

Blind SNR Estimation and Nonparametric Channel Denoising in Multi-Antenna mmWave Systems

Alexandra Gallyas-Sanhueza¹ and Christoph Studer²

¹*Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY; email: ag753@cornell.edu*

²*Dept. of Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland; email: studer@ethz.ch*

Abstract—We propose blind estimators for the average noise power, receive signal power, signal-to-noise ratio (SNR), and mean-square error (MSE), suitable for multi-antenna millimeter wave (mmWave) wireless systems. The proposed estimators can be computed at low complexity and solely rely on beamspace sparsity, i.e., the fact that only a small number of dominant propagation paths exist in typical mmWave channels. Our estimators can be used (i) to quickly track some of the key quantities in multi-antenna mmWave systems while avoiding additional pilot overhead and (ii) to design efficient nonparametric algorithms that require such quantities. We provide a theoretical analysis of the proposed estimators, and we demonstrate their efficacy via synthetic experiments and using a nonparametric channel-vector denoising task with realistic multi-antenna mmWave channels.

I. INTRODUCTION

Accurate knowledge of system parameters, such as average noise power, signal power, or the signal-to-noise ratio (SNR) is critical in communication systems, as many baseband processing tasks use these quantities [1]. Most conventional communication systems dedicate a training stage to estimate such system parameters. However, since the propagation conditions can change at fast rates, especially at millimeter-wave (mmWave) frequencies where blockers or interferers may appear quickly [2], it is important to develop low-complexity solutions that accurately track such parameters.

Fortunately, modern wireless communication systems deal with high-dimensional data. For example, massive multiple-input multiple-output (MIMO) basestations are expected to be equipped with hundreds of antennas, or orthogonal frequency-division multiplexing (OFDM) systems have thousands of subcarriers. Since many of the signals in such systems are structured (e.g., exhibit sparsity or are taken from a discrete set), one can design statistical methods that blindly estimate parameters, without the need of a dedicated training stage.

A. Prior Art in Blind and Nonparametric Estimation

Blind estimators rely on the signal statistics rather than pilot sequences. Many of the existing blind noise variance and SNR estimators exploit modulation-specific structure, such as the cyclic prefix redundancy in OFDM [3] or periodicity of synchronization sequences [4]. Other methods use expectation

maximization (EM) with sophisticated statistical models. For example, EM has been successfully used for blind SNR estimation [5] or to recover sparse signals [6]. However, the iterative nature of EM and its relatively high per-iteration complexity renders such methods unsuitable for real-time estimation in multi-antenna mmWave wireless systems that operate with high-dimensional data at gigabit per second sampling rates. In contrast, we propose a range of low-complexity blind estimators whose complexity is only $\mathcal{O}(D)$, where D is the dimension of the processed data.

In order to deal with algorithm parameters, nonparametric methods have been proposed recently. The nonparametric equalizer (NOPE) [7], for example, performs linear minimum mean-square error (MSE) estimation in massive MIMO systems without knowledge of the SNR. NOPE combines approximate message passing [8], [9] with Stein's unbiased risk estimate (SURE) [10] to automatically tune the algorithm parameters. The concept of estimating a subset of the algorithm parameters directly from the noisy measurements, has been used recently for adaptive denoising of mmWave [11]–[13] or OFDM [14] channel vectors. Denoising algorithms typically require two parameters (the noise power and a thresholding parameter), but since the threshold can be estimated using SURE from the noisy observations, such methods only need knowledge of the noise power. In this paper, we propose low-complexity blind estimators, which enable the design of nonparametric (i.e., parameter free) channel-vector denoisers.

B. Contributions

Our main contributions are as follows. We propose low-complexity blind estimators for the average noise power, signal power, SNR, and MSE after applying an element-wise function to the noisy observation. We provide a theoretical analysis by developing bounds on the accuracy in the large-dimension limit that depend on the SNR and sparsity. We provide simulation results with synthetic data to demonstrate the efficacy and limits of our estimators in finite dimensions. We showcase an application example of our blind estimators, which leads to a novel nonparametric channel-vector denoising algorithm for mmWave massive MIMO systems.

C. Notation

Lowercase and uppercase boldface letters denote column vectors and matrices, respectively. The d th entry of the

The work of AGS and CS was supported by ComSenTer, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. The work of CS was also supported by an ETH Research Grant and by the US NSF under grants CNS-1717559 and ECCS-1824379.

vector $\mathbf{a} \in \mathbb{C}^D$ is a_d , the real and imaginary parts are $\Re\{\mathbf{a}\}$ and $\Im\{\mathbf{a}\}$, respectively, and we use $\mathbf{b} = |\mathbf{a}|^2$ to refer to $b_d = |a_d|^2$ for $d = 1, \dots, D$. The soft-thresholding function with threshold τ is defined as $\eta(x; \tau) = x/|x| \max\{|x| - \tau, 0\}$ for $x \neq 0$ and $\eta(x; \tau) = 0$ for $x = 0$, and is applied element-wise to vectors. Statistical quantities are denoted by plain symbols, e.g., the variance $E_x = \frac{1}{D} \mathbb{E}[\|\mathbf{x}\|^2]$ of the random vector $\mathbf{x} \in \mathbb{C}^D$, where $\mathbb{E}[\cdot]$ denotes expectation; sample estimates are denoted by a bar, e.g., the sample variance $\bar{E}_x = \frac{1}{D} \|\mathbf{x}\|^2$; corresponding blind estimators are denoted by a hat, e.g., $\hat{E}_x$. For $x \in \mathbb{R}$, rounding towards plus and minus infinity is denoted by $\lceil x \rceil$ and $\lfloor x \rfloor$, respectively, and $[x]_+ = \max\{x, 0\}$. The big-O notation is $\mathcal{O}(\cdot)$.

