A Low-Complexity Data Detection Algorithm for Massive MIMO Systems

IMRAN A. KHOSO, XIAOMING DAI, M. NAUMAN IRSHAD, ALI KHAN, AND XIYUAN WANG, (Member, IEEE)
School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
Corresponding author: Xiaoming Dai (daixiaoming@ustb.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871029, in part by the Beijing Natural Science Foundation under Grant L172049, and in part by the Beijing Municipal Commission of Science and Technology under Grant D171100006317004 and Grant Z181100003218008.

ABSTRACT Achieving high spectral efficiency in realistic massive multiple-input multiple-output (M-MIMO) systems entail a significant increase in implementation complexity, especially with respect to data detection. Linear minimum mean-squared error (LMMSE) can achieve near-optimal performance but involves computationally expensive large-scale matrix inversions. This paper proposes a novel computationally efficient data detection algorithm based on the modified Richardson method. We first propose an antenna-dependent approach for the robust initialization of the Richardson method. It is shown that the proposed initializer outperforms the existing initialization schemes by a large margin. Then, the Chebyshev acceleration technique is proposed to overcome the sensitivity of the Richardson method to relaxation parameter while simultaneously enhancing its convergence rate. We demonstrate that the proposed algorithm mitigates multiuser interference and offers significant performance gains via the iterative cancellation of the bias term by prior estimation. Hence, each step of the iteration routine gives a new and better estimate of the solution. An asymptotic expression for the average convergence rate is also derived in this paper. The numerical results show that the proposed algorithm outperforms the existing methods and achieves near-LMMSE performance with a significantly reduced computational complexity.

INDEX TERMS Data detection, massive multiple-input multiple-output (MIMO), linear minimum mean square error (LMMSE), low-complexity, relaxation parameter, Chebyshev acceleration, Richardson iteration.

I. INTRODUCTION
Massive multiple-input multiple-output (M-MIMO), which employs the infrastructure base station (BS) with hundreds or thousands of antennas, offers significant improvements in link reliability and spectral efficiency compared to existing (small-scale) MIMO systems [1]–[7]. It is shown mathematically in [8] that as the number of receive antennas approaches infinity, the effect of non-correlated noise and fast fading is eliminated. In addition, M-MIMO helps to reduce power consumption at the BS and simplifies the signal processing operations. Therefore, M-MIMO is expected to be a core technology for the next generation of wireless communication systems. The theoretically predicted gains of the M-MIMO rely on proper multiuser signal separation at the receiver. Maximum likelihood (ML) is an optimum data detection method [9], [10]. Nevertheless, its complexity scales exponentially with the number of users which is not suitable for M-MIMO systems [11]. Therefore, sub-optimal detectors with reduced complexity are necessary. To obtain near-optimal performance, the fixed-complexity sphere decoding [12] and K-best [13] algorithms have been proposed for MIMO systems. However, the complexity of these algorithms scales unfavorably with the number of antennas which makes them impractical for M-MIMO systems in the low signal-to-noise ratio (SNR) regime [14]. Linear data detectors such as linear minimum mean squared error (LMMSE) and zero-forcing (ZF) [15]–[17] can achieve near-optimal bit error rate (BER) performance for M-MIMO systems with a large base station (BS)-to-user-antenna ratio (BUAR) [14]. But, both LMMSE and ZF detection methods involve complicated matrix inversion operations for
large antenna arrays. Therefore, novel low-complexity near-optimal data detection methods are of great interest [18]–[20].

By exploiting the channel hardening property of M-MIMO channel, the Neumann series expansion (NSE) based approximation approach [21] is proposed to convert the matrix inversion into a series of matrix-vector multiplications to alleviate the implementation burden of the matrix inversion. The performance of NSE method approaches that of the LMMSE for M-MIMO system with a large base station (BS)-to-user-antenna ratio (BUAR). However, the performance of NSE degrades significantly if BUAR is not large enough, e.g., less than four. Moreover, its complexity is even higher than the exact matrix inversion algorithm for more than three expansion terms. To reduce the computational complexity by replacing the NSE, several other low-complexity approximate data detection methods [22]–[36] based on classical iterative methods have been proposed. However, all these methods exhibit degraded performance when the number of users increases and requires a larger number of iterations to achieve better performance with even higher complexity than the exact matrix inversion (see the analysis in Section-III). Therefore, to successfully deploy M-MIMO in practical wireless systems, new and efficient data detection algorithms that achieve low error rate at low-complexity in realistic antenna scenarios are highly desirable.

A. CONTRIBUTIONS
We propose a novel computationally efficient iterative detection technique for M-MIMO systems that achieves a good performance-complexity tradeoff in practical antenna scenarios. In the proposed method, the high-dimensional matrix inversion problem is implemented by an improved Richardson iteration [37]. The channel matrix is asymptotically orthogonal for uplink M-MIMO systems, which enables us to design an efficient initializer using the number of BS and user antennas that significantly speed up the convergence rate. We consider the Chebyshev acceleration method to alleviate the sensitivity of conventional Richardson detector to relaxation parameter. The Chebyshev polynomials and iterative parameters are computed using the minimum and maximum eigenvalues. Given the difficulty of computing the eigenvalues, we approximate these eigenvalues using the number of BS and user antennas at virtually no loss in terms of error-rate performance. Hence, to guarantee the performance of the proposed algorithm in practical wireless systems all aforementioned parameters including initial solution are computed approximately using the number of antennas that further reduces the complexity. Our main contributions can be summarized as follows.

