Approximation-Tolerant Model-Based Compressive Sensing

Chinmay Hegde

CSAIL, MIT

Joint work with:
Ludwig Schmidt and Piotr Indyk

Approximation-Tolerant

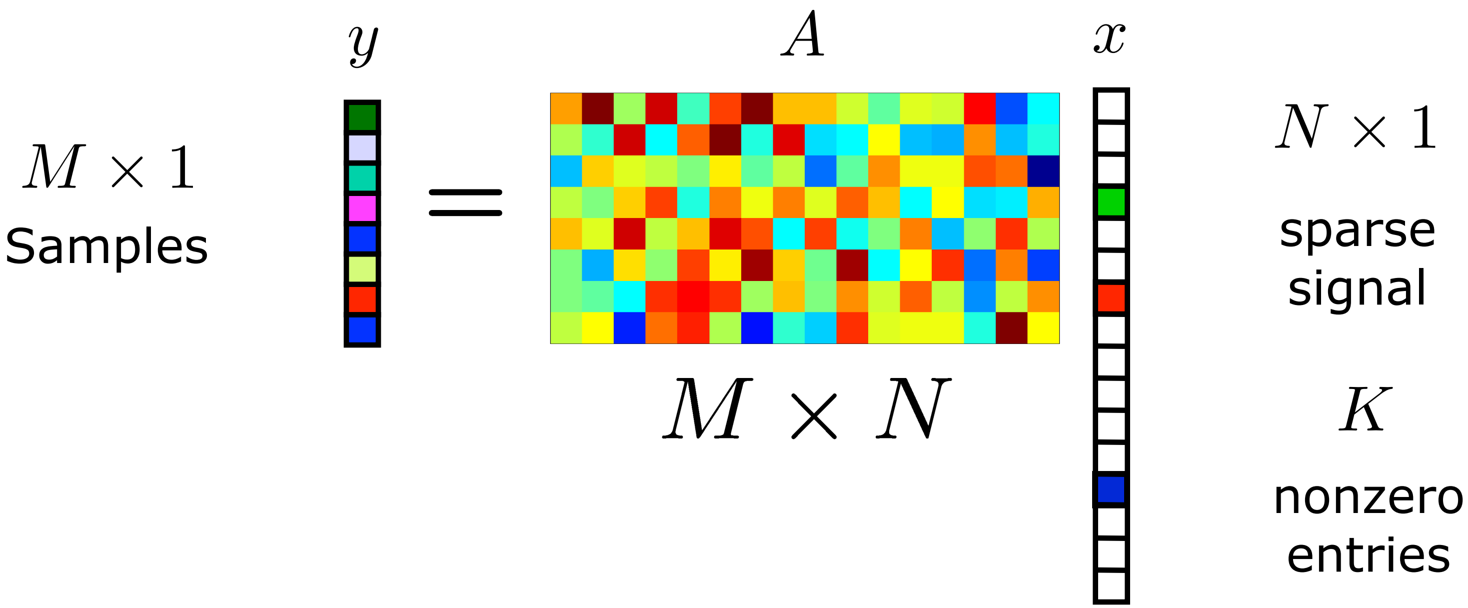
Model-Based

Compressive Sensing

Compressive Sensing

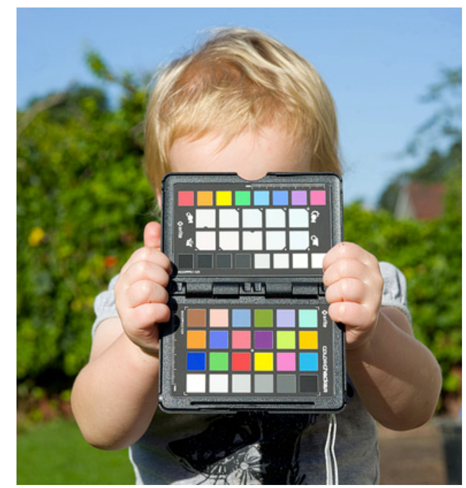
Compressive Sensing (CS)

Sampling and recovery of sparse signals ...

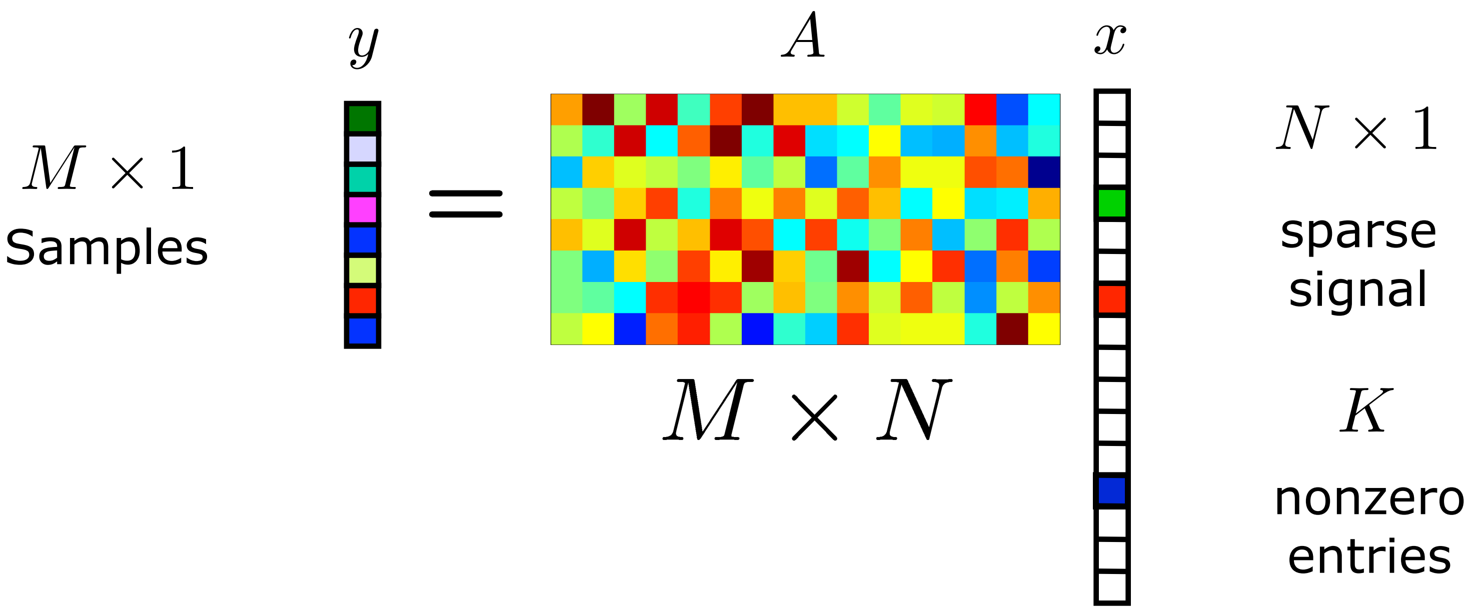


Compressive Sensing (CS)

Compressive Sensing (CS)



CS : Sampling

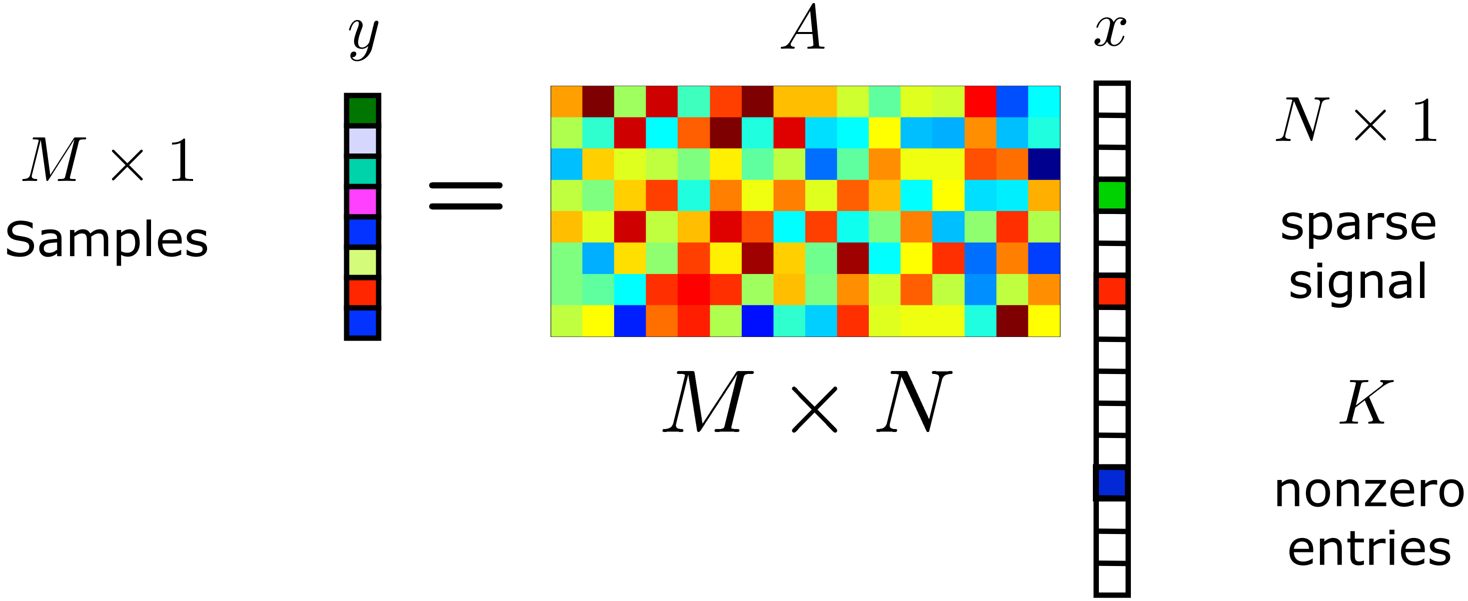


- *Random* sub-Gaussian matrix A has **RIP** w.h.p. if

$$M = O(K + \log \binom{N}{K}) = O(K \log(N/K))$$

[CRT06], [BDDeVW07]

CS : Recovery



- ℓ_1-optimization
  [CRT04]; [D04]

- Greedy algorithms
  - **iterated thresholding** [DDDeM04]; [BD07]
  - **CoSaMP** [NT09]; Subspace Pursuit [DM09]

Sparsity

- Sparsity doesn't tell the entire story ...

