The Annals of Statistics
2019, Vol. 47, No. 3, 1754–1775
https://doi.org/10.1214/18-AOS1736

SUPER-RESOLUTION ESTIMATION OF CYCLIC ARRIVAL RATES

BY NINGYUAN CHEN[*,1], DONALD K. K. LEE[†] AND SAHAND N. NEGAHBAN[†,2]

HKUST[*] and Yale University[†]

Exploiting the fact that most arrival processes exhibit cyclic behaviour, we propose a simple procedure for estimating the intensity of a nonhomogeneous Poisson process. The estimator is the super-resolution analogue to Shao (2010) and Shao and Lii [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** (2011) 99–122], which is a sum of p sinusoids where p and the amplitude and phase of each wave are not known and need to be estimated. This results in an interpretable yet flexible specification that is suitable for use in modelling as well as in high resolution simulations.

Our estimation procedure sits in between classic periodogram methods and atomic/total variation norm thresholding. Through a novel use of window functions in the point process domain, our approach attains super-resolution without semidefinite programming. Under suitable conditions, finite sample guarantees can be derived for our procedure. These resolve some open questions and expand existing results in spectral estimation literature.

1. Introduction. Real world arrival patterns typically exhibit cyclic (but not necessarily periodic) behaviour. Motivated by the need for tractable yet flexible functional forms for the arrival rate in queuing literature (Chen, Lee and Shen (2018)), we consider the following problem: Suppose we observe the jump times $\{t_j\}_j$ of a nonhomogeneous Poisson process (NHPP) $\{N(t) : t \geq 0\}$ in $[0, T]$. Here, $N(t)$ denotes the number of arrivals in $(0, t]$, and the intensity $\lambda(t)$ and the cumulative rate function $\Lambda(t)$ are defined as

$$\mathbb{E}N(t) = \int_0^t \lambda(u)\,du = \Lambda(t).$$

Our goal is to use the observed data to estimate arrival rates of the form

$$\lambda(t) = c_0^\lambda + \sum_{j=1}^{p/2} d_j^\lambda \cos\bigl(f_j^\lambda t + \phi_j^\lambda\bigr) = c_0^\lambda + \sum_{k=1}^{p} c_k^\lambda e^{2\pi i \nu_k^\lambda t}, \tag{1.1}$$

where the even number p of frequency components, the frequencies $\nu^\lambda = \{\nu_k^\lambda\}_k$ in a pre-specified band $[-B, +B]$, and the complex coefficients $c^\lambda = \{c_k^\lambda\}_k$ are all

Received June 2017; revised June 2018.

[1]Supported by the HKUST start-up fund R9382.

[2]Supported by the NSF Award DMS 1723128.

MSC2010 subject classifications. Primary 62M15, 90B22; secondary 60G55.

Key words and phrases. Spectral estimation, periodogram, window function, thresholding, nonhomogeneous Poisson process, queueing theory.

Algorithm 1 The proposed estimation procedure

1 Define the windowed periodogram for the point process as

$$|H(\nu)| = \frac{1}{T}\left|\sum_j w(t_j)e^{-2\pi i \nu t_j}\right|$$

for $|\nu| \le B$, and note that it is symmetric in ν. The sum can be computed efficiently using nonuniform FFT algorithms like Dutt and Rokhlin (1993).

2 Identify the frequency region $R = \{\nu : r \le |\nu| \le B, |H(\nu)| > \tau\}$ where the value of periodogram exceeds the threshold τ.

3 Set $\nu_0^\lambda = \hat{\nu}_0 = 0$, $k = 1$ and repeat the following steps:

- Find the highest stationary peak of the periodogram in R and set $\hat{\nu}_k$ as the corresponding frequency location. If no peaks exist then exit loop.
- Perform the updates $k \leftarrow k + 1$ and $R \leftarrow R\backslash(\hat{\nu}_k - r, \hat{\nu}_k + r)$. This removes a neighbourhood of radius r centred at $\hat{\nu}_k$ from R.

4 Compute the estimator (4.1) for c_k^λ.

unknown. Given the connections to Fourier series, this specification is very flexible and was introduced by (Shao (2010), Shao and Lii (2011)). They resolve the estimation problem under the classical setting where the frequencies are assumed to be spaced more than order $1/T$ apart. In this paper, we examine the problem from the super-resolution perspective: We propose a simple procedure for estimating (1.1) when the frequencies can be up to order $1/T$ of each other. This is the finest possible resolution in the sense that no estimator can generally resolve frequencies separated by less than $1/T$ in the presence of noise (Moitra (2015)).

Our approach modifies the classic periodogram and combines it with the super-resolution literature on total-variation/atomic norm regularization. Three ingredients (to be specified in Proposition 3) are used in Algorithm 1: (i) A window function $w(t)$ supported on $[0, T]$; (ii) a threshold $\tau > 0$; and (iii) a neighbourhood exclusion radius $r > 0$. The simple but elegant intuition behind the thresholding idea (Donoho and Johnstone (1994)) as applied to our situation is that the spectral energy [given by $|H(\nu)|$ as defined in the algorithm] should be concentrated at the signal frequencies $\nu_0^\lambda, \ldots, \nu_p^\lambda$. If the signals are strong enough that $|H(\nu_0^\lambda)|, \ldots, |H(\nu_p^\lambda)|$ exceed the ambient noise level, then setting τ above the noise level will result in the algorithm isolating a neighbourhood around each ν_k^λ (see Figure 1). It will be shown that if the frequencies are separated from one another by a gap (resolution) of at least $g(T)/T$ where $g(T) \ge 4$, then with high probability our procedure will recover each ν_k^λ with a precision of $2/T$, provided that the dynamic range of the amplitudes $\max_k |c_k^\lambda| / \min_k |c_k^\lambda|$ is less than 14.5. This can be dramatically relaxed as the frequency gap is increased: For example

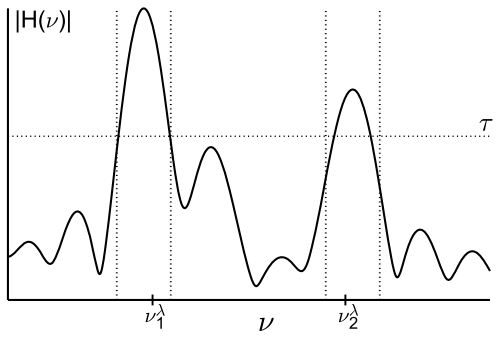
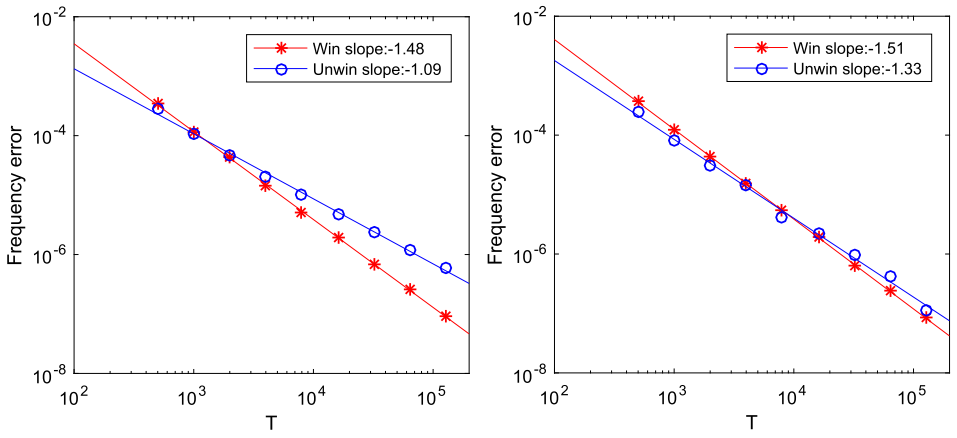


FIG. 1. *Visualization of Algorithm 1. In the depicted periodogram there are two signal frequencies* ν_1^λ *and* ν_2^λ. *Setting* τ (*horizontal line*) *above the ambient noise results in the algorithm selecting neighbourhoods* (*between the pairs of vertical lines*) *that contain* ν_1^λ *and* ν_2^λ.

if $g(T) \geq 6$, then the maximum allowable dynamic range exceeds 100. As discussed in Section 3 of Shao and Lii (2011), some sort of dynamic range condition is needed even in the classical setting where the frequency gap is larger than order $1/T$. Our analysis provides a way for quantifying the maximum allowable range when T is finite.

