A Statistical Linear Precoding Scheme Based on Random Iterative Method for Massive MIMO Systems

Zheng Wang, *Member, IEEE*, Robert M. Gower, Cheng Zhang, *Member, IEEE*, Shanxiang Lyu, Yili Xia, *Member, IEEE*, and Yongming Huang, *Senior Member, IEEE*

***Abstract*—In this paper, the random iterative method is introduced to massive multiple-input multiple-output (MIMO) systems for the efficient downlink linear precoding. By adopting the random sampling into the traditional iterative methods, the matrix inversion within the linear precoding schemes can be approximated statistically, which not only achieves a faster exponential convergence with low complexity but also experiences a global convergence without suffering from the various convergence requirements. Specifically, based on the random iterative method, the randomized iterative precoding algorithm (RIPA) is firstly proposed and we show its approximation error decays exponentially and globally along with the number of iterations. Then, with respect to the derived convergence rate, the concept of conditional sampling is introduced, so that further optimization and enhancement are carried out to improve both the convergence and the efficiency of the randomized iterations. After that, based on the equivalent iteration transformation, the modified randomized iterative precoding algorithm (MRIPA) is presented, which achieves a better precoding performance with low-complexity for various scenarios of massive MIMO. Finally, simulation results based on downlink precoding in massive MIMO systems are given to show the system gains of RIPA and MRIPA in terms of performance and complexity.**

***Index Terms*—Massive MIMO, large-scale MIMO, linear precoding, low complexity, iterative methods, matrix inversion, convergence analysis and enhancement.**

I. Introduction

As a promising technique for cellular network in the 5th generation mobile network (5G) and beyond 5G, massive multiple-input multiple-output (MIMO) systems have been widely investigated in the last decade, which boosts the network capacity on a much greater scale without extra bandwidth [1]–[4]. By deploying large-scale antenna array at the base station (BS), significant array gain and spatial resolution can be achieved, which greatly facilitates the MIMO system to serve numerous users simultaneously [5], [6]. Specifically, it has been shown that linear precoding techniques like zero-forcing (ZF), regularized ZF (RZF) are capable of achieving the capacity-approaching performance if the number of antennas at BS is sufficiently large [7]–[9]. For this reason, more efforts have been paid attention to decrease the complexity cost of the matrix inversion in linear precoding [10]–[15], which is computational expensive especially in high-dimensional systems. However, some of these low-complexity precoding schemes like Neumann series (NS), Newton iteration (NI) and so on suffer from some specific convergence requirements (i.e., the number of antennas at BS should be greatly larger than that of the user side), rendering them rather limited in the various scenarios of interest. Actually, besides BS, the total number of antennas at the user side also improves rapidly, which contains the increments of both the number of users and the number of antennas at each user equipment (UE) [16], [17]. This means the environment of wireless communications becomes much more complicated than before, making a flexible and effective precoding scheme highly desired.

In particular, to bypass the matrix inversion in linear precoding schemes, a straightforward approximation way is to resort to the polynomial expansion. In particular, the Neumann series and the Kapteyn series are introduced respectively, which transforms the matrix inversion into a precondition matrix with simple matrix multiplications and summations [18], [19]. With the increase of the polynomial items, an approximation of the matrix inversion can be achieved [20]. Unfortunately, the convergence of Neumann series is guaranteed only if the number of antennas at base station is much larger than that on the

Manuscript received 16 October 2021; revised 4 April 2022; accepted 8 June 2022. Date of publication 20 June 2022; date of current version 12 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61801216, Grant 61771124, and Grant 61720106003; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180420 and Grant BK20190337; in part by the State Key Laboratory of Integrated Services Networks (Xidian University) under Grant ISN21-31; in part by the Zhi Shan Young Scholar Program of Southeast University; and in part by the Fundamental Research Funds for the Central Universities under Grant 2242022k30002. The associate editor coordinating the review of this article and approving it for publication was H. Q. Ngo. *(Corresponding author: Zheng Wang.)*

Zheng Wang is with the School of Information Science and Engineering and the Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, China, and also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: z.wang@ieee.org).

Robert M. Gower is with the Center for Computational Mathematics, Flatiron Institute, Simons Foundation, New York, NY 10010 USA (e-mail: rgower@flatironinstitute.org).

Cheng Zhang, Yili Xia, and Yongming Huang are with the School of Information Science and Engineering and the Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, China.

Shanxiang Lyu is with the College of Cyber Security, Jinan University, Guangzhou 510632, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3182407.

Digital Object Identifier 10.1109/TWC.2022.3182407

user side, otherwise its performance will severely degrade [21]. Although Newton iteration was further employed with a better convergence performance than Neumann series [22], [23], it still suffers from this requirement in convergence [24]. On the other hand, a low-complexity RZF precoding is presented in [12], [13], which replaces the matrix inversion within it by a truncated polynomial expansion (TPE). Nevertheless, TPE precoding has to deal with a complicated parameter optimization problems.

Besides polynomial expansion, another general strategy to avoid matrix inversion is the iterative methods for solving the linear systems [25]. However, the convergence of these iterative methods are also somehow restricted [26], [27]. For example, in Jacobi iteration, the convergence is ensured if the symmetric positive matrix $\mathbf{A}$ (i.e., the matrix inversion of $\mathbf{A}$ is the target matrix) is strictly diagonally dominant [28]. As for Richardson iteration, the relaxation factor $0 < \omega < 2/\rho(\mathbf{A})$ that controls the matrix splitting during the iterations should be well selected to guarantee the convergence [29], [30], where $\rho(\cdot)$ denotes the spectral radius of a matrix. As for successive over-relaxation (SOR) and symmetric successive over-relaxation (SSOR), they both restrict the relaxation factor as $0 < \omega < 2$ to ensure the convergence [31], [32]. Meanwhile, these iterative methods generally have prohibitive computation overhead, which are not well suited for systems with moderate to large channel coherence time [10], especially for wireless communication systems with large bandwidth. Other low complexity precoding schemes can also be found in [33]–[35].

Nowadays, random sampling has emerged as a fundamental tool in solving various signal processing problems. From it, extra processing freedom can be introduced while considerable performance and complexity gain can be obtained [36]–[40]. In this paper, in order to achieve the low-complexity precoding with better and global convergence performance, a statistical precoding scheme based on random iterative method is proposed for various downlink scenarios of massive MIMO. First of all, by approximating the matrix inversion by incorporating a random sampling into the traditional iterative methods, the randomized iterative precoding algorithm (RIPA) is proposed for downlink massive MIMO systems. Meanwhile, its convergence rate for iteratively approximating the matrix inversion is derived and we show that it not only converges exponentially fast but also enjoys a global convergence. This means the convergence obstacle of the traditional iterative methods for massive MIMO systems is successfully overcome. Secondly, by resorting to the conditional sampling, theoretic analysis and optimization with respect to the randomized iteration are given. Finally, based on the equivalent iteration transformation, the modified randomized iterative precoding algorithm (MRIPA) is proposed to further improve the precoding performance and efficiency. To summarize, the proposed MRIPA not only achieves a faster convergence performance with low complexity cost, but also enjoys a global convergence to well suited the various cases of massive MIMO systems.

The rest of this paper is organized as follows. Section II briefly introduces the traditional linear precoding for downlink massive MIMO systems and reviews the low-complexity precoding schemes by polynomial expansion and iterative methods. In Section III, the proposed randomized iterative precoding algorithm (RIPA) is described and its convergence analysis is given to show the global and the exponential convergence performance. In Section IV, the concept of iteration transformation is proposed to lower the complexity of randomized iteration for precoding. Then, by adopting the conditional sampling into the randomized iteration, further optimization and enhancement are given. In Section V, the modified randomized iterative precoding algorithm (MRIPA) is presented to improve the convergence and efficiency. After that, simulations of the proposed precoding schemes for downlink massive MIMO are shown in Section VI. Finally, Section VII concludes the paper.

Notation: Matrices and column vectors are denoted by upper and lowercase boldface letters, and the transpose, conjugate transpose, inverse, pseudoinverse of a matrix $\mathbf{B}$ by $\mathbf{B}^T, \mathbf{B}^H, \mathbf{B}^{-1}$, and $\mathbf{B}^{\dagger}$, respectively. We use $\mathbf{b}_i$ for the ith column of the matrix $\mathbf{B}$, $b_{i,j}$ for the entry in the ith row and jth column of the matrix $\mathbf{B}$. Let $\langle \mathbf{X}, \mathbf{Y} \rangle_{F(\mathbf{W}^{-1})} \triangleq \text{Tr}(\mathbf{X}^H \mathbf{W}^{-1} \mathbf{Y} \mathbf{W}^{-1})$ denote the weighted Frobenius inner product, where $\mathbf{X}, \mathbf{Y} \in \mathbb{C}^{n \times n}$ and $\mathbf{W} \in \mathbb{C}^{n \times n}$ is a symmetric positive definite matrix. Furthermore, let $\|\mathbf{X}\|^2_{F(\mathbf{W}^{-1})} \triangleq \text{Tr}(\mathbf{X}^H \mathbf{W}^{-1} \mathbf{X} \mathbf{W}^{-1}) = \|\mathbf{W}^{-\frac{1}{2}} \mathbf{X} \mathbf{W}^{-\frac{1}{2}}\|^2_F$ where $\|\cdot\|_F$ is the standard Frobenius norm with identity matrix $\mathbf{I}$ and $\text{Tr}(\cdot)$ denotes the trace of the matrix. $\Re(\cdot)$ and $\Im(\cdot)$ indicate the real and imaginary components.

II. Preliminary

In this section, the linear precoding in downlink massive MIMO systems is reviewed, followed by the background of low-complexity linear precoding schemes derived by polynomial expansion and iterative methods.

A. Linear Precoding in the Downlink

The base station (BS) in the massive MIMO system we considered is equipped with N antennas, and simultaneously serves K single antenna user terminals (UT) ($N \geq K$). Throughout the context, we assume the channel matrix $\mathbf{H} \in \mathbb{C}^{N \times K}$ to be perfectly known at BS.

Specifically, let $\mathbf{s}$ denote the $K \times 1$ transmitted source information to K users during the downlink transmission, according to precoding, the signal vector $\mathbf{y}$ received at UTs are [9], [10]

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}^H \mathbf{G} \mathbf{s} + \mathbf{n}. \tag{1}$$

Here, $\mathbf{G} \in \mathbb{C}^{N \times K}$ is the precoding matrix, $\rho > 0$ is the average transmit power at BS, and $\mathbf{n}$ is a $K \times 1$ additive white Gaussian noise vector whose entries follow $\mathcal{CN}(0, \sigma^2)$. To satisfy the power constraint, $\mathbf{s}$ and $\mathbf{G}$ are selected according to $\|\mathbf{s}\|^2 = 1$ and $\text{tr}(\mathbf{G}\mathbf{G}^H) = 1$.

To eliminate the interference during the downlink transmission, ZF precoder is defined as

$$\mathbf{G}_{\text{zf}} = \beta \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1}, \tag{2}$$

which transmits the signal toward the intended user while nulling in the directions of other users. Here, β serves as a scaling factor to normalize the signal power. Furthermore, as an enhanced version of ZF, the regularized ZF (RZF) is given by

$$\mathbf{G}_{\text{rzf}} = \beta\mathbf{H}(\mathbf{H}^H\mathbf{H} + \xi\mathbf{I})^{-1} \tag{3}$$

according to mean square error (MSE) criterion, making it also know as the minimum mean-square error (MMSE) precoding. Here, $\mathbf{I}$ is the identity matrix and ξ is the scalar regularization coefficient. Typically, RZF precoding reduces to ZF if $\xi \to 0$ while RZF precoding is similar to maximal ratio transmission (MRT) when $\xi \to \infty$.

From (2) and (3), the precoding matrices $\mathbf{G}_{\text{zf}}$ and $\mathbf{G}_{\text{rzf}}$ are functions of the channel matrix, which both require the matrix inversion of large size. Theoretically, this costs $\mathcal{O}(K^3)$ computational complexity, which increases rapidly as the dimension of massive MIMO expands. To this end, a number of advanced low-complexity precoding schemes has been proposed to avoid the matrix inversion.

B. Low Complexity Iterative Precoding

Given RZF precoding, the truncated polynomial expansion (TPE) precoding is proposed to reduce the computational complexity while maintaining the similar performance. In particular, TPE precoding replaces the matrix inversion in RZF by a truncated polynomial expansion so that its precoding matrix is given by

$$\mathbf{G}_{\text{tpe}} = \sum_{k=0}^{J-1} \omega_k (\mathbf{H}^H\mathbf{H})^k \mathbf{H}^H, \tag{4}$$

where the coefficient ω_i denotes the precoder polynomial of order J. Intuitively, the selection of J enables a flexible trade-off and the approximation accuracy gradually improves with the increment of J.[1] However, finding the optimal choice of coefficients ω_i's turns out to be not easy in practice.

