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:

$$\widehat{x} = \mathsf{F}x, \qquad \mathsf{F}\text{: Fourier matrix}$$

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

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:

$$\widehat{x} = \mathsf{F}x, \qquad \mathsf{F}:\ \text{Fourier matrix}$$

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

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

Sampling mechanism

$$y_i = \langle a_i, x \rangle, \quad i = 1, \cdots, m.$$

We refer to m/n as the sampling rate.

(In time domain) $\qquad y = Ax.$

(In frequency domain) $\qquad y = A\mathsf{F}^* \widehat{x} = A_{\mathsf{F}} \widehat{x}.$

Sampling mechanism

$$y_i = \langle a_i, x \rangle, \quad i = 1, \cdots, m.$$

We refer to m/n as the sampling rate.

(In time domain)	$y = Ax.$
(In frequency domain)	$y = A\mathsf{F}^* \widehat{x} = A_{\mathsf{F}} \widehat{x}.$

Normalization

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

$$\rightarrow A = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \qquad \|a_i\|_2 = 1$$

Sampling schemes:

- Instantaneous sampling at equispaced times → rate = Nyquist rate
  [Shannon 1948]

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

Our scheme:

- Non-instantaneous sampling at random times → $m = k + o(n)$

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 = Nyquist rate

  [Shannon 1948]

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

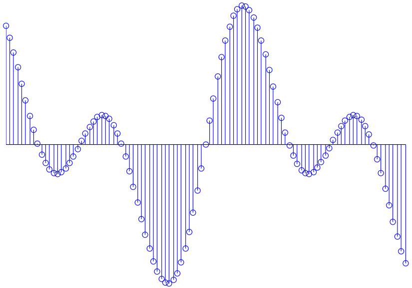
  [Candés, Romberg, Tao 2006, Candés, Plan 2011]

Our scheme:

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

Classical compressive sensing scheme

- Measurements: sample pointwise at random times



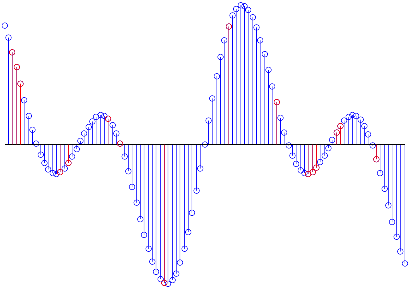
Fourier domain: random rows of DFT matrix.

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

▷ Reconstruction: Convex minimization (ℓ_1 minimization)

Classical compressive sensing scheme

- Measurements: sample pointwise at random times



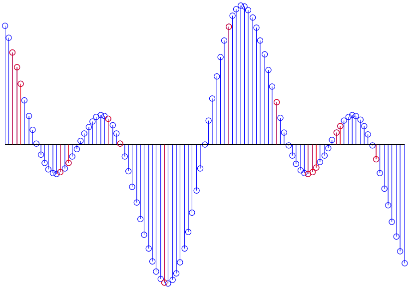
Fourier domain: random rows of DFT matrix.

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

▷ Reconstruction: Convex minimization (ℓ_1 minimization)

Classical compressive sensing scheme

- Measurements: sample pointwise at random times



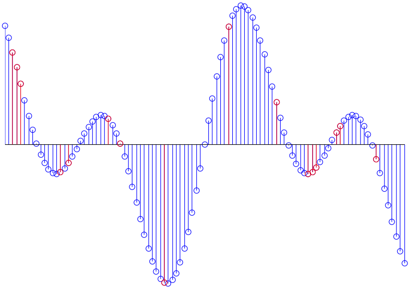
Fourier domain: random rows of DFT matrix.

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

▷ Reconstruction: Convex minimization (ℓ_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 (ℓ_1 minimization)

Our scheme

A different solution!

