

RESEARCH

Open Access



Low-complexity signal detection networks based on Gauss-Seidel iterative method for massive MIMO systems

Haifeng Yao¹, Ting Li¹, Yunchao Song², Wei Ji¹, Yan Liang¹ and Fei Li^{1*} 

*Correspondence:
lifei@njupt.edu.cn

¹ School of communication and information engineering, Nanjing University of Posts and Telecommunications, Nanjing, China

² College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China

Abstract

In massive multiple-input multiple-output (MIMO) systems with single- antenna user equipment (SAUE) or multiple-antenna user equipment (MAUE), with the increase of the number of received antennas at base station, the complexity of traditional detectors is also increasing. In order to reduce the high complexity of parallel running of the traditional Gauss-Seidel iterative method, this paper proposes a model-driven deep learning detector network, namely Block Gauss-Seidel Network (BGS-Net), which is based on the Gauss-Seidel iterative method. We reduce complexity by converting a large matrix inversion to small matrix inversions. In order to improve the symbol error ratio (SER) of BGS-Net under MAUE system, we propose Improved BGS-Net. The simulation results show that, compared with the existing model-driven algorithms, BGS-Net has lower complexity and similar the detection performance; good robustness, and its performance is less affected by changes in the number of antennas; Improved BGS-Net can improve the detection performance of BGS-Net.

Keywords: MIMO detection, Deep learning, Gauss-Seidel, SAUE system, MAUE system

1 Introduction

Beyond fifth generation (B5G) is a perfect process to solve some application scenarios and technologies of 5th generation mobile communication technology (5G). Artificial intelligence technology is the engine of B5G. In recent years, machine learning algorithms have been used in various fields such as healthcare, transportation, energy, and self-driving cars. These algorithms are also being used in communication technologies to improve system performance in terms of spectrum utilization, latency and security. With the rapid development of machine learning techniques, especially deep learning techniques, it is crucial to consider symbol error ratio (SER) and complexity when applying algorithms [1]. Massive multi-input multi-output (MIMO) is a key technology in B5G, where tens, hundreds or thousands of antennas are equipped at the base station. This system makes the signal detection problem a big challenge, because the computational complexity of the detector increases with the number of antennas [2, 3]. Therefore, how to find a balance between detection accuracy and complexity in massive MIMO signal detection has become a hot topic of research for domestic and foreign scholars.

In conventional signal detection methods, maximum likelihood (ML) is a nonlinear maximum likelihood detector. However, its complexity increases exponentially as the number of transmitting antennas increases, hindering its implementation in practical MIMO systems [2]. The spherical decoding (SD) detector [3] and the k-best detector [4] are two variants of ML detectors that balance computational complexity and SER by controlling the number of nodes in each search phase. Unfortunately, QR decomposition in these nonlinear detectors leads to high computational complexity and low parallelism because of the inclusion of unfavorable matrix operations, such as element elimination. In contrast, suboptimal linear detectors, such as minimum mean square error (MMSE) [5] and zero forcing (ZF) [6], provide a better trade-off between SER and computational complexity, but their complexity still reaches three times the number of transmitting antennas.

In order to reduce the complexity of matrix inversion, in 2013, Wu et al. proposed an approximate inversion-based uplink detection algorithm in [7]. Over the next few years, a large number of MIMO detectors designed for specific massive MIMO systems continued to appear. The main idea of the proposed method is to use iterative methods to approximate the inverse of a matrix or to avoid the computation of exact matrices. For example, the neumann series method (NS) [8], the newton iterative method (NI) [9], the Gauss-Seidel method (GS) [10], the successive super-relaxation method (SOR) [11], the jacobi (JA) method [12], the richardson method (RI) [13], the conjugate gradient method (CG) [14], the lanczos method (LA) [15], residual method [16], coordinate descent method (CD) [17], belief propagation method (BP) [18], etc., these algorithms successfully reduce the complexity to $O(M^2)$, but the SER is only close to MMSE.

In order to improve the performance of detection algorithms, [19–21] introduced deep learning into communication. These methods treat the functional blocks of wireless communication as black boxes and replace them with deep learning networks. The mapping relationship between input and output data is obtained by training a large amount of data in an offline training phase. However, deepening the network does not significantly improve the performance beyond a certain number of layers, for this reason [22] proposed a parallel detection network (PDN), which consists of several unconnected deep learning detection networks in parallel. By designing specific loss functions that reduce the similarity between detection networks, the PDN obtains considerable diversity results. These algorithms are pure black boxes and although they improve the performance of detection, they require a large amount of training data to learn a large number of parameters, the advantage of these algorithms is that they do not require the incorporation of communication knowledge. [23, 24] proposed a modern neural network structure suitable for this detection task, detection network (DetNet). The structure of DetNet is obtained by expanding the iterations of the projected gradient descent algorithm into the network. [25, 26] proposed the orthogonal approximate message passing (OAMPNet), a model-driven deep learning network for multiple-input multiple-output (MIMO) detection, and [27, 28] proposed MIMO detection network (MMNet), a deep learning MIMO detection scheme. The design of MMNet is based on the theory of iterative soft thresholding algorithms, which significantly outperforms existing methods on realistic channels of the same or lower computational complexity. These algorithms are purely white-box model iterative and have better performance than convolutional neural network (CNN), deep neural networks (DNN), but are not as applicable as its

wide range. In addition to these, there are BP-Net [29], CG-Net [30], which are networks changed based on approximation methods. [31] proposed a data-driven implementation of an iterative soft interference cancellation (SIC) algorithm, called DeepSIC. This method significantly outperforms model-based methods in the presence of channel state information (CSI) uncertainty, but the network is more complex and this method is a combination of black-box and white-box methods. Therefore, we know that deep learning methods can improve the performance of detection. However, when the number of antennas is large, deep learning not only requires high hardware requirements, but also requires training, which in reality, can have a significant delay.

Therefore, we consider using approximate inversion methods to reduce complexity while using deep learning methods to improve SER. However, most of the above-mentioned work is directed to a single-antenna user equipment (SAUE) system, and it is assumed that the channel matrix is independent, identically distributed (i.i.d) and obeys gaussian distribution. Unfortunately, in practice, because a user maybe equipped with several antennas, the antennas from the same user equipment (UE) are not sufficiently separated [32], so their transmission vectors are usually related. The spatial correlation between antennas is a key factor affecting the performance of the massive MIMO (M-MIMO) system. Therefore, this paper also considers the multiple-antenna user equipment(MAUE) system while considering the SAUE system.

In this paper, we propose a model-driven deep learning detector network Block Gauss-Seidel Network (BGS-Net) to solve the high complexity caused by the parallel operation of traditional Gauss-Seidel [10], which is based on the Gauss-Seidel iterative method . We reduce the complexity by converting large matrix inversions $(\mathbf{D} + \mathbf{L})^{-1}$ to small matrix inversions and converting matrix-by-matrix to matrix-by-vector. This paper considers SAUE and MAUE systems [32, 33] under Rayleigh channels. In order to improve SER of BGS-Net under MAUE system, we improve the initial solution of BGS-Net by replacing $\mathbf{x}_0 = \mathbf{D}^{-1}\mathbf{H}^T\mathbf{y}$ with $\mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{H}^T\mathbf{y}$. For \mathbf{A}^{-1} , we use block matrix approximation to reduce its complexity. Simulation results show that compared with existing model-driven algorithms, BGS-Net has lower complexity and similar SER; good robustness, its performance is less affected by changes in the number of antennas; SER is better than traditional Gauss-Seidel; Improved BGS-Net can improve the SER of BGS-Net.

This paper is organised as follows. Section 2 presents and analyses the channels required in this paper. Section 3 analyses the existing OAMPNet/MMNet-iid algorithms. Section 4 proposes the algorithm BGS-Net, and explains why BGS-Net is proposed. Section 5 analyses the problems that BGS-Net may encounter in MAUE systems and improve it, and proposes Improved BGS-Net. Section 6 analyses the complexity of BGS-Net and Improved BGS-Net in terms of their complexity. Experimental simulations and discussion are performed in Section 7. Section 8 concludes the full paper.

1.1 Notation

In this paper, lower-case and upper-case boldface letters are used to represent column vectors and matrices, respectively . \mathbf{I}_n denotes a unit matrix of size n. For any matrix \mathbf{A} , \mathbf{A}^T , \mathbf{A}^H , $tr(\mathbf{A})$, and \mathbf{A}^+ represent the transpose, conjugate transpose, trace, and pseudo-inverse of \mathbf{A} . $N_C(s_i; r_i, \tau_i^2)$ denotes the univariate Gaussian distribution of a random variable s_i with

mean r_i and variance τ_i^2 . The operator $\|\cdot\|$ denotes the vector/matrix parametrization. The notation $diag(\mathbf{x})$ creates a matrix with \mathbf{x} in the diagonal and $diag(\mathbf{X})$ is the vector of the diagonal elements of \mathbf{X} .

2 Background

2.1 SAUE System

Consider an uplink massive MIMO system that uses N_r antennas at the BS to serve N_t single-antenna user terminals simultaneously, where $N_r \gg N_t$. The SAUE system can be simply expressed as:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}_S \tilde{\mathbf{x}} + \tilde{\mathbf{n}} \tag{1}$$

where $\tilde{\mathbf{y}} \in \mathbb{C}^{N_r \times 1}$, $\tilde{\mathbf{H}}_S \in \mathbb{C}^{N_r \times N_t}$, $\tilde{\mathbf{x}} \in \mathbb{C}^{N_t \times 1}$, and $\tilde{\mathbf{n}} \in \mathbb{C}^{N_r \times 1}$ are the receive symbol, channel response, transmit symbol, and system noise respectively. N_r and N_t are the numbers of receive and transmit antennas, respectively. $\tilde{\mathbf{n}}$ is distributed as $CN(0, \sigma^2)$. For signal detection, the complex-valued system model (1) is converted to the corresponding real-valued system model as

$$\mathbf{y} = \mathbf{H}_S \mathbf{x} + \mathbf{n} \tag{2}$$

where $\mathbf{y} = [\Re(\tilde{\mathbf{y}})^T \ \Im(\tilde{\mathbf{y}})^T]^T \in \mathbb{R}^{2N_r \times 1}$, $\mathbf{x} = [\Re(\tilde{\mathbf{x}})^T \ \Im(\tilde{\mathbf{x}})^T]^T \in \mathbb{R}^{2N_t \times 1}$, all \mathbf{x} come from the discrete constellation diagram $S = \{s_1, s_2, \dots, s_M\}$,

$$\mathbf{n} = [\Re(\tilde{\mathbf{n}})^T \ \Im(\tilde{\mathbf{n}})^T]^T \in \mathbb{R}^{2N_r \times 1}, \quad \mathbf{H}_S = \begin{bmatrix} \Re(\tilde{\mathbf{H}}_S) & -\Im(\tilde{\mathbf{H}}_S) \\ \Im(\tilde{\mathbf{H}}_S) & \Re(\tilde{\mathbf{H}}_S) \end{bmatrix} \in \mathbb{R}^{2N_r \times 2N_t}, \quad N = 2N_r,$$

$M = 2N_t$. \mathbf{H}_S denotes the flat Rayleigh fading channel matrix whose entries are assumed to be independently and identically distributed (i.i.d.) with zero mean and variance $(1/N)\mathbf{I}$. Since each user is a single antenna, the correlation between them is not considered.

2.2 MAUE System

Consider an uplink massive MIMO with a multi-antenna user equipment (MAUE) system. A BS with N_r antennas communicates with m UEs, and each UE is equipped with m_{UE} antennas, as shown in Figure 1. The total number of antennas on the user side is $N_t = m \times m_{UE}$. The transmission vector is expressed as $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m]^T$, where $\mathbf{x}_i = [x_{i1}, \dots, x_{ij}, \dots, x_{im_{UE}}] \in \mathbb{R}^{1 \times m_{UE}}$, $E\{x_{ij}^2\} = 1$. $N = 2N_r$, $M = 2N_t$, the vector $\mathbf{y} \in \mathbb{R}^{N \times 1}$ received by the BS:

$$\mathbf{y} = \mathbf{H}_M \mathbf{x} + \mathbf{n} \tag{3}$$

where $\mathbf{H}_M = [\mathbf{H}_{M1}, \dots, \mathbf{H}_{Mi}, \dots, \mathbf{H}_{Mm}] \in \mathbb{R}^{N \times M}$, $\mathbf{H}_{Mi} = [\mathbf{H}_{Mi1}, \dots, \mathbf{H}_{Mim_{UE}}]$. $\mathbf{H}_{Mij} \in \mathbb{R}^{N \times 1}$, represents the uplink from the j th antenna of the i th UE to the BS, $\mathbf{n} \in \mathbb{R}^{N \times 1}$ is an additive white gaussian noise (AWGN) vector with a mean value of zero and a variance of $\sigma^2/2$. The Kronecker channel model [34] is: $\mathbf{H}_M = \mathbf{R}^{1/2} \mathbf{H}_S \mathbf{T}^{1/2}$. Spatial correlation matrix $\mathbf{R} \in \mathbb{R}^{N \times N}$ and $\mathbf{T} \in \mathbb{R}^{M \times M}$.

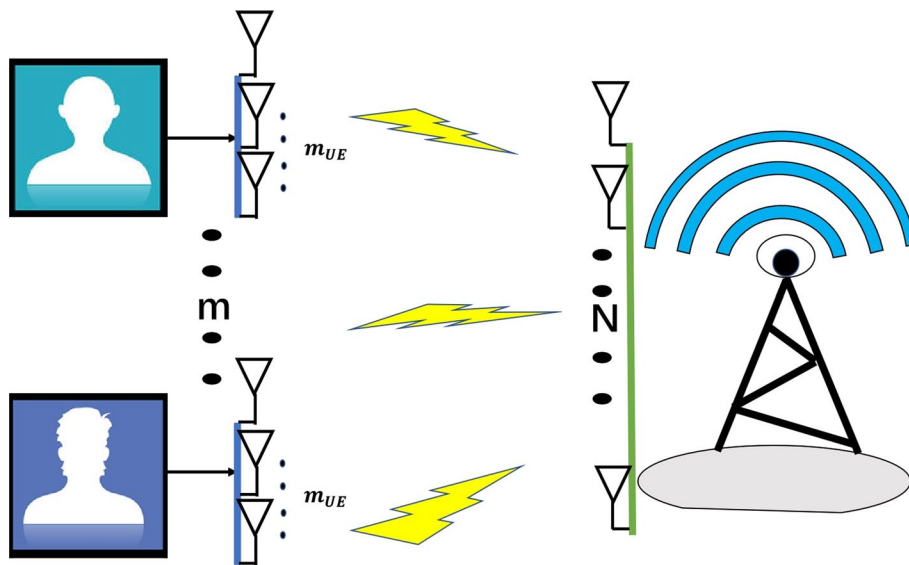


Fig. 1 An uplink M-MIMO system

$$R_{pq} = \begin{cases} (\xi_r e^{j\theta})^{2(q-p)} & \text{if } p \leq q \\ R_{qp}^* & \text{if } p > q \end{cases} \quad (4)$$

$$T_{pq} = \begin{cases} (\xi_t e^{j\theta})^{2(q-p)} & \text{if } p \leq q \\ T_{qp}^* & \text{if } p > q \end{cases} \quad (5)$$

where ξ_r and ξ_t are correlation coefficients, R_{pq} is the value of the receive antenna correlation matrix \mathbf{R} , T_{pq} is the value of the transmit antenna correlation matrix \mathbf{T} for each user. It can be seen that the transmitting antennas from the same terminal usually have correlation. However, most of the current papers do not consider the correlation between antennas from the same terminal, which is impractical and inaccurate. Therefore, according to the actual situation of propagation, ξ_r and ξ_t are respectively defined as the correlation factors of the receiving antenna and the transmitting antenna from the same terminal. Note that the correlation between different terminal antennas is ignored [32, 33].

