Dimensionality-reduced estimation of primaries by sparse inversion

Bander Jumah & Felix J. Herrmann

SLIM
University of British Columbia

Motivation

- Data-driven methods

  Estimation of Primaries by Sparse Inversion (EPSI)

- Curse of dimensionality

  In 3D these methods *suffer* from *exponential* growth in computational & storage *demands*

[Herrmann, F. J., 2010, Randomized sampling and sparsity]

SLIM

Objective

Reduction in *computational* and *storage* demands:

- *dimensionality*-reduction technique
- *adaptive* low-rank approximation
- *blackbox* model

Outline

- Estimation of Primaries by Sparse Inversion (EPSI)
- Dimensionality reduction via *Randomized* Singular Value Decompositions (r-SVD's)
- Results
- Conclusions

EPSI problem

recorded data   predicted data

$$\hat{\mathbf{P}} = \hat{\mathbf{G}}(\hat{\mathbf{Q}} + \hat{\mathbf{R}}\hat{\mathbf{P}})$$

$\hat{\mathbf{P}}$ total up-going wave-field

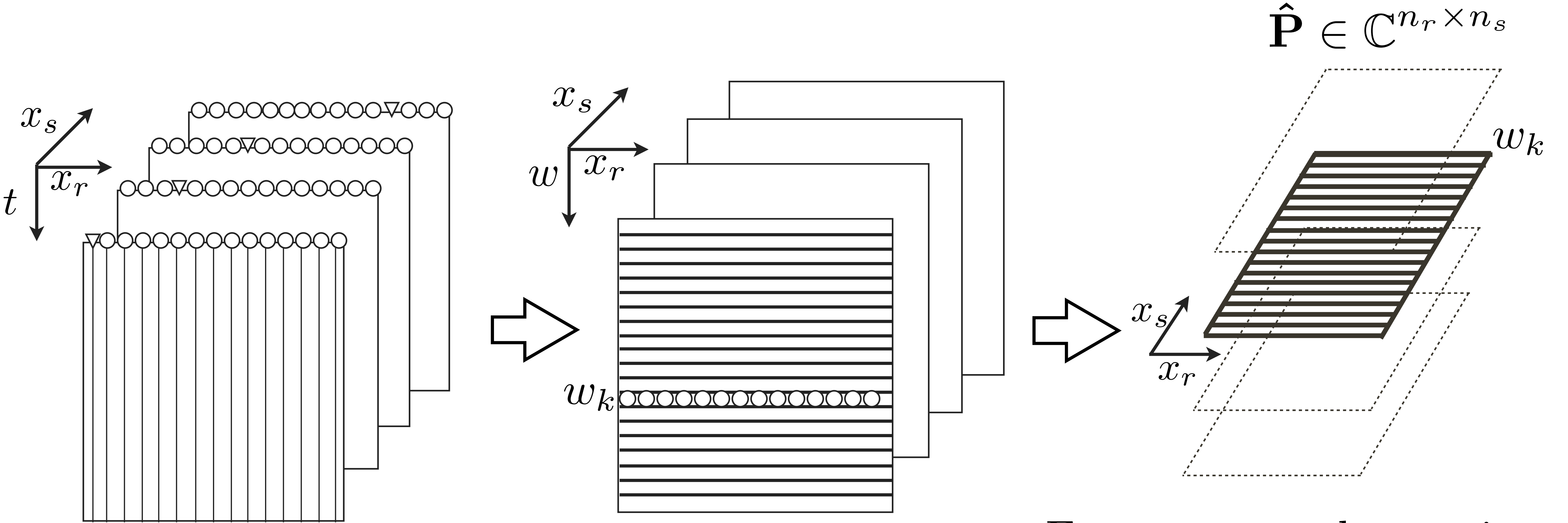
$\hat{\mathbf{Q}}$ down-going source signature

$\hat{\mathbf{R}}$ reflectivity of free surface

$\hat{\mathbf{G}}$ surface-free Green's function

[van Groenestijn and Verschuur, 2009]

Monochromatic “data matrices”



Fourier transform

Extract monochromatic shot records

[Ewoud van Dedem, 2002]

SLIM

EPSI problem

recorded data

predicted data

$$\hat{\mathbf{P}} = \hat{\mathbf{G}}(\hat{\mathbf{Q}} - \hat{\mathbf{P}})$$

$\hat{\mathbf{P}}$ “low-rank” matrix (known)

$\hat{\mathbf{Q}}$ full-rank diagonal matrix (known)

$\hat{\mathbf{R}}$ assume $-\mathbf{I}$

$\hat{\mathbf{G}}$ unknown

EPSI problem

EPSI *linear* algebra format:

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}$$

$$\tilde{\mathbf{x}} = \underbrace{\arg\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathbf{1}}}_{\text{sparsity promoting part}} \quad \text{subject to} \quad \underbrace{||\mathbf{A}\mathbf{x} - \mathbf{b}||_{\mathbf{2}} \leq \sigma}_{\text{data fitting part}}$$

EPSI problem

EPSI *linear* algebra format:

$$\underbrace{\mathbf{F}_t^* \begin{bmatrix} \left((\widehat{\mathbf{Q}} - \widehat{\mathbf{P}})_1^* \otimes \mathbf{I}\right) & & \\ & \ddots & \\ & & \left((\widehat{\mathbf{Q}} - \widehat{\mathbf{P}})_{n_f}^* \otimes \mathbf{I}\right) \end{bmatrix} \mathbf{F}_t}_{\widehat{\mathbf{U}}} \begin{bmatrix} \mathrm{vec}\left(\mathbf{G}_1\right) \\ \vdots \\ \mathrm{vec}\left(\mathbf{G}_{n_t}\right) \end{bmatrix} \approx \begin{bmatrix} \mathrm{vec}\left(\mathbf{P}_1\right) \\ \vdots \\ \mathrm{vec}\left(\mathbf{P}_{n_t}\right) \end{bmatrix}$$

Combination with sparsity promotion:

$\mathbf{A} = \widehat{\mathbf{U}}\mathbf{C}^* \quad \mathbf{C}$ is curvelet transform

$\mathbf{x}$: discrete curvelet representation of $\mathbf{G}$

$\mathbf{b}$: discrete representation of $\mathbf{P}$

EPSI problem

Data matrix $\hat{\mathbf{P}}$

- dense
- low-rank
- extremely large
  - each frequency is a $10^6 \times 10^6$ matrix where $n_r = n_s = 1000$
- expensive to access & store
- high mat-mat multiplication cost $O(N^3)$

EPSI problem

Challenges in solving the optimization problem

- multiple iterations
- multiple evaluations of $\mathbf{A}$, $\mathbf{A}^*$ and $\mathbf{A}^*\mathbf{A}$

Dimensionality-reduction via SVD

Approximate data matrix $\hat{\mathbf{P}}$ with *low-rank* factorization:

$$\hat{\mathbf{P}} = \hat{\mathbf{G}}(\hat{\mathbf{Q}} - \hat{\mathbf{P}})$$

$$\hat{\mathbf{P}} \approx \mathbf{U\Sigma V}^*$$

$\mathbf{U_{n_r \times k}}$ left singular vectors

$\mathbf{\Sigma_{k \times k}}$ singular values

$\mathbf{V_{n_s \times k}}$ right singular vectors

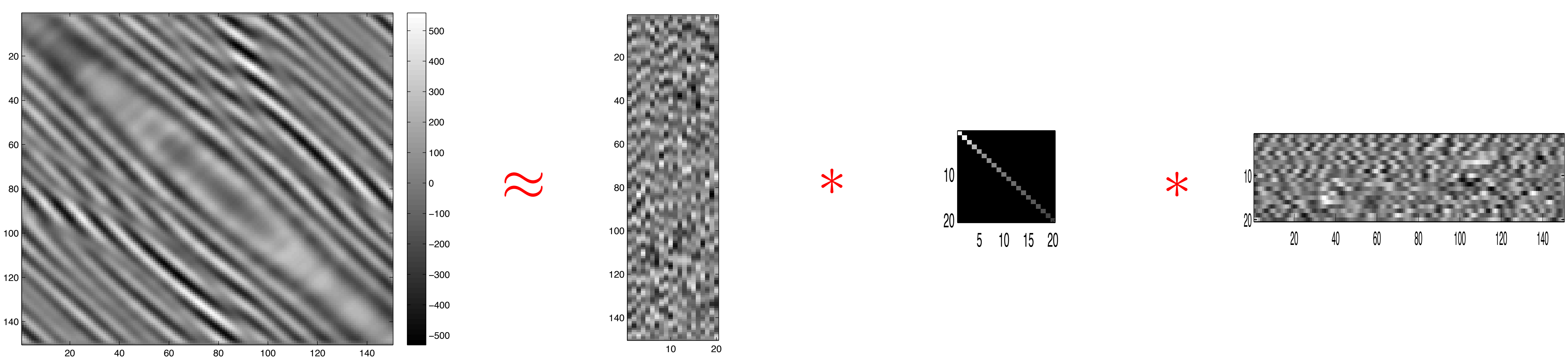
$\mathrm{k} :$ approximate rank

$k << min(n_r, n_s)$

Dimensionality Reduction Via SVD

Approximate data matrix $\hat{\mathbf{P}}$ with *low-rank* factorization:

$$\hat{\mathbf{P}} \approx \mathbf{U} * \mathbf{\Sigma} * \mathbf{V}^*$$



$n_r \times n_s$ $\quad$ $n_r \times k$ $\quad$ $k \times k$ $\quad$ $k \times n_s$

