Coherence Statistics of Structured Random Ensembles and Support Detection Bounds for OMP

Qiyou Duan, Taejoon Kim, *Member, IEEE*, Lin Dai, *Senior Member, IEEE*, and Erik Perrins, *Senior Member, IEEE*

***Abstract*—A structured random matrix ensemble that maintains constant modulus entries and unit-norm columns, often called a random phase-rotated (RPR) matrix, is considered in this letter. We analyze the coherence statistics of RPR measurement matrices and apply them to acquire probabilistic performance guarantees of orthogonal matching pursuit (OMP) for support detection (SD). It is revealed via numerical simulations that the SD performance guarantee provides a tight characterization, especially when the signal is sparse.**

***Index Terms*—Random phase-rotated (RPR) measurements, coherence statistics, structured random ensemble, support detection (SD), orthogonal matching pursuit (OMP).**

I. Introduction

RANDOM matrix ensembles have found wide applications in fields of wireless communications and signal processing [1]–[4]. Despite the fact that most studied Gaussian measurement ensembles offer trackable analyses and appealing results [5]–[7], they are of somewhat limited use in practical applications because the design of measurement matrices is usually subject to physical or other constraints provided by a specific system architecture. It is desirable to explore random matrix ensembles with hidden structure from a computational and an application-oriented point of view.

Coherence has been utilized to measure the quality of the measurement matrix [8]. Analysis of coherence statistics of random vectors/matrices plays an important role in solving a series of signal processing problems including the Grassmannian line packing [9], [10], random vector quantization [11], [12], and support detection (SD) [5], [13]–[15]. In particular, the performance of SD considerably varies with the characteristics of measurement matrices. There is a certain class of random matrix ensembles with hidden structures that can demonstrate an improvement in SD performance guarantees compared to Gaussian ensembles [5]. Distinguished from the Gaussian measurement matrix that does not contain hidden constraints, the random phase-rotated (RPR) measurement matrix, where each entry is drawn from the constant modulus uniform phase rotation distribution, brings the benefits of maintaining unit-norm columns and constant modulus entries of the measurement matrix. This measurement ensemble has been utilized in advanced beamforming and precoding for wireless communications [16], [17].

In this letter, we calculate high probability bounds on the coherence statistics of RPR measurement matrices and apply them to obtain SD performance guarantees for orthogonal matching pursuit (OMP), which is a low-complexity, greedy approach for SD [5], [18]. The performance bound is in terms of the required number of measurements for any given number of supports and system dimensions. A free variable is introduced, which is optimized to further tighten the performance bound. The main motivation is that previous work relying on the coherence property did not contain hidden constraints that are suitable for SD of OMP. Numerical evaluations demonstrate that the analyzed SD performance guarantee of OMP is tight, especially when the signal is sparse.

II. Coherence Statistics

Suppose a random measurement matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N] \in \mathbb{C}^{M\times N}$ with $\mathbf{a}_n \in \mathbb{C}^{M\times 1}$ being the nth column of $\mathbf{A}$. Each entry of $\mathbf{A}$ is constant modulus and drawn from the random phase rotation variable as

$$A_{mn} = \frac{1}{\sqrt{M}} e^{j\Theta_{mn}}, \tag{1}$$

where A_{mn} denotes the mth row and nth column entry of $\mathbf{A}$, $m = 1, \ldots, M$, $n = 1, \ldots, N$, and the phase Θ_{mn} is an independent and identically distributed (i.i.d.) uniform random variable, i.e., $\Theta_{mn} \sim \mathcal{U}[0, 2\pi)$. With the construction in (1), $\mathbf{A}$ maintains $\|\mathbf{a}_n\| = 1$, $\forall n$.

The coherence of $\mathbf{A}$ is the maximum absolute correlation between two distinct columns of $\mathbf{A}$ [19], which is given by

$$\mu(\mathbf{A}) \triangleq \max_{i\neq j} |\mathbf{a}_i^* \mathbf{a}_j|, \tag{2}$$

where $(\cdot)^*$ denotes the conjugate transpose. Characterizing the distribution of $\mu(\mathbf{A})$ is of interest - however, it is challenging to directly derive the distribution of $\mu(\mathbf{A})$ when $\mathbf{A}$ follows (1). To circumvent this difficulty, we relegate to find a lower bound on the cumulative distribution function (CDF) of $\mu(\mathbf{A})$ instead. We start by building a connection between the vector drawn from the distribution in (1) and the vector consisting of Bernoulli random variables.

