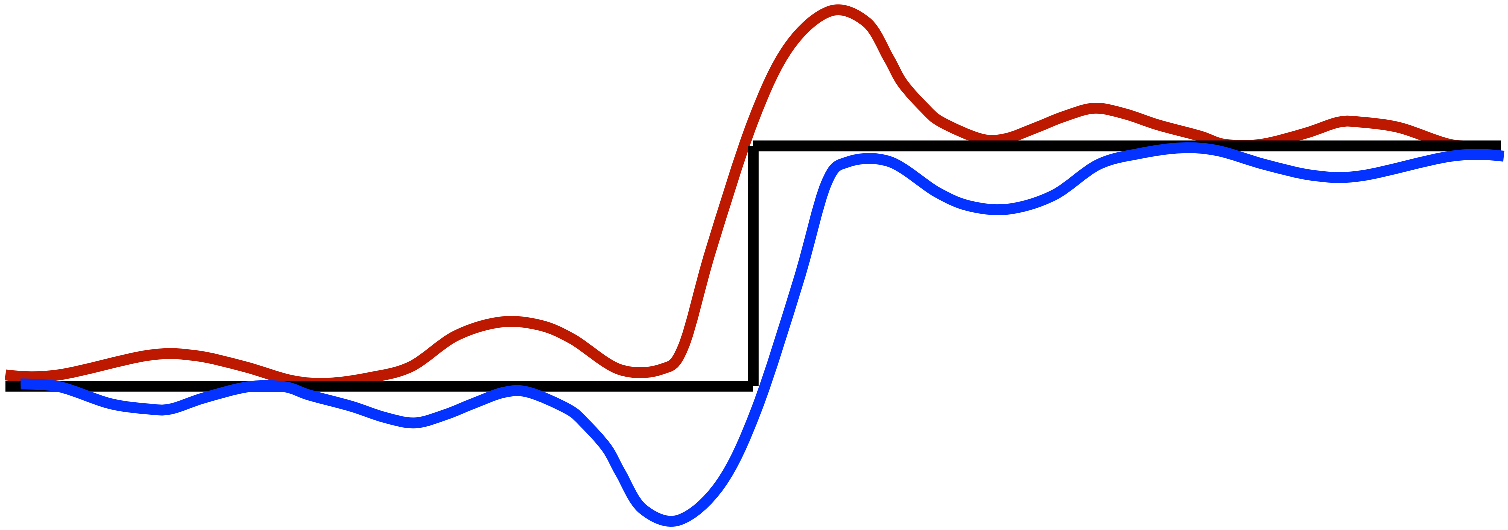
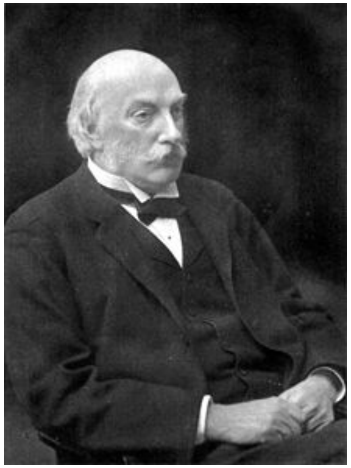
Super-resolution, Extremal Functions and the Condition Number of Vandermonde Matrices



Ankur Moitra

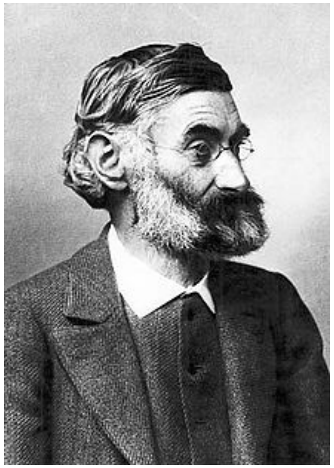
Massachusetts Institute of Technology

Limits to Resolution



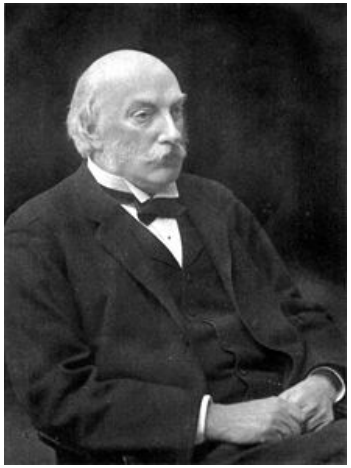
Lord Rayleigh
(1842-1919)

$$d = \frac{\lambda}{\underbrace{2n\sin\theta}_{\text{numerical aperture}}}$$



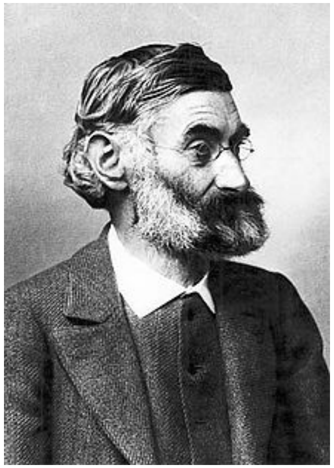
Ernst Abbe
(1840-1905)

Limits to Resolution



Lord Rayleigh
(1842-1919)

$$d = \underbrace{\frac{\lambda}{2n\sin\theta}}_{\text{numerical aperture}}$$



Ernst Abbe
(1840-1905)

In microscopy, it is difficult to observe sub-wavelength structures (**Rayleigh Criterion**, **Abbe Limit**, …)

Many devices are inherently **low-pass**

Many devices are inherently **low-pass**

Super-resolution: Can we recover **fine**-grained structure from **coarse**-grained measurements?

Many devices are inherently **low-pass**

Super-resolution: Can we recover **fine**-grained structure from **coarse**-grained measurements?

Applications in medical imaging, microscopy, astronomy, radar detection, geophysics, …

Many devices are inherently **low-pass**

Super-resolution: Can we recover **fine**-grained structure from **coarse**-grained measurements?

