SVD-free matrix completion for seismic data reconstruction

Rajiv Kumar, Oscar Lopez, Ernie Esser and Felix J. Herrmann

SLIM
University of British Columbia

SVD-free matrix completion for seismic data reconstruction

Rajiv Kumar, Oscar Lopez, Ernie Esser and Felix J. Herrmann


SLIM
University of British Columbia

Outline

- interpolation
  - comparison with curvelet-based reconstruction methods

- regularization
  - is binning the right approach?

Motivation

- acquisition challenges
  - missing data
  - irregular acquisition grid
- fully sampled data
  - simultaneous shot based FWI & migration
  - estimation of primaries by sparse inversion & SRME
- regularization
  - imaging and inversion algorithm require equi-spaced grid
- exploit *low-rank* structure of seismic data
  - *SVD-free* matrix factorization

Outline

- interpolation
  - comparison with curvelet-based reconstruction methods
- regularization
  - is binning the right approach?

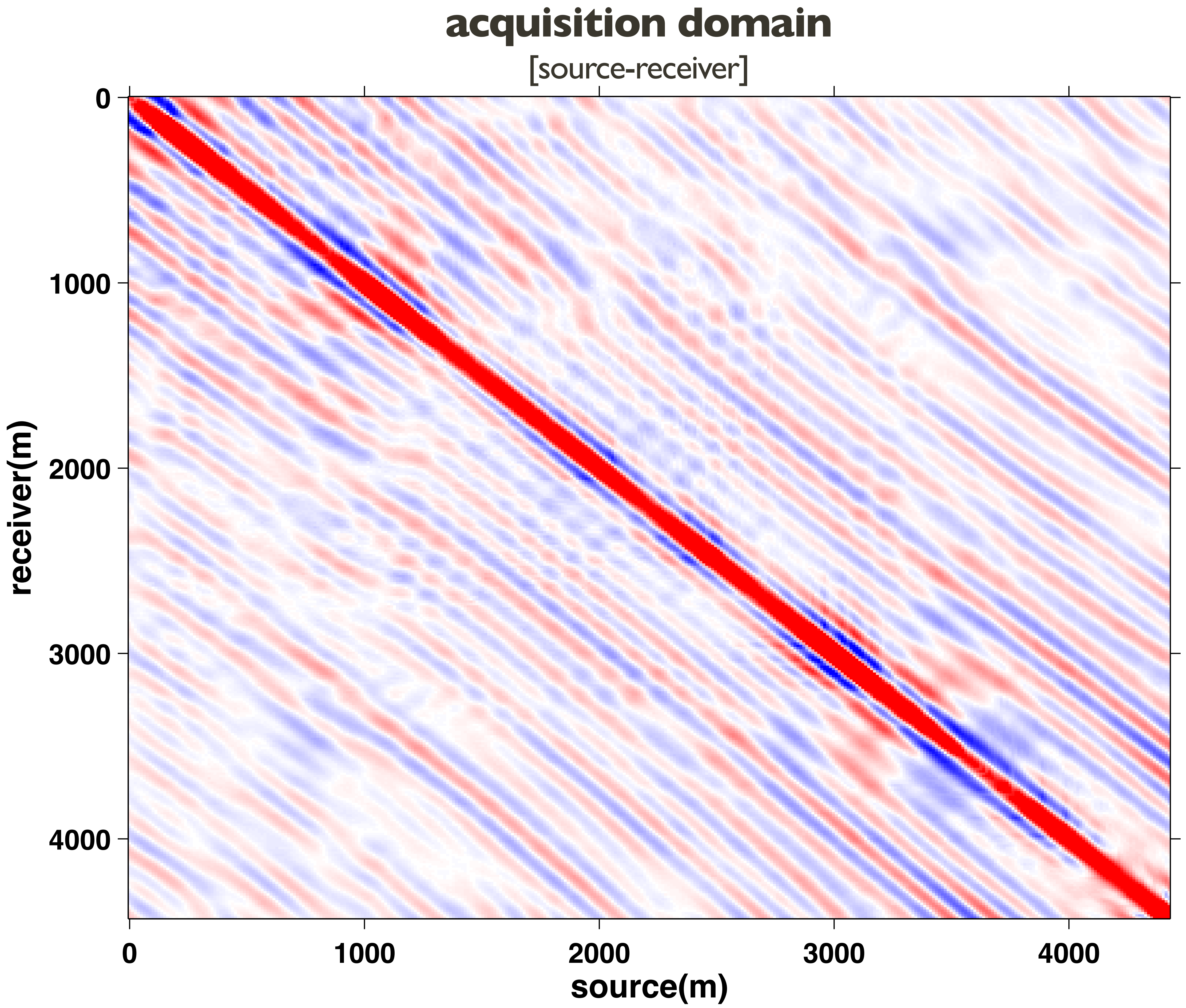
[Candes and Plan 2010, Oropeza and Sacchi 2011]

