A Summary and Analysis of Low Complexity Methods for Massive MIMO Systems

Siyu Chen
Department of Electronic Information Engineering, Sichuan University, No.4, First South Section, First Ring Road, Wuhou District, Chengdu City, Sichuan Province, China
114631646@qq.com

Abstract. The massive multiple-input multiple-output (MIMO) systems help increase the wireless channel capacity and spectral efficiency manifold, and the coverage as well as power consumption. In MIMO systems, the linear minimum mean-square error (MMSE) can achieve the near-performance. However, along with the huge number of the BS-to-user antennas at the BS station, the computational complexity becomes a serious problem due to computing matrix inversion. Many methods were proposed to improve complexity, for instance, Neumann series expansion method, Jacobi methods and conjugate gradient (CG) method, but it is still costly. And recently many other methods on account of the above methods were proposed, respectively. Such as Gauss-Seidel method, Symmetric Successive Over Relaxation (SSOR) method, further improved Jacobi method based on soft-output massive MIMO detection scheme, pre-conditioner conjugate gradient (PCG) based on SSOR, split pre-conditioned conjugate gradient (SPCG) method, etc, which all show the reduced complexity, improved convergence and BER performance.

1. Introduction
In the development of modern communication industry, the 4th generation (4G) technology has been very mature. However, due to the small number of antennas, which is generally 4 or 8, it limits the communication capacity. With the increase in application requirements, 5G, which has been developed on the basis of 4G, has introduced massive multiple-input multiple-out (MIMO), allowing the number of antennas to reach hundreds of thousands. Compared with the original MIMO systems, on the one hand, the massive MIMO systems’ communication capacity is significantly enhanced, as a result of the increase in the number of antennas, and also transmitted data information. Meanwhile in this system, having the characteristic of Beam Spatial Multiplexing, it can be simultaneously communicated at different locations under the same conditions, which greatly improves spectrum efficiency and enables efficient transmission. On the other hand, in order to meet the same range of communications, the number of antennas in massive MIMO system has extremely more than traditional ones, then the weighted synthesis gain increases. Therefore, the signal performance in the corresponding places is strengthened, the communication coverage is expanded as well. Because the communication range is proportional to the power consumption of the device, the loss is accordingly reduced.

However, it involves matrix inversion in the linear detection methods, particularly the minimum mean square error (MMSE) whose core part is to find the corresponding weight matrix, which has the tremendous computational complexity. To solve this problem, many methods have been proposed. Firstly, Cholesky (IC) factorization is generally taken as reference; Neumann series expansion method...
avoids the singularity problems in the inversion of system matrix equations, and its solutions are obtained without dividing many units; Gauss-Seidel method can achieve the computational complexity which reduced from $O(N^3)$ to $O(N^2)$ for arbitrary number of iterations, and the more is that a faster convergence rate can be expected. Moreover the required number of iterations to achieve a certain estimate accuracy becomes smaller; Jacobi method uses iterative matrix coefficients for iterative processing of matrices. The more popular CG method only uses the first derivative information of the matrix and the sparsity of the coefficient matrix. Based on the above methods, many other methods have been proposed. The SSOR method provides the appropriate relaxation parameters and the initial values. Especially the PCG and SPCG methods are greatly suitable for the conditions of positive symmetric of coefficient matrix, which take use of the diagonally dominant of matrix. They all use the main diagonal elements to replace the entire matrix.

The above methods, utilizing the features of the coefficient matrix, effectively solve the problem of computational complexity in different aspects. Nonetheless, these methods have their own disadvantages. Such as CG method’s convergence is so slow that its performance cannot be obviously improved to a certain extent. And as to the SSOR method, the relaxation factor $\omega$ is a pivotal factor but difficult to achieve.

The rest of the paper is organized as follows. Section II briefly describes the system model of massive MIMO system. Section III specifically introduces the low-complexity methods. Section IV performs the correlation analysis of these methods and discusses other methods to have the improvements. Finally, conclusions are drawn in Section V.

Notation: $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose, respectively; Lower-case and upper-case boldface letters denote vectors and matrices, respectively; $\mathbf{I}$ is the unit matrix.

2. System model

In a flat massive MIMO system with N BS antennas, which can establish communication with M independent users, usually we have $M \gg N$, the received signal vector $\mathbf{y}$ at the BS can be described as

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N]$ is the channel matrix whose i-th row and j-th column denote the complex channel gain from the i-th transmit antenna to the j-th receive one, $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$ is the transmitted symbol vector and $\mathbf{n}$ is the additive white Gaussian noise whose variance is $\sigma^2$.

