Block Gradient Descent Algorithm for Distributed Detection in Massive MIMO

Yunjie Chen, Qiqiang Chen, Lanxin He, Zheng Wang
School of Information Science and Engineering
Southeast University, Nanjing, China
Email: yunjchen@foxmail.com, {q.chen, lanxin_he}@seu.edu.cn, wznuaa@gmail.com

***Abstract*—In this paper, a block gradient descent (BGD) detection algorithm based on ring architecture is proposed for uplink distributed massive multiple-input multiple-output (MIMO) systems. Different from the existing stochastic gradient descent (SGD) algorithm that only allows one single antenna in each distributed unit (DU), flexible numbers of DU antennas are permitted in BGD detection scheme, which is applicable to a variety of practical scenarios. Meanwhile, with dynamic step size conditionally selected for the iterative update, theoretical analysis shows that BGD detection converges to the solution of linear detection schemes. Furthermore, we analyse the advantages of the proposed algorithm in reducing complexity and bandwidth cost. Finally, simulation results are also provided to compare the performance of BGD to those of other distributed detection schemes.**

***Index Terms*—Distributed signal detection, massive MIMO, low-complexity detection, ring architecture, gradient descent algorithms.**

I. Introduction

For the future wireless communications, massive MIMO has been regarded as a key technology to highly improve the spectral and energy efficiency [1]. As the number of antennas at the base station (BS) increases, there arises a pressing challenge in transferring and computing vast volumes of raw data in centralized detection, despite state-of-the-art hardware capabilities [2], [3]. To deal with these problems, a number of distributed detection schemes with low complexity have been proposed [4]–[12], which are designed to approach the performance of the traditional linear detections like zero forcing (ZF) and minimum mean square error (MMSE).

Specifically, the decentralized baseband processing (DBP) architecture was introduced in [4]. Based on DBP, the decentralized alternating direction method of multipliers (DeADMM) and conjugate gradient (DeCG) methods have been applied for MIMO detection [5]. Moreover, to minimize the latency issue of DBP, MMSE was employed on the fully decentralized (FD) and partially decentralized (PD) architecture, respectively, which retain only an unidirectional link from DUs to the central unit (CU) [6]. Additionally, with the aid of channel hardening, the decentralized Newton (DN) detection was designed for a lower complexity implementation [7]. Compared to the parallel implementation, daisy-chain architecture has a pipelined design [8]. Based on it, the recursive least square (RLS) detection algorithm operates as ZF detection but suffers from the high complexity and data bandwidth costs. To avoid matrix multiplications, the simplified recursive least square (SRLS) [9], SGD [10], and averaged stochastic gradient descent (ASGD) [11] have been proposed. However, these iterative methods work effectively only when the number of receive antennas is sufficiently greater than that of the transmit antennas. Moreover, they constrain each DU equipped with only one single antenna, making it inapplicable to most cases of interest. The general recursive least square (GRLS) [12] algorithm is an extension of RLS, which allows multiple DU antennas. Nevertheless, in GRLS, the first several DUs are faced with high computational cost.

Corresponding author: Zheng Wang (email: wznuaa@gmail.com)

In this paper, we extend the traditional SGD algorithm to BGD algorithm, which is based on the ring architecture but with flexible number of antennas in each DU. In other words, the conventional SGD algorithm is only a special case of the proposed BGD scheme. Then, we provide a convergence analysis to demonstrate its convergence to the solution of linear detection. Moreover, the complexity and data bandwidth of BGD are also evaluated to explicate its practicality. Finally, simulation results are provided to affirm its convergence and performance compared to other low-complexity distributed detection schemes.

II. Preliminary

Consider a large-scale MIMO system with N_r receive antennas at the BS, serving different N_t single antenna users ($N_r \geqslant N_t$). Then, the input-output relationship of the system can be represented as