In [18], Neumann series is applied into precoding to approximate the matrix inversion with low complexity cost. For example, let $\mathbf{A} = \mathbf{H}^H\mathbf{H}$ for ZF precoding or $\mathbf{A} = \mathbf{H}^H\mathbf{H} + \xi\mathbf{I}$ for RZF precoding so that the precoding matrix

$$\mathbf{G} = \beta\mathbf{H}\mathbf{A}^{-1}, \tag{5}$$

then the approximation of the matrix inversion of $\mathbf{A}$ becomes [7]

$$\mathbf{A}^{-1} = \sum_{k=0}^{\infty} (\mathbf{I} - \Theta\mathbf{A})^k \Theta, \tag{6}$$

where Θ is a $K \times K$ diagonal matrix and k denotes the iteration index. Based on Neumann series, several modifications like identity matrix NS (INS), diagonal matrix NS (DNS) [41], and identity-plus-column NS (ICNS) [14] have been proposed to further improve the approximation. However, the approximation in (6) holds only if

$$\lim_{k\to\infty} (\mathbf{I} - \Theta\mathbf{A})^k = \mathbf{0}, \tag{7}$$

which implies the condition $N \gg K$ is required in massive MIMO systems. More specifically, this condition is further specified as $N/K \geq 5.83$ in [21].

On the other hand, to approximate the matrix inversion of $\mathbf{A}$ contained in $\mathbf{G}_{\text{rzf}}$ or $\mathbf{G}_{\text{zf}}$, the following linear system can be established

$$\mathbf{A}\mathbf{t} = \mathbf{s} \tag{8}$$

with $\mathbf{t} = \mathbf{A}^{-1}\mathbf{s}$. In this way, the system model (1) in becomes

$$\mathbf{y} = \sqrt{\rho}\beta\mathbf{H}^H\mathbf{H}\mathbf{t} + \mathbf{n} \tag{9}$$

with the transmitted signal obtained by

$$\mathbf{x} = \mathbf{G}_{\text{rzf}}\mathbf{s} = \beta\mathbf{H}\mathbf{A}^{-1}\mathbf{s} = \beta\mathbf{H}\mathbf{t}, \tag{10}$$

where the problem of approximating the matrix inversion of $\mathbf{A}$ turns out to solve the linear systems about $\mathbf{t}$ in (8).

Therefore, by splitting $\mathbf{A}$ into $\mathbf{A} = \mathbf{P} + \mathbf{Q}$ (matrix $\mathbf{P} \in \mathbb{C}^{K\times K}$ is nonsingular, $\mathbf{Q} \in \mathbb{C}^{K\times K}$), the linear system in (8) can also be solved by the following iterations

$$\mathbf{t}^{(k)} = \mathbf{B}\mathbf{t}^{(k-1)} + \mathbf{f} \tag{11}$$

where $\mathbf{B} = -\mathbf{P}^{-1}\mathbf{Q} = \mathbf{I} - \mathbf{P}^{-1}\mathbf{A} \in \mathbb{C}^{K\times K}$ is known as the iteration matrix and $\mathbf{f} = \mathbf{P}^{-1}\mathbf{s} \in \mathbb{C}^K$. However, the iterative methods also have to confront the convergence problem by satisfying

$$\lim_{k\to\infty} \mathbf{B}^k = \mathbf{0}. \tag{12}$$

For this reason, the iterative precoding scheme like Newton iteration still suffers from the same convergence requirement as Neumann series.

As for Jacobi and Richardson iterative methods, the iteration matrices are set as $\mathbf{B}_{\text{Jacobi}} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$ (i.e., $\mathbf{P} = \mathbf{D}$) and $\mathbf{B}_{\text{Richardson}} = \mathbf{I} - \omega\mathbf{A}$ (i.e., $\mathbf{P} = \frac{1}{\omega}\mathbf{I}$) respectively, where $\mathbf{D} \in \mathbb{C}^{K\times K}$ is the diagonal component of the matrix $\mathbf{A}$ and $\omega > 0$ is known as the relaxation parameter. Typically, in order to guarantee the convergence, the matrix $\mathbf{A}$ in Jacobi iteration should be strictly diagonally dominant (SDD), i.e.,

$$|a_{i,i}| > \sum_{j\neq i} |a_{i,j}| \tag{13}$$

while the Richardson iteration is convergent if $0 < \omega < \frac{2}{\varrho(\mathbf{A})}$. $\varrho(\mathbf{A})$ is the spectral radius of matrix $\mathbf{A}$. Clearly, this is also related to a specific requirement of $\mathbf{A}$ [28] and ω should be well selected for the convergence. For a better convergence, the iteration method of successive overrelaxation (SOR) was introduced as [42]

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

with $\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{L}^H$, where $\mathbf{D}$, $\mathbf{L}$ and $\mathbf{L}^H$ respectively stand for the diagonal components, the strictly lower triangular components and the strictly upper triangular components of $\mathbf{A}$. However, the SOR method converges only if the relaxation parameter $0 < \omega < 2$, which should be carefully selected.

[1] TPE precoding is equivalent to the traditional MRT precoding when $J = 1$. On the other hand, RZF precoding will be achieved if J goes infinity.

III. Randomized Iterative Precoding Algorithm

In this section, the statistical linear precoding based on random iterative methods for downlink massive MIMO system is proposed. By capturing the advantages of random sampling, exponential and global convergence can be achieved by the random iterations, which leads to the randomized iterative precoding algorithm (RIPA).

A. Algorithm Description

Given the target matrix $\mathbf{A} \in \mathbb{C}^{K\times K}$ (i.e., $\mathbf{A} = \mathbf{H}^H\mathbf{H} + \xi\mathbf{I}$ for RZF and $\mathbf{A} = \mathbf{H}^H\mathbf{H}$ for ZF), in order to find its matrix inversion $\mathbf{X} \in \mathbb{C}^{K\times K}$ for $\mathbf{AX} = \mathbf{I}$, an auxiliary matrix $\mathbf{S} \in \mathbb{C}^{K\times q}$ can be introduced so that the problem of finding $\mathbf{A}^{-1}$ can be partially characterized by finding the solution to

$$\mathbf{S}^H\mathbf{AX} = \mathbf{S}^H. \tag{15}$$

Theoretically, if $\mathbf{S}$ is a $K \times K$ (i.e., $q = K$) invertible matrix, left multiplying $\mathbf{S}^H$ to the both sides in (15) will not change the solution, where solving (15) exactly yields the same result of solving $\mathbf{AX} = \mathbf{I}$. However, if $q \ll K$, then the dimension of (15) will be reduced greatly, making it easier to solve. Most importantly, the dimension reduction by the introduced matrix $\mathbf{S}$ comes at the price of various solutions to (15), which should be carefully taken care of.

Then, to seek for the desired result among numerous solutions of $\mathbf{X}$, a certain level of randomness can be introduced by $\mathbf{S}$ to calculate $\mathbf{X}$ iteratively and dynamically [43]–[45]. In this way, the most general solution of $\mathbf{X}$ that fits every random matrix $\mathbf{S} \in \mathbb{C}^{K\times q}$ in (15) will be obtained gradually, which means to exploit as much of information learned so far as possible. Meanwhile, in order to calibrate the norm of a matrix in the Euclidean space, the squared *Frobenius norm* is applied throughout the context, which cycles through all the entries of a matrix by

$$\|\mathbf{X}\|_F^2 = \sum_i^m \sum_j^n |x_{i,j}|^2 = \mathrm{Tr}(\mathbf{XX}^H). \tag{16}$$

Then, based on Frobenius norm, a symmetric positive definite matrix $\mathbf{W} \in \mathbb{C}^{K\times K}$ is introduced to serve as the weighted matrix, i.e.,

$$\|\mathbf{X}\|^2_{F(\mathbf{W}^{-1})} \triangleq \|\mathbf{W}^{-\frac{1}{2}}\mathbf{X}\mathbf{W}^{-\frac{1}{2}}\|_F^2. \tag{17}$$

From it, the problem in (15) can be expressed in an iterative way as

$$\begin{aligned}\mathbf{X}_{k+1} = \ &\underset{\mathbf{X}\in\mathbb{C}^{K\times K}}{\arg\min}\ \|\mathbf{X} - \mathbf{X}_k\|^2_{F(\mathbf{W}^{-1})} \\ &\text{subject to } \mathbf{S}_k^H\mathbf{AX} = \mathbf{S}_k^H,\end{aligned} \tag{18}$$

where k denotes the iteration index, $\mathbf{S}_k$ obeys a discrete distribution $\mathcal{D}$ with $r > 0$ outcomes, i.e., $\mathbf{S}_k \in \{\mathbf{M}_1, \ldots, \mathbf{M}_r\}$, and $\mathbf{M}_i \in \mathbb{C}^{K\times q_i}$, $1 \le i \le r$ is a full column rank matrix with probability

$$p_i \triangleq \mathcal{D}(\mathbf{S}_k = \mathbf{M}_i) > 0 \tag{19}$$

and $\sum_{i=1}^r p_i = 1$.

Based on (18), given the random chosen matrix $\mathbf{S}_k$, the update of $\mathbf{X}^{(k+1)}$ at each iteration is specified as the closest result to $\mathbf{X}^{(k)}$ with the weighted matrix, which results in a convergent iterations of $\mathbf{X}^{(k)}$. We point out that the distribution $\mathcal{D}$ and the weight matrix $\mathbf{W}$ are the system parameters, which should be carefully designed for the sake of convergence. More specifically, according to optimization theory, the dual formulation of (18) is

$$\mathbf{X}^{(k+1)} = \underset{\mathbf{X}\in\mathbb{C}^{K\times K},\mathbf{Y}\in\mathbb{C}^{K\times q}}{\arg\min} \|\mathbf{X}^{(k)} - \mathbf{A}^{-1}\|^2_{F(\mathbf{W}^{-1})} \tag{20}$$

$$\begin{aligned}&\text{subject to } \mathbf{X} = \mathbf{X}^{(k)} + \mathbf{WA}^H\mathbf{S}_k\mathbf{Y}^H, \\ &\mathbf{S}_k \sim \mathcal{D}.\end{aligned} \tag{21}$$

Then, by substituting (21) into (15), we can obtain that

$$\mathbf{Y}^H = (\mathbf{S}_k^H\mathbf{AWA}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H(\mathbf{I} - \mathbf{AX}^{(k)}). \tag{22}$$

Subsequently, by putting (22) back to (21), we can get an explicit expression of $\mathbf{X}^{(k+1)}$, namely

$$\begin{aligned}\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \mathbf{WA}^H\mathbf{S}_k(\mathbf{S}_k^H\mathbf{AWA}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H \\ \times (\mathbf{I} - \mathbf{AX}^{(k)})\end{aligned} \tag{23}$$

with $\mathbf{S}_k \sim \mathcal{D}$, which could be further expressed by

$$\mathbf{X}^{(k+1)} - \mathbf{A}^{-1} = (\mathbf{I} - \mathbf{WZ})(\mathbf{X}^{(k)} - \mathbf{A}^{-1}) \tag{24}$$

with the defined symmetric matrix $\mathbf{Z}$

$$\mathbf{Z} \triangleq \mathbf{A}^H\mathbf{S}_k(\mathbf{S}_k^H\mathbf{AWA}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H\mathbf{A} \in \mathbb{C}^{K\times K}. \tag{25}$$

From (24), it is clear to see that the approximation of $\mathbf{A}^{-1}$ by $\mathbf{X}^{(k)}$ proceeds iteratively under the introduced random sampling brought by $\mathbf{S}_k$.

In order to specify the randomized iteration in RIPA, the following choices about systems parameters $\mathcal{D}$ and $\mathbf{W}$ with respect to (23) are given. On one hand, let $\mathbf{W} = (\mathbf{A}^H\mathbf{A})^{-1}$, then the iteration in (23) turns out to be

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \mathbf{A}^{-1}\mathbf{S}_k(\mathbf{S}_k^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H(\mathbf{I} - \mathbf{AX}^{(k)}). \tag{26}$$

On the other hand, we set $\mathbf{S}_k = \mathbf{AX}^{(k)}_{:,q_i's}$, $\mathbf{X}^{(k)}_{:,q_i's}$ stands for a column concatenation of q_i columns of matrix $\mathbf{X}^{(k)}$ (i.e., $\mathbf{X}^{(k)}_{:,q_i's} = \mathbf{X}^{(k)}\mathbf{I}_{:,q_i's}$ where $\mathbf{I}_{:,q_i's}$ represents a column concatenation of q_i columns of $K \times K$ identity matrix $\mathbf{I}$) and the q_i columns are uniform randomly selected from $i \in \{1, \ldots, K\}$. Here, to enable an efficient sampling, the index set of multiple columns related to q_i is fixed, which forms a block operation. For example, each q_i corresponds to a set containing 3 column indices as follows

$$\{1, 2, 5\}_{q_1} \cup \ldots \cup \{4, 8, 12\}_{q_r} = \{1, \ldots, K\} \tag{27}$$

with sets $\{\cdot\}_{q_i} \cap \{\cdot\}_{q_j} = \emptyset$ and $\sum_{i=1}^r q_i = K$.

