Signal and Noise Separation: art and science

Tadeusz J. Ulrych *, Mauricio D. Sacchi † and Mike Graul ‡

* UBC, † U. of Alberta, ‡ Texseis, Inc.

Abstract

The separation of signal and noise is a central issue in seismic data processing. The noise is both random and coherent in nature, the coherent part often masquerading as signal. We present some approaches to signal isolation, in which stacking is a central concept. Our methodology is to transform the data to a domain where noise and signal are separable, a goal that we attain by means of inversion. We will illustrate our ideas with some of our favourite transformations, wavelets, eigenvectors, and Radon transforms. We end with the notion of risk, baseball and the Stein estimator.

INTRODUCTION

The purpose of this paper is to present some approaches to the ubiquitous problem of signal and noise separation that we have found useful. Let D represent the record of our seismic experiment in whatever form. D contains all the information that is at our disposal. Our task is to chisel away at D to unearth S. Of course, we must define S. We define signal as that energy that is coherent from trace to trace. Noise, N, on the other hand, is that energy that is incoherent from trace to trace. Unfortunately, the most expensive noise is also spatially coherent and we must modify the above definition. We define signal as that energy that is most coherent. This definition immediately introduces one of the central themes of this article, resolution. In order to extract the signal from the background of noise we must be able to resolve the difference, often very subtle and highly dependent on the acquisition, between the coherent signal and noise. We can now write the model

$$\mathbf{D} = \mathbf{S} + \mathbf{N}_c + \mathbf{N}_r \tag{1}$$

where $\mathbf{N}_c$ and $\mathbf{N}_r$ represent the coherent and incoherent noise components, respectively and $\mathbf{N}_c + \mathbf{N}_r = \mathbf{N}$.

METHODOLOGY

D lives in space and time, the $x - t$ domain. It is characteristic of this domain that S and N are intertwined and, consequently, are not only difficult to separate but also to identify. In order to accomplish these tasks, D must be mapped into a domain where the characteristics distinguishing signal and noise map S and N into separate spaces. In operator form

$$\mathcal{T}\mathbf{m} = \mathbf{d} \tag{2}$$

where $\mathcal{T}$ is the linear or nonlinear transformation, m is the vector of model parameters in the transformed domain and d is a data vector realization from D.
The transformation expressed by equation (2) is guided by a principle that is, to our minds, fundamental. The model m is non-uniquely related to the data. As such it may be recovered only when the inverse transformation is regularized.
Now comes the second part of our methodology. The manner of regularizing the transformation in equation (2), expressed in a form that emphasizes the inverse nature of the problem, follows Bayesian principles. We impose *a priori* information by means of Bayes' theorem, and in this manner obtain, from an infinity of possibilities, that model which honors the data and satisfies our prior beliefs.

PRINCIPLE OF STACKING

In much of what follows, the transformation in equation (2) is linked in some indirect, often obscure manner, to stacking, a concept that we return to at the end. Stacking involves a very special estimation of the first moment of a probability distribution. It is the maximum likelihood estimator, that we write as

$$\delta^0(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{3}$$

where the x_i are M samples of the random variable x. δ^0 is a very special estimator indeed. Statisticians have shown that, given $\mathbf{x}|\mu \overset{\text{ind}}{\sim} N(\mu, 1)$, δ^0 has lowest risk of any linear or nonlinear unbiased estimator, a characteristic that should surely appeal to all.
All of us have our favorite transformations. Here are a few of ours.

EIGENVECTOR

We deal briefly with two eigenvector decompositions that have proved to be of particular value.

Eigenimages

Consider our data D to be, as above, a $N \times M$ matrix comprising M traces and N points per trace. D admits the Lanczos decomposition

$$\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{M} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{4}$$

where U and V are the left and right eigenvector matrices, respectively and Σ is the matrix of singular values. In the summation representation, the σ_i are the individual, positive singular values so ordered that $\sigma_1 > \sigma_2 > \ldots > \sigma_m$. Equation (4) represents a projection of D onto an orthonormal basis, specifically a basis composed of weighted rank one matrices $\mathbf{u}_i\mathbf{v}_i^T$, called eigenimages. Clearly, the contribution of a particular eigenimage in the reconstruction of D is proportional to the magnitude of the associated singular value. It is often possible, depending of course on the data, to reconstruct the matrix D using only the first few eigenimages. For this reason, D is preprocessed prior to transformation so that the signal is as horizontal as possible. As a result, the signal exhibits maximum coherence from trace to trace, and often the first few eigenimages will contain the separated signal. Random noise is dispersed equally among all the eigenimages and thus, keeping only the first few eigenimages results in the attenuation of both N_r and N_c (please see Freire and Ulrych, (1988) for details). This process is very similar to that of stacking in the case where the noise is Gaussian and the signal is well aligned. The difference between eigenimage decomposition and stacking is, however, obvious when we consider AVO processing. In this case, stacking is forbidden, yet the first eigenimage attenuates noise as well as preserving the telltale amplitude variations.