II. LOW-COMPLEXITY BLIND ESTIMATORS

A. System Models

We say that a vector $\mathbf{s} \in \mathbb{C}^D$ is *sparse* if the number of nonzero entries K is smaller than the dimension of the vector D . As a sparsity measure, one can use the ℓ_0 pseudo norm $\|\mathbf{s}\|_0 = K$, which counts the number of nonzero entries of $\mathbf{s}$. As we will show in Section III, sparsity is one of the key ingredients to our blind estimators. In the rest of the paper, we focus on the following two system models.

System Model 1. Let $\mathbf{s} \in \mathbb{C}^D$ be a sparse signal with average energy $E_s = \frac{1}{D} \mathbb{E}[\|\mathbf{s}\|^2]$. We model a noisy observation of the sparse signal using the following input-output relation:

$$\mathbf{y} = \mathbf{s} + \mathbf{n}. \quad (1)$$

Here, $\mathbf{y} \in \mathbb{C}^D$ is the noisy observation and $\mathbf{n} \in \mathbb{C}^D$ models noise with circularly-symmetric i.i.d. complex Gaussian entries with variance N_0 . We assume that the sparse signal vector $\mathbf{s}$ and noise vector $\mathbf{n}$ are statistically independent.

In what follows, we do not make assumptions on the signal sparsity $\|\mathbf{s}\|_0 = K$, even though the performance of the proposed estimators depends on this parameter; see Section III for the details. Furthermore, unlike pilot-based estimators, we assume that the sparse vector $\mathbf{s}$ is unknown, which makes parameter estimation nontrivial in our scenario.

System Model 1 finds numerous applications in wireless communication systems. Prime examples are modeling the estimated channel vectors in multi-antenna mmWave systems [11]–[13] or in OFDM systems [14], where the beamspace representation or the delay-domain representation of these channel vectors are typically sparse, respectively.

System Model 2. Let $\mathbf{y}$ be a noisy observation as in System Model 1. Fix a weakly differentiable function $\mu : \mathbb{C} \rightarrow \mathbb{C}$ that operates element-wise on vectors. We model the output as

$$\mu(\mathbf{y}) = \mathbf{s} + \mathbf{e}, \quad (2)$$

where $\mathbf{e} \in \mathbb{C}^D$ contains residual distortion.

We emphasize that in (2), the residual distortion $\mathbf{e}$ is not necessarily independent of the sparse signal $\mathbf{s}$. System Model 2 is relevant in the following scenarios: (i) When estimating a

sparse signal $\mathbf{s}$ from a noisy measurement $\mathbf{y}$ by applying an (entry-wise) denoising or estimation function, producing the signal estimate $\hat{\mathbf{s}} = \mu(\mathbf{y})$. This scenario finds use for channel-vector denoising [11], [12], for example. (ii) When modeling nonlinearities caused by hardware impairments [15], in which case the distorted version of the noisy received signal can be expressed as $\mathbf{r} = \mu(\mathbf{y})$. This scenario finds use in systems with low-resolution data converters [16], [17], for example.

B. Low-Complexity Blind Estimators

The blind estimators proposed next make frequent use of the sample median, which we define as follows.

Definition 1 (Sample Median). Let $\mathbf{s} \in \mathbb{R}^D$ be a vector and $\mathbf{s}^{\text{sort}} \in \mathbb{R}^D$ be its sorted version (entries sorted in ascending order). Then, the sample median is defined as

$$\bar{m}(\mathbf{s}) = \frac{1}{2} \left(s_{\lfloor (D+1)/2 \rfloor}^{\text{sort}} + s_{\lceil (D+1)/2 \rceil}^{\text{sort}} \right). \quad (3)$$

The sample median is robust to outliers [18], [19], which makes it amenable to System Model 1, as the nonzero entries of the sparse vector $\mathbf{s}$ can be considered to be outliers for the purpose of estimating the noise level. We emphasize that the sample median can be computed at low complexity in $\mathcal{O}(D)$ average time using quickselect [20] or in $\mathcal{O}(D)$ deterministic time using the MedianOfNinthers algorithm [21].

Estimator 1 (Average Noise Power). Consider System Model 1. We propose the following blind estimator

$$\hat{N}_0 = \frac{\bar{m}(|\mathbf{y}|^2)}{\log(2)}, \quad (4)$$

to estimate the average noise power defined as

$$N_0 = \frac{1}{D} \mathbb{E}[\|\mathbf{n}\|^2]. \quad (5)$$

Estimator 1 only requires the absolute square entries of the noisy observation $\mathbf{y}$ in (1) and can be computed efficiently. The estimator exploits sparsity in the signal $\mathbf{s}$, but is independent of the sparsity level K , the signal power, or the statistical sparsity model. It is, however, important to understand that the accuracy of this estimator depends on all of these factors as it relies on the fact that the nonzero entries of the sparse vector $\mathbf{s}$ can be treated as outliers when estimating the average noise power. We note that this noise variance estimator can be seen as a complex-valued (and squared) generalization¹ of the conventional median absolute deviation (MAD) [18], where we use the assumption that the noise in System Model 1 is zero-mean. The intuition behind this estimator (and the $\log(2)$ scaling factor) is the fact that the entries $|n_d|^2/(N_0/2)$, $d = 1, \dots, D$ are χ^2 distributed with two degrees of freedom, which have a median of $2 \log(2)$. A detailed derivation of this blind estimator and an analysis are provided in Section III-C.

