Challenges & opportunities for Compressive Sensing in seismic acquisition

Felix J. Herrmann

SLIM

Seismic Laboratory for Imaging and Modeling
the University of British Columbia

[Candès et.al, Donoho, '06]

SLIM

Compressive sensing

New *rigorous* theory and *concrete* implementation with

- *sampling* & *compression* combined
- *recovery* by exploiting *structure*
- recovery *guarantees*

Major breakthrough in a wide range of fields

- signal/image processing
- MRI imaging
- scientific computing

SLIM

Premise

Signals in nature including seismic wavefields & sedimentary basins *exhibit* some sort of *structure*

- transform-domain sparsity
- low-rank property

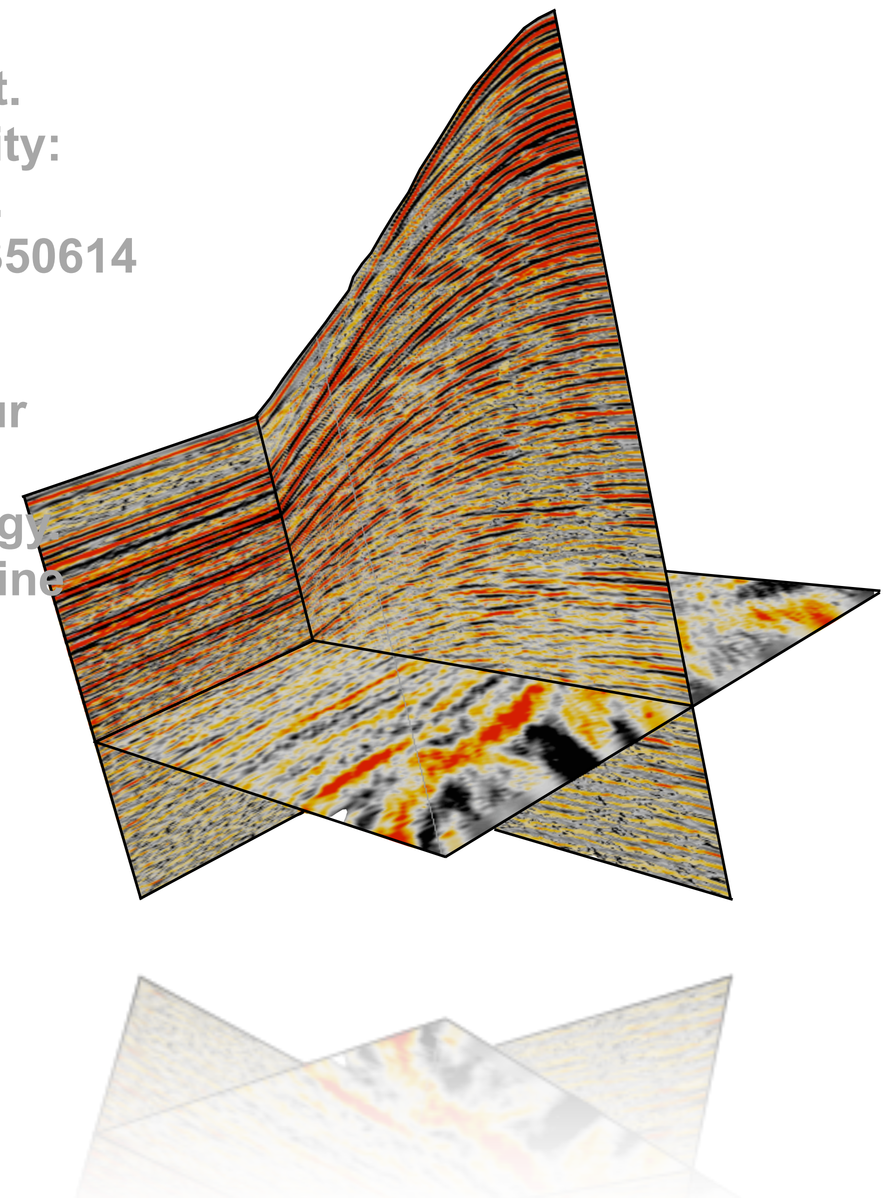
If this is true, can we use this observation to use these properties during sampling and inversion?

Compressive sensing *delivers* on this by coming up with a *rigorous* theory and sampling *criteria* that guarantee recovery from *severe* subsamplings.

Basics of compressive sensing

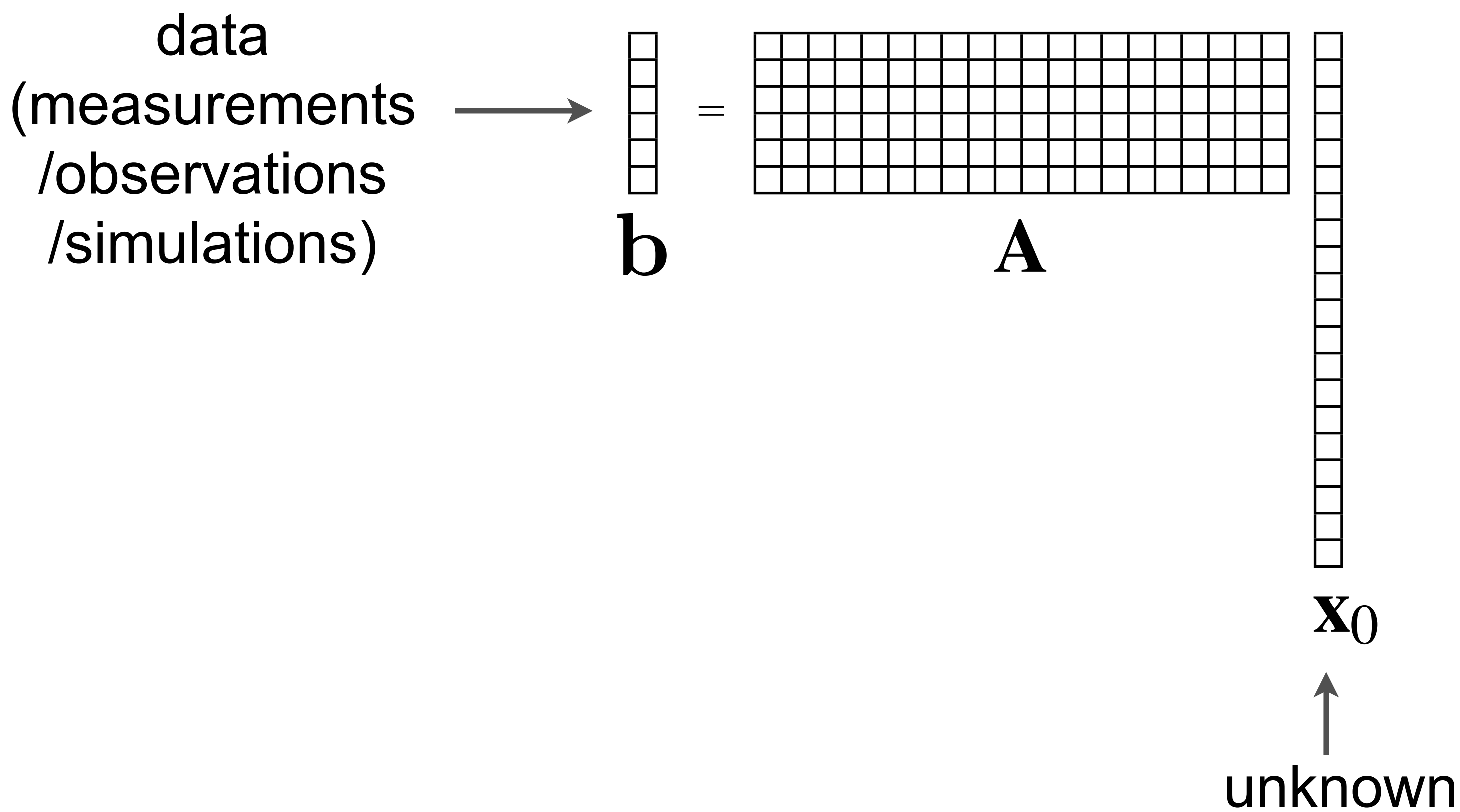
Felix J. Herrmann. UBC-EOS Technical Report. TR-2010-01. Randomized sampling and sparsity: getting more information from fewer samples. Geophysics 75, WB173 (2010); doi:10.1190/1.350614

Felix J.Herrmann, Michael P.Friedlander, Ozgur Yilmaz. Fighting the curse of dimensionality: compressive sensing in exploration seismology. 2011. In revision for Signal Processing Magazine



Problem statement

Consider the following (severely) *underdetermined* system of linear equations with **A** a *n* X *N* matrix with *n*<<*N*



Is it possible to recover $\mathbf{x}_0$ accurately from **b**

- in case $\mathbf{x}_0$ has *k* non-zeros?
- in case $\mathbf{x}_0$ is *compressible*, i.e., has *few* large entries and *many* small ones?

Recovery

Naive first guess would be to recover via

$$\underbrace{\min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_2} = \Big(\sum_n |x_i|^2\Big)^{1/2}}_{\text{energy}} \quad \text{subject to} \quad \underbrace{\mathbf{A}\mathbf{x} = \mathbf{y}}_{\text{perfect reconstruction}}$$

with analytic solution:

$$\tilde{\mathbf{x}} = \mathbf{A}^*\big(\mathbf{A}\mathbf{A}^*\big)^{-1}\mathbf{y}$$

will not find the k-sparse solution when n<<*N*.

Recovery by strict sparsity promotion

Better choice would be to recover via

$$\underbrace{\min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_0} = \#\text{nonzeros}\{\mathbf{x}\}}_{\text{strict sparsity}} \quad \text{subject to} \quad \underbrace{\mathbf{A}\mathbf{x} = \mathbf{y}}_{\text{perfect reconstruction}}$$

- if every n X n submatrix of **A** is nonsingular then this program recovers every k-sparse vector exactly when $k<n/2$
- We **only** need $n>2k$ measurements regardless of N.
- ***no*** analytic solution
- numerically unstable
- NP-hard problem

Recovery by relaxed sparsity promotion

Convexify via one-norm minimization

$$\underbrace{\min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_1} = \sum_n |x_i|}_{\text{"sparsity"}} \quad \text{subject to} \quad \underbrace{\mathbf{A}\mathbf{x} = \mathbf{y}}_{\text{perfect reconstruction}}$$

will recover *k*-sparse solutions with *overwhelming probability* from

$$n \geq c\,k \log(N/n) \quad \text{measurements}$$

- ***no*** analytic solution
- stable
- computationally feasible
- extends to compressible signals

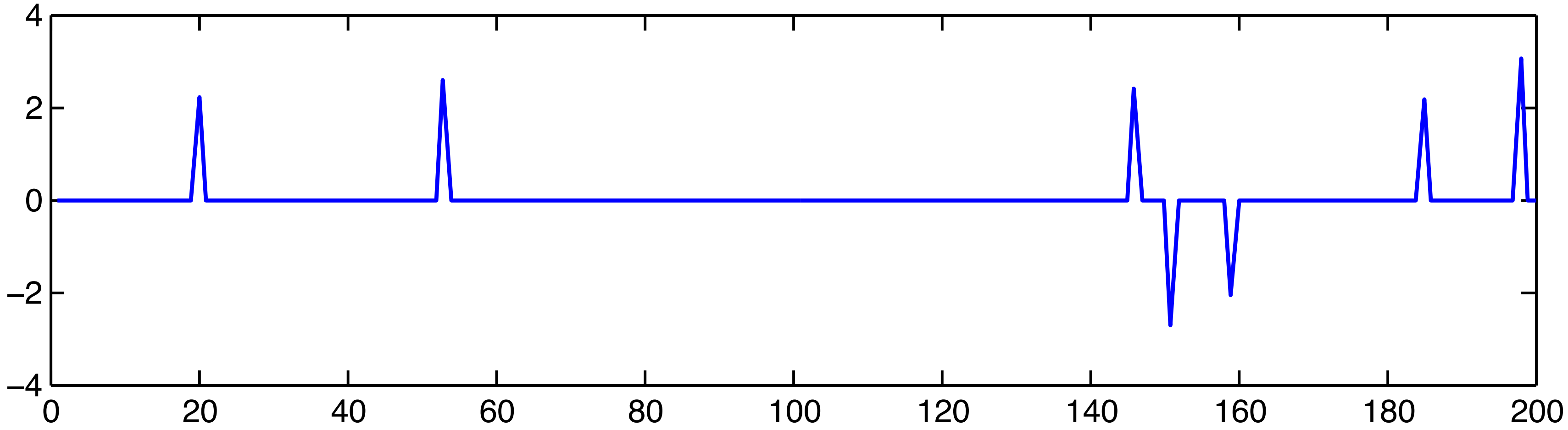
[Candès et al.'06]
[Donoho'06]

Example different recovery techniques

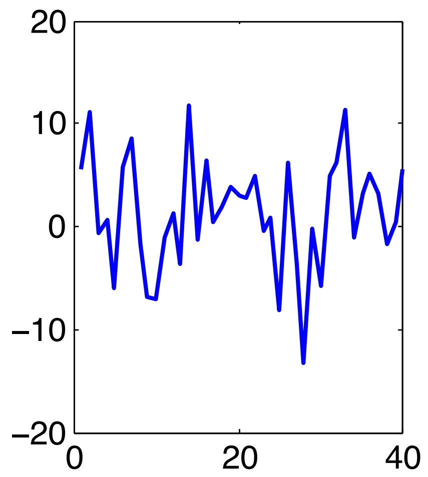
Setup example in Matlab and run comparison

```
% L2, L1 recovery comparison for sparse signals

A = randn(40,200);
x = sparsify(randn(200,1),7);   % 7 nonzero elements
plot(x)

y = A*x;
plot(y)

x_ell2 = lsqr(A,y);
plot(x_ell2)

x_ell1 = spgl1(A,y,0,1e-7,[]);
plot(x_ell1)
```