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

Here, $\bar{\mathbf{y}} \in \mathbb{C}^{N_r}$ denotes the received vector, and $\bar{\mathbf{x}} \in \mathcal{O}^{N_t}$ represents the transmitted vector from the discrete complex M-quadrature amplitude modulation (QAM) constellation set $\mathcal{O}^{N_t}$. $\bar{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t}$ represents the Rayleigh fading channel matrix whose entries follow $\mathcal{CN}(0,1)$ and $\bar{\mathbf{n}} \in \mathbb{C}^{N_r}$ denotes the addictive white Gaussian noise (AWGN) with zero mean and covariance matrix $\sigma_{\bar{n}}^2 \mathbf{I}_{N_r}$. The complex-valued model in (1) can be equivalently translated into a real-valued system of dimensions $2N_r \times 2N_t$ as follows:

$$\begin{bmatrix} \Re(\bar{\mathbf{y}}) \\ \Im(\bar{\mathbf{y}}) \end{bmatrix} = \begin{bmatrix} \Re(\bar{\mathbf{H}}) & -\Im(\bar{\mathbf{H}}) \\ \Im(\bar{\mathbf{H}}) & \Re(\bar{\mathbf{H}}) \end{bmatrix} \begin{bmatrix} \Re(\bar{\mathbf{x}}) \\ \Im(\bar{\mathbf{x}}) \end{bmatrix} + \begin{bmatrix} \Re(\bar{\mathbf{n}}) \\ \Im(\bar{\mathbf{n}}) \end{bmatrix}, \tag{2}$$

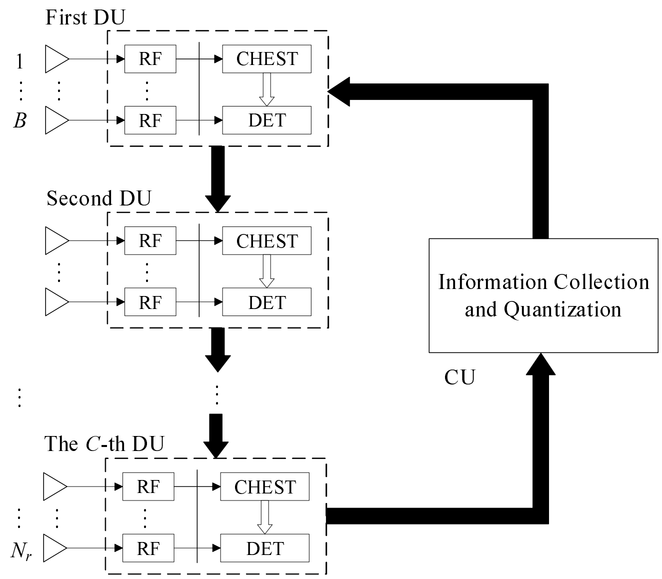

Fig. 1. Illustration of a distributed ring architecture with C DUs.

which can be succinctly expressed by

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

For simplicity of notation, let N and K denote $2N_r$ and $2N_t$ respectively, so as to $\mathbf{y} \in \mathbb{R}^N$ and $\mathbf{n} \in \mathbb{R}^N$. In this way, the complex constellation $\mathcal{O}^{N_t}$ is transformed into a real-valued $\sqrt{M}$-amplitude-shift keying (ASK) constellation set $\mathcal{X}^K$, so that we can obtain $\mathbf{x} \in \mathcal{X}^K$. Considering the equivalent channel matrix $\mathbf{H} \in \mathbb{R}^{N\times K}$ with entries distributed as $\mathcal{N}(0, 1/2)$, the traditional linear ZF detection method can be expressed as

$$\mathbf{x}_{\text{ZF}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{y} = \mathbf{A}^{-1}\mathbf{b}, \tag{4}$$

where the symmetric positive definite (SPD) matrix $\mathbf{A} = \mathbf{H}^T\mathbf{H}$ is the filtering matrix and $\mathbf{b} = \mathbf{H}^T\mathbf{y}$ is the matched filter output.

On the other hand, as for the decentralized architectures, the ring architecture was introduced as a solution to reduce both the computational complexity and data bandwidth costs of centralized detection [7]. As illustrated in Fig. 1, each DU is equipped with $B = N_r/C$ antennas and an independent computing fabric for channel estimation (CHEST) and detection (DET). The iterative process of signal detection is entirely performed in DUs, while the CU is only related to information collection in initialization and quantization before output. In distributed massive MIMO systems, the uplink received vector and channel matrix are uniformly divided into

$$\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, \cdots, \mathbf{y}_C^T\right]^T, \tag{5}$$

