CS 59300 – Algorithms for Data Science

Classical and Quantum approaches

Lecture 6 (09/18)

Super-resolution (I)

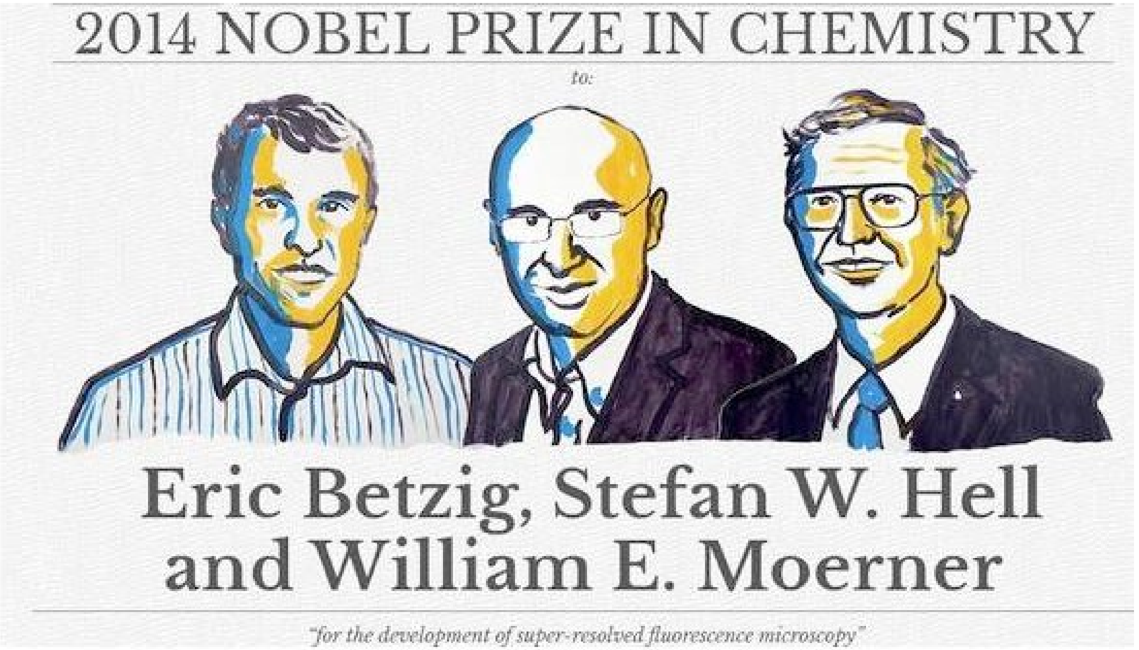
https://ruizhezhang.com/course_fall_2025.html

Limits to resolution

In optics, many devices are inherently low pass

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

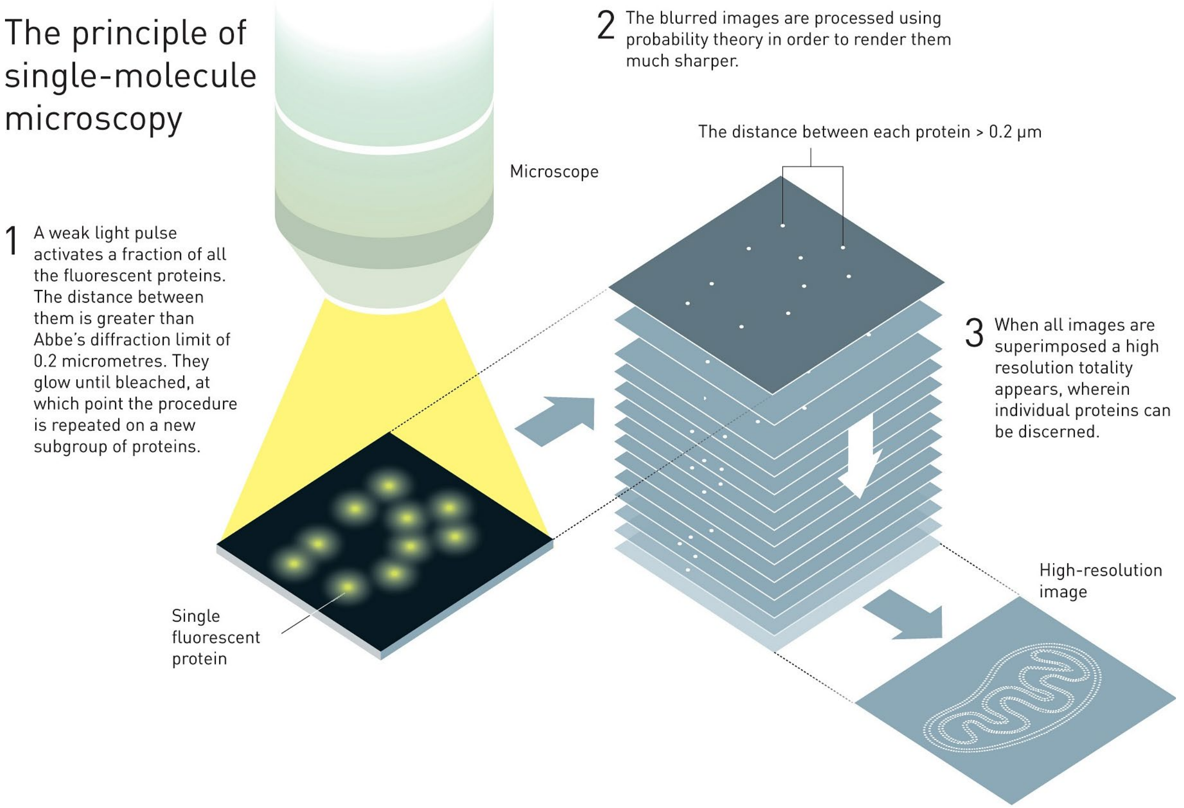
Applications in medical imaging, microscopy, astronomy, radar detection, geophysics, etc

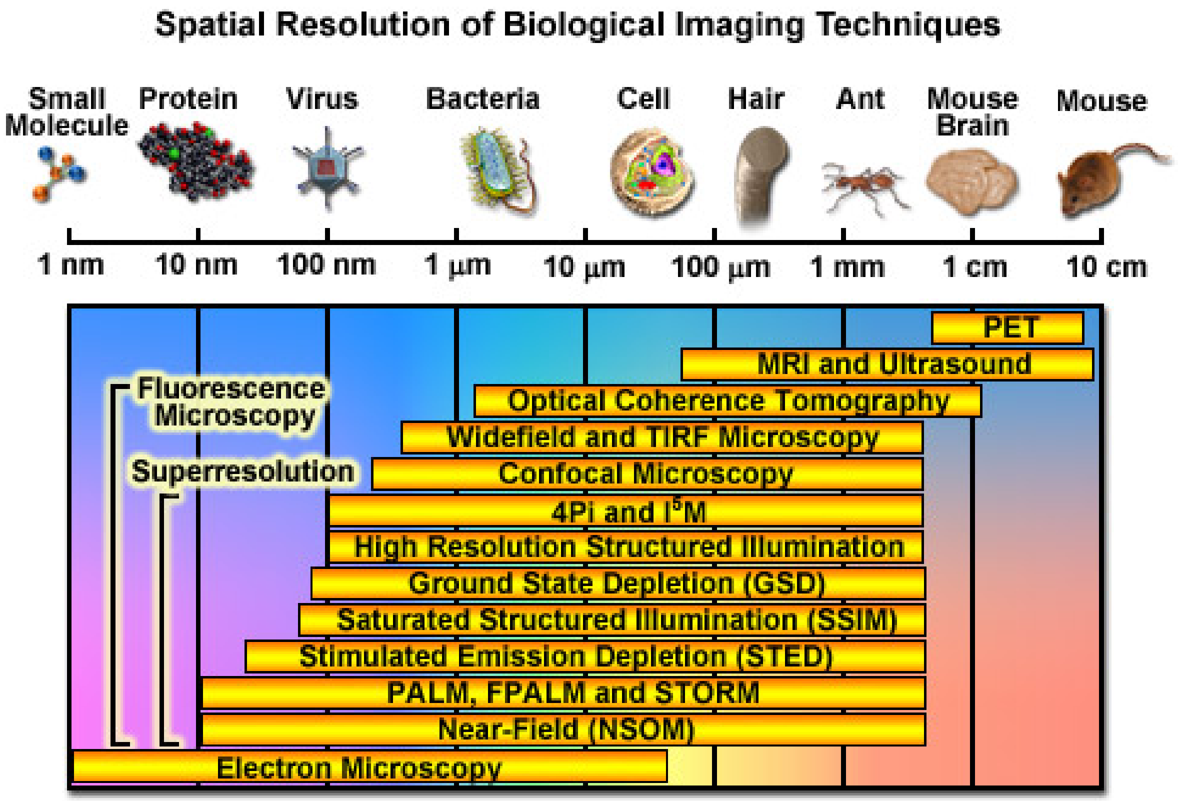


Super-resolution camera

$$\text{Abbe limit} = \frac{\lambda}{2\eta \sin(\alpha)} \approx 200 \text{ nm}$$

Super-resolution imaging

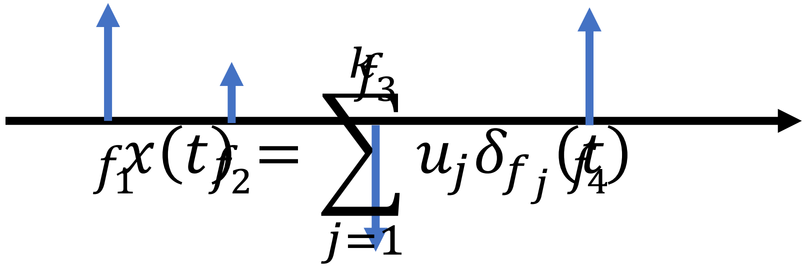




A mathematical framework

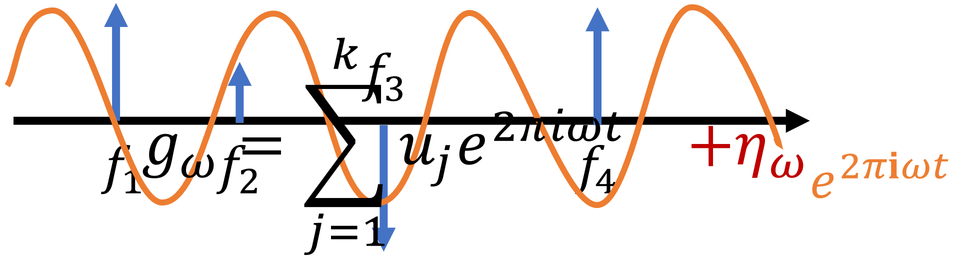
Introduced by (Donoho ’91)

- 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 low frequencies ω, up to cutoff $|\omega| \le n$



$$g_\omega = \sum_{j=1}^{k} u_j e^{2\pi i \omega t} + \eta_\omega$$

David Donoho

Can we recover the locations and coefficients?
With noise?

