General Recursive Least Square Algorithm For Distributed Detection In Massive MIMO

Qiqiang Chen, Zheng Wang, *Senior Member, IEEE*, Cong Ma, Xiaoming Dai, *Member, IEEE*, and Derrick Wing Kwan Ng, *Fellow, IEEE*

***Abstract*—In this paper, a general recursive least square (GRLS) detection algorithm is proposed for the uplink of distributed massive multiple-input multiple-output (MIMO) to alleviate the bottlenecks in both computational complexity and data bandwidth for interconnection. Different from the existing recursive least square (RLS) detection algorithm which only supports a single antenna in each distributed unit (DU), the proposed GRLS allows for multiple antennas in each DU, rendering it adaptable to a variety of practical scenarios. Moreover, among the total C DUs and with an integer parameter C_0, the complexity of $C-C_0$ DUs in GRLS can be significantly reduced by leveraging the channel hardening property. Through analysis, we demonstrate that the convergence of the GRLS algorithm is guaranteed if $C_0 \geq \left\lfloor \left(\sqrt{B/2}+\sqrt{K}\right)^2/B \right\rfloor$ holds, where K and B denote the numbers of antennas at the user side and each DU, respectively. Furthermore, based on the daisy-chain architecture, the proposed GRLS algorithm also enjoys excellent scalability, which can be easily extended with extra DUs for further improvement. Finally, the computational complexity and data bandwidth analysis are provided to unveil the superiority of GRLS compared to other distributed detection schemes for massive MIMO.**

***Index Terms*—Massive MIMO, distributed MIMO detection, decentralized signal detection, daisy-chain, RLS.**

I. Introduction

Due to its promising capacity, ultra-fast data rate, and high energy efficiency, massive MIMO has become a core technology for enabling beyond fifth-generation (5G) and sixth-generation (6G) wireless communications [1]. However, most of the existing detection schemes for massive MIMO are commonly implemented in a centralized manner. As the number of antennas at the base station (BS) increases to hundreds or thousands, there arises a pressing challenge in transferring the vast volumes of raw data for advanced signal processing, even with the state-of-the-art hardware capabilities [2]. Meanwhile, the rapid growths in both computational complexity and data storage requirement also render the single computing fabric difficult to satisfy practical demands. To this end, a number of distributed detection schemes for massive MIMO have been proposed [2]–[9].

Specifically, the decentralized baseband processing (DBP) architecture was introduced in [2]. It partitions the BS antennas into C individual DUs, where each DU contains B antennas and is equipped with an independent computing fabric. Based on DBP, the alternating direction method of multipliers (ADMM), conjugate gradient (CG) methods [3], and expectation propagation algorithm (EPA) [4] can be applied for distributed massive MIMO detection but at a high data bandwidth cost. Besides, the minimum mean square error (FD-MMSE) [5] and the Gaussian message passing (GMP) [6] algorithms were employed on the fully decentralized architecture, which retains only an unidirectional link from DUs to the central processing unit (CPU), thereby further reducing the data bandwidth cost at the expense of practical performance degradation. Furthermore, with the aid of channel hardening, the decentralized Newton (DN) detection was designed for a lower complexity implementation [7]. However, all of these methods require a large number of antennas in DU, i.e., $B \gg 1$, and rely on the CPU to process the partial results. In contrast, the daisy-chain architecture, presented in [8], [9], directly outputs the detection results to the CPU, thus freeing it from the detection process. Based on the daisy-chain architecture, the RLS detection algorithm operates as the traditional zero forcing (ZF) detection but in a distributed fashion. Nevertheless, this approach not only suffers from the high complexity and bandwidth costs, but also confine to DUs equipped with only a single antenna, rendering it impractical in the most of cases. Unfortunately, these issues persist in other detection methods, such as the stochastic gradient descent (SGD) and the averaged stochastic gradient descent (ASGD) detection schemes [8].

In this paper, we extend the distributed RLS detection to a more generalized one, named as GRLS, allowing multiple antennas to be deployed at each DU. Then, by exploiting the channel hardening property in massive MIMO, those computationally expensive operations, such as matrix multiplication and inversion, in the last $C - C_0$ DUs in GRLS can be greatly simplified, resulting in remarkable reductions in both complexity and data bandwidth. Meanwhile, to ensure the convergence of GRLS, we also provide a convergence analysis concerning the choice of C_0. Furthermore, the complexity and data bandwidth of GRLS are given to validate its advantages over various existing distributed massive MIMO detection schemes. Finally, simulation results are provided to affirm its improved trade-off between performance and complexity, as well as its notable scalability to very large antenna arrays,

This work was supported in part by the National Natural Science Foundation of China under Grants 61771124, 62371037, in part by the ZTE Industry-University-Institute Cooperation Funds under Grant No. HC-CN-20200628010. *(Corresponding author: Zheng Wang.)*

Q. Chen and Z. Wang are with School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: zheng_wang@seu.edu.cn). C. Ma is with ZTE Corporation, Shenzhen 518057, China (e-mail:ma.cong@zte.com.cn). X. Dai is with the University of Science and Technology Beijing, Beijing 100083, China (e-mail: daixiaoming@ustb.edu.cn). D. W. K. Ng is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: w.k.ng@unsw.edu.au).

