Ordered Iterative Methods for Low-Complexity Massive MIMO Detection

Beilei Gong, Ningxin Zhou, Zheng Wang
School of Information Science and Engineering
Southeast University, Nanjing, China
Email: beilei_gong@seu.edu.cn, ningxin_zhou@seu.edu.cn, z.wang@ieee.org

2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring) | 979-8-3503-1114-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/VTC2023-Spring57618.2023.10200673

***Abstract*—In this paper, two ordered iterative detection methods are proposed for better signal detection performance in massive multiple-input multiple-output (MIMO) systems. First of all, in order to reduce error propagation in the traditional iterative detection schemes with sequential order, the ordered iterative detection (OID) algorithm is proposed, which achieves a better detection performance with low complexity. Then, we show that the convergence performance chiefly depends on the residual component during the iterations. Therefore, a dynamic ordering strategy is given for further performance improvement, which leads to the modified ordered iterative detection (MOID) algorithm. After that, we extend the proposed MOID algorithm via deep learning network (DNN), and parameters like relaxation factor are trained to optimal for further performance gain.**

***Index Terms*—Massive MIMO detection, iterative detection, iteration methods, deep neural network.**

I. Introduction

As a key pillar in 5G communication networks, massive multiple-input multiple-output (MIMO) can significantly improve the spectral and energy efficiency [1]. However, the growing number of antennas at the both sides of MIMO systems also poses a pressing challenge upon the uplink signal detection. It has been shown in [2] that the near-optimal maximum-likelihood (ML) detection performance can be achieved by the traditional zero forcing (ZF) and minimum mean square error (MMSE) detection schemes when the number of antennas at the base station (denoted by N_r) is sufficiently larger than that of user equipments (denoted by N_t). Nevertheless, the implementation of ZF or MMSE is still challenging because of the matrix inversion with computational complexity $\mathcal{O}\left(N_t^3\right)$.

In order to reduce the complexity burden of linear detection, a number of suboptimal iterative detectors are proposed, where low complexity $\mathcal{O}\left(N_t^2\right)$ can be achieved. Specifically, the iterative detection based on Jacobi method in [3] passes from one iteration to the next by approaching one component of the vector estimation at a time. Compared to Jacobi iteration, the Gauss-Seidel (GS) iteration method is a similar but inherently sequential algorithm since each component of the latest iteration depends on all the previously updated computed components [4]. Based on GS, it is also shown that relaxation factor $0 < \omega < 2$ has a positive effect on the convergence if chosen wisely, and this leads to successive over relaxation (SOR) [5] iteration.

Corresponding author: Zheng Wang (e-mail: z.wang@ieee.org)

Nowadays, deep learning (DL) has revolutionized many research fields, which combines the internal structure of certain model-based algorithms with the remarkable power of deep neural networks (DNN). In massive MIMO detection, by unfolding projected gradient descent via DNN, the DetNet in [6] is able to achieve better performance than MMSE detector. On the other hand, the model-driven DL detectors try to exploit full domain knowledge to achieve comparable performance with few parameters optimized. For example, the OAMPNet algorithm in [7] outperforms the original orthogonal approximate message passing (OAMP) algorithm by adding only a few trainable parameters to the constructed network.

In this paper, in order to improve the detection performance, two ordered iterative detection algorithms are proposed, which is well compatible to the traditional sequential iteration methods. First of all, by reducing error propagation, the proposed OID algorithm has afforded us a better detection performance with low complexity. Then, we show that the convergence rate is essentially related to residual component during the iterations. Therefore, to further speed up the convergence, the MOID algorithm is proposed, with a dynamical ordering strategy. Moreover, we construct a DNN by unfolding the proposed MOID algorithm, where the relaxation factor is optimized as a trainable parameter.

