Stair Matrix and Its Applications to Massive MIMO Uplink Data Detection

Fan Jiang, *Student Member, IEEE*, Cheng Li, *Senior Member, IEEE*, Zijun Gong, *Student Member, IEEE*, and Ruoyu Su, *Member, IEEE*

***Abstract*— In this paper, we investigate low-complexity data detection scheme for massive multiple-input multiple-output (MIMO) uplink transmission. We propose to utilize the stair matrix, instead of diagonal matrix in existing proposals, for the development, and achieve near linear minimum mean-square error detection performance. We first demonstrate the applicability of the proposed method by showing that the probability (that the convergence conditions are met) approaches one as long as sufficiently large number of antennas are equipped at the base station. We then propose an iterative method to perform data detection and show that much improved performance can be achieved with the computational complexity remaining at the same level of existing iterative methods, where the diagonal matrix is adopted. Furthermore, we conduct numerical simulations, and the results validate the significant performance enhancement of using the stair matrix over the diagonal matrix in all performance aspects. Moreover, we apply the proposed scheme to a massive MIMO system, where the extended vehicular A channel data are generated. The performance improvement of the proposed scheme over existing proposals is also validated.**

***Index Terms*— Massive MIMO, stair matrix, iterative method, convergence condition.**

I. Introduction

THE development and successful applications of multiple-input multiple-output (MIMO) systems in modern wireless communications have brought the bright prospective of massive MIMO techniques in future 5G mobile communication systems [1]–[3]. It is foreseeable that massive MIMO, together with the millimeter wave frequency band [4], has been a promising candidate to meet the high rate, low latency 5G system requirements. Due to the huge potential multiplexing and diversity gain over the small-scale MIMO and single-antenna systems, massive MIMO can boom the system spectrum and energy efficiency [1], [5]–[7]. Along with the benefits of massive MIMO, however, the cost of high computational complexity required in signal processing (data detection, precoding, etc.) increases, which prohibits the application of the optimal detection methods, such as the maximum likelihood (ML), and maximum *a posteriori* (MAP) detection, in realization [8]–[26].

To achieve good tradeoff between the system performance and the computational complexity, much effort has been made in developing low-complexity detection schemes to achieve near-optimal performance [10]–[26]. For example, in [10] and [11], the authors extend a class of likelihood ascent search (LAS) based multi-user detection schemes to large-scale MIMO. The Monte-Carlo sampling based massive MIMO detection schemes are studied in [12] and [13], where the Markov chain Monte-Carlo (MCMC) strategy is introduced. A class of belief propagation (BP) based detection schemes for massive MIMO have been studied in [14] and [15], where the idea of message passing through factor graph is introduced for the development. Recently, adaptive reduced rank interference suppression based detection schemes are developed for large-scale MIMO uplink detection [16]–[18]. The idea behind these schemes is to project a large dimensional signal vector to a small dimensional one, and then multi-stage Winner filter (MSWF) is adopted to perform optimization of the desired signal [27]–[30]. It has been also reported that the complexity required for each symbol detection can be reduced to $\mathcal{O}(N_B \times N_U)$. ($N_B$ and N_U denotes the number of antennas at base station and the single-antenna user terminals, respectively.) Among these low complexity detection schemes for massive MIMO uplink, a class of near linear MMSE/ZF detection schemes are studied in [19]–[26]. Generally, all those schemes attempt to avoid matrix inversion operations in linear MMSE/ZF detection schemes, and the strategies can be summarized in two categories: the first one is to approach matrix inversion, and the other is to solve linear equations with iterative methods.

The first category is to approach the matrix inversion [19]–[21]. For example, in [19], the authors attempt to introduce Neumann series expansion to avoid the matrix inversion in linear MMSE detection. It has been shown that when the number of antennas at base station is much greater than the number of user equipment, the orders required for Neumann series expansion can be as few as 3 (for example, $r = N_B / N_U \geq 16$). In [20], the probability of the convergence condition that using the diagonal matrix in Neumann series expansion has been comprehensively discussed. However, Neumann series expansion suffers from matrix multiplications, and the computational complexity is comparable to the matrix inversion algorithm when the expansion order is more than

Manuscript received June 27, 2017; revised October 28, 2017; accepted December 22, 2017. Date of publication January 3, 2018; date of current version June 14, 2018. This work is supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada (Discovery Grant 293264-12, Engage Grant 486202-15), and InnovateNL SensorTECH Program (Grant number: 5404-2061-101). The associate editor coordinating the review of this paper and approving it for publication was R. C. de Lamare. *(Corresponding author: Cheng Li.)*

The authors are with the Department of Electrical and Computer Engineering, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada (e-mail: licheng@mun.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2017.2789211

two. In order to speed up the convergence rate, diagonal banded Newton iteration based matrix inversion approach is studied in [21], where the Newton iteration structure is used. Actually, the results after P iterations in Newton iteration can be seen as the Neumann series expansion of the order $2^P - 1$ [21]. Inevitably, matrix multiplications are involved in diagonal banded Newton iteration based matrix inversion approach, and the iterations are limited to 2 for computational complexity consideration. In summary, the methods that are to approach matrix inversion suffer from high computational complexity due to the matrix multiplications and the slow convergence rate when the ratio r is not sufficiently large.

The second category is to solve linear equations with iterative methods [22]–[26]. The basic idea of these methods is to transform the matrix inversion problem into solving linear equations. To solve the linear equations, an initial estimation is provided. Then following an iterative structure to converge, the final output is provided as the solutions to linear equations. For example, in [22], the Jacobi method is adopted, and by following the Jacobi iterative structure, the estimation eventually approaches the MMSE estimation. The Richardson iteration in massive MIMO uplink data detection has been studied in [23], and the authors have demonstrated that the iterative structure can converge even with zero initialization. However, as pointed out in [24] and [25], the convergence rate for both Jacobi method and Richardson iteration is slow, hence quite a few iterations are required for convergence. The application of Gauss-Seidel method to massive MIMO uplink data detection is studied in [24], and the convergence performance can be greatly improved. By providing an initial estimation that is close to the MMSE estimation, the joint steepest-decent and Jacobi method based data detection is proposed in [25], and the iterations are greatly reduced. In [26], the authors formulate the MMSE estimation as a minimization problem, and use the conjugate gradient to calibrate the next estimation. However, conjugate gradient-based data detection scheme involves many division operations, which is also computational costly. Compared to the first category which is to approach matrix inversion, solving linear equations with iterative methods is of less complexity due to the replacement of matrix multiplications with matrix-vector products. However, as summarized in [25], the overall computational complexity of the iterative methods, including the computations in both the initialization and iteration, is still high. It is worth pointing out that the convergence rate of the existing iterative methods can be speeded up by using preconditioning [31]. The potential direction to further reduce the computational complexity can be finding an iterative method that requires less computation in initialization and less iterations for convergence [32].

In both the previous mentioned two categories of near linear MMSE/ZF data detection schemes, we note that most proposals in existing literatures mainly utilize the diagonal matrix in the development. In [19] and [20], the applicability of using diagonal matrix to massive MIMO uplink data detection has been demonstrated. However, as we will show later, we find some limitations for using diagonal matrix. First of all, in the massive MIMO system configuration with small ratio of $r = N_B / N_U$, the convergence rate of using the diagonal matrix is slow. Alternatively, a few iterations (or orders in Neumann series expansion) are required to provide near-optimal system performance. Besides, the convergence conditions, which are critical for the both data detection schemes mentioned above, are met with a low probability when r is small. In other word, in some cases, the diagonal matrix may not be used to converge.

The motivation of this paper originates from achieving a better tradeoff between computational complexity and system performance in massive MIMO uplink data detection. Specifically, we propose to develop an iterative method to achieve near linear MMSE detection scheme use the stair matrix. Stair matrix is a special tridiagonal matrix where the off-diagonal elements in either the even or the odd row are zeros [33]. As far as we know, the applications of stair matrix in massive MIMO systems have not been studied. In the development of the iterative method using stair matrix, we need to address two fundamental issues: the probability that the convergence conditions are satisfied, and the convergence rate. The previous one tells whether the stair matrix is applicable or not (applicability) in massive MIMO, while the latter one reveals the advantages over the use of diagonal matrix. We address these issues in massive MIMO, and the contributions of this paper are summarized as follows:

- We show that when N_B grows to infinite, the probability that the convergence conditions are met approaches 1. As the antennas at base station in Massive MIMO systems can be hundreds, this conclusion demonstrates the applicability of the stair matrix in massive MIMO systems. In finite N_B region (or low r region), we show that by using the stair matrix, the probability that the convergence conditions are met can be greatly improved, and the cumulative distribution function of the maximum eigenvalue of the convergence matrix indicates that the convergence rate can be speeded up by using the stair matrix;
- We propose an iterative method with the use of the stair matrix, and demonstrate the proposed scheme has the same level of the computational complexity compared to the existing iterative methods where the diagonal matrix is applied;
- We apply the stair matrix in Neumann series expansion to approach matrix inverse, and demonstrate that the mean-square error of the truncated expansion can be greatly reduced compared to the use of diagonal matrix;
- The proposed iterative method is to approach linear MMSE estimation vector, and we show that the residual estimation error, between the iterative estimation and the linear MMSE estimation, is much less than that of the Jacobi method where the diagonal matrix is applied;
- We perform numerical simulations to evaluate the system BER performance, and show that the proposed iterative method achieves significant performance improvement over the proposals where diagonal matrix is adopted in the development;
- We apply the proposed scheme to a massive MIMO system with extended vehicular A (EVA) channel, and show significant BER performance enhancement over the reported near linear MMSE detection schemes.

The rest of this paper is organized as follows. Section II provides the system model, including the massive MIMO structure and the preliminary work of linear ZF/MMSE detection. In section III, the introduction to stair matrix and its applicability in massive MIMO will be presented. The implementation of stair matrix in massive MIMO data detection with iterative method is presented in section IV. In section V, we conduct the numerical simulations and present the results and discussion. Finally, the conclusions and future work are presented in section VI.

Notations: Throughout the paper, the lowercase and uppercase bold symbols denote the column vector and matrix, respectively. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, and $(\cdot)^{-1}$ are reserved for matrix transpose, conjugate transpose, and inverse, respectively. $\mathbb{C}$ and $\mathbb{N}$ are reserved for the sets of the complex and natural numbers, respectively. $\|\mathbf{A}\|_{\mathrm{F}}$ and $\|\mathbf{a}\|_2$ are the Frobenius-norm of a matrix $\mathbf{A}$ and the ℓ_2-norm of a vector $\mathbf{a}$. $\mathrm{E}\{\cdot\}$ and $\mathrm{cov}\{\cdot,\cdot\}$ denote the expectation, and covariance operation. $\exp(\cdot)$ and $\ln(\cdot)$ denote the exponential and natural logarithmic functions, respectively. $\mathbf{I}_L$ is reserved for the size L identity matrix and $\mathbf{e}_l$ represents the l-th column of $\mathbf{I}_L$; $\mathrm{diag}\{\mathbf{a}\}$ converts a column vector $\mathbf{a}$ to a diagonal matrix and $\mathrm{diag}\{\mathbf{A}\}$ obtains the diagonal elements in a matrix $\mathbf{A}$ to form a column vector. $\rho(\mathbf{A})$ is the spectral radius of the matrix $\mathbf{A}$.

II. System Model

We consider the massive MIMO uplink with N_B antennas at base station to simultaneously serve N_U single-antenna user equipment. The N_U bitstream from each user is first encoded, then interleaved, and fed into digital modulator. The modulated symbols are transmitted into massive MIMO channel, and the received signal vector at base station can be expressed as

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

where $\mathbf{y} = \left[y_1, y_2, \cdots, y_{N_B}\right]^{\mathrm{T}}$ is a complex-valued $N_B \times 1$ vector, with y_m denoting the received signal from the m-th receiving antenna. $\mathbf{x} = \left[x_1, x_2, \cdots, x_{N_U}\right]^{\mathrm{T}}$ with the transmitted symbol of user u denoted by x_u. The transmitted symbols are unit power normalized and independent from each other, i.e., $\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\} = \mathbf{I}_{N_U}$. $\mathbf{H} = \left[\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_{N_U}\right]$ denotes the channel matrix with $\mathbf{h}_u \in \mathbb{C}^{N_B \times 1}$ where each entry is independent and identically distributed (i.i.d.), modeled as the flat Rayleigh fading channel [1], [5], [24], [34]. $\mathbf{z} = \left[z_1, z_2, \cdots, z_{N_B}\right]^{\mathrm{T}}$ is the noise vector, satisfying $\mathrm{E}\left\{\mathbf{z}\mathbf{z}^{\mathrm{H}}\right\} = \sigma_z^2\mathbf{I}_{N_B}$ with each entry modeled as zero-mean complex Gaussian circularly symmetric (ZMCGCS) random variable. It is worth noting that in frequency selective fading channels, by applying the orthogonal frequency division multiplexing and single-carrier frequency division multiple access (OFDM/SC-FDMA) techniques, the signal model expressed in Equation (1) is established over each subcarrier.

A. Linear MMSE Data Detection

The multi-user data detector at the base station is to compute the *a posteriori* log likelihood ratio (LLR) of the bits associated with the modulated symbols. After the knowledge of the channel matrix (note that the channel matrix is obtained through channel estimator, where time domain and/or frequency domain training pilots are used for the channel estimation [35], [36]), the well-known linear MMSE data detection can be given by

$$\hat{\mathbf{x}} = \left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \sigma_z^2\mathbf{I}_{N_U}\right)^{-1}\mathbf{H}^{\mathrm{H}}\mathbf{y} = \mathbf{W}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{2}$$

where $\mathbf{y}^{\mathrm{MF}} = \mathbf{H}^{\mathrm{H}}\mathbf{y}$ can be seen as the matched-filter output, and the MMSE equalization matrix $\mathbf{W}$ can be expressed as

$$\mathbf{W} = \mathbf{G} + \sigma_z^2\mathbf{I}_{N_U}, \tag{3}$$

where $\mathbf{G} = \mathbf{H}^{\mathrm{H}}\mathbf{H}$ is the Gram matrix. It is worth noting that in high signal-to-noise ratio (SNR) region, if the Gram matrix is invertible, [1] Equation (2) can be reduced to

$$\hat{\mathbf{x}} = \mathbf{G}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{4}$$

which is the linear ZF data detection scheme, where the noise component is not considered in the equalization process.