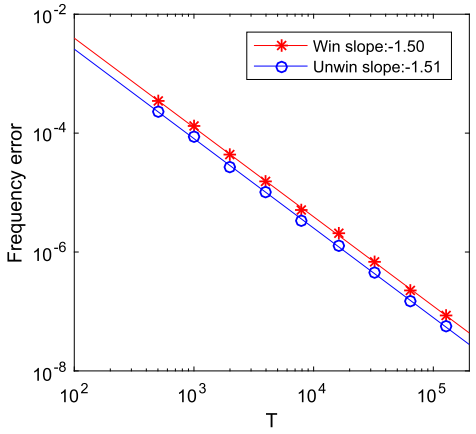
A notable aspect of our methodology is in introducing the windowed periodogram to the point process domain: Bartlett's (1963) classic "unwindowed" periodogram for point processes is essentially $|\sum_j e^{-2\pi i \nu t_j}|/T$, which is a special case of $|H(\nu)|$ when $w(t)$ is the rectangle window on $[0, T]$. We show that this window can and should be replaced with one that has faster decaying spectral tails. Doing so has two benefits. First, super-resolution can be achieved without needing to solve a semidefinite program. Second, even under the classical setting where the frequencies are spaced more than order $1/T$ apart [$g(T) \to \infty$ as $T \to \infty$], frequency estimation is more precise with a windowed periodogram. For example, Figure 2 presents a log–log plot of the frequency estimation error versus T for various choices of $g(T)$. The details of this experiment are elaborated upon in Section 5.1. The plots show that the rate of convergence increases with $g(T)$, with the windowed periodogram outperforming the unwindowed one until $g(T)$ reaches order $T^{1/2}$, whereupon both achieve the maximum rate of $\mathcal{O}(T^{-3/2})$ as predicted by theory.

The remainder of this paper is organized as follows. Our contributions to existing literature will be described below. Section 2 reviews some basic results from signal processing and shows how spectral leakage and windowing manifest themselves in arrivals data. This motivates the design of our estimation procedure. Frequency recovery is discussed in Section 3 where $w(t)$, τ and r are specified. Under the conditions given in the section, it will be shown that our procedure will recover all frequencies to within a precision of $2/T$ with high probability. We will also articulate the tradeoffs involved relative to methods designed for the classical



(a) $g(T) = 6$.

(b) $g(T) = T^{1/6}$.



(c) $g(T) = T^{1/2}$.

FIG. 2. *Frequency recovery error* $\max_k |\hat{\nu}_k - \nu_k^{\lambda}|$ *for the simulation* (5.4) *as a function of* T, *for* $g(T) \in \{6, T^{1/6}, T^{1/2}\}$. *"Win" refers to the windowed periodogram and "Unwin" refers to the classic one. The error rates are* $\max_k |\hat{\nu}_k - \nu_k^{\lambda}| \sim T^s$ *where* s *is the slope of the relevant fitted line.*

resolution setting. The estimator for the corresponding amplitudes and phases is given in Section 4. In Section 5, we use simulations to compare our procedure to the model selection approach in Shao (2010). Concluding remarks can be found in Section 6. The proofs of all results can be found in the Supplementary Material (Chen, Lee and Negahban, 2018).

Contributions to literature. Our estimation procedure sits in between two streams of literature on spectral estimation. At one end, the classic approach is to visually inspect the unwindowed periodogram for point processes (Bartlett (1963)) to find frequencies corresponding to peaks in the plot. If it is assumed that there is only one frequency (Lewis (1970), Vere-Jones (1982)), the frequency corre-

sponding to the largest peak of the periodogram is selected. Other approaches (Bebbington and Zitikis (2004), Belitser, Serra and van Zanten (2013), Helmers and Mangku (2003)) also exist but it is unclear if they generalize to the setting with multiple frequencies.

Under the classical setting where the frequency gap is assumed to be $1/o(T)$, Shao (2010), Shao and Lii (2011) extend the periodogram method to the multiple frequencies setting. Their procedure corresponds to setting $R = \{\nu : r \leq |\nu| \leq B\}$ in Algorithm 1 and running step 3 until p frequencies have been selected. The choice of p is determined using the AIC/BIC model selection criterion derived in the dissertation of Shao (2010). For BIC, Shao (2010) states that the probability of selecting the true p eventually approaches 1 as $T \to \infty$. Our procedure builds on Shao (2010), Shao and Lii (2011) in two directions. First, the use of windowing enables periodogram methods to achieve super-resolution, and this can be combined with either thresholding or model selection to estimate p. Second, finite sample performance bounds can be derived for our thresholding approach. This complements the BIC approach which does not come with high probability guarantees for finite data. Moreover, the bounds also provide a way for quantifying the allowable amplitude dynamic range when T is finite. As discussed in Section 3 of Shao and Lii (2011), some sort of dynamic range condition is needed even if the frequency gap is larger than order $1/T$. Of course, this will be more restrictive in the super-resolution setting, so there will be a cost to using our approach if the frequencies are in fact spaced far apart. This tradeoff will be discussed in Section 3.

The other related stream of work is the super-resolution literature that uses total-variation or atomic norm regularization to select frequencies (Bhaskar, Tang and Recht (2013), Candès and Fernandez-Granda (2013), Fernandez-Granda (2013), Tang, Bhaskar and Recht (2015)). These papers study a generic spectral estimation problem in a discrete time setting. They generalize the ℓ_1-norm for a finite number of variables to the case where there is a continuum of predictors $\{e^{i\omega t}\}_\omega$. An infinite dimensional extension of Lasso is then formulated and solved as a semidefinite program to select predictors and their coefficients. While the authors show that this method outperforms existing ones for the setting described, challenges arise when trying to adapt it to our problem. First, the arrival counts must be discretized into time bins, which introduces aliasing effects.[3] Second, the required computational effort is overly taxing[4] for the size of problems we consider. For example, Chen, Lee and Shen (2018) analyzes 652 days of arrivals data from an emergency department and used 5216 bins of 3-hour widths for the Lasso extension. The ADMM implementation recommended in Bhaskar, Tang and Recht

[3]This can, however, be overcome using bins narrower than $1/(2B)$ (Nyquist sampling).

[4]A Lasso approximation obtained from discretizing the frequency domain is suggested in Bhaskar, Tang and Recht (2013) as a speedup. However, this is still more difficult to implement than the periodogram method, along with the additional downside of a fixed discretized frequency grid.

(2013) takes at least ten days to run on a computer with Intel i7 6500 cores. By contrast, our procedure takes only a few minutes.

In terms of frequency recovery, the approaches in Fernandez-Granda (2013), Tang, Bhaskar and Recht (2015) are guaranteed to pick out one or more frequencies within some C/T of each signal frequency when the resolution is $4/T$. In the stochastic noise setting of Tang, Bhaskar and Recht (2015), the guarantee holds with high probability, and they further conjecture that it is possible to prevent the selection of spurious frequencies. We contribute to this literature by resolving the conjecture in the affirmative, since our procedure recovers exactly $p+1$ frequencies with high probability, one within $2/T$ of each true signal. The tradeoff with using a periodogram method is that a bound on the dynamic range of the amplitudes is needed. However, as mentioned earlier, this can be dramatically relaxed by widening the frequency gap slightly, from $4/T$ to $6/T$ for example.

2. Overview of the estimation approach. Let the continuum of complex exponentials $\{e^{2\pi i\nu t}\}_{|\nu|\le B}$ be our dictionary for constructing an arrival rate. Suppose the rate for the underlying NHPP is (1.1), which belongs in the collection

$$\left\{c_0+\sum_{k=1}^{p}c_k e^{2\pi i\nu_k t} : c_k\in\mathbb{C},\, p<\infty\right\}. \tag{2.1}$$