An ancient algorithm

Prony's method (Prony '1795):

Proposition. When there is no noise ($\eta_\omega = 0$), there is a polynomial time algorithm to recover the (u_j, f_j)'s exactly with $n = 2k + 1$, i.e., measurements at

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

Is it stable to noise?

Gaspard de Prony
(1755-1839)

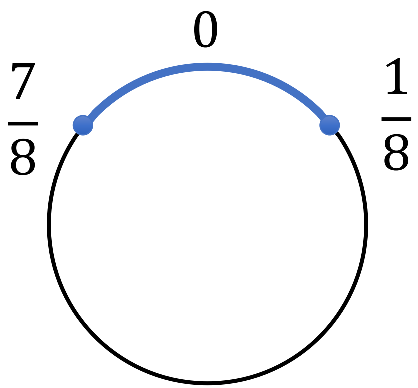
1d super-resolution: upper bound

Theorem (Moitra ’2015; Li-Liao-Fannjiang ’20).

There is a polynomial-time algorithm for super-resolution if $n \gtrsim 1/\Delta$, and otherwise it is statistically impossible.

separation condition

Wrap-around distance d_w:



$d_w(7/8\,, 1/8) = 1/4$

$$\boxed{\Delta := \min_{j \neq j' \in [k]} d_w\left(f_j, f_{j'}\right)}$$

1d super-resolution: upper bound

Theorem (Moitra '2015; Li-Liao-Fannjiang '20).

There is a polynomial-time algorithm to recover $\{(\hat{f}_j, \hat{u}_j)\}_{j\in[k]}$ such that

$$\min_{\pi\in\mathcal{S}_k} \max_{j\in[k]} d_w(\hat{f}_{\pi(j)}, f_j) + |\hat{u}_{\pi(j)} - u_j| \le \epsilon,$$

provided $|\eta_\omega| \le \mathrm{poly}(\epsilon, 1/n, 1/k)$ and $m > 1/\Delta + 1$

- The estimates converge to the ground-truth at an inverse polynomial rate, in terms of the magnitude of the noise

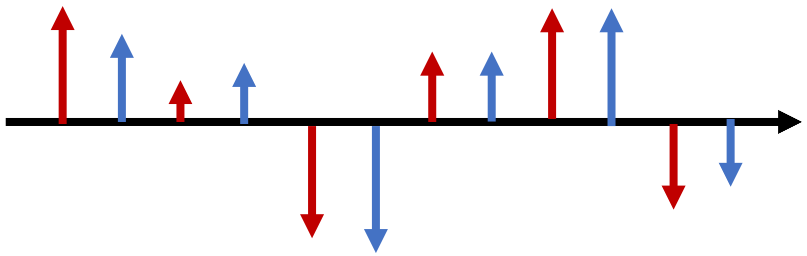
1d super-resolution: lower bound

Theorem (Moitra '15).

For any $m \le (1-\epsilon)/\Delta$, there is a pair of Δ-separated signals x and $\hat{x}$ such that

$$\left| \sum_{j=1}^{k} u_j e^{2\pi \mathbf{i} f_j \omega} - \sum_{j=1}^{k} \hat{u}_j e^{2\pi \mathbf{i} \hat{f}_j \omega} \right| \le 2^{-\epsilon k}$$

for any $|\omega| \le n$.



History

Donoho ’92:

Asymptotic bound for $n = 1/\Delta$ on-grid

Liao-Fannjiang ’14:

Algorithm for $n = \big(1 + o(\Delta)\big)/\Delta$ with noise

Moitra ’14:

Lower and upper bounds

Candes-Fernandez-Granda ’12:

Convex program for $n \geq 2/\Delta$ no noise

Fernandez-Granda ’13:

Convex program for $n \geq 2/\Delta$ with noise

The ESPRIT algorithm

Estimation of signal parameters via rotational invariance techniques

- It is one of the most effective spectral estimation method in practice.

-]

STE

ESPRIT—Estimation of Signal Parameters Via Rotational Invariance Techniques

RICHARD ROY AND THOMAS KAILATH, FELLOW, IEEE

***Abstract*—High-resolution signal parameter estimation is a problem of significance in many signal processing applications. Such applications include *direction-of-arrival* (DOA) estimation, system identification, and time series analysis. A novel approach to the general problem of signal parameter estimation is described. Although discussed in the context of direction-of-arrival estimation, ESPRIT can be applied to a wide variety of problems including accurate detection and estimation of sinusoids in noise. It exploits an underlying *rotational invariance* among signal subspaces induced by an array of sensors with a *translational invariance* structure. The technique, when applicable, manifests significant performance and computational advantages over previous algorithms such as MEM, Capon's MLM, and MUSIC.**

I. Introduction

In many practical signal processing problems, the objective is to estimate from measurements a set of *constant* parameters upon which the received signals depend. For example, high-resolution direction-of-arrival (DOA) estimation is important in many sensor systems such as radar, sonar, electronic surveillance, and seismic exploration. High-resolution frequency estimation is important in numerous applications, recent examples of which include the design and control of robots and large flexible space structures. In such problems, the functional form of the underlying signals can often be assumed to be known (e.g., narrow-band plane waves, cisoids). The quantities to be estimated are parameters (e.g., frequencies and DOA's of plane waves, cisoid frequencies) upon which the sensor outputs depend, and these parameters are assumed to be constant.[1]

There have been several approaches to such problems including the so-called maximum likelihood (ML) method of Capon (1969) and Burg's (1967) maximum entropy (ME) method. Although often successful and widely used, these methods have certain fundamental limitations (especially bias and sensitivity in parameter estimates), largely because they use an incorrect model (e.g., AR rather than special ARMA) of the measurements. Pisarenko (1973) was one of the first to exploit the structure of the data model, doing so in the context of estimation of parameters of cisoids in additive noise using a covariance approach. Schmidt (1977) and independently Bienvenu (1979) were the first to correctly exploit the measurement model in the case of sensor arrays of arbitrary form. Schmidt, in particular, accomplished this by first deriving a complete geometric solution in the absence of noise, then cleverly extending the geometric concepts to obtain a *reasonable* approximate solution in the presence of noise. The resulting algorithm was called MUSIC (MUltiple SIgnal Classification) and has been widely studied. In a detailed evaluation based on thousands of simulations, M.I.T.'s Lincoln Laboratory concluded that, among currently accepted high-resolution algorithms, MUSIC was the most promising and a leading candidate for further study and actual hardware implementation. However, although the performance advantages of MUSIC are substantial, they are achieved at a considerable cost in computation (searching over parameter space) and storage (of array calibration data).

In this paper, a new algorithm (ESPRIT) that dramatically reduces these computation and storage costs is presented. In the context of DOA estimation, the reductions are achieved by requiring that the sensor array possess a *displacement invariance*, i.e., sensors occur in matched pairs with identical displacement vectors. Fortunately, there are many practical problems in which these conditions are or can be satisfied. In addition to obtaining signal parameter estimates efficiently, *optimal* signal copy vectors for reconstructing the signals are elements of the ESPRIT solution as well. ESPRIT is also manifestly more robust (i.e., less sensitive) with respect to array imperfections than previous techniques including MUSIC [1].

To make the presentation as clear as possible, an attempt is made to adhere to a somewhat standard notational convention. Lowercase boldface italic characters will generally refer to vectors. Uppercase boldface italic characters will generally refer to matrices. For either real- or complex-valued matrices, $(\cdot)^*$ will be used to denote the Hermitian conjugate (or complex-conjugate transpose) operation. Eigenvalues of square Hermitian matrices are assumed to be ordered in decreasing magnitude, as are the

Manuscript received January 12, 1988; revised October 5, 1988. This work was supported in part by the Joint Services Program at Stanford University (U.S. Army, U.S. Navy, U.S. Air Force) under Contract DAAG29-85-K-0048, and the SDI/IST Program managed by the Office of Naval Research, under Contract N00014-85-K-0550.

The authors are with the Information Systems Laboratory, Stanford University, Stanford, CA 94305.

IEEE Log Number 8928125.

[1] Extensions to situations in which the parameters may be time varying can be made, however, they rely on an inherent *time-scale* or *eigenvalue separation* between the parameter dynamics and the dynamics of the signal process. Fundamentally, the assumption is made that over time intervals long enough to collect sufficient information from which to obtain accurate parameter estimates, the parameters have not changed significantly.

0096-3518/89/0700-0984$01.00 © 1989 IEEE

$\Longrightarrow$

ESPRIT-estimation of signal parameters via rotational invariance techniques

R Roy, T Kailath - IEEE Transactions on acoustics, speech, and signal …, 1989

An approach to the general problem of signal parameter estimation is described. The algorithm differs from its predecessor in that a total least-squares rather than a standard least-…

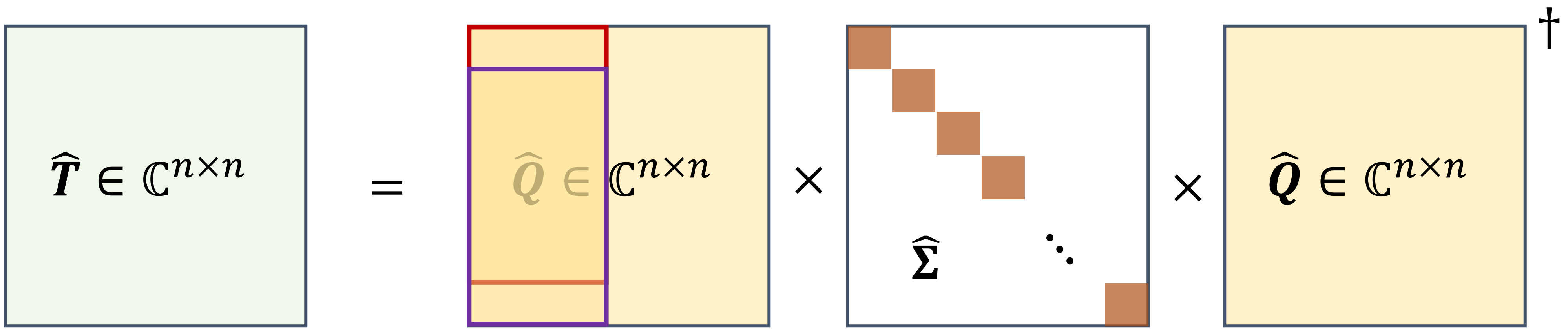
$g_{n-1} \quad g_{n-2} \quad g_{n-3} \quad g_0$

