Low-Complexity Randomized Iterative Detection For XL-MIMO Communications

Zheng Wang, *Senior Member, IEEE*, Yong Zeng, *Fellow, IEEE*, Yongming Huang, *Fellow, IEEE*, and Rui Zhang, *Fellow, IEEE*

***Abstract*—In this paper, complexity reduction for randomized iterative detection is pursued to facilitate the low-complexity linear detection in uplink extremely large-scale multiple-input multiple-output (XL-MIMO) communications. First, by introducing the step size into randomized iterations and properly selecting its value, we demonstrate that a faster convergence rate can be achieved. Additionally, it is shown that, based on the Taylor expansion, the computational complexity of randomized iterative detection can be significantly reduced with negligible performance loss. As a result, the proposed low-complexity randomized iterative detection (LCRID) algorithm greatly reduces the computational complexity of the baseline randomized iterations from $O(NK^{1.5}l)$ to $O(NKl)$ while achieving better detection performance, where N, K and l respectively indicate the numbers of antennas at the receiver and transmitter sides and the number of iterations. Furthermore, the technique of conditional sampling is applied to LCRID, which not only boosts its convergence rate but also simplifies the iterative operations. Finally, extensive simulation results about XL-MIMO are given to show the superiority of the LCRID algorithm in terms of both detection performance and complexity.**

***Index Terms*—Signal detection, XL-MIMO, complexity reduction, random sampling, iterative detection, randomized iterations.**

I. Introduction

EXTREMELY large-scale multiple-input multiple-output (XL-MIMO) communication has emerged as a promising technique for 6G by greatly improving the network capacity without extra bandwidth cost [1], [2], [3], [4], [5]. However, the rapid increase in system dimensions of XL-MIMO also imposes pressing challenges upon uplink signal detection [6], [7]. For example, given the numbers of the receive and transmit antennas N and K with $N \geq K$, the traditional linear detectors like zero forcing (ZF) and minimum mean-square error (MMSE) require matrix multiplication (complexity order $O(NK^2)$) and matrix inversion (complexity order $O(K^3)$), which become unaffordable with the system dimension rapidly scaling up. To this end, since massive MIMO in 5G, some low-complexity detection algorithms have been given to alleviate the complexity burden for signal detection in MIMO communications [8], [9], [10].

Specifically, regarding complexity reduction for linear detection, the computational operation of matrix inversion with complexity order $O(K^3)$ is firstly considered, such that many iterative detection methods are presented to circumvent the matrix inversion with quadratic complexity $O(K^2 l)$, where l denotes the number of iterations [11], [12], [13]. For example, Neumann series (NS) is applied to approximate the matrix inversion by series expansion, which can be further enhanced by Newton iteration (NI) with a faster convergence performance [14], [15], [16]. In [17], the truncated polynomial expansion (TPE) is reasonably embedded into the operations of linear detections, which leads to a low complexity order $O(NKl)$. Different from NS, NI and TPE, detection schemes based on the traditional iteration methods like Jacobi and Richardson iterations yield the detected signal via matrix splitting rather than series expansion [18], [19], [20], [21]. Subsequently, iterative detections based on Gauss-Seidel (GS) and successive overrelaxation (SOR) iterations are employed for further performance gains [22], [23], [24]. Meanwhile, an iterative block decision-feedback equalizer (IB-DFE) is also given for massive MIMO combined with single-carrier with frequency-domain equalization (SC-FDE), which allows for the incorporation of conventional iterative methods [25]. Some other low-complexity detection schemes based on message passing (MP), conjugate iteration and so on can also be found in [26], [27], [28], [29], [30], [31], [32], [33]. In addition, note that these low-complexity iterative methods can also be applied to precoding for MIMO downlink [34], [35].

Recently, random sampling has become an effective mathematical tool for addressing different challenging problems in MIMO systems [36], [37], [38], [39]. In [40], the randomized iterative detection algorithm (RIDA) is proposed, which leverages the random sampling technique to iteration processes for the extra degree of freedom. More specifically, RIDA

Received 5 April 2025; revised 25 November 2025 and 26 January 2026; accepted 11 February 2026. Date of publication 22 June 2026; date of current version 23 July 2026. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFC2205501; in part by the National Natural Science Foundation of China under Grant 62371124, Grant 62225107, and Grant 61720106003; and in part by the National Science and Technology Major Projects of China 2025ZD1301800. The associate editor coordinating the review of this article and approving it for publication was Cihan Tepedelenlioglu. *(Corresponding authors: Zheng Wang; Yongming Huang.)*

Zheng Wang, Yong Zeng, and Yongming Huang are with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: wznuaa@gmail.com).

Rui Zhang is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: elezhang@nus.edu.sg).

Digital Object Identifier 10.1109/TSP.2026.3706194

not only converges in an exponential manner but also enjoys convergence regardless of any specific conditions (known as *global convergence* because the convergence always holds). This makes the randomized iteration well suited to various scenarios of interest. Based on RIDA, the multistep conditional sampling is then applied, which leads to the modified randomized iterative detection algorithm (MRIDA) with faster convergence and lower complexity [40].

However, most of those low-complexity detection schemes mentioned above only consider the matrix inversion, leaving the complexity reduction of matrix multiplication as an open problem. To be more specific, the computationally expensive matrix multiplication with complexity order $O(NK^2)$ still does exist in those works, but is deemed as a preprocessing operation by default. To overcome this obstacle, new randomized iterative detection algorithm (NRIDA) is presented in [41], which takes the complexity reduction of both matrix multiplication and inversion into account. Typically, by jointly considering the matrix multiplication and inversion, NRIDA iteratively approaches the linear detection performance with complexity cost $O(NK^{1.5}l)$ in total, which eliminates the preprocessing stage of matrix multiplication. Nevertheless, with the explosive growth of antenna numbers, the complexity burden of NRIDA remains significant especially in XL-MIMO communications.

In this paper, we reduce the complexity cost of randomized iterations for linear detection from order $O(NK^{1.5}l)$ to $O(NKl)$, where both the complicated matrix multiplication and inversion are bypassed in a more efficient fashion. Benefiting from it, the complexity hurdle of signal detection in uplink XL-MIMO can be significantly alleviated. In a nutshell, we advance the researches on low-complexity linear detection for both far-field and non-stationary scenarios of XL-MIMO by the following ways.

- Firstly, the step size α is adopted in the randomized iterations for the convergence improvement. According to theoretic analysis, we demonstrate that the randomized iterations aided by step size not only exponentially converges to the target solution of linear detection but also enjoys a faster convergence rate. Meanwhile, the proper choice of α is also investigated in detail for the efficient implementation.
- Secondly, besides the step size, the classic Taylor expansion is also exploited in the randomized iterations for the remarkable complexity reduction with negligible performance loss, which leads to the proposed low-complexity randomized iterative detection (LCRID) algorithm. In particular, the computational complexities of each full iteration (indexed by l) in LCRID for far-field and non-stationary scenarios of XL-MIMO are $2NKt+2NK$, $t\leq 2$ and $(2NK+K)t+2NK+K$, $t\leq 2$ respectively, which reduce the complexity order of randomized iterations from $O(NK^{1.5}l)$ to $O(NKl)$.
- Thirdly, the technique of conditional sampling is employed into the proposed LCRID algorithm, which is further updated via the multistep conditional sampling. By taking the previous multiple sampling results into account, the randomized iteration of LCRID becomes deterministic, which is demonstrated to be greatly beneficial to its convergence and efficiency.

To sum up, in Table I we compare the new contributions to the related literature, where l denotes the iteration index while N and K, $N\geq K$, represent the antenna numbers at the receiver and transmitter, respectively.

The rest of this manuscript is organized as follows. Section II presents the background of linear detections for MIMO communications and describes the developments about efficient iterative detections. In Section III, the proposed LCRID algorithm is introduced with convergence improvement and complexity reduction, where its exponential convergence to the target solution is also demonstrated. In Section IV, by fully making use of multistep conditional sampling, additional convergence and efficiency gains can be attained by LCRID. In Section V, simulation results with respect to the proposed LCRID algorithm for the uplink signal detection in XL-MIMO systems are presented. Finally, Section VI concludes the paper.

Notation: Matrices and column vectors are represented by upper and lowercase boldface letters, and the conjugate transpose, inverse, pseudoinverse of a matrix $\mathbf{B}$ are denoted by $\mathbf{B}^H, \mathbf{B}^{-1}$, and $\mathbf{B}^{\dagger}$, respectively. We use $\mathbf{b}_i$ to denote the i-th column of the matrix $\mathbf{B}$, $b_{i,j}$ indicates the entry in the i-th row and j-th column of the matrix $\mathbf{B}$. $\|\cdot\|_F$ is the Frobenius norm and $\|\cdot\|$ is the Euclidean norm, i.e., $\|\cdot\|_2$, $\mathbf{I}$ is the identity matrix. $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ respectively represent the minimum and maximum eigenvalues of a matrix while $\sigma_{\min}(\cdot)$ and $\sigma_{\max}(\cdot)$ indicate the minimum and maximum singular values of a matrix. Finally, throughout this paper, the computational complexity is measured by the number of complex multiplications.

II. Preliminary

In this section, conventional linear detection methods for MIMO communications are described, followed by a brief review about the efficient detection algorithms based on iterations.

A. Uplink Linear Detection

Consider uplink signal detection for XL-MIMO communications with N antennas at the base station (BS), where user equipments with K antennas are served simultaneously, $N\geq K$ [42], [43], [43], [44]. Let $\mathbf{x}\in\mathcal{X}^K$ denote the transmitted signal with its i-th $(1\leq i\leq K)$ component being selected from the QAM constellation $\mathcal{X}$. Then, at BS the received signal $\mathbf{y}\in\mathbb{C}^N$ can be expressed as

$$\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{n}, \tag{1}$$

where $\mathbf{H}\in\mathbb{C}^{N\times K}$ denotes the channel matrix, $\mathbf{n}\in\mathbb{C}^N$ indicates the AWGN noise with each element following $\mathcal{CN}(0,\sigma^2)$. Clearly, to recover $\mathbf{x}$, the uplink signal detection corresponds to solving the integer least square (ILS) problem shown below

$$\widehat{\mathbf{x}}_{\text{ml}}=\arg\min_{\mathbf{x}\in\mathcal{X}^K}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|^2, \tag{2}$$

which accounts for the solution of maximum likelihood (ML) signal detection [36].

TABLE I
COMPARISON OF DIFFERENT ITERATIVE DETECTION ALGORITHMS

	Complexity of Preprocessing	Complexity of Iterative Detection	Global Convergence
NS [9]	$O(NK^2)$	$O(K^2 \cdot l)$ for $l \leq 2$	×
NI [14]	$O(NK^2)$	$O(K^2 \cdot l)$ for $l \leq 2$	×
Jacobi [19]	$O(NK^2)$	$O(K^2 \cdot l)$	×
Richardson [21]	$O(NK^2)$	$O(K^2 \cdot l)$	×
Gauss Seidel [23]	$O(NK^2)$	$O(K^2 \cdot l)$	√
SOR [22]	$O(NK^2)$	$O(K^2 \cdot l)$	×
RIDA [40]	$O(NK^2)$	$O(K^{2.5} \cdot l)$	√
MRIDA [40]	$O(NK^2)$	$O(K^2 \cdot l)$	√
NRIDA [41]	No need	$O(K^{1.5}N \cdot l)$	√
TPE [17]	No need	$O(KN \cdot l)$	×
LCRID (this work)	No need	$O(KN \cdot l)$	× but achieves superior convergence (better than NRIDA by Corollary 1)

Generally, to address the problem in (2), linear detection computes

$$\widetilde{\mathbf{x}}_{\text{linear}} = \mathbf{G}_{\text{linear}}\mathbf{y}, \tag{3}$$

where the linear equalization matrix $\mathbf{G} \in \mathbb{C}^{K\times N}$ takes the form of

$$\mathbf{G}_{\text{zf}} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H \tag{4}$$

or

$$\mathbf{G}_{\text{mmse}} = \left(\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{H}^H \tag{5}$$

for ZF and MMSE detectors respectively. Then, the final decisions $\widehat{\mathbf{x}}_{\text{linear}}$ are made by quantizing $\widetilde{\mathbf{x}}_{\text{linear}}$ based on the modulation constellation $\mathcal{X}^K$, namely,

$$\widehat{\mathbf{x}}_{\text{linear}} = \lceil\widetilde{\mathbf{x}}_{\text{linear}}\rfloor_Q \in \mathcal{X}^K. \tag{6}$$

Theoretically, linear detection (3) is designed for solving the least square (LS) problem shown below

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{C}^K} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2, \tag{7}$$

which is an approximation of ILS problem. Theoretically, when $N \gg K$ the optimal ML performance by solving the ILS problem in (2) can be obtained by linear detections, thus enabling linear detection widely accepted in large-scale MIMO systems [45].

B. Iterative Detection

The linear detection in (3) can be equivalently formulated as solving a linear equation

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{8}$$

with $\mathbf{b} = \mathbf{H}^H\mathbf{y} \in \mathbb{C}^K$, $\mathbf{A} = \mathbf{H}^H\mathbf{H} \in \mathbb{C}^{K\times K}$. Here, $\mathbf{A}$ is the matrix of ZF filtering where the matrix of MMSE filtering is $\mathbf{A} = \mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}$ accordingly.

To address the linear equation in (8) with low complexity, Neumann series (NS) was firstly applied to approximate the matrix inverse of $\mathbf{A}$ by [9], [16], [27]

$$\mathbf{A}^{-1} = \sum_{l=0}^{\infty}(\mathbf{I} - \mathbf{\Theta}\mathbf{A})^l\mathbf{\Theta}, \tag{9}$$

where $\mathbf{\Theta}$ represents a $K \times K$ diagonal matrix and l is the iteration index. Furthermore, the matrix iterations by NS can be updated by Newton iteration (NI) with faster convergence performance [14], [15].

