Selecting Optimal Sampling Rate for Stable Super-Resolution

Nuha Diab
Department of Applied Mathematics, Tel-Aviv University, Tel-Aviv, Israel
nuhadiab@tauex.tau.ac.il

***Abstract*—We investigate the recovery of nodes and amplitudes from noisy frequency samples in spike train signals, also known as the super-resolution (SR) problem. When the node separation falls below the Rayleigh limit, the problem becomes ill-conditioned. Admissible sampling rates, or decimation parameters, improve the conditioning of the SR problem, enabling more accurate recovery. We propose an efficient preprocessing method to identify the optimal sampling rate, significantly enhancing the performance of SR techniques.**

I. Introduction

Consider the following "spike train" signal:

$$\mu(x) := \sum_{i=1}^{n} a_j \delta_{x_j}, \quad a_j \in \mathbb{C}, \quad x_j \in \mathbb{T} := \mathbb{R} \mod 2\pi. \tag{1}$$

where δ_x is the Dirac δ-distribution. We refer to $\{x_j\}_{j=1}^n$ as the nodes and to $\{a_j\}_{j=1}^n$ as the coefficients. Given noisy band-limited Fourier measurements $\hat{\mu}_\epsilon(\omega) = \hat{\mu}(\omega) + e(\omega)$, where

$$\hat{\mu}(\omega) = \sum_{j=1}^{n} a_j \exp(i\omega x_j), \quad \omega \in [-\Omega, \Omega] \tag{2}$$

where $\Omega > 0$ and $\|e(\omega)\|_\infty \leq \epsilon$ for some $\epsilon > 0$, the goal is to recover $\{a_j, x_j\}_{j=1}^n$ in equation (1). This **Super-resolution** (SR) problem is an inverse problem of great theoretical and practical interest, with diverse applications in optics, imaging, inverse scattering, signal processing, spectroscopy, and data analysis in general [15], [6], [25], [1], [31], [28], [7].

Let the SR factor be defined as $SRF := (\Delta\Omega)^{-1}$, where Δ is the smallest separation of the nodes. When $SRF \gg 1$, the SR problem becomes very ill-conditioned. In this regime, the nodes form 'clusters'. An important goal is to prove that certain algorithms are optimal, attaining the optimal reconstruction bounds known as the Min-max error bounds [5]. Extensive research has been conducted on this model regarding its stability analysis and algorithms [24], [26], [12], [2], [11], [27], [14], [23], [16].

The utilization of the Decimation technique was pioneered in [5] to establish bounds on the Min-max error in the SR regime. The decimation method proceeds as follows: the spectral range $[-\Omega, \Omega]$ is uniformly sampled at a rate ρ, and subsequently, a system of equations of the form $\hat{\mu}(\rho k) = \sum_{j=1}^n a_j \exp(i\rho k x_j)$ is obtained. However, not all ρ's are desirable or suitable; some may result in collisions and the formation of new clusters, thus worsening the ill-conditioned nature of the problem. The existence of admissible ρ's within a certain interval was first proved in [5]. An optimal approach proposed in [5] involves employing an oracle to obtain the correct value of the decimation parameter ρ.

The Decimation method has already been applied in SR algorithms, such as the Decimated Prony Method (DP) [21] and VEXPA [8], to improve the conditioning of the problem (see also [29], [30], [9]). Similar ideas have also been explored in the sparse FFT literature [17], [20]. Using decimation as a preprocessing step in any SR method with a suitable ρ yields a set of solutions $\{e^{\imath\rho x_j}\}_{j=1}^n$ instead of $\{e^{\imath x_j}\}_{j=1}^n$. This new set of solutions forms a well-separated configuration compared to the undecimated one, making recovery easier and more accurate.

Furthermore, the optimality of DP is established due to the optimality of Prony's method for $\Omega = 2n$ combined with decimation [22]. *Despite its practical usage, there is currently no mathematical evidence supporting the identification of such admissible decimation parameter*.

In this paper, our objective is to develop a method to automatically select an optimal decimation parameter ρ within a given interval, ensuring effective separation between the nodes. Towards that goal, we find the spectrum of the square Vandermonde matrix of the samples (2), showing its dependence on Δ and on the number of clusters the nodes form (similarly to [3] for rectangular Vandermonde matrices with number of samples $N \gg 2n$). Consequently, we find the spectrum of the Toeplitz matrix of the samples (3), which is our main tool to select the optimal decimation parameter. Lastly, we present the Enhanced Decimated Prony's method (EDP) and the Decimated Matrix Pencil method (DMP), demonstrating their time-complexity advantage over both DP and the Matrix Pencil Method (MP) [19] and then show the numerical optimality of EDP.

