Recent Developments in Compressed Sensing

M. Vidyasagar

Distinguished Professor, IIT Hyderabad
m.vidyasagar@iith.ac.in, www.iith.ac.in/~m_vidyasagar/

ISL Seminar, Stanford University, 19 April 2018
Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm
Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm
Compressed sensing refers to the recovery of “high-dimensional but low-complexity” entities from a limited number of measurements.

**Examples:** High dimensional but sparse (or nearly sparse) vectors, large images with gradients (sharp changes) in only a few pixels, large matrices of low rank, partial realization problem in control theory.

**Manuscript:** *An Introduction to Compressed Sensing* to be published by SIAM (Society for Industrial and Applied Mathematics)

**Note:** Talk will focus only on vector recovery, not matrix recovery.
High level objective of compressed sensing: Recover an unknown sparse or nearly sparse vector $x \in \mathbb{R}^n$ from $m \ll n$ linear measurements of the form $y = Ax$, $A \in \mathbb{R}^{m \times n}$.

Sparse Regression: Given an underdetermined set of linear equations $y = Ax$, where $A \in \mathbb{R}^{m \times n}, y \in \mathbb{R}^m$ are given, find the most sparse solution for $x$.

Difference: In sparse regression, $A, y$ are given, and there need not be a “true but unknown” $x$. In compressed sensing, the matrix $A$ can be chosen by the user.
Illustrative Application: Decoding Linear Codes

Caution: Nonstandard notation!

Suppose $H \in \mathbb{F}_2^{m \times n}$ is the parity check matrix of a code. So $u \in \mathbb{F}_2^n$ is a code word if and only if $Hu = 0$. Suppose that $u$ is transmitted on a noisy channel, and the received signal is $v = u \oplus x$, where $x$ is a binary error vector (with limited support).

The vector $y = Hv = H(u + x) = Hx$ is called the “syndrome” in coding theory. The problem is to determine the most sparse $x$ that satisfies $y = Hx$. 
Illustrative Application: Maximum Hands-Off Control

Given a linear system

\[ x_{t+1} = Ax_t + Bu_t, \quad x_0 \neq 0, \]

and a final time \( T \), find the most sparse control sequence \( \{u_t\}_{t=0}^{T-1} \) such that \( x_T = 0 \) and \( |u_t| \leq 1 \) for all \( t \). Note that we want

\[ x_T = A^T x_0 + \sum_{t=0}^{T-1} A^{T-1-t} Bu_t = 0. \]

So we want the most sparse solution of

\[ \sum_{t=0}^{T-1} A^{T-1-t} Bu_t = -A^T x_0 \]

while satisfying the constraint \( |u_t| \leq 1 \) for all \( t \).
Notation: For an integer $n$, $[n]$ denotes $\{1, \ldots, n\}$.

If $x \in \mathbb{R}^n$, define its “support” as

$$\text{supp}(x) := \{i \in [n] : x_i \neq 0\}.$$ 

Given an integer $k$, define the set of $k$-sparse vectors as

$$\Sigma_k := \{x \in \mathbb{R}^n : |\text{supp}(x)| \leq k\}.$$ 

Given a norm $\| \cdot \|$ on $\mathbb{R}^n$, and an integer $k$, define the $k$-sparsity index of $x$ as

$$\sigma_k(x, \| \cdot \|) := \min_{z \in \Sigma_k} \|x - z\|.$$
Problem Formulation

Define \( A \in \mathbb{R}^{m \times n} \) as the “measurement map,” and \( \Delta : \mathbb{R}^m \to \mathbb{R}^n \) as the “decoder map.”

Measurement vector \( y = Ax \) or \( y = Ax + \eta \) (noisy measurements).

**Definition**

The pair \((A, \Delta)\) achieves **robust sparse recovery** of order \( k \) if there exist constants \( C \) and \( D \) such that

\[
\| x - \Delta(Ax + \eta) \|_2 \leq C\sigma_k(x, \| \cdot \|_1) + D\epsilon,
\]

where \( \epsilon \) is an upper bound for \( \| \eta \|_2 \).
Implications

In particular, robust sparse recovery of order $k$ implies

- With $k$-sparse vectors and noise-free measurements, we get
  \[ \Delta(Ax) = x, \forall x \in \Sigma_k, \]
  or exact recovery of $k$-sparse vectors.

