Structured tensor missing-trace interpolation in the Hierarchical Tucker format

Curt Da Silva and Felix J. Herrmann

Sept. 26, 2013
Motivation

3D seismic experiments - 5D data
• expensive to acquire, store
• sample at sub-Nyquist rates

Data exhibits low-rank structure
• exploit structure for interpolation

Fully sampled data
• simultaneous sources in wave-equation based inversion
• mitigating multiples
Low-rank matrix/tensor completion via nuclear norm projection [1]
- Require SVDs on huge data matrices
- Not scalable to large problem sizes

Data completion via Toeplitz embedding [2]
- Problem size - (# data points)$^2$
Goals

Generalization of Compressible Sensing to multiple dimensions
  • what can we learn from 1D/2D recovery?

Randomized source/receiver acquisition
  • reduce acquisition financial/time costs

Efficient solver
  • SVD-free, parallelizable
  • # parameters << # data points
7.34 Hz - 75% missing receivers
Common source gather

True data

Subsampled data
7.34 Hz - 75% missing receivers

Common source gather

True data

Recovered data - SNR 17.4 dB
Compressive sensing
with sparsity promotion

Successful reconstruction scheme

Signal structure
  • sparsity

Sampling
  • subsampling decreases sparsity

Optimization
  • look for sparsest solution
Multidimensional interpolation
with Hierarchical Tucker

Successful reconstruction scheme

Signal structure
  • Hierarchical Tucker

Sampling
  • subsampling increases hierarchical rank

Optimization
  • fit data in the Hierarchical Tucker format
Matricization

The matricization of a tensor $X$ with dimensions $1, \ldots, d$ along the dimensions $t = (t_1, \ldots, t_r)$ is the matrix formed by placing the dimensions $t$ along the rows and dimensions $t^c$ along the columns.

Denoted $X^{(t)}$
Example in Matlab

n1 = 20; n2 = 20; n3 = 20; n4 = 20;
% Tensor
x = randn(n1,n2,n3,n4);

% Matricization along dimensions 1 and 2
X^{(1,2)} \ x12 = reshape(x,n1 * n2, n3 * n4);

% Matricization along dimensions 3 and 4
X^{(3,4)} \ y34 = permute(x,[3 4 1 2]);
\ x34 = reshape(x, n3 * n4, n1 * n2);

% Matricization along dimensions 1 and 3
X^{(1,3)} \ y13 = permute(x,[1 3 2 4]);
\ x13 = reshape(x,n1 * n3, n2 * n4);
Hierarchical Tucker format

\[ X - n_1 \times n_2 \times n_3 \times n_4 \text{ tensor} \]

\[
\begin{array}{c}
\begin{array}{c}
X^{(1,2)} \\
\end{array}
\end{array}
\begin{array}{c}
\begin{array}{c}
U_{12} \\
k_{12}
\end{array}
\end{array}
\begin{array}{c}
\begin{array}{c}
B_{1234} \\
k_{34}
\end{array}
\end{array}
\begin{array}{c}
\begin{array}{c}
U_{34}^T \\
n_3n_4
\end{array}
\end{array}
\]

“SVD”-like decomposition
Hierarchical Tucker format

\[ X - n_1 \times n_2 \times n_3 \times n_4 \text{ tensor} \]
Hierarchical Tucker format

\[ X - n_1 \times n_2 \times n_3 \times n_4 \text{ tensor} \]
Hierarchical Tucker format

Intermediate matrices don’t need to be stored

\( U_t, B_t \) - small parameter matrices
  • specify the tensor completely

*Separating* groups of dimensions from each other
  • dimension tree
The geometry of hierarchical tensors

\[
\{1, 2, 3, 4, 5\} = t_r
\]

\[
U_{123} \quad B_{123} \quad U_{45} \quad B_{45}
\]

\[
\{1, 2, 3\} \quad \{4, 5\}
\]

\[
\{1\} \quad \{2, 3\} = t \quad \{4\} \quad \{5\}
\]

\[
\{2\} = t_1 \quad \{3\} = t_2
\]
Hierarchical Tucker format

Storage $\leq dNK + (d - 2)K^3 + K^2$

Compare to $N^d$ storage for the full tensor

Effectively breaking the curse of dimensionality when $K \ll N, \quad d \geq 4$

Low frequency data compresses in HT
Hierarchical Tucker example

For a $100 \times 100 \times 100 \times 100$ cube with max rank 20

$N = 100$, $d = 4$, $K = 20$

Full storage: $N^d = 10^8$ values

HTucker storage: 24400 values

Compression of a factor of 99.97%
Seismic Hierarchical Tucker

We consider a 3D seismic survey with coordinates
(src x, src y, rec x, rec y, time)

