RETHINKING SUPER-RESOLUTION: THE BANDWIDTH SELECTION PROBLEM

Dmitry Batenkov, Ayush Bhandari† and Thierry Blu⋆*

*Dept. of Mathematics, Massachusetts Institute of Technology, Cambridge MA, USA.
†Dept. of Electrical and Electronic Engineering, Imperial College London, UK.
⋆Dept. of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong.
Email: batenkov@MIT.edu • a.bhandari@imperial.ac.uk • thierry.blu@polytechnique.org

ABSTRACT

Super-resolution is the art of recovering spikes from their low-pass projections. Over the last decade specifically, several significant advancements linked with mathematical guarantees and recovery algorithms have been made. Most super-resolution algorithms rely on a two-step procedure: deconvolution followed by high-resolution frequency estimation. However, for this to work, exact bandwidth of low-pass filter must be known; an assumption that is central to the mathematical model of super-resolution. On the flip side, when it comes to practice, smoothness rather than bandlimitedness is a much more applicable property. Since smooth pulses decay quickly, one may still capitalize on the existing super-resolution algorithms provided that the essential bandwidth is known. This problem has not been discussed in literature and is the theme of our work. In this paper, we start with an experiment to show that super-resolution in the presence of noise is sensitive to bandwidth selection. This raises the question of how to select the optimal bandwidth. To this end, we propose a bandwidth selection criterion which works by minimizing a proxy of estimation error that is dependent of bandwidth. Our criterion is easy to compute, and gives reasonable results for experimentally acquired data, thus opening interesting avenues for further investigation, for instance the relationship to Cramér-Rao bounds.

***Index Terms*—** Sampling theory, sparsity, super-resolution, spectral estimation, sparse deconvolution, bandwidth.

1. INTRODUCTION

Recovering spikes from low-pass projections is a classical problem that arises in a variety of applications. In the field of signal processing, this problem is widely studied under the theme of (a) sparse-deconvolution [1], (b) sparse or finite rate of innovation sampling [2, 3] and, (c) super-resolution [4]. Given N time-domain, sampled measurements, $y(nT), n = 0, \ldots, N-1$ of the continuous signal

$$y(t) = \sum\nolimits_{k=0}^{K-1} c_k \phi(t - t_k), \quad (1)$$

the super-resolution problem seeks to recover the $2K$ unknowns $\{c_k, t_k\}_{k=0}^{K-1}$ assuming that [1–4]: (A1) K and ϕ are known; and (A2) ϕ is bandlimited (its Fourier transform is compactly supported). The notion of sparsity naturally finds its way in the super-resolution problem because $y(t) = (\phi * s)(t)$ where s is a continuous-time, K-sparse signal

$$s(t) = \sum\nolimits_{k=0}^{K-1} c_k \delta(t - t_k), \quad t_k \in [0, \tau). \quad (2)$$

Clearly when the consecutive t_k's are close to each other, the overlap between $c_k\phi(t - t_k)$ and $c_{k+1}\phi(t - t_{k+1})$ makes the identification of $\{c_k, t_k\}_{k=0}^{K-1}$ challenging and hence one must "super-resolve."

Over the course of time, several efforts have been made to understand the theoretical aspects of this problem which have also improved the algorithmic machinery linked with the super-resolution problem (cf. [5]). That said, from a practical perspective, physical models underlying the super-resolution problem make the application of theory, as is, challenging. By working with experimental data in several bands of the electro-magnetic spectrum, we have observed that the current model specified in (1) may be too simple to explain the experimental data. To elaborate on this aspect, we provide two concrete examples where deviations from the *de-facto* mathematical model (1) arise. In practice the pulse ϕ may be distorted due to propagation or transmission and hence the data is better explained by overlapping templates $\{c_k\phi_k(t - t_k)\}$ where each ϕ_k is related to some basic pulse ϕ. This aspect was studied in the time-varying model discussed in [6]. Other models addressing the same issue were considered in [7, 8]. Even when (1) is a valid model, the pulse ϕ may be non-parameteric or arbitrary. This aspect was covered in [9] where the requirement was that ϕ reproduces trigonometric moments.