Applications in medical imaging, microscopy, astronomy, radar detection, geophysics, …

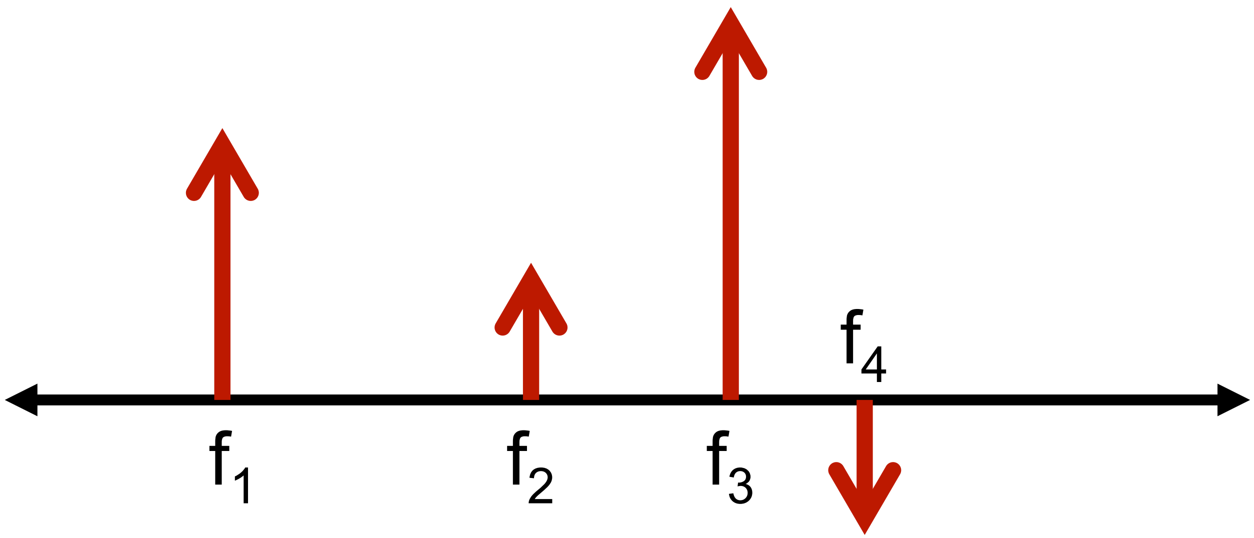
2014 Nobel Prize in Chemistry!

Super-resolution Cameras

Eric Betzig, Stefan Hell, William Moerner

A Mathematical Framework [Donoho, '91]:

Super-position of k spikes, each f_j in [0,1):



A Mathematical Framework [Donoho, '91]:

Super-position of k spikes, each f_j in [0,1):

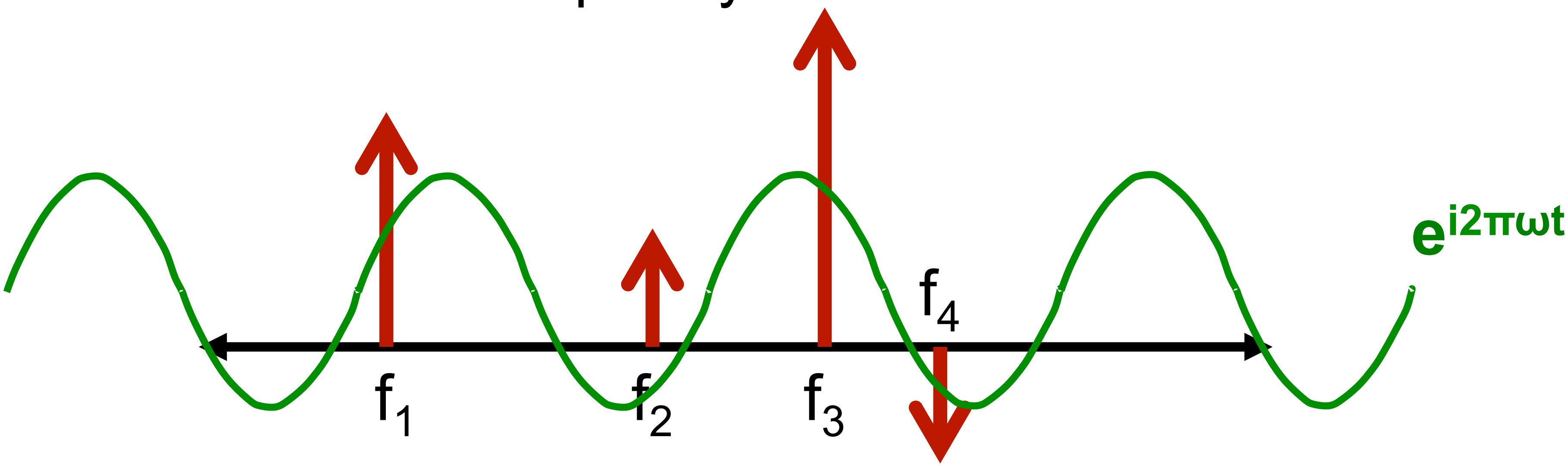
$$x(t) = \sum_{j=1}^{k} \underbrace{u_j}_{\text{coefficient}} \; \underbrace{\delta_{f_j}(t)}_{\text{delta function at } f_j}$$

A Mathematical Framework [Donoho, '91]:

Super-position of k spikes, each f_j in [0,1):

$$x(t) = \sum_{j=1}^{k} u_j \, \delta_{f_j}(t)$$

Measurement at frequency ω:



A Mathematical Framework [Donoho, '91]:

Super-position of k spikes, each f_j in [0,1):

$$x(t) = \sum_{j=1}^{k} u_j \, \delta_{f_j}(t)$$

Measurement at frequency ω:

$$v_\omega = \int_0^1 e^{i2\pi\omega t} \, x(t) \, dt$$

A Mathematical Framework **[Donoho, '91]:**

Super-position of k spikes, each f_j in [0,1):

$$x(t) = \sum_{j=1}^{k} u_j \, \delta_{f_j}(t)$$

Measurement at frequency ω:

$$v_\omega = \sum_{j=1}^{k} u_j \, e^{i2\pi f_j \omega}$$

A Mathematical Framework [Donoho, ‘91]:

Super-position of k spikes, each f_j in [0,1):

$$x(t) = \sum_{j=1}^{k} u_j \, \delta_{f_j}(t)$$

cut-off frequency

Measurement at frequency ω, $|\omega| \leq m$

$$v_\omega = \sum_{j=1}^{k} u_j \, e^{i2\pi f_j \omega}$$

A Mathematical Framework [Donoho, '91]:

Super-position of k spikes, each f_j in [0,1):

$$x(t) = \sum_{j=1}^{k} u_j \, \delta_{f_j}(t)$$

cut-off frequency

Measurement at frequency ω, $|\omega| \leq m$

noise

$$v_\omega = \sum_{j=1}^{k} u_j \, e^{i2\pi f_j \omega} + \eta_\omega$$

When can we recover the coeffs (u_j's) and locations (f_j's) from low frequency measurements?

When can we recover the coeffs (u_j's) and locations (f_j's) from low frequency measurements?

[Prony (1795), Pisarenko (1973), Matrix Pencil (1990),…]

Proposition 1: When there is no noise (η_ω=0), there is a polynomial time algorithm to recover the u_j's and f_j's exactly with $m = 2k + 1$ – i.e. measurements at

$$\omega = -k, -k+1, \ldots, k-1, k$$

When can we recover the coeffs (u_j's) and locations (f_j's) from low frequency measurements?

