Einstein from Noise: Statistical Analysis

Amnon Balanov, Wasim Huleihel, and Tamir Bendory

July 7, 2024

Abstract

"Einstein from noise" (EfN) is a prominent example of the model bias phenomenon: systematic errors in the statistical model that lead to erroneous but consistent estimates. In the EfN experiment, one falsely believes that a set of observations contains noisy, shifted copies of a template signal (e.g., an Einstein image), whereas in reality, it contains only pure noise observations. To estimate the signal, the observations are first aligned with the template using cross-correlation, and then averaged. Although the observations contain nothing but noise, it was recognized early on that this process produces a signal that resembles the template signal! This pitfall was at the heart of a central scientific controversy about validation techniques in structural biology.

This paper provides a comprehensive statistical analysis of the EfN phenomenon above. We show that the Fourier phases of the EfN estimator (namely, the average of the aligned noise observations) converge to the Fourier phases of the template signal, explaining the observed structural similarity. Additionally, we prove that the convergence rate is inversely proportional to the number of noise observations and, in the high-dimensional regime, to the Fourier magnitudes of the template signal. Moreover, in the high-dimensional regime, the Fourier magnitudes converge to a scaled version of the template signal's Fourier magnitudes. This work not only deepens the theoretical understanding of the EfN phenomenon but also highlights potential pitfalls in template matching techniques and emphasizes the need for careful interpretation of noisy observations across disciplines in engineering, statistics, physics, and biology.

1 Introduction

Model bias is a fundamental pitfall arising across a broad range of statistical problems, leading to consistent but inaccurate estimations due to systematic errors in the model. This paper focuses on the *Einstein from Noise* (EfN) experiment: a prototype example of model bias that appears in template matching techniques. Consider a scenario where scientists acquire observational data and genuinely believe their observations contain noisy, shifted copies of a known template signal. However, in reality, their data consists of pure noise with no actual signal present. To estimate the (absent) signal, the scientists align each observation by cross-correlating it with the template and then average the aligned observations. Remarkably, empirical evidence has shown, multiple times, that the reconstructed structure from this process is structurally similar to the template, even when all the measurements are

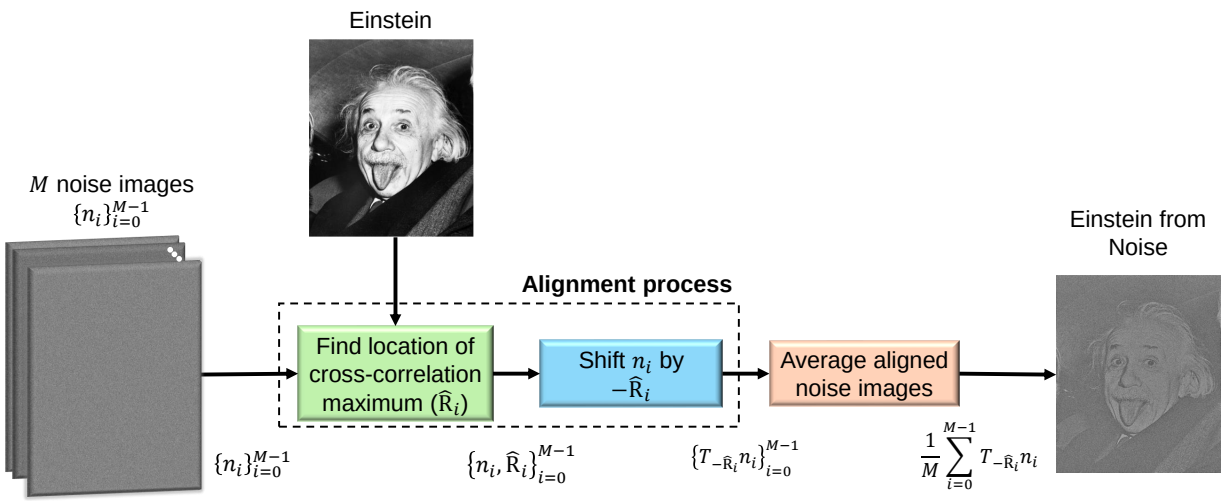

Figure 1: Einstein from Noise. The EfN estimator consists of three stages: (1) finding the index of the maximum of the cross-correlation ($\widehat{\mathrm{R}}_i$) between the i-th noise signal (n_i) and the template signal (e.g., Einstein's image); (2) cyclically shifting the noise signal by $-\widehat{\mathrm{R}}_i$; (3) averaging the shifted noise signals. In this paper, we precisely characterize the relationship between the output of this process—the EfN estimator—and the template signal.

pure noise [18, 36, 38]. This phenomenon stands in striking contrast to the prediction of the unbiased model, that averaging pure noise signals would converge towards a signal of zeros, as the number of noisy observations diverges. Thus, the above EfN estimation procedure is biased towards the template signal.

Although widely recognized, the EfN phenomenon has not been examined so far theoretically. This article fills this gap by characterizing precisely the relationship between the reconstructed and the template signals. The authors of the original article presenting the EfN phenomenon chose an image of Einstein as the template signal, and hence the name [36]. Consequently, we refer to the average of the aligned pure noise signals as the *EfN estimator*. The problem is formulated in detail in Section 3, and is illustrated in Figure 1.

Main results. The central results of this work are as follows. Our first result, stated in Theorem 4.1, shows that the Fourier phases of the EfN estimator converge to the Fourier phases of the template signal, as the number M of noisy observations converges to infinity. We also show that the corresponding mean squared error (MSE) decays to zero with a rate of $1/M$. Since it is known that the Fourier phases are responsible for the formation of geometrical image elements, like contours and edges [30, 37], this clarifies why the resulting EfN estimator image exhibits a structural similarity to the template, but not necessarily a full recovery. Our second result, stated in Theorem 4.3, proves that in the high-dimensional regime, where the dimension of the signal diverges, the convergence rate of the Fourier phases is inversely proportional to the square of the Fourier magnitudes of the template signal. In this case, the Fourier magnitudes of the EfN estimator converge to a scaled version of the template's Fourier magnitudes.

Organization. The rest of this paper is organized as follows. The next section discusses the connection between the EfN problem and single-particle cryo-electron microscopy (cryo-

EM)—the main motivation of this paper—and provides empirical demonstrations. Section 3 formulates the problem in detail. Our main results, Theorems 4.1 and 4.3, are presented in Section 4, and proved in Appendix B and Appendix C, respectively. Finally, our conclusions and outlook appear in Section 5.

2 Cryo-EM and Empirical Demonstration

Cryo-EM is a powerful tool of modern structural biology, offering advanced methods to visualize complex biological macromolecules with ever-increasing precision. One of its central advantages lies in its capability to resolve the structures of proteins that are hard to crystallize in traditional methods, especially, in a near-physiological environment, see e.g., [29, 40]. This advantage enables researchers to delve into the dynamic behaviors of proteins and their complexes, shedding light on fundamental biological processes.

Cryo-EM uses single-particle electron microscopy to reconstruct 3D structures from 2D tomographic projection images [6]. Typically, the 3D reconstruction involves two main steps: detecting and extracting single particle images using a particle picking algorithm, [35, 17, 8, 16], and then reconstructing the 3D density map [34, 33]. Most detection algorithms use template-matching techniques, which can introduce bias if improper templates are chosen, especially in low signal-to-noise ratio (SNR) conditions, which is the standard scenario in cryo-EM.

The EfN controversy. A publication of the 3D structure of an HIV molecule in PNAS in 2013 [27] initiated a fundamental controversy about validation techniques within the cryo-EM community, published as four follow-up PNAS publications [18, 46, 44, 26]. The EfN pitfall played a central role in this discussion. The primary question of the discussion was whether the collected datasets contained informative biological data or merely pure noise images. The core of the debate emphasized the importance of exercising caution and implementing cross-validation techniques when fitting data to a predefined model. This precautionary approach aims to mitigate the risk of erroneous fittings, which could ultimately lead to inaccuracies in 3D density map reconstruction. Model bias is still a fundamental problem in cryo-EM, as highlighted by an ongoing debate concerning validation tools, see for example, [43, 36, 19, 11, 12, 20, 21, 42].

Empirical demonstration. The EfN phenomenon depends on several key parameters: (1) the number of observations which we denote by M; (2) the dimension of the signal, denoted as d; and (3) the power spectral density (PSD) of the template signal. To demonstrate the dependency on these parameters and provide insight into our main results, Figures 2 and 3 show the convergence of the EfN estimator. Specifically, Figure 2 illustrates the behavior of the Fourier phases as a function of M. Figure 2(c) highlights that the convergence rate is proportional to $1/M$. It can be seen that the convergence rate is faster for higher spectral components. Figure 3 illustrates the impact of the PSD of the template signal on the cross-correlation between the template and the EfN estimator. Notably, a flatter PSD (i.e., a faster decay of the auto-correlation) leads to a higher correlation between the template and the estimator signals. These empirical results are proved theoretically in Theorems 4.1 and 4.3.

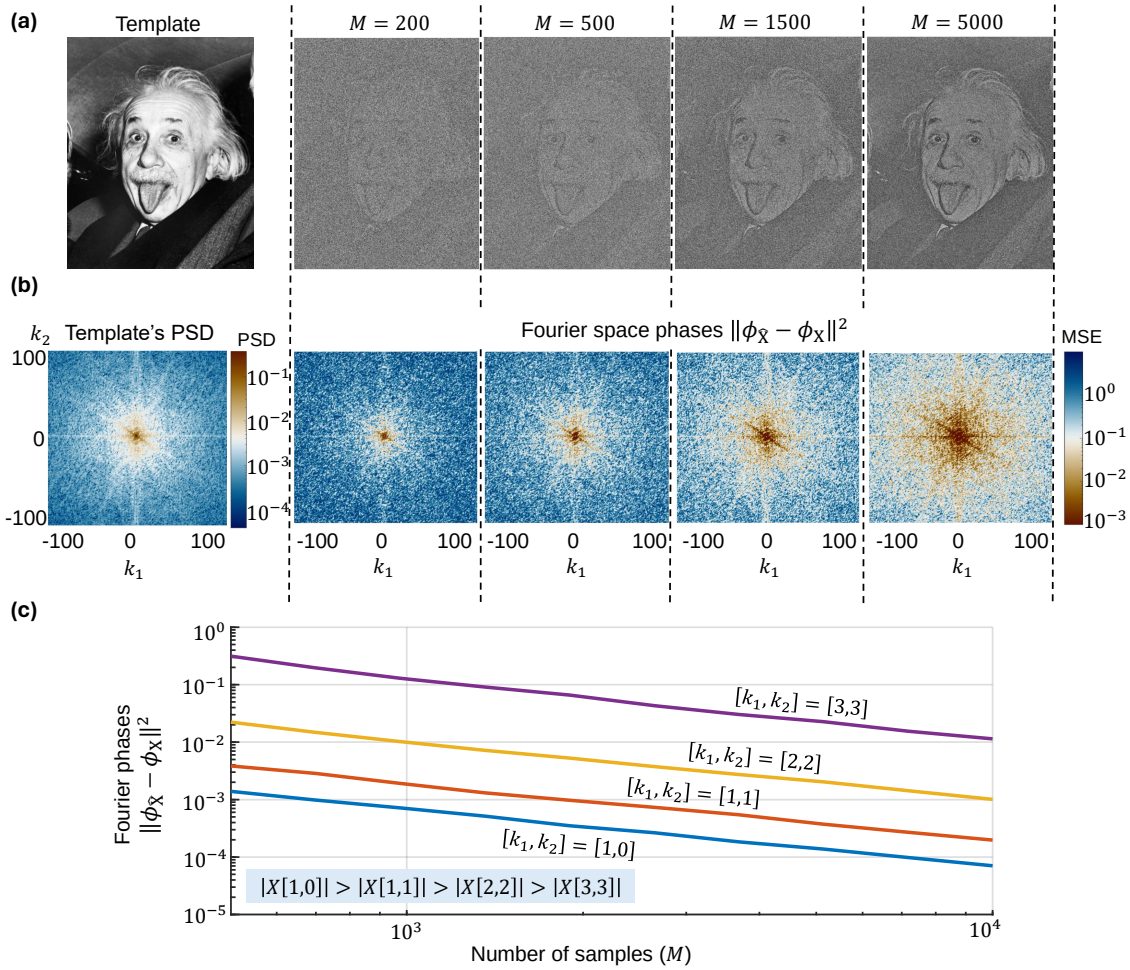

Figure 2: The impact of the number of noise observations on the EfN estimator. **(a)** The structural similarity between the EfN estimator and the template image as a function of the number of noise observations (M). **(b)** The MSE between the phases of the template image $\mathsf{X}[k_1, k_2]$ and the estimator $\hat{\mathsf{X}}[k_1, k_2]$ for $-100 \leq k_1, k_2 \leq 100$, with varying observation sizes ($M = 200, 500, 1500, 5000$). More observations lead to lower MSE. **(c)** The convergence rate of the MSE between the Fourier phases of the EfN estimator and the Fourier phases of the template as a function of the number of observations across different frequencies. The relationship between MSE and M is $\propto M^{-1}$. In addition, stronger spectral components lead to lower MSE. Figures **(b)** and **(c)** were generated through 200 Monte-Carlo trials.

More applications. The EfN phenomenon extends to various applications employing template matching, whether through a feature-based or direct template-based approach. For instance, template matching holds significance in computational anatomy, where it aids in discovering unknown diffeomorphism to align a template image with a target image [10]. Other areas include medical imaging processing [1], manufacturing quality control [3], and navigation systems for mobile robots [22]. This pitfall may also arise in the feature-based approach, which relies on extracting image features like shapes, textures, and colors to match a target image by neural networks and deep-learning classifiers [50, 28, 45, 24].

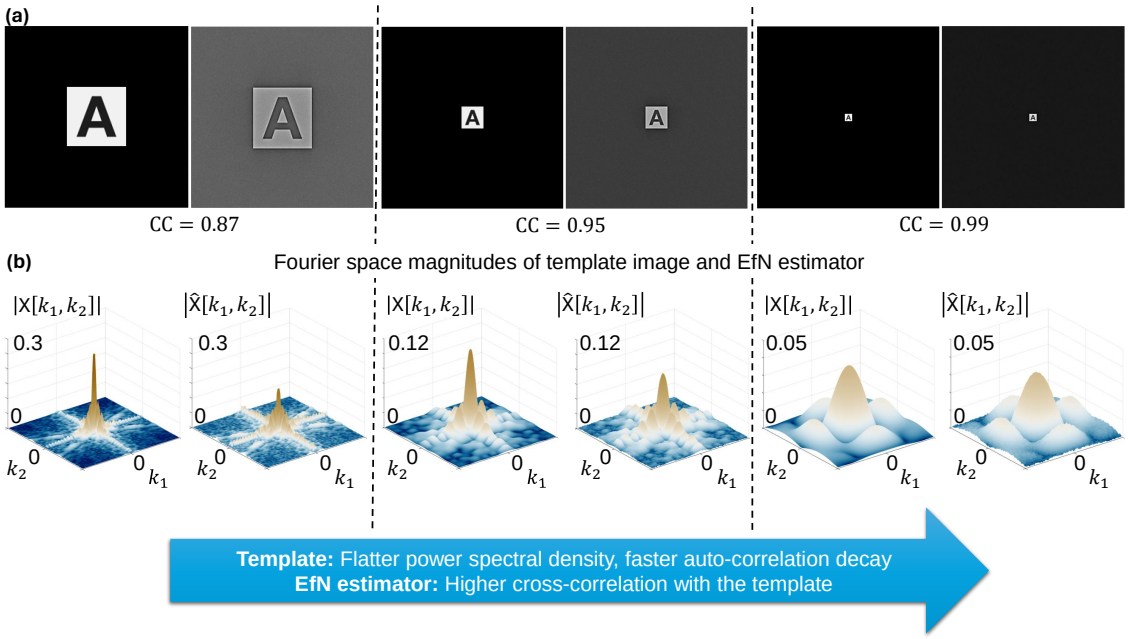

Figure 3: The influence of the PSD of the template signal on the correlation between the template and the EfN estimator. (a) Three images of the letter A are shown, with an increasing zero-padding ratio. As the zero-padding ratio increases, the PSD flattens, and the cross-correlation (CC) between the template and the EfN estimator increases. This higher cross-correlation is evident in both the image background and the colors of the letter A. **(b)** Flatter PSDs lead to EfN estimators whose Fourier magnitudes are closer to those of the template image. The EfN estimators in these experiments were generated using $M = 10^5$ observations.

