SUPER-RESOLUTION ESTIMATION OF CYCLIC ARRIVAL RATES

BY NINGYUAN CHEN†, DONALD K.K. LEE*, SAHAND N. NEGAHBAN*

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 and Lii [2010,2011], 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 et al., 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

$$EN(t) = \int_0^t \lambda(u)du = \Lambda(t).$$

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

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

where the even number $p$ of frequency components, the frequencies $\nu_k^\lambda = \{\nu_k^\lambda\}_k$ in a pre-specified band $[-B,+B]$, and the complex coefficients $c_k^\lambda = \{c_k^\lambda\}_k$.
\{c_k^\lambda\}_k are all 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 \(\tau\) above the noise level will result in the algorithm isolating a neighbourhood around each \(\nu_k^\lambda\) (see Figure 1.1). It will be shown that if the frequencies are separated from one another by a gap (resolution) of at least \(g(T)/T\) where \(g(T)\) ≥ 4, then with high probability our procedure will recover each \(\nu_k^\lambda\) 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 if \(g(T)\) ≥ 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 [1963]’s classic ‘unwindowed’ periodogram for point processes is essentially \(\left| \sum_j e^{-2\pi i \nu_j} \right| /T\), 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 1.2 presents a log-log
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 \(|\nu| \leq B\), and note that it is symmetric in \(\nu\). The sum can be computed efficiently using non-uniform FFT algorithms like Dutt and Rokhlin [1993].

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

3 Set \(\nu_0^\lambda = \nu_0 = 0\), \(k = 1\) 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 \(R \leftarrow R \setminus (\hat{\nu}_k - r, \hat{\nu}_k + r)\). This removes a neighbourhood of radius \(r\) centred at \(\hat{\nu}_k\) from \(R\).

4 Compute the estimator (4.1) for \(c_\lambda\).

---

Fig 1.1: Visualization of Algorithm 1. In the depicted periodogram there are two signal frequencies \(\nu_1^\lambda\) and \(\nu_2^\lambda\). Setting \(\tau\) (horizontal line) above the ambient noise results in the algorithm selecting neighbourhoods (between the pairs of vertical lines) that contain \(\nu_1^\lambda\) and \(\nu_2^\lambda\).
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 $O(T^{-3/2})$ as predicted by theory.

Fig 1.2: 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.

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 \( w(t), \tau, \) 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 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.

**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 corresponding to the largest peak of the periodogram is selected. Other approaches [Bebbington and Žitikis, 2004, Belitser et al., 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 $\ell_1$-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\(^1\). Second, the required computational effort is overly taxing\(^2\) for the size of problems we consider. For example, Chen et al. [2018] analyzes 652 days of arrivals data from an emergency department and used 5,216 bins of 3 hour widths for the Lasso extension. The ADMM implementation recommended in Bhaskar et al. [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 et al. [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 et al. [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 $\{e^{2\pi i\nu t}\}_{|\nu| \leq B}$ be our dictionary for constructing an arrival rate. Suppose the rate for the underlying NHPP is (1.1), which belongs

---

\(^1\)This can however be overcome using bins narrower than $1/(2B)$ (Nyquist sampling).

\(^2\)A Lasso approximation obtained from discretizing the frequency domain is suggested in Bhaskar et al. [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.
in the collection
\[
\left\{ c_0 + \sum_{k=1}^{p} c_k e^{2\pi i \nu_k t} : c_k \in \mathbb{C}, p < \infty \right\}.
\]

Since \(\lambda(t)\) is real-valued, (1.1) will lie in the subset where the presence of \((c_k, \nu_k)\) 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 \(\nu^\lambda\) of frequencies in (1.1), where \(p\) is even but unknown. Given these, the coefficients \(c^\lambda\) 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 \(\nu^\lambda\) 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
\[
\{dN(t)\}_{t \in [0, T]} = [d\Delta(t) + d\{N(t) - \Lambda(t)\}]I_{(0,T]}(t) = \lambda(t)I_{(0,T]}(t)dt + d\varepsilon(t)I_{(0,T]}(t),
\]
where \(I_{(0,T]}(t)\) is the indicator function of \(\{0 < t \leq T\}\). 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 \(\{\nu^\lambda_k\}_k\):
\[
\tilde{\lambda}(\nu) = \sum_{k=0}^{p} c^\lambda_k \delta(\nu - \nu^\lambda_k).
\]

