CS 59300 – Algorithms for Data Science

Classical and Quantum approaches

Lecture 7 (09/25)

Super-resolution (II)

https://ruizhezhang.com/course_fall_2025.html

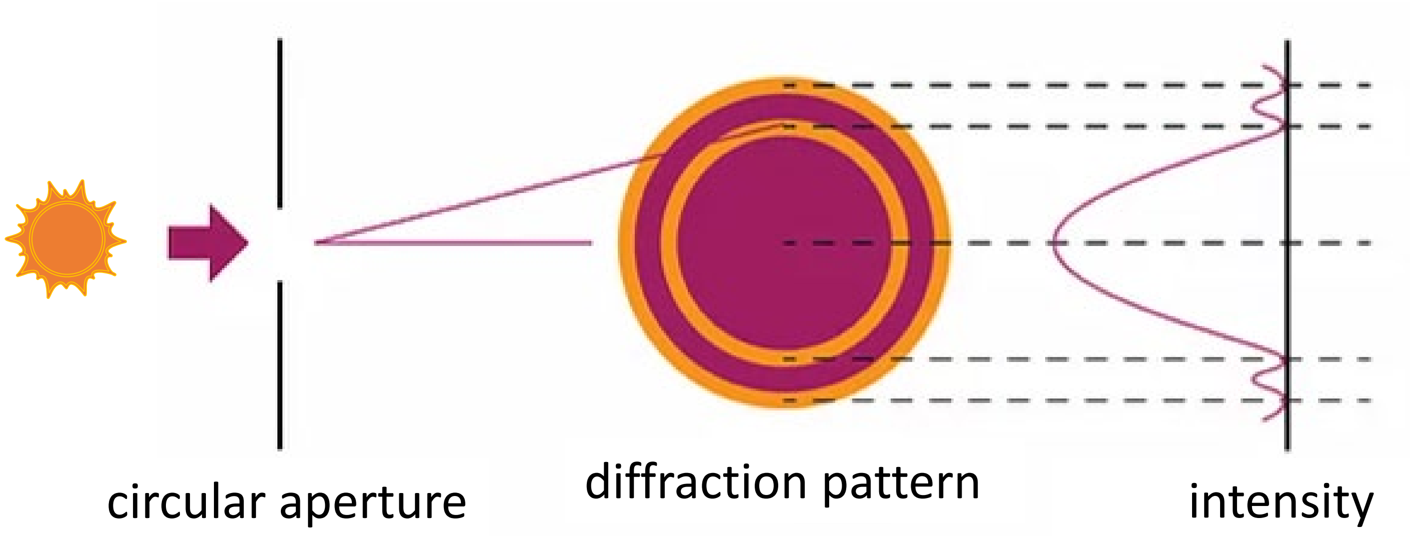
Last lecture, we discussed the 1-D super-resolution of point-sources.

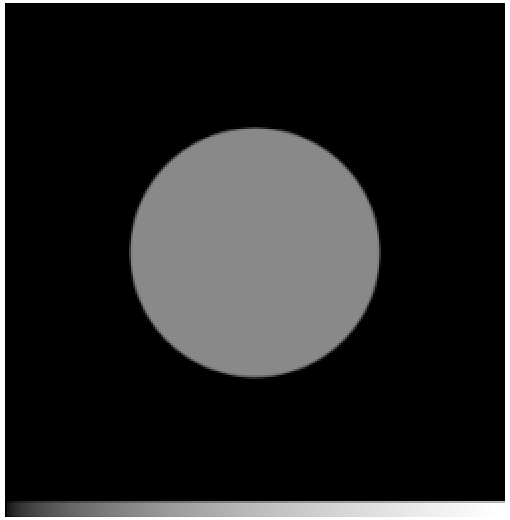
Today, we turn to a more physical problem—one with an even longer history:

- The diffraction limit in optics (a.k.a. learning mixture of Airy disks)

The physics of diffraction

When light from a point source passes through a small circular aperture, it does not produce a bright dot as an image, but rather a diffuse circular disc known as Airy disk

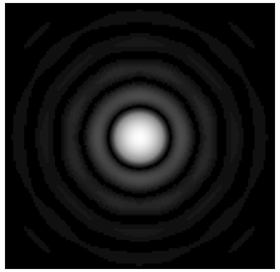




The physics of diffraction

The Airy disk has the following normalized intensity function:

$$I(x) = \frac{1}{\pi\sigma^2}\left(\frac{2J_1(\|x\|/\sigma)}{\|x\|/\sigma}\right)^2 \longleftrightarrow$$



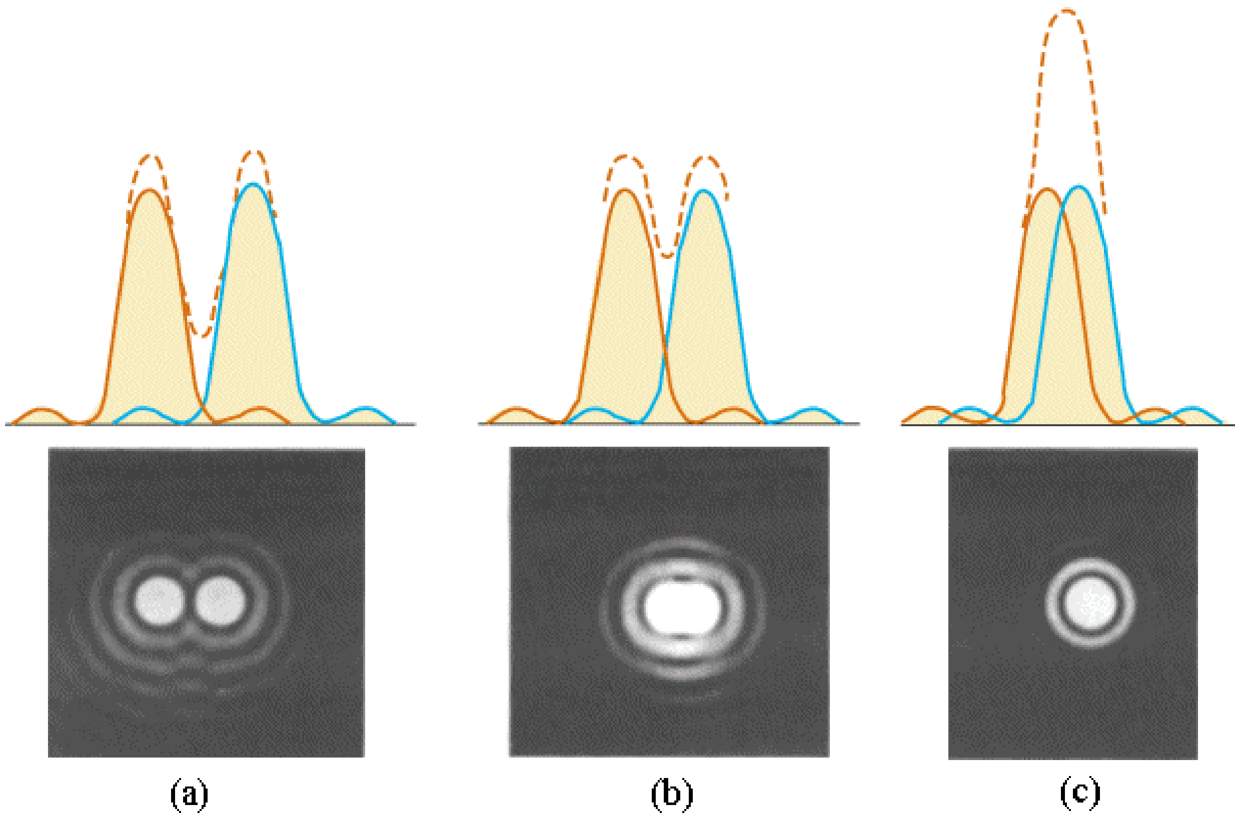
where J_1 is the Bessel function of the first kind, and σ is a spread parameter governed by physical properties (such as numerical aperture) that quantifies the degree of blur