k : approximate rank

$k << min(n_r, n_s)$

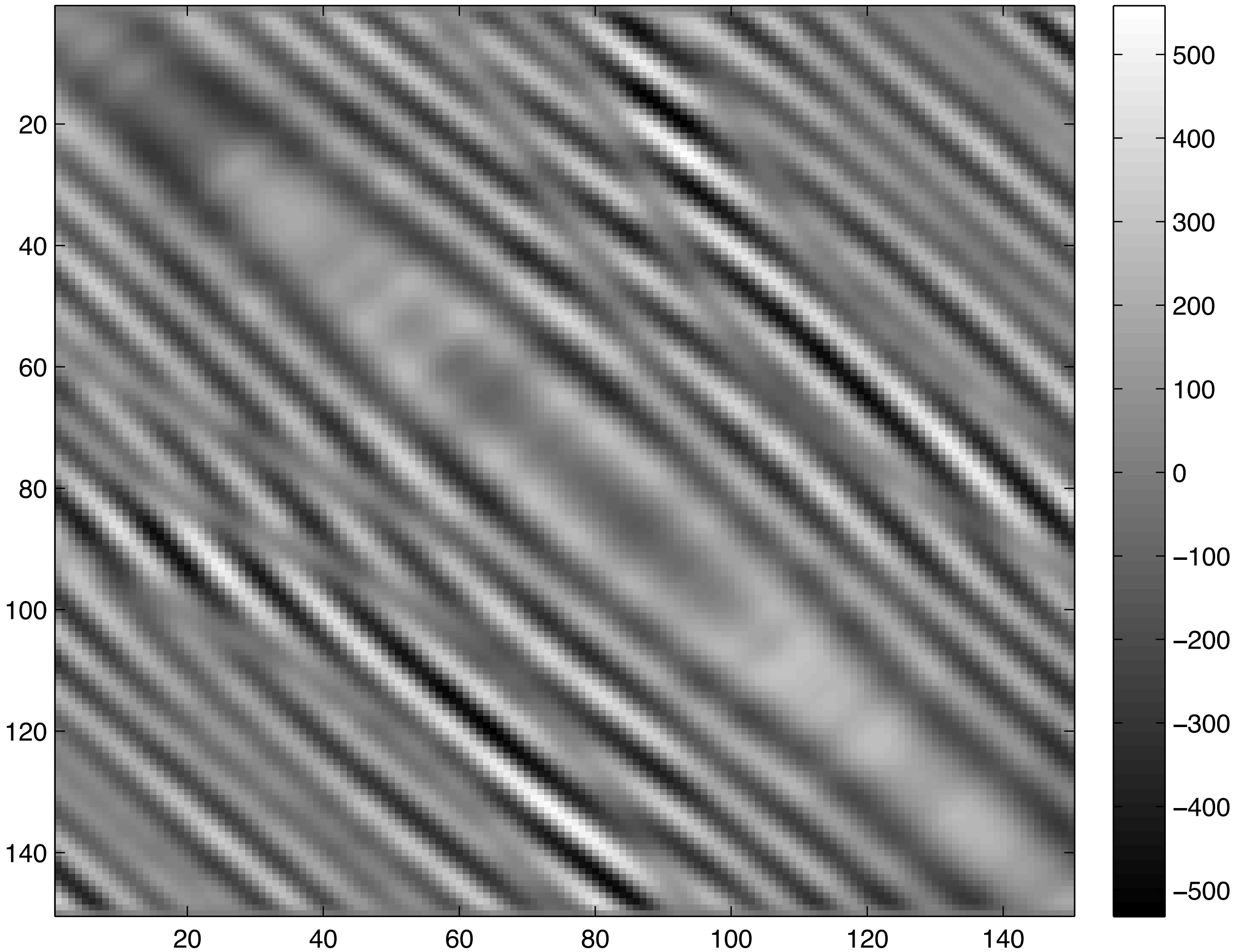
Dimensionality Reduction Via SVD

Advantages of using low rank factorization

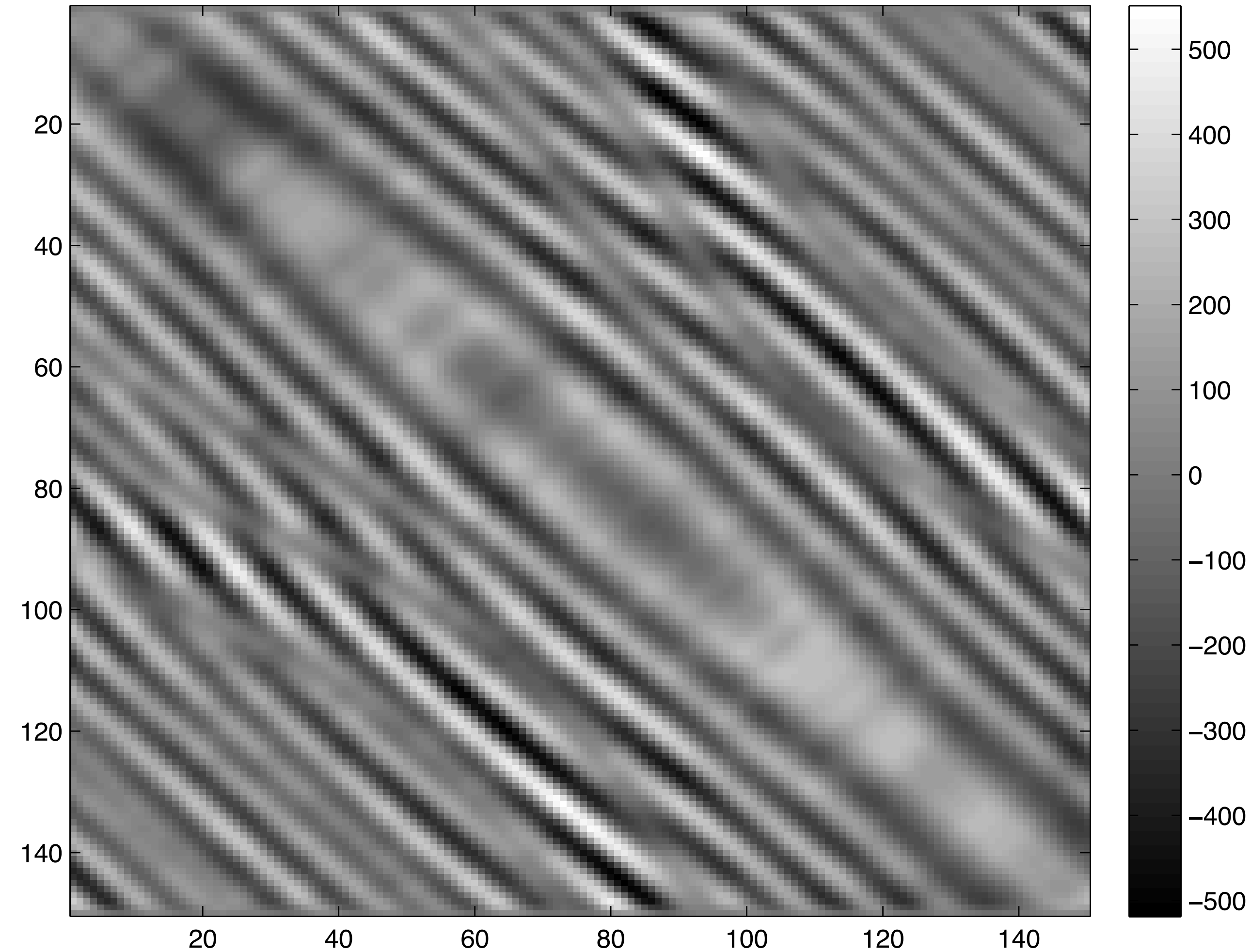
	Regular method	Low-rank approximation
Matrix-Matrix multiplication	$O(N^3)$	$O(kN^2)$
Storage (bytes)	$O(N^2)$	$O(2Nk + k^2)$