$$\mathbf{H} = \left[\mathbf{H}_1^T, \mathbf{H}_2^T, \cdots, \mathbf{H}_C^T\right]^T, \tag{6}$$

where $\mathbf{y}_c \in \mathbb{R}^Q$ and $\mathbf{H}_c \in \mathbb{R}^{Q\times K}$ with $Q = 2B$ represent the local received vector and corresponding channel matrix of the c-th DU, respectively. Typically, each DU updates the estimated vector serially as follows:

$$\mathbf{x}^t = h\left(\mathbf{x}^{t-1}, \gamma_t; \mathbf{y}_j, \mathbf{H}_j\right), \ \ t = 1, 2, \cdots, T_{\max}, \tag{7}$$

where $j = (t-1) \bmod C + 1$ and $T_{\max} = Ck$, k indicates the number of iteration loops over the ring topology. $h(\cdot)$ denotes a sequence of operations that depend on the specific algorithm used, and γ_t denotes relevant parameter. Additionally, the ring architecture can be transformed into daisy-chain architecture if $k = 1$. In this condition, the SGD algorithm updates $\mathbf{x}^t$ in each DU as follows [10]:

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \gamma\left(\mathbf{H}_t^T\mathbf{y}_t - \mathbf{H}_t^T\mathbf{H}_t\mathbf{x}^{t-1}\right), (Q = 2, T_{\max} = C) \tag{8}$$

with the optimal step size

$$\gamma = \frac{K}{2N}\log\left(2N \cdot \text{SNR}\right). \tag{9}$$

III. The Proposed BGD Algorithm

In this section, we first propose the BGD detection algorithm. Then, a convergence analysis is provided with respect to the choice of step size. Finally, its complexity and bandwidth are also discussed.

A. Convex Quadratic Optimization and BGD Algorithm

For low-complexity gradient descent algorithms, the linear detection solution to (3) can be obtained by solving a strictly convex quadratic optimization problem, i.e.,

$$\min_{\mathbf{x}\in\mathbb{R}^K} F(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x}, \tag{10}$$

where $\nabla F(\mathbf{x}^t) = \mathbf{A}\mathbf{x}^t - \mathbf{b}$ represents the global gradient for $\mathbf{x}^t$. Clearly, in distributed MIMO detection, each DU only calculates the local gradient instead of the global one, which is denoted by

$$\mathbf{g}_c^t = \mathbf{H}_c^T\mathbf{H}_c\mathbf{x}^t - \mathbf{H}_c^T\mathbf{y}_c = \mathbf{A}_c\mathbf{x}^t - \mathbf{b}_c. \tag{11}$$

Then, based on ring architecture, we propose the BGD detection algorithm by updating $\mathbf{x}^t$ as

$$\mathbf{x}^t = \mathbf{x}^{t-1} - \gamma_t\mathbf{g}_j^{t-1}, (Q \geqslant 2, T_{\max} = Ck) \tag{12}$$

with $j = (t-1) \bmod C + 1$. Different from the iterations of SGD shown in (8), the step size γ_t is dynamic, and the number of DU antennas can be flexibly configured for the practical need, i.e. $Q \geqslant 2$.

B. Convergence Analysis

In BGD, every DU updates the estimated vector once in one iteration loop. Approximately, this can be viewed as each DU is selected to update $\mathbf{x}^t$ with the same probability $\frac{1}{C}$. Based on this assumption, the expectation of $\mathbf{g}_j^{t-1}$ follows:

$$E\left(\mathbf{g}_j^{t-1}\right) = \frac{1}{C}\sum_{c=1}^{C}\mathbf{g}_c^{t-1} = \frac{\nabla F\left(\mathbf{x}^{t-1}\right)}{C}. \tag{13}$$

Then, by letting $\mathbf{x}^* = \mathbf{x}_{\text{ZF}} = \mathbf{A}^{-1}\mathbf{b}$ and $\delta_t = \left\|\mathbf{x}^t - \mathbf{x}^*\right\|^2$, with the assumption of uniformly sampling the index j, we can arrive at the following results about the convergence of the proposed BGD algorithm.