¹The squared median absolute deviation (MAD) [19] corresponds to $\bar{m}(|\mathbf{y}|)^2$ whereas we propose to use $\bar{m}(|\mathbf{y}|^2)$. While for even dimensions D we have $\bar{m}(|\mathbf{y}|)^2 \leq \bar{m}(|\mathbf{y}|^2)$, both estimators coincide when D is odd. What is more, our scaling factor $\log(2)$ differs from the widely-used scale factor of $(\Phi^{-1}(3/4))^2 \approx (0.6745)^2$ [22]. The latter is derived for variance estimation of *real-valued* Gaussians using MAD, while in our derivation we consider the case of *complex-valued* Gaussians.

Estimator 2 (Average Signal Power). *Consider System Model 1. We propose the following blind estimator*

$$\hat{E}_s = \left[\frac{\|\mathbf{y}\|^2}{D} - \hat{N}_0 \right]_+ \quad (6)$$

for the average signal power defined as

$$E_s = \frac{1}{D} \mathbb{E}[\|\mathbf{s}\|^2]. \quad (7)$$

Estimator 2 only uses the sample estimate of the average receive power $\bar{E}_y = \frac{1}{D} \|\mathbf{y}\|^2$ and the blind noise estimate $\hat{N}_0$ from Estimator 1. The derivation of this estimator and an analysis of its key properties are provided in Section III-D.

Estimator 3 (Signal-to-Noise Ratio). *Consider System Model 1. We propose the following blind estimator*

$$\widehat{SNR} = \left[\frac{\|\mathbf{y}\|^2}{D\hat{N}_0} - 1 \right]_+ \quad (8)$$

for the SNR defined as

$$SNR = \frac{\mathbb{E}[\|\mathbf{s}\|^2]}{\mathbb{E}[\|\mathbf{n}\|^2]}. \quad (9)$$

Estimator 3 is blind as it combines the sample estimate of the average receive power $\bar{E}_y = \frac{1}{D} \|\mathbf{y}\|^2$ and the blind estimate $\hat{N}_0$ from Estimator 1. The derivation of this estimator and an analysis of its key properties are provided in Section III-E.

Estimator 4 (Mean-Square Error). *Consider System Model 2 with a fixed function $\mu : \mathbb{C} \rightarrow \mathbb{C}$. We propose the following blind estimator*

$$\begin{aligned} \hat{E}_0 &= \frac{1}{D} \|\mu(\mathbf{y}) - \mathbf{y}\|_2^2 - \hat{N}_0 \\ &\quad + \frac{\hat{N}_0}{D} \sum_{d=1}^D \left(\frac{\partial \Re\{\mu(y_d)\}}{\partial \Re\{y_d\}} + \frac{\partial \Im\{\mu(y_d)\}}{\partial \Im\{y_d\}} \right) \end{aligned} \quad (10)$$

for the MSE defined as

$$E_0 = \frac{1}{D} \mathbb{E}[\|\mu(\mathbf{y}) - \mathbf{s}\|^2] = \frac{1}{D} \mathbb{E}[\|\mathbf{e}\|^2]. \quad (11)$$

Estimator 4 only uses the receive signal $\mathbf{y}$, the estimate $\hat{N}_0$ from Estimator 1, and the function μ . Note that if we use the identity function $\mu(\mathbf{y}) = \mathbf{y}$, then the MSE corresponds to $E_0 = N_0$ while the estimated MSE corresponds to $\hat{E}_0 = \hat{N}_0$, as expected. The derivation of this estimator and an analysis of its key properties are provided in Section III-F.

III. THEORY

A. Convergence of the Sample Median for $D \rightarrow \infty$

We will use the following definition of the median.

Definition 2 (Median). *Let X be an absolutely continuous random variable (RV) with cumulative distribution function (CDF) $F_X(x)$. Then, the median m_X of X is defined as*

$$F_X(m_X) = \frac{1}{2}. \quad (12)$$

We will frequently make use of the following result.

Lemma 1 (Lemma C.1 from [23]). *Suppose that $f_X(x)$ is a differentiable probability density function (PDF) in some neighborhood of the median m_X , and vector $\mathbf{x}$ contains i.i.d.*

samples of X . Then, for any $c > 0$ the sample median $\bar{m}(\mathbf{x})$ satisfies

$$\lim_{D \rightarrow \infty} \Pr[|\bar{m}(\mathbf{x}) - m_X| \geq c] = 0. \quad (13)$$

In words, Lemma 1 implies that in the large-dimension limit, i.e., when $D \rightarrow \infty$, the sample median $\bar{m}(\mathbf{x})$ converges to the median m_X . Hence, by observing a large number of samples, which is possible in modern multi-antenna mmWave or OFDM systems, we can accurately estimate the true median.

B. Statistical Model for Complex-Valued Sparse Vectors

In order to derive and analyze the blind estimators proposed in Section II, we need a suitable statistical model for the sparse signal $\mathbf{s}$. The statistical model should (i) have as few parameters as possible while being able to model a large class of complex-valued sparse vectors typically arising in communication systems and (ii) facilitate a theoretical analysis. In what follows, we consider Bernoulli complex Gaussian (BCG) random vectors [9], [24], which allow control over the signal sparsity and the signal power.

Definition 3 (BCG Random Vector). *Each entry in the sparse vector $\mathbf{s} \in \mathbb{C}^D$ is nonzero with activity rate $p \in (0, 1]$, and the nonzero entries are i.i.d. circularly-symmetric complex Gaussian with variance E_s/p . The PDF of each entry s_d , $d = 1, \dots, D$, is therefore given by*

$$f_S(s_d) = (1-p)\delta(s_d) + p \frac{1}{\pi E_s/p} e^{-\frac{|s_d|^2}{E_s/p}}, \quad (14)$$

where $\delta(\cdot)$ is the Dirac delta function.

With this statistical model, the expected number of nonzero entries (the sparsity) is $K = pD$ and the average power of the sparse signal vector $\mathbf{s}$ corresponds to $E_s = \frac{1}{D} \mathbb{E}[\|\mathbf{s}\|^2]$.