II. System Model

Consider the massive MIMO system with N_t transmit and N_r receive antennas. Let $\mathbf{x} \in \mathcal{A}^{N_t}$ denote the transmit signal vector, and the corresponding received signal vector $\mathbf{y} \in \mathbb{C}^{N_r}$ is given by

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

Here, $\mathcal{A} = \{\pm 1, \pm 3, ..., \pm\sqrt{M} - 1\}$ with M representing the modulation index of the corresponding Quadrature Amplitude Modulation (QAM), $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, $\mathbf{n} \in \mathbb{C}^{N_r}$ is the additive white Gaussian noise (AWGN) vector with zero mean and variance σ_n^2.

Theoretically, the optimal ML detection computes

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{A}^{N_t}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2, \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm. However, it is not feasible for massive MIMO systems due to the exponentially increased complexity. Therefore, traditional linear detection like ZF or MMSE turns out to be an effective alternative for massive MIMO systems by taking advantages of receive

979-8-3503-1114-3/23/$31.00 ©2023 IEEE

diversity when $N_r \geq N_t$. Specifically, the linear MMSE detector [2] firstly performs the following estimations

$$\tilde{\mathbf{x}}_{\text{MMSE}} = \left(\mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y}, \tag{3}$$

and the signal decision $\widehat{\mathbf{x}}_{\text{MMSE}}$ is then determined by

$$\widehat{\mathbf{x}}_{\text{MMSE}} = \lceil \tilde{\mathbf{x}}_{\text{MMSE}} \rfloor_Q \in \mathcal{A}^{N_t}, \tag{4}$$

where $\lceil \cdot \rfloor_Q$ denotes the direct rounding according to the discrete constellation $\mathcal{A}^{N_t}$. However, due to the matrix inversion in (3), the complexity of MMSE (or ZF) detector is of order $\mathcal{O}\left(N_t^3\right)$, which is still unaffordable especially in high dimensional systems.

In order to bypass the matrix inversion, a number of low-complexity detectors based on iterative methods are proposed to solve the following linear equation

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{5}$$

which is equivalent to (3) with $\mathbf{b} = \mathbf{H}^H\mathbf{y} \in \mathbb{C}^{N_t}$ and the MMSE filtering matrix $\mathbf{A} = \mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I} \in \mathbb{C}^{N_t \times N_t}$. In particular, the Jacobi iteration [3] updates each component of $\mathbf{x}$ in an iterative way as follows

$$x_i^{(t+1)} = x_i^{(t)} + \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(t)} - \sum_{j=i}^{n} a_{ij}x_j^{(t)}\right), \tag{6}$$

where t indicates the iteration index, $x_i^{(t)}$ represents the i-th component of the iterate $\mathbf{x}^{(t)}$, b_i denotes the i-th component of $\mathbf{b}$, and $a_{ij} \in \mathbb{R}$ stands for the element of matrix $\mathbf{A}$.

Different from Jacobi iteration, the iteration of $\mathbf{x}$ in GS iteration [4] is carried out element by element in a sequential order, where the newly updated elements of $\mathbf{x}^{(t+1)}$ (i.e., $x_j^{(t+1)}, 1 \leq j < i$) are also taken into account to update the current element $x_i^{(t+1)}$ as

$$x_i^{(t+1)} = x_i^{(t)} + \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(t+1)} - \sum_{j=i}^{n} a_{ij}x_j^{(t)}\right). \tag{7}$$

Besides, as an accelerated GS method with the aid of relaxation factor $0 < \omega < 2$ [5], SOR iteration updates $x_i^{(t+1)}$ as

$$x_i^{(t+1)} = x_i^{(t)} + \frac{\omega}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(t+1)} - \sum_{j=i}^{n} a_{ij}x_j^{(t)}\right), \tag{8}$$

with the optimal ω_{opt} computed by

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho^2(J)}}. \tag{9}$$

Here, $\rho^2(\mathbf{J})$ is the spectrum of a matrix $\mathbf{J} \in \mathbb{C}^{N_t \times N_t}$.