2.3 Channel characteristics

In this section, SAUE and MAUE system characteristics are analyzed. $N_t = 4$, $N_r = 32$ are used here. It can be seen from Figure 2 that when the SAUE system has a large-scale antenna and $\alpha = N_r/N_t$ is large, the channel appears channel hardening, and $\mathbf{H}_S^T \mathbf{H}_S$ is diagonally dominant, so the approximate inversion method is very suitable for this environment. Figure 3 shows that in the MAUE system environment with $\xi_r = 0$, $\xi_t = 0.2$, $m_{UE} = 2$, although $\mathbf{H}_M^T \mathbf{H}_M$ is still dominant diagonally and exhibits blockiness, it can be seen from the color depth that the non-diagonal elements on both sides of the diagonal elements have begun to have an impact on the diagonal elements and cannot be ignored. Figure 4 shows that in the MAUE system environment with $\xi_r = 0$, $\xi_t = 0.4$, $m_{UE} = 4$, the two sides of the diagonal element have had a great impact on the diagonal element, and it is difficult to get a

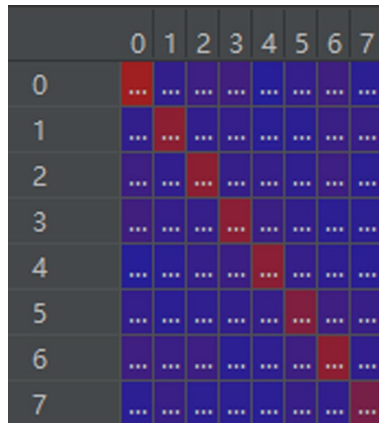


Fig. 2 $H_S^T H_S$ in single-antenna user equipment (SAUE) system

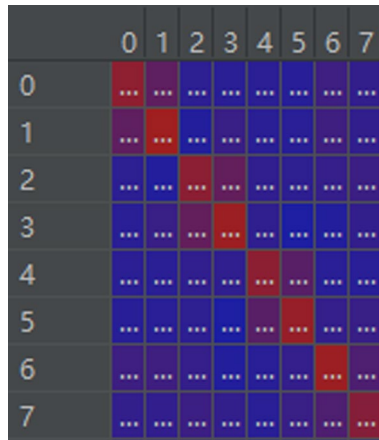


Fig. 3 $H_M^T H_M$ in multi-antenna user equipment (MAUE) with $\xi_r = 0, \xi_t = 0.2, m_{UE} = 2$

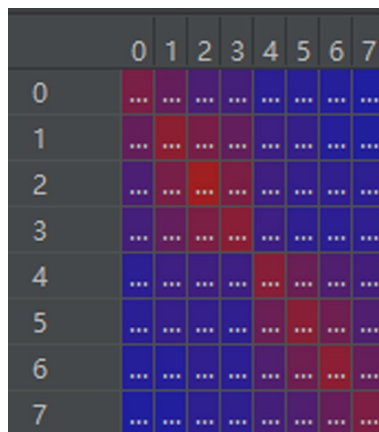


Fig. 4 $H_M^T H_M$ in multi-antenna user equipment (MAUE) with $\xi_r = 0, \xi_t = 0.4, m_{UE} = 4$

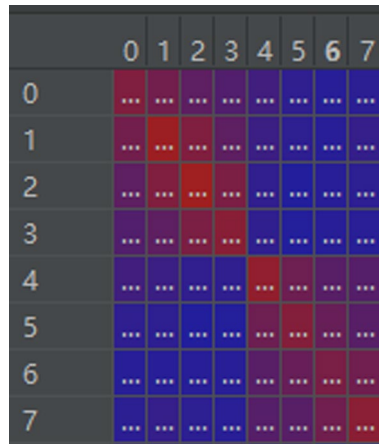


Fig. 5 $H_M^T H_M$ in multi-antenna user equipment (MAUE) with $\xi_r = 0.2, \xi_t = 0.4, m_{UE} = 4$

good approximation by the approximation method. Figure 5 shows that in the MAUE system environment with $\xi_r = 0.2, \xi_t = 0.4, m_{UE} = 4$, the non-diagonal elements around the diagonal elements have a serious impact on the diagonal elements, and the approximation effect is very poor.

The Marchenko-Pasture theorem of random matrix theory shows that when each element of the matrix channel \mathbf{H} is independently and identically distributed at zero mean and the variance is $1/N$, the number of rows N and the number of columns M tend to Infinity, that is, $M, N \rightarrow \infty$, and the ratio of the two tends to a constant ($N/M \rightarrow \beta$), the diagonal elements of the matrix $\mathbf{H}^T \mathbf{H}$ tend to a certain constant, and the off-diagonal elements tend to zero. The following analyzes the symmetry of $\mathbf{H}^T \mathbf{H}$ under SAUE and MAUE systems: Known from formulas (4) and (5): \mathbf{R} and \mathbf{T} are symmetric matrices, so suppose

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{R}_2^T & \mathbf{R}_1 \end{bmatrix} \tag{6}$$

and

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 & \cdots & 0 & 0 \\ \mathbf{T}_2^T & \mathbf{T}_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{T}_1 & \mathbf{T}_2 \\ 0 & 0 & \cdots & \mathbf{T}_2^T & \mathbf{T}_1 \end{bmatrix} \tag{7}$$

where $\mathbf{R}_1, \mathbf{R}_2 \in \mathbb{R}^{N/2 \times N/2}, \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 \\ \mathbf{T}_2^T & \mathbf{T}_1 \end{bmatrix} \in \mathbb{R}^{m_{UE} \times m_{UE}}, \mathbf{T}_1, \mathbf{T}_2 \in \mathbb{R}^{m_{UE}/2 \times m_{UE}/2}$.

(a) $\mathbf{H}_S^T \mathbf{H}_S$ under SAUE system. From equation (2) we know that:

$$\begin{aligned} \mathbf{H}_S^T \mathbf{H}_S &= \begin{bmatrix} \Re(\tilde{\mathbf{H}}_S)^T & \Im(\tilde{\mathbf{H}}_S)^T \\ -\Im(\tilde{\mathbf{H}}_S)^T & \Re(\tilde{\mathbf{H}}_S)^T \end{bmatrix} \begin{bmatrix} \Re(\tilde{\mathbf{H}}_S) & -\Im(\tilde{\mathbf{H}}_S) \\ \Im(\tilde{\mathbf{H}}_S) & \Re(\tilde{\mathbf{H}}_S) \end{bmatrix} \\ &= \begin{bmatrix} \Re(\tilde{\mathbf{H}}_S)^T \Re(\tilde{\mathbf{H}}_S) + \Im(\tilde{\mathbf{H}}_S)^T \Im(\tilde{\mathbf{H}}_S) & -\Re(\tilde{\mathbf{H}}_S)^T \Im(\tilde{\mathbf{H}}_S) + \Im(\tilde{\mathbf{H}}_S)^T \Re(\tilde{\mathbf{H}}_S) \\ -\Im(\tilde{\mathbf{H}}_S)^T \Re(\tilde{\mathbf{H}}_S) + \Re(\tilde{\mathbf{H}}_S)^T \Im(\tilde{\mathbf{H}}_S) & \Im(\tilde{\mathbf{H}}_S)^T \Im(\tilde{\mathbf{H}}_S) + \Re(\tilde{\mathbf{H}}_S)^T \Re(\tilde{\mathbf{H}}_S) \end{bmatrix} \end{aligned} \tag{8}$$

It can be seen from (8) that the upper left matrix and the lower right matrix of $\mathbf{H}_S^T \mathbf{H}_S$ are the same, and the matrix is symmetric about the main diagonal.

(b) $\mathbf{H}_M^T \mathbf{H}_M$ under MAUE system (only $\mathbf{T}^{1/2}$)

$$\mathbf{H}_M^T \mathbf{H}_M = (\mathbf{H}_S \mathbf{T}^{1/2})^T \mathbf{H}_S \mathbf{T}^{1/2} = \mathbf{T}^{1/2} \mathbf{H}_S^T \mathbf{H}_S \mathbf{T}^{1/2} \tag{9}$$

where $\mathbf{H}_S = [\mathbf{H}_{S_1}, \dots, \mathbf{H}_{S_l}, \dots, \mathbf{H}_{S_m}] \in \mathbb{R}^{N \times M}$. It can be seen from (8) that the upper left corner matrix and the lower right corner matrix of $\mathbf{H}_S^T \mathbf{H}_S$ are the same, and the matrix is symmetric about the main diagonal. We rewrite (8) formula $\mathbf{H}_S^T \mathbf{H}_S = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \\ \mathbf{Q}_2^T & \mathbf{Q}_1 \end{bmatrix}$, and rewrite (7) formula $\mathbf{T}^{1/2} = \begin{bmatrix} \mathbf{K}_1^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_1^{1/2} \end{bmatrix}$, where $\mathbf{K}_1 \in \mathbb{R}^{M/2 \times M/2}$, $\mathbf{Q}_1 \in \mathbb{R}^{M/2 \times M/2}$, $\mathbf{Q}_2 \in \mathbb{R}^{M/2 \times M/2}$. Bringing \mathbf{T} into equation (9), you can write equation (9) as

$$\begin{aligned} \mathbf{T}^{1/2} \mathbf{H}_S^T \mathbf{H}_S \mathbf{T}^{1/2} &= \begin{bmatrix} \mathbf{K}_1^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_1^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \\ \mathbf{Q}_2^T & \mathbf{Q}_1 \end{bmatrix} \begin{bmatrix} \mathbf{K}_1^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_1^{1/2} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{K}_1^{1/2} \mathbf{Q}_1 \mathbf{K}_1^{1/2} & \mathbf{K}_1^{1/2} \mathbf{Q}_2 \mathbf{K}_1^{1/2} \\ \mathbf{K}_1^{1/2} \mathbf{Q}_2^T \mathbf{K}_1^{1/2} & \mathbf{K}_1^{1/2} \mathbf{Q}_1 \mathbf{K}_1^{1/2} \end{bmatrix} \end{aligned} \tag{10}$$

Therefore, the upper left corner matrix of $\mathbf{H}_M^T \mathbf{H}_M$ is the same as the lower right corner matrix, and the matrix is symmetric about the main diagonal.

(c) $\mathbf{H}_M^T \mathbf{H}_M$ under MAUE system (only $\mathbf{T}^{1/2}$ and $\mathbf{R}^{1/2}$)

$$\begin{aligned} \mathbf{H}_M^T \mathbf{H}_M &= (\mathbf{R}^{1/2} \mathbf{H}_S \mathbf{T}^{1/2})^T (\mathbf{R}^{1/2} \mathbf{H}_S \mathbf{T}^{1/2}) \\ &= (\mathbf{T}^{1/2})^T (\mathbf{R}^{1/2} \mathbf{H}_S)^T \mathbf{R}^{1/2} \mathbf{H}_S \mathbf{T}^{1/2} \end{aligned} \tag{11}$$

where

$$\mathbf{R}^{1/2} \mathbf{H}_S = \begin{bmatrix} \mathbf{R}_1^{1/2} \Re(\tilde{\mathbf{H}}_S)^T - \mathbf{R}_2^{1/2} \Im(\tilde{\mathbf{H}}_S)^T & \mathbf{R}_1^{1/2} \Im(\tilde{\mathbf{H}}_S)^T + \mathbf{R}_2^{1/2} \Re(\tilde{\mathbf{H}}_S)^T \\ (\mathbf{R}_2^T)^{1/2} \Re(\tilde{\mathbf{H}}_S)^T - \mathbf{R}_1^{1/2} \Im(\tilde{\mathbf{H}}_S)^T & (\mathbf{R}_2^T)^{1/2} \Im(\tilde{\mathbf{H}}_S)^T + \mathbf{R}_1^{1/2} \Re(\tilde{\mathbf{H}}_S)^T \end{bmatrix} \tag{12}$$

After calculation, we can find that the matrix $(\mathbf{R}^{1/2} \mathbf{H}_S)^T \mathbf{R}^{1/2} \mathbf{H}_S$ is symmetric about the main diagonal, but the upper left matrix and the lower right matrix are not the same, we set

$$(\mathbf{R}^{1/2} \mathbf{H}_S)^T \mathbf{R}^{1/2} \mathbf{H}_S = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 \\ \mathbf{Z}_2^T & \mathbf{Z}_3 \end{bmatrix} \tag{13}$$

where $\mathbf{Z}_1 \in \mathbb{R}^{M/2 \times M/2}$, $\mathbf{Z}_2 \in \mathbb{R}^{M/2 \times M/2}$, $\mathbf{Z}_3 \in \mathbb{R}^{M/2 \times M/2}$. Substitute (13) into $\mathbf{H}_M^T \mathbf{H}_M$, get

$$\begin{aligned} \mathbf{H}_M^T \mathbf{H}_M &= \begin{bmatrix} \mathbf{K}_1^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_1^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 \\ \mathbf{Z}_2^T & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{K}_1^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_1^{1/2} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{K}_1^{1/2} \mathbf{Z}_1 \mathbf{K}_1^{1/2} & \mathbf{K}_1^{1/2} \mathbf{Z}_2 \mathbf{K}_1^{1/2} \\ \mathbf{K}_1^{1/2} \mathbf{Z}_2^T \mathbf{K}_1^{1/2} & \mathbf{K}_1^{1/2} \mathbf{Z}_3 \mathbf{K}_1^{1/2} \end{bmatrix} \end{aligned} \tag{14}$$

It can be seen from formula (14) that the upper left corner matrix and the lower right corner matrix of $\mathbf{H}_M^T \mathbf{H}_M$ are different, but the matrix is symmetric about the main diagonal.

3 Related work

The goal of the receiver is to calculate the maximum likelihood (ML) estimate $\hat{\mathbf{x}}$ of \mathbf{x} set as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{S}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2 \tag{15}$$

However, its complexity is too high. In the past few decades, researchers have been studying various detectors to reduce their complexity while maintaining their SER.

3.1 OAMPNet

OAMPNet is a model-driven DL algorithm for MIMO detection derived from orthogonal approximate matching tracking (OAMP). Compared with approximate message passing (AMP), the advantage of OAMP is that it can be applied to unitary invariant matrices, while AMP is only applicable to Gaussian measurement matrices. OAMPNet has better performance than OAMP and can be adapted to various channel environments by using a number of learnable variables. The algorithm for OAMPNet is as follows.

Step 1: We need to design a linear detector. v_t^2 is the variance of the nonlinear estimation error

$$v_t^2 = \frac{\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t\|_2^2 - N \frac{\sigma^2}{2}}{\text{tr}(\mathbf{H}^T \mathbf{H})} \tag{16}$$

here \mathbf{W}_t is the optimal \mathbf{W} in OAMP in [35]

$$\hat{\mathbf{W}}_t = v_t^2 \mathbf{H}^T \left(v_t^2 \mathbf{H} \mathbf{H}^T + \frac{\sigma^2}{2} \right)^{-1} \tag{17}$$

$$\mathbf{W}_t = \frac{M}{\text{tr}(\hat{\mathbf{W}}_t \mathbf{H})} \hat{\mathbf{W}}_t \tag{18}$$

In this way, the value of the linear estimate can be obtained

$$\mathbf{r}_t = \hat{\mathbf{x}}_t + \gamma_t \mathbf{W}_t (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t) \tag{19}$$

Step 2: We take the linear detection estimates as input and perform nonlinear detection

$$\mathbf{C}_t = \mathbf{I} - \theta_t^2 \mathbf{W}_t \mathbf{H} \tag{20}$$

τ_t^2 is the variance of the linear estimation error

$$\tau_t^2 = \frac{1}{M} \text{tr}(\mathbf{C}_t \mathbf{C}_t^T) \nu_t^2 + \frac{\theta_t^2 \sigma^2}{M} \text{tr}(\mathbf{W}_t \mathbf{W}_t^T) \tag{21}$$

$p(x_i)$ is the prior probability

$$p(x_i) = \sum_{j \in \mathcal{M}} \frac{1}{\sqrt{M}} \delta(x_i - s_j) \tag{22}$$

Bring τ_t^2 and $p(x_i)$ into $E\{\mathbf{x} | \mathbf{r}_t, \tau_t\}$

$$E\{x_{ii} | r_{ii}, \tau_t\} = \frac{\sum_{s_i \in \mathcal{S}} s_i N_c(s_i; r_{ii}, \tau_t^2) p(s_i)}{\sum_{s_i \in \mathcal{S}} N_c(s_i; r_{ii}, \tau_t^2) p(s_i)} \tag{23}$$

In this way, the input value of the (t+1) layer can be obtained

$$\hat{\mathbf{x}}_{t+1} = E\{\mathbf{x} | \mathbf{r}_t, \tau_t\} \tag{24}$$

The performance of the OAMPNet detection algorithm is very good, and there are only two training parameters (γ_t, θ_t^2), but each layer of $\hat{\mathbf{W}}_t$ needs to be calculated once, and each calculation requires a pseudo-inverse, which brings great complexity, not suitable for massive MIMO, suitable for medium-scale MIMO [25, 26].

3.2 MMNet-iid

The main idea of MMNet-iid is to introduce an appropriate degree of flexibility in the linear and denoising components of the iterative framework while maintaining its linear plus non-linear structure [27, 28].

Step 1: We need to design a linear detector to estimate \mathbf{z}_t .

$$\mathbf{z}_t = \hat{\mathbf{x}}_t + \theta_t^{(1)} \mathbf{H}^T (\mathbf{y} - \mathbf{H} \hat{\mathbf{x}}_t) \tag{25}$$

Step 2: We take \mathbf{z}_t as input and perform nonlinear detection. σ_t^2 is the variance of the linear detection estimation error

$$\sigma_t^2 = \frac{\theta_t^{(2)}}{M} \left(\frac{\|\mathbf{I} - \mathbf{A}_t \mathbf{H}\|_F^2}{\|\mathbf{H}\|_F^2} \left[\|\mathbf{y} - \mathbf{H} \hat{\mathbf{x}}_t\|_2^2 - N_r \sigma^2 \right]_+ + \frac{\|\mathbf{A}_t\|_F^2}{\|\mathbf{H}\|_F^2} \sigma^2 \right) \tag{26}$$

$\eta_t(\mathbf{z}_t; \sigma_t^2)$ is a nonlinear detection estimate

$$\eta_t(z_{ii}; \sigma_t^2) = \frac{1}{Z} \sum_{s_i \in \mathcal{S}} s_i \exp\left(-\frac{\|z_{ii} - s_i\|^2}{\sigma_t^2}\right) \tag{27}$$

In this way, the input of the next layer can be obtained

$$\hat{\mathbf{x}}_{t+1} = \eta_t(\mathbf{z}_t; \sigma_t^2) \tag{28}$$

where $Z = \sum_{s_i \in S} \exp(-\frac{\|z_{ii} - s_i\|^2}{\sigma_t^2})$, $\mathbf{A}_t = \theta_t^{(1)} \mathbf{H}^T$. It has two training parameters $(\theta_t^{(1)}, \theta_t^{(2)})$, and its complexity is much smaller than OAMPNet. MMNet-iid performs well when the number of antennas is large and the linear Gaussian channel is good, and has poor performance in correlated channels or when the number of antennas is low.

