Compressive Sampling Optimization for User Signal Parameter Estimation in Massive MIMO Systems

Yujie Gu and Yimin D. Zhang

Abstract

As the most promising technology in wireless communications, massive multiple-input multiple-output (MIMO) faces a significant challenge in practical implementation because of the high complexity and cost involved in deploying a separate front-end circuit for each antenna. In this paper, we apply the compressive sampling technique to reduce the number of required front-end circuits in the analog domain and the computational complexity in the digital domain. Unlike the commonly adopted random projections, we exploit the *a priori* probability distribution of the user positions to optimize the compressive sampling strategy, so as to maximize the mutual information between the compressed measurements and the direction-of-arrival (DOA) of user signals. With the optimized compressive sampling strategy, we further propose a compressive sampling Capon spatial spectrum estimator for DOA estimation. In addition, the user signal power is estimated by solving a compressed measurement covariance matrix fitting problem. Furthermore, the user sig-

☆Part of the results was presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing [1].

The authors are with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA, 19122 USA (e-mail: guyujie@hotmail.com; ydzhang@temple.edu).

nal waveforms are estimated from a robust adaptive beamformer through the reconstruction of an interference-plus-noise compressed covariance matrix. Simulation results clearly demonstrate the performance advantages of the proposed techniques for user signal parameter estimation as compared to existing techniques.
Keywords: Adaptive beamforming, compressive sampling optimization, massive MIMO, mutual information, parameter estimation.

1. Introduction

Massive multiple-input multiple-output (MIMO) is recognized as one of the most promising techniques in the next generation (5G and beyond) cellular wireless communication systems. By employing a high number of antennas at the base station, a massive MIMO system provides a number of advantages, including increased system capacity, energy efficiency, security and robustness [2, 3, 4, 5, 6]. While most current research activities mainly focus on communication applications, such as mobile broadband, low-power machine-type or ultra-reliable communications, massive MIMO is also very appealing for non-communication applications such as radar, sensing and positioning [7]. Among them, the location information of the users can provide added values, e.g., for location-aware services and smart cities [8]. Compared to previous cellular wireless communication techniques, massive MIMO enables high-accuracy user localization benefiting from its high number of antennas and large array aperture [9]. For instance, a fingerprinting-based positioning solution is proposed for locating mobile users in distributed massive MIMO systems [10]. A personal radar integrating millimeter-wave massive arrays in a smartphone or tablet enables accurate indoor simultaneous localization and mapping (SLAM) [11, 12]. An estimation of signal parameters via rotational invariance techniques (ESPRIT)-based approach and a beamspace approach are respectively proposed for two-dimensional (2-D) localization of incoherently distributed sources in massive MIMO systems in [13] and [14]. In [15], a source enumeration algorithm in the framework of the Bayesian information criterion is proposed to provide reliable detection of the number of sources in a large-scale adaptive antenna array. A direct source localization technique is proposed in [16] to jointly process the snapshots of data acquired at each base station for direct estimation of the source location.

Despite such advantages of massive MIMO for both communications and user localization, it remaines challenging to equip a dedicated radio frequency (RF) front-end chain and a high-resolution analog-digital converter (ADC) for each antenna in massive MIMO systems. An RF front-end chain generally consists of a band-pass filter, a low-noise amplifier, a mixer and a low-pass filter. The ADC then digitizes the analog signal to obtain the baseband digital signal to perform subsequent signal processing. Among them, the ADC is a significant limiting factor from the perspective of cost and power consumption. On the other hand, the computational complexity is another major concern in massive MIMO systems for optimized beamforming because of the high-dimensional matrix operation involved in either inversion or eigendecomposition. Hence, there is an urgent requirement to provide a hybrid analog-digital processing strategy that simultaneously reduces the required number of both RF front-end chains and ADCs in the analog domain and the corresponding computational complexity in the digital domain. As a widely recognized signal processing technique, compressive sampling is an effective and feasible solution to achieve this object in massive MIMO systems.

Compressive sampling, also well known as compressive sensing, is a popular signal acquisition and recovery technique which provides solutions to underdetermined linear systems for sparse signals [17]. The sparsity of a signal implies that the signal can be sparsely represented in some domain or basis. In this case, a high-dimensional signal can be reduced to some low-dimensional manifold through a proper transformation. According to the compressive sampling theory, the existence of signal sparsity allows signal recovery from far fewer samples than those required by the classical Shannon-Nyquist sampling theory. In order to solve the underdetermined signal recovery problem, the incoherence between the compressive measurement matrix and the signal's sparse representation basis is required to satisfy the restricted isometry property (RIP) [17, 18]. It is known that random compressive measurement matrices, such as Gaussian and Bernoulli matrices, satisfy this incoherence requirement with a high probability for accurate signal recovery [18, 19, 20].

In the past decade, the compressive sampling technique has been widely applied in array signal processing for direction-of-arrival (DOA) estimation [21, 22, 23, 24, 25, 26, 27] and adaptive beamformer design [28, 29, 30, 31]. For example, the compressive sampling technique is applied in the time domain to reduce the ADC sampling rate or the number of time samples in each array element for bearing estimation [21] and beamforming [28], respectively. In [22], compressing sampling is exploited in the spatial domain to compress the array of a large number of elements into an array of a much smaller number of elements. In [26], the signal received in a coprime array is compressively sampled to a lower-dimensional sketch to perform DOA estimation with a high resolution and an increased number of degrees-of-freedom. In all these approaches, random measurement matrices are adopted for compressive sampling to satisfy the incoherence requirement, which guarantees the recovery of the sparse signal from sub-Nyquist samples with a high probability [18, 19, 20]. Clearly, as a data-independent sampling scheme, random sampling is robust but does not exploit the *a priori* knowledge of the signals beyond the sparsity. In practical applications, however, prior knowledge of the signals of interest is usually available and can be used to optimize the compressive sampling [32, 33].

In massive MIMO systems, the *a priori* knowledge of the user signals, such as the distribution of their spatial positions, can be learned and updated from the observed data. Indeed, mobile users are often clustered but with time-varying patterns. By taking this fact into account, the objective of this paper is to achieve effective compressive sampling with a minimum loss in the output performance. Toward this end, we exploit the available *a priori* information of the spatial distribution of user positions to optimize the compressive sampling strategy for the estimation of user signal parameters. As such, both the number of required front-end circuits (i.e., the RF front-end chains and the ADCs) and the complexity of subsequent digital processing (i.e., the computational complexity) are significantly reduced, while the advantages offered by the large array aperture and array gain of the massive MIMO systems are maintained. Unlike random projections, the proposed compressive sampling is optimized to maximize the mutual information between the compressed measurements and the signal DOA, where the compressed measurements are parametrically characterized by a Gaussian mixture model through discretizing the *a priori* probability distribution of the signal DOA. Considering that the mutual information is not a convex function with respect to (w.r.t.) the compressive sampling matrix, we derive the approximate mutual information gradient to search for the optimal compressive sampling matrix in a gradient ascent direction.

To describe the users' information in a massive MIMO system, the key parameters of interest generally include the DOA, power and waveform of each user signal. After the compressive sampling strategy is optimized, we then develop a compressive sampling Capon (CS-Capon) spatial spectrum estimator for the DOA estimation of the user signals. Considering that the Capon spatial spectrum grossly underestimates the power in the case of limited snapshots [34], we then formulate a compressed measurement covariance matrix fitting problem to achieve improved power estimation. Based on the estimated DOAs of the user signals and their power, the waveforms of the user signals are then estimated from robust adaptive beamformers, which are designed based on the reconstruction of the interference-plus-noise compressed covariance matrix [34], to separate them from the *mixed* compressed measurements. Simulation results demonstrate the performance advantage of the proposed compressive sampling methods on user signal parameter estimation over the random projections.