Algorithm 1: OID

Input $\mathbf{A} = \mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I}, \mathbf{b} = \mathbf{H}^H\mathbf{y}, \mathbf{x}^{(0)} = \mathbf{0}, T$
Output near MMSE detection solution $\widehat{\mathbf{x}}^{(t)}$
for $t = 0, 1, ..., T-1$ **do**
  Select i coordinate in descending order of $|a_{ii}|$
  Update $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \frac{\omega}{a_{ii}} r_i^{(t)}\mathbf{e}_i$
end for
output $\widehat{\mathbf{x}}^{(t)}$ by rounding $\mathbf{x}^{(t)}$ based on constellation $\mathcal{A}^{N_t}$

III. Ordered Iterative Detection Scheme

Different from Jacobi iteration with parallel structure for implementation, the iterations in GS and SOR methods are carried out sequentially in a forwards or backwards order (e.g., from $i = 1$ to $i = N_t$). Nevertheless, we point out that the order of updating the components of $\mathbf{x}$ also has an inherent impact on the iteration performance [8]. For this reason, two ordering strategies for the traditional iterative detection are proposed in this section for better detection performance.

For a better presentation, the proposed OID and MOID are described based on the traditional SOR iteration, where the related extensions to GS iteration are straightforward with $\omega = 1$. Meanwhile, the update in (8) with respect to one component of $\mathbf{x}$ (i.e., x_i) can be expressed in a vector way

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \frac{\omega}{a_{ii}} r_i^{(t)}\mathbf{e}_i, \tag{10}$$

with the residual

$$r_i^{(t)} = b_i - \mathbf{a}_i^H\mathbf{x}^{(t)}. \tag{11}$$

Here, $\mathbf{e}_i$ is the i-th coordinate basis column vector, $\mathbf{a}_i^H$ denotes the i-th row vector of matrix $\mathbf{A}$. Clearly, according to (10), by performing the update of x_i, $1 \leq i \leq N_t$ in a certain order, a full iteration in GS or SOR method is finished, and now we are seeking for a better way to update x_i in what follows.

A. Ordered Iterative Detection

The first criterion we considered for the update order is to suppress the effect of noise, so that the underlying error propagation can be controlled in a reasonable way.

Specifically, according to (10) and (11), we have

$$\begin{aligned}\mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} + \frac{\omega}{a_{ii}}(b_i - \mathbf{a}_i^H\mathbf{x}^{(t)})\mathbf{e}_i \\ &= \mathbf{x}^{(t)} - \frac{\omega}{a_{ii}}\mathbf{a}_i^H\mathbf{x}^{(t)}\mathbf{e}_i + \frac{\omega}{a_{ii}} b_i\mathbf{e}_i,\end{aligned} \tag{12}$$

where the noise is contained by the vector $\mathbf{b}$ as follows

$$\mathbf{b} = \mathbf{H}^H\mathbf{y} = \mathbf{H}^H(\mathbf{H}\mathbf{x} + \mathbf{n}) = \mathbf{H}^H\mathbf{H}\mathbf{x} + \mathbf{H}^H\mathbf{n}. \tag{13}$$

Now using (13) we have

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \frac{\omega}{a_{ii}}\mathbf{a}_i^H\mathbf{x}^{(t)}\mathbf{e}_i + \frac{\omega}{a_{ii}}\mathbf{g}_i^H\mathbf{x}^{(t)}\mathbf{e}_i + \underbrace{\frac{\omega}{a_{ii}}(\mathbf{h}_i)^H\mathbf{n}\mathbf{e}_i}_{\text{noise part}}. \tag{14}$$

Here, $\mathbf{g}_i^H$ is the i-th row vector of matrix $\mathbf{G} = \mathbf{H}^H\mathbf{H}$, $\mathbf{h}_i$ denotes the i-th column vector of matrix $\mathbf{H}$.

