

Decimated Prony's Method for Stable Super-Resolution

Rami Katz, Nuha Diab, and Dmitry Batenkov

Abstract—We study recovery of amplitudes and nodes of a finite impulse train from noisy frequency samples. This problem is known as super-resolution under sparsity constraints and has numerous applications. An especially challenging scenario occurs when the separation between Dirac pulses is smaller than the Nyquist-Shannon-Rayleigh limit. Despite large volumes of research and well-established worst-case recovery bounds, there is currently no known computationally efficient method which achieves these bounds in practice. In this work we combine the well-known Prony’s method for exponential fitting with a recently established decimation technique for analyzing the super-resolution problem in the above mentioned regime. We show that our approach attains optimal asymptotic stability in the presence of noise, and has lower computational complexity than the current state of the art methods.

Index Terms—Prony’s method, decimation, sparse super-resolution, direction of arrival, sub Nyquist sampling, finite rate of innovation, exponential fitting.

I. INTRODUCTION

VARIOUS problems of signal reconstruction in multiple basic and applied settings can be reduced to recovering the amplitudes \( \{\alpha_k\}_{k=1}^n \) and nodes \( \{x_k\}_{k=1}^n \) of a finite impulse train

\[
f(x) = \sum_{k=1}^n \alpha_k \delta(x - x_k)
\]

from band-limited and noisy spectral measurements

\[
g(\omega) = \sum_{k=1}^n \alpha_k e^{2\pi i x_k \omega} + e(\omega), \quad \omega \in [-\Omega, \Omega],
\]

where \( \Omega > 0 \) and \( \|e\|_\infty \leq \epsilon \) for some \( \epsilon > 0 \). Due to its widespread applications, this problem has been studied under various guises including tauberian approximation [1], parametric spectrum estimation and direction of arrival [2], [3], time-delay estimation [4], sparse deconvolution [5], super-resolution (SR) [6], [7] and finite-rate-of-innovation sampling [8], [9]. Beyond the theoretical modelling, recent advances have shown (1) to be the work-horse for emerging areas such as super-resolution tomography and spectroscopy [10], [11], ultra-fast time-of-flight imaging [12] and unlimited sensing [13]; in such cases, efficient and robust solutions to (1) entail pushing real-world capabilities beyond the possibilities of conventional hardware.

Despite the theoretical advances on this topic, there are still fundamental research gaps that arise due to ill-posedness and instability of (1) in the presence of noise. In particular, an especially challenging regime occurs when the separation \( \Delta \) between two or more nodes is smaller than the Nyquist-Shannon-Rayleigh (NSR) limit \( 1/\Omega \). Recently, min-max error bounds for SR in the noisy regime were derived in the case when some nodes form a dense cluster [14], [15], [16], establishing the fundamental limits of recovery in the SR problem (cf. Section II). The analysis therein relies on a theoretical decimation technique [17], [18], [19], [20], [21]. Existence of an “admissible” decimation parameter was proved in [15] by measure-theoretic arguments, while obtaining such a parameter in practice remained open. The accuracy of Prony’s method [22] was established in [23], proving its optimality for fixed \( \Omega \) and \( \Delta \leq 1 \). However, a provably optimal and tractable algorithm in the general case of varying \( \Omega \) has been missing from the literature. In particular, the common ESPRIT algorithm [24] is known to only be almost-optimal, whereas no theoretical guarantees exist for the Matrix Pencil (MP) algorithm in the SR regime [15]. The main contributions of the present work are as follows.

1) We develop a novel tractable decimation-based algorithm for SR, dubbed the Decimated Prony’s method (DPM), which finds an admissible decimation parameter by solving multiple decimated sub-problems and a suitable binning procedure (cf. Section III).

2) We prove that DPM achieves optimal recovery in the case of a single cluster, making it the first algorithm provably attaining min-max bounds in this regime. For multiple clusters, numerical simulations provide evidence of optimal asymptotic stability and noise tolerance of DPM. Furthermore, we show that DPM has lower computational complexity than the ESPRIT algorithm.