Full vs approximated data

$\hat{\mathbf{P}}$

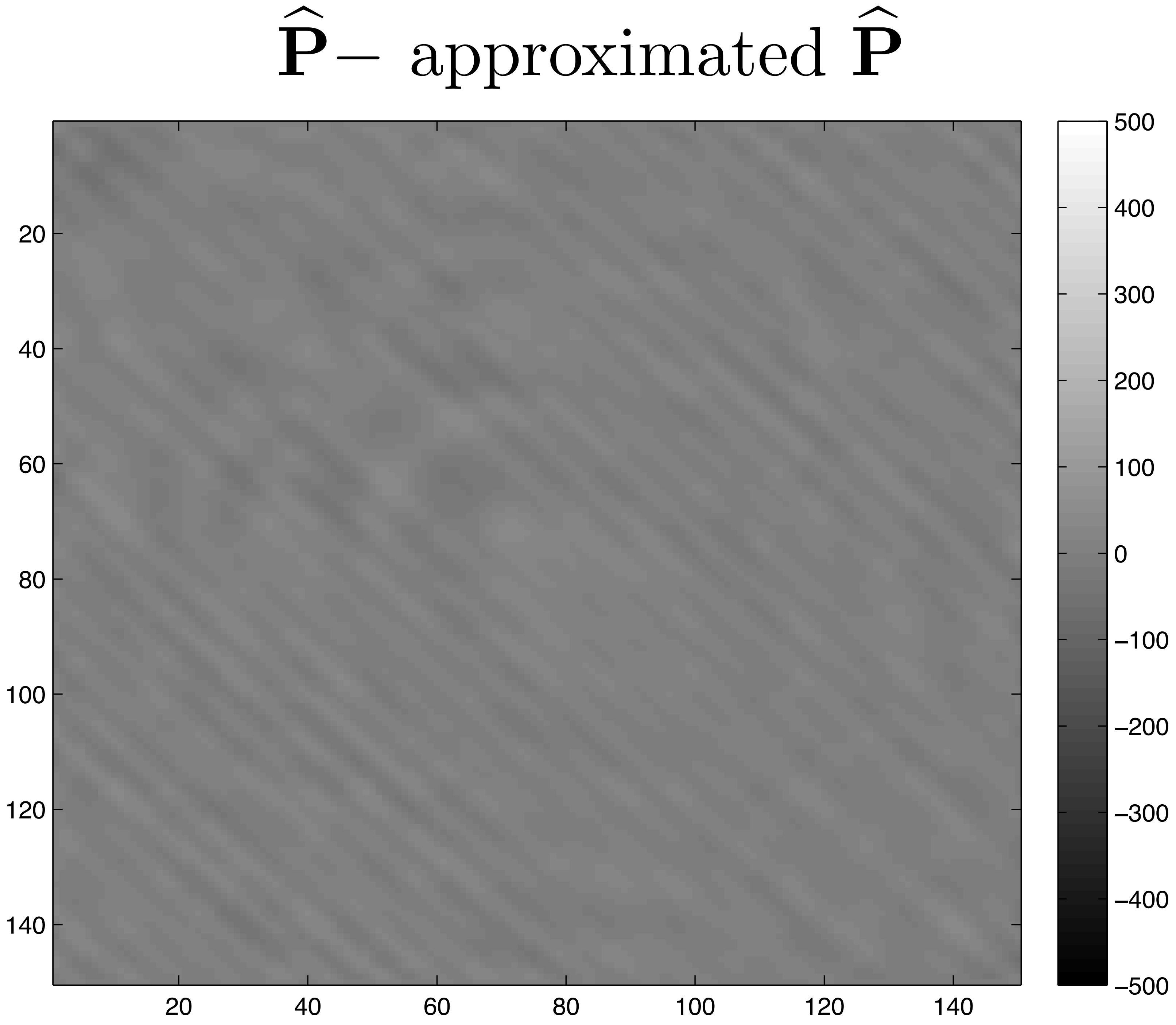


Approximated $\hat{\mathbf{P}}$



$$n_s = n_r = 150$$
$$k = 20 = 14\%$$
$$SNR = 16dB$$

Full vs approximated data



Multiplication speed up

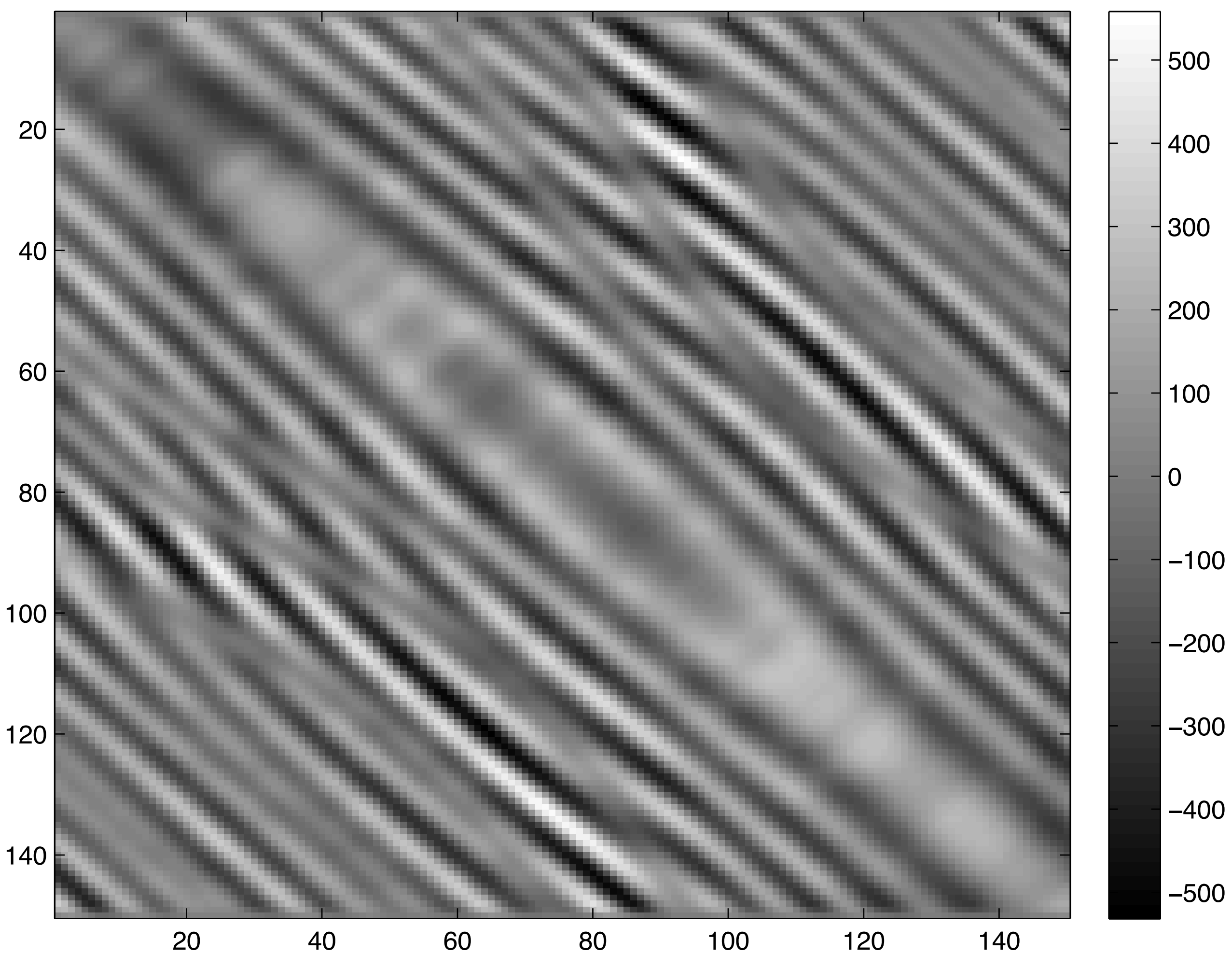
7.5 x

Memory usage

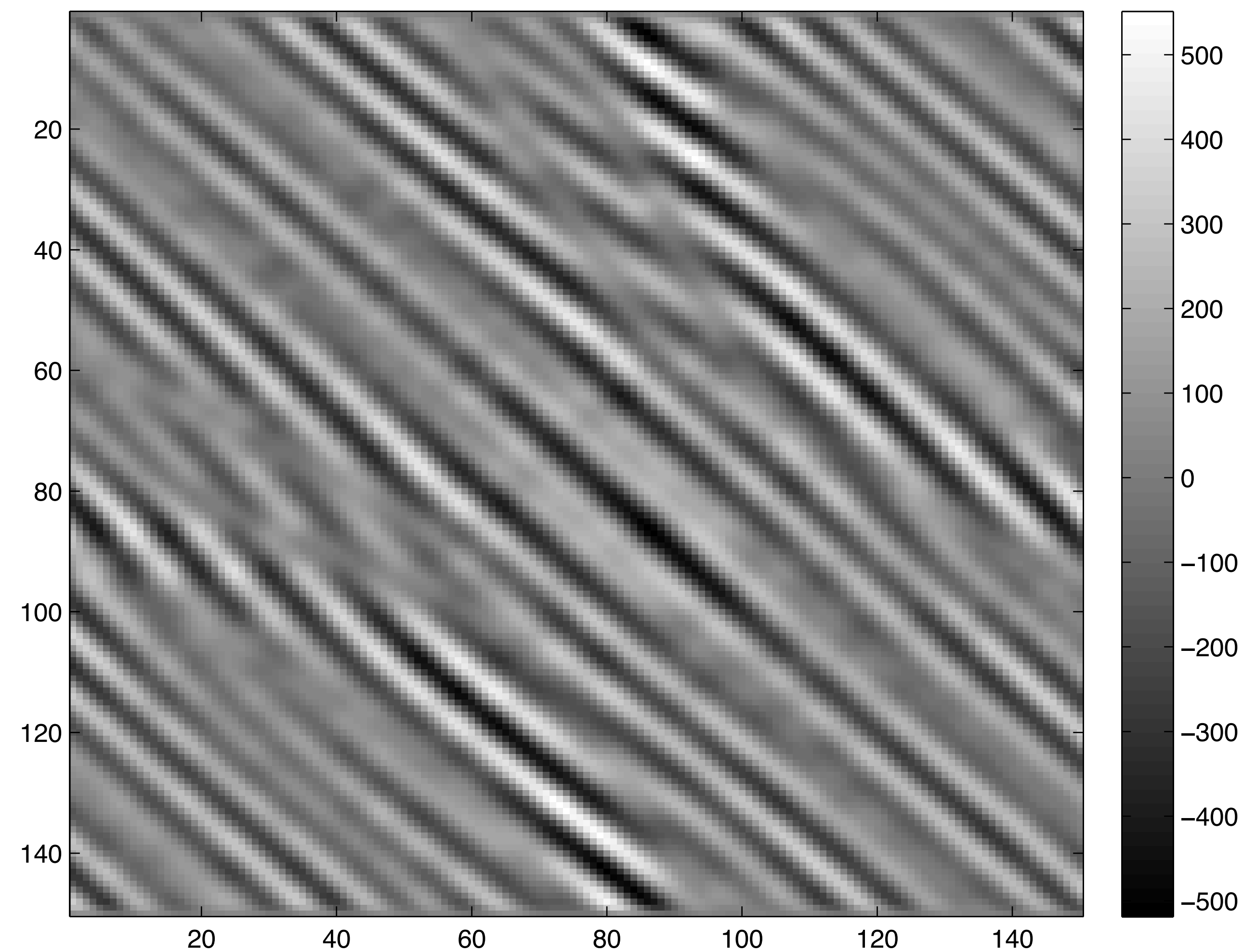
70% less

Full vs approximated data

$\hat{\mathbf{P}}$

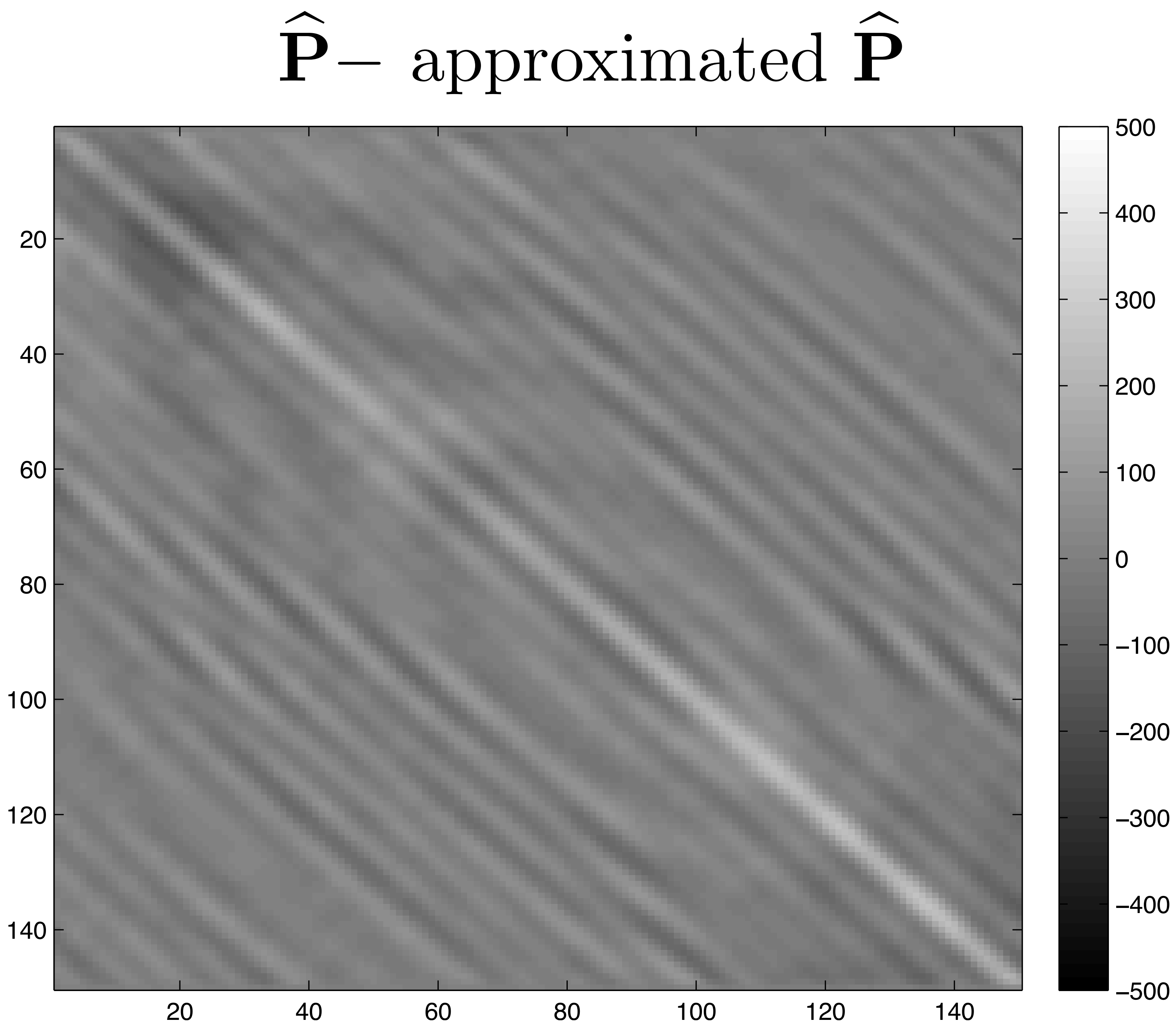


Approximated $\hat{\mathbf{P}}$



$$n_s = n_r = 150$$
$$k = 8 = 5\%$$
$$SNR = 8dB$$

Full vs approximated data



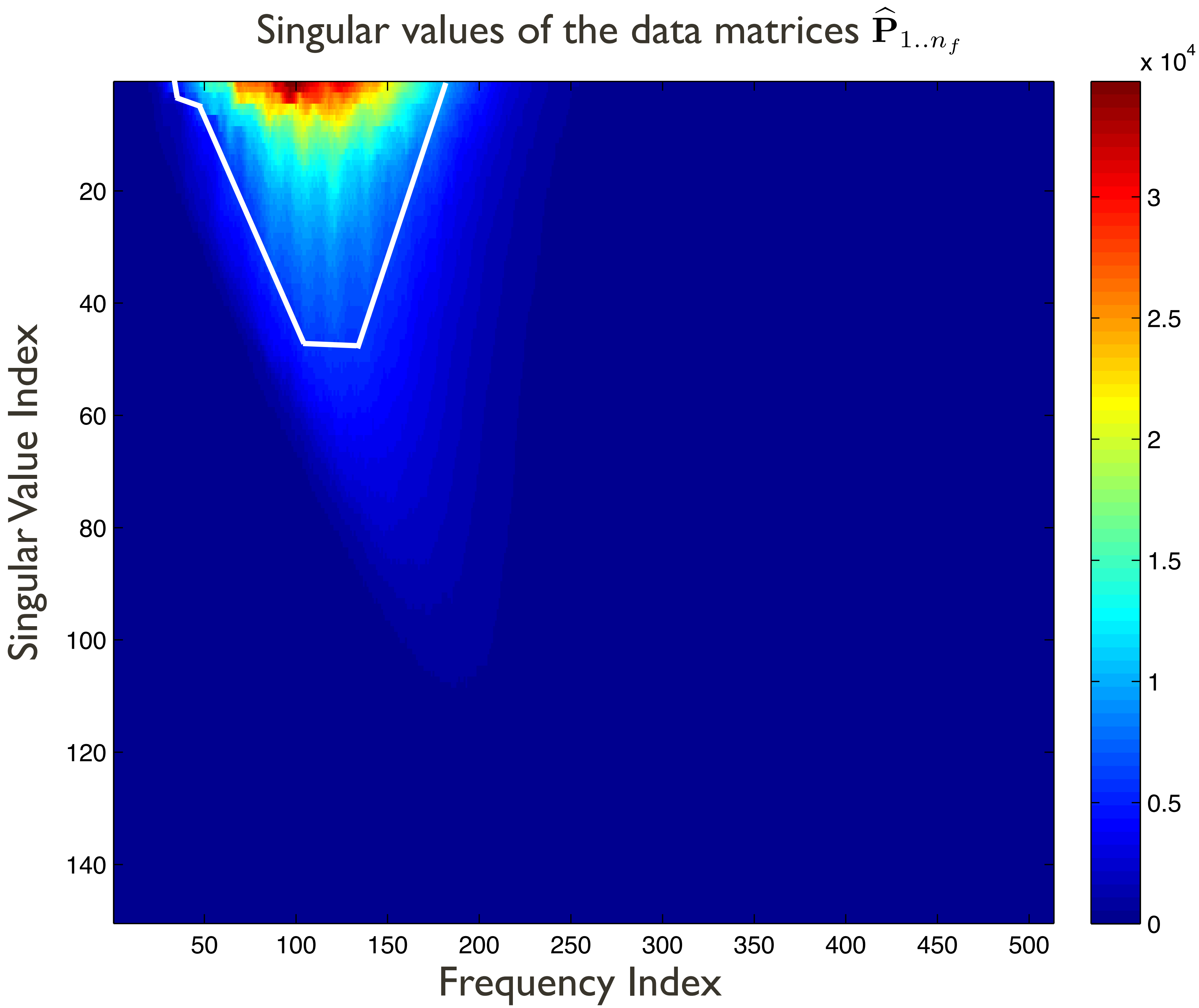
Multiplication speed up

20x

Memory usage

90% less

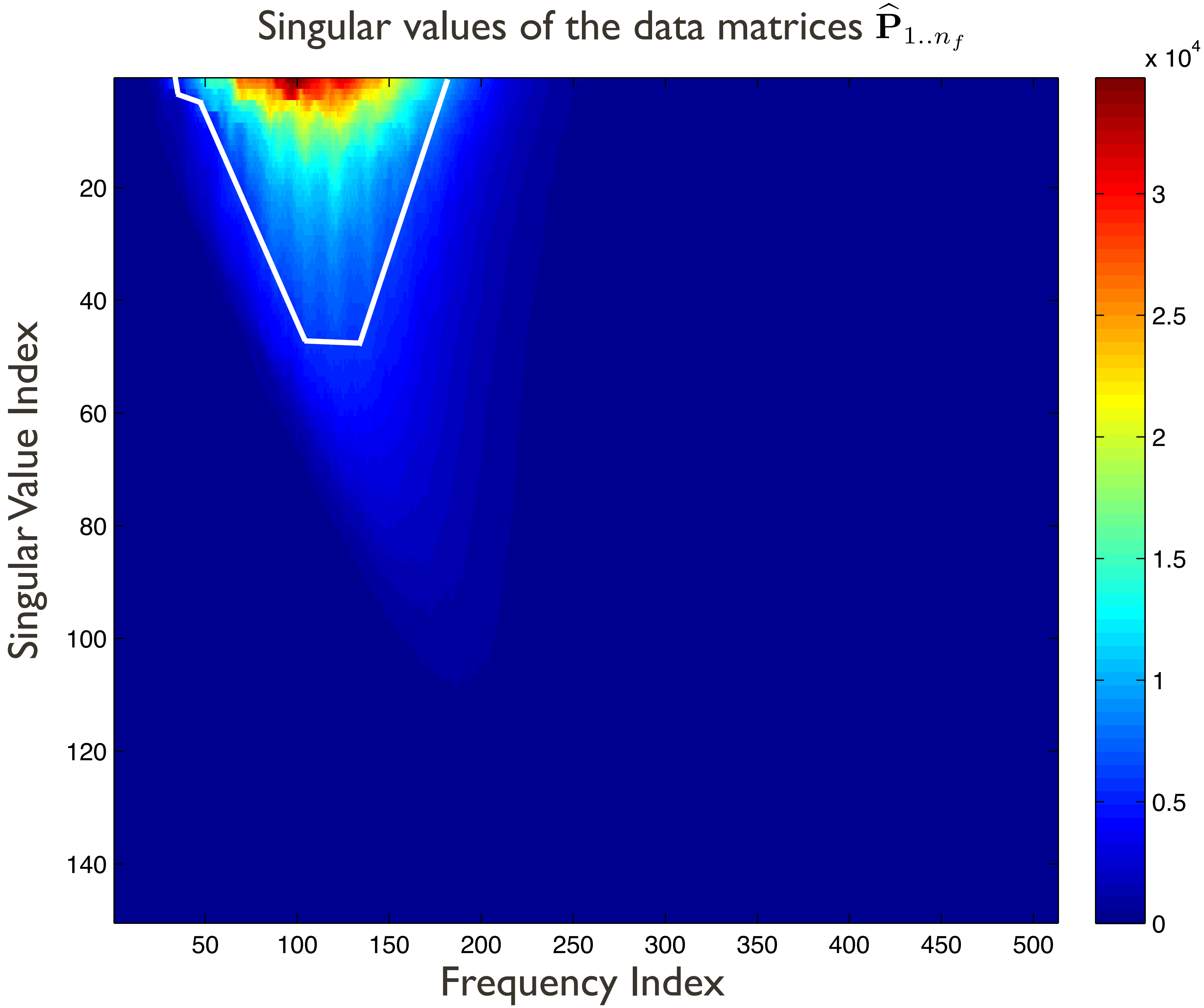
Singular values of the data matrix



Objective:

Approximate *all* frequency slices

Singular values of the data matrix



Challenge:

SVD IS SLOW

SLIM

Singular Value Decomposition (SVD)

Randomized SVD:

Requires action of data matrix on small number of *randomized* vectors (*simultaneous* shots)

Fast $O(mnlog(k))$

Classical SVD:

Slow $O(mnk)$

[Halko, N., P. G. Martinsson, and J. A. Tropp, 2011]

Singular Value Decomposition (SVD)

Randomized vs. Classical SVD

Example

# rows = # columns = 10000
k = 20% = 2000

SVD

$$O((10,000)^2 * 2000)$$

R-SVD

$$O((10,000)^2 * log(2000))$$ $200\times$ Faster !