3 Problem Formulation and Notation

This section outlines the probabilistic model behind the EfN experiment and delineates our main mathematical objectives. Although the EfN phenomenon is typically described for images, we will formulate and analyze it for one-dimensional signals, bearing in mind that the extension to two-dimensional images is straightforward (see Section 5 for more details).

Let $x \in \mathbb{R}^d$ denote the *template signal* (e.g., an Einstein image), and $n_i \sim \mathcal{N}(0, \sigma^2 I_{d\times d})$, for $i = 0, 1, \ldots, M-1$, be a set of M independent and identically distributed (i.i.d.) d-dimensional Gaussian noise vectors. Here, M denotes the number of *observations*, and without loss of generality, we will assume that d is even. To describe the EfN estimation process, we define the circular shift operator. Fix $0 \leq \ell \leq d-1$, and let $\mathcal{T}_\ell : \mathbb{R}^d \to \mathbb{R}^d$ denotes an operator which acts on $y \in \mathbb{R}^d$, and defined as $[\mathcal{T}_\ell y]_i \triangleq y_{(i-\ell) \bmod d}$, for $0 \leq i \leq d-1$. Throughout this paper, we assume that the template signal x is normalized, i.e., $\|x\|_2^2 = 1$, where $\|\cdot\|_2$ is the Euclidean norm, and further assume that its Fourier transform in non-vanishing. The first assumption is used for convenience and does not alter (up to a normalization factor) our main results in Theorems 4.1 and 4.3. The second assumption is essential for the theoretical analysis of the EfN process and is expected to hold in many applications, including cryo-EM. A similar assumption is frequently taken in related work, e.g., [5, 31, 7].

We are now in a position to define the EfN estimation process. First, for each noise

observation $i = 0, \ldots, M-1$, we compute the maximal correlation shift,

$$\hat{\mathsf{R}}_i \triangleq \arg\max_{0 \le \ell \le d-1} \langle n_i, \mathcal{T}_\ell x \rangle. \tag{3.1}$$

Then, the EfN estimator is given by the average of the noise observations, but each first aligned according to the above maximal shifts, i.e.,

$$\hat{x} \triangleq \frac{1}{M} \sum_{i=0}^{M-1} \mathcal{T}_{-\hat{\mathsf{R}}_i} n_i. \tag{3.2}$$

The EfN phenomenon states that, at least empirically, $\hat{x}$ and x appear "close" in some sense; our goal is to understand this phenomenon mathematically. To that end, we will consider the two asymptotic regimes where either $M \to \infty$ and d is fixed, or $M, d \to \infty$. It should be noted that since the spectrum of x is non-vanishing, $\hat{\mathsf{R}}_i$ is unique almost surely.

As will become clear in the next sections, it is convenient to work in the Fourier domain. Let $\phi_{\mathsf{Z}} \triangleq \sphericalangle \mathsf{Z}$ denote the phase of a complex number $\mathsf{Z} \in \mathbb{C}$, and recall that the discrete Fourier transform (DFT) of a d-length signal $y \in \mathbb{R}^d$ is given by,

$$\mathsf{Y}[k] \triangleq \mathcal{F}\{y\} = \sum_{\ell=0}^{d-1} y_\ell e^{-j\frac{2\pi}{d}k\ell}, \tag{3.3}$$

where $j \triangleq \sqrt{-1}$, and $0 \le k \le d-1$. Accordingly, we let X, $\hat{\mathsf{X}}$, and N_i, denote the DFTs of x, $\hat{x}$, and n_i, respectively, for $0 \le i \le M-1$. These DFT sequences can be equivalently represented in the magnitude-phase domain as follows,

$$\mathsf{X} = \{|\mathsf{X}[k]|\, e^{j\phi_{\mathsf{X}}[k]}\}_{k=0}^{d-1}, \quad \hat{\mathsf{X}} = \{|\hat{\mathsf{X}}[k]| e^{j\phi_{\hat{\mathsf{X}}}[k]}\}_{k=0}^{d-1}, \quad \mathsf{N}_i = \{|\mathsf{N}_i[k]|\, e^{j\phi_{\mathsf{N}_i}[k]}\}_{k=0}^{d-1}, \tag{3.4}$$

for $0 \le i \le M-1$. Note that the random variables $\{|\mathsf{N}_i[k]|\}_{k=0}^{d/2}$ and $\{\phi_{\mathsf{N}_i}[k]\}_{k=0}^{d/2}$ are two independent sequences of i.i.d. random variables, such that, $|\mathsf{N}_i[k]| \sim \mathsf{Rayleigh}\,(\sigma^2)$ has Rayleigh distribution, and the phase $\phi_{\mathsf{N}_i}[k] \sim \mathsf{Unif}[-\pi, \pi)$ is uniformly distributed over $[-\pi, \pi)$.

With the definitions above, we can express the estimation process in the Fourier domain. Since a shift in real-space corresponds to a linear phase shift in the Fourier space, it follows that,

$$\hat{\mathsf{X}}[k] = \frac{1}{M} \sum_{i=0}^{M-1} |\mathsf{N}_i[k]|\, e^{j\phi_{\mathsf{N}_i}[k]} e^{j\frac{2\pi k}{d}\hat{\mathsf{R}}_i}, \tag{3.5}$$

for $k = 0, 1, \ldots, d-1$. It is important to note that the location of the maximum correlation, i.e., $\hat{\mathsf{R}}_i$, captures the dependency on the template signal, as well as the connections between the different spectral components.

We mention that recent research has explored a related, but distinct, problem [47]. In this alternate problem formulation, rather than averaging over all shifted noisy signals, only the "most biased" members—those with the highest cross-correlation values with the template—are averaged. Finally, throughout the rest of this paper, we use $\xrightarrow{\mathcal{D}}$, $\xrightarrow{\mathcal{P}}$, $\xrightarrow{\text{a.s.}}$, and $\xrightarrow{\mathcal{L}^p}$, to denote the convergence of sequences of random variables in distribution, in probability, almost surely, and in $\mathcal{L}^p$ norm, respectively. We denote by $\mathbb{E}|\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]|^2$ the MSE of the Fourier phases of the k-th spectral component.

4 Main Results

In this section, we present our main results. We begin by analyzing the regime where $M \to \infty$ and fixed d. We show that the Fourier phases of the EfN estimator converge to the Fourier phases of the template signal. We also characterize the convergence of the magnitudes. Then, we consider the high-dimensional regime, where $d \to \infty$ as well. Here, we will prove stronger convergence results, provided that some additional assumptions are met. Throughout the following theorems, we assume that the template signal x has unit norm, and that its spectrum is non-vanishing, as discussed in the previous section.

Theorem 4.1. *Fix $d \geq 2$ and assume that $\mathsf{X}[k] \neq 0$, for all $0 < k \leq d-1$.*

1. *For any $0 \leq k \leq d-1$, we have,*

$$\phi_{\hat{\mathsf{X}}}[k] \xrightarrow{\text{a.s.}} \phi_{\mathsf{X}}[k], \tag{4.1}$$

as $M \to \infty$. Furthermore,

$$\lim_{M\to\infty} \frac{\mathbb{E}|\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]|^2}{1/M} = C_k, \tag{4.2}$$

for a finite constant $C_k < \infty$.

2. *For any $0 \leq k \leq d-1$, we have,*

$$|\hat{\mathsf{X}}[k]| \xrightarrow{\text{a.s.}} \mathbb{E}\left[|\mathsf{N}[k]| \cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \phi_{\mathsf{N}}[k] - \phi_{\mathsf{X}}[k]\right)\right], \tag{4.3}$$

as $M \to \infty$, where $\hat{\mathsf{R}}_1$ is defined in (3.1).

Theorem 4.1 captures two central properties. The first addresses the convergence of the EfN estimator's phases to those of the template signal. In addition, the corresponding convergence rate in MSE is proportional to $1/M$. The second result captures the convergence of the EfN estimator's magnitudes to the term given in the right-hand-side (r.h.s.) of (4.3). Interestingly, this term is not necessarily proportional to the magnitudes $|\mathsf{X}[k]|$ of the template signal.

Next, we consider the high-dimensional regime $d \to \infty$ (after taking $M \to \infty$). Here, we place several additional technical assumptions on the template signal. Roughly speaking, we need to control the decay rate of the auto-correlation function of the template signal x as a function of d; the auto-correlation function of the signal should decay faster than $1/\log d$, and each one of the template's Fourier magnitudes components should decay faster than $1/\sqrt{\log d}$. Note that both $\mathsf{X}[k]$ and $\hat{\mathsf{X}}[k]$ depend on d. We define the auto-correlation function of the signal x by R_{XX}, and recall the auto-correlation function is given by the Fourier transform of the PSD of x, i.e., $\mathsf{R}_{\mathsf{XX}} = \mathcal{F}\left\{|\mathsf{X}|^2\right\}$.

Assumption 4.2. *Let $\mathsf{X} = \mathcal{F}\{x\}$ and let R_{XX} be the auto-correlation of the signal x. We say that the template signal x satisfies Assumption 4.2 if the following hold:*

1. *The auto-correlation satisfies,*

$$\lim_{d\to\infty} \mathsf{R}_{\mathsf{XX}}[d] \cdot \log d = 0. \tag{4.4}$$

2. *The magnitudes satisfy,*

$$\lim_{d\to\infty} \left\{ \max_{0<k\leq d-1} \{|\mathsf{X}[k]|\} \cdot \sqrt{\log d} \right\} = 0. \tag{4.5}$$

3. *The signal's DC component is zero, i.e.,* $|\mathsf{X}[0]| = 0$.

Theorem 4.3. *Assume that* $\mathsf{X}[k] \neq 0$, *for all* $0 < k \leq d-1$, *and that* x *satisfies Assumption 4.2. Then,*

1. *For any* $0 \leq k \leq d-1$, *we have,*

$$\lim_{M,d\to\infty} \frac{\mathbb{E}|\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]|^2}{1/(M\log d)} \frac{1}{1/(4\,|\mathsf{X}[k]|^2)} = 1. \tag{4.6}$$

2. *For any* $0 \leq k \leq d-1$, *we have,*

$$\frac{1}{\sqrt{2\log d}} \frac{|\hat{\mathsf{X}}[k]|}{|\mathsf{X}[k]|} \xrightarrow{\text{a.s.}} 1, \tag{4.7}$$

as $M, d \to \infty$.

Based on Theorem 4.3, as $M, d \to \infty$, the convergence rate of the Fourier phases of the EfN estimator is inversely proportional to the Fourier magnitude square. In addition, unlike Theorem 4.1, the Fourier magnitudes of the EfN estimator converge to those of the template signal, up to a constant factor. Therefore, when $d \to \infty$, under Assumption 4.2, the signal $\hat{x}$ recovers the template signal, up to a known normalization factor. This, in turn, implies that the normalized cross-correlation between the template and the EfN estimator approaches unity.

Figure 4 exemplifies the results of Theorem 4.3 and shows that the convergence rate depends on the template signal's PSD. In particular, Figure 4(b) demonstrates how the vector length and spectral density impact the convergence of the magnitudes of the estimator, showing that longer vectors and flatter spectral densities yield higher cross-correlation between the estimator and the template signal. In addition, Figure 4(c) demonstrates that the estimator's phases convergence closely aligns with our analytical prediction as the PSD flattens. We remark that the last assumption in Assumption 4.2 about the zero DC component is not necessary empirically but is used as part of the proof.

Strategy of the proofs. We discuss briefly the ideas behind the proofs of our main results. The proof of Theorem 4.1 relies on two central ingredients. Specifically, it can be shown that the central limit theorem (CLT) and the law of large numbers (LLN) imply that as $M \to \infty$, we have $\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k] \xrightarrow{\mathcal{D}} \arctan(\mathsf{Q}_k)$, where Q_k is a Gaussian random variable with zero mean and variance $\sigma^2_{\mathsf{Q}}[k] = C_k/M$, with some explicit formula for C_k. Then, by using

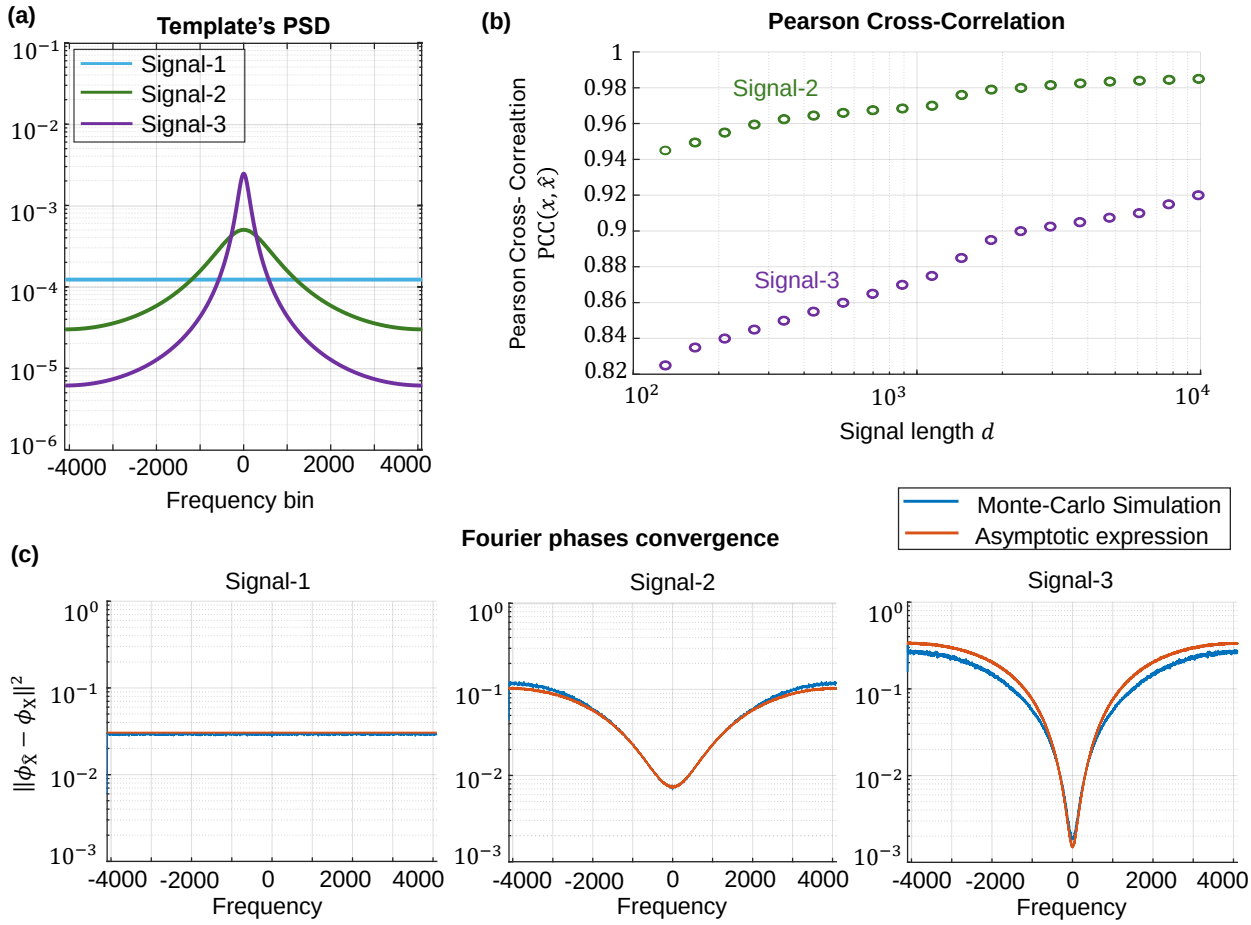

Figure 4: Comparison between analytic expression and Monte-Carlo simulations for asymptotic signal length, d, and for signals with varying power spectral densities. The analytic expressions for the Fourier phases convergence and Fourier magnitudes are given by (4.6) and (4.7), respectively. **(a)** Three signals with varying PSDs are examined. **(b)** Monte-Carlo simulation of the Pearson cross-correlation between the template, x, and the estimator, $\hat{x}$. As d increases, the correlation between the estimator and the template vector increases, especially for signals with faster-decaying auto-correlations. **(c)** The analytic expression (4.6) closely aligns with Monte-Carlo simulations and shows better agreement for signals characterized by faster-decaying auto-correlations. The simulations were conducted with $d = 8192$ and $M = 10000$, based on 2000 Monte-Carlo trials.

