Survey on Compressed Sensing and its Applications

Jan Vybíral

(Czech Technical University Prague, Czech Republic)

Lecture: Compressed Sensing

The setting of compressed sensing

Linear algebra revisited

"Simplest" equation in mathematics:

$y = Ax$ for (known) $m \times N$ matrix A and $y \in \mathbb{R}^m$

Task: recover $x \in \mathbb{R}^N$ from y

Studied from many points of view:

Linear algebra: existence, uniqueness

Numerical analysis: stability, speed

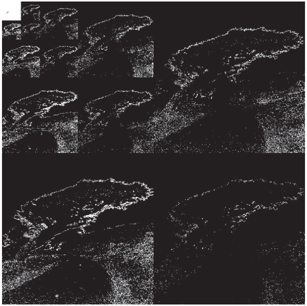
Special methods for structured matrices A

"New" point of view:

. . . we look for a solution x with special structure!

The world is compressible!

Natural images can be sparsely represented by wavelets!. . . JPEG2000



. . . today, we measure all the data (megapixels, i.e. millions), to throw the most of them away!

Sparse solutions

Simplified situation:

Let A be an $m \times N$ matrix, and let $x \in \mathbb{R}^N$ be sparse, i.e. with $\|x\|_0 := \#\{i : x_i \neq 0\}$ small. Recover x from $y = Ax$.

Natural assumption:

Given $x \in \mathbb{R}^N$. By experience, we “know” (i.e. expect) that there exists an orthonormal basis Φ with $x = \Phi c$ such that c is sparse

Task:

Let A be an $m \times N$ matrix, let $x = \Phi c \in \mathbb{R}^N$ with Φ an ONB and $\|c\|_0$ small. Recover x from $y = A\Phi c$.

Prony’s method (1795)

Let x be s-sparse, i.e. $\|x\|_0 \leq s$

Then x can be recovered by asking $2s$ (non-linear) queries:
- locations of non-zero positions
- and their value

$$\implies 2s \text{ degrees of freedom.}$$

Theorem (Prony, 1795):
Let $N \geq 2s$. Then every s-sparse vector $x \in \mathbb{R}^N$ can be recovered (by a “practical” procedure) from its first $2s$ discrete Fourier coefficients.

- not stable with respect to “defects” of sparsity, i.e. fails for “nearly sparse” vectors
- not robust with respect to noise of the measurements

Let's play: $\|x\|_0 = 1$, i.e. $x = \lambda e_j$

$$A := \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{m \times N} \quad \text{is just bad...}$$

$$A := \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & N \end{pmatrix}; A(\lambda e_j) = \begin{pmatrix} \lambda \\ j\lambda \end{pmatrix}.$$

But

$$A\Big(\frac{N-2}{N-1}, 0, \dots, 0, \frac{1}{N-1}\Big)^T = Ae_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \implies \text{bad stability}$$

Random matrices

Let $A = (a_{k,l})$, then $A(\lambda e_j) = \lambda a_{\cdot,j}$ is co-linear with $a_{\cdot,j}$, the jth column of A

If the columns of A are normalized, "nearly orthogonal", we can easily find j - in a stable way.

Random matrices

Let $A = (a_{k,l})$, then $A(\lambda e_j) = \lambda a_{\cdot,j}$ is co-linear with $a_{\cdot,j}$, the jth column of A

If the columns of A are normalized, "nearly orthogonal", we can easily find j - in a stable way.

Concentration of measure phenomenon:
If $a_{k,l}$ are i.i.d. random variables (Gaussian, Bernoulli, ...), then this is the case already for surprisingly small m's... $m \approx \log N$.

Most constructions in geometric functional analysis are random
Roman Vershynin (ICM 2010)

Sparse recovery

Natural minimization problem:

Given an $m \times N$ matrix A and $y \in \mathbb{R}^m$, solve

$$\min_x \|x\|_0 \;\text{ subject to }\; y = Ax$$

This minimization problem is NP-hard!