- $I(x)$ can be interpreted as the infinitesimal probability of detecting a photon at x (in quantum optics theory)

The physics of diffraction

For >150 years, it has been widely believed that physics imposes fundamental limits to resolution

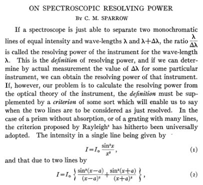
If two Airy disks are too close, the blur makes it impossible to distinguish them



Are there statistical/algorithmic limitations to how accurately we can estimate a mixture of Airy disks?

The diffraction limit

In particular, what is the minimum separation?



ON SPECTROSCOPIC RESOLVING POWER

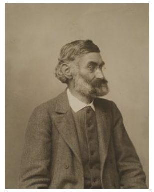
By C. M. SPARROW

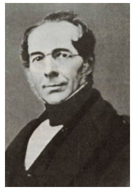
If a spectroscope is just able to separate two monochromatic lines of equal intensity and wave-lengths λ and $\lambda+\Delta\lambda$, the ratio $\frac{\lambda}{\Delta\lambda}$ is called the resolving power of the instrument for the wave-length λ. This is the *definition* of resolving power, and if we can determine by actual measurement the value of $\Delta\lambda$ for some particular instrument, we can obtain the resolving power of that instrument. If, however, our problem is to calculate the resolving power from the optical theory of the instrument, the *definition* must be supplemented by a *criterion* of some sort which will enable us to say when the two lines are to be considered as just resolved. In the case of a prism without absorption, or of a grating with many lines, the criterion proposed by Rayleigh[1] has hitherto been universally adopted. The intensity in a single line being given by

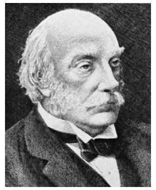
$$I = I_0 \frac{\sin^2 x}{x^2}, \qquad (1)$$

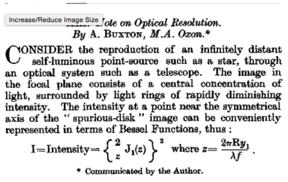
and that due to two lines by

$$I = I_0 \left\{ \frac{\sin^2(x-a)}{(x-a)^2} + \frac{\sin^2(x+a)}{(x+a)^2} \right\}, \qquad (2)$$









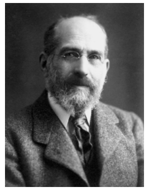
Note on Optical Resolution.

By A. Buxton, *M.A. Oxon.**

CONSIDER the reproduction of an infinitely distant self-luminous point-source such as a star, through an optical system such as a telescope. The image in the focal plane consists of a central concentration of light, surrounded by light rings of rapidly diminishing intensity. The intensity at a point near the symmetrical axis of the "spurious-disk" image can be conveniently represented in terms of Bessel Functions, thus:

$$I = \text{Intensity} = \left\{ \frac{2}{z} J_1(z) \right\}^2 \text{ where } z = \frac{2\pi R y_1}{\lambda f}.$$

* Communicated by the Author.



Sparrow Abbe Dawes Houston Rayleigh Buxton Schuster

0.94 1 1.02 1.03 1.22 1.46 2.44

pairwise separation ($\times\ \pi\sigma$)

Which, if any, of these criteria is the right one?

A persistent debate

In 1879 Lord Rayleigh proposed a heuristic that is still widely used

"This rule is convenient on account of its simplicity and it is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution."

Subsequently, many other refinements were proposed based on different sorts of arguments, with varying degrees of rigor

"It is obvious that the undulation condition should set an upper limit to the resolving power ... My own observations on this point have been checked by a number of friends and colleagues."

Carroll Sparrow, 1918

A persistent debate

Others pushed back on there being a diffraction limit at all

"It seems a little pedantic to put such precision into the resolving power formula … Actually, if sufficiently careful measurements of the exact intensity distribution over the diffracted image can be made, the fact that two sources make the spot can be proved [regardless of separation]."

Richard Feynman, 1964

Nevertheless, there is decades of empirical evidence that there actually does seem to be a limit to what we can resolve?

Can we put the diffraction limit on a rigorous foundation?

Learning mixture of Airy disks

Setup:

- There are k Airy disks centered at unknown points $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k \in \mathbb{R}^2$
- Density for the i-th Airy disk is $I(\boldsymbol{x} - \boldsymbol{\mu}_i)$
- The minimum separation $\Delta \coloneqq \min_{i \neq j \in [k]} \left\| \boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right\|$
- We get access to i.i.d. samples from the distribution

$$I(\boldsymbol{x}) = \frac{1}{\pi\sigma^2}\left(\frac{2J_1(\|\boldsymbol{x}\|/\sigma)}{\|\boldsymbol{x}\|/\sigma}\right)^2$$

$$\rho(\boldsymbol{x}) = \sum_{i=1}^{k} \lambda_i I(\boldsymbol{x} - \boldsymbol{\mu}_i)$$

$\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$

- **Goal:** estimate $\mu_1, \dots, \mu_K$

Main result 1

Theorem (Chen-Moitra ’20).

