Covariance Matrix Recovery From One-Bit Data With Non-Zero Quantization Thresholds: Algorithm and Performance Analysis

Yu-Hang Xiao , *Member, IEEE*, Lei Huang , *Senior Member, IEEE*, David Ramírez , *Senior Member, IEEE*, Cheng Qian , *Member, IEEE*, and Hing Cheung So , *Fellow, IEEE*

Abstract—Covariance matrix recovery is a topic of great significance in the field of one-bit signal processing and has numerous practical applications. Despite its importance, the conventional arcsine law with zero threshold is incapable of recovering the diagonal elements of the covariance matrix. To address this limitation, recent studies have proposed the use of non-zero clipping thresholds. However, the relationship between the estimation error and the sampling threshold is not yet known. In this article, we undertake an analysis of the mean squared error by computing the Fisher information matrix for a given threshold. Our results reveal that the optimal threshold can vary considerably, depending on the variances and correlation coefficients. As a result, it is inappropriate to adopt a constant threshold to encompass parameters that vary widely. To mitigate this issue, we present a recovery scheme that incorporates time-varying thresholds. Our approach differs from existing methods in that it utilizes the exact values of the threshold, rather than its statistical properties, to increase the estimation accuracy. Simulation results, including those of the direction-of-arrival estimation problem, demonstrate the efficacy of the developed scheme, especially in complex scenarios where the covariance elements are widely separated.

Index Terms—Covariance matrix estimation, mean squared error analysis, non-zero threshold, one-bit sampling.

I. INTRODUCTION

ONE-BIT analog-to-digital converters (ADCs) have garnered significant attention in recent years due to their unique merits over high-resolution ADCs. These advantages include cost-effectiveness, lower power consumption, and simpler hardware design. In addition, the reduced data flow associated with one-bit ADCs makes data storage and transmission more manageable. This has led to the widespread application of one-bit signal processing in various fields, such as multiple-input multiple-output communications [1], [2], [3], [4], [5], [6], array processing [7], [8], [9], [10], [11], [12], and radar [13], [14], [15], [16], [17], [18], [19].

Despite its numerous advantages, one-bit analog-to-digital conversion with zero sampling thresholds has created challenges in applications involving parameter estimation and detection. The loss of amplitude information has limited its use in areas that rely on second-order statistics, such as direction-of-arrival (DOA) estimation [20], spectrum sensing [21], [22], [23], and radar target detection [24], [25]. Therefore, the covariance matrix recovery has become a critical topic in one-bit processing research.

Under the assumption of zero-mean Gaussian inputs, the most frequently employed criterion for recovering one-bit covariance matrices is the arcsine law [26]. It can immediately translate the one-bit covariance matrix into that of the unquantized data matrix. It does, however, provide a normalized version of the covariance matrix, namely the correlation matrix,[1] rather than the original covariance matrix. That is, unless the diagonal elements of the covariance matrix are equal, the estimation is biased and inconsistent. It is because these systems adopt zero as the sampling threshold, meaning that the likelihood of the quantized signal has no bearing on the variance of the random variables. As a result, these samples cannot be used to estimate variances, i.e., the diagonal entries of the covariance matrix.

To address this issue, Liu and Lin [27] have employed a constant (non-zero) threshold to enable accurate and consistent estimates of the covariance matrix, which may be easily accomplished by adding a DC level to the input signal. With the addition of the non-zero threshold, the likelihood of the output

---

[1]This matrix contains all pairwise correlation coefficients.

being $+1$ or -1 is no longer fixed at $1/2$ but is instead a function of the ratio between the threshold and the standard deviation of the random variable. This allows the variance to be estimated.

In a parallel development, [28] and [29] use time-varying thresholds to estimate parameters of sinusoidal signals, including frequency, phase, and crucially, amplitude—a parameter not recoverable with the zero-threshold approach. Moreover, it is shown that employing time-varying thresholds can considerably increase the recovery capability of one-bit processing methodologies. For instance, [30] presents an innovative iterative hard thresholding algorithm devised for signal reconstruction from one-bit measurements. Besides, [31], [32], [33], [34], [35] proves that the use of adaptive or time-varying thresholds can mitigate the reconstruction error in one-bit compressed sensing, with [36] notably providing theoretical guarantees for these schemes in cases of a large number of samples. It has also been demonstrated that the use of time-varying sampling thresholds can increase the signal recovery accuracy in numerous one-bit signal recovery applications, including unlimited sampling [37], [38], phase retrieval [39], [40], low-rank matrix sensing [41], [42], among others. As illustrated in [41], a well-designed threshold can effectively enhance signal recovery. Moreover, the time-varying nature of the threshold allows it to be modified adaptively to align better with the recovery algorithm [43]. For instance, the work in [37] utilizes the concept of unlimited sampling to design an effective dithering scheme by minimizing the difference between the dynamic ranges of input signals and thresholds, thereby achieving better recovery performance compared to one-bit $\Sigma\Delta$ sampling. In recent works, [44], [45], [46], [47] suggest the use of a time-varying threshold for covariance matrix recovery. This approach, involving the addition of a Gaussian dithering signal to the original constant threshold, promises to deliver superior performance. Additionally, [33], [34] demonstrate that dithering enables the reconstruction of signals from non-Gaussian measurements and enhances robustness against noise. Finally, employing multiple sequences of dithering signals has been shown to further enhance performance, as evidenced in [30], [35], [37], [41], [44].

However, there is still no performance analysis conducted to derive the estimation error associated with the threshold and the population covariance matrix, making it impossible to optimize the threshold value to improve estimation performance. In addition, from a statistical sense, the approach in [45] is equivalent to modifying the population covariance matrix of the signal prior to quantizing with a constant threshold. Without such analysis, we cannot set the dithering signal properly to relocate the covariance matrix to an appropriate region.

In this article, we analyze the performance of the constant threshold estimator in [27], which is also compatible with the random threshold method in [45]. Due to the absence of closed-form estimators, it is prohibitive to define their statistical behavior using conventional methodologies. Our idea is to perform a Taylor's expansion and then apply the result to compute the mean squared error (MSE) of the estimators. It is found that a low threshold facilitates the estimation of the non-diagonal elements while diagonal ones favour thresholds comparable to their square roots. Therefore, it is inappropriate to adopt a constant threshold to deal with all elements in the covariance matrix, especially when the parameters are distinct from each other, as is typical when the dimension increases.

To address this issue, we present a novel approach based on a time-varying threshold, which differs from [44], [45], [46] since it uses the exact values of the threshold and not only its statistical properties. Exploiting Price's theorem [48], we calculate the gradient of the orthant probability with regard to the covariance matrix parameters and seek the maximum likelihood estimators (MLEs) of the parameters. The algorithm is also extended to complex-valued scenarios to accommodate array processing applications. Furthermore, we carry out performance analysis of the new method by computing the inverse of the Fisher information matrix, which allows us to predict the performance more efficiently than through Monte Carlo simulations.

Finally, simulation results are presented to demonstrate the effectiveness of our proposed approach. We consider the DOA estimation of coherent sources, which requires the reconstruction of the received signals covariance matrix, as an example. We first estimate the covariance matrix through different methods and then process the results with the Enhanced Principal-singular-vector Utilization for Modal Analysis (EPUMA) [20] algorithm to produce DOA estimates. It is shown that compared to constant and random threshold-based methods, our algorithm achieves significantly improved accuracy and stability.

The key contributions of this article are as follows:
1) We conduct a thorough performance analysis of the constant threshold approach by leveraging a Taylor's expansion to analyze the estimator, proving that it is challenging to use a constant threshold to effectively estimate parameters distributed over a wide range. This finding opens up the opportunity for optimization of the sampling threshold.
2) We introduce a new sampling strategy that utilizes time-varying thresholds and the corresponding recovery algorithm. In comparison to the existing constant and random threshold approaches, our solution offers higher estimation accuracy and demonstrates improved robustness against parameter unevenness and high correlation coefficients.
3) To further analyze the algorithm performance, we compute the Fisher information corresponding to each threshold value. Our results demonstrate that the Fisher information provides a precise performance indicator even when the likelihood function is inconsistent across different samples.
4) Finally, we extend the covariance matrix estimator to the complex-valued scenario and integrate it with the EPUMA for DOA estimation, highlighting the broad range of potential applications.

We proceed under the assumption of zero-mean Gaussian inputs, a practice that aligns with most of the research in this field. It is noteworthy that some studies do explore non-Gaussian settings. For instance, in the field of one-bit compressed sensing, [50], [51], [52], [53] focus on the recovery of sub-Gaussian signals by exploiting sparsity rather than statistical

distributions. On the other hand, in the field of covariance matrix recovery, [49] extends the arcsine law to all complex elliptically symmetric distributions. More recently, [47] has derived a sharp non-asymptotic error bound for the arcsine law estimator, and proposed a simplified reconstruction method for sub-Gaussian signals with theoretical guarantees, demonstrating that the probability of the estimation error exceeding a certain level decays exponentially. The effectiveness of this approach has further been validated in a channel estimation application [54].

Nonetheless, in our study, the spotlight remains on the Gaussian signal case, since the proposed method and analysis are based on the likelihood function. We also operate under the assumption of a large number of available samples—a common scenario in one-bit sampling [40], [41], facilitated by its simplicity that enables a high sampling rate. Moreover, our analysis diverges subtly from [47], [54] in that we scrutinize the individual MSE of the elements to be estimated, rather than the matrix norms. This focus is driven by our intent to understand how threshold settings individually influence the estimation accuracy of variance and covariance. The overall estimation error can then be evaluated by aggregating the individual MSEs.

Finally, we assume that the dimension of the covariance matrix to be recovered is significantly smaller than the number of samples. Namely, our analysis considers the case of $N \to \infty$ and $M/N \to 0$, where M and N represent the dimension of the covariance matrix and the number of samples, respectively. Worth mentioning, however, is [47] and a recent study [55] that show that the recovery of high-dimensional sparse covariance matrices under sub-Gaussian and heavy-tailed distributions is feasible. Yet, the task of a general high-dimensional one-bit covariance matrix recovery method remains an arduous challenge and is beyond the scope of this article.

Notation: Throughout this article, we use boldface uppercase letters for matrices, boldface lowercase letters for column vectors, and lightface lowercase letters for scalar quantities. The notation $\mathbf{A} \in \mathbb{R}^{p \times q}$ ($\mathbb{C}^{p \times q}$) indicates that $\mathbf{A}$ is a $p \times q$ real (complex) matrix. The operators $\mathbb{E}[a]$ and $\mathbb{V}[a]$ denote, respectively, the expectation and variance of random variable a, $\mathbb{C}[a, b]$ is the covariance between a and b, and $\sim$ means "distributed as". The superscript $\hat{a}$ denotes the estimate of a. Finally, the operators $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ extract the real and imaginary parts of their argument and $\imath = \sqrt{-1}$ is the imaginary unit.

II. PRELIMINARIES

In this section, we present the problem of one-bit covariance estimation and review existing methods based on various sampling schemes, including the zero threshold, constant threshold, and random threshold approaches.

A. Problem Formulation

Suppose $\mathbf{y} \in \mathbb{R}^{M \times 1}$ follows a zero-mean multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}})$. Assume we have N i.i.d. one-bit quantized observations of $\mathbf{y}$, $\mathbf{x}(t) = \mathrm{sign}(\mathbf{y}(t) - \mathbf{v}(t))$, $t = 1, \cdots, N$, where $\mathbf{x}(t) = [x_1(t), \cdots, x_M(t)]^T$, $\mathbf{y}(t) = [y_1(t), \cdots, y_M(t)]^T$, and $\mathbf{v}(t) = [v_1(t), \cdots, v_M(t)]^T$,

is the quantization threshold vector. The function $\mathrm{sign}(\cdot)$ is the quantization operator

$$\mathrm{sign}(x) = \begin{cases} +1, & x \geq 0, \\ -1, & x < 0. \end{cases} \quad (1)$$

Our aim is to recover the covariance matrix of the unquantized signal $\mathbf{y}$, $\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbb{E}[\mathbf{y}\mathbf{y}^T]$, given its one-bit quantized samples, i.e., $\mathbf{X} = [\mathbf{x}(1), \cdots, \mathbf{x}(N)]$. To simplify our discussion, we focus on the 2×2 case:

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \quad (2)$$

which can be easily extended to the general case in a pairwise manner.

There are various methods of setting the threshold $\mathbf{v}(t)$. Traditionally, it is fixed at $\mathbf{v}(t) = \mathbf{0}_2$, resulting in the complete loss of amplitude information and only the correlation coefficients can be obtained. In order to estimate the variance of the random variables, it is necessary to set $\mathbf{v}(t)$ to be nonzero by incorporating a control sequence at the input of the ADC. This control sequence can be a DC level [27], or taking a time-varying form, such as a sine wave [56], or a random sequence, which can be generated with thermal noise diodes [32], [36], [45].

B. Zero Threshold

When the sampling threshold is 0, the relationship between $\boldsymbol{\Sigma}_{\mathbf{x}}$ and $\boldsymbol{\Sigma}_{\mathbf{y}}$ can be described using the well-known *arcsine law* [26]:

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \frac{2}{\pi} \sin^{-1} \left(\mathbf{D}_{\mathbf{y}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\mathbf{y}} \mathbf{D}_{\mathbf{y}}^{-\frac{1}{2}} \right), \quad (3)$$

where $\mathbf{D}_{\mathbf{y}} = \mathrm{diag}(\boldsymbol{\Sigma}_{\mathbf{y}})$. Assuming that $\mathbf{D}_{\mathbf{y}}$ is the identity matrix, a natural estimator of $\boldsymbol{\Sigma}_{\mathbf{y}}$ is $\hat{\boldsymbol{\Sigma}}_{\mathbf{y}} = \sin\left(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\pi/2\right)$, where $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ is the sample covariance matrix of $\mathbf{x}$, $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/N$. In the complex-valued case, where the sampling process is modified as $\mathbf{x} = \mathcal{Q}(\mathbf{y}) = \mathrm{sign}(\mathrm{Re}(\mathbf{y}) - \mathbf{v}) + \imath\mathrm{sign}(\mathrm{Im}(\mathbf{y}) - \mathbf{v})$, the estimator is modified accordingly as

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{y}} = \sin\left(\frac{\pi}{4}\mathrm{Re}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}})\right) + \imath\sin\left(\frac{\pi}{4}\mathrm{Im}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}})\right). \quad (4)$$

Interestingly, the work [49] demonstrated that (4) holds not only for complex circular Gaussian distributions, but all complex elliptically symmetric distributions. However, a significant drawback of the arcsine law is that it is incapable of estimating the diagonal entries of $\boldsymbol{\Sigma}_{\mathbf{y}}$, as the likelihood function does not include these entries. That said, if the assumption of unit diagonal entries is violated, the arcsine law becomes biased and inconsistent.