Toeplitz matrix

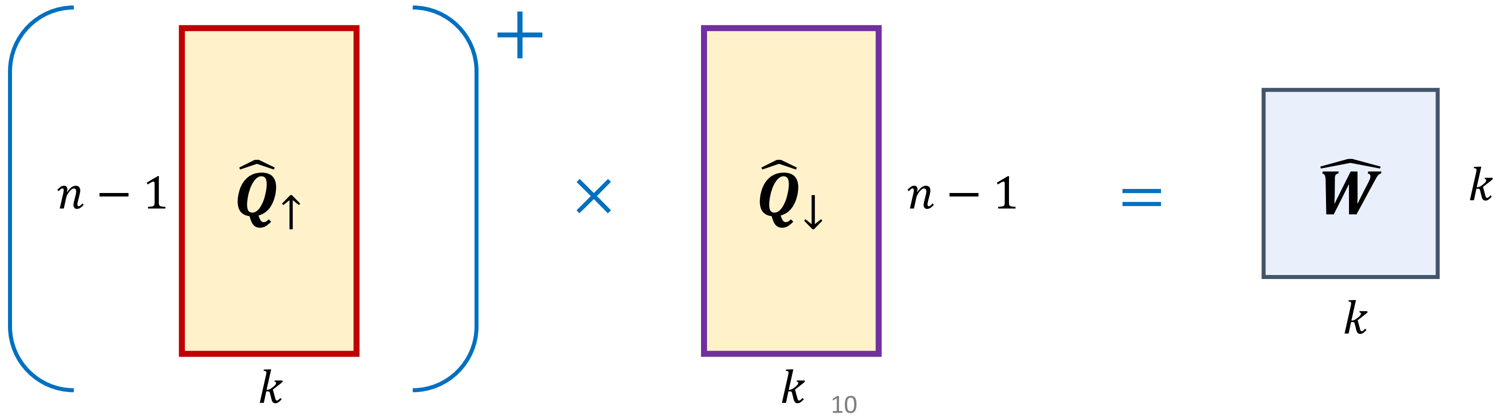
The ESPRIT algorithm

Estimation of signal parameters via rotational invariance techniques

STEP 2: Eigen-decomposition of $\widehat{\boldsymbol{T}}$



STEP 3: Comparing the sub-matrices of $\widehat{\boldsymbol{Q}}$



A^+: Moore–Penrose pseudo-inverse

The ESPRIT algorithm

Estimation of signal parameters via rotational invariance techniques

STEP 4: eigen-decomposition of $\widehat{\boldsymbol{W}}$

- Let $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k$ be the eigenvalues of $\widehat{\boldsymbol{W}}$
- Output $\left\{\hat{f}_j = \left(\arg \hat{\lambda}_j\right)/2\pi\right\}_{j=1}^{k}$ as the estimated locations

Why does ESPRIT work

Claim. When the signal is noise-free, i.e., $g_\omega = \sum_{j=1}^{k} u_j z_j^{\omega} \coloneqq \sum_{j=1}^{k} u_j e^{2\pi \mathbf{i} f_j \omega}$, then the ESPRIT algorithm can recover $\{z_j\}$ exactly (up to a permutation)

- Let $\boldsymbol{z} \coloneqq (z_1, z_2, \dots, z_k)$ and $\boldsymbol{u} \coloneqq (u_1, u_2, \dots, u_k)$
- The "clean" Toeplitz matrix $\boldsymbol{T}$ has a Vondermonde decomposition:

$$\begin{bmatrix} g_0 & \overline{g_1} & \overline{g_2} & \cdots & \overline{g_{n-1}} \\ g_1 & g_0 & \overline{g_1} & \cdots & \overline{g_{n-2}} \\ g_2 & g_1 & g_0 & \cdots & \overline{g_{n-3}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{n-1} & g_{n-2} & g_{n-3} & \cdots & g_0 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_k \\ z_1^2 & z_2^2 & \cdots & z_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{n-1} & z_2^{n-1} & \cdots & z_k^{n-1} \end{bmatrix}}_{\boldsymbol{V}_n(\boldsymbol{z})} \cdot \underbrace{\begin{bmatrix} u_1 & & & \\ & u_2 & & \\ & & \ddots & \\ & & & u_k \end{bmatrix}}_{\mathrm{diag}(\boldsymbol{u})} \cdot \underbrace{\begin{bmatrix} 1 & z_1^{-1} & z_1^{-2} & \cdots & z_1^{-n+1} \\ 1 & z_2^{-1} & z_2^{-2} & \cdots & z_2^{-n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_k^{-1} & z_k^{-2} & \cdots & z_k^{-n+1} \end{bmatrix}}_{\boldsymbol{V}_n(\boldsymbol{z})^\dagger}$$

- $\boldsymbol{T} = \boldsymbol{V}_n(\boldsymbol{z}) \cdot \mathrm{diag}(\boldsymbol{\mu}) \cdot \boldsymbol{V}_n(\boldsymbol{z})^\dagger = \boldsymbol{Q} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{Q}^\dagger = \boldsymbol{Q}_r \cdot \boldsymbol{\Sigma}_r \cdot \boldsymbol{Q}_r^\dagger$ (drop 0 eigenvalues)

  $\to$ $\mathrm{Range}(\boldsymbol{V}_n(\boldsymbol{z})) = \mathrm{Range}(\boldsymbol{Q}_r)$

Why the ESPRIT algorithm works

- $\text{Range}\big(\boldsymbol{V}_n(\boldsymbol{z})\big) = \text{Range}(\boldsymbol{Q}_r)$ implies that there exists an invertible $\boldsymbol{P}$ such that $\boldsymbol{Q}_r = \boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{P}$
- The sub-matrices of the Vandermonde matrix has the following structure:

$$\underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_k \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{n-2} & z_2^{n-2} & \cdots & z_k^{n-2} \end{bmatrix}}_{\boldsymbol{V}_n(\boldsymbol{z})_\uparrow} \cdot \underbrace{\begin{bmatrix} z_1 & & & \\ & z_2 & & \\ & & \ddots & \\ & & & z_k \end{bmatrix}}_{\text{diag}(\boldsymbol{z})} = \underbrace{\begin{bmatrix} z_1 & z_2 & \cdots & z_k \\ z_1^2 & z_2^2 & \cdots & z_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{n-1} & z_2^{n-1} & \cdots & z_k^{n-1} \end{bmatrix}}_{\boldsymbol{V}_n(\boldsymbol{z})_\downarrow}$$

- They imply that $\boldsymbol{W}$ is similar to $\text{diag}(\boldsymbol{z})$:

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{Q}_\uparrow^+ \boldsymbol{Q}_\downarrow = \boldsymbol{Q}_\uparrow^+ (\boldsymbol{V}_n(\boldsymbol{z})_\downarrow \boldsymbol{P}) = \boldsymbol{Q}_\uparrow^+ \big(\boldsymbol{V}_n(\boldsymbol{z})_\uparrow \text{diag}(\boldsymbol{z})\big)\boldsymbol{P} \\ &= \boldsymbol{Q}_\uparrow^+ (\boldsymbol{Q}_\uparrow \boldsymbol{P}^{-1})\text{diag}(\boldsymbol{z})\boldsymbol{P} \\ &= \boldsymbol{P}^{-1}\text{diag}(\boldsymbol{z})\boldsymbol{P} \end{aligned}$$

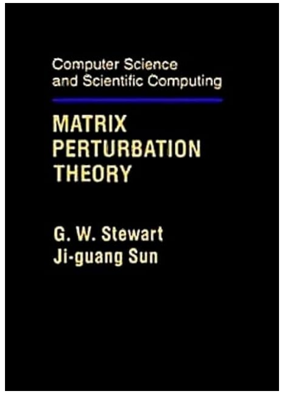
Noise stability of ESPRIT

- $\boldsymbol{T} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^\dagger$
- $\boldsymbol{W} = \boldsymbol{Q}_\uparrow^+ \boldsymbol{Q}_\downarrow$
- $\mathrm{eig}(\boldsymbol{W})$

- $\widehat{\boldsymbol{T}} := \boldsymbol{T} + \boldsymbol{E} = \widehat{\boldsymbol{Q}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{Q}}^\dagger$
- $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{Q}}_\uparrow^+ \widehat{\boldsymbol{Q}}_\downarrow$
- $\mathrm{eig}(\widehat{\boldsymbol{W}})$

Lemma (Li-Liao-Fannjiang ’20; Ding-Epperly-Lin-**Z.** ’24). There exists a unitary matrix $\boldsymbol{U}$ such that

$$\widehat{\boldsymbol{Q}} \approx \boldsymbol{Q}\boldsymbol{U} \quad \text{and} \quad \widehat{\boldsymbol{Q}}_\uparrow^+ \widehat{\boldsymbol{Q}}_\downarrow = \boldsymbol{U}^\dagger \boldsymbol{Q}_\uparrow^+ \boldsymbol{Q}_\downarrow \boldsymbol{U}$$



+

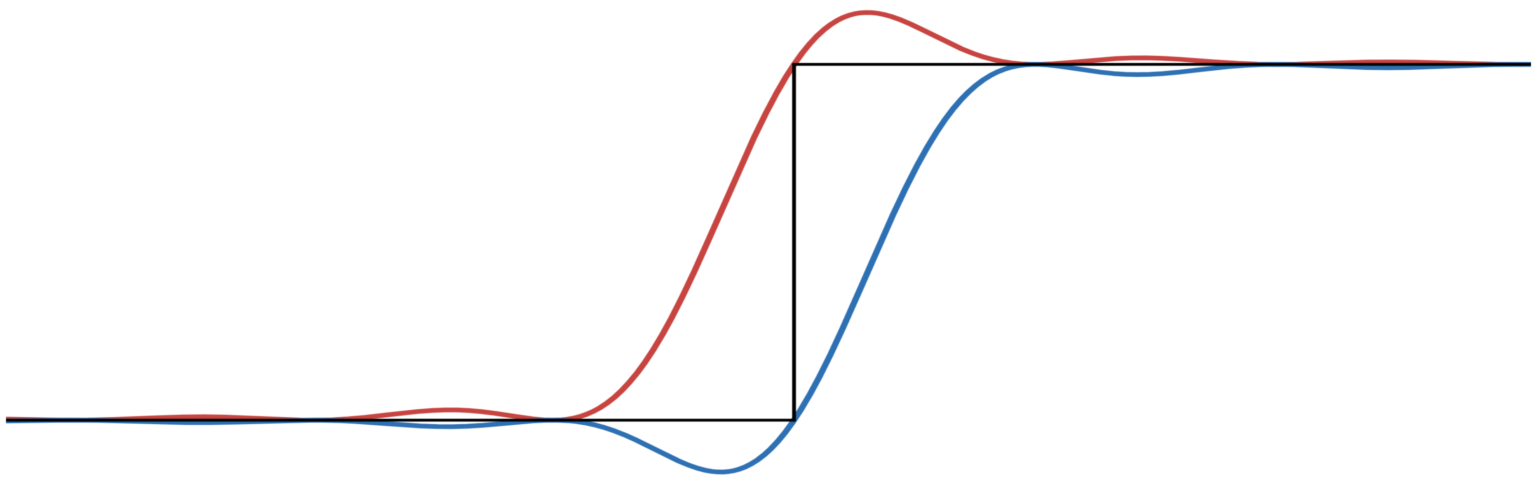
Condition number of the Vondermonde matrix