Lemma 1: Let $\mathbf{p} \in \mathbb{C}^{M\times 1}$ and $\mathbf{q} \in \mathbb{R}^{M\times 1}$ be random vectors with i.i.d. entries $p_m = 1/\sqrt{M} e^{j\theta_m}$, $\theta_m \in \mathcal{U}[0, 2\pi)$, and $q_m \in \{-1/\sqrt{M}, 1/\sqrt{M}\}$ with equal probability for $m = 1, \ldots, M$, respectively. Then, for any unit-norm vector $\mathbf{u} \in \mathbb{C}^{M\times 1}$, the

Manuscript received August 27, 2019; accepted September 14, 2019. Date of publication September 20, 2019; date of current version October 2, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Marco Felipe Duarte. *(Corresponding author: Qiyou Duan.)*

Q. Duan and L. Dai are with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong (e-mail: qyduan.ee@my.cityu.edu.hk; lindai@cityu.edu.hk).

T. Kim and E. Perrins are with the Department of Electrical Engineering and Computer Science, The University of Kansas, Lawrence, KS 66045 USA (e-mail: taejoonkim@ku.edu; esp@ieee.org).

Digital Object Identifier 10.1109/LSP.2019.2942737

following inequality holds

$$\mathbb{E}\left[|\mathbf{p}^*\mathbf{u}|^{2\,k}\right] \leq \mathbb{E}\left[|\mathbf{q}^*\bar{\mathbf{u}}|^{2\,k}\right], \tag{3}$$

where $\bar{\mathbf{u}} \in \mathbb{R}^{M\times 1}$ has each entry $\bar{u}_m = 1/\sqrt{M}$, $\forall m$, k is a nonnegative integer, and the expectations are taken over $\mathbf{p}$ and $\mathbf{q}$, respectively.

Proof: See Appendix A. ■

Based on Lemma 1, we characterize a bound on the distribution of $|\mathbf{p}^*\mathbf{u}|$ below.

Lemma 2: Suppose the vectors $\mathbf{p}$ and $\mathbf{u}$ defined in Lemma 1. Then, for any $\delta > 0$, the following inequality holds

$$\Pr(|\mathbf{p}^*\mathbf{u}| \geq \delta) \leq \left(1 - \frac{2}{g}\right)^{-\frac{1}{2}} e^{-\frac{\delta^2 M}{g}},\ g > 2. \tag{4}$$

Proof: See Appendix B. ■

Remark 1: It is also possible to derive an upper bound on $\Pr(|\mathbf{p}^*\mathbf{u}| \geq \delta)$ by leveraging the matrix Bernstein inequality [20, Theorem 1.6.2], which leads to $\Pr(|\mathbf{p}^*\mathbf{u}| \geq \delta) \leq 4e^{-\frac{3M\delta^2}{2\delta\sqrt{M}+6}}$. However, this bound is looser than that in (4).

A lower bound on the CDF of $\mu(\mathbf{A})$ in (2) can be found.

Theorem 1: Suppose a matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ consisting of i.i.d. entries $A_{mn} = 1/\sqrt{M}e^{j\Theta_{mn}}$, $\Theta_{mn} \in \mathcal{U}[0, 2\pi)$, $m = 1, \ldots, M$, $n = 1, \ldots, N$. Then, the following holds for $g > 2$,

$$\Pr(\mu(\mathbf{A}) < \delta) \geq \left(1 - \left(1 - \frac{2}{g}\right)^{-\frac{1}{2}} e^{-\frac{\delta^2 M}{g}}\right)^{\frac{N(N-1)}{2}}. \tag{5}$$

Proof: The inner product between two distinct column vectors of $\mathbf{A}$ satisfies

$$\mathbf{a}_{n_1}^*\mathbf{a}_{n_2} = \sum_{m=1}^{M} \frac{1}{M} e^{j\Delta\Theta_m} \overset{\mathrm{d}}{=} \sum_{m=1}^{M} \frac{1}{M} e^{j\xi_m} = \mathbf{p}^*\bar{\mathbf{u}}, \tag{6}$$

where $\Delta\Theta_m \triangleq \Theta_{mn_2} - \Theta_{mn_1}$, $n_1 \neq n_2$, is the difference between two independent uniform random variables, whose probability density function is given by

$$p(\Delta\Theta_m) = \begin{cases} \frac{2\pi - |\Delta\Theta_m|}{4\pi^2}, & \text{if } -2\pi \leq \Delta\Theta_m < 2\pi \\ 0, & \text{otherwise.} \end{cases}$$

