Singular Spectrum Analysis via Optimal Shrinkage of Singular Values

Mehdi Aharchaou and John E. Anderson, ExxonMobil Upstream Research Company; Steve Hughes and John Bingham, ExxonMobil International Limited

Summary

Singular Spectrum Analysis (SSA) is a widely used time series analysis technique. For seismic processing, SSA has proved to be highly effective at suppressing random (incoherent) noise, interpolating gaps, or also reconstructing randomly distributed traces. In its standard form, SSA is based on the truncated singular value decomposition (TSVD) of Hankel matrices for which we assume the rank to be known. We ask: is this the best rank reduction strategy? To answer this question, we exploit recent results from the field of low rank matrix denoising to improve the performance of SSA via optimal shrinkage of singular values. Examples are given on synthetic and field data. The result is a simple, completely data-driven algorithm that addresses successfully the trade-off between signal preservation and noise suppression.

Introduction

Given a multi-dimensional grid of traces, the SSA algorithm (Sacchi, 2009 and Oropeza *et al.*, 2011) works on patches as follows:

1. Apply the Fourier transform to each trace
2. For each frequency independently …
   2.a. Form a Hankel matrix (**H**) based on the frequency slice (process explained in figure 1)
   2.b. Reduce its rank
   2.c. Reform frequency slice by anti-diagonal averaging
3. Apply the inverse Fourier transform to each trace

Algorithm 1: SSA

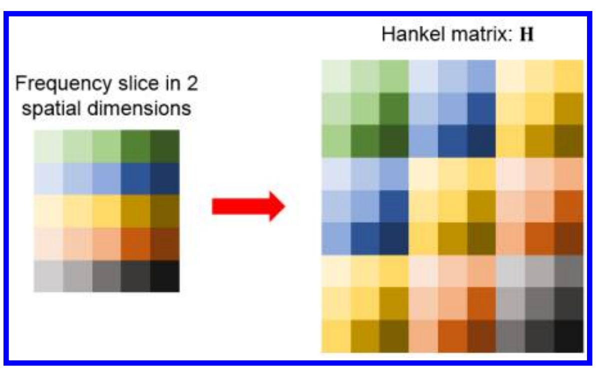

Figure 1 Block-Hankel matrix formed based on a frequency slice (process also called *Hankelization* or *embedding*). For two spatial dimensions, **H** is a Hankel matrix of Hankel matrices. In three spatial dimensions, **H** is a Hankel matrix of Hankel matrices of Hankel matrices, *etc*…. The size of matrix **H** grows much larger than the corresponding frequency slice as more spatial dimensions are included.

Regardless of how many spatial dimensions, **H** is always a 2D matrix that is nearly square (by design). We assume it here to be of size *m*-by-*n*, and take $p = \min(m, n)$. In the noise-free case, it can be rigorously shown that if the input grid of traces contains *r* distinct dips (linear move-outs), then **H** is *exactly* of rank *r*. In the noisy case, we assume that **H** contains coefficients which are related to a random "noise" matrix, for which we denote σ the standard deviation of the noise, and coefficients related to a "signal" matrix **X** that we assume to be non-random and of rank *r*. In this sense, **H** can be viewed as the "signal-plus-noise" matrix.

The SSA problem can be stated as: *We wish to recover* ***X*** *from* ***H*** *with some improvement in the mean square error (MSE).*

In the SSA problem, a standard and widely used strategy to reduce the rank of **H** (step 2.b. in Algorithm 1) is based on TSVD: write the singular value decomposition $\mathbf{H} = \sum_{i=1}^{p} h_i \mathbf{u}_i \mathbf{v}_i'$, where $\mathbf{u}_i \in \mathbf{C}^m$ and $\mathbf{v}_i \in \mathbf{C}^n$ are the left and right singular vectors of **H** corresponding to the observed (noisy) singular values h_i for $i = 1, \dots, p$. The TSVD solution when r is known is $\hat{\boldsymbol{X}}_r = \sum_{i=1}^{r} h_i \mathbf{u}_i \mathbf{v}_i'$. One can think of this as simply hard thresholding, where we keep (untouched) singular values larger than h_r and set to 0 the rest of singular values.