II. Preliminaries

We recall some definitions from [3].

Definition 1. *For $x \in \mathbb{R}$, we denote*

$$\|x\|_{\mathbb{T}} := \left|\arg(e^{\imath x})\right| = \left|x \mod (-\pi, \pi]\right|,$$

where $\arg(z)$ is the principal value of the argument of $z \in \mathbb{C} \setminus \{0\}$, taking values in $(-\pi, \pi]$.

For a set of n distinct nodes $X := \{x_j\}_j^n$ with $x_j \in (-\frac{\pi}{2}, \frac{\pi}{2}]$, we introduce the following definitions.

Definition 2. *Define the* ***minimal separation*** *of the set X as*

$$\Delta = \Delta(X) := \min_{i \neq j} \|x_i - x_j\|_{\mathbb{T}}.$$

In addition, for any $\rho \in \mathbb{R}$ we define

$$\Delta_\rho = \Delta(\rho X) := \min_{i \neq j} \|\rho x_i - \rho x_j\|_{\mathbb{T}}.$$

Definition 3 (Single cluster configuration)**.** *The set of nodes X is said to form an (Δ, ν, n)-cluster if*

$$\forall x, y \in X, x \neq y : \Delta \leq \|y - x\|_{\mathbb{T}} \leq \nu\Delta.$$

Definition 4 (Multi-cluster configuration)**.** *The set of nodes X is said to form an $((h^{(j)}, \nu^{(j)}, n^{(j)})_{j=1}^M, \eta)$-clustered configuration if there exist an M-partition $X = \bigcup_{j=1}^M \mathcal{C}^{(j)}$, such that for each $j \in \{1, ..., M\}$ the following conditions are satisfied:*

1) *$\mathcal{C}^{(j)}$ is an $(h^{(j)}, \nu^{(j)}, n^{(j)})$-cluster.*
2) *$\|x - y\|_{\mathbb{T}} \geq \eta > 0$, $\forall x \in \mathcal{C}^{(j)}, \forall y \in X \setminus \mathcal{C}^{(j)}$.*

Definition 5. *For a finite set of nodes X and sampling set S, we define the corresponding* ***Vandermonde*** *matrix as*

$$V(X; S) = \left[e^{ikx}\right]_{k \in S}^{x \in X}.$$

Let $\sigma_1(L) \geq \sigma_2(L) \geq \cdots \geq \sigma_n(L)$ denote the singular values of a matrix $L \in \mathbb{C}^{n \times n}$ and let $\lambda_1(L) \geq \lambda_2(L) \geq \cdots \geq \lambda_n(L)$ denote it's eigenvalues.

III. Identification of an Optimal Rate

In this section, we establish a measure for approximating the minimal separation Δ_ρ between decimated nodes with decimation parameter ρ. For M clusters, we prove in proposition 1 that $\sigma_{M+1}(T_\rho) \asymp \Delta_\rho^2$, where $\asymp$ denotes asymptotic scaling.

A. Main result and proofs

Let $\hat{\mu}_k := \hat{\mu}(k) = \sum_{j=1}^n a_j e^{ikx_j}$ be the Fourier measurement of model (1). For a set of nodes $X := \{x_1, ..., x_n\}$, the Toeplitz matrix of the samples is defined as:

$$T = T(X; n) := \begin{bmatrix} \mu_{n-1} & \mu_{n-2} & \cdots & \mu_0 \\ \vdots & \vdots & & \vdots \\ \mu_{2n-1} & \mu_1 & \cdots & \mu_{n-1} \end{bmatrix}. \quad (3)$$

and it can be decomposed into:

$$T = \tilde{V}_n \underbrace{diag(a_1, \ldots, a_n)}_{A} V_n^*,$$

where

$$\tilde{V}_n := V(X; \{n-1, \ldots, 2n-2\}),$$
$$V_n := V(X; \{0, \ldots, n-1\}).$$

We have

$$\tilde{V}_n := V_n diag(e^{i(n-1)x_1}, ..., e^{i(n-1)x_n}) = V_n E,$$

then we can write

$$T = V_n D V_n^*, \quad D = EA. \quad (4)$$

Theorem 1. *Let $Q \in \mathbb{C}^{n \times n}$ be a matrix of the form $Q = V^* D V$, where D is a diagonal complex matrix and $V \in \mathbb{C}^{n \times n}$. Then for $i = 1, \ldots, n$ we have:*

$$\sigma_i(Q) = \left|\lambda_i(Q)\right| = \theta_i \lambda_i(VV^*), \quad |D| := (DD^*)^{\frac{1}{2}}. \quad (5)$$

where $\lambda_n((DD^*)^{\frac{1}{2}}) \leq \theta_i \leq \lambda_1((DD^*)^{\frac{1}{2}})$.

