A Split Pre-Conditioned Conjugate Gradient Method for Massive MIMO Detection

Jiejun Jin[1,2], Ye Xue[1,2], Yeong-Luh Ueng[3], Xiaohu You[2], and Chuan Zhang[1,2,*]
[1]Lab of Efficient Architectures for Digital-communication and Signal-processing (LEADS)
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing, China
[3]Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan
Email: [2]{jjjin, yexue, xhyu, chzhang}@seu.edu.cn, [3]ylueng@ee.nthu.edu.tw

***Abstract*—In massive multiple-input multiple-output (MIMO) mobile system, the computational complexity of signal detection increases exponentially along with the growing number of antennas. For example, the sub-optimal linear detection schemes, such as zero forcing (ZF) detector and minimum mean square error (MMSE) detector, always have to balance the performance and complexity resulted from the large-scale matrix inversion operations. Recently, some iterative linear solvers, such as conjugate gradient (CG), have been proposed to address this issue. These series of detection algorithms offer a better tradeoff between error-rate performance and computational complexity by avoiding the computation-hungry operations like matrix inversion. However, when the the system loading factor ρ goes up, their results are no longer satisfactory. To solve the aforementioned issues, this paper 1) first introduces a novel, low-complexity pre-conditioner by exploring the properties of the equalization matrix and 2) then proposes a split pre-conditioned conjugate gradient (SPCG) method to speed up the convergence rate of detection. Both analytical and numerical results have demonstrated the performance and complexity advantages of the proposed algorithm over the sate-of-the-art ones. The proposed detector outperforms the conventional CG detector with around 2 dB for BER $= 10^{-4}$. When the number of user antennas is relatively large, its complexity is only 25% of the existing pre-conditioned conjugate gradient detector based on incomplete Cholesky decomposition (ICCG).**

***Keywords*—Massive MIMO, iterative linear detection, conjugate gradient, pre-conditioner.**

I. Introduction

Massive multiple-input multiple-output (MIMO) system [1], also known as large-scale MIMO system, refers to a system where hundreds of antennas are equipped at the base station (BS) to serve tens of users simultaneously. By making higher use of spatial resources via a large scale of antennas array, massive MIMO systems show obvious advantages in improving the system capacity without increasing the spectral resources and transmitting power [2]. Therefore, massive MIMO has been regarded as one of the core techniques for the next generation mobile communication.

However, one important issue which prevents massive MIMO from practical applications is the complexity lying in detection. The dramatically increasing number of antennas results in a considerably high complexity, which makes the problem very intractable. The optimal detection method, maximum likelihood (ML), will be prohibitive because of its exponential complexity. To reduce the complexity while maintaining a acceptable performance, a variety of algorithms are proposed, among which linear detection methods, such as zero forcing (ZF) and minimum mean square error (MMSE), can achieve sub-optimal performances with low complexity [1]. However, requiring exact matrix inversion operations with complexity up to the order of $\mathcal{O}(N^3)$, both ZF and MMSE suffer from an inherent high complexity, as the number of user antennas (N) grows [3].

To this end, some approximate methods or iterative linear solvers have been considered to avoid the matrix inversion. The former ones include the Neumann series expansion (NSE) [4–6], whereas the latter ones include conjugate gradient (CG) method and so on. Among them, NSE algorithm reduces the computation load by calculating the inverse of the diagonalized matrix with finite series terms other than Cholesky decomposition [7]. However, its complexity returns $\mathcal{O}(N^3)$ when the number of NSE terms becomes more than two. As for CG method proposed in [8–10], which is another matrix-inversion free approach, it is useful when solving the large-scale problem for its low computational complexity [8, 11]. But there is an obstacle: when the system loading factor, which refers to the user-to-BS antenna ratio increases, the convergence of CG approach becomes slower due to the growing condition number of the corresponding matrix [12].

Therefore, the aforementioned ineffective solvers do not meet the requirements which emphasize the balance between good performance and low complexity. Regarding the performance issue, one potential technique to accelerate the convergence of conventional iterative linear solvers is pre-condition. [13] introduces an incomplete Cholesky decomposition (IC) pre-conditioner into CG, which shows efficiency in improving the performance. However, IC based pre-conditioner will also lead to a high complexity due to the calculation of the pre-conditioned matrix. In addition, the lower-triangular pre-conditioned matrix requires more storage space during the detection process.