Lemma 1. *The proposed BGD detection algorithm for the distributed massive MIMO systems is convergent by*

$$E\left(\delta_t\right) \leqslant \phi\left(\gamma_t\right) \cdot \delta_{t-1}+\theta\left(\gamma_t\right) \tag{14}$$

with convergence rate term

$$\phi\left(\gamma_t\right)=1-\frac{2 \mu}{C} \gamma_t+\frac{L^2}{C^2} \gamma_t^2 \tag{15}$$

and error term

$$\theta\left(\gamma_t\right)=\epsilon \gamma_t^2, \tag{16}$$

where $\mu>0$ and $L>0$ represent the minimum and maximum eigenvalues of $\mathbf{A}$, respectively, and $\epsilon \geqslant 0$ is a constant.

Proof. First of all, (12) can be rewritten as

$$\mathbf{x}^t-\mathbf{x}^*=\left(\mathbf{x}^{t-1}-\mathbf{x}^*\right)-\gamma_t \mathbf{g}_j^{t-1}. \tag{17}$$

By taking the inner product on both sides of (17) with themselves, we have

$$\delta_t=\delta_{t-1}-2 \gamma_t\left\langle\mathbf{g}_j^{t-1}, \mathbf{x}^{t-1}-\mathbf{x}^*\right\rangle+\gamma_t^2\left\|\mathbf{g}_j^{t-1}\right\|^2. \tag{18}$$

Then, we take the expectation on the both sides of (18) as follows:

$$\begin{aligned} E\left(\delta_t\right)&=\delta_{t-1}-2 \gamma_t E\left(\left\langle\mathbf{g}_j^{t-1}, \mathbf{x}^{t-1}-\mathbf{x}^*\right\rangle\right)+\gamma_t^2 E\left(\left\|\mathbf{g}_j^{t-1}\right\|^2\right) \\ &\triangleq \delta_{t-1}-2 \gamma_t E_1+\gamma_t^2 E_2. \end{aligned} \tag{19}$$

After that, we try to find the bound of these two expectation terms on the right-hand side of (19), respectively. As for E_1, it follows that

$$\begin{aligned} E_1 &\stackrel{(a)}{=}\left\langle\frac{\nabla F\left(\mathbf{x}^{t-1}\right)-\nabla F\left(\mathbf{x}^*\right)}{C}, \mathbf{x}^{t-1}-\mathbf{x}^*\right\rangle \\ &=\frac{1}{C}\left\langle\mathbf{A}\left(\mathbf{x}^{t-1}-\mathbf{x}^*\right), \mathbf{x}^{t-1}-\mathbf{x}^*\right\rangle \\ &=\frac{1}{C}\left(\mathbf{x}^{t-1}-\mathbf{x}^*\right)^T \mathbf{A}\left(\mathbf{x}^{t-1}-\mathbf{x}^*\right) \\ &\stackrel{(b)}{\geqslant} \frac{\mu}{C} \delta_{t-1}. \end{aligned} \tag{20}$$

Here, equality (a) holds by (13) and $\nabla F\left(\mathbf{x}^*\right)=\mathbf{0}$. Inequality (b) follows

$$\mu \leqslant \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \tag{21}$$

for $\mathbf{x} \neq \mathbf{0}$.

As for E_2, according to Section 5.2 in [13], there exists a constant $\epsilon \geqslant 0$ such that the statistic variance of local gradient satisfies

$$E\left(\left\|\mathbf{g}_j-\frac{\nabla F(\mathbf{x})}{C}\right\|^2\right) \leqslant \epsilon. \tag{22}$$

Then, we can arrive at

$$\begin{aligned} E_2 &\leqslant E\left(\left\|\mathbf{g}_j^{t-1}-\frac{\nabla F\left(\mathbf{x}^{t-1}\right)}{C}\right\|^2\right)+\left\|\frac{\nabla F\left(\mathbf{x}^{t-1}\right)}{C}\right\|^2 \\ &\stackrel{(c)}{\leqslant} \epsilon+\frac{1}{C^2}\left\|\nabla F\left(\mathbf{x}^{t-1}\right)-\nabla F\left(\mathbf{x}^*\right)\right\|^2 \\ &\leqslant \epsilon+\frac{1}{C^2}\left(\|\mathbf{A}\| \cdot\left\|\mathbf{x}^{t-1}-\mathbf{x}^*\right\|\right)^2 \\ &\stackrel{(d)}{=} \epsilon+\frac{L^2}{C^2} \delta_{t-1}. \end{aligned} \tag{23}$$

Here, inequality (c) comes from the fact shown in (22) and equality (d) holds because $\mathbf{A}$ is a symmetric matrix.

