Received 13 August 2022, accepted 15 September 2022, date of publication 20 September 2022, date of current version 29 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3208155

METHODS

Improved Iterative Inverse Matrix Approximation Algorithm for Zero Forcing Precoding in Large Antenna Arrays

ARAL ERTUG ZORKUN[1], MIGUEL ALEJANDRO SALAS-NATERA[2], AND RAMON MARTÍNEZ RODRIGUEZ-OSORIO[2]

[1]ETSI Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain
[2]Information Processing and Telecommunications Center (IPTC), Universidad Politécnica de Madrid, 28040 Madrid, Spain

Corresponding author: Aral Ertug Zorkun (ar.zorkun@upm.es)

This work was supported by the Spanish Government, Ministry of Economy, and the National Program of Research, Development, and Innovation under the project New Array Antenna Tech. and digital processing for the FUTURE integrated terrestrial and space-based millimeter-wave radio systems, UPM-InTerSpaCE, under Grant PID2020-112545RB-C51.

ABSTRACT Precoding algorithms are used in massive multiple-input multiple-output (mMIMO) communication systems to ensure effective signal transmission. The zero-forcing (ZF) is one of the most common linear precoding algorithms used in MIMO and mMIMO systems. ZF precoding is complex to implement because it requires direct matrix inversion of the gram matrix. Iterative algorithms have been proposed for approximating matrix inversions. However, iterative algorithms require initial conditions and pre-computations to converge to the optimal transmitted signal vector. This paper proposes a new improved iterative algorithm that guarantees convergence under any circumstances without dependency on any optimized initial parameter or condition. The proposed algorithm is based on a three-step iterative and iterative generalized inverse matrix approximation algorithm. The proposed algorithm was verified using a new correlated channel model that included mutual coupling effects and gain and phase variances caused by radio frequency elements at a base station (BS). The computational complexity of the proposed algorithm was then computed. This study analyzes and compares the bit error rate (BER) performance of the proposed algorithm with that of prominent existing algorithms. Moreover, the sum-rate performance of the proposed algorithm was analyzed. Simulations were performed under both correlated and uncorrelated channel conditions, for comparison and analysis. The simulation results demonstrate that the proposed algorithm outperforms the compared algorithms in terms of the convergence and convergence rates.

INDEX TERMS Adaptive antenna system, approximate matrix inversion, iterative algorithms, massive MIMO, ZF precoding.

I. INTRODUCTION

In wireless communication systems, the demand for high-speed data access or transmission is inevitably growing with the exponentially increasing number of users. Fifth-generation (5G) communication systems are expected to provide an improved communication quality with extended coverage. Massive multiple-input multiple-output (mMIMO) technology is considered a key factor that enables people to use fifth-generation communication systems [1] and thing-to-thing prospective sixth-generation (6G) systems [2]. To simultaneously serve many users at the same time, mMIMO systems consist of a large number of antennas. The densely numbered antenna structure of mMIMO systems provides high transmission rate, spectral efficiency, and power efficiency [3], [4], [5]. However, a large number of antennas have drawbacks that must be carefully considered, such as pilot-signal contamination and interference at the base station (BS) and user equipment (UE) [1], [4], [6]. In mMIMO systems, precoding algorithms are used to

The associate editor coordinating the review of this manuscript and approving it for publication was Parul Garg.

provide seamless signal transmission and interference cancellation [7].

Massive MIMO systems track the instantaneous state of a channel by using pilot signals. In this sense, channel state information (CSI) is used to mitigate channel impairments to optimize signal transmission at the BS through precoding. Precoding algorithms are classified as linear or nonlinear. In terms of capacity, nonlinear precoding algorithms are superior to linear ones [8]. However, their complexities are considerably higher and their hardware implementation is costly. Thus, mMIMO systems generally employ linear precoding algorithms [5]. There are many linear precoding algorithms in the literature, the most prominent of which are the matched filter (MF), zero-forcing (ZF), regularized zero-forcing (RZF), and minimum mean square error (MMSE).

The linear precoding algorithms ZF, RZF, and MMSE provide optimal detection performance; however, these algorithms involve high-dimensional matrix inversion [9], [10]. Fast matrix inversion is required because the channel state changes rapidly, constantly adding or removing users from the system and processing them in a short time. However, the computational complexity of the direct inverse channel correlation matrix increases significantly as the number of antennas and users increases; hence, a fast matrix inversion becomes critical. Inverse matrix approximation algorithms are commonly used in linear precoding to provide simple and accurate implementation of precoding algorithms. Finally, there are two prominent approximation algorithms for computing inverse matrix calculations: truncated number series algorithms based on polynomial expansion, and iterative algorithms. The most widely used truncated number series algorithm is the Neumann [11] series, because it has a very simple hardware implementation. Although the Neuman series is favorable in terms of complexity, its convergence and accuracy are insufficient. On the other hand, the Taylor [12] and Kapteyn series [13] have better convergence and accuracy but at the cost of higher complexity. In addition, rapid matrix inversion updates are required to maintain an accurate seamless transmission when a user is added to or removed from the system [14]. However, the inverse matrix update slows down as the polynomial terms of the truncated number series increase. Moreover, the computation of the optimal coefficients of the matrix polynomial with the smallest possible number of terms places a higher burden on hardware [10]. However, iterative algorithms are less complex, have a faster convergence rate, and have fair accuracy; therefore, they are considered to be more efficient than linear precoding algorithms. In terms of approaches to the problem, iterative algorithms can generally be classified into three categories: approximate matrix inversion algorithms (AMIA), iterative approaches for solving linear equations (IASLE), and iterative algorithms for minimizing the residual norm (IMRN) [9]. To calculate the required approximate matrix inversion, AMIA is derived from a truncated number series, such as Newton-Schulz iteration (NI) and Chebyshev iteration (CI) [15], [16]. The IASLE algorithm approaches the matrix inversion problem by solving the system equation, and the optimal transmitted signal vector is calculated by applying iterative processes to decomposed matrix elements, such as the Gauss-Seidel (GS) algorithm and its derivative successive over-relaxation (SOR) algorithm [7]. Finally, IMRN algorithms focus on the minimization of the residual norm application order to bypass the approximate matrix inversion operations and directly find the transmitted signal vector, such as the conjugate residual (CR) algorithm (further improvement of the conjugate gradient algorithm) [17] and generalized minimal residual (GMRES) algorithm [18]. Moreover, iterative algorithms can be obtained by combining two or more algorithms, such as the joint CI and Neumann series (CI-NS) algorithms and the SOR-based approximate matrix inversion (SOR-AMI) algorithm [15]. Among the three types of iterative algorithms, the AMIA is the most inefficient. As the dimensions of the channel matrix increase, the number of polynomial terms or iterations increases to maintain the accuracy. Hence, the computational complexity of the algorithm increases significantly. It should be noted that the implementation complexity and computational complexity are different. Iterative algorithms have a tradeoff between implementation complexity and convergence, as shown in Fig. 1. However, under inappropriate initial conditions, such as when the channel matrix is nonsymmetric, positive, definitive, and strictly diagonally dominant, many IASLE and IMRN algorithms fail to converge to a proper solution [9].

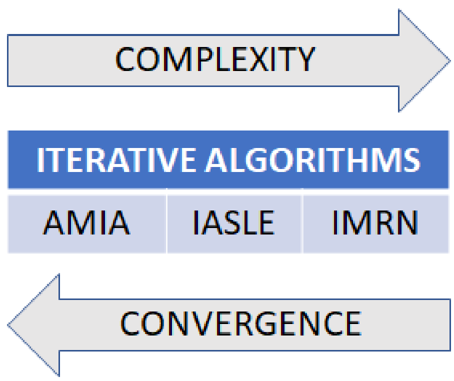


FIGURE 1. Trade-off between implementation complexity and convergence of iterative algorithms.

The three aforementioned types of algorithms require precomputations for the optimal initial parameters to guarantee convergence. The advantages and disadvantages of iterative algorithms can be found in [7] and with brief descriptions.

In general, one algorithm alone is not sufficient to satisfy the convergence, fast convergence rate, and accuracy requirements. However, most algorithms proposed in the literature require initial conditions and an appropriate initial matrix, known as the precondition matrix to guarantee convergence. Hence, in this study, we propose a new improved method based on the combination of Homeier's cubically

TABLE 1. Comparison of iterative Algorithms [7].

