Released to public domain under Creative Commons license type BY (https://creativecommons.org/licenses/by/4.0).
Copyright (c) 2013 SLIM group @ The University of British Columbia.

London '13

Seismic data interpolation and denoising using SVD-free low-rank matrix factorization

Rajiv Kumar, Sasha Aravkin, Hassan Mansour, Ben Recht and Felix J. Herrmann

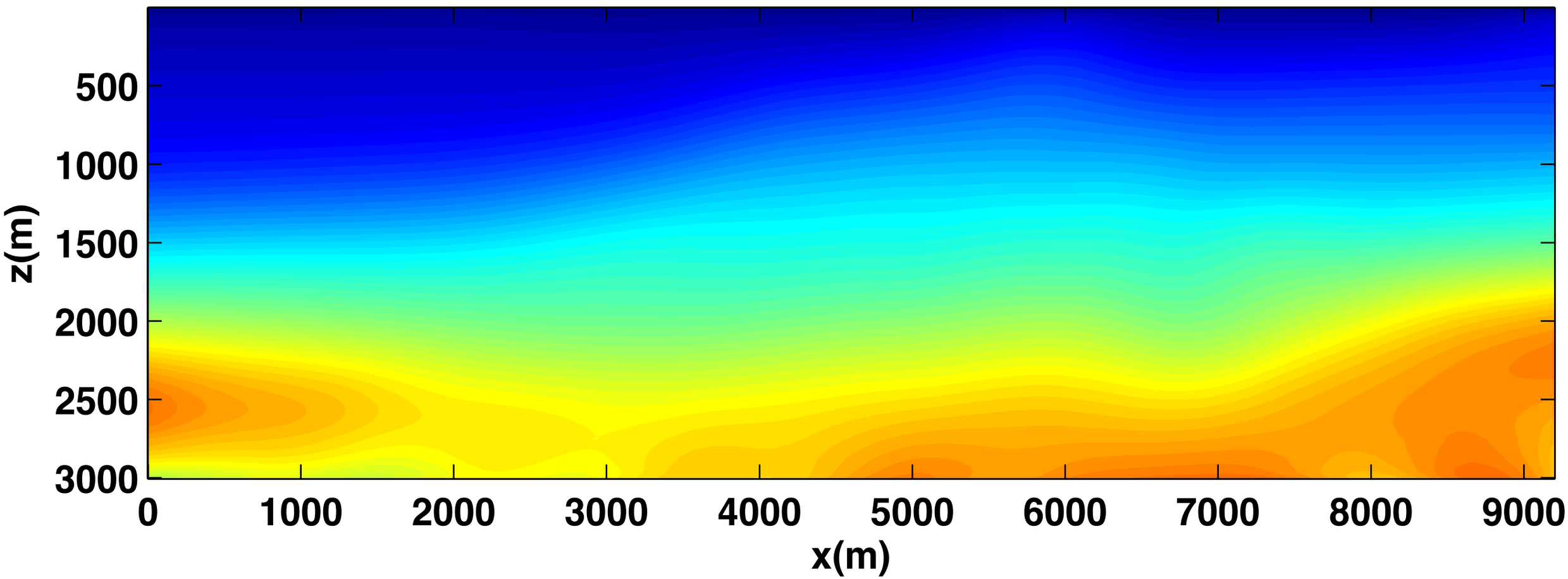
SLIM
University of British Columbia

Motivation

- acquisition challenges
  - missing data
  - noise
- fully sampled data
  - simultaneous shot based FWI & migration
  - estimation of primaries by sparse inversion & SRME
- exploit *low-rank* structure of seismic data
  - randomized sampling
  - *SVD-free* matrix factorization

Full waveform inversion

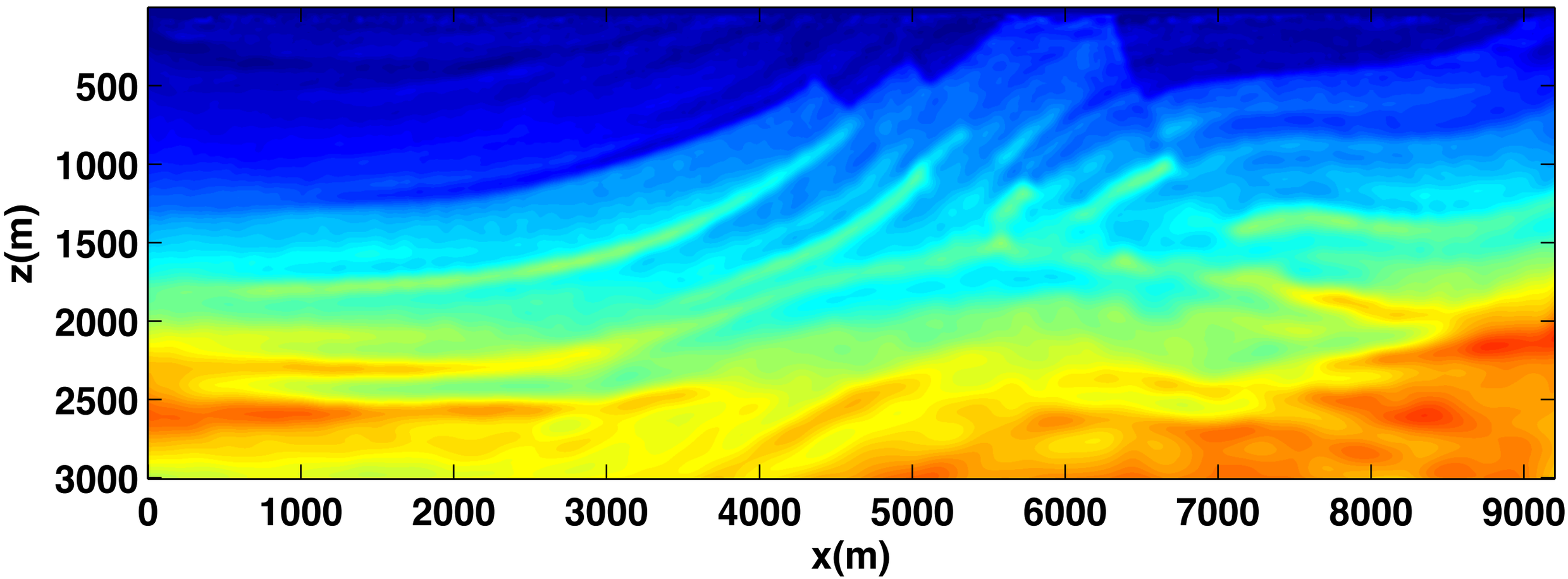
initial model



[Li et. al. 2012, van Leeuwen and Herrmann 2012]

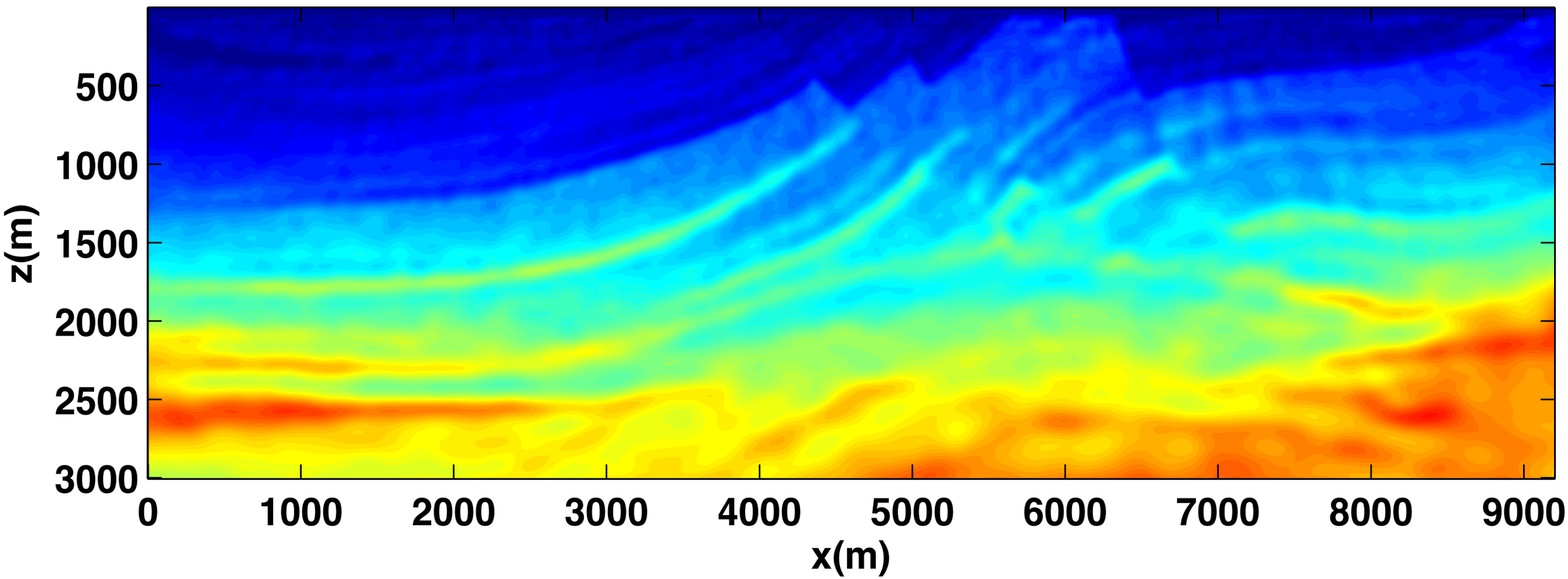
Full waveform inversion

[Ideal scenario, 60 sequential shots, redraw]



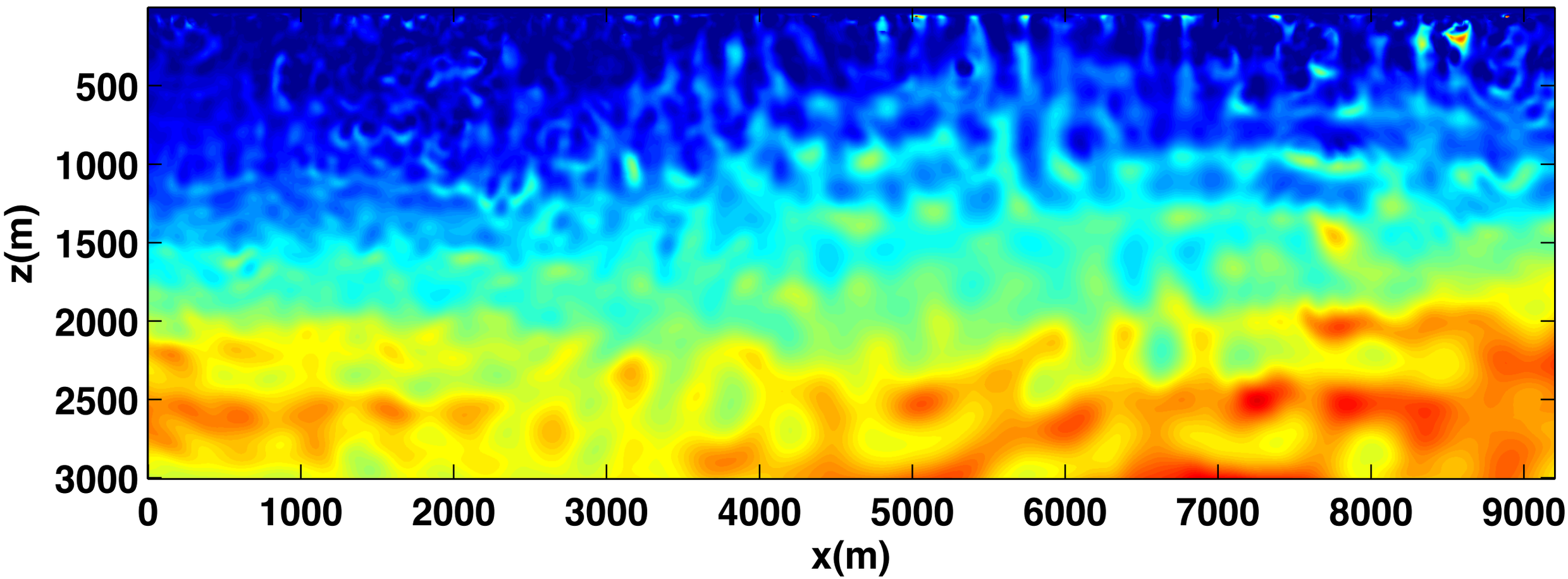
Full waveform inversion

[Ideal scenario, 20 simultaneous shots, redraw]



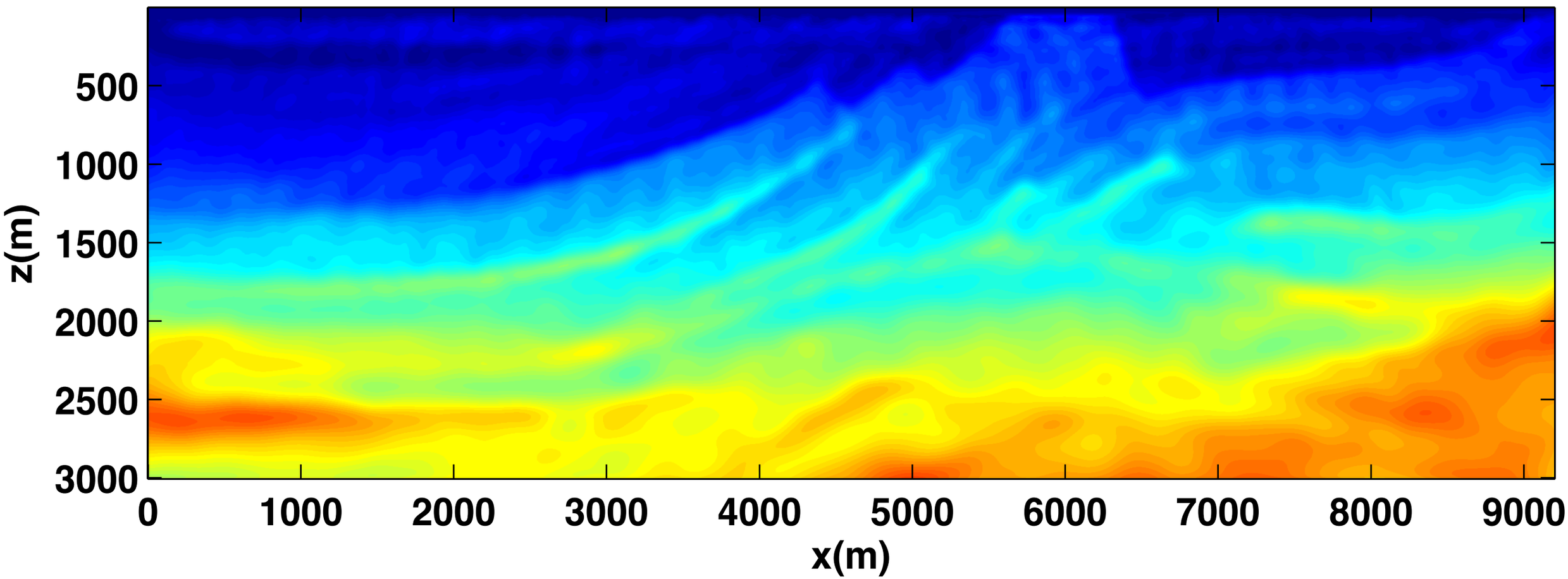
Full waveform inversion

[50% random missing shots, 60 sequential shots, redraw]



Full waveform inversion

[low-rank interpolation, 20 simultaneous shots, redraw]



[Candes and Donoho 2000, Herrmann 2008]

Compressive sensing

- signal structure
  - *sparse/compressible*
- sampling scheme
  - random missing traces make signal *less sparse* in transform domain
- recovery using *sparsity promoting* scheme

- **is *sparsity* the only inherent structure in seismic??**

[Candes and Plan 2010, Oropeza and Sacchi 2011]

Matrix completion

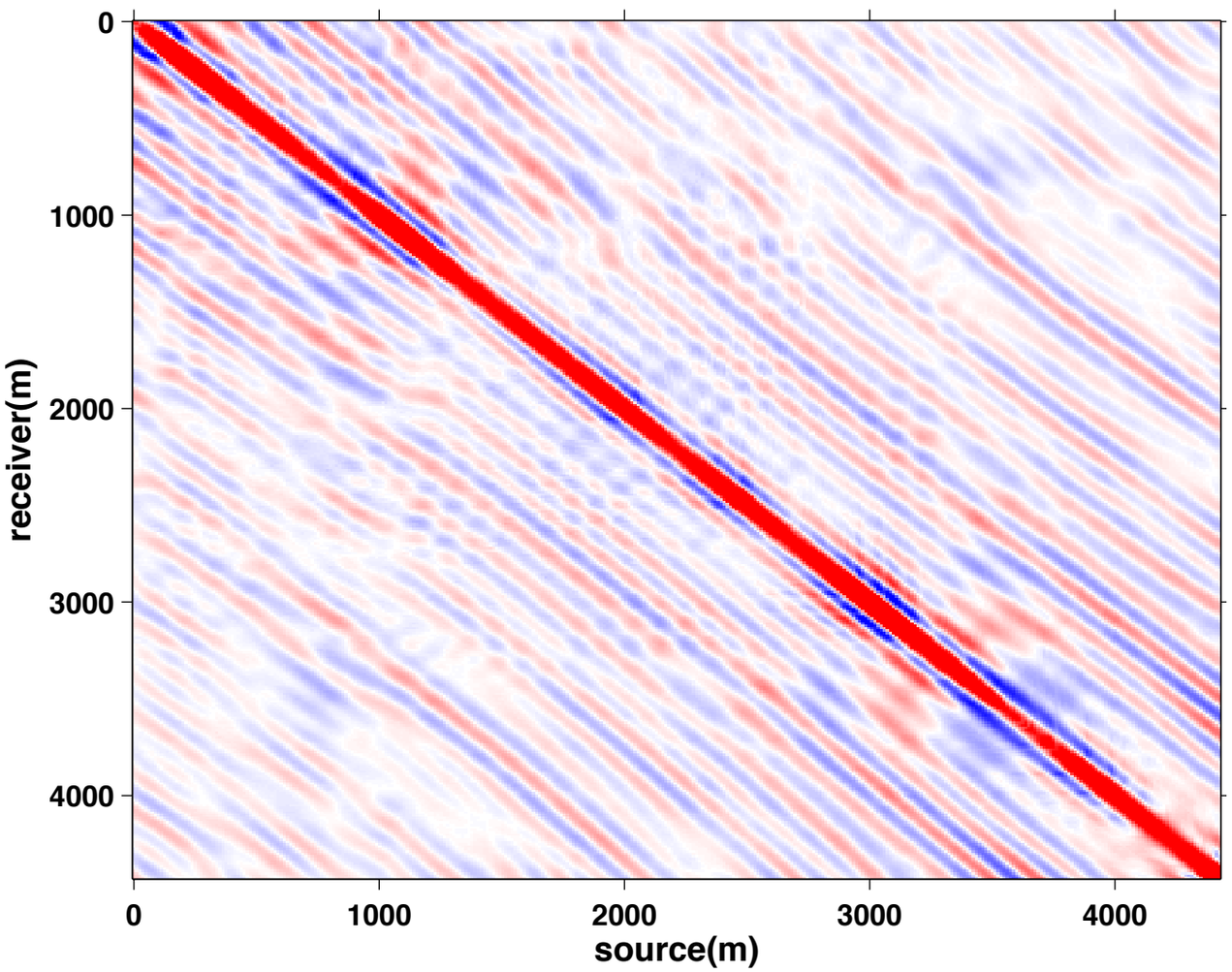
- signal structure
  - *low rank/fast decay* of singular values
- sampling scheme
  - missing data *increase* rank in "transform domain"
- recovery using *rank penalization* scheme

Low-rank structure

2-D acquisition

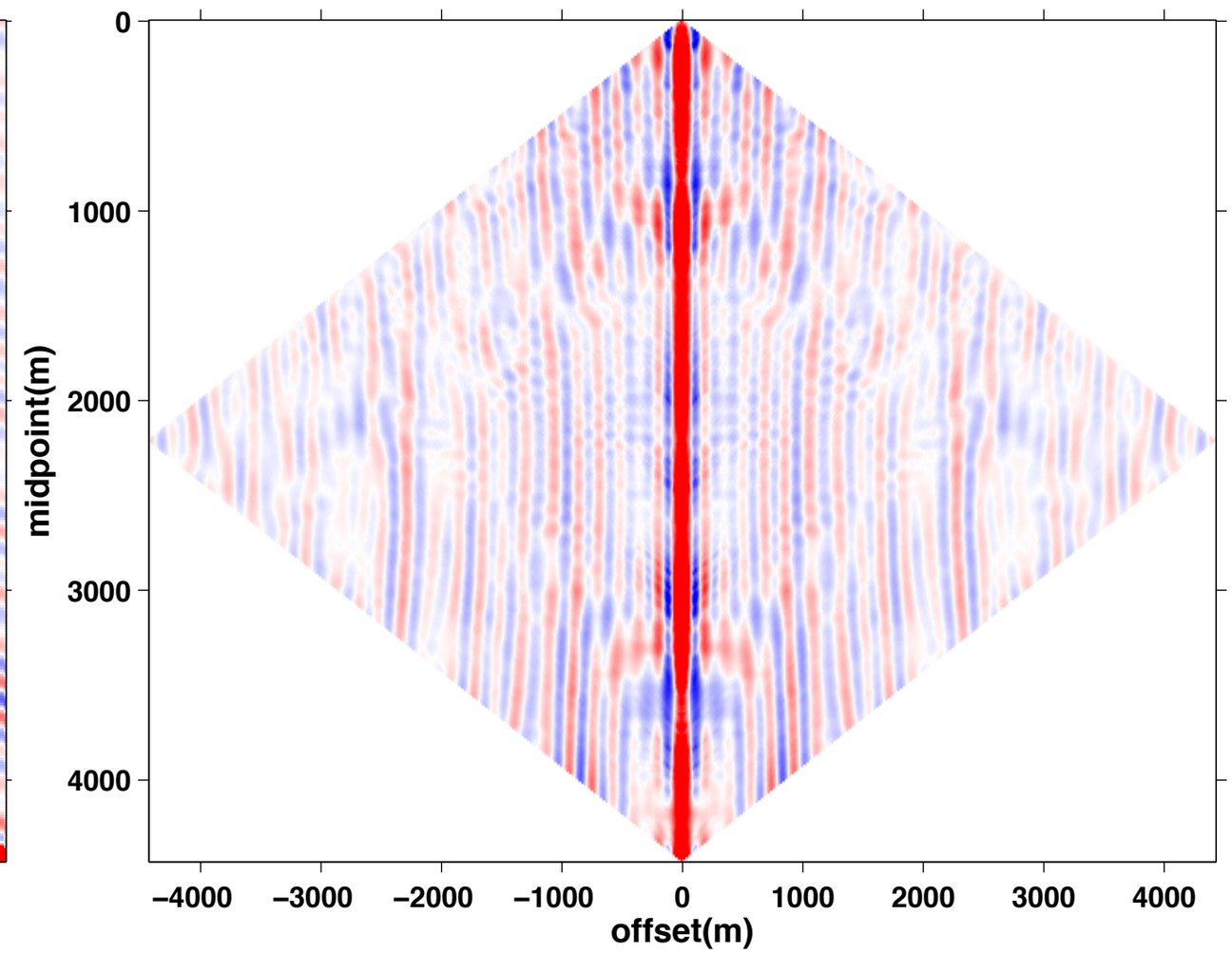
acquisition domain

[source-receiver]