certain properties of cyclo-stationary Gaussian processes, we show that $C_k < \infty$, for any $0 \leq k \leq d-1$, from which we deduce (4.1)–(4.2). Then, to obtain the refined convergence rate in Theorem 4.3 for $d \to \infty$ (namely, that $C_k = \frac{1}{4|\mathsf{X}[k]|^2 \log d}$), we utilize results from the theory of extrema of Gaussian processes, particularly, the convergence of the maximum of a stationary Gaussian process to the Gumbel distribution, see, e.g., [23, 9, 2, 4].

5 Conclusion and outlook

In this paper, we have shown that the Fourier phases of the EfN estimator converge to those of the template signal for an asymptotic number of observations. Since Fourier phases are crucial for perceiving image structure, the reconstructed image appears structurally similar to the template signal, even in cases where the estimator's spectral magnitudes differ from those of the template [30, 37]. We also show that the Fourier phases of spectral components with higher magnitudes converge faster, leading to faster structural similarity in the overall image perception.

Perspective. We anticipate that the findings of this paper will be beneficial in various fields. For example, the paper sheds light on a fundamental pitfall in template matching techniques, which may lead engineers and statisticians to misleading results. In addition, physicists and biologists working with data sets of low SNRs will benefit from understanding limitations and potential biases introduced by template matching techniques. More generally, this work provides a cautionary framework for the broader scientific community, highlighting the importance of exercising care when interpreting noisy observations.

Implications to cryo-EM. These findings have practical implications for cryo-EM. Typically, protein spectra exhibit rapid decay at low frequencies (known as the Guinier plot) and remain relatively constant at high frequencies, a behavior characterized by Wilson in [49] and known as Wilson statistics. Wilson statistics is used to sharpen 3-D structures [39]. To mitigate the risk of model bias, we suggest using templates with reduced high frequencies, recommending filtered, smooth templates. This insight may also relate to, or support the common practice of initializing the expectation-maximization (EM) algorithm for 3-D refinement with a smooth 3-D volume. Each iteration of the EM algorithm effectively applies a version of template matching multiple times, although projection images typically contain a signal, not just noise as in the EfN case.

Extension to higher dimensions and asymptotic regimes. While this paper focuses on one-dimensional signals, the analysis can be readily extended to higher dimensions. This extension involves replacing the one-dimensional DFT with its N-dimensional counterpart. The symmetry properties established in Theorem 4.1, including the results in Propositions B.1 and B.3, remain valid. For the high-dimensional case of Theorem 4.3, the conditions on the PSD adjust for the N-dimensional case. Specifically, the auto-correlation decay rate of the multidimensional array should be faster than $1/\log d$, in each dimension. Finally, while our focus in this paper was on the asymptotic regimes where either d is fixed and $M \to \infty$ or both $M, d \to \infty$, there are other challenging asymptotic regimes that should be studied. Most notably, it is interesting to understand what happens in the regime where both $M, d \to \infty$, but with a fixed ratio, say, $\frac{d}{M} \to c \in (0, \infty)$; it seems as though that more powerful techniques are needed to analyze this scenario.

Hard assignment algorithms and the EM algorithm. One promising avenue for future research involves examining hard-assignment algorithms. These algorithms iteratively refine estimates of an underlying signal from noisy observations, where the signal is obscured by high noise (unlike the pure noise scenario in EfN). The process begins by aligning observations with a template signal in the initial iteration and averaging them to improve the

template for subsequent iterations. A primary goal is to characterize the model bias in this iterative algorithm, particularly focusing on the relationship between the output and the initial model.

Another important direction is investigating the EM algorithm, a cornerstone of cryo-EM algorithms [34, 33]. EM maximizes the likelihood function of models incorporating nuisance parameters [14], a topic of significant recent interest [13, 48]. Unlike hard-assignment algorithms, EM operates iteratively as a soft assignment algorithm, assigning probabilities to various possibilities and computing a weighted average rather than selecting a single optimal alignment per observation. Further exploration of EM could provide deeper insights into iterative methodologies in cryo-EM and their associated model biases.

Acknowledgment

T.B. is supported in part by BSF under Grant 2020159, in part by NSF-BSF under Grant 2019752, and in part by ISF under Grant 1924/21. W.H. is supported by ISF under Grant 1734/21.

References

[1] Ashley Aberneithy. Automatic detection of calcified nodules of patients with tuberculous. *University College, London*, 2007.

[2] Robert J Adler and Jonathan E Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.

[3] MS Aksoy, Orhan Torkul, and Ismail Hakki Cedimoglu. An industrial visual inspection system that uses inductive learning. *Journal of Intelligent Manufacturing*, 15:569–574, 2004.

[4] Jean-Marc Azaïs and Mario Wschebor. *Level sets and extrema of random processes and fields*. John Wiley & Sons, 2009.

[5] Afonso S Bandeira, Ben Blum-Smith, Joe Kileel, Jonathan Niles-Weed, Amelia Perry, and Alexander S Wein. Estimation under group actions: recovering orbits from invariants. *Applied and Computational Harmonic Analysis*, 66:236–319, 2023.

[6] Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2):58–76, 2020.

[7] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on signal processing*, 66(4):1037–1050, 2017.

[8] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J Noble, and Bonnie Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods*, 16(11):1153–1160, 2019.

[9] Simeon M Berman. Limit theorems for the maximum term in stationary sequences. *The Annals of Mathematical Statistics*, pages 502–516, 1964.

[10] Gary E Christensen, Richard D Rabbitt, and Michael I Miller. Deformable templates using large deformation kinematics. *IEEE transactions on image processing*, 5(10):1435–1447, 1996.

[11] Jon Cohen. Is high-tech view of HIV too good to be true?, 2013.

[12] Pilar Cossio. Need for cross-validation of single particle cryo-EM. *Journal of Chemical Information and Modeling*, 60(5):2413–2418, 2020.

[13] Constantinos Daskalakis, Christos Tzamos, and Manolis Zampetakis. Ten steps of EM suffice for mixtures of two gaussians. In *Conference on Learning Theory*, pages 704–710. PMLR, 2017.

[14] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.

[15] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[16] Amitay Eldar, Keren Mor Waknin, Samuel Davenport, Tamir Bendory, Armin Schwartzman, and Yoel Shkolnisky. Object detection under the linear subspace model with application to cryo-EM images. *arXiv preprint arXiv:2405.00364*, 2024.

[17] Ayelet Heimowitz, Joakim Andén, and Amit Singer. APPLE picker: Automatic particle picking, a low-effort cryo-EM framework. *Journal of structural biology*, 204(2):215–227, 2018.

[18] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013.

[19] Richard Henderson, Andrej Sali, Matthew L Baker, Bridget Carragher, Batsal Devkota, Kenneth H Downing, Edward H Egelman, Zukang Feng, Joachim Frank, Nikolaus Grigorieff, et al. Outcome of the first electron microscopy validation task force meeting. *Structure*, 20(2):205–214, 2012.

[20] J Bernard Heymann. Validation of 3D EM reconstructions: The phantom in the noise. *AIMS biophysics*, 2(1):21, 2015.

[21] Gerard J Kleywegt, Paul D Adams, Sarah J Butcher, Catherine L Lawson, Alexis Rohou, Peter B Rosenthal, Sriram Subramaniam, Maya Topf, Sanja Abbott, Philip R Baldwin, et al. Community recommendations on cryoEM data archiving and validation. *IUCrJ*, 11(2), 2024.

[22] Theocharis Kyriacou, Guido Bugmann, and Stanislao Lauria. Vision-based urban navigation procedures for verbally instructed robots. *Robotics and Autonomous Systems*, 51(1):69–80, 2005.

[23] Malcolm R Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Science & Business Media, 2012.

[24] Yuhai Li, Jian Liu, Jinwen Tian, and Hongbo Xu. A fast rotated template matching based on point feature. In *MIPPR 2005: SAR and Multispectral Image Processing*, volume 6043, pages 453–459. SPIE, 2005.

[25] Sergio I Lopez and Leandro PR Pimentel. On the location of the maximum of a process: L'evy, gaussian and multidimensional cases. *arXiv preprint arXiv:1611.02334*, 2016.

[26] Youdong Mao, Luis R Castillo-Menendez, and Joseph G Sodroski. Reply to subramaniam, van heel, and henderson: Validity of the cryo-electron microscopy structures of the HIV-1 envelope glycoprotein complex. *Proceedings of the National Academy of Sciences*, 110(45):E4178–E4182, 2013.

[27] Youdong Mao, Liping Wang, Christopher Gu, Alon Herschhorn, Anik Désormeaux, Andrés Finzi, Shi-Hua Xiang, and Joseph G Sodroski. Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proceedings of the National Academy of Sciences*, 110(30):12438–12443, 2013.

[28] Amit Moscovich and Saharon Rosset. On the cross-validation bias due to unsupervised preprocessing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1474–1502, 2022.

[29] Eva Nogales. The development of cryo-EM into a mainstream structural biology technique. *Nature methods*, 13(1):24–27, 2016.

[30] Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

[31] Amelia Perry, Jonathan Weed, Afonso S Bandeira, Philippe Rigollet, and Amit Singer. The sample complexity of multireference alignment. *SIAM Journal on Mathematics of Data Science*, 1(3):497–517, 2019.

[32] Leandro PR Pimentel. On the location of the maximum of a continuous stochastic process. *Journal of Applied Probability*, 51(1):152–161, 2014.

[33] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature methods*, 14(3):290–296, 2017.

[34] Sjors HW Scheres. RELION: implementation of a bayesian approach to cryo-EM structure determination. *Journal of structural biology*, 180(3):519–530, 2012.

[35] Sjors HW Scheres. Semi-automated selection of cryo-EM particles in relion-1.3. *Journal of structural biology*, 189(2):114–122, 2015.

[36] Maxim Shatsky, Richard J Hall, Steven E Brenner, and Robert M Glaeser. A method for the alignment of heterogeneous macromolecules from electron microscopy. *Journal of structural biology*, 166(1):67–78, 2009.

[37] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[38] Fred J Sigworth. A maximum-likelihood approach to single-particle image refinement. *Journal of structural biology*, 122(3):328–339, 1998.

[39] Amit Singer. Wilson statistics: derivation, generalization and applications to electron cryomicroscopy. *Acta Crystallographica Section A: Foundations and Advances*, 77(5):472–479, 2021.

[40] Amit Singer and Fred J Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual review of biomedical data science*, 3:163–190, 2020.

[41] E. Slutsky. Über stochastische Asymptoten und Grenzwerte. 1925.

[42] Carlos OS Sorzano, JL Vilas, Erney Ramírez-Aportela, J Krieger, D Del Hoyo, David Herreros, Estrella Fernandez-Giménez, D Marchán, JR Macías, I Sánchez, et al. Image processing tools for the validation of CryoEM maps. *Faraday Discussions*, 240:210–227, 2022.

[43] Alex Stewart and Nikolaus Grigorieff. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy*, 102(1):67–84, 2004.

[44] Sriram Subramaniam. Structure of trimeric HIV-1 envelope glycoproteins. *Proceedings of the National Academy of Sciences*, 110(45):E4172–E4174, 2013.

[45] Itamar Talmi, Roey Mechrez, and Lihi Zelnik-Manor. Template matching with deformable diversity similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 175–183, 2017.

[46] Marin van Heel. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, 110(45):E4175–E4177, 2013.

[47] Shao-Hsuan Wang, Yi-Ching Yao, Wei-Hau Chang, and I-Ping Tu. Quantification of model bias underlying the phenomenon of "Einstein from noise". *Statistica Sinica*, 31:2355–2379, 2021.

[48] Ji Xu, Daniel J Hsu, and Arian Maleki. Global analysis of expectation maximization for mixtures of two gaussians. *Advances in Neural Information Processing Systems*, 29, 2016.

[49] SH Yü. Determination of absolute from relative X-ray intensity data. *Nature*, 150(3796):151–152, 1942.

[50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

Appendix A Preliminaries

Before we delve into the proofs of Theorems 4.1 and 4.3, we fix some notations and definitions. Recall the definition of the Fourier transform of x and n_i (3.4). Note that since x and n_i are real-valued, their Fourier coefficients satisfy the conjugate-symmetry relation:

$$\mathsf{X}[k] = \overline{\mathsf{X}[d-k]}, \quad \mathsf{N}_i[k] = \overline{\mathsf{N}_i[d-k]}.$$

In particular, $|\mathsf{N}_i[k]| = |\mathsf{N}_i[d-k]|$ and $\phi_{\mathsf{N}_i}[k] = -\phi_{\mathsf{N}_i}[d-k]$, which implies that only the first $d/2+1$ components of $\mathsf{N}[k]$ are statistically independent.

The definition of the maximal correlation in (3.1) can be represented in the Fourier domain as follows,

$$\hat{\mathsf{R}}_i = \arg\max_{0\le r\le d-1} \sum_{k=0}^{d-1} |\mathsf{X}[k]|\,|\mathsf{N}_i[k]|\,\cos\left(\frac{2\pi kr}{d} + \phi_{\mathsf{N}_i}[k] - \phi_{\mathsf{X}}[k]\right). \tag{A.1}$$

To simplify notation, we define

$$\mathsf{S}_i[r] \triangleq \sum_{k=0}^{d-1} |\mathsf{X}[k]|\,|\mathsf{N}_i[k]|\,\cos\left(\frac{2\pi kr}{d} + \phi_{\mathsf{N}_i}[k] - \phi_{\mathsf{X}}[k]\right), \tag{A.2}$$

for $0 \le r \le d-1$, and therefore, $\hat{\mathsf{R}}_i = \arg\max_{0\le r\le d-1} \mathsf{S}_i[r]$. We note that for any $0 \le i \le M-1$, the random vector $\mathsf{S}_i \triangleq (\mathsf{S}_i[0], \mathsf{S}_i[1], \ldots, \mathsf{S}_i[d-1])^T$ is Gaussian distributed, with zero mean vector, and a circulant covariance matrix; therefore, it is a cyclo-stationary random process.