Algorithm	Relevant Characteristic	Pros	Cons
NI and CI	-Symmetric matrix -Needs precondition	-If the optimal initial values are adapted convergence is fast and accurate	-Optimal initial values are cumbersome to calculate
GS	-Matrix decomposition is needed	-If the number of transmitter and receiver antennas are equal it gives near-optimal solution	-Unable to implement on parallel computing structure -Includes matrix decomposition
SOR	-Matrix decomposition is needed -Needs precondition	-If the ratio of BS antennas and UE antennas, is large, it gives near optimal solution	-Unable to implement on parallel computing structure -Includes matrix decomposition -Relaxation parameter is uncertain
CR	-Convergence only guaranteed with proper initial vector and high number of iterations	-If the ratio of BS antennas and UE antennas, is large, it gives near optimal solution	-It requires a large number of iterations and initial vector is crucial for convergence
GMRES	-Needs preconditioning to turn linearly dependent vectors into orthonormal vectors -Restart needed	-Suitable for non-symmetric matrices and very robust	-Hard to implement and complexity rises linearly as the number of iterations increase

iterative method [19] and the Karush Kuhn–Tucker (KKT) constraint, [20] resulting in an approach different from Xia's iterative method [21]. Thus, the proposed method converges globally. The proposed method is globally convergent, without the need for any preconditions, can use any square matrix without symmetry, and the initial matrix can be diagonally dominant. Hence, there is no need for any preconditions, and pre-computation is discarded. The proposed algorithm was divided into two parts. The first part computes the initial approximation, whereas the second part processes the iterative generalized inverse matrix approach. The two parts are described as follows:

- The first part of the algorithm was a three-step iterative method based on Homeier's method. However, in the third step, a secant approach was adopted to keep the polynomial order low to avoid a higher computational complexity [22]. After one iteration, the output of the three-step iterative method was passed as an input to the second iterative method. The second method is based on iteratively approximating the generalized inverse using the KKT conditions [21]. If the KKT conditions hold for a problem, optimality is guaranteed [23]. The advantage of this method is that no initial condition is needed for convergence because the Moore-Penrose rules hold for KKT [24]. Thus, nonsymmetric diagonally dominant matrix inverses can be computed more accurately without pre-computation to optimize the initial values.
- In the second part, there is typically a need for direct inversion of the acceleration scheme, as proposed in [21]. Otherwise, the convergence rate degrades as the dimensions of the input matrix increase. Hence, we replaced the direct inversion with a highly accurate iterative estimation method. The need for a proper precondition matrix to guarantee the convergence of the iterative algorithm based on Homeier's method in the literature is discarded here by applying KKT conditions, as in [21]. Thus, the proposed algorithm can be categorized as AMIA-type.

In the ZF precoding algorithm, interference is forced to zero. In this study, ZF precoding was selected because interference, rather than additive noise, is the dominant factor when the number of antennas at the BS increases [15]. The bit error rate (BER) and sum rate of the proposed algorithm were evaluated using both the correlated and uncorrelated channel models. One of the main focuses of this study was to evaluate and analyze the proposed algorithm under realistic conditions; thus, a new correlated channel model was considered. Referring to the channel model, a mutual coupling channel model [3] was used in this study, including array manifolds, which were modeled as in, [25] and extended to mMIMO. Array manifolds include manufacturing tolerances, active radio frequency (RF) element gain, phase variations, and mutual coupling, which are modeled using the k-nearest neighbor approach, as explained in detail in [25]. By contrast, the Rayleigh fading channel model was used for a simpler evaluation of the performance of the proposed algorithm. In addition, some of the most prominent iterative algorithms, such as the CI, SOR, CR, and GMRES algorithms, were investigated for a performance comparison with the proposed algorithm. The main contributions of this study are summarized as follows:

- First, we propose a channel model that produces correlated channels, including mutual coupling, gain, and phase variances, caused by RF component errors and manufacturing tolerances. A model was developed to evaluate the performance of the proposed algorithm under realistic conditions.
- Second, we propose an improved approximate matrix inversion algorithm for the ZF precoding. The proposed algorithm always converges without depending on any initial conditions, and is suitable for correlated channel conditions.

– Finally, the performance of the proposed method was evaluated with respect to different criteria, and its performance was compared with that of various iterative methods in the literature, as summarized in Table 1. Based on these results, the effectiveness of the proposed algorithm was discussed.

The remainder of this paper is organized as follows. In Section II, the system model and the ZF precoding matrix are described. In Section III, the investigated iterative algorithms and the proposed algorithms are described. In Section IV, a computational complexity analysis of the investigated iterative algorithms and proposed algorithms is presented. Simulations to demonstrate the performance of the proposed algorithm are presented in Section V. Finally, Section VI concludes this paper.

Notations: Upper-case and lower-case boldface letters denote the matrices and vectors, respectively. $(.)^T$, $(.)^H$, $(.)^{-1}$, represent the transpose, conjugate transpose (Hermitian), and inversion, respectively. $\mathbf{I}_K$ is an identity matrix of size K.

II. SYSTEM MODEL

This section describes the system model used in this study. We consider a downlink mMIMO system in which M transmit antennas at the BS are employed to serve K single UEs. In our model, an encoder digitally modulates the transmitted signal $\mathbf{s}$, and a precoder weights the information stream $\mathbf{x}$, at the BS. Fig. 2 shows the mMIMO system employed in this study.

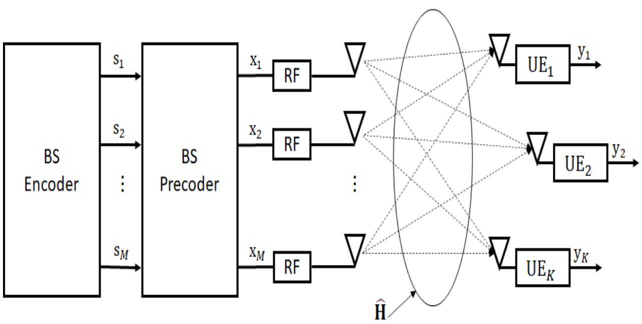


FIGURE 2. Block diagram of mMIMO system model with encoder and precoder.

In this study, we consider a mutual coupling channel model that includes an array manifold to create a more realistic channel model. The estimated imperfect channel between the antenna transmitter array and users is modeled as $\widehat{\mathbf{H}} \in C^{K\times M}$ and can be expressed as

$$\widehat{\mathbf{H}} = \sqrt{1-\tau^2}\widetilde{\mathbf{H}} + \tau\mathbf{v} \tag{1}$$

where $\tau \in [0, 1]$ is a scalar parameter denoting the imperfection of channel estimation. When $\tau = 0$, perfect channel estimation is obtained. The estimated channel noise $\mathbf{v} \in C^{1\times M}$ is independent and identically distributed over the real channel matrix $\widetilde{\mathbf{H}} = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \ldots, \tilde{\mathbf{h}}_K\right]$ which follows a Gaussian distribution with a zero mean and unit variance. The correlation coefficients of the real channel matrix $\widetilde{\mathbf{H}}$ are obtained according to the correlation channel model and can be expressed as

$$\tilde{\mathbf{h}}_k = \boldsymbol{\alpha}\mathbf{A}_M^H, \quad k = 1, 2, \ldots, K \tag{2}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2 \ldots, \alpha_L]$ denotes the complex gains, which are i.i.d. complex Gaussian distributions with zero mean and unit variance, and L denotes the number of propagation paths of the incoming signals to the BS. $\mathbf{A}_M \in C^{M\times L}$ is the array steering matrix of the BS antenna array and is given as follows:

$$\mathbf{A}_M = \left[\boldsymbol{a}\left(\theta_{M,1}, \varphi_{M,1}\right), \boldsymbol{a}\left(\theta_{M,2}, \varphi_{M,2}\right), \ldots, \boldsymbol{a}\left(\theta_{M,L}, \varphi_{M,L}\right)\right] \tag{3}$$

where (θ_M, φ_M) are the L scattered i.i.d. uniformly distributed angles of arrival, with an angular spread of 5°, The BS steering vectors A (θ, φ) can be calculated as follows:

$$\boldsymbol{a}(\theta, \varphi) = \frac{1}{\sqrt{M}} e^{-j\frac{2\pi}{\lambda}(x_m \sin\theta\cos\varphi + y_m \sin\theta\sin\varphi)}, \quad m = 1\ldots M \tag{4}$$

where x_m and y_m are the m-th antenna coordinates in the x-y plane, λ is the wavelength of the carrier frequency, and M is the number of antennas in the antenna array.