Since $\lambda(t)$ is real-valued, (1.1) will lie in the subset where the presence of (c_k,ν_k) implies its conjugate $(\bar{c}_k,-\nu_k)$, so in particular c_0 will be real and positive. The quantity of interest is the $(p+1)$-vector ν^λ of frequencies in (1.1), where p is even but unknown. Given these, the coefficients c^λ in (1.1) will be estimated by the complex-valued least squares solution (4.1) described in Section 4. Since $\lambda(t)$ is unobservable, we only see arrivals in the time window $[0,T]$. Estimating the intensity therefore becomes a question of recovering ν^λ from the frequency components in the trajectory $\{N(t)\}_{t\in[0,T]}$. To make the connection between the spectrums of the two quantities clearer, rewrite the latter in its Doob–Meyer form of signal and noise components

$$\begin{aligned}\{dN(t)\}_{t\in[0,T]} &= \big[d\Lambda(t)+d\{N(t)-\Lambda(t)\}\big]I_{(0,T]}(t)\\ &= \lambda(t)I_{(0,T]}(t)\,dt+d\varepsilon(t)I_{(0,T]}(t),\end{aligned} \tag{2.2}$$

where $I_{(0,T]}(t)$ is the indicator function of $\{0<t\le T\}$. Even in the absence of noise, the spectrum of the signal component $\lambda(t)I_{(0,T]}(t)$ is itself a distorted version of the one for $\lambda(t)$: Denoting the Fourier transform of $f(t)$ as

$$\tilde{f}(\nu)=\int f(t)e^{-2\pi i\nu t}\,dt,$$

we can write the spectrum of $\lambda(t)$ as the sum of the Dirac delta spikes centred at $\{\nu_k^\lambda\}_k$:

$$\tilde{\lambda}(\nu)=\sum_{k=0}^{p}c_k^\lambda\delta\big(\nu-\nu_k^\lambda\big).$$

On the other hand, $\lambda(t)I_{(0,T]}(t)$ is the result of truncating $\lambda(t)$ due to T being finite, a spectrum distorting operation known as leakage: Denote the convolution operator $*$ by $f * h(t) = \int f(s)h(t-s)\,ds$, the h-smoothed average of f about the point t. The spectrum of $\lambda(t)I_{(0,T]}(t)$ is

$$(\widetilde{\lambda \cdot I_{(0,T]}})(\nu) = (\tilde{\lambda} * \tilde{I}_{(0,T]})(\nu) = \sum_{k=0}^{p} c_k^{\lambda} \tilde{I}_{(0,T]}\big(\nu - \nu_k^{\lambda}\big), \tag{2.3}$$

a weighted average of $\lambda(t)$'s spectral values c_k^{λ} concentrated at $\{\nu_k^{\lambda}\}_k$. Thus truncation has the effect of smearing the frequency spikes in $\tilde{\lambda}(\nu)$ into a continuous spectrum: For $\nu \notin \bigcup_k \{\nu_k^{\lambda}\}$, $\widetilde{\lambda \cdot I_{(0,T]}}(\nu)$ can have a nonzero value, creating an artificial noise floor. The noise floor around strong signal frequencies may mask weaker neighbouring signals, leading to resolution loss and making it difficult to recover ν^{λ} from $\lambda(t)I_{(0,T]}(t)$. Leakage distortion is a manifestation of the uncertainty principle because perfect frequency localization requires $\tilde{I}_{(0,T]}(\nu) = \delta(\nu)$, but this is only possible if $I_{(0,T]}(t) = 1$, that is, an infinite time window is needed.

The key idea that Algorithm 1 uses to deal with leakage is to replace $I_{(0,T]}(t)$ with a suitably chosen window function $w(t)$ to obtain the weighted arrival process $dN^w(t) = w(t)\,dN(t)$. We see from (2.3) that the extent of leakage depends on the tail decay of $\tilde{I}_{(0,T]}$, as this dictates the influence that distant frequencies has on the local spectral value. Since $\lambda(t)$ can be truncated to $(0, T]$ using any $w(t)$ supported on $(0, T]$, we can multiply $\lambda(t)$ with one whose Fourier transform has lighter tails.

While the usual anti-leakage benefits of nonuniform windows is well known in signal processing, they are in fact needed in our procedure for attaining frequency resolutions of order $1/T$: The tail decay of $\tilde{I}_{(0,T]}(\nu)$ is of order $1/(T\nu)$. Thus if $\{\nu_k^{\lambda}\}_k$ are spaced $1/T$ apart, the leakage (2.3) around a neighbourhood of ν_k^{λ} from the other frequencies can be of order $\log p$ for the rectangle window. This can easily mask the periodogram spike at ν_k^{λ} when p is large enough. Hence the classic periodogram method is generally unable to attain frequency resolutions of order $1/T$. Interestingly, the window that is usually considered optimal for signal processing[5] is actually suboptimal for frequency recovery: Theorem 3.44 of Osipov, Rokhlin and Xiao (2013) shows that the spectral tail decay of the prolate spheroidal function is also of order $1/(T\nu)$ when it is time-limited to $[0, T]$.

[5]Optimal in the sense that its spectrum is the one that is most concentrated about the origin.

Returning to the problem of recovering ν^λ from (2.2), consider the $(1/T)$-scaled spectrum of the windowed data $dN^w(t) = w(t)\,dN(t)$:

$$\begin{aligned} H(\nu) &= \frac{1}{T}\int_0^T e^{-2\pi i \nu t}\,dN^w(t) \\ &= \frac{1}{T}\int_0^T e^{-2\pi i \nu t} w(t)\lambda(t)\,dt + \frac{1}{T}\int_0^T e^{-2\pi i \nu t} w(t)\,d\varepsilon(t) \\ &= \frac{1}{T}\sum_{k=0}^{p} c_k^\lambda \tilde{w}\big(\nu - \nu_k^\lambda\big) + \frac{\tilde{\varepsilon}^w(\nu)}{T}. \end{aligned} \tag{2.4}$$

Recall from Algorithm 1 that $|H(\nu)|$ is defined as the windowed periodogram. For ν sufficiently far from $\{\nu_k^\lambda\}_k$, the noise level outside the vicinity of these frequencies should be low for light-tailed $\tilde{w}$:

$$|H(\nu)| \le \|c^\lambda\|_\infty \sum_{k=0}^{p} \frac{|\tilde{w}(\nu - \nu_k^\lambda)|}{T} + \sup_{\nu\in[0,B]} \frac{|\tilde{\varepsilon}^w(\nu)|}{T}. \tag{2.5}$$

If the signal strengths c^λ are sufficiently strong, then intuitively a neighbourhood of $\bigcup_k\{\nu_k^\lambda\}$ can be isolated by simply excluding frequency regions in $[-B, +B]$ where $|H(\nu)|$ is below some threshold τ (see Figure 1). This is the idea behind step 2 of Algorithm 1. The analysis presented in the next section will guide our choices for $w(t)$, τ and r in our estimation procedure.

3. Frequency recovery. To guarantee that Algorithm 1 will recover the true signal frequencies ν^λ with high probability, we will assume that conditions A1 and A2 given in this section hold from the point they are stated. First, since no method can distinguish among frequencies that are clustered arbitrarily close together, we impose a minimum separation gap.

A1. For $0 \le k, k' \le p$, $\min_{k\neq k'} |\nu_k^\lambda - \nu_{k'}^\lambda| \ge \frac{g(T)}{T}$ for some $g(T) \ge 4$.

The gap $g(T)/T$ represents the frequency resolution for our procedure, and our recovery results cover all possible rates of growth for $g(T)$ as $T \to \infty$. The lower bound of $4/T$ benchmarks the frequency gap employed in the super-resolution literature (Fernandez-Granda (2013), Tang, Bhaskar and Recht (2015)). If instead the benchmark target is the classical setting in Shao (2010), Shao and Lii (2011), then A1 may be relaxed to $6/T$, see the remark following Proposition 3 below.

Under A1, we must localize each ν_k^λ to within a neighbourhood of radius $2/T$ to avoid possible ambiguity from overlapping. To achieve this with thresholding, note from (2.5) that if ν is at least $2/T$ away from the nearest ν_k^λ, then $|H(\nu)|$ is

strictly less than[6]

$$\underbrace{\frac{2}{T}\sum_{l=0}^{\infty}\sup_{|\nu|\geq\frac{2}{T}+\frac{4}{T}l}|\tilde{w}(\nu)|}_{S_1}\cdot\|c^{\lambda}\|_{\infty}+\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T}, \tag{3.1}$$

where the tail sum S_1 bounds the leakage noise floor outside the vicinity of $\{\nu_k^{\lambda}\}_k$, and the last term is the statistical noise level. The unknown $\|c^{\lambda}\|_{\infty}$ can be estimated using the highest peak of the periodogram: It is shown in the Supplementary Material that

$$\left(\frac{|\tilde{w}(0)|}{T}-\underbrace{\frac{2}{T}\sum_{l=1}^{\infty}\sup_{|\nu|\geq\frac{4}{T}l}|\tilde{w}(\nu)|}_{S_2}\right)\|c^{\lambda}\|_{\infty}-\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T}\leq\sup_{\nu\in[0,B]}|H(\nu)|, \tag{3.2}$$

$$\sup_{\nu\in[0,B]}|H(\nu)|\leq\max\left(S_1,\frac{|\tilde{w}(0)|}{T}+\frac{S_1+S_2}{2}\right)\|c^{\lambda}\|_{\infty}+\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T}. \tag{3.3}$$