Throughout the proofs, we condition on the k-th Fourier coefficient of the noise realization. Specifically, note that $\mathsf{S}_i|\mathsf{N}_i[k] \sim \mathcal{N}(\mu_{k,i}, \Sigma_{k,i})$, where

$$\mu_{k,i}[r] \triangleq \mathbb{E}\left[\mathsf{S}_i[r]|\mathsf{N}_i[k]\right] = 2\,|\mathsf{X}[k]|\,|\mathsf{N}_i[k]|\cos\left(\frac{2\pi kr}{d} + \phi_{\mathsf{N}_i}[k] - \phi_{\mathsf{X}}[k]\right) \tag{A.3}$$

for $0 \leq r \leq d-1$, and

$$\begin{aligned}\Sigma_{k,i}[r,s] &\triangleq \mathbb{E}\left[\left(\mathsf{S}_i[r]-\mathbb{E}\mathsf{S}_i[r]\right)\left(\mathsf{S}_i[s]-\mathbb{E}\mathsf{S}_i[s]\right)|\mathsf{N}_i[k]\right] \\ &= \sigma^2 \sum_{l=0}^{d-1} |\tilde{\mathsf{X}}_k[l]|^2 \cos\left(\frac{2\pi l}{d}(r-s)\right), \end{aligned} \tag{A.4}$$

for $0 \leq r, s \leq d-1$, where $\tilde{\mathsf{X}}_k$ is defined by:

$$\tilde{\mathsf{X}}_k[l] \triangleq \begin{cases} 0 & \text{if } l = k, d-k, \\ \mathsf{X}[l] & \text{if } l = 0, d/2, \\ \sqrt{2}\cdot \mathsf{X}[l] & \text{otherwise.} \end{cases} \tag{A.5}$$

Without loss of generality, we normalize the diagonal elements of the above covariance matrix by setting,

$$\sigma^2 = \left[\sum_{l=0}^{d-1} |\tilde{\mathsf{X}}_k[l]|^2\right]^{-1}. \tag{A.6}$$

Note that the conditional process $\mathsf{S}_i|\mathsf{N}_i[k]$ is Gaussian because it is given by a linear transform of i.i.d. Gaussian variables. Also, since its covariance matrix is circulant and depends only on the difference between the two indices, i.e., $\Sigma_{k,i}[r,s] = \sigma_{k,i}[|r-s|]$, it is cycle-stationary with a cosine trend. The eigenvalues of this circulant matrix are given by the DFT of its first row, and thus its ℓ-th eigenvalue equals to $|\tilde{\mathsf{X}}_k[\ell]|^2$, for $0 \leq \ell \leq d-1$. Finally, for simplicity of notation, whenever it is clear from the context, we will omit the dependence of the above quantities on the i-th observation and k-th frequency indices, and use $\mu[r]$ and $\Sigma[r,s]$, instead. Furthermore, for convenience, we will assume the template vector is normalized to unity, i.e., $\sum_{l=0}^{d-1} |\mathsf{X}[l]|^2 = 1$.

Our goal is to investigate the phase and magnitude of the estimator $\hat{\mathsf{X}}$ in (3.5). Simple manipulations reveal that, for any $0 \leq k \leq d-1$, the estimator's phases are given by,

$$\phi_{\hat{\mathsf{X}}}[k] = \phi_{\mathsf{X}}[k] + \arctan\left(\frac{\sum_{i=0}^{M-1} |\mathsf{N}_i[k]| \sin\left(\phi_{e,i}[k]\right)}{\sum_{i=0}^{M-1} |\mathsf{N}_i[k]| \cos\left(\phi_{e,i}[k]\right)}\right), \tag{A.7}$$

where we define,

$$\phi_{e,i}[k] \triangleq \frac{2\pi k \hat{\mathsf{R}}_i}{d} + \phi_{\mathsf{N}_i}[k] - \phi_{\mathsf{X}}[k], \tag{A.8}$$

and

$$|\hat{\mathsf{X}}[k]| = \frac{1}{M}\left|\sum_{i=0}^{M-1} |\mathsf{N}_i[k]|\, e^{j\phi_{e,i}[k]}\right|. \tag{A.9}$$

Appendix B Proof of Theorem 4.1

We will use the following two results. The proofs are provided in Sections B.1 and B.2.

Proposition B.1. *Fix $d \in \mathbb{N}$. Then, for any $0 \leq k \leq d-1$,*

$$\sqrt{M} \cdot \tan\left(\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]\right) \xrightarrow{\mathcal{D}} \mathsf{Q}_k, \tag{B.1}$$

as $M \to \infty$, where Q_k is distributed as,

$$\mathsf{Q}_k \sim \mathcal{N}\left(0, \frac{\mathbb{E}[|\mathsf{N}_1[k]| \sin(\phi_{e,1}[k])]^2}{\left(\mathbb{E}[|\mathsf{N}_1[k]| \cos(\phi_{e,1}[k])]\right)^2}\right), \tag{B.2}$$

and $\phi_{e,1}[k]$ is defined in (A.8).

Remark B.2. *Note that the above result implies that C_k in (4.2) is given by,*

$$C_k = \frac{\mathbb{E}[|\mathsf{N}_1[k]| \sin(\phi_{e,1}[k])]^2}{\mathbb{E}[|\mathsf{N}_1[k]| \cos(\phi_{e,1}[k])]^2}. \tag{B.3}$$

Proposition B.3. *Fix $d \in \mathbb{N}$, and assume that $\mathsf{X}[k] \neq 0$ for all $0 < k \leq d-1$. Then, for any $0 \leq k \leq d-1$,*

$$\mathbb{E}[|\mathsf{N}_1[k]| \cos(\phi_{e,1}[k])] > 0. \tag{B.4}$$

By definition, Proposition B.3 implies that there is a positive correlation between the EfN estimator $\hat{x}$ and the template x. We are now in a position to prove Theorem 4.1.

Proof of Theorem 4.1. We start with the convergence of the estimator's magnitudes. Following (A.9), by applying the strong law of large numbers (SLLN), we get,

$$\frac{1}{M}\sum_{i=0}^{M-1} |\mathsf{N}_i[k]| \, e^{j\phi_{e,i}[k]} \xrightarrow{\text{a.s.}} \mathbb{E}\left[|\mathsf{N}_1[k]| \cos\left(\phi_{e,1}[k]\right)\right] + j\mathbb{E}\left[|\mathsf{N}_1[k]| \sin\left(\phi_{e,1}[k]\right)\right], \tag{B.5}$$

where we have used the fact that the sequences of random variables $\{|\mathsf{N}_i[k]| \sin\left(\phi_{e,i}[k]\right)\}_{i=0}^{M-1}$ and $\{|\mathsf{N}_i[k]| \cos\left(\phi_{e,i}[k]\right)\}_{i=0}^{M-1}$ are i.i.d. with finite mean and variances. Now, using simple symmetry arguments, we show in the proof of Proposition B.1 (see, (B.38)), that,

$$\mathbb{E}\left[|\mathsf{N}_1[k]| \sin\left(\phi_{e,1}[k]\right)\right] = 0. \tag{B.6}$$

Together with (B.5), this proves the second item of Theorem 4.1.

Next, we prove the first item of Theorem 4.1, starting with (4.1). To this end, using (A.7) and using the continuous mapping theorem, it is evident that it suffices to prove that $\frac{\sum_{i=0}^{M-1}|\mathsf{N}_i[k]| \sin(\phi_{e,i}[k])}{\sum_{i=0}^{M-1}|\mathsf{N}_i[k]| \cos(\phi_{e,i}[k])} \xrightarrow{\text{a.s.}} 0$. This, however, follows by applying the SLLN,

$$\frac{\sum_{i=0}^{M-1} |\mathsf{N}_i[k]| \sin\left(\phi_{e,i}[k]\right)}{\sum_{i=0}^{M-1} |\mathsf{N}_i[k]| \cos\left(\phi_{e,i}[k]\right)} \xrightarrow{\text{a.s.}} \frac{\mu_A}{\mu_B}, \tag{B.7}$$

where $\mu_A \triangleq \mathbb{E}\left[|\mathsf{N}_1[k]|\sin(\phi_{e,1}[k])\right]$ and $\mu_B \triangleq \mathbb{E}\left[|\mathsf{N}_1[k]|\cos(\phi_{e,1}[k])\right]$. We already saw that $\mu_A = 0$, while by Proposition B.3, we have that $\mu_B > 0$, and thus their ratio converges a.s. to zero by the Continuous Mapping Theorem. Thus, we proved that $\phi_{\hat{\mathsf{X}}}[k] \xrightarrow{\text{a.s.}} \phi_{\mathsf{X}}[k]$.

Finally, we prove (4.2). To that end, we use Portmanteau Lemma, which states that if $\{D_\ell\}$ is a sequence of random variables such that $D_\ell \xrightarrow{\mathcal{D}} D_\infty$, then for any bounded and continuous function f we have $\mathbb{E}[f(D_\ell)] \to \mathbb{E}[f(D_\infty)]$, as $\ell \to \infty$. In our case, for any fixed $0 \le k \le d-1$, we let this sequence of random variables, indexed by M, be defined as $D_M \triangleq \sqrt{M} \cdot \tan\left(\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]\right)$, as in Proposition B.1, and we note that the index M is hidden also in $\phi_{\hat{\mathsf{X}}}[k]$. Then, Proposition B.1 implies that $D_M \xrightarrow{\mathcal{D}} \mathsf{Q}_k \triangleq D_\infty$, where we recall that $\mathsf{Q}_k \sim \mathcal{N}(0, C_k)$. The Portmanteau Lemma then implies that for any bounded f,

$$\mathbb{E}\left[f\left(\sqrt{M} \cdot \tan\left(\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]\right)\right)\right] \longrightarrow \mathbb{E}\left[f(\mathsf{Q}_k)\right]. \tag{B.8}$$

We are now ready to prove (4.2). Using (A.7) it is clear that,

$$\sqrt{M} \cdot \tan\left[\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]\right] = \frac{\frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} |\mathsf{N}_i[k]| \sin\left(\phi_{e,i}[k]\right)}{\frac{1}{M} \sum_{i=0}^{M-1} |\mathsf{N}_i[k]| \cos\left(\phi_{e,i}[k]\right)} \triangleq \frac{\mathsf{A}_M}{\mathsf{B}_M}, \tag{B.9}$$

and thus

$$\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k] = \arctan\left(\frac{1}{\sqrt{M}} \frac{\mathsf{A}_M}{\mathsf{B}_M}\right). \tag{B.10}$$

Therefore, we have,

$$\mathbb{E}\left|\phi_{\hat{\mathsf{X}}}[k] - \phi_{\mathsf{X}}[k]\right|^2 = \mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}} \frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\right]. \tag{B.11}$$

Theorem 4.1 is a direct application of the following result.

Lemma B.4. *Recall the definition of* Q_k *in* (B.2)*. Then,*

$$\lim_{M\to\infty} \frac{\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}} \frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\right]}{\frac{1}{M}\mathbb{E}\left[\mathsf{Q}_k^2\right]} = 1. \tag{B.12}$$

Proof of Lemma B.4. One should note that the denominator B_M can be zero with positive probability, and thus we need to control such an event. To that end, recall that $\mathsf{B}_M \xrightarrow{\text{a.s.}} \mu_B$. Then, for any $0 < \epsilon < \mu_B$, we decompose,

$$\begin{aligned}\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}} \frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\right] &= \mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}} \frac{\mathsf{A}_M}{\mathsf{B}_M}\right) \mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right] \\ &\quad + \mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}} \frac{\mathsf{A}_M}{\mathsf{B}_M}\right) \mathbb{1}_{|\mathsf{B}_M|<\epsilon}\right]. \end{aligned} \tag{B.13}$$

Next, we first show that the second term at the r.h.s. of (B.13) converges to zero with rate $O(\frac{1}{M^2})$. Indeed, since $\arctan(x) \le \frac{\pi}{2}$, for any $x \in \mathbb{R}$, we have

$$\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}} \frac{\mathsf{A}_M}{\mathsf{B}_M}\right) \mathbb{1}_{|\mathsf{B}_M|<\epsilon}\right] \le \frac{\pi^2}{4} \cdot \mathbb{E}\left[\mathbb{1}_{|\mathsf{B}_M|<\epsilon}\right] \tag{B.14}$$

$$\leq \frac{\pi^2}{4} \cdot \mathbb{P}\left(\mathsf{B}_M < \epsilon\right) \tag{B.15}$$

$$= \frac{\pi^2}{4} \cdot \mathbb{P}\left(\mathsf{B}_M - \mu_B < \epsilon - \mu_B\right) \tag{B.16}$$

$$\leq \frac{\pi^2}{4} \cdot \mathbb{P}\left(|\mathsf{B}_M - \mu_B| > \mu_B - \epsilon\right). \tag{B.17}$$

Let us denote the summand in the denominator in (B.9) by $V_i \triangleq |\mathsf{N}_i[k]| \cos\left(\phi_{e,i}[k]\right)$, for $0 \leq i \leq M-1$. Then, we note that,

$$\mathbb{E}(V_i^4) = \mathbb{E}\left[|\mathsf{N}_i[k]| \cos\left(\phi_{e,i}[k]\right)\right]^4 \leq \mathbb{E}\left[|\mathsf{N}_i[k]|\right]^4 < \infty. \tag{B.18}$$

Thus, by Chebyshev's inequality,

$$\mathbb{P}\left(|\mathsf{B}_M - \mu_B| > \mu_B - \epsilon\right) \leq \frac{\mathbb{E}\left[\mathsf{B}_\mathsf{M} - \mu_B\right]^4}{\left(\mu_B - \epsilon\right)^4}. \tag{B.19}$$

Now, by the definition of B_M, we have,

$$\mathbb{E}\left[\mathsf{B}_M - \mu_B\right]^4 = \frac{1}{M^4} \sum_{i,j,k,l=0}^{M-1} \mathbb{E}\left[\left(V_i - \mu_B\right)\left(V_j - \mu_B\right)\left(V_k - \mu_B\right)\left(V_l - \mu_B\right)\right] \tag{B.20}$$

$$= \frac{1}{M^4}\left[M \cdot \mathbb{E}\left[V_i - \mu_B\right]^4 + 3M(M-1)\left(\mathbb{E}\left[V_i - \mu_B\right]^2\right)^2\right]. \tag{B.21}$$

Thus, it is clear that there is a constant D_1 (which depends on the second and forth moment of V_i), such that,

$$\frac{\mathbb{E}\left[\mathsf{B}_\mathsf{M} - \mu_B\right]^4}{\left(\mu_B - \epsilon\right)^4} \leq \frac{D_1}{\left(\mu_B - \epsilon\right)^4 M^2}. \tag{B.22}$$

Thus, plugging (B.19) and (B.22) into (B.17) leads to,

$$\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}}\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\mathbb{1}_{|\mathsf{B}_M|<\epsilon}\right] \leq \frac{\pi^2}{4} \cdot \frac{D_1}{\left(\mu_B - \epsilon\right)^4 M^2}. \tag{B.23}$$

