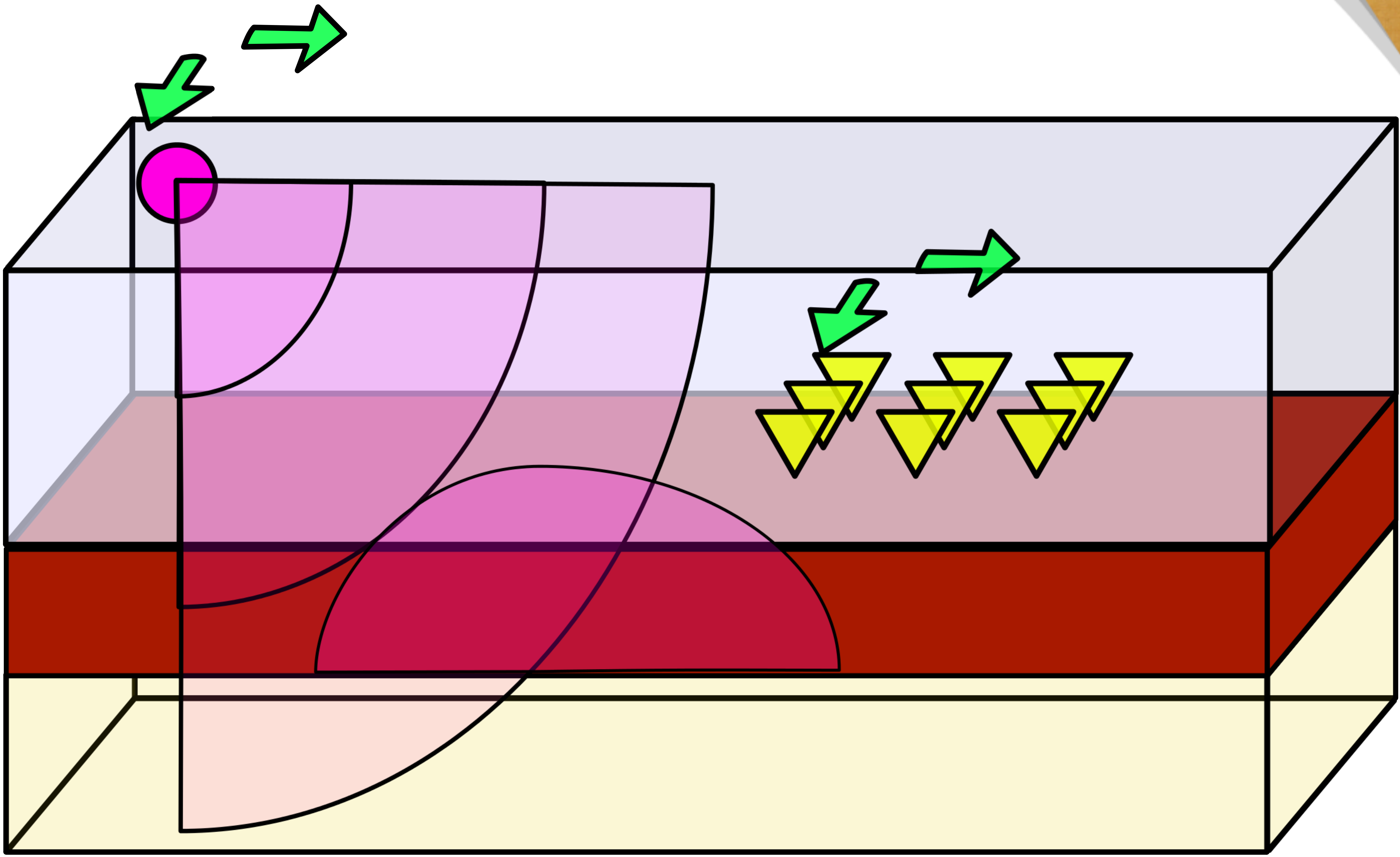
Hierarchical Tucker Tensor Optimization - Applications to Tensor Completion
SAMPTA 2013

Curt Da Silva and Felix J. Hermann

Seismic Laboratory for Imaging and Modeling

SLIM
University of British Columbia

3D Seismic Experiments



Motivation

- 5D data - (src x, src y, rec x, rec y, time)
  - Expensive to acquire, store, process
  - Curse of dimensionality

- Low (temporal) frequency data
  - Well represented as a structured tensor

SLIM

Goals

- Interpolation
  - Subsample data collection - reduce costs
- Matrix completion/Compressive sensing
  - Heuristic insights for tensor recovery

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 h-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 $X^{(t)}$

Multilinear product

Given $X - n_1 \times n_2 \times \ldots n_d$ tensor

$A_i \in \mathbb{C}^{m_i \times n_i}, i = 1, \ldots, d$ matrices

Multilinear product

$$Y = A_1 \circ_1 A_2 \circ_2 \ldots A_d \circ_d X$$

Unfolded

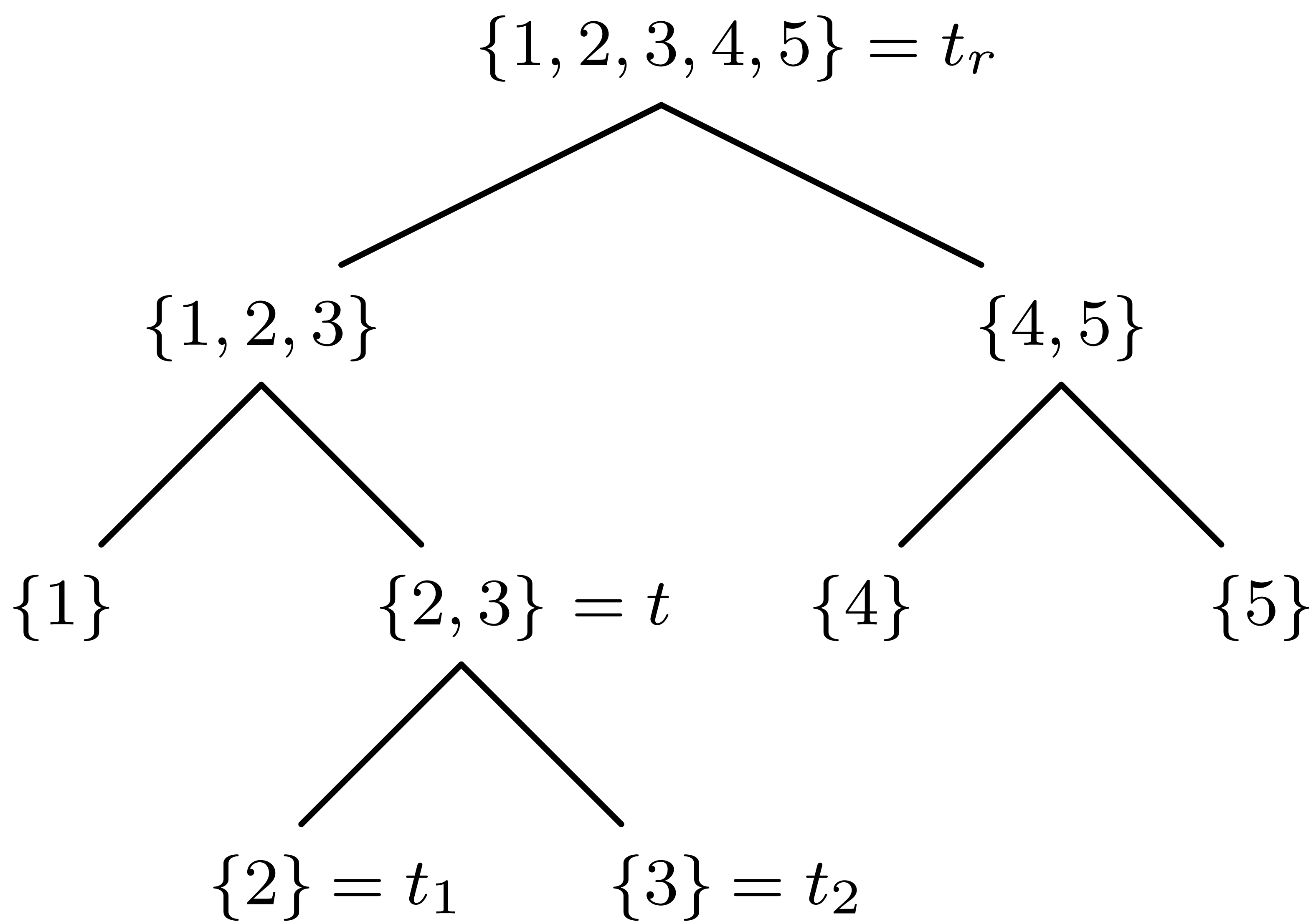
$$Y^{(i)} = A_i X^{(i)} A_d^T \otimes \ldots A_{i+1}^T \otimes A_{i-1}^T \otimes \ldots A_1^T$$

Hierarchical Tucker Format

- A *dimension tree* T for dimensions $\{1,\ldots,d\}$ is a non-trivial binary tree such that
  - the root, t_{root}, is has the label $\{1,\ldots,d\}$
  - each non-leaf node, t, can be written as $t = t_l \cup t_r$, $t_l \cap t_r = \emptyset$ where t_l is the left child of t, t_r is the right child of t

A. Uschmajew, B. Vandereycken. The geometry of algorithms using hierarchical tensors. Linear Algebra and its Applications, 2013.

Example



Hierarchical Tucker Format

- A tensor X can be written in the *Hierarchical Tucker format* corresponding to a dimension tree T and a vector of hierarchical ranks
$$(k_t)_{t \in T}, k_{\mathrm{root}} = 1$$
if it can be written as

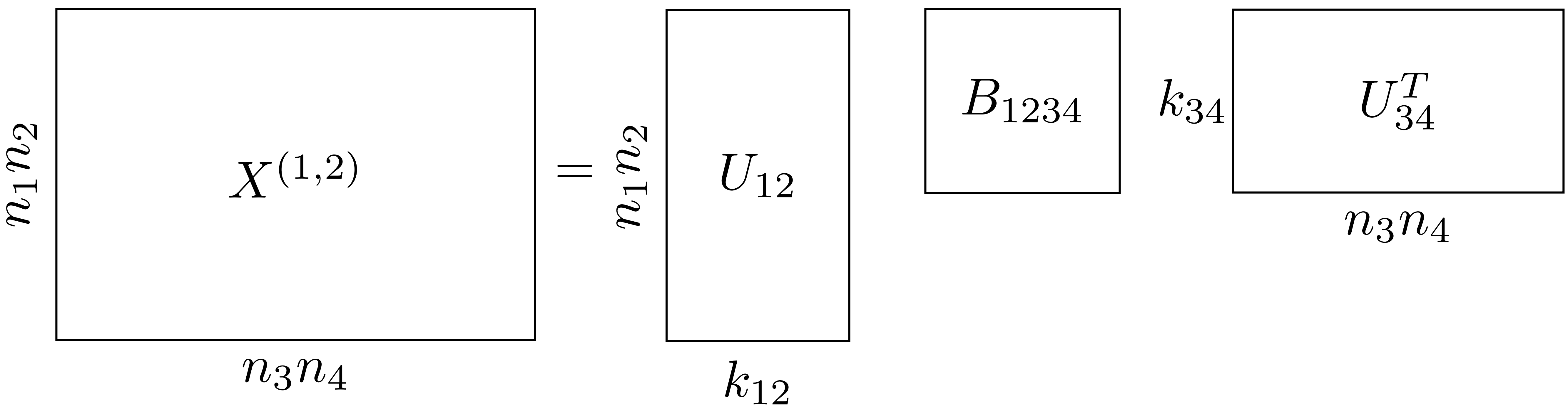
Hierarchical Tucker Format

$$X = U_{t_l} \circ_1 U_{t_r} \circ_2 B_{t_{\text{root}}} \qquad t = t_{\text{root}}$$

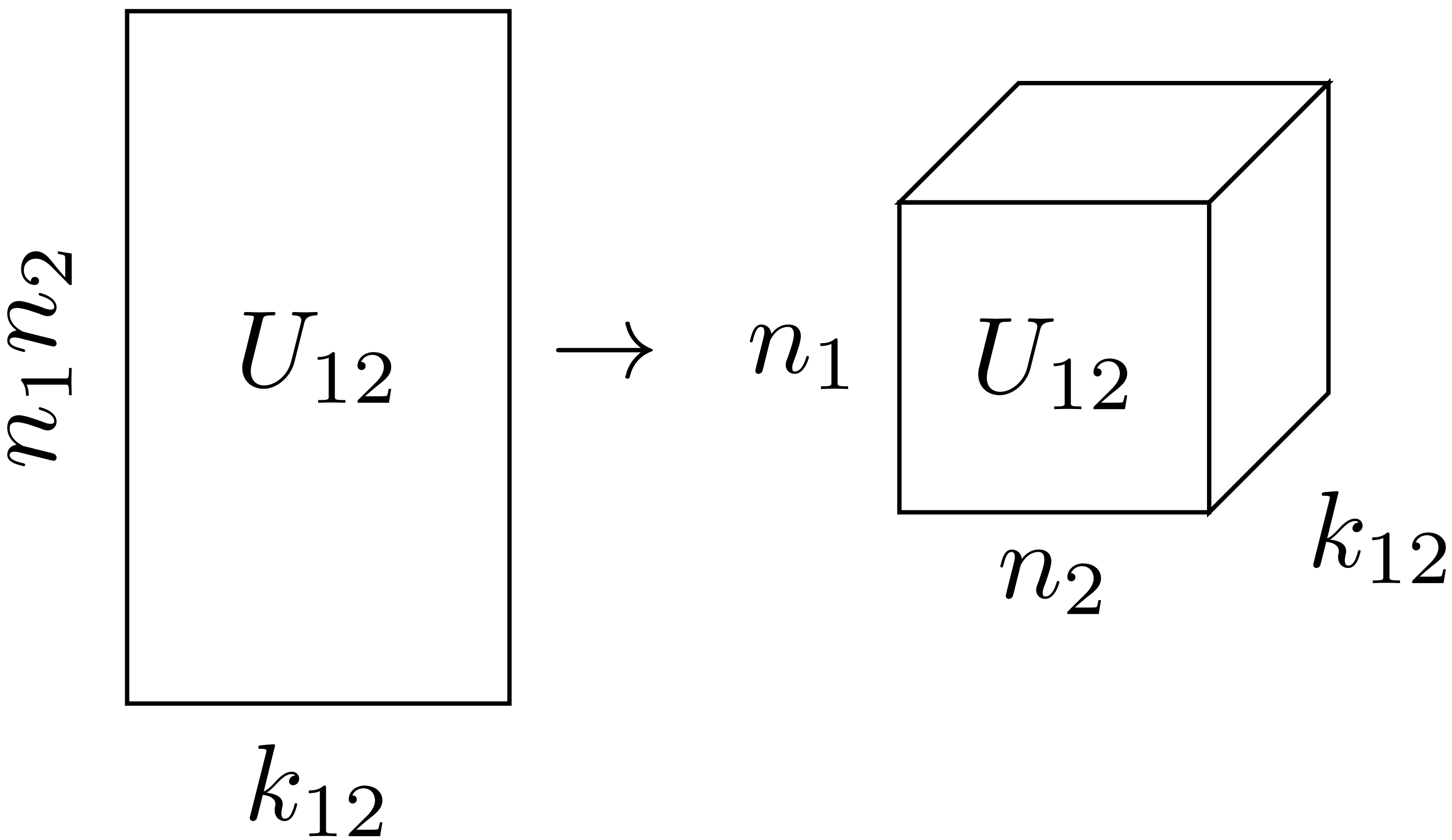
$$U_t = U_{t_l} \circ_1 U_{t_r} \circ_2 I_{k_t} \circ_3 B_t \quad t \text{ not a leaf}$$

$$U_t \in \mathbb{C}^{n_t \times k_t} \qquad B_t \in \mathbb{C}^{k_l \times k_r \times k_t}$$

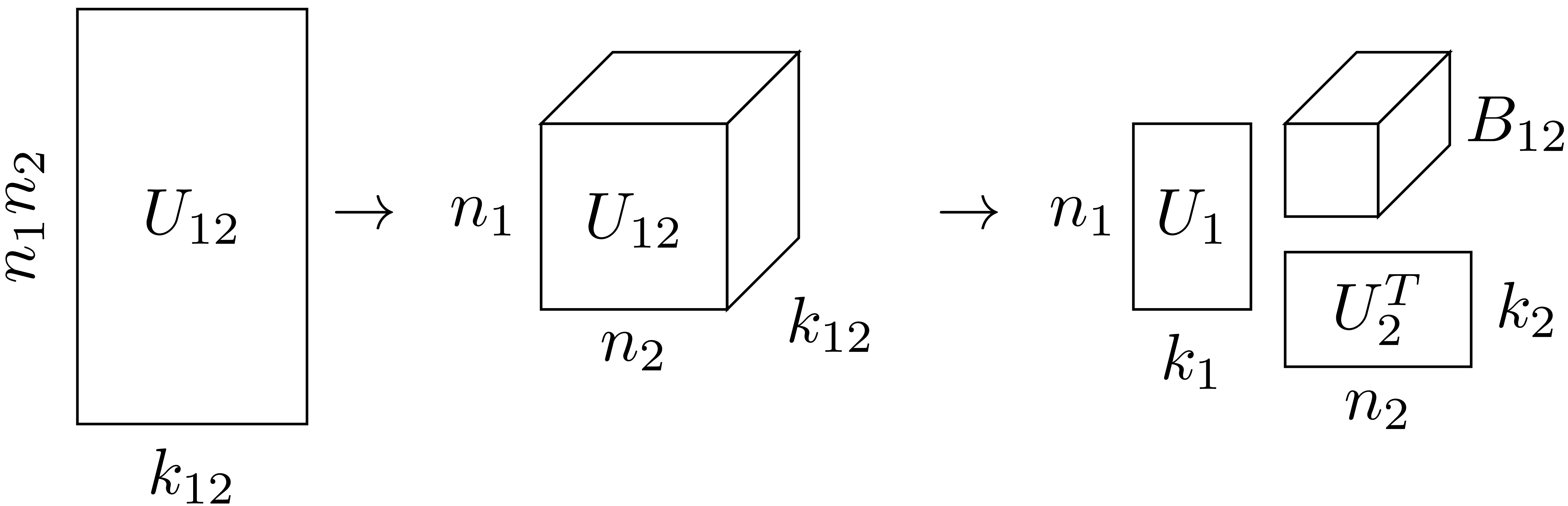
HT Format $X - n_1 \times n_2 \times n_3 \times n_4$ tensor



HT Format $X - n_1 \times n_2 \times n_3 \times n_4$ tensor



HT Format $X - n_1 \times n_2 \times n_3 \times n_4$ tensor



Hierarchical Tucker Format

- Intermediate matrices don't need to be stored
- U_t, B_t - small parameter matrices
  - specify the tensor completely