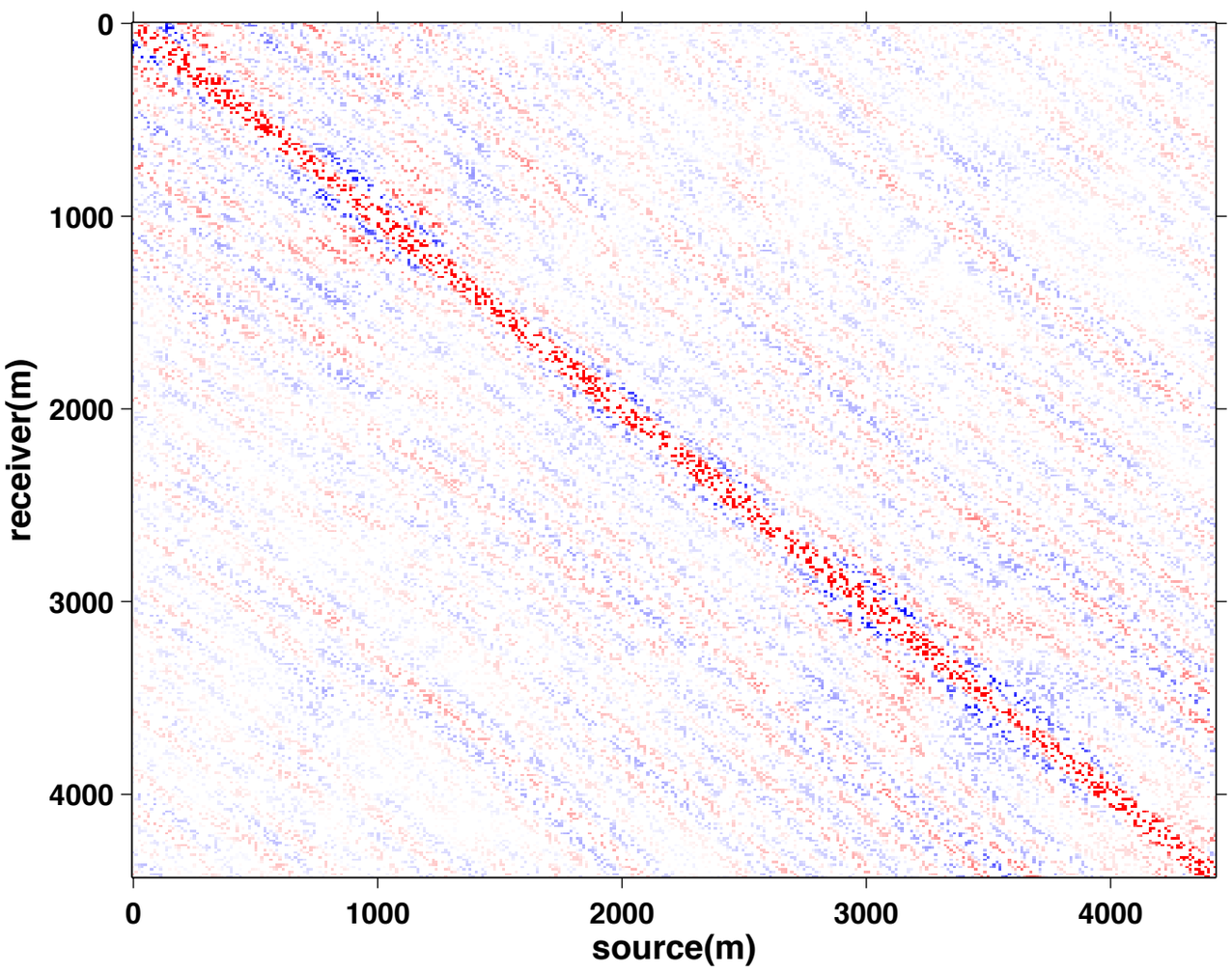
transform domain

[midpoint-offset]

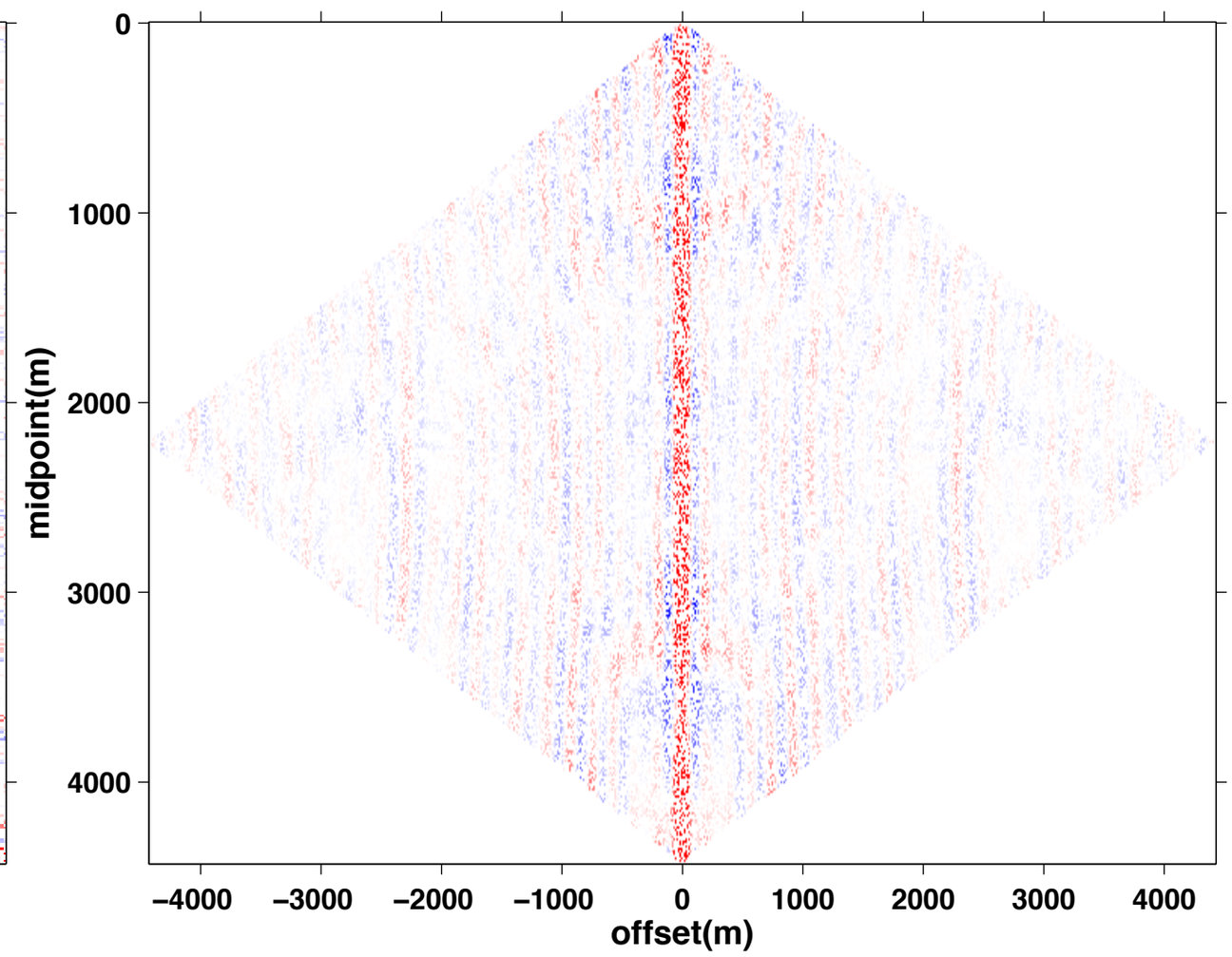


Matrix completion problem

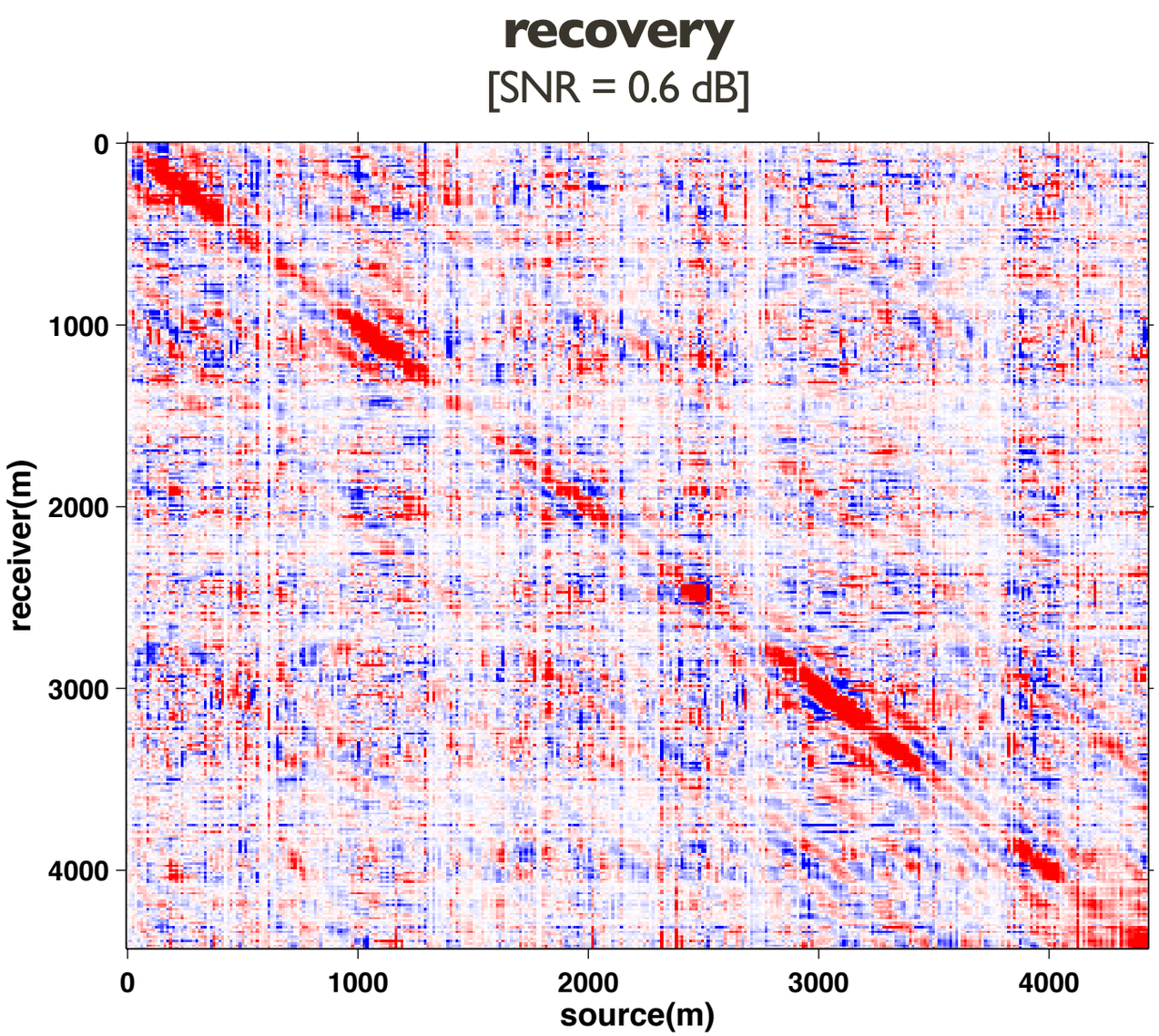
missing entries
[source-receiver]

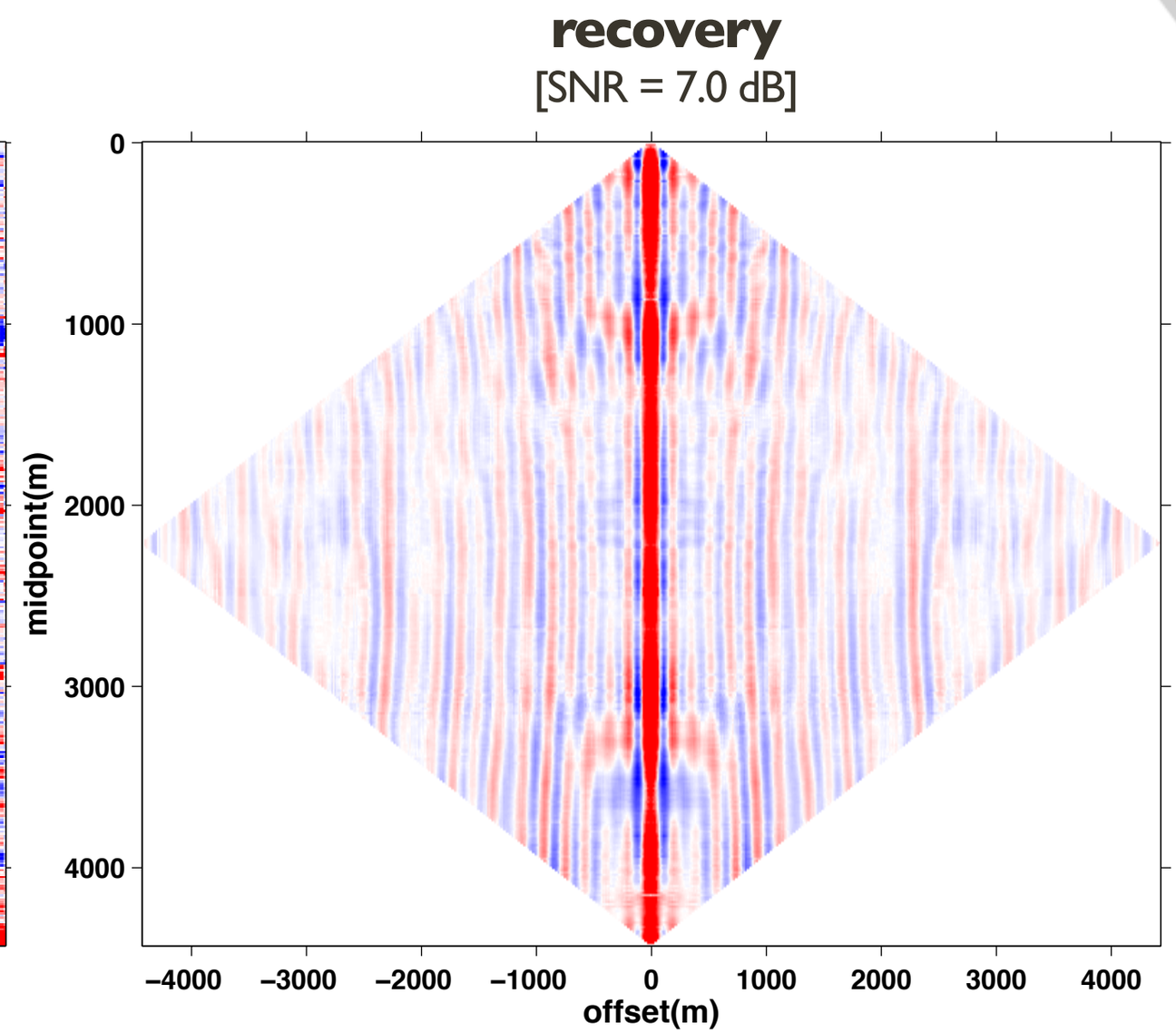


missing entries
[midpoint-offset]