II. TOWARDs OPTIMAL ALGORITHMS

Throughout the article we consider the number of nodes (resp. amplitudes) \( n \in \mathbb{N} \) (in 1) to be fixed. We assume that the nodes satisfy \( \{x_k\}_{k=1}^n \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right] \), without loss of generality (see Section IV in [15]). Let \( \{\tilde{x}_k\}_{k=1}^n \) and \( \{\tilde{\alpha}_k\}_{k=1}^n \) be the approximated parameters obtained via a theoretical reconstruction algorithm using the data (1).

Definition 1 ([15]): Let \( \{\alpha_k, x_k\}_{k=1}^n \subseteq U \). Given \( \epsilon > 0 \), the min-max recovery rates are

\[
\Lambda^{\epsilon}_{x, U, \Omega} = \inf_{A: g \rightarrow (\alpha_j, x_j)} \sup_{\{\alpha_j, x_j\}} \sup_{\|e\|_\infty \leq \epsilon} \|\cdot_j - \hat{\cdot}_j\|
\]



Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
II.1. \( \{x_k\}_{k=1}^n \) is a clustered configuration if there is a partition of the nodes into clusters such that the distances between the nodes in each cluster are on the order of \( \Delta \), while the inter-cluster distances are on the order of \( 1/\Omega \) (well-separated clusters) [14, 15].

**Theorem 1** ([15]): Let the super-resolution factor satisfy \( \text{SRF} := \frac{1}{\Omega} \geq 1 \), and \( \{\{x_k\}_{k=1}^n\} \) be (uniformly) bounded. Assume a single cluster \( \mathcal{X} \) of size \( 1 < \ell < n \), whereas all other (singleton) clusters are well-separated, and \( \epsilon \lesssim \Omega(\Delta^{2\ell-1}) \). Then

\[
\begin{align*}
\Lambda_{\epsilon,\mathcal{U},\Omega}^{\ell,j} &:= \frac{\text{SRF}^{2\ell-2}}{\epsilon} \quad x_j \in \mathcal{X}, \quad x_j \notin \mathcal{X}, \\
\Lambda_{\epsilon,\mathcal{U},\Omega}^{\ell,j} &:= \frac{\text{SRF}^{2\ell-2}}{\epsilon} \quad x_j \in \mathcal{X}, \quad x_j \notin \mathcal{X}.
\end{align*}
\]

Here, \( \lesssim \approx \) denote asymptotic inequalities/equivalence up to multiplying constants independent of \( \Omega, \Delta, \epsilon \), while \( U \) is the set of signals satisfying the assumptions above.

The min-max bounds establish fundamental recovery limits in any application modeled by (1) (e.g. [31]). Related results are known in the signal processing literature for Gaussian noise [27, 29, 32]. Despite a plethora of methods to solve this problem, up to date a provably tractable algorithm which achieves the min-max rates is missing from the literature. In particular, the widely used ESPRIT algorithm is sub-optimal, both in terms of the bounds and the threshold SNR [24].

The proof of Theorem 1 is non-constructive and employs the following decimation technique. Let \( \mathcal{J} := \left\{ \frac{1}{2^{2\ell-1}}, \frac{1}{2^{2\ell-2}} \right\} \). For a decimation parameter \( \lambda \in \mathcal{J} \), the measurements \( \{g(\lambda k)\}_{k=0}^{2n-1} \) yield the problem (1) with \( \{e^{\ell \pi j x_j}\}_{j=1}^n \) replaced by \( \{e^{2\ell \pi j x_j}\}_{j=1}^n \). By [15, Prop. 5.8] there exists an interval \( \mathcal{I} \subset \mathcal{J} \) of length \( |\mathcal{I}| \geq c \epsilon \) such that every \( \lambda \in \mathcal{I} \) satisfies \( |e^{2\ell \pi j x_j} - e^{2\ell \pi j x_j}| \geq \epsilon^{-2}, \) whenever \( x_j, x_j \) belong to different clusters (collision avoidance) [15, Prop. 5.12]. For a collision-avoiding \( \lambda \), the condition number of the (theoretical) solution map \( P_n \) which inverts (1) in the case \( e \equiv 0 \) and \( \omega \in \{\{x_j\}_{j=1}^n\} \) matches the min-max rates. Set \( \{\lambda_j, e^{2\ell \pi j x_j}\}_{j=1}^n \). Let the set of all aliased solutions corresponding to \( \{\tilde{y}_j\}_{j=1}^n \) be

\[
X_{\lambda} := \left\{ (\lambda, t) : t = \frac{\tilde{y}_j + m}{\lambda}, \ m \in \mathbb{Z}, |t| \leq \frac{1}{2} \right\},
\]

