Research Article

Joint Newton Iteration and Neumann Series Method of Convergence-Accelerating Matrix Inversion Approximation in Linear Precoding for Massive MIMO Systems

Lin Shao and Yunxiao Zu

School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Lin Shao; linshao2001@163.com

Due to large numbers of antennas and users, matrix inversion is complicated in linear precoding techniques for massive MIMO systems. Several approximated matrix inversion methods, including the Neumann series, have been proposed to reduce the complexity. However, the Neumann series does not converge fast enough. In this paper, to speed up convergence, a new joint Newton iteration and Neumann series method is proposed, with the first iteration result of Newton iteration method being employed to reconstruct the Neumann series. Then, a high probability convergence condition is established, which can offer useful guidelines for practical massive MIMO systems. Finally, simulation examples are given to demonstrate that the new joint Newton iteration and Neumann series method has a faster convergence rate compared to the previous Neumann series, with almost no increase in complexity when the iteration number is greater than or equal to 2.

1. Introduction

Massive multiple-input multiple-output (MIMO) [1–3] is one of the promising technologies for the fifth-generation communication system. Compared to traditional MIMO, hundreds of antennas are equipped in base stations (BSs) in massive MIMO systems in order to achieve orders of magnitude increases in spectral and energy efficiency [4–6]. Unfortunately, the application of the massive MIMO system still faces several challenging problems in practice. For example, linear precoding techniques, such as zero-forcing (ZF) precoding, are always involved with complicated matrix inversion due to large numbers of BS antennas and users.

In order to reduce the complexity, several approximated approaches have been recently proposed to avoid the exact matrix inversion. A low-complexity Richardson method has been proposed in [7], but it has an uncertain parameter which affects the convergence of the method. Conjugate gradient method was applied to reduce the complexity of data detection and precoding in the massive MIMO system with realistic antenna configurations [8]. However, many divisions are involved in the approach. Gauss-Seidel (GS) method was

employed in [9] to speed up convergence. Unfortunately, the matrix inversion is not obtained directly, which will add extra complexity to the calculation of the transmitted signal after precoding. Newton iteration method converges fast and the complexity can be controlled just by the number of iterations [10]. However, a complex calculation is always required to get an initial input to ensure convergence [10]. Truncated Neumann series in [11, 12] was proposed to obtain near-optimal performance. It contains matrix multiplication and matrix addition, which are preferable in hardware implementation. Nevertheless, how to speed up the convergence rate of the Neumann series is still a problem.

In this paper, we propose a new joint Newton iteration and Neumann series method, where Newton iteration method is utilized to provide an efficient searching direction for the Neumann series. Specifically, the first iteration result of Newton iteration method is employed to reconstruct the Neumann series expansion to accelerate convergence. Furthermore, a high probability convergence condition is derived to guarantee the convergence of the new approach. This condition is expected to contribute to the massive MIMO system in practice.

The remainder of the paper is organized as follows. First, Section 2 introduces the background including the system model, the Neumann series, and Newton iteration method. Then, the new joint Newton iteration and Neumann series method is proposed in Section 3. Simulation results are in Section 4. Finally, conclusions are drawn in Section 5.

2. Background

In this section, the system model, the Neumann series, and Newton iteration method are introduced.

2.1. System Model. Consider a downlink massive MIMO system with N antennas at the BS and $K \ll N$ single antenna users. The received $K \times 1$ vector $\mathbf{y}$ can be expressed as

