Super-resolution, subspace methods, and Fourier matrices

Weilin Li
Courant Institute
New York University
https://cims.nyu.edu/∼weilinli

MINDS Seminar
November 4, 2021

Outline

Mathematics of super-resolution

Non-harmonic Fourier matrices

Subspace methods

Fundamental limitations of super-resolution

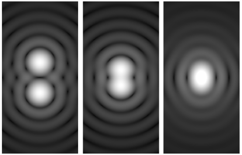
Multiple snapshot super-resolution

Conclusions

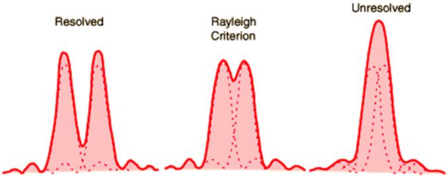
What is resolution?

The **Rayleigh length** of an imaging system is the smallest distance between two point sources that can be resolved.

Loosely referred to as the **resolution** of the imaging device.



(a) Airy disks



(b) One-dimensional plot

Resolution and Fourier transform

Spatial domain:

$$\Big(\text{imaged object}\Big) * \Big(\text{point spread function}\Big) = \Big(\text{low-resolution image}\Big)$$

Fourier domain:

$$\mathcal{F}\Big(\text{imaged object}\Big) \cdot \mathcal{F}\Big(\text{point spread function}\Big) = \mathcal{F}\Big(\text{low-resolution data}\Big)$$

Goal of super-resolution algorithms: Leverage prior information in order to overcome the inherent resolution limit of the imaging device – extract high-resolution features from observed low-resolution or coarse information.

Mathematical model of super-resolution [Donoho 1992]

Unknown: Atomic measure

$$\mu := \sum_{j=1}^{S} a_j \delta_{x_j}, \quad a_j \in \mathbb{C}, \quad x_j \in \mathbb{T} := [0,1).$$

Known: Perturbation of M consecutive Fourier coefficients

$$y := \mathcal{F}_M \mu + \eta \in \mathbb{C}^M$$

where

$$\mathcal{F}_M \mu(m) := \left\{ \widehat{\mu}(m) \right\}_{m=0}^{M-1}$$

and

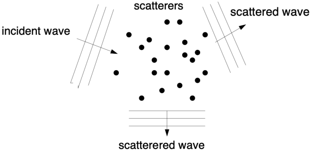
$$\widehat{\mu}(m) := \int_{\mathbb{T}} e^{-2\pi i m x} \, d\mu(x).$$

Goal: Recover μ. Primarily its support.

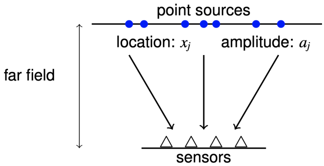
Imaging of point sources

Unknown: Atomic measure

$$\mu := \sum_{j=1}^{S} a_j \delta_{x_j}, \quad a_j \in \mathbb{C}, \quad x_j \in \mathbb{T} := [0,1).$$



(a) Inverse scattering



(b) Direction of arrival

Applications: Super-resolution microscopy, geophysics, astronomy, remote sensing, inverse scattering, direction of arrival, and line spectral estimation.

Mathematical connections: Sparse recovery, non-harmonic Fourier analysis, sub-Nyquist sampling, and diffraction limited imaging.

Three main difficulties

1. It is a non-linear inverse problem:

$$\{(x_j, a_j)\}_{j=1}^{S} \quad \leftrightarrow \quad \mu \quad \mapsto \quad \mathcal{F}_M(\mu).$$

   Well-posed in the sense that this map is injective if $M \geq 2S$ (Prony).

Three main difficulties

1. It is a non-linear inverse problem:

$$\left\{(x_j, a_j)\right\}_{j=1}^{S} \quad \leftrightarrow \quad \mu \quad \mapsto \quad \mathcal{F}_M(\mu).$$

   Well-posed in the sense that this map is injective if $M \geq 2S$ (Prony).

2. We deal with non-harmonic sums:

$$\widehat{\mu}(m) = \sum_{j=1}^{S} a_j e^{-2\pi i m x_j}; \quad \widehat{\mu}(t) = \sum_{j=1}^{S} a_j e^{-2\pi i t x_j}.$$

Three main difficulties

1. It is a non-linear inverse problem:

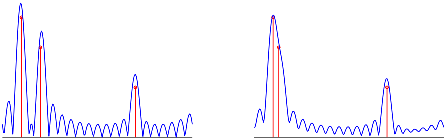
$$\{(x_j, a_j)\}_{j=1}^{S} \quad \leftrightarrow \quad \mu \quad \mapsto \quad \mathcal{F}_M(\mu).$$

   Well-posed in the sense that this map is injective if $M \geq 2S$ (Prony).