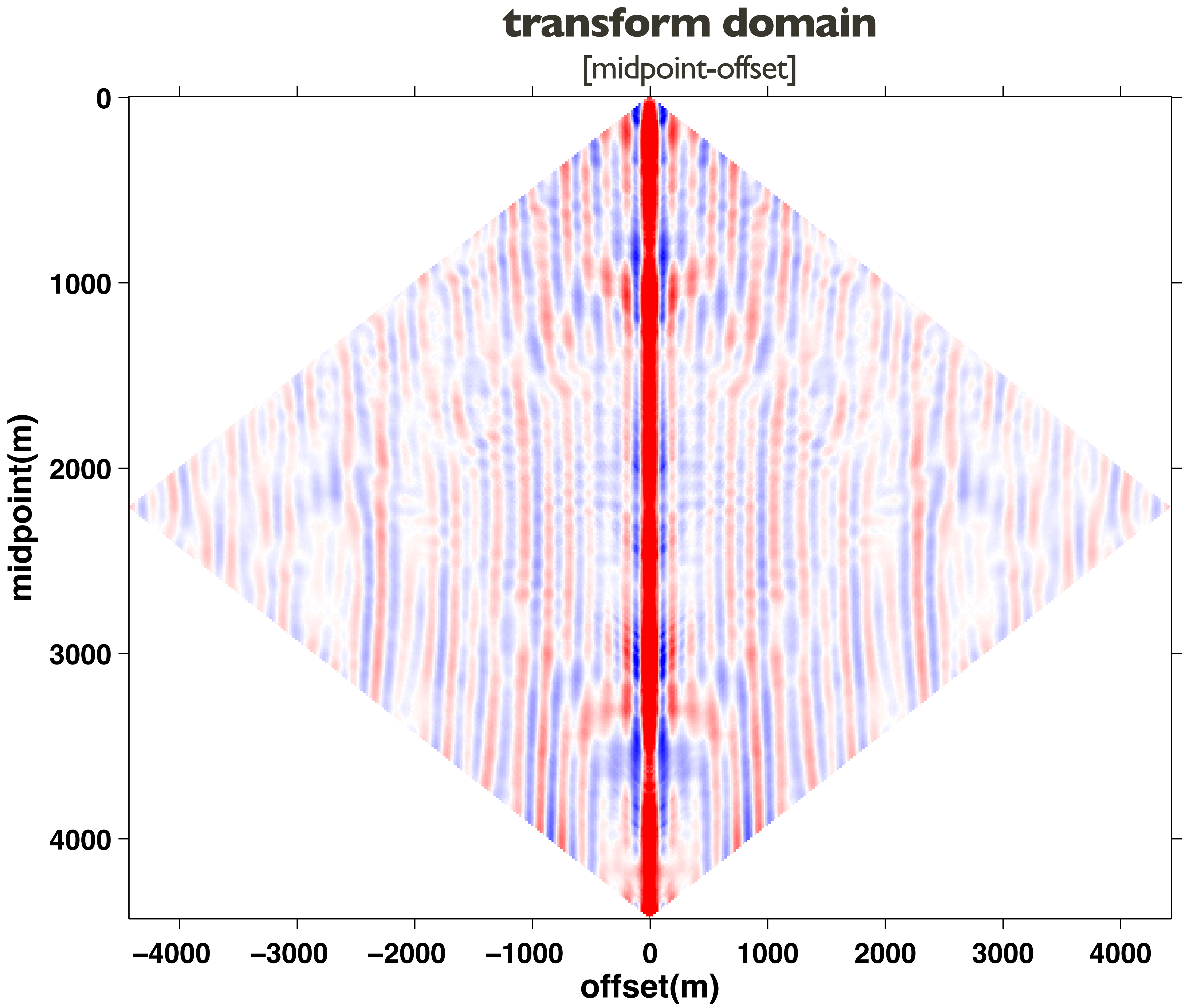
Matrix completion

- 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

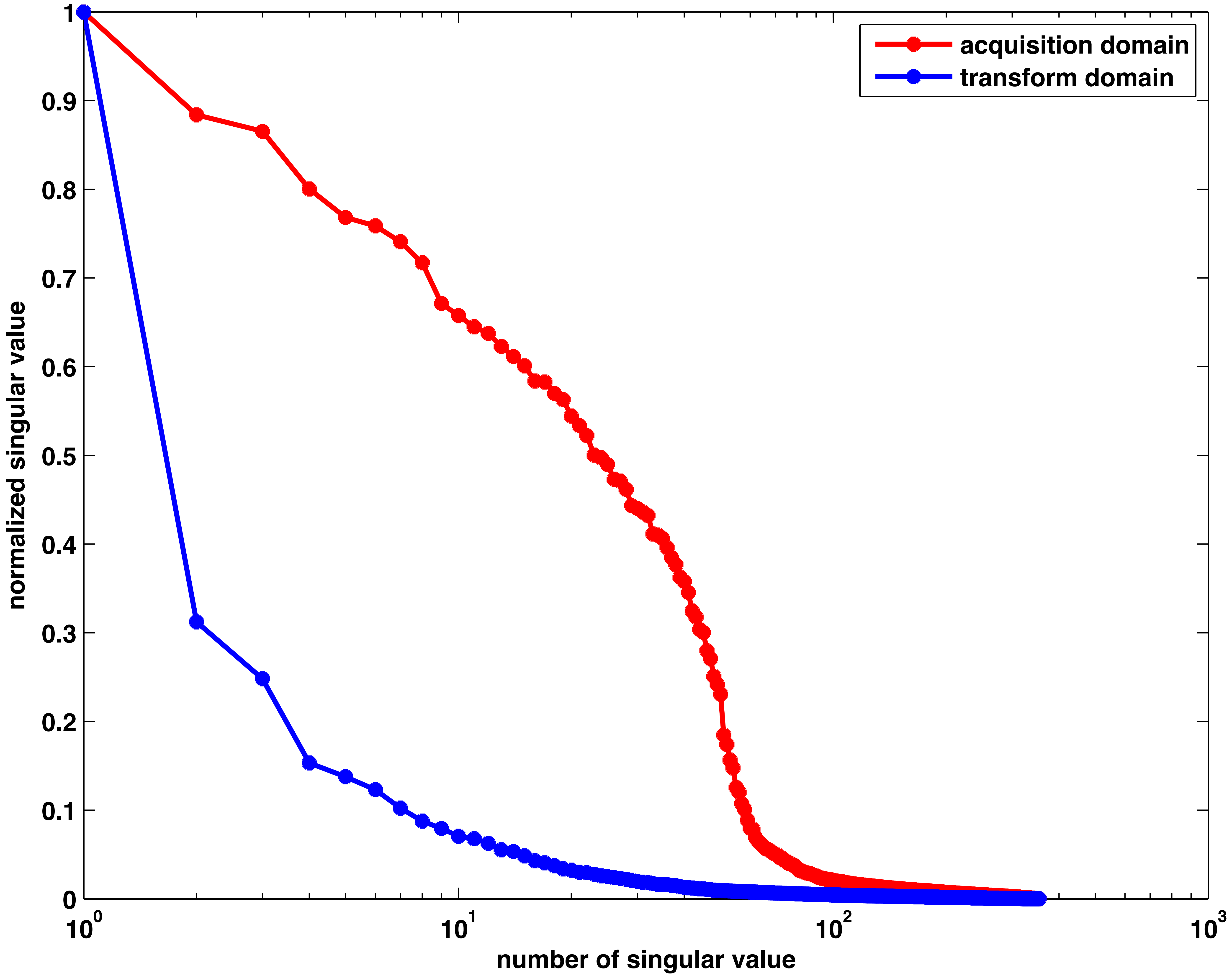
2-D acquisition





Singular value decay

2-D acquisition



Matrix completion

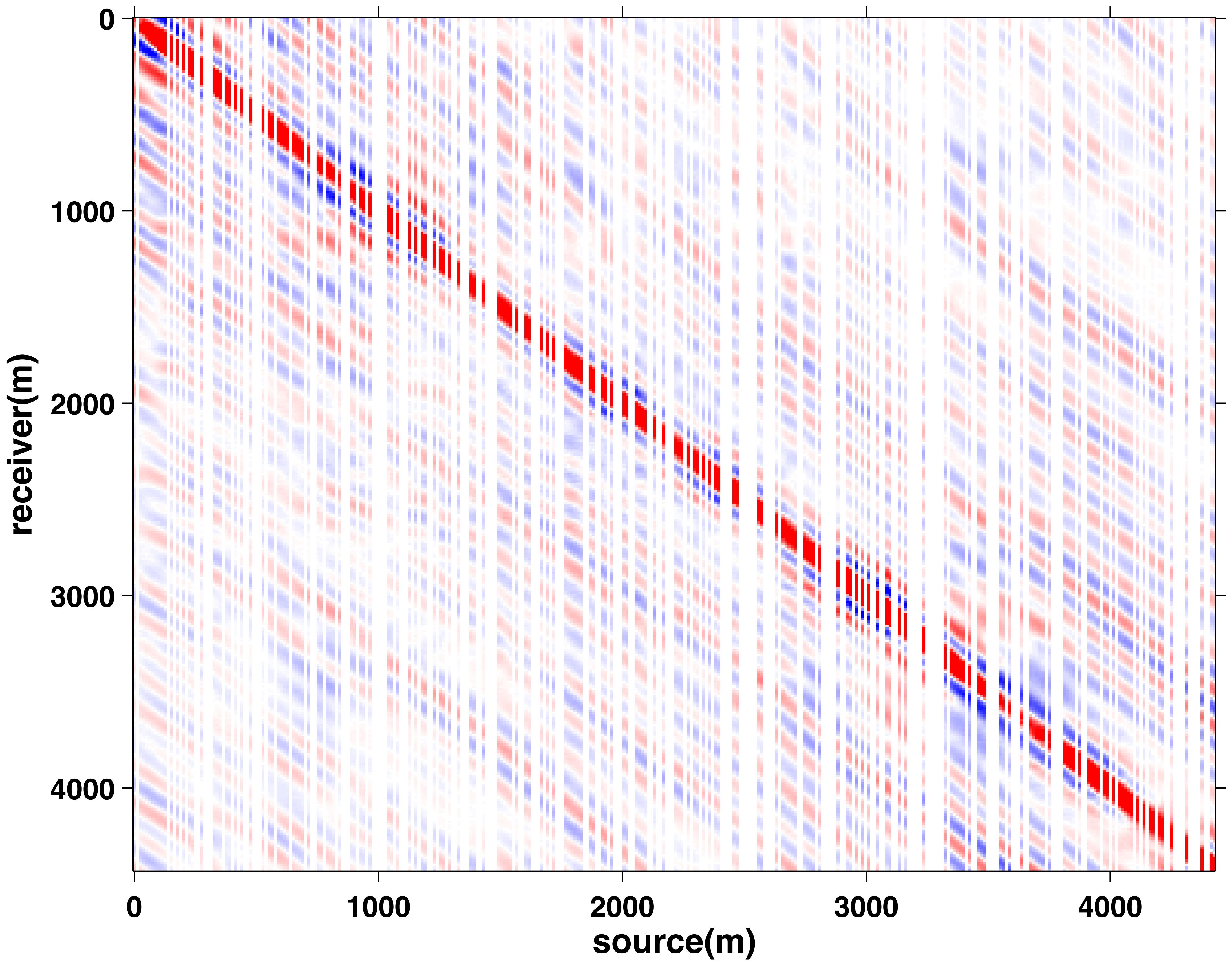
- signal structure
  - *low rank/fast decay* of singular values
- sampling scheme
  - missing data *increase* rank in "transform domain"
- recovery using *rank penalization* scheme