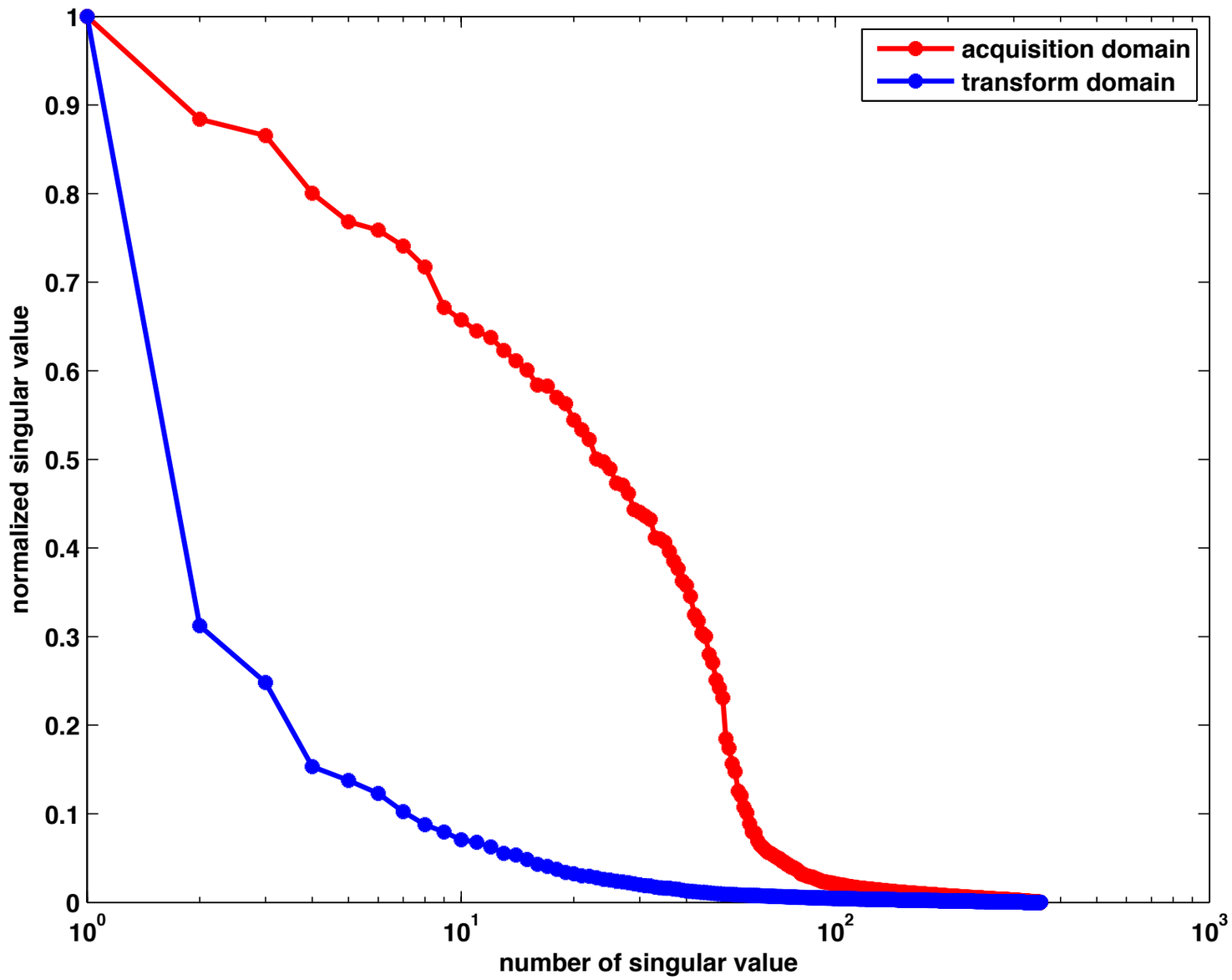
Low-rank interpolation





Singular value decay

2-D acquisition

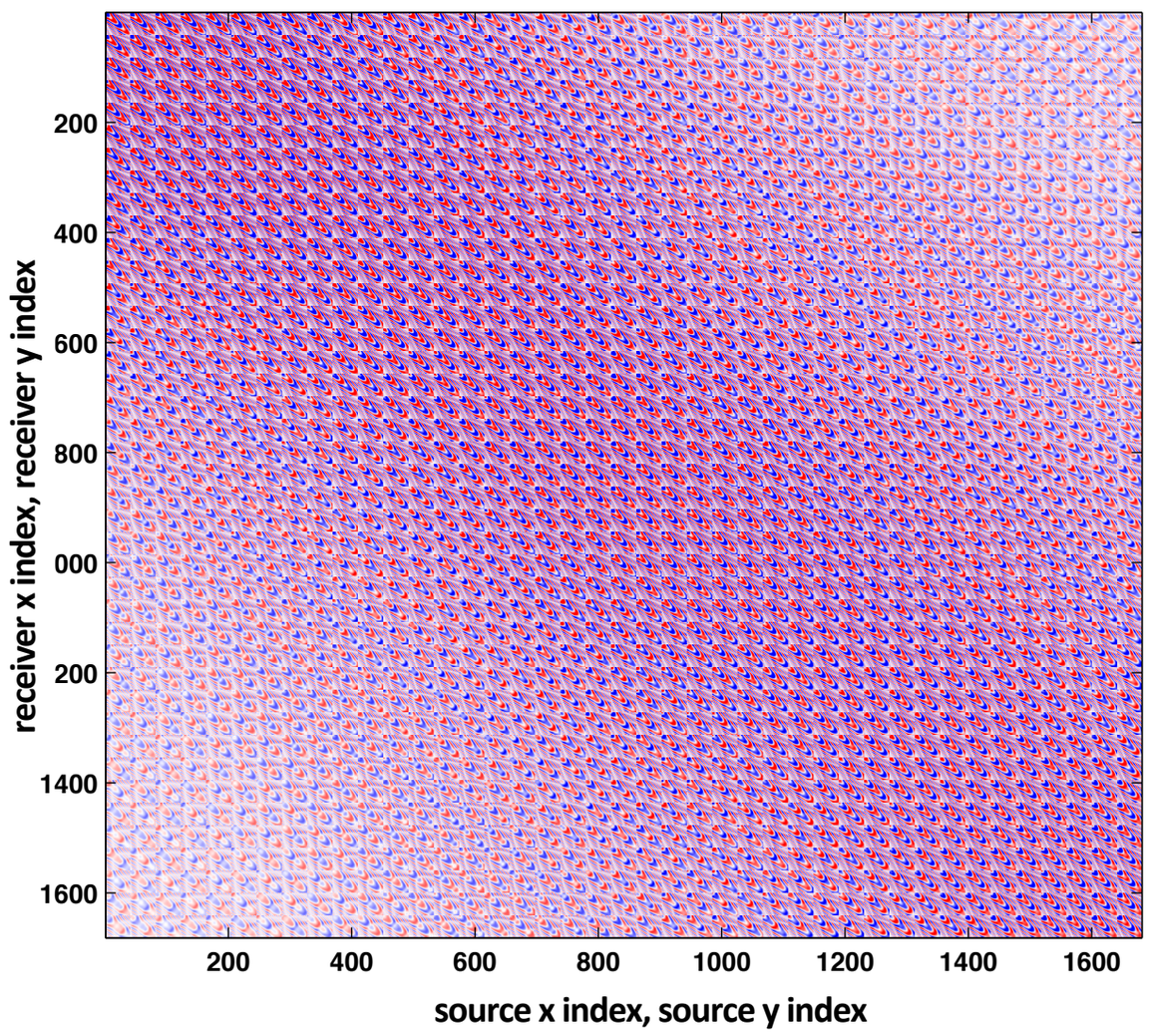


Low-rank structure

3-D acquisition

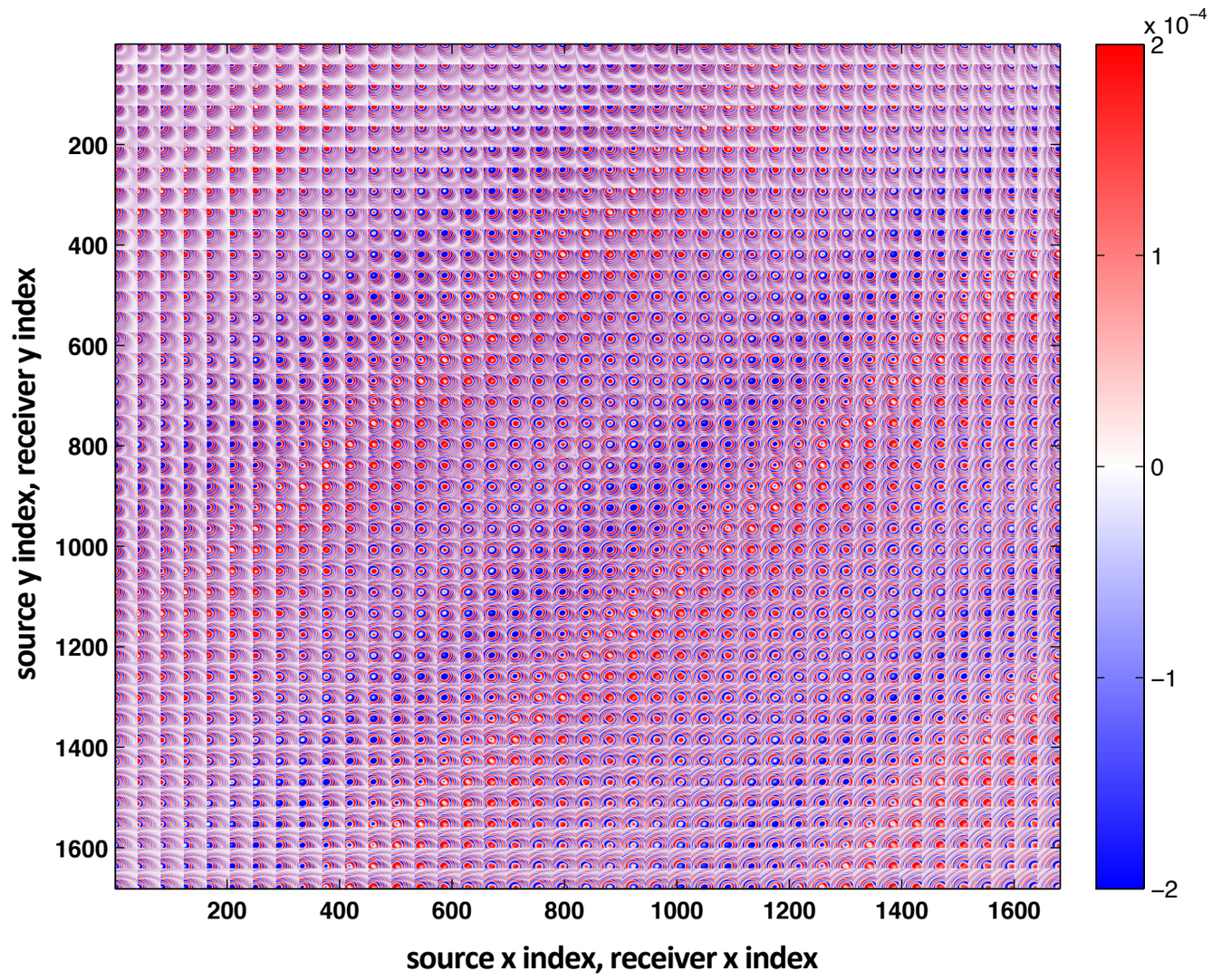
acquisition domain

[(source x,source y) & (receiver x, receiver y)]



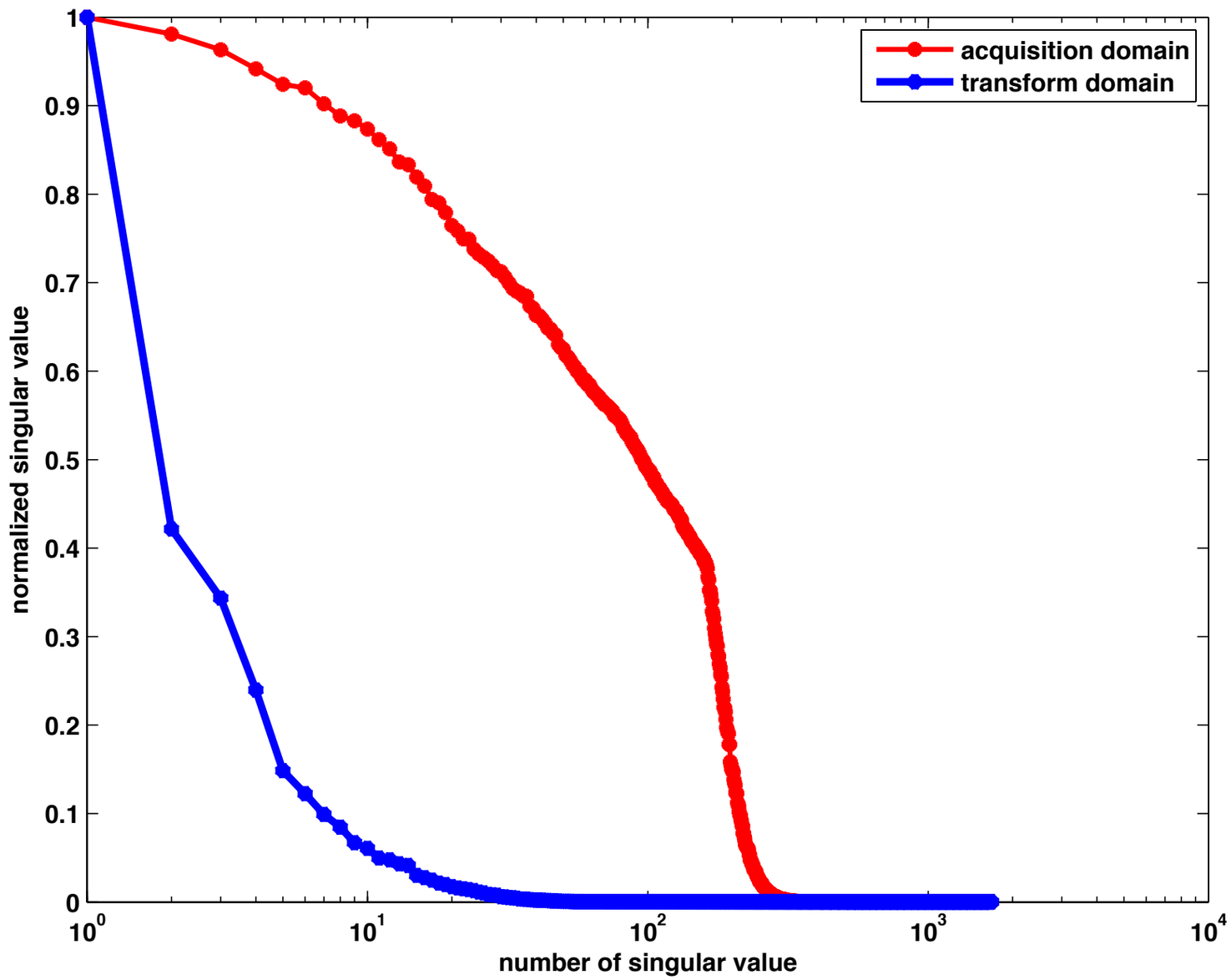
transform domain

[(source x,receiver x) & (source y, receiver y)]



Singular value decay

3-D acquisition



Matrix completion

- signal structure
  - *low rank/fast decay* of singular values

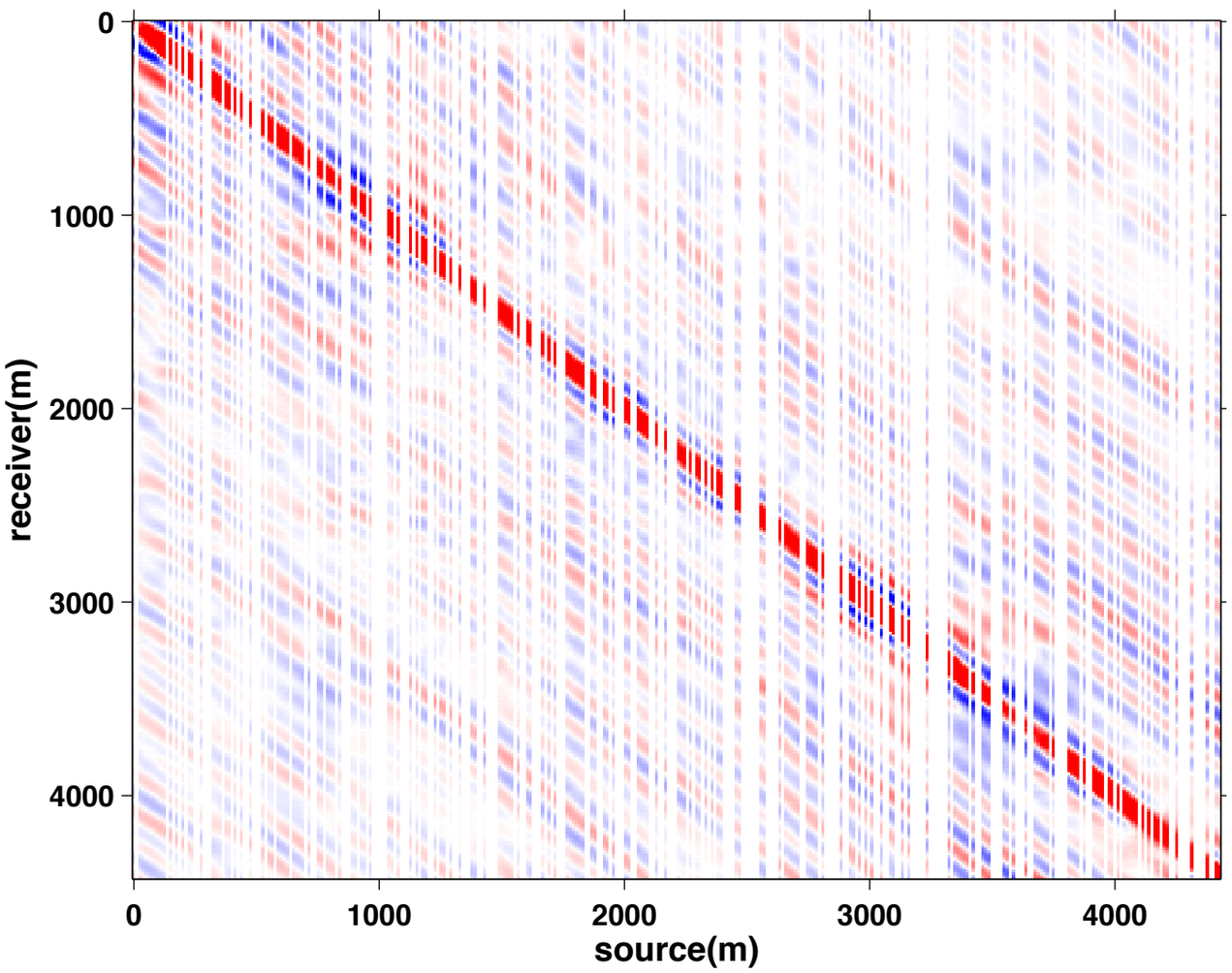
- sampling scheme
  - missing data *increase* rank in "transform domain"

- recovery using *rank penalization* scheme

2-D Acquisition
[randomized sampling]