1.1. Super-resolution is Sensitive to Bandwidth

Central to the problem of super-resolution is the assumption that ϕ is bandlimited [1–4]. The notion of exact bandlimitedness is far from practice and a better suited model would be the one that considers smooth pulses. From classical Fourier analysis, we know that smoothness in time-domain imposes a decay constraint on the Fourier coefficients and hence, *bandlimited approximation* (cf. Fig. 3 in [10]) is a viable solution. While smooth pulses are approximately bandlimited, choosing the correct bandwidth in presence of noise and model mismatch is a non-trivial aspect. This is the main theme of this paper.

To clarify this point, consider the Fourier domain version of (1). Let $\widehat{y}(\omega)$ denote the Fourier transform of $y(t)$. Then, from the convolution–product theorem, we have, $\widehat{y}(\omega) = \widehat{\phi}(\omega)\,\widehat{s}(\omega)$ where $\widehat{s}(\omega)$ is a sum of complex exponentials, or,

$$\widehat{s}(\omega) = \sum\nolimits_{k=0}^{K-1} c_k e^{-\jmath\omega t_k}, \quad (3)$$

and equivalently, with sampled measurements, one has $\{\widehat{y}(n\omega_0)\}_{n=0}^{N-1}$ with $\omega_0 = 2\pi/\tau$. Typical recovery procedure in the super-resolution problem exploits the structure in (3). This is done in two steps:

1. **Deconvolution**. Here $\widehat{s}(n\omega_0)$ is estimated by using,

$$\widehat{s}(n\omega_0) = \frac{\widehat{y}(n\omega_0)}{\widehat{\phi}(n\omega_0)}, \quad n\omega_0 \in [-\Omega, \Omega] \quad (4)$$

where Ω is the bandwidth of ϕ.

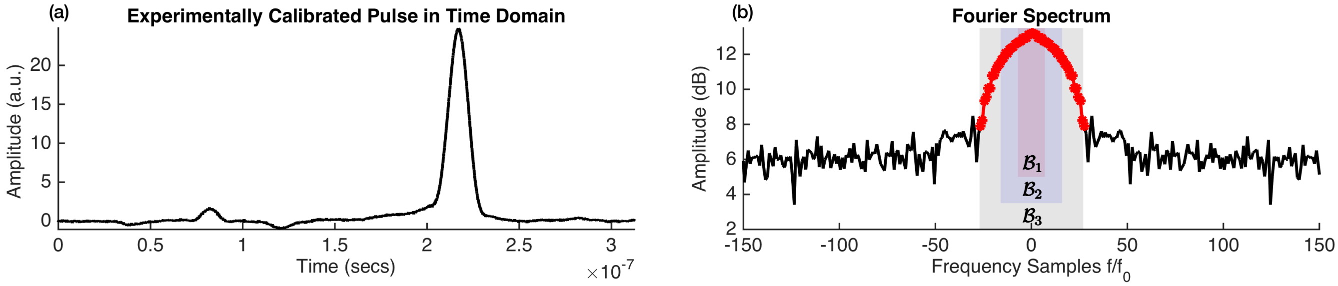

Fig. 1: (a) Time-domain samples of pulse $\{\phi(nT)\}_{n=1}^{4464}$ calibrated using time-resolved sensor in [10]. Note that the pulse is smooth and asymmetric. (b) Fourier domain samples. Here $\mathcal{B}_k = [-\Omega_k, \Omega_k]$, $k = 1, 2, 3$ are possible bandwidths (for deconvolution) that can be used for super-resolving spikes.

2. **Parameter Estimation**. Once $\widehat{s}(n\omega_0)$ is computed, its parametric form in (3) is then used for estimating unknowns $\{c_k, t_k\}_{k=0}^{K-1}$. To do so, one first estimates $\{t_k\}_{k=0}^{K-1}$ using high resolution spectral estimation methods [12] or fitting approaches [5]. With $\{t_k\}_{k=0}^{K-1}$ known, $\{c_k\}_{k=0}^{K-1}$ are estimated by solving least-squares problem in (3). The estimates $\{c_k, t_k\}_{k=0}^{K-1}$ correspond to the sparse signal in (2).

When ϕ is a smooth pulse, as is often the case in practice (cf. experimental examples in [6, 10, 13]) the selection criterion for bandwidth parameter Ω is *unclear*. Consider the case when measurements incorporate noise and read,

$$m(t) = y(t) + e(t) \tag{5}$$

where $e(t)$ is the contribution due to bounded noise.