Intuitively, from (14), because a larger size of a_{ii} would naturally lead to a smaller noise impact upon the element x_i, it is preferable to update x_i in the descending order of $|a_{ii}|$ so that the noise in b_i will be reasonably controlled. Otherwise, the possible error propagation will happen along the sequential updating, which is harmful to the iteration convergence. To make it more clear, the following analysis is presented to reveal the relationship between noise and a_{ii}.

$$\begin{aligned} \|\frac{\omega}{a_{ii}}(\mathbf{h}_i)^H\mathbf{n}\|^2 &\leq \omega^2\frac{\|\mathbf{h}_i\|^2}{|a_{ii}|^2}\|\mathbf{n}\|^2 \\ &= \omega^2\|\mathbf{n}\|^2\frac{|(\mathbf{h}_i)^H\mathbf{h}_i|}{|a_{ii}|^2} \\ &\overset{(a)}{=} \omega^2\|\mathbf{n}\|^2\frac{|a_{ii}-\sigma_n^2|}{|a_{ii}|^2} \\ &\propto \frac{1}{|a_{ii}|}, \end{aligned} \tag{15}$$

where the equality (a) comes from the fact $\mathbf{A}=\mathbf{H}^H\mathbf{H}+\sigma_n^2\mathbf{I}$. Clearly, this means the upper bound of the noise part partially depends on $|a_{ii}|$. More precisely, the larger $|a_{ii}|$, the stronger ability for noise suppression, so that the component x_i with larger size of $|a_{ii}|$ should be processed earlier. To summarize, this leads to the proposed OID algorithm for massive MIMO systems, which is outlined in details in Algorithm 1.

B. Modified Ordered Iterative Detection

Although the ordering strategy based on $|a_{ii}|$ provides a performance gain to the traditional iterative detection, it does have its own restriction due to the fixed order. Because of the cyclic traversal iteration by iteration, the update based on small $|a_{ii}|$ will still be carried out, which still limits the performance gain stemming from the ordering mechanism. Therefore, to avoid such a latent problem, a dynamic ordering strategy is proposed, which takes the updated residual r_i into account as well.

In particular, from (10), the difference between $\mathbf{x}^{(t+1)}$ and $\mathbf{x}^{(t)}$, i.e., $\Delta\mathbf{x}$, is determined by $\frac{\omega}{a_{ii}}r_i^{(t)}$, namely,

$$\mathbf{x}^{(t+1)}-\mathbf{x}^{(t)}=\Delta\mathbf{x}=\frac{\omega}{a_{ii}}r_i^{(t)}\mathbf{e}_i. \tag{16}$$

Therefore, from a perspective of iteration convergence, a large size $\Delta\mathbf{x}$ implies a large change of the iteration, which may have a positive impact upon the convergence. Motivated by this point, we propose to update x_i in a descending order of $\|\Delta\mathbf{x}\|$, which corresponds to a descending order of $|\frac{r_i^{(t)}}{a_{ii}}|$. Clearly, such a update mechanism is dynamic with the iteration of $\mathbf{x}^{(t)}$, which not only effectively avoids the problem of fixed order but also enables a better iteration convergence.

C. Complexity Analysis

As the traditional iterative detection schemes are completed by updating all the N_t components of $\mathbf{x}$ in a sequential order, for ease of comparison, we adopt N_t times iterations as an outer-loop and number of outer-loops is $k=T/N_t$. Note that better performance can be achieved by maximizing

Algorithm 2: MOID

Input $\mathbf{A}=\mathbf{H}^H\mathbf{H}+\sigma_n^2\mathbf{I}, \mathbf{b}=\mathbf{H}^H\mathbf{y}, \mathbf{x}^{(0)}=\mathbf{0}, T$
Output near MMSE detection solution $\widehat{\mathbf{x}}^{(t)}$

for $t=0,1...,T-1$ **do**
  Update the descending order $o(i)$ by $|\frac{r_i^{(t)}}{a_{ii}}|$ when $t=0,N_t,2N_t,...$
  Select i coordinate in descending order of $o(i)$
  Update $\mathbf{x}^{(t+1)}=\mathbf{x}^{(t)}+\frac{\omega}{a_{ii}}r_i^{(t)}\mathbf{e}_i$
end for
output $\widehat{\mathbf{x}}^{(t)}$ by rounding $\mathbf{x}^{(t)}$ based on constellation $\mathcal{A}^{N_t}$