Proof.

$$\begin{aligned} \left|\lambda_i(V^*DV)\right|^2 &= \lambda_i(V^*DV)\bar{\lambda}_i(V^*DV) \\ &= \lambda_i(V^*DV)\lambda_i(V^*D^*V). \end{aligned}$$

from properties of eigenvalues:

$$\mu_i := \lambda_i(V^*DV) = \lambda_i(VV^*D) = \lambda_i(D^{\frac{1}{2}}VV^*D^{\frac{1}{2}}),$$
$$\bar{\mu}_i := \lambda_i(V^*D^*V) = \lambda_i(VV^*D^*) = \lambda_i((D^*)^{\frac{1}{2}}VV^*(D^*)^{\frac{1}{2}}).$$

and thus

$$D^{\frac{1}{2}}VV^*D^{\frac{1}{2}}u_i = \mu_i u_i, \;\; u_i^*(D^*)^{\frac{1}{2}}VV^*(D^*)^{\frac{1}{2}} = \bar{\mu}_i u_i^*.$$

Note that these matrices are not necessarily Hermitian.

$$u_i^*(D^*)^{\frac{1}{2}}VV^*(D^*)^{\frac{1}{2}}D^{\frac{1}{2}}VV^*D^{\frac{1}{2}}u_i = \bar{\mu}_i\mu_i u_i^* u_i,$$
$$u_i^*(D^*)^{\frac{1}{2}}VV^*|D|VV^*D^{\frac{1}{2}}u_i = \bar{\mu}_i\mu_i u_i^* u_i,$$
$$u_i^*(D^*)^{\frac{1}{2}}VV^*D^{\frac{1}{2}}(D^*)^{\frac{1}{2}}VV^*D^{\frac{1}{2}}u_i = \bar{\mu}_i\mu_i u_i^* u_i,$$
$$\frac{\|(D^*)^{\frac{1}{2}}VV^*D^{\frac{1}{2}}u_i\|_2^2}{\|u_i\|_2^2} = \mu_i\bar{\mu}_i = \sigma_i^2((D^*)^{\frac{1}{2}}VV^*D^{\frac{1}{2}}),$$
$$\sigma_i((D^*)^{\frac{1}{2}}VV^*D^{\frac{1}{2}}) = \lambda_i((D^*)^{\frac{1}{2}}VV^*D^{\frac{1}{2}}).$$

Where the last equivalence is true because $(D^*)^{\frac{1}{2}}VV^*D^{\frac{1}{2}}$ is a positive semi-definite hermitian matrix. By Ostrowski's Theorem for Hermitian matrices (Theorem 4.5.9 in [18]), we have

$$\lambda_i(D^{\frac{1}{2}}VV^*(D^*)^{\frac{1}{2}}) = \theta_i\lambda_i(VV^*),$$

where

$$\lambda_n((DD^*)^{\frac{1}{2}}) \leq \theta_i \leq \lambda_1((DD^*)^{\frac{1}{2}}).$$

Lastly we have, $\left|\lambda_i(V^*DV)\right| = \sqrt{\lambda_i(V^*DV) \cdot \bar{\lambda}_i(V^*DV)} = \sqrt{\lambda_i(V^*D^*VV^*DV)} = \sigma_i(V^*DV)$. □

From now on, we choose X to form a $((h^{(j)}, \nu^{(j)}, n^{(j)})_{j=1}^M, \eta)$-clustered configuration, as defined in Definition (4), with $h^{(1)} = h^{(2)} = \cdots = h^{(M)} = \Delta(X)$.

Theorem 2. *There exist C_1, C_2 depending on $s := \max_j n^{(j)}$ such that for $\eta \geq C_1$ and $\Delta \leq \frac{C_2}{n^{2\nu}}, \nu := \max_k \nu^{(k)}$, for each $m = 1, ..., s$ there are precisely $\ell_m := \#\{1 \leq k \leq M : m \leq n^{(k)}\}$ singular values of $V_n := V_n(X)$ bounded below by*

$$C\Delta^{m-1},$$

where C doesn't depend on Δ.