In context of the experimentally calibrated pulse shown in Fig. 1(a), a reasonable method to choose Ω (equivalently, the spectral window $\mathcal{B} = [-\Omega, \Omega]$), entails the largest, contiguous set of frequencies for which $\widehat{\phi}(n\omega_0)$ is away from zero (see red samples in Fig. 1(b)). For the moment, let us denote this *heuristic* value of bandwidth by Ω_0. Varying Ω arbitrarily, leads to the following scenarios.

- When Ω is such that $N < 2K$, the parameter estimation by fitting (3) will fail as the system is under-determined.
- Gradually increasing Ω such that $2K\omega_0 \le \Omega < \Omega_0$ leads to *over-sampling* and hence to performance enhancement of the spectral estimation methods. The benefits of over-sampling are well known in literature. For instance, both Cadzow's method [14] and the matrix pencil method introduced by Hua and Sarkar [11] improves the performance of parameter estimation. This is also consistent with the Cramér-Rao bounds developed in [3].
- Understandably, when Ω approaches the heuristically chosen Ω_0, the deconvolution step in (4) becomes ill-posed. This is because smoothness of the pulse ϕ causes its spectrum to decay and the values of $\widehat{\phi}(n\omega_0)$ start approaching the noise floor. This causes the denominator in (4) to blow up.

The above observation is verified through a thought experiment in Fig. 2(a). Given noisy measurements $m(t)$ in (5) of a signal with $K = 2$ spikes separated by $\Delta = |t_2 - t_1| = 12.6$ ns, and convolved with the kernel in Fig. 1(a), further corrupted by 15 dB additive white Gaussian noise, we deconvolve and estimate $\{\widetilde{c}_k, \widetilde{t}_k\}_{k=0}^{K-1}$ using the matrix pencil method [11], by sweeping the bandwidth parameter Ω. Based on the estimates, we plot the maximal estimation error $\max_{k\in[0,K-1]} |\widetilde{t}_k - t_k|$. Corresponding to the above three possibilities, we choose $\Omega_1 = 2K\omega_0$, Ω_2 which minimizes the estimation error above, and Ω_3 is arbitrarily chosen to demonstrate the effect of reconstruction when deconvolution is unstable. While the case of Ω_1 suffers from not having enough samples to combat the effect of noise, the case of Ω_3 fails because unstable deconvolution leads to estimation of erroneous frequencies in (3), hence adversely affecting the reconstruction.

Clearly, Ω_2 leads to the optimal performance and benefits from over-sampling. Nonetheless, this is a result of *brute-force search* over all Ω with $\{t_k\}_{k=0}^{K-1}$ known. In practice, one does not have access to the ground truth $\{t_k\}_{k=0}^{K-1}$ and this necessitates a **bandwidth selection criterion** which will lead to an optimal performance. To the best of our knowledge, such a result in context of sparse super-resolution has not been studied in literature (cf. [1–4] and follow up work) and here we take a step towards this direction. In the remainder of this work, we develop and present our main result.

2. TOWARDS A BANDWIDTH SELECTION PRINCIPLE

In order to estimate the optimal bandwidth Ω_2, the blowup of the deconvolution error due to the vanishing denominator in (4) (as Ω increases from Ω_1 to Ω_3) must be balanced with the performance enhancement of the parameter estimation problem (3). While the former is directly related to the decay of $\widehat{\phi}$, the latter is in general an open problem in the theory of spectral estimation, especially in the finite sample case. Important progress in this direction has been made in recent years, starting with the seminal paper by Donoho [15], emphasizing *a-priori stability estimates* for the problem of super-resolution. Below we demonstrate how to use such kind of estimates for the problem of bandwidth selection.

To be more concrete, consider the model (2), and suppose that we are given the frequency domain measurements of the signal $m(t)$ in (5), where $e(t)$ is a bounded, white-noise process with $|e(t)| \le \eta$. Dividing $\widehat{m}(\omega)$ by $\widehat{\phi}$ (i.e. deconvolving), and using (4), we obtain

$$\frac{\widehat{m}(\omega)}{\widehat{\phi}(\omega)} = \sum_{k=0}^{K-1} c_k e^{-\jmath\omega t_k} + \widehat{e}_\phi(\omega), \quad |\omega| \leqslant \Omega \tag{6}$$

$$|\widehat{e}_\phi(\omega)| = \left|\frac{\widehat{e}(\omega)}{\widehat{\phi}(\omega)}\right| \leqslant \eta \cdot \underbrace{\left(\min_{|\omega|\leqslant\Omega} \left|\widehat{\phi}(\omega)\right|\right)^{-1}}_{:=\varepsilon_\Omega}. \tag{7}$$