- With $k$-sparse vectors and noisy measurements, we get
  \[ \|x - \Delta(Ax + \eta)\|_2 \leq D\|\eta\|_2, \]
i.e., residual error comparable to that achievable by an “oracle” that knows the support of $x$. 

Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm
Let $\|x\|_0$ denote the “$\ell_0$-norm,” i.e., the number of nonzero components of $x$.

**Sparse regression problem:** Find most sparse solution of $y = Ax$.

$$\hat{x} = \arg\min_z \|z\|_0 \text{ s.t. } Az = y.$$  

This problem is NP-hard!

So we replace $\| \cdot \|_0$ by its “convex envelope” (the largest convex function that is dominated by $\| \cdot \|_0$), which is $\| \cdot \|_1$. The problem now becomes

$$\hat{x} = \arg\min_z \|z\|_1 \text{ s.t. } Az = y.$$

This is called “basis pursuit” by Chen-Donoho-Saunders (1991).

This problem is tractable. But when does it solve the original problem?
Restricted Isometry Property

References: Candès-Tao (2005) and other papers by Candès, Donoho and co-authors.

Definition

A matrix $A \in \mathbb{R}^{m \times n}$ satisfies the **restricted isometry property (RIP)** of order $k$ with constant $\delta_k$ if

$$(1 - \delta_k) \|u\|_2^2 \leq \|Au\|_2^2 \leq (1 + \delta_k) \|u\|_2^2, \quad \forall u \in \Sigma_k.$$  

**Interpretation:** Every set of $k$ or fewer columns of $A$ forms a near-isometry.
Main Theorem

Given $A \in \mathbb{R}^{m \times n}$ and $y = Ax + \eta$ where $\|\eta\|_2 \leq \epsilon$, define the decoder

$$\Delta(y) = \hat{x} := \arg\min_z \|z\|_1 \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$ 

Theorem (Cai-Zhang 2014) Suppose that, for some $t > 1$, the matrix $A$ satisfies the RIP of order $tk$ with constant $\delta_{tk} < \sqrt{(t - 1)/t}$. Then $(A, \Delta)$ achieves robust sparse recovery of order $k$.

Theorem (Cai-Zhang 2014) For $t \geq 4/3$, the above bound is tight.
Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm

M. Vidyasagar
Recent Developments in Compressed Sensing
Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm
Let $X$ be a zero-mean, unit-variance sub-Gaussian random variable. This means that, for some constants $\gamma, \zeta$, we have that

$$\Pr\{|X| > t\} \leq \gamma \exp(-\zeta t^2), \forall t > 0.$$ 

Let $\Phi \in \mathbb{R}^{m \times n}$ consist of independent samples of $X$, and define $A = (1/\sqrt{m})\Phi$. Then $A$ satisfies the RIP of order $k$ with high probability (which can be quantified).
Suppose an integer $k$ and real numbers $\delta, \xi \in (0, 1)$ are specified, and that $A = (1/\sqrt{m})\Phi$, where $\Phi \in \mathbb{R}^{m \times n}$ consists of independent samples of a sub-Gaussian random variable $X$. Then $A$ satisfies the RIP of order $k$ with constant $\delta$ with probability $\geq 1 - \xi$ provided

$$m \geq \frac{1}{\tilde{c} \delta^2} \left( \frac{4}{3} k \ln \frac{en}{k} + \frac{14k}{3} + \frac{4}{3} \ln \frac{2}{\xi} \right).$$

Tighter bounds are available for pure Gaussian samples.
Some Observations

- With sub-Gaussian random variables, $m = O(k \ln(n/k))$ measurements suffice.
- *Any* matrix needs to have at least $m = O(k \ln(n/k))$ measurements (Kashin width property).
- Ergo, this approach is “order-optimal.”
The Stings in the Tail

1. No one bothers to specify the constant under the $O$ symbol! For values of $n < 10^4$ or so, $m > n!$ (No compression!)

