

Weighted Gauss-Seidel Precoder for Downlink Massive MIMO Systems

Jun-Yong Jang¹, Won-Seok Lee¹, Jae-Hyun Ro¹, Young-Hawn You² and Hyoung-Kyu Song^{1,*}

¹Department of Information and Communication Engineering, Convergence Engineering for Intelligent Drone, Sejong University, Seoul, 05006, Korea

²Department of Computer Engineering, Sejong University, Seoul, 05006, Korea

*Corresponding Author: Hyoung-Kyu Song. Email: songhk@sejong.ac.kr

Received: 20 November 2020; Accepted: 12 December 2020

Abstract: In this paper, a novel precoding scheme based on the Gauss-Seidel (GS) method is proposed for downlink massive multiple-input multiple-output (MIMO) systems. The GS method iteratively approximates the matrix inversion and reduces the overall complexity of the precoding process. In addition, the GS method shows a fast convergence rate to the Zero-forcing (ZF) method that requires an exact invertible matrix. However, to satisfy demanded error performance and converge to the error performance of the ZF method in the practical condition such as spatially correlated channels, more iterations are necessary for the GS method and increase the overall complexity. For efficient approximation with fewer iterations, this paper proposes a weighted GS (WGS) method to improve the approximation accuracy of the GS method. The optimal weights that accelerate the convergence rate by improved accuracy are computed by the least square (LS) method. After the computation of weights, the different weights are applied for each iteration of the GS method. In addition, an efficient method of weight computation is proposed to reduce the complexity of the LS method. The simulation results show that bit error rate (BER) performance for the proposed scheme with fewer iterations is better than the GS method in spatially correlated channels.

Keywords: Massive MIMO; GS; matrix inversion; complexity; weight

1 Introduction

Massive multiple-input multiple-output (MIMO) for a very large scale of multi-user (MU) MIMO is a significant technology that enables advancement to the fifth-generation (5G) cellular networks. Since hundreds of antennas at the base station (BS) induce significant multiplexing and diversity gain, massive MIMO systems provide high energy efficiency, spectral efficiency, and transmit reliability compared with small scale MU-MIMO systems. Also, simple linear processing such as the ZF method provides optimal performance by the channel hardening property that makes fading channels behave as deterministic channels in massive MIMO systems. It has been proved that the ZF method with accurate channel state information (CSI) can perfectly eliminate



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

intra-cell interference (ICI) in a single-cell system [1–6]. Indeed, there are various non-linear processing methods and those provide higher error performance than linear processing methods [7]. However, the complexity of non-linear processing as an existing problem has become higher in large scale MIMO systems. Therefore, the ZF method as the simple linear processing with optimal performance has been attracted attention for massive MIMO systems [5]. However, the ZF method requires the operation of direct matrix inversion whose computational complexity for multiplication is $O\{N_u^3\}$ where N_u is the number of users. The complexity of direct matrix inversion is unaffordable for the practical implementation of massive MIMO systems which serve tens of users as the number of BS antennas increases [6]. Several precoding and detection schemes with low complexity were surveyed in [8–11]. To balance between the complexity and performance, many studies that approximate the matrix inversion to reduce the complexity were conducted [12–24], and approximate methods can be classified into two categories: Neumann series expansion (NSE) and iterative method.

The NSE method approximates the matrix inversion as expansion terms of a matrix polynomial. The invertible matrix for an initial matrix is required to use the NSE method. The diagonal matrix was used as the initial matrix of the NSE method to achieve the error performance equivalent to the ZF method and low complexity in [12,13]. However, the NSE method has two limited conditions related to its practical application to massive MIMO systems. Firstly, the expansion of the NSE method does not exceed 2 terms since the expansion of 3 terms requires the same complexity of $O\{N_u^3\}$ as direct matrix inversion. Secondly, the NSE method can only achieve high convergence probability in the large ratio $\rho > 5.83$, where the ratio ρ is defined as $\rho = N_b/N_u$ and N_b is the number of BS antennas [13]. Thus, the NSE method is difficult to achieve both low complexity and high approximation accuracy in small ρ .

To obtain better approximation accuracy than the NSE method for the same complexity, the iterative method approximates a vector multiplied by an invertible matrix. Since the iterative method only requires matrix-vector multiplication operation, any iterations maintain the complexity of $O\{N_u^2\}$. The error performance of iterative methods is determined depending on the initial solution and iteration matrix. The GS method requires fewer iterations for demanded error performance compared with other iterative methods such as Jacobi et al. [9].

To improve the error performance of the GS method, the initial solutions were variously defined in [15–19]. The stair matrix was proposed as the initial solution in [15]. The one expansion term of the NSE method that uses the diagonal matrix as the initial matrix was proposed as the initial solution in [16]. Also, the two expansion terms of the NSE method were proposed as the initial solution in [17–19]. In addition to variant initial solutions, the successive over relaxation (SOR) methods that converge faster with the aid of a relaxation parameter compared with the GS method were proposed in [20–22]. The GS and SOR methods can converge to the error performance of the ZF method with fewer iterations under the ideal conditions such as the large ρ . However, the convergence rate and approximation accuracy are reduced under practical conditions such as the small ρ and spatially correlated channels. To solve these problems, the GS methods based on the soft-output detector for uplink systems were proposed in [16–18], and the SOR method applying adaptive relaxation parameters for each iteration was proposed in [22]. Another way to solve these problems applies different weights for each expansion term or iteration to improve the error performance. The NSE methods applying different weights were proposed in [23,24], and the approximate methods with properly computed weights can achieve better error performance. However, References [12–22] did not consider the performance improvement by weighting, and the weights for each iteration were only set to default value

as 1. In addition to weighting the approximate method, the consideration of practical channel environment is significant in terms of performance verification [6]. This paper especially verifies the error performance of iterative methods for spatially correlated channels.

In this paper, a WGS method to perform the ZF precoder is proposed and provides better error performance in spatially correlated channels. The weights that minimize the error between exact and approximate matrix inversion are computed by the LS method. For a feasible LS method, the process of the GS method should be transformed into the form of NSE. The optimal weights are only computed by the channel response of each channel coherence interval. However, the proposed scheme uses one weight vector in the optimal weights since it is more efficient in terms of computational complexity. Also, this paper proposes an efficient method to reduce the complexity for the computation of the weights. The simulation results show that the WGS precoder provides better error performance than the GS precoder.

This paper is organized as follows. Section 2 presents models for the downlink massive MIMO system and spatially correlated channels. Section 3 explains iterative methods, the GS precoder, and the improvement of the GS method. In Section 4, the WGS precoder and the efficient weight computation method are proposed. Also, Section 4 provides the computational complexity for iterative methods. Section 5 presents the simulation results for BER performance comparison of iterative methods. Finally, Section 6 gives brief conclusions.

2 System Model

This section presents a downlink massive MIMO system model and channel model assumed in this paper.