4 The proposed BGS-Net method

4.1 Gauss-Seidel iterative method

GS is one of the common iterative methods used to solve systems of linear equations. If a system of linear equations $\mathbf{Ax} = \mathbf{b}$ is required to be solved, it will be decomposed as follows.

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i \quad (i = 1, 2, \dots, n) \tag{29}$$

The iterative formula of GS [36] is

$$x_i^{(k+1)} = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)})/a_{ii} \tag{30}$$

$(i = 1, 2, \dots, n; k = 0, 1, 2, \dots, t)$

For each layer, b_i is used to subtract the updated $\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)}$ and the not yet updated $\sum_{j=i+1}^n a_{ij}x_j^{(k)}$, its matrix representation is

$$\mathbf{x}^{(k+1)} = \mathbf{D}^{-1}(-\mathbf{Lx}^{(k+1)} - \mathbf{Ux}^{(k)} + \mathbf{b}) \tag{31}$$

$$(\mathbf{D} + \mathbf{L})\mathbf{x}^{(k+1)} = -\mathbf{Ux}^{(k)} + \mathbf{b} \tag{32}$$

$$\mathbf{x}^{(k+1)} = (\mathbf{D} + \mathbf{L})^{-1}(-\mathbf{Ux}^{(k)} + \mathbf{b}) \tag{33}$$

GS is used in communication, that is, the Hermitian positive semi-definite matrix \mathbf{A} is decomposed into strictly lower triangular terms \mathbf{L} , strictly upper triangular terms \mathbf{U} and diagonal terms \mathbf{D}

$$\mathbf{A} = \mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I} \tag{34}$$

$$\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{U} \tag{35}$$

Then the problem we solve is

$$\mathbf{Ax} = \mathbf{H}^T \mathbf{y} \tag{36}$$

Then solve a set of linear equations by calculating the solution of the iterative behavior [37].

$$\hat{\mathbf{x}}^{(n)} = (\mathbf{D} + \mathbf{L})^{-1}[\hat{\mathbf{x}}_{MF} - \mathbf{U}\hat{\mathbf{x}}^{(n-1)}] \tag{37}$$

where $\hat{\mathbf{x}}^{(n)}$ is the estimated signal, iteratively refined in each iteration and $\hat{\mathbf{x}}_{MF} = \mathbf{H}^T \mathbf{y}$, replacing \mathbf{b} in (32). Here $\hat{\mathbf{x}}^{(0)}$ is initialised to $\mathbf{D}^{-1} \mathbf{H}^T \mathbf{y}$. Gauss-Seidel has good convergence, and is guaranteed to converge when \mathbf{A} is diagonally dominant or symmetrically positive definite. This is because in MAUE systems, \mathbf{A} is not guaranteed to be diagonally dominant, but \mathbf{A} is definitely symmetric positive definite. The following is a proof of convergence of Gauss-Seidel for diagonally dominant or symmetric positive times, respectively [38].

Theorem 1 *When \mathbf{A} is diagonally dominant, Gauss-Seidel can guarantee convergence.*

Proof of Theorem 1. For strictly diagonally dominant matrix \mathbf{A} , its diagonal elements $a_{ii} \neq 0, i = 1, 2, \dots, n$, so

$$|\mathbf{D} + \mathbf{L}| = \prod_{i=1}^n a_{ii} \neq 0 \tag{38}$$

Suppose $\mathbf{B}_G = -(\mathbf{D} + \mathbf{L})^{-1} \mathbf{U}$, the characteristic value is λ , then the characteristic equation is

$$\begin{aligned} |\lambda \mathbf{I} - \mathbf{B}_G| &= \left| \lambda \mathbf{I} + (\mathbf{D} + \mathbf{L})^{-1} \mathbf{U} \right| = \left| (\mathbf{D} + \mathbf{L})^{-1} \right| |\lambda(\mathbf{D} + \mathbf{L}) + \mathbf{U}| = 0 \\ &\Rightarrow |\lambda(\mathbf{D} + \mathbf{L}) + \mathbf{U}| = 0 \end{aligned} \tag{39}$$

When the determinant is zero, the equation has a non-zero solution. Use contradiction: suppose $|\lambda| \geq 1$,

$$\lambda(\mathbf{D} + \mathbf{L}) + \mathbf{U} = \begin{bmatrix} \lambda a_{11} & a_{12} & \dots & a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & a_{2n} \\ \vdots & \dots & \ddots & \vdots \\ \lambda a_{n1} & \lambda a_{n2} & \dots & \lambda a_{nn} \end{bmatrix} \tag{40}$$

It is a strictly diagonally dominant matrix, so it is non-singular, that is, $|\lambda(\mathbf{D} + \mathbf{L}) + \mathbf{U}| \neq 0$ contradicts the eigenvalue λ satisfying $|\lambda(\mathbf{D} + \mathbf{L}) + \mathbf{U}| = 0$. So $|\lambda| < 1$ is $\rho(\mathbf{B}_G) < 1$, when \mathbf{A} is diagonally dominant, Gauss-Seidel converges.

Theorem 2 *When \mathbf{A} is symmetrically positive, Gauss-Seidel can guarantee convergence.*

Proof of Theorem 2. Suppose $\mathbf{B}_G = -(\mathbf{D} + \mathbf{L})^{-1} \mathbf{U}$, the eigenvalue is λ , and \mathbf{x} is the eigenvector, then

$$\begin{aligned} -(\mathbf{D} + \mathbf{L})^{-1} \mathbf{U} \mathbf{x} &= \lambda \mathbf{x} \\ \rightarrow -\mathbf{U} \mathbf{x} &= \lambda(\mathbf{D} + \mathbf{L}) \mathbf{x} \\ \rightarrow -\mathbf{x}^T \mathbf{U} \mathbf{x} &= \lambda \mathbf{x}^T (\mathbf{D} + \mathbf{L}) \mathbf{x} \end{aligned} \tag{41}$$

Because \mathbf{A} is positive definite, so $p = \mathbf{x}^T \mathbf{D} \mathbf{x} > 0$, set $-\mathbf{x}^T \mathbf{U} \mathbf{x} = a$, then

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T (\mathbf{D} + \mathbf{L} + \mathbf{U}) \mathbf{x} = p - 2a > 0 \tag{42}$$

$$\lambda = \frac{-\mathbf{x}^T \mathbf{U} \mathbf{x}}{\mathbf{x}^T (\mathbf{D} + \mathbf{L}) \mathbf{x}} = \frac{a}{p - a} \tag{43}$$

$$\lambda^2 = \frac{a^2}{p^2 - 2pa + a^2} = \frac{a^2}{p(p - 2a) + a^2} < 1 \tag{44}$$

So $|\lambda| < 1$ is $\rho(\mathbf{B}_G) < 1$. When \mathbf{A} is symmetric positive definite, Gauss-Seidel converges, and we divide the numerator and denominator of (44) by a^2 at the same time, we get

$$\lambda^2 = \frac{1}{\left(\frac{p}{a} - 1\right)^2} \tag{45}$$

From (45) we can see that the more diagonally dominant, the smaller the λ^2 , the faster the convergence.

4.2 BGS-Net architecture

In this section, a model-driven DL detector network (called BGS-Net) is proposed. The signal detector uses the Gauss-Seidel method and nonlinear activation to improve the detection performance. The only training parameters is $\Omega = \gamma_t, \gamma_t \in \mathbb{R}^{M \times 1}$. In the algorithm $(\mathbf{D} + \mathbf{L})^{-1}$, $\hat{\mathbf{x}}_{MF}$, \mathbf{U} , $tr(\mathbf{H}^T \mathbf{H})$, $\frac{1}{M} tr(\mathbf{C}_t \mathbf{C}_t^T)$, $\frac{\sigma^2}{M} tr(\mathbf{W}_t \mathbf{W}_t^T)$, all of which need to be computed only once and then reused at each layer. In contrast, \mathbf{W}_t and \mathbf{A}_t for OAMPNet and MMNet-iid need to be calculated once per layer because of the training parameters present in them. The structure of BGS-Net is shown in Figure 6, which is an improved algorithm by adding a learnable vector variable γ_t . The network consists of L_{layer} cascaded layers, each of which has the same structure, including nonlinear estimator, error variance τ_t^2 , and tied weights. The input of the BGS-Net network is $\hat{\mathbf{x}}_{MF}$ and the initial value $\hat{\mathbf{x}}_0$, and the output is the final estimate of the signal $\hat{\mathbf{x}}_{L_{layer}}$. To make it easier to see the deep learning structure, see Figure 7. We first calculate $\hat{\mathbf{z}}_t$ and scalar τ_t^2 through GS detection block, plus the constellation map S as the input of the network. We introduce a

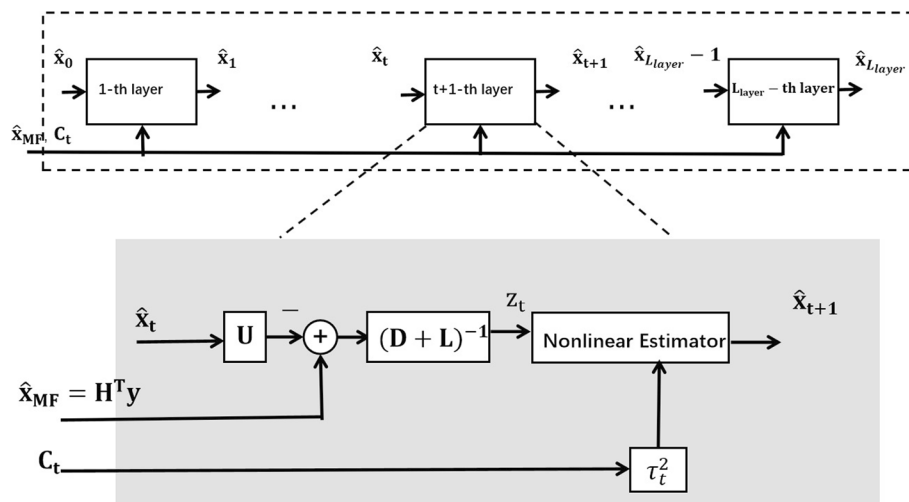


Fig. 6 Structure of BGS-Net

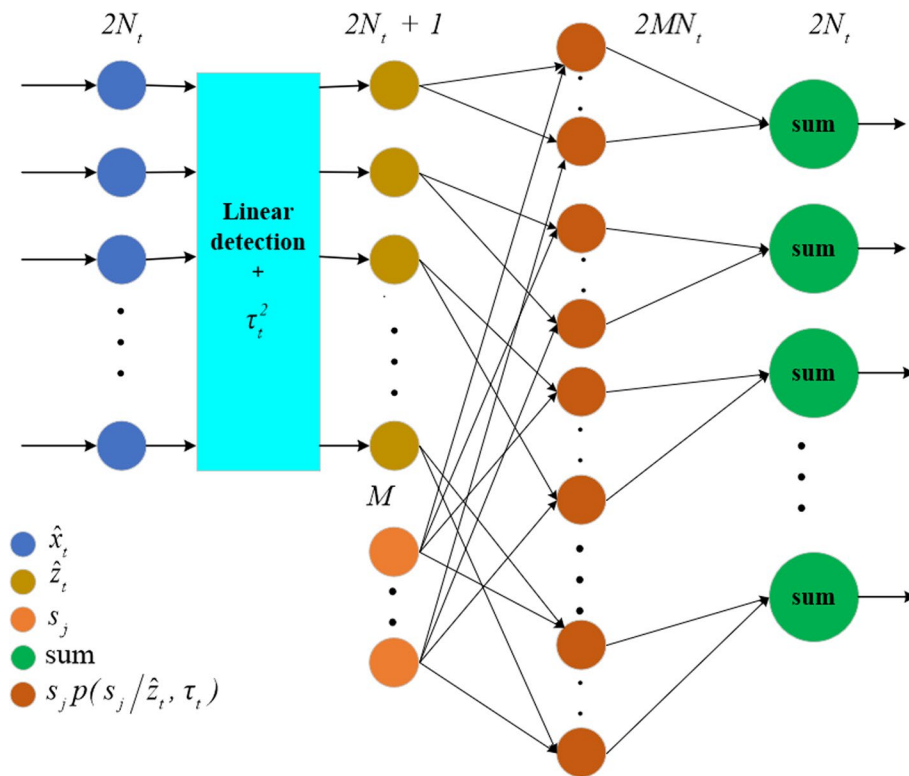


Fig. 7 Deep learning network neuron structure of BGS-Net

vector variable γ_t in nonlinear detection, and finally output $\hat{\mathbf{x}}_{t+1}$. The difference between model-driven and DNN is that many of the parameters of model-driven are fixed values obtained from past experience, while the parameters of DNN are all variable values.

Algorithm1: BGS-Net algorithm for MIMO detection

Input: Received signal \mathbf{y} , channel matrix \mathbf{H} , noise level $\sigma^2/2$

Initialize: $\hat{\mathbf{x}}_0 \leftarrow \mathbf{D}^{-1}\mathbf{H}^T\mathbf{y}$

$$1. \hat{\mathbf{z}}_t = (\mathbf{D} + \mathbf{L})^{-1}[\hat{\mathbf{x}}_{MF} - \mathbf{U}\hat{\mathbf{x}}_t]$$

$$2. v_t^2 = \frac{\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t\|_2^2 - N\frac{\sigma^2}{2}}{\text{tr}(\mathbf{H}^T\mathbf{H})}$$

$$3. v_t^2 = \max(v_t^2, 10^{-9})$$

$$4. \tau_t^2 = \frac{1}{M}\text{tr}(\mathbf{C}_t\mathbf{C}_t^T)v_t^2 + \frac{\sigma^2}{M}\text{tr}(\mathbf{W}_t\mathbf{W}_t^T)$$

$$5. \tau_t^2 = \frac{\tau_t^2}{\gamma_t}$$

$$6. \hat{\mathbf{x}}_{t+1} = E\{\mathbf{x}|\hat{\mathbf{z}}_t, \tau_t\}$$

where

$$\mathbf{W}_t = \mathbf{H}^T \tag{46}$$

$$\mathbf{C}_t = \mathbf{I} - \mathbf{W}_t\mathbf{H} \tag{47}$$

$$\begin{aligned}
 E\{x_{ti}|\hat{z}_{ti}, \tau_{ti}\} &= \sum_{s_j \in \mathcal{S}} s_j \times p(s_j|\hat{z}_{ti}, \tau_{ti}) \\
 &= \sum_{s_j \in \mathcal{S}} s_j \times \text{softmax}\left(\frac{-\|\hat{z}_{ti}-s_j\|^2}{\tau_{ti}^2}\right)
 \end{aligned}
 \tag{48}$$

where $\text{softmax}(V_i) = \frac{\exp^{V_i}}{\sum_j \exp^{V_j}}$. It can be seen from Algorithm1 that we only have one training parameter per layer, and vector γ_t is used to adjust the variance τ_t^2 which is an estimated value. Because $\frac{1}{M} \text{tr}(\mathbf{C}_t \mathbf{C}_t^T)$ used is a constant value, it is multiplied by \mathbf{v}_t^2 each time in τ_t^2 , which saves a large amount of calculation. What we need to pay attention to is that when $4 \rightarrow 5$, τ_t^2 's dimension expansion to a vector τ_t^2 .

4.3 Low-complexity algorithm for $(\mathbf{D} + \mathbf{L})^{-1}$

In this section, the complexity of $(\mathbf{D} + \mathbf{L})^{-1}$ will be reduced. The complexity in calculating $\hat{\mathbf{x}}_t = (\mathbf{D} + \mathbf{L})^{-1}[\hat{\mathbf{x}}_{MF} - \mathbf{U}\hat{\mathbf{x}}_{t-1}]$ is mainly concentrated in solving $(\mathbf{D} + \mathbf{L})^{-1}$. If the inverse is solved directly, the complexity of the algorithm will reach $O(M^3)$, so a circular nesting method is proposed to reduce its complexity, as described below :

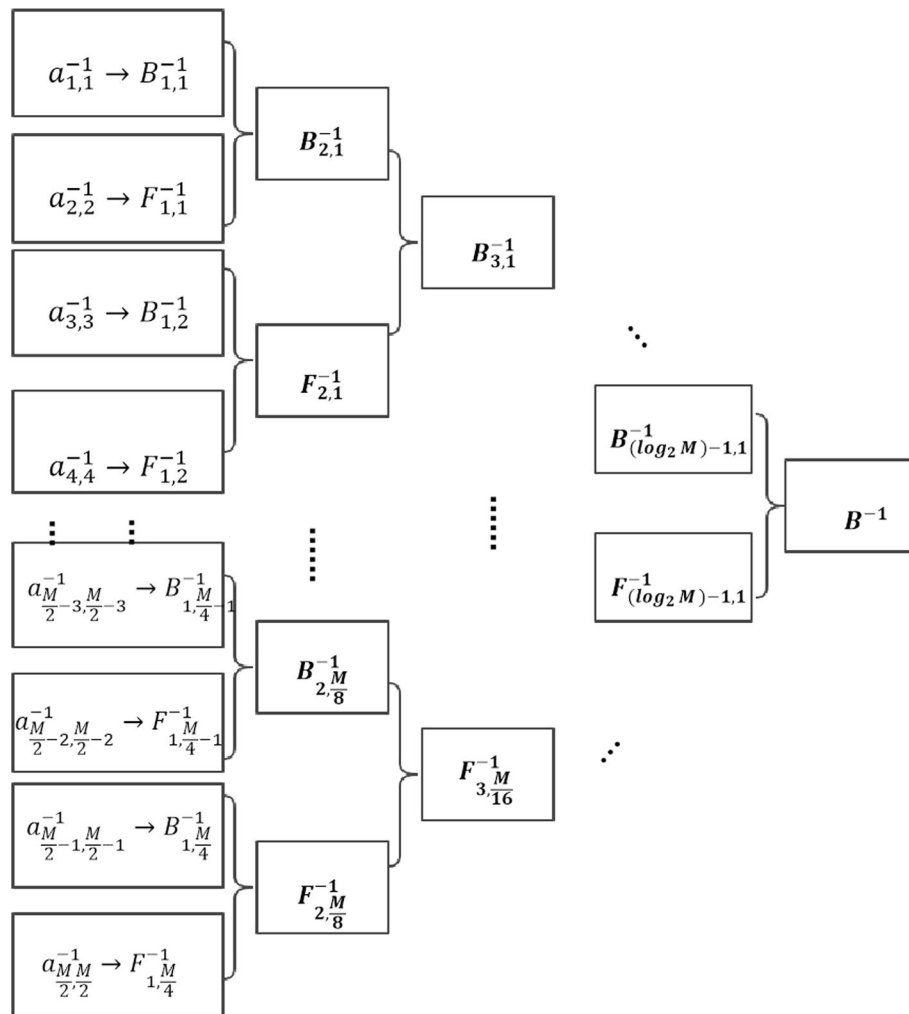


Fig. 8 Matrix solution for $(\mathbf{D} + \mathbf{L})^{-1}$

The first: From Eq. (30), the complexity of each row is $M - 1$ multiplication, 1 division, for a total of M rows, so the complexity of one iteration is M^2 . However, this method is not applicable to BGS-Net.

