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

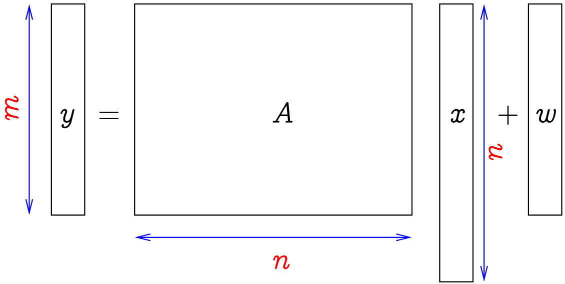
Adel Javanmard
with David Donoho and Andrea Montanari

Stanford University

November 10, 2012

# General problem

$$y = Ax + \text{noise}\,,$$




- $x$ high-dimensional but highly structured
- How many linear measurements are needed?

# Normalization

$\rightarrow$ $w \sim \mathrm{N}(0, \sigma^2 I_{m\times m})$

$\rightarrow$ $m, n \to \infty,\ m/n = \delta$

$\rightarrow$ $A = [A_1 | \cdots | A_n] \qquad \|A_i\|_2 = \Theta(1)$

# Compressed sensing: Basic insights

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

| | | |
|---|---|---|
| **Structure** | $\rightarrow$ | $\|x\|_0 \leq k$ adversarial |
| **Rate** | $\rightarrow$ | $m = C\, k \log(n/k)$ |
| **Reconstruction** | $\rightarrow$ | Convex optimization |
| **Measurements** | $\rightarrow$ | Random isotropic vectors |
| **Robustness** | $\rightarrow$ | MSE $\leq C\sigma^2$ |

Is this the optimal compression rate?

# Compressed sensing: Basic insights

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

| | | |
|---|---|---|
| **Structure** | $\rightarrow$ | $\|x\|_0 \leq k$ adversarial |
| **Rate** | $\rightarrow$ | $m = C\, k \log(n/k)$ |
| **Reconstruction** | $\rightarrow$ | Convex optimization |
| **Measurements** | $\rightarrow$ | Random isotropic vectors |
| **Robustness** | $\rightarrow$ | MSE $\leq C\sigma^2$ |

Is this the optimal compression rate?

# This paper

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

# This paper

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

# 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)

$$
\begin{aligned}
x &= (x_1, \ldots, x_n), \quad x_i \sim_{\text{i.i.d.}} p_X \,, \\
y &= Ax \,, \quad y \in \mathbb{R}^m \,,
\end{aligned}
$$

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

$p_X$ is known! Non-universal!

# A toy example (random signal)

$$
\begin{aligned}
x &= (x_1, \dots, x_n), \quad x_i \sim_{\text{i.i.d.}} p_X \,, \\
y &= Ax \,, \quad y \in \mathbb{R}^m \,,
\end{aligned}
$$

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

$p_X$ is known! Non-universal!

# A toy example (random signal)

$$
\begin{aligned}
x &= (x_1, \dots, x_n), \quad x_i \sim_{\text{i.i.d.}} p_X\,, \\
y &= Ax\,, \quad y \in \mathbb{R}^m\,,
\end{aligned}
$$

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

$p_X$ is known! Non-universal!

# How many measurements are needed?

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

- 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?

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

- 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?

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

- 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)

# What is 0.1 here?

**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 $X$

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

*Alternative characterization:*

- If

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

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

# What is 0.1 here?

**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:*

- 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*

$$m > \overline{d}(X)\, n + o(n)\,.$$

*(equivalently, $\delta > \overline{d}(X) + o(1)$).*

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

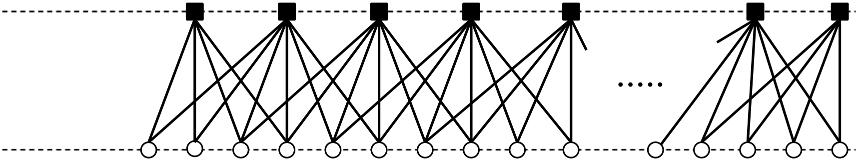
- Construction
- A rigorous proof
- Beyond random signals
- Robustness

# Spatially coupled sensing matrix

$$A = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & a_1 & a_2 & * & * & a_\ell & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & b_1 & b_2 & * & * & b_\ell & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_1 & c_2 & * & * & c_\ell & 0 & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

- $\sim$ independent entries
- $\sim$ band diagonal
- $m, n, \ell \to \infty$, with $m/n \to \delta \in (0, 1)$, $\ell/n \to 0$

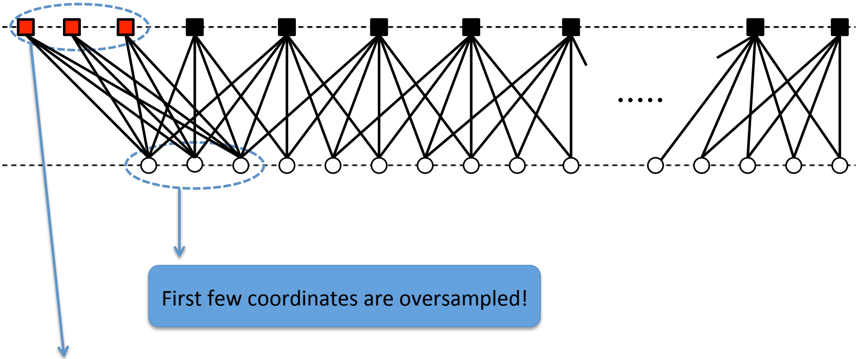
# How does spatial coupling work?



○ Coordinates of x

■ Coordinates of y

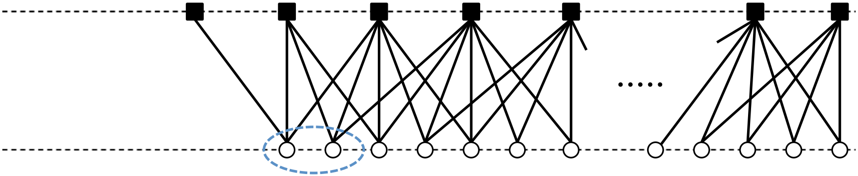
# How does spatial coupling work?



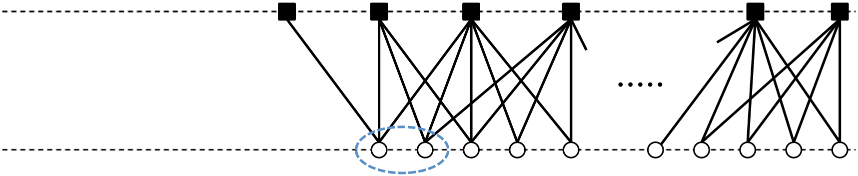
First few coordinates are oversampled!

Additional measurements associated to the first few coordinates

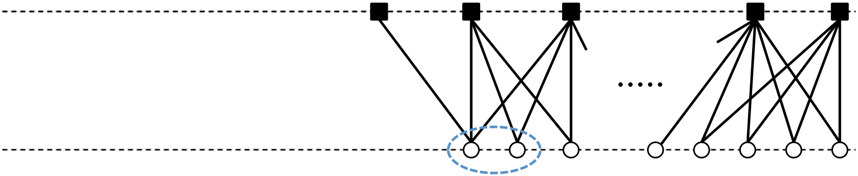
# How does spatial coupling work?



# How does spatial coupling work?



# How does spatial coupling work?



# Bayes-optimal AMP

$$
\begin{aligned}
x^{t+1} &= \eta_t(x^t + (Q_t \odot A)^* r^t)\,, \\
r^t &= y - Ax^t + \mathsf{b}_t \odot r^{t-1}\,.
\end{aligned}
$$

$Q_t, \mathsf{b}_t$ explicitly given normalizations

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

(reduces to simple expression in most cases)

# Bayes-optimal AMP

$$\begin{aligned} x^{t+1} &= \eta_t(x^t + (Q_t \odot A)^* r^t)\,, \\ r^t &= y - Ax^t + \mathsf{b}_t \odot r^{t-1}\,. \end{aligned}$$

$Q_t, \mathsf{b}_t$ 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*

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

*noiseless measurements.*

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

$$\mathrm{MSE} \leq C(p_X)\sigma^2\,.$$

[a] $\overline{D}(X) = \overline{d}(X)$ 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 $p_{x(n)} \to p_X$.*

*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[a] $m > \overline{D}(X)n + o(n)$, and measurements are noisy*

$$\mathrm{MSE} \leq C(p_X)\sigma^2\,.$$

[a] $\overline{D}(X) = \overline{d}(X)$ 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* $p_{x(n)} \to p_X$.

*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*[a] $m > \overline{D}(X)n + o(n)$, *and measurements are noisy*

$$\mathrm{MSE} \leq C(p_X)\sigma^2\,.$$

[a] $\overline{D}(X) = \overline{d}(X)$ 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* $p_{x(n)} \to p_X$.

*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*[a] $m > \overline{D}(X)n + o(n)$, *and measurements are noisy*

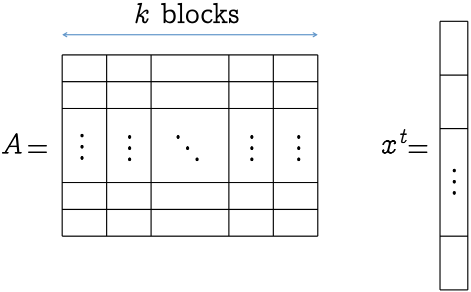
$$\mathrm{MSE} \leq C(p_X)\sigma^2\,.$$

[a] $\overline{D}(X) = \overline{d}(X)$ in most cases.

## Proof technique

# State evolution

A block Gaussian sensing matrix



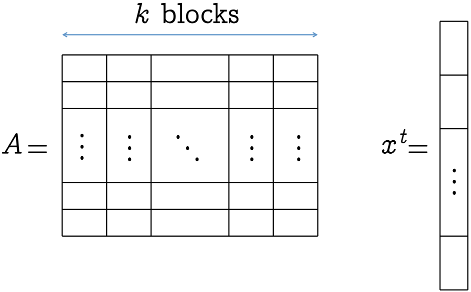

$$\mathsf{MSE}^{(t)} \in \mathbb{R}^k, \qquad \mathsf{MSE}^{(t)}(i) = \lim_{n\to\infty} \frac{n}{k} \|x^t_{B_i} - x_{B_i}\|^2.$$

We show a state evolution recursion:

$$\mathsf{MSE}^{(t+1)} = \mathcal{F}(\mathsf{MSE}^{(t)}; p_X)$$

# State evolution

A block Gaussian sensing matrix

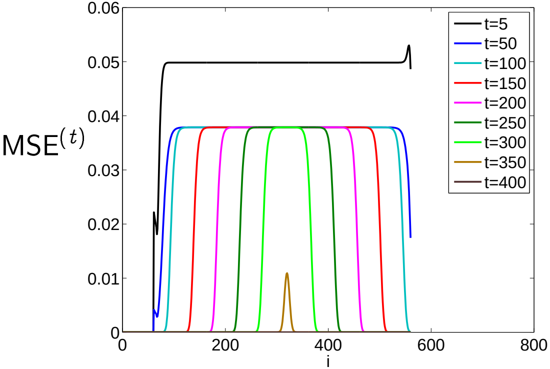



$$\mathsf{MSE}^{(t)} \in \mathbb{R}^k, \qquad \mathsf{MSE}^{(t)}(i) = \lim_{n\to\infty} \frac{n}{k} \|x^t_{B_i} - x_{B_i}\|^2.$$

We show a state evolution recursion:

$$\mathsf{MSE}^{(t+1)} = \mathcal{F}(\mathsf{MSE}^{(t)}; p_X)$$

# 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 $\Phi_\infty \to \nabla\mathcal{E}(\Phi_\infty) = 0$

# Supercooling

# Question

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!

# Question

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

I will discuss it in my talk on Thursday!

Thanks!