2.1 Downlink Massive MIMO System Model

Fig. 1 shows the system configuration for the downlink massive MIMO where the BS with N_b transmit antennas simultaneously communicates N_u users with a single antenna, and $N_b \gg N_u$ is assumed. The channel matrix \mathbf{H} between all transmit antennas and total users is $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \cdots \mathbf{h}_{N_u}]^T$ whose the m -th row vector is $\mathbf{h}_m = [h_{m1} h_{m2} \cdots h_{mN_b}]$. The received symbol y_m at the m -th user is as follows:

$$y_m = \sqrt{P} \mathbf{h}_m \frac{\mathbf{g}_m}{\|\mathbf{G}\|_F} s_m + \sum_{n=1, n \neq m}^{N_u} \sqrt{P} \mathbf{h}_m \frac{\mathbf{g}_n}{\|\mathbf{G}\|_F} s_n + z_m, \quad (1)$$

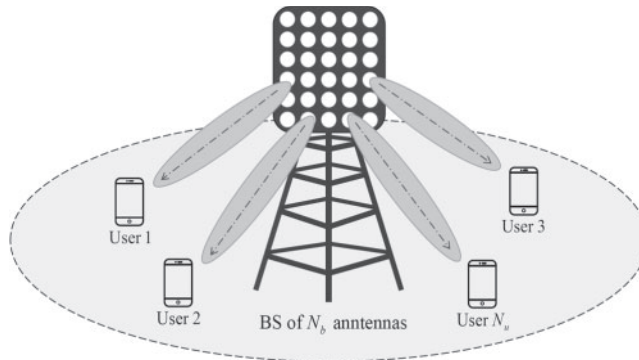


Figure 1: The downlink massive MIMO system configuration

where P is the downlink transmit power, \mathbf{G} is an $N_b \times N_u$ precoding matrix, \mathbf{g}_m is the m -th column of \mathbf{G} , s_m is the m -th complex transmit symbol with zero mean and unit variance, z_m is the m -th additive white Gaussian noise (AWGN) with zero mean and unit variance, and $\|\cdot\|_F$ is Frobenius norm operator.

2.2 Channel Model

For spatially correlated channels, this paper considers the exponential correlation matrix model which is expressed as [25]

$$r_{i,j} = \begin{cases} a^{|i-j|} & \text{if } i \geq j \\ (a^*)^{|i-j|}, & \text{otherwise} \end{cases} \quad (2)$$

where $r_{i,j}$ is the element in the i -th row and j -th column of the spatial correlation matrix \mathbf{R} , $(\cdot)^*$ is the complex conjugate, $a = \zeta e^{j\theta}$ is the correlation coefficient, ζ is the correlation coefficient magnitude with $0 < \zeta \leq 1$, and θ is randomly determined in interval $[0, \frac{\pi}{2}]$. The channel matrix \mathbf{H} is expressed as follows:

$$\mathbf{H} = \mathbf{R}_r^{\frac{1}{2}} \mathbf{H}_w \mathbf{R}_t^{\frac{1}{2}}, \quad (3)$$

where \mathbf{H}_w is an $N_u \times N_b$ Rayleigh fading channel matrix which elements are independent and identically distributed (i.i.d.) circular symmetric complex Gaussian random variables with zero mean and unit variance, \mathbf{R}_r is an $N_u \times N_u$ complex spatial correlation matrix for receive antennas, and \mathbf{R}_t is an $N_b \times N_b$ complex spatial correlation matrix for transmit antennas. The distance between adjacent antennas is related to the correlation coefficient magnitude and the larger distance than half a wavelength can omit the spatial correlation [22]. This paper considers single antenna users and assumes that the distance of users is larger than half a wavelength. Therefore, it is assumed that the magnitude of correlation coefficient of the user has zeros in this paper.

3 Approximate Methods

In this section, iterative methods and the relationship between iterative method and NSE are briefly explained, and the GS precoders are presented.

3.1 Iterative Methods

The iterative method approximates a signal vector \mathbf{x} to solve the linear equation of $\mathbf{W}\mathbf{x} = \mathbf{s}$. The iterative method can be expressed as follows:

$$\mathbf{M}\mathbf{x}_k = \mathbf{N}\mathbf{x}_{k-1} + \mathbf{s}, \quad (4)$$

where $\mathbf{W} = \mathbf{M} - \mathbf{N}$, \mathbf{M} is a nonsingular matrix, subscript k is the iteration number, and \mathbf{x}_0 is an initial solution that generally assumes zero vector. The expression of Eq. (4) with respect to the \mathbf{x}_k is expressed as follows:

$$\mathbf{x}_k = \mathbf{M}^{-1} (\mathbf{N}\mathbf{x}_{k-1} + \mathbf{s}) = \mathbf{B}\mathbf{x}_{k-1} + \mathbf{M}^{-1}\mathbf{s}, \quad (5)$$

where $\mathbf{B} = \mathbf{M}^{-1}\mathbf{N}$ is an iteration matrix. For the convergence of $\lim_{k \rightarrow \infty} \mathbf{x}_k = \mathbf{W}^{-1}\mathbf{s}$, the condition of $\rho(\mathbf{B}) < 1$ should be satisfied, and $\rho(\cdot)$ is the spectral radius of a matrix. In comparison to the iterative method, the NSE method is expressed as follows:

$$\mathbf{W}_k^{-1} = \sum_{n=0}^{k-1} \left(\mathbf{I}_N - \mathbf{M}^{-1}\mathbf{W} \right)^n \mathbf{M}^{-1}, \quad (6)$$

where \mathbf{I}_N is the $N \times N$ identity matrix, and the iterative method can be transformed as the form of NSE as follows:

$$\mathbf{x}_k = \mathbf{M}^{-1}\mathbf{N}\mathbf{x}_{k-1} + \mathbf{M}^{-1}\mathbf{s} = \sum_{n=0}^{k-1} \left(\mathbf{I}_N - \mathbf{M}^{-1}\mathbf{W} \right)^n \mathbf{M}^{-1}\mathbf{s}, \quad (7)$$

where the initial solution \mathbf{x}_0 is zero vector, and the \mathbf{M} is called as preconditioner that determines the error performance of approximate methods.

3.2 GS Precoder

For the expression of the GS precoder, the gram matrix \mathbf{W} is defined as follows:

$$\mathbf{W} = \mathbf{H}\mathbf{H}^H. \quad (8)$$

The preconditioner of the GS method is expressed as follows:

$$\mathbf{M}_{\text{GS}} = \mathbf{D} + \mathbf{L}, \quad (9)$$

where \mathbf{W} is decomposed as $\mathbf{W} = \mathbf{D} + \mathbf{L} + \mathbf{U}$, \mathbf{D} is the diagonal part, \mathbf{L} is the strictly lower part, and \mathbf{U} is the strictly upper part. The k -th GS solution \mathbf{x}_k is calculated as follows:

$$\mathbf{x}_k = (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{s} - \mathbf{U}\mathbf{x}_{k-1}), \quad (10)$$

where $\mathbf{s} = [s_1 s_2 \cdots s_{N_u}]^T$ is a $N_u \times 1$ modulated symbol vector. The downlink transmit symbol vector \mathbf{t} is calculated with matched filter as follows:

$$\mathbf{t} = \frac{1}{\sqrt{\gamma}} \mathbf{H}^H \mathbf{x}_k, \quad (11)$$

where γ is scaling factor to normalize the total transmit power.