2. Once a matrix is generated at random, checking whether it does indeed satisfy the RIP is NP-hard!

3. The matrices have no structure, so CPU time is enormous!

Need to look for alternate (deterministic) approaches.

Several deterministic methods exist, based on finite fields, expander graphs, algebraic coding, etc. Only a few are discussed here.
Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm

M. Vidyasagar
Recent Developments in Compressed Sensing
Suppose $A \in \mathbb{R}^{m \times n}$ is column-normalized, i.e., $\|a_j\|_2 = 1 \ \forall j \in [n]$. Then
\[
\mu(A) = \max_{i \neq j} |\langle a_i, a_j \rangle|
\]
is called the **one-column coherence** of $A$.

**Lemma**

*For each $k < 1/\mu(A) - 1$, the matrix $A$ satisfies the RIP of order $k$ with constant $\delta_k < (k - 1) \cdot \mu(A)$.***

To construct a matrix that satisfies RIP of order $k$ with constant $\delta$, we need to have
\[
(k - 1)\mu \leq \delta, \text{ or } \mu \leq \frac{\delta}{k - 1}.
\]
DeVore’s Construction

Let $q$ be a prime number or a prime power, and let $\mathbb{F}_q$ denote the corresponding finite field with $q$ elements.

Choose an integer $r \geq 3$, and let $\Pi_r$ denote the set of all polynomials of degree $r - 1$ or less over $\mathbb{F}_q$.

For each polynomial $\phi \in \Pi_r$, construct a column vector $a_\phi \in \{0, 1\}^{q^2 \times 1}$ as follows:

The vector $a_\phi \in \{0, 1\}^{q^2 \times 1}$ consists of $q$ blocks of $q \times 1$ binary vectors, each vector containing exactly one “1”, evaluated as follows.
DeVore’s Construction (Cont’d)

- Enumerate the elements of $\mathbb{F}_q$ in some order. If $q$ is a prime number, then $\{0, 1, \ldots, q - 1\}$ is natural.
- Let the indeterminate $x$ vary over $\mathbb{F}_q$. Suppose $x$ is the $l$-th element of $\mathbb{F}_q$ (in the chosen ordering), and that $\phi(x)$ is the $i$-th element of $\mathbb{F}_q$. Then the $l$-th block of $a_\phi$ has a “1” in row $i$ and zeros elsewhere.

**Example:** Let $q = 3$, $\mathbb{F}_q = \{0, 1, 2\}$, $r = 3$, and $\phi(x) = 2x^2 + 2x + 1$. Then $\phi(0) = 1$, $\phi(1) = 2$, and $\phi(2) = 1$. Therefore

$$a_\phi = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^T.$$

Define $A \in \{0, 1\}^{q^2 \times q^r}$ by

$$A = [a_\phi, \phi \in \Pi_r].$$
DeVore’s Theorem

**Theorem**

*(DeVore 2007)* If \( \phi, \psi \) are distinct polynomials in \( \Pi_r \), then

\[
\langle a_\phi, a_\psi \rangle \leq r - 1.
\]

**Corollary**

*(DeVore 2007)* With \( A \) as above, the column-normalized matrix \( (1/q)A \) satisfies the RIP of order \( k \) with constant \( \delta_k = (r - 1)/q \).
Number of Measurements Using the DeVore Matrix

We want \( \delta_{tk} < \sqrt{(t - 1)/t} \). Choose \( t = 1.5 \) (optimal choice), \( \delta = 0.5 < 1/\sqrt{3} \). Also choose \( r = 3 \). Then we need \( m = q^2 \) measurements where

\[
q = \max \left\{ \left\lceil 6k - 4 \right\rceil_p, n^{1/3} \right\}.
\]

Here \( \left\lceil s \right\rceil_p \) denotes the smallest prime number \( > s \).

Number of measurements is \( O(\max\{k^2, n^{1/3}\}) \), but in practice is smaller than with probabilistic methods.

