FFT-ESPRIT: a kernel-based subspace estimator for frequency super-resolution at quasi-linear time complexity

Shawn L. Kiser [1], Pierre Margerit [1], Marc Rébillat [1], Mikhail Guskov [1], and Nicolas Ranc [1]

[1]Affiliation not available

October 30, 2023

Abstract

We introduce two FFT-based ESPRIT algorithms for line spectral estimation which have lower time complexities than the original ESPRIT algorithm's O(N^3). The preferred method, named FFT-ESPRIT, can be characterized as being a kernel-based subspace estimator that achieves super-resolution at O(N log N) for frequency estimates. First, we demonstrate two estimations of the signal subspace via an integral transformation on the row space of the data matrix and the data matrix itself. The subspace-based methods are approximate in nature, and yet perturbation bounds reveal a noise regime in which FFT-ESPRIT exceeds ESPRIT's performance. We demonstrate the behavior of the algorithm across different SNR regimes and show that the estimated signal subspace is statistically efficient. Numerical simulations show that FFT-ESPRIT is more robust than the ESPRIT algorithm at the very low SNRs, and has a nearly identical performance as ESPRIT at higher SNRs.

FFT-ESPRIT: a kernel-based subspace estimator for frequency super-resolution at quasi-linear time complexity

PREPRINT – OCTOBER 2, 2022

Shawn L. Kiser[*,†]
Laboratoire PIMM
Arts et Métiers Sciences et Technologies
shawn_lee.kiser@ensam.eu

Pierre Margerit[*]
Laboratoire PIMM
Arts et Métiers Sciences et Technologies
pierre.margerit@ensam.eu

Marc Rébillat[*]
Laboratoire PIMM
Arts et Métiers Sciences et Technologies
marc.rebillat@ensam.eu

Mikhail Guskov[*]
Laboratoire PIMM
Arts et Métiers Sciences et Technologies
mikhail.guskov@ensam.eu

Nicolas Ranc[*]
Laboratoire PIMM
Arts et Métiers Sciences et Technologies
nicolas.ranc@ensam.eu

ABSTRACT

We introduce two FFT-based ESPRIT algorithms for line spectral estimation which have lower time complexities than the original ESPRIT algorithm's $\mathcal{O}(N^3)$. The preferred method, named FFT-ESPRIT, can be characterized as being a kernel-based subspace estimator that achieves super-resolution at $\mathcal{O}(N \log N)$ for frequency estimates. First, we demonstrate two estimations of the signal subspace via an integral transformation on the row space of the data matrix and the data matrix itself. The subspace-based methods are approximate in nature, and yet perturbation bounds reveal a noise regime in which FFT-ESPRIT exceeds ESPRIT's performance. We demonstrate the behavior of the algorithm across different SNR regimes and show that the estimated signal subspace is statistically efficient. Numerical simulations show that FFT-ESPRIT is more robust than the ESPRIT algorithm at the very low SNRs, and has a nearly identical performance as ESPRIT at higher SNRs.

Keywords Frequency estimation · Subspace methods · Fast Fourier Transform · Super-resolution · Kernel function

1 Introduction

The problem of line spectral estimation (LSE), i.e. frequency estimation, is a fundamental research domain in signal processing. Essential engineering problems have been adapted to the LSE problem including direction of arrival (DOA) of automotive radar [1], spectroscopy [2], simulations of molecular dynamics [3], and vibrational eigenmode analysis [4]. LSE is seen as the estimation of frequencies for P complex sinusoids of a signal $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$:

$$\widehat{x}[n] = \sum_{i=0}^{P-1} \beta_i e^{\mathrm{j}2\pi\omega_i[n]} + \varepsilon[n] = x[n] + \varepsilon[n] \tag{1}$$

where $n \in \mathbb{Z}$ is a discrete index and $\varepsilon[n]$ represents additive white Gaussian noise (AWGN) with variance η. The complex amplitude β_i is treated as a nuisance parameter, and $\{\omega_i \in [0,1); \omega_i \neq \omega_j : \forall i \neq j\}$ is the normalized frequency whose set is distinct. In matrix-vector notation, Eq. (1) is:

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

[*]CNRS, CNAM, HESAM Université
[†] Corresponding author

where $\boldsymbol{A} = [\boldsymbol{a}_{\omega_0} \cdots \boldsymbol{a}_{\omega_{P-1}}] \in \mathbb{C}^{N\times P}$ is a Vandermonde matrix with columns $\boldsymbol{a}_{\omega_i} = \begin{bmatrix}1 & e^{\mathrm{j}2\pi\omega_i} & \cdots & e^{\mathrm{j}2\pi(N-1)\omega_i}\end{bmatrix}^{\mathrm{T}}$, and the signal-to-noise ratio (SNR) defined as $\mathrm{SNR} \triangleq \|\boldsymbol{A}\boldsymbol{\beta}\|^2/\eta$. This model of the signal, originally used by Prony's method [5], has led to a multitude of parametric estimation techniques in the past half-century.

Many LSEs make use of the discrete Fourier transform (DFT), e.g. [6, 7], where they use a combination of interpolation, a peak picking algorithm, and iterative refinement to determine the spectral line from the frequency spectrum. Despite their low computational complexity via the usage of the Fast Fourier Transform (FFT) algorithm, any two spectral lines can only be resolved if they are separated by more than the Rayleigh limit [8]: this corresponds to $1/N$ for the discrete signal. Recently, the introduction of compressed sensing [9] and sparse regularization cast the line spectral problem of the signal from an infinite dimensional basis of sinusoids. Some notable algorithms include the usage of the atomic norm [10], a "newtonized" orthogonal matching pursuit [11], and a low complexity Bayesian method [12]. Sufficient recovery via convex optimization for exactly two spectral lines has been shown to require a separation of at least $\frac{4}{(N-1)}$ [13], and for $P \geq 2$ spectral lines a separation of at least $4.1 \exp\left(\frac{1}{\mathrm{SNR}}\right)^{\frac{1}{2N-2}}$ [14].

Alternatively, subspace methods are capable of resolving spectral lines beyond the Rayleigh limit, which include the notable Multiple Signal Classification (MUSIC) [15] and Estimation of Signal Parameters via Rotation Invariance Techniques (ESPRIT) [16] algorithms. Subspace methods decompose the finite data into signal and orthogonal (noise) subspaces via a singular-value decomposition (SVD) on a Hankel data matrix, or eigenvalue decomposition (EVD) on the covariance matrix. They do not require convex optimization and retrieve exactly the frequencies in the noiseless case. However, these algorithms are bounded by a computational complexity of $\mathcal{O}(N^3)$. Both these methods have been studied for their statistical efficiency, showing that they achieve the maximum likelihood (ML) estimate when $N \to \infty$ and/or $\mathrm{SNR} \to \infty$ [17]. More recently, ESPRIT and MUSIC have had their computational super-resolution limit defined [18, 19]. Coupled with the fact that MUSIC can only provide on-grid frequency estimates, these studies have shown that ESPRIT is more optimal than MUSIC for the line spectral problem.

For finite N and/or SNR degeneration, subspace methods have performance breakdowns below a threshold when compared to the Cramér-Rao lower bound (CRB) [20]. It's understood that beyond this threshold, often called the no-information SNR regime, the estimates of signal and orthogonal subspaces begin to swap a small portion of their relevant eigenvectors, i.e. subspace swap. A first approach for mitigating this breakdown was to introduce a pseudo-noise resampling technique to perturb the orthogonal subspace favorably. This combined with a bootstrap- or jackknife-like resampling scheme has led to a better threshold performance for subspace methods, e.g. for the Unitary ESPRIT algorithm [21]. An alternative approach uses an estimate of the lower bound of the probability of a subspace swap [22]. Given a high probability, the (non-convex) ML cost function is optimized using the subspace method's estimated frequencies. If no significant change exists between the refined estimates, the lack of refinement substantiates the lacking of subspace swap. These have led to more developed, but computationally taxing, algorithms, e.g. [23] which iteratively refines the covariance matrix using repeated ESPRIT DOA estimates.

In practical applications, it is common to estimate LSEs to quasi-stationary signals despite the basis mismatch. It has been shown by benchmarks in the literature, e.g. vibration analysis [24] and power system synchrophasors [25], that ESPRIT exceeded other state-of-the-art alternatives in performance metrics but at a much higher computational cost. Thus, for problems requiring real-time usage, subspace methods would benefit from fast estimations of the signal subspace. Multiple approaches in the literature leverage the fact that the span of the signal subspace is much smaller than its orthogonal subspace. In [26], they use the Power method coupled with the Lanczos algorithm to iteratively converge to the signal subspace. The work by [27] assumed that the signal subspace is an autoregressive process whose polynomials and rational functions form the sample covariance. Alternatively, the signal subspace is approximated in [28] by using the DFT and the discrete cosine transforms for use in the MUSIC algorithm. Lastly, in [29] the periodicity of an ideally long signal is estimated for, such that the estimated covariance matrix is approximately circulant, and thus the eigenvectors can be computed through an FFT, i.e. the subspace.

In this study, we introduce the FFT-ESPRIT algorithm for the line spectral problem, which aims to achieve super-resolution without an SVD on the $M \times L$-sized data matrix. This is in contrast to ESPRIT with partial SVDs and fast Hankel matrix-vector products presented in [30] which achieves time complexity $\mathcal{O}(MN \log N + M^3)$; or Nyström-based ESPRIT [31] which achieves time complexity $\mathcal{O}(MNK + MK^2)$ for $K \leq \min(M, N)$, but requires tuning of their sub-vector length K and does not approach the asymptotic behavior ESPRIT. The main novelty of FFT-ESPRIT resides in the signal subspace estimation strategy by efficient use of the kernel of the DFT matrix. We build off the work of [28], where a DFT-based signal subspace approximation is described, but its complexity suffers from searching the on-grid pseudo-spectrum of MUSIC and lacks a theoretical analysis. We go one step further by utilizing an iterative interpolated DFT algorithm (IIp-DFT) [7], generalizing the signal subspace estimate through analysis on the eigenspace

perturbation, and using ESPRIT for off-grid frequency estimates. Lastly, we give a computationally efficient version which leverages the FFT and achieves a quasi-linear time complexity with respect to its signal length.

Outline In Section 2 we present some preliminaries, namely the original ESPRIT algorithm and the definitions of power spectral density. In Section 3 we introduce our signal subspace estimation of FFT-ESPRIT via kernel transformation and establish the signal subspace estimation is statistically efficient for the asymptotic case. Based on subspace perturbations, we also show that FFT-ESPRIT can be improved. We then derive a low-complexity version and mention the numerical aspects to be considered when implemented. Numerical experiments are presented in Section 4 and conclusions are given in Section 5.