Compared to the preliminary results presented in [1], which primarily focuses on the optimization of the compressive sampling for DOA estimation, this paper provides user signal parameter estimation in a more general context. More specifically, we further propose the compressed measurement covariance matrix fitting optimization approach to achieve improved power estimation, and a robust adaptive beamforming technique is developed based on the reconstruction of the interference-plus-noise compressed covariance matrix to achieve improved waveform estimation.

The rest of the paper is organized as follows. In Section 2, we introduce the compressive array signal model in massive MIMO systems. In Section 3, we present the probabilistic signal model and propose an information-theoretic compressive sampling optimization. In Section 4, we propose compressive sampling-based technique to estimate the key user parameters, including DOA, power and waveform. In Section 5, we present simulation results to assess the parameter estimation performance of the proposed compressive sampling techniques which are compared to those based on conventional random projections. Conclusions are provided in Section 6.

2. Array structure and signal model

2.1. Array signal model

Assume D uncorrelated users (i.e., sources) which are operated in the same frequency band and impinge on a massive MIMO array equipped with N antennas from the directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_D]^{\mathrm{T}}$, where N is an order of magnitude more elements than in systems being built today, say 50 antennas or more. Here, $(\cdot)^{\mathrm{T}}$ denotes the transpose operation. The array received signal in the baseband measured at the t-th sampling time, $\boldsymbol{x}(t) \in \mathbb{C}^N$, can

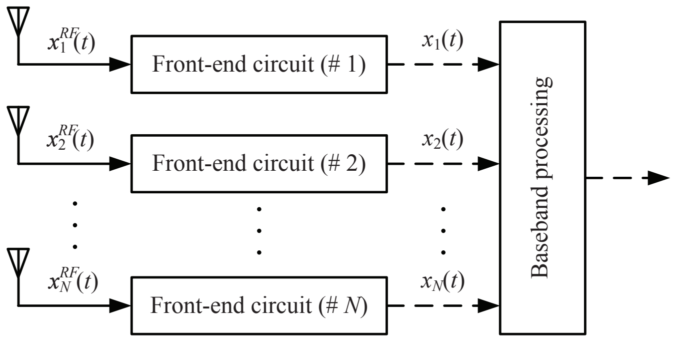


Figure 1: Block diagram of an antenna array, where solid lines denote analog signals and dashed lines denote digital signals.

be modeled as

$$\begin{aligned}\boldsymbol{x}(t) &= \sum_{d=1}^{D} \boldsymbol{a}(\theta_d) s_d(t) + \boldsymbol{n}(t) \\ &= \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{s}(t) + \boldsymbol{n}(t), \end{aligned} \tag{1}$$

where $\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_D)] \in \mathbb{C}^{N \times D}$ denotes the array manifold matrix whose column $\boldsymbol{a}(\theta_d) \in \mathbb{C}^N$ represents the steering vector of the d-th user with DOA θ_d, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_D(t)]^{\mathrm{T}} \in \mathbb{C}^D$ represents the signal waveform vector, and $\boldsymbol{n}(t) \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}\right)$ represents the zero-mean additive white Gaussian noise (AWGN) vector. Here, σ_n^2 denotes the noise power, and $\boldsymbol{I}$ denotes an identity matrix with appropriate dimensions.

As shown in the system block diagram of Fig. 1, each antenna of the receive array is equipped with a dedicated separate front-end chain. The front-end chain transforms the received RF signal to the digital baseband signal by performing low-noise amplification, down-conversion, low-pass filtering and analog-digital conversion in turn, and the result is then fed to a digital signal processor for further digital signal processing. In the case

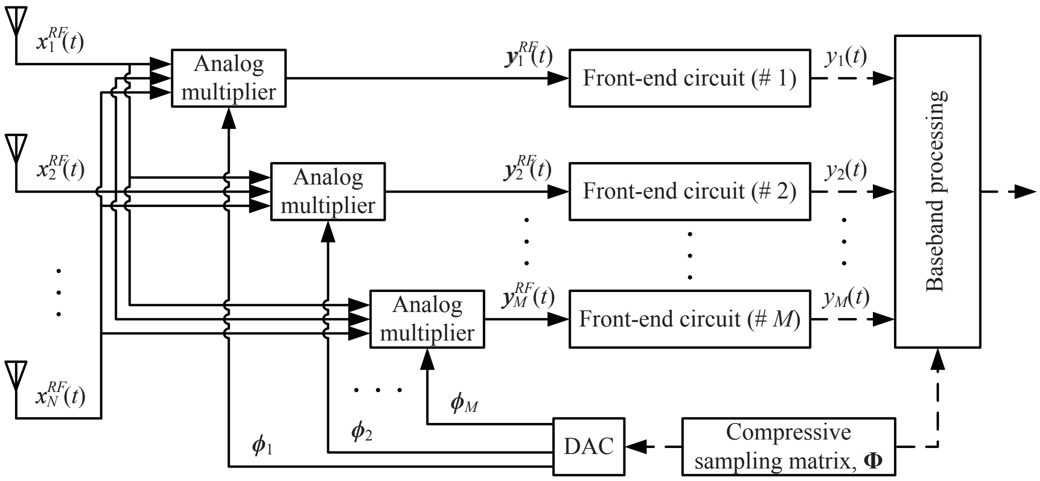


Figure 2: Block diagram of a compressive sampling antenna array, where solid lines denote analog signals and dashed lines denote digital signals.

of massive MIMO systems with $N \gg 1$ antennas at the base station, it requires a huge overhead to exploit the high number of front-end circuits and implement the high-dimensional matrix operations.

2.2. Compressive array signal model

In massive MIMO systems, the number of users is typically much less than the number of antennas at the base station, not to mention the number of possible positions in the spatial domain. It implies that the array received signals have a sparse representation and can be limited to a low-dimensional signal subspace manifold. Motivated by the successful applications of compressive sampling in the past decade, in this paper, we consider compressive sampling of the array received signal as depicted in the system block diagram in Fig. 2, where the array received signals are compressed in the analog domain before passing through much fewer number of front-end chains. As such, the number of required front-end chains is greatly reduced, thus subsequently decreasing the computational complexity drastically.

In specific, we measure $M \ll N$ linear projections of the array received signal (in the analog domain), $\boldsymbol{x}^{RF}(t)$, onto a set of measurement kernels $\{\boldsymbol{\phi}_m, m = 1, \cdots, M\}$. The m-th measurement $y_m^{RF}(t)$ is the projection of the array received signal $\boldsymbol{x}^{RF}(t)$ onto the m-th measurement kernel $\boldsymbol{\phi}_m$, expressed as

$$\begin{aligned} y_m^{RF}(t) &= \langle \boldsymbol{\phi}_m, \boldsymbol{x}^{RF}(t) \rangle \\ &= \sum_{n=1}^{N} \phi_{m,n} x_n^{RF}(t), \end{aligned} \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product operator, and $\boldsymbol{\phi}_m = [\phi_{m,1}, \phi_{m,2}, \cdots, \phi_{m,N}] \in \mathbb{C}^{1\times N}$ is the discrete representation of the analog measurement kernel. Here, the analog multiplication is generally implemented by an attenuator on each branch, or simply by a phase shifter in case of a Bernoulli distribution for the entries of the measurement kernel. In order not to destroy the essential information of the original array received signal, the number of compressed branches, M, should be selected to ensure that it is greater than the number of expected users.

