Preserving signal: Automatic rank determination for noise suppression

Stewart Trickett, Absolute Imaging Inc.*

Summary

Matrix rank reduction filters such as Cadzow / SSA applied to constant-frequency slices have become popular for suppressing random noise on seismic data sets. A critical parameter for such methods is the matrix rank. A small rank gives strong noise suppression that may damage signal, and is best suited for noisy data and simple geological structures. A large rank gives weak noise suppression that preserves signal well, and is best suited for cleaner data and complex geological structures. Even within a single seismic survey, however, conditions change with time, space, and frequency, making a fixed rank inappropriate. Here I describe how the rank can be automatically determined for each matrix, allowing the filter to adapt to changing conditions. Examples are given on synthetic and real data. The result is an easy-to-use noise suppressor that finds a reasonable balance between signal preservation and noise removal throughout the section.

Introduction

Matrix rank reduction filters for removing random noise from constant frequency slices work as follows: beginning with a grid of uniformly spaced seismic traces, possibly in more than one spatial dimension, do the following:

```
Take the Discrete Fourier Transform (DFT) of each trace.
For each frequency of interest…
 {
    Form a complex-valued matrix from the frequency slice.
    Reduce the rank of the matrix.
    Reform the frequency slice from the rank-reduced
      matrix.
 }
Take the inverse DFT of each matrix.
```

The entire seismic data set is not filtered in one go. Rather the data is typically divided into overlapping tiles in time and space, each tile noise suppressed independently, and then the filtered tiles tapered and summed together to reform the complete data set. Typical dimensions of a tile might be 15 traces in each spatial direction and 400 ms in the time direction. Tiling ensures that each filtering operation is carried out on local data, limiting the number of dips present.

There are numerous ways to form the matrix from a frequency slice, resulting in methods such as eigenimage (Trickett, 2003), Cadzow or SSA (Trickett, 2002; Sacchi 2009), multi-dimensional Cadzow or MSSA (Trickett, 2008), and hybrid methods (Trickett and Burroughs, 2009).

Here's how rank reduction can be done. Assume for simplicity that **A** is a square nxn matrix, although rectangular matrices can also be handled. Its Singular Value Decomposition (SVD) is

$$\mathbf{A} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^H \qquad (1)$$

where **U** and **V** are nxn complex unitary matrices and **S** is an nxn real diagonal matrix with diagonal entries

$$s_1 \geq s_2 \geq \cdots \geq s_n \geq 0 \qquad (2)$$

called the *singular values* of matrix **A**. Reduce its rank to k by zeroing all but the first k singular values and then reforming the matrix from equation (1).

The most critical parameter is the matrix rank. The previously cited methods have the following property:

> *If a noiseless seismic section is the sum of plain waves having at most k distinct dips then filtering with rank k preserves the signal exactly.*

In practice, a small rank gives strong noise suppression that may damage signal, and is best suited for noisy data and simple geological structures. A large rank gives weak noise suppression that preserves signal well, and is best suited for cleaner data and complex geological structures. It is common to test various ranks on part of a seismic data set and then choose a rank that maximizes the noise suppression while still preserving signal. Difference plots can show when the filter is removing significant coherent energy, and thus has a rank which is too low.

Even within a single seismic survey, however, signal and noise conditions change with time, space, and frequency, making the use of a fixed rank throughout inappropriate. It would be best to vary the rank to adapt to changing conditions. Picking the rank manually, as one might pick stacking velocities for example, is expensive and processor dependent. It is preferable to automatically determine the rank for each matrix in a way that maximizes the amount of noise removed while still preserving signal. This talk describes such a method and demonstrates its performance on synthetic and real data.

Automatic rank determination has been proposed before in seismic processing. Harris and White (1997) did so using a Cadzow-like technique for improving f-x prediction filtering, based (interestingly) on the shape of the singular vectors rather than the magnitude of the singular values. Dack, Trickett, and Milton (2014) used automatic rank determination to remove noise from first arrivals. Finally, some tensor interpolation schemes automatically determine rank (e.g., Kreimer, Stanton, and Sacchi, 2013), although at the cost of adding other parameters requiring selection.

