Stable Signal Recovery from Incomplete and Inaccurate Measurements

Emmanuel Candès, Justin Romberg, Terence Tao (2005)

Background

- Authors
- Signal Reconstruction
- Nyquist-Shannon Sampling Theorem
- Compression

Emmanuel Candès

- Born Apr 27th, 1970, PhD 1998 at Stanford under David Donoho
- Dissertation "Ridgelets: Theory and Applications"
- Research Interests: Wavelet Theory, Curvelets, Compressed Sensing
- Professor of Mathematics and Statistics at Stanford University
- Alan T. Waterman Award with Terence Tao and many other honors

Justin Romberg

- PhD 2004 at Rice University under Richard G. Baraniuk
- Dissertation "Multiscale Geometric Image Processing"
- Research Interests: Imaging Inverse Problems, Data Compression, Sparse Approximation
- Professor at the School of Electrical and Computer Engineering Georgia Tech

Terence Tao (陶哲軒)

- Born July 17th, 1975, PhD 1996 at Princeton under Elias Stein
- Dissertation "Three Regularity Results in Harmonic Analysis"
- Research Interests: harmonic analysis, PDE, geometric combinatorics, arithmetic combinatorics, analytic number theory, compressed sensing, and algebraic combinatorics
- Professor of Mathematics at UCLA
- Fields Medal (2006) for contributions to PDEs, combinatorics, harmonic analysis and additive number theory

Signal Reconstruction: Problem (1)

- Consider a continuous time signal

$$x : [0, T] \to \mathbb{R},$$

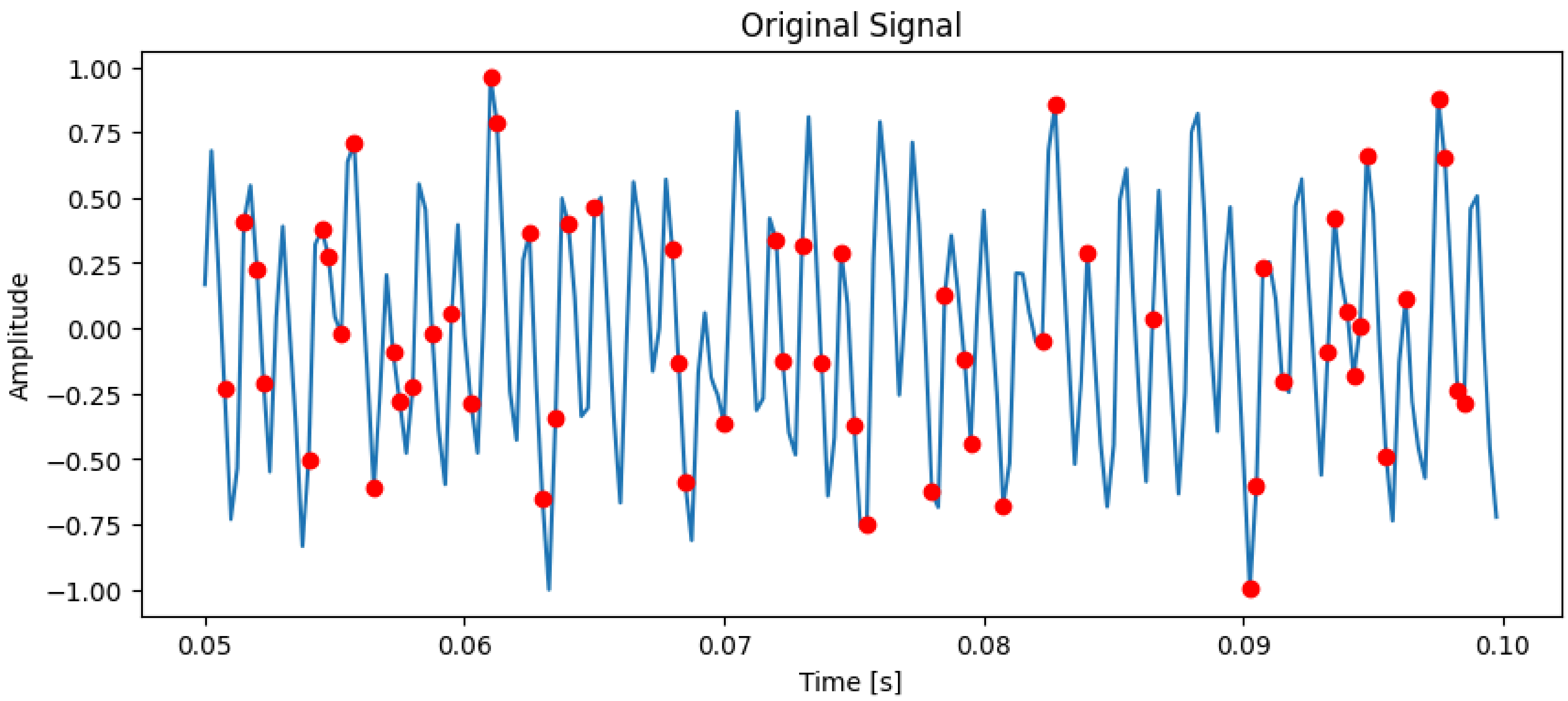
  of which samples $x_0, x_1, \ldots, x_n$ at known time points $t_0, t_1, \ldots, t_n$ are available.
- The goal of signal reconstruction is to recover the signal $x : t \mapsto x(t)$.
- Motivation: E.g. recording of audio signals (analog to digital, 1D time domain).
- Other signals could be e.g. images $[0, W] \times [0, H] \to [0, 1]^3$ (2D spatial domain), geological or medical scans (3D spatial domain).

Signal Reconstruction: Problem (2)

- $C\left[0,T\right]$ is infinite-dimensional $\Rightarrow$ clearly impossible without constraints/model assumptions.

```
Original Signal:
```

0:00 / 0:00



Nyquist-Shannon Sampling Theorem (1)

Let $x \in L^2(\mathbb{R}) \cap C(\mathbb{R})$ with Fourier transform

$$X(f) = \int_{-\infty}^{\infty} x(t) \exp(-i2\pi ft) dt$$

such that $\operatorname{supp}(X) \subseteq [-B, B]$ and $T < \frac{1}{2B}$, then

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \operatorname{sinc}\left(\frac{t - nT}{T}\right)$$

with

$$\operatorname{sinc}(x) = \begin{cases} \frac{\sin(\pi z)}{\pi z} & z \neq 0, \\ 1 & z = 0, \end{cases}$$

and this series converges absolutely and uniformly.

Nyquist-Shannon Sampling Theorem (2)

- **Model**: Signal only contains frequencies smaller than B.
- Derived in 1940 and published by Claude E. Shannon in 1948 and 1949, implied by work from Nyquist in 1928.
- Many other claimants are discussed, including E. T. Whittaker, J. M. Whittaker, D. Gabor, and É. Borel.
- Application Example: Humans can perceive frequencies in the $[20\mathrm{Hz}, 20\mathrm{kHz}]$-range. Audio CDs have a sampling rate of $44.1\mathrm{kHz}$.