2. We deal with non-harmonic sums:

$$\widehat{\mu}(m) = \sum_{j=1}^{S} a_j e^{-2\pi i m x_j}; \quad \widehat{\mu}(t) = \sum_{j=1}^{S} a_j e^{-2\pi i t x_j}.$$

3. Sources can be closely spaced:



(a) separation is $4/M$ (b) separation is $0.6/M$

Figure: $|\mu * D_M|$, where $D_M(x) = \dfrac{1}{M}\sum_{m=0}^{M-1} e^{2\pi i m x}$

Geometric notions

Minimum separation:

$$\Delta := \min_{j \neq k} |x_j - x_k|_{\mathbb{T}}$$

Rayleigh length: $\approx 1/M$

Super-resolution factor: Rayleigh length divided by the target resolution

$$SRF := \frac{\text{standard resolution}}{\text{target resolution}} := \frac{1/M}{\Delta} = \frac{1}{M\Delta}.$$

Sub-Nyquist or diffraction limited regime:

$$SRF \geq 1 \quad \text{or equivalently} \quad \Delta \leq \frac{1}{M}.$$

Existing methods (up to around 2017–2018)

	$\Delta \gg 1/M$ or $SRF < 1$	$\Delta \ll 1/M$ or $SRF \gg 1$
1a. TV-min (μ complex)	$O(\|\eta\|_2)$	Can fail even if $\eta = 0$
1b. TV-min (μ positive)	$O(\|\eta\|_2)$	$O(\Delta^{-2S+1}\|\eta\|_2)$
2. Greedy methods	$O(\|\eta\|_2)$	Can fail if $\eta \neq 0$
3. Subspace methods	$O(\|\eta\|_2)$	Only numerical results
4. "Best" algorithm		$O(SRF^{2S-1}\|\eta\|_2)$

References:

1. TV-min: [Candès, Fernandez-Granda 2013, 2014], [Tang, Bhaskar, Shah, Recht 2013], [Duval, Peyré 2015], [Morgenshtern, Candès 2016], [Denoyelle, Duval, Peyré 2016], [Schiebinger, Robeva, Recht 2017]
2. Greedy: [Duarte, Baraniuk 2013], [Fannjiang, Liao 2012]
3. Subspace: [Liao, Fannjiang 2016], [Moitra 2015]
4. SR Limit: [Donoho 1992], [Demanet, Nguyen 2015]

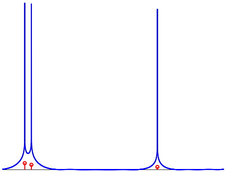
Main questions

Study the $\Delta \ll 1/M$ or $SRF \gg 1$ regime.

Main questions

Study the $\Delta \ll 1/M$ or $SRF \gg 1$ regime.

1. **Computational:** Are there accurate and efficient algorithms?
2. **Computational analysis:** What are the performance guarantees of said algorithms?
3. **Information theoretic:** What is the best possible recovery rate independent of algorithm?

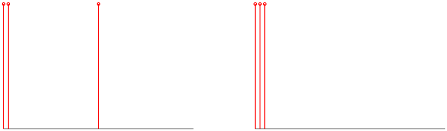


(a) $\Delta = 0.6/M$, no noise

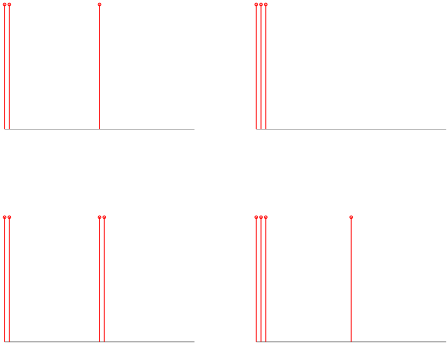
(b) $\Delta = 0.6/M$, $\sigma = 0.01$

Figure: Output of the MUSIC algorithm

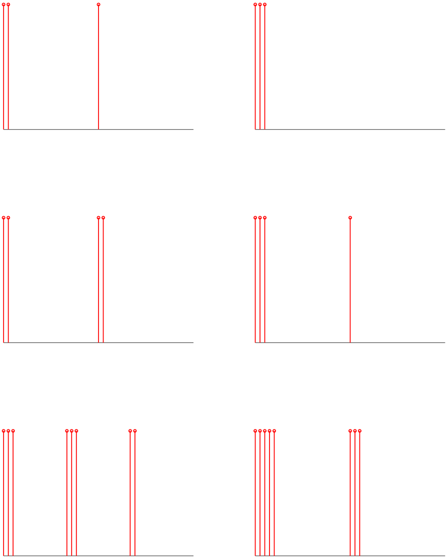
Role of geometry; which is harder?



Role of geometry; which is harder?



Role of geometry; which is harder?



Outline

Mathematics of super-resolution

Non-harmonic Fourier matrices

Subspace methods

Fundamental limitations of super-resolution

Multiple snapshot super-resolution

Conclusions

Fourier matrices

Unknown: Atomic measure $\mu := \sum_{j=1}^{S} a_j \delta_{x_j}$, where $a_j \in \mathbb{C}$ and $x_j \in \mathbb{T} := [0, 1)$.