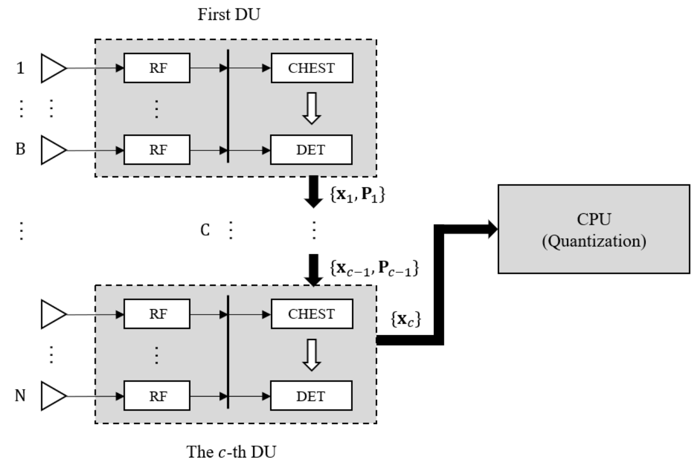


Fig. 1. Illustration of a decentralized daisy-chain architecture with C DUs. Each DU is equipped with $B = N/C$ antennas and an independent computing fabric for channel estimation (CHEST) and detection (DET), while quantization ($\mathcal{Q}$) is performed centrally.

i.e., XL-MIMO [10].

II. RLS Detection Algorithm based on Decentralized Daisy-chain Architecture

Considering a massive MIMO scenario with N antennas at the BS that serves K single antenna users $(N \gg K)$, the input-output relation for the uplink is

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

Here, $\mathbf{y} \in \mathbb{C}^N$ is the received vector, $\mathbf{x} \in \mathcal{O}^K$ is the transmitted vector from the discrete QAM constellation $\mathcal{O}^K$, $\mathbf{H} \in \mathbb{C}^{N \times K}$ represents the Rayleigh fading channel matrix whose entries follow $\mathcal{CN}(0,1)$ and $\mathbf{n} \in \mathbb{C}^N$ denotes the addictive white Gaussian noise (AWGN) with mean $\mathbf{0}$ and covariance matrix $\sigma^2 \mathbf{I}_N$. Given the system model in (1), the traditional linear detection methods, i.e., ZF and MMSE are presented as

$$\mathbf{x}_{\text{ZF}} = \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{y}, \tag{2}$$

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

where the final detection output $\widehat{\mathbf{x}}$ is acquired by quantizing $\widehat{\mathbf{x}} = \lceil \mathbf{x}_{\text{ZF}} \rfloor_{\mathcal{Q}} \in \mathcal{O}^K$ or $\widehat{\mathbf{x}} = \lceil \mathbf{x}_{\text{MMSE}} \rfloor_{\mathcal{Q}} \in \mathcal{O}^K$. Theoretically, due to the *favorable propagation* in massive MIMO systems, the optimal maximum likelihood (ML) detection performance can be approximated by ZF or MMSE [1].

On the other hand, based on the *daisy-chain* architecture shown in Fig. 1, the RLS detection algorithm serves as a recursive version of ZF detection by [8]

$$\mathbf{x}_c = \mathbf{x}_{c-1} + \mathbf{P}_c \mathbf{h}_c^H (y_c - \mathbf{h}_c \mathbf{x}_{c-1}), \tag{4}$$

where

$$\mathbf{P}_c = \mathbf{P}_{c-1} - \frac{\mathbf{P}_{c-1} \mathbf{h}_c^H \mathbf{h}_c \mathbf{P}_{c-1}^H}{1 + \mathbf{h}_c \mathbf{P}_{c-1} \mathbf{h}_c^H} \in \mathbb{C}^{K \times K} \tag{5}$$

is the weight matrix. Here, $\mathbf{h}_c \in \mathbb{C}^{1 \times K}$, y_c, and $\mathbf{x}_c \in \mathbb{C}^K$ refer to the local channel information, received signal, and estimated vector at the c-th DU, respectively.

However, the RLS detection method restricts each DU in daisy-chain to equip with only one single antenna, i.e., $B = 1$, which severely limits its piratical applications. Moreover, the matrix multiplications in computing $\mathbf{P}_c$ in (5) also incur a high complexity cost. In addition, updating $\mathbf{x}_c$ on the c-th DU requires $\mathbf{P}_{c-1}$ and $\mathbf{x}_{c-1}$ from the previous DU, which accounts for the high data bandwidth consumption by conveying a $K \times K$ matrix and a $K \times 1$ vector, respectively.

III. The Proposed GRLS Detection Algorithm

A. Extension of the Traditional RLS Algorithm

We now extend the existing RLS detection algorithm to a more generalized one such that it allows for having multiple antennas at each DU over iterations as

$$\mathbf{x}_c = \mathbf{x}_{c-1} + \mathbf{P}_c \mathbf{H}_c^H (\mathbf{y}_c - \mathbf{H}_c \mathbf{x}_{c-1}), \tag{6}$$

with

$$\mathbf{R}_c = \left(\mathbf{I}_B + \mathbf{H}_c \mathbf{P}_{c-1} \mathbf{H}_c^H\right)^{-1} \tag{7}$$

and

$$\mathbf{P}_c = \mathbf{P}_{c-1} - \mathbf{P}_{c-1} \mathbf{H}_c^H \mathbf{R}_c \mathbf{H}_c \mathbf{P}_{c-1}^H, \tag{8}$$