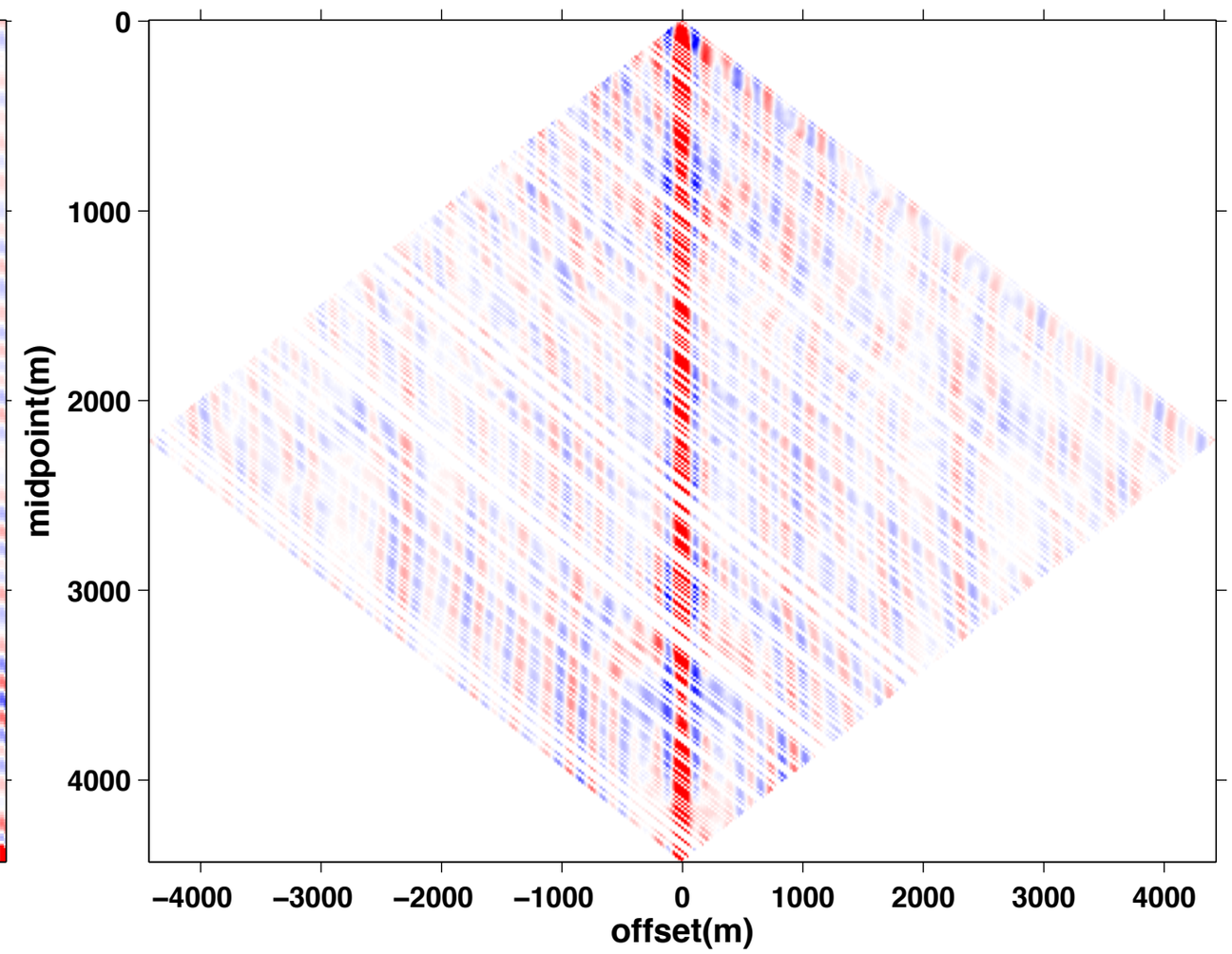
acquisition domain

missing columns *do not* increase rank

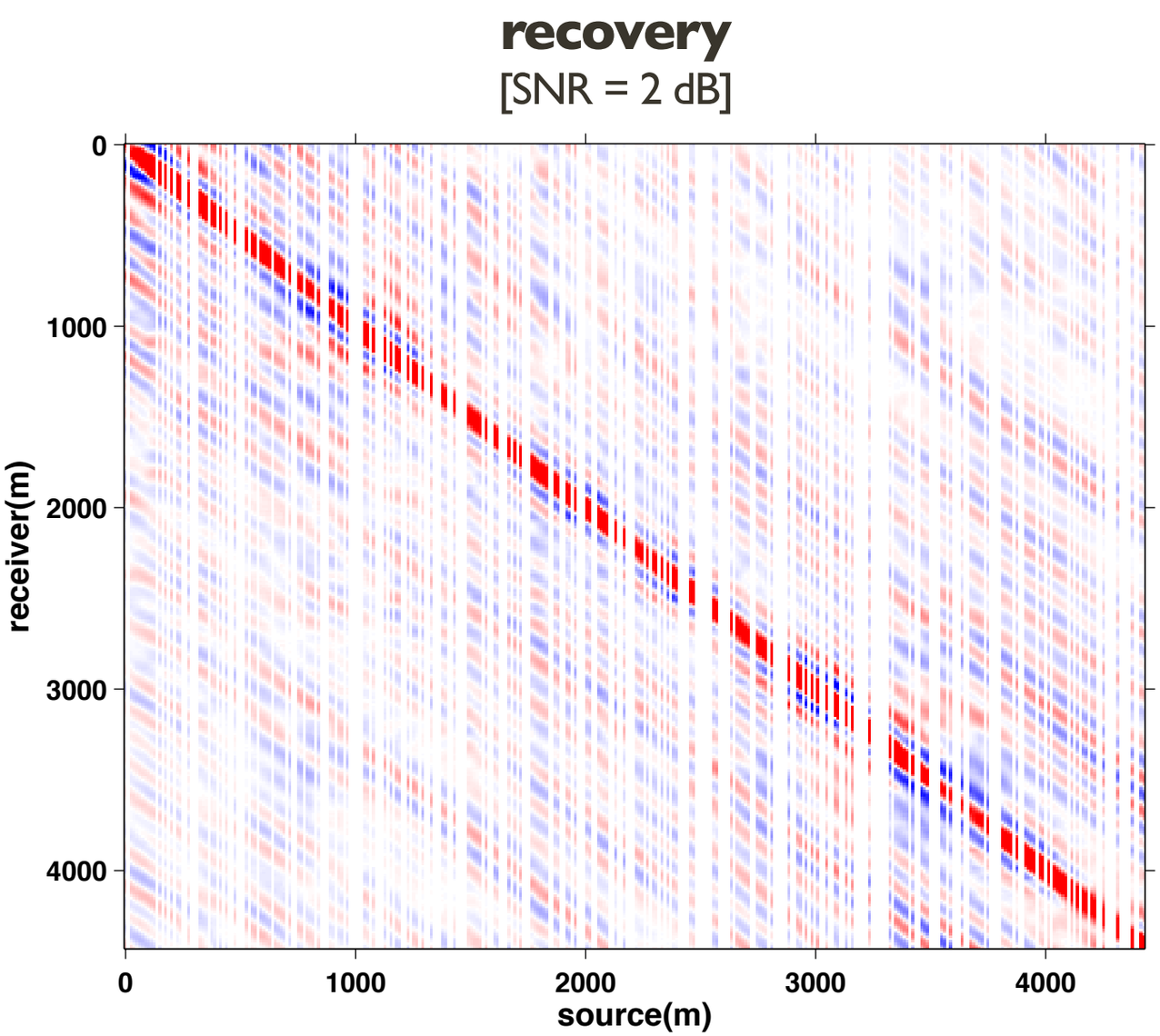


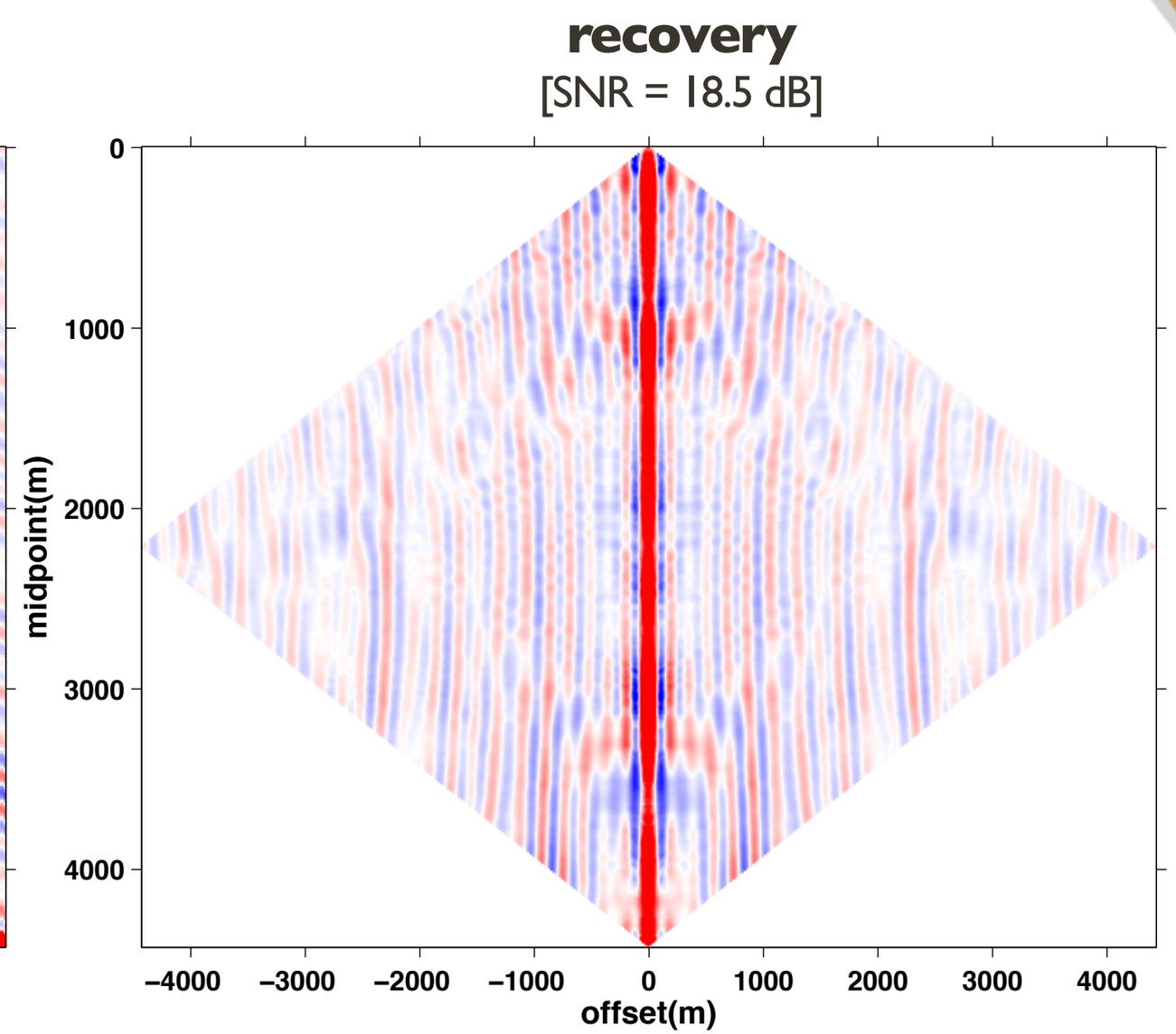
transform domain

missing columns *do* increase rank



Low-rank interpolation

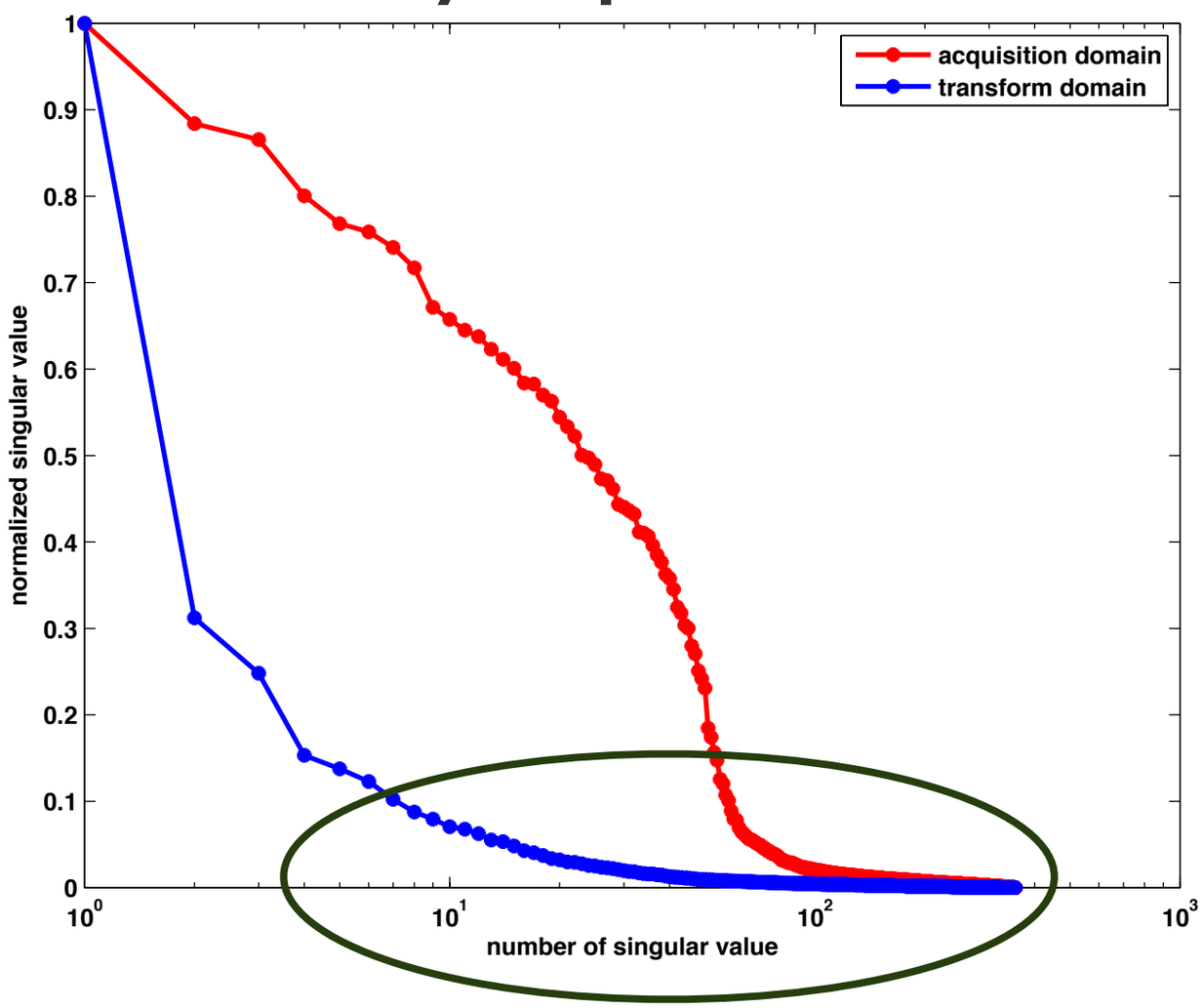




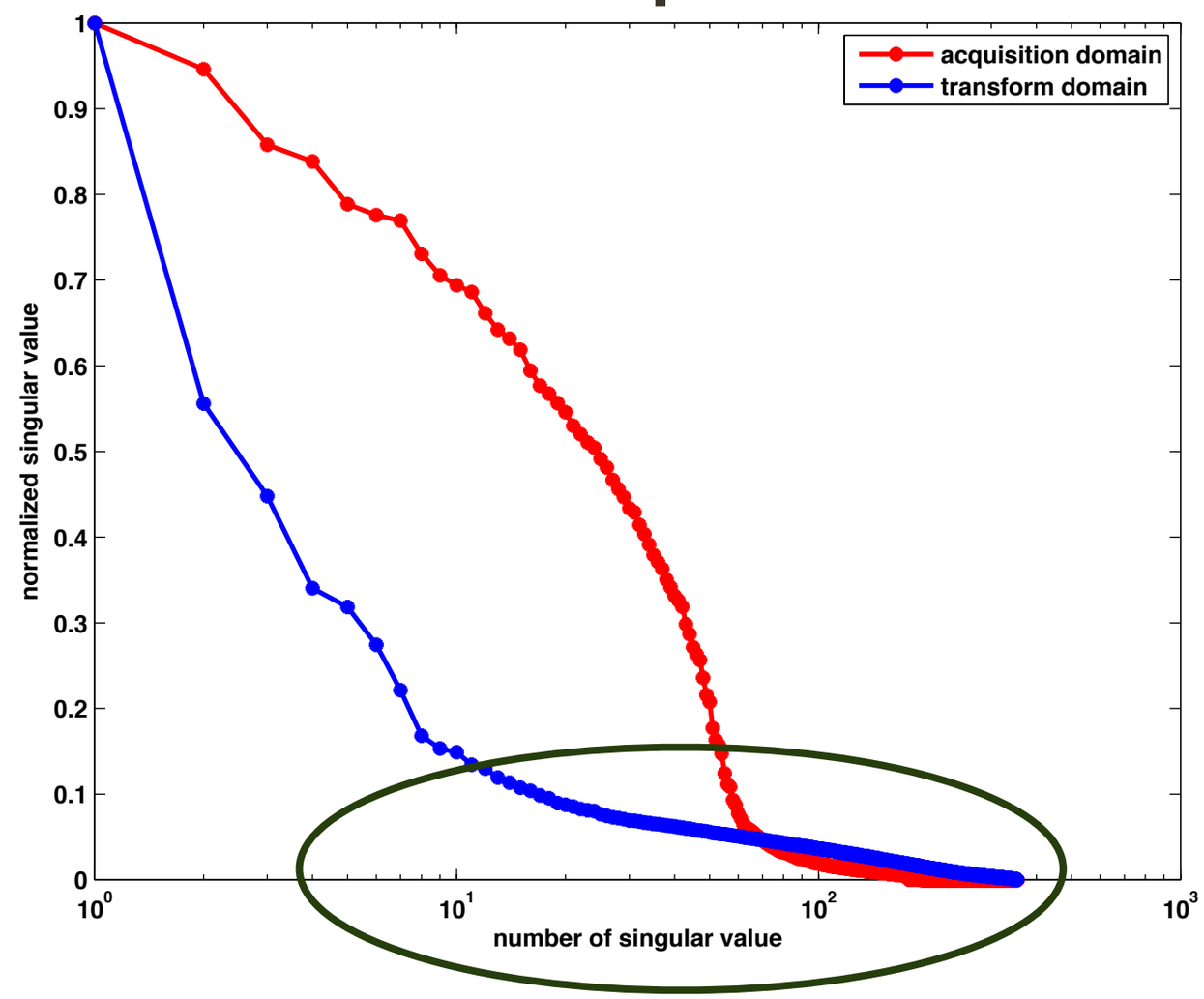
Randomized sampling

singular value decay

fully sampled data



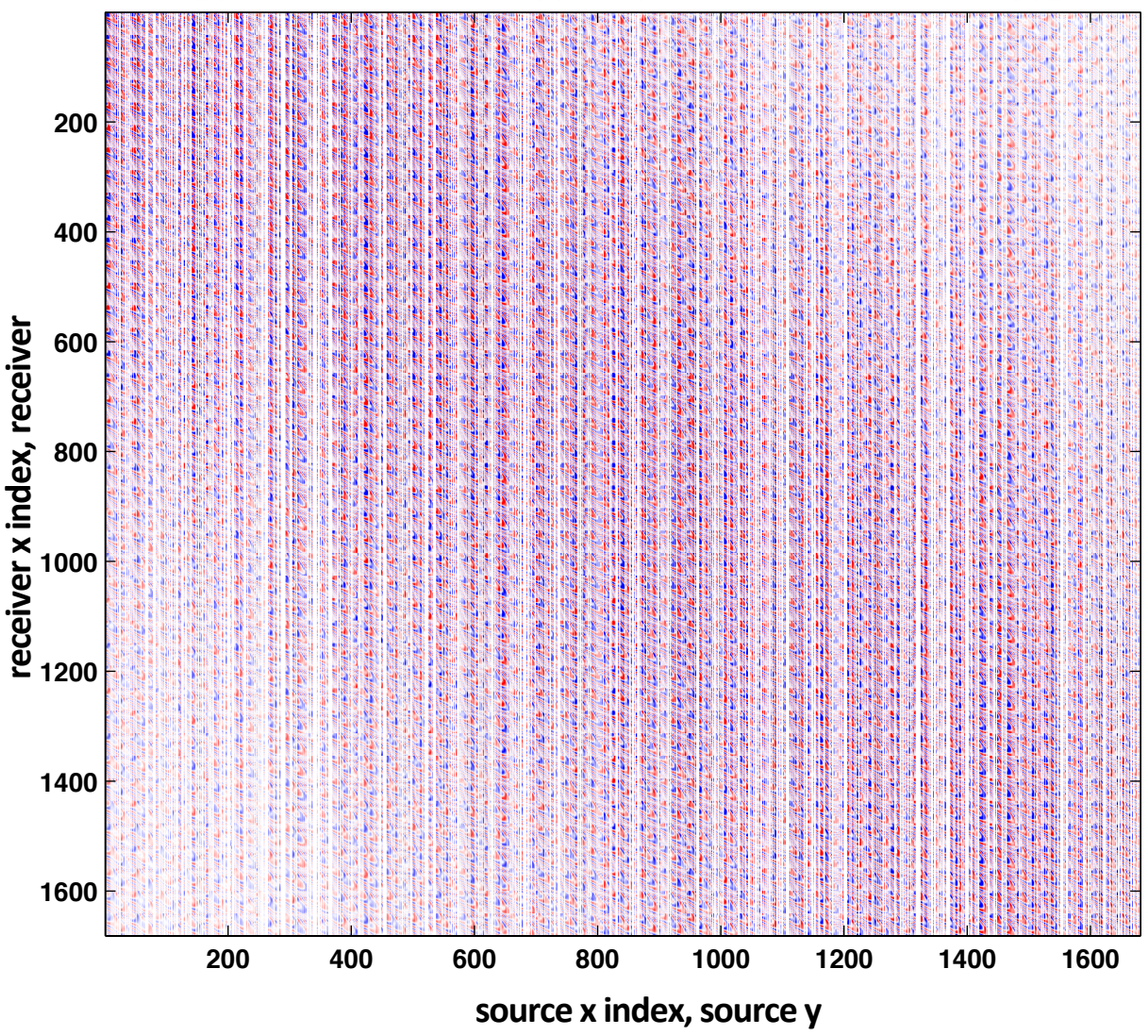
random sampled data



3-D Acquisition
[randomized sampling]

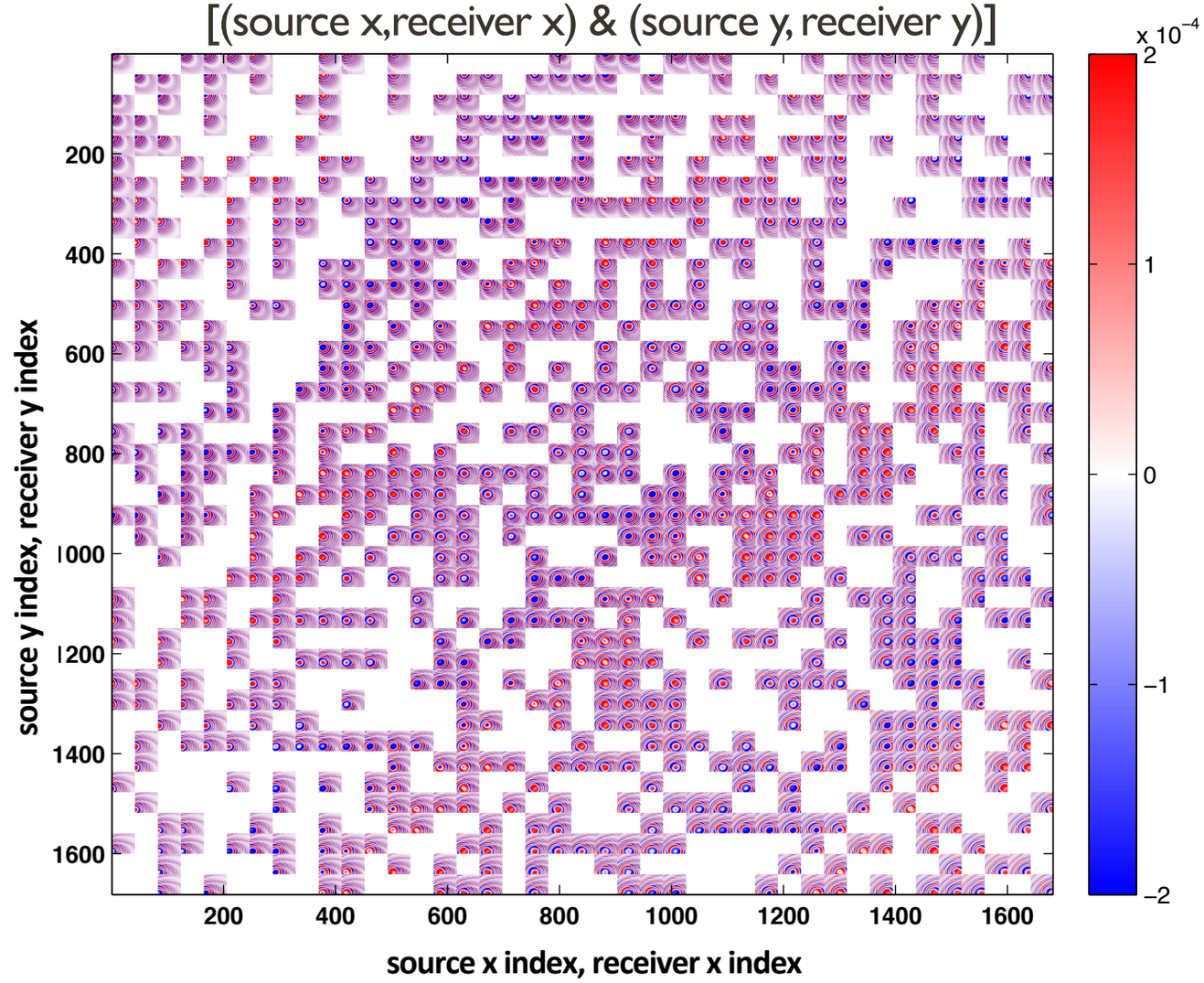
acquisition domain

[(source x,source y) & (receiver x, receiver y)]



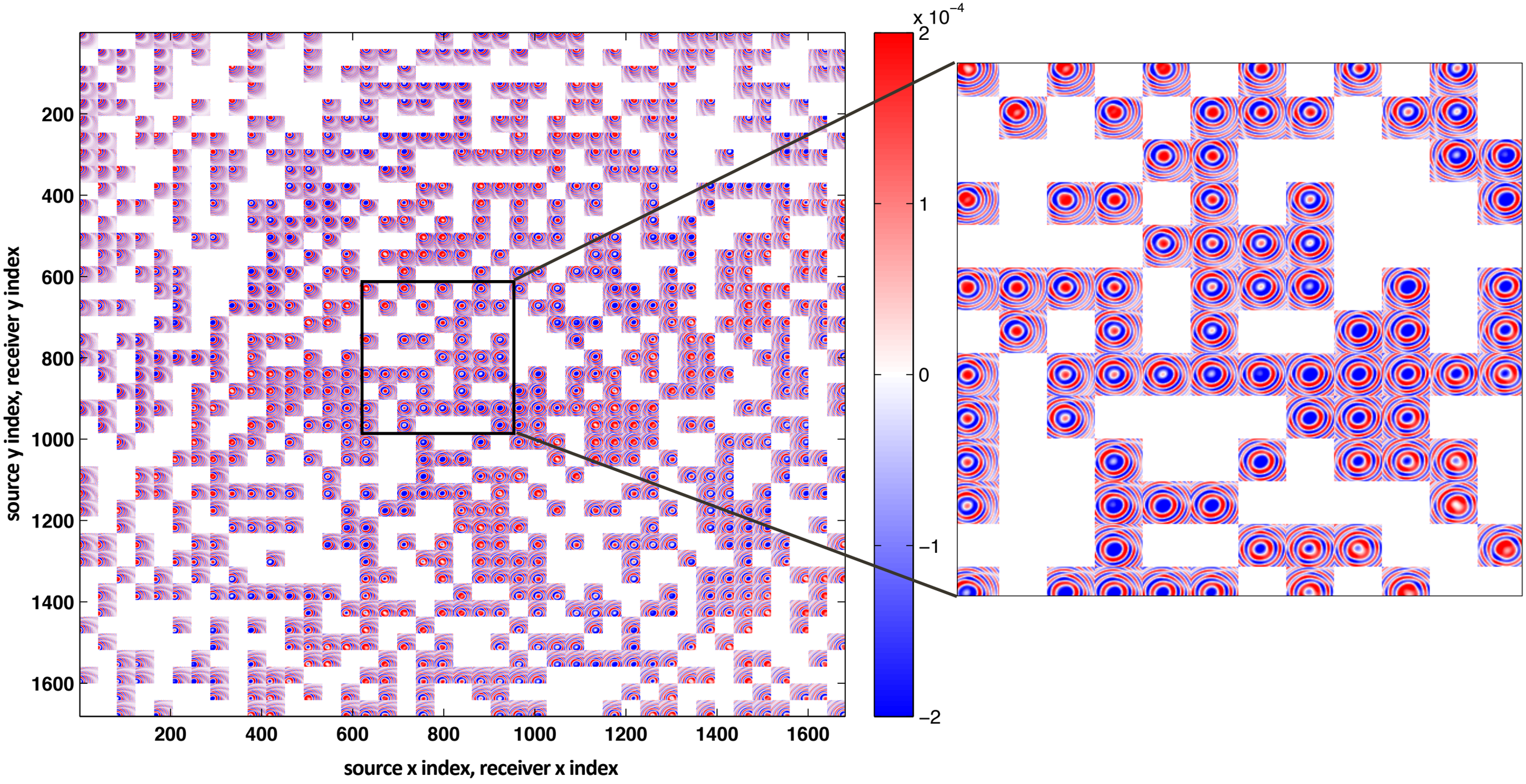
transform domain

[(source x,receiver x) & (source y, receiver y)]



