This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2021.3121445, IEEE Communications Letters

An Efficient Linear Detection Scheme Based on L-BFGS Method for Massive MIMO Systems

Lin Li, and Jianhao Hu, *Member, IEEE*

***Abstract*—For massive multiple-input multiple-output (MIMO) systems, minimum mean square error (MMSE) detection is near-optimal, but requires high-complexity matrix inversion. To avoid matrix inversion, we formulate MMSE detection as a strictly convex quadratic optimization problem, which can be solved iteratively by the recognized most efficient Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method. According to special properties of massive MIMO systems, we propose a novel limited-memory BFGS (L-BFGS) scheme for MMSE detection with one correction search, unit step length, and simplified initialization, which can greatly reduce the storage and computation cost compared to BFGS method. Simulation results finally verify the effectiveness of the proposed scheme.**

***Index Terms*—Massive MIMO, MMSE detection, quasi-Newton method, L-BFGS.**

I. Introduction

In wireless communication systems, massive multiple-input multiple-output (MIMO) has been a key technology for improving the spectral efficiency and system capacity by deploying hundreds of antennas at the base station (BS) to simultaneously serve tens of user equipments (UEs) in the same frequency band [1], [2]. For uplink massive MIMO systems, linear detection schemes, such as the minimum mean square error (MMSE) detection, have been demonstrated to be near-optimal with relatively lower computational complexity than nonlinear methods[3]. However, they will involve high-dimensional matrix inversion with the increasing number of UEs, causing difficulties for the practical implementation.

Consequently, many efforts have been devoted to avoiding exact matrix inversion in linear massive MIMO detection, which can be roughly divided into approximation methods and iterative methods [4]. Neumann series approximation (NSA) algorithms are proposed to convert matrix inversion into the accumulation of a series of matrix-vector multiplications. However, its computational complexity is even higher than exact matrix inversion, when the number of NSA terms is larger than two [5]. The iterative methods, including Gauss-Seidel (GS) [6], successive over relaxation (SOR) [7], and symmetric successive over relaxation (SSOR) [8], are derived by splitting and rearranging the system matrix. Their disadvantage is that they focus more on precision and the complexity is still high. Conjugate gradient (CG) method [9] is a more efficient iterative algorithm, which transforms the problem of solving the linear equation into a strictly convex quadratic optimization problem. Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, a popular quasi-Newton method to build up an inverse Hessian approximation, is acknowledged as the most efficient and robust iterative algorithms for solving optimization problems [10]. Nevertheless, it requires to store the inverse Hessian approximation, resulting in high storage and computation cost for the large-scale optimization, which makes it unsuitable for linear massive MIMO detection.

In this letter, we exploit and improve the limited-memory BFGS (L-BFGS) method to iteratively realize MMSE-based massive MIMO detection without matrix inversion. The basic idea underlying L-BFGS method is to update the inverse Hessian approximation implicitly by the latest m correction vector pairs, so as to reduce the storage and be computationally efficient compared with BFGS method [11]. To further reduce the complexity, based on the special property that the filtering matrix of MMSE detection is symmetric positive definite (SPD), we demonstrate that L-BFGS method simplified with $m = 1$ is mathematically identical to BFGS method under exact line search, and the application of unit step length can also ensure convergence. Moreover, since the asymptotic orthogonality of massive MIMO channels makes the MMSE filtering matrix diagonally dominant, an appropriate initialization strategy is proposed to accelerate convergence. Numerical simulations finally verify that the proposed scheme can closely approach to the performance of exact MMSE detection with much lower complexity compared to the existing algorithms reported in this letter.

Notations: The bold uppercase and lowercase denote the matrix and vector, respectively. The conjugate transpose, transpose, and inverse of a matrix are denoted by $[\cdot]^H$, $[\cdot]^T$, and $[\cdot]^{-1}$, respectively. $\mathbb{R}$, $\mathbb{C}$ denote a set of real and complex numbers, while $\Re\{\cdot\}$, $\Im\{\cdot\}$ denote the real and imaginary part of the complex number, respectively. $\mathcal{CN}(\cdot,\cdot)$ represents the complex Gaussian distribution, where the two arguments are the mean and variance or covariance matrix. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. O represents order of.