The general receiver's detector is expressed as

$$\hat{x} = w^T\mathbf{y}$$

where $w$ is the linear transformation matrix and $\hat{x}$ is the estimated value of the transmitted symbol vector. The error between the actual transmitted signal and the detected value definitely exists. Through the MMSE detection method, considering the noise factor, the $\hat{x}$ can be written as

$$\hat{x} = \mathbf{A}^{-1}\mathbf{y}$$

where $\mathbf{y} = \mathbf{H}^H\mathbf{y}$, and the MMSE detection matrix $\mathbf{A}$ can be shown as

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

It can be known that $\mathbf{H}$ is asymptotically orthogonal in the massive MIMO systems, indicating that $\mathbf{A}$ is the diagonally dominant which is called symmetric positive definite (SPD).

3. Proposed low-complexity methods

In this section, the concrete implementation steps of each method have been given. In this way they can provide the basis for discussion and analysis in the next section.

3.1. Neumann Series Expansion Method
The MMSE detection performs better than ZF. However, it needs to solve the matrix inversion so that the computational complexity is \( O(N^3) \), which is a huge loss. People proposed the Neumann series expansion method to reduce the complexity. The \( A^{-1} \) can be renewed as

\[
A^{-1} = \sum_{n=0}^{\infty} (X^{-1}(X - A))^n X
\]

when \( A \) is approximate to invertiable matrix \( X \). \( A \) can be decomposed as

\[
A = D + E
\]

where \( D \) is a diagonal matrix and \( E \) is a hollow matrix. As a result of that \( A \) is SPD, we can have \( A \approx D \). Then equation (5) will be rewritten as

\[
A^{-1} = \sum_{n=0}^{\infty} (-D^{-1}E)^n D^{-1}
\]

And its i-th of the Neumann series expansion is

\[
A_i^{-1} = \sum_{n=0}^{i-1} (-D^{-1}E)^n D^{-1}
\]

From the above equations (7) and (8), it can be concluded that a lower complexity can be obtained when the number of items \( i \) is small, through using the series expansion approach, for instance, its complexity is \( O(N^2) \) as \( i = 2 \), while it costs high when \( i \) is large.

Here we set the number of antennas to be 64 × 16 and 128 × 16, respectively. Through comparing the BER performance with the changes of SNR under different iteration times \( i \), we can find that the BER performance is gradually improved with the increase of \( i \), when SNR keeps constant as SNR = 14 dB. However, the difference between it and the ideal MMSE detection algorithm is large, which can only reach 10\(^{-1}\) when it is 64×16, indicating that the superiority of this method is limited. In addition, in the condition of two different antenna configurations and \( i = 4 \), the required SNR is more than 14 dB when the antenna configuration is 64×16, while it only needs 12 dB to obtain BER performance of 10\(^{-4}\) when it is 128×16. All of the above show that Neumann series expansion method is suitable for the massive MIMO systems, which need small number of iterations.

3.2. Gauss-Seidel (GS) method

The GS method is an improvement over Jacobi iteration method. As is known from equation (4), the coefficient matrix \( A \) is SPD and also Hermitian positive definite. According to this characteristic, then the \( A \) can be decomposed as

\[
A = D + L + L^H
\]

where \( D, L, L^H \) is the diagonal matrix, the strictly lower triangular matrix and the strictly upper triangular matrix of \( A \), respectively. Then we can deduce the estimated transmitted signal vector \( \hat{S} \) as

\[
s^{(0)} = (D + L)^{-1} (\hat{Y} - L^H s^{(0)})
\]

where \( i \) is the number of iterations, \( i = 1, 2, 3... \) and \( s^{(0)} \) is the initial solution, usually setting to zero vector.

From equation (10), it can be known that it is assumed that the data \( s^{(0)} \) obtained this time performs better than \( s^{(i-1)} \), namely, each component in \( s^{(0)} \) is superior to the corresponding component in \( s^{(i-1)} \). And when solving for the i-th component in \( s^{(0)} \), the (i-1)-th components have been found. To put it simple, it replaces the old elements with new ones. Therefore, it can be used directly, which has faster convergence than Jacobi for any initial solution in most cases.

Using this method to compute the likelihood ratio (LLR), it can obtain the near-optimal BER performance as well. Comparing the BER performance at different iteration times \( i \) with changes in SNR, simulations are performed with this application for the same exact LLRs. It can be concluded that
under the situation that the SNR is constant, such as SNR = 16 dB, the BER curve can only gain nearly $10^{-4}$ when $i = 2$. However, when $i = 4$, it can achieve the approximate depth as MMSE algorithm, which is $10^{-3}$, with merely 0.1 dB difference. And the result under the condition that $i = 3$ is similar to that in $i = 4$. It is seen that it can be barely obtained under the certain conditions.

We can also know the complexity of GS method through simulating the number of complex multiplications, which usually made the Cholesky decomposition as the baseline. When the number of users N is 32 and the iterative times $i = 2$, 3 or 4, even any number, the complexity is extremely lower than the Cholesky, which shows great advantages.