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where ρ is the signal-to-noise ratio (SNR) in the downlink, $\mathbf{H} \in \mathbb{C}^{K \times N}$ denotes the Rayleigh fading channel matrix whose entries are independent and identically distributed (i.i.d.) zero-mean unit-variance complex Gaussian variables, $\mathbf{n} \in \mathbb{C}^{K \times 1}$ is the additive white Gaussian noise vector, and $\mathbf{x} \in \mathbb{C}^{N \times 1}$ presents the transmitted signal vector after precoding and is obtained by

$$\mathbf{x} = \mathbf{P}\mathbf{s}, \tag{2}$$

where $\mathbf{s}$ is $K \times 1$ vector of QAM symbols for transmission [13] and $\mathbf{P} \in \mathbb{C}^{N \times K}$ denotes the precoding matrix which specifically for ZF precoding can be represented as

$$\mathbf{P} = \beta\mathbf{H}^{\dagger}, \tag{3}$$

where

$$\mathbf{H}^{\dagger} = \mathbf{H}^{H}\left(\mathbf{H}\mathbf{H}^{H}\right)^{-1} \triangleq \mathbf{H}^{H}\mathbf{W}^{-1} \tag{4}$$

denotes the pseudoinverse of $\mathbf{H}$ and β is a scalar which is always chosen as

$$\beta = \sqrt{\frac{K}{\text{tr}\left(\mathbf{W}^{-1}\right)}}, \tag{5}$$

where $\text{tr}(\mathbf{W}^{-1})$ denotes the trace of $\mathbf{W}^{-1}$. The main computation complexity for ZF precoding lies in the inversion of $K \times K$ matrix $\mathbf{W}$, so several approximated approaches including the Neumann series have been investigated recently, which are shown in Section 1.

2.2. Neumann Series. Based on the Neumann series, the matrix inversion of $K \times K$ matrix $\mathbf{W}$ is transformed to the sum of matrix polynomials as

$$\mathbf{W}^{-1} = \sum_{n=0}^{+\infty} \left(\mathbf{I}_K - \mathbf{\Theta}\mathbf{W}\right)^n \mathbf{\Theta}, \tag{6}$$

where $\mathbf{\Theta}$ is $K \times K$ matrix. Note that, according to [14], the premise condition of (6) is

$$\lim_{n \to \infty} \left(\mathbf{I}_K - \mathbf{\Theta}\mathbf{W}\right)^n = \mathbf{0}_K. \tag{7}$$

For practical use, $\mathbf{W}^{-1}$ is approximated as [14]

$$\mathbf{W}^{-1} \approx \sum_{n=0}^{L} \left(\mathbf{I}_K - \mathbf{\Theta}\mathbf{W}\right)^n \mathbf{\Theta}, \tag{8}$$

where L is the iteration number.

Remark 1. Based on (2), (3), and (4), $\mathbf{x}$ is obtained as

$$\mathbf{x} = \beta\mathbf{H}^{H}\mathbf{W}^{-1}\mathbf{s}. \tag{9}$$

Let $\mathbf{f} \triangleq \mathbf{W}^{-1}\mathbf{s}$ and $\mathbf{f}$ can be regarded as the solution of the linear equation

$$\mathbf{W}\mathbf{f} = \mathbf{s}. \tag{10}$$

Recently, Richardson method [7], conjugate gradient method [8], and GS method [9] are all applied to solve (10) in an iterative way. But they approximate $\mathbf{W}^{-1}\mathbf{s}$ instead of $\mathbf{W}^{-1}$, while the Neumann series approximates $\mathbf{W}^{-1}$ directly.

2.3. Newton Iteration Method. Newton iteration method can be employed to estimate $\mathbf{W}^{-1}$ in an iterative way. Suppose $\mathbf{Z}_0$ is the original estimation of $\mathbf{W}^{-1}$ and the following condition is satisfied:

$$\left\|\mathbf{I}_K - \mathbf{W}\mathbf{Z}_0\right\| < 1. \tag{11}$$

Then, the nth iteration estimation of $\mathbf{W}^{-1}$ is expressed as

$$\mathbf{Z}_n = \mathbf{Z}_{n-1}\left(2\mathbf{I}_K - \mathbf{W}\mathbf{Z}_{n-1}\right). \tag{12}$$

Newton iteration method converges fast, but formula (11) usually needs a complex calculation to get an appropriate $\mathbf{Z}_0$ [10].

3. Joint Newton Iteration and Neumann Series Method

It can be seen from (7) that, as the first item of the series, $\mathbf{\Theta}$ greatly affects convergence. How to choose $\mathbf{\Theta}$ plays a key role in the Neumann series. In [13], $\mathbf{\Theta}$ is initially set as $(1/(N + K))\mathbf{I}_K$ and the Neumann series converges when N and K grow to infinity. In [11], $\mathbf{\Theta}$ is then chosen as the matrix inversion of $K \times K$ diagonal matrix $\mathbf{D}$ whose entries are the main diagonal elements of $\mathbf{W}$, resulting in a faster convergence rate. In this paper, Newton iteration method is employed to provide an appropriate $\mathbf{\Theta}$ to speed up the convergence of the Neumann series. Moreover, a high probability convergence condition about N/K ratio is derived for the joint Newton iteration and Neumann series method. Finally, the complexity of the new approach is analyzed.

3.1. Joint Newton Iteration and Neumann Series Method and Its Convergence Condition. The fast convergence property of Newton iteration method inspired us to use $\mathbf{Z}_1$ instead of $\mathbf{D}^{-1}$ as $\mathbf{\Theta}$ to speed up the convergence rate of the Neumann series. Note that formula (11) must be satisfied before the application of Newton iteration method. According to [10], setting $\mathbf{Z}_0$ to

be $\mathbf{D}^{-1}$ makes it easy to meet (11) for massive MIMO systems. Therefore, we get the joint Newton iteration and Neumann series method as follows.

First, let $\mathbf{Z}_0 = \mathbf{D}^{-1}$ and then

$$\mathbf{Z}_1 = \mathbf{Z}_0\left(2\mathbf{I}_K - \mathbf{W}\mathbf{Z}_0\right) = \mathbf{D}^{-1}\left(2\mathbf{I}_K - \mathbf{W}\mathbf{D}^{-1}\right). \quad (13)$$

Second, use $\mathbf{Z}_1$ to reconstruct the Neumann series as follows:

$$\mathbf{W}^{-1} \approx \sum_{n=0}^{L}\left(\mathbf{I}_K - \mathbf{Z}_1\mathbf{W}\right)^n \mathbf{Z}_1. \quad (14)$$

It is worth pointing out that, in order to guarantee the convergence of the Neumann series with $\mathbf{\Theta} = \mathbf{Z}_1$, the accurate lower bound of N/K ratio needs to be determined. For this end, a high probability convergence condition about N/K ratio is given in the following lemma.

Lemma 2. *For downlink massive MIMO systems, the Neumann series with $\mathbf{\Theta} = \mathbf{Z}_1$ converges with high probability when*

$$\alpha \triangleq \frac{N}{K} > \frac{1}{\left(\sqrt{2}-1\right)^2}. \quad (15)$$

Proof. $\sum_{n=0}^{\infty}(\mathbf{I}_K - \mathbf{Z}_1\mathbf{W})^n\mathbf{Z}_1$ converges:

⇔ $\lim_{n\to\infty}(\mathbf{I}_K - \mathbf{Z}_1\mathbf{W})^n = \mathbf{0}_K$,

⇔ $\rho(\mathbf{I}_K - \mathbf{Z}_1\mathbf{W}) < 1$, where $\rho(\mathbf{I}_K - \mathbf{Z}_1\mathbf{W})$ denotes the spectral radius of $\mathbf{I}_K - \mathbf{Z}_1\mathbf{W}$,

⇔ $|\lambda(\mathbf{A})| < 1$, where $\lambda(\mathbf{A})$ is any eigenvalue of $\mathbf{A}$ and $\mathbf{A} = \mathbf{I}_K - \mathbf{Z}_1\mathbf{W}$.

Note that

$$\mathbf{A} = \mathbf{I}_K - \mathbf{Z}_1\mathbf{W} = \mathbf{I}_K - \mathbf{D}^{-1}\left(2\mathbf{I}_K - \mathbf{W}\mathbf{D}^{-1}\right)\mathbf{W}$$
$$= \mathbf{I}_K - 2\mathbf{D}^{-1}\mathbf{W} + \left(\mathbf{D}^{-1}\mathbf{W}\right)^2 = \left(\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W}\right)^2 \quad (16)$$

and $\sum_{n=0}^{\infty}(\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W})^n\mathbf{D}^{-1}$ converges ⇔ $|\lambda(\mathbf{B})| < 1$, where $\mathbf{B} = \mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W}$.