Compression

- JPEG (1992):
  - Partition image into 8×8 blocks.
  - Apply the Discrete Cosine Transform (DCT) to each block (spatial $\to$ frequency domain).
  - Most frequencies are near-zero, keep only important frequency.
- MP3 (1991):
  - Apply Modified Discrete Cosine Tranform (MDCT) to overlapping segments of signal.
  - Most of the signal is concentrated on small parts of the frequency spectrum.
  - Discard negligible (very small) and inaudible (based on psychoacoustic models) frequencies.
- Well compressible in this way $\Leftrightarrow$ Sparse approximation in frequency domain exists.
- Photos and audio recordings are known to compress well.

Motivation

- Compressed Sensing
- Uncertainty Principle $\rightarrow$ Unique Sparse Representation
- Sparsity through L1 Norm

Motivation: Compressed Sensing

- **Suggests better model**: Signal is (approximately) sparse in frequency domain.
- "One can regard the possibility of digital compression as a failure of sensor design. If it is possible to compress measured data, one might argue that too many measurements were taken." -David Brady
- Idea of **compressed sensing**: Measure only the important data.
- But
    - not obvious, that it can be known a priori, which parts of the signal are important.
    - not obvious, that sparse solutions of underdetermined systems are unique.

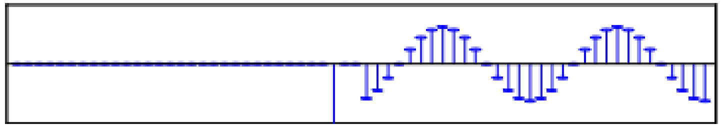
Motivation: Uncertainty Principles (Donoho-Starck, 1989)

Let $x : \mathbb{N}_{\leq N} \to \mathbb{R}$ be a discrete, non-zero signal, $\hat{x} : \mathbb{N}_{\leq N} \to \mathbb{R}$ be its discrete Fourier transform, i.e. $\hat{x} = \Psi x$, then

$$\|x\|_0 \cdot \|\hat{x}\|_0 \geq N.$$

Corollary:

$$\|x\|_0 + \|\hat{x}\|_0 \geq 2\sqrt{N}.$$



Source

Uncertainty Principle $\to$ Unique Sparse Representation

Let y_{full} be the full signal of which we measured some subset of indices $S \subset \{1, \ldots, m\}$ as signal $y = I_S y_{\text{full}}$. Assume

$$y_{\text{full}} = \Psi^* x_0,$$

where Ψ^* is the inverse discrete orthonormal fourier transform, and

$$\|x_0\|_0 < \sqrt{N}.$$

Let $A := I_S \Psi^*$, then x_0 is the unique solution of:

$$\min \|x\|_0 \quad \text{s.t.} \quad Ax = y$$

Proof: Let x' be a solution and assume $x' \neq x_0$, then $\|x'\|_0 \leq \|x_0\|_0$ and for $h := x_0 - x'$, we see $\|h\|_0 < 2\sqrt{N}$ and $\|Ah\|_0 = 0$, which contradicts the uncertainty principal.

$\Rightarrow$ Signal is recoverable in principle under the sparsity in frequency model assumption.

Sparsity through L1 norm

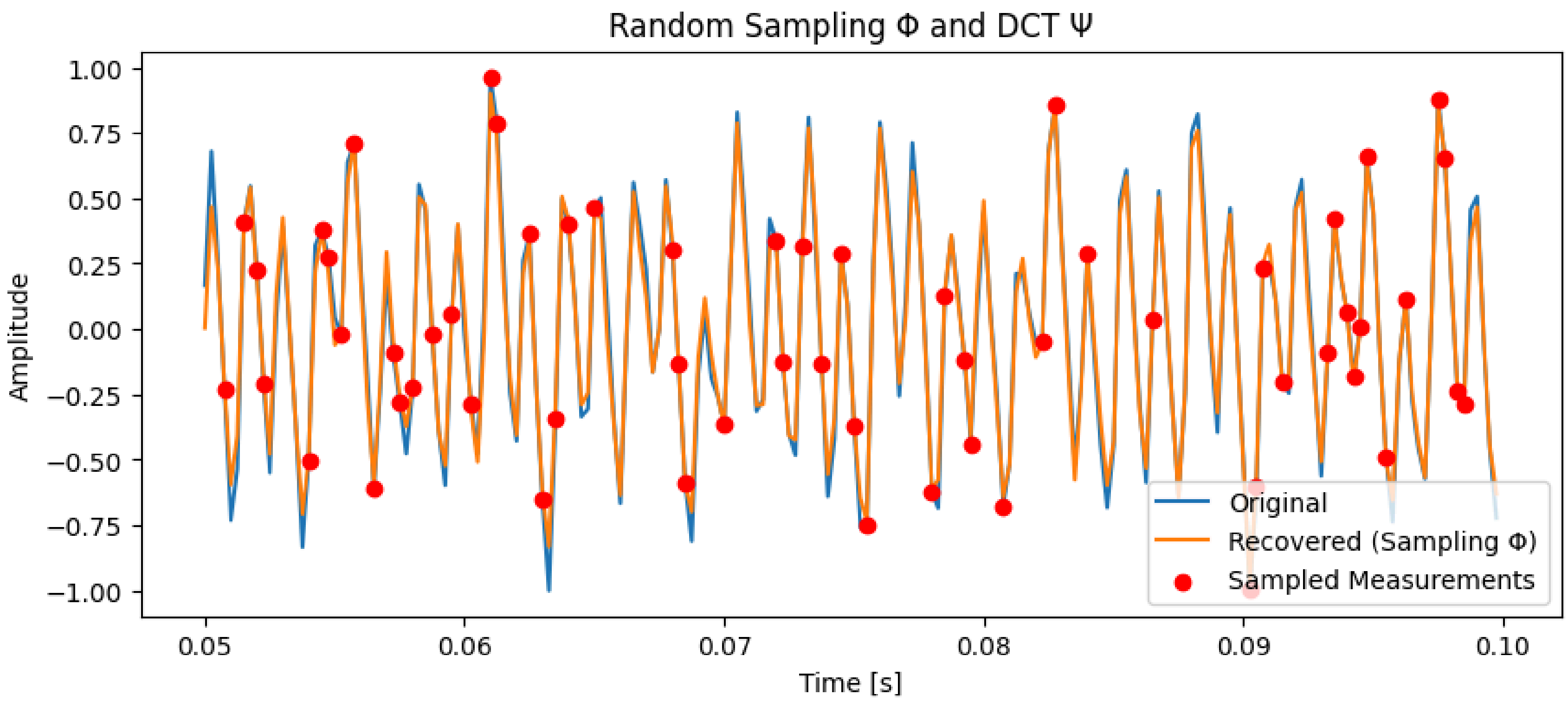
- $\min \|x\|_0$ objective is non-convex and combinatorial $\Rightarrow$ computationally intractable.
- $\min \|x\|_1$ objective is convex and induces sparsity in many settings (see e.g. LASSO).
- In settings, where the solution of

$$\min \|x\|_1 \quad \text{s.t.} \quad Ax = y$$

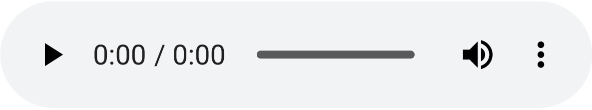
  can be shown to recover the unique sparse solution, compressed sensing is practical.
- In the 2000s such settings where identified by Donoho, Candès, Tao, Romberg and others.