Example different recovery techniques

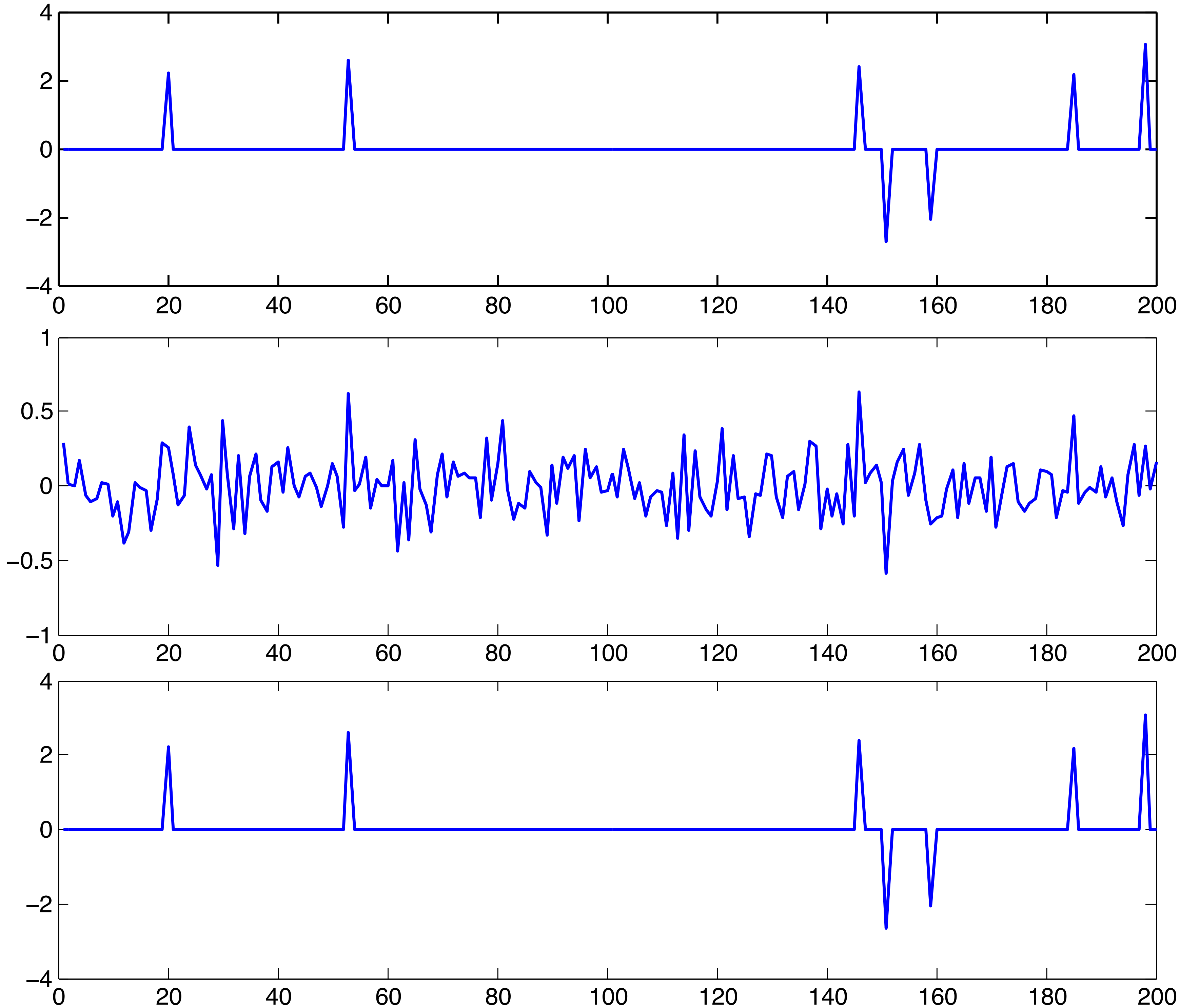


Original signal
x



sampled signal
y

Example different recovery techniques



Original signal

$\mathbf{x}$

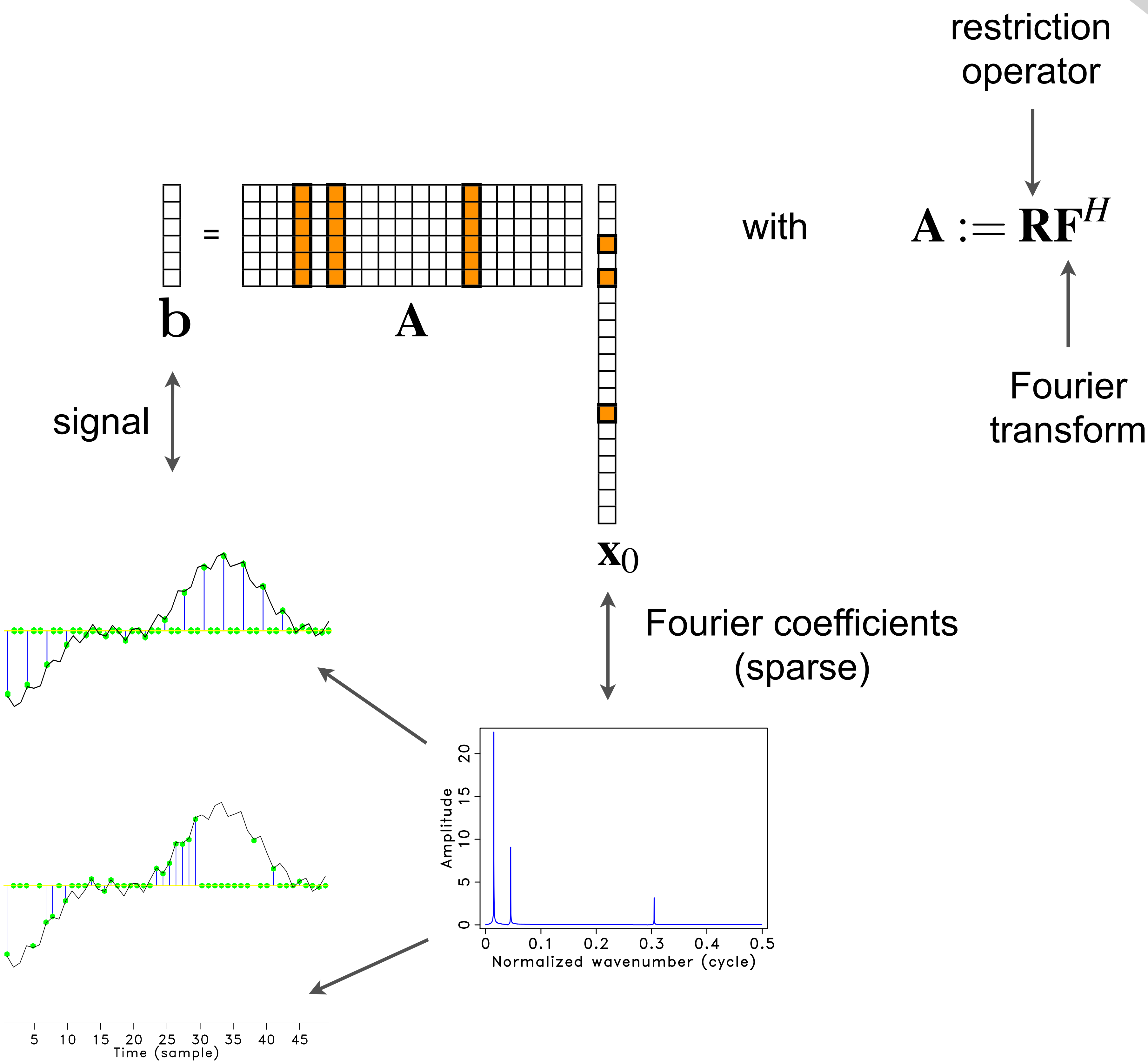
energy recovery

$\tilde{\mathbf{x}}_{\ell_2}$

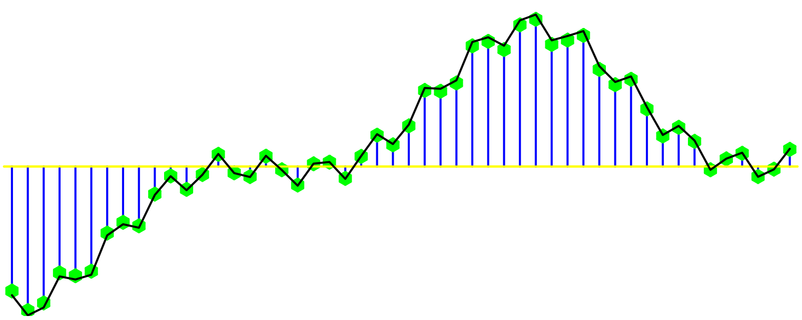
one-norm recovery

$\tilde{\mathbf{x}}_{\ell_1}$

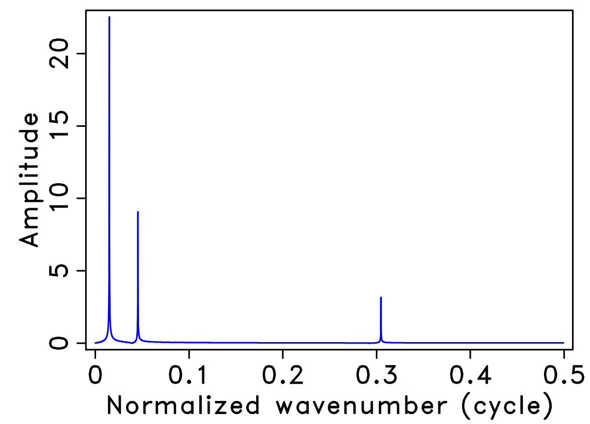
Sparse recovery



Coarse sampling schemes

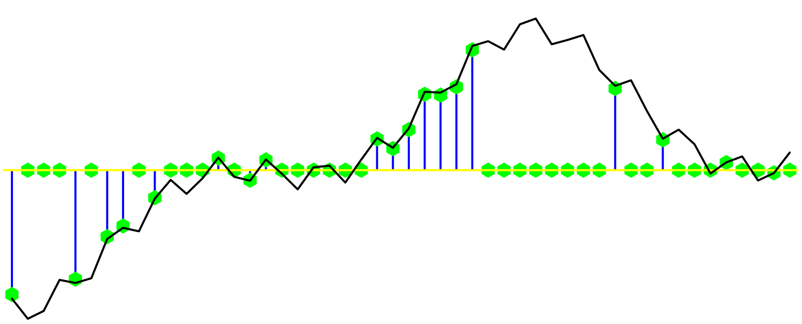


Fourier → transform

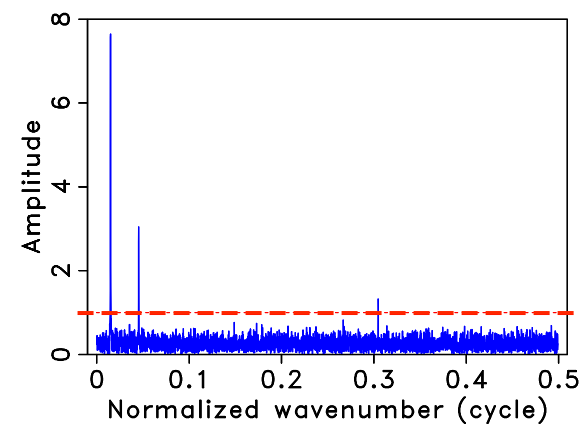


few significant coefficients

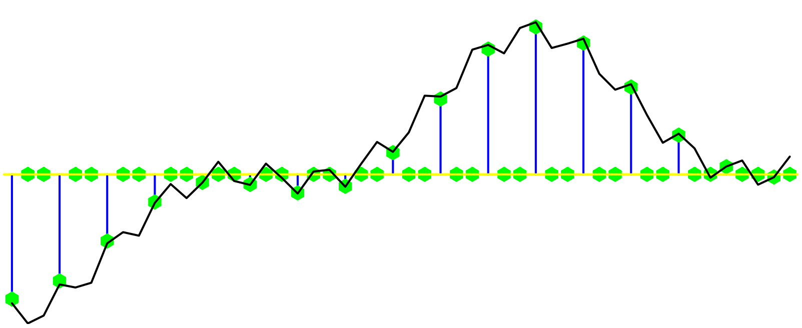
3-fold under-sampling



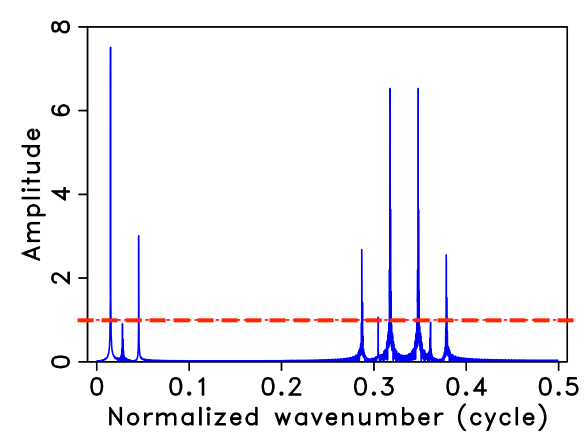
Fourier → transform



✓ significant coefficients detected



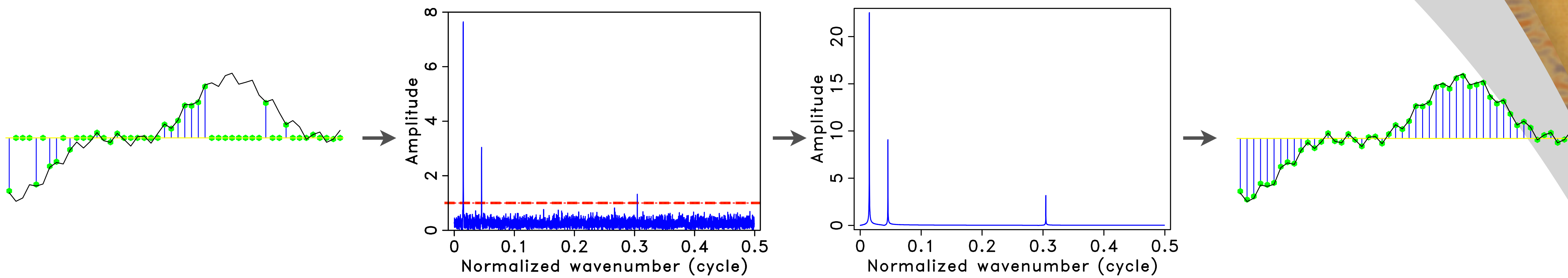
Fourier → transform



✗ ambiguity

[Hennenfent & Herrmann, ’08]

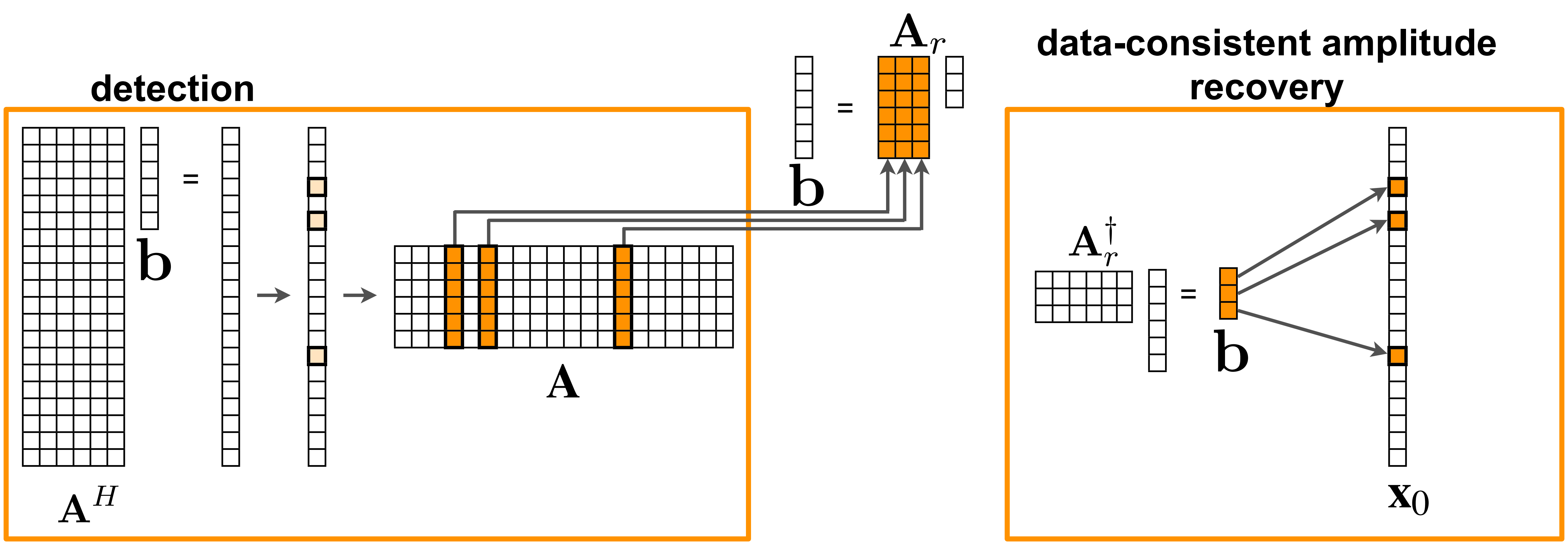
NAIVE sparsity-promoting recovery



inverse Fourier transform

detection + data-consistent amplitude recovery

Fourier transform

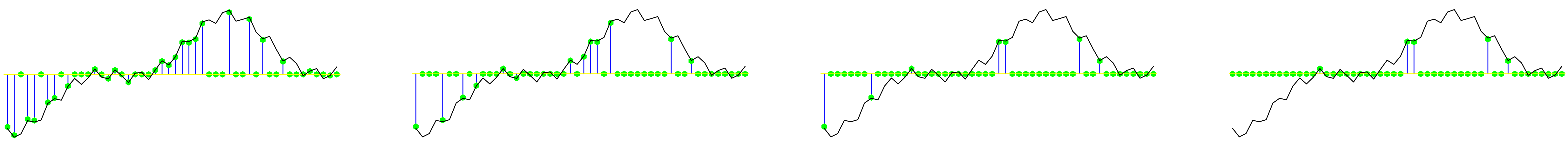


Undersampling “noise”

“noise” *interferences:*

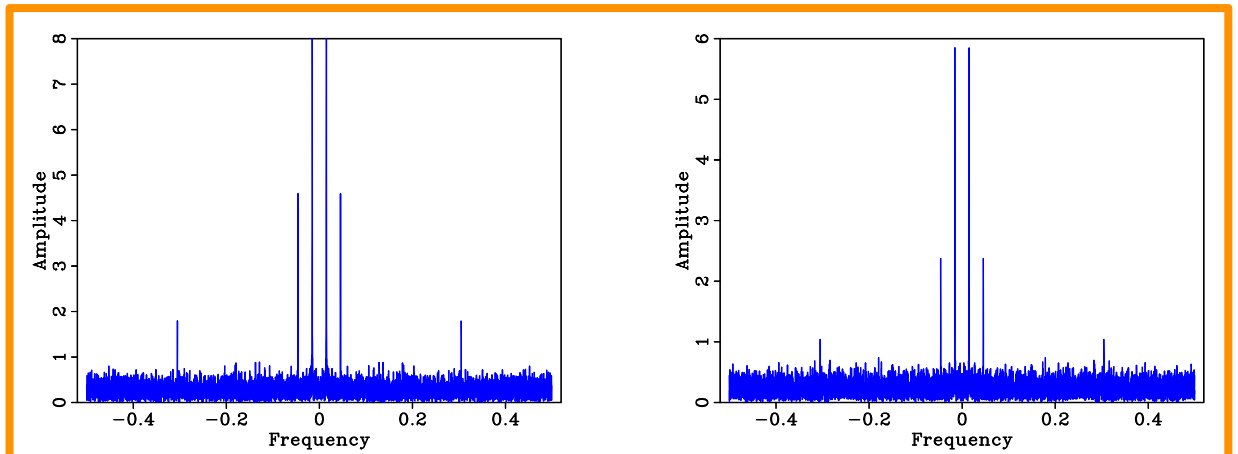
- due to $\mathbf{A}^H\mathbf{A} \neq \mathbf{I}$ (Gram matrix)
- defined by $\mathbf{A}^H\mathbf{A}\mathbf{x}_0 - \alpha\mathbf{x}_0 = \mathbf{A}^H\mathbf{b} - \alpha\mathbf{x}_0$

1 out of 2 | 1 out of 4 | 1 out of 6 | 1 out of 8

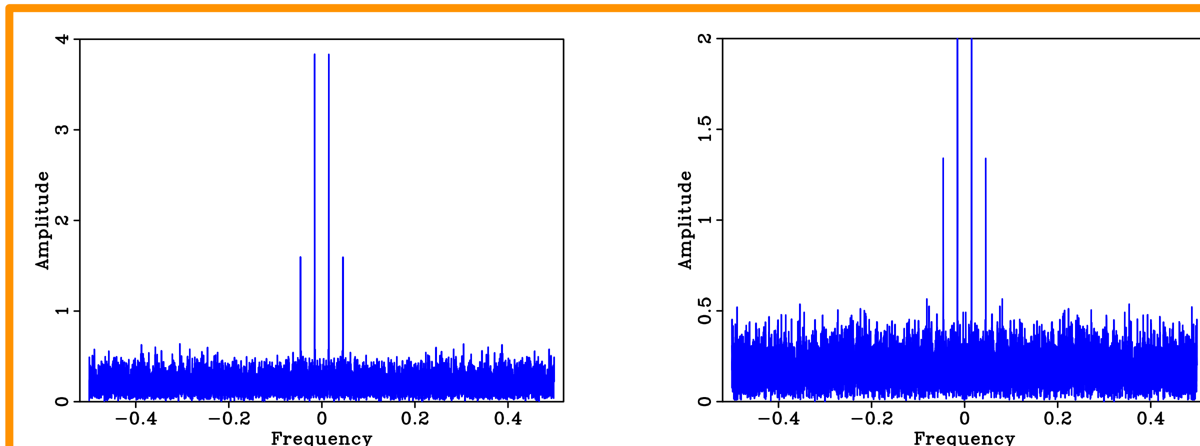


less acquired data →

3 detectable Fourier modes



2 detectable Fourier modes



SLIM

Mutual coherence

Compressive Sensing is all about *controlling* the *off-diagonals* of the *Gram* matrix

