dangerous in real large-scale problems, when only few iterations are affordable.

The left side of Figure 6 shows some of the first iterations and the final result of inverse interpolation with shaping regularization, where the shaping operator $\mathbf{S}$ was chosen to be Gaussian smoothing with the impulse response width of about 10 samples. The final result is smoother, and the iteration is both fast-converging and producing reasonable results at the very first iterations. Thanks to the fact that the smoothing operation is applied at each iteration, the estimated model is guaranteed to have the prescribed shape.

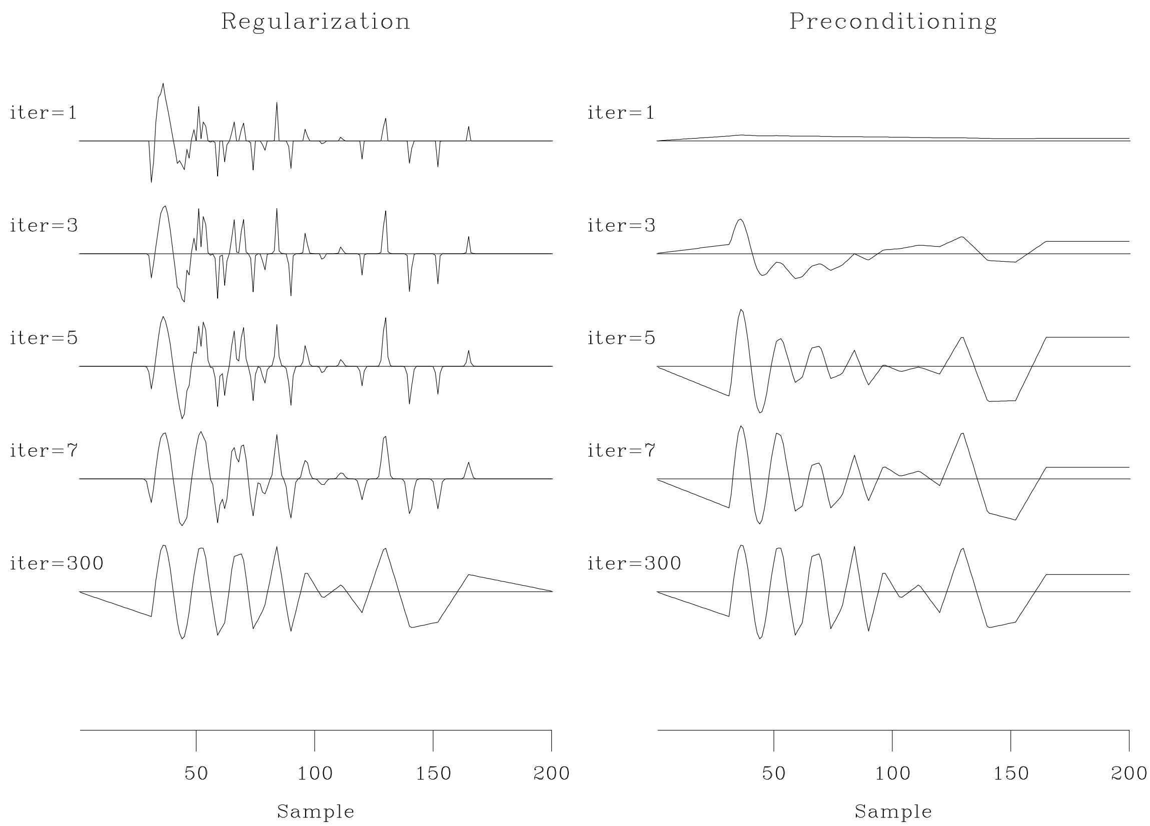

Figure 5: The first iterations and the final result of inverse interpolation with Tikhonov's regularization using equation 1 (left) and with model preconditioning using equation 3 (right). The regularization operator $\mathbf{D}$ is the first finite difference. The preconditioning operator $\mathbf{P} = \mathbf{D}^{-1}$ is causal integration. The number of iterations is indicated in the plot labels.

