OPTIMIZED COMPRESSIVE SENSING-BASED DIRECTION-OF-ARRIVAL ESTIMATION IN MASSIVE MIMO

Yujie Gu[1], *Yimin D. Zhang*[1], *and Nathan A. Goodman*[2]

[1] Department of Electrical & Computer Engineering, Temple University, Philadelphia, PA 19122, USA
[2] School of Electrical & Computer Engineering, The University of Oklahoma, Norman, OK 73019, USA

ABSTRACT

As a new emerging technology for wireless communications, massive multiple-input multiple-output (MIMO) faces a significant challenge to deploy a separate receiver chain of front-end circuits in a dense circuit board. In this paper, we apply the compressive sensing technique to reduce the required number of front-end circuits and the overall computational complexity. Unlike the commonly adopted random projections, we utilize the *a priori* probability distribution of the directions-of-arrival (DOAs) of the signals to optimize compressive sensing kernels for massive MIMO systems, such that the mutual information between the compressed measurement and the DOA is maximized. With the optimized sensing matrix, we present a compressive sensing spatial spectrum estimator under the minimum variance distortionless response criterion. Simulation results demonstrate performance advantages of the proposed optimal sensing kernel over random sensing kernels.

Index Terms— Compressive sensing, DOA estimation, kernel optimization, massive MIMO, spatial spectrum.

1. INTRODUCTION

Direction-of-arrival (DOA) estimation using an array of multiple sensors is a fundamental problem in a variety of fields, such as radar, sonar, wireless communications, acoustics, seismology, and radio astronomy [1,2]. Recently, with the development of millimeter wave technology, massive multiple-input multiple-output (MIMO) equipped with a large number of sensors has been a hot research topic for next generation wireless communications [3–8]. Although a massive MIMO system does provide a higher resolution because of its larger array aperture, it is a great challenge to equip a large number of front-end circuit chains. Furthermore, the computational complexity is also a major concern because of the high-dimensional matrix operation (e.g., inversion or eigen-decomposition). In order to reduce the complexity, compressive sensing is a viable solution for massive MIMO systems.

In the past decade, the compressive sensing technique has been widely applied in a variety of signal processing applications, such as DOA estimation [10–12], beamforming [13–16], and radar imaging [17, 18]. It is well known that random sensing matrix, such as Gaussian or Bernoulli matrix, satisfies the incoherence property with any fixed signal representation basis, which guarantees the reliable signal recovery. Obviously, such random matrices do not exploit the potential prior knowledge of the signals beyond the sparsity. In practical massive MIMO applications, prior knowledge of DOAs of the signals of interest is usually available.

In this paper, we exploit the available *a priori* information to optimize the compressive sensing kernel for DOA estimation in a massive MIMO system. More specifically, the compressed measurements are characterized by a Gaussian mixture model through discretizing the *a priori* probability distribution of the DOAs. Based on this observation, the compressive sensing matrix can be optimized by maximizing the mutual information between the compressed measurements and the DOAs to be estimated. With the optimized sensing matrix, we present a compressive sensing minimum variance distortionless response (CS-MVDR) spatial spectrum estimator, from which both the DOAs and the power of the signals can be estimated. Simulation results demonstrate that the DOA estimation accuracy is improved by using the optimized sensing matrix. Furthermore, the CS-MVDR spatial spectrum estimator with the optimized sensing matrix provides more accurate power estimation than the standard MVDR spatial spectrum estimator.

2. SIGNAL MODEL

Consider D far-field uncorrelated narrowband signal sources impinging on an N-element massive array from directions $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_D]^{\mathrm{T}}$, where $(\cdot)^{\mathrm{T}}$ denotes the transpose operation. The array received signal at the kth sampling time can be modeled in complex baseband as

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

The work of Y. Gu and Y. D. Zhang is supported in part by the National Science Foundation (NSF) under grant AST-1547420. The work of Y. Gu and N. A. Goodman is supported in part by the Defense Advanced Research Projects Agency (DARPA) via grant #N66001-10-1-4079.

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 steering matrix with the dth column representing the dth source's steering vector $\boldsymbol{a}(\theta_d) \in \mathbb{C}^N$, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_D(t)]^{\mathrm{T}} \in \mathbb{C}^D$ denotes the signal waveform vector, and $\boldsymbol{n}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma_{\boldsymbol{n}}^2 \mathbf{I}_N)$ denotes the zero-mean additive white Gaussian noise vector. Here, $\mathbf{I}_N$ is the $N \times N$ identity matrix.