Different from the matrix iterations based on NS or NI, the matrix splitting $\mathbf{A} = \mathbf{M} + \mathbf{N}$ is used with $\mathbf{N} \in \mathbb{C}^{K\times K}$ and $\mathbf{M} \in \mathbb{C}^{K\times K}$ such that the iterative methods are performed regarding to the detected signal $\mathbf{x}$ by [46]

$$\mathbf{x}^{l+1} = \mathbf{Q}\mathbf{x}^l + \mathbf{g} \tag{10}$$

with $\mathbf{Q} = -\mathbf{M}^{-1}\mathbf{N} = \mathbf{I} - \mathbf{M}^{-1}\mathbf{A} \in \mathbb{C}^{K\times K}$ and $\mathbf{g} = \mathbf{M}^{-1}\mathbf{b} \in \mathbb{C}^K$. Here, $\mathbf{Q}$ is the *iteration matrix*, and the convergence of (10) is ensured when [13]

$$\lim_{l\to\infty}\mathbf{Q}^l = \mathbf{0}, \tag{11}$$

implying that $\mathbf{x}^{l+1}$ will gradually converge to $\widetilde{\mathbf{x}}_{\text{linear}}$.

Undoubtedly, how to choose $\mathbf{M}$ and $\mathbf{N}$ with respect to $\mathbf{A} = \mathbf{M} + \mathbf{N}$ is critical to iterative detections. Typically, in Jacobi method, $\mathbf{M} = \mathbf{D}$ and $\mathbf{N} = \mathbf{L} + \mathbf{U}$ are applied, which result in the iteration expression [31]

$$\mathbf{D}\mathbf{x}^{l+1} = -(\mathbf{U} + \mathbf{L})\mathbf{x}^l + \mathbf{b}. \tag{12}$$

Here, $\mathbf{D} \in \mathbb{C}^{K\times K}$, $\mathbf{U} \in \mathbb{C}^{K\times K}$ and $\mathbf{L} \in \mathbb{C}^{K\times K}$ are respectively the diagonal components, upper triangular components and lower triangular components of $\mathbf{A}$ (i.e., $\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{U}$ and $\mathbf{U} = \mathbf{L}^H$). Regarding to Richardson method, matrices $\mathbf{N} = \mathbf{A} - \frac{1}{\omega}\mathbf{I}$ and $\mathbf{M} = \frac{1}{\omega}\mathbf{I}$ with relaxation factor $\omega > 0$ are selected [15]. To achieve better convergence, GS method with $\mathbf{N} = \mathbf{L}$ and $\mathbf{M} = \mathbf{D} + \mathbf{U}$ works as [23]

$$(\mathbf{D} + \mathbf{U})\mathbf{x}^{l+1} = -\mathbf{L}\mathbf{x}^l + \mathbf{b}. \tag{13}$$

After that, SOR iteration is further proposed to improve the convergence, which adapts a relaxation factor $2 > \omega > 1$ to GS method by [11]

$$(\mathbf{D} + \omega\mathbf{U})\mathbf{x}^{l+1} = [(1-\omega)\mathbf{D} - \omega\mathbf{L}]\mathbf{x}^l + \omega\mathbf{b}. \tag{14}$$

C. Randomized Iterative Detection

Although efficient linear detection performance can be realized via iterative methods, a few necessary requirements (e.g., $N \gg K$) have to be satisfied for ensuring the iteration convergence of NS, NI, Jacobi, Richardson, SOR and so on. This would greatly impede their applications in the cases of interest. Motivated by this, by leveraging the sampling operation into iterations, randomized iterative detection algorithm (RIDA) is given in [40]. It not only entails a lower computational complexity due to the exponential convergence, but also works freely without any restrictions (i.e., global convergence), making it more suitable to various scenarios in practice.

Specifically, RIDA follows the random iteration

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{A}^H\mathbf{S}_k(\mathbf{S}_k^H\mathbf{A}\mathbf{A}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H(\mathbf{b} - \mathbf{A}\mathbf{x}^k) \quad (15)$$

with $\mathbf{x}^0 = \mathbf{D}^{-1}\mathbf{b}$ and the iteration index k, where $\mathbf{S}_k \in \{\mathbf{I}_{:,\mathcal{Q}_1}, \ldots, \mathbf{I}_{:,\mathcal{Q}_r}\}$ is generated randomly according to a designed discrete distribution $\mathcal{D}$, i.e.,

$$p(\mathbf{S}_k = \mathbf{I}_{:,\mathcal{Q}_i}) = \frac{1}{r}. \quad (16)$$

Here, $\mathbf{I}_{:,\mathcal{Q}_i}, 1 \le i \le r$ indicates a column subsection of identity matrix $\mathbf{I}_{K\times K}$, and $\mathcal{Q}_i = \{\text{index } 1, \ldots, \text{index } q_i\} \subseteq \{1, \ldots, K\}$ represents a set of indices with set size $|\mathcal{Q}_i| = q_i$, i.e.,

$$\mathbf{I}_{:,\mathcal{Q}_i} = [\mathbf{I}_{:,\text{index}1}, \ldots, \mathbf{I}_{:,\text{index}q_i}]. \quad (17)$$

As a result, this leads to a block formation, e.g.,

$$\underbrace{\{1,3,7\}}_{\mathcal{Q}_1} \cup \ldots \cup \underbrace{\{2,5,8\}}_{\mathcal{Q}_r} = \{1, \ldots, K\} \quad (18)$$

with $\mathcal{Q}_i \cap \mathcal{Q}_j = \emptyset, 1 \le i \ne j \le r$. More specifically, the indices in $\mathcal{Q}_i$ can be set initially such that the upcoming sampling with $\mathbf{S}_k$ is performed in a block-wise manner (i.e., $\mathbf{I}_{:,\mathcal{Q}_i}$). For the sake of simplicity, the equal block size $|\mathcal{Q}_1| = \ldots = |\mathcal{Q}_r| = K/r = q$ (corresponds to $q_1 = \ldots = q_r = q$) is normally applied. Moreover, from (15), it can be verified that at each iteration indexed by k only q rather than K components of $\mathbf{x}$ are updated. Towards this end, K/q iterations of (15) are grouped as a full iteration indexed by l to facilitate the fair comparison with other iterative detection schemes indexed by l.

Subsequently, based on RIDA, modified randomized iterative detection algorithm (MRIDA) is presented to improve the iteration convergence and efficiency [40]. In particular, the iteration of MRIDA works by

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{S}_k(\mathbf{S}_k^H\mathbf{A}\mathbf{S}_k)^{-1}(\mathbf{S}_k^H\mathbf{b} - \mathbf{S}_k^H\mathbf{A}\mathbf{x}^k), \quad (19)$$

where $\mathbf{S}_k = \mathbf{D}^{-\frac{1}{2}}\mathbf{I}_{:,\mathcal{Q}_i} \in \mathbb{C}^{K\times q}$ and $\mathbf{S}_k \notin \{\mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-r+1}\}$. More precisely, given $1 \le q \le \sqrt{K}$, the complexity order of MRIDA in a full iteration (i.e., indexed by l) is $O(K^2)$, which is competitive in iterative detection algorithms.

However, in both iterative and randomized iterative detections, computing the linear filtering matrix $\mathbf{A} = \mathbf{H}^H\mathbf{H}$ or $\mathbf{A} = \mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}$ as well as $\mathbf{b} = \mathbf{H}^H\mathbf{y} \in \mathbb{C}^K$ is normally viewed as a preprocessing operation, which means the related complexity costs are not counted in the algorithm design. Therefore, in order to alleviate these complexities caused by matrix multiplication, new randomized iterative detection algorithm (NRIDA) obeying the following iteration [41]

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{I}_{:,\mathcal{Q}_i}(\mathbf{H}_{:,\mathcal{Q}_i}^H\mathbf{H}_{:,\mathcal{Q}_i})^{-1}\mathbf{H}_{:,\mathcal{Q}_i}^H(\mathbf{y} - \mathbf{H}\mathbf{x}^k) \quad (20)$$

is proposed, where the index set $\mathcal{Q}_i$ is sampled from the set $\{\mathcal{Q}_1, \ldots, \mathcal{Q}_r\}$ according to

$$p_i \triangleq p(\mathcal{Q}_i) = \frac{1}{r}. \quad (21)$$

Note that similar to RIDA, the indices in set $\mathcal{Q}_i$, $1 \le i \le r$ in NRIDA are determined initially while a simple way is to group these indices in a sequential order, i.e., $\mathbf{H}_{:,\mathcal{Q}_i} = [\mathbf{h}_{(i-1)q+1}, \ldots, \mathbf{h}_{iq}]$, $\mathcal{Q}_i = \{(i-1)q+1, \ldots, iq\}$. Intuitively, given $\mathbf{H}$ and $\mathbf{y}$, NRIDA works without the need of $\mathbf{A}$ and $\mathbf{b}$, which leads to further complexity reduction in total. As illustrated in [41], the computational complexity of the iteration in (20) is $q^3 + qN + 2q^2N + KN$, which corresponds to $K^2 + NK + 3NK^{1.5}$ (when $q = \sqrt{K}$) in a full iteration with complexity order $O(NK^{1.5})$.

III. Low-Complexity Randomized Iterative Detection Algorithm

In this section, based on NRIDA, low-complexity randomized iterative detection (LCRID) algorithm is proposed for XL-MIMO systems to achieve better convergence performance with less complexity cost.

A. Convergence Improvement via Step Size α

Step size plays an important role in iteration-based algorithms to facilitate the convergence performance [46]. As for NRIDA, by adopting the step size $\alpha > 0$, the iteration in (20) becomes

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha\mathbf{I}_{:,\mathcal{Q}_i}(\mathbf{H}_{:,\mathcal{Q}_i}^H\mathbf{H}_{:,\mathcal{Q}_i})^{-1}\mathbf{H}_{:,\mathcal{Q}_i}^H(\mathbf{y} - \mathbf{H}\mathbf{x}^k). \quad (22)$$

From it, we now show that with a proper choice of α the exponential convergence still can be achieved by the randomized iterations in (22) in terms of the norm of the expected error, where the original iteration in (20) is only a special case with $\alpha = 1$.

Here, to simplify the upcoming proof, allow us to make the following definitions

$$\mathbf{Z} \triangleq \mathbf{I}_{:,\mathcal{Q}_i}(\mathbf{H}_{:,\mathcal{Q}_i}^H\mathbf{H}_{:,\mathcal{Q}_i})^{-1}\mathbf{I}_{:,\mathcal{Q}_i}^H \quad (23)$$

and $\mathbf{x}^* = \mathbf{H}^\dagger\mathbf{y} = \mathbf{x}_{\text{zf}}$ denoting the desired detection solution of (7). According to Gram matrix $\mathbf{H}^H\mathbf{H}$, the expectation of $\mathbf{Z}$, i.e., $E[\mathbf{Z}]$, turns out to be positive definite due to[1]

$$\begin{aligned} E[\mathbf{Z}] &= \sum_{i=1}^{r} p_i\mathbf{I}_{:,\mathcal{Q}_i}(\mathbf{H}_{:,\mathcal{Q}_i}^H\mathbf{H}_{:,\mathcal{Q}_i})^{-1}\mathbf{I}_{:,\mathcal{Q}_i}^H \\ &= \frac{1}{r}\left(\sum_{i=1}^{r}\mathbf{I}_{:,\mathcal{Q}_i}(\mathbf{H}_{:,\mathcal{Q}_i}^H\mathbf{H}_{:,\mathcal{Q}_i})^{-\frac{1}{2}}(\mathbf{H}_{:,\mathcal{Q}_i}^H\mathbf{H}_{:,\mathcal{Q}_i})^{-\frac{1}{2}}\mathbf{I}_{:,\mathcal{Q}_i}^H\right) \end{aligned}$$

[1]In MIMO detection, the full rank of $\mathbf{H}$ is generally a default configuration to ensure the unique solution of $\mathbf{x}$. To be specific, any two UEs do not have linear dependent channel vectors, e.g., $\mathbf{h}_i \neq \kappa\mathbf{h}_j$, κ is a constant.

$$
\begin{aligned}
&= \frac{1}{r}(\mathbf{IJ})(\mathbf{JI}^H) \\
&= \frac{1}{r}\mathbf{J}^2
\end{aligned} \tag{24}
$$

with invertible block diagonal matrix $\mathbf{J} \in \mathbb{C}^{K\times K}$, i.e.,

$$
\mathbf{J}{=}\mathrm{diag}((\mathbf{H}^H_{:,\mathcal{Q}_1}\mathbf{H}_{:,\mathcal{Q}_1})^{-\frac{1}{2}},\ldots,(\mathbf{H}^H_{:,\mathcal{Q}_r}\mathbf{H}_{:,\mathcal{Q}_r})^{-\frac{1}{2}}). \tag{25}
$$

As a result, the matrix product $E[\mathbf{Z}]\mathbf{H}^H\mathbf{H}$ is symmetric positive definite as well.