Known: Consecutive M noisy Fourier coefficients

$$y := \mathcal{F}_M \mu + \eta, \quad \mathcal{F}_M \mu(m) := \int_{\mathbb{T}} e^{-2\pi i m x} \, d\mu(x) \quad \text{for} \quad m = 0, 1, \ldots, M-1.$$

Fourier matrix associated with $X = \{x_j\}_{j=1}^{S}$:

$$\Phi_M := \Phi_M(X) := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-2\pi i x_1} & e^{-2\pi i x_2} & \cdots & e^{-2\pi i x_S} \\ \vdots & \vdots & & \vdots \\ e^{-2\pi i (M-1) x_1} & e^{-2\pi i (M-1) x_2} & \cdots & e^{-2\pi i (M-1) x_S} \end{bmatrix},$$

and so

$$y = \Phi_M(X) a + \eta.$$

Fourier matrices and super-resolution

Fourier matrix associated with $X = \{x_j\}_{j=1}^S$:

$$\Phi_M := \Phi_M(X) := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-2\pi i x_1} & e^{-2\pi i x_2} & \cdots & e^{-2\pi i x_S} \\ \vdots & \vdots & & \vdots \\ e^{-2\pi i (M-1) x_1} & e^{-2\pi i (M-1) x_2} & \cdots & e^{-2\pi i (M-1) x_S} \end{bmatrix}.$$

Minimum singular value $\sigma_S(\Phi_M(X))$ **controls:**

- Robustness of subspace methods (will return to this later).
- Min-max error or information theoretic limit (will return to later).

Connections between super-resolution and Fourier matrices is hinted at by earlier work [Donoho 1992], [Demanet, Nguyen 2015], [Moitra 2015].

Known results about Fourier matrices

Dichotomy:

- If $\Delta \geq C/M$ for some $C > 1$, then $\sigma_S(\Phi_M) \gtrsim \sqrt{M}$.
  [Vaaler 1985], [Moitra 2015], based on work by Beurling and Selberg.
  (Trivially $\sigma_1(\Phi_M) \leq \sqrt{MS}$)

Known results about Fourier matrices

Dichotomy:

- If $\Delta \geq C/M$ for some $C > 1$, then $\sigma_S(\Phi_M) \gtrsim \sqrt{M}$.
  [Vaaler 1985], [Moitra 2015], based on work by Beurling and Selberg.
  (Trivially $\sigma_1(\Phi_M) \leq \sqrt{MS}$)
- If $\Delta < 1/M$, then it appears that $\sigma_S(\Phi_M)$ can be as small as $O(\Delta^{S-1})$.
  [Gautschi 1962], [Gautschi, Inglese 1987], [Bazán 2000].

Known results about Fourier matrices

Dichotomy:

- If $\Delta \geq C/M$ for some $C > 1$, then $\sigma_S(\Phi_M) \gtrsim \sqrt{M}$.
  [Vaaler 1985], [Moitra 2015], based on work by Beurling and Selberg.
  (Trivially $\sigma_1(\Phi_M) \leq \sqrt{MS}$)
- If $\Delta < 1/M$, then it appears that $\sigma_S(\Phi_M)$ can be as small as $O(\Delta^{S-1})$.
  [Gautschi 1962], [Gautschi, Inglese 1987], [Bazán 2000].
  For example, [Gautschi 1962] showed that for a square general Vandermonde matrix with nodes $Z = \{z_1, \ldots, z_S\} \subseteq \mathbb{C}$,

$$\sigma_S(\Phi_S(Z)) \geq \min_{1 \leq j \leq S} \prod_{k=1,\, k \neq j}^{S} \frac{|z_k - z_j|}{1 + |z_k|},$$

  where equality holds if and only if $z_1, \ldots, z_S$ lie on the same line in $\mathbb{C}$.

Known results about Fourier matrices

Dichotomy:

- If $\Delta \geq C/M$ for some $C > 1$, then $\sigma_S(\Phi_M) \gtrsim \sqrt{M}$.
  [Vaaler 1985], [Moitra 2015], based on work by Beurling and Selberg.
  (Trivially $\sigma_1(\Phi_M) \leq \sqrt{MS}$)
- If $\Delta < 1/M$, then it appears that $\sigma_S(\Phi_M)$ can be as small as $O(\Delta^{S-1})$.
  [Gautschi 1962], [Gautschi, Inglese 1987], [Bazán 2000].
  For example, [Gautschi 1962] showed that for a square general Vandermonde matrix with nodes $Z = \{z_1, \ldots, z_S\} \subseteq \mathbb{C}$,

$$\sigma_S(\Phi_S(Z)) \geq \min_{1\leq j\leq S} \prod_{k=1,\, k\neq j}^{S} \frac{|z_k - z_j|}{1 + |z_k|},$$

  where equality holds if and only if $z_1, \ldots, z_S$ lie on the same line in $\mathbb{C}$.
  This bound is sharp, but can be improved under certain configurations.

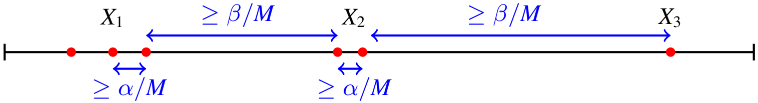
Separated clumps model

Separated clumps: (Two-scale model) A set $X = \{x_j\}_{j=1}^S$ consists of separated clumps with parameters (R, M, S, α, β) if we have the disjoint union

$$X = \bigcup_{r=1}^{R} X_r$$

where each X_r is contained in an interval of length $1/M$ and

1. (Intra-clump separation) $\Delta \geq \alpha/M$ where $\alpha \leq 1$.
2. (Inter-clump separation) If $R > 1$, then $\text{dist}(X_j, X_k) \geq \beta/M$ where $\beta \geq 1$.