Stacking the M measurement kernels as a compressive sampling matrix $\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1^{\mathrm{T}}, \boldsymbol{\phi}_2^{\mathrm{T}}, \cdots, \boldsymbol{\phi}_M^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times N}$ yields an M-dimensional compressed measurement vector in baseband, $\boldsymbol{y}(t) = [y_1(t), y_2(t), \cdots, y_M(t)]^{\mathrm{T}} \in \mathbb{C}^M$, expressed as

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}\boldsymbol{x}(t) = \boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{\Phi}\boldsymbol{n}(t), \tag{3}$$

where $\boldsymbol{\Phi}\boldsymbol{A}(\theta) = [\boldsymbol{\Phi}\boldsymbol{a}(\theta_1), \boldsymbol{\Phi}\boldsymbol{a}(\theta_2), \cdots, \boldsymbol{\Phi}\boldsymbol{a}(\theta_D)] \in \mathbb{C}^{M\times D}$ is a compressed array manifold matrix with a significantly reduced dimension because $M \ll N$, and $\boldsymbol{\Phi}\boldsymbol{n}(t) \sim \mathcal{CN}\big(\mathbf{0}, \sigma_n^2 \boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}}\big)$ is the zero-mean additive Gaussian noise vector with dimension $M \times 1$. Here, $(\,\cdot\,)^{\mathrm{H}}$ denotes the Hermitian transpose operation.

Unlike using random projections, our objective is to optimize the compressive sampling matrix $\boldsymbol{\Phi}$ according to the prior knowledge of the user distribution in the spatial domain to further improve the estimation accuracy of user signal parameters including direction, power and waveform, from the compressed measurements.

3. Compressive sampling optimization

By exploiting the available probability distribution of the signal DOA θ, we adopt the mutual information maximization criterion in this section to optimize the compressive sampling matrix $\boldsymbol{\Phi}$ for the parameter estimation of user signals.

3.1. Probabilistic signal model

In our compressive sampling optimization, the DOA θ is treated as a random variable with *a priori* known probability density function (pdf) $f(\theta)$. Such probability distribution can be estimated either from the previously observed data or from the tracking information of current users in the massive MIMO system.

Applying the law of total probability, the pdf of the compressed measurement vector $\boldsymbol{y}$ can be expressed as

$$f(\boldsymbol{y}) = \mathbb{E}_\theta \left\{ f(\boldsymbol{y}|\theta) \right\} = \int_{\theta\in\Theta} f(\boldsymbol{y}|\theta) f(\theta)\, d\theta, \tag{4}$$

where $\mathbb{E}_\theta\{\cdot\}$ denotes the statistical expectation operator w.r.t. the variable θ, and Θ is the region of observation. Discretizing the pdf $f(\theta)$ into K angular bins with an equal width of $\Delta\bar{\theta}$, the pdf of the measurement vector $\boldsymbol{y}$ can be approximated as

$$f(\boldsymbol{y}) \approx \sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}|\bar{\theta}_k), \tag{5}$$

where $p_k = f(\bar{\theta}_k)\Delta\bar{\theta}$ denotes the probability of the k-th angular bin with $\sum_{k\in\mathcal{K}} p_k = 1$, and $f(\boldsymbol{y}|\bar{\theta}_k)$ denotes the corresponding conditional pdf of the measurement vector $\boldsymbol{y}$ given the specific DOA $\bar{\theta}_k$. Here, $\mathcal{K} = \{1, 2, \cdots, K\}$ denotes the index set of the angular bins.

The signal arrived in a specific angular bin can be regarded as a mixture of different user signals, and is modeled as a zero-mean complex Gaussian random variable $s(t) \sim \mathcal{CN}(0, \sigma_s^2)$, where σ_s^2 denotes the signal power. When the signal impinges from the k-th angular bin with a nominal DOA $\bar{\theta}_k$, the compressed measurement vector is given by

$$\boldsymbol{y}|_{\theta=\bar{\theta}_k} = \boldsymbol{\Phi}\big(\boldsymbol{a}(\bar{\theta}_k)s(t) + \boldsymbol{n}(t)\big) \tag{6}$$

with conditional pdf

$$f(\boldsymbol{y}|\bar{\theta}_k) = \frac{1}{\pi^M \left|\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}\right|} e^{-\boldsymbol{y}^{\mathrm{H}}\boldsymbol{C}^{-1}_{\boldsymbol{yy}|\bar{\theta}_k}\boldsymbol{y}}, \tag{7}$$

where

$$\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k} = \boldsymbol{\Phi}\left(\sigma_s^2 \boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \sigma_n^2 \boldsymbol{I}\right)\boldsymbol{\Phi}^{\mathrm{H}} \tag{8}$$

is the compressed measurement covariance matrix given the specific DOA $\bar{\theta}_k$, and $|\cdot|$ denotes the determinant of a matrix.

As a result, the pdf of the compressed measurement vector $\boldsymbol{y}$ is approximated as a weighted sum of K zero-mean Gaussian distributions $\{f(\boldsymbol{y}|\bar{\theta}_k); k \in$

$\mathcal{K}\}$, which forms a Gaussian mixture distribution with the parameters $\{p_k, \mathbf{0}, \boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}; k \in \mathcal{K}\}$. When the angular bin width $\Delta\bar{\theta}$ is chosen to be sufficiently small, the approximation in (5) approaches to equality.

3.2. Optimization of the compressive sampling matrix

According to the Gaussian mixture distribution of the compressed measurement vector, we adopt the mutual information maximization criterion [35] to optimize the compressive sampling matrix $\boldsymbol{\Phi}$ such that the user signal parameters can be estimated more accurately. Considering the fact that the optimization variable $\boldsymbol{\Phi}$ is a high-dimensional matrix, we choose a gradient ascent approach to search for the optimal compressive sampling matrix.

The gradient of the mutual information $I(\boldsymbol{y}; \theta)$ between the compressed measurement vector $\boldsymbol{y}$ and the signal DOA θ w.r.t. the compressive sampling matrix $\boldsymbol{\Phi}$ is computed as

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y}; \theta) = \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}) - \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}|\theta), \tag{9}$$

where $\nabla_{\boldsymbol{\Phi}}\{\,\cdot\,\}$ is the gradient operator w.r.t. $\boldsymbol{\Phi}$, $h(\boldsymbol{y}) = -\mathbb{E}_{\boldsymbol{y}}\{\log[f(\boldsymbol{y})]\}$ denotes the differential entropy of the compressed measurement vector $\boldsymbol{y}$, and $h(\boldsymbol{y}|\theta) = -\mathbb{E}_{\boldsymbol{y},\theta}\{\log[f(\boldsymbol{y}|\theta)]\}$ represents the conditional differential entropy of the compressed measurement vector $\boldsymbol{y}$ given the signal DOA θ. In general, there is no closed-form expression for the mutual information gradient. However, the discretization of the pdf of DOA makes it easy to derive an approximate mutual information gradient.

The conditional pdf of the mean value of the compressed measurement vector $\boldsymbol{y}$ ($\boldsymbol{y}_0 = \mathbf{0}$) given the signal DOA θ is given by

$$f(\boldsymbol{y}_0|\theta) = \frac{1}{\pi^M \left|\boldsymbol{C}_{\boldsymbol{yy}|\theta}\right|}, \tag{10}$$

where

$$\boldsymbol{C}_{\boldsymbol{yy}|\theta} = \boldsymbol{\Phi}\big(\sigma_s^2 \boldsymbol{a}(\theta)\boldsymbol{a}^{\mathrm{H}}(\theta) + \sigma_n^2 \boldsymbol{I}\big)\boldsymbol{\Phi}^{\mathrm{H}} \tag{11}$$

is the compressed measurement covariance matrix given the DOA θ. Using the approximation in (5), the mutual information can be expressed as

$$\begin{aligned} I(\boldsymbol{y};\theta) &= -\int f(\boldsymbol{y})\log f(\boldsymbol{y})d\boldsymbol{y} + \iint f(\boldsymbol{y},\theta)\log f(\boldsymbol{y}|\theta)d\boldsymbol{y}d\theta \\ &\approx -\log\sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}_0|\bar{\theta}_k) + \iint f(\boldsymbol{y},\theta)\log f(\boldsymbol{y}_0|\theta)\, d\boldsymbol{y}d\theta \\ &= -\log\sum_{k\in\mathcal{K}} \frac{p_k}{\pi^M \left|\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}\right|} + \int \log\frac{1}{\pi^M\left|\boldsymbol{C}_{\boldsymbol{yy}|\theta}\right|} f(\theta)d\theta \\ &\approx -\log\sum_{k\in\mathcal{K}} \frac{p_k}{\left|\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}\right|} - \sum_{k\in\mathcal{K}} p_k \log\left|\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}\right|, \end{aligned} \tag{12}$$

where the first approximation is based on the first-order Taylor series expansion of $\log f(\boldsymbol{y})$ and $\log f(\boldsymbol{y}|\theta)$ around the mean value $\boldsymbol{y}_0 = \boldsymbol{0}$, respectively.