In order to eliminate the aforementioned shortcomings of existing detection methods, this paper devotes itself in proposing a new pre-conditioner for a better tradeoff between complexity and performance. It has been shown that the proposed pre-conditioner costs very little complexity and storage space while playing an efficient role in speeding up the convergence of the iterative algorithm. This method makes full use of the diagonal-dominate property of matrix $\mathbf{H}^H\mathbf{H}$, and a new matrix is constructed to optimize the original one. The new matrix is obtained through matrix multiplication to normalize the diagonal elements in the corresponding matrix. Since the

978–1–5386–0446–5/17/$31.00 ©2017 IEEE

multiplication is operated at both sides of the original matrix, we entitle the proposed algorithm the split pre-conditioned conjugate gradient (SPCG) method. The convergence is analysed based on *Gershgorin Disc Theorem*. Numerical results have shown that this proposed SPCG detection method outperforms benchmark solutions in massive MIMO. Furthermore, this pre-conditioned method also requires less storage space than the state-of-the-art pre-conditioners. In fact, the operations of acquiring the proposed pre-conditioned matrix are more accessible, and consequently SPCG is supposed to be more efficiently when applied to the massive MIMO systems.

The remainder of this paper is organized as follows. Section II describes the massive MIMO system model as well as the MMSE detection scheme. Section III proposes the SPCG algorithm. Its complexity analysis is given in the same section. Numerical results and the corresponding explanation are given in Section IV. Finally, Section V concludes the entire paper.

Notation: In this paper, lower-case and upper-case boldface letters indicate column vector and matrix, respectively. The operations $(\cdot)^T$, $(\cdot)^H$, and $\mathbb{E}\{\cdot\}$ stand for transpose, conjugate transpose and expectation, respectively. The entry in the i-th row and j-th column of $\mathbf{A}$ is $\mathbf{A}_{(i,j)}$; the k-th entry of $\mathbf{a}$ is a_k. The vector $\mathbf{a}$ in the kth iteration is $\mathbf{a}^{(k)}$. The computational complexity is denoted in terms of the required number of complex-valued multiplications in this paper.

II. Preliminaries

In this paper, we focus on a massive MIMO system with M antennas at BS concurrently receiving N data streams ($M \geq N$) from the users terminal (UT).

A. Uplink Massive MIMO System

Define the system loading factor $\rho = N/M$ $(0 < \rho \leq 1)$. For typical massive MIMO uplink case, the received symbol $\mathbf{y} \in \mathbb{C}^M$ at BS can be described by:

$$\mathbf{y} = \sum_{k=1}^{N} \mathbf{h}_k s_k + \mathbf{n}, \tag{1}$$

where $s_k \in \mathcal{O}^N$ $(k = 1, 2, ..., N)$ represents the transmitted symbol of the k-th user at UT and $\mathcal{O}$ corresponds to the 2^Q-QAM constellation. It is assumed that transmitting power of each user is $\mathbb{E}\{\|s_k\|^2\} = E_s = 1, \forall s$. $\mathbf{n} \in \mathbb{C}^M$ is the addictive white Gaussian noise (AWGN) with zero-mean and a variance of σ^2. Since $\mathbf{h}_k$ denotes a vector, we can rewrite the Eq. (1) in the form of combination:

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

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ...\mathbf{h}_N] \in \mathbb{C}^{M\times N}$ is the flat Rayleigh fading channel matrix, whose entries are assumed to be independent identical distributed (i.i.d.) with zero-mean and unit variance.

B. MMSE Detection Scheme

In order to minimize the mean square error (MSE) between the estimated symbol $\hat{\mathbf{s}}$ and the real transmitted symbol $\mathbf{s}$, one can resort to the MMSE equalizer at BS according to [14]. As a result, the estimated symbol $\hat{\mathbf{s}}$ obtained after the MMSE detector can be written as follows:

$$\hat{\mathbf{s}} = (\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^H\mathbf{y}, \tag{3}$$

where $\mathbf{I}$ represents a $N \times N$ identity matrix. For the sake of convenience, we denote:

$$\tilde{\mathbf{y}} = \mathbf{H}^H\mathbf{y}, \tag{4}$$

and the MMSE detection matrix can be written as:

$$\mathbf{A} = \mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}. \tag{5}$$

In this way, the Eq. (3) can be simplified as $\hat{\mathbf{s}} = \mathbf{A}^{-1}\tilde{\mathbf{y}}$. Since we assume that $\mathbf{H}$ and σ^2 are perfectly known by the receiver, and the average SNR per receive antenna can be computed as $\zeta = \mathbb{E}\{\|s_k\|^2\}/\sigma^2$ $(\forall s_k \in \mathbf{s})$, the result of the estimated symbol $\hat{\mathbf{s}}$ can be obtained theoretically by solving linear equations. However, if the exact matrix inversion method such as Cholesky decomposition is chosen to deal with this problem, the outcome will be disastrous because of the overwhelming complexity as user number grows, and it impedes MMSE detection for massive MIMO system.