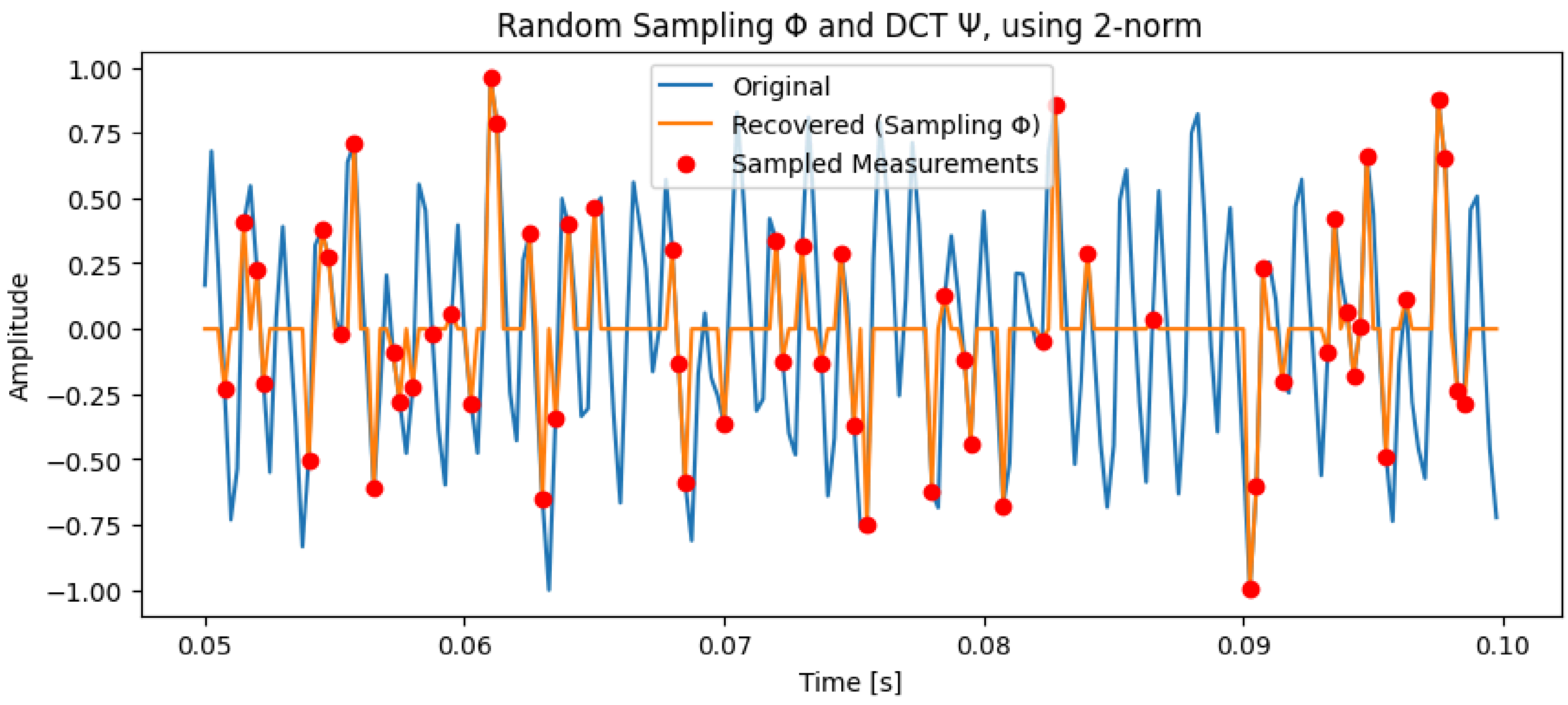
```
interactive(children=(FloatSlider(value=1.0, description='p', max=3.0, m
in=0.1), Output()), _dom_classes=('wid…

<function __main__.plot_solution(p)>
```




```
Recovered Signal (Sampling Φ, DCT Ψ):
```






```
Recovered Signal (Sampling Φ, DCT Ψ):
```

0:00 / 0:00

Results

- Restricted Isometry Property
- Theorem 1: $\|\cdot\|_1$-recovery of inaccurate, S-sparse signal under $4S$-RIP
- Theorem 2: $\|\cdot\|_1$-recovery of inaccurate, approximately S-sparse signal under $4S$-RIP.

Restricted Isometry Property: Definition (1)

Let $A \in \mathbb{R}^{n\times m}$ and for any $T \subset \{1,...,m\}$ denote by $A_T \in \mathbb{R}^{n\times|T|}$ the submatrix with the columns with indices in T extracted from A.

For $S \in \mathbb{N}_{\leq m}$ and $\delta_S \in \mathbb{R}$, if

$$(1-\delta_S)\,\|c\|_2^2 \leq \|A_T c\|_2^2 \leq (1+\delta_s)\,\|c\|_2^2$$

for all $T \subset \{1,...,m\}$ such that $|T| \leq S$, and all $c \in \mathbb{R}^{|T|}$, then we say, A satisfies the S-Restricted Isommetry Property (S-RIP) with restricted isommetry constant δ_S.

Restricted Isometry Property: Definition (2)

- With the same notations A satisfies the S-RIP with δ_S if

$$\|A_T^* A_T - I_{|T|}\|_2 \leq \delta_S$$

- Alternatively, if all eigenvalues of $A_T^* A_T$ are in $[1-\delta_S, 1+\delta_S]$.

$||\cdot||_0$-Recovery under $2S$-RIP

Let A satisfy a $2S$-RIP with $\delta_{2S} < 1$, $Ax_0 = y$, $||x_0||_0 < S$. Then x_0 is the unique solution of

$$\min ||x||_0 \quad \text{s.t.} \quad Ax = y.$$

Proof:

Let $x^\sharp$ be a solution and $h := x_0 - x^\sharp$.

1. (Subspace constraint): By feasibility, $Ah = A\left(x_0 - x^\sharp\right) = y - y = 0$.
2. (Sparsity constraint): $||h||_0 \leq ||x_0||_0 + ||x^\sharp||_0 < 2S$.