3.3 Improvement of GS Method

The initial solution \mathbf{x}_0 is variously defined according to the system requirement, which can improve the convergence rate and approximation accuracy of iterative methods. Among various initial solutions, the recent initial solution for improvement of the GS method is expressed as [18]

$$\mathbf{x}_0 = \sum_{n=0}^1 \mathbf{D}^{-1} \left(-\mathbf{E}\mathbf{D}^{-1} \right)^n \mathbf{s}, \quad (12)$$

where the matrix \mathbf{E} is defined as $\mathbf{E} = \mathbf{L} + \mathbf{U}$. Another way to improve the GS method is the usage of a relaxation parameter. The relaxation parameter α to the GS method to work well at any ρ is applied as [19]

$$\mathbf{x}_k = \left(\mathbf{L} + \frac{1}{\alpha} \mathbf{D} \right)^{-1} \left[\mathbf{s} + \left(\left(\frac{1}{\alpha} - 1 \right) \mathbf{D} - \mathbf{U} \right) \mathbf{x}_{k-1} \right]. \quad (13)$$

Eq. (13) is called as the SOR method, and the GS method is special case of the SOR method whose the value of α is equal to 1. The GS precoder with the initial solution of Eq. (12) and the SOR precoder with the appropriately selected α provides better error performance than the GS precoder with the initial solution of zero vector.

4 Proposed Downlink Precoding Scheme

In Section 2, the spatially correlated channel is considered. The problem is that the spatial correlation degrades the error performance of approximate methods. Although direct elimination of the spatial correlation is difficult, the improvement of error performance can alleviate this problem to some extent. In this section, a WGS method is designed to improve the error performance of the GS precoder for downlink massive MIMO systems. The WGS method is inspired by the weighted Neumann series in [23,24]. Also, an efficient method to reduce the complexity of weight computation is proposed.

4.1 Weighted GS Precoder

The proposed scheme that applies different weights to each iteration of the GS method is expressed as follows:

$$\mathbf{x}_k = (\mathbf{D} + \mathbf{L})^{-1} (u_{k-1} \mathbf{s} - \mathbf{U} \mathbf{x}_{k-1}), \quad (14)$$

where $\{u_n\}_{n=0}^{k-1}$ are complex weights to decrease approximation error, and the initial solution \mathbf{x}_0 is zero vector. For the computation of the optimal weights, the WGS method is expressed similarly with Eq. (6) as follows:

$$\mathbf{x}_k = (\mathbf{D} + \mathbf{L})^{-1} (u_{k-1} \mathbf{s} - \mathbf{U} \mathbf{x}_{k-1}) = \sum_{n=0}^{k-1} u_{k-n-1} \left(\mathbf{I}_K - (\mathbf{D} + \mathbf{L})^{-1} \mathbf{W} \right)^n (\mathbf{D} + \mathbf{L})^{-1} \mathbf{s}, \quad (15)$$

where the weights are applied in reverse order to each term of matrix polynomial, and the weighted \mathbf{W}_k^{-1} of the GS method is expressed as follows:

$$\mathbf{W}_k^{-1} = \sum_{n=0}^{k-1} a_n \left(\mathbf{I}_K - (\mathbf{D} + \mathbf{L})^{-1} \mathbf{W} \right)^n (\mathbf{D} + \mathbf{L})^{-1}, \quad (16)$$

where a_n is the weight of reverse order to simplify the subscript. The weights can be computed by the LS method that determines the solution of the linear model in which the given data is the channel response [26]. Firstly, the optimal weights should minimize the value of the problem as follows:

$$P_{opt} = \left\| \mathbf{W} \mathbf{W}_k^{-1} - \mathbf{I}_{N_u} \right\|_F. \quad (17)$$

For the computation of the optimal weights, the j -th term of the matrix polynomial of the $\mathbf{W}\mathbf{W}_k^{-1}$ except the weight parameter is defined as follows:

$$\mathbf{C}^{(j)} = \mathbf{W} \left((\mathbf{I}_K - \mathbf{P}^{-1} \mathbf{W})^{(j-1)} \right) \mathbf{P}^{-1} = [\mathbf{c}_1^{(j)} \mathbf{c}_2^{(j)} \cdots \mathbf{c}_{N_u}^{(j)}], \quad (18)$$

where $\mathbf{c}_\ell^{(j)}$ is the ℓ -th column vector of $\mathbf{C}^{(j)}$ and the ℓ -th column vector of \mathbf{I}_{N_u} is defined as \mathbf{e}_ℓ . Secondly, the optimization problem in Eq. (17) can be modified with the different form that computes the weights as the solution of linear equation. The modified optimization problem is defined as follows:

$$\mathbf{a}_{opt} = \arg \min_{\mathbf{a} \in \mathbb{C}} P_{opt} = \arg \min_{\mathbf{a} \in \mathbb{C}} \|\mathbf{Z}\mathbf{a} - \mathbf{b}\|_F, \quad (19)$$

where $\mathbf{a} = [a_0 \ a_1 \ \cdots \ a_{k-1}]^T$ is a $k \times 1$ weight vector $\mathbf{b} = [\mathbf{e}_1^T \ \mathbf{e}_2^T \ \cdots \ \mathbf{e}_{N_u}^T]^T$ is the $N_u^2 \times 1$ vector, and \mathbf{Z} is a $N_u^2 \times k$ combined matrix whose the (i,j) -th element is $\mathbf{c}_i^{(j)}$ as follows:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{c}_1^{(1)} & \mathbf{c}_1^{(2)} & \cdots & \mathbf{c}_1^{(k)} \\ \mathbf{c}_2^{(1)} & \mathbf{c}_2^{(2)} & \cdots & \mathbf{c}_2^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_{N_u}^{(1)} & \mathbf{c}_{N_u}^{(2)} & \cdots & \mathbf{c}_{N_u}^{(k)} \end{bmatrix}. \quad (20)$$

Finally, the weight vector computed by the LS method is expressed as follows:

$$\mathbf{a}_{opt} = (\mathbf{Z}^H \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{b}. \quad (21)$$

In this way, the computation of the optimal weights should be performed at every channel coherence interval, and the computational complexity is higher than direct matrix inversion. To reduce the complexity, the WGS precoder uses the only one optimal weight vector computed early in the system. The computation of the weights is executed only once and not included in approximation complexity. Since the optimal weights are determined only within a certain distribution, any weights in the distribution can sufficiently improve the error performance of the GS precoder. Fig. 2 shows the distribution of the optimal weights for 5 expansion terms of matrix polynomial in complex plane. To illustrate the effect of the spatial correlation and the ratio $\rho = N_b/N_u$ on weights, three different cases are presented. In Fig. 2, n is the order of the weights, the N_b is fixed at 200. The case (a) is that N_u is 20 and ζ is 0, the case (b) is that N_u is 40 and ζ is 0, and the case (c) is that N_u is 40 and ζ is 0.3. The number of optimal weights representing the distribution is 200 for each term of the matrix polynomial. The distribution of the case (c) is the largest in Fig. 2, but is enough small since the width of the distribution for real and imaginary values is close to 1. The optimal weights change according to the ρ and ζ since the correlation degree of the channels is different. The weights of the first and second order overlap since those values are close to 1. In contrast, the absolute value of the weights for the latter order increases as the correlation degree of the channels increases and is clearly observed.

