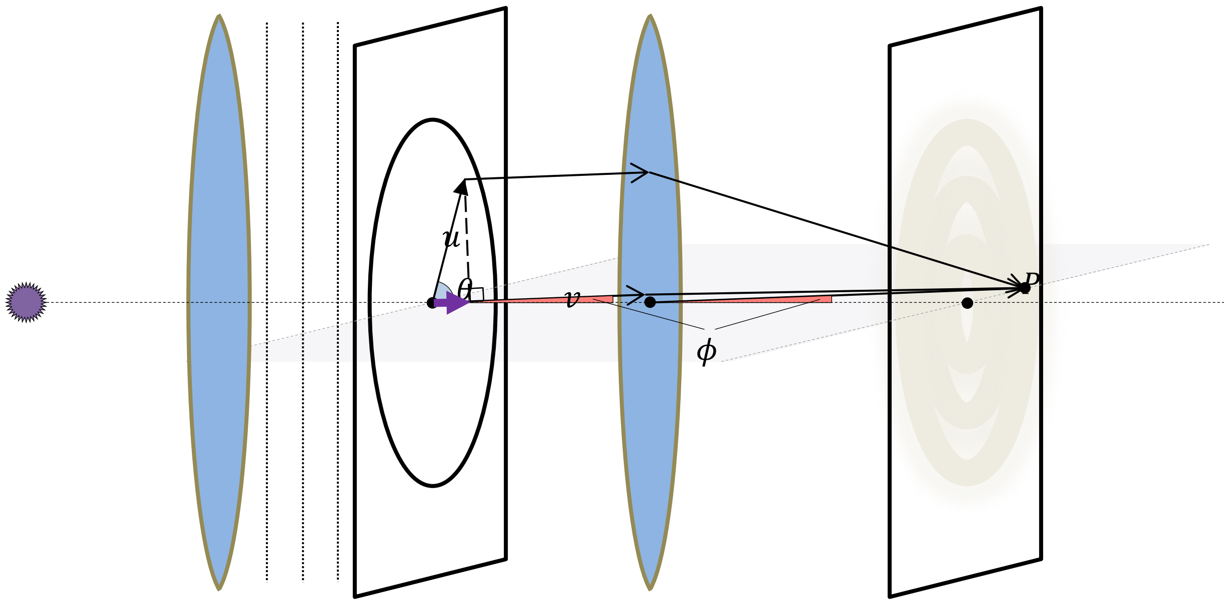
Algorithmic Foundations for the Diffraction Limit

Ankur Moitra (MIT)

UT Austin Machine Learning Lab

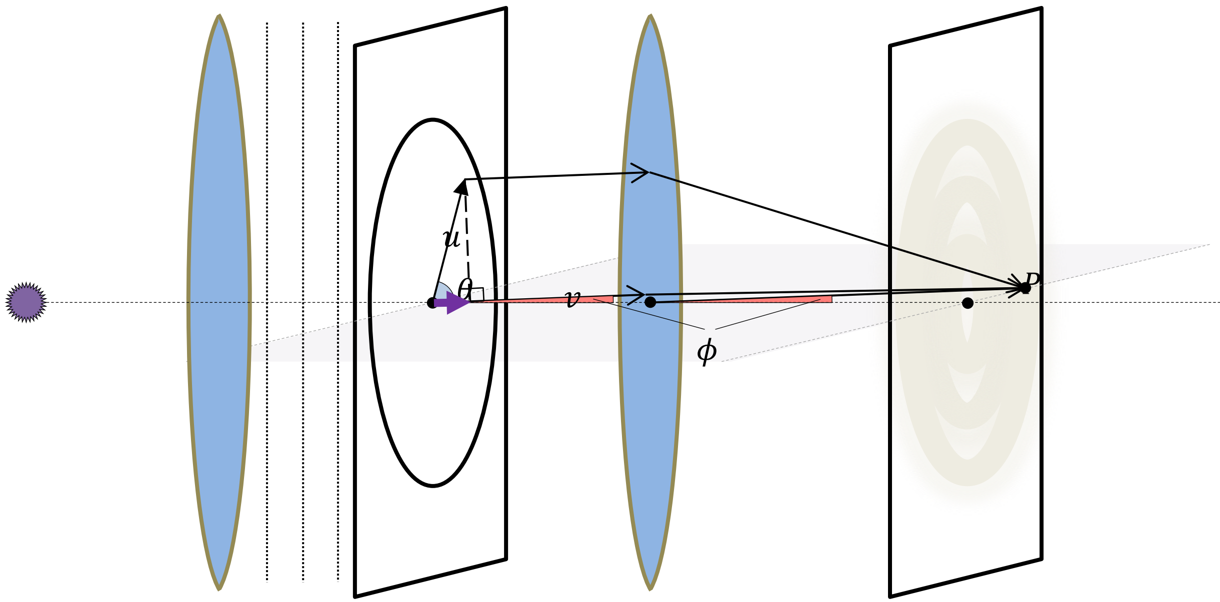
THE PHYSICS OF DIFFRACTION

Observe incoherent illumination from far-way point sources through a circular aperture



THE PHYSICS OF DIFFRACTION

Observe incoherent illumination from far-way point sources through a circular aperture

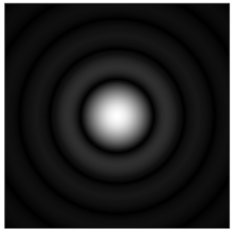


The normalized intensity is called an **Airy disk**

THE PHYSICS OF DIFFRACTION

Definition: The Airy disk is the function

$$I(x) = \frac{1}{\pi\sigma^2}\left(\frac{2J_1(\|x\|_2/\sigma)}{\|x\|_2/\sigma}\right)^2 \longleftrightarrow$$

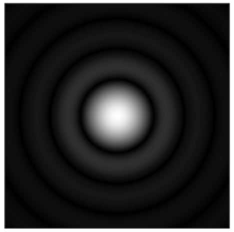


where J_1 is a **Bessel function of the first kind** and σ is the blur and is determined by physical properties (e.g. **numerical aperture**)

THE PHYSICS OF DIFFRACTION

Definition: The Airy disk is the function

$$I(x) = \frac{1}{\pi\sigma^2}\left(\frac{2J_1(\|x\|_2/\sigma)}{\|x\|_2/\sigma}\right)^2 \quad \longleftrightarrow$$



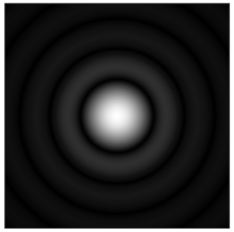
where J_1 is a **Bessel function of the first kind** and σ is the blur and is determined by physical properties (e.g. **numerical aperture**)

Interpretation: It is the infinitesimal probability of detecting a photon at some position in the observation plane

THE PHYSICS OF DIFFRACTION

Definition: The Airy disk is the function

$$I(x) = \frac{1}{\pi\sigma^2}\left(\frac{2J_1(\|x\|_2/\sigma)}{\|x\|_2/\sigma}\right)^2 \longleftrightarrow$$