Proof. Recall the following Theorem:

Theorem 3. *[Proposition 7.1 in [4]] Let $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_n$ denote the union of all the singular values of the matrices $V_n^{(j)} := V(\mathcal{C}^{(j)}; \{0, \ldots, n-1\})$ in non-increasing order, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ denote the singular values of V_n. Then*

there exist $C_1(s), C_2(s)$ depending only on $s := \max_j n^{(j)}$ such that for $\eta \geq C_1$ and $\Delta \leq \frac{C_2}{n^2\nu}, \nu := \max_k \nu^{(k)}$ we have

$$\sigma_j \geq \frac{1}{2}\tilde{\sigma}_j, \quad j = 1, ..., n.$$

We have to find the scaling of the singular values of each $V_n^{(j)}$ to finish the proof. For simplicity, assume that $n = 2p+1$. Thus we can write,

$$U_p^{(j)} = \left[e^{ikx}\right]_{k\in\mathcal{G}_p}^{x\in\mathcal{C}^{(j)}}.$$

where $\mathcal{G}_p := \{-p, -p+1, ..., 0, ..., p\}$ is the symmetric one-dimensional grid. Let $G_j := \left(U_p^{(j)}\right)^* U_p^{(j)}$ for $j = 1, ..., M$. Now we use the following Lemma and write it for the univariate case:

Lemma 1 (Lemma 3 in [13])**.** *Let $V_{\leq m}(\mathcal{G}) := [g^k]_{g\in\mathcal{G}}^{0\leq k\leq m}$. Then the sampling set $\mathcal{G}_p$ satisfies the condition $rank(V_{\leq n^{(j)}-1}(\mathcal{G}_p)) \geq n^{(j)}$ and the eigenvalues of G_j split into $n^{(j)}$ groups (each group has one element):*

$$\lambda_0 = \Delta^0(\tilde{\lambda}_0 + O(\Delta)), \quad \lambda_1 = \Delta^2(\tilde{\lambda}_1 + O(\Delta)), ...,$$
$$\lambda_{n^{(j)}-1} = \Delta^{2(n^{(j)}-1)}(\tilde{\lambda}_{n^{(j)}} + O(\Delta)).$$

where $\tilde{\lambda}_i$ doesn't depend on Δ.

Since $\left(V_n^{(j)}\right)^* V_n^{(j)} = (D_j^{-1})^* G_j D_j^{-1}$ for $D_j := diag(e^{-ipx})_{x\in\mathcal{C}^{(j)}} \in \mathbb{C}^{n^{(j)}\times n^{(j)}}$, we have that $\sigma_k\left(V_n^{(j)}\right) = \sqrt{\lambda_k\left(G_j\right)} \asymp \Delta^k$. Finally, for each $k = 1, ..., s$ we have ℓ_k singular values that scale like Δ^k, obtained from each cluster. Combining Theorem 3, we finish the proof of Theorem 2. □

Theorem 4. *The singular values of the Vandermonde matrix V_n scale as follows:*

$$\{\sigma_{1,i}(V_n)\}_{i=1}^{\ell_1} \asymp \Delta^0, \ldots, \{\sigma_{s,i}(V_n)\}_{i=1}^{\ell_s} \asymp \Delta^{s-1},$$

where $s := \max_k n^{(j)}$.

Proof. The determinant of the Vandermonde matrix V_n is:

$$det(V_n) = \prod_{k\neq j} |e^{ix_j} - e^{ix_k}|.$$

Using Lemma 5.6 in [5], for $|x - x'| \leq \frac{\pi}{2}$ we have

$$\frac{2}{\pi}|x - x'| \leq |e^{ix} - e^{ix'}| \leq |x - x'|.$$

Thus we get that

$$det(V_n) = \prod_{j=1}^{M} \prod_{x,y\in\mathcal{C}^{(j)}} |e^{ix} - e^{iy}| \prod_{j\neq i} \prod_{x\in\mathcal{C}^{(j)}, y\in\mathcal{C}^{(i)}} |e^{ix} - e^{ix'}|$$
$$= C(\eta, \nu, n) \prod_{j=1}^{M} \prod_{x,y\in\mathcal{C}^{(j)}} |e^{ix} - e^{iy}| = C\Delta^{\sum_{j=1}^{M}\binom{n^{(j)}}{2}}.$$