• We propose a novel low-complexity efficient data detection algorithm based on improved Richardson iteration method, which enables near-LMMSE error rate performance for uplink M-MIMO systems.
• We introduce an antenna-dependent approximate initial method to make the proposed algorithm more attractive for practical implementations and show that it leads to the fastest convergence among existing counterparts.

• We show that by considering the Chebyshev acceleration method the proposed algorithm overcomes the sensitivity of conventional Richardson detector to relaxation parameter while remarkably accelerating the convergence rate.
• An asymptotic expression for the average convergence rate is also derived for the proposed algorithm.
• We analyze the computational complexity of different methods and demonstrate the performance differences with numerical simulations.

Our results illustrate that the proposed algorithm achieves superior error rate performance compared to state-of-the-art methods [21]–[25] and achieves near-LMMSE performance with a significantly reduced computational complexity.

B. PAPER OUTLINE
The rest of this paper is organized as follows. Section II introduces the M-MIMO system model. Section III reviews the prior work. Section IV introduces the proposed low-complexity iterative method. Symbol error rate (SER) performance and computational complexity are presented in section V. Finally, concluding remarks are drawn in section VI.

C. NOTATION
Boldface capital letters and lowercase letters represent matrices and vectors, respectively. For a matrix A, we denote its inverse and Hermitian transpose by \( A^{-1} \) and \( A^H \), respectively. \( I_{NT} \) denotes the \( NT \times NT \) identity matrix, and \( |.| \) denotes the absolute operator.

II. SYSTEM MODEL
We consider an uncoded M-MIMO system, where the BS is equipped with \( NR \) antennas and serves \( NT \) single-antenna users simultaneously. Let \( \mathbf{s} \) denotes the \( NT \times 1 \) transmitted signal vector containing the transmitted data symbols from all users. \( \mathbf{H} \in \mathbb{C}^{NR \times NT} \) denotes the channel matrix, in which each element is independent and identically distributed (i.i.d) with zero mean and unit variance. Then the \( NR \times 1 \) received signal vector \( \mathbf{y} \), using the general input-output relation, can be written as follows:

\[
\mathbf{y} = \mathbf{Hs} + \mathbf{n},
\]

where \( \mathbf{n} \) is the complex additive white Gaussian noise vector with zero mean and variance \( N_0 \) per complex entry.

A. LMMSE DETECTION
The estimate of the transmitted signal \( \hat{\mathbf{s}} \) based on the LMMSE principle is given by

\[
\hat{\mathbf{s}} = \left( \mathbf{H}^H \mathbf{H} + \frac{N_0}{E_s} \mathbf{I}_{NT} \right)^{-1} \mathbf{H}^H \mathbf{y} = \mathbf{A}^{-1} \hat{\mathbf{y}},
\]
where $\hat{y} = H^T y$ can be interpreted as output of the matched filter, and the LMMSE filtering matrix $A$ denoted by

$$A = H^T H + \frac{N_0}{E_s} I_{N_T} = G + \frac{N_0}{E_s} I_{N_T},$$

(3)

where $G = H^T H$ is gram matrix and $E_s$ is average energy of the symbol.

Owing to the fact that $N_R \gg N_T$ for M-MIMO systems, the column vectors of channel matrix $H$ are asymptotically orthogonal. Therefore, $H$ is a well-conditioned matrix. Based on this property, the LMMSE detection technique is able to attain near-optimal error rate performance [14].

Based on this property, the LMMSE detection technique has been proposed to reduce the computing load of matrix inversion. Matrix inversion based on the NSE method [21] can be expressed as follows:

$$A_{\text{iter}}^{-1} = \sum_{m=0}^{N_{\text{iter}}-1} \left(-D^{-1}(L + U)^m\right) D^{-1},$$

(4)

where $N_{\text{iter}}$ denotes the terms of NSE, $D$, $L$, and $U$ denote the diagonal, the lower, and the upper triangular components of matrix $A$, respectively. The NSE achieves better detection accuracy. However, when $N_{\text{iter}} > 3$, complexity is even more than that of the LMMSE exact inversion method.

Joint steepest descent and Jacobi (JSDJ) based detection method has been proposed [22] to improve the convergence rate. The first Jacobi iteration using the hybrid iteration can be rewritten as follows:

$$s^{(2)} = s^{(1)} + D^{-1} \left(\hat{y} - As^{(1)}\right) = s^{(1)} + D^{-1} r^{(1)},$$

(5)

where $r^{(1)} = \hat{y} - As^{(1)}$. Suppose that the result of the steepest decent algorithm is $s^{(1)} = s^{(0)} + ur^{(0)}$, then

$$r^{(1)} = \hat{y} - A \left(s^{(0)} + ur^{(0)}\right) = r^{(0)} - up^{(0)},$$

(6)

where $s^{(0)}$ is initial solution and $u = \frac{r^{(0)} H r^{(0)}}{\lambda_{\text{max}}(A) \lambda_{\text{min}}(A)}$. Substituting (6) and $s^{(1)} = s^{(0)} + ur^{(0)}$ into (5), steepest decent and first Jacobi iteration can be merged into the following hybrid iteration

$$s^{(2)} = s^{(0)} + ur^{(0)} + D^{-1} \left(r^{(0)} - up^{(0)}\right).$$

(7)

The JSDJ method has a fast convergence rate, but in higher-order antenna scenarios, its performance becomes poor.