Accomplished by a *combination* of

- *randomization*
- with *spreading* of *sampling* vectors in the *sparsifying* domain
  - e.g. Fourier *vs* Dirac

[Candès et.al, Donoho, '06]

SLIM

Sparse recovery

Solve *optimization* problem:

$$\min_{\mathbf{x}} \overbrace{\|\mathbf{x}\|_1}^{\text{detection}} \quad \text{subject to} \quad \overbrace{\mathbf{b} = \mathbf{A}\mathbf{x}}^{\text{data-consistent amplitude recovery}}$$

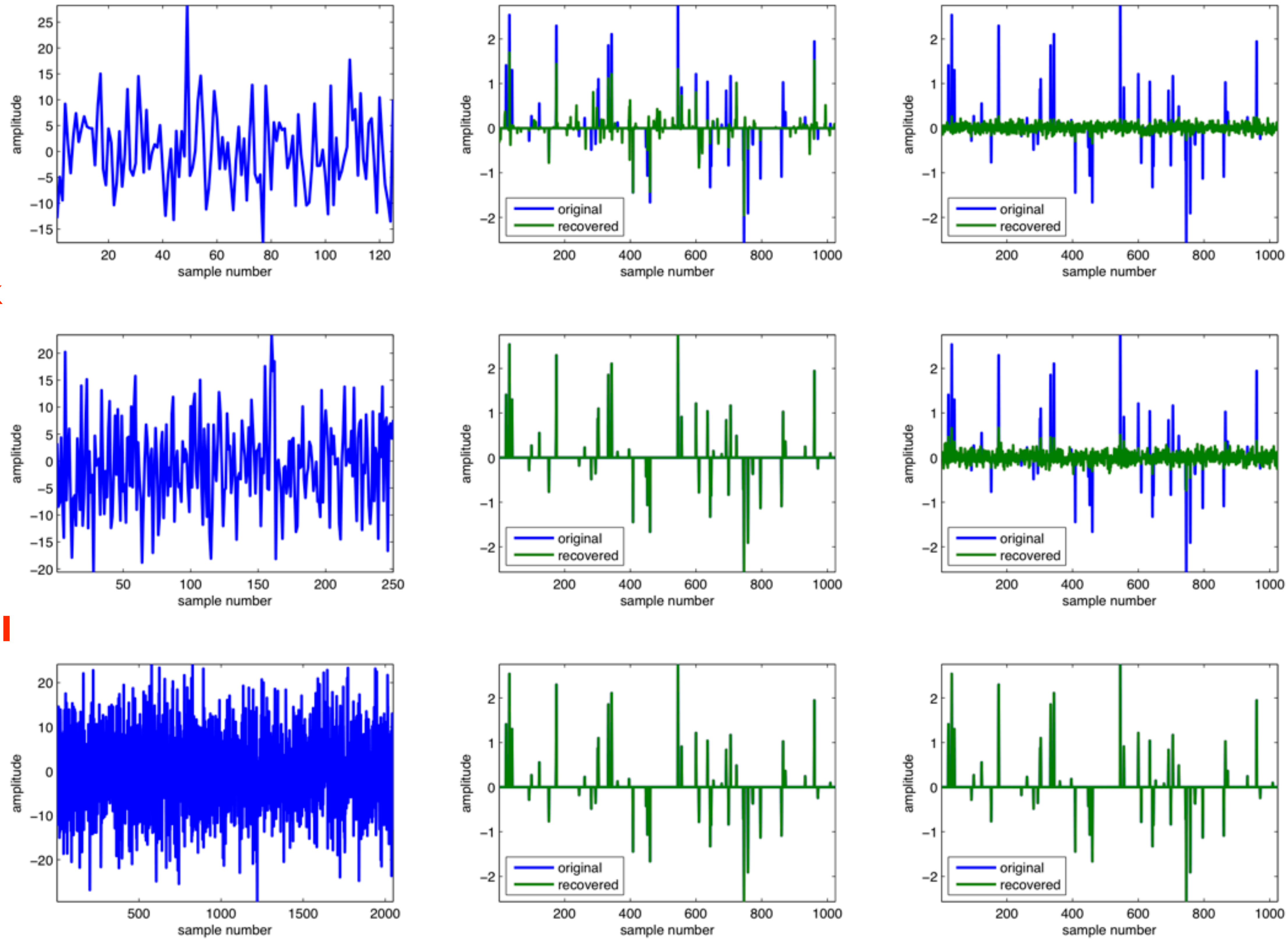
- *convexification* of the NP-hard *zero-norm* problem
- *suite* of different large-scale solvers available
- recovery *quality* depends on *coherence* & *aspect ratio* of **A** & *sparsity* of **x**

Example [signal length 1024, 50 non zeros]

2.5 X k | **measurements** | ℓ_1 **recovery** | ℓ_2 **recovery**

5 X k

20 X l



Observations

Compressive Sensing breaks *coherent/periodic interferences*

- *randomization & incoherence*

Sparsity-promoting recovery hinges on

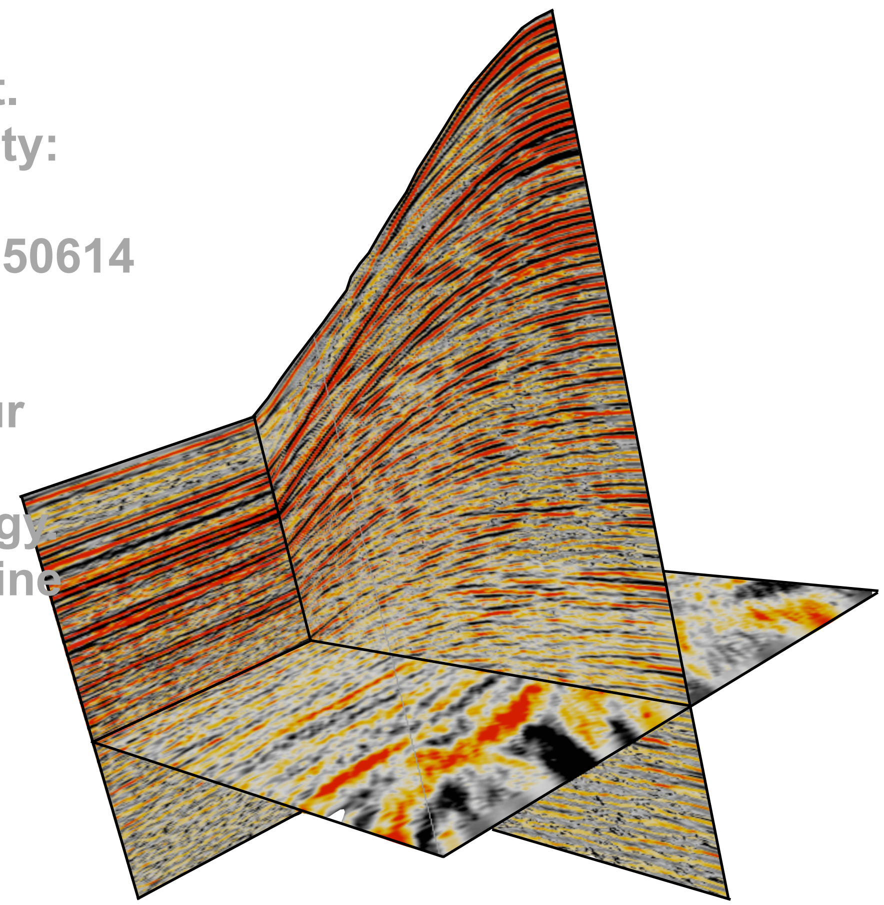
- *aspect ratio* of $\mathbf{A}$ & *sparsity* of $\mathbf{x}$
- *mutual coherence* of $\mathbf{A}$

Compressive Sensing is a *design* problem seeking *sampling* & *sparsifying* transforms that *act* as *random Gaussian matrices*...

Compressive acquisition

Felix J. Herrmann. UBC-EOS Technical Report. TR-2010-01. Randomized sampling and sparsity: getting more information from fewer samples. Geophysics 75, WB173 (2010); doi:10.1190/1.350614

Felix J.Herrmann, Michael P.Friedlander, Ozgur Yilmaz. Fighting the curse of dimensionality: compressive sensing in exploration seismology. 2011. In revision for Signal Processing Magazine



Compressive acquisition

Challenge:

Acquisition costs are determined by Nyquist

Key idea: *Randomize* acquisition, *subsample*, and *sparse* recovery

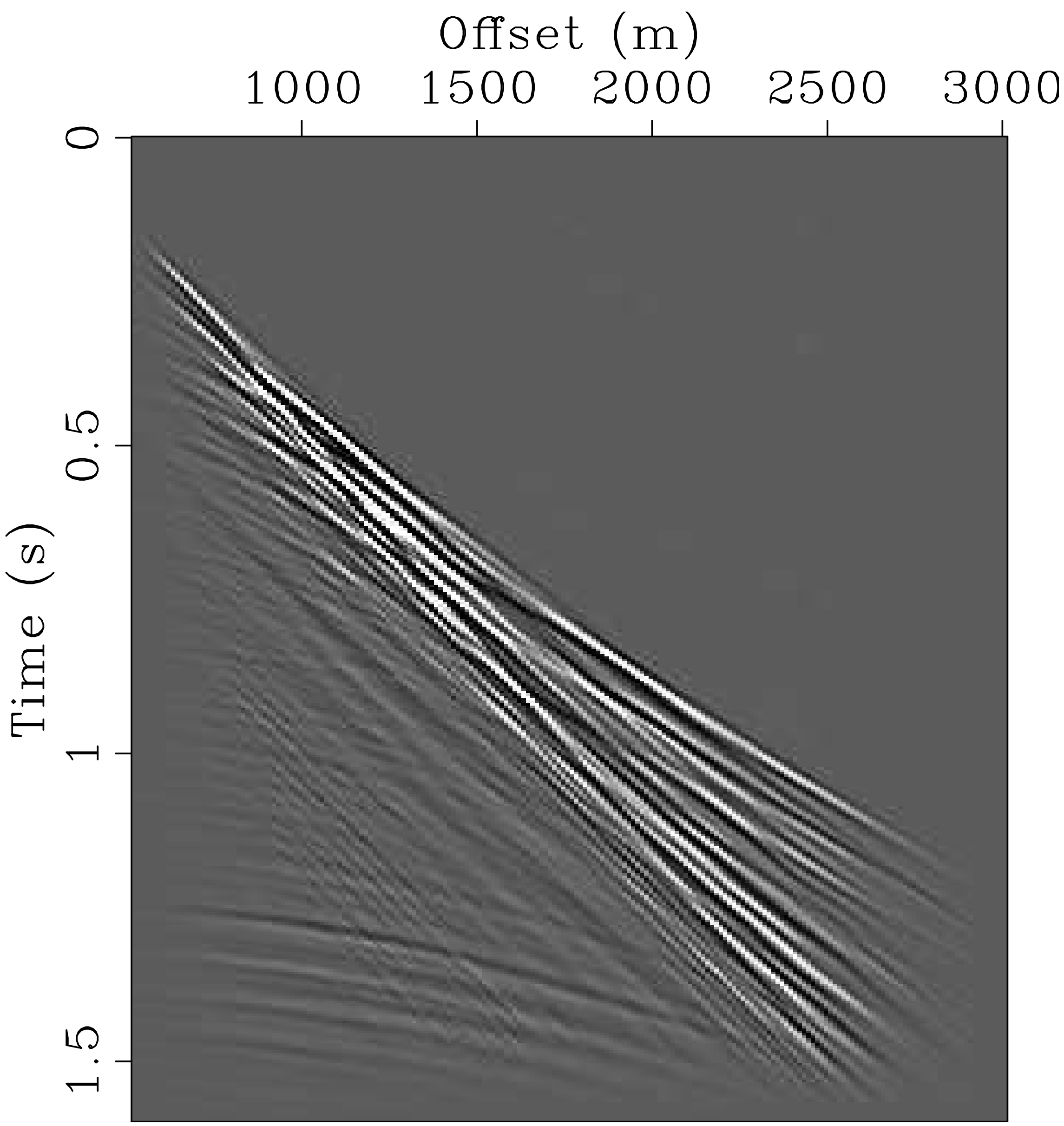
Intelligent reduction of acquisition *costs* via *randomized*

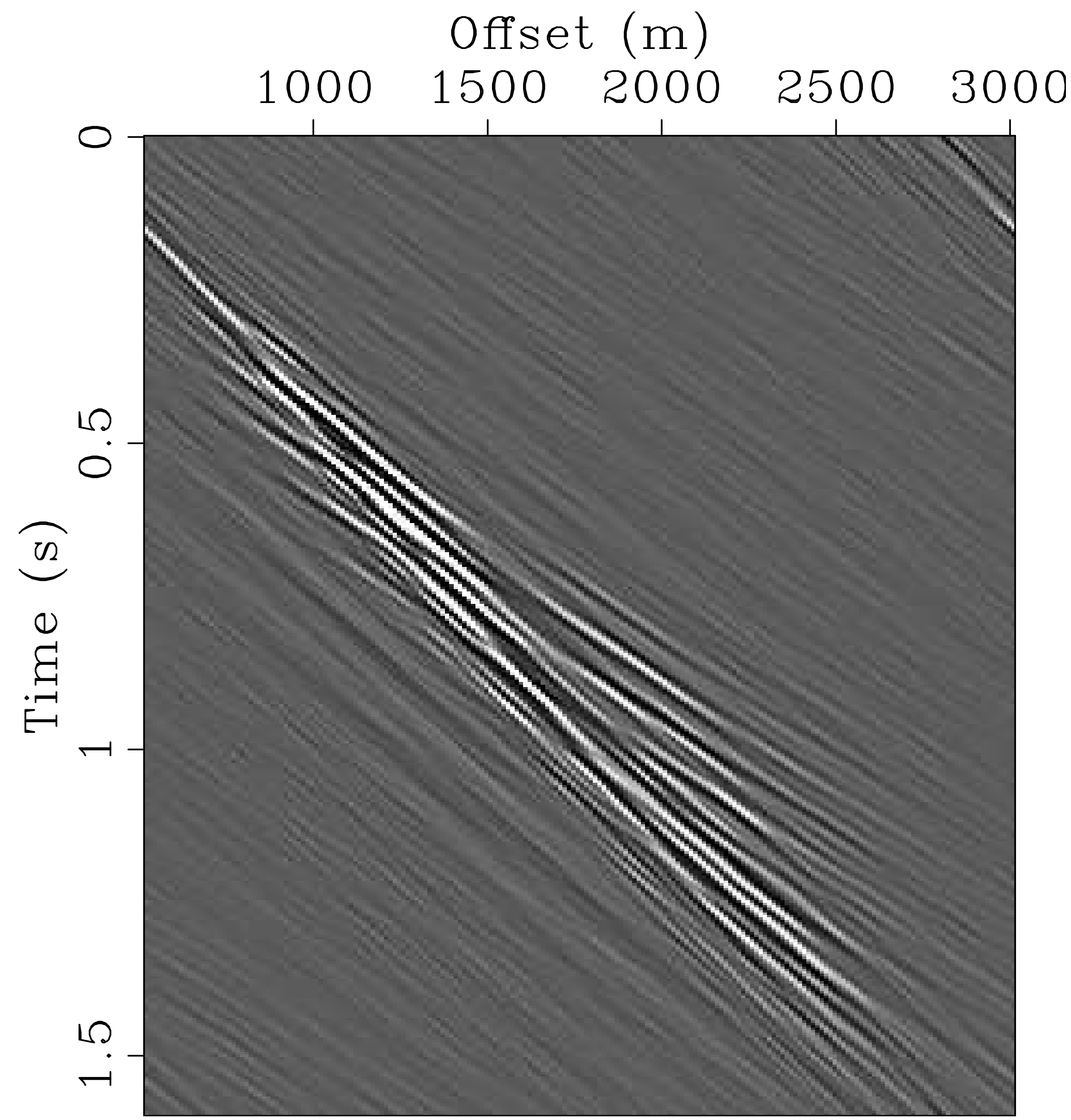
- *jitter* & *coil* acquisition
  [Hennenfent & FJH, 08-‘; Moldoveanu ’10-]
- amplitude/phase-encoded *simultaneous* ‘land’ acquisition
  [Krohn et. al., 2006]
- *ditter* continuous ‘marine’ acquisition
  [Beasley, ’98, Berkhout, ’08, Blacquiere, ’10; Abma, ’10, Mansour & FJH, ’11]

CS design principles

- sparsifying transform
  - typically **localized** in the time-space domain to handle the complexity of seismic data
- advantageous coarse randomized sampling
  - generates incoherent random undersampling “noise” in the sparsifying domain
- sparsity-promoting solver
  - requires few matrix-vector multiplications