Karhunen-Loéve (KL)

The KL transform was introduced in order to decorrelate the expansion coefficients, a_i, in the decomposition

$$\mathbf{x} = \sum_{i=1}^{N} a_i \mathbf{v}_i \tag{5}$$

where $\mathbf{v}_i$ are the eigenvectors associated with R, the Toeplitz autocovariance matrix of the process $x(t)$. Since the eigenvectors form an othonormal basis, the coefficients, a_i, are indeed uncorrelated.
In the example illustrated here, we adapt the KL to a multi-channel input by computing R as the average matrix for all traces. The filtering is accomplished by retaining only a subset of the computed eigenvectors. Results are shown in Figure 1 that considers the filtering of events with an AVO signature. Figure 1a and 1b show the input S and D, respectively. Figures 1c and 1d show the filtered output and the residual section. In this case, unlike the case of eigenimage filtering, NMO correction is not required prior to decomposition.

RADON

The signal in D can, to first order, be approximated by hyperbolic events. As such, we seek a mapping from $x - t$ to a domain where a hyperbolic event will map into a point. In principle, the Radon hyperbolic transform will achieve this end. It turns out that the hyperbolic form of the Radon transform, compared to its parabolic cousin, is very much more time consuming to implement (Hampson, 1986). One prefers, therefore, to transform via the parabolic form after a NMO correction to D. We define the parabolic Radon transform as a linear transformation from the data space into the parabolic Radon domain for a particular frequency. In matrix notation we have

$$\tilde{\mathbf{m}} = \mathbf{L}\mathbf{d} \tag{6}$$

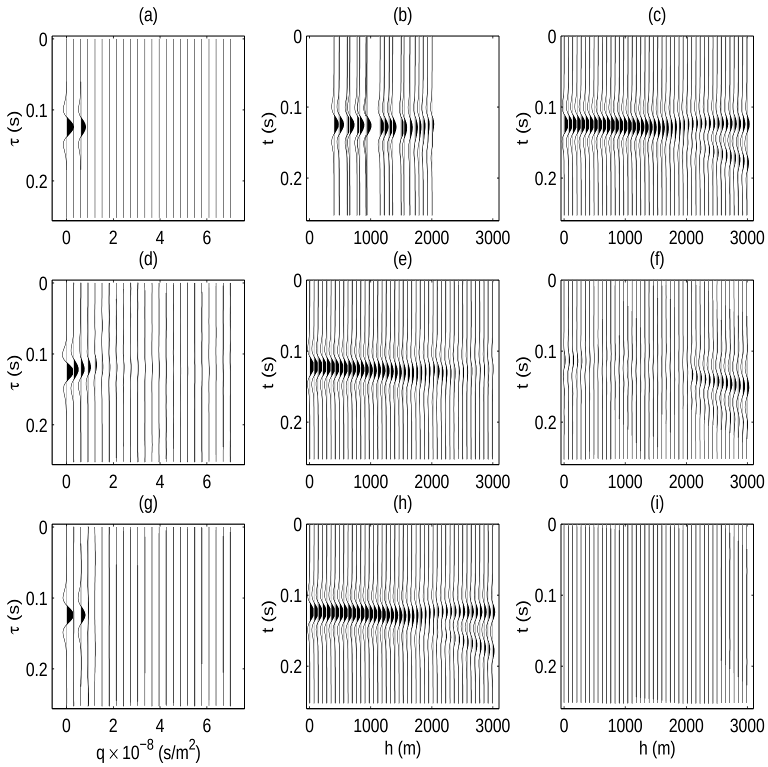

Figure 1: Radon transform and multiple attenuation

where m̃ and d are vectors that define the model space (the Radon domain) and the data (a common mid point or shot gather), respectively. We can say that L is an operator that stacks the data along parabolic paths. We have placed a ˜ over the m to indicate that the mapping according to equation (6) yields a low resolution or smooth model. In the stacking process, reflections with different moveout should collapse into points in the Radon space. Unfortunately, the operator L does not possess enough resolution to properly distinguish events with similar moveout. Instead of using the operator L to map reflections into the parabolic Radon domain, we prefer to pose the mapping as an inverse problem where now, d, the data, are viewed as the result of the transformation (Sacchi and Ulrych, 1995)

$$\mathbf{d} = \mathbf{L}'\mathbf{m} \tag{7}$$

$\mathbf{L}'$ is an operator that maps a point in the parabolic Radon space into a data parabola. The advantages of using equation (7) over equation (6) are twofold. First, by posing our problem as an inverse problem we can choose a strategy to enhance the resolution of the transformation. Second, by selecting the proper regularization scheme random noise can be attenuated.

