Iterative Learning for Optimized Compressive Measurements in Massive MIMO Systems

Yimin D. Zhang
Department of Electrical and Computer Engineering
Temple University, Philadelphia, PA 19122, USA

***Abstract*—Massive multiple-input multiple-output (MIMO) is a key enabling technology for next-generation wireless communications and future radar systems. The high number of antennas in a massive MIMO system makes it difficult to process array signals with separate radio frequency chains. We have previously developed a compressive measurement technique that substantially reduces the required number of front-end circuits and the overall processing complexity. By utilizing the *a priori* distribution of signal directions, the compressive measurement matrix is optimized by maximizing the mutual information between the compressive measurement output and the signal directions-of-arrival (DOAs). In practice, however, the prior distribution of the signal DOAs may not be available, thereby limiting the applications of this technique. In this paper, we develop an iterative learning scheme that uses the estimated DOA spectrum as the prior for the next iteration. By removing the hurdle of requiring the prior information of signal arrivals, the proposed technique makes optimized compressive measurement feasible for broader applications in practice.**

Keywords: Massive MIMO, compressive measurement, DOA estimation, mutual information, iterative learning.

I. INTRODUCTION

With the rapid development of wireless communications, particularly millimeter wave communication systems, massive multiple-input multiple-output (MIMO) equipped with a high number of antennas has become a key enabling technology for 5G communications and beyond [1–6]. On the other hand, while large-size arrays are commonly used in radar in and out of the MIMO radar context [7], the success of massive MIMO in wireless communications has also inspired massive MIMO radar [8, 9]. Massive MIMO technology is also exploited in the development of large intelligent surface (LIS)-based communications and sensing [11, 12].

Despite the excellent processing performance of massive MIMO systems, one of the significant problems that massive MIMO communication and radar systems commonly face is the high complexity that increases with the number of antennas. In particular, processing each antenna with a separate radio frequency (RF) chain is often impractical. One of the effective approaches is to reduce the high dimension of the array measurements into a lower dimension by performing analog-domain compressive sampling so that a much fewer RF front-end circuits and analog-to-digital converters can be used. While such compressive sampling can be performed using different ways, commonly used random sensing matrices [13–15] suffer from significant information loss [16].

In [17, 18], we have proposed a novel hybrid analog-digital processing strategy that mitigates such information loss while effectively reducing the signal dimensions. By utilizing a coarse *a priori* probability distribution of the directions-of-arrival (DOAs) of the signals, this technique optimizes the compressive measurement matrix through the maximization of the mutual information between the compressive measurement output and the signal DOAs. Once the compressive measurement matrix is optimized, we can estimate the signal DOAs and power using a spatial spectrum estimator, such as the minimum variance distortionless response (MVDR), and the results enable robust beamforming for interference-free signal extraction [18, 19]. In [20], this approach is extended to difference coarray-based DOA estimation exploiting sparse arrays.

In practice, however, a coarse prior distribution of the signal DOAs may not always be available, thereby limiting the application of this technology. In this paper, we develop an iterative learning approach that uses the estimated DOA spectrum as the prior for the next iteration. In so doing, the proposed technique removes the main hurdle of requiring the prior information of the signal arrivals, thereby making the optimized compressive measurement much more feasible in practice.

Related works are recently reported in the literature in which DOA estimation is considered as an initial alignment stage for massive MIMO channel estimation. In [21], a Bayesian learning-based adaptive beamforming strategy is developed to select the beamforming vectors from the hierarchical codebook based on the posterior DOA distribution. In [22], a deep neural network-based approach is exploited to design a sequence of adaptive sensing vectors. Both approaches consider the single-user single-path scenario with only a single RF chain involved. In the signal model considered in this work, on the other hand, the number of antennas N and the dimension of the compressed measurement vector $M (\ll N)$ can take arbitrary values, and no codebook is needed.