Hierarchical Tucker Format

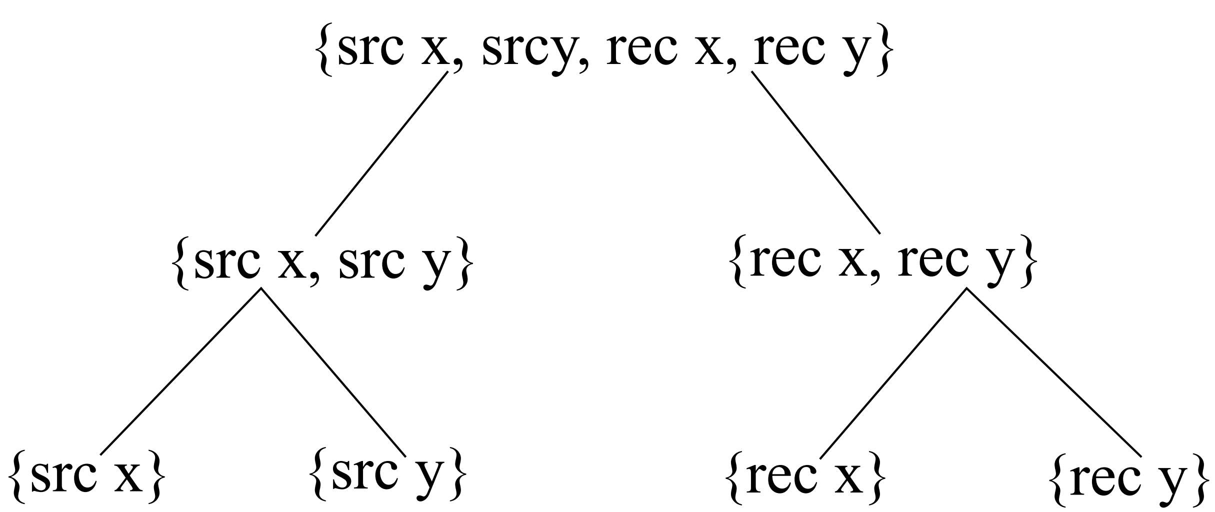
- Storage $\leq dNK + (d-2)K^3 + K^2$
- Compare to N^d storage for the full tensor
- Effectively breaking the curse of dimensionality when $K \ll N$

Seismic HTucker

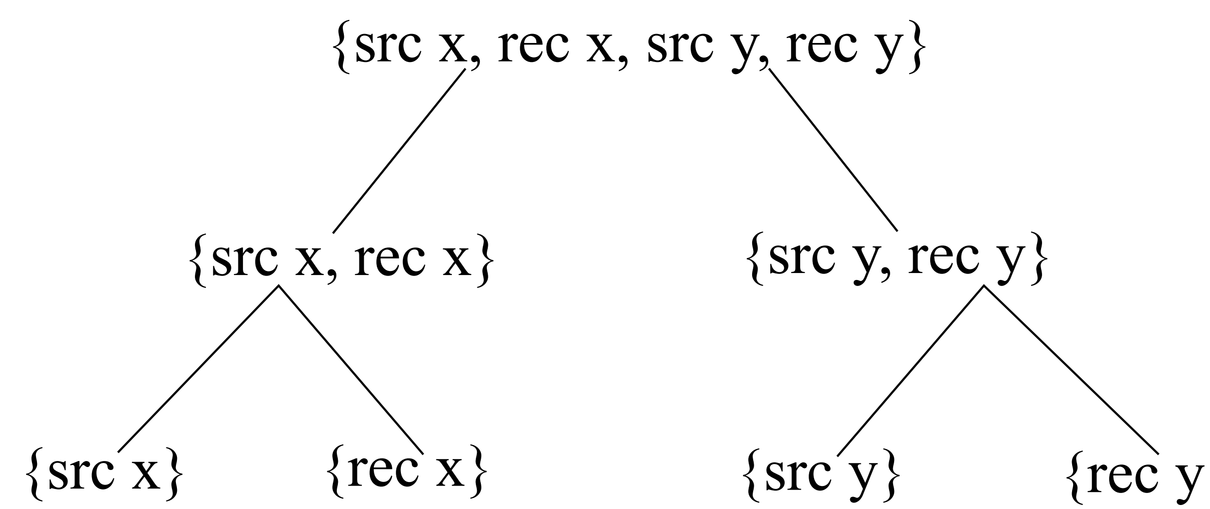
- We consider a 3D seismic survey with coordinates (src x, src y, rec x, rec y, time)
- We take a Fourier transform in time and restrict ourselves to a single frequency slice

Seismic HTucker

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

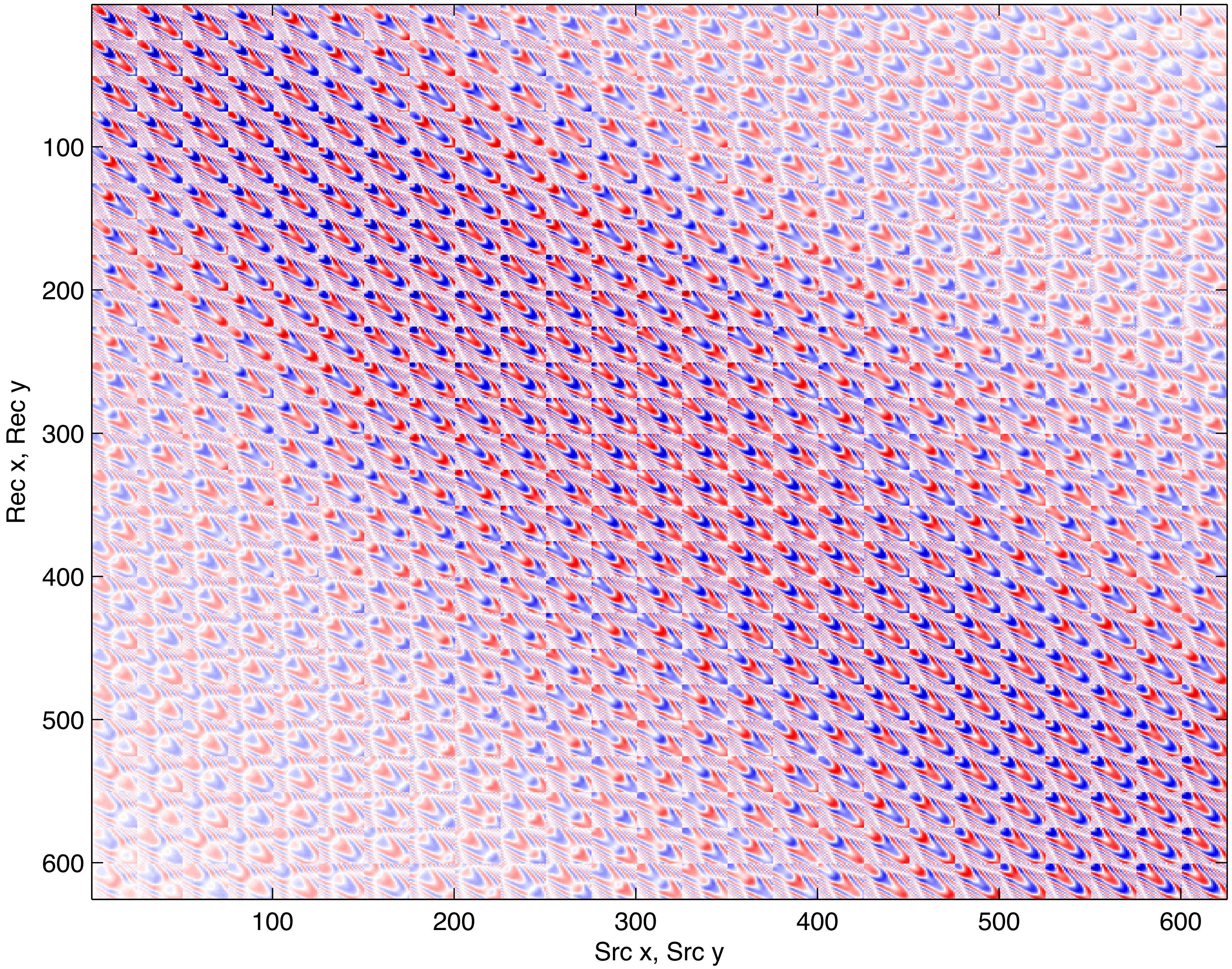


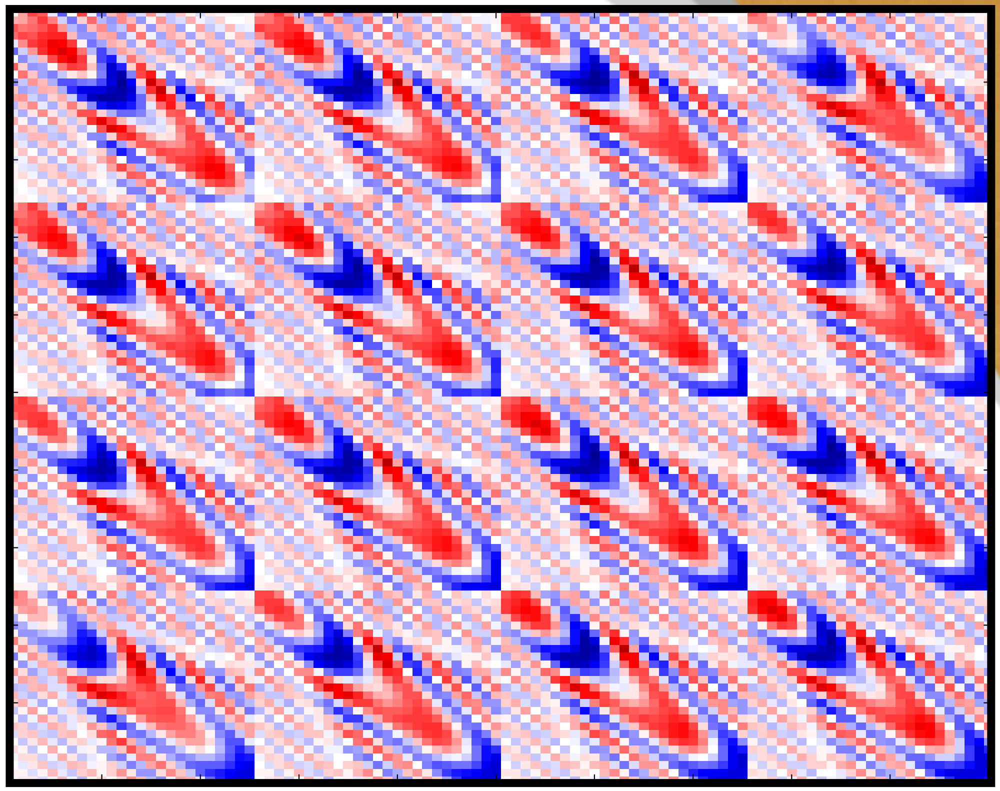
Canonical Decomposition

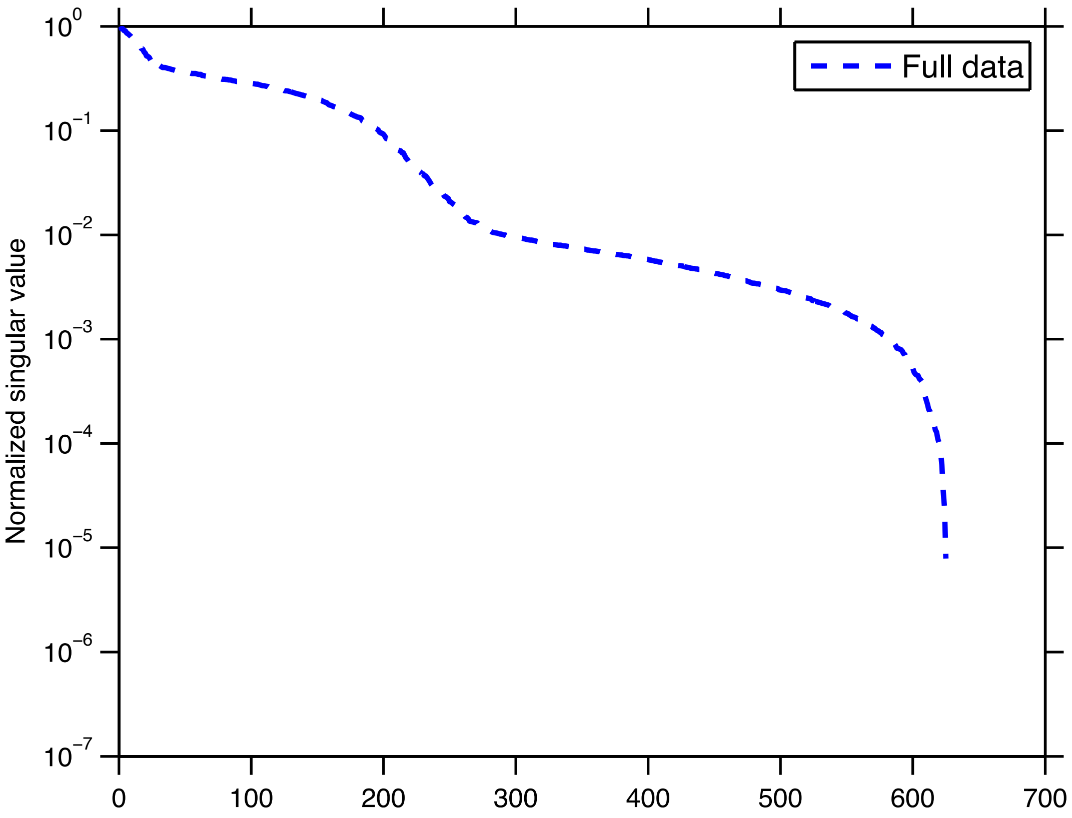


Non-canonical Decomposition

Matricizations

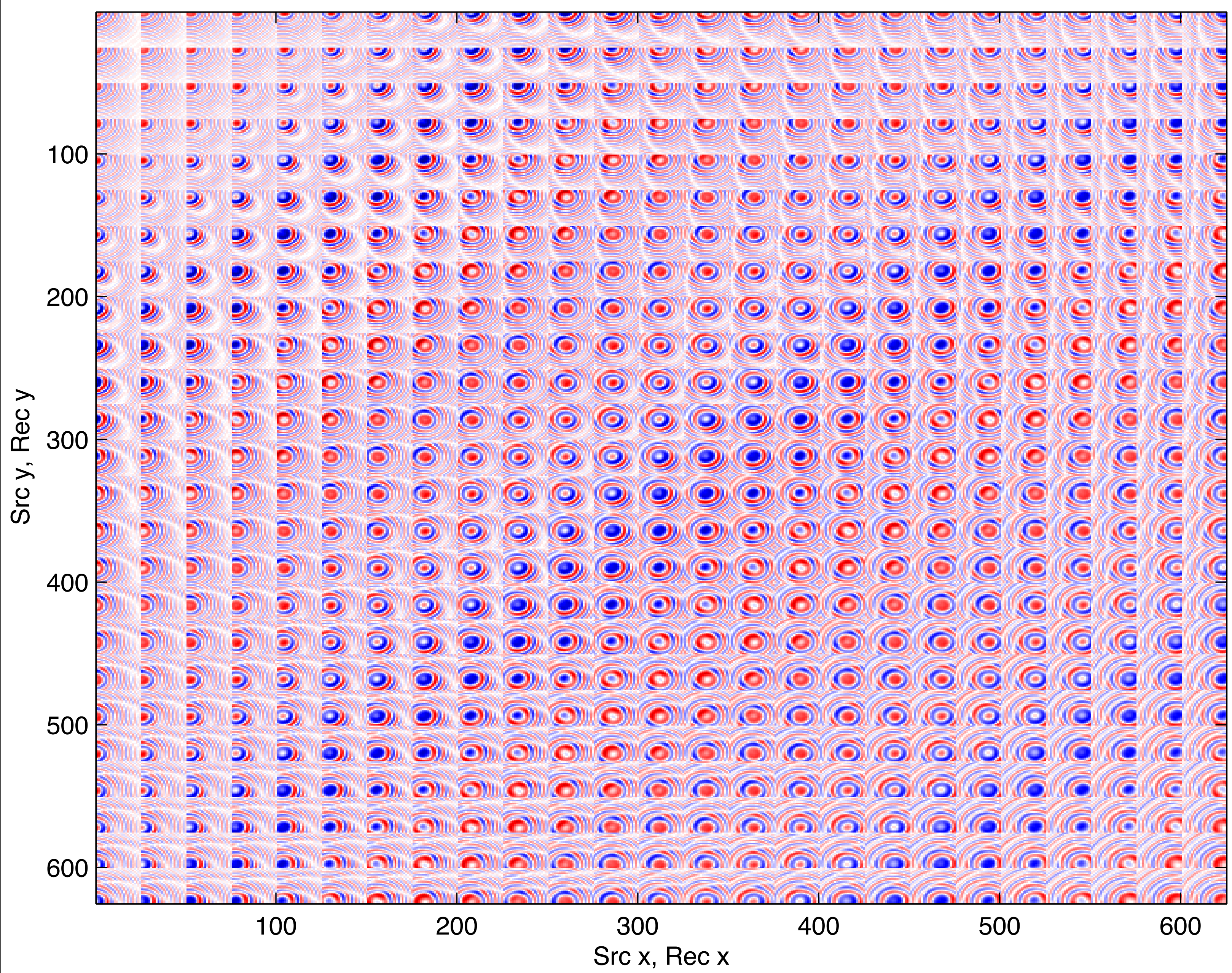


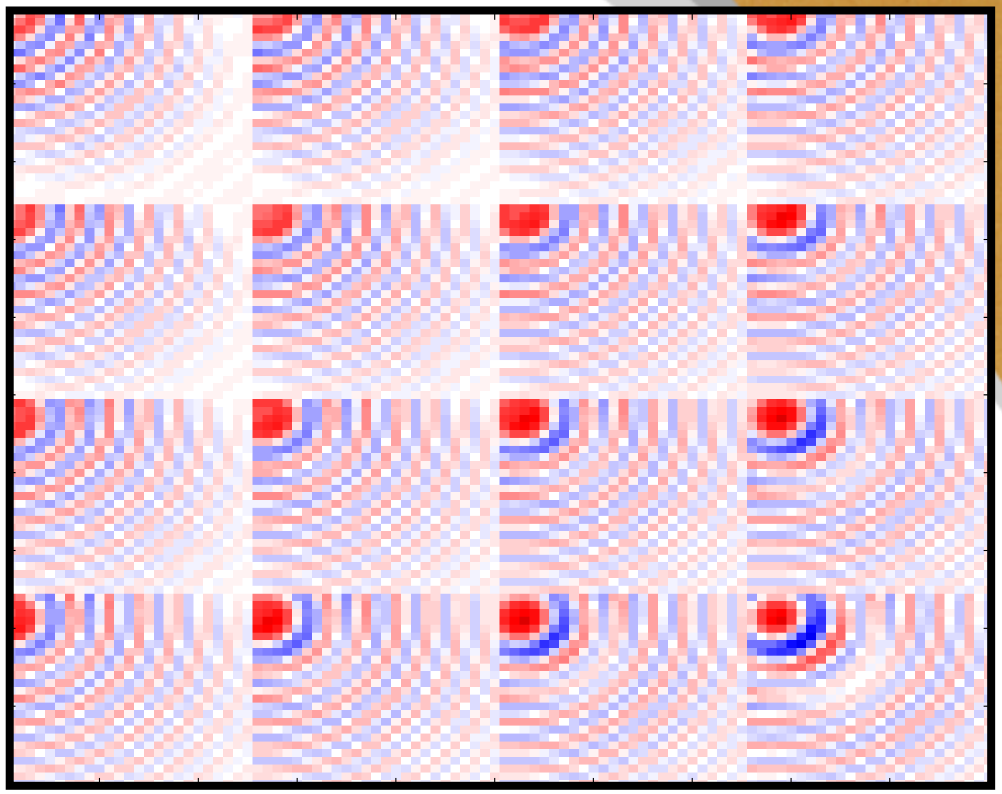


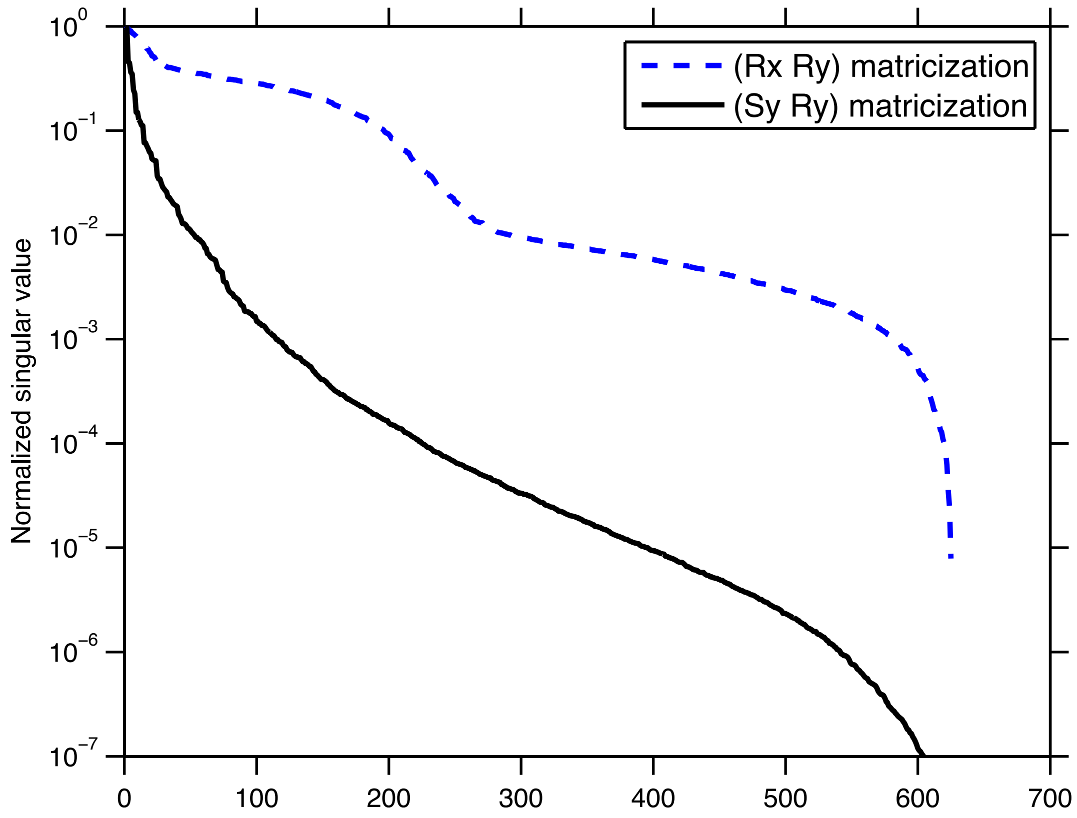


(Rec x, Rec y) matricization - full data

Matricizations







(Src y, Rec y) matricization - full data

Multidimensional interpolation

Successful reconstruction scheme

- Signal structure - *Hierarchical Tucker*
- **Sampling - *subsampling increases h-rank***
- Optimization - *fit data in the Hierarchical Tucker format*