To obtain the *a posteriori* LLR of the bits associated with the modulated symbols, we write the estimation in Equation (2) as

$$\hat{x}_u = \mathbf{e}_u^{\mathrm{H}}\hat{\mathbf{x}} = \rho_u x_u + \xi_u, \tag{5}$$

where the equivalent channel gain ρ_u and the *a posteriori* noise-plus-interference (NPI) ξ_u can be given by

$$\rho_u = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{G}\mathbf{e}_u, \tag{6}$$

$$\xi_u = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{G}\left(\mathbf{x} - x_u\mathbf{e}_u\right) + \mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{H}^{\mathrm{H}}\mathbf{z}. \tag{7}$$

The covariance of the NPI $\upsilon_u^2 = \mathrm{cov}\left(\xi_u, \xi_u\right)$ is given by

$$\begin{aligned}\upsilon_u^2 &= \mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{G}\mathbf{G}\mathbf{W}^{-1}\mathbf{e}_u + \sigma_z^2\mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{G}\mathbf{W}^{-1}\mathbf{e}_u - \rho_u^2 \\ &= \rho_u - \rho_u^2.\end{aligned} \tag{8}$$

Given Equations (5), (6), and (8), we derive the *max-log* approximated LLR of the bits associated with x_u, given by

$$L\left(b_{u,k}\right) = \gamma_u\left(\min_{s\in\chi_k^0}\left|\frac{\hat{x}_u}{\rho_u} - s\right|^2 - \min_{s'\in\chi_k^1}\left|\frac{\hat{x}_u}{\rho_u} - s'\right|^2\right), \tag{9}$$

where $b_{u,k}$ is the k-th mapping bit associated with x_u; $\gamma_u = \rho_u^2/\upsilon_u^2$ is the *a posteriori* signal-to-noise-plus-interference ratio (SINR); $\chi_k^b \triangleq \{s \,|\, s \in \chi, q_k = b\}$ denotes the subset of χ, where the k-th mapping bit associated with the constellation symbol s, i.e. q_k, is b; χ is the constellation symbols set. After data detection of all users, the LLRs are fed into the soft-input channel decoder for decoding process.

B. Neumann Series Expansion

In the previous subsection, we note that the matrix inverse operations are involved in linear MMSE/ZF data detection. The matrix inverse is computational costly especially when the matrix size is large. One of the promising practical solutions to address the matrix inverse issue is to employ the Neumann

[1] According to the random matrix theory [1], [37], in massive MIMO, the probability that $\mathbf{G}$ is invertible is high.

series expansion [19], [38]. The complete Neumann series expansion of the matrix inverse $\mathbf{W}^{-1}$ is given by [19], [38]

$$\mathbf{W}^{-1} = \sum_{l=0}^{\infty} \left(\mathbf{X}^{-1}\left(\mathbf{X}-\mathbf{W}\right)\right)^{l} \mathbf{X}^{-1}, \tag{10}$$

where $\mathbf{X} \neq \mathbf{W}$, and the following conditions are satisfied:

$$\lim_{l\to\infty} \left(\mathbf{I} - \mathbf{X}^{-1}\mathbf{W}\right)^{l} = \mathbf{0}. \tag{11}$$

When the high orders are ignored, the truncated Neumann series expansion can be expressed as

$$\mathbf{W}_L^{-1} = \sum_{l=0}^{L-1} \left(\mathbf{X}^{-1}\left(\mathbf{X}-\mathbf{W}\right)\right)^{l} \mathbf{X}^{-1}. \tag{12}$$

The significance of Neumann series expansion is to approach matrix inversion by using matrix multiplications with known $\mathbf{X}^{-1}$. Specifically, if the inverse of the selected matrix $\mathbf{X}^{-1}$ is easy to obtain, then the truncated expansion $\mathbf{W}_L^{-1}$ can be used to approximate $\mathbf{W}^{-1}$. Generally, when we select the matrix $\mathbf{X}$ that is close to $\mathbf{W}$, a few L orders expansion $\mathbf{W}_L^{-1}$ in Equation (12) can be close to $\mathbf{W}^{-1}$. Fortunately, in massive MIMO systems, the gram matrix $\mathbf{G}$ is diagonally dominant; hence the diagonal matrix, i.e., $\mathbf{D} = \text{diag}\left\{\left[\mathbf{W}_{(0,0)}, \mathbf{W}_{(1,1)}, \cdots, \mathbf{W}_{(N_U-1,N_U-1)}\right]^{\mathrm{T}}\right\}$ can be selected as $\mathbf{X}$, then the L order expansion $\mathbf{W}_L^{-1}$ is given by

$$\mathbf{W}_L^{-1} = \sum_{l=0}^{L-1} \left(\mathbf{D}^{-1}\left(\mathbf{D}-\mathbf{W}\right)\right)^{l} \mathbf{D}^{-1}. \tag{13}$$

In [19], the authors have provided the upper bound of the residual estimation error [2] using $\mathbf{W}_L^{-1}$ to approach $\mathbf{W}^{-1}$, i.e.,

$$\left\| \left(\mathbf{W}^{-1} - \mathbf{W}_L^{-1}\right)\mathbf{y}^{\mathrm{MF}} \right\|_2 \leq \left\| \mathbf{I} - \mathbf{D}^{-1}\mathbf{W} \right\|_{\mathrm{F}}^{L} \left\| \hat{\mathbf{x}} \right\|_2, \tag{14}$$

where $\|\mathbf{A}\|_{\mathrm{F}}$ and $\|\mathbf{a}\|_2$ are the Frobenius norm of a matrix $\mathbf{A}$ and the ℓ_2-norm of a vector $\mathbf{a}$. From Equation (14), we can see that the upper bound of residual estimation error decreases as the increase of the expansion order and N_B. In other words, if the number of antennas at the base station is sufficiently large, even with a small order expansion, the residual estimation error will be small. Particularly, when N_B is sufficiently large and the expansion order $L \leq 2$, the computation required for the Neumann series expansion will be much reduced, compared to the matrix inverse operations. These two evidences support the use of the diagonal matrix in Neumann series expansion for massive MIMO systems.

C. Jacobi Method

In Neumann series expansion, if the expansion order is greater than 2, the matrix multiplication operations are involved; hence, the computational complexity is comparable with that of the matrix inverse operations. On the other hand, as we can see in Equation (14), if N_B is not sufficiently large, with the expansion order that is less than 2, the residual estimation error is still considerable. These two factors limit the applications of diagonal matrix in Neumann series expansion.

To avoid the matrix multiplication operations, but maintain a reasonable orders of expansion, we can use the iterative methods. To be specific, we first rewrite the MMSE estimation in Equation (2) as

$$\mathbf{W}\hat{\mathbf{x}} = \mathbf{y}^{\mathrm{MF}}. \tag{15}$$

By transforming the matrix inverse problem into the format of Equation (15), we can adopt the iterative methods to solve linear equations. Generally, the iterative methods follow the following process:

(1) Provide an initial estimation;
(2) Follow an iterative structure to obtain the next estimation;
(3) When the estimation converges, output the final estimation.

In Jacobi method, we have the initial estimation as

$$\mathbf{x}^{(0)} = \mathbf{D}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{16}$$

which is the common selection in most of the existing literature. The iterative structure is given by

$$\begin{aligned}\mathbf{x}^{(i+1)} &= \mathbf{D}^{-1}\left(\left(\mathbf{D}-\mathbf{W}\right)\mathbf{x}^{(i)} + \mathbf{y}^{\mathrm{MF}}\right) \\ &= \mathbf{x}^{(i)} - \mathbf{D}^{-1}\mathbf{W}\mathbf{x}^{(i)} + \mathbf{D}^{-1}\mathbf{y}^{\mathrm{MF}},\end{aligned} \tag{17}$$

where $\mathbf{x}^{(i)}$ denotes the i-th estimation. According to the iterative structure in Equation (17), and use the initial estimation given by Equation (16), we can derive the i-th estimation given by

$$\mathbf{x}^{(i)} = \sum_{l=0}^{i} \left(\mathbf{D}^{-1}\left(\mathbf{D}-\mathbf{W}\right)\right)^{l} \mathbf{D}^{-1}\mathbf{y}^{\mathrm{MF}}. \tag{18}$$

That is to say, by selecting the initial estimation given by (16), after i iterations following Jacobi iterative structure, we have the same estimation results as the $(i+1)$-th order expansion in Neumann series. Therefore, the convergence conditions, the residual estimation error, and the estimation results are the same as those in the previous subsection. However, as we can see from Equation (16) to Equation (17), only matrix-vector product operations are involved; therefore, Jacobi method has low complexity compared to the Neumann series expansion with the same iterations (or orders in Neumann series expansion).

III. Stair Matrix and Its Applicability to Massive MIMO Systems

In this section, we will first introduce the stair matrix and its properties. And then, we will demonstrate the applicability of the stair matrix to massive MIMO systems.

A. Stair Matrix and Its Properties

To begin with, we have the following definitions.

Definition 1 [33]: In an $N \times N$ matrix $\mathbf{A}$, if its entry $\mathbf{A}_{(m,n)} = \mathbf{e}_m^H \mathbf{A} \mathbf{e}_n$, $m, n = 1, 2, \cdots, N$, satisfies $\mathbf{A}_{(m,n)} = 0$

[2] The residual estimation error defined in [19] is to evaluate the mean-square error between the estimation using the L order expansion and the exact linear MMSE estimation, since the previous one is to approximate the latter one.

where $n \notin \{m-1, m, m+1\}$, we then call it as a tridiagonal matrix, denoted by $\mathbf{A} = \text{tridiag}\left(\mathbf{A}_{(m,m-1)}, \mathbf{A}_{(m,m)}, \mathbf{A}_{(m,m+1)}\right)$.

A typical tridiagonal matrix $\mathbf{A}$ can be given as

$$\begin{bmatrix} \mathbf{A}_{(1,1)} & \mathbf{A}_{(1,2)} & & & & \\ \mathbf{A}_{(2,1)} & \mathbf{A}_{(2,2)} & \mathbf{A}_{(2,3)} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \mathbf{A}_{(N-1,N-2)} & \mathbf{A}_{(N-1,N-1)} & \mathbf{A}_{(N-1,N)} \\ & & & \mathbf{A}_{(N,N-1)} & \mathbf{A}_{(N,N)} \end{bmatrix}$$

Definition 2 [33], [39]: If a tridiagonal matrix satisfies one of the following conditions:

(I) $\mathbf{A}_{(m,m-1)} = 0$, $\mathbf{A}_{(m,m+1)} = 0$, where $m = 2k-1$, $k = 1, 2, \cdots, \lfloor (N+1)/2 \rfloor$. Alternatively, the non-diagonal elements in the odd rows of tridiagonal matrix are zeros;

(II) $\mathbf{A}_{(m,m-1)} = 0$, $\mathbf{A}_{(m,m+1)} = 0$, where $m = 2k$, $k = 1, 2, \lfloor N/2 \rfloor$. In other words, the non-diagonal elements in the even rows of tridiagonal matrix are zeros;

we then call it as a stair matrix, denoted by $\mathbf{A} = \text{stair}\left(\mathbf{A}_{(m,m-1)}, \mathbf{A}_{(m,m)}, \mathbf{A}_{(m,m+1)}\right)$.

In accordance, a stair matrix is of type I if the condition (I) is satisfied and is of type II if the condition (II) is satisfied. For example, a 5×5 stair matrix has the following forms:

$$\mathbf{A} = \begin{bmatrix} \times & & & & \\ \times & \times & \times & & \\ & & \times & & \\ & & \times & \times & \times \\ & & & & \times \end{bmatrix} \text{ or } \mathbf{A} = \begin{bmatrix} \times & \times & & & \\ & \times & & & \\ & \times & \times & \times & \\ & & & \times & \\ & & & \times & \times \end{bmatrix}.$$

The previous one is of type I and the latter one is of type II. Next, we provide the following properties of the stair matrix in ***Corollary*** 1 and 2.

Corollary 1: Let $\mathbf{A}$ be a stair matrix. Then $\mathbf{A}^{\mathrm{H}}$ is also a stair matrix. In addition, if $\mathbf{A}$ is of type I, then $\mathbf{A}^{\mathrm{H}}$ is of type II, and vice verse.

Proof: Using the definition, it is straightforward to obtain *Corollary* 1. ∎

Corollary 1 shows that the properties of the stair matrix of type I and type II are almost the same; therefore, we only consider the stair matrix of type I hereafter except for specification.

Corollary 2: Let $\mathbf{A}$ be a stair matrix. $\mathbf{A}$ is nonsingular if and only if $\mathbf{A}_{m,m}$, $m = 1, 2, \cdots, N$, is nonzero. Furthermore, the inverse of $\mathbf{A}$, i.e., $\mathbf{A}^{-1}$ is also a stair matrix of the same type, given by $\mathbf{A}^{-1} = \mathbf{D}^{-1}(2\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}$, where $\mathbf{D} = \text{diag}\left(\left[\mathbf{A}_{(1,1)}, \mathbf{A}_{(2,2)}, \cdots, \mathbf{A}_{(N,N)}\right]\right)$ is the diagonal matrix extracted from $\mathbf{A}$.

Proof: Since $\det(\mathbf{A}) = \prod_{m=1}^{N} \mathbf{A}_{(m,m)}$, we can see that $\mathbf{A}$ is nonsingular if and only if $\mathbf{A}_{(m,m)}$, $m = 1, 2, \cdots, N$, is nonzero.

Following the matrix multiplications, we can obtain that $\mathbf{D}^{-1}(2\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}\mathbf{A} = \mathbf{I}_N$. Moreover, we can easily derive that $\mathbf{A}^{-1}$ is also a stair matrix and of the same type as $\mathbf{A}$. ∎

From *Corollary* 2, we have the **Algorithm 1** to obtain $\mathbf{A}^{-1}$. It is clear from **Algorithm 1** that the complexity to obtain the inverse of a stair matrix is $\mathcal{O}(N)$, which is the same order of the computation of $\mathbf{D}^{-1}$.

Algorithm 1 Compute the Inverse of a Stair Matrix

Input: The Stair Matrix $\mathbf{A} = \text{stair}\left(\mathbf{A}_{(m,m-1)}, \mathbf{A}_{(m,m)}, \mathbf{A}_{(m,m+1)}\right)$

Output: $\mathbf{A}^{-1} = \mathbf{B} = \text{stair}\left(\mathbf{B}_{(m,m-1)}, \mathbf{B}_{(m,m)}, \mathbf{B}_{(m,m+1)}\right)$

1.for $m = 1 : 1 : N$
2. $\mathbf{B}_{(m,m)} = 1 / \mathbf{A}_{(m,m)}$
3. end
4. for $m = 2 : 2 : 2\lfloor N/2 \rfloor$
5. $\mathbf{B}_{(m,m-1)} = -\mathbf{A}_{(m,m-1)}\mathbf{B}_{(m,m)}\mathbf{B}_{(m-1,m-1)}$;
6. $\mathbf{B}_{(m,m+1)} = -\mathbf{A}_{(m,m+1)}\mathbf{B}_{(m,m)}\mathbf{B}_{(m+1,m+1)}$;
7. end

Return B.