It is also much faster due to the sparsity and binary nature of \( A \).
Let $p$ be a prime, and $\mathbb{Z}_p = \mathbb{Z}/(p)$. For each $x, y \in \mathbb{Z}_p$, define $C_{x,y}: \mathbb{Z}_p \to \mathbb{Z}_p$ as follows:

$$C_{x,y}(t) = (-1)^{ty} \exp[i\pi(2x + yt)t/p].$$

Define $C \in \mathbb{C}^{p \times p^2}$ by varying $t$ over $\mathbb{Z}_p$ to generate the rows, and $x, y$ over $\mathbb{Z}_p$ to generate the columns.

The matrix $C$ contains only various $p$-th roots of unity.
Properties of the Chirp Matrix

Theorem

Suppose $p$ is a prime number. Then

$$|\langle C_{x_1,y_1}, C_{x_2,y_2} \rangle|^2 = \begin{cases} \frac{p^2}{p} & \text{if } x_1 = x_2, y_1 = y_2, \\ p & \text{if } y_1 \neq y_2, \\ 0 & \text{if } y_1 = y_2, x_1 \neq x_2. \end{cases}$$

Thus $A = \left(1/\sqrt{p}\right)C$ has $\mu(A) = 1/\sqrt{p}$.

Number of measurements $m = \lceil (3k - 2)^2 \rceil_p$.

Compare with $m = \left(\lceil 6k - 4 \rceil_p \right)^2$ for DeVore's method, which is roughly four times larger.
### Sample Complexity Estimates

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Probabilistic</th>
<th>Deterministic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$ $k$</td>
<td>$m_G$ $m_{SG}$ $m_A$</td>
<td>$m_D$ $m_C$</td>
</tr>
<tr>
<td>10,000 5</td>
<td>5,333 28,973 3,492</td>
<td>841 197</td>
</tr>
<tr>
<td>10,000 10</td>
<td>8,396 48,089 5,796</td>
<td>3,481 787</td>
</tr>
<tr>
<td>100,000 10</td>
<td>10,025 57,260 6,901</td>
<td>3,481 787</td>
</tr>
<tr>
<td>100,000 20</td>
<td>17,781 104,733 12,622</td>
<td>16,129 3,371</td>
</tr>
<tr>
<td>1,000,000 5</td>
<td>7,009 38,756 4,671</td>
<td>10,201 1,009</td>
</tr>
<tr>
<td>1,000,000 30</td>
<td>30,116 177,635 21,407</td>
<td>32,041 7,753</td>
</tr>
<tr>
<td>1,000,000 50</td>
<td>47,527 283,042 34,110</td>
<td>94,249 21,911</td>
</tr>
<tr>
<td>1,000,000 100</td>
<td>88,781 534,210 64,378</td>
<td>358,801 88,807</td>
</tr>
</tbody>
</table>

For “pure” Gaussian, sub-Gaussian, bipolar random variables, DeVore and chirp constructions.
Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm
Let $n = 10^4$, $k = 6$, and construct a $k$-sparse vector $x_0 \in \mathbb{R}^n$.

$$\text{supp}(x_0) = \{887, 4573, 4828, 5779, 9016, 9694\},$$

$$\begin{bmatrix} 0.6029 \\ -0.3323 \\ -0.7458 \\ 0.1071 \\ 0.3198 \\ -0.5214 \end{bmatrix}, \| (x_0)_S \|_1 = 2.6293.$$
Construction of Vectors (Cont’d)

Construct non-sparse vectors

\[ x_i = x_0 + \epsilon_i N(0, 1), \quad i = 1, 2, 3, \]

where \( \epsilon_1 = 0.02, \epsilon_2 = 0.002, \epsilon_3 = 0.0002 \). However, the components of \( x_0 \) belonging to the set \( S \) were not perturbed. Thus

\[
\sigma_6(x_1, \| \cdot \|_1) = 159.5404, \quad \sigma_6(x_2, \| \cdot \|_1) = 15.95404,
\]

\[
\sigma_6(x_1, \| \cdot \|_1) = 1.595404,
\]

Compare with \( \| (x_0)_S \|_1 = 2.6293 \).
Figure: The “true” vector $x_1$ with $n = 10^4$, $k = 6$. It consists of a $k$-sparse vector perturbed by additive Gaussian noise with variance $\epsilon_1 = 0.02$. 