We 'smear out' the samples in the time domain

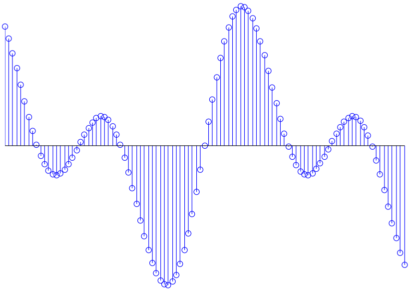
$$\{t_1, \cdots, t_m\}, \quad \{\omega_1, \cdots, \omega_m\}, \quad \omega_i = 2\pi i/m.$$

Our scheme

A different solution!

We 'smear out' the samples in the time domain

$$\{t_1, \cdots, t_m\}, \quad \{\omega_1, \cdots, \omega_m\}, \quad \omega_i = 2\pi i/m.$$

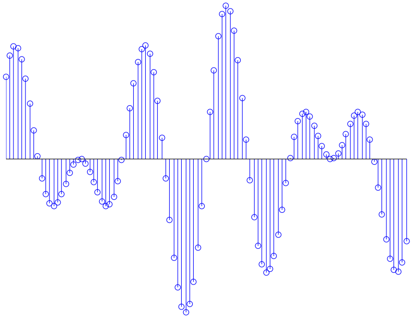


Our scheme

A different solution!

We ‘smear out’ the samples in the time domain

$$\{t_1, \cdots, t_m\}, \quad \{\omega_1, \cdots, \omega_m\}, \quad \omega_i = 2\pi i/m.$$



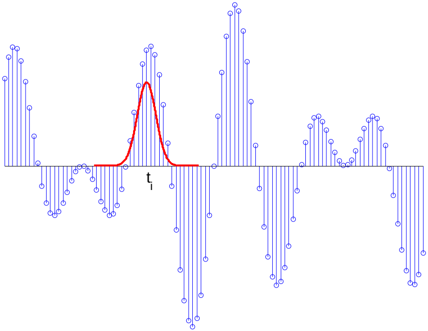
modulate with ω_i

Our scheme

A different solution!

We 'smear out' the samples in the time domain

$$\{t_1, \cdots, t_m\}, \quad \{\omega_1, \cdots, \omega_m\}, \quad \omega_i = 2\pi i/m.$$



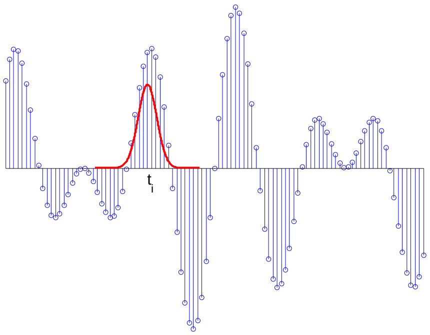
integrate over a window (of size ℓ) around t_i

Our scheme

A different solution!

We 'smear out' the samples in the time domain

$$\{t_1, \cdots, t_m\}, \quad \{\omega_1, \cdots, \omega_m\}, \quad \omega_i = 2\pi i/m.$$



$$y_i = \langle b_{\omega_i, t_i}, x \rangle, \quad i = 1, \cdots, m. \qquad b_{\omega_*, t_*}(t) \equiv \exp\left\{ i\omega_* t - \frac{(t-t_*)^2}{2\ell^2} \right\}.$$

Our scheme (Cont'd)

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

$\Longrightarrow$ 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 ω_*.

$\Longrightarrow$ 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 ω_*.

$\Longrightarrow A_{\mathsf{F}}$ is roughly band-diagonal !

- Reconstruction: Bayesian AMP

Why should it work?

Classical scheme	Our scheme
Fourier coefficients (Delocalized measurements)	Gabor coefficients (Band-diagonal sensing matrix)

This is an implementation of a broader idea $\rightarrow$ 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?

Classical scheme	Our scheme
Fourier coefficients (Delocalized measurements)	Gabor coefficients (Band-diagonal sensing matrix)

This is an implementation of a broader idea $\rightarrow$ Spatial Coupling!