To obtain a more realistic channel model, array manifolds including a coupling matrix were included in this study. The array manifold channel model $\mathbf{H} \in C^{K\times M}$ is expressed as follows:

$$\mathbf{H} = \widetilde{\mathbf{H}}\boldsymbol{\Delta}_M \tag{5}$$

where $\Delta_M \in C^{M\times M}$ denotes the BS antenna array manifold, and, which considers the mutual coupling matrix. The array manifold matrix $\Delta(\theta, \varphi)$ is given [25] by:

$$\boldsymbol{\Delta}(\theta, \varphi) = \mathbf{C}\left(\boldsymbol{\Delta}\boldsymbol{G}_{RF}\boldsymbol{\Delta}\boldsymbol{G}(\theta, \varphi)\boldsymbol{\Delta}\boldsymbol{A}(\theta, \varphi)\right) \cdot \boldsymbol{G}_{RF}\left(g(\theta, \varphi)\boldsymbol{a}(\theta, \varphi)\right) \tag{6}$$

where $\boldsymbol{G}_{RF}$, $\Delta\boldsymbol{G}_{RF}$, $\Delta\boldsymbol{A}(\theta, \varphi)$ and $\Delta\boldsymbol{G}(\theta, \varphi)$ are $M \times M$ diagonal matrices with complex elements representing each (θ, φ), is the gain of the RF circuit, and the effect of the gain and phase uncertainty sources owing to the active antenna array components and antenna elements, respectively [25], [26]. In addition, $\boldsymbol{g}(\theta, \varphi) \in C^{M\times 1}$ denotes the amplitude and phase of the m-th element, and $\boldsymbol{a}(\theta, \varphi) \in C^{M\times 1}$ denotes the ideal steering vector of the array containing the information for each (θ, φ) [25]. The mutual coupling matrix $\mathbf{C} \in C^{M\times M}$ consists of the amplitude and phase coupling coefficients $\boldsymbol{C}_{m,k}$, k is the k-th neighbor of the antenna element m. The coefficient of mutual coupling $\mathbf{C}$ can be calculated as follows:.

$$\boldsymbol{C}_{m,k} = 1 + e^{-j\frac{2\pi}{\lambda}(x_m \sin\theta\cos\varphi + y_m \sin\theta\sin\varphi)} \cdot \left(\boldsymbol{a}_0^T \boldsymbol{s}_{(m)}\right) \tag{7}$$

where a_0 is the steering vector that contains all the k-neighboring antenna elements of m-th antenna element, and

$\mathbf{s}_{(m)} = [S_{m1}, \ldots, S_{km}]$ is the $1 \times k$ vector of the scattering coefficients of the m-th antenna element.

The received complex baseband signal $\mathbf{y} \in C^{K\times 1}$ is given as

$$\mathbf{y} = \sqrt{\rho}\,\mathbf{Hx} + \mathbf{n} \tag{8}$$

where ρ is the normalized average transmit symbol energy by the number of transmit antennas, which denotes the signal-to-noise power ratio (SNR); $\mathbf{n} \in C^{1\times K}$ denotes the additive white Gaussian vector; and $\mathbf{x} \in C^{M\times 1}$ represents the transmitted signal vector after precoding and can be expressed as

$$\mathbf{x} = \mathbf{Gs} \tag{9}$$

where $\mathbf{s} \in C^{K\times 1}$ is the symbol vector of the constellation symbols to be transmitted and $\mathbf{G} \in C^{M\times K}$ is the precoding matrix, which can be expressed as

$$\mathbf{G} = \beta \widehat{\mathbf{H}}^H \left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H\right)^{-1} \tag{10}$$

where the scalar β is chosen to satisfy the equation $\|\mathbf{G}\|_F^2 = \operatorname{tr}\left(\mathbf{GG}^H\right) = P$. P is the total transmitted power. The BER and sum rate were considered as criteria for measuring the performance of the precoding algorithms.

After precoding, the sum-rate capacity [5] of the mMIMO system can be calculated as

$$\mathrm{C} = \sum_{k=1}^{K} \log_2 (1 + \gamma_k) \tag{11}$$

where $\gamma_k = \rho/K\, |g_{kk}|^2$, and g_{kk} is the k-th row and k-th column of matrix $\mathbf{G}$.

III. ITERATIVE ALGORITHMS

A. THE APPROXIMATE MATRIX INVERSION ALGORTIHMS

Approximate matrix inversion algorithms are based on number series. These algorithms are derived from the higher-order recursions [27] in (12), which can be expressed as:

$$\mathbf{X}_{i+1} = \mathbf{X}_i \left(\mathbf{I}_\mathrm{K} + (\mathbf{I}_\mathrm{K} - \mathbf{WX}_i) + \cdots + (\mathbf{I}_\mathrm{K} - \mathbf{WX}_i)^{p-1}\right) \tag{12}$$

where i is the number of iterations, p is the order of the polynomial series, $\mathbf{X}_i \in C^{K\times K}$ is the estimated inverse matrix and when $i = 0$, $\mathbf{X}_0$ is defined as the preconditioning matrix when $\mathrm{i} = 0$.

Newton Iteration and Chebyshev Iteration Algorithms: The Newton iteration algorithm converges to the inverse matrix when $p = 2$ [16] and if the inequality condition $\|\mathbf{I}_\mathrm{K} - \mathbf{WX}_0\| < 1$ is satisfied. The estimated inverse matrix at the i-th iteration can be expressed as:

$$\mathbf{X}_{i+1} = \mathbf{X}_i (2\mathbf{I}_\mathrm{K} - \mathbf{WX}_i) \tag{13}$$

However, if $p = 3$ [27] and the same inequality condition for NI is satisfied, this algorithm is called the Chebyshev iteration, and can be expressed as

$$\mathbf{X}_{i+1} = \mathbf{X}_i (3\mathbf{I}_\mathrm{K} - \mathbf{WX}_i (3\mathbf{I}_\mathrm{K} - \mathbf{WX}_i)) \tag{14}$$

The convergence of the approximate matrix inversion algorithms depends on the number of iterations i, the order p and as well as on the preconditioning matrix $\mathbf{X}_0$. A Better convergence can be achieved as the i and p increase with the cost of increased complexity. However, the initial values of the preconditioning matrix also affect the convergence and complexity. In [16], the optimized initial values of $\mathbf{X}_0$ were calculated for both the NI and CI algorithms, and it was shown that both algorithms had better convergence with the optimized values.

B. THE ITERATIVE APPROACHES FOR SOLVING LINEAR EQAUTIONS

Iterative approaches for solving linear equations iteratively solve the linear equation $\mathbf{Wz} = \mathbf{s}$ to approximate a solution, where $\mathbf{z} \in C^{K\times 1}$ is an unknown vector solution. First, the gram matrix $\mathbf{W}$ is decomposed into its subcomponents, and then, an iterative process is applied to the decomposed parts of the $\mathbf{W}$.

Gauss-Seidel Algorithm and Successive Over-Relaxation Algorithm: The GS and SOR algorithms were similar. The Gram matrix is decomposed into $\boldsymbol{W} = \boldsymbol{D} + \boldsymbol{L} + \boldsymbol{U}$, where $\boldsymbol{D} \in C^{K\times K}$ is a diagonal matrix containing the diagonal elements of matrix $\boldsymbol{W}$, $\boldsymbol{L} \in C^{K\times K}$ contains the lower triangular components, and the $\boldsymbol{U} \in C^{K\times K}$ contains the upper triangular component matrix $\boldsymbol{W}$. Subsequently, the estimated transmitted signal vector z is computed iteratively using the decomposed components. The GS algorithm can be expressed as

$$\mathbf{z}_{i+1} = (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{s} - \mathbf{U})\, \mathbf{z}_i \tag{15}$$

The SOR algorithm was derived from the GS algorithm with a slight difference. The iteratively calculated estimated transmitted signal vector z using the SOR algorithm is given as

$$\mathbf{z}_{i+1} = \left(\frac{1}{\omega}\mathbf{D} + \mathbf{L}\right)^{-1} \left(\mathbf{s} + \left(\left(\frac{1}{\omega} - 1\right)\mathbf{D} - \mathbf{U}\right)\mathbf{z}_i\right) \tag{16}$$

where ω denotes the relaxation parameter, which is crucial to the performance of the SOR algorithm. The relaxation parameter must be between $0 < \omega < 2$ to satisfy the convergence [7]. However, in [15], an optimized value for the relaxation parameter is given as

$$\alpha_{opt} = 0.404e^{\left(-0.323\frac{M}{K}\right)} + 1.035 \tag{17}$$

In [15], the authors combined the SOR and approximate matrix inversion (AMI) algorithms and proposed a joint SOR-AMI to increase the convergence of the SOR algorithm. Moreover, the authors combined the CI algorithm with the SOR-AMI and proposed a joint CI-SOR-AMI to increase the convergence rate. In the first step, one iteration is spared for the CI, and then the SOR-AMI is performed. The structure is similar to that of our proposed algorithm; however, the CI-SOR-AMI is highly dependent on the initial conditions and relaxation parameters.