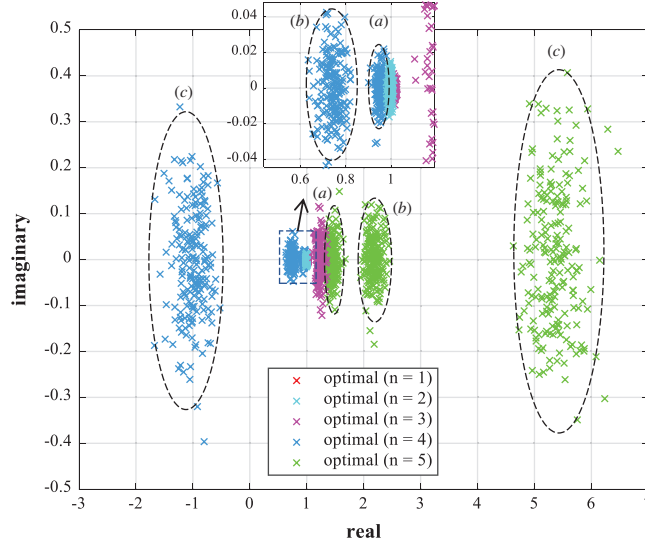


Figure 2: The distribution of optimal weights in three different cases

4.2 Efficient Weight Computation Method

As seen above process, the LS method can efficiently compute the weights. However, the computation method of the optimal weights requires very high computational complexity $\mathcal{O}(N_u^4)$. This paper proposes an efficient method to compute the weights with less complexity. The complexity can be reduced by modifying the \mathbf{Z} and \mathbf{b} in Eq. (21). The modified \mathbf{Z} is defined as the $N_u \times k$ combined matrix \mathbf{X} as follows:

$$\mathbf{X} = \begin{bmatrix} \text{diag}(\mathbf{C}^{(1)}) & \text{diag}(\mathbf{C}^{(2)}) & \cdots & \text{diag}(\mathbf{C}^{(k)}) \end{bmatrix}. \quad (22)$$

where $\text{diag}(\cdot)$ forms the diagonal elements of a matrix to a column vector, and the modified \mathbf{b} is defined as the $k \times 1$ vector \mathbf{d} whose all elements have the value of 1. The efficient weight computation method is expressed similarly with Eq. (21) as follows:

$$\mathbf{a}_{diag} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{d}. \quad (23)$$

Additionally, only the real part of the weights is used to exclude the influence on imaginary values when the weights are determined close to the distribution boundary on the imaginary axis. In terms of the number of users, the weight computation method using only the diagonal elements for each term of the matrix polynomial reduces the complexity to $\mathcal{O}(N_u^2)$. The weights computed by Eq. (23) are close enough to the optimal weights and significantly improve the error performance of the GS method. Tab. 1 expresses the meaning of parameters in the WGS algorithm that the summary of the WGS precoder using the real part of one \mathbf{a}_{diag} is expressed as an algorithm as follows:

Step 1: Input parameter $\mathbf{s}, \mathbf{H}, \mathbf{K}$

Step 2: Initialization

$$\mathbf{W} = \mathbf{H}\mathbf{H}^H = \mathbf{D} + \mathbf{L} + \mathbf{U}$$

$$\mathbf{x}_0 = 0$$

Step 3: Weight computation

$$\mathbf{X} = []$$

$$\mathbf{d} = \text{ones}(N_u)$$

$$\mathbf{B} = (\mathbf{D} + \mathbf{L})^{-1}$$

for $k = 1 : 1 : K$

$$\mathbf{X} = \left[\mathbf{X} \text{diag} \left(\mathbf{W} (\mathbf{I}_{N_u} - \mathbf{B}\mathbf{W})^{(k-1)} \mathbf{B} \right) \right]$$

end for

$$\mathbf{a} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{d}$$

Step 4: GS iteration

for $k = 1 : 1 : K$

$$\mathbf{x}_k = (\mathbf{D} + \mathbf{L})^{-1} (\text{real}(a_{K-k+1}) \mathbf{s} - \mathbf{U} \mathbf{x}_{k-1})$$

end for

Step 5: GS precoder

$$\gamma = \text{trace}(\mathbf{x}_k \mathbf{s}^H)$$

$$\mathbf{t} = \frac{1}{\sqrt{\gamma}} \mathbf{H}^H \mathbf{x}_K$$

Step 6: Output parameter \mathbf{t}

Table 1: The meaning of parameters in the WGS algorithm

Parameters	Meaning
K	The number of iterations
$[]$	The empty matrix
$\text{ones}(N_u)$	The $N_u \times 1$ vector whose all elements are one
$\text{real}()$	The real part of the complex value
$\text{trace}()$	The sum of diagonal elements

4.3 Complexity Analysis

Tab. 2 expresses the complexity of iterative precoders compared in simulation results. The number of complex multiplications for only calculating \mathbf{x}_k is considered, and the complexity of the weight computation is excluded since the weight vector is not iteratively computed during the precoding process. The difference between the WGS and the GS method is the application of the weights as a coefficient of symbol vector in Eq. (14). The \mathbf{x}_k of the WGS method is expressed in terms of elements as follows:

$$x_i^{(k)} = \frac{1}{w_{i,i}} \left(u_{k-1} s_i - \sum_{j=1}^{i-1} w_{i,j} x_j^{(k)} - \sum_{j=i+1}^{N_u} w_{i,j} x_j^{(k-1)} \right), \quad (24)$$

where $x_i^{(k)}$, $x_i^{(k-1)}$, and s_i are the i -th element of \mathbf{x}_k , \mathbf{x}_{k-1} , and \mathbf{s} respectively, and $w_{i,j}$ is the element in the i -th row and j -th column of \mathbf{W} . The multiplication required for one element of \mathbf{x}_k can be divided as 4 parts. The number of multiplications outside parenthesis is 1. In parenthesis, The number of multiplications for the first term is 1, the number of multiplications for the second term is $i-1$, and the number of multiplication for the third term is N_u-i . Since the size of \mathbf{x}_k is $N_u \times 1$, the total number of multiplications for one iteration is $N_u^2 + N_u$.