Theorem 1: For uplink XL-MIMO detection, with $0 < \alpha < 2/\lambda_{\max}(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})$, the randomized iteration in (22) converges as

$$
\|E[\mathbf{x}^k - \mathbf{x}^*]\|^2 \le \rho_\alpha^k\|\mathbf{x}^0 - \mathbf{x}^*\|^2, \tag{26}
$$

where its convergence rate follows

$$
\rho_\alpha = \max_i|1 - \alpha\lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})| < 1. \tag{27}
$$

Proof: First of all, the iteration in (22) can be reformulated as

$$
\begin{aligned}
\mathbf{x}^{k+1} &= \mathbf{x}^k + \alpha\mathbf{Z}\mathbf{H}^H(\mathbf{y} - \mathbf{H}\mathbf{x}^k) \\
&\overset{(a)}{=} \mathbf{x}^k + \alpha\mathbf{Z}(\mathbf{H}^H\mathbf{H}\mathbf{H}^\dagger\mathbf{y} - \mathbf{H}^H\mathbf{H}\mathbf{x}^k) \\
&= \mathbf{x}^k + \alpha\mathbf{Z}\mathbf{H}^H\mathbf{H}(\mathbf{x}^* - \mathbf{x}^k),
\end{aligned} \tag{28}
$$

where the equality $\mathbf{H}^H\mathbf{H}\mathbf{H}^\dagger\mathbf{y} = \mathbf{H}^H\mathbf{y}$ is applied in (a). Then, by adding the term $-\mathbf{x}^*$ on the both sides of (28), we can arrive at

$$
\mathbf{x}^k - \mathbf{x}^* = (\mathbf{I} - \alpha\mathbf{Z}\mathbf{H}^H\mathbf{H})(\mathbf{x}^{k-1} - \mathbf{x}^*). \tag{29}
$$

After that, considering the expectation on the both sides of (29), we can have

$$
E[\mathbf{x}^k - \mathbf{x}^*] = (\mathbf{I} - \alpha E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})E[\mathbf{x}^{k-1} - \mathbf{x}^*]. \tag{30}
$$

Moreover, it follows

$$
\begin{aligned}
\|E[\mathbf{x}^k{-}\mathbf{x}^*]\|^2 &= \|(\mathbf{I} - \alpha E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})E[\mathbf{x}^{k-1} - \mathbf{x}^*]\|^2 \\
&\le \|\mathbf{I} - \alpha E[\mathbf{Z}]\mathbf{H}^H\mathbf{H}\|^2\|E[\mathbf{x}^{k-1} - \mathbf{x}^*]\|^2 \\
&= \sigma^2_{\max}(\mathbf{I} - \alpha E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})\|E[\mathbf{x}^{k-1} - \mathbf{x}^*]\|^2 \\
&= \left(\max_i|1 - \alpha\sigma_i^2(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})|\right) \\
&\quad \|E[\mathbf{x}^{k-1} - \mathbf{x}^*]\|^2 \\
&= \left(\max_i|1 - \alpha\lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})|\right) \\
&\quad \|E[\mathbf{x}^{k-1} - \mathbf{x}^*]\|^2.
\end{aligned} \tag{31}
$$

Clearly, letting $\rho_\alpha = \max_i|1 - \alpha\lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})|$, to ensure the convergence, we must fulfill the following condition, i.e.,

$$
\rho_\alpha = \max_i|1 - \alpha\lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})| < 1. \tag{32}
$$

For notational simplicity, here we use $\lambda_{\max}$ and $\lambda_{\min}$ to represent the maximum and minimum eigenvalues of $E[\mathbf{Z}]\mathbf{H}^H\mathbf{H}$. Meanwhile, considering the fact that $\lambda_{\min} > 0$ for the symmetric positive definite matrix, from (32) we can readily attain the following requirements:

$$
1 - \alpha\lambda_{\min} < 1 \text{ and } 1 - \alpha\lambda_{\max} > -1, \tag{33}
$$

which account for

$$
0 < \alpha < \frac{2}{\lambda_{\max}}. \tag{34}
$$

Therefore, with $0 < \alpha < \frac{2}{\lambda_{\max}}$, we can express (31) as

$$
\begin{aligned}
\|E[\mathbf{x}^k - \mathbf{x}^*]\|^2 &\le \rho_\alpha\|E[\mathbf{x}^{k-1} - \mathbf{x}^*]\|^2 \\
&\le \qquad \vdots \\
&\le \rho_\alpha^{k-1}\|E[\mathbf{x}^1 - \mathbf{x}^*]\|^2 \\
&\le \rho_\alpha^k\|\mathbf{x}^0 - \mathbf{x}^*\|^2
\end{aligned} \tag{35}
$$

with $\rho_\alpha < 1$, where $\mathbf{x}^0$ is determined at the very beginning as an initial setup. □

According to Theorem 1, with the proper choice of α, the convergence of the iteration in (22) always holds due to $\rho_\alpha < 1$. Next, given the range $0 < \alpha < 2/\lambda_{\max}(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})$, we discuss the optimal choice of the step size α, i.e., α_{optimal}, to accelerate the convergence performance, namely,

$$
\alpha_{\text{optimal}} = \arg\min_{0<\alpha<\frac{2}{\lambda_{\max}}}\{\max_i|1 - \alpha\lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})|\}. \tag{36}
$$

More specifically, this corresponds to

$$
\min_{0<\alpha<\frac{2}{\lambda_{\max}}}\{\max\{|1 - \alpha\lambda_{\min}|, |1 - \alpha\lambda_{\max}|\}\}, \tag{37}
$$

where α_{optimal} is reached if the following the following coincidence happens [13]

$$
|1 - \alpha\lambda_{\min}| = |1 - \alpha\lambda_{\max}|. \tag{38}
$$

Intuitively, this leads to

$$
\alpha_{\text{optimal}} = \frac{2}{\lambda_{\max} + \lambda_{\min}} \tag{39}
$$

with the optimal convergence rate

$$
\rho_{\text{optimal}} = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}. \tag{40}
$$

Consequently, with the choice of α_{optimal} in (39), we can arrive at the following result, where the introduced convergence gain brought by the step size α can be readily attained.

Corollary 1: With the step size α_{optimal} in (39), further convergence improvement is achieved by the randomized iterations in (22) due to a faster convergence rate

$$
\rho_{\text{optimal}} \le \rho, \tag{41}
$$

where ρ denotes the convergence rate of the original NRIDA with $\alpha = 1$.

Proof: Specifically, we have

$$
\begin{aligned}
\rho_{\text{optimal}} &= \max_i|1 - \alpha_{\text{optimal}}\lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})| \\
&= \min_{0<\alpha<\frac{2}{\lambda_{\max}}}\{\max_i|1 - \alpha\lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})|\} \\
&\le \max_i|1 - \lambda_i(E[\mathbf{Z}]\mathbf{H}^H\mathbf{H})| \\
&= \rho,
\end{aligned} \tag{42}
$$

thus completing the proof. □

B. Complexity Reduction via Taylor Expansion

To start with, the randomized iterations in (22) can be expressed as

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{I}_{:,\mathcal{Q}_i} \mathbf{r}^k \tag{43}$$

with

$$\mathbf{r}^k \triangleq (\mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i})^{-1} \mathbf{b}^k \in \mathbb{C}^q \tag{44}$$

and

$$\mathbf{b}^k = \mathbf{H}^H_{:,\mathcal{Q}_i} (\mathbf{y} - \mathbf{H}\mathbf{x}^k) \in \mathbb{C}^q. \tag{45}$$

Clearly, the computation of the randomized iterations in (43) chiefly lies on calculating the term $\mathbf{r}^k$. To this end, we now introduce the classic Taylor expansion to approximate it in a low complexity way.

In theory, according to Taylor expansion[2]

$$\mathbf{A}^{-1} = \lim_{t\to\infty} \sum_{i=0}^{t} (\mathbf{I} - \mathbf{A})^i, \tag{46}$$

when $\lim_{t\to\infty}(\mathbf{I} - \mathbf{A})^t = \mathbf{0}$, the term $\mathbf{r}^k$ in (44) can be obtained by

$$\mathbf{r}^k = \sum_{t=0}^{\infty} (\mathbf{I} - \mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i})^t \mathbf{b}^k. \tag{47}$$

Here, to enhance the convergence performance of Taylor expansion, an auxiliary $q \times q$ diagonal matrix $\mathbf{D}^k$ is introduced into (46), i.e.,

$$\mathbf{A}^{-1} = \lim_{t\to\infty} \sum_{i=0}^{t} (\mathbf{I} - \mathbf{D}^k \mathbf{A})^i \mathbf{D}^k, \tag{48}$$

so as to

$$\mathbf{r}^k = \sum_{t=0}^{\infty} (\mathbf{I} - \mathbf{D}^k \mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i})^t \mathbf{D}^k \mathbf{b}^k. \tag{49}$$

In principle, to ensure the convergence of (49), the following condition should be satisfied

$$\lim_{t\to\infty} (\mathbf{I} - \mathbf{D}^k \mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i})^t = \mathbf{0}, \tag{50}$$

which essentially accounts for

$$\max_i |1 - \lambda_i(\mathbf{D}^k \mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i})| < 1. \tag{51}$$

Therefore, the choice of the diagonal matrix $\mathbf{D}^k$ should be carefully considered, which are studied in detail for both far-field and non-stationary channels of XL-MIMO (see subsection C and D).

Towards the term $\mathbf{r}^k$ in (49), an iteration can be established to approach it, i.e.,

$$\mathbf{r}_t = \left(\mathbf{I} - \mathbf{D}^k \mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i}\right) \mathbf{r}_{t-1} + \mathbf{D}^k \mathbf{b}^k, \tag{52}$$

so as to $\mathbf{r}_\infty = \mathbf{r}^k$. By doing this, the iteration in (43) naturally becomes

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{I}_{:,\mathcal{Q}_i} \mathbf{r}_t, \tag{53}$$

where the matrix inversion of $\mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i}$ is iteratively approximated along with the index t. Clearly, the computational complexity of $\mathbf{D}^k \mathbf{H}^H_{:,\mathcal{Q}_i} \mathbf{H}_{:,\mathcal{Q}_i} \mathbf{r}_{t-1}$ in (52) is $2Nq + q$.

We then consider the complexity of obtaining the term $\mathbf{b}^k$ in (52). According to (22), in essence, there are only q elements of $\mathbf{x}$ (denoted by $\mathbf{x}_{\mathcal{Q}_j}$) being updated at each iteration indexed by k, namely,

$$\mathbf{x}^k - \mathbf{x}^{k-1} = [0, \ldots, 0, \mathbf{x}^H_{\mathcal{Q}_j}, 0, \ldots, 0]^H. \tag{54}$$

From it, the computation of $\mathbf{b}$ can be expressed by

$$\begin{aligned}\mathbf{b}^k &= \mathbf{H}^H_{:,\mathcal{Q}_i} (\mathbf{y} - \mathbf{H}\mathbf{x}^k) \\ &= \mathbf{H}^H_{:,\mathcal{Q}_i} (\mathbf{y} - \mathbf{H}[\mathbf{x}_{[-\mathcal{Q}_j]}; \mathbf{x}_{\mathcal{Q}_j}]) \\ &= \mathbf{H}^H_{:,\mathcal{Q}_i} \left(\mathbf{y} - \sum_{a \in [-\mathcal{Q}_j]} \mathbf{h}_a x_a - \sum_{b \in \mathcal{Q}_j} \mathbf{h}_b x_b \right),\end{aligned} \tag{55}$$

where $\mathbf{x}_{[-\mathcal{Q}_j]}$ represents the other unchanged $K - q$ elements in the update of $\mathbf{x}^k$. In other words, this means the computation between $\mathbf{H}$ and $\mathbf{x}_{[-\mathcal{Q}_j]}$ (i.e., $\sum_{a \in [-\mathcal{Q}_j]} \mathbf{h}_a x_a$) can be avoided as it has already been calculated in the previous iterations indexed by k. Consequently, the complexity of computing $\mathbf{b}^k$ turns out to be $2Nq$, which corresponds to $2Nq + q$ for obtaining $\mathbf{D}^k \mathbf{b}^k$.

As a result, by approximating $\mathbf{r}$ via the sub-iteration in (52), the computational complexity of the iteration in (53) is $(2Nq + q)t + 2Nq + q$ in total. Clearly, with a mild size of t, it is much smaller than that (i.e., $q^3 + qN + 2q^2N + KN$) of the original NRIDA. For a fair comparison, the full iteration indexed by l that includes K/q times of (53) is still applied, and this corresponds to the complexity $(2NK + K)t + 2NK + K$, which reduces the complexity order of randomized iterations from $O(NK^{1.5}l)$ to $O(NKl)$.

C. Optimizations to Far-Field Channel Model in XL-MIMO

When all the user equipments locate far away from the BS, the far-field channel model should be considered since the assumption of planar wave still holds. In this condition, the channel matrix $\mathbf{H}$ behaves as an extension of that in massive MIMO by obeying the classic Rayleigh fading. Based on it, the following optimizations with respect to both the choices of α and $\mathbf{D}^k$ in the proposed LCRID algorithm are carried out.

On one hand, as for the choice of α, from (39), the bounds of α_{optimal} can be estimated by

$$\frac{1}{\lambda_{\min}} \geq \alpha_{\text{optimal}} \geq \frac{1}{\lambda_{\max}}. \tag{56}$$

Moreover, considering the following relationships

$$\lambda_{\max}(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A}) \lambda_{\max}(\mathbf{B}) \tag{57}$$

and

$$\frac{1}{\lambda_{\min}(\mathbf{AB})} \leq \frac{1}{\lambda_{\min}(\mathbf{A})} \frac{1}{\lambda_{\min}(\mathbf{B})} \tag{58}$$

for any two symmetric positive definite matrices $\mathbf{A}$ and $\mathbf{B}$, the inequality in (56) can be expressed as

$$\frac{1}{\lambda_{\min}(E[\mathbf{Z}])} \frac{1}{\lambda_{\min}(\mathbf{H}^H\mathbf{H})} \geq \alpha_{\text{optimal}} \geq \frac{1}{\lambda_{\max}(E[\mathbf{Z}]) \lambda_{\max}(\mathbf{H}^H\mathbf{H})}. \tag{59}$$

[2]It can also be interpreted by Neumann series (NS) in (9).