Condition number bounds

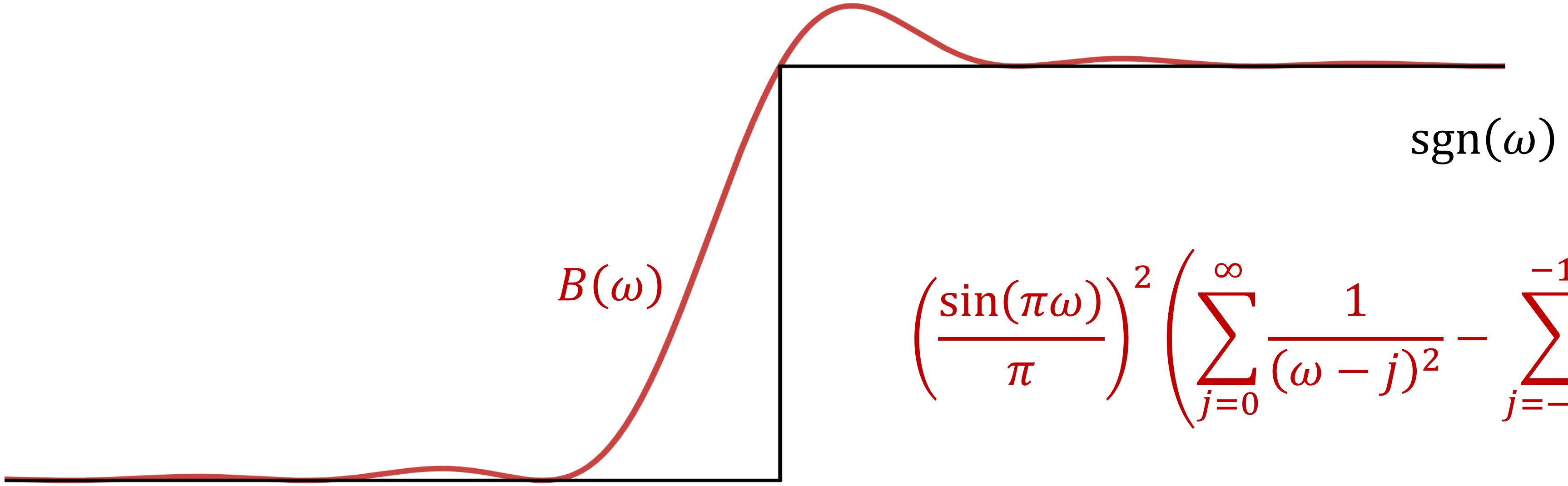
Proposition.

For any $\boldsymbol{u}$, $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 = (n - 1 \pm 1/\Delta)\|\boldsymbol{u}\|^2$, provided that $n > 1/\Delta + 1$

Main technical tool: extremal functions



The Beurling-Selberg majorant



$$\left(\frac{\sin(\pi\omega)}{\pi}\right)^2 \left(\sum_{j=0}^{\infty} \frac{1}{(\omega-j)^2} - \sum_{j=-\infty}^{-1} \frac{1}{(\omega-j)^2} + \frac{2}{\omega}\right)$$

Properties:

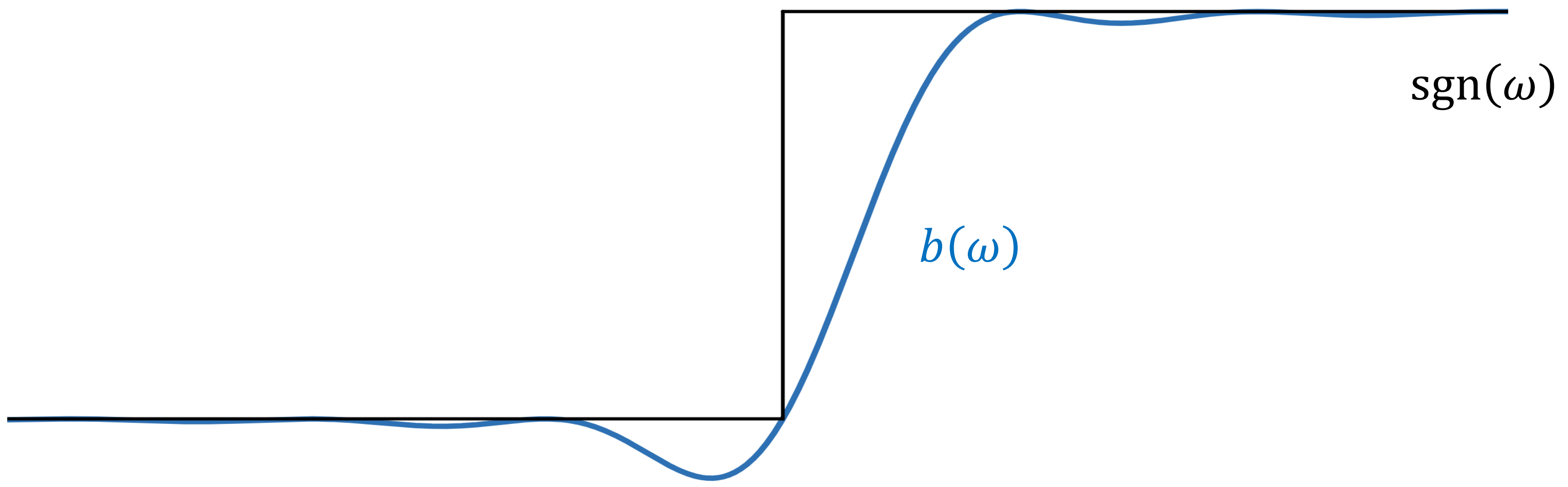
1) $\operatorname{sgn}(\omega) \leq B(\omega)$
2) $\hat{B}(t)$ supported in $[-1,1]$
3) $\int_{-\infty}^{\infty} B(\omega) - \operatorname{sgn}(\omega)\, d\omega = 1$

$B(\omega)$ is an extremal function:

- For any $F(\omega)$ satisfying 1) and 2),

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

The Beurling-Selberg minorant



Properties:

1) $\mathrm{sgn}(\omega) \geq b(\omega)$
2) $\hat{b}(t)$ supported in $[-1,1]$
3) $\int_{-\infty}^{\infty} \mathrm{sgn}(\omega) - b(\omega) d\omega = 1$

Approximate the indicator function of an interval

Corollary.

There are functions $C_E(\omega)$ and $c_E(\omega)$ for $E = [0, m-1]$ that satisfy:

- $c_E(\omega) \leq I_E(\omega) \leq C_E(\omega)$
- $\widehat{c_E}(t)$ and $\widehat{C_E}(t)$ supported in $[-\Delta, \Delta]$
- $\int_{-\infty}^{\infty} C_E(\omega) - I_E(\omega) d\omega = \int_{-\infty}^{\infty} I_E(\omega) - c_E(\omega) d\omega = 1/\Delta$

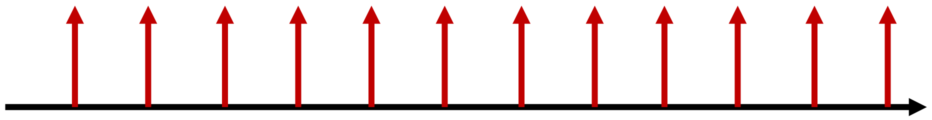
Condition number bounds

Proposition.

For any $\boldsymbol{u}$, $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 = (n - 1 \pm 1/\Delta)\|\boldsymbol{u}\|^2$, provided that $n > 1/\Delta + 1$

Proof.

- $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 = \sum_{\omega=0}^{n-1} |g_\omega|^2$
- Let $h(\omega) \coloneqq \sum_{t=-\infty}^{\infty} \delta_t(\omega)$ be the Dirac comb



- Then, we have

$$\sum_{\omega=0}^{n-1} |g_\omega|^2 = \int_{-\infty}^{\infty} h(\omega) I_E(\omega) |g_\omega|^2 d\omega \leq \int_{-\infty}^{\infty} h(\omega) C_E(\omega) |g_\omega|^2 d\omega$$

$$= \sum_{j,j' \in [k]} u_j \overline{u_{j'}} \int_{-\infty}^{\infty} h(\omega) C_E(\omega) e^{2\pi \mathbf{i}\left(f_j - f_{j'}\right)\omega} d\omega$$

Condition number bounds

Proposition.

For any $\boldsymbol{u}$, $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 = (n - 1 \pm 1/\Delta)\|\boldsymbol{u}\|^2$, provided that $n > 1/\Delta + 1$

Proof.

- $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 = \sum_{\omega=0}^{n-1} |g_\omega|^2$
- Let $h(\omega) \coloneqq \sum_{t=-\infty}^{\infty} \delta_t(\omega) = \sum_{t=-\infty}^{\infty} e^{2\pi \mathbf{i} \omega t}$ (Fourier transform of a comb is a comb)
- Then, we have

$$\sum_{\omega=0}^{n-1} |g_\omega|^2 = \int_{-\infty}^{\infty} h(\omega) I_E(\omega) |g_\omega|^2 d\omega \leq \int_{-\infty}^{\infty} h(\omega) C_E(\omega) |g_\omega|^2 d\omega$$

$$= \sum_{j,j' \in [k]} \sum_{t=-\infty}^{\infty} u_j \overline{u_{j'}} \int_{-\infty}^{\infty} e^{2\pi \mathbf{i} \omega t} C_E(\omega) e^{2\pi \mathbf{i} \left(f_j - f_{j'}\right)\omega} d\omega = \sum_{j,j',t} u_j \overline{u_{j'}} \widehat{C_E}\left(f_j - f_{j'} + t\right)$$

Condition number bounds

Proposition.