TABLE I
COMPUTATIONAL COMPLEXITY OF ITERATIVE DETECTION SCHEMES

MIMO detection	Multiplication	Summation
SOR [12]	$N_t^2+N_t$	$N_t^2+N_t$
OID	$N_t^2+2N_t$	$N_t^2+N_t$
MOID	$2N_t^2+3N_t$	$N_t^2+N_t$

$|\frac{r_i^{(t)}}{a_{ii}}|$ at each iteration. However, it costs $\mathcal{O}\left(N_t^3\right)$ in an outer-loop with the residual calculation (11) involved. To reduce the complexity burden, we give the loosely dynamic ordering strategy by sorting the components of $\mathbf{x}$ in a descending order of $|\frac{r_i^{(t)}}{a_{ii}}|$ for every outer-loop, where the related details can be found in Algorithm 2.

Considering the additional computational complexity introduced by sort operation is no more than $\mathcal{O}\left(N_t^2\right)$ and the sort operation of $|a_{ii}|$ can be viewed as preprocess in OID algorithm, further reducing the complexity. To summarize, the computational complexities of OID and MOID algorithm at each outer-loop actually maintain $\mathcal{O}\left(N_t^2\right)$, which are much lower than the classic Ordered Successive Interference Cancellation (OSIC) with $\mathcal{O}\left(N_t^3\right)$ [9] and are still competitive compared to the traditional SOR iteration. To be more specific, the multiplication and addition as a rough measurement for the computational complexity of iterative detection schemes per outer-loop are listed in Table I.

IV. MOID-NET

For a better detection performance, we further upgrade the proposed MOID algorithm with DNN, where the parameter optimization and nonlinear projection operations are trained via DL. Typically, model-driven DL detectors were carried out based on AMP with complexity $\mathcal{O}\left(N_tN_r\right)$ [10], while we now try to introduce such a lower-complexity version into the model-based algorithm.

First of all, the complex-valued system model in (1) should be converted to an equivalent real-valued one constrained in the DL environment [11], where the related transformation process is omitted due to the simplicity. Here, to reduce the training burden, the number of layers in our network is equivalent to the number of outer-loops k.

According to (16), the relaxation factor ω plays an important role in the proposed MOID algorithm. However, the related

Algorithm 3: MOID-Net

Require: $\mathbf{A} = \mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I}, \mathbf{b} = \mathbf{H}^H\mathbf{y}, \mathbf{x}_0^{(0)} = \mathbf{0}, K, T$
Ensure: near MMSE detection solution $\widehat{\mathbf{x}}_k$

for $k = 0, 1..., K-1$ **do**
  Update the descending order $o(i)$ by $|\frac{r_i^{(0)}}{a_{ii}}|$
  for $t = 0, 1..., T-1$ **do**
    Select i coordinate according to $i = o(t)$
    Update $\mathbf{x}_k^{(t+1)} = \mathbf{x}_k^{(t)} + \frac{\omega_k}{a_{ii}} r_i^{(t)} \mathbf{e}_i$
  end for
  $\mathbf{z}_k = \text{ReLU}(\mathbf{W}_z \mathbf{x}_k^{(T)} + \mathbf{p}_z)$
  $\mathbf{x}_{oh,k+1} = \mathbf{W}_x \mathbf{z}_k + \mathbf{p}_x$
  $\mathbf{x}_{k+1}^{(0)} = f_{oh}(\mathbf{x}_{oh,k+1})$
end for
output $\widehat{\mathbf{x}}_k$ by rounding $\mathbf{x}_k^{(0)}$ based on constellation $\mathcal{A}^{N_t}$