For smooth pulses $\phi(t)$, the quantity ε_Ω will grow rapidly with Ω. This is because $\widehat{\phi}$ decays quickly (due to smoothness). On the other hand, increasing the sampling bandwidth in (6) also increases the accuracy of estimating t_k, however, the exact form of this dependency is both kernel and signal dependent. It is exactly at this point that the

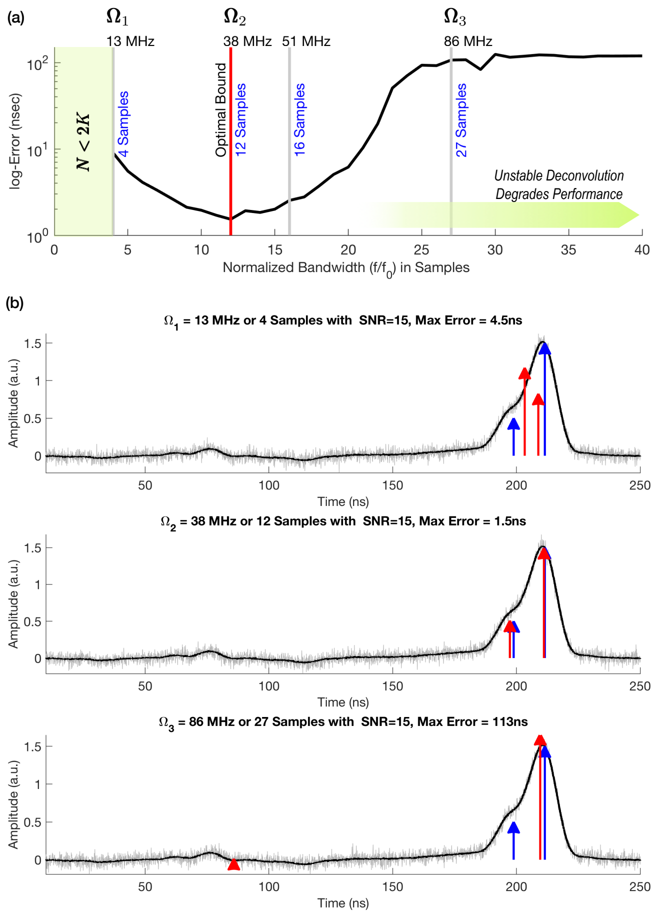

Fig. 2: Super-resolution is sensitive to bandwidth selection. (a) Estimation error as a function of bandwidth Ω. Let $\widetilde{t}_{k,\Omega}$ be the estimated spike support for a given Ω. The metric for error is chosen to be $\max|\widetilde{t}_{k,\Omega} - t_k|$ where t_k is the ground truth and the spikes are separated by $\Delta = |t_2 - t_1| = 12.6$ ns. Given noisy measurements $m(t)$ (5) corrupted by 15 dB additive white Gaussian noise, this plot is obtained by sweeping bandwidth Ω in the range $\omega_0 \in [2K, N]$ and averaged over 200 realizations. The optimal bandwidth in this case amounts minimum value of the error curve. (b) Corresponding reconstruction for three scenarios: (1) Critical sampling with $\Omega_1/\omega_0 = 2K$. Here, the reconstruction fails because of noise. (2) Optimal Ω_2 obtained by sweeping Ω and observing the minimum of the error in (a). This scheme benefits from over-sampling. (3) Excessive over-sampling when Ω_3 approaches the noise floor. Here reconstruction fails because unstable deconvolution leads to erroneous estimates of sinusoidal parameters in (3). We use matrix pencil method [11] for frequency estimation.

importance of *a-priori stability bounds* for super-resolution becomes apparent.

2.1. Super-resolution condition number

In order to bound the possible error in resolving the spike locations $\{t_k\}$ from the noisy data (6), we consider the *noise amplification factor*, or, in the language of numerical analysis, the *condition number* for the problem. In the following definition (first used in [16] in a somewhat heuristic way) we propose a certain quantity, which can be thought of as a first-order proxy for the stability of the nonlinear least squares fit of the sum-of-exponentials model from noisy data with arbitrary bounded noise.