The relationship between the singular values of **X** (the noise-free matrix) and those of **H** (the noisy matrix) is a non-trivial one (Benaych-Georges *et al.*, 2012). One can think of the observed singular values $\{h_i\}_{i \in \{1,\dots,p\}}$ as being "inflated" by the noise (see top part in figure 2). Hence some type of "shrinkage" of these singular values is needed to recover **X** properly. Moreover, the optimal threshold to cut out the observed (noisy) singular values is not necessarily *r* (as shown in Nadakuditi *et al.*, 2013).

Why TSVD is not necessarily the best choice:

The TSVD solves the problem of finding the best rank *r* approximation of **H** in the least squares sense. It does not solve the fundamental denoising problem of how to best recover a low rank matrix buried in noise (*i.e.* estimate **X** from **H)**, even though users of SSA sometimes invoke it as though it does. Thus, we should not expect TSVD to be the best solution to the SSA problem.

In practice, rank *r* is assumed to be known and is fixed to be constant for all frequency slices and across the entire survey. This can lead to TSVD-based SSA either under- or over-performing depending on how complex the geology is, and this even after a lot of trial and error in searching for the best rank r. Hence, an *automated* treatment of the singular values is needed to leverage the performance of SSA and properly manage the tradeoff between signal preservation and noise suppression.

These arguments motivated the investigation of strategies other than TSVD for SSA. We present a simple-to-use and completely data-driven algorithm that improves the MSE performance of SSA and makes this technique more automatic.

Previous efforts exist that attempt to improve rank determination for rank reduction methods, and they are based on TSVD. Trickett (2015) proposed to automatically estimate the rank by computing the standard deviation of the noise from the data in an iterative scheme. A rank selection strategy was proposed by Porsani *et al* (2012) and Chen *et al.* (2016) based, interestingly, on local dip decomposition. Kreimer *et al.* (2013) proposed tensor completion for seismic trace reconstruction, and their method automatically estimates the rank but at the expense of some tuning parameters. Finding more automated ways with minimal user intervention to estimate the rank, which is a parameter of fundamental importance in all rank reduction techniques and not just SSA, is clearly needed.

Method

We use results in Gavish and Donoho (2014) and Nakaduti (2013) who deal with low rank matrix recovery from noisy observations. They proposed, in an asymptotic framework where matrix size (=*mn*) is large compared to its rank (=*r*), an optimal hard threshold and an optimal shrinkage technique of singular values (called *OptShrink*) that work under reasonably broad conditions. For SSA, the asymptotic framework is particularly relevant as we have more and more spatial dimensions in the input grid of traces (this increases the size of the Hankel matrix **H**, as depicted in figure 1).

The proposed method aims to improve step 2.b. in Algorithm 1. We first compute the singular value decomposition of **H**. Even though the cost of computation can be prohibitive for large matrices, we find that a divide-and-conquer implementation of SVD is reasonably fast (Gu *et. al.*, 1994). Note that in this method we need all p singular values. Parameter $\beta = \frac{\min(m,n)}{\max(m,n)}$ characterizes the shape of matrix **H**.

Step 1 *(optional)* Estimate an (effective) rank $\hat{r}$ of X based on the hard threshold: $\boldsymbol{\tau_H = \omega(\beta)\, h_{med}}$ (Gavish and Donoho, 2014)

This threshold assumes no knowledge of the noise level σ. It relies on the median singular value h_{med} (derived from the set of "noisy" singular values $\{h_i\}_{i\in\{1,\dots,p\}}$) to adapt to the noise present in **H**. Obviously, h_{med} is not fixed and varies with each frequency slice in the SSA algorithm. The constant $\omega(\beta)$ can be approximated without much loss in precision with the cubic polynomial $\omega(\beta) = .56\beta^3 - 0.95\beta^2 + 1.82\beta + 1.43$. We estimate the effective $\hat{r}$ as the number of singular values in $\{h_i\}_{i\in\{1,\dots,p\}}$ that exceed τ_H. This estimation is very simple to implement and requires no tuning parameters.