Method

Numerous methods have been proposed for automatic rank determination for a given matrix. I will work from papers by Gavish and Donoho (2014) and Donoho and Gavish (2014), which have references to previous rank-determination techniques.

The first step is to estimate the L largest singular values $s_{i,}, i = 1, \ldots, L$ and singular vectors of the matrix, where L is greater than the expected rank of the signal. Calculating the complete SVD is generally too expensive, particularly for multi-dimensional filtering where matrices are large. Instead I recommend using a partial Lanczos bidiagonalization followed by an SVD of the inner bidiagonal matrix (Trickett, 2003, Appendix A; Gao, Sacchi, and Chen, 2013).

The next step is to estimate the standard deviation σ of the noise. Gavish and Donoho (2014) suggest estimating it from the median singular value. Since we only have the first L singular values, however, I propose the following:

$\sigma = .7 s_q / \sqrt{\tilde{n}}$ where $q = 3L/4$ (an initial guess).

Iterate about 5 times…

{

$$E = \sum_{i,j} ||A_{i,j}||^2 + \sum_{i=1}^{L} (w_i - 2) w_i s_i^2$$

$$\sigma = \tilde{n} \sqrt{E/(mn-1)} / (\tilde{n} - \sum_{i=1}^{L} w_i)$$

}

where **A** is mxn, $\tilde{n} = \min(m,n)$, $\hat{n} = \max(m,n)$, $w_i = \eta^*(y_i)/y_i$, $y_i = s_i / (\sigma \sqrt{\hat{n}})$, and η^* is the operator-norm loss function from equation (5) of Donoho and Gavish (2014) with $\beta = \tilde{n}/\hat{n}$. This method iterates between using soft thresholding of the singular values to estimate the energy of the noise and thus σ, and using σ to estimate the degree of soft thresholding.

Finally, set the matrix rank k to the number of singular values exceeding

$$s^* = \sqrt{2(\beta+1) + \frac{8\beta}{\beta+1+\sqrt{\beta^2+14\beta+1}}} \; \sigma \sqrt{\hat{n}} . \quad (3)$$

For squares matrices ($\beta = 1$) this is simply $4/\sqrt{3}\ \sigma \sqrt{\hat{n}}$. Gavish and Donoho (2014) prove that s^* minimizes the Frobenius norm between the pure signal matrix and the rank-reduced noisy matrix as $m, n \to \infty$. They also empirically verify that it works well for relatively small matrices.

Although this is a theoretically satisfying approach, in extreme noise it can damage signal. Sometimes the noise is so strong that all singular values (and thus the frequency slice) are zeroed, even when signal is apparent in the input data. There are countless ways to avoid this. I propose that we ensure that at least some singular components are preserved by limiting the threshold

$$s^* = \min\ (cs_1,\ s^*) \quad (4)$$

where c is some value less than 1 (e.g., .75). I'll refer to this as *threshold capping*. On clean and moderately noisy frequency slices, threshold capping has little or no effect.

Examples

Let's test this on a synthetic example. Our model is in two spatial dimensions (Figure 1). In the horizontal direction, the events are flat on the left and become progressively curving on the right. In the vertical direction the model is flat, but the amplitude of bandpassed random Gaussian noise increases from top to bottom. Three horizontal slices are analyzed: clean, moderately noisy, and very noisy.

Figure 2 shows the automatically determined rank averaged over 15 to 70 Hz. As expected, the rank increases as the section becomes cleaner and more structurally complex. This is exactly the behavior we seek.

Figure 3 shows the three horizontal slices unfiltered, and with f-xy Cadzow filtering (Trickett, 2008) using a fixed rank of 3, a fixed rank of 5, and threshold-capped automatic rank determination. Overlapping 15 x 15 trace tiles were used for each filter.

Figure 4 shows the difference between the clean signal and these filtered noisy sections. Automatic rank determination does the best overall job of preserving signal, and in addition suppresses more noise than the rank-5 filter in the flatter regions.