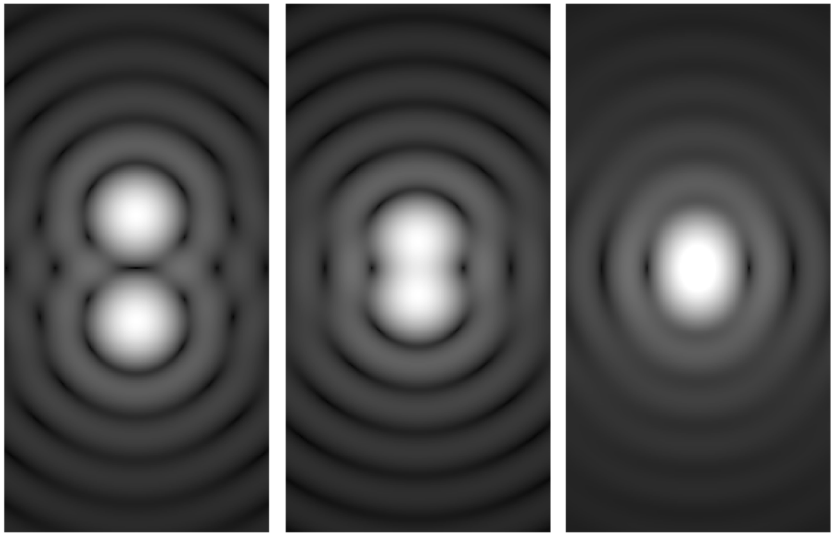
where J_1 is a **Bessel function of the first kind** and σ is the blur and is determined by physical properties (e.g. **numerical aperture**)

Interpretation: It is the infinitesimal probability of detecting a photon at some position in the observation plane

First explicitly computed by Sir George Biddell Airy in 1835

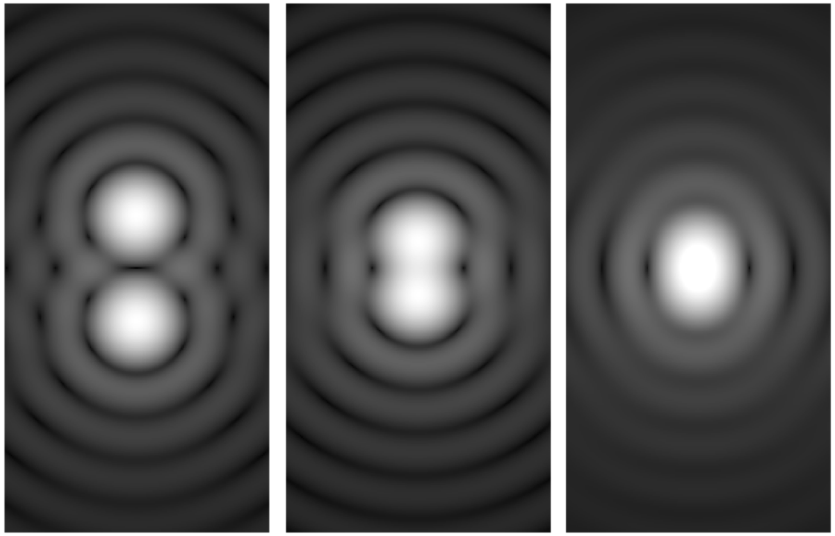
THE DIFFRACTION LIMIT

For more than a century and a half it has been widely *believed* that **physics imposes fundamental limits to resolution**



THE DIFFRACTION LIMIT

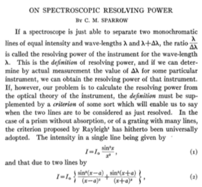
For more than a century and a half it has been widely *believed* that **physics imposes fundamental limits to resolution**



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

THE DIFFRACTION LIMIT

In particular, how should the **minimum separation** that you can resolve depend on the parameters of the optical system?



ON SPECTROSCOPIC RESOLVING POWER

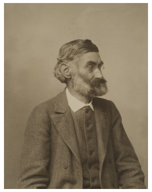
By C. M. SPARROW

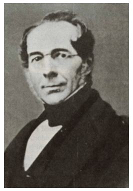
If a spectroscope is just able to separate two monochromatic lines of equal intensity and wave-lengths λ and $\lambda+\Delta\lambda$, the ratio $\frac{\lambda}{\Delta\lambda}$ is called the resolving power of the instrument for the wave-length λ. This is the *definition* of resolving power, and if we 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

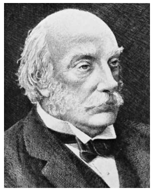
$$I=I_0\frac{\sin^2 x}{x^2}, \tag{1}$$

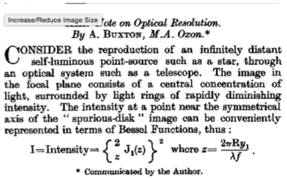
and that due to two lines by

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









Increase/Reduce Image Size

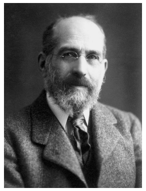
Note on Optical Resolution.

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:

$$\mathrm{I}=\text{Intensity}=\left\{\frac{2}{z}\mathrm{J}_1(z)\right\}^2 \quad \text{where } z=\frac{2\pi \mathrm{R}y_1}{\lambda f}.$$

* Communicated by the Author.

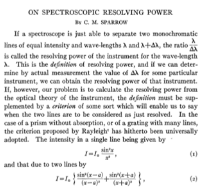


Sparrow	Abbe	Dawes	Houston	Rayleigh	Buxton	Schuster
0.94	1	1.02	1.03	1.22	1.46	2.44

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

THE DIFFRACTION LIMIT

In particular, how should the **minimum separation** that you can resolve depend on the parameters of the optical system?



ON SPECTROSCOPIC RESOLVING POWER

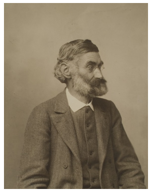
By C. M. SPARROW

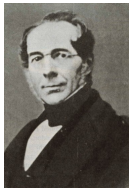
If a spectroscope is just able to separate two monochromatic lines of equal intensity and wave-lengths λ and $\lambda+\Delta\lambda$, the ratio $\frac{\lambda}{\Delta\lambda}$ is called the resolving power of the instrument for the wave-length λ. This is the *definition* of resolving power, and if we 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

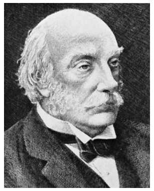
$$I = I_0 \frac{\sin^2 x}{x^2}, \qquad (1)$$

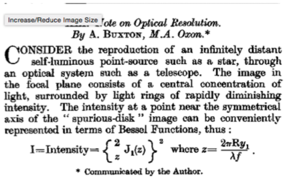
and that due to two lines by

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









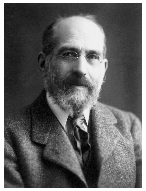
Increase/Reduce Image Size ...ote on Optical Resolution.

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:

$$\mathrm{I} = \text{Intensity} = \left\{ \frac{2}{z} \mathrm{J}_1(z) \right\}^2 \text{ where } z = \frac{2\pi R y_1}{\lambda f}.$$

* 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?

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- **The Lost Art of Debate**
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

A PERSISTENT DEBATE

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

A PERSISTENT DEBATE

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

> “ **The rule is convenient on account of its simplicity and is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution.** ”

A PERSISTENT DEBATE

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

> “ **The rule is convenient on account of its simplicity and 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

A PERSISTENT DEBATE

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

> “ **The rule is convenient on account of its simplicity and 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

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

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?

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?

