

# DECIMATED PRONY'S METHOD FOR STABLE SUPER-RESOLUTION

Rami Katz

Nuha Diab, Dmitry Batenkov

School of Electrical Engineering  
Tel-Aviv University  
Tel-Aviv, Israel

Department of Applied Mathematics  
Tel-Aviv University  
Tel-Aviv, Israel

## ABSTRACT

We study recovery of amplitudes and nodes of a finite impulse train from a limited number of equispaced noisy frequency samples. This problem is known as super-resolution (SR) under sparsity constraints and has numerous applications. Prony's method is an algebraic technique which fully recovers the signal parameters in the absence of measurement noise. In the presence of noise, Prony's method may experience significant loss of accuracy, especially when the separation between Dirac pulses is smaller than the Nyquist-Shannon-Rayleigh (NSR) limit. In this work we combine Prony's method with a recently established decimation technique for analyzing the SR problem in the regime where the distance between two or more pulses is much smaller than the NSR limit. We show that our approach attains optimal asymptotic stability in the presence of noise. Our result challenges the conventional belief that Prony-type methods tend to be highly numerically unstable.

**Index Terms**— Prony's method, decimation, sparse super-resolution

## 1. INTRODUCTION

The (noisy) SR problem is to recover 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

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

where  $\Omega > 0$  and  $\|e\|_\infty \leq \epsilon$  for  $\epsilon > 0$ . The SR problem belongs to a class of inverse problems with multiple applications, including parametric spectral estimation, direction of arrival estimation, finite rate of innovation

sampling, time-of-flight imaging and unlimited sensing [1, 2, 3, 4, 5]. High-order versions of (1) are considered in e.g. [6, 7].

In the noiseless regime (i.e., when  $e(\omega) \equiv 0$ ), Prony's method [8] recovers the parameters  $\{\alpha_k, x_k\}_{k=1}^n$  from  $2n$  equispaced samples  $g(k)$ ,  $k = 0, \dots, 2n - 1$ . This is achieved by reducing the SR problem to a three-step procedure, which involves solution of two linear systems, in combination with a root-finding step (See Algorithm 2.1 below). In the presence of measurement noise, however, the conventional wisdom is that Prony's method is ill-conditioned, meaning that it suffers from a significant loss of accuracy. An especially challenging regime for Prony's method 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 [9, 10]. These theoretical bounds establish the fundamental limits of recovery in the SR problem. However, a tractable algorithm which achieves these theoretical bounds has been missing in the literature.

In this work we develop a reconstruction algorithm inspired by a decimation approach first proposed in [2]. Our procedure relies on sampling the spectrum of  $g$  at  $2n$  equispaced and maximally separated frequencies, and solving the SR problem thus obtained by applying the Prony's method. For success of the approach, care needs to be taken to avoid node aliasing and collisions. Inspired by [10], this is achieved by considering sufficiently many decimated sub-problems. The result is a tractable algorithm, which has lower computational complexity than the well-established and frequently used ESPRIT algorithm (see e.g. [11]). We further provide theoretical motivation and numerical examples which show that the proposed algorithm achieves the optimal asymptotic stability and noise tolerance guaranteed in the literature.

ND and DB are supported by the Israel Science Foundation Grant 1793/20 and a collaborative grant from the Volkswagen Foundation.

## 2. TOWARDS OPTIMAL ALGORITHMS

Throughout the paper we consider the number of nodes (resp. amplitudes)  $n \in \{1, 2, \dots\}$  in (1) to be fixed. We assume that the nodes satisfy  $\{x_k\}_{k=1}^n \subseteq [-\frac{1}{2}, \frac{1}{2}]$ . By rescaling the data (1), this assumption poses no loss of generality (see Section 4 in [10]). Let  $\{\tilde{x}_k\}_{k=1}^n$  and  $\{\tilde{\alpha}_k\}_{k=1}^n$  be the approximated parameters obtained via a reconstruction algorithm using the data (1).