III. Proposed Split Pre-Conditioned Conjugate Gradient Method

In this section, the traditional CG algorithm for MMSE linear detection is reviewed first, followed by pointing out its disadvantages when applying in systems where loading factor ρ is comparatively large. Then, an improved iterative method called SPCG is introduced. We compare this new way of pre-conditioner with other approaches. The proof of convergence and analysis of its complexity are also discussed in this section.

A. Conjugate Gradient Method

Without the loss of generality, as it is described in Section II-B, the MMSE detector forms the detection problem into a typical linear equation $\hat{\mathbf{s}} = \mathbf{A}^{-1}\tilde{\mathbf{y}}$, which is a linear function problem. The conventional CG method is one of the most effective ways to solve the linear equations, and the ith iteration result of the estimated symbol $\hat{\mathbf{s}}$ can be obtained as follows:

$$\hat{\mathbf{s}}^{(i)} = \hat{\mathbf{s}}^{(i-1)} + \frac{\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)}{\left(\mathbf{A}\tilde{\mathbf{p}}^{(i-1)} \cdot \tilde{\mathbf{p}}^{(i-1)}\right)}\tilde{\mathbf{p}}^{(i-1)}, \tag{6}$$

$$\tilde{\mathbf{y}}^{(i)} = \tilde{\mathbf{y}}^{(i-1)} - \frac{\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)}{\left(\mathbf{A}\tilde{\mathbf{p}}^{(i-1)} \cdot \tilde{\mathbf{p}}^{(i-1)}\right)}\mathbf{A}\tilde{\mathbf{p}}^{(i-1)}, \tag{7}$$

$$\tilde{\mathbf{p}}^{(i)} = \tilde{\mathbf{y}}^{(i)} + \frac{\left(\tilde{\mathbf{y}}^{(i)} \cdot \tilde{\mathbf{y}}^{(i)}\right)}{\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)}\tilde{\mathbf{p}}^{(i-1)}. \tag{8}$$

Moreover, the channel matrix $\mathbf{H}$ proves to be asymptotically orthogonal when M is much larger than N [1], which implies that the matrix $\mathbf{A}$ obtained through MMSE detection can be regarded as a diagonally dominant one [15]. This property of $\mathbf{A}$ contributes much to the proposed pre-condition process. Furthermore, $\mathbf{A}$ has a trait of being Hermitian positive definite (HPD) [16]. All these factors make it reasonable to use CG algorithm in the process, since it is quite suitable to cope with such question with conditions providing a very small loading factor.

However, when ρ increases, the situation is different. Such detection results tend to be unsatisfactory because of the slow convergence rate of CG algorithm when N becomes comparable as M. It is clearly shown in [13] that if we keep the iteration number $k = 3$, when $N = 16$, the BER performance of the classical CG method is over 10^{-4} at SNR $= 20$ dB. On the contrary, the BER performances are only approximate to 10^{-2} and 10^{-1} for the numbers of user antennas are 32 and 48, respectively, indicating the ineffectiveness under such circumstances.

Also, as the convergence of CG algorithm is decided by the condition number of $\mathbf{A}$ as described in [12], the performance of this iterative method is closely related to Gram matrix of MMSE, $\mathbf{A}$. If the condition number of $\mathbf{A}$ tends to be lager, the results may even not converge at all in certain extreme cases.

B. Proposed SPCG Method

To improve the performance of CG algorithm when ρ is medium, a large number of pre-conditioned methods have been introduced. One of them is based on incomplete Cholesky (IC) factorization [13], which contains the difficulty of determining the pre-conditioned matrix $\mathbf{L}$. In addition to the complex calculation algorithm, more space on the hardware is needed to store the lower triangle matrix $\mathbf{L}$ in detection process.

Since the lack of robustness is a widely recognized weakness of CG iteration method, and the ICCG process mentioned above still suffers from some disadvantages, here we propose a proper pre-conditioner which takes the characteristic of Gram matrix $\mathbf{A}$ into account. This pre-conditioned algorithm, whose name is the split pre-conditioned conjugate gradient (SPCG) method, will transform the original question into a similar but faster convergent problem that can be solved by CG scheme finally.

In general, this algorithm is derived from the consideration of the condition number of matrix. Since the condition number of $\mathbf{A}$ plays an important role in convergence rate, the main aim is to control the condition number through pre-conditioner. Due to the fact that Gram matrix $\mathbf{A}$ is diagonally dominant, the main diagonal elements are greater than the others. Consequently, conclusion can be drawn that the condition number of $\mathbf{A}$ largely depends on the elements of main diagonal. Moreover, if the difference between the main diagonal elements becomes greater, the value of condition number of $\mathbf{A}$ will accordingly increases, which leads to a harder convergence. The proof of this viewpoint will be shown in Section III-C.