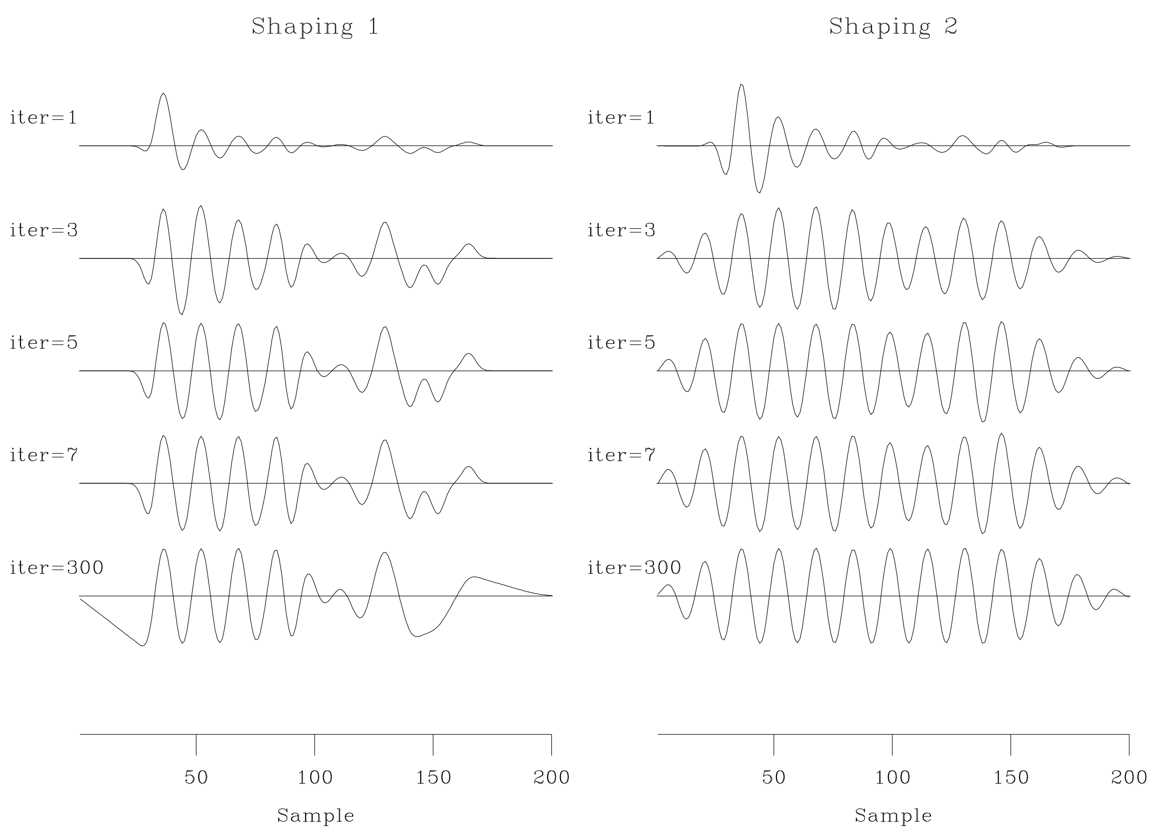

Figure 6: The first iterations and the final result of inverse interpolation with shaping regularization using equation 13. Left: the shaping operator **H** is lowpass filtering with a Gaussian smoother. Right: the shaping operator **H** is bandpass filtering with a shifted Gaussian. Shaping by bandpass filtering recovers the sinusoidal shape of the estimated model. The number of iterations is indicated in the plot labels.