Thus, the second term at the r.h.s. of (B.13) indeed converges to zero as $1/M^2$. Next, we analyze the first term at the r.h.s. of (B.13). We will show that,

$$1 \leq \lim_{M\to\infty} \frac{\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}}\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right]}{\frac{1}{M}\mathbb{E}\left[\mathsf{Q}_k^2\right]} \leq \frac{\mu_B^2}{\epsilon^2}. \tag{B.24}$$

As this is true for every $\epsilon < \mu_B$, it would imply that,

$$\lim_{M\to\infty} \frac{\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}}\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right]}{\frac{1}{M}\mathbb{E}\left[\mathsf{Q}_k^2\right]} = 1. \tag{B.25}$$

Let use denote by $\mathsf{G}_M \triangleq \frac{\mathsf{A}_M}{\mathsf{B}_M}\mathbb{1}_{|\mathsf{B}_M|>\epsilon}$. Since $[\arctan(x)]^2/x^2 \to 1$, as $x \to 0$, and because $\frac{1}{\sqrt{M}}\frac{\mathsf{A}_M}{\mathsf{B}_M} \xrightarrow{\text{a.s.}} 0$, then the following holds from the Taylor expansion of $\arctan(x)$, around $x = 0$,

$$\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}}\cdot \mathsf{G}_M\right)\right] = \mathbb{E}\left[\sum_{k=0}^{\infty}(-1)^k\frac{\left[\frac{\mathsf{G}_M}{\sqrt{M}}\right]^{2k+1}}{2k+1}\right]^2. \tag{B.26}$$

As the sum in (B.26) converges for every M, and converges to zero as $M \to \infty$, then,

$$\mathbb{E}\left[\sum_{k=0}^{\infty}(-1)^k\frac{\left[\frac{\mathsf{G}_M}{\sqrt{M}}\right]^{2k+1}}{2k+1}\right]^2 = \frac{1}{M}\mathbb{E}\left[\mathsf{G}_M^2\right] + \sum_{(k_1,k_2)\neq(0,0)}\frac{(-1)^{k_1+k_2}}{(2k_1+1)(2k_2+1)}\mathbb{E}\left[\left(\frac{\mathsf{G}_M}{\sqrt{M}}\right)^{(2k_1+2k_2+2)}\right]. \tag{B.27}$$

Now, since the term at the left-hand-side of (B.27) as well as the first term at the right-hand-side of (B.27), are bounded for every M and converges to zero, then also the last term at the right-hand-side of (B.27) is bounded for every M and converge to zero as $M \to \infty$. Specifically, we note that the last term converges to zero with rate $1/M^2$, while the first term in the right-hand-side converges to zero with rate $1/M$. Thus,

$$\lim_{M\to\infty}\frac{\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}}\cdot\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right]}{\frac{1}{M}\mathbb{E}\left[\left(\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)^2\mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right]} = 1. \tag{B.28}$$

Also, note that

$$\lim_{M\to\infty}\mathbb{E}\left[\left(\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)^2\mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right] \leq \lim_{M\to\infty}\frac{1}{\epsilon^2}\mathbb{E}\left[\mathsf{A}_M^2\mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right] = \frac{\sigma_A^2}{\epsilon^2}. \tag{B.29}$$

Thus, combining (B.28) and (B.29) we get,

$$\lim_{M\to\infty}\frac{\mathbb{E}\left[\arctan^2\left(\frac{1}{\sqrt{M}}\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)\mathbb{1}_{|\mathsf{B}_M|>\epsilon}\right]}{\frac{1}{M}\mathbb{E}\left[\mathsf{Q}_k^2\right]} \leq \lim_{M\to\infty}\frac{\frac{1}{\epsilon^2 M}\sigma_A^2}{\frac{1}{M}\mathbb{E}\left[\mathsf{Q}_k^2\right]} = \frac{\mu_B^2}{\epsilon^2}, \tag{B.30}$$

which proves the upper bound in (B.24). Similarly, since $\mathsf{B}_M \xrightarrow{\text{a.s.}} \mu_B$, for any $\epsilon_2 > 0$, we have,

$$\lim_{M\to\infty}\mathbb{E}\left[\left(\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)^2\mathbb{1}_{\{\mathsf{B}_M>\epsilon\}}\right] \geq \lim_{M\to\infty}\mathbb{E}\left[\left(\frac{\mathsf{A}_M}{\mathsf{B}_M}\right)^2\mathbb{1}_{\{\mathsf{B}_M>\epsilon\}\wedge\{\mathsf{B}_M<\mu_B+\epsilon_2\}}\right] \tag{B.31}$$

$$\geq \frac{\sigma_A^2}{(\mu_B+\epsilon_2)^2}. \tag{B.32}$$

Since (B.32) is true for every $\epsilon_2 > 0$, we get the lower bound in (B.24), which concludes the proof of (B.25). Substituting (B.23) and (B.25) in (B.13), leads to the proof of the lemma. □

□

B.1 Proof of Proposition B.1

Recall the relation in (B.9) and the definitions of A_M and B_M. Since $\{\mathsf{N}_i\}_{i=0}^{M-1}$ is an i.i.d. sequence of random variables, and because each $\phi_{e,i}$ depends on N_i solely (in particular, independent of N_j, for $j \neq i$), we have that $\{|\mathsf{N}_i[k]| \sin(\phi_{e,i}[k])\}_{i=0}^{M-1}$ and $\{|\mathsf{N}_i[k]| \cos(\phi_{e,i}[k])\}_{i=0}^{M-1}$ are two sequences of i.i.d. random variables. Accordingly, let,

$$\mu_{\mathsf{A}} \triangleq \mathbb{E}\left[|\mathsf{N}_1[k]| \sin(\phi_{e,1}[k])\right], \tag{B.33}$$

$$\sigma_{\mathsf{A}}^2 = \mathsf{Var}\left(|\mathsf{N}_1[k]| \sin(\phi_{e,1}[k])\right), \tag{B.34}$$

which are the mean value and variance of A_M, as defined in (B.9). Then, by the CLT:

$$\left(\mathsf{A}_M - \sqrt{M}\mu_{\mathsf{A}}\right) \xrightarrow{\mathcal{D}} \mathsf{A}, \tag{B.35}$$

where $\mathsf{A} \sim \mathcal{N}(0, \sigma_{\mathsf{A}}^2)$.

Next, we show that $\mu_{\mathsf{A}} = 0$. Indeed, let $\mathsf{D}[k] \triangleq \phi_{\mathsf{X}}[k] - \phi_{\mathsf{N}_1}[k]$, and recall the definition of $\hat{\mathsf{R}}_i$ in (A.1). Note that $\hat{\mathsf{R}}_1$ depends on $\phi_{\mathsf{N}_1}[k]$ and $\phi_{\mathsf{X}}[k]$ only through $\mathsf{D}[k]$. Accordingly, viewing $\hat{\mathsf{R}}_1$ as a function of $\mathsf{D}[k]$, we have,

$$\hat{\mathsf{R}}_1\left(-\mathsf{D}[0], -\mathsf{D}[1], \ldots, -\mathsf{D}[d-1]\right) = -\hat{\mathsf{R}}_1\left(\mathsf{D}[0], \mathsf{D}[1], \ldots, \mathsf{D}[d-1]\right), \tag{B.36}$$

namely, by flipping the signs of all the phases, the location of the maximum flips its sign as well. Then, by the law of total expectation,

$$\begin{aligned}\mu_{\mathsf{A}} &= \mathbb{E}\left[|\mathsf{N}_1[k]| \sin\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \phi_{\mathsf{N}_1}[k] - \phi_{\mathsf{X}}[k]\right)\right] \\ &= \mathbb{E}\left\{|\mathsf{N}_1[k]| \cdot \mathbb{E}\left[\sin\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \phi_{\mathsf{N}_1}[k] - \phi_{\mathsf{X}}[k]\right) \middle| \{|\mathsf{N}_1[k]|\}_{k=0}^{d-1}\right]\right\}.\end{aligned} \tag{B.37}$$

The inner expectation in (B.37) is taken w.r.t. the randomness of the phases $\{\phi_{\mathsf{N}_1}[k]\}_{k=0}^{d-1} \in [-\pi, \pi)$. However, due to (B.36), and since the sine function is odd around zero, the integration in (B.37) nullifies. Therefore,

$$\mathbb{E}\left[\sin\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \phi_{\mathsf{N}_1}[k] - \phi_{\mathsf{X}}[k]\right) \middle| \{|\mathsf{N}_1[k]|\}_{k=0}^{d-1}\right] = 0, \tag{B.38}$$

and thus $\mu_{\mathsf{A}} = 0$.

Next, we analyze the denominator in (B.9). Specifically, we already saw that $\{|\mathsf{N}_i[k]| \cos(\phi_{e,i}[k])\}_{i=0}^{M-1}$ form a sequence of i.i.d. random variables, and thus by the SLLN we have $\mathsf{B}_M \xrightarrow{\text{a.s.}} \mu_{\mathsf{B}}$, where,

$$\mu_{\mathsf{B}} \triangleq \mathbb{E}\left[|\mathsf{N}_1[k]| \cos(\phi_{e,1}[k])\right]. \tag{B.39}$$

We will prove in Proposition B.3 that $\mu_{\mathsf{B}} > 0$. Finally, applying Slutsky's Theorem on the ratio $\frac{\mathsf{A}_M}{\mathsf{B}_M}$, we obtain,

$$\frac{\mathsf{A}_M}{\mathsf{B}_M} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma_{\mathsf{A}}^2}{\mu_{\mathsf{B}}^2}\right) = \mathsf{Q}_k, \tag{B.40}$$

which concludes the proof.

B.2 Proof of Proposition B.3

To prove Proposition B.3, we will first establish some notations and state two auxiliary results. Let $\mathsf{S}^{(1)} \sim \mathcal{N}(\mu, \Sigma)$ and $\mathsf{S}^{(2)} \sim \mathcal{N}(-\mu, \Sigma)$ be two d-dimensional Gaussian vectors, where Σ is the circulant covariance matrix defined in (A.4). We define the entries of μ as,

$$\mu_\ell \triangleq [\mu]_\ell = \cos\left(\frac{2\pi k}{d}\ell + \varphi\right), \tag{B.41}$$

for $\varphi \in [0, 2\pi)$, and $0 \leq \ell \leq d-1$. Note that $-\mu_\ell = \cos\left(\frac{2\pi k}{d}\ell + \varphi + \pi\right)$, for $0 \leq \ell \leq d-1$. Define,

$$\hat{\mathsf{R}}^{(1)} = \underset{0 \leq \ell \leq d-1}{\arg\max}\, \mathsf{S}_\ell^{(1)}, \tag{B.42}$$

$$\hat{\mathsf{R}}^{(2)} = \underset{0 \leq \ell \leq d-1}{\arg\max}\, \mathsf{S}_\ell^{(2)}. \tag{B.43}$$

We further claim that due to the assumption that $\mathsf{X}[k] \neq 0$, for all $0 < k \leq d-1$ in Proposition B.3, it follows that $\mathsf{rank}(\Sigma) \geq d/2$. Indeed, for the rank of the covariance matrix to be larger than $d/2$, at least half of its eigenvalues should be non-zero. Now, as mentioned in Subsection A, the eigenvalues of Σ are given by $|\tilde{\mathsf{X}}_k[\ell]|^2$, for $0 \leq \ell \leq d-1$. Thus, assuming that the spectrum of X is not vanishing clearly implies that $\mathsf{rank}(\Sigma) \geq d/2$; in fact it implies that $\mathsf{rank}(\Sigma) \geq d-2$, which is larger than $d/2$, for $d \geq 4$. We have the following result.

Lemma B.5. *Consider the definitions in* (B.41)–(B.43). *Fix* $0 \leq \ell \leq d-1$. *If* $\mu_\ell > 0$, *then,*

$$\mathbb{P}\left[\hat{\mathsf{R}}^{(1)} = \ell\right] > \mathbb{P}\left[\hat{\mathsf{R}}^{(2)} = \ell\right], \tag{B.44}$$

otherwise, if $\mu_\ell < 0$, *then,*

$$\mathbb{P}\left[\hat{\mathsf{R}}^{(1)} = \ell\right] < \mathbb{P}\left[\hat{\mathsf{R}}^{(2)} = \ell\right]. \tag{B.45}$$

In particular, for any $\varphi \in [0, 2\pi)$ *and* $0 \leq k \leq d-1$,

$$\mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}^{(1)} + \varphi\right)\right] + \mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}^{(2)} + \varphi + \pi\right)\right] > 0. \tag{B.46}$$

Before proving Lemma B.5, we use its result to establish Proposition B.3. Our goal is to prove that $\mathbb{E}[|\mathsf{N}_1[k]|\cos(\phi_{e,1}[k])] > 0$. By the law of total expectation, we have,

$$\begin{aligned}\mathbb{E}[|\mathsf{N}_1[k]|\cos(\phi_{e,1}[k])] &= \mathbb{E}\left[|\mathsf{N}_1[k]| \cdot \mathbb{E}\left(\cos(\phi_{e,1}[k])|\,\mathsf{N}_1[k]\right)\right] \\ &= \mathbb{E}\left[|\mathsf{N}_1[k]| \cdot \mathbb{E}\left(\cos\left(\frac{2\pi k\hat{\mathsf{R}}_1}{d} + \phi_{\mathsf{N}_1}[k] - \phi_{\mathsf{X}}[k]\right)\middle|\,\mathsf{N}_1[k]\right)\right]. \end{aligned} \tag{B.47}$$

More explicitly, we can write,

$$\begin{aligned}&\mathbb{E}[|\mathsf{N}_1[k]|\cos(\phi_{e,1}[k])] = \\ &\frac{1}{2\pi}\int_0^\infty \mathrm{d}n n f_{|\mathsf{N}_1[k]|}(n)\int_{-\pi}^{\pi}\mathrm{d}\varphi\mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \varphi\right)\middle|\,|\mathsf{N}_1[k]| = n, \phi_{\mathsf{N}_1}[k] = \phi_{\mathsf{X}}[k] + \varphi\right].\end{aligned} \tag{B.48}$$

Now, note that the inner integral can be written as,

$$\begin{aligned}&\int_{-\pi}^{\pi}\mathrm{d}\varphi\mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \varphi\right)\middle|\,|\mathsf{N}_1[k]| = n, \phi_{\mathsf{N}_1}[k] = \phi_{\mathsf{X}}[k] + \varphi\right] \\ &\quad = \frac{1}{\pi}\int_0^{\pi}\mathrm{d}\varphi\;\mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \varphi\right)\middle|\,|\mathsf{N}_1[k]| = n, \phi_{\mathsf{N}_1}[k] = \phi_{\mathsf{X}}[k] + \varphi\right] + \\ &\quad + \frac{1}{\pi}\int_0^{\pi}\mathrm{d}\varphi\;\mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}_1 + \varphi + \pi\right)\middle|\,|\mathsf{N}_1[k]| = n, \phi_{\mathsf{N}_1}[k] = \phi_{\mathsf{X}}[k] + \varphi + \pi\right].\end{aligned} \tag{B.49}$$

The main observation here is that the conditional distribution of $\hat{\mathsf{R}}_1$ on the event $\{|\mathsf{N}_1[k]| = n, \phi_{\mathsf{N}_1}[k] = \phi_{\mathsf{X}}[k] + \varphi\}$ coincides with that of $\hat{\mathsf{R}}^{(1)}$ in (B.42), and similarly the conditional distribution of $\hat{\mathsf{R}}_1$ on the event $\{|\mathsf{N}_1[k]| = n, \phi_{\mathsf{N}_1}[k] = \phi_{\mathsf{X}}[k] + \varphi + \pi\}$ coincides with that of $\hat{\mathsf{R}}^{(2)}$ in (B.43). Thus, the sum of the integrands at the r.h.s. of (B.49) is exactly the left-hand-side (l.h.s.) of (B.46), and thus by Lemma B.5, this sum is positive for every $\varphi \in [0, \pi]$. Together with (B.48), this concludes the proof of Proposition B.3. It is left to prove Lemma B.5.