Substituting the bound (3.2) for $\|c^{\lambda}\|_{\infty}$ into (3.1) shows that the threshold level τ in Algorithm 1 should be

$$\frac{S_1}{|\tilde{w}(0)|/T-S_2}\sup_{\nu\in[0,B]}|H(\nu)|+\left(\frac{S_1}{|\tilde{w}(0)|/T-S_2}+1\right)\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T} \tag{3.4}$$

in order to remove from the region R all frequencies not within $2/T$ of any ν_k^{λ}. Our procedure will then select a unique frequency within $2/T$ of each ν_k^{λ} if $|H(\nu_k^{\lambda})|>\tau$, so we can set $r=2/T$. In view of (2.4) and (3.3), a sufficient condition for $|H(\nu_k^{\lambda})|>\tau$ is

$$\begin{aligned}&\frac{|\tilde{w}(0)|}{T}|c_k^{\lambda}|-S_2\|c^{\lambda}\|_{\infty}-\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T}\\&\quad>\frac{S_1}{|\tilde{w}(0)|/T-S_2}\left\{\max\left(S_1,\frac{|\tilde{w}(0)|}{T}+\frac{S_1+S_2}{2}\right)\|c^{\lambda}\|_{\infty}+\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T}\right\}\\&\qquad+\left(\frac{S_1}{|\tilde{w}(0)|/T-S_2}+1\right)\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T}\end{aligned}$$

for $k=0,\dots,p$, or equivalently

$$\begin{aligned}\frac{|\tilde{w}(0)|}{T}\min_k|c_k^{\lambda}|>&\left\{S_2+\frac{S_1\max(S_1,\frac{|\tilde{w}(0)|}{T}+\frac{S_1+S_2}{2})}{|\tilde{w}(0)|/T-S_2}\right\}\max_k|c_k^{\lambda}|\\&+2\left(\frac{S_1}{|\tilde{w}(0)|/T-S_2}+1\right)\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^{w}(\nu)|}{T}.\end{aligned} \tag{3.5}$$

[6]The sum to infinity is needed as p is unknown.

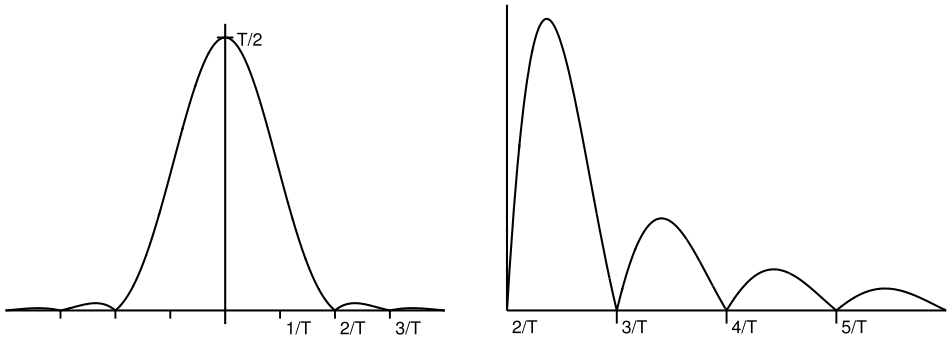


FIG. 3. *Plot of* $|\tilde{w}(\nu)|$ *for the Hann window* (3.6). *Left panel*: *Most of the energy is concentrated in the main lobe between* $\nu = \pm\frac{2}{T}$. *Right panel*: *The side lobes are of width* $1/T$ *and have successively lower peaks*.

It will be shown that the first two terms are dominant. Hence to first order, as the tail sums S_1 and S_2 become small relative to $|\tilde{w}(0)|/T$, a larger margin of separation between signal and leakage noise is attained in frequency domain. Therefore, window functions with rapidly decaying spectral tails are desired. Of the commonly used continuous time windows presented in Table 3.1 of Prabhu (2013) with spectral energy concentrated inside $|\nu| < 2/T$, the time-shifted Hann window has the lightest spectral tails [order $1/(T\nu)^3$]:

$$
\begin{aligned}
w(t) &= \left(\sin^2\frac{\pi t}{T}\right) I_{[0,T]}(t) \leftrightarrow \tilde{w}(\nu) \\
&= \begin{cases} T/2 & \nu = 0, \\ -T/4 & \nu = \pm\dfrac{1}{T}, \\ \dfrac{T}{2} e^{-i\pi T\nu} \dfrac{\operatorname{sinc}(T\nu)}{1-(T\nu)^2} & \text{else}, \end{cases}
\end{aligned} \tag{3.6}
$$

where $\operatorname{sinc}(\nu) = \sin(\pi\nu)/(\pi\nu)$ is the sinc kernel. Note from Figure 3 that $|\tilde{w}(\nu)|$ is symmetric and most of its energy is concentrated inside the main lobe between $\nu = \pm\frac{2}{T}$. The side lobes are of width $1/T$ and have successively lower peaks. The following lemma provides estimates for S_1 and S_2.

LEMMA 1. *For the Hann window*,

$$
0.02843 < S_1 = \frac{2}{T}\sum_{l=0}^{\infty} \sup_{|\nu| \geq \frac{2}{T} + \frac{4}{T}l} \left|\tilde{w}(\nu)\right| < 0.02844,
$$

$$
0.00464 < S_2 = \frac{2}{T}\sum_{l=1}^{\infty} \sup_{|\nu| \geq \frac{4}{T}l} \left|\tilde{w}(\nu)\right| < 0.00465.
$$

Furthermore, if we define $\tilde{w}'(\nu) = \frac{d\tilde{w}}{d\nu}(\nu)$, *then for any* $\nu \in (\nu_k^\lambda - \frac{2}{T}, \nu_k^\lambda + \frac{2}{T})$,

$$\frac{1}{T}\sum_{l\neq k}|\tilde{w}(\nu-\nu_l^\lambda)| < \frac{4}{g(T)^3}, \qquad \frac{1}{T}\sum_{l\neq k}|\tilde{w}'(\nu-\nu_l^\lambda)| < \frac{29T}{g(T)^3}.$$

The remaining quantity not yet examined in (3.4) and (3.5) is the supremum spectral density $\sup_{\nu\in[0,B]}|\tilde{\varepsilon}^w(\nu)|$ of the windowed statistical noise. Noting that $|w(t)| \le 1$ and $|w'(t)| = |\frac{dw}{dt}(t)| \le \pi/T < \infty$, the following lemma shows that the scaled spectral noise level is of order $(\log T/T)^{1/2}$.

LEMMA 2. *Define* $\bar{\Lambda}_T = \Lambda(T)/T$ *and* $\bar{N}_T = N(T)/T$, *and suppose that* $\sup_{t\in[0,T]}|w(t)| \le 1$, $\sup_{t\in(0,T)}|w'(t)| < \infty$. *Then for any* $\beta > 0$, $\gamma > 1$, *and* $\alpha \ge \gamma/(\gamma-1)$, *with probability*

$$1 - 8\gamma\pi B\big[1/T^{(\frac{\gamma-1}{\gamma}\alpha)^2-1} + T\exp\{-(\Lambda(T)\log T)^{1/2}\}\big] - 2e^{-\Lambda(T)\beta^2/4}$$

we have

$$(1-\beta)\bar{\Lambda}_T < \bar{N}_T < (1+\beta)\bar{\Lambda}_T$$

and

$$\sup_{\nu\in[0,B]}\frac{|\tilde{\varepsilon}^w(\nu)|}{T} < 4\alpha\bar{\Lambda}_T^{1/2}\left(\frac{\log T}{T}\right)^{1/2} < \frac{4\alpha\bar{N}_T^{1/2}}{(1-\beta)^{1/2}}\left(\frac{\log T}{T}\right)^{1/2}.$$

Lemmas 1 and 2 can be used in (3.4) to define the data-driven threshold

$$\tau = 0.0574\sup_{\nu\in[0,B]}|H(\nu)| + \frac{4.23\alpha\bar{N}_T^{1/2}}{(1-\beta)^{1/2}}\left(\frac{\log T}{T}\right)^{1/2} \tag{3.7}$$

for the Hann window. Deriving the sufficient condition for frequency recovery (3.5) for this τ and the Hann window yields:

A2. There exist $\beta > 0$, $\gamma > 1$, and $\alpha \ge \gamma/(\gamma-1)$ such that

$$\min_k|c_k^\lambda| > 0.0686\max_k|c_k^\lambda| + 16.9\alpha\left\{1+\left(\frac{1+\beta}{1-\beta}\right)^{1/2}\right\}\bar{\Lambda}_T^{1/2}\left(\frac{\log T}{T}\right)^{1/2}.$$