II. System Model

We consider an uplink massive MIMO system equipped with N_r antennas at the BS to simultaneously serve N_t single-antenna UEs ($N_t \leq N_r$). The received signal $\overline{\mathbf{y}} \in \mathbb{C}^{N_r \times 1}$ is

$$\overline{\mathbf{y}} = \overline{\mathbf{H}}\overline{\mathbf{s}} + \overline{\mathbf{n}}, \tag{1}$$

where $\overline{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t}$ is the flat Rayleigh fading channel matrix, whose entry is the independent and identically distributed (i.i.d) Gaussian random variable with zero mean and unit variance. $\overline{\mathbf{s}} \in \mathbb{C}^{N_t \times 1}$ is the transmitted signal vector with the average power normalized to $1/N_t$. $\overline{\mathbf{n}} \in \mathbb{C}^{N_r \times 1} \sim \mathcal{CN}\left(0, \sigma_n^2 \mathbf{I}_{N_r}\right)$ denotes the additive white Gaussian noise (AWGN) vector

The authors are with National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, 611731 China e-mail: jhhu@uestc.edu.cn.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2021.3121445, IEEE Communications Letters

with $\sigma_{\bar{n}}^2$ being the average noise power. Thus the signal-to-noise ratio (SNR) is calculated as $1/\sigma_{\bar{n}}^2$. Without losing generality, we assume $\overline{\mathbf{H}}$ is perfectly known at the BS.

For ease of discussion, given $N = 2N_r$ and $K = 2N_t$, the complex-valued system model can be converted to the model in the real domain, i.e.,

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

by $\mathbf{y} = \begin{bmatrix} \Re\{\bar{\mathbf{y}}\} \\ \Im\{\bar{\mathbf{y}}\} \end{bmatrix} \in \mathbb{R}^{N\times 1}$, $\mathbf{s} = \begin{bmatrix} \Re\{\bar{\mathbf{s}}\} \\ \Im\{\bar{\mathbf{s}}\} \end{bmatrix} \in \mathbb{R}^{K\times 1}$, $\mathbf{n} = \begin{bmatrix} \Re\{\bar{\mathbf{n}}\} \\ \Im\{\bar{\mathbf{n}}\} \end{bmatrix} \in \mathbb{R}^{N\times 1}$, and $\mathbf{H} = \begin{bmatrix} \Re\{\overline{\mathbf{H}}\} & -\Im\{\overline{\mathbf{H}}\} \\ \Im\{\overline{\mathbf{H}}\} & \Re\{\overline{\mathbf{H}}\} \end{bmatrix} \in \mathbb{R}^{N\times K}$.

In classical MMSE detection, the estimation of the transmitted signal $\hat{\mathbf{s}}$ can be expressed as

$$\hat{\mathbf{s}} = \left(\mathbf{H}^T\mathbf{H} + \sigma_n^2\mathbf{I}_K\right)^{-1}\mathbf{H}^T\mathbf{y} = \mathbf{A}^{-1}\mathbf{b}, \tag{3}$$

where $\sigma_n^2 = 1/2\sigma_{\bar{n}}^2$ and $\mathbf{A} = \mathbf{H}^T\mathbf{H} + \sigma_n^2\mathbf{I}_K$, $\mathbf{G} = \mathbf{H}^T\mathbf{H}$, $\mathbf{b} = \mathbf{H}^T\mathbf{y}$ denote the MMSE filtering matrix, Gram matrix and matched-filter vector, respectively. Note that the computational complexity of the exact matrix inversion $\mathbf{A}^{-1}$ is $O(K^3)$, which causes high computational burden for massive MIMO systems.

III. Proposed Detection Scheme

In order to avoid the matrix inversion, MMSE detection can be interpreted as solving the following linear equation

$$\mathbf{A}\hat{\mathbf{s}} = \mathbf{b}. \tag{4}$$

First, we have

$$\mathbf{G}^T = (\mathbf{H}^T\mathbf{H})^T = \mathbf{H}^T\mathbf{H} = \mathbf{G}, \tag{5}$$

which shows $\mathbf{G}$ is symmetric. In addition, for an arbitrary non-zero vector $\mathbf{x} \in \mathbb{R}^{K\times 1}$, we can derive that

$$\mathbf{x}^T\mathbf{G}\mathbf{x} = \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x} = (\mathbf{H}\mathbf{x})^T(\mathbf{H}\mathbf{x}) \geq 0, \tag{6}$$

which further shows Gram matrix $\mathbf{G}$ is positive semi-definite. Since the average noise power σ_n^2 is positive, the MMSE filtering matrix $\mathbf{A} = \mathbf{G} + \sigma_n^2\mathbf{I}_K$ is SPD. Based on this property, equation (4) can be equivalent to minimizing a strictly convex quadratic function

