Atomic Norm Denoising-Based Joint Channel Estimation and Faulty Antenna Detection for Massive MIMO

Peng Zhang, Lu Gan, Cong Ling and Sumei Sun

Abstract—We consider joint channel estimation and faulty antenna detection for massive multiple-input multiple-output (MIMO) systems operating in time-division duplexing (TDD) mode. For systems with faulty antennas, we show that the impact of faulty antennas on uplink (UL) data transmission does not vanish even with unlimited number of antennas. However, the signal detection performance can be improved with a priori knowledge on the indices of faulty antennas. This motivates us to propose the approach for simultaneous channel estimation and faulty antenna detection. By exploiting the fact that the degrees of freedom of the physical channel matrix are smaller than the number of free parameters, the channel estimation and faulty antenna detection can be formulated as an extended atomic norm denoising problem and solved efficiently via the alternating direction method of multipliers (ADMM). Furthermore, we improve the computational efficiency by proposing a fast algorithm and show that it is a good approximation of the corresponding extended atomic norm minimization method. Numerical simulations are provided to compare the performances of the proposed algorithms with several existing approaches and demonstrate the performance gains of detecting the indices of faulty antennas.

Index Terms—Channel estimation, atomic norm, massive MIMO, antenna failure.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) or large antenna systems have been proposed in [1], where each base station (BS) is equipped with a large number of antennas that greatly exceeds the number of user terminals (UTs). As a potential enabler for the development of future broadband network, it offers many advantages over conventional point-to-point MIMO, such as energy-efficiency, security, and robustness [2]. However, the spatial multiplexing gains and the array gains of massive MIMO rely on accurate channel state information at the transmitter (CSIT) [2], [3].

In time-division duplexing (TDD) massive MIMO systems, only CSI for the uplink needs to be estimated due to the assumption of channel reciprocity. Conventional methods on

P. Zhang is with the Institute for Infocomm Research, A*STAR, Singapore 138632, and also with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK (e-mail: zhangp@i2r.a-star.edu.sg; p.zhang12@imperial.ac.uk).

L. Gan is with the College of Engineering, Design and Physical Science, Brunel University, London, UB8 3PH, UK (e-mail: lu.gan@brunel.ac.uk).

C. Ling is with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK (e-mail: cling@ieee.org) .

S. Sun is with the Institute for Infocomm Research, A*STAR, Singapore, 138632, Singapore (e-mail: sunsm@i2r.a-star.edu.sg).

massive MIMO channel estimation include least squares (LS) and minimum mean squared error (MMSE) [4]–[6]. A compressed sensing (CS) based approach has been proposed to address the channel estimation for massive MIMO systems with a physical finite scattering channel model [7]. By employing random Bernoulli pilots at each user terminal (UT), the channel estimation was converted into a low-rank approximation problem and solved via quadratic semidefinite programming (SDP). Recently, a novel atomic norm denoising-based channel estimation method was proposed to further improve the performance [8]. In [8], it is demonstrated that the atomic norm denoising-based approach can outperforms the least squares, the least squares scaled least squares [9] and the compressed sensing-based methods [7] [10]. Similar atomic norm denoising-based channel estimation method has also been used for OFDM systems in [11], [12].

However, an important issue overlooked by research on channel estimation for massive MIMO is the impact of faulty antennas at the BS. The use of large-scale antenna arrays brings many remarkable features, such as propagation loss mitigation and low inter-user interference [1]–[3]. For the consideration of network deployment, it might be more attractive and practical to use large-scale antenna arrays with inexpensive antenna elements. Thus, there is always a possibility of failure of one or more elements in the large array [13]–[16]. It is therefore important to analyze the impact of faulty antennas on the channel estimation accuracy and data transmission quality. Moreover, it would be significant to promptly identify the faulty antennas, which is crucial for replacement or reconfiguration. We note that the influence of hardware impairments (e.g., amplifier nonlinearities, I/Q imbalances, and quantization errors) has been considered in [17]–[21]. Specifically, [21] analyzed the aggregate impacts of different hardware impairments on massive MIMO systems. However, the problem of hardware impairments is different from the impact of faulty antennas: the distortion noise caused by the hardware impairments is modeled as additive noise (e.g., Gaussian distribution conditioned on given channel realization) and applied to all the antennas at the BS [21]–[23]; on the other hand, the corruption noise due to faulty antennas usually appears only on a small subset of the antenna array, but with unknown distribution and unbounded magnitudes (see Section II for the system model).

In this paper, we consider the channel estimation for massive MIMO systems under the impact of faulty antennas. We analyze the mean squared error of the signal detection with

maximum ratio combining (MRC) and zero-forcing (ZF) receivers on the uplink (UL) in the large system limit. It has been shown in [1] that a linear receiver can be used to eliminate the effect of uncorrelated noise and small-scale fading when considering an infinite number of fault-free antennas. Here, we prove that the effect of faulty antennas does not vanish even with unlimited number of antennas. Furthermore, we show that the knowledge on the indices of faulty antennas can be used to modify the receiver for better signal detection performance. In particular, the resultant MSE is always smaller than the original one in the large system limit. It is noted that the improvement on signal detection depends on the ability to acquire the indices of faulty antennas. Our main contribution is an atomic denoising-based method that can estimate the channel and detect faulty antennas simultaneously in the training phase. Relying on this advantage, the system will be able to: 1) monitor the antenna array; 2) identify the faulty antennas; and 3) maintain good signal detection performance before repair. In addition, we consider reducing the computational complexity of the proposed method for practical implementation. We propose a fast algorithm based on the group lasso and prove that it can achieve nearly the same performance as the atomic norm-based method.

Additional related work. Detection of the faulty elements in large antenna arrays has been considered in [24]–[28]. In these works, different techniques are proposed to identity the faulty elements based on the measurements of the "fault-free" and "distorted" radiation patterns, e.g., back-propagation [25], genetic algorithms [26] and Bayesian compressed sensing [28]. In contrast, our work focuses on simultaneous channel estimation and faulty antenna detection from the received signals in pilot-aided training. In addition, our atomic norm denoising-based approaches are novel in the faulty antennas detection literature. On the other hand, antenna selection has been studied to help to reduce the system complexity and cost by using a smaller number of RF chains than the antennas [29]–[32]. This is different from our work since they assume that all the antennas are working and the CSI over all the antennas is available.

Notations. We denote matrices and vectors by uppercase and lowercase boldface letters respectively. For a vector $\mathbf{a}$, we denote by a_i, $(i \in [n] = \{0, ..., n-1\})$, its i-th element. We represent a sequence of vectors by $\mathbf{a}_0, ..., \mathbf{a}_{n-1}$. For a matrix $\mathbf{A}$, $\mathbf{A}_{(j,k)}$ denotes the element on its j-th row and k-th column. The vector obtained by taking the j-th row (k-th column) of $\mathbf{A}$ is represented by $\mathbf{A}_{(j,:)}$ ($\mathbf{A}_{(:,k)}$). $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$ and $\mathbf{A}^*$ represent the inverse, the pseudo-inverse and the conjugate transpose of $\mathbf{A}$. The Frobenius norm and spectral norm of $\mathbf{A}$ is denoted by $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}(\mathbf{A}^*\mathbf{A})}$ and $\|\mathbf{A}\|_2$ respectively.

Outline. The rest of the paper is organized as follows. In Section II, we first introduce the system and channel models, then review the concept of atomic norm and its application in channel estimation for faultless antennas. In Section III, we consider the channel estimation under faulty antennas. In particular, we analyse the impact of faulty antennas on UL data transmission, propose an approach to simultaneously estimate the channel and identify the faulty antennas, and provide a fast algorithm that can well approximate the atomic norm-based approach. Section IV demonstrates the performance of the proposed approaches through a series of numerical simulations, followed by discussions and conclusions in Section V and VI respectively.

II. PRELIMINARIES

A. Massive MIMO system model

For a massive MIMO system operating in TDD mode, only CSI for the uplink needs to be estimated due to the assumption of channel reciprocity. There is one BS equipped with a large number of antennas M. In the uplink, K autonomous single-antenna UTs send signals to the BS ($M \geq K$). At the t-th symbol time, the received signal $\mathbf{y}_t \in \mathbb{C}^M$ at the BS can be expressed as

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{n}_t, \qquad (1)$$

where $\mathbf{H} \in \mathbb{C}^{M \times K}$ is the flat-fading quasi-static channel matrix, $\mathbf{x}_t \in \mathbb{C}^K$ denotes the transmitted vector of the K users and $\mathbf{n}_t \in \mathbb{C}^M$ is the complex Gaussian noise with zero mean and unit variance. In the training phase, each user sends a training sequence of length L. The signal model (1) can be equivalently written as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}_0, \qquad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L] \in \mathbb{C}^{K \times L}$ is the training matrix, $\mathbf{Y} \in \mathbb{C}^{M \times L}$ is the received signal matrix and $\mathbf{N}_0 \in \mathbb{C}^{M \times L}$ is the noise matrix with independent and identically distributed (i.i.d.) $\mathcal{CN}(0, \sigma^2)$ entries. The training matrix $\mathbf{X} = [\boldsymbol{\phi}_1^T; \boldsymbol{\phi}_2^T; \ldots; \boldsymbol{\phi}_K^T]$ can be regarded as the collection of K length-L training sequences from the UTs, where $\boldsymbol{\phi}_k$ denotes the training sequence transmitted from user k.

In this paper, we consider systems with faulty antennas at the BS. Let $S = \gamma M$, $\gamma \in [0, 1]$, denote the number of faulty antennas, then at the t-th symbol time, the distortion caused by faulty antennas is denoted by an S-sparse vector $\mathbf{w}_t$, in which case the received signal becomes $\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{w}_t + \mathbf{n}_t$. The support of $\mathbf{w}_t$'s indicates the indices of faulty antennas, which is assumed to be arbitrary and static for a relative long period. However, at different symbol times, the magnitude of the sparse vectors may vary arbitrarily. The signal model in the training phase is given by

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}_0 + \mathbf{N}_0, \qquad (3)$$

where $\mathbf{W}_0 \in \mathbb{C}^{M \times L}$ is a matrix with S non-zero rows to account for the corruption caused by the S faulty antennas. The indices of non-zero rows of $\mathbf{W}_0$ are equivalent to the positions of faulty antennas at the BS. Our main task is to estimate the channel matrix $\mathbf{H}$ and the positions of faulty antennas based on the training matrix $\mathbf{X}$ and the received signal matrix $\mathbf{Y}$. We note that the system model (3) represents a general scenario, where no statistical knowledge on the channel matrix or the corruption is required.

B. Finite scattering channel model

In this paper, we consider a physical finite scattering model that has been frequently used in the literature for massive

MIMO systems [6], [7], [33]–[36]. This practical channel model is motivated by two reasons. First, it characterizes a poor scattering propagation environment, where the number of physical objects is limited [37]–[40]. Second, it can also describe the propagation channel where the scatters appear in groups (a.k.a. clusters) with similar delays, angle-of-arrivals (AoAs) and angle-of-departures (AoDs) [5], [41]. In this channel model, the angular domain is divided into a finite number of directions P, which is fixed and smaller than the number of BS antennas. For uniform linear arrays (ULA), the steering vector associated with each angle-of-arrival (AOA) $\theta_p \in [-\pi/2, \pi/2]$, $p = 1, \ldots, P$, is given by [34]

$$\mathbf{a}(\theta_p) \triangleq [1 \quad e^{-j2\pi\frac{D}{\lambda}\sin(\theta_p)} \ldots \quad e^{-j2\pi\frac{(M-1)D}{\lambda}\sin(\theta_p)}]^T, \quad (4)$$

where D is the antenna spacing at the BS and λ is the signal wavelength. The channel vector from UT k to the BS is defined as

$$\mathbf{h}_k = \frac{1}{\sqrt{P}}\sum_{p=1}^{P} g_{kp}\mathbf{a}(\theta_p),$$

where g_{kp} is the random propagation coefficient from user k to the BS. Hence, the channel matrix $\mathbf{H}$ can be written as

$$\mathbf{H} = \frac{1}{\sqrt{P}}\mathbf{A}\mathbf{G},$$

where $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_P)] \in \mathbb{C}^{M \times P}$ contains P steering vectors and the path gain matrix $\mathbf{G} \in \mathbb{C}^{P \times K}$ is assumed to have independent entries with $\mathbf{G}_{(p,k)} \sim \mathcal{CN}(0, \delta_{pk}^2)$. The variance δ_{pk}^2 is the channel's average attenuation including the path loss and shadowing effects [5], [42]. In the rest of this paper, we assume that $\delta_{pk}^2 = 1$ just for the simplicity. We note that the analysis can be easily extended to different values of δ_{pk} with more general distance-based power-law decaying path-loss distributions.