As T grows the last term in A2 vanishes, so to first order the condition $\min_k|c_k^\lambda| > 0.0686\max_k|c_k^\lambda|$ requires the dynamic range of the amplitudes to be less than 14.5. The smaller the tailsums S_1 and S_2 are, the larger the allowable range. In particular, if the gap in A1 is slightly relaxed from $4/T$ to $6/T$, the value of 14.5 can be increased to over 100 by replacing the Hann window with the lighter spectral-tailed $\cos^4$ window (Prabhu (2013)). Thus windows with light spectral tails provide a solution for detecting weak frequency signals in the presence of strong ones. This addresses a point mentioned in passing on page 110 of Shao and Lii (2011): Issues with the periodogram method arise when the dynamic range is large, even in

the classical setting where the frequency gap is $1/o(T)$. Our analysis provides a way for quantifying this for both the windowed and unwindowed periodograms when T is finite. In the special case where all the frequencies have the same amplitude $|c_1^\lambda| = \cdots = |c_p^\lambda|$, A2 simplifies to requiring the amplitude to be larger than a multiple of the statistical noise level (last term of A2).

The main frequency recovery result can now be stated under A1 and A2.

PROPOSITION 3. *Let $w(t)$ in Algorithm 1 be the Hann window (3.6), and set $r = 2/T$ and τ as (3.7). Then with probability at least*

$$1 - 8\gamma\pi B\big[1/T^{(\frac{\gamma-1}{\gamma}\alpha)^2-1} + T\exp\{-(\Lambda(T)\log T)^{1/2}\}\big] - 2e^{-\Lambda(T)\beta^2/4}$$

our procedure will select exactly $p+1$ frequencies $\hat{\nu} = \{\hat{\nu}_k\}_k$ with precision $\|\nu^\lambda - \hat{\nu}\|_\infty < 2/T$. Furthermore,

$$\big\|\nu^\lambda - \hat{\nu}\big\|_\infty < \min\left\{\frac{2}{T}, \frac{2\epsilon(T)}{T}\right\}$$

if

$$\epsilon(T) \triangleq \frac{348(\|c^\lambda\|_\infty + \alpha\bar{\Lambda}_T^{1/2})}{\min_k |c_k^\lambda|} \max\left\{\frac{1}{g(T)^3}, \left(\frac{\log T}{T}\right)^{1/2}\right\} \le \frac{87}{40}.$$

REMARK. Through the use of windowing, we obtain the first periodogram peak-hunting method that is able to achieve super-resolution. Note from the definition of $\epsilon(T)$ that if $g(T) \to \infty$ then the procedure will recover all frequencies with precision $o(1/T)$. In particular, if $g(T)$ is $\mathcal{O}(T^{1/6})$ or greater then the estimation error is $\mathcal{O}(T^{-3/2})$ up to a log factor. For the unwindowed periodogram in the closely related time series setting, Theorem 6.8b of Li (2014) shows that the same rate is achieved when $g(T)$ is greater than $\mathcal{O}(T^{1/2})$. This is because $T^{3/2}(\nu^\lambda - \hat{\nu})$ has a bias of $\mathcal{O}(T^{1/2}/g(T))$ due to the slower spectral tail decay of the rectangle window [Remark 6.14 of Li (2014)]. Thus even under the classical resolution setting, windowing is still beneficial since it sharpens the precision of the frequency estimates.

REMARK. In applications, α, β and γ are chosen to balance a number of considerations. First is the expected dynamic range (A2) for the particular problem being considered. Second is the bandwidth B: If A2 holds for values of α, γ satisfying $\alpha(\gamma-1)/\gamma \ge \sqrt{2}$, then the probability bound above is $1 - 8\gamma\pi B/T - 2e^{-\Lambda(T)\beta^2/4}$ to leading order in B/T, in which case B has the same asymptotic scaling as Shao (2010), Shao and Lii (2011), Vere-Jones (1982). Third is the desired recovery probability. One possible choice that balances these considerations is $\alpha = 2$, $\beta = 2\sqrt{\log T/T}$, and $\gamma = 4$, which simplifies the probability bound to

$$1 - 32\pi B\big[T^{-5/4} + T\exp\{-(\Lambda(T)\log T)^{1/2}\}\big] - 2T^{-\bar{\Lambda}_T}.$$

REMARK. When p is known, no thresholding is necessary, and the asymptotic normality results in Shao and Lii (2011) for the classic periodogram can be extended to the windowed one. The details are provided in the Supplementary Material.

When does the approach of Shao (2010), Shao and Lii (2011) perform better? If $\{\nu_k^\lambda\}_k$ are in fact spaced more than order $1/T$ apart from one another, then it follows from (2.5) that the leakage outside a $\mathcal{O}(g(T)/T)$-neighbourhood of the frequencies is of order $1/g(T)^3 \to 0$ for the Hann-windowed periodogram. Hence the threshold (3.7) is conservative in this setting. While it will still work within the dynamic range implied by A2, we expect the method in Shao (2010), Shao and Lii (2011), which was specifically designed for the classical resolution setting, to recover more of the frequencies with amplitude less than $1/14.5$ of the largest one. Of course, the Hann window analyzed here can also be used with the method in Shao (2010), Shao and Lii (2011).

Connection to super-resolution literature. There are clear connections between our results and those arising from the work on super-resolution recovery of discrete time signals (Bhaskar, Tang and Recht (2013), Candès and Fernandez-Granda (2013), Fernandez-Granda (2013), Tang, Bhaskar and Recht (2015)). In that setting, the authors assume a discrete time signal $x = \sum_{k=1}^p c_k^\lambda e^{2\pi i \nu_k t} \in \mathbb{R}^n$ and the observations are of the form $y = x + e$ where $e \in \mathbb{R}^n$ is a noise vector.

For a bounded e, Candès and Fernandez-Granda (2013), Fernandez-Granda (2013) establish signal and support recovery guarantees for their semidefinite programming approach. On the other hand for $e_i \sim N(0, \sigma^2)$, the related AST approach (Bhaskar, Tang and Recht (2013), Tang, Bhaskar and Recht (2015)) achieves near minimax rates. Furthermore, if $\min_k |c_k^\lambda|$ is larger than some multiple of $\sigma p(\log n/n)^{1/2}$, then with high probability AST is guaranteed to pick out one or more frequencies within some C/n of each signal frequency. The authors conjecture that it is possible to prevent the selection of spurious frequencies, and that the sparsity p can be dropped from the lower bound on $\min_k |c_k^\lambda|$. The following corollary shows that our procedure resolves these conjectures in the affirmative when applied to this setting.

COROLLARY 4. *Suppose T is replaced by n and $\bar{\Lambda}_T$ is replaced by σ^2 in* A2. *Under the discrete time setting above, with high probability our procedure will select exactly p frequencies within distance $\|\widehat{\nu} - \nu^\lambda\|_\infty \le 4/n$ of the true ones.*

Modified threshold. The last term in the threshold (3.7) comes from the spectral noise bound in Lemma 2, whose constant $4\bar{N}_T^{1/2}$ may be conservative. As a result, we observe in experiments that a large value of T is sometimes needed for the guarantees to hold with high probability. To obtain a tighter estimate, one idea is to approximate the spectral noise level of the underlying nonhomogeneous Poisson

process with that of a homogeneous one. This is motivated by the fact that the noise bound in Lemma 2 depends on $\lambda(t)$ only through the average rate $\bar{N}_T$, regardless of whether the Poisson process is homogeneous or not. Thus for a given $\xi > 0$, consider the modified threshold

$$\tau_\xi = (0.0574 + \xi) \sup_{\nu \in [0,B]} |H(\nu)| + 1.06 \min\left\{ \hat{\chi}_T, \frac{4\alpha \bar{N}_T^{1/2}}{(1-\beta)^{1/2}} \left(\frac{\log T}{T} \right)^{1/2} \right\}, \tag{3.8}$$

where $\hat{\chi}_T$ is the simulated $\sup_{\nu \in [0,B]} |\tilde{\varepsilon}^w(\nu)|/T$ for the homogeneous Poisson process with rate $\bar{N}_T$ over $[0, T]$. It is equivalent to applying the expression (5.2) to simulated data. Clearly, if the second quantity in the curly bracket is smaller then we effectively recover (3.7). In experiments, we find that thresholding with τ_ξ performs better than τ in practice. The following corollary provides a large sample recovery guarantee for τ_ξ.