Step 2 Shrink singular values based on 'OptShrink' (Nakaduti, 2013)

The estimation of $\hat{r}$ enables the following algorithm to select (roughly) the 'signal portion' of the singular value spectrum and the 'noise portion'. Note that $\hat{r}$ does not need to be exact.

Given $\hat{r}$ and the SVD of **H** ($= \sum_{i=1}^{p} h_i \mathbf{u}_i \mathbf{v}_i'$), $p = \min(m,n)$
1. Compute $\mathbf{\Sigma}_{\hat{r}} = diag(h_{\hat{r}+1,\dots,h_p})$ ('noise portion')
2. For $i = 1, \dots, \hat{r}$
   2.a. Compute (scalars) $D(h_i, \mathbf{\Sigma}_{\hat{r}})$ and $D'(h_i, \mathbf{\Sigma}_{\hat{r}})$
   2.b. Compute $w_i = -2\, D(h_i, \mathbf{\Sigma}_{\hat{r}})/D'(h_i, \mathbf{\Sigma}_{\hat{r}})$
3. Return $\sum_{i=1}^{\hat{r}} w_i \mathbf{u}_i \mathbf{v}_i'$ (denoised estimate of **H**)

Algorithm 2: OptShrink

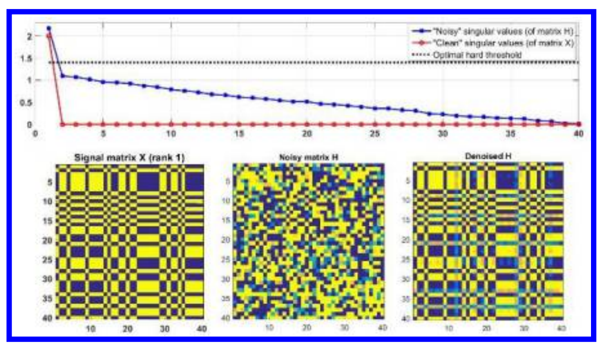

Figure 2 Top: comparison between singular values of **H** (noisy matrix) and **X** (signal matrix). The optimal hard threshold τ_H is represented by the dashed black line. Bottom, from left to right: 40-by-40 "signal" matrix **X** (of rank 1), "signal-plus-noise" matrix **H** (after adding lots of random noise), denoised **H** based on OptShrink. The denoising has successfully recovered the main features in **X**.

$D(h_i, \mathbf{\Sigma}_{\hat{r}})$ and $D'(h_i, \mathbf{\Sigma}_{\hat{r}})$ are scalars and they can be computed in a straightforward way using (16a) and (16b) in Nakaduti (2013). Algorithm 2 (OptShrink), embedded inside Algorithm 1 (SSA), yields a completely data-driven algorithm (we call OptShrink SSA), and uses *information* provided by the $(min(m,n) - \hat{r})$ singular values in the 'noise portion' to re-weight the singular vectors in the 'signal portion' in a way that minimizes the MSE. In an asymptotic framework, where the size of **H** is much larger than its rank (which is equivalent to the signal of interest being *sparse*), it is clear that the number of singular values in the 'noise portion' is usually high, meaning these values can provide information to *better guide* our separation of the signal from the noise. An impressive feature of this algorithm, as explained in Nakaduti (2013) and as we tested on several examples, is that it *largely mitigates the effect of over-estimating rank* r.

Examples

We present examples on synthetic and field seismic data, contrasting the performance of the two versions of SSA:TSVD and OptShrink.

