Rank-Reduction Filtering
In Seismic Exploration

Stewart Trickett
Calgary, Alberta, Canada
Mauricio Sacchi + SAIG consortium
University of Alberta
Edmonton, Alberta
2008 to present

Felix Herrmann + SLIM consortium
University of British Columbia
Vancouver, British Columbia
2012 to present

Myself + others
Calgary, Alberta
2001 to present
Topics

1. Seismic exploration
2. Rank-reduction filtering on constant-frequency slices
3. Cadzow / SSA filtering
4. Multidimensional filtering
5. Computational speed
6. Robust filtering
7. Automatic rank determination
8. Interpolation
9. Dealiasing interpolation
10. Tensor Interpolation
1. Seismic Exploration
Seismic exploration uses generated sound waves traveling within the earth to locate and develop petroleum and other mineral resources.

$15$ billion / year industry.
A time series recorded by a single receiver from a single source is called a **trace**.
Marine Exploration
Number of recorded traces in a single seismic survey...

10 million to 100 billion. 
**Trillion** trace surveys being considered.

Single survey covers an area between 5 sq km to 10,000 sq km.
A recorded seismic trace has *four spatial dimensions* describing its geographical location...

After massive computation, traces are transformed into an image of the subsurface in two spatial dimension and in time.
(client seismic slides removed)
2. Rank-Reduction Filtering On Constant-Frequency Slices
Suppose we have a grid of seismic traces in any number of spatial dimensions...

Take the Discrete Fourier Transform (DFT) of every trace.
For every frequency of interest...

{  
  Form a complex-valued trajectory matrix from the constant-frequency slice (somehow).
  
  Reduce its rank.
  
  Place the elements of the matrix back into the frequency slice (any repeated elements are averaged).
}

Take the inverse DFT of every trace.

*Used for ... Random noise suppression.*
*Trace interpolation.*
Given a noiseless multi-dimensional grid of traces made up of the sum of plane waves having at most k distinct dips (slopes), rank-reduction filtering with rank k preserves the signal exactly. 

(at least for the trajectory matrices we will consider)

Only holds when filtering on constant-frequency slices, not when filtering in the time domain!
Why does the exactness property hold?

1. A single plane wave becomes a complex multi-dimensional sinusoid on each frequency slice.

2. A complex sinusoid has rank 1 in all of the trajectory matrices and tensors we will consider.

These filters extract a few dominant sinusoids from a noisy background.

It takes two ranks to model a real sinusoid, one rank to model a complex sinusoid.
Typical tile dimensions: 0.4 s in time, 15 traces in each spatial direction. Reduces number of dips (and thus the signal rank).
3. Cadzow / SSA Filtering
Cadzow / SSA Filtering

Complex frequency slice in 1 spatial dimension

Hankel Matrix

Example In 1 Spatial Dimension

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
</tr>
</thead>
</table>
| Time  | ![Raw Time](image)
| Difference | ![Difference Rank 1](image) | ![Difference Rank 2](image) | ![Difference Rank 3](image) |
Take the Discrete Fourier Transform (DFT) of every trace. 
For every frequency of interest...

\{ 
  
  Form a complex-valued trajectory matrix from the constant-frequency slice.
  
  Reduce its rank.
  
  Place the elements of the matrix back into the frequency slice (any repeated elements are averaged).

\} 

Iterate

Take the inverse DFT of every trace.

Cadzow / SSA?

Q: Would finding the rank-k Hankel matrix closest to the original give a better noise suppression?
4. Multidimensional Filtering
Raw seismic traces have *four spatial dimensions*.

Filtering in many spatial dimensions at once greatly increases the power of the filter. We can remove far more noise while preserving signal.