$$f(\mathbf{s}) = \frac{1}{2}\mathbf{s}^T\mathbf{A}\mathbf{s} - \mathbf{b}^T\mathbf{s}, \tag{7}$$

where $\mathbf{A}$ is the Hessian matrix. That is to say, the transmitted signal $\hat{\mathbf{s}}$ can be estimated by solving the unconstrained quadratic optimization problem

$$\min_{\mathbf{s}\in\mathbb{R}^{K\times 1}} f(\mathbf{s}). \tag{8}$$

A. BFGS Method

BFGS method is among the most effective and stable quasi-Newton methods for optimization problems. Therefore, it can be used to avoid matrix inversion in MMSE detection. Starting from an initial vector $\mathbf{s}_0$, the iterative formula used to solve (8) is often expressed as

$$\mathbf{s}_{k+1} = \mathbf{s}_k + \alpha_k\mathbf{d}_k, \tag{9}$$

where α_k is the step length, $\mathbf{d}_k$ is the search direction, and k denotes the k-th iteration.

It is known that quasi-Newton methods can attain finite termination on quadratic optimization problems with exact line search [12]. Hence, we compute α_k such that

$$\begin{aligned}\alpha_k &= \arg\min_{\alpha\geq 0} f(\mathbf{s}_k + \alpha\mathbf{d}_k) \\ &= -\frac{\mathbf{g}_k^T\mathbf{d}_k}{\mathbf{d}_k^T\mathbf{A}\mathbf{d}_k},\end{aligned} \tag{10}$$

where $\mathbf{g}_k$ is the gradient of $f(\mathbf{s})$ at $\mathbf{s}_k$.

In quasi-Newton methods, the search direction $\mathbf{d}_k$ at iteration k is generated by

$$\mathbf{d}_k = -\mathbf{B}_k\mathbf{g}_k, \tag{11}$$

where $\mathbf{B}_k$ is a SPD matrix to approximate the inverse Hessian matrix $\mathbf{A}$. The basic requirement for updating $\mathbf{B}_k$ is that the following secant equation is satisfied at each iteration.

$$\mathbf{B}_{k+1}\mathbf{q}_k = \mathbf{p}_k, \tag{12}$$

where $\mathbf{p}_k = \mathbf{s}_{k+1} - \mathbf{s}_k$ and $\mathbf{q}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$. Initially, $\mathbf{B}_0$ can be any SPD matrix, and the updating formula of $\mathbf{B}_k$ by a rank-two correction is expressed as

$$\begin{aligned}\mathbf{B}_{k+1} &= \mathbf{B}_k - \frac{\mathbf{p}_k\mathbf{q}_k^T\mathbf{B}_k + \mathbf{B}_k\mathbf{q}_k\mathbf{p}_k^T}{\mathbf{p}_k^T\mathbf{q}_k} + \left(1 + \frac{\mathbf{q}_k^T\mathbf{B}_k\mathbf{q}_k}{\mathbf{p}_k^T\mathbf{q}_k}\right)\frac{\mathbf{p}_k\mathbf{p}_k^T}{\mathbf{p}_k^T\mathbf{q}_k} \\ &= \left(\mathbf{I} - \frac{\mathbf{p}_k\mathbf{q}_k^T}{\mathbf{p}_k^T\mathbf{q}_k}\right)\mathbf{B}_k\left(\mathbf{I} - \frac{\mathbf{q}_k\mathbf{p}_k^T}{\mathbf{p}_k^T\mathbf{q}_k}\right) + \frac{\mathbf{p}_k\mathbf{p}_k^T}{\mathbf{p}_k^T\mathbf{q}_k} \\ &= \mathbf{V}_k^T\mathbf{B}_k\mathbf{V}_k + \rho_k\mathbf{p}_k\mathbf{p}_k^T,\end{aligned} \tag{13}$$

where $\rho_k = 1/\mathbf{p}_k^T\mathbf{q}_k$ and $\mathbf{V}_k = \mathbf{I} - \rho_k\mathbf{q}_k\mathbf{p}_k^T$.

Obviously, as the inverse Hessian approximation $\mathbf{B}_k$ iteratively becomes dense, the storage and computation of BFGS method will become expensive, which is not practical to implement for massive MIMO systems.