[Prony (1795), Pisarenko (1973), Matrix Pencil (1990),…]

Proposition 1: When there is no noise (η_ω=0), there is a polynomial time algorithm to recover the u_j's and f_j's exactly with m = 2k +1 – i.e. measurements at

$$\omega = -k, -k+1, \ldots, k-1, k$$

What is possible in the noise-free vs. the noisy setting will turn out to be **fundamentally** different…

What if there is noise?

What if there is noise?

Under what conditions is there an estimator

$$\hat{u}_j \longrightarrow u_j \quad \text{and} \quad \hat{f}_j \longrightarrow f_j$$

which converges at a polynomial-rate (in $|\eta_\omega|$)?

What if there is noise?

Under what conditions is there an estimator

$$\widehat{u}_j \longrightarrow u_j \quad \text{and} \quad \widehat{f}_j \longrightarrow f_j$$

which converges at a polynomial-rate (in $|\eta_\omega|$)?

And is there an algorithm?

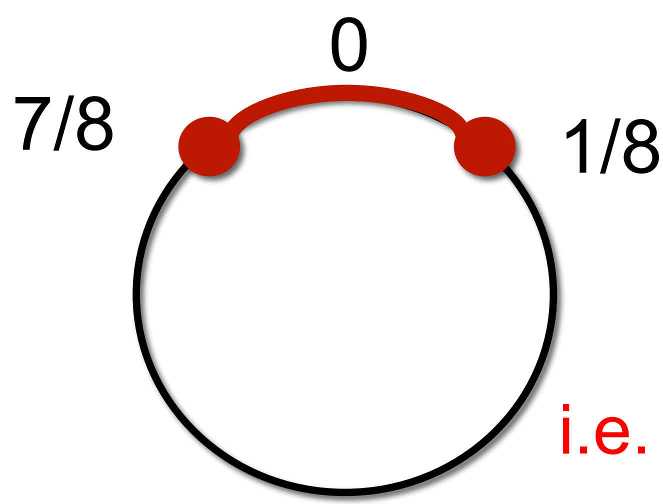
Theorem: There is a polynomial time algorithm for noisy super-resolution if $m > 1/\Delta+1$

separation condition

Theorem: There is a polynomial time algorithm for noisy super-resolution if $m > 1/\Delta+1$

separation condition

…where d_w is the "wrap-around" distance:

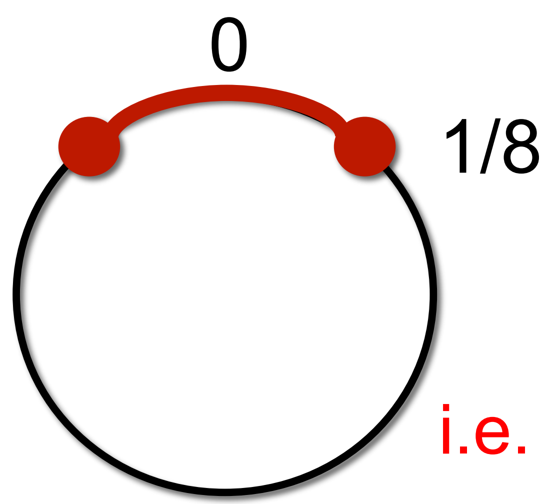


i.e. $d_w(1/8,7/8) = 1/4$

Theorem: There is a polynomial time algorithm for noisy super-resolution if $m > 1/\Delta+1$

separation condition

…where d_w is the "wrap-around" distance:



and $\Delta = \min_{i \neq j} d_w(f_i, f_j)$

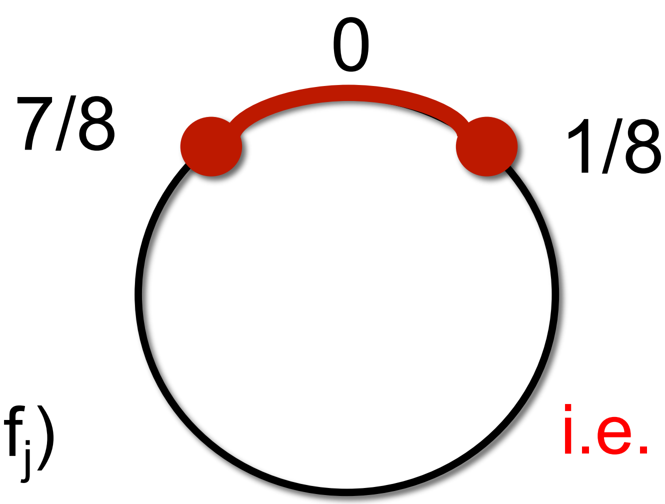
i.e. $d_w(1/8, 7/8) = 1/4$

Theorem: There is a polynomial time algorithm to recover estimates where

$$\min_{\text{matchings } \sigma} \max_{j} \left| \widehat{f}_{\sigma(j)} - f_j \right| + \left| \widehat{u}_{\sigma(j)} - u_j \right| \leq \varepsilon$$

provided $|\eta_\omega| \leq \text{poly}(\varepsilon, 1/m, 1/k)$, and $m > 1/\Delta + 1$

…where d_w is the "wrap-around" distance:



and $\Delta = \min_{i \neq j} d_w(f_i, f_j)$

i.e. $d_w(1/8, 7/8) = 1/4$

Theorem: For any $m \leq (1-\varepsilon)/\Delta$ and k, there is a pair of Δ-separated signals x and $\hat{x}$ where

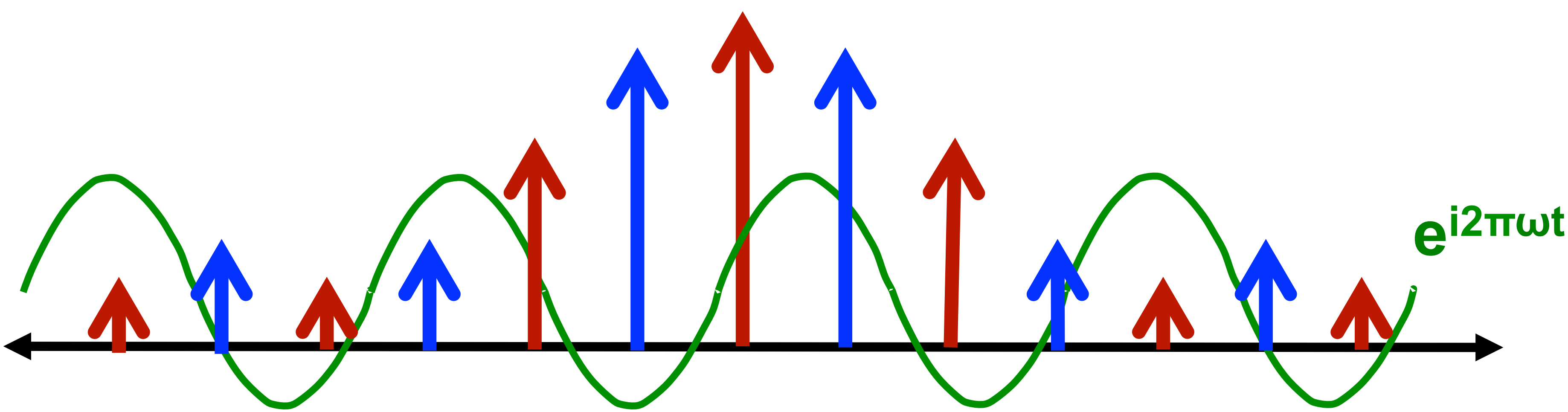
$$\left| \sum_{j=1}^{k} u_j\, e^{i2\pi f_j \omega} - \sum_{j=1}^{k} \hat{u}_j\, e^{i2\pi \hat{f}_j \omega} \right| \leq e^{-\varepsilon k}$$