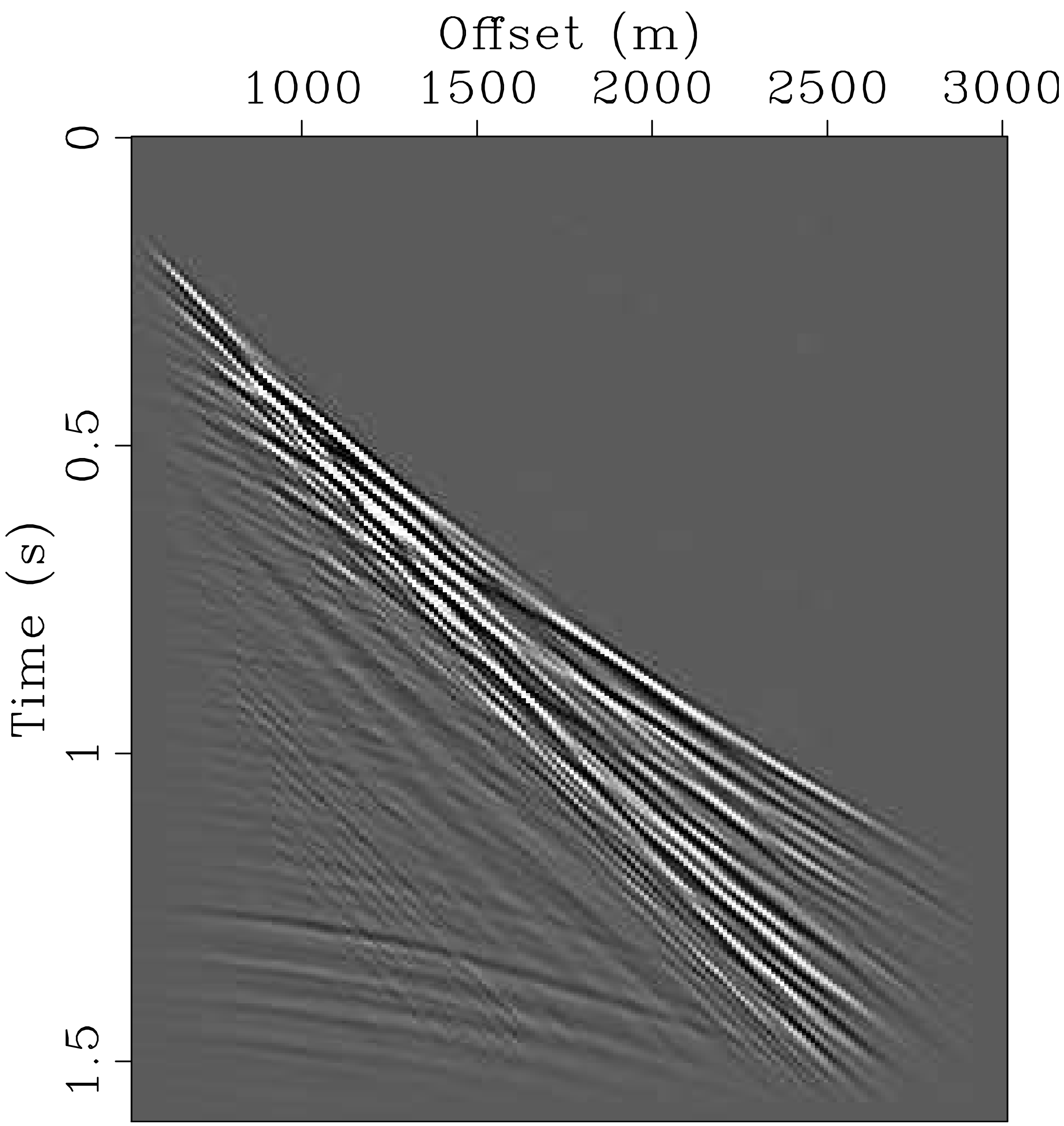
Fourier reconstruction

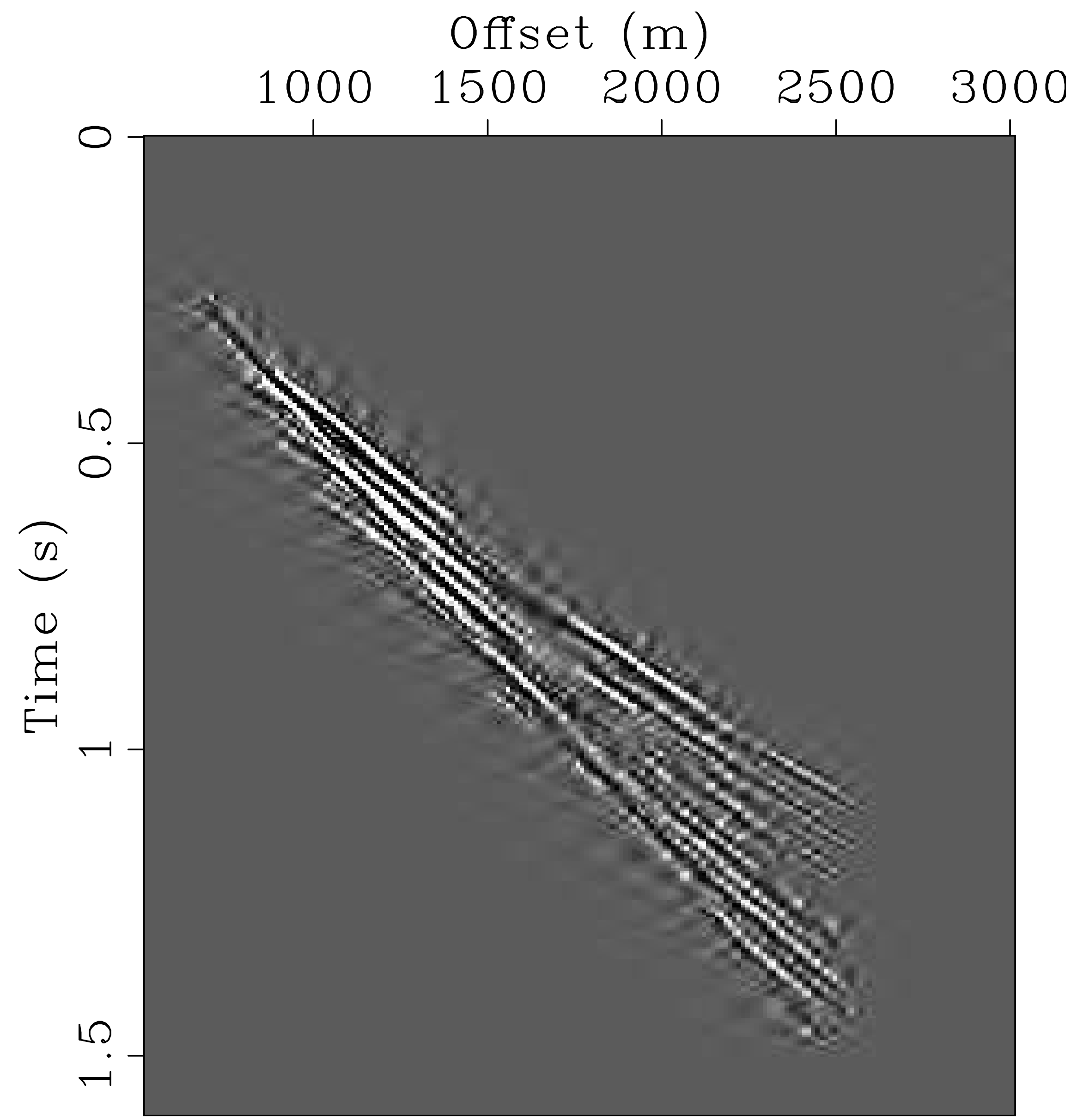




1 % of coefficients

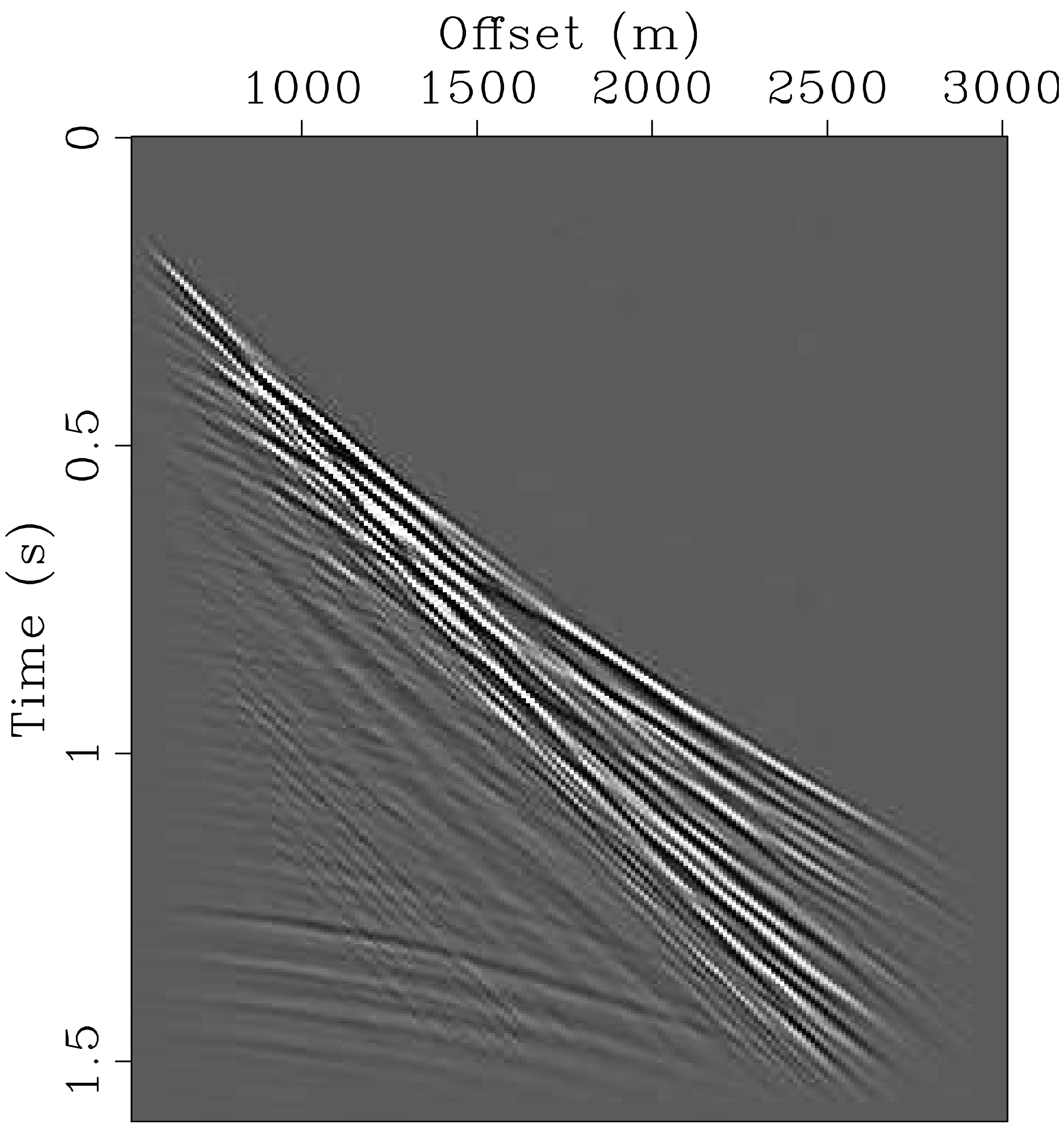
Wavelet reconstruction

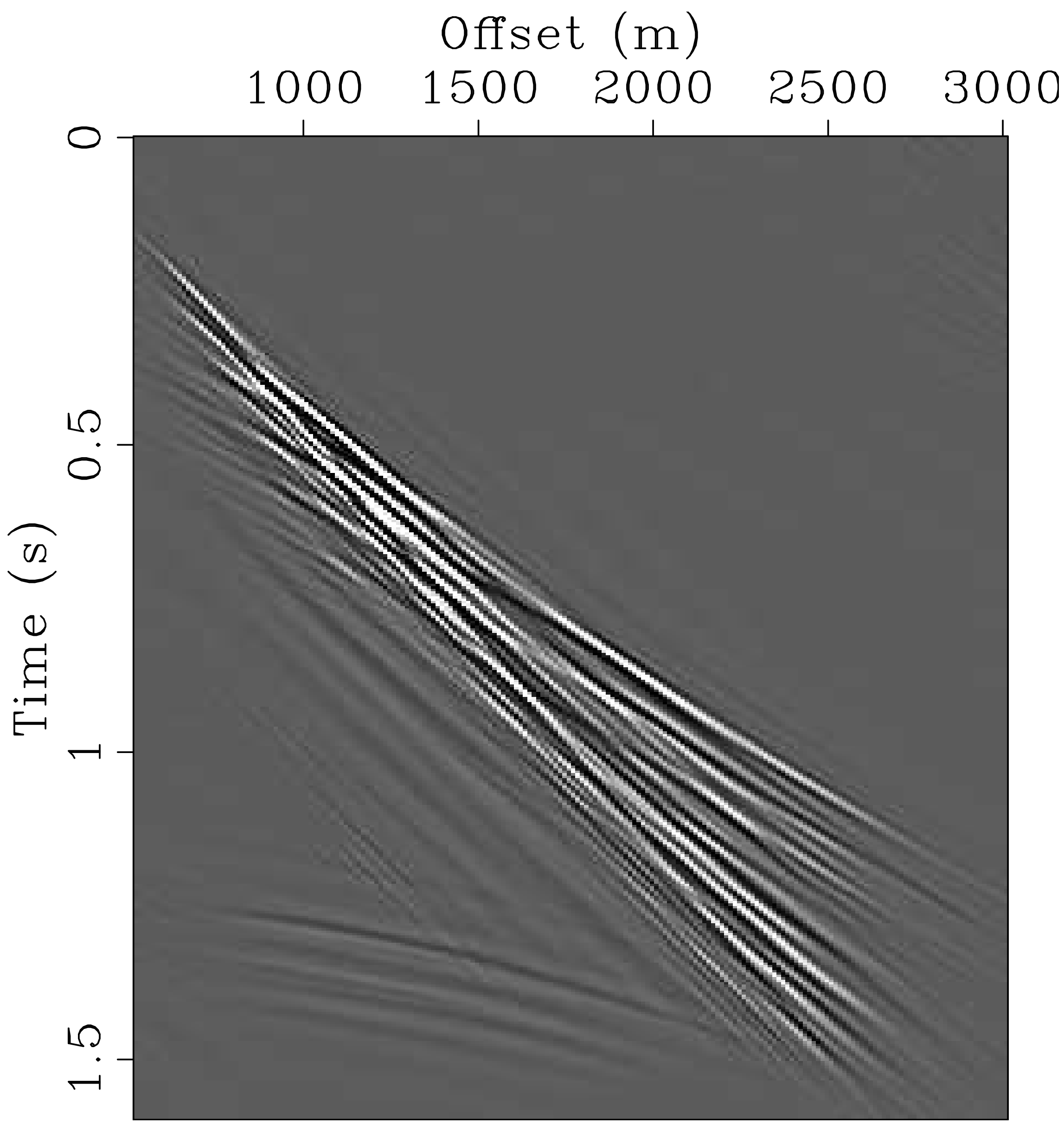




1 % of coefficients

Curvelet reconstruction



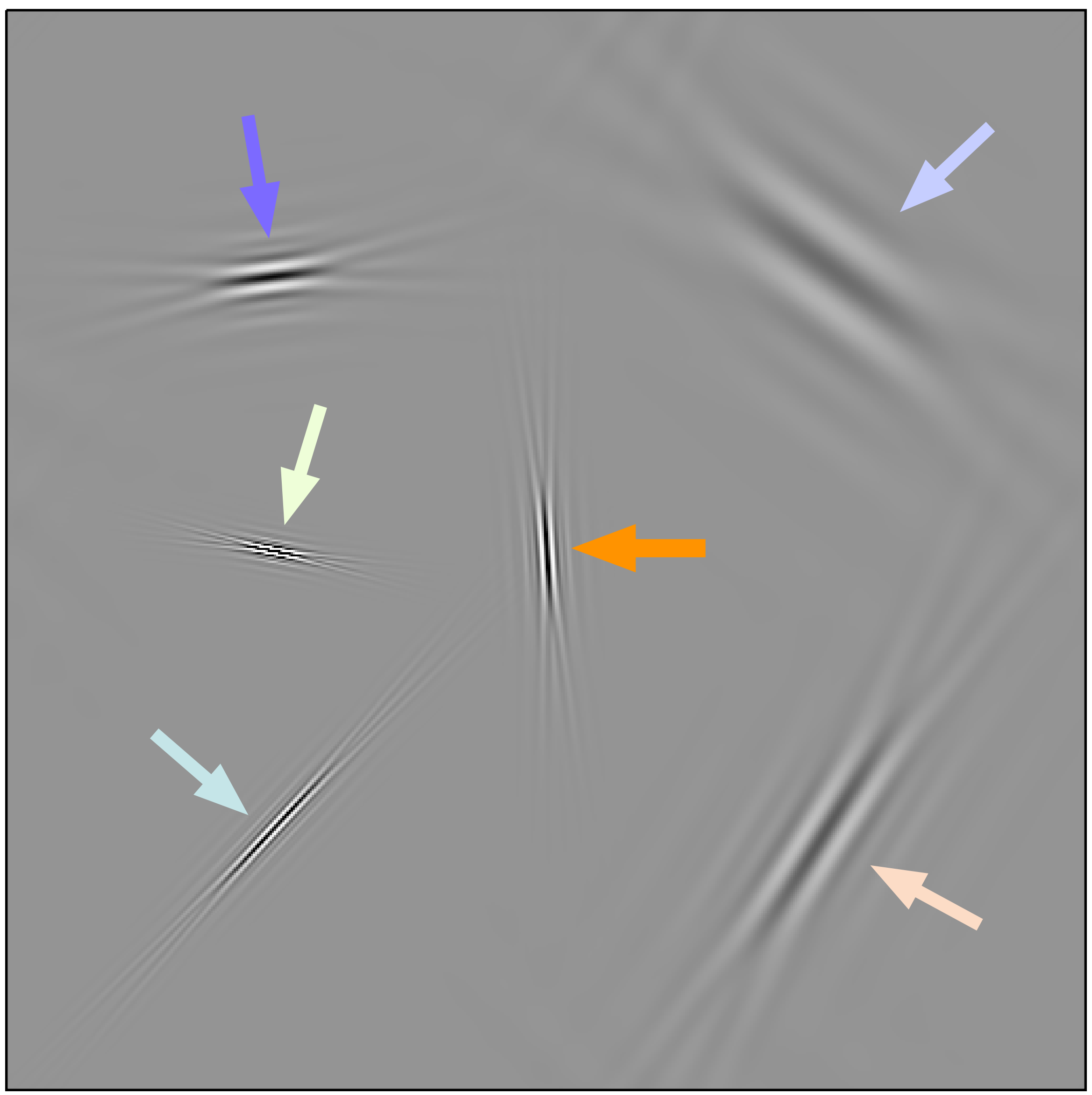


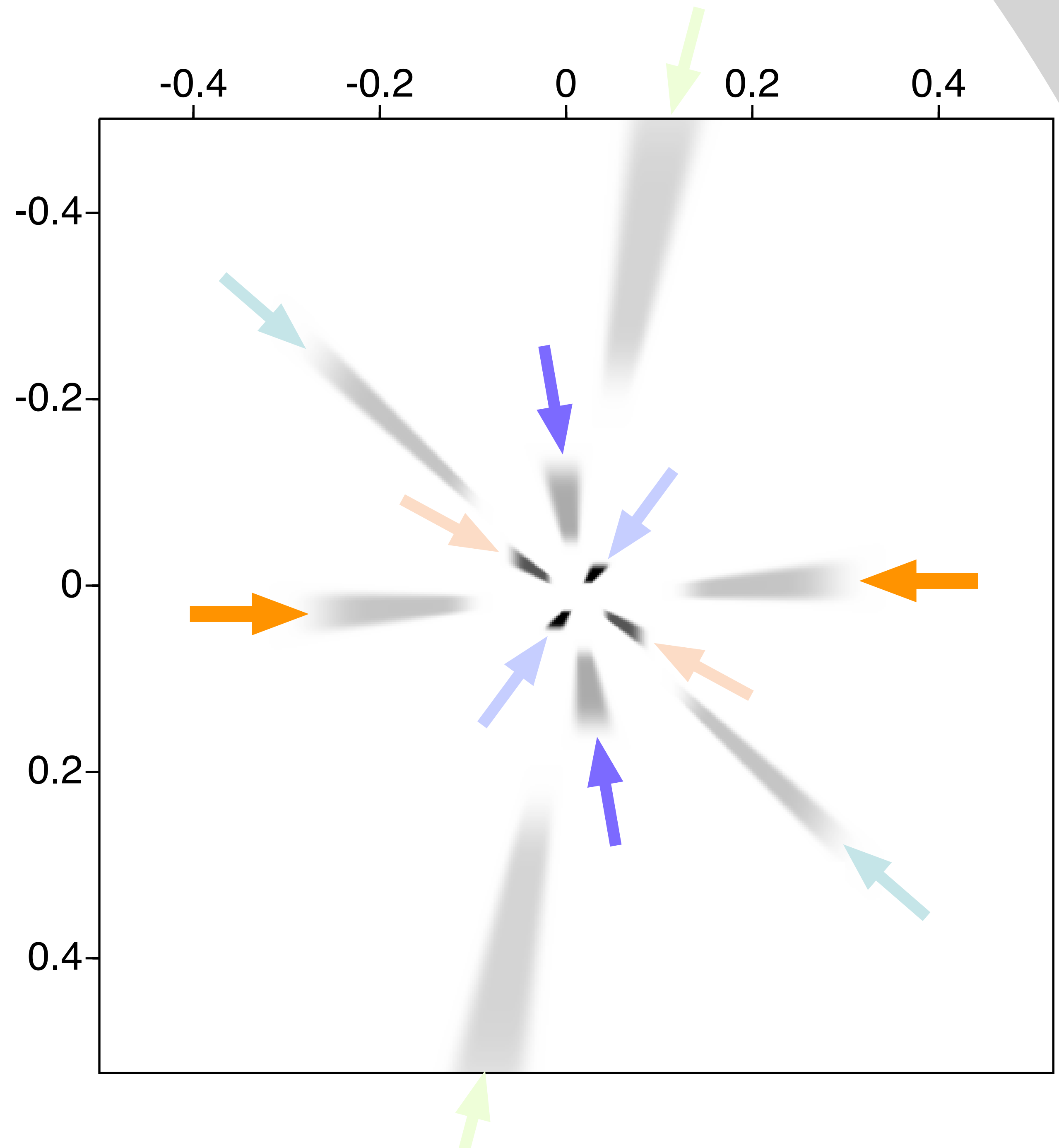
1 % of coefficients

SLIM

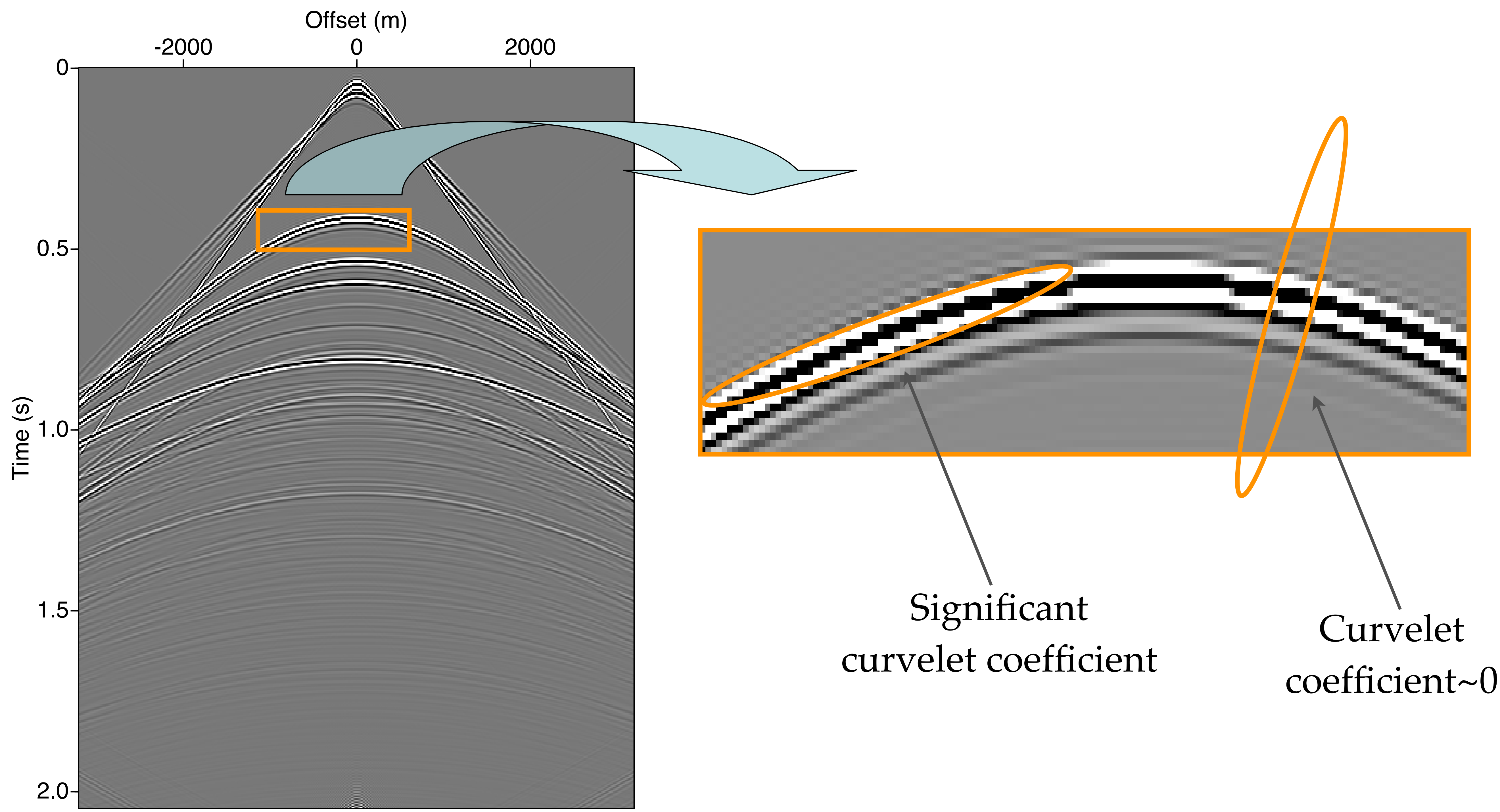
[Demanet et. al., ‘06]

Curvelets





Detect the wavefronts



Empirical performance analysis

Selection of the appropriate sparsifying transform

- nonlinear approximation error

$$\mathrm{SNR}(\rho) = -20 \log \frac{\|\mathbf{f} - \mathbf{f}_\rho\|}{\|\mathbf{f}\|} \quad \text{with} \quad \rho = k/P$$