Therefore, based on $\mathbf{W} = (\mathbf{A}^H\mathbf{A})^{-1}$ and $\mathbf{S}_k = \mathbf{AX}^{(k)}_{:,q_i's}$, the iteration in the proposed RIPA can be expressed as

$$\begin{aligned}\mathbf{X}^{(k+1)} &= \mathbf{X}^{(k)} + \mathbf{X}^{(k)}_{:,q_i's}(\mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H\mathbf{AX}^{(k)}_{:,q_i's})^{-1}\mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H \\ &\quad \times (\mathbf{I} - \mathbf{AX}^{(k)}) \\ &= \mathbf{X}^{(k)} + \mathbf{X}^{(k)}_{:,q_i's}(\mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H\mathbf{AX}^{(k)}_{:,q_i's})^{-1} \\ &\quad \times (\mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H - \mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H\mathbf{AX}^{(k)})\end{aligned} \tag{28}$$

where the sizes of the index set are set equally $q_i = \ldots = q_r = q$ for the sake of implementation simplicity.

Algorithm 1 Randomized Iterative Precoding Algorithm (RIPA) for Downlink Massive MIMO Systems

Require: $\mathbf{A}=\mathbf{H}^H\mathbf{H}$ or $\mathbf{A}=\mathbf{H}^H\mathbf{H}+\xi\mathbf{I}$, $\mathbf{X}^{(0)}=\mathbf{D}^{-1}$, Q, β
Ensure: near ZF or RZF precoding matrix $\mathbf{G}=\beta\mathbf{H}\mathbf{X}^{(Q)}$
for $k=0,\ldots,Q-1$ **do**
  randomly sample q_i column indexes according to (27)
  update $\mathbf{X}^{(k+1)}$ according to (28)
end for
output $\mathbf{G}=\beta\mathbf{H}\mathbf{X}^{(Q)}$

Therefore, by updating $\mathbf{X}^{(k)}$ iteratively according to (28), $\mathbf{A}^{-1}$ can be asymptotically approximated. As for the initial choice of $\mathbf{X}^{(0)}$, it actually can be set arbitrarily and we take $\mathbf{X}^{(0)}=\mathbf{D}^{-1}$ as an alternative. To summarize, the operations of the proposed randomized iterative precoding algorithm (RIPA) for downlink massive MIMO systems is outlined in Algorithm 1.

We then consider the computational complexity of RIPA. In particular, as for the iteration in (28), the computational complexity of calculating $(\mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H\mathbf{A}\mathbf{X}^{(k)}_{:,q_i's})^{-1}$ is $q^3+q^2K+2qK^2$; the computational complexity of $\mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H-\mathbf{X}^{(k)H}_{:,q_i's}\mathbf{A}^H\mathbf{A}\mathbf{X}^{(k)}$ is q^2K+3qK^2 while multiplying it with the former terms costs qK^2+q^2K. Therefore, the total computational complexity of RIPA at each iteration can be approximated by $q^3+3q^2K+6qK^2$, which can be expressed as $O(qK^2)$. Considering $q\ll K$, it turns out to be competitive compared to other low-complexity precoding schemes.

B. Convergence Analysis

We now study the convergence of RIPA in terms of the expectation of the error norm, i.e., $E[\|\mathbf{X}^{(k)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})}]$. For the sake of convenience, let $\mathbf{M}=[\mathbf{M}_1,\ldots,\mathbf{M}_r]\in\mathbb{C}^{K\times\sum_{i=1}^r q_i}$ and assume $\mathbf{M}$ is full row rank. Actually, because the discrete distribution $\mathcal{D}$ can be designed arbitrarily, this weak assumption is easily achieved [46].

According to (24), the random ingredients are contained in the matrix $\mathbf{Z}$. Specifically, given the matrix $\mathbf{S}_k$ randomly sampled from $\mathcal{D}$, the expectation of $\mathbf{Z}$ can be derived as

$$\begin{aligned}E[\mathbf{Z}] &= \sum_{i=1}^r p_i\mathbf{A}^H\mathbf{M}_i(\mathbf{M}_i^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{M}_i)^{-1}\mathbf{M}_i^H\mathbf{A}\\ &= \mathbf{A}^H\left(\sum_{i=1}^r\mathbf{M}_i p_i^{\frac{1}{2}}(\mathbf{M}_i^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{M}_i)^{-\frac{1}{2}}\right.\\ &\quad\left.\times(\mathbf{M}_i^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{M}_i)^{-\frac{1}{2}}p_i^{\frac{1}{2}}\mathbf{M}_i^H\right)\mathbf{A}\\ &= (\mathbf{A}^H\mathbf{M}\mathbf{J})(\mathbf{J}\mathbf{M}^H\mathbf{A})\end{aligned}\tag{29}$$

with the invertible block diagonal matrix $\mathbf{J}\in\mathbb{C}^{K\times K}$

$$\begin{aligned}\mathbf{J}=\mathrm{diag}(&p_1^{\frac{1}{2}}(\mathbf{M}_1^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{M}_1)^{-\frac{1}{2}},\ldots,p_r^{\frac{1}{2}}\\ &\times(\mathbf{M}_r^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{M}_r)^{-\frac{1}{2}}).\end{aligned}\tag{30}$$

Intuitively, even under the random sampling, $E[\mathbf{Z}]$ is still symmetric by structure.

On the other hand, considering the Rayleigh fading channels in massive MIMO systems, because the channel matrix $\mathbf{H}$ is full rank matrix, the multiplication $\mathbf{H}\mathbf{v}$ for vector $\mathbf{v}\in\mathbb{C}^K$ equals to $\mathbf{0}$ if and only if $\mathbf{v}$ is a zero vector. Therefore, it is straightforward to check that the Gram matrix $\mathbf{H}^H\mathbf{H}$ in matrix $\mathbf{A}$ is positive definite due to

$$\mathbf{v}^H(\mathbf{H}^H\mathbf{H})\mathbf{v}=(\mathbf{H}\mathbf{v})^H\mathbf{H}\mathbf{v}>0,\tag{31}$$

which leads to a positive definite matrix $\mathbf{A}=\mathbf{H}^H\mathbf{H}+\xi\mathbf{I}$. Moreover, given the full row rank matrix $\mathbf{M}_i^H\mathbf{A}$ and the invertible diagonal matrix $\mathbf{J}$, we can verify that $E[\mathbf{Z}]$ in (29) is symmetric positive definite.

Based on the symmetric positive definite $E[\mathbf{Z}]$, the following results about the convergence of RIPA can be demonstrated while the proof partially follows [47].

Theorem 1: With respect to the downlink precoding in massive MIMO systems, let $\mathbf{S}_k$ be randomly sampled from the discrete distribution $\mathcal{D}$, the proposed randomized iteration following (23) converges by

$$E[\|\mathbf{X}^{(k)}-\mathbf{A}^{-1}\|^2_{F(\mathbf{W}^{-1})}]\le\rho^k\|\mathbf{X}^{(0)}-\mathbf{A}^{-1}\|^2_{F(\mathbf{W}^{-1})}\tag{32}$$

with exponential convergence rate

$$\rho=1-\lambda_{\min}(\mathbf{W}^{\frac{1}{2}}E(\mathbf{Z})\mathbf{W}^{\frac{1}{2}})<1,\tag{33}$$

where $\lambda_{\min}(\cdot)$ stands for the minimum eigenvalue of a matrix.

Proof: To start with, in order to concisely state the result, the following definitions are made

$$\mathbf{R}_k=\mathbf{W}^{-\frac{1}{2}}(\mathbf{X}^{(k)}-\mathbf{A}^{-1})\mathbf{W}^{-\frac{1}{2}}\tag{34}$$

and

$$\widehat{\mathbf{Z}}=\mathbf{W}^{\frac{1}{2}}\mathbf{Z}\mathbf{W}^{\frac{1}{2}}.\tag{35}$$

Then, by multiplying $\mathbf{W}^{-\frac{1}{2}}$ on the both sides of (24), it follows that

$$\mathbf{R}_{k+1}=(\mathbf{I}-\widehat{\mathbf{Z}})\mathbf{R}_k,\tag{36}$$

and from which we can have

$$\begin{aligned}E[\|\mathbf{X}^{(k+1)}-\mathbf{A}^{-1}\|^2_{F(\mathbf{W}^{-1})}] &= E[\|\mathbf{R}_{k+1}\|_F^2]\\ &\overset{(a)}{=} E[E[\|\mathbf{R}_{k+1}\|_F^2|\mathbf{R}_k]]\end{aligned}\tag{37}$$

where equality (a) holds according to the *law of total probability* for expectation (i.e., $E[E(A|B)]=E(A)$).

Then, regarding to the term $\|\mathbf{R}_{k+1}\|_F^2$ in the above equation, we have the following derivations

$$\begin{aligned}\|\mathbf{R}_{k+1}\|_F^2 &= \|(\mathbf{I}-\widehat{\mathbf{Z}})\mathbf{R}_k\|_F^2\\ &\overset{(b)}{=} \mathrm{Tr}((\mathbf{I}-\widehat{\mathbf{Z}})\mathbf{R}_k\mathbf{R}_k^H(\mathbf{I}-\widehat{\mathbf{Z}})^H)\\ &\overset{(c)}{=} \mathrm{Tr}((\mathbf{I}-\widehat{\mathbf{Z}})(\mathbf{I}-\widehat{\mathbf{Z}})\mathbf{R}_k\mathbf{R}_k^H)\\ &\overset{(d)}{=} \mathrm{Tr}((\mathbf{I}-\widehat{\mathbf{Z}})\mathbf{R}_k\mathbf{R}_k^H)\\ &= \|\mathbf{R}_k\|_F^2-\mathrm{Tr}(\widehat{\mathbf{Z}}\mathbf{R}_k\mathbf{R}_k^H).\end{aligned}\tag{38}$$

Here, equality (b) comes from (16), equality (c) holds due to the symmetry of matrices $\mathbf{W}$ and $\widehat{\mathbf{Z}}$. Besides, equality (d) comes from the fact that $\widehat{\mathbf{Z}}\widehat{\mathbf{Z}} = \widehat{\mathbf{Z}}$ due to

$$\begin{aligned}&\underbrace{\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\mathbf{S}_k(\mathbf{S}_k^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H\mathbf{A}\mathbf{W}^{\frac{1}{2}}}_{\widehat{\mathbf{Z}}}\\ &\quad\times \underbrace{\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\mathbf{S}_k(\mathbf{S}_k^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H\mathbf{A}\mathbf{W}^{\frac{1}{2}}}_{\widehat{\mathbf{Z}}}\\ &= \mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\mathbf{S}_k(\mathbf{S}_k^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H\mathbf{A}\mathbf{W}^{\frac{1}{2}}\\ &= \mathbf{W}^{\frac{1}{2}}\mathbf{Z}\mathbf{W}^{\frac{1}{2}}\\ &= \widehat{\mathbf{Z}}. \end{aligned}\tag{39}$$

Next, based on (38), it follows that

$$\begin{aligned}E[\|\mathbf{R}_{k+1}\|_F^2|\mathbf{R}_k] &= \|\mathbf{R}_k\|_F^2 - \mathrm{Tr}(E[\widehat{\mathbf{Z}}]\mathbf{R}_k\mathbf{R}_k^H)\\ &\overset{(e)}{\leq} \|\mathbf{R}_k\|_F^2 - \lambda_{\min}(E[\widehat{\mathbf{Z}}])\mathrm{Tr}(\mathbf{R}_k\mathbf{R}_k^H)\\ &= (1-\lambda_{\min}(E[\widehat{\mathbf{Z}}])\|\mathbf{R}_k\|_F^2\\ &= (1-\lambda_{\min}(\mathbf{W}^{\frac{1}{2}}E(\mathbf{Z})\mathbf{W}^{\frac{1}{2}})\|\mathbf{R}_k\|_F^2\\ &= \rho\|\mathbf{R}_k\|_F^2\end{aligned}\tag{40}$$

where the inequality (e) relies on the symmetric property of $E[\widehat{\mathbf{Z}}]$.