for any $|\omega| \leq m$

Theorem: For any $m \leq (1-\varepsilon)/\Delta$ and k, there is a pair of Δ-separated signals x and $\hat{x}$ where

$$\left| \sum_{j=1}^{k} u_j\, e^{i2\pi f_j \omega} - \sum_{j=1}^{k} \hat{u}_j\, e^{i2\pi \hat{f}_j \omega} \right| \leq e^{-\varepsilon k}$$

for any $|\omega| \leq m$

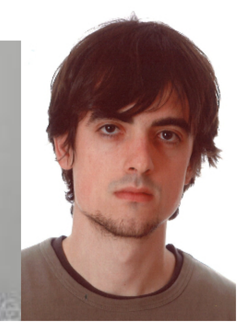


[Donoho, ’91]:

Asymptotic bounds
for $m = 1/\Delta$, on a grid

[Donoho, '91]:
Asymptotic bounds
for m = 1/Δ, on a grid

[Candes, Fernandez-Granda, '12]:
Convex program for m ≥ 2/Δ, no noise

[Donoho, '91]:
Asymptotic bounds
for m = 1/Δ, on a grid

[Candes, Fernandez-Granda, '12]:
Convex program for m ≥ 2/Δ, no noise

[Fernandez-Granda, '13]:
Convex program for m ≥ 2/Δ, with noise

[Donoho, ’91]:
Asymptotic bounds
for m = 1/Δ, on a grid

[Liao, Fannjiang, ’14]: (concurrent)
Algorithm for m = (1+C(Δ))/Δ, with noise

[Candes, Fernandez-Granda, ’12]:
Convex program for m ≥ 2/Δ, no noise

[Fernandez-Granda, ’13]:
Convex program for m ≥ 2/Δ, with noise

The Noise-free Case

Vandermonde Matrices

$$V_m^k = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & & \alpha_k \\ \alpha_1^2 & \alpha_2^2 & & \alpha_k^2 \\ \vdots & \vdots & & \vdots \\ \alpha_1^{m-1} & \alpha_2^{m-1} & \cdots & \alpha_k^{m-1} \end{bmatrix} \qquad \alpha_j \overset{\text{def}}{=} e^{i2\pi f_j}$$

Vandermonde Matrices

$$V_m^k = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & & \alpha_k \\ \alpha_1^2 & \alpha_2^2 & & \alpha_k^2 \\ \vdots & \vdots & & \vdots \\ \alpha_1^{m-1} & \alpha_2^{m-1} & \cdots & \alpha_k^{m-1} \end{bmatrix} \qquad \alpha_j \overset{\text{def}}{=} e^{i2\pi f_j}$$

This matrix plays a key role in many *exact* **inverse problems**

Vandermonde Matrices

$$V_m^k = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & & \alpha_k \\ \alpha_1^2 & \alpha_2^2 & & \alpha_k^2 \\ \vdots & \vdots & & \vdots \\ \alpha_1^{m-1} & \alpha_2^{m-1} & \cdots & \alpha_k^{m-1} \end{bmatrix} \qquad \alpha_j \overset{\text{def}}{=} e^{i2\pi f_j}$$

This matrix plays a key role in many *exact* **inverse problems**

e.g. polynomial interpolation, sparse recovery, inverse moment problems, …

Matrix Pencil Method

$$A = \boxed{V}\ \boxed{D_u}\ \boxed{V^H}$$

diagonal of u_j's

Matrix Pencil Method

$$A = \boxed{V}\ \boxed{D_u}\ \boxed{V^H}$$

diagonal of u_j's

Claim 1: The entries of A correspond to measurements with $-m+1 \leq \omega \leq m$

Matrix Pencil Method

$$A = \boxed{V}\ \boxed{D_u}\ \boxed{V^H} \qquad B = \boxed{V}\ \boxed{D_u}\ \boxed{D_\alpha}\ \boxed{V^H}$$

diagonal of u_j's

diagonal of α_j's

Claim 1: The entries of A correspond to measurements with $-m+1 \leq \omega \leq m$

Matrix Pencil Method

$$A = \boxed{V}\ \boxed{D_u}\ \boxed{V^H} \qquad B = \boxed{V}\ \boxed{D_u}\ \boxed{D_\alpha}\ \boxed{V^H}$$

diagonal of u_j's **diagonal of α_j's**

Claim 1: The entries of A and B correspond to measurements with $-m+1 \leq \omega \leq m$

Matrix Pencil Method

$$A = \boxed{V}\;\boxed{D_u}\;\boxed{V^H} \qquad B = \boxed{V}\;\boxed{D_u}\;\boxed{D_\alpha}\;\boxed{V^H}$$

diagonal of u_j's **diagonal of α_j's**

Claim 1: The entries of A and B correspond to measurements with $-m+1 \le \omega \le m$

Claim 2: If α_j's are distinct and $m \ge k$ and u_j's are non-zero, the unique solns to $Ax = \lambda Bx$ are $\lambda = 1/\alpha_j$

Noise Stability?

Robust Recovery?

Robust Recovery?

Fact: The Vandermonde has full (column) rank iff α_j's are distinct, and this is enough for **noise-free** recovery

Robust Recovery?

Fact: The Vandermonde has full (column) rank iff α_j's are distinct, and this is enough for **noise-free** recovery

When is the Matrix Pencil Method robust to noise?

Robust Recovery?

Fact: The Vandermonde has full (column) rank iff α_j's are distinct, and this is enough for **noise-free** recovery

When is the ~~Matrix Pencil Method~~ robust to noise?

Generalized Eigenvalue Problem

Robust Recovery?

Fact: The Vandermonde has full (column) rank iff α_j's are distinct, and this is enough for **noise-free** recovery

When is the ~~Matrix Pencil Method~~ robust to noise?