[Kudekar, Pfister, 2010]

[Krzakala, Mézard, Sausset, Sun, Zdeborova, 2011]

[cf. also Felstrom, Zigangirov, 1999; Kudekar, Richardson, Urbanke 2009-2011]

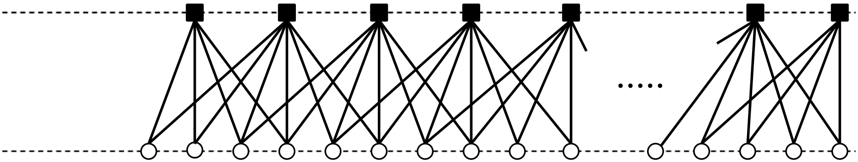
An overview on spatial coupling

Spatially coupled sensing matrix

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

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

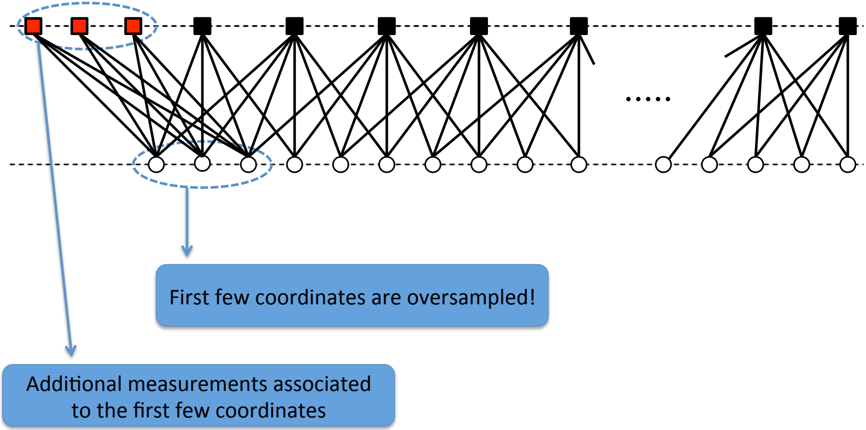
How does spatial coupling work?



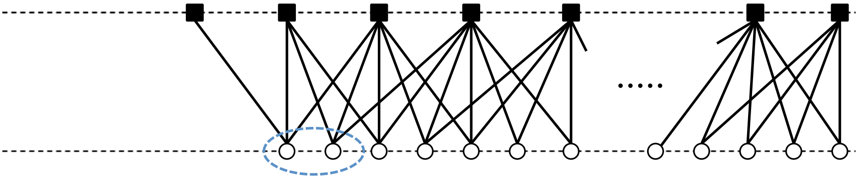
○ Coordinates of x

■ Coordinates of y

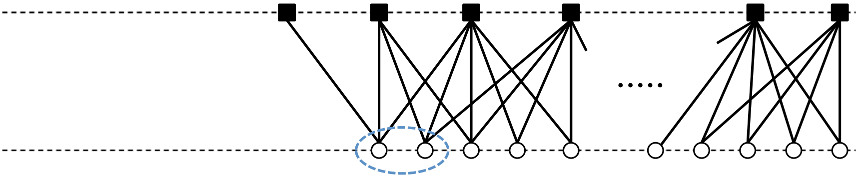
How does spatial coupling work?



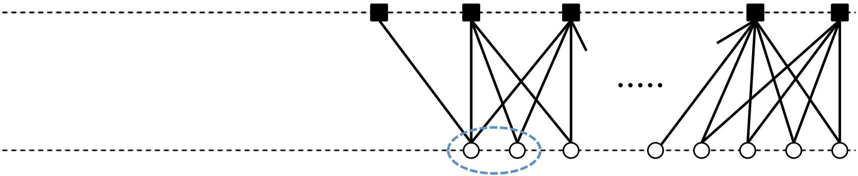
How does spatial coupling work?



How does spatial coupling work?



How does spatial coupling work?



Bayes-optimal AMP

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

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

$Q_t, \mathsf{b}_t, \mathsf{d}_t$ 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

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

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

$Q_t, \mathsf{b}_t, \mathsf{d}_t$ 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

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

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.

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 $\overline{d}(X)$ (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
ν_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]

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
ν_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]

Question

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

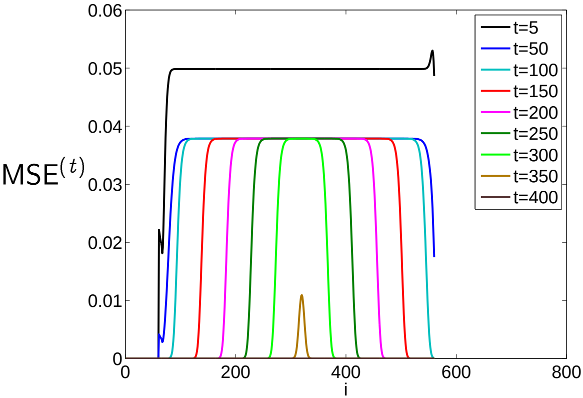
Experiments

Experiment

- $x(1), \ldots, x(n) \sim_{\text{i.i.d.}} (1-\varepsilon)\delta_0 + \varepsilon\,\text{Normal}(0,1)$

- Will it work for $m \geq n\varepsilon + o(n)$?

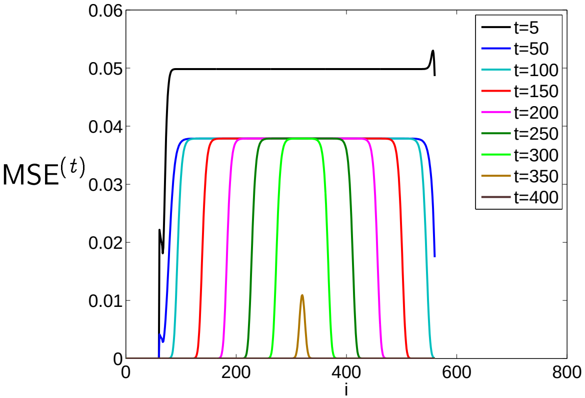
$\varepsilon = 0.1,\ m = 0.15\, n$



$$\mathsf{MSE}^{(t)} \in \mathbb{R}^n, \quad \mathsf{MSE}^{(t)}(i) = \mathbb{E}\{|\widehat{x}_i^t - \widehat{x}_i|^2\}.$$

- ℓ_1 minimization requires $m \gtrsim 0.33\, n$!

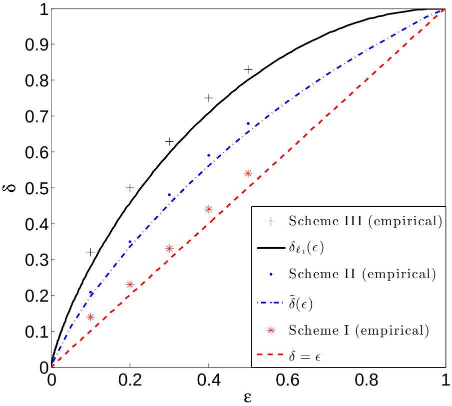
$\varepsilon = 0.1,\ m = 0.15\,n$



$$\mathrm{MSE}^{(t)} \in \mathbb{R}^n, \quad \mathrm{MSE}^{(t)}(i) = \mathbb{E}\{|\widehat{x}_i^t - \widehat{x}_i|^2\}.$$

- ℓ_1 minimization requires $m \gtrsim 0.33\,n$!

Phase transition



- Scheme I : Bayesian AMP, Random Gabor.
- Scheme II: Bayesian AMP, Random Fourier.
- Scheme III: ℓ_1, Random Gabor.

Conclusion

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

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

Thanks!