To further specify the bounds in (59), the classical *random matrix theory* is applied here. Typically, the smallest and the largest eigenvalues of the Gram matrix $\mathbf{H}^H\mathbf{H}$ regarding to the Gaussian channel $\mathbf{H}\in\mathbb{C}^{N\times K}$ would converge as follows [47]

$$\lambda_{\max}(\mathbf{H}^H\mathbf{H})\to N\left(1+\sqrt{\frac{K}{N}}\right)^2 \tag{60}$$

and

$$\lambda_{\min}(\mathbf{H}^H\mathbf{H})\to N\left(1-\sqrt{\frac{K}{N}}\right)^2, \tag{61}$$

when $N\to\infty$ and $K\to\infty$. On the other hand, because of

$$E[\mathbf{Z}]=\sum_{i=1}^{r}p_i\mathbf{I}_{:,\mathcal{Q}_i}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})^{-1}\mathbf{I}^H_{:,\mathcal{Q}_i}, \tag{62}$$

it follows that

$$\begin{aligned}\lambda_{\max}(E[\mathbf{Z}])&\le\lambda_{\max}(p_1\mathbf{I}_{:,\mathcal{Q}_1}(\mathbf{H}^H_{:,\mathcal{Q}_1}\mathbf{H}_{:,\mathcal{Q}_1})^{-1}\mathbf{I}^H_{:,\mathcal{Q}_1})+\cdots\\&\quad+\lambda_{\max}(p_r\mathbf{I}_{:,\mathcal{Q}_r}(\mathbf{H}^H_{:,\mathcal{Q}_r}\mathbf{H}_{:,\mathcal{Q}_r})^{-1}\mathbf{I}^H_{:,\mathcal{Q}_r})\\&\le\max_{1\le i\le r}\{\lambda_{\max}((\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})^{-1})\}\\&=\max_{1\le i\le r}\{1/\lambda_{\min}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})\}\end{aligned} \tag{63}$$

and

$$\begin{aligned}\lambda_{\min}(E[\mathbf{Z}])&\ge\lambda_{\min}(p_1\mathbf{I}_{:,\mathcal{Q}_1}(\mathbf{H}^H_{:,\mathcal{Q}_1}\mathbf{H}_{:,\mathcal{Q}_1})^{-1}\mathbf{I}^H_{:,\mathcal{Q}_1})+\cdots\\&\quad+\lambda_{\min}(p_r\mathbf{I}_{:,\mathcal{Q}_r}(\mathbf{H}^H_{:,\mathcal{Q}_r}\mathbf{H}_{:,\mathcal{Q}_r})^{-1}\mathbf{I}^H_{:,\mathcal{Q}_r})\\&\ge\min_{1\le i\le r}\{\lambda_{\min}((\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})^{-1})\}\\&=\min_{1\le i\le r}\{1/\lambda_{\max}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})\}.\end{aligned} \tag{64}$$

Moreover, similar to (61) and (60), the following results hold

$$\lambda_{\max}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})\to N\left(1+\sqrt{\frac{q}{N}}\right)^2, \tag{65}$$

$$\lambda_{\min}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})\to N\left(1-\sqrt{\frac{q}{N}}\right)^2 \tag{66}$$

when $N\to\infty$ and $q\to\infty$. Therefore, by combining (63), (64), (65) and (66), the following inequality holds

$$\frac{1}{N(1+\sqrt{\frac{q}{N}})^2}\le\lambda_{\min}(E[\mathbf{Z}])\le\lambda_{\max}(E[\mathbf{Z}])\le\frac{1}{N(1-\sqrt{\frac{q}{N}})^2}. \tag{67}$$

Consequently, by substituting (61), (60) and (67) into (59), the choice of α_{optimal} can be roughly bounded by

$$\frac{\left(1+\sqrt{q/N}\right)^2}{\left(1-\sqrt{K/N}\right)^2}\ge\alpha_{\text{optimal}}\ge\frac{\left(1-\sqrt{q/N}\right)^2}{\left(1+\sqrt{K/N}\right)^2}, \tag{68}$$

thus greatly simplifying the choice of α in practice.

On the other hand, regarding to the choice of $\mathbf{D}^k$ in (49) given the facts in (65) and (66) for Gaussian channels $\mathbf{H}_{:,\mathcal{Q}_i}$, let us set

$$\mathbf{D}^k=\frac{1}{N+q}\mathbf{I}, \tag{69}$$

such that

$$\lambda_{\max}(\mathbf{D}^k\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})\to1+\frac{2\sqrt{Nq}}{N+q}, \tag{70}$$

$$\lambda_{\min}(\mathbf{D}^k\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})\to1-\frac{2\sqrt{Nq}}{N+q} \tag{71}$$

when $N\to\infty$, $q\to\infty$ and $q<N$. Accordingly, the iteration in (52) becomes

$$\mathbf{r}_t=\left(\mathbf{I}-\frac{1}{N+q}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}\right)\mathbf{r}_{t-1}+\frac{1}{N+q}\mathbf{b}^k. \tag{72}$$

Note that other choices of $\mathbf{D}^k$ are also applicable for the convergence improvement (e.g., the diagonal matrix of $\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}$) but extra complexity cost will be incurred.

Clearly, based on (70) and (71), it is straightforward to verify that the convergence requirement in (51) always holds if $N\ge q$. Interestingly, the size q in randomized iterations can be freely adjusted. Generally, considering the choice $1\le q\le K$ in randomized iterations, the condition of $N\gg q$ can be readily fulfilled. Moreover, the value $\frac{2\sqrt{Nq}}{N+q}$ could be very close to 0, which implies that (50) tends to converge very rapidly. For this reason, $t\le2$ is rather sufficient for the approximation of $\mathbf{r}^k$ in (49). In fact, the approximation with $t=1$ is accurate enough, which will be observed in simulations with different system configurations in Section V. Note that with the choice of $\mathbf{D}^k$ in (69), the complexity cost of approximating $\mathbf{r}^k$ by iterations of $\mathbf{r}_t$ in (72) could be further simplified as $2Nqt+2Nq$. Overall, the complexity cost of LCRID under far-field channels is $(2NKt+2NK)l$.

D. Optimizations to Non-Stationary Channel Model in XL-MIMO

Due to the huge number of antennas, the channel in XL-MIMO may exhibit spatial correlation and non-stationarity, such that the non-stationary channel model was developed [43], [48]. Specifically, in non-stationary channel model, the channel of each user is modeled as [34], [35]

$$\mathbf{h}_i=\sqrt{N}\Theta_i^{\frac{1}{2}}\mathbf{z}_i, \tag{73}$$

where $\mathbf{h}_i$ follows $\mathcal{CN}(\mathbf{0},\Theta_i)$, $\mathbf{z}_i$ obeys $\mathcal{CN}(\mathbf{0},\frac{1}{N}\mathbf{I})$, and Θ_i denotes the channel covariance matrix with

$$\Theta_i=\mathbf{F}_i^{\frac{1}{2}}\mathbf{R}_i\mathbf{F}_i^{\frac{1}{2}}. \tag{74}$$

Here, $\mathbf{R}_i\in\mathbb{C}^{N\times N}$ is the spatial correlation matrix and $\mathbf{F}_i\in[0,1]^{N\times N}$ is an indicator diagonal matrix. To be more specific, when the j-th element of $\mathbf{F}_i$ is 1, i.e., $[\mathbf{F}_i]_{j,j}=1$, it means the transmitted signal from user i is received by the antenna j, and vice-versa, which is known as visibility region (VR).

Intuitively, compared to Rayleigh fading channel, the non-stationary channel behaves more complicated. However, even under non-stationary channels, as demonstrated in Theorem 1 that the iteration in (22) always converges, and the convergence improvement introduced by α still can be obtained. On the other hand, given the non-stationary channel $\mathbf{H}$, we have to point out that the related convergence analysis of Taylor expansion is not easy to perform. Nevertheless, effective remedy still does exist to modify the convergence therein, which is described in the following.

Typically, considering the Taylor expansion in (46) with $\mathbf{A} = \mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}$, its related convergence condition turns out to be

$$\max_i |1-\lambda_i(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})| < 1, \tag{75}$$

which corresponds to $\mathbf{D}^k = \mathbf{I}$ in (51). Theoretically, according to (75), a better convergence of Taylor expansion can be achieved by a smaller ratio $\lambda_{\max}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})/\lambda_{\min}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})$. Fortunately, this can be achieved by lowering the dimension size q of the matrix $\mathbf{H}_{:,\mathcal{Q}_i} \in \mathbb{C}^{N\times q}$.

Lemma 1: Given two matrices $\mathbf{H}_{:,\mathcal{Q}_i} \in \mathbb{C}^{N\times q_i}$ and $\mathbf{H}_{:,\mathcal{Q}_j} \in \mathbb{C}^{N\times q_j}$, if $\mathbf{H}_{:,\mathcal{Q}_i}$ is a submatrix of $\mathbf{H}_{:,\mathcal{Q}_j}$ with $q_i < q_j$, then it follows

$$\frac{\lambda_{\max}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})}{\lambda_{\min}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})} \leq \frac{\lambda_{\max}(\mathbf{H}^H_{:,\mathcal{Q}_j}\mathbf{H}_{:,\mathcal{Q}_j})}{\lambda_{\min}(\mathbf{H}^H_{:,\mathcal{Q}_j}\mathbf{H}_{:,\mathcal{Q}_j})}. \tag{76}$$

Proof: To start with, as $\mathbf{H}_{:,\mathcal{Q}_i}$ is a $N \times q_i$ submatrix of $\mathbf{H}_{:,\mathcal{Q}_j} \in \mathbb{C}^{N\times q_j}$, the following relationship about the singular values holds

$$\sigma_{q_j}(\mathbf{H}_{:,\mathcal{Q}_j}) \leq \sigma_{q_i}(\mathbf{H}_{:,\mathcal{Q}_i}) \leq \ldots \leq \sigma_1(\mathbf{H}_{:,\mathcal{Q}_i}) \leq \sigma_1(\mathbf{H}_{:,\mathcal{Q}_j}), \tag{77}$$

where the equality achieves if $\mathbf{H}_{:,\mathcal{Q}_j}$ is an orthogonal matrix. Then, based on it, we have

$$\frac{\sigma_{\max}(\mathbf{H}_{:,\mathcal{Q}_i})}{\sigma_{\min}(\mathbf{H}_{:,\mathcal{Q}_i})} = \frac{\sigma_1(\mathbf{H}_{:,\mathcal{Q}_i})}{\sigma_{q_i}(\mathbf{H}_{:,\mathcal{Q}_i})} \leq \frac{\sigma_1(\mathbf{H}_{:,\mathcal{Q}_j})}{\sigma_{q_j}(\mathbf{H}_{:,\mathcal{Q}_j})} = \frac{\sigma_{\max}(\mathbf{H}_{:,\mathcal{Q}_j})}{\sigma_{\min}(\mathbf{H}_{:,\mathcal{Q}_j})}, \tag{78}$$

which indicates that $\mathbf{H}_{:,\mathcal{Q}_i}$ has a better condition than $\mathbf{H}_{:,\mathcal{Q}_j}$ in terms of *condition number*. From (78), we can arrive at

$$\frac{\lambda_{\max}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})}{\lambda_{\min}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})} \leq \frac{\lambda_{\max}(\mathbf{H}^H_{:,\mathcal{Q}_j}\mathbf{H}_{:,\mathcal{Q}_j})}{\lambda_{\min}(\mathbf{H}^H_{:,\mathcal{Q}_j}\mathbf{H}_{:,\mathcal{Q}_j})} \tag{79}$$

according to $\lambda_i(\mathbf{A}^H\mathbf{A}) = \sigma_i^2(\mathbf{A})$ for matrix $\mathbf{A}$, thus completing the proof. □

According to Lemma 1, a better conditioned matrix of $\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}$ can be realized by simply decreasing q, which recommends us a small q to ensure the convergence of Taylor expansion under non-stationary channels. This also explains the reason why the Taylor expansion about $(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})^{-1}$ enjoys a faster convergence performance than that of $(\mathbf{H}^H\mathbf{H})^{-1}$.

Besides, we claim that the introduced auxiliary diagonal matrix $\mathbf{D}^k \in \mathbb{C}^{q\times q}$ in (48) is still helpful, and we now provide an alternative choice of it. In particular, decompose $\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}$ as

$$\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i} = \mathbf{D}_i + \mathbf{E}_i, \tag{80}$$

where $\mathbf{D}_i \in \mathbb{C}^{q\times q}$ is a diagonal matrix including the diagonal elements of $\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}$, and $\mathbf{E}_i \in \mathbb{C}^{q\times q}$ is a hollow matrix including the off-diagonal elements of $\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}$. Then, based on (80), we set

$$\mathbf{D}^k = \mathbf{D}_i^{-1} \tag{81}$$

with respect to $\mathcal{Q}_i$, which leads to

$$\mathbf{r}_t = \left(\mathbf{I} - \mathbf{D}_i^{-1}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i}\right)\mathbf{r}_{t-1} + \mathbf{D}_i^{-1}\mathbf{b}^k \tag{82}$$

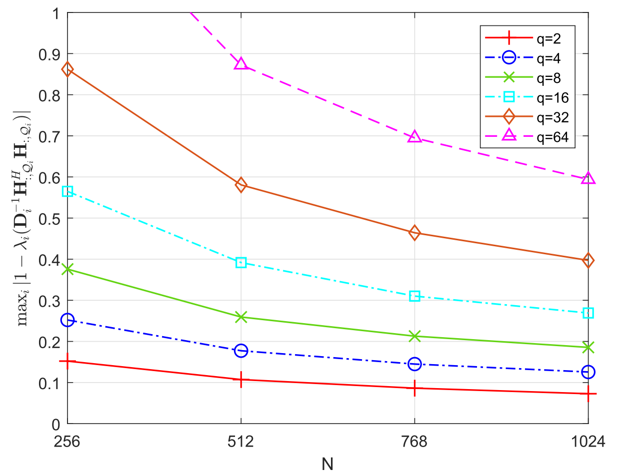

Fig. 1. The comparison of values $\max_i |1-\lambda_i(\mathbf{D}_i^{-1}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})|$ with different sizes of $q = 2, 4, 8, 16, 32, 64$ for $N \times 64$ XL-MIMO systems under non-stationary channels.

with the convergence requirement

$$\max_i |1-\lambda_i(\mathbf{D}_i^{-1}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})| < 1. \tag{83}$$

Here, for a better understanding, following the setup of non-stationary channels in [34], [35], [43], [48], Fig. 1 is presented to illustrate the value $\max_i |1-\lambda_i(\mathbf{D}_i^{-1}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})|$ in (83) in $N \times 64$ XL-MIMO systems by Monte Carlo simulations. As for the visibility region, we assume only half number of antennas at BS contribute each user's connection. Clearly, with fixed N, the value $\max_i |1-\lambda_i(\mathbf{D}_i^{-1}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})|$ decreases gradually with the decrement of q, which leads to the enhanced convergence performance. As demonstrated in Lemma 1, this is because the channel condition of $\mathbf{H}_{:,\mathcal{Q}_i} \in \mathbb{C}^{N\times q}$ is improved by lowering q. Note that the case[3] with $q = 64$ and $N = 256$ fails to satisfy the convergence condition in (83) but such an issue is overcome by smaller sizes of q. On the other hand, one can also observe from Fig. 1 that all values of $\max_i |1-\lambda_i(\mathbf{D}_i^{-1}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})|$ with different block sizes $q = 2, 4, 8, 16, 32, 64$ drop gradually with the increment of the received antenna number N at BS. Intuitively, this also can be attributed by modifying the channel condition $\mathbf{H}_{:,\mathcal{Q}_i} \in \mathbb{C}^{N\times q}$ but with increasing N and fixed q.