3-D Acquisition

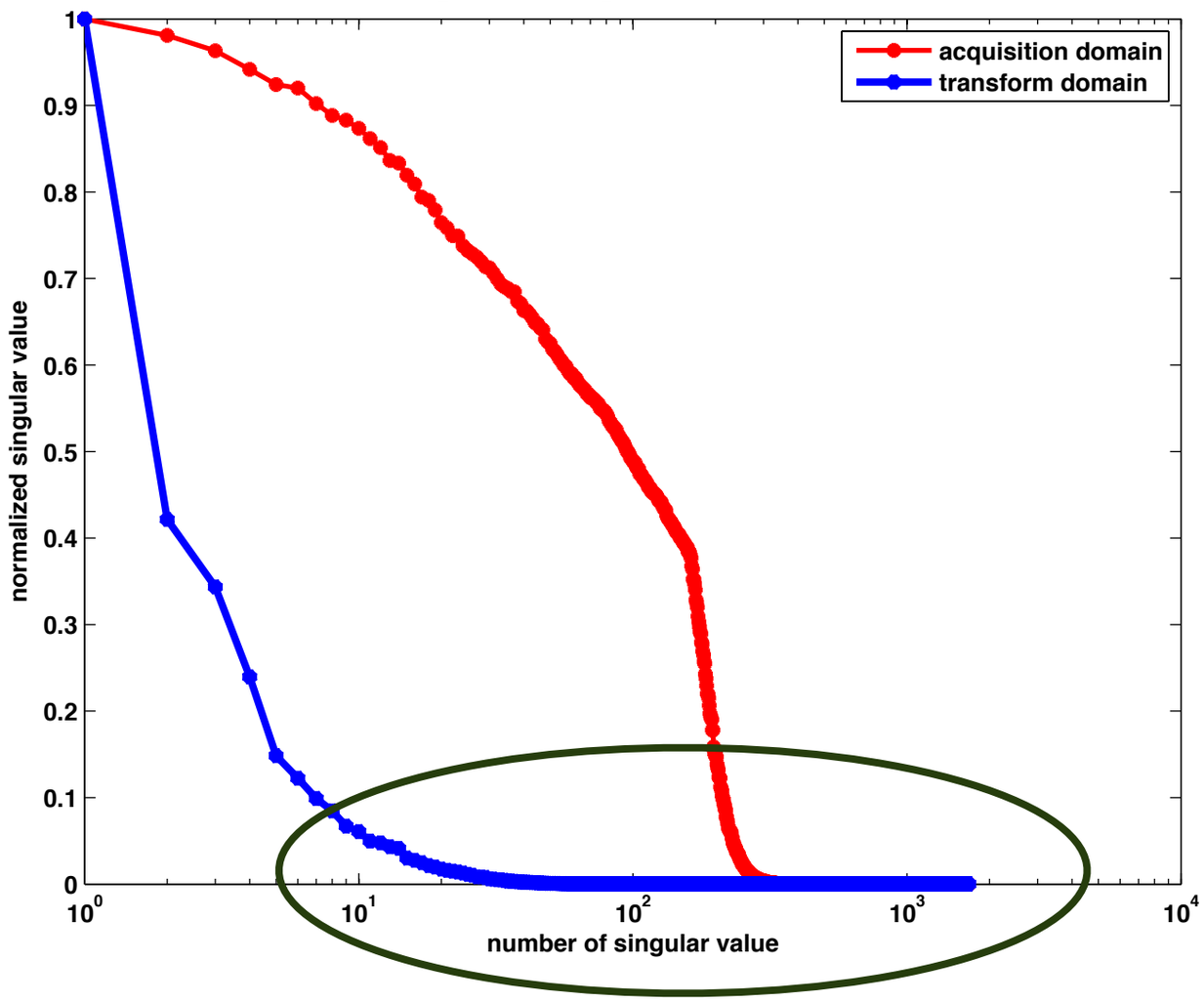
[randomized sampling - “transform” domain]



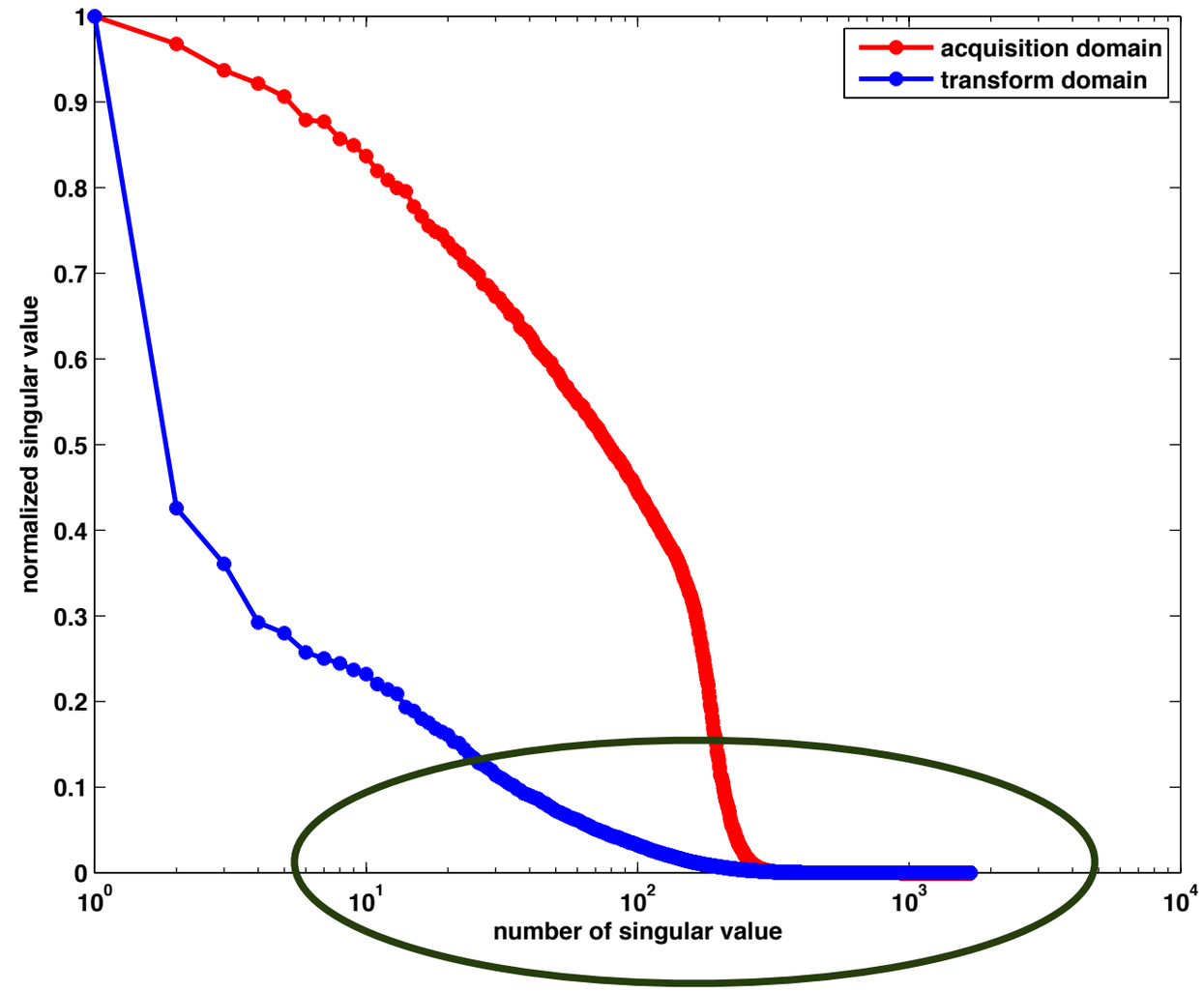
Randomized sampling

singular value decay

fully sampled data



random sampled data



Low-rank domain

- 2-D acquisition
  - midpoint-offset

- 3-D acquisition
  - (source x, receiver x) and (source y, receiver y)
  - midpoint-offset-azimuth

Observations

- sampling become *incoherent* in "transform" domain
- *slow decay* of singular values in "transform" domain

Matrix completion

- signal structure
  - *low rank/fast decay* of singular values
- sampling scheme
  - missing data *increase* rank in "transform domain"
- recovery using *rank penalization* scheme

Rank minimization

- given a set of measurements $\mathbf{b}$, aim is to solve

$$(BPDN_\sigma) \qquad \min_{\mathbf{X}} \quad rank(\mathbf{X}) \quad \text{s.t.} \ ||\mathcal{A}(\mathbf{X}) - \mathbf{b}||_2^2 \leq \sigma$$

where

$rank(\mathbf{X}) =$ number of singular values of $\mathbf{X}$

- $\mathcal{A}$ is the transform-sampling operator defined as

$$\mathcal{A} = \mathbf{R}\mathbf{M}\mathcal{S}^H$$

where

$\mathbf{R}$: restriction operator
$\mathbf{M}$: measurement operator
$\mathcal{S}^H$: transform operator

Rank minimization

- prohibitively *expensive*
  - do not know rank value in advance
  - search over all possible values of rank

- instead solve nuclear-norm minimization
  - convex relaxation of rank-minimization

[Recht et. al. 2010]

[Recht et. al. 2010]

SLIM

Nuclear-norm minimization

- we want to solve

$$(BPDN_\sigma) \qquad \min_{\mathbf{X}} ||\mathbf{X}||_* \quad \text{s.t.} \quad ||\mathcal{A}(\mathbf{X}) - \mathbf{b}||_2^2 \leq \sigma$$

where

$$\|\mathbf{X}\|_* = \sum_{i=1}^{m} \lambda_i = \|\lambda\|_1$$

where λ_i are the *singular* values

Challenges

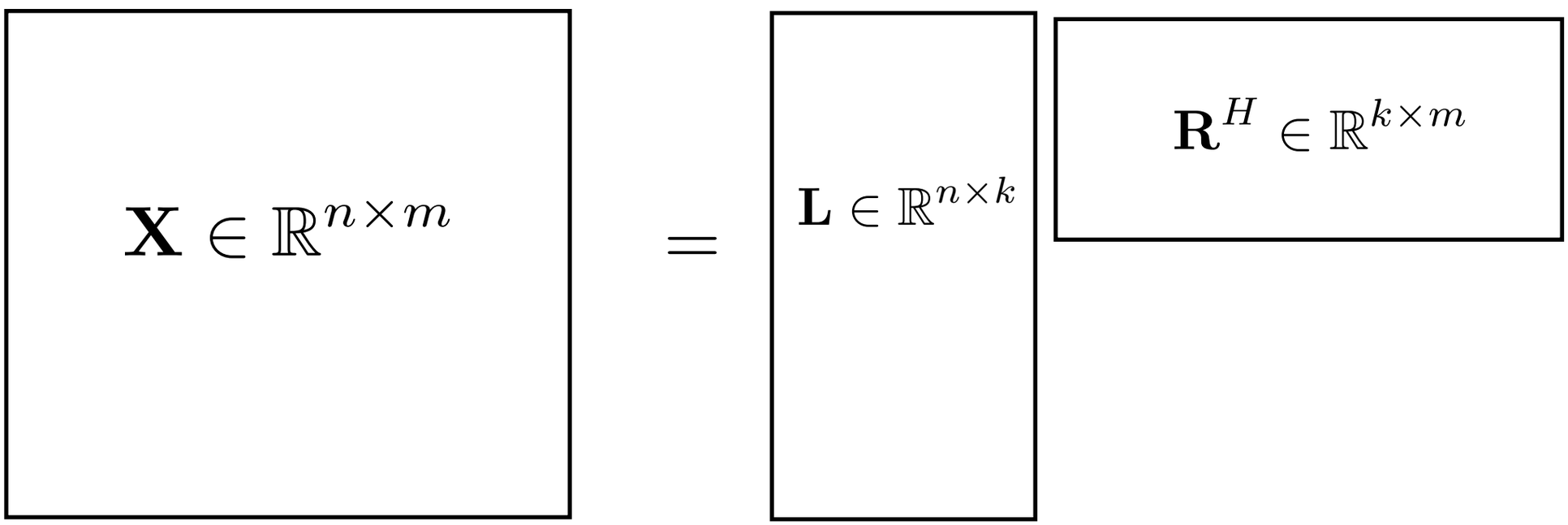
- requires repeated application of *SVD* for projection

- expensive to compute for large system
  - curse of dimensionality

- can we exploit rank structure "*SVD* free"

[Rennie and Srebro 2005, Lee et. al. 2010, Recht and Re 2011]

Factorized formulation



$$\mathbf{X} = \mathbf{L}\mathbf{R}^H$$

[Berg and Friedlander 2008, Aravkin et al. 2012b]

Factorized formulation

- reformulate $BPDN_\sigma$ formulation

$$\min_{\mathbf{L},\mathbf{R}} ||\mathbf{L}\mathbf{R}^H||_* \quad \text{s.t.} \ ||\mathcal{A}(\mathbf{L}\mathbf{R}^H) - \mathbf{b}||_2^2 \leq \sigma$$

- approximately solve a series of $LASSO_\tau$ formulation

$$v(\tau) = \min_{\mathbf{L},\mathbf{R}} ||\mathcal{A}(\mathbf{L}\mathbf{R}^H) - \mathbf{b}||_2^2 \quad \text{s.t.} \ \|\mathbf{L}\mathbf{R}^H\|_* \leq \tau$$

where τ is a rank regularization parameter

[Rennie and Srebro 2005]

Factorized formulation

- nuclear norm is then define as

$$\|\mathbf{L}\mathbf{R}^H\|_* \leq \frac{1}{2}\left\|\begin{bmatrix}\mathbf{L}\\ \mathbf{R}\end{bmatrix}\right\|_F^2$$

where $\|\cdot\|_F^2$ is sum of squares of all entries

- choose rank k *explicitly* & avoid costly *SVD's*

Denoising

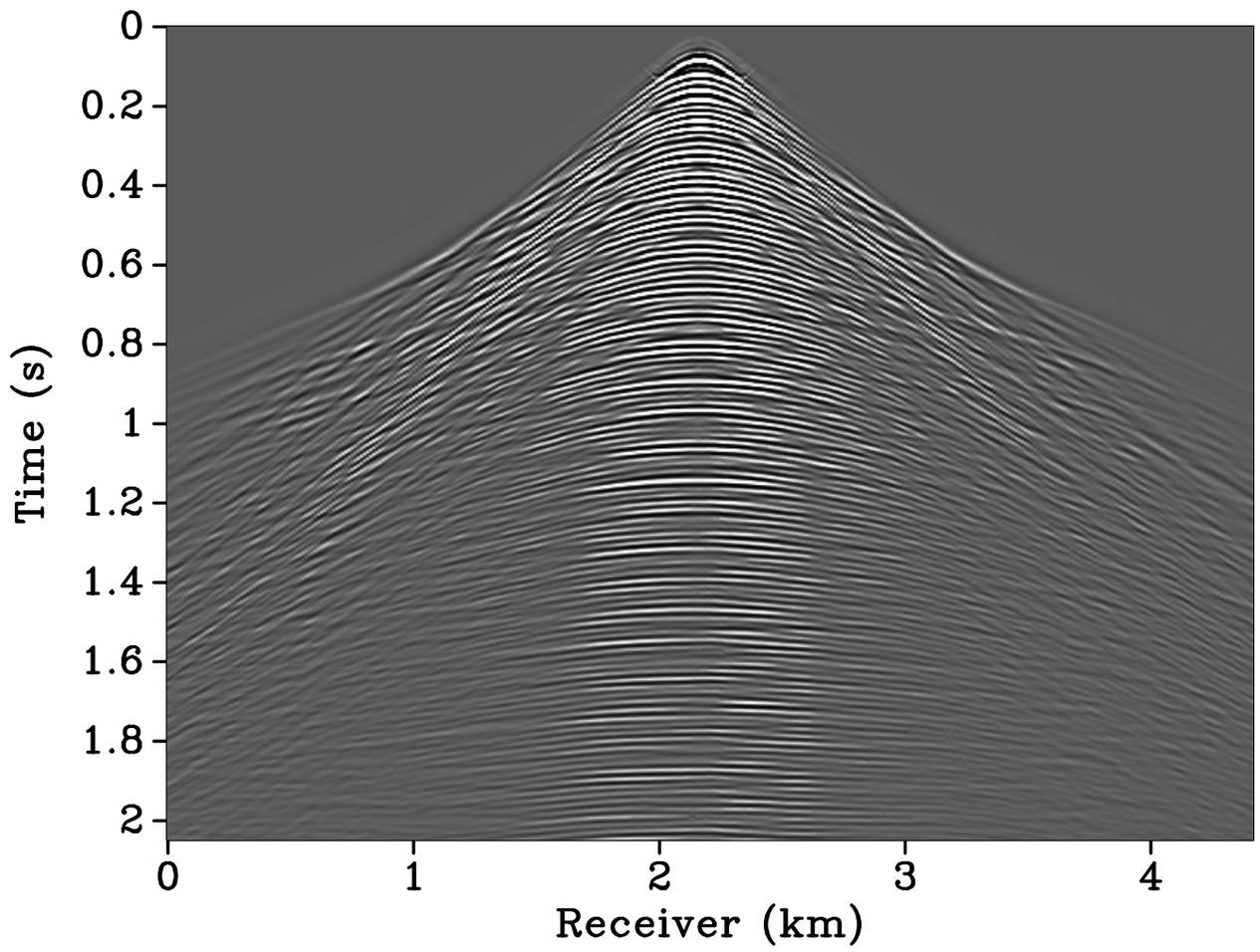
- Gulf of Suez
  - 2-D seismic line
  - 50% missing
  - 10% high amplitude noise

- Gulf of Mexico
  - monochromatic data slice at 2 Hz
  - low-amplitude noise

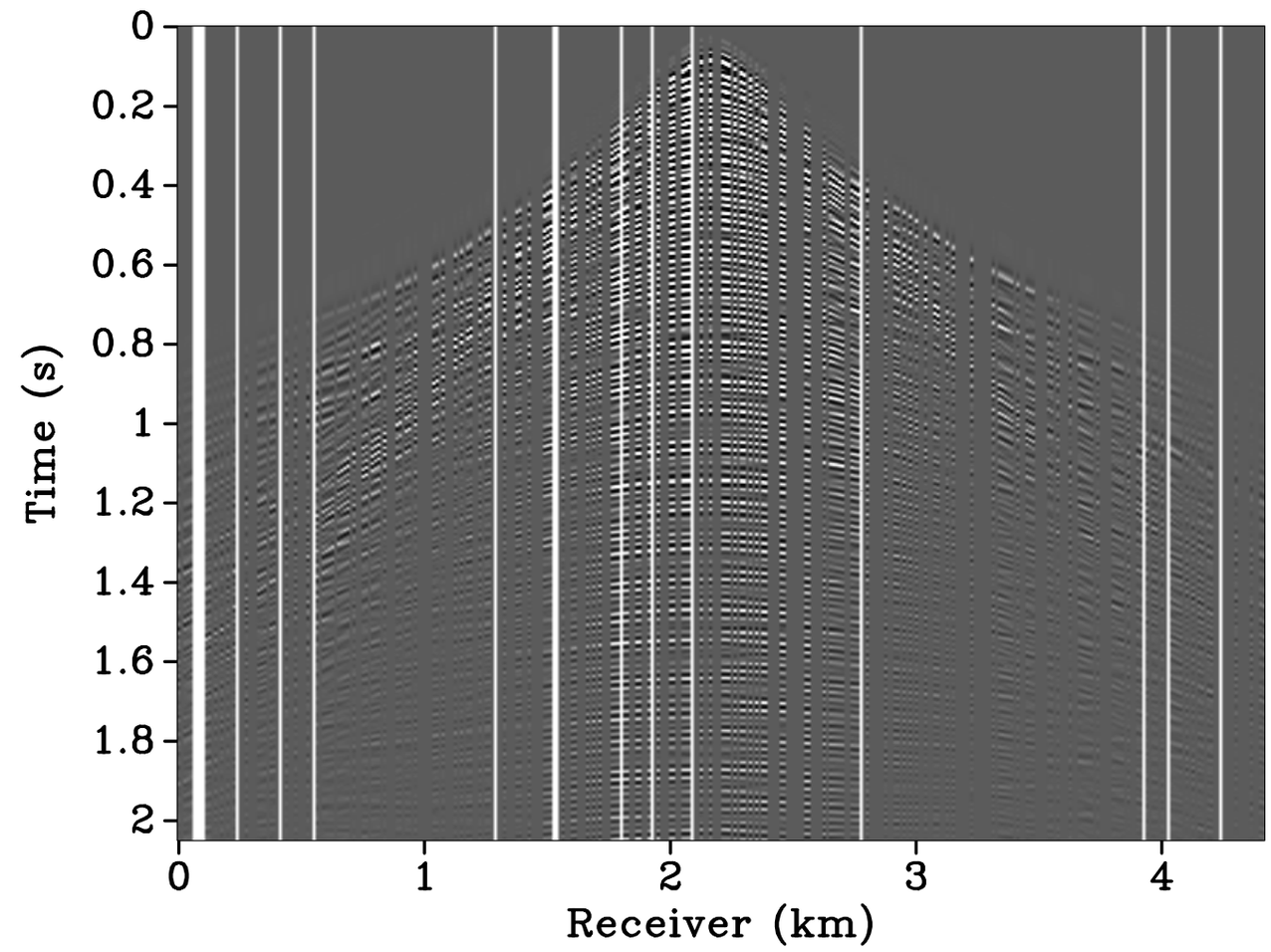
Gulf of Suez
[erratic gaussian noise]