Dimensionality Reduction Via *RSVD*

Two-stage approach:

1. *capture* action of the data $\hat{\mathbf{P}}$ matrix on $k+l$ *random vectors*

$$\hat{\mathbf{Y}} = \hat{\mathbf{P}}\hat{\mathbf{W}}$$

$\hat{\mathbf{W}}$: Gaussian random matrix

l is a small over sampling parameter (1-8)

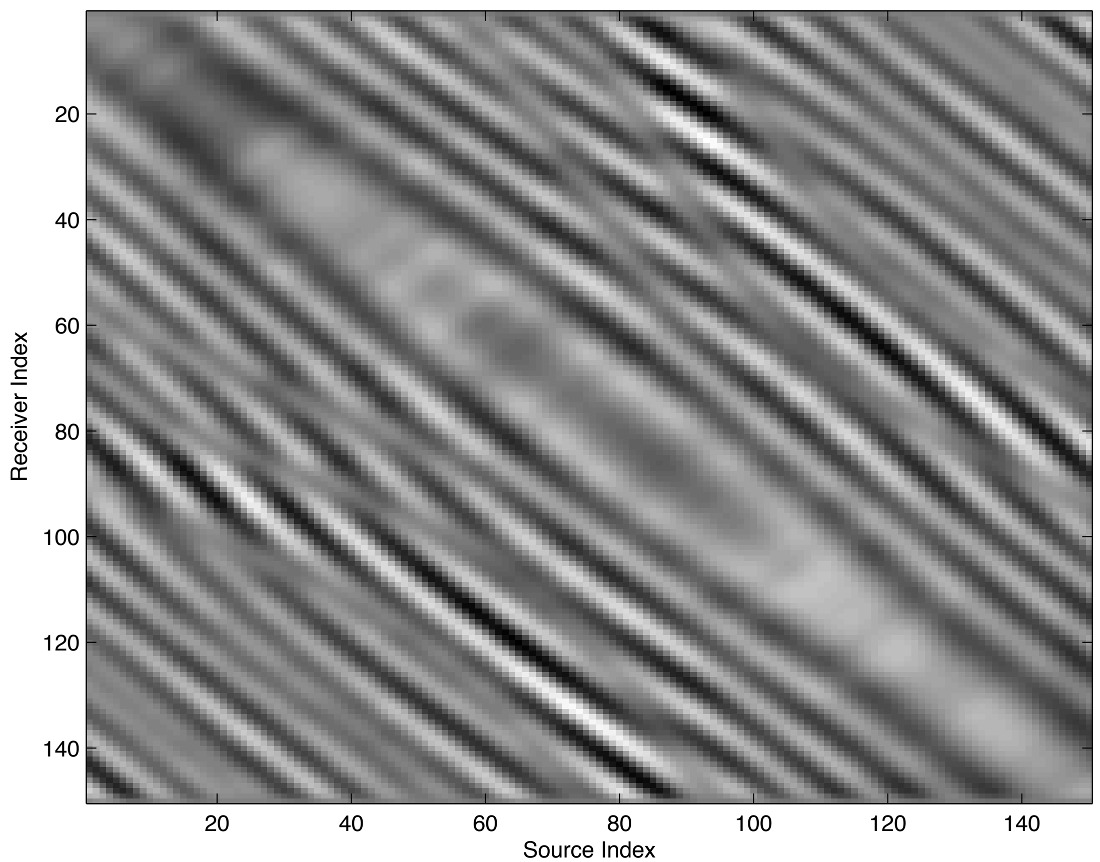
2. *form* a SVD on $\hat{\mathbf{Y}}$

[Halko, N., P. G. Martinsson, and J. A. Tropp, 2011]

Dimensionality Reduction Via *RSVD*

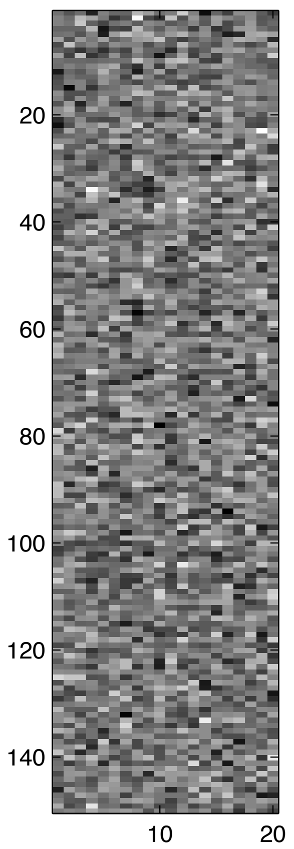
Stage 1: Capturing the action of $\hat{\mathbf{P}}$

$\hat{\mathbf{P}}$



$n_r \times n_s$

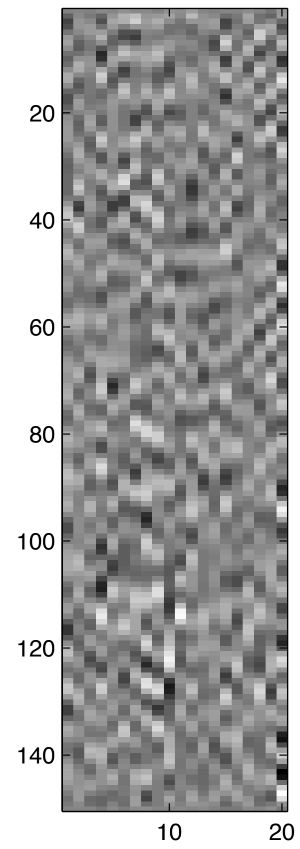
$\hat{\mathbf{W}}$



$n_s \times k$

$=$

$\hat{\mathbf{Y}} = \hat{\mathbf{P}}\hat{\mathbf{W}}$



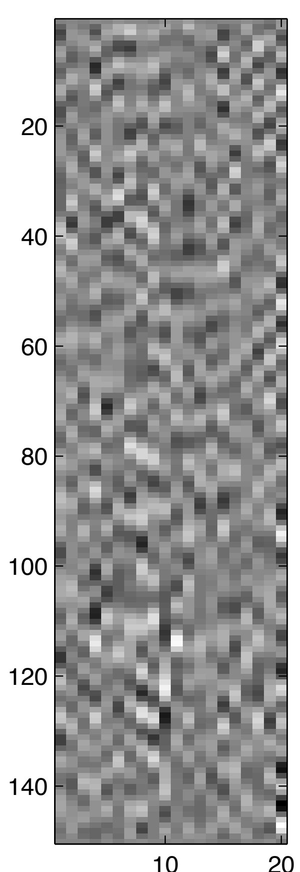
$n_r \times k$

Dimensionality Reduction Via *RSVD*

Stage 1: Capturing the action of $\hat{\mathbf{P}}$

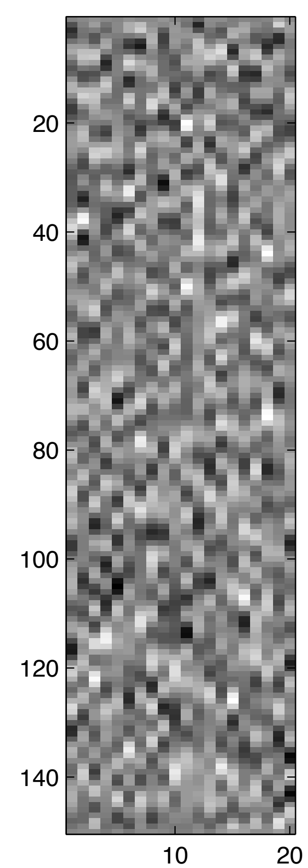
2. Form a low-rank QR factorization $\hat{\mathbf{Y}} \approx \mathbf{QR}$

$\hat{\mathbf{Y}}$ $\quad = \quad$ $\mathbf{Q}$ $\quad$ $\mathbf{R}$

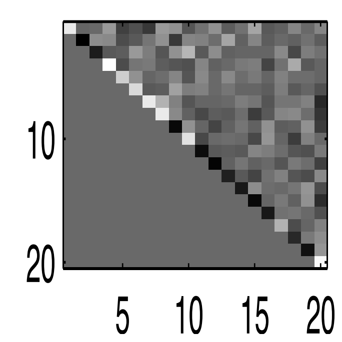


$n_r \times k$

$=$



$n_r \times k$



$k \times k$

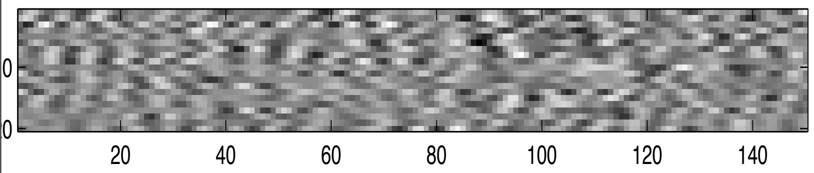
SLIM

Dimensionality Reduction Via *RSVD*

Stage 2 : Compute an approximate SVD of $\hat{\mathbf{P}}$

1. Form $\mathbf{B} = \mathbf{Q}^*\hat{\mathbf{P}}$

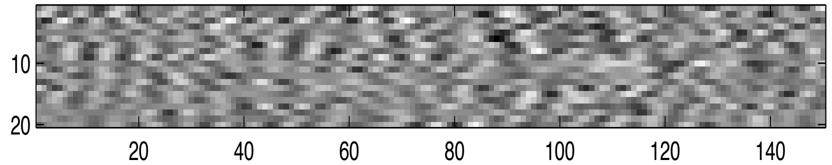
$\mathbf{B}$



$k * n_s$

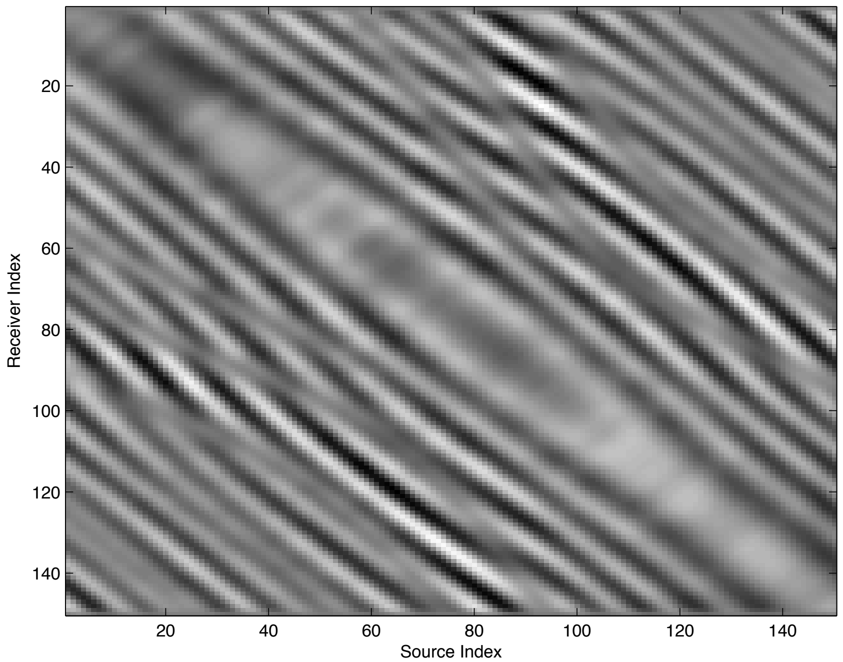
=

$\mathbf{Q}^*$



$k \times n_r$

$\hat{\mathbf{P}}$



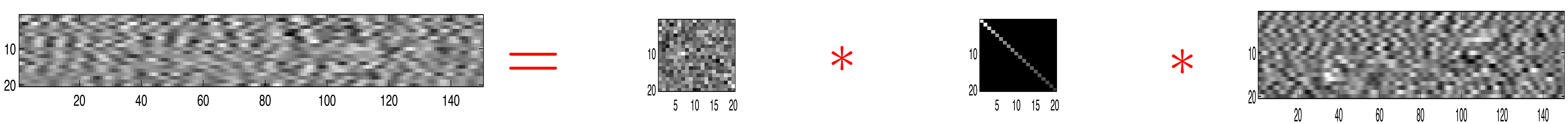
$n_r \times n_s$

Dimensionality Reduction Via *RSVD*

Stage 2 : Compute an approximate SVD of $\hat{\mathbf{P}}$

2. Compute **SVD** of the small matrix $\mathbf{B} = \widetilde{\mathbf{U}}\mathbf{\Sigma}\mathbf{V}^*$

$\mathbf{B}$ $\quad$ $\widetilde{\mathbf{U}}$ $\quad$ $\mathbf{\Sigma}$ $\quad$ $\mathbf{V}^*$