Matrix Completion

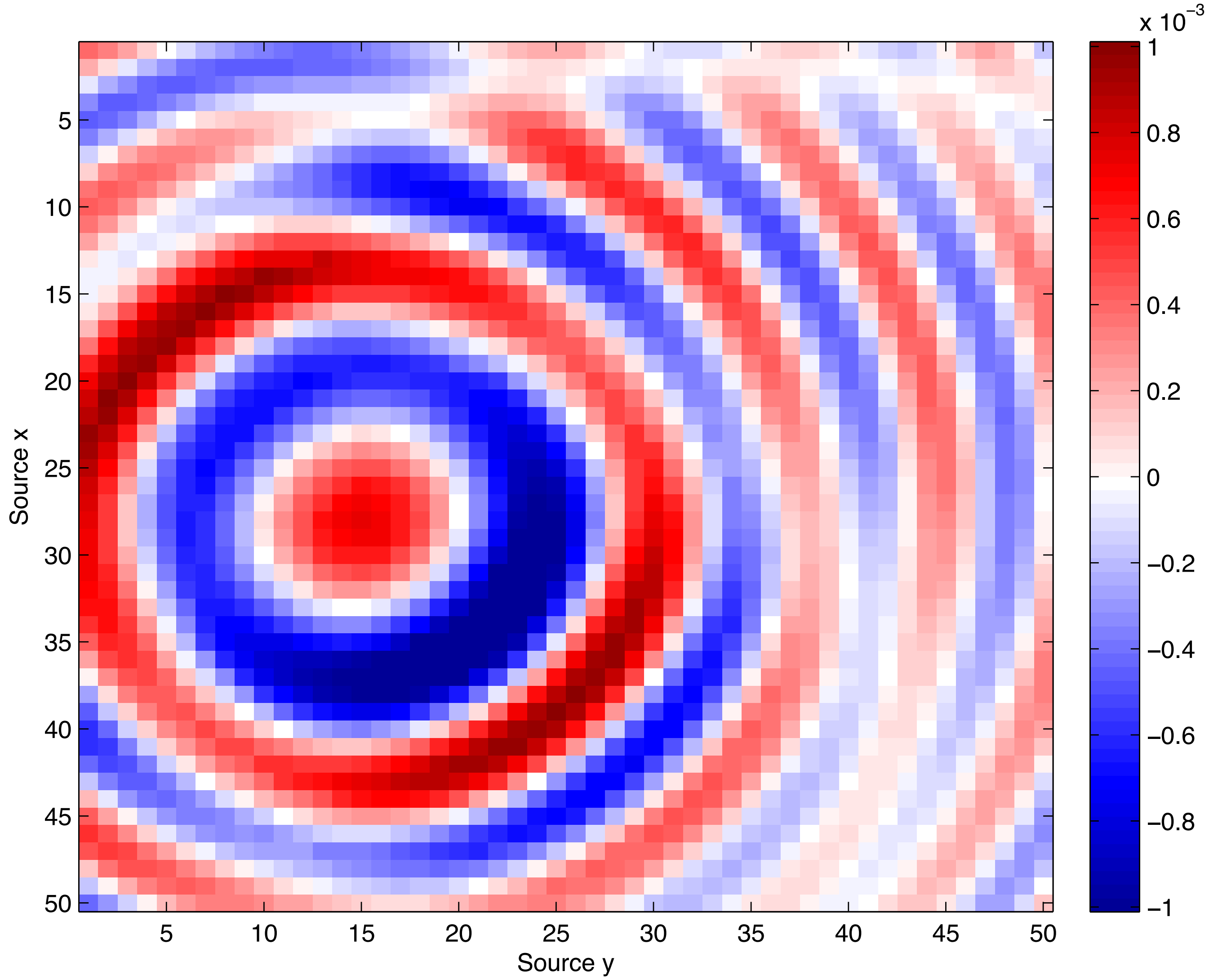
- *Structure* - recover a *matrix* $\mathbf{X}$ which is low rank
- *Sampling* - random removal of points tends to *increase* rank

Tensor Completion

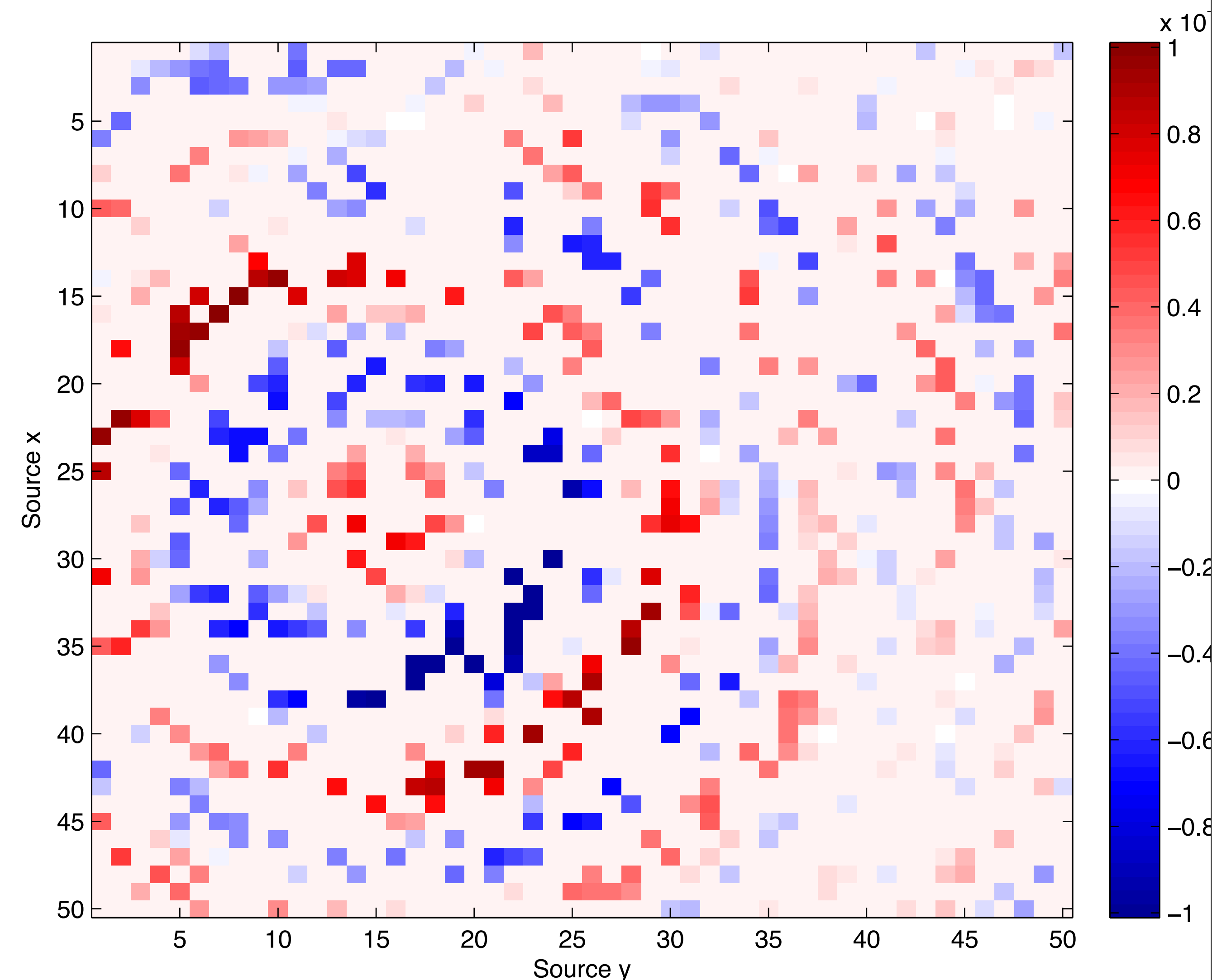
- *Structure* - recover a *tensor* $\mathbf{X}$ which has low *hierarchical* rank
    - Well represented in HT
- *Sampling* - random removal of entries *increases* rank
    - Poorly represented in HT
    - Idealized sampling