B. Using Stair Matrix in Neumann Series Expansion

As linear MMSE detection scheme is known to have better system performance compared to linear ZF detection scheme, we show the use of the stair matrix in approaching linear ZF estimation in this Section. In Section V, we show the corresponding results of using stair matrix in approaching linear MMSE estimation.

To begin with, we extract the stair matrix from the Gram matrix, i.e., $\mathbf{S} = \text{stair}\left(\mathbf{G}_{u,u-1}, \mathbf{G}_{u,u}, \mathbf{G}_{u,u+1}\right)$, given by

$$\mathbf{S}_{(u,v)} = \begin{cases} \mathbf{G}_{(u,v)}, & u \in \mathbb{U}_1, v = u; \\ \mathbf{G}_{(u,v)}, & u \in \mathbb{U}_2, v \in \{u-1, u, u+1\}; \\ 0, & \text{otherwise}, \end{cases}$$

where $\mathbb{U} \triangleq \{n | n \in \mathbb{N}, n \leq N_U + 1\}$, denoting the positive integer numbers less than N_U; $\mathbb{U}_1$ and $\mathbb{U}_2$ are subsets of $\mathbb{U}$, defined as $\mathbb{U}_1 \triangleq \{n | n \in \mathbb{U}, n = 2k-1, k \in \mathbb{N}\}$, denoting the odd numbers less than N_U, and $\mathbb{U}_2 \triangleq \{n | n \in \mathbb{U}, n = 2k, k \in \mathbb{N}\}$, denoting the even numbers less than N_U, respectively.

Applying the stair matrix in Neumann series expansion, similar to Equation (10), we have

$$\mathbf{G}^{-1} = \sum_{k=0}^{\infty} \left(\mathbf{I} - \mathbf{S}^{-1}\mathbf{G}\right)^k \mathbf{S}^{-1}, \tag{19}$$

where $\mathbf{X}$ is replaced with the stair matrix $\mathbf{S}$ and the Gram matrix is considered. The convergence condition for Equation (19) is

$$\lim_{k \to \infty} \left(\mathbf{I} - \mathbf{S}^{-1}\mathbf{G}\right)^k = \mathbf{0}, \tag{20}$$

or equivalently

$$\rho(\mathbf{B}) = |\lambda_0| < 1, \tag{21}$$

i.e., the maximum eigenvalue of the convergence matrix $\mathbf{B}$ is less than one. $\mathbf{B} = \mathbf{I} - \mathbf{S}^{-1}\mathbf{G}$, and $\rho(\mathbf{B})$ is the spectral radius of the matrix $\mathbf{B}$. $|\lambda_0| \geq |\lambda_1| \geq \cdots \geq |\lambda_{N_U - 1}|$ denote the N_U eigenvalues of $\mathbf{B}$.

The convergence condition is critical for the application of the stair matrix in massive MIMO systems. In order to investigate the maximum eigenvalue of $\mathbf{B}$, we suppose N_U is

$$\mathbf{B}_{(u,v)} = \begin{cases} -\dfrac{\mathbf{G}_{(u,v)}}{\mathbf{G}_{(u,u)}}, & u \in \mathbb{U}_1, v \neq u; \\ 0, & u \in \mathbb{U}_1, v = u; \\ -\dfrac{\mathbf{G}_{(u,v)}}{\mathbf{G}_{(u,u)}} + \dfrac{\mathbf{G}_{(u,u-1)} \cdot \mathbf{G}_{(u-1,v)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}} + \dfrac{\mathbf{G}_{(u,u+1)} \cdot \mathbf{G}_{(u+1,v)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}}, & u \in \mathbb{U}_2, v \neq u; \\ \dfrac{\mathbf{G}_{(u,u-1)} \cdot \mathbf{G}_{(u-1,u)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}} + \dfrac{\mathbf{G}_{(u,u+1)} \cdot \mathbf{G}_{(u+1,u)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}}, & u \in \mathbb{U}_2, v = u. \end{cases} \tag{22}$$

odd, [3] and derive each entry in $\mathbf{B}$ given by Equation (22), as shown at the top of this page, where **Algorithm 1** is used to compute the matrix inverse of the stair matrix.

We have the following theorem:

Theorem 1: $\mathbf{B}_{(u,v)}$ is given by Equation (22), and $|\lambda_0| \geq |\lambda_1| \geq \cdots \geq \left|\lambda_{N_U-1}\right|$ denote the N_U eigenvalues of $\mathbf{B}$. We have

$$\lim_{N_B \to \infty} \Pr\{|\lambda_0| < 1\} = 1. \tag{23}$$

Proof: See Appendix B. ∎

Theorem 1 shows that when the number of antennas at base station is sufficiently large, the convergence conditions in Equation (20) can be satisfied; therefore, using stair matrix in Neumann series expansion can be established. As we show in Section V, for the system with 25 active users under service, $N_B = 150$ is sufficient to meet the convergence conditions.

Hence we demonstrate the applicability of the stair matrix in massive MIMO systems.

C. Residual Estimation Error

We now investigate the residual estimation error by using the truncated Neumann series expansion. According to Equation (12), we have

$$\mathbf{G}_L^{-1} = \sum_{l=0}^{L-1} \left(\mathbf{S}^{-1}(\mathbf{S} - \mathbf{G})\right)^l \mathbf{S}^{-1}. \tag{24}$$

Replacing $\mathbf{G}^{-1}$ with $\mathbf{G}_L^{-1}$ in Equation (4), we have

$$\hat{\mathbf{x}}^{(L)} = \mathbf{G}_L^{-1}\mathbf{y}^{\mathrm{MF}}. \tag{25}$$

Therefore, the residual estimation error[4] $J = \left\|\hat{\mathbf{x}}^{(L)} - \hat{\mathbf{x}}\right\|_2$, is bounded as

$$\begin{aligned} J &= \left\|\left(\mathbf{G}^{-1} - \mathbf{G}_L^{-1}\right)\mathbf{y}^{\mathrm{MF}}\right\|_2 \\ &= \left\|\sum_{l=L}^{\infty}\left(\mathbf{S}^{-1}(\mathbf{S} - \mathbf{G})\right)^l \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}\right\|_2 \\ &= \left\|\left(\mathbf{S}^{-1}(\mathbf{S} - \mathbf{G})\right)^L \mathbf{G}^{-1}\mathbf{y}^{\mathrm{MF}}\right\|_2 \\ &\leq \|\mathbf{B}\|_{\mathrm{F}}^L \left\|\hat{\mathbf{x}}\right\|_2, \end{aligned} \tag{26}$$

where the inequality holds since $\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{x}\|_2$. As shown in Appendix B, when $N_B \to \infty$, $\Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2 < 1\right\} \to 1$ and $\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\} \to 0$. That is to say, the residual estimation error will approach 0 as indicated by inequality (26). Inequality (26) also indicates that increasing the truncation order in Neumann series expansion, the upper bound of the residual estimation error can be reduced. This evidence, together with high probability with the convergence conditions to be met, supports the applications of the stair matrix to massive MIMO systems.

[3] When N_U is even, the difference in the expression of $\mathbf{B}$ is only present in the last row. However, the general result is also expected.

[4] The residual estimation error is consistent with the definition in [19], which is to evaluate the difference between the linear MMSE estimation and the approximate estimation using truncated Neumann series expansion.

IV. Implementation of the Stair Matrix in Iterative Method

Due to the involvement of matrix multiplications, the truncation order in Neumann series expansion is limited to three; otherwise, the computational complexity is comparable with matrix inversion algorithm. Besides, we note that in existing work, the computation of the LLR is obtained by utilizing the NPI after the first truncation order in Neumann series expansion (or first iteration in iterative method). This implementation, however, causes significant performance loss when N_B is not sufficiently large (or $r = N_B/N_U$ is not large, for example, $r < 8$). In this section, we address these issues by developing of an iterative method using stair matrix.

A. Stair Matrix in Iterative Method

Compared to the linear ZF detection, linear MMSE detection achieves a better balance in consideration of interference and noise. Therefore, we will introduce an iterative method to approach the linear MMSE detection.

To start with, we extract the stair matrix $\mathbf{S} = \mathrm{stair}\left(\mathbf{W}_{(u,u-1)}, \mathbf{W}_{(u,u)}, \mathbf{W}_{(u,u+1)}\right)$. It is worth noting that compared to the stair matrix we discussed in previous section, the diagonal elements in the new stair matrix has increased by σ_z^2 according to Equation (3), which brings negligible computational cost. According to Equation (17), we have

$$\begin{aligned} \mathbf{x}^{(i+1)} &= \mathbf{S}^{-1}\left((\mathbf{S} - \mathbf{W})\mathbf{x}^{(i)} + \mathbf{y}^{\mathrm{MF}}\right) \\ &= \mathbf{x}^{(i)} - \mathbf{S}^{-1}\mathbf{W}\mathbf{x}^{(i)} + \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}, \end{aligned} \tag{27}$$

where $\mathbf{x}^{(i)}$ is the i-th estimation.

In accordance, if the initial estimation $\mathbf{x}^{(0)}$ is selected as

$$\mathbf{x}^{(0)} = \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{28}$$

following the iterative process in Equation (27), we can derive

$$\mathbf{x}^{(i)} = \sum_{l=0}^{i} \left(\mathbf{S}^{-1}(\mathbf{S} - \mathbf{W})\right)^l \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{29}$$

which indicates that the iterative method in Equation (27) can be seen as truncated Neumann series expansion based method by selecting $\mathbf{X} = \mathbf{S}$. However, in Equation (27), only matrix-vector product is involved, hence the overall computational complexity is of the order $\mathcal{O}\left(KN_U^2\right)$, where K denotes the iteration numbers.

B. Computation of the LLR

After the estimation of transmitted vector $\mathbf{x}$, we need to compute the LLRs of the associated bits for the soft-input channel decoder. After K iterations, the equivalent channel gain $\rho_u^{(K)}$ and the covariance of the NPI $\left|v_u^{(K)}\right|^2$ can be respectively given by

$$\rho_u^{(K)} = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{G}\mathbf{e}_u, \tag{30}$$

$$\left|v_u^{(K)}\right|^2 = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{G}\mathbf{G}\mathbf{W}_K^{-1}\mathbf{e}_u + \sigma_z^2\mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{G}\mathbf{W}_K^{-1}\mathbf{e}_u - \left|\rho_u^{(K)}\right|^2 \tag{31}$$

Apparently, Equations (30) and (31) requires matrix multiplications if $K \geq 2$. Therefore, in [19] and [24]–[26], $\mathbf{D}^{-1}$, where $\mathbf{D} = \mathrm{diag}\left\{\left[\mathbf{W}_{(0,0)}, \mathbf{W}_{(1,1)}, \cdots, \mathbf{W}_{(N_U-1,N_U-1)}\right]^{\mathrm{T}}\right\}$ is considered for the simplification. This approximation, however, as we will show in the next section, has caused a significant performance loss.

As we can see from Equation (8), the exact *a posteriori* covariance of the NPI in linear MMSE estimation can be derived if the equivalent channel gain is obtained. However, in [19], the authors have claimed that this relationship is not supported in truncated Neumann series expansion. The main reason for that claim is attributed to the fact that $\mathbf{W}_K^{-1}$ is far away from $\mathbf{W}^{-1}$ with small K. In previous section, we introduce the iterative method for detection, and the iteration numbers can be sufficiently large since the computational complexity in one iteration is of the order $\mathcal{O}\left(N_U^2\right)$. With sufficiently large iterations, $\mathbf{W}_K^{-1}$ can be quite close to $\mathbf{W}^{-1}$ (we will show this in the next section); hence, we can used Equation (8) to derive the covariance of the NPI. The next question is how to maintain low computational complexity to obtain the equivalent channel gain.

We rewrite the equivalent channel gain in Equation (8) as $\rho_u = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{G}\mathbf{e}_u = 1 - \sigma_z^2\mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{e}_u$. In addition, we replace $\mathbf{W}^{-1}$ with $\mathbf{W}_K^{-1}$, leading to

$$\rho_u^{(K)} = 1 - \sigma_z^2\mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{e}_u. \tag{32}$$

That is to say, we need obtain the diagonal elements in $\mathbf{W}_K^{-1}$ to compute $\rho_u^{(K)}$.

If N_B and r are sufficiently large, the Gram matrix $\mathbf{G}$ and $\mathbf{W}$ will become diagonal dominant; therefore, $\mathbf{D}^{-1}$ can be a good approximation of $\mathbf{W}^{-1}$, and we have the approximation to $\rho_u^{(K)}$ given by

$$\rho_u^{(K)} \approx 1 - \sigma_z^2\mathbf{D}_{(u,u)}^{-1}, \tag{33}$$

and $\left|v_u^{(K)}\right|^2$ is approximated as

$$\left|v_u^{(K)}\right|^2 \approx \rho_u^{(K)}\left(1 - \rho_u^{(K)}\right). \tag{34}$$

Algorithm 2 Proposed Iterative Method Using Stair Matrix

Input: $\mathbf{y}$, $\mathbf{H}$, σ_z^2, and Iteration number K;
Output: LLRs of the associated bits $L\left(b_{u,k}\right)$.
Initialization:
1.$\mathbf{G} = \mathbf{H}^{\mathrm{H}}\mathbf{H}$, $\mathbf{W} = \mathbf{G} + \sigma_z^2\mathbf{I}_{N_U}$, $\mathbf{y}^{\mathrm{MF}} = \mathbf{H}^{\mathrm{H}}\mathbf{y}$;
2. $\mathbf{S} = \mathrm{stair}\left(\mathbf{W}_{(u,u-1)}, \mathbf{W}_{(u,u)}, \mathbf{W}_{(u,u+1)}\right)$;
3. Compute $\mathbf{S}^{-1}$ through **Algorithm 1**, and $\mathbf{D}^{-1} = \mathrm{diag}\left(\mathbf{S}^{-1}\right)$;
4. Initial estimation: $\mathbf{x}^{(0)} = \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}$;
Iteration:
5. for $i = 1 : 1 : K$
6. $\mathbf{x}^{(i+1)} = \mathbf{S}^{-1}\left(\left(\mathbf{S} - \mathbf{W}\right)\mathbf{x}^{(i)} + \mathbf{y}^{\mathrm{MF}}\right)$;
7. end
LLR Computation:
8. $\rho_u^{(K)} = 1 - \sigma_z^2\mathbf{D}_{(u,u)}^{-1}$, $\gamma_u^{(K)} = \frac{\rho_u^{(K)}}{1-\rho_u^{(K)}}$;
9. $L\left(b_{u,k}\right) = \gamma_u^{(K)}\left(\min_{s\in\chi_k^0}\left|\frac{\hat{x}_u^{(K)}}{\rho_u^{(K)}} - s\right|^2 - \min_{s'\in\chi_k^1}\left|\frac{\hat{x}_u^{(K)}}{\rho_u^{(K)}} - s'\right|^2\right)$.
Return $L\left(b_{u,k}\right)$.