Therefore, by casting (20) and (23) into (19), we can obtain

$$\begin{aligned} E\left(\delta_t\right) &\leqslant\left(1-\frac{2 \mu}{C} \gamma_t+\frac{L^2}{C^2} \gamma_t^2\right) \delta_{t-1}+\epsilon \gamma_t^2 \\ &=\phi\left(\gamma_t\right) \cdot \delta_{t-1}+\theta\left(\gamma_t\right), \end{aligned} \tag{24}$$

which completes the proof. □

According to Lemma 1, to guarantee the exact convergence of BGD, $0<\phi\left(\gamma_t\right)<1$ must be set, leading to

$$0<\gamma_t<\frac{2 \mu C}{L^2}. \tag{25}$$

Moreover, error term $\theta\left(\gamma_t\right)$ should gradually decay to zero. Following the selection of dynamic step size in [14], here we set γ_t as

$$\gamma_t=\frac{2 \mu \cdot \min (K / 4, C)}{L^2} \cdot \frac{C+K}{C+K+t}, \tag{26}$$

which leads to the following result.

Theorem 1. *The choice of γ_t in (26) ensures that the solution of BGD algorithm converges to $\mathbf{x}^*$, i.e.,*

$$\lim _{t \rightarrow \infty} E\left(\left\|\mathbf{x}^t-\mathbf{x}^*\right\|^2\right)=0. \tag{27}$$

Note that in massive MIMO systems, μ and L will converge to the deterministic values as N and K increase [15], which are given by

$$\mu \rightarrow \frac{N}{2}\left(1-\sqrt{\frac{K}{N}}\right)^2, L \rightarrow \frac{N}{2}\left(1+\sqrt{\frac{K}{N}}\right)^2. \tag{28}$$

Therefore, the values of γ_t can be pre-calculated with (26) and (28), to avoid solving eigenvalues of $\mathbf{A}$ directly.

To get a proper initial solution which can accelerate convergence, during pre-processing, the sum of $\mathbf{b}_c$ can be accumulated in DUs. Specifically, the first DU conveys $\mathbf{b}_1$ to the second DU, the second conveys the result of $\mathbf{b}_1+\mathbf{b}_2$ to the third, and so on. Eventually, $\mathbf{b}=\sum_{c=1}^C \mathbf{b}_c$ is transmitted to the CU. Similar to the choice of initial solution in [16], $\mathbf{x}^0$ can be obtained by

$$\mathbf{x}^0=\gamma_0 \mathbf{b}. \tag{29}$$

In summarize, the proposed distributed BGD detection algorithm for uplink massive MIMO system is presented in Algorithm 1.

Algorithm 1 The Proposed BGD Algorithm

Input: $\mathbf{H}_c, \mathbf{y}_c, c = 1, 2, \cdots, C$, $T_{\max}$

Output: estimated transmit signal $\hat{\mathbf{x}}$

Pre-processing: $\mathbf{b}_c = \mathbf{H}_c^T \mathbf{y}_c$, $\mathbf{b} = \sum_{c=1}^{C} \mathbf{b}_c$ // in each DU

Calculate μ and L via (28) // in CU

Initialization: calculate $\mathbf{x}^0$ via (26) and (29) // in CU

for $t = 1, 2, \cdots, T_{\max}$ **do**

  Update $\mathbf{x}^t$ by (12) with step size γ_t in (26) // in DU

end for

Output $\hat{\mathbf{x}} = \lceil \mathbf{x}^{T_{\max}} \rfloor_{\mathcal{Q}} \in \mathcal{X}^K$ // in CU

Additionally, $\mathbf{x}^*$ can be extended to MMSE solution via some simple transformations. One possible way is to extend $\mathbf{H}$ and $\mathbf{y}$ as follows [17]:

$$\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \sigma_n \mathbf{I}_K \end{bmatrix} \text{ and } \underline{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_K \end{bmatrix}, \tag{30}$$

where $\sigma_n = \sigma_{\bar{n}}/\sqrt{2}$. In this way, one can obtain

$$\mathbf{x}^* = \mathbf{x}_{\text{MMSE}} = \left(\underline{\mathbf{H}}^T \underline{\mathbf{H}}\right)^{-1} \underline{\mathbf{H}}^T \underline{\mathbf{y}} = \left(\mathbf{H}^T \mathbf{H} + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{H}^T \mathbf{y}, \tag{31}$$

so that $\mathbf{x}^t$ will converge to the MMSE detection result.