After that, by simply substituting (40) into (37), we can easily arrive at the following result by induction

$$\begin{aligned}E[\|\mathbf{X}^{(k)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})}^2] &\leq \rho\|\mathbf{R}_{k-1}\|_F^2\\ &= \rho E[\|\mathbf{X}^{(k-1)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})}^2]\\ &\leq \rho^k E[\|\mathbf{X}^{(0)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})}^2]\\ &= \rho^k\|\mathbf{X}^{(0)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})},\end{aligned}\tag{41}$$

where the matrix $\mathbf{X}^{(0)}$ is given at the beginning as an initial setup.

Moreover, since $E(\mathbf{Z})$ is a symmetric positive definite matrix in massive MIMO systems, all the eigenvalues of it are positive, i.e. $\lambda_{\min}(\mathbf{E}(\mathbf{Z})) > 0$. Therefore, we have

$$\rho = 1-\lambda_{\min}(\mathbf{W}^{\frac{1}{2}}\mathbf{E}(\mathbf{Z})\mathbf{W}^{\frac{1}{2}}) < 1 \tag{42}$$

because the weighted matrix $\mathbf{W}$ is also symmetric positive definite. ∎

By Theorem 1, the proposed randomized iteration in RIPA converges exponentially fast to $\mathbf{A}^{-1}$. More importantly, we point out that such an exponential convergence works globally without suffering from the requirement about the ratio N/K. Meanwhile, the requirements about system parameters $\mathbf{W}$ and $\mathcal{D}$ are easy to achieve. Moreover, in order to guarantee the approximation error smaller than a given value $0 < \epsilon < 1$

$$E[\|\mathbf{X}^{(k)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})}^2] \leq \epsilon\|\mathbf{X}^{(0)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})}^2, \tag{43}$$

the required number of iterations is lower bounded by[2]

$$k \geq \frac{1}{1-\rho}\log(\frac{1}{\epsilon}) \tag{44}$$

[2]The inequality $\ln(1-\delta) < -\delta$ for $0 < \delta < 1$ is applied here.

so as to a tractable iteration. Besides, according to (32), a closer choice of $\mathbf{X}^{(0)}$ to the target solution $\mathbf{A}^{-1}$ is beneficial to boost the convergence, which is rather recommended in practice.

IV. Optimization and Enhancement

In this section, the iteration transformation with respect to the randomized iteration is firstly proposed for the sake of efficiency. Then, by adopting the conditional sampling into iterations, the previous samplings from $\mathcal{D}$ can be learnt by the current randomized iteration, where considerable convergence gain can be exploited. Moreover, the mechanism driven by conditional sampling can be further strengthened to a pseudorandom iteration, which leads to better convergence and efficiency.

A. Iteration Transformation

Given (24), it is possible to convert the iteration about matrix $\mathbf{X}$ into the iteration about vector $\mathbf{t}$ by right multiplying $\mathbf{s}$ as

$$\mathbf{t}^{(k+1)} - \mathbf{A}^{-1}\mathbf{s} = (\mathbf{I}-\mathbf{W}\mathbf{Z})(\mathbf{t}^{(k)}-\mathbf{A}^{-1}\mathbf{s}). \tag{45}$$

By doing this, the approximation of $\mathbf{A}^{-1}\mathbf{s}$ would be outputted by $\mathbf{t}^{(k+1)}$ in an iterative way, which follows the counterpart of the system model shown in (9). Similar to (15), this corresponds to solving the linear system shown below

$$\mathbf{S}^H\mathbf{A}\mathbf{t} = \mathbf{S}^H\mathbf{s} \tag{46}$$

with the auxiliary matrix $\mathbf{S} \in \mathbb{C}^{K\times q}$.

Subsequently, based on (45), the related iteration in (23) naturally becomes

$$\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + \mathbf{W}\mathbf{A}^H\mathbf{S}_k(\mathbf{S}_k^H\mathbf{A}\mathbf{W}\mathbf{A}^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H(\mathbf{s}-\mathbf{A}\mathbf{t}^{(k)}) \tag{47}$$

with $\mathbf{S}_k \sim \mathcal{D}$. More specifically, it is straightforward to verify that the exponential convergence of the iteration in (47) still holds globally, and we have the following result with omitted proof due to simplicity.

Corollary 1: With respect to the downlink precoding in massive MIMO systems, let $\mathbf{S}_k$ be randomly sampled from the discrete distribution $\mathcal{D}$, the proposed randomized iteration following (47) converges by

$$E[\|\mathbf{t}^{(k)}-\mathbf{A}^{-1}\mathbf{s}\|_{F(\mathbf{W}^{-1})}^2] \leq \rho^k\|\mathbf{t}^{(0)}-\mathbf{A}^{-1}\mathbf{s}\|_{F(\mathbf{W}^{-1})}^2 \tag{48}$$

with exponential convergence rate

$$\rho = 1-\lambda_{\min}(\mathbf{W}^{\frac{1}{2}}E(\mathbf{Z})\mathbf{W}^{\frac{1}{2}}) < 1, \tag{49}$$

where $\lambda_{\min}(\cdot)$ stands for the minimum eigenvalue of a matrix.

It is clear that both the iterations in (23) and (47) regarding to $\mathbf{X}$ and $\mathbf{t}$ respectively are able to provide the effective solutions to the downlink precoding even with the same convergence expression. Note that the convergence rates ρ in (53) and (49) are the same, where the difference of convergence performance lies on the different Frobenius norms of $\|\mathbf{X}^{(0)}-\mathbf{A}^{-1}\|_{F(\mathbf{W}^{-1})}$ and $\|\mathbf{t}^{(0)}-\mathbf{A}^{-1}\mathbf{s}\|_{F(\mathbf{W}^{-1})}$. Nevertheless, we claim that the latter one is a better choice since the computational complexity of each iteration can be reduced. In particular, the computational complexities of $\mathbf{s}-\mathbf{At}^{(k)}$ in (47) is much smaller than $\mathbf{I}-\mathbf{AX}^{(k)}$ in (23), thus leading to a more efficient iteration scheme. For this reason, the following optimization and enhancement are carried out based on the iteration transformation about $\mathbf{t}$ in (45).

B. Optimization by Conditional Random Sampling

Although the random sampling of $\mathbf{S}_k$ from $\mathcal{D}$ offers an effective way for the linear system in (46), it does have a side effect during the iteration. Because the randomness during the sampling is difficult to control, the sampling diversity could be constrained if a certain sampling choice $\mathbf{M}_i$ is obtained repeatedly by two consecutive sampling operations $\mathbf{S}_{k-1}$ and $\mathbf{S}_k$. This is straightforward to understand since the convergence comes from various choices of $\mathbf{S}_k$ rather than a single one.

In order to overcome such an issue, it is possible to update the sampling probability in (19) as a conditional one, so that the prior knowledge from the last sampling can be utilized, i.e.,

$$\begin{aligned}\overline{p}_i &\triangleq \mathcal{D}(\mathbf{S}_k=\mathbf{M}_i|\mathbf{S}_{k-1}=\mathbf{M}_j), \quad i\neq j\\ &= \frac{p_i}{1-p_j}, \quad i\neq j.\end{aligned} \tag{50}$$

Intuitively, from (50), the sampling $\mathbf{S}_{k-1}$ of the last iteration is taken into account by the sampling at the current iteration, where the sampling choice $\mathbf{M}_j$ is discarded by the current sampling for $\mathbf{S}_k$. In this way, the aforementioned issue that $\mathbf{S}_{k-1}=\mathbf{S}_k$ can be avoided, which results in a better convergence for the randomized iterations.

We now consider the convergence of the conditional randomized iteration. Typically, according to $\overline{p}_i$ in (50), the conditional expectation of $\mathbf{Z}$ given the last sampling choice $\mathbf{S}_{k-1}=\mathbf{M}_j$ can be written as

$$\begin{aligned}E[\mathbf{Z}|\mathbf{S}_{k-1}] &= \sum_{i=1}^{j-1}\overline{p}_i\mathbf{A}^H\mathbf{M}_i(\mathbf{M}_i^H\mathbf{AWA}^H\mathbf{M}_i)^{-1}\mathbf{M}_i^H\mathbf{A}\\ &\quad+\sum_{i=j+1}^{r}\overline{p}_i\mathbf{A}^H\mathbf{M}_i(\mathbf{M}_i^H\mathbf{AWA}^H\mathbf{M}_i)^{-1}\mathbf{M}_i^H\mathbf{A}\\ &= (\mathbf{A}^H\overline{\mathbf{M}}\overline{\mathbf{J}})(\overline{\mathbf{J}}\overline{\mathbf{M}}^H\mathbf{A})\in\mathbb{C}^{K\times K},\end{aligned} \tag{51}$$

which still turns out to be symmetric definite positive with $\overline{\mathbf{M}}=[\mathbf{M}_1,\ldots,\mathbf{M}_{j-1},\mathbf{M}_{j+1},\ldots,\mathbf{M}_r]$ and $\overline{\mathbf{J}}=\text{diag}(\overline{p}_1^{\frac{1}{2}}(\mathbf{M}_1^H\mathbf{AWA}^H\mathbf{M}_1)^{-\frac{1}{2}},\ldots,\overline{p}_{j-1}^{\frac{1}{2}}(\mathbf{M}_{j-1}^H\mathbf{AWA}^H\mathbf{M}_{j-1})^{-\frac{1}{2}},\overline{p}_{j+1}^{\frac{1}{2}}(\mathbf{M}_{j+1}^H\mathbf{AWA}^H\mathbf{M}_{j+1})^{-\frac{1}{2}},\ldots,\overline{p}_r^{\frac{1}{2}}(\mathbf{M}_r^H\mathbf{AWA}^H\mathbf{M}_r)^{-\frac{1}{2}})$. Then, it is clear to confirm the convergence of the conditional randomized iteration so that the following Theorem can be achieved, where the related proof is omitted because of simplicity.

Theorem 2: Given the sampling choice $\mathbf{S}_{k-1}=\mathbf{M}_j$, $\mathbf{M}_j\in\{\mathbf{M}_1,\ldots,\mathbf{M}_r\}$, let $\mathbf{S}_k$ be randomly sampled from $\mathcal{D}$ according to the conditional sampling probability $\overline{p}_i$ defined in (50), then the conditional randomized iteration following (47) converges by

$$E[\|\mathbf{t}^{(k)}-\mathbf{A}^{-1}\mathbf{s}\|^2_{F(\mathbf{W}^{-1})}]\leq\overline{\rho}\|\mathbf{t}^{(k-1)}-\mathbf{A}^{-1}\mathbf{s}\|^2_{F(\mathbf{W}^{-1})} \tag{52}$$

with exponential convergence rate

$$\overline{\rho}=1-\lambda_{\min}(\mathbf{W}^{\frac{1}{2}}E(\mathbf{Z}|\mathbf{S}_{k-1})\mathbf{W}^{\frac{1}{2}})<1. \tag{53}$$

Note that the convergence rate $\overline{\rho}$ varies at each iteration given the conditional sample $\mathbf{S}_{k-1}$. After that, regarding to the convergence of the conditional randomized iteration, we can arrive at the following result by optimization.