Estimate of the transmitted signal based on the Chebyshev iteration [23] method is given by

$$s^{(i)} = s^{(i-1)} + \mu^{(i-1)},$$

(8)

where the superscript $i$ denotes the iteration number and $\mu^{(i-1)}$ is given as follows:

$$\mu^{(i)} = \delta^{(i)} r^{(i)} + \eta^{(i)} \mu^{(i-1)},$$

(9)

where $\mu^{(0)} = \frac{1}{\beta} r^{(0)}$, the residual $r^{(i)} = \hat{y} - As^{(i)}$, and the parameters $\delta^{(i)}$ and $\eta^{(i)}$ are computed as:

$$\delta^{(i)} = 2 \frac{\alpha}{\beta} \frac{T^{(i)}(\alpha)}{T^{(i+1)}(\alpha)}; \quad \eta^{(i)} = \frac{T^{(i-1)}(\alpha)}{T^{(i+1)}(\alpha)},$$

(10)

where $T^{(i)}$ are the Chebyshev polynomials (will be defined in IV-C section) and $\alpha$ and $\beta$ are iterative parameters whose values are calculated as follows:

$$\alpha = \lambda_{\text{max}} + \lambda_{\text{min}}, \quad \beta = \frac{\lambda_{\text{max}} + \lambda_{\text{min}}}{2},$$

(11)

where $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ are the smallest and largest eigenvalues respectively of matrix $A$. This method reduces the computing load and achieves potential parallelism of matrix inversions at the cost of reduction in detection accuracy.

Matrix inversion based on Newton iteration (NI) and diagonal band Newton iteration (DBNI) [24] is given as follows:

$$s^{(i+1)} = s^{(i)} \left(2I - (D + \hat{E}) s^{(i)}\right).$$

(12)

where $\hat{E}$ denotes the bandwidth, which states that only $\hat{E}$ elements adjacent with the diagonal element in a row for both right and left side are considered. Although DBNI has good precision but its complexity increases monotonically as $\hat{E}$ increases, and the main drawback of DBNI method is that for more number of iterations it reaches high error floor.

Richardson iteration (RI) [25]–[27] achieves the LMMSE data detection in an iterative manner as follows:

$$s^{(i)} = s^{(i-1)} + w \left(\hat{y} - As^{(i-1)}\right).$$

(13)

where $w$ denotes the relaxation parameter. Although, its complexity is low but it is overly sensitive to the relaxation parameter. The conventional RI detectors [25]–[27] use a constant value as relaxation parameter which is suitable for a certain antenna configuration. Furthermore, to achieve good performance a large number of iterations is required, which substantially increases the computation complexity. Therefore, conventional RI detectors are less attractive for practical implementations.

In addition, Gauss-Seidal (GS) based approximate LMMSE detection methods have been proposed to avoid large-scale matrix inversion [28]–[30]. The proposed GS based detectors enable low error rate at low-complexity. However, the internal iterations within each GS iteration make it hard for parallel implementation. Moreover, the conjugate gradient (CG) method [31] also suffers from significant correlation issues, because it computes communication-intensive inner products to determine the recurrence coefficients, thus resulting in low parallelism. Recently, low-complexity ZF detector based on Newton-Schultz iterative (NSI) method [32] has been developed. The NSI algorithm achieves good detection accuracy and strong numerical stability. Unfortunately, the complexity of NSI is cubic per iteration, which is not feasible for M-MIMO systems. The iterative detector based on accelerated overrelaxation (AOR) method [33] reduces the complexity. However, it is overly
sensitive to the acceleration and relaxation parameters. Iterative refinement methods such as detection and precoding based on iterative refinement [34], NI and DBNI with iterative refinement [35] and refinement Jacobi [36] have been used to achieve high precision for M-MIMO detection. The extra refinement steps increase the cost of each iteration, thereby resulting in significantly increased complexity.

All these methods exploit approximate approaches to realize complicated matrix inversion and multiplication, though, ongoing challenges still exist. Therefore, we need an efficient detection scheme to achieve better performance with feasible computational complexity.

IV. PROPOSED DATA DETECTION ALGORITHM FOR M-MIMO SYSTEMS
We now develop a novel data detection algorithm for M-MIMO system. We start by discussing the low-complexity approximate-LMMSE algorithm based on Richardson iteration method. Next, an antenna-dependent approximate approach is presented for a quick start of the iterative detection algorithm, and Chebyshev acceleration method is proposed to overcome the sensitivity of Richardson method to relaxation parameter. Finally, an asymptotic expression for the average convergence rate is derived.

A. PROPOSED ALGORITHM
Since computation of $A^{-1}$ using direct methods is prohibitive for M-MIMO systems due to large matrix dimensions, we focus on iterative methods in order to arrive at the computationally efficient algorithm. Classical iterative methods can be utilized to solve large linear systems of equations that is the main requirement for LMMSE data detection in M-MIMO systems (2). The LMMSE filtering matrix $A$ is symmetric positive definite (SPD) for uplink M-MIMO systems, this special property motivates us to exploit the Richardson iterative technique for data detection [38]. It could potentially be applied to achieve low error rate at low-complexity for M-MIMO data detection system as described in (13), i.e., $s^{(i)} = s^{(i-1)} + w(\hat{y} - A s^{(i-1)})$.

However, many unsolved problems persist, which place constraints on the practical implementation of the Richardson detector. First, it is hard to determine a perfect initial solution for the Richardson method. Second, it is overly sensitive to the relaxation parameter. Third, the conventional Richardson detector is feasible for a particular antenna scenario while in realistic M-MIMO systems antenna configuration varies. Our goal is to solve these problems while simultaneously accelerating the convergence, hence, enabling practical implementation of the proposed algorithm.