C. THE ITERATIVE ALGORITHMS FOR MINIMIZING RESIDUAL NORM

As mentioned previously, the IMRN algorithms focus on minimizing the residual norm rather than approximating a direct solution [7]. This type of algorithm directly estimates the transmitted signal vector without computing matrix inversion. In IMRN algorithms, the norm of the residual vector $\mathbf{r}_i$ is reduced until the desired tolerance is obtained or the direct solution of $\mathbf{z}$ is obtained.

Conjugate Residual Algorithm and Generalized Minimal Residual Algorithm: The CR algorithm is derived from the well-known conjugate gradient (CG) algorithm to achieve better BER performance than the CG algorithm [7]. The CR algorithm can be expressed as

$$\alpha_i = \frac{\mathbf{r}_i^{\mathrm{T}}\mathbf{W}\mathbf{r}_i}{(\mathbf{W}\mathbf{p}_i)^{\mathrm{T}}\,\mathbf{W}\mathbf{p}_i} \tag{18}$$

$$\mathbf{z}_{i+1} = \mathbf{z}_i + \alpha_k \mathbf{p}_i \tag{19}$$

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \alpha_i \mathbf{W}\mathbf{p}_i \tag{20}$$

$$\beta_i = \frac{\mathbf{r}_{i+1}^{\mathrm{T}}\mathbf{W}\mathbf{r}_{i+1}}{\mathbf{r}_i^{\mathrm{T}}\mathbf{W}\mathbf{r}_i} \tag{21}$$

$$\mathbf{p}_{i+1} = \mathbf{r}_{i+1} + \beta_i \mathbf{p}_i \tag{22}$$

$$\mathbf{W}\mathbf{p}_{i+1} = \mathbf{W}\mathbf{r}_{i+1} + \beta_i \mathbf{W}\mathbf{p}_i \tag{23}$$

where $\mathbf{z}_0$ denotes some arbitrary initial guess, $\mathbf{r}_0 = \mathbf{y} - \mathbf{W}\mathbf{z}_0$ and $\mathbf{p}_0 = \mathbf{r}_0$.

On the other hand, the residual vector $\mathbf{r}_i$ is defined as $\mathbf{r}_i = \mathbf{H}\mathbf{z}_i - \mathbf{y}$ in GMRES algorithm. The solution vector $\mathbf{z}_i \in \tau_i$, τ_i is a set that contains all linearly independent combinations of vectors, is given as

$$\tau_i = \operatorname{span}\left\{\mathbf{y}, \mathbf{H}\mathbf{y}, \ldots, \mathbf{H}^{i-1}\mathbf{y}\right\} \tag{24}$$

In some cases, the vectors $\mathbf{y}, \mathbf{H}\mathbf{y}, \ldots, \mathbf{H}^{i-1}\mathbf{y}$ are not fully linearly independent. Therefore, orthogonalization algorithms, such as Arnoldi and Householder algorithms, are applied to find orthonormal vectors to form an orthonormal basis [28]. The formulation of GMRES can be found in [18].

D. THE PROPOSED IMPROVED ITERATIVE THREE-STEP GENERALIZED INVERSE MATRIX APPROXIMATION ALGORITM

A three-step generalized inverse matrix approximation (TSGIM) algorithm was proposed. First, the system model determined the features required for the proposed algorithm. According to Section II, unequal couplings and different array manifolds distort the conjugate symmetric matrix property of the gram matrix, $\mathbf{W}$. As mentioned previously, when algorithms deal with nonsymmetric positive definitive matrices, convergence cannot be ensured. Among all algorithms, only GMRES offers a robust solution in this case. However, the implementation of the GMRES is extremely difficult and requires many restarts to ensure convergence. The computational complexity increases rapidly as the number of iterations increases. These drawbacks of GMRES make it reluctant to use it, and we propose an approximate matrix inversion algorithm; hence, matrix-vector operations are irrelevant to our case. Convergence, low complexity, and easy implementation are key requirements of approximation algorithms, which are commonly used to fulfill each other's inadequacies.

AMIA has fast convergence, fair precision, and easy implementation but at the cost of increased computational complexity. Thus, in the first step, we selected a fifth-order ($p = 5$) three-step iterative algorithm based on (12). However, the computational complexity of the algorithm increased with the number of orders. In [22], a three-step iterative algorithm was proposed based on two-step cubically iterative Homeier and secant algorithms. The proposed three-step algorithm for solving any function that equals zero ($f(x) = 0$) is given as

$$y_i = x_i - f'(x_i)^{-1} f(x_i) \tag{25}$$

$$z_i = x_i - \frac{1}{2} f(x_i)\left(f'(x_i)^{-1} + f'(y_i)^{-1}\right) \tag{26}$$

$$x_{i+1} = z_i - (f[z_i, x_i])^{-1} f(z_i) \tag{27}$$

where $f[z_i, x_i] = (z_i - x_i)^{-1}(f(z_i) - f(x_i))$ is the two-point divided difference.

To iteratively approximate the matrix inversion, $f(x) = x^{-1} - \mathbf{W}$ was applied to the above equations. The iterative process can then be expressed as:

$$\mathbf{X}_{i+1} = \frac{1}{2}\mathbf{X}_i\,[9\mathbf{I}_{\mathrm{K}} - \mathbf{W}\mathbf{X}_i\,(16\mathbf{I}_{\mathrm{K}} - \mathbf{W}\mathbf{X}_i(14\mathbf{I}_{\mathrm{K}} - \mathbf{W}\mathbf{X}_i(6\mathbf{I}_{\mathrm{K}} - \mathbf{W}\mathbf{X}_i)))] \tag{28}$$

In the second part of the algorithm, we adopt a novel iterative algorithm to compute the generalized inverse matrix in [21]. It applies the KKT condition to minimize the Frobenius norm and iteratively solves the Moore-Penrose generalized inverse conditions [21] with vector-matrix multiplications. The KKT condition was used for the convex optimization of the Frobenius norm. In [21], an acceleration scheme that replaces vector-matrix multiplications with matrix-matrix multiplications was proposed. Unlike the algorithms in, the proposed solution does not require an initial precondition matrix, norm condition, or symmetric or conjugate symmetric matrix, as in the Newton and Chebyshev algorithms. It is important to highlight that their algorithm for the acceleration scheme uses a direct inversion matrix computation. Thus, the inverse matrix approximation problem has not been properly solved. The first part of the proposed algorithm replaces the direct inversion matrix using a robust iterative algorithm.

Computation of the initial values of the initial matrix allows for faster computation and easier implementation, as is the case with the proposed algorithm. However, the inverse matrix approximation algorithm must converge under any circumstances, such as uncertainties in the gain and phase as well as mutual coupling between adjacent antenna elements in a real implementation.