The second: Solve the lower triangular matrix in parallel, the structure is shown in Figure 8. For the inversion of the lower triangular matrix, we have the following properties:

$$Y = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{C} & \mathbf{F} \end{bmatrix} \rightarrow Y^{-1} = \begin{bmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ -\mathbf{F}^{-1}\mathbf{C}\mathbf{B}^{-1} & \mathbf{F}^{-1} \end{bmatrix} \tag{49}$$

where $\mathbf{B}, \mathbf{C}, \mathbf{F}$ have the same size. The main complexity of (37) is to solve $(\mathbf{D} + \mathbf{L})^{-1}$, which we know to be a lower triangular matrix, and use the above property for its inverse. It can be known from (8),(10) and (14) that when the system is SAUE or MAUE (only \mathbf{T}), it has the following (50) properties; when MAUE (both \mathbf{T} and \mathbf{R}), there is no (50) property:

$$= \begin{bmatrix} a_{1,1} & 0 & \dots & 0 & 0 \\ a_{2,1} & a_{2,2} & \dots & 0 & 0 \\ \vdots & \dots & \ddots & \dots & \vdots \\ a_{\frac{M}{2}-1,1} & a_{\frac{M}{2}-1,2} & \dots & a_{\frac{M}{2}-1,\frac{M}{2}-1} & 0 \\ a_{\frac{M}{2},1} & a_{\frac{M}{2},2} & \dots & a_{\frac{M}{2},\frac{M}{2}-1} & a_{\frac{M}{2},\frac{M}{2}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{\frac{M}{2}+1,\frac{M}{2}+1} & 0 & \dots & 0 & 0 \\ a_{\frac{M}{2}+2,\frac{M}{2}+1} & a_{\frac{M}{2}+2,\frac{M}{2}+2} & \dots & 0 & 0 \\ \vdots & \dots & \ddots & \dots & \vdots \\ a_{M-1,\frac{M}{2}+1} & a_{M-1,\frac{M}{2}+2} & \dots & a_{M-1,M-1} & 0 \\ a_{M,\frac{M}{2}+1} & a_{M,\frac{M}{2}+2} & \dots & a_{M,M-1} & a_{M,M} \end{bmatrix} \tag{50}$$

We can see from Figure 8 that the specific steps of the loop nesting method are as follows.

Step 1: Find the reciprocals of $a_{1,1}, a_{2,2}, a_{3,3}, \dots, a_{\frac{M}{2},\frac{M}{2}}$ and assign them to $\mathbf{B}_{1,t}^{-1}$ and $\mathbf{F}_{1,t}^{-1}$, $t \in (1, \frac{M}{4})$ respectively.

Step 2: Bring the resulting $\mathbf{B}_{i,t}^{-1}$ and $\mathbf{F}_{i,t}^{-1}$, and the corresponding $\mathbf{C}_{i,t}$, $t \in (1, \frac{M}{2^{i+1}})$ into (49), we can get $\mathbf{B}_{i+1,t}^{-1}$ and $\mathbf{F}_{i+1,t}^{-1}$, $t \in (1, \frac{M}{2^{i+2}})$. If $\mathbf{B}_{i+1,t}^{-1}$ is an $\frac{M}{2} \times \frac{M}{2}$ matrix, then the next step, otherwise loop the second step.

*Step 3:*In the case of Section 2.3 (a) (b), assign $\mathbf{B}^{-1} = \mathbf{B}_{i+1,t}^{-1}$ to \mathbf{F}^{-1} ; otherwise, the same method as for \mathbf{B}^{-1} , solve \mathbf{F}^{-1} . In this way, we can obtain $(\mathbf{D} + \mathbf{L})^{-1} = \begin{bmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ -\mathbf{F}^{-1}\mathbf{C}\mathbf{B}^{-1} & \mathbf{F}^{-1} \end{bmatrix}$

Note that we don't need to find the value of $(\mathbf{D} + \mathbf{L})^{-1}$, just take the following formula to solve the linear detection term. (Because $\mathbf{P}_1 \in \mathbb{R}^{P \times Q}$, $\mathbf{P}_2 \in \mathbb{R}^{Q \times K}$, $\mathbf{b} \in \mathbb{R}^{K \times 1}$, we know $(\mathbf{P}_1\mathbf{P}_2)\mathbf{b} = \mathbf{P}_1(\mathbf{P}_2\mathbf{b})$. Let $\mathbf{b} = \hat{\mathbf{x}}_{MF} - \mathbf{U}\hat{\mathbf{x}}_t = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$, where $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^{\frac{M}{2} \times 1}$. So

$$\begin{aligned} (\mathbf{D} + \mathbf{L})^{-1}[\hat{\mathbf{x}}_{MF} - \mathbf{U}\hat{\mathbf{x}}_t] &= \begin{bmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ -\mathbf{F}^{-1}\mathbf{C}\mathbf{B}^{-1} & \mathbf{F}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{B}^{-1}\mathbf{c}_1 \\ -\mathbf{F}^{-1}\mathbf{C}\mathbf{B}^{-1}\mathbf{c}_1 + \mathbf{F}^{-1}\mathbf{c}_2 \end{bmatrix} \end{aligned} \tag{51}$$

4.4 Error analysis

In this section, we will study the reasons why BGS-Net can improve performance. The analysis of the error ($\hat{\mathbf{x}}_t - \mathbf{x}$) is as follows:

Define the output error $\mathbf{e}_t^{lin} = \mathbf{z}_t - \mathbf{x}$ for the linear phase at iteration t and the output error at the previous iteration $t - 1$ as $\mathbf{e}_{t-1}^{den} = \hat{\mathbf{x}}_t - \mathbf{x}$. We can rewrite the update equation of Algorithm 1 based on these two output errors as:

$$\begin{aligned}
 \mathbf{e}_t^{lin} &= (\mathbf{D} + \mathbf{L})^{-1} \mathbf{H}^T \mathbf{y} - (\mathbf{D} + \mathbf{L})^{-1} \mathbf{U} \hat{\mathbf{x}}_t - \mathbf{x} \\
 &= (\mathbf{D} + \mathbf{L})^{-1} \mathbf{H}^T (\mathbf{H} \mathbf{x} + \mathbf{n}) - (\mathbf{D} + \mathbf{L})^{-1} \mathbf{U} \hat{\mathbf{x}}_t - \mathbf{x} \\
 &= (\mathbf{D} + \mathbf{L})^{-1} \mathbf{H}^T (\mathbf{H} \mathbf{x} + \mathbf{n}) - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{A} - \mathbf{D} - \mathbf{L}) \hat{\mathbf{x}}_t - \mathbf{x} \\
 &= \mathbf{e}_{t-1}^{den} + (\mathbf{D} + \mathbf{L})^{-1} \mathbf{H}^T (\mathbf{H} \mathbf{x} + \mathbf{n}) - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I}) \hat{\mathbf{x}}_t \\
 &= (\mathbf{I} - (\mathbf{D} + \mathbf{L})^{-1} \mathbf{H}^T \mathbf{H}) \mathbf{e}_{t-1}^{den} + (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{n} - \frac{\sigma^2}{2} \hat{\mathbf{x}}_t) \\
 &= (\mathbf{I} - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I})) \mathbf{e}_{t-1}^{den} + (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{n} - \frac{\sigma^2}{2} \mathbf{x})
 \end{aligned} \tag{52}$$

and

$$\mathbf{e}_t^{den} = E\{\mathbf{x} | \hat{\mathbf{z}}_t, \tau_t\} - \mathbf{x} \tag{53}$$

From Figure 2, we know that under channel hardening conditions, the first term of equation (52) $(\mathbf{I} - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I}))$ tends to 0; the second term is divided into the effect of \mathbf{n} and the effect of \mathbf{x} , for the noise, where $\mathbf{n} = \sqrt{\sigma^2/2} * N(0, 1)$, so when the signal-to-noise ratio(SNR) is small, $\mathbf{H}^T \mathbf{n}$ becomes larger, $\frac{\sigma^2}{2} \mathbf{x}$ also becomes larger; when the SNR is large, $\mathbf{H}^T \mathbf{n}$ becomes smaller and $\frac{\sigma^2}{2} \mathbf{x}$ also becomes smaller. And $(\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{n} - \frac{\sigma^2}{2} \mathbf{x})$ can in turn be reduced to $(\mathbf{D} + \mathbf{L})^{-1} \mathbf{H}^T \mathbf{y} - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{H} \mathbf{x} + \frac{\sigma^2}{2} \mathbf{x}) = (\mathbf{D} + \mathbf{L})^{-1} \mathbf{H}^T \mathbf{y} - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I}) \mathbf{x}$

which is approximated under good channel hardening as $\mathbf{x}_{MMSE} - \mathbf{x}$, and as far as we know, the gap between $\mathbf{x}_{MMSE} - \mathbf{x}$ decreases as the channel hardens more. Under channel hardening conditions the error \mathbf{e}_{t-1}^{den} from the previous stage, suppressed by $(\mathbf{I} - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I}))$, is significantly attenuated. These calculations explain why BGS-Net has good performance on i.i.d Gaussian channels. Moreover, it is better than MMNet-iid's $\mathbf{I} - \theta_t^{(1)} \mathbf{H}^T \mathbf{H}$ on correlated channels, where channel hardening disappears when the channel is correlated and there is no way for $\mathbf{I} - \theta_t^{(1)} \mathbf{H}^T \mathbf{H}$ to converge to $\mathbf{0}$ as the number of antennas increases, while $\mathbf{I} - (\mathbf{D} + \mathbf{L})^{-1} (\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I})$, since \mathbf{A} is symmetric and $\mathbf{D} + \mathbf{L}$ itself contains all the information in \mathbf{A} . When the number of antennas increases, it can be approximated as $\mathbf{I} - \mathbf{A}^{-1} \mathbf{A}$, tending to $\mathbf{0}$ but not to $\mathbf{0}$.

For the effect of the nonlinear activation function, $E\{\mathbf{x} | \hat{\mathbf{z}}_t, \tau_t\} - \mathbf{x}$ in (53) reduces the difference of $\hat{\mathbf{x}}_{t+1} - \mathbf{x}$. The proof is as follows:

$$E\{x_{ti} | \hat{z}_{ti}, \tau_{ti}\} - x_{ti} = \sum_{s_j \in S} s_j \times \text{softmax}(\frac{-\|\hat{z}_{ti} - s_j\|^2}{\tau_{ti}^2}) - x_{ti} \tag{54}$$

Assuming that the true value x_{ti} is s_1 , then the above formula is equal to

$$\sum_{s_j \in S} s_j \times p(s_j / \hat{z}_{ti}, \tau_{ti}) - s_1 \tag{55}$$

We know that the softmax soft decision uses an exponent that allows judgments with larger probabilities to become larger and smaller probabilities to become smaller, but the total probability is still 1. It can be seen that as the probability of judging s_1 increases, the first term of equation (55) gets closer and closer to s_1 , so using this activation function can further reduce the error.

5 Improved BGS-Net method

5.1 Analysis of the problem

Under the MAUE system, \mathbf{A} cannot be approximated as a diagonal matrix \mathbf{D} , which has a great influence on $\hat{\mathbf{x}}_0 \leftarrow \mathbf{D}^{-1} \mathbf{H}^T \mathbf{y}$. $\hat{\mathbf{x}}_0$ is a initial solution. If $\hat{\mathbf{x}}_0$ is given well, then the number of iterations required is small. In the SAUE system, the value of $\hat{\mathbf{x}}_0 \leftarrow \mathbf{D}^{-1} \mathbf{H}^T \mathbf{y}$ is approximately equal to $(\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{2} \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y}$, so the number of iterations is small, which can also explain why fast iterations converge under channel hardening conditions. In the MAUE system, $\mathbf{H}^T \mathbf{H}$ loses the diagonal dominance, and the sum of the other elements in the same line is no longer much smaller than the diagonal elements. $\mathbf{D}^{-1} \mathbf{H}^T \mathbf{y}$ cannot approach the real solution \mathbf{x} , here $\hat{\mathbf{x}}_0 \leftarrow \mathbf{A}^{-1} \mathbf{H}^T \mathbf{y}$ replaces $\hat{\mathbf{x}}_0 \leftarrow \mathbf{D}^{-1} \mathbf{H}^T \mathbf{y}$, so that no matter how the channel changes, a good initial solution can be extracted.

However, calculating \mathbf{A}^{-1} has a high complexity, so the low-complexity method is used to approximate the solution of \mathbf{A}^{-1} .

5.2 Improved BGS-Net Design

In this section, BGS-Net is improved to adapt to the MAUE system by replacing $\hat{\mathbf{x}}_0 \leftarrow \mathbf{D}^{-1} \mathbf{H}^T \mathbf{y}$ with $\hat{\mathbf{x}}_0 \leftarrow \mathbf{A}^{-1} \mathbf{H}^T \mathbf{y}$.

5.2.1 Approximation of \mathbf{A}^{-1}

From Figures 3, 4 and 5, it can be seen that as the correlation coefficient increases, the channel is block-like in character. To approximate \mathbf{A}^{-1} , as in Figure 9, divide the diagonal of matrix \mathbf{A} into $\frac{M}{m_{UE}}$ small matrices, the size of the small matrix is $m_{UE} \times m_{UE}$, each

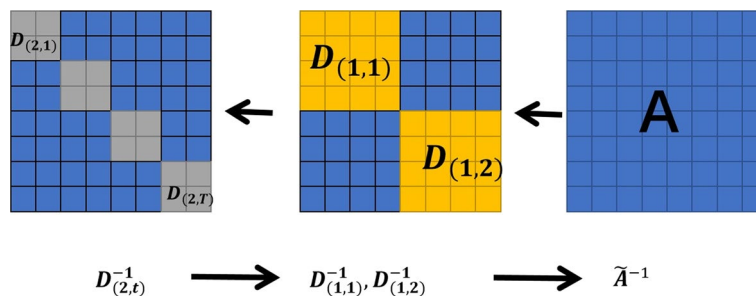


Fig. 9 Block matrix approximation of \mathbf{A}

small matrix in order is set to $\mathbf{D}_{(2,t)} \in \mathbb{R}^{m_{UE} \times m_{UE}}$, ($t \in [1 : T]$) respectively. The matrix \mathbf{A} is again divided into 4 matrices, with the upper left and lower right matrices set to $\mathbf{D}_{(1,1)}, \mathbf{D}_{(1,2)} \in \mathbb{R}^{M/2 \times M/2}$ respectively. To ensure convergence of the Neumann Series, here, unlike in [32], the parameter $\alpha_{opt} = 1 + \eta$, $\eta = M/N$ was introduced [39]. We do the following for all small matrices $\mathbf{D}_{(2,t)}$ as $\mathbf{D}_{(2,t)} = \mathbf{D}_{(2,t)} \times \alpha_{opt}$ and find the inverse of each small matrix $\mathbf{D}_{(2,t)}^{-1}$. In order to get $\mathbf{D}_{(1,1)}^{-1}$, we don't want to find it directly, because its complexity is $O(M^3/4)$, at this time $\mathbf{D}_{(1,1)}^{-1}$ has both diagonal elements and non-diagonal elements, unlike only diagonal elements under channel hardening conditions. In this paper, the Neumann Series is used to approximate it [32, 40] by the following method:

$$\mathbf{N} = \text{diag}[\mathbf{D}_{2,1}^{-1}, \mathbf{D}_{2,2}^{-1}, \dots, \mathbf{D}_{2, \frac{T}{2}}^{-1}] \tag{56}$$

$$\tilde{\mathbf{D}}_{1,1} = \text{diag}[\mathbf{D}_{2,1}, \mathbf{D}_{2,2}, \dots, \mathbf{D}_{2, \frac{T}{2}}] \tag{57}$$

use \mathbf{E} to approximate the off-diagonal block part of $\mathbf{D}_{(1,1)}$

$$\mathbf{E} = \tilde{\mathbf{D}}_{1,1} - \mathbf{D}_{(1,1)} \tag{58}$$

and use \mathbf{N} to approximate $\mathbf{D}_{(1,1)}^{-1}$

$$\mathbf{D}_{(1,1)}^{-1} = \mathbf{N} \tag{59}$$

So we can approximate $\mathbf{D}_{(1,1)}^{-1}$ by k_N times Neumann method.

$$k_N \text{ times} : \mathbf{D}_{(1,1)}^{-1} = \mathbf{NED}_{(1,1)}^{-1} + \mathbf{N} \tag{60}$$

As above, we can obtain $\mathbf{D}_{(1,2)}^{-1}$, and then splice $\mathbf{D}_{(1,1)}^{-1}$ and $\mathbf{D}_{(1,2)}^{-1}$ together

$$\mathbf{D}^{-1} = \text{diag}[\mathbf{D}_{(1,1)}^{-1}, \mathbf{D}_{(1,2)}^{-1}] \tag{61}$$

In this way we can get the required \mathbf{D}^{-1} . The same can be obtained:

$$\mathbf{E} = \mathbf{D}^{-1} - \mathbf{A} \tag{62}$$

$$\tilde{\mathbf{A}}^{-1} = \mathbf{D}^{-1} \tag{63}$$

$$k_F \text{ times} : \tilde{\mathbf{A}}^{-1} = \mathbf{D}^{-1}\mathbf{E}\tilde{\mathbf{A}}^{-1} + \mathbf{D}^{-1} \tag{64}$$

In this way, we get $\tilde{\mathbf{A}}^{-1}$. Considering the high complexity of Eq. (64), we rewrite Eq. (64) without changing its principle [41]

$$\mathbf{S}_t = \vartheta \mathbf{S}_{t-1} + \mathbf{S}_0 \quad (k_F \geq t \geq 1) \tag{65}$$

where $\mathbf{S}_0 = \mathbf{D}^{-1}(\mathbf{H}^T \mathbf{y})$, $\vartheta = \mathbf{D}^{-1}\mathbf{E}$. In this way, we can bypass the high complexity of solving Eq. (64) and directly obtain $\hat{\mathbf{x}}_0 \leftarrow \mathbf{A}^{-1}\mathbf{H}^T \mathbf{y}$ with low complexity.