In (6), $\bar{\mathbf{u}}$ follows the same definition in Lemma 1, and we use the fact that $e^{j\Delta\Theta_m} = e^{j\,\mathrm{mod}(\Delta\Theta_m, 2\pi)}$ and $\xi_m \triangleq \mathrm{mod}(\Delta\Theta_m, 2\pi)$, in which $\mathrm{mod}(a, b)$ is the modulo b of a. Note that $\xi_m \sim \mathcal{U}[0, 2\pi)$ and it verifies that $\mathbf{a}_{n_1}^*\mathbf{a}_{n_2}$ has the same distribution as $\mathbf{p}^*\bar{\mathbf{u}}$ in (6), where $\overset{\mathrm{d}}{=}$ is the equality in distribution.

By Lemma 2, we now have $\Pr(|\mathbf{a}_{n_1}^*\mathbf{a}_{n_2}| < \delta) = \Pr(|\mathbf{p}^*\bar{\mathbf{u}}| < \delta) \geq 1 - (1 - 2/g)^{-1/2} e^{-\delta^2 M/g}$. Then, the maximum order statistic of $|\mathbf{a}_{n_1}^*\mathbf{a}_{n_2}|$ is lower bounded by

$$\Pr\left(\max_{n_1 \neq n_2} |\mathbf{a}_{n_1}^*\mathbf{a}_{n_2}| < \delta\right) = \Pr(\mu(\mathbf{A}) \leq \delta)$$
$$\geq \left(1 - \left(1 - \frac{2}{g}\right)^{-\frac{1}{2}} e^{-\frac{\delta^2 M}{g}}\right)^{\frac{N(N-1)}{2}}.$$

This completes the proof. ■

Remark 2: Because Bernoulli random matrices with each entry filled with $\pm\frac{1}{\sqrt{M}}$ can be regarded as a special case of the RPR matrices in (1) when $\Theta_{mn} \in \{0, \pi\}$ with equal probability, $\forall m, n$, the coherence statistic in (5) also holds for the Bernoulli random matrix.

Algorithm 1: OMP for SD.

Input: $\mathbf{A}$, $\mathbf{y}$, and K.
Output: $\hat{\mathcal{S}}$.

1: Initialization: Set iteration number $t = 1$, $\mathbf{r}_0 = \mathbf{y}$, and $\mathcal{S}_0 = \phi$.
2: Select the active index: $i_t = \mathrm{argmax}_{n\in\mathcal{S}_{t-1}^C} |\mathbf{a}_n^*\mathbf{r}_{t-1}|$.
3: Update the active support set: $\mathcal{S}_t = \mathcal{S}_{t-1} \cup \{i_t\}$.
4: Estimate the signal vector: $\hat{\mathbf{x}}_t = \mathrm{argmin}_{\mathbf{z}:\mathrm{supp}(\mathbf{z})=\mathcal{S}_t} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2$.
5: Update the residual: $\mathbf{r}_t = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_t = \mathbf{y} - \mathbf{A}_{\mathcal{S}_t}\hat{\mathbf{x}}_{\mathcal{S}_t}$.
6: **if** $|\mathcal{S}_t| = K$ **then** terminate and
7: **return** $\hat{\mathcal{S}} = \mathcal{S}_t$.
8: **else** $t = t + 1$ and
9: **return** to Step 2.

III. Support Detection Bounds for OMP

In this section, the coherence statistics of RPR measurement matrices are applied to obtain the probability bounds of SD for OMP.

A. Measurement Model and OMP Algorithm

Suppose a measurement model

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{7}$$

where each entry of $\mathbf{A} \in \mathbb{C}^{M\times N}$ follows (1). Here, the assumption is that the number of measurements M is smaller than the signal dimension N, i.e., $M < N$. The signal $\mathbf{x} \in \mathbb{C}^{N\times 1}$ in (7) has K nonzero elements (supports) whose indexes are defined by the support set

$$\mathcal{S} = \mathrm{supp}(\mathbf{x}) = \{n_1, \ldots, n_K | x_{n_k} \neq 0, n_k \in \{1, \ldots, N\}\}, \tag{8}$$

where $|\mathcal{S}| = K \ll M$. The goal is to detect the support set $\mathcal{S}$ from the measurement $\mathbf{y} \in \mathbb{C}^{M\times 1}$ in (7).

