# APPENDIX A
# PROOF OF LEMMA 6

*Proof:* We only prove the first equation. Let $D = \{t : x(t) \geq 0\}$ be the set of timestamps at which observations are available. Then, it is easy to see that $|S_i^+| = \sum_{t \in S_i \bigcap D} x_0(t)$. Now, let us partition $S_i$ into

$$S_{i,j} = \{i + jT, i + (j + T_0)T, i + (j + 2T_0)T, \ldots\},$$

where $j = \{0, \ldots, T_0 - 1\}$. Since each subsequence $\{x_0(t) : t \in S_{i,j}\}$ is the realization of a single mixing process, we have w.p. 1 that

$$\begin{aligned}\lim_{n\to\infty} \frac{|S_i^+|}{n} &= \lim_{n\to\infty} \frac{\sum_{j=0}^{T_0-1} \sum_{t\in S_{i,j} \bigcap D} x_0(t)}{|S_i \bigcap D|} \cdot \frac{|S_i \bigcap D|}{n} \\ &= \frac{\rho_f \cdot \sum_{j=0}^{T_0-1} p^{T_0}_{\mathcal{F}_{T_0}(i+j\times T)}}{T_0 T},\end{aligned}$$

where we use $\lim_{n\to\infty} \frac{|S_i \bigcap D|}{n} = \frac{\rho_f}{T}$ for the last equality. Also, since the random process can be decomposed into $T_0$ mixing processes, we have w.p. 1 that $\lim_{n\to\infty} |S^+|/n = \frac{\rho_f}{T_0} \sum_{k=0}^{T_0-1} p_k^{T_0}$. Therefore,

$$\begin{aligned}\lim_{n\to\infty} \mu^+_{\mathcal{X}}(I,T) &= \lim_{n\to\infty} \frac{|S_I^+|/n}{|S^+|/n} = \lim_{n\to\infty} \frac{\sum_{i\in I} |S_i^+|/n}{|S^+|/n} \\ &= \sum_{i\in I} \left( \frac{1}{T} \sum_{j=0}^{T_0-1} \frac{p^{T_0}_{\mathcal{F}_{T_0}(i+j\times T)}}{\sum_{k=0}^{T_0-1} p_k^{T_0}} \right).\end{aligned}$$

□

# APPENDIX B
# PROOF OF THEOREM 3

*Proof:* Define $c_i^+ = \frac{p_i^{T_0}}{\sum_{k=0}^{T_0-1} p_k^{T_0}} - \frac{1}{T_0}$, it is easy to see that the value $\lim_{n\to\infty} \gamma^+_{\mathcal{X}}(T_0)$ is achieved by $I^* = \{i \in [0, T_0 - 1] : c_i^+ > 0\}$. So it suffices to show that for any $T \in \mathbb{Z}$ and $I \in \mathcal{I}_T$,

$$\lim_{n\to\infty} \Delta^+_{\mathcal{X}}(I,T) \leq \lim_{n\to\infty} \Delta^+_{\mathcal{X}}(I^*, T_0) = \sum_{i\in I^*} c_i^+.$$

Meanwhile, from Lemma 6, we have

$$\begin{aligned}\lim_{n\to\infty} \Delta^+_{\mathcal{X}}(I,T) &= \frac{1}{T} \sum_{i\in I} \left\{ \sum_{j=0}^{T_0-1} \left( \frac{p^{T_0}_{\mathcal{F}_{T_0}(i+j\times T)}}{\sum_{k=0}^{T_0-1} p_k^{T_0}} \right) - 1 \right\} \\ &= \frac{1}{T} \sum_{i\in I} \sum_{j=0}^{T_0-1} \left( \frac{p^{T_0}_{\mathcal{F}_{T_0}(i+j\times T)}}{\sum_{k=0}^{T_0-1} p_k^{T_0}} - \frac{1}{T_0} \right) \\ &= \frac{1}{T} \sum_{i\in I} \sum_{j=0}^{T_0-1} c^+_{\mathcal{F}_{T_0}(i+j\times T)} \\ &\leq \frac{1}{T} \sum_{i\in I} \sum_{j=0}^{T_0-1} \max(c^+_{\mathcal{F}_{T_0}(i+j\times T)}, 0) \\ &\leq \frac{1}{T} \sum_{j=0}^{T_0 T-1} \max(c^+_{\mathcal{F}_{T_0}(j)}, 0) \\ &= \frac{1}{T} \times T \sum_{i\in I^*} c_i^+ = \sum_{i\in I^*} c_i^+,\end{aligned}$$

where the fourth equality uses the definition of $I^*$. So the proof is complete. □

# APPENDIX C
# DISCUSSION ON THE PERFORMANCE OF FOURIER TRANSFORM

Fourier transform has long been regarded as one of the standard tools for periodicity analysis. Therefore, some readers may find it rather surprising that it actually performs much worse than our method, especially with randomly generated periodic behaviors. Also, as shown in Figure 12(a) and 12(b), Fourier transform is not as robust as our method w.r.t missing observations. To provide some further understandings of these important issues, next we give a brief review of Fourier transform.