Notations We denote $\boldsymbol{y}$ and $\boldsymbol{Y}$ as vectors and matrices respectively. The ith entry of the vector $\boldsymbol{y}$ is denoted as $y[i]$. The ith column and row vector of $\boldsymbol{Y}$ is $\boldsymbol{y}_i$ and $\vec{\boldsymbol{y}}_i$ respectively. Round parentheses of $y(i)$ denotes $i \in \mathbb{R}$ whereas square parentheses $y[i]$ denotes $i \in \mathbb{Z}$. The complex conjugate, Hermitian transpose, and Moore-Penrose pseudo-inverse of $\boldsymbol{Y} \in \mathbb{C}$ is denoted as $\boldsymbol{Y}^*$, $\boldsymbol{Y}^{\mathrm{H}}$, and $\boldsymbol{Y}^{\ddagger}$ respectively. The inner product for $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{C}$ is defined as $\boldsymbol{y}^{\mathrm{H}}\boldsymbol{z}$. A noise perturbed y is denoted $\widehat{y}$. The expected value of y is denoted $\mathbb{E}\{y\}$. The variance of the white noise is denoted η. The ℓ_2 and Frobenius norms are denoted $\|\boldsymbol{Y}\|$ and $\|\boldsymbol{Y}\|_{\mathrm{F}}$ respectively. The Hadamard product between $\boldsymbol{y}, \boldsymbol{z}$ is denoted $\boldsymbol{y} \circ \boldsymbol{z}$.

2 Preliminaries

The signal subspace estimation and approach we present in the following are a combination of classical parametric and non-parametric estimators in the literature. We have curated these ideas introduced in this section to obtain an ESPRIT-based LSE that can be implemented with lower complexity. The signal's model is assumed to be of Eq. (1), is second-order stationary $\mathbb{E}\{\boldsymbol{x}\} = 0$, and the model order P is known a priori.

2.1 Original ESPRIT algorithm

The original ESPRIT algorithm can be described with a data matrix:

$$\boldsymbol{X} = \begin{bmatrix} x[n] & x[n+1] & \cdots & x[n+L] \\ x[n+1] & x[n+2] & \cdots & x[n+L+1] \\ \vdots & \vdots & \ddots & \vdots \\ x[n+M-1] & x[n+M] & \cdots & x[n+N] \end{bmatrix} \tag{3}$$

where $\boldsymbol{X} \in \mathbb{C}^{M \times L}$ is a Hankel matrix, M is a chosen sub-vector size such that:

$$P < M < N + 1 - P; \quad 2P \leq N + 1; \quad M \leq N/2 \tag{4}$$

Thus the remaining length is $L = N - M + 1$. The ML form of the sample covariance $\tilde{\boldsymbol{R}} \in \mathbb{C}^{M \times M}$ is the statistically efficient approximation of the true covariance $\boldsymbol{R}$:

$$\tilde{\boldsymbol{R}} = \frac{1}{L}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} = \frac{1}{L}\sum_{n=0}^{L-1} \boldsymbol{x}_n \boldsymbol{x}_n^{\mathrm{H}} \tag{5}$$

If the SVD of $\boldsymbol{X}/\sqrt{L}$ is $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$, then the EVD on the covariance yields:

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\mathrm{H}} \tag{6}$$

where $\boldsymbol{\Sigma}$ corresponds to a diagonal matrix that contains the eigenvalues on the main diagonal, sorted in descending order, and $\boldsymbol{U}$ the (left) subspace. For two sample-shifted overlapping subspaces:

$$\begin{aligned} \boldsymbol{U}^{\uparrow} &= [\boldsymbol{I}_{M-1} \quad \boldsymbol{0}]\,\boldsymbol{U} = \boldsymbol{\Gamma}^{\uparrow}\boldsymbol{U} \\ \boldsymbol{U}_{\downarrow} &= [\boldsymbol{0} \quad \boldsymbol{I}_{M-1}]\,\boldsymbol{U} = \boldsymbol{\Gamma}_{\downarrow}\boldsymbol{U} \end{aligned} \tag{7}$$

one can show the rotational invariance property:

$$\boldsymbol{\Gamma}^{\uparrow}\boldsymbol{U}\boldsymbol{D} = \boldsymbol{\Gamma}_{\downarrow}\boldsymbol{U} \tag{8}$$

where matrix $\boldsymbol{D} = \mathrm{diag}\{e^{\mathrm{j}2\pi\omega_0}, \cdots, e^{\mathrm{j}2\pi\omega_{P-1}}\}$ and $\boldsymbol{I}_{M-1}$ is an identity matrix of size $M-1$. There exists a unitary matrix $\boldsymbol{F}$ [16] such that:

$$\boldsymbol{\Gamma}^{\uparrow}\boldsymbol{U}\boldsymbol{F}\boldsymbol{D} = \boldsymbol{\Gamma}_{\downarrow}\boldsymbol{U}\boldsymbol{F} \tag{9}$$

and since $\boldsymbol{U}$ is orthonormal, i.e. $\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{I}$:

$$\boldsymbol{\Gamma}^{\uparrow}\boldsymbol{U}(\boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^{\mathrm{H}}) = \boldsymbol{\Gamma}_{\downarrow}\boldsymbol{U} \tag{10}$$

which admits the eigenvalue problem $\boldsymbol{\Psi} = \boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^{\mathrm{H}}$. Thus, the ESPRIT algorithm [16] solves in the least-squares sense for $\boldsymbol{\Psi}$ where its eigenvalues are projected onto the complex unit circle to give the frequencies of $\boldsymbol{x}$:

$$\boldsymbol{\Psi} = \left(\widehat{\boldsymbol{U}}^{\uparrow}\right)^{\ddagger}\boldsymbol{U}_{\downarrow} \tag{11}$$

if and only if the signal subspace is used for the left P column vectors of $\boldsymbol{U} = [\boldsymbol{u}_0 \quad \cdots \quad \boldsymbol{u}_{P-1}]$ of Eq. (7).

2.2 Power spectral analysis

The probabilistic view of signals has led to the first non-parametric spectral analysis which makes use of Parseval's theorem [32]. The power spectral density (PSD) is defined as the discrete-time Fourier Transform (DTFT) of the (auto)covariance sequence $r(n) = x(n) \circledast x(-n)$:

$$\phi(\omega) = \sum_{n=-\infty}^{\infty} r(n)e^{-\mathrm{j}2\pi\omega(n)} \tag{12}$$

A discrete signal only has a finite average power, so the periodogram offers an approximation:

$$\phi_{\mathrm{p}}(\omega) \coloneqq \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-\mathrm{j}2\pi\omega[n]}\right|^2 \tag{13}$$

which is simply the squared magnitude of the DFT. It's well known that the periodogram has a non-zero variance asymptotically and its resolution is the Rayleigh limit $1/N$. It can be shown asymptotically for two infinitesimally close frequencies, the periodogram of those estimates is uncorrelated:

$$\lim_{N\to\infty} \mathbb{E}\left\{\left(\phi_{\mathrm{p}}\left(\omega_i\right) - \phi\left(\omega_i\right)\right)\left(\phi_{\mathrm{p}}\left(\omega_j\right) - \phi\left(\omega_j\right)\right)\right\} = \begin{cases} \phi^2\left(\omega_1\right), & \omega_i = \omega_j \\ 0, & \omega_i \neq \omega_j \end{cases}$$

In the finite case, the covariance of two frequencies begins to become negligible if $|\omega_i - \omega_j| > 2/N$ [32], which has led to methods of decorrelating estimates through smoothing in the frequency domain.

3 FFT-ESPRIT

In this section, we introduce the FFT-ESPRIT algorithm and contrast it with the original ESPRIT. First, we detail the problem of estimating the signal subspace and give insight into the statistical and perturbation behavior. We go one step further by using a priori information suspected in the signal subspace to generate a truncated DFT matrix. Lastly, an efficient implementation of FFT-ESPRIT is given that obtains $\mathcal{O}(N \log N)$.

3.1 Signal subspace estimation via kernels

Let the ith eigenvalue problem corresponding to the M descending eigenvalues $\{\lambda_0 \geq \cdots \geq \lambda_{M-1}\}$ be:

$$\boldsymbol{R}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i \tag{14}$$

Substitution of the sample covariance $\tilde{\boldsymbol{R}}$ of Eq. (5) into Eq. (14) yields:

$$\frac{1}{L}\sum_{n=0}^{L-1}\left(\boldsymbol{x}_n^{\mathrm{H}}\boldsymbol{v}_i\boldsymbol{x}_n\right) \approx \lambda_i\boldsymbol{v}_i \tag{15}$$

Eq. (15) shows that the ith unit eigenvector $\boldsymbol{v}_i$ is approximately a linear combination of $\boldsymbol{x}_n$. The linear combination of $\boldsymbol{x}_n$ has a maximum norm of λ_1 which corresponds to the first eigenvector $\boldsymbol{v}_1$. The next linear combination with the maximum norm λ_2 corresponds to $\boldsymbol{v}_2$ which is orthogonal to $\boldsymbol{v}_1$, etc. To avoid computing Eq. (14), we seek an estimate of $\boldsymbol{x}_n^{\mathrm{H}}\boldsymbol{v}_i$ that conserves the eigenvectors of the problem by a set of L vectorized kernels $\{\|\boldsymbol{\kappa}_i\| = c\,;\, \boldsymbol{\kappa}_i^{\mathrm{H}}\boldsymbol{\kappa}_j = 0 : \forall i \neq j\}$:

$$\operatorname{span}\{c\lambda_0\boldsymbol{v}_0, \cdots, c\lambda_{M-1}\boldsymbol{v}_{M-1}\} \subseteq \operatorname{span}\left\{\sum_{n=0}^{L-1}\left(\kappa_0[n]\,\boldsymbol{x}_n\right), \cdots, \sum_{n=0}^{L-1}\left(\kappa_{M-1}[n]\,\boldsymbol{x}_n\right)\right\} \tag{16}$$

i.e. the linear transformation that yields the eigenvector and eigenvalue must correspond to the product between the kernel and the column space of the data matrix. Let $\boldsymbol{\mathcal{K}}$ be the matrix operator associated with the kernel of Eq. (16), where the QR decomposition of their matrix product is:

$$\boldsymbol{\mathcal{K}}\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\mathfrak{R}} \tag{17}$$

ESPRIT is invariant to the choice of orthonormal basis, which can be seen from the following:

Proposition 1. *Let $\boldsymbol{Y}_P \in \mathbb{C}^{L\times P}$ of rank P where $L \geq P$. Let its QR decomposition give:*

$$\boldsymbol{Y}_P = \boldsymbol{Q}\boldsymbol{\mathfrak{R}} \tag{18}$$

where $\boldsymbol{Q} \in \mathbb{C}^{L\times P}$, $\boldsymbol{\mathfrak{R}} \in \mathbb{C}^{P\times P}$, and $\boldsymbol{Q}$ is a unitary matrix whose columns are sorted descending by eigenvalues. Then it's rudimentary to see the singular values for the upper triangular matrix $\boldsymbol{\mathfrak{R}}$ are the same as the singular values of $\boldsymbol{Y}_P$, i.e. $\{\sigma_0(\boldsymbol{\mathfrak{R}}), \cdots, \sigma_{P-1}(\boldsymbol{\mathfrak{R}})\} = \{\sigma_0(\boldsymbol{Y}_P), \cdots, \sigma_{P-1}(\boldsymbol{Y}_P)\}$.