Figure 5 shows portions of NMO-corrected cross-spread gathers from a real 3D data set. In the simpler region with only a few dips, f-xy Cadzow filtering with automatic rank determination behaves like a rank-3 filter, giving more aggressive noise suppression than rank 6. In the more complex region with many conflicting dips, automatic rank determination behaves like a rank-6 filter, preserving coherency better than rank 3.

Conclusions and Future Work

I have described a method of automatic rank determination which allows noise suppression filters to adapt to changing geological and noise conditions throughout a section. Although based on Gavish and Donoho (2014), their method could not be used straight out of the box. Instead I had to address three problems: (1) avoiding the expense of calculating a full SVD, (2) estimating the noise level when the full SVD is unavailable, and (3) making the filter less harsh for extreme noise in order to preserve signal. The result is an easy-to-use noise suppressor that finds an appropriate balance between signal preservation and noise removal throughout the section.

Donoho and Gavish (2014) also describe soft thresholding schemes, where singular values are scaled rather than accepted or rejected. In theory this is superior to hard thresholding, and has the additional benefit of making the output continuously dependent on the input. I have not addressed these schemes here for the sake of brevity, but to summarize I have found that the less aggressive criteria such as minimizing the L2 operator norm give results similar to hard thresholding, while the Frobenius or (especially) nuclear norms tend to damage complex structure in ways that are noticeable in difference plots. I may address this in future papers.

Acknowledgements

Thanks to Apache Corporation and NuVista Energy Ltd. for permission to show their data. Also thanks to the crew at Absolute Imaging Inc. in Calgary for all of their help.

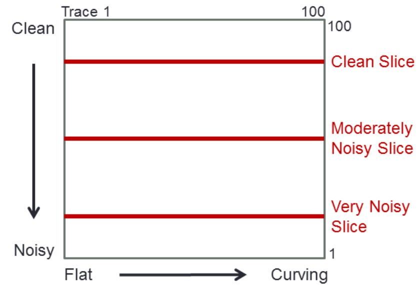

Figure 1: 100 x 100 synthetic trace grid. Event curvature increases from left to right, and noise amplitude increases from top to bottom. Three horizontal trace slices (shown in red) are analyzed.

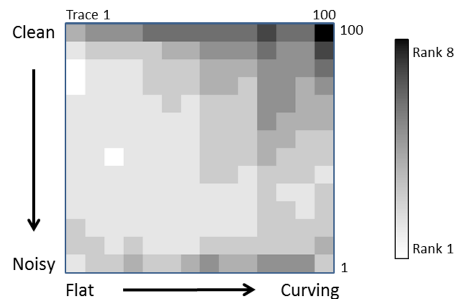

Figure 2: Automatically determined rank averaged over 15 to 70 Hz. The rank increases as the section gets cleaner and more structured. The effects of threshold capping are seen at the bottom.