**Definition 1** ([10]). *Let  $\{(\alpha_k, x_k)\}_{k=1}^n \subseteq U$ . Given  $\epsilon > 0$ , the min-max recovery rates are*

$$\Lambda_{\epsilon, U, \Omega}^{x, j} = \inf_{\mathcal{A}: g \mapsto \{\tilde{\alpha}_j, \tilde{x}_j\}} \sup_{\{\alpha_j, x_j\}} \sup_{\|e\|_\infty \leq \epsilon} |x_j - \tilde{x}_j|,$$

$$\Lambda_{\epsilon, U, \Omega}^{\alpha, j} = \inf_{\mathcal{A}: g \mapsto \{\tilde{\alpha}_j, \tilde{x}_j\}} \sup_{\{\alpha_j, x_j\}} \sup_{\|e\|_\infty \leq \epsilon} |\alpha_j - \tilde{\alpha}_j|.$$

Here  $U$  is some fixed subset in the parameter space, whereas the infimum is over the set of all reconstruction algorithms  $\mathcal{A}$  which employ the data (1).

The min-max rates define the optimal recovery rates achievable by a reconstruction algorithm, in the presence of measurement noise of magnitude less than  $\epsilon$  and when the node and amplitude pairs belong to  $U$ .

It has been well established that the difficulty of SR is related to the minimal separation between nodes,  $\Delta := \min_{s \neq k} |x_s - x_k|$ . A particular case of interest, both theoretically and from an applications perspective, concerns signals whose nodes are densely clustered, i.e.  $\Delta \ll \frac{1}{\Omega}$  [12, 13, 14, 15, 16, 17].

**Definition 2.** *We call  $\{x_k\}_{k=1}^n$  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). [9, 10].*

**Theorem 1** ([10]). *Let the super-resolution factor satisfy  $\text{SRF} := \frac{1}{\Omega\Delta} \gtrsim 1$ , and  $\{|\alpha_k|\}_{k=1}^n$  be (uniformly) bounded. Let there exist 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*

$$\Lambda_{\epsilon, U, \Omega}^{x, j} \asymp \begin{cases} \text{SRF}^{2\ell-2} \frac{\epsilon}{\Omega} & x_j \in \mathcal{X}, \\ \frac{\epsilon}{\Omega} & x_j \notin \mathcal{X}, \end{cases}$$

$$\Lambda_{\epsilon, U, \Omega}^{\alpha, j} \asymp \begin{cases} \text{SRF}^{2\ell-1} \epsilon & x_j \in \mathcal{X}, \\ \epsilon & x_j \notin \mathcal{X}. \end{cases}$$

Here,  $\lesssim, \gtrsim, \asymp$  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. [18]). Related results are known in the signal processing literature for Gaussian noise [14, 16, 19]. Despite a plethora of methods to solve this problem, up to date a tractable algorithm which achieves the min-max rates is missing from the literature. In particular, the frequently used ESPRIT algorithm is sub-optimal, both in terms of the bounds and the threshold SNR [11].

While the proof of Theorem 1 is non-constructive, it motivates a design of an algorithm achieving the optimal rates in practice. Let  $\mathcal{J} := \left[ \frac{1}{2} \frac{\Omega}{2n-1}, \frac{\Omega}{2n-1} \right]$ . For a decimation parameter  $\lambda \in \mathcal{J}$ , the measurements  $\{g(\lambda k)\}_{k=0}^{2n-1}$  yield the problem (1) with  $\{e^{2\pi j x_j}\}_{j=1}^n$  replaced by  $\{e^{2\pi j \lambda x_j}\}_{j=1}^n$ . By [10, Prop. 5.8] there exists an interval  $\mathcal{I} \subset \mathcal{J}$  of length  $|\mathcal{I}| \geq c\Omega$  such that every  $\lambda \in \mathcal{I}$  satisfies  $|e^{2\pi j \lambda x_i} - e^{2\pi j \lambda x_j}| \geq \frac{1}{n^2}$ , whenever  $x_i, x_j$  belong to different clusters (collision avoidance). By [10, Prop. 5.12], for a collision-avoiding  $\lambda$ , the condition number of the (theoretical) solution map  $\mathcal{P}_n$  which inverts (1) in the case  $e \equiv 0$  and  $\omega \in \{k\lambda\}_{k=0}^{2n-1}$  matches the min-max rates. Set  $\{\tilde{\alpha}_j, e^{2\pi j \tilde{y}_{\lambda, j}}\}_{j=1}^n = \mathcal{P}_n \{g(\lambda k)\}_{k=0}^{2n-1}$ . Let the set of all aliased solutions corresponding to  $\{\tilde{y}_{\lambda, j}\}_{j=1}^n$  be