As a consequence, the *a posteriori* SINR is approximated as

$$\gamma_u^{(K)} \approx \frac{\left|\rho_u^{(K)}\right|^2}{\left|v_u^{(K)}\right|^2} = \frac{\rho_u^{(K)}}{1 - \rho_u^{(K)}}. \tag{35}$$

$\rho_u^{(K)}$ and $\gamma_u^{(K)}$ are used in Equation (9) to compute $L\left(b_{u,k}\right)$.

The accuracy of the approximation in Equation (33) is important for the overall system performance. In order to show the accuracy of Equation (33), we define

$$MSE = \mathrm{E}\left\{\frac{1}{N_U}\sum_{u=0}^{N_U-1}\left|\rho_u^{(K)} - \rho_u\right|^2\right\}, \tag{36}$$

to show the normalized mean-square error between the approximation $\rho_u^{(K)}$ given by Equation (33) and the exact equivalent channel gain ρ_u given by Equation (6). The results are shown in Figure 1. The results are obtained by randomly generating 2000000 channel realization, and ρ_u is computed by Equation (6), while the approximation is given by Equation (33). It is clear from Figure 1 that the approximation in Equation (33) is quite close to the exact ρ_u as MSE is at very low level. Besides, Figure 1 also indicates that with the increase of N_B or average SNR at receiver, the approximation will be more close. These results indicate our approximation given by Equation (33) works well in approaching linear MMSE detection method. We will also validate this result in next Section in BER evaluation.

It is worth pointing out that although we utilize the diagonal matrix to estimate the equivalent channel gain, the computation of $\gamma_u^{(K)}$ in Equation (35) indicates that we try to approach the SINR in linear MMSE detection to derive the LLRs of the associated bits. This is quite different from the existing work [19], [24]–[26], where the SINR after the first iteration (or the first truncation order in Neumann series expansion method) is adopted. In fact,

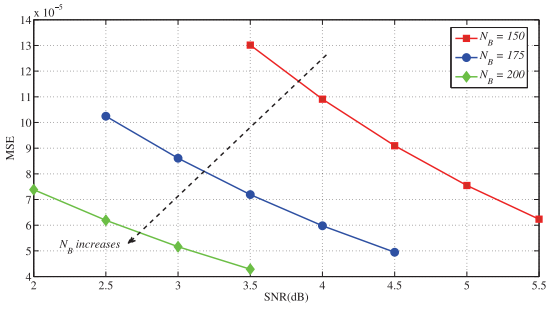


Fig. 1. The normalized mean-square error for the approximation in Equation (33), $N_U = 25$.

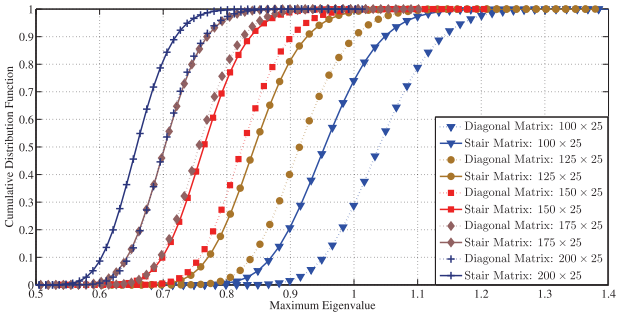


Fig. 2. Cumulative distribution function of the maximum eigenvalue $N_U = 25$.

as the iterations increase, the covariance of the NPI will decrease, and our proposed approximation method is more efficient and accurate. In numerical simulations, we also validate that our approximation in (33) and (35) outperforms the approximation in existing work.

To summarize, we present **Algorithm 2** for the proposed iterative method using stair matrix.

C. Computational Complexity Analysis

We consider the number of real number multiplications/divisions to evaluate the computational complexity. In initialization steps, the computation of $\mathbf{W}$ and $\mathbf{y}^{\mathrm{MF}}$ requires $2N_B N_U^2$ and $4N_B N_U$ real number multiplications, respectively. According to **Algorithm 1**, the computation of $\mathbf{S}^{-1}$ requires $3\left(N_U - 1\right)$ real number multiplications and N_U real number divisions. The initial estimation, provided in Step 4 of **Algorithm 2**, requires $\frac{N_U+1}{2} \times 2 + \frac{N_U-1}{2} \times (8+2) = 6N_U - 4$ real number multiplications. Therefore, the total computational complexity in initialization steps is $2N_B N_U^2 + 4N_B N_U + 10N_U - 7$.

The iteration steps in **Algorithm 2** involves matrix-vector production. The computation of $(\mathbf{S} - \mathbf{W})\,\mathbf{x}^{(i)}$ requires $\frac{N_U+1}{2} \times 4\left(N_U - 1\right) + \frac{N_U-1}{2} \times 4\left(N_U - 3\right) = 4(N_U - 1)^2$ real number multiplications. The resultant vector is multiplied by a stair matrix, and additional $6N_U - 4$ real number multiplications are required. Therefore, the total computational complexity in iteration steps is $K\left(4N_U^2 - 2N_U\right)$. That is to say, the computational complexity of the iterative process is of $\mathcal{O}\left(N_U^2\right)$, which is the same as the existing iterative methods where the diagonal matrix is applied.

Last, to obtain $L\left(b_{k,u}\right)$, we need the computation of $\rho_u^{(K)}$, and the proposed approximation method only requires the diagonal elements in $\mathbf{D}$, which is obtained in step 3. Compared to the existing work in [19] and [24]–[26], our proposed scheme saves computational complexity in this stage.

To summarize, the overall computational complexity is the same level of the existing work [19], [24]–[26]. However, as we will see in next Section, the stair matrix outperforms the diagonal matrix at all round.

V. Numerical Simulations and Performance Evaluation

A. Convergence Conditions

We first investigate the convergence condition using the stair matrix. Using Monte-Carlo method, we generate 2e7 random channel matrix $\mathbf{H}$. For each $\mathbf{H}$, we extract the diagonal matrix $\mathbf{D}$ and the stair matrix $\mathbf{S}$, and compute the maximum eigenvalues of the matrix $\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}$, and $\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}$, respectively. Using numerical simulations, we eventually obtain the cumulative distribution function (CDF) of the maximum eigenvalues, given by Figure 2. In Figure 2, we evaluate the scenario that 25 users are in service and we increase the number of antennas at base station from 100 to 200. The following observations can be found:

- With the increase of antennas at base station, the probability that the convergence conditions are met, i.e., $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right) < 1\right\}$ and $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right) < 1\right\}$ will increase. Specifically, for the usage of the diagonal matrix, the probability that the convergence conditions are met, is increase from 0.29 when $N_B = 100$, to 1 when $N_B = 200$. In accordance, for the usage of the stair matrix, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right) < 1\right\}$ is increased from 0.74 when $N_B = 100$, to 1 when $N_B = 200$;
- In low $r = N_B/N_U \leq 5$ region, the usage of the stair matrix can increase the probability that the convergence conditions are met. For example, when $N_B = 100$, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right) < 1\right\}$ is only 0.29, while $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right) < 1\right\}$ becomes 0.76. This indicates that in some low r region, the diagonal matrix is not applicable while the stair matrix can be used;
- In any system configuration, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right) < a\right\} \geq \Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right) < a\right\}$, $a \in (0,1)$. As the maximum eigenvalue determines the convergence rate, we can conclude that by using the stair matrix, the convergence rate is more likely faster compared to the use of the diagonal matrix.

The above observations validate the applicability of the usage of the stair matrix and diagonal matrix in massive MIMO systems. Besides, the results reveal that by using stair matrix, we can increase the probability that the convergence conditions are met in low r region compared to the usage of the diagonal matrix. In high r region, it is also shown that the probability that the convergence conditions are satisfied is almost 1 using both the stair matrix and the diagonal matrix. This is consistent with the discoveries in [19] where the applicability of the diagonal matrix is demonstrated. It is also revealed in [20] and [37] that in high r region, the probability that the maximum eigenvalue less than 1 is near 1. Furthermore, we also find that by using the stair matrix,

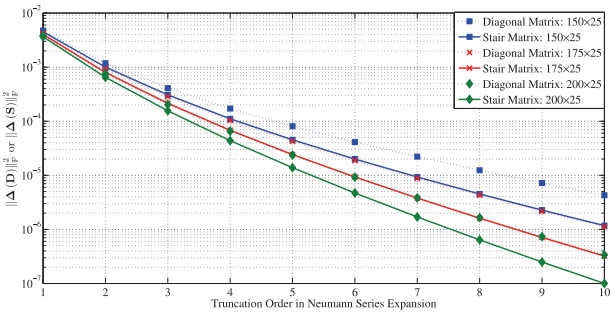


Fig. 3. Normalized mean-square error for the matrix inverse approximation.

the convergence rate is more likely accelerated than the use of the diagonal matrix.

B. Matrix Inverse

We now investigate the performance of the stair matrix in Neumann series expansion to approach the matrix inverse.[5] We define $\Delta(\mathbf{S}) = \left(\mathbf{I} - \sum_{l=0}^{L-1}\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right)^{l}\mathbf{S}^{-1}\mathbf{G}\right)$, where $\mathbf{S} = \text{stair}\left(\mathbf{G}_{u,u-1}, \mathbf{G}_{u,u}, \mathbf{G}_{u,u+1}\right)$, to show the error matrix between the identity matrix and the corresponding matrix multiplication when the approximate matrix inverse is adopted. Accordingly, $\Delta(\mathbf{D}) = \left(\mathbf{I} - \sum_{l=0}^{L-1}\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right)^{l}\mathbf{D}^{-1}\mathbf{G}\right)$ where $\mathbf{D} = \text{diag}\left(\left[\mathbf{G}_{(0,0)}, \mathbf{G}_{(1,1)}, \cdots, \mathbf{G}_{(N_U-1,N_U-1)}\right]^{\mathrm{T}}\right)$. In addition, we have $\mathrm{E}\{\frac{1}{N_U^2}\|\Delta(\mathbf{S})\|_{\mathrm{F}}^2\}$ and $\mathrm{E}\{\frac{1}{N_U^2}\|\Delta(\mathbf{D})\|_{\mathrm{F}}^2\}$ to indicate the normalized mean-square error[6] for the approximation using the stair matrix and the diagonal matrix, respectively. With different truncation order, we present the results in Figure 3. The following observations can be found:

- With the increase of the truncation order, the normalized mean-square error is decreased. This indicates that the more truncation orders used in Neumann series expansion, the more close of the resulting approximation of the matrix inverse is obtained;
- By using the stair matrix, the normalized mean-square error is always less than that of using the diagonal matrix in the same system configuration. This indicates that the use of the stair matrix always achieves better approximation performance with the same truncation order compared to the use of the diagonal matrix;
- By using the stair matrix, less iterations are required to achieve the same level of the normalized mean-square error in using the diagonal matrix. As the truncation order is equivalent to the iteration number in iterative method, the less iterations indicate less computational complexity in implementation.

[5]In implementation, we propose the iterative method as shown in section IV. However, the results of the iterative method can be seen as the Neumann series expansion.

[6]The normalized mean-square error here is consistent with the definition in [21]. This term is defined to evaluate the difference between the approximate and exact matrix inverse.

To summarize, we conclude that the use of the stair matrix outperforms the use of the diagonal matrix in terms of the matrix inverse approximation performance. As we showed in section IV.A, the truncation order is equivalent to the iterations in iterative method; therefore, the results in Figure 3 help to interpret the convergence performance of the proposed iterative method.

C. Residual Estimation Error

In iterative method, the estimation is to approach the estimation vector in linear ZF/MMSE method. In section III.C, an upper bound of the residual estimation error[7] for the use of the stair matrix in approaching linear ZF detection is presented. In order to differentiate the residual estimation error for the use of stair matrix and the diagonal matrix in linear ZF and MMSE detection, we define the following terms:

$$J(\mathbf{D}_1) = \left\| \left(\mathbf{D}_1^{-1}(\mathbf{D}_1 - \mathbf{G})\right)^{L}\mathbf{G}^{-1}\mathbf{y}^{\mathrm{MF}} \right\|_2,$$
$$J(\mathbf{D}_2) = \left\| \left(\mathbf{D}_2^{-1}(\mathbf{D}_2 - \mathbf{W})\right)^{L}\mathbf{W}^{-1}\mathbf{y}^{\mathrm{MF}} \right\|_2,$$
$$J(\mathbf{S}_1) = \left\| \left(\mathbf{S}_1^{-1}(\mathbf{S}_1 - \mathbf{G})\right)^{L}\mathbf{G}^{-1}\mathbf{y}^{\mathrm{MF}} \right\|_2,$$
$$J(\mathbf{S}_2) = \left\| \left(\mathbf{S}_2^{-1}(\mathbf{S}_2 - \mathbf{W})\right)^{L}\mathbf{W}^{-1}\mathbf{y}^{\mathrm{MF}} \right\|_2,$$

where $\mathbf{W} = \mathbf{G} + \sigma_z^2\mathbf{I}_{N_U}$, and

$$\mathbf{D}_1 = \text{diag}\left\{\left[\mathbf{G}_{(0,0)}, \mathbf{G}_{(1,1)}, \cdots, \mathbf{G}_{(N_U-1,N_U-1)}\right]^{\mathrm{T}}\right\},$$
$$\mathbf{D}_2 = \text{diag}\left\{\left[\mathbf{W}_{(0,0)}, \mathbf{W}_{(1,1)}, \cdots, \mathbf{W}_{(N_U-1,N_U-1)}\right]^{\mathrm{T}}\right\},$$
$$\mathbf{S}_1 = \text{stair}\left(\mathbf{G}_{m,m-1}, \mathbf{G}_{m,m}, \mathbf{G}_{m,m+1}\right),$$
$$\mathbf{S}_2 = \text{stair}\left(\mathbf{W}_{m,m-1}, \mathbf{W}_{m,m}, \mathbf{W}_{m,m+1}\right),$$

According to Equation (26), we can see that $J(\mathbf{D}_1)$ and $J(\mathbf{D}_2)$ denote the residual estimation error for the use of the diagonal matrix in approaching linear ZF and MMSE detection, respectively. $J(\mathbf{S}_1)$ and $J(\mathbf{S}_2)$ denote the residual estimation error for the use of the stair matrix in approaching linea ZF and MMSE detection, respectively. For a given system configuration and average receiving SNR, we present the residual estimation error performance in Figure 4. The results are present with 2000000 randomly generated channel realizations, and the following observations are found:

- From Figure 4(a) and 4(b), $\mathrm{E}\{J(\mathbf{S}_1)\}$ is always less than $\mathrm{E}\{J(\mathbf{D}_1)\}$, and $\mathrm{E}\{J(\mathbf{S}_2)\}$ is always less than $\mathrm{E}\{J(\mathbf{D}_2)\}$ after the same iteration numbers. These results reflect that after the same iterations, using the stair matrix in iterative method can approach both the linear ZF and MMSE estimation more closely compared to the use of the diagonal matrix;
- In Figure 4(a), we note that, for the use of the diagonal matrix, the residual estimation error decreases slowly

[7]We have shown in Section IV.A that the iterative estimation is equivalent to the estimation using the truncated Neumann series expansion; therefore, we use the term, residual estimation error, defined in [19] to evaluate the difference between the linear MMSE estimation and the iterative estimation.