Due to the $2S$-RIP,

$$0 = ||Ah||_2^2 \geq (1 - \delta_{2S})\, ||h||_2^2$$

$$\Rightarrow ||h||_2^2 = 0 \Rightarrow x^\sharp = x_0.$$

$||\cdot||_1$-Recovery under $3S$-RIP

Let A satisfy a $3S$-RIP with $\delta_{3S} < \frac{1}{3}$, otherwise same setting, then x_0 is the unique solution of

$$\min ||x||_1 \quad \text{s.t.} \quad Ax = y.$$

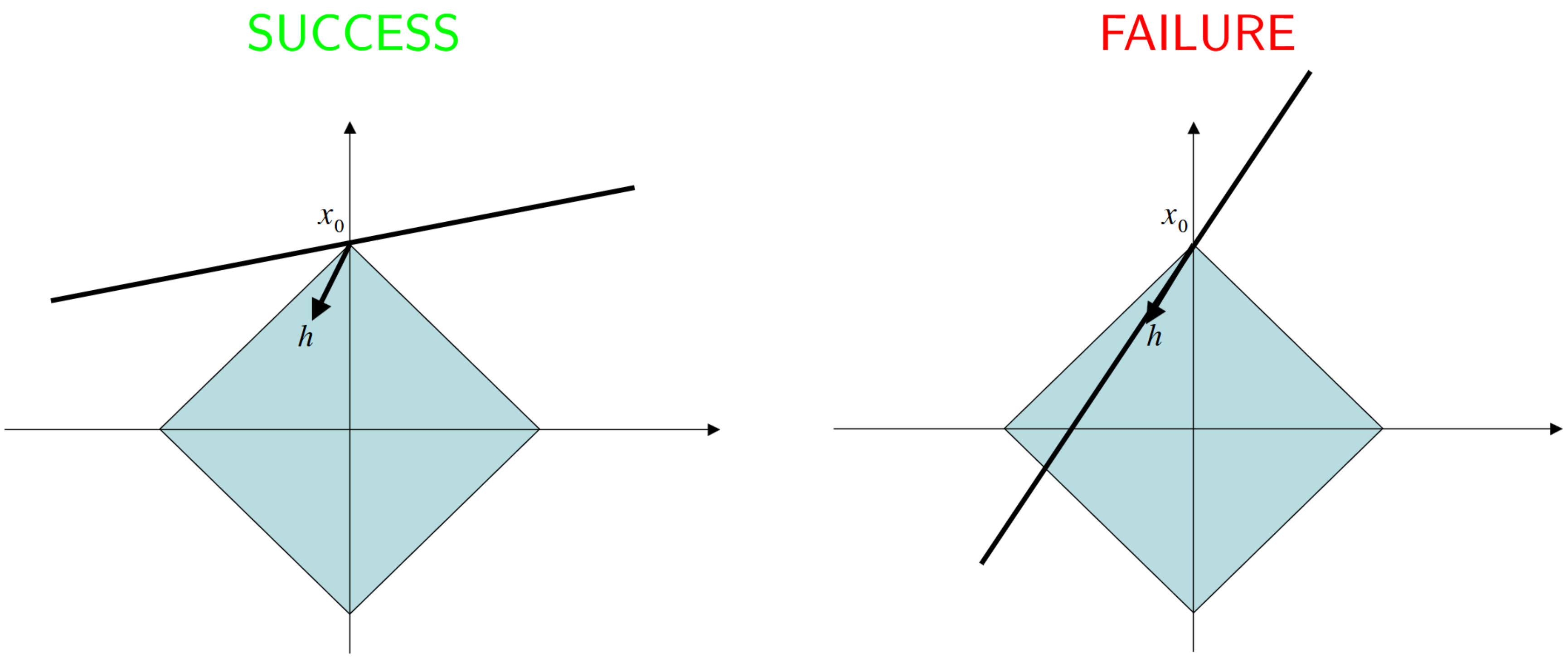
Proof:

Let T_0 be the support of x_0.

1. (Subspace constraint): Again, $Ah = 0$ by feasibility.
2. (Cone constraint): $||h_{T_0}||_1 \geq ||h_{T_0^C}||_1$, due to being a descent direction.

Combining 1. and 2. will again imply $h = 0$.

Constraints visualisation



Source

- Black line: Subspace $Ax = y$.
- The components of h on T_0 (support of x_0) are larger than the other components.

Proof: Cone Constraint

- h must point along subspace but also be a decent direction (cone constraint).

$$\begin{aligned}
\|x_0\|_1 \geq \left\|x^\sharp\right\|_1 &= \|x_0 + h\|_1 \\
&= \left\|x_0 + h_{T_0^C} + h_{T_0}\right\|_1 \\
&\geq \left\|x_0 + h_{T_0^C}\right\|_1 - \|h_{T_0}\|_1 \\
&= \|x_0\|_1 + \left\|h_{T_0^C}\right\|_1 - \|h_{T_0}\|_1
\end{aligned}$$

$$\Rightarrow \|h_{T_0}\|_1 \geq \|h_{T_0^C}\|_1$$

Proof: Using $3S$-RIP

- Let $T_1, T_2, \ldots$ be the indices of $2S$ largest, next largest, ... terms of $h_{T_0^C}$.

$$
\begin{aligned}
0 = \|Ah\|_2 &= \left\| A\left(\sum_{j\geq 0} h_{T_j}\right)\right\|_2 \\
&\geq \|A(h_{T_0} + h_{T_1})\|_2 - \sum_{j\geq 2} \|Ah_{T_j}\|
\end{aligned}
$$

$$
\|A(h_{T_0} + h_{T_1})\|_2 \leq \sum_{j\geq 2} \|Ah_{T_j}\|_2
$$

Apply $3S$-RIP on both sides.

$$
\begin{aligned}
\sqrt{1-\delta_{3S}}\|h_{T_0} + h_{T_1}\|_2 &\leq \sqrt{1+\delta_{2S}} \sum_{j\geq 2} \|h_{T_j}\|_2 \\
\|h_{T_0} + h_{T_1}\|_2 &\leq \frac{\sqrt{1+\delta_{2S}}}{\sqrt{1-\delta_{3S}}} \sum_{j\geq 2} \|h_{T_j}\|_2
\end{aligned}
$$

Proof: Show $h = 0$ (1)

$$\frac{\|h_{T_0}\|_1}{\sqrt{S}} \le \|h_{T_0}\|_2 \le \|h_{T_0} + h_{T_1}\|_2 \le \frac{\sqrt{1+\delta_{2S}}}{\sqrt{1-\delta_{3S}}} \sum_{j\ge 2} \|h_{T_j}\|_2$$

$$\le \frac{\sqrt{1+\delta_{2S}}}{\sqrt{1-\delta_{3S}}} \sum_{j\ge 2} \sqrt{2S}\|h_{T_j}\|_\infty$$

$$\le \frac{\sqrt{1+\delta_{2S}}}{\sqrt{1-\delta_{3S}}} \sum_{j\ge 2} \frac{\|h_{T_j}\|_1}{\sqrt{2S}}$$

$$\le \frac{\sqrt{1+\delta_{2S}}}{\sqrt{1-\delta_{3S}}} \frac{\left\|h_{T_0^C}\right\|_1}{\sqrt{2S}}$$

$$\Rightarrow \|h_{T_0}\|_1 \le \rho \left\|h_{T_0^C}\right\|_1 \text{ with } \rho := \frac{\sqrt{1+\delta_{2S}}}{\sqrt{2\left(1-\delta_{3S}\right)}}$$

$$\delta_{2S} \le \delta_{3S} < \frac{1}{3} \Rightarrow \rho < 1$$

Proof: Show $h = 0$ (2)