Proof of Lemma B.5. By definition, it is clear that,

$$\mathbb{P}\left[\hat{\mathsf{R}}^{(m)} = \ell\right] = \mathbb{P}\left[\mathsf{S}_\ell^{(m)} \geq \max_{i\neq\ell}\mathsf{S}_i^{(m)}\right], \tag{B.50}$$

for $m = 1, 2$ and $0 \leq \ell \leq d-1$. Since $\mathsf{S}^{(1)}$ and $\mathsf{S}^{(2)}$ can be decomposed as $\mathsf{S}^{(1)} = \mathsf{Z} + \mu$ and $\mathsf{S}^{(2)} = \mathsf{Z} - \mu$, where Z is a cyclo-stationary process, and μ is defined in (B.41). Then,

$$\mathbb{P}\left[\mathsf{S}_\ell^{(1)} \geq \max_{i\neq\ell}\mathsf{S}_i^{(1)}\right] = \mathbb{P}\left[\mathsf{Z}_\ell + \mu_\ell \geq \max_{m\neq\ell}\mathsf{Z}_m + \mu_m\right], \tag{B.51}$$

and,

$$\mathbb{P}\left[\mathsf{S}_\ell^{(2)} \geq \max_{i\neq\ell}\mathsf{S}_i^{(2)}\right] = \mathbb{P}\left[\mathsf{Z}_\ell - \mu_\ell \geq \max_{m\neq\ell}\mathsf{Z}_m - \mu_m\right]. \tag{B.52}$$

We will show that for any ℓ such that $\mu_\ell > 0$, we have,

$$\mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m\neq\ell}\left\{\mathsf{Z}_m + \mu_m - \mu_\ell\right\}\right] > \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m\neq\ell}\left\{\mathsf{Z}_m - \mu_m + \mu_\ell\right\}\right], \tag{B.53}$$

which in turn implies that $\mathbb{P}\{\hat{\mathsf{R}}^{(1)} = \ell\} > \mathbb{P}\{\hat{\mathsf{R}}^{(2)} = \ell\}$.

By definition, since Z is a cyclo-stationary random process, its cumulative distribution function F_{Z} is invariant under cyclic shifts, i.e.,

$$F_{\mathsf{Z}}\left(z_0, z_1, \ldots, z_{d-1}\right) = F_{\mathsf{Z}}\left(z_\tau, z_{\tau+1}, \ldots, z_{\tau+d-1}\right), \tag{B.54}$$

for any $\tau \in \mathbb{Z}$, where the indices are taken modulo d. Furthermore, the time indices can be reverted and the distribution will remain the same, namely,

$$F_{\mathsf{Z}}\left(z_0, z_1, \ldots, z_{\ell-1}, z_\ell, z_{\ell+1}, \ldots, z_{d-1}\right) = F_{\mathsf{Z}}\left(z_{d-1}, z_{d-2}, \ldots, z_{\ell+1}, z_\ell, z_{\ell-1}, ..., z_0\right). \tag{B.55}$$

Combining (B.54) and (B.55) yields,

$$F_{\mathsf{Z}}\left(z_\ell, z_{\ell+1}, z_{\ell+2}, \ldots, z_{\ell-2}, z_{\ell-1}\right) = F_{\mathsf{Z}}\left(z_\ell, z_{\ell-1}, z_{\ell-2}, \ldots, z_{\ell+2}, z_{\ell+1}\right). \tag{B.56}$$

Accordingly, let us define the Gaussian vectors $\mathsf{Z}^{(1)}$ and $\mathsf{Z}^{(2)}$, such that their m-th entry is,

$$[\mathsf{Z}^{(1)}]_m = \mathsf{Z}_{\ell+m} - \mathsf{Z}_\ell, \tag{B.57}$$

$$[\mathsf{Z}^{(2)}]_m = \mathsf{Z}_{\ell-m} - \mathsf{Z}_\ell, \tag{B.58}$$

for $1 \leq m \leq d-1$. It is clear from (B.56) that $\mathsf{Z}^{(1)}$ and $\mathsf{Z}^{(2)}$ have the same cumulative distribution function, i.e.,

$$F_{\mathsf{Z}^{(1)}} = F_{\mathsf{Z}^{(2)}}. \tag{B.59}$$

Therefore, the following holds,

$$\begin{aligned}
&\mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m\neq 0}\left\{\mathsf{Z}_{\ell+m} + \mu_{\ell+m} - \mu_\ell\right\}\right] \\
&\qquad = \mathbb{P}\left[0 \geq \max_{m\neq 0}\left\{\mathsf{Z}_{\ell+m} - \mathsf{Z}_\ell + \mu_{\ell+m} - \mu_\ell\right\}\right] \\
&\qquad = \mathbb{P}\left[0 \geq \max_{m\neq 0}\left\{\mathsf{Z}_{\ell-m} - \mathsf{Z}_\ell + \mu_{\ell+m} - \mu_\ell\right\}\right] \\
&\qquad = \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m\neq 0}\left\{\mathsf{Z}_{\ell-m} + \mu_{\ell+m} - \mu_\ell\right\}\right],
\end{aligned} \tag{B.60}$$

where the second equality follows from (B.59). Next, we note that for every $0 < m \leq d-1$ and $\mu_\ell > 0$,

$$\mu_{\ell-m} + \mu_{\ell+m} = 2\mu_\ell \cos\left(\frac{2\pi k}{d} m\right). \tag{B.61}$$

Therefore,

$$\mu_\ell - \mu_{\ell-m} + \mu_\ell - \mu_{\ell+m} = 2\mu_\ell \left(1 - \cos\left(\frac{2\pi k}{d} m\right)\right) \geq 0, \tag{B.62}$$

which implies

$$\mu_\ell - \mu_{\ell-m} \geq \mu_{\ell+m} - \mu_\ell, \tag{B.63}$$

or, equivalently,

$$\mu_\ell - \mu_{\ell+m} \geq \mu_{\ell-m} - \mu_\ell. \tag{B.64}$$

Note that the above inequality is strict if $\cos\left(\frac{2\pi k}{d} m\right) \neq 0$, which implies that most of the inequalities are strict (at least $d/2$ of the inequalities for $0 \leq m \leq d-1$). Thus, using (B.63) and (B.64), we get for $\mu_\ell > 0$ that,

$$\max_{m \neq 0} \left\{ \mathsf{Z}_{\ell-m} - \mu_{\ell+m} + \mu_\ell \right\} > \max_{m \neq 0} \left\{ \mathsf{Z}_{\ell-m} + \mu_{\ell-m} - \mu_\ell \right\}, \tag{B.65}$$

and

$$\max_{m \neq 0} \left\{ \mathsf{Z}_{\ell-m} - \mu_{\ell-m} + \mu_\ell \right\} > \max_{m \neq 0} \left\{ \mathsf{Z}_{\ell-m} + \mu_{\ell+m} - \mu_\ell \right\}, \tag{B.66}$$

where the strict inequality arises because there are at least $d/2$ arguments from the l.h.s. that are greater than their corresponding counterparts on the r.h.s. Therefore, a direct consequence of (B.66) is that since $\mathsf{rank}(\Sigma) \geq d/2$, and thus,

$$\mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq 0} \left\{ \mathsf{Z}_{\ell-m} + \mu_{\ell+m} - \mu_\ell \right\}\right] > \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq 0} \left[\mathsf{Z}_{\ell-m} - \mu_{\ell-m} + \mu_\ell \right\}\right]. \tag{B.67}$$

Combining (B.60) and (B.67) leads to,

$$\begin{aligned} \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq 0} \left\{ \mathsf{Z}_{\ell+m} + \mu_{\ell+m} - \mu_\ell \right\}\right] &= \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq 0} \left\{ \mathsf{Z}_{\ell-m} + \mu_{\ell+m} - \mu_\ell \right\}\right] \\ &> \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq 0} \left\{ \mathsf{Z}_{\ell-m} - \mu_{\ell-m} + \mu_\ell \right\}\right], \end{aligned} \tag{B.68}$$

or, equivalently,

$$\mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq \ell} \left\{ \mathsf{Z}_m + \mu_m - \mu_\ell \right\}\right] > \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq \ell} \left\{ \mathsf{Z}_m - \mu_m + \mu_\ell \right\}\right]. \tag{B.69}$$

A similar result can be obtained for the case where $\mu_\ell < 0$, i.e.,

$$\mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq \ell} \left\{ \mathsf{Z}_m + \mu_m - \mu_\ell \right\}\right] < \mathbb{P}\left[\mathsf{Z}_\ell \geq \max_{m \neq \ell} \left\{ \mathsf{Z}_m - \mu_m + \mu_\ell \right\}\right], \tag{B.70}$$

which completes the proofs of (B.44)–(B.45).

Finally, we prove (B.46). By definition, it is clear that

$$\mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}^{(1)}+\varphi\right)\right]+\mathbb{E}\left[\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}^{(2)}+\varphi+\pi\right)\right]$$
$$=\sum_{\ell=0}^{d-1}\cos\left(\frac{2\pi k}{d}\ell+\varphi\right)\left[\mathbb{P}\left(\hat{\mathsf{R}}^{(1)}=\ell\right)-\mathbb{P}\left(\hat{\mathsf{R}}^{(2)}=\ell\right)\right], \tag{B.71}$$

where we have used the fact that $\cos(\alpha+\pi)=-\cos\alpha$, for any $\alpha\in\mathbb{R}$. By Lemma B.5, for any $0\leq\ell\leq d-1$ such that $\mu_\ell=\cos\left(\frac{2\pi k}{d}r+\varphi\right)>0$ it holds that $\mathbb{P}(\hat{\mathsf{R}}^{(1)}=\ell)>\mathbb{P}(\hat{\mathsf{R}}^{(2)}=\ell)$, otherwise, for $0\leq\ell\leq d-1$ such that $\mu_\ell=\cos\left(\frac{2\pi k}{d}r+\varphi\right)<0$, it holds that $\mathbb{P}(\hat{\mathsf{R}}^{(1)}=\ell)<\mathbb{P}(\hat{\mathsf{R}}^{(2)}=\ell)$. Therefore,

$$\sum_{\ell=0}^{d-1}\cos\left(\frac{2\pi k}{d}\ell+\varphi\right)\left[\mathbb{P}\left(\hat{\mathsf{R}}^{(1)}=\ell\right)-\mathbb{P}\left(\hat{\mathsf{R}}^{(2)}=\ell\right)\right]>0, \tag{B.72}$$

which in light of (B.71) concludes the proof.

□

Appendix C Proof of Theorem 4.3

Remark on notation. In this section, we omit the dependence on $0\leq i\leq M-1$, where this is clear from the context, i.e., $\mathsf{N}_i=\mathsf{N}$ and $\hat{\mathsf{R}}_i=\hat{\mathsf{R}}$.

The proof of Theorem 4.3 relies on the following central result. To state the result, it is convenient to define the functions,

$$f_1(r)\triangleq|\mathsf{N}[k]|\cos\left(\frac{2\pi k}{d}r+\phi_{\mathsf{N}}[k]-\phi_{\mathsf{X}}[k]\right), \tag{C.1}$$

and

$$f_2(r)\triangleq|\mathsf{N}[k]|^2\sin^2\left(\frac{2\pi k}{d}r+\phi_{\mathsf{N}}[k]-\phi_{\mathsf{X}}[k]\right), \tag{C.2}$$

for $0\leq r\leq d-1$. Note that f_1 and f_2 correspond to the denominator and nominator in (B.3), respectively.

Lemma C.1. *Assume the template signal x satisfies Assumption 4.2, and that $\mathsf{X}[k]\neq 0$, for all $0<k\leq d-1$. Then, the expected value of the functions $f_1(\hat{\mathsf{R}})$ and $f_2(\hat{\mathsf{R}})$, as defined in (C.1) and (C.2), converge for asymptotic d to:*

$$\lim_{d\to\infty}\frac{1}{2a_d\cdot\sigma^2}\frac{\mathbb{E}[f_1(\hat{\mathsf{R}})]}{|\mathsf{X}[k]|}=1, \tag{C.3}$$

and

$$\lim_{d\to\infty}\frac{1}{\sigma^2}\mathbb{E}[f_2(\hat{\mathsf{R}})]=1, \tag{C.4}$$

where $a_d\triangleq\sqrt{2\log d}$.

The proof of Lemma C.1 relies on the following proposition.

Proposition C.2. *Let $\mathsf{S} \sim \mathcal{N}(\mu, \Sigma)$ be a d-dimensional Gaussian random vector, with mean μ and a covariance matrix Σ. Assume that $|\Sigma_{ij}| = \rho_{|i-j|}$, where $\{\rho_\ell\}_{\ell \in \mathbb{N}}$ is a sequence of real-valued numbers such that $\rho_0 = 1$, $\rho_d < 1$, and $\rho_\ell \log \ell \to 0$, as $\ell \to \infty$. Assume also that $\sqrt{\log d} \cdot \max_{1 \le i \le d} |\mu_i| \to 0$, as $d \to \infty$, and let $\hat{\mathsf{R}} \triangleq \arg\max \{\mathsf{S}_0, \mathsf{S}_1, \ldots, \mathsf{S}_{d-1}\}$. Then, for a bounded deterministic function $f : \{0, 1, \ldots, d-1\} \to \mathbb{R}$, we have,*

$$\lim_{d \to \infty} \mathbb{E}[f(\hat{\mathsf{R}})] - \frac{\sum_{r=0}^{d-1} f(r) e^{\mu_r a_d}}{\sum_{r=0}^{d-1} e^{\mu_r a_d}} = 0, \tag{C.5}$$

where $a_d \triangleq \sqrt{2 \log d}$.

We prove the above results in Section C.2. Before that, let us show that it implies Theorem 4.3.