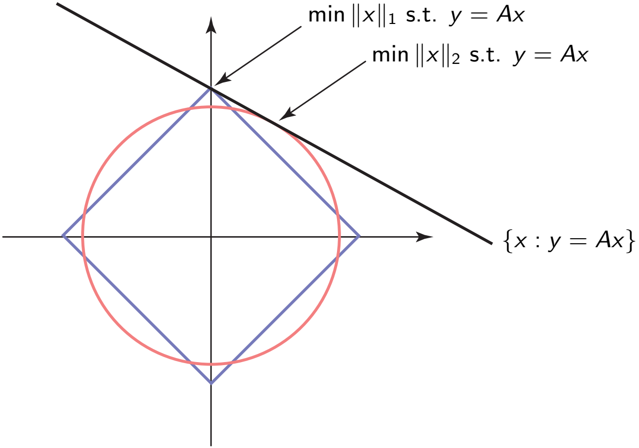
$$\|x\|_p = \left(\sum_{j=1}^{N} |x_j|^p\right)^{1/p} : \begin{cases} p \leq 1 - \text{promotes sparsity} \\ p \geq 1 - \text{convex problem} \end{cases}$$

Basis pursuit (ℓ_1-minimization; Chen, Donoho, Saunders - 1998):

$$\min_x \|x\|_1 \;\text{ subject to }\; y = Ax$$

$\longrightarrow$ This can be solved by linear programming!

ℓ_1 promotes sparsity



red: $x_1^2 + x_2^2 \leq \alpha$ blue: $|x_1| + |x_2| \leq \beta$

Summary of the introduction

Situation:

Given an $m \times N$ matrix A and a sparse $x \in \mathbb{R}^N$,
recover x from $y = Ax$!

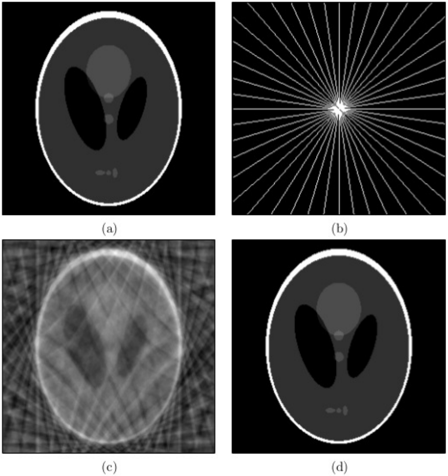
'Initial' papers:

- E. Candès, J. Romberg, T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), 1207–1223.
- D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52** (2006), 1289–1306.

Basic message:

Sparse high-dimensional signals can be recovered efficiently
from a small set of linear, non-adaptive measurements!

. . . random measurements, ℓ_1-minimization



(a) Logan-Shepp phantom, (b) Sampling Fourier coef. along 22 radial lines, (c) ℓ_2 reconstruction, (d) total variation minimization

Source: Candès, Romberg, Tao

Backgrounds of compressed sensing

Directions

Situation:

Given an $m \times N$ matrix A and an s-sparse $x \in \mathbb{R}^N$, recover x from $y = Ax$!

Fundamental (theoretical) questions:

- What is the minimal number $m = m(s, N)$ of measurements?
- For which sensing matrices is the task (uniquely) solvable?
- "Good" algorithms for recovery of x?
- Stability - i.e. "nearly sparse" x's?
- Robustness - i.e. noisy measurements?

Notation

Sparsity: $x \in \mathbb{R}^N$ is s-sparse, if

$$\|x\|_0 \leq s.$$

We often write: $\Sigma_s = \{x \in \mathbb{R}^N : x \text{ is } s\text{-sparse}\}$.

Compressibility: $x \in \mathbb{R}^N$ is compressible, if it can be well approximated by sparse vectors, i.e. when its best s-term approximation

$$\sigma_s(x)_p := \min_{\tilde{x} \in \Sigma_s} \|x - \tilde{x}\|_p$$

is small.

Null Space Property

Definition:

$A \in \mathbb{R}^{m \times N}$ has the Null Space Property (NSP) of order s if

$$\|1_\Lambda h\|_1 < \tfrac{1}{2}\|h\|_1 \quad \text{for all } h \in \operatorname{kern}(A) \setminus \{0\} \text{ and for all } \#\Lambda \leq s.$$

Theorem (*Cohen, Dahmen, DeVore - 2008*):

Let $A \in \mathbb{R}^{m \times N}$ and $s \in \mathbb{N}$. TFAE:

(i) Every $x \in \Sigma_s$ is the unique solution of

$$\min_z \|z\|_1 \ \text{subject to} \ Az = y,$$

where $y = Ax$.