- **noise level = 1000 times the maximum amplitude in data**

Regular sampled data



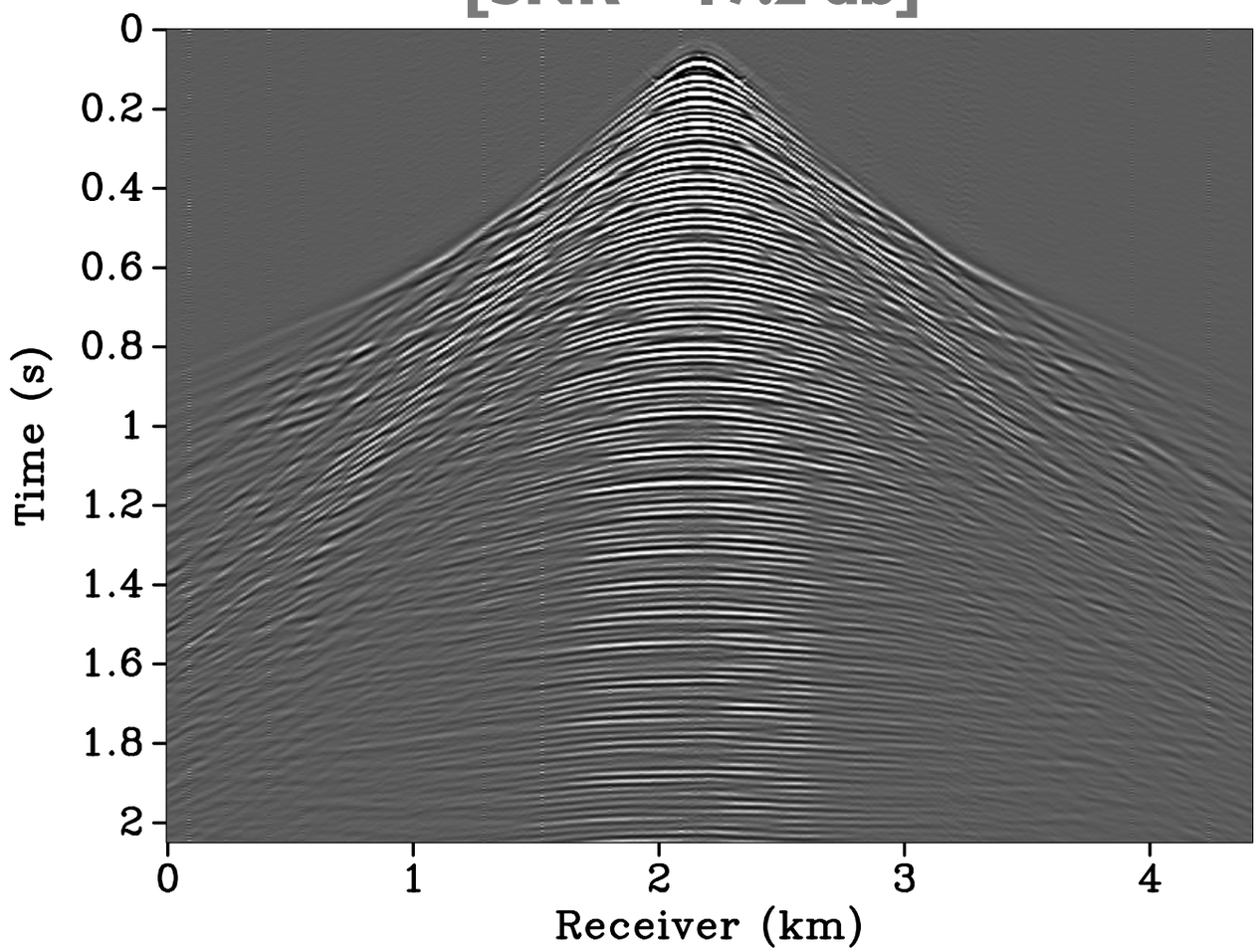
50% missing and 10% erratic noise



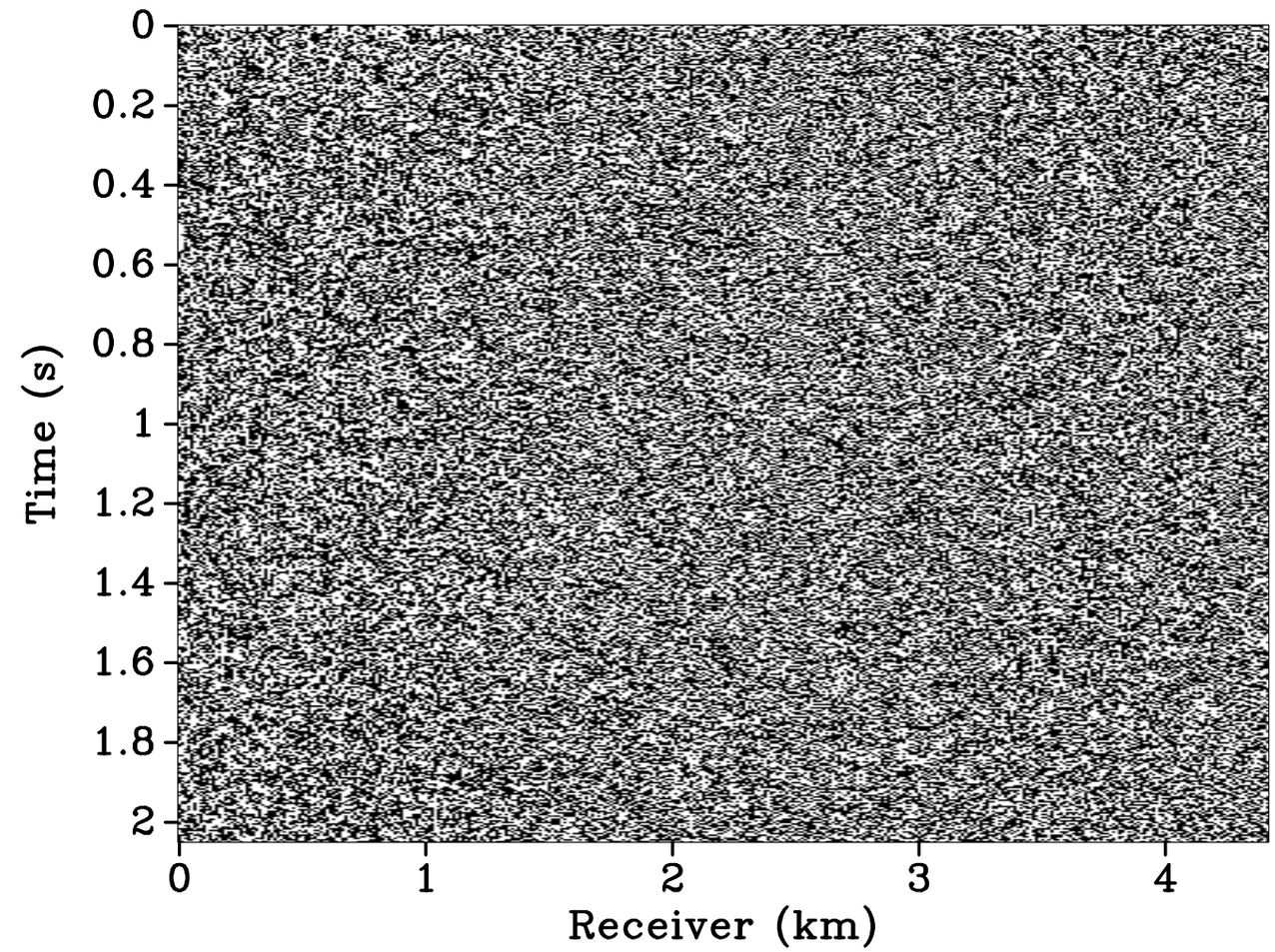
[Aravkin et. al. 2013]

Recovery - Shot Gather

Robust Penalty
[SNR = 17.2 db]

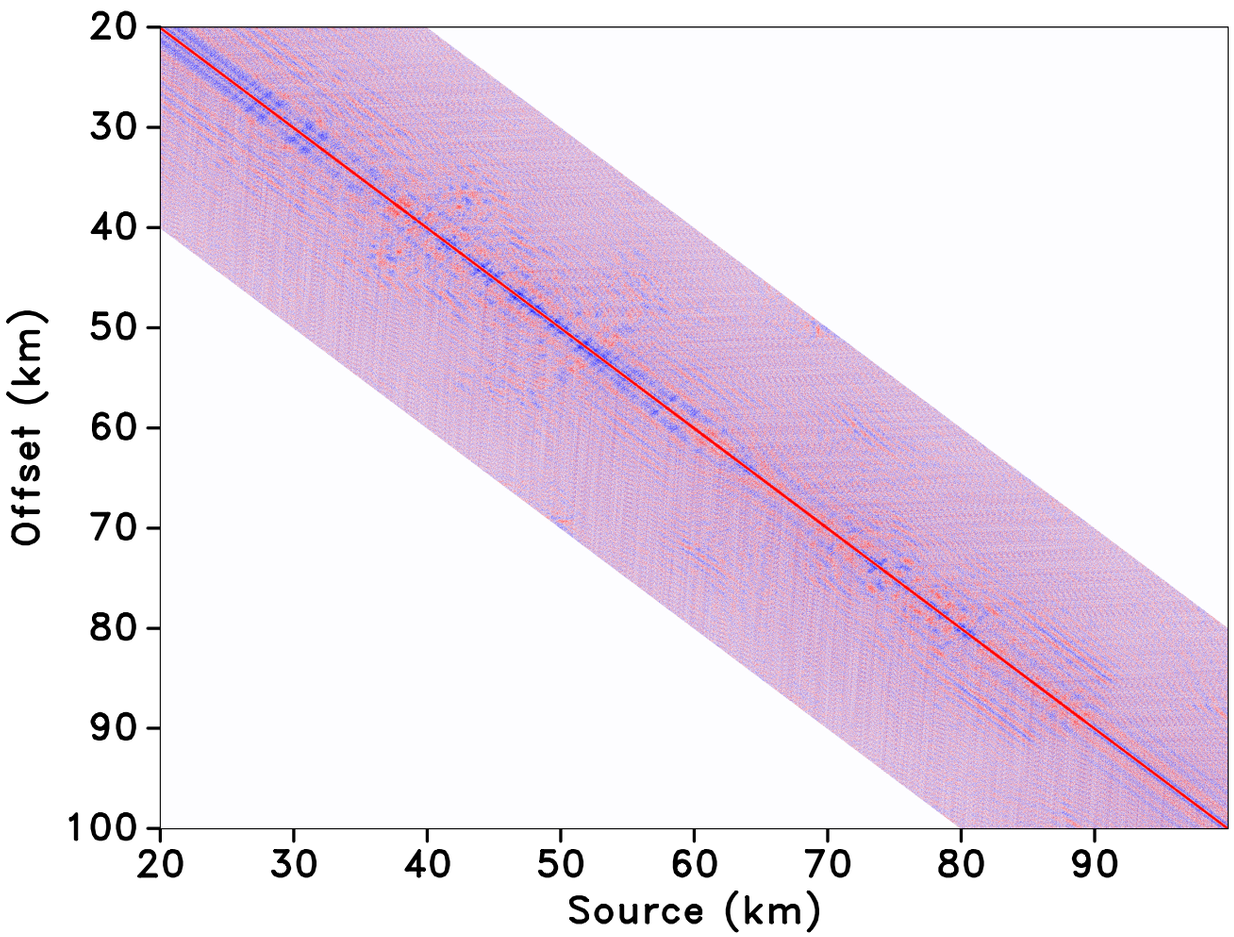


Least-squares Penalty

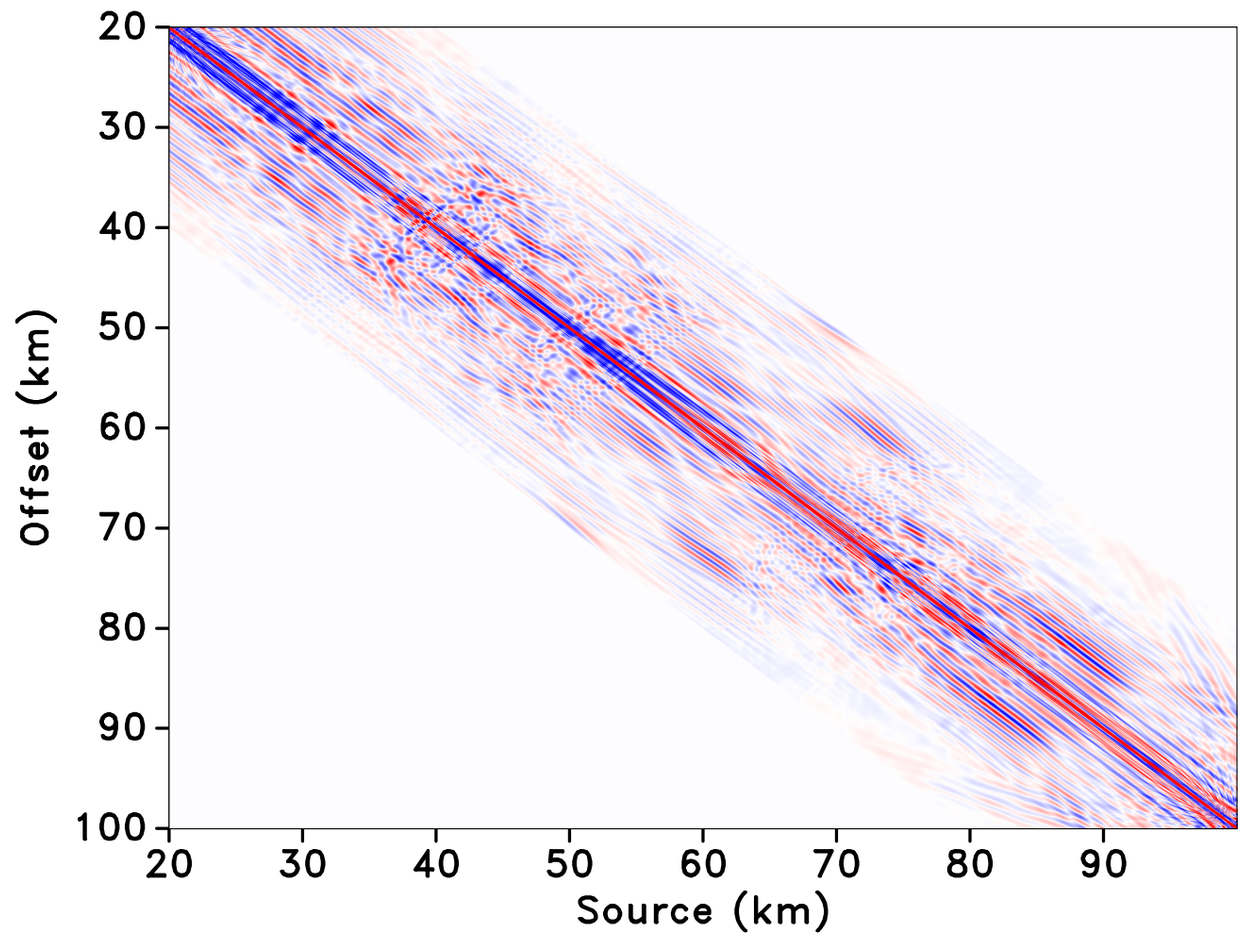


Gulf of Mexico
[2 Hz]

before denoising

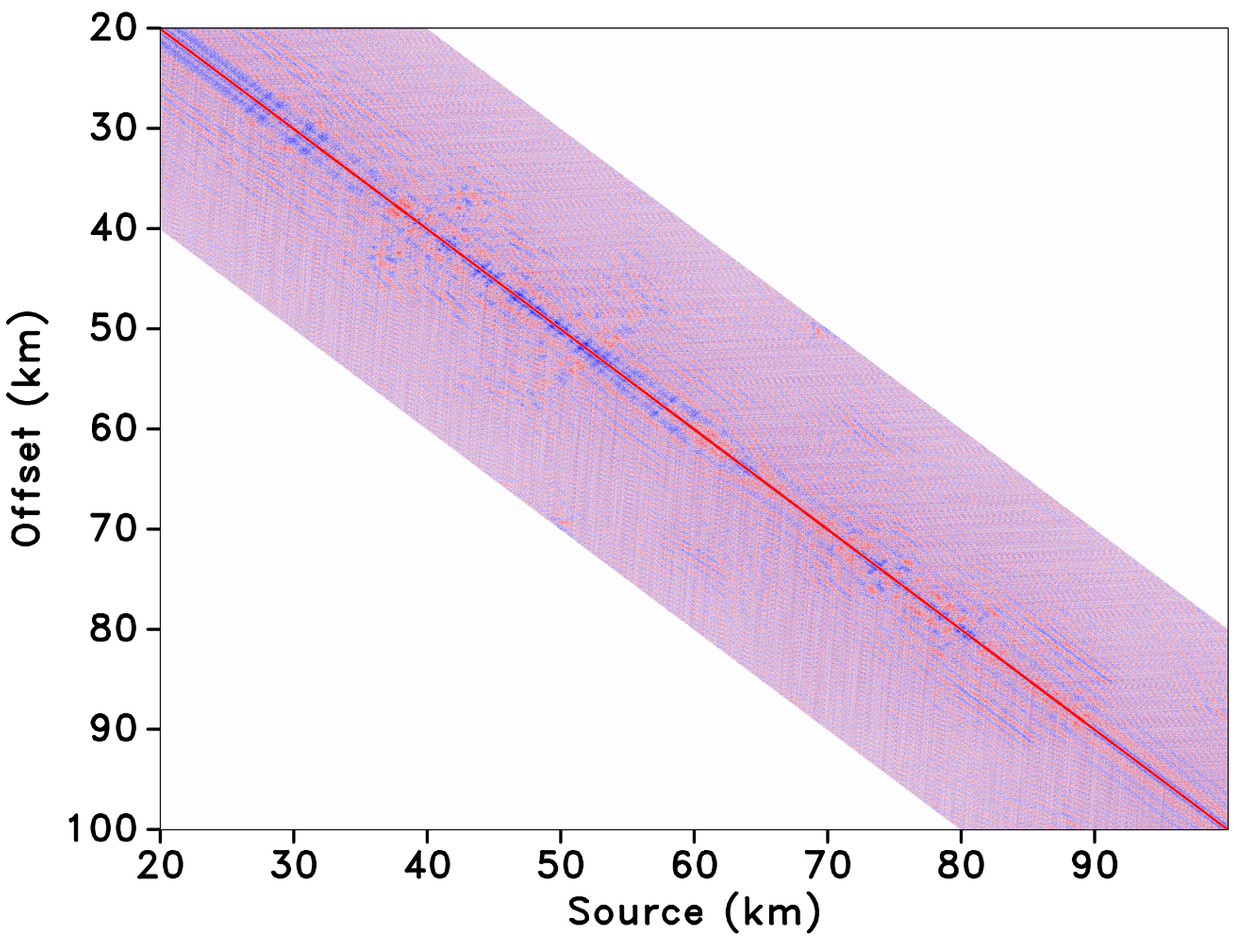


after denoising
[curvelet]