B. L-BFGS Method

L-BFGS method is an adaptation of BFGS method to large-scale problems [13]. It aims to store no more than m pairs $\{\mathbf{p}_i, \mathbf{q}_i\}_{i=k-\hat{m}}^{k}$, where $\hat{m} = \min\{k, m-1\}$, for corrections of the initial matrix $\mathbf{B}_k^0$ to generate $\mathbf{B}_{k+1}$. Based on the formula (13), the iterative formula of L-BFGS method is derived as

$$\begin{aligned}\mathbf{B}_{k+1} = &\left(\mathbf{V}_k^T\cdots\mathbf{V}_{k-\hat{m}}^T\right)\mathbf{B}_k^0\left(\mathbf{V}_{k-m}\cdots\mathbf{V}_k\right) \\ &+ \rho_{k-\hat{m}}\left(\mathbf{V}_k^T\cdots\mathbf{V}_{k-\hat{m}+1}^T\right)\mathbf{p}_{k-\hat{m}}\mathbf{p}_{k-\hat{m}}^T\left(\mathbf{V}_{k-\hat{m}+1}\cdots\mathbf{V}_k\right) \\ &\vdots \\ &+ \rho_k\mathbf{p}_k\mathbf{p}_k^T,\end{aligned} \tag{14}$$

where $\mathbf{B}_k^0$ is always a simple SPD matrix given in advance or generated automatically at each iteration. And we will discuss how to choose it in the next subsection.

Using the formula (14), we can create a two-loop recursion to compute the search direction $\mathbf{d}_{k+1} = -\mathbf{B}_{k+1}\mathbf{g}_{k+1}$ without storing $\mathbf{B}_k$ explicitly. The efficient two-loop recursive procedure is shown as below.

Algorithm 1 L-BFGS two-loop recursion

Input: a certain number m of $\{\mathbf{p}_i, \mathbf{q}_i\}$, and $\rho_i = 1/\left(\mathbf{p}_i^T \mathbf{q}_i\right)$ for all $i \in k-m+1, \cdots, k$, the current gradient $\mathbf{g}_{k+1}$;
Output: new search direction $\mathbf{d}_{k+1} = -\mathbf{B}_{k+1}\mathbf{g}_{k+1}$;

$\mathbf{r} = \mathbf{g}_{k+1}$;
for $i = k$ to $k-m+1$ **do**
  $\alpha_i = \rho_i \mathbf{p}_i^T \mathbf{r}$;
  $\mathbf{r} = \mathbf{r} - \alpha_i \mathbf{q}_i$;
end for
$\mathbf{r} = \mathbf{B}_k^0 \mathbf{r}$;
for $i = k-m+1$ to k **do**
  $\beta = \rho_i \mathbf{q}_i^T \mathbf{r}$;
  $\mathbf{r} = \mathbf{r} + \mathbf{p}_i\left(\alpha_i - \beta\right)$;
end for
Return: $\mathbf{d}_{k+1} = -\mathbf{r}$.

It can be seen that the two-loop recursion requires only $(4m+1)K+2m$ multiplications. And L-BFGS method need to store $2mK + O(K)$ numbers, which can reduce the memory cost significantly compared to BFGS method with $O(K^2)$ memory requirement. Practical experience has indicated that small values of m (between 3 and 7, say) often give satisfactory results for general large-scale unconstrained problems.

Based on the analysis above, the computation process of L-BFGS method for MMSE detection is summarized in Algorithm 2.

Algorithm 2 L-BFGS method for MMSE detection

Input: $\mathbf{H}$, $\mathbf{y}$, σ_n^2;
Output: $\hat{\mathbf{s}}$;
Initialization:
Initialize starting vector $\mathbf{s}_0$, inverse Hessian approximation $\mathbf{B}_0$, accuracy threshold ϵ, iteration number L, correction number m; Compute $\mathbf{A} = \mathbf{H}^T\mathbf{H} + \sigma_n^2 \mathbf{I}_K$, $\mathbf{b} = \mathbf{H}^T\mathbf{y}$, $\mathbf{g}_0 = \mathbf{A}\mathbf{s}_0 - \mathbf{b}$, $\mathbf{d}_0 = -\mathbf{B}_0\mathbf{g}_0$; Set $k = 0$;
Step 1: If $\|\mathbf{g}_k\| \le \epsilon$ or $k = L$, stop and return $\hat{\mathbf{s}} = \mathbf{s}_k$;
Step 2: Find the step length α_k with the formula (10);
Step 3: Compute the new iteration $\mathbf{s}_{k+1} = \mathbf{s}_k + \alpha_k \mathbf{d}_k$;
Step 4: Compute the new gradient $\mathbf{g}_{k+1} = \mathbf{A}\mathbf{s}_{k+1} - \mathbf{b}$;
Step 5: Update $\mathbf{p}_k = \mathbf{s}_{k+1} - \mathbf{s}_k$ and $\mathbf{q}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$;
Step 6: If $k{>}m$, discard vector pair $\mathbf{p}_{k-m}$, $\mathbf{q}_{k-m}$ from storage;
Step 7: Use Algorithm 1 to compute the new search direction $\mathbf{d}_{k+1} = -\mathbf{B}_{k+1}\mathbf{g}_{k+1}$;
Step 8: $k := k+1$ and go to **Step 1.**