Figure 3 shows a 2D gather containing two plane waves with severe random noise. SSA based on OptShrink with no rank specification gives a cleaner output than TSVD-based SSA with rank 2 (the correct rank value, equal to the number of distinct dips). One may wonder why OptShrink is superior even though we use the *correct* rank value for TSVD-based SSA? There are two reasons for that: 1) OptShrink SSA does not just keep the two first singular values as they are, but it shrinks them to counteract the effect of the noise, 2) In frequencies where noise dominates signal, the effective rank estimated by the optimal threshold (step 1) is in this case 1 (instead of 2). This enabled more noise suppression at those frequencies where signal is almost non-existent.

Figure 4 is based on a migrated 3D stack. SSA based on TSVD is compromised in this case: with a low rank, it performs well for the flat geology but loses dipping events in the more complex zones; with a high rank, it performs better in the complex zones but leaves noise in the simple geology. A rank value of 7 was the best choice in this case (value confirmed after trial and error). The proposed algorithm based on the optimal threshold and OptShrink required no specification of rank. The denoised stack based on the proposed method is cleaner with very little coherent energy showing in the difference (noise model). Checking the noise energy levels on power spectra, the gain in denoising from OptShrink SSA is about +3 dBs relative to SSA based on TSVD.

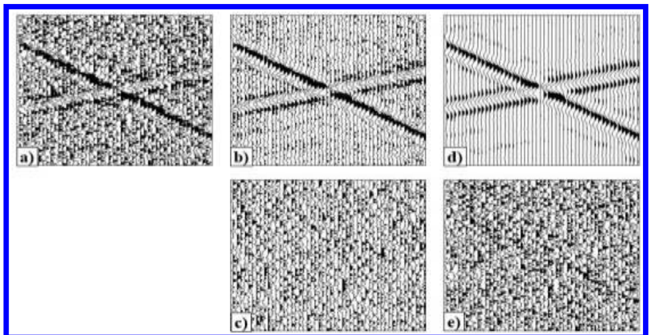

Figure 3: a) 2D gather made of two linear events and random noise, b) denoised using SSA based on TSVD (rank=2), c) difference between a) and b), d) denoised using SSA based on proposed method (no rank specification), e) difference between a) and d).

Conclusions

We have described a simple-to-use, completely data driven method of optimal selection-and-shrinkage of singular values that improves the SSA algorithm for random noise suppression. Even though it was not the focus here, the proposed algorithm can be used iteratively for better 'gap filling' and random trace interpolation of seismic volumes. OptShrink SSA showed to *always be superior* to TSVD, and it shines in that it can *largely mitigate* the effect of over estimating the rank. This makes the SSA algorithm more *automated* with the patch size as the only remaining parameter to be specified. Since our method does not exploit the Hankel structure, it can be used to improve other rank reduction techniques such as the method of Eigen Images or also the Karhunen-Loeve transform. We may address this in future papers.

Acknowledgments

The authors would like to thank Exxon Mobil Corporation for permission to publish this work.