- recovery error

$$\mathrm{SNR}(\delta) = -20 \log \frac{\|\mathbf{f} - \tilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|} \quad \text{with} \quad \delta = n/N$$

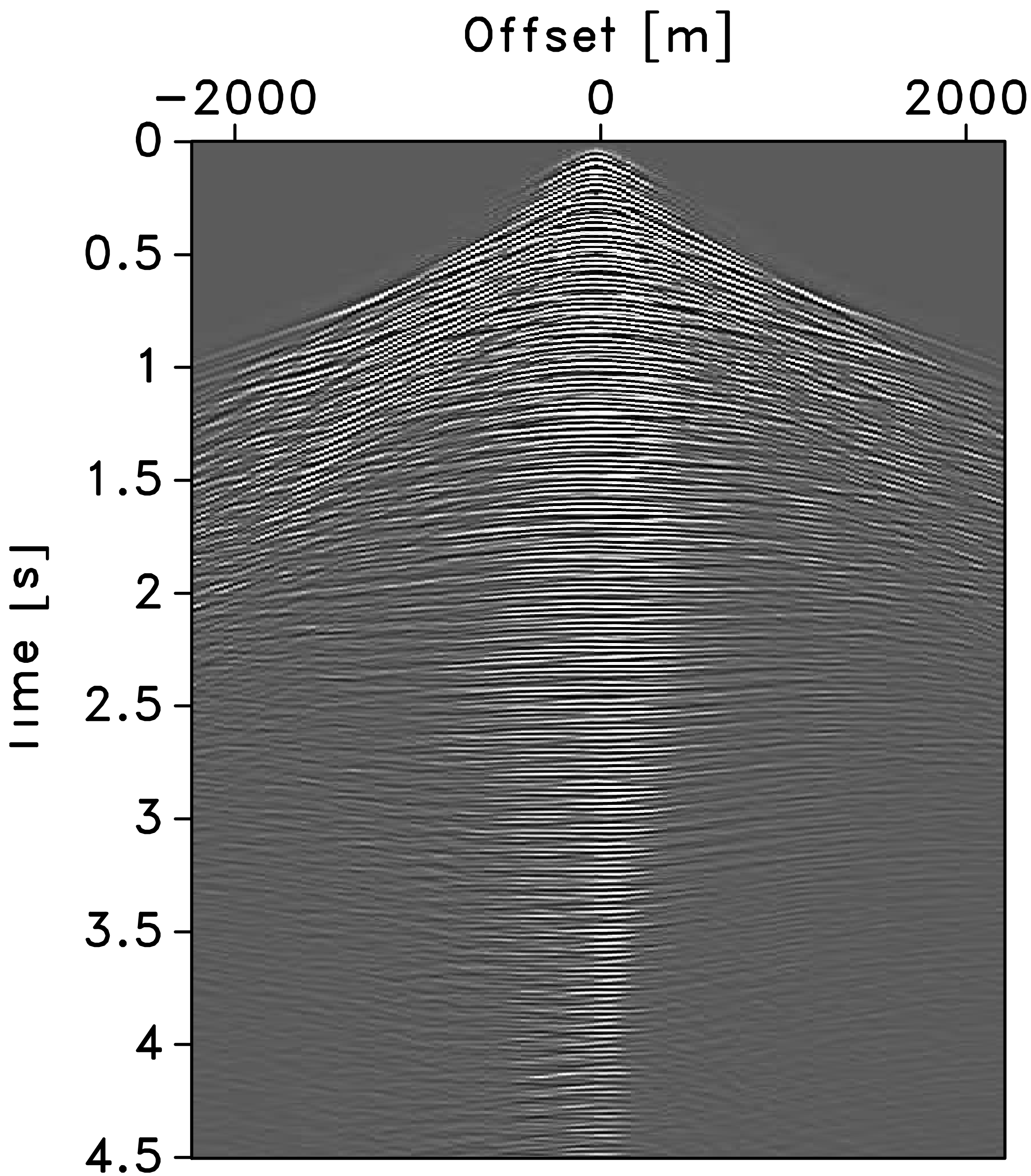
- oversampling ratio

$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\mathrm{SNR}}(\delta) \leq \mathrm{SNR}(\tilde{\rho})\}$$

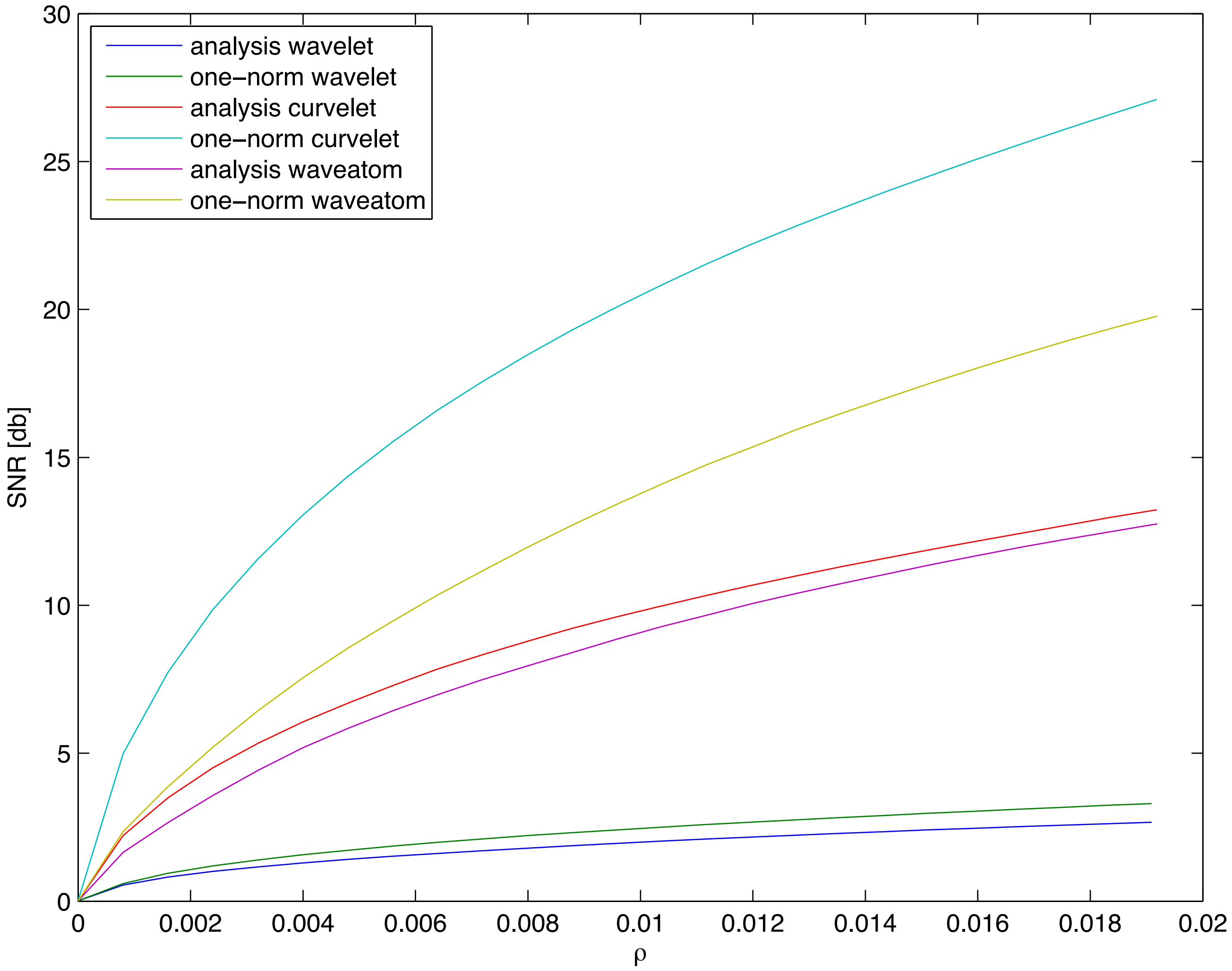
[FJH, '10]

Nonlinear approximation error

common receiver gather



recovery error



[FJH, '10]

CS design principles

☑ sparsifying transform

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☐ advantageous coarse randomized sampling