where $\mathbf{H}_c \in \mathbb{C}^{B \times K}$ and $\mathbf{y}_c \in \mathbb{C}^B$ with $B \geq 1$. Meanwhile, the weight matrix $\mathbf{P}_c$ is used to approximate $(\mathbf{H}_{\text{Acum}}^H \mathbf{H}_{\text{Acum}})^{-1}$, while $\mathbf{H}_{\text{Acum}} = [\mathbf{H}_1; \mathbf{H}_2; \ldots; \mathbf{H}_c] \in \mathbb{C}^{N_c \times K}$ denotes the accumulated channel information from the previous $N_c = c \times B$ received antennas. Note that matrix $\mathbf{H}_{\text{Acum}}$ is equivalent to $\mathbf{H}$ and N_c is equal to N when $c = C$. Furthermore, according to (6), (7), and (8), $\mathbf{x}_c$ can be reformulated as

$$\mathbf{x}_c = [\mathbf{P}_0^{-1} + \sum_{i=1}^{c} \mathbf{H}_i^H \mathbf{H}_i]^{-1} [\mathbf{P}_0^{-1} \mathbf{x}_0 + \sum_{i=1}^{c} \mathbf{H}_i^H \mathbf{y}_i], \tag{9}$$

which naturally leads to the following convergence result.

Theorem 1. *Based on the iterations in (6), (7), and (8),* $\mathbf{x}_c$ *will gradually converge to the MMSE detection solution in* (3) *with the initial setup* $\mathbf{x}_0 = \mathbf{0}$ *and* $\mathbf{P}_0 = \frac{1}{\sigma^2} \mathbf{I}_K$.

Concerning the least squares problem, the least absolute shrinkage and selection operator (LASSO) [11] and box-relaxation optimization (BRO) [12] methods can be applied for further improvement. However, as shown in (6), (7), and (8), each iteration of such an extension of RLS involves the complicated operations such as matrix multiplication and matrix inversion. In order to effectively reduce the computational complexity burden, we aim to incorporate the property of channel hardening into the iterations.

B. Complexity Reduction by Adopting Channel Hardening

With $N \gg K$, the Gram matrix $\mathbf{H}^H \mathbf{H}$ is diagonally dominant due to channel hardening, which is one of the key properties in massive MIMO systems [1]. Here, since the weight matrix $\mathbf{P}_c$ is an approximation of the inverse of $\mathbf{H}_{\text{Acum}}^H \mathbf{H}_{\text{Acum}}$, it is supposed to be diagonally dominant as well when c is sufficiently large, which inspires us to construct the matrix $\mathbf{P}_c$ only by its diagonal elements, i.e.,

$$\bar{\mathbf{P}}_c = \text{diag}\left(\text{diag}\left(\bar{\mathbf{P}}_{c-1} - \bar{\mathbf{P}}_{c-1} \mathbf{H}_c^H \bar{\mathbf{R}}_c \mathbf{H}_c \bar{\mathbf{P}}_{c-1}^H\right)\right). \tag{10}$$

Algorithm 1: The Proposed GRLS Detection Algorithm

Input : $\mathbf{y}_c$, $\mathbf{H}_c$, $c = 1, 2, \cdots C, \sigma^2, \beta$ and C_0
Output : estimated transmit signal $\widehat{\mathbf{x}}$

Initialize: $\mathbf{x}_0 = \mathbf{0}, \mathbf{P}_0 = \frac{1}{\sigma^2}\mathbf{I}_K$
for $c = 1, 2, \cdots, C_0$ **do**
  $\mathbf{R}_c = \left(\mathbf{I}_B + \mathbf{H}_c\mathbf{P}_{c-1}\mathbf{H}_c^H\right)^{-1}$
  $\mathbf{P}_c = \mathbf{P}_{c-1} - \mathbf{P}_{c-1}\mathbf{H}_c^H\mathbf{R}_c\mathbf{H}_c\mathbf{P}_{c-1}^H$
  $\mathbf{x}_c = \mathbf{x}_{c-1} + \mathbf{P}_c\mathbf{H}_c^H(\mathbf{y}_c - \mathbf{H}_c\mathbf{x}_{c-1})$
  if $c = C_0$ **then**
    $\bar{\mathbf{P}}_c = \text{diag}\left(\text{diag}\left(\mathbf{P}_c\right)\right)$
  end if
end for
for $c = C_0 + 1, C_0 + 2 \cdots, C$ **do**
  $\bar{\mathbf{R}}_c = \text{diag}\left(\text{diag}\left(\left(\mathbf{I}_B + \mathbf{H}_c\bar{\mathbf{P}}_{c-1}\mathbf{H}_c^H\right)^{-1}\right)\right)$
  $\bar{\mathbf{P}}_c = \text{diag}\left(\text{diag}\left(\bar{\mathbf{P}}_{c-1} - \bar{\mathbf{P}}_{c-1}\mathbf{H}_c^H\bar{\mathbf{R}}_c\mathbf{H}_c\bar{\mathbf{P}}_{c-1}^H\right)\right)$
  $\mathbf{x}_c = \mathbf{x}_{c-1} + (1-\beta)\left(\bar{\mathbf{P}}_c\mathbf{H}_c^H(\mathbf{y}_c - \mathbf{H}_c\mathbf{x}_{c-1})\right)$
end for
output $\widehat{\mathbf{x}} = \lceil \mathbf{x}_C \rfloor_{\mathcal{Q}} \in \mathcal{O}^K$