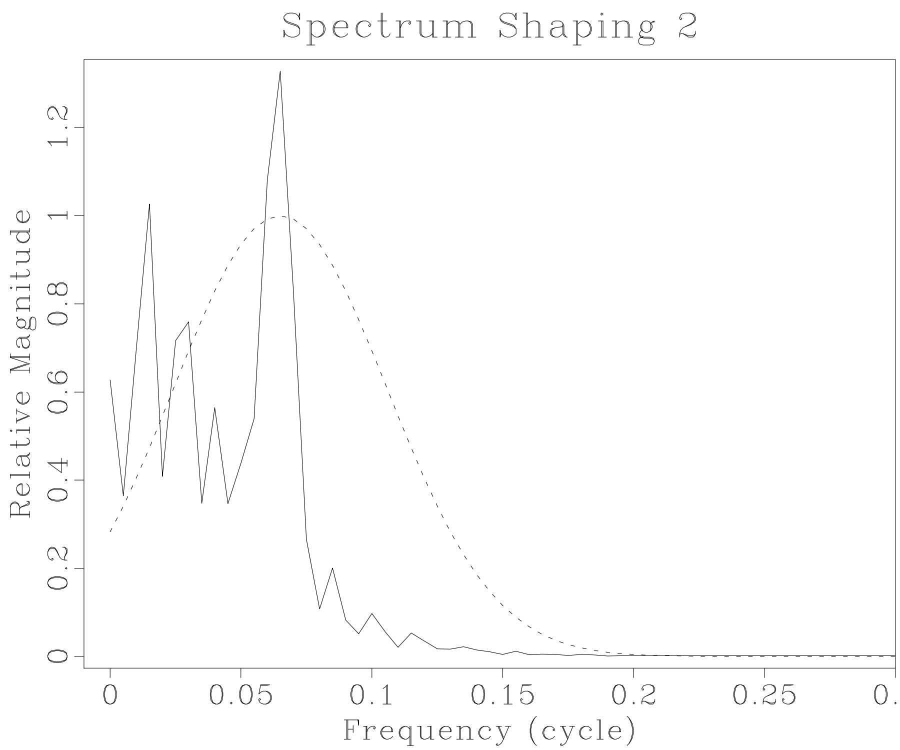

Figure 7: Spectrum of the estimated model (solid curve) fitted to a shifted Gaussian (dashed curve). The Gaussian band-limited filter defines a shaping operator for recovering a band-limited signal.

Examining the spectrum of the final result (Figure 7), one can immediately notice the peak at the dominant frequency of the initial sinusoid. Fitting a Gaussian shape to the peak defines a data-adaptive shaping operator as a bandpass filter implemented in the frequency domain (dashed curve in Figure 7). Inverse interpolation with the estimated shaping operator recovers the original sinusoid (right side of Figure 6). Analogous ideas in the model preconditioning context were proposed by Liu and Sacchi (2001).

Velocity estimation

The second example is an application of shaping regularization to seismic velocity estimation. Figure 8 shows a time-migrated image from a historic Gulf of Mexico dataset (Claerbout, 2006). The image was obtained by velocity continuation (Fomel, 2003). The corresponding migration velocity is shown in the right plot of Figure 8. Shaping regularization was used for picking a smooth velocity profile from semblance gathers obtained in the process of velocity continuation.