- generates incoherent random undersampling “noise” in the sparsifying domain

☐ sparsity-promoting solver

- requires few matrix-vector multiplications

Different sampling schemes

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

restriction matrix — measurement matrix — sparsity matrix

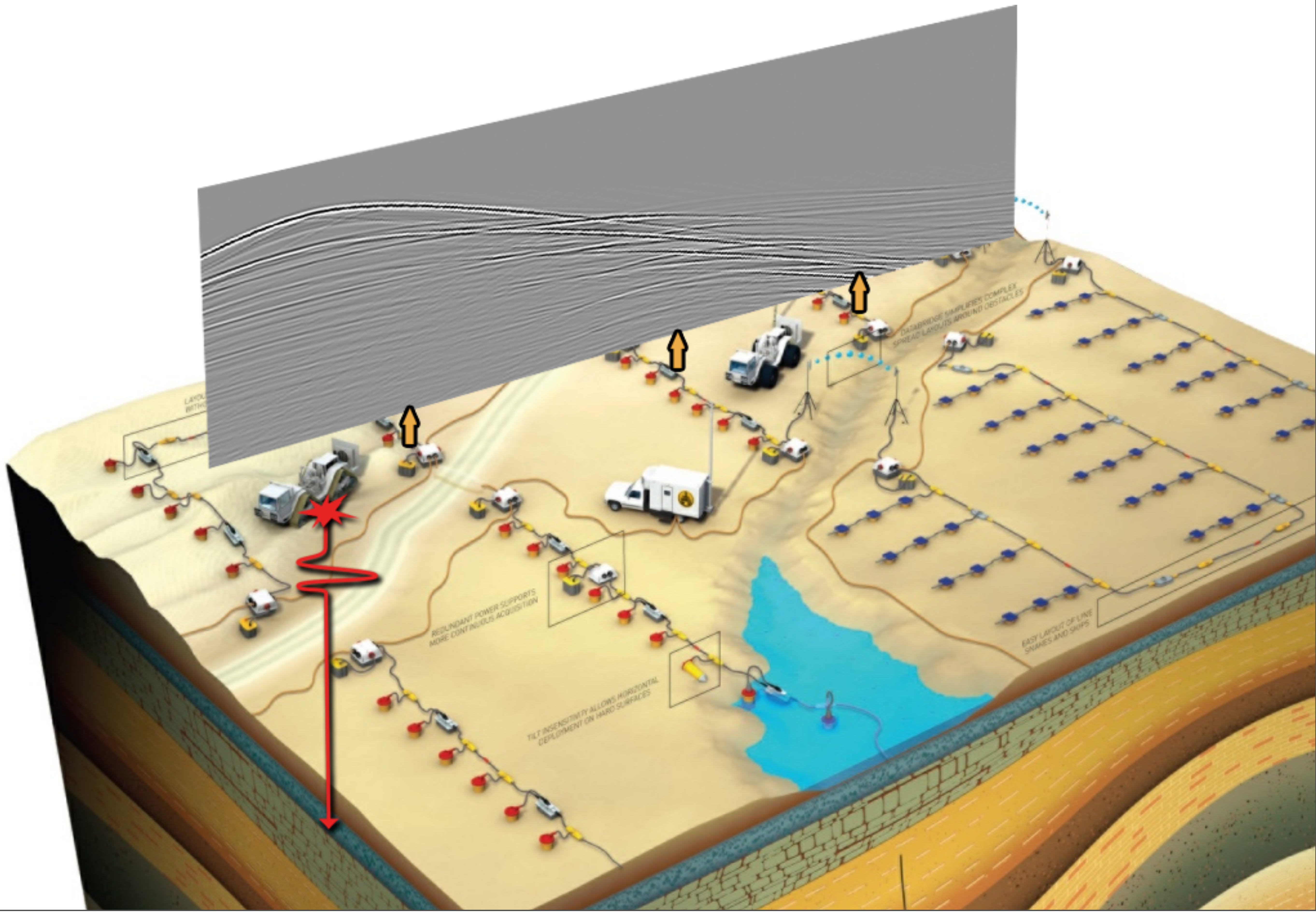
with

$$\begin{aligned}\mathbf{R} &= \left(\mathbf{R}^{\Sigma} \otimes \mathbf{I}^{\tau}\right) \\ \mathbf{M} &= \left(\mathbf{I}^{\Sigma} \otimes \mathbf{I}^{\tau}\right) \quad \text{or} \quad \left(\mathbf{G}^{\Sigma} \otimes \mathbf{I}^{\rho}\right) \\ \mathbf{S}^{*} &= \mathbf{C}_2^{*}\end{aligned}$$