For any $\boldsymbol{u}$, $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 = (n - 1 \pm 1/\Delta)\|\boldsymbol{u}\|^2$, provided that $n > 1/\Delta + 1$

Proof.

$$\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 \leq \sum_{\omega=0}^{n-1} |g_\omega|^2 = \sum_{j,j' \in [k]} \sum_{t=-\infty}^{\infty} u_j \overline{u_{j'}} \widehat{C_E}\left(f_j - f_{j'} + t\right)$$

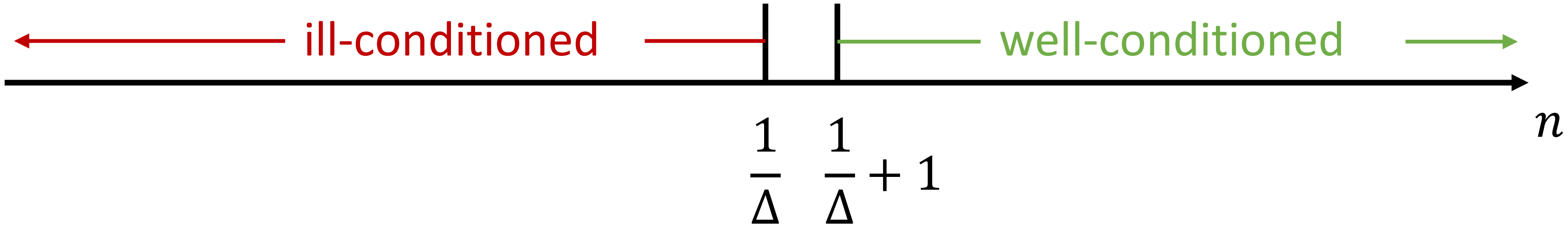
- By the separation condition, $f_j - f_{j'} + t \notin [-\Delta, \Delta]$ for any integer $t \neq 0$
- Hence,
$$\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 \leq \|\boldsymbol{u}\|^2 \widehat{C_E}(0) = \|\boldsymbol{u}\|^2(|E| + 1/\Delta) = \|\boldsymbol{u}\|^2(n - 1 + 1/\Delta)$$
- Using $b_E(\omega)$, we can show that
$$\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|^2 \geq \|\boldsymbol{u}\|^2(n - 1 - 1/\Delta)$$

■

Sharp phase transition of the condition number

Proposition.

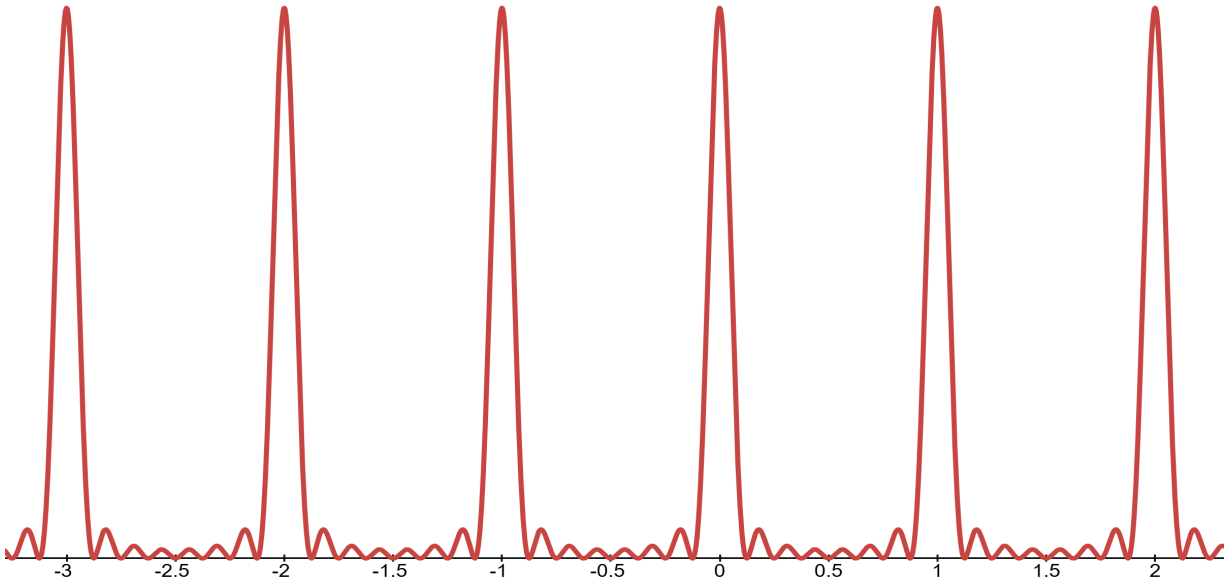
If $n = (1-\epsilon)/\Delta$, then there exists a Δ-separated $\boldsymbol{z}$ such that $\boldsymbol{V}_n(\boldsymbol{z})$ has condition number $2^{\Omega(\epsilon k)}$



- Main technical tool: Fejer kernel

Fejer kernel

$$K_L(\omega) = \frac{1}{L^2}\sum_{t=-L}^{L}(L-|t|)e^{2\pi\mathbf{i}\omega t}$$
$$= 1/L^2\left(\frac{\sin(\pi L\omega)}{\sin(\pi\omega)}\right)^2$$



Properties:

1) $\widehat{K_L}(t) \geq 0$, supported on $\{-L, -L+1, \ldots, L-1, L\}$, and sum to 1

2) $K_L(\omega) \leq \frac{1}{4L^2\omega^2}$ for $\omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$

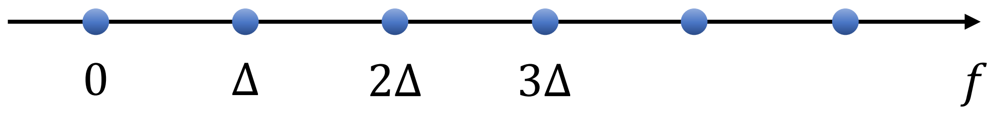
Sharp phase transition of the condition number

Proposition.

If $n = (1-\epsilon)/\Delta$, then there exists a Δ-separated $\boldsymbol{z}$ such that $\boldsymbol{V}_n(\boldsymbol{z})$ has condition number $2^{\Omega(\epsilon k)}$

Proof.

- Let $z_j = e^{2\pi \mathbf{i} j\Delta}$ for $j \in [k]$. Our goal is to construct a vector $\boldsymbol{u}$ such that $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\| = 2^{-\Omega(\epsilon k)}\|\boldsymbol{u}\|$
- Let $L = \left\lceil \frac{4}{\epsilon} \right\rceil$ and $r = \left\lfloor \frac{k-1}{2L} \right\rfloor$ so that $2rL + 1 \le k$



- Let $a_t \coloneqq \widehat{K_L^r}(t)$ for $|t| \le rL$. Then, $\sum_t a_t = 1$
- Define

$$u_j = \begin{cases} a_{j-rL} e^{-\mathbf{i}\pi(j-rL)} e^{-\mathbf{i}\pi jn\Delta}, & 0 \le j \le 2rL \\ 0, & 2rL < j < k \end{cases}$$

Sharp phase transition of the condition number

Proposition.

If $n = (1-\epsilon)/\Delta$, then there exists a Δ-separated $\boldsymbol{z}$ such that $\boldsymbol{V}_n(\boldsymbol{z})$ has condition number $2^{\Omega(\epsilon k)}$

Proof.

$$u_j = \begin{cases} a_{j-rL} e^{-\mathbf{i}\pi(j-rL)} e^{-\mathbf{i}\pi j n\Delta}, & 0 \le j \le 2rL \\ 0, & 2rL < j < k \end{cases}$$

- $\|u\|_1 = \sum_t a_t = 1$. Thus, $\|u\|_2 \ge 1/\sqrt{k}$
- We have

$$(\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u})_l = \sum_{j=0}^{k-1} u_j e^{2\pi \mathbf{i} j l\Delta} = \sum_{t=-rL}^{rL} a_t e^{-\mathbf{i}\pi t} e^{-\mathbf{i}\pi(t+rL)n\Delta} e^{2\pi\mathbf{i}(t+rL)l\Delta} = e^{\mathbf{i}\phi(l)} \underbrace{\sum_{t=-rL}^{rL} a_t e^{2\pi i t\left(l\Delta - \frac{n\Delta+1}{2}\right)}}_{K_L^r\left(l\Delta - \frac{n\Delta+1}{2}\right)}$$

Sharp phase transition of the condition number

Proposition.

If $n = (1-\epsilon)/\Delta$, then there exists a Δ-separated $\boldsymbol{z}$ such that $\boldsymbol{V}_n(\boldsymbol{z})$ has condition number $2^{\Omega(\epsilon k)}$

Proof.

$$u_j = \begin{cases} a_{j-rL} e^{-\mathbf{i}\pi(j-rL)} e^{-\mathbf{i}\pi j n \Delta}, & 0 \le j \le 2rL \\ 0, & 2rL < j < k \end{cases}$$

- $\|u\|_1 = \sum_t a_t = 1$. Thus, $\|u\|_2 \ge 1/\sqrt{k}$
- We have

$$|(\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u})_l| = K_L^r(l\Delta - (n\Delta + 1)/2) \le (4L^2(\epsilon/4)^2)^{-r}$$

Sharp phase transition of the condition number

Proposition.

If $n = (1-\epsilon)/\Delta$, then there exists a Δ-separated $\boldsymbol{z}$ such that $\boldsymbol{V}_n(\boldsymbol{z})$ has condition number $2^{\Omega(\epsilon k)}$

Proof.

$$u_j = \begin{cases} a_{j-rL} e^{-\mathbf{i}\pi(j-rL)} e^{-\mathbf{i}\pi j n\Delta}, & 0 \le j \le 2rL \\ 0, & 2rL < j < k \end{cases}$$

- $\|u\|_1 = \sum_t a_t = 1$. Thus, $\|u\|_2 \ge 1/\sqrt{k}$
- We have

$$|(\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u})_l| \le (4L^2(\epsilon/4)^2)^{-r} = \exp\big(-\Omega(r)\big) = \exp\big(-\Omega(\epsilon k)\big)$$

- Therefore, $\|\boldsymbol{V}_n(\boldsymbol{z})\boldsymbol{u}\|_\infty = \exp\big(-\Omega(\epsilon k)\big)$

$\blacksquare$

Why "super-resolution"?

- Physically, it means that we can resolve real-world objects bypassing the Abbe limit ($\approx 200\ \mathrm{nm}$)
- In our setting, the signal is represented by a superposition of point sources,

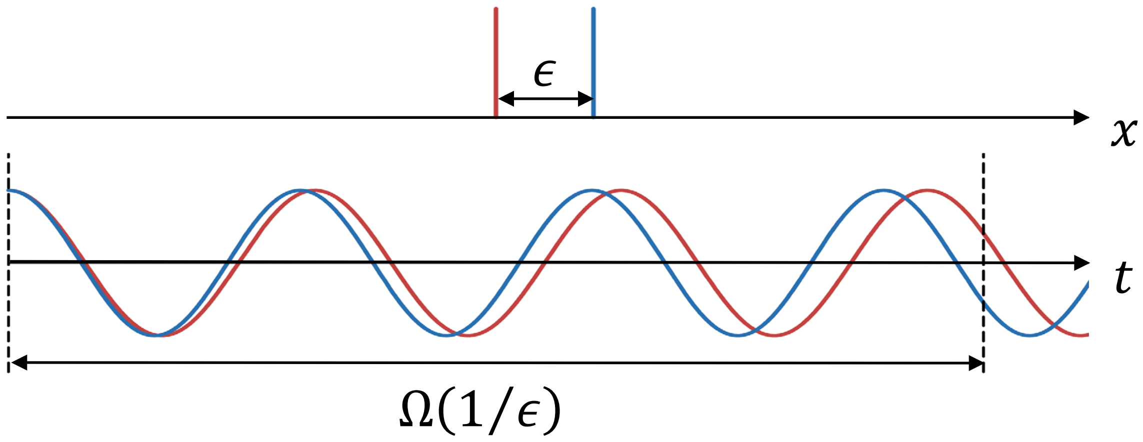
$$x(t) = \sum_{j=1}^{k} u_j \delta_{f_j}(t)$$

  which should be understood as a purely mathematical idealization

- The way to understand super-resolution is from the error scaling in terms of the number of measurements

Error scaling of spectral estimation

To estimate the locations $\{f_j\}$ upto ϵ error, how many measurements (n) do we need?