5% sparse image

Structure

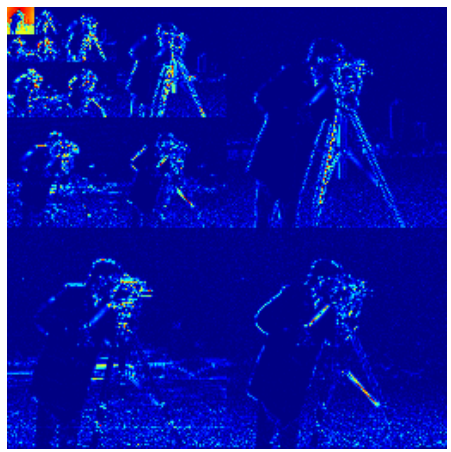
- ... since several signals exhibit **additional structure**

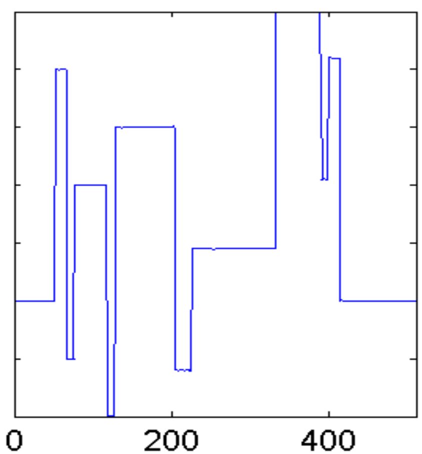
5% sparse image

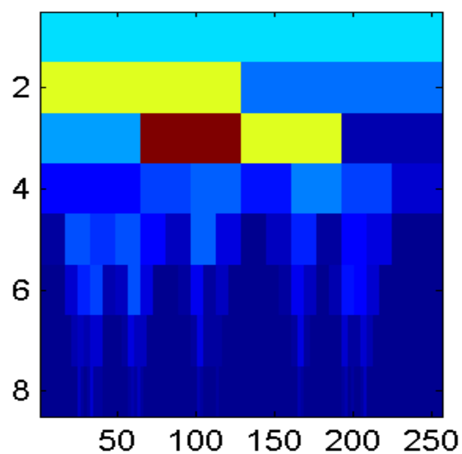
Also, a 5% sparse image! But the support is highly structured...

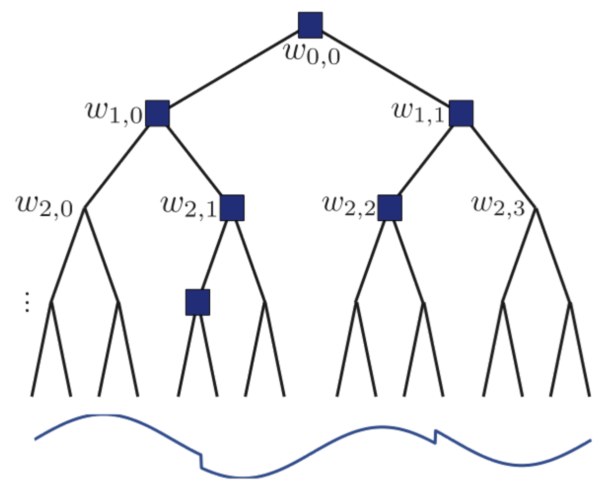
Examples of Structure

- Tree-sparsity model (in the wavelet domain) for natural images, piecewise polynomial signals..



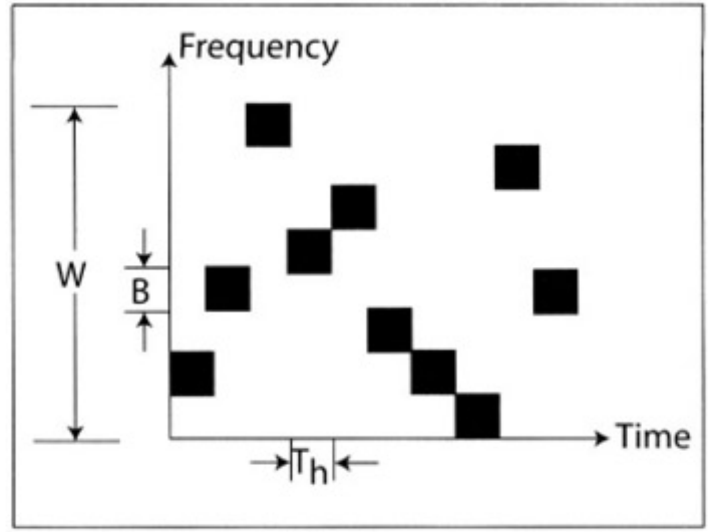




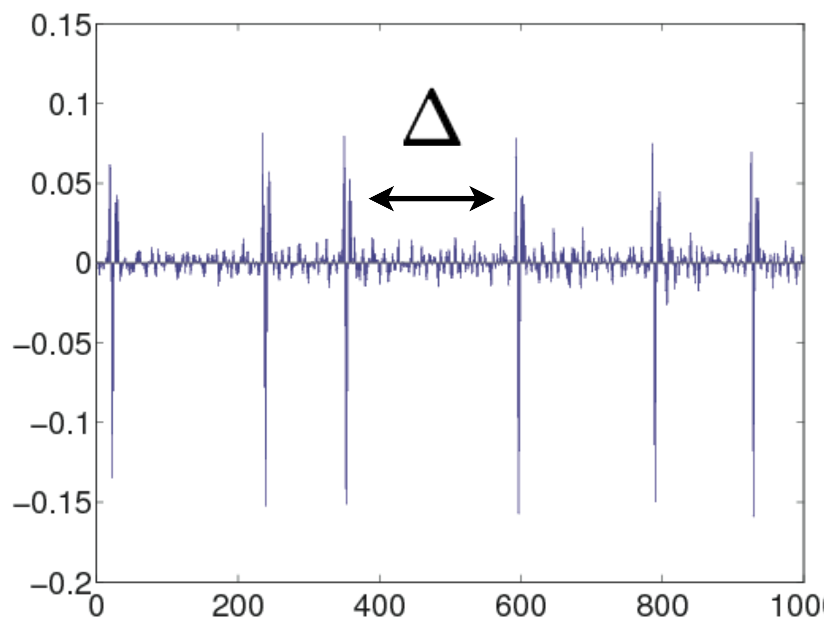


(More) Examples of Structure

- Block-sparsity model for wireless transmissions / sensor networks/ speech recordings/ gene expression data, ...



- Δ-separated spikes for neuronal recordings, electrophysiological signals, ...

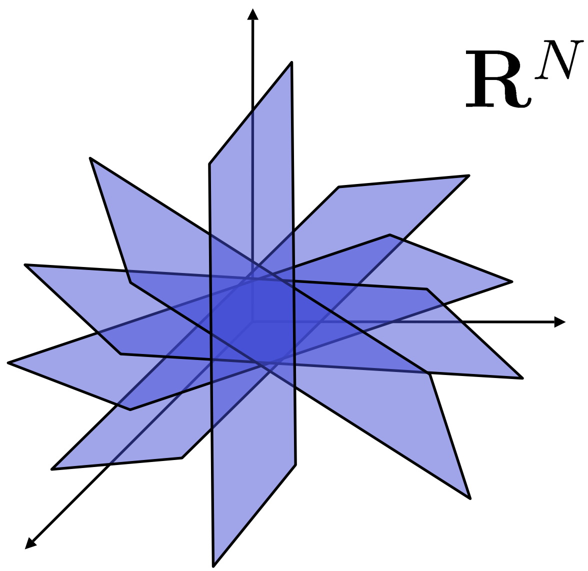


Model-Based

Compressive Sensing

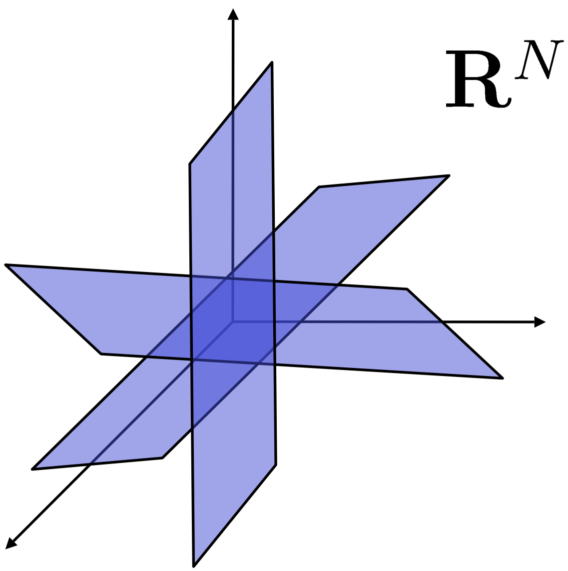
Sparse signals

- ***K*-sparse signals** comprise signals with ***all possible supports*** of size *K*



Model-sparse signals

- **Def**: A ***K*-sparse structured-sparsity *model*** comprises a particular (*reduced*) set of L_K supports