B. ANTENNA-DEPENDENT BASED APPROXIMATE INITIAL METHOD
The good initialization $s^{(0)}$, which is close to the final LMMSE estimate $\hat{s}$, would impact the convergence rate greatly and affect both complexity and accuracy of the final solution. In particular, for complexity constrained practical implementations. Since traditional zero-vector based initial solution is generally far away from the final solution, we propose an antenna-dependent approximate initial method to obtain a closer-to-LMMSE-estimate initialization solution. The channel matrix $H$ is asymptotically orthogonal when $N_R \gg N_T$, which indicates that the matrix $A$ is diagonally dominant, such that when the number of BS and user antennas grow, the eigenvalues of LMMSE filtering matrix $A$ converge to a fixed deterministic distribution known as the Marchenko-Pastur distribution. The $\lambda_{\max}$ and $\lambda_{\min}$ of the matrix $A$ converge to [14]

$$\begin{align*}
\lambda_{\max}(\frac{\varphi}{1 + \varphi} - A) &\to 1 + 2\sqrt{\frac{\varphi}{1 + \varphi}}, \\
\lambda_{\min}(\frac{\varphi}{1 + \varphi} - A) &\to 1 - 2\sqrt{\frac{\varphi}{1 + \varphi}}.
\end{align*}$$

(14)

where $\varphi$ is the ratio of the number of BS to the number of user antennas $\varphi = N_R/N_T$. Therefore, the eigenvalues of $I_N - \frac{\varphi}{1 + \varphi}A = I_N - A/(N_R + N_T)$ lie approximately in the range $[-2\sqrt{\varphi}/(1 + \varphi), 2\sqrt{\varphi}/(1 + \varphi)]$. Motivated by [14], the LMMSE filtering matrix $A$ should satisfy the following equation

$$A_{ij} = \begin{cases} N_R + N_T = \gamma, & i=j, \\ 0, & i \neq j. \end{cases}$$

(15)

Each diagonal entry of the LMMSE filtering matrix $A$ is approximately equal to $\gamma$. Based on this principle, the initial solution $s^{(0)}$ can be approximated as:

$$s^{(0)} = A\gamma = \frac{1}{N_R + N_T}H^Hy = \frac{1}{\gamma}\hat{y}.$$  

(16)

Due to low approximation error, the proposed antenna-dependent based initialization (16) is able to guarantee a remarkable performance improvement over the traditional zero-vector and diagonal initialization. Our simulations in Section V-A validate that this approximation based initial method enables a faster convergence rate while significantly reducing the computational complexity in practical antenna scenarios. Since the LMMSE filtering matrix $A$ is Hermitian positive definite, the proposed iterative method is convergent for any initial solution [38].

C. CHEBYSHEV ACCELERATION METHOD
We now consider Chebyshev acceleration method in order to solve the sensitivity problem of conventional Richardson method to relaxation parameter. The first iteration of the proposed algorithm implemented as follows:

$$s^{(1)} = s^{(0)} + z(\hat{y} - As^{(0)}),$$

(17)

where $s^{(0)}$ is initial solution defined in (16) and $z$ is given by

$$z = \frac{2}{\lambda_{\max} + \lambda_{\min}}.$$ 

(18)
In general, it is difficult to calculate the eigenvalues, we use approximated values \( \hat{\lambda}_{\min}, \hat{\lambda}_{\max} \) as:

\[
\hat{\lambda}_{\min} = N_R \left( 1 - \sqrt{\frac{N_T}{N_R}} \right)^2; \hat{\lambda}_{\max} = N_R \left( 1 + \sqrt{\frac{N_T}{N_R}} \right)^2.
\]  

(19)

It can be observed that, to guarantee the performance of (17) in practice, \( z \) only depends on MIMO system configuration (i.e., the number of BS antennas and number of users). We next use the Chebyshev acceleration to estimate the transmitted signal (i.e., \( i \geq 2 \)) given by

\[
s^{(i)} = \psi^{(i-1)} \left( \mathbf{y} - A s^{(i-1)} \right) + \tau^{(i-1)} \left( s^{(i-2)} - s^{(i-1)} \right) + s^{(i-1)},
\]

(20)

where \( \psi^{(i-1)} \) and \( \tau^{(i-1)} \) are iterative parameters whose values are calculated from the Chebyshev polynomials (will be defined later) and eigenvalues of \( \mathbf{A} \). This observation (i.e., \( s^{(i-1)} - s^{(i-2)} \)) helps to gain some understanding about how the iterative procedure can offer significant performance gains. The partial cancellation of the bias term by the prior solution makes it safe to be more daring and choose bigger values of iterative parameters in subsequent iterations. The new \( s^{(i)} \) estimate is of reduced variance and an even better estimate for the bias cancellation term is available in the next iteration.

The parameters \( \psi^{(i-1)} \) and \( \tau^{(i-1)} \) are given by [37]

\[
\psi^{(i-1)} = \frac{4T^{(i-1)}(\alpha)}{(\lambda_{\max} - \lambda_{\min})T^{(i)}(\alpha)},
\]

(21)

\[
\tau^{(i-1)} = \frac{T^{(i-2)}(\alpha)}{T^{(i)}(\alpha)},
\]

(22)

where \( T_n \) are Chebyshev polynomials and \( \alpha \) is defined in (11). At each iteration, for fixed \( \hat{\lambda}_{\min} \) and \( \hat{\lambda}_{\max} \), one always does as well as possible in reducing the error in the minimax sense. Since the coefficients in (20) (i.e., \( \psi^{(i-1)} \) and \( \tau^{(i-1)} \)) are less than unity, each step of the iteration routine provides a higher level of accuracy so that desired performance is achieved with a few number of iterations. The Chebyshev polynomials are given by

\[
T^{(0)}(\alpha) = 1, \quad T^{(1)}(\alpha) = \alpha, \quad T^{(i)}(\alpha) = 2\alpha T^{(i-1)}(\alpha) - T^{(i-2)}(\alpha), \quad i \geq 2.
\]

(23)

Keeping the past history of two vectors instead of only one, the proposed algorithm overcomes the sensitivity of relaxation parameter. Besides, the proposed approach is applicable for different antenna scenarios and is more robust to higher-order modulation schemes compared to conventional Richardson method (i.e. only suitable for a certain antenna configuration) which has been verified via numerical results in the results section.