$$X_\lambda := \bigcup_{j=1}^n \left\{ (\lambda, t) : t = \frac{\tilde{y}_{\lambda, j} + m}{\lambda}, m \in \mathbb{Z}, |t| \leq \frac{1}{2} \right\}, \quad (2)$$

where the aliasing follows by periodicity of  $y \mapsto e^{2\pi j y}$ . The arguments above imply that  $X_\lambda$  contains at least one element  $(\lambda, t)$  with  $t \approx x_j$  for each  $j = 1, \dots, n$  (call it Property P\*). Thus, to obtain a constructive procedure for recovery, we propose the following general approach:

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

In this work we tackle steps 1 and 3. For step 2, we propose to use the classical Prony's method [8], which provides an exact solution to the problem in the noiseless regime (see Alg. 2.1). Define the node/amplitude error amplification factors

$$\mathcal{K}_{x, j} := \epsilon^{-1} \Omega |x_j - \tilde{x}_j|, \quad \mathcal{K}_{\alpha, j} := \epsilon^{-1} |\alpha_j - \tilde{\alpha}_j|.$$

The use of Prony's method is motivated by the simulations in Fig. 1. Fixing  $\Omega = 2n - 1$  and testing Alg. 2.1 for  $n \in \{3, 4, 5\}$  and cluster sizes  $\ell \in \{2, 3, 5\}$  in the regime  $\Delta \ll 1$  (i.e.,  $\text{SRF} \gg 1$  in Theorem 1), and in the presence of measurement noise shows that

Prony's method asymptotically achieves the min-max rates required for step 2.

**Remark 1.** In all numerical tests in this paper, we follow [10] and consider  $x_k$  to be successfully recovered if the error  $|x_k - \tilde{x}_k|$  is smaller than one third of the distance between  $x_k$  and its nearest neighbor.

---

**Algorithm 2.1:** The Classical Prony method

---

**Input** : Sequence  $\{\tilde{m}_k \equiv g(k)\}_{k=0}^{2n-1}$   
**Output** : Estimates  $\{\tilde{x}_k, \tilde{\alpha}_k\}_{k=1}^n \in \mathbb{C}^m$ .  
**Notation:**  $\text{col}\{y_k\}_{k=1}^m = [y_1, \dots, y_m]^\top \in \mathbb{C}^m$ .

- 1 Construct  $\tilde{H}_n = (\tilde{m}_{i+j})_{0 \leq i, j \leq n-1}$
- 2 Solve the linear least squares problem

$$\text{col}\{q_k\}_{k=0}^{n-1} = \arg \min_{q \in \mathbb{C}^n} \|\tilde{H}_n q + \text{col}\{\tilde{m}_k\}_{k=n}^{2n-1}\|_2$$

- 3 Compute  $\{\tilde{z}_k\}_{k=1}^n$  as the roots of the (perturbed) Prony polynomial  $q(z) := z^n + \sum_{j=0}^{n-1} q_j z^j$ .
- 4 Recover  $\{\tilde{x}_k\}_{k=1}^n$  from  $\tilde{z}_k$  via  $\tilde{x}_k = \frac{\text{Arg}(\tilde{z}_k)}{2\pi}$ .
- 5 Construct  $\tilde{V} = (\tilde{z}_k^i)_{i=0, \dots, n-1}^{k=1, \dots, n}$  and solve  $\text{col}\{\tilde{\alpha}_k\}_{k=1}^n = \arg \min_{\alpha \in \mathbb{C}^n} \|\tilde{V}\alpha - \text{col}\{\tilde{m}_k\}_{k=0}^{n-1}\|_2$

6 **return** the estimated parameters  $\{\tilde{x}_k, \tilde{\alpha}_k\}_{k=1}^n$