2-D acquisition

uniform-random sampling

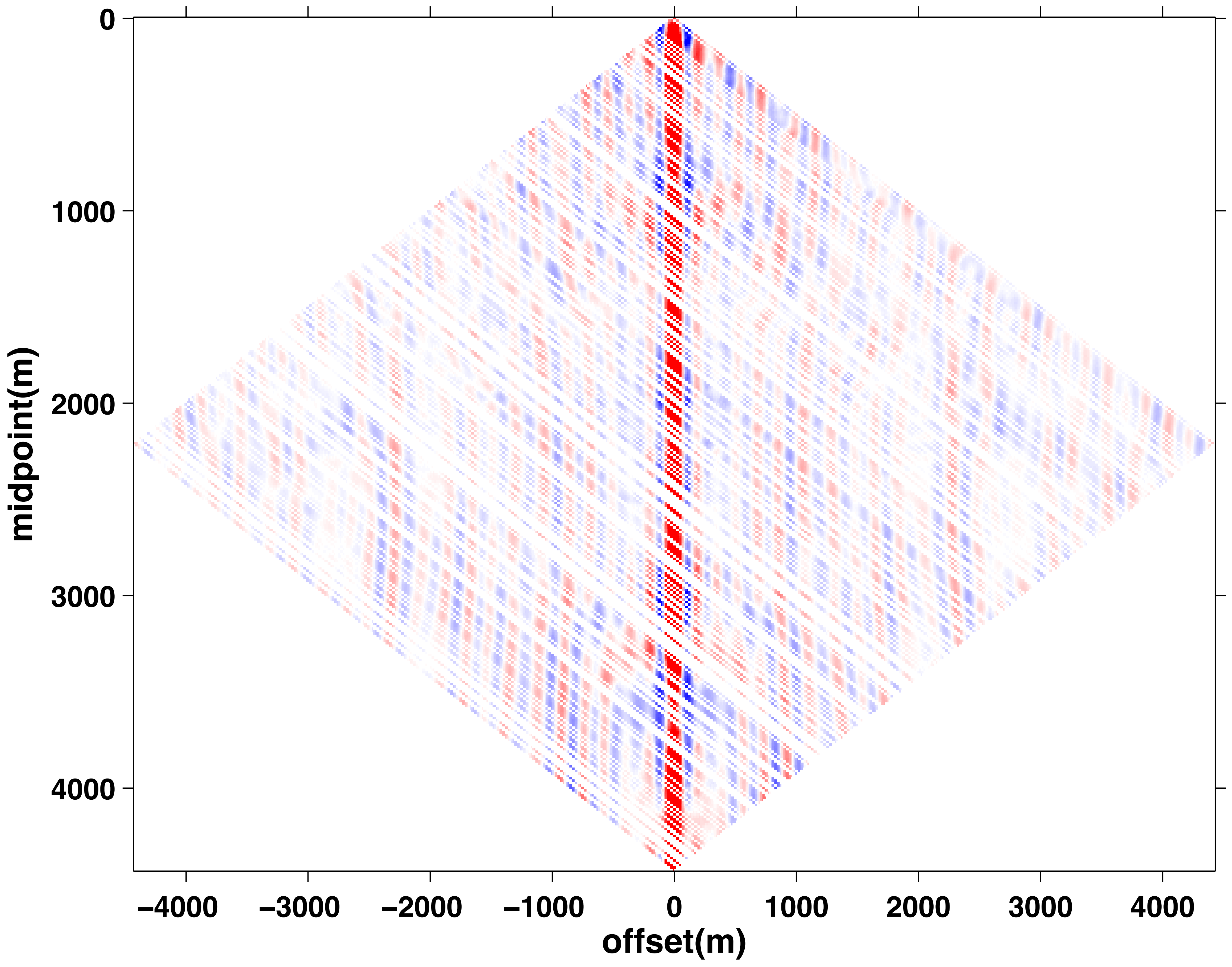
acquisition domain

missing columns *do not* increase rank

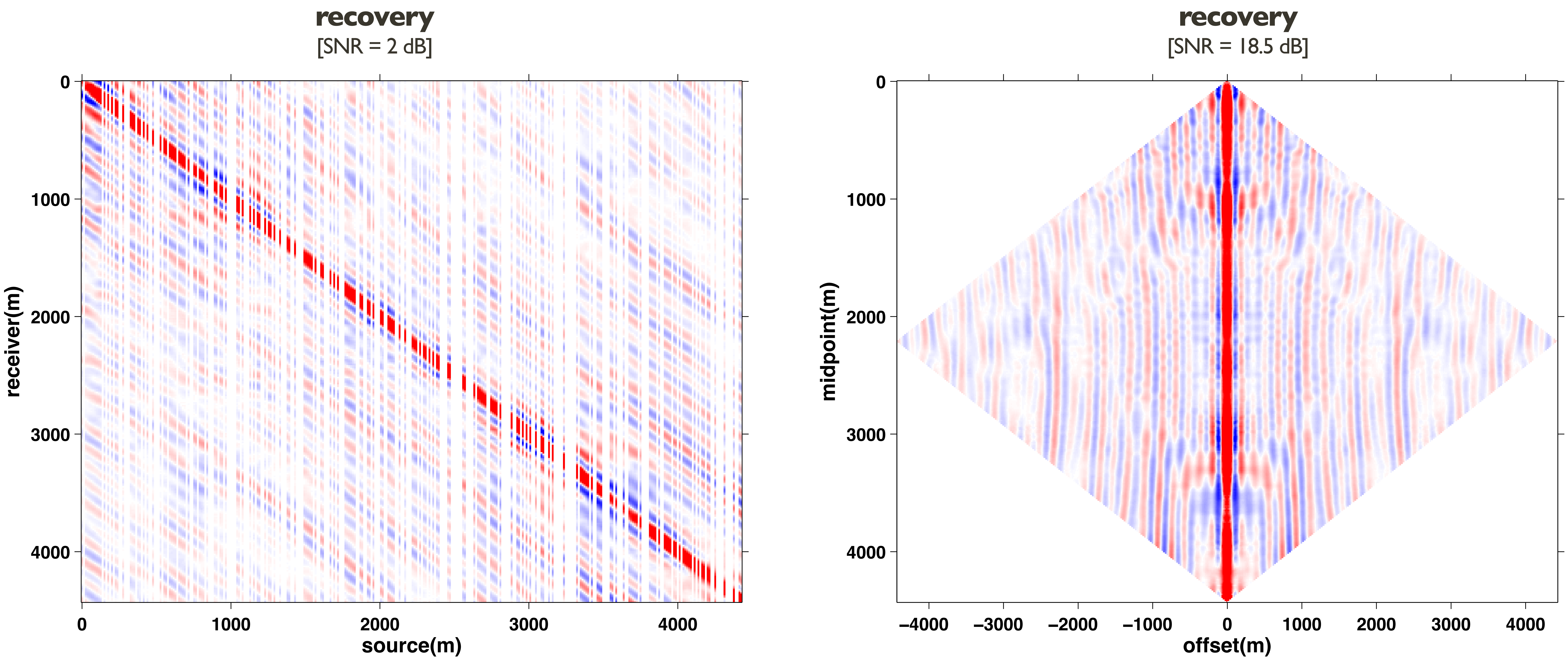


transform domain

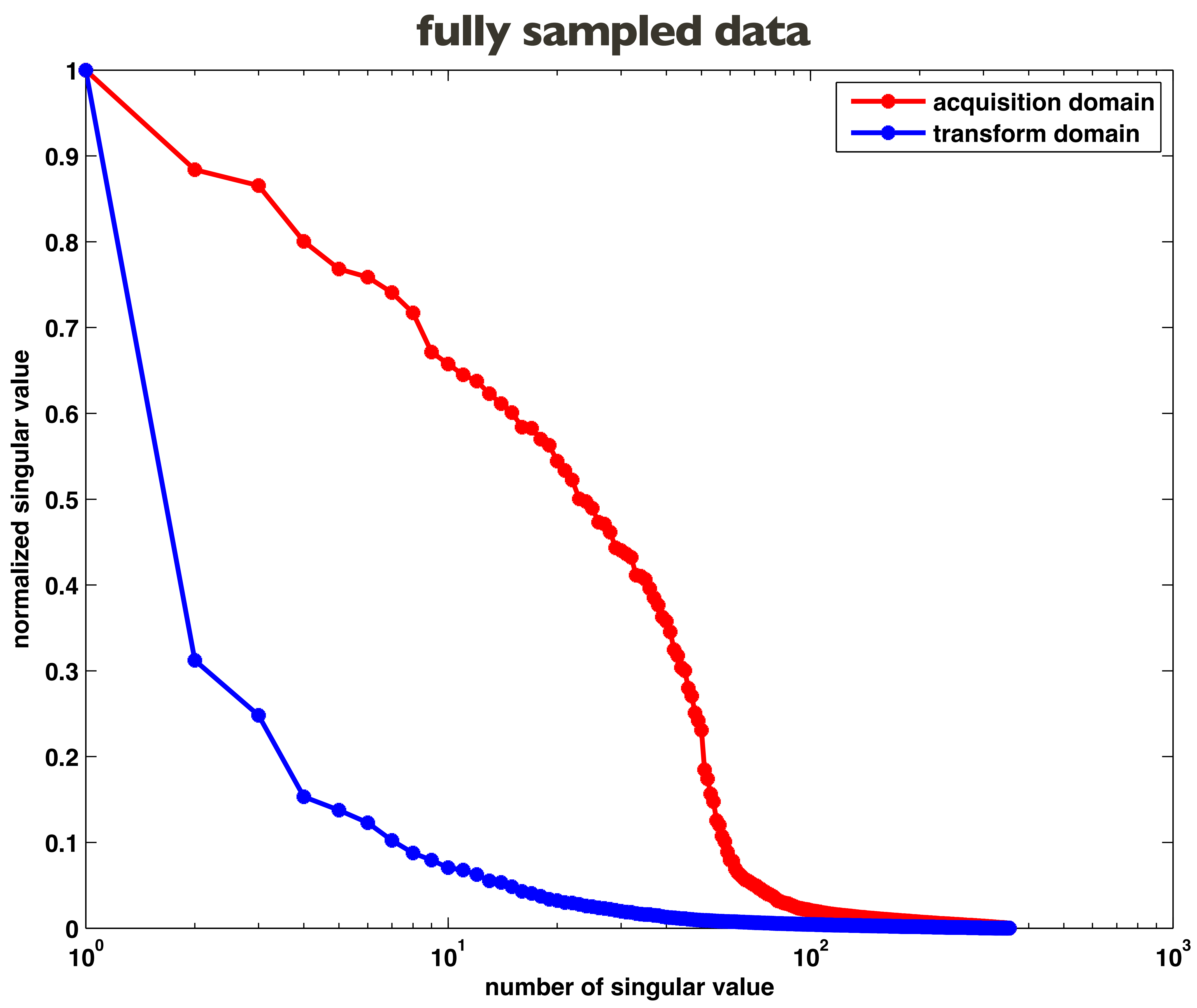
missing columns *do* increase rank

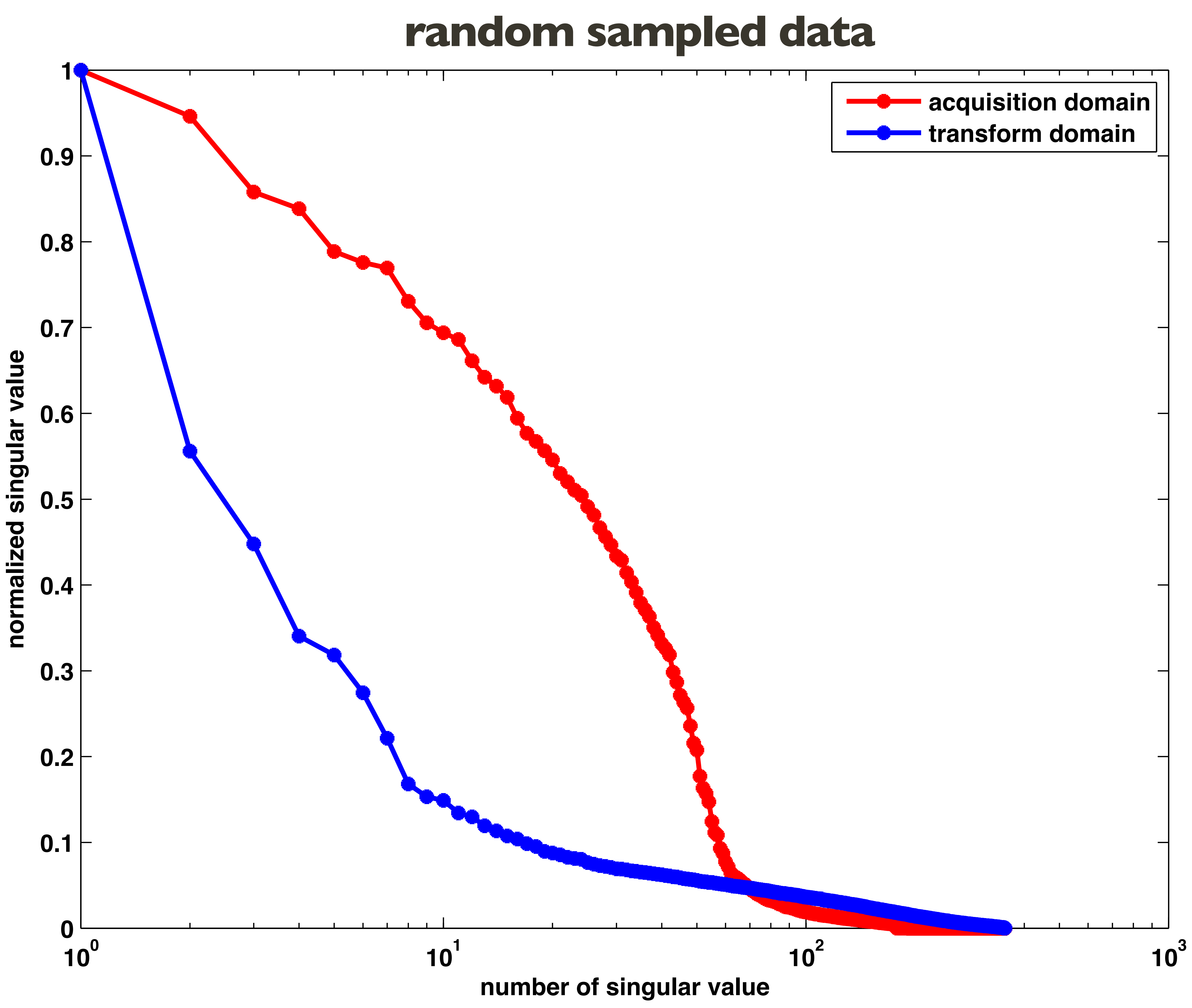


Low-rank interpolation



Randomized sampling
singular value decay





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

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

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

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

$$\mathcal{A} = \mathbf{RM}\mathcal{S}^H$$

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]

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 projections

- 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

$$\mathbf{X} \in \mathbb{R}^{n\times m} \quad = \quad \mathbf{L} \in \mathbb{R}^{n\times k} \quad \mathbf{R}^H \in \mathbb{R}^{k\times m}$$

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

[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

- Upper-bound on nuclear norm is defined 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 k *explicitly* & avoid costly *SVD's*

Computational cost

with and without SVD

		50%		75%	
	σ	0.1	0.08	0.1	0.08
Matrix completion w/ SVD	SNR (dB)	17.3	18.3	11.6	11.5
	time (sec)	812	937	790	765
Matrix completion w/o SVD	SNR (dB)	17.6	18.4	12.6	13.3
	time (sec)	8	10	8	7

Upcoming paper

Check https://www.slim.eos.ubc.ca soon!

Computational cost

matrix completion v/s curvelet-based methods

		50%		75%	
	σ	0.1	0.08	0.1	0.08
Matrix completion w/ SVD	SNR (dB)	17.3	18.3	11.6	11.5
	time (sec)	812	937	790	765
Matrix completion w/o SVD	SNR (dB)	17.6	18.4	12.6	13.1
	time (sec)	8	10	8	7
Curvelet-based sparsity promotion	SNR (dB)	17.4	18.6	12.5	12.8
	time (sec)	879	989	817	1010

Observation

matrix completion v/s curvelet-based methods

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

Take-away message

- can avoid *"SVD"*
- faster compare to curvelet-based sparsity promotion techniques
- memory efficient compare to curvelet-based techniques

Outline

- interpolation
  - comparison with curvelet-based reconstruction methods

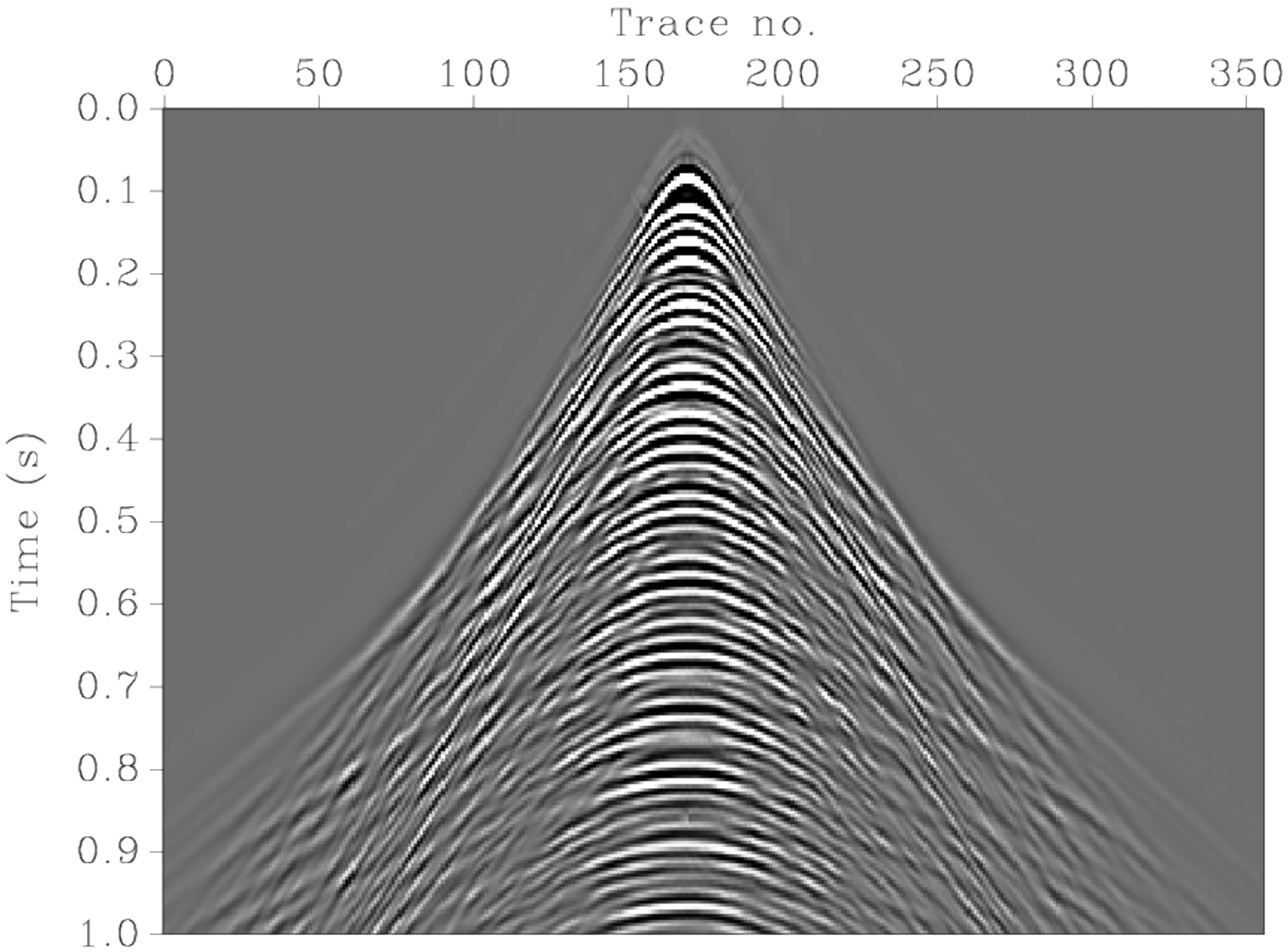
- regularization
  - is binning the right approach?

Regularization

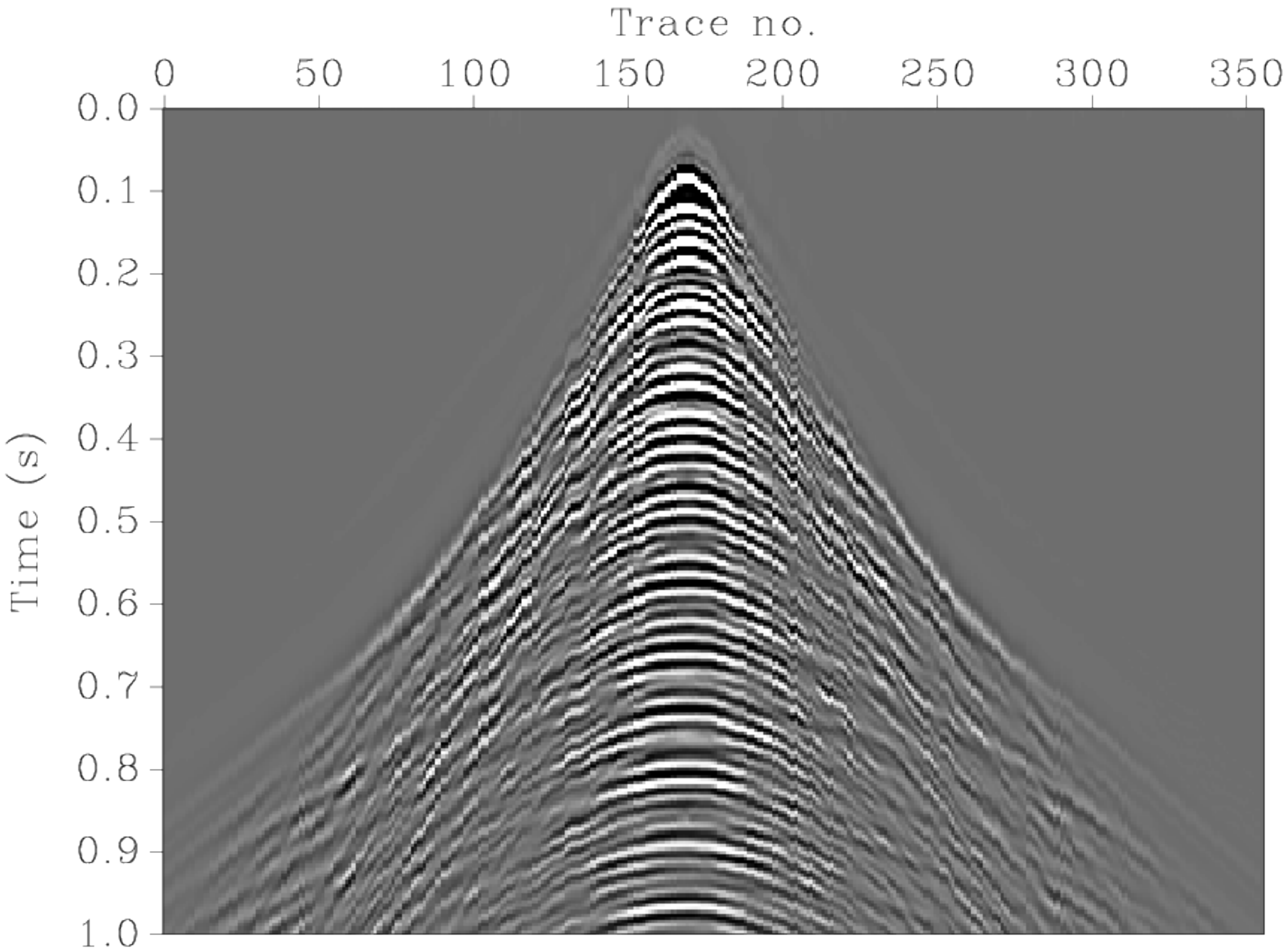
- unstructured acquisition grid
- imaging and inversion algorithm
  - regularly sampled data
- binning
  - does not preserve the data-structure