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 \(\ast\) by \(f \ast h(t) = \int f(s)h(t-s)ds\), the \(h\)-smoothed average of \(f\) about the point \(t\). The spectrum of \(\lambda(t)I_{(0,T]}(t)\) is
\[
(\lambda \cdot I_{(0,T]})(\nu) = \left(\tilde{\lambda} \ast \tilde{I}_{(0,T]}\right)(\nu) = \sum_{k=0}^{p} c^\lambda_k \tilde{I}_{(0,T]}(\nu - \nu^\lambda_k),
\]
a weighted average of \(\lambda(t)\)'s spectral values \(c^\lambda_k\) concentrated at \(\{\nu^\lambda_k\}_k\). Thus truncation has the effect of smearing the frequency spikes in \(\tilde{\lambda}(\nu)\) into a continuous spectrum: For \(\nu \notin \cup_k \{\nu^\lambda_k\}\), \(\lambda \cdot I_{(0,T]}(\nu)\) can have a non-zero 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 $\nu^\lambda$ from $\lambda(t)I_{(0,T]}(t)$. Leakage distortion is a manifestation of the uncertainty principle because perfect frequency localization requires $\hat{I}_{(0,T]}(\nu) = \delta(\nu)$, but this is only possible if $I_{(0,T]}(t) = 1$, i.e. 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 $\hat{I}_{(0,T]}(\nu)$, as this dictates the influence that distant frequencies have 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 non-uniform 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 $\hat{I}_{(0,T]}(\nu)$ is of order $1/(TV)$ for $\nu$sufficiently far from $\{\nu^\lambda_k\}_{k}$. Thus if $\{\nu^\lambda_k\}_{k}$ are spaced $1/T$ apart, the leakage (2.3) around a neighbourhood of $\nu^\lambda_k$ from the other frequencies can be of order $\log p$ for the rectangle window. This can easily mask the periodogram spike at $\nu^\lambda_k$ 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 is actually suboptimal for frequency recovery: Theorem 3.44 of Osipov et al. [2013] shows that the spectral tail decay of the prolate spheroidal function is also of order $1/(TV)$ when it is time-limited to $[0,T]$.

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

$$
H(\nu) = \frac{1}{T} \int_{0}^{T} e^{-2\pi i \nu t} dN^w(t)
= \frac{1}{T} \int_{0}^{T} e^{-2\pi i \nu t} w(t)\lambda(t)dt + \frac{1}{T} \int_{0}^{T} e^{-2\pi i \nu t} w(t)d\varepsilon(t)
= \frac{1}{T} \sum_{k=0}^{p} c_k^\lambda \hat{w}(\nu - \nu^\lambda_k) + \frac{\hat{\varepsilon}^w(\nu)}{T}.
$$

(2.4)

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

$$\begin{align*}
|H(\nu)| & \leq \|c^\lambda\|_\infty \sum_{k=0}^{p} \left| \frac{\tilde{w}(\nu - \nu_k^\lambda)}{T} \right| + \sup_{\nu \in [0, B]} \left| \frac{\tilde{\varepsilon}_w(\nu)}{T} \right|.
\end{align*}$$

If the signal strengths $c^\lambda$ are sufficiently strong, then intuitively a neighbourhood of $\cup_k \{\nu_k^\lambda\}$ can be isolated by simply excluding frequency regions in $[-B, +B]$ where $|H(\nu)|$ is below some threshold $\tau$ (see Figure 1.1). This is the idea behind step 2 of Algorithm 1. The analysis presented in the next section will guide our choices for $w(t)$, $\tau$, and $r$ in our estimation procedure.