Compared to the iteration in (52), extra computational cost (i.e., Nq) is required to obtain the diagonal matrix $\mathbf{D}_i$ for each $\mathcal{Q}_i$, thus resulting in additional complexity NK in total for all cases of $\mathcal{Q}_i, 1 \leq i \leq r$. Note that these computations of $\mathcal{Q}_i$'s can be performed only once as a preprocessing to serve for the upcoming iterations. Overall, the complexity cost of LCRID under non-stationary channels is $[(2NK + K)t + 2NK + K]l + NK$. Interestingly, we can see that the flexible choice of q does not alter the complexity burden of LCRID (indexed by the full iteration), which is in sharp contrast to the traditional randomized iterative detections. Therefore, we recommend a

[3]The case of $q = 64$ can be viewed as an application of Neumann series (NS) since the channel matrix $\mathbf{H}$ is directly considered, i.e., $\max_i |1-\lambda_i(\mathbf{D}_i^{-1}\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})| = \max_i |1-\lambda_i([\text{diag}(\mathbf{H}^H\mathbf{H})]^{-1}\mathbf{H}^H\mathbf{H})|$.

TABLE II
COMPLEXITIES OF THE PROPOSED LCRID ALGORITHM FOR BOTH FAR-FIELD AND NON-STATIONARY XL-MIMO SYSTEMS

	Complexity of Each Full Iteration	Complexity of $\mathbf{D}^k$	Total Complexity
Far-Field	$2NKt + 2NK, t \leq 2$	1	$(2NKt + 2NK) \cdot l + 1, t \leq 2$
Non-Stationary	$(2NK + K)t + 2NK + K, t \leq 2$	NK	$[(2NK+K)t+2NK+K] \cdot l+NK, t \leq 2$

smaller q for LCRID in non-stationary channels than that in far-field channels, which significantly boosts the convergence of Taylor expansion to get a small t. To summarize, the complexity costs of the proposed LCRID algorithm for both far-field and non-stationary channels are illustrated in Table II.

Additionally, two aspects about the complexity of LCRID should be pointed out. On one hand, due to the rapid convergence of the introduced Taylor expansion, $t = 1$ is sufficient to ensure the approximation accuracy of $\mathbf{r}^k$ in (49), which can be verified in both far-field and non-stationary XL-MIMO systems by simulations results in Section V. For this reason, $t = 1$ is recommended to LCRID in practice. On the other hand, the complexity of sampling operations with respect to the set $\mathcal{Q}_i, 1 \leq i \leq r$ is omitted here, and the reason of it is two-fold. Firstly, since the uniform sampling distribution is employed, the sampling cost of $\mathcal{Q}_i$ is trivial. Secondly, by introducing the multistep conditional sampling into LCRID, we show that the sampling process of randomized iterations can be removed but with improved convergence and efficiency, which is fully described in the following.

IV. Convergence Enhancement and Efficiency Improvement

In this section, motivated by the technique of conditional sampling, the sampling probability about the index set $\mathcal{Q}_i$ at each iteration of LCRID is investigated for further convergence and efficiency improvements for XL-MIMO detection.

A. Enhancement by Conditional Sampling

Here, to better serve for the illustration of convergence analysis, we claim that the following theoretical derivations regarding to the proposed LCRID algorithm are carried out with respect to the randomized iteration in (22) rather than the one in (53). This is not difficult to understand since the latter serves as an approximation of the former but with lower complexity.

Typically, different from the sampling probability p_i in (21), the probability $\overline{p}_i$ of conditional sampling in LCRID is defined as

$$\begin{aligned}\overline{p}_i &\triangleq p(\mathcal{Q}_i|\text{sampling result } \mathcal{Q}_j \text{ at last iteration}) \\ &= \frac{p_i}{1-p_j},\ i \neq j,\end{aligned} \tag{84}$$

such that the previous sampling result $\mathcal{Q}_j$ (e.g., at the $(k-1)$-th iteration) is removed from the sampling candidate list of $\mathcal{Q}_i$ at the current iteration (e.g., at the k-th iteration). In this way, it is clear to see that the sampling repetitions over two consecutive iterations are effectively avoided. Since the sampling diversity is improved, extra convergence gain can be successfully obtained thereafter.

Theoretically, let $E[\mathbf{Z}|\mathcal{Q}_j]$ denote the expectation of the matrix $\mathbf{Z}$ with conditional sampling probability $\overline{p}_i$ in (84), the exponential convergence of the random iterations in LCRID still can be demonstrated, where the proof is omitted here due to the similarity with Theorem 1.

Theorem 2: For uplink XL-MIMO detection, according to the conditional sampling probability $\overline{p}_i$ in (84), then with $0 < \alpha < 2/\lambda_{\max}(E[\mathbf{Z}|\mathcal{Q}_j]\mathbf{H}^H\mathbf{H})$ the randomized iteration in (22) converges as

$$\|E[\mathbf{x}^k - \mathbf{x}^*]\|^2 \leq \overline{\rho}_\alpha \|\mathbf{x}^{k-1} - \mathbf{x}^*\|^2, \tag{85}$$

where its convergence rate follows

$$\overline{\rho}_\alpha = \max_i |1 - \alpha\lambda_i(E[\mathbf{Z}|\mathcal{Q}_j]\mathbf{H}^H\mathbf{H})| < 1. \tag{86}$$

Subsequently, let $\overline{\lambda}_{\text{max}}$ and $\overline{\lambda}_{\text{min}}$ denote the maximum and minimum eigenvalues of $E[\mathbf{Z}|\mathcal{Q}_j]\mathbf{H}^H\mathbf{H}$, the optimal choice of α turns out to be

$$\alpha_{\text{optimal}} = \frac{2}{\overline{\lambda}_{\text{max}} + \overline{\lambda}_{\text{min}}}. \tag{87}$$

From it, the convergence gain brought by the usage of conditional sampling can be confirmed in what follows.

Corollary 2: With sampling probability $\overline{p}_i$ in (84), the randomized iterations based on (22) achieves an improved convergence than that with p_i in (21) by a smaller upper bound of convergence rate.

Proof: First of all, given the optimal choice α_{optimal} in (87), the optimal convergence rate $\overline{\rho}_{\text{optimal}}$ can be obtain as

$$\overline{\rho}_{\text{optimal}} = \frac{\overline{\lambda}_{\text{max}} - \overline{\lambda}_{\text{min}}}{\overline{\lambda}_{\text{max}} + \overline{\lambda}_{\text{min}}} = 1 - \frac{2}{\overline{\lambda}_{\text{max}}/\overline{\lambda}_{\text{min}} + 1} \propto \frac{\overline{\lambda}_{\text{max}}}{\overline{\lambda}_{\text{min}}}. \tag{88}$$

Therefore, the comparison between convergence $\overline{\rho}_{\text{optimal}}$ in (88) and ρ_{optimal} in (40) can be performed with respect to $\overline{\lambda}_{\max}/\overline{\lambda}_{\min}$ and $\lambda_{\max}/\lambda_{\min}$.

On the other hand, for any two symmetric positive definite matrices $\mathbf{A}$ and $\mathbf{B}$, the following inequalities hold [49]

$$\lambda_{\max}(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})\lambda_{\max}(\mathbf{B}), \tag{89}$$

$$\lambda_{\min}(\mathbf{AB}) \geq \lambda_{\min}(\mathbf{A})\lambda_{\min}(\mathbf{B}). \tag{90}$$

Then, from these, we can arrive at

$$\frac{\overline{\lambda}_{\text{max}}}{\overline{\lambda}_{\text{min}}} \leq \frac{\lambda_{\max}(E[\mathbf{Z}|\mathcal{Q}_j])\lambda_{\max}(\mathbf{H}^H\mathbf{H})}{\lambda_{\min}(E[\mathbf{Z}|\mathcal{Q}_j])\lambda_{\min}(\mathbf{H}^H\mathbf{H})} \tag{91}$$

and

$$\frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \leq \frac{\lambda_{\max}(E[\mathbf{Z}])\lambda_{\max}(\mathbf{H}^H\mathbf{H})}{\lambda_{\min}(E[\mathbf{Z}])\lambda_{\min}(\mathbf{H}^H\mathbf{H})}. \tag{92}$$

Next, as for $E[\mathbf{Z}|\mathcal{Q}_j]$, it follows that

$$E[\mathbf{Z}|\mathcal{Q}_j] = \sum_{i\neq j} \overline{p}_i \mathbf{I}_{:,\mathcal{Q}_i}(\mathbf{H}^H_{:,\mathcal{Q}_i}\mathbf{H}_{:,\mathcal{Q}_i})^{-1}\mathbf{I}^H_{:,\mathcal{Q}_i} = \frac{1}{r-1}\overline{\mathbf{J}}^2 \tag{93}$$

with invertible block diagonal matrix $\overline{\mathbf{J}} \in \mathbb{C}^{K\times K}$, i.e.,

$$\begin{aligned}\overline{\mathbf{J}} = \mathrm{diag}\big(&(\mathbf{H}^H_{:,\mathcal{Q}_1}\mathbf{H}_{:,\mathcal{Q}_1})^{-\frac{1}{2}}, \ldots, (\mathbf{H}^H_{:,\mathcal{Q}_{j-1}}\mathbf{H}_{:,\mathcal{Q}_{j-1}})^{-\frac{1}{2}}, \mathbf{0}_{q\times q}, \qquad (1)\\ &(\mathbf{H}^H_{:,\mathcal{Q}_{j+1}}\mathbf{H}_{:,\mathcal{Q}_{j+1}})^{-\frac{1}{2}}, \ldots, (\mathbf{H}^H_{:,\mathcal{Q}_r}\mathbf{H}_{:,\mathcal{Q}_r})^{-\frac{1}{2}}\big).\end{aligned} \tag{94}$$

Clearly, compared to the matrix $\mathbf{J}^2$ in (24), the matrix $\overline{\mathbf{J}}^2$ zeros the corresponding part related to $(\mathbf{H}^H_{:,\mathcal{Q}_j}\mathbf{H}_{:,\mathcal{Q}_j})^{-1}$. From it, the following relationships hold

$$\lambda_{\max}(\mathbf{J}^2) \geq \lambda_{\max}(\overline{\mathbf{J}}^2) \text{ and } \lambda_{\min}(\mathbf{J}^2) \leq \lambda_{\min}(\overline{\mathbf{J}}^2), \tag{95}$$

which further corresponds to

$$\frac{\lambda_{\max}(E[\mathbf{Z}|\mathcal{Q}_j])}{\lambda_{\min}(E[\mathbf{Z}|\mathcal{Q}_j])} \leq \frac{\lambda_{\max}(E[\mathbf{Z}])}{\lambda_{\min}(E[\mathbf{Z}])}. \tag{96}$$

Finally, by simply substituting (96) into (91) and (92), it is straightforward to see that $\overline{\lambda}_{\max}/\overline{\lambda}_{\min}$ achieves a smaller upper bound than $\lambda_{\max}/\lambda_{\min}$, so as to the comparison result between the upper bounds of $\overline{\rho}_{\text{optimal}}$ and ρ_{optimal}. □

In addition, according to (91) and (92), we can also observe that a small condition number of matrix $\mathbf{H}$ is naturally beneficial to the convergence of LCRID by enabling a smaller value of $\overline{\rho}_{\text{optimal}}$ or ρ_{optimal}.

B. Extension to Multistep Conditional Sampling

In order to speed up the convergence by conditional sampling, we commence to consider more previous results into the sampling, which leads to the multistep conditional sampling in LCRID as

$$\overline{p}^f_i \triangleq p(\mathcal{Q}_i|\text{sampling results of } \mathcal{Q} \text{ in last } f \text{ iterations}) \tag{97}$$

where $\mathcal{Q}_i$ differs from all the sampling results in the last f iterations. Here, for the sake of notational simplicity, we describe this relationship as follows

$$\mathcal{Q}^k = \mathcal{Q}_i \notin \{\mathcal{Q}^{k-1}, \ldots, \mathcal{Q}^{k-f}\}, \tag{98}$$

such that the probability $\overline{p}^f_i$ of multistep conditional sampling can be rewritten as

$$\overline{p}^f_i = p(\mathcal{Q}^k = \mathcal{Q}_i|\mathcal{Q}^{k-1}, \ldots, \mathcal{Q}^{k-f}) \tag{99}$$

with $\mathcal{Q}_i \notin \{\mathcal{Q}^{k-1}, \ldots, \mathcal{Q}^{k-f}\}$. From it, we can see that $1 \leq f \leq r-1$ actually represents the number of the involved previous steps while $\overline{p}_i$ in (84) turns out to be a case of $\overline{p}^f_i$ by $f = 1$.

Here, let $E[\mathbf{Z}|\mathcal{Q}^{k-1}, \ldots, \mathcal{Q}^{k-f}]$ represent the expectation of the matrix $\mathbf{Z}$ with $\overline{p}^f_i$ in (99), we can get the following convergence result, where the proof details are ignored because of similarity.

Theorem 3: For uplink XL-MIMO detection, according to multistep conditional sampling probability $\overline{p}^f_i$ in (99), then with $0 < \alpha < 2/\lambda_{\max}(E[\mathbf{Z}|\mathcal{Q}^{k-1}, \ldots, \mathcal{Q}^{k-f}]\mathbf{H}^H\mathbf{H})$ the randomized iteration in (22) converges as

$$\|E[\mathbf{x}^k - \mathbf{x}^*]\|^2 \leq \overline{\rho}^f_\alpha\|\mathbf{x}^{k-1} - \mathbf{x}^*\|^2, \tag{100}$$

where its convergence rate follows

$$\overline{\rho}^f_\alpha = \max_i |1 - \alpha\lambda_i(E[\mathbf{Z}|\mathcal{Q}^{k-1}, \ldots, \mathcal{Q}^{k-f}]\mathbf{H}^H\mathbf{H})| < 1. \tag{101}$$

Note that the original probability p_i in (21) works as a case of $\overline{p}^f_i$ by $f = 0$, which extends multistep conditional sampling as $0 \leq f \leq r-1$. Then, by simply extending Corollary 2, we can get the following result, where the related proof is omitted for similarity.