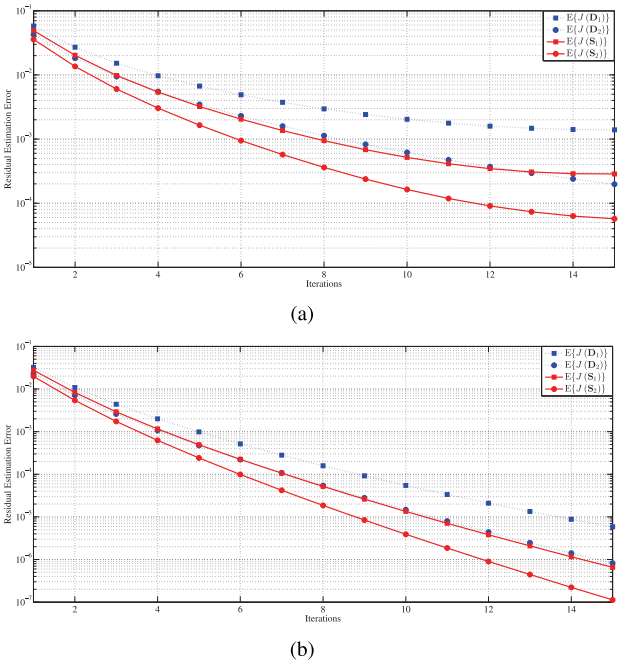


Fig. 4. Residual Estimation Error: (a) $N_B = 150$, $N_U = 25$, average SNR= 5dB; (b) $N_B = 200$, $N_U = 25$, average SNR= 3.5dB.

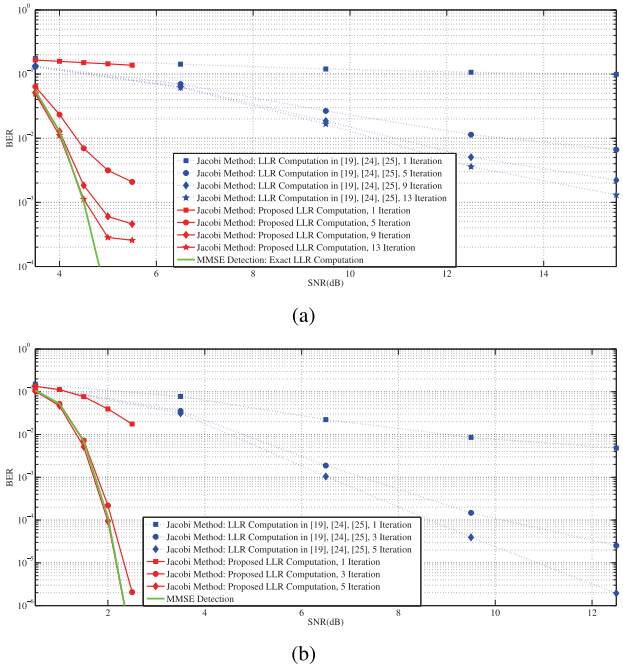


Fig. 5. BER performance: (a) $N_B = 150$, $N_U = 25$; (b) $N_B = 250$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation.

and remains a comparatively high level even with large iteration numbers. However, by using the stair matrix, we can speed up the decreasing rate and achieve a comparatively lower estimation error level. These results are consistent with the previous numerical results where we demonstrate that the use of the diagonal matrix may not be applicable in low r ratio.

- From Figure 4(a) and Figure 4(b), we can see that, with the increase of the receiving antennas at base station, the performance gain with the use of the stair matrix over the use of the diagonal matrix becomes small. These results are reasonable as N_B increases, $\mathbf{G}$ and $\mathbf{W}$ both become diagonal dominant. However, we can also achieve comparatively lower residual estimation error by using the stair matrix in iterative method.

To summarize, we conclude that the use of the stair matrix outperforms the use of the diagonal matrix in terms of the residual estimation error. The performance gain is more significant in low r ratio, but still obvious in high r ratio.

D. BER Performance

We now evaluate the system BER performance. In the system, the base station is simultaneously serving $N_U = 25$ users. For each user, a LDPC code with code length 64800, code rate 1/2 is considered for channel code scheme.[8] We consider 64QAM modulation, and a block independent channel is considered for the evaluation.

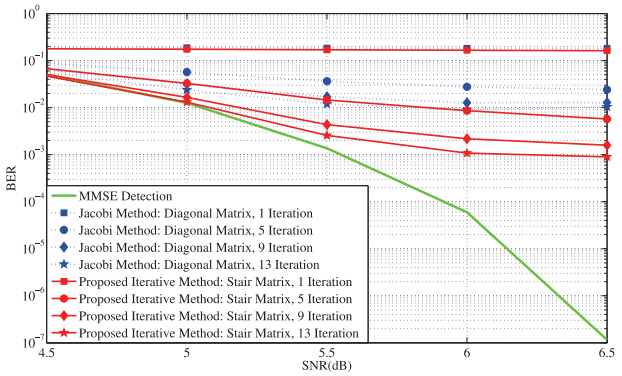
To begin with, we investigate the proposed LLR computation given by (35), and the equivalent channel gain ρ_u and the covariance of the NPI υ_u are approximated by (33) and (34).

[8]LDPC code has been an agreed standard for long code in 5G

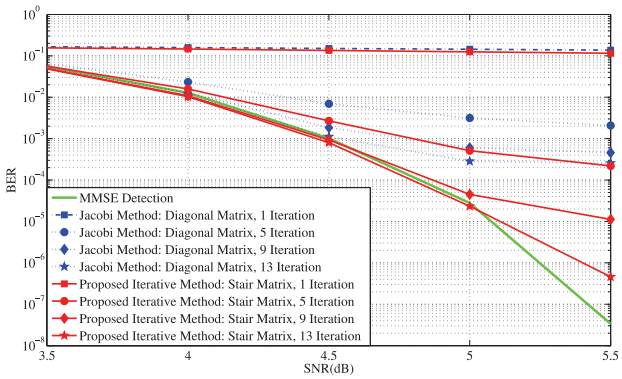
For comparison, we provide the linear MMSE detection as a benchmark, where the LLR computation is given by Equation (9) with ρ_u and υ_u given by Equation (6) and (8), respectively. The LLR computation in existing work such as [19], [24], and [25] is to compute the covariance of the NPI after the first iteration. It is worth pointing out that the iterative methods in [24] and [25] requires less iterations to approach the linear MMSE detection; however, the LLR computation used in MMSE detection is not computed from the exact NPI of the MMSE detection, but the NPI after the first iteration. In Figure 5, we can see that the BER performance of the Jacobi method with the LLR computation in [19], [24], and [25] is far away from the BER performance of the MMSE detection with the exact LLR computation. This is consistent with our previous analysis, where we pointed out that the covariance of the NPI will decrease with iterations. However, we note that the proposed LLR computation can greatly improve the BER performance of the Jacobi method by approximating the covariance of the NPI of the MMSE detection. Hereafter, we only utilize the proposed LLR computation for the BER performance comparison.

We now present the results with low $r = N_B/N_U$ region, and the results are presented in Figure 6. The following observations are found.

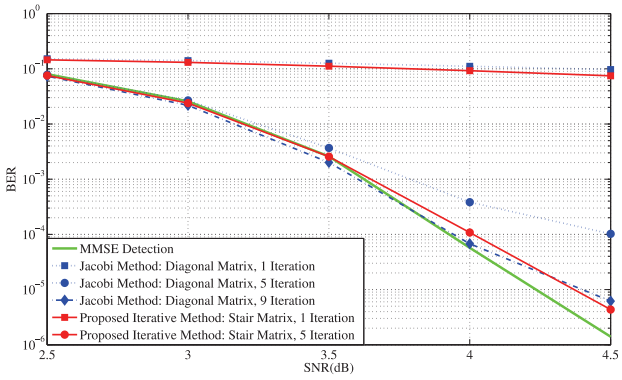
- From Figure 6(a), we note that the BER performance improvement with the proposed stair matrix compared to the diagonal matrix is obvious. However, the system performance is still far away from the MMSE detection even with sufficiently large iterations. Specially, for the use of the diagonal matrix, the performance levels off after 9 iterations; for the use of the stair matrix, the performance is greatly improved, but a leveled off



(a)



(b)



(c)

Fig. 6. BER performance: (a) $N_B = 125$, $N_U = 25$; (b) $N_B = 150$, $N_U = 25$; (c) $N_B = 175$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation.

performance still appears. These are attributed to the slow convergence rate and not a 100 percent convergence conditions satisfied;

- From Figure 6(b) and Figure 6(c), we can see that the BER performance eventually converges to the performance of the MMSE detection. Specifically, in the system configuration $N_B = 150$, $N_U = 25$, at SNR$= 5dB$, the BER performance of the proposed iterative method after 13 iterations is almost the same as the performance of the MMSE detection. In the system configuration $N_B = 175$, $N_U = 25$, at SNR$= 4dB$, the BER performance of the Jacobi method after 9 iterations approaches the performance of the MMSE detection;

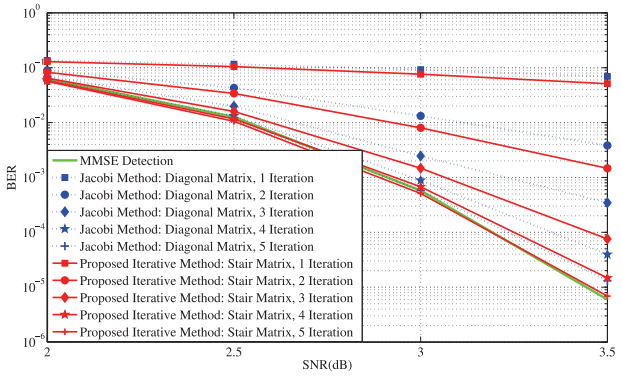


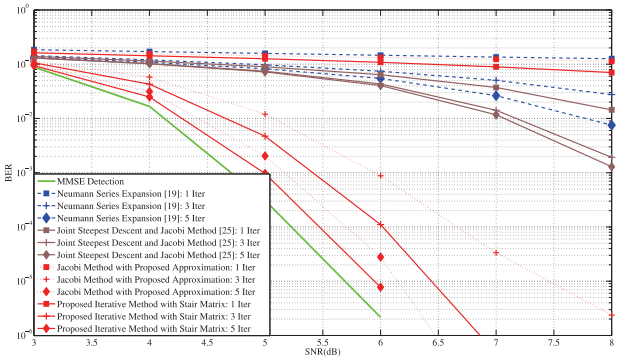
Fig. 7. BER performance: $N_B = 200$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation.

- From Figure 6(a) to Figure 6(c), we can see that the convergence rate of the proposed iterative method is faster than that of the Jacobi method. These results are consistent with the previous analysis. With faster convergence rate, fewer iterations are required for the proposed iterative method, hence reducing the overall system computational complexity.

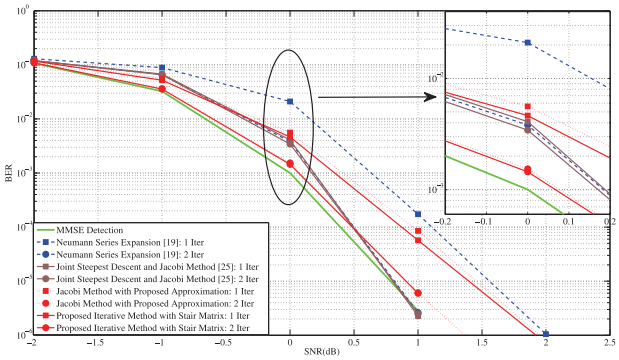
Next, we evaluate the BER performance in the system configuration with high $r = N_B / N_U$ region, and the results are shown in Figure 7. It is clear that both the uses of the diagonal matrix and stair matrix require few iterations to converge. However, as indicated by the cumulative distribution function of the maximum eigenvalue, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right) < a\right\} \geq \Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right) < a\right\}$, $a \in (0,1)$, we can conclude that the convergence rate of the proposed iterative method using the stair matrix is faster than that of the Jacobi method using the diagonal matrix. The results validate these conclusions.

In order to show the applications of the proposed scheme in a practical system, we evaluate the system BER performance in an OFDM system where the extended vehicular A (EVA) channels are used to generate the channel data for each user [40], [41]. We assume the channel data remains constant during one OFDM symbol block, and varies independently from one block to another [11], [42]. For each user, the LDPC encoder with code length 1152, code rate 0.5, is adopted. We consider 64QAM modulation, and all users takes up 192 subcarriers for data transmission. As we mentioned in Section II, the received signal model is established at each subcarrier. Therefore, the estimation is performed over individual subcarrier. We also provid the existing proposals in [19] and [25] for comparison. The results are presented in Figure 8, with the following observations found:

- In Figure 8(a), the number of antennas at base station is 150, and the user number under service is 25. In this comparatively low r region, we can see that the performance of existing proposals are far away from that of the exact linear MMSE detection scheme. This is obvious as we pointed out previously that the NPI in the initial estimation is far away from the exact NPI in linear MMSE detection. However, both Jacobi method with the diagonal matrix in the development and the proposed iterative method with the stair matrix can eventually approach



(a)



(b)

Fig. 8. BER performance: (a) $N_B = 150$, $N_U = 25$; (b) $N_B = 400$, $N_U = 25$. LDPC code: code length 1152, code rate 0.5. 64QAM modulation.

the performance of the linear MMSE detection scheme. In addition, we can see that the BER performance using stair matrix is always much better than that using the diagonal matrix, which is consistent with our previous demonstration;

- In Figure 8(b), $N_B = 400$ and $N_U = 25$. In this scenario, $r = 16$, which falls the application region of Neumann series expansion based detection scheme [19]. We can see that after two iterations, the performance of the Neumann series expansion based detection scheme converges to linear MMSE detection even with the NPI in the initial estimation for LLR computation. These results are consistent with the observations in [19]. In this comparatively high r region, it is clear that the BER performance of all detection methods eventually converge to that of the linear MMSE detection scheme.

The results in Figure 8 validate the performance enhancement of the proposed iterative method using stair matrix in the development over the existing proposals where the diagonal matrix is adopted, especially in a comparatively low r region. This is consistent with our previous analysis. It is also foreseeable that applying the proposed scheme in SC-FDMA system, the performance enhancement can be achieved since the received signal model in Equation (1) holds over each subcarrier as well.