**Problem Formulation**

Approach Based on $\ell_1$-Norm Minimization

Construction of Measurement Matrices

Numerical Examples

Statistical Recovery

A Non-Iterative Algorithm
For each method, a corresponding number of measurements $m$ was chosen, and a measurement matrix $A \in \mathbb{R}^{m \times n}$ was constructed. Each component of $Ax$ was perturbed by additive Gaussian noise with zero mean and standard deviation of 0.01. The $\ell_2$-norm of the error was estimated using this fact.

For each method, an estimate $\hat{x}$ was constructed as

$$\hat{x} = \arg\min_{z} \|z\|_1 \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$
Recovery of Exactly Sparse Vector

- For using a measurement matrix consisting of random Gaussian samples, the number of samples is 5,785.
- For DeVore’s matrix, the prime number $q = 37$, and $m = q^2 = 1,369$.
- For the Chirp matrix method, the prime number $p = 257$, and $m = p = 257$.
- The CPU time was 14 seconds with DeVore’s matrix, four minutes with the Chirp matrix, and four hours with Gaussian samples. This is because the Gaussian samples have no structure.
- All three measurement matrices with $\ell_1$-norm minimization recovered $x_0$ perfectly.
Recovery of Non-Sparse Vector Using DeVore’s Matrix

Figure: The true vector $x_1$, and the corresponding recovered vector using DeVore’s matrix, illustrating “support recovery”.

M. Vidyasagar

Recent Developments in Compressed Sensing
Figure: The true vector $x_1$, and the corresponding recovered vector using the chirp matrix. Again the support is recovered.
Problem Formulation
Approach Based on $\ell_1$-Norm Minimization
Construction of Measurement Matrices
Numerical Examples
Statistical Recovery
A Non-Iterative Algorithm

Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm
Until now we have studied *guaranteed* recovery of *all* sufficiently sparse vectors, and derived sufficient conditions.

What happens if we settle for *statistical* recovery of all but a fraction $1 - \epsilon$ of sparse vectors, with respect to a suitable probability measure?

The number of samples reduces drastically, from $m = O(k \log(n/k))$, to $m = O(k)$.

A really superficial overview is given in the next few slides.
Encoder can be nonlinear as well as the decoder.
Unknown $x$ is generated according to a known probability $p_X$ (and is sparse with high probability)
Algorithm is expected to work with high probability.

Statistical recovery is possible if and only if

$$m \geq n \bar{d}(p_X) + o(n),$$

where $\bar{d}(p_X)$ is the upper Rényi Information Dimension and is $O(k/n)$. 

(Reference: Wu-Verdu, T-IT 2012)
Approximate Message Passing

References: Several papers by Donoho et al.

An iterative alternative to $\ell_1$-norm minimization. Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is smooth, and define $\phi : \mathbb{R}^n \to \mathbb{R}^n$ componentwise. Let $A$ consist of i.i.d. samples of normal Gaussians. Set $x^0 = 0$, and then

$$x^{t+1} = \phi(A^T w^t + x^t),$$

$$w^t = y - Ax^t + \frac{1}{\delta}w^{t-1} (\phi'(A^T w^{t-1} + x^{t-1})), $$

where $\phi'$ denotes the derivative of $\phi$.

Phase transitions in the $\delta = m/n$, and $\rho = k/m$ space are comparable to those with $\ell_1$-norm minimization.
Bound Based on Descent Cone

Reference: Amelunxen et al., 2014.