Notations: Lower- and upper-case bold characters are used to describe vectors and matrices, respectively. In particular, $\boldsymbol{I}_N$ stands for the $N \times N$ identity matrix, and $\mathbf{0}$ is a vector with a proper dimension. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ respectively denote the transpose and conjugate transpose (Hermitian) of a matrix or vector. $|\boldsymbol{A}|$ represents the determinant of matrix $\boldsymbol{A}$, and $\mathrm{E}_x(\cdot)$ denotes the statistical expectation with respect to x.

II. Signal Model

Consider an N-element massive MIMO receiver in which D far-field uncorrelated sources impinge from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_D]^{\mathrm{T}}$. The received baseband signal vector of the array at the discrete time t can be modeled as

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

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 with the dth column representing the dth source's steering vector $\boldsymbol{a}(\theta_d) \in \mathbb{C}^N$. Note that no specific array configuration is assumed in this model. In addition, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_D(t)]^{\mathrm{T}} \in \mathbb{C}^D$ denotes the signal waveform vector, and $\boldsymbol{n}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathrm{n}}^2 \mathbf{I}_N)$ denotes the zero-mean complex additive white Gaussian noise vector.

Denote $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_M]^{\mathrm{T}} \in \mathbb{C}^{M\times N}$ with $M \ll N$ as the compressive measurement matrix, as shown in Fig. 1. We introduce a row-orthonormal constraint on $\boldsymbol{\Phi}$, i.e., $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} = \mathbf{I}_M$ so that the noise power does not change after applying the compressive measurement and the mutual information does not fluctuate as the result that the compressive measurement matrix $\boldsymbol{\Phi}$ scales up or down. When $\boldsymbol{\Phi}$ is applied to the massive MIMO receiver, the received signal vector $\boldsymbol{x}(t) \in \mathbb{C}^N$ is compressed into an M-dimensional measurement vector $\boldsymbol{y}(t)$, expressed as

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

Hence, $\boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\theta}) \in \mathbb{C}^{M\times D}$ represents a sketch of the massive MIMO array manifold with a significantly reduced dimension. The objective of the proposed work is to jointly optimize the compressive measurement matrix $\boldsymbol{\Phi}$ and estimate the signal DOAs without any prior information of the signal angular distribution. Toward this end, an iterative learning approach is proposed. In this approach, a random compressive measurement matrix is used in the first iteration, and the estimated spatial spectrum obtained in each iteration is used as the prior information of the sigal angular distribution in the subsequent iteration of the iterative optimization of the compressive measurement matrix.

III. Probabilistic Optimization of Compressive Measurement Matrix

In this section, we briefly summarize the optimization scheme of the compressive measurement matrix using a probabilistic model based on [17, 18].

A. Probabilistic signal model

We treat the signal DOA θ as a random variable that occupies a region Θ with a probability density function (PDF), denoted as $f(\theta)$. Unlike in [17, 18] which assume $f(\theta)$ to be known *a priori*, we consider in this paper that no prior information about $f(\theta)$ is available but will be estimated.

We first discretize the PDF $f(\theta)$ into K angular bins with an equal width of $\Delta\bar{\theta}$, where the nominal DOA of the kth angular bin is denoted as $\bar{\theta}_k$. The discretized probability mass function (PMF) of each angular bin is denoted as $p_k = f(\bar{\theta}_k)\Delta\bar{\theta}$ with $\sum_{k\in\mathcal{K}} p_k = 1$, where $\mathcal{K} = \{1, 2, \cdots, K\}$.

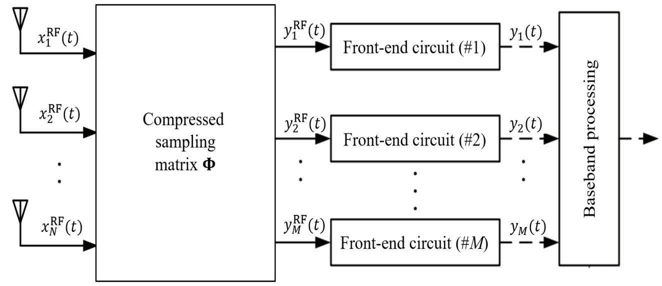


Fig. 1. Block diagram of compressive measurement sampling scheme. Solid lines denote analog signal flows and dashed lines denote digital signal flows.