$e^{2\pi\mathbf{i}\lambda t}$ vs. $e^{2\pi\mathbf{i}(\lambda+\epsilon)t}$

Nyquist scaling

Noiseless

Error scaling of spectral estimation

To estimate the locations $\{f_j\}$ upto ϵ error, how many measurements (n) do we need?

- "Super-resolution" means that the error scaling is asymptotically better than the Nyquist scaling $(1/\epsilon)$
- What if the signal has larger noise?

Super-resolution scaling

Noise $\sim \mathrm{poly}(\epsilon)$

Nyquist scaling

Noise $\sim O(1)$

Central-limit scaling

Noise $\sim O(1)$

$O(1/\Delta)$ | $O(\epsilon^{-1})$ | $O(\epsilon^{-2})$ | n

(Donoho '92; Candès-Romberg-Tao '06; Liao-Fannjiang '14; Moitra '15; …)

(Price-Song '15)

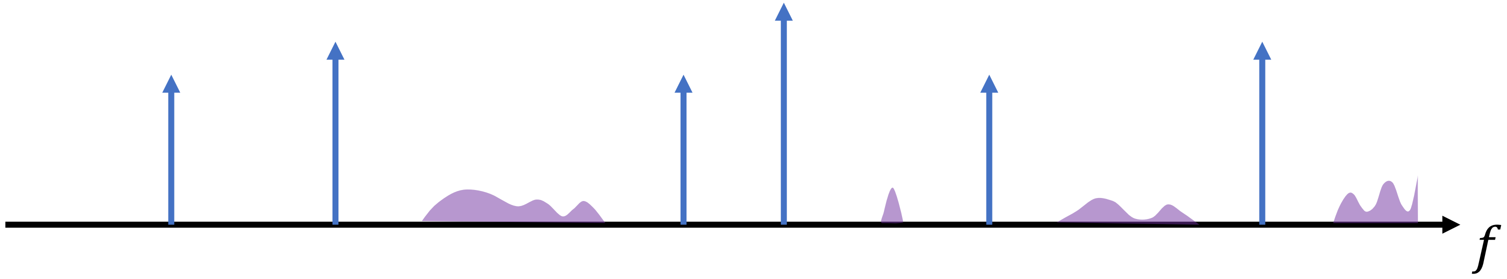
(Li-Liao-Fannjiang '20)

Noisy super-resolution

An algorithm satisfies noisy super-resolution scaling if it can recover the locations up to error strictly superior to the Nyquist error scaling, i.e., $\epsilon = o(n^{-1})$.

Is it possible to achieve a **noisy super-resolution** scaling for solving the spectral estimation problem with bias and large measurement noise?

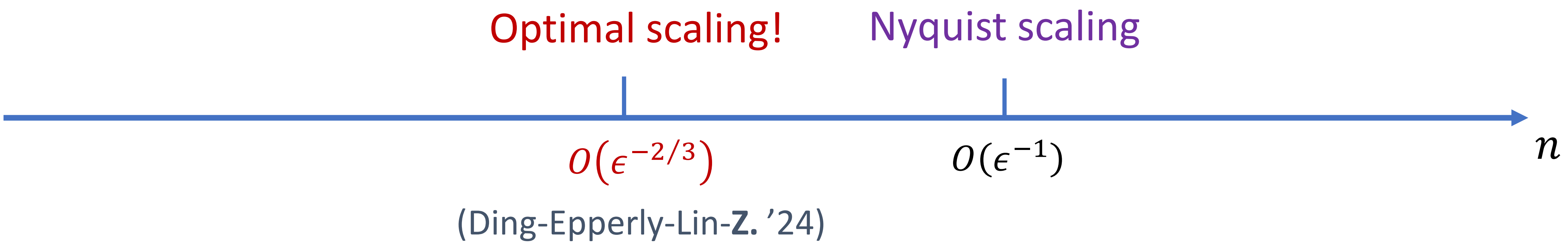
$$g_\omega = \underbrace{\sum_{j=1}^{k} u_j z_j^\omega}_{\text{signal}} + \underbrace{\sum_{j=k+1}^{d} u_j z_j^\omega}_{\text{bias}} + \underbrace{E_\omega}_{\text{random noise}}$$



Noisy super-resolution

An algorithm satisfies noisy super-resolution scaling if it can recover the locations up to error strictly superior to the Nyquist error scaling, i.e., $\epsilon = o(n^{-1})$.

Is it possible to achieve a **noisy super-resolution** scaling for solving the spectral estimation problem with bias and large measurement noise?



ESPRIT for noisy super-resolution

Setup:

- Let the dominant locations and intensities be $\boldsymbol{z}_{\text{dom}} \coloneqq (z_1, \dots, z_k)$ and $\boldsymbol{u}_{\text{dom}} \coloneqq (u_1, \dots, u_k)$.
- Let the tail locations and intensities be $\boldsymbol{z}_{\text{tail}} \coloneqq (z_{k+1}, \dots, z_d)$ and $\boldsymbol{\mu}_{\text{tail}} \coloneqq (z_{k+1}, \dots, z_d)$.
- We have $\widehat{\boldsymbol{T}} = \boldsymbol{T} + \boldsymbol{E}$, $\boldsymbol{E} = \boldsymbol{E}_{\text{tail}} + \boldsymbol{E}_{\text{random}}$, where

$$\begin{aligned}
\boldsymbol{T} \quad &\coloneqq \boldsymbol{V}_n(\boldsymbol{z}_{\text{dom}}) \cdot \operatorname{diag}(\boldsymbol{u}_{\text{dom}}) \cdot \boldsymbol{V}_n(\boldsymbol{z}_{\text{dom}})^\dagger \\
\boldsymbol{E}_{\text{tail}} \quad &\coloneqq \boldsymbol{V}_n(\boldsymbol{z}_{\text{tail}}) \cdot \operatorname{diag}(\boldsymbol{u}_{\text{tail}}) \cdot \boldsymbol{V}_n(\boldsymbol{z}_{\text{tail}})^\dagger \\
\boldsymbol{E}_{\text{random}} &\coloneqq \operatorname{Toep}\big((E_0, E_1, \dots, E_{n-1})\big)
\end{aligned}$$

- We use the matching distance to quantify the estimation error:

$$\operatorname{md}(\boldsymbol{z}, \hat{\boldsymbol{z}}) = \min_{\pi \in \mathcal{S}_k} \max_{1 \le i \le k} \left| z_i - \hat{z}_{\pi(i)} \right|$$

Assumptions

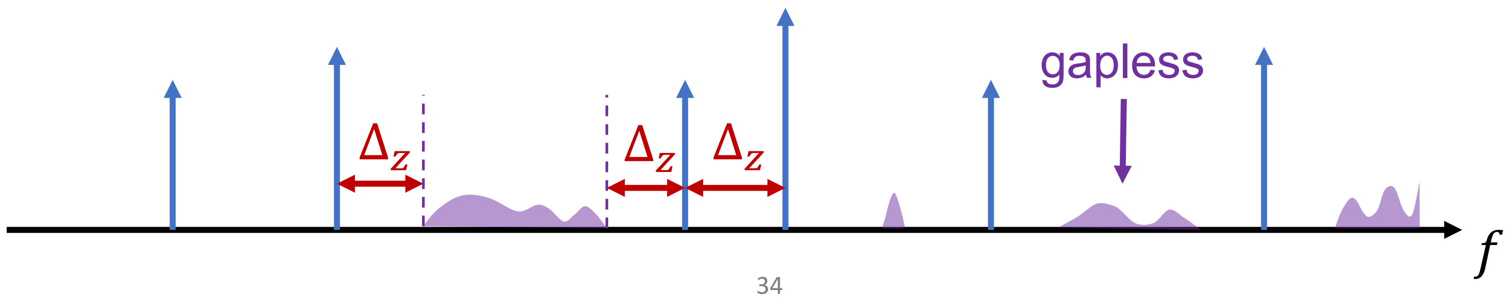
A. Separation of locations. All dominant locations are separated from each other and from non-dominant locations:

$$\Delta_{\mathbf{z}} := \min_{1 \le i \le k, 1 \le j \le d, i \ne j} |z_i - z_j| > 0$$

B. Separation of intensities. We assume the cumulative intensity of non-dominant locations is bounded:

$$u_{\text{tail}} := u_{r+1} + u_{r+2} + \cdots + u_d \ll u_{\min} := \min_{j \in [k]} u_j$$

C. Random measurement noise. We assume that $\{E_j\}_{j \in [d]}$ are independent complex random variables with zero mean and α-sub-Gaussian tail decay ($\alpha > 0$ is the noise level).



Central limit error scaling of ESPRIT

Theorem.

Under **Assumptions A-C**, for sufficiently large cutoff frequency ($n \gg 1/\Delta_{\boldsymbol{z}}$), with high probability, the location estimation of the ESPRIT algorithm satisfies:

$$\mathrm{md}(\hat{\boldsymbol{z}}, \boldsymbol{z}_{\mathrm{dom}}) = \tilde{O}\left(\frac{\alpha}{\mu_r \sqrt{n}}\right)$$

- It recovers the traditional super-resolution error scaling: setting the noise level $\alpha = O(\epsilon \cdot \mu_r \sqrt{n})$ suffices to achieve $\mathrm{md}(\hat{\boldsymbol{z}}, \boldsymbol{z}_{\mathrm{dom}}) \leq \epsilon$ for any $\epsilon > 0$.

Optimal error scaling of ESPRIT

Theorem (Ding-Epperly-Lin-**Z**. '24).