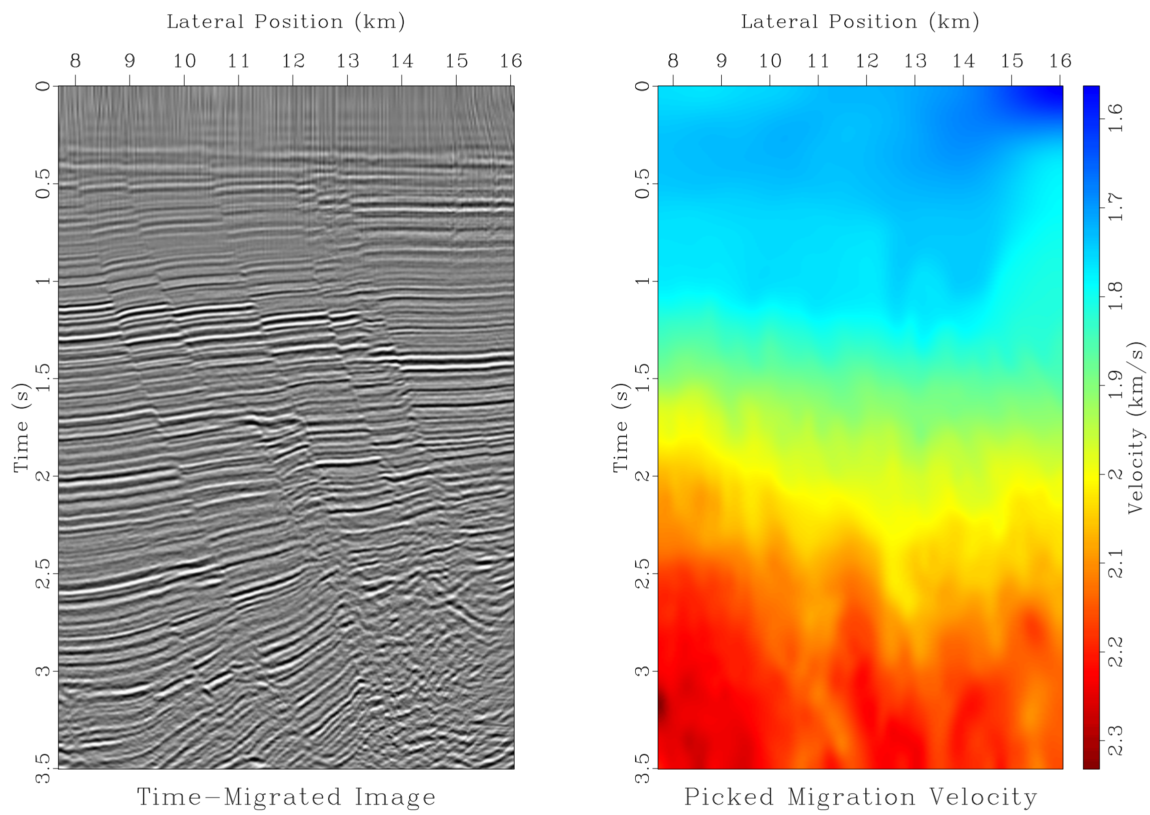

Figure 8: Left: time-migrated image. Right: The corresponding migration velocity from automatic picking.

The task of this example is to convert the time migration velocity to the interval velocity. I use the simple approach of Dix inversion (Dix, 1955) formulated as a regularized inverse problem (Valenciano et al., 2004). In this case, the forward operator **L** in equation 11 is a weighted time integration. There is a choice in choosing the shaping operator **H**.

Figure 9 shows the result of inversion with shaping by triangle smoothing. While the interval velocity model yields a good prediction of the measured velocity, it may not appear geologically plausible because the velocity structure does not follow the structure of seismic reflectors as seen in the migrated image.

Following the ideas of steering filters (Clapp et al., 1998, 2004) and plane-wave construction (Fomel and Guitton, 2006), I estimate local slopes in the migration image using the method of plane-wave destruction (Fomel, 2002) and define a triangle plane-wave shaping operator **H** using the method of the previous section. The result of inversion, shown in Figures 10 and 11, makes the estimated interval velocity follow the geological structure and thus appear more plausible for direct interpretation. Similar results were obtained by Fomel and Guitton (2006) using model parameterization by plane-wave construction but at a higher computational cost. In the case of shaping regularization, about 25 efficient iterations were sufficient to converge to the machine precision accuracy.