Corollary 3: With the increase of $0 \leq f \leq r-1$, the convergence of the randomized iterations in (22) with $\overline{p}^f_i$ in (99) improves because of the decreasing convergence rate upper bound.

From Corollary 3, the best convergence performance is obtained by the proposed LCRID algorithm when $f = r-1$ is applied. More interestingly, with $f = r-1$, there is only one choice for the current sampling $\mathcal{Q}^k$ when $k > r-1$, thus rendering the sampling operation about $\mathcal{Q}^k$ deterministic. Intuitively, this is significantly helpful to the proposed LCRID by avoiding the random sampling without degrading the performance, resulting in a better choice of LCRID. Hence, $\overline{p}^f_i$ with $f = r-1$ is greatly desired in LCRID for better convergence and efficiency, which replaces p_i in (21) with $\overline{p}^f_i$ in (99). For this reason, the complexity of sampling operations has no need to be taken into count in the proposed LCRID algorithm. To summarize, the proposed LCRID algorithm for XL-MIMO systems is shown in Algorithm 1.

Algorithm 1: LCRID Algorithm for XL-MIMO Systems.

Require: $\mathbf{H}, \mathbf{y}, \mathbf{x}^0 = \mathbf{0}, \alpha, q, L$
Ensure: detection results $\widehat{\mathbf{x}}$
for $l = 1, \ldots, L$ **do**
    for $k = 1, \ldots, K/q$ **do**
        obtain $\mathcal{Q}_i$ by $\overline{p}^{r-1}_i$ in (99)
        compute $\mathbf{b}_k$ based on (55)
        initialize $\mathbf{r}_0 = \frac{1}{N+q}\mathbf{b}^k$ or $\mathbf{r}_0 = \mathbf{D}^{-1}_k\mathbf{b}^k$
        for $t = 1, 2$ **do**
            update $\mathbf{r}_t$ according to (72) or (82)
        end for
        update $\mathbf{x}$ based on (53)
    end for
end for
output $\widehat{\mathbf{x}} = \lceil \mathbf{x}^L \rfloor_Q \in \mathcal{X}^K$

V. Simulations Results

In this section, the proposed LCRID algorithm for XL-MIMO systems is examined in detail by simulations, where both far-field (Figs. 2–8) and non-stationary (Figs. 9–13) scenarios of XL-MIMO are considered. For the sake of fair comparison, all the randomized iterations (except RIDA) are implemented under sampling probability $\overline{p}^f_i$ with $f = r-1$, $q = 32$

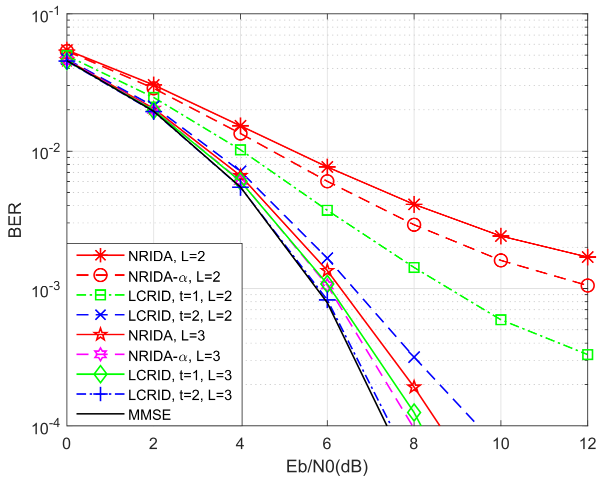

Fig. 2. BER performance comparison for 256×64 far-field XL-MIMO using 16-QAM.

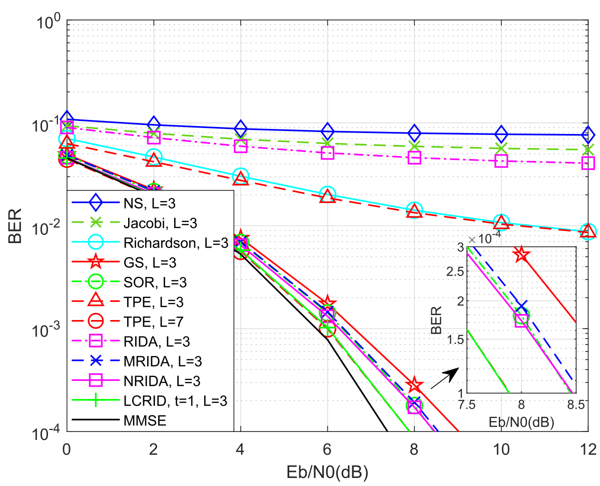

Fig. 3. BER performance comparison for 256×64 far-field XL-MIMO using 16-QAM.

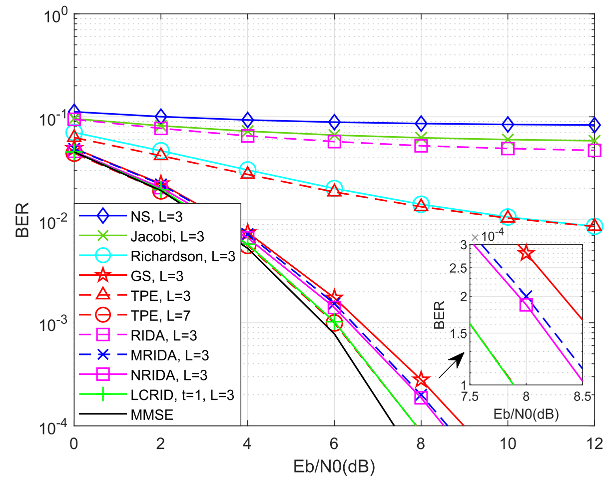

Fig. 4. BER performance comparison for 512×128 far-field XL-MIMO using 16-QAM.

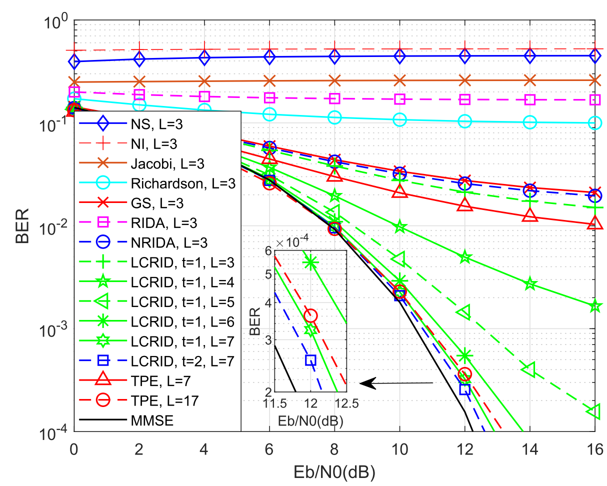

Fig. 5. BER performance comparison for 512×256 far-field XL-MIMO using 16-QAM.

(far-field) and $q = 16$ (non-stationary), where $l = L$ denotes the number of full iterations. Meanwhile, according to (68) and (87), the step size α in LCRID is set to 1.1 as a suboptimal but much efficient solution. Additionally, note that in non-stationary channels only half number of antennas at BS are assumed to contribute each user's connection due to visibility region.

In Fig. 2, the bit error rate (BER) detection performance of LCRID algorithm is illustrated in 256×64 far-field XL-MIMO with 16-QAM. Here, NRIDA-α denotes the step size added-randomized iterations given in Section III-A, which essentially corresponds to LCRID but without the complexity reduction via Taylor series. As can be seen from Fig. 2, the detection performance of all the randomized iterations improves gradually with the increase of full iterations, which is in accordance with the convergence results given in Theorems 1 and 3. To be precise, due to the introduced step size α, the BER detection performance of NRIDA-α outperforms the original NRIDA. Another interesting point is that LCRID is able to achieve better detection performance than NRIDA-α. This is because in both NRIDA and NRIDA-α the block size q is set as 8 to balance the performance and complexity, while in LCRID a large size q can be set freely (here use $q = 32$) to enhance the convergence without altering the complexity cost. Therefore, in this case, extra performance gain can be introduced by the larger q in LCRID.

As a counterpart of Figs. 2 and 3 is given to illustrate the performance comparison between LCRID and other low complexity iterative detection schemes in 256×64 far-field XL-MIMO using 16-QAM. Different from MMSE detection, the conventional iterative detections like NS in [9], Richardson method in [21], Jacobi method in [19], successive over-relaxation (SOR) method in [22], Gauss Seidel method in [23], TPE method in [17], are employed. Besides, for a better comparison, randomized iterative detection algorithms such as RIDA and MRIDA in [40], and NRIDA in [41] are also considered. Clearly, under

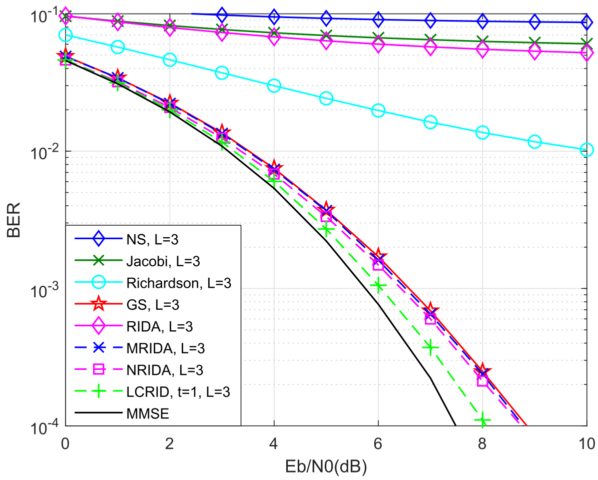

Fig. 6. BER performance comparison for 1024×256 far-field XL-MIMO using 16-QAM.

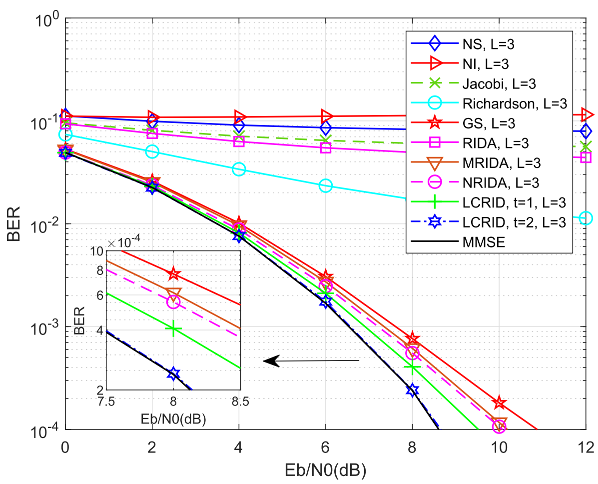

Fig. 7. BER performance comparison for 256×64 far-field XL-MIMO with the imperfect CSI using 16-QAM.

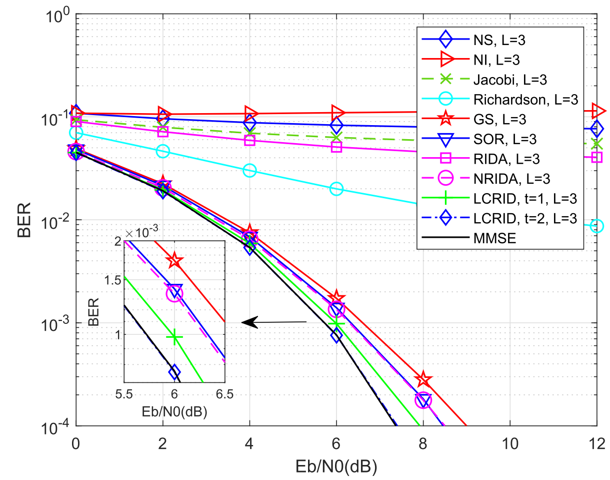

Fig. 8. BER performance comparison for 256×64 far-field XL-MIMO using 16-QAM with common pilot based channel estimation.

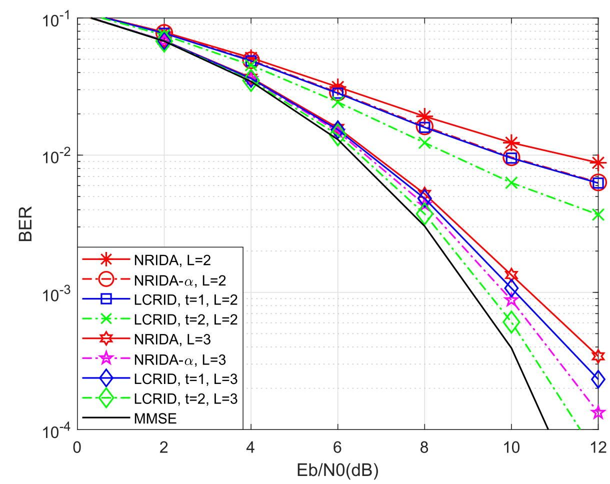

Fig. 9. BER performance comparison for 256×64 non-stationary XL-MIMO using 16-QAM.

the same number of iterations with $L = 3$, the proposed LCRID algorithm even with $t = 1$ achieves the best detection BER performance among them. Here, besides $L = 3$, the performance curve of TPE detector in [17] with $L = 7$ is also given for a comparison. As the complicated matrix operation is also avoided in TPE, more iterations can be recalled to improve its performance accordingly with complexity cost $(2L - 1)NK + LK$.

Fig. 4 presents the BER performance comparison about the proposed LCRID algorithm in a 512×128 far-field XL-MIMO system with 16-QAM. As expected, the performance superiority of LCRID compared to other detection schemes can be easily confirmed. Note that detection schemes like NI fail to work in this case. Furthermore, we update the antenna ratio as $N/K = 2$ in Fig. 5. As can be seen clearly, all the BER detection performance of those iterative detection schemes degrade accordingly. This is straightforward to interpret because the condition number of the channel matrix becomes larger with the change of the antenna ratio. In other words, given the constant N, the columns of $\mathbf{H}$ are getting more correlated to each other with the increment of K. In this condition, detection schemes like NS, NI, Jacobi all fail to converge unfortunately while randomized iterations like MRIDA, NRIDA and LCRID still work but with slower convergence performance. Nevertheless, we can see that the low complexity linear detection performance still can be realized via LCRID by improving the number of iterations.