Theorem 3: Given the sampling choice $\mathbf{S}_{k-1}=\mathbf{M}_j$, $\mathbf{M}_j\in\{\mathbf{M}_1,\ldots,\mathbf{M}_r\}$, the convergence rate of the conditional randomized iteration takes the form

$$\overline{\rho}=1-\frac{\lambda_{\min}(\overline{\mathbf{M}}^H\mathbf{AWA}^H\overline{\mathbf{M}})}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2} \tag{54}$$

if the sampling probability $\overline{p}_i$ follows

$$\overline{p}_i=\frac{\text{Tr}(\mathbf{M}_i^H\mathbf{AWA}^H\mathbf{M}_i)}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2}, \quad i\neq j. \tag{55}$$

Proof: To start with, for the sake of notational simplicity, let $t_i=\text{Tr}(\mathbf{M}_i^H\mathbf{AWA}^H\mathbf{M}_i)$ and $y_i=(\mathbf{M}_i^H\mathbf{AWA}^H\mathbf{M}_i)^{-1}$, then we have

$$\overline{\mathbf{J}}^2=\frac{\text{diag}(t_1y_1,\ldots,t_{j-1}y_{j-1},t_{j+1}y_{j+1},\ldots,t_ry_r)}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2} \tag{56}$$

so that

$$\begin{aligned}\lambda_{\min}(\overline{\mathbf{J}}^2) &= \frac{1}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2}\min_{i\neq j}\left\{\frac{t_i}{\lambda_{\max}(\mathbf{M}_i^H\mathbf{AWA}^H\mathbf{M}_i)}\right\}\\ &\overset{(f)}{\geq}\frac{1}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2},\end{aligned} \tag{57}$$

where inequality (f) holds due to the fact that the trace of a matrix equals the sum of its eigenvalues, i.e.,

$$\text{Tr}(\mathbf{A})=\sum_i\lambda_i(\mathbf{A}) \quad \text{and } \text{Tr}(\mathbf{A})\geq\lambda_{\max}(\mathbf{A})\geq\lambda_{\min}(\mathbf{A}). \tag{58}$$

Therefore, according to (57) and (51), we can arrive at the following derivation

$$\begin{aligned}\lambda_{\min}(\mathbf{W}^{\frac{1}{2}}E(\mathbf{Z}|\mathbf{S}_{k-1})\mathbf{W}^{\frac{1}{2}}) &= \lambda_{\min}(\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\overline{\mathbf{J}}^2\overline{\mathbf{M}}^H\mathbf{AW}^{\frac{1}{2}})\\ &= \lambda_{\min}(\overline{\mathbf{M}}^H\mathbf{AWA}^H\overline{\mathbf{M}}\overline{\mathbf{J}}^2)\\ &\overset{(g)}{\geq}\lambda_{\min}(\overline{\mathbf{M}}^H\mathbf{AWA}^H\overline{\mathbf{M}})\lambda_{\min}(\overline{\mathbf{J}}^2)\\ &\geq\frac{\lambda_{\min}(\overline{\mathbf{M}}^H\mathbf{AWA}^H\overline{\mathbf{M}})}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2},\end{aligned} \tag{59}$$

where (g) holds because $\lambda_{\min}(\mathbf{EF})\geq\lambda_{\min}(\mathbf{E})\lambda_{\min}(\mathbf{F})$ if matrices $\mathbf{E}$, $\mathbf{F}\in\mathbb{C}^{K\times K}$ are both positive definite. Intuitively, considering the fact that the system parameter $\mathbf{W}$ is symmetric positive definite by default, the matrix multiplication $\overline{\mathbf{M}}^H\mathbf{AWA}^H\overline{\mathbf{M}}$ is also positive definite, and so is the diagonal matrix $\overline{\mathbf{J}}^2$.

Finally, by substituting (59) into (53), we can easily obtain that

$$\overline{\rho} = 1 - \frac{\lambda_{\min}(\overline{\mathbf{M}}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \overline{\mathbf{M}})}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2} \tag{60}$$

with the sampling probability

$$\overline{p}_i = \frac{\mathrm{Tr}(\mathbf{M}_i^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M}_i)}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2}, \quad i \neq j, \tag{61}$$

completing the proof. ■

Based on Theorem 3, it is possible to improve the convergence rate $\overline{\rho}$ through optimizing the choice of the matrix $\overline{\mathbf{M}} = [\mathbf{M}_1, \ldots, \mathbf{M}_{j-1}, \mathbf{M}_{j+1}, \ldots, \mathbf{M}_r]$. More specifically, the convergence rate $\overline{\rho}$ can be lower bounded as

$$\begin{aligned}\overline{\rho} &= 1 - \frac{\lambda_{\min}(\overline{\mathbf{M}}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \overline{\mathbf{M}})}{\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\overline{\mathbf{M}}\|_F^2} \\ &= 1 - \frac{\lambda_{\min}(\overline{\mathbf{M}}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \overline{\mathbf{M}})}{\mathrm{Tr}(\overline{\mathbf{M}}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \overline{\mathbf{M}})} \\ &\geq 1 - \frac{1}{\sum_{i=1}^{j-1} q_i + \sum_{i=j+1}^{r} q_i},\end{aligned} \tag{62}$$

where the lower bound in (62) holds if and only if $\overline{\mathbf{M}}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \overline{\mathbf{M}} = \mathbf{I}$. On the other hand, with respect to the original randomized iteration in RIPA, the lower bound of its convergence rate ρ given sampling probability $p_i = \mathrm{Tr}(\mathbf{M}_i^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M}_i)/\|\mathbf{W}^{\frac{1}{2}}\mathbf{A}^H\mathbf{M}\|_F^2$ can be derived by simple substitution

$$\rho = 1 - \frac{\lambda_{\min}(\mathbf{M}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M})}{\mathrm{Tr}(\mathbf{M}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M})} \geq 1 - \frac{1}{\sum_{i=1}^{r} q_i}, \tag{63}$$

where the lower bound is achieved when $\mathbf{M}^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M} = \mathbf{I}$. Moreover, due to $\sum_{i=1}^{j-1} q_i + \sum_{i=j+1}^{r} q_i < \sum_{i=1}^{r} q_i = K$, we have the following result.

Corollary 2: According to (62) and (63), the conditional randomized iteration outperforms the randomized iteration due to a smaller convergence lower bound.

C. Enhancement by Multi-Step Conditional Sampling

Given the convergence gain brought by the conditional sampling, a straightforward enhancement for randomized iterations is to recall more previous samplings, namely,

$$\overline{p}_L \triangleq \mathcal{D}(\mathbf{S}_k = \mathbf{M}_i | \mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-L}) \tag{64}$$

where $1 \leq L \leq r-1$ indicates the number of previous samplings with $\mathbf{M}_i \notin \{\mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-L}\}$. In this way, the introduced conditional randomized iteration can be actually viewed as a special case of the multi-step conditional randomized iteration with $L = 1$. Similarly, it is straightforward to verify the exponential convergence of L-step conditional randomized iteration. To this end, we have the following Corollary, where its proof is omitted.

Corollary 3: Given L-step sampling choices $\mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-L}$, let $\mathbf{S}_k$ be randomly sampled according to the conditional sampling probability $\overline{p}_L$ in (64), the L-step conditional randomized iteration following (47) converges by

$$E[\|\mathbf{t}^{(k)} - \mathbf{A}^{-1}\mathbf{s}\|_{F(\mathbf{W}^{-1})}^2] \leq \overline{\rho}_L \|\mathbf{t}^{(k-1)} - \mathbf{A}^{-1}\mathbf{s}\|_{F(\mathbf{W}^{-1})}^2 \tag{65}$$

with exponential convergence rate

$$\overline{\rho}_L = 1 - \lambda_{\min}(\mathbf{W}^{\frac{1}{2}} E(\mathbf{Z}|\mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-L})\mathbf{W}^{\frac{1}{2}}) < 1. \tag{66}$$

Meanwhile, following Theorem 3, it is clear to see that the convergence performance of L-step conditional randomized iteration improves gradually with the increase of L. Therefore, it is encouraged to set $L = r-1$ to fully take advantages of conditional sampling by

$$\overline{\rho}_{r-1} = 1 - \frac{\lambda_{\min}(\mathbf{M}_i^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M}_i)}{\mathrm{Tr}(\mathbf{M}_i^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M}_i)} \tag{67}$$

$$\geq 1 - \frac{1}{q_i}. \tag{68}$$

Clearly, the lower bound of $\overline{\rho}_{r-1}$ is smaller than that of $\overline{\rho}_1$ in (62), which leads to a better convergence performance. Apart from the convergence gain, with the increase of L, the L-step conditional randomized iteration gradually becomes deterministic in selection of $\mathbf{S}_k$. More specifically, when $k > r-1$ and $L = r-1$, there is only one sampling option left for $\mathbf{S}_k$ given these $r-1$ previous sampling choices of $\mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-r+1}$. Interestingly, such a pseudorandom or derandomized sampling is also beneficial to the implementation of L-step conditional randomized iteration, making it more efficient in practice. Therefore, with $L = r-1$, the following modified randomized iterative precoding algorithm (MRIPA) based on the multi-step conditional randomized iteration is proposed based on the randomized iteration in (47).

V. Modified Randomized Iterative Precoding Algorithm

In this section, based on iteration transformation and multi-step conditional sampling, the modified randomized iterative precoding algorithm (MRIPA) is proposed for a better convergence and efficiency.

A. Algorithm Description

In particular, from (68), in order to obtain the convergence lower bound of $\overline{\rho}_{r-1}$ in L-step conditional randomized iteration, the condition $\mathbf{M}_i^H \mathbf{A}\mathbf{W}\mathbf{A}^H \mathbf{M}_i = \mathbf{I}$ given $\mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-L}$ with $\mathbf{M}_i \notin \{\mathbf{S}_{k-1}, \ldots, \mathbf{S}_{k-L}\}$ should be fulfilled. Therefore, a straightforward way to simplify the randomized iteration is to apply $\mathbf{W} = \mathbf{A}^{-1}$ and $\mathbf{M}_i^H \mathbf{A}^H \mathbf{M}_i = \mathbf{I}$ respectively. Unfortunately, finding the optimal $\mathbf{M}_i = \mathbf{A}^{-\frac{1}{2}}\mathbf{I}_{:,q_i's} = \mathbf{A}_{:,q_i's}^{-\frac{1}{2}} \in \mathbb{C}^{K \times q_i}$ is hard to realize in practice. Therefore, given the symmetric positive matrix $\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{L}^H$, we use the diagonal matrix $\mathbf{D}$ as an approximation, namely,

$$\overline{\mathbf{M}}_i = \mathbf{D}^{-\frac{1}{2}}\mathbf{I}_{:,q_i's} \in \mathbb{C}^{K \times q_i}. \tag{69}$$

Consequently, the related convergence rate can be derived as

$$\overline{\rho}_{r-1} = 1 - \frac{\lambda_{\min}(\overline{\mathbf{M}}_i^H \mathbf{A}\overline{\mathbf{M}}_i)}{\mathrm{Tr}(\overline{\mathbf{M}}_i^H \mathbf{A}\overline{\mathbf{M}}_i)}. \tag{70}$$

Intuitively, the above convergence rate obeys the lower bound in (68) as well. Similar to RIPA, the setup $q_1 = \ldots = q_r = q$ is also applied in the proposed MRIPA for the sake of simplicity.

To summarize, based on the $r-1$ previous samplings $\mathbf{S}_{k-1},\ldots,\mathbf{S}_{k-r+1}$, the sampling choice $\mathbf{S}_k = \overline{\mathbf{M}}_i$ can be efficiently determined especially when $k > r-1$, and the randomized iteration of MRIPA with $\mathbf{W}=\mathbf{A}^{-1}$ becomes

$$\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + \overline{\mathbf{M}}_i(\overline{\mathbf{M}}_i^H\mathbf{A}\overline{\mathbf{M}}_i)^{-1}(\overline{\mathbf{M}}_i^H\mathbf{s} - \overline{\mathbf{M}}_i^H\mathbf{A}\mathbf{t}^{(k)}) \tag{71}$$

with

$$\overline{\mathbf{M}}_i \notin \{\mathbf{S}_{k-1},\ldots,\mathbf{S}_{k-r+1}\}. \tag{72}$$

On the other hand, according to (52) in Theorem 2 and (65) in Corollary 2, a closer choice of $\mathbf{t}^{(0)}$ to $\mathbf{A}^{-1}\mathbf{s}$ is able to significantly accelerate the convergence, which leads to a more efficient randomized iterative precoding. For this reason, the matrix $\mathbf{D}$ can be applied here to serve as the initial choice of $\mathbf{t}^{(0)}$, i.e.,

$$\mathbf{t}^{(0)} = \mathbf{D}^{-1}\mathbf{s}, \tag{73}$$

where the approximation improves with the increase of the ratio N/K and vice versa. Note that because $\mathbf{D}$ is a diagonal matrix, the computational complexity of $\mathbf{t}^{(0)}$ is rather low, which only requires K multiplications. Overall, the operation procedures of the proposed MRIPA for downlink massive MIMO systems are outlined in Algorithm 2. In addition, we point out that the usage of multi-step conditional sampling is also compatible to the proposed RIPA scheme despite of the different choices of system parameters $\mathbf{W}$ and $\mathbf{S}_k$.