Therefore, the orthonormal basis of $\boldsymbol{Q}$ is used instead of the SVD's $\boldsymbol{U}$, which has the eigenvalue decomposition whose M largest $\|\boldsymbol{q}_i\|$ yield the eigenvalues:

$$\boldsymbol{Q} = \boldsymbol{Y}\boldsymbol{\Lambda}\boldsymbol{Y}^{\mathrm{H}} \approx c\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{H}} \tag{19}$$

where $\boldsymbol{Y}$ is some arbitrary eigenvector basis, $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_0, \cdots\ \lambda_{M-1}\}$, and $\boldsymbol{V}$ is the eigenvector basis formed by $\mathrm{span}\{\lambda_0\boldsymbol{v}_1, \cdots, \lambda_{M-1}\boldsymbol{v}_{M-1}\}$. One may realize that the problem of finding the appropriate set of kernels $\{\boldsymbol{\kappa}_i\}$ leads to a computationally intractable M dimensional optimization. An insight to avoid this lies in the fact that the Toeplitz equivalent[2] of the sample covariance $\tilde{\boldsymbol{R}}$ can be shown to approach the eigenvalues of the PSD by application of Szegö's theorem [34].

Theorem 1. *Let $\boldsymbol{T}_N$ be the Toeplitz equivalent of $\tilde{\boldsymbol{R}}$ defined in Eq.* (5) *of the same size. In the asymptotic case:*

$$\lim_{N\to\infty} \mathbb{E}\{\tilde{\boldsymbol{R}}\} = \boldsymbol{T}_N$$

Proof. See [35].

Thus, an asymptotic relationship exists between the ML sample covariance and its Toeplitz equivalent, which allows use of Szegö's theorem:

Theorem 2. *Let $\boldsymbol{T}_N$ be any mth order Toeplitz matrix with descending eigenvalues $\lambda_{N,k}$:*

$$\boldsymbol{T}_N = \begin{bmatrix}
t[0] & t[-1] & t[-2] & \cdots & t[-m] & & & & & \\
t[1] & t[0] & t[-1] & t[-2] & \cdots & t[-m] & & & & \\
\vdots & & & \ddots & & & & & & \\
t[m] & t[m-1] & t[m-2] & \cdots & t[0] & t[-1] & \cdots & t[-m] & & \\
 & t[m] & & \cdots & t[1] & t[0] & t[-1] & \cdots & t[-m] & \\
 & & \ddots & & & \ddots & & & \ddots & \\
 & & & t[m] & \cdots & t[1] & t[0] & t[-1] & \cdots & t[-m] \\
 & & & & \ddots & & & & & \vdots \\
 & & & & & t[m] & \cdots & t[1] & t[0] & t[-1] \\
 & & & & & & t[m] & \cdots & t[1] & t[0]
\end{bmatrix}$$

Let $\boldsymbol{C}_N$ with entries $c[n]$ be a circulant matrix with a proper circulant structure, i.e. when every row of $\boldsymbol{T}_N$ is a right cyclic shift of the row above it such that $t[n] = t[i-n]$ for $i = 1, 2, \cdots m-1$. $\boldsymbol{T}_N$ and $\boldsymbol{C}_N$ are defined as asymptotically equivalent, if and only if, for some finite bound ϵ:

$$\|\boldsymbol{T}_N\| \leq \epsilon < \infty, \quad \|\boldsymbol{C}_N\| \leq \epsilon < \infty$$

and:

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \|\boldsymbol{T}_N - \boldsymbol{C}_N\|_{\mathrm{F}} = 0$$

Let the asymptotic equivalent circulant matrix $\boldsymbol{C}$ have entries $c[n]$. The eigenvalues of $\boldsymbol{C}$ are:

$$\mu_{N,k} = \sum_{n=0}^{N-1} c[n] e^{-\mathrm{j}2\pi k[n]/N}$$

[2]The true covariance matrix is Hermitian and Toeplitz for a stationary process. In practice, the sample covariance of Eq. (5) is rarely Toeplitz. Different "Toeplitzations" of Eq. (5) have been explored in [33] with respect to the true covariance in the Frobenius norm sense.

where $\mu_{N,k}$ indicates the kth eigenvalue for matrix size N. In the asymptotic case, the so-called eigenvalue moment exists such that they are equal:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\mu_{N,k}=\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\lambda_{N,k}$$

Then Szegö's theorem [34] gives for any function g continuous on the range of ϕ:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}g(\lambda_{N,k})=\frac{1}{2\pi}\int_0^{2\pi}g\left(\phi(\tau)\right)d\tau \tag{20}$$

where $\phi(\tau)$ is the definition of the PSD Eq. (12) *with $\tau=2\pi k/N$.*

Proof. See [36].

In essence, one would choose $g(x)=x$ which would result in Eq. (20) becoming:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\lambda_{N,k}=\frac{1}{2\pi}\int_0^{2\pi}\phi(\tau)d\tau \tag{21}$$

Eq. (21) shows the eigenvalues of $\boldsymbol{T}_N$ can be expected to have the same distribution as the PSD, and when N is large, the eigenvalues are close to the values of the PSD. By extension of Theorem 1, this is true for the sample covariance $\tilde{\boldsymbol{R}}$. We demonstrate this between the true covariance and the circulant (PSD) approximation for the distribution of eigenvalues for a finite signal in Appendix A.

Thus, it follows that the DFT used inside Eq. (13) can reproduce the kernel of Eq. (16) since it respects the eigenvalue distribution of Eq. (5) and the Hankel property of $\boldsymbol{X}$. We demonstrate this relation by rewriting Eq. (16):

$$\begin{aligned}\operatorname{span}\{c\lambda_0\boldsymbol{v}_0,\,\cdots,\,c\lambda_{M-1}\boldsymbol{v}_{M-1}\}&\subseteq\operatorname{span}\left\{\sum_{n=0}^{L-1}\left(\kappa_0[n]\,\boldsymbol{x}_n\right),\,\cdots,\,\sum_{n=0}^{L-1}\left(\kappa_{M-1}[n]\,\boldsymbol{x}_n\right)\right\}\\&\subseteq\operatorname{span}\left\{\sum_{n=0}^{L-1}\left(e^{-\mathrm{j}2\pi[0][n]/L}\boldsymbol{x}_n\right),\,\cdots,\,\sum_{n=0}^{L-1}\left(e^{-\mathrm{j}2\pi[M-1][n]/L}\,\boldsymbol{x}_n\right)\right\}\end{aligned} \tag{22}$$

Thus in matrix-vector form:

$$\operatorname{span}\left\{\sum_{n=0}^{L-1}\left(e^{-\mathrm{j}2\pi[0][n]/L}\boldsymbol{x}_n\right),\,\cdots,\,\sum_{n=0}^{L-1}\left(e^{-\mathrm{j}2\pi[M-1][n]/L}\,\boldsymbol{x}_n\right)\right\}\approx\operatorname{span}\left\{\boldsymbol{W}_L\vec{\boldsymbol{x}}_0,\,\cdots,\,\boldsymbol{W}_L\vec{\boldsymbol{x}}_{L-1}\right\} \tag{23}$$

where $\boldsymbol{W}_L$ is an $L\times L$ DFT matrix and $\vec{\boldsymbol{x}}_n$ is the nth row vector. Therefore, a kernel which reproduces the product $\boldsymbol{x}_n^{\mathrm{H}}\boldsymbol{v}_i$ with $\boldsymbol{x}_n$ can be approximated by the kernel of the (integral) Fourier transform on the row-space of the Hankel matrix $\boldsymbol{X}$. One may realize that the form of Eq. (23) is the same as the feature mapping procedure used in the kernel principal component analysis (PCA) if the covariance matrix is used instead.

We show this kernel-based subspace estimator implemented in Algorithm 1, using MATLAB functions. One may note that the balance between resolution and the (PSD) approximation of the eigenvalues by the DFT matrix is maximized when $M=L$, i.e. it follows that the performance of this Algorithm 1 is tied to this constraint on $\boldsymbol{X}$ being square. This is similarly shown for SVD-based methods by [37] and proven for ESPRIT by [18] where better performance is achieved. Note that Algorithm 1 has an asymptotic complexity of $\mathcal{O}(N^2\log N)$ due to lines 2-4.

Algorithm 1 ESPRIT with subspace approximation by DFT kernel

Input: $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$, M, P
Output: $\boldsymbol{\omega}$
$\boldsymbol{X} \leftarrow (\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{L-1})$
for $i = 0$ to $M-1$ **do**
  $\boldsymbol{y}_i \leftarrow \texttt{fft}(\vec{\boldsymbol{x}}_i)$
end for
$\boldsymbol{Y} \leftarrow \texttt{sortMax}(\|\boldsymbol{y}_0\|, \cdots, \|\boldsymbol{y}_{M-1}\|)$
$\boldsymbol{Y}_P \leftarrow (\boldsymbol{y}_0, \cdots, \boldsymbol{y}_{P-1})$
$\boldsymbol{Q} \leftarrow \texttt{qr}(\boldsymbol{Y}_P, \texttt{'econ'})$
$\boldsymbol{Q}^{\uparrow} \leftarrow (\vec{\boldsymbol{q}}_0; \cdots; \vec{\boldsymbol{q}}_{M-2})$
$\boldsymbol{Q}_{\downarrow} \leftarrow (\vec{\boldsymbol{q}}_1; \cdots; \vec{\boldsymbol{q}}_{M-1})$
$\boldsymbol{\Psi} \leftarrow \left(\boldsymbol{Q}^{\uparrow}\right)^{\ddagger}\boldsymbol{Q}_{\downarrow}$
$\boldsymbol{\omega} \leftarrow \texttt{mod}(\texttt{angle}(\texttt{eig}(\boldsymbol{\Psi})/2\pi, 1))$
return $\boldsymbol{\omega}$

From Algorithm 1, it can be seen that the calculation of $\boldsymbol{\Psi}$ from the basis spanned by $\boldsymbol{Q}$ contains the linear combinations of $\boldsymbol{A\beta}$. The resulting eigenvalue decomposition admits the dominant ω_i, with the other smaller signal components perturbing the final eigenvalue. To gauge this effect, one can invoke the theorems of Davis-Kahan and Wedin [38] from ℓ_2 eigenspace perturbation theory. Let the ith principle angle between two subspace matrices $\boldsymbol{S}$, $\widehat{\boldsymbol{S}}$ with dimension $M \times L$ of be defined as:

$$\theta_i \coloneqq \arccos(|\widehat{\boldsymbol{s}}_i^{\mathrm{H}}\boldsymbol{s}_i|) = \arccos(\sigma_i) \tag{24}$$

where $\{\sigma_i\}$ denotes the ith descending singular value $\{\sigma_0 \geq \cdots \geq \sigma_i \geq \cdots \geq \sigma_{M-1}\}$ and let:

$$\boldsymbol{\Theta}\big(\widehat{\boldsymbol{S}}, \boldsymbol{S}\big) \coloneqq \big\{\theta_0, \cdots, \theta_{M-1}\big\} \tag{25}$$

Theorem 3. *Let $\boldsymbol{X}, \widehat{\boldsymbol{X}} \in \mathbb{C}^{M\times L}$, where $\widehat{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{\mathcal{E}}$ and has rank P. Let their SVDs be:*

$$\boldsymbol{X} = \sum_{i=0}^{\min\{M-1,L-1\}} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{H}} = \begin{bmatrix}\boldsymbol{U} & \boldsymbol{U}_{\perp}\end{bmatrix} \begin{bmatrix}\boldsymbol{\Sigma} & \\ & \boldsymbol{\Sigma}_{\perp}\end{bmatrix} \begin{bmatrix}\boldsymbol{V}^{\mathrm{H}} \\ \boldsymbol{V}_{\perp}^{\mathrm{H}}\end{bmatrix} \tag{26}$$

$$\widehat{\boldsymbol{X}} = \sum_{i=0}^{\min\{M-1,L-1\}} \widehat{\sigma}_i \widehat{\boldsymbol{u}}_i \widehat{\boldsymbol{v}}_i^{\mathrm{H}} = \begin{bmatrix}\widehat{\boldsymbol{U}} & \widehat{\boldsymbol{U}}_{\perp}\end{bmatrix} \begin{bmatrix}\widehat{\boldsymbol{\Sigma}} & \\ & \widehat{\boldsymbol{\Sigma}}_{\perp}\end{bmatrix} \begin{bmatrix}\widehat{\boldsymbol{V}}^{\mathrm{H}} \\ \widehat{\boldsymbol{V}}_{\perp}^{\mathrm{H}}\end{bmatrix} \tag{27}$$

whose ith descending singular values is $\{\sigma_i\}$. It follows from Wedin's $\sin\boldsymbol{\Theta}$ theorem that if $2\|\boldsymbol{\mathcal{E}}\| \leq \sigma_{P-1}(\widehat{\boldsymbol{X}})$, then it holds:

$$\sin\theta_0 \leq \frac{2\|\boldsymbol{\mathcal{E}}\|}{\sigma_{P-1}(\widehat{\boldsymbol{X}})} \tag{28}$$

Proof. See [38].