Now we show that

$$\sum_{j=1}^{M}\binom{n^{(j)}}{2} = \sum_{j=1}^{M}\frac{n^{(j)}(n^{(j)}-1)}{2} = \sum_{j=1}^{M}\sum_{i=1}^{n^{(j)}}(i-1)$$
$$= \sum_{i=1}^{s}\ell_i(i-1).$$

where the last step is due to the definition of ℓ_j. Thus we have that $det(V_n) = C \cdot \Delta^{\sum_{i=1}^{s}\ell_i(i-1)}$. It's known that $|det(V_n)| = \prod_{i=1}^{n}\sigma_i(V_n)$. Thus on the one hand we have $\prod_{i=1}^{n}\sigma_i(V_n) = C \cdot \Delta^{\sum_{i=1}^{s}\ell_i(i-1)}$ and on the other hand (by Theorem 2) we have that

$$\prod_{i=1}^{n}\sigma_i(V_n) \geq \prod_{i=1}^{s}\prod_{k=1}^{\ell_k} C_{i,k}\sigma_{i,k}(V_n) = \prod_{i=1}^{s}\prod_{k=1}^{\ell_k} C_{i,k}\Delta^{i-1}$$
$$= \tilde{C}\Delta^{\sum_{i=1}^{s}\ell_i(i-1)}.$$

Since C doesn't depend on Δ, this forces the singular values to have exact scaling (i.e. $C_{i,k}$ don't depend on Δ). □

Now we state our main result.

Proposition 1. *For any $\rho > 0$, the singular values of the Toeplitz matrix $T_\rho := T(\rho X; n)$ scales as follows:*

$$\{\sigma_{1,i}(T_\rho)\}_{i=1}^{\ell_1} \asymp \Delta_\rho^0, \ldots, \{\sigma_{s,i}(T_\rho)\}_{i=1}^{\ell_s} \asymp \Delta_\rho^{2(s-1)},$$

where $s := \max_j n^{(j)}$ and $\Delta_\rho := \Delta(\rho X)$.

Proof. Combining Theorems 1 and 4, we get the desired result. □

Remark 1. *The constants in Proposition 1 also depend on the amplitudes $\{a_j\}_{j=1}^{n}$.*

B. Numerical Validation

Let $k \in \mathbb{N}$ be the number of clusters in a configuration with nodes in X. To validate Proposition 1, we plotted the minimal separation of decimated nodes for all decimation parameters ρ in the interval $\mathcal{I} := \left[\frac{1}{2}\frac{\Omega}{2n-1}, \frac{\Omega}{2n-1}\right]$ against the $k+1$-th singular value of the Toeplitz matrix T_ρ. Figure 1 confirms that $\sigma_{k+1}(T_\rho)$ scales proportionally to Δ_ρ^2.

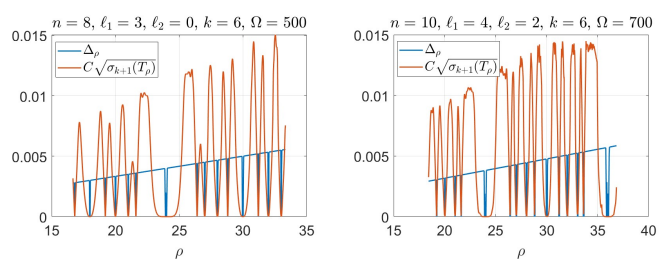

Fig. 1: (left) X has one cluster of size ℓ_1 with $SRF = 6$. (right) X has two clusters of sizes ℓ_1 and ℓ_2 with $SRF = 3$. The plotted $\sqrt{\sigma_{k+1}(T_\rho)}$ values are scaled by $C := \frac{n}{\Omega}$ for better visualization.

Remark 2. *We expect our estimates in Proposition 1 to hold for noisy samples, using standard perturbation analysis for singular values. We leave it for future work.*

IV. Algorithm

A decimation parameter $\lambda \in \mathcal{I}$ is said to be admissible if it attains good separation properties, in particular, when the cluster nodes are separated by at least $O(\Omega\Delta)$ and the non-cluster nodes by $\geq \frac{1}{2n^2}$ (for the full details, see Proposition 5.8 in [5]).