The signal arrival $s(t)$ in an angular bin is modeled as a zero-mean complex Gaussian random variable with variance σ_s^2, i.e., $s(t) \sim \mathcal{CN}(0, \sigma_s^2)$. Considering the kth angular bin with nominal DOA $\bar{\theta}_k$, the compressed measurement vector is given as

$$\boldsymbol{y}(t)|_{\theta=\bar{\theta}_k} = \boldsymbol{\Phi}\big(\boldsymbol{a}(\bar{\theta}_k)s(t) + \boldsymbol{n}(t)\big). \quad (3)$$

With the law of total probability, the PDF of the compressed measurement vector $\boldsymbol{y}(t)$ can be expressed as

$$f(\boldsymbol{y}) = \mathrm{E}_\theta\{f(\boldsymbol{y}|\theta)\} = \int_{\theta\in\Theta} f(\boldsymbol{y}|\theta) f(\theta)\,\mathrm{d}\theta \approx \sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}|\bar{\theta}_k), \quad (4)$$

where the corresponding conditional PDF $f(\boldsymbol{y}|\bar{\theta}_k)$ is given as

$$f(\boldsymbol{y}|\bar{\theta}_k) = \frac{1}{\pi^M \left|\mathbf{C}_{\boldsymbol{yy}|\bar{\theta}_k}\right|} e^{-\boldsymbol{y}^{\mathrm{H}}\mathbf{C}^{-1}_{\boldsymbol{yy}|\bar{\theta}_k}\boldsymbol{y}}, \quad (5)$$

with

$$\mathbf{C}_{\boldsymbol{yy}|\bar{\theta}_k} = \boldsymbol{\Phi}\big(\sigma_s^2 \boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \sigma_{\mathrm{n}}^2\mathbf{I}_N\big)\boldsymbol{\Phi}^{\mathrm{H}} \quad (6)$$

denoting the covariance matrix of $\boldsymbol{y}(t)|_{\theta=\bar{\theta}_k}$. In this case, the PDF of the compressed measurement vector $\boldsymbol{y}(t)$ is a weighted sum of K Gaussian distributions, thereby forming a Gaussian mixture distribution.

B. Compressive sensing matrix optimization

We optimize the compressive measurement matrix $\boldsymbol{\Phi}$ through the maximization of the mutual information between the compressed signal vector $\boldsymbol{y}(t)$ and the DOA of the signal arrivals [23], and the gradient-based strategy [24] is used. The mutual information $I(\boldsymbol{y};\theta)$ between $\boldsymbol{y}(t)$ and θ is given as

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

where $h(\boldsymbol{y}) = -\mathrm{E}_{\boldsymbol{y}}\big\{\log[f(\boldsymbol{y})]\big\}$ denotes the differential entropy of $\boldsymbol{y}$ and $h(\boldsymbol{y}|\theta) = -\mathrm{E}_{\boldsymbol{y},\theta}\big\{\log[f(\boldsymbol{y}|\theta)]\big\}$ is the conditional differential entropy of $\boldsymbol{y}$ given signal DOA θ. The corresponding gradient of the mutual information $I(\boldsymbol{y};\theta)$ with respect to the compressive measurement matrix $\boldsymbol{\Phi}$ becomes

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\theta) = \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}) - \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}|\theta), \quad (8)$$

where $\nabla_{\mathbf{\Phi}}\{\cdot\}$ denotes the gradient operator with respective to $\mathbf{\Phi}$.