Algorithm 2 Modified Randomized Iterative Precoding Algorithm (MRIPA) for Downlink Massive MIMO Systems

Require: $\mathbf{A}=\mathbf{H}^H\mathbf{H}$ or $\mathbf{A}=\mathbf{H}^H\mathbf{H}+\xi\mathbf{I}$, $\mathbf{t}^{(0)}=\mathbf{D}^{-1}\mathbf{s}$, Q, β
Ensure: near ZF or RZF precoding solution $\mathbf{G}\mathbf{s}=\beta\mathbf{H}\mathbf{t}^{(Q)}$
1: **for** $k=0,\ldots,Q-1$ **do**
2: randomly sample q_i column indexes according to (72)
3: update $\mathbf{t}^{(k+1)}$ according to (71)
4: **end for**
5: output $\mathbf{G}\mathbf{s}=\beta\mathbf{H}\mathbf{t}^{(Q)}$

B. Complexity Reduction of MRIPA

As for the computational complexity of MRIPA, the computational complexities of computing $(\overline{\mathbf{M}}_i^H\mathbf{A}\overline{\mathbf{M}}_i)^{-1}$ and $\overline{\mathbf{M}}_i^H\mathbf{s} - \overline{\mathbf{M}}_i^H\mathbf{A}\mathbf{t}^{(k)}$ in (71) are $q^2K+qK^2+q^3$ and $2qK+qK^2$ respectively, and multiplying these several terms costs q^2K+qK. Therefore, the total complexity of MRIPA at each iteration can be approximated by $3qK+q^3+2q^2K+2qK^2$, which is smaller than that of RIPA. Moreover, in the following we show that the computational complexity of MRIPA can be further reduced by well taking advantages of the initial setup given in (73).

According to (69), the matrix $\overline{\mathbf{M}}_i$ has a special structure, which can be further expressed as $\overline{\mathbf{M}}_i^H$, shown at the bottom of the page. Clearly, the operations of $\overline{\mathbf{M}}_i$ or $\overline{\mathbf{M}}_i^H$ are essentially performed by the $q\times q$ nonzero submatrix within it. Meanwhile, we can observe that the $q\times q$ submatrix only contains the nonzero diagonal elements, where the rest of elements in it are also 0. Therefore, further complexity reduction can be achieved by well exploiting these special structures of $\overline{\mathbf{M}}_i$.

In particular, the computational complexity of calculating $(\overline{\mathbf{M}}_i^H\mathbf{A}\overline{\mathbf{M}}_i)^{-1}$ can be reduced to $qK+q^2+q^3$; the computational complexity of $\overline{\mathbf{M}}_i^H\mathbf{s} - \overline{\mathbf{M}}_i^H\mathbf{A}\mathbf{t}^{(k)}$ is $q+2qK$ and multiplying them together costs q^2+qK. To summarize, the total complexity of MRIPA at each iteration is reduced as $q+q^2+q^3+4qK$. Meanwhile, to fulfill the requirement of $q \ll K$ in practice, we set $1<q\leq\sqrt{K}$ as a solution, which means the reduced computational complexity of MRIPA at each iteration is no more than $O(K^{1.5})$. We point out that the average complexity of each iteration in RIPA can also be reduced as $O(K^2)$ in the same way. According to (71), in the proposed MRIPA only q components of $\mathbf{t}$ are updated at each iteration. Therefore, for a fair comparison, K/q times iterations are needed as a full iteration to update all the K elements of $\mathbf{t}$. Nevertheless, the complexity of MRIPA in a full iteration is $O(K^2)$ with $1<q\leq\sqrt{K}$, making it still competitive compared to traditional iterative precoding schemes.

VI. Simulations

In this section, the proposed randomized iterative precoding schemes for downlink massive MIMO systems is examined by simulations. Here, we assume a flat fading environment, and the channel matrix $\mathbf{H}\in\mathbb{C}^{N\times K}$ is perfectly known at the base station, which contains uncorrelated complex Gaussian fading gains with unit variance and remains constant over each frame duration.

In Fig. 1, the proposed randomized iterative precoding algorithm (RIPA) is evaluated in a uncoded massive MIMO system with $N=128$ and $K=32$, where the average achievable rate per user terminal (i.e., rate $=\frac{1}{K}\sum_i^K E[\log_2(1+\text{SINR}_i)]$) is applied as the comparison measurement. More precisely, given the corresponding precoding matrix $\mathbf{G}$, the signal to interference and noise ratio (SINR) at the i-th user is expressed as [12], [48]

$$\text{SINR}_i = \frac{\mathbf{h}_i^H\mathbf{g}_i\mathbf{g}_i^H\mathbf{h}_i}{\mathbf{h}_i^H\mathbf{G}\mathbf{G}^H\mathbf{h}_i+\sigma^2}, \tag{74}$$

$$\overline{\mathbf{M}}_i^H = \begin{bmatrix} 0 & \cdots & 0 & m_{1,(i-1)*q+1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & 0 & \ddots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \ddots & \ddots & 0 & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & m_{q,(i-1)*q+q} & 0 & \cdots & 0 \end{bmatrix}.$$

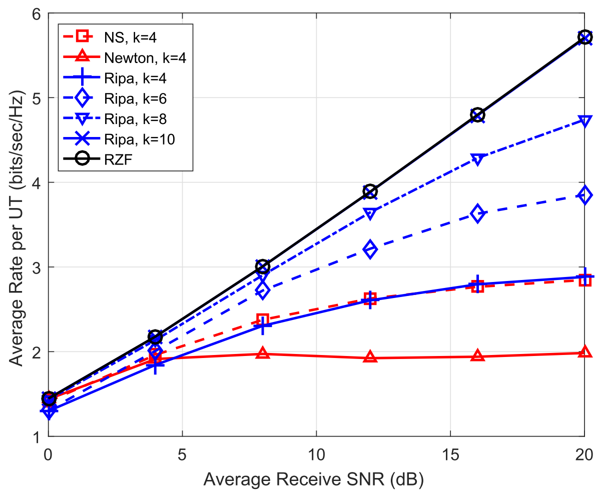


Fig. 1. Average achievable rate per UT versus average receive SNR for a uncoded massive MIMO system with $N = 128$ and $K = 32$.

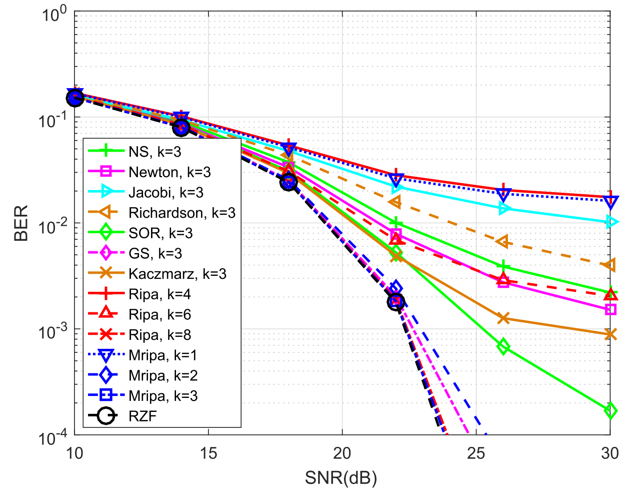


Fig. 3. Bit error rate versus average SNR for a uncoded massive MIMO system with $N = 128$ and $K = 16$ using 64-QAM.

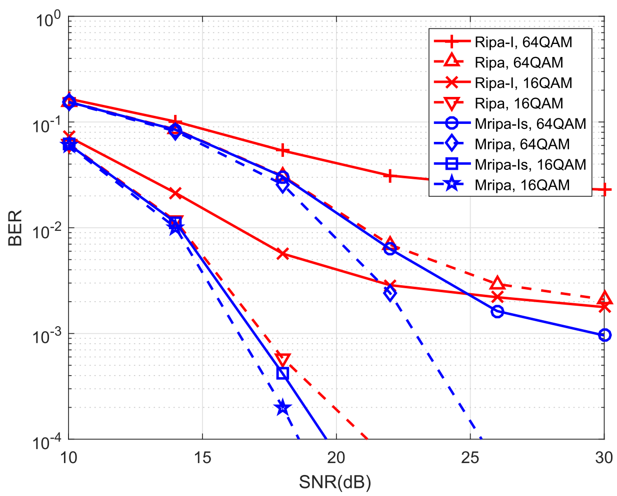


Fig. 2. Bit error rate versus average SNR for a uncoded massive MIMO system with $N = 128$ and $K = 16$.

where $\mathbf{g}_i$ is the i-th column of the precoding matrix $\mathbf{G}$. For a better illustration, RZF precoding scheme is employed as the performance benchmark while the Neumann series (NS) in [21] and Newton iteration (NI) in [22] are also applied as the comparison baselines. Clearly, under $k = 4$ iterations, a comparable performance can be achieved by RIPA with $q = 8$ compared to Neumann series. On the contrary, the Newton iteration does not work well in this case as its convergence requirement of $N \gg K$ is not satisfied. With the increment of k, we can observe that the performance of RIPA improves gradually, which is accordance with the convergence result in Theorem 1. To further study the proposed RIPA and MRIPA schemes with more details, the bit error rate (BER) performance is employed in the following simulations.

In Fig. 2, the initial setups of RIPA and MRIPA are examined by BER performance in a uncoded massive MIMO system with $N = 128$ and $K = 16$, where 16-QAM and 64-QAM are applied respectively for a better understanding. Specifically, two initial choices of $\mathbf{X}^{(0)} = \mathbf{I}$ and $\mathbf{X}^{(0)} = \mathbf{D}^{-1}$ are evaluated for RIPA while two choices of $\mathbf{t}^{(0)} = \mathbf{Is}$ and $\mathbf{t}^{(0)} = \mathbf{D}^{-1}\mathbf{s}$ are estimated for MRIPA. Meanwhile, with $q = 4$, we set $k = 6$ for RIDA and $k = 2$ for MRIPA respectively, where MRIPA applies a full iteration. Typically, from Theorem 1 and Corollary 3, the convergence performance of RIPA and MRIPA are also determined by the choice of $\mathbf{X}^{(0)}$ and $\mathbf{t}^{(0)}$. Therefore, by offering better approximations, the choices $\mathbf{X}^{(0)} = \mathbf{D}^{-1}$ and $\mathbf{t}^{(0)} = \mathbf{D}^{-1}\mathbf{s}$ for RIPA and MRIPDA achieve better performance than those with $\mathbf{X}^{(0)} = \mathbf{I}$ and $\mathbf{t}^{(0)} = \mathbf{Is}$. Given the low complexity of computing $\mathbf{D}^{-1}$, such choices of $\mathbf{X}^{(0)} = \mathbf{D}^{-1}$ and $\mathbf{t}^{(0)} = \mathbf{D}^{-1}\mathbf{s}$ are highly recommended in practice, and we also apply them in the following simulations by default. Meanwhile, as expected, MRIPA achieves a better BER performance than RIPA, which is due to its faster convergence shown in Corollary 2.

In Fig. 3, the BER performance comparison between RIPA, MRIPA and other conventional iteration precoding schemes are presented with respect to a 128×16 uncoded massive MIMO system with 64-QAM. Besides the Neumann series and Newton iteration, the Jacobi iteration, Richardson iteration, SOR iteration, Gauss Seidel method and random Kaczmarz iterations [40] are also employed for a better comparison, where RZF serves a performance benchmark. Note that the condition $N \gg K$ is satisfied here, so that these traditional iterative methods like Neumann series, Newton iteration and so on work as usual. In order to show the convergence behaviour in a better way, the iteration numbers of RIPA and MRIPA are set as $k = 4, 6, 8$ and $k = 1, 2, 3$ respectively with $q = 4$. As can be seen clearly, with the increment of iterations, both the precoding performance of RIPA and MRIPA improve gradually, thus confirming the convergence of the proposed randomized iterations. More specifically, with $k = 8$ and $k = 3$, near RZF performance will be obtained by RIPA and MRIPA respectively while MRIPA achieves a better performance than RIPA. This is line with the afore-mentioned analysis about the convergence rate and MRIPA does have a better convergence by optimization and enhancement.

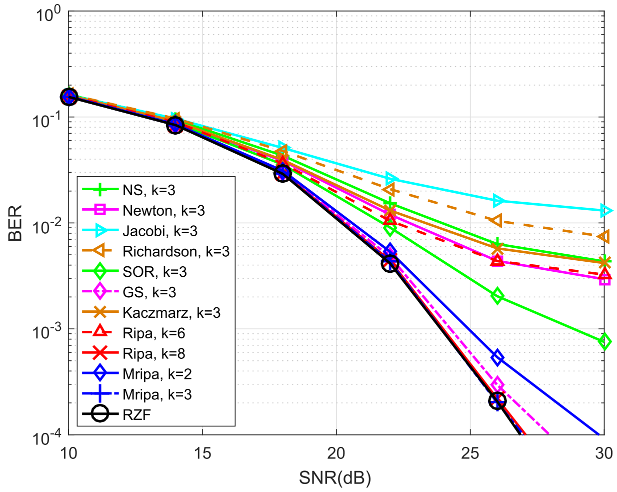


Fig. 4. Bit error rate versus average SNR for a uncoded massive MIMO system with $N = 128$ and $K = 16$ using 64-QAM under imperfect CSI.