(2)

where the aliasing follows by periodicity of \( y \rightarrow e^{2\ell \pi j y} \). The arguments above imply that \( X_{\lambda} \) contains at least one element \( (\lambda, t) \) with \( t \approx x_j \) for each \( j = 1, \ldots, n \) (call it Property P*) and suggest the following heuristic recovery procedure:

- Find a collision-avoiding \( \lambda \in \mathcal{J} \);
- Compute \( \{\tilde{y}_j, e^{2\ell \pi j x_j}\}_{j=1}^n = P_n\{g(\lambda k)\}_{k=0}^{2n-1} \) with optimal stability/accuracy;
- Find \( \{\lambda, x_j\}_{j=1}^n \subset X_{\lambda} \) s.t. \( x_j \approx \tilde{x}_j \) (dealiasing).

Since the analysis in [15] provides no constructive method of obtaining a collision-avoiding \( \lambda \), any conversion of [15] into a tractable algorithm must tackle this challenge.

In this work we propose a constructive algorithm aimed at detecting a collision-avoiding \( \lambda \) and achieving dealiasing. For that purpose, we solve multiple decimated sub-problems and employ a suitable binning procedure. For the middle step, we propose to use Prony’s method [22] (Algorithm II.1), which provides an exact solution to the problem in the noiseless regime. The use of Prony’s method is motivated by our recent results in [23] which prove its optimality in the regime \( \Delta \lesssim 1 \) and \( \Omega \) fixed (corresponding to SRF \( \lesssim 1 \) in Theorem 1).

**Theorem 2** ([23]): Suppose \( \ell_\ast \) is the largest cluster size, and each \( x_j \) belongs to a cluster of size \( \ell_j \). For \( \epsilon < \Omega^{2\ell_j - 1} \), the output of Algorithm II.1 satisfies \( |x_j - \tilde{x}_j| \lesssim \Delta^{2\ell_j - 1} \epsilon \) for the nodes, and \( |\alpha_j - \tilde{\alpha}_j| \lesssim \Delta^{\frac{\ell_j}{2}} \epsilon \) for the amplitudes, where \( \ell_j = 1 - 2\ell_j \) if \( \ell_j > 1 \), and \( \ell_j = 0 \) if \( \ell_j = 1 \).

Here we consider the more general case of varying \( \Omega \). Hence, [23] is insufficient to justify the proposed algorithm. We prove that our algorithm achieves optimal recovery in the case of a single cluster. For multiple clusters, we perform numerical simulations that show optimal asymptotic stability and noise tolerance of DPM, as they appear in [15].

Define the node/amplitude error amplification factors

\[
K_{x, j} := -\Omega |x_j - \tilde{x}_j|, \quad K_{\alpha, j} := -|\alpha_j - \tilde{\alpha}_j|.
\]

Fixing \( \Omega = 2n - 1 \) for \( n \in \{3, 4, 5\} \) and \( \ell_\ast \in \{2, 3, 5\} \), Fig. 1 shows that both \( K_{x, j} \) and \( K_{\alpha, j} \) computed by Algorithm II.1 scale as the min-max rates above.
Algorithm II.1: The Classical Prony Method.