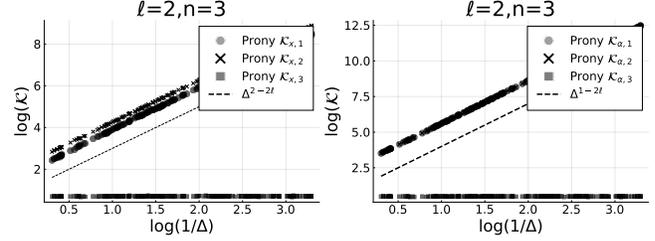
---

### 3. DECIMATED PRONY'S METHOD

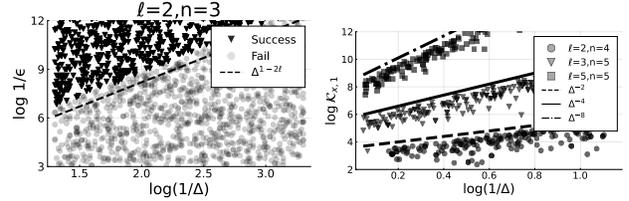
Here we develop the Decimated Prony's Method (DPM) (Alg. 3.1). To find a collision-avoiding  $\lambda$ , we consider  $X_\lambda$  as in (2) for each  $\lambda \in G$  where  $G = \text{linspace}(\mathcal{J}, N_\lambda)$  is the uniform grid of size  $N_\lambda \in \mathbb{N}$  (the choice of  $N_\lambda$  is motivated in Remark 2. Cf. Fig. 2 for a numerical justification of this approach) [steps 1–5]. Next, we compute the histogram of  $\{x : (\lambda, x) \in \bigcup_{\lambda \in G} X_\lambda\}$  with  $N_b$  bins (cf. Remark 2) and find the  $n$  bins  $\{B_j\}_{j=1}^n$  with largest counts [step 6]. Based on [10] we expect the set

$$\Lambda := \bigcap_{k=1}^n \left\{ \lambda : (\lambda, x) \in \bigcup_{\lambda \in G} X_\lambda \wedge x \in B_k \right\} \quad (3)$$

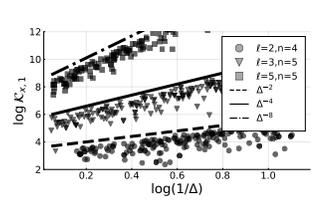
to contain only collision-avoiding  $\lambda$ 's. In particular, if  $\lambda$  is not collision-avoiding, **Property P\*** will not be satisfied since at least two nodes will be ill-conditioned. Furthermore, the proof of [10, Prop. 5.17] suggests that if  $\lambda_1 \neq \lambda_2$  are collision-avoiding, 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  $j =$



(a) Error amplification factors for  $\ell = 2, n = 3$



(b) Recovery SNR threshold



(c)  $\mathcal{K}_{x,1}$  for various  $\ell, n$

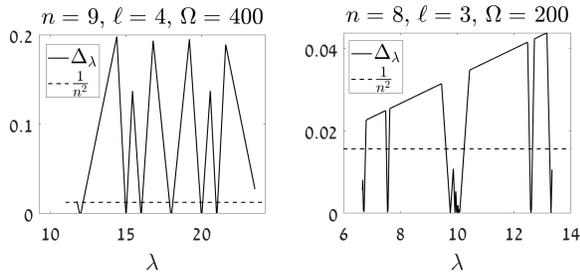
**Fig. 1.** Classical Prony method - asymptotic optimality. (a) For cluster nodes,  $\{\mathcal{K}_{x,j}\}$  (left) scale like  $\Delta^{2-2\ell}$ , while  $\{\mathcal{K}_{\alpha,j}\}$  (right) scale like  $\Delta^{1-2\ell}$ . For the non-cluster node  $j=3$ ,  $\{\mathcal{K}_{x,j}\}$  and  $\{\mathcal{K}_{\alpha,j}\}$  are lower bounded by a constant. (b) Noise threshold for recovery of cluster nodes scales like  $\Delta^{2\ell-1}$ . (c)  $\mathcal{K}_{x,j} \asymp \Delta^{2-2\ell}$  for cluster nodes holds for different  $\ell, n$ . The number of tests is (a) 200, (b) 2000, (c) 200 for each  $\ell, n$ .