To bridge the connection between Davis-Kahan and Wedin's theorems with the singular values of the QR decomposition one can refer to Proposition 1.

Theorem 3 provides bounds that subspace methods must balance the perturbations $\boldsymbol{\mathcal{E}}$ caused to the signal subspace without affecting the smallest singular value of the signal subspace σ_{P-1}. For the original ESPRIT algorithm, perturbations are usually caused by noise in practice, which can lead to subspace swap of the smallest signal singular value and the largest noise singular value [22]. For Algorithm 1 in the noiseless case, this perturbation is characterized by the difference between the Hermitian product between the kernel and row space of the data matrix and the true subspace, i.e. the transformation by the kernel must aim to approach the true eigenvalue distribution under finite support.

To improve the efficiency of Algorithm 1, we propose a kernel of a truncated DFT matrix, in which the columns are chosen by their energy; this corresponds to an "eigenfilter" on the subspace. To demonstrate this idea, assume that $\widehat{\boldsymbol{X}}$ is rank-deficient such that its Vandermonde decomposition exists. Thus, $\widehat{\boldsymbol{X}}$ is decomposed as:

$$\begin{aligned}\widehat{\boldsymbol{X}} &= \boldsymbol{A}_P \boldsymbol{B} \boldsymbol{A}_L + \boldsymbol{\mathcal{E}} \\ &= \boldsymbol{X} + \boldsymbol{\mathcal{E}}\end{aligned} \tag{29}$$

where $\boldsymbol{A}_P \in \mathbb{C}^{M\times P}$ and $\boldsymbol{A}_L \in \mathbb{C}^{P\times L}$ are Vandermonde matrices, $\boldsymbol{B} = \operatorname{diag}\{\boldsymbol{\beta}\}$, and $\mathbb{E}\{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^{\mathrm{H}}\} = M\eta\boldsymbol{I}_M$ is the matrix form of the AWGN. Let us define the following matched distances of $\widehat{\boldsymbol{\Psi}} = \left(\widehat{\boldsymbol{U}}^{\uparrow}\right)^{\ddagger}\widehat{U}_{\downarrow}$ which is diagonalizable with eigenvalues $\{\widehat{\lambda}_i = e^{-\mathrm{j}2\pi\widehat{\omega}_i}\}_{i=0}^{P-1}$:

$$\operatorname{md}(\widehat{\boldsymbol{\Psi}}, \boldsymbol{\Psi}) \coloneqq \min_{\psi_i\in\boldsymbol{\Psi}} \max_i \left|\widehat{\lambda}_{\psi_i} - e^{-\mathrm{j}2\pi\omega_i}\right| \tag{30}$$

and subsequently the frequencies $\{\widehat{\omega}_i = -\angle\widehat{\lambda}_i/2\pi\}_{i=0}^{P-1}$:

$$\operatorname{md}(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega}) \coloneqq \min_{\psi_i\in\boldsymbol{\Psi}} \max_i |\widehat{\omega}_{\psi_i} - \omega_i| \tag{31}$$

These two matched distanced have the relationship, proven in [18, Lemma 2]:

$$\operatorname{md}(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega}) \leq \frac{1}{2}\operatorname{md}(\widehat{\boldsymbol{\Psi}}, \boldsymbol{\Psi}) \leq \|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\| \tag{32}$$

To have a direct comparison with the original ESPRIT, the authors in [18] proved that ESPRIT has bounds with respect to its minimum singular value σ_{P-1} of the signal subspace:

Theorem 4. *Let the constraints of Eq.* (4) *be fixed. The original ESPRIT has the following bounds if the noise is moderate and bounded such that:*

$$\|\boldsymbol{\mathcal{E}}\| \leq \frac{\beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{U}^{\uparrow})}{4\sqrt{2P}} \tag{33}$$

then its eigenspace has the stability:

$$\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\| \leq \frac{14\sqrt{2P}\|\boldsymbol{\mathcal{E}}\|}{\beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}^2(\boldsymbol{U}^{\uparrow})} \tag{34}$$

and the performance on the frequency matched distance:

$$\operatorname{md}(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega}) = \frac{20P^2\sqrt{M+1}\|\boldsymbol{\mathcal{E}}\|}{\beta_{\min}\sigma_{P-1}^2(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}^2(\boldsymbol{U}^{\uparrow})} \tag{35}$$

Proof. See [18].

The original ESPRIT is better understood through the stability of $\boldsymbol{\Psi}$, since it's computed from $\boldsymbol{U}$. What is notable about Theorem 4 are the roles of the smallest singular value of the Vandermonde matrices $\boldsymbol{A}_P, \boldsymbol{A}_L$ and the signal subspace matrix $\boldsymbol{U}^{\uparrow}$ (with its last row removed) on the stability and performance bounds. In [39], an accurate lower bound is given for $\sigma_{P-1}(\boldsymbol{A}_{\square})$, and in [18] for $\sigma_{P-1}(\boldsymbol{U}^{\uparrow})$.

By selectively multiplying $\widehat{\boldsymbol{X}}$ with $\boldsymbol{W}_P$ which corresponds to the kernel of the truncated DFT matrix with P frequencies, i.e. a matrix of size $L \times P$, Eq. (29) becomes:

$$\begin{aligned}\widehat{\boldsymbol{X}}\boldsymbol{W}_P &= \boldsymbol{A}_P\boldsymbol{B}\boldsymbol{A}_L\boldsymbol{W}_P + \boldsymbol{\mathcal{E}}\boldsymbol{W}_P \\ &= \boldsymbol{X}\boldsymbol{W}_P + \boldsymbol{\mathcal{E}}\boldsymbol{W}_P\end{aligned} \tag{36}$$

Given this, and the fact we opt to use the QR decomposition as opposed to an SVD, we can provide similar bounds:

Theorem 5. *Let the constraints of Eq.* (4) *be fixed. FFT-ESPRIT of Algorithm 2 with the truncated DFT matrix $\boldsymbol{W}_P \in \mathbb{C}^{L\times P}$ has the following bounds if the noise is moderate and bounded such that:*

$$\|\boldsymbol{\mathcal{E}}\| \leq \frac{\beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{W}_P)\sigma_{P-1}(\boldsymbol{Q}^{\uparrow})}{4\sqrt{2P}\|\boldsymbol{W}_P\|} \tag{37}$$

then its eigenspace has the stability:

$$\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\| \leq \frac{14\sqrt{2P}\|\boldsymbol{\mathcal{E}}\|\|\boldsymbol{W}_P\|}{\beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{W}_P)\sigma_{P-1}^2(\boldsymbol{Q}^{\uparrow})} \tag{38}$$

and the performance on the frequency matched distance:

$$\operatorname{md}(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega}) = \frac{20P^2\sqrt{M+1}\|\boldsymbol{\mathcal{E}}\|\|\boldsymbol{W}_P\|}{\beta_{\min}\sigma_{P-1}^2(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{W}_P)\sigma_{P-1}^2(\boldsymbol{Q}^{\uparrow})} \tag{39}$$

Proof. See Appendix B combined with [18, Lemma 2].

One can see that in the moderate to high SNR regime, FFT-ESPRIT will offer a lower quality estimate, since $\|\boldsymbol{W}_P\| \geq \sigma_{P-1}(\boldsymbol{W}_P)$. Between the no-information to low SNR regime, i.e. $2\|\boldsymbol{\mathcal{E}}\| \geq \sigma_{P-1}(\widehat{\boldsymbol{X}})$, one can show that the Wedin's bounds of Eq. (28) give $\sin\theta_0 \leq \frac{\sigma_{P-1}(\widehat{\boldsymbol{X}})}{2\|\boldsymbol{\mathcal{E}}\|}$. It is rudimentary to see that the moderate SNR regime performance bounds of Theorems 4 and 5 are inversed in this low SNR regime. We propose that there is a regime between the no-information and low SNR where FFT-ESPRIT can have better performance than the original ESPRIT, which we demonstrate numerically in the following section. The criticality of when $\boldsymbol{W}_P = \boldsymbol{A}_L$ is chosen exactly[3] can be understood in two senses: one, by invoking Theorem 5 realize that its performance in the low SNR regime is proportional to:

$$\mathrm{md}(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega}) \propto \frac{\sigma_{P-1}^2(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{W}_P)}{\|\boldsymbol{\mathcal{E}}\|\|\boldsymbol{W}_P\|} \tag{40}$$

FFT-ESPRIT will offer a better estimate since the subspace of the DFT matrix obeys $\|\boldsymbol{W}_P\| \geq \sigma_{P-1}(\boldsymbol{W}_P)$; two, that the SNR is improved in Frobenius-sense such that: $\|\boldsymbol{X}\boldsymbol{W}_P\|_{\mathrm{F}}/\|\boldsymbol{\mathcal{E}}\boldsymbol{W}_P\|_{\mathrm{F}} \geq \|\boldsymbol{X}\|_{\mathrm{F}}/\|\boldsymbol{\mathcal{E}}\|_{\mathrm{F}}$. When $\boldsymbol{W}_P$ deviates, i.e. the frequencies are no longer aligned with the true $\boldsymbol{A}_P$, the inequality may no longer hold.

We present the previous analysis implemented as FFT-ESPRIT in Algorithm 2. For use in an algorithm, the columns $\boldsymbol{W}_P$ should correspond only to the true frequencies. Therefore, we suggest that an FFT-based algorithm should be used which performs off-grid estimates for the columns of $\boldsymbol{W}_P$: e.g. an IIp-DFT estimator [7], which is one of many DFT peak interpolation algorithms that can be used in Algorithm 2. Additionally, one may realize that $\boldsymbol{W}_P$ must at minimum span P in order to lie in the signal subspace of $\boldsymbol{U}$, which does not allow one to underestimate the model order. From this point, we just notate $\boldsymbol{A}_P$ instead of $\boldsymbol{W}_P$ when using variable names in the algorithms. Note, the IIp-DFT algorithm is programmed such that its frequency estimates are sorted descending by peak prominence.