Similarly, the matrix $\mathbf{R}_c$ in (7) can also be approximated as

$$\bar{\mathbf{R}}_c = \text{diag}\left(\text{diag}\left(\left(\mathbf{I}_B + \mathbf{H}_c\bar{\mathbf{P}}_{c-1}\mathbf{H}_c^H\right)^{-1}\right)\right). \tag{11}$$

Thanks to the diagonal structure of matrices $\bar{\mathbf{P}}_c$ and $\bar{\mathbf{R}}_c$, significant reductions in both computational complexity and data bandwidth can be achieved. On one hand, the computation of the matrix inversion is simplified as computing only the reciprocal of the diagonal elements, while the matrix multiplication is also simplified as only the diagonal elements need to be preserved. On the other hand, the $K \times K$ diagonal matrix $\bar{\mathbf{P}}_c$, which needs to convey by each DU, can be refined as a $K \times 1$ diagonal vector, thus greatly reducing the required data bandwidth.

C. Accumulation of Channel Information

However, due to the daisy-chain architecture, the matrix $\mathbf{P}_c$ at the c-th DU only contains the information from the previous N_c received antennas. As a result, the diagonal elements of $\mathbf{P}_c$ become dominant gradually with the increment of c, thus rendering the complexity reduction driven by channel hardening ineffective during the early stages in daisy-chain.

To address this issue, we divide the iterations of the proposed GRLS into two stages:

- *The first C_0 DUs*: Due to the lack of channel information for exploiting the channel hardening, perform the iteration in (6) based on $\mathbf{R}_c$ in (7) and $\mathbf{P}_c$ in (8).
- *The rest of $C - C_0$ DUs*: Based on the accumulated channel information for channel hardening, perform the iteration in (6) based on $\bar{\mathbf{R}}_c$ in (11) and $\bar{\mathbf{P}}_c$ in (10).

In addition, for better detection performance, the technique of damping factor $\beta \in (0, 1)$ (i.e., here we use $\beta = 0.1$) is also introduced to the second stage by

$$\mathbf{x}_c = \beta\mathbf{x}_{c-1} + (1-\beta)\,\mathbf{x}_c. \tag{12}$$

Finally, at the CPU, the detected signal

$$\widehat{\mathbf{x}} = \lceil \mathbf{x}_C \rfloor_{\mathcal{Q}} \in \mathcal{O}^K \tag{13}$$

is outputted as the detection solution of GRLS.

Based on the daisy-chain architecture, the GRLS algorithm employs local processing of data where it is generated to avoid the additional overhead for transmission and storage. Furthermore, leveraging its sequential, regular, and modular structure, extending GRLS to larger antenna arrays can be easily achieved by increasing the number of DUs, which demonstrates its excellent scalability. To summarize, the proposed general recursive least square (GRLS) detection algorithm for distributed uplink massive MIMO systems is outlined in Algorithm 1.

IV. Convergence Analysis

Undoubtedly, how to reasonably set the parameter C_0 is the key to GRLS. In what follows, we provide the convergence analysis with respect to the choice of C_0.

Lemma 1. *For the flat Rayleigh fading matrix $\mathbf{H}_c \in \mathbb{C}^{B\times K}$ whose entries follow $\mathcal{CN}(0, 1)$, it follows that*

$$\mathbf{A}_c = \text{E}[\mathbf{H}_c^H\mathbf{H}_c] = B\mathbf{I}_K. \tag{14}$$

Proof. To start with, let h_{mn} denote the element in the m-th row and n-th column of matrix $\mathbf{H}_c$. Meanwhile, define $\mathbf{D}_c = \mathbf{H}_c^H\mathbf{H}_c$ with its elements $d_{mn} = \sum_{l=1}^{B} h_{lm}^* h_{ln}$. Then, in the case of $m = n$, we can find that $h_{lm}^* h_{lm} = \Re(h_{lm})^2 + \Im(h_{lm})^2$, while $\Re(h_{lm})^2$ and $\Im(h_{lm})^2$ both follow the *Gamma distribution* with the shape parameter $\alpha = 0.5$ and the scale parameter $\gamma = 1$ [13] (i.e., $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts, respectively), namely,

$$h_{lm}^* h_{lm} \sim \Gamma(0.5, 1) + \Gamma(0.5, 1) = \Gamma(1, 1). \tag{15}$$

Building upon it, d_{mm}, as the summation of $h_{lm}^* h_{lm}$, obeys the following distribution

$$d_{mm} = \sum_{l=1}^{B} h_{lm}^* h_{lm} \sim \Gamma(B, 1). \tag{16}$$

Subsequently, in the case of $m \neq n$, h_{lm} and h_{ln} are independent entries with mean zero, and this leads to

$$\text{E}\left[d_{mn}\right] = \begin{cases} B & \text{if } m = n, \\ 0 & \text{if } m \neq n, \end{cases} \tag{17}$$

which completes the proof. □

Theorem 2. *The proposed GRLS detection algorithm for the distributed massive MIMO systems converges when*

$$C_0 \geq \left\lfloor \left(\sqrt{\frac{B}{2}} + \sqrt{K}\right)^2 / B \right\rfloor, \tag{18}$$

where $\lfloor x \rfloor$ rounds to the closest integer smaller than or equal to x.

Proof. According to Theorem 1, the convergence of the first stage about the first C_0 DUs in GRLS is guaranteed unconditionally. Therefore, we pay our attentions on the convergence of the second stage for the rest of $C - C_0$ DUs.