Let compressive sensing matrix $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1^{\mathrm{T}}, \boldsymbol{\phi}_2^{\mathrm{T}}, \cdots, \boldsymbol{\phi}_M^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M\times N}$ $(M \ll N)$ consist of M row-orthonormal sensing kernels, i.e., $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} = \mathbf{I}_M$, where $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose. When $\boldsymbol{\Phi}$ is applied to the massive array, its received signal vector $\boldsymbol{x}(t) \in \mathbb{C}^N$ can be compressed into an M-dimensional measurement vector

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}\boldsymbol{x}(t) = \boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{\Phi}\boldsymbol{n}(t). \tag{2}$$

Hence, $\boldsymbol{\Phi}\boldsymbol{a}(\theta) \in \mathbb{C}^M$ can be regarded as a *sketch* of the massive array steering vector with a significantly reduced size of dimension. Our objective is to optimize the compressive sensing matrix $\boldsymbol{\Phi}$ based on the prior knowledge of DOA distribution to obtain DOA estimates with an improved accuracy from the compressed measurement vector $\boldsymbol{y}(t)$.

3. PROBABILISTIC OPTIMIZATION OF COMPRESSIVE SENSING MATRIX

3.1. Probabilistic signal model

In this section, we optimize the compressive sensing matrix using a probabilistic model. The DOA θ is treated as a random variable that occupies a region Θ with a probability density function (pdf) $f(\theta)$ which is known *a priori*. Such distribution can be, e.g., estimated from the previous observed data, determined by the mission of operation, or decided based on the previous states of the sources.

With the law of total probability, the pdf of the measurement vector can be expressed as

$$f(\boldsymbol{y}) = \mathrm{E}_\theta\left\{f(\boldsymbol{y}|\theta)\right\} = \int_{\theta\in\Theta} f(\boldsymbol{y}|\theta) f(\theta)\, d\theta, \tag{3}$$

where $\mathrm{E}_\theta\{\cdot\}$ denotes the statistical expectation with respect to θ. Discretize the pdf $f(\theta)$ into K angular bins with an equal width of $\Delta\bar{\theta}$, and denote $p_k = f(\bar{\theta}_k)\Delta\bar{\theta}$ with $\sum_{k\in\mathcal{K}} p_k = 1$, where $\mathcal{K} = \{1, 2, \cdots, K\}$. Then, the above expression can be approximated as

$$f(\boldsymbol{y}) \approx \sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}|\bar{\theta}_k). \tag{4}$$

Consider the kth angular bin with a nominal DOA $\bar{\theta}_k$. The signal arrival in this angular bin, $s(t)$, is considered as the mixture of different sources and is modeled as a zero-mean complex Gaussian random variable, i.e., $s(t) \sim \mathcal{CN}(0, \sigma_s^2)$. In this case, the corresponding compressed measurement vector is

$$\boldsymbol{y}|_{\theta=\bar{\theta}_k} = \boldsymbol{\Phi}\big(\boldsymbol{a}(\bar{\theta}_k)s(t) + \boldsymbol{n}(t)\big), \tag{5}$$

and the conditional pdf is

$$f(\boldsymbol{y}|\bar{\theta}_k) = \frac{1}{\pi^M \left|\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}\right|} e^{-\boldsymbol{y}^{\mathrm{H}}\mathbf{C}^{-1}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}\boldsymbol{y}}, \tag{6}$$

where $\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} = \boldsymbol{\Phi}\big(\sigma_s^2 \boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \sigma_{\boldsymbol{n}}^2\mathbf{I}_N\big)\boldsymbol{\Phi}^{\mathrm{H}}$ is the covariance matrix of $\boldsymbol{y}|_{\theta=\bar{\theta}_k}$. Hence, the pdf of the compressed measurement vector is a weighted sum of K Gaussian distributions, which forms a Gaussian mixture distribution. If the bin width $\Delta\bar{\theta}$ is chosen to be sufficiently small, then the approximation in (4) approaches to equality at the cost of infinite components in the mixture.