Algorithm 2 FFT-ESPRIT

Input: $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$, M, P
Output: $\boldsymbol{\omega}$
1: $\boldsymbol{X} \leftarrow (\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{M-1})$
2: $\widehat{\boldsymbol{\omega}} \leftarrow \texttt{IIp-DFT}(\boldsymbol{x}, P)$
3: $\boldsymbol{A}_P \leftarrow (\boldsymbol{a}_{-\widehat{\omega}_0}, \cdots, \boldsymbol{a}_{-\widehat{\omega}_{P-1}})$
4: $\boldsymbol{Y}_P \leftarrow \boldsymbol{X}\boldsymbol{A}_P$
5: $\boldsymbol{Q} \leftarrow \texttt{qr}(\boldsymbol{Y}_P, \texttt{'econ'})$
6: $\boldsymbol{Q}^{\uparrow} \leftarrow (\vec{\boldsymbol{q}}_0; \cdots; \vec{\boldsymbol{q}}_{M-2})$
7: $\boldsymbol{Q}_{\downarrow} \leftarrow (\vec{\boldsymbol{q}}_1; \cdots; \vec{\boldsymbol{q}}_{M-1})$
8: $\boldsymbol{\Psi} \leftarrow \left(\boldsymbol{Q}^{\uparrow}\right)^{\ddagger}\boldsymbol{Q}_{\downarrow}$
9: $\boldsymbol{\omega} \leftarrow \texttt{mod}(\texttt{angle}(\texttt{eig}(\boldsymbol{\Psi})/2\pi, 1))$
10: **return** $\boldsymbol{\omega}$

3.2 Fast version of FFT-ESPRIT

In this section we aim to accelerate the computations of Algorithm 2, namely to reduce the quadratic complexity dependency on M imposed by the matrix product of line 4 and to prevent inefficient computations of the pseudo-inverse of line 8. Due to the reduction of this time complexity, it enables one to create a fast version which obtains $\mathcal{O}(N\log N)$.

First, we will approach line 4 of Algorithm 2 and aim to leverage the properties of the Hankel matrix $\boldsymbol{X}$ to achieve a faster matrix product. In Algorithm 3 we show that the complexity can be reduced to $\approx 2P\mathcal{N}\log\mathcal{N}$, where $\mathcal{N}$ is the next power of two of the signal length N. Additionally, the Hankel matrix does not have to be explicitly formed. We take advantage of the circulant nature of the DFT, where the Hankel matrix can be Fourier transformed into a Toeplitz matrix, and the Vandermonde matrix can be Fourier transformed by taking its FFT. The product of these transformed matrices can then be obtained by element-wise multiplications, followed by an inverse FFT and a row-wise truncation. Note that lines 8-12 of Algorithm 3 are parallelizable given there are P computing processors available.

For line 8 of Algorithm 2, the pseudo-inverse used for the orthogonal basis of $\boldsymbol{Q}^{\uparrow}$ which is size $\approx M \times P$. One choice is to naively use MATLAB's `pinv`, which relies on calculating a thin SVD on $\boldsymbol{Q}^{\uparrow}$. However, the minimal norm least-squares solution is not useful since $\boldsymbol{Q}^{\uparrow}$ is never rank-deficient. A preferred alternative relies on a QR solver for the

[3] If and only if $\boldsymbol{W}_P$ corresponds to the true frequencies, i.e. the frequencies are chosen by their true PSD.

Algorithm 3 Fast Hankel Matrix-Matrix product

Input: $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$, $\boldsymbol{A}_P = (\boldsymbol{a}_{-\widehat{\omega}_0}, \cdots, \boldsymbol{a}_{-\widehat{\omega}_{P-1}})$, M, P
Output: $\boldsymbol{Y}_P$
$\boldsymbol{c} \leftarrow (x[M-1], \cdots, x[0])$
$\boldsymbol{r} \leftarrow (x[N-2], \cdots, x[M])$
$\mathcal{N} \leftarrow \texttt{nextPow2}(N-1)$
$\gamma \leftarrow 2^{\mathcal{N}} - N + 1$
$\boldsymbol{0} \leftarrow (0[0], \cdots, 0[\gamma])$
$\boldsymbol{t} \leftarrow (\boldsymbol{c}, \boldsymbol{0}, \boldsymbol{r})$
$\boldsymbol{t}_{\mathrm{f}} \leftarrow \texttt{fft}(\boldsymbol{t})$
for $i = 0$ to $P-1$ **do**
    $\boldsymbol{b} \leftarrow \texttt{fft}(\boldsymbol{a}_{-\widehat{\omega}_i}, 2^{\mathcal{N}})$
    $\boldsymbol{k} \leftarrow \boldsymbol{b} \circ \boldsymbol{t}_{\mathrm{f}}$
    $\boldsymbol{y}_i \leftarrow \texttt{ifft}(\boldsymbol{k})$
end for
$\boldsymbol{Y}_P \leftarrow (\vec{\boldsymbol{y}}_0; \cdots; \vec{\boldsymbol{y}}_{M-1})$
$\boldsymbol{Y}_P \leftarrow \texttt{flipud}(\boldsymbol{Y}_P)$
return $\boldsymbol{Y}_P$

least-squares solution, which is used by MATLAB's `mldivide` (\). Both these methods avoid the usage of the normal equations, which means their condition number is linear: $\kappa(\boldsymbol{Q}^{\uparrow})$.

If one uses the normal equations, a faster version can be derived starting with:

$$\boldsymbol{\Psi} = \left(\boldsymbol{Q}^{\uparrow}\right)^{\ddagger}\boldsymbol{Q}_{\downarrow} = \left(\left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}^{\uparrow}\right)^{-1}\left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}_{\downarrow}$$

when $\boldsymbol{Q}^{\uparrow}$ is full-column rank, and that $\left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}^{\uparrow}$ is non-singular. By invoking the Woodbury matrix identity [40] and storing the product $\left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}_{\downarrow}$ in memory, one obtains a rank-one modification:

$$\boldsymbol{\Psi} = \left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}_{\downarrow} + \left(\vec{\boldsymbol{q}}^{\mathrm{H}}\left(\vec{\boldsymbol{q}}\left(\left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}_{\downarrow}\right)\right)\right)\left(\frac{1}{\vec{\boldsymbol{q}}^{\mathrm{H}}\vec{\boldsymbol{q}}}\right) \tag{41}$$

where $\vec{\boldsymbol{q}}$ corresponds to the last row of $\boldsymbol{Q}$. While Eq. (41) is a faster method since it has a lower flops requirement, its performance relies on the orthogonality of the matrix, i.e. $\left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}^{\uparrow} = \boldsymbol{I}_P$, and therefore its condition number is quadratic: $\kappa((\boldsymbol{Q}^{\uparrow})^{\mathrm{H}}\boldsymbol{Q}^{\uparrow})$. The numerical error that is propagated due to corrupted data and/or poor orthogonality will lose twice as many digits of accuracy compared to the QR- or SVD-based methods. Therefore, the QR decomposition of $\boldsymbol{Y}_P$ is recommended to use a Householder pseudo-reflection variant (for complex numbers) described by [40, Theorem 2.1.13] as opposed to a Gram-Schmidt variant. We compare all LS options in Table 1 with respect to their flops and opt to use Eq. (41). Thus, when using Algorithm 3 and Eq. (41) for FFT-ESPRIT of Algorithm 2, one achieves the fast variant of FFT-ESPRIT with an asymptotic complexity of $\mathcal{O}(N \log N)$ since $P \ll M \propto N$.

Algorithm 4 Fast FFT-ESPRIT

Input: $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$, P, M
Output: $\boldsymbol{\omega}$
$\widehat{\boldsymbol{\omega}} \leftarrow \texttt{IIp-DFT}(\boldsymbol{x}, P)$
$\boldsymbol{A}_P \leftarrow (\boldsymbol{a}_{-\widehat{\omega}_0}, \cdots, \boldsymbol{a}_{-\widehat{\omega}_{P-1}})$
$\boldsymbol{Y}_P \leftarrow \texttt{Algorithm-3}(\boldsymbol{x}, \boldsymbol{A}_P, M, P)$
$\boldsymbol{Q} \leftarrow \texttt{qr}(\boldsymbol{Y}_P, \texttt{'econ'})$
$\boldsymbol{Q}^{\uparrow} \leftarrow (\vec{\boldsymbol{q}}_0; \cdots; \vec{\boldsymbol{q}}_{M-2})$
$\boldsymbol{Q}_{\downarrow} \leftarrow (\vec{\boldsymbol{q}}_1; \cdots; \vec{\boldsymbol{q}}_{M-1})$
$\boldsymbol{Z} \leftarrow \left(\boldsymbol{Q}^{\uparrow}\right)^{\mathrm{H}}\boldsymbol{Q}_{\downarrow}$
$\boldsymbol{\Psi} \leftarrow \boldsymbol{Z} + \left(\vec{\boldsymbol{q}}^{\mathrm{H}}\left(\vec{\boldsymbol{q}}\boldsymbol{Z}\right)\right)\left(\frac{1}{\vec{\boldsymbol{q}}^{\mathrm{H}}\vec{\boldsymbol{q}}}\right)$
$\boldsymbol{\omega} \leftarrow \texttt{mod}(\texttt{angle}(\texttt{eig}(\boldsymbol{\Psi})/2\pi, 1))$
return $\boldsymbol{\omega}$

Table 1: Comparison of LS algorithms with respect to flops

Algorithm	Computation	Flops (multiplications and additions)
SVD-based [41]	$\texttt{pinv}(\boldsymbol{Q}^{\uparrow})\boldsymbol{Q}_{\downarrow}$	$2M^2P + M^2 + 6MP^2 - \frac{4}{3}P^3 - MP + P$
QR-based	$\boldsymbol{Q}^{\uparrow}\backslash\boldsymbol{Q}_{\downarrow}$	$4MP^2 + \frac{1}{3}P^3 - P^2$
Woodbury-based	Eq. (41)	$2MP^2 + 3P^2 + P - 1$

4 Simulation results

We compare the previous FFT-based ESPRIT algorithms with respect to two performance measures on the signal model of Eq. (1) with length $N = 2^7$. Specifically, only the fast variant of FFT-ESPRIT (Algorithm 4) and Algorithm 1 are included. For all simulations, the complex amplitudes β_i are generated independent and identically distributed whose magnitudes are then normalized to unity.