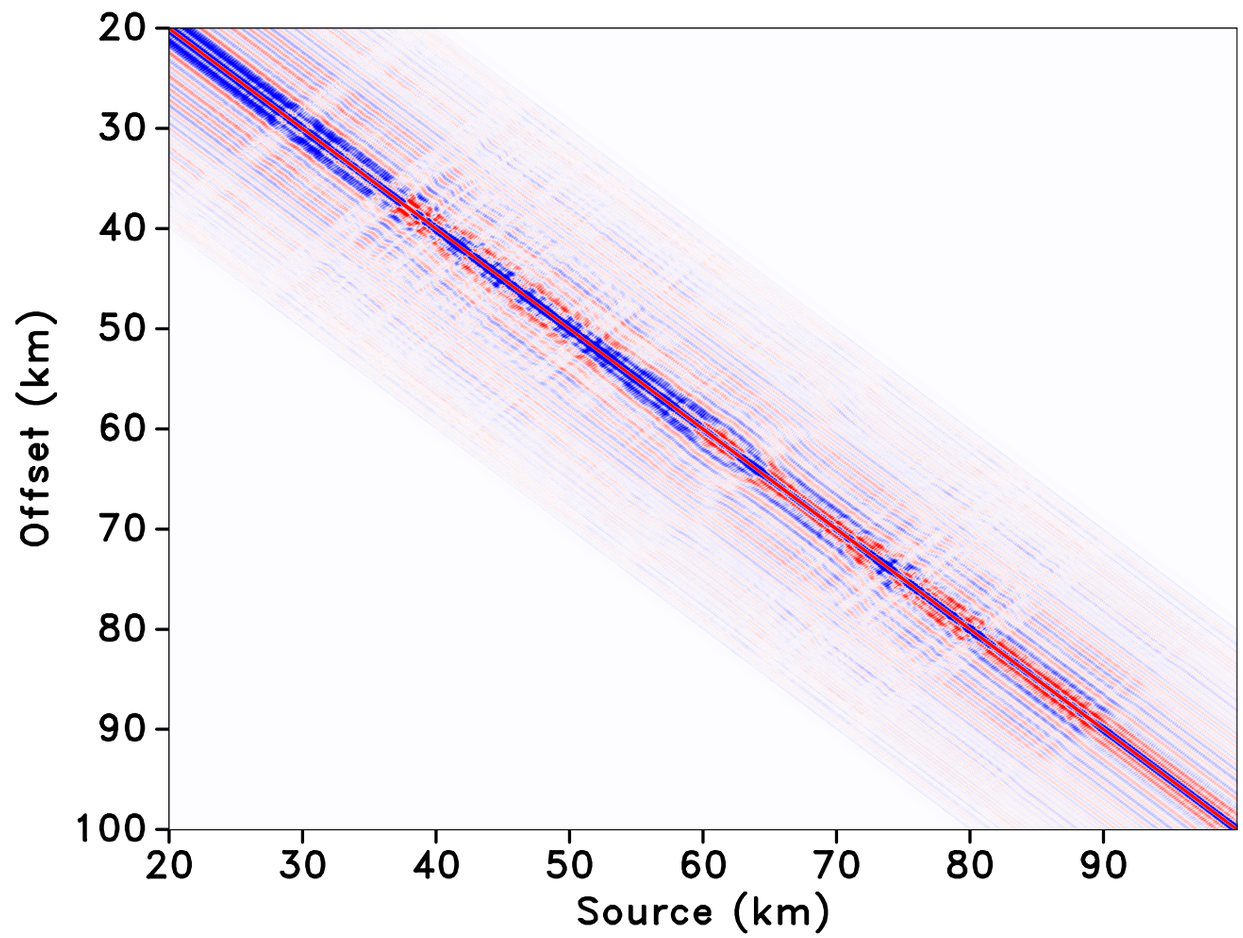


Gulf of Mexico
[2 Hz]

before denoising

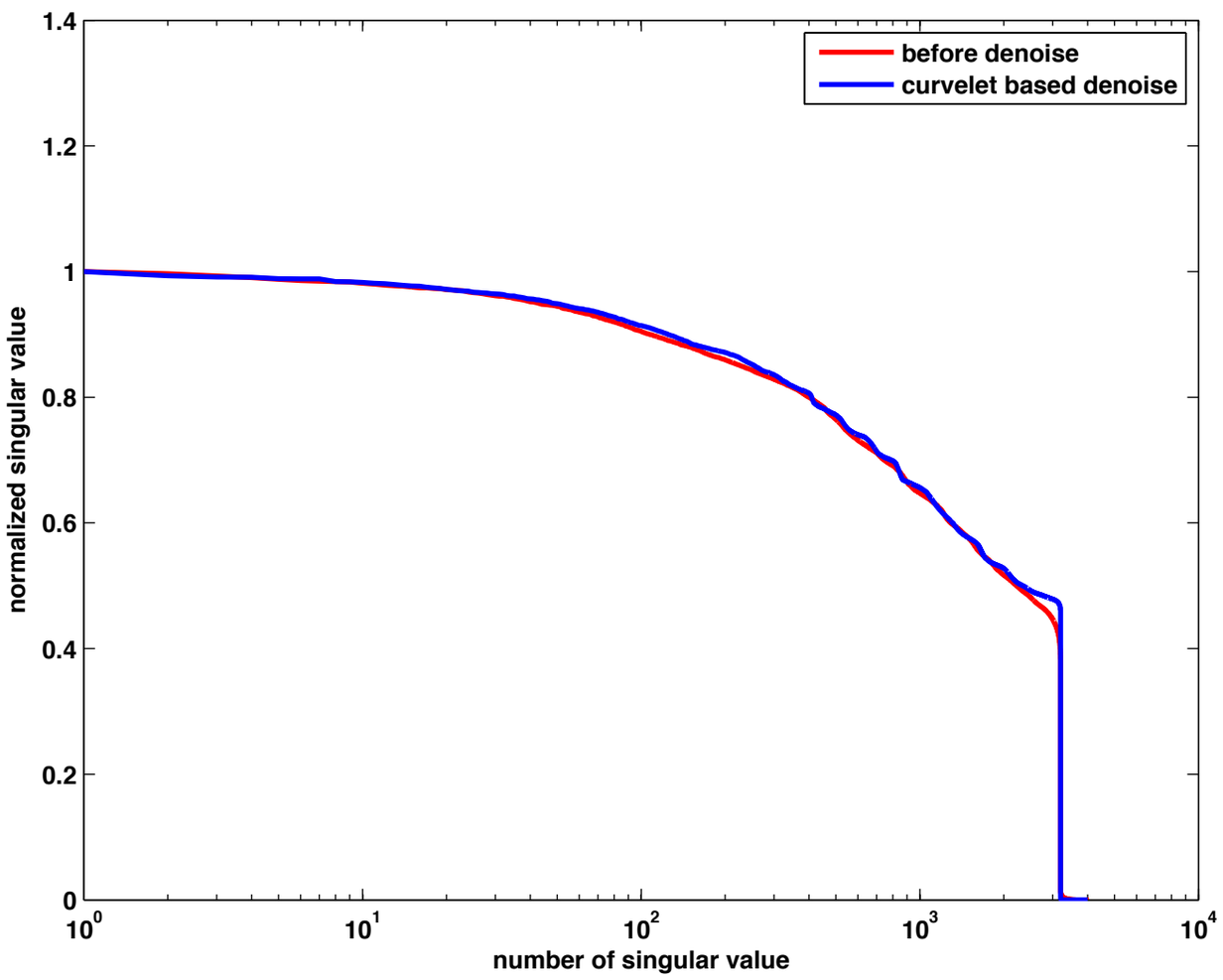


after denoising
[low-rank]

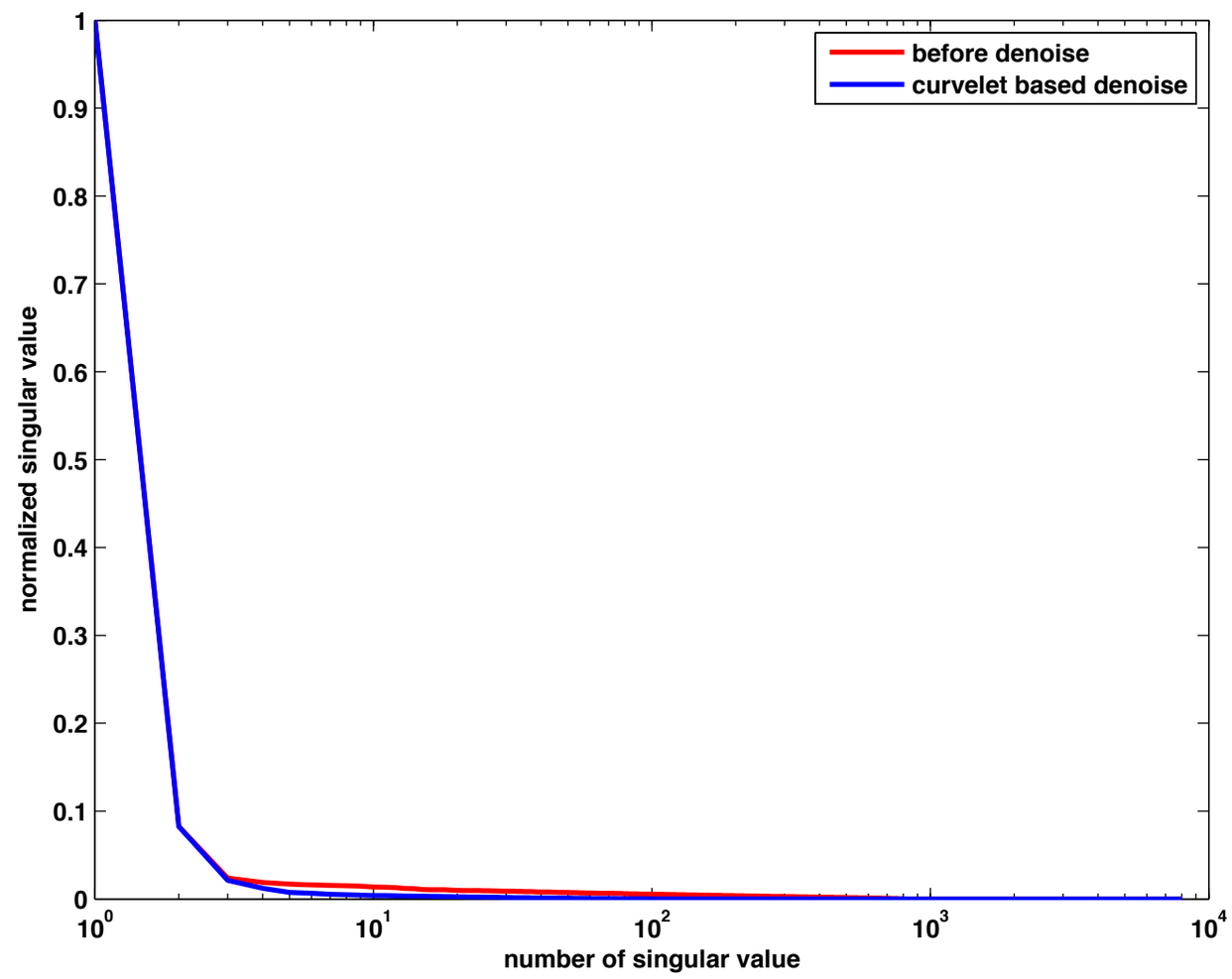


No Free Lunch

acquisition domain
[source-receiver]



transform domain
[midpoint-offset]



Comparison

	Low-rank	Curvelet
computational time	$O(minutes)$	$O(hours)$
storage	$k \times (n+m)$	$8 \times nm$

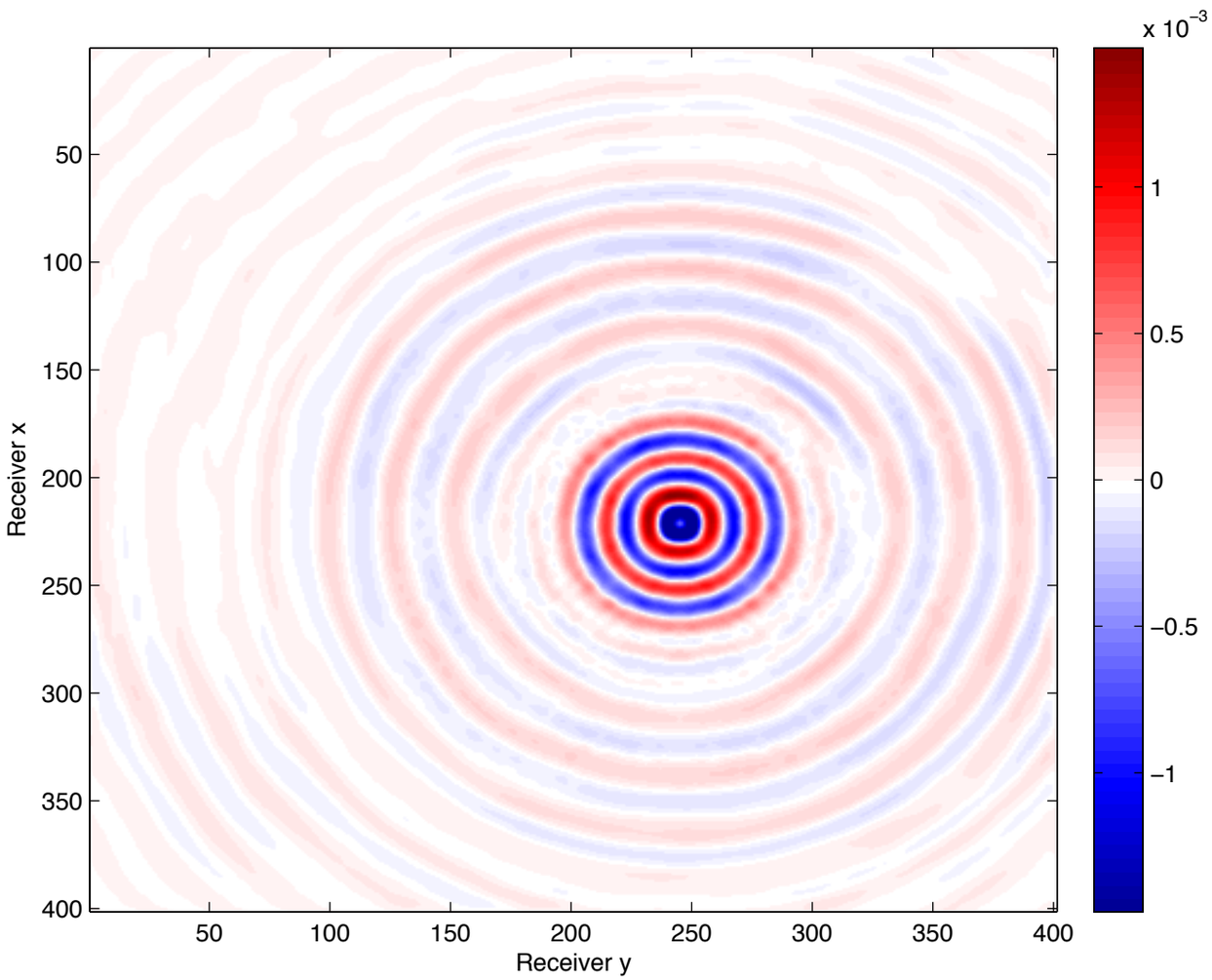
Interpolation

- 3-D synthetic BG data set
- 4624 shots spaced by 150m
- 401 x 401 receivers spaced by 25m
- 25% missing shots
- monochromatic data slices at 4.8 Hz

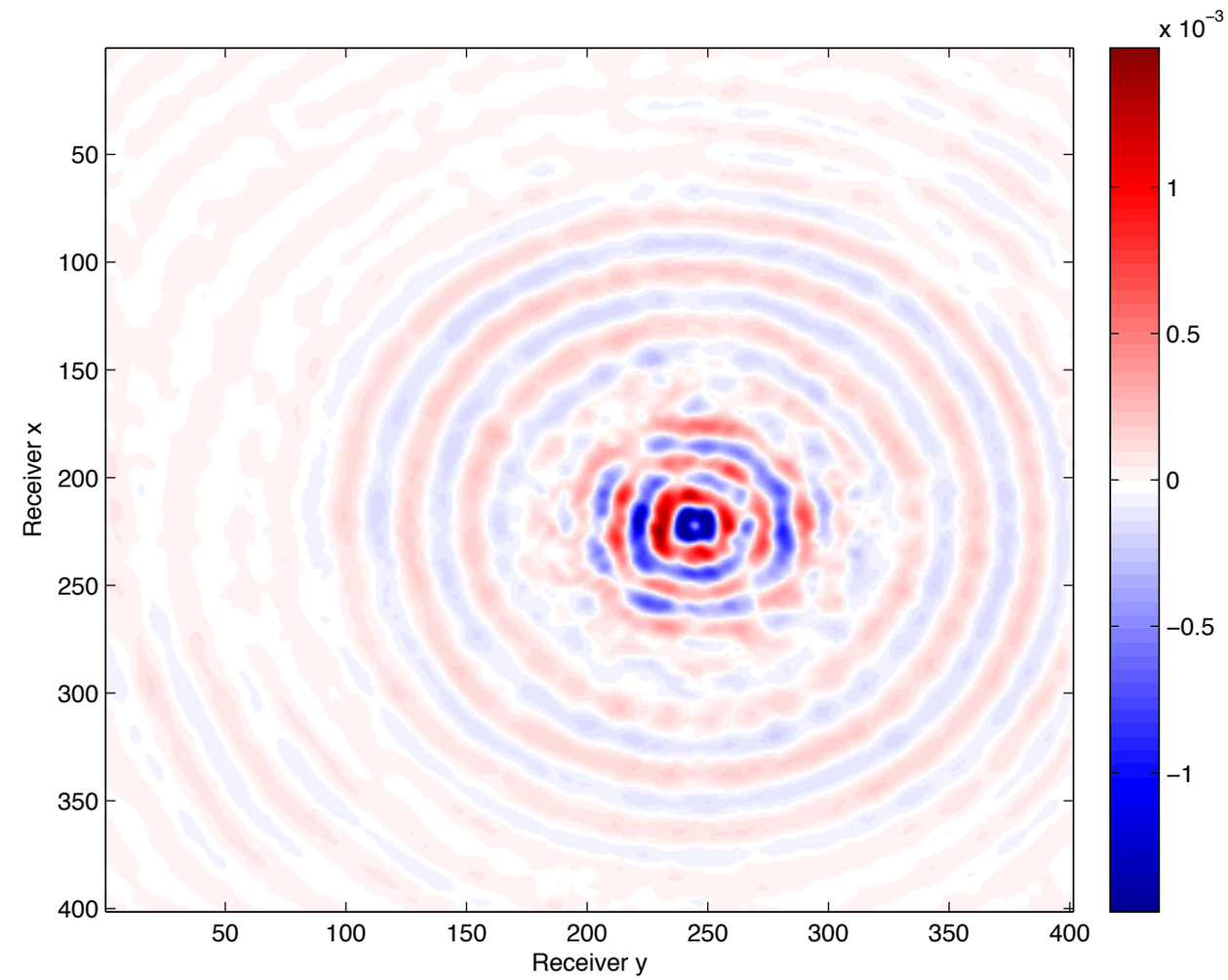
Recovery - Shot Gather

[Curvelet, 1000 iterations, 4.8 Hz]

true data
(src x, src y = 38,42)



