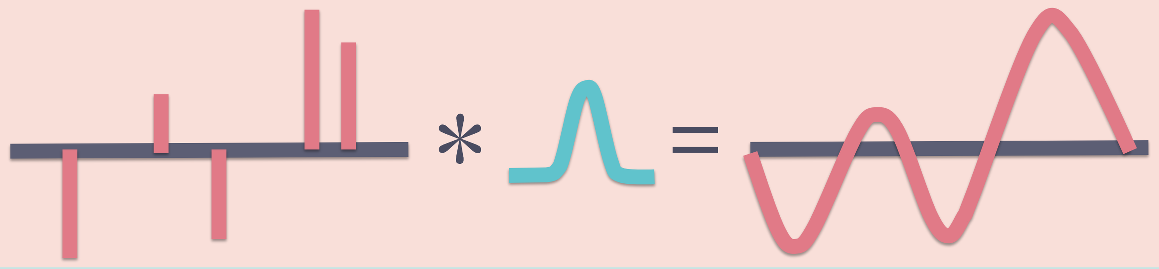


Super-resolution off the grid

Qingqing Huang
MIT

Sham M Kakade
University of Whashington

Introduction

✧ **Problem:** recover a superposition of point sources

(k-sum d-dim point sources) $x(t) = \sum_{j=1}^{k} w_j \delta_{\mu^{(j)}}$.

using bandlimited and noise corrupted measurements.

(Fourier measurements) $f(s) = \int_{t\in\mathbb{R}^d} e^{i\pi<t,s>} x(dt) = \sum_{j=1}^{k} w_j e^{i\pi<\mu^{(j)},s>}$.

(Measurement noise) $\widetilde{f}(s) = f(s) + z(s), \quad |z(s)| \le \epsilon_z, \forall s.$

✧ **Goal:** 1. use coarse Fourier measurements (cutoff freq)
2. take small number of measurements
3. run quickly.

(Minimum separation) $\Delta = \min_{j\neq j'} \|\mu^{(j)} - \mu^{(j')}\|_2$

✧ **Stable recovery:**

$$\min_{\pi} \max \left\{ \|\widehat{\mu}^{(j)} - \mu^{(\pi(j))}\|_2 : j \in [k] \right\} \le \mathrm{poly}(d,k)\epsilon_z,$$

Previous works

✧ Super-resolution, 1-d

✧ Take uniform samples on a grid
$\{f(-N), f(-N+1), \dots, f(N-1), f(N)\}$

✧ Cut-off freq determined by the minimum separation
$N = \Omega(\frac{1}{\Delta})$

✧ Compressed sensing off the grid
use $k\log(k)$ random samples to recover $2N$ samples

✧ Multiple dimensions

✧ Project to multiple 1-d problems (reduces separation)

✧ Grid the multi-dim space (computationally expensive)
both approaches $\Omega(poly(k)^d)$

Basic ideas of our work

✧ Run Prony's method on **random samples** (Matrix−Pencil / MUSIC / ESPRIT)

✧ Skip the intermediate step of recovering all the $\Omega(N^d)$ samples on the hyper-grid, random samples $\rightarrow$ estimate point sources

✧ Sample at $\mathcal{S}$ such that F admits the tensor decomposition:

$F = V_{S'} \otimes V_{S'} \otimes (V_2 D_w)$, (Rank-k 3-way tensor)

$$V_S = \underbrace{\begin{bmatrix} e^{i\pi<\mu^{(1)},s^{(1)}>} & \dots & e^{i\pi<\mu^{(k)},s^{(1)}>} \\ e^{i\pi<\mu^{(1)},s^{(2)}>} & \dots & e^{i\pi<\mu^{(k)},s^{(2)}>} \\ \vdots & \dots & \vdots \\ e^{i\pi<\mu^{(1)},s^{(m)}>} & \dots & e^{i\pi<\mu^{(k)},s^{(m)}>} \end{bmatrix}}_{d\times k}.$$

(Vandermonde Matrix with complex nodes)

✧ Prony's method as a special case

Algorithm

Input: Cutoff freq R, number of measurements m.