Two Extremes:

- $R = S$: each clump is a single point and $\Delta \geq 1/M$.
- $R = 1$: one clump contains all S points.

Lower bound for separated clumps

Theorem (L., W. Liao, ACHA 2021)

Suppose X consists of separated clumps with parameters (R, M, S, α, β), where $M \geq S^2$, $\alpha \in (0,1)$, and $\beta \geq \max_r 20\sqrt{S}\,\lambda_r^{5/2}/\sqrt{\alpha}$ where $\lambda_r = |X_r|$. Then there exist explicit constants $\{C_r\}_{r=1}^R$ that do not depend on M and α such that

$$\sigma_S(\Phi_M(X)) \geq \sqrt{M}\left(\sum_{r=1}^{R} C_r^2 \left(\frac{1}{\alpha}\right)^{2\lambda_r - 2}\right)^{-1/2}.$$

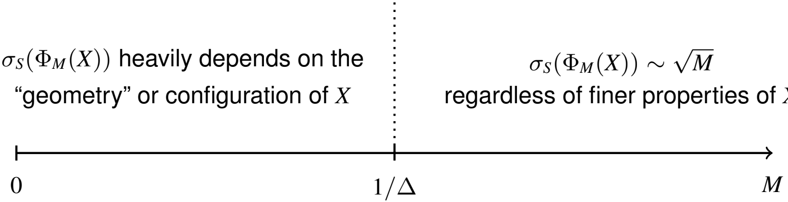
Rough estimate: Letting $\lambda = \max_r \lambda_r$, note that $SRF = \frac{1/M}{\alpha/M} = 1/\alpha$, so

$$\sigma_S(\Phi_M(X)) \geq C(\lambda)\sqrt{\frac{M}{R}}\, SRF^{-\lambda+1} = C(\lambda)\sqrt{\frac{M}{R}}\,(M\Delta)^{\lambda-1}.$$

Phase transition:

$\sigma_S(\Phi_M(X))$ heavily depends on the "geometry" or configuration of X

$\sigma_S(\Phi_M(X)) \sim \sqrt{M}$ regardless of finer properties of X



Related work

Theorem (L., W. Liao, ACHA 2021)

If X consists of separated clumps with parameters (R, M, S, α, β) where $M \geq S^2$, $\alpha \in (0,1)$, and $\beta \geq \max_r 20\sqrt{S}\,\lambda_r^{5/2}/\sqrt{\alpha}$ where $\lambda_r = |X_r|$, there exist explicit constants $\{C_r\}_{r=1}^R$ where C_r only depends on λ_r and

$$\sigma_S(\Phi_M(X)) \geq \sqrt{M}\left(\sum_{r=1}^R C_r^2 \left(\frac{1}{\alpha}\right)^{2\lambda_r-2}\right)^{-1/2}.$$

Brief history on "geometrically" aware lower bounds on $\sigma_S(\Phi_M)$:

- For general Vandermonde matrices, see [Gautschi 1962] for square case, and [Bazán 2000] for rectangular ones.
- For rectangular Fourier matrices, this result appeared in 2018, concurrently with [Batenkov, Demanet, Goldman, Yomdin, 2020].
- Follow up papers [Kunis, Nagel, 2020], [Demanet, Goldman, Yomdin, 2021].

Proof technique

Duality: If $v \in \mathbb{C}^S$ is the S-th right singular vector of $\Phi_M(X)$, then

$$\sigma_S(\Phi_M(X)) = \max_{\substack{f \text{ trig poly degree } M-1 \\ f(X)=v}} \|f\|_{L^2(\mathbb{T})}^{-1}.$$

Method: Construct trigonometric polynomials $\{L_{X,j}\}_{j=1}^S$

1. Each $L_{X,j}$ has degree $M-1$
2. $L_{X,j}(x_j) = 1$
3. $L_{X,j}(x_k) = 0$ for all x_k in the same clump as x_j
4. $L_{X,j}$ decays quickly away from x_j.

$$\Bigg\| \underbrace{\sum_{j=1}^S v_j L_{X,j}}_{\text{approximate interpolant}} \Bigg\|_{L^2(\mathbb{T})} \le \Big(\sum_{j=1}^S \|L_{X,j}\|_{L^2(\mathbb{T})}^2 \Big)^{1/2}.$$

By a robust version of duality,

$$\sigma_S(\Phi_M(X)) \gtrsim \Big(\sum_{j=1}^S \|L_{X,j}\|_{L^2(\mathbb{T})}^2 \Big)^{-1/2}$$

Outline

Subspace methods

Subspace (algebraic) methods:

1. MUSIC [Schmidt 1986]
2. ESPRIT [Roy, Kailath 1989]
3. MPM [Hua, Sarkar 1990]

Hankel matrix: Suppose $M \geq 2S + 1$ and $L \approx M/2$.

$$H(y) := \begin{bmatrix} y_0 & y_1 & \cdots & y_{M-1-L} \\ y_1 & y_2 & \cdots & y_{M-L} \\ \vdots & \vdots & & \vdots \\ y_L & y_{L+1} & \cdots & y_{M-1} \end{bmatrix}.$$