**Input:** Sequence \( \{\hat{m}_k\} \equiv g(k) 2^{n-1} \)

**Output:** Estimates \( \hat{x}_k, \hat{\alpha}_k \)

**Notation:** \( \text{col} \{y_k\}^n_{k=1} = [y_1, \ldots, y_n]^\top \in \mathbb{C}^m \)

1. Construct \( \tilde{H}_n = (\hat{m}_{i+j})_{0 \leq i,j \leq n-1} \)
2. Solve the linear least squares problem
   \[
   \text{col} \{q_k\}^n_{k=1} = \arg \min_{q \in \mathbb{C}^m} \|H_n q + \text{col} \{\hat{m}_k\}^n_{k=n} \|_2
   \]
3. Compute \( \tilde{z}_k \) as the roots of the (perturbed) Prony polynomial \( g(z) = z^n + \sum_{j=1}^n q_j z^j \).
4. Recover \( \tilde{x}_k \) from \( \tilde{z}_k \) via \( \tilde{x}_k = \frac{\text{Arg}(\tilde{z}_k)}{\pi} \).
5. Construct \( \tilde{V} = (\tilde{z}_k^i)_{k=0}^{n-1} \) and solve
   \[
   \text{col} \{\hat{\alpha}_k\}^n_{k=1} = \arg \min_{\alpha \in \mathbb{C}^m} \|V \alpha - \text{col} \{\hat{m}_k\}^n_{k=0} \|_2
   \]
6. return the estimated parameters \( \{\tilde{x}_k, \tilde{\alpha}_k\}^n_{k=1} \)

**Remark 1:** Throughout the article we follow [15] and consider \( x_k \) to be **successfully recovered** if \( |x_k - \tilde{x}_k| \) is smaller than one third of the distance between \( x_k \) and its nearest neighbor.

III. Decimated Prony’s Method

We develop the Decimated Prony’s Method (DPM) (Algorithm III.1). In order to identify a collision-avoiding \( \lambda \), we propose to solve multiple Prony problems and apply a binning procedure. We provide a description of DPM and its underlying reasoning.

Let \( G = \text{linspan}(J, N_\lambda) \) be the uniform grid of size \( N_\lambda \) (the choice of \( N_\lambda \) is motivated in Remark 1. See Fig. 2 for a numerical justification). For each \( \lambda \in G \) recall that \( X_\lambda \) in (2) is the set of aliased solutions corresponding to \( \lambda \). We consider the set \( \{x : (\lambda, x) \in \bigcup_{\lambda \in G} X_\lambda\} \) and compute its histogram with \( N_\lambda = 3\Delta^{-1} \) bins. This choice is motivated by the criterion for successful node recovery (see Remark 1).

We find the \( n \) bins \( \{B_j\}^n_{j=1} \) with largest counts. The elements in each of the bins are considered as candidates for a valid approximation of one of the nodes \( \{x_j\}^n_{j=1} \) and will be reduced further in the next step.

Existence of a collision-avoiding \( \lambda \) is guaranteed by [15]. In particular, if \( \lambda \) is not collision-avoiding, Property P* will not be satisfied since at least two nodes will be ill-conditioned.

Furthermore, [15, Prop. 5.17] suggests that if \( \lambda_1 \neq \lambda_2 \) are collision-avoiding then, with high probability, \((\lambda_1, t_1) \in X_{\lambda_1}\) and \((\lambda_2, t_2) \in X_{\lambda_2}\) with \( t_1 \approx t_2 \) implies \( t_1 \approx t_2 \approx x_j \) for some \( j = 1, \ldots, n \). Hence, collision-avoiding values of \( \lambda \) are expected to contribute exactly one candidate to each of the bins \( \{B_j\}^n_{j=1} \) and the set

\[
\Lambda := \bigcap_{k=1}^n \{\lambda : (\lambda, x) \in \bigcup_{\lambda \in G} X_\lambda \land x \in B_k\}
\]

should yield only collision-avoiding \( \lambda \)’s.

Assuming \( \Lambda \neq \emptyset \) (otherwise the algorithm fails), let \( \lambda^* = \max \Lambda \). This \( \lambda^* \) yields maximal in-cluster separation of \( \{e^{2\pi i \lambda^* x_j}\}^n_{j=1} \). The corresponding \( \{\tilde{x}_j\}^n_{j=1} \) are obtained by choosing \( \tilde{x}_j \in B_j \) s.t. \( \langle \lambda^*, \tilde{x}_j \rangle \in X_{\lambda^*} \). Finally, the amplitude approximations are found by solving a Vandermonde system.

To further motivate the algorithm, we prove (see Appendix) its correctness in case of a single cluster (i.e. \( \ell = n \)).

**Theorem 3:** In the notations of Theorem 1, suppose \( \ell = n \). Under a further technical assumption (to be elaborated in the proof), and with the choice of \( N_\lambda = O(\Omega) \) and \( N_\theta = 3\Delta^{-1} \), Algorithm III.1 attains the bounds of Theorem 1.