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

1. **Take measurements at random s:**

   Let $\mathcal{S} = \{s^{(1)}, \dots, s^{(m)}\}$ be m i.i.d. samples from the Gaussian $\mathcal{N}(0, R^2 I_{d\times d})$. Set $s^{(m+n)} = e_n$ 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 $\widetilde{F} \in \mathbb{C}^{m'\times m'\times 3}$:

   $$\widetilde{F}_{n_1,n_2,n_3} = \widetilde{f}(s)|_{s=s^{(n_1)}+s^{(n_2)}+v^{(n_3)}}.$$

2. **Tensor Decomposition:** Set $(\widehat{V}_{S'}, \widehat{D}_w) = \mathrm{TensorDecomp}(\widetilde{F})$.

   For $j = 1, \dots, k$, set $[\widehat{V}_{S'}]_j = [\widehat{V}_{S'}]_j / [\widehat{V}_{S'}]_{m',j}$

3. **Read of estimates:** For $j = 1, \dots, k$, set $\widehat{\mu}^{(j)} = Real(\log([\widehat{V}_S]_{[m+1:m+d,j]})/(i\pi))$.

4. Set $\widehat{W} = \arg\min_{W\in\mathbb{C}^k} \|\widetilde{F} - \widehat{V}_{S'} \otimes \widehat{V}_{S'} \otimes \widehat{V}_d D_w\|_F$.

✧ 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 Δ

	$d=1$			$d\ge 1$		
	cutoff freq	measurements	runtime	cutoff freq	measurements	runtime
SDP	$\frac{1}{\Delta}$	$k\log(k)\log(\frac{1}{\Delta})$	$poly(\frac{1}{\Delta}, k)$	$\frac{C_d}{\Delta_\infty}$	$(\frac{1}{\Delta_\infty})^d$	$poly((\frac{1}{\Delta_\infty})^d, k)$
MP	$\frac{1}{\Delta}$	$\frac{1}{\Delta}$	$(\frac{1}{\Delta})^3$	-	-	-
Ours	$\frac{1}{\Delta}$	$(k\log(k))^2$	$(k\log(k))^2$	$\frac{\log(kd)}{\Delta}$	$(k\log(k)+d)^2$	$(k\log(k)+d)^2$

Table 1: Comparison of results. We are implicitly using $O(\cdot)$ notation here.

Main results

✧ **Main Theorem** (stability of our algorithm)

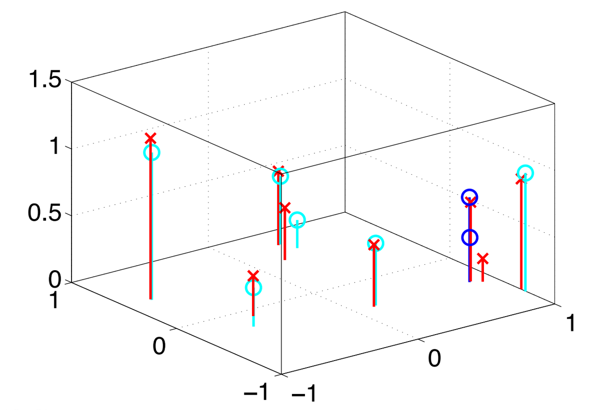
Theorem 1. *For a fixed error probability δ, the algorithm achieves stable recovery with number of measurements and runtime both bounded by $O((k\log(k)+d)^2)$. The frequency range of the measurements are bounded by $O(1/\Delta)$ (ignoring log factors).*

✧ **Key Lemma** (condition number of random Vandermonde)