In particular, we can express the error in the second stage (i.e., $c > C_0$) as

$$\begin{aligned}\mathbf{e}_C =& \mathbf{x}_C - \mathbf{x}_{\text{MMSE}} \\ =& \mathbf{x}_{C-1} + \bar{\mathbf{P}}_C \mathbf{H}_C^H (\mathbf{y}_C - \mathbf{H}_C \mathbf{x}_{C-1}) - \mathbf{x}_{\text{MMSE}} \\ =& (\mathbf{I}_K - \bar{\mathbf{P}}_C \mathbf{H}_C^H \mathbf{H}_C)(\mathbf{x}_{C-1} - \mathbf{x}_{\text{MMSE}}) + \bar{\mathbf{P}}_C \mathbf{H}_C^H (\mathbf{y}_C - \mathbf{H}_C \mathbf{x}_{\text{MMSE}}) \\ =& (\mathbf{I}_K - \bar{\mathbf{P}}_C \mathbf{H}_C^H \mathbf{H}_C) \mathbf{e}_{C-1} + \bar{\mathbf{P}}_C \mathbf{H}_C^H \mathbf{n}_C. \end{aligned} \tag{19}$$

Moreover, given the fact that $\mathbf{H}_c$ is statistically independent of $\bar{\mathbf{P}}_c$ and $\mathbf{e}_c$, then based on (14) in Lemma 1, by taking expectation on (19), we have

$$\begin{aligned}\mathrm{E}[\mathbf{e}_C] &= \mathrm{E}[(\mathbf{I}_K - \bar{\mathbf{P}}_C \mathbf{H}_C^H \mathbf{H}_C)\mathbf{e}_{C-1}] + \mathrm{E}[\bar{\mathbf{P}}_C \mathbf{H}_C^H \mathbf{n}_C] \\ &= (\mathbf{I}_K - \bar{\mathbf{P}}_C \mathbf{A}_C) \mathrm{E}[\mathbf{e}_{C-1}] \\ &\triangleq \prod_{c=C_0+1}^{C} \mathbf{F}_c \mathrm{E}[\mathbf{e}_{C_0}], \end{aligned} \tag{20}$$

where $\mathbf{F}_c \triangleq \mathbf{I}_K - B\bar{\mathbf{P}}_c \in \mathbb{C}^{K\times K}$ is the *iteration matrix*. Then, to guarantee the convergence of (20) for diminishing the error over the iterations, the spectral radius of $\mathbf{F}_c$ should be smaller than 1 [14], namely,

$$\rho(\mathbf{F}_c) = \max_{1\le k\le K} |\lambda_k(\mathbf{F}_c)| = \max_{1\le k\le K} |1 - B\lambda_k(\bar{\mathbf{P}}_c)| < 1, \tag{21}$$

where $\lambda_k(\cdot)$ denotes the k-th eigenvalue. Since the matrix $\bar{\mathbf{P}}_c$ is approximated to the inverse of a positive-definite matrix $\mathbf{H}_{\text{Acum}}^H \mathbf{H}_{\text{Acum}}$, its eigenvalues are all larger than 0 [1], which implies that the condition in (21) can be further expressed by

$$\lambda_{\max}(\bar{\mathbf{P}}_c) < \frac{2}{B}. \tag{22}$$

On the other hand, at the c-th DU, $\lambda_{\min}\left(\left(\mathbf{H}_{\text{Acum}}^H \mathbf{H}_{\text{Acum}}\right)^{-1}\right)$ approaches $N_c\left(1-\sqrt{K/N_c}\right)^2$ with $N_c \gg K$ [14] such that the following approximation holds

$$\lambda_{\max}\left(\bar{\mathbf{P}}_c\right) \approx \frac{1}{\lambda_{\min}\left(\left(\mathbf{H}_{\text{Acum}}^H \mathbf{H}_{\text{Acum}}\right)^{-1}\right)} \approx \frac{1}{N_c\left(1-\sqrt{\frac{K}{N_c}}\right)^2}. \tag{23}$$

Therefore, by substituting (23) into (22), we arrive at the following convergence requirement

$$N_c > \left(\sqrt{\frac{B}{2}} + \sqrt{K}\right)^2. \tag{24}$$

To satisfy (24) with $N_c = c \times B$ for $c = C_0+1, C_0+2, \ldots, C$, we get the following condition

$$C_0 \ge \left\lceil \frac{N_c}{B} - 1 \right\rceil = \left\lfloor \left(\sqrt{\frac{B}{2}} + \sqrt{K}\right)^2 / B \right\rfloor, \tag{25}$$

completing the proof. □

According to Theorems 1 and 2, it is evident that the proposed GRLS offers a flexible trade-off between the detection performance and complexity reduction, where a better detection performance can be obtained with the increment of C_0 at the expense of computational complexity. When $C_0 = C$, GRLS will exactly output the performance of MMSE detection.

V. Complexity And Data Bandwidth Analysis

We now study the complexity and data bandwidth of the proposed GRLS algorithm for distributed detection in massive MIMO, where the computational complexity is evaluated in terms of the required number of complex multiplications [4]. For example, computing the inversion of a $K \times K$ complex-valued matrix demands complexity $0.5K^3$.