Generalized Eigenvalue Problem

[Stewart, Sun]: Various stability bounds for generalized eigenvalues/vectors based on the condition number

Robust Recovery?

Fact: The Vandermonde has full (column) rank iff α_j's are distinct, and this is enough for **noise-free** recovery

When is the ~~Matrix Pencil Method~~ robust to noise?

Generalized Eigenvalue Problem

[Stewart, Sun]: Various stability bounds for generalized eigenvalues/vectors based on the condition number

We show a sharp **phase-transition** for the condition number of the Vandermonde matrix

We use **extremal functions** to bound the condition number of the Vandermonde matrix

We use **extremal functions** to bound the condition number of the Vandermonde matrix

Such functions are used to prove sharp inequalities (on exponential sums) in **analytic number theory**

We use **extremal functions** to bound the condition number of the Vandermonde matrix

Such functions are used to prove sharp inequalities (on exponential sums) in **analytic number theory**

Theorem: $||V_m^k u||^2 = (m\text{-}1 \pm 1/\Delta)\ ||u||^2$

We use **extremal functions** to bound the condition number of the Vandermonde matrix

Such functions are used to prove sharp inequalities (on exponential sums) in **analytic number theory**

Theorem: $||V_m^k u||^2 = (m\text{-}1 \pm 1/\Delta)\ ||u||^2$

Moreover a direct construction based on the **Fejer kernel** shows this is tight…

We use **extremal functions** to bound the condition number of the Vandermonde matrix

Such functions are used to prove sharp inequalities (on exponential sums) in **analytic number theory**

Theorem: $||V_m^k u||^2 = (m\text{-}1 \pm 1/\Delta)\ ||u||^2$

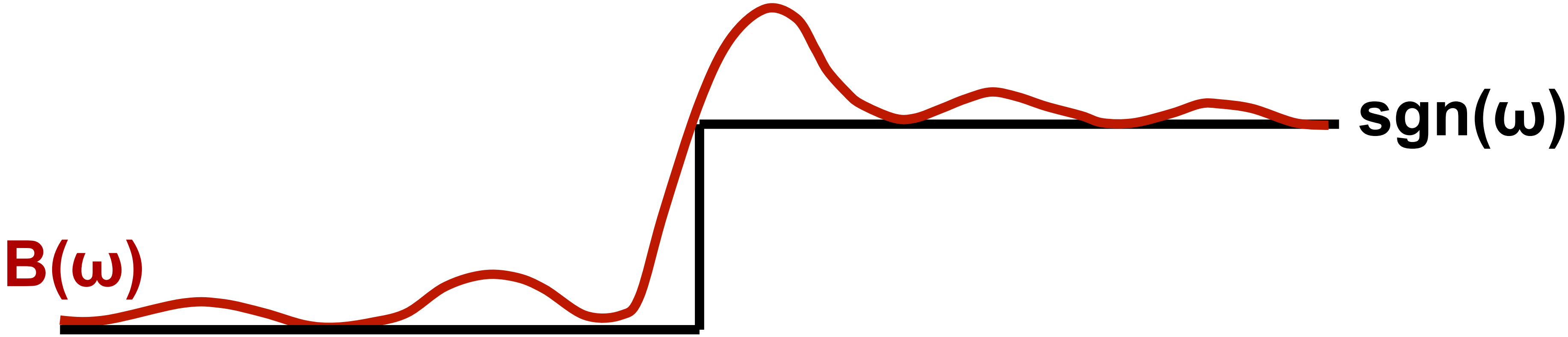
Moreover a direct construction based on the **Fejer kernel** shows this is tight…

Theorem: If $m = (1\text{-}\varepsilon)/\Delta$, there is a choice of α_j's, u_j's s.t.

$$||V_m^k u||^2 \leq e^{-\varepsilon k}\ ||u||^2$$

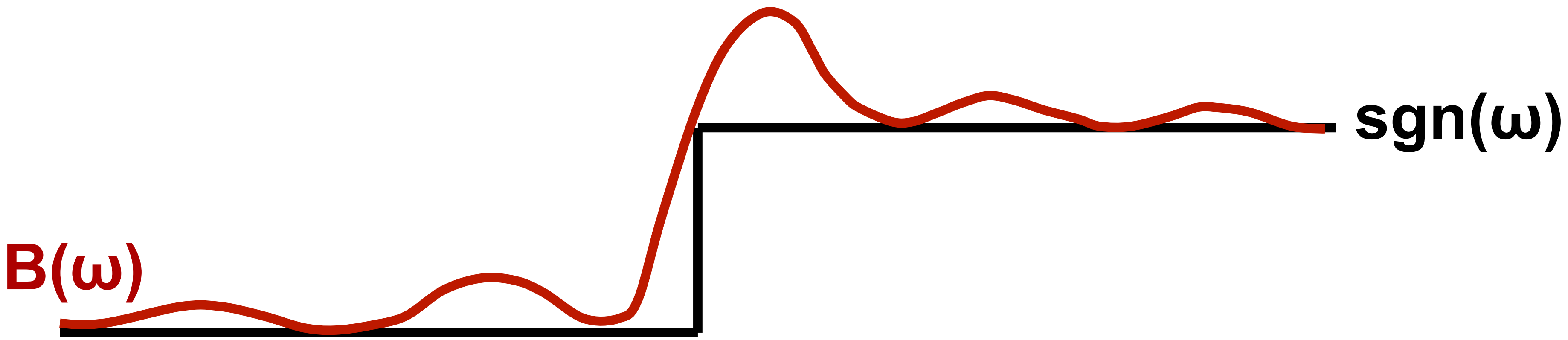
An Interlude

The **Beurling-Selberg majorant**:



An Interlude

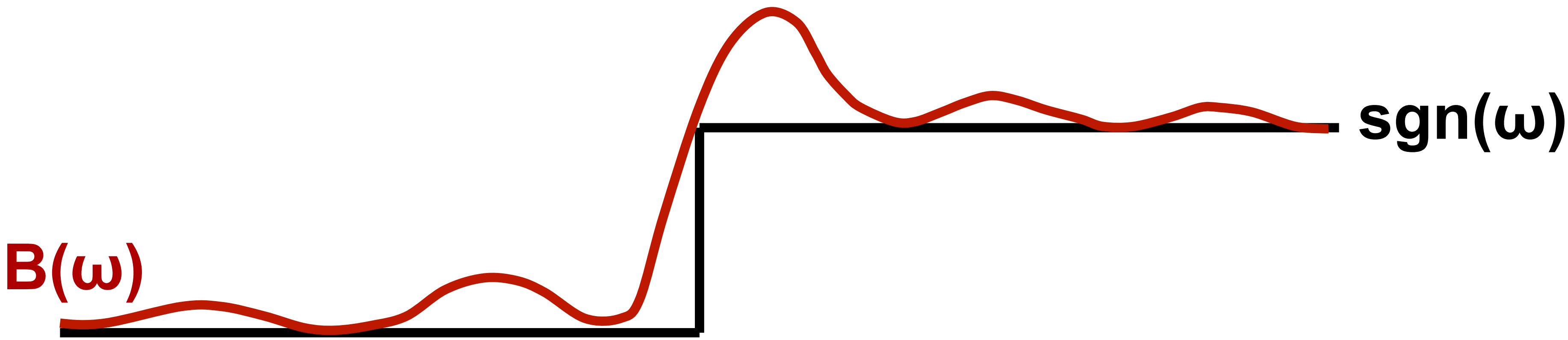
The **Beurling-Selberg majorant**:



Properties: (1) $\mathrm{sgn}(\omega) \leq B(\omega)$

An Interlude

The **Beurling-Selberg majorant**:



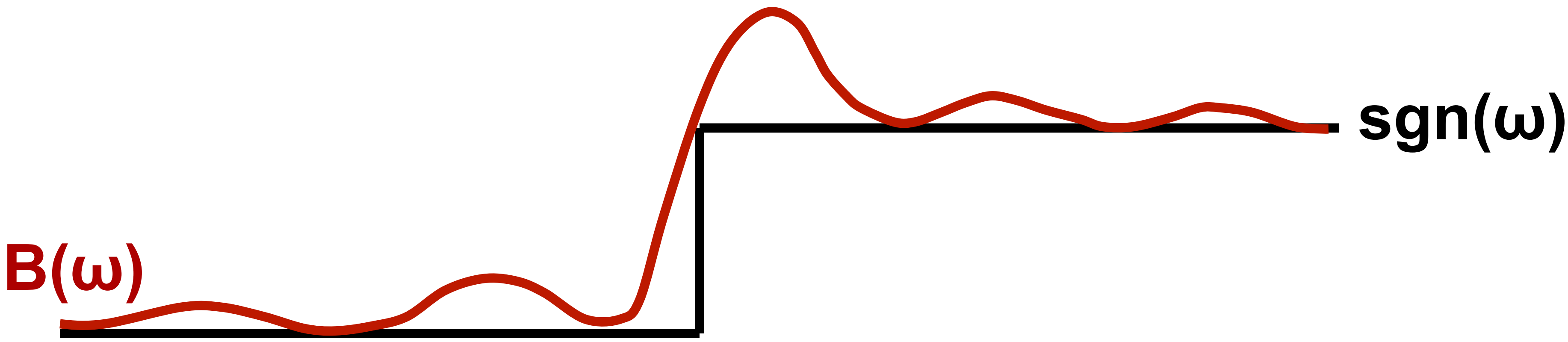
Properties:

(1) $\text{sgn}(\omega) \leq B(\omega)$

(2) $\widehat{B}(x)$ supported in $[-1,1]$

An Interlude

The **Beurling-Selberg majorant**:



Properties:

(1) $\mathrm{sgn}(\omega) \leq B(\omega)$

(2) $\hat{B}(x)$ supported in $[-1,1]$

(3) $\displaystyle\int_{-\infty}^{\infty} B(\omega) - \mathrm{sgn}(\omega)\, d\omega = 1$

An Interlude

The **Beurling-Selberg majorant**:

$$\left(\frac{\operatorname{sign}(\pi\omega)}{\pi}\right)^2 \left(\sum_{j=1}^{\infty} (\omega - j)^{-2} - \sum_{j=-\infty}^{-1} (\omega - j)^{-2} + \frac{2}{\omega} \right)$$

Properties:

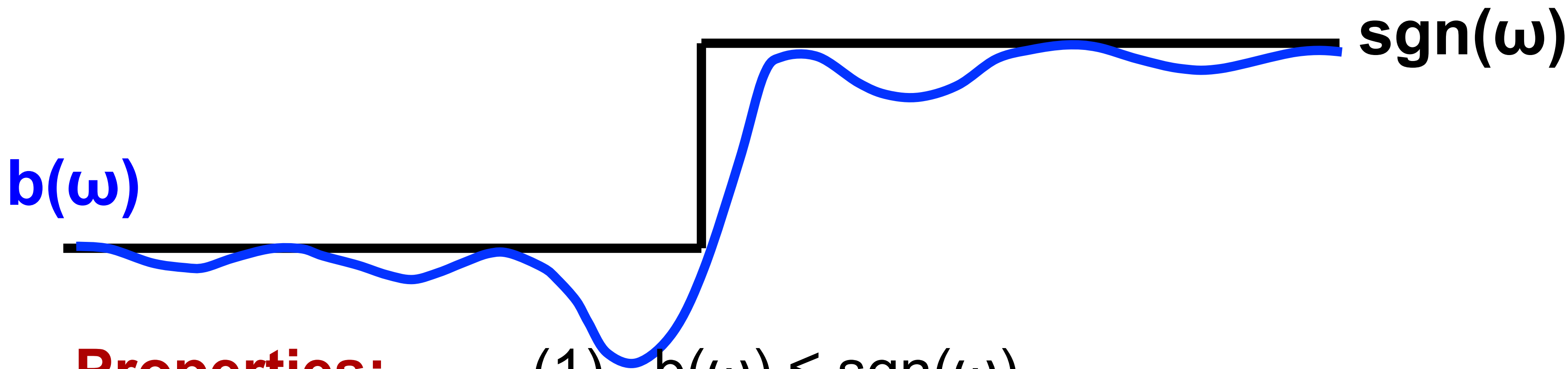
(1) $\operatorname{sgn}(\omega) \leq B(\omega)$

(2) $\hat{B}(x)$ supported in $[-1,1]$

(3) $\int_{-\infty}^{\infty} B(\omega) - \operatorname{sgn}(\omega)\, d\omega = 1$

An Interlude

The **Beurling-Selberg minorant**:



Properties:

(1) $b(\omega) \le \mathrm{sgn}(\omega)$

(2) $\hat{b}(x)$ supported in $[-1,1]$

(3) $\displaystyle\int_{-\infty}^{\infty} \mathrm{sgn}(\omega) - b(\omega)\, d\omega = 1$

Proof Omitted

Taking a Step Back

Many inverse problems are well-studied in the **exact** case

Taking a Step Back

Many inverse problems are well-studied in the **exact** case

When is the solution robust to noise?

Taking a Step Back

Many inverse problems are well-studied in the **exact** case

When is the solution robust to noise?