The gradient of the approximate mutual information $I(\boldsymbol{y};\theta)$ in (12) w.r.t. the compressive sampling matrix $\boldsymbol{\Phi}$ is then computed as

$$\begin{aligned} \nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\theta) \approx{}& \frac{\displaystyle\sum_{k\in\mathcal{K}} \frac{p_k}{\left|\frac{\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_n^2}\right|} \left[\frac{\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_n^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_s^2}{\sigma_n^2}\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \boldsymbol{I}\right)}{\displaystyle\sum_{k\in\mathcal{K}} p_k \left|\frac{\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_n^2}\right|^{-1}} \\ &- \sum_{k\in\mathcal{K}} p_k \left[\frac{\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_n^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_s^2}{\sigma_n^2}\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \boldsymbol{I}\right), \end{aligned} \tag{13}$$

where σ_s^2/σ_n^2 denotes the signal-to-noise ratio (SNR) of the input signal, and

$$\frac{\boldsymbol{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_n^2} = \boldsymbol{\Phi}\left(\frac{\sigma_s^2}{\sigma_n^2}\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \boldsymbol{I}\right)\boldsymbol{\Phi}^{\mathrm{H}}. \tag{14}$$

To maximize the mutual information, we use the approximate mutual information gradient $\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\theta)$ in (13) to iteratively search for the optimal compressive sampling matrix in the gradient-ascent direction according to

$$\tilde{\mathbf{\Phi}} = \mathbf{\Phi} + \gamma \nabla_{\mathbf{\Phi}} I(\boldsymbol{y}; \theta), \tag{15}$$

where $\gamma > 0$ is a small step size. To avoid increasing the mutual information simply by scaling up $\mathbf{\Phi}$, we introduce a row-orthonormal constraint [1] on $\mathbf{\Phi}$, i.e., $\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}} = \boldsymbol{I}$. Hence, the procedure is iterated by re-orthonormalizing the rows of the updated compressive sampling matrix $\tilde{\mathbf{\Phi}}$ (15) and computing the approximate mutual information gradient $\nabla_{\tilde{\mathbf{\Phi}}} I(\boldsymbol{y}; \theta)$ (13). The iterative process continues until either the convergence or the maximum number of iterations is reached.

Compared to random projections which has been widely used in the compressive sensing literature [17, 21, 22], the proposed compressive sampling is optimized in the information-theoretic framework based on the *a priori* knowledge of the user spatial distribution. Considering that the computational complexity of obtaining the approximate mutual information gradient is $\mathcal{O}(KMN^2)$, the computational complexity of the proposed compressive sampling optimization is $\mathcal{O}(JKMN^2)$, where J denotes the number of iterations.

4. User signal parameter estimation

In massive MIMO systems, the user signal parameters which can be estimated at the base station generally include DOA (i.e., relative position to the

[1] In order to avoid increasing the mutual information by scaling up the compressive sampling matrix, another feasible constraint is $\mathrm{Tr}(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}}) = M$, where $\mathrm{Tr}(\,\cdot\,)$ denotes the trace of a matrix.

array), power (i.e., received signal strength), and waveform (i.e, embedded information symbol).

With a given (either random or optimized) compressive sampling matrix $\mathbf{\Phi}$, the sample covariance matrix of the compressed measurements is computed as

$$\hat{\boldsymbol{R}}_{\boldsymbol{yy}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}(t)\boldsymbol{y}^{\mathrm{H}}(t) = \mathbf{\Phi}\hat{\boldsymbol{R}}_{\boldsymbol{xx}}\mathbf{\Phi}^{\mathrm{H}}, \tag{16}$$

where $T \geq 1$ denotes the number of snapshots, and

$$\hat{\boldsymbol{R}}_{\boldsymbol{xx}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t) \tag{17}$$

is the sample covariance matrix of the array received signals without any compression as shown in Fig. 1. Due to the time-varying nature of wireless channel, the calculation of sample covariance matrix is time-consuming. For reducing complexity and latency, it can be easily estimated using the training pilots only [36]. In addition to the sample covariance matrix, some structural information on the covariance matrix can be further exploited, which may improve the estimation performance of user signal parameters [37, 38, 39, 40]. For example, in [37], the asymptotic maximum likelihood estimate of a structured covariance matrix makes use of the extended invariance principle for parameter estimation. In [38], a maximum likelihood solution is derived for a structured covariance matrix to provide faster convergence.

4.1. DOA estimation

Because the compressive sampling matrix $\mathbf{\Phi}$ is row-orthonormal, the compressed noise vector $\mathbf{\Phi}\boldsymbol{n}(t)$ is still zero-mean white Gaussian, i.e., $\mathbf{\Phi}\boldsymbol{n}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\boldsymbol{I})$. Based on the minimum variance distortionless response (MVDR) criterion, we propose a compressive sampling MVDR spatial spectrum estimator as

$$P_{\text{CS-Capon}}(\theta) = \frac{1}{N} \frac{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}\hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\boldsymbol{a}(\theta)}, \tag{18}$$

which is also referred to as the CS-Capon spatial spectrum estimator. When there is no compression (i.e., $\boldsymbol{\Phi} = \boldsymbol{I}$), the proposed CS-Capon spatial spectrum estimator degenerates into the standard Capon spatial spectrum estimator

$$P_{\text{Capon}}(\theta) = \frac{1}{\boldsymbol{a}^{\mathrm{H}}(\theta)\hat{\boldsymbol{R}}_{\boldsymbol{xx}}^{-1}\boldsymbol{a}(\theta)} \tag{19}$$

because $\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{a}(\theta) = N$ for any direction θ. By searching the peaks of the proposed CS-Capon spatial spectrum $P_{\text{CS-Capon}}(\theta)$, the DOAs of user signals can be estimated. Other high-resolution DOA estimation methods, such as multiple signal classification (MUSIC) [41], ESPRIT [42], sparse signal reconstruction [43] and off-grid DOA estimation [27], can also be used.

By compressive sampling the array received signals, the high-dimensional sample covariance matrix $\hat{\boldsymbol{R}}_{\boldsymbol{xx}} \in \mathbb{C}^{N\times N}$ is compressed to a much lower-dimensional compressed measurement covariance matrix $\hat{\boldsymbol{R}}_{\boldsymbol{yy}} \in \mathbb{C}^{M\times M}$ because $M \ll N$. Hence, the computational complexity of the proposed CS-Capon spatial spectrum estimator in (18) is $\mathcal{O}(M^3)$, reduced from $\mathcal{O}(N^3)$, the computational complexity of the Capon spatial spectrum estimator in (19). Meanwhile, in order to avoid the ill-conditioned matrix inversion, the number of required snapshots for calculating the compressed measurement covariance matrix is $T \geq M$, which is much smaller than $T \geq N$ required for Nyquist sampling without any compression.