C. Constant Threshold Approach

The use of a constant threshold has been introduced in [27] for covariance matrix recovery. The reconstruction can be accomplished based on the following probabilities:

$$p_i = \mathrm{Pr}\{x_i = +1\} = Q\left(\frac{v}{\sigma_i}\right), \quad i = 1, 2, \quad (5)$$

$$p_{12} = \Pr\{x_1 = +1, x_2 = +1\}$$
$$= \int_{\frac{v}{\sigma_1}}^{\infty} \int_{\frac{v}{\sigma_2}}^{\infty} f\left(y_1, y_2 \Big| \frac{\sigma_{12}}{\sigma_1 \sigma_2}\right) dy_1 dy_2, \qquad (6)$$

where v is the threshold, $f(y_1, y_2|\rho)$ is the probability density function of bivariate Gaussian distribution with unit variances and correlation coefficient ρ, given by

$$f(y_1, y_2|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{2(1-\rho^2)}\right), \quad (7)$$

and

$$Q(a) = \int_a^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt. \qquad (8)$$

The MLEs of the probabilities are:

$$\hat{p}_i = \frac{\sum_{t=1}^{N} [x_i(t) + 1]}{2N}, \quad i = 1, 2, \qquad (9)$$

$$\hat{p}_{12} = \frac{\sum_{t=1}^{N} [x_1(t) + 1][x_2(t) + 1]}{4N}. \qquad (10)$$

As a consequence, and using the invariance property of the MLE, the MLEs of the variances are $\hat{\sigma}_i = v/Q^{-1}(\hat{p}_i)$. On the other hand, the right hand side of (6) can be rewritten as the following infinite polynomial form:[2]

$$\hat{p}_{12} = \frac{e^{-\frac{v^2}{\hat{\sigma}_1 \hat{\sigma}_2}}}{\pi} \sum_{k=0}^{\infty} \frac{H_k\left(\frac{v}{\sqrt{2}\hat{\sigma}_1}\right) H_k\left(\frac{v}{\sqrt{2}\hat{\sigma}_2}\right)}{2^{k+1}(k+1)!} \rho^{k+1} + \frac{\mu_1 \mu_2}{4}, \qquad (11)$$

where

$$\mu_i = 2Q\left(\frac{v}{\hat{\sigma}_i}\right) - 1, \quad i = 1, 2, \qquad (12)$$

and

$$H_k(a) = (-1)^k e^{a^2} \frac{d^k}{da^k} e^{-a^2}, \qquad (13)$$

is the Hermite polynomial of order k. The correlation coefficient ρ can then be estimated numerically by solving the equation omitting higher-order terms of the polynomial. It is worth mentioning that, although no theoretical guarantee was reported in [27], simulations show that the approximation is reasonably accurate for $|\rho| < 0.6$. This will be evidenced by our numerical results in Section IV-F.

D. Random Threshold

In [44] and [45], the use of a random threshold with a Gaussian distribution $\mathcal{N}(d\mathbf{1}_M, \boldsymbol{\Sigma}_t)$ is suggested. This is equivalent to adding a zero-mean dithering signal to the constant sampling threshold $d\mathbf{1}_M$, resembling the "dithered quantization" in the traditional multi-bit scenarios [57]. Consequently, the covariance matrix is adjusted to $\boldsymbol{\Sigma}'_\mathbf{y} = \boldsymbol{\Sigma}_\mathbf{y} + \boldsymbol{\Sigma}_t$. When designed appropriately, this dithering signal has the potential to reduce estimation errors.

[2]Note that the result here is slightly modified, as opposed to the original version in [27], to cope with the non-uniform variances.

Interestingly, [44] and [45] presented a *modified arcsine law*:

$$\boldsymbol{\Sigma}_\mathbf{x}(i,j) = \frac{e^{\frac{-d^2(\sigma_i + \sigma_j - 2\sigma_{ij})}{2(\sigma_i \sigma_j - \sigma_{ij}^2)}}}{\pi\sqrt{(\sigma_i \sigma_j - \sigma_{ij}^2)}} \left\{ \int_0^{\frac{\pi}{2}} \frac{1}{\beta_n} + \sqrt{\frac{\pi}{\beta_n}} \frac{\alpha_n}{2\beta_n} e^{\frac{\alpha_n^2}{4\beta_n}} \right.$$
$$\left. - \sqrt{\frac{\pi}{\beta_n}} \frac{\alpha_n}{\beta_n} Q\left(\frac{\alpha_n}{\sqrt{2\beta_n}}\right) e^{\frac{\alpha_n^2}{4\beta_n}} d\theta \right\} - 1, \qquad (14)$$

with

$$\alpha_n = \frac{d\left(\sigma_i \sin\theta + \sigma_j \cos\theta - \sigma_{ij}(\cos\theta + \sin\theta)\right)}{\left(\sigma_i \sigma_j - \sigma_{ij}^2\right)}, \qquad (15a)$$

$$\beta_n = \frac{\sigma_j \cos^2\theta + \sigma_i \sin^2\theta - \sigma_{ij}\sin 2\theta}{2\left(\sigma_i \sigma_j - \sigma_{ij}^2\right)}, \qquad (15b)$$

and the population parameters σ_i and σ_{ij} are drawn from $\boldsymbol{\Sigma}'_\mathbf{y}$. Furthermore, it was shown in [45] that σ_{ij} could be determined through minimization of the cost function:

$$G(\sigma_{ij}) \triangleq \log\left(\left|\hat{\boldsymbol{\Sigma}}_\mathbf{x}(i,j) - H_n(\sigma_i^\star, \sigma_j^\star, \sigma_{ij})\right|^2\right). \qquad (16)$$

In this case, $H_n(\sigma_i^\star, \sigma_j^\star, \sigma_{ij})$ serves as an approximation to the aforementioned modified arcsine law. Importantly, [45] proposed three distinct approximations. The first is based on the Padé approximation and results in a non-convex optimization problem. The remaining two methods rely on the Gauss-Legendre quadrature approximation and Monte-Carlo integration, which respectively generate a convex optimization problem ensuring a global optimum.

Notably, [45] presented a modified version of the Bussgang law [58], [59], which is useful for analyzing the correlation between the quantized signal $\mathbf{x}$ and unquantized signal $\mathbf{y}$. Specifically, the cross-covariance matrix between $\mathbf{x}$ and $\mathbf{y}$ is:

$$\boldsymbol{\Sigma}_\mathbf{xy}(i,j) = \boldsymbol{\Sigma}_\mathbf{xv}(i,j) + [\varepsilon_1 \sigma_{ij} - \varepsilon_2 d(\sigma_j - \sigma_{ij})]. \qquad (17)$$

Here, $\boldsymbol{\Sigma}_\mathbf{xv}$, representing the cross-covariance matrix between $\mathbf{x}$ and the random threshold $\mathbf{v}$, can be estimated from samples. The parameters ε_1 and ε_2 are given by

$$\varepsilon_1 = \sqrt{\frac{2}{\pi\sigma_j}} \Gamma\left(1, \frac{d^2}{2\sigma_j}\right) - \frac{d}{\sqrt{\pi\sigma_j^2}} \left(\Gamma\left(\frac{1}{2}, \frac{d^2}{2\sigma_j}\right) - \sqrt{\pi}\right), \qquad (18)$$

$$\varepsilon_2 = -\frac{1}{\sigma_j} \operatorname{erf}\left(\frac{d}{\sqrt{2\sigma_j}}\right), \qquad (19)$$

where $\Gamma(\cdot)$ denotes the Gamma function. It is also worth mentioning that the above results have been extended to stationary Gaussian inputs where the covariance matrix is Toeplitz, also with convex programs containing a global solution, as detailed in [46].

In general, non-zero threshold approaches surpass the arcsine law as they allow for the full recovery of the covariance matrix. However, it remains unclear whether a constant threshold is optimal. Particularly, no performance analysis has been conducted to determine whether estimating $\boldsymbol{\Sigma}_\mathbf{y}$ or $\boldsymbol{\Sigma}'_\mathbf{y}$ provides smaller MSE, which makes it impossible to determine the

shifting matrix Σ_t. In addition, it is unknown which threshold provides optimum estimation for different diagonal and non-diagonal elements.

In this article, we first analyze the MSE of the constant threshold estimator, revealing that the optimal threshold for estimating different variances and covariances are distinct. We then present a recovery algorithm based on time-varying thresholds, where the thresholds are known deterministic values instead of random variables, as opposed to [45].

III. PERFORMANCE ANALYSIS OF CONSTANT THRESHOLD APPROACH

In this section, we analyze the MSE of the constant-threshold-based approach with regard to both variance and co-variance estimation. The analysis is conducted by applying a Taylor's expansion to the expressions of the estimators. For variance estimation, a Taylor's expansion up to the second order is applied, while for the estimation of covariances, a first-order expansion is employed due to the complexity of the estimator.

A. MSE of Diagonal Elements

The approximation is made under the assumption that N is large, which is a common scenario in one-bit systems as the sampling rate is typically very high. Furthermore, as it has been proved that the bias of MLE approaches 0 as $N \to \infty$ [60], the MSE of the detector becomes equivalent to the variance of the estimators.

Recall that the estimator for $\sigma_i (i = 1, 2)$ is

$$\hat{\sigma}_i = \frac{v}{Q^{-1}(\hat{p}_i)}. \tag{20}$$

We first compute the second-order Taylor's expansion of the estimator. For simplicity, we define

$$h(a) = \frac{v}{Q^{-1}(a)}, \tag{21}$$

and the second-order Taylor's expansion of $h(a)$ at $a = p_i$ is:

$$h(a) = h(p_i) + h'(p_i)(a - p_i)$$
$$+ \frac{1}{2} h''(p_i)(a - p_i)^2 + \mathcal{O}((a - p_i)^3), \tag{22}$$

where

$$h'(p_i) = \frac{\sqrt{2\pi}\sigma_i^2}{v} \exp\left(\frac{v^2}{2\sigma_i^2}\right), \tag{23}$$

$$h''(p_i) = \exp\left(\frac{v^2}{\sigma_i^2}\right)\left(\frac{4\pi\sigma_i^3}{v^2} - 2\pi\sigma_i\right). \tag{24}$$

Proof: See Appendix A. $\qquad\square$

According to (22), the variance of $\hat{\sigma}_i$ can be approximated as:

$$\mathbb{V}(\hat{\sigma}_i) \approx (h'(p_i) - h''(p_i)p_i)^2 \mathbb{V}(\hat{p}_i) + \frac{1}{4}[h''(p_i)]^2 \mathbb{V}(\hat{p}_i^2)$$
$$+ (h'(p_i) - h''(p_i)p_i)h''(p_i)\mathbb{C}(\hat{p}_i, \hat{p}_i^2). \tag{25}$$

Next, we calculate the terms $\mathbb{V}(\hat{p}_i)$, $\mathbb{V}(\hat{p}_i^2)$ and $\mathbb{C}(\hat{p}_i, \hat{p}_i^2)$, which requires us to first compute the second- to fourth-order moments of $\hat{p}_i$. Since $N_i = N\hat{p}_i$ follows a binomial distribution, its moments can be evaluated by the following lemma [61].

Lemma 1: The cth order moment of a binomial distributed random variable ϑ with success probability p_i and number of trials N is:

$$\mathbb{E}[\vartheta^c] = \sum_{k=0}^{c} S_k^c N^{\underline{k}} p_i^k, \tag{26}$$

where S_k^c is the Stirling number of the second kind:

$$S_k^c = \sum_{j=1}^{k} (-1)^{k-j} \frac{j^{c-1}}{(j-1)!(k-j)!}, \tag{27}$$

and $N^{\underline{k}}$ is the $k-$th falling power of N:

$$N^{\underline{k}} = N(N-1)\cdots(N-k+1). \tag{28}$$

Using Lemma 1 with $\vartheta = N\hat{p}_i$, the required moments of $\hat{p}_i$ are:

$$m_2 = \mathbb{E}[\hat{p}_i^2] = \frac{p_i + p_i^2(N-1)}{N}, \tag{29}$$

$$m_3 = \mathbb{E}[\hat{p}_i^3] = \frac{p_i + 3p_i^2(N-1) + p_i^3(N-1)(N-2)}{N^2}, \tag{30}$$

$$m_4 = \mathbb{E}[\hat{p}_i^4] = \frac{p_i + 7p_i^2(N-1) + 6p_i^3(N-1)(N-2)}{N^3}$$
$$+ \frac{p_i^4(N-1)(N-2)(N-3)}{N^3}, \tag{31}$$

where m_k denotes the $k-$th order moment of $\hat{p}_i$. Therefore, we have

$$\mathbb{V}(\hat{p}_i) = m_2 - p_i^2, \tag{32}$$
$$\mathbb{V}(\hat{p}_i^2) = m_4 - m_2^2, \tag{33}$$
$$\mathbb{C}(\hat{p}_i, \hat{p}_i^2) = m_3 - p_i m_2. \tag{34}$$

Substituting (32)–(34) into (25) results in the variance of $\hat{\sigma}_i$.

B. MSE of Non-Diagonal Elements

The analysis of the covariance estimator is more complex compared to the variance estimator as it depends not only on $\hat{p}_{ij}$, but also on the estimated variances $\hat{\sigma}_i$ and $\hat{\sigma}_j$. Therefore, a second-order analysis is not feasible and a first-order analysis is conducted instead. This involves constructing a linear approximation of $\hat{\sigma}_{ij}$, resulting in a simplified representation of its behavior. The result is summarized in the following lemma.