We define the following metrics: the mean square error (MSE) is used as a statistical measure of error:

$$\text{MSE} \coloneqq \frac{1}{P}\sum_{j=1}^{P}\left(\min_{\widehat{\omega}_i\in\widehat{\boldsymbol{\omega}}}|\widehat{\omega}_i - \omega_j|\right)^2 \tag{42}$$

An approximate CRB is used alongside the MSE since it characterizes the asymptotic behavior at large N and/or high SNR [32] for a single sinusoid. This is given as:

$$\text{CRB} \coloneqq \frac{6\eta}{\|\boldsymbol{A}\boldsymbol{\beta}\|^2(N^2-1)} \tag{43}$$

The MSE can be misleading since it does not give an indication of the probability of failures and can be skewed by such outliers. Therefore, the frequency success rate (FSR) is defined as follows:

$$\text{FSR} \coloneqq \frac{\sum_{\widehat{\omega}_i\in\widehat{\boldsymbol{\omega}}}\mathcal{S}\left(\widehat{\omega}_i,\boldsymbol{\omega}\right) + \sum_{\omega_j\in\boldsymbol{\omega}}\mathcal{S}\left(\omega_j,\widehat{\boldsymbol{\omega}}\right)}{2P} \tag{44}$$

with the success function defined as:

$$\mathcal{S}(i,\boldsymbol{j}) \coloneqq \mathbb{1}\left[\min_{j_k\in\boldsymbol{j}}|i-j_k| < \frac{1}{2N}\right] \tag{45}$$

where $\mathbb{1}[\square]$ denotes the indicator function. An FSR of 1 is obtained if all estimated frequencies are near one or more simulated frequencies and all simulated frequencies are near one or more estimated frequencies. All metrics are averaged over all independent runs.

To benchmark our algorithms, we use the original ESPRIT algorithm [16] which is given the same parameters as FFT-ESPRIT and Algorithm 1, i.e. $M = N/2$. The original ESPRIT and Algorithm 1 use a QR-based solver for $\boldsymbol{\Psi}$, whereas FFT-ESPRIT uses the Woodbury-based solution of Eq. (41). We also include the admission of the IIp-DFT algorithm to contrast the effect of its estimation performance on FFT-ESPRIT's performance.

4.1 Single sinusoid

In the first simulation, we perform 10^4 Monte Carlo simulations of $P = 1$ complex sinusoid with zero-mean and is corrupted by AWGN with a varying SNR. In Fig. 1, the MSE is normalized by the CRB of Eq. (43). It can be seen that in the no-information to low SNR regimes, FFT-ESPRIT has better performance over the original ESPRIT in both MSE and FSR. As the SNR increases, this advantage slowly inverts, and ESPRIT has better asymptotic performance. Algorithm 1 can be seen to have the worst MSE in the single sinusoidal case. When contrasted with its FSR, one can conclude its performance is limited by the subspace swap introduced by using the eigenvalue approximation of the full DFT matrix for a subspace estimate. The effect of IIp-DFT can be seen to offer subpar MSE relative to FFT-ESPRIT throughout the SNR regimes. However, its estimate provides FFT-ESPRIT an advantage with an "eigenfilter" property which allows it to outperform ESPRIT from the no-information regime up to ≈ -3 dB.

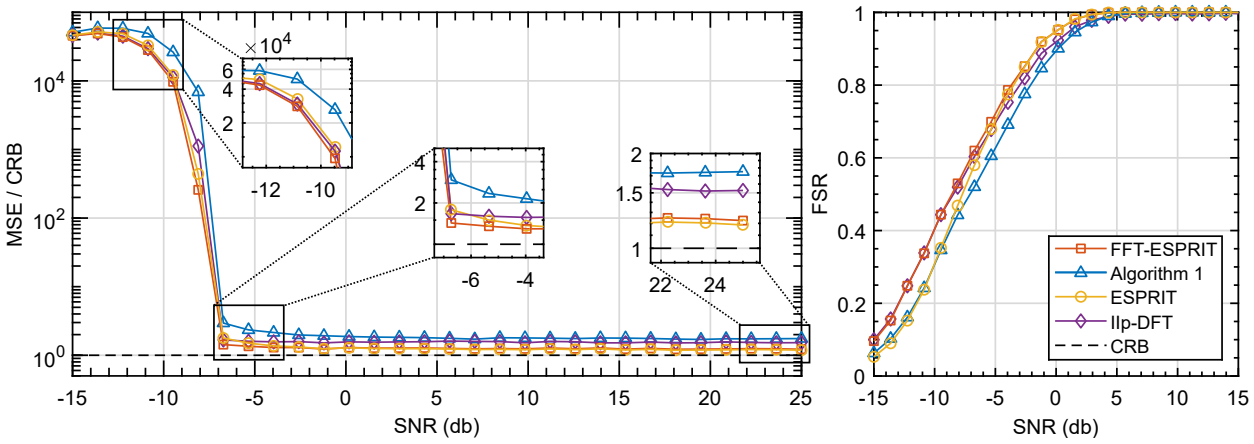

Figure 1: Simulation results for varying SNR. The signal length is $N = 2^7$ with $P = 1$ sinusoid.

4.2 Super-resolution of two closely-spaced sinusoids

Here, the performance of the estimators with respect to their super-resolution is analyzed, i.e. ability to resolve closely-spaced frequencies beyond the Rayleigh limit $1/N$ (the best resolution obtainable by the DFT). We perform 10^4 Monte Carlo simulations of $P = 2$ complex sinusoids with one of the frequencies to be closely situated within a multiple of the Rayleigh limit $1/N$. The signals have zero mean and are chosen to have an SNR of 10 dB. In Fig. 2, one can observe that Algorithm 1 outperforms other estimators until $\approx 0.5/N$. After this threshold, all estimators but ESPRIT have a sharp transition in MSE. At this regime, FFT-ESPRIT, as well as Algorithm 1, has a performance advantage over the original ESPRIT up until $\approx 0.7/N$. Afterward, FFT-ESPRIT, Algorithm 1, and ESPRIT all closely follow similar MSEs demonstrating the superiority of subspace methods and their ability of super-resolution. At first, one may be surprised by the MSE performance of IIp-DFT below $\approx 0.5/N$. However, its lower MSE can be ascribed to a precise, yet inaccurate frequency estimate, indicated by its very low FSR. This bias introduced by interpolation [7], results in plateaus in its MSE and FSR. In terms of the FSR, all three subspace methods follow a similar trajectory, with Algorithm 1 lagging behind after $\approx 0.7/N$.

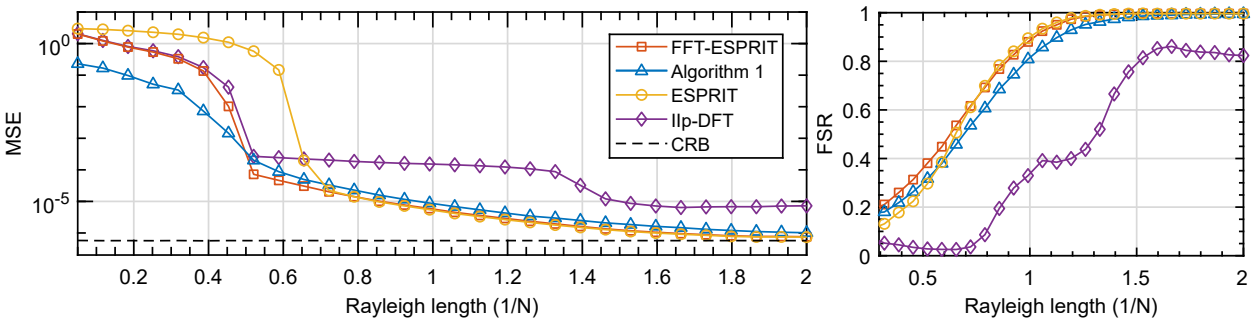

Figure 2: Simulation results for a pair of closely-spaced sinusoids. The signal length is $N = 2^7$ with where one complex sinusoid is closely situated at a multiple of the Rayleigh limit.

4.3 Bias of two, well-separated sinusoids

To contrast with the previous section, the bias of other well-separated sinusoids with respect to performance is investigated. As mentioned in Section 2.2, the covariance between two sinusoids starts to become negligible when $|\omega_i - \omega_j| > 2/N$ [32]. Therefore, one can expect a compounding effect if an estimator is sensitive to the number of sinusoids. We perform 10^4 Monte Carlo simulations of $P = 2$ complex sinusoids which are ensured to be well-separated of at least $2/N$: the set of well-separated frequencies are drawn from a uniform distribution $\{\omega_i \in [0, 1) : \min\{|\omega_i - \omega_j|, |1 - |\omega_i - \omega_j||\} \geq 2/N : \forall i \neq j\}$. The signals have zero mean and are corrupted by AWGN with varying SNRs.

In Fig. 3, the MSE is normalized by the CRB of Eq. (43). Similarly to the single sinusoidal case, FFT-ESPRIT can be seen to have an advantage over the other estimators in the no-information to the low-SNR regime, until ≈ 2 dB. This behavior was conjectured after Theorem 5, where the truncated DFT matrix enhanced the SNR at a bounded noise perturbation, yielding the performance increase shown by Eq. (40). ESPRIT has a better MSE than FFT-ESPRIT after this inflection point. FFT-ESPRIT, being an approximation of the subspace, obtains nearly identical performance as ESPRIT in the medium SNR regime and above. Differently from the single sinusoidal case, Algorithm 1 begins to suffer greatly in its threshold transition from ≈ -10 dB to 2 dB: the approximation of the eigenvalues provided by the kernel of the full DFT matrix includes a small amount of subspace swap per sinusoid, in addition to the subspace swap of noise, thus compounding the effect. When looking at the FSR of Algorithm 1, one can see a relatively normal trajectory, indicating that the poor performance in its MSE is due to a spurious outlier.

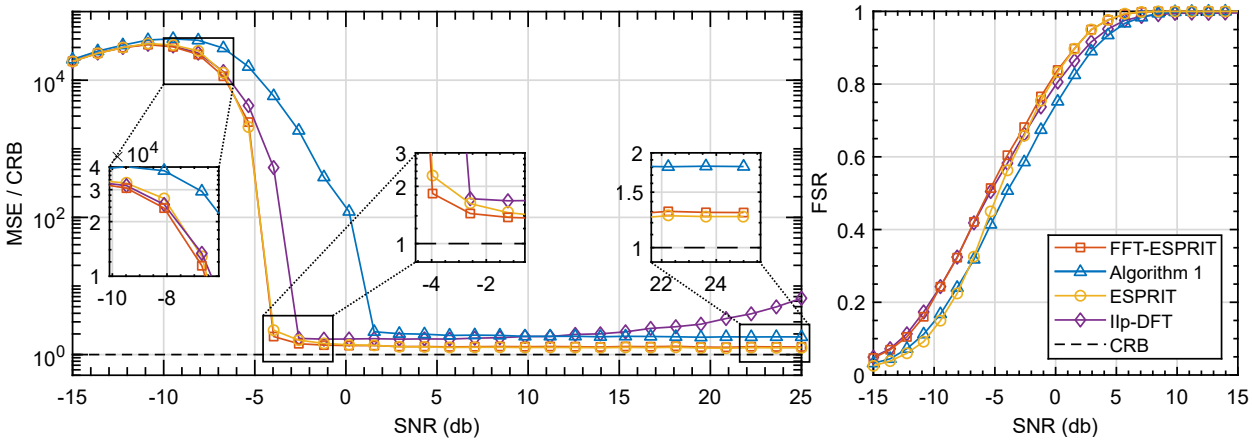

Figure 3: Simulation results for varying SNR. The signal length is $N = 2^7$ with $P = 2$ well-separated sinusoids.