approximated data
SNR=3.69 db

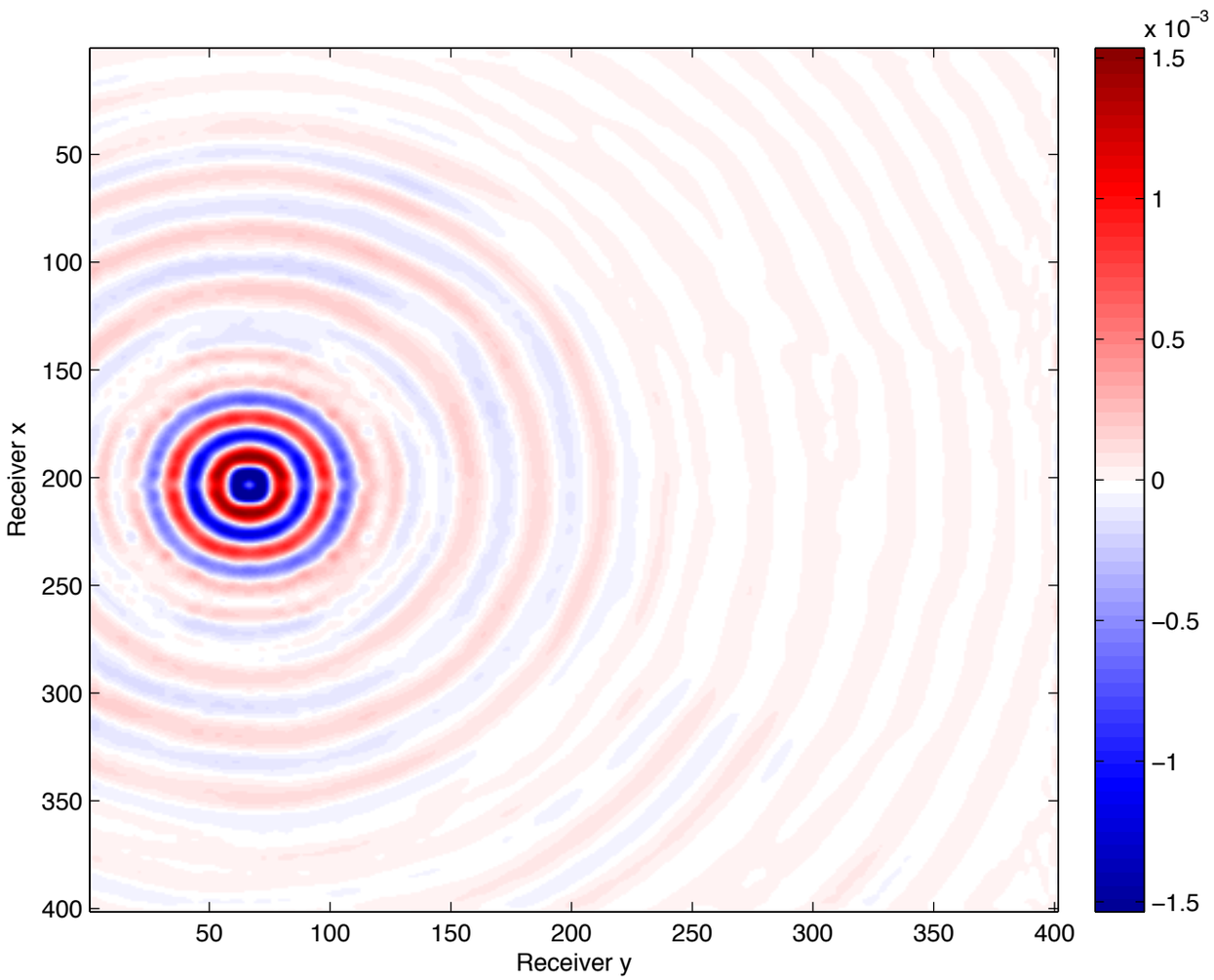


- average SNR = 3.12 dB

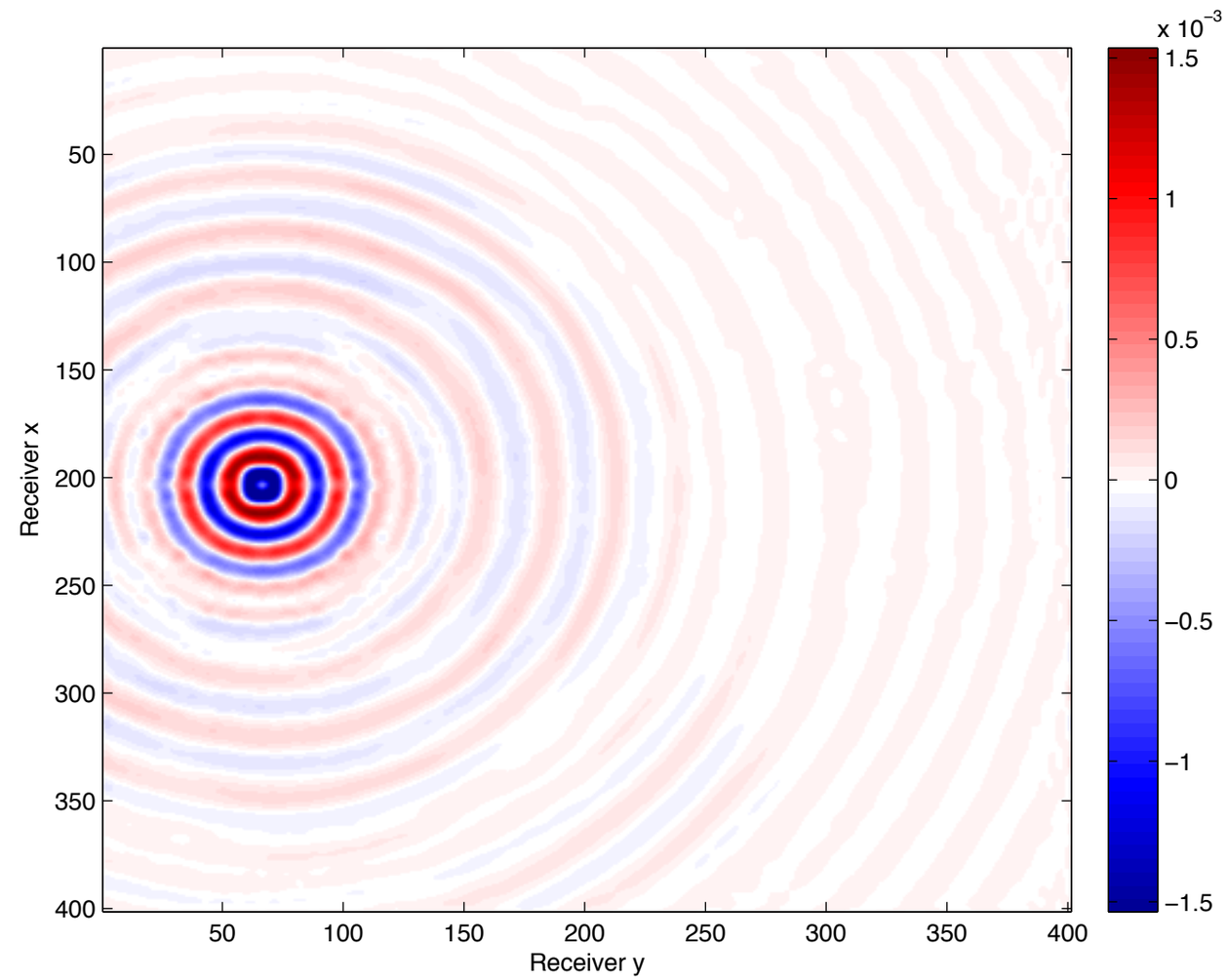
Recovery - Shot Gather

[low-rank, rank=120, 1000 iterations, 4.8 Hz]

true data
(src x, src y = 35,12)



approximated data
SNR=24.0 db

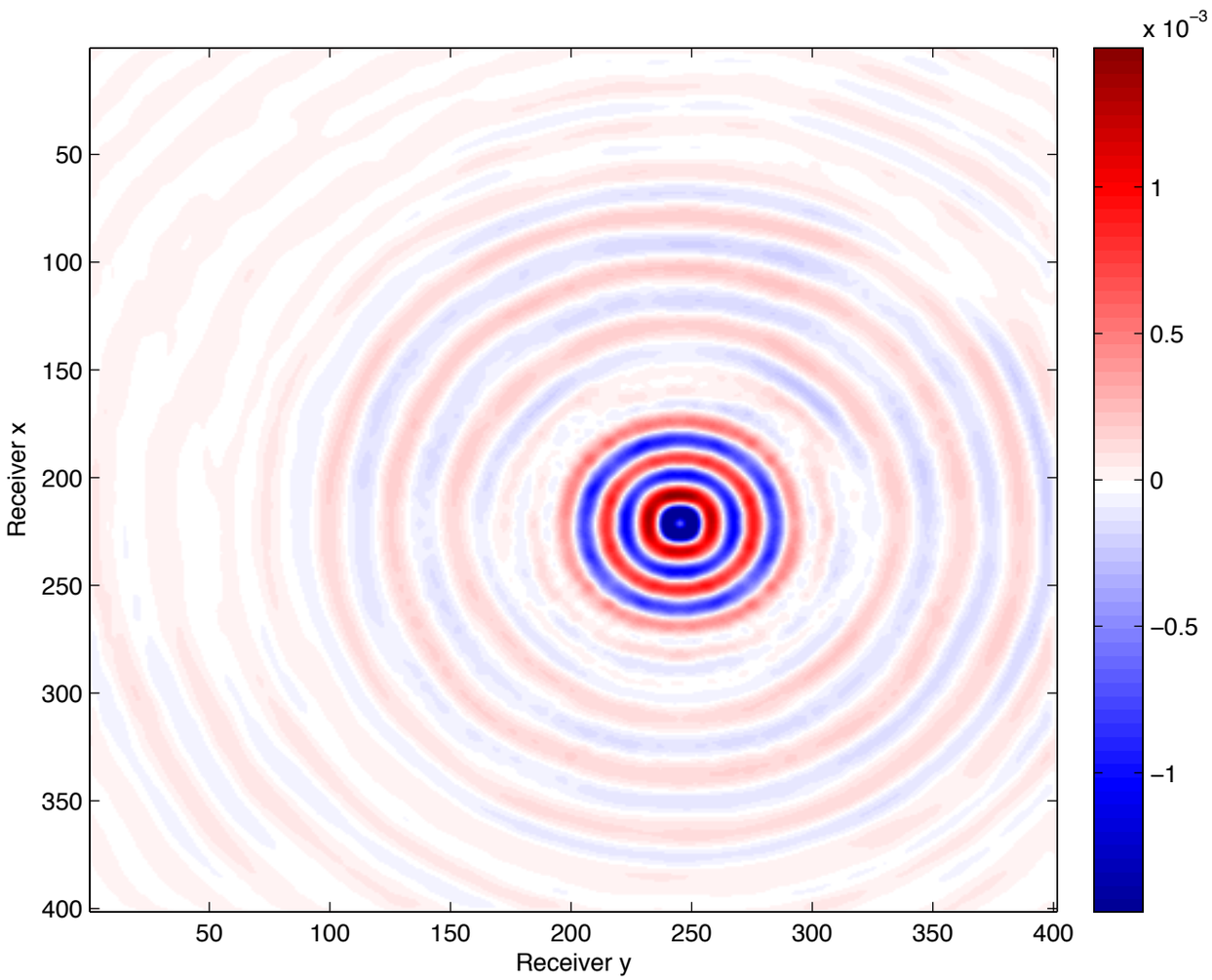


- average SNR = 21.5 dB

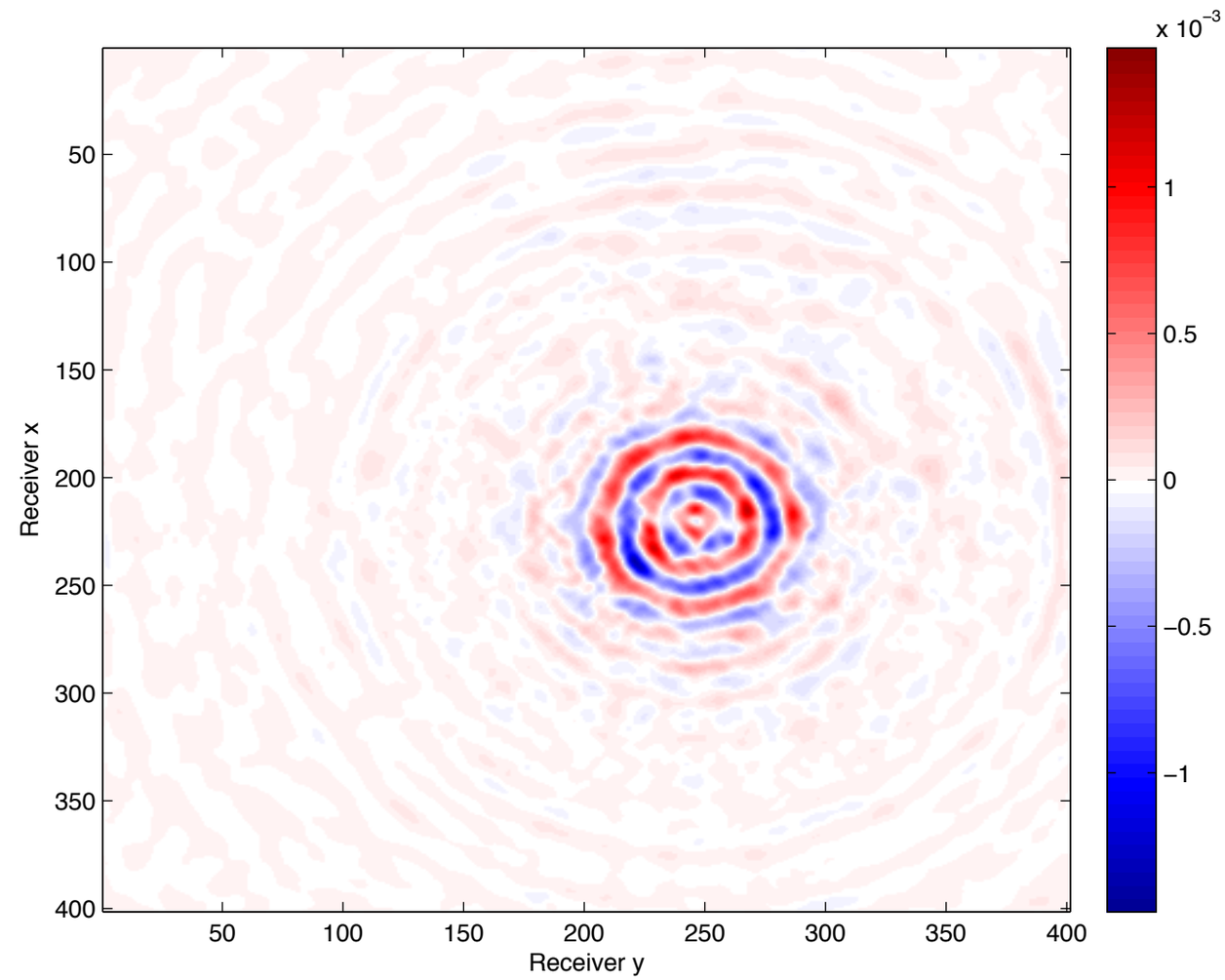
Recovery - Shot Gather

[Curvelet, 1000 iterations, 4.8 Hz]

true data
(src x, src y = 38,42)



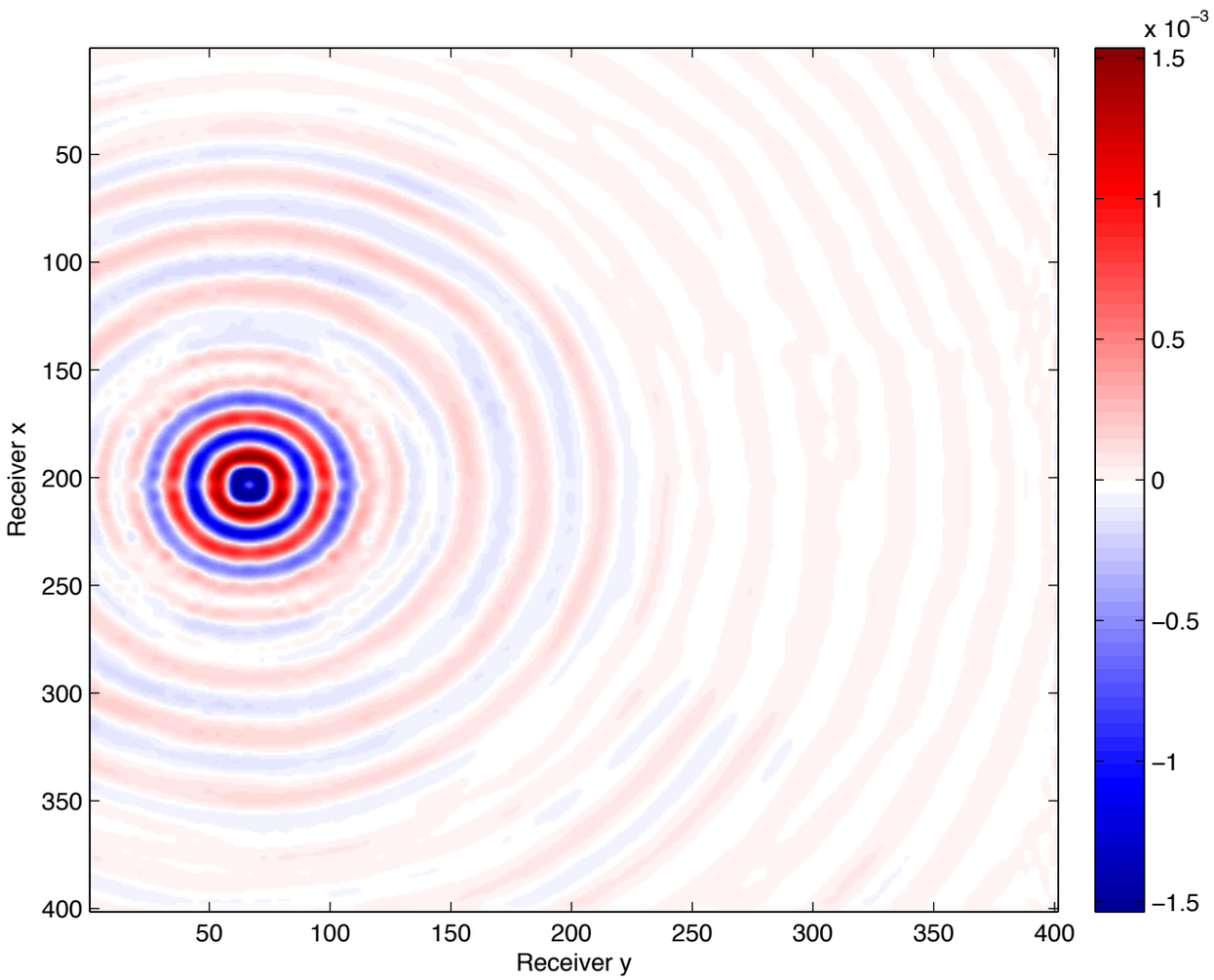
difference



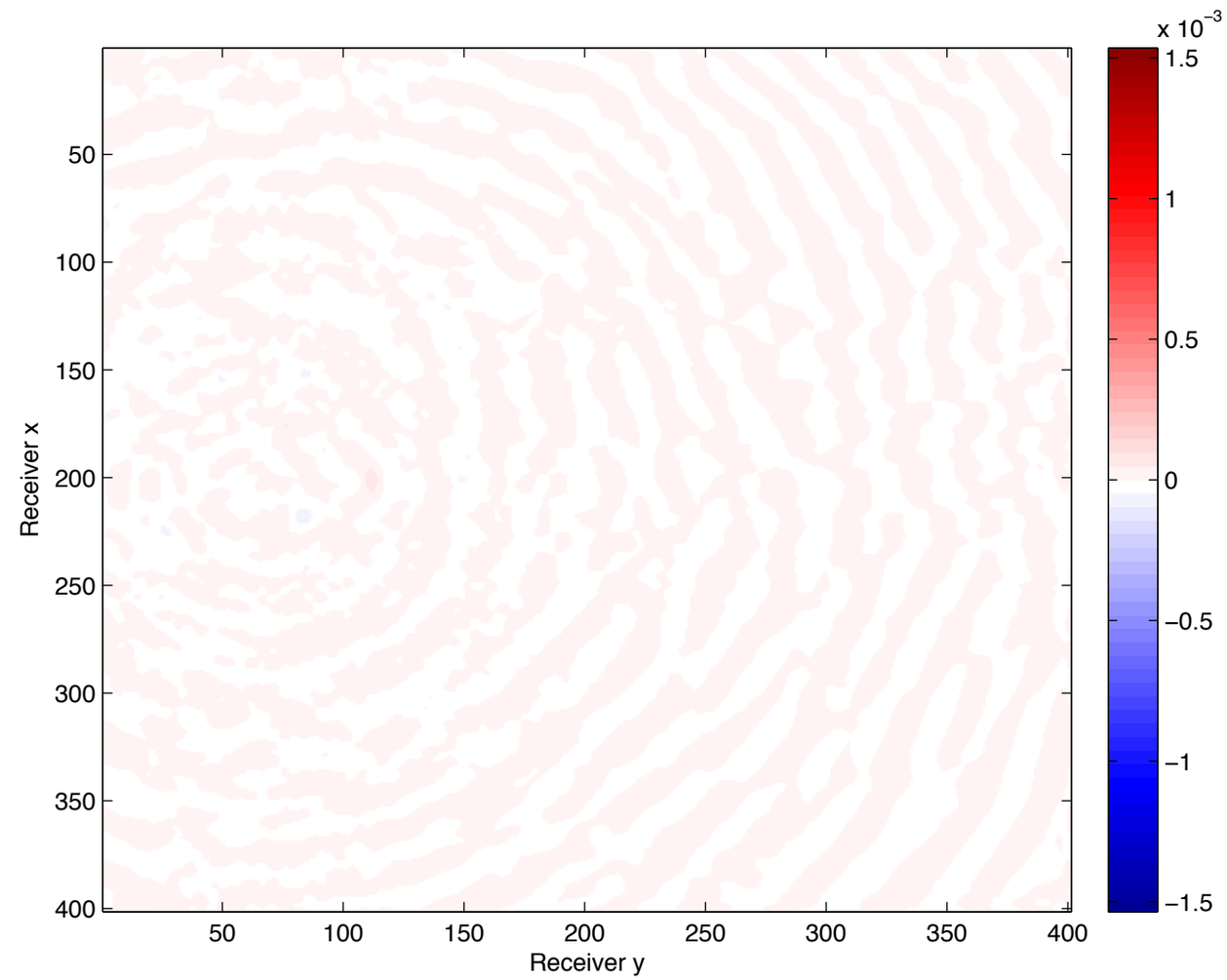
Recovery - Shot Gather

[low-rank, rank=120, 1000 iterations, 4.8 Hz]

true data
(src x, src y = 35,12)



difference



Comparison

	Low-rank	Curvelet
computational time	$O(hours)$	$O(days)$
storage	$k \times (n+m)$	$27 \times nm$

Conclusion

- matrix factorization allows *SVD-free* robust rank penalty methods that work fast on large data
- robust formulations are able to ignore *LARGE* unexplained artifacts in the data
- low-rank structure holds promise for data recovery and more compact representation.

References

- **Candes, E. J., and D. L. Donoho, 2000,** Curvelets - a surpris- ingly effective nonadaptive representation for objects with edges, *in* Curves and Surfaces: Vanderbilt University Press, Nashville, TN.
- **Rennie, J.D.M. and Srebro, N. [2005]** *ICML '05 Proceedings of the 22nd international conference on Machine learning.*
- **Berg, E.v. and Friedlander, M.P. [2008]** Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2), 890–912.
- **Herrmann, F. J., 2008**, Compressive sampling: a new paradigm for seismic data acquistion and processing?: Presented at the ION.
- **Herrmann, F. J., and G. Hennenfent, 2008**, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, 173, 233–248.
- **Lee, J., B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp, 2010**, Practical large- scale optimization for max-norm regularization: Analysis, 23, 1–9.
- **Recht, B., Fazel, M. and Parrilo, P. [2010]** Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3), 471–501
- **Candes, E., and Y. Plan, 2010,** Matrix completion with noise: Proceedings of the IEEE, 98, 925 –936.

References

- **Oropeza, V. and Sacchi, M. [2011]** Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics*, 76(3), V25–V32.

- **Recht, B. and Ré, C. [2011]** Parallel stochastic gradient algorithms for large-scale matrix completion. *Optimization Online*.

- **Aravkin, A., Burke, J. and Friedlander, M. [2012b]** Variational properties of value functions. *submitted to SIAM J. Opt., arXiv:1211.3724*.

- **Aravkin, A., Friedlander, M.P., Herrmann, F. and T. van Leeuwen, 2012,** Variational properties of value functions. submitted to SIAM J. Opt., arXiv:1211.3724

- **Xiang Li, Aleksandr Y. Aravkin, Tristan van Leeuwen, and Felix J. Herrmann, 2012,** Fast randomized full-waveform inversion with compressive sensing, Geophysics, vol. 77, p. A13-A17.

- **Aleksandr Y. Aravkin, Rajiv Kumar, Hassan Mansour, Ben Recht, Felix J. Herrmann, 2013,** A robust SVD-free approach to matrix completion, with applications to interpolation of large scale data, arXiv:1302.4886

Acknowledgements

Thanks everyone for your attention

SINBAD

NSERC CRSNG

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, ION GXT, CGGVeritas, PGS, Woodside,Total SA, and WesternGeco.