Sparse Support Recovery via Covariance Estimation

Lekshmi Ramesh

(Joint work with Prof. Chandra R. Murthy)

Indian Institute of Science
Bangalore

July 13, 2018
Outline

- Setup
  - Multiple measurement vector setting
  - Support recovery problem

- Support recovery as covariance estimation
  - Covariance matching, Gaussian approximation
  - Maximum likelihood-based estimation
  - Solution using non negative quadratic programming
  - Simulation results

- Remarks on non negative sparse recovery

- Conclusion
Multiple measurement vector (MMV) model:
Observations $\{y_i\}_{i=1}^L$ are generated from the following linear model:

$$y_i = \Phi x_i + w_i, \quad i \in [L],$$

where $\Phi \in \mathbb{R}^{m \times N}$ ($m < N$), $x_i \in \mathbb{R}^N$ unknown, random and noise $w_i \sim \mathcal{N}(0, \sigma^2 I)$

$x_i$ are $k$-sparse with common support

$\text{supp}(x_i) = T$ for some $T \subset [N]$ with $|T| \leq k$, $\forall i \in [L]$

Goal: Recover the common support $T$ given $\{y_i\}_{i=1}^L$, $\Phi$

Applications in hyperspectral imaging, sensor networks
Problem setup

- Generative model for $x_i$
  
  Assumption: Non-zero entries uncorrelated

$$p(x_i; \gamma) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi \gamma_j}} \exp \left( -\frac{x_{ij}^2}{2\gamma_j} \right)$$

i.e., $x_i \overset{iid}{\sim} \mathcal{N}(0, \Gamma)$ where $\Gamma = \text{diag}(\gamma)$

- Note:
  - $\text{supp}(x_i) = \text{supp}(\gamma) = T$ (since $\gamma_j = 0 \Leftrightarrow x_{ij} = 0$ a.s.)
  - $y_i \sim \mathcal{N}(0, \Phi \Gamma \Phi^\top + \sigma^2 I)$

- Equivalent problem: Recover $\text{supp}(\gamma)$ given $\{y_i\}_{i=1}^{L}$, $\Phi$
\[ \mathbf{x}_i \overset{iid}{\sim} \mathcal{N}(0, \Gamma) \]

\[ \begin{align*}
\mathbf{x}_1 & \quad \mathbf{x}_2 \\
\vdots & \quad \vdots \\
\mathbf{x}_L & \quad 
\end{align*} \]

\[ \begin{align*}
\mathbf{y}_1 & \quad \mathbf{y}_2 \\
\vdots & \quad \vdots \\
\mathbf{y}_L & \quad 
\end{align*} \]

\[ \Sigma = \Phi \Gamma \Phi^\top + \sigma^2 \mathbf{I} \]
Support recovery as covariance estimation

- Use the sample covariance matrix $\hat{\Sigma} = \frac{1}{L} \sum_{i=1}^{L} y_i y_i^\top$ to estimate $\Gamma$

- Express $\hat{\Sigma}$ as

$$\hat{\Sigma} = \Sigma + E,$$

where $E$: Noise/Error matrix

For the noiseless case ($\sigma^2 = 0$)

$$\hat{\Sigma} = \Phi \Gamma \Phi^\top + E$$

$$\underbrace{\text{vectorize}}\downarrow$$

$$r = (\Phi \odot \Phi) \gamma + e$$

where $\odot$ denotes the Khatri-Rao product

- Use Gaussian approximation for $e$

- Find the maximum likelihood estimate of $\gamma$

$^1$details for noisy case can be found in the paper
Noise statistics

■ Mean

\[ \mathbb{E}(E) = \frac{1}{L} \sum_{i=1}^{L} \mathbb{E} y_i y_i^\top - \Sigma = 0 \]

■ Covariance

\[ \text{cov}(\text{vec}(E)) = \frac{1}{L} (\Phi \otimes \Phi) (\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}}) \text{cov}(\text{vec}(zz^\top)) (\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}}) (\Phi \otimes \Phi)^\top, \]

where \( z \sim \mathcal{N}(0, I_N) \)
Example: $N=3$

Let $\mathbf{z} = [z_1, z_2, z_3]^\top$ with $z_i \overset{iid}{\sim} \mathcal{N}(0, 1)$. Then,

$$\mathbf{z} \mathbf{z}^\top = \begin{bmatrix}
    z_1^2 & z_1z_2 & z_1z_3 \\
    z_1z_2 & z_2^2 & z_2z_3 \\
    z_1z_3 & z_2z_3 & z_3^2
\end{bmatrix} \xrightarrow{\text{vectorize}} \begin{bmatrix}
    z_1^2 \\
    z_1z_2 \\
    z_1z_3 \\
    z_1^2 \\
    z_2^2 \\
    z_2z_3 \\
    z_1z_3 \\
    z_2z_3 \\
    z_3^2
\end{bmatrix}$$
Example: \( N=3 \)

- The covariance matrix \( B \) of \( \text{vec}(zz^\top) \) will be of size \( 9 \times 9 \) with \( B_{i,j} \in \{0, 1, 2\}, \ 1 \leq i, j \leq 3 \).