We now have

$$\|h_{T_0}\|_1 \geq \|h_{T_0^C}\|_1$$

and

$$\|h_{T_0}\|_1 \leq \rho \left\| h_{T_0^C} \right\|_1$$

with $\rho < 1$.

$$\Rightarrow h = 0.$$

Matrices satisfying RIP (1)

RIPs with $\delta_{3S} + 3\delta_{4S} < 2$ holds for A with high probability if

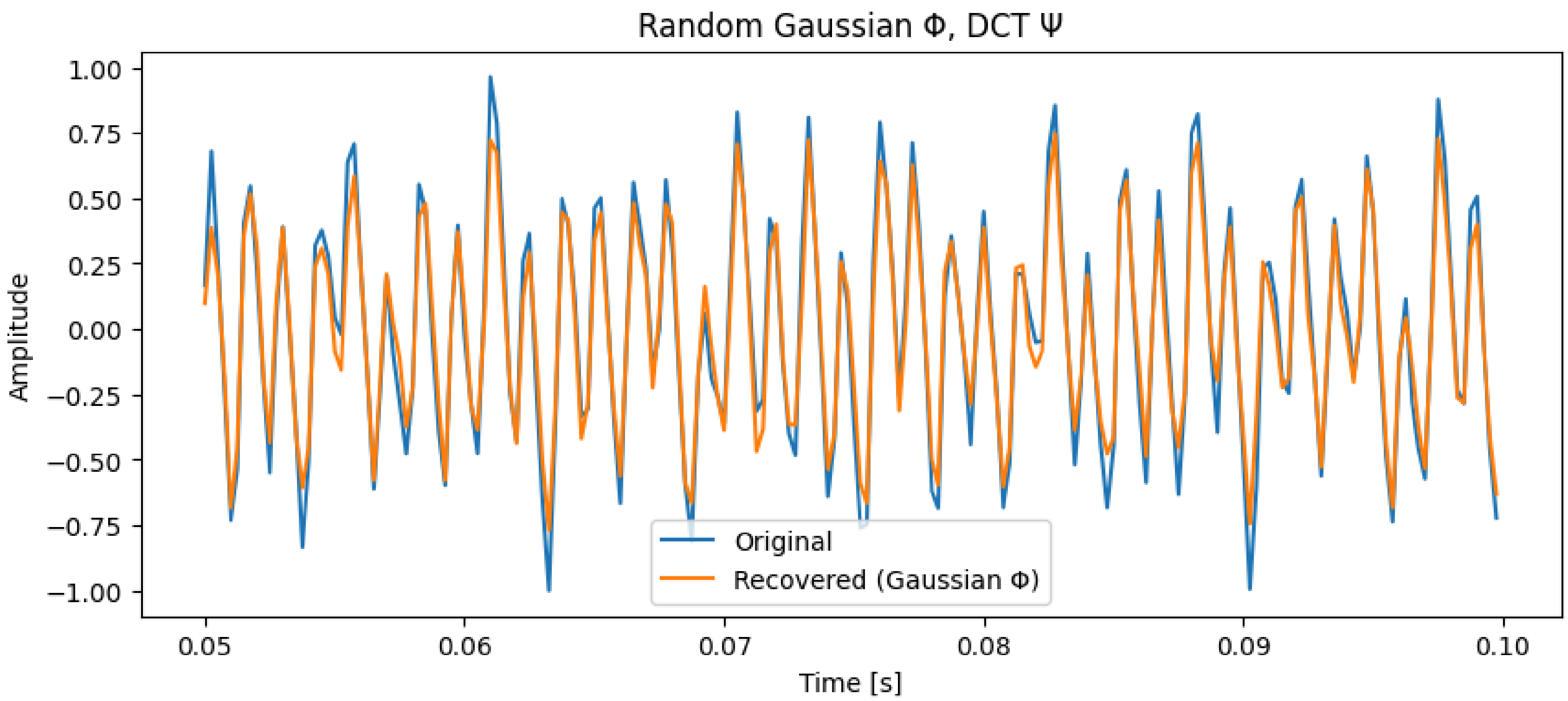
1. entries are i.i.d. $\mathcal{N}\left(0, \frac{1}{n}\right)$ and $S \leq C \cdot \frac{n}{\log \frac{m}{n}}$.
2. entries are i.i.d. $\pm\frac{1}{\sqrt{n}}$ with $p = \frac{1}{2}$.
3. n rows are randomly chosen from discrete $m \times m$ Fourier transform, columns renormalized and $S \leq C \cdot \frac{n}{\log m}$.
4. n rows are randomly chosen from orthonormal basis U, columns renormalised and $S \leq C \cdot \frac{1}{\mu^2} \cdot \frac{n}{(\log m)^6}$ with $\mu := \sqrt{m} \max_{i,j} |U_{i,j}|$.

Matrices satisfying RIP (2)

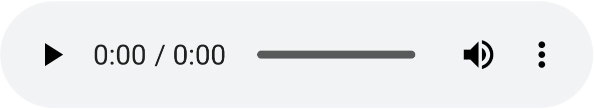
- Note a special case of 4.: $U = \Phi\Psi^*$, with
    - orthonormal basis Φ (measurements)
    - orthonormal basis Ψ in which the signal is sparse.
    - Then

      $$\mu = \sqrt{m}\,\max|\langle \phi_k, \psi_j \rangle|$$

      is a coherence measure (uncertainty principle: canonical and Fourier basis are incoherent).
- This allows using different models Ψ for the same measurement Φ.