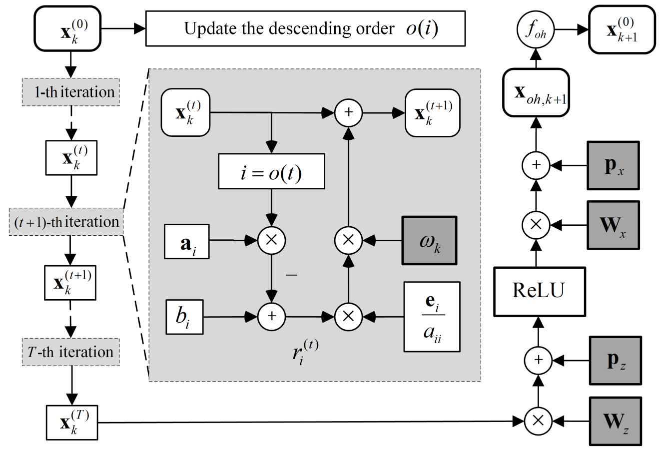

Fig. 1. The architecture of the MOID-Net detector.

calculation of $\rho(J)$ in (9) involves sophisticated matrix factorization, which is unaffordable in practice. Alternatively, $\rho(J)$ is usually approximated by [13]

$$\rho(J) = \left(1 + \sqrt{\frac{N_t}{N_r}}\right)^2 - 1, \tag{17}$$

which suffers from requirements for a certain antenna ratio N_t/N_r.

Instead of using other computationally expensive method to find an optimal ω, we use a DL approach to provide an appropriate relaxation factor, which is transformed into a layer-dependent learnable parameter ω_k. Moreover, a nonlinear projection operation is also constructed to enable the proposed MOID-Net to outperform the original MOID algorithm in terms of the detection performance. More specifically, the following operations are carried out.

$$\mathbf{x}_k^{(T)} = \mathbf{x}_k^{(T-1)} + \frac{\omega_k}{a_{ii}} r_i^{(T)} \mathbf{e}_i, \tag{18}$$

$$\mathbf{z}_k = \text{ReLU}(\mathbf{W}_z \mathbf{x}_k^{(T)} + \mathbf{p}_z), \tag{19}$$

$$\mathbf{x}_{oh,k+1} = \mathbf{W}_x \mathbf{z}_k + \mathbf{p}_x, \tag{20}$$

$$\mathbf{x}_{k+1}^{(0)} = f_{oh}(\mathbf{x}_{oh,k+1}). \tag{21}$$

where the architecture is depicted in Fig. 1. Here, $T = 2N_t$ denotes the total number of iterations in a layer, the weights $\mathbf{W}_z \in \mathbb{R}^{2N_t \times 2N_t}$, $\mathbf{W}_x \in \mathbb{R}^{|\mathcal{A}|\cdot 2N_t \times 2N_t}$ and the bias $\mathbf{p}_z \in \mathbb{R}^{2N_t}$, $\mathbf{p}_x \in \mathbb{R}^{|\mathcal{A}|\cdot 2N_t}$ are trainable parameters, $\text{ReLU}(\cdot)$ is the rectified linear activation function, $\mathbf{x}_{oh} \in \{0,1\}^{|\mathcal{A}|\cdot 2N_t}$ stands for the one-hot vector mapped to $\mathbf{x} \in \mathcal{A}^{N_t}$, f_{oh} is the mapping function to transform the one-hot vector into the scalar estimate [11]. Overall, the parameters of MOID-Net optimized during the learning phase are

$$\theta = \{\mathbf{W}_z, \mathbf{W}_x, \mathbf{p}_z, \mathbf{p}_x, \omega_k\}. \tag{22}$$

Then, the learnable parameters are trained by minimizing the following mean squared error (MSE) loss function

$$l(\mathbf{x}; \hat{\mathbf{x}}) = \sum_{k=1}^{K} log(k) \parallel \mathbf{x}_l - \hat{\mathbf{x}}_k \|^2, \tag{23}$$

where $\mathbf{x}_l$ denotes the training label. Meanwhile, the outputs of all layers are taken into account in the weighted structure to alleviate the vanishing gradient problem in back-propagation (BP) procedure [14]. For a better understanding, the MOID-Net algorithm is summarized in Algorithm 3.