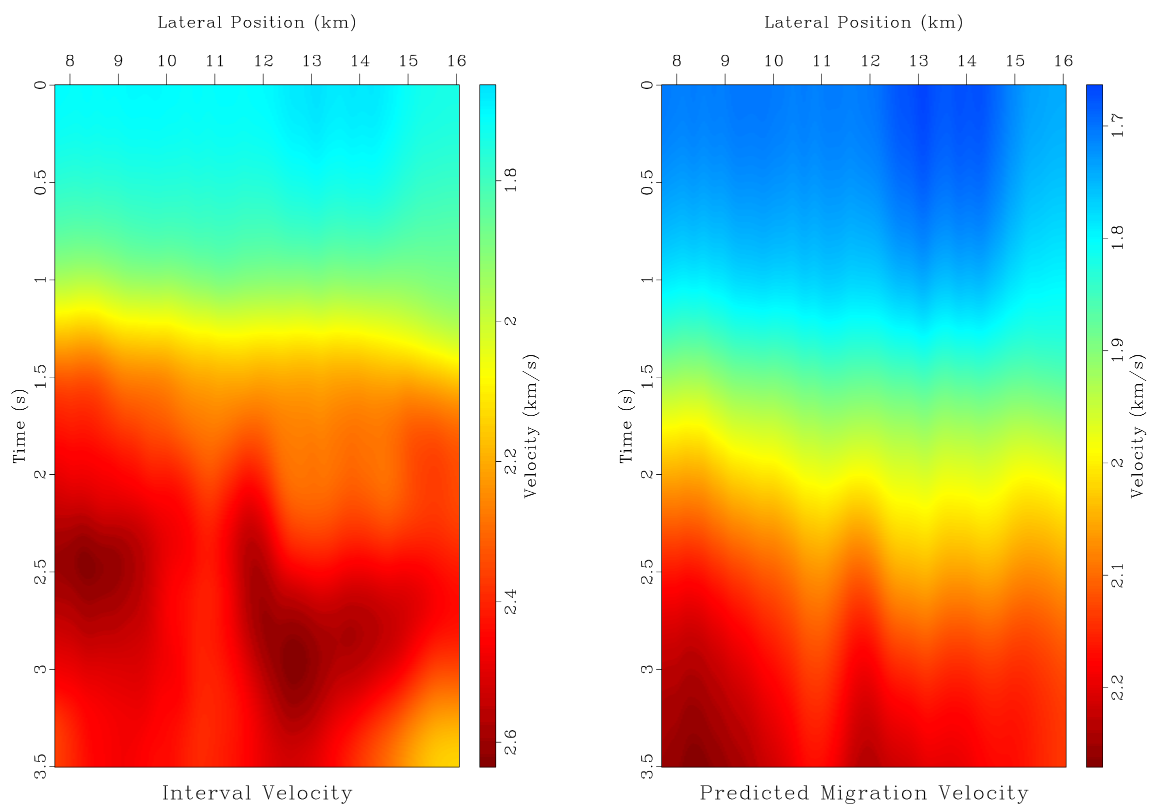

Figure 9: Left: estimated interval velocity. Right: predicted migration velocity. Shaping by triangle smoothing.

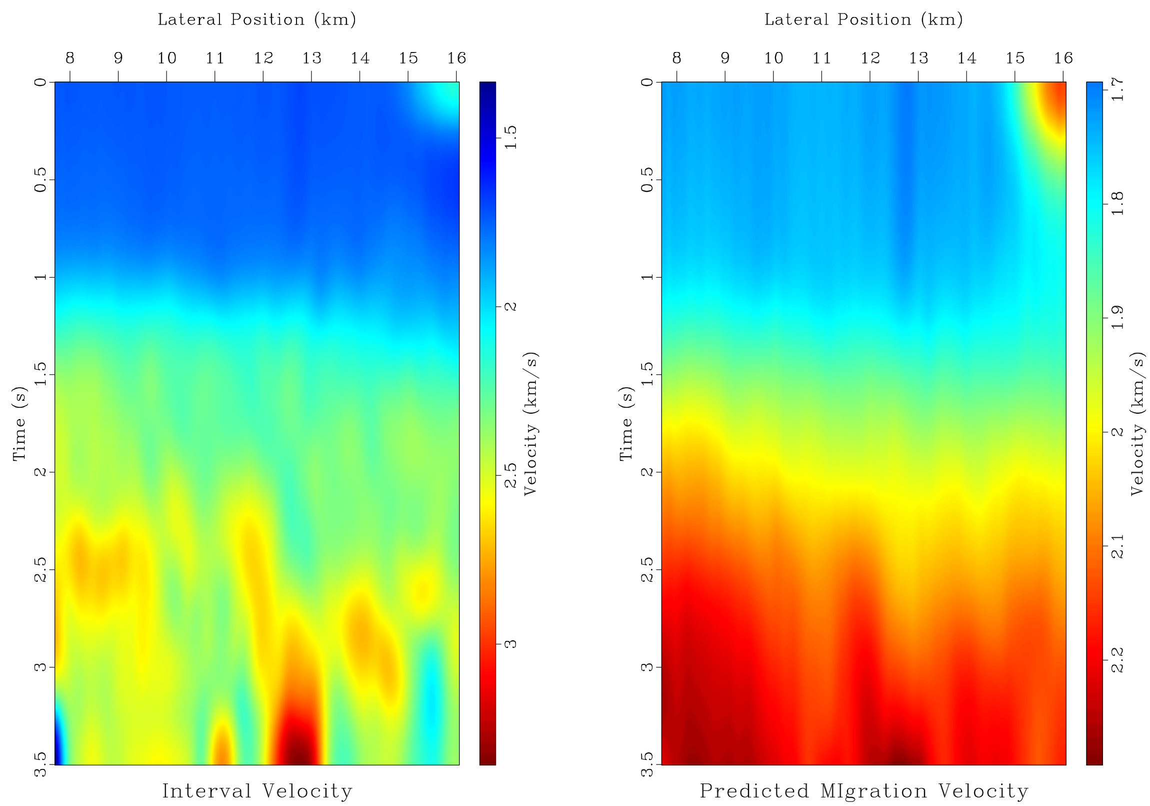

Figure 10: Left: estimated interval velocity. Right: predicted migration velocity. Shaping by triangle local plane-wave smoothing creates a velocity model consistent with the reflector structure.