3.2. Compressive sensing matrix optimization

We adopt the maximum mutual information criterion [19] to optimize the compressive sensing matrix $\boldsymbol{\Phi}$ such that a massive array can estimate the DOAs of the sources more accurately. Considering that the optimization variables are represented as a high-dimensional fat matrix, we prefer a gradient-based strategy as in [20] to search for the optimal sensing matrix. This procedure requires the computation of the gradient of the Shannon mutual information $I(\boldsymbol{y};\theta)$ between the compressed measurement vector $\boldsymbol{y}$ and the DOA θ with respect to the sensing matrix $\boldsymbol{\Phi}$, described as

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\theta) = \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}) - \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}|\theta), \tag{7}$$

where $\nabla_{\boldsymbol{\Phi}}\{\cdot\}$ stands for the gradient operator with respective to $\boldsymbol{\Phi}$, $h(\boldsymbol{y}) = -\mathrm{E}_{\boldsymbol{y}}\big\{\log[f(\boldsymbol{y})]\big\}$ denotes the differential entropy of the compressed measurement vector $\boldsymbol{y}$, and $h(\boldsymbol{y}|\theta) = -\mathrm{E}_{\boldsymbol{y},\theta}\big\{\log[f(\boldsymbol{y}|\theta)]\big\}$ represents the conditional differential entropy of the compressed measurement vector $\boldsymbol{y}$ given the DOA θ. Although there is no closed-form information gradient expression for practical array applications, the discretization of the pdf of DOAs makes it possible to obtain an approximate gradient.

The conditional pdf of the mean of the compressed measurement vector $\boldsymbol{y}$, given DOA θ, is given by

$$f(\boldsymbol{y}_0 = \mathbf{0}|\theta) = \frac{1}{\pi^M \left|\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\theta}\right|}, \tag{8}$$

where $\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\theta} = \boldsymbol{\Phi}\big(\sigma_s^2\boldsymbol{a}(\theta)\boldsymbol{a}^{\mathrm{H}}(\theta) + \sigma_{\boldsymbol{n}}^2\mathbf{I}_N\big)\boldsymbol{\Phi}^{\mathrm{H}}$. Hence, we can compute the approximate mutual information as:

$$\begin{aligned} I(\boldsymbol{y};\theta) &\approx h(\boldsymbol{y}) + \iint f(\boldsymbol{y},\theta)\log f(\boldsymbol{y}_0|\theta)\, d\boldsymbol{y}d\theta \\ &= h(\boldsymbol{y}) + \int \log f(\boldsymbol{y}_0|\theta) f(\theta)\, d\theta \\ &= h(\boldsymbol{y}) - \int \log\left|\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\theta}\right| f(\theta) d\theta - M\log\pi \\ &\approx -\log\sum_{k\in\mathcal{K}} p_k \left|\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}\right|^{-1} - \sum_{k\in\mathcal{K}} p_k \log\left|\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}\right|, \end{aligned} \tag{9}$$

where the first approximation of $h(\boldsymbol{y}|\theta)$ is based on a first-order Taylor series expansion of $\log f(\boldsymbol{y}|\theta)$ around the mean value $\boldsymbol{y}_0 = \mathbf{0}$, and the last approximation of the differential entropy $h(\boldsymbol{y})$ can be found in [20].

By taking the gradient of the resulting approximated mutual information $I(\boldsymbol{y};\theta)$ in (9) with respect to the sensing matrix $\boldsymbol{\Phi}$, we have

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\theta) \approx \frac{\sum_{k\in\mathcal{K}} p_k \left|\frac{\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}}{\sigma_{\boldsymbol{n}}^2}\right|^{-1} \left[\frac{\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}}{\sigma_{\boldsymbol{n}}^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_s^2}{\sigma_{\boldsymbol{n}}^2}\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k)+\mathbf{I}_N\right)}{\sum_{k\in\mathcal{K}} p_k \left|\frac{\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}}{\sigma_{\boldsymbol{n}}^2}\right|^{-1}} - \sum_{k\in\mathcal{K}} p_k \left[\frac{\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}}{\sigma_{\boldsymbol{n}}^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_s^2}{\sigma_{\boldsymbol{n}}^2}\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k)+\mathbf{I}_N\right), \quad (10)$$

where $\sigma_s^2/\sigma_{\boldsymbol{n}}^2$ denotes the input signal-to-noise ratio (SNR) of the signal, and $\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}/\sigma_{\boldsymbol{n}}^2 = \boldsymbol{\Phi}\left((\sigma_s^2/\sigma_{\boldsymbol{n}}^2)\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k)+\mathbf{I}_N\right)\boldsymbol{\Phi}^{\mathrm{H}}$.