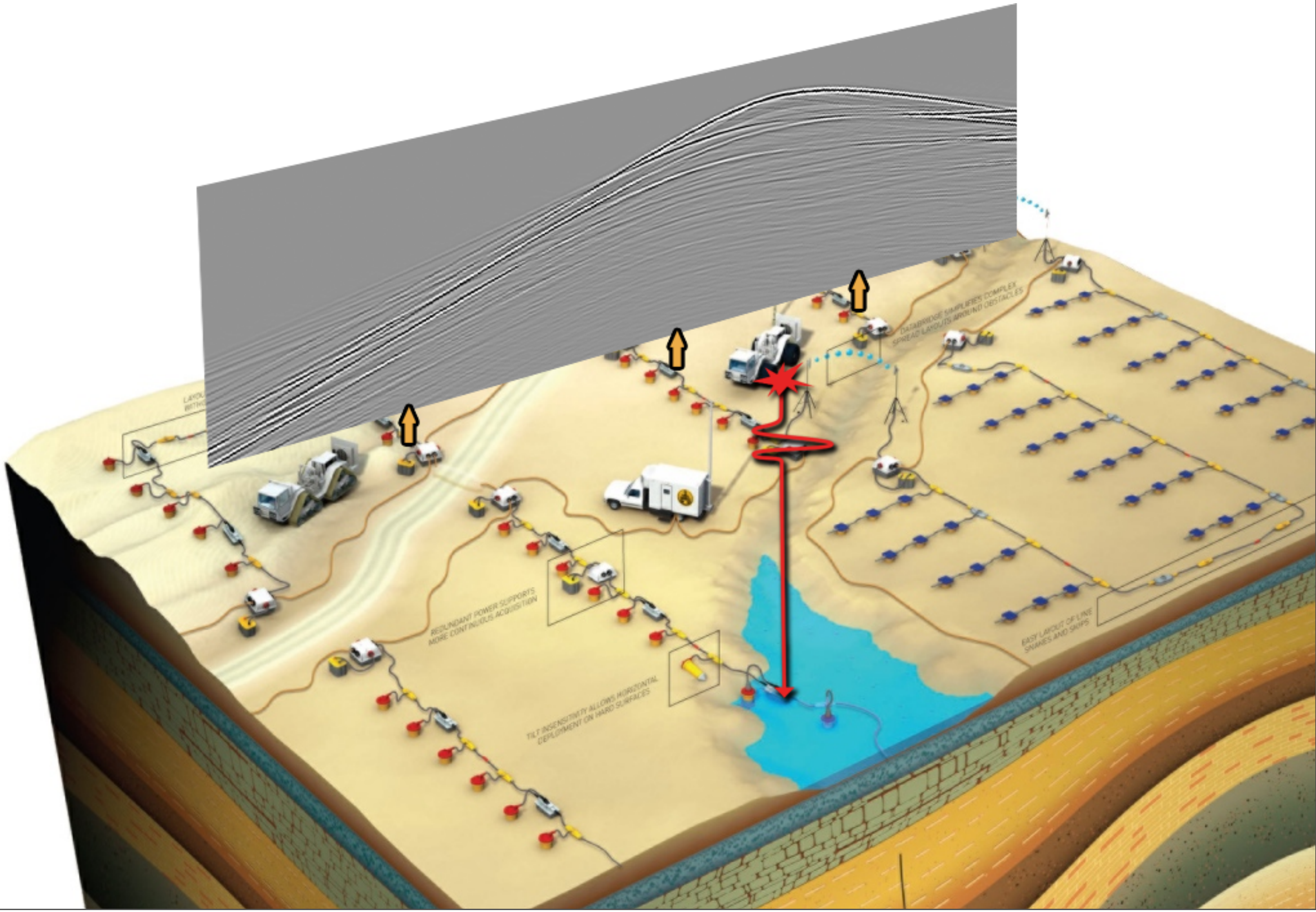
Solve

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

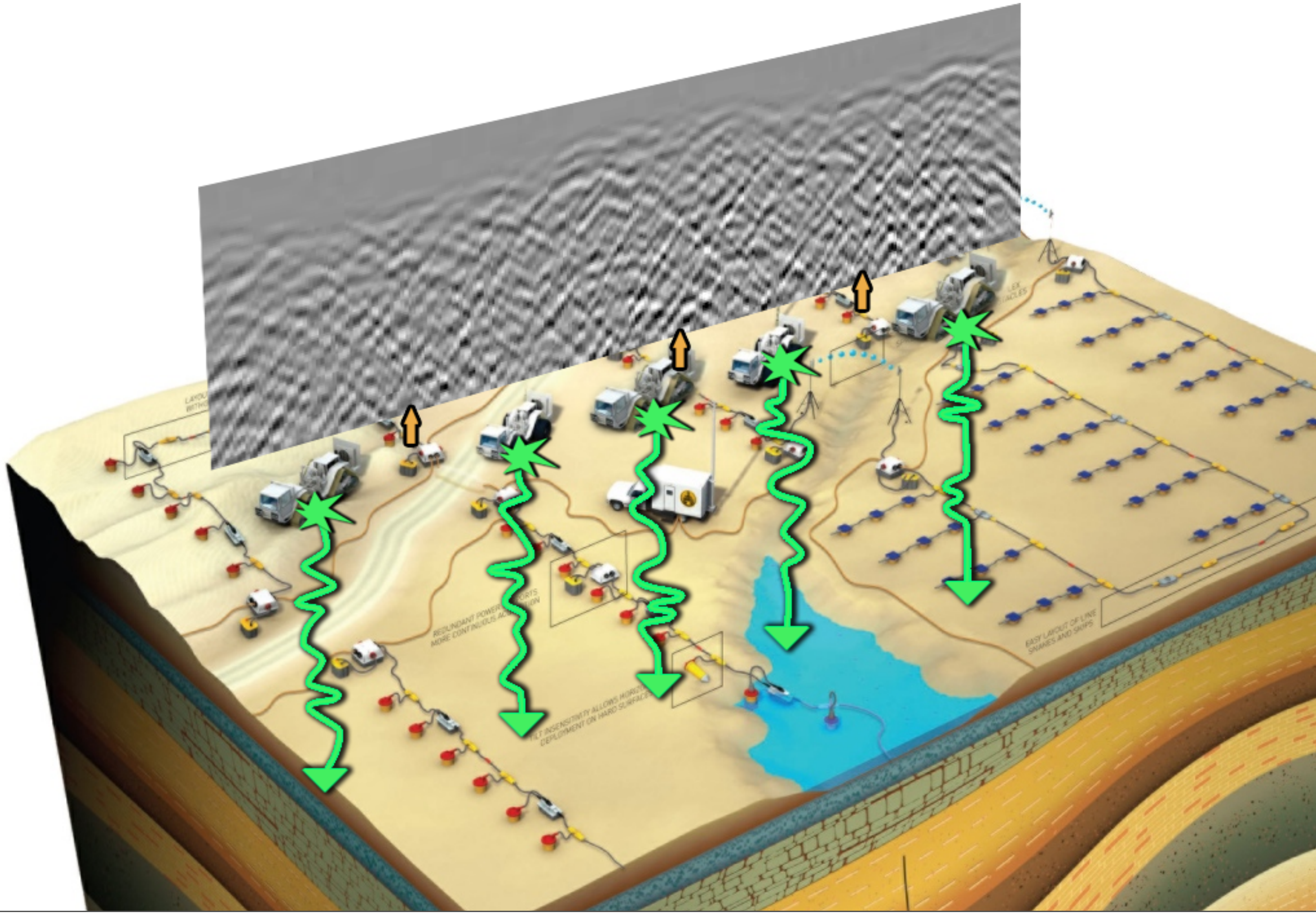
Individual shots



Individual shots



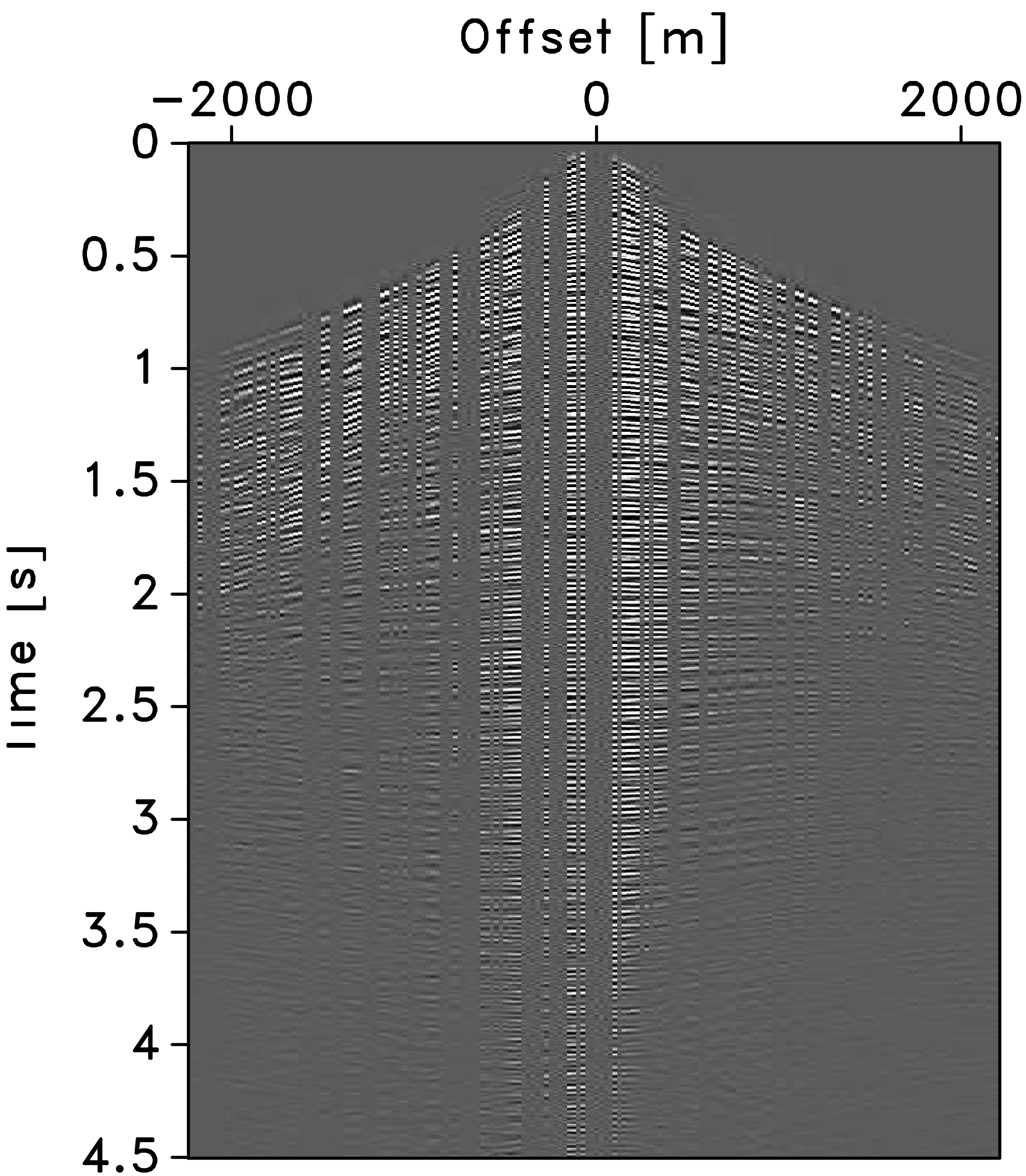
Simultaneous & incoherent sources



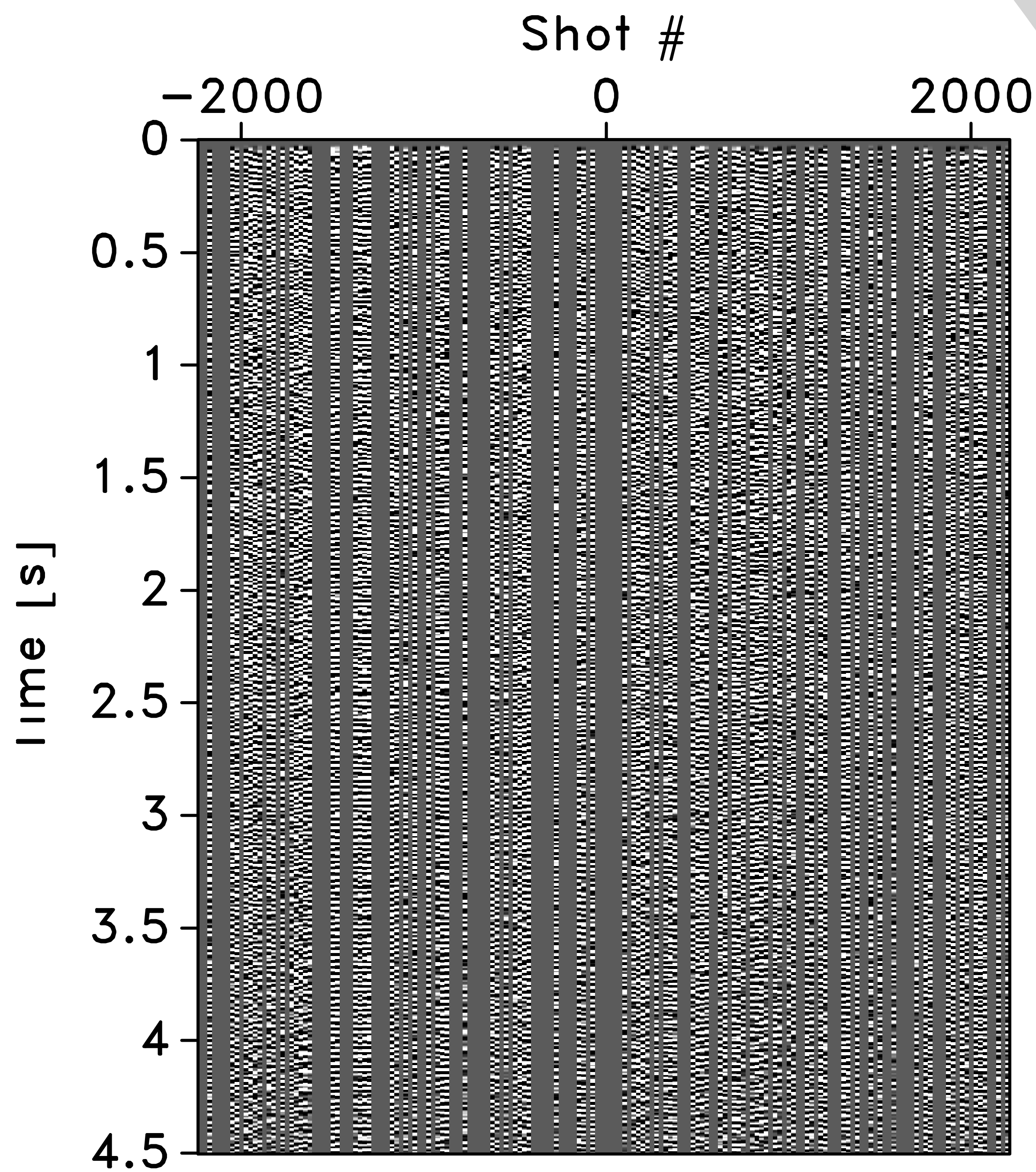
Data

SLIM

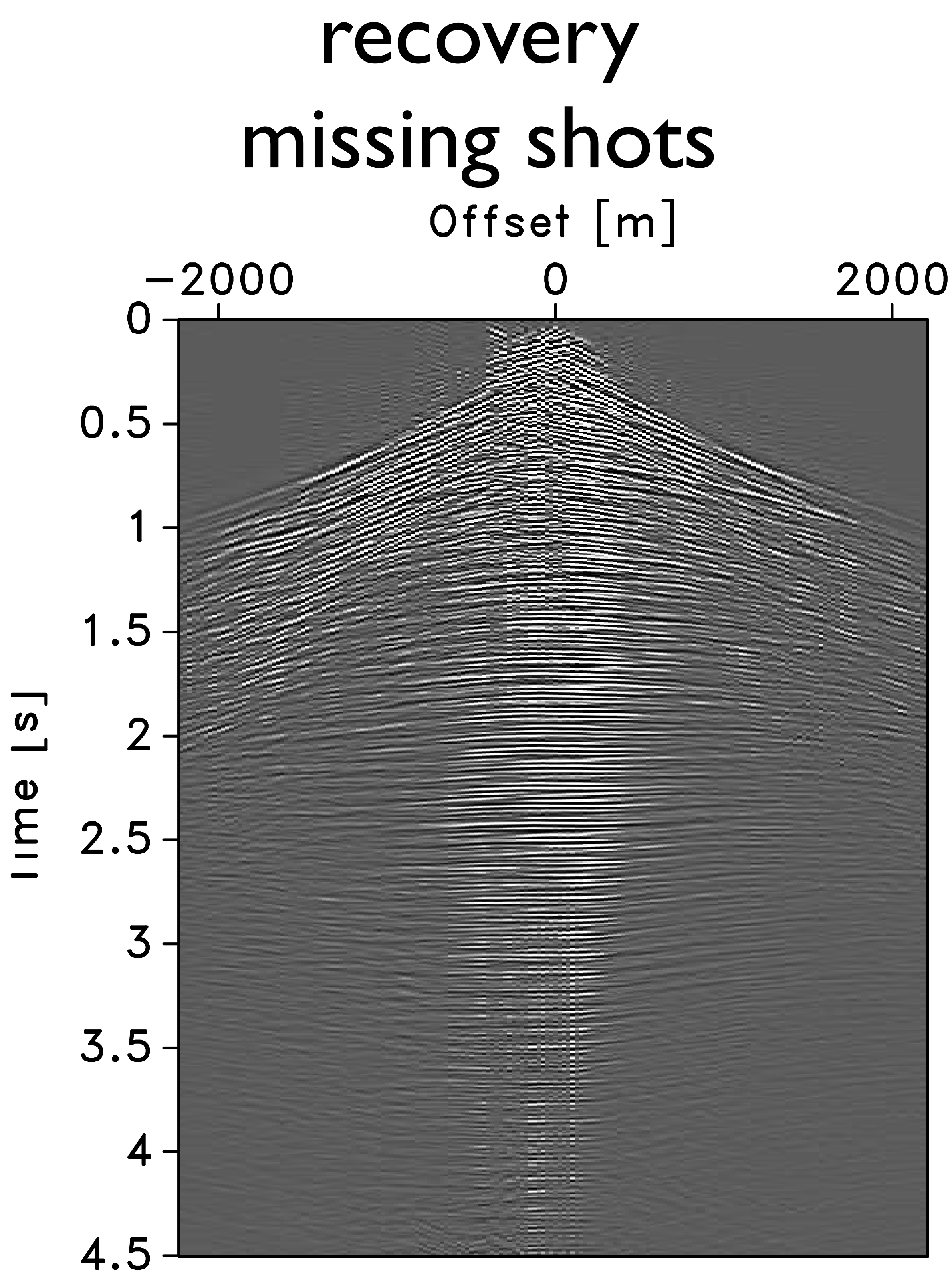
missing shots

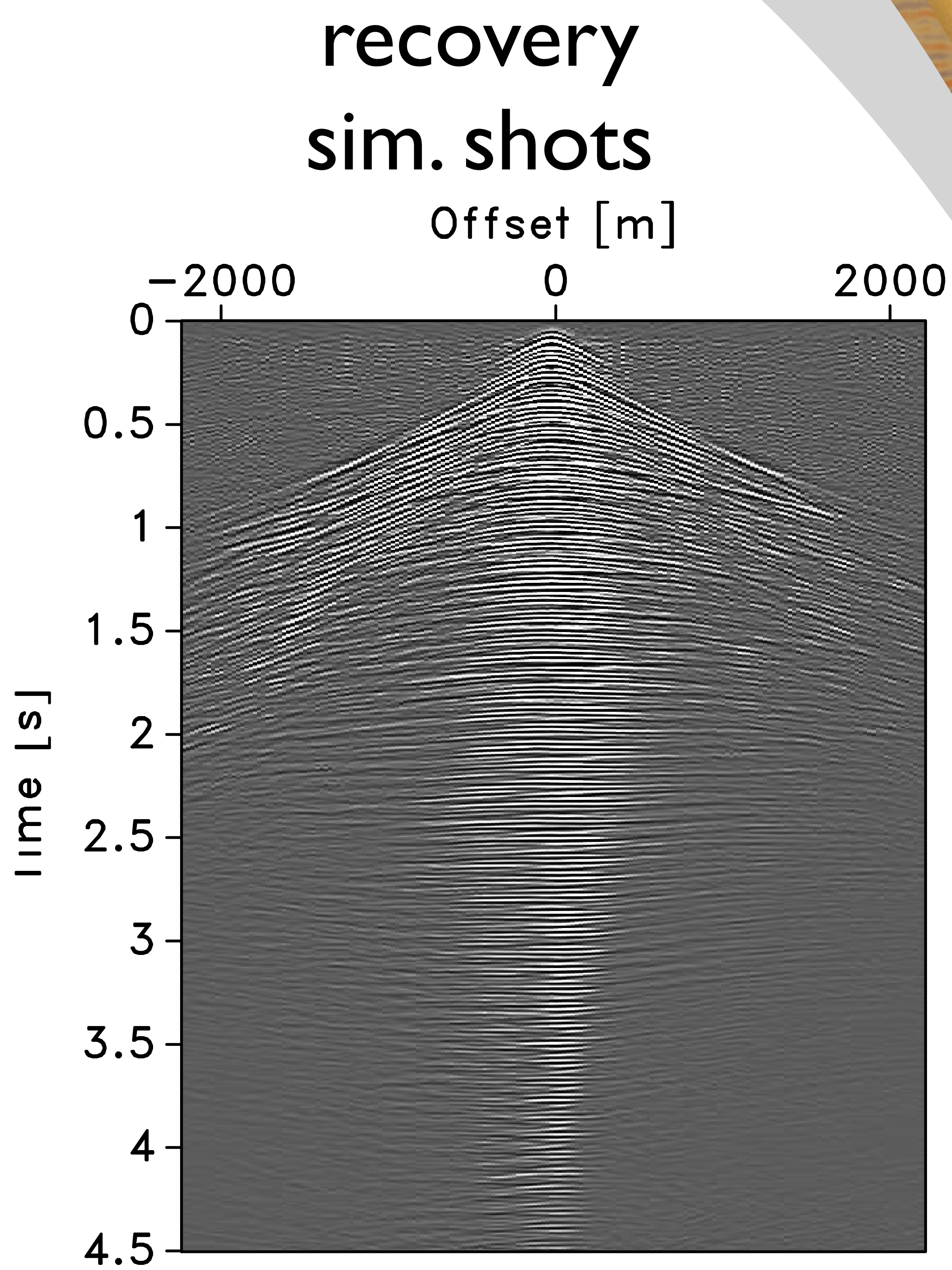


sim. shots

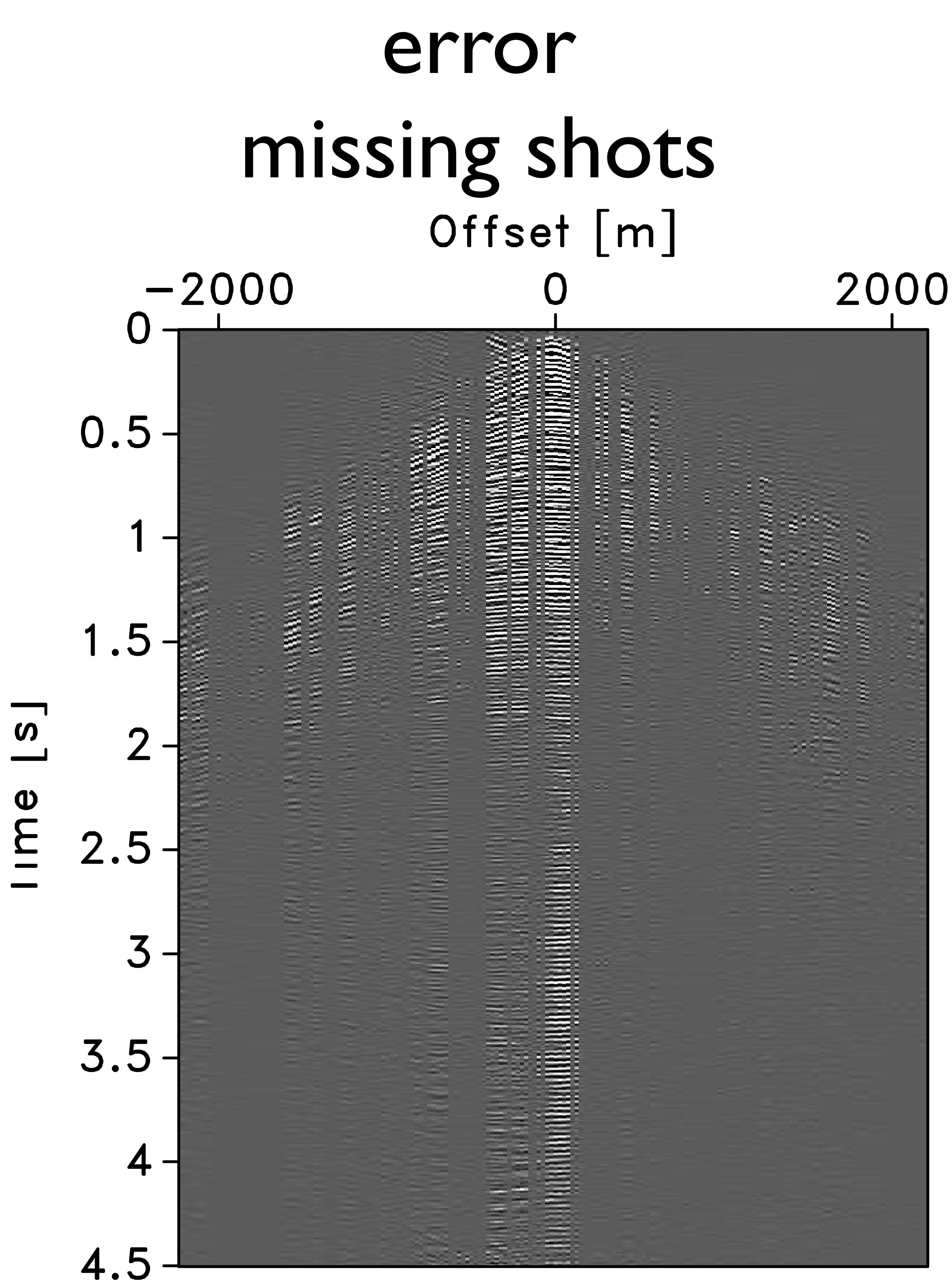


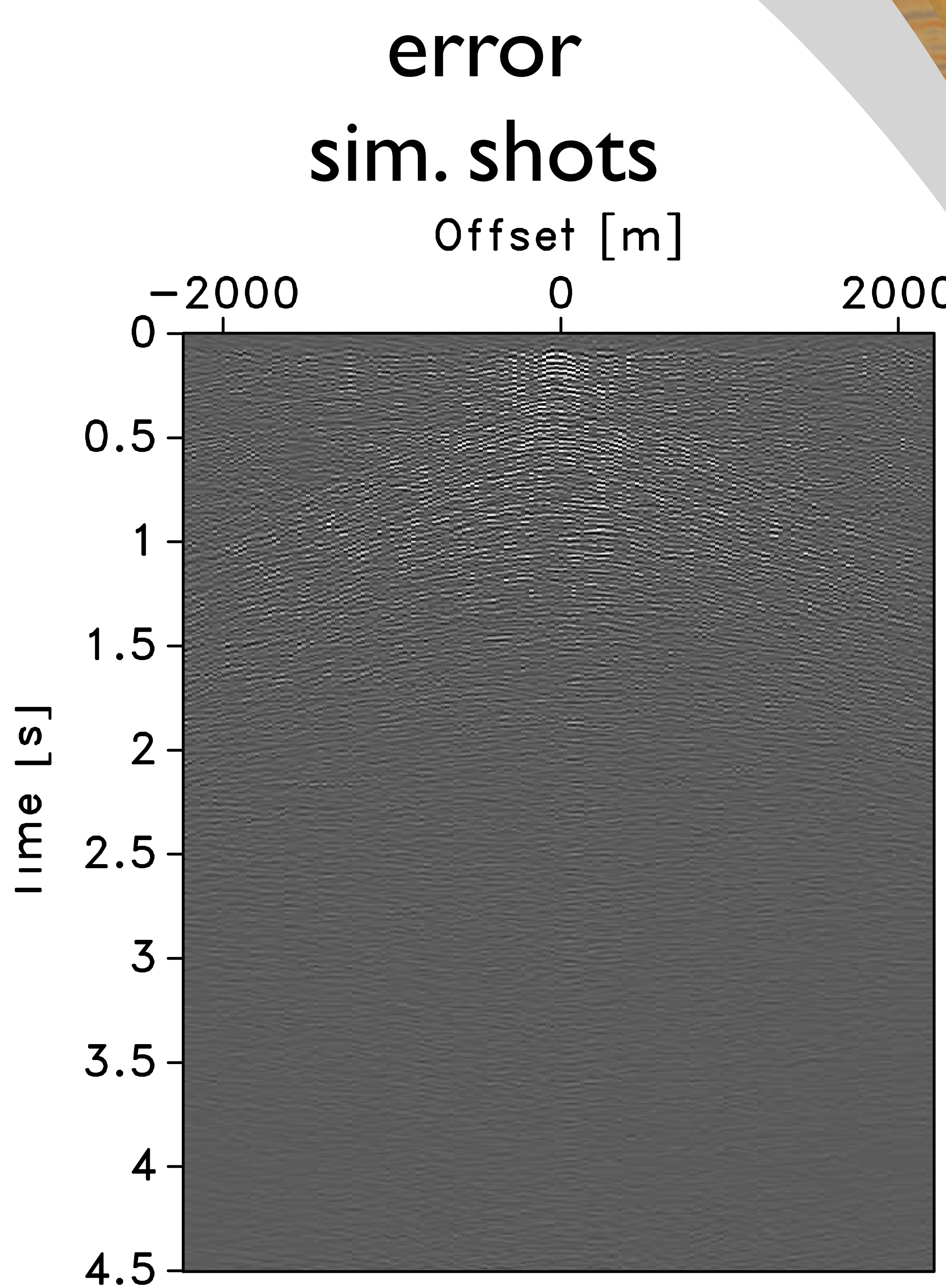
Sparse recovery





Sparse recovery error





SLIM

Empirical performance analysis

Selection of the appropriate sparsifying transform

- nonlinear approximation error

$$\mathrm{SNR}(\rho) = -20\log\frac{\|\mathbf{f}-\mathbf{f}_\rho\|}{\|\mathbf{f}\|} \quad \text{with} \quad \rho = k/P$$