COROLLARY 5. *Suppose* A2 *is slightly strengthened to* $\min_k |c_k^\lambda| > (0.0686 + 4\xi) \max_k |c_k^\lambda|$, *and* T *is large enough that* $\alpha(\frac{1+\beta}{1-\beta} \cdot \frac{\bar{\Lambda}_T \log T}{T})^{1/2} \leq \frac{30\xi}{28+25\xi} \|c^\lambda\|_\infty$. *Then with probability at least*

$$1 - 8\gamma\pi B\big[1/T^{(\frac{\gamma-1}{\gamma}\alpha)^2-1} + T\exp\big\{-\big(\Lambda(T)\log T\big)^{1/2}\big\}\big] - 2e^{-\Lambda(T)\beta^2/4}$$

all frequencies will be recovered with the precision stated in Proposition 3 when we threshold with τ_ξ.

REMARK. If the second condition is to ever hold, ξ must then be at least of order $\|c^\lambda\|_\infty^{-1} \sqrt{\bar{\Lambda}_T \log T / T}$.

4. Amplitude and phase estimation. As noted by Rice and Rosenblatt (1988) for the case of cyclic time series and Shao (2010), Shao and Lii (2011) for the case of cyclic Poisson processes, it is necessary for the estimated frequencies $\hat{\nu}$ to be within $o(1/T)$ of ν^λ if we wish to estimate the coefficients c^λ consistently. We will therefore let $g(T) \to \infty$ in Proposition 3 so that $\epsilon(T) \to 0$. Our estimator is the complex-valued least squares solution to (2.2) in the limit $dt \to 0$:

$$\hat{c} = \hat{\Gamma}^{-1} y, \tag{4.1}$$

where the jth entry of the $(p+1)$-vector y is $\frac{1}{T}\int_0^T e^{-2\pi i \hat{\nu}_j t}\, dN(t)$, and the (j,k)-entry of the $(p+1) \times (p+1)$ matrix $\hat{\Gamma}$ is

$$\hat{\Gamma}_{jk} = \frac{1}{T}\int_0^T e^{-2\pi i(\hat{\nu}_j - \hat{\nu}_k)t}\, dt = \frac{1}{T}\tilde{I}_{(0,T]}(\hat{\nu}_j - \hat{\nu}_k),$$

where $\tilde{I}_{(0,T]}(\nu) = Te^{-i\pi T\nu}\mathrm{sinc}(T\nu)$ is the Fourier transform of the rectangle. Since $\{\hat{\nu}_k\}_k$ are symmetric about zero, it can be shown for $\hat{\nu}_k = -\hat{\nu}_l$ that $\hat{c}_k$ and $\hat{c}_l$ are conjugate pairs, hence the estimator for $\lambda(t)$ is always real-valued. We note

that the corresponding estimator in Shao (2010), Shao and Lii (2011) can be recovered by setting $\hat{\Gamma}$ to the identity matrix, which is asymptotically valid because $\hat{\Gamma}$ converges to an orthonormal design as $g(T) \to \infty$. Our choice of $\hat{\Gamma}$ provides a second-order correction when T is finite.

PROPOSITION 6. *Suppose the conditions for Proposition 3 hold with* $\epsilon(T) \leq 87/40$, *and that*

$$\Gamma_{jk} = \frac{1}{T}\int_0^T e^{-2\pi i(\nu_j^\lambda - \nu_k^\lambda)t}\,dt = \frac{1}{T}\tilde{I}_{(0,T]}(\nu_j^\lambda - \nu_k^\lambda)$$

is invertible. Then with probability at least

$$1 - 8\gamma\pi B\big[1/T^{(\frac{\gamma-1}{\gamma}\alpha)^2-1} + T\exp\{-(\Lambda(T)\log T)^{1/2}\}\big] - 2e^{-\Lambda(T)\beta^2/4}$$

(i) $\hat{\Gamma}$ *is also invertible for sufficiently large* T; *and* (ii)

$$\|\hat{c} - c^\lambda\|_\infty < 2\{(\pi + 2\alpha)\|\hat{\Gamma}^{-1}\|\max(\|c^\lambda\|_1, 1)\}\epsilon(T).$$

5. Numerical examples. We use simulations to compare our thresholding procedure (based on the modified threshold) to the windowed periodogram combined with BIC model selection, and also to the classic periodogram in Shao (2010), Shao and Lii (2011) combined with BIC. We also use our procedure to analyze arrivals data from an academic emergency department in the United States. We focus on the BIC because it is asymptotically consistent, and the corresponding penalized log-likelihood for Poisson processes is derived in Section 3.3.4 of Shao (2010):

$$-2\left(\sum_{j=1}^{N(T)}\log\lambda(t_j) - \Lambda(T)\right) + (5p+1)\log T. \tag{5.1}$$

The algorithm for using the windowed periodogram with BIC selection corresponds to setting $R = \{\nu : r \leq |\nu| \leq B\}$ in Algorithm 1 and running step 3 until p frequencies have been selected. The value of p is chosen to minimize (5.1).

Since by default the frequency $\nu = 0$ is always selected, we work with the centralized version of $|H(\nu)|$ instead:

$$\begin{aligned}|H_c(\nu)| &= \frac{1}{T}\left|\int_0^T e^{-2\pi i\nu t}w(t)\left(dN(t) - \frac{N(T)}{T}\,dt\right)\right| \\ &= \frac{1}{T}\left|H(\nu) - \frac{N(T)}{T}\tilde{w}(\nu)\right|,\end{aligned} \tag{5.2}$$

which is one way to generalize the centralized unwindowed periodogram given by equation (4) in Shao and Lii (2011). This approximately removes from the windowed periodogram the peak at the origin.

The asymptotic analysis in Shao (2010), Shao and Lii (2011) recommends a minimum exclusion radius[7] of $r = 3/T$, which corresponds to assuming a frequency gap of at least $6/T$ in the finite T setting. Hence, in order to compare the methods, we also set $r = 3/T$ in Algorithm 1 and assume that $g(T) \geq 6$. The modified threshold τ_ξ corresponding to (3.8) is then

$$(5.3) \qquad (0.0180 + \xi) \sup_{\nu \in [0,B]} |H_c(\nu)| + 1.02 \min\left\{ \hat{\chi}_T, \frac{4\alpha \bar{N}_T^{1/2}}{(1-\beta)^{1/2}} \left(\frac{\log T}{T} \right)^{1/2} \right\},$$

where we choose $\xi = 0.0001$ to be small. In all our analyses, $\hat{\chi}_T$ turns out to be always smaller than the lower bound $4(\bar{N}_T \log T/T)^{1/2}$ for the second quantity in the curly bracket. When $g(T) \geq 6$, the maximum allowable dynamic range widens to 47 under the Hann window.

5.1. *Frequency recovery error rate.* We use the following simulation to empirically study the error rates for frequency recovery in Proposition 3, which shows that when $g(T)$ is constant the error $\|\hat{\nu} - \nu^\lambda\|_\infty$ is no greater than $\mathcal{O}(1/T)$. As remarked after the proposition, the error rate becomes $\mathcal{O}(T^{-3/2})$ for $g(T)$ equal to or greater than $\mathcal{O}(T^{1/6})$. In the closely related time series setting, the unwindowed periodogram achieves the same rate when $g(T)$ is greater than $\mathcal{O}(T^{1/2})$ [Theorem 6.8b of Li (2014)]. We will therefore examine $\|\hat{\nu} - \nu^\lambda\|_\infty$ as a function of T at the frequency resolutions corresponding to $g(T) \in \{6, T^{1/6}, T^{1/2}\}$. Consider the following class of arrival rates:

$$(5.4) \qquad \lambda(t) = 7.5 + \sum_{k=1}^{5} a_k \cos\left(2\pi \left(0.1 + (k-1)\frac{g(T)}{T} \right) t + \phi_k \right)$$

whose frequencies are spaced apart by $g(T)/T$. The amplitudes a_k are drawn randomly from $U[1, 1.5]$, and the phases ϕ_k from $U[0, 2\pi)$. For each combination of T and $g(T)$, we sample 100 sets of the amplitudes and phases, and then use each set to simulate the corresponding arrival process up to time T.