In particular, the computational complexity of GRLS consists of two parts. As for the first stage (i.e., $c \le C_0$), computing $\mathbf{R}_c$ involves matrix multiplications between $\mathbf{P}_{c-1} \in \mathbb{C}^{K\times K}$ and $\mathbf{H}_c^H \in \mathbb{C}^{K\times B}$, between $\mathbf{H}_c \in \mathbb{C}^{B\times K}$ and $\mathbf{P}_{c-1}\mathbf{H}_c^H \in \mathbb{C}^{K\times B}$, and a $B \times B$ matrix inversion, which corresponds to complexity $K^2B + KB^2 + 0.5B^3$. Similarly, the complexities of computing $\mathbf{P}_c$ and $\mathbf{x}_c$ are $K^2B + KB^2$ and $K^2B + 2KB$, respectively.

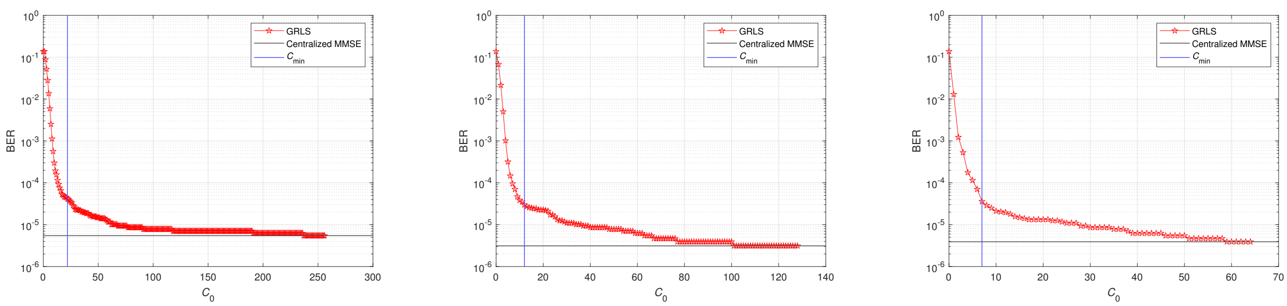
On the other hand, as for the second stage (i.e., $c > C_0$), the matrix multiplications between $\bar{\mathbf{P}}_{c-1}$ and $\mathbf{H}_c^H$, and between $\mathbf{H}_c$ and $\bar{\mathbf{P}}_{c-1}\mathbf{H}_c^H$ in (11) only need to take the diagonal elements into account, resulting in complexity $2KB$. Subsequently, the inversion operation requires the complexity B since it only involves calculating the reciprocals of the diagonal elements. The complexities of (11), (10), and (6) are reduced to $2KB + B$, $2KB$, and $3KB$, respectively.

To summarize, the total computational complexity of GRLS is $(3K^2B+2KB^2+2KB+0.5B^3)C_0+(7KB+B)(C-C_0)$. For ease of illustration, the complexity comparison is shown in Table I. Throughout the context, the numbers of DUs of RLS, SGD, and ASGD algorithms in [8] are set to $C = N$ due to the restriction $B = 1$, and the ADMM, CG, EPA, FD-MMSE, GMP, and DN algorithms in [3]–[7] with $C = 8$ are applied as the comparison, which corresponds to $B = N/8$ in each DU. As recommended in [8], the step-size of SGD is configured at 0.03 and the onset of the averaging procedure in ASGD is set to $k_0 = N/2$. Meanwhile, the numbers of inner iterations for EPA and GMP algorithms and the numbers of outer iterations for ADMM, CG, EPA, and DN algorithms are set to $I = 3$ and $T = 3$, respectively. From Table I, the proposed GRLS algorithm exhibits a much lower computational complexity than other schemes except the simple SGD and ASGD algorithms. More precisely, compared to the traditional RLS algorithm, more than 75% complexity is reduced for the case of 256×16, where $B = 4$, $C = 64$, and $C_0 = 7$, the minimum value in (18), are employed in GRLS, making it highly suitable for practical hardware implementation.

As summarized in Table I, the data bandwidth for interconnection of these distributed massive MIMO detection schemes is determined by the averaged complex values transferred on each link (e.g., K for SGD algorithm), which means the actual overhead in interface and may restricts its applications in practice [8]. Intuitively, GRLS achieves the comparable data bandwidth to the simple SGD and ASGD algorithms, along with a 78% reduction compared to the traditional RLS algorithm in the case of 256×16. Moreover, for the 256×32 scenario, while GRLS requires more data bandwidth, it still brings a reduction of 76% over RLS, making it more easily accommodated by existing hardware interfaces.

