Information-Theoretically Optimal
Compressed Sensing via Spatial Coupling and
Approximate Message Passing

Adel Javanmard
with David Donoho and Andrea Montanari

Stanford University

November 10, 2012
General problem

\[ y = Ax + \text{noise}, \]

- \( x \) high-dimensional but highly structured
- How many linear measurements are needed?
Normalization

\[ w \sim N(0, \sigma^2 I_{m \times m}) \]

\[ m, n \to \infty, \ m/n = \delta \]

\[ A = [A_1 | \cdots | A_n], \quad \|A_i\|_2 = \Theta(1) \]
**Compressed sensing: Basic insights**

Donoho, Candés, Romberg, Tao, Indyk, Gilbert, ... [2005-...]

<table>
<thead>
<tr>
<th>Structure</th>
<th>$|x|_0 \leq k$ adversarial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>$m = C k \log(n/k)$</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>Convex optimization</td>
</tr>
<tr>
<td>Measurements</td>
<td>Random isotropic vectors</td>
</tr>
<tr>
<td>Robustness</td>
<td>$\text{MSE} \leq C\sigma^2$</td>
</tr>
</tbody>
</table>
Compressed sensing: Basic insights

Donoho, Candés, Romberg, Tao, Indyk, Gilbert, ... [2005-...]

<table>
<thead>
<tr>
<th>Structure</th>
<th>$|x|_0 \leq k$ adversarial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>$m = C k \log(n/k)$</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>Convex optimization</td>
</tr>
<tr>
<td>Measurements</td>
<td>Random isotropic vectors</td>
</tr>
<tr>
<td>Robustness</td>
<td>$\text{MSE} \leq C\sigma^2$</td>
</tr>
</tbody>
</table>

Is this the optimal compression rate?
This paper

<table>
<thead>
<tr>
<th>Structure</th>
<th>$x = x_{\text{discr}} + x_{\text{other}}$; $|x_{\text{other}}|_0 \leq k$ oblivious</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>$m = k + o(n)$</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>Bayesian AMP</td>
</tr>
<tr>
<td>Measurements</td>
<td>Spatially coupled matrices</td>
</tr>
<tr>
<td>Robustness</td>
<td>$\text{MSE} \leq C(x)\sigma^2$</td>
</tr>
</tbody>
</table>
This paper

| Structure → | $x = x_{\text{discr}} + x_{\text{other}}$; $\|x_{\text{other}}\|_0 \leq k$ oblivious |
| Rate → | $m = k + o(n)$ |
| Reconstruction → | Bayesian AMP |
| Measurements → | Spatially coupled matrices |
| Robustness → | $\text{MSE} \leq C(x)\sigma^2$ |
Outline

- A toy example (random signal).
  - Results.
    - ‘Spatially coupled’ sensing matrices
    - How does spatial coupling work?
    - Bayes-optimal AMP
  - Proof technique.
    - State evolution
- Supercooling.
Outline

▶ A toy example (random signal).

▶ Results.
  ▶ ‘Spatially coupled’ sensing matrices
  ▶ How does spatial coupling work?
  ▶ Bayes-optimal AMP

▶ Proof technique.
  ▶ State evolution

▶ Supercooling.
Outline

- A toy example (random signal).

- Results.
  - ‘Spatially coupled’ sensing matrices
  - How does spatial coupling work?
  - Bayes-optimal AMP

- Proof technique.
  - State evolution

- Supercooling.
Outline

▶ A toy example (random signal).

▶ Results.
  ▶ ‘Spatially coupled’ sensing matrices
  ▶ How does spatial coupling work?
  ▶ Bayes-optimal AMP

▶ Proof technique.
  ▶ State evolution

▶ Supercooling.
A toy example (random signal)
A toy example (random signal)

\[ x = (x_1, \ldots, x_n), \quad x_i \sim \text{i.i.d. } p_X, \]

\[ y = Ax, \quad y \in \mathbb{R}^m, \]

\[ p_X = 0.2 \delta_0 + 0.3 \delta_1 + 0.2 \delta_{-1} + 0.2 \delta_3 + 0.1 \text{Uniform}(-2, 2). \]

\( p_X \) is known! Non-universal!
A toy example (random signal)

\[ x = (x_1, \ldots, x_n), \quad x_i \sim \text{i.i.d. } p_X, \]

\[ y = Ax, \quad y \in \mathbb{R}^m, \]

\[ p_X = 0.2 \delta_0 + 0.3 \delta_1 + 0.2 \delta_{-1} + 0.2 \delta_3 + 0.1 \text{Uniform}(-2, 2). \]
A toy example (random signal)

\[ x = (x_1, \ldots, x_n), \quad x_i \sim_{\text{i.i.d.}} p_X, \]
\[ y = Ax, \quad y \in \mathbb{R}^m, \]

\[ p_X = 0.2 \delta_0 + 0.3 \delta_1 + 0.2 \delta_{-1} + 0.2 \delta_3 + 0.1 \text{Uniform}(-2, 2). \]

\( p_X \) is known! Non-universal!
How many measurements are needed?

\[ p_X = 0.2\delta_0 + 0.3\delta_1 + 0.2\delta_{-1} + 0.2\delta_3 + 0.1 \text{Uniform}(-2, 2). \]

- Classical compressed sensing: \( m = 0.97n + o(n) \)
  (Donoho 2006, universal, Donoho-Maleki-M. 2011 uniformly robust)

- This talk: \( m = 0.1n + o(n) \)
  (non-universal, robust)
How many measurements are needed?

\[ p_X = 0.2 \delta_0 + 0.3 \delta_1 + 0.2 \delta_{-1} + 0.2 \delta_3 + 0.1 \text{Uniform}(-2, 2). \]

- Classical compressed sensing: \( m = 0.97 n + o(n) \)
  (Donoho 2006, universal, Donoho-Maleki-M. 2011 uniformly robust)

- This talk: \( m = 0.1 n + o(n) \)
  (non-universal, robust)
How many measurements are needed?

\[ p_X = 0.2 \delta_0 + 0.3 \delta_1 + 0.2 \delta_{-1} + 0.2 \delta_3 + 0.1 \text{Uniform}(-2, 2). \]

- Classical compressed sensing: \( m = 0.97 n + o(n) \)
  (Donoho 2006, universal, Donoho-Maleki-M. 2011 uniformly robust)

- This talk: \( m = 0.1 n + o(n) \)
  (non-universal, robust)
Definition (Renyi’s Information Dimension)

For $X \sim p_X$, let $\langle X \rangle_m = \lfloor 2^m X \rfloor / 2^m$ be an $m$-digits rounding of $X$

$$
\overline{d}(X) \equiv \limsup_{m \to \infty} \frac{H(\langle X \rangle_m)}{m}.
$$

Alternative characterization:

1. If

$$
p_X = (1 - \varepsilon) \cdot \text{discrete} + \varepsilon \cdot \text{abs. continuous},
$$

then $\overline{d}(X) = \varepsilon$. 

**Definition (Renyi’s Information Dimension)**

For $X \sim p_X$, let $\langle X \rangle_m = \lfloor 2^m X \rfloor / 2^m$ be an $m$-digits rounding of $X$

$$\overline{d}(X) \equiv \lim_{m \to \infty} \sup_{m \to \infty} \frac{H(\langle X \rangle_m)}{m}.$$ 

**Alternative characterization:**

- If

  $$p_X = (1 - \varepsilon) \cdot \text{discrete} + \varepsilon \cdot \text{abs. continuous},$$

then $\overline{d}(X) = \varepsilon$. 
Why is this important?

Theorem (Verdú, Wu, 2010)

Under mild regularity hypotheses, non-adaptive compressed sensing is possible if and only if

\[ m > d(X) n + o(n). \]

(equivalently, \( \delta > d(X) + o(1) \)).

Shannon-theoretic argument. Exhaustive-search reconstruction :-(

Results
Two tricks

- ‘Spatially coupled’ sensing matrix.
  [Kudekar, Pfister, 2010]
  [cf. also Felstrom, Zigangirov, 1999; Kudekar, Richardson, Urbanke 2009-2011]

- AMP reconstruction, Posterior-expectation denoiser
  [Donoho, Maleki, Montanari 2009]

- Spatial coupling + MP reconstruction
  [Krzakala, Mézard, Sausset, Sun, Zdeborova, 2011]
  [no proof :-()]

Two tricks

- ‘Spatially coupled’ sensing matrix.
  [Kudekar, Pfister, 2010]
  [cf. also Felstrom, Zigangirov, 1999; Kudekar, Richardson, Urbanke 2009-2011]

- AMP reconstruction, Posterior-expectation denoiser
  [Donoho, Maleki, Montanari 2009]

- Spatial coupling + MP reconstruction
  [Krzakala, Mézard, Sausset, Sun, Zdeborova, 2011]
  [no proof :-()]

Our contributions

- Construction
- A rigorous proof
- Beyond random signals
- Robustness
Spatially coupled sensing matrix

\[ A = \begin{bmatrix}
0 & 0 & 0 & 0 & a_1 & a_2 & \ast & \ast & a_\ell & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & b_1 & b_2 & \ast & \ast & b_\ell & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & c_1 & c_2 & \ast & \ast & c_\ell & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix} \]

- independent entries
- band diagonal
- \( m, n, \ell \to \infty, \text{ with } m/n \to \delta \in (0, 1), \ell/n \to 0 \)
How does spatial coupling work?

- Coordinates of $x$
- Coordinates of $y$
How does spatial coupling work?

First few coordinates are oversampled!

Additional measurements associated to the first few coordinates
How does spatial coupling work?
How does spatial coupling work?
How does spatial coupling work?
Bayes-optimal AMP

\[
x^{t+1} = \eta_t( x^t + (Q_t \odot A)^* r^t ) ,
\]
\[
r^t = y - A x^t + b^t \odot r^{t-1} .
\]

Q_t, b_t explicitly given normalizations

\[
\eta_t(y) \equiv \mathbb{E}\{X | X + r_t Z = y\}
\]

(reduces to simple expression in most cases)
Bayes-optimal AMP

\[ x^{t+1} = \eta_t(x^t + (Q_t \circ A)^* r^t), \]
\[ r^t = y - Ax^t + b_t \odot r^{t-1}. \]

\(Q_t, b_t\) explicitly given normalizations

\[ \eta_t(y) \equiv \mathbb{E}\{X|X + \tau_t Z = y\} \]

(reduces to simple expression in most cases)
A theorem

Theorem (Donoho, Javanmard, Montanari, 2011)

Let \{(x(n), y(n))\}_{n \geq 0} be a sequence of instances and assume the empirical distributions converge \( p_x(n) \to p_X \).

Using Gaussian spatially-coupled matrices, Bayes-optimal AMP recovers \( x(n) \) with high probability from

\[
m > \bar{d}(X) n + o(n),
\]

noiseless measurements.

Further, if \( m > \bar{D}(X) n + o(n) \), and measurements are noisy

\[
\text{MSE} \leq C(p_X) \sigma^2.
\]

\(^a\bar{D}(X) = \bar{d}(X)\) in most cases.
A theorem

**Theorem (Donoho, Javanmard, Montanari, 2011)**

Let \( \{ (x(n), y(n)) \}_{n \geq 0} \) be a sequence of instances and assume the empirical distributions converge \( p_{x(n)} \rightarrow p_X \).

Using Gaussian spatially-coupled matrices, Bayes-optimal AMP recovers \( x(n) \) with high probability from

\[
m > \overline{d}(X) n + o(n),
\]

noiseless measurements.

Further, if\( m > \overline{D}(X) n + o(n) \), and measurements are noisy

\[
\text{MSE} \leq C(p_X) \sigma^2.
\]

\( \overline{D}(X) = \overline{d}(X) \) in most cases.
A theorem

Theorem (Donoho, Javanmard, Montanari, 2011)

Let \( \{(x(n), y(n))\}_{n \geq 0} \) be a sequence of instances and assume the empirical distributions converge \( p_{x(n)} \rightarrow p_X \).

Using Gaussian spatially-coupled matrices, Bayes-optimal AMP recovers \( x(n) \) with high probability from

\[
m > \bar{d}(X) n + o(n),
\]

noiseless measurements.

Further, if \( m > \overline{D}(X)n + o(n) \), and measurements are noisy

\[
\text{MSE} \leq C(p_X)\sigma^2.
\]

\( a \overline{D}(X) = \bar{d}(X) \) in most cases.
Theorem (Donoho, Javanmard, Montanari, 2011)

Let \( \{(x(n), y(n))\}_{n \geq 0} \) be a sequence of instances and assume the empirical distributions converge \( p_{x(n)} \rightarrow p_X \).

Using Gaussian spatially-coupled matrices, Bayes-optimal AMP recovers \( x(n) \) with high probability from

\[
m > \bar{d}(X) n + o(n),
\]

noiseless measurements.

Further, if\(^a\) \( m > \bar{D}(X)n + o(n) \), and measurements are noisy

\[
\text{MSE} \leq C(p_X)\sigma^2.
\]

\(^a\bar{D}(X) = \bar{d}(X)\) in most cases.
Proof technique
State evolution

A block Gaussian sensing matrix

\[ A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \end{bmatrix} \]

\[ x^t = \begin{bmatrix} \vdots \end{bmatrix} \]

\[ k \text{ blocks} \]

\[ MSE^{(t)} \in \mathbb{R}^k, \quad MSE^{(t)}(i) = \lim_{n \to \infty} \frac{n}{k} \left\| x_{B_i}^t - x_{B_i} \right\|^2. \]

We show a state evolution recursion:

\[ MSE^{(t+1)} = \mathcal{F}(MSE^{(t)}; p_X) \]
State evolution

A block Gaussian sensing matrix

\[ A = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \quad k \text{ blocks} \]

\[ x^t = \begin{bmatrix} x^t_1 \\ \vdots \\ x^t_k \end{bmatrix} \]

\[ \text{MSE}^{(t)} \in \mathbb{R}^k, \quad \text{MSE}^{(t)}(i) = \lim_{n \to \infty} \frac{n}{k} \| x^t_{B_i} - x_{B_i} \|^2. \]

We show a state evolution recursion:

\[ \text{MSE}^{(t+1)} = \mathcal{F}(\text{MSE}^{(t)}; p_X) \]
An illustration
Steps of the proof

- Analysis of the state evolution
- Continuum state evolution
- An energy functional $\mathcal{E}(\cdot)$
  - Fixed point of the state evolution $\Phi_\infty \rightarrow \nabla \mathcal{E}(\Phi_\infty) = 0$
Supercooling
Does the spatial coupling phenomenon survive for physically constrained sensing matrices?

I will discuss it in my talk on Thursday!

Thanks!
Question

Does the spatial coupling phenomenon survive for physically constrained sensing matrices?

I will discuss it in my talk on Thursday!

Thanks!
Does the spatial coupling phenomenon survive for physically constrained sensing matrices?

I will discuss it in my talk on Thursday!

Thanks!