*How do we extend rank-reduction filtering to many dimensions?*
Multidimensional Embedding

2D frequency slice

Multidimensional Cadzow / SSA (2008)

Unstructured (2003)

Hybrid (2009)
Properties

Mild noise suppression
Trace grid need not be uniformly spaced
Can handle random x- and y-consistent time shifts in the seismic traces

Unstructured

Strong noise suppression
Trace grid must be uniformly spaced
Can’t handle but random time shifts

Cadzow / SSA

Medium noise suppression
Those dimensions treated as unstructured
have same properties as unstructured

Hybrid
Two Spatial Dimensions

- Signal (2 spatial dimensions)
- Signal + Noise
- Prediction Filtering
- Projection Filtering
- Cadzow / SSA Filtering
Two Spatial Dimensions

Raw (two spatial dimensions)

Cadzow / SSA  Difference
Two Spatial Dimensions

- Raw (2 spatial dimensions)
- Hybrid
- Cadzow / SSA
Three Spatial Dimensions

Raw
(3 spatial dimensions)

Hybrid

Cadzow / SSA
5. Computational Speed
Approximating The SVD

Classic method of matrix rank reduction is the **Truncated Singular Value Decomposition** (TSVD).

*Extremely expensive* for multidimensional filtering.

TSVD approximated by a **partial Lanczos bidiagonalization**:

\[
\text{TSVD}_k(A) \approx P B Q^H
\]

- **Complex matrices with** $k$
  - **orthonormal columns**
- **Real $k \times k$**
  - **bidiagonal matrix**
Partial Lanczos Bidiagonalization captures the largest singular values, but also some *small (spurious) singular values*.

Can remove spurious singular values by…

1. Calculate the rank \( k+r \) partial bidiagonalization, where \( r \) is some small integer.

2. Perform TSVD on the inner bidiagonal matrix \( B \) to reduce it to rank \( k \).

Can prove you get a better approximation to the rank-\( k \) TSVD as \( r \) increases (in exact arithmetic).
Lanczos diagonalization requires many \textit{matrix-vector products}.

Hankel matrix is part of a \textit{cyclical convolution} matrix. Can do matrix-vector product using \textit{Discrete Fourier Transforms}. Extends to multiple spatial dimensions using a \textit{multidimensional DFT}.

Summing along the anti-diagonals is a \textit{non-cyclical convolution} of the singular vectors, and can also be done using multidimensional DFTs.

Want to minimize padding in multi-dimensional DFTs. Need a Fast Fourier Transform for \textit{small arbitrary lengths}. \textit{FFTW} isn’t quite up to the task (e.g., 19, 23, 29). \textit{James Von Buskirk’s algorithms} come to the rescue.

http://home.comcast.net/~kmbtib/index.html

Q: Apart from Monte Carlo methods, are there faster ways to perform rank reduction?
6. Robust Filtering
Erratic Noise

Truncated SVD is a least-squares fit, well suited for Gaussian noise.

In seismic data, noise is often erratic due to:

Air blast, powerline and other cultural noises, recording and parity errors, uncorrected polarity reversals, isolated noise bursts (often due to deconvolved spikes, zeroes, and clips), misfired shots, scattered shot noise, poor surface conditions, disabled or poorly coupled geophones, wind, rain, simultaneous shooting…

Need a statistically robust version!
Iterative Winsorization

Let $S = \text{Raw frequency slice}$

$T = \text{Filtered frequency slice}$

$T = S$

Iterate until $T$ stops changing

$\{$

$\quad T = \text{Weighted sum of } S \text{ and } T$

$\quad \text{Rank reduction filter } T$

$\}$

$T = \text{Weighted sum of } S \text{ and } T \text{ (optional)}$

Remove erratic noise only
Let $S$ = Raw frequency slice
$T$ = Filtered frequency slice

Then the weighting is...

$$T_i = w_i S_i + (1 - w_i) T_i$$

where

$$w_i = \begin{cases} (1 - u_i^2)^2 & u_i < 1 \\ 0 & \text{otherwise} \end{cases}$$

(Tukey’s biweight)

$$u_i = \frac{|S_i - T_i|}{\varepsilon}$$

$$\varepsilon = \text{robust estimate of scale of } |S_i - T_i|$$

(e.g., MADN)

Q: *Iterative Winsorization works well in practice, but what are its convergence properties?*
Synthetic Example

Signal (2 spatial dimensions)

Gaussian noise

Erratic noise

Input

Standard Cadzow / SSA

Robust Cadzow / SSA

Erratic-Only Cadzow / SSA
Real Example (Final Image)

- Raw Final Image
- Standard Cadzow / SSA
- Robust Cadzow / SSA
7. Automatic Rank Determination
What Rank Do We Want?