Measurement is linear: $y = Ax$, consisting of i.i.d. samples of normal Gaussians, and decoder is

$$\hat{x} = \arg\min_z f(z) \text{ s.t. } y = Az,$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex. Define the descent cone of $f$ as

$$D(f, x) := \bigcup_{\tau > 0} \{ h \in \mathbb{R}^n : f(x + \tau h) \leq f(x) \}.$$

Define the statistical dimension $\delta$ of a cone.
Theorem

Define \( a(\epsilon) := \sqrt{8 \log(4/\epsilon)} \). With all other symbols as above, if

\[
m \leq \delta(\mathcal{D}(f, x)) - a(\epsilon) \sqrt{n},
\]

then the decoding algorithm fails with probability \( \geq 1 - \epsilon \). If

\[
m \geq \delta(\mathcal{D}(f, x)) + a(\epsilon) \sqrt{n},
\]

then the decoding algorithm succeeds with probability \( \geq 1 - \epsilon \).
Outline

1. Problem Formulation
2. Approach Based on $\ell_1$-Norm Minimization
3. Construction of Measurement Matrices
   - Probabilistic Construction
   - Deterministic Construction
4. Numerical Examples
5. Statistical Recovery
6. A Non-Iterative Algorithm
Summary of the Method

- Based on expander graphs, and a part of doctoral research of Mahsa Lotfi.
- Unlike $\ell_1$-norm minimization, this algorithm is noniterative – one simply “reads off” the unknown vector! Hence hundreds of times faster than $\ell_1$-norm minimization.
- The measurement matrix is the same as the DeVore construction, but with about half of the number of measurements.
- Works even with “nearly” sparse vectors, and also with burst measurement errors.
- Noise model is similar to that in error-correcting coding.
The Measurement Matrix (Same as DeVore)

**Notation:** \( [s]_p \) denotes the smallest *prime* number \( \geq s \).

Given integers \( n \) (dimension of unknown vector) and \( k \) (sparsity count), choose a prime number \( q \) such that

\[
q = \lceil 4k - 2 \rceil_p, \quad n \leq q^r.
\]

Form DeVore’s measurement matrix \( A \in \{0, 1\}^{q^2 \times q^r} \) as before. Define the measurement vector \( y = Ax \).

**Recall:** For \( \ell_1 \)-norm minimization, \( q = \lceil 6k - 4 \rceil_p \), or about 1.5 times higher. Since \( m = q^2 \), \( \ell_1 \)-norm minimization requires roughly \( 1.5^2 = 2.25 \) times more measurements.
Key Idea: The “Reduced” Vector

The measurement \( y = Ax \in \mathbb{R}^{q^2} \). For each index \( j \in [n] \), construct a “reduced” vector \( y_j \in \mathbb{R}^q \) as follows:

Each column of \( A \) contains \( q \) elements of 1 and the rest are zero. For each column \( j \in [q^r] = \{1, \ldots, q^r\} \), identify the \( q \) indices \( v_1(j), \ldots, v_q(j) \) corresponding to the locations of the “1” entries. Define the “reduced” vector \( \bar{y} \in \mathbb{R}^q \) as

\[
\bar{y}_j = [y_{v_1(j)} \cdots y_{v_q(j)}]^{\top}.
\]

Note that the reduced vector picks off different rows of \( y \) for each column index \( j \).
Suppose $q = 2$, and

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \in \{0, 1\}^{4 \times 6}.$$ 

Suppose $y \in \mathbb{R}^4$. Then

$$\bar{y}_1 = (y_1, y_2), \bar{y}_2 = (y_2, y_3), \bar{y}_3 = (y_3, y_4),$$

$$\bar{y}_4 = (y_2, y_4), \bar{y}_5 = (y_1, y_4), \bar{y}_6 = (y_1, y_3).$$
Suppose $x \in \Sigma_k$, and let $y = Ax$. Then

1. If $j \notin \text{supp}(x)$, then $\bar{y}_j$ contains no more than $k(r - 1)$ nonzero components.

2. If $j \in \text{supp}(x)$, then at least $q - (k - 1)(r - 1)$ elements of $\bar{y}_j$ equal $x_j$. 
The New Algorithm

Choose \( q = \lceil 2(2k - 1) \rceil_p \), and construct the DeVore matrix \( A \). Suppose \( x \in \Sigma_k \) and let \( y = Ax \). Run through \( j \) from 1 to \( n \). For each \( j \), check to see how many nonzero entries \( \bar{y}_j \) has.