4.2. Power estimation

It is pointed out in [44] that there exists the peak underestimation problem of the Capon spatial power spectrum estimator, that is, the Capon spatial spectrum estimator grossly underestimates the power in the case of small number of snapshots or imprecise steering vector.

To avoid this problem and obtain an accurate power estimation, we make use of the estimated DOAs as the support information. With DOA estimates $\hat{\boldsymbol{\theta}} = \big[\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_D\big]$, the user signal power estimation can be formulated as a compressed measurement covariance matrix fitting problem as

$$\min_{\boldsymbol{P}(\hat{\boldsymbol{\theta}}),\sigma_n^2} \left\| \hat{\boldsymbol{R}}_{\boldsymbol{yy}} - \boldsymbol{\Phi}\boldsymbol{A}(\hat{\boldsymbol{\theta}})\boldsymbol{P}(\hat{\boldsymbol{\theta}})\boldsymbol{A}^{\mathrm{H}}(\hat{\boldsymbol{\theta}})\boldsymbol{\Phi}^{\mathrm{H}} - \sigma_n^2\boldsymbol{I} \right\|_F^2, \tag{20}$$

where the optimization variable $\boldsymbol{P}(\hat{\boldsymbol{\theta}}) = \mathrm{Diag}\big(\boldsymbol{p}(\hat{\boldsymbol{\theta}})\big)$ with the power $\boldsymbol{p}(\hat{\boldsymbol{\theta}}) \in \mathbb{R}_{++}^D$ over the DOA support $\hat{\boldsymbol{\theta}}$, and $\boldsymbol{\Phi}\boldsymbol{A}\big(\hat{\boldsymbol{\theta}}\big) = \big[\boldsymbol{\Phi}\boldsymbol{a}\big(\hat{\theta}_1\big), \boldsymbol{\Phi}\boldsymbol{a}\big(\hat{\theta}_2\big), \cdots, \boldsymbol{\Phi}\boldsymbol{a}\big(\hat{\theta}_D\big)\big] \in \mathbb{C}^{N\times D}$ is the corresponding compressed array manifold matrix. Here, $\|\cdot\|_F^2$ denotes the Frobenius norm of a matrix, and $\mathbb{R}_{++}^D$ denotes the set of positive real D-dimensional vectors. For the sake of simplicity , the noise power σ_n^2 can be estimated as the minimum eigenvalue of $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}$, i.e., $\hat{\sigma}_n^2$.

Note that $\boldsymbol{P}(\hat{\boldsymbol{\theta}}) = \mathrm{Diag}\big(\boldsymbol{p}(\hat{\boldsymbol{\theta}})\big)$ and $\boldsymbol{p}(\hat{\boldsymbol{\theta}}) = \mathrm{diag}\big(\boldsymbol{P}(\hat{\boldsymbol{\theta}})\big)$. Therefore, the compressed measurement covariance matrix fitting problem (20) is equivalent to

$$\min_{\boldsymbol{p}(\hat{\boldsymbol{\theta}})} \left\| \mathrm{vec}\left(\hat{\boldsymbol{R}}_{\boldsymbol{yy}} - \hat{\sigma}_n^2\boldsymbol{I}\right) - \left(\boldsymbol{\Phi}\boldsymbol{A}(\hat{\boldsymbol{\theta}}) \odot \boldsymbol{\Phi}\boldsymbol{A}(\hat{\boldsymbol{\theta}})\right)\boldsymbol{p}(\hat{\boldsymbol{\theta}}) \right\|_2^2, \tag{21}$$

where vec(·) denotes the vectorization operator, and $\odot$ stands for the Khatri-Rao product. It is clear that the closed-form power estimation is given by

$$\boldsymbol{p}(\hat{\boldsymbol{\theta}}) = \big[\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\big]^{-1}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{z}, \tag{22}$$

where matrix $\boldsymbol{G} = \boldsymbol{\Phi}\boldsymbol{A}(\hat{\boldsymbol{\theta}}) \odot \boldsymbol{\Phi}\boldsymbol{A}(\hat{\boldsymbol{\theta}}) = \Big[\mathrm{vec}\big(\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_1)\boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_1)\boldsymbol{\Phi}^{\mathrm{H}}\big), \mathrm{vec}\big(\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_2)\boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_2)\boldsymbol{\Phi}^{\mathrm{H}}\big), \cdots,$ $\mathrm{vec}\big(\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_D)\boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_D)\boldsymbol{\Phi}^{\mathrm{H}}\big)\Big] \in \mathbb{C}^{M^2 \times D}$ and vector $\boldsymbol{z} = \mathrm{vec}\big(\hat{\boldsymbol{R}}_{\boldsymbol{yy}} - \hat{\sigma}_n^2\boldsymbol{I}\big) \in \mathbb{C}^{M^2}$ are obtained by stacking the tensor outer products of the compressed signal steering vectors and the compressed measurement covariance matrix subtracted by the estimated noise covariance matrix, respectively.

The computational complexity of the power estimation in (22) is $\mathcal{O}\left(\max(D^2M^2, D^3)\right)$, which is much less than $\mathcal{O}\left(D^2N^2\right)$, the computational complexity of the power estimation without any compression.

4.3. Waveform estimation

In order to decode the embedded information, the mixed signals from different directions can be separated out from the compressed measurements through spatial filtering, namely, adaptive beamforming. According to the estimated signal DOA and power information, the adaptive beamformer based on the interference-plus-noise covariance matrix sparse reconstruction [34] is adopted, which provides a near-optimal output performance regardless of the input SNR. In detail, using the D-sparse spatial spectrum $\boldsymbol{p}(\hat{\boldsymbol{\theta}})$, the interference-plus-noise compressed covariance matrix for the d-th user signal with the estimated DOA $\hat{\theta}_d$ is reconstructed as

$$\hat{\boldsymbol{R}}_{\mathrm{CS},i+n,d} = \sum_{q=1,q\neq d}^{D} p(\hat{\theta}_q)\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_q)\boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_q)\boldsymbol{\Phi}^{\mathrm{H}} + \hat{\sigma}_n^2\boldsymbol{I},\ \forall\ d = \{1, \cdots, D\}, \tag{23}$$

where $\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_q)\boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_q)\boldsymbol{\Phi}^{\mathrm{H}}$ is the tensor outer product of the q-th interference compressed steering vector $\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_q)$. Using the MVDR criterion, the robust adaptive beamformer for the d-th user signal is given by

$$\boldsymbol{w}_d = \frac{\hat{\boldsymbol{R}}_{\mathrm{CS},i+n,d}^{-1}\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_d)}{\boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_d)\boldsymbol{\Phi}^{\mathrm{H}}\hat{\boldsymbol{R}}_{\mathrm{CS},i+n,d}^{-1}\boldsymbol{\Phi}\boldsymbol{a}(\hat{\theta}_d)},\ \forall\ d = \{1, \cdots, D\}, \tag{24}$$

whose output

$$\hat{s}_d(t) = \boldsymbol{w}_d^{\mathrm{H}} \boldsymbol{y}(t) \tag{25}$$

is the estimate of the signal waveform impinged from θ_d.

The computational complexity of computing the interference-plus-noise compressed covariance matrix reconstruction-based adaptive beamformer is $\mathcal{O}\left(\max(DMN, DM^2, M^3)\right)$, which is much less than $\mathcal{O}\left(N^3\right)$, the computational complexity of computing the interference-plus-noise covariance matrix reconstruction-based adaptive beamformer.