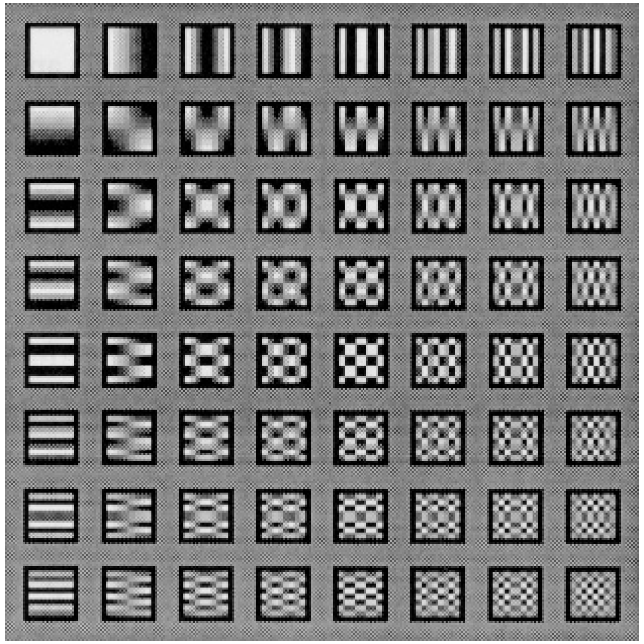

```
Recovered Signal (Gaussian Φ, DCT Ψ):
```



Matrices satisfying RIP (3)

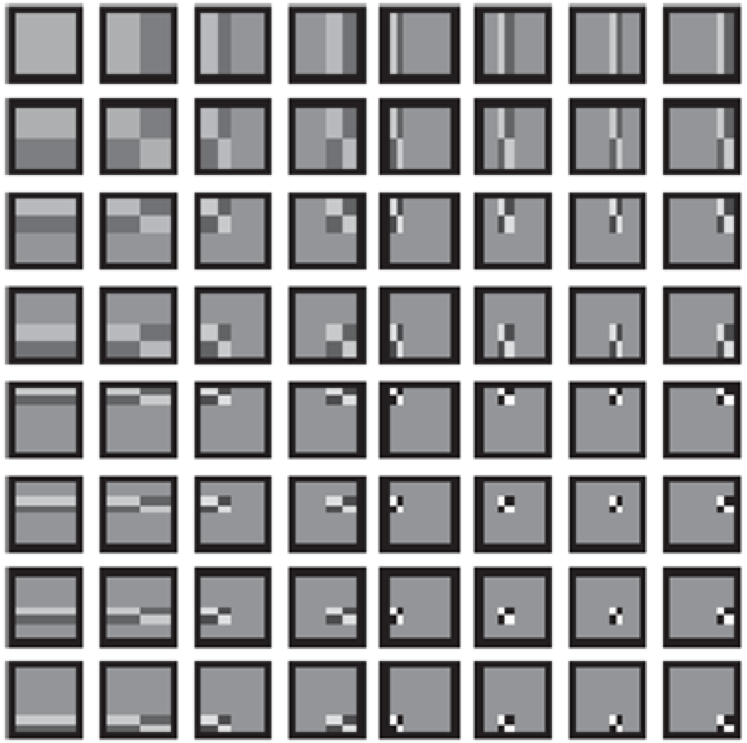
- Examples for Ψ
  - Discrete Fourier Transform (DFT)
  - (Modified) Discrete Cosine Transform ((M)DCT)
  - Block-wise DCT (JPEG, 8×8 blocks)
  - Wavelets, e.g. Daubechies wavelet (after Prof. Ingrid Daubechies, DB4, DB8, etc.), used in MRI, geological scans, etc.
- Examples for Φ
  - Canonical Basis/Identity Matrix (single pixels/samples)
  - (normalised) Binary matrix (single pixel camera)
  - Gaussian ensemble
  - Partial Fourier measurements (MRI)

Example Ψ - 8×8 DCT



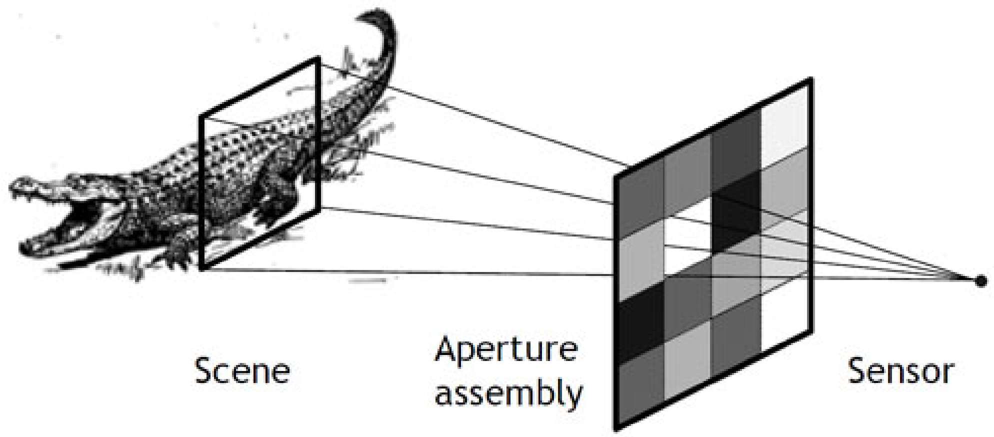
Source

Example Ψ - Haar Wavelet



Source

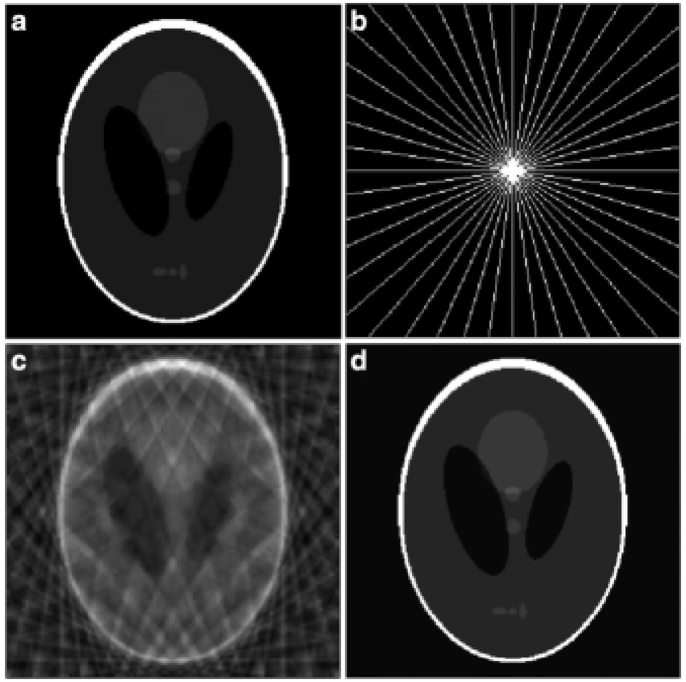
Example Φ - I.i.d. entries/Single Pixel Camera



Source

- Note that such a pattern represents one row of Φ and many random patterns are needed.