Table 2: The complexity of iterative precoders

Scheme	The number of complex multiplications
WGS	$C_{WGS} = k(N_u^2 + N_u) + N_b N_u$ $\mathbf{x}_k: k(N_u^2 + N_u)$ $\mathbf{H}^H \mathbf{x}_k: N_b N_u$
GS with the initial solution of zero vector	$C_{GS_1} = k N_u^2 + N_b N_u$ $\mathbf{x}_k: k N_u^2$ $\mathbf{H}^H \mathbf{x}_k: N_b N_u$
GS with the initial solution of Eq. (12)	$C_{GS_2} = (k+3) N_u^2 - 2N_u + N_b N_u$ $\mathbf{x}_k: (k+3) N_u^2 - 2N_u$ [18] $\mathbf{H}^H \mathbf{x}_k: N_b N_u$
SOR	$C_{SOR} = k(N_u^2 + 2N_u) + N_b N_u$ $\mathbf{x}_k: k(N_u^2 + 2N_u)$ [20] $\mathbf{H}^H \mathbf{x}_k: N_b N_u$
ZF	$C_{ZF} = N_u^3 + 2N_u^3 + N_b N_u$ $\mathbf{W}^{-1}: N_u^3 + N_u^2$ (Gauss–Jordan) [14] $\mathbf{H}^H (\mathbf{W}^{-1} \mathbf{s}): N_u^2 + N_b N_u$

5 Performance Evaluation and Discussion

In this section, the BER performances of iterative precoders according to signal to noise ratio (SNR) are presented. The BER performance of the exact ZF precoder is provided as a benchmark. To confirm the difference in weight usage for BER performance, the GS precoder with the initial solution of zero vector is compared with the WGS precoder. Also, the BER performances of the GS precoder with the initial solution of Eq. (12) and the SOR precoder are provided to compare with the WGS precoder. In all simulations, the number of antennas at BS is fixed at 200 while the number of users is 20, 30 and 40 respectively, and 64-quadrature amplitude modulation (QAM) is used. The downlink channels are modelled as the Rayleigh fading channel with or without spatial correlation. In simulation results, K indicates the iteration number.

Fig. 3 shows the comparison of the use of different weight vectors for the WGS method. The number of users is 30, the correlation magnitude is 0.5, and the number of iterations is 5. In Fig. 3, using all elements means the use of one \mathbf{a}_{opt} , using diagonal elements means the use of the real part of one \mathbf{a}_{diag} , and optimal weights mean the use of optimal weight vector that

changes at each coherence interval. Fig. 3a presents the distance between weight values computed once by other methods and optimal weight values in complex plane according to the weight order. In Fig. 3a, the distance is averaged by 1000 optimal weight vectors. The graph shape can be different depending on the weight vector which is computed once. However, the difference in the graph shape does not change significantly since the one weight vector is determined close to the certain distribution as shown Fig. 2. The distance value calculated by comparison with the one \mathbf{a}_{opt} is smaller than using \mathbf{a}_{diag} , and the distance value of \mathbf{a}_{diag} on all weight order is smaller than 1. Furthermore, the BER performances are nearly same for all weight vectors in Fig. 3b. Thus, the use of \mathbf{a}_{diag} for the WGS precoder is reasonable in terms of the complexity and the BER performance, and the weight vector of the WGS precoder presented in all of following simulation results is based on the real part of the \mathbf{a}_{diag} .

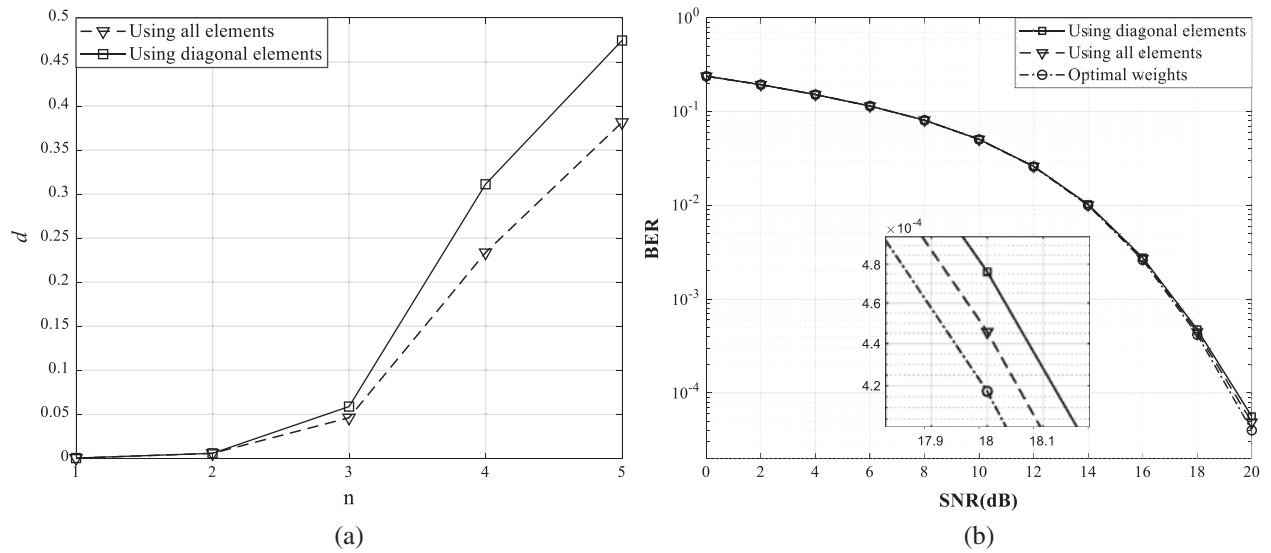


Figure 3: The comparison of using different weights in spatially correlated channel

Fig. 4 shows the average magnitude for the diagonal dominance of \mathbf{W} according to the different number of users and correlation magnitude. The average magnitude for the diagonal dominance of \mathbf{W} is calculated as follows:

$$v = \frac{1}{N_u} \sum_{i=1}^{N_u} \frac{|w_{i,i}|}{\sum_{j=1, j \neq i}^{N_u} |w_{i,j}|}, \quad (25)$$

where $w_{i,j}$ is the element in the i -th row and j -th column of \mathbf{W} . Since the low ratio ρ and high correlation magnitude weaken the channel hardening property, the average magnitude for the diagonal dominance of \mathbf{W} is decreased.

Fig. 5 shows the BER performances of the GS precoder with the initial solution of zero vector and the WGS precoder in ideal Rayleigh fading channel with no spatial correlation. In order to compare the WGS precoder with the GS precoder only with respect to the number of users, the number of users is chosen as 20, 30 and 40. For all number of users, the BER

performance of the WGS precoder is close to the exact ZF precoder in 3 iterations. However, the WGS and GS precoders provide nearly the same BER performance as the exact ZF precoder in Fig. 5a since the ratio ρ is large. In Figs. 5b and 5c, the difference of 2 dB is shown in the range of 18–20 dB. The difference in the BER performances is increasingly clear as the number of users increases.