$1, \dots, n$ . Thus, provided  $\Lambda \neq \emptyset$  (otherwise the algorithm fails), we choose  $\lambda^* = \max \Lambda$  to obtain maximal in-cluster separation of  $\{e^{2\pi j \lambda^* x_j}\}_{j=1}^n$  [step 7] and recover the corresponding  $\{\tilde{x}_j\}_{j=1}^n$  by choosing  $\tilde{x}_j \in B_j$  s.t.  $(\lambda^*, \tilde{x}_j) \in X_{\lambda^*}$  [step 8]. Finally, the amplitude approximations are found by solving a Vandermonde system [step 9].

**Remark 2.** Following the criterion for successful node recovery (see Remark 1),  $N_b$  is set to be at least  $3\Delta^{-1}$ . We further conjecture that  $N_\lambda \gtrsim \Omega$  is sufficient to ensure correctness of the algorithm. We leave the rigorous justification of this claim to future work. As a guideline, increasing  $N_\lambda$  is expected to improve the robustness of the approximations generated by DPM.

Next, we analyze the time complexity of DPM by addressing each step of Algorithm 3.1. Below we use the notation  $O = O_n$  (recall that  $n$  is considered fixed). The classical Prony's method has complexity  $O(1)$ , since it depends only on  $n$ . For every  $\lambda \in G$  we apply Algorithm 2.1 with the samples  $\{g(\lambda k)\}$  and compute  $X_\lambda$  in (2), which costs  $O(N_\lambda + \lambda N_\lambda)$ . Computing the histogram with  $N_b$  bins for data of size  $\sum_{i=1}^{N_\lambda} n \lambda_i$  costs  $O(N_\lambda \Omega + N_b)$ . Finding the bins  $\{B_k\}_{k=1}^n$  with  $n$  largest counts costs  $O(N_b)$ . Computing  $\Lambda$  in (3)

and  $\lambda^* = \max \Lambda$ , together with finding the estimated  $\{\tilde{x}_k\}_{k=1}^n$  costs  $O(N_\lambda)$ . Finally, solving an  $n$ -order Vandermonde system costs  $O(1)$  [20]. Thus, the *total complexity* is  $O(N_\lambda \Omega + N_b)$ . As mentioned in Remark 2,  $N_\lambda \gtrsim \Omega$  and  $N_b \gtrsim \Delta^{-1}$  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  $N_\lambda = O(\text{SRF})$  to have maximal robustness. Otherwise, for large  $\Omega$  the dominating factor is  $O(\Omega^2)$ . For comparison, the widely used ESPRIT method [1] has a time complexity of order  $O(\Omega^3)$ : it includes three SVD decompositions and several matrix multiplications of order  $O(\Omega) \times O(\Omega)$ .



**Fig. 2.** The separation  $\Delta_\lambda := \min_{s \neq k} \left| \arg \left( e^{2\pi j \lambda (x_k - x_s)} \right) \right|$  as a function of  $\lambda$ , for the single (left) and multi-cluster (right) cases. We see that most  $\lambda \in G$  satisfy  $\Delta_\lambda > \frac{1}{n^2}$ , i.e. are collision-avoiding. Here we set  $N_\lambda = \Omega$ .

---

**Algorithm 3.1:** Decimated Prony Method

---

**Data :**  $N_\lambda, n, \Omega, N_b > \frac{3}{\Delta}$   
**Input :**  $g(\omega)$  as in (1)  
**Output:** Estimates  $\{\tilde{x}_k, \tilde{\alpha}_k\}_{k=1}^n$

- 1 **for**  $\lambda \in G := \text{linspace}(\mathcal{J}, N_\lambda)$  **do**
- 2      $\tilde{m}^{(\lambda)} := \left\{ \tilde{m}_k^{(\lambda)} = g(\lambda k) \right\}_{k=0}^{2n-1}$
- 3      $\{e^{2\pi j \lambda y_{\lambda,k}}, \tilde{\alpha}_{\lambda,k}\} \leftarrow \text{Prony}(\tilde{m}^{(\lambda)})$
- 4     **Compute**  $X_\lambda$  as in (2)
- 5  $X \leftarrow \bigcup_{\lambda \in G} X_\lambda$
- 6 **Compute**  $H$  - Histogram of  $\{x : (\lambda, x) \in X\}$   
with  $N_b$  bins. Set  $\{B_k\}_{k=1}^n = \text{ArgMax}(H, n)$
- 7 **Compute**  $\Lambda$  as in (3) and  $\lambda^* \leftarrow \max\{\lambda : \lambda \in \Lambda\}$
- 8  $\{\tilde{x}_j\}_{j=1}^n \leftarrow \{x : (\lambda^*, x) \in X_{\lambda^*} \wedge x \in B_j\}$
- 9 **Construct**  $\tilde{V} = (e^{2\pi j \tilde{x}_j k \lambda^*})_{k=0, \dots, n-1}^{j=1, \dots, n}$  **and solve**  