In System Model 1, we assumed that the noise vector $\mathbf{n}$ is i.i.d. circularly-symmetric complex Gaussian with variance N_0 per complex entry. Hence, the PDF of each entry is given by $f_N(n_d) = \frac{1}{\pi N_0} e^{-|n_d|^2/N_0}$. Consequently, the PDF of the noisy observation vector $\mathbf{y} = \mathbf{s} + \mathbf{n}$ in (1) is as follows.

Definition 4 (Noisy BCG Random Vector). *The PDF of the entries y_d , $d = 1, \dots, D$, of a BCG random vector per Definition 3 observed as in System Model 1 is given by*

$$f_Y(y_d) = (1-p) \frac{1}{\pi N_0} e^{-\frac{|y_d|^2}{N_0}} + p \frac{1}{\pi(N_0 + E_s/p)} e^{-\frac{|y_d|^2}{N_0 + E_s/p}}. \quad (15)$$

For this signal and observation model, we are now able to derive and analyze the proposed blind estimators.

C. Analysis of Estimator 1

We start with the blind noise variance estimator defined in Estimator 1. We have the following key result. The proof is given in Appendix A.

Theorem 1. *Let $\mathbf{y}$ be a noisy BCG random vector with PDF as in Definition 4 and with activity rate*

$$p \leq \frac{1/2 - e^{-2}}{1 - e^{-2}} \approx 0.421. \quad (16)$$

Then, the average noise variance N_0 satisfies

$$\frac{m_Z}{\min\{\log(\frac{2-2p}{1-2p}), \log(2)(1+SNR)\}} \leq N_0$$

$$\leq \frac{m_Z}{\log(2)} \left((1-p) + \frac{p^2}{p+SNR} \right), \quad (17)$$

where m_Z is the median of an entry z_d of $\mathbf{z} = |\mathbf{y}|^2$.

It is important to realize that for $SNR \rightarrow 0$ or $p \rightarrow 0$, the lower and upper bounds in (17) coincide and equal $m_Z / \log(2)$. Furthermore, we reiterate that in the large-dimension limit, i.e., when $D \rightarrow \infty$, the sample median $\overline{m}(\mathbf{z})$ converges to the true median m_Z as established by Lemma 1. Thus, in these two cases the proposed estimate corresponds to the true noise variance $N_0 = m_Z / \log(2)$. Hence, by assuming the large-dimension limit, Theorem 1 has the following key implications on Estimator 1: (i) Since

$$N_0 \stackrel{(a)}{\leq} \frac{m_Z}{\log(2)} \left((1-p) + \frac{p^2}{p+SNR} \right) \stackrel{(b)}{\leq} \frac{m_Z}{\log(2)} = \hat{N}_0, \quad (18)$$

where (a) follows from (17) and (b) holds for every valid value of p and SNR , the proposed blind estimate $\hat{N}_0$ bounds the average noise variance N_0 from above, i.e., we have developed a pessimistic estimator. (ii) By letting $p \rightarrow 0$, we have that $N_0 = m_Z / \log(2)$ and the inequalities in (18) hold with equality; this implies that the average noise variance will approach the value of the blind estimator for sparse signals (irrespective of the SNR). (iii) By letting $SNR \rightarrow 0$, we have that $N_0 = m_Z / \log(2)$ and the inequalities in (18) hold with equality; this implies that the noise variance will approach the value of the blind estimator at low SNR (irrespective of the signal's sparsity). In summary, the proposed noise variance estimate is pessimistic but accurate for sparse vectors or in low-SNR scenarios for high-dimensional data. While the above observations are only true for our blind estimator in the large-dimension limit and for the noisy BCG model in Definition 4, we will use simulations in Section IV to demonstrate the accuracy of Estimator 1 for finite (and small) dimensions D and showcase an application example for channel-vector denoising in multi-antenna mmWave communication systems.

D. Analysis of Estimator 2

For the blind estimate $\hat{E}_s$ of the average signal power E_s , we use the following standard result, which follows from the fact that the entries of the vector $\mathbf{z} = |\mathbf{y}|^2$ are i.i.d. with expected value of $\mathbb{E}[z_d] = E_s + N_0$, $d = 1, \dots, D$.

Lemma 2. *Let $\mathbf{y}$ be a noisy BCG random vector with PDF as in Definition 4. Then, we have that*

$$\lim_{D \rightarrow \infty} \frac{1}{D} \|\mathbf{y}\|^2 \xrightarrow{a.s.} E_s + N_0, \quad (19)$$

where $\xrightarrow{a.s.}$ implies almost sure convergence.

By defining the RV $W_D = \frac{1}{D} \|\mathbf{y}\|^2 - N_0$, we have that $\lim_{D \rightarrow \infty} W_D \xrightarrow{a.s.} E_s$. By replacing the average noise power N_0 by the blind estimate $\hat{N}_0$ in (4) from Estimator 1 and by clipping the result, we obtain Estimator 2 in (6). Since, in the large-dimension limit, the noise power estimate $\hat{N}_0$ is

overestimating the true average noise power, the blind estimate $\hat{E}_s$ in (6) tends to underestimate the signal power. From Theorem 1 it follows that for $p \rightarrow 0$ or $SNR \rightarrow 0$, the blind signal power estimate is exact. While the above observations only hold for $D \rightarrow \infty$, we showcase their accuracy for finite dimensions D in Section IV.

E. Analysis of Estimator 3

The blind SNR estimator is obtained by simply taking the ratio of $\hat{E}_s$ in (6) and $\hat{N}_0$ in (4). For $D \rightarrow \infty$, the blind signal estimate underestimates the average signal power and the noise power estimate overestimates the average noise power, which means that the blind SNR estimate in (8) underestimates the SNR. From Theorem 1 it follows that for $D \rightarrow \infty$ with either $p \rightarrow 0$ or $SNR \rightarrow 0$ the blind SNR estimate is exact. We provide simulation results for finite dimensions D in Section IV.