C. Improvement Strategy

To take full advantage of the massive MIMO properties and further reduce the iteration cost, we have studied three improvement strategies for L-BFGS method based MMSE detection, including reducing the number of correction vector pairs, evaluating the impact of initialization and choosing the appropriate step length.

1) One Correction Search: Based on the discussion in the previous subsection, it can be derived that the smaller m is, the more computationally efficient L-BFGS method is. Consider the case that $m = 1$, the transformation of the formula (14) can be expressed as

$$\mathbf{B}_{k+1} = \mathbf{V}_k^T \mathbf{B}_k^0 \mathbf{V}_k + \rho_k \mathbf{p}_k \mathbf{p}_k^T. \tag{15}$$

By comparing the formula (15) with the formula (13), it is clear that L-BFGS method with $m = 1$ is a special case of BFGS method in which $\mathbf{B}_k^0$ is updated at each iteration.

Our primary concern is the search direction $\mathbf{d}_{k+1}$, thus we can simply multiply $-\mathbf{g}_{k+1}$ by the right side of $\mathbf{B}_{k+1}$. Note that with the exact line search, $\mathbf{p}_k^T \mathbf{g}_{k+1} = 0$. Then we can obtain

$$\begin{aligned}
\mathbf{d}_{k+1}^{LBFGS} &= -\mathbf{B}_{k+1}\mathbf{g}_{k+1} \\
&= -\left(\mathbf{V}_k^T \mathbf{B}_k^0 \mathbf{V}_k + \rho_k \mathbf{p}_k^T \mathbf{p}_k\right)\mathbf{g}_{k+1} \\
&= -\left(\mathbf{I} - \rho_k \mathbf{p}_k \mathbf{q}_k^T\right)\mathbf{B}_k^0\left(\mathbf{I} - \rho_k \mathbf{q}_k \mathbf{p}_k^T\right)\mathbf{g}_{k+1} \\
&= -\mathbf{B}_k^0 \mathbf{g}_{k+1} + \rho_k \mathbf{p}_k \mathbf{q}_k^T \mathbf{B}_k^0 \mathbf{g}_{k+1}.
\end{aligned} \tag{16}$$

In a similar way, refer to the updating formula (13) of BFGS method, we can derive

$$\mathbf{d}_{k+1}^{BFGS} = -\mathbf{B}_k \mathbf{g}_{k+1} + \rho_k \mathbf{p}_k \mathbf{q}_k^T \mathbf{B}_k \mathbf{g}_{k+1}. \tag{17}$$

In the context of BFGS method, we consider the computation of $\mathbf{B}_j^{BFGS}\mathbf{g}_{k+1}$ $(0 \le j \le k)$, i.e.

$$\mathbf{B}_j^{BFGS}\mathbf{g}_{k+1} = \mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1} + \rho_k \mathbf{p}_{j-1}\mathbf{q}_{j-1}^T \mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1}. \tag{18}$$

Assume that $\mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1} = \mathbf{B}_0\mathbf{g}_{k+1}$ and combine the BFGS property $\mathbf{g}_j^T \mathbf{B}_0 \mathbf{g}_{k+1} = 0$ for convex quadratic problems [13], then we calculate

$$\begin{aligned}
\mathbf{q}_{j-1}^T \mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1} &= \left(\mathbf{g}_j - \mathbf{g}_{j-1}\right)^T \mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1} \\
&= \mathbf{g}_j^T \mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1} - \mathbf{g}_{j-1}^T \mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1} \\
&= \mathbf{g}_j^T \mathbf{B}_0 \mathbf{g}_{k+1} - \mathbf{g}_{j-1}^T \mathbf{B}_0 \mathbf{g}_{k+1} \\
&= 0.
\end{aligned} \tag{19}$$