Regularization

Ground truth



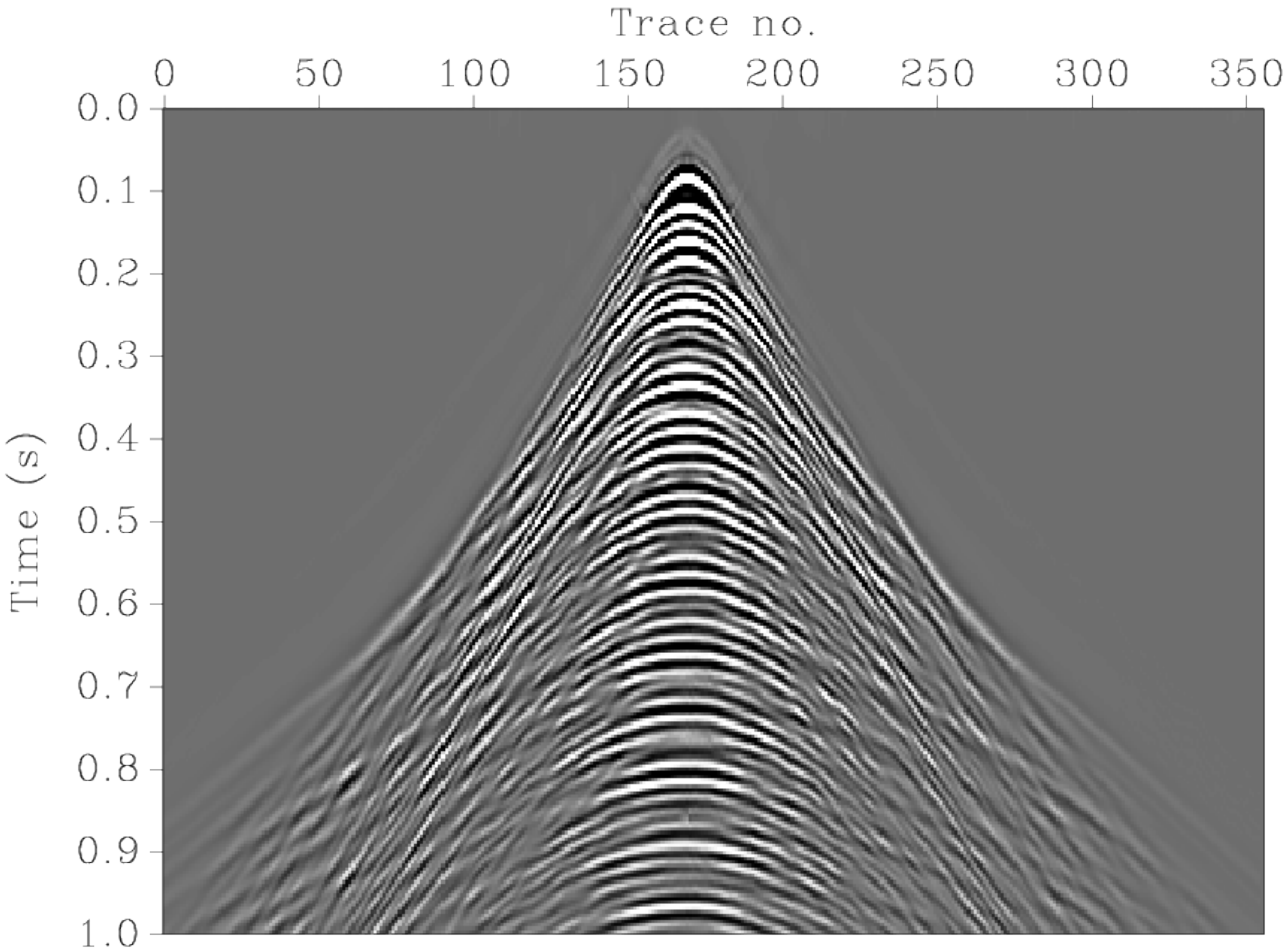
Observed data



Regularization

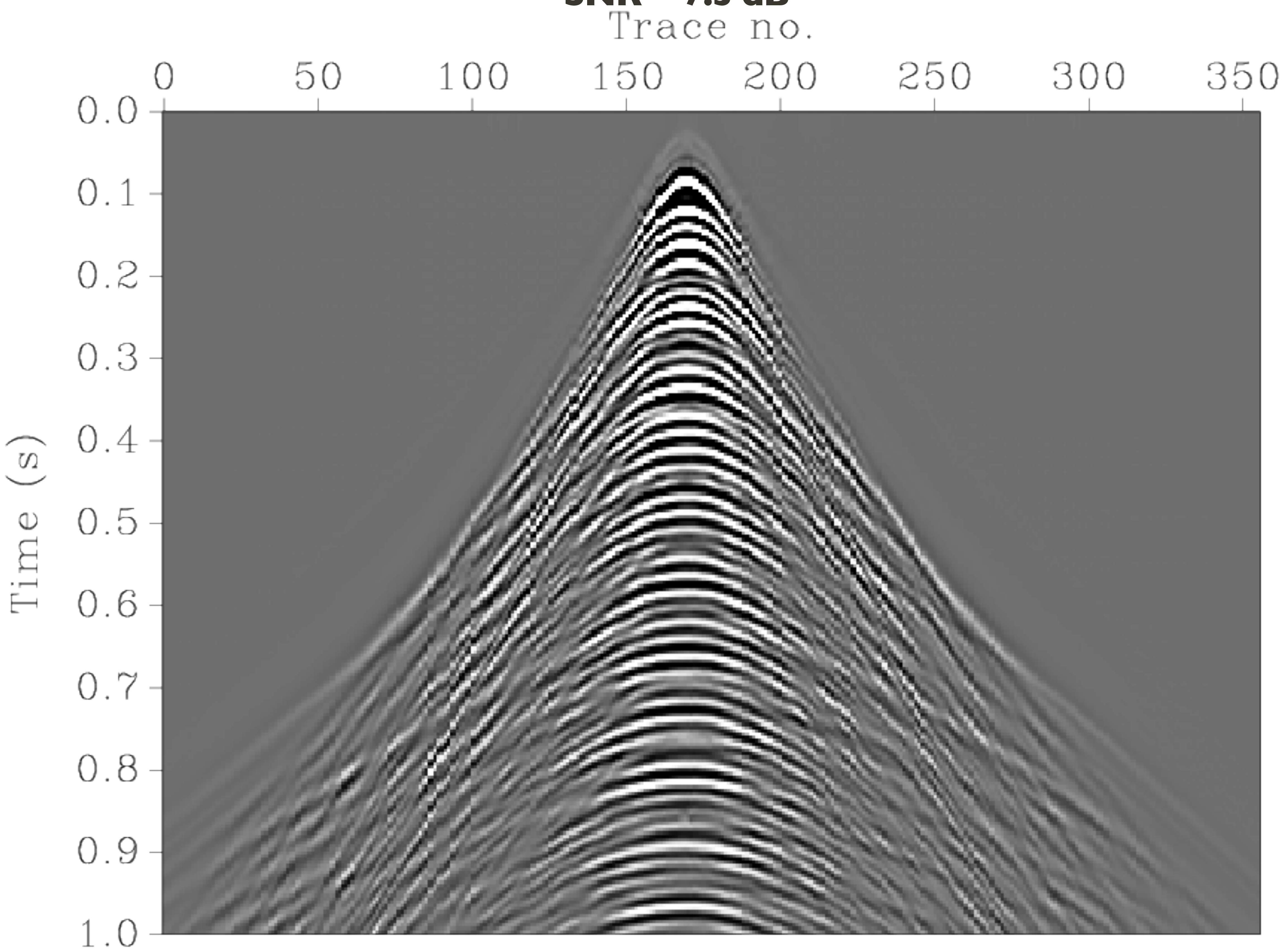
binning

Ground truth



Recovery

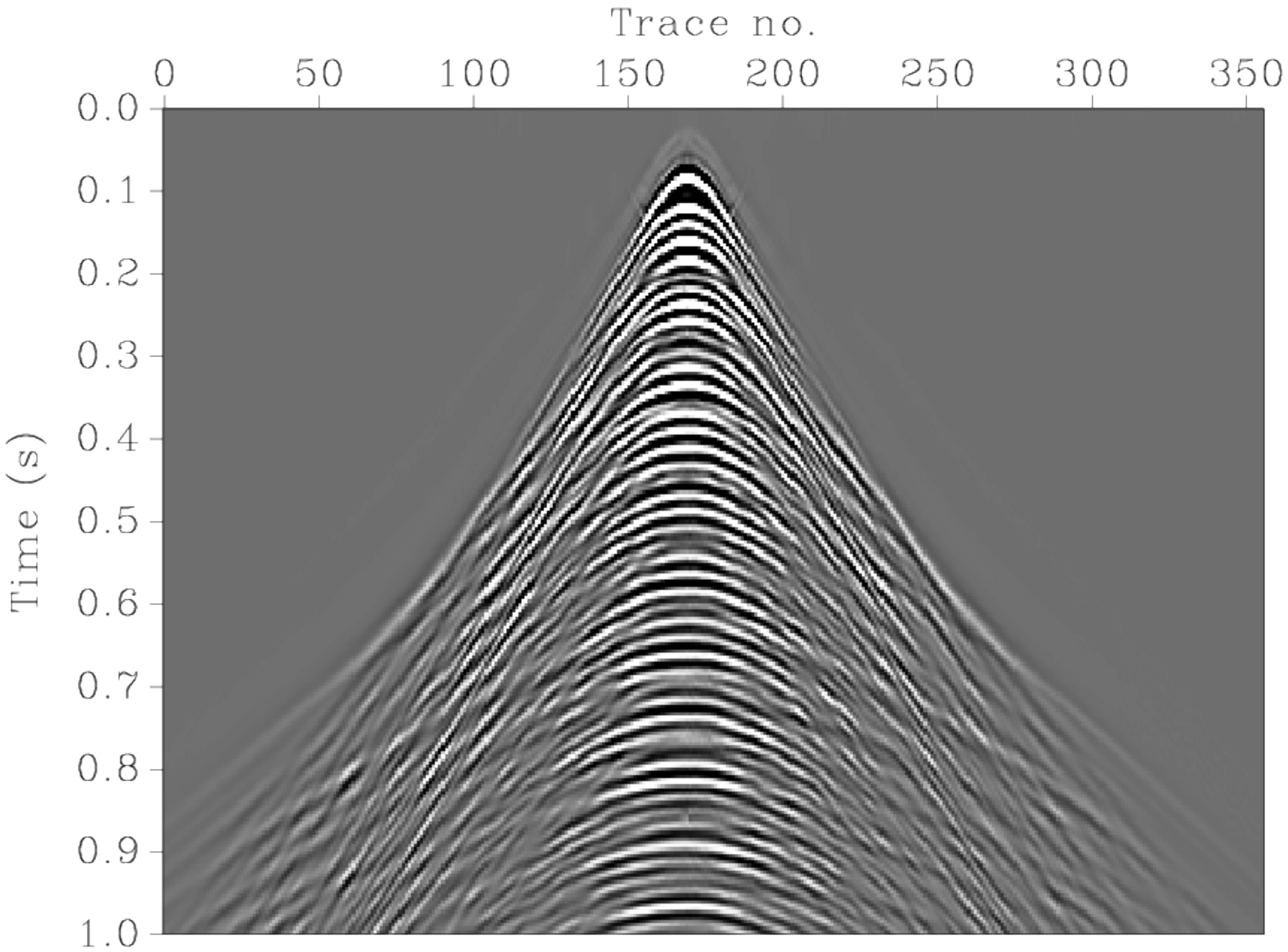
SNR = 7.3 dB



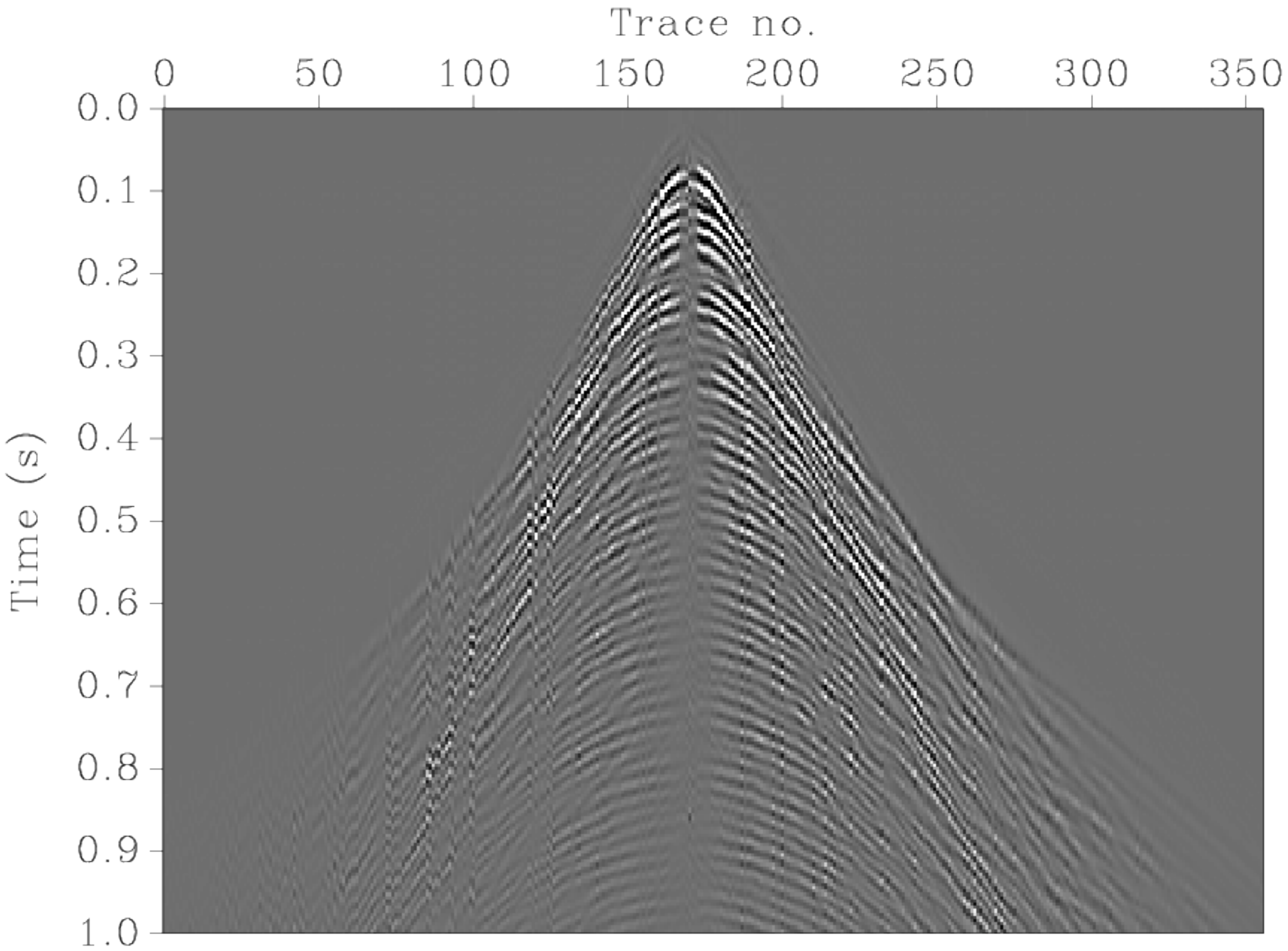
Regularization

binning

Ground truth

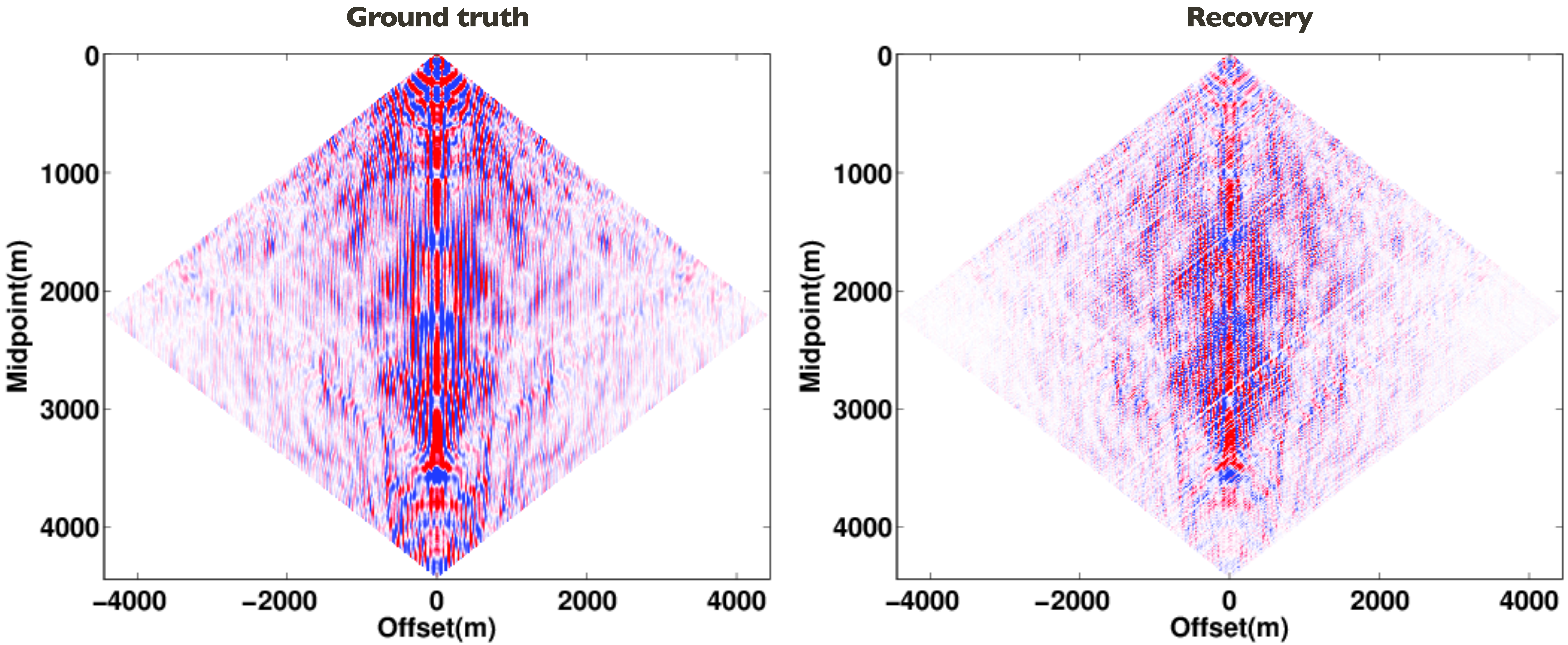


Residual



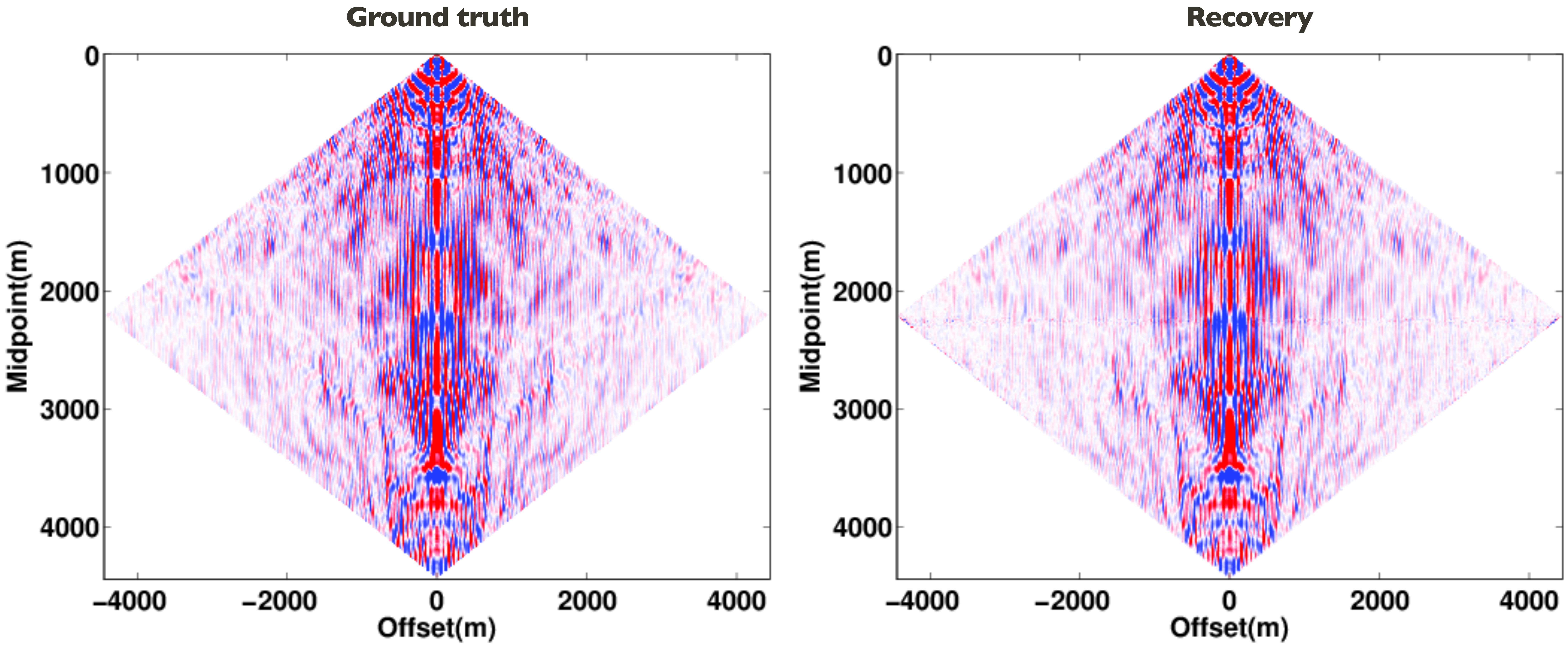
Low-rank structure