**Conjecture 1:** \( N_\lambda \gtrsim \Omega \) and \( N_\theta \gtrsim \Delta^{-1} \) ensure correctness of the algorithm in the general case. We leave the rigorous proof of this claim to future work. Note that increasing \( N_\lambda \) is expected to improve the robustness of the DPM.

Next, we analyze the time complexity of DPM by addressing each step of Algorithm III.1. We use the notation \( O = O_n \) (recall that \( n \) is fixed). The classical Prony’s method has complexity \( O(1) \), since it depends only on \( n \). For every \( \lambda \in G \) we apply Algorithm II.1 with the samples \( \{g(k\lambda)\} \) and compute \( X_\lambda \) in (2), which costs \( O(N_\lambda + \lambda N_\theta) \). Computing the histogram with \( N_\theta \) bins for data of size \( \sum_{k=1}^n m_k \) costs \( O(N_\theta \Omega + N_\theta) \). Finding the bins \( \{B_k\}^n_{k=1} \) with \( n \) largest counts costs \( O(N_\theta) \). Computing \( \Lambda \) in (3) and \( \lambda^* = \max \Lambda \), together with finding \( \{\tilde{x}_k\}^n_{k=1} \) costs \( O(N_\theta) \). Finally, solving an \( n \)-order Vandermonde system costs \( O(1) \) [33]. Thus, the total complexity is \( O(N_\theta \Omega + N_\theta) \). As mentioned in Remark 1, \( N_\lambda \gtrsim \Omega \) and \( N_\theta \gtrsim \Delta^{-1} \) are expected to be sufficient for correctness of DPM. This gives overall complexity \( O(\Omega^2) + O(\Delta^{-1}) \). For small \( \Omega \) the dominating factor is \( \Delta^{-1} \), in which case we may take \( N_\lambda = O(SRF) \) to have maximal robustness. Otherwise, for large \( \Omega \) the dominating factor is \( O(\Omega^2) \). For comparison, the widely used ESPRIT method [2], known to be nearly-optimal in the SR regime [24], has \( O(\Omega^3) \) time complexity: it includes three SVD decompositions and several matrix multiplications of order \( O(\Omega^3) \times O(\Omega) \).

We perform reconstruction tests of a signal with random complex amplitudes and measurement noise in a two-cluster configuration with \( \ell = 2 \), where \( \epsilon, \Omega, \Delta \) are chosen uniformly at random. The results appear in Fig. 3. We also investigate the noise threshold \( \epsilon \gtrsim SRF^{1-2\delta} \) for successful recovery (see Remark 1), and compare to Theorem 1 by recording the success/failure result of each experiment. These results provide numerical validation of the optimality of DPM both in terms of the SNR threshold and the attained estimation accuracy.

Finally, we compare the performance of DPM with ESPRIT and MP. For \( 2 \times 2 \) configuration with \( \Delta = 10^{-2.8}, \epsilon = 10^{-5} \),
Algorithm III.1: Decimated Pruning Method.

Data : $N_{\lambda}, n, \Omega, N_b$

Input : $g(\omega)$ as in (1)

Output: Estimates $\{\tilde{x}_{k}, \tilde{\alpha}_{k}\}_{k=1}^{n}$

for $\lambda \in G := \text{linspace}(\mathcal{J}, N_b)$ do

$\hat{m}(\lambda) := \left\{\tilde{m}_{k}(\lambda) = g(\lambda k)\right\}_{k=0}^{2n-1}$

$\left\{e^{2\pi i j k_{x}, \tilde{\alpha}_{k,k}}\right\} \leftarrow \text{Proxy} \left(\hat{m}_{k}(\lambda)\right)$

Compute $X_{\lambda}$ as in (2)

$X \leftarrow \bigcup_{\lambda \in G} X_{\lambda}$

Compute $\tilde{\mathcal{H}}$ - Histogram of $\{x : (\lambda, x) \in X\}$ with $N_b$ bins.