Lemma 2: The first-order Taylor's expansion of σ_{12} as a function of $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_{12}$ is

$$\sigma_{12} - \hat{\sigma}_{12} \approx \mathbf{b} [p_1 - \hat{p}_1, p_2 - \hat{p}_2, p_{12} - \hat{p}_{12}]^T, \tag{35}$$

where

$$\mathbf{b} = \left[-\frac{\partial \sigma_{12}}{\partial p_{12}} \frac{\partial p_{12}}{\partial \sigma_1} h'(p_1), -\frac{\partial \sigma_{12}}{\partial p_{12}} \frac{\partial p_{12}}{\partial \sigma_2} h'(p_2), \frac{\partial \sigma_{12}}{\partial p_{12}} \right], \tag{36}$$

with

$$\frac{\partial p_{12}}{\partial \sigma_{12}} = \left[\frac{\partial \sigma_{12}}{\partial p_{12}}\right]^{-1} = \frac{1}{\sigma_1 \sigma_2} f\left(\frac{v}{\sigma_1}, \frac{v}{\sigma_2}\Big|\rho\right), \tag{37}$$

$$\frac{\partial p_{12}}{\partial \sigma_1} = \frac{1}{\sigma_1} g\left(\frac{v}{\sigma_1}, \frac{v}{\sigma_2}, \rho\right), \tag{38}$$

$$\frac{\partial p_{12}}{\partial \sigma_2} = \frac{1}{\sigma_2} g\left(\frac{v}{\sigma_2}, \frac{v}{\sigma_1}, \rho\right), \tag{39}$$

a. $[\sigma_1, \sigma_2, \sigma_{12}] = [0.25, 0.6, -0.08]$ b. $[\sigma_1, \sigma_2, \sigma_{12}] = [0.9, 1.2, 0.2]$ c. $[\sigma_1, \sigma_2, \sigma_{12}] = [1.5, 2, -0.3]$
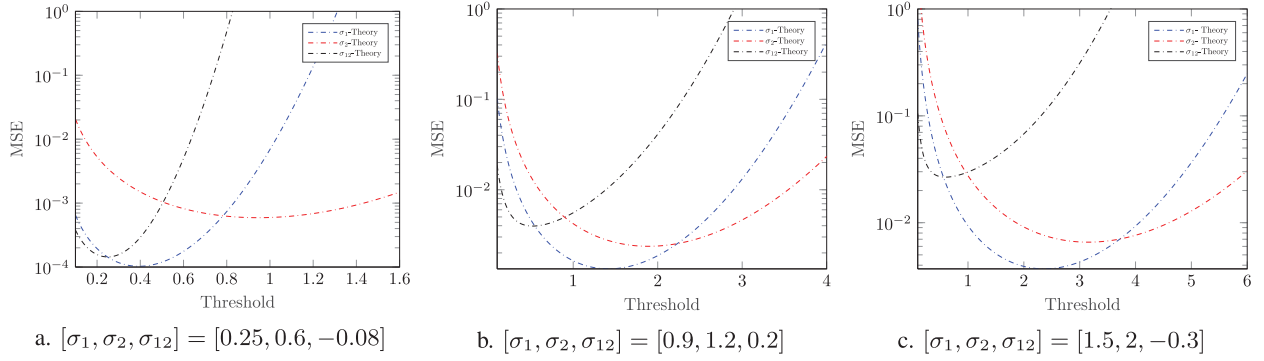
Fig. 1. Mean squared error versus threshold.

and

$$g(\kappa_1, \kappa_2, \varrho)$$
$$= \frac{\kappa_1}{\sqrt{2\pi}} \exp\left(-\frac{\kappa_1^2}{2}\right) Q\left(\frac{\kappa_2 - \varrho\kappa_1}{\sqrt{1-\varrho^2}}\right) - \varrho f(\kappa_1, \kappa_2 | \varrho). \quad (40)$$

Proof: See Appendix B. □

Using the previous lemma, the variance of $\hat{\sigma}_{12}$ can be computed as

$$\mathbb{V}[\hat{\sigma}_{12}] \approx \mathbf{b}\mathbf{R}\mathbf{b}^T, \quad (41)$$

where $\mathbf{R}$ is the covariance matrix of the random vector $[\hat{p}_1, \hat{p}_2, \hat{p}_{12}]^T$:

$$\mathbf{R} = \frac{1}{N}\begin{bmatrix} p_1\bar{p}_1 & p_{12} - p_1p_2 & p_{12}\bar{p}_1 \\ p_{12} - p_1p_2 & p_2\bar{p}_2 & p_{12}\bar{p}_2 \\ p_{12}\bar{p}_1 & p_{12}\bar{p}_2 & p_{12}\bar{p}_{12} \end{bmatrix}, \quad (42)$$

with $\bar{p}_1 = 1 - p_1$, $\bar{p}_2 = 1 - p_2$, and $\bar{p}_{12} = 1 - p_{12}$.

Proof: See Appendix C. □

Having obtained the theoretical performance of the constant-threshold estimator, we now conduct a simulation to study how the estimation errors of the standard deviation and co-variance fluctuate with regard to the threshold value. As illustrated in Fig. 1, three different groups of $\sigma_1, \sigma_2, \sigma_{12}$ are selected: $[0.25, 0.6, -0.08]$, $[0.9, 1.2, 0.2]$, and $[1.5, 2, -0.3]$. Correspondingly, the threshold ranges from 0.1 to 1.6, 4, and 6. It is clearly demonstrated that the optimal threshold for the three parameters can vary significantly. In this simulation, the optimal threshold value for the standard deviation estimation is approximately 1.6 times the population standard deviation, whereas the estimation of covariance prefers a low threshold. As a result, it is difficult to use a single threshold to deal with all the parameters. This issue is further compounded in real-world applications, where the parameters may be distributed over a broad range as the dimension increases. Consequently, recovery schemes incorporating time-varying thresholds are needed.

IV. PROPOSED COVARIANCE RECOVERY SCHEME

In this section, we propose the implementation of a time-varying, known sampling threshold in lieu of constant or random sampling thresholds. Specifically, the sampling period is divided into l sub-intervals of length n, with each sub-interval employing a distinct constant threshold. Compared to [27], [45],

our approach has the potential to increase robustness, particularly in situations where the diagonal entries differ significantly or the correlation coefficients are high. To achieve this, we first establish the MLEs of σ_1 and σ_2 using the data from their respective channels, and then search for the MLE of σ_{12} with the previously estimated σ_1 and σ_2 fixed. Then, the obtained values are used as the starting point for an iterative process that ultimately yields the joint MLE of $\boldsymbol{\theta} = [\sigma_1, \sigma_2, \sigma_{12}]^T$. Finally, we prove that the joint MLE is numerically close to the initial estimates when the number of sub-intervals is small. In such cases, we can omit using the joint MLE with negligible performance loss and significant computational savings.

A. Diagonal Entries

Without loss of generality, we study the MLE of σ_i based on $\mathbf{x}_i = [x_i(1), \cdots, x_i(N)]$. The log-likelihood of σ_i can be written as:

$$\mathcal{L}(\mathbf{x}_i; \sigma_i) = \sum_{t=1}^{N} \log\left(Q\left[\frac{x_i(t)v_i(t)}{\sigma_i}\right]\right). \quad (43)$$

Consequently, the MLE of σ_i is the solution of the following equation:

$$\frac{\partial \mathcal{L}(\mathbf{x}_i; \sigma_1)}{\partial \sigma_i} = \sum_{t=1}^{N} \frac{\Delta_{1,t}(\sigma_i)}{q_{i,t}(\sigma_i)} = 0, \quad (44)$$

where

$$\Delta_{1,t}(\sigma_i) = \frac{v_i(t)}{\sqrt{2\pi}\sigma_i^2} \exp\left(-\frac{v_i^2(t)}{2\sigma_i^2}\right), \quad (45)$$

$$q_{i,t}(\sigma_i) = \frac{x_i(t) - 1}{2} + p_{i,t}(\sigma_i), \quad (46)$$

with

$$p_{i,t}(\sigma_i) = Q\left(\frac{v_i(t)}{\sigma_i}\right). \quad (47)$$

We then obtain the ML estimate of σ_i by the following Newton's iteration:

$$\hat{\sigma}_i^{(u+1)} = \hat{\sigma}_i^{(u)} - \left. \frac{\partial \mathcal{L}(\mathbf{x}_i; \sigma_i)}{\partial \sigma_i} \middle/ \frac{\partial^2 \mathcal{L}(\mathbf{x}_i; \sigma_i)}{\partial \sigma_i^2} \right|_{\sigma_i = \hat{\sigma}_i^{(u)}}, \quad (48)$$

where the second-order derivative is calculated as:

$$\frac{\partial^2 \mathcal{L}(\mathbf{x}_i; \sigma_i)}{\partial \sigma_i^2} = \sum_{t=1}^{N} \frac{q_t(\sigma_i)\Delta_{2,t}(\sigma_i) - \Delta_{1,t}^2(\sigma_i)}{q_t^2(\sigma_i)}, \quad (49)$$

with

$$\Delta_{2,t}(\sigma_i) = \frac{v_i^3(t) - 2v_i(t)\sigma_i^2}{\sqrt{2\pi}\sigma_i^5} \exp\left(-\frac{v_i^2(t)}{2\sigma_i^2}\right). \quad (50)$$

B. Non-Diagonal Entries

After obtaining the MLEs of σ_1 and σ_2, the covariance σ_{12} can be estimated by assuming $\sigma_1 = \hat{\sigma}_1$ and $\sigma_2 = \hat{\sigma}_2$. Therefore, we have

$$p_{12,t}(\tilde{\rho}) = \int_{\frac{v_1(t)}{\hat{\sigma}_1}}^{\infty} \int_{\frac{v_2(t)}{\hat{\sigma}_2}}^{\infty} f\left(y_1, y_2 \big| \tilde{\rho}\right) dy_1 dy_2. \quad (51)$$

where $\tilde{\rho} = \sigma_{12}/(\hat{\sigma}_1\hat{\sigma}_2)$. According to the Price theorem [27], [48], the derivative of p_{12} with respect to $\tilde{\rho}$ is calculated as:

$$\frac{\partial p_{12,t}(\tilde{\rho})}{\partial \tilde{\rho}} = f\left(\frac{v_1(t)}{\hat{\sigma}_1}, \frac{v_2(t)}{\hat{\sigma}_2} \big| \tilde{\rho}\right). \quad (52)$$

Then, the log-likelihood function is

$$\mathcal{L}(\mathbf{X}; \tilde{\boldsymbol{\theta}}) = \sum_{t=1}^{N} \log\left(o_t(\tilde{\boldsymbol{\theta}})\right), \quad (53)$$

where $\tilde{\boldsymbol{\theta}} = [\hat{\sigma}_1, \hat{\sigma}_2, \sigma_{12}]^T$ and

$$o_t(\tilde{\boldsymbol{\theta}}) = \begin{cases} p_{12,t}(\tilde{\rho}), & \mathbf{x}(t) = [+1, +1]^T, \\ p_{1,t}(\hat{\sigma}_1) - p_{12,t}(\tilde{\rho}), & \mathbf{x}(t) = [+1, -1]^T, \\ p_{2,t}(\hat{\sigma}_2) - p_{12,t}(\tilde{\rho}), & \mathbf{x}(t) = [-1, +1]^T, \\ 1 - p_{1,t}(\hat{\sigma}_1) - p_{2,t}(\hat{\sigma}_2) + p_{12,t}(\tilde{\rho}), & \mathbf{x}(t) = [-1, -1]^T. \end{cases} \quad (54)$$

The first-order derivative of the log-likelihood is

$$\frac{\partial \mathcal{L}(\mathbf{X}; \tilde{\boldsymbol{\theta}})}{\partial \sigma_{12}} = \sum_{t=1}^{N} \frac{\Delta'_{1,t}(\rho)}{o_t(\tilde{\boldsymbol{\theta}})}, \quad (55)$$

where

$$\Delta'_{1,t}(\tilde{\rho}) = \frac{x_1(t)x_2(t)f\left(w_1(t), w_2(t) \big| \tilde{\rho}\right)}{\hat{\sigma}_1\hat{\sigma}_2}, \quad (56)$$

with

$$w_1(t) = \frac{v_1(t)}{\hat{\sigma}_1}, \qquad w_2(t) = \frac{v_2(t)}{\hat{\sigma}_2}. \quad (57)$$

In addition, the second-order derivative can be computed as

$$\frac{\partial^2 \mathcal{L}(\mathbf{X}; \tilde{\boldsymbol{\theta}})}{\partial \sigma_{12}^2} = \sum_{t=1}^{N} \frac{o_t(\tilde{\boldsymbol{\theta}})\Delta'_{2,t}(\tilde{\rho}) - [\Delta'_{1,t}(\tilde{\rho})]^2}{o_t^2(\tilde{\boldsymbol{\theta}})}, \quad (58)$$

where

$$\Delta'_{2,t}(\tilde{\rho}) = \frac{1}{2\pi\hat{\sigma}_1\hat{\sigma}_2\sqrt{1 - \tilde{\rho}^2}} \left[\frac{\tilde{\rho} + w_1(t)w_2(t)}{1 - \tilde{\rho}^2} - \frac{\tilde{\rho}u_t(\tilde{\rho})}{(1 - \tilde{\rho}^2)^2}\right]$$
$$\times \exp\left[-\frac{u_t(\tilde{\rho})}{2(1 - \tilde{\rho}^2)}\right], \quad (59)$$

with

$$u_t(\tilde{\rho}) = w_1^2(t) + w_2^2(t) - 2\tilde{\rho}w_1(t)w_2(t). \quad (60)$$

Similarly, we construct the Newton's iteration algorithm to solve this problem, which is:

$$\hat{\sigma}_{12}^{(u+1)} = \hat{\sigma}_{12}^{(u)} - \frac{\partial^2 \mathcal{L}(\mathbf{X}; \tilde{\boldsymbol{\theta}})}{\partial \sigma_{12}} \bigg/ \frac{\partial^2 \mathcal{L}(\mathbf{X}; \tilde{\boldsymbol{\theta}})}{\partial \sigma_{12}^2} \bigg|_{\sigma_{12} = \hat{\sigma}_{12}^{(u)}}. \quad (61)$$

C. Joint MLE

Having obtained the initial estimates, we now seek the joint MLE of σ_1, σ_2, and σ_{12}, which can be achieved using a gradient descent approach. Following the argument in (38) and (39), it is easy to obtain the gradients of the log-likelihood with respect to σ_1 and σ_2 as

$$\frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \sigma_1} = \sum_{t=1}^{N} \frac{1}{\sigma_1 o_t(\boldsymbol{\theta})} g\left(z_1(t), z_2(t), x_1(t)x_2(t)\rho\right), \quad (62)$$

$$\frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \sigma_2} = \sum_{t=1}^{N} \frac{1}{\sigma_2 o_t(\boldsymbol{\theta})} g\left(z_2(t), z_1(t), x_1(t)x_2(t)\rho\right), \quad (63)$$

where $z_i(t) = v(t)x_i(t)/\sigma_i$. Furthermore, since

$$\frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \sigma_{12}} = \sum_{t=1}^{N} \frac{x_1(t)x_2(t)}{\sigma_1\sigma_2 o_t(\boldsymbol{\theta})} f\left(z_1(t), z_2(t)|\rho\right), \quad (64)$$

the iterative procedure is

$$\hat{\boldsymbol{\theta}}^{(u+1)} = \hat{\boldsymbol{\theta}}^{(u)} + \gamma^{(u)} \frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(u)}}, \quad (65)$$

where $\gamma^{(u)}$ is the learning rate at the uth iteration.

However, when the number of sub-intervals l is small, i.e., the number of different thresholds is small, the above iterative process can be omitted with minimal performance loss and the estimates are given by those in previous sections. This assertion is proved in Appendix D and, in the next section, it is also verified by numerical simulations.

Remark 1: When considering dimension greater than 2, our approach involves recursively employing the 2×2 algorithm to reconstruct the covariance matrix. Theoretically, as the dimensions increase, there is a potential risk for the resulting matrix to be not positive-semidefinite. However, it is worth noting that throughout our extensive simulations, we have yet to encounter this particular issue. Special attention should be given to this potential limitation, especially in applications involving extremely high-dimensional settings.

D. Complex-Valued Case

We now assume $\mathbf{x}$ follows a multivariate complex Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$. We perform the widely linear transformation [62], namely, stacking the real and imaginary parts of $\mathbf{x}$ as $\underline{\mathbf{x}} = [\mathbf{w}^T, \mathbf{z}^T]^T$, where $\mathbf{w} = \text{Re}(\mathbf{x})$ and $\mathbf{z} = \text{Im}(\mathbf{x})$. Then, the covariance matrix of $\underline{\mathbf{x}}$ is

$$\boldsymbol{\Sigma}_{\underline{\mathbf{x}}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{ww}} & \boldsymbol{\Sigma}_{\mathbf{wz}} \\ \boldsymbol{\Sigma}_{\mathbf{zw}} & \boldsymbol{\Sigma}_{\mathbf{zz}} \end{bmatrix}. \quad (66)$$

Accordingly, we perform the same procedure to transform the one-bit samples $\mathbf{y}$ into $\underline{\mathbf{y}}$. Then, $\boldsymbol{\Sigma}_{\underline{\mathbf{x}}}$ is estimated from $\underline{\mathbf{y}}$ via the algorithm in the previous subsection. Finally, we reconstruct the covariance matrix of $\mathbf{x}$ from $\hat{\boldsymbol{\Sigma}}_{\underline{\mathbf{x}}}$ as

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{ww}} + \hat{\boldsymbol{\Sigma}}_{\mathbf{zz}} + \imath(\hat{\boldsymbol{\Sigma}}_{\mathbf{zw}} - \hat{\boldsymbol{\Sigma}}_{\mathbf{wz}}). \quad (67)$$

E. Performance Analysis of the Estimator

This section delves into the analysis of the MSE for the proposed time-varying threshold-based approach. To summarize our findings, we present the following theorem.

Theorem 1: The MSE matrix of the MLE can be approximated asymptotically ($N \to \infty$) by

$$\mathbf{Q} = \mathbf{F}^{-1}(\boldsymbol{\theta}_0). \tag{68}$$

Here, $\mathbf{F}(\boldsymbol{\theta})$ denotes the Fisher information matrix (FIM) defined as:

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right]. \tag{69}$$

Furthermore, $\boldsymbol{\theta}_0 = [\sigma_1, \sigma_2, \sigma_{12}]^T$ represents the genuine parameter vector.

Proof: See Appendix E.   □

Since the samples are mutually independent, we can compute the Fisher information contributed by each sample separately. Using the first-order derivatives in (62)–(64) and the fact that $\mathbf{x}(t) \in \{\pm 1, \pm 1\}$, for $t = 1, \cdots, N$, the FIM is computed as

$$\mathbf{F}(\boldsymbol{\theta}) = \sum_{t=1}^{N} \sum_{\mathbf{x}(t) \in \{\pm 1, \pm 1\}} o_t(\boldsymbol{\theta}) \left[\frac{\partial \mathcal{L}(\mathbf{x}(t))}{\partial \boldsymbol{\theta}} \frac{\partial \mathcal{L}(\mathbf{x}(t))}{\partial \boldsymbol{\theta}^T}\right]. \tag{70}$$

Building upon Theorem 1, the asymptotic MSE for the individual components can be gleaned from the diagonal entries of $\mathbf{F}^{-1}(\boldsymbol{\theta}_0)$.

Remark 2: A crucial application of the above result is the optimization of the sampling threshold. This threshold can be dynamically adjusted in practice. More specifically, we can assess the MSE by utilizing the current covariance matrix estimate, and then adopt the threshold value to minimize the MSE in the subsequent stages of the observation process. Though optimizing the overall MSE can be a challenging task due to the presence of the matrix inverse operator, a more straightforward but approximate strategy could be minimizing the most significant MSE among all elements. As observed in Fig. 1, the optimal threshold value for variance estimation is approximately 1.6 times the standard deviation. In contrast, determining the optimal threshold for covariance estimation is more complicated. While it can be approximated numerically by creating a plot akin to Fig. 1, the exploration of potential numerical optimization techniques could prove beneficial for future work.

F. Complexity Analysis

During the iterative process, the primary computational load originates from two sources: 1) computation of the Q function, denoted as C_1, and 2) computation of the 2-dimensional orthant probability (2-D OP), denoted as C_2. We segment the entire sampling process into distinct sub-intervals, and within each, a separate Q function is computed for the evaluation of p_1 and p_2. Furthermore, we calculate an additional Q function alongside a 2-D OP for p_{12}. The total number of iterations for the ith variance estimation is designated as t_i, while that for the covariance is denoted as t_{12}. The total computational load for variance estimation is thus $l(t_1 + t_2) \times C_1$, while that for

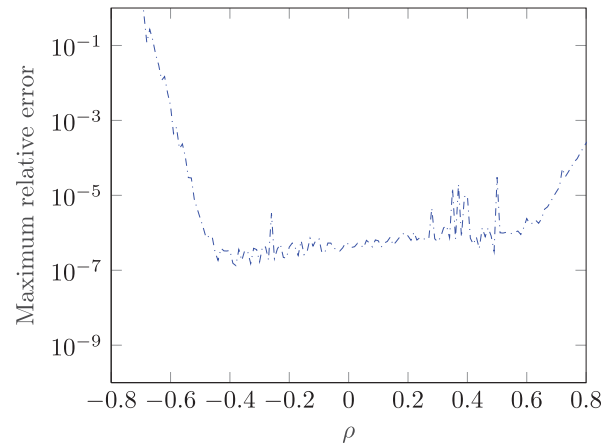
Fig. 2.   Maximum relative error versus correlation coefficient.

covariance estimation equates to $lt_{12} \times C_2$. In the joint MLE process, a total computation of $lr \times (C_1 + C_2)$ is required, where r is the number of iterations in this step. Consequently, the total computational load primarily comprises:

$$C = l(t_1 + t_2 + r)C_1 + l(t_{12} + r)C_2. \tag{71}$$

Evidently, C is influenced by two factors: the number of iterations and the computation of the Q function and the 2-D OP.

We begin our study with the number of iterations using simulations. We establish a threshold that ranges from 0.1 to 1, increasing in increments of 0.1, with each value maintained for $1/10$ of the sampling period. The standard deviations σ_1 and σ_2 are independently generated within $[0.5, 1.5]$, while ρ falls within $[-0.95, 0.95]$, all according to a uniform distribution. After conducting $10,000$ simulations, we find the average number of iterations for variance estimation to be 4.3 and 4.7 for covariance estimation. We then explore the number of iterations necessary for joint MLE. To showcase the efficacy of employing separate MLEs as initial values in reducing the iteration numbers, we also explore the scenario where separate MLEs are not adopted as the initial value, opting for $\mathbf{I}_2$ instead. Our findings show that an average of 5.2 iterations are needed with initial estimates, while 270.2 iterations are required without them. Thus, utilizing separate MLEs limits the total number of iterations effectively.

Next, we delve into the computational loads C_1 and C_2. Given the Q function's computational maturity and the availability of numerous efficient algorithms, its computational load is relatively low. In contrast, the 2-D OP requires numerical evaluation or the application of the Hermite polynomial method as detailed in [27]. To compare the accuracy and evaluation times between these two methods, we first plot the relative error of the Hermite polynomial method. From (11), it is clear that accuracy is primarily influenced by the parameter ρ. Hence, ρ is varied from -0.95 to 0.95, and the two variances are independently generated uniformly within $[0.5, 1.5]$ with a threshold of 1. We carry out 1000 trials at each point, documenting the maximum relative error. The results, depicted in Fig. 2, affirm the polynomial method's reasonable accuracy for $|\rho| < 0.6$. As a result, in each iteration step, based on the current value of ρ,

we opt for numerical integration if $|\rho| \geq 0.6$ and use the Hermit polynomial method when $|\rho| < 0.6$.

Secondly, we study the average elapsed time, utilizing a computer equipped with a 3.8 GHz Intel i9 processor and 32 GB RAM for the simulation. The average elapsed time is 4.86 ms for numerical integration and a significantly lower 0.014 ms for the Hermite polynomial method. We can thus express the total computational cost as:

$$C = l(t_1 + t_2 + r)C_1 + l(t'_{12} + r')C'_2 + l(t''_{12} + r'')C''_2. \tag{72}$$

Here, C'_2 denotes the cost of numerical integration and C''_2 signifies the cost of the Hermite polynomial method. The variables t'_{12} and r' encapsulate the numbers of instances when $|\rho| \geq 0.6$ for the estimation of σ_{12} and the joint MLE, respectively. Meanwhile, t''_{12} and r'' represent the number of occurrences when $|\rho| < 0.6$.

Lastly, performing the joint MLE for dimensions higher than 2 can prove quite expensive due to the rapidly increasing complexity of higher-order orthant probabilities with the dimension [63]. Hence, in practical scenarios, it is often sufficient to recover the covariance matrix by first estimating the variances, followed by pairwise covariance estimation. Under such circumstances, the total computational cost can be represented as:

$$C = \sum_{i=1}^{m} l t_i C_1 + \sum_{i=1}^{m} \sum_{j=1}^{i-1} l(t'_{ij} C'_2 + t''_{ij} C''_2). \tag{73}$$

In this formulation, t_i represents the number of iterations in the estimation of the ith variance, t'_{ij} represents the number of iterations in the estimation of the (i, j)th covariance where numerical integration is employed, and t''_{ij} signifies the instance where the Hermite polynomial method is applied.

V. NUMERICAL RESULTS

In this section, we conduct numerical simulations to compare the proposed recovery scheme with existing results using constant [27] and random [45], [46] thresholds in both stationary and non-stationary scenarios. Additionally, we validate the accuracy of our MSE analysis. Each result represents a Monte Carlo simulation based on 10^5 independent tests.

A. Usefulness of Exact Threshold Values

We commence our simulation by examining the benefits of using the exact values of the thresholds over their statistical properties to decrease the MSE. Here we establish σ_1 at 0.7, σ_2 at 0.9, and σ_{12} at 0.25, and the number of samples N to vary from 1000 to 3000. The threshold adheres to a Gaussian distribution $\mathcal{N}(0.5\mathbf{1}_2, 0.1\mathbf{I}_2)$. The difference is that our proposed method uses the exact values of the threshold, whereas [45] perceives the thresholds as random and utilizes their statistical attributes. As the outcome in Fig. 3 reveals, the MSEs of σ_1, σ_2 and σ_{12} all undergo a reduction, thus affirming the advantage of employing exact threshold values. This superiority arises from the fact that exact values inherently contain more useful information than statistical properties, thereby facilitating a
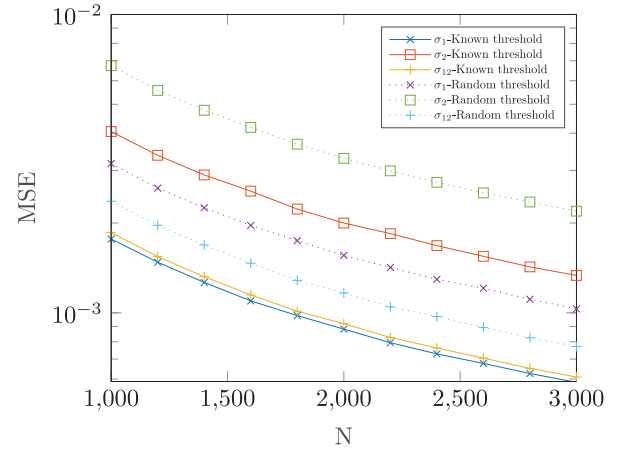

Fig. 3.   Mean squared error versus number of samples.

more accurate and efficient estimation process that leads to a reduction in MSE.

B. Comparison of Mean Squared Errors

Subsequently, we compare the overall MSE of our proposed method with [27], [45], [46] in both stationary and non-stationary cases. We begin with the more general non-stationary case. In Fig. 4(a), the population parameters are chosen as $\sigma_1 = 0.25$, $\sigma_2 = 0.6$, and $\sigma_{12} = -0.08$, and the number of samples is $N = 1000$. Our approach employs a threshold that varies from 0.1 to 1, with increments of 0.1, and each value is maintained for $1/10$ of the sampling period. The constant threshold approach takes a different value between 0.1 and 1 for each simulation. For the random threshold method, the thresholds are combined with a dithering signal following $\mathcal{N}(\mathbf{0}_2, 0.15 \cdot \mathbf{I}_2)$. We plot the total MSE of σ_1, σ_2, and σ_{12}. The results show that the time-varying threshold provides a lower MSE than any constant threshold value, as it can effectively estimate parameters over a wider range. It also outperforms the random threshold approach as it exploits the exact values of the threshold rather than their statistical properties.

Next, we compare our method with [27] and [46] in the context of stationary signals. Specifically, the covariance matrix of stationary signals is a Toeplitz matrix, and we have set its first column to $[1, -0.3, 0.1, 0]^T$. For our method, the stationary property is exploited by averaging the elements on each diagonal of the estimated covariance matrix. The threshold of our method is set as linearly spaced within the range of 0.2 to 0.3, comprised of 10 points, each lasting for $1/10$ of the sampling period. We collect the total MSE of the 4 parameters and the result is plotted in Fig. 4(b). Yet again, we observe that our proposed method excels both constant and random threshold approaches across any threshold value. However, the performance improvement decreases compared with the non-stationary case. This effect will be elaborated upon in Subsection D.

C. Influence of Correlation Coefficient

Next, we examine the impact of the correlation coefficient on estimation accuracy. We set $\sigma_1 = 0.25$ and $\sigma_2 = 0.6$, while
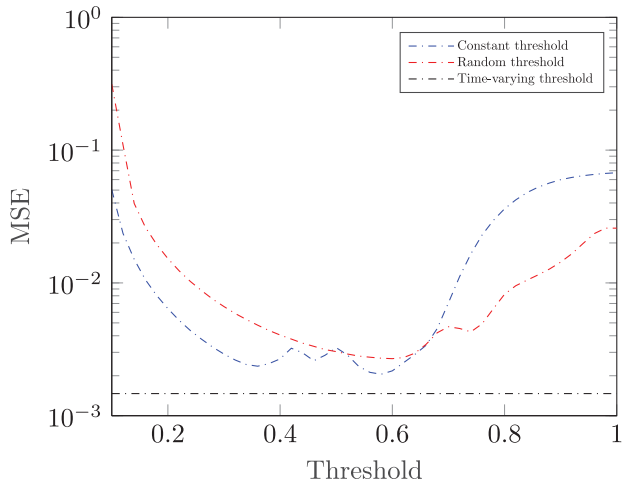
(a) Non-stationary signals
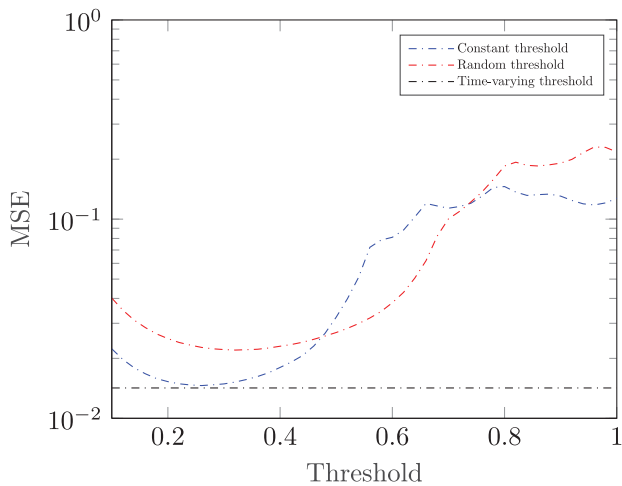
(b) Stationary signals

Fig. 4. Mean squared error versus number of samples.

Fig. 5. Mean squared error versus correlation coefficient.
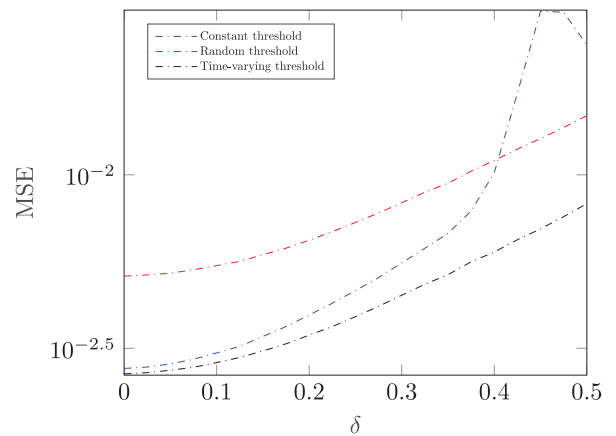
Fig. 6. Mean squared error versus variance separation level.

the correlation coefficient ranges from -0.95 to 0.95, and the number of samples is still $N = 1000$. The constant threshold approach employs a threshold value of 0.5, while the dithering signal corresponding to the random threshold approach and the threshold for our approach remain as in the previous experiment.

Compared to fixed or random thresholds, our method generally yields smaller MSE and demonstrates greater robustness, as shown in Fig. 5. The dithering approach is also more stable than the constant threshold although it yields a higher MSE on average.

D. Influence of Variance Unevenness

As illustrated in Fig. 1, the optimal threshold for variance estimation is approximately 1.6 times the standard deviation. Therefore, different variances will make the estimation more challenging for a constant threshold. In the next experiment, we set $\sigma_1 = 0.6 + \delta$ and $\sigma_2 = 0.6 - \delta$. The correlation coefficient is set to 0.5 and $N = 1000$.
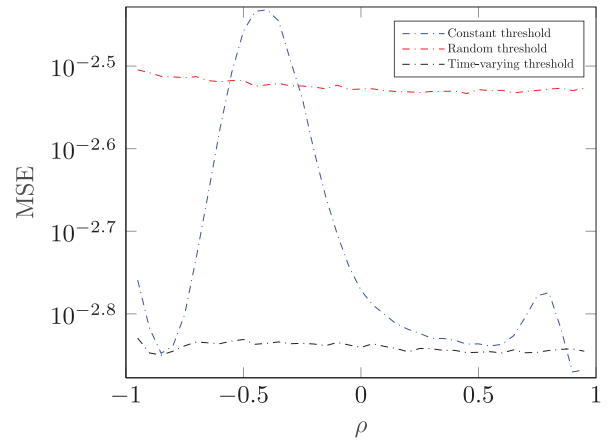
It is clear that all three approaches experience degradation in performance as the level of unevenness increases, as Fig. 6 shows. However, the time-varying threshold approach demonstrates the smallest increase in estimation error, which highlights its robustness when estimating covariance matrices with diverse parameters, which is a common in real-world applications.

E. Influence of Threshold Strategy

Now, we investigate the impact of the threshold strategy on our proposed time-varying known threshold approach. Specifically, we study two aspects: the range window and the resolution of the threshold. For this purpose, we set $\sigma_1 = 0.5$, $\sigma_2 = 0.7$, and $\sigma_{12} = 0.08$. The threshold window is in the form of $[0.1 + \delta, 1 + \delta]$, with δ ranging in a linspace between $[0, 0.8]$ with a length of 10. We experiment with a high-resolution strategy with 10 different threshold values, as well as a low-resolution strategy with 3 values. A total of $N = 3000$ samples are collected. The results are presented in Fig. 7, indicating that the MSE of σ_1 and σ_2 in the $l = 10$ case is relatively stable, while that in the $l = 3$ case exhibits more fluctuations. Interestingly, the MSE of σ_{12} in the $l = 10$ case increases more quickly than in the $l = 3$ case. This is because the $l = 3$ case
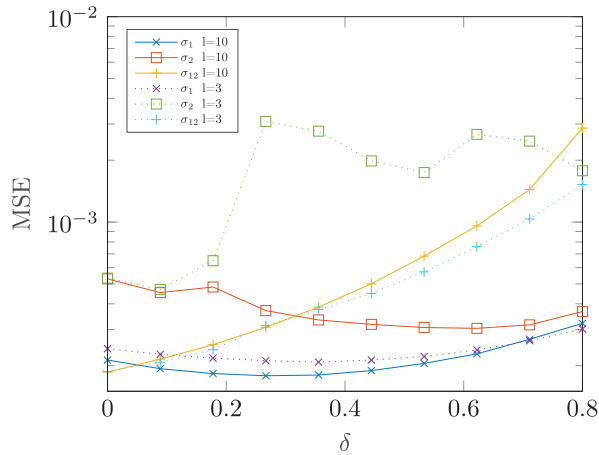
Fig. 7.   Mean squared error versus threshold shift level.

TABLE I
ABSOLUTE INITIAL GRADIENT AND MSE COMPARISON BETWEEN
JOINT AND SEPARATE MLES

|	Largest gradient	MSE (Separate)	MSE (Joint)
σ_1	8.382×10^{-3}	2.291×10^{-4}	2.241×10^{-4}
σ_2	9.213×10^{-4}	1.024×10^{-3}	1.023×10^{-3}
σ_{12}	8.496×10^{-7}	2.160×10^{-4}	2.137×10^{-4}

retains more low thresholds as δ increases. Therefore, in practice, it is important to set a portion of low thresholds to improve covariance recovery. Moreover, the analysis reveals that the original setting with $\delta = 0$, where the threshold ranges linearly between 0.1 and 1 with $l = 10$, offers robust performance for all σ_1, σ_2 and σ_{12}, thus presenting itself as an effective choice in practical applications.

### F. Influence of Joint MLE

In this subsection, we verify the effectiveness of estimating variances separately versus seeking the joint MLE. We collect the largest gradients that emerged in the iterative procedure in (65) and compare the MSE with and without this process. The results are presented in Table I for $\sigma_1 = 0.25$, $\sigma_2 = 0.6$, $\rho = 0.5$, and $N = 1000$. We observe that even the largest gradients exhibit negligible values, indicating that the iterative process for joint MLE has a minimal impact on the estimation result. Furthermore, the initial estimates provide nearly identical MSE values as the joint MLE, implying that the iterative process for joint MLE can be safely omitted without any adverse effects on performance as shown in Appendix D.

G. Theoretical Mean Squared Error

Now we examine the accuracy of the theoretical MSE of the variance estimator and covariance estimator obtained by inverting the FIM in (70). The population parameters are set as $\sigma_1 = 0.8$, $\sigma_2 = 0.9$, $\sigma_{12} = 0.25$, and $N = 1000$. We begin by investigating the theoretical performance of our approach in Fig. 8(a), where the sampling thresholds remain unchanged as previously. The result corresponding to the constant threshold
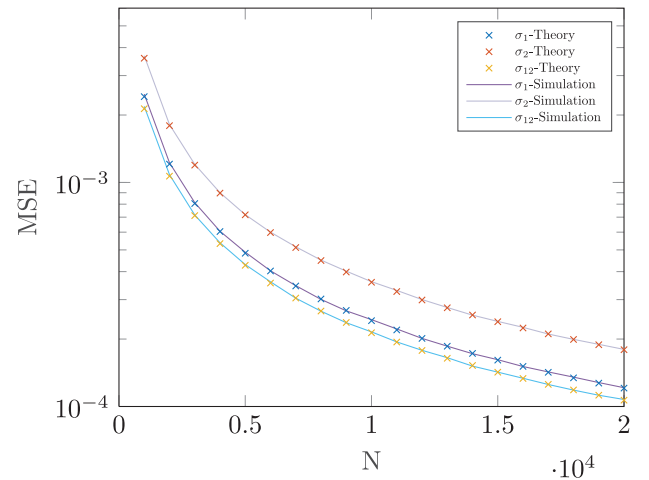
(a) Time-varying threshold
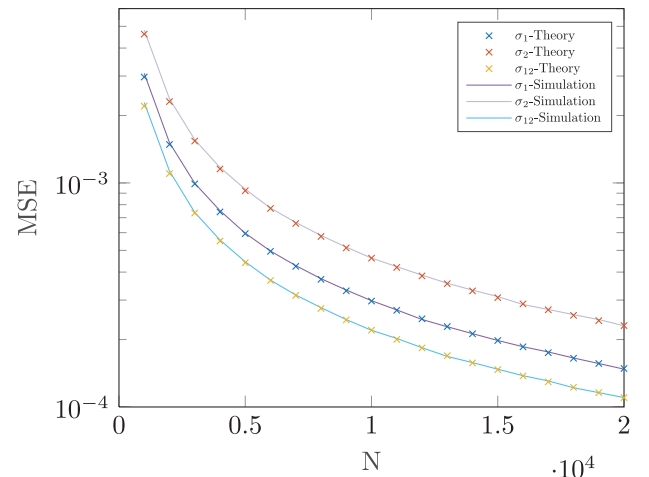
(b) Constant threshold

Fig. 8.   Mean squared error versus number of samples.

is illustrated in Fig. 8(b). It is worth noting that the covariance matrix of the dithering signal in the random threshold approach can be incorporated into that of the signal part, thus, the performance of the random threshold approach is predictable by the result of the constant threshold approach, eliminating the need for a different simulation.