Example #1: Polynomial Interpolation

Lagrange Interpolation: Can recover a polynomial p(x) from its evaluations at deg(p)+1 points

Lagrange Interpolation: Can recover a polynomial p(x) from its evaluations at deg(p)+1 points

$$\underbrace{\begin{bmatrix} p_0 & \cdots & p_{m-1} \end{bmatrix}}_{\text{coeffs of poly. } p(x)}$$

Lagrange Interpolation: Can recover a polynomial p(x) from its evaluations at deg(p)+1 points

$$\underbrace{\begin{bmatrix} p_0 & \cdots & p_{m-1} \end{bmatrix}}_{\textbf{coeffs of poly. } \mathbf{p(x)}} \begin{bmatrix} 1 & \cdots & 1 \\ \alpha_1 & & \alpha_k \\ \alpha_1^2 & \cdots & \alpha_k^2 \\ \vdots & & \vdots \\ \alpha_1^{m-1} & \cdots & \alpha_k^{m-1} \end{bmatrix}$$

Lagrange Interpolation: Can recover a polynomial p(x) from its evaluations at deg(p)+1 points

$$\underbrace{\begin{bmatrix} p_0 & \cdots & p_{m-1} \end{bmatrix}}_{\textbf{coeffs of poly. p(x)}} \begin{bmatrix} 1 & \cdots & 1 \\ \alpha_1 & & \alpha_k \\ \alpha_1^2 & \cdots & \alpha_k^2 \\ \vdots & & \vdots \\ \alpha_1^{m-1} & \cdots & \alpha_k^{m-1} \end{bmatrix} = \underbrace{\begin{bmatrix} p(\alpha_1) & \cdots & p(\alpha_k) \end{bmatrix}}_{\textbf{evals of p(x)}}$$

Lagrange Interpolation: Can recover a polynomial p(x) from its evaluations at deg(p)+1 points

$$\underbrace{\begin{bmatrix} p_0 & \cdots & p_{m-1} \end{bmatrix}}_{\text{coeffs of poly. } p(x)} \begin{bmatrix} 1 & \cdots & 1 \\ \alpha_1 & & \alpha_k \\ \alpha_1^2 & \cdots & \alpha_k^2 \\ \vdots & & \vdots \\ \alpha_1^{m-1} & \cdots & \alpha_k^{m-1} \end{bmatrix} = \underbrace{\begin{bmatrix} p(\alpha_1) & \cdots & p(\alpha_k) \end{bmatrix}}_{\text{evals of } p(x)}$$

Often highly unstable

Lagrange Interpolation: Can recover a polynomial p(x) from its evaluations at deg(p)+1 points

$$\underbrace{\begin{bmatrix} p_0 & \cdots & p_{m-1} \end{bmatrix}}_{\text{coeffs of poly. } p(x)} \begin{bmatrix} 1 & \cdots & 1 \\ \alpha_1 & & \alpha_k \\ \alpha_1^2 & \cdots & \alpha_k^2 \\ \vdots & & \vdots \\ \alpha_1^{m-1} & \cdots & \alpha_k^{m-1} \end{bmatrix} = \underbrace{\begin{bmatrix} p(\alpha_1) & \cdots & p(\alpha_k) \end{bmatrix}}_{\text{evals of } p(x)}$$

Often highly unstable (**over the reals**), but not if the α_j's are complex roots of unity (**DFT matrix)**

Taking a Step Back

Many inverse problems are well-studied in the **exact** case

When is the solution robust to noise?

Example #1: Polynomial Interpolation

Taking a Step Back

Many inverse problems are well-studied in the **exact** case

When is the solution robust to noise?

Example #1: Polynomial Interpolation

Example #2: Sums of Exponentials (i.e. super-resolution)

Taking a Step Back

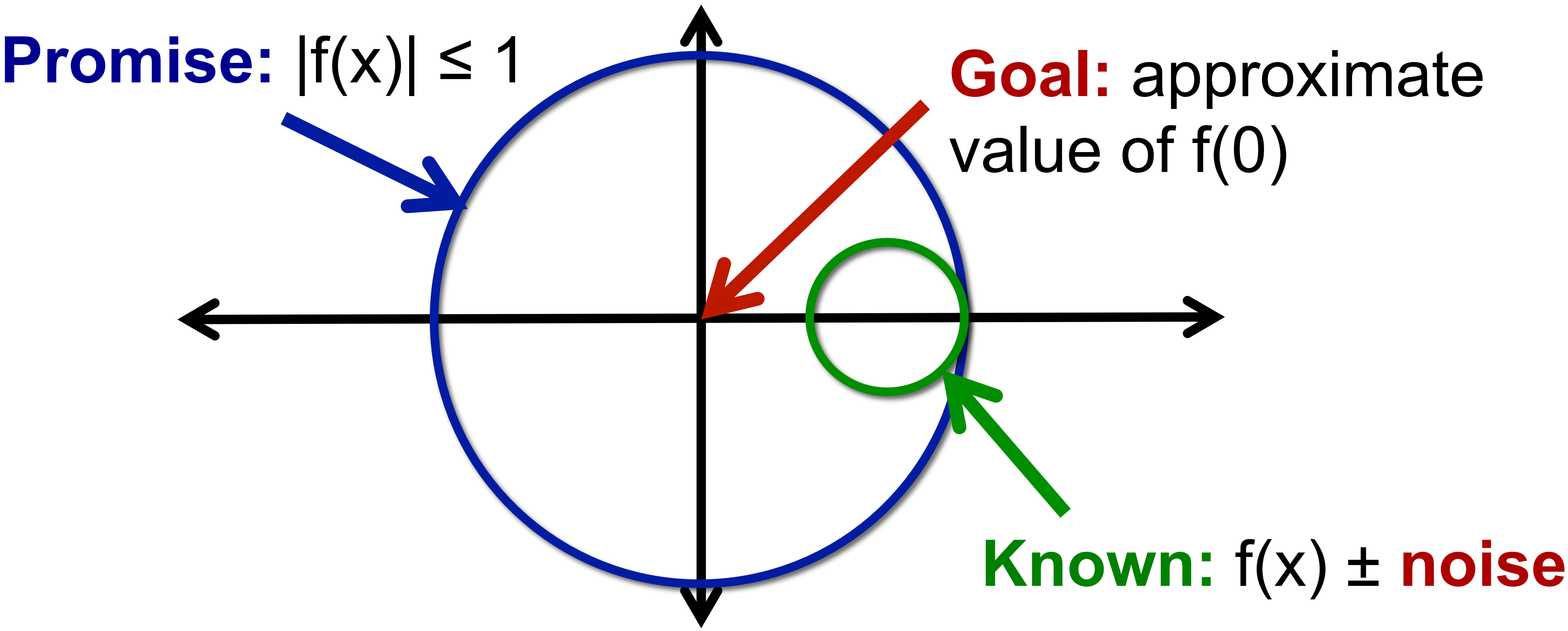
Many inverse problems are well-studied in the **exact** case

When is the solution robust to noise?