Using the discretized PMF approximation described in Section III-A, the expression of $\nabla_{\mathbf{\Phi}} I(\boldsymbol{y};\theta)$ can be analytically derived as [17, 18]:

$$\begin{aligned}
&\nabla_{\mathbf{\Phi}} I(\boldsymbol{y};\theta)\\
&\approx \frac{\sum_{k\in\mathcal{K}} p_k \left|\frac{\mathbf{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_{\rm n}^2}\right|^{-1} \left[\frac{\mathbf{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_{\rm n}^2}\right]^{-1} \mathbf{\Phi}\left(\gamma \boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\rm H}(\bar{\theta}_k)+\mathbf{I}_N\right)}{\sum_{k\in\mathcal{K}} p_k \left|\frac{\mathbf{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_{\rm n}^2}\right|^{-1}}\\
&-\sum_{k\in\mathcal{K}} p_k \left[\frac{\mathbf{C}_{\boldsymbol{yy}|\bar{\theta}_k}}{\sigma_{\rm n}^2}\right]^{-1} \mathbf{\Phi}\left(\gamma \boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\rm H}(\bar{\theta}_k)+\mathbf{I}_N\right), \qquad (9)
\end{aligned}$$

where $\gamma = \sigma_s^2/\sigma_{\rm n}^2$ denotes the input signal-to-noise ratio (SNR) of the signal.

Using the above information gradient $\nabla_{\mathbf{\Phi}} I(\boldsymbol{y};\theta)$, we can iteratively optimize the compressive measurement matrix using the following expression:

$$\mathbf{\Phi} \leftarrow \mathbf{\Phi} + \alpha \nabla_{\mathbf{\Phi}} I(\boldsymbol{y};\theta), \qquad (10)$$

where $\alpha > 0$ is a step size. This procedure is iterated and the rows of $\mathbf{\Phi}$ are orthonormalized in each iteration.

IV. Proposed Iterative Learning Scheme

Because of the analogy and certain shared properties between a normalized power spectrum density (PSD) and a PDF, a normalized PSD is often exploited in lieu of a PDF in signal processing [25]. In this work, we treat the normalized spatial power spectrum estimate as the posterior angular distribution of the signals and feed it as the prior information for the subsequent iteration. Other learning methods, such as those based on Bayesian learning [21] and deep learning [22], can also be used for this purpose but generally with much higher complexity.

When T samples are available, the sample covariance matrix of the compressed measurement vector $\boldsymbol{y}(t)$ is expressed as $\hat{\mathbf{R}}_{\boldsymbol{yy}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}(t)\boldsymbol{y}^{\rm H}(t)$. Then, an MVDR estimator computes the spatial power spectrum using the following expression:

$$P^{(i)}(\theta) = \frac{\boldsymbol{a}^{\rm H}(\theta)(\mathbf{\Phi}^{(i)})^{\rm H}\mathbf{\Phi}^{(i)}\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\rm H}(\theta)(\mathbf{\Phi}^{(i)})^{\rm H}\left(\hat{\mathbf{R}}^{(i)}_{\boldsymbol{yy}}\right)^{-1}\mathbf{\Phi}^{(i)}\boldsymbol{a}(\theta)}, \qquad (11)$$

where superscript $^{(i)}$ is added to indicate the ith iteration. Based on this result, we set the prior distribution in the subsequently iteration as

$$\hat{p}_k^{(i+1)} = \frac{P^{(i)}(\theta_k)}{\sum_{j\in\mathcal{K}} P^{(i)}(\theta_j)}. \qquad (12)$$

When there is no prior information about the signal arrivals in the initial conditions, we use a random compressive measurement matrix consisting of Gaussian random variables in the first iteration where each row is orthonormal.

To summarize, the proposed iterative learning scheme contains two loops. In the inner loop, the compressive measurement matrix $\mathbf{\Phi}$ is optimized using (9) and (10), whereas the outer loop learns $\hat{p}_k$ based on (11) and (12).