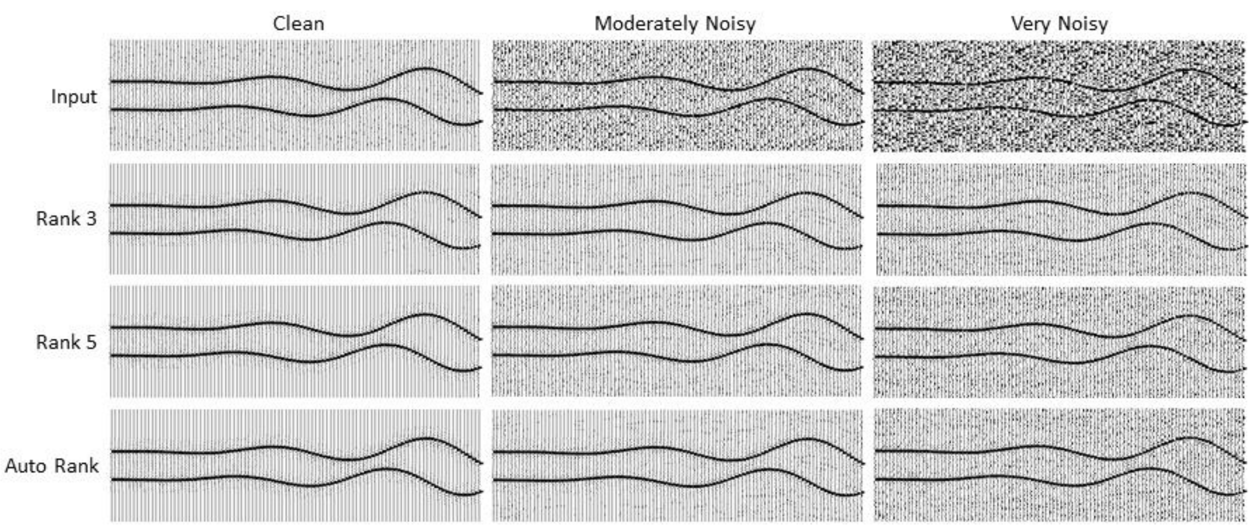

Figure 3: Synthetic tests 400 ms in the time direction. The three raw horizontal slices are shown along the top. Below this are f-xy Cadzow filtered versions with rank 3, rank 5, and automatic rank determination with threshold capping.

Automatic Rank Determination

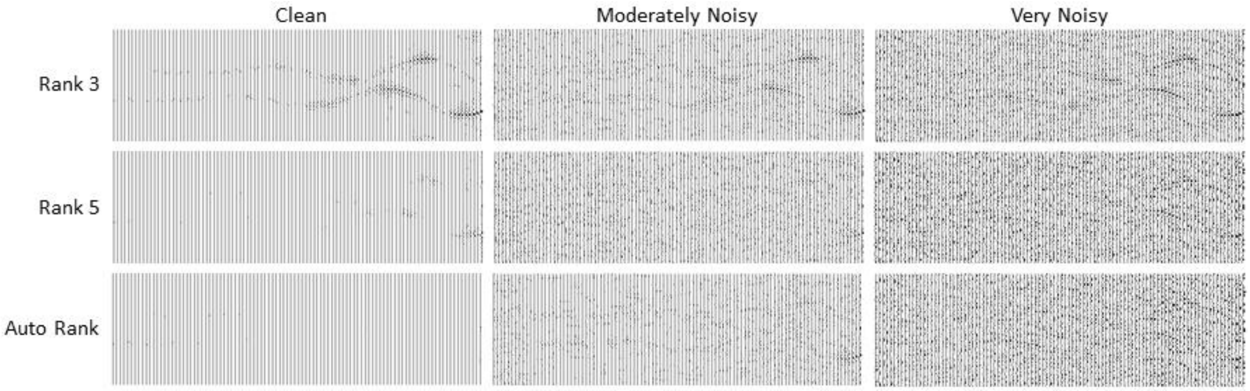

Figure 4: The difference between the noiseless signal and the various filtered versions of the noisy data from the previous figure. Automatic rank determination has removed the least amount of signal while suppressing more noise than the rank-5 filter in the flatter regions.