binning, midpoint-offset domain



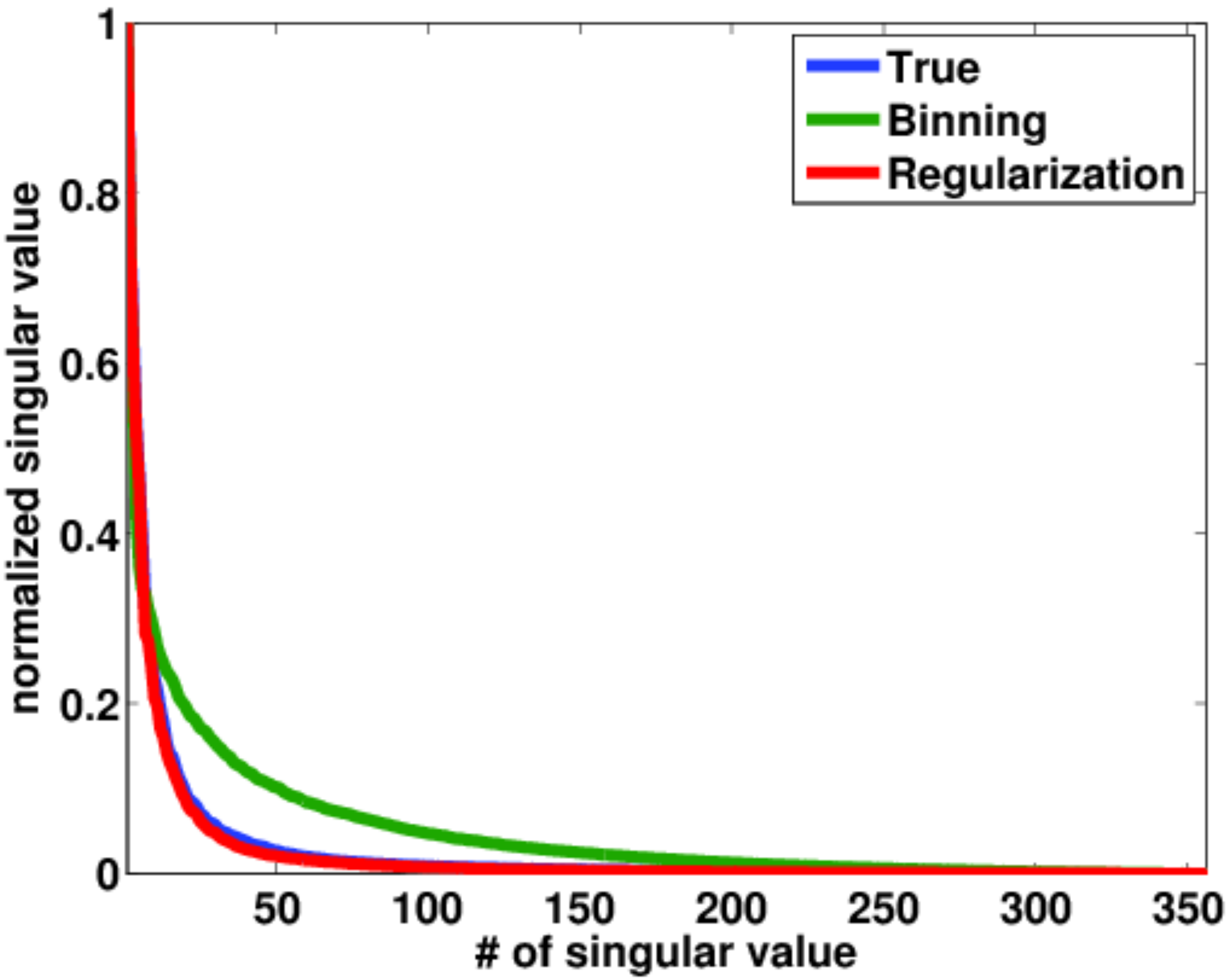
Regularization

matrix completion, midpoint-offset domain



Singular value decay

regularization v/s binning



Methodology
matrix completion

- given a regularization operator $\mathbf{N} : \mathbb{C}^{n\times m} \to \mathbb{C}^{n\times m}$ so that $\mathbf{N}(\mathbf{X}_r) = (\mathbf{X}_{ir})$, transform-sampling operator is redefine as