- For e.g.,

\[
\begin{align*}
B_{1,1} &= \text{cov}(z_1^2, z_1^2) = \mathbb{E}z_1^4 - (\mathbb{E}z_1^2)^2 = 3 - 1 = 2 \\
B_{1,2} &= \text{cov}(z_1^2, z_1z_2) = \mathbb{E}z_1^3z_2 - \mathbb{E}z_1^2\mathbb{E}z_1z_2 = 0 \\
B_{2,4} &= \text{cov}(z_1z_2, z_1z_2) = \mathbb{E}z_1^2z_2^2 - \mathbb{E}z_1z_2\mathbb{E}z_1z_2 = 1
\end{align*}
\]
Example: $N=3$

$$B = \text{cov}(\text{vec}(zz^\top)) = \begin{bmatrix}
2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
\end{bmatrix}$$
We now have the following model

\[ r = A \gamma + e, \]  

where

\[ A = (\Phi \odot \Phi), \]
\[ E[e] = 0, \]
\[ \text{cov}(e) = W = \frac{1}{L} (\Phi \otimes \Phi)(\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}})B(\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}})(\Phi \otimes \Phi)^\top. \]

Remarks

- The noise term vanishes as \( L \to \infty \)
- The noise covariance depends on the parameter to be estimated
- \( r, \Phi \odot \Phi \) and \( e \) have redundant entries – restrict to the \( \frac{m(m+1)}{2} \) distinct entries
New model, Gaussian approximation

- Pre-multiply (1) by $P \in \mathbb{R}^{\frac{m(m+1)}{2} \times m^2}$, formed using a subset of the rows of $I_{m^2}$, that picks the relevant entries. Thus,

$$r_P = A_P \gamma + e_P,$$

where $r_P := Pr$, $A_P := PA$, and $e_P := Pe$.

- Further, we approximate the distribution of $n_P$ by $\mathcal{N}(0, W_P)$, where $W_P = PW_P^\top$

- Thus, $r_P \sim \mathcal{N}(A_P \gamma, W_P)$
ML estimation of $\gamma$

- Denote the ML estimate of $\gamma$ by $\gamma_{\text{ML}}$

$$\gamma_{\text{ML}} = \arg \max_{\gamma \geq 0} p(r_P; \gamma),$$  \hspace{1cm} (2)

where

$$p(r_P; \gamma) = \frac{1}{(2\pi)^{m(m+1)/4} |W_P|^{1/2}} \exp \left( \frac{-(r_P - A_P \gamma)^\top W_P^{-1} (r_P - A_P \gamma)}{2} \right).$$

- Simplifying (2), we get

$$\gamma_{\text{ML}} = \arg \min_{\gamma \geq 0} \log |W_P| + (r_P - A_P \gamma)^\top W_P^{-1} (r_P - A_P \gamma).$$ \hspace{1cm} (3)

- For a fixed $W_P$, (3) can be solved using Non Negative Quadratic Programming (NNQP)
Algorithm 1 MRNNQP

1: Input: Measurement matrix $\Phi$, vectorized sample covariance $r$, initial value $\Gamma^{(0)} = \text{diag}(\gamma^{(0)})$, $i = 1$

2: While (not converged) do

3: $W_P^{(i)} \leftarrow \frac{1}{L} P (\Phi \otimes \Phi) B (\Gamma^{(i-1)} \otimes \Gamma^{(i-1)}) (\Phi \otimes \Phi)^\top P^\top$

4: $b^{(i)} \leftarrow -A_P^\top W_P^{(i)-1} r_P$

5: $Q^{(i)} \leftarrow A_P^\top W_P^{(i)-1} A_P$

6: $\gamma^{(i)} \leftarrow \text{NNQP}(Q^{(i)}, b^{(i)})$

7: $\Gamma^{(i)} \leftarrow \text{diag}(\gamma^{(i)})$

8: $i \leftarrow i + 1$

9: end While

10: Output: support of $\gamma^{(i)}$
The MSBL algorithm

- \( X = [x_1, \cdots, x_L], \ Y = [y_1, \cdots, y_L] \)
- Posterior moments
  \[ R = \text{cov}(x_i | y_i; \gamma); \ M = [\mu_1, \cdots, \mu_L] \]

Algorithm 2 MSBL\(^2\)

1: Input: Measurement matrix \( \Phi \), observations \( Y \), initial value \( \Gamma^{(0)} = \text{diag}(\gamma^{(0)}), i = 1 \)
2: While (not converged) do
3: \[ R^{(i)} \leftarrow \Gamma^{i-1} - \Gamma^{(i-1)} \Phi^\top (\Sigma^{(i-1)})^{-1} \Phi \Gamma^{(i-1)} \]
4: \[ M^{(i)} \leftarrow \Gamma^{(i-1)} \Phi^\top (\Sigma^{(i-1)})^{-1} Y \]
5: \[ \gamma_j^{(i)} \leftarrow \frac{1}{L} \| \mu_j^{(i)} \|_2^2 + R_{jj}^{(i)} \]
6: \( i \leftarrow i + 1 \)
7: end While
8: Output: \( \hat{x}_j = \mu_j^{(i)} \)