Given samples from a Δ-separated mixture of k Airy disks where each relative intensity is at least λ, there is an algorithm that takes

$$\mathrm{poly}\left((k\sigma/\Delta)^{k^2}, 1/\lambda, 1/\epsilon\right)$$

samples and learns within error ϵ with high probability

Remark. For two Airy disks (the focus of the debate), there is no fundamental limitation to what can be resolved!

Main result 2

When the number of centers is large there is a phase transition

Theorem (Chen-Moitra '20).

- If the k Airy disks are $1.53\pi\sigma$-separated, there is a polytime algorithm that takes
$$\mathrm{poly}(k, 1/\Delta\,, 1/\lambda\,, 1/\epsilon)$$
samples and learns within error ϵ with high probability

- There are $< 1.15\pi\sigma$-separated mixtures of k Airy disks that require $\exp\left(\Omega\left(\sqrt{k}\right)\right)$ samples to learn

Remark. With any reasonable physical setup, there really is a fundamental limit to resolving many point sources

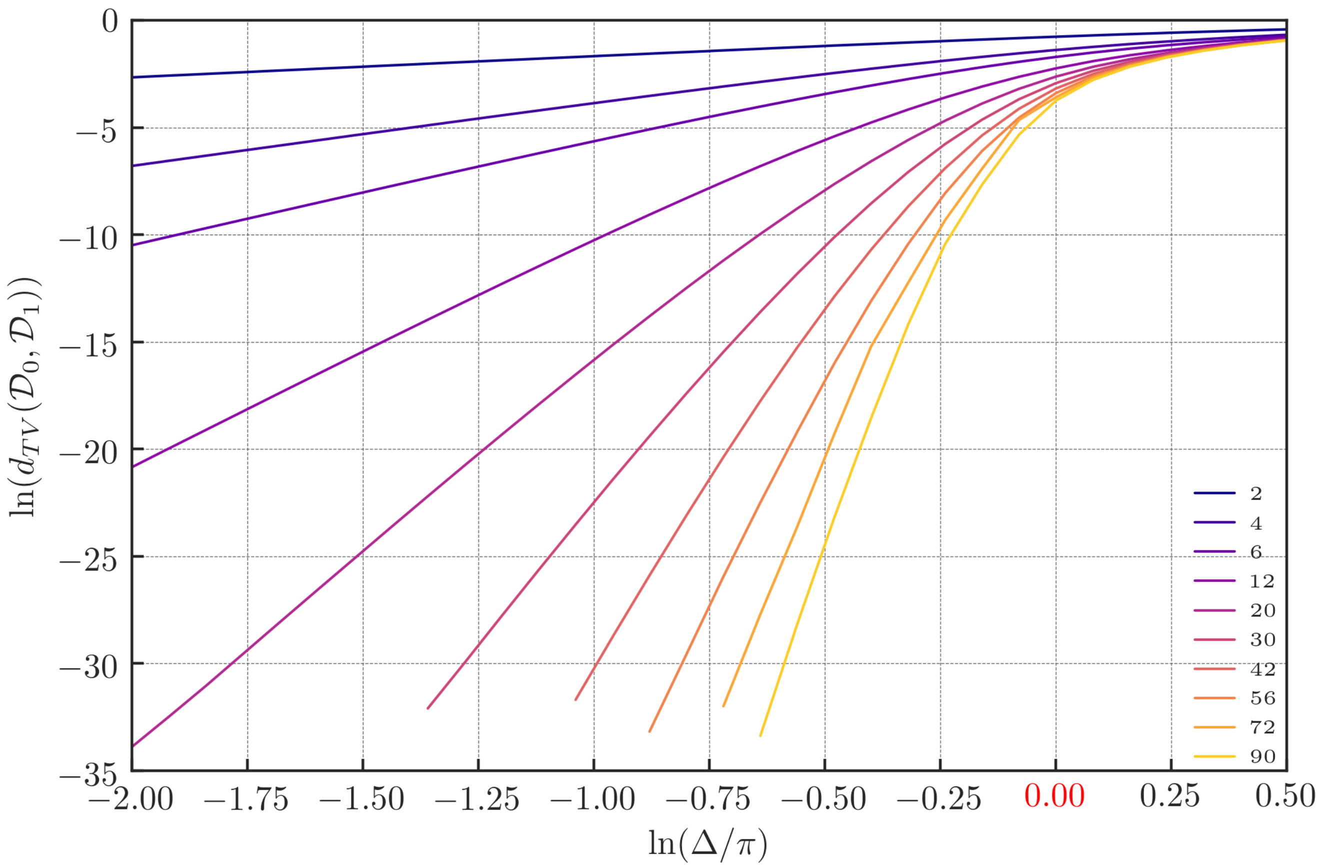
Interpretation

Opposing views on the diffraction limit:

- In domains where there are few close-by sources (e.g. astronomy), it is possible to resolve below the diffraction limit
- In domains where there are many close-by sources (e.g. microscopy), diffraction imposes fundamental limit on resolution

Visualizing the diffraction limit

In 1-D, we know the precise threshold ($\Delta = \pi\sigma$), the Abbe limit, and can visualize how resolution undergoes a phase transition



Deconvolution

Last lecture, we showed how to learn the locations and intensities of a 1-D Fourier signal

$$g(\omega) = \sum_{j=1}^{k} u_j e^{2\pi \mathbf{i} f_j \omega}$$

- Diffracted image:

$$\rho(\boldsymbol{x}) = \sum_{j=1}^{k} \lambda_j I(\boldsymbol{x} - \boldsymbol{\mu}_j)$$