(ii) A satisfies the null space property of order s.

Restricted Isometry Property

Definition:
$A \in \mathbb{R}^{m \times N}$ has the Restricted Isometry Property (RIP) of order s with RIP-constant $\delta_s \in (0,1)$ if

$$(1-\delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_s)\|x\|_2^2 \qquad \forall x \in \Sigma_s.$$

Theorem (*Cohen, Dahmen, DeVore - 2008; Candès - 2008*):
Let $A \in \mathbb{R}^{m \times N}$ with RIP of order $2s$ with $\delta_{2s} < 1/3$. Then A has NSP of order s.

Sensing matrices

Random matrices (*Candès, Donoho, et al.; 2006–2011*)
Let A be an $m \times N$-matrix with independent subgaussian entries. If

$$m \geq C\delta^{-2} s \log(N/s),$$

then A satisfies the RIP of order s with $\delta_s \leq \delta$ with prob. at least

$$1 - 2\exp(-c\delta^2 m) \quad \text{'overwhelmingly high probability'}.$$

Optimality (through high-dimensional geometry):
Stable recovery of s-sparse vectors is possible only for
$m \geq Cs \log(N/s)$.

Sensing matrices

Deterministic matrices:

$m \times N$-matrices *(Bourgain, DeVore, Haupt, et al.; 2007–2011)*:

$$m = O(s^2 \log N) \quad \text{or} \quad m = O(sN^\alpha), \quad \text{but } m \text{ must be large.}$$

Structured random matrices:

Random partial Fourier matrices

Random circulant matrices

Other constructions involving limited randomness and quick running time ...

$$m \geq Cs \log^2(s) \log^2(N)$$

Krahmer, Mendelson, Rauhut (2012)

Sparse recovery algorithms: ℓ_1-minimization

- Basis pursuit:

$$\min_x \|x\|_1 \text{ subject to } y = Ax$$

- Quadratically constrained basis pursuit:

$$\min_x \|x\|_1 \text{ subject to } \|Ax - y\|_2^2 \leq \varepsilon$$

- Unconstrained version:

$$\min_x \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_1$$

- LASSO (Least Absolute Shrinkage and Selection Operator)

$$\min_x \|Ax - y\|_2^2 \quad \text{s.t.} \quad \|x\|_1 \leq \tau$$

$\longrightarrow$ *Specialized algorithms for Compressed Sensing!*

$\longrightarrow$ `www.acm.caltech.edu/l1magic` *and* `sparselab.stanford.edu`!

Sparse recovery algorithms: greedy and combinatorial

Greedy algorithms:

- Orthogonal matching pursuit (OMP)
- Compressive sampling matching pursuit (CoSaMP)
- Iterative hard thresholding (IHT)
- Hard thresholding pursuit (HTP)
- ...

Combinatorial algorithms:

- Combinatorial group testing
- Data streams
- ...

Stability, robustness

The theory can be easily generalized to include

- *stability* (x not sparse but compressible) and
- *robustness* (measurements with noise)

Let $y = Ax + e$, $\|e\|_2 \leq \eta$, where A has the *Robust Null Space Property* of order s. Then

$$x^{\#} := \arg\min_{x} \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \eta$$

satisfies

$$\|x - x^{\#}\|_1 \leq C\sigma_s(x)_1 + D\sqrt{s}\,\eta$$

and

$$\|x - x^{\#}\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta.$$

Extensions, tricks, "small" applications

"Matrix completion", or low-rank matrix recovery

The theory applies to other sorts of sparsity!
x sparse means, that some (unknown) of its possible degrees of freedom are not used (i.e. equal to zero)

The same is true for low-rank matrices!

E. Candès and T. Tao. The power of convex relaxation: near-optimal matrix completion, IEEE Trans. Inform. Theory, 56(5), pp. 2053 - 2080 (2010)

E. Candès and B. Recht. Exact matrix completion via convex optimization, Found. of Comp. Math., 9 (6). pp. 717-772 (2009)

D. Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Inform. Theory 57(3), pp. 1548-1566 (2011)

Low-rank matrix recovery

Let $X \in \mathbb{C}^{n_1 \times n_2}$ be a matrix of rank at most r.
Let $y = \mathcal{A}(X) \in \mathbb{C}^m$ be the (linear) measurements of X.