Here, based on Figs. 3–5, a complexity comparison between the proposed LCRID algorithm and TPE detector in [17] is presented. We claim that other iterative schemes are excluded as they have high complexity order than both of LCRID and TPE (i.e., $O(NK \cdot l)$). Specifically, for a fair comparison, the performance curves of TPE detector with $L = 7$, $L = 7$ and $L = 17$ have also plotted in Figs. 3–5 respectively to obtain the comparable performance as LCRID. Clearly, in all cases, given the similar BER performance, the proposed LCRID algorithm requires less complexity costs than TPE detector, which paves

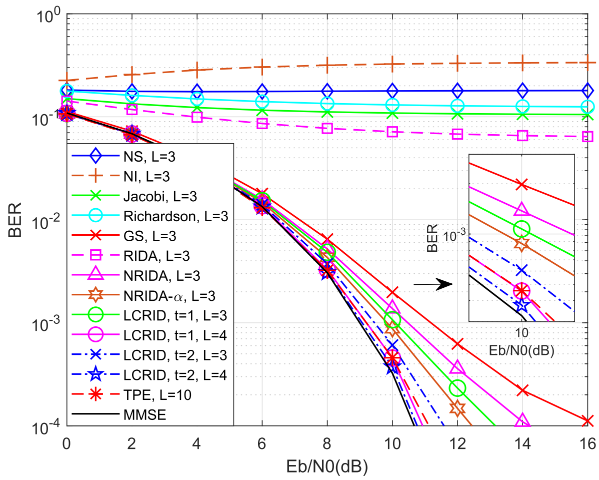

Fig. 10. BER Performance comparison for 256×64 non-stationary XL-MIMO using 16-QAM.

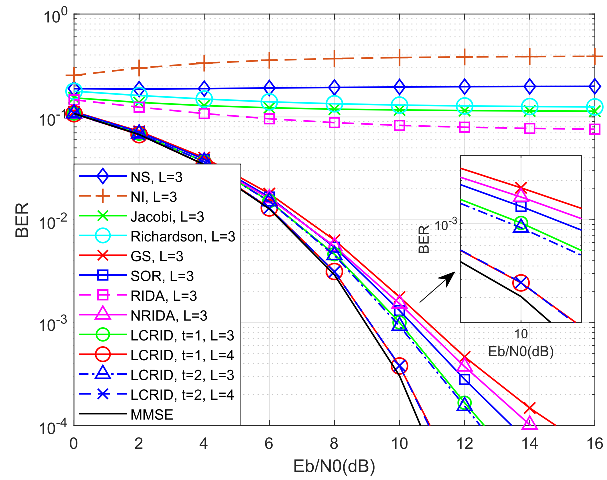

Fig. 12. BER performance comparison for 1024×256 non-stationary XL-MIMO using 16-QAM.

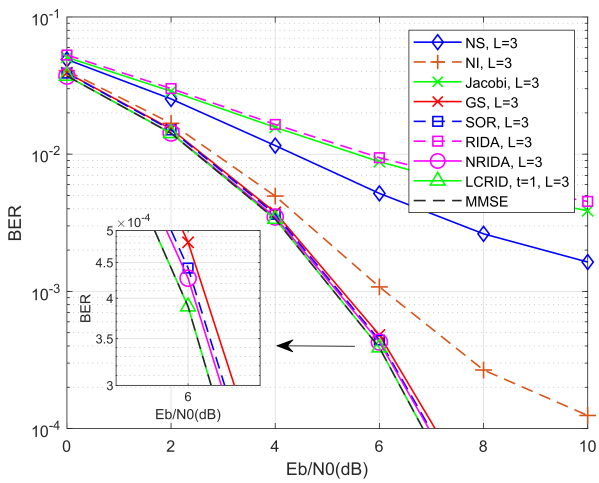

Fig. 11. BER performance comparison for 512×64 non-stationary XL-MIMO using 16-QAM.

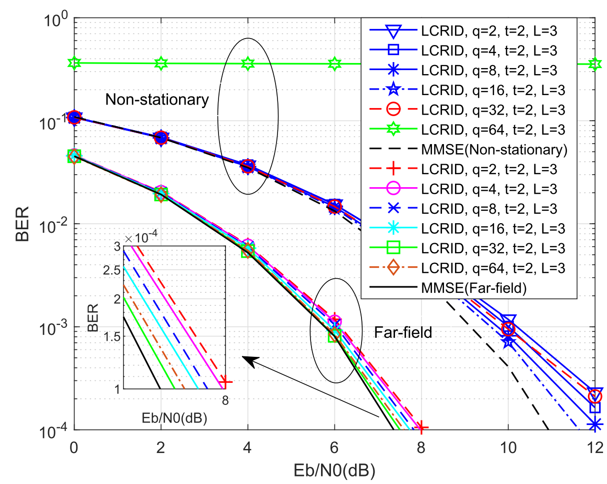

Fig. 13. BER performance comparison for 256×64 far-field and non-stationary XL-MIMO systems using 16-QAM.

TABLE III
COMPLEXITY COMPARISON OF LCRID AND TPE IN [17]

	256×64	512×128	512×256
LCRID, $t = 1$	$L = 3$: 196608	$L = 3$: 786432	$L = 7$: 3670016
TPE [17]	$L = 7$: 213440	$L = 7$: 852864	$L = 17$: 4329728

the way for the development of randomized iterations in XL-MIMO systems as shown in Table III.

In Fig. 6, the comparison of BER detection performance about the proposed LCRID algorithm in a 1024×256 far-field XL-MIMO system with 16-QAM is presented. Undoubtedly, LCRID greatly outperforms other iterative detection schemes with the best detection performance and the lowest computational complexity. Because of such an incomparable detection trade-off between performance and complexity, LCRID turns out to be a perspective choice to the uplink signal detection of XL-MIMO systems. Note that in all these cases of XL-MIMO, the convergence of LCRID always can be guaranteed with the proper choice of step size α. More specifically, this is in line with the derived results given in Theorems 1 and 3, which also makes LCRID well suited to different scenarios of XL-MIMO systems.

Fig. 7 is presented to illustrate the BER performance of the proposed LCRID without the perfect channel state information (CSI) in 256×64 far-field XL-MIMO systems using 16-QAM. In particular, $\widehat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$ denotes imperfect CSI at the receiver, where $\Delta\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2 \mathbf{I}_N)$ represents the channel estimation errors with $\sigma_e^2 = \frac{K}{n_p \cdot E_p}$ [50]. Here, n_p and E_p respectively stand for the number and power of pilot symbols, and we apply $\sigma_e^2 = 0.1$ in the simulation. Compared to the detection performance under perfect CSI in Fig. 3, the BER performance of detection schemes with imperfect CSI decline accordingly in Fig. 7. Nevertheless, the performance gain of LCRID still can be verified. Furthermore, in Fig. 8, the BER performance comparison based on the common pilot channel estimation is shown with respect to 256×64 far-field XL-MIMO systems using 16-QAM. In particular, we consider a pilot matrix $\mathbf{X}$ composed of K columns of the $\tau \times \tau$ discrete Fourier transform (DFT) operator and we set $\tau = 64$. Based on it, the model of channel estimation is formulated as $\mathbf{Y}_p = \mathbf{H}\mathbf{X} + \mathbf{W}_p$ where $\mathbf{W}_p$ and $\mathbf{Y}_p$ denote the noise matrix and the receive signal matrix during the training period respectively. Then, the classic least-squares (LS) channel estimate $\widehat{\mathbf{H}} = \mathbf{Y}_p\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}$ is applied as the imperfect channel state information (CSI) with the estimation errors $\Delta\mathbf{H} = \widehat{\mathbf{H}} - \mathbf{H} = \mathbf{W}_p\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}$ [51], [52], [53]. Similar to the imperfect CSI case in Fig. 7, the BER performance of all detection schemes in Fig. 8 are degraded compared to the perfect case in Fig. 3. This is easy to understand as the errors $\Delta\mathbf{H}$ in channel estimation are taken into account.

In Fig. 9, the bit error rate (BER) detection performance of the proposed LCRID algorithm is presented in 256×64 non-stationary XL-MIMO with 16-QAM. As can be seen clearly, the detection performance of all the randomized iterations improve gradually with the increase of full iterations, which is in accordance with the convergence results given in Theorems 1 and 3. To be precise, due to the introduced step size α, the BER detection performance of NRIDA-α outperforms the original NRIDA. Nevertheless, as pointed out previously, the approximation driven by Taylor series converges rapidly, such performance of NRIDA-α can be well approximated by LCRID. Moreover, LCRID could also achieve a better detection performance than NRIDA-α by simply increasing the block size q, so that more convergence gain can be introduced.

As a counterpart of Figs. 9 and 10 is given to show the performance comparison between LCRID and other low complexity iterative detection schemes in 256×64 non-stationary XL-MIMO using 16-QAM. Compared to the far-field scenario shown in Fig. 3, the detection performance of all schemes degrade accordingly in non-stationary XL-MIMO. This is intuitive to understand as the channel model in non-stationary becomes more complicated with inferior channel condition. Note that similar to the case in Fig. 3, detection schemes like NS, NI, Jacobi, Richardson still fail to work. Clearly, under the same number of iterations with $L = 3$, the proposed LCRID algorithm even with $t = 1$ achieves competitive detection BER performance, where MMSE performance can be obtained by LCRID with $L = 4$. For case $L = 4$, the performance gap of LCRID between $t = 1$ and $t = 2$ is negligible, implying LCRID with $t = 1$ (corresponds to complexity $(4NK + 2K)L + NK$) a better choice in practice. Additionally, here the BER performance of TPE detector in [17] is also added. In particular, to obtain the same performance, LCRID $t = 1$ with $L = 4$ (i.e., $17NK + 8K$) requires less complexity cost than TPE detector with $L = 10$ (i.e., $19NK + 10K$).

In Fig. 11, the comparison of BER performance about LCRID is extended to 512×64 non-stationary XL-MIMO systems with 16-QAM. Clearly, compared to the case 256×64, the antenna ratio N/K in case of 512×64 increases. This generally enables a better channel condition, which relaxes the convergence requirements upon some iterative detection schemes. To this end, we can see that schemes like NS, NI and Jacobi work normally. On the other hand, due to the extra receive diversity gain brought by the higher antenna ratio N/K, a smaller condition number of $\mathbf{H}$ can be realized. As studied before, this is helpful to the convergence performance of randomized iterations, such that the performance gap between them and MMSE becomes smaller than that in case of 256×64 under the same iteration number. Note that MMSE detection performance can be achieved by the proposed LCRID algorithm with $t = 1$ and $L = 3$.

In Fig. 12, the comparison of BER detection performance about LCRID in a 1024×256 non-stationary XL-MIMO system with 16-QAM is given. Undoubtedly, LCRID outperforms other iterative detection schemes with the best detection performance and the lowest computational complexity. Because of such a competitive detection trade-off between performance and complexity, LCRID turns out to be a prospective choice to the uplink signal detection of XL-MIMO systems. Besides, given the choice of $\mathbf{D}_k$ in (81), by setting the size $q = 16$ to facilitate the convergence of Taylor expansion, we can see that there is negligible performance gap between cases of $t = 1$ and $t = 2$ of LCRID, making LCRID with $t = 1$ more preferred in practice.

Finally, in Fig. 13, the choice of the block size q in LCRID is studied in full details under both far-field and non-stationary 256×64 XL-MIMO systems with 16-QAM. In particular, in MRIDA [40] or NRIDA [41], it has been shown that a larger size q would improve the detection performance but with the increased complexity burden of each full iteration accordingly. However, as shown in Table II, different from them, the variation of q in LCRID does not alter the complexity cost in a full iteration. Interestingly, this means that the choice of q can be set with much more degrees of freedom. More specifically, from Fig. 13, the detection performance can be improved by adjusting the size of q. Nevertheless, due to the approximation introduced by Taylor series, a larger size q does not always correspond to more performance gain, which means that the choice of q should also be carefully selected. Typically, as illustrated in Fig. 1, the related Taylor expansion with $q = 64$ fails to converge, but such an issue can be resolved by lowering the block size q. This is in line with the fact that we observe in Fig. 13. Moreover, when the convergence of Taylor is ensured, a large size q is welcome to enable the convergence of randomized iterations. Hence, there is a latent convergence trade-off between the randomized iterations of $\mathbf{x}^k$ and the Taylor expansion of $\mathbf{r}_t$.

VI. Conclusion and Future Work

This paper proposes new complexity reduction approaches for randomized iterations to enable low-complexity linear detection in uplink XL-MIMO systems. First, we incorporate the step size into randomized iterations to accelerate the convergence rate. Then, through Taylor series expansion, considerable computational complexity reduction in randomized iterations is achieved with negligible performance loss. Building on the above, the proposed LCRID algorithm is introduced. In addition, we apply conditional sampling techniques to LCRID for further accelerating convergence speed and improving iteration efficiency. Finally, simulation results are given to validate the superior gains of LCRID algorithm in terms of both performance and complexity as compared to benchmark schemes, demonstrating its competitiveness for signal detection in both far-field and non-stationary XL-MIMO communications.

Based on the proposed LCRID algorithm, the future research directions may be carried out in the following directions. First of all, given LCRID, further complexity reduction about randomized iterations is worthy of being investigated, especially for near-field XL-MIMO systems with highly correlated channels. Secondly, the joint signal processing between channel estimation and signal detection can be considered, where LCRID can be well designed to facilitate the interaction among different system modules. Thirdly, benefiting from the merits of LCRID, how to adopt it with deep learning in a reasonable way is also an interesting topic.

References

[1] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[2] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, "Toward the 6G network era: Opportunities and challenges," *IT Prof.*, vol. 22, no. 1, pp. 34–38, 2020.

[3] Z. Wang et al., "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1560–1605, 3rd Quart. 2024.

[4] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[5] H. Lu et al., "A tutorial on near-field XL-MIMO communications toward 6G," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 4, pp. 2213–2257, 4th Quart. 2024.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* New York, NY, USA: Cambridge Univ. Press, 2005.