Idealized recovery

75% random entries removed
Fixed source coordinates

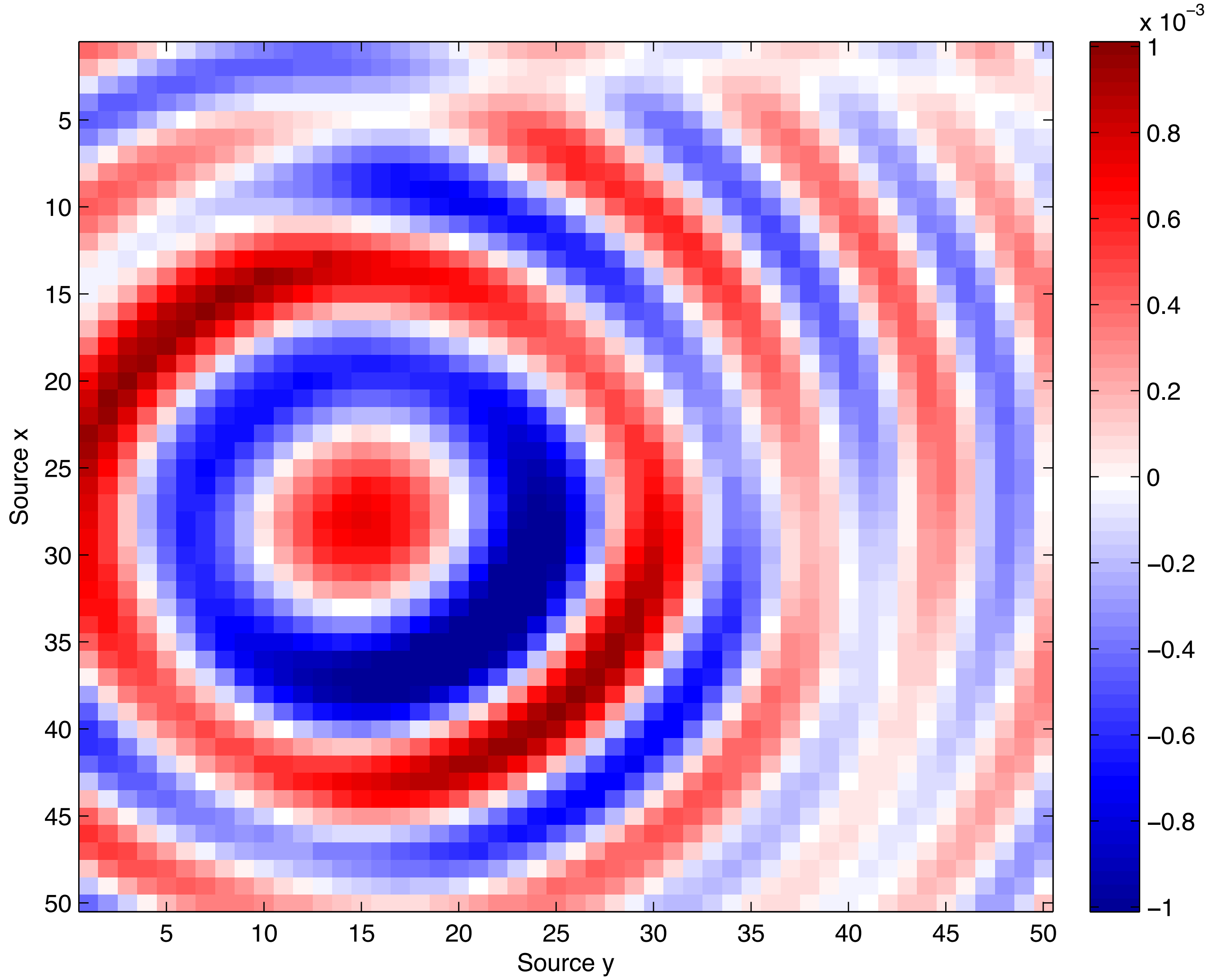


True Data

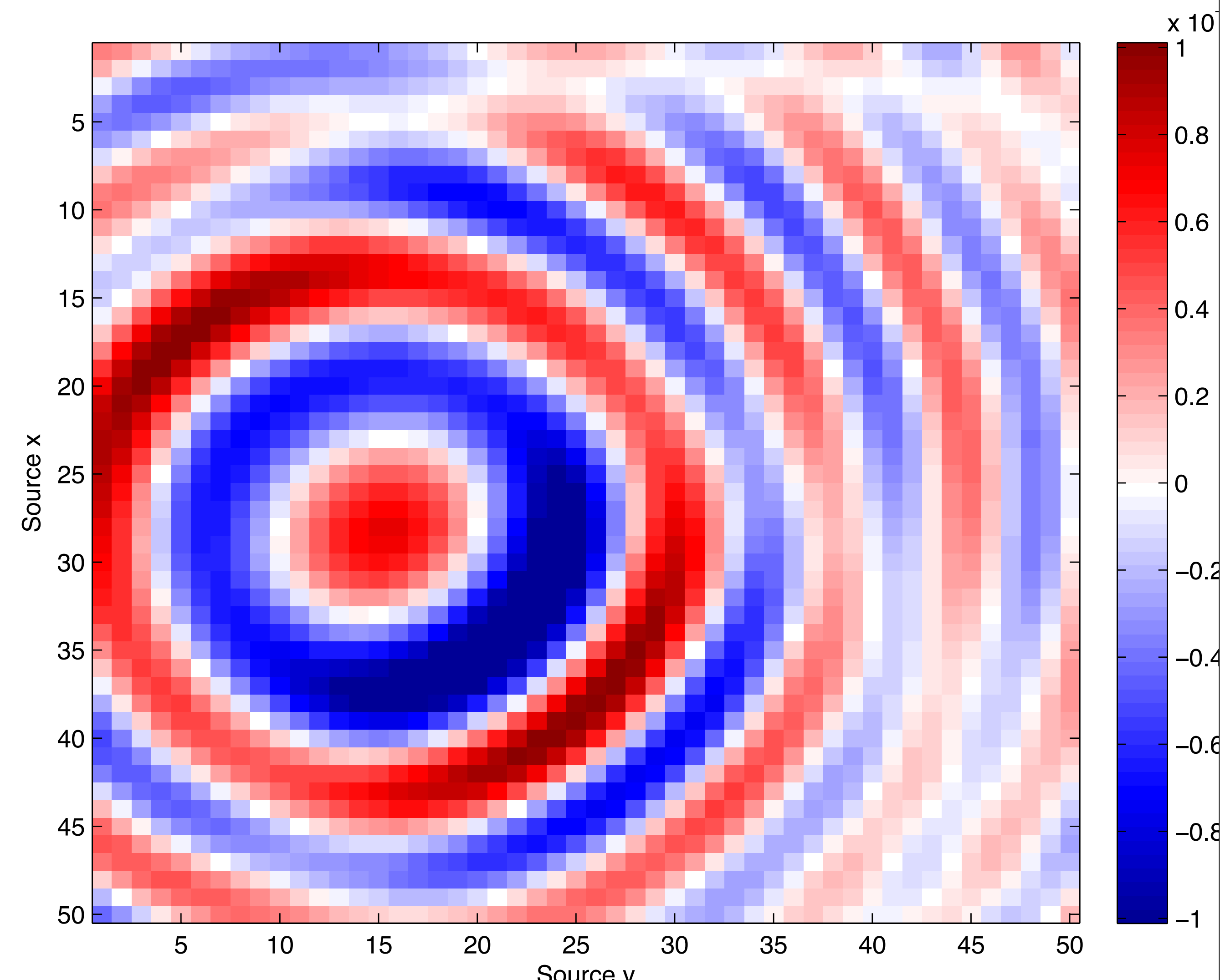


Subsampled Data

Idealized recovery
75% random entries removed
Fixed source coordinates



True Data



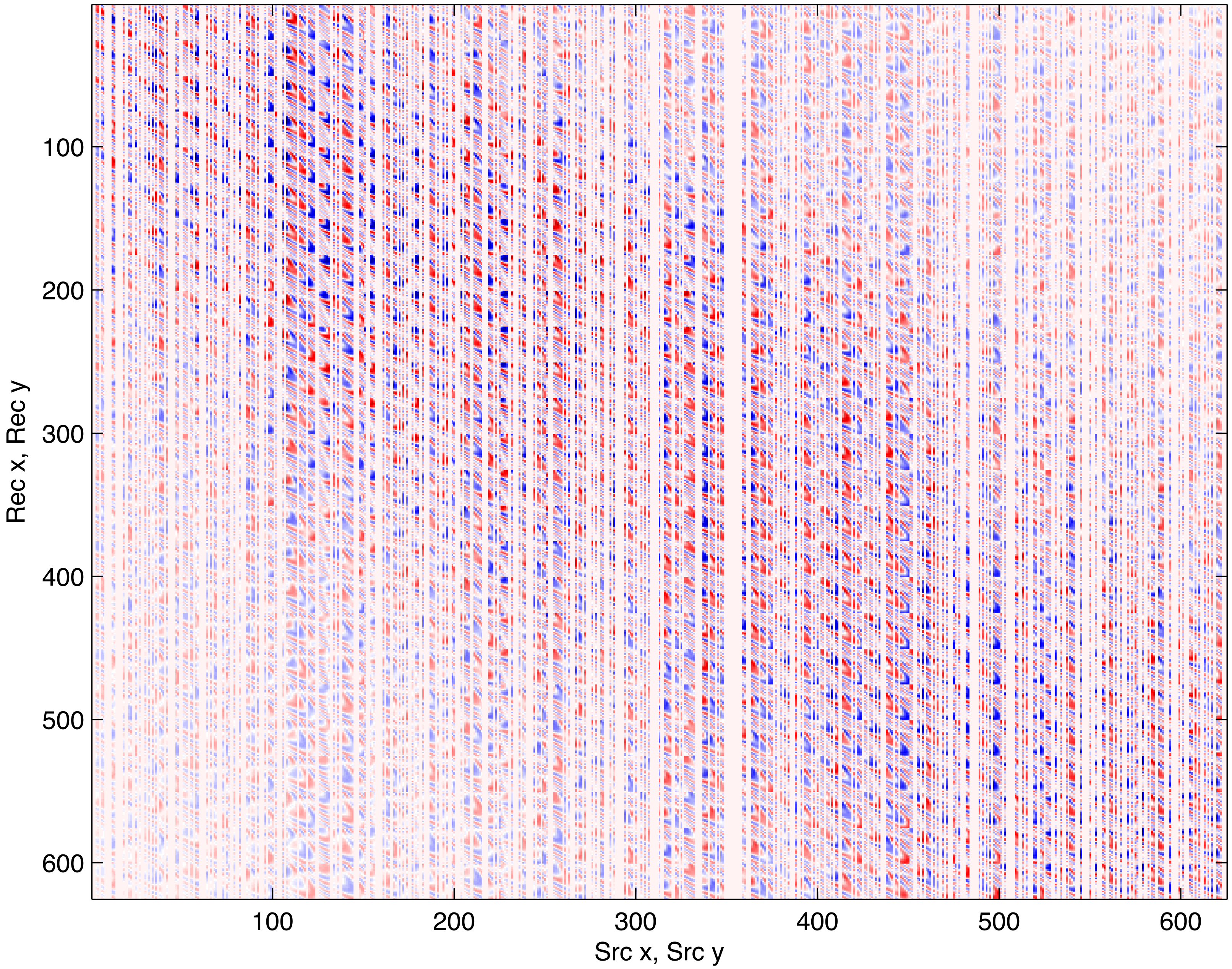
Recovered Shot

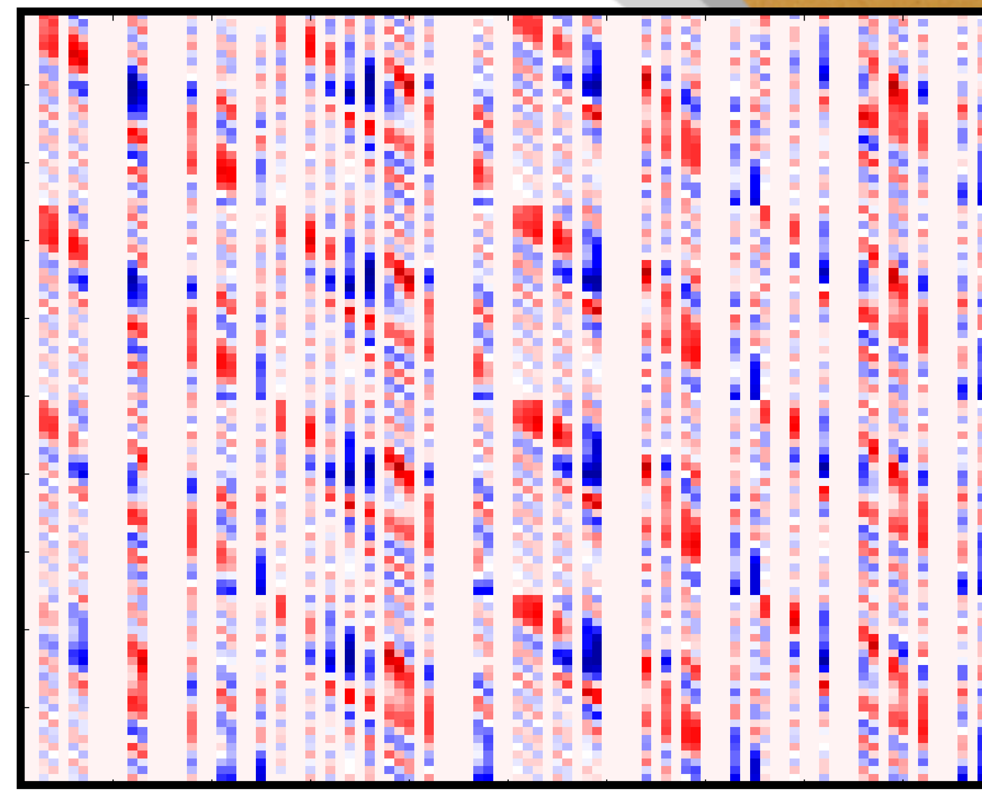
SNR 19.3 dB

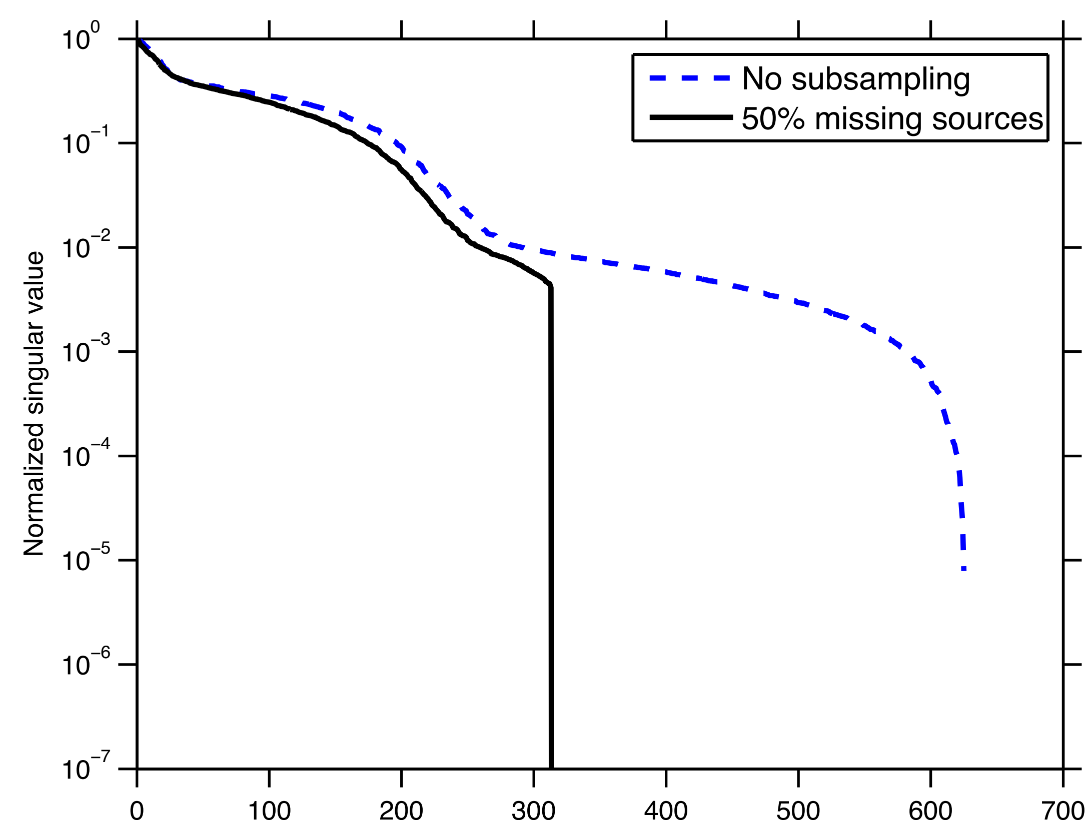
Sampling

- Sampling (src x, src y, rec x, rec y) points
  - Not physical
- How does our data behave under randomly missing sources or receivers?
  - Physically realizable

Sampling



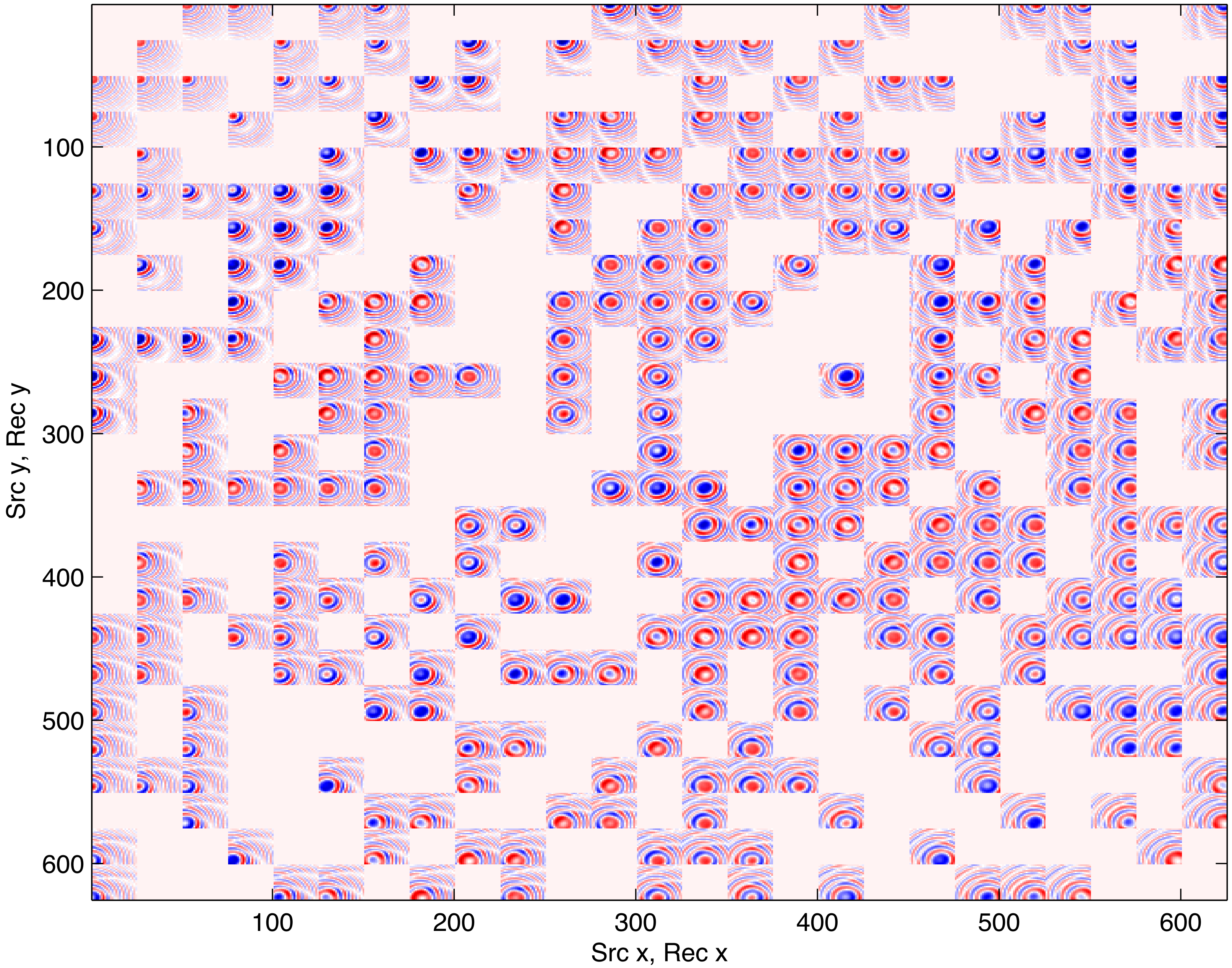


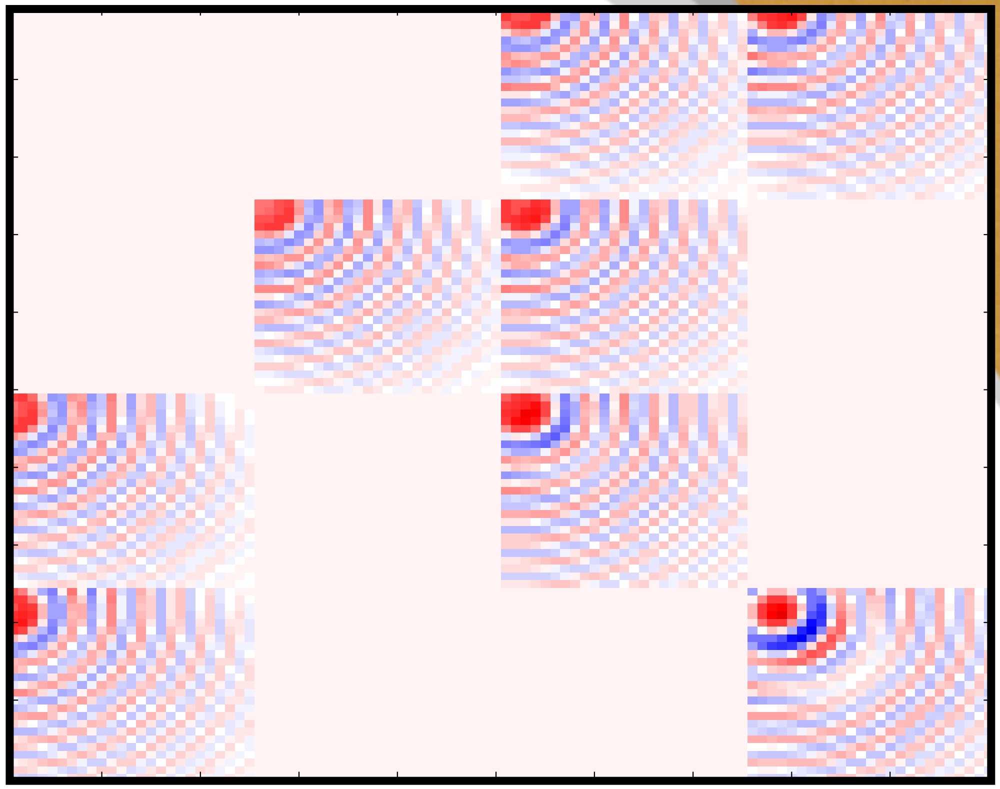


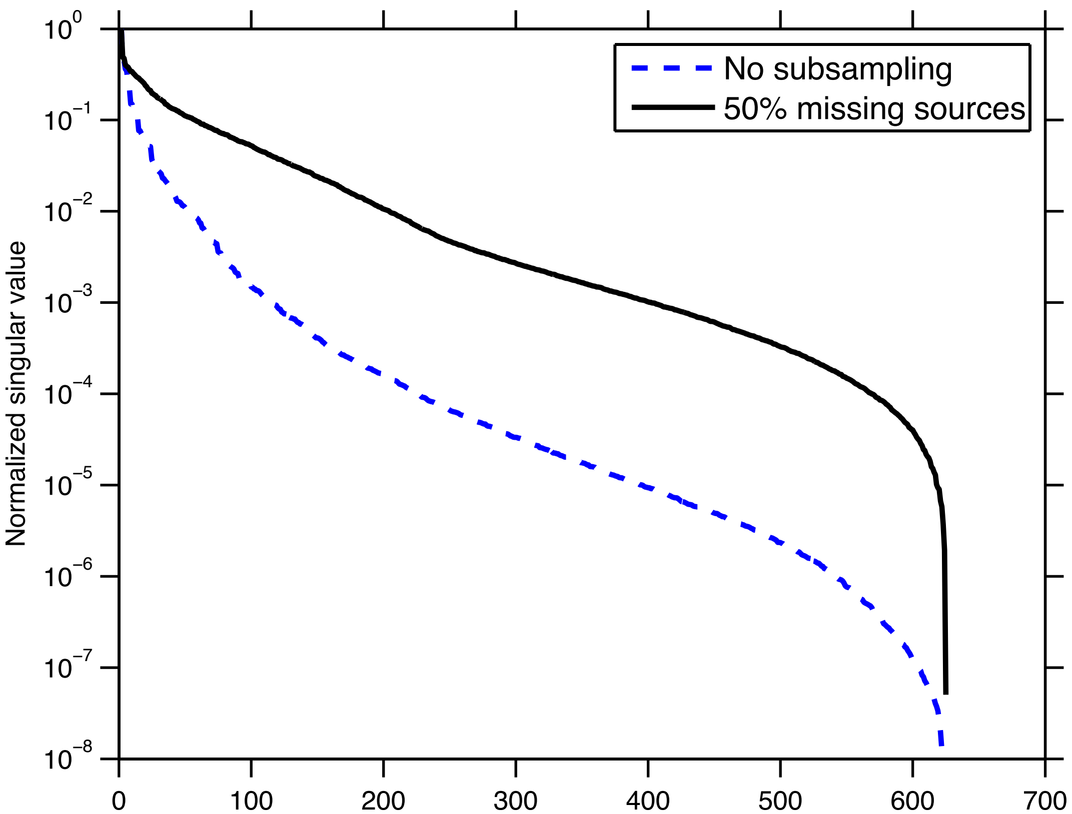
(Rec x, Rec y) matricization

50% randomly missing sources

Sampling







(Src y, Rec y) matricization

50% randomly missing sources

Multidimensional interpolation

Successful reconstruction scheme

- Signal structure - *Hierarchical Tucker*
- Sampling - *subsampling increases h-rank*
- **Optimization - *fit data in the Hierarchical Tucker format***

Optimization Format

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

$$\phi : (U_t, B_t) \to \mathbb{R}^{n_1 \times \dots n_d}$$

solve

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

SLIM

P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton Univ Press, 2008.
A. Uschmajew, B. Vandereycken. *The geometry of algorithms using hierarchical tensors*. Linear Algebra and its Applications, 2013

Differential Geometry

- Representations $x = (U_t, B_t)$ of HT tensors are non-unique
- $y = (U_t A_t, A_{t_l}^{-1} \circ_1 A_{t_r}^{-1} \circ_2 A_t \circ_3 B_t)$ under ϕ maps to $\phi(x)$ for $A_t \in \mathrm{GL}(k_t)$
- Group action on parameters -> Quotient manifold

SLIM

P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton Univ Press, 2008.
A. Uschmajew, B. Vandereycken. *The geometry of algorithms using hierarchical tensors*. Linear Algebra and its Applications, 2013

Differential Geometry

- HT tensors are a *submanifold* of $\mathbb{R}^{n_1 \times \dots n_d}$

- "If you like it, then you should put a ... [Riemannian metric] on it " - Beyoncé

Riemannian metric

At $x = (U_t, B_t)$, given horizontal vectors $dx = (\delta U_t, \delta B_t)$, $dy = (\delta V_t, \delta C_t)$

$$\langle dx, dy \rangle_x = \sum_{t \in T} \langle dx_t, dy_t \rangle_t$$

$$\langle dx_t, dy_t \rangle_t = \langle (U_t^T U_t)^{-1} \delta U_t^T \delta V_t \rangle$$ t is a leaf

$$\langle dx_t, dy_t \rangle_t = \langle \delta B_t, U_{t_l}^T U_{t_l} \circ_1 U_{t_r}^T U_{t_r} \circ_2 (U_t^T U_t)^{-1} \circ_3 \delta C_t \rangle$$

otherwise

Riemannian metric

- Metric respects the group action
  - Well-defined on the quotient manifold
- Orthogonal parameters
  - Standard Euclidean inner product
- Riemannian gradient

Derivatives

- Derivatives of a particular node with respect to its children can be computed efficiently
- The chain rule gives the gradient of the function ϕ

Riemannian gradient

Algorithm 2 The Riemannian gradient $\nabla_x^R f$ at a point $x = (U_t, B_t) \in \mathcal{M}_{\mathcal{T},\mathbf{k}}$

Require: $x = (U_t, B_t)$ parameter representation of the current point
  Compute $\mathbf{X} = \phi(x)$
  Compute $\nabla_x f(\mathbf{X})$, the Euclidean gradient of f
  $Z_{\mathrm{t_{root}}} \leftarrow \nabla_x f(\mathbf{X})$
  for $t \in T$, visiting parents before their children **do**
    if $t \in L$ **then**
      $\delta U_t = Z_t$
    else
      $Z_{t_l} = P^{\perp}_{U_{t_l}} \langle \delta U_{t_r}^T \circ_{k_r} Z_t, B \rangle_{(k_r,k_t),(k_r,k_t)}$
      $Z_{t_r} = P^{\perp}_{U_{t_r}} \langle \delta U_{t_l}^T \circ_{k_l} Z_t, B \rangle_{(k_l,k_t),(k_l,k_t)}$
      if $t = \mathrm{t_{root}}$ **then**
        $\delta B_t = (U_{t_l}^T, U_{t_r}^T) \circ Z_{\mathrm{t_{root}}}$
      else
        $\delta B_t = (P^{\perp}_{B^{(k_l,k_r)}} \circ_{k_l,k_r} I_{k_t} \circ_{k_t}) \circ (U_{t_l}^T \circ_{k_l} U_{t_r}^T \circ_{k_r} I_{k_t} \circ_{k_t}) \circ Z_t$
      end if
    end if
  end for
  return $\delta x \leftarrow (\delta U_t, \delta B_t)$