The proposed algorithm is summarized in Algorithm 1

### Algorithm 1: Proposed data detection Algorithm for M-MIMO Systems

1. **Input**: \( N_R \times N_T \) channel matrix, \( \mathbf{H} \);  
2. \( N_R \times 1 \) received signal vector at the BS, \( \mathbf{y} \);  
3. Number of iteration, \( N_{\text{iter}} \);  
4. **Output**:  
   5. Detected signal, \( \hat{s} \);  
6. **Initialization**:  
   7. \( \hat{\lambda}_{\min} = N_R \left( 1 - \sqrt{\frac{N_T}{N_R}} \right)^2, \hat{\lambda}_{\max} = N_R \left( 1 + \sqrt{\frac{N_T}{N_R}} \right)^2 \);  
   8. \( \alpha = \frac{\lambda_{\max} + \lambda_{\min}}{2} \);  
   9. \( T^{(0)}(\alpha) = 1, T^{(1)}(\alpha) = \alpha \);  
   10. \( \gamma = N_R + N_T \);  
   11. \( \hat{\mathbf{y}} = \mathbf{H}^H \mathbf{y} \);  
   12. \( s^{(0)} = \frac{1}{\gamma} \hat{s} \);  
   13. \( s^{(1)} = s^{(0)} + z(\mathbf{y} - A s^{(0)}) \);  
14. **Iteration**:  
   15. for \( i = 1, \ldots, N_{\text{iter}} - 1 \) do  
   16. \( T^{(i+1)}(\alpha) = 2\alpha T^{(i)}(\alpha) - T^{(i-1)}(\alpha) \);  
   17. \( \psi^{(i)} = \frac{4T^{(i)}(\alpha)}{(\lambda_{\max} - \lambda_{\min})T^{(i+1)}(\alpha)} \);  
   18. \( \tau^{(i)} = \frac{T^{(i+1)}(\alpha)}{T^{(i)}(\alpha)} \);  
   19. \( s^{(i+1)} = \psi^{(i)}(\mathbf{y} - A s^{(i)}) + \tau^{(i)}(s^{(i)} - s^{(i-1)}) + s^{(i)} \);  
20. **End for**

### D. Asymptotic Expression for the Average Convergence Rate

In this section, it is shown mathematically how the proposed iterative method realizes low-approximation error that approaches zero. In an iterative procedure, it is generally useful to anticipate the required number of iterations to obtain the specified precision. Therefore, an asymptotic expression for the average convergence rate is derived. Approximate error at the \( i \)th stage can be expressed as:

\[
\mathbf{e}^{(i)} = s^{(i)} - \mathbf{s} = e^{(i-1)} - w^{(i)} \mathbf{A} e^{(i-1)} = \left( I - w^{(i)} \mathbf{A} \right) e^{(i-1)} = \sum_{k=1}^{i} \left( I - w^{(k)} \mathbf{A} \right) \mathbf{e}^{(0)},
\]

(24)

so that

\[
\mathbf{e}^{(i)} = P^{(i)}(\mathbf{A}) \mathbf{e}^{(0)},
\]

(25)

where \( P^{(i)}(\mathbf{A}) \) is \( i \)th degree polynomial in the matrix \( \mathbf{A} \). If \( n \) eigenvalues of the matrix \( \mathbf{A} \) are \( \lambda^{(k)} \) with corresponding eigenvectors \( v^{(k)} \), then \( \mathbf{e}^{(0)} \) is represented as:

\[
\mathbf{e}^{(0)} = \sum_{k=1}^{n} a^{(k)} v^{(k)},
\]

(26)
where \( a^{(k)} \) are constants. Since \( P^{(i)}(\lambda^{(k)}) \) are eigenvalues and \( v^{(k)} \) are eigenvectors of \( P^{(i)}(A) \), then for any integer \( i \) we obtain

\[
e^{(i)} = P^{(i)}(A)e^{(0)}
= \sum_{k=1}^{n} a^{(k)} P^{(i)}(A)v^{(k)}
= \sum_{k=1}^{n} a^{(k)} P^{(i)}(\lambda^{(k)})v^{(k)}.
\tag{27}
\]

Assume that the eigenvalues of the matrix \( A \) are positive and can be bounded by \( a \) and \( b \) such that

\[
0 < a < \lambda^{(k)} < b < \infty.
\tag{28}
\]

Instead of making \( |P^{(i)}(\lambda^{(k)})| \) small for \( k = 1, 2, ..., n \) we try to make \( |P^{(i)}(s)| \) small for the whole interval \([a, b]\), that is to minimize \( \max_{a \leq s \leq b} |P^{(i)}(s)| \) under the constraint

\[
P^{(i)}(0) = 1.
\tag{29}
\]

Then, by Markoff theorem [39]

\[
P^{(i)}(s) = \frac{T^{(i)}(\frac{b+a-2s}{b-a})}{T^{(i)}(\frac{b+a}{b-a})},
\tag{30}
\]

where \( T^{(i)}(s) \) is the Chebyshev polynomial of order \( i \) given by

\[
T^{(i)}(\cos\theta) = \cos(i\cos^{-1}(s)).
\]