Applying (19) to (18), we get

$$\mathbf{B}_j^{BFGS}\mathbf{g}_{k+1} = \mathbf{B}_{j-1}^{BFGS}\mathbf{g}_{k+1} = \mathbf{B}_0\mathbf{g}_{k+1}. \tag{20}$$

Finally, we can conclude

$$\mathbf{d}_k^{LBFGS} = \mathbf{d}_k^{BFGS}. \tag{21}$$

That is, L-BFGS method with $m = 1$ and $\mathbf{B}_k^0 = \mathbf{B}_0$ can generate the identical search directions to BFGS method under the exact line search for MMSE detection. It can be observed that when m is set to 1 in Algorithm 1, the inner iteration of the two-loop recursion will be omitted, which can reduce the storage and complexity significantly.

2) Simplified Initial Approximation Matrix: Because the choice of $\mathbf{B}_k^0$ will influence the behavior of L-BFGS method, it is worth the investigation in this letter.

A simple initialization way of L-BFGS method is to set $\mathbf{B}_k^0 = \mathbf{I}$, but often suffers from slow convergence. Due to that the two-loop recursion makes $\mathbf{B}_k^0$ isolated from the rest of the computations, $\mathbf{B}_k^0$ is allowed to be chosen freely and vary at each iteration. Thus, a common and effective alternative for L-BFGS method is to set $\mathbf{B}_k^0 = \gamma_k \mathbf{I}$ [14], where

$$\gamma_k = \frac{\mathbf{q}_k^T \mathbf{p}_k}{\mathbf{q}_k^T \mathbf{q}_k}. \tag{22}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2021.3121445, IEEE Communications Letters

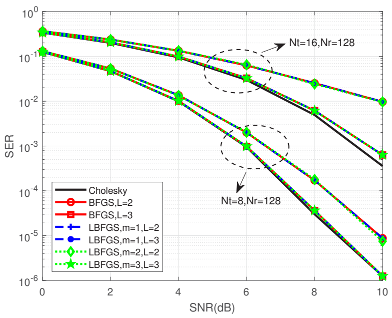


Fig. 1. Performance comparison of different m.

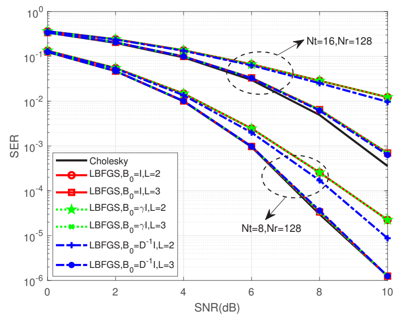


Fig. 2. Performance comparison of different $\mathbf{B}_0$.

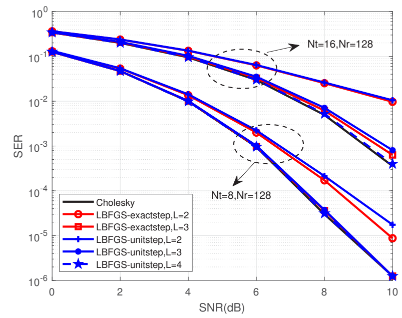


Fig. 3. Performance comparison of different α.

For massive MIMO systems, it is known that $\mathbf{A}$ is diagonally dominant due to the asymptotic orthogonality of the channel $\mathbf{H}$, so the inverse Hessian approximation $\mathbf{B}_k^0$ can be simply taken as $\mathbf{D}^{-1}$ herein, where $\mathbf{D}$ represents the diagonal matrix of $\mathbf{A}$. Obviously, the complexity of computing $\mathbf{D}^{-1}$ is low. And the simulation in the subsequent section will compare the performance of these initialization methods.

3) Unit Step length: As seen from Algorithm 2, the exact line research is a major computational burden of L-BFGS method. Fortunately, quasi-Newton methods are robust with a good property [14]:

Theorem 1: Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the strictly convex quadratic function as expressed in equation (7). For $\alpha_k \equiv 1$, BFGS methods converge globally and superlinearly to $\mathbf{A}^{-1}\mathbf{b}$.

We have previously shown that MMSE detection can be transformed into a strictly convex quadratic optimization problem and L-BFGS method is a special case of BFGS method. Therefore, with the property above, we can directly set the step length as $\alpha_k = 1$ to further reduce the computational complexity of L-BFGS method for MMSE detection.