A. Selecting an optimal sampling rate

By Proposition 1, for any $\rho \in \mathbb{N}$, we have $\sigma_{M+1}(T_\rho) \asymp \Delta_\rho^2$, where M is the number of clusters with multiplicity at least 1. For a set $\Lambda = \{\rho_j,\}_{j=1}^m \subset \mathbb{N}$, we select $\rho = \arg\max_j\{\sigma_{M+1}(T_{\rho_j})\}_{j=1}^m$ as the optimal decimation parameter with the best separation properties.

Since decimation alone is insufficient for recovery, any SR method applied to samples at rate ρ gives nodes $\{e^{i\rho x_j}\}_{j=1}^n$, with each $e^{i\rho x_j}$ having ρ candidates. To resolve this ambiguity, we use the technique from [10] for de-aliasing, which employs a second shifted sample set $\mathcal{M}_{ds} := \{\mu(\rho k + t)\}_{k=0}^{2n-1}$ alongside $\mathcal{M}_d := \{\mu(\rho k)\}_{k=0}^{2n-1}$, where ρ and t are co-prime. Let $\Phi_j := \exp(ix_j)$. In the noiseless case, we have $\mu(\rho k) = \sum_{j=1}^n a_j \Phi_j^{\rho k}$ and $\mu(\rho k + t) = \sum_{j=1}^n (a_j \Phi_j^t)\Phi_j^{\rho k}$. To recover $\{\Phi_j^t\}_{j=1}^n$, we first recover the amplitudes $\{a_j\}_{j=1}^n$ and $\{a_j\Phi_j^t\}_{j=1}^n$ from the sampling sets $\mathcal{M}_d$ and $\mathcal{M}_{ds}$, respectively. Then compute $\{\Phi_j^t\}_{j=1}^n = \left\{\frac{a_j\Phi_j^t}{a_j}\right\}_{j=1}^n$. Finally, since ρ and t are co-prime, the intersection of the candidate sets for $e^{i\rho x_j}$ and e^{itx_j} contains exactly one element. For noisy samples, we first match the aliased nodes $\{\tilde{\Phi}_j^\rho\}$ and $\{\hat{\Phi}_j^\rho\}$ before division. A more stable solution can be achieved using additional shifted sample sets; see [10], [8] for the full details.

Summarizing the above, Algorithm 1 can be applied to any SR method to generate its decimated version.

Algorithm 1 Decimated SR method

Require: M, $\{\hat{\mu}(\omega)\}_{\omega\in[-\Omega,\Omega]}$, Ω, n, SRmethod$(\cdot)$.

Select Admissible Decimation parameter:
- (a) Select $\rho := \arg\max_{\rho'\in\mathcal{I}\cap\mathbb{N}}\{\sigma_{M+1}(T_{\rho'})\}$.
- (b) Find t such that ρ and t are co-prime.

Solve for Decimated and Shifted samples:
- (a) $\{\tilde{\Phi}_j^\rho, \tilde{a}_j\}_{j=1}^n =$ SRmethod$(\mathcal{M}_d)$.
- (b) $\{\hat{\Phi}_j^\rho, \hat{a}_j\hat{\Phi}_j^t\}_{j=1}^n =$ SRmethod$(\mathcal{M}_{ds})$.

Perform De-aliasing for $\{\tilde{\Phi}_j^\rho, \hat{\Phi}_j^t\}$ (see the text above).

Return the estimates $\{\tilde{x}_j, \tilde{a}_j\}$.

Remark 3. *Using Proposition 5.8 from [5], we can derive a lower bound of $\frac{1}{n^2}$ for selecting an admissible decimation parameter from $\mathcal{I}$, enabling randomness in algorithm 1.*

We introduce the Enhanced Decimated Prony Method (EDP) and the Decimated Matrix Pencil method (DMP) as improvements over DP and MP, respectively. EDP applies Algorithm 1 with Prony's method as the SR method, eliminating the need to solve multiple sub-Prony problems and constructing a histogram, as done in DP. Similarly, DMP applies Algorithm 1 with MP as the SR method.

DP and MP have a time complexity of $O(\Omega^2) + O(\Delta^{-1})$ and $O(\Omega^3)$, respectively. For EDP and DMP, there are $O(\Omega)$ natural decimation parameters in the interval $\mathcal{I}$. For each, we construct an $n\times n$ Toeplitz matrix of samples and compute its $M+1$-th singular value, which requires $O(n^3\Omega)$. Finding a co-prime $t \neq 1$ for ρ has a worst-case complexity of $O(\Omega\log(\Omega))$, as gcd computations take $O(\log(\Omega))$, and the search might iterate up to $\Omega - 1$. Since matching the aliased nodes, applying Prony's method, MP (for $2n$ samples) and de-aliasing depend only on n (fixed), the total time complexity of EDP and DMP is $O(\Omega\log(\Omega))$.