Now we consider the average rate of convergence as [39]:

\[
R = -\left(\frac{1}{\alpha}\right) \log \Lambda^{(i)},
\tag{31}
\]

where \( \Lambda^{(i)} \) is the maximum of the absolute values of the eigenvalues of the linear transformation associated with the process. As we have already exhibited the linear transformation operating on the error vector in (25). It has also been noted that the eigenvalues are positive and given by

\[
p^{(i)}(\lambda^{k}), k = 1, 2, ..., n.
\tag{32}
\]

Therefore, from (30)

\[
|P^{(i)}(\lambda^{(k)})| \leq \max_{s \in [a, b]} |P^{(i)}(s)| = \frac{1}{T^{(i)}(\alpha)}.
\tag{33}
\]

Since \( T^{(i)}((b+a-2s)/(b-a)) \leq 1 \) with equality for \( s = a \). Moreover, if \( a \) is the smallest eigenvalue of the matrix \( A \), then

\[
p^{(i)}(a) = \frac{1}{T^{(i)}(\alpha)}.
\tag{34}
\]

so that in fact \( 1/T^{(i)}(a) \) is the generally desired \( \Lambda^{(i)} \) for the transformation \( P^{(i)}(A) \).

Since \( \alpha > 1 \), again by [39]

\[
T^{(i)}(\alpha) = \frac{(\alpha + \sqrt{\alpha^2-1})^i + (\alpha - \sqrt{\alpha^2-1})^{-i}}{2}.
\tag{35}
\]

The asymptotic value of (35) as \( i \to \infty \) is

\[
T^{(i)}(\alpha) \sim \frac{(\alpha + \sqrt{\alpha^2-1})^i}{2}.
\tag{36}
\]

So, we obtain an asymptotic expression for the average convergence rate as follows:

\[
R \sim \log \left(\frac{\alpha + \sqrt{\alpha^2-1}}{\alpha - \sqrt{\alpha^2-1}}\right).
\tag{37}
\]

V. NUMERICAL RESULTS

We now provide simulation results to demonstrate the validity of the proposed detector and compare it with recently introduced M-MIMO data detectors [21]–[25]. The simulations are conducted using Matlab for various antenna scenarios. We provide the SER performance comparison against different values of SNR using higher-order (i.e. 16-QAM and 64-QAM) modulation schemes. We consider i.i.d. flat Rayleigh fading channel model, and average the SER over \( 1 \times 10^5 \) Monte-Carlo trials. The impact of imperfect CSI on the error-rate performance is also analyzed for \( N_R \times N_T = 128 \times 16 \) using 16-QAM modulation scheme. The performance of ZF and LMMSE detector is included as a baseline. The computational complexity of the proposed algorithm is compared with different M-MIMO detectors [21]–[25].

A. COMPARISON OF DIFFERENT INITIALIZERS

Fig. 1 shows the performance comparison of the proposed algorithm using an antenna-dependent based initialization scheme and conventional (zero-vector and diagonal) initialization methods for \( 256 \times 64 \) antenna configuration with 16-QAM modulation. It is illustrated in Fig. 1 that the proposed initialization scheme can significantly accelerate the convergence rate compared to the existing counterparts. Note that the proposed method with \( i = 3 \) even outperforms the conventional zero-vector initialization with \( i = 4 \) by approximately 1.5 dB at SER of \( 1 \times 10^{-3} \). Moreover, the proposed

![FIGURE 1. SER performance comparison of the proposed method against SNR using different initial solutions for \( N_R \times N_T = 256 \times 64 \).](image-url)
method with \( i = 3 \) outperforms the conventional diagonal initial solution by approximately 0.45 dB at SER of \( 1 \times 10^{-3} \). This translates to reduced latency, which is appealing for delay stringent applications.

### B. ERROR-RATE PERFORMANCE

We now compare the SER performance of several data detectors for a range of system configurations with higher-order modulation schemes. Fig. 2 shows the uncoded SER curves versus the signal-to-noise ratio (SNR) for various M-MIMO scenarios (i.e., \( \phi = 8, 4 \) and 3) with 16-QAM modulation. For DBNI, the adopted bandwidth is \( \hat{E} = 2 \) for 128 \( \times \) 16 and \( \hat{E} = 3 \) for all other antenna configurations. In Fig. 2a we consider the case of \( N_R = 128 \) with 16 users and compare the performance of the proposed algorithm with recently reported approximate detection algorithms (namely, RI, NSE, DBNI, PCI, and JSDJ). Results show that the performance of all iterative methods improves as the number of iterations increases. We find that, at high SNR values, the performance gap of the RI and NSE remarkably increases compared to the performance of all other methods. Furthermore, it can be seen that the PCI and DBNI methods show performance improvements over RI and NSE. However, the proposed algorithm achieves the best performance closely followed by JSDJ compared to all aforementioned iterative algorithms and attains almost similar performance as that of LMMSE and ZF.

For the 256 \( \times \) 64 system, Fig. 2b shows that the performance of the proposed algorithm is very close to a system with LMMSE exact matrix inversion and ZF. We observe that the RI and NSE perform strictly worse for 256 \( \times \) 64 scenario compared to the system in Fig. 2a. This is due to the fact that these algorithms require higher BUAR value to achieve the desired performance. We see that PCI and JSDJ significantly outperform RI, NSE and DBNI methods, which exhibit high error floor. We further note that PCI shows performance improvement over JSDJ for \( i \geq 4 \), the performance gap is about 0.35 dB for a target SER of \( 10^{-3} \). However, PCI entails a 0.6 dB SNR loss compared to the proposed algorithm for the same SER target. Hence, the proposed algorithm significantly outperforms all iterative detectors in terms of SER with a smaller number of iterations.