Proof of Theorem 4.3. The signal x, which satisfies the conditions of Theorem 4.3, also satisfies the conditions of Lemma C.1. By definition, the constant C_k in (B.3) can be rewritten as,

$$a_d^2 \cdot C_k = a_d^2 \cdot \frac{\mathbb{E}[|\mathsf{N}[k]| \sin(\phi_e[k])]^2}{(\mathbb{E}[|\mathsf{N}[k]| \cos(\phi_e[k])])^2} = a_d^2 \cdot \frac{\mathbb{E}[f_2(\hat{\mathsf{R}})]}{\mathbb{E}^2[f_1(\hat{\mathsf{R}})]}. \tag{C.6}$$

Accordingly, Lemma C.1 and (C.6) imply that,

$$\lim_{d \to \infty} \sigma^2 \cdot a_d^2 \cdot |\mathsf{X}[k]|^2 \cdot C_k = \frac{1}{4}. \tag{C.7}$$

Thus, to prove (4.6) it is left to show that $\sigma^2 \to 1/2$, as $d \to \infty$. By Assumption 4.2, and the definition of the eigenvalues of the covariance matrix $\Sigma_k[r, s]$ (see, (A.5)), it follows that,

$$\lim_{d \to \infty} \sum_{l=0}^{d-1} \left| 2 |\mathsf{X}[l]|^2 - \left| \tilde{\mathsf{X}}_k[l] \right|^2 \right| = 0. \tag{C.8}$$

Therefore, since we assume that the template signal is normalized, i.e., $\sum_{l=0}^{d-1} |\mathsf{X}[l]|^2 = 1$, it follows from (A.6) and (C.8) that,

$$\lim_{d \to \infty} \sigma^2 = \lim_{d \to \infty} \left[\sum_{l=0}^{d-1} |\tilde{\mathsf{X}}_k[l]|^2 \right]^{-1} = \lim_{d \to \infty} \left[\sum_{l=0}^{d-1} 2 |\mathsf{X}_k[l]|^2 \right]^{-1} = \frac{1}{2}, \tag{C.9}$$

as claimed. Combining the last result with (C.7), we obtain that,

$$\lim_{d \to \infty} a_d^2 \cdot |\mathsf{X}[k]|^2 \cdot C_k = \frac{1}{2}, \tag{C.10}$$

which proves (4.6).

Finally, we prove (4.7). To that end, we apply Theorem 4.1 and (4.3), and notice that,

$$\frac{1}{2 a_d \cdot \sigma^2} \frac{|\hat{\mathsf{X}}[k]|}{|\mathsf{X}[k]|} \xrightarrow{\text{a.s.}} \frac{1}{2 a_d \cdot \sigma^2} \frac{\mathbb{E}[f_1(\hat{\mathsf{R}})]}{|\mathsf{X}[k]|} \to 1, \tag{C.11}$$

as $M, d \to \infty$, where the second passage follows from (C.3). Then, (4.7) follows by combining the last result with the fact that $\sigma^2 \to 1/2$, as $d \to \infty$, which concludes the proof. □

C.1 Proof of Lemma C.1

Our goal is to prove (C.3) and (C.4). By the law of total expectation, we have,

$$\frac{1}{a_d}\mathbb{E}[f_1(\hat{\mathsf{R}})] = \frac{1}{a_d}\mathbb{E}\left[\mathbb{E}\left[f_1(\hat{\mathsf{R}})\middle|\,\mathsf{N}[k]\right]\right], \tag{C.12}$$

and

$$\mathbb{E}[f_2(\hat{\mathsf{R}})] = \mathbb{E}\left[\mathbb{E}\left[f_2(\hat{\mathsf{R}})\middle|\,\mathsf{N}[k]\right]\right]. \tag{C.13}$$

Accordingly, we will prove

$$\frac{1}{a_d \cdot |\mathsf{X}[k]|}\mathbb{E}\left[f_1(\hat{\mathsf{R}})\middle|\,\mathsf{N}[k]\right] \xrightarrow{\mathcal{L}^1} |\mathsf{N}[k]|^2, \tag{C.14}$$

and,

$$\mathbb{E}\left[f_2(\hat{\mathsf{R}})\middle|\,\mathsf{N}[k]\right] \xrightarrow{\mathcal{L}^1} \frac{1}{2}|\mathsf{N}[k]|^2, \tag{C.15}$$

which yields the desired result. To that end, we will use Proposition C.2, which is proved in Section C.2.

Recall our definition for the vector S in (A.2). When conditioned on $\mathsf{N}[k]$, this random vector is Gaussian with mean $\mu_k[r]$, defined in (A.3), and a covariance matrix $\mathsf{\Sigma}_k[r,s]$, defined in (A.4). Recall that we assume that these mean vectors and covariance matrices satisfy Assumption 4.2, and we claim that these satisfy the conditions of Proposition C.2 as well. Indeed, note that:

1. By the definition of the covariance matrix of $\mathsf{S}|\mathsf{N}[k]$ in (A.4), which is circulant and entirely defined by the eigenvalues given by $\tilde{\mathsf{X}}_k[l]$ (Eq. (A.5)) for $0 \le l \le d-1$, we have $\rho_m = \mathsf{R}_{\mathsf{XX}}[m]$, where R_{XX} is the Fourier transform of the PSD of x (or equivalently, the auto-correlation of x), and ρ_m is defined in Proposition C.2. By Assumption 4.2, $\mathsf{R}_{\mathsf{XX}}[d] \cdot \log d \to 0$, for asymptotic d; thus, satisfying the requirement of the decay rate of the covariance entries. One should note that the eigenvalues of the covariance matrix of $\mathsf{S}|\mathsf{N}[k]$ are $|\tilde{\mathsf{X}}_k[l]|^2$ and not $|\mathsf{X}[l]|^2$, yet also $\tilde{\mathsf{X}}_k[l]$ satisfies the conditions of Proposition C.2. Also,

$$\lim_{d\to\infty}\sum_{l=0}^{d-1}\left|2\,|\mathsf{X}[l]|^2 - |\tilde{\mathsf{X}}_k[l]|^2\right| = 0, \tag{C.16}$$

   which follows from the definition of $\tilde{\mathsf{X}}_k[l]$ in (A.5). Therefore, $\tilde{\mathsf{X}}_k[l]$ satisfies the conditions of Lemma C.3.

2. By Assumption 4.2, we have $|\mathsf{X}[k]|\sqrt{\log d} \to 0$, as $d \to \infty$, for every $0 \le k \le d-1$, implying that $\sqrt{\log d}\max|\mu_k[r]| \to 0$, where the $|\mathsf{N}[k]|$ term in $\mu_k[r]$ is finite and independent of d.

We invoke Proposition C.2 and (C.5) on $f_1(\hat{\mathsf{R}})$ and $f_2(\hat{\mathsf{R}})$, conditioned on $\mathsf{N}[k]$. Since the conditions of Proposition C.2 are satisfied, it implies that,

$$\mathbb{E}\left[f_1(\hat{\mathsf{R}}) \middle| \mathsf{N}[k] \right] - \frac{\sum_{r=0}^{d-1} f_1(r) e^{\mu_k[r] a_d}}{\sum_{r=0}^{d-1} e^{\mu_k[r] a_d}} \xrightarrow{\text{a.s.}} 0, \tag{C.17}$$

and

$$\mathbb{E}\left[f_2(\hat{\mathsf{R}}) \middle| \mathsf{N}[k] \right] - \frac{\sum_{r=0}^{d-1} f_2(r) e^{\mu_k[r] a_d}}{\sum_{r=0}^{d-1} e^{\mu_k[r] a_d}} \xrightarrow{\text{a.s.}} 0, \tag{C.18}$$

where $\mu_k[r] = \frac{f_1(r)}{2|\mathsf{X}[k]|}$. Next, we evaluate the terms at the left-hand-side of (C.17) and (C.18). Specifically, we first prove that

$$\frac{1}{2a_d} \frac{\sum_{r=0}^{d-1} \mu_k[r] \exp\{\mu_k[r] a_d\}}{\sum_{r=0}^{d-1} \exp\{\mu_k[r] a_d\}} \frac{1}{|\mathsf{X}[k]|^2} \xrightarrow{\text{a.s.}} |\mathsf{N}[k]|^2 . \tag{C.19}$$

The definition of $f_1(r)$ implies that,

$$\sum_{r=0}^{d-1} f_1(r) = 0, \tag{C.20}$$

almost surely. By definition,

$$\mu_k[r] a_d = 2a_d \, |\mathsf{X}[k]| \, |\mathsf{N}[k]| \cos\left(\frac{2\pi k}{d} r + \phi_{\mathsf{N}}[k] - \phi_{\mathsf{X}}[k] \right). \tag{C.21}$$

Now, by Assumption 4.2, we have $a_d \, |\mathsf{X}[k]| \to 0$ as $d \to \infty$. Then, (C.21) and the continuous mapping theorem it imply that,

$$\mu_k[r] a_d \xrightarrow{\text{a.s.}} 0. \tag{C.22}$$

Therefore, using the fact that $\sum_{r=0}^{d-1} \mu_k[r] = 0$, we deduce from the continuous mapping theorem and (C.22) that,

$$\frac{\sum_{r=0}^{d-1} \mu_k[r] \exp\{\mu_k[r] a_d\}}{a_d \sum_{r=0}^{d-1} [\mu_k[r]]^2} \xrightarrow{\text{a.s.}} 1. \tag{C.23}$$

Similarly, applying the continuous mapping theorem we also have,

$$\frac{\sum_{r=0}^{d-1} \exp\{\mu_k[r] a_d\}}{d} \xrightarrow{\text{a.s.}} 1. \tag{C.24}$$

Now, by definition,

$$\sum_{r=0}^{d-1} [\mu_k[r]]^2 = 4 \left(|\mathsf{X}[k]| \, |\mathsf{N}[k]| \right)^2 \sum_{r=0}^{d-1} \cos^2\left(\frac{2\pi k}{d} r + \phi_{\mathsf{N}}[k] - \phi_{\mathsf{X}}[k] \right), \tag{C.25}$$

and as so,

$$\frac{1}{d}\sum_{r=0}^{d-1}\cos^2\left(\frac{2\pi k}{d}r+\phi_{\mathsf{N}}[k]-\phi_{\mathsf{X}}[k]\right)\xrightarrow{\text{a.s.}}\frac{1}{2}, \tag{C.26}$$

as $d\to\infty$. Therefore, combining (C.23)–(C.26), we get,

$$\frac{1}{2a_d}\frac{\mathbb{E}\left[\mu_k(\hat{\mathsf{R}})\middle|\,\mathsf{N}[k]\right]}{|\mathsf{X}[k]|^2}=\frac{1}{2a_d}\frac{\sum_{r=0}^{d-1}[\mu_k[r]]\exp\{[\mu_k[r]]a_d\}}{\sum_{r=0}^{d-1}\exp\{[\mu_k[r]]a_d\}}\frac{1}{|\mathsf{X}[k]|^2}\xrightarrow{\text{a.s.}}|\mathsf{N}[k]|^2, \tag{C.27}$$

which proves (C.19).

Let us now deduce (C.3). Note that,

$$\max_{0\le r\le d-1}|\mu_k[r]|=2\,|\mathsf{X}[k]|\,|\mathsf{N}[k]|, \tag{C.28}$$

and

$$f_1(\hat{\mathsf{R}})=|\mathsf{N}[k]|\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}+\phi_{\mathsf{N}}[k]-\phi_{\mathsf{X}}[k]\right)=\frac{\mu_k(\hat{\mathsf{R}})}{2\,|\mathsf{X}[k]|}. \tag{C.29}$$

Therefore, substituting (C.19), (C.28), and (C.29), in (C.17), yields,

$$\frac{1}{a_d\cdot|\mathsf{X}[k]|}\mathbb{E}\left[|\mathsf{N}[k]|\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}+\phi_{\mathsf{N}}[k]-\phi_{\mathsf{X}}[k]\right)\middle|\,\mathsf{N}[k]\right]=\frac{1}{a_d}\frac{\mathbb{E}[\mu_k(\hat{\mathsf{R}})\big|\,\mathsf{N}[k]]}{2\,|\mathsf{X}[k]|^2}\xrightarrow{\text{a.s.}}|\mathsf{N}[k]|^2. \tag{C.30}$$

Denote the term at the left-hand-side of (C.27) by G_d, and the right-hand-side by G, and so $G_d\xrightarrow{\text{a.s.}}G$. By definition, note that $|G_d|\le|\mathsf{N}[k]|^2$, and it is clear that $\mathbb{E}\,|\mathsf{N}[k]|^2<\infty$. Therefore, by the dominated convergence theorem, we have $G_d\xrightarrow{\mathcal{L}^1}G$, and in particular,

$$\mathbb{E}\left[\frac{1}{2a_d}\frac{\mathbb{E}[\mu(\hat{\mathsf{R}})|\mathsf{N}[k]]}{|\mathsf{X}[k]|^2}\right]\to\mathbb{E}\left[|\mathsf{N}[k]|^2\right]. \tag{C.31}$$

Thus, by the law of total expectation, we obtain,

$$\frac{1}{2a_d\cdot|\mathsf{X}[k]|\cdot\sigma^2}\mathbb{E}\left[|\mathsf{N}[k]|\cos\left(\frac{2\pi k}{d}\hat{\mathsf{R}}+\phi_{\mathsf{N}}[k]-\phi_{\mathsf{X}}[k]\right)\right]\to\frac{1}{\sigma^2}\mathbb{E}\left[|\mathsf{N}[k]|^2\right]=2, \tag{C.32}$$

as $d\to\infty$, which proves (C.3).

Next, we prove (C.4). Since $a_d\cdot\max_{0\le r\le d-1}|\mu_k[r]|\xrightarrow{\text{a.s.}}0$, the numerator in (B.2) converges to,

$$\mathbb{E}[|\mathsf{N}[k]|\sin(\phi_e[k])]^2=\frac{1}{2}\mathbb{E}\,|\mathsf{N}[k]|^2-\mathbb{E}[|\mathsf{N}[k]|^2\cos(2\phi_e[k])]\to\frac{1}{2}\mathbb{E}\,|\mathsf{N}[k]|^2, \tag{C.33}$$

as $d\to\infty$, where the last transition is because $\mathbb{E}[\cos(2\phi_e[k])|\mathsf{N}[k]]\xrightarrow{\text{a.s.}}0$. Thus,

$$\frac{1}{\sigma^2}\mathbb{E}[|\mathsf{N}[k]|\sin(\phi_e[k])]^2\to\frac{1}{2\sigma^2}\mathbb{E}\,|\mathsf{N}[k]|^2=1, \tag{C.34}$$

which concludes the proof.

C.2 Proof of Proposition C.2

The proof of Proposition C.2 relies on an auxiliary result, which we prove in Section C.3. To state this result, we introduce some additional notations. Let $\mathsf{S}(r)$, for $r \in \{0, 1, \ldots, d-1\}$, be a discrete stochastic process. We define the function $h^{(\alpha)}(r)$ as follows,

$$h^{(\alpha)}(r) \triangleq \mathsf{S}(r) + \alpha f(r), \tag{C.35}$$

where $f(r)$ is a bounded deterministic function, and $\alpha \in \mathbb{R}$. We further define,

$$\mathsf{M}_d(\alpha) \triangleq \max_r h^{(\alpha)}(r), \tag{C.36}$$

and

$$\hat{\mathsf{R}}(\alpha) \triangleq \arg\max_r h^{(\alpha)}(r). \tag{C.37}$$

Note that $\mathsf{M}_d(a)$ and $\hat{\mathsf{R}}(a)$ are random variables. Finally, we denote $\hat{\mathsf{R}} \triangleq \hat{\mathsf{R}}(0)$. We have the following result.

Lemma C.3. *The following holds,*

$$\mathbb{E}[f(\hat{\mathsf{R}})] = \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbb{E}[\mathsf{M}_d(\alpha)]\bigg|_{\alpha=0}. \tag{C.38}$$

Lemma C.3 implies that finding the expected value of $f(\hat{\mathsf{R}})$ is related directly to the derivative of the expected value of the maximum around zero. Thus, the problem of finding the expected value of $f(\hat{\mathsf{R}})$ is related to finding the expected value of the maximum of the stochastic process. In our case, S will be a Gaussian vector with mean given by (A.3) and a covariance matrix given by (A.4). Thus, our goal now is to find the expected value of the maximum of S. For this purpose, we will recall some well-known results on the maximum of Gaussian processes.

