Compressed Sensing of Continuous Index Signals and of Low Rank Matrices

Yoram Bresler
Coordinated Science Laboratory and Department of ECE
University of Illinois at Urbana-Champaign

Acknowledgement:
NSF Grant CCF 06-35234
Outline

• Part I: compressive sensing of continuous-index signals (work with Ping Feng and Raman Venkataramani)

• Part II: Guaranteed low-rank matrix approximation from linear observations (work with Kiryung Lee)
Applications:

Real-world Signals are continuous indexed:

- Communications
- ELINT
- Fourier imaging: MRI, CT, SAR
  - Sparse image $X(f)$ $f \in \mathbb{R}^d$
  - Sampling in $t \in \mathbb{R}^d$
 Compression On-The-Fly

Conventional

F. T. Imaging Physics

Compression On-The-Fly

1998 - IT Workshop
Spectrum-Sparse Continuous-IndexSignals

- **1-D continuous-time signals**
  \[ x(t) \leftrightarrow X(f) \quad t, f \in \mathbb{R} \]

- **1-D discrete-time signals** (infinite sequences)
  \[ x(n) \leftrightarrow X(f) \quad n \in \mathbb{Z}, \quad f \in \mathbb{R} \]

Spectrum-Sparse:

\[ X(f) = 0, \quad f \notin \mathcal{F}, \quad \mathcal{F} = \bigcup_{i=1}^{n} [a_i, b_i] \subseteq [f_{\min}, f_{\max}] \]

\[ \frac{\lambda(\mathcal{F})}{f_{\max} - f_{\min}} \leq \Omega < 1 \]
Questions

• Sampling rate requirements ?
  – Known \( F \)
  – Unknown \( F \)

• Sampling at the minimum rate ?
  – Design of sampling scheme
  – Reconstruction

• How achieve:
  – Universal (non-adaptive sampling)
  – Perfect (or robust) reconstruction
  – Computation linear in the data size

• Relationship to compressive sensing?
Sampling Rate

Throw in the towel (sufficient condition)

- Nyquist sampling:  \( f_{Nyq} = f_{max} - f_{min} \)

Necessary conditions

- Landau lower bound (1967)  
  - Arbitrary pointwise sampling
  - Sufficient for known \( \mathcal{F} \) packable or not (Kahn & Liu, 1965)
    \[ D^- \geq \lambda(\mathcal{F}) \]

Sufficient conditions, unknown \( \mathcal{F} \) (Feng & Bresler, 1996):

\[ D^- \approx \begin{cases} 2\Omega f_{Nyq} & \text{for all signals} \\ \Omega f_{Nyq} & \text{for almost all signals} \end{cases} \]
Periodic Sampling

- Periodic sampling pattern: $C = \{c_i : 1 \leq i \leq p\}$
- Period = $L$
Spectral support at resolution $L$

Active spectral cells: $\mathcal{F}_i \cap \mathcal{F} \neq \emptyset, i = 1, 2, \ldots q$

$\Omega_L \triangleq \frac{q}{L} \approx \Omega$

$\mathcal{F}_0 = [0, \frac{1}{LT}]$, \hspace{1cm} $\mathcal{F}_l \triangleq \mathcal{F}_0 + \frac{l}{L}, \hspace{1cm} l = 1, 2, \ldots, L - 1,$

Vectorized signal spectrum

$X_l(f) \triangleq X(f + \frac{l}{LT}) \chi(f; \mathcal{F}_0)$

$\zeta(f) \triangleq [X_1(f), \ldots, X_L(f)]', \hspace{1cm} f \in \mathcal{F}_0$

Vectorized sample spectrum

$x_i(m) = x(Lm + c_i), \hspace{1cm} m \in \mathbb{Z} \leftrightarrow X_i(f)$

$y(f) \triangleq LT[e^{-j2\pi c_1 f^T} X_1(LT f), \ldots, e^{-j2\pi c_1 f^T} X_p(LT f)]', \hspace{1cm} p = \frac{q}{L}$

$y(f) = A \zeta(f), \hspace{1cm} f \in \mathcal{F}_0, \hspace{1cm} A \in \mathbb{C}^{p \times L} =$ Submatrix of DFT $L \times L$
Blind Reconstruction & Compressive Sensing

\[ y(f) = A\zeta(f), \quad f \in \mathcal{F}_0, \quad A \in \mathbb{C}^{p \times L} = \text{Submatrix of DFT} \]

\[ \|\zeta(f)\|_0 \leq \Omega_{LL}, \quad f \in \mathcal{F}_0 \]

**P0:** for each \( f \in \mathcal{F}_0 \)

\[
\min_{\zeta(f)} \|\zeta(f)\|_0 \\
\text{subject to } y(f) = A\zeta(f)
\]

- Common sparsity pattern for all \( f \in \mathcal{F}_0 \)
- *But, uncountably infinite number of simultaneous measurement vectors*
Spectral Support Recovery