D. Complexity Analysis

The computational complexity of the proposed scheme is analyzed in terms of the required number of multiplications. Generally, the whole complexity comes from two parts. Thereinto, one part is from the computation of the MMSE filtering matrix $\mathbf{A}$ and matched-filter output $\mathbf{b}$, which exists in almost all the conventional MMSE detection. Hence, we focus on the computational complexity of another part, i.e. the iteration process. The computational complexity of the proposed scheme in one iteration is listed as below.

1) Compute the step length α_k: Refer to the formula (10), the step length α_k can be computed through a multiplication of a $K\times K$ matrix and a $K\times 1$ vector with two inner products of two $K\times 1$ vectors. Thus the total complexity of this calculation is K^2+2K. However, it can be avoided when we set $\alpha_k = 1$, greatly reducing the whole complexity.

2) Compute the new iteration $\mathbf{s}_{k+1}$: The new iteration $\mathbf{s}_{k+1}$ needs to multiply a scalar α_k by the $K\times 1$ vector $\mathbf{d}_{k+1}$, which contains the complexity of K. It is worth noting that when $\alpha_k = 1$, this calculation can also be omitted.

3) Compute the new gradient $\mathbf{g}_{k+1}$: The calculation of $\mathbf{g}_{k+1}$ only includes one multiplication of a $K\times K$ matrix $\mathbf{A}$ and a $K\times 1$ vector $\mathbf{s}_{k+1}$, the complexity of which is K^2.

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY

Detection Method	Number of Multiplications
Cholesky	$2K^3+2K^2$
NSA	$2K^2-2K\ (L=2),(L-2)K^3-K\ (L>2)$
SSOR	$(4K^2+4K)L$
CG	$(2K^2+7K)L$
BFGS	$(6K^2+7K)L$
Proposed L-BFGS	$(K^2+5K+2)L$

4) Compute the two-loop recursion to obtain the new direction $\mathbf{d}_{k+1}$: As mentioned earlier, the computational complexity of the L-BFGS recursion is $4Km+K+2m$. According to the discussion about m, which can be set as 1 to be equivalent to BFGS method, the complexity can be reduced to $5K+2$.

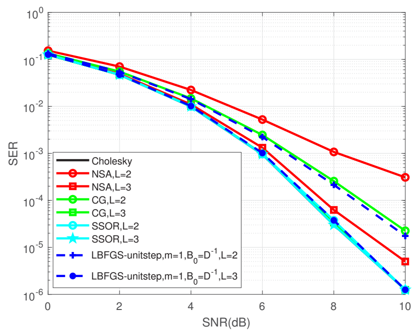
Therefore, the overall complexity of the proposed L-BFGS method for MMSE detection is $2K^2+8K+2$ and can be further reduced to K^2+5K+2 with $\alpha_k = 1$.

Table I shows the complexity comparison among the proposed L-BFGS method and other classical linear detection methods. The complexity of Cholesky method for exact MMSE detection is $O(K^3)$. When the NSA term is 2, NSA method achieves minimal complexity, but when it is larger than 2, the complexity increases even higher than Cholesky method. The iterative methods can reduce the complexity of exact MMSE detection to $O(K^2)$, and the proposed method has the lowest complexity in an iteration process. The smaller required number of iterations L means the lower total complexity. Moreover, the proposed L-BFGS method like the other existing methods only requires $O(K)$ storage space.

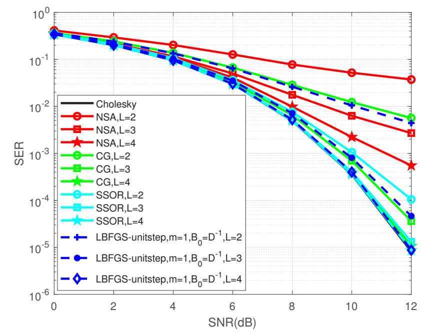
IV. SIMULATION RESULTS

We now provide simulation results to verify the effectiveness of the proposed scheme and compare the average symbol error rate (SER) performance to conventional detection algorithms against different values of SNR under 16-QAM modulation. The performance of Cholesky method based MMSE detection is used as the benchmark for comparison. Simulations are conducted using Matlab for various antenna scenarios in i.i.d. flat Rayleigh fading channel.

We can see from Fig. 1 that L-BFGS method with $m = 1$ is identical to BFGS method as well as L-BFGS method with



(a) 8×128



(b) 16×128

Fig. 4. SER performance comparison of the proposed method with existing methods under 16-QAM modulation for different antenna configurations.

m which is increased up to the required number of iteration under 8×128 and 16×128 antenna configurations. Thus, we can set $m = 1$ for reliable performance and lower complexity.