Based on the discussion above, we apply SPCG algorithm for the purpose of normalizing the diagonal elements of Gram matrix. First, we can use a diagonal matrix $\mathbf{B}$ to transform $\mathbf{A}$. The diagonal elements of $\mathbf{B}$ can be attained as:

$$\mathbf{B}_{(i,i)} = \frac{1}{\sqrt{\mathbf{A}_{(i,i)}}}, \tag{9}$$

where $i = 1, 2, ..., N$.

Since the pre-conditioned matrix here is an $N \times N$ diagonal one, the space to store $\mathbf{B}$ is smaller than that for matrix $\mathbf{L}$ in IC method.

After preparation in Eq. (9), the new matrix $\mathbf{A}' = \mathbf{BAB}$ with unit diagonal elements can be used as a substitute for the original $\mathbf{A}$. To guarantee an equivalent form of the initial linear equation $\mathbf{A}\hat{\mathbf{s}} = \tilde{\mathbf{y}}$, $\hat{\mathbf{s}}' = \mathbf{B}^{-1}\hat{\mathbf{s}}$ and $\tilde{\mathbf{y}}' = \mathbf{B}\tilde{\mathbf{y}}$ are introduced to replace $\hat{\mathbf{s}}$ and $\tilde{\mathbf{y}}$, respectively in subsequent operations. As a result, the equation converts to $\mathbf{A}'\hat{\mathbf{s}}' = \tilde{\mathbf{y}}'$, which can be solved by the conventional CG method now, and the solution $\hat{\mathbf{s}}$ to the original problem can be denoted as $\hat{\mathbf{s}} = \mathbf{B}\hat{\mathbf{s}}'$.

In the following, the proposed SPCG algorithm is given in detail:

Algorithm 1 Proposed SPCG Algorithm

Input:

Matrix $\mathbf{A}$, vector $\tilde{\mathbf{y}}$

1: Compute $\mathbf{B}$ by Eq. (9).

2: $\hat{\mathbf{s}}^{(0)} = 0, \tilde{\mathbf{y}}^{(0)} = \mathbf{B}\tilde{\mathbf{y}}, \tilde{\mathbf{p}}^{(0)} = \tilde{\mathbf{y}}^{(0)}$

3: **for** $i = 1, \ldots, k$ **do**

4: $\quad \hat{\mathbf{s}}^{(i)} = \hat{\mathbf{s}}^{(i-1)} + \frac{\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)}{\left(\mathbf{BAB}\tilde{\mathbf{p}}^{(i-1)} \cdot \tilde{\mathbf{p}}^{(i-1)}\right)} \mathbf{B}\tilde{\mathbf{p}}^{(i-1)}$

5: $\quad \tilde{\mathbf{y}}^{(i)} = \tilde{\mathbf{y}}^{(i-1)} - \frac{\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)}{\left(\mathbf{BAB}\tilde{\mathbf{p}}^{(i-1)} \cdot \tilde{\mathbf{p}}^{(i-1)}\right)} \mathbf{BAB}\tilde{\mathbf{p}}^{(i-1)}$

6: $\quad \tilde{\mathbf{p}}^{(i)} = \tilde{\mathbf{y}}^{(i)} + \frac{\left(\tilde{\mathbf{y}}^{(i)} \cdot \tilde{\mathbf{y}}^{(i)}\right)}{\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)} \tilde{\mathbf{p}}^{(i-1)}$

7: **end for**

Output:

$\hat{\mathbf{s}} = \hat{\mathbf{s}}^{(k)}$

C. Proof of Convergence

It has been proved in [17] that the convergence of CG method largely depends on the condition number of $\mathbf{A}$ represented in the form of 2-norm. As described in Section III-A, the matrix $\mathbf{A}$ has a feature of being HPD, which means the condition number of $\mathbf{A}$ can be written as:

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}, \tag{10}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ represent the largest and the smallest one of the eigenvalues of $\mathbf{A}$, respectively.

More exactly, if we assume that $\hat{\mathbf{s}}^{(*)}$ is the accurate solution of $\hat{\mathbf{s}}$ in the equation $\mathbf{A}\hat{\mathbf{s}} = \tilde{\mathbf{y}}$, the error estimation of the kth iteration can be illustrated as follows:

$$\|\hat{\mathbf{s}}^{(*)} - \hat{\mathbf{s}}^{(k)}\|_{\mathbf{A}} \leq 2\left[\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right]^k \|\hat{\mathbf{s}}^{(*)} - \hat{\mathbf{s}}^{(0)}\|_{\mathbf{A}}, \tag{11}$$

where $\|\mathbf{x}\|_{\mathbf{A}} = \sqrt{\mathbf{Ax} \cdot \mathbf{x}}$.