C. LS and MMSE channel estimation

The conventional pilot-aided channel estimation is accomplished by transmitting a training sequence with length $L \geq K$ from each user to the BS. Based on the received signal matrix $\mathbf{Y}$ and the training matrix $\mathbf{X}$ as in (3), the least squares (LS), LS scaled least squares (LS-SLS) channel estimates can be written as [9]

$$\widehat{\mathbf{H}}_{LS} = \mathbf{Y}\mathbf{X}^{\dagger} = \mathbf{Y}\mathbf{X}^*(\mathbf{X}\mathbf{X}^*)^{-1} \quad (5)$$

$$\widehat{\mathbf{H}}_{LS-SLS} = \frac{K\text{tr}(\mathbf{Y}\mathbf{Y}^H)}{\mathcal{P}(\sigma^2 KM + \text{tr}(\mathbf{Y}\mathbf{Y}^H))}\mathbf{Y}\mathbf{X}^*, \quad (6)$$

and the minimum mean-square-error (MMSE) channel estimate is given by [33]

$$\widehat{\mathbf{H}}_{MMSE} = \frac{K}{\mathcal{P}}\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{Y}\mathbf{X}^*, \quad (7)$$

where the total power is constrained by $\mathcal{P} = \|\mathbf{X}\|_F^2$. We note that the MMSE method assumes that the matrix $\mathbf{A}$ is available, whereas our proposed channel estimation and faulty antenna detection schemes do not require such assumption.

D. Atomic norm

Suppose a high-dimensional signal $\mathbf{x}$ can be formed as [43]

$$\mathbf{x} = \sum_{i=1}^{r} c_i\mathbf{a}_i, \quad \mathbf{a}_i \in \mathcal{A}, c_r \geq 0,$$

where $\mathcal{A}$ is a set of atoms that consists of simple building blocks of general signals (a.k.a atomic set). $\mathbf{x}$ is said to be simple when it can be expressed as sparse linear combination of atoms from some atomic set (i.e., r is relatively small). The atomic set $\mathcal{A}$ can be very general: for example, if $\mathbf{x}$ is a sparse vector, $\mathcal{A}$ could be the finite set of unit-norm one-sparse vectors; if $\mathbf{x}$ is a low-rank matrix, $\mathcal{A}$ could be the set of all unit-norm rank-1 matrices. The atomic norm is defined as below.

Definition II.1 (Atomic Norm). *[43] The atomic norm $\| \cdot \|_{\mathcal{A}}$ of $\mathcal{A}$ is the Minkowski functional (or the gauge function) associated with the convex hull of $\mathcal{A}$ (conv($\mathcal{A}$)):*

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf\{t > 0 | \mathbf{x} \in t\text{conv}(\mathcal{A})\}. \quad (8)$$

For example, when $\mathcal{A}$ is the set of unit-norm one-sparse vectors in $\mathbb{C}^n$, the atomic norm $\| \cdot \|_{\mathcal{A}}$ is the l_1 norm. When $\mathcal{A}$ is the set of unit-norm rank-1 matrices, the atomic norm is the nuclear norm (sum of singular values).

The atomic norm denoising problem is to use atomic norm regularization to denoise a signal known to be a sparse nonnegative combination of atoms from a set $\mathcal{A}$. Suppose we have an observation $\mathbf{y} = \mathbf{x} + \mathbf{n}$ and that we know in a priori that $\mathbf{x}$ can be expressed as a linear combination of a small number of atoms from $\mathcal{A}$. Intuitively, we could search over all the possible short linear combinations from $\mathcal{A}$ to select the one which minimizes the error $\|\mathbf{y} - \mathbf{x}\|_2^2$. However, this problem is NP-hard. We can estimate the signal through the following convex optimization [44]

$$\underset{\mathbf{x}}{\text{argmin}}\frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \tau\|\mathbf{x}\|_{\mathcal{A}}, \quad (9)$$

where τ is a regularization parameter.

E. Atomic norm of the channel matrix

In this part, we describe the atomic norm of the channel matrix and review our atomic norm denoising-based channel estimation scheme proposed in [8]. Suppose $\{\mathbf{u}_1\mathbf{u}_2\ldots\mathbf{u}_P\}$ is a set of K-dimensional vectors with unit-norm, i.e. $\|\mathbf{u}_p\|_2 = 1$. We define an atom as

$$A(\theta, \mathbf{u}) = \frac{1}{\sqrt{M}}\mathbf{a}(\theta)\mathbf{u}^*,$$

where $\mathbf{a}(\theta)$ follows (4) and the associated atomic set is given by

$$\mathcal{A} = \{A(\theta, \mathbf{u}) | \theta \in [-\pi/2, \pi/2], \|\mathbf{u}\|_2 = 1\}. \quad (10)$$

Then, the channel matrix $\mathbf{H}$ can be represented as a non-negative combination of P elements from the atomic set

$$\mathbf{H} = \frac{1}{\sqrt{P}}\mathbf{A}\mathbf{G} = \sum_{p=1}^{P} c_p A(\theta_p, \mathbf{u}_p),$$

where $\mathbf{A}$ contains P steering vectors and $c_p = \sqrt{\frac{M}{P}}\|\mathbf{G}_{(p,:)}\|_2$. The atomic norm of an arbitrary matrix $\mathbf{V}$ with respect to $\mathcal{A}$ can be expressed as

$$\|\mathbf{V}\|_{\mathcal{A}} = \inf\{t > 0 : \mathbf{V} \in t\text{conv}(\mathcal{A})\}$$
$$= \inf\{\sum_k c_k | \mathbf{V} = \sum_k c_k A(\theta_k, \mathbf{u}_k), c_k \geq 0\},$$

where $\text{conv}(\mathcal{A})$ is the convex hull of $\mathcal{A}$. Based on this definition, the atomic norm of the channel matrix can be written as

$$\|\mathbf{H}\|_{\mathcal{A}} \leq \sum_{p=1}^{P} c_p = \sum_{p=1}^{P} \sqrt{\frac{M}{P}}\|\mathbf{G}_{(p,:)}\|_2.$$

For the rest of this paper, the atomic norm $\|\cdot\|_{\mathcal{A}}$ refers to the one associated with the atomic set (10).

Next, we review the atomic norm denoising-based channel estimation scheme. For an arbitrary row-wise orthonormal matrix $\boldsymbol{\Phi} \in \mathbb{C}^{K \times L}$ ($L \geq K$), i.e. $\boldsymbol{\Phi}\boldsymbol{\Phi}^* = \mathbf{I}$, our training matrix $\mathbf{X} \in \mathbb{C}^{K \times L}$ is given by

$$\mathbf{X} = \sqrt{\frac{\mathcal{P}}{K}}\boldsymbol{\Phi}. \tag{11}$$

Multiplying (2) by $\boldsymbol{\Phi}^*$, we obtain

$$\mathbf{Z} \triangleq \frac{\mathbf{Y}\boldsymbol{\Phi}^*}{\sqrt{\mathcal{P}/K}} = \mathbf{H} + \mathbf{N}, \tag{12}$$

where $\mathbf{N} \triangleq \frac{1}{\sqrt{\mathcal{P}/K}}\mathbf{N}_0\boldsymbol{\Phi}^*$ has i.i.d. $\mathcal{CN}(0, \frac{K\sigma^2}{\mathcal{P}})$ entries. Here, the channel estimation is reduced to the estimation of a simple model $\mathbf{H}$ from its noisy observations $\mathbf{Z}$, which turns out to be an atomic norm denoising problem (9). The estimated channel matrix $\widehat{\mathbf{H}}_{AD}$ (AD is short for atomic norm denoising) can be obtained by the following atomic norm regularized algorithm:

$$\widehat{\mathbf{H}}_{AD} = \underset{\mathbf{H}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{H} - \mathbf{Z}\|_F^2 + \tau\|\mathbf{H}\|_{\mathcal{A}}, \tag{13}$$

where τ denotes a regularization parameter, the Frobenius norm $\|\mathbf{H} - \mathbf{Z}\|_F^2$ is employed to promote the structure of the additive noise and the penalization on the atomic norm is due to the fact that the channel can be expressed as a non-negative combination of a small number of elements from the atomic set (i.e., P is small.). In [45], it is shown that (13) can be efficiently implemented via the alternating direction method of multipliers (ADMM). The channel estimation error has been proved in [8, Proposition 1].

III. CHANNEL ESTIMATION WITH FAULTY ANTENNAS

In this section, we consider massive MIMO systems with failure of arbitrary antennas at the BS. The uncertainty on the indices of faulty antennas and the associated additive corruption noise can affect the channel estimation accuracy, and thus degrade the data transmission performance. We first analyze the signal detection MSE with a linear receiver (MRC, ZF) on the uplink in the large system limit, and show that the effect of faulty antennas does not vanish even with an infinite number of BS antennas. We then prove that a modified receiver can always achieve better signal detection performance

provided that the indices of faulty antennas are available. This motivates the study of identifying faulty antennas before the data transmission.

In the second part, we propose an atomic norm-based method that can estimate the channel and detect faulty antennas simultaneously in the training phase. By treating the channel matrix as a low rank matrix, a stable principle component pursuit (PCP)-based approach can also be applied. We provide a comparison between the proposed method and the stable PCP-based one.

Lastly, we present a fast algorithm and prove that it is a good approximation of the atomic norm-based method.

A. Asymptotic Analysis

In the UL data transmission, the received signal is

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w} + \mathbf{n}, \tag{14}$$

where ρ is the signal transmit power, the transmitted signal from K users $\mathbf{x} \in \mathbb{C}^K$ is a random vector with zero mean and covariance $\mathbf{I}_K$, the sparse vector $\mathbf{w} \in \mathbb{C}^M$ denotes the corruption error due to faulty antennas at the BS, and $\mathbf{n} \in \mathbb{C}^M$ is the independent additive noise with $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}_M)$.

We employ MRC (or ZF) receiver for the signal detection at the BS, and assume perfect CSI $\mathbf{H}$ for the asymptotic analysis. The mean squared error of signal detection $\mathbb{E}\{\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2\}$ is used as the performance measure, where the expectation is with respect to $\mathbf{x}$, $\mathbf{n}$, and the channel matrix $\mathbf{H}$. We assume that $\mathbf{x}$, $\mathbf{n}$, and the channel matrix $\mathbf{H}$ are independent from each other. We consider the large system limit, where S and M grow infinitely large while keeping the ratio $\gamma = S/M$ fixed. In other words, the number of faulty antennas increases proportionally with the number of total antennas. Here, $M \to \infty$ means $M \to \infty$, $S \to \infty$, while γ, K and P are fixed.

We first present the asymptotic analysis in detail on MRC receiver. Then, the same technique is employed to analyze the signal detection performance of ZF receiver.

1) Asymptotic analysis on (modified) MRC receiver: Two important preliminaries:

Lemma III.1. *[46, Lemma 3.2] Let* $\mathbf{s} = [s_1 \ldots s_N]^T$ *where* s_i's *are i.i.d. zero mean, unit variance random variables with finite fourth moment. Let* $\mathbf{A}$ *be a deterministic symmetric positive definite matrix. Then*

$$\mathbb{E}\{\mathbf{s}^*\mathbf{A}\mathbf{s}\} = \text{tr}\mathbf{A} \tag{15}$$

Based on the finite scattering channel model with $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_P)] \in \mathbb{C}^{M \times P}$ containing P steering vectors, we have [33, Equation (27-28)], for $p \neq q$,

$$\frac{1}{M}\mathbf{a}(\theta_p)^*\mathbf{a}(\theta_q) = \frac{1}{M}\sum_{m=0}^{M-1} e^{j2\pi\frac{D}{\lambda}(\sin\theta_p - \sin\theta_q)m}$$
$$= \frac{1}{M}\frac{1 - e^{j2\pi\frac{D}{\lambda}(\sin\theta_p - \sin\theta_q)M}}{1 - e^{j2\pi\frac{D}{\lambda}(\sin\theta_p - \sin\theta_q)}}$$
$$\overset{M\to\infty}{\to} 0,$$

and

$$\frac{1}{M}\mathbf{a}(\theta_p)^*\mathbf{a}(\theta_p) = 1.$$

In other words,

$$\frac{1}{M}\mathbf{A}^*\mathbf{A} \overset{M\to\infty}{\Rightarrow} \mathbf{I}_P \qquad (16)$$

By MRC, the estimation on the transmitted signal $\mathbf{x}$ is

$$\widehat{\mathbf{x}} = \frac{\mathbf{H}^*\mathbf{y}}{\sqrt{\rho}M} = \frac{\mathbf{H}^*\mathbf{H}}{M}\mathbf{x} + \frac{\mathbf{H}^*\mathbf{w}}{\sqrt{\rho}M} + \frac{\mathbf{H}^*\mathbf{n}}{\sqrt{\rho}M} \qquad (17)$$

The MSE can be decomposed as (18) at the top of the next page.

Due to independence and (16)

$$\mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{w}}{M}\right\} = 0, \quad \mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{n}}{M}\right\} = 0,$$

$$\mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{w}}{M^2}\right\} = 0, \quad \mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{n}}{M^2}\right\} = 0,$$

$$\mathbb{E}\left\{\frac{\mathbf{w}^*\mathbf{H}\mathbf{H}^*\mathbf{n}}{M^2}\right\} = 0.$$

We have

$$\mathbb{E}\left\{\frac{\mathbf{n}^*\mathbf{H}\mathbf{H}^*\mathbf{n}}{M^2}\right\} = \frac{\mathbb{E}\{\mathrm{tr}(\mathbf{H}\mathbf{H}^*)\}}{M^2}$$
$$= \frac{\mathbb{E}\{\mathrm{tr}(\mathbf{A}\mathbf{G}\mathbf{G}^*\mathbf{A}^*)\}}{PM^2}$$
$$= \frac{\mathrm{tr}(\mathbf{A}\mathbb{E}\{\mathbf{G}\mathbf{G}^*\}\mathbf{A}^*)}{PM^2}$$
$$= \frac{\mathrm{tr}(\mathbf{A}\mathbf{A}^*)K}{PM^2}$$
$$= \frac{MPK}{PM^2} \overset{M\to\infty}{\Rightarrow} 0, \qquad (19)$$

where the first equality is due to Lemma III.1.

$$\mathbb{E}\{\|\mathbf{x}\|_2^2\} = K \qquad (20)$$

$$\mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{x}}{M}\right\} = \frac{1}{P}\mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{G}^*\mathbf{A}^*\mathbf{A}\mathbf{G}\mathbf{x}}{M}\right\}$$
$$= \frac{1}{P}\mathrm{tr}\left(\frac{\mathbb{E}\{\mathbf{G}^*\mathbf{A}^*\mathbf{A}\mathbf{G}\}}{M}\right)$$
$$\overset{M\to\infty}{\Rightarrow} \frac{1}{P}\mathrm{tr}(\mathbb{E}\{\mathbf{G}^*\mathbf{G}\}) = K, \qquad (21)$$

where the last two steps are due to Lemma III.1 and (16) respectively.

Similarly,

$$\mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{H}\mathbf{x}}{M^2}\right\} = \frac{\mathbb{E}\{\mathrm{tr}(\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{H})\}}{M^2}$$
$$= \frac{\mathbb{E}\{\mathrm{tr}(\mathbf{G}^*\mathbf{A}^*\mathbf{A}\mathbf{G}\mathbf{G}^*\mathbf{A}^*\mathbf{A}\mathbf{G})\}}{P^2M^2}$$
$$\overset{M\to\infty}{\Rightarrow} \frac{\mathbb{E}\{\mathrm{tr}(\mathbf{G}^*\mathbf{G}\mathbf{G}^*\mathbf{G})\}}{P^2}$$
$$= \frac{K^2}{P} + \frac{K}{P} + K, \qquad (22)$$

$$\mathbb{E}\left\{\frac{\mathbf{w}^*\mathbf{H}\mathbf{H}^*\mathbf{w}}{M^2}\right\} = \frac{\mathbb{E}\{\mathbf{w}^*\mathbf{A}\mathbf{G}\mathbf{G}^*\mathbf{A}^*\mathbf{w}\}}{PM^2}$$
$$= \frac{\mathbf{w}^*\mathbf{A}\mathbb{E}\{\mathbf{G}\mathbf{G}^*\}\mathbf{A}^*\mathbf{w}}{PM^2}$$
$$= \frac{K}{PM^2}\|\mathbf{A}^*\mathbf{w}\|_2^2. \qquad (23)$$

Combine the above equations, we have

$$\mathbb{E}\{\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2\} \overset{M\to\infty}{\Rightarrow} \frac{K^2}{P} + \frac{K}{P} + \frac{K}{\rho PM^2}\|\mathbf{A}^*\mathbf{w}\|_2^2. \qquad (24)$$

For any $\mathbf{a}(\theta_i)$, $i \in [P]$, we have $\mathbf{a}(\theta_i)^*\mathbf{w} \leq \gamma M\|\mathbf{w}\|_\infty$. Hence,

$$0 \leq \frac{K}{\rho PM^2}\|\mathbf{A}^*\mathbf{w}\|_2^2 \leq \frac{K\gamma^2}{\rho}\|\mathbf{w}\|_\infty^2. \qquad (25)$$

Next, we analyze the asymptotic MSE when the indices of faulty antennas are known. We denote the support of $\mathbf{w}$ by $\Omega \in \{1, \ldots, M\}$, i.e., $|\Omega| = S = \gamma M$. Suppose Ω is known in a priori at the BS, we shall prove that this knowledge can be utilized to obtained better signal detection performance in terms of asymptotic MSE. Let $\overline{\mathbf{A}}$ denote the matrix obtained by removing the rows of $\mathbf{A}$ indexed by Ω, $\overline{\mathbf{A}} \in \mathbb{C}^{(1-\gamma)M \times P}$, and $\overline{\mathbf{n}} \in \mathbb{C}^{(1-\gamma)M}$ represent the i.i.d. additive noise vector with zero mean and unit variance.

Since the indices of faulty antennas Ω are known, we can omit the readings received from those antennas at the BS. In this way, the impact of $\mathbf{w}$ is removed and the received signal is

$$\overline{\mathbf{y}} = \sqrt{\rho}\overline{\mathbf{H}}\mathbf{x} + \overline{\mathbf{n}}, \qquad (26)$$

where $\overline{\mathbf{H}} := \frac{1}{\sqrt{P}}\overline{\mathbf{A}}\mathbf{G}$.

By MRC, the transmitted signal can be recovered by

$$\overline{\mathbf{x}} = \frac{\overline{\mathbf{H}}^*\overline{\mathbf{y}}}{(1-\gamma)\sqrt{\rho}M} = \frac{\overline{\mathbf{H}}^*\overline{\mathbf{H}}\mathbf{x}}{(1-\gamma)M} + \frac{\overline{\mathbf{H}}^*\overline{\mathbf{n}}}{(1-\gamma)M\sqrt{\rho}}$$

For $\overline{\mathbf{A}} = [\overline{\mathbf{a}}(\theta_1), \overline{\mathbf{a}}(\theta_2), \cdots, \overline{\mathbf{a}}(\theta_P)]$ with $\overline{\mathbf{a}}(\theta_i)$'s being the vectors obtained by removing the Ω-th entries of $\mathbf{a}(\theta_i)$, we have, for $p \neq q$,

$$\frac{1}{(1-\gamma)M}\overline{\mathbf{a}}(\theta_p)^*\overline{\mathbf{a}}(\theta_q) = \frac{1}{(1-\gamma)M}\sum_{m\in[M]\backslash\Omega} e^{j2\pi\frac{D}{\lambda}(\sin\theta_p - \sin\theta_q)m}$$
$$\overset{M\to\infty}{\Rightarrow} 0,$$

and

$$\frac{1}{(1-\gamma)M}\overline{\mathbf{a}}(\theta_p)^*\overline{\mathbf{a}}(\theta_p) = 1.$$

Hence,

$$\frac{1}{(1-\gamma)M}\overline{\mathbf{A}}^*\overline{\mathbf{A}} \overset{M\to\infty}{\Rightarrow} \mathbf{I}_P \qquad (27)$$

Similarly, the asymptotic MSE can be obtained by applying Lemma III.1 and (27)

$$\mathbb{E}\{\|\mathbf{x} - \overline{\mathbf{x}}\|_2^2\}$$
$$= \mathbb{E}\{(\mathbf{x} - \overline{\mathbf{x}})^*(\mathbf{x} - \overline{\mathbf{x}})\}$$
$$= \mathbb{E}\left\{\|\mathbf{x}\|_2^2 - 2\frac{\mathbf{x}^*\overline{\mathbf{H}}^*\overline{\mathbf{H}}\mathbf{x}}{(1-\gamma)M} - 2\frac{\mathbf{x}^*\overline{\mathbf{H}}^*\overline{\mathbf{n}}}{(1-\gamma)M\sqrt{\rho}} + \frac{\mathbf{x}^*\overline{\mathbf{H}}^*\overline{\mathbf{H}}\overline{\mathbf{H}}^*\overline{\mathbf{H}}\mathbf{x}}{(1-\gamma)^2M^2}\right.$$
$$\left. + 2\frac{\mathbf{x}^*\overline{\mathbf{H}}^*\overline{\mathbf{H}}\overline{\mathbf{H}}^*\overline{\mathbf{n}}}{(1-\gamma)^2M^2\sqrt{\rho}} + \frac{\overline{\mathbf{n}}^*\overline{\mathbf{H}}\overline{\mathbf{H}}^*\overline{\mathbf{n}}}{(1-\gamma)^2M^2\sqrt{\rho}}\right\}$$
$$\overset{M\to\infty}{\Rightarrow} \frac{K^2}{P} + \frac{K}{P} \qquad (28)$$

Hence, in the large system limit, the MSE obtained by appealing to the knowledge on the indices of faulty antennas is

$$\mathbb{E}\{\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2\} = \mathbb{E}\{(\mathbf{x} - \widehat{\mathbf{x}})^*(\mathbf{x} - \widehat{\mathbf{x}})\}$$

$$= \mathbb{E}\left\{(\mathbf{x} - \frac{\mathbf{H}^*\mathbf{H}}{M}\mathbf{x} - \frac{\mathbf{H}^*\mathbf{w}}{\sqrt{\rho}M} - \frac{\mathbf{H}^*\mathbf{n}}{\sqrt{\rho}M})^*(\mathbf{x} - \frac{\mathbf{H}^*\mathbf{H}}{M}\mathbf{x} - \frac{\mathbf{H}^*\mathbf{w}}{\sqrt{\rho}M} - \frac{\mathbf{H}^*\mathbf{n}}{\sqrt{\rho}M})\right\}$$

$$= \mathbb{E}\left\{\|\mathbf{x}\|_2^2 - 2\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{x}}{M} - 2\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{w}}{\sqrt{\rho}M} - 2\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{n}}{\sqrt{\rho}M} + \frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{H}\mathbf{x}}{M^2} + 2\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{w}}{\sqrt{\rho}M^2}\right.$$

$$\left. + 2\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{n}}{\sqrt{\rho}M^2} + \frac{\mathbf{w}^*\mathbf{H}\mathbf{H}^*\mathbf{w}}{\rho M^2} + 2\frac{\mathbf{w}^*\mathbf{H}\mathbf{H}^*\mathbf{n}}{\rho M^2} + \frac{\mathbf{n}^*\mathbf{H}\mathbf{H}^*\mathbf{n}}{\rho M^2}\right\} \tag{18}$$

always smaller than the one given by the conventional method (compare (24) and (28)). Some remarks are in order.

1) The difference of the two MSEs ranges from 0 to $\frac{K\gamma^2}{\rho}\|\mathbf{w}\|_\infty^2$ (25). The result is consistent with the intuitive understanding: when $\gamma = 0$, i.e. all antennas are fault-free, the two MSEs are identical; in the worst case, the difference $\frac{K\gamma^2}{\rho}\|\mathbf{w}\|_\infty^2$ is proportional to the number of users K and inversely proportional to the signal transmit power ρ.

2) In the large system limit with fault-free antennas, MRC has been shown to mitigate the effect of uncorrelated noise and small-scale fading [1]. When considering faulty-antennas, our analysis indicates that the impact of faulty-antennas does not vanish with unlimited number of antennas. However, with the knowledge on the indices of faulty antennas, this effect can be eliminated by a simple modified MRC receiver.

3) In this paper, our analysis does not rely on any assumption on the statistical model of the distortion noise caused by faulty antennas. If the statistical information on either the positions of faulty antennas or the magnitude of the distortion caused by faulty antennas is available/etimatable, other receivers (e.g. MMSE) could be investigated.

4) When the channel matrix $\mathbf{H}$ is i.i.d. Gaussian, similar result can be obtained by following the derivations above.

2) Asymptotic analysis on (modified) ZF receiver: With a similar technique, we shall show that a modified ZF receiver with the knowledge on the indices of faulty antennas can provide better signal detection performance in the large system limit. The ZF receiver can suppress interuser interference and perform well at high SNR. For the ZF receiver, we assume that $P \geq K$ [1]. Here, we use $\mathbb{E}\{\|\mathbf{x} - \widehat{\mathbf{x}}\|_2\}$ as the performance metric for convenience. Without a priori knowledge on the indices of faulty antennas, the signal recovered by ZF receiver can be written as

$$\widehat{\mathbf{x}} = \frac{\mathbf{H}^\dagger\mathbf{y}}{\sqrt{\rho}} = \mathbf{x} + \frac{\mathbf{H}^\dagger\mathbf{w}}{\sqrt{\rho}} + \frac{\mathbf{H}^\dagger\mathbf{n}}{\sqrt{\rho}},$$

---

[1]This is to ensure that $\mathbf{G}^*\mathbf{G}$ is invertible. The same assumption has been used in the literature, e.g. [33].

where $\mathbf{H}^\dagger \triangleq (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*$. Denoting $\mathbf{G}^\dagger \triangleq (\mathbf{G}^*\mathbf{G})^{-1}\mathbf{G}^*$, the asymptotic error is given by

$$\mathbb{E}\{\|\mathbf{x} - \widehat{\mathbf{x}}\|_2\} \leq \frac{1}{\sqrt{\rho}}\mathbb{E}\{\|\mathbf{H}^\dagger\mathbf{w}\|_2\} + \mathbb{E}\{\|\mathbf{H}^\dagger\mathbf{n}\|_2\}$$

$$= \frac{1}{\sqrt{\rho}}\mathbb{E}\left\{\left\|\left(\frac{\mathbf{G}^*\mathbf{A}^*\mathbf{A}\mathbf{G}}{P}\right)^{-1}\frac{\mathbf{G}^*\mathbf{A}^*}{\sqrt{P}}\mathbf{w}\right\|_2\right\}$$

$$= \frac{\sqrt{P}}{\sqrt{\rho}M}\mathbb{E}\left\{\left\|\left(\frac{\mathbf{G}^*\mathbf{A}^*\mathbf{A}\mathbf{G}}{M}\right)^{-1}\mathbf{G}^*\mathbf{A}^*\mathbf{w}\right\|_2\right\}$$

$$\overset{M\rightarrow\infty}{\rightarrow} \frac{\sqrt{P}}{\sqrt{\rho}M}\mathbb{E}\left\{\left\|(\mathbf{G}^*\mathbf{G})^{-1}\mathbf{G}^*\mathbf{A}^*\mathbf{w}\right\|_2\right\}$$

$$\leq \frac{\sqrt{P}}{\sqrt{\rho}M}\mathbb{E}\{\|\mathbf{G}^\dagger\|_2\}\|\mathbf{A}^*\mathbf{w}\|_2$$

$$\overset{(a)}{\leq} \frac{\sqrt{P}}{\sqrt{\rho}M(\sqrt{P} - \sqrt{K})}\|\mathbf{A}^*\mathbf{w}\|_2$$

$$\overset{(b)}{\leq} \frac{\gamma P\|\mathbf{w}\|_\infty}{\sqrt{\rho}(\sqrt{P} - \sqrt{K})},$$

where (a) is based on Gordon's theorem for Gaussian matrices ([47, Theorem 5.32]) and the fact that $s_{min}(\mathbf{G}) = 1/\|\mathbf{G}^\dagger\|$ with $s_{min}(\mathbf{G})$ denoting the smallest singular value of $\mathbf{G}$; (b) is due to a similar technique to obtain (25). Like the modified MRC receiver, we can omit the readings from the faulty antennas if their locations are available, and apply ZF receiver to the signal (26). We have

$$\overline{\mathbf{x}} = \frac{\overline{\mathbf{H}}^\dagger\overline{\mathbf{y}}}{\sqrt{\rho}} = \mathbf{x} + \frac{\overline{\mathbf{H}}^\dagger\overline{\mathbf{n}}}{\sqrt{\rho}}.$$

The asymptotic error is $\mathbb{E}\{\|\mathbf{x} - \overline{\mathbf{x}}\|_2\} \overset{M\rightarrow\infty}{\rightarrow} 0$, which demonstrates the improved signal detection performance of the modified ZF receiver.

Remark III.2. Consider the case when the number of dominant AoAs P also grows infinitely large and assume that M grows at a greater rate than P (as in [33, Remark 1]). Here, $M, P \rightarrow \infty$ means $M \rightarrow \infty$, $S \rightarrow \infty$, $P \rightarrow \infty$, while γ and K are fixed. We have

$$\frac{1}{M}\mathbf{A}^*\mathbf{A} \overset{M,P\rightarrow\infty}{\rightarrow} \mathbf{I}_P,$$

$$\frac{1}{P}\mathbf{G}^*\mathbf{G} \overset{M,P\rightarrow\infty}{\rightarrow} \mathbf{I}_K.$$

For the MRC receiver without the knowledge on the indices of the faulty antennas, the affected term is (22), which is now given by

$$\mathbb{E}\left\{\frac{\mathbf{x}^*\mathbf{H}^*\mathbf{H}\mathbf{H}^*\mathbf{H}\mathbf{x}}{M^2}\right\} \overset{M,P\rightarrow\infty}{\rightarrow} K. \tag{29}$$

Therefore, we have $\mathbb{E}\{\|\mathbf{x}-\widehat{\mathbf{x}}\|_2^2\} \overset{M,P\rightarrow\infty}{\rightarrow} \frac{K}{\rho P M^2}\|\mathbf{A}^*\mathbf{w}\|_2^2$. It is noted that this term is bounded by (25). If the indices of the faulty antennas are available, it can be proven that $\mathbb{E}\{\|\mathbf{x}-\overline{\mathbf{x}}\|_2^2\} \overset{M,P\rightarrow\infty}{\rightarrow} 0$ by following the same steps.

For the ZF receiver, it can be shown that the ZF receiver is equivalent to the MRC receiver when $M, P \rightarrow \infty$ and M growing at a greater rate than P. We have

$$\begin{aligned}
\widehat{\mathbf{x}} &= \frac{\mathbf{H}^\dagger\mathbf{y}}{\sqrt{\rho}} \\
&= \frac{(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{y}}{\sqrt{\rho}} \\
&= \frac{1}{M}\left(\frac{\mathbf{G}^*\mathbf{A}^*\mathbf{A}\mathbf{G}}{MP}\right)^{-1}\frac{\mathbf{H}^*\mathbf{y}}{\sqrt{\rho}} \\
&\overset{M,P\rightarrow\infty}{\rightarrow} \frac{\mathbf{H}^*\mathbf{y}}{\sqrt{\rho}M},
\end{aligned}$$

which is the same as the estimation given by the MRC receiver (17). Hence, the asymptotic signal detection MSE follows the results for the MRC receiver.

B. The proposed method

Since the magnitude of the sparse distortion $\mathbf{w}$ may change arbitrarily for different symbol intervals, estimation of the exact vector $\mathbf{w}$ that affects data transmission is impossible. From the asymptotic analysis in the previous subsection, we may estimate the channel matrix and detect the indices of faulty antennas simultaneously in the training phase and apply the modified MRC/ZF receiver to guarantee signal detection performance. We therefore propose the following method.

We assume that the training matrix $\mathbf{X}$ is constructed as (11). Multiplying (3) by $\mathbf{\Phi}^*$, we obtain

$$\mathbf{Z} \triangleq \frac{\mathbf{Y}\mathbf{\Phi}^*}{\sqrt{\mathcal{P}/K}} = \mathbf{H} + \mathbf{W} + \mathbf{N}, \tag{30}$$

where $\mathbf{N} \triangleq \frac{1}{\sqrt{\mathcal{P}/K}}\mathbf{N}_0\mathbf{\Phi}^*$ has i.i.d. $\mathcal{CN}(0, \frac{K\sigma^2}{\mathcal{P}})$ entries and $\mathbf{W} \triangleq \frac{1}{\sqrt{\mathcal{P}/K}}\mathbf{W}_0\mathbf{\Phi}^* \in \mathbb{C}^{M\times K}$ is still a matrix with S nonzero rows. Also, the indices of nonzero rows of $\mathbf{W}$ are the same as those of $\mathbf{W}_0$. Now, our problem is to decompose the given matrix $\mathbf{Z}$ into a sum of three component matrices, each of which is 'simple' in a sense described by the corresponding norm. To recover $\mathbf{H}$ and $\mathbf{W}$, we propose to solve the following *extended* atomic norm minimization algorithm:

$$\underset{\mathbf{H},\mathbf{W}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{H}+\mathbf{W}-\mathbf{Z}\|_F^2 + \tau_1\|\mathbf{H}\|_\mathcal{A} + \tau_2\|\mathbf{W}\|_{2,1}. \tag{31}$$

Here, the function $\|\cdot\|_{2,1}$ representing the sum of the Euclidean norms of the rows of the matrix is applied since $\mathbf{W}$ is a matrix with a small number of nonzero rows. Although the dimensions of $\mathbf{W}_0$ and $\mathbf{W}$ are different, recovering $\mathbf{W}$ is good

enough since its indices of nonzero rows are the same as those of $\mathbf{W}_0$. To implement the proposed algorithm, we reformulate it into a matrix decomposition problem

$$\begin{aligned}
&\underset{\mathbf{H},\mathbf{W},\mathbf{N}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{N}\|_F^2 + \tau_1\|\mathbf{H}\|_\mathcal{A} + \tau_2\|\mathbf{W}\|_{2,1} \tag{32} \\
&\text{s.t.} \quad \mathbf{H} + \mathbf{W} + \mathbf{N} = \mathbf{Z},
\end{aligned}$$

which can be solved based on exchange and ADMM [48]. Detail steps are included in Algorithm 1. In every iteration, each of the matrices $\mathbf{H}, \mathbf{W}, \mathbf{N}$ is updated by a convex optimization algorithm. As shown in Algorithm 1, the updates on $\mathbf{W}$ and $\mathbf{N}$ have simple closed form solutions. While for the update on $\mathbf{H}$,

$$\mathbf{H} = \underset{\mathbf{H}}{\operatorname{argmin}} \tau_1\|\mathbf{H}\|_\mathcal{A} + \frac{1}{2}\|\mathbf{H}-\mathbf{V}_1^k\|_F^2 \tag{33}$$

is exactly the atomic norm regularization algorithm in (13), which can be efficiently implemented by the ADMM as described in [45].

Remark III.3. In the case of $K = 1$, the problem (30) is reduced to the line spectrum estimation of a vector signal that is corrupted by sparse noise and dense noise. Currently, the theoretical guarantee has been proved only when the dense noise is absent [49], [50]. When both the sparse noise and the dense noise exist, a vector version of the proposed extended atomic norm minimization algorithm is adapted to recover the signal [50]. In addition, the vector version has been applied in the delay-doppler estimation for OFDM passive radar [51].

C. The stable PCP-based approach

In the proposed method, we can simultaneously estimate the channel and detect faulty antennas by mainly exploring the knowledge that the intrinsic information of the channel matrix is small, that is the channel matrix is a linear combination of a small number of elements from the atomic set (10). We note that similar idea of exploring the intrinsic information for channel estimation has been considered based on compressed sensing. In [7], [10], [52], where no faulty antenna is considered, channel estimation is transformed into a low-rank matrix recovery problem by noticing the fact that the channel matrix is low rank for finite scattering channel models. Comparing with our proposed method, it naturally leads to another algorithm for channel estimation and faulty antenna detection. By relying on the knowledge that the channel matrix $\mathbf{H}$ is low rank, we can solve the matrix decomposition problem (30) by the following algorithm

$$\begin{aligned}
&\underset{\mathbf{H},\mathbf{W},\mathbf{N}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{N}\|_F^2 + \lambda_1\|\mathbf{H}\|_* + \lambda_2\|\mathbf{W}\|_{2,1} \tag{34} \\
&\text{s.t.} \quad \mathbf{H} + \mathbf{W} + \mathbf{N} = \mathbf{Z},
\end{aligned}$$

where $\|\cdot\|_*$ denotes the nuclear norm. In other words, the algorithm is to recover a low rank matrix which is corrupted by a row-sparse matrix and additive noise. We note that the above algorithm is quite similar to the stable principle component pursuit (PCP) algorithm [53]. The only difference is that in [53] $\mathbf{W}$ is a sparse matrix, whereas here it is a row-sparse

Algorithm 1 Implementation of the extended atomic norm minimization algorithm

---

Input: $\mathbf{Z}$, τ_1, τ_2
Initialize: $\mathbf{H}^{(0)}$, $\mathbf{W}^{(0)}$, $\mathbf{N}^{(0)}$, $\mathbf{U}^{(0)}$

   for $k = 1, 2, \ldots$ **do**

$$\overline{\mathbf{X}}^k = \frac{1}{3}(\mathbf{H}^k + \mathbf{W}^k + \mathbf{N}^k)$$
$$\mathbf{V}_1^k = \mathbf{H}^k - \overline{\mathbf{X}}^k + (1/3)\mathbf{Z} - \mathbf{U}^k$$
$$\mathbf{V}_2^k = \mathbf{W}^k - \overline{\mathbf{X}}^k + (1/3)\mathbf{Z} - \mathbf{U}^k$$
$$\mathbf{V}_3^k = \mathbf{N}^k - \overline{\mathbf{X}}^k + (1/3)\mathbf{Z} - \mathbf{U}^k$$
$$\mathbf{H}^{k+1} = \underset{\mathbf{H}}{\arg\min}\ \tau_1\|\mathbf{H}\|_{\mathcal{A}} + \frac{1}{2}\|\mathbf{H} - \mathbf{V}_1^k\|_F^2$$
$$\mathbf{W}^{k+1} = \underset{\mathbf{W}}{\arg\min}\ \tau_2\|\mathbf{W}\|_{2,1} + \frac{1}{2}\|\mathbf{W} - \mathbf{V}_2^k\|_F^2$$
$$\mathbf{N}^{k+1} = \underset{\mathbf{N}}{\arg\min}\ \frac{1}{2}\|\mathbf{N}\|_F^2 + \frac{1}{2}\|\mathbf{N} - \mathbf{V}_3^k\|_F^2$$
$$\overline{\mathbf{X}}^{k+1} = \frac{1}{3}(\mathbf{H}^{k+1} + \mathbf{W}^{k+1} + \mathbf{N}^{k+1})$$
$$\mathbf{U}^{k+1} = \mathbf{U}^k + \overline{\mathbf{X}}^{k+1} - (1/3)\mathbf{Z}$$

    Solve for $\mathbf{H}^{k+1}$ by another ADMM algorithm as in [45]
    Solve for $\mathbf{W}^{k+1}$ and $\mathbf{N}^{k+1}$

$$\mathbf{W}_{(m,:)}^{k+1} = (1 - \frac{\tau_2}{\|(\mathbf{V}_2^k)_{(m,:)}\|_2})_+ (\mathbf{V}_2^k)_{(m,:)}$$
$$\mathbf{N}^{k+1} = \frac{1}{2}\mathbf{V}_3^k$$

                               ▷ $m = 1, \ldots, M$
               ▷ $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ otherwise
   end for
   return $\mathbf{H}$, $\mathbf{W}$

---

matrix. This stable PCP-based algorithm (34) can be efficiently implemented via ADMM[2]. The update steps are in general the same as those of our proposed method (see Algorithm 1). The difference is on the update of $\mathbf{H}$, where the following algorithm is implemented,

$$\mathbf{H} = \underset{\mathbf{H}}{\arg\min}\ \lambda_1\|\mathbf{H}\|_* + \frac{1}{2}\|\mathbf{H} - \mathbf{V}_1^k\|_F^2. \tag{35}$$

D. A fast algorithm

Comparing with LS-based channel estimation methods, the proposed atomic norm-based method has higher computational complexity. The computational complexity for LS and LS-SLS methods are $\mathcal{O}(MLK + K^3)$ and $\mathcal{O}(M^2L + MLK)$ respectively. As shown in Algorithm 1, the main step of our proposed algorithm is to update the estimated channel matrix $\mathbf{H}$ (33), while the rest updates have closed form solutions. The dominate operation in each iteration for solving (33) is the eigenvalue decomposition of a $(M + K) \times (M + K)$ matrix, hence the complexity is $\mathcal{O}((M + K)^3)$. Similarly, for

---

[2]In the literature, the stable PCP can also be solved by standard SDP and fast proximal gradient algorithm. However, the state-of-art solver is by ADMM [53], [54].

the stable PCP-based approach (34), the main step is also to update the estimated channel matrix by solving (35). In this algorithm, the dominate operation in each iteration is the singular value decomposition (SVD) for a $M \times K$ matrix, which has computational complexity $\mathcal{O}(M^2K + K^3)$. We note that the computational complexity of the stable PCP-based approach is slightly better than our proposed method, however, its performance is degraded. In this part, we present a fast algorithm for the proposed atomic norm-based approach. We prove that the fast algorithm performs as well as the atomic norm-based approach, and has a much lower computational complexity than both the atomic norm and stable PCP-based approaches.

Recall that the steering vector associated with each angle-of-arrival (AOA) is given by

$$\mathbf{a}(\theta) \triangleq [1\quad e^{-j2\pi\frac{D}{\lambda}\sin(\theta)} \ldots \quad e^{-j2\pi\frac{(M-1)D}{\lambda}\sin(\theta)}]^T,$$

where $\theta \in [-\pi/2, \pi/2]$ and $\frac{D}{\lambda} < 1$. For each $\theta \in [-\pi/2, \pi/2]$, we can find a corresponding $f \in [0, 1)$ such that the steering vector can be written as

$$\mathbf{a}(f) \triangleq [1\quad e^{j2\pi f} \ldots \quad e^{j2\pi(M-1)f}]^T.$$

Accordingly, the atomic set in (10) can be expressed as

$$\mathcal{A} = \{A(f, \mathbf{u}) = \frac{1}{\sqrt{M}}\mathbf{a}(f)\mathbf{u}^* | f \in [0, 1), \|\mathbf{u}\|_2 = 1\}. \tag{36}$$

To present the fast algorithm, we divide the interval $[0, 1)$ into N uniform grids and define a new atomic set

$$\mathcal{A}_N = \{A_N(f, \mathbf{u}) | f \in \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}, \|\mathbf{u}\|_2 = 1\}, \tag{37}$$

where $A_N(f, \mathbf{u}) = \frac{1}{\sqrt{M}}\mathbf{a}(f)\mathbf{u}^*$. Based on the new atomic set, the channel matrix can be written as

$$\mathbf{H} = \frac{1}{\sqrt{P}}\overline{\mathbf{F}}\overline{\mathbf{G}} = \sum_{i=1}^{N} c_i A_N(f_i, \mathbf{u}_i), \tag{38}$$

where $\overline{\mathbf{F}} = [\frac{1}{\sqrt{M}}\mathbf{a}(0), \ldots, \frac{1}{\sqrt{M}}\mathbf{a}(\frac{N-1}{N})] \in \mathbb{C}^{M \times N}$ $(M < N)$ is the first M rows of a normalized discrete Fourier transform (DFT) matrix, $\overline{\mathbf{G}} \in \mathbb{C}^{N \times K}$ and $c_i = \sqrt{\frac{M}{P}}\|\overline{\mathbf{G}}_{(i,:)}\|_2$. If the channel matrix is related to a finite scattering channel model with P paths, the above expression indicates that $\overline{\mathbf{G}}$ is a matrix with P non-zero rows. The atomic norm of the channel matrix associated with the new atomic set is given by

$$\|\mathbf{H}\|_{\mathcal{A}_N} = \inf\{\sum_i c_i | \mathbf{H} = \sum_i c_i A_N(f_i, \mathbf{u}_i)\} \tag{39}$$

$$= \inf\{\sum_i \sqrt{\frac{M}{P}}\|\overline{\mathbf{G}}_{(i,:)}\|_2 | \mathbf{H} = \frac{1}{\sqrt{P}}\overline{\mathbf{F}}\overline{\mathbf{G}}\} \tag{40}$$

$$= \inf\{\sqrt{\frac{M}{P}}\|\overline{\mathbf{G}}\|_{2,1} | \mathbf{H} = \frac{1}{\sqrt{P}}\overline{\mathbf{F}}\overline{\mathbf{G}}\}. \tag{41}$$

Now, we can solve the matrix decomposition problem (30) by the following algorithm

$$\underset{\mathbf{H}, \mathbf{W}, \mathbf{N}}{\arg\min}\ \frac{1}{2}\|\mathbf{N}\|_F^2 + \nu_1\|\mathbf{H}\|_{\mathcal{A}_N} + \nu_2\|\mathbf{W}\|_{2,1} \tag{42}$$

$$\text{s.t.}\quad \mathbf{H} + \mathbf{W} + \mathbf{N} = \mathbf{Z}.$$

One important question is that when the frequencies $f \in [0, 1)$ of the channel are not on the uniform grids, the new atomic norm $\|\mathbf{H}\|_{\mathcal{A}_N}$ will deviate from the original one $\|\mathbf{H}\|_{\mathcal{A}}$ and lead to additional estimation error. Here, we address this question by proving that $\|\mathbf{H}\|_{\mathcal{A}_N}$ is a good approximation of $\|\mathbf{H}\|_{\mathcal{A}}$ by setting N a constant factor larger than MK. In particular, we have (See Appendix A for the proof)

$$(1 - \frac{2\pi MK}{N})^{\frac{1}{2}} \|\mathbf{H}\|_{\mathcal{A}_N} \leq \|\mathbf{H}\|_{\mathcal{A}} \leq \|\mathbf{H}\|_{\mathcal{A}_N}. \quad (43)$$

The above result indicates that $\|\mathbf{H}\|_{\mathcal{A}_N}$ is a good approximation of $\|\mathbf{H}\|_{\mathcal{A}}$ when $N = \mathcal{O}(MK)$ and the gap between them approaches zero as $N \to \infty$. We note that the term $(1 - \frac{2\pi MK}{N})^{\frac{1}{2}}$ is not tight, which means $N \geq 2\pi MK$ is not necessarily required in practice.

Next, we proceed to explain the computational efficiency of the new algorithm. Clearly, the algorithm (42) can still be implemented by exchange and ADMM, as the one for algorithm (31) in Algorithm 1. The key difference is to change the update step of $\mathbf{H}$ to

$$\mathbf{H} = \underset{\mathbf{H}}{\operatorname{argmin}} \ \nu_1 \|\mathbf{H}\|_{\mathcal{A}_N} + \frac{1}{2} \|\mathbf{H} - \mathbf{V}_1^k\|_F^2, \quad (44)$$

which is equivalent to

$$\overline{\mathbf{G}} = \underset{\overline{\mathbf{G}}}{\operatorname{argmin}} \ \nu_1 \|\overline{\mathbf{G}}\|_{2,1} + \frac{1}{2} \|\overline{\mathbf{F}}\overline{\mathbf{G}} - \mathbf{V}_1^k\|_F^2 \quad (45)$$

$$\mathbf{H} = \overline{\mathbf{F}}\overline{\mathbf{G}}, \quad (46)$$

where the first algorithm is a group lasso for the recovery of a matrix with small number of non-zero rows. It can be solved iteratively with each iteration dominated by the matrix multiplication by $\overline{\mathbf{F}}$. Hence, the computational complexity of the fast algorithm is $\mathcal{O}(MK \log(MK))$ by relying on the efficiency of the FFT.

IV. NUMERICAL RESULTS

In this section, we conduct numerical experiments on massive MIMO systems with faulty antennas to demonstrate the performance of our proposed extended atomic norm denoising-based method (exAD) (31) and the fast atomic norm denoising-based method (fsAD) (42). We also include the simulation results of the LS-based methods, the MMSE method and the stable PCP-based algorithm (stPCP) (34) for comparison.

We consider the signal model in (3) and the pilot matrix given by (11). The additive noise term $\mathbf{N}_0$ is a $M \times L$ matrix with i.i.d. $\mathcal{CN}(0, \sigma^2)$ entries. For the distortion matrix $\mathbf{W}_0$, the indices of the nonzero rows are selected uniformly at random and each nonzero entry is an i.i.d. Bernoulli random variable scaled by $4\sqrt{\rho}$, where ρ represents the power of the transmit symbol. Here, the setting of the distortion matrix is rather arbitrary and aims at generating the synthetic data for the distortion noise caused by faulty antennas. The magnitude $4\sqrt{\rho}$ is selected such that the distortion noise caused at any faulty antenna is proportional to the signal power received at that antenna. However, we remark that our proposed method does not rely on the distribution of nonzero rows or entries of the distortion matrix. The steering vector is set to $D/\lambda = 0.3$,

AOA $\theta_p = -\pi/2 + (p-1)\pi/P$, $p = 1, 2, \ldots, P$, as in [4] [6] [7]. The channel estimation error (dB) is measured by the following normalized norm

$$10 \log_{10}(\frac{1}{MK} \|\mathbf{H} - \widehat{\mathbf{H}}\|_F^2). \quad (47)$$

To measure the performances of different algorithms on faulty antenna detection, we define the faulty antenna detection error in the following way. Associated with the distortion matrix $\mathbf{W}_0$, we define a length-M binary vector $\mathbf{s}$ with entries 1 indicating the locations of faulty antennas and 0 for fault-free antennas. For example, if only the second row of $\mathbf{W}_0$ is nonzero, then we have $\mathbf{s} = 0, 1, 0, \ldots, 0$. Similarly, we can obtain an estimated binary vector $\hat{\mathbf{s}}$ based on the estimated distortion matrix $\widehat{\mathbf{W}}$. The faulty antenna detection error is defined as the hamming distance between $\mathbf{s}$ and $\hat{\mathbf{s}}$, i.e., $\mathrm{hamming}(\mathbf{s}, \hat{\mathbf{s}})$, which accounts for both miss detection and false alarm. Hence, the value of the faulty antenna detection error can range from 0 to M.

Our proposed algorithms (exAD, fsAD) and the stPCP algorithm essentially solve a matrix decomposition problem based on $\mathbf{Z}$. The selection of each pair of regularization parameters follows the approach in [48] as below. From (49), (50) and (43), we have

$$\|\mathbf{Z}\|_{\mathcal{A}_N}^* \leq \|\mathbf{Z}\|_{\mathcal{A}}^* \leq \|\mathbf{Z}\|_2, \quad (48)$$

where $\|\mathbf{Z}\|_2$ is the dual norm of $\|\mathbf{Z}\|_*$. The dual norm of $\|\mathbf{Z}\|_{2,1}$ is $\|\mathbf{Z}\|_{2,\infty} := \max_i \|\mathbf{Z}(i, :)\|_2$. The first and second regularization parameters of each algorithm are respectively upper bounded by $\|\mathbf{Z}\|_2$ and $\|\mathbf{Z}\|_{2,\infty}$, above which the optimal values of $\mathbf{H}$ and $\mathbf{W}$ are zero. Hence, the regularization parameters are set to a down scaled version of $\|\mathbf{Z}\|_2$ and $\|\mathbf{Z}\|_{2,\infty}$. We use $N = MK$ for the fast algorithm.

In the first simulation, we set the number of BS antennas $M = 120$, the number of faulty antennas $S = \lfloor 0.05M \rfloor$, number of paths $P = 20$, length of pilot sequences $L = 10$, the number of users $K = 10$ and SNR$= 10$ dB. To compare the performances of different regularization algorithms, we vary their first regularization parameters (i.e. τ_1, ν_1 and λ_1) at different scaling levels and keep the second regularization parameter fixed. The reason is that the first regularization parameters are associated with the penalties on the corresponding norms of the estimated channel matrix, which are different for each algorithm. On the other hand, the norms of the distortion matrix ($\|\mathbf{W}\|_{2,1}$) are the same for all three algorithms. We vary the scalar $\alpha = [0 : 0.05 : 1]$, and set $\{\tau_1, \nu_1, \lambda_1\} = \alpha\|\mathbf{Z}\|_2$ (exAD, fsAD and stPCP), $\{\tau_2, \nu_2, \lambda_2\} = 0.5\|\mathbf{Z}\|_{2,\infty}$. For each scalar and each algorithm, 10 iterations are run to obtain the average channel estimation error and faulty antenna detection error.

Fig. 1a shows the channel estimation error for different methods at different scaling levels of the first regularization parameter[3]. Obviously, the performances of LS, LS-SLS and MMSE based methods are independent of the value of the first regularization parameter. We note that the LS based method converges to the LS-SLS based one, while the MMSE method

---

[3]The x-axis denotes the scalar α, while the actual value of the first regularization parameter is $\alpha\|\mathbf{Z}\|_2$.
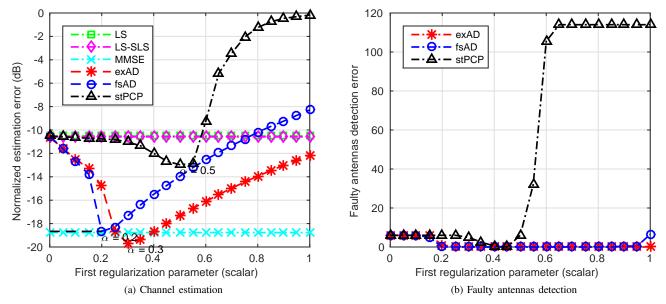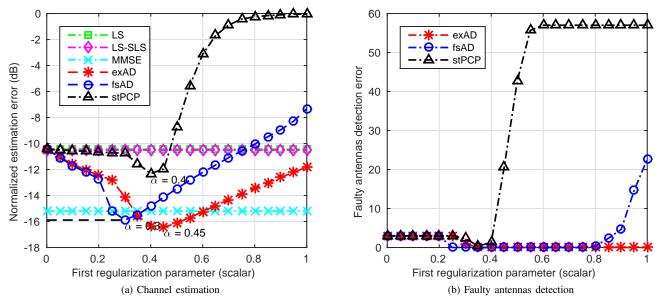
Fig. 1. Comparison of the channel estimation methods for different values of $\alpha \in [0:0.05:1]$ with $\{\tau_1, \nu_1, \lambda_1\} = \alpha \|\mathbf{Z}\|_2$, $\{\tau_2, \nu_2, \lambda_2\} = 0.5 \|\mathbf{Z}\|_{2,\infty}$, $M = 120$, $S = \lfloor 0.05M \rfloor$, $K = 10$, $P = 20$ and $L = 10$.

Fig. 2. Comparison of the channel estimation methods for different values of $\alpha \in [0:0.05:1]$ with $\{\tau_1, \nu_1, \lambda_1\} = \alpha \|\mathbf{Z}\|_2$, $\{\tau_2, \nu_2, \lambda_2\} = 0.5 \|\mathbf{Z}\|_{2,\infty}$, $M = 60$, $S = \lfloor 0.05M \rfloor$, $K = 10$, $P = 20$ and $L = 10$.

provides a better performance due to its extra assumption on knowing $\mathbf{A}$. For the regularization algorithms, the first regularization parameter traces out a path of solutions. The original channel matrices are overestimated when the scalar is too small, and over-shrunken when the scalar is too large. At scaling levels $\alpha = \{0.3, 0.2, 0.5\}$, the exAD, fsAD and stPCP respectively achieve the smallest channel estimation error. Although the MMSE method utilizes extra knowledge on the matrix $\mathbf{A}$, it is interesting to note that our proposed algorithms are comparable to the MMSE method. In addition, it indicates that both of them provide better channel estima-

tion performance than the stPCP algorithm. Furthermore, the smallest channel estimation error given by the fast algorithm (fsAD) is comparable with that from the original algorithm (exAD), which is consistent with our theoretical results in (43). In addition to channel estimation, the regularization algorithms simultaneously provide estimation on the locations of faulty antennas at the BS. Fig. 1b shows their performances on detecting faulty antennas. It indicates that the algorithms can faithfully detect faulty antennas when the scalar is within certain ranges.

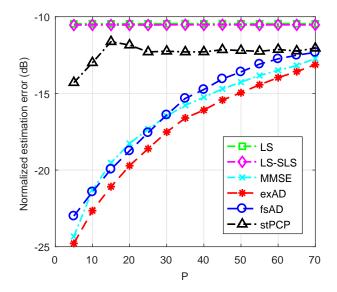For the second simulation, the setting is the same as the

Fig. 3. Comparison of the channel estimation methods for different values of P with $\alpha = \{0.3, 0.2, 0.5\}$, $\{\tau_1, \nu_1, \lambda_1\} = \alpha\|\mathbf{Z}\|_2$, $\{\tau_2, \nu_2, \lambda_2\} = 0.5\|\mathbf{Z}\|_{2,\infty}$, $M = 120$, $S = \lfloor 0.05M \rfloor$, $K = 10$, $P = 20$ and $L = 10$.

Fig. 4. Comparison of the channel estimation methods for different values of $\alpha \in [0 : 0.05 : 1]$ with $\{\tau_1, \nu_1, \lambda_1\} = \alpha\|\mathbf{Z}\|_2$, $\{\tau_2, \nu_2, \lambda_2\} = 0.5\|\mathbf{Z}\|_{2,\infty}$, $M = 120$, $S = 0$, $K = 10$, $P = 20$ and $L = 10$.

first one but the number of antennas at the BS reduced to $M = 60$. As shown in Fig. 2a, the scaling levels at which the regularization algorithms achieve smallest channel estimation error are slightly changed to $\alpha = \{0.45, 0.3, 0.4\}$, which are rather stable considering the large reduction on M. Also, Fig. 2b indicates that the ranges within which each algorithm can faithfully detect faulty antennas cover the corresponding points $\{0.45, 0.3, 0.4\}$.

The impact of the number of AoAs. In this simulation, we investigate the channel estimation performances of different algorithms with respect to different number of directions P. The setting is the same as that of the first simulation, but the scaling levels of the first regularization parameters are fixed at $\alpha = \{0.3, 0.2, 0.5\}$. As shown in Fig. 3, the performances of LS based and LS-SLS based methods are independent of the value of P. On the other hand, the performances of the rest algorithms degrade as P increases.

The impact of the faulty antennas on MMSE. In the above simulations, it is noted that the proposed algorithms can sometimes perform better than the MMSE estimation method. When the noise term is Gaussian distributed, using MMSE estimation is optimal [55]. In the above simulations, the received signal is affected by both the additive noise $\mathbf{N}_0$ and the corruption noise $\mathbf{W}_0$ due to the faulty antennas. Therefore, even with the knowledge on the angular support $\mathbf{A}$, using MMSE estimation may not always provide optimal performance since the noise term $\mathbf{N}_0 + \mathbf{W}_0$ is not Gaussian distributed. To further elaborate this point, simulation results considering zero faulty antenna are included here. In this case, the only noise term $\mathbf{N}_0$ is Gaussian distributed. The setting of this simulation is the same as the first one except for two things: the number of faulty antennas is zero, i.e. $S = 0$, and the values of the second regularization parameters $(\tau_2, \nu_2, \lambda_2)$ are set to infinity such that the estimated number of faulty
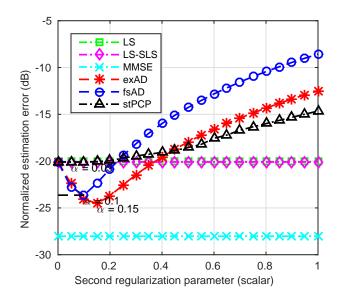
antennas would always be zero. As shown in Fig. 4, MMSE estimation exhibits the best performance, which is consistent with the theoretical understanding.

The impact of random AoAs. Next, we demonstrate the impact of the distribution of the AoAs on the performance. The simulation follows the settings of the first one except that the AoAs are chosen randomly in $[-\pi/2, \ldots, \pi/2]$. Three different resolution settings are selected for the fsAD algorithm: fsAD1_10 with $N = MK/10$, fsAD with $N = MK$ and fsAD4 with $N = 4MK$. As shown in Fig. 5, fsAD ($N = MK$) is still able to provide a low normalized estimation error. For $N = 4MK$, the performance is further improved and becomes very close to that of exAD. On the other hand, the normalized estimation error increases when $N = MK/10$, which indicates that the resolution of the grid (i.e. N) does influence the performance. The simulation results are consistent with the theoretical implications.

The effect of the second regularization parameters. In previous simulations, we vary the first regularization parameters (i.e. τ_1, ν_1 and λ_1) at different scaling levels and keep the second regularization parameters fixed. By doing so, we are able to demonstrate the effects of other factors (i.e., M, P, and AoAs). We include additional simulation results here to demonstrate the impact of the second regularization parameters on the performance. We fix the first regularization parameters and vary the second regularization parameters at different scaling levels. As shown in Fig. 6, the performances of the proposed methods (exAD and fsAD) are relatively stable when the scalars of second regularization parameters are in the range of $(0.15, 0.55)$.

Signal detection performance Lastly, we examine the signal detection performances associated with different channel estimation methods. In this simulation, we apply the same settings as in the first two but fix $\alpha = \{0.4, 0.25, 0.4\}$ and $\{\tau_2, \nu_2, \lambda_2\} = 0.5\|\mathbf{Z}\|_{2,\infty}$. We obtain the estimated channel
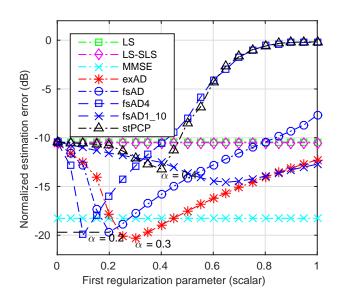
Fig. 5. Comparison of the channel estimation methods under random AoAs for different values of $\alpha \in [0:0.05:1]$ with $\{\tau_1, \nu_1, \lambda_1\} = \alpha\|\mathbf{Z}\|_2$, $\{\tau_2, \nu_2, \lambda_2\} = 0.5\|\mathbf{Z}\|_{2,\infty}$, $M = 120$, $S = \lfloor 0.05M \rfloor$, $K = 10$, $P = 20$ and $L = 10$.
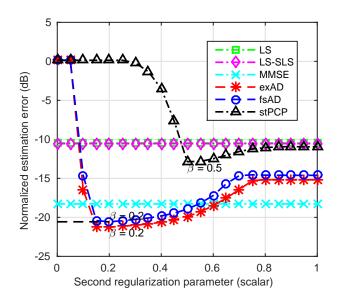

Fig. 6. Comparison of the channel estimation methods for different values of $\beta \in [0:0.05:1]$ with $\alpha = \{0.3, 0.2, 0.5\}$, $\{\tau_1, \nu_1, \lambda_1\} = \alpha\|\mathbf{Z}\|_2$, $\{\tau_2, \nu_2, \lambda_2\} = \beta\|\mathbf{Z}\|_{2,\infty}$, $M = 120$, $S = \lfloor 0.05M \rfloor$, $K = 10$, $P = 20$ and $L = 10$.

matrices by different channel estimation methods for different numbers of BS antennas ranging from 60 to 120. Then, each estimated channel matrix is employed for signal detection over 10^5 signal samples. For channel matrices estimated by LS or LS-SLS methods, the traditional ZF receiver is employed in the signal detection process due to its good performance at high SNR. Since our proposed method can estimate the channel matrix and locate the indices of faulty antennas simultaneously, the improved ZF receiver is applied for signal detection.

Fig. 7a shows the channel estimation error for different

methods. It can be seen that the proposed approaches achieve better performance. The symbol error rate for 8-PSK modulation based on different estimated channel matrices are shown in Fig. 7b. The figure also presents two oracle performances: 'PChn' represents the results obtained by assuming perfect knowledge on the channel matrix and using the traditional ZF receiver; 'PChn+O' represents the results obtained by assuming perfect knowledge on both the channel matrix and the indices of faulty antennas and employing the improved ZF receiver. The performance gap between the two oracle curves verifies our asymptotic analysis on the benefit of knowing the indices of faulty antennas in Section III. In addition, the performance gap between the MMSE and our proposed algorithms further demonstrates such a benefit. It can be seen that our proposed methods provide better signal detection performance than other methods.

V. DISCUSSIONS

A. SDP characterizations of the algorithms

Intuitively, the superior performance of our proposed method can be attributed to its better exploration on the intrinsic information, that is, the application of atomic norm instead of nuclear norm. Specifically, this can be seen through the SDP characterizations of algorithms (33) and (35).

For the atomic norm-based approach (33), we have the following SDP characterization

$$\underset{\mathbf{a},\mathbf{B},\mathbf{H}}{\operatorname{argmin}} \frac{1}{2}\operatorname{Tr}(\mathcal{T}(\mathbf{a})) + \frac{1}{2}\operatorname{Tr}(\mathbf{B}) + \mu_1\|\mathbf{H} - \mathbf{V}_1^k\|_F^2 \quad (49)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathcal{T}(\mathbf{a}) & \mathbf{H} \\ \mathbf{H}^* & \mathbf{B} \end{bmatrix} \succeq 0,$$

where $\mathcal{T}(\mathbf{a})$ is the Hermitian Toeplitz matrix with vector $\mathbf{a}$ as its first column and μ_1 depends only on τ_1.

The SDP characterization of (35) is given by

$$\underset{\mathbf{P},\mathbf{Q},\mathbf{H}}{\operatorname{argmin}} \frac{1}{2}\operatorname{Tr}(\mathbf{P}) + \frac{1}{2}\operatorname{Tr}(\mathbf{Q}) + \mu_2\|\mathbf{H} - \mathbf{V}_1^k\|_F^2 \quad (50)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{P} & \mathbf{H} \\ \mathbf{H}^* & \mathbf{Q} \end{bmatrix} \succeq 0,$$

where μ_2 depends only on λ_1. It can be observed that (50) is a relaxation of (49) by dropping the Toeplitz constraint of the first diagonal block in the positive semi-definite (PSD) feasibility condition, which explains why our proposed atomic norm-based method outperforms the stable PCP-based one.

B. Regularization parameters tuning in practice

In practical applications, the proposed atomic norm denoising-based algorithms (exAD and fsAD) depend on unknown regularization parameters that need to be chosen carefully, which is known as "parameter tuning". For general parameterized semidefinite convex optimization problems, there have been some studies on performing the tuning step offline by computing an approximated regularization path [56]–[58] and cross-validation algorithms [59]. In this part, we propose a hybrid approach to choose the regularization parameters for the atomic norm denoising algorithms. The
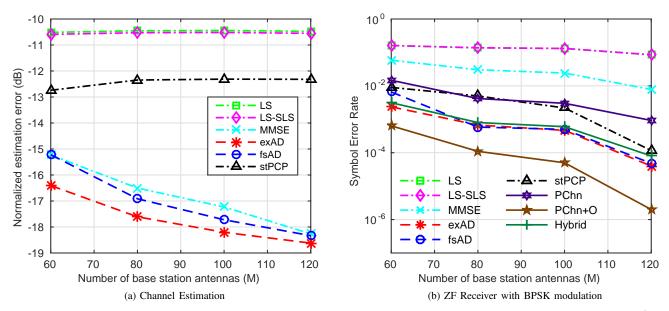
Fig. 7. Comparison of the signal detection performances associated with channel estimation methods for different values of BS antennas M with $S = \lfloor 0.05M \rfloor$, $K = 10$, $P = 20$ and $L = 10$.

difficulty of choosing the regularization parameters in practice is the lack of a reference. In other words, for each pair of regularization parameters, the quality of the estimated channel matrix can not be verified without the knowledge on the true channel matrix. From the first two simulations, it is observed that the channel estimation results given by MMSE are close to those of the proposed algorithms (exAD and fsAD). Thus, our hybrid approach is as below.

1) Assume that $\mathbf{A}$ is known, and estimate the channel matrix via MMSE.
2) Use the MMSE estimated channel matrix as a reference, and test exAD/fsAD with different pairs of regularization parameters within in a certain range.
3) Find the pair of regularization parameters such that the exAD/fsAD estimated channel matrix is close to the MMSE one.
4) The estimated indices of faulty antennas given by that pair of regularization parameters can be obtained simultaneously.
5) Do signal detection with the modified MRC/ZF receiver.

The hybrid approach does not require a priori knowledge of the regularization parameters. It uses the MMSE estimation result as a reference and searches for the suitable pair of regularization parameters. In Fig. 7b, the symbol error rate of the hybrid approach with exAD is depicted. It can be seen that the result is close to that given by the exAD method.

VI. CONCLUSION

In this paper, we have considered channel estimation and faulty antenna detection for massive MIMO systems. We have proven that the negative effect of faulty antennas on signal detection does not vanish even with unlimited number of BS antennas. To mitigate this effect, an atomic norm denoising-based approach has been proposed to simultaneously acquire

the CSI and detect the locations of faulty antennas. In addition, we have proposed a fast algorithm which has been shown to be a good approximation of the original algorithm. It has been demonstrated that both the proposed approaches outperform existing ones including conventional linear estimators and a stable PCP-based method.

We envision several directions for future work. Firstly, analysis on the effect of faulty antennas on other receivers (e.g., MMSE) and possible improvements could be investigated. Secondly, our analysis and algorithm have considered the worst case scenario, i.e, no statistical knowledge on the channel or the distortion caused by faulty antennas is assumed. With more measurement results for massive MIMO systems with faulty antennas, it may provide a statistical model for the distortion noise, and hence help develop better algorithms. Lastly, the effect of faulty antennas could be analyzed together with other factors in practical massive MIMO systems, e.g., mutual coupling, element pattern, pilot contamination, and colored noise model.

APPENDIX A
BOUND ON THE NEW ATOMIC NORM $\|\mathbf{H}\|_{\mathcal{A}_N}$

By definition, we can write the dual norm of $\|\mathbf{H}\|_{\mathcal{A}}$ as

$$
\begin{aligned}
\|\mathbf{V}\|_{\mathcal{A}}^* &= \sup_{\|\mathbf{H}\|_{\mathcal{A}} \leq 1} \langle \mathbf{V}, \mathbf{H} \rangle \\
&= \sup_{\substack{f \in [0,1) \\ \|\mathbf{u}\|_2 = 1}} \left| \left\langle \mathbf{u}, \frac{1}{\sqrt{M}} \mathbf{V}^* \mathbf{a}(f) \right\rangle \right| \\
&= \sup_{f \in [0,1)} \left\| \frac{1}{\sqrt{M}} \mathbf{V}^* \mathbf{a}(f) \right\|_2 \\
&= \sup_{f \in [0,1)} \sqrt{\sum_{k=1}^{K} |v_k(f)|^2},
\end{aligned} \tag{51}
$$

where $\langle \mathbf{V}, \mathbf{H} \rangle := \mathrm{Re}(\mathrm{Tr}(\mathbf{V}^*\mathbf{H}))$ and

$$v_k(f) = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \mathbf{V}^*_{(i,k)} \mathrm{e}^{j2\pi f(i-1)}. \tag{52}$$

Similarly, the dual norm of $\|\mathbf{H}\|_{\mathcal{A}_N}$ is given by

$$\|\mathbf{V}\|^*_{\mathcal{A}_N} = \sup_{f \in [(N-1)/N]} \sqrt{\sum_{k=1}^{K} |v_k(f)|^2}, \tag{53}$$

where $[(N-1)/N] := \{0, 1/N, \ldots, (N-1)/N\}$.

For $f_1, f_2 \in [0, 1)$, by Bernstein's theorem (Lemma A.1) and a result in [60, Appendix B], we have

$$|v_k(f_1)| - |v_k(f_1)| \le 2\pi M |f_1 - f_2| \sup_{f \in [0,1)} |v_k(f)| \tag{54}$$

Lemma A.1 ([61]). *Let $q(z)$ be any polynomial of degree n with complex coefficients and derivative $q'(z)$. Then,*

$$\sup_{|z|\le 1} |q'(z)| \le n \sup_{|z|\le 1} |q(z)|. \tag{55}$$

Thus,

$$\sum_k |v_k(f_1)|^2 - \sum_k |v_k(f_2)|^2 \tag{56}$$

$$\le \sum_k (|v_k(f_1)| + |v_k(f_2)|)2\pi M|f_1 - f_2| \sup_{f\in[0,1)} |v_k(f)|$$

$$\le \sum_k 4\pi M|f_1 - f_2| \sup_{f\in[0,1)} |v_k(f)|^2$$

$$\le 4\pi MK|f_1 - f_2| \sup_{f\in[0,1)} \sum_k |v_k(f)|^2$$

$$= 4\pi MK|f_1 - f_2|(\|\mathbf{V}\|^*_{\mathcal{A}})^2. \tag{57}$$

Note that for any $f_1 \in [0, 1)$, there is a $f_2 \in [(N-1)/N]$ with $|f_1 - f_2| \le \frac{1}{2N}$, then (57) gives

$$\sup_{f\in[0,1)} \sum_k |v_k(f)|^2 \le \sup_{f\in[(N-1)/N]} \sum_k |v_k(f)|^2$$
$$+ \frac{2\pi MK}{N}(\|\mathbf{V}\|^*_{\mathcal{A}})^2.$$

By substituting (51) and (53), it becomes

$$(\|\mathbf{V}\|^*_{\mathcal{A}})^2 \le (\|\mathbf{V}\|^*_{\mathcal{A}_N})^2 + \frac{2\pi MK}{N}(\|\mathbf{V}\|^*_{\mathcal{A}})^2.$$

Hence,

$$\|\mathbf{V}\|^*_{\mathcal{A}} \le (1 - \frac{2\pi MK}{N})^{-\frac{1}{2}} \|\mathbf{V}\|^*_{\mathcal{A}_N}. \tag{58}$$

From (51) and (53), we have

$$\|\mathbf{V}\|^*_{\mathcal{A}_N} \le \|\mathbf{V}\|^*_{\mathcal{A}}. \tag{59}$$

Thus,

$$\|\mathbf{V}\|^*_{\mathcal{A}_N} \le \|\mathbf{V}\|^*_{\mathcal{A}} \le (1 - \frac{2\pi MK}{N})^{-\frac{1}{2}} \|\mathbf{V}\|^*_{\mathcal{A}_N}, \tag{60}$$

and

$$(1 - \frac{2\pi MK}{N})^{\frac{1}{2}} \|\mathbf{H}\|_{\mathcal{A}_N} \le \|\mathbf{H}\|_{\mathcal{A}} \le \|\mathbf{V}\|_{\mathcal{A}_N}, \tag{61}$$

which completes the proof.

We note that the above techniques of bounding the dual norm associated with frequencies defined on uniform grids have been used in [44], [45], [60]. Our proof generalizes the line spectral estimation in [44], [60] into a mutiple measurement vectors model and improves the lower bound in [45].

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, February 2014.

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[4] J. Hoydis, S. Ten Brink, M. Debbah *et al.*, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[6] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *IEEE ICASSP'11, Prague*, 2011, pp. 3464–3467.

[7] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *IEEE WCNC'13*, 2013, pp. 2890–2895.

[8] P. Zhang, L. Gan, S. Sun, and C. Ling, "Atomic norm denoising-based channel estimation for massive multiuser MIMO systems," in *IEEE ICC'15, London*, June 2015, pp. 4564–4569.

[9] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, 2006.

[10] X. Rao and V. K. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, 2014.

[11] S. Pejoski and V. Kafedziski, "Estimation of sparse time dispersive channels in pilot aided OFDM using atomic norm," *IEEE Wireless Commun. Letters*, vol. 4, no. 4, pp. 397–400, 2015.

[12] ——, "Joint atomic norm based estimation of sparse time dispersive simo channels with common support in pilot aided OFDM systems," *Mobile Networks and Applications*, pp. 1–11, 2016.

[13] T. J. Peters, "A conjugate gradient-based algorithm to minimize the sidelobe level of planar arrays with element failures," *IEEE Trans. Antennas and Propagation*, vol. 39, no. 10, pp. 1497–1504, 1991.

[14] B.-K. Yeo and Y. Lu, "Array failure correction with a genetic algorithm," *IEEE Trans. Antennas and Propagation*, vol. 47, no. 5, pp. 823–828, 1999.

[15] O. P. Acharya, A. Patnaik, and S. N. Sinha, "Null steering in failed antenna arrays," *Applied Computational Intelligence and Soft Computing*, vol. 2011, p. 4, 2011.

[16] R. C. de Lamare, "Massive MIMO systems: signal processing challenges and future trends," *URSI Radio Science Bulletin*, 2013.

[17] E. Bjornson, P. Zetterberg, M. Bengtsson, and B. Ottersten, "Capacity limits and multiplexing gains of MIMO channels with transceiver impairments," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 91–94, 2013.

[18] T. Koch, A. Lapidoth, and P. P. Sotiriadis, "Channels that heat up," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3594–3612, 2009.

[19] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 303–313, 2013.

[20] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, 2013.

[21] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[22] T. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer, 2008.

[23] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-rf impairments," in *Proc. ITG/IEEE Workshop on Smart Antennas (WSA)*. IEEE, 2010, pp. 189–196.

[24] J. A. Rodríguez-González, F. Ares-Pena, M. Fernández-Delgado, R. Iglesias, and S. Barro, "Rapid method for finding faulty elements in antenna arrays using far field pattern samples," *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1679–1683, 2009.

[25] J. Lee, E. M. Ferren, P. D. Woollen, and K. M. Lee, "Near-field probe used as a diagnostic tool to locate defective elements in an array antenna," *IEEE Trans. Antennas Propag.*, vol. 36, no. 6, pp. 884–889, 1988.

[26] O. Bucci, A. Capozzoli, and G. D'elia, "Diagnosis of array faults from far-field amplitude-only data," *IEEE Trans. Antennas Propag.*, vol. 48, no. 5, pp. 647–652, 2000.

[27] M. D. Migliore, "A compressed sensing approach for array diagnosis from a small set of near-field measurements," *IEEE Trans. Antennas Propag.*, vol. 59, no. 6, pp. 2127–2133, 2011.

[28] G. Oliveri, P. Rocca, and A. Massa, "Reliable diagnosis of large linear arrays—a Bayesian compressive sensing approach," *IEEE Trans. Antennas Propag.*, vol. 60, no. 10, pp. 4627–4636, 2012.

[29] A. Dua, K. Medepalli, and A. J. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2353–2357, 2006.

[30] S. Mahboob, R. Ruby, and V. Leung, "Transmit antenna selection for downlink transmission in a massively distributed antenna system using convex optimization," in *Broadband, Wireless Computing, Communication and Applications (BWCCA), 2012 Seventh International Conference on*. IEEE, 2012, pp. 228–233.

[31] X. Gao, O. Edfors, J. Liu, and F. Tufvesson, "Antenna selection in measured massive MIMO channels using convex optimization," in *Globecom Workshops (GC Wkshps), 2013 IEEE*. IEEE, 2013, pp. 129–134.

[32] X. Gao, O. Edfors, F. Tufvesson, and E. Larsson, "Massive MIMO in real propagation environments: do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. PP, no. 99, pp. 1–1, 2015.

[33] H. Ngo, E. Larsson, and T. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, June 2013.

[34] J.-A. Tsai, R. M. Buehrer, and B. D. Woerner, "The impact of AOA energy distribution on the spatial fading correlation of linear antenna array," in *Proc. IEEE VTC'02*, vol. 2, May 2002, pp. 933–937.

[35] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Energy-efficient downlink transmission for multicell massive DAS with pilot contamination," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1209–1221, 2017.

[36] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Spatial DCT-based channel estimation in multi-antenna multi-cell interference channels," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1404–1418, 2015.

[37] R. R. Muller, "A random matrix model of communication via antenna arrays," *IEEE Transactions on information theory*, vol. 48, no. 9, pp. 2495–2506, 2002.

[38] A. G. Burr, "Capacity bounds and estimates for the finite scatterers MIMO wireless channel," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 812–818, 2003.

[39] J. Fuhl, A. F. Molisch, and E. Bonek, "Unified channel model for mobile radio systems with smart antennas," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 145, no. 1, pp. 32–41, 1998.

[40] P. Petrus, J. H. Reed, and T. S. Rappaport, "Geometrical-based statistical macrocell channel model for mobile environments," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 495–502, 2002.

[41] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 361–368, 2002.

[42] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 264–273, 2013.

[43] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[44] B. N. Bhaskar and B. Recht, "Atomic norm denoising with applications to line spectral estimation," in *IEEE Allerton'11*, 2011, pp. 261–268.

[45] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *arXiv preprint arXiv:1408.2242*, 2014.

[46] D. N. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 171–188, 2000.

[47] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Chapter 5 of Compressed sensing: theory and applications*, Cambridge University Press, 2012.

[48] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[49] Y. Chi, "Convex relaxations of spectral sparsity for robust super-resolution and line spectrum estimation."

[50] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *arXiv preprint arXiv:1609.02247*, 2016.

[51] L. Zheng and X. Wang, "Super-resolution delay-doppler estimation for OFDM passive radar," *arXiv preprint arXiv:1610.04218*, 2016.

[52] M. D. Migliore, "Minimum trace norm regularization (MTNR) in electromagnetic inverse problems," *IEEE Trans. Antennas Propag.*, vol. 64, no. 2, pp. 630–639, 2016.

[53] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, "Stable principal component pursuit," in *2010 IEEE International Symposium on Information Theory*. IEEE, 2010, pp. 1518–1522.

[54] N. S. Aybat and G. Iyengar, "An alternating direction method with increasing penalty for stable principal component pursuit," *Computational Optimization and Applications*, vol. 61, no. 3, pp. 635–668, 2015.

[55] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[56] A. d'Aspremont, F. R. Bach, and L. E. Ghaoui, "Full regularization path for sparse principal component analysis," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 177–184.

[57] J. Giesen, M. Jaggi, and S. Laue, "Regularization paths with guarantees for convex semidefinite optimization." in *AISTATS*, 2012, pp. 432–439.

[58] G. Mateos and G. B. Giannakis, "Robust PCA as bilinear decomposition with outlier-sparsity regularization," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5176–5190, 2012.

[59] S. Arlot, A. Celisse *et al.*, "A survey of cross-validation procedures for model selection," *Statistics surveys*, vol. 4, pp. 40–79, 2010.

[60] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, 2015.

[61] A. Schaeffer, "Inequalities of A. Markoff and S. Bernstein for polynomials and related functions," *Bulletin of the American Mathematical Society*, vol. 47, no. 8, pp. 565–579, 1941.