$$\mathcal{A} = \mathbf{RMN}^H \mathcal{S}^H$$

where

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

SLIM

Theorem

matrix completion

Let $\mathbf{X_r} \in \mathbb{C}^{n\times m}, \hat{\mathbf{X}}_r \in \mathcal{S}$ and $\mathbf{b} = \mathbf{RM}(\mathbf{X}_{ir}) + e$ with $\|e\| \leq \eta$. Let $\tilde{\mathbf{X}}$ be the solution of BPDN_σ, then

$$\|\mathcal{S}(\mathbf{X}_r - \tilde{\mathbf{X}})\| \leq \underbrace{\frac{C_1}{\sqrt{k}} \sum_{j=k+1}^{l} \sigma_j(\hat{\mathbf{X}}_r)}_{\textbf{interpolation error}} + \underbrace{(\frac{C_1}{\sqrt{k}} B_2 + 1)\|P\|_F}_{\textbf{regularization error}} + \underbrace{C_2\eta}_{\textbf{noise}}$$

where

$$P = \mathbf{N}^{-1}(\mathbf{X}_{ir}) - \mathbf{X}_r$$

$$l = min\{n, m\}$$

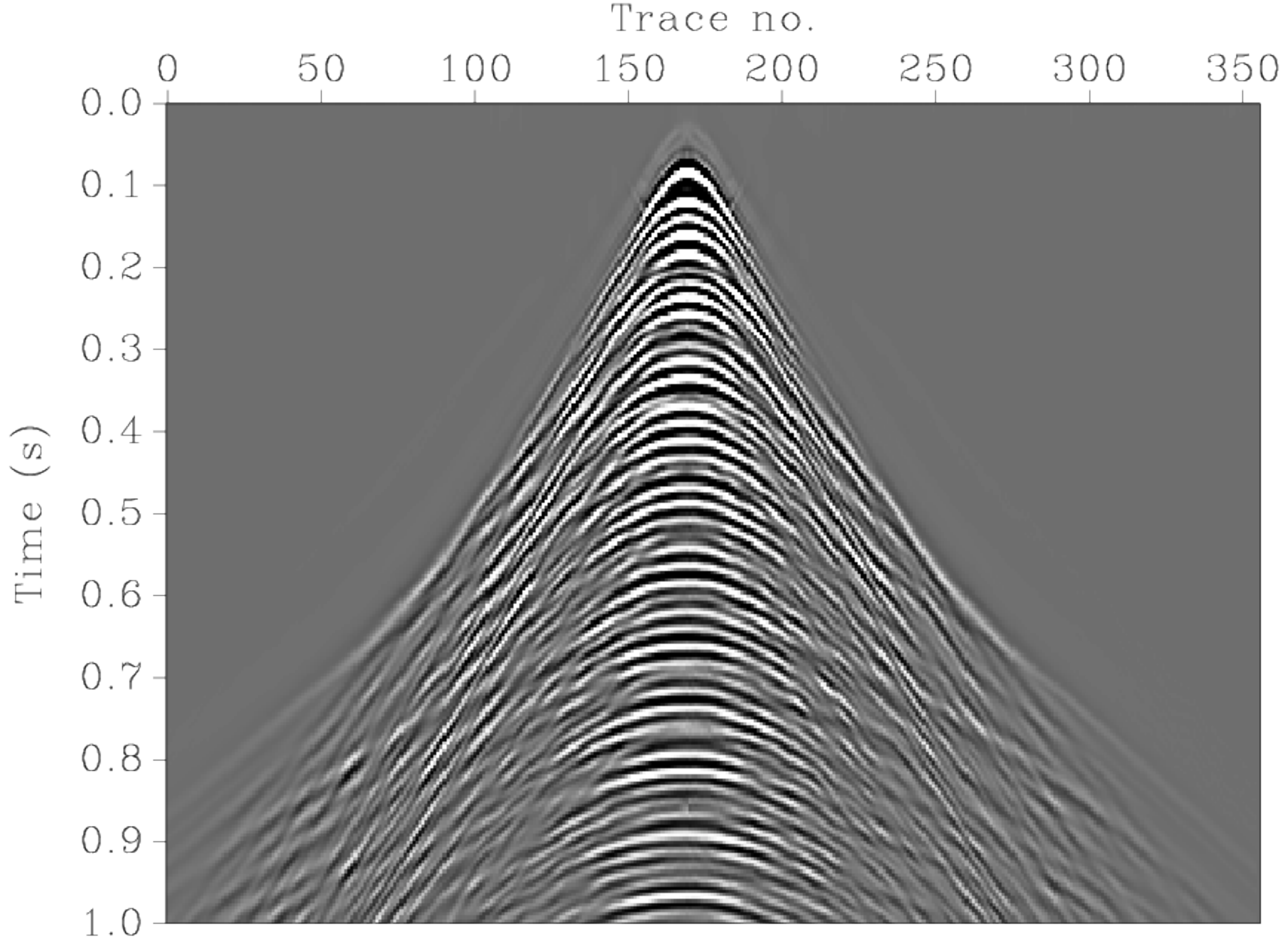