Fourier decomposition of Hankel matrix

$$H(\mathcal{F}_M \mu) = \Phi_L(X)\,\mathrm{diag}(a)\,\Phi_{M-1-L}(X)^t.$$

Which implies

$$H(y) = H(\mathcal{F}_M \mu) + H(\eta).$$

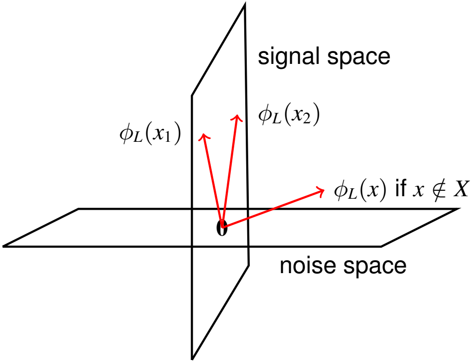
Subspace methods continued

Compute the best rank S approximation of $H(y)$ by SVD truncation:

$$\widetilde{U}\widetilde{\Sigma}\widetilde{V}^* = \text{best rank } S \text{ approximation of } H(y).$$

If the noise is sufficiently small,

$$\text{Range}(\widetilde{U}) \approx \text{Range}(\Phi_L(X))$$



Estimation of signal parameters via rotational invariance techniques

ESPRIT algorithm:

- **Noiseless case:**
  1. Compute a matrix U whose columns form an orthonormal basis for the range of $\Phi_M(X)$.
  2. Let U_0 and U_1 be the first and last L rows of U, respectively.
  3. The eigenvalues of $U_0^\dagger U_1$ are $\{e^{2\pi i x_j}\}_{j=1}^S$, from which we extract $\{x_j\}_{j=1}^S$.

- **Noise case:**
  1. Compute a matrix $\widetilde{U}$ whose columns form an orthonormal basis an approximation of the range of $\Phi_M(X)$.
  2. Let $\widetilde{U}_0$ and $\widetilde{U}_1$ be the first and last L rows of U, respectively.
  3. Calculate the eigenvalues of $\widetilde{U}_0^\dagger \widetilde{U}_1$, project them to the complex unit circle, and extract their arguments.

Estimation of signal parameters via rotational invariance techniques

ESPRIT algorithm:

- **Noiseless case:**
  1. Compute a matrix U whose columns form an orthonormal basis for the range of $\Phi_M(X)$.
  2. Let U_0 and U_1 be the first and last L rows of U, respectively.
  3. The eigenvalues of $U_0^\dagger U_1$ are $\{e^{2\pi i x_j}\}_{j=1}^S$, from which we extract $\{x_j\}_{j=1}^S$.

- **Noise case:**
  1. Compute a matrix $\widetilde{U}$ whose columns form an orthonormal basis an approximation of the range of $\Phi_M(X)$.
  2. Let $\widetilde{U}_0$ and $\widetilde{U}_1$ be the first and last L rows of U, respectively.
  3. Calculate the eigenvalues of $\widetilde{U}_0^\dagger \widetilde{U}_1$, project them to the complex unit circle, and extract their arguments.

Subspace methods require knowing S, or one estimates S through some other knowledge such as noise level. Recent work [P. Liu, H. Zhang 2021]

Theorem (L., W. Liao, A. Fannjiang, IEEE IT 2020)

For any S-atomic μ with support X and any η such that $\|H(\eta)\|_2$ is sufficiently small, if $\widetilde{X}$ is the output of ESPRIT, we have

$$\textit{md}(X, \widetilde{X}) \leq \underbrace{\frac{C(M,S)}{(\min_j |a_j|)\sigma_S^2(\Phi_L(X))}}_{\textit{numerical conditioning}} \|\eta\|_2.$$

Theorem (L., W. Liao, A. Fannjiang, IEEE IT 2020)

For any S-atomic μ with support X and any η such that $\|H(\eta)\|_2$ is sufficiently small, if $\widetilde{X}$ is the output of ESPRIT, we have

$$\textit{md}(X, \widetilde{X}) \leq \underbrace{\frac{C(M,S)}{(\min_j |a_j|)\sigma_S^2(\Phi_L(X))}}_{\text{numerical conditioning}} \|\eta\|_2.$$

Under the clumps model (under additional assumptions on X),

$$\textit{md}(X, \widetilde{X}) \leq \underbrace{\frac{C(R,M,S,\lambda)SRF^{2\lambda-2}}{\min_j |a_j|}}_{\text{numerical conditioning}} \|\eta\|_2.$$

Comparison: Previous best bounds for ESPRIT are

$O(\|\eta\|_2\sigma_S^{-5}(\Phi_{M/2}))$ [Aubel, Bölcskei, 2016],

$O(\|H(\eta)\|_2\sigma_S^{-4}(\Phi_{M/2}))$ [Fannjiang, 2016].

ESPRIT is near optimal: Upper bound is sharp in terms of SRF, and is one M factor away from min-max lower bound in [Batenkov, Goldman, Yomdin, 2020].

ESPRIT automatically adapts to the geometry: The algorithm does not require knowledge of the clump parameters!

Non-harmonic uncertainty principle

Concentration: Given a S-atomic measure μ and $L \geq S$, the quantity

$$C_L(\mu) := \frac{|\widehat{\mu}(0)|^2}{\sum_{m=0}^{L} |\widehat{\mu}(m)|^2}.$$

Uncertainty principle: If μ is S-atomic, expect its Fourier coefficients to not be perfectly localized.