We implemented the EDP, DP, DMP and MP methods, assigning the number of nodes for both MP and DMP as an input and using $3n$ decimated samples in DMP. The implementation is in MATLAB. We recall that for DP, N_ρ is the number of decimation parameters being tested in the interval $\mathcal{I}$ and N_b is the number of bins in the constructed histogram of node candidates. The noise in the samples is generated from a Cauchy distribution. Each method was tested 10 times across four different SRF values. In Figure 2, we present the mean reconstruction error of the cluster node x_1 as a function of SRF. The results show that all methods achieve high accuracy, with EDP and DMP being the fastest, offering a speed improvement of up to two orders of magnitude.

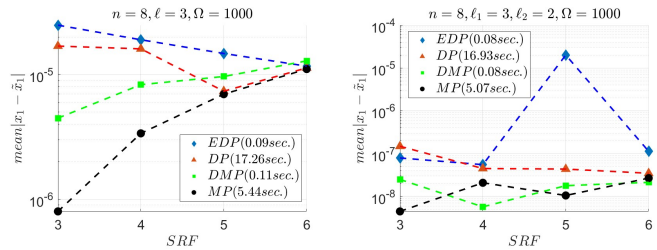

Fig. 2: (left) single cluster configuration. (right) multi-cluster configuration. For both experiments the noise level is 10^{-6}, $N_\rho = 900$ and $N_b = 3\Delta^{-1}$.

Remark 4. *To obtain higher accuracy for DMP, we can use $N \gg 2n$ decimated samples if possible.*

B. Optimality of EDP

In Figure 3, we numerically show that EDP is optimal, meaning it achieves the min-max error bounds (Theorem 2.8 in [5]) in the multi-cluster geometry. We plot the node/amplitude error amplification factors (6) as a function of SRF.

$$\mathcal{K}_{x_j} := \epsilon^{-1}\Omega|x_j - \tilde{x}_j|, \quad \mathcal{K}_{a_j} := \epsilon^{-1}|a_j - \tilde{a}_j|. \tag{6}$$

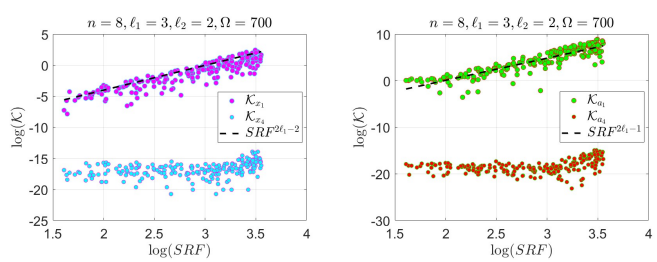

Fig. 3: EDP - asymptotic optimality. For cluster node x_1, $\mathcal{K}_{x_1}$ (left) scales like $SRF^{2\ell_1-2}$, while the $\mathcal{K}_{a_1}$ (right) scales like $SRF^{2\ell_1-1}$. For the non-cluster node x_4, both $\mathcal{K}_{x_4}$ and $\mathcal{K}_{a_4}$ are lower bounded by a constant. These scaling rates are optimal.

Acknowledgment. We thank Dr. Dmitry Batenkov and Prof. Ronen Talmon for their insightful comments.

References

[1] Habib Ammari, Ekaterina Iakovleva, and Dominique Lesselier. A MUSIC algorithm for locating small inclusions buried in a half-space from the scattering amplitude at a fixed frequency. *Multiscale Modeling & Simulation*, 3(3):597–628, 2005.

[2] Dmitry Batenkov and Nuha Diab. Super-resolution of generalized spikes and spectra of confluent Vandermonde matrices. *Applied and Computational Harmonic Analysis*, 65:181–208, 2023.

[3] Dmitry Batenkov, Benedikt Diederichs, Gil Goldman, and Yosef Yomdin. The spectral properties of Vandermonde matrices with clustered nodes. *Linear Algebra and its Applications*, 609:37–72, 2021.

[4] Dmitry Batenkov and Gil Goldman. Single-exponential bounds for the smallest singular value of Vandermonde matrices in the sub-Rayleigh regime. *Applied and Computational Harmonic Analysis*, 55:426–439, 2021.