Using the approximate information gradient $\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\theta)$ (10), we can iteratively search for the optimal compressive sensing matrix from

$$\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} + \gamma \nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\theta), \quad (11)$$

where $\gamma > 0$ is a step size. This procedure can be iterated by re-orthonormalizing the rows of $\tilde{\boldsymbol{\Phi}}$ and computing the information gradient $\nabla_{\tilde{\boldsymbol{\Phi}}} I(\boldsymbol{y};\theta)$. In contrast to randomly generated compressive sensing kernels as commonly used in the compressive sensing literature [9–11], the compressive sensing kernel in the proposed framework can be optimized based on the *a priori* knowledge of the distribution of the DOAs. The detailed optimization process and convergence criterion can be found in [20–23]. Because the complexity of computing the information gradient in (10) is $\mathcal{O}(KMN^2)$, the overall computational complexity of the proposed compressive sensing matrix optimization is $\mathcal{O}(JKMN^2)$, where J denotes the number of iterations.

4. CS-MVDR SPATIAL SPECTRUM ESTIMATOR

Because the compressive sensing matrix $\boldsymbol{\Phi}$ is row-orthonormal, the compressed additive noise vector $\boldsymbol{\Phi}\boldsymbol{n}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma_{\boldsymbol{n}}^2\mathbf{I}_M)$ is still white Gaussian. Based on the compressed measurements of (2), we propose the following CS-MVDR spatial spectrum estimator:

$$P_{\text{CS-MVDR}}(\theta) = \frac{1}{N}\frac{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}\hat{\mathbf{R}}_{\boldsymbol{y}\boldsymbol{y}}^{-1}\boldsymbol{\Phi}\boldsymbol{a}(\theta)}, \quad (12)$$

where $\hat{\mathbf{R}}_{\boldsymbol{y}\boldsymbol{y}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}(t)\boldsymbol{y}^{\mathrm{H}}(t) = \boldsymbol{\Phi}\hat{\mathbf{R}}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\Phi}^{\mathrm{H}}$ is the compressed measurement sample covariance matrix. Here, $T \geq 1$ is the number of snapshots, and $\hat{\mathbf{R}}_{\boldsymbol{x}\boldsymbol{x}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)$ is the massive array sample covariance matrix. It is worth mentioning that some structural information on the covariance matrix can be exploited to enhance the estimation accuracy, which may achieve better final performance [24–27]. When $\boldsymbol{\Phi} = \mathbf{I}_N$, namely, Nyquist sampling is applied, the above CS-MVDR spatial spectrum estimator degenerates to the standard MVDR spatial spectrum estimator

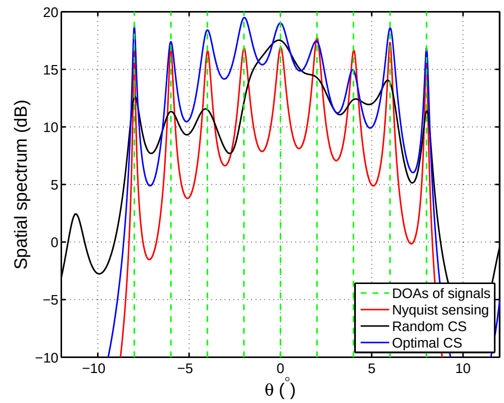


Fig. 1. CS-MVDR spatial spectra comparison.

$$P_{\text{MVDR}}(\theta) = \frac{1}{\boldsymbol{a}^{\mathrm{H}}(\theta)\hat{\mathbf{R}}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{a}(\theta)}, \quad (13)$$