V. SIMULATIONS

In this section, the performance of the proposed OID and MOID algorithms for massive MIMO systems are fully investigated by simulations.

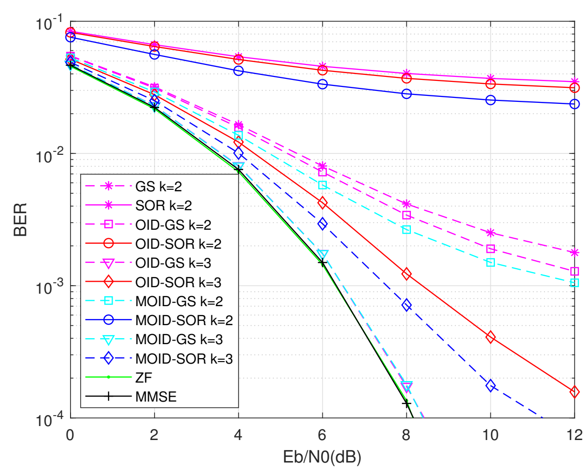

Fig. 2. Performance comparison under 64-QAM scheme of size 128 × 16.

The training procedure works on the DL library PyTorch. We draw on 10,000 data for each SNR per bit, and train MOID-Net with Adam Optimizer [15] using a batch size of 100. The learning rate is set as 0.002 and would decay by 0.97 after each epoch. In our simulations, the overall convergence of the training needs nearly 100 epochs.

The detection performance is evaluated in terms of the bit error rates (BERs). Fig. 2 demonstrates that our modifications in traditional iterative detection are reasonable. As can be seen, both the OID and MOID outperform the conventional iterative algorithm in a 128×16 MIMO system with 64-QAM while MMSE detection is applied as the baseline. To be more specific, MOID achieves a better convergence performance than OID under the same number of iterations. This is in line with the afore-mentioned analysis about the OID and MOID.

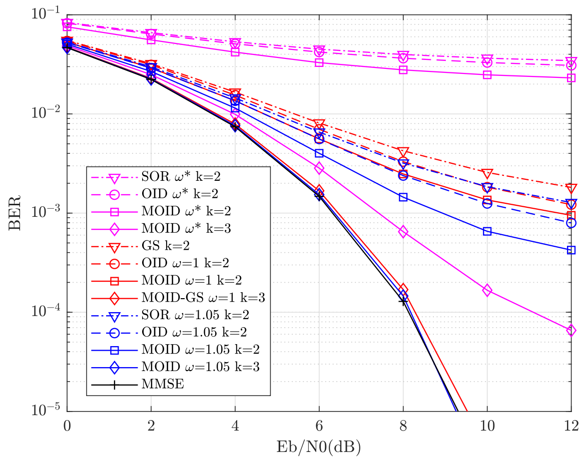

Fig. 3. Performance comparison under 64-QAM scheme of size 128×16 with different ω.

In Fig. 3, the choices of relaxation factor ω for the proposed iterative algorithm schemes are investigated. We can observe that ω has a great effect on the convergence, however, the traditional ω^* computed according to (9) and (17) is not ideal in massive MIMO systems. For this reason, we propose the MOID-Net based on DNN expansion with ω as a learnable parameter. In Fig. 4, as expected, considerable performance gain can be confirmed.