H. DOA Estimation of Coherent Sources

Finally, we assess the performance of the three methods in a real-world application, namely, the DOA estimation of coherent sources. The covariance matrix is first reconstructed using each of the three methods, and then processed by the EPUMA [20] algorithm. A total of 6 antennas are considered and there are three sources located at $15°$, $45°$, and $75°$, with a signal-to-noise-ratio (SNR) of 20dB. The source signal is generated from a circular complex Gaussian distribution, wherein the first two sources are coherent, while the last source is independent of the first two. The number of samples is 10000, and a total of 20 simulations were conducted. For comparative purposes, we also consider a one-bit DOA estimation method designed for

a. Time-varying threshold   b. Constant threshold
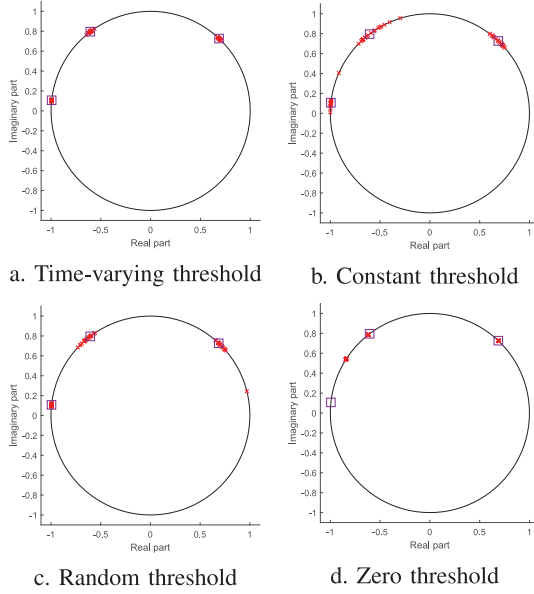

c. Random threshold   d. Zero threshold

Fig. 9.   Comparison of estimated DOA.

zero thresholds, as proposed by [64]. Fig. 9 shows that our time-varying threshold approach provides the most accurate and reliable results compared to the constant threshold and random threshold methods. This is due to the fact that the parameters of the actual covariance matrix can span a wide range, making robustness a crucial factor in ensuring estimation precision. Additionally, it is observed that the zero-threshold method produces very stable estimates, but the DOA of the last source exhibits a bias. This occurs because, in the case of correlated sources, the diagonal elements of the covariance matrix become non-uniform, and the zero-threshold hinders their estimation. This insight underscores the practical benefits of employing non-zero thresholds.

VI. Conclusion

The results of this article demonstrate the importance of threshold selection in one-bit estimation of covariance matrices. By examining the limitations of a static threshold approach, a novel time-varying threshold-based recovery scheme is developed to achieve improved accuracy in the estimation of covariance matrices. The superior performance is demonstrated through both theoretical analysis and numerical simulations, and the results show significantly reduced MSE and enhanced robustness in complex scenarios. This study opens the door for future research to further optimize the threshold selection based on the derived theoretical results of the MSE. The results of this study also have a wide range of potential applications in many areas, including array processing and communications.

Appendix A
Proof of (22)

The first- and second-order derivatives of $h(a)$ are

$$h'(p_i) = -\frac{v}{[Q^{-1}(p_i)]^2} \frac{\partial Q^{-1}(p_i)}{\partial p_i}, \tag{74}$$

$$h''(p_i) = \frac{v}{[Q^{-1}(p_i)]^3} \left(2\left[\frac{\partial Q^{-1}(p_i)}{\partial p_i}\right]^2 - Q^{-1}(p_i)\frac{\partial^2 Q^{-1}(p_i)}{\partial p_i^2} \right). \tag{75}$$

Using the formulas of the derivative of inverse functions, we have:

$$\frac{\partial Q^{-1}(a)}{\partial a} = \frac{1}{Q'(Q^{-1}(a))}, \tag{76}$$

$$\frac{\partial^2 Q^{-1}(a)}{\partial a^2} = \frac{Q''(Q^{-1}(a))}{[Q'(Q^{-1}(a))]^3}. \tag{77}$$

Now, taking into consideration that

$$Q^{-1}(p_i) = \frac{v}{\sigma_i}, \tag{78}$$

and

$$Q'(a) = \frac{\partial Q(a)}{\partial a} = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right), \tag{79}$$

$$Q''(a) = \frac{\partial^2 Q(a)}{\partial a^2} = \frac{a}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right), \tag{80}$$

the derivatives become

$$\frac{\partial Q^{-1}(p_i)}{\partial p_i} = -\sqrt{2\pi} \exp\left(\frac{v^2}{2\sigma_i^2}\right), \tag{81}$$

$$\frac{\partial^2 Q^{-1}(p_i)}{\partial p_i^2} = \frac{2\pi v}{\sigma_i} \exp\left(\frac{v^2}{\sigma_i^2}\right). \tag{82}$$

Substituting (81), (82), and (78) into (74) and (75) yields

$$h'(p_i) = \frac{\sqrt{2\pi}\sigma_i^2}{v} \exp\left(\frac{v^2}{2\sigma_i^2}\right), \tag{83}$$

$$h''(p_i) = \exp\left(\frac{v^2}{\sigma_i^2}\right) \left(\frac{4\pi\sigma_i^3}{v^2} - 2\pi\sigma_i\right). \tag{84}$$

Appendix B
Proof of Lemma 2

We first establish the first-order Taylor's expansion p_{12} at $\hat{p}_{12}$:

$$
\begin{aligned}
\hat{p}_{12} &= p_{12}(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_{12}) \\
&\approx p_{12}(\sigma_1, \sigma_2, \sigma_{12}) + \frac{\partial p_{12}}{\partial \sigma_1}(\hat{\sigma}_1 - \sigma_1) \\
&\quad + \frac{\partial p_{12}}{\partial \sigma_2}(\hat{\sigma}_2 - \sigma_2) + \frac{\partial p_{12}}{\partial \sigma_{12}}(\hat{\sigma}_{12} - \sigma_{12}).
\end{aligned} \tag{85}
$$

Rearranging terms, we get:

$$
\begin{aligned}
&\sigma_{12} - \hat{\sigma}_{12} \\
&\approx \frac{\partial \sigma_{12}}{\partial p_{12}} \left[p_{12} - \hat{p}_{12} - \frac{\partial p_{12}}{\partial \sigma_1}(\sigma_1 - \hat{\sigma}_1) - \frac{\partial p_{12}}{\partial \sigma_2}(\sigma_2 - \hat{\sigma}_2) \right],
\end{aligned} \tag{86}
$$

where we have used the inverse function rule. In the previous subsection, we obtained

$$\sigma_i - \hat{\sigma}_i = h'(p_i)(p_i - \hat{p}_i) + \mathcal{O}((p_i - \hat{p}_i)^2), \quad i = 1, 2. \tag{87}$$

Combining (86) and (87) we have the following linear function:

$$\sigma_{12} - \hat{\sigma}_{12} \approx \frac{\partial \sigma_{12}}{\partial p_{12}} \left[p_{12} - \hat{p}_{12} - \frac{\partial p_{12}}{\partial \sigma_1} h'(p_1)(p_1 - \hat{p}_1) \right.$$
$$\left. - \frac{\partial p_{12}}{\partial \sigma_2} h'(p_2)(p_2 - \hat{p}_2) \right]$$
$$= \mathbf{b} \left[p_1 - \hat{p}_1, p_2 - \hat{p}_2, p_{12} - \hat{p}_{12} \right]^T. \tag{88}$$

Moreover, since

$$p_{12} = \int_{\frac{v}{\sigma_1}}^{\infty} \int_{\frac{v}{\sigma_2}}^{\infty} f(x_1, x_2 | \rho) \, dx_1 dx_2, \tag{89}$$

the partial derivative $\partial p_{12} / \partial \sigma_1$ is computed via the following integration:

$$\frac{\partial p_{12}}{\partial \sigma_1} = \frac{v}{\sigma_1^2} \int_{\frac{v}{\sigma_2}}^{\infty} f\left(\frac{v}{\sigma_1}, x_2 \Big| \rho\right) dx_2$$
$$- \frac{\rho}{\sigma_1} \frac{\partial}{\partial \rho} \int_{\frac{v}{\sigma_1}}^{\infty} \int_{\frac{v}{\sigma_2}}^{\infty} f(x_1, x_2 | \rho) \, dx_1 dx_2$$
$$= \frac{v}{\sigma_1^2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2\sigma_1^2}\right) Q\left(\frac{v/\sigma_2 - \rho v/\sigma_1}{\sqrt{1 - \rho^2}}\right)$$
$$- \frac{\rho}{\sigma_1} f\left(\frac{v}{\sigma_1}, \frac{v}{\sigma_2} \Big| \rho\right)$$
$$= \frac{1}{\sigma_1} g\left(\frac{v}{\sigma_1}, \frac{v}{\sigma_2}, \rho\right), \tag{90}$$

where we have used Leibniz integral rule and also (52) to compute the last term. Similarly, we could obtain $\partial p_{12} / \partial \sigma_2$. Finally, and using again the derivative $\partial p_{12} / \partial \rho$ that has been calculated in (52), it is straightforward to obtain

$$\frac{\partial p_{12}}{\partial \sigma_{12}} = \frac{\partial p_{12}}{\partial \rho} \frac{\partial \rho}{\partial \sigma_{12}} = \frac{1}{\sigma_1 \sigma_2} f\left(\frac{v}{\sigma_1}, \frac{v}{\sigma_2} \Big| \rho\right). \tag{91}$$

APPENDIX C
PROOF OF (42)

As $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_{12}$ are scaled binomial random variables, the diagonal entries of $\mathbf{R}$ are easily determined as:

$$[\mathbf{R}]_{1,1} = \frac{p_1 - p_1^2}{N}, \tag{92}$$

$$[\mathbf{R}]_{2,2} = \frac{p_2 - p_2^2}{N}, \tag{93}$$

$$[\mathbf{R}]_{3,3} = \frac{p_{12} - p_{12}^2}{N}. \tag{94}$$

The covariance between $\hat{p}_1$ and $\hat{p}_{12}$ is

$$\mathbb{C}(\hat{p}_1, \hat{p}_{12}) = \mathbb{E}[\hat{p}_1 \hat{p}_{12}] - p_1 p_{12}$$
$$= \mathbb{E}[N_1 N_{12}]/N^2 - p_1 p_{12}, \tag{95}$$

where $N_1 = N\hat{p}_1$ and $N_{12} = N\hat{p}_{12}$. The value of $\mathbb{E}[N_1 N_{12}]$ is

$$\mathbb{E}[N_1 N_{12}] = \sum_{k,l=1}^{N} \mathbb{E}\left[\frac{x_1(k)+1}{2} \frac{x_1(l)+1}{2} \frac{x_2(l)+1}{2}\right]$$
$$= \sum_{k=1}^{N} \mathbb{E}\left[\left(\frac{x_1(k)+1}{2}\right)^2 \frac{x_2(k)+1}{2}\right]$$

$$+ \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \mathbb{E}\left[\frac{x_1(k)+1}{2} \frac{x_1(l)+1}{2} \frac{x_2(l)+1}{2}\right]$$
$$= N p_{12} + N(N-1) p_1 p_{12}$$
$$= N^2 p_1 p_{12} + N p_{12}(1 - p_1), \tag{96}$$

where we have used the independence between $x_1(k)$ and $x_2(l)$. Therefore, the covariance becomes

$$\mathbb{C}(\hat{p}_1, \hat{p}_{12}) = \frac{p_{12}(1 - p_1)}{N}. \tag{97}$$

Similarly, we can obtain

$$\mathbb{C}(\hat{p}_2, \hat{p}_{12}) = \frac{p_{12}(1 - p_2)}{N}. \tag{98}$$

Finally, since $\mathbb{E}[N_1 N_2]$ is

$$\mathbb{E}[N_1 N_2] = \sum_{k,l=1}^{N} \mathbb{E}\left[\frac{x_1(k)+1}{2} \frac{x_2(l)+1}{2}\right]$$
$$= \sum_{k=1}^{N} \mathbb{E}\left[\frac{x_1(k)+1}{2} \frac{x_2(k)+1}{2}\right]$$
$$+ \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \mathbb{E}\left[\frac{x_1(k)+1}{2} \frac{x_2(l)+1}{2}\right]$$
$$= N p_{12} + N(N-1) p_1 p_2, \tag{99}$$

where we have used again the independence between $x_1(k)$ and $x_2(l)$, the last covariance is

$$\mathbb{C}(\hat{p}_1, \hat{p}_2) = \mathbb{E}[N_1 N_2]/N^2 - p_1 p_2$$
$$= \frac{p_{12} - p_1 p_2}{N}. \tag{100}$$

The proof is complete.