VI. Conclusions and Future Work

In this paper, we propose the application of the stair matrix in massive MIMO systems. To begin with, we demonstrate that with sufficiently large number of antennas at base station, the probability that the convergence conditions are met with the use of the stair matrix approaches 1. We then propose an iterative method to reduce the computational complexity and show that the overall computational complexity is of the same level as the existing iterative methods where the diagonal matrix is applied. Furthermore, we evaluate the performance of the stair matrix in terms of the probability that the convergence conditions are met, the normalized mean-square error in Neumann series expansion to approach the matrix inverse, the residual estimation error of the iterative method to approach the linear ZF/MMSE estimation, and the system BER performance. Numerical simulations show that performance enhancement by using the stair matrix over the diagonal matrix is presented in all performance metrics. In future work, we will take channel estimation error and fast time-varying channel in consideration for the development.

Appendix

A. Preliminaries

We first present the preliminary lemmas.

Lemma 1: Let $a_k \sim CN(0,1)$, we then have

$$\mathrm{E}\left\{|a_k|^2\right\} = 1, \tag{37}$$

$$\mathrm{E}\left\{|a_k|^4\right\} = 2, \tag{38}$$

$$\mathrm{E}\left\{|a_k|^6\right\} = 6, \tag{39}$$

$$\mathrm{E}\left\{|a_k|^8\right\} = 24, \tag{40}$$

Proof: We first obtain the joint probability density function (PDF) $f(a_k) = \frac{1}{\pi}\exp\left(-|a_k|^2\right)$, and then apply $\mathrm{E}\{g(a_k)\} = \int_{\mathbb{C}} g(a_k)\cdot f(a_k)\,da_k$ to obtain the results in Equations (37) - (40). ■

Lemma 2: Let $\mathbf{a} = \left[a_1, a_2, \cdots, a_{N_B}\right]^{\mathrm{T}}$ with each entry $a_k \sim CN(0,1)$, independent and identically distributed (i.i.d.). We then have

$$\mathrm{E}\left\{\mathbf{a}^{\mathrm{H}}\mathbf{a}\right\} = N_B, \tag{41}$$

$$\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^4\right\} = A_3, \tag{42}$$

$$\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^{-4}\right\} = \frac{1}{B_1}, \tag{43}$$

where

$$A_3 = 24N_B + N_B(N_B-1)(N_B-2)(N_B-3) + 36N_B(N_B-1) + 12N_B(N_B-1)(N_B-2), \tag{44}$$

$$B_1 = (N_B-1)(N_B-2)(N_B-3)(N_B-4). \tag{45}$$

Proof: Through $\mathrm{E}\left\{\mathbf{a}^{\mathrm{H}}\mathbf{a}\right\} = \mathrm{E}\left\{\sum_{k=1}^{N_B}|a_k|^2\right\} = \sum_{k=1}^{N_B}\mathrm{E}\left\{|a_k|^2\right\}$ and the results from Equation (37), we can derive the results in Equation (41).

We first write $\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^4 = \sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}|a_k|^2|a_l|^2|a_m|^2|a_n|^2$. Therefore, we have $\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^4\right\}$ given by Equation (46), as shown at the top of the next page. The results can be derived

$$\begin{aligned}\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^4\right\} &= \mathrm{E}\left\{\sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}|a_k|^2|a_l|^2|a_m|^2|a_n|^2\right\}\\ &= N_B\mathrm{E}\left\{|a_k|^8\right\} + 4N_B\left(N_B-1\right)\mathrm{E}\left\{|a_k|^6\right\}\mathrm{E}\left\{|a_l|^2\right\} + 6N_B\left(N_B-1\right)\left(N_B-2\right)\mathrm{E}\left\{|a_k|^4\right\}\mathrm{E}\left\{|a_l|^2\right\}\mathrm{E}\left\{|a_m|^2\right\}\\ &\quad + 3N_B\left(N_B-1\right)\mathrm{E}\left\{|a_k|^4\right\}\mathrm{E}\left\{|a_l|^4\right\} + N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)\mathrm{E}\left\{|a_k|^2\right\}\mathrm{E}\left\{|a_l|^2\right\}\mathrm{E}\left\{|a_m|^2\right\}\mathrm{E}\left\{|a_n|^2\right\}\end{aligned} \tag{46}$$

$$\begin{aligned}\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^8\right\} &= \sum_{k=1}^{N_B}\sum_{k_1=1}^{N_B}\sum_{l=1}^{N_B}\sum_{l_1=1}^{N_B}\sum_{m=1}^{N_B}\sum_{m_1=1}^{N_B}\sum_{n=1}^{N_B}\sum_{n_1=1}^{N_B}\mathrm{E}\left\{a_k^*a_{k_1}^*a_la_{l_1}a_m^*a_{m_1}^*a_na_{n_1}\right\}\mathrm{E}\left\{b_kb_{k_1}b_l^*b_{l_1}^*b_mb_{m_1}b_n^*b_{n_1}^*\right\}\\ &= \alpha_1\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^8 + \alpha_2\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^4\left(\mathrm{E}\left\{|a_l|^4\right\}\right)^2 + \alpha_3\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^2\left(\mathrm{E}\left\{|a_l|^6\right\}\right)^2\\ &\quad + \alpha_4\left(\mathrm{E}\left\{|a_k|^4\right\}\right)^4 + \alpha_5\mathrm{E}\left\{|a_k|^8\right\}\mathrm{E}\left\{|a_l|^8\right\}\end{aligned} \tag{52}$$

$$\begin{aligned}\mathrm{E}\left\{|A|^4\right\} &= \sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}\sum_{o=1}^{N_B}\sum_{p=1}^{N_B}\sum_{q=1}^{N_B}\sum_{r=1}^{N_B}\mathrm{E}\left\{a_k^*a_ma_o^*a_qb_kb_l^*b_m^*b_nb_ob_p^*b_q^*b_rc_lc_n^*c_pc_r^*\right\}\\ &= \alpha_1\left(\mathrm{E}\left\{|a_k|^4\right\}\right)^2\mathrm{E}\left\{|a_l|^8\right\} + \alpha_2\left(\mathrm{E}\left\{|a_k|^4\right\}\right)^4 + \alpha_3\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^4\left(\mathrm{E}\left\{|a_l|^4\right\}\right)^2\\ &\quad + \alpha_4\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^3\mathrm{E}\left\{|a_l|^4\right\}\mathrm{E}\left\{|a_m|^6\right\} + \alpha_5\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^8 + \alpha_6\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^6\mathrm{E}\left\{|a_l|^4\right\}\end{aligned} \tag{55}$$

as follows. Since the elements in $\mathbf{a}$ are i.i.d., the non-zero terms come from 5 cases:

- **Case1:** $\mathrm{E}\left\{|a_k|^8\right\}$, which corresponds to $k=l=m=n$;
- **Case2:** $\mathrm{E}\left\{|a_k|^6\right\}\mathrm{E}\left\{|a_l|^2\right\}$, which corresponds to $k=m=n$ and $k\neq l$;
- **Case3:** $\mathrm{E}\left\{|a_k|^4\right\}\mathrm{E}\left\{|a_l|^2\right\}\mathrm{E}\left\{|a_m|^2\right\}$, which corresponds to $k=n, k\neq l\neq m$;
- **Case4:** $\mathrm{E}\left\{|a_k|^4\right\}\mathrm{E}\left\{|a_l|^4\right\}$, which corresponds to $k=m, l=n, k\neq l$;
- **Case5:** $\mathrm{E}\left\{|a_k|^2\right\}\mathrm{E}\left\{|a_l|^2\right\}\mathrm{E}\left\{|a_m|^2\right\}\mathrm{E}\left\{|a_n|^2\right\}$, which corresponds to $k\neq l\neq m\neq n$.

After expansion, we can see there are:

- N_B terms in **Case1**;
- $\binom{4}{3}N_B\left(N_B-1\right)$ terms in **Case2**;
- $\binom{4}{2}N_B\left(N_B-1\right)\left(N_B-2\right)$ terms in **Case3**;
- $\frac{1}{2}\binom{4}{2}N_B\left(N_B-1\right)$ terms in **Case4**;
- $N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$ terms in **Case5**.

Using the results in Equations (37) - (40), we derive the result in Equation (42).

Equation (43) is obtained by noting that $\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^{-1}$ follows an inverse-Gamma distribution, and the 4-th moment is obtained in [19]. ■

Lemma 3: Let $\mathbf{a} = \left[a_1, a_2, \cdots, a_{N_B}\right]^{\mathrm{T}}$, $\mathbf{b} = \left[b_1, b_2, \cdots, b_{N_B}\right]^{\mathrm{T}}$, with each entry $a_k \sim CN\left(0,1\right)$, $b_k \sim CN\left(0,1\right)$, and i.i.d., we then have

$$\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^4\right\} = A_1, \tag{47}$$

$$\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^8\right\} = A_5. \tag{48}$$

where

$$A_1 = 2N_B\left(N_B+1\right), \tag{49}$$

$$\begin{aligned}A_5 &= 576N_B + 24N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)\\ &\quad + 864N_B\left(N_B-1\right) + 288N_B\left(N_B-1\right)\left(N_B-2\right).\end{aligned} \tag{50}$$

Proof: We first write

$$\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^2 = \left(\sum_{k=1}^{N_B}a_k^*b_k\right)\left(\sum_{l=1}^{N_B}a_lb_l^*\right) = \sum_{k=1}^{N_B}\sum_{l=1}^{N_B}a_k^*a_lb_l^*b_k.$$

In addition, we have

$$\begin{aligned}\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^4\right\} &= \mathrm{E}\left\{\sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}a_k^*a_la_m^*a_nb_l^*b_kb_n^*b_m\right\}\\ &= \sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}\mathrm{E}\left\{a_k^*a_la_m^*a_nb_l^*b_kb_n^*b_m\right\}\\ &= 2N_B\left(N_B-1\right)\mathrm{E}\left\{|a_k|^2\right\}^2\mathrm{E}\left\{|b_k|^2\right\}^2\\ &\quad + N_B\mathrm{E}\left\{|a_k|^4\right\}\mathrm{E}\left\{|b_k|^4\right\}.\end{aligned} \tag{51}$$

In (51), the i.i.d. assumption is used for the derivation and only non-zero terms are considered. With the results in Equations (37) and (38), we derive the result in Equation (47). Moreover, we write $\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^{8}\right\}$ as Equation (52), as shown at the top of the previous page, where only the non-zero terms are considered, and the i.i.d. assumption is used for the derivation. The coefficients α_i, $i = 1, 2, \cdots, 5$, account for the number of the non-zero terms, and are given by

$$\begin{aligned}
\alpha_1 &= \binom{4}{1}\cdot\binom{3}{1}\cdot\binom{2}{1} N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)\\
&= 24N_B\left(N_B-1\right)\left(N_B-2\right),\\
\alpha_2 &= \binom{4}{2}\cdot\binom{4}{2}\cdot 2N_B\left(N_B-1\right)\left(N_B-2\right)\\
&= 72N_B\left(N_B-1\right)\left(N_B-2\right),\\
\alpha_3 &= \binom{4}{1}\cdot\binom{4}{1} N_B\left(N_B-1\right) = 16N_B\left(N_B-1\right),\\
\alpha_4 &= \frac{1}{2}\binom{4}{2}\cdot\binom{4}{2} N_B\left(N_B-1\right) = 18N_B\left(N_B-1\right),\\
\alpha_5 &= N_B.
\end{aligned}$$

With the results in Equations (37) - (40), we have the result in Equation (48). ∎

Lemma 4: Let $A = \mathbf{a}^{\mathrm{H}}\mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{c}$, where $\mathbf{a} = \left[a_1, a_2, \cdots, a_{N_B}\right]^{\mathrm{T}}$, $\mathbf{b} = \left[b_1, b_2, \cdots, b_{N_B}\right]^{\mathrm{T}}$, and $\mathbf{c} = \left[c_1, c_2, \cdots, c_{N_B}\right]^{\mathrm{T}}$, with each entry $a_k \sim CN(0,1)$, $b_k \sim CN(0,1)$, and $c_k \sim CN(0,1)$, and i.i.d., we then have

$$\mathrm{E}\left\{|A|^4\right\} = A_2, \tag{53}$$

where

$$\begin{aligned}
A_2 = {} & 96N_B + 4N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)\\
& + 144N_B\left(N_B-1\right) + 48N_B\left(N_B-1\right)\left(N_B-2\right),
\end{aligned} \tag{54}$$

Proof: We first write

$$A = \mathbf{a}^{\mathrm{H}}\mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{c} = \sum_{k=1}^{N_B} a_k^* b_k \sum_{l=1}^{N_B} b_l^* c_l = \sum_{k=1}^{N_B}\sum_{l=1}^{N_B} a_k^* b_k b_l^* c_l,$$

and then we have

$$|A|^2 = \sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B} a_k^* a_m b_k b_l^* b_m^* b_n c_l c_n^*.$$

Therefore, we have $\mathrm{E}\left\{|A|^4\right\} = \mathrm{E}\left\{|A|^2\cdot|A|^2\right\}$ given by Equation (55), where

$$\begin{aligned}
\alpha_1 &= N_B,\\
\alpha_2 &= N_B\left(N_B-1\right),\\
\alpha_3 &= 4N_B\left(N_B-1\right)\left(N_B-2\right) + 8N_B\left(N_B-1\right),\\
\alpha_4 &= 8N_B\left(N_B-1\right),\\
\alpha_5 &= 4N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right),\\
\alpha_6 &= 16N_B\left(N_B-1\right)\left(N_B-2\right).
\end{aligned}$$

In Equation (55), as shown at the top of the previous page, only the non-zero terms are considered in the expectation and the i.i.d. assumption is used for the derivation. We summarize the non-zero terms in TABLE I. It worth noting that we only provides an example to show how the non-zero term is generated, and the coefficients correspond to similar cases that a non-zero term is generated. For example, for the non-zero term $\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^8$, except for the cases in TABLE I, we also have the following cases to generate the non-zero term.

$$\begin{aligned}
&k = q, m = o, l = n, p = r, \quad k \neq o \neq l \neq p;\\
&k = m, o = q, l = r, p = n, \quad k \neq o \neq l \neq p;\\
&k = q, m = o, l = r, p = n, \quad k \neq o \neq l \neq p.
\end{aligned}$$

That is why the coefficient 4 is present. The same for other non-zero terms in TABLE I.