Gauss-Newton Hessian

$$\min_x \frac{1}{2}\|A\phi(x) - b\|_2^2$$

Linearize

$$\min_{dx} \frac{1}{2}\|AD\phi(x_k)dx - b\|_2^2$$

Gauss-Newton Hessian

Solution

$$D\phi(x)^* A D\phi(x)\delta x = -D\phi(x)^* r(x)$$

Ill-conditioned

Instead solve

$$D\phi(x)^* D\phi(x)\delta x = -D\phi(x)^* r(x)$$

Gauss-Newton Hessian

$$D\phi(x) : \mathcal{H}_x M \to \mathbb{R}^{n_1 \times \cdots \times n_d}$$

$$D\phi(x)^* : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathcal{H}_x M$$

$$H_{GN} = D\phi(x)^* \circ D\phi(x) : \mathcal{H}_x M \to \mathcal{H}_x M$$

Gauss-Newton Hessian

$$H_{GN}(\delta U_t, \delta B_t) = (\delta U_t G_t, G_t \circ_3 B_t)$$

G_t are the spd *Gramian* matrices, which satisfy

$$\lambda_i(G_t) = \lambda_i(X^{(t)}(X^{(t)})^T)$$

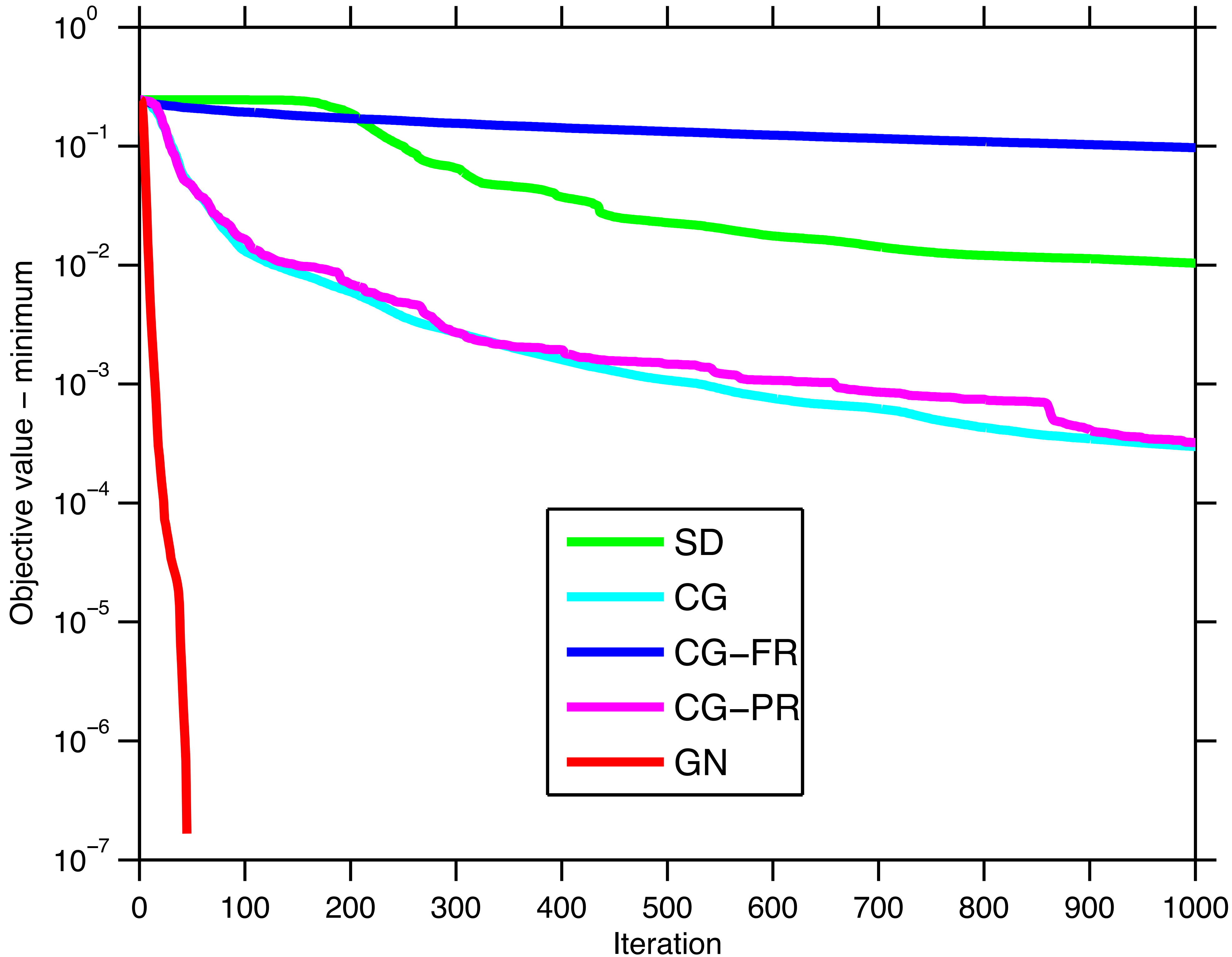
Christine Tobler, Low Rank Tensor Methods for Linear Systems and Eigenvalue Problems, PhD Thesis, ETH Zürich, 2012

1. R3MC: A Riemannian three-factor algorithm for low-rank matrix completion. B. Mishra and R. Sepulchre, Tech Report, University of Liège, June 2013.
2. B. Mishra, K. Adithya Apuroop, and R. Sepulchre. A Riemannian geometry for low-rank matrix completion, 2012.

Gauss-Newton Hessian

- Much faster to apply, invert H_{GN} than using the naive method
  - No large intermediate vectors
- This is the approach in [1], [2] for d=2
  - Explains why their preconditioning scheme works

Convergence



Results

Synthetic BG Data

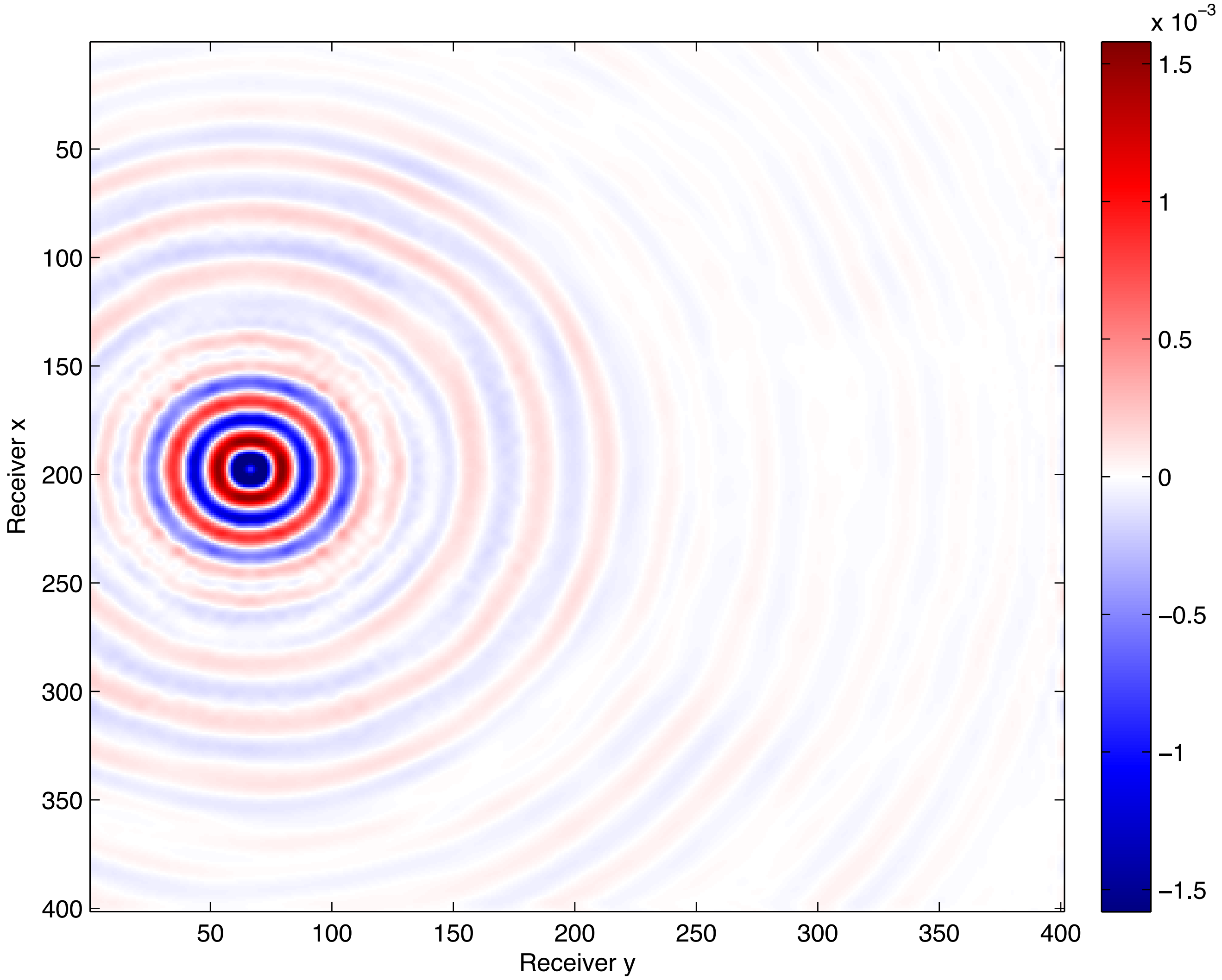
- Unknown model
- 68 x 68 sources with 401 x 401 receivers, data at 4.68Hz, 7.34 Hz
  - Receivers subsampled, Fourier interpolated back to 401 x 401 to produce figures

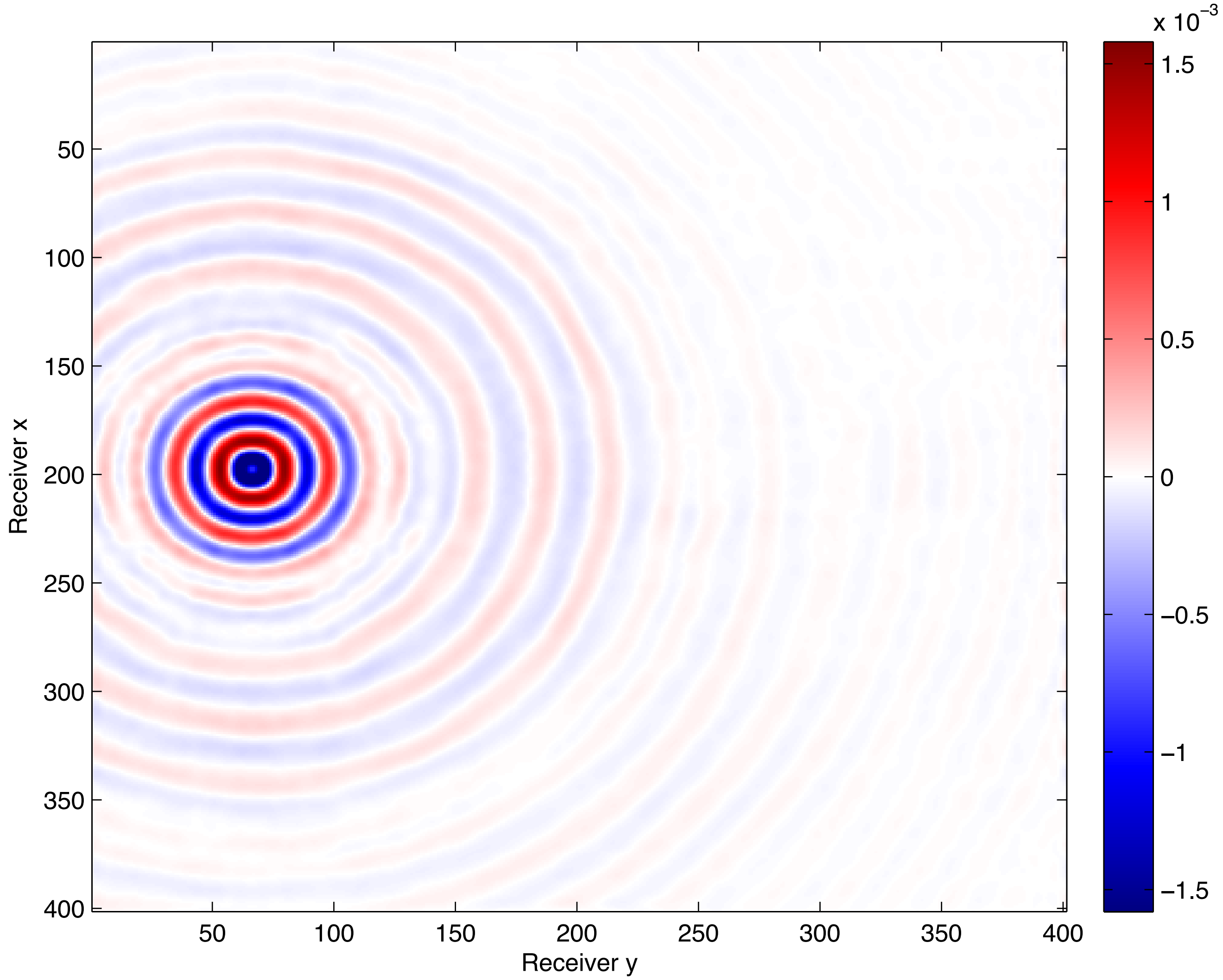
Synthetic BG Data

- Single frequency slice (real part) - scaled to unit norm
- Varying percentages of sources have been randomly removed
- Recovered with GN, CG