- Noisy data, simple geology → Low rank
- Clean data, complex geology → High rank

Even within a single seismic survey, noise and geological conditions change with time, space, and frequency.

So how do we choose the rank?

Answer: **Automatically change the rank throughout to best suit the conditions.**
Given $\sigma$, the standard deviation of the noise, set the matrix rank to the number of singular values exceeding:

$$s^* = \frac{4}{\sqrt{3}} \sigma \sqrt{n}$$

Assumes matrix is $n \times n$. Formula for rectangular matrices more complicated.

Called **hard thresholding**.

**Soft thresholding** (shrinking the singular values) is better in theory. In practice not much improvement.
Three Problems

Two problems to overcome...

1. Estimating noise level $\sigma$ when full SVD is unavailable.
2. Making the filter less harsh for extreme noise.
Estimating The Noise Level

Requires an estimate of $\sigma$, standard deviation of the noise.

Donoho and Gavish suggest estimating it from the median singular value. But we only have the first $k$ singular values!

I developed an iterative method for estimating $\sigma$:

1. Rough estimate of $\sigma$
2. Estimate signal energy from $\sigma$ and first $k$ singular values
3. Estimate $\sigma$ from signal energy

Details in “Trickett, 2015, Preserving Signal: Automatic rank determination in noise suppression, Submitted to the SEG Annual Convention”
In extreme noise, automatic rank determination can recommend not keeping any singular values (zeroing the frequency slice).

Optimum in theory. Unacceptable in practice.

Countless ways to fix this. Suggest limiting threshold to something less than the first singular value $s_1$:

\[ s^{\uparrow*} = \min (s^{\uparrow*}, 0.75 s_{\downarrow 1}) \]
Synthetic Model

- **Trace 1**
  - Clean
  - Noisy
- **100 x 100 trace grid**
  - Flat
  - Curving
- **Time**
  - X
Rank Averaged Over All Frequencies

Trace 1

Increasing Rank

Clean

Noisy

Flat

Curving

Rank 1

Rank 8
Synthetic Model

Input

Clean

Moderately noisy

Very noisy

Rank 3

Rank 5

Auto Rank

Difference between pure signal and filtered data

Rank 3

Rank 5

Auto Rank

Higher rank in clean curving areas

Lower rank in noisy flat areas
Q: Is there anything special we should do for structured matrices? For robust filtering? For interpolation?

Q: Is soft thresholding (singular value shrinkage) worth it?
8. Interpolation
Migration

"Migration" uses the wave equation to back propagate the sound waves, as if they only travelled in one direction.

Geological model

What we actually see. Two-way travel produces a warped image.

For a good migration, sources and receivers must be regularly and densely placed.

Interpolation in four spatial dimensions tries to produce regular, dense shooting.

Interpolation

To interpolate we do *matrix completion*, filling in the unknown entries by assuming the completed signal has low rank.
General Completion Algorithm

Loop...
{
  Perform rank-reduction filtering.
  Accelerate convergence.
  Replace the known filtered elements of the frequency slice with the known original elements.
}
Perform rank-reduction filtering.

Some Accelerators

A. Nothing (works well when a small percentage of elements missing).
B. Over relaxation.
C. Scale the filtered frequency slice to least-squares fit the known elements.
D. Scale each singular component to least-squares fit the known elements.
E. Change the singular vectors to least-squares fit the known elements (ALS).
Interpolation

Model in four spatial dimensions
(one-dimensional slice shown)

Sloping in two spatial dimensions

Curving in two spatial dimensions

After interpolation in four spatial dimensions
Interpolation

Before interpolation

After interpolation in 4 spatial dimensions
Migrated Time Slice

Without interpolation

With interpolation in four spatial dimensions
9. Dealiasing Interpolation
When every 2\textsuperscript{nd} trace is missing, interpolating on single frequency slices creates ambiguities. This is called \textit{aliasing}. 

\begin{itemize}
  \item Random traces missing (1 spatial dimension)
  \item Interpolated
  \item Every 2\textsuperscript{nd} trace missing
  \item Interpolated
\end{itemize}
Solution: Look at more than one frequency slice at once!