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- **Rigorous Foundations and Visualizations**

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

OUR RESULTS

We give the first provable algorithms for learning mixtures of Airy disks that have non-asymptotic guarantees

OUR RESULTS

We give the first provable algorithms for learning mixtures of Airy disks that have non-asymptotic guarantees

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

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

samples and learns within error ϵ with failure probability δ

OUR RESULTS

We give the first provable algorithms for learning mixtures of Airy disks that have non-asymptotic guarantees

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

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

samples and learns within error ϵ with failure probability δ

Many arguments for the existence of a diffraction limit stem from reasoning about mixtures of two Airy disks --- but there is no fundamental limitation to what can be resolved in this setting!

OUR RESULTS, CONTINUED

At the same time, when the number of centers is large there is a phase transition

OUR RESULTS, CONTINUED

At the same time, when the number of centers is large there is a phase transition

Let $\gamma_+ \triangleq \dfrac{2j_{0,1}}{\pi} \approx 1.53$ and $\gamma_- \triangleq \sqrt{\dfrac{4}{3}} \approx 1.15$

OUR RESULTS, CONTINUED

At the same time, when the number of centers is large there is a phase transition

Let $\gamma_+ \triangleq \frac{2j_{0,1}}{\pi} \approx 1.53$ and $\gamma_- \triangleq \sqrt{\frac{4}{3}} \approx 1.15$

Theorem [Chen, Moitra '20]: If the mixture is $\gamma_+ \pi \sigma$-separated there is a polytime algorithm that takes

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

samples and learns within error ϵ with failure probability δ.

OUR RESULTS, CONTINUED

At the same time, when the number of centers is large there is a phase transition

Let $\gamma_+ \triangleq \frac{2j_{0,1}}{\pi} \approx 1.53$ and $\gamma_- \triangleq \sqrt{\frac{4}{3}} \approx 1.15$

Theorem [Chen, Moitra '20]: If the mixture is $\gamma_+\pi\sigma$-separated there is a polytime algorithm that takes

$$\text{poly}\left(k, 1/\Delta, 1/\lambda, 1/\epsilon\right)$$

samples and learns within error ϵ with failure probability δ.

Conversely there are $(1-\epsilon)\gamma_-\pi\sigma$-separated mixtures of k Airy disks that require exponentially many samples to learn

OUR RESULTS, CONTINUED

At the same time, when the number of centers is large there is a phase transition

Let $\gamma_+ \triangleq \frac{2j_{0,1}}{\pi} \approx 1.53$ and $\gamma_- \triangleq \sqrt{\frac{4}{3}} \approx 1.15$

With any reasonable physical setup (finite exposure times, finite precision in recording locations of photons) there really is a fundamental limit to resolving many point sources

Thus it is simultaneously possible that:

(1) **In domains where there are few close-by sources (e.g. astronomy) super resolution is possible**

Thus it is simultaneously possible that:

1. **In domains where there are few close-by sources (e.g. astronomy) super resolution is possible**

2. **In domains where there are many close-by sources (e.g. microscopy) super resolution is impossible**

Thus it is simultaneously possible that:

(1) In domains where there are few close-by sources (e.g. astronomy) super resolution is possible

(2) In domains where there are many close-by sources (e.g. microscopy) super resolution is impossible

This is borne out empirically b/c there are successful heuristics for resolving double-stars in astronomy, but in microscopy new experimental techniques really were needed

Thus it is simultaneously possible that:

(1) **In domains where there are few close-by sources (e.g. astronomy) super resolution is possible**

(2) **In domains where there are many close-by sources (e.g. microscopy) super resolution is impossible**

This is borne out empirically b/c there are successful heuristics for resolving double-stars in astronomy, but in microscopy new experimental techniques really were needed, e.g.

2014 Nobel Prize in Chemistry!

Super-resolution through stimulated emission

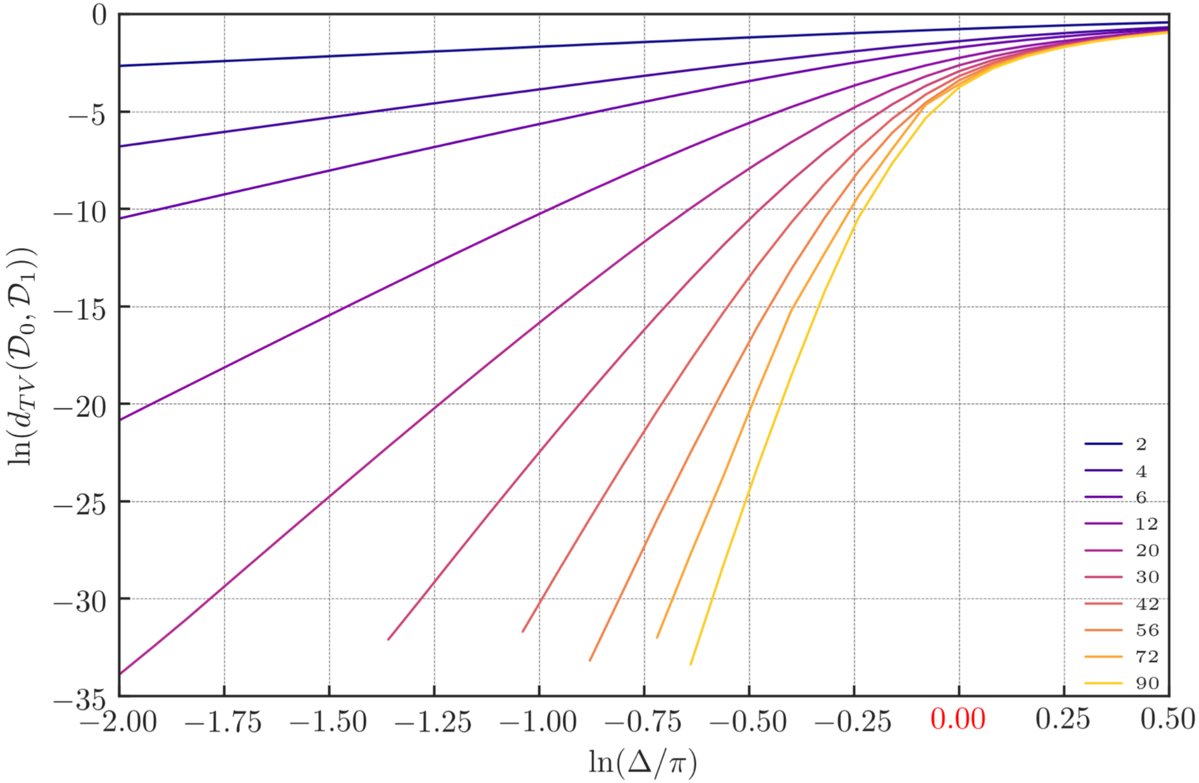
Eric Betzig, Stefan Hell, William Moerner

VISUALIZING THE DIFFRACTION LIMIT

In 1-D we can pinpoint the diffraction limit (it's the **Abbe limit**) and can visualize how resolution undergoes a phase transition

VISUALIZING THE DIFFRACTION LIMIT

In 1-D we can pinpoint the diffraction limit (it's the **Abbe limit**) and can visualize how resolution undergoes a phase transition



OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- **Deconvolution via the Fourier Transform**
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

DECONVOLUTION

There is a natural strategy for deconvolving by an Airy disk via the Fourier transform --- **division!**

DECONVOLUTION

There is a natural strategy for deconvolving by an Airy disk via the Fourier transform --- **division!**