APPENDIX D
PROOF OF THE VANISHING GRADIENT WITH SMALL NUMBER OF SUB-INTERVALS

Let us denote the original estimates by $\hat{\theta}$, obtained in Sections IV-A and IV-B, and the joint MLE by $\hat{\theta}'$, obtained in Section IV-C after the gradient-based algorithm converges. We start by considering the first sub-interval, which is of length n and define the following random variables:

$$K_1 = \frac{\sum_{t=1}^{n} [x_1(t)+1][x_2(t)+1]}{4n}, \tag{101}$$

$$K_2 = \frac{\sum_{t=1}^{n} [x_1(t)+1][x_2(t)-1]}{4n}, \tag{102}$$

$$K_3 = \frac{\sum_{t=1}^{n} [x_1(t)-1][x_2(t)-1]}{4n}, \tag{103}$$

$$K_4 = \frac{\sum_{t=1}^{n} [x_1(t)-1][x_2(t)+1]}{4n}, \tag{104}$$

which estimate the probability of $\mathbf{x}(t) = \boldsymbol{\epsilon}_i$, with

$$\boldsymbol{\epsilon}_1 = [+1, +1]^T, \quad \boldsymbol{\epsilon}_2 = [+1, -1]^T,$$
$$\boldsymbol{\epsilon}_3 = [-1, -1]^T, \quad \boldsymbol{\epsilon}_4 = [-1, +1]^T. \tag{105}$$

Then, the derivative of the log-likelihood with respect to σ_1 evaluated at the original estimate is

$$\sum_{t=1}^{n} \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \sum_{i=1}^{4} n K_i \frac{\partial \mathcal{L}(\mathbf{x} = \boldsymbol{\epsilon}_i; \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

$$= \sum_{i=1}^{4} \frac{n K_i}{q_i} g\left(\frac{v(t)\epsilon_{i,1}}{\hat{\sigma}_1}, \frac{v(t)\epsilon_{i,2}}{\hat{\sigma}_2}, \hat{\rho}\right), \tag{106}$$

where $\hat{\rho} = \hat{\sigma}_{12}/\hat{\sigma}_1\hat{\sigma}_2$, and

$$q_i = \int_{\frac{\tau_i v_1(t)}{\hat{\sigma}_1}}^{\infty} \int_{\frac{\tau_i v_2(t)}{\hat{\sigma}_2}}^{\infty} f\left(y_1, y_2 \middle| \tau_i \hat{\rho}\right) dy_1 dy_2, \tag{107}$$

is the probability that $\mathbf{x} = \boldsymbol{\epsilon}_i$, with $\tau_i = \epsilon_{i,1}\epsilon_{i,2}$. Recalling the definition of $g(z_1, z_2, \rho)$ in (40), it is easily seen that

$$g(\kappa_1, \kappa_2, \varrho) = -g(-\kappa_1, \kappa_2, -\varrho). \tag{108}$$

Therefore, it can be shown that

$$\sum_{t=1}^{n} \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = n\left(\frac{K_1}{q_1} - \frac{K_4}{q_4}\right) g\left(\frac{v(t)}{\hat{\sigma}_1}, \frac{v(t)}{\hat{\sigma}_2}, \hat{\rho}\right)$$

$$+ n\left(\frac{K_2}{q_2} - \frac{K_3}{q_3}\right) g\left(\frac{v(t)}{\hat{\sigma}_1}, -\frac{v(t)}{\hat{\sigma}_2}, -\hat{\rho}\right). \tag{109}$$

Since (nK_1, nK_2, nK_3, nK_4) follows a multinomial distribution with probabilities (q_1, q_2, q_3, q_4), the random variables $\frac{K_1}{q_1}, \frac{K_2}{q_2}, \frac{K_3}{q_3}, \frac{K_4}{q_4}$ follow asymptotically a Gaussian distribution $\mathcal{N}(\mathbf{1}_4, \mathbf{C})$, where

$$[\mathbf{C}]_{i,j} = \begin{cases} \frac{1-q_i}{n q_i}, & i = j, \\ -\frac{1}{n}, & i \neq j. \end{cases} \tag{110}$$

Then, we have

$$\frac{K_1}{q_1} - \frac{K_4}{q_4} = \mathcal{O}\left(n^{-\frac{1}{2}}\right), \tag{111}$$

$$\frac{K_2}{q_2} - \frac{K_3}{q_3} = \mathcal{O}\left(n^{-\frac{1}{2}}\right), \tag{112}$$

and (109) becomes

$$\sum_{t=1}^{n} \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathcal{O}\left(n^{\frac{1}{2}}\right). \tag{113}$$

Note, that this derivative is not zero because $\hat{\sigma}_1$ was obtained using the likelihood of $x_1(t), t = 1, \ldots, N$. To proceed, we apply a first-order Taylor's expansion to the derivative of the log-likelihood, which results in

$$\sum_{t=1}^{N} \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}'} = \sum_{t=1}^{N} \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

$$+ (\hat{\sigma}_1' - \hat{\sigma}_1) \sum_{t=1}^{N} \frac{\partial^2 \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1^2}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{114}$$

Since $\hat{\sigma}_1'$ is the solution to the equation

$$\sum_{t=1}^{N} \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}'} = 0, \tag{115}$$

we have

$$\hat{\sigma}_1' - \hat{\sigma}_1 \approx -\frac{\sum_{t=1}^{N} \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}}{\sum_{t=1}^{N} \frac{\partial^2 \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1^2}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}}. \tag{116}$$

Now we investigate the second-order derivative. When n is large, it becomes

$$\sum_{t=1}^{n} \frac{\partial^2 \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1^2}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \to n\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1^2}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right], \tag{117}$$

which is of order n since $\mathbb{E}\left[\partial^2 \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})/\partial \sigma_1^2|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right] = \mathcal{O}(1)$. Therefore, the numerator in (116) is a summation of l terms of order $n^{\frac{1}{2}}$ while the denominator is a summation of l terms of order n, where l is the number of sub-intervals. As a result, we obtain

$$\hat{\sigma}_1' - \hat{\sigma}_1 \approx \mathcal{O}\left(n^{-\frac{1}{2}}\right). \tag{118}$$

This implies that when l is small and $n = N/l$ is large, the estimated σ_1 in the joint MLE is close to the initial estimate. Similarly, we can obtain $\hat{\sigma}_2' - \hat{\sigma}_2 \approx \mathcal{O}(n^{-\frac{1}{2}})$. Furthermore, since $\hat{\sigma}_{12}$ is obtained using the two-channel data by solving

$$\frac{\partial \mathcal{L}(\mathbf{X}; \hat{\sigma}_1, \hat{\sigma}_2, \sigma_{12})}{\partial \sigma_{12}}\bigg|_{\sigma_{12}=\hat{\sigma}_{12}} = 0, \tag{119}$$

its initial gradient is already 0. With $\hat{\sigma}_1$ and $\hat{\sigma}_2$ remaining almost unchanged, the gradient of σ_{12} is also negligible. Then, the original estimate $\hat{\boldsymbol{\theta}}$ and the joint MLE by $\hat{\boldsymbol{\theta}}'$ are close.

APPENDIX E
PROOF OF THEOREM 1

We first prove that, for each sample vector $\mathbf{x}(t)$, $(t = 1, \cdots, N)$, the regularity condition holds, namely:

$$\mathbb{E}\left[\frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}. \tag{120}$$

Then the result naturally holds for the collection of all samples. At first, we have

$$\mathbb{E}\left[\frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\right]$$

$$= \sum_{\mathbf{x}(t) \in \{\pm 1, \pm 1\}} o_t(\boldsymbol{\theta}) \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}$$

$$= \sum_{\mathbf{x}(t) \in \{\pm 1, \pm 1\}} \frac{1}{\sigma_1} g\left(\frac{v(t)x_1(t)}{\sigma_1}, \frac{v(t)x_2(t)}{\sigma_2}, x_1(t)x_2(t)\rho\right). \tag{121}$$

Taking into account (108), it can be shown that

$$\frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\mathbf{x}(t)=\boldsymbol{\epsilon}_1} = -\frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\mathbf{x}(t)=\boldsymbol{\epsilon}_4},$$

$$\frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\mathbf{x}(t)=\boldsymbol{\epsilon}_2} = -\frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_1}\bigg|_{\mathbf{x}(t)=\boldsymbol{\epsilon}_3}, \tag{122}$$

which yields $\mathbb{E}\left[\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})/\partial \sigma_1\right] = 0$. This verifies the regularity condition for σ_1, which can be easily extended to σ_2. Similarly, we have

$$
\mathbb{E}\left[\frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_{12}}\right] = \sum_{\mathbf{x}(t) \in \{\pm 1, \pm 1\}} o_t(\boldsymbol{\theta}) \frac{\partial \mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \sigma_{12}}
$$
$$
= \sum_{\mathbf{x}(t) \in \{\pm 1, \pm 1\}} \frac{x_1(t) x_2(t)}{\sigma_1 \sigma_2} f\left(\frac{v(t) x_1(t)}{\sigma_1}, \frac{v(t) x_2(t)}{\sigma_2} \Big| \rho\right).
\tag{123}
$$

Since $f(z_1, z_2 | \rho) = f(-z_1, z_2 | -\rho)$, following the same process as above, we can prove that the summation in (123) is 0. Given the above, and by referencing [60, Appendix 3A], we can conclude that

$$
\mathbb{E}\left[\frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right] = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right].
\tag{124}
$$

We next assert the consistency of the proposed estimator. Our proof strategy mirrors that of [60, Appendix 7B]. Specifically, for a large n and the kth sub-interval, we can write:

$$
\frac{1}{n} \sum_{t=\nu_k}^{\eta_k} \ln o_t^k(\boldsymbol{\theta}) \to \mathbb{E}\left[\ln o_t^k(\boldsymbol{\theta})\right]
$$
$$
= \sum_{\mathbf{x} \in \{\pm 1, \pm 1\}} o_t^k(\boldsymbol{\theta}_0) \ln o_t^k(\boldsymbol{\theta}),
\tag{125}
$$

where $\nu_k = (k-1)n$, $\eta_k = kn$ and o_t^k denotes the likelihood function on the kth subinterval. Thus, we arrive at:

$$
\mathcal{L}(\mathbf{x}(t); \boldsymbol{\theta}) = \sum_{k=1}^{l} \sum_{\mathbf{x} \in \{\pm 1, \pm 1\}} o_t^k(\boldsymbol{\theta}_0) \ln o_t^k(\boldsymbol{\theta}).
\tag{126}
$$

Let us consider two estimators: $\hat{\boldsymbol{\theta}}_1 = \boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_2 \neq \boldsymbol{\theta}_0$. Utilizing the non-negativity of the Kullback-Leibler divergence, we have:

$$
\sum_{\mathbf{x} \in \{\pm 1, \pm 1\}} o_t^k(\boldsymbol{\theta}_0) \ln \frac{o_t^k(\boldsymbol{\theta}_0)}{o_t^k(\hat{\boldsymbol{\theta}}_2)} \geq 0,
\tag{127}
$$

which leads to:

$$
\sum_{\mathbf{x} \in \{\pm 1, \pm 1\}} o_t^k(\boldsymbol{\theta}_0) \ln o_t^k(\boldsymbol{\theta}_0) \geq \sum_{\mathbf{x} \in \{\pm 1, \pm 1\}} o_t^k(\boldsymbol{\theta}_0) \ln o_t^k(\hat{\boldsymbol{\theta}}_2).
\tag{128}
$$

From this, we can infer that $\hat{\boldsymbol{\theta}}_1 = \boldsymbol{\theta}_0$ maximizes $o_t^k(\boldsymbol{\theta})$ for $k = 1, \cdots, l$, and subsequently, the likelihood function in (126). Therefore, the MLE converges to the true parameter, proving the estimator's consistency.

Applying the first-order Taylor's expansion, we get:

$$
\left. \frac{\partial \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left. \frac{\partial \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
$$
$$
+ \left. \frac{\partial^2 \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),
\tag{129}
$$

where $\boldsymbol{\theta}_0 < \tilde{\boldsymbol{\theta}} < \hat{\boldsymbol{\theta}}$. Using its consistency property, it is evident that $\tilde{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$. Moreover, from

$$
\left. \frac{\partial \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0},
\tag{130}
$$

we derive

$$
\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left[\left. \frac{\partial^2 \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right]^{-1} \left. \frac{\partial \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.
\tag{131}
$$

With $n \to \infty$, it follows that

$$
-\sum_{t=\nu_k}^{\eta_k} \frac{\partial^2 \ln p(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \to -\mathbb{E}\left[\sum_{t=\nu_k}^{\eta_k} \frac{\partial^2 \ln p(\mathbf{x}(t); \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right]
$$
$$
= \mathbf{F}_k(\boldsymbol{\theta}),
\tag{132}
$$

where $\mathbf{F}_k(\boldsymbol{\theta})$ is the FIM in the kth subinterval and we have used (124). Thus,

$$
-\left[\left. \frac{\partial^2 \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] \to \sum_{k=1}^{l} \mathbf{F}_k(\boldsymbol{\theta}_0) = \mathbf{F}(\boldsymbol{\theta}_0).
\tag{133}
$$

From (131), it can be shown that

$$
\mathbf{Q} = \mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\right] = \mathbf{F}(\boldsymbol{\theta}_0)^{-1},
\tag{134}
$$

which completes the proof.

REFERENCES

[1] E. Balevi and J. Andrews, "One-bit OFDM receivers via deep learning," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4326–4336, Jun. 2019.

[2] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Deep learning for massive MIMO with 1-bit ADCs: When more antennas need fewer pilots," *IEEE Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1273–1277, Aug. 2020.

[3] J. Mo and R. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[4] C. Qian, X. Fu, and N. Sidiropoulos, "Amplitude retrieval for channel estimation of MIMO systems with one-bit ADCs," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1698–1702, Nov. 2019.

[5] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[6] J. Choi, J. Mo, and R. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[7] O. Bar-Shalom and A. Weiss, "DOA estimation using one-bit quantized measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 868–884, Jul. 2002.

[8] K. Yu, Y. Zhang, M. Bao, Y. Hu, and Z. Wang, "DOA estimation from one-bit compressed array data via joint sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 8, pp. 1279–1283, Sep. 2016.

[9] C. Liu and P. Vaidyanathan, "One-bit sparse array DOA estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 3126–3130.

[10] M. Stein, K. Barbe, and J. Nossek, "DOA parameter estimation with 1-bit quantization bounds, methods and the exponential replacement," in *Proc. 20th Int. ITG Workshop Smart Antennas*, Munich, Germany, 2016, pp. 1–6.

[11] S. Sedighi, B. Shankar, M. Soltanalian, and B. Ottersten, "One-bit DOA estimation via sparse linear arrays," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Barcelona, Spain, 2020, pp. 9135–9139.

[12] S. Sedighi, B. R, M. Soltanalian, and B. Ottersten, "On the performance of one-bit DOA estimation via sparse linear arrays," *IEEE Trans. Signal Process.*, vol. 69, pp. 6165–6182, 2021.

[13] J. Ren and J. Li, "One-bit digital radar," in *Proc. IEEE Asilomar Conf. Signal, Syst., Comput.*, Pacific Grove, USA, 2017, pp. 1142–1146.

[14] Y.-H. Xiao, D. Ramírez, P. Schreier, C. Qian, and L. Huang, "One-bit target detection in collocated MIMO radar and performance degradation analysis," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9363–9374, Sep. 2022.

[15] F. Xi, Y. Xiang, Z. Zhang, S. Chen, and A. Nehorai, "Joint angle and doppler frequency estimation for MIMO radar with one-bit sampling: A maximum likelihood-based method," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 6, pp. 4734–4748, Dec. 2020.

[16] F. Xi, Y. Xiang, S. Chen, and A. Nehorai, "Gridless parameter estimation for one-bit MIMO radar with time-varying thresholds," *IEEE Trans. Signal Process.*, vol. 68, pp. 1048–1063, 2020.