An iterative procedure of OMP for SD is depicted in Algorithm 1 for the measurement model in (7). To make sure that the active index determined in Step 2 is a true support, the following sufficient condition [19] should be met,

$$\rho(\mathbf{r}_{t-1}) \triangleq \frac{\|\mathbf{A}_{\mathcal{S}^C}^*\mathbf{r}_{t-1}\|_\infty}{\|\mathbf{A}_{\mathcal{S}}^*\mathbf{r}_{t-1}\|_\infty} < 1, \tag{9}$$

where $\mathbf{A}_{\mathcal{S}} \in \mathbb{C}^{M\times K}$ is the submatrix formed by taking the columns of $\mathbf{A}$ indexed by $\mathcal{S}$ and $\mathbf{A}_{\mathcal{S}^C} \in \mathbb{C}^{M\times(N-K)}$ is the complementary submatrix of $\mathbf{A}_{\mathcal{S}}$. The nonzero coefficients $\hat{\mathbf{x}}_{\mathcal{S}_t} \in \mathbb{C}^{t\times 1}$ estimated in Step 5 are formed by extracting the nonzero elements of $\hat{\mathbf{x}}_t \in \mathbb{C}^{N\times 1}$ indexed by $\mathcal{S}_t$ and given by $\hat{\mathbf{x}}_{\mathcal{S}_t} = (\mathbf{A}_{\mathcal{S}_t}^*\mathbf{A}_{\mathcal{S}_t})^{-1}\mathbf{A}_{\mathcal{S}_t}^*\mathbf{y}$. It is crucial to recognize that the updated residual $\mathbf{r}_t$ is orthogonal to the columns of $\mathbf{A}_{\mathcal{S}_t}$. The OMP detects one support at each iteration and runs for exactly K iterations.

B. Support Detection Performance Guarantee

We provide the SD performance guarantee of the OMP in Algorithm 1 as follows.

Theorem 2: Suppose the measurement model in (7) with the RPR measurement matrix $\mathbf{A}$ based on (1). Then, the OMP in Algorithm 1 detects the K supports of $\mathbf{x}$ for any (M, N) with

$$\Pr(\mathcal{V}_{\mathrm{SSD}}) \geq 1 - \left(1 - \frac{2}{g}\right)^{-\frac{1}{2}} KN \cdot e^{-\frac{M}{gK^2}},\ g > 2, \tag{10}$$

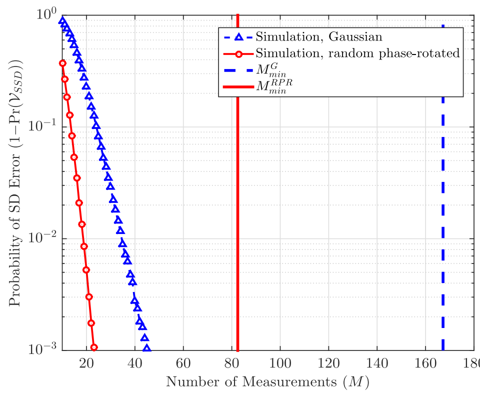

Fig. 1. SD performance guarantees of OMP with the RPR and Gaussian measurements when $N = 200$, $K = 2$, $\epsilon = 10^{-1}$, and $g^{\text{opt}} = 2.1020$.

where $\mathcal{V}_{\text{SSD}}$ is the event of successful SD (SSD) after K iterations. When the number of measurements M satisfies

$$M \geq gK^2 \ln\left(KN/(\epsilon\sqrt{1-2/g})\right), \; g > 2, \tag{11}$$

for $\epsilon \in (0, 1)$, Algorithm 1 satisfies $\Pr(\mathcal{V}_{\text{SSD}}) \geq 1 - \epsilon$.

Proof: See Appendix C. ■

To further tighten the lower bound in (11), we optimize the free variable g by minimizing the right hand side (r.h.s.) of (11) such that

$$g^{\text{opt}} = \underset{g>2}{\text{argmin}}\, f(g) \triangleq gK^2 \ln\left(KN/(\epsilon\sqrt{1-2/g})\right). \tag{12}$$

Theorem 3: The objective function $f(g)$ in (12) is convex for $g > 2$ and a closed-form expression of g^{opt} is given by

$$g^{\text{opt}} = \frac{2}{1 + \left(W_{-1}(-(\frac{\epsilon}{KN})^2 e^{-1})\right)^{-1}}, \tag{13}$$

where $W_{-1}(\cdot)$ is the lower branch of the Lambert W function [21], defined by $z = W_{-1}(ze^z)$ for $z < -1$.