C. Complexity and Bandwidth Analysis

We now investigate the computational complexity and data bandwidth of the proposed BGD detection algorithm, where the computational complexity is evaluated in terms of the required number of real multiplications.

In particular, during the pre-processing and initialization stage, the computational complexity of computing $\mathbf{b}_c = \mathbf{H}_c^T \mathbf{y}_c$ in each DU is $CQK = NK$, the complexity of obtaining the initial solution $\mathbf{x}^0$ is K. As for the iteration in (12), there are two approaches for DUs to compute $\mathbf{A}_j \mathbf{x}^{t-1} = \mathbf{H}_j^T \mathbf{H}_j \mathbf{x}^{t-1}$ as follows:

- First calculate $\mathbf{A}_j$, and then obtain $\mathbf{A}_j \mathbf{x}^{t-1}$, which costs $QK^2 + K^2$. $\mathbf{A}_j$ can be reused for the subsequent iteration loops so that the cost to update $\mathbf{x}^t$ decreases to K^2.
- First multiply $\mathbf{H}_j$ with $\mathbf{x}^{t-1}$, and then multiply $\mathbf{H}_j^T$ with $\mathbf{H}_j \mathbf{x}^{t-1}$, which costs $2QK$ for every iteration.

The first approach is employed in algorithms over the star topology for lower latency. However, in pipelined signal processing, the second approach is recommended for lower computational complexity, especially when $K \geqslant 2Q$. Here we adopt the second one, which can also be used in other gradient descent algorithms like SGD and ASGD. Moreover, the complexity of multiplying γ_t with $\mathbf{g}_j^{t-1}$ needs to be considered, which is K, while the complexity of other operations is negligible.

In summarize, the total computational complexity of the proposed BGD detection algorithm with $T_{\max}$ times iterations is $(N+1)K + (2QK + K)Ck$, namely $\mathcal{O}(NK \cdot k)$, which is rather competitive compared to other distributed detection schemes. On the other hand, as only the vector $\mathbf{x}^t \in \mathbb{R}^K$ needs to be unidirectionally conveyed through DUs, the data bandwidth cost is K, which is also appealing.

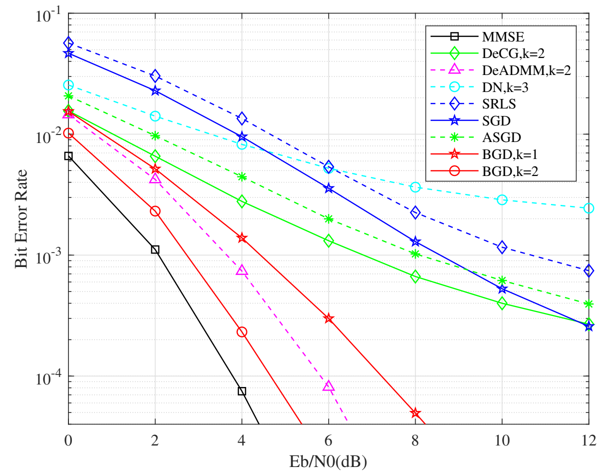

Fig. 2. BER comparison in 128×16 massive MIMO system with $C = 16$ using 16-QAM.

IV. Simulation

In this section, a comprehensive simulation study is conducted to assess the performance of the BGD detection algorithm, in terms of bit error rate (BER), and computational complexity.