We pose our inverse problem as one of inference and build our cost function by allowing prior information to guide us to a solution that, while honoring the data, exhibits the characteristics that we require. In order to incorporate our prior constraints, we use Bayes' theorem.

It is $p(\mathbf{m})$, the prior model p.d.f., that gives the solution that special flavor. We wish to impose sparseness or limited support that will simulate an extended aperture and allow us to identify and, hopefully attenuate, N $_c$. Choosing a Cauchy distribution for $p(\mathbf{m})$ allows us to attain this goal.

In Figure 2 we consider a model that consists of a primary and multiple event, shown in the $\tau - q$ domain in Figure 2a (the primary is the first event) This model is mapped into the $x - t$ domain in Figure 2c to simulate two reflections with parabolic moveout. The data are unevenly sampled and severely aperture-limited as shown in Figure 2b. Our task is summarized as follows: given the data in the $x - t$ domain (Figure 2b) recover the $\tau - q$ model and from the model reconstruct the evenly sampled and 'infinite' aperture data (Figure 2c). In Figure 2d we show the $\tau - q$ model computed using the least-squares approach and in Figure 2g the $\tau - q$ model retrieved using the Cauchy prior. It is clear that the Cauchy prior yields a result that is more consistent with the true model, whereas the least-squares approach fails to distinguish the positions of the two events.

Stein processing

We began this paper with some comments regarding the basics of stacking. We end on the same note. This section is aimed, primarily, for cerebral stimulation.
There is no doubt that, given a realization of M samples from $N(\mu,1)$, the only admissible estimator of μ is δ^0 defined in equation (3). The question is, is this also true if more than one realization is available? An illuminating example, from the scientific discipline of baseball, is given by Efron and Morris (1977). After the first 45 at bats in the 1970 season Thurmon Munson was in an early slump and managed only an average of .178. Faced with the question, what will the average be at the end of the season, the only admissible answer is .178. Munson, however, was batting with many other batters. James and Stein in the early 60's proved a very controversial theorem. An average estimated by δ^0 after forty at bats, given that there are more than two other batters in the league, is not an admissible estimator. In other words, Munson's average is better estimated taking into account what other batters are doing. Put in a different but equivalent manner, for this case δ^0 is not the estimator with lowest risk.
We are in the process of testing this method of stacking on seismic data. Specifically, our attempt is to improve the SNR in D without stacking for the purpose of, say, AVO analysis. We use the Stein estimator as a preprocessing step that allows us to shrink the noise variance in the section. Each batter becomes a time point associated with a trace. We thus consider M batters for each of the N points in the NMO corrected gather, compute the appropriate shrinkage locally and then the new sample value.

CONCLUSIONS
There is no manner of increasing the information content of the recorded D. All we can hope to do is to expose S by attenuating N. We attempt to do this by mapping D into a domain where, because of the different characteristics of S, N_c and N_r, such as coherence, bandwidth, fractal dimension, scale etc., we can separate these quantities into distinct subspaces and thereby attenuate the undesired components. In all cases, the forward and inverse mappings, derived on the basis of scientific principles, require some parameters to make them work. Setting these parameters is, in general, an art. It is these details that, according to Arthur Weglein, seldom see the light of publication.
We have briefly explored various techniques of signal to noise enhancement. Stacking is one such mapping that is, however, not always desirable, since it may destroy important signal attributes. Eigenimage and KL decomposition, Radon transformation, wavelet thresholding (another method that we explore) and Stein processing can lead to signal and noise separation without, we hope, signal distortion. REFERENCES

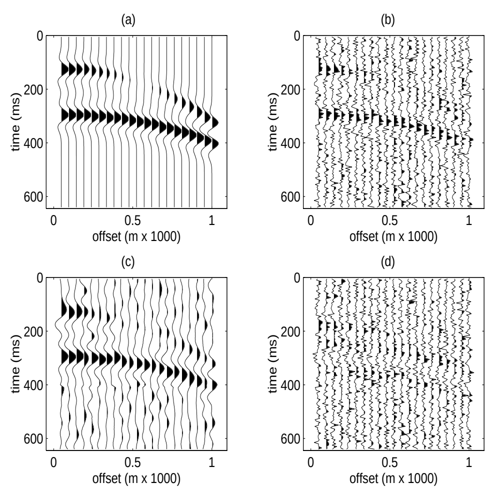

Figure 2: Karhunen-Loéve filtering

Efron, B., and Morris, C., 1977, Stein's paradox in statistics: Scientific American, May, 119–128.

Freire, S.L., and Ulrych, T.J., 1988, Application of singular value decomposition to vertical seismic profiling: Geophysics, 53, 778-785.

Hampson, D., 1986, Inverse velocity stacking for multiple elimination: J. Can. Soc. Expl. Geophys., 22, 44-55.

Sacchi, M.D., and Ulrych, T.J., 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, 60, 1169-1177.