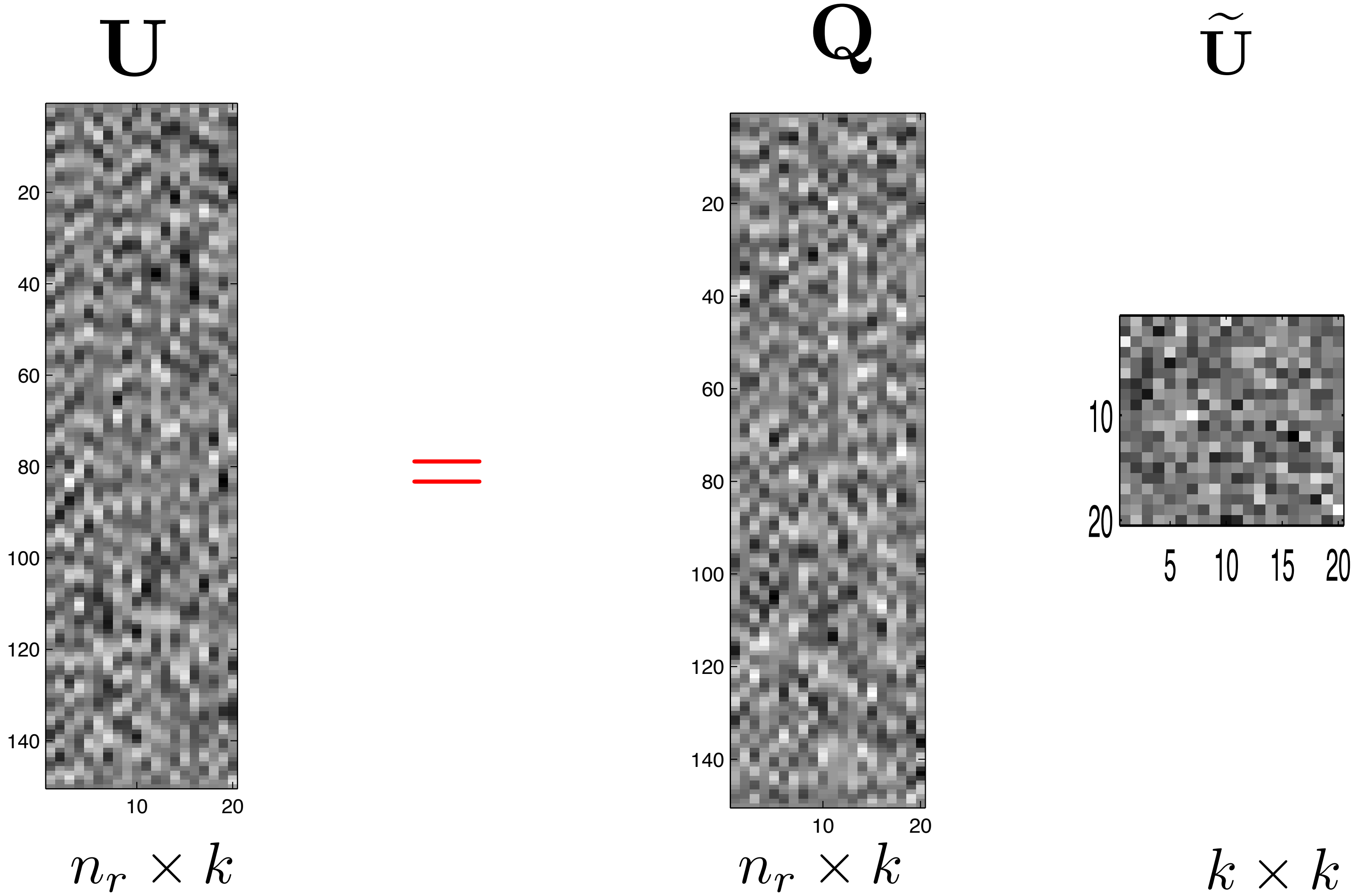


$k * n_s$ $\quad$ $k * k$ $\quad$ $k * k$ $\quad$ $k * n_s$

Dimensionality Reduction Via *RSVD*

Stage 2 : Compute an approximate SVD of $\hat{\mathbf{P}}$

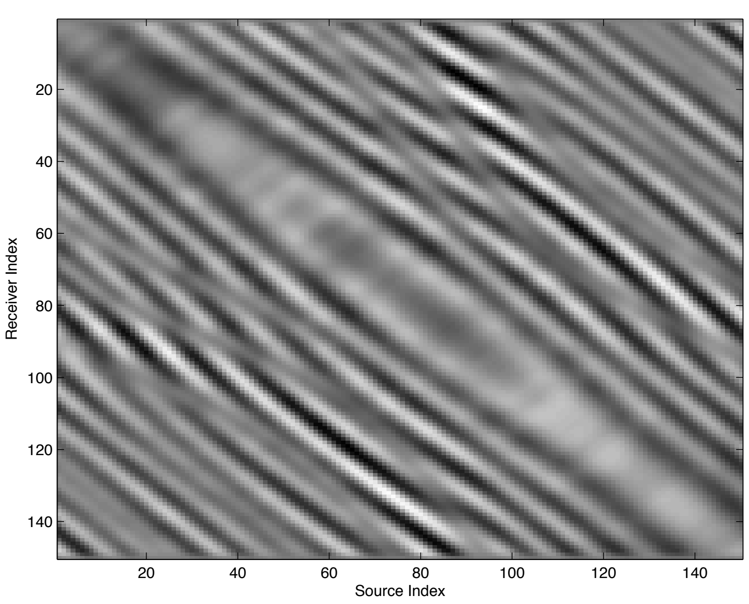
3. Compute $\mathbf{U} = \mathbf{Q}\widetilde{\mathbf{U}}$



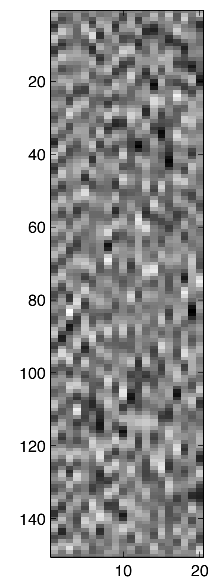
Dimensionality Reduction Via *RSVD*

Stage 2 : Compute an approximate SVD of $\hat{\mathbf{P}}$

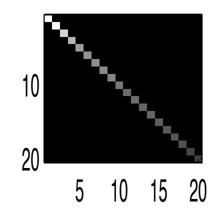
approximated $\hat{\mathbf{P}}$ $\approx$ $\mathbf{U}$ $*$ $\mathbf{\Sigma}$ $*$ $\mathbf{V}^*$



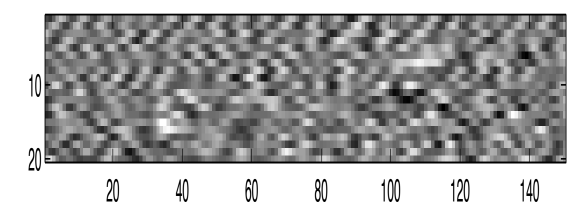
$n_r \times n_s$



$n_r \times k$



$k \times k$



$k \times n_s$

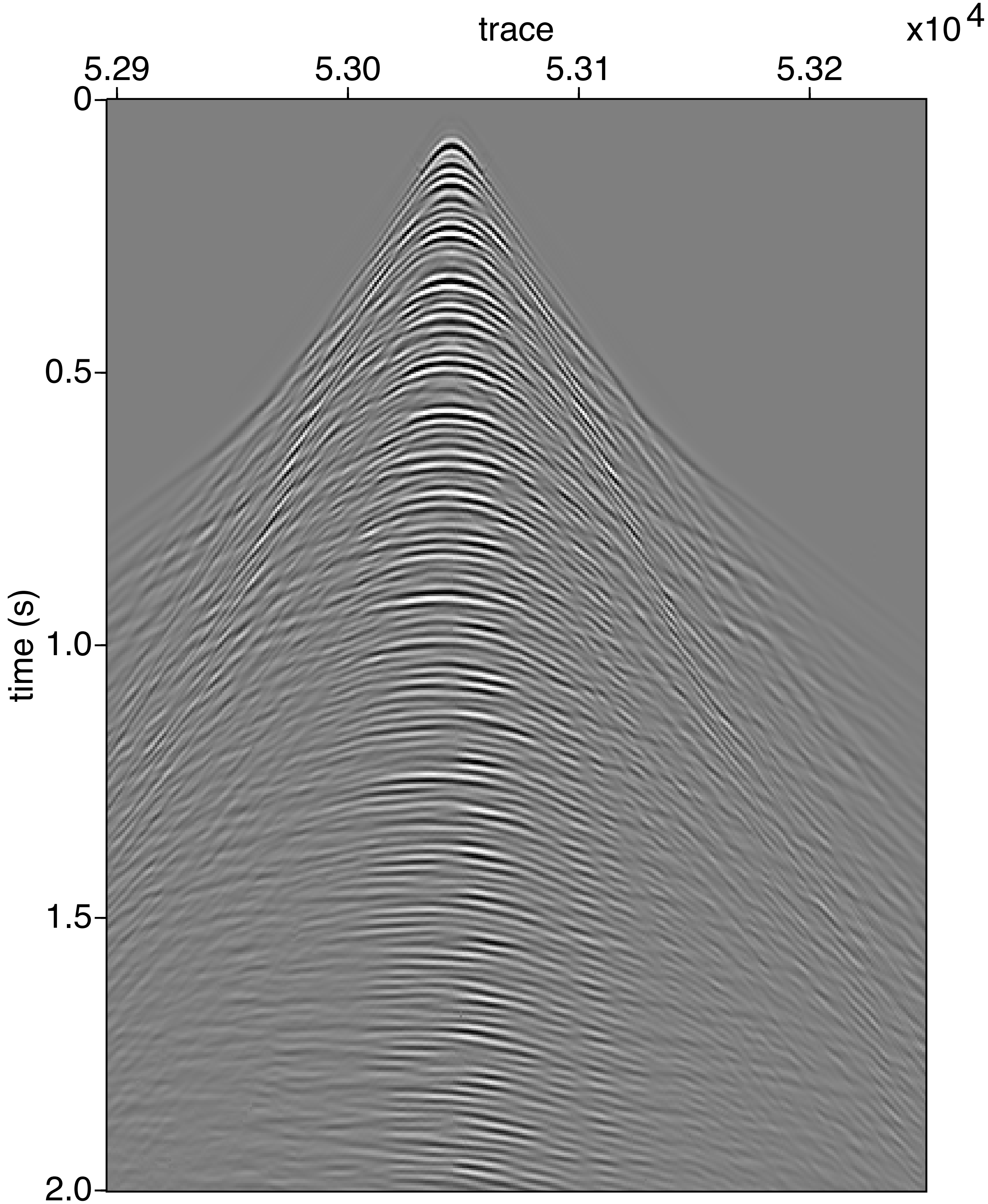
Summery

EPSI Formulation

R-SVD

Results

- 2D seismic line (Gulf of Suez)
  - $n_s = n_r = 355$
  - $n_t = 1024,\ dt = .004s$
- Adaptive approximation
- Compare results from EPSI using full vs. approximated data