Set $\{B_{k}\}_{k=1}^{n} = \lambda \rightarrow \text{gMax}(\mathcal{H}, n)$

Compute $\Lambda$ as in (3) and $\Lambda^* = \max(\Lambda : \lambda \in \Lambda)$

$\{\tilde{x}_{k}\}_{j=1}^{n} \leftarrow \{x : (\lambda^*, x) \in X_{\lambda^*} \land x \in B_{j}\}$

Construct $\tilde{\mathcal{V}} = \left(e^{2\pi i j k_{x}}k_{x}^{*}\right)_{j=0, \ldots, n-1}$ and solve

$\text{col}\left\{\tilde{\alpha}_{k,k}^{n} \leftarrow = \arg\min_{\tilde{\alpha}_{k,k}} \left\|\tilde{\mathcal{V}} x - \text{col} \left(\tilde{m}_{k}(\lambda^{*})\right)\right\|^2_{2}\right\}$

return the estimates $\{\tilde{x}_{k}, \tilde{\alpha}_{k,k}\}_{k=1}^{n}$

Fig. 3. DPM - asymptotic optimality. (a) For cluster nodes, $\{K_{x,j}\}$ (left) scale like $\text{SRF}^{2k-2}$, while the $\{K_{z,j}\}$ (right) scale like $\text{SRF}^{2k-1}$ (SRF = $(\Omega \Delta)^{-1}$). For the non-cluster node $j = 5$, both $\{K_{x,5}\}$ and $\{K_{z,5}\}$ are lower bounded by a constant. Here $N_{\lambda} = 50$ and number of tests = 300. (b) Noise threshold for recovery of cluster nodes scales with $\text{SRF}^{2k-1}$. Here $N_{\lambda} = 150$ and number of tests = 500. (c) Comparison of accuracy and runtime with ESPRIT/MP (see text).

IV. DISCUSSION

Future research avenues include a rigorous proof of the algorithm’s correctness in the general case and improving its robustness, building upon the theory developed in [15], [23]. We believe our method can be extended to higher dimensions, along the lines of recent works such as [34].

ACKNOWLEDGMENT

Dmitry Batenkov would like to thank Prof. Ayush Bhandari for useful discussions.

APPENDIX

Here we prove Theorem 3. For the case of a single cluster, each $\lambda \in \mathcal{J}$ is collision-avoiding. Hence, we only need to show that the bins $\{B_{i}\}_{i=1}^{n}$ contain $N_{\lambda}$ valid node approximations.

Proposition 4: Let $h, \eta > 0$ be arbitrary. There exist constants $K_1(n), K_2(n)$ such that for each $h < |c| \leq \eta/6$ and each $\Omega \in \left[\frac{N_{\lambda}}{\eta}, \frac{N_{\lambda}}{\eta 2}\right]$ there exists an interval $I \subset \mathcal{J}$ of length $|I| = \eta^{-1}$ satisfying $|c - \lambda_{m}^{-1} k| > h$ for all $\lambda \in I, k \in \mathbb{Z}$.

Proof: This is just a simplified version of [15, Prop. F.3]. In the proof, the interval $I_{1}$ in case 1 may be replaced by any $I \subset \mathcal{J}$ of length $\eta^{-1}$ and appropriate adjustment of the constants $K_1, K_2$; in case 2 the interval $I$ can be taken as $I_{2}$ itself.

Proof of Theorem 3: Without loss of generality we assume that $x_{j} \in [-h, h]$ for each $j = 1, \ldots, n$ where $h = \tau \Delta$. Now suppose $\lambda_{m} \in G$ and introduce the auxiliary parameter $a = \min(1/6, \tau)$. Letting $\varepsilon \leq \Delta^{2n-1}$ be small enough and employing Theorem 2, the set $X_{\lambda_{0}}$ in (2) has the form $X_{\lambda_{0}} = \bigcup_{m \in R(\lambda_{0})} X_{\lambda_{0} + m}$, where $R(\lambda_{0}) = \left\{\left\lfloor\frac{-2m}{\Delta}\right\rfloor, \ldots, \left\lfloor\frac{-m}{\Delta}\right\rfloor\right\}$ and

$X_{\lambda_{0} + m} = \left\{x_{j, \lambda_{0} + \frac{m}{\Delta}} \middle| j = 1, \ldots, n\right\}.$

Recall step 6 in Algorithm III.1. We make the following Genericity Assumption: The distance from $x_{j}$ to its closest bin edge is at least $a \Delta$.