Diffracted Image: $$\rho(x) = \sum_{j=1}^{k} \lambda_j I(x - \mu_j)$$

DECONVOLUTION

There is a natural strategy for deconvolving by an Airy disk via the Fourier transform --- **division!**

Diffracted Image: $$\rho(x) = \sum_{j=1}^{k} \lambda_j I(x - \mu_j)$$

Its Fourier Transform: $$\widehat{\rho}(\omega) = \sum_{j=1}^{k} \lambda_j \widehat{I}(\omega) e^{-2\pi i \langle \mu_j, \omega \rangle}$$

where $$\widehat{I}(\omega) = \frac{2}{\pi}\Big(\arccos(\pi\sigma\|\omega\|) - \pi\sigma\|\omega\|\sqrt{1 - \pi^2\sigma^2\|\omega\|^2}\Big)$$

DECONVOLUTION

There is a natural strategy for deconvolving by an Airy disk via the Fourier transform --- **division!**

Diffracted Image: $$\rho(x) = \sum_{j=1}^{k} \lambda_j I(x - \mu_j)$$

Its Fourier Transform: $$\widehat{\rho}(\omega) = \sum_{j=1}^{k} \lambda_j \widehat{I}(\omega) e^{-2\pi i \langle \mu_j, \omega \rangle}$$

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

Fact: $\widehat{I}(\omega)$ is nonzero on the disk of radius $\dfrac{1}{\pi\sigma}$ centered at zero

DECONVOLUTION

There is a natural strategy for deconvolving by an Airy disk via the Fourier transform --- **division!**

Diffracted Image: $$\rho(x) = \sum_{j=1}^{k} \lambda_j I(x - \mu_j)$$

Its Fourier Transform: $$\widehat{\rho}(\omega) = \sum_{j=1}^{k} \lambda_j \widehat{I}(\omega) e^{-2\pi i \langle \mu_j, \omega \rangle}$$

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

Now can we **remove** the $\widehat{I}(\omega)$ term, at least in the region where it is nonzero?

Lemma: For any $\|\omega\| < \frac{1}{\pi\sigma}$ we can simulate noisy access to the exponential sum

$$f(\omega) = \sum_{j=1}^{k} \lambda_j e^{-2\pi i \langle \mu_j, \omega \rangle}$$

Lemma: For any $\|\omega\| < \frac{1}{\pi\sigma}$ we can simulate noisy access to the exponential sum

$$f(\omega) = \sum_{j=1}^{k} \lambda_j e^{-2\pi i \langle \mu_j, \omega \rangle}$$

This is achieved via a simple procedure:

- Draw samples to construct an empirical estimate of $\rho(x)$
- Form a Kernel Density Estimate (i.e. smooth by convolving with a small variance Gaussian)
- Take the Fourier transform and pointwise divide by $\widehat{I}(\omega)$

Lemma: For any $\|\omega\| < \frac{1}{\pi\sigma}$ we can simulate noisy access to the exponential sum

$$f(\omega) = \sum_{j=1}^{k} \lambda_j e^{-2\pi i \langle \mu_j, \omega \rangle}$$

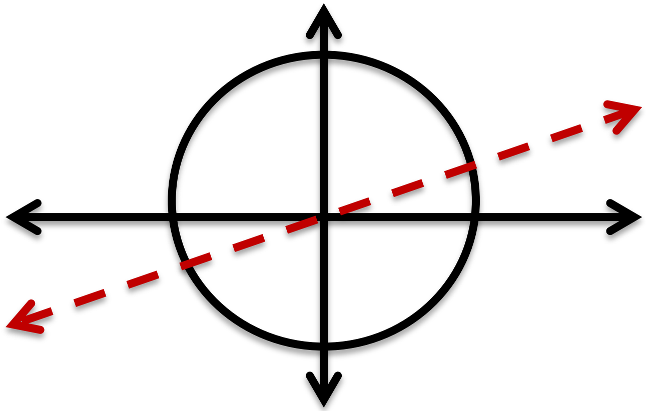
This is achieved via a simple procedure:

- Draw samples to construct an empirical estimate of $\rho(x)$
- Form a Kernel Density Estimate (i.e. smooth by convolving with a small variance Gaussian)
- Take the Fourier transform and pointwise divide by $\widehat{I}(\omega)$

Now, can we estimate the centers from the exponential sum?

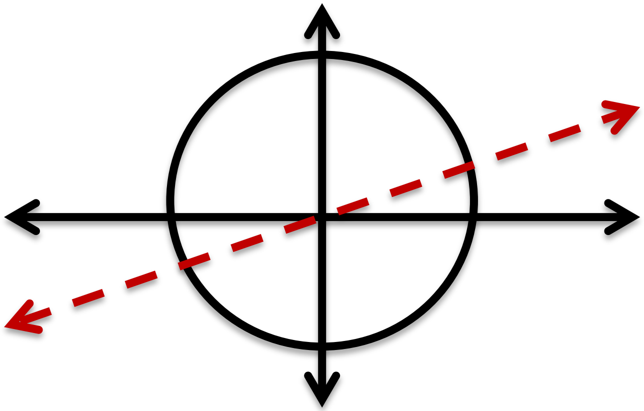
REDUCING TO ONE DIMENSION

What if we only query $f(\omega)$ on a line?

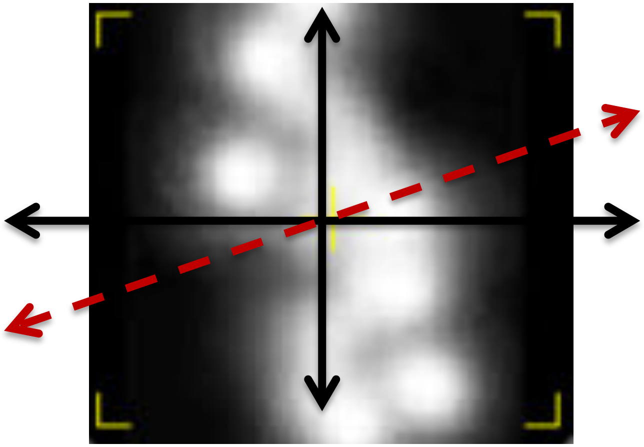


REDUCING TO ONE DIMENSION

What if we only query $f(\omega)$ on a line?

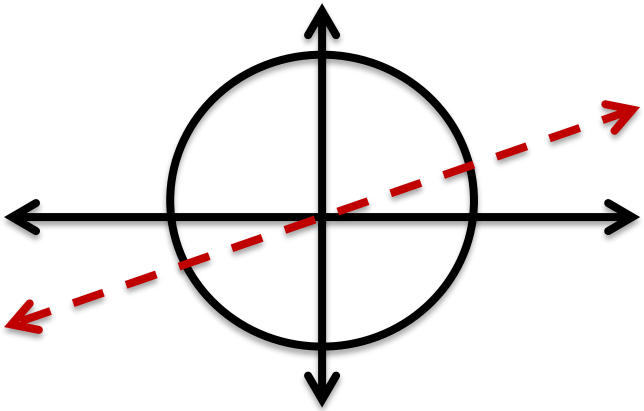


Equivalent to **projecting** the Airy disks onto a line

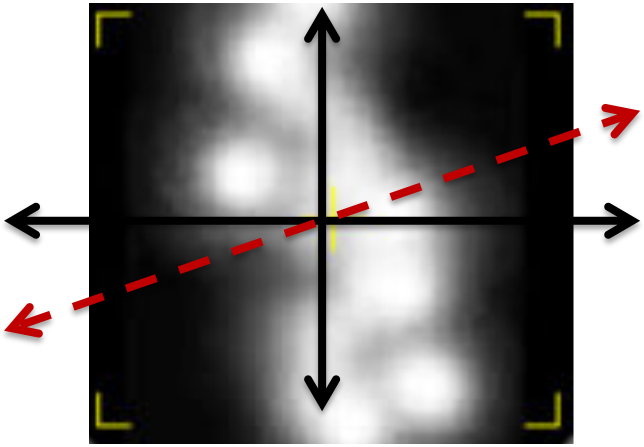


REDUCING TO ONE DIMENSION

What if we only query $f(\omega)$ on a line?