TABLE I
COMPUTATION COMPLEXITY COMPARISON OF DECENTRALIZED ALGORITHMS

Algorithm	Number of complex multiplications	Averaged complex values transferred on each link	256×16 system Complexity	256×16 system Bandwidth	256×32 system Complexity	256×32 system Bandwidth
ADMM [3]	$(2K^2B + KB + K + 0.5K^3)\,C + ((K^2 + 2K)\,C + K)\,T$	$(2T - 1)\,K$	158640	80	690016	160
CG [3]	$(K^2B + KB)C + (K^2 + 7K)CT - 4KC$	$(2T + 2)\,K$	77952	128	299264	256
EPA [4]	$((8\mathcal{O} + 2K + 1)NKI + 5\mathcal{O}K)\,T + NK + \mathcal{O}$	$(2T - 1)\,K$	5943056	80	14245392	160
FD-MMSE [5]	$(K^2B + KB + 3.5K^3 + K^2 + 4K)C$	$2K$	186880	32	1197056	64
GMP [6]	$(8\mathcal{O} + 2K + 1)NKI + NK + 6\mathcal{O}K$	$2K$	1984000	32	4754432	64
DN [7]	$(K^2B + KB + K)\,C + (K^2C + K)T$	Star: $(4T - 2)K$ Ring: $2KT + 4KT/C$	75952	160 120	295264	320 240
RLS [8]	$(3K^2 + 4K + 1)\,N$	$K^2 + K$	213248	272	819456	1056
SGD [8]	$(2K + 1)\,N$	K	8448	16	16640	32
ASGD [8]	$(2K + 1)\,k_0 + (3K + 1)\,(N - k_0)$	$2K$	10496	32	20736	64
Proposed GRLS	$(3K^2B + 2KB^2 + 2KB + 0.5B^3)C_0 + (7KB + B)(C - C_0)$	$((K^2 + K)C_0 + 2K(C - C_0))\,/C$	$C_{\min}$: 51972 $2 \times C_{\min}$: 75016 $3 \times C_{\min}$: 98060	58 84 110	210000 362404 514800	250 436 622



(a) $B = 1$, $C = 256$. (b) $B = 2$, $C = 128$. (c) $B = 4$, $C = 64$.

Fig. 2. GRLS detection algorithm convergence with SNR = 2 dB for the uncoded 256×16 massive MIMO.

VI. SIMULATION

In this section, to illustrate the convergence of the proposed GRLS algorithm, we first verify the effectiveness of the setting of parameter C_0 in Theorem 2. Then we compare GRLS with other distributed detection methods in bit error rate (BER) performance. Finally, we assess the performance of the GRLS across varying system sizes to evaluate its scalability. In all simulations, we consider the 16-QAM modulation scheme and uncoded systems. Generally, in massive MIMO, the channel hardening property is deemed to hold if $N/K \geq 8$ [15], [16].

Fig. 2 illustrates the convergence performance of the proposed GRLS detection algorithm under different antenna configurations at each DU for 256×16 massive MIMO systems at signal-to-noise ratio (SNR) = 2 dB. As can be seen clearly, when C_0 exceeds the minimum value in Theorem 2 ($C_{\min}$), the GRLS detection algorithm converges rapidly and achieves the near MMSE performance. It is worth noting that as C_0 increases to C, GRLS works as the MMSE detection solution, which is in line with the result derived in Theorem 1.

The BER performance comparison between GRLS and other distributed detection schemes, employed the same parameters as Section V, is presented in Fig. 3 with respect to a 256×16 massive MIMO system. Despite their extremely high computational complexity and data bandwidth, the EPA and RLS algorithms achieve the MMSE performance. Also, the ADMM and CG detection schemes incur a slight performance loss over the centralized MMSE detection but at a substantial data bandwidth cost. Due to the high scattering of BS antennas, the local estimation results obtained by FD-MMSE, GMP, and DN algorithms are inaccurate, resulting in significant performance degradation. Besides, the SGD and ASGD algorithms require quite low computational complexity and data bandwidth, making it easy to apply, but their performance decreases at about 7.0 dB and 3.5 dB over the centralized MMSE detection, respectively. Note that the proposed GRLS detection algorithm with $C_{\min}$ outperforms CG, DN, FD-MMSE, and GMP detection schemes at a lower cost while achieving near MMSE performance with the increment of C_0.

In Fig. 4, we extend the BER performance comparison to a 256×32 massive MIMO system. Intuitively, all the convergence performance of FD-MMSE, GMP, and DN algorithms degrades further, leading to unsatisfactory detection performance in uplink massive MIMO systems. This is attributed to the fact that the limited number of antenna resources in each DU fails to satisfy the condition $B \gg K$. In contrast, the convergence of GRLS works well as expected, because its convergence always hold if (18) is satisfied. Clearly, with the increment of C_0, the BER performance of GRLS improves gradually. For example, the proposed GRLS detection algorithm with two and three times $C_{\min}$ achieves gains of nearly 1.5 dB and 2.0 dB over that with $C_{\min}$ at the BER of 10^{-5}, respectively. However, it also leads to a rise in

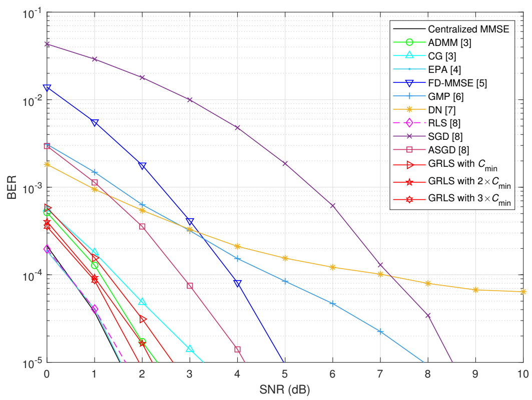


Fig. 3. BER performance comparison of different methods for the uncoded 256×16 massive MIMO.