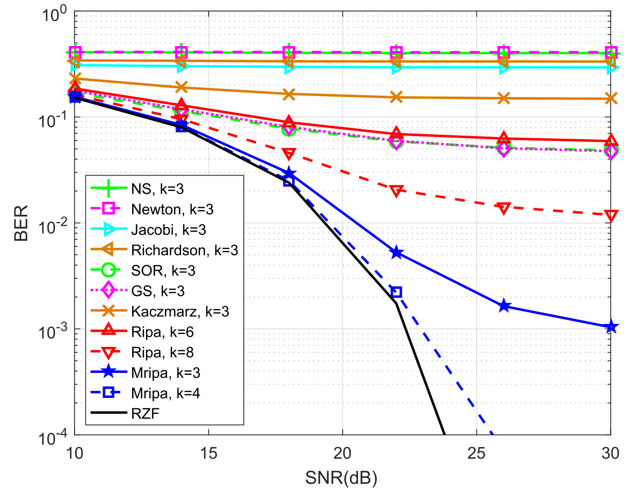


Fig. 5. Bit error rate versus average SNR for the uncoded 128×16 massive MIMO using 64-QAM with normalized correlation index $\psi = 0.05$.

As a counterpart of Fig. 3, Fig. 4 is presented to evaluate the BER performance of the proposed RIPA and MRIPA without perfect channel state information (CSI) in a 128×16 uncoded massive MIMO systems using 64-QAM. Specifically, let $\widehat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$ stand for the channel matrix with imperfect CSI. Here, $\Delta\mathbf{H}$ denotes the channel estimation errors and each of its elements follows $\mathcal{CN}(0, \sigma_e^2)$ with $\sigma_e^2 = 0.1$ [49], [50]. Due to the imperfect CSI, the BER performance of all the precoding schemes deteriorate in general compared to the case with perfect CSI in Fig. 3. Nevertheless, the performance gain of MRIPA still can be verified while near RZF performance can be achieved by it with the increment of k.

Besides the independent, identically distributed (i.i.d.) channels, the impact of correlated channels is also studied to reveal the convergence performance of the proposed RIPA and MRIPA schemes with $q = 4$. Specifically, following the setups of correlation channels in [51], [52], the correlated channel matrix is set by $\mathbf{R}_{\text{b}}^{\frac{1}{2}}\mathbf{H}\mathbf{R}_{\text{u}}^{\frac{1}{2}}$, where $\mathbf{R}_{\text{b}} \in \mathbb{C}^{N\times N}$ and $\mathbf{R}_{\text{u}} \in \mathbb{C}^{K\times K}$ denote the correlation matrices at base station side and user side respectively. Here, the normalized correlation coefficient $1 \geq \psi \geq 0$ controls the correlation degree of the channels, where $\psi = 0$ means an uncorrelated scenario and $\psi = 1$ implies a fully correlated one. Compared to the i.i.d. case in Fig. 3, the precoding performance of RZF slightly degrades with normalized correlation index $\psi = 0.05$ in Fig. 5. However, the performance of traditional iteration methods like Neumann series, Newton, Jacobi, Richardson are terrible because their convergence severely suffer from the correlated channels. More precisely, this is because a more correlated channel naturally leads to a large *condition number*, which has a negative effect upon their convergence. On the contrary, both the proposed RIPA and MRIPA work as usual but with slower convergence rates. Meanwhile, their precoding performance improve gradually with the increase of k, which is accordance with the derived convergence results. Note that under the same iteration number MRIPA achieves a better performance than GS, SOR and random Kaczmarz. In Fig. 6, the similar observations can also be found, where the channel matrix becomes more correlated with the normalized correlation index $\psi = 0.1$. Due to the illness of the channel matrix, the performance of RZF also gets worse than before. In this case, traditional iterations like Neumann series, Newton, Jacobi, Richardson do not work any more. Different from them, the convergence of RIPA and MRIPA are ensured even though more number of iterations are required to achieve the near RZF performance. Nevertheless, considerable performance gain still can be verified by MRIPA compared to GS, SOR and so on under the same iterations.

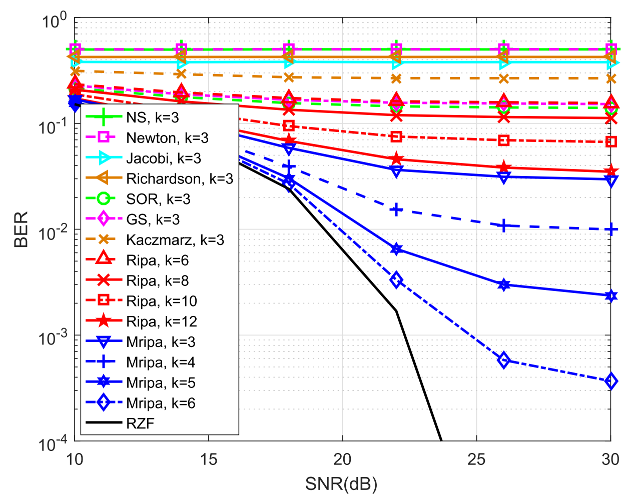


Fig. 6. Bit error rate versus average SNR for the uncoded 128×16 massive MIMO using 64-QAM with normalized correlation index $\psi = 0.1$.

In Fig. 7, the BER performance comparison with respect to RIPA and MRIPA is presented in a uncoded massive MIMO system with $N = 128$ and $K = 32$ using 16-QAM. For a better comparison, the precoding schemes like RZF, Neumann series, Newton iteration, Jacobi iteration, Richardson iteration and so on are added as well. Intuitively, as the convergence requirement $N \gg K$ is not fulfilled in this case,

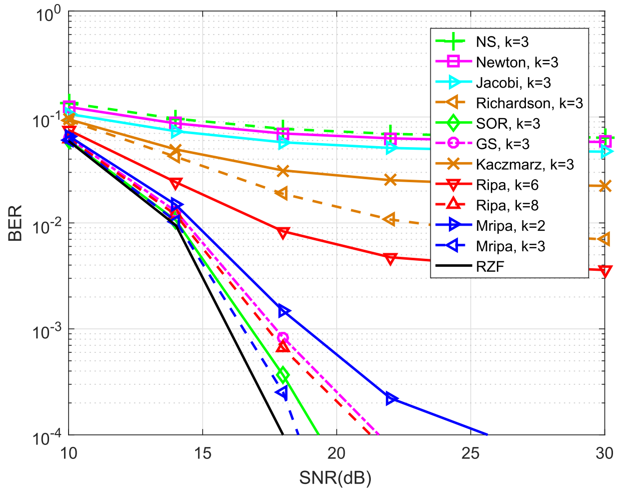


Fig. 7. Bit error rate versus average SNR for a uncoded massive MIMO system with $N = 128$ and $K = 32$ using 16-QAM.

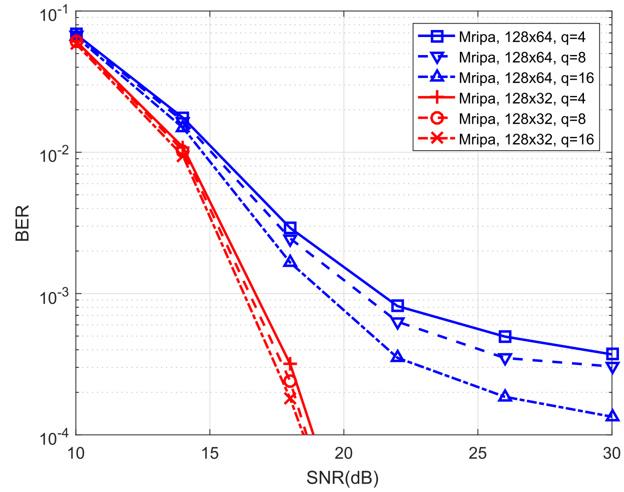


Fig. 9. Bit error rate versus average SNR for the uncoded massive MIMO systems using 16-QAM.

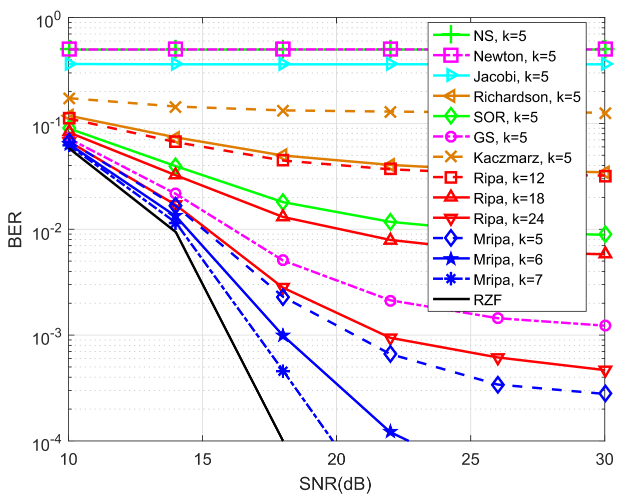


Fig. 8. Bit error rate versus average SNR for a uncoded massive MIMO system with $N = 128$ and $K = 64$ using 16-QAM.

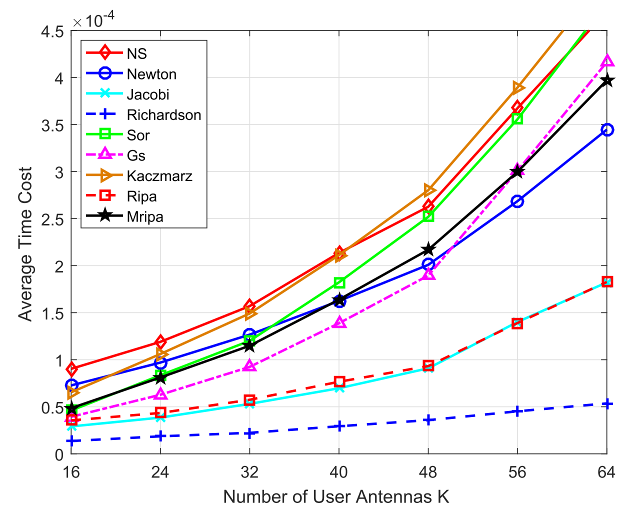


Fig. 10. Complexity comparison in average time cost for the uncoded $128 \times K$ massive MIMO system using 16-QAM at SNR $= 18$dB.

the convergence performance of Neumann series, Newton iteration and Jacobi iteration are poor, which result in the terrible BER performance in the downlink massive MIMO systems. In contrast, the convergence of both RIPA and MRIPA with $q = 8$ work well as they converge globally according to Theorem 1. Meanwhile, under the same iterations, the performance of MRIPA outperforms GS, SOR and random Kaczmarz. With the increase of the iteration number, the precoding performance of RIPA and MRIPA improve gradually while the RZF precoding performance can be achieved subsequently.

In Fig. 8, we extend the precoding performance comparison to a uncoded massive MIMO system with $N = 128$ and $K = 64$ using 16-QAM. Compared to Fig. 7, the antenna ratio N/K between the base station side and the user side gets smaller. Clearly, we can observe that the conventional iteration schemes like Neumann series, Newton iteration and Jacobi iterations do not converge any more so that the precoding schemes based on them do not work at all. In sharp contrast with them, the proposed RIPA and MRIPA with $q = 8$ work as usual, and their BER performance gradually improve with the increase of the number of iterations. Similarly, under the same iterations, MRIPA achieves a better BER performance than GS, SOR and random Kaczmarz.

Fig. 9 is shown to evaluate the different choices of q for MRIPA in both 128×64 and 128×32 uncoded massive MIMO systems with 16-QAM. Specifically, the choices of $q = 4, 8, 16$ are applied for a better understanding. Meanwhile, the numbers of iterations are set as $k = 3$ for case of 128×32 and $k = 5$ for case of 128×64 respectively. Intuitively, the precoding performance of MRIPA improves gradually with the increase of size q since more components of $\mathbf{x}$ can be updated together. However, according to (71), a larger size q also requires more computational cost so that a reasonable size $1 \leq q \leq \sqrt{K}$ should be carefully selected, which will be one of our work in future.

In Fig. 10, the complexity comparison in average elapsed running times per iteration is given to show the computational costs of the proposed RIPA and MRIPA. The number of antennas at the base station side is set as $N = 128$ while 16-QAM is applied with SNR = 18dB. Meanwhile, the simulation is conducted by MATLAB R2019a on a single computer with an Intel Core i7 processor at 2.8GHz, a RAM of 8GB and Windows 10 Enterprise Service Pack operating system. Clearly, the average elapsed running times per iteration of all the precoding schemes grow accordingly with the increment of antennas at the user side (i.e., K). We point out that the computational complexity of MRIPA in (71) is significantly less than that of RIPA in (28) due to the usage of iteration transformation and optimization. However, a full iteration that contains K/q single iterations of (71) is applied to MRIPA for a fair comparison with other traditional iteration schemes. Nevertheless, MRIPA is still competitive compared to GS and SOR iteration schemes. On the other hand, it is clear to see that Newton iteration, Jacobi iteration and Richardson iteration have smaller average running times than MRIPA. However, MIRPA not only archives a better convergence performance than them but also avoids suffering from any specific convergence requirement.