V. Simulation Results

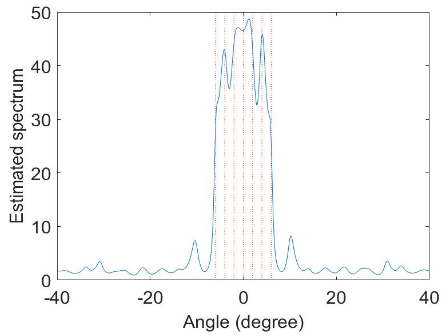
Two simulation examples are presented to verify the effectiveness of the proposed approach. In the first example, a uniform linear array (ULA) with half-wavelength interelement spacing is used, whereas the second example exploits a sparse linear array in which the sensors are randomly placed on a half-wavelength grid.

We first consider a massive MIMO system with $N = 50$ omni-directional antennas separated by half-wavelength. 7 far-field uncorrelated sources impinge from directions $[-6°, -4°, -2°, 0°, 2°, 4°, 6°]$. All sources have the same input SNR of 10 dB and the number of snapshots is $T = 100$. The dimension of the compressed measurement vector $\boldsymbol{y}(t)$ is $M = 10$, rendering the compression ratio to be $N/M = 5$. We uniformly discretize the PDF of DOA with a width of $\Delta\bar{\theta} = 0.1°$ over all azimuth directions $0 \le \theta \le 180°$, resulting in $K = 1,801$ components in the Gaussian mixture model. The step size used in the inner optimization is set as $\alpha = 0.001$. The maximum number of iterations used in the inner loop is 100.

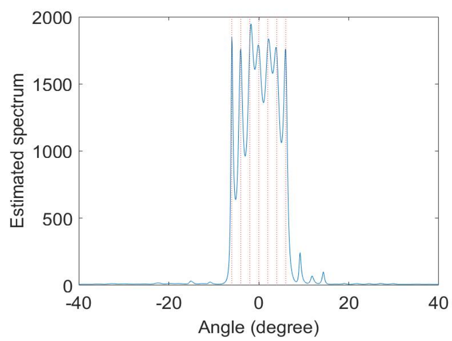
The MVDR spatial spectra obtained from the proposed iterative learning approach in first three iterations are shown in Fig. 2. In the first iteration, a random compressive measurement matrix is used. As shown in Fig. 2(a), because of the information loss as the result of using an unoptimized compressive measurement matrix, the estimated spatial spectrum does not resolve all the sources. The yielding result, however, can be used as a coarse distribution of the signal arrivals that can aid the optimization of the compressive measurement matrix in the subsequent outer iterations. It is clear in Figs. 2(b) and 2(c) that the estimated spatial spectrum converges very fast and the proposed approach correctly estimates the DOAs of all 7 sources in both the second and third iterations. The results imply that the compressive measurement matrix $\mathbf{\Phi}$ is effectively optimized through iterative learning. The comparison between Figs. 2(a) and 2(c) clearly demonstrates the effectiveness of the optimized compressive measurement matrix.

For reference, we also show the results of two cases when no compressive measurement is made. In the first case, as shown in Fig. 3(a), we use 10 antennas that are connected to 10 RF chains. Because of the small number of antennas, the 7 sources are not resolved in this case. In the second case, we consider 50 antennas that are connected to 50 RF chains. As shown in Fig. 3(b), this case achieves the best performance and it shows clean sidelobe region values. However, as it requires full number of RF chains, its complexity is significantly higher.