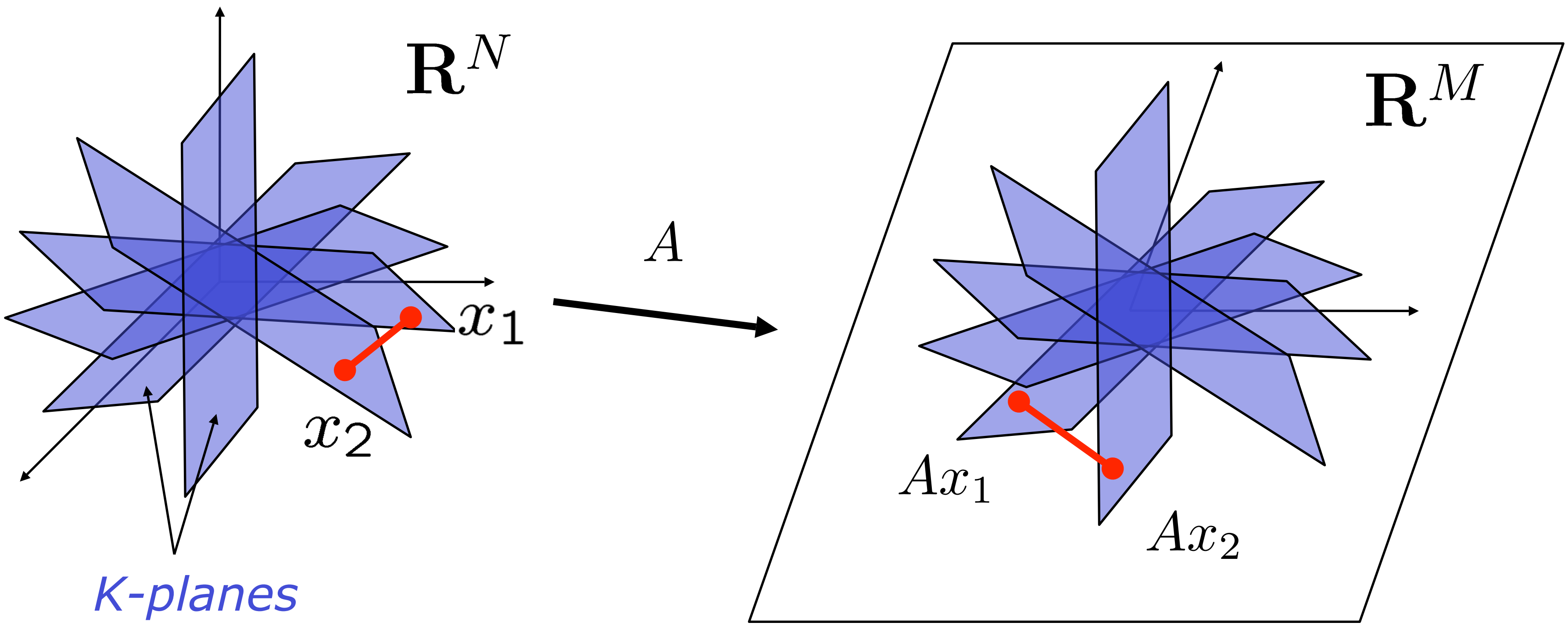


For our purposes,

$$L_K = \Theta(2^{O(K)})$$

Sampling

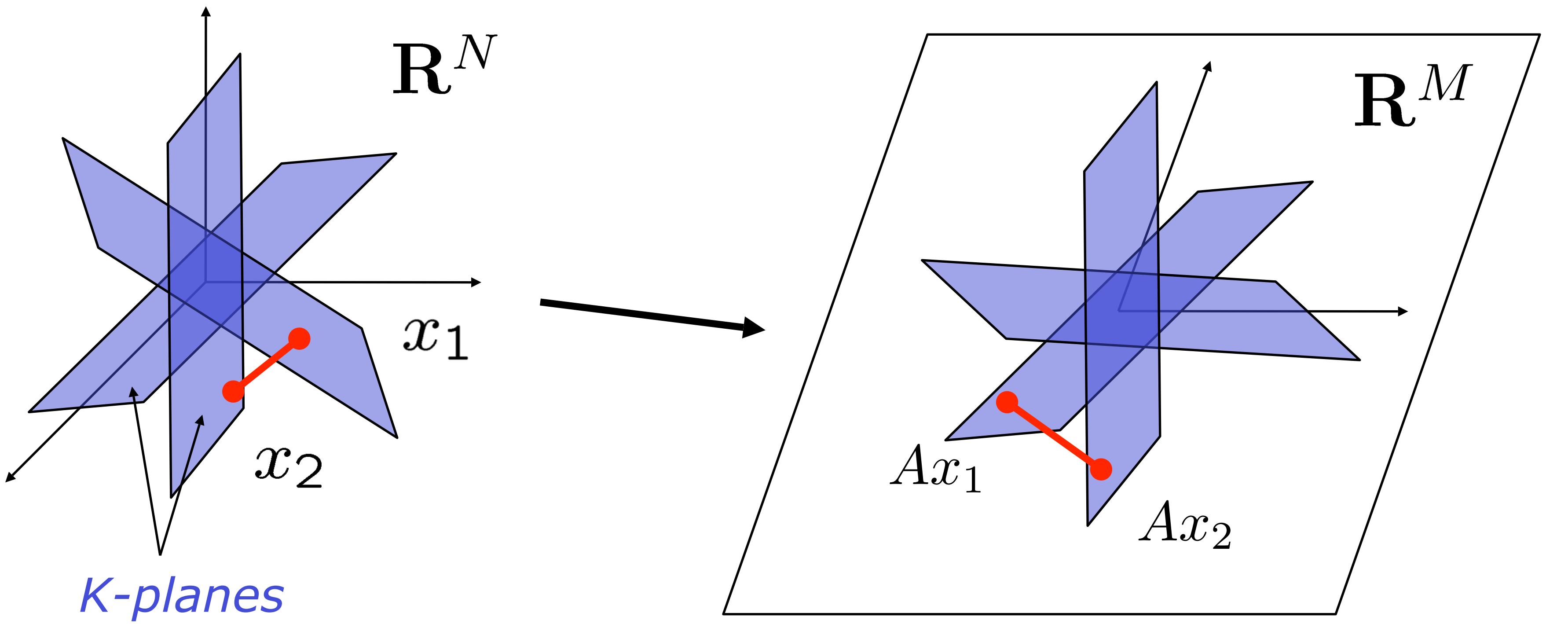
- RIP: stable embedding for *K*-sparse signals



$$M = O(K + \log \binom{N}{K}) = O(K \log(N/K))$$

Model-Based Sampling

- **Model-RIP:** embedding for **model**-sparse signals
  [B, D]; [B,D,DeV,W]



$$M = O(K + \log(L_K))$$

Sparse Recovery

- (IHT) given $y = Ax$, recover x

  iterate:

  $$x_{i+1} \leftarrow \mathrm{thresh}(x_i + A^T(y - Ax_i))$$

  where:

  $$\mathrm{thresh}(x_0, K) \leftarrow K\text{-largest elements of } x_0$$

Model-Based Recovery

- **(M-IHT)** given $y = Ax$, recover x

iterate:

$$x_{i+1} \leftarrow \mathbb{M}(x_i + A^T(y - Ax_i))$$

where:

$$\mathbb{M}(x) = x_\Omega, \text{ where } \Omega = \arg\min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

$\mathbb{M}(\cdot)$: ***Model-projection oracle***

Model-Based CS

Theorem [BCDH10]: For ***any arbitrary*** structured sparsity model, M-IHT rapidly converges to the correct answer, i.e.,

$$\|x - x_{i+1}\|_2 \leq \frac{1}{2}\|x - x_i\|_2$$

- For tree-sparsity, $M = O(K)$
- Since $M = K$ measurements are *necessary*, this scaling is info-theoretic optimal

- Similar gains for other models

Daubechies/CoSaMP - K = 6000 M = 30000

SNR = 13.1361dB

Daubechies/Tree CoSaMP - K = 6000 M = 30000

SNR = 17.8263dB

Beyond Structured Sparsity

- Along identical lines, this principle can be applied to ***virtually any signal model***:
  - Low-rank matrices [LB09],[JMD09]
  - Arbitrary unions-of-subspaces [Blu10]
  - Low-dimensional manifolds [SC10]
  - Mixtures of manifolds [HB11]

  - <insert your favorite model>

- **Very general principle** for solving inverse problems

Recipe for Model-CS

1. **An RIP-matrix for that model**

$$y = Ax$$

2. An exact model-projection oracle

Recipe for Model-CS

1. An RIP-matrix for that model

$$y = Ax$$

2. An exact model-projection oracle

$$\mathbb{M}(x) = x_\Omega, \text{ where } \Omega = \arg\min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

Challenge

- Model-projection, in general, can be computationally very challenging
  - **Sometimes even NP-hard** ☹

$$\mathbb{M}(x) = x_\Omega, \text{ where } \Omega = \arg\min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

Challenge

- Model-projection, in general, can be computationally very challenging

  - **Sometimes even NP-hard** ☹

$$\mathbb{M}(x) = x_\Omega, \text{ where } \Omega = \arg\min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

- Idea: Instead of an ***exact*** optimization, can we use an ***approximation algorithm*** instead?

This idea makes sense..

- For a number of known NP-hard optimization problems, **approximation algorithms exist**

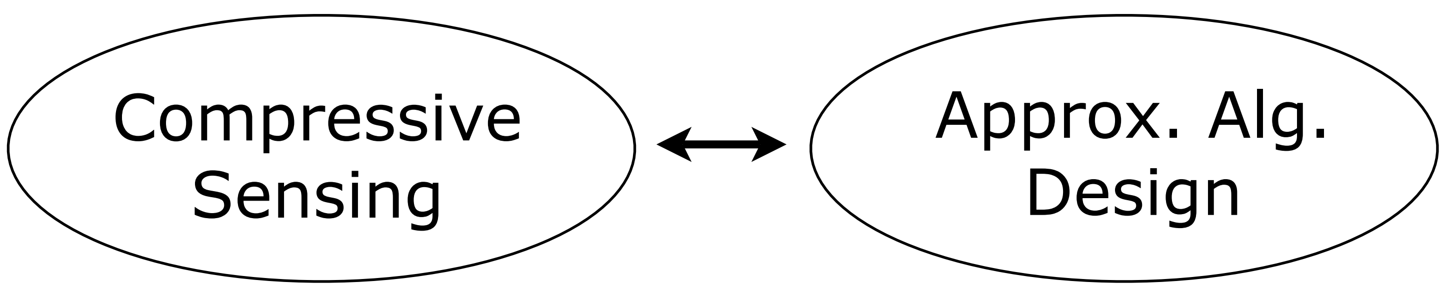
- Even if the exact optimization problem was *poly-time*, it can be impractical for real-world problems
  - e.g. a run-time of $O(N^3)$ is impractical for even a mega-pixel size image

This idea makes sense..