Proof: See Appendix D. ■

IV. Numerical Simulations

To verify the SD performance guarantee in (11), we perform Monte Carlo simulations in Fig. 1, where the probability of SD error, i.e., $1 - \Pr(\mathcal{V}_{\text{SSD}})$, across different numbers of measurements M for $N = 200$ and $K = 2$, is evaluated. In the simulation, the signal $\mathbf{x}$ is generated by randomly choosing K supports with each support having $x_n = 1$, for $n \in \mathcal{S}$, and we compare with the existing coherence-based SD performance guarantee for the Gaussian random measurement matrix [5]. In Fig. 1, the vertical lines denote the minimum required M to guarantee the SD error rate $\epsilon = 10^{-1}$, where these values are given by the r.h.s. of (11) for the RPR measurements ($M_{min}^{RPR} = 82$ with $g^{\text{opt}} = 2.1020$ according to (13)), and $M \geq CK\ln(N/\epsilon)$ for the Gaussian case ($M_{min}^{G} = 168$ with $C = 11$) [5], respectively. Seen from Fig. 1, the obtained SD performance guarantee of RPR matrices provides a tighter characterization than the Gaussian case when the signal is sparse, i.e., K is small.

V. Conclusion and Discussion

The coherence statistics of RPR matrices were analyzed and applied to obtain the SD performance guarantees of OMP. The introduced free variable was optimized to further tighten the SD bound. Numerical simulations corroborated the theoretical findings and revealed that including the constant modulus and unit-norm structure for random measurement ensembles is desirable for SD using OMP.

In this work, we focused on the coherence statistics of RPR matrices to show the SD performance guarantees of OMP. In particular, we proved that OMP can achieve SSD with high probability, provided $M = O(K^2 \ln(KN))$ RPR measurements. It is of interest to compare our coherence-based analysis with the restricted isometry property (RIP)-based result since they are two main techniques in analyzing the performance guarantees of SD for OMP. By using the concentration inequality in [4, Theorem 2] and the method of proving the RIP for random matrices in [22, Theorem 5.2], one can obtain that $M \geq 16K\ln(N/K)/\delta^2$ is sufficient for the RPR matrices to satisfy the RIP with high probability, where $\delta \in (0, 1)$ is the restricted isometry constant. With $\delta < \frac{1}{\sqrt{K}}$ being a strict condition of SSD for OMP [23], the RIP-based SD bound can be given by $M \geq 16\, K^2 \ln(N/K)$, which is on par with our coherence-based results in Theorem 2.

Finally, one limitation of the work is that the SD bound becomes loose as K grows. Seen from Fig. 1, there is still room for further improvement by investigating a new structure of random measurement ensembles, which is subject to future research.

Appendix A
Proof of Lemma 1

Proof: The left hand side (l.h.s.) and r.h.s. of (3) can be rewritten as $\mathbb{E}[|\mathbf{p}^*\mathbf{u}|^{2\,k}] = \mathbb{E}[|\sum_{m=1}^{M} u_m e^{-j\theta_m}|^{2\,k}]/M^k$ and $\mathbb{E}[|\mathbf{q}^*\bar{\mathbf{u}}|^{2\,k}] = \mathbb{E}[|\sum_{m=1}^{M} \zeta_m|^{2\,k}]/M^{2\,k}$, respectively, where $\zeta_m \in \{1, -1\}, \forall m$, with equal probability. Thus, showing the inequality in (3) is equivalent to showing

$$M^k \mathbb{E}\left[\left|\sum_{m=1}^{M} u_m e^{-j\theta_m}\right|^{2\,k}\right] \leq \mathbb{E}\left[\left|\sum_{m=1}^{M} \zeta_m\right|^{2\,k}\right]. \tag{14}$$

The l.h.s. of (14) can be simplified as

$$\begin{aligned} &M^k \mathbb{E}\left[\left|\sum_{m=1}^{M} u_m e^{-j\theta_m}\right|^{2\,k}\right] \\ &\overset{(a)}{=} M^k \mathbb{E}\left[\left(\left|\sum_{m=1}^{M} u_m \cos(\theta_m)\right|^2 + \left|\sum_{m=1}^{M} u_m \sin(\theta_m)\right|^2\right)^k\right] \\ &\overset{(b)}{=} M^k \mathbb{E}\left[\left(\sum_{m=1}^{M} |u_m|^2\right)^k\right] \overset{(c)}{=} M^k, \end{aligned} \tag{15}$$