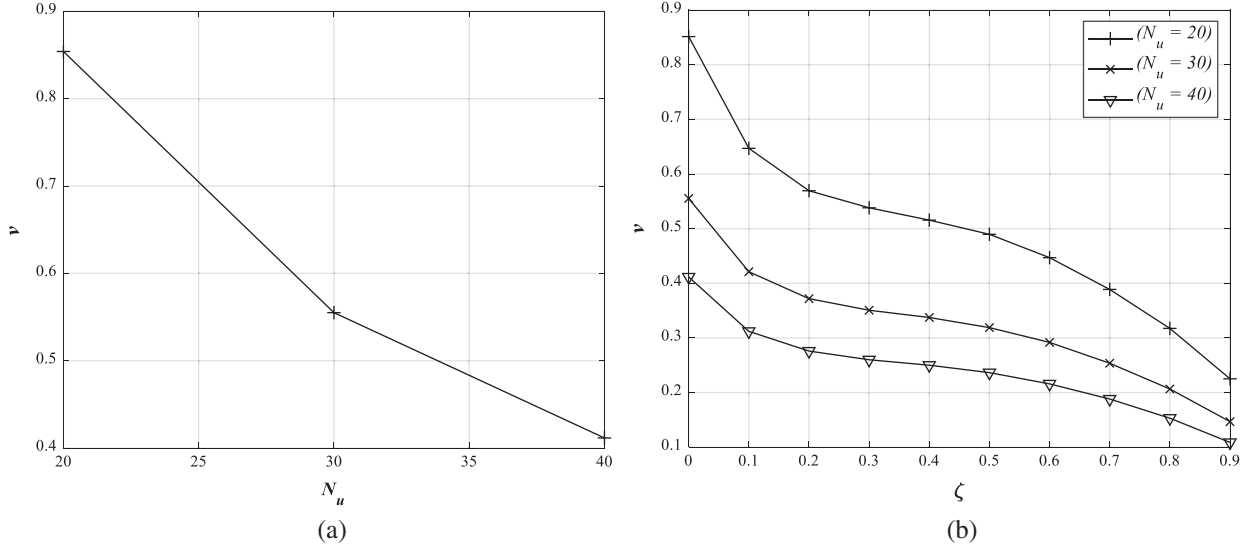


Figure 4: The average magnitude for diagonal dominance of \mathbf{W} according to the different number of users and correlation magnitude

Fig. 6 shows the BER performances of the GS precoder with the initial solution of zero vector and the WGS precoder in spatially correlated channel. The number of users is 20, and the correlation magnitude is 0.5. For $K = 4$, the BER performance of the WGS precoder is close to the exact ZF precoder. The WGS precoder provides better BER performance than the GS precoder in the same iterations since the weights accelerate the convergence rate. The difference in the BER performances is more obvious compared with Fig. 5 since the spatial correlation significantly degrades the BER performance of the GS precoder.

Fig. 7 shows the difference of approximation accuracy and convergence rate for the iterative methods according to the iteration number. In Fig. 7a, the approximation error between $\mathbf{W}\mathbf{W}_k^{-1}$ and \mathbf{I}_{N_u} for the iterative methods is shown. The approximation error is calculated as follows:

$$e = \left\| \mathbf{W}\mathbf{W}_k^{-1} - \mathbf{I}_{N_u} \right\|_F. \quad (26)$$

In Fig. 7b, the comparison of BER performances is shown for the SNR of 20 dB. The number of users is 40, the correlation magnitude is 0.5, and the range of iterations is from 2 to 8 to clearly confirm that the error is gradually decreased and the WGS method completely converges to the ZF method. The nearly optimal relaxation parameter of SOR method is empirically chosen for each configuration and correlation magnitude in Fig. 7 and following simulation results. It is confirmed that the graph shape in Fig. 7a is similar with that in Fig. 7b. Therefore, the decrease of the approximation error is directly related to the increase in BER performance. The WGS method shows the fastest convergence rate among the compared methods

since the weights decrease the approximation error. Also, the approximation error and BER performance for WGS method with $K = 5$ are smaller than other methods with $K = 6$.

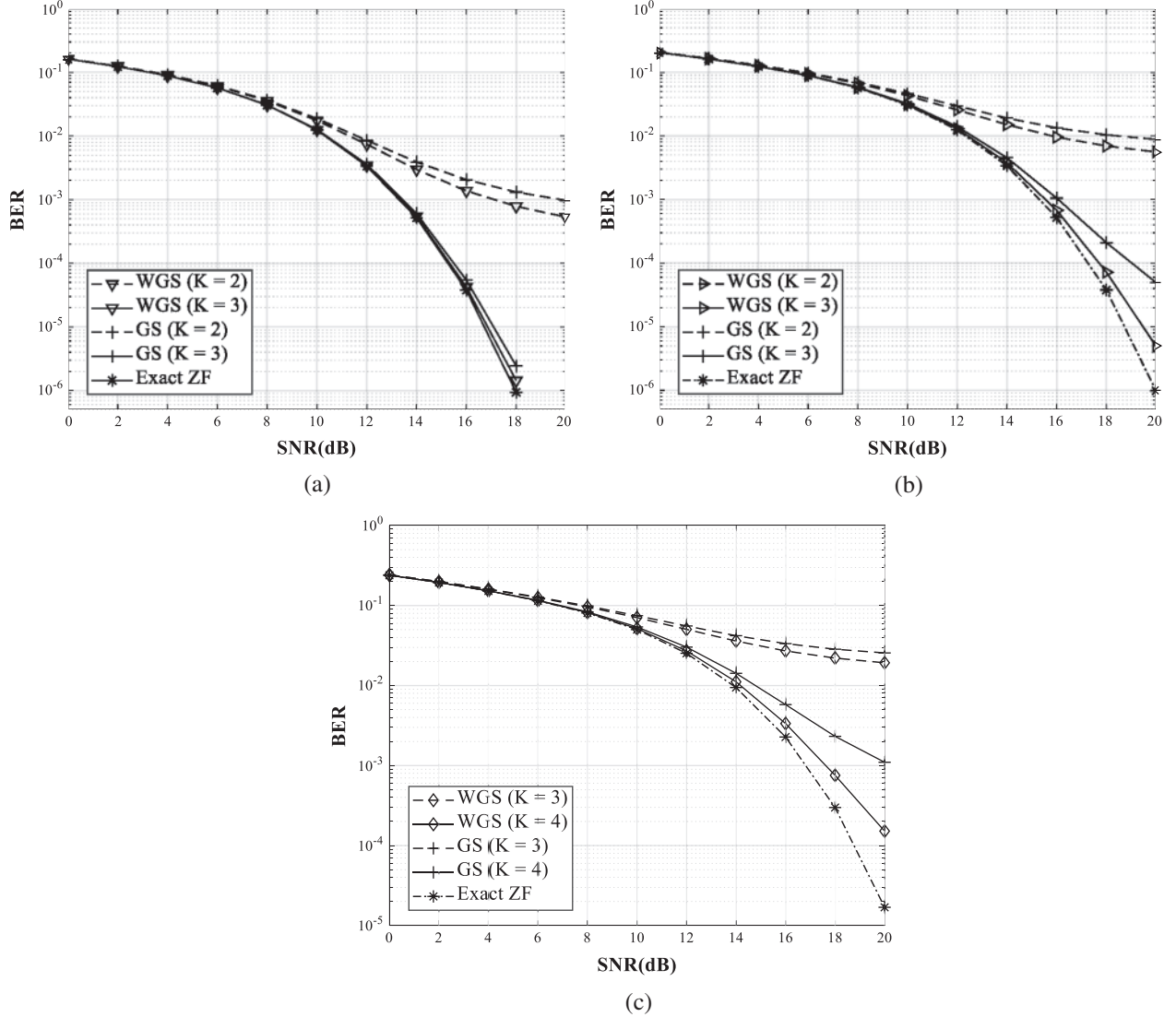


Figure 5: The BER performance comparison of the WGS, GS and exact ZF precoders in ideal Rayleigh fading channel with the different number of users. (a) $N_b = 200$, $N_u = 20$, $\zeta = 0$. (b) $N_b = 200$, $N_u = 30$, $\zeta = 0$. (c) $N_b = 200$, $N_u = 40$, $\zeta = 0$