Example Φ - Incomplete Fourier samples/MRI



Source

Results: Theorem 1

Assume

$$y = Ax_0 + e \quad \text{with} \quad \|e\|_2 \leq \epsilon$$

A satisfies a RIP with $\delta_{3S} + 3\delta_{4S} < 2$, $\|x_0\|_0 \leq S$, then for the solution $x^\sharp of$

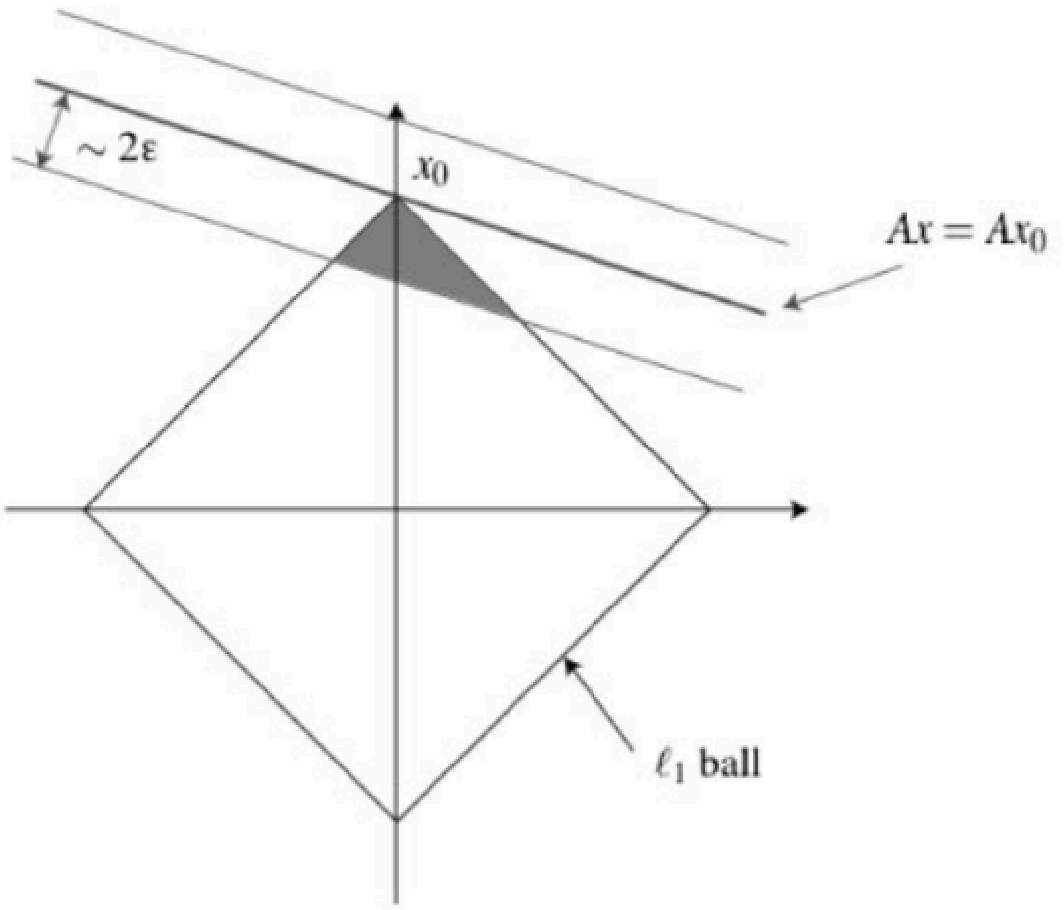
$$\min \|x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \epsilon.$$

it holds that $\|x^\sharp - x_0\|_2 \leq C_S \cdot \epsilon$, and C_S depends only on δ_{4S}.

- For e.g. $\delta_{4S} = \frac{1}{5}$, we have $C_S \approx 8.82$, for $\delta_{4S} = \frac{1}{4}$, we have $C_S \approx 10.47$.
- If the support of x_0 were known, the predictor $(A^*_{T_0} A_{T_0})^{-1} A^*_{T_0} y$ would also have a 2-norm error on the order of ε, so we cannot do much better in this model.

Proof Idea: Theorem 1

1. Previos subspace constraint becomes tube constraint $\|Ah\|_2 \leq 2\epsilon$.
2. Cone constraint $\|h_{T_0}\|_1 \geq \|h_{T_0^C}\|_1$ still holds.



Proof Idea: Theorem 1 (2)

- Similarly to exact case, decompose h into $3|T_0|$ support subvectors and estimate

$$\|Ah\|_2 \geq \sqrt{1-\delta_{4S}}\|h_{T_{01}}\|_2 - \sqrt{1+\delta_{3S}}\sum_{j\geq 2}\|h_{T_j}\|_2.$$

- Again, similarly to the exact case, estimate

$$\sum_{j\geq 2}\|h_{T_j}\|_2 \leq \sqrt{\frac{1}{3}}\|h_{T_0}\|_2$$

Proof Idea: Theorem 1 (3)

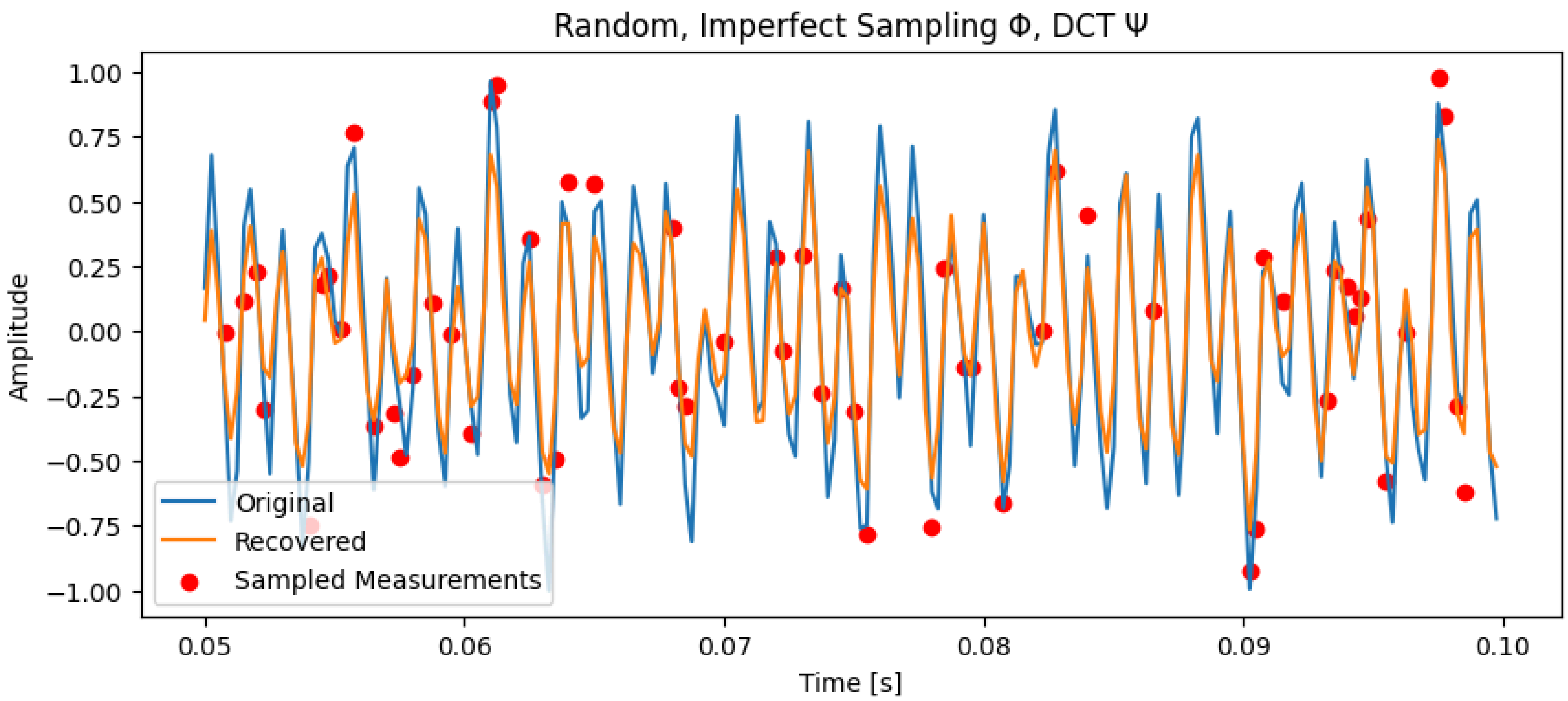
- Putting these together, we get $\|Ah\|_2 \geq C' \|h_{T_{01}}\|_2$ with