$$B_2 = (1 - \frac{k}{l})\sqrt{l}$$

$$C_1 \text{ and } C_2 > 0$$

Regularization

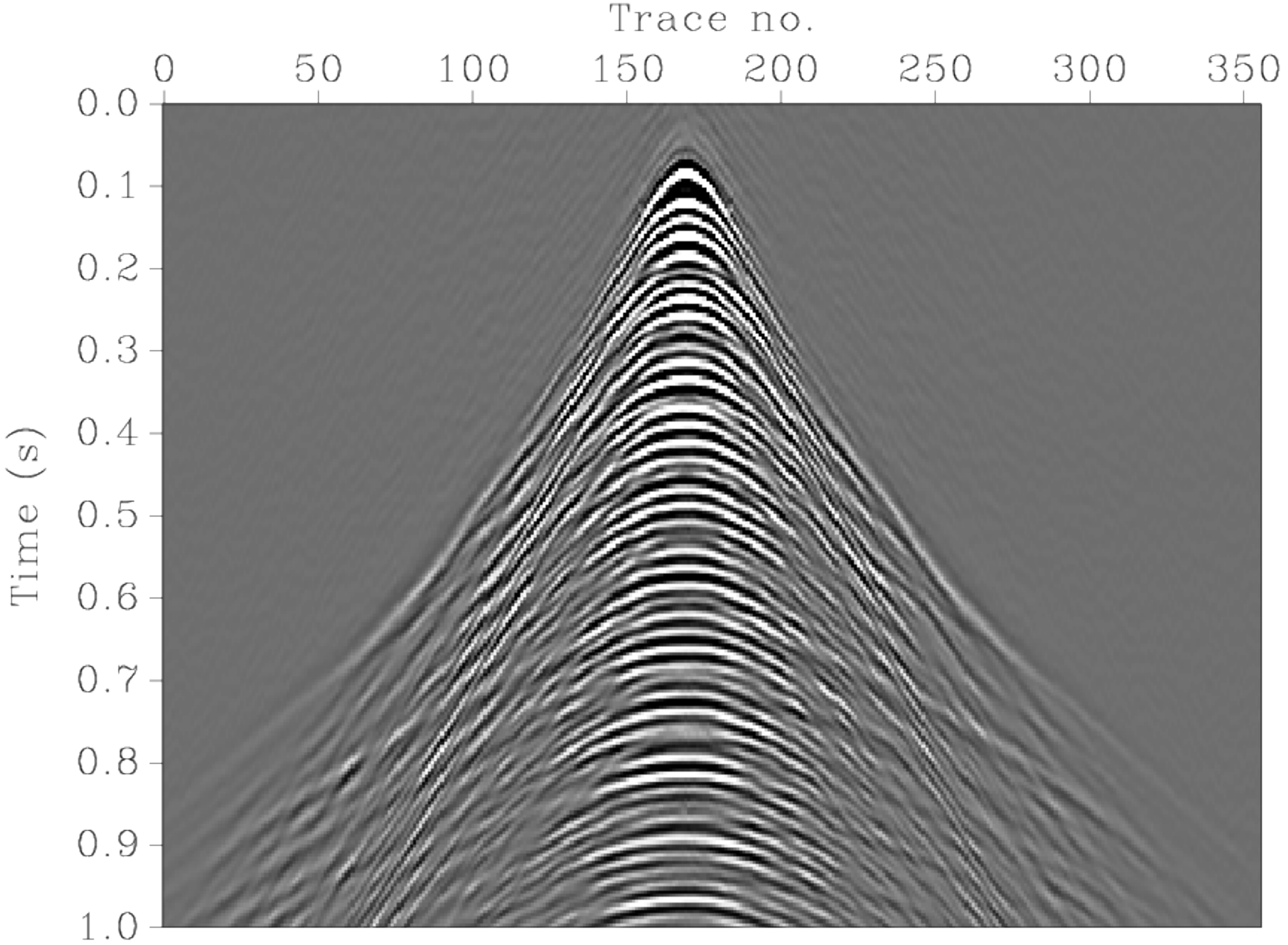
matrix completion

Ground truth



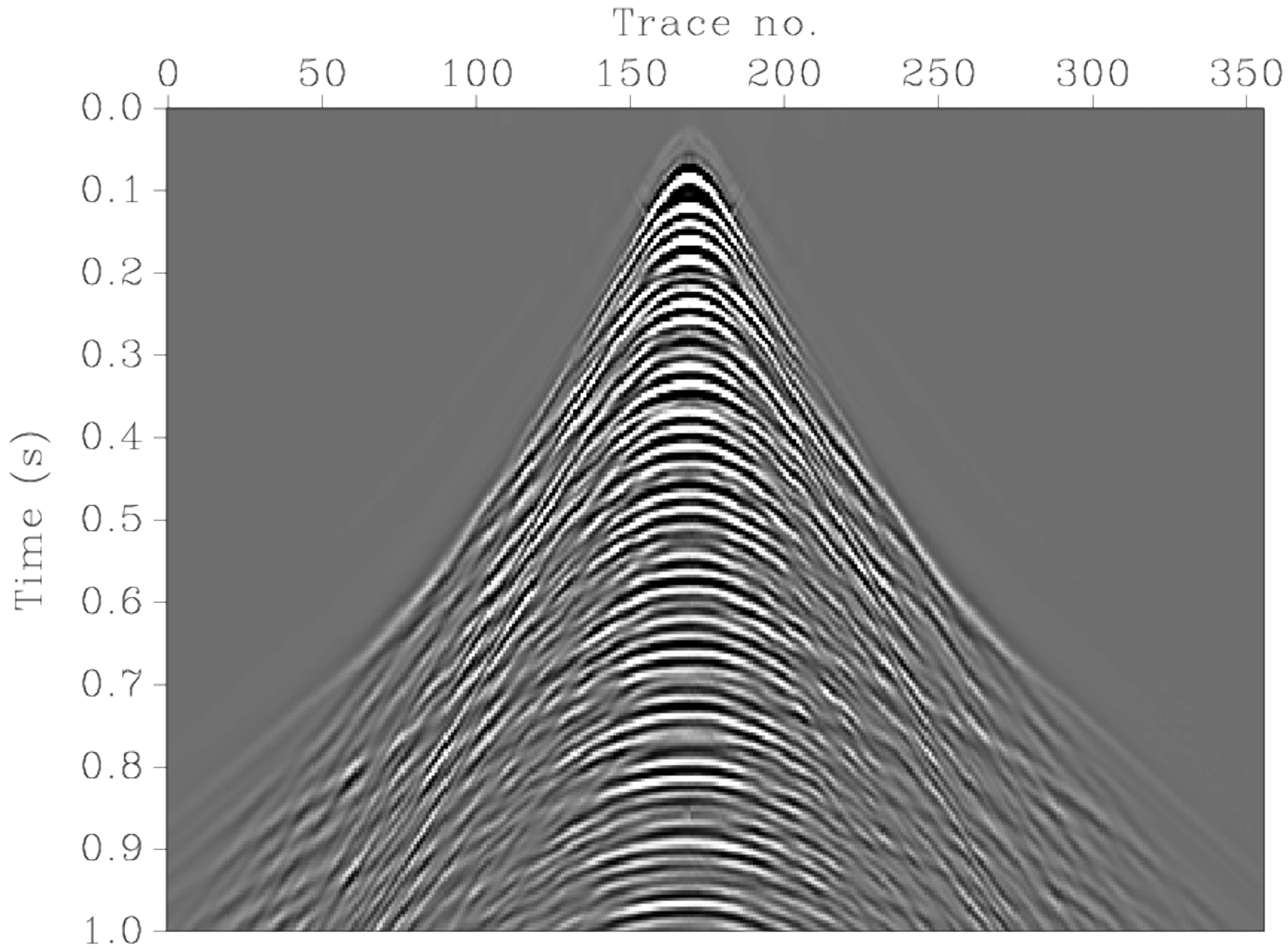
Recovery

SNR = 18.6 dB

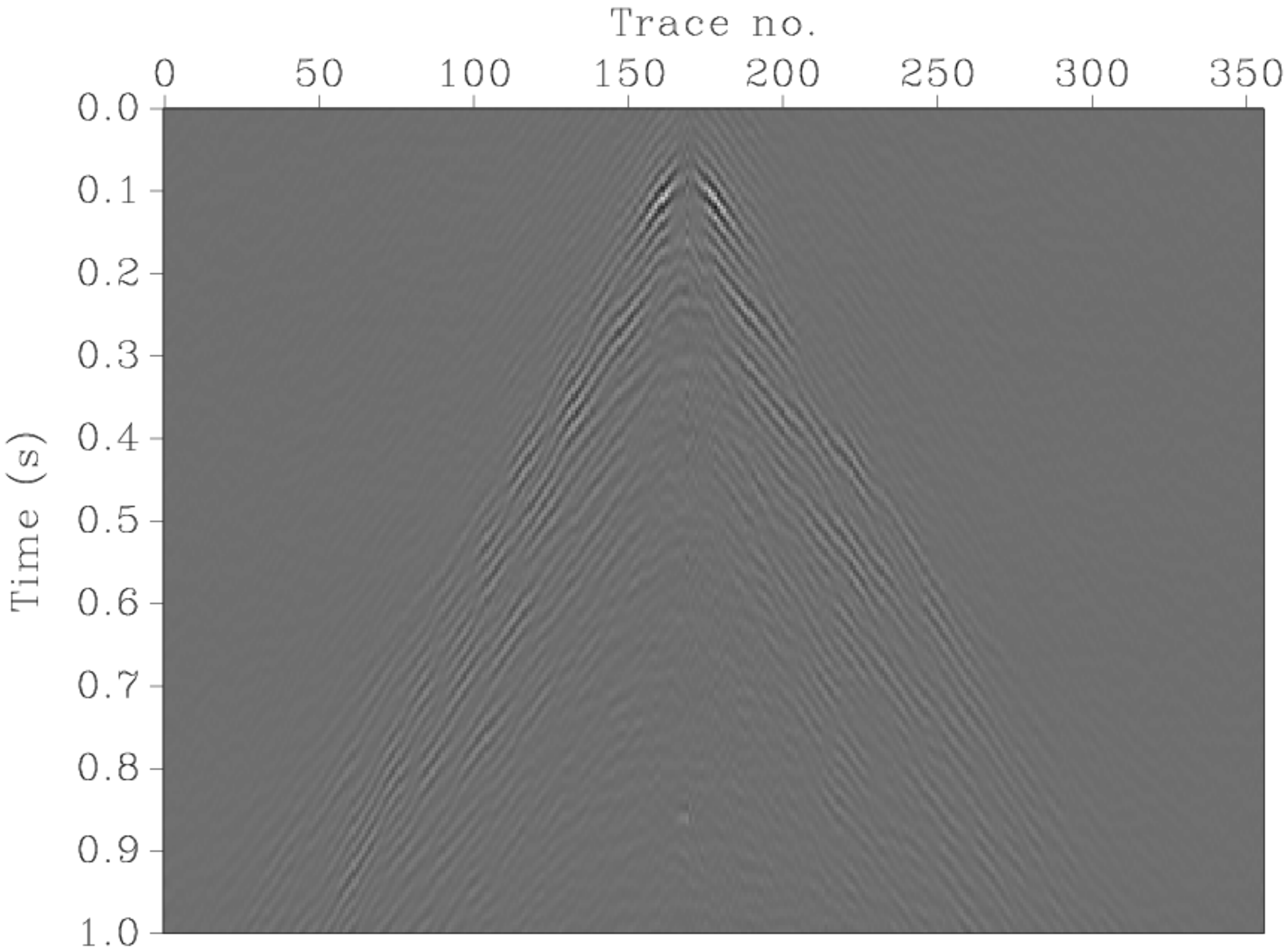


Regularization
matrix completion

Ground truth

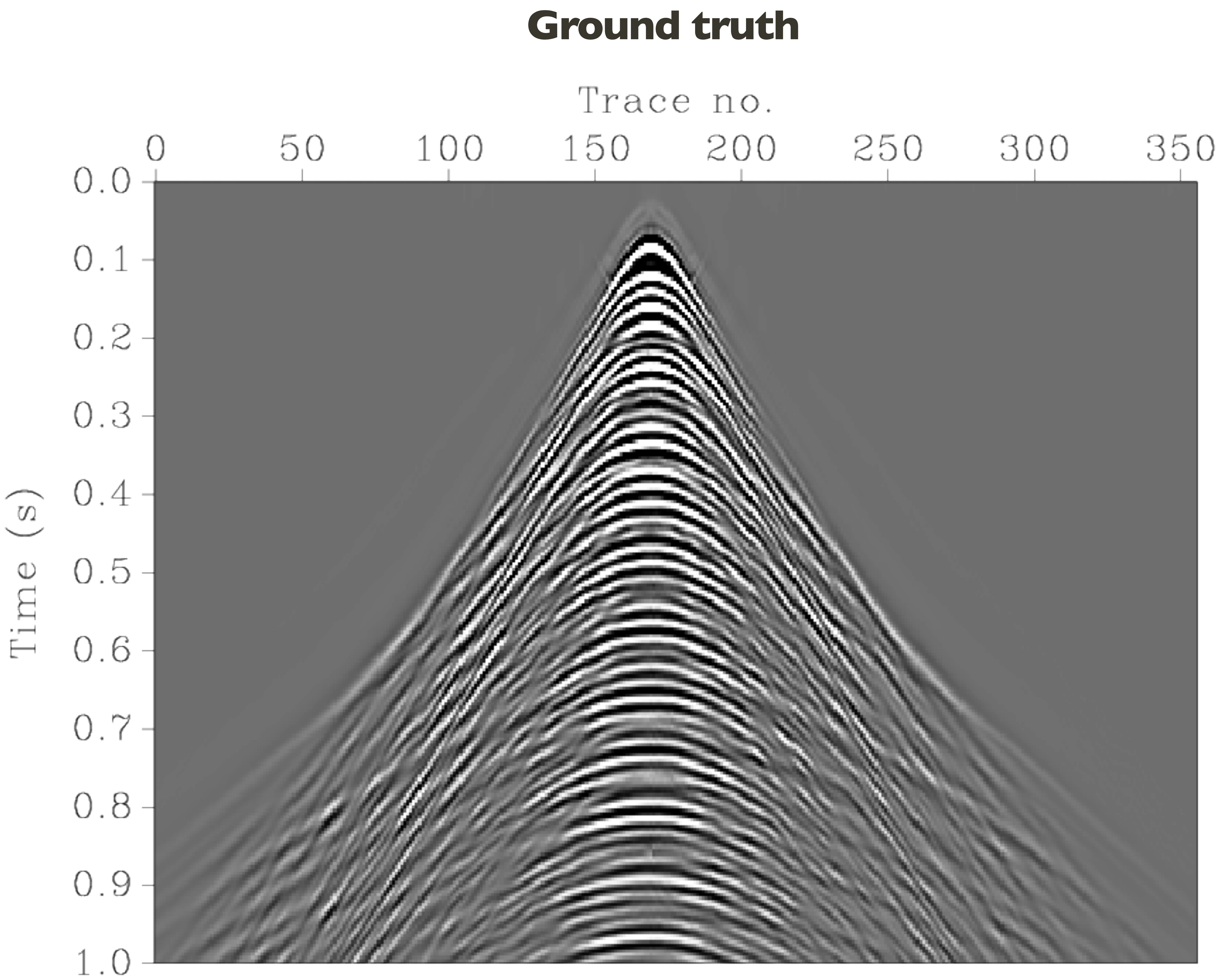


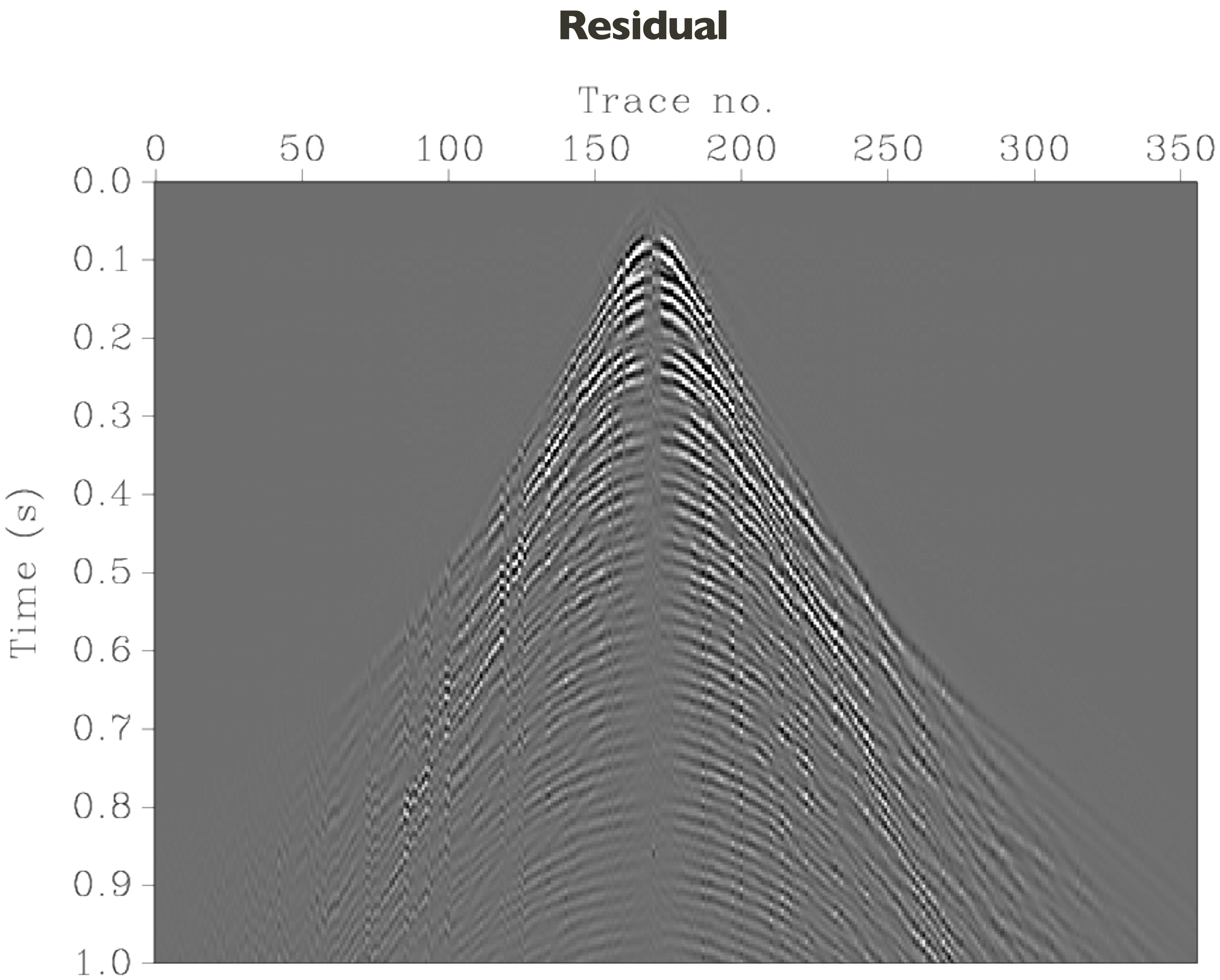
Residual



Regularization

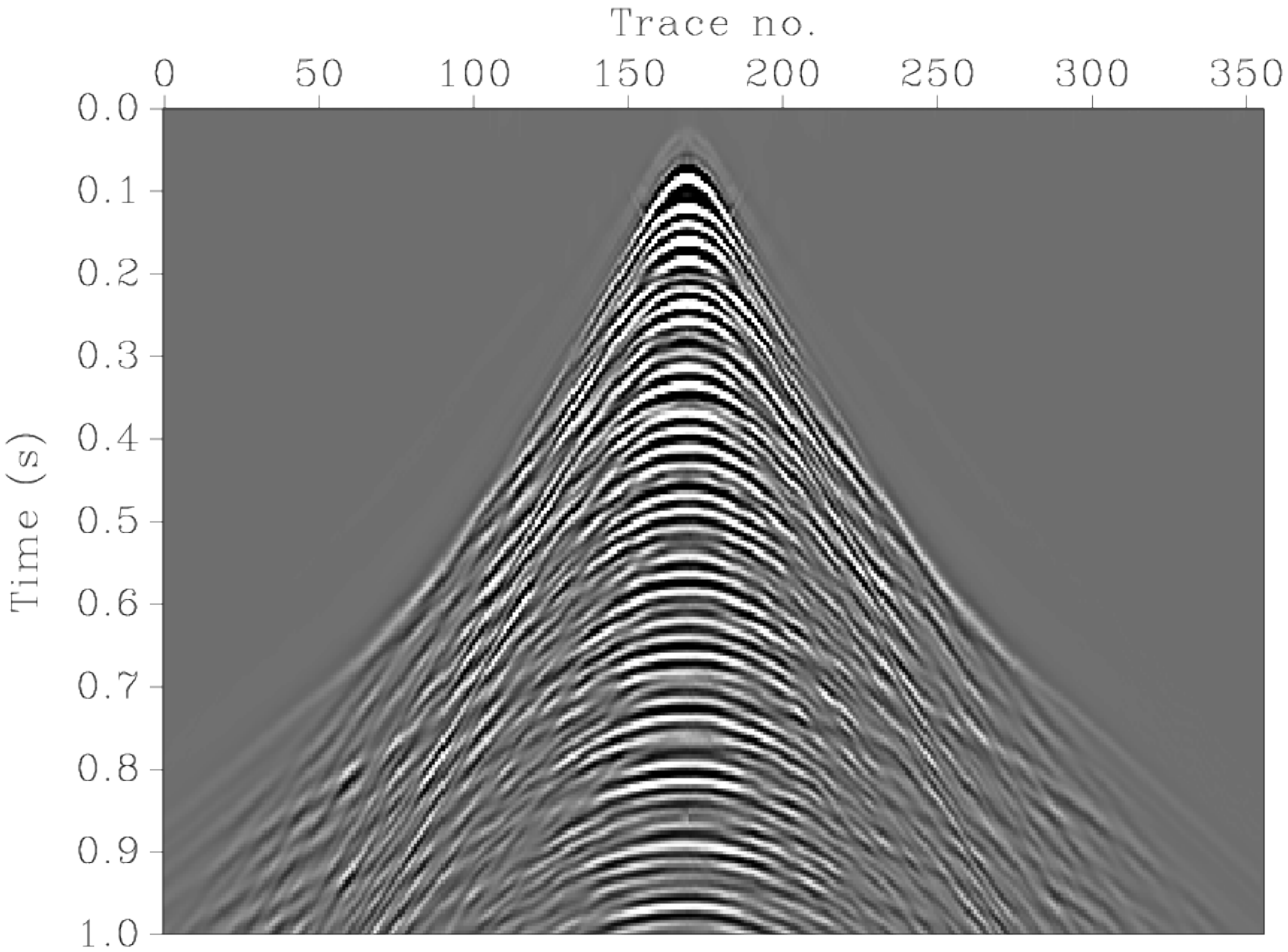
binning



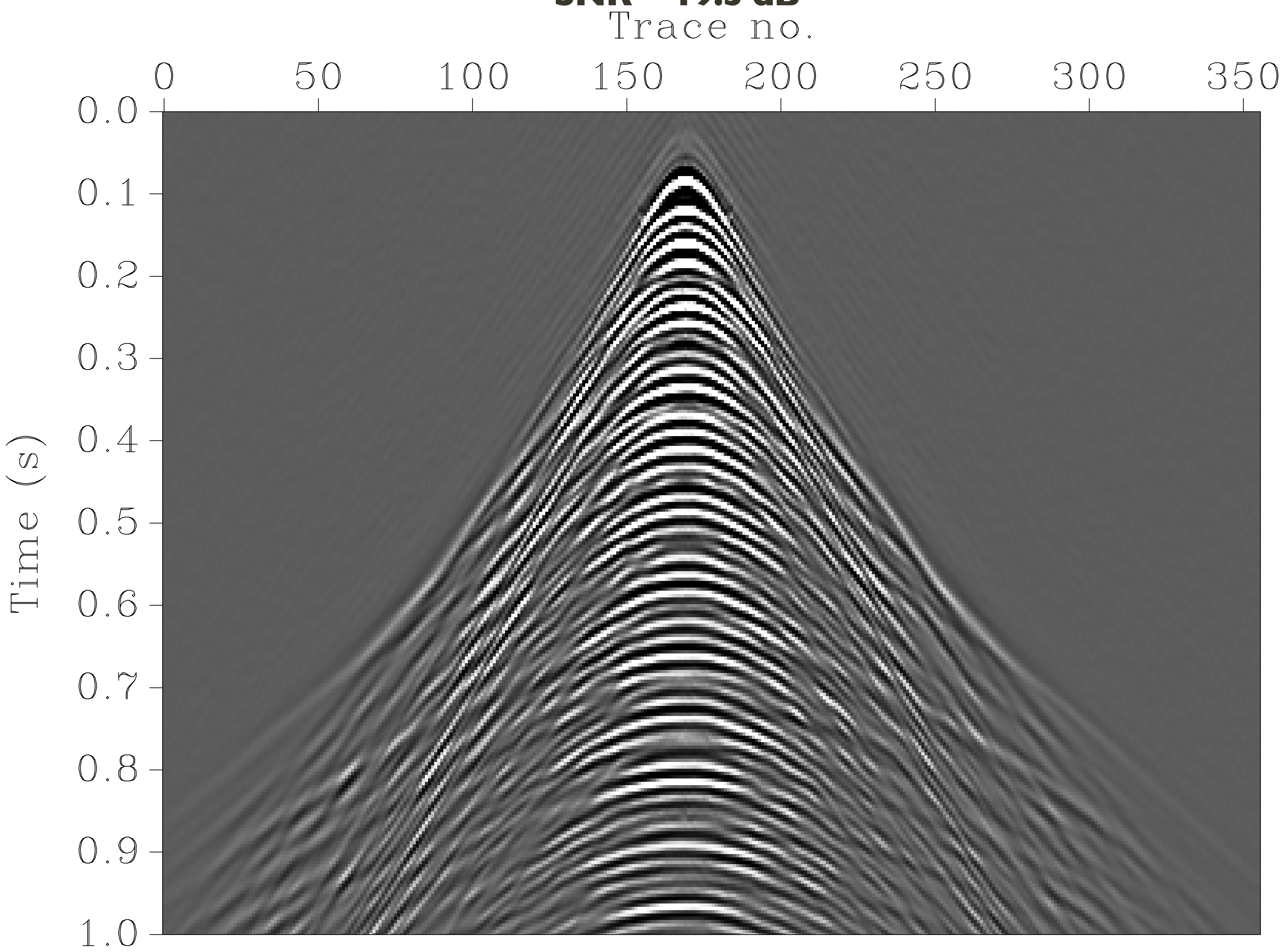


Regularization & Interpolation
matrix completion

Ground truth



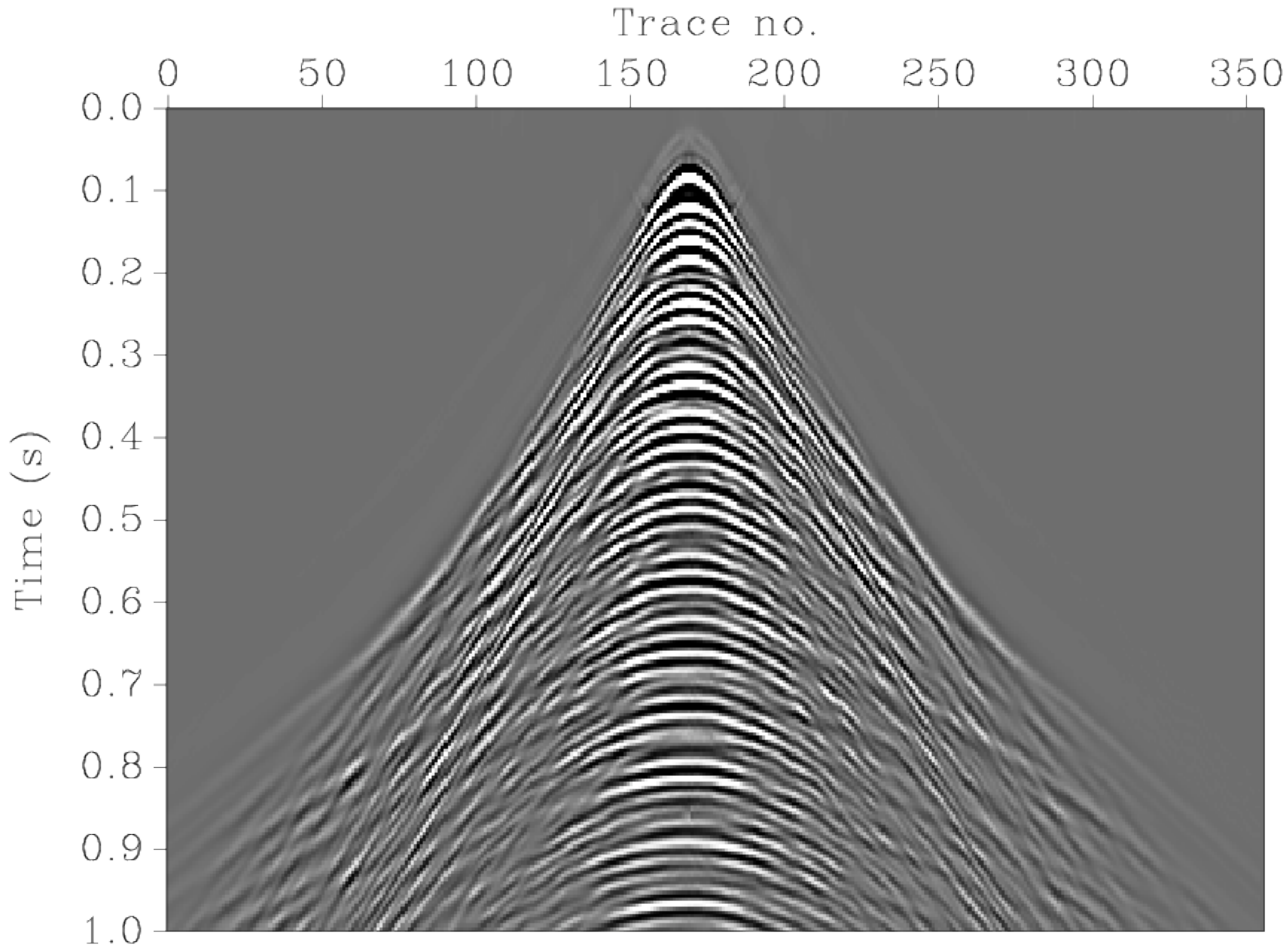
Recovery
SNR = 19.3 dB



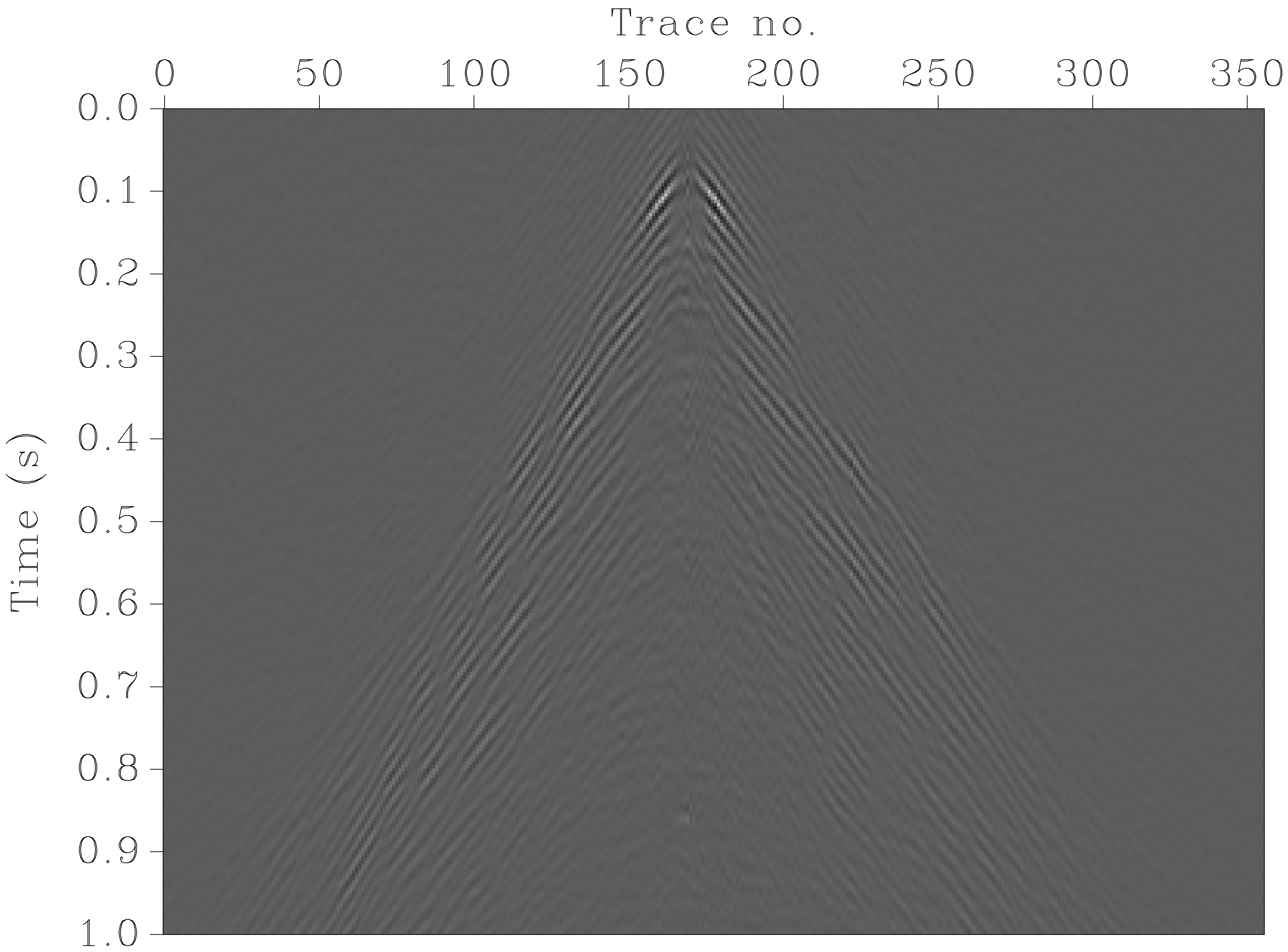
Regularization & Interpolation

matrix completion

Ground truth



Residual



Conclusion

- matrix factorization allows SVD-free low-rank methods that work fast on large data
- reconstruction quality is as good as curvelet-based techniques but computationally more feasible then curvelet
- matrix-factorization promise more compact representation
- able to handle data at unstructured grids

Future work

- incorporate irregularity along both sources & receivers coordinates
- extension to 5D seismic data volumes
- testing of matrix-factorization based methods on real-data
- comparison with tensor-based interpolation methods

SLIM

Acknowledgements

We need Real data set

Thank you for your attention !

https://www.slim.eos.ubc.ca/

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, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, WesternGeco, and Woodside.