F. Analysis of Estimator 4

In order to analyze Estimator 4, we first assume that the average noise power N_0 is known. For this scenario, we can borrow the following two theorems from [13].

Theorem 2 (Thm. 1 of [13]). *Let $\mathbf{y}$ be a noisy random vector of observations of $\mathbf{s}$ as in System Model 1, and apply a weakly differentiable function $\mu : \mathbb{C} \rightarrow \mathbb{C}$ to the entries of $\mathbf{y}$ as in System Model 2. Then, Stein's unbiased risk estimate given by*

$$SURE = \frac{1}{D} \|\mu(\mathbf{y}) - \mathbf{y}\|_2^2 - N_0$$

$$+ \frac{N_0}{D} \sum_{d=1}^D \left(\frac{\partial \Re\{\mu(y_d)\}}{\partial \Re\{y_d\}} + \frac{\partial \Im\{\mu(y_d)\}}{\partial \Im\{y_d\}} \right), \quad (20)$$

is an unbiased estimate of the MSE so that $\mathbb{E}[SURE] = E_0$.

Under the same assumptions, we have the following result which characterizes the behavior in the large-dimension limit.

Theorem 3 (Thm. 3 of [13]). *In the large-dimension limit, i.e., when $D \rightarrow \infty$, SURE in (20) converges to the MSE in (11), i.e., we have $\lim_{D \rightarrow \infty} SURE = E_0$.*

These results imply that if N_0 were known perfectly, one can estimate the MSE without knowing the sparse signal vector $\mathbf{s}$ in the large-dimension limit when $D \rightarrow \infty$. To obtain the blind version in Estimator 4, we have replaced the true average noise power N_0 by its estimate $\hat{N}_0$. Consequently, for $D \rightarrow \infty$ and either $p \rightarrow 0$ or $SNR \rightarrow 0$, Theorem 1 implies that $\hat{N}_0$ will be exact, which implies that Estimator 4 will be exact in this scenario. To demonstrate the efficacy of this estimator in finite dimensions, we will show synthetic results and an application for mmWave channel-vector denoising in Section IV.

IV. NUMERICAL RESULTS

A. Synthetic Results

We perform Monte-Carlo simulations with 10,000 trials to characterize the accuracy of the estimators proposed in Section II-B. We use the sparse signal model in Definition 4 and fix $N_0 = 1$, without loss of generality. We show the

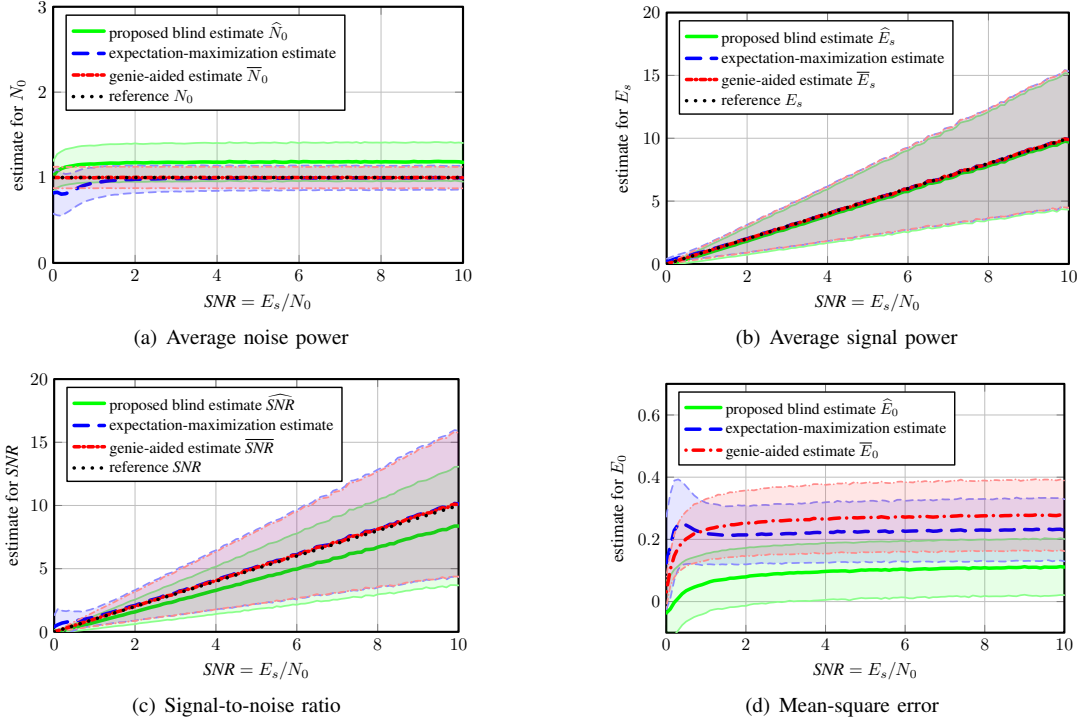


Fig. 1. Effect of varying the SNR on the proposed low-complexity blind estimators for the average noise power, signal power, SNR, and MSE.

effect of the $SNR = E_s/N_0$ on the proposed estimators for an activity rate of $p = 0.1$ and a dimension of $D = 64$.