$$\text{col} \{ \tilde{\alpha}_k \}_{k=1}^n = \arg \min_{\alpha \in \mathbb{C}^n} \left\| \tilde{V} \alpha - \text{col} \left\{ \tilde{m}_k^{(\lambda^*)} \right\}_{k=0}^{n-1} \right\|_2$$
- return the estimates**  $\{\tilde{x}_k, \tilde{\alpha}_k\}_{k=1}^n$

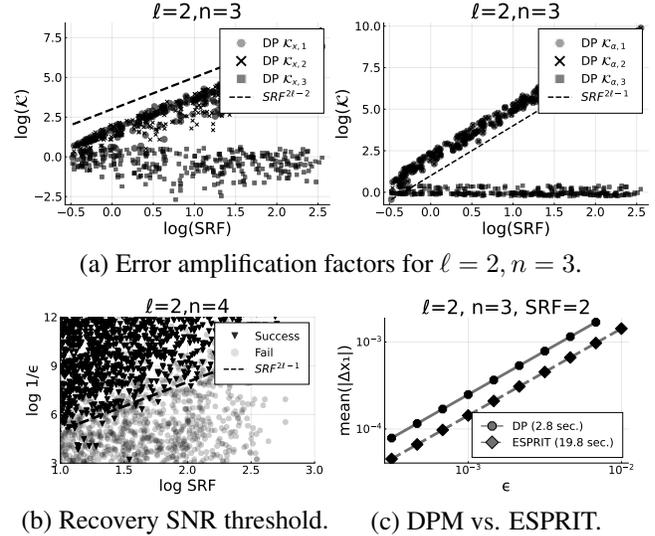
---

We perform reconstruction tests of a signal with

random complex amplitudes and measurement noise in a single-cluster configuration, where  $\epsilon, \Omega, \Delta$  are chosen uniformly at random. The results appear in Fig. 3. We also investigate the noise threshold  $\epsilon \gtrsim \text{SRF}^{1-2\ell}$  for successful recovery (see Remark 1), and compare to Thm. 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 the ESPRIT method. Fixing  $\ell = 2, n = 3, \Delta = 10^{-2.8}, \Omega = 10^{2.5}$  and running 50 tests for each of 10 values of  $\epsilon$  between  $10^{-3.5}$  and  $10^{-2}$ , the mean absolute error in recovering the cluster node is comparable between the two methods. However, DPM with  $N_\lambda = 50$  runs about 7 times faster (Fig. 3(c)).

Future research avenues include providing a rigorous proof of the algorithm's correctness and improving its robustness, and proving that Prony's method indeed achieves the optimal recovery rates. We believe our method can be extended to higher dimensions, along the lines of recent works such as [21] and references therein.



**Fig. 3.** DPM - asymptotic optimality. (a) For cluster nodes,  $\{\mathcal{K}_{x,j}\}$  (left) scale like  $\text{SRF}^{2\ell-2}$ , while the  $\{\mathcal{K}_{\alpha,j}\}$  (right) scale like  $\text{SRF}^{2\ell-1}$  ( $\text{SRF} = (\Omega\Delta)^{-1}$ ). For the non-cluster node  $j=3$ , both  $\{\mathcal{K}_{x,j}\}$  and  $\{\mathcal{K}_{\alpha,j}\}$  are lower bounded by a constant. Here  $N_\lambda = 10$  and number of tests=300. (b) Noise threshold for recovery of cluster nodes scales with  $\text{SRF}^{2\ell-1}$ . Here  $N_\lambda = 2500$  and number of tests=5000. (c) Comparison of accuracy and runtime with the ESPRIT method. While the mean absolute errors (50 tests for each  $\epsilon$ ) scale the same and are comparable, the DPM runs 7 times faster ( $N_\lambda = 50$ ).