Definition 2.1. Let $\underline{\theta} := [c_0, t_0, \ldots, c_{K-1}, t_{K-1}]^\top \in \mathbb{R}^{2K}$ be an unknown parameter vector, and denote by $\mathsf{F}(\underline{\theta}) : \mathbb{R}^{2K} \to \mathbb{C}^{2N+1}$ the parametric mapping

$$\mathsf{F}_{\underline{\theta}} = \left\{\sum\nolimits_{k=0}^{K-1} c_k e^{-\jmath n \omega_0 t_k}\right\}_{|n|\leqslant N}. \tag{8}$$

For each $i = 1, \ldots, 2K$, the *linearized condition number* corresponding to the i'th parameter, $\kappa^{(i)}(\mathsf{F}_{\underline{\theta}})$ is the ℓ_1 norm of the i-th row of the matrix $\mathbf{J}^\dagger$, where $\mathbf{J}$ is the Jacobian matrix of $\mathsf{F}_{\underline{\theta}}$, and $(\cdot)^\dagger$ is the Moore-Penrose pseudo-inverse.

Remark 2.1. It can be shown that the deterministic condition numbers $\kappa^{(i)}$ are equivalent to the Cramér-Rao Lower Bounds (CRLB) for the parameter identification problem (6), (7) in the particular case when $\widehat{e}_\phi(\omega)$ is white Gaussian noise. Unfortunately, the CRLB do not appear to be easily applicable for our question—estimating the accuracy of recovering t_k—as we numerically demonstrate in Fig. 3(b). See related works [9, 17].

Proposition 2.1. *Suppose that η in* (7) *is sufficiently small*[1], *and consider the solution of the nonlinear least squares problem*

$$\theta^{(\mathsf{LS})} = \arg\min_{\underline{\theta}} \underbrace{\sum\nolimits_{|n|\leq N} |\mathsf{F}_{\underline{\theta}}(n) - \widehat{s}(n\omega_0)|_2^2}_{:=\ell(\underline{\theta})}.$$

Then, in the sufficiently small neighborhood of the original parameter vector $\underline{\theta} = \underline{\theta}^{(0)}$ and $\eta = 0$, we have, for some constant $c > 0$, the following upper bound for each component of the solution:

$$\left|\theta_i^{(\mathsf{LS})} - \theta_i^{(0)}\right| \leq c\kappa^{(i)}\varepsilon_\Omega \eta. \tag{9}$$

Proof outline. In the neighborhood of the optimal value $\underline{\theta}^{(0)}$, the least squares loss function $\ell(\underline{\theta})$ is twice continuously differentiable. Furthermore, the Hessian matrix $\mathbf{H}$ of $\ell(\underline{\theta})$ is precisely $\mathbf{H} = \mathcal{J}^*\mathcal{J}$, where $\mathcal{J}$ is the Jacobian matrix of the "realification" of the map $\mathsf{F}_{\underline{\theta}}$ from (8) (i.e. replacing each complex coordinate with corresponding two real coordinates). It can be verified that $\mathcal{J}$ is full-rank whenever the spike locations are pairwise different and the amplitudes are bounded from zero. Therefore, $\mathbf{H}$ is positive definite. Using the stability theory of nonlinear programming (see e.g. [18, Corollary 3.2.3], based on the implicit function theorem), there exists a unique continuously differentiable solution function $\theta^{(\mathsf{LS})}$ (of both $\underline{\theta}^{(0)}$ and η) whose gradient vector w.r.t η, $\nabla_\eta \theta^{(\mathsf{LS})}$ can be shown to be $\mathbf{H}^{-1} \cdot \mathcal{J}^* \cdot \nabla_\eta \widehat{s}(n\omega_0)$, and therefore its i-th entry is bounded by the ℓ_1 norm of the corresponding row of $\mathcal{J}^\dagger$. Since $\mathcal{J}^\dagger$ and $\mathbf{J}^\dagger$ are related (cf. [19]) it can be shown that there exists a constant c such that

$$\left|\nabla_\eta \theta^{(\mathsf{LS})}|_{\eta=0}\right|_i \leq c\kappa^{(i)}\varepsilon_\Omega.$$

The bound (9) follows by Taylor expansion of $\theta^{(\mathsf{LS})}$ w.r.t. η. □

2.2. The bandwidth selection criterion

The condition numbers $\kappa^{(i)}$ will in general depend on both Ω and the point $\underline{\theta}$. However, provided some a-priori knowledge on the possible parameter ranges of the components of $\underline{\theta}$, it is sometimes possible to obtain bounds which are *uniform* over these ranges. We shall consider several examples in the next section, but let us now return to the bandwidth selection problem.