where (a) follows from the equality $e^{-j\theta_m} = \cos(\theta_m) - j\sin(\theta_m)$, (b) is due to the fact that $\mathbb{E}[\cos(\theta_{m_1})\cos(\theta_{m_2})] = \mathbb{E}[\sin(\theta_{m_1})\sin(\theta_{m_2})] = 0$ for $m_1 \neq m_2$, and (c) holds because $\|\mathbf{u}\|_2 = 1$. Expanding the r.h.s. of (14) leads to

$$\begin{aligned} \mathbb{E}\left[\left|\sum_{m=1}^{M} \zeta_m\right|^{2\,k}\right] &= \mathbb{E}\left[(M + G(M))^k\right] \\ &= \mathbb{E}\left[\sum_{i=0}^{k} \binom{k}{i} M^{k-i} G(M)^i\right] \geq M^k, \end{aligned} \tag{16}$$

where $G(M) \triangleq \sum_{m_1=1}^{M}\sum_{m_2=1,m_2\neq m_1}^{M} \zeta_{m_1}\zeta_{m_2}$. The inequality in (16) becomes the equality only if $k=0,1$ because $\mathbb{E}[\zeta_{m_1}\zeta_{m_2}]=0$ for $m_1\neq m_2$. On the other hand, when $k>1$, the strict inequality in (16) holds because $\mathbb{E}[\zeta_{m_1}^{2l_1}\zeta_{m_2}^{2l_2}]=1$ for any positive integers l_1, l_2, leading to $\mathbb{E}[G(M)^i]>0$ for $i>1$. Combining (15) and (16) results in (14). ■

Appendix B
Proof of Lemma 2

Proof: By using Markov's inequality, we have for $h\geq 0$,

$$\Pr(|\mathbf{p}^*\mathbf{u}|\geq\delta)=\Pr(|\mathbf{p}^*\mathbf{u}|^2\geq\delta^2)\leq\mathbb{E}\left[e^{h|\mathbf{p}^*\mathbf{u}|^2}\right]e^{-h\delta^2}. \tag{17}$$

The term $\mathbb{E}[e^{h|\mathbf{p}^*\mathbf{u}|^2}]$ in (17) can further be upper bounded for $h\in[0,\ M/2)$ by

$$\mathbb{E}\left[e^{h|\mathbf{p}^*\mathbf{u}|^2}\right]\leq\mathbb{E}\left[e^{h|\mathbf{q}^*\bar{\mathbf{u}}|^2}\right]\leq(1-2\,h/M)^{-\frac{1}{2}}, \tag{18}$$

where the first inequality is due to the Taylor series expansion of $\mathbb{E}[e^{h|\mathbf{p}^*\mathbf{u}|^2}]=\sum_{k=0}^{\infty}\frac{h^k}{k!}\mathbb{E}[|\mathbf{p}^*\mathbf{u}|^{2\,k}]$ and Lemma 1 applied to $\mathbb{E}[|\mathbf{p}^*\mathbf{u}|^{2\,k}]$, and $\bar{\mathbf{u}}$ follows the same definition in Lemma 1. The last step in (18) follows from the inequality $\mathbb{E}[e^{h|\mathbf{q}^*\mathbf{u}|^2}]\leq 1/\sqrt{1-2\,h/M}$ for $h\in[0,\ M/2)$ in [24, Lemma 5.2].

Inserting (18) into (17) leads to

$$\Pr(|\mathbf{p}^*\mathbf{u}|\geq\delta)\leq(1-2\,h/M)^{-\frac{1}{2}}e^{-h\delta^2}. \tag{19}$$

Because the inequality holds for any $h\in[0,M/2)$, substituting $h=M/g$, $g>2$, into (19) completes the proof. ■

Appendix C
Proof of Theorem 2

Proof: The proof is inspired by a similar theorem in [5, Theorem 6] and refines the results for the RPR measurement ensembles in conjunction with Lemma 2 and Theorem 1. We first elaborate two events: 1) $\mathcal{V}_{\text{SSD}}$ is defined on the basis of the condition in (9) as $\mathcal{V}_{\text{SSD}}\triangleq\{\max_{t=1,\ldots,K}\rho(\mathbf{r}_{t-1})=\frac{\|\mathbf{A}_{\mathcal{S}^C}^*\mathbf{r}_{t-1}\|_\infty}{\|\mathbf{A}_{\mathcal{S}}^*\mathbf{r}_{t-1}\|_\infty}<1\}$; and 2) The event that $\mu(\mathbf{A}_{\mathcal{S}})$ is bounded by $1/K$, i.e., $\mathcal{D}\triangleq\{\mu(\mathbf{A}_{\mathcal{S}})<1/K\}$. The event $\mathcal{D}$ is to restrict the $\mathcal{V}_{\text{SSD}}$ on a special class of $\mathbf{A}$ to ease the bound analysis below.