- Active spectral cell indices:
  \[ \mathbf{k} = [k_1, k_2, \ldots, k_q] : \mathcal{F}_{k_r} \cap \mathcal{F} \neq \emptyset, \]
  \[ k_r \in \{1, 2, \ldots, L\} \quad r = 1, \ldots, q, \]

\[ P_1: \hat{\mathbf{k}} = \arg \min_{\| \mathbf{k} \|_0 = q} \int_{f \in \mathcal{F}_0} \left\| P_{\text{Range}^\perp(A_k)} \mathbf{y}(f) \right\|_2^2 df \]

\[ P_2: \hat{\mathbf{k}} = \arg \min_{\| \mathbf{k} \|_0 = q} \text{tr}(P_{\text{Range}^\perp(A_k)} \mathbf{R}) \]

\[ \mathbf{R} \triangleq \int_{f \in \mathcal{F}_0} \mathbf{y}(f) \mathbf{y}^*(f) df \]

\[ (\mathbf{R})_{kl} = \langle x_k(n - \frac{c_k}{L}), x_l(n - \frac{c_l}{L}) \rangle. \]

- Finite dimensional problem!
Finite-dimensional optimization

\[ P2: \hat{k} = \arg \min_{q, \|k\|_0 = q} \text{tr} \left[ P_{\text{Range}^\perp(A_k)} R \right] \]

• Solution of P2 using greedy algorithm (Feng & Bresler, 1996; Venkataramani & Bresler, 1998)
More equivalent problems

**P3:** Given $R \in \mathbb{C}^{p \times p}$ find the smallest integer $q$ and spectral cell index vector $k$ of length $q$ such that

$$R = A_k Z A_k^*$$

for some $Z \in \mathbb{C}^{q \times q}$, $Z \succeq 0$

- Let $R = U_s \Lambda_s U_s^*$, $U \in \mathbb{C}^{p \times r}$, $r = \text{rank}(R) = \text{rank}(Z)$

**P4:** Given $U$ find the smallest integer $q$ and spectral cell index vector $k$ of length $q$ such that

$$U = A_k Q$$

for some $Q \in \mathbb{C}^{q \times q}$

- **Proposition:** Problems P0 -- P4 are equivalent

- P4 is the (now) classical MMV problem!
- Mishali+Eldar (2007) proposed to solve it using MMV compressed sensing methods.
Sampling Conditions

**Theorem** (Feng & Bresler, 1996). Let $\mathcal{X}$ consist of signals with spectral support of spectral occupancy at most $\Omega_L$ at resolution $L$. Then

1. all signals in $\mathcal{X}$ can be uniquely reconstructed from their samples on a periodic pattern with average rate $\approx 2\Omega_L f_{\text{max}}$.

2. almost all signals in $\mathcal{X}$ can be reconstructed from samples at rate $\approx \Omega_L f_{\text{max}}$.

Proof: (1) For bunched pattern with $p \geq 2q + 1$

$$\text{Spark}(A) \geq 2q + 1$$

(b) Almost all $Z$ have full rank.
Numerical Experiments
2D Spectrum-Blind Sampling*

(a) (b)

*ICIP 1998
Conclusions

• Spectrum-Blind sampling generalizes Compressed Sensing, and reduces the infinite dimensional problem to a finite dimensional problem without discretization.

• Compressed sensing with DFT measurement and time-sparse signals in Euclidean spaces is a special case of spectrum-blind sampling.

• Compressed Sensing theory provides new insights and provable, efficient techniques to solve the spectrum-blind sampling problem.
Guaranteed Low-Rank Matrix Approximation from Linear Observations
Work with Kiryung Lee
ECE/CSL UIUC

\[ A : \mathbb{C}^{m \times n} \rightarrow \mathbb{C}^p, \quad X \in \mathbb{C}^{m \times n}, \quad y \in \mathbb{C}^p \]

\[ y = AX + v, \]

\[ \min \text{rank}(X), \text{subject to } \|y - AX\|_2 < \varepsilon \]
Rank Minimization: Analogy – Previous Works

- $\ell_0$-Norm Minimization
- Rank Minimization

<table>
<thead>
<tr>
<th>Performance guarantee</th>
<th>Sparsity (or low-rank)</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak</td>
<td>Exact</td>
<td>Exact - noiseless</td>
</tr>
<tr>
<td>Strong</td>
<td>Approximate</td>
<td>Noisy</td>
</tr>
</tbody>
</table>

- Basis Pursuit
- CoSaMP / Subspace Pursuit
- Orthogonal Matching Pursuit
- Matching Pursuit

Weak performance guarantee
Strong performance guarantee
Weak performance guarantee

Computationally efficient
Rank Minimization: Analogy – New Contribution

- $\ell_0$-Norm Minimization
- Rank Minimization

- Basis Pursuit
- CoSaMP / Subspace Pursuit
- Orthogonal Matching Pursuit
- Matching Pursuit

Weak performance guarantee
Strong performance guarantee

Computationally efficient

Performance guarantee | Sparsity (or low-rank) | Measurements
---|---|---
Weak | Exact | Exact - noiseless
Strong | Approximate | Noisy
Vector Atomic Decomposition

\( x \in \mathbb{C}^n \)

\[ x = c_1 \psi_i + c_2 \psi_j + c_3 \psi_k \]