- For a number of known NP-hard optimization problems, **approximation algorithms exist**

- Even if the exact optimization problem was *poly-time*, it can be impractical for real-world problems

- Extensive body of research in Theory of Computing, Computational Geometry, *et al.*



Approximation-Tolerant

Model-Based

Compressive Sensing

A Version of M-IHT

- A natural notion of approximation would be the (imperfect) oracle *T*(*x*) :

$$\|x - T(x)\|_2 \leq C \min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

  - In words: the oracle returns *T*(*x*) with an error *close to the minimum possible tail error*

A Version of M-IHT

- A natural notion of approximation would be the (imperfect) oracle $T(x)$:

$$\|x - T(x)\|_2 \leq C \min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

- Let us plug this into M-IHT:

$$x_{i+1} \leftarrow T(x_i + A^T(y - Ax_i))$$

A Version of M-IHT

- A natural notion of approximation would be the (imperfect) oracle *T*(*x*) :

$$\|x - T(x)\|_2 \leq C \min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

- Let us plug this into M-IHT:

$$\cancel{x_{i+1} \leftarrow T(x_i + A^T(y - Ax_i))}$$

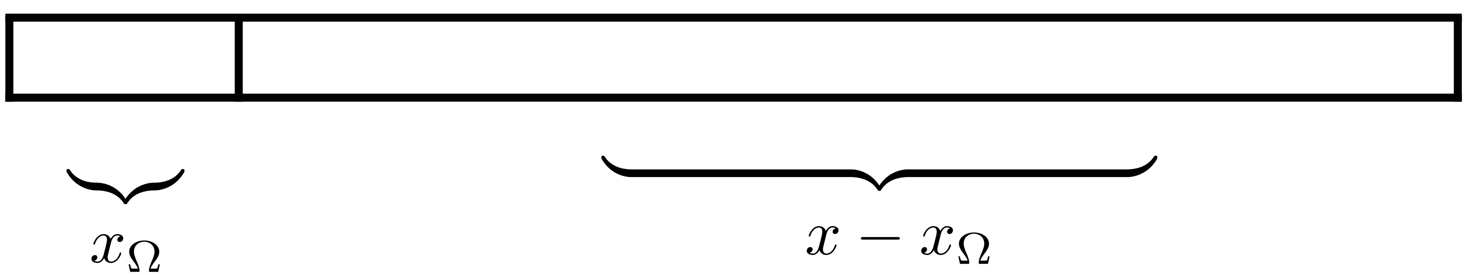
- **Unfortunately, this doesn't work ☹☹☹**

A Negative Result

- **Theorem [HIS14]**: For any **constant** value of *C*, there is an instance of M-IHT that **never** converges

$$x_{i+1} \leftarrow T(x_i + A^T(y - Ax_i))$$

- Proof intuition: Start with the zero signal; if the first signal estimate $A^T y$ has a really large tail, then M-IHT gets stuck at zero ...



A Subtle Property

- For any model, consider the **exact** projection oracle:

$$\mathbb{M}(x) = x_\Omega, \text{ where } \Omega = \arg\min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

  – i.e., the estimate minimizes the norm of the "tail"

A Subtle Property

- For any model, consider an exact projection oracle:

$$\mathbb{M}(x) = x_\Omega, \text{ where } \Omega = \arg\min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

  - i.e., the estimate **minimizes** the norm of the **"tail"**

- **Equivalent to the condition:**

$$\mathbb{M}(x) = x_\Omega, \text{ where } \Omega = \arg\max_{\Omega \in \mathcal{M}} \|x_\Omega\|_2$$

  - i.e., the estimate ***maximizes*** *the norm of the* ***"head"***

Tails vs. Heads

- Therefore, an exact projection oracle **simultaneously** optimizes for both head- and tail-problems

Tails vs. Heads

- Therefore, an exact projection oracle **simultaneously** optimizes for both head- and tail-problems

- However, an approximation oracle defined in terms of the tail error **says nothing** about the head

$$\|x - T(x)\|_2 \leq C \min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

  - $\|T(x)\|_2$ can be arbitrarily small (even zero)

A New Recipe

1. (As before) assume an **RIP-matrix** for the model

A New Recipe

1. (As before) assume an **RIP-matrix** for the model

2. Assume an imperfect **tail oracle:**

$$\|x - T(x)\|_2 \leq C_t \min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

A New Recipe

1. (As before) assume an **RIP-matrix** for the model

2. Assume an imperfect **tail oracle:**

$$\|x - T(x)\|_2 \leq C_t \min_{\Omega \in \mathcal{M}} \|x - x_\Omega\|_2$$

3. Assume a **second**, also imperfect **head oracle:**

$$\|H(x)\|_2 \geq C_h \max_{\Omega \in \mathcal{M}} \|x_\Omega\|_2$$

Approximation-Tolerant M-IHT

- (AM-IHT) given $y = Ax$, recover x

iterate: $$x_{i+1} \leftarrow T(x_i + H(A^T(y - Ax_i)))$$

Approximation-Tolerant M-IHT

- (AM-IHT) given $y = Ax$, recover x

iterate: $x_{i+1} \leftarrow T(x_i + H(A^T(y - Ax_i)))$

- **Theorem [HIS14]**: If *A* satisfies the model-RIP with constant δ, then the iterates of AM-IHT satisfy

$$\|x - x_{i+1}\|_2 \leq (1 + c_T)\left(\frac{\sqrt{1 - c_H^2}(1+\delta) + \delta}{c_H} + 2\delta\right)\|x - x_i\|_2$$

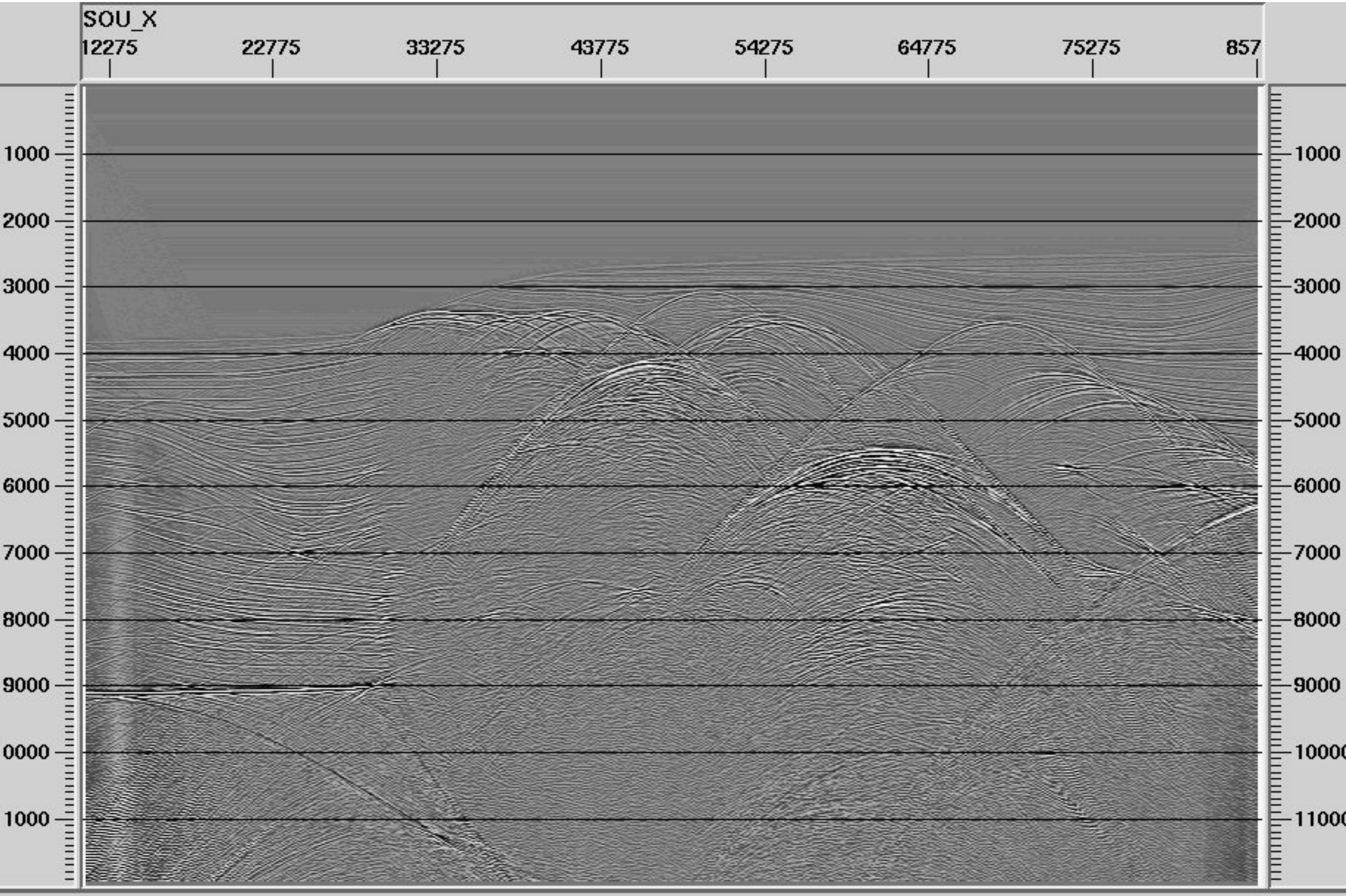
* *Extension to CoSaMP [NT08] easy, also works in noise*