Non-harmonic uncertainty principle

Concentration: Given a S-atomic measure μ and $L \geq S$, the quantity

$$C_L(\mu) := \frac{|\widehat{\mu}(0)|^2}{\sum_{m=0}^{L} |\widehat{\mu}(m)|^2}.$$

Uncertainty principle: If μ is S-atomic, expect its Fourier coefficients to not be perfectly localized.

Theorem (L., W. Liao, A. Fannjiang, IEEE IT 2021)

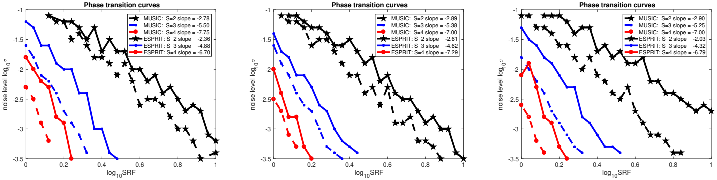
If $L \geq S > 1$, then

$$C_{L,S} := \sup_{\substack{\mu \text{ is } S\text{-atomic} \\ \mu \neq 0 \\ \mu \text{ complex}}} C_L(\mu) \leq 1 - 4^{-S},$$

$$C_{L,S,\mathbb{R}} := \sup_{\substack{\mu \text{ is } S\text{-atomic} \\ \mu \neq 0 \\ \mu \text{ real}}} C_L(\mu) \leq 1 - (8S-1)^{-1}.$$

Can be seen as a quantitative version of a result in [Donoho, Stark 1989].

Numerical simulations



(a) 1 clump (b) 2 clumps (c) 3 clumps

Figure: The phase transition curves below which the algorithm succeeds (defined to be successful if $\mathrm{md}(X,\widetilde{X}) < \Delta/2$) with probability at least 95% for $\lambda = 2, 3, 4$ with respect to $\log_{10}(SRF)$ (x-axis) and $\log_{10}\sigma$ (y-axis). The slopes are computed by least squares.

Numerical simulations continued

	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	Numerical	Theoretical
1-clump: MUSIC	2.78	5.50	7.75	$2.49\lambda - 2.11$	$2\lambda - 2$
2-clump: MUSIC	2.89	5.38	7.00	$2.06\lambda - 1.08$	$2\lambda - 2$
3-clump: MUSIC	2.90	5.25	7.00	$2.05\lambda - 1.10$	$2\lambda - 2$
4-clump: MUSIC	3.01	5.12	8.50	$2.75\lambda - 2.70$	$2\lambda - 2$
1-clump: ESPRIT	2.36	4.88	6.70	$2.17\lambda - 1.86$	$2\lambda - 2$
2-clump: ESPRIT	2.61	4.62	7.29	$2.34\lambda - 2.18$	$2\lambda - 2$
3-clump: ESPRIT	2.03	4.32	6.79	$2.38\lambda - 2.76$	$2\lambda - 2$
4-clump: ESPRIT	1.81	4.34	6.43	$2.31\lambda - 2.74$	$2\lambda - 2$

Table: Slopes extracted from the previous phase transition curves of MUSIC and ESPRIT.

Outline

Min-max error for sparse model on the grid

Grid model: Suppose $X \subseteq \{\frac{n}{N}\}_{n=0}^{N-1}$ for some large N. Every S-atomic μ can be identified with a S-sparse vector $u \in \mathbb{C}_S^N$.

Min-max error: Accuracy of the "best" possible algorithm (including those that have exponential run-time),

$$E(M,N,S,\delta) = \inf_{\substack{\phi:\, y \to \mathbb{C}_S^N \\ y = \mathcal{F}_M u + \eta}} \sup_{\substack{u \in \mathbb{C}_S^N :\, \|u\|_2 \le 1 \\ \eta \in \mathbb{C}^M :\, \|\eta\|_2 \le \delta}} \|\phi(u,\eta) - u\|_2.$$

No algorithm can beat the min-max error, and in particular,

ESPRIT error for this sparsity model $\geq E(M,N,S,\delta)$.

Sharp estimate on the min-max error

Theorem (L., W. Liao, ACHA 2021)

Let $S \geq 1$ and $M \geq 4S$. For N sufficiently large,

$$\frac{1}{\sqrt{M}} SRF^{2S-1}\delta \lesssim_{M,S} E(M,N,S,\delta) \lesssim_{M,S} \frac{1}{\sqrt{M}} SRF^{2S-1}\delta,$$

where $SRF = N/M$.

Sharp estimate on the min-max error

Theorem (L., W. Liao, ACHA 2021)
Let $S \geq 1$ and $M \geq 4S$. For N sufficiently large,

$$\frac{1}{\sqrt{M}} SRF^{2S-1}\delta \lesssim_{M,S} E(M,N,S,\delta) \lesssim_{M,S} \frac{1}{\sqrt{M}} SRF^{2S-1}\delta,$$

where $SRF = N/M$.

Remarks:

1. Non-asymptotic bounds for the min-max error with explicit constants.
2. Single clump with S points is the worst case scenario?
3. Proof relies on showing that $E(M,N,S,\delta)$ is related to $\sigma_S(\Phi_M(X \cup \widetilde{X}))$, where $X \cup \widetilde{X}$ is a set of cardinality at most $2S$.

Related work: [Donoho 1992], [Demanet, Nguyen 2015], [Batenkov, Goldman, Yomdin, 2020]

Outline

How to incorporate additional information?

Unknown: Time-dependent atomic measure, where the locations are fixed, but the amplitudes are time varying

$$\mu_t := \sum_{j=1}^{S} a_j(t)\delta_{x_j}, \quad a_j(t) \in \mathbb{C}, \quad x_j \in \mathbb{T} := [0, 1).$$

Known: Perturbed consecutive M Fourier coefficients at times $t_1, \ldots, t_n$,

$$y(t_\ell) := \mathcal{F}_M(\mu_{t_\ell}) + \eta(t_\ell), \quad \ell = 1, \ldots, n.$$

This is the **multiple-snapshot** problem, in contrast to the **single-snapshot** version earlier.

Multiple snapshot ESPRIT

ESPRIT readily extends to the multi-snapshot case:

1. Empirical covariance matrix

$$Y := \frac{1}{n}\sum_{\ell=1}^{n} y(t_\ell)y(t_\ell)^*.$$

2. $\widetilde{U}$ is the best rank S approximation of Y.
3. Find the eigenvalues of $\widetilde{U}_0^\dagger \widetilde{U}_1$, project to the unit complex circle, and extract their arguments.

What is the performance of ESPRIT

Is there any advantage of taking Fourier measurements over time?

1. Without any additional assumptions on $a(t)$ and $\eta(t)$,

   multi-snapshot error $=$ single-snapshot error.

What is the performance of ESPRIT

Is there any advantage of taking Fourier measurements over time?

1. Without any additional assumptions on $a(t)$ and $\eta(t)$,

$$\text{multi-snapshot error} = \text{single-snapshot error.}$$

2. Assume that $\eta(t_\ell)$ is complex i.i.d. Gaussian vector with covariance $\nu^2 I$. In expectation, we expect

$$\text{multi-snapshot error} \leq \frac{1}{\sqrt{n}} \cdot \text{single-snapshot error.}$$

What is the performance of ESPRIT

Is there any advantage of taking Fourier measurements over time?

1. Without any additional assumptions on $a(t)$ and $\eta(t)$,

$$\text{multi-snapshot error} = \text{single-snapshot error.}$$

2. Assume that $\eta(t_\ell)$ is complex i.i.d. Gaussian vector with covariance $\nu^2 I$. In expectation, we expect

$$\text{multi-snapshot error} \leq \frac{1}{\sqrt{n}} \cdot \text{single-snapshot error.}$$

This comes from just doing naive averaging:

$$\text{avg}(y(t_1), \ldots, y(t_n)) = \mathcal{F}_M\Big(\sum_{j=1}^{S} \text{avg}(a_j(t_1), \ldots, a_j(t_n))\delta_{x_j}\Big) + \underbrace{\frac{1}{n}\sum_{\ell=1}^{n} \eta(t_\ell)}_{\sqrt{n}\text{ cancellation}}.$$

Theorem (L., Z. Zhu, W. Gao, W. Liao, preprint 2021)

1. *Assume that each* $\eta(t_\ell) \sim CN(0, \nu^2 I)$ *(this can be relaxed significantly).*
2. *Assume that* $a(t_1), \dots, a(t_n)$ *span* $\mathbb{C}^S$*. Or equivalently the amplitude covariance matrix*

$$A = \frac{1}{n} \sum_{\ell=1}^{n} a(t_\ell) a(t_\ell)^*$$

   is strictly positive definite.

Theorem (L., Z. Zhu, W. Gao, W. Liao, preprint 2021)

1. *Assume that each* $\eta(t_\ell) \sim CN(0, \nu^2 I)$ *(this can be relaxed significantly).*
2. *Assume that* $a(t_1), \ldots, a(t_n)$ *span* $\mathbb{C}^S$*. Or equivalently the amplitude covariance matrix*

$$A = \frac{1}{n} \sum_{\ell=1}^{n} a(t_\ell) a(t_\ell)^*$$

   is strictly positive definite.

For n *sufficiently large, the output* $\widetilde{X}$ *of ESPRIT satisfies, with probability at least* $1 - \delta$*,*

$$md(X, \widetilde{X}) \le \underbrace{\frac{C(M,S)}{\sqrt{n}\sqrt{\lambda_S(A)}\sigma_S(\Phi_M)}}_{\text{numerical conditioning}} \nu \left(1 + \sqrt{\frac{\log(1/\delta)}{M}}\right).$$

Theorem (L., Z. Zhu, W. Gao, W. Liao, preprint 2021)

1. *Assume that each* $\eta(t_\ell) \sim CN(0, \nu^2 I)$ *(this can be relaxed significantly).*
2. *Assume that* $a(t_1), \dots, a(t_n)$ *span* $\mathbb{C}^S$*. Or equivalently the amplitude covariance matrix*

$$A = \frac{1}{n}\sum_{\ell=1}^{n} a(t_\ell)a(t_\ell)^*$$

   is strictly positive definite.