For each $\lambda' \in G$, $|x_{j, \lambda_{0} + m} - x_{j, \lambda'}| \leq 2an\Delta < \varepsilon/3$. Therefore, both $x_{j, \lambda_{0}}, x_{j, \lambda'}$ must belong to the same bin. In particular, we are guaranteed the existence of $n$ bins containing at least $N_{\lambda}$ elements each. Fix $m \in R(\lambda_{0}) \setminus \{0\}$. The choice of $a$ guarantees $X_{\lambda_{0} + m} \subset c + [-2h, 2h]$ where $c := \frac{m}{\Delta}$. Put $h = 6h$. Since $|c| > \lambda_{0}^{-1} \geq \frac{2n}{12\Delta}$, the condition $|c| > h$ holds whenever $\text{SRF} > 2n$. Applying Proposition 4 with $\eta = 3$ and $h, c$ as above, there exists $I_{m} \subset \mathcal{J}$ of length $|I_{m}| = 1/3$ s.t.

$$\forall \lambda \in I_{m}, \forall k \in \mathbb{Z}: \left|\lambda_{0}^{-1} m - \lambda_{m}^{-1} k\right| > 6h. \quad (4)$$

By choosing $N_{\lambda} \geq C_{1} \Omega$ for sufficiently large $C_{1}$, we can ensure that there exists $\lambda_{m} \in I_{m} \setminus G$. Therefore, for each $k \in \mathbb{Z}$, (4) implies that $X_{\lambda_{m} + k} \subset [-1/2, 1/2] \setminus \left\{\frac{m}{\lambda_{0}} + [-2h, 2h]\right\}$.

Since $X_{\lambda_{m} + k} \subset \frac{m}{\lambda_{0}} + [-2h, 2h]$, we conclude that $x_{j, \lambda_{m} + k}$ cannot belong to the same bin. In particular, the bin containing $x_{j, \lambda_{m} + k}$ contains at most $N_{\lambda} - 1$ elements.

Since $\lambda_{0}$ and $m$ were arbitrary, we proved that for $j = 1, \ldots, n$ the bins containing $\{x_{j, \lambda} : \lambda \in G\}$ have counts at least $N_{\lambda}$, while all other bins have strictly smaller counts.

Remark 2: The genericity assumption is a technical and not an essential restriction. Algorithm III.1 can be easily modified to account for the case that all valid approximations to a node $x_{j}$ belong to two neighboring bins.
REFERENCES

[1] R. J. P. D. Figueiredo and C.-L. Hu, “Waveform feature extraction based on tauberian approximation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 105–116, Mar. 1982.

[2] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice Hall, 2005.

[3] R. Badeau, B. David, and G. Richard, “High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials,” *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1341–1350, Apr. 2006.

[4] I. Krestein, “High resolution time delay estimation,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*. 1987, pp. 451–454.

[5] L. Li and T. P. Speed, “Parametric deconvolution of positive spike trains,” *Ann. Statist.*, vol. 28, no. 5, pp. 1279–1301, 2000.

[6] D. L. Donoho, “Superresolution via sparsity constraints,” *SIAM J. Math. Anal.*, vol. 23, no. 5, pp. 1309–1331, 1992.

[7] E. J. Candès and C. Fernandez-Granda, “Towards a mathematical theory of super-resolution,” *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, Jun. 2014.

[8] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[9] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008.

[10] C. S. Seelamantula and S. Mulley, “Super-resolution reconstruction in frequency-domain optical-coherence tomography using the finite-rate-of-innovation principle,” *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5020–5029, Oct. 2014.

[11] S. Mulley et al., “Super-resolved nuclear magnetic resonance spectroscopy,” *Sci. Rep.*, vol. 7, no. 1, Aug. 2017, Art. no. 9651.

[12] A. Bhandari and R. Raskar, “Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-D imaging,” *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 45–58, Sep. 2016.