[1] Here η is considered as a small parameter such that all the relevant quantities are differentiable with respect to it. For futher details see [18].

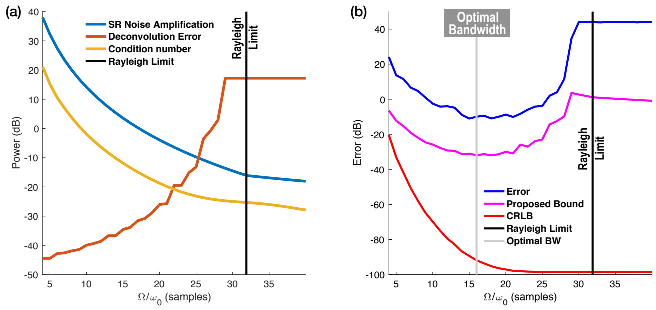

Fig. 3: In order to demonstrate our method of optimal bandwidth selection, we have generated synthetic signals using the ToF kernel, with $K = 2$ spikes separated by $\Delta = 9.8$ ns, and corrupted by bounded noise with SNR $= 20$ dB. (a) The curves corresponding to 1) ε_Ω (red), the blowup of the deconvolution error (6), 2) the actual condition number $\kappa^{(i)}$ (yellow) obtained from the Definition 2.1, and 3) the simplified bound $G(\Omega)$ combining (11) and (12) (blue). (b) The actual error $|\widetilde{t}_{k,\Omega} - t_k|$ (blue), and the bound $G(\Omega)\varepsilon_\Omega$ (magenta), obtained from multiplying the corresponding curves in (a). The Rayleigh threshold is approximately 32 samples. We can see that both the actual optimal bandwidth, and the predicted one, are pretty close to each other, and both of them are below the Rayleigh limit. For comparison, the CRLB for the parameter identification problem $\widehat{m}(n\omega_0) = \widehat{\phi}(n\omega_0)\widehat{s}(n\omega_0) + \widehat{e}(n\omega_0)$, with $|n| \leq \Omega/\omega_0$ and $e(t)$ white Gaussian noise, is shown in red, and it is difficult to identify the optimal value of Ω based on it alone. The divergence of CRLB from the actual error may be partly due to the fact that the noise covariance matrix of $\widehat{e}_\phi$ in (7) becomes badly conditioned for large Ω, and matrix pencil being non-optimal in such a regime.

For simplicity, suppose that we are only interested in recovery of the spike locations ($i = 2k+1$, for $k = 0, \dots, K-1$). Using (9) and (7), and assuming that there exists an explicitly computable function $G(\Omega, D)$ such that

$$\sup_{\underline{\theta} \in D, k=0,\dots,K-1} \kappa^{(2k+1)}(\underline{\theta}, \Omega) \leq G(\Omega, D),$$

uniformly in some parameter domain $\underline{\theta} \in D \subset \mathbb{R}^{2K}$, we finally introduce the bandwidth selection criterion:

$$\Omega_{\text{opt}}(D) = \arg\min_{\Omega} \left\{ G(\Omega, D) \cdot \varepsilon_\Omega \right\}, \qquad (10)$$

where ε_Ω is defined in (7). For shortness of notation, we shall omit D and write $G(\Omega)$, when D should be obvious from the context (for instance, one such D is explicitly defined in Theorem 2.1 below).

2.3. Simple bounds on $G(\Omega)$

Obtaining tight bounds on the stability of the super-resolution problem is a hard theoretical question, considered in various forms in many publications [4, 15, 20–22], including by the first author [16, 23–27]. By now it is well-understood that there exist two very different stability regimes, depending on whether the minimal separation between the spikes is above or below the Rayleigh threshold $\frac{1}{\Omega}$.

Here we present a certain simple version of such a bound for $\kappa^{(i)}$ defined above, based on the recent results in [16]. A more general result, for "partially clustered" configurations, is proved in the forthcoming paper [26], using recent progress made in [27].