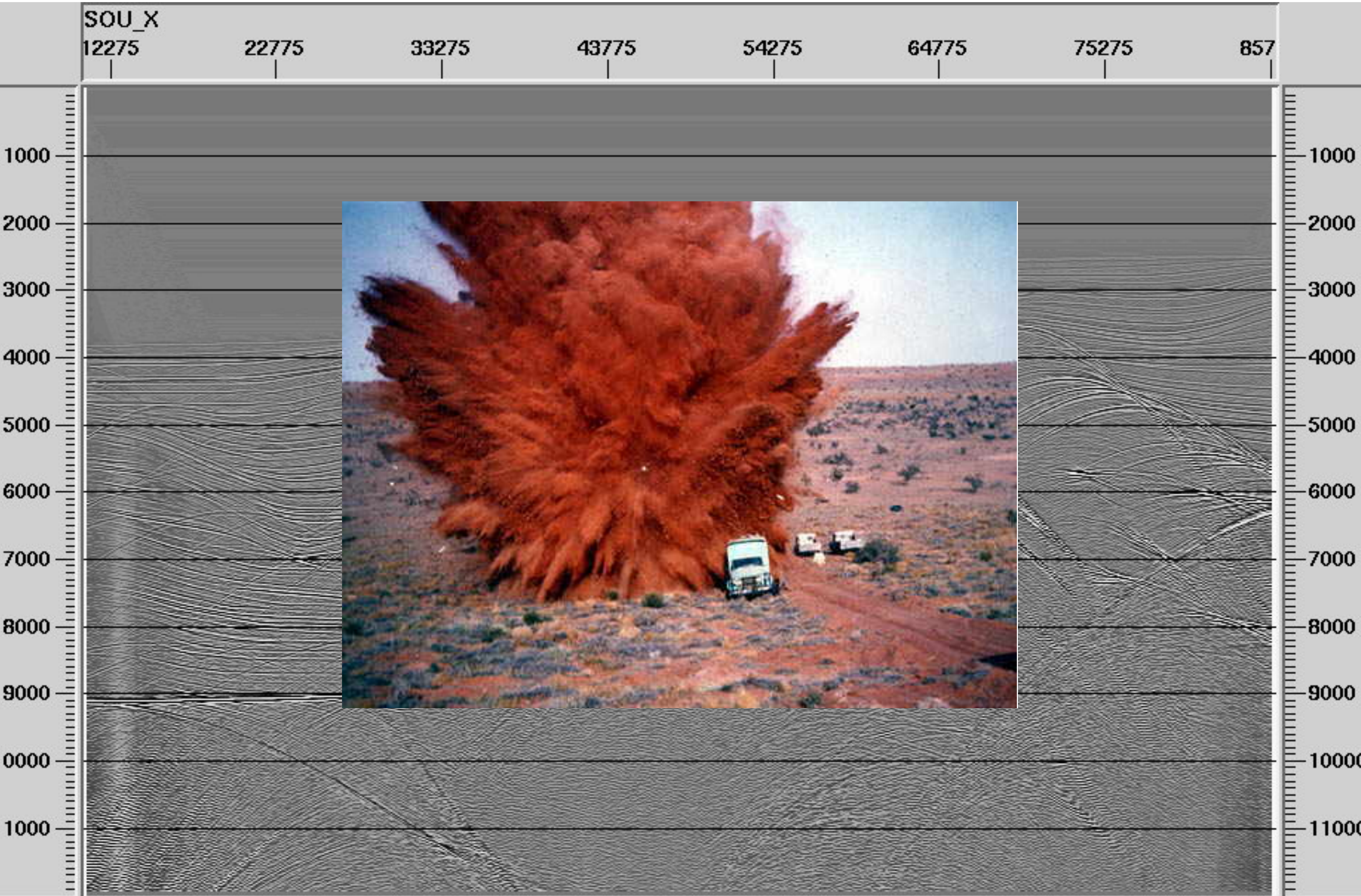
Putting our work in context

- Approximate oracles have been explored in the literature before
  - Blumensath [11] : Oracles with *additive* approx w.r.t. tail
    - **Weak** notion of approximation
  - Kyrillidis+Cevher [12]: Oracles with *multiplicative* approx w.r.t. head only
    - Convergence not guaranteed
  - Giryes+Elad [13]: Oracles with *multiplicative* approx w.r.t. tail only
    - Needs an assumption much stronger than RIP
  - Davenport+Needell+Wakin [13]: Oracles with *multiplicative* approx w.r.t. head and tail
    - Similar to ours (but somewhat more stringent)

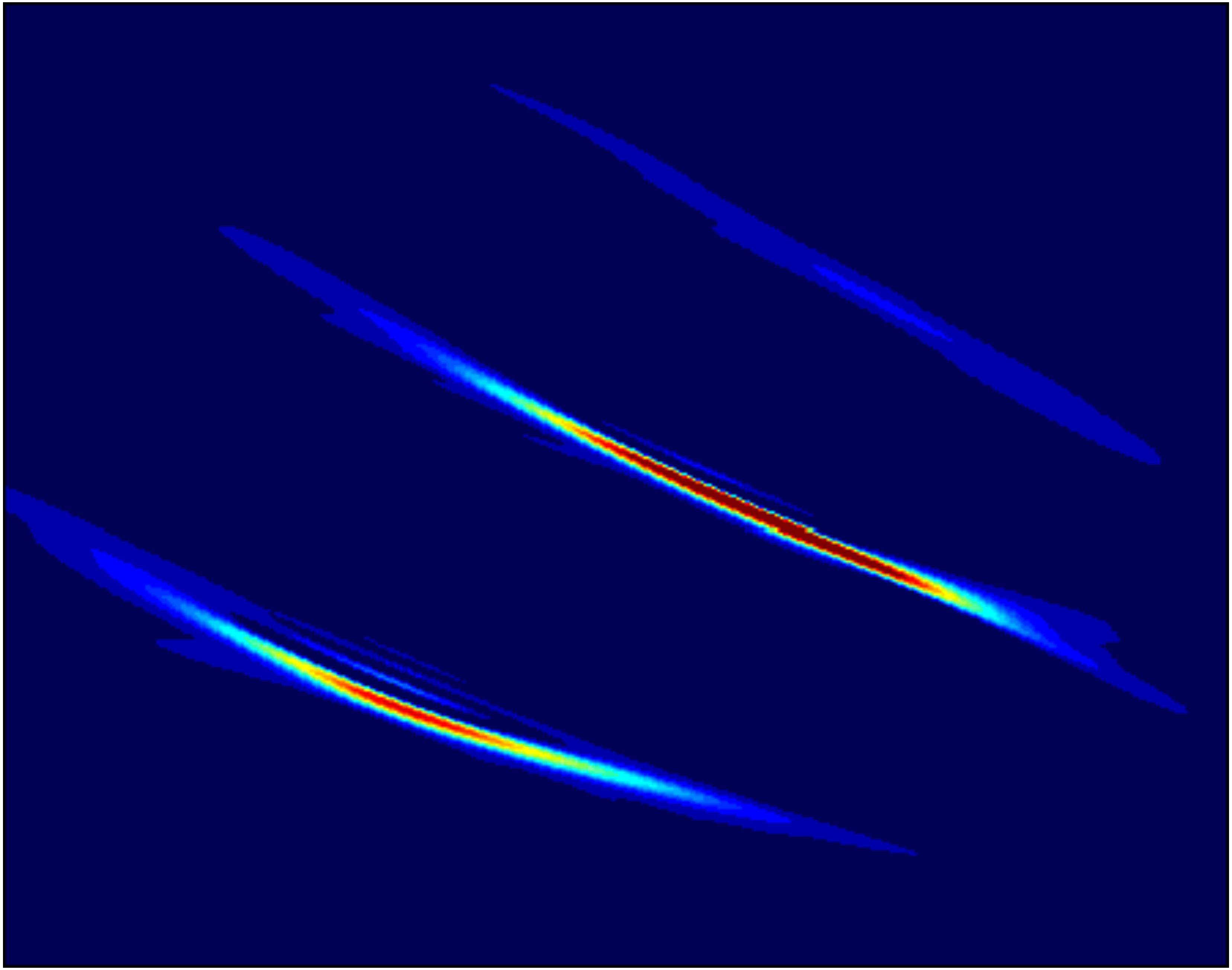
Approximation-Tolerant Model-Based Compressive Sensing:

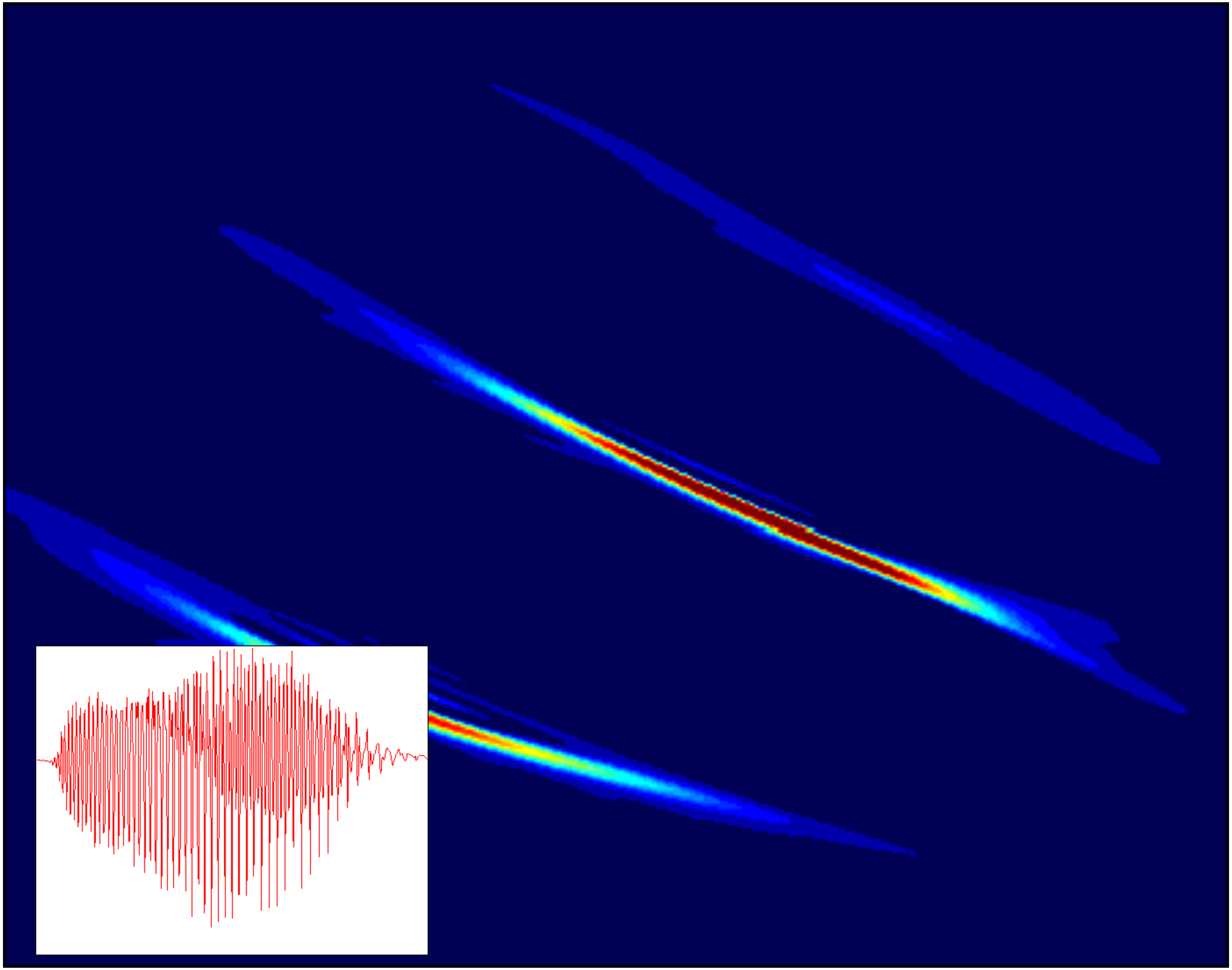
A Case Study

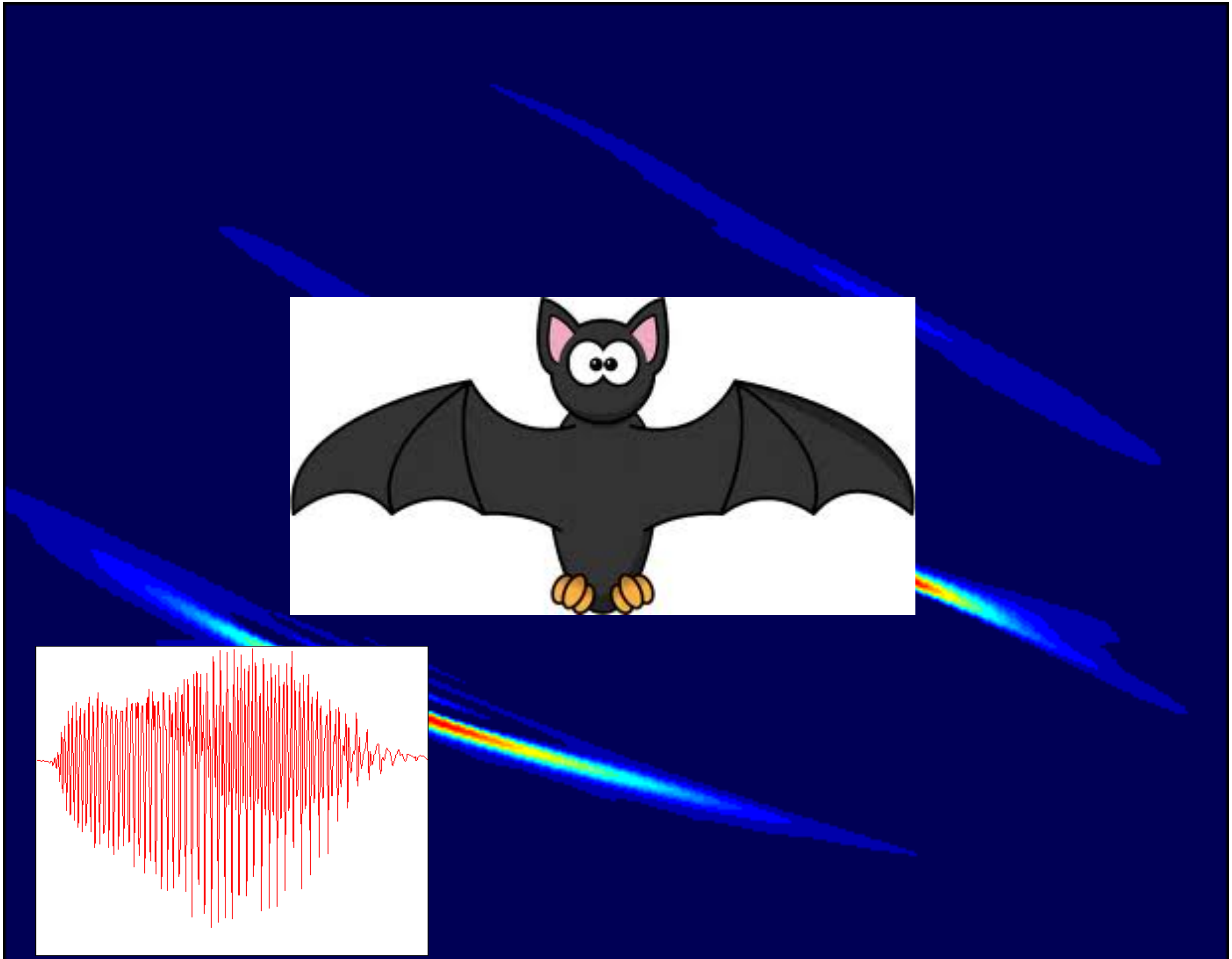

SOU_X
12275
22775
33275
43775
54275
64775
75275
857
1000
2000
3000
4000
5000
6000
7000
8000
9000



SOU_X
12275
22775
33275
43775
54275
64775
75275
857
1000
2000
3000
4000
5000
6000
7000
8000
9000
0000
1000
10000
11000

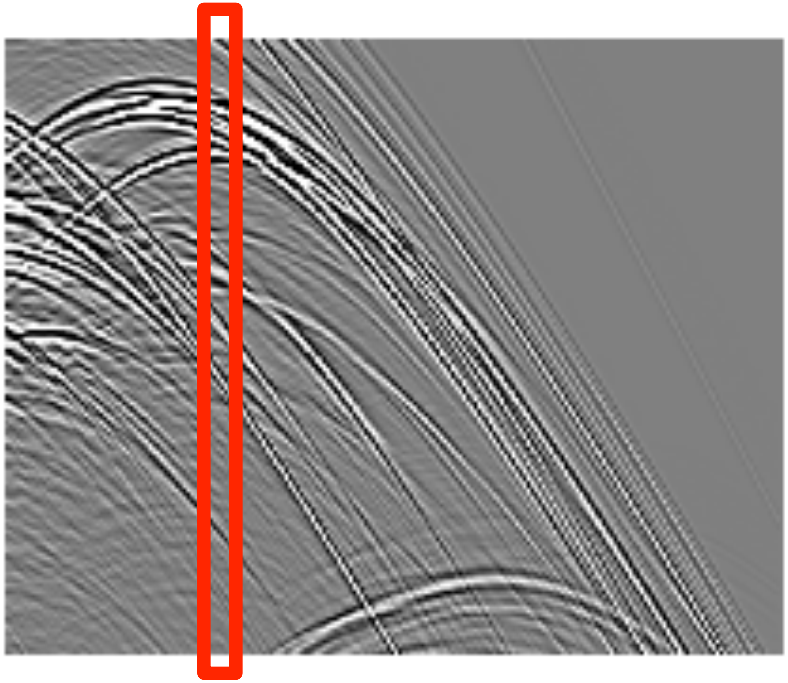




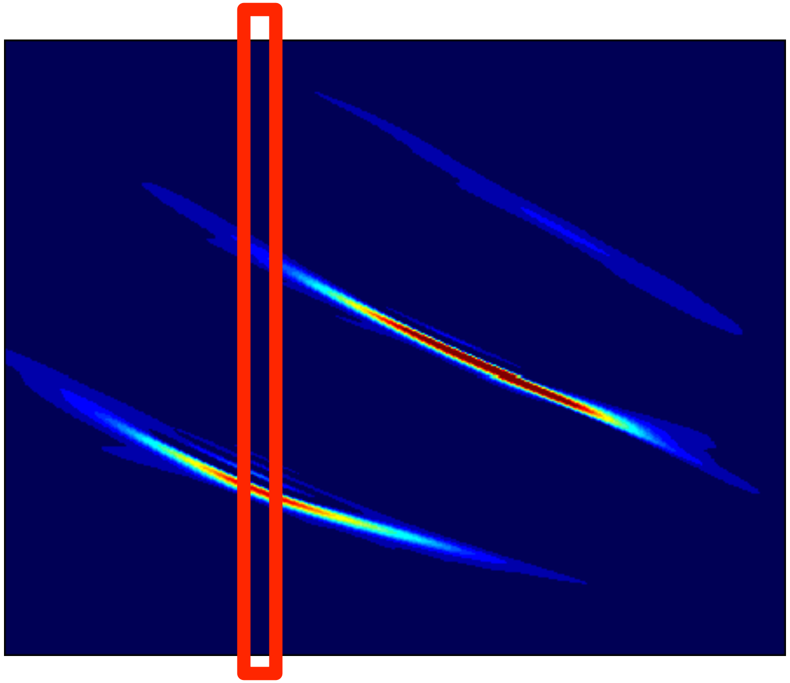


What’s Common?