Fig.2 shows the performance comparison of the different initialization methods for the proposed L-BFGS detection scheme under different antenna configurations. As seen from Fig.2, the initial method $\mathbf{B}_0^k = \mathbf{D}^{-1}$ can accelerate the convergence compared with its counterparts, and the acceleration performance is more significant under 8×128 antenna configuration due to the channel hardening. The convergence performance of $\mathbf{B}_0^k = \mathbf{I}$ and $\mathbf{B}_0^k = \gamma_k \mathbf{I}$ is almost the same, but the former has the lowest complexity. Thus, we choose the initial methods $\mathbf{B}_0^k = \mathbf{D}^{-1}$ for its better trade-off between complexity and performance.

As seen from Fig. 3, L-BFGS method with unit step length can nearly achieve the detection performance with the exact line research. It also can be seen that with few iterations, L-BFGS method with unit step length can achieve the exact MMSE detection performance. Based on the simulations above, we compare the detection performance of the proposed scheme with that of the existing linear detection algorithms.

Fig. 4 compares the SER performance of the proposed L-BFGS detection scheme with different detection algorithms (namely, Cholesky, NSA, CG, SSOR) with 16-QAM modulation under 8×128 and 16×128 antenna configurations, respectively. Results show that the performance of all iterative methods improves as the number of iterations increases. And they can approach to exact MMSE performance with a small number of iterations. From this figure, we can see that NSA works as the worst one among all cases. The performance of the proposed scheme is almost consistent with CG method, and even slightly better in some cases. Although SSOR method can achieve the best performance, the performance gain is minor with much higher complexity than the proposed method.

In summary, all simulations verify the efficiency and robustness of the proposed L-BFGS scheme for MMSE detection.

V. Conclusion

In this letter, by fully utilizing the special property of massive MIMO systems, we propose a novel and efficient detection scheme based on L-BFGS method. Through derivation and simulation, we show that the proposed scheme can achieve near-MMSE detection performance with significant reduction in complexity. Additionally, the idea of using L-BFGS method to efficiently avoid matrix inversion can be applied to other signal processing problems involving complicated matrix inversion. Our future work is to extend the proposed linear detection scheme into more realistic channel models for massive MIMO systems and achieve efficient hardware design.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.
[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.
[3] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2017.
[4] M. A. Albreem, W. Salah, A. Kumar, M. H. Alsharif, A. H. Rambe, M. Jusoh, and A. N. Uwaechia, "Low complexity linear detectors for massive MIMO: A comparative study," *IEEE Access*, vol. 9, pp. 45 740–45 753, 2021.
[5] B. Nagy, M. Elsabrouty, and S. Elramly, "Fast converging weighted Neumann Series precoding for massive MIMO systems," *IEEE Wireless Communication Letters*, vol. 7, no. 2, pp. 154–157, 2017.
[6] C. Zhang, Z. Wu, C. Studer, Z. Zhang, and X. You, "Efficient soft-output Gauss-Seidel data detector for massive MIMO systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, pp. 1–12, 2018.
[7] S. Hashima and O. Muta, "Fast matrix inversion methods based on Chebyshev and Newton iterations for zero forcing precoding in massive MIMO systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, pp. 1–12, 2020.
[8] T. Xie, L. Dai, X. Gao, X. Dai, and Y. Zhao, "Low-complexity SSOR-based precoding for massive MIMO systems," *IEEE Communications Letters*, vol. 20, no. 4, pp. 744–747, 2016.
[9] J. Jin, Y. Xue, Y.-L. Ueng, X. You, and C. Zhang, "A split pre-conditioned conjugate gradient method for massive MIMO detection," in *2017 IEEE International Workshop on Signal Processing Systems*. IEEE, 2017, pp. 1–6.
[10] J. Nocedal, S. J. Wright, T. V. Mikosch, S. I. Resnick, and S. M. Robinson, "Numerical optimization," *Springer*, 1999.
[11] J. Nocedal, "Updating quasi-newton matrices with limited storage," *Mathematics of Computation*, vol. 35, no. 151, pp. 773–782, 1980.
[12] D. Ek and A. Forsgren, "Exact linesearch limited-memory quasi-newton methods for minimizing a quadratic function," *Computational Optimization and Applications*, pp. 1–28, 2021.
[13] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, 1989.
[14] J. E. Dennis and J. J. More, "Quasi-newton methods, motivation and theory," *Siam Review*, vol. 19, no. 1, pp. 46–89, 1977.