$$C' = \sqrt{1 - \delta_{4S}} - \sqrt{\frac{1}{3}} \sqrt{1 + \delta_{3S}}$$

  Under the RIP, this is positive.

- Finally,

$$\|h\|_2 \leq \frac{4}{3} \|h_{T_{01}}\|_2 \leq \frac{\sqrt{\frac{4}{3}}}{C'} \|Ah\|_2 \leq 2 \frac{\sqrt{\frac{4}{3}}}{C'} \epsilon$$



Recovered, Imperfect Signal (Sampling Φ, DCT Ψ):

0:00 / 0:00

Results: Theorem 2

Under the same setting without the sparsity assumption for x_0, the solution $x^\sharp of$

$$\min \|x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \epsilon.$$

satisfies $\|x^\sharp - x_0\|_2 \leq C_{1,S} \cdot \epsilon + C_{2,S} \cdot \frac{\|x_0 - x_{0,S}\|_1}{\sqrt{S}}$, with $x_{0,S}$ being the subvector of x_0 with the S largest entries (by absolute value). The constants depend only on δ_{4S}.

- For e.g. $\delta_{4S} = \frac{1}{5}$, we have $C_{1,S} \approx 12.04$ and $C_{2,S} \approx 8.77$.
- If x_0 is compressible in the sense $|x_0|_{(k)} \leq C_r \cdot k^{-r}$, then

$$\frac{\|x_0 - x_{0,S}\|_1}{\sqrt{S}} \leq C_r' \cdot S^{-r+\frac{1}{2}}$$

and

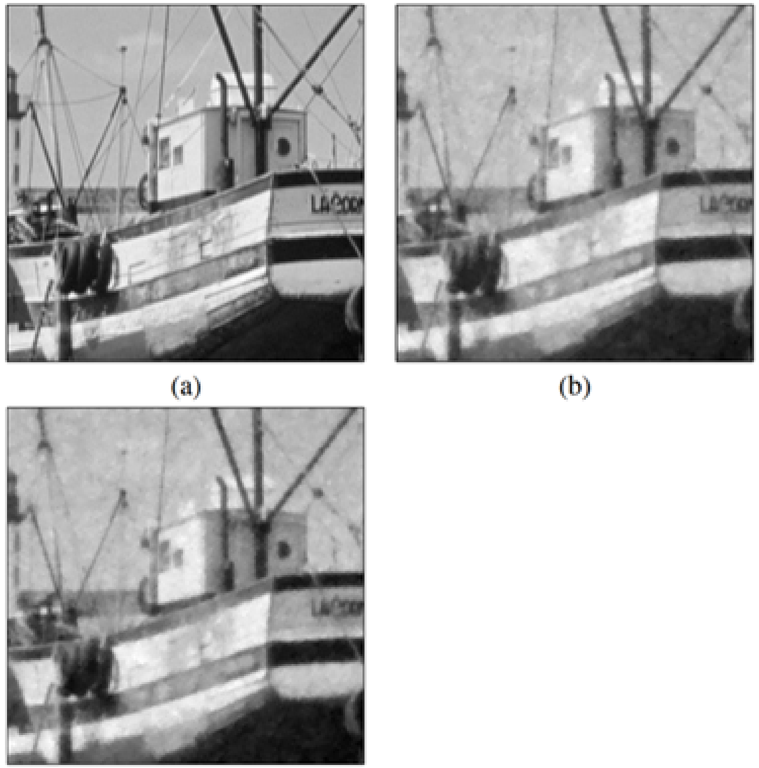
$$\|x_0 - x_{0,S}\|_2 \leq C_r'' \cdot S^{-r+\frac{1}{2}}$$

Proof Idea: Theorem 2

- Similar idea, cone constraint becomes generalized cone constraint

$$\|h_{T_0^C}\|_1 \leq \|h_{T_0}\|_1 + 2\|x_{0,T_0^C}\|_1$$

Numerical Examples



(a) (b) (c)

FIGURE 3.2. (a) Original 256×256 *Boats* image. (b) Recovery via (TV) from $n = 25\,000$ measurements corrupted with Gaussian noise. (c) Recovery via (TV) from $n = 25\,000$ measurements corrupted by roundoff error. In both cases, the reconstruction error is less than the sum of the nonlinear approximation and measurement errors.

Numerical Examples (2)

TABLE 3.1. Recovery results for sparse one-dimensional signals. Gaussian white noise of variance σ^2 was added to each of the $n = 300$ measurements, and (P_2) was solved with ϵ chosen such that $\|e\|_2 \leq \epsilon$ with high probability (see (3.1)).

σ	0.01	0.02	0.05	0.1	0.2	0.5
ϵ	0.19	0.37	0.93	1.87	3.74	9.34
$\|x^\sharp - x_0\|_2$	0.25	0.49	1.33	2.55	4.67	6.61

TABLE 3.2. Recovery results for compressible one-dimensional signals. Gaussian white noise of variance σ^2 was added to each measurement, and (P_2) was solved with ϵ as in (3.1).

σ	0.01	0.02	0.05	0.1	0.2	0.5
ϵ	0.19	0.37	0.93	1.87	3.74	9.34
$\|x^\sharp - x_0\|_2$	0.69	0.76	1.03	1.36	2.03	3.20

Lectures on Compressed Sensing by the Authors

- 11 Lecture Course by Justin Romberg at Tsinghua: Playlist with detailed motivation, intuition, proofs, etc. This paper is covered in lecture 8. Course Page
- 2008 Lecture by Terence Tao on Compressed Sensing at NTNU: Playlist. A lot of discussion of applications.
- 2017 Abel Prize Lecture by Emmanuel Candès at the University of Oslo: Wavelets, sparsity and its consequences