In Fig. 2, the BER performance comparison among various low-complexity distributed detection schemes is shown in an uncoded 128×16 massive MIMO system using 16-QAM. Here, the number of DUs, i.e. C, in schemes of DeCG [4], DeADMM [5], DN [7] and BGD is set as 16. Different from them, SRLS [9], SGD [10] and ASGD [11] schemes restrict $C = N_r = 128$, which are deemed as a default setup in the following simulation. As for SRLS, SGD and ASGD schemes, k is set as 1 due to the daisy-chain architecture. To make a fair comparison, the matched filter output vector $\mathbf{b}$ is pre-calculated and employed in the initialization of each algorithm. Moreover, the BER performance of centralized MMSE is also given to serve as a performance benchmark. As shown in Fig. 2, the proposed BGD scheme with $k = 2$ achieves a better detection performance than other distributed schemes, which is near-optimal for MMSE detection. Compared to the traditional SGD scheme, the proposed BGD algorithm obtains considerable performance gain owing to the usage of dynamic step size.

Fig. 3 further extends the BER performance comparison of these schemes to the 128×32 massive MIMO system using 16-QAM. Different from Fig. 2, $C = 32$ is applied in schemes of DeCG, DeADMM, DN and BGD. In particular, compared to the case of 128×16, the performance of all the detection schemes degrade as the transmit antennas increase. Nevertheless, the proposed BGD algorithm approach the MMSE performance with $k = 3$ and still outperforms other schemes under a different antenna configuration in DUs. Overall, BGD maintains its high convergence rate without the condition $N \gg K$ and has flexibility on the number of DU antennas.

In Fig. 4, the complexity comparison among different distributed detection schemes is presented with respect to

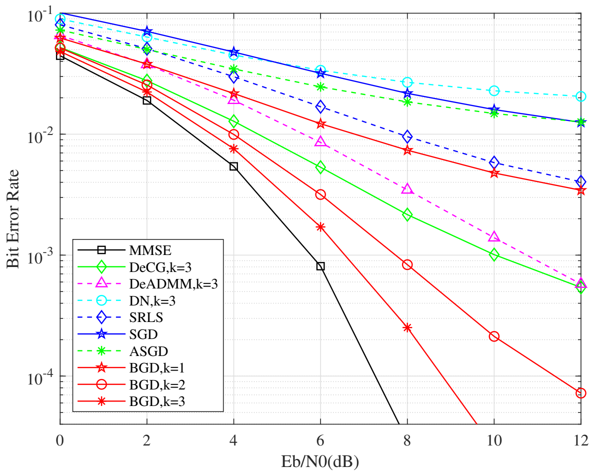

Fig. 3. BER comparison in 128×32 massive MIMO system with $C = 32$ using 16-QAM.

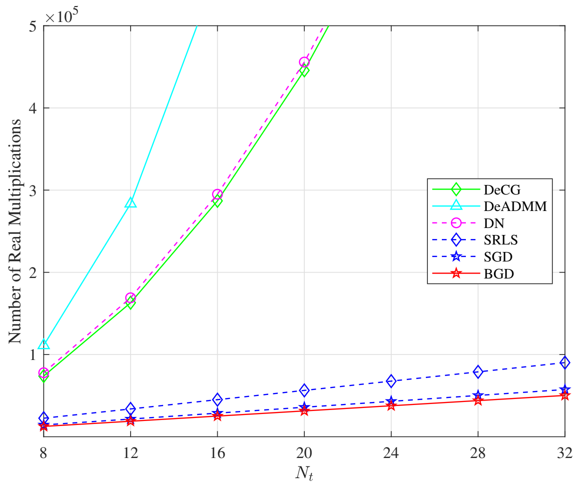

Fig. 4. Complexity comparison in terms of the number of real multiplications for $128 \times N_t$ massive MIMO systems with $C = 16$ and $k = 1$.

$128 \times N_t$ massive MIMO systems. Here, $C = 16$ is fixed and the computational complexity is evaluated by the number of real multiplications in a single iteration loop (i.e. $k = 1$) for all algorithms. To be specific, the computational cost of DeADMM is quite high due to matrix inversion, which turns out to be unaffordable in practice. CG and DN have the similar complexity, which is much higher than SRLS and SGD. As the deployment of multiple DU antennas is allowed, BGD can have less iterations in one iteration loop, which makes its complexity cost lower than SRLS and SGD. Therefore, the proposed BGD is quite a competitive algorithm for distributed massive MIMO detection.

V. Conclusion

In this paper, by extending and improving the existing SGD detection algorithm, we proposed a novel low-complexity algorithm, namely BGD, for distributed massive MIMO detection. The convergence, complexity, and bandwidth analysis for the proposed BGD detection algorithm are also provided. When the dynamic step size γ_t is conditionally selected, BGD can rapidly converge to the linear detection solution. Finally, the simulation results not only confirm its convergence superiority but also demonstrate its advantages in complexity and BER performance, compared to other distributed detection algorithms.