For the values of T considered, all frequencies are detected by the three methods. Hence BIC selection and thresholding produce the same results when applied to the windowed periodogram. Figure 2 plots on a log–log scale the error $\|\hat{\nu} - \nu^\lambda\|_\infty$ averaged across the 100 simulations for each combination of T and $g(T)$. The slopes of the fitted lines estimate the error rate. For this example, the windowed periodogram performs even better than what the theory predicts, achieving an error rate of almost $\mathcal{O}(T^{-3/2})$ even for $g(T) = 6$. All methods attain this rate when $g(T) = T^{1/2}$.

[7] In terms of angular frequency, this corresponds to the diameter of $12\pi/T$ in Shao and Lii (2011).

TABLE 1

Results for the sawtooth intensity (5.5). Column UBIC is the unwindowed periodogram combined with BIC selection, WBIC is the windowed periodogram with BIC selection, and WThres is the windowed periodogram with thresholding. Averages over 100 simulations are reported (standard errors in parentheses)

	UBIC	**WBIC**	**WThres**
MSE	0.17 (0.03)	0.19 (0.03)	0.17 (0.03)
#correct frequencies	3.05 (0.86)	2.77 (0.74)	4.41 (1.32)
#spurious frequencies	0.01 (0.10)	0.00 (0.00)	0.36 (0.77)

5.2. *Misspecified arrival rate in the classical resolution setting.* The sawtooth wave in Shao and Lii (2011) provides a nice example for testing the robustness of the methods to misspecifications to (1.1). Consider the arrival rate

$$\lambda(t) = 0.1 + 0.5\mathrm{mod}(t, 2\pi) = 0.1 + 0.5\pi - \sum_{k=1}^{\infty} \frac{\sin(2\pi(\frac{k}{2\pi})t)}{k}, \tag{5.5}$$

which has an infinite number of Fourier series frequencies spaced $1/(2\pi)$ apart. We simulate 100 realizations of the arrival process up to time $T = 1000$, which is well within the classical setting where the frequencies are spaced $1/o(T)$ apart. To assess the accuracies of the three methods at estimating $\lambda(t)$, we use the average of the MSE $\frac{1}{T}\int_0^T \{\lambda(t) - \hat{\lambda}(t)\}^2\,dt$ across the 100 samples as the performance metric. We also report the average number of correct and spurious frequencies[8] recovered by each method in Table 1.

Per the discussion in Section 3 regarding when the classic periodogram method should outperform our approach, (5.5) fits the bill since the frequencies are separated by much more than order $1/T$. Interestingly, the differences in performance among the methods are not statistically significant for this example.

5.3. *A super-resolution example with varying dynamic range.* The following arrival rate is inspired by Professor E. H. Kaplan's analysis of arrivals data to a psychiatric ward, where the existence of a lunar and a monthly cycle are verified:

$$\lambda(t) = (2r + 2) + 2r\cos\left(\frac{2\pi}{30}t + 2.6\right) + 2\cos\left(\frac{2\pi}{28}t + 4.5\right). \tag{5.6}$$

The two frequencies at $1/28$ and $1/30$ are separated by a gap that is slightly larger than $6/T$ when $T = 3000$. The monthly cycle is r times stronger than the lunar one, meaning that leakage from the former can easily mask the latter when r is large. The left panels of Figure 4 display the centralized windowed periodograms

[8] A correct recovery is defined as one that is within $3/T$ of one of the Fourier frequencies.

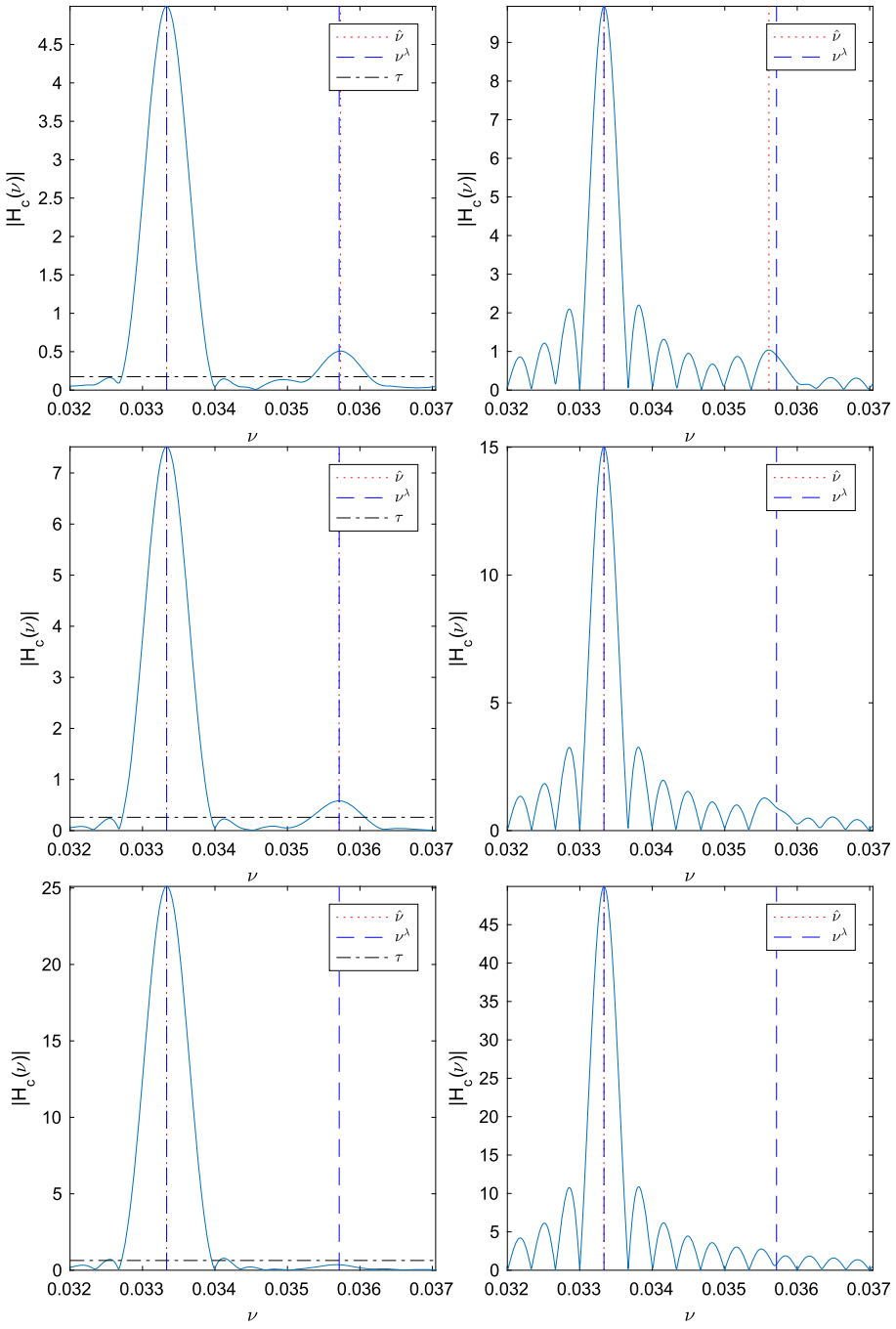


FIG. 4. *Results for* (5.6). Left panels: *The centralized Hann-windowed periodograms. The threshold is represented by the horizontal line, and the locations of the frequencies and their estimates are given by the vertical ones*. Right panels: *The unwindowed centralized periodograms. Top row*: $r = 10$, *middle row*: $r = 15$, *bottom row*: $r = 50$.

for different values of the dynamic range r, and the right panels display the corresponding unwindowed periodograms. Here, we apply thresholding to the windowed periodograms; BIC selection performs similarly.

For $r = 10$ (top row), both periodograms are able to resolve the two frequencies. For $r = 15$ (middle row), only the windowed periodogram is able to detect the weaker lunar cycle. Both methods fail to identify the lunar cycle when $r = 50$ (bottom row), although the windowed periodogram is still able to do so for $r = 45$ (not shown). This illustrates the role of windowing in suppressing leakage, thereby allowing for super-resolution frequency recovery. Moreover, our findings match the calculations at the beginning of this section that show the Hann-windowed periodogram has a maximum allowable dynamic range of 47 when $g(T) \geq 6$. If there are actually more frequencies in (5.6) that are $\mathcal{O}(1/T)$ away from the lunar cycle, then the leakage around $1/28$ in the classic periodogram will be of order $\log p$ as explained in Section 2. In such cases, the classic periodogram may not be able to detect the lunar cycle even if the dynamic range is 1.

5.4. *Patient arrivals to an emergency department.* Our last example analyzes arrivals data from the emergency department of an academic hospital in the United States. We focus in particular on the arrivals of 66,240 mid-acuity level[9] patients from 2014 to Q3 of 2015 ($T = 652$ days).