Lastly, IIp-DFT suffers in MSE for multiple sinusoids, as opposed to the single sinusoidal case in both the SNR transitionary regime and the medium SNR regime and beyond. Interpolation methods are known to achieve ML performance in the single sinusoidal asymptotic case [32], but the iterative steps to detect and subtract multiple sinusoids cannot completely remove this bias effect for IIp-DFT.

4.4 Computation times

In Fig. 4 we show algorithm runtimes for varying signal lengths, for single and multiple sinusoidal cases. The results are obtained from MATLAB's `timeit`, using an Intel Core i7-12800H processor. As discussed in Section 3.2, FFT-ESPRIT has an asymptotic time-complexity of $\mathcal{O}(N \log N)$, which is juxtaposed onto the plot at larger N. For larger signals, the time-complexity of ESPRIT ($\mathcal{O}(N^3)$) and Algorithm 1 ($\mathcal{O}(N^2 \log N)$) are evident. For the case of $P = 1$ sinusoid, ESPRIT and Algorithm 1 require ≈ 50 seconds at signal lengths $N = 2^{13}$ and $N = 2^{16}$ respectively. FFT-ESPRIT has a clear advantage due to the fast-multiplication of Algorithm 3, only taking ≈ 0.05 seconds at signal length $N = 2^{16}$.

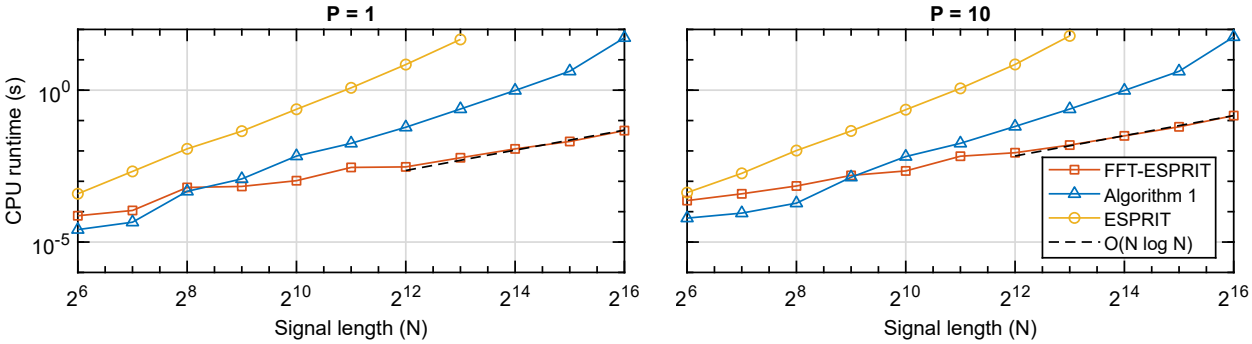

Figure 4: CPU runtimes versus the number of sinusoids $P = 1$ (left) and $P = 10$ (right). The timing in seconds is measured with MATLAB's `timeit`.

5 Conclusions

We have presented two low-complexity FFT-based ESPRIT algorithms that utilize the kernels of the DFT matrix and truncated DFT matrix, which make use of Szegö's theorem for eigenvalue approximations. Both proposed algorithms fall in the category of subspace approximation methods for line spectral estimation and keep their flops with respect to the signal length below $\mathcal{O}(N^4)$. If the covariance matrix form is used, the method is similar in form to the popular kernel PCA.

Algorithm 1 is posited to have a subspace swap introduced by using the eigenvalue approximation of the full DFT matrix for a subspace estimate. This is seen in Section 4 by its shifted SNR transitional threshold in MSE compared to other estimators in simulations, as well as lower FSRs for all cases. This is remediated in FFT-ESPRIT (Algorithm 4) when incorporating a priori knowledge into the kernel of the DFT matrix, i.e. the kernel of a truncated DFT matrix.

FFT-ESPRIT combines certain non-parametric and parametric features of its underlying algorithms, specifically efficient eigenvalue and subspace approximation followed by utilization of the subspace rotation property. The perturbation bounds of Theorem 3 and performance bounds of Theorem 4 allow one to conjecture a performance increase of FFT-ESPRIT over the original ESPRIT algorithm in the no-information to low SNR regimes. This is validated in numerical simulations for both single and multiple sinusoidal cases. FFT-ESPRIT has computational and performance advantages over Algorithm 1: achieving a quasi-linear time complexity of $\mathcal{O}(N \log N)$ due to the fast Hankel matrix-matrix multiplication of Algorithm 3; and having a closer asymptotic performance to the original ESPRIT at high SNRs.

The conclusions following from Section 4 indicate that FFT-ESPRIT can be used instead of the original ESPRIT for line spectral problems with real-time hardware constraints and/or large signal lengths. Contrary to iterative, interpolated DFT algorithms, FFT-ESPRIT provides superior super-resolution performance and does not suffer from a bias effect in higher SNR regimes when the number of sinusoids increase, as seen in Fig. 4.

Future research avenues lie in enhancing the signal subspace estimate through alternative kernels whose fast integral transform exists, providing deterministic bounds (as opposed to only asymptotic and perturbation bounds), and generalizing the approach to compressed sensing or multi-dimensional problems.

A Distribution of eigenvalues for Toeplitz matrix and the PSD

Because of the similarity between circulant and Toeplitz matrices and Theorems 1 and 2, for a finite signal of length N, the approximate eigenvalue distribution can be replicated via the PSD.

Take the (auto)covariance sequence $r[n] = e^{-2|n|}$, which can be obtained from the convolution of the signals $r[n] = x[n] \circledast x[-n]$. The PSD $\phi(\omega)$ for $\omega \in [0, 2\pi)$ is juxtaposed with the eigenvalues λ of $\tilde{\boldsymbol{R}}$. The eigenvalues are plotted to be the closest match of the spectrum $\phi(\omega)$. Fig. 5 shows the eigenvalue distribution and the PSD as N increases from $N = 2^5 \to 2^7$.

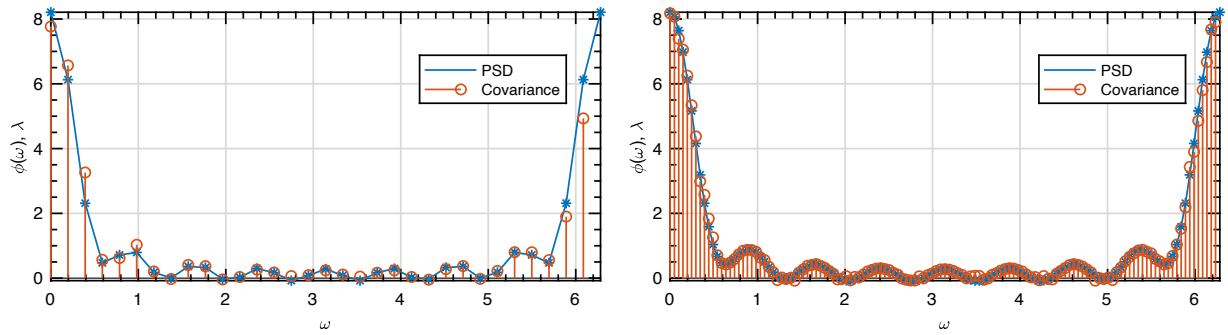

Figure 5: Comparison between the eigenvalues of $\tilde{\boldsymbol{R}}$ and $\phi(\omega)$ for $N = 2^5$ (left) and $N = 2^7$ (right).

B Proof of Theorem 5

Here we will follow, almost identically, to the derivation of [18] for certain perturbation bounds for FFT-ESPRIT of Algorithm 2, which has the form:

$$\begin{aligned}\widehat{\boldsymbol{X}}\boldsymbol{W}_P &= \boldsymbol{A}_P\boldsymbol{B}\boldsymbol{A}_L\boldsymbol{W}_P + \boldsymbol{\mathcal{E}}\boldsymbol{W}_P \\ &= \boldsymbol{X}\boldsymbol{W}_P + \boldsymbol{\mathcal{E}}\boldsymbol{W}_P\end{aligned}$$

First Eq. (28) is rewritten from Theorem 3 with respect to the truncated DFT matrix, which states that if $2\|\boldsymbol{\mathcal{E}}\boldsymbol{W}_P\| \leq 2\|\boldsymbol{\mathcal{E}}\|\|\boldsymbol{W}_P\| \leq \sigma_{P-1}(\widehat{\boldsymbol{X}}\boldsymbol{W}_P)$:

$$\sin\theta_0 \leq \frac{2\|\boldsymbol{\mathcal{E}}\boldsymbol{W}_P\|}{\sigma_{P-1}(\widehat{\boldsymbol{X}}\boldsymbol{W}_P)} \leq \frac{2\|\boldsymbol{\mathcal{E}}\|\|\boldsymbol{W}_P\|}{\sigma_{P-1}(\widehat{\boldsymbol{X}}\boldsymbol{W}_P)} \tag{46}$$

Given that ESPRIT is invariant to the choice of orthonormal basis by Proposition 1, we work with $\boldsymbol{Q} \leftarrow \boldsymbol{X}\boldsymbol{W}_P$. Eq. (46) shows that when the rotation between $\boldsymbol{Q}, \widehat{\boldsymbol{Q}}$ are small, the column spaces are close when the noise is negligible.

Lemma 1. *Let the size constraints of the data matrix Eq.* (4) *be fixed. If* $2\|\boldsymbol{\mathcal{E}}\boldsymbol{W}_P\| \leq \beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{W}_P)$*, then:*

$$\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\| \leq \frac{2\sqrt{2P}\|\boldsymbol{\mathcal{E}}\boldsymbol{W}_P\|}{\beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{W}_P)} \leq \frac{2\sqrt{2P}\|\boldsymbol{\mathcal{E}}\|\|\boldsymbol{W}_P\|}{\beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{W}_P)}$$

Proof. For $k = 0, \cdots, P-1$:

$$\begin{aligned}\|\widehat{\boldsymbol{q}}_k - \boldsymbol{q}_k\|^2 &= 4\sin^2\left(\frac{\theta_k}{2}\right) \\ &= 2\left(1-\cos\theta_k\right) \leq 2\left(1-\cos^2\theta_k\right) \leq 2\sin^2\theta_k\end{aligned}$$

By the properties of matrix norms, and the above inequality, one gets:

$$\begin{aligned}\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\| \leq \|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\|_{\mathrm{F}} &= \left(\sum_{k=0}^{P-1}\|\widehat{\boldsymbol{q}}_k - \boldsymbol{q}_k\|^2\right)^{1/2} \\ &\leq \left(2P\sin^2\theta_0\right)^{1/2} = \sqrt{2P}\sin\theta_0\end{aligned}$$

Given the definition of the ℓ_2 norm on $\boldsymbol{X}\boldsymbol{W}_P$:

$$\|\boldsymbol{X}\boldsymbol{W}_P\| = \|\boldsymbol{A}_P\boldsymbol{B}\boldsymbol{A}_L\boldsymbol{W}_P\| \geq \beta_{\min}\sigma_{P-1}(\boldsymbol{A}_P)\sigma_{P-1}(\boldsymbol{A}_L)\sigma_{P-1}(\boldsymbol{W}_P)$$

This combined with Eq. (46) completes the proof. ■

Thus, the next step is to relate $\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\|$ with $\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\|$ since $\boldsymbol{\Psi} \leftarrow \boldsymbol{Q}$.