3. Frequency recovery. To guarantee that Algorithm 1 will recover the true signal frequencies $\nu^\lambda$ 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 $0 \leq k, k' \leq p$, $\min_{k \neq k'} |\nu_k^\lambda - \nu_{k'}^\lambda| \geq \frac{g(T)}{T}$ for some $g(T) \geq 4$.

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 et al., 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 $\nu_k^\lambda$ 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 $\nu$ is at least $2/T$ away from the nearest $\nu_k^\lambda$, then $|H(\nu)|$ is strictly less than\(^4\)

$$\begin{align*}
\frac{2}{T} \sum_{t=0}^{\infty} \sup_{|\nu| \geq \frac{2}{T} + \frac{2}{t}} \left| \tilde{w}(\nu) \right| \cdot \|c^\lambda\|_\infty + \sup_{\nu \in [0, B]} \left| \frac{\tilde{\varepsilon}_w(\nu)}{T} \right|,
\end{align*}$$

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

---

\(^4\)The sum to infinity is needed as $p$ is unknown.
the supplement that
\begin{equation}
\left(\frac{|\tilde{w}(0)|}{T} - \frac{2}{T} \sum_{l=1}^{\infty} \sup_{|\nu| \geq \frac{l}{T}} |\tilde{w}(\nu)|\right) \|c^\lambda\|_\infty - \sup_{\nu \in [0,B]} \frac{|\tilde{\varepsilon}(\nu)|}{T} \leq \sup_{\nu \in [0,B]} |H(\nu)|,
\end{equation}

(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}(\nu)|}{T}.

Substituting the bound (3.2) for \(\|c^\lambda\|_\infty\) into (3.1) shows that the threshold level \(\tau\) in Algorithm 1 should be
\begin{equation}
(3.3) \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}(\nu)|}{T}
\end{equation}
in order to remove from the region \(R\) all frequencies not within \(2/T\) of any \(\nu^\lambda_k\). Our procedure will then select a unique frequency within \(2/T\) of each \(\nu^\lambda_k\) if \(|H(\nu^\lambda_k)| > \tau\), so we can set \(r = 2/T\). In view of (2.4) and (3.3), a sufficient condition for \(|H(\nu^\lambda_k)| > \tau\) is
\begin{align*}
\frac{|\tilde{w}(0)|}{T} |c^\lambda_k| - S_2 \|c^\lambda\|_\infty - \sup_{\nu \in [0,B]} \frac{|\tilde{\varepsilon}(\nu)|}{T} > & \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}(\nu)|}{T} \right\} \\
& + \left( \frac{S_1}{|\tilde{w}(0)|/T - S_2} + 1 \right) \sup_{\nu \in [0,B]} \frac{|\tilde{\varepsilon}(\nu)|}{T}
\end{align*}
for \(k = 0, \ldots, p\), or equivalently
\begin{equation}
(3.5) \frac{|\tilde{w}(0)|}{T} \min_k |c^\lambda_k| > \left\{ S_2 + \frac{S_1 \max \left( S_1, \frac{|\tilde{w}(0)|}{T} + \frac{S_1 + S_2}{2} \right) }{ |\tilde{w}(0)|/T - S_2 } \right\} \max_k |c^\lambda_k| \\
+ 2 \left( \frac{S_1}{|\tilde{w}(0)|/T - S_2} + 1 \right) \sup_{\nu \in [0,B]} \frac{|\tilde{\varepsilon}(\nu)|}{T}.
\end{equation}
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.
Fig 3.1: 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.

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 $|\nu| < \frac{2}{T}$, the time-shifted Hann window has the lightest spectral tails (order $1/(T\nu)^3$):

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

where $\text{sinc}(\nu) = \sin(\pi \nu) / (\pi \nu)$ is the sinc kernel. Note from Figure 3.1 that $|\tilde{w}(\nu)|$ is symmetric and most of its energy is concentrated inside the main lobe between $\nu = \pm \frac{2}{T}$. The sidelobes 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

$$0.02843 < S_1 = \frac{2}{T} \sum_{l=0}^{\infty} \sup_{|\nu| \geq \frac{2}{T} + \frac{1}{T} \nu} |\tilde{w}(\nu)| < 0.02844,$$