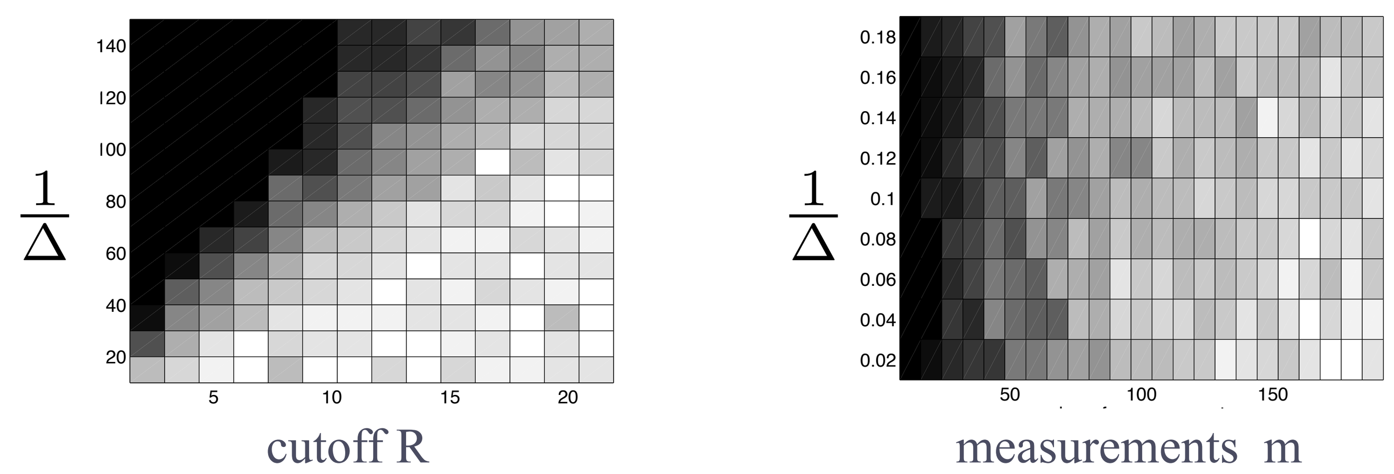
Lemma 1. *For fixed ϵ_x, fix R s.t. $R \ge \frac{\sqrt{2\log(k/\epsilon_x)}}{\pi\Delta}$ for $d \ge 2$, and $R \ge \frac{\sqrt{2\log(1+2/\epsilon_x)}}{\pi\Delta}$ for $d = 1$. Let $\mathcal{S} = \{s^{(1)}, \dots, s^{(m)}\}$ be m independent samples of the Gaussian vector $s \sim \mathcal{N}(0, R^2 I_{d\times d})$. For $m \ge \frac{k}{\epsilon_x}\sqrt{8\log\frac{k}{\delta_s}}$, with probability at least $1-\delta_s$ over the random sampling, we can bound the condition number of the factor matrix V_S by $cond_2(V_S) \le \sqrt{\frac{1+2\epsilon_x}{1-2\epsilon_x}}$.*

Simulation Results

✧ 2-D example



✧ Phase transitions



✧ Runs fast, just one SVD, no SDP!

Application to Learning GMMs

✧ **Setup:** k-mixture of d-dim spherical Gaussians

(Parameters) $\{(w_j, \mu^{(j)}, \Sigma^{(j)} = \sigma^2 I_{d\times d})\}_{i\in[k]}$

✧ **Problem:** what condition permits efficient learning algorithm

(Minimum Separation) $\Delta_G = \frac{\min_{j\neq j'} \|\mu^{(j)} - \mu^{(j')}\|_2}{\sigma}$,

(Moment generating function)

$$\phi_X(s) = \mathbb{E}[e^{i<x,s>}] = \sum_{j\in[k]} w_j e^{-\frac{1}{2}\sigma^2\|s\|_2^2 + i<\mu^{(j)},s>}.$$

(Empirical MGF) $\widehat{\phi}_X(s) = \frac{1}{N}\sum_{l\in[N]} e^{i<x_l,s>}$.

✧ **Corollary:** Recover the scaling result of (Dasgupta 99)

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

Future Works

✧ Apply the idea to learn general case well-separated GMMs

✧ Reduce sample complexity to info optimal $\Omega(kd)$

Selected references

✧ *G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. Information Theory, IEEE Transactions on, 59(11):7465–7490, 2013.*

✧ *E. J. Cand`es and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. Communications on Pure and Applied Mathematics, 67(6):906–956, 2014.*

✧ *S. Leurgans, R. Ross, and R. Abel. A decomposition for three-way arrays. SIAM Journal on Matrix Analysis and Applications, 14(4):1064–1083, 1993.*

✧ *S. Dasgupta. Learning mixtures of gaussians. In Foundations of Computer Science, 1999. 40th Annual Symposium on, pages 634–644. IEEE, 1999.*