According to the relationship in the Eq. (11), the convergence rate increases as the condition number of $\mathbf{A}$ approaches 1. In other words, the main target is to reduce the gap between the largest and the smallest eigenvalue of $\mathbf{A}$. Therefore, SPCG method is applied to solve the problem.

The reason for introducing such a method can be obtained rigorously using some principals in mathematics. The well-known Gershgorin disc theorem indicates the relationship between diagonal elements and eigenvalues of a matrix. Supposing a complex $N \times N$ matrix $\mathbf{A}$, for $i = 1, 2, ..., N$, let $\mathbf{R}_i = \sum_{j=1, j\neq i}^{N} |\mathbf{A}_{(i,j)}|$ be the sum of the absolute value of the non-diagonal entries in the i-th row. We note $\mathbf{D}(\mathbf{A}_{(i,i)}, \mathbf{R}_i)$ as the closed disc centered at $\mathbf{A}_{(i,i)}$ with radius of $\mathbf{R}_i$ and such a disc is called a Gershgorin disc. The theorem points out that every eigenvalue of $\mathbf{A}$ lies within at least one of the Gershgorin discs, or in other words, all the eigenvalues including the largest and the smallest are located in the area which is composed by the sum aggregate of all discs.

In the signal detection process, Gram matrix of MMSE, $\mathbf{A}$, accords with the condition in the theorem, so a similar result can be obtained easily. And what has been done in the pre-condition algorithm is to normalize the diagonal elements in the matrix so that all of them become the same number, that is $\mathbf{A}'_{(i,i)} = 1$, which indicates all centers of the Gershgorin disc are moved to the same point. Furthermore, as the order of magnitude in $\mathbf{A}$ is almost the same as that of antennas at BS [18], the radius of every disc $\mathbf{R}_i$ also becomes smaller. All these changes will lead to a much smaller area containing the eigenvalues of the new matrix $\mathbf{A}'$ compared to the previous one. It is quite definite that the distance between the biggest element and the smallest one decreases considerably in the process, and as a result, the condition number of the new matrix that is calculated by the ratio of the largest to the smallest eigenvalue is smaller than the former one, which provides a better performance illustrated by simulation results in Section IV-A.

It is shown clearly from Fig. 1 that the method can reduce the condition number, confirming the proof in mathematics.

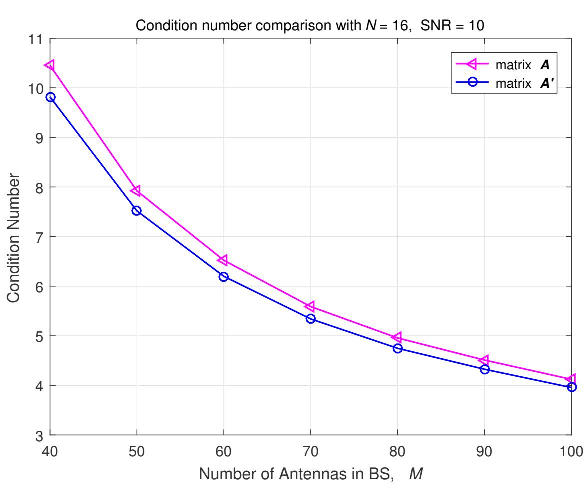

Fig. 1. Results of condition numbers of MMSE detection $\mathbf{A}$ without pre-conditioned method and matrix $\mathbf{A}'$ using the proposed pre-conditioner, respectively.

In Fig. 1, comparison of the condition numbers of matrix $\mathbf{A}$ without pre-conditioned method and matrix $\mathbf{A}'$ using the proposed pre-conditioner is displayed. Here, we fix the number of antennas at UT as well as the value of SNR while changing the value of M to picture the curve. It is illustrated that as the number of antennas at BS increases, both condition numbers show a downward trend. However, with a reduction of 6%, the condition number of new matrix $\mathbf{A}'$ is always smaller than that of the original one. Without the loss of generality, the data shown in the graph are obtained from the statistical results which are acquired by changing the channel matrix $\mathbf{H}$ which leads to changes in the MMSE detection matrix $\mathbf{A}$ and using the average finally.

D. Complexity Analysis

In this part, we will compute the algorithm complexity in terms of the required number of complex-valued multiplications in each step. Owing to the diagonal pre-conditioned matrix $\mathbf{B}$, the complexity of multiplication reduces much in each step. The results provide a clear demonstration about the advantages of this SPCG algorithm.

