Super-resolution off the grid

Qingqing Huang
MIT

Sham M Kakade
University of Washington

Introduction

Problem: recover a superposition of point sources using bandlimited and noise corrupted measurements.

Input: Cutoff freq \( R \), number of measurements \( m \).

Output: Estimates \( \{ \tilde{w}_j, \tilde{\mu}(j) : j \in [k] \} \).

1. Take measurements at random \( s \):
   - Let \( S = \{ s(1), \ldots, s(m) \} \) be \( m \) i.i.d. samples from the Gaussian \( N(0, R^2 I_{d \times d}) \).
   - Set \( s(m+n) = e_0 \) for all \( n \in [d] \) and \( s(m+n+1) = 0 \).
   - Denote \( m' = m + d + 1 \).
   - Take another random samples \( v \) from the unit sphere, set \( v(1) = v \) and \( v(2) = 2v \).
   - Construct a tensor \( F \in \mathbb{C}^{m' \times m' \times 3} \):
     \[
     \tilde{F}_{n_1, n_2, n_3} = \tilde{f}(s) x_{s(n_1)+1, s(n_2)+1, s(n_3)+1}.
     \]

2. Tensor Decomposition: Set \( (\tilde{V}_S, \tilde{D}_w) = \text{TensorDecomp}(\tilde{F}) \).
   - For \( j = 1, \ldots, k \), set \( \tilde{V}_s[j] = \tilde{V}_S[j] / \| \tilde{V}_s[j] \|_2 \).

3. Read of estimates: For \( j = 1, \ldots, k \), set \( \tilde{\mu}(j) = \text{Real}(\| \tilde{V}_S[m+1:m+d,j] \|_2) / (\pi \| \tilde{V}_s[j] \|_2) \).

4. Set \( \tilde{W} = \arg \min_{W \in \mathbb{C}^*} \| \tilde{F} - \tilde{V}_S \otimes \tilde{V}_S \otimes \tilde{D}_w F \|_F \).

Basic ideas of our work

- Run Prony’s method (Matrix-Pencil / MUSIC / ESPRIT) on random samples.
- Skip the intermediate step of recovering all the \( \Omega(N^d) \) samples on the hypergrid, random samples \( \rightarrow \) estimate point sources.
- Sample at \( S \) such that \( F \) admits the tensor decomposition:
  \[
  V_S = \text{Vandermonde Matrix with complex nodes}.
  \]

Algorithm

Sample complexity determined by condition number of \( F \) and \( V_S \).

- Random samples with frequency range \( R = O(1/\Delta) \).
- Number of measurements does not depend on \( \Delta \).

<table>
<thead>
<tr>
<th>( d = 1 )</th>
<th>( d \geq 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>cutoff freq</td>
<td>( \frac{1}{(k \log(k))^2} )</td>
</tr>
<tr>
<td>measurements</td>
<td>( \frac{1}{(k \log(k))^2} )</td>
</tr>
<tr>
<td>runtime</td>
<td>( \frac{1}{(k \log(k))^2} )</td>
</tr>
</tbody>
</table>

Table 1: Comparison of runtime. We are implicitly using \( \mathcal{O}(\cdot) \) notation here.

Main results

Main Theorem (stability of our algorithm)

Theorem 1. For a fixed error probability \( \delta \), the algorithm achieves stable recovery with number of measurements and runtime both bounded by \( \Omega(k \log(k) + d)^2 \). The frequency range of the measurements are bounded by \( O(1/\Delta) \) (ignoring log factors).

Key Lemma (condition number of random Vandermonde)

Lemma 1. For fixed \( \epsilon_2 \), fix \( R \) s.t. \( R \geq \frac{\sqrt{\log(k)/\epsilon_2}}{2 \log(1+2\epsilon_2)} \) for \( d \geq 2 \), and \( R \geq \frac{\sqrt{2\log(1+2\epsilon_2)}}{\epsilon_2} \) for \( d = 1 \). Let \( S = \{ s(1), \ldots, s(m) \} \) be \( m \) independent samples of the Gaussian vector \( s \sim N(0, R^2 I_{d \times d}) \). For \( m \geq \frac{\sqrt{2 \log(k)/\epsilon_2}}{\epsilon_2} \), with probability at least \( 1 - \delta \), over the random sampling, we can bound the condition number of the matrix \( V_S \) by \( \text{cond}_2(V_S) \leq \sqrt{\frac{2 \log(k)}{\epsilon_2}} \).

Application to Learning GMMs

Setup: \( k \)-mixture of \( d \)-dim spherical Gaussians

- Parameters: \( \{ (w_j, \mu(j), \Sigma(j) = \sigma^2 I_{d \times d}) \} \in [k] \)
- Problem: what condition permits efficient learning algorithm
- Minimum Separation: \( \Delta_G = \min_{s, j} \| \mu(j) - \mu(s) \|_2 / \sigma \)

- Moment generating function: \( \phi_X(s) = E[e^{\epsilon s^T x}] = \sum_{j \in [k]} w_j e^{-\frac{1}{2} \sigma^2 \| x \|^2} e^{\epsilon s^T \mu(j)} \)
- Empirical MGF: \( \hat{\phi}(s) = \frac{1}{n} \sum_{i \in [n]} e^{\epsilon s^T x(i)} \)

Corollary: Recover the scaling result of (Dasgupta 99)

poly time algorithm if \( \Delta_G \geq \Omega(d^{1/2}) \)

Future Works

- Apply the idea to learn general case well-separated GMMs
- Reduce sample complexity to info optimal \( \Omega(kd) \)

Selected references