Design a “rank-reduction operator” on every known trace of the half frequency, and apply it to the current frequency.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Trace</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 Hz</td>
<td><img src="image1.jpg" alt="Trace Design for 20 Hz" /></td>
</tr>
<tr>
<td>40 Hz</td>
<td><img src="image2.jpg" alt="Trace Design for 40 Hz" /></td>
</tr>
</tbody>
</table>

- Design rank reduction on known traces of 20 Hz
- Apply to 40 Hz
Designing and Applying

\[ A_{20} = \text{Trajectory matrix for known traces of 20 Hz.} \]
\[ A_{40} = \text{Trajectory matrix for 40 Hz.} \]

TSVD of \( A_{20} \)

Rank reduction of \( A_{40} \)

Imposes \( A_{20} \) pattern on \( A_{40} \).
Dealiasing Example

Every 2\textsuperscript{nd} Trace Missing

Dealiased Interpolation

10. Tensor Interpolation
Hankel Tensors

Frequency slice $S$

$4^{th}$-order Hankel tensor $T$

\[ S (i+j-1, m+n-1) = T (i, j, m, n) \]

Four Spatial Dimensions

\[ S (i+j-1, m+n-1, p+1-1, r+s-1) = T (i, j, m, n, p, q, r, s) \]

Every spatial dimension produces two tensor orders!
Define a rank-$k$ tensor as the sum of $k$ outer products.

Then the Exactness Theorem still holds:

If a noiseless multi-dimensional grid of traces is made up of the sum of plane waves having at most $k$ distinct dips then the Hankel tensor is at most rank $k$. 
For a 15 x 15 x 15 x 15 spatial grid, each outer product (plane wave) is described by…

- Hankel matrix: 8192 complex values
- Hankel tensor: 64 complex values

→ Fewer parameters to estimate.

*Improved performance over large gaps or when very sparse.*
Methods for tensor rank reduction:

- Higher Order SVD (HOSVD)
- Nuclear Norm Minimization
- Parallel Matrix Factorization
- PARAFAC using Alternating Least Squares (ALS)

Find a rank-$k$ tensor which is the best least-squares fit to the known elements of the raw tensor.
Efficiency

Reduction to rank $k$ of an $n \times n \times n \times n$ trace grid takes...

Hankel matrix: $O(k n^4 \log n)$
Hankel tensor: $O(k n^4)$ (PARAFAC ALS)

Despite the fact that the number of elements in the tensor is $O(n^8)$!

Q: Can other approaches for tensor rank reduction (e.g., HOSVD) be made as efficient?
2D Gap Filling

Hankel Matrix Interpolation

Hankel Tensor Interpolation

Time
4D Sparse Interpolation

Raw data in 4 spatial dimensions

90%  
95%  
97%  

Hankel Matrix Interpolation

Hankel Tensor Interpolation
4D Tensor Interpolation

Final image with no interpolation

Final image with tensor interpolation in four spatial dimensions
Thanks to...

Canadian Society of Exploration Geophysicists

Society of Exploration Geophysicists
Purpose of inner loop is to find the rank-\(k\) Hankel matrix closest to the original data Hankel matrix.

Inner loop proven to converge, but not necessarily to the closest rank-\(k\) Hankel matrix.

Multiplied cost without improving result.

Sacchi (2009) pointed out that without the inner loop, “Cadzow filtering” was equivalent Singular Spectrum Analysis (SSA).

Now we don’t know what to call the method. Cadzow / SSA?

Q: Would finding the rank-\(k\) Hankel matrix closest to the original give a better noise suppression?