Acknowledgment

This work is supported in part by the National Key R&D Program of China under Grants No.2023YFC2205501, and in part by National Natural Science Foundation of China under Grants No.62371124.

References

[1] J. Tian, Y. Han, S. Jin, J. Zhang, and J. Wang, “Analytical channel modeling: From mimo to extra large-scale mimo,” *Chinese Journal of Electronics*, vol. 34, no. 1, pp. 1–15, 2025.

[2] M. A. Albreem, A. Alhabbash, A. M. Abu-Hudrouss, and T. A. Almohamad, “Data detection in decentralized and distributed massive MIMO networks,” *Comput. Commun.*, vol. 189, pp. 79–99, 2022.

[3] N. Li and P. Fan, “Distributed cell-free massive mimo versus cellular massive mimo under ue hardware impairments,” *Chinese Journal of Electronics*, vol. 33, no. 5, pp. 1274–1285, 2024.

[4] K. Li, Y. Chen, R. Sharan, T. Goldstein, J. R. Cavallaro, and C. Studer, “Decentralized data detection for massive MU-MIMO on a Xeon Phi cluster,” in *Proc 50th Asilomar Conf. Signals Syst. Comput.*, 2016, pp. 468–472.

[5] K. Li, R. R. Sharan, Y. Chen, T. Goldstein, J. R. Cavallaro, and C. Studer, “Decentralized baseband processing for massive MU-MIMO systems,” *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 7, no. 4, pp. 491–507, 2017.

[6] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, “Decentralized equalization with feedforward architectures for massive MU-MIMO,” *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4418–4432, 2019.

[7] A. Kulkarni, M. A. Ouameur, and D. Massicotte, “Hardware topologies for decentralized large-scale MIMO detection using newton method,” *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 68, no. 9, pp. 3732–3745, 2021.

[8] J. R. Sanchez, F. Rusek, M. Sarajlic, O. Edfors, and L. Liu, “Fully decentralized massive MIMO detection based on recursive methods,” in *Proc IEEE Int. Workshop Signal Process. Syst. SiPS*, 2018, pp. 53–58.

[9] Q. Liu, H. Liu, Y. Yan, and P. Wu, “A distributed detection algorithm for uplink massive MIMO systems,” in *Proc IEEE Int. Workshop Signal Process. Syst. SiPS*, 2019, pp. 213–217.

[10] J. R. Sanchez, F. Rusek, O. Edfors, M. Sarajlic, and L. Liu, “Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms,” *IEEE Trans. Signal Process.*, vol. 68, pp. 687–700, 2020.

[11] J. R. Sanchez, F. Rusek, M. Sarajlic, O. Edfors, and L. Liu, “Fully decentralized massive MIMO detection based on recursive methods,” in *Proc IEEE Int. Workshop Signal Process. Syst. SiPS*, 2018, pp. 53–58.

[12] Q. Chen, Z. Wang, C. Ma, X. Dai, and D. W. K. Ng, “General recursive least square algorithm for distributed detection in massive MIMO,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 8, pp. 12 137–12 142, 2024.

[13] L. Bottou, F. E. Curtis, and J. Nocedal, “Optimization methods for large-scale machine learning,” *SIAM Rev.*, vol. 60, no. 2, pp. 223–311, 2018.

[14] Z. Wang, C. Pan, Y. Huang, S. Jin, and G. Caire, “Randomized iterative algorithms for distributed massive MIMO detection,” *IEEE Trans. Signal Process.*, vol. 73, pp. 2304–2319, 2025.

[15] A. Edelman, “Eigenvalues and condition numbers of random matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 4, pp. 543–560, 1988.

[16] J. Tu, M. Lou, J. Jiang, D. Shu, and G. He, “An efficient massive MIMO detector based on second-order richardson iteration: From algorithm to flexible architecture,” *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 67, no. 11, pp. 4015–4028, 2020.

[17] D. Wubben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, “Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice-reduction,” in *Proc IEEE Int. Conf. Commun.*, vol. 2, 2004, pp. 798–802.