5.2.2 Offline training algorithm

A relatively good initial solution is obtained from the approximation of \mathbf{A} in 5.2.1.

Algorithm2: Improved BGS-Net offline training

Input: Received signal \mathbf{y} , channel matrix \mathbf{H} , noise level $\sigma^2/2$

Initialize: $\hat{\mathbf{x}}_0 \leftarrow \bar{\mathbf{A}}^{-1} \mathbf{H}^T \mathbf{y}$

$$1. \hat{\mathbf{z}}_t = (\mathbf{D} + \mathbf{L})^{-1} [\hat{\mathbf{x}}_{MF} - \mathbf{U} \hat{\mathbf{x}}_t]$$

$$2. v_t^2 = \frac{\|\mathbf{y} - \mathbf{H} \hat{\mathbf{x}}_t\|_2^2 - N \frac{\sigma^2}{2}}{\text{tr}(\mathbf{H}^T \mathbf{H})}$$

$$3. v_t^2 = \max(v_t^2, 10^{-9})$$

$$4. \tau_t^2 = \frac{1}{M} \text{tr}(\mathbf{C}_t \mathbf{C}_t^T) v_t^2 + \frac{\sigma^2}{M} \text{tr}(\mathbf{W}_t \mathbf{W}_t^T)$$

$$5. \tau_t^2 = \frac{\tau_t^2}{\gamma}$$

$$6. \hat{\mathbf{x}}_{t+1} = E\{\mathbf{x} | \hat{\mathbf{z}}_t, \tau_t\}$$

6 Complexity analysis

6.1 Complexity of $(\mathbf{D} + \mathbf{L})^{-1}$

In this section, the complexity of (37) is analysed. $\hat{\mathbf{x}}_t = (\mathbf{D} + \mathbf{L})^{-1} [\hat{\mathbf{x}}_{MF} - \mathbf{U} \hat{\mathbf{x}}_{t-1}]$ in which the complexity is mainly solving $(\mathbf{D} + \mathbf{L})^{-1}$. If the inverse is solved directly, the complexity of the algorithm will reach $O(M^3)$, so we use a nested loop method to reduce its complexity.

So the complexity calculation formula of $(\mathbf{D} + \mathbf{L})^{-1}$ is:

(a) When the system is SAUE or MAUE (only \mathbf{T})

$$\begin{aligned} & \frac{M}{2} + (2 \times (2^0)^3 \times \frac{M}{8} \times 2 + 2 \times (2^1)^3 \times \frac{M}{16} \times 2 + \dots \\ & + 2 \times (2^{\log_2 M - 3})^3 \times \frac{M}{2^{\log_2 M}} \times 2) + \frac{M^3}{32} = \frac{1}{24} M^3 + \frac{1}{3} M \end{aligned} \tag{66}$$

Because $(\mathbf{D} + \mathbf{L})^{-1}$ only needs to be calculated once, the complexity of formula (37) is $\frac{1}{24} M^3 + \frac{3}{4} t M^2 + (\frac{1}{3} + t) M$.

(b) When MAUE (both \mathbf{T} and \mathbf{R})

$$\begin{aligned} & 2 \times [\frac{M}{2} + (2 \times (2^0)^3 \times \frac{M}{8} \times 2 + 2 \times (2^1)^3 \times \frac{M}{16} \times 2 + \dots \\ & + 2 \times (2^{\log_2 M - 3})^3 \times \frac{M}{2^{\log_2 M}} \times 2) + \frac{M^3}{32}] = \frac{1}{12} M^3 + \frac{2}{3} M \end{aligned} \tag{67}$$

So the complexity of formula (37) is $\frac{1}{12} M^3 + \frac{3}{4} t M^2 + (\frac{2}{3} + t) M$.

6.2 Complexity of $\hat{\mathbf{x}}_0$

In this section, the complexity analysis of $\hat{\mathbf{x}}_0$ is carried out. The initial solutions of Improved BGS-Net and BGS-Net are different:

(1) The complexity of $\hat{\mathbf{x}}_0 \leftarrow \mathbf{D}^{-1} \mathbf{H}^T \mathbf{y}$ is: $MN + M$.

Table 1 Computational complexity comparison

Algorithm	Number of multiplications
Cholesky MMSE [42]	$O(2M^3 + (N + M)M)$
OAMPNet [25]	$O(2tN^3)$
MMNet-iid [27]	$O(2tMN + tM)$
TL-BD-INSA [32]	$O(\frac{k_N+k_E}{4}M^3 + \frac{k_N-2}{2}m_{UE}M^2 + (N + M + 2m_{UE}^2)M)$
BGS-Net	$\begin{cases} a) & O(\frac{1}{24}M^3 + \frac{3}{4}tM^2 + MN + M) \\ b) & O(\frac{1}{12}M^3 + \frac{3}{4}tM^2 + MN + M) \end{cases}$
Improved BGS-Net	$\begin{cases} a) & O(\frac{4+3k_N}{24}M^3 + \frac{4k_F+(k_N-2)m_{UE}+2+3t}{4}M^2 + (N + m_{UE}^2 + t)M) \\ b) & O(\frac{1+3k_N}{12}M^3 + \frac{4k_F+2(k_N-2)m_{UE}+2+3t}{4}M^2 + (N + 2m_{UE}^2 + t)M) \end{cases}$

(2) The complexity of $\hat{\mathbf{x}}_0 \leftarrow \mathbf{A}^{-1}\mathbf{H}^T\mathbf{y}$ is: From [32] we can see that the complexity of (60) is $\frac{k_N-1}{8}M^3 + \frac{k_N-2}{4}m_{UE}M^2 + Mm_{UE}^2$, and the complexity of ϑ is $M^2\frac{M}{2} - M(\frac{M}{2})^2 = \frac{1}{4}M^3$, then the complexity of equation (64) becomes $\frac{1}{4}M^3 + (k_F + \frac{1}{2})M^2 + MN$.

- (a) When the system is SAUE or MAUE (only **T**): $\frac{k_N+1}{8}M^3 + (\frac{k_N-2}{4}m_{UE} + k_F + \frac{1}{2})M^2 + Mm_{UE}^2 + MN$
- (b) When MAUE (both **T** and **R**): $\frac{k_N}{4}M^3 + (\frac{k_N-2}{2}m_{UE} + k_F + \frac{1}{2})M^2 + 2Mm_{UE}^2 + MN$

6.3 Complexity comparison

This section will analyze the computational complexity according to the number of multiplications of different algorithms. Here $N_r = 256$, $N_t = 32$, $k_N = k_F = 2$, $m_{UE} = 4$. From Table 1, we can know that in the environment of a (when the system is SAUE or MAUE (only **T**)), the complexity of BGS-Net is about 16% of MMSE, 0.01% of OAMPNet, 30% of TL-BD-INSA, and 20% of MMNet-iid; the complexity of Improved BGS-Net is about 38% of MMSE, 0.02% of OAMPNet, 42% of TL-BD-INSA, and 65% of MMNet-iid. In the environment of b (when MAUE (with both **T** and **R**)), the complexity of BGS-Net is approximately 19% of MMSE, 0.01% of OAMPNet, 32% of TL-BD-INSA, and 21% of MMNet-iid; the complexity of Improved BGS-Net is about 46% of MMSE, 0.02% of OAMPNet, 80% of TL-BD-INSA, and 52% of MMNet-iid. So our algorithm complexity is very low. From [28] we can know that the complexity of MMNet-iid nonlinear detection is $O(M^2)$, while the complexity of BGS-Net’s nonlinear detection is much smaller than that of MMNet-iid nonlinear detection, and the complexity of nonlinear detection of MMNet-iid is much smaller than that of linear detection, so we ignore its complexity.

7 Numerical results and discussion

In this section we give simulation results for MIMO detection of BGS-Net and Improved BGS-Net, and evaluate the performance by the symbol error rate(SER) for different signal-to-noise ratios(SNR). The SNR of the system, defined as

$$\text{SNR} = \frac{E\|\mathbf{H}\mathbf{x}\|_2^2}{E\|\mathbf{n}\|_2^2} \tag{68}$$

7.1 Experimental description

BGS-Net and Improved BGS-Net simulations were implemented using Tensorflow. The number of layers T was set to 4. The training data consisted of a number of randomly generated pairs (x, y). where the data x is generated by QPSK modulation symbols. We trained the network for 1000 iterations using stochastic gradient descent and the Adam optimiser. The learning rate was set to 0.001. $k_N = k_F = 2$ and in the experimental setup we chose l_{2loss} as the cost function.

$$l_{2loss} = \frac{1}{L_{layer}} \sum_{t=1}^{L_{layer}} \|\hat{\mathbf{x}}_t - \mathbf{x}\|_2^2 \tag{69}$$

The different detectors are described in detail below, and in order to reduce the high latency of the deep learning method, all networks and iterations are set to 4 layers.

- *MMSE*: MMSE detector uses $\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H} + \frac{\sigma^2}{2}\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}$.
- *Gauss-Seidel*: The maximum number of iterations is set to layer=4.
- *TL_BD_INSA* [32]: TL_BD_INSA is an improved Neumann series approximation algorithm based on two-level block diagonal.
- *OAMPNet*: It is a DL-based detector that develops the OAMP detector. In our simulation, the number of layers of OAMPNet is set to 4 layers, and each layer has 2 learnable variables.
- *MMNet-iid*: MMNet-iid is specially designed for i.i.d. Gaussian channels. In our simulation, the number of layers of MMNet-iid is set to 4, and each layer has 2 learnable variables.

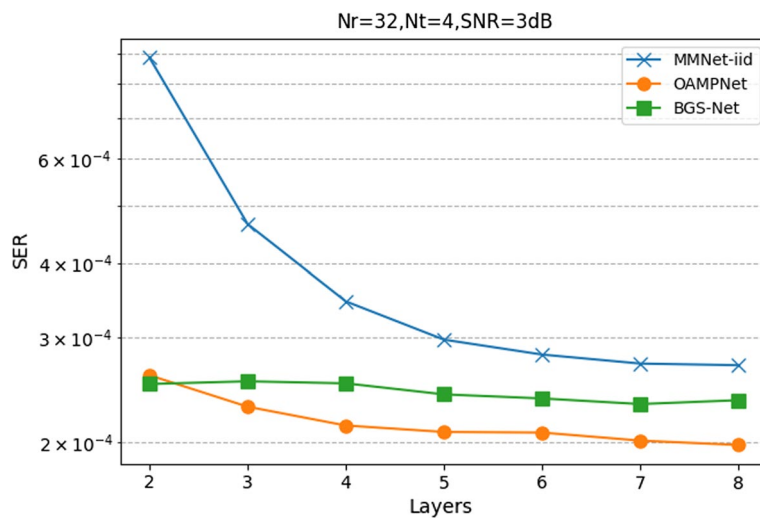


Fig. 10 Effect of the number of network layers on SER under $N_t = 4, N_r = 32$, QPSK, SAUE system

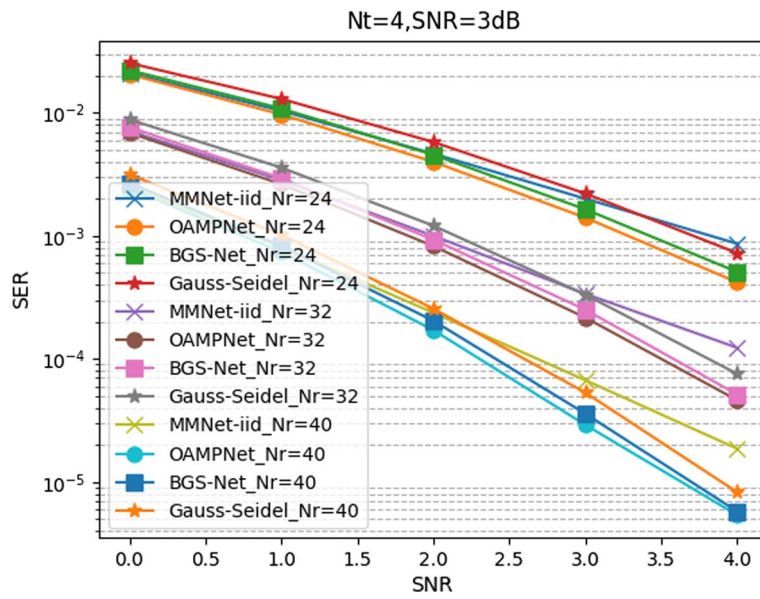


Fig. 11 Effect of the ratio α between the receiving and transmitting antennas on the performance of the algorithm for $N_t = 4$, $SNR=3dB$, QPSK, SAUE system

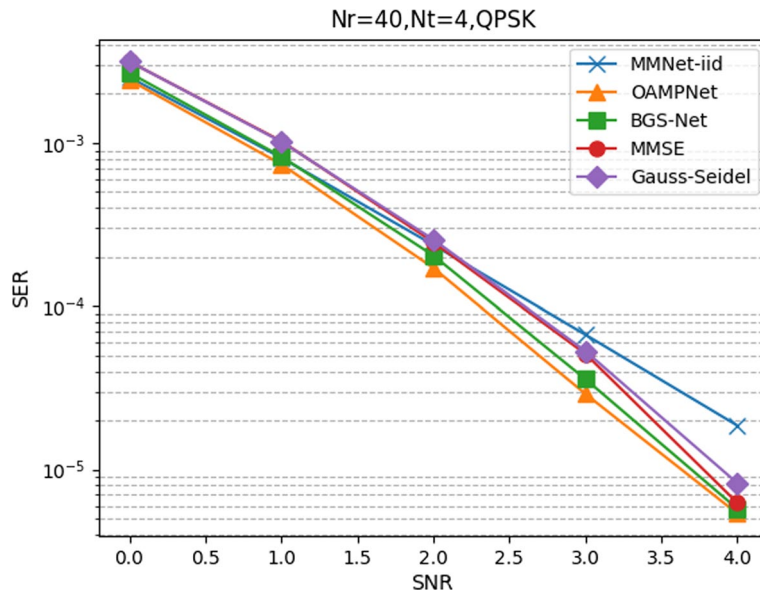


Fig. 12 SER performance comparison of different algorithms under $N_t = 4, N_r = 40$, QPSK, SAUE system

- *GS-Net*: The structure of BGS-Net is shown in Section 4. It has 4 layers and each layer has 1 learnable vector variable.
- *Improved BGS-Net*: The structure of Improved BGS-Net is shown in Section 5, it has 4 layers with 1 learnable vector variable per layer. $m_{UE} = 2$ when $N_t = 4$ or 8; $m_{UE} = 4$ when $N_t = 16$ or 32.