- If \( \bar{y}_j \) has fewer than \( (q - 1)/2 \) nonzero entries, then \( j \notin \text{supp}(x) \).
- If \( \bar{y}_j \) has more than \( (q + 1)/2 \) nonzero entries, then \( j \in \text{supp}(x) \). In this case, at least \( (q + 1)/2 \) entries of \( \bar{y}_j \) will be equal, and that value is \( x_j \).

**Note:** No optimization or iterations are required! The nonzero components of \( x \) are simply read off!

The method extends to the case of “burst noise,” where the noise has limited support.
Suppose $x \in \Sigma_k$, and that $y = Ax + \eta$ where $\|\eta\|_0 \leq M$ (burst noise). Then

1. If $j \notin \text{supp}(x)$, then $\bar{y}_j$ contains no more than $k(r - 1) + M$ nonzero components.

2. If $j \in \text{supp}(x)$, then $\bar{y}_j$ contains at least $q - [(k - 1)(r - 1) + M]$ components that are all equal to $x_j$.

Extensions to the case where $x$ is “nearly sparse” but not exactly sparse can also be proven.
Comparison of Sample Complexity and Speed

All methods require \( m = q^2 \) measurements.

<table>
<thead>
<tr>
<th>Method</th>
<th>( \ell_1 )-norm Min.</th>
<th>Expander Graphs</th>
<th>New Alg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bound: ( q \geq )</td>
<td>( 6k - 4 )</td>
<td>( 8(2k - 1) )</td>
<td>( 4k - 2 )</td>
</tr>
<tr>
<td>( q ) with ( k = 6 )</td>
<td>37</td>
<td>89</td>
<td>29</td>
</tr>
<tr>
<td>( m ) with ( k = 6 )</td>
<td>1,369</td>
<td>7,921</td>
<td>841</td>
</tr>
</tbody>
</table>

Table: Number of measurements for various approaches with \( n = 20,000 \).

New algorithm is about 200 times faster than \( \ell_1 \)-norm minimization and 1,000 times faster than expander graph (Xu-Hassibi) algorithm.
We chose \( n = 20,000 \) and \( k = 6 \), constructed \( A \) with \( q = 29 \) for new algorithm and \( q = 37 \) for \( \ell_1 \)-norm minimization. We chose a random vector \( x \in \Sigma_k \) and constructed the measurement vector \( Ax \). Then we chose \( M = 6 \), and perturbed the measurement vector \( Ax \) in \( M \) locations with a random number of variance \( \alpha \).

As \( \alpha \) is increased, the new algorithm recovers \( x \) \textit{perfectly} no matter how large \( \alpha \) is, whereas \( \ell_1 \)-norm minimization fails to recover the true \( x \).
### Computational Results

<table>
<thead>
<tr>
<th>Alpha</th>
<th>New Algorithm</th>
<th>( \ell_1 )-norm minimization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Err.</strong></td>
<td><strong>Time</strong></td>
</tr>
<tr>
<td>10(^{-5})</td>
<td>0</td>
<td>0.1335</td>
</tr>
<tr>
<td>10(^{-4})</td>
<td>0</td>
<td>0.1325</td>
</tr>
<tr>
<td>10(^{-3})</td>
<td>0</td>
<td>0.1336</td>
</tr>
<tr>
<td>10(^{-2})</td>
<td>0</td>
<td>0.1357</td>
</tr>
<tr>
<td>10(^{-1})</td>
<td>0</td>
<td>0.1571</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.1409</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0.1494</td>
</tr>
</tbody>
</table>

**Table:** Performance of new algorithm and \( \ell_1 \)-norm minimization with additive burst noise
Some Topics Not Covered

- Recovery of group sparse vectors
- Matrix recovery
- Alternatives to the $\ell_1$-norm
- One-bit compressed sensing
- Applications to image recovery, control systems

All of these are covered in the book.
Some Interesting Open Questions

**Caution:** Heavily biased by my own preference for deterministic approaches!

- Is there a deterministic procedure for designing measurement matrices that is *order-optimal* with $m = O(k \log n)$?
- Can deterministic *vector recovery* be extended seamlessly to problems of *matrix recovery*?
- Can the partial realization problem of control theory (which is a problem of completing a Hankel matrix to minimize its rank) be tackled as a matrix completion problem?
Questions?