Equivalent to **projecting** the Airy disks onto a line



This could decrease the separation and make resolution harder, but let's figure out what our queries look like, mathematically

Fact: Suppose we sample $f(\omega)$ at the sequence of points on a line

$$a, a+b, a+2b, \cdots$$

Fact: Suppose we sample $f(\omega)$ at the sequence of points on a line

$$a, a+b, a+2b, \cdots$$

Then our vector of measurements can be expressed as

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_k \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_k^2 \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

Fact: Suppose we sample $f(\omega)$ at the sequence of points on a line

$$a, a+b, a+2b, \cdots$$

Then our vector of measurements can be expressed as

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_k \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_k^2 \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

where $\alpha_j = e^{-2\pi i \langle \mu_j, b \rangle}$ and $c_j = \lambda_j e^{-2\pi i a}$

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- **The Matrix Pencil Method**
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

THE MATRIX PENCIL METHOD

There is an algorithm called the **Matrix Pencil Method** based on solving a generalized eigenvalue problem that works w/o noise

THE MATRIX PENCIL METHOD

There is an algorithm called the **Matrix Pencil Method** based on solving a generalized eigenvalue problem that works w/o noise

Theorem: It is possible to recover the parameters of a sum of k exponentials with 2k+1 noiseless measurements

THE MATRIX PENCIL METHOD

There is an algorithm called the **Matrix Pencil Method** based on solving a generalized eigenvalue problem that works w/o noise

Theorem: It is possible to recover the parameters of a sum of k exponentials with 2k+1 noiseless measurements

But in our setup noise is unavoidable

THE MATRIX PENCIL METHOD

There is an algorithm called the **Matrix Pencil Method** based on solving a generalized eigenvalue problem that works w/o noise

Theorem: It is possible to recover the parameters of a sum of k exponentials with 2k+1 noiseless measurements

But in our setup noise is unavoidable

Does the algorithm still work with noisy measurements?

THE MATRIX PENCIL METHOD

There is an algorithm called the **Matrix Pencil Method** based on solving a generalized eigenvalue problem that works w/o noise

Theorem: It is possible to recover the parameters of a sum of k exponentials with 2k+1 noiseless measurements

But in our setup noise is unavoidable

Does the algorithm still work with noisy measurements?

Lemma [Moitra '15]: The stability of the Matrix Pencil Method depends on the condition number of the Vandermonde matrix

Lemma: If we choose a line to restrict to at random and the measurements are finely spaced then

$$\sigma_{min} \geq \left(\frac{\Delta}{k}\right)^{ck^2}$$

with high probability

Lemma: If we choose a line to restrict to at random and the measurements are finely spaced then

$$\sigma_{min} \geq \left(\frac{\Delta}{k}\right)^{ck^2}$$

with high probability

The intuition is the Vandermonde matrix is merely exponentially ill-conditioned (i.e. finding coefficients from querying a polynomial)

Lemma: If we choose a line to restrict to at random and the measurements are finely spaced then

$$\sigma_{min} \geq \left(\frac{\Delta}{k}\right)^{ck^2}$$

with high probability

The intuition is the Vandermonde matrix is merely exponentially ill-conditioned (i.e. finding coefficients from querying a polynomial)

Thus if we take an exponential in k number of samples, we can accurately determine the projected centers of the disks

Lemma: If we choose a line to restrict to at random and the measurements are finely spaced then

$$\sigma_{min} \geq \left(\frac{\Delta}{k}\right)^{ck^2}$$

with high probability

The intuition is the Vandermonde matrix is merely exponentially ill-conditioned (i.e. finding coefficients from querying a polynomial)

Thus if we take an exponential in k number of samples, we can accurately determine the projected centers of the disks

Then repeat for new lines and piece together the estimates

BETTER BOUNDS?

In 1-D the smallest singular value of the Vandermonde matrix is much larger when the projected centers are separated

BETTER BOUNDS?

In 1-D the smallest singular value of the Vandermonde matrix is much larger when the projected centers are separated

Theorem [Moitra '15]: If the cutoff frequency (in our case $\frac{1}{\pi\sigma}$) is at least $\frac{1}{\Delta} + 1$ then

$$\sigma_{min} \geq \left(\omega_{max} - 1 - \frac{1}{\Delta}\right)^{1/2}$$

If instead it is at most $\frac{1-\epsilon}{\Delta}$ then $\sigma_{min} \leq 2^{-\epsilon k}$

BETTER BOUNDS

In 1-D the smallest singular value of the Vandermonde matrix is much larger when the projected centers are separated

Theorem [Moitra '15]: If the cutoff frequency (in our case $\frac{1}{\pi\sigma}$) is at least $\frac{1}{\Delta} + 1$ then

$$\sigma_{min} \geq \left(\omega_{max} - 1 - \frac{1}{\Delta}\right)^{1/2}$$

If instead it is at most $\frac{1-\epsilon}{\Delta}$ then $\sigma_{min} \leq 2^{-\epsilon k}$

The approach was based on a 1-D extremal function from number theory called the **Beurling-Selberg Majorant** and is the best smooth approximation to the sgn function

BETTER BOUNDS?

In 1-D the smallest singular value of the Vandermonde matrix is much larger when the projected centers are separated

Theorem [Moitra '15]: If the cutoff frequency (in our case $\frac{1}{\pi\sigma}$) is at least $\frac{1}{\Delta} + 1$ then

$$\sigma_{min} \geq \left(\omega_{max} - 1 - \frac{1}{\Delta}\right)^{1/2}$$

If instead it is at most $\frac{1-\epsilon}{\Delta}$ then $\sigma_{min} \leq 2^{-\epsilon k}$

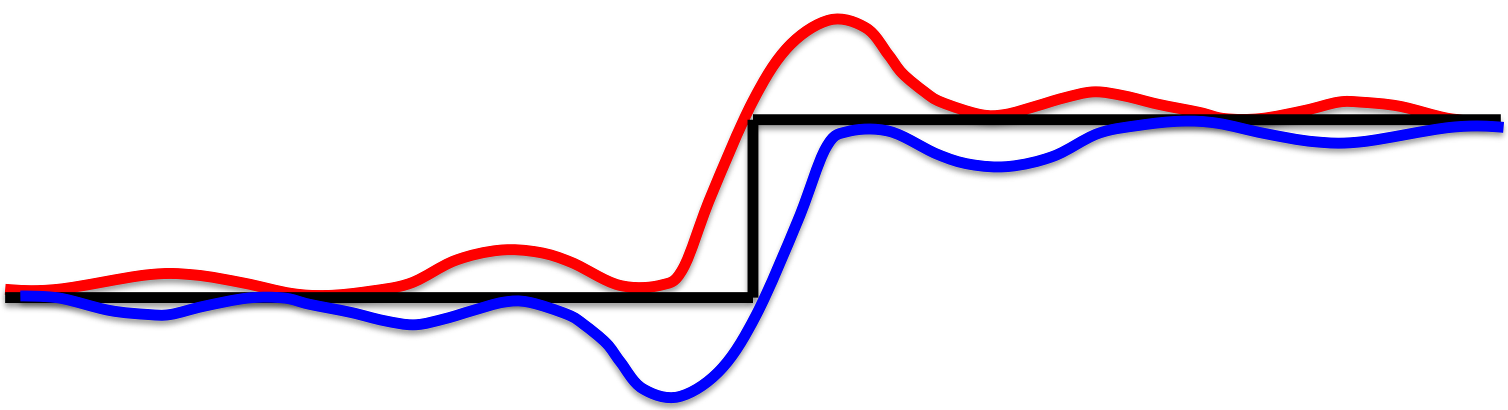
BETTER BOUNDS?