To adopt an iterative algorithm for computing the generalized inverse matrix, significant algorithms in the literature are based on the Moore-Penrose condition, [24] which denotes that for any matrix $\mathbf{A} \in C^{a\times b}$, there exists only one matrix $\mathbf{P} \in C^{b\times a}$ that satisfies the following equations:

$$\begin{cases} \mathbf{A} = \mathbf{APA} \\ \mathbf{P} = \mathbf{PAP} \\ (\mathbf{AP})^T = \mathbf{AP} \\ (\mathbf{XP})^T = \mathbf{XP} \end{cases} \tag{29}$$

where $\mathbf{P}$ is the generalized inverse of matrix $\mathbf{A}$. Subsequently, the generalized inverse problem can be solved by minimizing the Frobenius norm with equality constraints. The optimization problem for $a \geq b$ with solution $\mathbf{P}^*$ can be expressed as:

$$\begin{aligned} &minimize\ \frac{1}{2}\|\mathbf{P}\|_F^2 \\ &subject\ to\ \mathbf{A}^T\mathbf{AP} = \mathbf{A}^T \end{aligned} \tag{30}$$

Applying the KKT condition to (30), the optimum solution $\mathbf{P}^*$ exists if and only if $\mathbf{R}^* \in C^{b\times a}$ exists, and $(\mathbf{P}^*, \mathbf{R}^*)$ satisfies below equations:

$$\begin{aligned} \mathbf{P}^* &= \mathbf{A}^T\mathbf{AR}^* \\ \mathbf{A}^T &= \left(\mathbf{A}^T\mathbf{A}\right)^2\mathbf{R}^* \end{aligned} \tag{31}$$

Subsequently, given matrix $\mathbf{R} \in C^{a\times a}$, as well as a positive scalar ϵ, the iterative algorithm is given as

$$\mathbf{R}_{i+1} = \mathbf{R}_i + \mathbf{M}^{-1}\left(\mathbf{A}^T - \mathbf{B}_1\mathbf{R}_i\right) \tag{32}$$

where $\mathbf{B}_1 = \left(\mathbf{A}^T\mathbf{A}\right)^2$ and $\mathbf{M} = \left(\mathbf{I}_\mathrm{K} + \left(\mathbf{A}^T\mathbf{A}\right)^2\right)$ and stopping criteria is $\left\|\mathbf{A}^T - \mathbf{B}_1\mathbf{R}_i\right\|_F^2 \leq \epsilon$ or when the set maximum number of iterations is reached. Inverse matrix $\mathbf{M}$ must be computed only once because it is not an iterative process. Proof of the global convergence of (32) can be found in [21].

From (32), we obtain the following algorithm. However, because $\mathbf{M}^{-1}$ is a direct inverse matrix, an iterative algorithm cannot be applied. Hence, we first approximated the $\mathbf{M}^{-1}$ using (28), where $\mathbf{M} = \left(\mathbf{I}_\mathrm{K} + \left(\mathbf{W}^T\mathbf{W}\right)^2\right)$. Matrix $\mathbf{A}$ is replaced with a gram matrix because we are interested in approximating the inverse of the gram matrix $\mathbf{W}$. Subsequently, matrix $\mathbf{B}_1$ becomes $\mathbf{B}_1 = \left(\mathbf{W}^T\mathbf{W}\right)^2$, and the proposed algorithm can be expressed as:

$$\mathbf{W}_{i+1}^{-1} = \mathbf{W}_i^{-1} + \mathbf{X}\left(\mathbf{W}^T - \mathbf{B}_1\mathbf{W}_i^{-1}\right) \tag{33}$$

where $\mathbf{X}$ is the estimated inverse matrix of $\mathbf{M}$ by (28).

The proposed improved iterative algorithm is described in detail below, where r is the number of iterations, and $\mathbf{W}^{-1}$ is the estimated inverse matrix as the output of the algorithm.

Algorithm 1 TSGIM Algorithm

Input: $\mathbf{W}$, r.
Output: $\mathbf{W}^{-1}$

```
X = (I_K + (W^T W)^2)
T = WX
Z = 14I− (6I−T) // intermediate step
X = (1/2)X(9I − T (16I − TZ)) // first step
// second step
B = W^T W
W^{-1} = D^{-1}
for i = 1 to r do
    W^{-1} = W^{-1} + X (W^T − B^2 W^{-1})
end for
Output: W^{-1}
```

IV. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section, we evaluate the computational complexity of the proposed algorithm in terms of the number of complex matrix multiplications required.

The computational complexity of ZF precoding with the direct inverse matrix technique is based on the evaluation of matrix $\mathbf{G}$. According to (10), if the pseudo-inverse of the term $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H$ is calculated directly, then the complexity of the ZF precoding technique can be determined using the following steps: We assume that the constant β is known. First, the complexity of matrix multiplication $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H$ contains K^2M complex multiplications. The computational complexity of the direct inversion of the resultant square matrix includes K^3 complex multiplications. The complexity of multiplying the resultant direct inverse matrix by $\widehat{\mathbf{H}}^H$ includes K^2M multiplications. Then, the number of complex multiplications required to calculate the $\mathbf{B}$s is MK where $\mathbf{B}$ denotes $\widehat{\mathbf{H}}^H\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H\right)^{-1}$ in this case. In the final step, the resultant M-by-1 vector is multiplied by scalar β and M complex multiplications. To calculate the transmitted signal vector $\mathbf{x}$ with ZF precoding $K^3+2K^2M+KM+M$ complex multiplications are required.

In our algorithm, the computational complexity in terms of the required number of complex matrix multiplications is calculated as follows: We assume $\mathbf{A}^\mathrm{T}\mathbf{A}$, $6\mathbf{I}_\mathrm{K}$, $9\mathbf{I}_\mathrm{K}$, $14\mathbf{I}_\mathrm{K}$, $16\mathbf{I}_\mathrm{K}$ and the constant β are known. According to (28), $5K^3 + K^2$ complex multiplications are required because the equation includes five matrix-matrix multiplications and one matrix-scalar multiplication. In the second part of the proposed algorithm, because matrix B is assumed to be known, only $2K^3$ complex multiplications are required. In total $7K^3 + 2K^2M + K^2 + KM + M$ complex multiplication was required to calculate the transmitted signal vector $\mathbf{x}$. According to the above analysis, the computational complexities for the first iteration ($i = 1$) of the NI and TSGIM algorithms and the computational complexity of the ZF precoding with DMI are listed in Table 2.

TABLE 2. Computational complexity.

Precoding Algorithm	Computational Complexity
ZF-Direct Matrix Inversion (DMI)	$K^3 + 2K^2M + KM + M$
ZF-NI	$2K^3 + 2K^2M + KM + M$
ZF-TSGIM (Proposed)	$7K^3 + K^2(M+1) + KM + M$

ZF precoding with DMI was chosen as the reference and ZF with the NI algorithm was chosen because it is the simplest algorithm in terms of implementation among the algorithms mentioned in Section III. It is observed that ZF with direct matrix conversion precoding has the lowest computational complexity. However, as previously mentioned, direct inversion is unfavorable for hardware. Although the proposed algorithm has more computational complexity than the NI algorithm, as described in Section III, it converges without depending on any initial conditions. However, convergence is only possible with appropriate initial conditions in inverse-matrix approximation algorithms including the NI algorithm. Therefore, there is a trade-off between robustness and computational complexity for iterative inverse matrix approximation algorithms. Fig. 3 shows the computational complexity in terms of the number of complex multiplications versus the number of BS antennas M, comparing the ZF precoding with the different inverse-matrix approximation algorithms mentioned in this section. In the case of fixing the number of UE antennas to $K = 16$ and increasing only the number of BS antennas, the proposed algorithm has a lower computational complexity. This situation is favorable, because mMIMO systems require hundreds of antenna elements. Fig. 4 depicts the computational complexity in terms of the number of complex multiplications versus the number of UE antennas K, comparing the ZF precoding with different inverse matrix approximation algorithms with the different algorithms mentioned in this section. In this case, our proposed algorithm has a slightly higher computational complexity than that of the NI algorithm. The number of BS antennas is fixed at M = 256, as shown in Fig. 4.

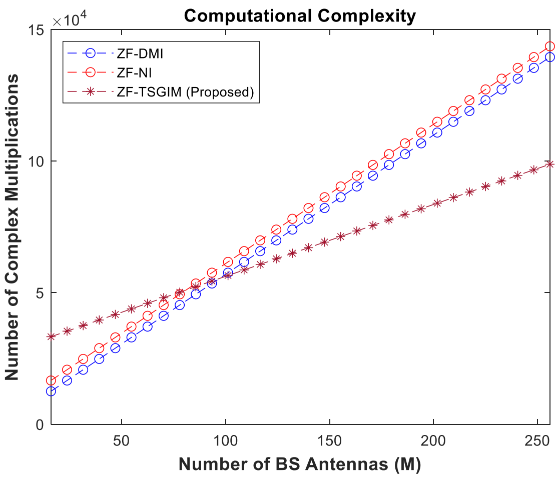


FIGURE 3. Computational complexity ($K = 16, i = 1$).