Figure 1 shows the accuracy when varying the SNR. The results for our blind estimators are shown in green, where thick solid lines refer to the average performance and light green areas indicate the standard deviation. For the MSE E_0 , we pick the soft-thresholding function $\mu(\mathbf{y}) = \eta(\mathbf{y}; \tau)$. Furthermore, for each received vector $\mathbf{y}$, we adaptively select the denoising threshold $\tau \geq 0$ that minimizes the estimated MSE $\hat{E}_0$ as done in [11]. As a comparison, we include an EM implementation specialized for circularly-symmetric complex Gaussian mixtures (average performance shown with a blue dashed line and standard deviation with a light blue area) with a maximum of 30 iterations and early stopping if the total parameter change is below 0.1%. We also show the accuracy of genie-aided estimators (average performance shown with a red dash-dotted line and standard deviation with a light red area) that have separate knowledge of $\mathbf{n}$ and $\mathbf{s}$. We compute $\bar{E}_s = \frac{1}{D} \|\mathbf{s}\|^2$, $\bar{N}_0 = \frac{1}{D} \|\mathbf{n}\|^2$, $\bar{SNR} = \frac{\bar{E}_s}{\bar{N}_0}$, $\bar{E}_0 = \frac{1}{D} \|\eta(\mathbf{y}; \tau) - \mathbf{s}\|^2$. The reference parameters used in our simulations are shown with black dotted lines. Note that there is no reference value for E_0 , as the adaptive threshold prevents us from computing E_0 analytically.

From Figure 1, we observe the following facts: (i) The standard deviation of the proposed blind estimators is comparable to that of the genie-aided methods that have separate access to $\mathbf{n}$ and $\mathbf{s}$. (ii) Even though the sample size is small ($D = 64$), our estimators are quite accurate with a standard deviation comparable to that of the genie-aided estimators; increasing D would further reduce the standard deviation of all considered

estimators. (iii) As predicted by our theory, the average noise power is overestimated while the signal power and SNR are underestimated. At low SNR, the three become exact. (iv) For the considered scenario, the blind MSE estimate has a negative offset. However, the key requirement for adaptive parameter tuning is that the estimated MSE $\hat{E}_0$ (which is the function to be minimized) has a similar shape as E_0 and thus, for this purpose, we do not clip negative values. As we show next, the MSE estimate still performs well in practice.

In comparison with EM, our method provides a less-accurate estimate at higher SNRs, but requires significantly lower complexity. The complexity (in terms of the number of real-valued additions, real-valued multiplications, and exponentials) of EM is more than $N(16D + 12) + 3D$ operations, where N is the number of EM iterations—the average number of iterations observed in our simulations ranges from 8 to 28 depending on the SNR. In contrast, our proposed median-based noise estimator has an average complexity of no more than $7.7D + 9$ operations, when computing the median using quickselect [20]. Hence, our proposed blind estimator is more than $17\times$ less complex than EM (and avoids the evaluation of complex operations such as exponentials and divisions), which renders our method suitable for (i) low-complexity parameter estimation and (ii) as a potential initializer for EM-based methods—the latter aspect is part of ongoing research.

B. Application to Nonparametric Channel-Vector Denoising

We now show an application of Estimator 4 for beamspace channel estimation. As in [11], we simulate an uplink massive MIMO system in which 8 single-antenna user equipments (UEs) transmit pilots and data to a BS equipped with a uniform

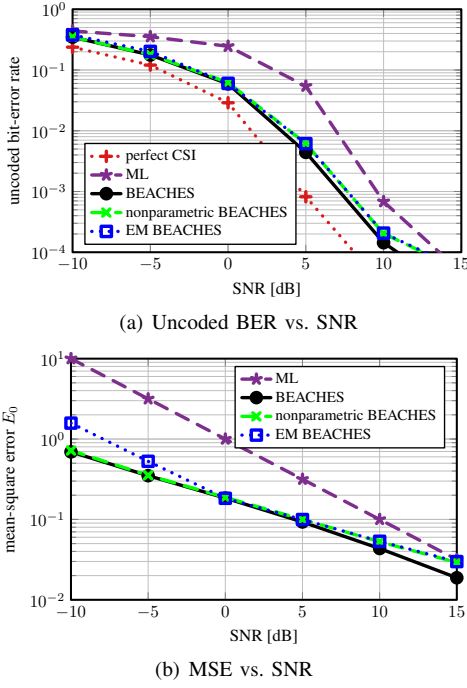


Fig. 2. Uncoded BER (a) and MSE (b) of mmWave channel estimation methods, including the nonparametric BEACHES variant which estimates the noise variance and denoising parameter directly from the receive vector.

linear array of $D = 128$ antenna elements. We consider channel estimation with orthogonal pilots. The maximum likelihood (ML) estimate of the channel matrix is obtained by right-multiplying the received pilot sequence with the inverse of the orthogonal pilot matrix. In this case, the beamspace representation of the estimated channel matrix $\tilde{\mathbf{H}}$ is given by $\tilde{\mathbf{H}} = \mathbf{H} + \tilde{\mathbf{N}}$, where $\mathbf{H} \in \mathbb{C}^{128 \times 8}$ is the beamspace channel matrix, $\tilde{\mathbf{N}}$ is complex Gaussian noise, and $\tilde{\mathbf{H}}$ is the beamspace ML channel estimate, which is a noisy observation of $\mathbf{H}$.

Since electromagnetic waves at high frequencies experience strong attenuation and little scattering or diffraction, typical mmWave channels consist only of a small number of dominant propagation paths. As each row index in the beamspace channel matrix corresponds to an angle-of-arrival to the BS, each column of the beamspace channel matrix $\mathbf{H}$ (which is the *channel vector* of one UE) will be sparse. As a consequence, by writing each column of $\tilde{\mathbf{H}}$ as an independent equation, we can express the channel estimation problem in the form of System Model 1, where each column of the channel matrix (that contains only few nonzero entries) corresponds to the sparse signal $\mathbf{s}$. This observation implies that we can perform denoising in order to improve the ML channel estimate.