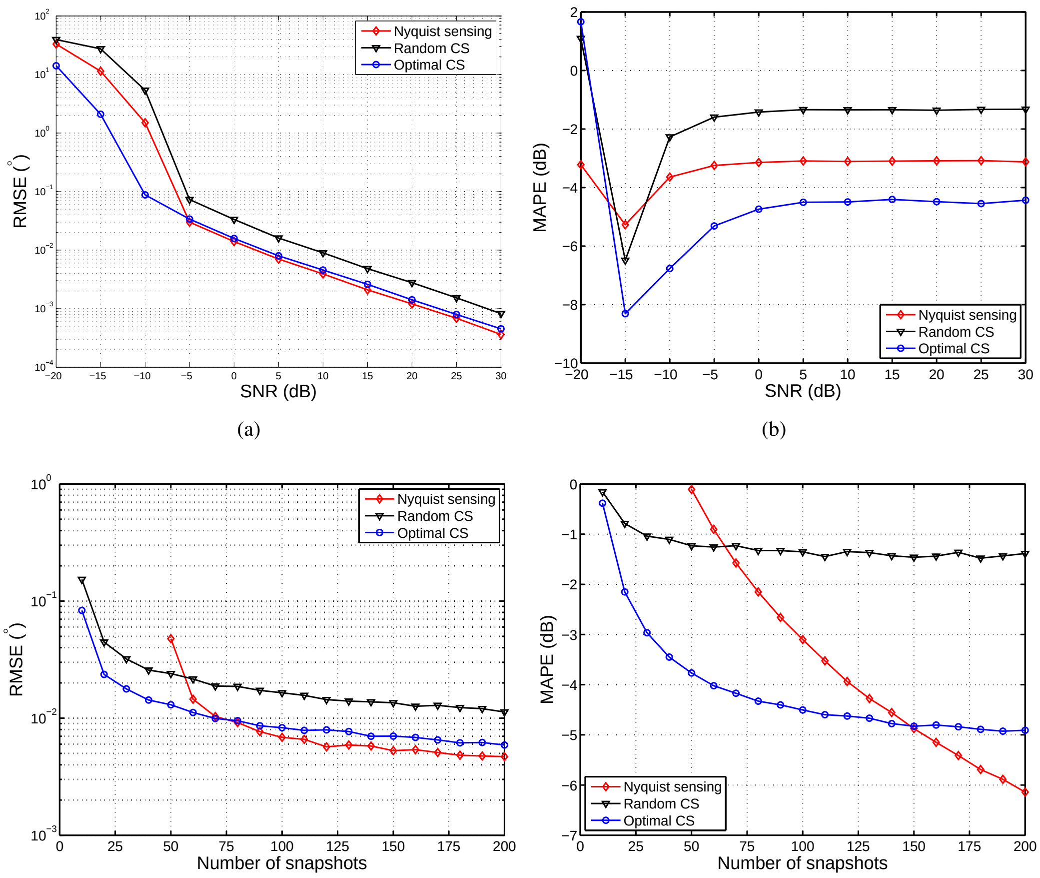
because $\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{a}(\theta) = N$ for any direction θ.

Benefiting from compressive sampling, the original high-dimensional sample covariance matrix $\hat{\mathbf{R}}_{\boldsymbol{x}\boldsymbol{x}} \in \mathbb{C}^{N\times N}$ is reduced to a much lower-dimensional measurement covariance matrix $\hat{\mathbf{R}}_{\boldsymbol{y}\boldsymbol{y}} \in \mathbb{C}^{M\times M}$ because $M \ll N$. Hence, the corresponding computational complexity of the MVDR spatial spectrum estimator is reduced to $\mathcal{O}(M^3)$ from the original $\mathcal{O}(N^3)$. Meanwhile, to avoid the ill-conditioned inversion, the required number of snapshots for compressive sampling is reduced to $T \geq M$ from that for Nyquist sampling, which is $T \geq N$.

5. SIMULATION RESULTS

We assume a massive array with $N = 50$ omnidirectional sensors and half-wavelength inter-element spacing. The compression ratio is chosen to be $N/M = 5$, namely, the dimension of the compressed measurement vector $\boldsymbol{y}(t)$ is $M = 10$. Without loss of generality, the DOAs of the sources for compressive sensing matrix optimization are assumed to follow a Gaussian distribution $\mathcal{N}(0°, (5°)^2)$. We uniformly discretize the pdf of DOA with a width of $\Delta\bar{\theta} = 0.1°$ for sensing matrix optimization regardless of the SNR. Namely, there are $K = 1,801$ components in the Gaussian mixture (4). In the process of sensing matrix optimization, a step size of $\gamma = 0.001$ is chosen for the gradient-based search.