V. SIMULATION RESULTS

Simulations based on the system model explained in Section II, were conducted to verify the effectiveness of the proposed algorithm. The BER, sum rate, and convergence performance of the proposed TSGIM algorithm are evaluated. The correlated channel model proposed in this study, as described in Section II was used in the simulations. Furthermore, the Rayleigh uncorrelated channel model was used to evaluate algorithms in the literature. The proposed TGSIM

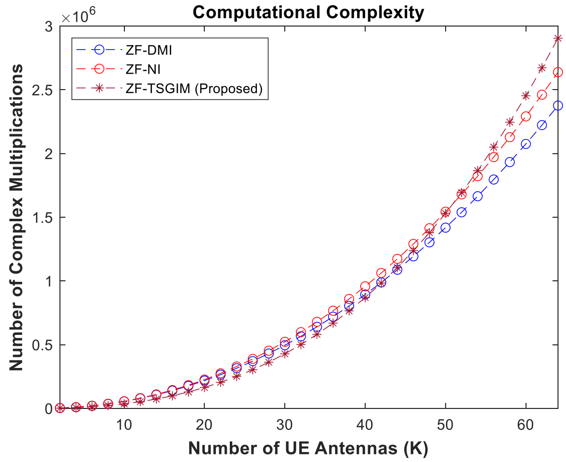


FIGURE 4. Computational complexity ($M = 256, i = 1$).

algorithm was compared to different types of algorithms (AMIA, IASLE, and IMRN). The optimized initial values in [16] were used for the CI algorithm, whereas the optimized relaxation parameter according to (16) was used for the SOR and joint SOR-AMI algorithms. In addition, a noisy random $\mathbf{x}$ vector is used as the initial vector, a diagonal matrix that contains the inverse diagonal elements of the resultant matrix $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H$ is used for the CR and SOR algorithms, and a diagonal matrix that contains the inverse diagonal elements of the resultant matrix $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H$. ZF precoding, which contains direct matrix inversion, was included as a benchmark.

With respect to the use case, a typical downlink massive MIMO configuration with $M \times K = 256 \times 32$ [15], [16] is considered for the BER analysis. For all combinations and analyses, there were $L = 8$ propagation paths. In this study, QPSK was used as the modulation scheme; however, any modulation technique can be used. The transmitted signal was normalized in the BER and sum-rate analyzes. The SNR, denoted as ρ in Section II, is the ratio of transmitted signal power to received noise power for BER and sum-rate analysis, and can simply be expressed as;

$$\rho = \frac{\left\|\widehat{\mathbf{H}}\right\|_F^2}{\sigma_n^2} \tag{34}$$

where $\left\|\widehat{\mathbf{H}}\right\|_F^2$ Frobenius norm of the channel and σ_n^2 is the noise variance.

The parameters for the correlated channel model are as follows: the mutual coupling between adjacent antenna elements is uniformly distributed between −20 dB and −10 dB, as used in [26] for the first tier of neighboring elements; gain and phase variations are ±5% and manufacturing tolerance is ±10% which affects the inter-antenna element spacing, d.

The BER performances of the different algorithms are compared in Fig. 5. The algorithms are simulated using the proposed correlated channel model. The number of iterations for all algorithms was $i = 1$. ZF precoding with direct matrix inversion was used as the benchmark. According to the results, the BER performances of the CI, SOR, CR, and

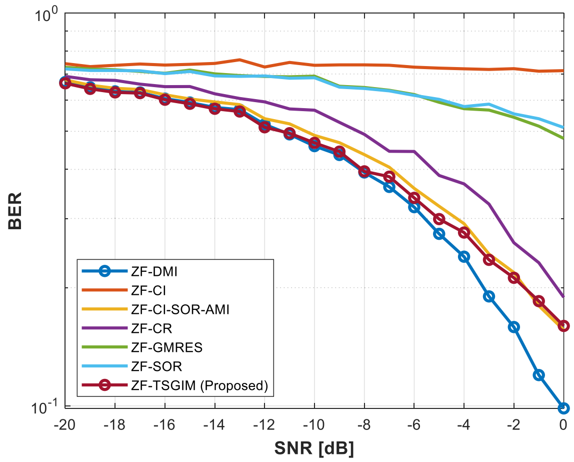


FIGURE 5. BER performance comparison, correlated channel model ($M = 256, K = 32, i = 1$).

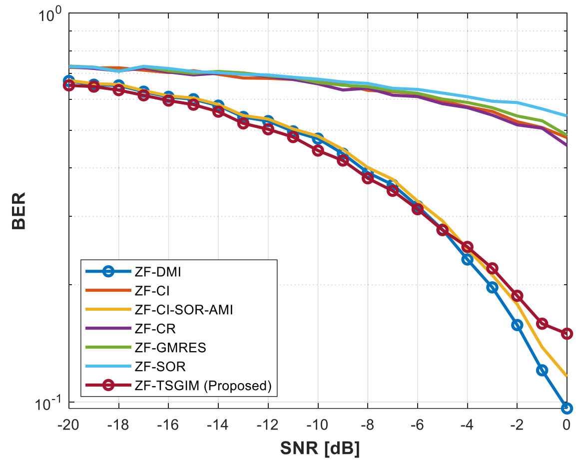


FIGURE 6. BER performance comparison, correlated channel model ($M = 256, K = 32, i = 3$).

GMRES algorithms were insufficient. However, the joint CI-SOR-AMI and proposed TSGIM algorithms performed well. This is because the CI, SOR, and CR are highly dependent on the initial conditions. In this case, the Gram matrix is not a symmetric or conjugate symmetric matrix; hence, the BER performances of CI and SOR are poor. In addition, because we did not consider multiple restarts, the BER performance of GMRES was insufficient. However, it can be seen that the BER performance of the CR algorithm is fair because the random initial vector is appropriate.

As shown in Fig. 6, when the number of iterations was increased to $i = 3$, all algorithms performed better BER performance but the CR algorithm. Although the number of iterations was higher, the BER performance of the CR algorithm degraded without a suitable initial vector. The BER performances of the joint CI-SOR-AMI and the proposed TSGIM algorithms converged to the optimal ZF as the number of iterations increased.

Fig. 7 shows that when the channels are uncorrelated, the Gram matrix becomes conjugate-symmetric; hence, the BER performance of the CI algorithm is satisfactory. However, more iterations are required to achieve a better BER performance. In addition, CI-SOR-AMI has a near-optimal BER performance because the initial parameters and conditions are optimal in this case. The relaxation parameter was optimized according to, [15] and the optimum initial matrix was, as in [16] the CI-SOR-AMI. The initial matrix for the proposed algorithm was a random square matrix. There is no optimization of the initial values for our proposed algorithm; hence, CI-SOR-AMI performs better. Although all the initial parameters of CI-SOR-AMI are optimized, CI-SOR-AMI performs better than the proposed algorithm under uncorrelated channel conditions. The proposed TSGIM algorithm also exhibits good BER performance under uncorrelated channels near the optimum. The SOR and CR algorithms require more iterations and more suitable initial vectors. Finally, the GMRES algorithm requires more restarts and iterations to improve BER performance.

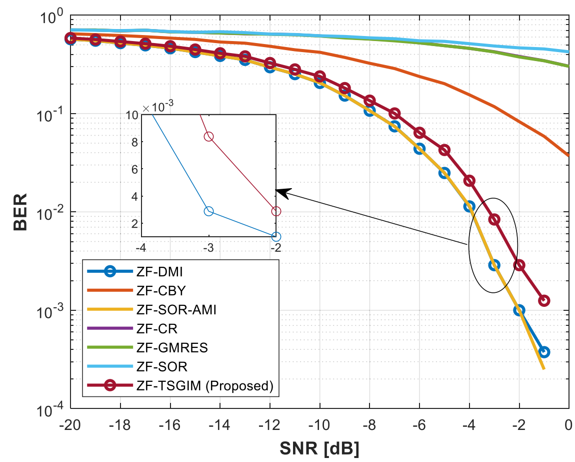


FIGURE 7. BER performance comparison, uncorrelated channel model ($M = 256, K = 32, i = 3$).

Fig. 8 shows that if the channel estimation is not perfect, i.e., CSI is incomplete, then the performance of the proposed algorithm decreases. Performance degradation due to the imperfect channel estimation is valid for all algorithms even if they are not shown in Fig. 8. In reality, perfect channel estimation is impossible but near-perfect approximations can be made. Referring to (1), $\tau = 0$ states that perfect channel estimation and $\tau = 0.3$ states that $\approx 95\%$ of the channel is estimated correctly.