Conditioned on the event $\mathcal{D}$, the probability of SSD can be lower bounded by

$$\Pr(\mathcal{V}_{\text{SSD}})\geq\Pr(\mathcal{V}_{\text{SSD}}\cap\mathcal{D})=\Pr(\mathcal{V}_{\text{SSD}}|\mathcal{D})\Pr(\mathcal{D}). \tag{20}$$

From Theorem 1, $\Pr(\mathcal{D})$ in (20) can be lower bounded by

$$\begin{aligned}\Pr(\mathcal{D})=&\ \Pr\left(\mu(\mathbf{A}_{\mathcal{S}})\leq\frac{1}{K}\right)\\ \geq&\ \left[1-\left(1-\frac{2}{g}\right)^{-\frac{1}{2}}e^{-\frac{M}{gK^2}}\right]^{\frac{K(K-1)}{2}},\end{aligned} \tag{21}$$

where $g>2$. The conditional probability on the r.h.s. of (20) can be lower bounded by

$$\begin{aligned}\Pr(\mathcal{V}_{\text{SSD}}|\mathcal{D})=&\ \Pr\left(\max_t\frac{\|\mathbf{A}_{\mathcal{S}^C}^*\mathbf{r}_{t-1}\|_\infty}{\|\mathbf{A}_{\mathcal{S}}^*\mathbf{r}_{t-1}\|_\infty}<1\middle|\mathcal{D}\right)\\ \overset{(a)}{\geq}&\ \Pr\left(\max_t\frac{\sqrt{K}\max_{j\in\mathcal{S}^C}|\mathbf{a}_j^*\mathbf{r}_{t-1}|}{\|\mathbf{A}_{\mathcal{S}}^*\mathbf{r}_{t-1}\|_2}<1\middle|\mathcal{D}\right)\\ \overset{(b)}{\geq}&\ \Pr\left(\max_t\max_{j\in\mathcal{S}^C}\left|\mathbf{a}_j^*\mathbf{b}_{t-1}\right|<\frac{1}{K}\middle|\mathcal{D}\right)\\ \overset{(c)}{=}&\ \prod_{j\in\mathcal{S}^C}\Pr\left(\max_t\left|\mathbf{a}_j^*\mathbf{b}_{t-1}\right|<\frac{1}{K}\middle|\mathcal{D}\right)\\ \overset{(d)}{\geq}&\ \left[1-\left(1-\frac{2}{g}\right)^{-\frac{1}{2}}e^{-\frac{M}{gK^2}}\right]^{K(N-K)}\end{aligned} \tag{22}$$

where (a) is due to the inequality $\|\mathbf{u}\|_\infty\geq\|\mathbf{u}\|_2/\sqrt{K}$ for $\mathbf{u}\in\mathbb{C}^{K\times 1}$, (b) comes from $\mathbf{b}_{t-1}\triangleq\tilde{\mathbf{b}}_{t-1}/\|\tilde{\mathbf{b}}_{t-1}\|_2$ where $\tilde{\mathbf{b}}_{t-1}\triangleq\mathbf{r}_{t-1}/(\sqrt{K}\|\mathbf{A}_{\mathcal{S}}^*\mathbf{r}_{t-1}\|_2)$ and $\|\tilde{\mathbf{b}}_{t-1}\|_2\leq 1$ because $\|\mathbf{A}_{\mathcal{S}}^*\mathbf{r}_{t-1}\|_2/\|\mathbf{r}_{t-1}\|_2\geq\sqrt{\lambda_{min}(\mathbf{A}_{\mathcal{S}}^*\mathbf{A}_{\mathcal{S}})}\geq\sqrt{1-(K-1)\mu(\mathbf{A}_{\mathcal{S}})}\geq 1/\sqrt{K}$ by applying Gershgorin disc theorem [25], (c) holds due to the fact that the $N-K$ columns of $\mathbf{A}_{\mathcal{S}^C}$ are independent, and (d) is due to Lemma 2.