The normalized discrete Fourier transform (DFT) of a sequence $\{x(t)\}_{t=0}^{n-1}$ is a sequence of complex numbers $X(f)$:

$$X(f_{k/n}) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} x(t) e^{-\frac{j2\pi kt}{n}},\ k = 0, \ldots, n-1, \quad (18)$$

where the subscript $k/n$ denotes the frequency (normalized to $[0, 1]$) that each coefficient captures. Since we are dealing with real signals, the Fourier coefficients are symmetric around 0.5. Most importantly, the Fourier transform aims to represent the original signal as a linear combination of the complex sinusoids, given by the inverse Fourier transform:

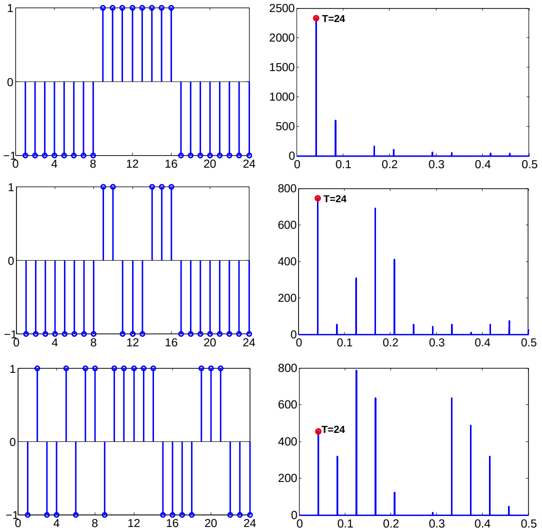
$$x(t) = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} X(f_{k/n}) e^{\frac{j2\pi kt}{n}}, \quad t = 0, \ldots, n-1. \quad (19)$$

To discover potential periodicities in the input sequence $\mathcal{X}$, one needs to examine its power spectrum. Mathematically, this is given by the peroidogram $P$, whose values are the squared length of the Fourier coefficients:

$$P(f_{k/n}) = \|X(f_{k/n})\|^2, \quad k = 0, 1, \ldots, \lceil \frac{n-1}{2} \rceil \quad (20)$$

Then, the dominant period of $\mathcal{X}$ is obtained assuming that *it corresponds to the frequency at which the periodogram achieves its highest value.*

In Figure 18, we show the periodograms of three synthetic sequences, all generated with $T = 24$, $TN = 1000$, $\eta = 1$, $\alpha = 1$ and $\beta = 0$. As one can see, when the periodic behavior is regular (Figure 18, first row), the dominant frequency does correspond to the actual period, suggesting that the time-series can be well-approximated by a sinusoid with period 24. However, this is not true when the periodic behavior is highly irregular (Figure 18, third row). in which case the periodogram is dominated by higher frequencies. This explains why Fourier transform performs miserably with randomly generate periodic behaviors. To the contrary, our method does not make any assumption

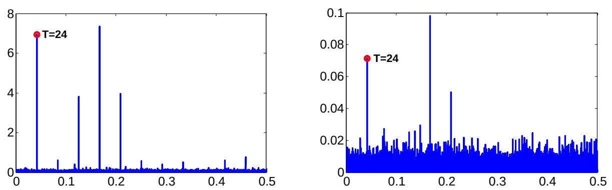



(a) $x(1:T)$ (b) Periodogram

Fig. 18. Sequences with the same period ($T = 24$) may have very different power spectrum. The dominant frequency does not necessarily indicate the true period.

on the periodic behaviors, and is guaranteed to work with any sequence as long as it is generated by some periodically mixing process.

In addition, in Figure 19 we show the periodogram of two sequences, which are generated by sampling the sequence shown in the second row of Figure 18 at sampling rate $\eta = 0.1$ and $\eta = 0.01$, respectively. As one can see, the dominate frequency of the periodogram no longer corresponds to the true period in these cases. This example illustrates that Fourier transform may be sensitive to missing observations.




(a) Periodogram, $\eta = 0.1$ (b) Periodogram, $\eta = 0.01$

Fig. 19. Effect of missing observations on FFT.