In the first example, we compare the CS-MVDR spatial spectra in Fig. 1, where the number of snapshots is $T = 100$



(a) (b) (c) (d)

Fig. 2. Performance comparison of different sensing kernels. (a) RMSE versus SNR; (b) MAPE versus SNR; (c) RMSE versus number of snapshots; (d) MAPE versus number of snapshots.

and nine sources have the same SNR of 20 dB. It is clear that the optimized sensing kernel can clearly identify the nine sources as the Nyquist sampling, while the random sensing kernels do not provide sufficient resolution to identify all nine sources.

In the next example, we assume that there is one signal following the same distribution as that for sensing matrix optimization. In Fig. 2, we compare the root mean square error (RMSE) of the DOA estimation, and the mean absolute percentage error (MAPE) of the power estimation. The number of snapshots is fixed to $T = 100$ in Figs. 2(a) and 2(b) when comparing the performance versus SNR, and the SNR is fixed at 5 dB in Figs. 2(c) and 2(d) when comparing the performance versus the number of snapshots. For each scenario, 1,000 Monte-Carlo runs are performed. It is clear that the optimized sensing kernel consistantly outperforms the random sensing kernels in both DOA and power estimations. For example, we can see from Fig. 2(a) that the optimized kernel provides an SNR advantage of at least 5 dB compared to random kernels in order to achieve the same DOA estimation accuracy. Furthermore, it also performs better than the Nyquist sampling either at low SNRs, benefiting from the *a priori* knowledge, or with few snapshots, benefiting from the low-dimensional covariance matrix estimation.

6. CONCLUSION

In this paper, we have considered the compressive sensing matrix optimization for DOA estimation in a massive MIMO system. Different from the commonly used random projections, the proposed compressive sensing kernel is optimized by utilizing the prior knowledge of the probabilities of signal DOAs to maximize the Shannon mutual information between the compressed measurements and the DOAs. The optimized sensing matrix is then used to estimate the DOAs by searching the CS-MVDR spatial spectrum. In addition to achieving significant complexity reduction, simulation results also demonstrate that the optimized sensing kernel offers significant performance improvement compared with random sensing kernels. Furthermore, the optimized compressive sensing kernel can obtain better estimation accuracy than the standard Nyquist sampling at low SNRs or with few snapshots.

7. REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part IV: Optimum Array Processing*. New York, NY: John Wiley, 2002.

[2] Z. Shi, C. Zhou, Y. Gu, N. A. Goodman, and F. Qu, "Source estimation using coprime array: A sparse reconstruction perspective," to appear in *IEEE Sensors Journal*.

[3] R. C. de Lamare, "Massive MIMO systems: Signal processing challenges and research trends," in arXiv:i3i0.7282, Oct. 2013.

[4] 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, Jan. 2013.

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

[6] 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.

[7] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[8] F. Jiang, J. Chen, A. L. Swindlehurst, and J. A. Lṕez-Salcedo, "Massive MIMO for wireless sensing with a coherent multiple access channel," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3005–3017, June 2015.

[9] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[11] V. Cevher, A. Gurbuz, J. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation," in *Proc. IEEE ICASSP*, Las Vegas, NV, March–April 2008.

[12] 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.

[13] A. Gürbüz, J. McClellan, and V. Cevher, "A compressive beamforming method," in *Proc. IEEE ICASSP*, Las Vegas, NV, March–April 2008.

[14] 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.

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

[16] 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.

[17] L. C. Potter, E. Ertin, J. T. Parker, and M. Çetin, "Sparsity and compressed sensing in radar imaging," *Proc. IEEE*, vol. 98, no. 6, pp. 1006–1020, June 2010.

[18] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "High-resolution passive SAR imaging exploiting structured Bayesian compressive sensing," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 8, pp. 1484–1497, Dec. 2015.

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

[20] 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.

[21] Y. Gu and N. A. Goodman, "Compressed sensing kernel design for radar range profiling," in *Proc. IEEE Radar Conf.*, Ottawa, Canada, April–May 2013.

[22] Y. Gu and N. A. Goodman, "Compressive sensing kernel optimization for time delay estimation," in *Proc. IEEE Radar Conf.*, Cincinnati, OH, May 2014.

[23] Y. Gu and N. A. Goodman, "Time domain CS kernel design for mitigation of wall reflections in urban radar," in *Proc. IEEE SAM Workshop*, A Coruña, Spain, June 2014.

[24] 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.

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

[26] 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.

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