We "want" to solve

$$\arg\min_{Z \in \mathbb{C}^{n_1 \times n_2}} \operatorname{rank}(Z) \quad s.t.\ \mathcal{A}(Z) = y.$$

$\operatorname{rank}(Z) = \|(\sigma_1(Z), \sigma_2(Z), \dots)\|_0$ gets replaced by
the nuclear norm $\|Z\|_* = \|(\sigma_1(Z), \sigma_2(Z), \dots)\|_1 = \sum_i |\sigma_i(Z)|$.

The convex relaxation is then

$$\arg\min_{Z \in \mathbb{C}^{n_1 \times n_2}} \|Z\|_* \quad s.t.\ \mathcal{A}(Z) = y.$$

Phase retrieval

Setting:
Reconstruct the signal x from the magnitude of its discrete Fourier transform $\hat{x}$

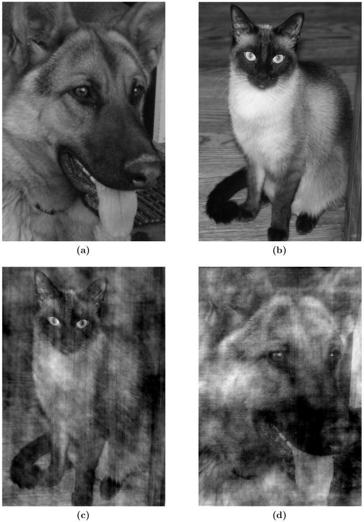
General setting:
x given, $b_k = |\langle a_k, x\rangle|^2, k = 1, \ldots, m$ known, recover x!

Frequent problem (i.e. astronomy, crystallography, optics), different algorithms exist...

PhaseLift:
quadratic measurements of x are “lifted up” and become linear measurements of the matrix $X := xx^*$:

$$|\langle a_k, x\rangle|^2 = \mathrm{Tr}(x^* a_k a_k^* x) = \mathrm{Tr}(a_k a_k^* x x^*) = \mathrm{Tr}(A_k X) = \langle A_k, X\rangle_F,$$

where $A_k := a_k a_k^*$


(a) (b) (c) (d)

Exchanging Fourier phase while keeping the magnitude

picture: Osherovich

PhaseLift

The "intuitive" problem

$$\begin{array}{ll} \text{find} & X \\ \text{subject to} & (\mathrm{Tr}(A_k X))_{k=1}^m = (b_k)_{k=1}^m \\ & X \geq 0 \\ & \mathrm{rank}(X) = 1 \end{array}$$

gets replaced by a "convex" problem

$$\begin{array}{ll} \text{minimize} & \sout{\mathrm{rank}(X)} \quad \|X\|_* \\ \text{subject to} & (\mathrm{Tr}(A_k X))_{k=1}^m = (b_k)_{k=1}^m \\ & X \geq 0. \end{array}$$

...Matrix recovery problem!

Results

E. Candès, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. SIAM J. on Imaging Sciences 6(1), pp. 199–225, 2011
E. Candès, T. Strohmer and V. Voroninski. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. Comm. Pure and Appl. Math. 66, pp. 1241–1274, 2011

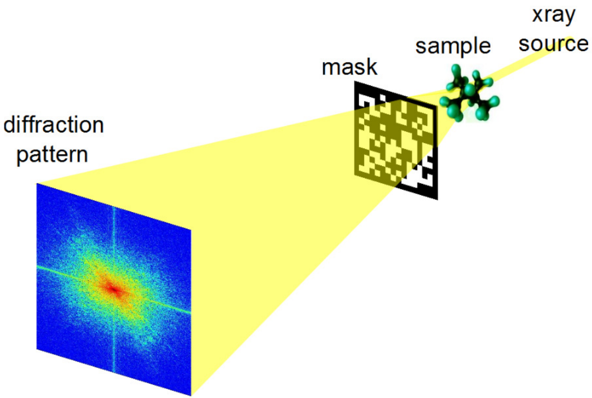
E. Candès and X. Li. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. To appear in Found. of Comp. Math.

Theorem (Candès, Li, Strohmer, Voroninski, 2011)

If a_k's are chosen independently on the sphere and $m \geq CN$ (not $N \log N$!), then the unique solution of the convex problem is $X = xx^*$ with high probability.