Lemma 2. *Let the size constraints of the data matrix Eq.* (4) *be fixed. Since* $\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\| \leq \sigma_{P-1}(\boldsymbol{Q}^{\uparrow})/2$*, then:*

$$\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\| \leq \frac{7\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\|}{\sigma^2_{P-1}(\boldsymbol{Q}^{\uparrow})}$$

Proof. The perturbation of $\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\|$ can be decomposed via triangle inequalities in ℓ_2 space [42]:

$$\begin{aligned}\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\| &= \left\|\left((\widehat{\boldsymbol{Q}}^{\uparrow})^{\ddagger} - (\boldsymbol{Q}^{\uparrow})^{\ddagger}\right)\widehat{\boldsymbol{Q}}_{\downarrow} + (\boldsymbol{Q}^{\uparrow})^{\ddagger}(\widehat{\boldsymbol{Q}}_{\downarrow} - \boldsymbol{Q}_{\downarrow})\right\| \\ &\leq \left\|(\widehat{\boldsymbol{Q}}^{\uparrow})^{\ddagger} - (\boldsymbol{Q}^{\uparrow})^{\ddagger}\right\|\left\|\widehat{\boldsymbol{Q}}_{\downarrow}\right\| + \left\|(\boldsymbol{Q}^{\uparrow})^{\ddagger}\right\|\left\|\widehat{\boldsymbol{Q}}_{\downarrow} - \boldsymbol{Q}_{\downarrow}\right\| \\ &\leq \left\|(\widehat{\boldsymbol{Q}}^{\uparrow})^{\ddagger} - (\boldsymbol{Q}^{\uparrow})^{\ddagger}\right\| + \left\|(\boldsymbol{Q}^{\uparrow})^{\ddagger}\right\|\left\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\right\|\end{aligned}$$

since $\|\widehat{\boldsymbol{Q}}^{\uparrow}\| \leq \|\widehat{\boldsymbol{Q}}\| = 1$ and $\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\| \leq \|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\|$. When assuming (well-conditioned) signal sEubspaces:

$$\|\widehat{\boldsymbol{Q}}^{\uparrow} - \boldsymbol{Q}^{\uparrow}\| \leq \|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\| \leq \frac{1}{2\sigma_{P-1}(\boldsymbol{Q}^{\uparrow})}$$

This allows application of the truncated SVD theorem [43]:

$$\begin{aligned}\left\|(\widehat{\boldsymbol{Q}}^{\uparrow})^{\ddagger}-(\boldsymbol{Q}^{\uparrow})^{\ddagger}\right\| &\leq \frac{3\|\widehat{\boldsymbol{Q}}^{\uparrow}-\boldsymbol{Q}^{\uparrow}\|}{\sigma_{P-1}(\boldsymbol{Q}^{\uparrow})\big(\sigma_{P-1}(\boldsymbol{Q}^{\uparrow})-\|\widehat{\boldsymbol{Q}}^{\uparrow}-\boldsymbol{Q}^{\uparrow}\|\big)}\\ &\leq \frac{6\|\widehat{\boldsymbol{Q}}-\boldsymbol{Q}\|}{2\sigma_{P-1}^{2}(\boldsymbol{Q}^{\uparrow})}\end{aligned}$$

Therefore, one can relate the eigenspace with the signal subspace:

$$\begin{aligned}\|\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}\| &\leq \left(\frac{6}{\sigma_{P-1}^{2}(\boldsymbol{Q}^{\uparrow})}+\frac{1}{\sigma_{P-1}(\boldsymbol{Q}^{\uparrow})}\right)\|\widehat{\boldsymbol{Q}}-\boldsymbol{Q}\|\\ &\leq \frac{7\|\widehat{\boldsymbol{Q}}-\boldsymbol{Q}\|}{\sigma_{P-1}^{2}(\boldsymbol{Q}^{\uparrow})}\end{aligned}$$

This combined with triangle inequality completes the proof. ■

References

[1] Y. Wu, C. Li, Y. T. Hou, and W. Lou, "Real-time DoA Estimation for Automotive Radar," in *2021 18th European Radar Conference (EuRAD)*, pp. 437–440, Apr. 2022.

[2] V. Viti, C. Petrucci, and P. Barone, "Prony methods in NMR spectroscopy," *International Journal of Imaging Systems and Technology*, vol. 8, no. 6, pp. 565–571, 1997.

[3] X. Andrade, J. N. Sanders, and A. Aspuru-Guzik, "Application of compressed sensing to the simulation of atomic systems," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 13928–13933, Aug. 2012.

[4] K. Ege, X. Boutillon, and B. David, "High-resolution modal analysis," *Journal of Sound and Vibration*, vol. 325, pp. 852–869, Sept. 2009.

[5] G. R. de Prony, "Essai expérimental et analytique: Sur les lois de la dilatabilité de fluides elastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes temperatures," *Jour. de L'Ecole Polytechnique*, vol. 1, pp. 24–76, 1795.

[6] Jian Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 281–295, Feb./1996.

[7] S. Ye and E. Aboutanios, "An algorithm for the parameter estimation of multiple superimposed exponentials in noise," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (South Brisbane, Queensland, Australia), pp. 3457–3461, IEEE, Apr. 2015.

[8] Rayleigh, "XXXI. *Investigations in optics, with special reference to the spectroscope*," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 8, pp. 261–274, Oct. 1879.

[9] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[10] B. N. Bhaskar and B. Recht, "Atomic norm denoising with applications to line spectral estimation," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, (Monticello, IL), pp. 261–268, IEEE, Sept. 2011.

[11] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized Orthogonal Matching Pursuit: Frequency Estimation Over the Continuum," *IEEE Transactions on Signal Processing*, vol. 64, pp. 5066–5081, Oct. 2016.

[12] T. L. Hansen, B. H. Fleury, and B. D. Rao, "Superfast Line Spectral Estimation," *IEEE Transactions on Signal Processing*, vol. 66, pp. 2511–2526, May 2018.

[13] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, (Cancun, Mexico), pp. 41–44, IEEE, Dec. 2015.

[14] P. Liu and H. Zhang, "A Theory of Computational Resolution Limit for Line Spectral Estimation," *IEEE Transactions on Information Theory*, vol. 67, pp. 4812–4827, July 2021.

[15] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, Mar. 1986.

[16] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 984–995, July 1989.

[17] P. Stoica and T. Soderstrom, "Statistical analysis of MUSIC and ESPRIT estimates of sinusoidal frequencies," in *[Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, (Toronto, Ont., Canada), pp. 3273–3276 vol.5, IEEE, 1991.

[18] W. Li, W. Liao, and A. Fannjiang, "Super-Resolution Limit of the ESPRIT Algorithm," *IEEE Transactions on Information Theory*, vol. 66, pp. 4593–4608, July 2020.

[19] W. Liao and A. Fannjiang, "MUSIC for single-snapshot spectral estimation: Stability and super-resolution," *Applied and Computational Harmonic Analysis*, vol. 40, pp. 33–67, Jan. 2016.

[20] J. Thomas, L. Scharf, and D. Tufts, "The probability of a subspace swap in the SVD," *IEEE Transactions on Signal Processing*, vol. 43, pp. 730–736, Mar. 1995.

[21] A. Gershman and M. Haardt, "Improving the performance of Unitary ESPRIT via pseudo-noise resampling," *IEEE Transactions on Signal Processing*, vol. 47, no. 8, pp. 2305–2308, Aug./1999.

[22] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: Prediction and cure," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 6, (Salt Lake City, UT, USA), pp. 4005–4008, IEEE, 2001.

[23] S. F. B. Pinto and R. C. de Lamare, "Multistep Knowledge-Aided Iterative ESPRIT: Design and Analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, pp. 2189–2201, Oct. 2018.

[24] S. L. Kiser, M. Rébillat, M. Guskov, and N. Ranc, "Real-time sinusoidal parameter estimation for damage growth monitoring during ultrasonic very high cycle fatigue tests," *Mechanical Systems and Signal Processing*, vol. 182, p. 109544, Jan. 2023.

[25] V. A. Lacerda, P. I. Barbalho, R. M. Monaro, and D. V. Coury, "Signal processing techniques for synchrophasors considering short-circuit signals: A comparative study," *IET Generation, Transmission & Distribution*, vol. 14, no. 19, pp. 3962–3971, 2020.

[26] D. Tufts and C. Melissinos, "Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies," in *ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 10, pp. 320–323, Apr. 1985.

[27] S. Kay and A. Shaw, "Frequency estimation by principal component AR spectral estimation method without eigendecomposition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 95–101, Jan. 1988.

[28] J. Karhunen and J. Joutsensalo, "Sinusoidal frequency estimation by signal subspace approximation," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2961–2972, Dec. 1992.

[29] O. Das, J. Abel, and J. Smith, "FAST MUSIC-An efficient implementation of the MUSIC algorithm for frequency estimation of approximately periodic signals," Sept. 2018.

[30] D. Potts and M. Tasche, "Fast ESPRIT algorithms based on partial singular value decompositions," *Applied Numerical Mathematics*, vol. 88, pp. 31–45, Feb. 2015.

[31] C. Qian, L. Huang, and H. So, "Computationally efficient ESPRIT algorithm for direction-of-arrival estimation based on Nyström method," *Signal Processing*, vol. 94, pp. 74–80, Jan. 2014.

[32] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, N.J: Pearson/Prentice Hall, 2005.

[33] T. J. Suffridge and T. L. Hayden, "Approximation by a Hermitian Positive Semidefinite Toeplitz Matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 14, pp. 721–734, July 1993.

[34] U. Grenander, G. Szegö, and M. Kac, "Toeplitz Forms and Their Applications," *Physics Today*, vol. 11, p. 38, Jan. 1958.

[35] J.-P. Delmas, "Asymptotic normality of sample covariance matrix for mixed spectra time series: Application to sinusoidal frequencies estimation," *IEEE Transactions on Information Theory*, vol. 47, pp. 1681–1687, May 2001.

[36] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Transactions on Information Theory*, vol. 18, pp. 725–730, Nov. 1972.

[37] S. Van Huffel, "Enhanced resolution based on minimum variance estimation and exponential data modeling," *Signal Processing*, vol. 33, pp. 333–355, Sept. 1993.

[38] P.-Å. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT*, vol. 12, pp. 99–111, Mar. 1972.

[39] W. Li and W. Liao, "Stable super-resolution limit and smallest singular value of restricted Fourier matrices," *Applied and Computational Harmonic Analysis*, vol. 51, pp. 118–156, Mar. 2021.

[40] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge ; New York: Cambridge University Press, 2nd ed ed., 2012.

[41] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences, Baltimore: The Johns Hopkins University Press, fourth edition ed., 2013.

[42] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Spectral Methods for Data Science: A Statistical Perspective," *Foundations and Trends® in Machine Learning*, vol. 14, no. 5, pp. 566–806, 2021.

[43] P. C. Hansen, "The truncated SVD as a method for regularization," *BIT*, vol. 27, pp. 534–553, Dec. 1987.