In 1-D the smallest singular value of the Vandermonde matrix is much larger when the projected centers are separated

Theorem [Moitra '15]: If the cutoff frequency (in our case $\frac{1}{\pi\sigma}$) is at least $\frac{1}{\Delta} + 1$ then

$$\sigma_{min} \geq \left(\omega_{max} - 1 - \frac{1}{\Delta}\right)^{1/2}$$

If instead it is at most $\frac{1-\epsilon}{\Delta}$ then $\sigma_{min} \leq 2^{-\epsilon k}$



BETTER BOUNDS?

So can we improve the dependence on k in 2-D when the centers are separated?

BETTER BOUNDS?

So can we improve the dependence on k in 2-D when the centers are separated?

Sometimes projection just doesn't work!

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- **Tackling the Two Dimensional Problem**

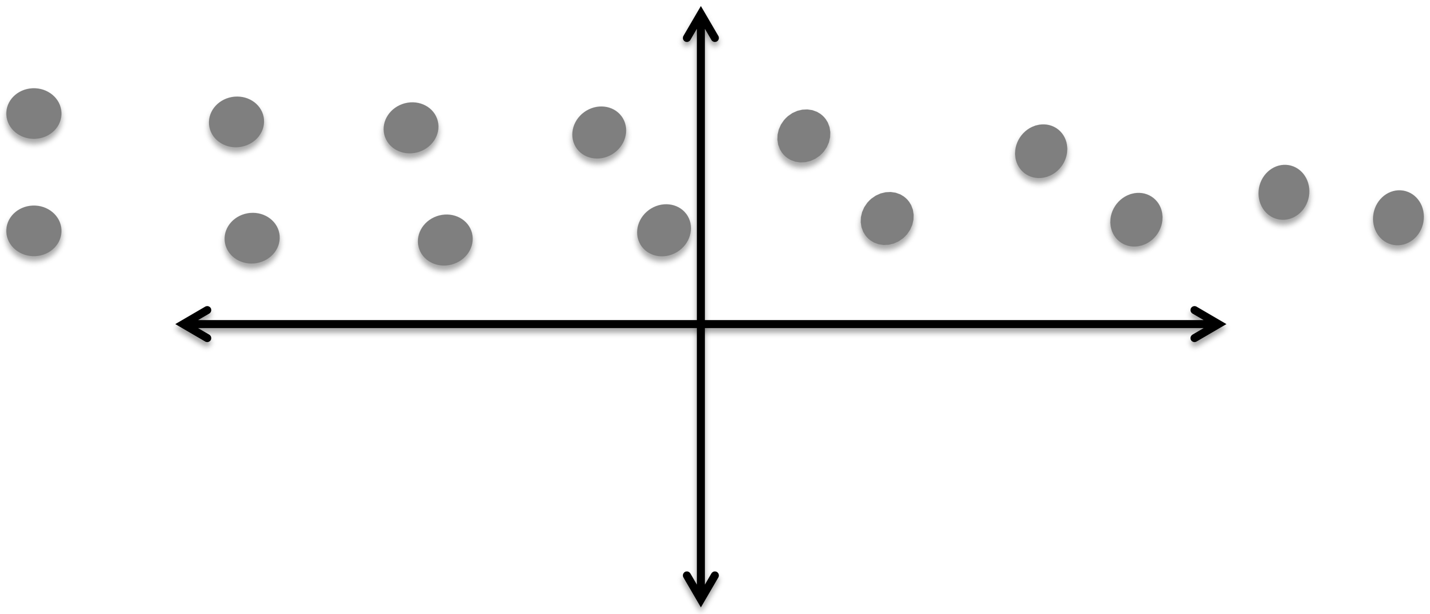
Part III: Connections to Mixtures of Gaussians

A NO-GO EXAMPLE

There are 2-D configurations where there is no 1-D projection that even approximately preserves the min separation

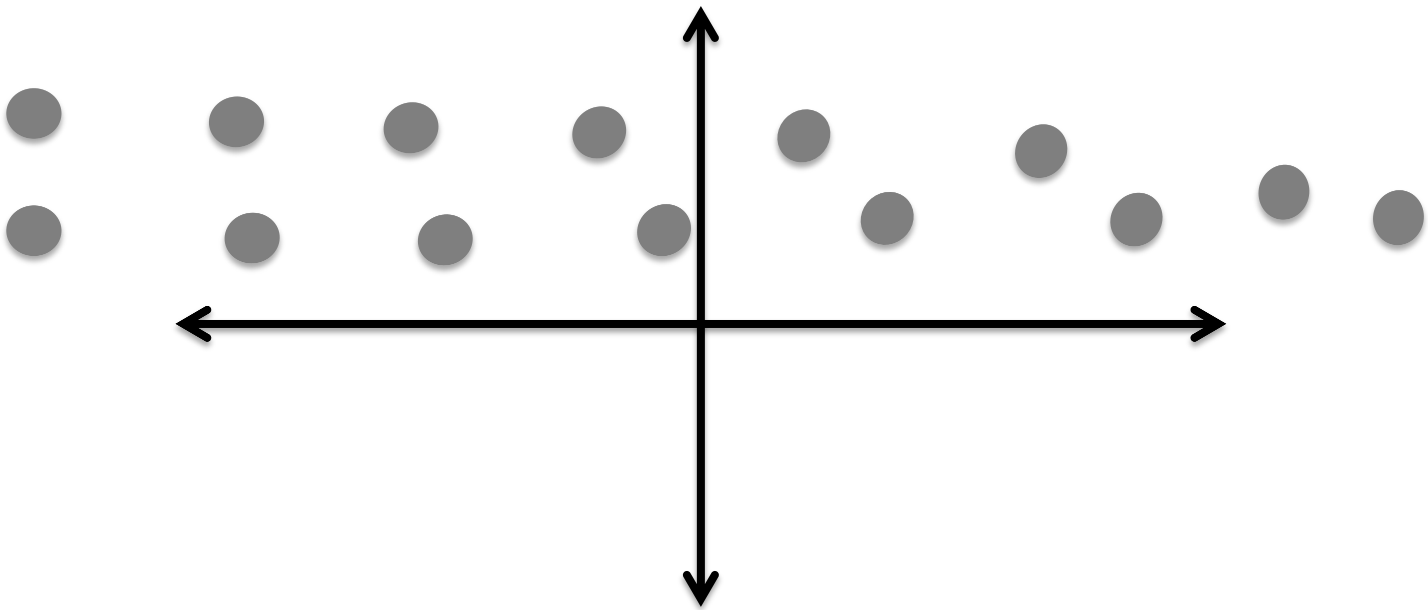
A NO-GO EXAMPLE

There are 2-D configurations where there is no 1-D projection that even approximately preserves the min separation



A NO-GO EXAMPLE

There are 2-D configurations where there is no 1-D projection that even approximately preserves the min separation



It's not just a failure of the technique! In fact the true threshold for 2-D problem is bounded away from that of the 1-D problem

MORE EXTREMAL FUNCTIONS

Instead we use very recent progress on 2-D extremal functions

MORE EXTREMAL FUNCTIONS

Instead we use very recent progress on 2-D extremal functions

Theorem [Gonclaves '18]: For all $\frac{2j_{0,1}}{\pi} < r < \frac{2j_{1,1}}{\pi}$ there is a function that satisfies...