It is known that for an i.i.d. sequence of normally distributed random variables $\{\xi_n\}$, the asymptotic distribution of the maximum $\mathsf{M}_n \triangleq \max\{\xi_1, \xi_2, ..., \xi_n\}$ is the Gumbel distribution, i.e., for any $x \in \mathbb{R}$,

$$\mathbb{P}\left[a_n(\mathsf{M}_n - b_n) \leq x\right] \to e^{-e^{(-x)}}, \tag{C.39}$$

as $n \to \infty$, where,

$$a_n \triangleq \sqrt{2\log n} \tag{C.40}$$

and,

$$b_n \triangleq \sqrt{2\log n} - \frac{1}{2}\frac{\log\log n + \log 4\pi}{\sqrt{2\log n}}. \tag{C.41}$$

It turns out that the above convergence result remains valid even if the sequence $\{\xi_n\}$ is not independent and normally distributed. Specifically, as shown in [23, Theorem 6.2.1], a similar result holds for Gaussian random variables $\{\xi_n\}$ with a covariance matrix that decays such that $\lim_{n\to\infty} \rho_n \cdot \log n = 0$, and with a mean vector whose maximum value decays faster than $\lim_{n\to\infty} \max_{0\leq m\leq n-1} |\mu_m| \cdot \sqrt{\log n} = 0$. These conditions precisely match those specified in Theorem 4.3.

Proof of Proposition C.2. Conditioned on $\mathsf{N}[k]$, the Gaussian vector S (see, (A.3) and (A.4)) can be represented as,

$$\mathsf{S}|\mathsf{N}[k] = \mathsf{Z} + \mu, \tag{C.42}$$

where Z is a zero mean Gaussian random vector with covariance matrix given by (A.4) and μ is given by (A.3). Define,

$$h^{(\alpha)}(r) \triangleq \mathsf{Z}(r) + \mu(r) + \alpha f(r), \tag{C.43}$$

where we use the same notations as in Lemma C.3. Then, using Lemma C.3,

$$\mathbb{E}[f(\hat{\mathsf{R}})] = \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbb{E}\mathsf{M}_d(\alpha)\bigg|_{\alpha=0}, \tag{C.44}$$

where $\mathsf{M}_d(\alpha) = \max_r \{\mathsf{Z}(r) + \mu(r) + \alpha f(r)\}$. Therefore, our goal is now to find the derivative of $\mathbb{E}\mathsf{M}_d(\alpha)$.

Using [23, Theorem 6.2.1], under the assumptions of Proposition C.2, for a sufficiently small value of α such that $\lim_{d\to\infty} |\alpha| \max_r |f(r)| \cdot \sqrt{\log d} = 0$, we have for any $x \geq 0$,

$$\lim_{d\to\infty} \mathbb{P}\left[a_d(\mathsf{M}_d(\alpha) - b_d - m_d^\star(\alpha)) \leq x\right] = e^{-e^{(-x)}}, \tag{C.45}$$

where a_d and b_d are given in (C.40) and (C.41), respectively, and

$$m_d^\star(\alpha) \triangleq a_d^{-1} \log\left(d^{-1}\sum_{i=0}^{d-1} e^{a_d(\mu_i + \alpha f(i))}\right). \tag{C.46}$$

For brevity, we denote,

$$\mathsf{T}_d(\alpha) \triangleq a_d \cdot [\mathsf{M}_d(\alpha) - b_d - m_d^\star(\alpha)], \tag{C.47}$$

and we note that,

$$\mathsf{T}_d(\alpha) - \mathsf{T}_d(0) = a_d[(\mathsf{M}_d(\alpha) - m_d^\star(\alpha)) - (\mathsf{M}_d(0) - m_d^\star(0))], \tag{C.48}$$

and so,

$$\Delta_d(\alpha) \triangleq \frac{1}{a_d}\frac{\mathsf{T}_d(\alpha) - \mathsf{T}_d(0)}{\alpha} = \frac{\mathsf{M}_d(\alpha) - \mathsf{M}_d(0)}{\alpha} - \frac{m_d^\star(\alpha) - m_d^\star(0)}{\alpha}, \tag{C.49}$$

for any $\alpha \neq 0$. The following result shows $\Delta_d(\alpha)$ converges zero in the $\mathcal{L}^1$ sense.

Lemma C.4. *For any $\alpha \neq 0$,*

$$\lim_{d\to\infty} |\Delta_d(\alpha)| = 0, \tag{C.50}$$

i.e., $\Delta_d(\alpha) \xrightarrow{\mathcal{L}^1} 0$, as $d \to \infty$.

Proof of Lemma C.4. To prove (C.50), we will first show that $\Delta_d(\alpha)$ converges to zero in probability. Because $\Delta_d(\alpha)$ is uniformly integrable, this is sufficient for the desired $\mathcal{L}^1$ convergence above. Specifically, recall from (C.45) that $\mathsf{T}_d(\alpha)$ converges in distribution to the Gumbel random variable Gum with location zero and unit scale, i.e., $\mathsf{T}_d(\alpha) \xrightarrow{\mathcal{D}} \mathsf{Gum}$, as $d \to \infty$. Furthermore, it is clear that $\frac{1}{a_d} = \frac{1}{\sqrt{2\log d}} \to 0$, as $d \to \infty$. Thus, Slutsky's theorem [41] implies that,

$$\frac{\mathsf{T}_d(\alpha)}{a_d} \xrightarrow{\mathcal{D}} 0. \tag{C.51}$$

It is known that convergence in distribution to a constant implies also convergence in probability to the same constant [15], and thus,

$$\frac{\mathsf{T}_d(\alpha)}{a_d} \xrightarrow{\mathcal{P}} 0. \tag{C.52}$$

Therefore, the above result together with the continuous mapping theorem [15] implies that,

$$\Delta_d(\alpha) \xrightarrow{\mathcal{P}} 0, \tag{C.53}$$

for every $\alpha \neq 0$.

Next, we show that $\Delta_d(\alpha)$ is bounded with probability one. Indeed, by the definition of $\mathsf{M}_d(\alpha)$ in (C.36), we have,

$$\left|\frac{\mathsf{M}_d(\alpha) - \mathsf{M}_d(0)}{\alpha}\right| \leq \max_{0\leq r\leq d-1} |f(r)| < \mathsf{C} < \infty, \tag{C.54}$$

for some $\mathsf{C} > 0$, where we have used the fact that f is bounded. Furthermore, note that,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} m_d^\star(\alpha) = \frac{\sum_{i=0}^{d-1} f(i)\exp\{a_d(\mu_i + \alpha f(i))}{\sum_{i=0}^{d-1}\exp\{a_d(\mu_i + \alpha f(i))\}}, \tag{C.55}$$

which is bounded because,

$$\left|\frac{\sum_{i=0}^{d-1} f(i)\exp\{a_d(\mu_i + \alpha f(i))}{\sum_{i=0}^{d-1}\exp\{a_d(\mu_i + \alpha f(i))\}}\right| \leq \max_{0\leq r\leq d-1} |f(r)| < \mathsf{C} < \infty. \tag{C.56}$$

Combining (C.49), (C.54) and (C.56), leads to,

$$|\Delta_d(\alpha)| \leq \left|\frac{\mathsf{M}_d(\alpha) - \mathsf{M}_d(0)}{\alpha}\right| + \left|\frac{m_d^\star(\alpha) - m_d^\star(0)}{\alpha}\right| \leq 2\max_{0\leq r\leq d-1} |f(r)| < \infty. \tag{C.57}$$

Now, since $\Delta_d(\alpha)$ is bounded, it is also uniformly integrable, and thus when combined with (C.53) we may conclude that,

$$\Delta_d(\alpha) \xrightarrow{\mathcal{L}^1} 0, \tag{C.58}$$

as claimed. □

We continue with the proof of Proposition C.2. First, we show that,

$$\lim_{d\to\infty}\lim_{\alpha\to 0}\mathbb{E}\left[\Delta_d(\alpha)\right] = \lim_{\alpha\to 0}\lim_{d\to\infty}\mathbb{E}\left[\Delta_d(\alpha)\right]. \tag{C.59}$$

Indeed, note that,

$$\lim_{d\to\infty}\lim_{\alpha\to 0}\mathbb{E}\left[\Delta_d(\alpha)\right] = \lim_{d\to\infty}\lim_{\alpha\to 0}\int\left[\frac{\mathsf{T}_d(\alpha)-\mathsf{T}_d(0)}{\alpha}\right]\mathrm{d}\mu, \tag{C.60}$$

where $\mathrm{d}\mu$ is the probability measure associated with T_d. From (C.57) we know that $\Delta_d(\alpha)$ is bounded. Thus, applying the dominated convergence theorem, we obtain,

$$\lim_{d\to\infty}\lim_{\alpha\to 0}\int\left[\frac{\mathsf{T}_d(\alpha)-\mathsf{T}_d(0)}{\alpha}\right]\mathrm{d}\mu = \lim_{d\to\infty}\int\lim_{\alpha\to 0}\left[\frac{\mathsf{T}_d(\alpha)-\mathsf{T}_d(0)}{\alpha}\right]\mathrm{d}\mu. \tag{C.61}$$

Since the integral at the right-hand-side of (C.61) is finite and bounded for each value of α, and for each value of d, the order of the limits can be exchanged, thus leading to (C.59). Therefore, from (C.49) and (C.59), we have,

$$\lim_{\alpha\to 0}\lim_{d\to\infty}\mathbb{E}\left[\Delta_d(\alpha)\right] = \lim_{d\to\infty}\lim_{\alpha\to 0}\left[\frac{\mathbb{E}[\mathsf{M}_d(\alpha)-\mathsf{M}_d(0)]}{\alpha} - \frac{[m_d^\star(\alpha)-m_d^\star(0)]}{\alpha}\right] \tag{C.62}$$

$$= \lim_{d\to\infty}\left[\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbb{E}\mathsf{M}_d(\alpha) - \frac{\mathrm{d}}{\mathrm{d}\alpha}m_d^\star(\alpha)\right]\Bigg|_{\alpha=0}. \tag{C.63}$$

Now, Lemma C.4 implies that the left-hand-side of (C.62) nullifies, and thus,

$$\lim_{d\to\infty}\left[\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbb{E}\mathsf{M}_d(\alpha) - \frac{\mathrm{d}}{\mathrm{d}\alpha}m_d^\star(\alpha)\right]\Bigg|_{\alpha=0} = 0. \tag{C.64}$$

Finally, combining (C.55) and (C.64), we obtain (C.5), which concludes the proof.

□

C.3 Proof of Lemma C.3

The proof technique of Lemma C.3 is similar to the technique used in [32, 25], but with a non-trivial adaption to the discrete case. To prove this lemma, we will first establish a deterministic counterpart of (C.38). Specifically, we define,

$$h^{(\alpha)}(r) \triangleq X(r) + \alpha f(r), \tag{C.65}$$

where $r \in \{0, 1, \ldots, d-1\}$. The functions $X : \{0, 1, \ldots, d-1\} \to \mathbb{R}$, and $f : \{0, 1, \ldots, d-1\} \to \mathbb{R}$ are assumed bounded and deterministic. We further assume that X is injective, i.e., for $z_i \neq z_j$, we have $X(z_i) \neq X(z_j)$. Define,

$$s(\alpha) \triangleq \max_r\{h^{(\alpha)}(r)\}, \tag{C.66}$$

and note that $s(\alpha)$ is well-defined over the supports of X and f, and it is a continuous function of α around $\alpha = 0$. Finally, we let,

$$Z_{\max}^{(\alpha)} \triangleq \arg\max_r\{h^{(\alpha)}(r)\}. \tag{C.67}$$

We have the following result.

Lemma C.5. *The following relation holds,*

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} s(\alpha)\bigg|_{\alpha=0} = f(Z_{\max}^{(0)}). \tag{C.68}$$

Proof of Lemma C.5. Note that,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha} s(\alpha)\bigg|_{\alpha=0} &= \lim_{\alpha\to 0} \frac{s(\alpha)-s(0)}{\alpha} \\ &= \lim_{\alpha\to 0} \frac{\max_r[X(r)+\alpha f(r)] - \max_r X(r)}{\alpha}.\end{aligned} \tag{C.69}$$

By the definition of $Z_{\max}^{(\alpha)}$, we have,

$$\max_r[X(r)+\alpha f(r)] = X(Z_{\max}^{(\alpha)}) + \alpha f(Z_{\max}^{(\alpha)}), \tag{C.70}$$

and

$$\max_r X(z) = X(Z_{\max}^{(0)}). \tag{C.71}$$

Now, the main observation here is that for a sufficiently small value of α around zero, we must have that $Z_{\max}^{(\alpha)}$ and $Z_{\max}^{(0)}$ equal because $Z_{\max}^{(0)}$ can take discrete values only, and it is unique. Thus, for $\alpha \cdot \max_r |f(r)| < \min_{i\neq j} |X(z_i) - X(z_j)|$, we have,

$$Z_{\max}^{(\alpha)} = Z_{\max}^{(0)}. \tag{C.72}$$

Combining (C.69)–(C.72) yields,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} s(\alpha)\bigg|_{\alpha=0} = \lim_{\alpha\to 0} \frac{X(Z_{\max}^{(\alpha)}) + \alpha f(Z_{\max}^{(\alpha)}) - X(Z_{\max}^{(0)})}{\alpha} \tag{C.73}$$

$$= f(Z_{\max}^{(0)}), \tag{C.74}$$

which concludes the proof. □

We are now in a position to prove Lemma C.3. Similarly to the deterministic case, we define the random function,

$$h^{(\alpha)}(r) = \mathsf{S}(r) + \alpha f(r), \tag{C.75}$$

where $\mathsf{S} : \{0, 1, \ldots, d-1\} \to \mathbb{R}$ is a discrete stochastic process, and f is a deterministic function. We assume that S has a continuous probability distribution without any single point with a measure greater than 0. Using Lemma C.5, for each realization of $\mathsf{S}(r)$, such that $\mathsf{S}(r)$ is injective, we have,

$$f(\hat{\mathsf{R}}) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\, \mathsf{M}_d(\alpha)\bigg|_{\alpha=0}. \tag{C.76}$$

Under the assumption above of $\mathsf{S}(r)$, the measure of the set of events that S is not injective is zero. Therefore, the fact that $\frac{\mathsf{M}_d(\alpha)-\mathsf{M}_d(0)}{\alpha}$ is bounded (see, (C.54)) and (C.76), imply that,

$$\begin{aligned}
\mathbb{E}[f(\hat{\mathsf{R}})] &= \int f(\hat{\mathsf{R}})\mathrm{d}\mu \\
&= \int \frac{\mathrm{d}}{\mathrm{d}\alpha}\, \mathsf{M}_d(\alpha)\Big|_{\alpha=0} \mathrm{d}\mu \\
&= \int \lim_{\alpha\to 0}\left[\frac{\mathsf{M}_d(\alpha)-\mathsf{M}_d(0)}{\alpha}\right]\mathrm{d}\mu \\
&= \lim_{\alpha\to 0}\int \left[\frac{\mathsf{M}_d(\alpha)-\mathsf{M}_d(0)}{\alpha}\right]\mathrm{d}\mu \\
&= \frac{\mathrm{d}}{\mathrm{d}\alpha}\, \mathbb{E}\mathsf{M}_d(\alpha)\Big|_{\alpha=0}, \qquad \text{(C.77)}
\end{aligned}$$

which concludes the proof.