1) Compute $\mathbf{B}$*:* The calculations of the square root as well as its inverse in this algorithm can be implemented through the look-up-table (LUT) by hardware. Therefore, according to the definition of complexity here, there is no required number of complex-valued multiplication, which indicates no need to consider the complexity. In addition, since the diagonal elements of $\mathbf{A}$ are essentially on the same order of magnitude, the memory for LUT is acceptable.

2) Compute $\tilde{\mathbf{y}}^{(0)}$ *and* $\tilde{\mathbf{p}}^{(0)}$*:* In Algorithm 1, $\tilde{\mathbf{y}}^{(0)}$ and $\tilde{\mathbf{p}}^{(0)}$ both equal $\mathbf{B}\tilde{\mathbf{y}}$, and as a result, the complexity of the multiplication of an $N \times N$ diagonal matrix with an $N \times 1$ vector is N.

3) Compute $\hat{\mathbf{s}}^{(i)}$*:* The computation of $\hat{\mathbf{s}}^{(i)}$ includes a multiplication of an $N \times N$ diagonal matrix with an $N \times N$ matrix: $\mathbf{BA}$, a multiplication of an $N \times N$ matrix with an $N \times 1$ vector: $\mathbf{BA}(\mathbf{B}\tilde{\mathbf{p}}^{(i-1)})$, a multiplication of an $N \times N$ diagonal matrix with an $N \times 1$ vector: $\mathbf{B}\tilde{\mathbf{p}}^{(i-1)}$, 2 inner products of two $N \times 1$ vectors and a scalar multiplication with an $N \times 1$ vector. In conclusion, the complexity in this step is $(N^2 + N^2 + N + 2N + N = 2N^2 + 4N)$.

4) Compute $\tilde{\mathbf{y}}^{(i)}$*:* As can been seen in the expression of $\tilde{\mathbf{y}}^{(i)}$, the value of $\frac{\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)}{\left(\mathbf{BAB}\tilde{\mathbf{p}}^{(i-1)} \cdot \tilde{\mathbf{p}}^{(i-1)}\right)}$ and $\mathbf{BAB}\tilde{\mathbf{p}}^{(i-1)}$ have been calculated in Section III-D3 already, so the remaining task is to consider a scalar multiplication which contains the complexity of N.

5) Compute $\tilde{\mathbf{p}}^{(i)}$*:* Having got the value of $\left(\tilde{\mathbf{y}}^{(i-1)} \cdot \tilde{\mathbf{y}}^{(i-1)}\right)$ in step Section III-D3, what needs to discuss is an inner product of two $N \times 1$ vectors and a scalar multiplication with an $N \times 1$ vector which leads to the complexity of $2N$.

Therefore, the overall complexity of employing SPCG method is given by $(2kN^2 + (7k + 1)N)$.

IV. Numerical Results

In this section, the numerical examples of the BER performance as well as the complexity results are given. All these data show the advantages of this split pre-conditioned CG method from different sides.

A. BER Performance

As explained in Section III-C, the traditional CG method loses its superiorities when the number of antennas at UT becomes comparable to that at BS. Thus, we introduce the

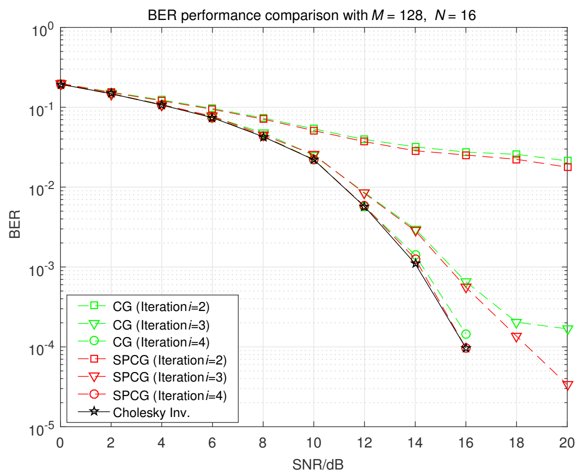

Fig. 2. BER performance of different decoding methods with $M = 128$ and $N = 16$.

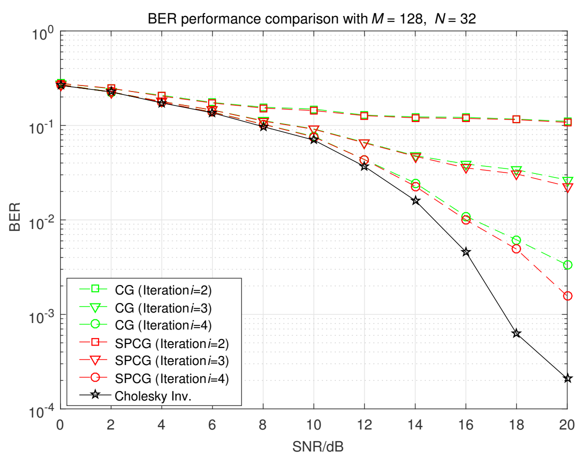

Fig. 3. BER performance of different decoding methods with $M = 128$ and $N = 32$.

pre-conditioner to accelerate the convergence rate under such conditions.