[7] 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.

[8] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 916–929, Oct. 2014.

[9] F. Rusek et al., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[10] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, 4th Quart. 2019.

[11] A. Greenbaum, *Iterative Methods for Solving Linear Systems.* Philadelphia, PA, USA: SIAM, 1997.

[12] A. Bjorck, *Numerical Methods for Least Squares Problems.* Philadelphia, PA, USA: SIAM, 1996.

[13] Y. Saad, *Iterative Methods for Sparse Linear Systems.* Philadelphia, PA, USA: SIAM, 2nd Ed., 2003.

[14] C. Tang, C. Liu, L. Yuan, and Z. Xing, "High precision low complexity matrix inversion based on Newton iteration for data detection in the massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 490–493, Mar. 2016.

[15] F. Jin, Q. Liu, H. Liu, and P. Wu, "A low complexity signal detection scheme based on improved Newton iteration for massive MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 748–751, Apr. 2019.

[16] D. Zhu, B. Li, and P. Liang, "On the matrix inversion approximation based on Neumann series in massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 1763–1769.

[17] K. Izadinasab, A. W. Shaban, and O. Damen, "Low-complexity detectors for uplink massive MIMO systems leveraging truncated polynomial expansion," *IEEE Access*, vol. 10, pp. 91610–91621, 2022.

[18] Y. Zhang, A. Yu, X. Tan, Z. Zhang, X. You, and C. Zhang, "Adaptive damped Jacobi detector and architecture for massive MIMO uplink," in *Proc. IEEE Asia Pacific Conf. Circuits Syst. (APCCAS)*, 2018, pp. 203–206.

[19] X. Qin, Z. Yan, and G. He, "A near-optimal detection scheme based on joint steepest descent and Jacobi method for uplink massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 276–279, Feb. 2016.

[20] J. Tu, M. Lou, J. Jiang, D. Shu, and G. He, "An efficient massive MIMO detector based on second-order Richardson iteration: From algorithm to flexible architecture," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 67, no. 11, pp. 4015–4028, Nov. 2020.

[21] X. Gao, L. Dai, C. Yuen, and Y. Zhang, "Low-complexity MMSE signal detection based on Richardson method for large-scale MIMO systems," in *Proc. IEEE 80th Veh. Technol. Conf. (VTC2014-Fall)*, 2014, pp. 1–5.

[22] P. Zhang, L. Liu, G. Peng, and S. Wei, "Large-scale MIMO detection design and FPGA implementations using SOR method," in *Proc. 8th IEEE Int. Conf. Commun. Softw. Netw. (ICCSN)*, 2016, pp. 206–210.

[23] L. Dai, X. Gao, X. Su, S. Han, I. C.-L, and Z. Wang, "Low-complexity soft-output signal detection based on Gauss-Seidel method for uplink multiuser large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4839–4845, Oct. 2015.

[24] M. A. Albreem, H. A. Saleem, S. Abdallah, and S. Shahabuddin, "A low complexity data detection scheme for massive MIMO systems using the TOR algorithm," *IEEE Signal Process. Lett.*, vol. 32, pp. 2429–2433, 2025.

[25] S. Berra, S. Chakraborty, and R. Dinis, "A low complexity iterative SC-FDE receiver for massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 10405–10420, Jul. 2024.

[26] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and Y. Su, "Convergence analysis and assurance for Gaussian message passing iterative detector in massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6487–6501, Sep. 2016.

[27] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.

[28] T. L. Narasimhan and A. Chockalingam, "Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 847–860, Oct. 2014.

[29] C. Yang, L. Tang, and Z. Wang, "Enhanced channel hardening-exploiting message passing algorithm for large-scale MIMO detection," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2021, pp. 1–5.

[30] Y. Xue, C. Zhang, S. Zhang, Z. Wu, and X. You, "Steepest descent method based soft-output detection for massive MIMO uplink," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, 2016, pp. 273–278.

[31] J. Minango, C. de Almeida, and C. Daniel Altamirano, "Low-complexity MMSE detector for massive MIMO systems based on damped Jacobi method," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, 2017, pp. 1–5.

[32] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1497–1511, Dec. 2011.

[33] Y. Hu, Z. Wang, X. Gaol, and J. Ning, "Low-complexity signal detection using CG method for uplink large-scale MIMO systems," in *Proc. IEEE Int. Conf. Commun. Syst.*, 2014, pp. 477–481.

[34] S. Berra, A. Benchabane, S. Chakraborty, K. Maruta, R. Dinis, and M. Beko, "A low complexity linear precoding method for extremely large-scale MIMO systems," *IEEE Open J. Veh. Technol.*, vol. 6, pp. 240–255, 2025.

[35] B. Xu, Z. Wang, H. Xiao, J. Zhang, B. Ai, and D. W. Kwan Ng, "Low-complexity precoding for extremely large-scale MIMO over non-stationary channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 6516–6522.

[36] Z. Wang and C. Ling, "On the geometric ergodicity of Metropolis-Hastings algorithms for lattice Gaussian sampling," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 738–751, Feb. 2018.

[37] Z. Wang, Y. Xia, C. Ling, and Y. Huang, "Randomized iterative sampling decoding algorithm for large-scale MIMO detection," *IEEE Trans. Signal Process.*, vol. 72, pp. 580–593, 2024.

[38] Z. Wang and C. Ling, "Lattice Gaussian sampling by Markov chain Monte Carlo: Bounded distance decoding and trapdoor sampling," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3630–3645, Jun. 2019.

[39] Z. Wang, C. Ling, S. Jin, Y. Huang, and F. Gao, "Probabilistic searching for MIMO detection based on lattice Gaussian distribution," *IEEE Trans. Commun.*, vol. 72, no. 1, pp. 85–100, Jan. 2024.

[40] Z. Wang, R. M. Gower, Y. Xia, L. He, and Y. Huang, "Randomized iterative methods for low-complexity large-scale MIMO detection," *IEEE Trans. Signal Process.*, vol. 70, pp. 2934–2949, 2022.

[41] Z. Wang, W. Xu, Y. Xia, Q. Shi, and Y. Huang, "A new randomized iterative detection algorithm for uplink large-scale MIMO systems," *IEEE Trans. Commun.*, vol. 71, no. 9, pp. 5093–5107, Sep. 2023.

[42] J. Tian, Y. Han, S. Jin, J. Zhang, and J. Wang, "Analytical channel modeling: From MIMO to extra large-scale MIMO," *Chin. J. Electronics*, vol. 34, no. 1, pp. 1–15, Jan. 2025.

[43] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[44] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 40–46, Jan. 2023.

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

[46] A. Björck, *Numerical Methods in Matrix Computations.* Berlin, Germany: Springer, 2015.

[47] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications.* Now Foundations Trends, 2004.

[48] A. Ali, E. D. Carvalho, and R. W. Heath, "Linear receivers in non-stationary massive MIMO channels with visibility regions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 885–888, Jun. 2019.

[49] F. Zhang and Q. Zhang, "Eigenvalue inequalities for matrix product," *IEEE Trans. Autom. Control*, vol. 51, no. 9, pp. 1506–1509, Sep. 2006.

[50] Q. Chen, S. Zhang, S. Xu, and S. Cao, "Efficient MIMO detection with imperfect channel knowledge - A deep learning approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2019, pp. 1–6.

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

[52] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, 2020.

[53] Z. Gao et al., "Data-driven deep learning based hybrid beamforming for aerial massive MIMO-OFDM systems with implicit CSI," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 2894–2913, Oct. 2022.

Zheng Wang (Senior Member, IEEE) received the B.S. degree in electronic and information engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009, the M.S. degree in communications from the University of Manchester, Manchester, U.K., in 2010, and the Ph.D. degree in communication engineering from the Imperial College London, U.K., in 2015. Since 2021, he has been an Associate Professor with the School of Information and Engineering, Southeast University, Nanjing. From 2015 to 2016, he served as a Research Associate with the Imperial College London. From 2016 to 2017, he was a Senior Engineer with the Radio Access Network R&D division, Huawei Technologies Company. From 2017 to 2020, he was an Associate Professor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing. From 2023 to 2024, he has been recognized as a Highly Cited Chinese Researcher by Elsevier for exceptional research performance in the field of information and communications engineering. He received the Spark Award from Huawei Technologies Company in 2023 for the research works towards Spark Challenge about Efficient Baseband Matrix Processing. He is also the Youth Editor of the Chinese Journal of Electronics (CJE). His research interests include massive MIMO systems, machine learning, and data analytic over wireless networks, and lattice theory for wireless communications.

Yong Zeng (Fellow, IEEE) received the Bachelor of Engineering (Hons.) and Ph.D. degrees from Nanyang Technological University, Singapore. He is a Chief Young Professor with the National Mobile Communications Research Laboratory, Southeast University, China, and also with the Purple Mountain Laboratories, Nanjing, China. From 2013 to 2018, he was a Research Fellow and a Senior Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore. From 2018 to 2019, he was a Lecturer with the School of Electrical and Information Engineering, University of Sydney, Australia. He was listed as a Highly Cited Researcher by the Clarivate Analytics for six consecutive years (2019–2024). He has published more than 200 papers, which have been cited by more than 34 000 times based on Google Scholar. He was elevated to IEEE Fellow "for contributions to UAV communications and wireless power transfer." He was a recipient of the Australia Research Council (ARC) Discovery Early Career Researcher Award (DECRA), the 2020 and 2024 IEEE Marconi Prize Paper Award in Wireless Communications, the 2018 IEEE Communications Society Asia-Pacific Outstanding Young Researcher Award, the 2020 and 2017 IEEE Communications Society Heinrich Hertz Prize Paper Award, the 2021 IEEE ICC Best Paper Award, and the 2021 China Communications Best Paper Award. He was the Symposium Chair of IEEE Globecom 2021 Track on Aerial Communications, the Workshop Co-Chair of ICC 2018–2023 Workshop on UAV Communications, and the Tutorial Speaker of Globecom 2018/2019 and ICC 2019 tutorials on UAV communications. He proposed the concept of channel knowledge map (CKM) and the transmission method of delay-Doppler alignment modulation (DDAM). He serves as an Editor for IEEE Transactions on Communications, IEEE Transactions on Mobile Computing, IEEE Communications Letters, and IEEE Open Journal of Vehicular Technology. He serves as a Leading Guest Editor for IEEE Wireless Communications on "Integrating UAVs into 5G and Beyond" and China Communications on "Network-Connected UAV Communications."

Yongming Huang (Fellow, IEEE) received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from the Southeast University, Nanjing, in 2007. Since 2007, he has been a faculty with the School of Information Science and Engineering, Southeast University, where he is currently a Full Professor. From 2008 to 2009, he visited the Signal Processing Lab, Royal Institute of Technology, Stockholm, Sweden. He has authored or coauthored more than 200 peer-reviewed papers, and holds more than 80 invention patents. His research interests include intelligent 5G/6G mobile communications and millimeter wave wireless communications. He submitted around 20 technical contributions to IEEE standards, and was awarded a certificate of appreciation for outstanding contribution to the development of IEEE standard 802.11aj. He was an Associate Editor for IEEE Transactions on Signal Processing and a Guest Editor of IEEE Journal Selected Areas in Communications. He is currently an Editor-at-Large of IEEE Open Journal of the Communications Society and an Associate Editor for IEEE Wireless Communications Letters.

Rui Zhang (Fellow, IEEE) received the B.Eng. (Hons.) and M.Eng. degrees from the National University of Singapore, Singapore, and the Ph.D. degree from Stanford University, Stanford, CA, USA, all in electrical engineering. From 2007 to 2009, he was a Research Scientist with the Institute for Infocomm Research, ASTAR, Singapore. In 2010, he joined the Department of Electrical and Computer Engineering, National University of Singapore, where he is currently a Provost's Chair Professor. He is also an Adjunct Professor with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China. He has published over 600 papers, all in the field of wireless communications and networks. Since 2015, he has been listed as a Highly Cited Researcher by Thomson Reuters/Clarivate Analytics. His research interests include intelligent surfaces, reconfigurable antennas, radio mapping, non-terrestrial communications, wireless power transfer, AI, and optimization methods. He is a fellow of the Academy of Engineering Singapore. He was a recipient of the Sixth IEEE Communications Society Asia Pacific Region Best Young Researcher Award in 2011, the Young Researcher Award of National University of Singapore in 2015, the Wireless Communications Technical Committee Recognition Award in 2020, the IEEE Signal Processing and Computing for Communications (SPCC) Technical Recognition Award in 2021, the IEEE Communications Society Technical Committee on Cognitive Networks (TCCN) Recognition Award in 2023, and the IEEE James Evans Avant Garde Award in 2025. His works received 18 IEEE Best Journal Paper Awards, including the IEEE Marconi Prize Paper Award in Wireless Communications in 2015 and 2020; the IEEE Signal Processing Society Best Paper Award in 2016; the IEEE Communications Society Heinrich Hertz Prize Paper Award in 2017, 2020, and 2022; and the IEEE Communications Society Stephen O. Rice Prize in 2021. He served for over 30 international conferences as the TPC Co-Chair and an organizing committee member. He was an Elected Member of the IEEE Signal Processing Society SPCOM Technical Committee from 2012 to 2017 and SAM Technical Committee from 2013 to 2015. He served as the Vice Chair for the IEEE Communications Society Asia Pacific Board Technical Affairs Committee from 2014 to 2015 and a member for the Steering Committee of IEEE WIRELESS COMMUNICATIONS LETTERS from 2018 to 2021 and the IEEE Communications Society Wireless Communications Technical Committee (WTC) Award Committee from 2023 to 2025. He is the Chair of the IEEE Communications Society Wireless Communications Technical Committee (WTC) Award Committee. He served as an Editor for several IEEE journals, including IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2012 to 2016, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: GREEN COMMUNICATIONS AND NETWORKING SERIES from 2015 to 2016, IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2013 to 2017, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING from 2016 to 2020, and IEEE TRANSACTIONS ON COMMUNICATIONS from 2017 to 2022. He serves as an Editorial Board Member for npj Wireless Technology. He was a Distinguished Lecturer of the IEEE Signal Processing Society and the IEEE Communications Society from 2019 to 2020.