Fig. 8 shows the BER performance for the WGS precoder, the GS precoder with the initial solution of zero vector, the GS precoder with the initial solution of Eq. (12), and the SOR precoder. In order to clarify the difference in BER performance among iterative precoders, the range of SNR is increased by 2 dB and 4 dB respectively. The correlation magnitude is fixed at 0.5, and the number of users is 30 and 40. In this situation, the WGS precoder requires more iteration to approach the BER performance of the exact ZF precoder since the convergence rate is sufficiently decreased. However, the WGS precoder outperforms the GS precoder with the

initial solution of zero vector. In Fig. 8a, the WGS precoder is close to the exact ZF precoder in 5 iterations, and the WGS precoder provides better BER performance than the other iterative precoders in the same iterations. In Fig. 8b, the WGS precoder is close to the exact ZF precoder in 6 iterations, and even the WGS precoder with 5 iterations provides better BER performance than the other iterative precoders with 6 iterations.

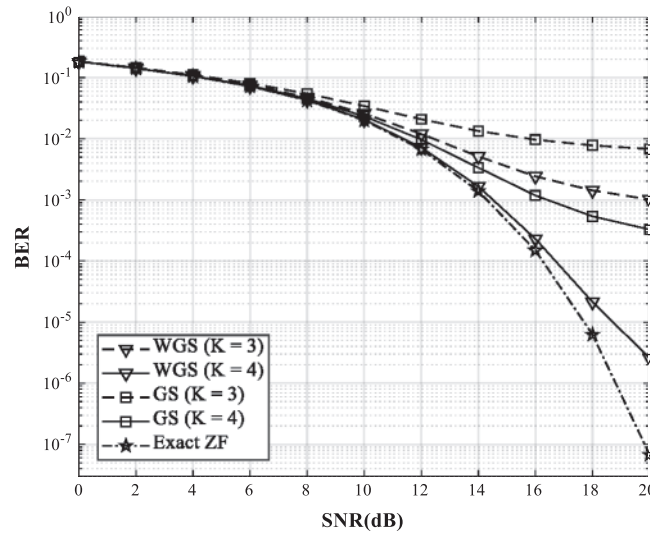


Figure 6: The BER performance comparison of the WGS, GS and exact ZF precoders in spatially correlated channel

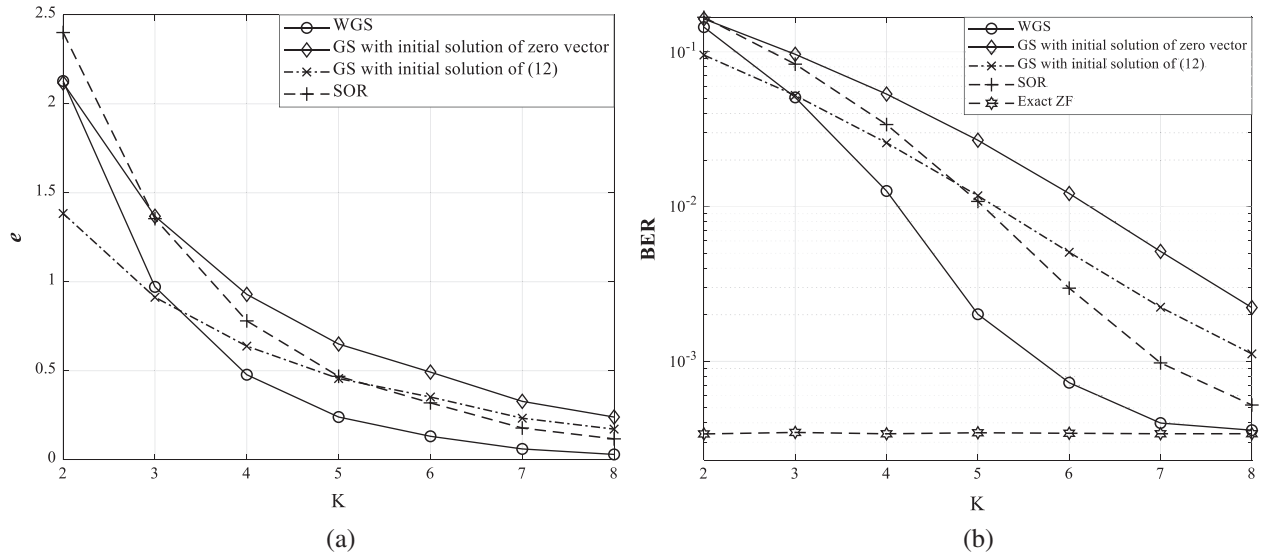


Figure 7: The approximation error and BER performances for the iterative methods according to the number of iterations