Both Fig.s 2 and 3 show the simulation results of SPCG method as well as the traditional CG method without pre-conditioned process with different loading factors. The C-holesky factorization simulation is provided as an ideal result.

Fig. 2 illustrates BER performance comparison between SPCG algorithm and CG algorithm while the number of antennas at BS (M) and UT (N) equal 128 and 16, respectively. As can be seen, SPCG method outperforms the traditional one and demonstrates a better result with around 2 dB performance improvement at BER of 10^{-4}. Above all, the result of this method is the same as that of the ideal Cholesky factorization when the iteration number reaches four.

In Fig. 3, the number of antennas at UT changes to 32, which provides a larger loading factor ρ. Due to the increase in condition number, both iteration methods slow their iteration rate compared to the results shown in Fig. 2. However, SPCG method still performs better between these methods,

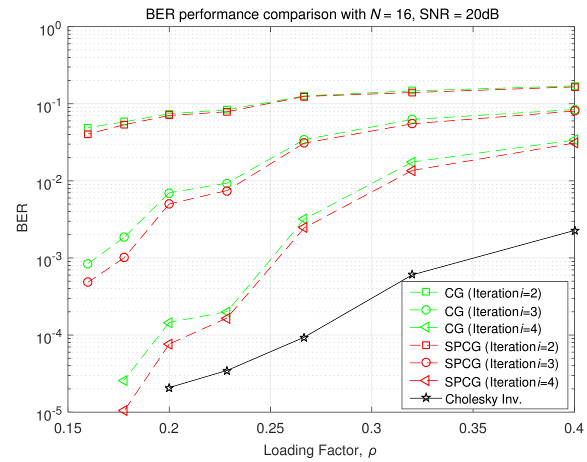

Fig. 4. BER performance of different decoding methods with various loading factors (ρ).

showing its advantages over the traditional CG under such circumstance.

More specifically, if the number of antennas at UT is fixed and that at BS varies only, results in Fig. 4 will present the change rule of BER performance with different ρ directly. Here, we keep $N = 16$ and the SNR $= 20$ dB. A great deal of information can be obtained from Fig. 4. First, it is clearly illustrated that BER performance of CG method becomes less satisfactory when the value of ρ increases. More importantly, although the convergence of SPCG method also becomes harder when ρ is relatively larger, it still outperforms than CG method, revealing its wider range of application.

From all these three graphs, Fig.s 2, 3, and 4, it is clearly displayed that SPCG method indeed outperforms the tradition-al algorithm whether the loading factor is large or relatively small. From this side, it is concluded that SPCG algorithm have a better BER performance than the non-preconditioned CG method.

B. Computational Complexity

Having considered the complexity in Section III-D, in this part, comparison is made between SPCG method, the traditional CG method and the previous introduced ICCG method. The result of the ideal Cholesky factorization method will still be used as a frame of reference.

In Table I, the computational complexity results of different CG methods as well as the Cholesky inverse algorithm are shown, respectively. According to the table, since SPCG and ICCG both include pre-conditioner, their numbers of multiplications are larger than that of CG. But compared with each other, the complexity of SPCG method does not include N^3, which decreases the difficulty in computation when N is large.

Fig. 5 provides some numerical results of comparison with a fixed value of antennas as BS (M) and SNR while a changing amount of antennas at UT (N). As the complexity of CG is obvious in the light of analysis above, here the results of SPCG and ICCG are presented which illustrate that the proposed method shows an advantage over ICCG especially when the

TABLE I. NUMBER OF COMPLEX-VALUED MULTIPLICATIONS

	Symbol $\hat{\mathbf{s}}$
Cholesky Inverse	$\frac{5}{6}N^3+\frac{3}{4}N^2+\frac{4}{3}N$
Traditional CG	$k(N^2+6N)$
ICCG	$\frac{(N^2-2S)^{3/2}-(N^2-2S)^{1/2}}{6}+$ $(2k+1)N^2+(7k-1)N-(2k+2)S$
Proposed	$2kN^2+(7k+1)N$

[13] calculates the complexity of the ICCG method where S represents the number of zero in the lower triangle matrix $\mathbf{L}$.

number of antennas at UT is large. 40% complexity reduction from ICCG method can be easily achieved when N is about 85 and over 75% reduction can be reached when N is approximate to 100. When N is relatively small, although the computational complexity of SPCG method may not stand out compared to the other CG methods, it still saves a lot when it comes to the Cholesky inverse algorithm. As a result, the conclusion is that this proposed method balances the complexity and its performance.