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

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

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

The remaining quantity not yet examined in (3.4) and (3.5) is the supremum spectral density $\sup_{\nu \in [0,B]} |\tilde{\varepsilon}(\nu)|$ of the windowed statistical noise.
Noting that $|w(t)| \leq 1$ and $|w'(t)| = \left| \frac{dw(t)}{dt} \right| \leq \pi / T < \infty$, 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 $\sup_{t \in [0,T]} |w(t)| \leq 1$, $\sup_{t \in (0,T)} |w'(t)| < \infty$. Then for any $\beta > 0$, $\gamma > 1$, and $\alpha \geq \gamma / (\gamma - 1)$, with probability

$$1 - 8\gamma\pi B \left[ 1 / T \left( \frac{\gamma - 1}{\gamma} \right)^2 \right] - 2e^{-\Lambda(T)\beta^2 / 4}$$

we have

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

and

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

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

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

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

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

$$\min_k |c_k^\lambda| > 0.0686 \max_k |c_k^\lambda| + 16.9\alpha \left\{ 1 + \left( \frac{1 + \beta}{1 - \beta} \right)^{1/2} \right\} \bar{\Lambda}_T^{1/2} \left( \frac{\log T}{T} \right)^{1/2}.$$
$T$ 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 $\tau$ as (3.7). Then with probability at least

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

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

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

if

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

**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 $O(T^{1/6})$ or greater then the estimation error is $O(T^{-3/2})$ 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 $O(T^{1/2})$. This is because $T^{3/2}(\nu^\lambda - \hat{\nu})$ has a bias of $O(T^{1/2}/g(T))$ 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, $\alpha$, $\beta$, and $\gamma$ 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 $\alpha, \gamma$ satisfying $\alpha(\gamma - 1)/\gamma \geq \sqrt{2}$, 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

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

**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 supplement.

**When does the approach of Shao [2010], Shao and Lii [2011] perform better?** If $\{\nu_k^T\}_k$ are in fact spaced more than order $1/T$ apart from one another, then it follows from (2.5) that the leakage outside a $O(g(T)/T)$-neighbourhood of the frequencies is of order $1/g(T)^3 \to 0$ for the Ham-mwindowed 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 et al., 2013, Candès and Fernandez-Granda, 2013, Fernandez-Granda, 2013, Tang et al., 2015]. In that setting the authors assume a discrete time signal $x = \sum_{k=1}^p c_k e^{2\pi i \nu_k t} \in \mathbb{R}^n$ 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 et al., 2013, Tang et al., 2015] achieves near minimax rates. Furthermore if $\min_k |c_k|$ 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 $\min_k |c_k|$. 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 $\tilde{\Lambda}_T$ is replaced by $\sigma^2$ in A2. Under the discrete time setting above, with high probability our procedure
will select exactly \( p \) frequencies within distance \( \| \hat{\nu} - \nu^\lambda \|_\infty \leq 4/n \) 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

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

where \( \hat{\chi}_T \) is the simulated \( \sup_{\nu \in [0,B]} |\hat{\varepsilon}^{\nu}(\nu)|/T \) 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 \( \tau_\xi \) performs better than \( \tau \) in practice. The following corollary provides a large sample recovery guarantee for \( \tau_\xi \).

**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 \( \alpha \left( \frac{1 + \beta}{1 - \beta} \cdot \frac{\bar{N}_T \log T}{T} \right)^{1/2} \leq \frac{30e}{28 + 2\xi} \| c^\lambda \|_\infty \). Then with probability at least

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

all frequencies will be recovered with the precision stated in Proposition 3 when we threshold with \( \tau_\xi \).

**Remark.** If the second condition is to ever hold, \( \xi \) must then be at least of order \( \| c^\lambda \|_\infty^{-1} \sqrt{\bar{N}_T \log T}/T \).