➡ recovery error

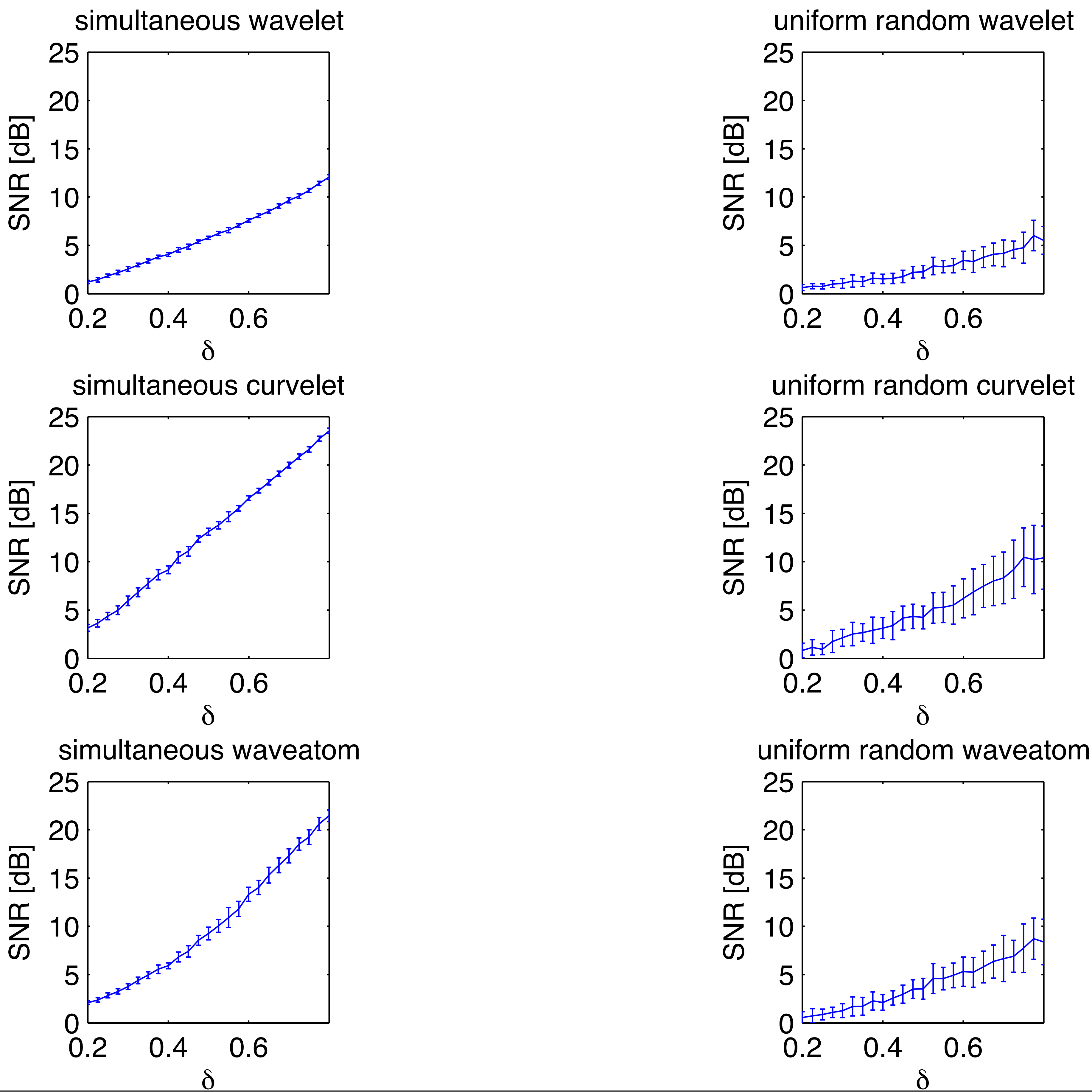
$$\mathrm{SNR}(\delta) = -20\log\frac{\|\mathbf{f}-\tilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|} \quad \text{with} \quad \delta = n/N$$

- oversampling ratio

$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\mathrm{SNR}}(\delta) \leq \mathrm{SNR}(\tilde{\rho})\}$$

[FJH, '10]

Multiple experiments



CS design principles

☑ sparsifying transform

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☑ advantageous coarse randomized sampling

- generates incoherent random undersampling "noise" in the sparsifying domain
- does not create coherent interferences in simultaneous acquisition
- does not create large gaps for measurement in the physical domain

☐ sparsity-promoting solver

- requires few matrix-vector multiplications

Reality check

"When a traveler reaches a fork in the road, the l_1-norm tells him to take either one way or the other, but the l_2-norm instructs him to head off into the bushes."

John F. Claerbout and Francis Muir, 1973

Approaches

- quadratic programming [many references!]

$$\mathrm{QP}_\lambda : \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

- basis pursuit denoise [Chen et al.'95]

$$\mathrm{BP}_\sigma : \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma$$

- LASSO [Tibshirani'96]

$$\mathrm{LS}_\tau : \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau$$

Approaches

- quadratic programming [many references!]

$$\mathrm{QP}_\lambda: \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

- basis pursuit denoise [Chen et al.'95]

$$\mathrm{BP}_\sigma: \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \le \sigma$$

- LASSO [Tibshirani'96]

$$\mathrm{LS}_\tau: \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \le \tau$$

Approaches

- quadratic programming [many references!]

$$\mathrm{QP}_\lambda: \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \boxed{\lambda}\|\mathbf{x}\|_1$$

- basis pursuit denoise [Chen et al.'95]

$$\mathrm{BP}_\sigma: \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \leq \boxed{\sigma}$$

- LASSO [Tibshirani'96]

$$\mathrm{LS}_\tau: \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau$$

Approaches

- quadratic programming [many references!]

$$\mathrm{QP}_\lambda : \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \boxed{\lambda}\|\mathbf{x}\|_1$$

- basis pursuit denoise [Chen et al.'95]

$$\mathrm{BP}_\sigma : \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \boxed{\sigma}$$

- LASSO [Tibshirani'96]

$$\mathrm{LS}_\tau : \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \boxed{\tau}$$

Approaches

- quadratic programming [many references!]

$$\mathrm{QP}_\lambda : \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

- basis pursuit denoise [Chen et al.'95]

$$\mathrm{BP}_\sigma : \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma$$

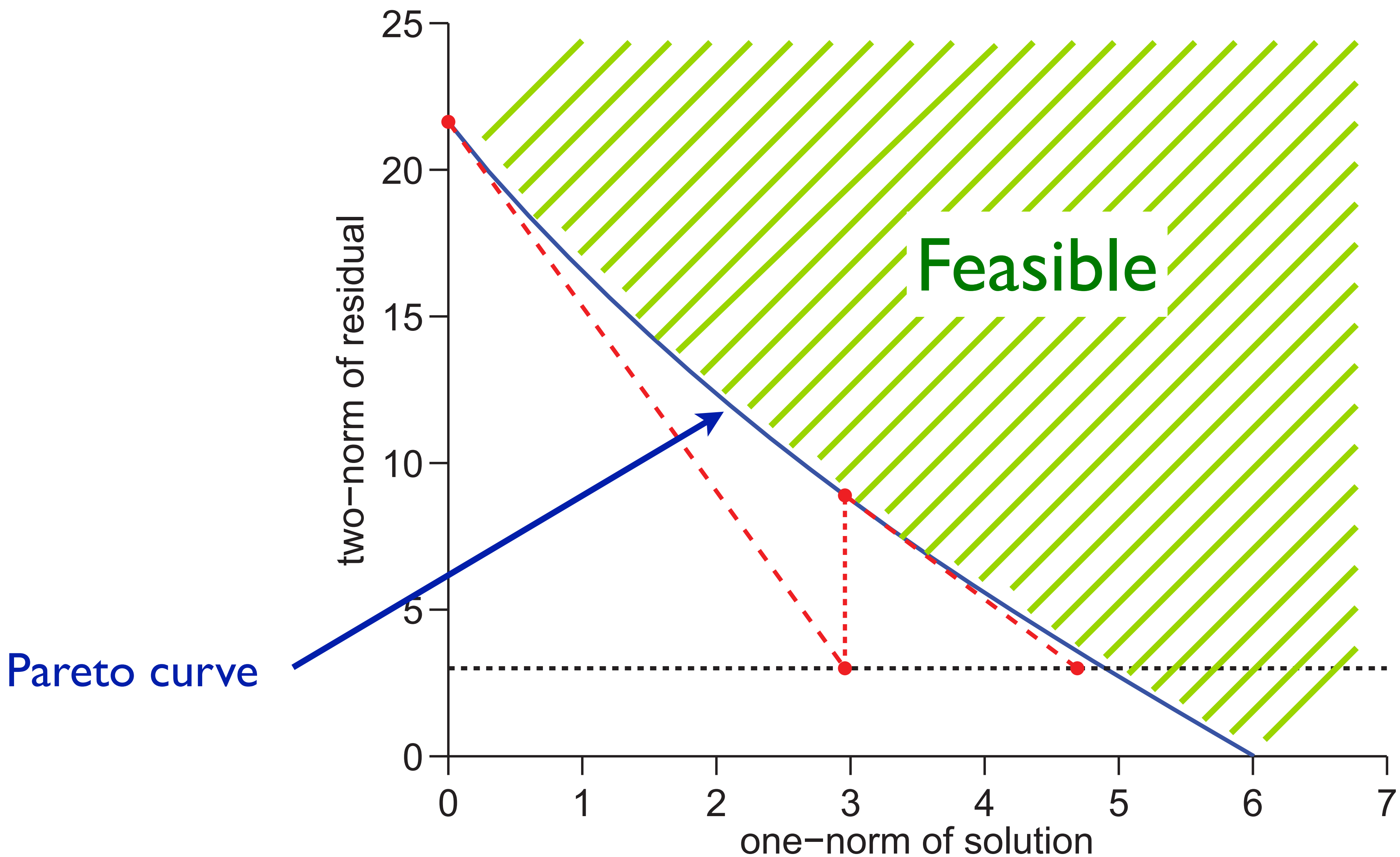
- LASSO [Tibshirani'96]

$$\mathrm{LS}_\tau : \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau$$

Pareto curve

$$\begin{array}{ll}\text{minimize} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2 \leq \sigma\end{array}$$

Look at the solution space and the line of optimal solutions (Pareto curve)

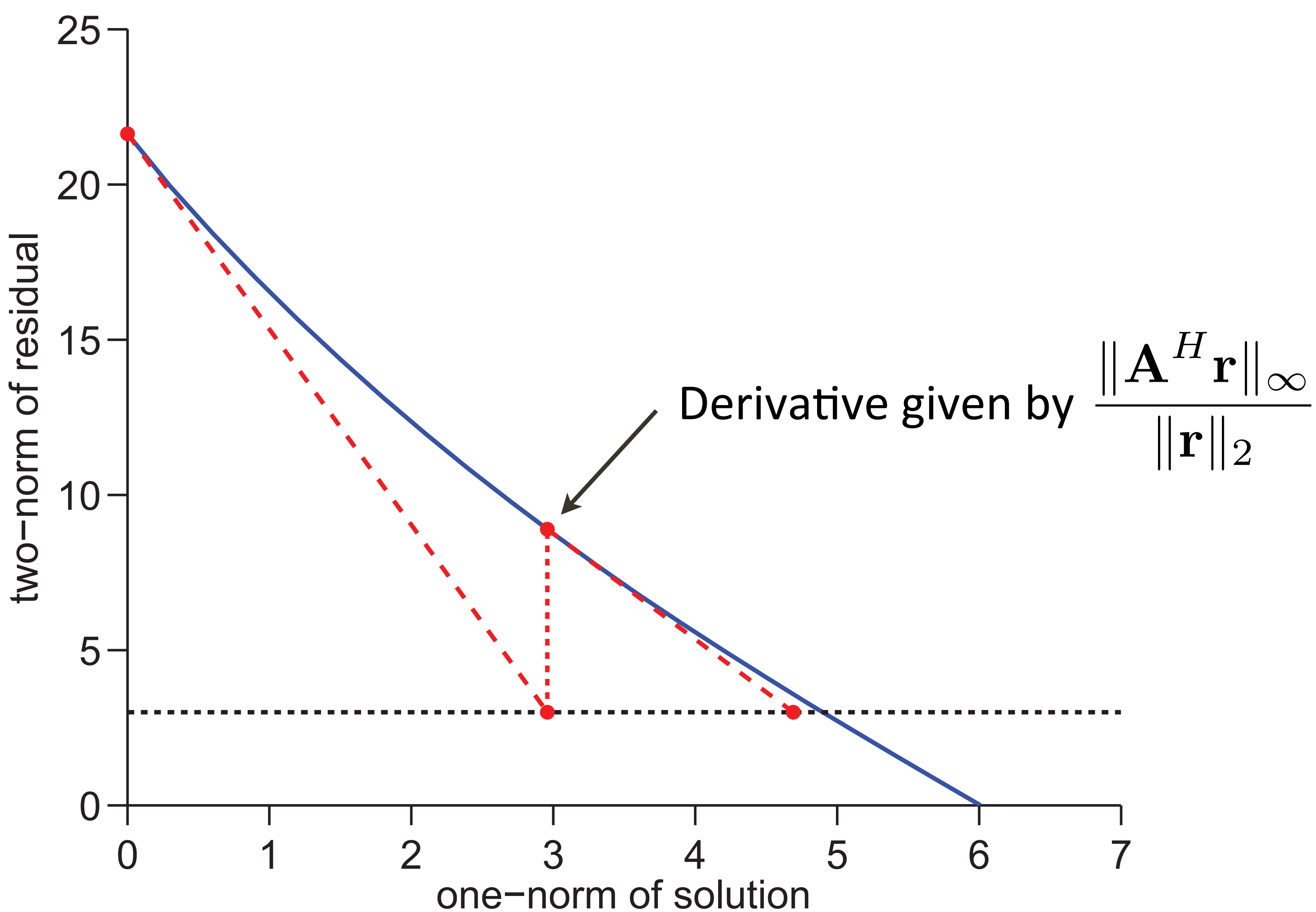


(van den Berg, Friedlander, 2008)

Pareto curve

$$\text{minimize} \quad \|x\|_1$$
$$\text{subject to} \quad \|Ax - b\|_2 \leq \sigma$$

Look at the solution space and the line of optimal solutions (Pareto curve)

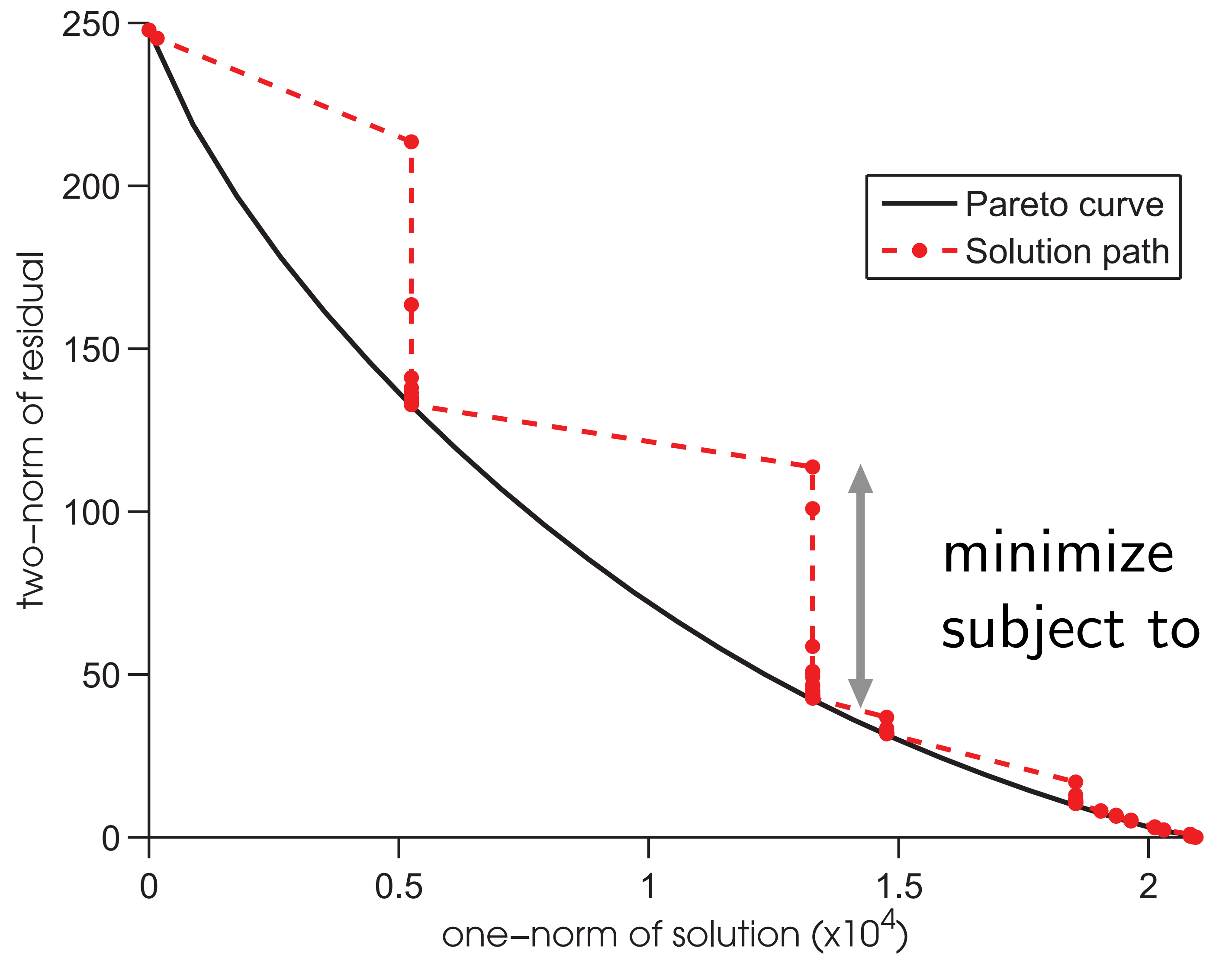


[van den Berg & Friedlander, '08]
[Hennenfent, FJH, et. al, '08]

Pareto curve

minimize $\|x\|_1$
subject to $\|Ax - b\|_2 \leq \sigma$

Look at the solution space and the line of optimal solutions (Pareto curve)



minimize $\|Ax - b\|_2$
subject to $\|x\|_1 \leq \tau$

solve with SPG
(spectral projected gradients)

CS design principles

☑ sparsifying transform

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☑ advantageous coarse randomized sampling

- generates incoherent random undersampling "noise" in the sparsifying domain
- does not create coherent interferences in simultaneous acquisition
- does not create large gaps for measurement in the physical domain

☑ sparsity-promoting solver

- requires few matrix-vector multiplications

Opportunities & challenges

CS offers a framework to *design* the *next*-generation of seismic *acquisition* technology.

Difficult to derive *engineering* principles because *sampling* matrices are prohibitively *large*.

Scale up to 3D data is a challenge

- seek *higher* dimensional transforms that exploit low *rankness*
- seek *optimization* techniques that exploit this *property*

SLIM

Opportunities & challenges

CS relies on a careful calibration

- affects of round-off errors can not be offset by increasing sampling rates [Saab & Yilmaz]
- errors in the sampling matrix are detrimental for recovery by sparsity promotion

Looking into

- classification of errors in relation to matrix type
- robust norms