Under **Assumptions A-C**, for sufficiently large n, with high probability, the location estimation of the ESPRIT algorithm satisfies:

$$\mathrm{md}(\hat{\boldsymbol{z}}, \boldsymbol{z}_{\mathrm{dom}}) = \tilde{O}\left(\frac{r^{1.5}\alpha^3}{\mu_r^3 \Delta_z^{1.5} n^{1.5}}\right)$$

And the intensity estimation satisfies:

$$\mathrm{md}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}_{\mathrm{dom}}) = \tilde{O}\left(\frac{r^{2.5}\alpha^3}{\mu_r^3 \Delta_z^{1.5} n^{0.5}}\right)$$

Optimal error scaling of ESPRIT

Theorem (Ding-Epperly-Lin-**Z**. '24).

Under **Assumptions A-C**, for sufficiently large n, with high probability, the location estimation of the ESPRIT algorithm satisfies:

$$\mathrm{md}(\hat{\boldsymbol{z}}, \boldsymbol{z}_{\mathrm{dom}}) = \tilde{O}\left(\frac{r^{1.5}\alpha^3}{\mu_r^3 \Delta_z^{1.5} n^{1.5}}\right)$$

Proof roadmap

Central limit error scaling

Upgrade with novel matrix perturbation results

Optimal error scaling

Key steps for proving the central limit scaling

Our goal is to prove that the eigenvalues of $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{Q}}_{\uparrow}^{+}\widehat{\boldsymbol{Q}}_{\downarrow}$ are close to the eigenvalues of $\boldsymbol{W} = \boldsymbol{Q}_{\uparrow}^{+}\boldsymbol{Q}_{\downarrow}$.
The key idea is to find a similarity transformation $\boldsymbol{A} \mapsto \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}$ to "align" $\widehat{\boldsymbol{W}}$ and $\boldsymbol{W}$.

1. Establishing a quantitative estimate that relates $\boldsymbol{Q}_r$ and $\widehat{\boldsymbol{Q}}_r$:

Eigenvector comparison, weak estimate: There exists a unitary matrix $\boldsymbol{U}_r \in \mathbb{C}^{r\times r}$ such that

$$\left\|\widehat{\boldsymbol{Q}}_r - \boldsymbol{Q}_r\boldsymbol{U}_r\right\|_2 \sim 1/\sqrt{n} \quad \text{and} \quad \left\|\widehat{\boldsymbol{Q}}_{\uparrow}^{+}\widehat{\boldsymbol{Q}}_{\downarrow} - \boldsymbol{U}_r^{\dagger}\boldsymbol{Q}_{\uparrow}^{+}\boldsymbol{Q}_{\downarrow}\boldsymbol{U}_r\right\|_2 \sim 1/\sqrt{n}$$

Main ingredients of the proof:

→ Bounds on the singular values of Vandermonde matrices (Moitra '15):

$$\sigma_i\big(\boldsymbol{V}_n(\boldsymbol{z}_{\mathrm{dom}})\big) \in \sqrt{n - 1 \pm 2\pi/\Delta_{\boldsymbol{z}}} \qquad \forall 1 \le i \le r.$$

→ Standard matrix perturbation theory (Stewart-Sun '90)

→ Matrix concentration inequality: $\|\boldsymbol{E}_{\mathrm{random}}\| = \tilde{O}(\alpha\sqrt{n})$

Key steps for proving the central limit scaling

Our goal is to prove that the eigenvalues of $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{Q}}_\uparrow^\dagger \widehat{\boldsymbol{Q}}_\downarrow$ are close to the eigenvalues of $\boldsymbol{W} = \boldsymbol{Q}_\uparrow^\dagger \boldsymbol{Q}_\downarrow$.
The key idea is to find a similarity transformation $\boldsymbol{A} \mapsto \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}$ to "align" $\widehat{\boldsymbol{W}}$ and $\boldsymbol{W}$.

1. Establishing a quantitative estimate that relates $\boldsymbol{Q}_r$ and $\widehat{\boldsymbol{Q}}_r$:

> Eigenvector comparison, weak estimate: There exists a unitary matrix $\boldsymbol{U}_r \in \mathbb{C}^{r\times r}$ such that
> $$\left\|\widehat{\boldsymbol{Q}}_r - \boldsymbol{Q}_r\boldsymbol{U}_r\right\|_2 \sim 1/\sqrt{n} \quad \text{and} \quad \left\|\widehat{\boldsymbol{Q}}_\uparrow^\dagger \widehat{\boldsymbol{Q}}_\downarrow - \boldsymbol{U}_r^\dagger \boldsymbol{Q}_\uparrow^\dagger \boldsymbol{Q}_\downarrow \boldsymbol{U}_r\right\|_2 \sim 1/\sqrt{n}$$

2. Converting eigenvector perturbations to ESPRIT's estimation error:

> From $\boldsymbol{Q}_\uparrow^\dagger \boldsymbol{Q}_\downarrow$ bounds to location estimation: For any invertible, near-isometry matrix $\boldsymbol{P}$, i.e., $\|\boldsymbol{P}\|_2, \|\boldsymbol{P}^{-1}\|_2 = O(1)$,
> $$\mathrm{md}(\hat{\boldsymbol{z}}, \boldsymbol{z}_{\mathrm{dom}}) \propto \left\|\widehat{\boldsymbol{Q}}_\uparrow^\dagger \widehat{\boldsymbol{Q}}_\downarrow - \boldsymbol{P}^{-1}\boldsymbol{Q}_\uparrow^\dagger \boldsymbol{Q}_\downarrow \boldsymbol{P}\right\|_2$$

→ Taking $\boldsymbol{P} = \boldsymbol{U}_r$ yields the $\tilde{O}(n^{-0.5})$ error scaling.

Towards the optimal error scaling

Eigenvector comparison, strong estimate: There exists an invertible, near-isometry matrix $\boldsymbol{P} \in \mathbb{C}^{r\times r}$ such that

$$\left\|\widehat{\boldsymbol{Q}}_{\uparrow}^{+}\widehat{\boldsymbol{Q}}_{\downarrow} - \boldsymbol{P}^{-1}\boldsymbol{Q}_{\uparrow}^{+}\boldsymbol{Q}_{\downarrow}\boldsymbol{P}\right\|_2 \sim n^{-1.5}$$

Combined with the second step of the central limit scaling proof, we obtain the optimal $n^{-1.5}$ scaling

- The matrix $\boldsymbol{P}$ is **not** unitary. We believe that if $\boldsymbol{P}$ is restricted to be unitary, then the best possible scaling would be $1/\sqrt{n}$
- This result cannot be proven by directly using standard matrix perturbation theory results
  We need:
  - → a novel eigenspace perturbation result
  - → a careful series expansion of $\boldsymbol{P}\widehat{\boldsymbol{Q}}_{\uparrow}^{+}\widehat{\boldsymbol{Q}}_{\downarrow}\boldsymbol{P}^{-1}$
  - → the Toeplitz structure of the error terms in the perturbation

Structure lemma on eigenspace perturbation

Second-order perturbation for dominant eigenspace: There exists a unitary matrix $\boldsymbol{U}_r$ such that

$$\widehat{\boldsymbol{Q}}_r \boldsymbol{U}_r = \boldsymbol{Q}_r + \underbrace{\sum_{k=0}^{\infty} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^{\perp}} \left(\boldsymbol{E}_{\text{tail}} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^{\perp}} \right)^k \boldsymbol{E}_{\text{random}} \boldsymbol{Q}_r (\boldsymbol{\Sigma}_r^{-1})^{k+1} + \tilde{O}(n^{-0.5}) \cdot \boldsymbol{\Pi}_{\boldsymbol{Q}_r} \widetilde{\boldsymbol{Q}}_1}_{\text{first order terms}} + \underbrace{\tilde{O}(n^{-1}) \cdot \widetilde{\boldsymbol{Q}}_2}_{\text{second order term}}$$

- Here, $\boldsymbol{\Pi}_{\boldsymbol{Q}_r} = \boldsymbol{Q}_r \boldsymbol{Q}_r^{\dagger}$ is the projector onto the column space of $\boldsymbol{Q}_r$, $\boldsymbol{\Pi}_{\boldsymbol{Q}_r^{\perp}} = \mathbf{I}_r - \boldsymbol{\Pi}_{\boldsymbol{Q}_r}$, and $\widetilde{\boldsymbol{Q}}_1, \widetilde{\boldsymbol{Q}}_2$ are matrices with $O(1)$ spectral norm.
- This lemma intuitively says that $\widehat{\boldsymbol{Q}}_r$ can be expressed as the sum of four parts (up to a unitary):
  1. The eigenvectors $\boldsymbol{Q}_r$
  2. A term of size $1/\sqrt{n}$ that is orthogonal to $\boldsymbol{Q}_r$
  3. A term of size $1/\sqrt{n}$ that is in the range of $\boldsymbol{Q}_r$
  4. Second-order terms of size $1/n$

Structure lemma on eigenspace perturbation

Second-order perturbation for dominant eigenspace: There exists a unitary matrix $\boldsymbol{U}_r$ such that

$$\widehat{\boldsymbol{Q}}_r \boldsymbol{U}_r = \boldsymbol{Q}_r + \underbrace{\sum_{k=0}^{\infty} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^{\perp}} \left(\boldsymbol{E}_{\text{tail}} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^{\perp}} \right)^k \boldsymbol{E}_{\text{random}} \boldsymbol{Q}_r (\boldsymbol{\Sigma}_r^{-1})^{k+1} + \tilde{O}(n^{-0.5}) \cdot \boldsymbol{\Pi}_{\boldsymbol{Q}_r} \widetilde{\boldsymbol{Q}}_1}_{\text{first order terms}} + \underbrace{\tilde{O}(n^{-1}) \cdot \widetilde{\boldsymbol{Q}}_2}_{\text{second order term}}$$

- Using this lemma, we can explicitly construct an invertible matrix $\boldsymbol{P}$ such that $\widehat{\boldsymbol{Q}}_r \boldsymbol{P}^{-1} - \boldsymbol{Q}_r$ is almost orthogonal to $\boldsymbol{Q}_r$, up to an error of size $n^{-1.5}$.
- And the orthogonal parts will be approximately cancelled in $\left\| \boldsymbol{P} \widehat{\boldsymbol{Q}}_{\uparrow}^{+} \widehat{\boldsymbol{Q}}_{\downarrow} \boldsymbol{P}^{-1} - \boldsymbol{Q}_{\uparrow}^{+} \boldsymbol{Q}_{\downarrow} \right\|_2$, which implies the strong estimate for the eigenvector comparison.