Under the compressive sampling framework, the output performance of the adaptive beamformer is evaluated by the compressive sampling output signal-to-interference-plus-noise ratio (SINR), defined as

$$\mathrm{SINR}_{\mathrm{CS}} = \frac{\sigma_{s_d}^2 |\boldsymbol{w}^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{a}(\theta_d)|^2}{\boldsymbol{w}^{\mathrm{H}} \boldsymbol{R}_{\mathrm{CS},i+n,d} \boldsymbol{w}}, \tag{26}$$

where $\sigma_{s_d}^2 = \mathbb{E}\left\{|s_d(t)|^2\right\}$ is the signal power of the d-th user, and

$$\begin{aligned} \boldsymbol{R}_{\mathrm{CS},i+n,d} &= \boldsymbol{\Phi} \mathbb{E}\left\{\boldsymbol{x}_{i+n,d}(t) \boldsymbol{x}_{i+n,d}^{\mathrm{H}}(t)\right\} \boldsymbol{\Phi}^{\mathrm{H}} \\ &= \boldsymbol{\Phi} \boldsymbol{R}_{i+n,d} \boldsymbol{\Phi}^{\mathrm{H}} \end{aligned} \tag{27}$$

is the corresponding interference-plus-noise compressed covariance matrix. Here, $\boldsymbol{x}_{i+n,d}(t) = \sum_{q=1,q\neq d}^{D} \boldsymbol{a}(\theta_q) s_q(t) + \boldsymbol{n}(t)$ is the array received interference-plus-noise term which excludes the desired signal $\boldsymbol{a}(\theta_d) s_d(t)$, i.e., the array received signal impinged from θ_d.

When the signal waveforms from different directions are separated out, the time of arrival (TOA) can be estimated via cross-correlation in terms of time delay for multipath signals that are correlated to each other, thus rendering further improvement of the localization accuracy of the users.

5. Simulation results

In this section, we compare the parameter estimation performance of the proposed compressive sampling with that of Gaussian random projections. We assume $N = 50$ omnidirectional sensors with a half-wavelength inter-element spacing in a massive MIMO system. The compression ratio is chosen to be $N/M = 5$, namely, the number of front-end chains is reduced to 10 from 50. Meanwhile, the dimension of the compressed measurement vector $\boldsymbol{y}(t)$ is reduced to $M = 10$ from $N = 50$. Without loss of generality, the DOAs of the users are assumed to follow a Gaussian distribution $\mathcal{N}(0^\circ, (5^\circ)^2)$. We uniformly discretize the pdf of the signal DOA with a width of $\Delta\bar{\theta} = 0.1^\circ$ for the compressive sampling matrix optimization regardless of the input SNR, rendering $K = 1,801$ components in the Gaussian mixture (5). In the process of sampling matrix optimization, a step size of $\gamma = 0.001$ is chosen for the gradient-based search.

In the first example, we compare the CS-Capon spatial spectra, computed from (18), between the optimized compressive sampling and the random compressive sampling. The results are shown in Fig. 3, where the number of snapshots is $T = 100$ and nine users impinging from $-8^\circ, -6^\circ, -4^\circ, -2^\circ, 0^\circ, 2^\circ, 4^\circ, 6^\circ$ and 8° have the same input SNR of 20 dB. As a reference, the standard Capon spatial spectrum obtained from (19) is also plotted. It is evident that, benefiting from the *a priori* knowledge of the user spatial distribution, the optimized compressive sampling can clearly identify the nine users as the Nyquist sampling, while the random compressive sampling does not provide sufficient resolution to identify all nine users.

In the second example, we consider one signal following the same dis-

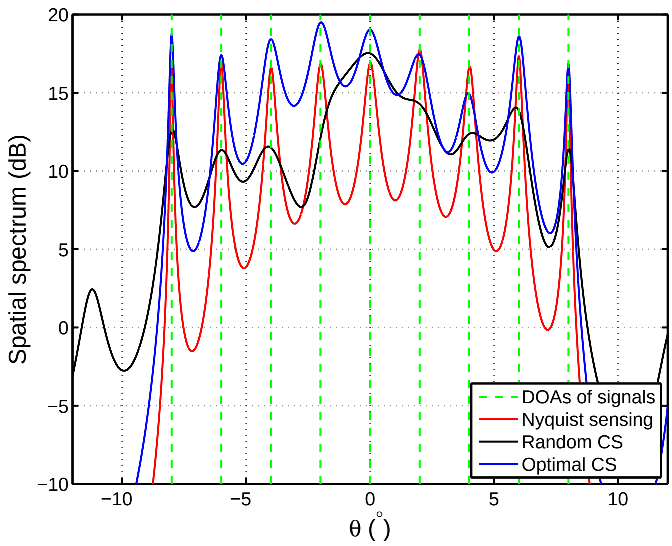


Figure 3: Capon spatial spectra comparison.

tribution as that for the compressive sampling optimization. In Fig. 4, we compare the root mean square error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{1}{N_{MC}} \sum_{l=1}^{N_{Mc}} \left(\hat{\theta}_l - \theta_l \right)^2}, \tag{28}$$

to evaluate the DOA estimation performance, and the mean absolute percentage error (MAPE), defined as

$$\text{MAPE} = \frac{1}{N_{MC}} \sum_{l=1}^{N_{Mc}} \frac{|\sigma_l^2 - \hat{\sigma}_l^2|}{\sigma_l^2}, \tag{29}$$

to evaluate the power estimation performance, where $\hat{\theta}_l$ and $\hat{\sigma}_l^2$ are respectively the estimate of DOA θ_l and the estimate of power σ_l^2, for the l-th Monte-Carlo trial. The number of snapshots is fixed to $T = 100$ in Figs. 4(a) and 4(b) when comparing the performance versus the input SNR, and the input SNR is fixed at 5 dB in Figs. 4(c) and 4(d) when comparing the per-

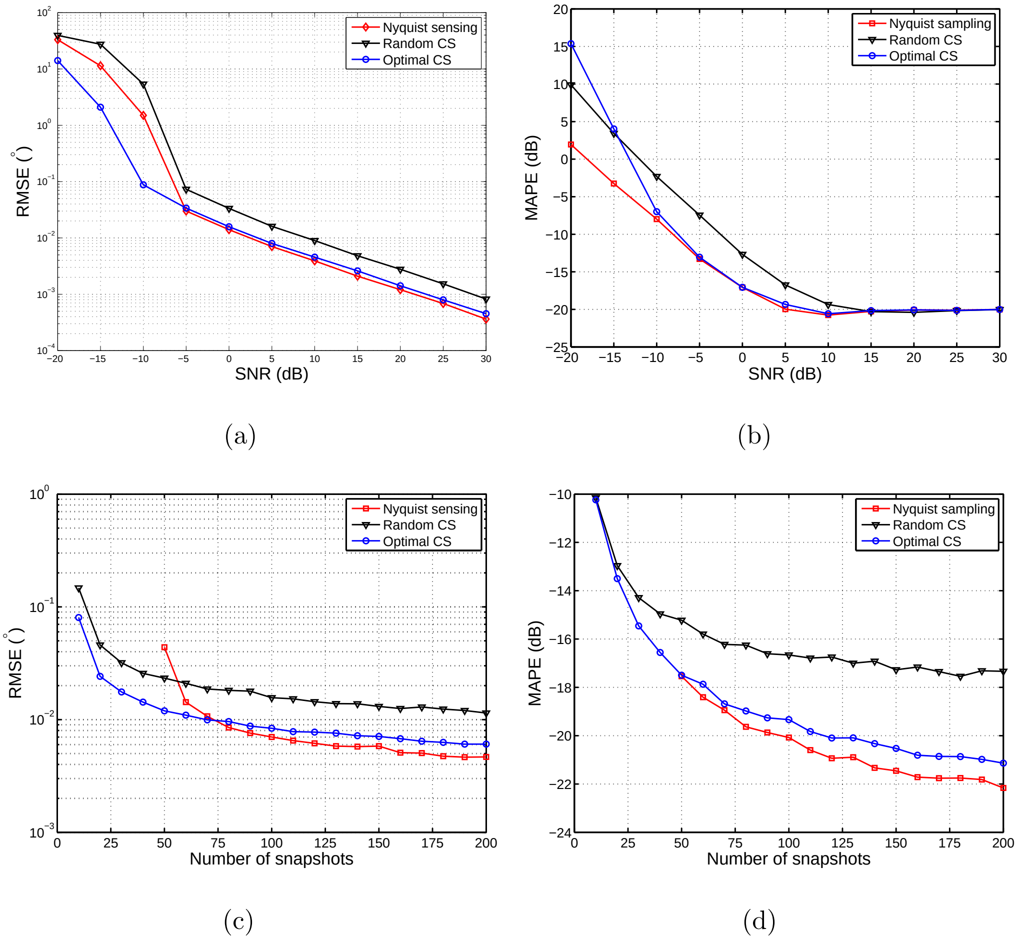


Figure 4: Performance comparison of different sampling schemes. (a) RMSE versus SNR; (b) MAPE versus SNR; (c) RMSE versus number of snapshots; (d) MAPE versus number of snapshots.

formance versus the number of snapshots. For each scenario, $N_{MC} = 1,000$ Monte-Carlo runs are performed.

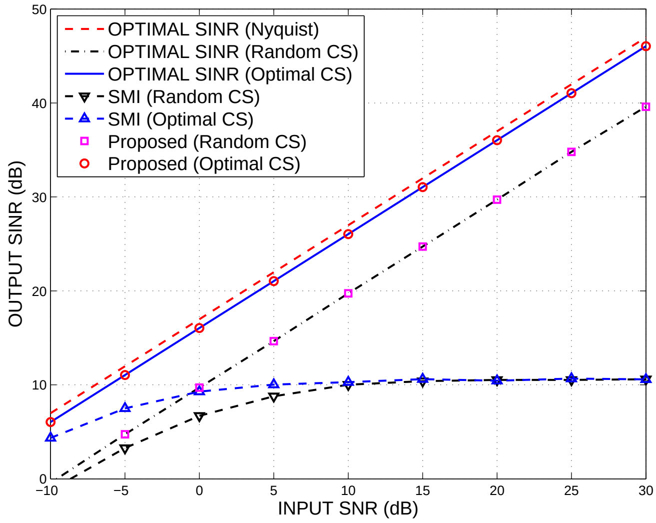
It is clear that the optimized compressive sampling outperforms random projections in both DOA and power estimations. For example, we observe from Fig. 4(a) that the optimized compressive sampling provides an SNR advantage of at least 5 dB compared to random projections in order to achieve the same DOA estimation accuracy. Furthermore, it also performs better than the Nyquist sampling either at the low SNR region, benefiting from the *a priori* knowledge, or with few snapshots, benefiting from the low-dimensional covariance matrix estimation. For the power estimation, the estimation accuracy can be further improved by increasing the number of snapshots as depicted in Fig. 4(d). On the other hand, when the number of snapshots is fixed, the power estimation accuracy is further improved with the increase of SNR when the SNR is lower than 10 dB, and the power estimation accuracy reaches a floor when the input SNR is higher than 10 dB. Compared to the CS-Capon spatial spectrum estimator presented in [1], the proposed compressed measurement covariance matrix fitting problem provides a better power estimation accuracy.

In the third example, we compare the waveform estimation performance in the terms of the output SINR versus the input SNR and the number of snapshots. The optimal SINR with Nyquist sampling defined as

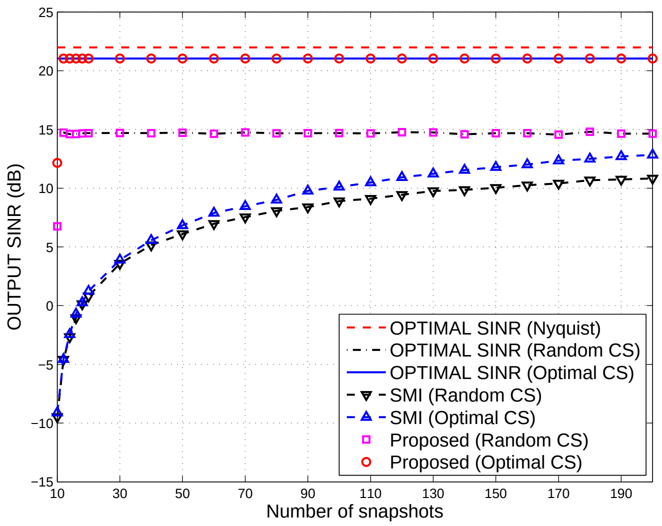
$$\max \ \mathrm{SINR} = \frac{\sigma_s^2 |\boldsymbol{w}^{\mathrm{H}} \boldsymbol{a}(\theta_d)|^2}{\boldsymbol{w}^{\mathrm{H}} \boldsymbol{R}_{i+n} \boldsymbol{w}}, \tag{30}$$

is also shown in all figures for the reference, where $\boldsymbol{R}_{i+n} = \mathbb{E}\left\{\boldsymbol{x}_{i+n}(t)\boldsymbol{x}_{i+n}^{\mathrm{H}}(t)\right\}$ is the interference-plus-noise covariance matrix corresponding to the d-th desired signal. In this example, the desired signal and the interferer are assumed to impinge from the known direction $\theta_s = 5^\circ$ and $\theta_i = -6^\circ$, respectively. The input interference-to-noise ratio (INR) in each sensor is equal to 10 dB. When comparing the output SINR versus the input SNR, the number of snapshots is fixed to $T = 100$. While comparing the output SINR versus the number of snapshots, the input SNR in each sensor is fixed at 5 dB. For each scenario, 1,000 Monte-Carlo runs are performed. The proposed compressive sampling matrix is assumed to be optimized when the input SNR is 10 dB.

From Fig. 5(a), it is clear that there is more than 7 dB performance loss for random projections compared with the Nyquist sampling, while the output performance of the optimized compressive sampling is very close to that of the Nyquist sampling. Compared with the classical sample matrix inversion (SMI) adaptive beamformer, the proposed interference-plus-noise compressed covariance matrix reconstruction-based adaptive beamformer can reach the optimal output performance regardless of the input SNR by excluding the desired signal component from the compressed covariance matrix for the adaptive beamformer design. Furthermore, as Fig. 5(b) shows, the proposed interference-plus-noise compressed covariance matrix reconstruction-based beamforming algorithm also has a faster convergence rate than the SMI beamforming algorithm, and it converges to the optimal output as long as the number of snapshots is larger than the number of compressed measurements. This is another attractive feature of the proposed technique to make quick response on the environment change.



(a)



(b)

Figure 5: Output SINR comparison of different beamforming algorithms with different compressive sampling schemes. (a) Output SINR versus input SNR; (b) output SINR versus number of snapshots.

6. Conclusion

In this paper, we proposed a compressive sampling optimization for user signal parameter estimation in massive MIMO systems. Different from random projections that are commonly used in the compressive sampling literatures, the proposed compressive sampling is optimized by exploiting the *a priori* knowledge of the probability of spatial distribution of the user signals to maximize the mutual information between the compressed measurements and the user DOAs. The optimized compressive sampling matrix is then used to estimate the DOAs by searching the proposed CS-Capon spatial spectrum, and the power by solving a compressed measurement covariance matrix fitting problem. The proposed robust adaptive beamformer separates each user signal waveform from the compressed measurements. In addition to achieving significant complexity reduction both in the number of front-end chains and in the computational complexity, simulation results demonstrate that the proposed compressive sampling techniques offer significant performance improvement compared with random projections. Furthermore, the proposed compressive sampling provides the better accuracy on DOA estimation than the standard Nyquist sampling particularly when the input SNR is low or when only few snapshots are available.

References

[1] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive MIMO," in *Proc. IEEE ICASSP*, New Orleans, LA, Mar. 2017, pp. 3181–3185.

[2] F. Rusek, D. Persson, Buon 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, Jan. 2013.

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

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] Y. Gu and Y. D. Zhang, “Pilot design for Gaussian mixture channel estimation in massive MIMO,” in *Proc. IEEE ICASSP*, Calgary, Canada, Apr. 2018, pp. 3266–3270.

[6] Y. Gu and Y. D. Zhang, “Information-theoretic pilot design for downlink channel estimation in FDD massive MIMO systems,” *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2334–2346, May 2019.

[7] E. Björnson, L. Sanguinetti, H. Wymeersch et al., “Massive MIMO is a reality - what is next? five promising research directions for antenna arrays,” *Digit. Signal Process.*, in press.

[8] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, “Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G,” *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 102–112, 2014.

[9] A. Guerra, F. Guidi, and D. Dardari, “Position and orientation error bound for wideband massive antenna arrays,” in *Proc. IEEE ICC*, London, UK, June 2015, pp. 853–858.

[10] V. Savic and E. G. Larsson, "Fingerprinting-based positioning in distributed massive MIMO systems," in *Proc. IEEE 82nd VTC Fall*, Boston, MA, Sept. 2015, pp. 1–5.

[11] F. Guidi, A. Guerra, and D. Dardari, "Millimeter-wave massive arrays for indoor SLAM," in *Proc. IEEE ICC*, Sydney, Australia, June 2014, pp. 114–120.

[12] F. Guidi, A. Guerra, and D. Dardari, "Personal mobile radars with millimeter-wave massive arrays for indoor mapping," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1536–1233, June 2016.

[13] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[14] T. Lv, F. Tan, H. Gao, and S. Yang, "A beamspace approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *Signal Process.*, vol. 121, pp. 30–45, Apr. 2016.

[15] L. Huang, Y. Xiao, K. Liu, H. C. So, and J.-K. Zhang, "Bayesian information criterion for source enumeration in large-scale adaptive antenna array," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3018–3032, May 2016.

[16] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, M. Coulon, "Direct localization for massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2475–2487, May 2017.

[17] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[18] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[19] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, Apr. 2007.

[20] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[21] V. Cevher, A. Gurbuz, J. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation," in *Proc. IEEE ICASSP*, Las Vegas, NV, Mar./Apr. 2008, pp. 2497–2500.

[22] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *Proc. IEEE SSP*, Cardiff, UK, Aug. 2009, pp. 626–629.

[23] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 3967–3971.

[24] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 419–430, Jan. 2014.

[25] X. Wu, W.-P. Zhu, and J. Yan, "A Toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8223–8237, Sept. 2017.

[26] C. Zhou, Y. Gu, Y. D. Zhang, Z. Shi, T. Jin, and X. Wu, "Compressive sensing-based coprime array direction-of-arrival estimation," *IET Commun.*, vol. 11, no. 11, pp. 1719–1724, Nov. 2017.

[27] X. Wu, W.-P. Zhu, J. Yan, and Z. Zhang, "Two sparse-based methods for off-grid direction-of-arrival estimation," *Signal Process.*, vol. 142, pp. 87–95, Jan. 2018.

[28] A. Gürbüz, J. McClellan, and V. Cevher, "A compressive beamforming method," in *Proc. IEEE ICASSP*, Las Vegas, NV, Mar./Apr. 2008.

[29] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4643–4657, Sept. 2012.

[30] A. Xenakia, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," *J. Acoust. Soc. Amer.*, vol. 136, no. 1, pp. 260–271, July 2014.

[31] Y. Gu, C. Zhou, N. A. Goodman, W.-Z. Song, and Z. Shi, "Coprime array adaptive beamforming based on compressive sensing virtual array signal," in *Proc. IEEE ICASSP*, Shanghai, China, March 2016, pp. 2981–2985.

[32] Y. Gu, N. A. Goodman, and A. Ashok, "Radar target profiling and recognition based on TSI-optimized compressive sensing kernel," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3194–3207, June 2014.

[33] Y. Gu and N. A. Goodman, "Information-theoretic compressive sensing kernel optimization and Bayesian Cramér-Rao bound for time delay estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4525–4537, Sept. 2017.

[34] Y. Gu, N. A. Goodman, S. Hong, and Y. Li, "Robust adaptive beamforming based on interference covariance matrix sparse reconstruction," *Signal Process.*, vol. 96, Part B, pp. 375–381, Mar. 2014.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory (2nd Edition)*. New York, NY: John Wiley, 2006.

[36] H. Lin, F. Gao, S. Jin, and Geoffrey Y. Li, "A new view of multi-user hybrid massive MIMO: Non-orthogonal angle division multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2268–2280, Oct. 2017.

[37] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, May 1999.

[38] M. Steiner and K. Gerlach, "Fast converging adaptive processor for a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, Oct. 2000.

[39] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Maximum likelihood estimation of a structured covariance matrix with a condition number constraint," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, June 2012.

[40] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6182–6189, Dec. 2012.

[41] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[42] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, July 1989.

[43] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[44] P. Stoica, J. Li, and X. Tan, "On spatial power spectrum and signal estimation using the Pisarenko framework," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5109–5119, Oct. 2008.

References

]Yujie Gu received the Ph.D. degree in electronic engineering from Zhejiang University, Hangzhou, China, in 2008. After graduation, he held multiple research positions in China, Canada, Israel, and the USA. He is currently a Postdoctoral Research Associate with Temple University, Philadelphia, PA, USA. His research interests are in statistical and array signal processing including adaptive beamforming, compressive sensing, MIMO systems, radar imaging, target localization, waveform design, etc. Dr. Gu is currently a Subject Editor for *Electronics Letters*, an Associate Editor for *Signal Processing*, *IET Signal Processing*, *Circuits, Systems and*

Signal Processing, and *EURASIP Journal on Advances in Signal Processing*. He is also the Lead Guest Editor of special issue on Source Localization in Massive MIMO for the *Digital Signal Processing*. He has been a member of the Technical Program Committee of multiple international conferences including IEEE ICASSP. He is also a Special Sessions Co-Chair of the 2020 IEEE Sensor Array and Multichannel Signal Processing Workshop. Dr. Gu is an IEEE Senior Member.

References

]Yimin D. Zhang received the Ph.D. degree from the University of Tsukuba, Tsukuba, Japan, in 1988. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, College of Engineering, Temple University, Philadelphia, PA, USA. From 1998 to 2015, he was a Research Faculty with the Center for Advanced Communications, Villanova University, Villanova, PA, USA. His research interests lie in the areas of statistical signal and array processing, including compressive sensing, machine learning, convex optimization, time-frequency analysis, MIMO systems, radar imaging, direction finding, target localization and tracking, wireless and cooperative networks, and jammer suppression, with applications to radar, wireless communications, and satellite navigation. He has authored/coauthored more than 340 journal and conference papers and 14 book chapters in these areas. Dr. Zhang is an Associate Editor for the *IEEE Transactions on Signal Processing* and an Editor for the *Signal Processing* journal. He was an Associate Editor for the *IEEE Signal Processing Letters* during 2006–2010, and an Associate Editor for the *Journal of the Franklin Institute* during 2007–2013. He is a member of the Sensor Array and Multichannel Technical Committee and the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society. He was the Technical Co-Chair of the 2018 IEEE Sensor Array and Multichannel Signal Processing Workshop. He was the recipient of the 2016 IET Radar, Sonar & Navigation Premium Award and the 2017 IEEE Aerospace and Electronic Systems Society Harry Rowe Mimno Award, and coauthored a paper that received the 2018 IEEE Signal Processing Society Young Author Best Paper Award. Dr. Zhang is an IEEE Fellow.