With the results in Equations (37) - (40), we derive the result in Equation (53). ∎

Lemma 5: Let $A = \mathbf{a}^{\mathrm{H}}\mathbf{a}\mathbf{b}^{\mathrm{H}}\mathbf{b}\mathbf{c}^{\mathrm{H}}\mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{d}\mathbf{a}^{\mathrm{H}}\mathbf{c}\mathbf{d}^{\mathrm{H}}\mathbf{a}$, where $\mathbf{a} = \left[a_1, a_2, \cdots, a_{N_B}\right]^{\mathrm{T}}$, $\mathbf{b} = \left[b_1, b_2, \cdots, b_{N_B}\right]^{\mathrm{T}}$, $\mathbf{c} = \left[c_1, c_2, \cdots, c_{N_B}\right]^{\mathrm{T}}$, and $\mathbf{d} = \left[d_1, d_2, \cdots, d_{N_B}\right]^{\mathrm{T}}$, with each entry $a_k \sim CN(0,1)$, $b_k \sim CN(0,1)$, $c_k \sim CN(0,1)$, and $d_k \sim CN(0,1)$, and i.i.d., we then have

$$E\left\{A^2\right\} = A_4, \tag{56}$$

where

$$\begin{aligned}
A_4 = {} & N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^3\\
& + 26N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2\\
& + 46N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2\\
& + 4N_B\left(N_B-1\right)^2\left(N_B-2\right)^3\left(N_B-3\right)\\
& + 220N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)\\
& + 48N_B\left(N_B-1\right)^2\left(N_B-2\right)^2\left(N_B-3\right)\\
& + 808N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)\\
& + 128N_B\left(N-1\right)^2\left(N_B-2\right)\left(N_B-3\right)\\
& + 832N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)\\
& + 40N_B\left(N_B-1\right)^2\left(N_B-2\right)^3\\
& + 600N_B\left(N_B-1\right)\left(N_B-2\right)^3\\
& + 4N_B\left(N_B-1\right)^3\left(N_B-2\right)^2\\
& + 576N_B\left(N_B-1\right)^2\left(N_B-2\right)^2\\
& + 3480N_B\left(N_B-1\right)\left(N_B-2\right)^2\\
& + 64N_B\left(N_B-1\right)^3\left(N_B-2\right)\\
& + 2592N_B\left(N_B-1\right)^2\left(N_B-2\right)\\
& + 8064N_B\left(N_B-1\right)\left(N_B-2\right)\\
& + 256N_B\left(N_B-1\right)^3 + 4352N_B\left(N_B-1\right)^2\\
& + 9888N_B\left(N_B-1\right) + 2304N_B.
\end{aligned} \tag{57}$$

Proof: We first write

$$A = \sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}\sum_{o=1}^{N_B}\sum_{p=1}^{N_B} |a_k|^2|b_l|^2 c_m^* b_m b_n^* d_n a_o^* c_o d_p^* a_p,$$

we then have $\mathrm{E}\left\{|A|^2\right\}$ given by Equation (58), where the i.i.d. assumption is used for the derivation. Similar to the process in ***Lemma 4***, we only consider the non-zero terms in the expectation, and summarize the non-zero terms in TABLE II.

TABLE I
NON-ZERO TERMS IN EQUATION (55)

Non-zero Terms	Cases
$\left(\mathrm{E}\left\{\left\|a_k\right\|^4\right\}\right)^2\mathrm{E}\left\{\left\|a_l\right\|^8\right\}$	$k=l=m=n=o=p=q=r$: N_B
$\left(\mathrm{E}\left\{\left\|a_k\right\|^4\right\}\right)^4$	$k=m=o=q, l=n=p=r, k\neq l$: $N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^4\left(\mathrm{E}\left\{\left\|a_l\right\|^4\right\}\right)^2$	$k=m=o=q, l=n, p=r, k\neq l\neq p: 2\cdot 2N_B\left(N_B-1\right)\left(N_B-2\right)$ or $k=m=l=n, o=q=p=r, k\neq o: 4\cdot 2N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^3\mathrm{E}\left\{\left\|a_l\right\|^4\right\}\mathrm{E}\left\{\left\|a_m\right\|^6\right\}$	$k=m=o=q=l=n, p=r, k\neq p: 4\cdot 2N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^8$	$k=m, o=q, l=n, p=r, k\neq o\neq l\neq p: 4N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^6\mathrm{E}\left\{\left\|a_l\right\|^4\right\}$	$k=m, o=q=l=n, p=r, k\neq o\neq p: 16N_B\left(N_B-1\right)\left(N_B-2\right)$

We take the non-zero term $\left(\mathrm{E}\left(\left|a_k\right|^2\right)\right)^{12}$ as an example to show the number of terms as follows. In Equation (58), as shown at the bottom of this page, the following cases contribute to the expected non-zero term:

$$\begin{aligned}
&o=p=m=n, o_1=p_1=m_1=n_1,\\
&\qquad k\neq k_1\neq o\neq o_1, l\neq l_1\neq o\neq o_1;\\
&o=o_1=n=n_1, p=p_1=m=m_1,\\
&\qquad k\neq k_1\neq o\neq p, l\neq l_1\neq o\neq p;\\
&o=o_1, p=p_1, m=m_1, n=n_1,\\
&\qquad k\neq k_1\neq o\neq p, l\neq l_1\neq m\neq n;\\
&o=o_1=n=n_1, p=p_1, m=m_1,\\
&\qquad k\neq k_1\neq o\neq p, l\neq l_1\neq o\neq p;\\
&o=o_1, n=n_1, p=p_1=m=m_1,\\
&\qquad k\neq k_1\neq o\neq p, l\neq l_1\neq o\neq p.
\end{aligned}$$

$N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2$ terms for each of the first two cases, and $N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^3$ terms for the third case, and $N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2$ terms for the lase two cases, which leads to the final number of terms in TABLE II for the non-zero term $\left(\mathrm{E}\left(\left|a_k\right|^2\right)\right)^{12}$. Following the similar analysis, we can derive the rest non-zero terms in TABLE II.

Using the results in Equations (37) - (40), we have the result in (56). ■

B. Proof of *Theorem 1*

To begin with, we have the following Lemma:

Lemma 6: $\mathbf{B}_{(u,v)}$ is given by Equation (22). When $N_B>4$, we have

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}\leq\sqrt{\frac{A_1}{B_1}}, \tag{59}$$

if $u\in\mathbb{U}_1, v\neq u$;

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}\leq\frac{\sqrt{A_2}}{B_1}, \tag{60}$$

if $v\in\{u-1,u+1\}$;

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}\leq\sqrt{\frac{12A_2A_3+6A_1A_3^2+24A_4+48\sqrt{A_1A_2A_3^3}}{B_1^3}} \tag{61}$$

if $u\in\mathbb{U}_2$, $v\notin\{u-1,u,u+1\}$;

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,u)}\right|^2\right\}\leq\sqrt{\frac{16A_3A_5}{B_1^3}}, \tag{62}$$

if $u\in\mathbb{U}_2$, $v=u$. A_1-A_5 and B_1 are respectively given by Equations (49), (54), (44), (57), (50), (45).

Proof: The derivation of inequations (59), (60), (61), and (62), are respectively detailed in Appendix C, D, E, F. ■

With the results in ***Lemma 6***, we have

$$\begin{aligned}
\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\}&=\sum_{u=1}^{N_U}\sum_{v=1}^{N_U}\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}\\
&\leq\frac{N_U^2-1}{2}\sqrt{\frac{A_1}{B_1}}+\left(N_U-1\right)\frac{\sqrt{A_2}}{B_1}\\
&\quad+\frac{\left(N_U-1\right)}{2}\sqrt{\frac{16A_3A_5}{B_1^3}}\\
&\quad+\frac{N_U^2-4N_U+3}{2}\\
&\quad\times\sqrt{\frac{12A_2A_3+6A_1A_3^2+24A_4+48\sqrt{A_1A_2A_3^3}}{B_1^3}}
\end{aligned} \tag{63}$$

Apparently, at the right hand side of the inequality (63), as the power in numerator is much less than that in denominator,

$$\begin{aligned}
\mathrm{E}\left\{|A|^2\right\}=&\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\sum_{p=1}^{N}\sum_{k_1=1}^{N}\sum_{l_1=1}^{N}\sum_{m_1=1}^{N}\sum_{n_1=1}^{N}\sum_{o_1=1}^{N}\sum_{p_1=1}^{N}\mathrm{E}\left\{\left|a_k\right|^2\left|a_{k_1}\right|^2a_o^*a_pa_{o_1}a_{p_1}^*\right\}\mathrm{E}\left\{\left|b_l\right|^2\left|b_{l_1}\right|^2b_mb_n^*b_{m_1}^*b_{n_1}\right\}\\
&\cdot\mathrm{E}\left\{c_m^*c_oc_{m_1}c_{o_1}^*\right\}\mathrm{E}\left\{d_nd_p^*d_{n_1}^*d_{p_1}\right\}
\end{aligned} \tag{58}$$

TABLE II
NON-ZERO TERMS IN EQUATION (58)

Non-zero Terms	Number of Terms
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{12}$	$2N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2+N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^3$ $+2N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{10}\mathrm{E}\left\{\lvert a_l\rvert^4\right\}$	$12N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2+20N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$ $+2N_B(N_B-1)^2(N_B-2)^3\left(N_B-3\right)+20N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{9}\mathrm{E}\left\{\lvert a_l\rvert^6\right\}$	$8N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)+4N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2$ $+4N_B(N_B-1)^2\left(N_B-2\right)^2\left(N_B-3\right)+8N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{8}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2$	$5N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2+6N_B(N_B-1)^2\left(N_B-2\right)^2\left(N_B-3\right)$ $+45N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)+8N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$ $+50N_B\left(N_B-1\right)\left(N_B-2\right)^2+N_B(N_B-1)^3(N_B-2)^2+10N_B(N_B-1)^2(N_B-2)^3$ $+8N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)+50N_B\left(N_B-1\right)\left(N_B-2\right)^3$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{8}\mathrm{E}\left\{\lvert a_l\rvert^8\right\}$	$2N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{7}\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$8N_B(N_B-1)^2\left(N_B-2\right)\left(N_B-3\right)+28N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$ $+40N_B\left(N_B-1\right)\left(N_B-2\right)+4N_B(N_B-1)^3\left(N_B-2\right)$ $+24N_B(N_B-1)^2(N_B-2)^2+40N_B\left(N_B-1\right)\left(N_B-2\right)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{6}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^3$	$38N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)+34N_B(N_B-1)^2(N_B-2)^2$ $+4N_B(N_B-1)^2\left(N_B-2\right)\left(N_B-3\right)+50N_B\left(N_B-1\right)\left(N_B-2\right)^3$ $+40N_B\left(N_B-1\right)\left(N_B-2\right)+2N_B(N_B-1)^3\left(N_B-2\right)+40N_B\left(N_B-1\right)\left(N_B-2\right)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{6}\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$4N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)+2N_B(N_B-1)^2\left(N_B-2\right)+10N_B\left(N_B-1\right)\left(N_B-2\right)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{6}\left(\mathrm{E}\left\{\lvert a_l\rvert^6\right\}\right)^2$	$8N_B\left(N_B-1\right)+4N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)+4N_B(N_B-1)^3$ $+8N_B(N_B-1)^2\left(N_B-2\right)+8N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{5}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$12N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)+60N_B(N_B-1)^2\left(N_B-2\right)+40N_B\left(N_B-1\right)\left(N_B-2\right)^2$ $+16N_B\left(N_B-1\right)+4N_B(N_B-1)^3+16N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{5}\mathrm{E}\left\{\lvert a_l\rvert^6\right\}\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$4N_B(N_B-1)^2+4N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{4}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^4$	$N_B(N_B-1)^2(N_B-2)^2+8N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)+65N_B\left(N_B-1\right)\left(N_B-2\right)^2$ $+32N_B(N_B-1)^2\left(N_B-2\right)+8N_B\left(N_B-1\right)+N_B(N_B-1)^3+8N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{4}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$24N_B\left(N_B-1\right)\left(N_B-2\right)+2N_B(N_B-1)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{4}\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\left(\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\right)^2$	$16N_B(N_B-1)^2+8N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{4}\left(\mathrm{E}\left\{\lvert a_l\rvert^8\right\}\right)^2$	$N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{3}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^3\mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$4N_B(N_B-1)^2\left(N_B-2\right)+36N_B\left(N_B-1\right)\left(N_B-2\right)+24N_B(N_B-1)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{3}\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\mathrm{E}\left\{\lvert a_n\rvert^8\right\}$	$8N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{2}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^5$	$2N_B(N_B-1)^2\left(N_B-2\right)+28N_B\left(N_B-1\right)\left(N_B-2\right)+8N_B(N_B-1)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{2}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^3\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$2N_B\left(N_B-1\right)\left(N_B-2\right)+8N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^{2}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\left(\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\right)^2$	$4N_B(N_B-1)^2+4N_B\left(N_B-1\right)$
$\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^4\mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$4N_B(N_B-1)^2+8N_B\left(N_B-1\right)$
$\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\mathrm{E}\left\{\lvert a_n\rvert^8\right\}$	$4N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^4\right\}\right)^6$	$N_B\left(N_B-1\right)^2+4N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^4\right\}\right)^4\mathrm{E}\left\{\lvert a_l\rvert^8\right\}$	$2N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^4\right\}\right)^2\left(\mathrm{E}\left\{\lvert a_l\rvert^8\right\}\right)^2$	N_B

we can derive

$$\lim_{N_B\to\infty}\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\}=0. \tag{64}$$

Applying the Markov's inequality, we have

$$\Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2<1\right\}=1-\Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\geq 1\right\}\geq 1-\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\}. \tag{65}$$

As $\|\mathbf{B}\|_{\mathrm{F}}^2=\sum_{i=0}^{N_U-1}|\lambda_i|^2$, we have

$$\Pr\left\{|\lambda_0|<1\right\}>\Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2<1\right\}\geq 1-\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\}. \tag{66}$$

Therefore, we have

$$\lim_{N_B\to\infty}\Pr\left\{|\lambda_0|<1\right\}=1. \tag{67}$$

This complete the proof to ***Theorem 1***.

C. Derivation of Inequation (59)

For $u\in\mathbb{U}_1, v\neq u$, from Equation (22), we have

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}=\mathrm{E}\left\{\frac{\left|\mathbf{W}_{(u,v)}\right|^2}{\left|\mathbf{W}_{(u,u)}\right|^2}\right\}\leq\sqrt{\mathrm{E}\left\{\left|\mathbf{W}_{(u,v)}\right|^4\right\}\cdot\mathrm{E}\left\{\left|\mathbf{W}_{(u,u)}\right|^{-4}\right\}}, \tag{68}$$

where the Cauchy-Schwarz inequality is applied [19]. From ***Lemma 3*** and ***Lemma 2***, we have

$$\mathrm{E}\left\{\left|\mathbf{W}_{u,v}\right|^4\right\} = A_1, \tag{69}$$

$$\mathrm{E}\left\{\left|\mathbf{W}_{(u,u)}\right|^{-4}\right\} = \frac{1}{B_1}. \tag{70}$$

Therefore, inequation (59) is established.

