Compressive Sampling of Ensembles of Correlated Signals

Justin Romberg
Collaborator: Ali Ahmed
Georgia Tech, School of ECE

Duke Workshop on Sensing and Analysis of High-Dimensional Data
July 26, 2011
Durham, North Carolina
Goal: acquire an \textit{ensemble} of $M$ signals

Bandlimited to $W/2$

“Correlated” $\rightarrow$ $M$ signals are $\approx$ linear combinations of $R$ signals
Goal: acquire an ensemble of $M$ signals

Bandlimited to $W/2$

“Correlated” $\rightarrow M$ signals are $\approx$ linear combinations of $R$ signals
Sensor arrays
Framework

- “Wired” local arrays that may or may not share a (multiplexed) ADC
- Sparsity has nothing to do with it (but makes a guest appearance...)
- Correlation structure is unknown (low-rank recovery problem)
- Interested in systems with clear “implementation potential”
Components

- Analog vector-matrix multiplier spreads energy across channels
- Modulators spread energy across frequency
- Filters spread energy in one channel across time
- We will use both uniform and non-uniform ADCs
Known correlation structure → whiten then sample

\[
\begin{bmatrix}
-0.82 & -1.31 \\
1.09 & 0.27 \\
1.05 & 1.81 \\
-0.74 & -0.31 \\
-0.97 & 0.94 \\
1.19 & 2.19 \\
\end{bmatrix}
\]

Suppose the “mixing matrix” \( A \) is known and has SVD

\[
A = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} \Sigma \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix}
\]

then an efficient sampling structure is to “whiten” with \( U^T \), then sample
Known correlation structure $\rightarrow$ whiten then sample

- Requires $R$ ADCs and a total of $RW$ samples
- Recover samples of original using $X = UY$
Sampling correlated signals

\[
M = \begin{bmatrix}
-0.82 & -1.31 \\
1.09 & 0.27 \\
1.05 & 1.81 \\
-0.74 & -0.31 \\
-0.97 & 0.94 \\
1.19 & 2.19 \\
\end{bmatrix}
\]
Sampling correlated signals

Bandlimited $\Rightarrow$ this is just a low-rank recovery problem

Sampling each channel separately takes $MW$ total samples, we want strategies that take $\sim RW$ total samples
Given \( p \) linear samples of a matrix,

\[
y = A(X_0), \quad y \in \mathbb{R}^p, \quad X_0 \in \mathbb{R}^{M \times W}
\]

we solve

\[
\min_X \|X\|_* \quad \text{subject to} \quad A(X) = y
\]

where \( \|X\|_* \) is the nuclear norm: the sum of the singular values of \( X \).

An “optimal” sampler \( A \) would (stably) recover \( X_0 \) from \( y \) when

\[
\#\text{samples} \gtrsim R \cdot \max(M, W)
\]

\[
\gtrsim RW \quad \text{(in our case)}
\]
Architecture 1: One non-uniform ADC per channel

- \( M \) individual nonuniform-ADCs with average rate \( \theta \)
- Same as choosing \( M\theta \) random samples from \( M \times W \) matrix
Matrix completion

- Results of Candes, Recht, Tao, Keshavan, Montenari, Oh, Plan, ...
  Given a small number of entries in a low-rank matrix, we can “fill in” the missing entries

\[ R \times W \text{ matrix } X = U \Sigma V^T \text{ is rank } R \]
\[ \mu = \max(\text{max}_i \| U^T e_i \|_2, \text{max}_i \| V^T e_i \|_2, \| UV^T \|_2) \]
\[ \text{then we can recover } X \text{ whp from randomly chosen samples when } \#	ext{samples} \geq \text{Const} \cdot \mu \cdot RW \log_2(W) \]
Matrix completion

- Results of Candes, Recht, Tao, Keshavan, Montenari, Oh, Plan, ...

  Given a small number of entries in a low-rank matrix, we can “fill in” the missing entries

- Recht ’09: Suppose $M \times W$ matrix $X = U\Sigma V^T$ is rank $R$ with

\[
\mu = \max \left( \frac{M}{R} \max_i \|U^T e_i\|_2^2, \frac{W}{R} \max_i \|V^T e_i\|_2^2, \frac{MW}{R} \|UV^T\|_\infty \right)
\]

then we can recover $X$ whp from randomly chosen samples when

\[
\#\text{samples} \geq \text{Const} \cdot \mu \cdot RW \log^2(W)
\]

using nuclear norm minimization
Architecture 1: One non-uniform ADC per channel

- Direct application of these results: we can recover "incoherent" ensembles when

\[ \text{total samples} = M\theta \geq \text{Const.} \cdot RW \cdot \log^2(W) \]

so we can take \( \theta \sim \frac{R}{M}W \) instead of \( W \).

- Incoherent \( \Rightarrow \)

  \textit{signal energy is spread out evenly across time and channels}
Architecture 1: One non-uniform ADC per channel

Drawbacks:
- Incoherence assumptions (not universal)
- Requires $M$ ADCs (time-multiplexing would be delicate...)
Spreading the signals out

\[ X = U\Sigma V^T \quad \Rightarrow \quad \tilde{X} = \tilde{U}\Sigma\tilde{V}^T \]

- Take \( A \ M \times M \) and orthogonal, 
  \( H = \text{circ}(h[n]) \) orthogonal:
  \[
  H = F^H \Lambda F, \quad \Lambda = \text{diag}\{\lambda_i\}, \quad |\lambda_i| = 1
  \]
  then

\[ X = U\Sigma V^T \quad \Rightarrow \quad \tilde{X} = \tilde{U}\Sigma\tilde{V}^H, \quad \tilde{U} = AU, \quad \tilde{V} = HV \]
We can recover the ensemble $\tilde{X}$ when
\[ \text{total samples } \gtrsim RW \log^4(W) \]

From $\tilde{X}$, we recover $X$ using
\[ X = A^T \tilde{X} H \]

Universal, but still using an ADC for every channel...
Multiplexing onto one channel

- We can always combine $M$ channels into 1 by multiplexing in either time or frequency.

Frequency multiplexer:

- Replace $M$ ADCs running at rate $W$ with 1 ADC at rate $MW$. 

\[ \text{modulator} \cos(Wt) \quad \text{modulator} \cos(2Wt) \quad \text{ADC rate } 3W \]
If the signals are spread out uniformly in time, then the ADC and modulators can run at rate

$$\varphi \gtrsim RW \log^{3/2}(MW)$$

This requires a (milder) “incoherence across time” assumption
Prefiltering spreads signals out over time (low-bandwidth filters)

Modulate and sum diversifies and then combines the channels

We use one standard ADC operating at rate $\varphi$ (modulation rate is the same as the ADC sample rate)

How big does $\varphi$ need to be to recover $X$?
Matrix formulation

\[ y = \begin{bmatrix} H_1 & H_2 & H_3 & \cdots & H_M \end{bmatrix} P \text{vec}(X) \]

\[ (\varphi \times \varphi M \text{ random matrix})(\text{samples of } X \text{ at rate } \varphi) \]

We have structured random linear measurements of a rank \( R \) matrix...
Compressive multiplexing theory

- Recht et al. '07:

  *Recovery is possible given the matrix-RIP:*

  \[(1 - \delta)\|X\|_F^2 \leq \|A(X)\|_2^2 \leq (1 + \delta)\|X\|_F^2, \quad \forall X : \text{rank}(X) \leq 2R,\]
Compressive multiplexing theory

Recht et al ’07 Candes and Plan ’09

The mRIP can follow from a certain concentration bound.

If for any fixed \( M \times W \) matrix \( X \) and some \( 0 < t < 1 \) we have

\[
P \left\{ \left| \| \mathbf{A}(\mathbf{X})\|_2^2 - \|\mathbf{X}\|_F^2 \right| > t\|\mathbf{X}\|_F^2 \right\} \leq C e^{-\frac{p}{\mu}},
\]

then \( \delta < .307 \) for

\[
p \gtrsim \mu RW
\]
Compressive multiplexing theory

- Krahmer and Ward ’10:
  *Modulating columns of a sparse-RIP matrix yields concentration.*

Suppose \( \Phi \) satisfies

\[
(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2 \quad \forall \text{ } K\text{-sparse } x.
\]

Set \( \Phi' = \Phi P \). Then there is a \( t < 1 \) s.t. for any fixed \( x \)

\[
P \left\{ \left| \left\| \Phi' x \right\|_2^2 - \|x\|_2^2 \right| > t\|x\|_2^2 \right\} \leq Ce^{-K/c}
\]
R '09:

*Concatenated random Toeplitz matrices obey a sparse-RIP.*

Take

\[ \Phi = \begin{bmatrix} H_1 & H_2 & \cdots & H_M \end{bmatrix} \]

then with high probability

\[
(1 - \delta) \| x \|_2^2 \leq \| \Phi x \|_2^2 \leq (1 + \delta) \| x \|_2^2 \quad \forall \text{ } K\text{-sparse } x
\]

when

\[ \varphi \gtrsim K \cdot \log^4(\varphi M) \]
We can stably recover a rank-$R$ ensemble $X$ when the modulators and ADC operate at rate

$$\varphi \gtrsim RW \log^4(MW)$$

This architecture is universal in that it works for any low-rank correlation structure.
Numerical experiments

A few data points form a rough idea of how this works in practice:

<table>
<thead>
<tr>
<th></th>
<th>$M \times W$</th>
<th>$R$</th>
<th>sample threshold</th>
<th>factor above $RW$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>300 $\times$ 1000</td>
<td>5</td>
<td>18400</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>300 $\times$ 1000</td>
<td>7</td>
<td>20800</td>
<td>2.97</td>
</tr>
<tr>
<td>A3</td>
<td>400 $\times$ 1000</td>
<td>2</td>
<td>7600</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>400 $\times$ 1000</td>
<td>4</td>
<td>10680</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td>400 $\times$ 1000</td>
<td>7</td>
<td>14000</td>
<td>2</td>
</tr>
<tr>
<td>A4</td>
<td>100 $\times$ 1000</td>
<td>4</td>
<td>6600</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>100 $\times$ 1000</td>
<td>7</td>
<td>10000</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>300 $\times$ 1000</td>
<td>5</td>
<td>13000</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>300 $\times$ 1000</td>
<td>7</td>
<td>14500</td>
<td>2.07</td>
</tr>
</tbody>
</table>

A2 is non-uniform sampling (matrix samples)
A3 is modulated multiplexing
A4 is prefilter+modulated multiplexing
We saw several *compressive sampling architectures* for acquiring ensembles of *correlated signals* where the total number of samples we take scales like

$$(\text{bandwidth}) \times (\text{rank})$$

(to within log factors)