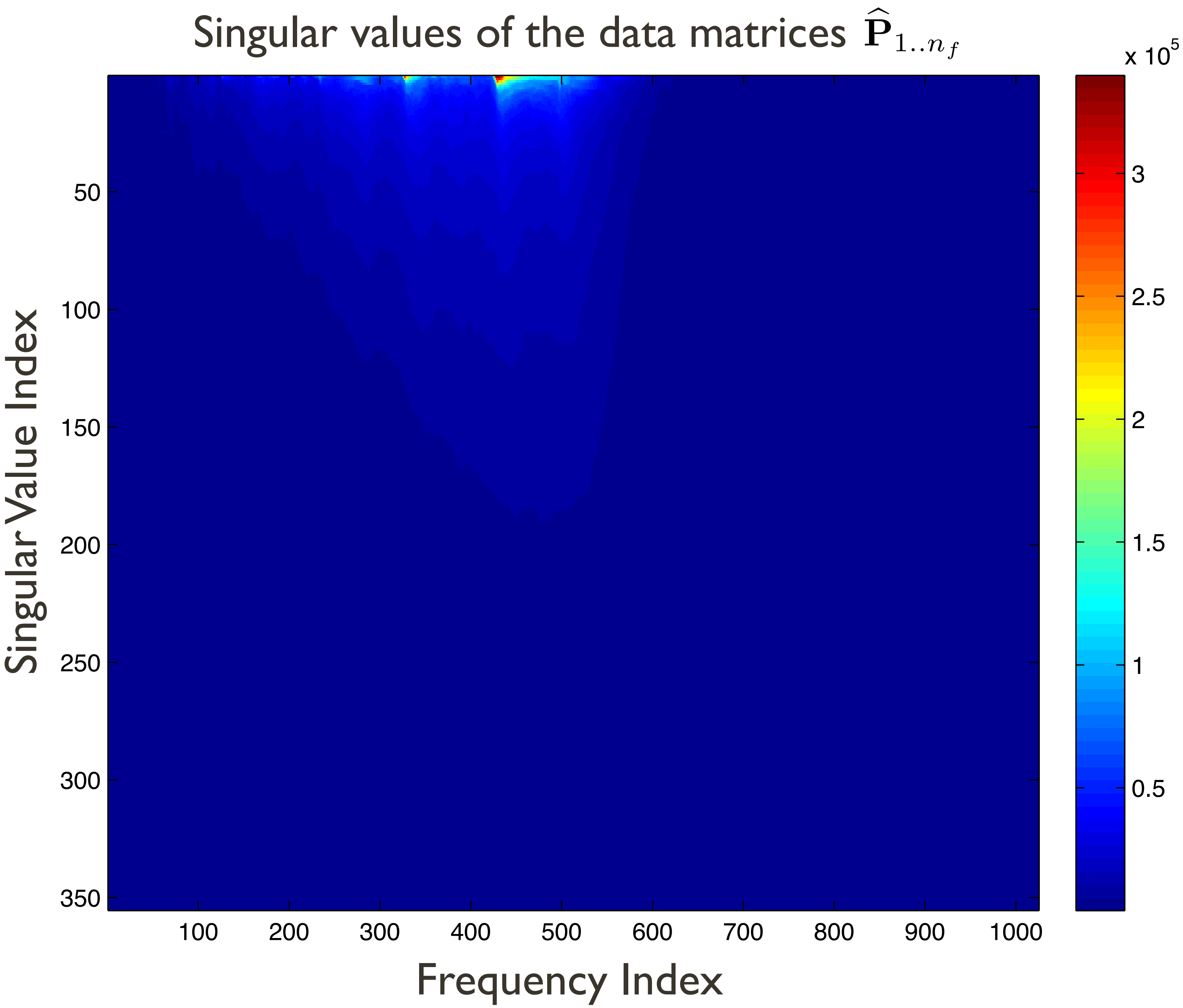
Gulf of Suez

Total Data

shot gather

$$n_r = 355$$
$$n_s = 355$$
$$n_t = 1024$$
$$dt = .004s$$

Singular values of the data matrix



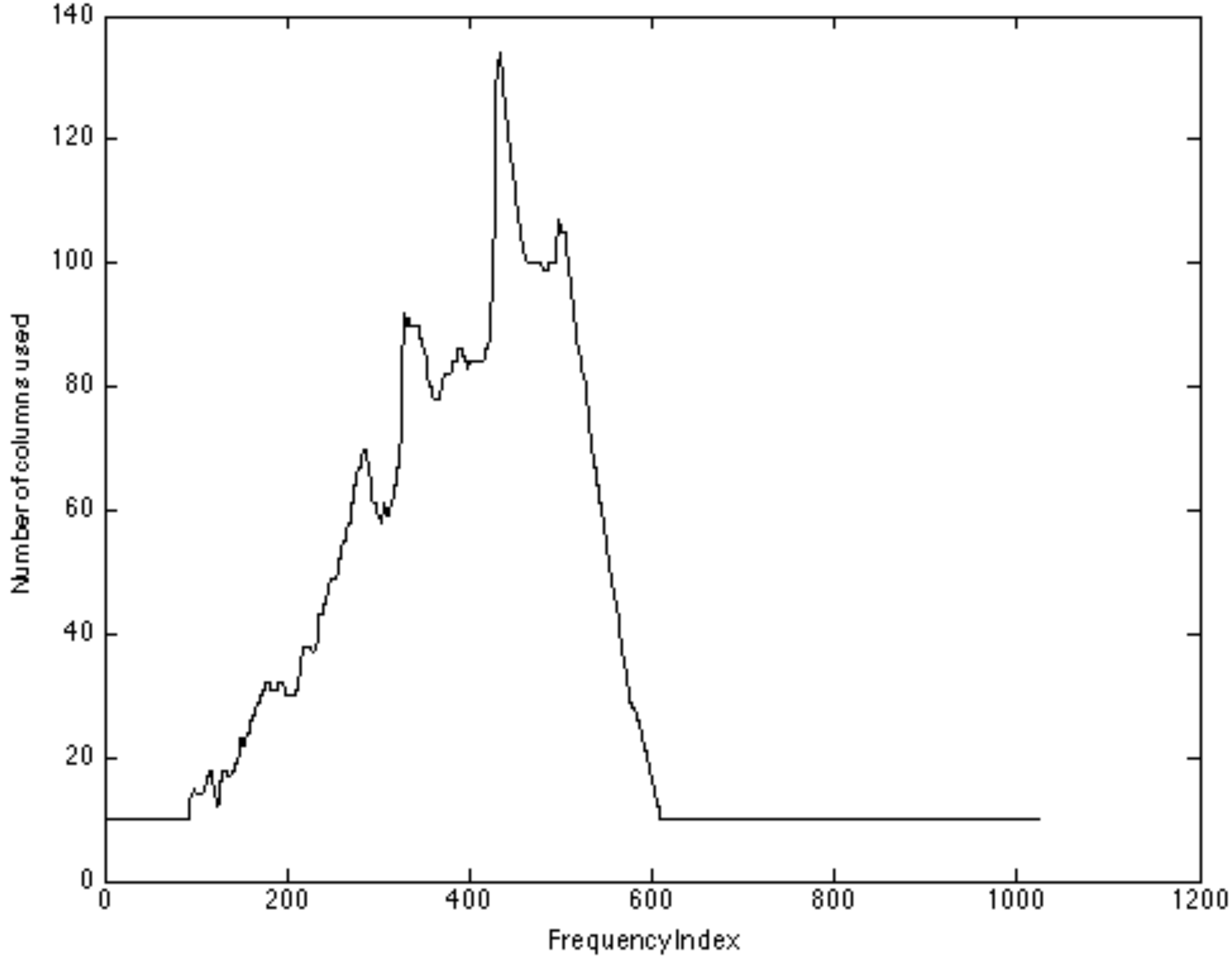
Adaptive approximation

For each frequency find rank k such that

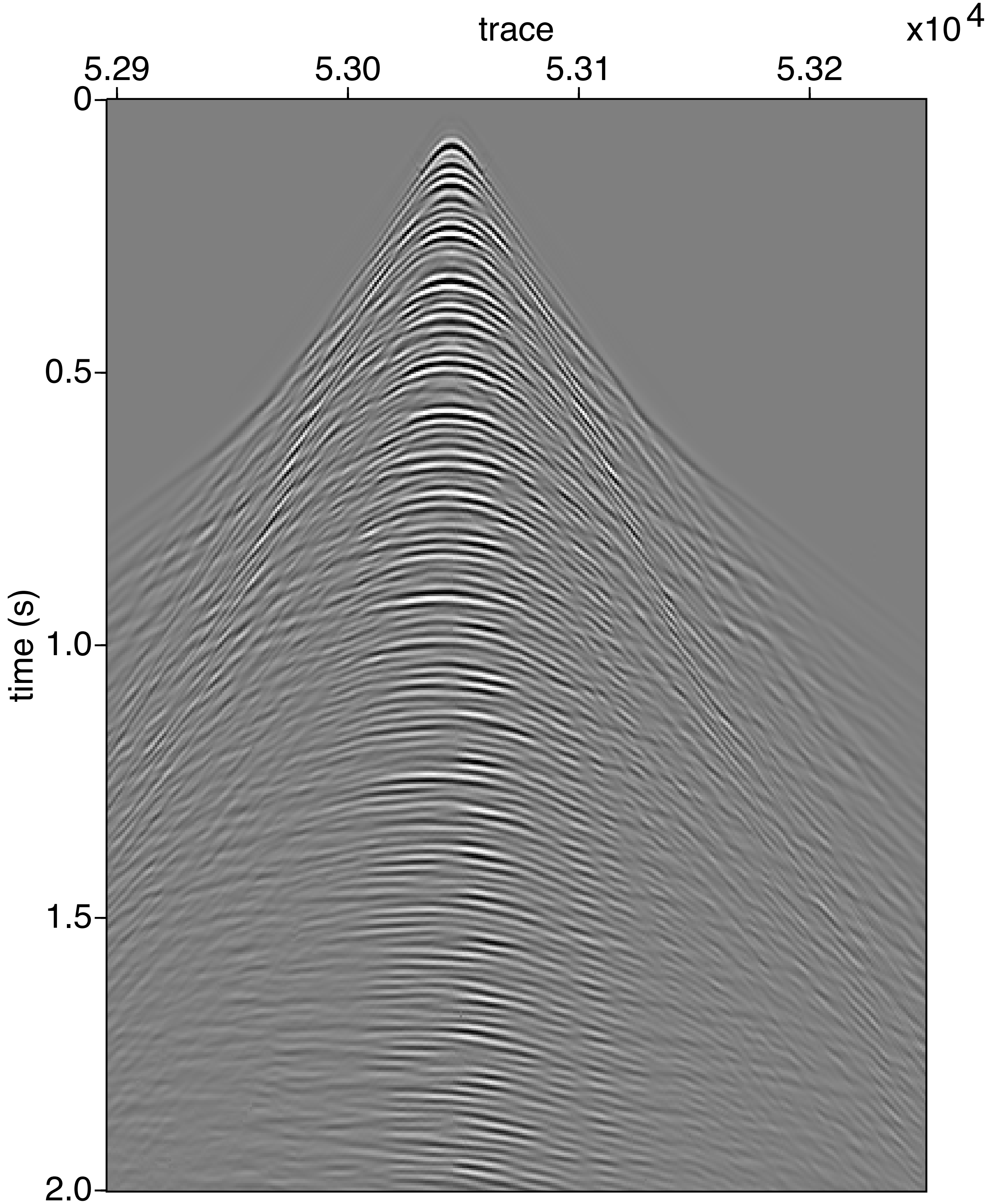
$$\|\mathbf{P} - \mathbf{USV}^*\|_{\mathbf{s}} \leq \epsilon$$

Rank budget = total number of columns

Adaptive rank selection



sum of all k ranks used is 9% of total number of columns

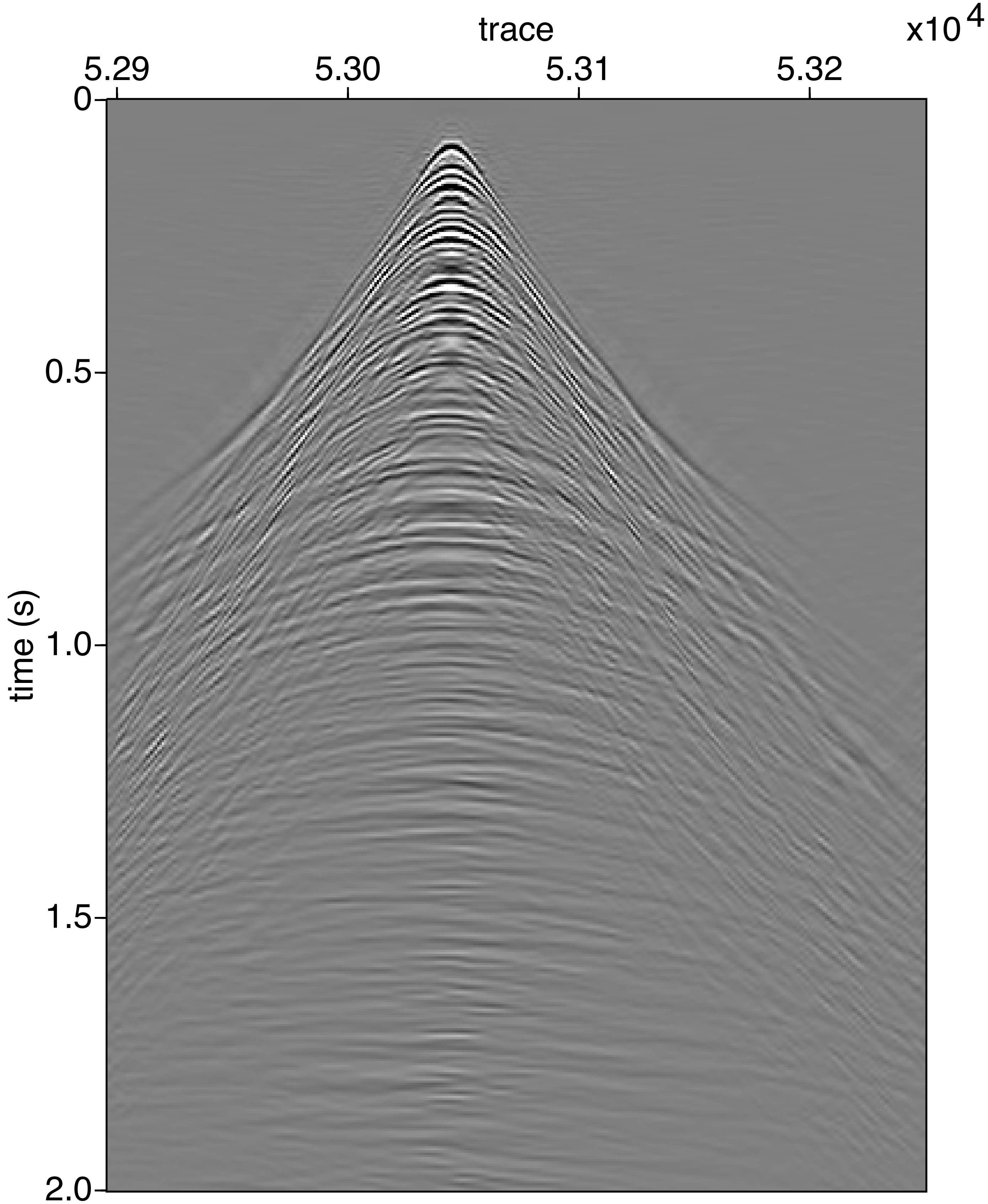


Gulf of Suez

Total Data

shot gather

$$n_r = 355$$
$$n_s = 355$$
$$n_t = 1024$$
$$dt = .004s$$



Gulf of Suez

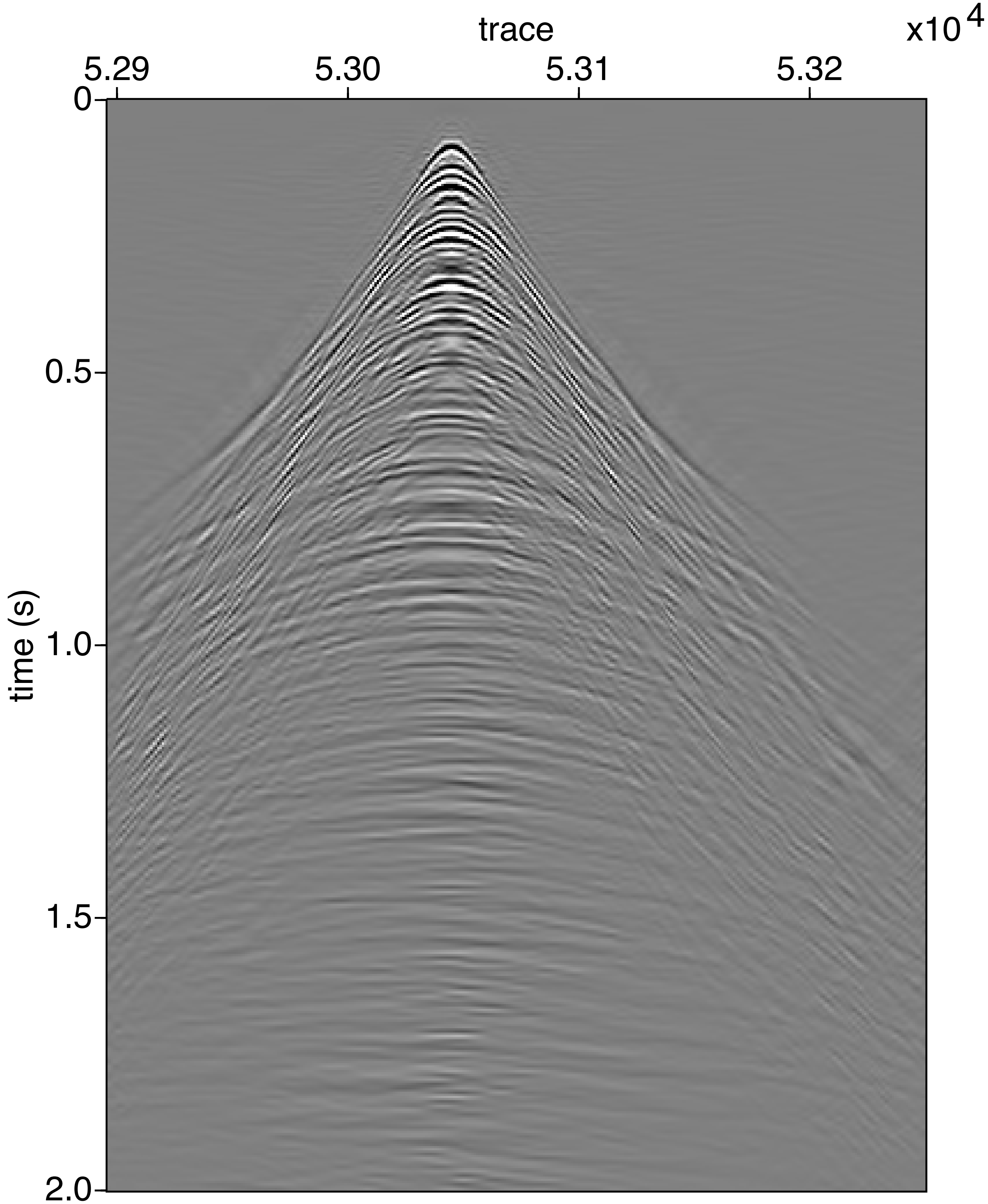
Full Data

Primary IR (G)

shot gather

2D Curvelet (Src-Rcv)