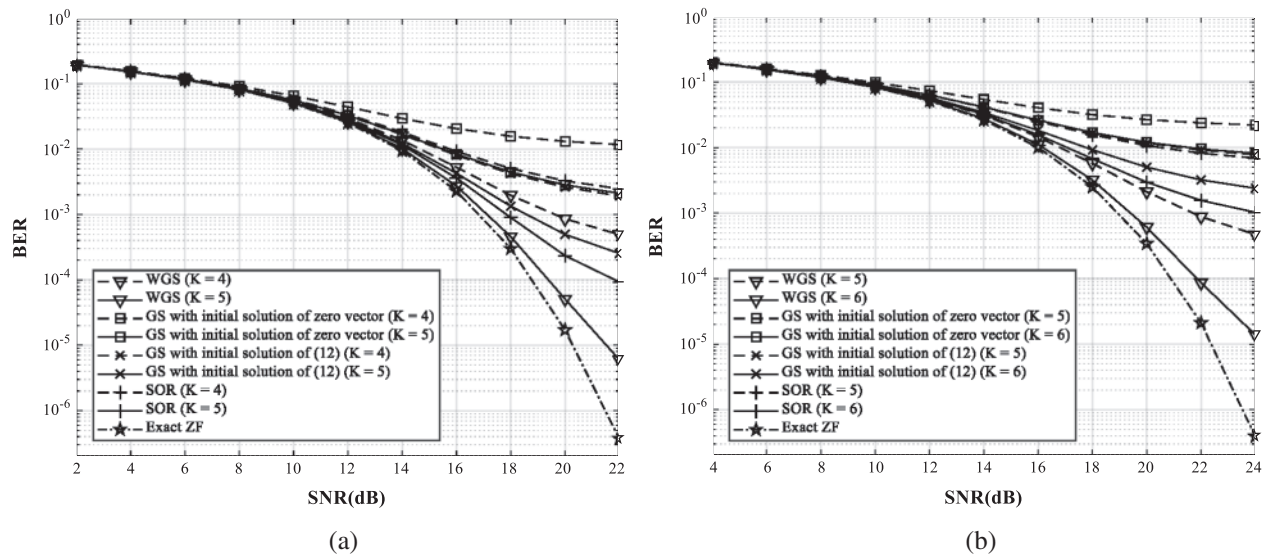


Figure 8: The BER performance comparison of the iterative precoders and the exact ZF precoders in spatially correlated channel with the different number of users. (a) $N_b = 200$, $N_u = 30$, $\zeta = 0.5$. (b) $N_b = 200$, $N_u = 40$, $\zeta = 0.5$

The improvement of the BER performance for the WGS precoder is noticeable in the large ratio ρ and spatially correlated channels. The WGS precoder maintains a faster convergence rate than the other iterative precoders, but large number of iterations can be required to converge to the BER performance of the exact ZF precoder shown in Fig. 8. However, the overall complexity is decreased by reducing the iteration number to satisfy demanded BER performance. In addition, the WGS precoder still maintains lower complexity than the exact ZF precoder according to the complexity analysis in Tab. 2.

6 Conclusion

The approximate methods based on iterative algorithm significantly reduce the computational complexity of direct matrix inversion and are efficient for the practical implementation of massive MIMO systems. In this paper, a weighted approximate method based on iterative algorithm is proposed to improve the approximation accuracy of the GS method. The WGS method uses different weights for each iteration of the GS method. To avoid the process that computes the optimal weights on every channel coherence interval, the weights are computed only once. The certain part of elements that is necessary for the computation of the weights is used to reduce the computational complexity. The WGS method maintains a fast convergence rate in spatially correlated channels. The simulation results show that the WGS method outperforms the GS method and provides better BER performance compared with other iterative methods.

Funding Statement: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2018-0-01423) supervised by the IITP (Institute for Information & communications Technology Promotion). Also, this research was supported by Basic Science Research Program through

the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03038540).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [2] H. Q. Ngo, E. G. Larsson and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.
- [3] E. G. Larsson, O. Edfors, F. Tufvesson and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.
- [5] T. Marzetta, E. Larsson, H. Yang and H. Ngo, *Fundamentals of Massive MIMO*. Cambridge, UK: Cambridge University Press, 2016.
- [6] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.
- [7] M. Mohaisen, H. An and K. Chang, "Detection techniques for MIMO multiplexing: A comparative review," *KSII Transactions on Internet and Information Systems*, vol. 3, no. 6, pp. 647–666, 2009.
- [8] M. A. Albreem, M. Juntti and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.
- [9] M. A. M. Albreem, "Approximate matrix inversion methods for massive MIMO detectors," in *2019 IEEE 23rd Int. Sym. on Consumer Technologies (ISCT)*, Ancona, Italy, pp. 87–92, 2019.
- [10] Z. Zhang, J. Wu, X. Ma, Y. Dong, Y. Wang *et al.*, "Reviews of recent progress on low complexity linear detection via iterative algorithms for massive MIMO systems," in *IEEE/CIC Int. Conf. on Communications in China (ICCC Workshops)*, Chengdu, China, pp. 1–6, 2016.
- [11] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Philadelphia: Society for Industrial and Applied Mathematics, 2003.
- [12] H. Prabhu, J. Rodrigues, O. Edfors and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear precoding systems," in *IEEE Wireless Communications and Networking Conf. (WCNC)*, Shanghai, China, pp. 2710–2715, 2013.
- [13] D. Zhu, B. Li and P. Liang, "On the matrix inversion approximation based on Neumann series in massive MIMO systems," in *IEEE Int. Conf. on Communications (ICC)*, London, UK, pp. 1763–1769, 2015.
- [14] J. Ro, W. Lee, J. Ha and H. Song, "An efficient precoding method for improved downlink massive MIMO system," *IEEE Access*, vol. 7, pp. 112318–112326, 2019.
- [15] M. Albreem, M. Juntti and S. Shahabuddin, "Efficient initialisation of iterative linear massive MIMO detectors using a stair matrix," *Electronics Letters*, vol. 56, no. 1, pp. 50–52, 2020.
- [16] L. Dai, X. Gao, X. Su, S. Han, C. I. *et al.*, "Low-complexity soft output signal detection based on gauss-seidel method for uplink multiuser large-scale MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4839–4845, 2015.
- [17] Z. Wu, C. Zhang, Y. Xue, S. Xu and X. You, "Efficient architecture for soft-output massive MIMO detection with gauss-seidel method," in *2016 IEEE Int. Sym. on Circuits and Systems (ISCAS)*, Montreal, Canada, pp. 1886–1889, 2016.

- [18] C. Zhang, Z. Wu, C. Studer, Z. Zhang and X. You, "Efficient soft-output gauss-seidel data detector for massive MIMO systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, pp. 1–12, 2018.
- [19] J. Zeng, J. Lin and Z. Wang, "An improved gauss-seidel algorithm and its efficient architecture for massive MIMO systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 9, pp. 1194–1198, 2018.
- [20] X. Gao, L. Dai, Y. Hu, Z. Wang and Z. Wang, "Matrix inversion-less signal detection using SOR method for uplink large-scale MIMO systems," in *IEEE Global Communications Conf.*, Austin, Texas, USA, pp. 3291–3295, 2014.
- [21] A. Yu, C. Zhang, S. Zhang and X. You, "Efficient sor-based detection and architecture for large-scale MIMO uplink," in *IEEE Asia Pacific Conf. on Circuits and Systems (APCCAS)*, Jeju, Korea (South), pp. 402–405, 2016.
- [22] A. Yu, S. Jing, X. Tan, Z. Wu, Z. Yan *et al.*, "Efficient successive over relaxation detectors for massive MIMO," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 2128–2139, 2020.
- [23] B. Nagy, M. Elsabrouty and S. Elramly, "Fast converging weighted Neumann series precoding for massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 154–157, 2018.
- [24] X. Liu, Z. Zhang, X. Wang, J. Lian and X. Dai, "A low complexity high performance weighted Neumann series-based massive MIMO detection," in *2019 28th Wireless and Optical Communications Conf. (WOCC)*, Beijing, China, pp. 1–5, 2019.
- [25] J. Choi and D. J. Love, "Bounds on eigenvalues of a spatial correlation matrix," *IEEE Communications Letters*, vol. 18, no. 8, pp. 1391–1394, 2014.
- [26] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York City, USA: Springer-Verlag, 2009.