The reconstruction is robust w.r.t. noise!
Version for x sparse!

Implementation of random measurements



Separating features in video's

Some videos (security cameras) can be divided into two parts
- background (= "low rank" component)
- movements (= "sparse" component)

The "intuitive" program

$$\arg\min_{L,S}(\mathrm{rank} L + \lambda\|S\|_0), \quad \text{s.t. } L + S = X.$$

gets replaced by a convex program

$$\arg\min_{L,S}(\|L\|_* + \lambda\|S\|_1), \quad \text{s.t. } L + S = X.$$

E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis?, Journal of ACM 58(1), 1-37 (2009)

Data from S. Becker (Caltech)

Separating features in video's: Example

Advanced Background Subtraction

First row:

Left: original image
Middle: low-rank (i.e. predictable) component
Right: sparse component

Second row: similar, quantization effects taken into account, i.e. another term with Frobenius norm added.

ℓ_1-SVM

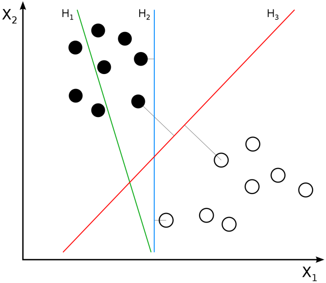
For $\{x_1, \ldots, x_m\} \subset \mathbb{R}^N$ and $\{y_1, \ldots, y_m\} \subset \{-1, 1\}$, the *Support Vector Machine* wants to separate the sets

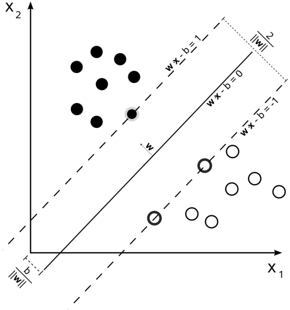
$$\{x_i : y_i = -1\} \quad \text{and} \quad \{x_i : y_i = +1\}$$

by a linear hyperplane, i.e. finds $w \in \mathbb{R}^N$ and $b \in \mathbb{R}$ with

$$\begin{aligned} \langle w, x_i \rangle - b > 0 \qquad & \text{for} \quad y_i = 1, \\ \langle w, x_i \rangle - b < 0 \qquad & \text{for} \quad y_i = -1. \end{aligned}$$

It maximizes the size of the margin around the separating hyperplane.





$$\min_{w \in \mathbb{R}^N} \sum_{i=1}^{m} (1 - y_i \langle w, x_i \rangle)_+ + \lambda \|w\|_2^2$$

ℓ_1-SVM replaces $\|w\|_2^2$ by $\|w\|_1$ - promotes the sparsity of w!

Zhu, Rosset, Hastie and Tibshirani (2003)

In bioinformatics: the (few) non-zero components of a sparse w are the "markers" of a disease

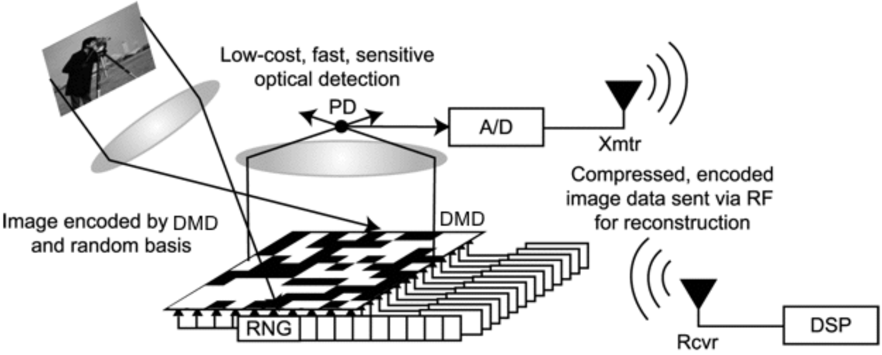
One-pixel camera

Dep. of Electrical and Computer Engineering, Rice University

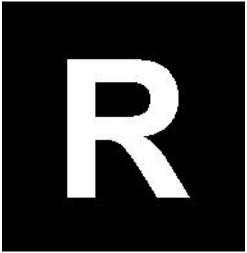
Digital micromirror device (DMD):
linear projections onto pseudorandom patterns
Random number generator (RNG):
creating random patterns
Single photon detector (PD):
“single pixel”