We take a Fourier transform in time and restrict ourselves to a single
frequency slice
Seismic Hierarchical Tucker

For a frequency slice with coordinates \((\text{src } x, \text{src } y, \text{rec } x, \text{rec } y)\), there are essentially two choices of dimension splitting (by reciprocity):

- **Canonical Decomposition**
  - \(\{\text{src } x, \text{src } y, \text{rec } x, \text{rec } y\}\)
  - \(\{\text{src } x, \text{src } y\}\)
  - \(\{\text{rec } x, \text{rec } y\}\)
  - \(\{\text{src } x\}\)
  - \(\{\text{src } y\}\)

- **Non-canonical Decomposition**
  - \(\{\text{src } x, \text{rec } x, \text{src } y, \text{rec } y\}\)
  - \(\{\text{src } x, \text{rec } x\}\)
  - \(\{\text{src } y, \text{rec } y\}\)
  - \(\{\text{src } x\}\)
  - \(\{\text{rec } x\}\)
  - \(\{\text{src } y\}\)
  - \(\{\text{rec } y\}\)
Matricizations

(Rec x, Rec y) matricization - Canonical ordering
Matricizations

(Src x, Rec x) matricization - Noncanonical ordering
Multidimensional interpolation
with Hierarchical Tucker

Successful reconstruction scheme

Signal structure
- Hierarchical Tucker

Sampling
- subsampling increases hierarchical rank

Optimization
- fit data in the Hierarchical Tucker format
Matrix Completion

\[
\mathbf{X} = \begin{bmatrix}
* & * & * & * & * \\
* & * & * & * & * \\
* & * & * & * & * \\
* & * & * & * & * \\
* & * & * & * & *
\end{bmatrix}
\]

\[
\text{Normalized singular value}
\]

\[
\begin{array}{cccccc}
0 & 10 & 20 & 30 & 40 & 50 \\
10^{-1} & & & & & \\
10^{-2} & & & & & \\
10^{-3} & & & & & \\
10^{-4} & & & & & \\
\end{array}
\]
Matrix Completion

\[
A(X) = \begin{bmatrix}
* & * & * & 0 & * \\
* & 0 & 0 & * & 0 \\
* & * & * & * & * \\
* & * & 0 & * & * \\
0 & * & * & * & 0 \\
\end{bmatrix}
\]

![Normalized singular value plot](image)
Tensor Completion

Structure - recover a tensor $X$ which has low hierarchical rank
- Well represented in HT

Sampling - random removal of points increases rank
- Poorly represented in HT
- Idealized sampling
Idealized recovery
75% random entries removed
Common receiver gather

True data

Subsampled data
Idealized recovery
75% random entries removed
Common receiver gather
Idealized recovery
75% random entries removed
Common receiver gather

True data

Recovered data - SNR 19.3 dB
Sampling

\((x_{src}, y_{src}, x_{rec}, y_{rec})\) points from the data
- idealized recovery
- impossible to physically implement

\((x_{rec}, y_{rec})\) points from the data
- less idealized
- possible to acquire data - e.g. ocean bottom nodes
Realistic recovery
50% random receivers removed

(Rec x, Rec y) matricization - Canonical ordering
Realistic recovery
50% random receivers removed

\((\text{Src } x, \text{Rec } x)\) matricization - Noncanonical ordering
Data organization

In summary:

(rec x, rec y) organization
- High rank
- Missing sources operator - removes columns
- Poor recovery scenario

(src x, rec x) organization
- Low rank
- Missing sources operator - removes blocks
- Closer to ideal recovery scenario
Multidimensional interpolation
with Hierarchical Tucker