150 iterations



Gulf of Suez

20% of rank budget

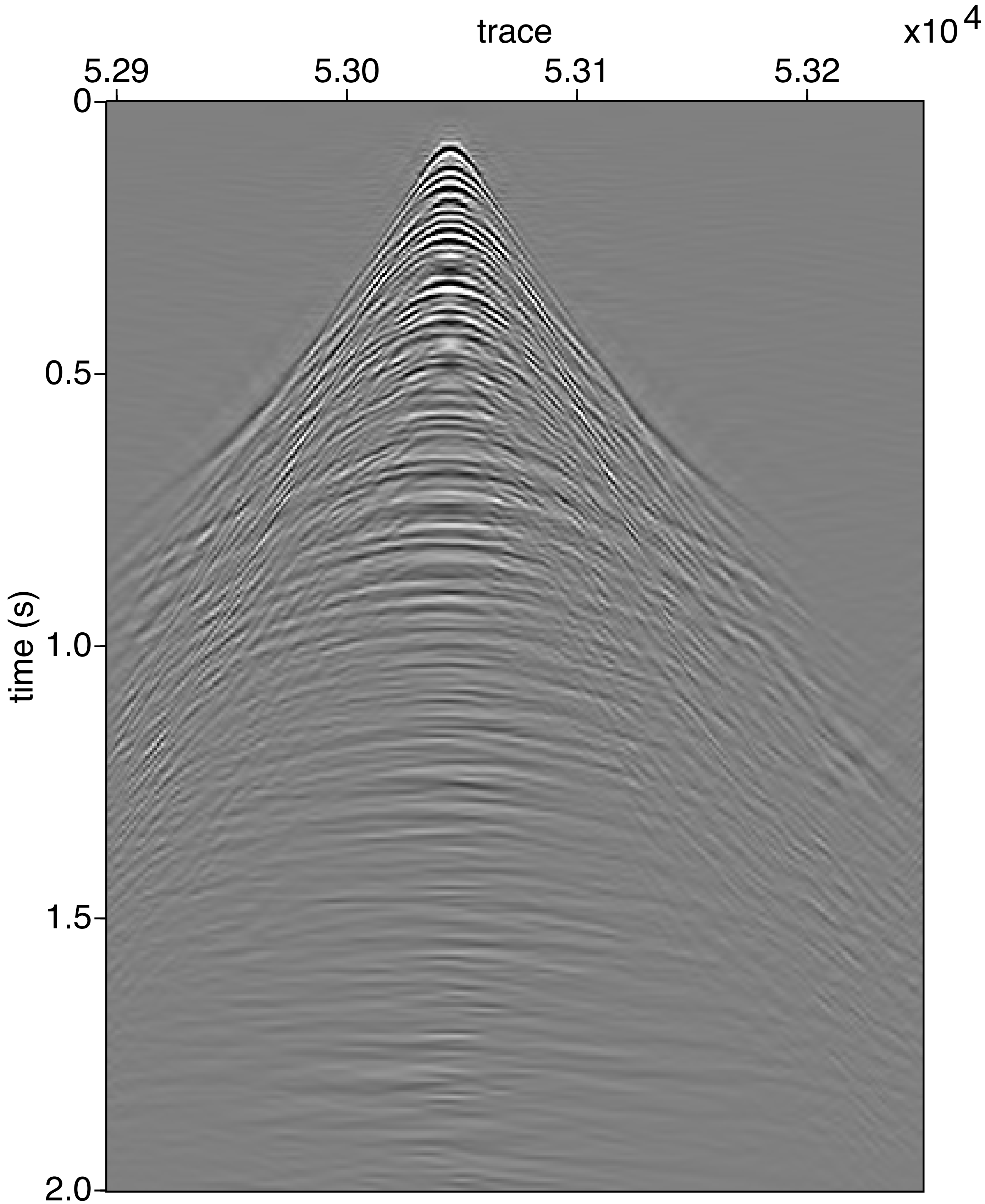
Primary IR (G)

SNR = 27dB

shot gather

2D Curvelet (Src-Rcv)

150 iterations



Gulf of Suez

12% of rank budget

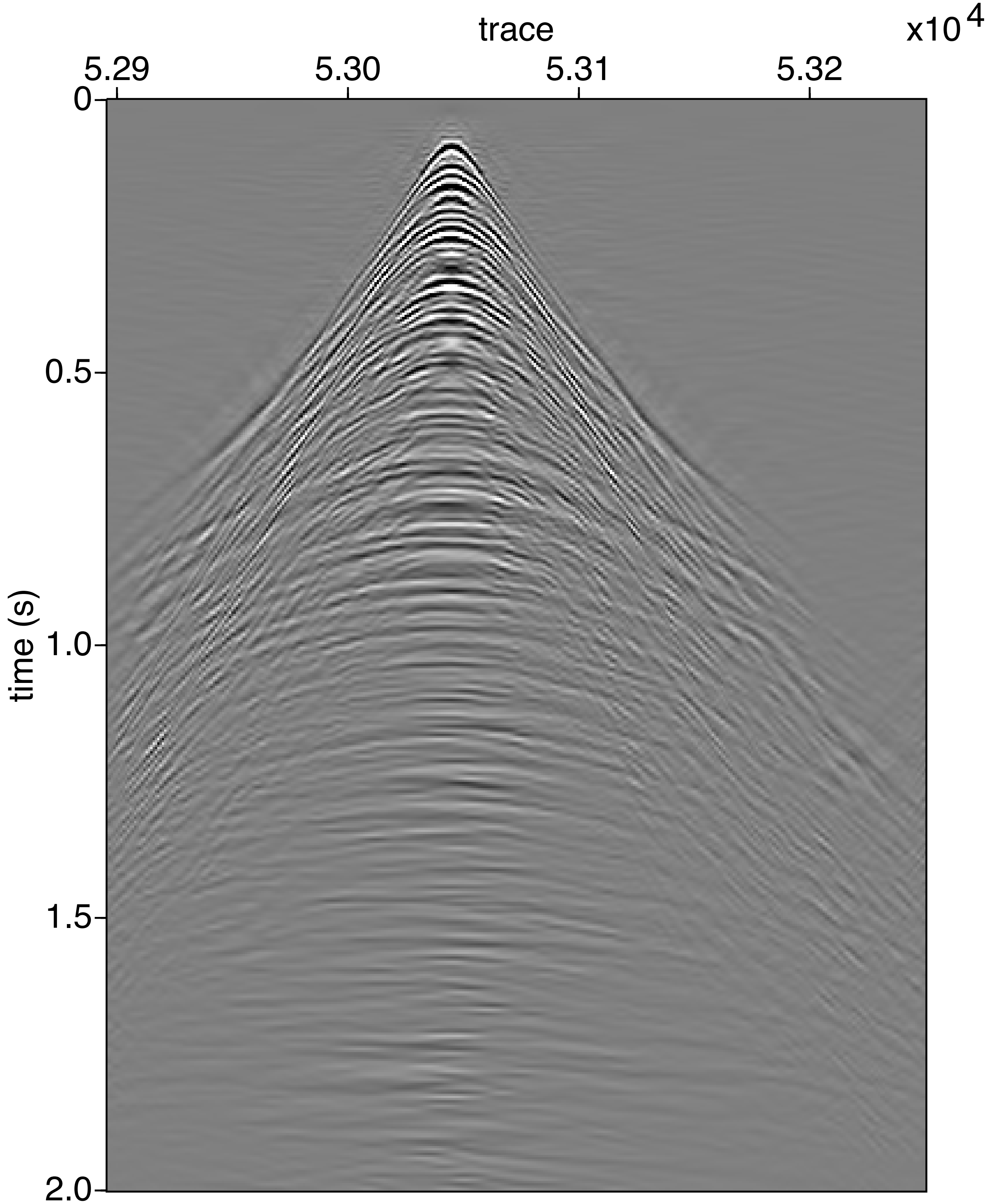
Primary IR (G)

SNR = 17dB

shot gather

2D Curvelet (Src-Rcv)

150 iterations



Gulf of Suez

8% of rank budget

Primary IR (G)

SNR = 12dB

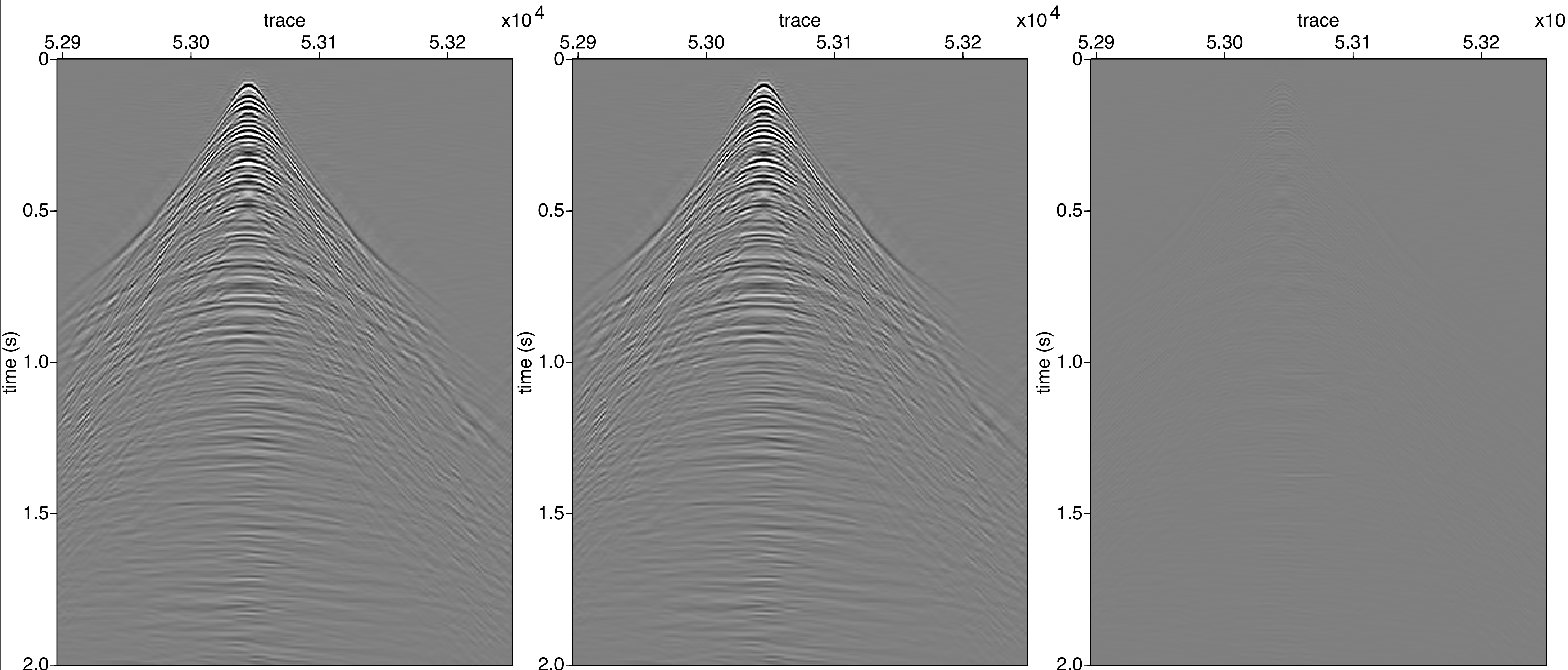
shot gather

2D Curvelet (Src-Rcv)