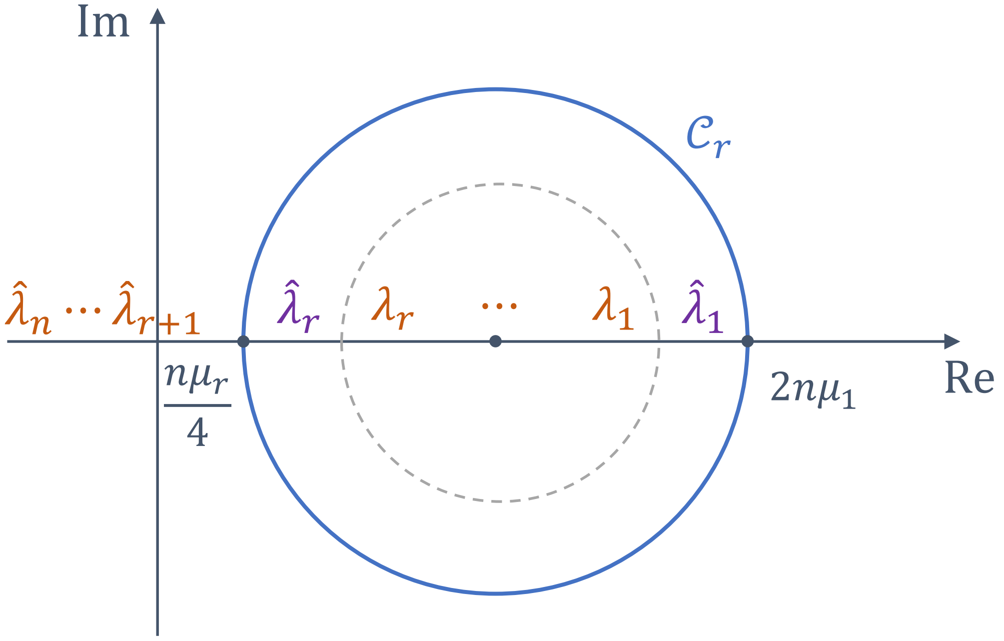
Interlude: perturbation theory via resolvents

- Assume $\boldsymbol{A} = \sum_{i=1}^{r} \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^\dagger \in \mathbb{C}^{n\times n}$ and $\boldsymbol{A} + \boldsymbol{E} = \sum_{i=1}^{n} \widehat{\lambda}_i \widehat{\boldsymbol{q}}_i \widehat{\boldsymbol{q}}_i^\dagger$ be the perturbation.
- Let $\mathcal{C}$ be a simple closed curve in $\mathbb{C}$ such that $\lambda_1, \ldots, \lambda_r, \widehat{\lambda_1}, \ldots, \widehat{\lambda_r}$ are inside $\mathcal{C}$ and all other eigenvalues are outside $\mathcal{C}$.
- Assume that $\|\boldsymbol{E}(\xi \boldsymbol{I} - \boldsymbol{A})^{-1}\| < 1$ for all $\xi \in \mathbb{C}$.
- Then, denoting $\boldsymbol{\Pi} = \sum_{i=1}^{r} \boldsymbol{q}_i \boldsymbol{q}_i^\dagger$ and $\widehat{\boldsymbol{\Pi}} = \sum_{i=1}^{r} \widehat{\boldsymbol{q}}_i \widehat{\boldsymbol{q}}_i^\dagger$, we have

$$\widehat{\boldsymbol{\Pi}} = \boldsymbol{\Pi} + \sum_{k=1}^{\infty} \frac{1}{2\pi \mathbf{i}} \oint_{\mathcal{C}} (\xi \boldsymbol{I} - \boldsymbol{A})^{-1} (\boldsymbol{E}(\xi \boldsymbol{I} - \boldsymbol{A})^{-1})^k d\xi \ .$$

Expansion of spectral projector

$$\mathbf{\Pi}_{\widehat{\boldsymbol{Q}}_r} = \mathbf{\Pi}_{\boldsymbol{Q}_r} + \sum_{k=1}^{\infty} \frac{1}{2\pi\mathbf{i}} \oint_{\mathcal{C}_r} (\xi\boldsymbol{I} - \boldsymbol{T})^{-1}(\boldsymbol{E}(\xi\boldsymbol{I} - \boldsymbol{T})^{-1})^k d\xi$$

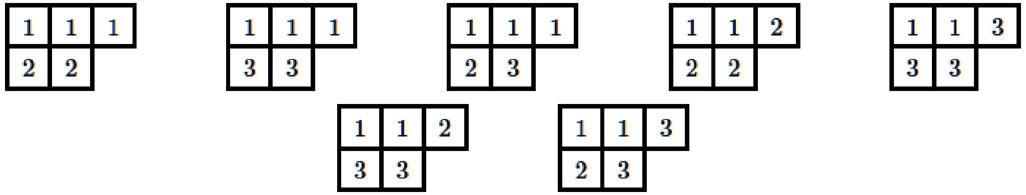


Evaluating the expansion

$$\mathbf{\Pi}_{\widehat{\boldsymbol{Q}}_r} = \mathbf{\Pi}_{\boldsymbol{Q}_r} + \sum_{k=1}^{\infty} \frac{1}{2\pi\mathbf{i}} \oint_{\mathcal{C}_r} (\xi - \boldsymbol{T})^{-1}(\boldsymbol{E}(\xi - \boldsymbol{T})^{-1})^k d\xi$$
$$= \mathbf{\Pi}_{\boldsymbol{Q}_r} + \underbrace{\mathbf{Poly}_{\widehat{\boldsymbol{Q}}_r}\left(\boldsymbol{T}^+, \mathbf{\Pi}_{\boldsymbol{Q}_r^\perp}, \boldsymbol{E}\right)}_{\text{An explicit formula for the contour integrals}}$$

- Expand the eigenvector projections within the integrals
- Combine the terms to form $\boldsymbol{T}^+$
- Simplify the expressions using Schur polynomials

$$s_\lambda(x_1, \ldots, x_n) = \prod_{1 \le i < j \le n} (x_i - x_j)^{-1} \begin{vmatrix} x_1^{\lambda_1+n-1} & x_2^{\lambda_1+n-1} & \ldots & x_n^{\lambda_1+n-1} \\ x_1^{\lambda_2+n-2} & x_2^{\lambda_2+n-2} & \ldots & x_n^{\lambda_2+n-2} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{\lambda_n} & x_2^{\lambda_n} & \ldots & x_n^{\lambda_n} \end{vmatrix}$$



Bounding the higher-order terms

$$
\begin{aligned}
\boldsymbol{\Pi}_{\widehat{\boldsymbol{Q}}_r} &= \boldsymbol{\Pi}_{\boldsymbol{Q}_r} + \sum_{k=1}^{\infty} \frac{1}{2\pi\mathbf{i}} \oint_{\mathcal{C}_r} (\xi \boldsymbol{I} - \boldsymbol{T})^{-1} (\boldsymbol{E}(\xi \boldsymbol{I} - \boldsymbol{T})^{-1})^k d\xi \\
&= \boldsymbol{\Pi}_{\boldsymbol{Q}_r} + \mathbf{Poly}_{\widehat{\boldsymbol{Q}}_r}\left(\boldsymbol{T}^+, \boldsymbol{\Pi}_{\boldsymbol{Q}_r^\perp}, \boldsymbol{E}\right) \\
&= \boldsymbol{\Pi}_{\boldsymbol{Q}_r} + \sum_{k=1}^{\infty} \left((\boldsymbol{T}^+)^k \boldsymbol{E}_{\mathrm{random}} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^\perp} \left(\boldsymbol{E} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^\perp}\right)^{k-1} + h.c. \right) + O(1/n)
\end{aligned}
$$

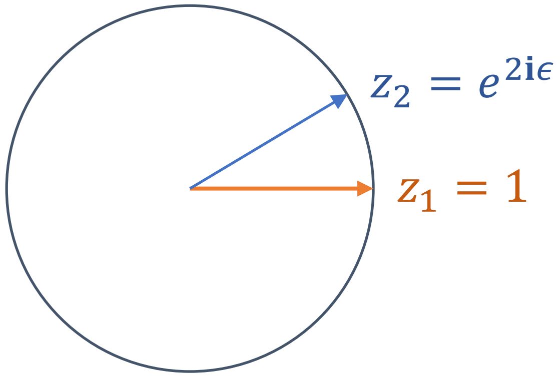
By connecting the angle between subspaces $\boldsymbol{\Pi}_{\widehat{\boldsymbol{Q}}_r}$ and $\boldsymbol{\Pi}_{\boldsymbol{Q}_r}$ to the distance between $\widehat{\boldsymbol{Q}}_r$ and $\boldsymbol{Q}_r$ (up to a unitary transformation), we prove the structure lemma:

Second-order perturbation for dominant eigenspace: There exists a unitary matrix $\boldsymbol{U}_r$ such that

$$
\widehat{\boldsymbol{Q}}_r \boldsymbol{U}_r = \boldsymbol{Q}_r + \underbrace{\sum_{k=0}^{\infty} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^\perp} \left(\boldsymbol{E}_{\mathrm{tail}} \boldsymbol{\Pi}_{\boldsymbol{Q}_r^\perp}\right)^k \boldsymbol{E}_{\mathrm{random}} \boldsymbol{Q}_r (\boldsymbol{\Sigma}_r^{-1})^{k+1} + \tilde{O}(n^{-0.5}) \cdot \boldsymbol{\Pi}_{\boldsymbol{Q}_r} \widetilde{\boldsymbol{Q}}_1}_{\text{first order terms}} + \underbrace{\tilde{O}(n^{-1}) \cdot \widetilde{\boldsymbol{Q}}_2}_{\text{second order term}}
$$

Spectral estimation lower bound

- Suppose there is an algorithm $\mathcal{A}$ that can estimate the locations within ϵ error.
- Consider the following one-sparse signals and noisy measurements:



$$g_j = 1 + \mathcal{N}(0,1) \qquad j = 1,2,\dots,n$$
$$g_j' = e^{2\mathbf{i}\epsilon j} + \mathcal{N}(0,1)$$

- $|z_1 - z_2| > \epsilon$; thus, $\mathcal{A}$ should be able to distinguish these two signals.
- To distinguish two Gaussians $\mathcal{N}(\mathbf{1}, \boldsymbol{I}_n)$ and $\mathcal{N}\left(\left(e^{2\mathbf{i}\epsilon j}\right), \boldsymbol{I}_n\right)$ with constant success probability,

$$d_{\mathrm{TV}}\left(\mathcal{N}(\mathbf{1}, \boldsymbol{I}_n), \mathcal{N}\left(\left(e^{2\mathbf{i}\epsilon j}\right), \boldsymbol{I}_n\right)\right) \leq \left\|\mathbf{1} - \left(e^{2\mathbf{i}\epsilon j}\right)\right\|_2 = O(n^3\epsilon^2) = \Omega(1) \implies \epsilon \leq n^{-1.5}.$$

Recap

1-D super-resolution upper bound:

- The ESPRIT algorithm
- The Beurling-Selberg majorant and minorant to bound the condition number of Vandermonde matrix

1-D super-resolution lower bound:

- Fejer kernel to construct ill-conditioned Vandermonde matrix

Extension to the high-noise regime

- Optimal sample complexity