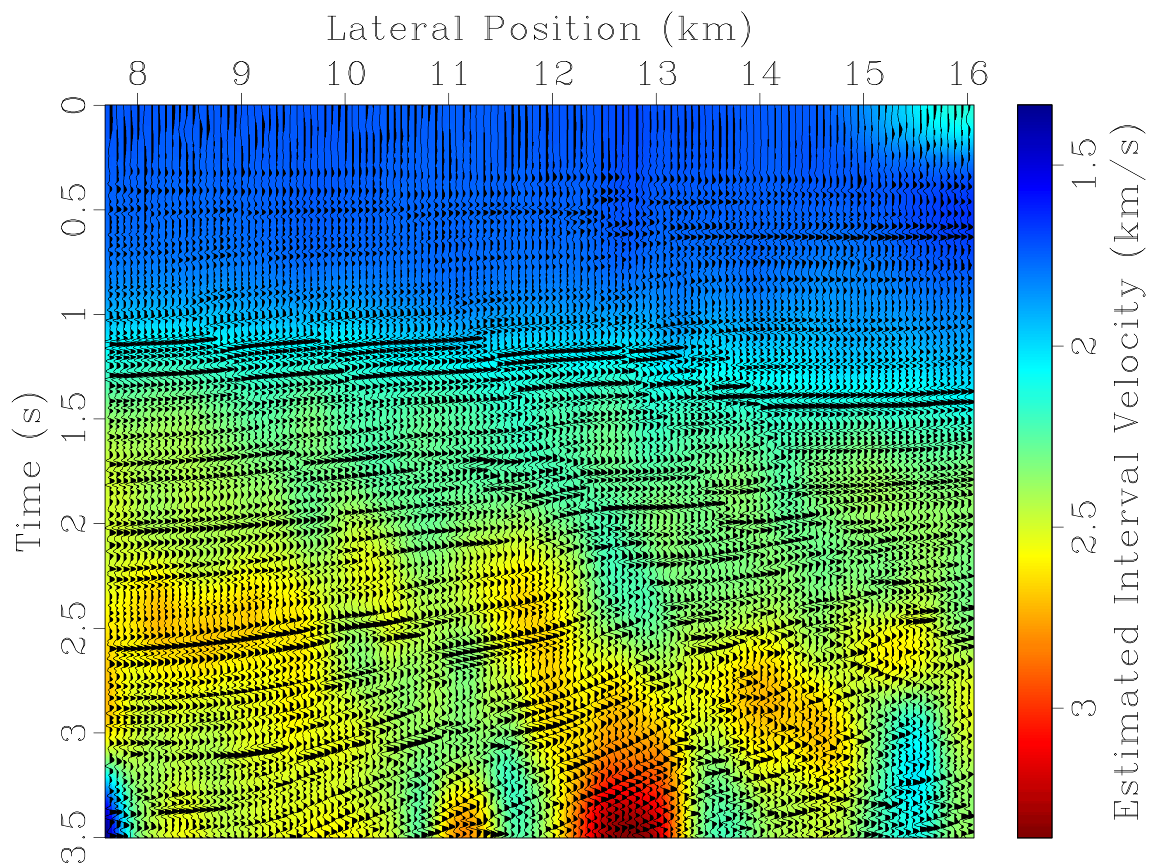

Figure 11: Seismic image from Figure 8 overlaid on top of the interval velocity model estimated with triangle plane-wave shaping regularization.

CONCLUSIONS

Shaping regularization is a new general method for imposing regularization constraints in estimation problems. The main idea of shaping regularization is to recognize shaping (mapping to the space of acceptable functions) as a fundamental operation and to incorporate it into iterative inversion.

There is an important difference between shaping regularization and conventional (Tikhonov's) regularization from the user prospective. Instead of trying to find and specify an appropriate regularization operator, the user of the shaping regularization algorithm specifies a shaping operator, which is often easier to design. Shaping operators can be defined following a triangle construction from local predictions or by combining elementary shapers.

I have shown two simple illustrations of shaping applications. The examples demonstrate a typical behavior of the method: enforced model compliance to the specified shape at each iteration and, in many cases, fast iterative convergence of the conjugate gradient iteration. The model compliance behavior follows from the fact that shaping enters directly into the iterative process and provides an explicit control on the shape of the estimated model.

ACKNOWLEDGMENTS

I would like to thank Pierre Hardy and Mauricio Sacchi for inspiring discussions, TOTAL for partially supporting this research, and three reviewers for helpful suggestions.