For n *sufficiently large, the output* $\widetilde{X}$ *of ESPRIT satisfies, with probability at least* $1-\delta$*,*

$$md(X,\widetilde{X}) \le \underbrace{\frac{C(M,S)}{\sqrt{n}\sqrt{\lambda_S(A)}\sigma_S(\Phi_M)}}_{\text{numerical conditioning}} \nu\left(1+\sqrt{\frac{\log(1/\delta)}{M}}\right).$$

Under the clumps model,

$$md(X,\widetilde{X}) \le \underbrace{\frac{C(M,S)SRF^{\lambda-1}}{\sqrt{n}\sqrt{\lambda_S(A)}}}_{\text{numerical conditioning}} \nu\left(1+\sqrt{\frac{\log(1/\delta)}{M}}\right).$$

Single and multi snapshot

- **Single snapshot:**

$$\text{ESPRIT error} \lesssim_{M,S,a} \frac{\text{noise}}{\sigma_S^2(\Phi_M)} \lesssim_{M,S,a} SRF^{2\lambda-2}\text{noise}.$$

Dependence on SRF and noise match the min-max rate, so ESPRIT is optimal in this sense.

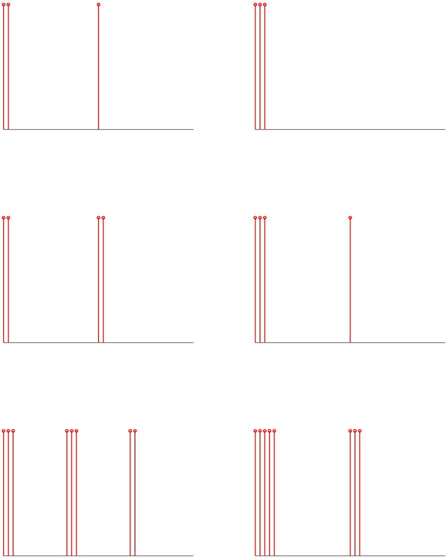
- **Multi snapshot:**

$$\text{ESPRIT error} \lesssim_{M,S,a} \frac{\text{noise}}{\sqrt{n}\sigma_S(\Phi_M)} \lesssim_{M,S,a} \frac{1}{\sqrt{n}} SRF^{\lambda-1}\text{noise}.$$

Dependence on SRF, number of snapshots n, and noise match a Cramer-Rao lower bound [L., Zhu, Gao, Liao 2021], so ESPRIT is optimal in this sense.

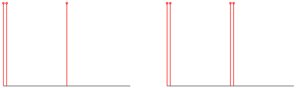
Outline

Difficulty of super-resolution

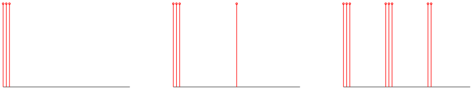


Difficulty of super-resolution

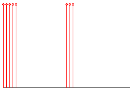
Order from least to most challenging in the limit $\Delta \to 0$:



(a) $O(SRF^2 \|\eta\|_2)$



(b) $O(SRF^4 \|\eta\|_2)$



(c) $O(SRF^8 \|\eta\|_2)$

Some remarks

- Classical work by Rayleigh and Abbe study the simplest scenario where there are only two point sources. Our results give a description for more complicated arrangements.

Some remarks

- Classical work by Rayleigh and Abbe study the simplest scenario where there are only two point sources. Our results give a description for more complicated arrangements.
- Our results show that super-resolution is possible in more situations than previously thought, since the error is $O(SRF^{2\lambda-2}\|\eta\|_2)$ as opposed to $O(SRF^{2S-2}\|\eta\|_2)$.

Some remarks

- Classical work by Rayleigh and Abbe study the simplest scenario where there are only two point sources. Our results give a description for more complicated arrangements.
- Our results show that super-resolution is possible in more situations than previously thought, since the error is $O(SRF^{2\lambda-2}\|\eta\|_2)$ as opposed to $O(SRF^{2S-2}\|\eta\|_2)$.
- ESPRIT is provably optimal (in terms of SRF), so criticism about "instability of subspace methods" is not fair.

Some remarks

- Classical work by Rayleigh and Abbe study the simplest scenario where there are only two point sources. Our results give a description for more complicated arrangements.
- Our results show that super-resolution is possible in more situations than previously thought, since the error is $O(SRF^{2\lambda-2}\|\eta\|_2)$ as opposed to $O(SRF^{2S-2}\|\eta\|_2)$.
- ESPRIT is provably optimal (in terms of SRF), so criticism about "instability of subspace methods" is not fair.
- Since $\sigma_S(\Phi_M(X)) > 0$ for any set X, an error bound of the form

  $$\text{error} \lesssim \frac{\text{noise}}{\sigma_S^2(\Phi_M(X))}$$

  means that (at least for point sources and Fourier measurements) separation is never the real bottleneck of super-resolution and that noise is the real culprit.

References

- L., and Wenjing Liao. Stable super-resolution limit and smallest singular value of restricted Fourier matrices. *Applied and Computational Harmonic Analysis*, 2021.
- L., Wenjing Liao, and Albert Fannjiang. Super-resolution limit of the ESPRIT algorithm. *IEEE Transactions on Information Theory*, 2020.
- L., Zengying Zhu, Weiguo Gao, and Wenjing Liao. Stability and super-resolution of MUSIC and ESPRIT for multi-snapshot spectral estimation. arXiv:2105.14304.

Thank you!