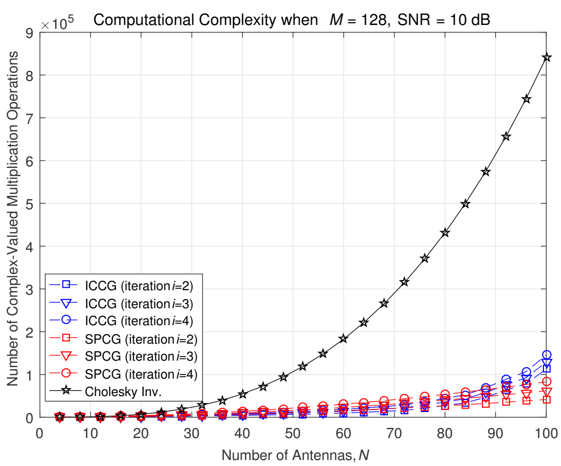

Fig. 5. Complexity comparison with $M = 128$ and SNR $= 10$ dB.

V. CONCLUSION

In this paper, a novel pre-conditioned CG method named SPCG method has been proposed for massive MIMO signal detection. Explanation about the motivation to introduce such pre-conditioner and the detailed algorithm to realize such method are given. Both theoretical and numerical results have proved that this algorithm can reach a tradeoff between BER performance and computational complexity. Further work will be directed towards the hardware architecture and implementation of the SPCG method.

ACKNOWLEDGMENT

This work is partially supported by NSFC under grant 61501116, International Science & Technology Cooperation Program of China under grant 2014DFA11640, Huawei HIRP Flagship under grant YB201504, Jiangsu Provincial NSF under grant BK20140636, Intel Collaborative Research Institute for MNC, the Fundamental Research Funds for the Central Universities, State Key Laboratory of ASIC & System under grant 2016KF007, the Project Sponsored by the SRF for the Returned Overseas Chinese Scholars of State Education Ministry, and Student Research Training Program of Southeast University.

REFERENCES

[1] 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 Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[3] B. Yin, M. Wu, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "A 3.8Gb/s large-scale MIMO detector for 3GPP LTE-Advanced," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 3879–3883.

[4] F. Wang, C. Zhang, J. Yang, X. Liang, X. You, and S. Xu, "Efficient matrix inversion architecture for linear detection in massive MIMO systems," in *Proc. IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, pp. 248–252.

[5] F. Wang, C. Zhang, X. Liang, Z. Wu, S. Xu, and X. You, "Efficient iterative soft detection based on polynomial approximation for massive MIMO," in *Proc. International Conference on Wireless Communications Signal Processing (WCSP)*, Oct 2015, pp. 1–5.

[6] X. Liang, C. Zhang, S. Xu, and X. You, "Coefficient adjustment matrix inversion approach and architecture for massive MIMO systems," in *Proc. IEEE International Conference on ASIC (ASICON)*, Nov 2015, pp. 1–4.

[7] H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear pre-coding systems," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 2710–2715.

[8] B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, "Conjugate gradient-based soft-output detection and precoding in massive MIMO systems," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2014, pp. 3696–3701.

[9] ——, "VLSI design of large-scale soft-output MIMO detection using conjugate gradients," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1498–1501.

[10] Y. Hu, Z. Wang, X. Gaol, and J. Ning, "Low-complexity signal detection using CG method for uplink large-scale MIMO systems," in *Proc. IEEE International Conference on Communication Systems (ICCS)*, 2014, pp. 477–481.

[11] H. MR, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, no. 5, pp. 303–320, 1969.

[12] Y. Saad, *Iterative methods for sparse linear systems*. SIAM, 2003.

[13] Y. Xue, C. Zhang, S. Zhang, and X. You, "A fast-convergent pre-conditioned conjugate gradient detection for massive MIMO uplink," in *Proc. IEEE International Conference on Digital Signal Processing (DSP)*, 2016, pp. 331–335.

[14] A. Sibille, C. Oestges, and A. Zanella, *MIMO: from theory to implementation*. Academic Press, 2010.

[15] G. H. Golub and C. F. V. Loan, *Matrix computations*. Oxford Academic Press, 1984.

[16] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge University Press, 2007.

[17] A. van der Sluis and H. van der Vorst, "The rate of convergence of conjugate gradients," *Numerische Mathematik*, vol. 48, no. 5, pp. 543–560, 1986.

[18] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 916–929, 2014.