As shown in the left panel of Figure 5, three intraday frequencies and five week-based ones are selected from the centralized periodogram. The intraday frequencies include a daily cycle ($\hat{\nu}_1 = 1.00$), a 12-hour cycle ($\hat{\nu}_2 = 2.00$), and an 8-hour cycle ($\hat{\nu}_3 = 3.00$). The week-based ones include a weekly cycle ($\hat{\nu}_4 = 0.142$), a half week cycle ($\hat{\nu}_5 = 0.286$), a $1/5$ week cycle ($\hat{\nu}_6 = 0.714$), a $1/6$ week cycle ($\hat{\nu}_7 = 0.857$) and a $1/8$ week cycle ($\hat{\nu}_8 = 1.143$). Given that the fitted rate has a weekly period, we can compare it to the average arrival rate for each of the 168 hours of the week (right panel of Figure 5). Overall, we see that using 8 frequencies to model the arrival rate does almost as well as using 168 piecewise constant hourly fits. Moreover, the sinusoidal estimate reveals two intraday peaks, the first at around 11 a.m. and the second at around 5 p.m. We also see that the intensity of arrivals fade steadily into the weekend.

6. Discussion. By a novel use of windowing, this paper shows that simple periodogram methods can in fact achieve super-resolution frequency recovery for cyclic arrival rates. This improves the resolution of classic periodograms, while being much faster to compute than the SDP approach in super-resolution literature. Under mild assumptions on the dynamic range of the frequency amplitudes, our approach guarantees that no spurious frequencies will be recovered. To establish the consistency of the coefficient estimates, our finite sample results show that if

[9]Defined as Emergency Severity Index (ESI) level 2.

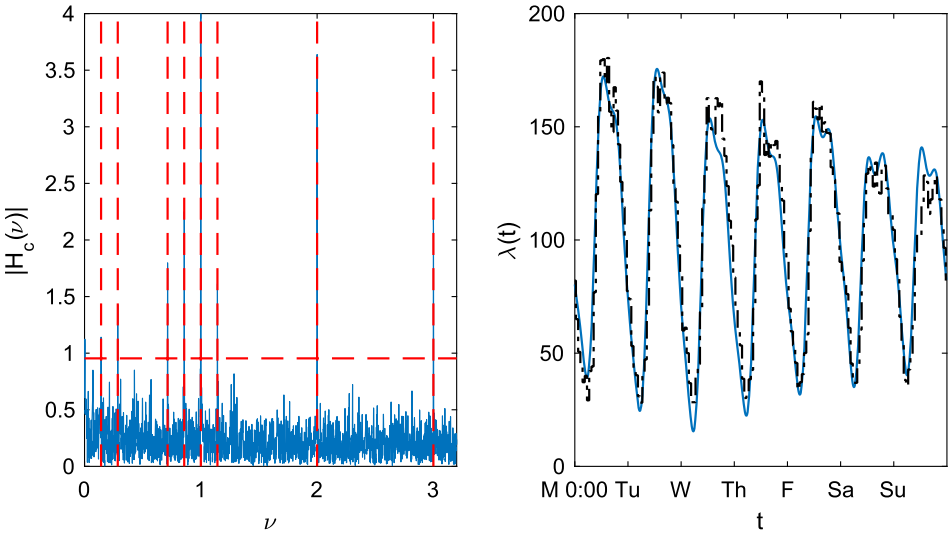


FIG. 5. *ESI level* 2 *arrivals*. Left panel: *The centralized windowed periodogram. The selected threshold is represented by the dashed horizontal line, and the location of the frequency estimates are given by the vertical ones*. Right panel: *The estimated arrival rate* (*arrivals per day*) *over the course of a week is given by the solid line. The dash-dot line represents the empirical average arrival rate for each hour of the week.*

the frequency gap is $1/o(T)$, then the frequencies can be recovered with precision $o(1/T)$ as required. Whether the gap can be relaxed to order $1/T$ is a question that is left for future research.

Another area for future research is to extend the cyclic specification (1.1) to allow for higher order noncyclical components as well. One approach is to add wavelets to the basis of complex exponentials. It might then be possible to leverage the rate-optimal procedure in Brown et al. (2010) to estimate the time-localized components of the arrival rate.

Acknowledgments. The review team provided many insightful comments that significantly improved the paper. Special thanks to Ed Kaplan and Don Green for stimulating discussions on spectral analysis. The emergency department arrivals data was kindly provided by Dr. Kito Lord.

SUPPLEMENTARY MATERIAL

Proofs and asymptotic normality (DOI: 10.1214/18-AOS1736SUPP; .pdf). The proofs of all results presented in this paper are provided in Appendix A of the supplement. Appendix B establishes the asymptotic normality of the windowed periodogram estimator.

REFERENCES

BARTLETT, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25** 264–296. MR0171334

BEBBINGTON, M. and ZITIKIS, R. (2004). A robust heuristic estimator for the period of a Poisson intensity function. *Methodol. Comput. Appl. Probab.* **6** 441–462. MR2108562

BELITSER, E., SERRA, P. and VAN ZANTEN, H. (2013). Estimating the period of a cyclic non-homogeneous Poisson process. *Scand. J. Stat.* **40** 204–218. MR3066411

BHASKAR, B. N., TANG, G. and RECHT, B. (2013). Atomic norm denoising with applications to line spectral estimation. *IEEE Trans. Signal Process.* **61** 5987–5999. MR3138795

BROWN, L., CAI, T., ZHANG, R., ZHAO, L. and ZHOU, H. (2010). The root-unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probab. Theory Related Fields* **146** 401–433. MR2574733

CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2013). Super-resolution from noisy data. *J. Fourier Anal. Appl.* **19** 1229–1254. MR3132912

CHEN, N., LEE, D. K. K. and NEGAHBAN, S. N. (2019). Supplement to "Super-resolution estimation of cyclic arrival rates." DOI:10.1214/18-AOS1736SUPP.

CHEN, N., LEE, D. K. K. and SHEN, H. (2018). Can customer arrival rates be modelled by sine waves? Working paper.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

DUTT, A. and ROKHLIN, V. (1993). Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.* **14** 1368–1393. MR1241591

FERNANDEZ-GRANDA, C. (2013). Support detection in super-resolution. In *Proceedings of the* 10*th International Conference on Sampling Theory and Applications* 145–148.

HELMERS, R. and MANGKU, I. W. (2003). On estimating the period of a cyclic Poisson process. In *Mathematical Statistics and Applications*: *Festschrift for Constance van Eeden*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **42** 345–356. IMS, Beachwood, OH. MR2138302

LEWIS, P. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. *J. Sound Vibration* **12** 353–375.

LI, T.-H. (2014). *Time Series with Mixed Spectra*. CRC Press, Boca Raton, FL. MR3098581

MOITRA, A. (2015). Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *STOC'*15*—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 821–830. ACM, New York. MR3388262

OSIPOV, A., ROKHLIN, V. and XIAO, H. (2013). *Prolate Spheroidal Wave Functions of Order Zero*: *Mathematical Tools for Bandlimited Approximation*. *Applied Mathematical Sciences* **187**. Springer, New York. MR3136430

PRABHU, K. (2013). *Window Functions and Their Applications in Signal Processing*. CRC Press, New York.

RICE, J. A. and ROSENBLATT, M. (1988). On frequency estimation. *Biometrika* **75** 477–484. MR0967586

SHAO, N. (2010). Modeling almost periodicity in point processes. Ph.D. thesis, Univ. California, Riverside.

SHAO, N. and LII, K.-S. (2011). Modelling non-homogeneous Poisson processes with almost periodic intensity functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 99–122. MR2797738

TANG, G., BHASKAR, B. N. and RECHT, B. (2015). Near minimax line spectral estimation. *IEEE Trans. Inform. Theory* **61** 499–512. MR3299978

VERE-JONES, D. (1982). On the estimation of frequency in point-process data. *J. Appl. Probab.* **19A** 383–394. MR0633207

N. CHEN
INDUSTRIAL ENGINEERING & DECISION ANALYTICS
HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY
CLEAR WATER BAY, KOWLOON
HONG KONG
E-MAIL: nychen@ust.hk

D. K. K. LEE
SCHOOL OF MANAGEMENT
AND
DEPARTMENT OF STATISTICS & DATA SCIENCE
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: donald.lee@yale.edu

S. N. NEGAHBAN
DEPARTMENT OF STATISTICS & DATA SCIENCE
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06510
USA
E-MAIL: sahand.negahban@yale.edu