- Both images are **column-sparse..**



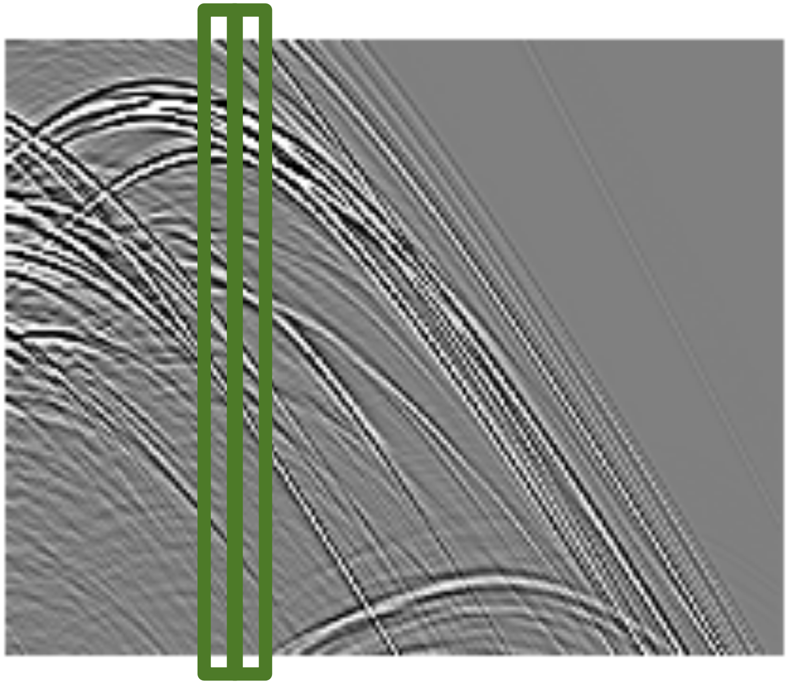
Seismic shot gathers



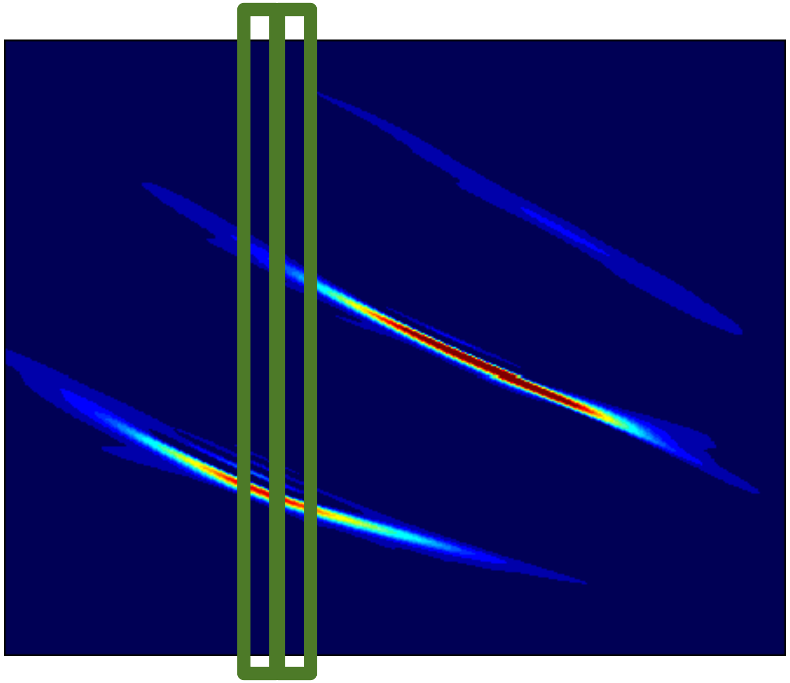
Bat-chirps (Time-frequency)

What's Common?

- ...and adjacent columns **share similar supports.**



Seismic shot gathers

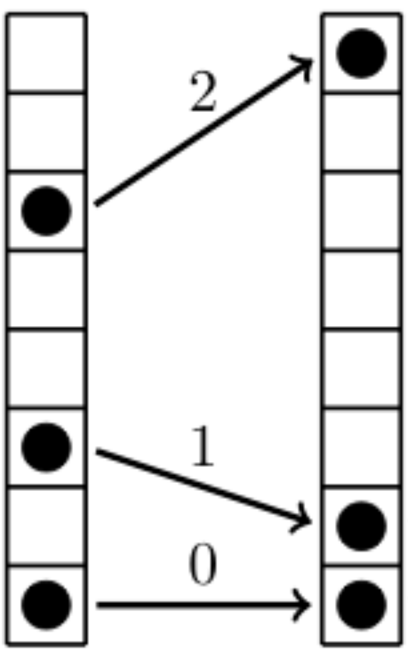


Bat-chirps (Time-frequency)

A Measure of Support Similarity

- **Earth Mover's Distance** (EMD)
  - Classical tool, used extensively in statistics, computational geometry, etc

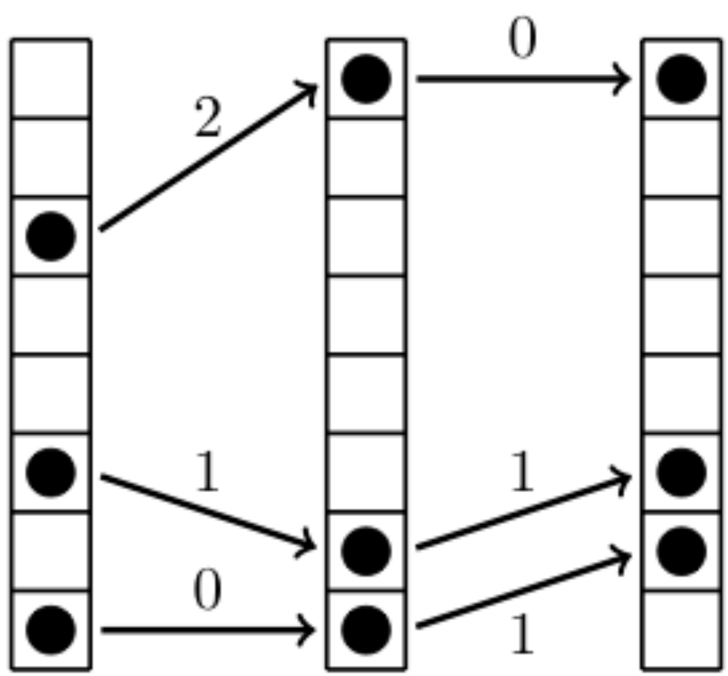
E.g. (Sparsity) k = 3, sEMD = 3



A Measure of Support Similarity

- **Earth Mover's Distance** (EMD)
  - Classical tool, used extensively in statistics, computational geometry, etc

E.g. (Sparsity) k = 3, sEMD = 5

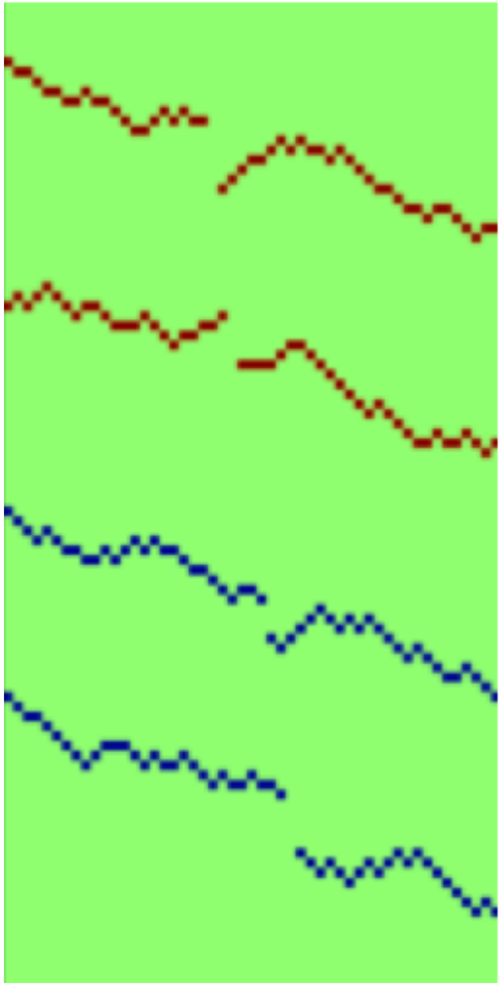


Extension to multiple columns is inductively defined

A New Signal Model

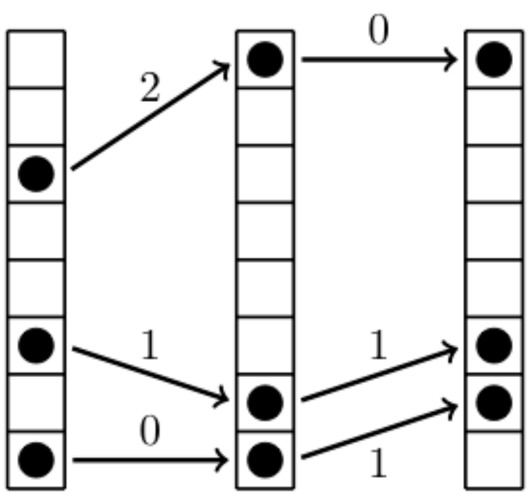
- **Def**: The **Constrained-EMD** model is the set of 2D signals $\mathcal{M}_{k,B}$ of size *N* = *h* x *w* parameterized by:
  - Column sparsity (at most) *k* (i.e., total sparsity *K* = *k* x *w*)
  - Cumulative Support-EMD (at most) *B* ("**EMD-budget**")