Support recovery performance

$N = 40, m = 20, k = 25$; exact recovery over 200 trials

Figure 1: Support recovery performance of the NNQP-based approach.
Support recovery performance

\[ N = 70, m = 20, L = 50, 1000; \text{ exact recovery over 200 trials} \]

Figure 2: Support recovery performance of the NNQP-based approach.
Phase transition

Figure 3: Phase transition. $N = 20, L = 200$
Observations

- Exact support recovery possible for $k < m$ regime with “small” $L$
  For $k \geq m$, recovery possible with “large” $L$

- Dependence of computational complexity on parameters
  - $L$: in computing $\hat{\Sigma}$ (offline)
  - $m, N$: scales as $m^4N^2$

- Comparison with Co-LASSO, MSBL
  - Improvement in performance by accounting for error due to $\hat{\Sigma}$
  - Only a one time computation of $\hat{\Sigma}$ is required whereas MSBL uses the entire set of measurements $\{y_i\}_{i=1}^L$ in every iteration of EM
Remarks on non negative sparse recovery

■ Inner loop in the ML estimation problem

\[
\arg \min_{\gamma \geq 0} (r_P - A_P\gamma)^\top W_P^{-1}(r_P - A_P\gamma)
\]

Note: no sparsity-inducing regularizer

■ Implicit regularization property of NNQP has been noted before\(^3,4\)

■ For successful recovery, require conditions on sign pattern of vectors in null space of \(A\)

---


Concluding remarks

- Sparse support recovery can be done using maximum likelihood-based covariance estimation.
- Support recovery possible even when \( k > m \).
- No explicit sparsity promoting regularizer needed.
- Recovery guarantees depend on properties of null space of \( \Phi \odot \Phi \).
Thank you
Non-negative quadratic program

\[
\text{minimize } \gamma \geq 0 \quad (r_P - A_P \gamma)^\top W_P^{-1} (r_P - A_P \gamma)
\]

Solution (entry-wise update equation for \(\gamma\)):

\[
\gamma_j^{(i+1)} = \gamma_j^{(i)} \left( \frac{-b_j + \sqrt{b_j^2 + 4(Q + \gamma_j^{(i)})_j(Q - \gamma_j^{(i)})_j}}{2(Q + \gamma_j^{(i)})_j} \right),
\]

where \(b = -A_P^\top W_P^{-1} r_P\), \(Q = A_P^\top W_P^{-1} A_P\),

\[
Q_{ij}^+ = \begin{cases} Q_{ij}, & \text{if } Q_{ij} > 0, \\ 0, & \text{otherwise} \end{cases}, \quad Q_{ij}^- = \begin{cases} -Q_{ij}, & \text{if } Q_{ij} < 0, \\ 0, & \text{otherwise} \end{cases}.
\]

---

Noise statistics

- Covariance

\[
\text{cov}(E) = \text{cov} \left( \sum_{i=1}^{L} \left( \frac{y_i y_i^\top}{L} - \frac{\Sigma}{L} \right) \right) \\
= L \text{cov} \left( \frac{y_1 y_1^\top}{L} - \frac{\Sigma}{L} \right) \quad \text{(sum of } L \text{ indep. random matrices)} \\
= \frac{1}{L} \text{cov}(y_1 y_1^\top - \Sigma) \\
= \frac{1}{L} \text{cov}(yy^\top)
\]

- Represent \( y \) as

\[
y = Cz,
\]

where \( z \sim \mathcal{N}(0, I) \) and \( \Sigma = CC^\top \)
Noise statistics

\[ \text{cov}(E) = \frac{1}{L} \text{cov}(yy^\top) \]

- For \( \sigma^2 = 0 \), \( \Sigma = \Phi \Gamma \Phi^\top \); can take \( C = \Phi \Gamma^{\frac{1}{2}} \)

- Using properties of Kronecker products:

\[
\begin{align*}
\text{cov}(\text{vec}(E)) &= \frac{1}{L} \text{cov}(\text{vec}(C zz^\top C^\top)) \\
&= \frac{1}{L} \text{cov}((C \otimes C) \text{vec}(zz^\top)) \\
&= \frac{1}{L} (C \otimes C) \text{cov}(\text{vec}(zz^\top)) (C \otimes C)^\top \\
&= \frac{1}{L} (\Phi \otimes \Phi) (\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}}) \text{cov}(\text{vec}(zz^\top)) (\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}}) (\Phi \otimes \Phi)^\top
\end{align*}
\]

- Last step: use \( (A \otimes B)(C \otimes D) = AB \otimes CD \)