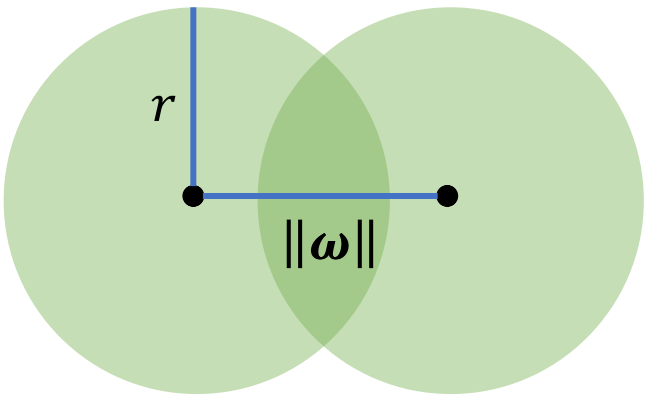
- Its Fourier transform:

$$\hat{\rho}(\boldsymbol{\omega}) = \sum_{j=1}^{k} \lambda_j \hat{I}(\boldsymbol{\omega}) e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega} \rangle}$$

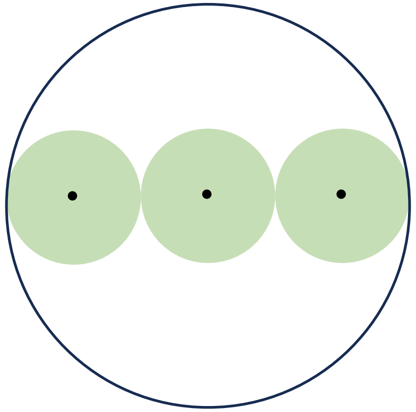
where $\hat{I}(\boldsymbol{\omega}) = \frac{2}{\pi}\left(\arccos(\pi\sigma\|\boldsymbol{\omega}\|) - \pi\sigma\|\boldsymbol{\omega}\|\sqrt{1 - \pi^2\sigma^2\|\boldsymbol{\omega}\|^2}\right)$

$$= 4\pi\sigma^2 \cdot 1_{B(r)}(\boldsymbol{\omega}) \star 1_{B(r)}(\boldsymbol{\omega}) \text{ with } r = \frac{1}{2\pi\sigma}$$

2-D convolution:



Deconvolution via division



- The support of $\hat{I}(\boldsymbol{\omega})$ is $B\left(\frac{1}{\pi\sigma}\right)$ (wlog, assume $\sigma = 1/\pi$)
- Thus, for $\boldsymbol{\omega} \in \mathbb{R}^2$ with $\|\boldsymbol{\omega}\| \le 1$, we can simulate the 2-D Fourier signal:

$$g(\boldsymbol{\omega}) := \frac{\hat{\rho}(\boldsymbol{\omega})}{\hat{I}(\boldsymbol{\omega})} = \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_j, \boldsymbol{\omega}\rangle}$$

- We only get samples from $\rho(\boldsymbol{x})$. How to get access to $\hat{\rho}(\boldsymbol{\omega})$?

$$\hat{\rho}(\boldsymbol{\omega}) \approx \frac{1}{N}\sum_i \cos(2\pi\langle \boldsymbol{\omega}, \boldsymbol{x}_i\rangle)$$

- $\mathbb{E}_{\boldsymbol{x}\sim\rho}[\cos(2\pi\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle)] = \Re\left(\mathbb{E}_{\boldsymbol{x}\sim\rho}\left[e^{-2\pi\mathbf{i}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle}\right]\right) = \Re\left(\hat{\rho}(\boldsymbol{\omega})\right) = \hat{\rho}(\boldsymbol{\omega})$

2-D super-resolution

Setup:

- Given access to measurements with $\|\boldsymbol{\omega}\| < 1$:

$$g(\boldsymbol{\omega}) = \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_j, \boldsymbol{\omega}\rangle} + \eta_{\boldsymbol{\omega}}$$

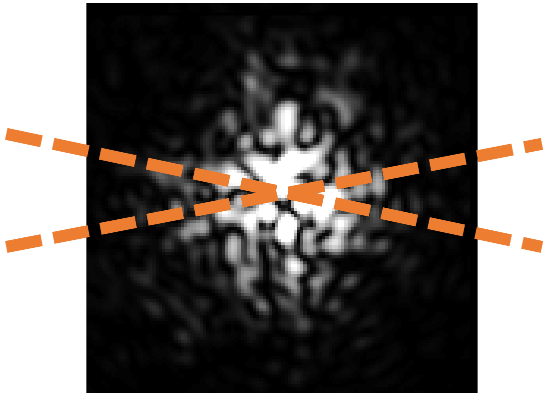
- **Goal:** recover $\{(\boldsymbol{\mu}_j, \lambda_j)\}_{j\in[k]}$

Two regimes:

- k is constant. Reduce to two 1-D super-resolution instances and piece the estimates together
- k is large and well-separated ($\Delta > 1.53$). Tensor-decomposition-based algorithm

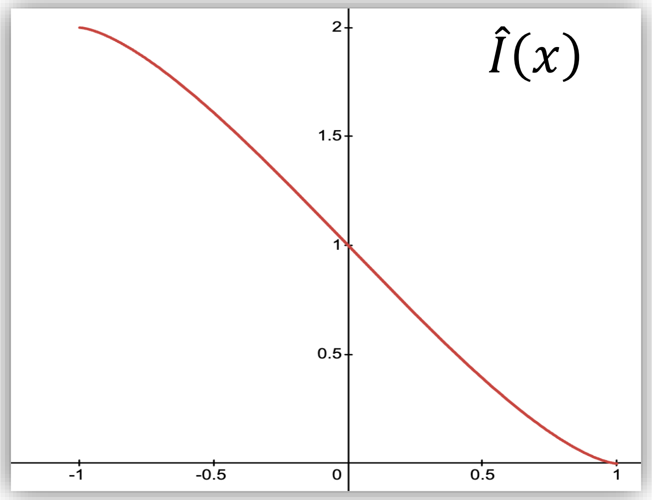
2-D super-resolution: constant k

Projection to 1-D:



1. Sample a unit vector $\boldsymbol{v} \in \mathbb{R}^2$
2. $g\left(l \frac{\boldsymbol{v}}{4k}\right) = \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{v}/(4k) \rangle l} + \eta_l = \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} f_j l} + \eta_l$
   for $l = 0, 1, \ldots, 2k-1$
3. Run ESPRIT to recover $\left\{(\lambda_j', f_j')\right\}_{j \in [k]}$
4. Repeat 1-3 and obtain $\left\{(\lambda_j'', f_j'')\right\}_{j \in [k]}$

$$\boxed{|\eta_l| \le \frac{\eta}{|\hat{I}(l/4k)|}}$$



2-D super-resolution: constant k

Piece together the 1-D estimates:

$$\begin{bmatrix} \boldsymbol{v}_1^{(1)} & \boldsymbol{v}_2^{(1)} \\ \boldsymbol{v}_1^{(2)} & \boldsymbol{v}_2^{(2)} \end{bmatrix} \begin{bmatrix} (\boldsymbol{\mu}_j)_1 \\ (\boldsymbol{\mu}_j)_2 \end{bmatrix} = \begin{bmatrix} f_j' \\ f_j'' \end{bmatrix}$$

The noise-stability of ESPRIT depends on the condition number of the Vandermonde matrix:

$$V_k = \begin{bmatrix} z_1^0 & \cdots & z_k^0 \\ \vdots & \ddots & \vdots \\ z_1^{k-1} & \cdots & z_k^{k-1} \end{bmatrix} \quad \text{where} \quad z_j := e^{-2\pi \mathbf{i} f_j}$$

- The locations are gapless: $\Delta' := \min_{i \neq j} |f_i - f_j| = \min_{i \neq j} |\langle \boldsymbol{\mu}_i - \boldsymbol{\mu}_j, \boldsymbol{v} \rangle| / 4k \sim \Delta / k$

$$V_k = \begin{bmatrix} z_1^0 & \cdots & z_k^0 \\ \vdots & \ddots & \vdots \\ z_1^{k-1} & \cdots & z_k^{k-1} \end{bmatrix} \quad \text{where} \quad z_j \coloneqq e^{-2\pi \mathbf{i} f_j}$$

- For any unit vector $\lambda \in \mathbb{C}^k$,

$$\|V_k \lambda\|^2 = \sum_{l=0}^{k-1} \left| \sum_{j=1}^{k} \lambda_j z_j^l \right|^2 \leq k^2$$

  Thus, $\sigma_{\max}(V_k) \leq k$

- For $\sigma_{\min}(V_k)$, we have

$$\sigma_{\min}(V_k) \geq \left(\prod_i \sigma_i(V_k) \right) \Big/ \left(\sigma_{\max}(V_k) \right)^{k-1} \geq |\det(V_k)| / k^{k-1}$$

$$|\det(V_k)| = \prod_{i<j} \left| z_i - z_j \right| \geq \left| e^{2\pi \mathbf{i} \Delta'} - 1 \right|^{\binom{k}{2}} \geq (\Delta')^{k(k-1)/2}$$

- Thus, $\kappa(V_k) \leq k^k (\Delta')^{-k^2} \sim (\Delta/k)^{k^2}$

Recap: phase transition of Vandermonde condition number

In last lecture, we showed that the condition number of $V_n(\boldsymbol{z})$ exhibits a sharp phase transition:

- If $n > \frac{1}{\Delta} + 1$, then $\kappa\big(V_n(\boldsymbol{z})\big) \leq \sqrt{\frac{n-1+1/\Delta}{n-1-1/\Delta}}$
- If $n < (1-\epsilon)\frac{1}{\Delta}$, then $\kappa\big(V_n(\boldsymbol{z})\big) = 2^{\Omega(\epsilon k)}$ in the worst-case

That is, if we use sufficiently high-frequency measurements, we can **always** estimate the locations with high-accuracy in 1-D

In the Airy disk case, we only get $\kappa(V_k) \leq k^k(\Delta')^{-k^2}$, which means the noise $|\eta_{\boldsymbol{\omega}}| \leq k^{-k}(\Delta')^{k^2}$, i.e. exponentially many samples are needed

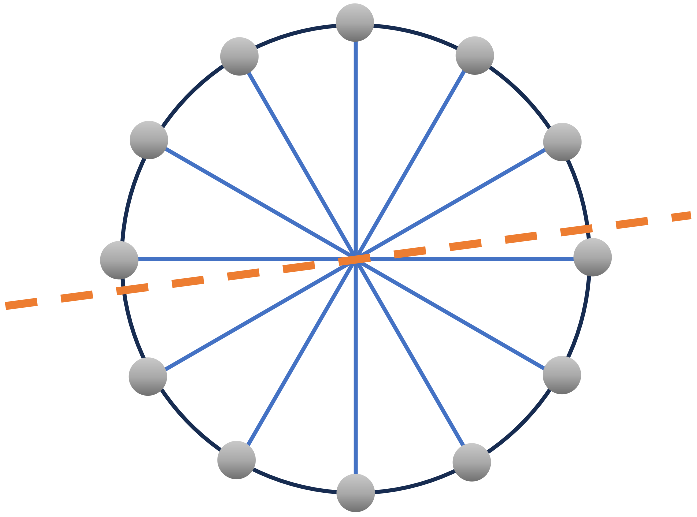
Can we get a better bound here?

1-D projection will not work for large k

1. The reduction from mixture of Airy disks to Fourier signal only works for small measurement frequencies, i.e., $\|\boldsymbol{\omega}\| \leq 1$
   → 1-D projection has cut-off frequency $\leq \mathcal{O}(1)$

2. There exists a 2-D configuration with min-separation Δ such that for every direction $\boldsymbol{v}$, the 1-D projection has min-separation $\mathcal{O}(\Delta/k)$

Always below the 1-D super-resolution limit!



2-D super-resolution: large k

- We sample random 2-D vectors $\boldsymbol{\omega}^{(1)}, \dots, \boldsymbol{\omega}^{(m)} \sim B(R)$ and $\boldsymbol{v} \sim B(1/2 - R)$
- Let $\boldsymbol{v}^{(1)} = \frac{1}{2}\boldsymbol{v}$ and $\boldsymbol{v}^{(2)} = \boldsymbol{v}$
- Construct an order-3 tensor $T \in \mathbb{C}^{m \times m \times 2}$

$$T_{abc} \coloneqq g\left(\boldsymbol{\omega}^{(a)} + \boldsymbol{\omega}^{(b)} + \boldsymbol{v}^{(c)}\right) = \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega}^{(a)} + \boldsymbol{\omega}^{(b)} + \boldsymbol{v}^{(c)} \rangle}$$

- T has the following tensor decomposition:

$$T = \sum_{j=1}^{k} V_j \otimes V_j \otimes (\lambda_j W_j)$$

where $V_j \coloneqq \left(e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega}^{(a)} \rangle}\right)_{a \in [m]} \in \mathbb{C}^m$ and $W_j \coloneqq \left(e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{v}^{(1)} \rangle} \quad e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{v}^{(2)} \rangle}\right) \in \mathbb{C}^2$

2-D super-resolution: large k

$$V = \begin{bmatrix} e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_1, \boldsymbol{\omega}^{(1)}\rangle} & \cdots & e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_k, \boldsymbol{\omega}^{(1)}\rangle} \\ \vdots & \ddots & \vdots \\ e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_1, \boldsymbol{\omega}^{(m)}\rangle} & \cdots & e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_k, \boldsymbol{\omega}^{(m)}\rangle} \end{bmatrix} \in \mathbb{C}^{m\times k}$$

- We need to upper bound the condition number $\kappa(V)$
- For any $\lambda \in \mathbb{C}^k$, we have

$$\lambda^\dagger V^\dagger V \lambda = \sum_{a=1}^{m} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_j, \boldsymbol{\omega}^{(a)}\rangle} \right|^2$$

2-D super-resolution: large k

$$V = \begin{bmatrix} e^{-2\pi\mathbf{i}\langle \boldsymbol{\mu}_1, \boldsymbol{\omega}^{(1)}\rangle} & \cdots & e^{-2\pi\mathbf{i}\langle \boldsymbol{\mu}_k, \boldsymbol{\omega}^{(1)}\rangle} \\ \vdots & \ddots & \vdots \\ e^{-2\pi\mathbf{i}\langle \boldsymbol{\mu}_1, \boldsymbol{\omega}^{(m)}\rangle} & \cdots & e^{-2\pi\mathbf{i}\langle \boldsymbol{\mu}_k, \boldsymbol{\omega}^{(m)}\rangle} \end{bmatrix} \in \mathbb{C}^{m\times k}$$

- We need to upper bound the condition number $\kappa(V)$
- For any unit vector $\lambda \in \mathbb{C}^k$, we have

$$\mathbb{E}_{\boldsymbol{\omega}^{(1)},\dots,\boldsymbol{\omega}^{(m)}}\left[\lambda^\dagger V^\dagger V \lambda\right] = m \int_{B(R)} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi\mathbf{i}\langle \boldsymbol{\mu}_j, \boldsymbol{\omega}\rangle} \right|^2 d\psi(\boldsymbol{\omega})$$

- $\|V_{i:}\| = \sqrt{k}$ and $\left\|V_{i:}^\dagger V_{i:}\right\| = k$
- By matrix Chernoff bound applied to the random matrices $V_{1:}^\dagger V_{1:}, \dots, V_{m:}^\dagger V_{m:}$

$$\Pr\left[\left\|V^\dagger V - \mathbb{E}\left[V^\dagger V\right]\right\| \geq \sqrt{m}kt\right] \leq ke^{-\Omega(t^2)}$$

- Thus, $\lambda V^\dagger V \lambda \in \mathbb{E}\left[\lambda V^\dagger V \lambda\right] \pm \tilde{O}(\sqrt{m}k)$ with high probability

Interlude: Matrix Chernoff (Bernstein) bound

Let $X_1, \dots, X_n \in \mathbb{C}^{k\times k}$ be independent, random, self-adjoint matrices satisfying:

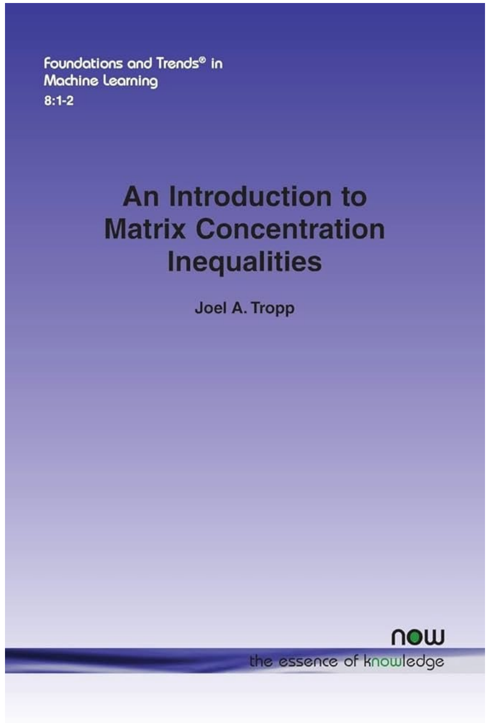
$$\mathbb{E}[X_i] = 0 \text{ and } \|X_i\| \leq R$$

Define the variance proxy:

$$\sigma^2 := \left\|\sum_i X_i^2\right\|$$

Then, for any $t > 0$,

$$\Pr\left[\left\|\sum_i X_i\right\| > t\right] \leq d \cdot \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right)$$



2-D super-resolution: large k

Lemma. We have

$$\Omega(\Delta - \overline{\gamma}) \leq \int_{B(R)} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega} \rangle} \right|^2 d\psi(\boldsymbol{\omega}) \leq k,$$

where Δ is the minimum separation of $\{\boldsymbol{\mu}_j\}$, $\overline{\gamma} \coloneqq \frac{2 j_{0,1}}{\pi} \approx 1.53$, and $R \coloneqq \frac{\overline{\gamma}}{\overline{\gamma}+\Delta} < \frac{1}{2}$

Upper bound:

$$\int_{B(R)} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega} \rangle} \right|^2 d\psi(\boldsymbol{\omega}) \leq k \|\lambda\|^2 \int_{B(R)} d\psi(\boldsymbol{\omega}) = k$$

2-D super-resolution: large k

Lower bound:

$$\int_{B(R)} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega} \rangle} \right|^2 d\psi(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega} \rangle} \right|^2 \frac{1}{\pi R^2} \mathbf{1}_{B(R)}(\boldsymbol{\omega}) d\boldsymbol{\omega}$$

- As in the 1-D super-resolution, we'll use a 2-D minorant for $\psi(\boldsymbol{\omega})$

Gonçalves '18: There exists a function $M(\omega)$ such that:

1. $M(\omega) \leq \psi(\boldsymbol{\omega})$, i.e. it minorizes the ball
2. $\operatorname{supp}\left(\widehat{M}(x)\right) \subset B(\Delta)$, i.e. it is smooth
3. $\widehat{M}(0) = \Omega(\Delta - \overline{\gamma})$, i.e. it is a non-trivial approximation

2-D super-resolution: large k

$$\int_{\mathbb{R}^2} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega} \rangle} \right|^2 \psi(\boldsymbol{\omega}) d\boldsymbol{\omega} \geq \int \sum_{j,j'} \lambda_{j'}^* \lambda_j e^{-2\pi \mathbf{i} \left\langle \boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'}, \boldsymbol{\omega} \right\rangle} M(\boldsymbol{\omega}) d\boldsymbol{\omega}$$

$$\int e^{-2\pi \mathbf{i} \left\langle \boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'}, \boldsymbol{\omega} \right\rangle} M(\boldsymbol{\omega}) d\boldsymbol{\omega} = \widehat{M}\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'}\right)$$

- Since $\left\| \boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'} \right\| > \Delta$ for $j \neq j'$, we have $\widehat{M}\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'}\right) = 0$
- Hence,

$$\sum_{j,j'} \lambda_{j'}^* \lambda_j \widehat{M}\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j'}\right) = \widehat{M}(0) \|\lambda\|^2 = \widehat{M}(0) = \Omega(\Delta - \overline{\gamma})$$

- Thus, $\int_{B(R)} \left| \sum_{j=1}^{k} \lambda_j e^{-2\pi \mathbf{i} \langle \boldsymbol{\mu}_j, \boldsymbol{\omega} \rangle} \right|^2 d\psi(\boldsymbol{\omega}) = \Omega(\Delta - \overline{\gamma})$

$\blacksquare$

2-D super-resolution: large k

- Recall that by matrix Chernoff bound, w.h.p.

$$\lambda V^\dagger V \lambda \in \mathbb{E}\left[\lambda V^\dagger V \lambda\right] \pm \tilde{O}(\sqrt{m}k)$$

- And we just proved that

$$m(\Delta - \overline{\gamma}) \leq \mathbb{E}\left[\lambda V^\dagger V \lambda\right] \leq mk$$

- Thus, the condition number can be upper-bounded by:

$$\kappa(V)^2 := \frac{\max_{\|\lambda\|=1} \left|\lambda V^\dagger V \lambda\right|}{\min_{\|\lambda\|=1} \left|\lambda V^\dagger V \lambda\right|} \leq \frac{k}{\Delta - \overline{\gamma}}$$

Diffraction limit

Theorem (Chen-Moitra '20).

For any $\epsilon > 0$, let $\Delta \coloneqq (1-\epsilon)\cdot \underline{\gamma}\pi\sigma = (1-\epsilon)\cdot \sqrt{\frac{4}{3}}\pi\sigma$. There exist two Δ-separated mixtures of k Airy disks that require $\exp\left(\Omega\left(\epsilon\sqrt{k}\right)\right)$ samples to learn

- The key step is to construct Δ-separated $\{\boldsymbol{\mu}_j\}$ and $\{\boldsymbol{\mu}'_j\}$ such that:

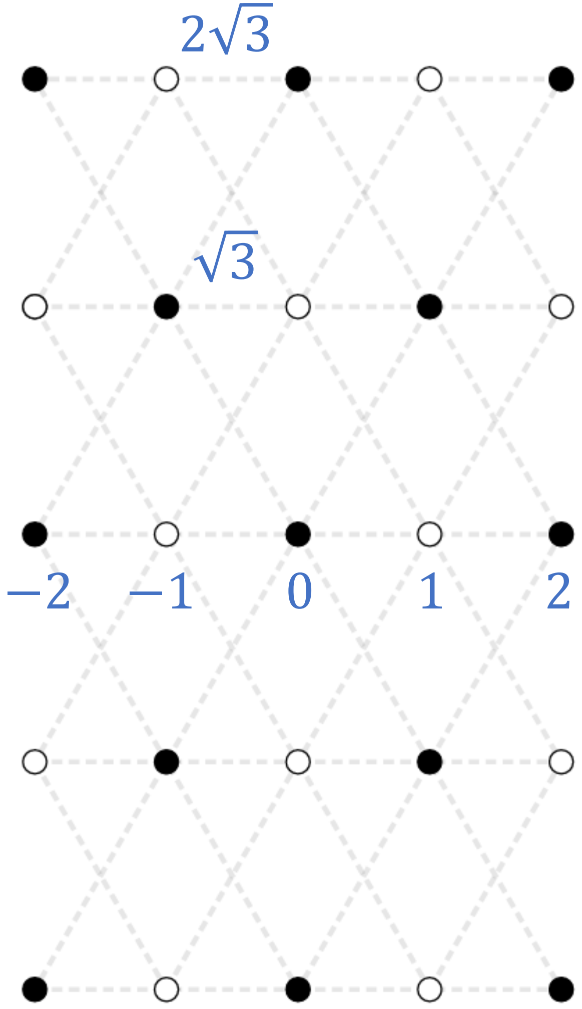
$$\left|\sum_{j=1}^{k} u_j e^{-2\pi\mathbf{i}\langle\boldsymbol{\mu}_j,\boldsymbol{\omega}\rangle} - \sum_{j=1}^{k} u'_j e^{-2\pi\mathbf{i}\langle\boldsymbol{\mu}'_j,\boldsymbol{\omega}\rangle}\right|^2 \le e^{-\Omega(\epsilon\sqrt{k})} \qquad \forall\ \|\boldsymbol{\omega}\| \le 1$$

Diffraction limit: construction

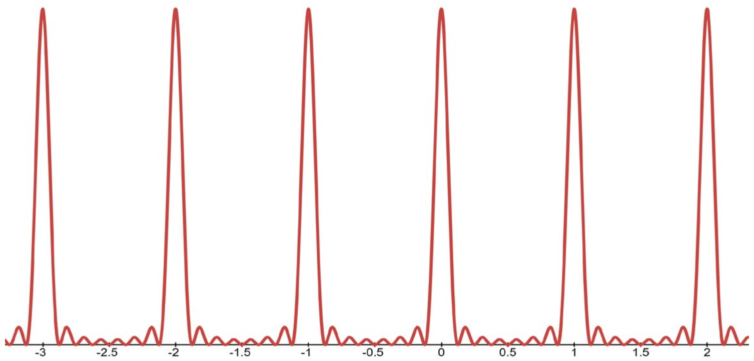
$$\nu_{j_1,j_2} \coloneqq \frac{\Delta}{2} \cdot \left(j_1, \sqrt{3} j_2\right)$$

$$j_1, j_2 \in \mathcal{J} \coloneqq \left\{-\frac{\sqrt{k}-1}{2}, \dots, \frac{\sqrt{k}-1}{2}\right\}$$