\( \forall: \) set of atoms  \( \forall: \) set of atomic spaces

\[ V_1 = \text{span}(\psi_1) \]

\[ V_2 = \text{span}(\psi_2) \]

\[ \vdots \]

\[ V_n = \text{span}(\psi_n) \]

finite

\[ \|x\|_0 \leq s \quad \text{iff} \quad x \text{ is spanned by } s \text{ atoms} \]

\[ x \in \bigcup_{i=1}^n \left( \sum_{k=1}^s V_{i_k} \right) = \bigcup_{i=1}^n \text{span} \left( \{\psi_{i_k}\}_{k=1}^s \right) \]
Matrix Atomic Decomposition

$X \in \mathbb{C}^{m \times n}$

$X = c_1 \psi_\alpha + c_2 \psi_\beta$

\[ \Theta: \text{set of atoms} \quad \mathbb{A}: \text{set of atomic spaces} \]

\[ V_\alpha = \text{span}(\psi_\alpha) \]
\[ V_\beta = \text{span}(\psi_\beta) \]

\[ \text{rank}(X) \leq r \quad \text{iff} \quad X \text{ is spanned by } r \text{ atoms} \]

\[ X \in \bigcup_{|\Gamma| \leq r} \left( \sum_{\alpha \in \Gamma} V_\alpha \right) = \bigcup_{\Psi \subset \Theta, |\Psi| \leq r} \text{span} (\Psi) \]
Vector Correlation Maximization

\[ A \in \mathbb{C}^{p \times n}, \ y \in \mathbb{C}^p \]

\( a_k : k\)-th column of \( A \)

\( e_k : k\)-th column of \( I_n (n \times n \text{ identity matrix}) \)

\( P_{e_k} : \text{projection onto span}(e_k) \)

\[
|\langle y, a_k \rangle| = |\langle y, Ae_k \rangle| = |\langle A^H y, e_k \rangle| = \left\| P_{e_k} A^H y \right\|_2
\]

Find a column \( a_k \) that maximizes the correlation

\[ \iff \text{Find an atom } e_k \text{ that maximizes } \left\| P_{e_k} A^H y \right\|_2 \]
Matrix Correlation Maximization

\[ \mathcal{A} : \mathbb{C}^{m \times n} \rightarrow \mathbb{C}^p, y \in \mathbb{C}^p \]

\( \mathcal{A}^* \): adjoint of \( \mathcal{A} \)

\( \psi \): an atom

\( P_\psi \): projection onto \( \text{span}(\psi) \)

\[ |\langle y, \mathcal{A} \psi \rangle| = |\langle \mathcal{A}^* y, \psi \rangle| = \left\| P_\psi \mathcal{A}^* y \right\|_F \]

Find an atom \( \psi \) that maximizes the correlation \( |\langle y, \mathcal{A} \psi \rangle| \)

\( \Leftrightarrow \) Find an atom \( \psi \) that maximizes \( \left\| P_\psi \mathcal{A}^* y \right\| \)

- Can be done by SVD of matrix \( \mathcal{A}^* y \)
Matrix Correlation Maximization

• Greedy algorithms (MP, OMP, CoSaMP, SP) are extended from the vector case to the matrix case

• In particular, we propose **ADMiRA** (Atomic Decomposition for Minimum Rank Approximation) that is **efficient** in computation and has a **performance guarantee**.
ADMiRA

• Analogous to CoSaMP/SP
• Iterative subspace selection
• Assumptions
  – Target rank is fixed to $r$
  – Linear operator satisfies RIP: $\delta_{7r} < 0.056$
  – Measurement is obtained by $b = AX + \nu$
ADMiRA

• Performance guarantee
  – After $6(r+1)$ iterations, ADMiRA provides a rank-$r$ approximation satisfying
    \[
    \|X - \hat{X}\|_F \leq 20 \left( \|X - X_r\|_F + \frac{1}{\sqrt{r}} \|X - X_r\|_* + \|\nu\|_2 \right)
    \]
    $X_r$ : best rank-$r$ approximation, $\|\|_*$ : nuclear norm

• Efficient computation
  – most computation is (truncated) SVD
  – Randomized algorithms are applicable
Rank Minimization: Analogy – New Contribution

\( \ell_0 \)-Norm Minimization

- Basis Pursuit
- CoSaMP / Subspace Pursuit
- Orthogonal Matching Pursuit
- Matching Pursuit

Computationally efficient

Strong performance guarantee

Weak performance guarantee

Rank Minimization

- Nuclear Norm Minimization*
- ADMiRA
- OMP variation
- MP variation

Strong performance guarantee

Computationally efficient

Performance guarantee | Sparsity (or low-rank) | Measurements
---|---|---
Weak | Exact | Exact - noiseless
Strong | Approximate | Noisy

* Note: after the presentation we learned about the paper by M. Fazel, E. Candès, B. Recht, & P. Parrilo, "Compressed Sensing and Robust Recovery of Low Rank Matrices," Asilomar Conference, Nov. 2008, which also presented a strong performance guarantee for nuclear norm minimization, although with different constants than ours.