The Frobenius norm errors of the CI, SOR-AMI, and proposed TSGIM algorithms are compared in Fig. 9 and 10 according to the increasing number of BS antennas. In addition, in Fig. 11 the number of BS antennas is fixed and the Frobenius norm errors are compared according to the increasing number of iterations for the algorithms. A total of 1,000 Monte Carlo (MC) trials were conducted under correlated channel conditions. It is worth noting that the BER performances of the other algorithms were not considered in this analysis because they skipped the inverse matrix calculation. In the Frobenius norm error analysis, we consider

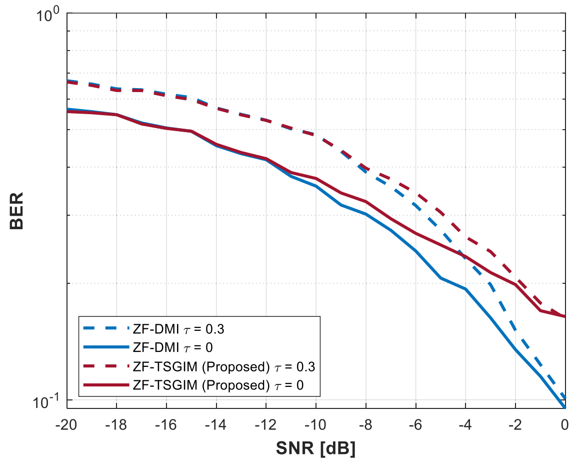


FIGURE 8. **BER performance perfect vs. imperfect channel estimation, correlated channel model ($M = 256, K = 32, i = 3$).**

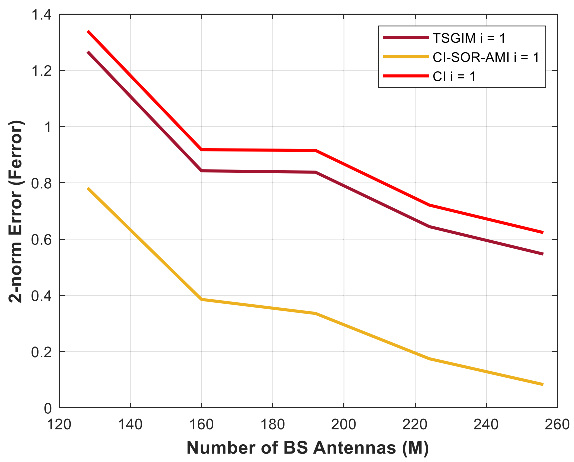


FIGURE 9. **MSE comparison of approximate matrix inversion algorithms, correlated channel condition ($K = 8, i = 1$).**

the mean squared error (MSE) between the ideal inverse of the Gram matrix $\mathbf{W}^{-1}$, and the estimated inverse of the Gram matrix $\widehat{\mathbf{W}}^{-1}$. The Frobenius norm error is $F_{error} = \|\mathbf{W}^{-1} - \widehat{\mathbf{W}}^{-1}\|_F$. In Fig. 9 and 10, for the sake of brevity, only the first iterations, $i = 1$, and fifth iteration, $i = 5$, of the algorithms were considered. As shown in Fig. 9, under the correlated channel condition, the MSE of the proposed CI-SOR-AMI algorithm outperformed those of the CI and TSGIM algorithms in the first iteration. In this analysis, the initial parameters for the CI, CI-SOR-AMI, and TSGIM algorithms were optimized. However, Fig. 10 shows that, after five iterations, TSGIM outperformed CI and CI-SOR-AMI. As shown in Fig. 11, the convergence rate of the proposed TSGIM algorithm outperformed those of CI and CI-SOR-AMI. After four iterations, the Forbenius norm errors of the proposed TSGIM algorithm, CI, and CI-SOR-AMI were 6.37×10^{-4}, 4.09×10^{-2} and 2.67×10^{-2} respectively. Moreover, after ten iterations proposed TSGIM and CI converge to the exact matrix inversion, while CI-SOR-AMI still has the same convergence magnitude.

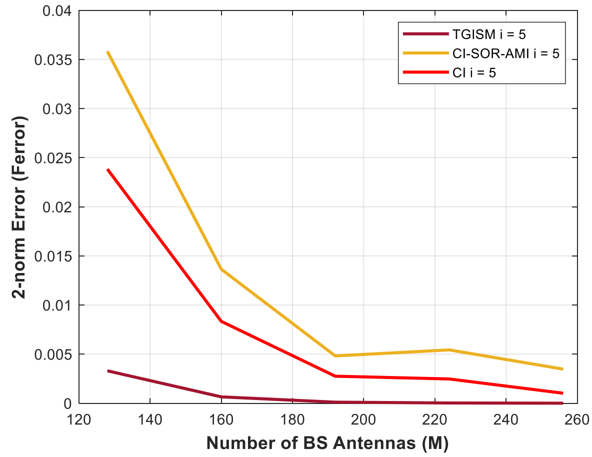


FIGURE 10. **MSE comparison of approximate matrix inversion algorithms, correlated channel condition ($K = 8, i = 5$).**

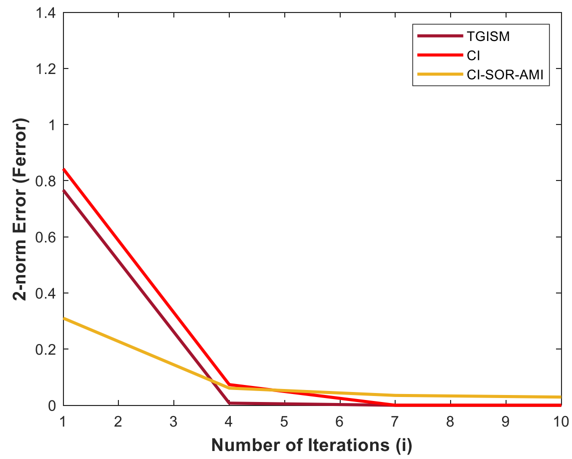


FIGURE 11. **MSE comparison of approximate matrix inversion algorithms, correlated channel condition ($M = 256, K = 8$).**

Fig. 12 compares the sum-rate performance of the ZF precoding with the proposed TGSIM algorithm and the ZF precoding with direct inverse matrix calculation. In this case, ZF precoding with direct inversion was the benchmark. It can be seen that in case of imperfect channel estimation, the

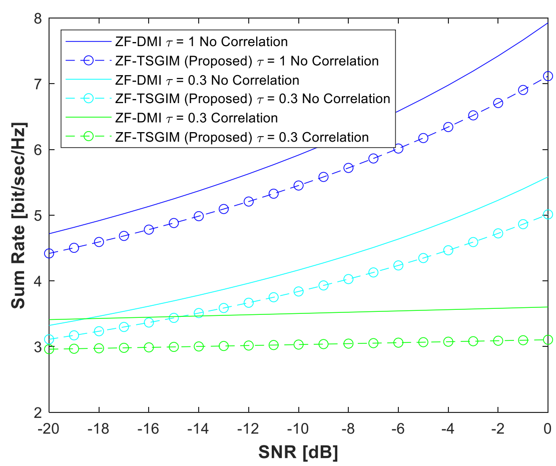


FIGURE 12. **Sum-rate vs SNR ($M = 256, K = 32, i = 1$).**

sum-rate performance of the precoding algorithms decreases. Moreover, if the estimated channel is imperfect and the channels are correlated, including the antenna array manifolds, the sum-rate performance decreases significantly, and although the SNR increases, it becomes almost stationary. However, in each case, the sum-rate performance of the proposed algorithm was close to the optimum.

VI. CONCLUSION

The main focus of this study was to propose a matrix inverse approximation algorithm for precoding transmissions in 5G mMIMO systems. The proposed algorithm outperformed the existing techniques in terms of BER, convergence speed, and Frobenius norm error when the number of BS antennas increased. The algorithm is based on a three-step approximation method using Homeier's approach and an iterative generalized inverse matrix approximation algorithm employing KKT conditions.