MORE EXTREMAL FUNCTIONS

Instead we use very recent progress on 2-D extremal functions

Theorem [Gonclaves '18]: For all $\frac{2j_{0,1}}{\pi} < r < \frac{2j_{1,1}}{\pi}$ there is a function that satisfies

(1) $M(x) \leq 1_{B(0,1)}$ **i.e. it minorizes the unit ball**

MORE EXTREMAL FUNCTIONS

Instead we use very recent progress on 2-D extremal functions

Theorem [Gonclaves '18]: For all $\frac{2j_{0,1}}{\pi} < r < \frac{2j_{1,1}}{\pi}$ there is a function that satisfies

(1) $M(x) \leq 1_{B(0,1)}$ **i.e. it minorizes the unit ball**

(2) $\mathrm{supp}(\widehat{M}) \subseteq B(0,r)$ **i.e. it is smooth**

MORE EXTREMAL FUNCTIONS

Instead we use very recent progress on 2-D extremal functions

Theorem [Gonclaves '18]: For all $\frac{2j_{0,1}}{\pi} < r < \frac{2j_{1,1}}{\pi}$ there is a function that satisfies

(1) $M(x) \leq 1_{B(0,1)}$ **i.e. it minorizes the unit ball**

(2) $\mathrm{supp}(\widehat{M}) \subseteq B(0,r)$ **i.e. it is smooth**

(3) $\widehat{M}[0] > 0$ **i.e. it is a non-trivial approximation**

MORE EXTREMAL FUNCTIONS

Instead we use very recent progress on 2-D extremal functions

Theorem [Gonclaves '18]: For all $\frac{2j_{0,1}}{\pi} < r < \frac{2j_{1,1}}{\pi}$ there is a function that satisfies

(1) $M(x) \leq 1_{B(0,1)}$ **i.e. it minorizes the unit ball**

(2) $\text{supp}(\widehat{M}) \subseteq B(0,r)$ **i.e. it is smooth**

(3) $\widehat{M}[0] > 0$ **i.e. it is a non-trivial approximation**

These bounds arise through the study of de Branges spaces of entire functions

[Huang, Kakade '15]: Introduced a tensor method for recovering exponential sums whose analysis depends

$$\begin{bmatrix} e^{-2\pi i\langle \mu_1,\omega_1\rangle} & e^{-2\pi i\langle \mu_1,\omega_2\rangle} & \cdots & e^{-2\pi i\langle \mu_1,\omega_m\rangle} \\ e^{-2\pi i\langle \mu_2,\omega_1\rangle} & e^{-2\pi i\langle \mu_2,\omega_2\rangle} & & e^{-2\pi i\langle \mu_2,\omega_m\rangle} \\ \vdots & & \ddots & \vdots \\ e^{-2\pi i\langle \mu_k,\omega_1\rangle} & e^{-2\pi i\langle \mu_k,\omega_2\rangle} & & e^{-2\pi i\langle \mu_k,\omega_m\rangle} \end{bmatrix}$$

and its condition number

[Huang, Kakade '15]: Introduced a tensor method for recovering exponential sums whose analysis depends

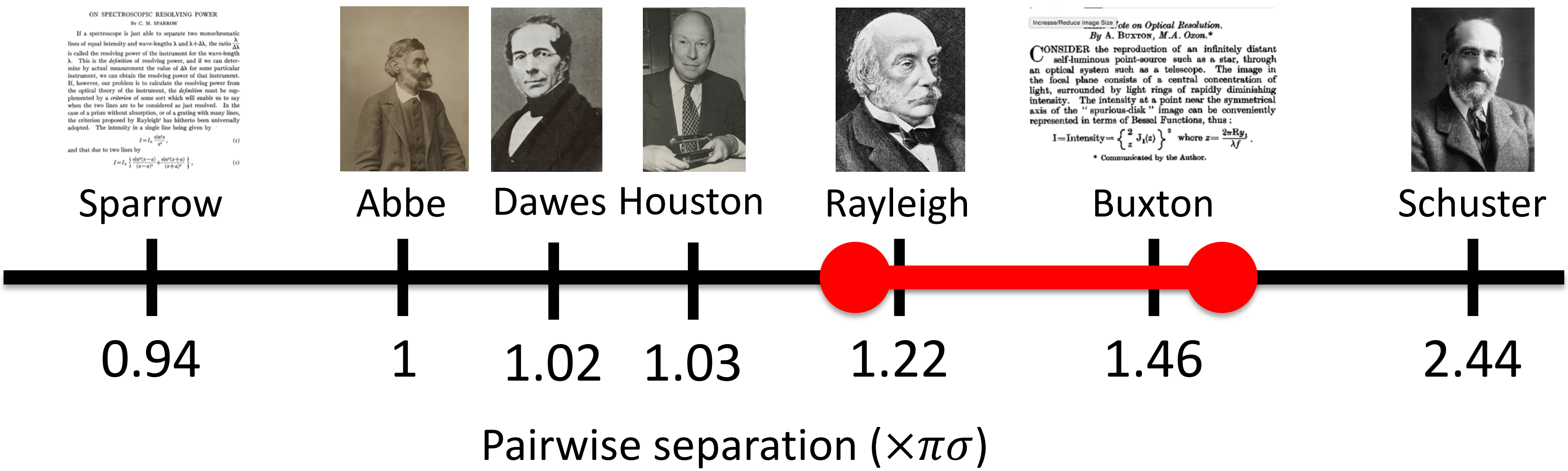
$$\begin{bmatrix} e^{-2\pi i\langle \mu_1,\omega_1\rangle} & e^{-2\pi i\langle \mu_1,\omega_2\rangle} & \cdots & e^{-2\pi i\langle \mu_1,\omega_m\rangle} \\ e^{-2\pi i\langle \mu_2,\omega_1\rangle} & e^{-2\pi i\langle \mu_2,\omega_2\rangle} & & e^{-2\pi i\langle \mu_2,\omega_m\rangle} \\ \vdots & & \ddots & \vdots \\ e^{-2\pi i\langle \mu_k,\omega_1\rangle} & e^{-2\pi i\langle \mu_k,\omega_2\rangle} & & e^{-2\pi i\langle \mu_k,\omega_m\rangle} \end{bmatrix}$$

and its condition number

We can use the 2-D extremal functions to show that random ω_j's from the $B(0, \gamma_+)$ have bounded condition number whp