D. Derivation of Inequation (60)

For $u \in \mathbb{U}_2$, $v = u-1$, from Equation (22), we have

$$\mathbf{B}_{(u,u-1)} = \frac{\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,u-1)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}}.$$

Applying the Cauchy-Schwarz inequality, we have

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,u-1)}\right|^2\right\} \le \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,u-1)}\right|^4\right\}} \cdot \sqrt{\mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right)^{-1}\right|^4\right\}} \tag{71}$$

According to ***Lemma 4*** and ***Lemma 2***, we have

$$\mathrm{E}\left\{\left|\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,u-1)}\right|^4\right\} = A_2, \tag{72}$$

$$\begin{aligned}\mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right)^{-1}\right|^4\right\} &= \mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\right)^{-1}\right|^4\right\} \\ &\quad \cdot \mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\right)^{-1}\right|^4\right\} \\ &= \frac{1}{B_1^2}.\end{aligned} \tag{73}$$

For $u \in \mathbb{U}_2$, $v = u+1$, following the similar process, we have the same result above. Therefore, we complete the derivation of the inequation (60).

E. Derivation of Inequation (61)

For $u \in \mathbb{U}_2$, $v \notin \{u-1, u, u+1\}$, from Equation (22), we have $\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}$ given by Equation (74), as shown at the top of the next page, where the Cauchy-Schwarz inequality is applied. Next, we have the first expectation in the right hand side of the inequality (74) given by (75), as shown at the top the next page, where

$$\begin{aligned}A &= \left|\mathbf{G}_{(u+1,u+1)}\right|^2\left|\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\right|^2, \\ B &= \left|\mathbf{G}_{(u-1,u-1)}\right|^2\left|\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)}\right|^2, \\ C &= \left|\mathbf{G}_{(u-1,u-1)}\right|^2\left|\mathbf{G}_{(u+1,u+1)}\right|^2\left|\mathbf{G}_{(u,v)}\right|^2, \\ D &= 2\,\mathrm{Re}\left(\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,u+1)}\mathbf{G}^*_{(u+1,v)}\right) \\ &\quad \cdot \mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u-1,u-1)}, \\ E &= -2\,\mathrm{Re}\left(\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,v)}\right) \\ &\quad \cdot \left|\mathbf{G}_{(u+1,u+1)}\right|^2\mathbf{G}_{(u-1,u-1)}, \\ F &= -2\,\mathrm{Re}\left(\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)}\mathbf{G}^*_{(u,v)}\right) \\ &\quad \cdot \left|\mathbf{G}_{(u-1,u-1)}\right|^2\mathbf{G}_{(u+1,u+1)}.\end{aligned}$$

The inequality (75) holds by noting that

$$\begin{aligned}&(A+B+C+D+E+F)^2 \\ &\quad \le 6\left(A^2+B^2+C^2+D^2+E^2+F^2\right),\end{aligned}$$

where A, B, C, D, E, F are both real numbers. Next, we derive the expectations as follows individually.

With the results in ***Lemma 2*** and ***Lemma 4***, we have $\mathrm{E}\left(A^2\right) = \mathrm{E}\left(B^2\right)$ given by

$$\mathrm{E}\left(A^2\right) = \mathrm{E}\left(B^2\right) = A_2A_3. \tag{76}$$

$\mathrm{E}\left(C^2\right)$ is given by

$$\mathrm{E}\left\{C^2\right\} = A_1{A_3}^2. \tag{77}$$

where the results in ***Lemma 2*** and ***Lemma 3*** are applied.

By using $(\mathrm{Re}\,(a))^2 \le |a|^2$, we derive the result of $\mathrm{E}\left\{D^2\right\}$, given by (78), as shown at the top of the next page, where A_4 is obtained through ***Lemma 5***.

Applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathrm{E}\left\{E^2\right\} &\le 4\mathrm{E}\left\{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,v)}\right|^2\right\} \\ &\quad \cdot \mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\right|^4\right\} \\ &\le 4\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\right|^4\right\}\sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\right|^4\right\}} \\ &\quad \cdot \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,v)}\right|^4\right\}}\end{aligned} \tag{79}$$

With the results in ***Lemma 2***, ***Lemma 3***, and ***Lemma 4***, we derive the result of $\mathrm{E}\left\{E^2\right\} = \mathrm{E}\left\{F^2\right\}$, given by

$$\mathrm{E}\left\{E^2\right\} = \mathrm{E}\left\{F^2\right\} \le 4A_3\sqrt{A_1A_2A_3}. \tag{80}$$

Therefore, we derive

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} \le \sqrt{\frac{12A_2A_3 + 6A_1A_3^2 + 24A_4 + 48\sqrt{A_1A_2A_3^3}}{B_1^3}} \tag{81}$$

F. Derivation of Inequation (62)

For $u \in \mathbb{U}_2$, $v = u$, from Equation (22), we have $\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}$ given by (82), as shown at the top of the next page, where the Cauchy-Schwarz inequality is applied. By using $|a+b|^2 \le 2\left(|a|^2+|b|^2\right)$, we have

$$\begin{aligned}&\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2 + \mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^2 \\ &\le 2\left(\left|\mathbf{G}_{(u+1,u+1)}\right|^2\left|\mathbf{G}_{(u,u-1)}\right|^4 + \left|\mathbf{G}_{(u-1,u-1)}\right|^2\left|\mathbf{G}_{(u,u+1)}\right|^4\right), \\ &\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2 + \mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^4 \\ &\le 8\left(\left|\mathbf{G}_{(u+1,u+1)}\right|^4\left|\mathbf{G}_{(u,u-1)}\right|^8 + \left|\mathbf{G}_{(u-1,u-1)}\right|^4\left|\mathbf{G}_{(u,u+1)}\right|^8\right).\end{aligned} \tag{83}$$

Therefore, we derive

$$\begin{aligned}&\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2 + \mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^4\right\} \\ &\quad \le 8\mathrm{E}\left(\left|\mathbf{G}_{(u+1,u+1)}\right|^4\right)\mathrm{E}\left(\left|\mathbf{G}_{(u,u-1)}\right|^8\right) \\ &\qquad + 8\mathrm{E}\left(\left|\mathbf{G}_{(u-1,u-1)}\right|^4\right)\mathrm{E}\left(\left|\mathbf{G}_{(u,u+1)}\right|^8\right).\end{aligned} \tag{84}$$

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} = \mathrm{E}\left\{\frac{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)} + \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)} - \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,v)}\right|^2}{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right|^2}\right\}$$
$$\leq \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)} + \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)} - \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,v)}\right|^4\right\}} \cdot \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right|^{-4}\right\}} \tag{74}$$

$$\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)} + \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)} - \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,v)}\right|^4\right\}$$
$$= \mathrm{E}\left\{(A+B+C+D+E+F)^2\right\} \leq 6\mathrm{E}\left\{A^2+B^2+C^2+D^2+E^2+F^2\right\} \tag{75}$$

$$\mathrm{E}\left\{D^2\right\} \leq 4\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,u+1)}\mathbf{G}^*_{(u+1,v)}\right|^2\right\} = 4A_4 \tag{78}$$

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,u)}\right|^2\right\} = \mathrm{E}\left\{\frac{\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2 + \mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^2}{\left|\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\right|^2}\right\}$$
$$\leq \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2 + \mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^4\right\} \cdot \mathrm{E}\left\{\left|\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\right|^{-4}\right\}} \tag{82}$$

With the results in ***Lemma 2*** and ***3***, we have

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,u)}\right|^2\right\} \leq \sqrt{\frac{16A_3A_5}{B_1^3}}. \tag{85}$$

REFERENCES

[1] 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.

[2] 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.

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

[4] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[6] 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.

[7] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[8] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 4th Quart., 2015.

[9] 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.

[10] K. V. Vardhan, S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "A low-complexity detector for large MIMO systems and multicarrier CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 473–485, Apr. 2008.

[11] S. Mohammed, A. Zaki, A. Chockalingam, and B. S. Rajan, "High-rate space-time coded large-MIMO systems: Low-complexity detection and channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 958–974, Dec. 2009.

[12] T. Datta, N. A. Kumar, A. Chockalingam, and B. S. Rajan, "A novel monte-Carlo-sampling-based receiver for large-scale uplink multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3019–3038, Sep. 2013.

[13] L. Bai, T. Li, J. Liu, Q. Yu, and J. Choi, "Large-scale MIMO detection using MCMC approach with blockwise sampling," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3697–3707, Sep. 2016.

[14] 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.

[15] 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.

[16] Y. Cai, R. C. de Lamare, B. Champagne, B. Qin, and M. Zhao, "Adaptive reduced-rank receive processing based on minimum symbol-error-rate criterion for large-scale multiple-antenna systems," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4185–4201, Nov. 2015.

[17] R. C. de Lamare, "Adaptive and iterative multi-branch MMSE decision feedback detection algorithms for multi-antenna systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5294–5308, Oct. 2013.

[18] L. Arevalo, R. C. de Lamare, M. Haardt, and R. Sampaio-Neto. (Nov. 17, 2016). "Decoupled signal detection for the uplink of large-scale mimo systems in heterogeneous networks." [Online]. Available: https://arxiv.org/abs/1611.05597

[19] 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.

[20] 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)*, London, U.K., Jun. 2015, pp. 1763–1769.

[21] 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.

[22] F. Wang, C. Zhang, X. Liang, Z. Wu, S. Xu, and X. You, "Efficient iterative soft detection based on polynomial approximation for massive MIMO," in *Proc. IEEE Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Oct. 2015, pp. 1–5.

[23] X. Gao, L. Dai, Y. Ma, and Z. Wang, "Low-complexity near-optimal signal detection for uplink large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 18, pp. 1326–1328, 2014.

[24] L. Dai *et al.*, "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.

[25] 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.

[26] B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, "Conjugate gradient-based soft-output detection and precoding in massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 3696–3701.

[27] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.

[28] M. L. Honig and J. S. Goldstein, "Adaptive reduced-rank interference suppression based on the multistage Wiener filter," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 986–994, Jun. 2002.

[29] R. C. de Lamare and R. Sampaio-Neto, "Reduced-rank space–time adaptive interference suppression with joint iterative least squares algorithms for spread-spectrum systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1217–1228, Mar. 2010.

[30] Y. Cai and R. C. de Lamare, "Adaptive linear minimum BER reduced-rank interference suppression algorithms based on joint and iterative optimization of filters," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 633–636, Apr. 2013.

[31] J. M. Burgerscentrum, "Iterative solutions methods," *Appl. Numer. Math.*, vol. 51, no. 4, pp. 437–450, 2011.

[32] F. Jiang, C. Li, and Z. Gong, "A low complexity soft-output data detection scheme based on Jacobi method for massive MIMO uplink transmission," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–5.

[33] H. Lu, "Stair matrices and their generalizations with applications to iterative methods I: A generalization of the successive overrelaxation method," *SIAM J. Numer. Anal.*, vol. 37, no. 1, pp. 1–17, 1999.

[34] C. Li, F. Jiang, C. Meng, and Z. Gong, "A new turbo equalizer conditioned on estimated channel for MIMO MMSE receiver," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 957–960, Apr. 2017.

[35] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[36] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb. 2016.

[37] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Breda, The Netherlands: Now Publishers, 2004.

[38] G. W. Stewart, *Matrix Algorithms: Basic Decompositions*, vol. 1. Philadelphia, PA, USA: SIAM, 1998.

[39] H.-B. Li, T.-Z. Huang, Y. Zhang, X.-P. Liu, and T.-X. Gu, "Chebyshev-type methods and preconditioning techniques," *Appl. Math. Comput.*, vol. 218, no. 2, pp. 260–270, 2011.

[40] F. Jiang and C. Li, "Soft input soft output MMSE-SQRD based turbo equalization for MIMO-OFDM systems under imperfect channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6, doi: 10.1109/GLOCOM.2015.7417541.

[41] *Evolved Universal Terrestrial Radio Access (E-UTRA): Physical Channels and Modulation*. document 3GPP TS 36.211, 2010. [Online]. Available: http://www.3gpp.org

[42] J. Wu and Y. R. Zheng, "Low complexity soft-input soft-output block decision feedback equalization," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 281–289, Feb. 2008, doi: 10.1109/JSAC.2008.080205.

Fan Jiang received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2010, and the M.E. degree in communication and information systems from Southeast University, Nanjing, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Memorial University of Newfoundland. His research interests include localization of wireless sensor networks, signal processing for wireless and underwater acoustic communication systems. He was a recipient of the Best Paper Award at the IEEE Global Telecommunications Conference, Singapore, in 2017.

Cheng Li (M'04–SM'07) received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1992 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from the Memorial University of Newfoundland, St. John's, NL, Canada, in 2004. He is currently a Full Professor with the Department of Electrical and Computer Engineering, Faculty of Engineering and Applied Science, Memorial University of Newfoundland. His research interests include mobile ad hoc and wireless sensor networks, wireless communications and mobile computing, switching and routing, and broadband communication networks. He is a registered Professional Engineer (P.Eng.) in Canada and a senior member of the IEEE and its communications, computer, ocean engineering, and vehicular technology societies. He is an Editorial Board member of *Wiley Wireless Communications and Mobile Computing*, *China Communications*, an Associate Editor of *Wiley Security and Communication Networks*, and an Editorial Board member of the *Journal of Networks*, and *KSII Transactions on Internet and Information Systems*. He was a recipient of the Best Paper Award at the IEEE International Conference on Communications, Cape Town, in 2010, and the IEEE Global Telecommunications Conference (Globecom), Singapore, in 2017. He has served as a General Co-Chair for WINCOM 2017, and he has served as a Technical Program Committee Co-Chair for the QBSC 2010, IEEE WiMob 2011, MSWIM 2013, ACM MSWIM 2014, and WiCON 2017. He has also served as a Co-Chair for various technical symposia or tracks of many international conferences, including the IEEE GLOBECOM, ICC, WCNC, VTC, and IWCMC. He is currently the Elected Chair of the IEEE Communications Society Ad Hoc and Sensor Networks Technical Committee.

Zijun Gong received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Memorial University of Newfoundland, St. John's, NL, Canada. His research interests include WLAN fingerprint-based indoor localization systems, radio propagation modeling, localization of WSN, and localization of underwater vehicles. He was involved in some research on channel estimation in massive MIMO, and millimeter wave communications. He was a recipient of the Best Paper Award at the IEEE Global Telecommunications Conference, Singapore, in 2017.

Ruoyu Su received the B.Eng. and M.Eng. degrees from the Nanjing University of Aeronautics and Astronautics, China, in 2006 and 2009, respectively, and the Ph.D. degree in electrical and computer engineering from Memorial University of Newfoundland, St. John's, Canada, in 2015. He held a post-doctoral position in electrical and computer engineering at the Memorial University of Newfoundland in 2016. His research interests include localization of sensor networks, energy-efficient design for underwater acoustic sensor networks, and cross-layer design for wireless sensor networks.