Subsampling at Information Theoretically Optimal Rates

Adel Javanmard, Andrea Montanari

Stanford University

July 5, 2012
A classical compressive sensing application

Sampling a random signal with sparse support in frequency domain.
Notation

- **Time domain:**

  \[ x = (x(1), x(2), \cdots, x(t), \cdots, x(n)) \in \mathbb{C}^n. \]

- **Fourier domain:**

  \[ \hat{x} = Fx, \quad F: \text{Fourier matrix} \]

  \[ \hat{x}(\omega) = \sum_{t=1}^{n} \frac{1}{\sqrt{n}} e^{-i\omega t} x(t), \quad \omega \in \{2\pi k/n\}_{k=0}^{n-1}. \]

Sparse structure: \( \hat{x} \) has \( k \) nonzero entries (\( k \ll n \)).
Notation

- **Time domain:**

\[ x = (x(1), x(2), \cdots, x(t), \cdots, x(n)) \in \mathbb{C}^n. \]

- **Fourier domain:**

\[ \hat{x} = Fx, \quad F: \text{Fourier matrix} \]

\[ \hat{x}(\omega) = \sum_{t=1}^{n} \frac{1}{\sqrt{n}} e^{-i\omega t} x(t), \quad \omega \in \{2\pi k/n\}_{k=0}^{n-1}. \]

Sparse structure: \( \hat{x} \) has \( k \) nonzero entries (\( k \ll n \)).
Sampling mechanism

\[ y_i = \langle a_i, x \rangle, \quad i = 1, \ldots, m. \]

We refer to \( m/n \) as the sampling rate.

(In time domain) \quad y = Ax.

(In frequency domain) \quad y = AF^* \hat{x} = A_F \hat{x}.
Sampling mechanism

\[ y_i = \langle a_i, x \rangle, \quad i = 1, \ldots, m. \]

We refer to \( m/n \) as the sampling rate.

(In time domain) \[ y = Ax. \]
(In frequency domain) \[ y = A \hat{F}^* \hat{x} = A \hat{F} \hat{x}. \]
Normalization

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

\[ A = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \quad \|a_i\|_2 = 1 \]
Sampling schemes:

- **Instantaneous** sampling at equispaced times $\rightarrow$ rate $= \text{Nyquist rate}$
  
  [Shannon 1948]

- **Instantaneous** sampling at random times $\rightarrow m = Ck \log n$
  
  [Candés, Romberg, Tao 2006, Candés, Plan 2011]

Our scheme:

- **Non-instantaneous** sampling at random times $\rightarrow m = k + o(n)$
Sampling schemes:

- **Instantaneous** sampling at equispaced times \( \rightarrow \) rate = Nyquist rate
  
  [Shannon 1948]

- **Instantaneous** sampling at random times \( \rightarrow m = Ck \log n \)
  
  [Candés, Romberg, Tao 2006, Candés, Plan 2011]

Our scheme:

- **Non-instantaneous** sampling at random times \( \rightarrow m = k + o(n) \)
Sampling schemes:

▶ **Instantaneous** sampling at equispaced times $\rightarrow$ rate $= \text{Nyquist rate}$

[Shannon 1948]

▶ **Instantaneous** sampling at random times $\rightarrow m = Ck \log n$

[Candés, Romberg, Tao 2006, Candés, Plan 2011]

Our scheme:

▶ **Non-instantaneous** sampling at random times $\rightarrow m = k + o(n)$
Classical compressive sensing scheme

- Measurements: sample pointwise at random times

Fourier domain: random rows of DFT matrix.
- Probes all freq. with the same weight. (Delocalized measurements)

▷ Reconstruction: Convex minimization ($\ell_1$ minimization)
Classical compressive sensing scheme

- Measurements: sample pointwise at random times

Fourier domain: random rows of DFT matrix.

- Probes all freq. with the same weight. (Delocalized measurements)

Reconstruction: Convex minimization ($\ell_1$ minimization)
Classical compressive sensing scheme

- Measurements: sample pointwise at random times

Fourier domain: random rows of DFT matrix.

- Probes all freq. with the same weight. (Delocalized measurements)

Reconstruction: Convex minimization ($\ell_1$ minimization)
Classical compressive sensing scheme

- Measurements: sample pointwise at random times

Fourier domain: random rows of DFT matrix.

- Probes all freq. with the same weight. (Delocalized measurements)

▷ Reconstruction: Convex minimization ($\ell_1$ minimization)
Our scheme

A different solution!

We ‘smear out’ the samples in the time domain

\[ \{t_1, \cdots, t_m\}, \quad \{\omega_1, \cdots, \omega_m\}, \quad \omega_i = 2\pi i/m. \]
Our scheme

A different solution!

We ‘smear out’ the samples in the time domain

\[ \{ t_1, \cdots, t_m \}, \quad \{ \omega_1, \cdots, \omega_m \}, \quad \omega_i = \frac{2\pi i}{m}. \]
Our scheme

A different solution!

We ‘smear out’ the samples in the time domain

\[ \{t_1, \ldots, t_m\}, \quad \{\omega_1, \ldots, \omega_m\}, \quad \omega_i = \frac{2\pi i}{m}. \]

modulate with \(\omega_i\)
Our scheme

A different solution!

We ‘smear out’ the samples in the time domain

\[ \{t_1, \cdots, t_m\}, \quad \{\omega_1, \cdots, \omega_m\}, \quad \omega_i = \frac{2\pi i}{m}. \]

integrate over a window (of size $\ell$) around $t_i$
Our scheme

A different solution!

We ‘smear out’ the samples in the time domain

\[ \{t_1, \ldots, t_m\}, \quad \{\omega_1, \ldots, \omega_m\}, \quad \omega_i = 2\pi i/m. \]

\[ y_i = \langle b_{\omega_i, t_i}, x \rangle, \quad i = 1, \ldots, m. \quad b_{\omega, t}(t) \equiv \exp \left\{ i \omega t - \frac{(t-t_*)^2}{2\ell^2} \right\}. \]
Our scheme (Cont’d)

Fourier domain:
... integrating over freq. within a window of size $\ell^{-1}$ around $\omega_*$.

$\Rightarrow A_F$ is roughly band-diagonal!

- Reconstruction: Bayesian AMP
Our scheme (Cont’d)

Fourier domain:
... integrating over freq. within a window of size $l^{-1}$ around $\omega_*$.

$\implies A_F$ is roughly band-diagonal!

Reconstruction: Bayesian AMP
Fourier domain:
... integrating over freq. within a window of size $\ell^{-1}$ around $\omega_*$. 

$\implies A_F$ is roughly band-diagonal!

- Reconstruction: Bayesian AMP
Why should it work?

<table>
<thead>
<tr>
<th>Classical scheme</th>
<th>Our scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier coefficients (Delocalized measurements)</td>
<td>Gabor coefficients (Band-diagonal sensing matrix)</td>
</tr>
</tbody>
</table>

This is an implementation of a broader idea → Spatial Coupling!

[Kudekar, Pfister, 2010]
[Krzakala, Mézard, Sausset, Sun, Zdeborova, 2011]
[cf. also Felstrom, Zigangirov, 1999; Kudekar, Richardson, Urbanke 2009-2011]
Why should it work?

<table>
<thead>
<tr>
<th>Classical scheme</th>
<th>Our scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier coefficients</td>
<td>Gabor coefficients</td>
</tr>
<tr>
<td>(Delocalized measurements)</td>
<td>(Band-diagonal sensing matrix)</td>
</tr>
</tbody>
</table>

This is an implementation of a broader idea → **Spatial Coupling**!

[Kudekar, Pfister, 2010]
[Krzakala, Mézard, Sausset, Sun, Zdeborova, 2011]
[cf. also Felstrom, Zigangirov, 1999; Kudekar, Richardson, Urbanke 2009-2011]
An overview on spatial coupling
**Spatially coupled sensing matrix**

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

- \(\sim\) 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

[Donoho, Maleki, Montanari 2009]
[Donoho, Javanmard, Montanari 2011]

\[ x^{t+1} = \eta_t(x^t + (Q_t \circ A_F)^* r^t) , \]
\[ r^t = y - A_F x^t + b_t \circ r^{t-1} + d_t \circ \bar{r}^{t-1} . \]

\( Q_t, b_t, d_t \) 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

[Donoho, Maleki, Montanari 2009]
[Donoho, Javanmard, Montanari 2011]

\[
x^{t+1} = \eta_t(x^t + (Q_t \odot A_F)r^t), \\
r^t = y - A_Fx^t + b_t \odot r^{t-1} + d_t \odot \tilde{r}^{t-1}.
\]

\[Q_t, b_t, d_t\] 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.
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.
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.
Rényi information dimension

Characterization of $\bar{d}(X)$ (Rényi)

Let $p_X$ be a probability measure over $\mathbb{R}$, and $X \sim p_X$.
Let

$$p_X = (1 - \varepsilon)\nu_d + \varepsilon\tilde{\nu}$$

with

$\nu_d$: a discrete distribution (i.e. with countable support)

$\tilde{\nu}_d$: an absolutely continuous

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

In particular, if $\mathbb{P}\{X \neq 0\} \leq \varepsilon$ then $\bar{d}(X) \leq \varepsilon$.

[cf. Wu, Verdú]
Rényi information dimension

Characterization of $\overline{d}(X)$ (Rényi)

Let $p_X$ be a probability measure over $\mathbb{R}$, and $X \sim p_X$. Let

$$p_X = (1 - \varepsilon)\nu_d + \varepsilon\tilde{\nu}$$

with

$\nu_d$: a discrete distribution (i.e. with countable support)

$\tilde{\nu}_d$: an absolutely continuous

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

In particular, if $\mathbb{P}\{X \neq 0\} \leq \varepsilon$ then $\overline{d}(X) \leq \varepsilon$.

[cf. Wu, Verdú]
Does the spatial coupling phenomenon survive for physically constrained sensing matrices?
Experiments
$x(1), \ldots, x(n) \sim \text{i.i.d.} (1 - \varepsilon)\delta_0 + \varepsilon \text{Normal}(0, 1)$

Will it work for $m \geq n\varepsilon + o(n)$?
\( \epsilon = 0.1, \ m = 0.15 \ n \)

\[
\text{MSE}^{(t)}(i) = \mathbb{E}\{|\hat{x}^t_i - \bar{x}_i|^2\}.
\]

\( \ell_1 \) minimization requires \( m \geq 0.33 \ n \)!
\[ \varepsilon = 0.1, \ m = 0.15 \ n \]

\[
\text{MSE}^{(t)}(i) = \mathbb{E}\{ |\hat{x}_i^t - \tilde{x}_i |^2 \}.
\]

- $\ell_1$ minimization requires $m \gtrsim 0.33 \ n$!
Phase transition

- Scheme I: Bayesian AMP, Random Gabor.
- Scheme II: Bayesian AMP, Random Fourier.
- Scheme III: $\ell_1$, Random Gabor.
Conclusion

- “Spatially-coupled measurements + Bayesian AMP” achieves the information theoretically optimal rate.

- The power of this scheme also applies to the physically constrained sensing matrices.

Thanks!
Conclusion

▶ “Spatially-coupled measurements + Bayesian AMP” achieves the information theoretically optimal rate.

▶ The power of this scheme also applies to the physically constrained sensing matrices.

Thanks!
Conclusion

- “Spatially-coupled measurements + Bayesian AMP” achieves the information theoretically optimal rate.

- The power of this scheme also applies to the physically constrained sensing matrices.

Thanks!
Conclusion

- "Spatially-coupled measurements + Bayesian AMP" achieves the information theoretically optimal rate.

- The power of this scheme also applies to the physically constrained sensing matrices.

Thanks!