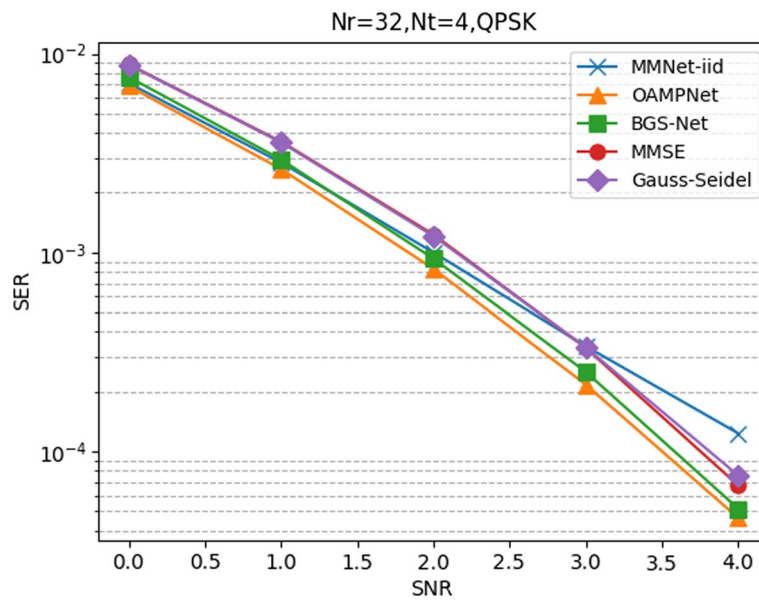


Fig. 13 SER performance comparison of different algorithms under $N_t = 4, N_r = 32$, QPSK, SAUE system

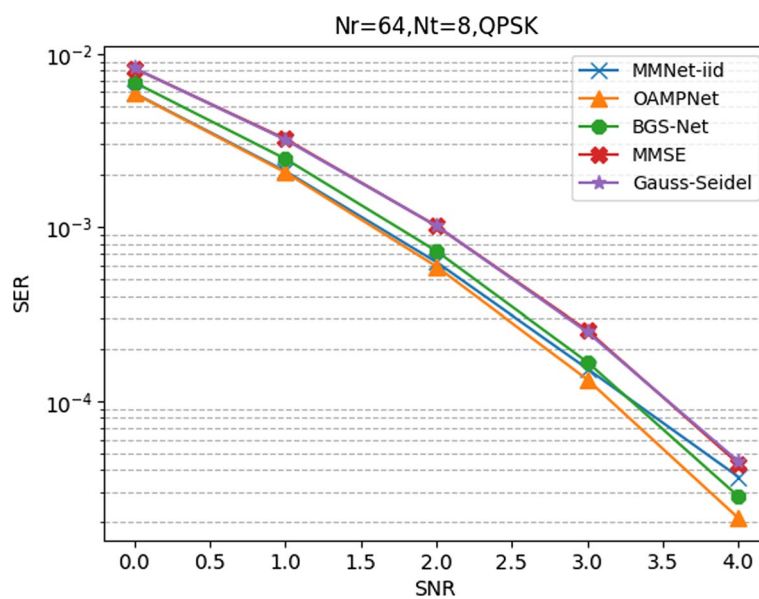


Fig. 14 SER performance comparison of different algorithms under $N_t = 8, N_r = 64$, QPSK, SAUE system

7.2 SAUE system

The performance of BGS-Net was tested on the SAUE system using QPSK modulation; the SNR was 3dB during training.

7.2.1 Convergence analysis

As shown in Figure 10, the convergence speed of BGS-Net was tested for different network layers with the same SNR of 3dB and the same number of antennas $N_r = 32, N_t = 4$. It can be observed that the 3-layer BGS-Net has converged, while MMNet-iid

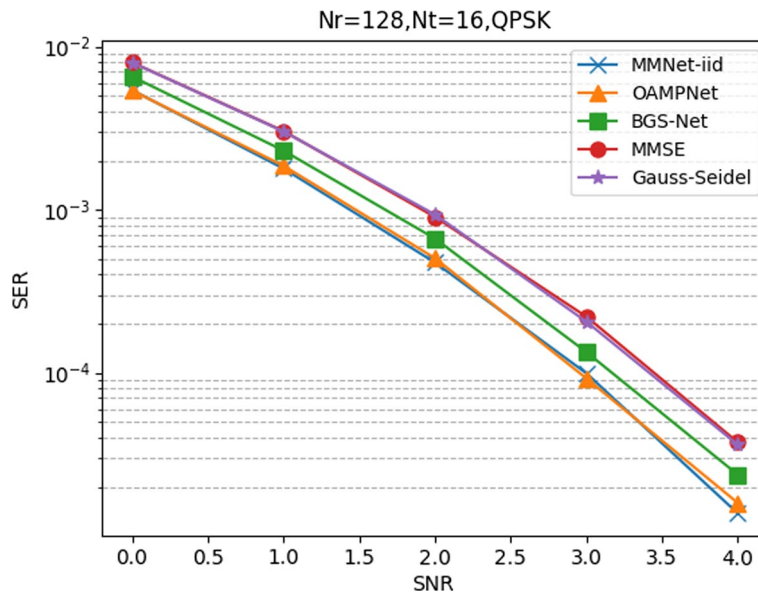


Fig. 15 SER performance comparison of different algorithms under $N_t = 16, N_r = 128, \text{QPSK, SAUE system}$

needs at least 7 layers to converge and OAMPNet needs 4 layers which indicates that BGS-Net converges fastest in the SAUE system environment. The SER of BGS-Net is much better than that of MMNet-iid in the SAUE system, but there is still a gap between the SER of BGS-Net and that of OAMPNet.

7.2.2 Impact of ratio α

This section analyzes the effect of the ratio α of the receiving antenna and the transmitting antenna on the performance of the algorithm. We set $N_t = 4, \text{SNR}=3\text{dB}$, and compare the performance of the algorithm when $N_r = 24, 32, 40$, respectively. As shown in Figure 11, it is found that as the ratio α increases, the lower the SER, the smaller the gap between the algorithms. Studying $N_t = 4, N_r = 40$ separately, as shown in Figure 12, BGS-Net can approximate the performance of OAMPNet with much lower complexity. When $N_r = 24$, the gap between BGS-Net and OAMPNet is 9×10^{-5} ; when $N_r = 40$, the gap between BGS-Net and OAMPNet is 3×10^{-7} .

7.2.3 Impact of the number of antennas

This section analyzes the influence of the number of antennas on the performance of the algorithm, and sets the ratio α as a fixed value, that is, $\alpha = 8$. As shown in Figure 13, Figure 14, and Figure 15, when the number of antennas increases, the performance of all algorithms is improving, the performance of MMNet-iid changes the most, and the performance of BGS-Net changes little. Affected by the number of antennas is small, which shows that BGS-Net is very robust. At the same time, BGS-Net has always been better than Gauss-Seidel, which shows that the nonlinear activation function improves the performance of Gauss-Seidel.

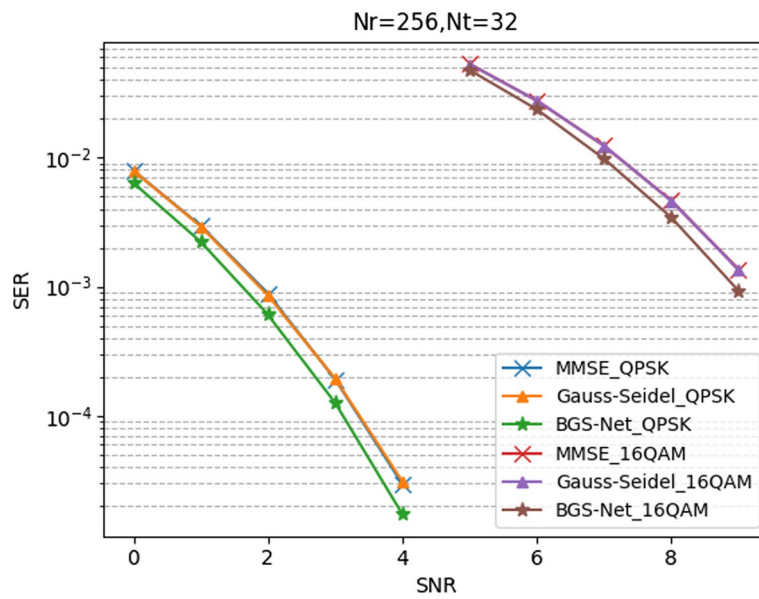


Fig. 16 Comparison of the effect of different modulation methods on SER for $N_t = 32, N_r = 256$, SAUE system

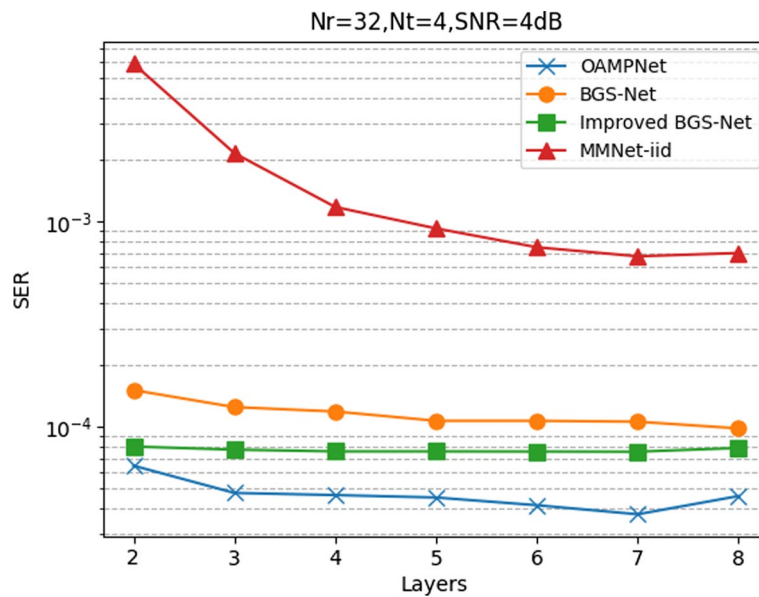


Fig. 17 Effect of number of network layers on SER for $N_t = 4, N_r = 32$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$

7.2.4 Effect of modulation order

This section analyzes the impact of modulation methods on algorithm performance. We compare the performance of MMSE, Gauss-Seidel, and BGS-Net under QPSK and 16QAM. When the test SNR is set to 5-9dB, the training test ratio is set to 7dB. As

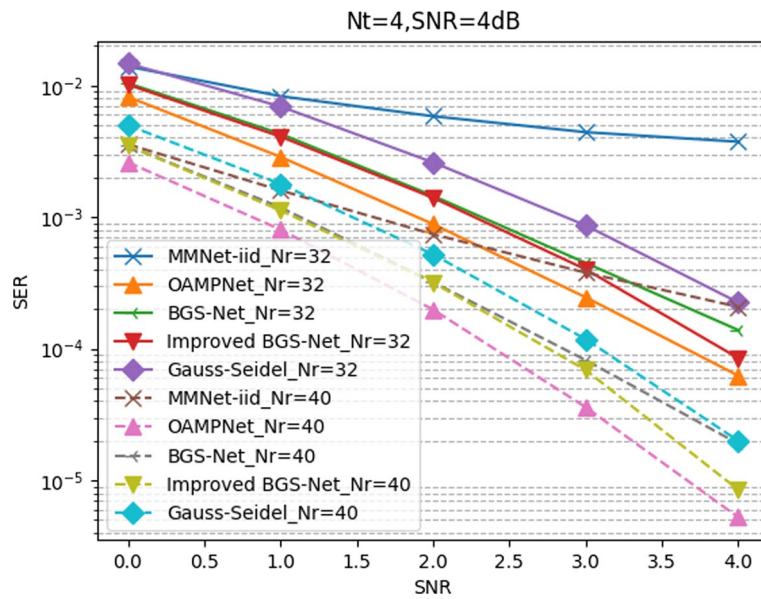


Fig. 18 Effect of the ratio α between the receiving and transmitting antennas on the performance of the algorithm for $N_t = 4$, $\text{SNR}=4\text{dB}$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$

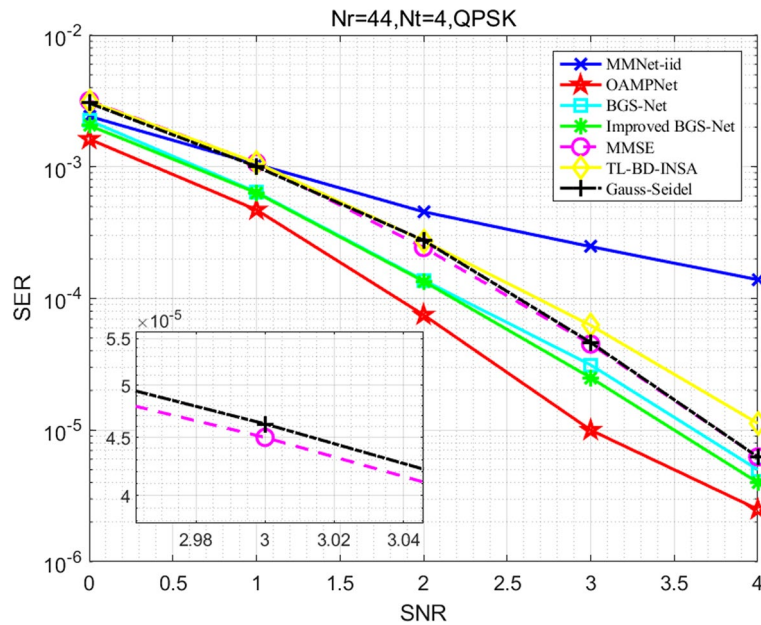


Fig. 19 SER performance comparison of different algorithms for $N_t = 4, N_r = 44$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$

shown in Figure 16, it is found that when the modulation order increases, the performance of the algorithm decreases, but the performance of BGS-Net is always better than Gauss-Seidel, and the performance of Gauss-Seidel is close to that of MMSE.

7.3 MAMU system

The MAUE system uses QPSK modulation, and the SNR during training is 4dB.

7.3.1 Convergence analysis

In order to study the performance of the algorithm, what is the difference between the MAUE system and the SAUE system. As shown in Figure 17, we tested the convergence speed of BGS-Net and Improved BGS-Net at different network layers under the same SNR of 4dB and the same number of antennas $N_t = 4, N_r = 32$, and found that the 3-layer Improved BGS-Net has converged, and the BGS-Net and OAMPNet require 4-layer can converge, while MMNet-iid needs 7-layer network to converge. The performance of MMNet-iid is much lower than that of other algorithms, while the performance of BGS-Net and Improved BGS-Net maintains a slight performance gap with OAMPNet, and the performance of Improved BGS-Net is better than BGS-Net.

7.3.2 Impact of ratio α

This section analyzes the effect of α on the performance of the algorithm under $\xi_r = 0, \xi_t = 0.2$, we set $N_t = 4, \text{SNR}=4\text{dB}$, and compare the performance of the algorithm when $N_r = 32, 40$ respectively. As shown in Figure 18, it is found that as α increases, the performance of Improved BGS-Net is closer to that of OAMPNet. As shown in Figure 19, when the antenna ratio $\alpha = 11$, the performance gap between Improved BGS-Net and OAMPNet is 2.5×10^{-6} . This shows that as long as α is large enough, the performance of Improved BGS-Net can approach OAMPNet with low complexity.

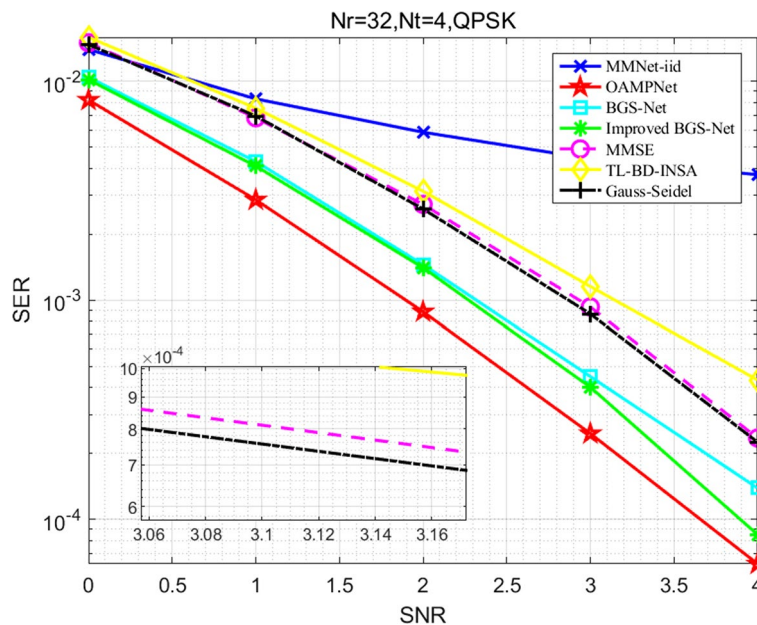


Fig. 20 SER performance comparison of different algorithms for $N_t = 4, N_r = 32$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$

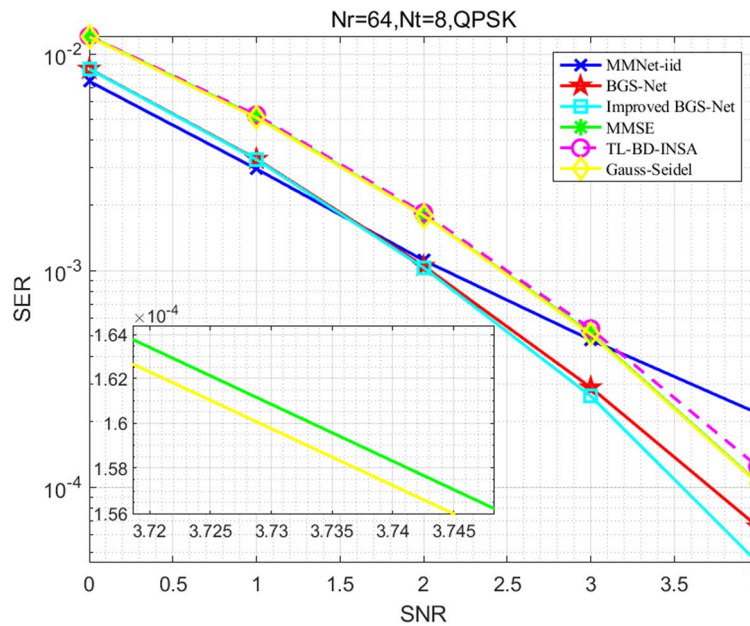


Fig. 21 SER performance comparison of different algorithms for $N_t = 8, N_r = 64$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$