REFERENCES

Bube, K., and R. Langan, 1999, On a continuation approach to regularization for crosswell tomography: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1295–1298.

Claerbout, J., and M. Brown, 1999, Two-dimensional textures and prediction-error filters: 61st Mtg., Eur. Assn. Geosci. Eng., Session:1009.

Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.

———, 2006, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

Clapp, R. G., B. Biondi, and J. F. Claerbout, 2004, Incorporating geologic information into reflection tomography: Geophysics, **69**, 533–546.

Clapp, R. G., B. L. Biondi, S. B. Fomel, and J. F. Claerbout, 1998, Regularizing velocity estimation using geologic dip information: 68th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1851–1854.

Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68–86.

Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.

Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

———, 2003, Time-migration velocity analysis by velocity continuation: Geophysics, **68**, 1662–1672.

Fomel, S., and J. Claerbout, 2003, Multidimensional recursive filter preconditioning in geophysical estimation problems: Geophysics, **68**, 577–588.

Fomel, S., and A. Guitton, 2006, Regularizing seismic inverse problems by model re-parameterization using plane-wave construction: Geophysics, **71**, S209–S217.

Fomel, S., P. Sava, J. Rickett, and J. Claerbout, 2003, The Wilson-Burg method of spectral factorization with application to helical filtering: Geophysical Prospecting, **51**, 409–420.

Harlan, W. S., 1995, Regularization by model redefinition: http://billharlan.com/pub/papers/regularization.pdf.

Hestenes, M. R., and E. Steifel, 1952, Methods of conjugate gradients for solving linear systems: J. Res. Nat. Bur. Stand., **49**, 409–436.

Jackson, D. D., 1972, Interpretation of inaccurate, insufficient and inconsistent data: Geophys. J. R. Astron. Soc., **28**, 97–109.

Liu, B., and M. Sacchi, 2001, Minimum weighted norm interpolation of seismic data with adaptive weights: 71st Ann. Internat. Mtg, Soc. of Expl. Geophys., 1921–1924.

Portniaguine, O., and J. Castagna, 2004, Inverse spectral decomposition, *in* 74th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1786–1789.

Saad, Y., 2003, Iterative methods for sparse linear systems: SIAM.

Sinoquet, D., 1993, Modeling a priori information on the velocity field in reflection tomography: 63rd Ann. Internat. Mtg, Soc. of Expl. Geophys., 591–594.

Tarantola, A., 2004, Inverse problem theory and methods for model parameter estimation: SIAM.

Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Sovet Math. Dokl., 1035–1038.

Trad, D., T. Ulrych, and M. Sacchi, 2003, Latest views of the sparse Radon transform: Geophysics, **68**, 386–399.

Valenciano, A. A., M. Brown, A. Guitton, and M. D. Sacchi, 2004, Interval velocity estimation using edge-preserving regularization, *in* 74th Ann. Internat. Mtg: Soc. of Expl. Geophys.

van der Vorst, H. A., 2003, Iterative Krylov methods for large linear systems: Cambridge University Press.

Versteeg, R., and W. W. Symes, 1993, Geometric constraints on seismic inversion: 63rd Ann. Internat. Mtg, Soc. of Expl. Geophys., 595–598.

Woodward, M. J., P. Farmer, D. Nichols, and S. Charles, 1998, Automated 3-D tomographic velocity analysis of residual moveout in prestack depth migrated common image point gathers: 68th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1218–1221.

Zhdanov, M. S., 2002, Geophysical inverse theory and regularization problems: Else-

vier Science.
Zhou, H., S. Gray, J. Young, D. Pham, and Y. Zhang, 2003, Tomographic residual curvature analysis: The process and its components: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 666–669.

APPENDIX A

CONJUGATE-GRADIENT ALGORITHM

A complete algorithm for conjugate-gradient iterative inversion with shaping regularization is given below. The algorithm follows directly from combining equation 13 with the classic conjugate-gradient algorithm of Hestenes and Steifel (1952).