Fig. 2c illustrates the SER performance for M-MIMO system with 200 users and 600 BS antennas. We observe almost similar trends as for the system considered in Fig. 2b. Note, however, that the relative performance difference between the proposed algorithm and the existing iterative detectors increases for a growing number of antennas (i.e., smaller BUAR values). Moreover, JSDJ performs strictly worse and exhibits high error floor compared to a system in Fig. 2b. Interestingly, PCI shows performance improvement as the number of iterations grows, however, the proposed algorithm outperforms the PCI for the same iteration count. We note that the proposed detector approaches the near-optimal performance of LMMSE and ZF with a slightly increased number of iterations (i.e., \( i = 7 \)) compared to Fig. 2a and Fig. 2b. Yet, the proposed algorithm still substantially reduces the computational complexity compared to LMMSE, which shows the superiority of the proposed algorithm.

**FIGURE 2.** Uncoded SER with 16-QAM modulation for various M-MIMO detectors as a function of SNR for antenna configurations of (a) 128 \( \times \) 16, (b) 256 \( \times \) 64 and (c) 600 \( \times \) 200.
In Fig. 3, we show the uncoded SER as a function of the SNR for antenna configurations of 160 × 32 and 600 × 150 with 64-QAM modulation. We consider DBNI bandwidth $\hat{E} = 3$ for both cases. Fig. 3a shows that the proposed algorithm provides a significant performance improvement over iterative detectors for a system with $N_T = 32$ users and $N_R = 160$ BS antennas. Fig. 3a demonstrates the following observations: (i) it is seen that the NSE and classical RI perform worse compared to other methods; (ii) the performance of DBNI method deteriorates at high SNR values; (iii) PCI outperforms other LMMSE approximations like RI, NSE, DBNI, and JSDJ in terms of SER; (iv) the proposed algorithm achieves almost similar performance as that of LMMSE and ZF with only 5 iterations.

For a system with $N_T = 150$ users and $N_R = 600$ BS antennas, Fig. 3b shows a similar trend for RI and NSE, i.e., they attain high error floor. Note that the systems in Fig. 2b and Fig. 3b have the same BUAR value. However, we observe that JSDJ detector performs slightly worse for similar BUAR value with 64-QAM modulation. It shows that the BUAR threshold for JSDJ to achieve a desirable performance is higher for 16-QAM than for 64-QAM. Additionally, results show that the proposed algorithm provides approximately 1 dB gain compared to PCI at SER of $10^{-3}$ for $i = 6$. The proposed algorithm approaches the performance of LMMSE and ZF with only a few numbers of iterations in this scenario.

Analytical results are shown in Fig. 4 for the considered $N_R \times N_T = 128 \times 16$ M-MIMO system using 16-QAM modulation for the cases of imperfect and perfect CSI. The SER performance of the proposed algorithm is compared with LMMSE in this scenario. Note that, with imperfect CSI, the performance of both methods (proposed and LMMSE) is degraded with respect to the case when perfect CSI is available. It can be observed that the proposed algorithm achieves near-LMMSE performance with a smaller number of iterations for the cases of imperfect and perfect CSI.

We finally conclude that the proposed data detector has excellent scalability to support hundreds of antennas with higher modulation schemes. Additionally, it may yield satisfactory SER performance if the number of antennas is further increased. Hence, compared to state-of-the-art approaches, we consider the proposed algorithm to be the preferred detector for realistic M-MIMO systems.

### C. COMPLEXITY ANALYSIS

We now compare the computational complexity of the proposed and existing data detectors [21]–[25]. The overall complexity is dominated by multiplications, thus the number of complex-valued multiplications is employed as a roughly estimated complexity. Since all the methods involve calculating the initial matrices such as the matrix $\mathbf{A}$ and the matched filter output $\hat{\mathbf{y}}$, we mainly focus on the approximate computations involved in the later steps of the methods. For the proposed method, the first set of computations originates from antenna-dependent based initial method, which involves one multiplication of $1/p$ with a vector $\hat{\mathbf{y}}$ and one real multiplication...
to compute $\frac{1}{\gamma}$. Hence, the required multiplications to compute initial solution is $N_T + 1$. The second set of calculations originates from (17), which involves one multiplication of $N_T \times N_T$ matrix $A$ with a vector $s^{(0)}$ of size $N_T \times 1$ and one multiplication of $z$ with a $N_T \times 1$ vector $(\hat{y} - As^{(0)})$. The computation of parameter $z$ requires one real multiplication, and two real multiplications in (19) to compute $\psi_i$. The $\lambda_{\min}$, $\lambda_{\max}$ and the parameter $\tau$ are just computed once and stored in memory as a constant. Hence, the number of real multiplications is negligible. Therefore, the computation of $s^{(1)}$ requires $N_T^2 + N_T$ number of multiplications. The third set of computations comes from (20) to estimate the $s^{(i)}$ for $i \geq 2$, which involves one multiplication of $N_T \times N_T$ matrix $A$ with a vector $s^{(i-1)}$ of size $N_T \times 1$, one multiplication of $\psi^{(i)}$ with $(\hat{y} - As^{(i-1)})$, and one multiplication of $\tau^{(i)}$ with $(s^{(i-1)} - s^{(i-2)})$. The final set of calculations is performed to compute $\psi^{(i)}$, $\tau^{(i)}$ and $T^{(i)}(\alpha)$ according to (21)–(23). The computation of $\psi^{(i)}$, $\tau^{(i)}$ and $T^{(i)}(\alpha)$ require three, one and two real multiplications, respectively. The required number of multiplications is $N_T^2 + 2N_T + 5$ per iteration for $i \geq 2$.