4.86 Hz - 50% missing sources

Fixed source coordinates - no data originally



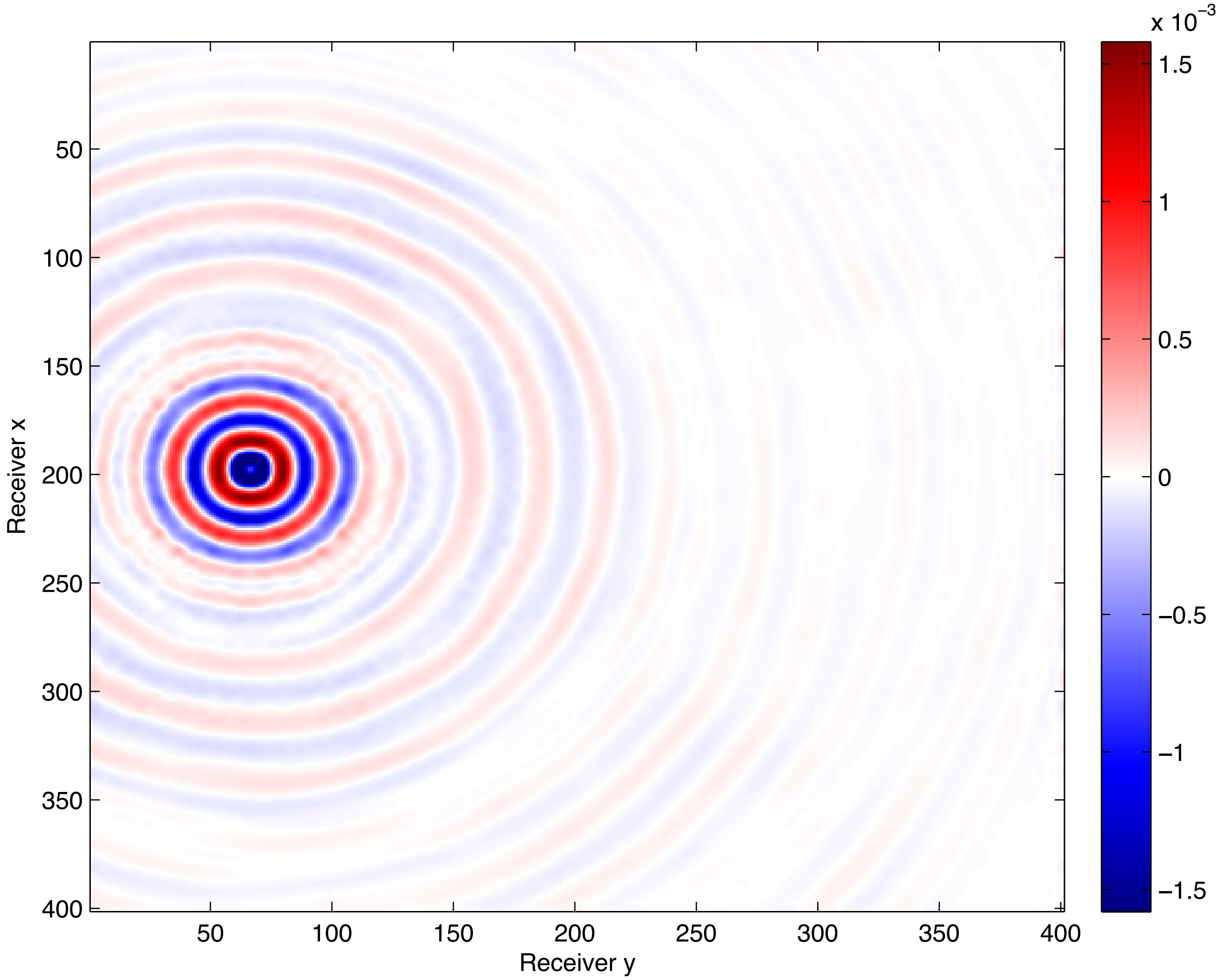


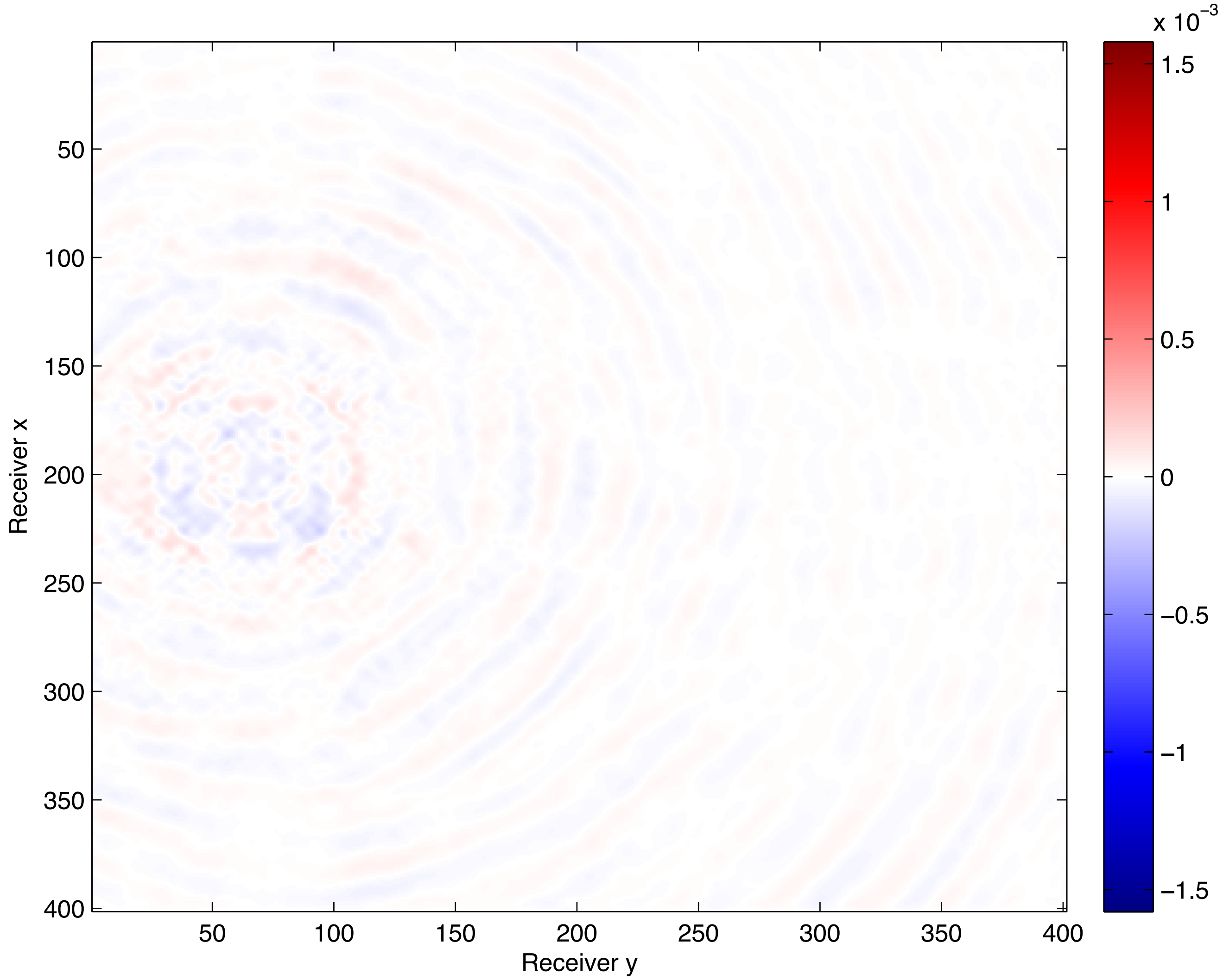
$(x_{src}, y_{src}) = (34,17)$

Interpolated Data
SNR 17.5 dB

4.86 Hz - 50% missing sources

Fixed source coordinates - no data originally



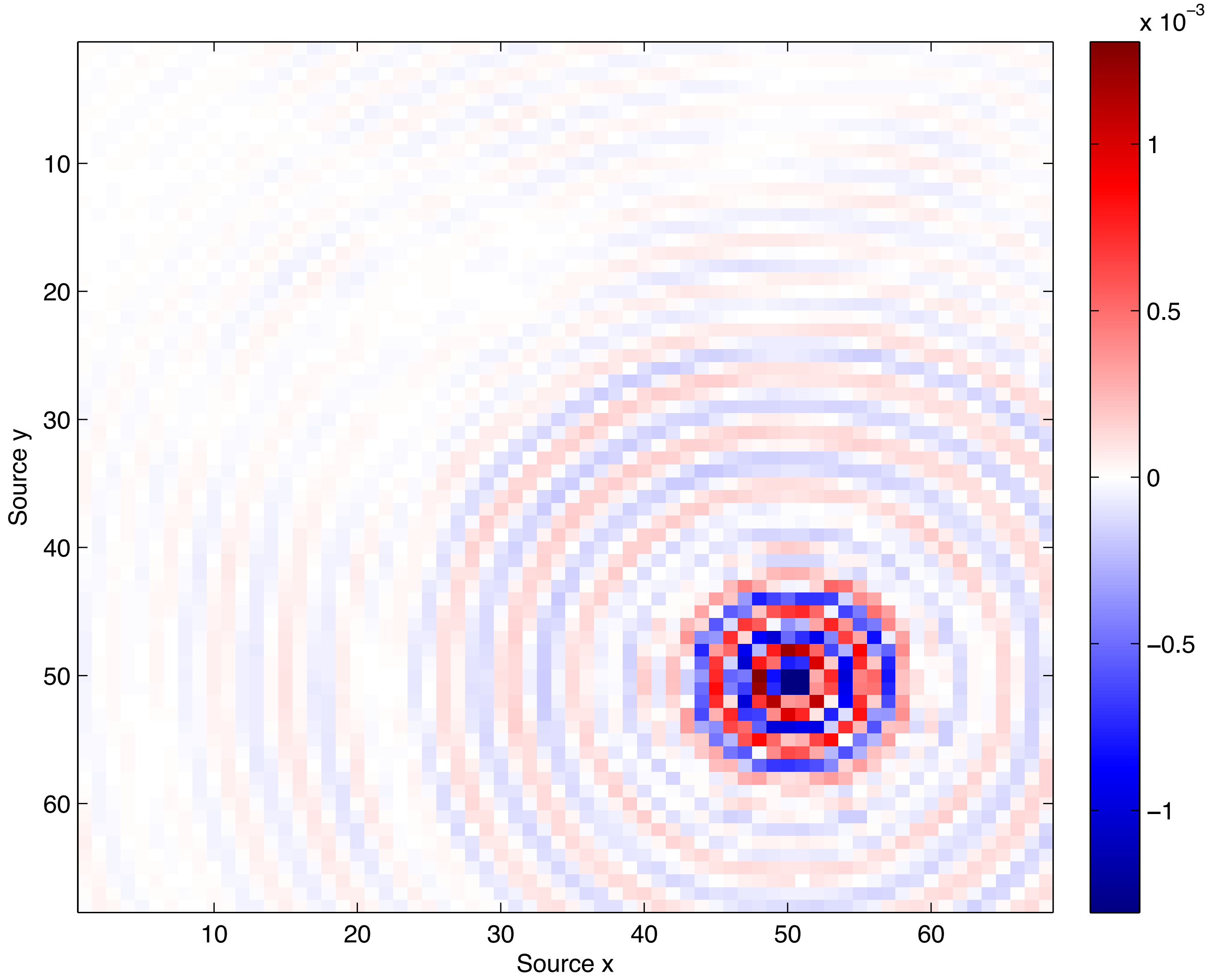


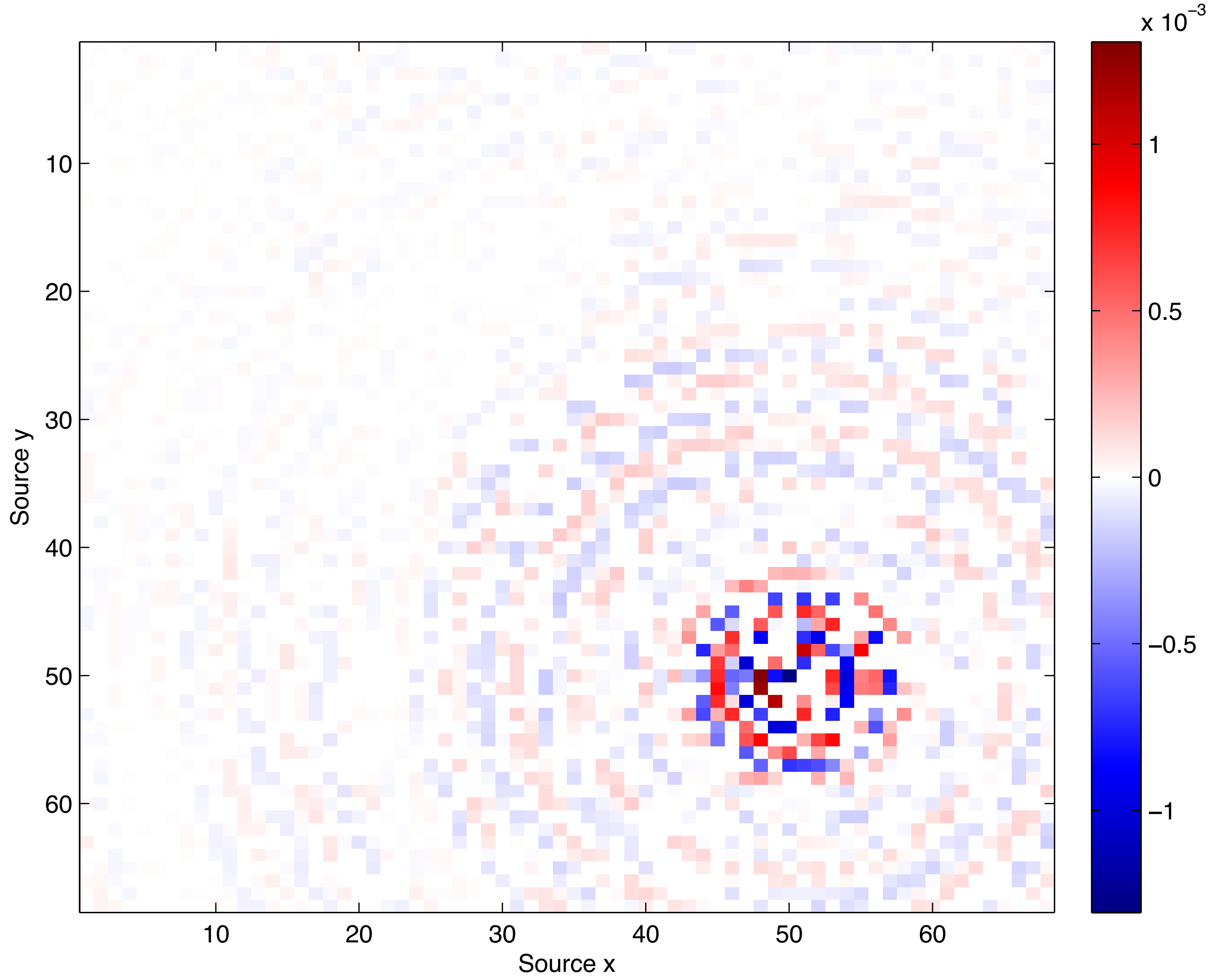
$(x_{src}, y_{src}) = (34,17)$

Difference

4.86 Hz - 50% missing sources

Fixed receiver coordinates



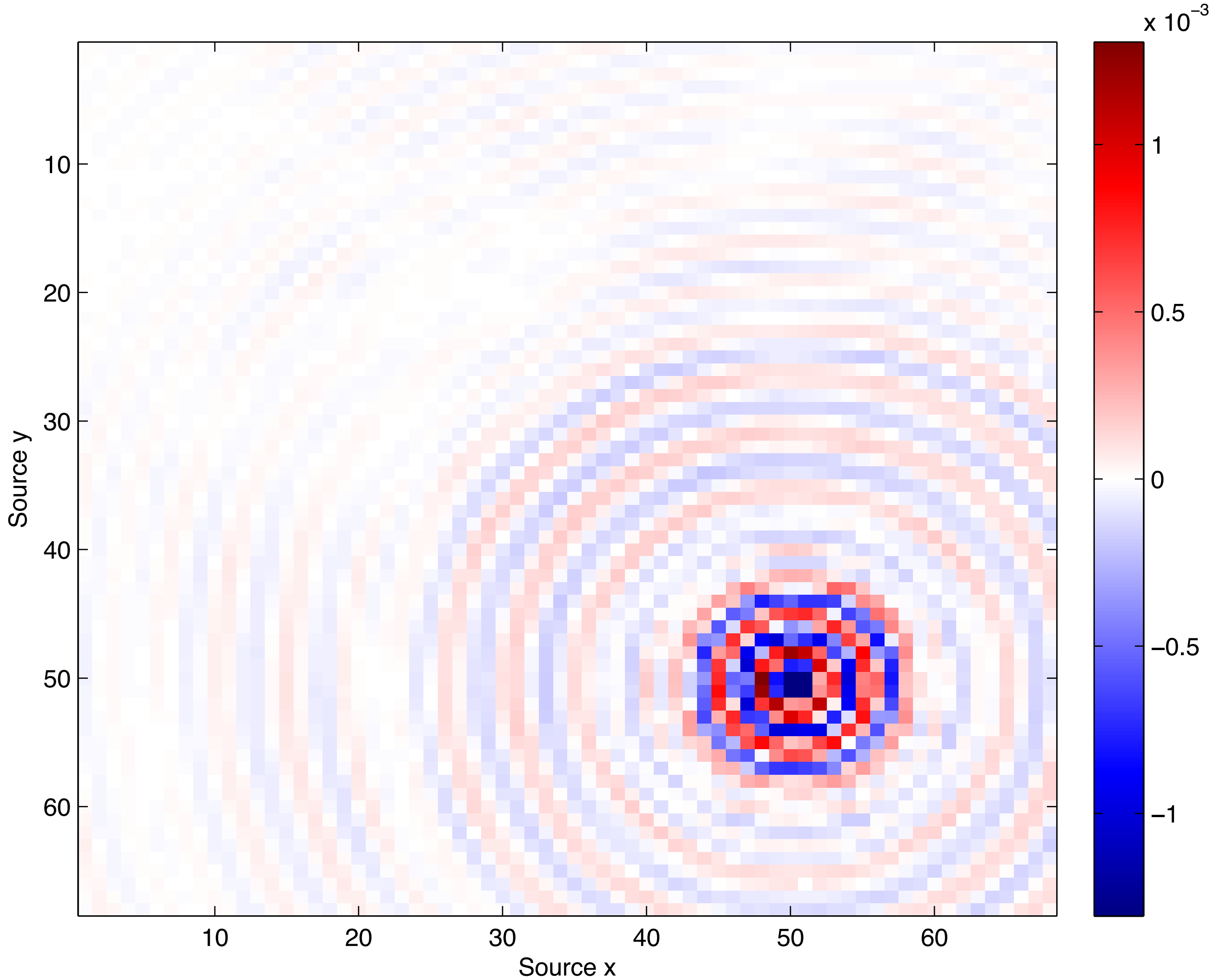


$(x_{rec}, y_{rec}) = (75,75)$

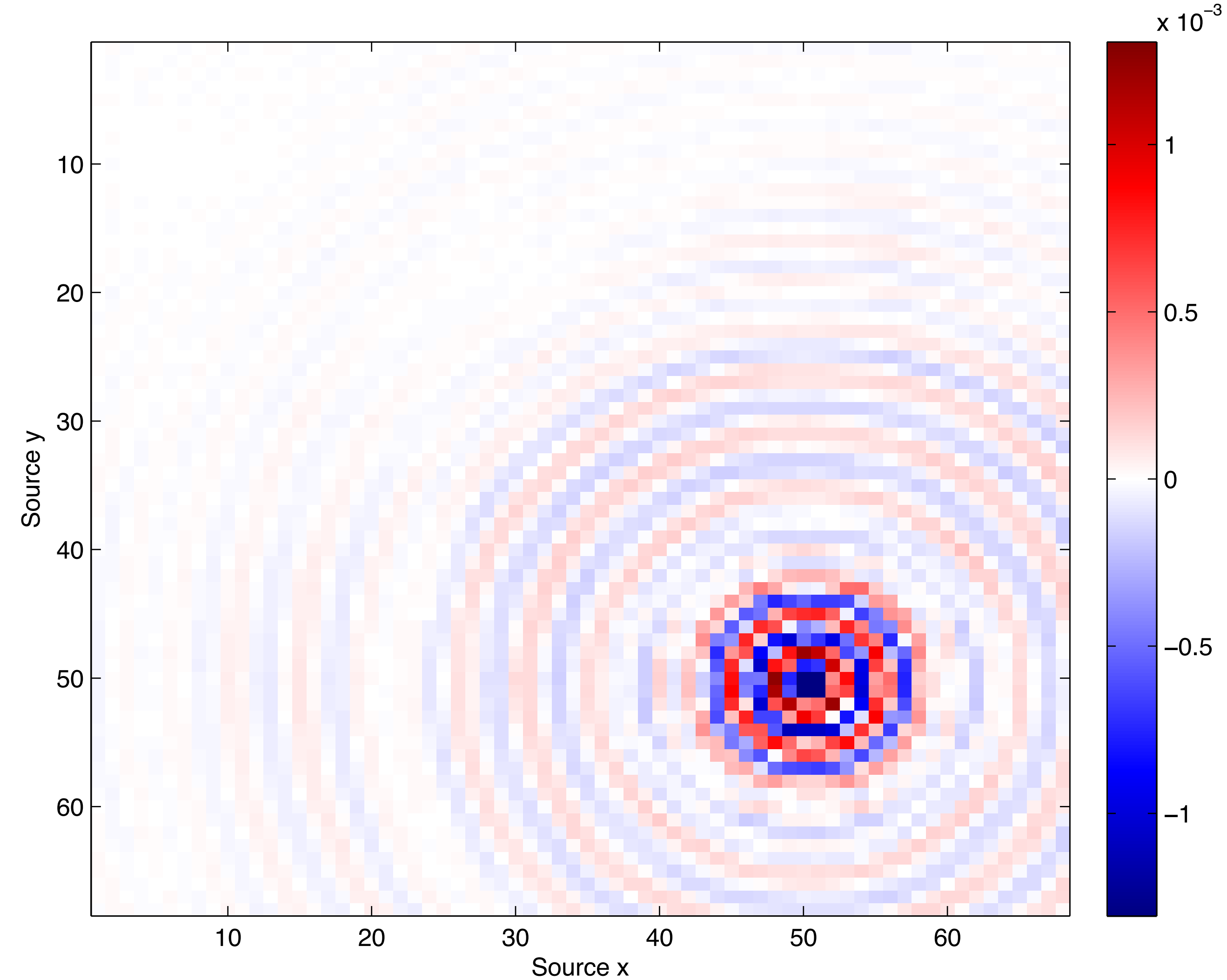
Subsampled Data

4.86 Hz - 50% missing sources

Fixed receiver coordinates



$(x_{rec}, y_{rec}) = (75,75)$

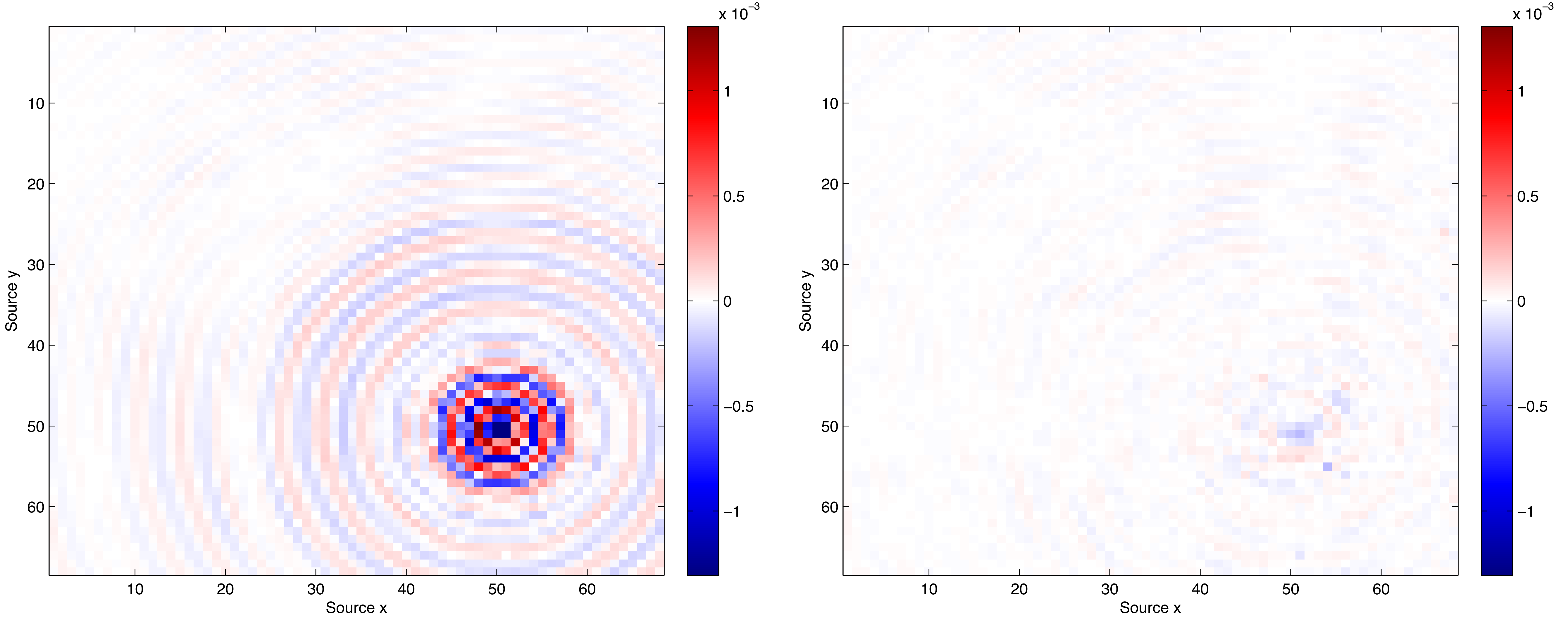


Interpolated Data

SNR 17.3 dB

4.86 Hz - 50% missing sources

Fixed receiver coordinates