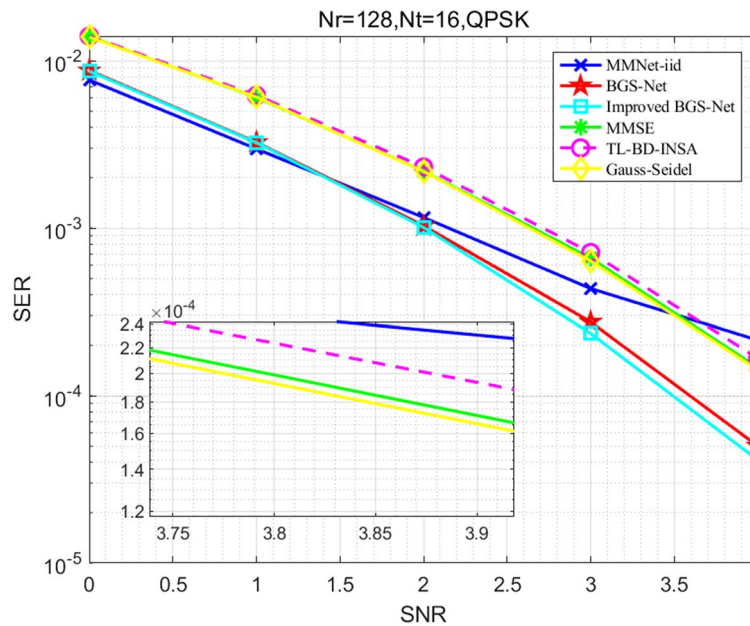


Fig. 22 SER performance comparison of different algorithms for $N_t = 16, N_r = 128$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$

7.3.3 Impact of the number of antennas

This section analyzes the influence of the number of antennas on the performance of the algorithm, we set α to a constant value, i.e. $\alpha = 8$. The performance of the algorithm is compared for $N_t = 4, N_r = 32/N_t = 8, N_r = 64/N_t = 16, N_r = 128$ respectively, as

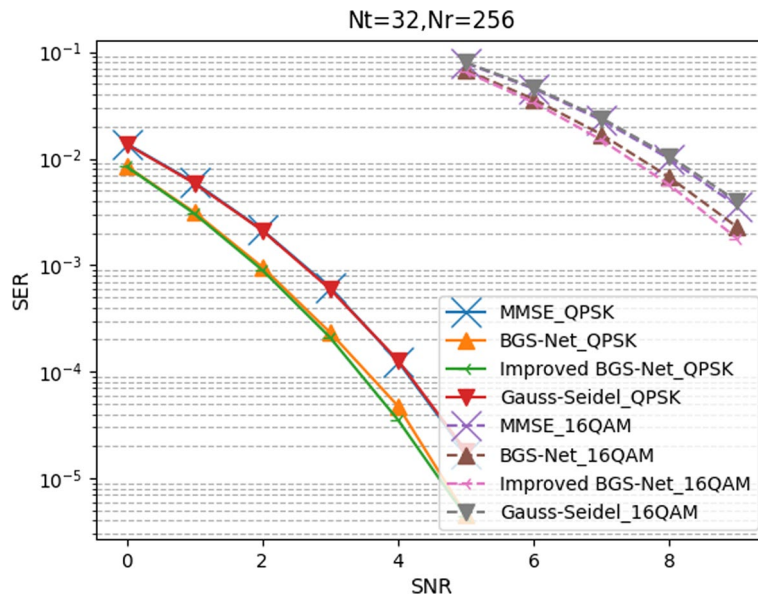


Fig. 23 Effect of different modulation orders on SER performance for $N_t = 32, N_r = 256$, MAUE system with $\xi_r = 0, \xi_t = 0.2$

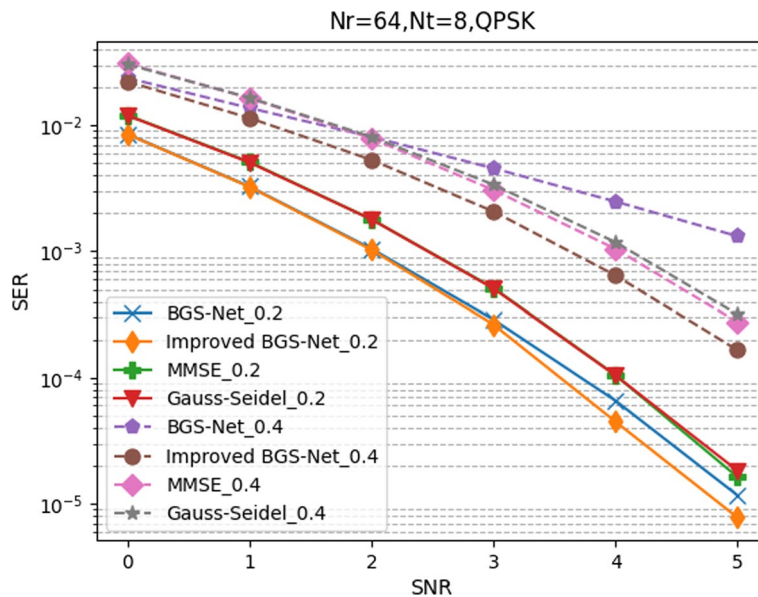


Fig. 24 SER performance comparison for $N_t = 8, N_r = 64$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$ or 0.4

shown in Figures. 20, 21, and 22. It is found that the performance gap between Improved BGS-Net and BGS-Net decreases as the number of antennas increases in the $\xi_r = 0, \xi_t = 0.2$ environment, while at the same time the MMNet-iid performance improves much faster than the others, suggesting that the impact brought by correlation can be improved by increasing the number of antennas in this environment. The fact that our proposed algorithm is consistently better than TL_BD_INSA suggests that Improved

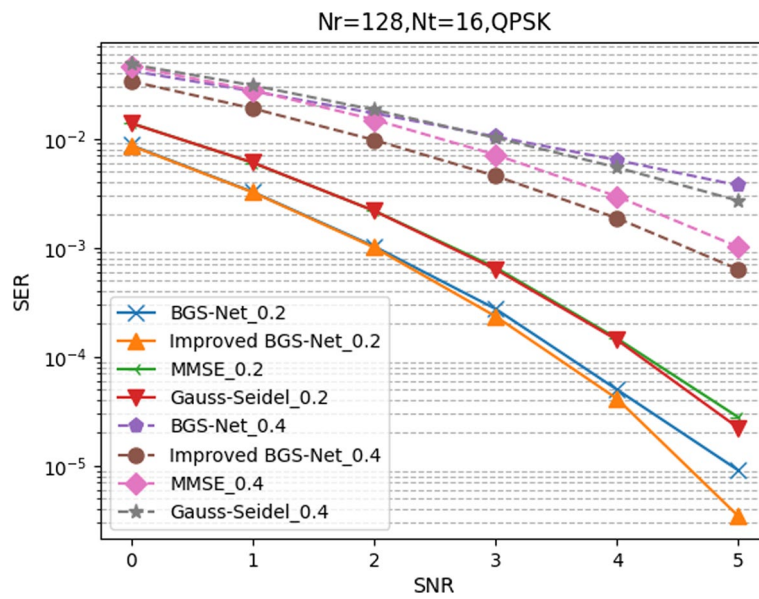


Fig. 25 SER performance comparison of different algorithms for $N_t = 16, N_r = 128$, QPSK, MAUE system with $\xi_r = 0, \xi_t = 0.2$ or 0.4

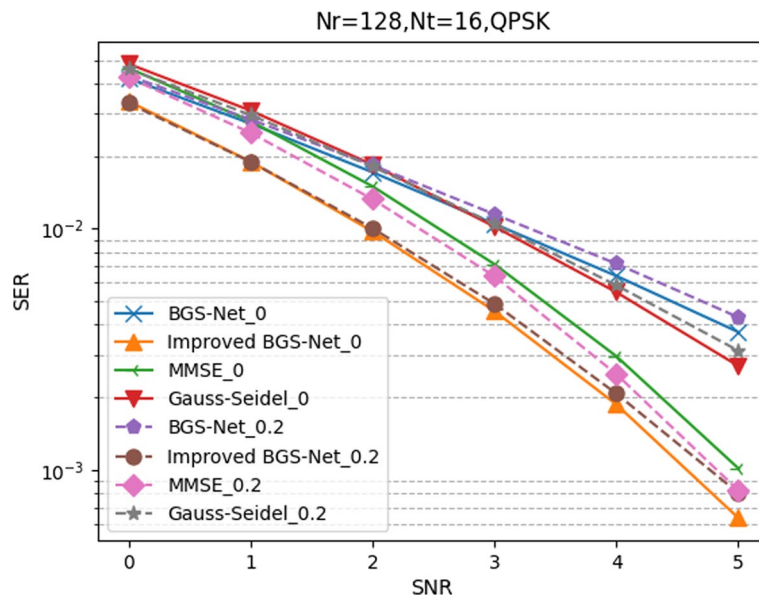


Fig. 26 SER performance comparison of different algorithms for $N_t = 16, N_r = 128$, QPSK, MAUE system with $\xi_r = 0$ or $0.2, \xi_t = 0.4$

BGS-Net does improve the performance of the algorithm based on using TL_BD_INSA as the initial solution.

7.3.4 Effect of modulation order

This section analyzes the effect of modulation on the performance of the algorithm. We compare the performance of MMSE, Gauss-Seidel, BGS-Net, and Improved BGS-Net

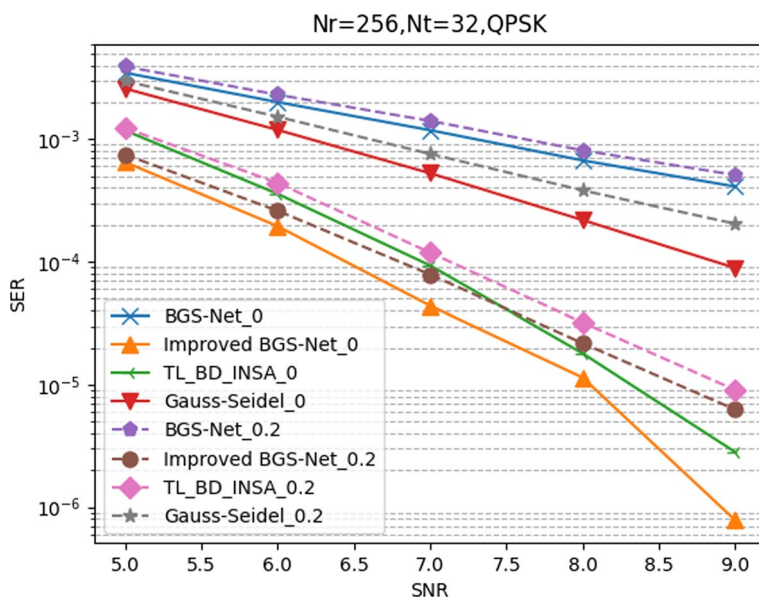


Fig. 27 SER performance comparison of different algorithms for $N_t = 32, N_r = 256$, QPSK, MAUE system with $\xi_r = 0$ or $0.2, \xi_t = 0.4$

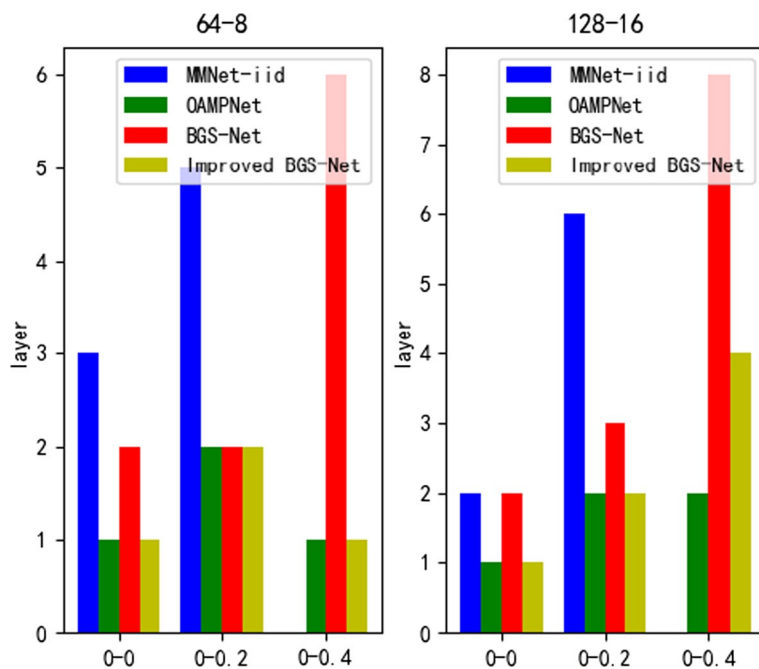


Fig. 28 Under different antenna configurations, when the SER reaches 10^{-5} in the SAUE system and the MAUE ($\xi_r = 0, \xi_t = 0.2, \mu_e=2$) system, the SER reaches 10^{-4} in the MAUE ($\xi_r = 0, \xi_t = 0.4, \mu_e=4$) system Required number of network layers

under QPSK and 16QAM. As shown in Figure 23, it is found that the larger the modulation order, the lower the performance of the algorithm, and the gap between Improved BGS-Net and BGS-Net has increased a bit. Under QPSK, BGS-Net coincides with

Table 2 When $N_t = 8, N_r = 64$, the required complexity corresponding to Figure 28 (calculated according to Table 1, based on the complexity of MMSE, NAN indicates that the performance cannot be achieved at this time)

Algorithms	Computing Complexity		
	$\xi_r = 0, \xi_k = 0$	$\xi_r = 0, \xi_k = 0.2$	$\xi_r = 0, \xi_k = 0.4$
	<i>MMNet – iid</i>	1.93	3.23
<i>OAMPNet</i>	327.68	655.36	327.68
<i>BGS – Net</i>	0.40	0.40	0.52
<i>Improved BGS – Net</i>	0.61	0.64	0.61

Table 3 When $N_t = 16, N_r = 128$, the required complexity corresponding to Figure 28 (calculated according to Table 1, based on the complexity of MMSE, NAN indicates that the performance cannot be achieved at this time)

Algorithms	Computing Complexity		
	$\xi_r = 0, \xi_k = 0$	$\xi_r = 0, \xi_k = 0.2$	$\xi_r = 0, \xi_k = 0.4$
	<i>MMNet – iid</i>	0.78	2.35
<i>OAMPNet</i>	399.61	799.22	799.22
<i>BGS – Net</i>	0.25	0.27	0.36
<i>Improved BGS – Net</i>	0.46	0.48	0.52

Improved BGS-Net at 5dB; under 16QAM, Improved BGS-Net has always been better than BGS-Net.

7.3.5 Effect of transmit correlation

To explore the effect of the correlation between the user’s multiple antennas on the performance of the algorithm, we made two sets of comparisons, one for the performance of the algorithm with $N_t = 8, N_r = 64, \xi_r = 0, \xi_t = 0.2$ or 0.4 , as shown in Figure 24. one for the performance of the algorithm with $N_t = 16, N_r = 128, \xi_r = 0,$

$\xi_t = 0.2$ or 0.4 , as shown in Figure 25. it was found that when the greater the correlation between multiple antennas, the lower the performance of the algorithm and the greater the difference between Improved BGS-Net and BGS-Net, which suggests that our improvements can make BGS-Net better adapted to the MAUE system environment.