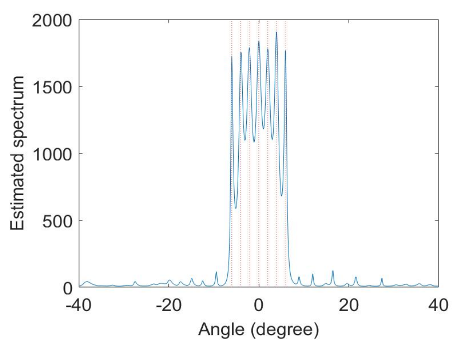
In the second example, we randomly place the $N = 50$ omni-directional antennas on a half-wavelength grid between 0 and 100 wavelengths. We consider 7 far-field un-



(a) First iteration



(b) Second iteration

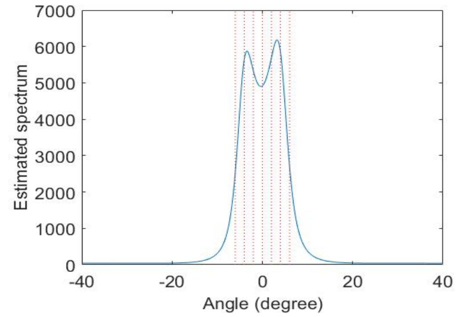


(c) Third iteration

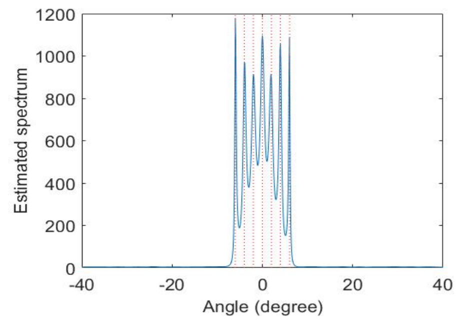
Fig. 2. Estimated spatial spectrum through iterative compressive measurement matrix optimization for the ULA case.

correlated sources which are closely located at directions $[-3^\circ, -2^\circ, -1^\circ, 0^\circ, 1^\circ, 2^\circ, 3^\circ]$. The other settings remain the same as the first example, e.g., the dimension of the compressed measurement vector $\boldsymbol{y}(t)$ is $M = 10$ and the input SNR of all sources is 10 dB.

Because of the large array aperture, all 7 sources can be resolved in the first iteration in Fig. 4(a), despite of the the closer angular separation between the sources. However, as a result of the sparse array configuration, we observe high sidelobes that spread across all spatial directions [26, 27]. By applying the proposed approach, as shown in Figs. 4(b) and 4(c), the floor is effectively reduced and more consistent spatial spectrum estimation is achieved, thereby demonstrating improved compressive measurement through the iterative learning process.



(a) Using 10 antennas and 10 RF chains



(b) Using 50 antennas and 50 RF chains

Fig. 3. Estimated spatial spectrum without performing compressive measurement.

VI. Conclusion

In this paper, we developed an iterative learning scheme for effective reduction of the array dimensionality in a massive MIMO system. The proposed scheme jointly optimizes the compressive measurement matrix and estimates the signal DOAs, thus achieving effective dimension reduction for low-complexity data acquisition and processing with minimum performance loss in the absence of prior information of the signal arrival distributions. The optimized compressive measurement matrix achieves comparable estimation accuracy with the full array case, and outperforms compressive measurements using random matrices.

VII. References

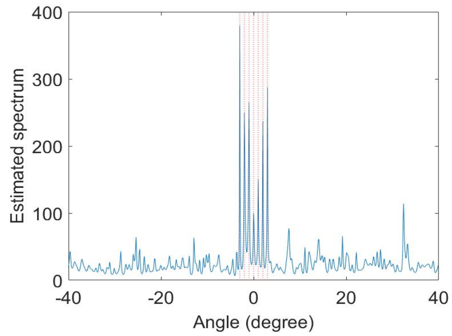
[1] R. C. de Lamare, "Massive MIMO systems: Signal processing challenges and research trends," *URSI Radio Science Bulletin*, Dec. 2013.

[2] 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.