Conjugate gradients with shaping$(\mathbf{L}, \mathbf{H}, \mathbf{d}, \lambda, tol, N)$

$\mathbf{p} \leftarrow \mathbf{0}$
$\mathbf{m} \leftarrow \mathbf{0}$
$\mathbf{r} \leftarrow -\mathbf{d}$
for $n \leftarrow 1, 2, \ldots, N$
  do
    $\mathbf{g}_m \leftarrow \mathbf{L}^T\,\mathbf{r} - \lambda\,\mathbf{m}$
    $\mathbf{g}_p \leftarrow \mathbf{H}^T\,\mathbf{g}_m + \lambda\,\mathbf{p}$
    $\mathbf{g}_m \leftarrow \mathbf{H}\,\mathbf{g}_p$
    $\mathbf{g}_r \leftarrow \mathbf{L}\,\mathbf{g}_m$
    $\rho \leftarrow \mathbf{g}_p^T\,\mathbf{g}_p$
    if $n = 1$
      then $\beta \leftarrow 0$
        $\rho_0 \leftarrow \rho$
      else $\beta \leftarrow \rho/\hat{\rho}$
        if $\beta < tol$ or $\rho/\rho_0 < tol$
          then return m
    $$\begin{bmatrix} \mathbf{s}_p \\ \mathbf{s}_m \\ \mathbf{s}_r \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{g}_p \\ \mathbf{g}_m \\ \mathbf{g}_r \end{bmatrix} + \beta \begin{bmatrix} \mathbf{s}_p \\ \mathbf{s}_m \\ \mathbf{s}_r \end{bmatrix}$$
    $\alpha \leftarrow \rho/\left[\mathbf{s}_r^T\,\mathbf{s}_r + \lambda\,(\mathbf{s}_p^T\,\mathbf{s}_p - \mathbf{s}_m^T\,\mathbf{s}_m)\right]$
    $$\begin{bmatrix} \mathbf{p} \\ \mathbf{m} \\ \mathbf{r} \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{p} \\ \mathbf{m} \\ \mathbf{r} \end{bmatrix} - \alpha \begin{bmatrix} \mathbf{s}_p \\ \mathbf{s}_m \\ \mathbf{s}_r \end{bmatrix}$$
    $\hat{\rho} \leftarrow \rho$
return m

The iteration terminates after N iterations or upon reaching convergence to the specified tolerance *tol*. It uses auxiliary vectors $\mathbf{p}$, $\mathbf{r}$, $\mathbf{s}_p$, $\mathbf{s}_m$, $\mathbf{s}_r$, $\mathbf{g}_p$, $\mathbf{g}_m$, $\mathbf{g}_r$ and applies operators $\mathbf{L}$, $\mathbf{H}$ and their adjoints only once per each iteration.

APPENDIX B

COMBINING SHAPING OPERATORS

General rules can be developed to combine two or more shaping operators for the cases when there are several features in the model that need to be characterized simultaneously. A general rule for combining two different shaping operators $\mathbf{S}_1$ and $\mathbf{S}_2$ can have the form

$$\mathbf{S}_{12} = \mathbf{S}_1 + \mathbf{S}_2 - \mathbf{S}_1\,\mathbf{S}_2\ , \tag{B-1}$$

where one adds the responses of the two shapers and then subtracts their overlap. An example is shown in Figure B-1, where an impulse response for oriented smoothing in two different directions is constructing from smoothing in each of the two directions separately.

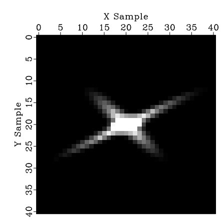

Figure B-1: Impulse response for a combination of two shaping operators smoothing in two different directions.

Combining two operators that work in orthogonal directions can be accomplished with a simple tensor product, as follows:

$$\mathbf{S}_{xy} = \mathbf{S}_x\,\mathbf{S}_y\ , \tag{B-2}$$

where $\mathbf{S}_x$ and $\mathbf{S}_y$ are shaping operators that apply in orthogonal x- and y-directions, and $\mathbf{S}_{xy}$ is a combined operator that works in both directions. An example is shown in Figure B-2, where two two-dimensional shapers working in orthogonal directions are combined to produce an impulse response of 3-D shaping operator that applies smoothing along a three-dimensional plane.

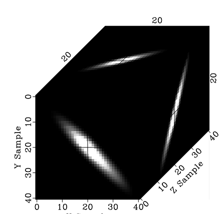

Figure B-2: 3-D impulse response for a combination of two 2-D shaping operators smoothing in in-line and cross-line directions.

Constructing multidimensional recursive filters for helical preconditioning (Fomel and Claerbout, 2003) is significantly more difficult. It involves helical spectral factorization, which may create long inefficient filters (Fomel et al., 2003).