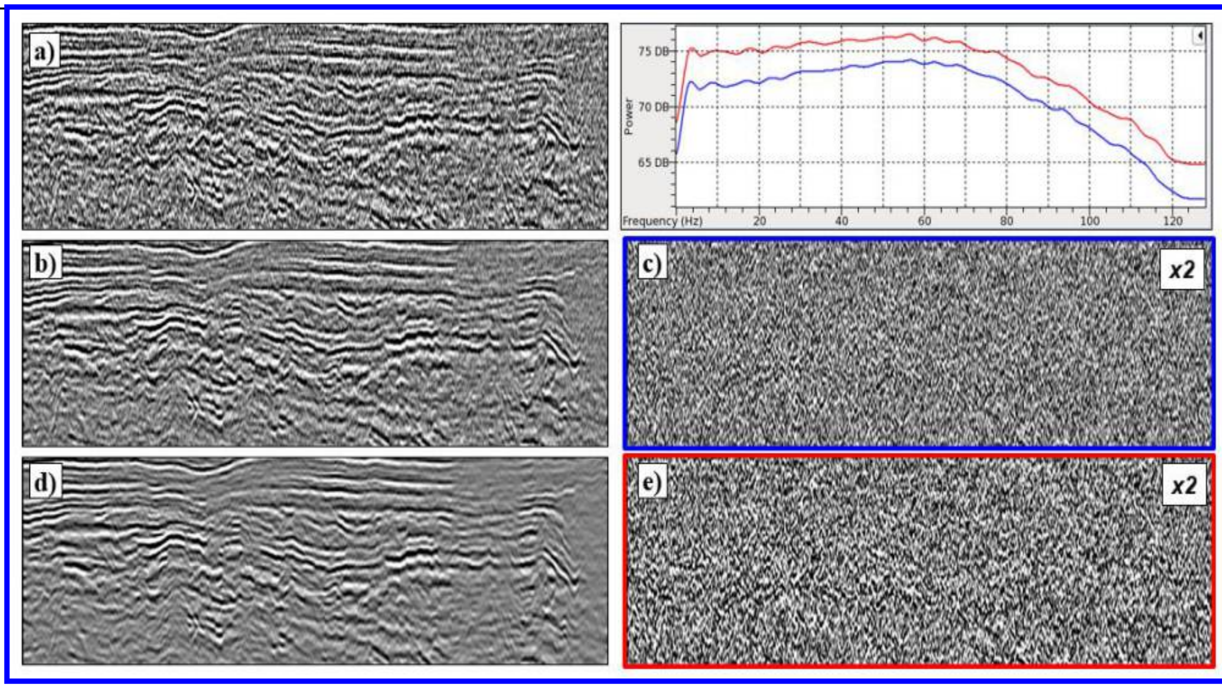

Figure 4: a) Cross-line slice of 3D stack exhibiting some random noise, b) denoised using 3D SSA based on TSVD (rank=7), c) difference magnified by a factor of 2 between a) and b), d) denoised using 3D SSA based on proposed algorithm (no rank specification), e) difference magnified by a factor of 2 between a) and d). All denoised outputs were generated based on patches of size 15-by-15 traces in the two spatial dimensions (inline and cross-line).

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Benaych-Georges, F., and R. R. Nadakuditi, 2012, The singular values and vectors of low rank perturbations of large rectangular random matrices: Journal of Multivariate Analysis, **111**, 120–135.

Chen, Y., W. Chen, S. Zu, W. Huang, and D. Zhang, 2016, Adaptive rank selection for singular spectrum analysis via dip decomposition: 86th Annual International Meeting, SEG, Expanded Abstracts, 4808–4813, https://doi.org/10.1190/segam2016-13712677.1.

Gavish, M., and D. Donoho, 2014, The optimal hard threshold for singular values is 4/(√3): IEEE Transactions on Information Theory, **60**, 5040–5053.

Gu, M., J. Demmel, and I. Dhilon, 1994, Efficient computation of the singular value decomposition with applications to least squares problems: Technical Report LBL-36201, October 1994.

Kreimer, N., A. Stanton, and M. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, no. 6, V273–V284, https://doi.org/10.1190/geo2013-0022.1.

Nadakuditi, R. R., 2013, Optshrink-low-rank signal matrix denoising via optimal, data-driven singular value shrinkage: IEEE Transactions on Information Theory, **60**, 3002–3018, https://doi.org/10.1109/TIT.2014.2311661.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, no. 3, V25–V32, https://doi.org/10.1190/1.3552706.

Porsani, M. J., B. Ursin, M. G. Silva, and P. E. M. Melo, 2012, Dip-adaptive singular-value decomposition filtering for seismic reflection enhancements: Geophysical Prospecting, **61**, 42–52, https://doi.org/10.1111/j.1365-2478.2012.01059.x.

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

Trickett, S., 2015, Preserving signal: Automatic rank determination for noise suppression: 85th Annual International Meeting, SEG, Expanded Abstracts, 4703–4707, https://doi.org/10.1190/segam2015-5873373.1.