Theorem 2.1. *Suppose that for all $\underline{\theta}$ in the domain $D \subset \mathbb{R}^{2K}$, the amplitudes are bounded: $0 < A_1 \leq |c_k| \leq A_2$, and also that the minimal distance between the Diracs is bounded by Δ, that is $\min_{j \neq k} |t_j - t_k| \geq \Delta > 0$. There exist constants c_1, c_2, c_3, depending on A_1, A_2, K, such that the following bounds hold.*

1. *If $\Delta > \frac{c_1}{\Omega}$ (**well-separated regime**), then*

$$\kappa^{(i)} \leq \frac{c_2}{\Omega}, \quad i = 1, 3, \dots, 2K-1. \qquad (11)$$

2. *If $\max_{j \neq k} |t_j - t_k| < \frac{2\pi K}{\Omega}$ (**single cluster regime**), then*

$$\kappa^{(i)} \leq \frac{c_3}{\Omega} \left(\frac{1}{\Omega\Delta} \right)^{2K-2}, \quad i = 1, 3, \dots, 2K-1. \qquad (12)$$

Outline of proof. The bound (11) follows directly from the first part of Theorem 6.1 in [16]. To prove (12), we factorize the Jacobian matrix of $\mathsf{F}_{\underline{\theta}}$ into a *Pascal-Vandermonde* matrix $\mathbf{U}$ (close relative of the confluent Vandermonde matrix, [23]) and a simple block-diagonal term. Then we can use Gautschi's estimates for the row-wise norms of $\mathbf{U}_0^{-1}$ (where $\mathbf{U}_0$ is an appropriate square Pascal-Vandermonde matrix) from [28], combining it with the "decimation" technique developed in [16]. Full details of the proof shall be provided later. □

Remark 2.2. Under the minimal separation assumption (corresponding to (11)), the super-resolution problem can be solved with provable guarantees by a variety of methods, including non-parametric and parametric spectral estimation [12], and more recently convex programming techniques [4]. Under the single cluster scenario, a nearly-optimal algorithm is presented in [29] using the "decimation" technique.

Remark 2.3. The factor $\frac{1}{\Delta\Omega}$ is called the "super-resolution factor" (SRF) in the literature, and it quantifies the difficulty of super-resolution under sparsity constraints. The bound (12) depends polynomially on SRF, but *exponentially* on K (cf. [30].) Still, provided the perturbation bound η is small enough, super-resolution is possible, and it is an important open problem to quantify *exactly how large* is η allowed to be. Initial results have been presented in [25], and further details will be available in [26].

The usefulness of the above bounds is demonstrated numerically, see Fig. 3.

3. CONCLUSIONS AND FUTURE WORK

Inspired by practical applications where spikes are filtered with smooth pulses (as opposed to exact bandlimited kernels), in this work we posed the problem of *optimal bandwidth selection*. The notion of optimal or essential bandwidth allows for extending the utility of off-the-shelf super-resolution algorithms to the case of smooth filters. By devising a bandwidth selection criterion which is easy to compute and provides reasonable performance, we take a first step towards the direction of achieving *optimal performance limits*, especially in the finite sample/sub-Rayleigh case. The results presented in this work and empirical studies performed on experimental data raise a number of interesting questions for future investigation. On the theoretical side, the relationship between our bounds and Cramér-Rao Lower Bounds (CRLB) is an interesting open question. Another important issue is to develop *robust* selection criteria, which will allow stable recovery even when the a-priori parameter domain D is known only approximately (such as in a Bayesian framework), and incorporating possible information about the noise distribution (note a related recent study of super-resolution from the angle of numerical analytic continuation [31]). From an applications standpoint, tailoring our work to other applications such as source localization and direction-of-arrival estimation [32] are interesting future directions.

4. REFERENCES

[1] L. Li and T. P. Speed, "Parametric deconvolution of positive spike trains," *Annals of Statistics*, pp. 1279–1301, 2000.

[2] 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, 2002.

[3] 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, pp. 31–40, 2008.

[4] E. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[5] H. Pan, T. Blu, and M. Vetterli, "Towards generalized FRI sampling with an application to source resolution in radioastronomy," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 821–835, Feb. 2017.

[6] A. Bhandari and T. Blu, "FRI sampling and time-varying pulses: Some theory and four short stories," in *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Mar. 2017.

[7] G. Plonka, K. Stampfer, and I. Keller, "Reconstruction of stationary and non-stationary signals by the generalized prony method," *Analysis and Applications*, pp. 1–32, jul 2018.