To evaluate the performance of the proposed algorithm under realistic mMIMO conditions, we propose a correlated channel model based on antenna array manifolds that includes mutual coupling, RF and radiating element impairments, and multipath channels with an angular spread. We then investigate the correlated channel effects on the ZF precoding algorithm. The proposed method required four iterations to converge to a direct inverse matrix with a Frobenius norm error magnitude of 10^{-4}. However, at least seven iterations are required for the CI algorithm to achieve the same error magnitude, and the CI-SOR-AMI algorithm cannot achieve the same performance even after ten iterations. The proposed TSGIM algorithm is suitable for approximating the inverse of non-symmetric matrices without any preconditions for convergence.

Finally, algorithms such as SOR and CI-SOR-AMI require matrix decomposition and the calculation of the optimum relaxation parameter, which increases the computation time. The compared algorithms, including SOR and CI-SOR-AMI, require pre-calculations to fulfill the pre-conditions to guarantee convergence to the exact matrix inverse. The proposed TSGIM algorithm has better BER performance and less computation time than the compared algorithms. Therefore, the proposed algorithm is feasible for use in 5G and other communications systems.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electron.*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[3] K. Zheng, O. Suling, and Y. Xuefeng, "Massive MIMO channel models: A survey," *Int. J. Antennas Propag.*, vol. 2014, pp. 1–11, 2014.

[4] E. Ali, M. Ismail, R. Nordin, and N. F. Abdulah, "Beamforming techniques for massive MIMO systems in 5G: Overview, classification, and trends for future research," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 6, pp. 753–772, 2017.

[5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2012.

[6] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. D. Porter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.

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

[8] V. L. Babu, L. Mathews, and S. S. Pillai, "Performance analysis of linear and nonlinear precoding in MIMO systems," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 373–376, Jun. 2015.

[9] H. Wu, B. Shen, S. Zhao, and P. Gong, "Low-complexity soft-output signal detection based on improved Kaczmarz iteration algorithm for uplink massive MIMO system," *Sensor*, vol. 20, no. 6, p. 1564, 2020.

[10] A. Mueller, A. Kammoun, E. Bjõrnson, and M. Debbah, "Linear precoding based on polynomial expansion: Reducing complexity in massive MIMO," *EURASIP J. wireless Commun. Netw.*, vol. 2016, no. 1, pp. 1–22, 2016.

[11] S. M. Abbas and T. Chi-Ying, "Approximate matrix inversion for linear pre-coders in massive MIMO," in *Proc. FIP/IEEE Int. Conf. Very Large Scale Integr.-Syst. Chip*. Cham, Switzerland: Springer, 2016, pp. 192–212.

[12] W. Zhang, R. C. de Lamare, C. Pan, M. Chen, B. Wu, and X. Bao, "Correlation-driven optimized Taylor expansion precoding for massive MIMO systems with correlated channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[13] X. Xue, Y. Man, S. Xing, Y. Liu, B. Li, and Q. Wu, "Improved massive MIMO RZF precoding algorithm based on truncated Kapteyn series expansion," *Information*, vol. 10, no. 4, p. 136, 2019.

[14] F. Rosário, F. A. Monteiro, and A. Rodrigues, "Fast matrix inversion updates for massive MIMO detection and precoding," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 75–79, Jan. 2016.

[15] S. Hashima and O. Muta, "Fast matrix inversion methods based on Chebyshev and Newton iterations for zero forcing precoding in massive MIMO systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–12, 2020.

[16] C. Zhang, Z. Li, L. Shen, F. Yan, M. Wu, and X. Wang, "A low-complexity massive MIMO precoding algorithm based on Chebyshev iteration," *IEEE Access*, vol. 5, pp. 22545–22551, 2017.

[17] 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.*, Dec. 2014, pp. 3696–3701.

[18] A. Abdaoui, M. Berbineau, and H. Snoussi, "GMRES interference canceler for doubly iterative MIMO system with a large number of antennas," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2007, pp. 449–453.

[19] H. H. H. Homeier, "On Newton-type methods with cubic convergence," *J. Comput. Appl. Math.*, vol. 176, no. 2, pp. 425–432, 2005.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K. : Cambridge Univ. Press, 2006.

[21] Y. Xia, T. Chen, and J. Shan, "A novel iterative method for computing generalized inverse," *Neural Comput.*, vol. 26, no. 2, pp. 449–465, Feb. 2014.

[22] T. Faezeh and F. Soleymani, "An iterative method for computing the approximate inverse of a square matrix and the Moore–Penrose inverse of a non-square matrix," *Appl. Math. Comput.*, vol. 224, pp. 671–680, Nov. 2013.

[23] J. Brinkhuis and V. Tikhomirov, *Optimization: Insights and Applications*. Princeton, NJ, USA: Princeton Univ. Press, 2005.

[24] K. M. Prasad and R. B. Bapat, "The generalized Moore–Penrose inverse," *Linear Algebra Appl.*, vol. 65, pp. 59–69, Mar. 1992.

[25] M. A. Salas-Natera, R. M. Rodríguez-Osorio, L. D. H. Ariet, and M. Sierra-Pérez, "Novel reception and transmission calibration technique for active antenna array based on phase center estimation," *IEEE Trans. Antennas Propag.*, vol. 65, no. 10, pp. 5511–5522, Oct. 2017.

[26] M. A. Salas-Natera, R. M. Rodriguez-Osorio, and L. de Haro, "Procedure for measurement, characterization, and calibration of active antenna arrays," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 2, pp. 377–391, Feb. 2013.

[27] M. Wu, B. Yin, K. Li, C. Dick, J. R. Cavallaro, and C. Studer, "Implicit vs. Explicit approximate matrix inversion for wideband massive MU-MIMO data detection," *J. Signal Process. Syst.*, vol. 90, no. 10, pp. 1311–1328, Oct. 2018.

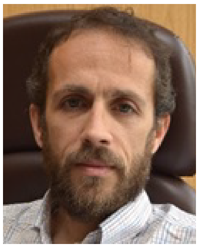
[28] H. F. Walker, "Implementation of the GMRES method using householder transformations," *SIAM J. Sci. Stat. Comput.*, vol. 9, no. 1, pp. 152–163, Jan. 1988.

ARAL ERTUG ZORKUN received the B.S. degree in electrical and electronics engineering from Sabancı University, in 2013, and the M.S. degree in electrical and electronics engineering from the TOBB University of Economics and Technology, in 2018. He is currently pursuing the Ph.D. degree with the Universidad Politécnica de Madrid, Madrid, Spain.

He is also with the Radiation Group, Universidad Politécnica de Madrid.

MIGUEL ALEJANDRO SALAS-NATERA received the master's degree in space technologies, in 2011, and the Ph.D. degree in technologies and communications systems from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2011.

In 2012, he was the Technical Director of Antenna System Solutions Company (spin-off of UPM) for two years. He is currently an Assistant Professor and a Researcher with the UPM Radiation Group. In addition, he has been responsible for a special session on SATCOM since its inception in URSI2016 in Spain. He led several international and local projects about the development of new antenna technologies and the antenna testing, validation, and analysis of SATCOM systems. He works in the areas of uncertainty analysis and calibration methods for antenna arrays, as well as in the design of ultracompact reflector antennas, including novel FSS and manufacturing technologies. He is actively participating in the development of algorithms for resource allocation of flexible transponders for classical GEO satellite systems as well as for massive LEO satellites and high-altitude platform (HAP) systems. He is currently leading two international projects: ''Development of one X/Ka single aperture antenna system'' and ''Development of high polarization purity antenna calibration source on board CubeSats under ESA program.''



RAMON MARTÍNEZ RODRIGUEZ-OSORIO received the Ph.D. degree in electrical engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2004.

In 2002, he joined the Telecommunication Engineering School, UPM (ETSIT-UPM), where he has been a Full Professor, since 2019, and a member of the Information Processing and Telecommunications Center. He has been a Lecturer with SatCom Systems, since 2004. He is currently the Principal Investigator of FUTURE-RADIO, a project to evaluate radio technologies for future communication systems including 5G. In recent years, he has worked on the optimization of high-throughput satellite systems with the use and operation of flexible payloads using machine learning techniques. He has led and participated in various international projects related to space technology and has collaborated with ESA and space communication industries in technology transfer programs. He has led the design of antenna systems for intersatellite links in nanosatellite missions. His research interests include wireless and satellite communications and application of antenna array processing systems.

Prof. Martínez Rodriguez-Osorio received awards for his Ph.D. dissertation, which focused on the application of smart antennas in cellular communications and the UPM Innovative Teaching Award.