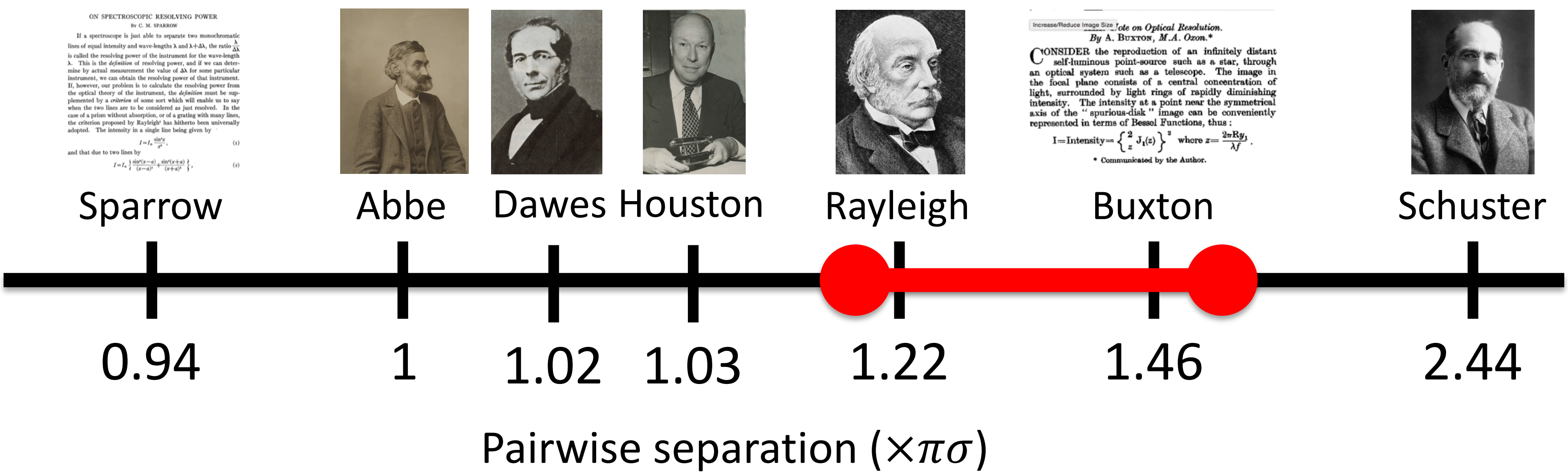
SUMMARY

With all due apologies to Carroll Sparrow's friends and colleagues:



SUMMARY

With all due apologies to Carroll Sparrow's friends and colleagues:



Finding the sharp diffraction limit remains a challenging problem in harmonic analysis

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

Part III: Connections to Mixtures of Gaussians

OUTLINE

Part I: Introduction

- The Diffraction Limit as an Inverse Problem
- The Lost Art of Debate
- Rigorous Foundations and Visualizations

Part II: Learning Mixtures of Airy Disks

- Deconvolution via the Fourier Transform
- The Matrix Pencil Method
- Tackling the Two Dimensional Problem

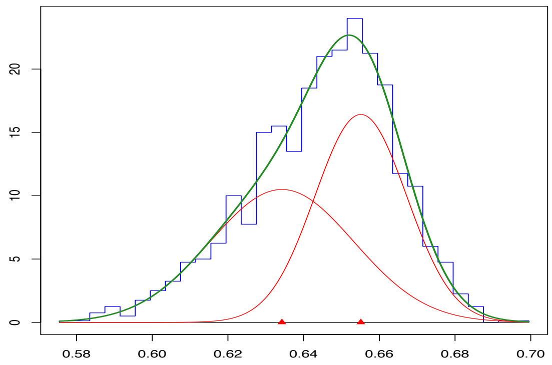
Part III: Connections to Mixtures of Gaussians

TAKING A STEP BACK

In theoretical machine learning we have provable algorithms for parameter learning, e.g. mixtures of gaussians

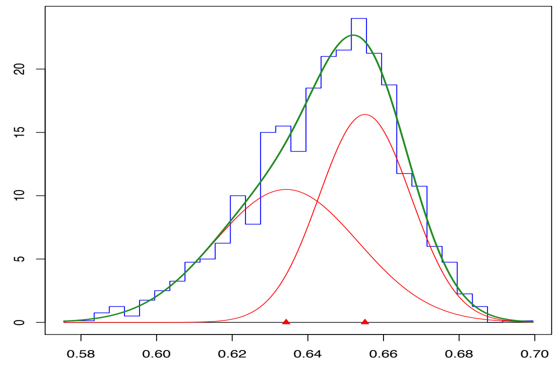
TAKING A STEP BACK

In theoretical machine learning we have provable algorithms for parameter learning, e.g. mixtures of gaussians



TAKING A STEP BACK

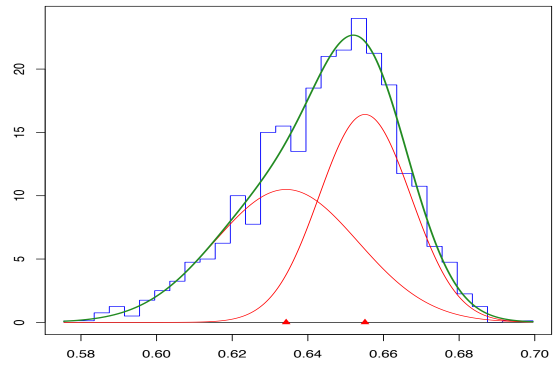
In theoretical machine learning we have provable algorithms for parameter learning, e.g. mixtures of gaussians



Equivalently we can solve inverse problems on the **heat equation with parametric assumptions**

TAKING A STEP BACK

In theoretical machine learning we have provable algorithms for parameter learning, e.g. mixtures of gaussians

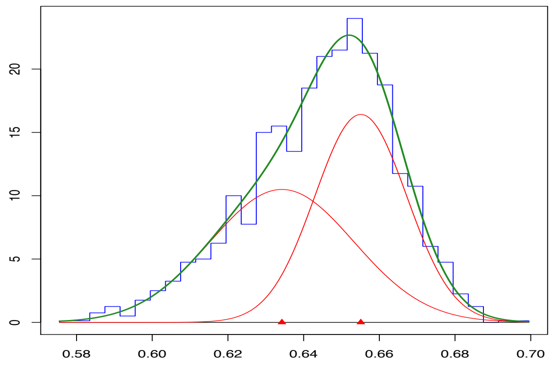


Equivalently we can solve inverse problems on the **heat equation with parametric assumptions**

In optics, resolution is an inverse problem for a different differential equation, but where many ideas can be adapted

TAKING A STEP BACK

In theoretical machine learning we have provable algorithms for parameter learning, e.g. mixtures of gaussians



Equivalently we can solve inverse problems on the **heat equation with parametric assumptions**

In optics, resolution is an inverse problem for a different differential equation, but where many ideas can be adapted

Do tools from theoretical machine learning have more to say about provable algorithms for inverse problems in science?

Summary:

- The diffraction limit is an inverse problem
- Fundamental limits in optics can be understood as a statistical phase transition
- **Are other inverse problems in the Sciences amenable to tools from theoretical ML?**

Summary:

- The diffraction limit is an inverse problem
- Fundamental limits in optics can be understood as a statistical phase transition
- **Are other inverse problems in the Sciences amenable to tools from theoretical ML?**

Thanks! Any Questions?