Figure 2 shows simulation results for 10,000 Monte-Carlo trials with line-of-sight (LoS) mmMAGIC QuaDRiGa mmWave channels [25]. For different channel estimation methods we compute the MSE, and the bit-error rate (BER) with linear minimum MSE equalization using the estimated channels and uncoded 16-QAM transmission. We simulate beamspace channel estimation (BEACHES) as in [11], which denoises the columns of the beamspace ML estimate $\tilde{\mathbf{H}}$ by applying the soft-thresholding function $\eta(x; \tau)$. The thresholding

parameter τ is adaptively selected for each noisy observation by minimizing SURE with perfect knowledge of the average noise power N_0 using an $\mathcal{O}(D \log(D))$ algorithm. We compare this to a new, nonparametric BEACHES variant, where we apply soft-thresholding denoising to the beamspace channel vectors and use the (nonparametric) threshold τ that minimizes Estimator 4, which is a nonparametric version of SURE. We also include a variant that we call EM BEACHES, where we use Estimator 4 but replace $\hat{N}_0$ by the EM noise power estimate. The methods described above, after denoising the beamspace channel vectors, use the inverse Fourier transform to obtain an antenna-domain channel estimate. As a reference, we show the performance of perfect channel state information (CSI) that uses the ground truth (noiseless) channel vector, and maximum likelihood (ML) estimation that simply takes the inverse Fourier transform of the beamspace noisy observation $\tilde{\mathbf{H}}$ as the antenna-domain channel estimate.

From Figure 2, we observe that the nonparametric BEACHES algorithm achieves virtually the same performance as that of the original BEACHES algorithm which requires knowledge of N_0 (except at high SNR where Estimator 1 tends to overestimate N_0). We reiterate that the nonparametric BEACHES algorithm requires no parameters and exhibits exactly the same complexity as the original algorithm as the latter already sorts the entries of $|\mathbf{y}|^2$, which we can reuse to compute Estimator 1. EM BEACHES achieves higher (worse) MSE, as realistic channels deviate from the BCG model in Definition 3, and exhibits higher complexity than our nonparametric method.

V. CONCLUSIONS

We have proposed blind estimators for the average noise power, signal power, SNR, and MSE. Our estimators can be calculated in $\mathcal{O}(D)$ time and only require the noisy observation vector, which avoids the need for additional pilot signals. We have analyzed our estimators for a complex Bernoulli-Gaussian sparsity model and evaluated their accuracy via simulations. Using a channel-vector denoising task in multi-antenna mmWave systems, we have demonstrated that our blind estimators lead to a novel nonparametric denoiser that achieves comparable performance and the same complexity as BEACHES in [11], [13] which requires knowledge of the average noise power. We believe that the proposed blind estimators find potential use in a large number of other applications in wireless communication systems that contain sparse signals and require low complexity.

APPENDIX A PROOF OF THEOREM 1

A. Prerequisites

In what follows, we need the distribution of $\mathbf{z} = |\mathbf{y}|^2$. Since the absolute-square entry of a circularly-symmetric complex Gaussian RV Q with variance E_q is exponentially distributed with CDF $F_Q(q) = 1 - e^{-\frac{q}{E_q}}$, $q \geq 0$, the CDF of each entry of the absolute-square noisy observation is as follows.

Definition 5 (Noisy BCG Power RV). Let $\mathbf{y}$ be as in Definition 4 and let $\mathbf{z} = |\mathbf{y}|^2$. Then, for $z \geq 0$, the CDF of each entry of $\mathbf{z}$ is given by

$$F_Z(z_d) = (1-p)(1 - e^{-\frac{z_d}{N_0}}) + p(1 - e^{-\frac{z_d}{N_0 + E_s/p}}). \quad (21)$$

B. Upper Bounds on the Median

We start with the following upper bound on the median m_Z of a noisy BCG power RV Z with CDF given in (21).

Lemma 3. For a noisy BCG power RV in Definition 5 with $p < 0.5$, the median is bounded from above by

$$N_0 \log\left(\frac{2-2p}{1-2p}\right) \geq m_Z. \quad (22)$$

Proof. Using (21), we obtain the expression for the median of a RV Z with CDF in (21) according to equation (12):

$$(1-p)(1 - e^{-\frac{m_Z}{N_0}}) + p(1 - e^{-\frac{m_Z}{N_0 + E_s/p}}) = \frac{1}{2}. \quad (23)$$

Since the second term is nonnegative, we can omit it to obtain the following inequality:

$$(1-p)(1 - e^{-\frac{m_Z}{N_0}}) \leq \frac{1}{2}. \quad (24)$$

Note that this bound will be useful for vectors $\mathbf{s}$ that are sparse, i.e., where p is small. We can simplify (24) as follows

$$\frac{m_Z}{N_0} \leq -\log\left(\frac{1-2p}{2-2p}\right) \quad (25)$$

which leads to an upper bound on the median m_Z . In order to take the logarithm in (25), we require $p \in (0, 1/2)$. $\square$

Lemma 4. For a noisy BCG power RV Z in Definition 5 with $p \leq \frac{1/2 - e^{-2}}{1 - e^{-2}}$, the median is bounded from above by

$$\log(2)(N_0 + E_s) \geq m_Z. \quad (26)$$

Proof. From (21), we have that

$$\frac{1}{2} = (1-p)e^{-\frac{m_Z}{N_0}} + pe^{-\frac{m_Z}{N_0 + E_s/p}}. \quad (27)$$

Let us define the function $g(r) = e^{-1/r}$ with $r > 0$. We can now rewrite (27) as follows:

$$\frac{1}{2} = (1-p)g\left(\frac{N_0}{m_Z}\right) + pg\left(\frac{N_0 + E_s/p}{m_Z}\right). \quad (28)$$