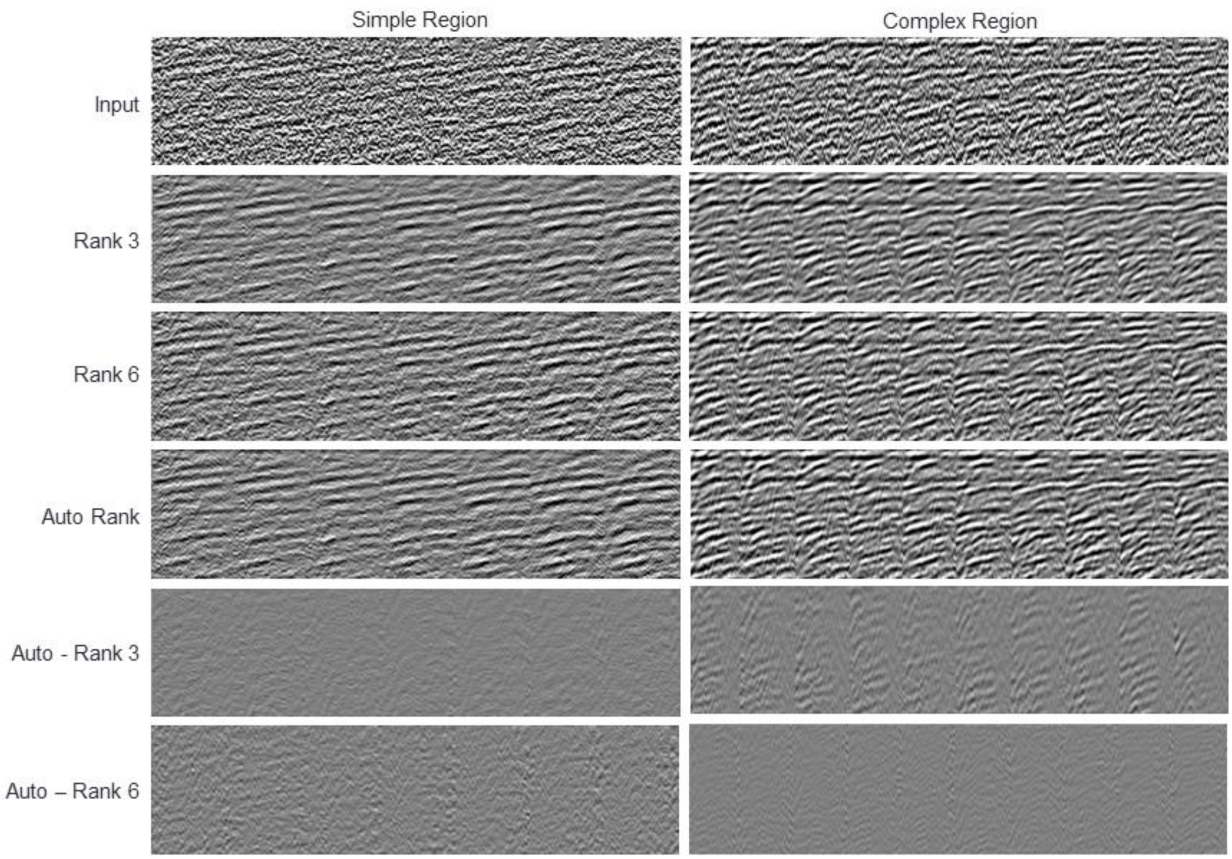

Figure 5: Two different regions of NMO-corrected cross-spread gathers from a real 3D data set. In the simpler region on the left, f-xy Cadzow with automatic rank determination behaves like a rank-3 filter, giving more aggressive noise suppression than rank 6. In the more complex region on the right, automatic rank determination behaves like a rank-6 filter, preserving coherency better than rank 3.

References

Dack, R., S. Trickett, and A. Milton, 2014, Cleaning up noisy first arrivals: SEG Post Convention Workshop – Towards robust first arrival picking.

Donoho, D. L. and M. Gavish, 2014, Optimal shrinkage of singular values: Dept. Statistics, Stanford University, Technical Report 2014-8.

Gao, J., M. Sacchi, and X. Chen, 2013, A fast rank reduction method for the reconstruction of 5D seismic volumes: Geophysics, **78**, V21-V30.

Gavish, M., and D. L. Donoho, 2014, The Optimal Hard Threshold for Singular Values is 4/sqrt(3): IEEE Transactions on Information Theory, **60**, 5040-5053.

Harris, P. E., and R. E. White, 1997, Improving the performance of f-x prediction filtering at low signal-to-noise ratios: Geophysical Prospecting, **45**, 269-302.

Kreimer, N., A. Stanton, and M. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, V273-V284.

Sacchi, M., 2009, FX singular spectrum analysis: CSPG CSEG CWLS Convention, Abstracts, 392–395.

Trickett, S., 2002, F-x eigenimage noise suppression: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2166–2169.

Trickett, S. R., 2003, F-xy eigenimage noise suppression: Geophysics, **68**, 751–759.

Trickett, S., 2008, F-xy Cadzow noise suppression: 78th Annual International Meeting, SEG, Expanded Abstracts, 2586–2590.

Trickett, S., and L. Burroughs, 2009, Prestack rank-reduction-based noise suppression: CSEG Recorder, **34**, no. 9, 24–31.