[5] Dmitry Batenkov, Gil Goldman, and Yosef Yomdin. Super-resolution of near-colliding point sources. *Information and inference*, 10(2):515–572, 2021.

[6] Mario Bertero and Christine De Mol. III Super-resolution by data inversion. In *Progress in optics*, volume 36, pages 129–178. Elsevier, 1996.

[7] Ayush Bhandari and Ramesh Raskar. Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-D imaging. *IEEE Signal Processing Magazine*, 33(5):45–58, 2016.

[8] Matteo Briani, Annie Cuyt, Ferre Knaepkens, and Wen-shin Lee. VEXPA: Validated EXPonential Analysis through regular sub-sampling. *Signal processing*, 177:107722–, 2020.

[9] Annie Cuyt, Yuan Hou, and Wen shin Lee. Validated analysis of modulated signals: From de Prony to Padé and beyond. *Journal of Computational and Applied Mathematics*, 413:114346, 2022.

[10] Annie Cuyt and Wen-shin Lee. How to get high resolution results from sparse and coarsely sampled data. *Applied and Computational Harmonic Analysis*, 48(3):1066–1087, 2020.

[11] Laurent Demanet and Nam Nguyen. The recoverability limit for superresolution via sparsity. *arXiv preprint arXiv:1502.01385*, 2015.

[12] Nadiia Derevianko, Gerlind Plonka, and Markus Petz. From ESPRIT to ESPIRA: estimation of signal parameters by iterative rational approximation. *IMA Journal of Numerical Analysis*, 43(2):789–827, 2023.

[13] Nuha Diab and Dmitry Batenkov. Spectral Properties of Infinitely Smooth Kernel Matrices in the Single Cluster Limit, with Applications to Multivariate Super-Resolution. *arXiv preprint arXiv:2407.10600*, 2024.

[14] Benedikt Diederichs. *Sparse frequency estimation: Stability and algorithms*. PhD thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky, 2018.

[15] D.L. Donoho. Superresolution via Sparsity Constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.

[16] Zhiyuan Fan and Jian Li. Efficient algorithms for sparse moment problems without separation. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 3510–3565. PMLR, 2023.

[17] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Nearly optimal sparse Fourier transform. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 563–578, 2012.

[18] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[19] Yingbo Hua and Tapan K Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):814–824, 1990.

[20] Piotr Indyk and Michael Kapralov. Sample-optimal sparse Fourier transform in any constant dimension. In *FOCS*, 2014.

[21] Rami Katz, Nuha Diab, and Dmitry Batenkov. Decimated Prony's method for stable super-resolution. *IEEE Signal Processing Letters*, 2023.

[22] Rami Katz, Nuha Diab, and Dmitry Batenkov. On the accuracy of Prony's method for recovery of exponential sums with closely spaced exponents. *Applied and Computational Harmonic Analysis*, 73:101687, 2024.

[23] Stefan Kunis and Dominik Nagel. On the condition number of Vandermonde matrices with pairs of nearly-colliding nodes. *Numerical Algorithms*, 87:473–496, 2021.

[24] Weilin Li, Wenjing Liao, and Albert Fannjiang. Super-resolution limit of the ESPRIT algorithm. *IEEE transactions on information theory*, 66(7):4593–4608, 2020.

[25] Jari Lindberg. Mathematical concepts of optical superresolution. *Journal of Optics*, 14(8):083001, 2012.

[26] Ping Liu and Hai Zhang. A theory of computational resolution limit for line spectral estimation. *IEEE Transactions on Information Theory*, 67(7):4812–4827, 2021.

[27] Veniamin I Morgenshtern and Emmanuel J Candes. Super-resolution of positive sources: The discrete setup. *SIAM Journal on Imaging Sciences*, 9(1):412–444, 2016.

[28] Satish Mulleti, Amrinder Singh, Varsha P Brahmkhatri, Kousik Chandra, Tahseen Raza, Sulakshana P Mukherjee, Chandra Sekhar Seelamantula, and Hanudatta S Atreya. Super-resolved nuclear magnetic resonance spectroscopy. *Scientific reports*, 7(1):9651, 2017.

[29] Gerlind Plonka and Therese von Wulffen. Deterministic sparse sublinear FFT with improved numerical stability. *Results in Mathematics*, 76(2):53, 2021.

[30] Gerlind Plonka, Katrin Wannenwetsch, Annie Cuyt, and Wen-shin Lee. Deterministic sparse FFT for M-sparse vectors. *Numerical algorithms*, 78:133–159, 2018.

[31] Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.