The function $g(r)$ is concave for $r \geq 1/2$, which holds when

$$2N_0 \geq m_Z. \quad (29)$$

We first verify when (29) holds. From (27), we have that

$$\frac{1}{2} \geq (1-p)e^{-\frac{2N_0}{m_Z}} + pe^{-\frac{2N_0}{N_0 + E_s/p}} \quad (30)$$

$$\frac{1}{2} - e^{-2} \geq p(1 - e^{-2}), \quad (31)$$

which implies that the condition (16) ensures concavity of $g(r)$. By assuming that the condition (16) holds, we can use Jensen's inequality on the expression in (28) to get

$$\frac{1}{2} \leq g\left((1-p)\frac{N_0}{m_Z} + p\frac{N_0 + E_s/p}{m_Z}\right) = e^{-m_Z \frac{1}{N_0 + E_s}}. \quad (32)$$

We can now simplify this expression to

$$\log(1/2) \leq -m_Z \frac{1}{N_0 + E_s} \quad (33)$$

$$\log(2)(N_0 + E_s) \geq m_Z, \quad (34)$$

which is what we show in Lemma 4. $\square$

C. Lower Bound on the Median

We now establish the following lower bound on the median.

Lemma 5. For a noisy BCG power RV Z in Definition 5 with $p \in (0, 1]$, the median is bounded from below by

$$\frac{\log(2)N_0}{(1-p) + \frac{p^2}{p+SNR}} \leq m_Z. \quad (35)$$

Proof. Since the exponential CDF $F_Q(q) = 1 - e^{-\frac{q}{E_q}}$ for $E_q \geq 0$ is concave in q , Jensen's inequality leads to

$$(1-p)(1 - e^{-\frac{m_Z}{N_0}}) + p(1 - e^{-\frac{m_Z}{N_0 + E_s/p}}) = \frac{1}{2} \quad (36)$$

$$1 - e^{-(1-p)\frac{m_Z}{N_0} - p\frac{m_Z}{N_0 + E_s/p}} \geq \frac{1}{2}. \quad (37)$$

We can simplify this expression to obtain the following bound

$$\frac{1}{2} \geq e^{-(1-p)\frac{m_Z}{N_0} - p\frac{m_Z}{N_0 + E_s/p}} \quad (38)$$

$$\log(2)N_0 \leq m_Z \left((1-p) + \frac{p^2}{p+SNR} \right), \quad (39)$$

which is what we have in (35) $\square$

D. Combining the Results

Finally, we can combine Lemma 3 with Lemma 4 and Lemma 5 to obtain the desired result in (17).

REFERENCES

- [1] T. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer Science & Business Media, 2008.
- [2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [3] F. Socheleau, D. Pastor, A. Aissa-El-Bey, and S. Houcke, "Blind noise variance estimation for OFDMA signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2009, pp. 2581–2584.
- [4] M. Zivkovic and R. Mathar, "Preamble-based SNR estimation in frequency selective channels for wireless OFDM systems," in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, Apr. 2009, pp. 1–5.
- [5] A. Das and B. D. Rao, "SNR and noise variance estimation for MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3929–3941, Apr. 2012.
- [6] S. Huang and T. D. Tran, "1-bit compressive sensing via approximate message passing with built-in parameter estimation," *arXiv preprint: 2007.07679*, Jul. 2020.
- [7] R. Ghods, C. Jeon, G. Mirza, A. Maleki, and C. Studer, "Optimally-tuned nonparametric linear equalization for massive MU-MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2118–2122.
- [8] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," in *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, Nov. 2009, pp. 18914–18919.
- [9] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2011, pp. 2168–2172.
- [10] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.
- [11] R. Ghods, A. Gallyas-Sanhueza, S. H. Mirfarshbafan, and C. Studer, "BEACHES: Beamspace channel estimation for multi-antenna mmWave systems and beyond," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.
- [12] A. Gallyas-Sanhueza, S. H. Mirfarshbafan, R. Ghods, and C. Studer, "Sparsity-adaptive beamspace channel estimation for 1-bit mmWave massive MIMO systems," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, May 2020, pp. 1–5.
- [13] S. H. Mirfarshbafan, A. Gallyas-Sanhueza, R. Ghods, and C. Studer, "Beamspace channel estimation for massive MIMO mmWave systems: Algorithm and VLSI design," *IEEE Trans. Circuits Sys. I (TCAS-I)*, vol. 67, no. 12, pp. 5482–5495, Sep. 2020.

- [14] K. Upadhyay, C. S. Seelamantula, and K. Hari, "A risk minimization framework for channel estimation in OFDM systems," *Elsevier Signal Process.*, vol. 128, pp. 78–87, Nov. 2016.
- [15] S. Jacobsson, U. Gustavsson, G. Durisi, and C. Studer, "Massive MU-MIMO-OFDM uplink with hardware impairments: Modeling and analysis," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1829–1835.
- [16] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. 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.
- [17] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Apr. 2017.
- [18] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *J. American Stat. Association*, vol. 88, no. 424, pp. 1273–1283, Sep. 1993.
- [19] P. J. Huber, *Robust statistics*. Wiley, 2004.
- [20] R. J. Tibshirani, "Fast computation of the median by successive binning," *arXiv preprint: 0806.3301*, May 2009.
- [21] A. Alexandrescu, "Fast deterministic selection," in *Int. Symp. Exp. Algor. (SEA)*, vol. 75, Aug. 2017, pp. 24:1–24:19.
- [22] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Sep. 1994.
- [23] A. Merberg and S. J. Miller, "The sample distribution of the median," Course Notes for Math 162: Mathematical Statistics, Williams College, Tech. Rep., Feb. 2008.
- [24] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2011, pp. 799–803.
- [25] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "QuaDRiGa - quasi deterministic radio channel generator, user manual and documentation," Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2.2.0, Jun. 2019.