Advantages:
short exposure time
beyond visible spectrum
special applications (astronomy)

Setting


Low-cost, fast, sensitive
optical detection
PD
A/D
Xmtr
Compressed, encoded
image data sent via RF
for reconstruction
DMD
Image encoded by DMD
and random basis
RNG
Rcvr
DSP

Results



Original image with 16384 pixels
Image obtained by 1600 (10%) measurements
Image obtained by 3200 (20%) measurements

Magnetic Resonance Imaging

MRI exhibits several important features, which suggest using CS:

1. MRI images are naturally sparse (in an appropriate transform and domain).

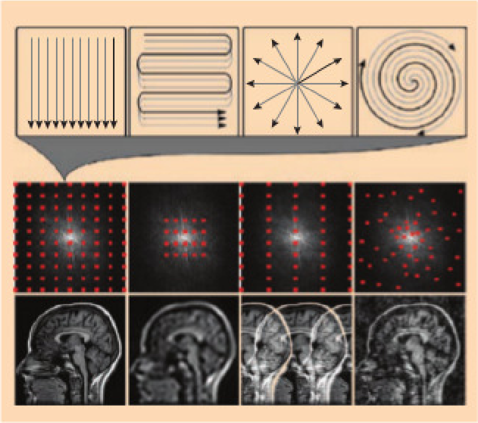
2. MRI scanners acquire encoded samples, rather than direct pixel samples.

3. Sensing is "expensive" (damage to patient, costs).

4. Processing time does not play much role.

MRI applies additional magnetic fields on top of a strong static magnetic field. The signal measured $s(t)$ is the Fourier transform of the object sampled at certain frequency $\bar{k}(t)$.

How to choose the frequencies, to allow for fast and high-quality recovery?



Different shapes in the k space correspond to sampling of different Fourier coefficients

MRI - state of the art

M. Lustig, D. Donoho and J. M. Pauly (2007)

Several groups around the world (NYU, Berkeley, MRB Würzburg & Siemens Medical Erlangen, Stanford, . . .)

Clinical testing in reach

Dream: Speed up to get videos?!

Antenna sends out a signal (radar pulse) and measures the response influenced by scattered objects.

Finite-dimensional model:
Translation and modulation operators

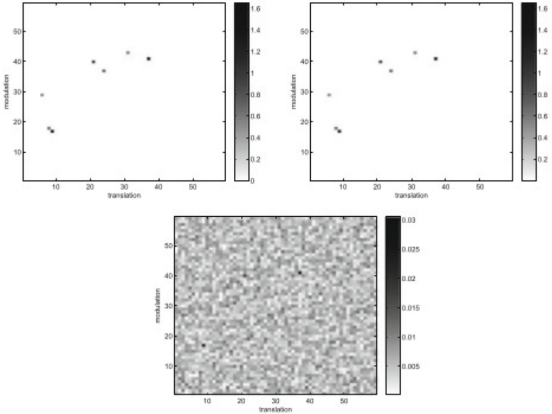
$$(T_k z)_j = z_{j-k \mod n}, \quad (M_l z)_j = \exp(2\pi i l j/n) z_j$$

Original signal is transformed to the measured signal by

$$B = \sum_{k,l=1}^{n} x_{k,l} T_k M_l$$

...we expect the coefficient vector x to be sparse.
"Random design" of the radial phase is often replaced by Alltop window:

$$g_l = \exp(2\pi i l^3/m), \quad l = 1, \ldots, m.$$



Top left: 7-sparse coef. vector in translation-modulation plane, top right: reconstruction by ℓ_1-minimization with Alltop window, bottom: reconstruction by ℓ_2; source: Foucart, Rauhut

Literature

- S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Birkhäuser/Springer 2013
    - Recent, but standard textbook
    - Detailed presentation
- H. Boche, R. Calderbank, G. Kutyniok, and J. V., *A Survey of compressed sensing*, Birkhäuser/Springer, 2015.
    - Short survey
    - 25 pages of basic theory
    - 15 pages of extensions
    - The most important proofs simplified as much as possible
    - Freely available
- Video-lecture of E. Candès from ICM 2014, available on youtube

Thank you for your attention!