The proposed antenna-dependent initializer reduces the computational complexity compared to the diagonal initial solution. However, its complexity is slightly higher than the conventional zero-vector initializer. Unfortunately, conventional zero-vector initial method sacrifice error-rate performance for complexity, which is not desirable for high-performance receiver implementations. The computational cost of the proposed initializer is negligible compared to the performance improvements. Additionally, the initial solution is just computed once and stored in memory as a constant. Hence, we consider $K$ iterations for a fair comparison. It is well known that the computational complexity of the conventional LMMSE is $O(N_T^3)$. However, the complexity of $O(N_T^3)$ is unacceptable for M-MIMO systems as the number of antennas increases significantly. Table 1 compares the computational complexity of various M-MIMO data detectors [21]–[25]. The complexity of NSE and DBNI is $O(N_T^2)$ for $i \geq 2$. Thus, $i = 2$ is considered for NSE and DBNI. It can be concluded that the NSE and DBNI can reduce the complexity from $O(N_T^3)$ to $O(N_T^2)$ when $i \leq 2$. However, usually a large number of $i$ is required to achieve the desired detection accuracy for both methods, which has already been verified via numerical results. Furthermore, the complexity of DBNI monotonically increases as the bandwidth $E$ increases. Therefore, only a marginal reduction in complexity can be achieved by NSE and DBNI. From Table 1, it can be seen that the complexity of JSDF and PCI methods is larger than the proposed algorithm. The complexity of the proposed algorithm and RI is the same for $i = 1$. Moreover, the proposed algorithm and RI have the same coefficient for the highest order term $N_T^2$ when $i \geq 2$. However, the coefficient of $N_T$ in the proposed algorithm is slightly greater than that of the conventional RI. In addition, it needs five more real multiplications. Though, the overall complexity is not much affected by the term $N_T$ and few real multiplications. The computational cost of few extra multiplications of the proposed algorithm is negligible compared to the significant performance improvements. Furthermore, as the proposed antenna-dependent initialization is close to the actual solution, the proposed algorithm can converge fast to the final estimate. Consequently, to achieve a certain estimation accuracy, the required number of iterations becomes smaller, which means the complexity of the proposed method can be reduced further.

**VI. CONCLUSION**

In this paper, we proposed a new detection scheme based on improved Richardson method to implement the LMMSE detection for uplink M-MIMO. We developed an antenna-dependent approach to generate a promising initial solution for a quick start of the proposed detector. We showed, in particular, that the proposed algorithm can efficiently solve a sensitivity problem of the conventional Richardson detector to relaxation parameter via Chebyshev acceleration. We derived an asymptotic expression for the average convergence rate. Numerical results validate the proposed algorithm outperforms the existing iterative detection algorithms for varying BUAR values with high-order modulations. In addition, we have shown that it achieves near-LMMSE performance with significantly reduced computational complexity. Thus, the proposed algorithm enables a realistic uplink M-MIMO detector implementation for modern wireless communication systems. The extension of the proposed algorithm to other signal processing problems involving high-dimensional matrix inversion for emerging communication technologies, such as downlink preceding for M-MIMO systems is straightforward.

**REFERENCES**


**TABLE 1. Computational Complexity.**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richardson iteration</td>
<td>$(N_T^2 + N_T)i$</td>
</tr>
<tr>
<td>Neumann series ($i = 2$)</td>
<td>$12N_T^2 + 4N_T$</td>
</tr>
<tr>
<td>Diagonal band NI ($i = 2$, $E = 2$)</td>
<td>$20N_T^2 + 8N_T - 4$</td>
</tr>
<tr>
<td>Joint steepest decent and Jacobi</td>
<td>$(4N_T^2 - 2N_T)i + 10N_T$</td>
</tr>
<tr>
<td>Parallelizable Chebyshev iteration ($i = 1$)</td>
<td>$(2N_T^2 + N_T + 14)i$</td>
</tr>
<tr>
<td>Proposed ($i \geq 2$)</td>
<td>$(N_T^2 + 2N_T + 5)i$</td>
</tr>
</tbody>
</table>


IMRAN A. KHOSO received the B.S. degree in electrical engineering from the COMSATS Institute of Information Technology, Lahore, Pakistan, in 2013. He is currently pursuing the M.S. degree in information and communication engineering with the School of Computer and Communication Engineering, University of Science and Technology Beijing, China. His research interests include signal processing and wireless communications.

XIAOMING DAI received the M.S. and Ph.D. degrees in electrical engineering and automation from Shanghai Jiao Tong University, Shanghai, China, in 1999 and 2002, respectively. He is currently a Professor with the Department of Communications, University of Science and Technology Beijing. His research activity has led to numerous publications in leading international journals and conference proceedings and to fruitful industrial applications, most notably the pattern division multiple access (PDMA).

M. NAUMAN IRSHAD received the B.Sc. degree in electronics from GC University, Lahore, Pakistan, in 2015. He is currently pursuing the master’s degree in information and communication engineering with the University of Science and Technology Beijing, China. His research interests include next-generation 5G wireless networks, the Internet of Things (IoT), SDN, virtualization of mobile communication, and multiple-input-multiple-output techniques.
ALI KHAN received the M.S. degree in information and communication engineering from the University of Science and Technology Beijing, in 2017, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include wireless networks, flying ad-hoc networks, self-organized networking, and routing in ad-hoc networks.

XIYUAN WANG (S’10–M’13) received the Ph.D. degree in control sciences and engineering from the Department of Automation, Tsinghua University, Beijing, China, in 2013. His research interests include information theory, array signal processing, and signal processing for wireless communications.