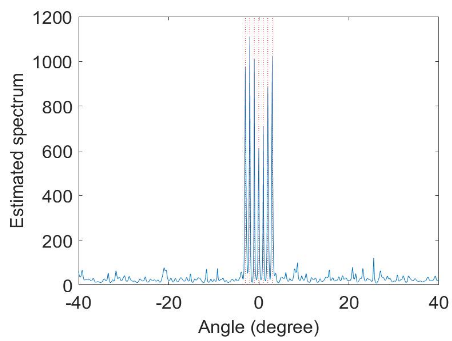
[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[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.

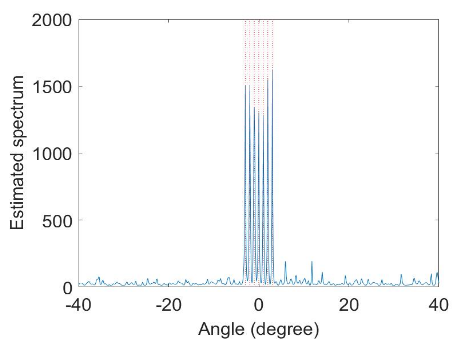
[5] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular



(a) First iteration



(b) Second iteration



(c) Third iteration

Fig. 4. Estimated spatial spectrum of the randomly spaced array case.

systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[6] 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.

[7] S. Wang, V. M. Patel, and A. Petropulu, "Multidimensional sparse Fourier transform based on the Fourier projection-slice theorem," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 54–69, Jan. 2019.

[8] S. Fortunati, L. Sanguinetti, F. Gini, M. S. Greco, and B. Himed, "Massive MIMO radar for target detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 859–871, 2020.

[9] A. Bose, A. Ghauri, and M. Soltanalian, "Limits of transmit beamforming for massive MIMO radar," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2020, pp. 1157–1161.

[10] S. Sun and Y. D. Zhang, "Multi-frequency sparse array-based massive MIMO radar for autonomous driving," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2020, pp. 1167–1171.

[11] W. Yan, X. Yuan, Z. He, and X. Kuai, "Large intelligent surface aided multiuser MIMO: Passive beamforming and information transfer," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–7.

[12] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304-44321, 2021.

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

[14] V. Cevher, A. Gurbuz, J. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Las Vegas, NV, March–April 2008.

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

[16] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of Fisher information and the Cramer-Rao bound for nonlinear parameter estimation after compressed sensing," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, 2013, pp. 6630-6634.

[17] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive array," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, New Orleans, LA, March 2017.

[18] Y. Gu and Y. D. Zhang, "Compressive sampling optimization for user signal parameter estimation in massive MIMO systems," *Digital Signal Process.*, vol. 94, pp. 105-113, Oct. 2019.

[19] Y. Gu, N. A. Goodman, and Y. D. Zhang, "Adaptive beamforming via sparsity-based reconstruction of covariance matrix," in A. De Maio, Y. C. Eldar, and A. Haimovich (eds.), *Compressed Sensing in Radar Signal Processing*, Cambridge Univ. Press, 2019.

[20] M. Guo, Y. D. Zhang, and T. Chen, "DOA estimation using compressed sparse array," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4133-4146, Aug. 2018.

[21] N. Ronquillo, S. Chiu, and T. Javidi, "Sequential learning of CSI for mmWave initial alignment," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2019, pp. 1278-1283.

[22] F. Sohrabi, Z. Chen, and W. Yu, "Deep active learning approach to adaptive beamforming for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2347-2360, Aug. 2021.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory (2nd Ed.)*. Wiley, 2006.

[24] 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.

[25] X. Yin and X. Cheng, *Propagation Channel Characterization, Parameter Estimation, and Modeling for Wireless Communications*, IEEE Press, 2016.

[26] D. H. Johnson and D. E. Dudgeon, *rray Signal Processing: Concepts and Techniques*, Prentice Hall, 1993.

[27] S. Sun and Y. D. Zhang, "4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE Sel. Top. Signal Process.*, vol. 15, no. 4, pp. 879–891, June 2021.