- $\{\boldsymbol{\mu}_j\} \coloneqq \{\nu_{j_1,j_2} \mid j_1 + j_2 \text{ even}\}$
- $\{\boldsymbol{\mu}'_j\} \coloneqq \{\nu_{j_1,j_2} \mid j_1 + j_2 \text{ odd}\}$



Diffraction limit: construction



- Let $\ell = \frac{4}{\epsilon}$, $r = \frac{\sqrt{k}-1}{2\ell} = \Theta\left(\epsilon\sqrt{k}\right)$, and $m = \frac{2}{\Delta}$
- Define

$$H(\boldsymbol{\omega}) \coloneqq K_\ell^r\left(\frac{\omega_1}{m} - \frac{1}{2}\right) \cdot K_\ell^r\left(\frac{\sqrt{3}\omega_2}{m} - \frac{1}{2}\right) \quad \forall\, \boldsymbol{\omega} \in \mathbb{R}^2$$

- Fourier transform:

$\widehat{K_\ell^r}(j)$

$$\widehat{H}(\boldsymbol{t}) = \sum_{j_1, j_2 \in \mathcal{J}} \frac{m^2}{\sqrt{3}} e^{-\pi \mathbf{i} m\left(t_1 + t_2/\sqrt{3}\right)} a_{j_1} a_{j_2} \delta(m t_1 - j_1) \delta\left(m t_2/\sqrt{3} - j_2\right)$$

$$= \sum_{j_1, j_2 \in \mathcal{J}} (-1)^{j_1 + j_2} a_{j_1} a_{j_2} \delta\left(\boldsymbol{t} - \nu_{j_1, j_2}\right)$$

$$= \sum_{j} a_{j_1} a_{j_2} \delta(t - \boldsymbol{\mu}_j) - \sum_{j} a_{j_1} a_{j_2} \delta\left(t - \boldsymbol{\mu}'_j\right)$$

Diffraction limit

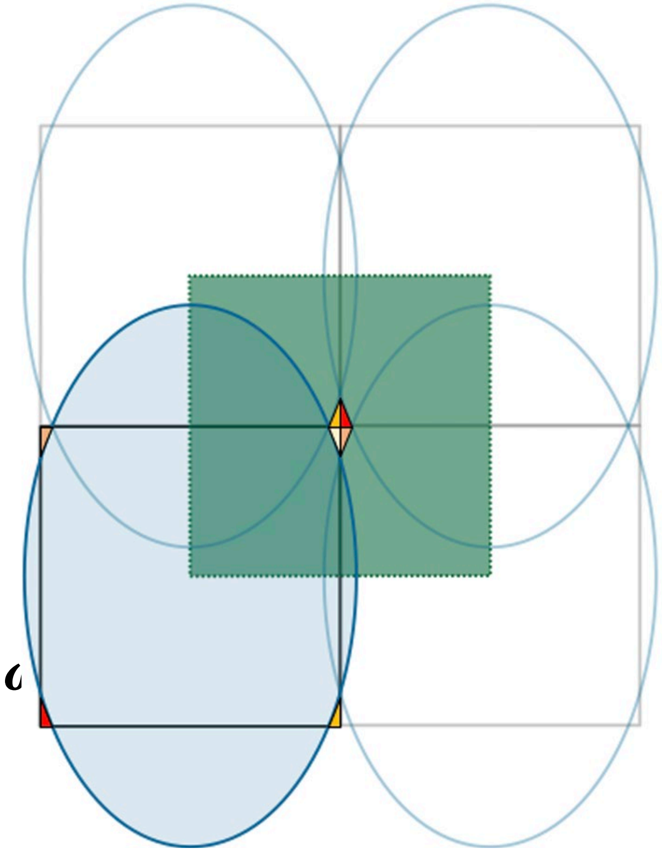
$$H(\boldsymbol{\omega}) = \int \widehat{H}(\boldsymbol{t}) e^{2\pi \mathbf{i}\langle \boldsymbol{\omega}, \boldsymbol{t}\rangle} d\boldsymbol{t} = \sum_j a_{j_1} a_{j_2} e^{2\pi \mathbf{i}\langle \boldsymbol{\omega}, \boldsymbol{\mu}_j\rangle} - \sum_j a_{j_1} a_{j_2} e^{2\pi \mathbf{i}\left\langle \boldsymbol{\omega}, \boldsymbol{\mu}'_j\right\rangle}$$

- Since $K_\ell(\omega) \le \frac{1}{4\ell^2\omega^2}$ for $\omega \in [-1/2, 1/2]$, we can show that

$$|H(\boldsymbol{\omega})| = \left| K_\ell^r\left(\frac{\omega_1}{m} - \frac{1}{2}\right) \cdot K_\ell^r\left(\frac{\sqrt{3}\omega_2}{m} - \frac{1}{2}\right)\right| \le \exp\left(-\Omega\left(\epsilon\sqrt{k}\right)\right)$$

- $\sum_j |u_j| + \sum_j |u'_j| = \sum_{j_1, j_2} a_{j_1} a_{j_2} = 1$ and $\sum_j u_j + u'_j = H(0) = 0$
- Thus, $\|u\|_1 = \|u'\|_1 = \Omega(1)$
- Hence, we have

$$\left|\sum_{j=1}^{k} u_j e^{-2\pi \mathbf{i}\langle \boldsymbol{\mu}_j, \boldsymbol{\omega}\rangle} - \sum_{j=1}^{k} u'_j e^{-2\pi \mathbf{i}\left\langle \boldsymbol{\mu}'_j, \boldsymbol{\omega}\right\rangle}\right|^2 \le e^{-\Omega(\epsilon\sqrt{k})} \qquad \forall \|\omega$$



Recap

We explored algorithms and hardness results for learning mixtures of 1-D point sources and mixtures of 2-D Airy disks

Topic not covered: Sparse Fourier transform

- **Goal:** given access to $x \in \mathbb{C}^N$, compute $\overline{x} \approx \hat{x}$
- ℓ_2/ℓ_2 guarantee:
$$\|\overline{x} - \hat{x}\|_2 \leq (1+\epsilon) \min_{k-\text{sparse } \hat{x}_k} \|\hat{x} - \hat{x}_k\|_2$$
- SOTA results: $\tilde{O}(k)$ samples and $\tilde{O}(k)$ time (sublinear algorithms)
- Generalization: continuous signals, gapless signals (Fourier interpolation), structured signals, high-dimensional SFT, …