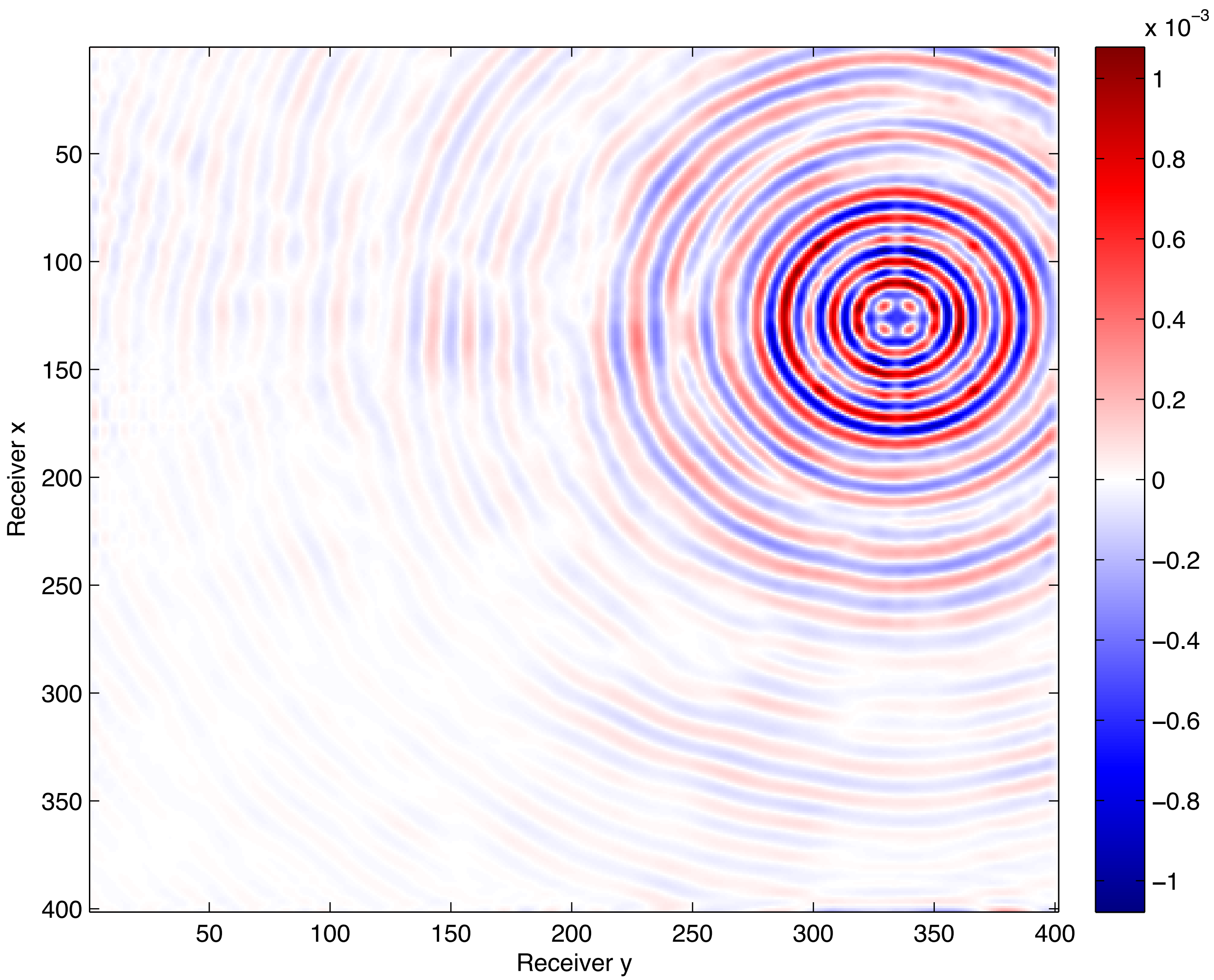


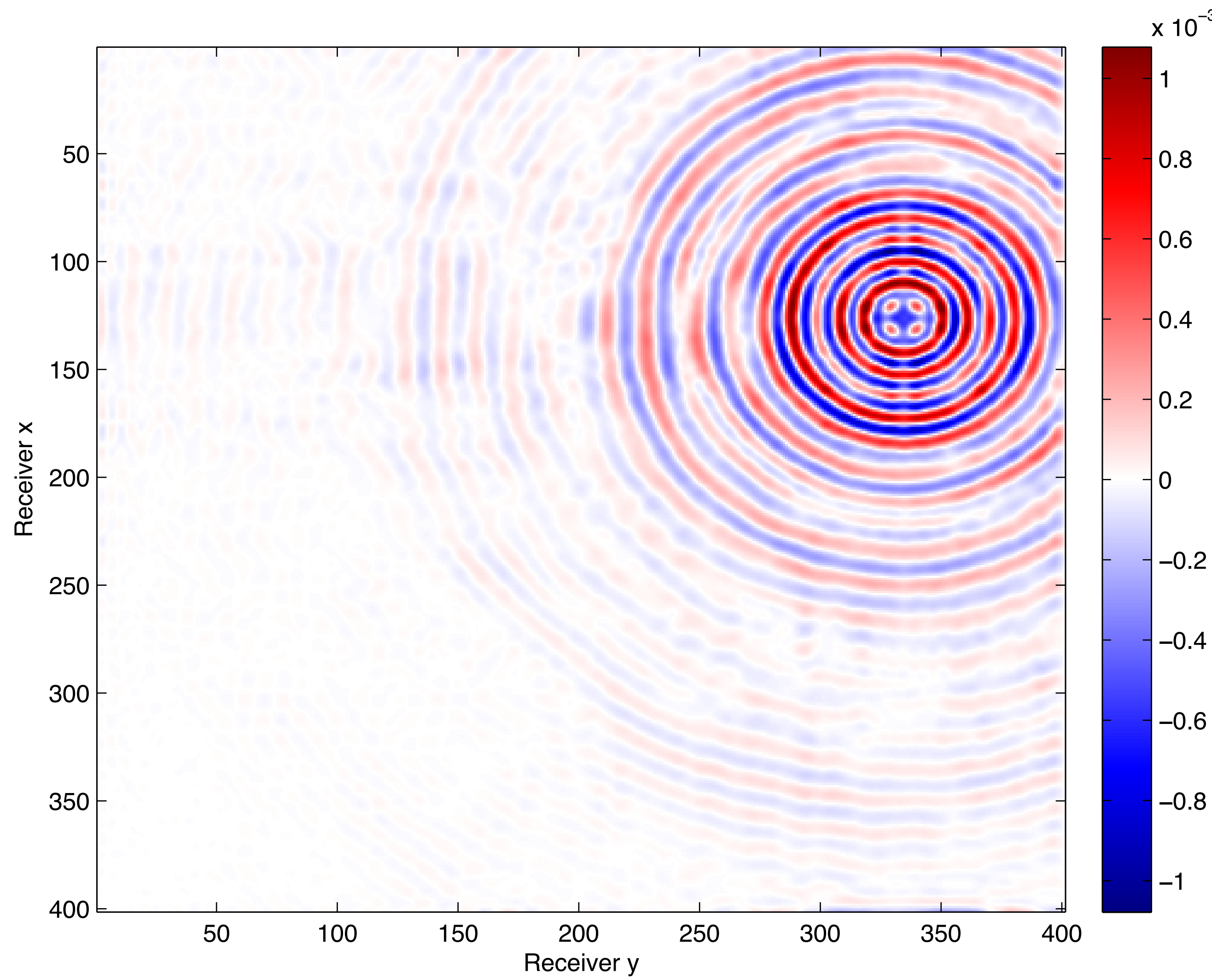
$(x_{rec}, y_{rec}) = (75,75)$

Difference

7.34 Hz - 75% missing sources

Fixed source coordinates - no data originally





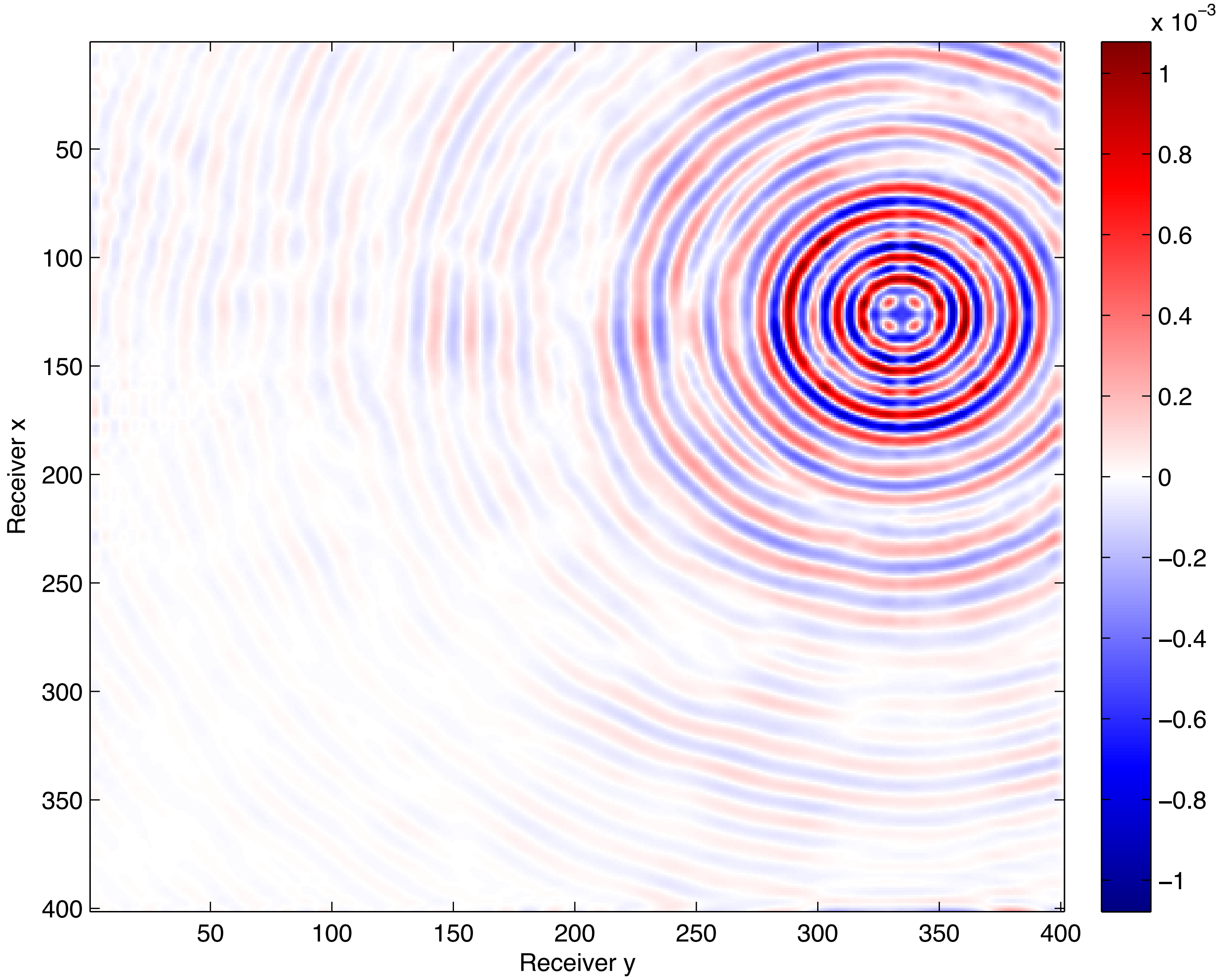
$(x_{src}, y_{src}) = (22,57)$

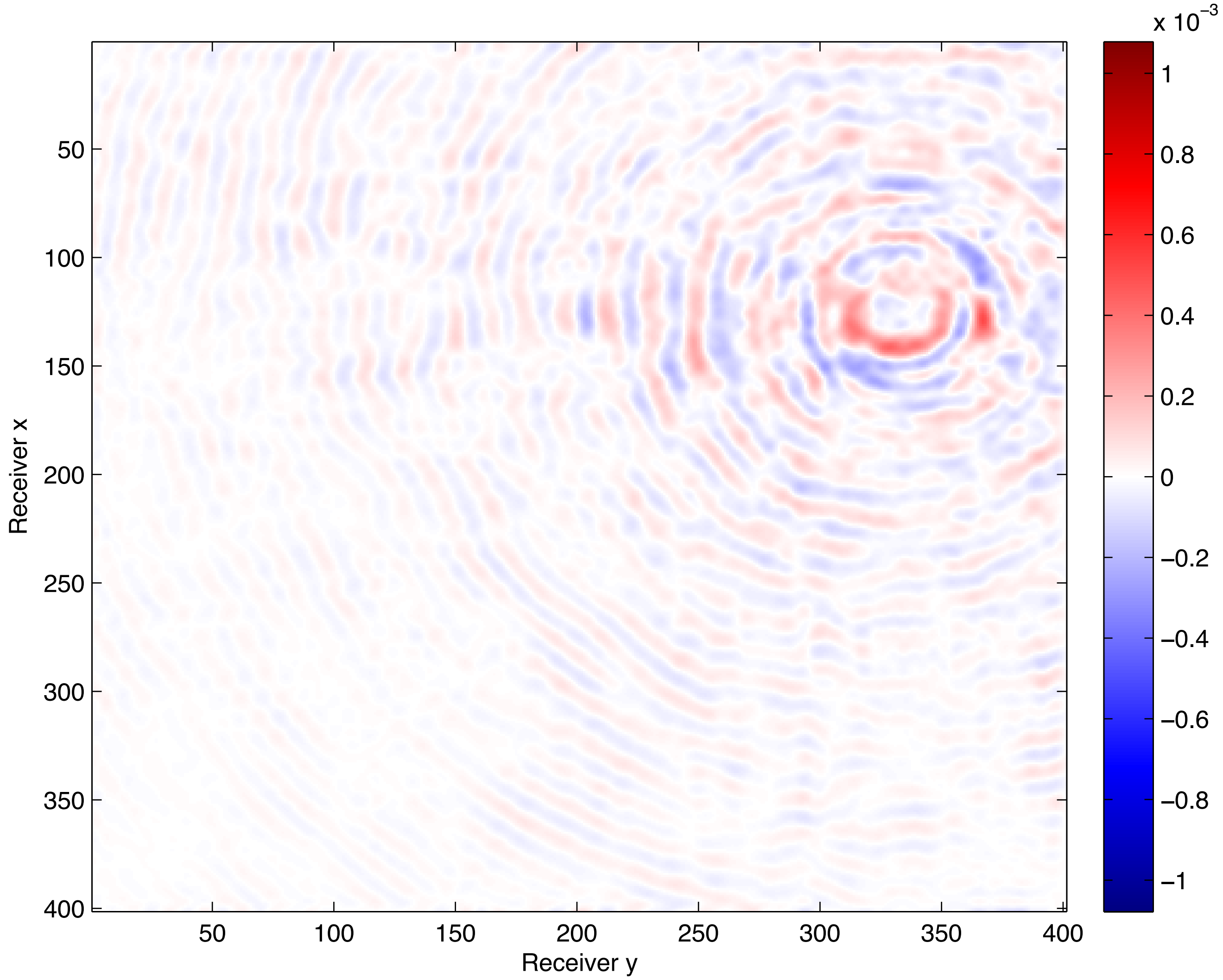
Interpolated Data

SNR 10.6 dB

7.34 Hz - 75% missing sources

Fixed source coordinates - no data originally



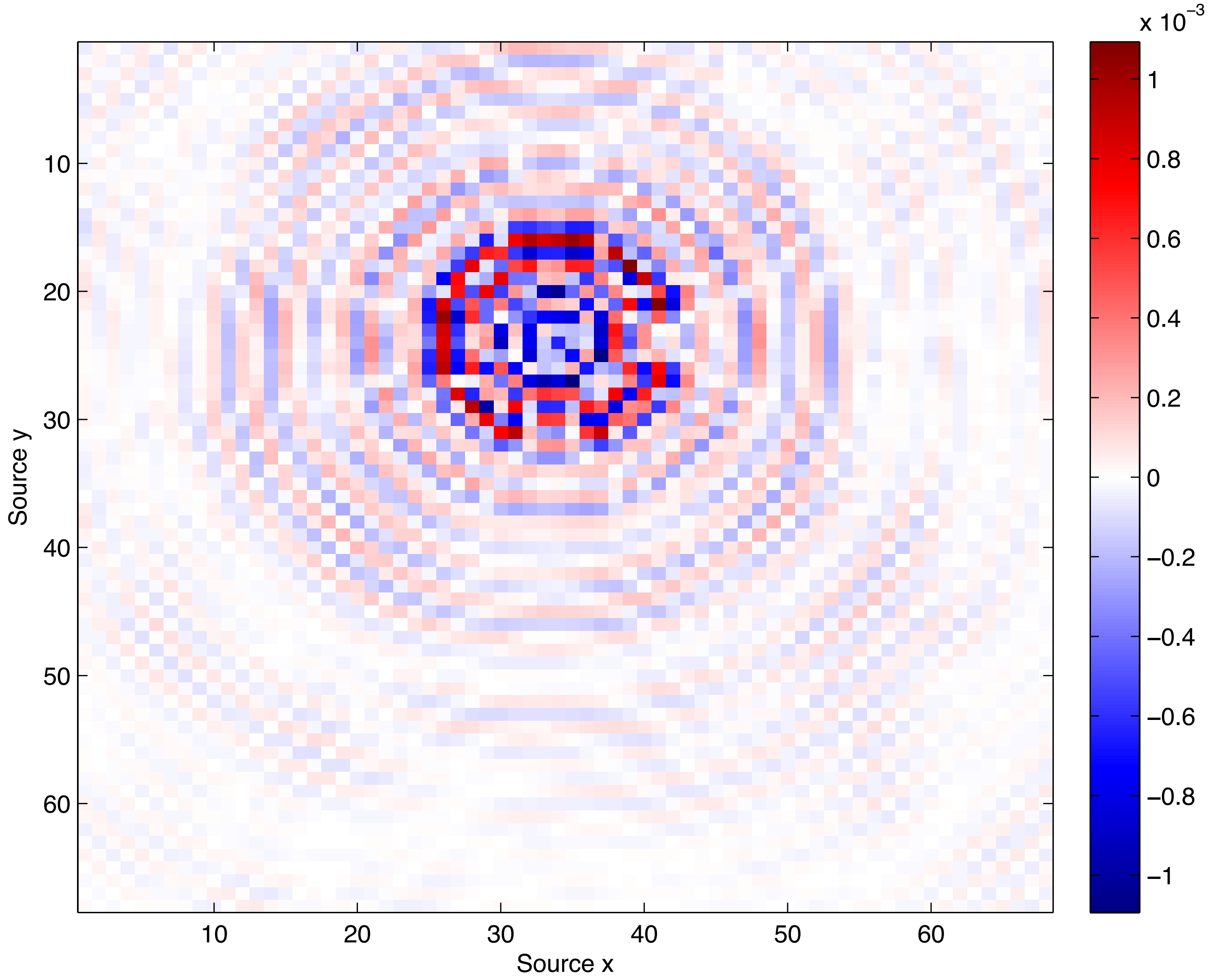


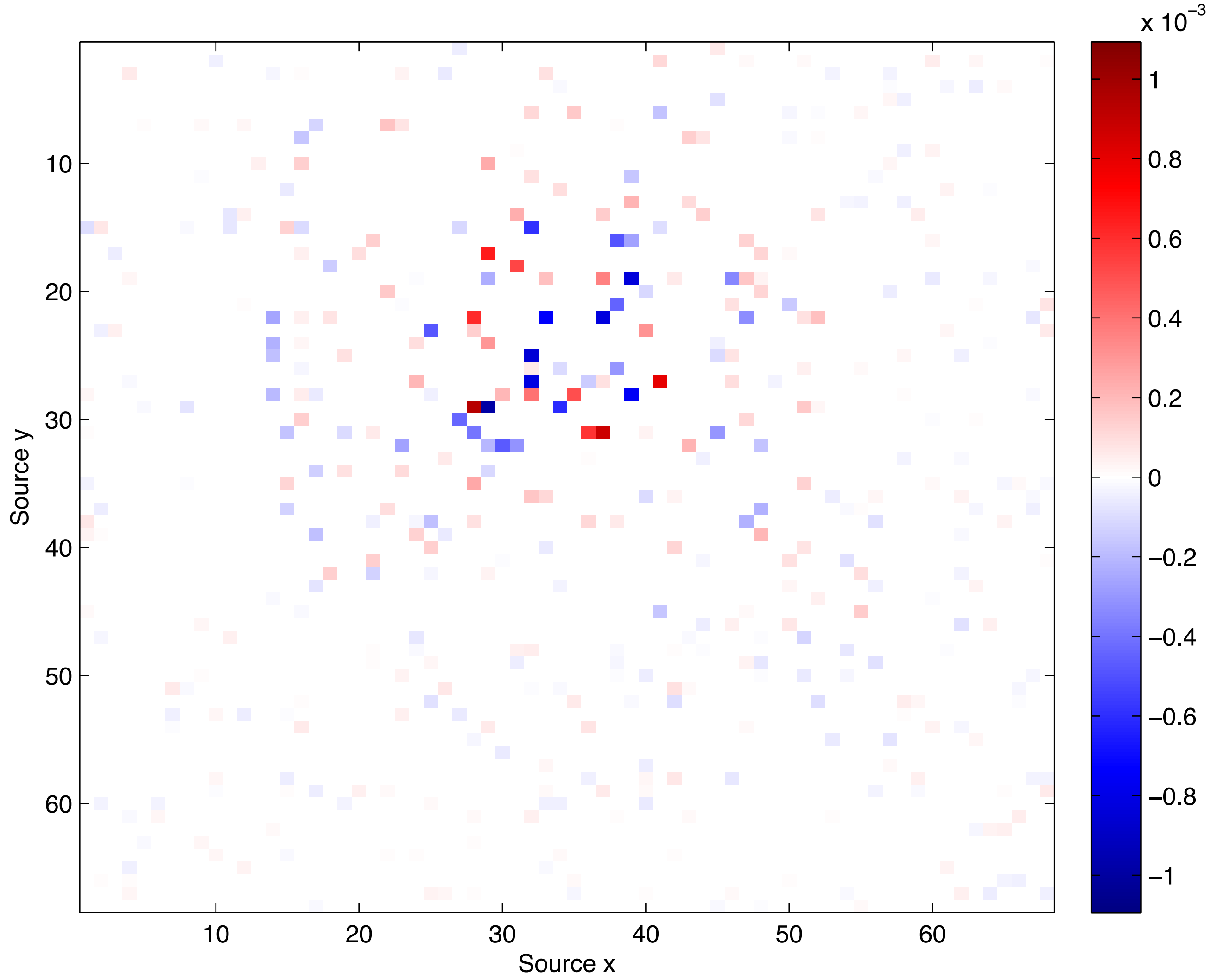
$(x_{src}, y_{src}) = (22,57)$

Difference

7.34 Hz - 75% missing sources

Fixed receiver coordinates



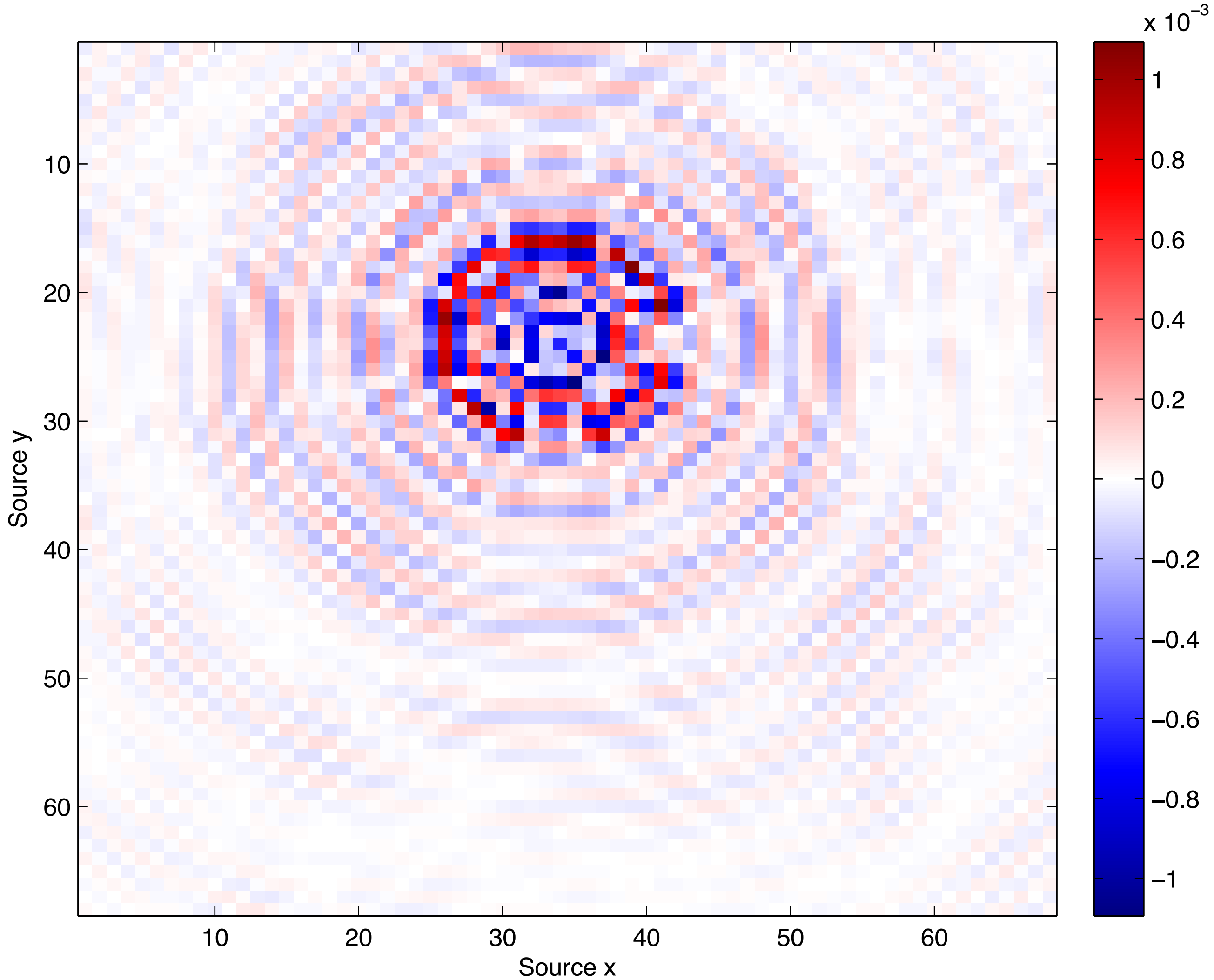


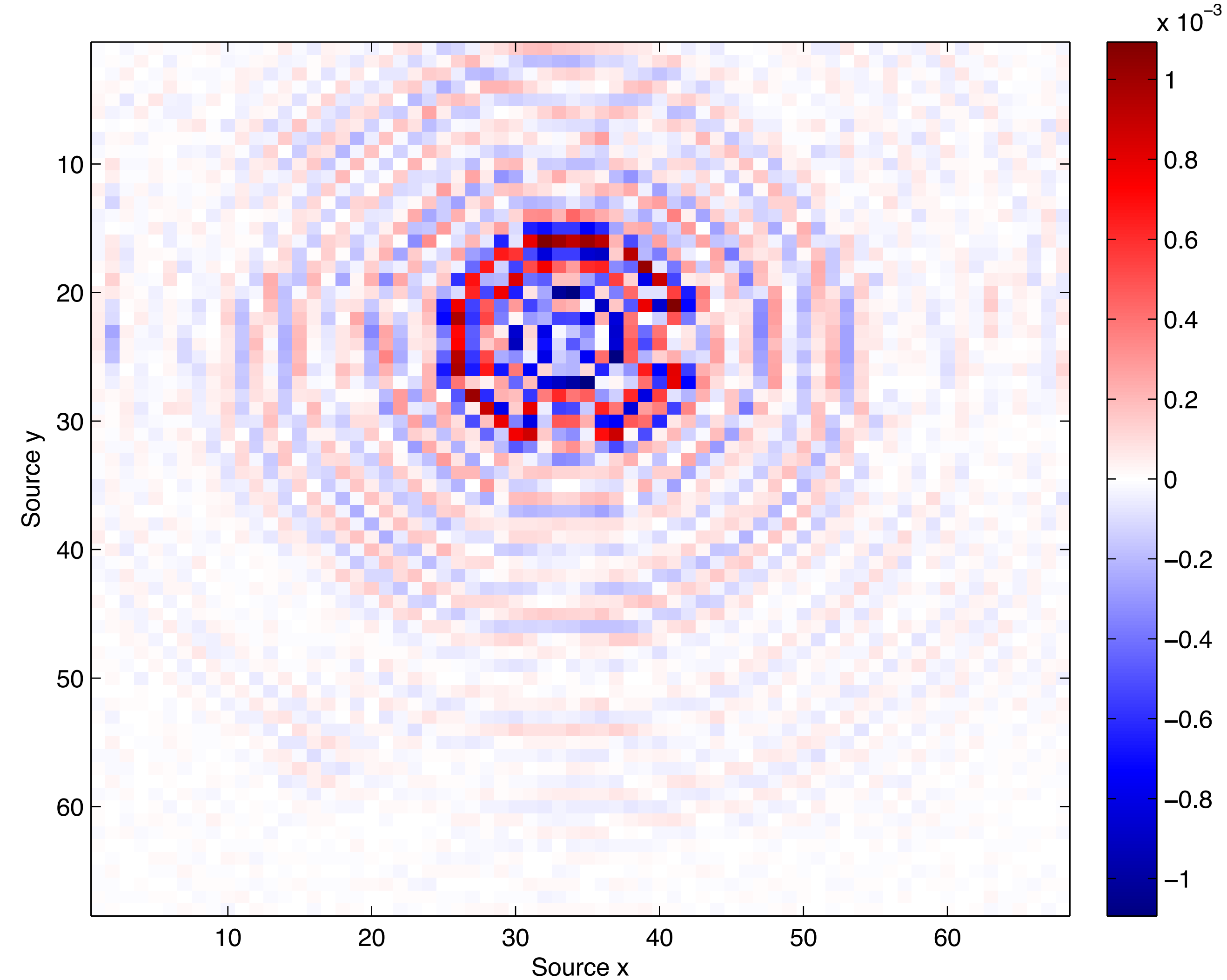
$(x_{rec}, y_{rec}) = (35,50)$

Subsampled Data

7.34 Hz - 75% missing sources

Fixed receiver coordinates





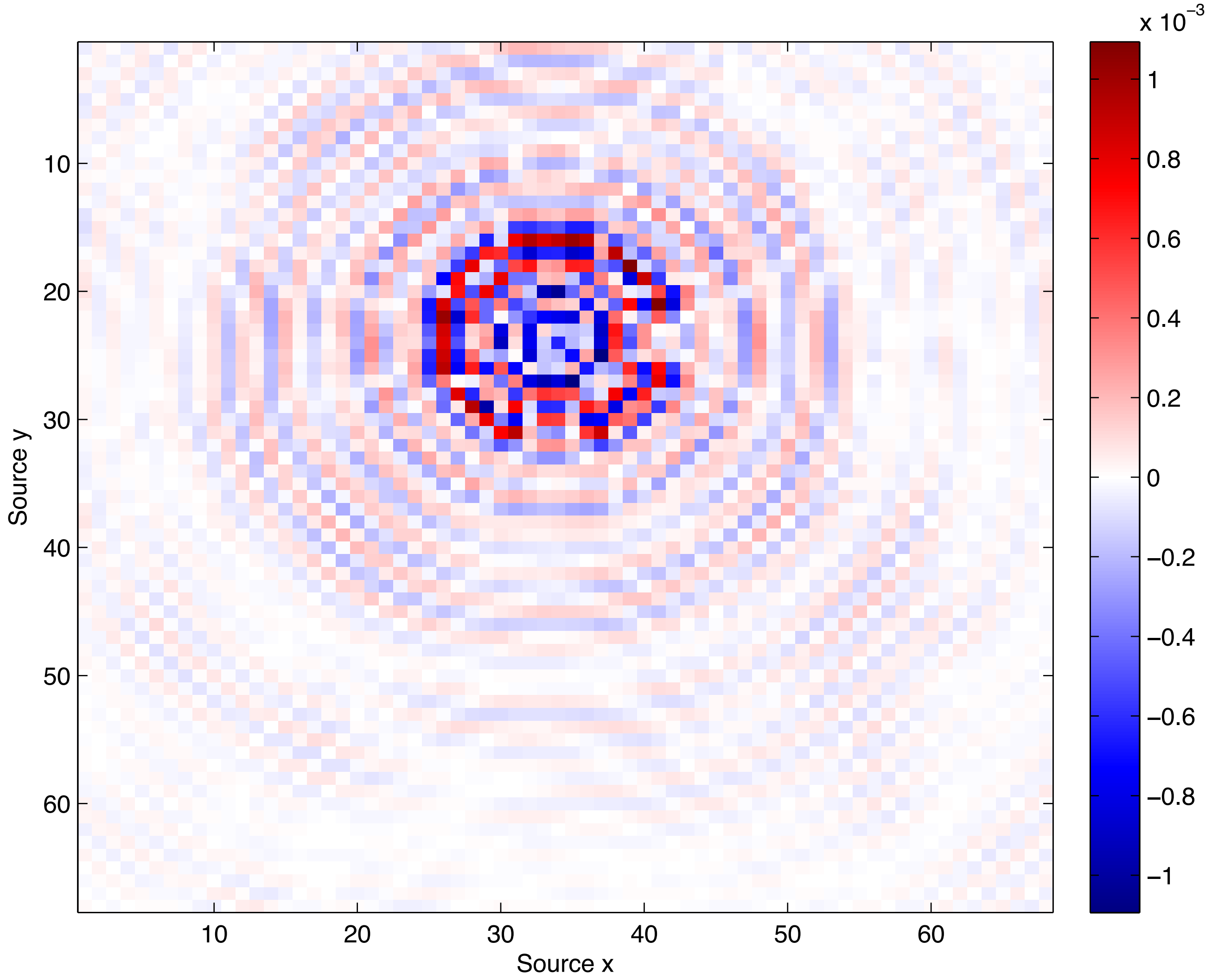