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 \( \nu^\lambda \) if we wish to estimate the
coefficients $c^\lambda$ consistently. We will therefore let $g(T) \rightarrow \infty$ in Proposition 3 so that $\epsilon(T) \rightarrow 0$. Our estimator is the complex-valued least squares solution to (2.2) in the limit $dt \rightarrow 0$:

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

where the $j$-th 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

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

where $\tilde{I}_{(0, T]}(\nu) = Te^{-i\pi T \nu \text{sinc}(T \nu)}$ 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 $g(T) \rightarrow \infty$. 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

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

is invertible. Then with probability at least

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

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

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

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]:

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

\[ |H_c(\nu)| = \frac{1}{T} \left| \int_0^T e^{-2\pi i \nu t} w(t) \left( dN(t) - \frac{N(T)}{T} dt \right) \right| \]
\[ = \frac{1}{T} \left| H(\nu) - \frac{N(T)}{T} \tilde{w}(\nu) \right|, \] (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\(^5\) 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 \( \tau_\xi \) corresponding to (3.8) is then

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

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 \( 4(\hat{N}_T \log T/T)^{1/2} \) 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

\( ^5 \)In terms of angular frequency, this corresponds to the diameter of \( 12\pi/T \) in Shao and Lii [2011].
\( \mathcal{O}(1/T) \). 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

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

whose frequencies are spaced apart by \( g(T)/T \). The amplitudes \( a_k \) are drawn randomly from \( U[1, 1.5] \), and the phases \( \phi_k \) 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 1.2 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 \( g(T) = T^{1/2} \).

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

\[
\lambda(t) = 0.1 + 0.5 \text{mod}(t, 2\pi) = 0.1 + 0.5\pi - \sum_{k=1}^{\infty} \frac{\sin \left( 2\pi \left( \frac{k}{2\pi} \right) t \right)}{k}
\]

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 \( T = 1,000 \), 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\(^6\) recovered by each method in Table 1.

\(^6\)A correct recovery is defined as one that is within \( 3/T \) of one of the Fourier frequencies.
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.

<table>
<thead>
<tr>
<th></th>
<th>UBIC</th>
<th>WBIC</th>
<th>WThres</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.17 (0.03)</td>
<td>0.19 (0.03)</td>
<td>0.17 (0.03)</td>
</tr>
<tr>
<td>#correct frequencies</td>
<td>3.05 (0.86)</td>
<td>2.77 (0.74)</td>
<td>4.41 (1.32)</td>
</tr>
<tr>
<td>#spurious frequencies</td>
<td>0.01 (0.10)</td>
<td>0.00 (0.00)</td>
<td>0.36 (0.77)</td>
</tr>
</tbody>
</table>

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).

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:

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

The two frequencies at 1/28 and 1/30 are separated by a gap that is slightly larger than $6/T$ when $T = 3,000$. 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 5.1 display the centralized windowed periodograms 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 $O(1/T)$ 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\(^7\) patients from 2014 to Q3 of 2015 ($T = 652$ days).

As shown in the left panel of Figure 5.2, 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 5.2). 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 11am and the second at around 5pm. 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 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 non-cyclical 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.

\(^7\)Defined as Emergency Severity Index (ESI) level 2.
Fig 5.1: 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: $r = 50$. 

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>2.5</th>
<th>3</th>
<th>3.5</th>
<th>4</th>
<th>4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta^*$</td>
<td>0.032</td>
<td>0.033</td>
<td>0.034</td>
<td>0.035</td>
<td>0.036</td>
<td>0.037</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig 5.1: 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: $r = 50$. 

Fig 5.2: 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.
ACKNOWLEDGEMENTS

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. SNN acknowledges support from NSF Award DMS 1723128.

SUPPLEMENTARY MATERIAL

Proofs and asymptotic normality

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


N.Y. Chen
Industrial Engineering & Decision Analytics
Hong Kong University of Science & Technology
E-mail: nychen@ust.hk

D.K.K. Lee
School of Management, and
Department of Statistics & Data Science
Yale University
E-mail: donald.lee@yale.edu

S.N. Negahban
Department of Statistics & Data Science
Yale University
E-mail: sahand.negahban@yale.edu