#### 4. REFERENCES

- [1] P. Stoica and R.L. Moses, *Spectral Analysis of Signals*, Pearson/Prentice Hall, 2005.
- [2] Dmitry Batenkov and Yosef Yomdin, “Algebraic signal sampling, Gibbs phenomenon and Prony-type systems,” in *Proceedings of the 10th International Conference on Sampling Theory and Applications (SAMPTA)*, 2013.
- [3] 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 Processing Magazine*, vol. 33, no. 5, Sept. 2016.
- [4] Ayush Bhandari, Felix Kraemer, and Thomas Poskitt, “Unlimited Sampling From Theory to Practice: Fourier-Prony Recovery and Prototype ADC,” *IEEE Transactions on Signal Processing*, vol. 70, 2022.
- [5] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse Sampling of Signal Innovations,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, Mar. 2008.
- [6] Dmitry Batenkov and Nuha Diab, “Super-resolution of generalized spikes and spectra of confluent Vandermonde matrices,” *arXiv:2203.11923 [cs, math]*, Mar. 2022.
- [7] Roland Badeau, Bertrand David, and Gaël Richard, “High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials,” *Signal Processing, IEEE Transactions on*, vol. 54, no. 4, 2006.
- [8] R. Prony, “Essai experimental et analytique,” *J. Ec. Polytech.(Paris)*, vol. 2, 1795.
- [9] Dmitry Batenkov, Laurent Demanet, Gil Goldman, and Yosef Yomdin, “Conditioning of Partial Nonuniform Fourier Matrices with Clustered Nodes,” *SIAM Journal on Matrix Analysis and Applications*, vol. 44, no. 1, Jan. 2020.
- [10] Dmitry Batenkov, Gil Goldman, and Yosef Yomdin, “Super-resolution of near-colliding point sources,” *Information and Inference: A Journal of the IMA*, vol. 10, no. 2, June 2021.
- [11] Weilin Li, Wenjing Liao, and Albert Fannjiang, “Super-resolution limit of the ESPRIT algorithm,” *IEEE Transactions on Information Theory*, 2020.
- [12] Dmitry Batenkov, Benedikt Diederichs, Gil Goldman, and Yosef Yomdin, “The spectral properties of Vandermonde matrices with clustered nodes,” *Linear Algebra and its Applications*, vol. 609, Jan. 2021.
- [13] Stefan Kunis and Dominik Nagel, “On the smallest singular value of multivariate Vandermonde matrices with clustered nodes,” *Linear Algebra and its Applications*, vol. 604, Nov. 2020.
- [14] Harry B. Lee, “The Cramér-Rao bound on frequency estimates of signals closely spaced in frequency,” *IEEE Transactions on Signal Processing*, vol. 40, no. 6, 1992.
- [15] Weilin Li and Wenjing Liao, “Stable super-resolution limit and smallest singular value of restricted Fourier matrices,” *Applied and Computational Harmonic Analysis*, vol. 51, Mar. 2021.
- [16] Petre Stoica, Virginija Šimonyte, and Torsten Söderström, “On the resolution performance of spectral analysis,” *Signal Processing*, vol. 44, no. 2, June 1995.
- [17] Ping Liu and Hai Zhang, “A Theory of Computational Resolution Limit for Line Spectral Estimation,” *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
- [18] D. Batenkov, A. Bhandari, and T. Blu, “Rethinking Super-resolution: The Bandwidth Selection Problem,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019.
- [19] M. Shahram and P. Milanfar, “On the resolvability of sinusoids with nearby frequencies in the presence of noise,” *IEEE Transactions on Signal Processing*, vol. 53, no. 7, July 2005.
- [20] Åke Björck and Victor Pereyra, “Solution of vandermonde systems of equations,” *Mathematics of computation*, vol. 24, no. 112, pp. 893–903, 1970.
- [21] Benedikt Diederichs, Mihail N. Kolountzakis, and Effie Papageorgiou, “How many Fourier coefficients are needed?,” *Monatshefte für Mathematik*, Oct. 2022.