$(x_{rec}, y_{rec}) = (35,50)$

Interpolated Data

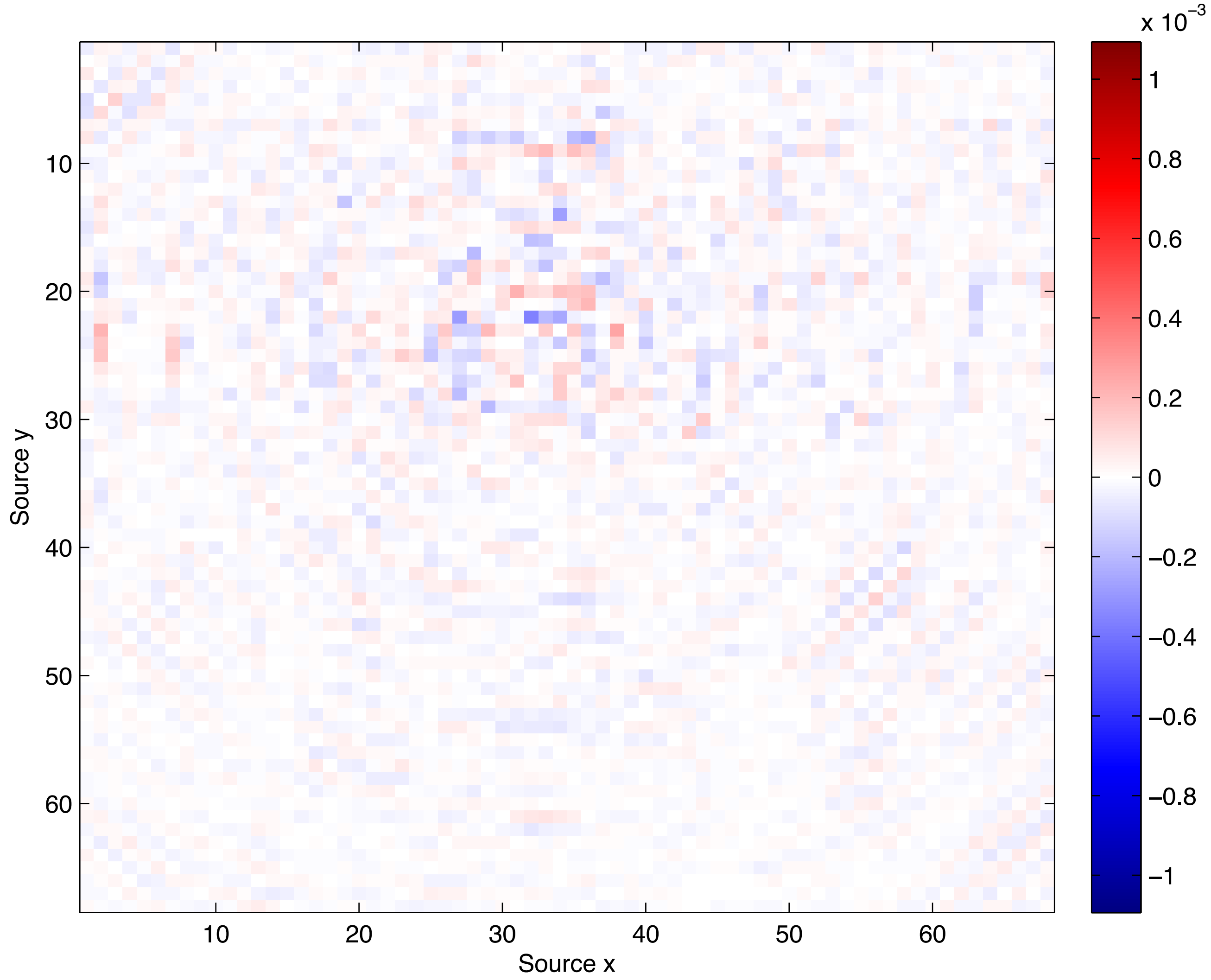
SNR 12.1 dB

7.34 Hz - 75% missing sources

Fixed receiver coordinates



$(x_{rec}, y_{rec}) = (35,50)$



Difference

SLIM

Summary - SNR - 4.86 Hz

% Missing srcs	SNR Fit (dB)	SNR Recovered (dB)	SNR Recovered LMAFit (dB)
25%	20.3	**18.9**	14.4
50%	18.0	**17.0**	14.4
75%	19.5	**16.2**	7.6

Summary - SNR - 7.34 Hz

% Missing srcs	SNR Fit (dB)	SNR Recovered (dB)	SNR Recovered LMAFit (dB)
25%	15.9	**14.4**	9.4
50%	16.8	**14.1**	9.4
75%	16.4	**11.9**	9.1

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
  - Efficient optimization framework
- Achieve good results from largely subsampled data (75% missing sources)

Acknowledgements

Thank you for your attention

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, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.