VII. Conclusion

In this paper, the downlink precoding in massive MIMO systems is studied, and two statistical linear precoding schemes based on random iterative method are proposed. First of all, by introducing random sampling into iteration schemes, the randomized iterative precoding algorithm (RIPA) is proposed with low computational complexity. Then, we demonstrate that RIPA achieves an exponential convergence, and its convergence rate is also derived. Meanwhile, we show that RIPA enjoys a global convergence without suffering from the convergence requirement like the antenna ratio on both transmitter and receiver sides. This significantly extends the applications of low-complexity precoding schemes in downlink massive MIMO systems. After that, based on iteration transformation and conditional sampling, further optimization and enhancement are given, where the modified randomized iterative precoding algorithm (MRIPA) is proposed for better convergence and efficiency. Therefore, by simply tuning the number of iterations, flexible precoding trade-off between performance and complexity can be achieved by MRIPA in various scenarios of the downlink massive MIMO systems.

References

[1] 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, Jan. 2020.

[2] 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.

[3] 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, Oct. 2019.

[4] 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.

[5] Y. Wu, C.-K. Wen, C. Xiao, X. Gao, and R. Schober, "Linear precoding for the MIMO multiple access channel with finite alphabet inputs and statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 983–997, Feb. 2015.

[6] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and J. Wang, "Joint MU-MIMO precoding and resource allocation for mobile-edge computing," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1639–1654, Mar. 2021.

[7] 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.

[8] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[9] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Syst. J.*, vol. 12, no. 1, pp. 3920–3931, Dec. 2017.

[10] T. Xie, L. Dai, X. Gao, X. Dai, and Y. Zhao, "Low-complexity SSOR-based precoding for massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 744–747, Apr. 2016.

[11] Y. Liu, J. Liu, Q. Wu, Y. Zhang, and M. Jin, "A near-optimal iterative linear precoding with low complexity for massive MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 6, pp. 1105–1108, Jun. 2019.

[12] A. Kammoun, A. Müller, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 861–875, Oct. 2014.

[13] A. Benzin, G. Caire, Y. Shadmi, and A. M. Tulino, "Low-complexity truncated polynomial expansion DL precoders and UL receivers for massive MIMO in correlated channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1069–1084, Feb. 2019.

[14] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Performance analysis for massive MIMO downlink with low complexity approximate zero-forcing precoding," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3848–3864, Sep. 2018.

[15] J.-C. Chen, C.-J. Wang, K.-K. Wong, and C.-K. Wen, "Low-complexity precoding design for massive multiuser MIMO systems using approximate message passing," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5707–5714, Jul. 2016.

[16] *User Equipment (UE) Conformance Specification*, Standard TS 38.521-4 v.15.0.0, 3GPP, Tech. Spec., May 2019, pp. 1–5.

[17] *Study on Scenarios and Requirements for Next Generation Access Technologies*, Standard TR 38.913 v.14.3.0, Tech. Rep., Jun. 2017, pp. 1–41.

[18] B. Nagy, M. Elsabrouty, and S. Elramly, "Fast converging weighted Neumann series precoding for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 154–157, Apr. 2018.

[19] Y. Man, Z. Li, F. Yan, S. Xing, and L. Shen, "Massive MIMO pre-coding algorithm based on truncated Kapteyn series expansion," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Dec. 2016, pp. 1–5.

[20] S. Zarei, W. Gerstacker, and R. Schober, "Low-complexity widely-linear precoding for downlink large-scale MU-MISO systems," *IEEE Commun. Lett.*, vol. 19, no. 4, pp. 665–668, Apr. 2015.

[21] 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)*, Jun. 2015, pp. 1763–1769.

[22] H. V. Nguyen, V.-D. Nguyen, and O.-S. Shin, "Low-complexity precoding for sum rate maximization in downlink massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 186–189, Apr. 2017.

[23] Y. Man, C. Zhang, Z. Li, F. Yan, S. Xing, and L. Shen, "Massive MIMO pre-coding algorithm based on improved Newton iteration," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.

[24] C. Tang *et al.*, "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.

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

[26] D. Kwon, W. Y. Yeo, and D. K. Kim, "A new precoding scheme for constructive superposition of interfering signals in multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 2047–2050, Nov. 2014.

[27] J. Minango and C. de Almeida, "A low-complexity linear precoding algorithm based on Jacobi method for massive MIMO systems," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–5.

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

[29] Z. Lu, J. Ning, Y. Zhang, T. Xie, and W. Shen, "Richardson method based linear precoding with low complexity for massive MIMO systems," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–4.

[30] Y. Saad, *Methods for Sparse Linear Systems*. Philadelphia, PA, USA: SIAM, 1997.

[31] T. Xie, Q. Han, H. Xu, Z. Qi, and W. Shen, "A low-complexity linear precoding scheme based on SOR method for massive MIMO systems," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.

[32] X. Gao, L. Dai, J. Zhang, S. Han, and I. Chih-Lin, "Capacity-approaching linear precoding with low-complexity for large-scale MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1577–1582.

[33] P. Yang and H. Yang, "A low-complexity linear precoding for MIMO channels with finite constellation inputs," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1415–1418, Oct. 2019.

[34] W. Zhang *et al.*, "Widely linear precoding for large-scale MIMO with IQI: Algorithms and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3298–3312, May 2017.

[35] S. Zarei, W. H. Gerstacker, R. Weigel, M. Vossiek, and R. Schober, "Robust MSE-balancing hierarchical linear/Tomlinson-Harashima precoding for downlink massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7309–7324, Nov. 2018.

[36] S. Liu, C. Ling, and D. Stehle, "Decoding by sampling: A randomized lattice algorithm for bounded distance decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5933–5945, Sep. 2011.

[37] 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.

[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] V. Croisfelt, A. Amiri, T. Abrao, E. de Carvalho, and P. Popovski, "Accelerated randomized methods for receiver design in extra-large scale MIMO arrays," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6788–6799, Jul. 2021.

[40] M. N. Boroujerdi, S. Haghighatshoar, and G. Caire, "Low-complexity statistically robust precoder/detector computation for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6516–6530, Oct. 2018.

[41] H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear pre-coding systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 2710–2715.

[42] A. Björck, *Numerical Methods in Matrix Computations*. Cham, Switzerland: Springer, 2015.

[43] R. Mansel Gower and P. Richtarik, "Stochastic dual ascent for solving linear systems," 2015, *arXiv:1512.06890*.

[44] D. Leventhal and A. S. Lewis, "Randomized methods for linear constraints: Convergence rates and conditioning," *Math. Oper. Res.*, vol. 35, no. 3, pp. 641–654, Aug. 2010.

[45] R. M. Gower and P. Richtárik, "Randomized quasi-Newton updates are linearly convergent matrix inversion algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 38, no. 4, pp. 1380–1409, Jan. 2017.

[46] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices* (Cambridge Studies in Advanced Mathematics). Cambridge, U.K.: Cambridge Univ. Press, 2009.

[47] R. M. Gower, "Sketch and project: Randomized iterative methods for linear systems and inverting matrices," Ph.D. thesis, School Math., Univ. Edinburgh, Edinburgh, U.K., 2016.

[48] A. Müller, A. Kammoun, E. Bjornson, and M. Debbah, "Efficient linear precoding for massive MIMO systems using truncated polynomial expansion," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2014, pp. 273–276.

[49] N. Lee, O. Simeone, and J. Kang, "The effect of imperfect channel knowledge on a MIMO system with interference," *IEEE Trans. Commun.*, vol. 60, no. 8, pp. 2221–2229, Aug. 2012.

[50] D. L. Colon, F. H. Gregorio, and J. Cousseau, "Linear precoding in multi-user massive MIMO systems with imperfect channel state information," in *Proc. 16th Workshop Inf. Process. Control (RPIC)*, Oct. 2015, pp. 1–6.

[51] H. R. Bahrami, T. Le-Ngoc, A. M. Nasri Nasrabadi, and S. H. Jamali, "Precoder design based on correlation matrices for MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dec. 2005, pp. 2001–2005.

[52] B. Costa, A. Mussi, and T. Abrao, "MIMO detectors under correlated channels," *Semina, Ciencias Exatas e Tecnologicas*, vol. 37, no. 1, pp. 3–12, 2016.

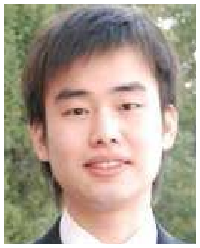
Zheng Wang (Member, IEEE) received the B.S. degree in electronic and information engineering from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2009, the M.S. degree in communications from The University of Manchester, Manchester, U.K., in 2010, and the Ph.D. degree in communication engineering from Imperial College London, U.K., in 2015.

From 2015 to 2016, he has worked as a Research Associate at Imperial College London. From 2016 to 2017, he was a Senior Engineer with the Radio Access Network Research and Development Division, Huawei Technologies Company. From 2017 to 2020, he was an Associate Professor at the College of Electronic and Information Engineering, NUAA. Since 2021, he has been an Associate Professor with the School of Information and Engineering, Southeast University, Nanjing. His current research interests include massive MIMO systems, machine learning and data analytics over wireless networks, and lattice theory for wireless communications.

Robert M. Gower received the bachelor's and master's degrees in applied mathematics from the State University of Campinas, Brazil, and the Ph.D. degree in applied mathematics from The University of Edinburgh. In his thesis, he introduced the new sketch-and-project methods for solving linear systems. He is currently a Research Scientist at the Flatiron Institute. Before that, he was a Visiting Scientist and Google Brain (2021) and Facebook AI Research (2020) New York. He has been an Assistant Professor at Télécom Paris, Institut Polytechnique, since 2017. He has designed the current state-of-the-art algorithms for automatically calculating high order derivatives using back-propagation at the State University of Campinas.

Cheng Zhang (Member, IEEE) received the B.Eng. degree from Sichuan University, Chengdu, China, in June 2009, the M.Sc. degree from the Xi'an Electronic Engineering Research Institute (EERI), Xi'an, China, in May 2012, and the Ph.D. degree from Southeast University (SEU), Nanjing, China, in December 2018. From June 2012 to August 2013, he was a Radar Signal Processing Engineer with Xian EERI. From November 2016 to November 2017, he was a Visiting Student with the University of Alberta, Edmonton, AB, Canada. He is currently an Associate Professor with SEU and is supported by the Zhishan Young Scholar Program of SEU. His research interests include MIMO wireless communications and signal processing. He was a recipient of the Excellent Doctoral Dissertation of the China Education Society of Electronics in December 2019.

Shanxiang Lyu received the B.S. and M.S. degrees in electronic and information engineering from the South China University of Technology, Guangzhou, China, in 2011 and 2014, respectively, and the Ph.D. degree from the Electrical and Electronic Engineering Department, Imperial College London, U.K., in 2018. He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou. His research interests are in lattice codes, wireless communications, and cryptography. He was a recipient of the 2021 CIE Information Theory Society Yong-Star Award and the 2020 Superstar Supervisor Award of the National Crypto-Math Challenge of China. He is the Organizing Chair of Inscrypt 2020.



Yili Xia (Member, IEEE) received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2006, the M.Sc. degree (Hons.) in communications and signal processing from the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2007, and the Ph.D. degree in adaptive signal processing from Imperial College London in 2011.

Since 2013, he has been an Associate Professor in signal processing with the School of Information Science and Engineering, Southeast University, where he is currently the Deputy Head of the Department of Information and Signal Processing Engineering. His research interests include complex and hyper-complex statistical analysis, detection and estimation, linear and nonlinear adaptive filters, as well as their applications on communications and power systems. He was a recipient of the Best Student Paper Award at the International Symposium on Neural Networks (ISNN) in 2010 (coauthor) and the Education Innovation Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2019. He is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Yongming Huang (Senior Member, 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 Southeast University, Nanjing, in 2007.

Since March 2007, he has been a Faculty with the School of Information Science and Engineering, Southeast University, where he is currently a Full Professor. During 2008–2009, he has visited the Signal Processing Laboratory, 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 has 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 the IEEE TRANSACTIONS ON SIGNAL PROCESSING and a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is an Editor-at-Large of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY and an Associate Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS.