SD-Based Low-Complexity Signal Detection Algorithm in Massive MIMO

Zhang Lihuan[1](✉) and Jiang Xiaolin[1,2]

[1] Heilongjiang University of Science and Technology, Harbin 150000, China
2655517184@qq.com
[2] Jinhua Advanced Research Institute, Jinhua 321000, China

Abstract. The steepest descent algorithm (SD) itself can find a better convergence direction, but its own convergence speed is relatively slow, resulting in multiple iterations to approach the true solution. This paper proposes an improvement method for this problem. The principle is to improve the approximate solution obtained after every time the steepest descent algorithm is performed, so as to change the iterative formula and speed up the convergence speed of the algorithm, and then divide the constellation into four regions based on the idea of region division. The points are extracted directly for judgment and no longer participate in subsequent iterations. It is proved by simulation that when the number of iterations of the improved algorithm is consistent with the number of iterations of the steepest descent method, the improved algorithm has an order of magnitude higher detection performance than the original SD algorithm. The improved SD iterative algorithm has 2 iterations and a signal-to-noise ratio greater than 4 dB. The bit error rate is lower than that of the Minimum Mean Square Error (MMSE) detection algorithm, and its own complexity is only 67.4% of that of the SD iterative algorithm 2 iterations. The greater the difference between the number of user antennas and the number of base station antennas, the complexity of the improved algorithm can be even greater low.

Keywords: Steepest descent method · Approximate solution · Minimum Mean Square Error · Number of iterations · Complexity

1 Introduction

With the increase of mobile smart terminals, the traditional MIMO system has been unable to meet the user's demand for data services [1–3], thus a massive MIMO system is proposed. The massive MIMO system improves the channel capacity and spectrum utilization of the entire communication system by configuring a large number of antennas in the base station to serve multiple users. In the traditional algorithm, the Maximum Likelihood (ML) [4] detection algorithm finds the optimal estimation value by traversing all the constellation points, so the bit error rate is the lowest but its complexity is exponentially related to the modulation order and the number of antennas. Although the Sphere Decoding algorithm [5] can reduce the complexity of the algorithm by reducing

X. Jiang (Ed.): MLICOM 2021, LNICST 438, pp. 214–225, 2022.
https://doi.org/10.1007/978-3-031-04409-0_20

the number of traversed constellation points, the detection performance is close to the detection performance of the ML algorithm, it is still relatively complex compared to other algorithms, while the Zero Forcing (ZF) detection algorithm and The Minimum Mean Square Error (MMSE) [6] detection algorithm as a linear detection algorithm involves matrix inversion. Due to the rapid increase in the number of antennas in a massive MIMO system, the four detection algorithms mentioned are too high complex to be implemented in a massive MIMO system, so searching for low-complexity signal detection algorithms has become a research hotspot in massive MIMO wireless communication technology.

As the number of antennas increases, the channel matrix in the wireless communication system will have channel hardening characteristics, and each sub-channel will gradually be orthogonal [7]. Therefore the MMSE detection algorithm has better detection performance, but the algorithm involves matrix inversion operation. The complexity is too high and is not suitable for using in massive MIMO systems, so it is generally used as a comparison standard for detection performance. In order to reduce the complexity of the algorithm, many scholars have proposed many improved algorithms based on MMSE. At present, the more frequently used MMSE optimization algorithms can be divided into two categories, they are iterative method and approximation method.

The more common iterative methods are: Jacobi iteration method, Gauss-Seidel iteration method [8], Steepest descent method, Relaxation iteration method [9], Conjugate gradient algorithm [10], etc. The iteration method was previously used for linear equations. The principle of solving is to continuously loop through the iterative formula, so that the value obtained in each iteration continuously approaches the real solution, it can avoid matrix inversion. The approximation method is mainly the Neuman Series expansion method [11]. The approximation method generally uses the expanded polynomial to approximate the matrix inversion, but the Neuman Series also has larger defects. When the number of expansion items is 2 or less, although the computational complexity of the algorithm is lower than the complexity of calculating matrix inversion, the bit error rate is higher. When the number of expansion items is greater than 2, the detection performance is improved but the algorithm complexity is increased. Therefore the Neuman Series expansion method generally takes the expansion term as 1 as the approximate matrix inverse, and the solution obtained in this way is used as the initial value of the iterative method to accelerate the iteration.

In response to the above problems, this paper proposes an improved algorithm based on the steepest descent method. This algorithm uses the solution of linear equations, the final exact solution will be equal to the sum of the rough solution and the residual solution, and the idea of region division is used further improvement, If a certain component of the estimated value x obtained after iteration falls into the reliable region, it is directly extracted for judgment and no longer participates in subsequent iterations, so that the matrix dimension of the next iteration will be reduced, and the complexity will also be reduced. Simulation proves that when the number of iterations of the improved algorithm is lower than that of the original algorithm, the detection performance is very close. When the number of iterations of the improved algorithm is consistent with the number of iterations of the original algorithm, the detection performance of the improved algorithm is better than that of the original algorithm.

2 System Model

The object of this paper is the uplink of a massive MIMO system. The system model can be expressed as:

$$y = Hx + n = h_1x_1 + h_2x_2 + \cdots + h_Kx_K + n \tag{1}$$

The base station in this system is configured with N antennas, and K single antenna users. Since K << N in the massive MIMO system, the sub-channels also tend to be orthogonal. The user sending vector and the base station receiving vector are arrespectively: $x = [x_1, x_2, \cdots, x_K]^T y = [y_1, y_2, \cdots, y_N]^T$, Where x_j represents the signal transmitted by the j-th user, y_i and represents the signal received by the i-th antenna. $n = [n_1, n_1, \cdots, n_N]^T$ Represents the Gaussian white noise that obeys the Gaussian distribution, and the Mathematical Expectation is 0.variance isσ^2.

Where H represents the channel matrix with dimension N * K, h_j represents the j-th column of the channel matrix.

The ZF detection algorithm eliminates interference between users through a weighted matrix. The weighting matrix is the inverse of the signal matrix and can be expressed as:

$$W_{ZF} = (H^H H)^{-1} H^H \tag{2}$$

The user transmitted vector estimated from the received signal vector can be expressed as:

$$x_{ZF} = W_{ZF} y = x + n_{ZF} \tag{3}$$

Since the ZF detection algorithm does not consider the influence of noise, its detection performance has a great relationship with the power of n_{ZF}. In the case of low signal-to-noise ratio, the error detection rate of the ZF detection algorithm is not very ideal. In order to reduce noise interference, the MMSE detection algorithm makes noise compensation on the basis of ZF. The weighting matrix of MMSE is:

$$W_{MMSE} = (H^H H + \sigma^2 I)^{-1} H^H \tag{4}$$

I is the identity matrix, and the estimated user transmission vector can be expressed as:

$$x_{MMSE} = W_{MMSE} y \tag{5}$$

Since the MMSE detection algorithm has made up for noise in the weighting matrix, the performance of the MMSE detection algorithm is better than that of the ZF detection algorithm. we can convert the MMSE detection algorithm into the form of solving linear equations, it can be expressed as:

$$Ax = b \tag{6}$$

Among them $A = H^H H + \sigma^2 I$, $b = H^H y$.

3 Improvement of SD Iterative Algorithm

3.1 Application of SD Algorithm

The iterative algorithm is to make the estimated value gradually close to the signal vector sented by the user through multiple iterations, so as to avoid the inversion of the matrix and reduce the complexity. Compared with other iterative algorithms, SD iterative algorithm can find a better iterative direction, but its own convergence speed is poor.

Theorem 1: Assuming that the matrix A is a symmetric matrix, the quadratic function f(x) = (Ax, x) − 2(b, x), if and only if x is the solution of the equation system Ax = b, point x makes the quadratic function take the minimum value.

Theorem 2: Assuming that the matrix A is a symmetric matrix, and the quadratic functional method f(x) = (Ax, x) − 2(b, x), if and only if x is the solution of the equations, x is the center of the ellipsoid f(x) = c.

Since C is symmetric and positive definite, according to Theorem 1 and Theorem 2, the solution of the equation system Ax = b is the center of f(x) = c of the ellipsoid, which also makes the quadratic function f(x) = (Ax, x) − 2(b, x) obtain the minimum value.

Take a random point x0 as the initial value, x0 must be on the spherical surface of the ellipsoid f(x) = (Ax0, x0) − 2(b, x0), passing through the point x0 along the fastest descending direction r0 to make a straight line x = x0 + t ∗ r0, and then look for the point x1 on this straight line that can make f(x) the minimum value. Assuming that the function f(x) can get the minimum value when t = a, Then there is x = x0 + a ∗ r0, so only a need be required., which becomes the problem of solving the minimum value of the function f(x0 + t ∗ r0) on the variable t. Derivation the function f(x0 + tr0) can be expressed as:

$$\frac{df(x_0 + tr_0)}{dt} = 2t(Ar_0, r_0) - 2(r_0, r_0) = 0 \tag{7}$$

$$\mathrm{t} = \frac{(r_0, r_0)}{(Ar_0, r_0)} = a \tag{8}$$

So we can get the next approximate solution

$$r_0 = b - Ax_0 \tag{9}$$

$$a_0 = \frac{(r_0, r_0)}{(Ar_0, r_0)} \tag{10}$$

$$x_1 = x_0 + a_0 r_0 \tag{11}$$

3.2 Neuman Series Expansion Algorithm

In order to avoid solving the inverse of matrix A directly, Neuman series expansion is used to approximate the inverse of A, when X and A are close and satisfy the condition:

$$\lim_{n \to \infty} (I - AX)^n = 0 \tag{12}$$

Then the inverse of matrix A can be expressed as:

$$A^{-1} = \sum_{n=0}^{\infty} X\,(X^{-1} - A)^{n} X \tag{13}$$

Decompose matrix A into A = D + E, where D is the matrix formed by the main diagonal of matrix A. Taking the first t term instead of the inverse of matrix A

$$A^{-1} = \sum_{n=0}^{t-1} (-D^{-1}\mathrm{E})^{n} D^{-1} \tag{14}$$

3.3 Improvements to the Iterative Structure of the SD Algorithm

Assuming that x1 is an approximate solution of the equation Ax = b, then r1 = b – A * x1 represents the residual vector. Solving the equation A ∗ d1 = r1 obtains the residual solution as d1 = $A^{-1}r1$, x2 = x1 + d1 then x2 is the equation Ax = b exact solution. Because the complexity of matrix inversion is too high when we solve d1, Neuman series expansion can be substituted matrix inversion to reduce the complexity. Although the exact value cannot be obtained directly, x2 is more accurate than x1.

$$x_1 = x_0 + a_0 r_0 \tag{15}$$

Then improve the approximate value obtained when the number of Neuman series expansion items is 1:

$$r_1 = b - Ax_1 = b - A(x_0 + a_0 r_0) = r_0 - a_0 A r_0 \tag{16}$$

$$x_2 = x_1 + D^{(-1)} r_1 = x_1 + D^{(-1)}(r_0 - a_0 A r_0) \tag{17}$$

x2 as the initial value of the next iteration,

Improvement when the number of Neuman series expansion items is 2:

$$x_2 = x_1 + A_2^{(-1)} r_1 = x_1 + A_2^{(-1)}(r_0 - a_0 A r_0) \tag{18}$$

where A_2 represents the approximate inverse of the matrix when the number of Neuman series expansion items is 2. It is also possible to use when the expansion series is 3, but when the expansion series is 3, the complexity is too high and not applicable.

3.4 Further Improvement Based on the Idea of Area Division

In the actual iteration process, the speed at which the estimated value of each user approaches the true value is inconsistent. The estimated value of some users is already very close to the true value. The estimated value continued iteration will not change much, and these values can be directly judged. The estimated value of some users is still far from the true value, for the overall accuracy, the estimated value of the user who is already close to the true solution must continue to iterate with the estimated value of other users, which causes unnecessary calculations. In order to reduce the complexity

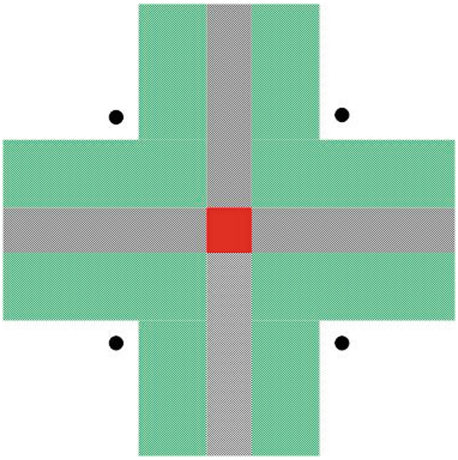

Fig. 1. Schematic diagram of 4QAM modulation area division

and detection performance of the algorithm, this paper divides the constellation map area into four different areas: reliable area, normal iteration area, mildly unreliable area, and unreliable area.

Figure 1 is a schematic diagram of the 4QAM modulation area division. The black dots represent the constellation points, the white area represents the reliable area, the cyan area represents the normal iteration area, the shaded area represents the mildly unreliable area, and the red area represents the unreliable area. The user's estimated value falling in the reliable area means that it does not need to iterate, these values can be directly judged. The user's estimated value falls in the normal iterative area, we don't do any disposal and let it continue to iterate. The user's estimated value falls in a mildly unreliable area, indicating that the estimated value is not much different from the two surrounding constellation points, which means that these two values may be the true value of the user's send vector. In order to reduce the possibility of misjudgment, the point with the smallest cost function is selected as the estimated value by traversing these two points. The user's estimate falls in the unreliable area, indicating that all four surrounding points may be true solutions, so all four surrounding points are traversed, and the point with the smallest cost function is selected as the estimated value.

The two thresholds used in this paper to distinguish each region are 0.04 and 0.198 respectively. The red area indicates that the real and imaginary parts of the estimated value are both less than 0.04, the cyan area indicates that the real or imaginary part is less than 0.04, the white area indicates that both the real and imaginary parts are greater than 0.19, and the rest are shaded parts.

After each iteration of the SD-NM algorithm, the estimated value obtained is divided into regions. If the estimated value falls in the reliable area, the value can be directly judged, and the corresponding rows and columns are deleted. If the estimated value falls in the continuous iteration area, no processing is done and the iteration continues. When the iteration is completed, if the estimated value falls in the shadow area, the two constellation points next to the shadow area are traversed, and if the estimated value falls in the red area, the four constellation points around the red area are all traversed.

In the iterative process, if the estimated value falls in the shaded area or the red area, no processing is performed. Assuming that the initial value is close to 1-i and the true solution is 1 + i, the estimated value obtained after one iteration may falls in the shaded area or red area, the estimated value after one iteration continue iterate, the estimated value may be in the reliable area. In order to avoid this situation, the points that fall in the shadow area or the red area after each iteration are not processed. Only after the iteration is completed, if the estimated value falls in the shaded area or the red area, it will be processed.

3.5 Maximum Likelihood Criterion

Assuming that estimated value of the j-th user falls in the shaded area, it means that excepting the point corresponding to the hard decision, an additional point must be traversed. Due to Hx $= h_1x_1 + h_2x_2 + \cdots + h_kx_k$, the extra traversed points are the same as the points corresponding to the hard decision, except for x_j, the other values are the same. In order to reduce the amount of calculation, the minimum cost function can be transformed as:

$$x_{ML} = argmin\|y - Hx\|^2 = argmin\|y - \sum_{i=1, i\neq j}^{k} h_ix_i - h_jx_j\|^2 \quad (19)$$

$$x_{ML} = argmin\|y - h_jx_j\|^2 \quad (20)$$

In formula (19), the $y - \sum_{i=1, i\neq j}^{k} h_ix_i$ is the same, the only difference is h_jx_j, this means that formula (19) is equivalent to formula (20).

Among them, using formula (20) to calculate the maximum likelihood cost function can greatly reduce the complexity of the algorithm. The formula (19) must first calculate Hx, and then calculate the inner product. The calculation of Hx can be regards as a 2k * 2k matrix multiplying a 2k * 1 vector. The complexity of multiplying the two is $4k^2$. The calculation of the inner product can be regarded as the multiplication of two 2k * 1 vectors. The complexity is 2k. Using Using formula (20) replace formula (19), the complexity of calculation the maximum likelihood cost function is only 4k. Therefore, every time a constellation point is traversed, the added complexity is 4k.

4 Simulation Results and Analysis

This section gives the MATLAB simulation result graph, and analyzes and explains the simulation result graph. In the simulation, the number of users is 16, the number of base station antennas is 64, and the size of channel matrix is 64 * 16. If there is no additional statement in the subsequent analysis, the default channel size is 64 * 16. The t in the simulation diagram represents the number of iterations of the detection algorithm, and n represents the number of items expanded by the Neuman series (Fig. 2).

The performance of the MMSE detection algorithm in the traditional algorithm is better, but it is only used as a standard to measure the detection performance of other

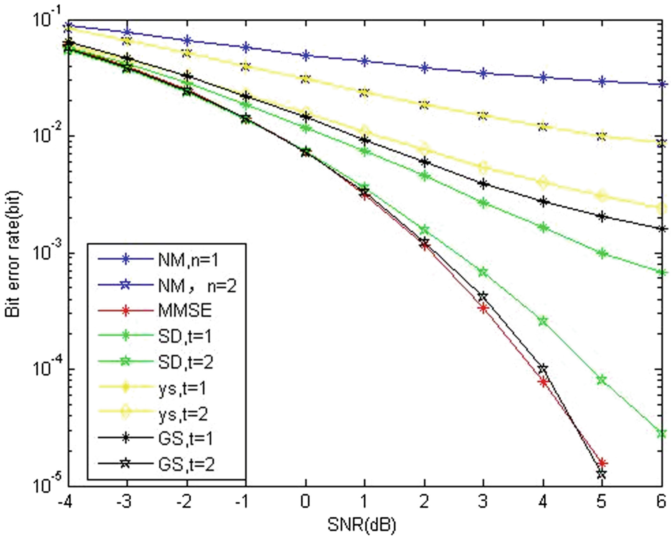

Fig. 2. Bit error rate of traditional detection algorithm

iterative algorithms. When the number of iterations of the Gauss Seidel iteration method is 2, the detectability can approach the performance of the MMSE detection algorithm. The detection performance of the SD algorithm when the number of iterations is 1 is better than the detection performance of the Gauss Seidel iteration method when the number of iterations is 1, Due to the problem of convergence speed, the detection performance of the SD algorithm is poor when the number of iterations of both is 2. The detection performance is not very good when the number of expansion items of the Neuman series expansion method is 1 and 2. When the signal-to-noise ratio is 6 dB, the bit error rate is still 10^{-2}. When the expansion item is 3, the complexity of Neuman series is close to complexity of the MMSE algorithm, so the Neuman series expansion method still has big flaws.

Figure 3 shows that SD is the iteration of the steepest descent method, and SD-NM is an improved algorithm for improving the approximate solution obtained in each iteration of SD, where t represents the number of iterations, and n represents the number of items expanded by the Neuman series. The simulation results show that the SD-NM iterative algorithm has detection performance is close to the detection performance of the traditional SD iterative algorithm when the number of iterations of SD-NM is 1, and the number of items expanded by the Neuman series is 1, the number of iterations of SD iterative algorithm is 2. This shows that improving the approximate solution obtained after each iteration of SD can speed up the convergence speed of the original algorithm. The signal-to-noise ratio is 5 dB, the bit error rate of SD-NM iterative algorithm at $t = 1$ and $n = 2$ is $8.9 * 10^{-4}$, and the bit error rate is $1.25 * 10^{-4}$ when $t = 1$ and $n = 1$, and when $t = 2$ and $n = 1$ bit error rate is $1.25 * 10^{-5}$.Therefore, when the difference in the number of receiving and transmitting antennas is more small, the number of Neuman series expansion items should be selected as 2, so that the iterative algorithm can converge faster.

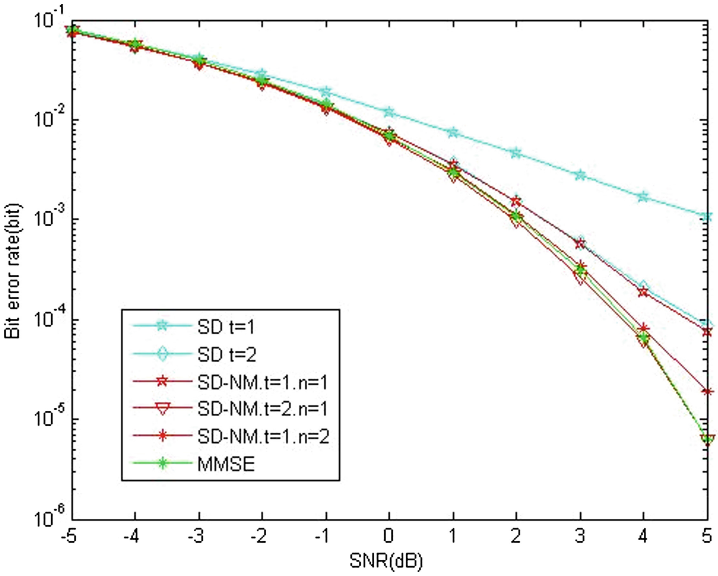

Fig. 3. Comparison of bit error rate between SD-NM algorithm and original algorithm

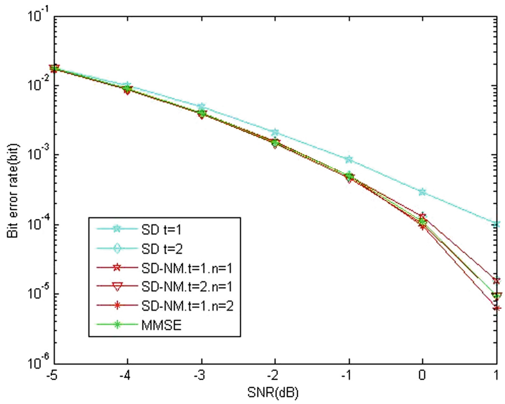

Fig. 4. The influence of the number of antennas on the detection performance of SD-NM algorithm

Figure 4 shows the comparison of the detection performance of each algorithm when the number of receiving and transmitting antennas is 128 * 16. As shown in Fig. 4, the detection performance of t = 1, n = 1 and t = 1, n = 2 in the SD-NM algorithm are consistent with the detection performance of the MMSE algorithm, but when the number of items expanded by the Neuman series is 2, the algorithm complexity is higher. Therefore, when the number of receiving and transmitting antennas differ greatly, the number of items in the Neuman series expansion should be selected as 1.

SD-NM-RC in Fig. 5 represents an algorithm that uses the idea of area division to further improve SD-NM. When the signal-to-noise ratio is lower than 2 dB, the

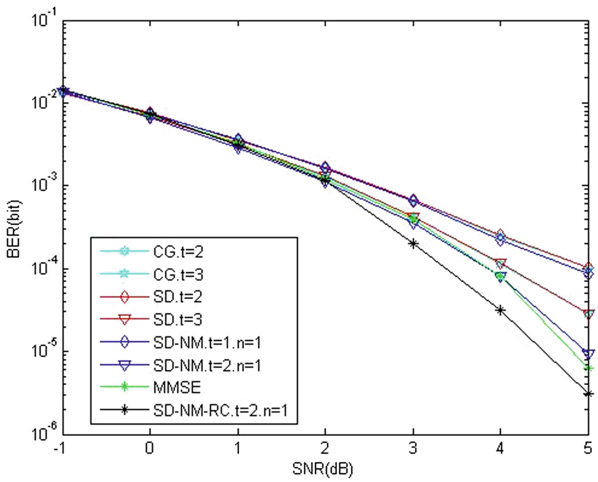

Fig. 5. Comparison of bit error rate between SD-NM-RC algorithm and other algorithms

detection performance of each algorithm is not much different. Because the SD-NM-RC algorithm uses the ML criterion for the unreliable area to traverse the surrounding points, the detection performance of the SD-NM-RC detection algorithm is better. when the signal-to-noise is greater than 3 dB, the bit error rate of SD-NM-RC at t = 2 and n = 1 is lower than that of the MMSE detection algorithm and SD-NM algorithm (Fig. 6).

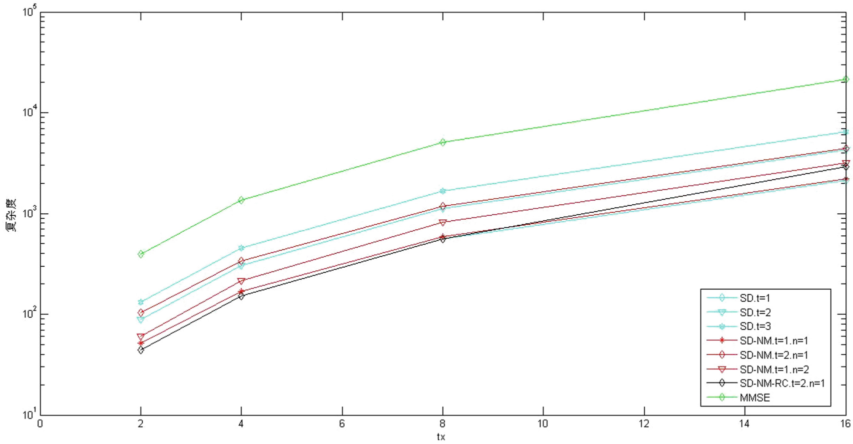

Fig. 6. Comparison of the complexity of each algorithm

Since the signal-to-noise ratio has an impact on the complexity of the SD-NM-RC algorithm, but has no effect on the complexity of other algorithms, in this simulation, the signal-to-noise ratio is set to 5 dB and the number of receiving antennas is fixed to 64. Change the number of transmitting antennas to compare the complexity of each algorithm. Because the complexity of each iteration of the SD-NM algorithm when n = 1 is only higher than the complexity of one iteration of the SD algorithm. When the SD-NM algorithm t = 2 and n = 1, the complexity of the two iterations of the SD algorithm is almost the same. When the SD-NM-RC algorithm has a large difference in the number of receiving antennas, the estimated value area is divided after one iteration. Almost all values fall in the reliable area, so there will be no second iteration or x need to be iterated. The portion is extremely small. When the number of receiving antennas is very close, the channel hardening characteristic is not obvious, and the convergence speed is slow. Therefore, the points falling in the reliable area will be reduced, and the points falling in the unreliable area will increase, which leads to the SD-NM-RC algorithm The complexity increases, but when the number of receiving antennas is 64 * 16, the complexity of the two iterations of the SD-NM-RC algorithm is only 67.4% of that of the two iterations of the SD algorithm.

5 Concluding Remarks

Due to the large number of antennas in the Massive MIMO system, the dimensionality of the channel matrix is high, while the traditional ZF detection algorithm and the MMSE detection algorithm are more complex. Aiming at this problem, this paper proposes an improved algorithm based on the steepest descent method. The algorithm can consider estimating the vector sent by the user as solving a linear equation set and solving it in an iterative manner, which avoids the inversion of a high-dimensional matrix and reduces the complexity. The improved algorithm is to change the iterative formula to speed up the iterative process by improving the approximate solution obtained after each iteration of the SD iterative method, and use area division to further improve. Because the things that need to be calculated in the process of improving the approximate solution obtained by SD have already been calculated when the SD algorithm is carried out, too much additional complexity will not be added, and the idea of region division will be used to further improve the algorithm complexity. Lower, every point falls in the reliable area, which will make the matrix dimension will be reduced by 1 in the subsequent iterations. On the whole, the improved algorithm estimates the user's transmission vector with lower complexity. The improved algorithm also has certain shortcomings. The threshold setting for area division plays a decisive role in the detection performance and complexity of the entire algorithm. At present, an accurate formula for calculating the threshold has not been found.

References

1. Lu, L., Li, G., Swindlehurst, A., Ashikhmin, A., Zhang, R.: An overview of massive MIMO: benefits and challenges. IEEE J. **8**, 742–758 (2014)
2. Wu, M., Yin, B., Wang, G., Dick, C., Cavallaro, J., Studer, C.: Large-scale MIMO detection for 3GPP LTE: algorithms and FPGA implementation. IEEE J. **8**, 916–929 (2014)
3. Nakamura, T., et al.: Trends in small cell enhancements in LTE advanced. IEEE Commun. Mag. **51**(2), 98–105 (2013)
4. Noh, H., Myoungseok, K., Jaesang, H., Chungyong, L.: A practical MMSE-ML detector for a MIMO SC-FDMA system. IEEE Commun. Lett. **13**, 902–904 (2009)

5. Nguyen, V.-H., Berder, O., Scalart, P.: On the efficiency of sphere decoding for linearly MIMO systems. In: Wireless Communications and Networking Conference (WCNC), pp. 4021–4025. IEEE, Shanghai (2013)
6. Tran, X.N., Le, A.T., Fujino, T.: Performance comparison of MMSE-SIC and MMSE-ML multiuser detectors in a STBC-OFNM system. In: Proceedings of the IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1050–1054. Berlin, Germany, 11–14 Sept 2005
7. Narasimhan, T.L.: Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems. IEEE J. Select. Top. Sign. Process. **8**(5), 847–860 (2014)
8. Zhou, J., Ye, Y., Hu, J.: Biased MMSE soft-output detection based on Jacobi method in massive MIMO. In: 2014 IEEE International Conference on Communication Problem-Solving, pp. 442–445. IEEE (2014)
9. Wang, F., Zhang, C., Liang, X., et al.: Efficient iterative soft detection based on polynomial approximation for massive MIMO. In: 2015 International Conference on Wireless Communications and Signal Processing (WCSP). IEEE (2015)
10. Zhou, J., Hu, J., Chen, J., et al.: Biased MMSE soft-output detection based on conjugate gradient in massive MIMO. In: 2015 IEEE 11th International Conference on ASIC (ASICON), pp. 1–4. IEEE (2015)
11. Rosario, F., Monteiro, F.A., Rodrigues, A.: Fast matrix inversion updates for massive MIMO detection and precoding. IEEE Sign. Process. Lett. **23**(1), 75–79 (2016)