150 iterations

Difference in EPSI Result

20% rank budget



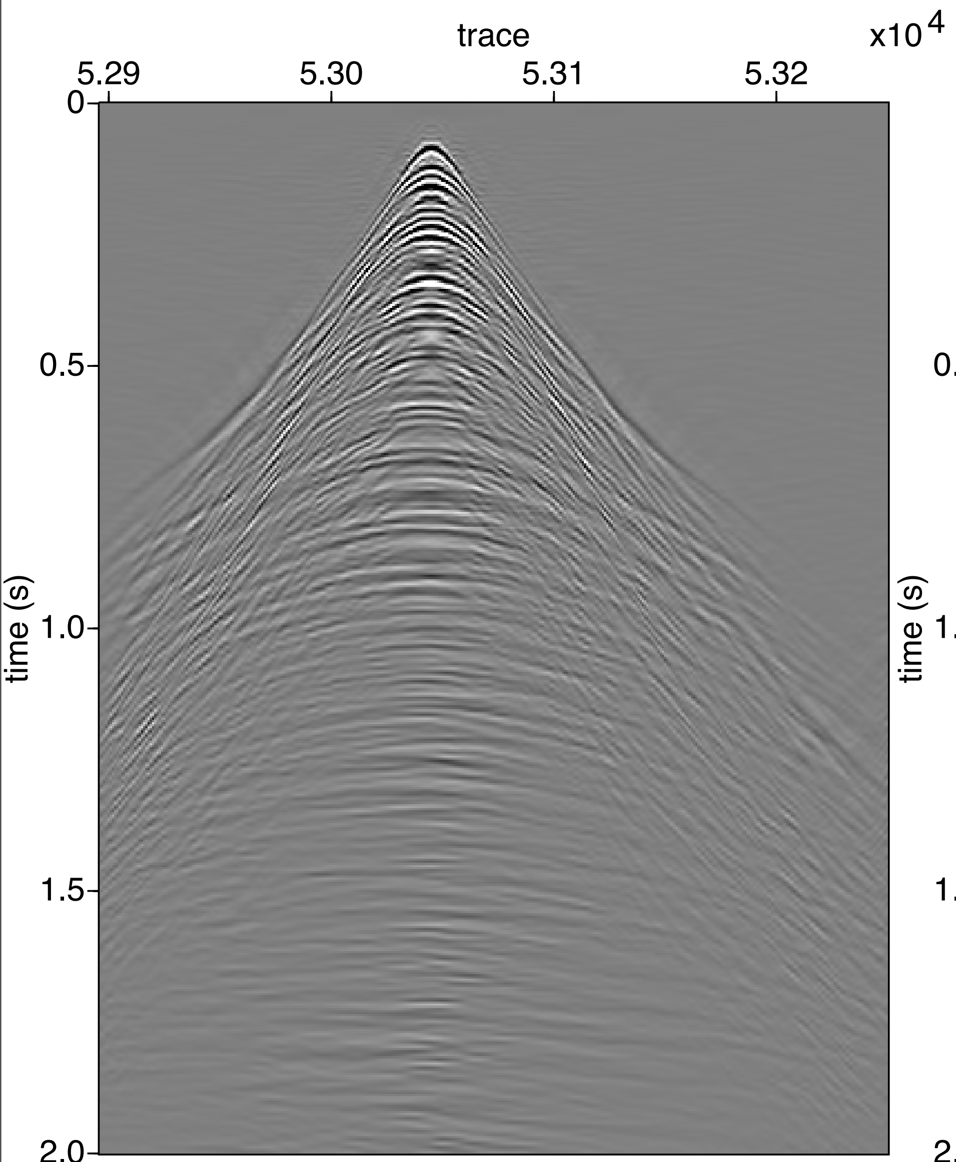
Primary IR full data

Primary IR approximated Data

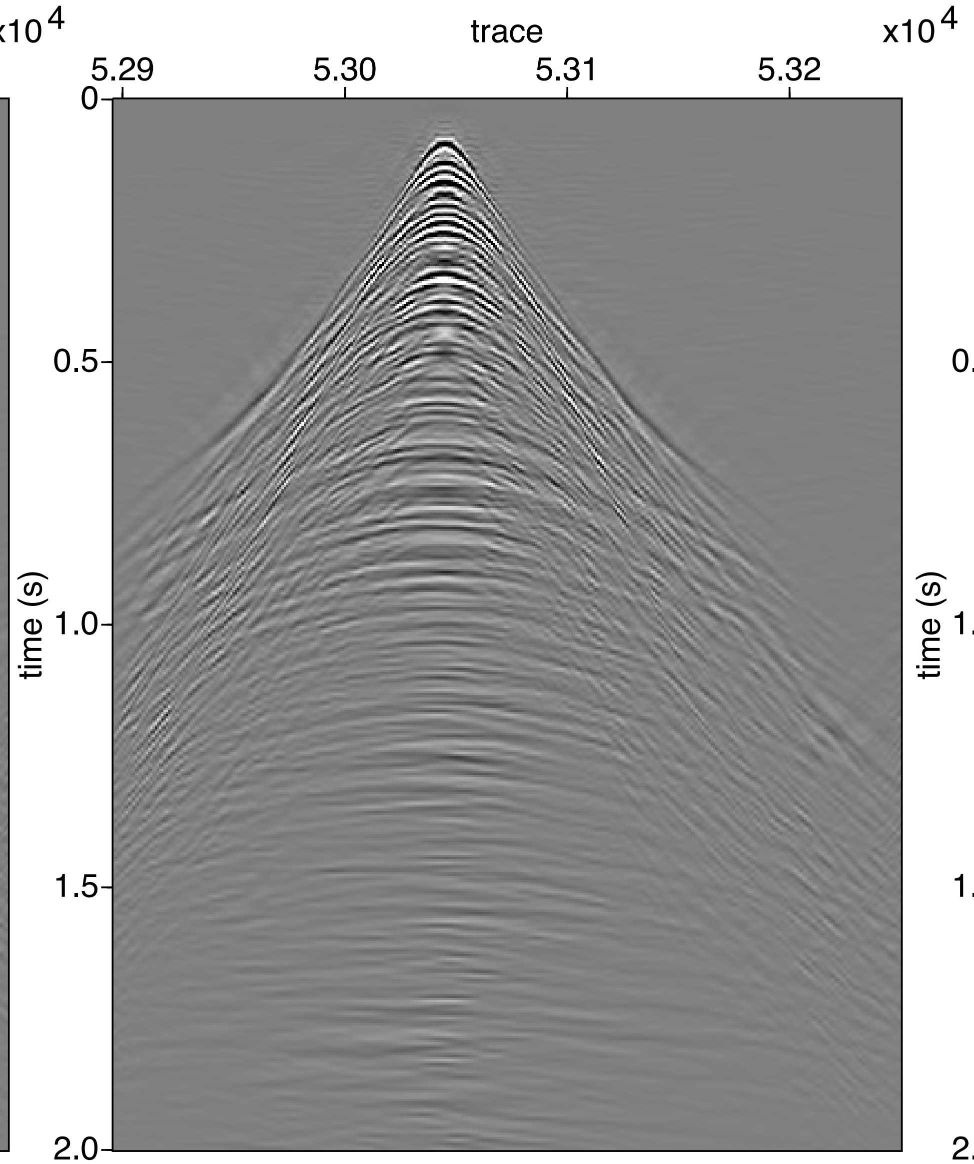
Difference

Difference in EPSI Result

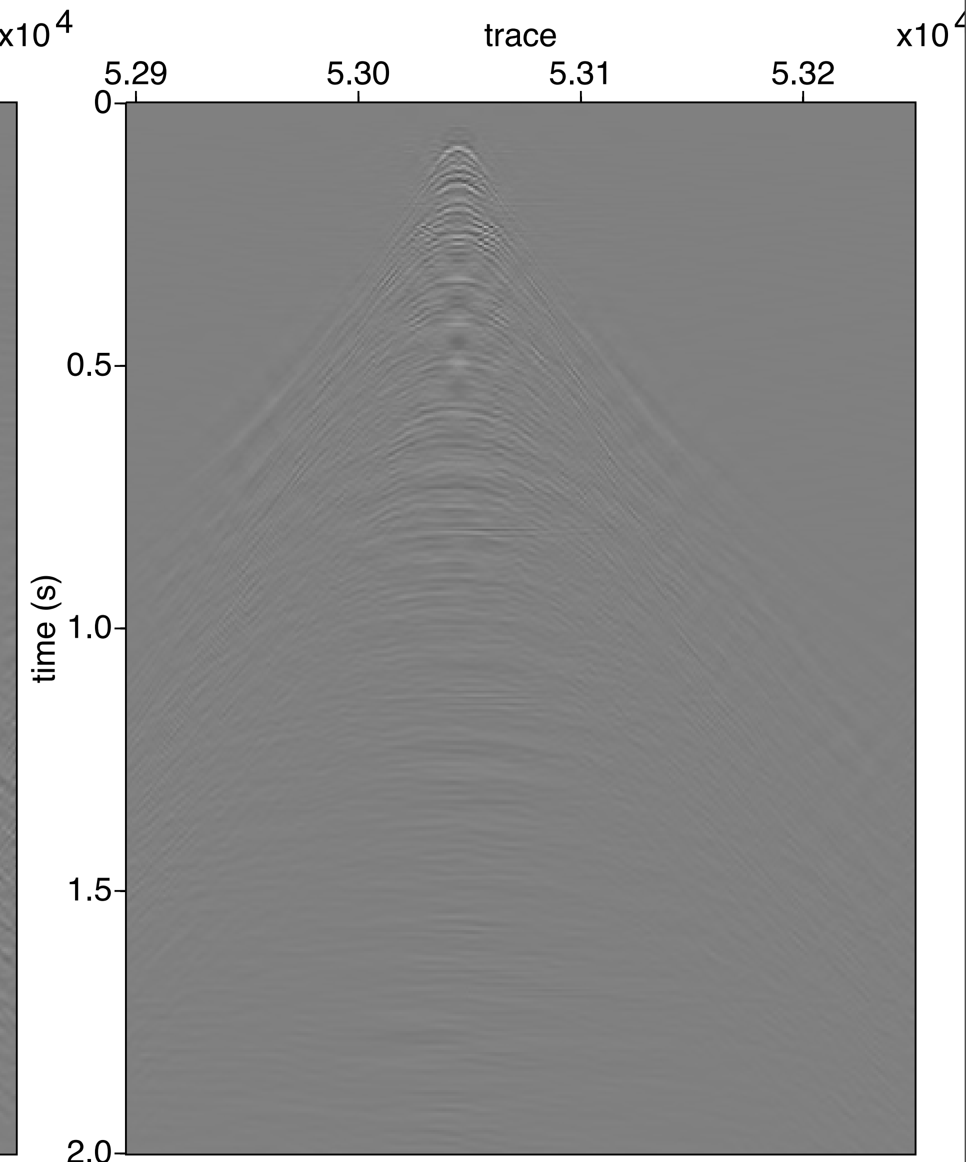
12% rank budget



Primary IR
full data



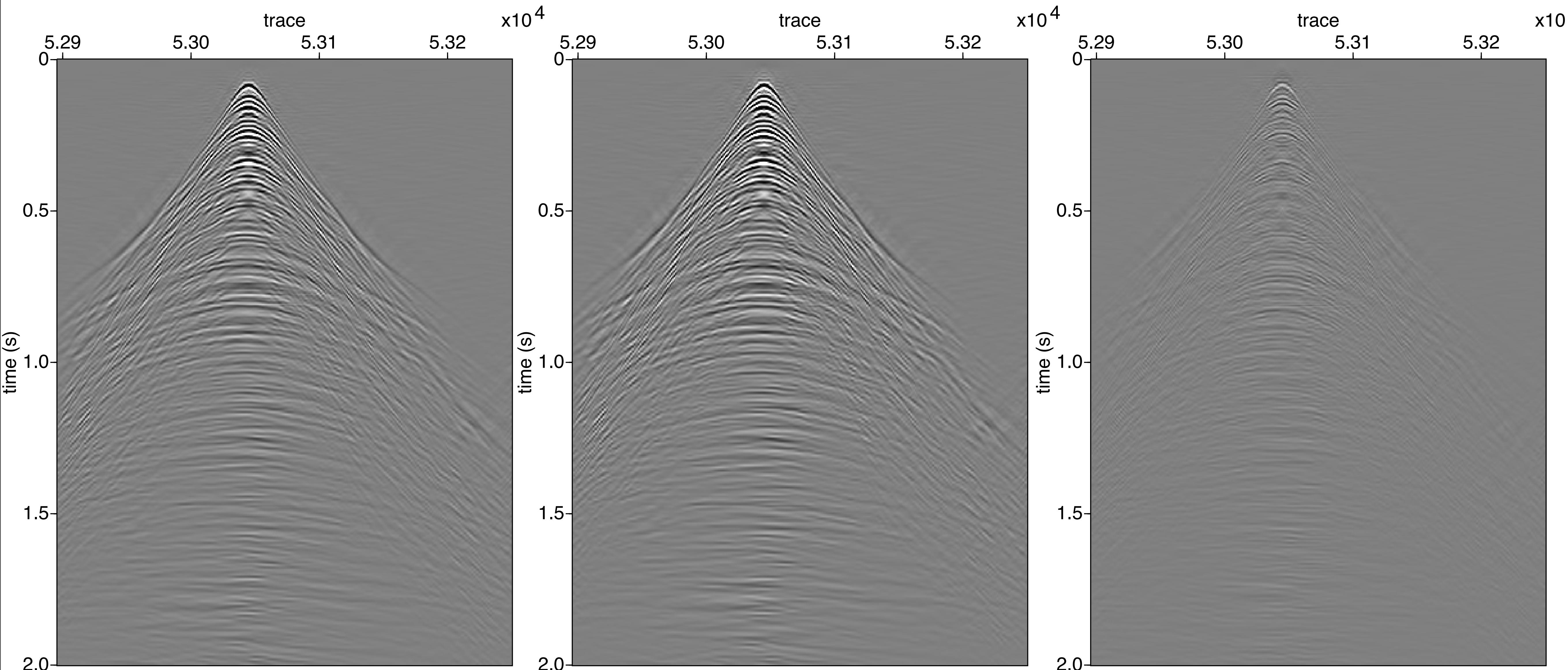
Primary IR
approximated Data



Difference

Difference in EPSI Result

8% rank budget



Primary IR full data

Primary IR approximated Data

Difference

Performance Summary

Rank Percentage	50%	20%	12%	8%
SNR (dB)	30	27	17	12
Multiplication Speedup	1.6x	2x	3.5x	5.7x
Memory savings	40%	50%	71%	82%

Conclusions

- Data driven methods - e.g. EPSI - suffers from the 'curse of dimensionality' when moving to 3D
- We utilize insights from random matrix theory to approximate action of the data matrix
- Reductions in multiplication and storage costs
- Up-Front cost of RSVD is O(mnlog(k))

SLIM

References

Berg, E. v., and M. P. Friedlander, 2008, Probing the Paretofrontier for basis pursuit solutions: SIAM Journal on Scien-tific Computing, 31, 890–912.

Berkhout, A. J., and D. J. Verschuur, 1997, Estimation of multi-ple scattering byiterative inversion, part I: theoretical considerations: Geophysics, 62, 1586–1595.

Candes, E., and B. Recht, 2009, Exact matrix completion viaconvex optimization: Foundations of Computational Mathe-matics, 9, 717–772.

Gandy, S., B. Recht, and I. Yamada, 2011, Tensor completionand low-n-rank tensor recovery via convex optimization:Inverse Problems, 27, 025010.Habashy, T. M., A. Abubakar, G. Pan, and A. Belani, 2010

Halko, N., P. G. Martinsson, and J. A. Tropp, 2011, Find-ing structure with randomness: Probabilistic algorithms forconstructing approximate matrix

Herrmann, F. J., 2010, Randomized sampling and sparsity:Getting more information from fewer samples: Geophysics,75, WB173–WB187.

References

Herrmann, F. J., and D. Wang, 2008, Seismic wavefield inversion with curvelet-domain sparsity promotion: SEG Technical Program Expanded Abstracts, SEG, 2497–2501.

Lin, T., and F. J. Herrmann, 2009, Unified compressive sensingframework for simultaneous acquisition with primary esti-mation: SEG Technical Program Expanded Abstracts, SEG,3113–3117.

Minato, S., T. Matsuoka, T. Tsuji, D. Draganov, J. Hunziker,and K. Wapenaar, 2011, Seismic interferometry using multidimensional deconvolution and crosscorrelation for crosswell seismic reflection data without borehole sources: Geophysics, 76, SA19–SA34.

van Groenestijn, G. J. A., and D. J. Verschuur, 2009, Estimating primaries by sparse inversion and application to near-offsetdata reconstruction: Geophysics, 74, A23–A28.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics,

57, 1166–1177.

Acknowledgements

Special thanks to:

SLIM group members

Saudi Aramco

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

Thank you!

slim.eos.ubc.ca