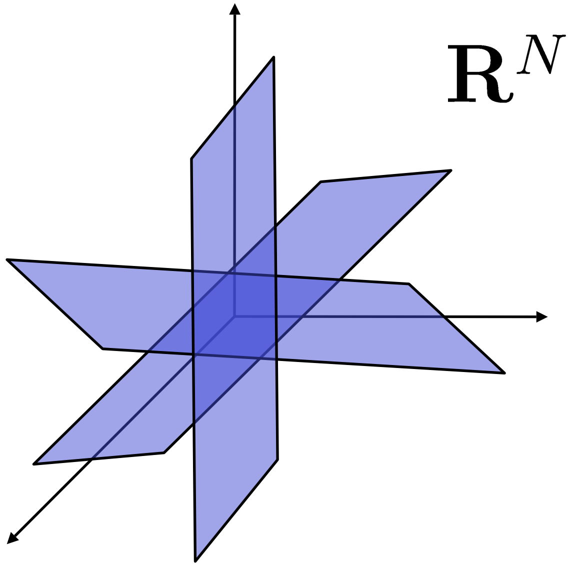
- Visualization: think of *k paths* in the plane from left to right



Ingredient #1: RIP Matrix

- Boils down to counting the total number of admissible supports in the CEMD model, L_K





- **Theorem [HIS14]**: For not-too-large values of EMD budget *B*, the number of measurements required to satisfy RIP scales as $M = O(K + k\log(B/k))$

Ingredient #2: Tail Oracle

- We want to (approximately) solve the problem

$$\text{minimize}\, \|X - X_\Omega\|\,, \quad \text{s. t.}$$

$$\text{col-sparsity}(\Omega) \leq k, \;\; \text{sEMD}(\Omega) \leq B$$

- Intuition:, consider the ***Lagrange relaxation***

$$\text{minimize}\, \|X - X_\Omega\|_2^2 + \lambda\, \text{sEMD}(\Omega), \quad \text{s. t.}$$

$$\text{col-sparsity}(\Omega) \leq k$$

indexed by the relaxation parameter λ

Ingredient #2: Tail Oracle

- Each Lagrange relaxation can be embedded into a ***min-cost flow*** problem on a specific graph
- Wrap everything up with a Pareto curve argument

- **Theorem [HIS14]**: There exists a poly-time algorithm that, for any arbitrary *X*, returns an estimate that satisfies:

$$\|X - X_i\|_2^2 \leq 2 \min_{X' \in \mathcal{M}_{k,B}} \|X - X'\|_2^2$$

Ingredient #3: Head Oracle

- Can be efficiently achieved by a ***greedy*** **approximation algorithm**

- Intuition: pick the single dominant path from left to right, subtract, rinse & repeat

- **Theorem [HIS14]** : There exists a poly-time algorithm that, for any arbitrary *X*, returns an estimate that satisfies:

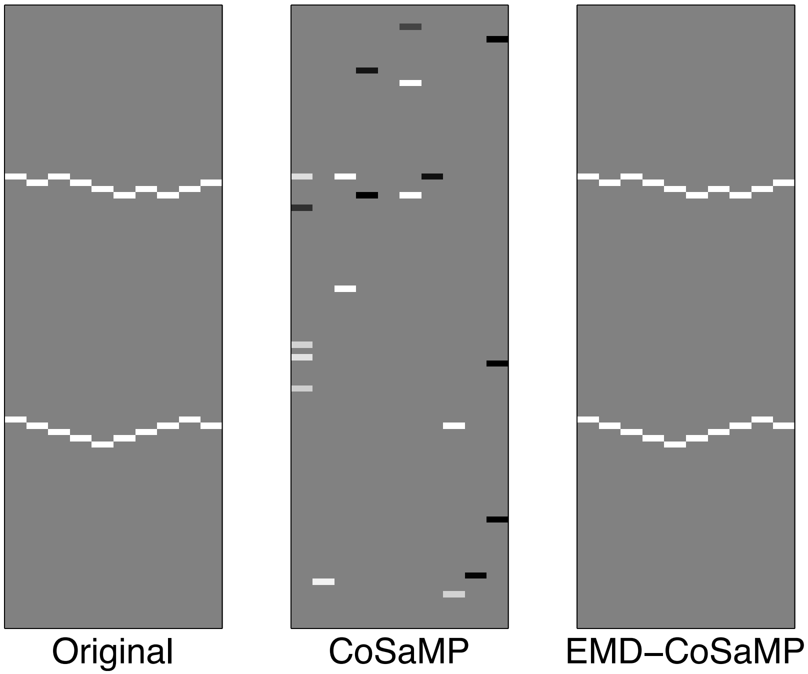
$$\|X_i\|_2^2 \geq \frac{3}{4} \max_{\Omega \in \mathcal{M}_{k,B}} \|X_\Omega\|_2^2$$

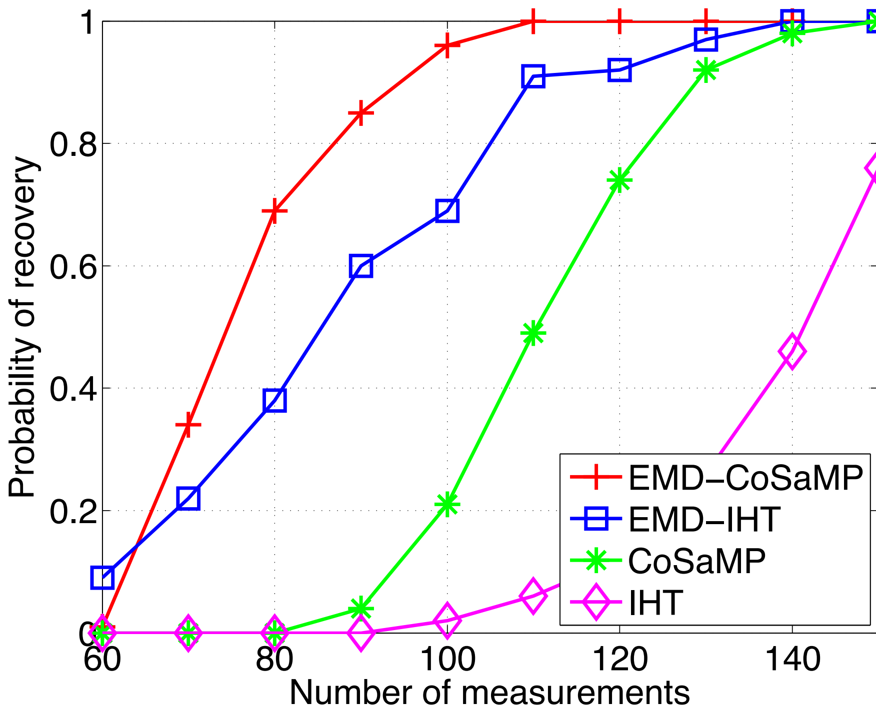
Putting the Dish Together

- RIP matrix + Tail-approximation + Head-approximation = **New CS recovery algorithm for Constrained-EMD signals**

- **Theorem [HIS14]** : If $M = O(K \log\log K)$, and EMD-budget (B) not-too-large, then AM-IHT can stably and rapidly recover any signal in the Constrained EMD model

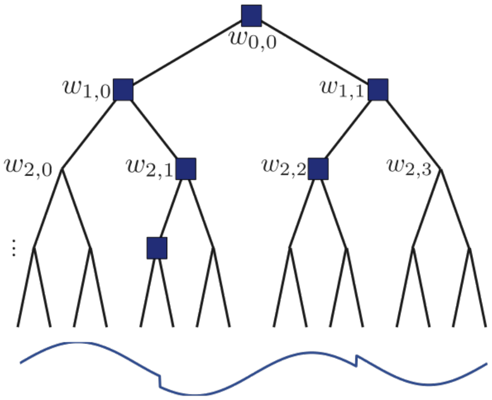
Numerical Results





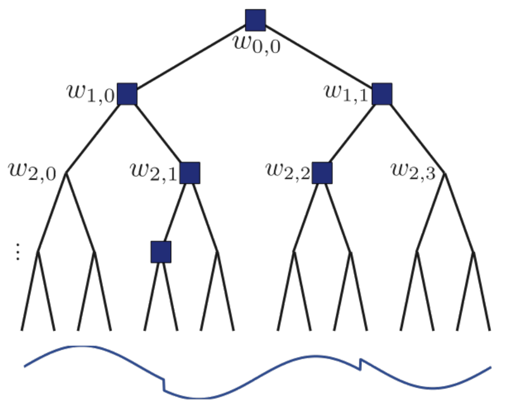
Summary

- **Model-CS**: A framework to incorporate structure into compressive sensing algorithms



Summary

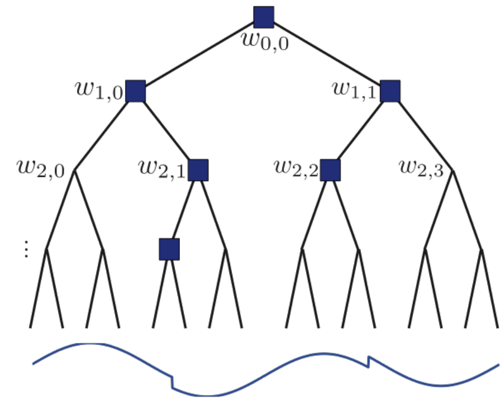
- **Model-CS**: A framework to incorporate structure into compressive sensing algorithms



- **Approximation Tolerant Model-CS**: A new way to do Model-CS by leveraging *approximation* algorithms

Summary

- **Model-CS**: A framework to incorporate structure into compressive sensing algorithms



- **Approximation Tolerant Model-CS**: A new way to do Model-CS by leveraging *approximation* algorithms

- **Constrained EMD-Model**: A new signal model for sparse signals with spatially-correlated supports

References

- [BCDH10] Baraniuk, Cevher, Duarte, Hegde, “Model-Based Compressive Sensing”, IEEE Info. Theory, 2010.

- [HIS14] Hegde, Indyk, Schmidt, “Approximation-Tolerant Model-Based Compressive Sensing”, SODA 2014.