[17] B. Liu, B. Chen, and M. Yang, "Parameter estimation and CRB analysis of 1-bit colocated MIMO radar," *IET Radar Sonar Navigation*, vol. 1, no. 13, pp. 1–13, Mar. 2021.

[18] B. Jin, J. Zhu, Q. Wu, Y. Zhang, and Z. Xu, "One-bit IFMCW radar: Spectrum analysis and target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 4, pp. 2732–2750, Aug. 2020.

[19] M. Stein, A. Kurzl, A. Mezghani, and J. Nossek, "Asymptotic parameter tracking performance with measurement data of 1-bit resolution," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6086–6095, Nov. 2015.

[20] C. Qian, L. Huang, N. Sidiropoulos, and H. So, "Enhanced PUMA for direction-of-arrival estimation and its performance analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4127–4137, Aug. 2016.

[21] L. Wei and O. Tirkkonen, "Spectrum sensing in the presence of multiple primary users," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1268–1277, May 2012.

[22] Y. Xiao, L. Huang, J. Xie, and H. So, "Approximate asymptotic distribution of locally most powerful invariant test for independence: Complex case," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1784–1799, Mar. 2018.

[23] Y. Zhao, X. Ke, B. Zhao, Y. Xiao, and L. Huang, "One-bit spectrum sensing based on statistical covariances: Eigenvalue moment ratio approach," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2474–2478, Nov. 2021.

[24] Y. Xiao, L. Huang, J. Zhang, J. Xie, and H. So, "Performance analysis of locally most powerful invariant test for sphericity of Gaussian vectors in coherent MIMO radar," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5868–5882, Jul. 2018.

[25] W. Liu, Y. Wang, J. Liu, W. Xie, H. Chen, and W. Gu, "Adaptive detection without training data in colocated MIMO radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 2469–2479, Jul. 2015.

[26] J. Vleck and D. Middleton, "The spectrum of clipped noise," *Proc. IEEE*, vol. 54, no. 1, pp. 2–19, Jan. 1966.

[27] C.-L. Liu and Z.-M. Lin, "One-bit autocorrelation estimation with nonzero thresholds," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Toronto, ON, Canada, 2021, pp. 4520–4524.

[28] J. Ren, T. Zhang, J. Li, and P. Stoica, "Sinusoidal parameter estimation from signed measurements via majorization-minimization based relax," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2173–2186, Apr. 2019.

[29] C. Gianelli, L. Xu, J. Li, and P. Stoica, "One-bit compressive sampling with time-varying thresholds: Maximum likelihood and the Cramer-Rao bound," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 399–403.

[30] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, Apr. 2013.

[31] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, "Exponential decay of reconstruction error from binary measurements of sparse signals," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3368–3385, Jun. 2017.

[32] K. Knudson, R. Saab, and R. Ward, "One-bit compressive sensing with norm estimation," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2748–2758, May 2016.

[33] S. Dirksen and S. Mendelson, "Non-Gaussian hyperplane tessellations and robust one-bit compressed sensing," *J. Eur. Math. Soc.*, vol. 23, no. 9, pp. 2913–2947, 2021.

[34] S. Dirksen and S. Mendelson, "Robust one-bit compressed sensing with partial circulant matrices," *Ann. Appl. Probability*, vol. 33, no. 3, pp. 1874–1903, 2023.

[35] A. Eamaz, F. Yeganegi, D. Needell, and M. Soltanalian, "One-bit quadratic compressed sensing: From sample abundance to linear feasibility," 2023, *arXiv:2303.09594*.

[36] A. Eamaz, F. Yeganegi, D. Needell, and M. Soltanalian, "Harnessing the power of sample abundance: Theoretical guarantees and algorithms for accelerated one-bit sensing," 2023, *arXiv:2308.00695*.

[37] A. Eamaz, K. Mishra, F. Yeganegi, and M. Soltanalian, "UNO: Unlimited sampling meets one-bit quantization," 2022, *arXiv:2301.10155*.

[38] A. Eamaz, K. Mishra, F. Yeganegi, and M. Soltanalian, "Unlimited sampling via one-bit quantization," in *Proc. Int. Conf. Sampling Theory Appl.*, 2023, pp. 1–5.

[39] A. Eamaz, F. Yeganegi, and M. Soltanalian, "Opera: Leveraging the sample size and complexity trade-off towards efficient one-bit phase retrieval," in *Proc. 56th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, USA, 2022, pp. 81–85.

[40] A. Eamaz, F. Yeganegi, and M. Soltanalian, "One-bit phase retrieval: More samples means less complexity?" *IEEE Trans. Signal Process.*, vol. 70, pp. 4618–4632, 2022.

[41] A. Eamaz, F. Yeganegi, D. Needell, and M. Soltanalian, "ORKA: Accelerated Kaczmarz algorithms for signal recovery from one-bit samples," *arXiv:2301.03467*.

[42] A. Eamaz, F. Yeganegi, and M. Soltanalian, "One-bit matrix completion with time-varying sampling thresholds," in *Proc. Int. Conf. Sampling Theory Appl.*, 2023, pp. 1–5.

[43] S. Dirksen and J. Maly, "Tuning-free one-bit covariance estimation using data-driven dithering," 2023. [Online]. Available: https://arxiv.org/abs/2307.12613

[44] A. Eamaz, F. Yeganegi, and M. Soltanalian, "Modified arcsine law for one-bit sampled stationary signals with time-varying thresholds," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 5459–5463.

[45] A. Eamaz, F. Yeganegi, and M. Soltanalian, "Covariance recovery for one-bit sampled non-stationary signals with time-varying sampling thresholds," *IEEE Trans. Signal Process.*, vol. 70, pp. 5222–5236, 2022.

[46] A. Eamaz, F. Yeganegi, and M. Soltanalian, "Covariance recovery for one-bit sampled stationary signals with time-varying sampling thresholds," *Signal Process.*, vol. 206, no. 1, pp. 1–12, 2023, Art. no. 108899.

[47] S. Dirksen, J. Maly, and H. Rauhut, "Covariance estimation under one-bit quantization," *Ann. Statist.*, vol. 50, no. 6, pp. 3538–3562, Dec. 2022.

[48] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inf. Theory*, vol. 4, no. 2, pp. 69–72, Jun. 1958.

[49] C.-L. Liu and P. Vaidyanathan, "One-bit normalized scatter matrix estimation for complex elliptically symmetric distributions," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Barcelona, Spain, 2020, pp. 9130–9134.

[50] A. Ai, A. Lapanowski, Y. Plan, and R. Vershynin, "One-bit compressed sensing with non-Gaussian measurements," *Linear Algebra Appl.*, vol. 441, pp. 222–239, Jan. 2014.

[51] Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tessellations," *J. Eur. Math. Soc.*, vol. 23, no. 9, pp. 2913–2947, 2021.

[52] H. Jung, J. Maly, L. Palzer, and A. Stollenwerk, "Quantized compressed sensing by rectified linear units," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 4125–4149, Jun. 2021.

[53] C. Thrampoulidis and A. Rawat, "The generalized lasso for sub-Gaussian measurements with dithered quantization," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2487–2500, Apr. 2020.

[54] T. Yang, J. Maly, S. Dirksen, and G. Caire, "Plug-in channel estimation with dithered quantized signals in spatially non-stationary massive MIMO systems," 2023, *arXiv:2301.04641*.

[55] J. Chen, C.-L. Wang, M. Ng, and D. Wang, "High dimensional statistical estimation under uniformly dithered one-bit quantization," *IEEE Trans. Inf. Theory*, vol. 69, no. 8, pp. 5151–5187, 2023.

[56] B. Zhao, L. Huang, and W. Bao, "One-bit SAR imaging based on single-frequency thresholds," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7017–7032, Sep. 2019.

[57] R. Gray and T. Stockham, "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.

[58] J. Bussgang, "Cross-correlation function of amplitude-distorted Gaussian signals," Research Laboratory of Electronics, MIT Press, Cambridge, MA, USA, Tech. Rep. 216, Mar. 1952.

[59] J. Minkoff, "The role of AM-to-PM conversion in memoryless nonlinear systems," *IEEE Trans. Commun.*, vol. 33, no. 2, pp. 139–144, Feb. 1985.

[60] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[61] A. Knoblauch, "Closed-form expressions for the moments of the binomial probability distribution," *SIAM J. Appl. Math.*, vol. 69, no. 1, pp. 197–204, 2008.

[62] P. Schreier and L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Non-Circular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[63] T. Miwa, A. Hayter, and S. Kuriki, "The evaluation of general non-centred orthant probabilities," *J. Roy. Statistical Soc. B*, vol. 65, no. 1, pp. 223–234, 2003.

[64] L. Feng, L. Huang, Q. Li, Z.-Q. He, and M. Chen, "An off-grid iterative reweighted approach to one-bit direction of arrival estimation," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8134–8139, Jan. 2023.

Yu-Hang Xiao (Member, IEEE) was born in Anhui, China, in January 20, 1992. He received the B.E. degree in microelectronics from Harbin Engineering University, China, in 2012, and the master's and Ph.D. degrees in information and communication engineering from Harbin Institute of Technology (HIT), China, in 2014 and 2018, respectively. From October 2015 to February 2017, he was a Visiting Ph.D. student with the Department of Electrical and Computer Engineering, McMaster University, Canada. From February 2019 to 2021, he was a Postdoctoral with the Signal and System Theory Group, Paderborn University, Germany. Since March 2021, he has been an Assistant Professor with the College of Electronics and Information Engineering, Shenzhen University. His research interests include statistical signal processing, hypotheses testing, and one-bit signal processing.

Lei Huang (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees in electronic engineering from Xidian University, Xian, China, in 2000 and 2005, respectively. He is currently with the College of Electronics and Information Engineering, Shenzhen University, as a Distinguished Professor, and established the Shenzhen Key Laboratory of Advanced Navigation Technology (ANT), as the Founding Director. He is now the Executive Dean of the College of Electronics and Information Engineering, Shenzhen University. His research interests include spectral estimation, array signal processing, statistical signal processing, and its applications in radar, navigation, and wireless communications. In these areas, he has published 100 IEEE journal papers, and undertaken 20 national and provincial key projects, such as the Key Project of the National Natural Science Foundation of China (NSFC) and Joint Project of NSFC-RGC (Hong Kong). He was the winner of the Distinguished Young Scientists of NSFC, Program of the New Century Excellent Talents in University of Ministry of Education (MOE) of China, and Leading Talent of Special Support Program of Guangdong Province. He won the first prize of Technological Invention Award of the Chinese Institute of Electronics (CIE), and the second prize of Natural Science Award in University of MOE of China. He is currently serving as a Senior Area Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING, and was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING (2015–2019). He also was on the editorial boards of *Elsevier-Digital Signal Processing* (2012–2019), the *IET Signal Processing* (2017–present), and an elected member of Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society (2016–present). He was elected as an IET Fellow in 2018.

David Ramírez (Senior Member, IEEE) received the Telecommunication Engineer degree and the Ph.D. degree in electrical engineering from the Universidad de Cantabria, Spain, in 2006 and 2011, respectively. From 2006 to 2011, he was with the Communications Engineering Department, University of Cantabria, Spain. In 2011, he joined as a Research Associate with the University of Paderborn, Germany, and later on, he became an Assistant Professor (Akademischer Rat). He is now an Associate Professor with the Universidad Carlos III de Madrid. He has been a Visiting Researcher with the University of Newcastle, Australia, and the University College London. His current research interests include signal processing for wireless communications, statistical signal processing, changepoint management, and signal processing over graphs. He has been involved in several national and international research projects on these topics. He was the recipient of the 2012 IEEE Signal Processing Society Young Author Best Paper Award and the 2013 extraordinary Ph.D. award of the University of Cantabria. Moreover, he served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and was Publications Chair of the 2018 IEEE Workshop on Statistical Signal Processing. Finally, he is a member of the IEEE Technical Committee on Signal Processing Theory and Methods and EURASIP Technical Area Committee on Theoretical and Methodological Trends in Signal Processing.

Cheng Qian (Member, IEEE) was born in Deqing, Zhejiang, China. He received the bachelor's degree in communication engineering from Hangzhou Dianzi University, Hangzhou, China, in 2011, and the master's and Ph.D. degrees in communication and information engineering from Harbin Institute of Technology (HIT), China, in 2013 and 2017, respectively. He was a Postdoctoral Researcher with the University of Minnesota Twin Cities and the University of Florida, from April 2017 to August 2019. Since September 2019, he has been a Senior Machine Learning Scientist with the Analytics Center of Excellence, IQVIA Inc. He is also an Associate Editor of Elsevier Signal Processing. His research interests are in deep learning, machine learning for healthcare, natural language processing, and signal processing.

Hing Cheung So (Fellow, IEEE) was born in Hong Kong. He received the B.Eng. degree from the City University of Hong Kong and the Ph.D. degree from The Chinese University of Hong Kong in electronic engineering, in 1990 and 1995, respectively. From 1990 to 1991, he was an Electronic Engineer with the Research and Development Division, Everex Systems Engineering Ltd., Hong Kong. From 1995 to 1996, he was a Postdoctoral Fellow with The Chinese University of Hong Kong. From 1996 to 1999, he was a Research Assistant Professor with the Department of Electronic Engineering, City University of Hong Kong, where he is currently a Professor. His research interests include detection and estimation, fast and adaptive algorithms, multidimensional harmonic retrieval, robust signal processing, source localization, and sparse approximation. He has been on the editorial boards of *IEEE Signal Processing Magazine* (2014–2017), IEEE TRANSACTIONS ON SIGNAL PROCESSING (2010–2014), *Signal Processing* (2010–), and *Digital Signal Processing* (2011–). He was also a Lead Guest Editor for IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, special issue on "Advances in Time/Frequency Modulated Array Signal Processing," in 2017. In addition, he was an elected member in Signal Processing Theory and Methods Technical Committee (2011–2016) of the IEEE Signal Processing Society, where he was Chair in the awards subcommittee (2015–2016).