3.3. Conjugate Gradient (CG) Method

This CG method is an iterative method, taking use of the sparseness and SPD of matrix $A$ to avoid the inversion, which is mainly applicable to large-scale systems ($M \gg N$). It combines the conjugate property with the Steepest Descent method, uses a gradient at the known point to construct a set of conjugate directions, then searches along this set of directions and finds the minimum point of the objective function.

It needs to find an initial value $x^{(0)}$, then according to equation (3) and the above steps, we can have the i-th iteration as follows:

$$
\hat{x}^{(i)} = x^{(i-1)} + \frac{(y^{(i-1)} \cdot \tilde{y}^{(i-1)})}{(Ap^{(i-1)} \cdot \tilde{p}^{(i-1)})} p^{(i-1)}
$$

(11)

$$
\tilde{y}^{(i)} = y^{(i-1)} - \frac{(y^{(i-1)} \cdot \tilde{y}^{(i-1)})}{(Ap^{(i-1)} \cdot \tilde{p}^{(i-1)})} Ap^{(i-1)}
$$

(12)

$$
\tilde{p}^{(i)} = y^{(i)} + \frac{(y^{(i)} \cdot \tilde{y}^{(i)})}{(y^{(i-1)} \cdot \tilde{y}^{(i-1)})} \tilde{p}^{(i-1)}
$$

(13)

From equations (11), (12) and (13), we can have the CG method do not need to pre-estimate other parameters. Furthermore, it is an easy method to solve the equations and even obtain the accurate approximate solutions with very few iteration.

However, the deficiency is that this method’s convergence speed is related to the distribution of the solutions of the coefficient matrix $A$. When the eigenvalues are relatively concentrated, the number of conditions of $A$ decreases, the convergence tends to be quick and gets better performance such as BER performance, which is close to MMSE algorithm. Otherwise, the speed and performance are neither fast nor satisfied.

Through the simulation of the traditional CG method with different number of iterations $i$ and users $M$, when the number of BS antennas $N$ is 128, the characteristics and changes of corresponding BER performance associated with changes in SNR are obtained. For example, in the case of SNB = 18 dB and keeping $i = 4$, we can see that the BER curve can only get nearly $10^{-1}$ when $N = 48$, while it can achieve the result close to $10^{-3}$ when $N = 32$. Not to mention that when $N = 16$, it can achieve the BER curve of less than $10^{-5}$. It is obvious that the convergence rate slows when $N$ is approaching to $M$, which is the shortcoming of CG method.

3.4. Pre-condition Conjugate Gradient (PCG) Method

The PCG method is an optimization solution of CG method. It improves the slow convergence speed by introducing a pre-conditioner, which makes the speed depend on its matrix $M$. Through choosing a good pre-conditioner, the number of matrix conditions is greatly reduced to ensure the efficient calculation, less consumption and complexity.

A good pre-conditioning matrix $M$ should be symmetry positive definite, similar to the sparseness of matrix $A$ and easy to solve linear equations after introducing it. The commonly used pre-conditioning
methods include incomplete Cholesky pre-conditioning, Symmetric Successive Over Relaxation (SSOR)-based and Jacobi-based methods, etc. Here is the method based on SSOR.

We can split \( A \) into the patterns in equation (9), a diagonal matrix, a strict lower triangular matrix and a strictly upper triangular matrix. Then we can iteratively compute the lower and upper triangular matrices with the new values \( s^{(i)} \) each time. It uses the SOR method, which is the process of taking a weighted average between the new and old values with lower triangular matrices, to compute the iteration can be shown as

\[
(D + \omega L)s^{(i+1/2)} = (1-\omega)Ds^{(i)} - \omega L^H s^{(i)} + \omega \tilde{y} 
\]

The iteration with upper triangular matrix is

\[
(D + \omega L^H)s^{(i+1/2)} = (1-\omega)Ds^{(i+1/2)} - \omega L s^{(i+1/2)} + \omega \tilde{y} 
\]

where \( \omega \) is the relaxation parameter, and when \( 0 < \omega < 2 \), it is guaranteed to be convergent.

One of the problems with this method is to gain the corresponding \( \omega \), which decides whether the systems have the high convergence and convergence speed. Although it seems that people have found the appropriate value in theory, the selection of it is still difficult for the large matrices in practice. As to the complexity, we can do the same simulation as in the GS method, and the result can be shown that its complexity is significantly reduced comparing to ZF and Neumann-based signal detection at any number of iterations, which the complex multiplications' times are all lower than \( 10^{-4} \).