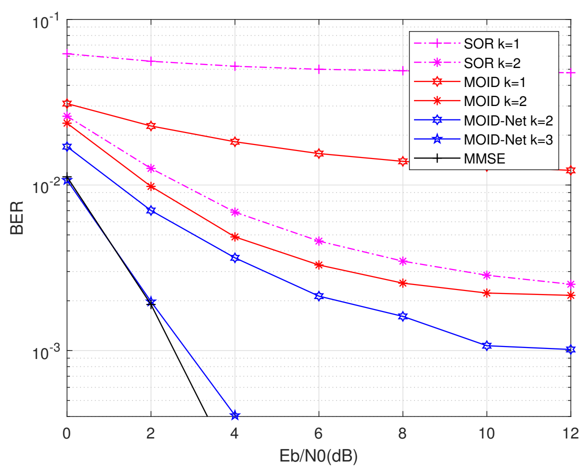

Fig. 4. Performance comparison under 16-QAM scheme of size 128×16.

On the other hand, Fig. 5 is give to illustrate the detection performance comparison in a 32×16 massive MIMO system with 4-QAM while the antenna ratio N_r/N_t gets smaller. Clearly, the proposed MOID and MOID-Net still work but more iteration numbers are needed compared to the case 128×16 with the loss of receive diversity.

VI. Acknowledgment

This work was supported in part by National Natural Science Foundation of China under Grants No. 61801216.

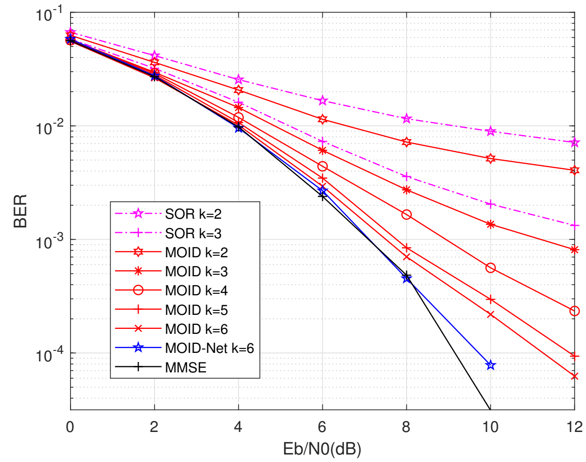

Fig. 5. Performance comparison under 4-QAM scheme of size 32×16.

References

[1] 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.

[2] E. G. Larsson, T. Marzetta, F. Rusek *et al.*, "Scaling up mimo: Opportunities and challenges with very large arrays," *Sig Process Mag IEEE*, vol. 30, no. 1, pp. 40–60, 2013.

[3] F. Jiang, C. Li, and Z. Gong, "A low complexity soft-output data detection scheme based on Jacobi method for massive MIMO uplink transmission," in *2017 IEEE international conference on communications (ICC)*. IEEE, 2017, pp. 1–5.

[4] L. Dai, X. Gao, X. Su, S. Han, I. Chih-Lin, and Z. Wang, "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, 2014.

[5] A. Yu, S. Jing, X. Tan, Z. Wu, Z. Yan, Z. Zhang, X. You, and C. Zhang, "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.

[6] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2017, pp. 1–5.

[7] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 584–588.

[8] Z. Wang, R. M. Gower, Y. Xia, L. He, and Y. Huang, "Randomized iterative methods for low-complexity large-scale mimo detection," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2934–2949, 2022.

[9] M. Mandloi, M. A. Hussain, and V. Bhatia, "Adaptive multiple stage k-best successive interference cancellation algorithm for mimo detection," *Telecommunication Systems*, vol. 66, pp. 1–16, 2017.

[10] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser mimo-ofdm systems using approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 902–915, 2014.

[11] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2554–2564, 2019.

[12] 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 *2014 IEEE Global Communications Conference*. IEEE, 2014, pp. 3291–3295.

[13] Å. Björck, *Numerical methods in matrix computations*. Springer, 2015, vol. 59.

[14] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," pp. 1310–1318, 2013.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2014.