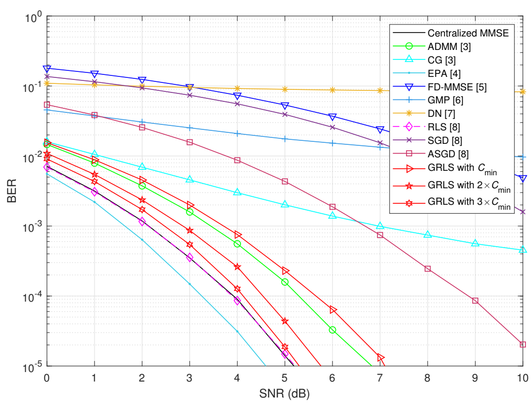


Fig. 4. BER performance comparison of different methods for the uncoded 256×32 massive MIMO.

computational complexity and data bandwidth costs by 72% and 74% for $2 \times C_{\min}$, and 145% and 148% for $3 \times C_{\min}$, as indicated in Table I, respectively. In summary, a larger C_0 brings considerable performance gains but with higher complexity and bandwidth cost such that it should be selected according to the practical implementation requirement.

In Fig. 5, to illustrate the scalability of our proposed GRLS algorithm, we scale the total number of BS antennas $N = C \times B$ by increasing the number of DU from $C = 128$, $C = 192$, to $C = 256$ with fixed $B = 4$. As can be seen clearly, GRLS consistently exhibits near-MMSE performance, suggesting its effectiveness and scalability in distributed massive MIMO systems.

VII. Conclusion

In this paper, by exploiting the potential of the RLS detection algorithm, we proposed a novel near-MMSE algorithm, termed GRLS, for distributed massive MIMO detection with low computational complexity and data bandwidth. The convergence, complexity, and bandwidth analysis for the proposed GRLS detection algorithm are also provided. When the derived convergence condition about C_0 is satisfied, the proposed GRLS achieves rapid convergence and is capable of attaining the near MMSE performance. Additionally, the simulation results not only confirm its desirable trade-off between complexity and performance but also validate its excellent scalability, making it well-suited for practical implementations.

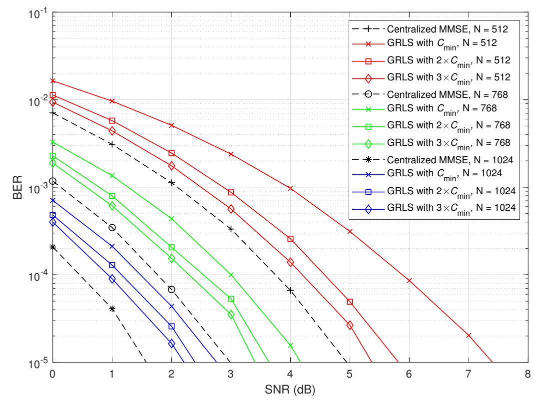


Fig. 5. BER performance of the GRLS algorithm with a fixed number of users $K = 64$.

References

[1] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO Detection Techniques: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 4, pp. 3109-3132, Aug. 2019.

[2] 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. 2016 50th Conf. Rec. Asilomar Conf. Signals Syst. Comput.*, pp. 468-472, Mar. 2016.

[3] 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. Emerging Sel. Top. Circuits Syst.*, vol. 7, no. 4, pp. 491-507, Nov. 2017.

[4] Z. Zhang, H. Li, Y. Dong, X. Wang, and X. Dai, "Decentralized Signal Detection via Expectation Propagation Algorithm for Uplink Massive MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11233-11240, Oct. 2020.

[5] 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, Jul. 2019.

[6] Z. Zhang, Y. Dong, K. Long, X. Wang, and X. Dai, "Decentralized Baseband Processing With Gaussian Message Passing Detection for Uplink Massive MU-MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 2152-2157, Feb. 2022.

[7] A. Kulkarni, M. A. Ouameur, and D. Massicotte, "Hardware Topologies for Decentralized Large-Scale MIMO Detection Using Newton Method," *IEEE Trans. Circuits Syst.*, vol. 68, no. 9, pp. 3732-3745, Jul. 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. 2018 IEEE Workshop Signal. Process. Syst. SiPS Des. Implement.*, pp. 53-58, Oct. 2018.

[9] J. R. Snchez, 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, Jan. 2020.

[10] Z. Wang, J. Zhang, and H. Du, "Extremely Large-Scale MIMO: Fundamentals, Challenges, Solutions, and Future Directions," *IEEE Wireless Commun.*, pp. 1-9, Apr. 2023.

[11] A. Bereyhi, S. Asaad, B. Gade, R. R. Muller, and H. V. Poor, "Detection of Spatially Modulated Signals via RLS: Theoretical Bounds and Applications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2291-2304, Apr. 2022.

[12] C. Thrampoulidis, W. Xu, and B. Hassibi, "Symbol Error Rate Performance of Box-Relaxation Decoders in Massive MIMO," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3377-3392, Jul. 2018.

[13] A. Papoulis, *Probability, random variables, and stochastic processes*, 4th ed. New York, NY, USA: McGraw-Hill, Dec. 2002.

[14] F. Rusek, D. Persson, B. K. Lau, and E. G. Larsson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Process Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.

[15] Z. Wang, R. M. Gower, Y. Xia, L. He and Y. Huang, "Randomized Iterative Methods for Low-Complexity Large-Scale MIMO Detection," *IEEE Trans. Signal Process.*, vol. 70, pp. 2934-2949, Jun. 2022.

[16] Z. Wang, W. Xu, Y. Xia, Q. Shi and Y. Huang, "A New Randomized Iterative Detection Algorithm for Uplink Large-Scale MIMO Systems," *IEEE Trans. Commun.*, vol. 71, no. 9, pp. 5093-5107, Sept. 2023.