Substitute (21) and (22) into (20) yields

$$\begin{aligned}\Pr(\mathcal{V}_{\text{SSD}})\geq&\left[1-\left(1-\frac{2}{g}\right)^{-\frac{1}{2}}e^{-\frac{M}{gK^2}}\right]^{K(N-K)+\frac{K(K-1)}{2}}\\ \overset{(a)}{\geq}&\ 1-\left(1-\frac{2}{g}\right)^{-\frac{1}{2}}\left[K(N-K)+\frac{K(K-1)}{2}\right]e^{-\frac{M}{gK^2}}\\ \geq&\ 1-\left(1-\frac{2}{g}\right)^{-\frac{1}{2}}KNe^{-\frac{M}{gK^2}},\end{aligned}$$

where (a) holds because $(1-2/g)^{-1/2}e^{-\frac{M}{gK^2}}<1$ and $K(N-K)+K(K-1)/2>1$. Setting $(1-2/g)^{-1/2}KNe^{-\frac{M}{gK^2}}\leq\epsilon$ and taking the natural logarithm of both sides reveals that $\Pr(\mathcal{V}_{\text{SSD}})\geq 1-\epsilon$ when $M\geq gK^2\ln(KN/(\epsilon\sqrt{1-2/g}))$. ■

Appendix D
Proof of Theorem 3

Proof: We first claim that the objective function $f(g)$ in (12) is convex for $g>2$. To show this, we check the second-order condition $f''(g)>0$, where $f''(g)$ is the second-order derivative of $f(g)$. After some algebraic manipulations, the first and second-order derivatives of $f(g)$ can be written, respectively, as $f'(g)=K^2\ln(\frac{KN}{\epsilon\sqrt{1-\frac{2}{g}}})-\frac{K^2}{g-2}$ and $f''(g)=\frac{2\,K^2}{g(g-2)^2}$. Because $f''(g)>0$ for $g>2$, $f(g)$ is convex.

The optimality condition of (12) can now be described by using the first-order optimality condition $f'(g^{\text{opt}})=0$ as

$$f(g^{\text{opt}})=g^{\text{opt}}K^2/(g^{\text{opt}}-2). \tag{23}$$

Let $\alpha=1-\frac{2}{g^{\text{opt}}}$, equivalently $\frac{1}{g^{\text{opt}}-2}=\frac{1-\alpha}{2\alpha}$. Then, by (12), the equality in (23) can be rewritten as $(\frac{\epsilon}{KN})^2e^{-1}=\frac{1}{\alpha}e^{-\frac{1}{\alpha}}$. This yields $\alpha=-(W_{-1}(-(\frac{\epsilon}{KN})^2e^{-1}))^{-1}$, which follows from the definition of the lower branch of the Lambert W function $W_{-1}(-\frac{1}{\alpha}e^{-\frac{1}{\alpha}})=-\frac{1}{\alpha}$ and $\alpha<1$ [21]. Now, by the equality $g^{\text{opt}}=\frac{2}{1-\alpha}$, we finally have (13). This completes the proof. ■

REFERENCES

[1] Y. Liang, G. Pan, and Z. D. Bai, "Asymptotic performance of mmse receivers for large systems using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4173–4190, Nov. 2007.

[2] R. Menon, P. Gerstoft, and W. S. Hodgkiss, "Asymptotic eigenvalue density of noise covariance matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3415–3424, Jul. 2012.

[3] K. Elkhalil, A. Kammoun, T. Y. Al-Naffouri, and M. Alouini, "Measurement selection: A random matrix theory approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4899–4911, Jul. 2018.

[4] W. Zhang, T. Kim, D. J. Love, and E. Perrins, "Leveraging the restricted isometry property: Improved low-rank subspace decomposition for hybrid millimeter-wave systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5814–5827, Nov. 2018.

[5] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[6] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit: A Brownian motion analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1010–1021, Mar. 2012.

[7] N. Lee, "Map support detection for greedy sparse signal recovery algorithms in compressive sensing," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 4987–4999, Oct. 2016.

[8] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[9] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[10] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[11] N. Jindal, "Mimo broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[12] C. K. Au-yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a miso system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.

[13] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[14] A. Bracher, G. Pope, and C. Studer, "Coherence-based probabilistic recovery guarantees for sparsely corrupted signals," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2012, pp. 307–311.

[15] E. Malhotra, K. Gurumoorthy, and A. Rajwade, "Stronger recovery guarantees for sparse signals exploiting coherence structure in dictionaries," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2017, pp. 6085–6089.

[16] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[17] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2015, pp. 146–150.

[18] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[19] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[20] J. A. Tropp, "An introduction to matrix concentration inequalities," *Found. Trends Mach. Learn.*, vol. 8, no. 1-2, pp. 1–230, 2015.

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Amsterdam, Netherland: Elsevier/Academic, 2007.

[22] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[23] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.

[24] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, Jun. 2003.

[25] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.