[13] A. Bhandari, F. Krahmer, and T. Poskitt, “Unlimited sampling from theory to practice: Fourier-Prony recovery and prototype ADC,” *IEEE Trans. Signal Process.*, vol. 70, pp. 1131–1141, 2022.

[14] D. Batenkov, L. Demanet, G. Goldman, and Y. Yomdin, “Conditioning of partial nonuniform Fourier matrices with clustered nodes,” *SIAM J. Matrix Anal. Appl.*, vol. 44, no. 1, pp. 199–220, 2020.

[15] D. Batenkov, G. Goldman, and Y. Yomdin, “Super-resolution of near-colliding point sources,” *Inf. Inference: A J. IMA*, vol. 10, no. 2, pp. 515–572, Jun. 2021.

[16] D. Batenkov and N. Diab, “Super-resolution of generalized spikes and spectra of confluent vandermonde matrices,” *Appl. Comput. Harmon. Anal.*, vol. 65, pp. 181–208, Jul. 2023.

[17] D. Batenkov, “Stability and super-resolution of generalized spike recovery,” *Appl. Comput. Harmon. Anal.*, vol. 45, no. 2, pp. 299–323, Sep. 2018.

[18] D. Batenkov, “Decimated generalized Prony systems,” 2013, arXiv:1308.0753.

[19] D. Batenkov and Y. Yomdin, “Algebraic signal sampling, Gibbs phenomenon and Prony-type systems,” in *Proc. 10th Int. Conf. Sampling Theory Appl.*, 2013, pp. 137–140.

[20] I. Maravic and M. Vetterli, “Sampling and reconstruction of signals with finite rate of innovation in the presence of noise,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[21] M. Briani, A. Cuyt, F. Knaepkens, and W.-S. Lee, “VEXPA: Validated EXPonential analysis through regular sub-sampling,” *Signal Process.*, vol. 177, Dec. 2020, Art. no. 107722.

[22] R. Prony, “Essai experimental et analytique,” *J. de l’Ecole Polytechnique*, vol. 2, 1795, Art. no. 929.

[23] R. Katz, N. Diab, and D. Batenkov, “On the accuracy of Prony’s method for recovery of exponential sums with closely spaced exponents,” 2023, arXiv:2302.05883.

[24] W. Li, W. Liao, and A. Fannjiang, “Super-resolution limit of the EP-PRIT algorithm,” *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4593–4608, Jul. 2020.

[25] D. Batenkov, B. Diederichs, G. Goldman, and Y. Yomdin, “The spectral properties of vandermonde matrices with clustered nodes,” *Linear Algebra Appl.*, vol. 609, pp. 37–72, Jan. 2021.

[26] S. Kunis and D. Nagel, “On the smallest singular value of multivariate vandermonde matrices with clustered nodes,” *Linear Algebra Appl.*, vol. 604, pp. 1–20, Nov. 2020.

[27] H. B. Lee, “The Cramér-Rao bound on frequency estimates of signals closely spaced in frequency,” *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1507–1517, Jun. 1992.

[28] W. Li and W. Liao, “Stable super-resolution limit and smallest singular value of restricted Fourier matrices,” *Appl. Comput. Harmon. Anal.*, vol. 51, pp. 118–156, Mar. 2021.

[29] P. Stoica, V. Šimonyte, and T. Söderström, “On the resolution performance of spectral analysis,” *Signal Process.*, vol. 44, no. 2, pp. 153–161, Jun. 1995.

[30] P. Liu and H. Zhang, “A theory of computational resolution limit for line spectral estimation,” *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4812–4827, Jul. 2021.

[31] D. Batenkov, A. Bhandari, and T. Blu, “Rethinking super-resolution: The bandwidth selection problem,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 5087–5091.

[32] M. Shahram and P. Milanfar, “On the resolvability of sinusoids with nearby frequencies in the presence of noise,” *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2579–2588, Jul. 2005.

[33] Å. Björck and V. Pereyra, “Solution of Vandermonde systems of equations,” *Math. Computation*, vol. 24, no. 112, pp. 893–903, 1970.

[34] B. Diederichs, N. M. Kolountzakis, and E. Papageorgiou, “How many Fourier coefficients are needed?,” *Monatshefte für Mathematik*, vol. 200, no. 1, pp. 23–42, 2022.