[8] S. Rudresh, S. Nagesh, and C. S. Seelamantula, "Asymmetric pulse modeling for FRI sampling," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2027–2040, apr 2018.

[9] J. A. Uriguen, T. Blu, and P. L. Dragotti, "FRI sampling with arbitrary kernels," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5310–5323, 2013.

[10] A. Bhandari and R. Raskar, "Signal processing for time-of-flight imaging sensors," *IEEE Signal Process. Mag.*, vol. 33, no. 4, pp. 2–16, Sep. 2016.

[11] Y. Hua and T. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, may 1990.

[12] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson/Prentice Hall, 2005.

[13] A. Bhandari, E. M. Strohm, M. C. Kolios, and R. Raskar, "Photoacoustic ToF tomography of blood cells: From mathematical approximation to super-resolution," in *Intl. Conf. on Sampling Theory and Applications (SampTA)*, 2017.

[14] J. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, 1988.

[15] D. Donoho, "Superresolution via sparsity constraints," *SIAM Journal on Mathematical Analysis*, vol. 23, no. 5, pp. 1309–1331, 1992.

[16] D. Batenkov, "Stability and super-resolution of generalized spike recovery," *Applied and Computational Harmonic Analysis*, vol. 45, no. 2, pp. 299–323, Sep. 2018.

[17] S. Basu and Y. Bresler, "The stability of nonlinear least squares problems and the Cramer-Rao bound," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3426–3436, Dec. 2000.

[18] A. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Academic Pr, 1983.

[19] L. Sorber, M. Barel, and L. Lathauwer, "Unconstrained Optimization of Real Functions in Complex Variables," *SIAM Journal on Optimization*, vol. 22, no. 3, pp. 879–898, Jan. 2012.

[20] L. Demanet and N. Nguyen, "The recoverability limit for superresolution via sparsity," 2014.

[21] V. Morgenshtern and E. Candès, "Super-Resolution of Positive Sources: The Discrete Setup," *SIAM Journal on Imaging Sciences*, vol. 9, no. 1, pp. 412–444, Jan. 2016.

[22] W. Li and W. Liao, "Stable super-resolution limit and smallest singular value of restricted Fourier matrices," *arXiv:1709.03146 [cs, math]*, Sep. 2017.

[23] D. Batenkov and Y. Yomdin, "On the accuracy of solving confluent Prony systems," *SIAM J. Appl. Math.*, vol. 73, no. 1, pp. 134–154, 2013.

[24] ——, "Algebraic signal sampling, Gibbs phenomenon and Prony-type systems," in *Intl. Conf. on Sampling Theory and Applications (SampTA)*, 2013.

[25] A. Akinshin, D. Batenkov, and Y. Yomdin, "Accuracy of spike-train Fourier reconstruction for colliding nodes," in *2015 International Conference on Sampling Theory and Applications (SampTA)*, May 2015, pp. 617–621.

[26] D. Batenkov, G. Goldman, and Y. Yomdin, "Super-resolution of near-colliding point sources," *Manuscript in preparation*, 2018.

[27] D. Batenkov, L. Demanet, G. Goldman, and Y. Yomdin, "Stability of partial Fourier matrices with clustered nodes," *arXiv:1809.00658 [cs, math]*, Sep. 2018.

[28] W. Gautschi, "On inverses of Vandermonde and confluent Vandermonde matrices," *Numerische Mathematik*, vol. 4, no. 1, pp. 117–123, 1962.

[29] D. Batenkov, "Accurate solution of near-colliding Prony systems via decimation and homotopy continuation," *Theoretical Computer Science*, vol. 681, pp. 27–40, Jun. 2017.

[30] A. Moitra, "Super-resolution, Extremal Functions and the Condition Number of Vandermonde Matrices," in *Proc. of Forty-Seventh ACM Symp. on Theory of Computing*, ser. STOC '15, New York, NY, USA, 2015, pp. 821–830.

[31] D. Batenkov, L. Demanet, and H. N. Mhaskar, "Stable soft extrapolation of entire functions," *Inverse Problems*, vol. 35, no. 1, p. 015011, Jan. 2019.

[32] T. E. Tuncer and B. Friedlander, *Classical and Modern Direction-of-Arrival Estimation*. London: Academic, 2009, oCLC: 299711034.