Successful reconstruction scheme

Signal structure
  • Hierarchical Tucker

Sampling
  • subsampling increases hierarchical rank

Optimization
  • fit data in the Hierarchical Tucker format
Optimization

Given data $b$ with missing sources and/or receivers, subsampling operator $A$, full tensor expansion operator

$$
\phi : (U_t, B_t) \rightarrow \mathbb{C}^{n_1 \times \cdots n_d}
$$

solve

$$
\min_{x=(U_t,B_t)} \frac{1}{2} ||A\phi(x) - b||_2^2
$$
Differential geometry

HT tensors parametrize a submanifold of full tensor space $\mathbb{C}^{n_1 \times \ldots \times n_d}$

- Nonlinear, nonconvex space
- Generalization of curved surfaces

Steepest Descent, Conjugate gradient, Gauss-Newton

- *without* SVDs in the full tensor space


C. Da Silva and F. J. Herrmann, *Optimization on the Hierarchical Tucker manifold - applications to tensor completion*, 2013
Optimization program

Parameter space

$\mathbb{C}^D$  
$x = (U_t, B_t)$

$x_{\text{best}}$

Full-tensor space

$\mathbb{C}^{n_1 \times \ldots \times n_d}$

$\phi(x)$

$\phi(x_{\text{best}})$
Optimization program

\[ A\phi(x) \]

\[ \mathbb{C}^{n_1 \times \ldots n_d} \]

\[ \phi(x) \]

\[ \phi(x_{\text{best}}) \]

\[ -\nabla f \]
Derivatives

Derivatives of a particular node with respect to its children can be computed efficiently, i.e. via

\[
(I - U_{t_l} U^{H}_{t_l}) \langle U^T_{t_r} \circ_2 Z, B_t \rangle_{(2,3),(2,3)}
\]

\[
(I - U_{t_r} U^{H}_{t_r}) \langle U^T_{t_l} \circ_1 Z, B_t \rangle_{(1,3),(1,3)}
\]

\[
P \circ_i Q\text{ multiplies } Q\text{ by } P\text{ in dimension } i, \langle X, Y \rangle_{(1,3),(1,3)} = \sum_{i_1, i_3} \overline{X}_{i_1, \cdot, i_3} Y_{i_1, \cdot, i_3}
\]

The chain rule gives the gradient of the function \( \phi \).
Derivatives

Only involves matrix-matrix multiplications of small matrices compared to the full-tensor space

Parallelizable - multilinear product can be done in parallel

SVD-free - no large-scale SVDs, unlike nuclear norm-based methods
Results
Synthetic BG Group data

Unknown model
  - 68 x 68 sources with 401 x 401 receivers, data at 4.68Hz, 7.34 Hz

Receivers subsampled to 201 x 201

Recovered with Gauss-Newton
4.68 Hz - 75% missing receivers

Common source gather

True data

Subsampled data
4.68 Hz - 75% missing receivers
Common source gather

True data

Recovered data - SNR 19.7 dB
4.68 Hz - 75% missing receivers
Common source gather

True data

Difference
4.68 Hz - 75% missing receivers
Common receiver gather - no data initially

True data

Recovered data - SNR 23 dB
4.68 Hz - 75% missing receivers
Common receiver gather - no data initially

True data

Difference
7.34 Hz - 75% missing receivers
Common source gather

True data

Subsampled data
7.34 Hz - 75% missing receivers
Common source gather

True data

Recovered data - SNR 17.6 dB
7.34 Hz - 75% missing receivers

Common source gather

True data

Difference
7.34 Hz - 75% missing receivers
Common receiver gather - no data initially

True data

Recovered data - SNR 16.2 dB
7.34 Hz - 75% missing receivers
Common receiver gather - no data initially

True data

Difference
**7.34 Hz - simultaneous receivers - 90% data reduction**

*Common source gather*

**True data**

**Input data -** $A^T Ab$

$A$ - subsampling operator    $b$ - full data
7.34 Hz - simultaneous receivers - 90% data reduction

**Common source gather**

**True data**

**Recovered data - SNR 16.4 dB**
7.34 Hz - simultaneous receivers - 90% data reduction

Common source gather

True data

Difference
Conclusion

3D seismic data has an underlying structure that we can exploit for interpolation (Hierarchical Tucker format)

Different schemes for organizing data - important for recovery
Conclusion

We can interpolate HT tensors with missing entries using the Riemannian manifold structure of the HT format.

Achieve good results from largely subsampled data (75% missing receivers).

Can use this method to create full volumes from subsampled data:
- Migration, multiple removal, etc.
Acknowledgements

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Total SA, WesternGeco, Statoil, and Woodside.

Thank you for your attention