7.3.6 Effect of receive correlation

This section analyzes the impact of the correlation between the multiple antennas of the BS on the performance of the algorithm. We have made two comparisons, one is the performance of the algorithm under $N_t = 16, N_r = 128, \xi_r = 0$ or $0.2, \xi_t = 0.4$, As shown in Figure 26. One group is the performance of the algorithm under $N_t = 32, N_r = 256, \xi_r = 0$ or $0.2, \xi_t = 0.4$, as shown in Figure 27. It is found that the greater the correlation between the multiple antennas of the BS, the lower the

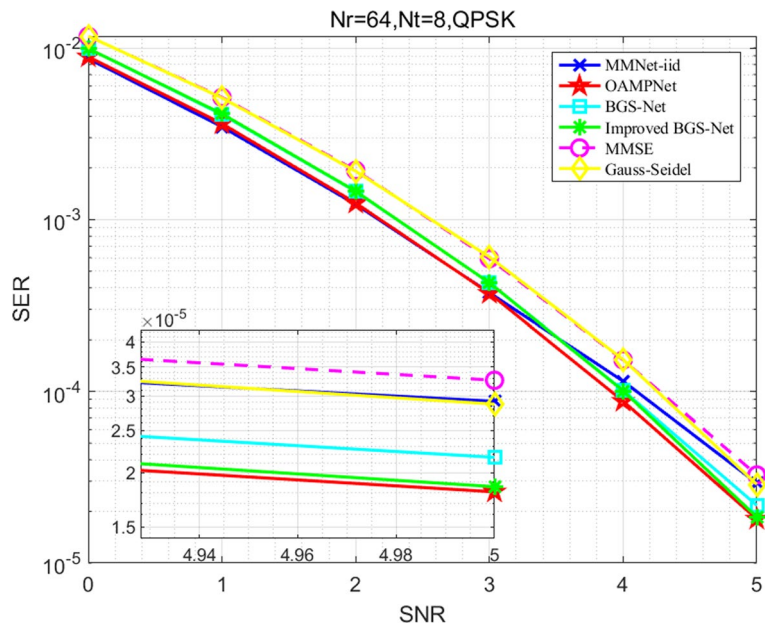


Fig. 29 SER performance of the proposed detection algorithm when there is a channel estimation error in an uplink multi-user massive MIMO system, where $N_t = 8, N_r = 64, \sigma_\epsilon^2 = 0.1$

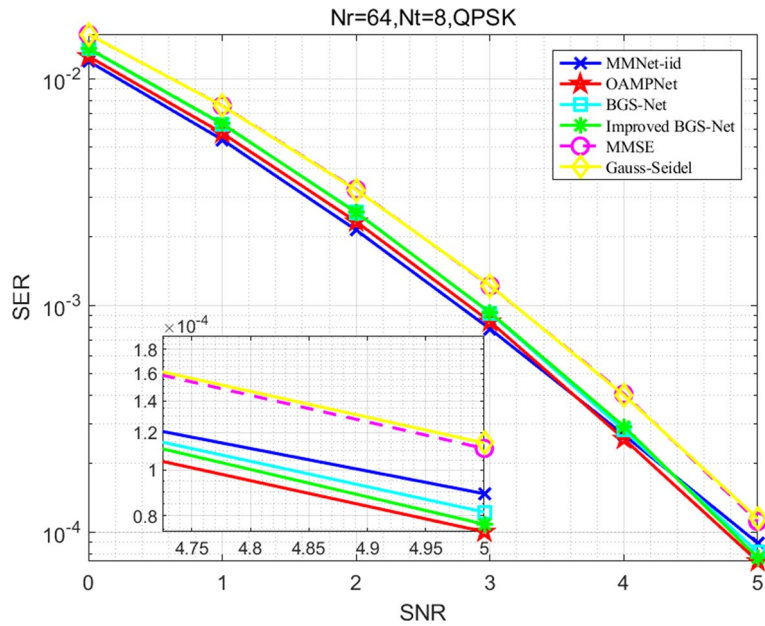


Fig. 30 The SER performance of the proposed detection algorithm when there is a channel estimation error in an uplink multi-user massive MIMO system, where $N_t = 8, N_r = 64, \sigma_\epsilon^2 = 0.2$

performance of the algorithm, and the greater the gap between Improved BGS-Net and BGS-Net. And in this environment, Improved BGS-Net and MMSE are very close, so our algorithm can only be applied to low and medium correlations.

7.4 Other performance analysis

7.4.1 Comprehensive analysis of complexity and performance

From Figure 28, Table 2, and Table 3, we can see that when the number of antennas is the same, as the correlation degree increases, the proposed algorithm requires more layers to converge; To achieve the same performance, although BGS-Net and Improved BGS-Net require more layers than OAMPNet, their required complexity is much lower than OAMPNet. When the number of antennas is increased, the performance of the algorithm should be better, but since the number of individual terminals is changed from 2 to 4, more layers are required to converge.

7.4.2 SER performance with channel estimation error

In the presence of channel estimation errors, the performance of the proposed algorithm in uplink multi-user massive MIMO systems is investigated. The estimated channel gain matrix is given by

$$\hat{\mathbf{H}} = \tilde{\mathbf{H}} + \Delta\tilde{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t} \tag{70}$$

where $\Delta\tilde{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t}$ is the error matrix of the complex Gaussian terms of iid with zero mean and variance σ_ϵ^2 .

As shown in Figures 29 and 30, when there is a channel estimation error, the Improved BGS-Net performance is very close to OAMPNet; With the increase of channel estimation error, the performance of all algorithms decreases, but the proposed detection algorithm still has good SER performance and is more robust to channel estimation error.

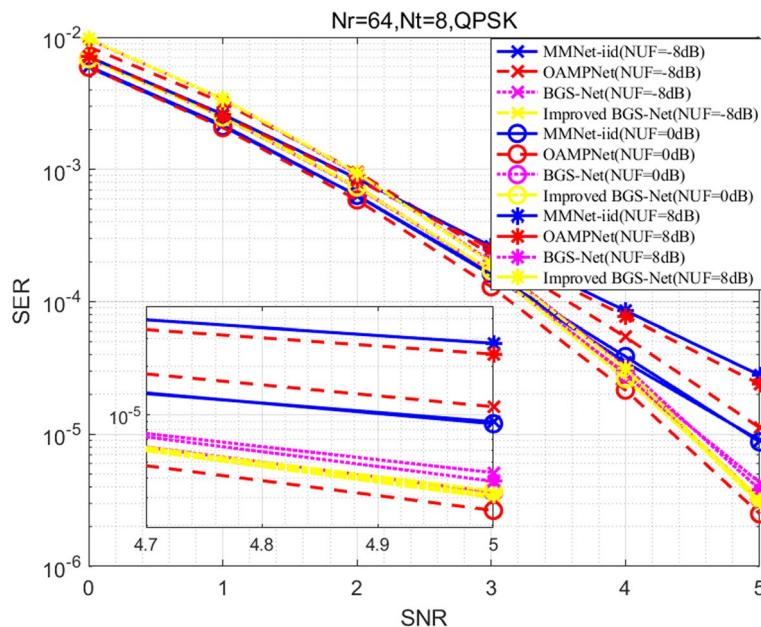


Fig. 31 SER performance of the proposed detection algorithm when there is noise estimation error in an uplink multi-user massive MIMO system, where $N_t = 8, N_r = 64$

7.4.3 SER performance with noise uncertainty

Next, we investigate the effect of noise variance uncertainty on the performance of different DL detectors. It is assumed that the noise variance is unknown in both training and testing phases. Therefore, when evaluating performance on test data, the noise variance is not the same as when training. Suppose the estimated noise variance is $\hat{\sigma}^2 = \eta\sigma^2$. We also define the noise uncertainty factor (NUF) as $\text{NUF} = 10\log_{10}\eta$.

As can be seen in Figure 31, both MMNet-iid and OAMPNet incur considerable performance losses when the estimated noise variance deviates from the true variance. When the estimation of noise variance is inaccurate, the performance gap between OAMPNet and BGS-Net, Improved BGS-Net becomes more obvious. At the same time, BGS-Net and Improved BGS-Net are hardly affected by inaccurate estimation of noise variance and have good robustness.

8 Conclusion

We propose a new model-driven deep learning network for MIMO detection, BGS-Net, and build on it with Improved BGS-Net. The network is based on Gauss-Seidel, coupled with a non-linear activation function, and exhibits excellent performance. The network needs to be optimised with few adjustable parameters and the training process is simple and fast. In this paper, single-antenna user equipment (SAUE) and multiple-antenna user equipment (MAUE) systems are considered under Rayleigh channels. Simulation results show that the performance of BGS-Net is significantly better than that of the Gauss-Seidel algorithm; the proposed scheme is suitable for massive MIMO with low complexity, and the performance can be improved by increasing the ratio between the receiving and transmitting antennas; the robustness of BGS-Net is good, and the performance is little affected by the variation of the number of antennas; under the MAUE system, the performance of Improved BGS-Net is better than that of BGS-Net, and both are suitable for low-and medium-correlation MAUE systems.

Abbreviations

BS:	Base station
BGS-Net:	Block Gauss-Seidel network
MAUE:	Multiple-antenna user equipment
DL:	Deep learning;
SAUE:	Single-antenna user equipment
MMSE:	Minimum mean square error
ZF:	Zero forcing
MIMO:	Multiple-input multiple-out
B5G:	Beyond 5G
ML:	Maximum likelihood
SD:	Spherical decoding
NS:	Neumann series
NI:	Newton iterative
GS:	Gauss-Seidel
SOR:	Successive super-relaxation
JA:	Jacobi
RI:	Richardson
CG:	Conjugate gradient
LA:	Lanczos
CD:	Coordinate descent
BP:	Belief propagation
PDN:	Parallel detection network
SIC:	Soft interference cancellation
UE:	User equipment
M-MIMO:	Massive multiple-input multiple-out

QR: Normal orthogonal matrix Q and upper triangular matrix R
 SER: Symbol error rate
 SNR: Signal to noise ratio
 OAMP: Orthogonal approximate matching tracking

Acknowledgements

Thanks to the help of the intelligent information processing team and This work is supported by National Natural Science Foundation of China under Grants No.61871238 and No.61771254.

Author contributions

Haifeng Yao conceived and designed the methods. Haifeng Yao performed the experiments and wrote the paper. Ting Li analyzed the simulation data. Fei Li and Wei Ji gave valuable suggestions on the structure of the paper. Yan Liang and Yunchao Song revised the original manuscript. All authors read and agreed the manuscript.

Availability of data and materials

Mostly, I got the writing material from different journals as presented in the references. A python tool has been used to simulate my concept.

Declarations

Ethics approval and consent to participate

No ethical issues.

Consent for publication

I agree to publication.

Competing interests

The authors declare no conflict of interest.

Received: 17 February 2022 Accepted: 3 June 2022

Published online: 21 June 2022

References

1. F.O. Catak, M. Kuzlu, E. Catak, U. Cali, D. Unal, Security concerns on machine learning solutions for 6g networks in mmwave beam prediction. *Phys. Commun.* (2022). <https://doi.org/10.1016/j.phycom.2022.101626>
2. X. Gao, L. Dai, Y. Hu, Y. Zhang, Z. Wang, Low-complexity signal detection for large-scale mimo in optical wireless communications. *IEEE J. Selected Areas Commun.* **33**(9), 1903–1912 (2015)
3. M.A. Albreem, N.A.H.B. Ismail, A review: detection techniques for lte system. *Telecommun. Syst.* **63**(2), 153–168 (2016)
4. S. Shahabuddin, O. Silvén, M. Juntti, Programmable asips for multimode mimo transceiver. *J. Signal Process. Syst.* **90**(10), 1369–1381 (2018)
5. C.D. Altamirano, J. Minango, C. De Almeida, N. Orozco, On the asymptotic ber of mmse detector in massive mimo systems, in *International Conference on Applied Technologies*, pp. 57–68 (2019)
6. C.D. Altamirano, J. Minango, H.C. Mora, C. De Almeida, Ber evaluation of linear detectors in massive mimo systems under imperfect channel estimation effects. *IEEE Access.* **7**, 174482–174494 (2019)
7. M. Wu, B. Yin, A. Vosoughi, C. Studer, J.R. Cavallaro, C. Dick, Approximate matrix inversion for high-throughput data detection in the large-scale mimo uplink, in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2155–2158 (2013)
8. M.A. Albreem, Approximate matrix inversion methods for massive mimo detectors, in *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*, pp. 87–92 (2019)
9. C. Tang, C. Liu, L. Yuan, Z. Xing, Approximate iteration detection with iterative refinement in massive mimo systems. *IET Commun.* **11**(7), 1152–1157 (2017)
10. Z. Wu, Y. Xue, X. You, C. Zhang, Hardware efficient detection for massive mimo uplink with parallel gauss-seidel method, in *2017 22nd International Conference on Digital Signal Processing (DSP)*, pp. 1–5 (2017)
11. Q. Deng, L. Guo, C. Dong, J. Lin, D. Meng, X. Chen, High-throughput signal detection based on fast matrix inversion updates for uplink massive multiuser multiple-input multi-output systems. *IET Commun.* **11**(14), 2228–2235 (2017)
12. Y. Lee, Decision-aided jacobi iteration for signal detection in massive mimo systems. *Electron. Lett.* **53**(23), 1552–1554 (2017)
13. B. Kang, J.-H. Yoon, J. Park, Low-complexity massive mimo detectors based on richardson method. *ETRI J.* **39**(3), 326–335 (2017)
14. J. Jin, Y. Xue, Y.-L. Ueng, X. You, C. Zhang, A split pre-conditioned conjugate gradient method for massive mimo detection, in *2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 1–6 (2017)
15. X. Jing, A. Li, H. Liu, A low-complexity lanczos-algorithm-based detector with soft-output for multiuser massive mimo systems. *Digit. Signal Process.* **69**, 41–49 (2017)
16. Y. Yang, Y. Xue, X. You, C. Zhang, An efficient conjugate residual detector for massive mimo systems, in *2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 1–6 (2017)
17. M. Wu, C. Dick, J.R. Cavallaro, C. Studer, High-throughput data detection for massive mu-mimo-ofdm using coordinate descent. *IEEE Trans. Circuits Syst. I Regul. Papers.* **63**(12), 2357–2367 (2016)

18. J. Yang, C. Zhang, X. Liang, S. Xu, X. You, Improved symbol-based belief propagation detection for large-scale mimo, in *2015 IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 1–6 (2015)
19. H. Hua, X. Wang, Y. Xu, Signal detection in uplink pilot-assisted multi-user mimo systems with deep learning, in *2019 Computing, Communications and IoT Applications (ComComAp)*, pp. 369–373 (2019)
20. J. Xia, K. He, W. Xu, S. Zhang, L. Fan, G.K. Karagiannidis, A mimo detector with deep learning in the presence of correlated interference. *IEEE Trans. Veh. Technol.* **69**(4), 4492–4497 (2020)
21. H. Ye, G.Y. Li, B.-H. Juang, Power of deep learning for channel estimation and signal detection in ofdm systems. *IEEE Wirel. Commun. Lett.* **7**(1), 114–117 (2017)
22. X. Jin, H.-N. Kim, Parallel deep learning detection network in the mimo channel. *IEEE Commun. Lett.* **24**(1), 126–130 (2019)
23. N. Samuel, T. Diskin, A. Wiesel, Learning to detect. *IEEE Trans. Signal Process.* **67**(10), 2554–2564 (2019)
24. N. Samuel, T. Diskin, A. Wiesel, Deep mimo detection, in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5 (2017)
25. H. He, C.-K. Wen, S. Jin, G.Y. Li, A model-driven deep learning network for mimo detection, in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 584–588 (2018)
26. H. He, C.-K. Wen, S. Jin, G.Y. Li, Model-driven deep learning for mimo detection. *IEEE Trans. Signal Process.* **68**, 1702–1715 (2020)
27. M. Khani, M. Alizadeh, J. Hoydis, P. Fleming, Exploiting channel locality for adaptive massive mimo signal detection, in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8565–8568 (2020)
28. M. Khani, M. Alizadeh, J. Hoydis, P. Fleming, Adaptive neural signal detection for massive mimo. *IEEE Trans. Wirel. Commun.* **19**(8), 5635–5648 (2020)
29. X. Tan, W. Xu, K. Sun, Y. Xu, Y. Be'ery, X. You, C. Zhang, Improving massive mimo message passing detectors with deep neural network. *IEEE Trans. Veh. Technol.* **69**(2), 1267–1280 (2019)
30. Y. Wei, M.-M. Zhao, M. Hong, M.-J. Zhao, M. Lei, Learned conjugate gradient descent network for massive mimo detection. *IEEE Trans. Signal Process.* **68**, 6336–6349 (2020)
31. N. Shlezinger, R. Fu, Y.C. Eldar, DeepSic: deep soft interference cancellation for multiuser mimo detection. *IEEE Trans. Wirel. Commun.* **20**(2), 1349–1362 (2020)
32. H. Wang, Y. Ji, Y. Shen, W. Song, M. Li, X. You, C. Zhang, An efficient detector for massive mimo based on improved matrix partition. *IEEE Trans. Signal Process.* **69**, 2971–2986 (2021)
33. J. Li, R. Chen, C. Li, W. Liu, D. Chen, Lattice-reduction-aided detection in spatial correlated mimo channels. *J. Xidian Univ. Nat. Sci.* **39**(1) (2012)
34. A. Van Zelst, J. Hammerschmidt, A single coefficient spatial correlation model for multiple-input multiple-output (mimo) radio channels. *Proc. 27th General Assembly of the Int. Union of Radio Science (URSI)*. (2002)
35. J. Ma, L. Ping, Orthogonal amp. *IEEE Access.* **5**, 2020–2033 (2017)
36. Z. Zhang, J. Wu, X. Ma, Y. Dong, Y. Wang, S. Chen, X. Dai, Reviews of recent progress on low-complexity linear detection via iterative algorithms for massive mimo systems, in *2016 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, pp. 1–6 (2016)
37. M.A. Albreem, W. Salah, A. Kumar, M.H. Alsharif, A.H. Rambe, M. Jusoh, A.N. Uwaechia, Low complexity linear detectors for massive mimo: a comparative study. *IEEE Access.* **9**, 45740–45753 (2021)
38. Chongqing Jiaotong University, Numerical Analysis for Graduate Students (12) Gauss-Seidel Iterative Method (2021). <https://wenku.baidu.com/view/1ae31ade998fcc22bcd10df0.html> Accessed September 24, 2021
39. W. Zhang, R.C. de Lamare, C. Pan, M. Chen, J. Dai, B. Wu, Simplified matrix polynomial-aided block diagonalization precoding for massive mimo systems, in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1–5 (2016)
40. Y. Ji, Z. Wu, Y. Shen, J. Lin, Z. Zhang, X. You, C. Zhang, A low-complexity massive mimo detection algorithm based on matrix partition, in *2018 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 158–163 (2018)
41. F. Wang, C. Zhang, X. Liang, Z. Wu, S. Xu, X. You, Efficient iterative soft detection based on polynomial approximation for massive mimo, in *2015 International Conference on Wireless Communications & Signal Processing (WCSP)*, pp. 1–5 (2015)
42. C.F. Van Loan, G. Golub, Matrix computations (johns hopkins studies in mathematical sciences). *Matrix Comput.* (1996)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.