Then the BER performance is simulated and the comparison is taken among the Neumann-based, ZF and SSOR-based detector. Under the condition that \( M = 256 \) and \( N = 32 \), all of these methods can have the deeper BER curve with the changes of SNR, which varies from 0 to 30 dB. However, the SSOR-based method can gain the better BER performance. For example, when the number of iterations \( i = 2 \), it can only gain the BER curve of \( 10^{-2} \) for Neumann-based signal detector, while the BER curve can achieve \( 10^{-5} \), which is a difference of three orders of magnitude. And the BER will have a sharp drop when \( i \) increases. Much more, it can be concluded that the performance of this method can be very close to the ZF detector with small \( i \).

4. Analysis and discussion

From the methods described in section III, it can be known that GS, CG and PCG method are the different ways of iterating, which can effectively solve large-scale linear equations, except for the Neumann series expansion method that only applies to cases with small numbers of iterations, which avoids the loss of performance. On the one hand, they have high stability, good convergence and can achieve the detection performance level of MMSE within a small number of iterations. On the other hand, they do not require any external parameters. Especially all the improved method based on CG use the peculiarity of matrix \( A \) that is diagonally dominant and we can use the main diagonal elements to replace the whole matrix \( A \), which greatly reduce the complexity.

In terms of the above-mentioned CG method, a suitable initial estimate is needed at the beginning in order to achieve faster convergence in a reasonable and short period of time. There are many methods for the pre-conditioner, which all can speed up the convergence to some extent and provide less computational complexity, such as implement LU, SSOR, Vanka, etc. Therefore, I consider that we can also use the idea of Neumann series expansion to construct a pre-conditioner.

It is known that the coefficient matrix \( A \) converges quickly when there are only a few irrelevant eigenvalues. Consequently, we can use the Neumann series expansion method to convert \( A \) into a matrix with only a few irrelevant eigenvalues, and then use the CG method to solve it. Firstly according to its diagonally dominant feature, we can transform \( A \) into the diagonal matrix \( D \) and \( E \) as in equation (6), which has better spectral properties to form a diagonal pre-conditioner matrix, and then resolve equation (3) with as few iterations as possible, using the iterative CG method.

In this case, it is possible to remove relatively unnecessary elements since the diagonal elements carry most of the important information of this matrix. We can not only decrease the number of
iterations significantly comparing to the traditional CG method, save computing time or reduce computational complexity, but also do not occupy more storage for the data processing computer.

5. Conclusions
In this paper, four methods for solving linear equations in massive MIMO systems are introduced. These methods are all effective to reduce the computational complexity and increase the convergence speed, but they are still not completely satisfied to the developing of the systems. We discuss a new improving method, which combines Neumann series expansion method with PCG method. In this process, the Neumann method provides a matrix decomposition approach for the pre-conditioner, and then it uses CG method to do iterations as before. Furthermore, there are other aspects for us to explore to reduce the complexity and get the better performance like BER and LLRs, as a result to improve the communication system. And also these proposed methods can be applied to other systems which involve the matrix inversion in the further development.

References
[1] Xue Y, Zhang C, et al 2017 A fast-convergent pre-conditioned conjugate gradient detection for massive MIMO uplink IEEE International Conference on Digital Signal Processing (Beijing) IEEE p 331-335
[2] Dai L, Gao X, Su X, et al. 2015 Low-Complexity Soft-Output Signal Detection Based on Gauss–Seidel Method for Uplink Multiuser Large-Scale MIMO Systems IEEE Transactions on Vehicular Technology 64(10) 4839-4845
[3] Ning J, Lu Z, Xie T, et al. 2016 Low complexity signal detector based on SSOR method for massive MIMO systems International Conference on Communicatinis and Networking in China (Ghent) IEEE p 193-202
[4] Zhu D, Li B, Liang P 2015 On the matrix inversion approximation based on Neumann series in massive MIMO systems IEEE International Conference on Communications (London) IEEE p 1763-1769
[5] Jin J, Xue Y, Ueng Y L, et al 2017 A split pre-conditioned conjugate gradient method for massive MIMO detection IEEE International Workshop on Signal Processing Systems (Lorient) IEEE p 1-6
[6] Jiang F, Li C, Gong Z 2017 A low complexity soft-output data detection scheme based on Jacobi method for massive MIMO uplink transmission IEEE International Conference on Communications (Paris) IEEE p 1-5
[7] Vardhan K V, Mohammed S K, Chockalingam A, et al 2008 A low-complexity detector for large MIMO systems and multicarrier CDMA systems IEEE Journal on Selected Areas in Communications, 26(3) 473-485
[8] Zhao S, Shen B, Yang F 2017 Low complexity hybrid iterative algorithm based signal detection in massive MIMO system Telecommunications Science
[9] Shu J 1998 Effectiveness comparison for preconditioned conjugate gradient algorithm solving reservoir simulerical problem Numerical Mathematics A Journal of Chinese Universities 25(4) 95—115