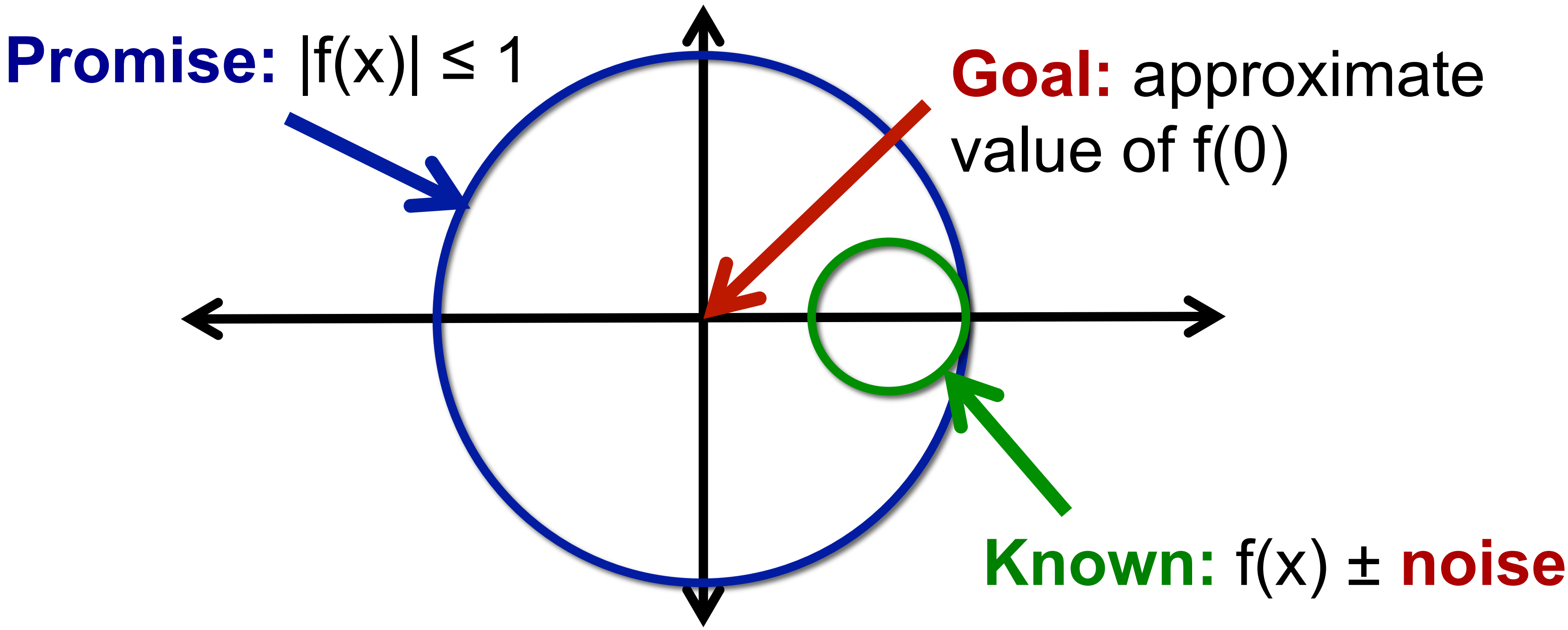
Example #1: Polynomial Interpolation

Example #2: Sums of Exponentials (i.e. super-resolution)

Example #3: Extrapolation with Boundary Conditions (lossy population recovery **[Moitra, Saks]**)



Promise: $|f(x)| \leq 1$
Goal: approximate value of $f(0)$
Known: $f(x) \pm$ **noise**



Hadamard Three Circle Theorem: Can extrapolate f(0) from evaluations on inner circle, if f is bounded on the outter circle

Theme: noisy inverse problems are better posed over the complex plane

Theme: noisy inverse problems are better posed over the complex plane

Are there other examples of this phenomenon?

Theme: noisy inverse problems are better posed over the complex plane

Are there other examples of this phenomenon?

We also give other connections between **test functions** in harmonic analysis and **preconditioners**

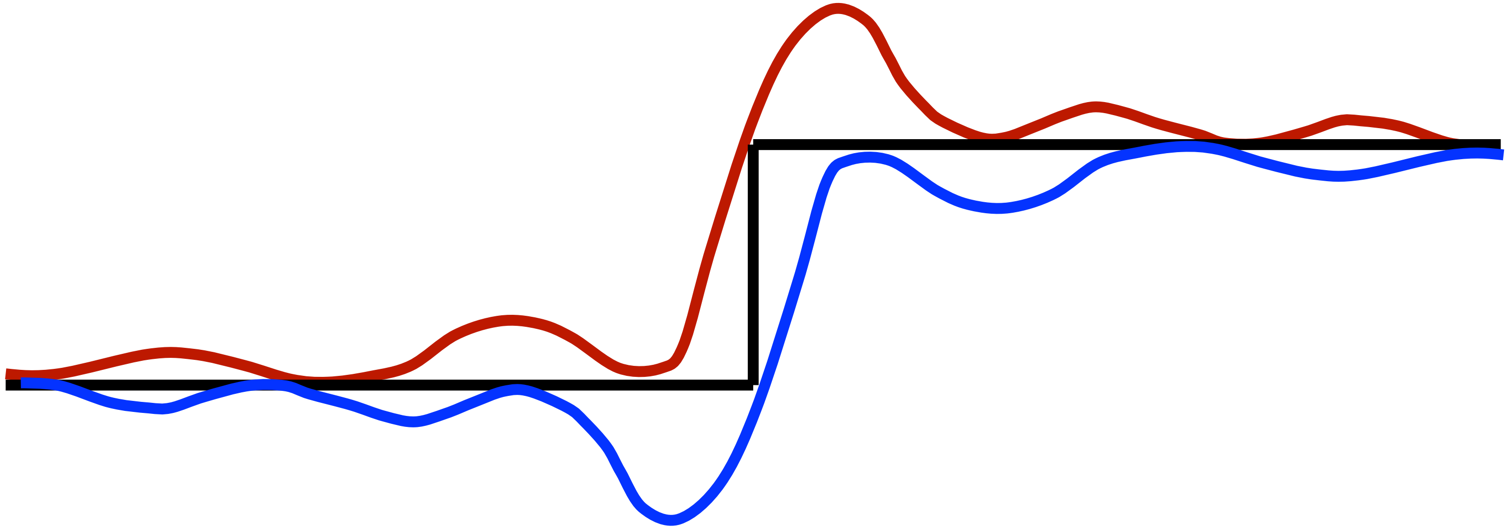
Theme: noisy inverse problems are better posed over the complex plane

Are there other examples of this phenomenon?

We also give other connections between **test functions** in harmonic analysis and **preconditioners**

These functions give a way to **obliviously** rescale rows of an unknown Vandermonde to make it nearly orthogonal

Thanks!



Any Questions?