Since $\mathbf{A} = \mathbf{B}^2$ and $\lambda(\mathbf{A}) = (\lambda(\mathbf{B}))^2$, $\sum_{n=0}^{\infty}(\mathbf{I}_K - \mathbf{Z}_1\mathbf{W})^n\mathbf{Z}_1$ converges ⇔ $\sum_{n=0}^{\infty}(\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W})^n\mathbf{D}^{-1}$ converges.

According to [14], for $\sum_{n=0}^{\infty}(\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W})^n\mathbf{D}^{-1}$, a high probability convergence condition in terms of α is

$$\alpha > \frac{1}{\left(\sqrt{2}-1\right)^2}. \quad (17)$$

Thus, the Neumann series with $\mathbf{\Theta} = \mathbf{Z}_1$, that is, $\sum_{n=0}^{\infty}(\mathbf{I}_K - \mathbf{Z}_1\mathbf{W})^n\mathbf{Z}_1$, converges with a high probability when $\alpha > 1/(\sqrt{2}-1)^2$. □

By (15), the maximum value of K can be calculated for a specific N to achieve a high convergence probability for the Neumann series with $\mathbf{\Theta} = \mathbf{Z}_1$. Figure 1 shows the convergence probability with the configurations in Table 1.

As seen in Figure 1, the joint Newton iteration and Neumann series method can achieve a very high probability

TABLE 1: Maximum K values corresponding to different BS antenna numbers under the condition of (15).

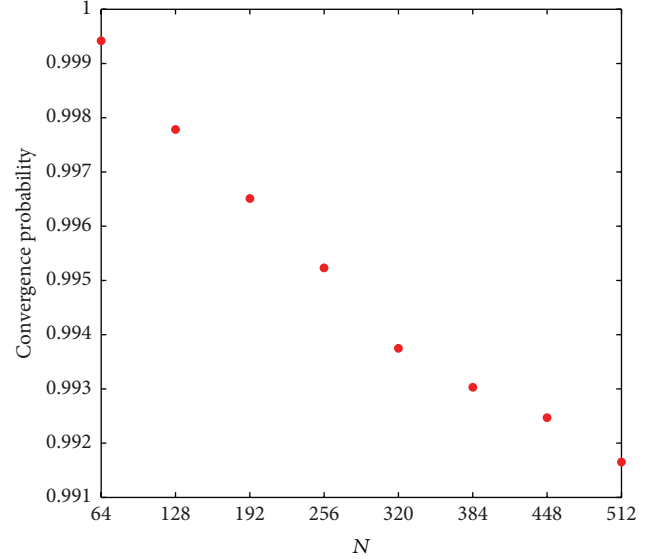| N	64	128	192	256	320	384	448	512
K	10	21	32	43	54	65	76	87

FIGURE 1: Convergence probability of Neumann series with $\mathbf{\Theta} = \mathbf{Z}_1$ under configurations in Table 1.

of convergence under the condition of (15), with the convergence probability as high as 0.999. Furthermore, when N is relatively small, the performance is slightly better. Also, as Figure 1 is obtained with the smallest allowed α under (15), it can be deduced that the convergence probability will be higher as α increases.

Lemma 2 has potential applications in massive MIMO systems. For the two typical downlink massive MIMO configurations $N \times K = 256 \times 16$ and $N \times K = 256 \times 32$ with $\alpha = 16$ and $\alpha = 8$, respectively [9], by Lemma 2, it can be concluded that the new joint Newton iteration and Neumann series method is convergent in both of these scenarios.

3.2. Complexity Analysis. The number of complex-valued multiplications is employed as a roughly estimated complexity of an algorithm. When $\mathbf{\Theta} = \mathbf{D}^{-1}$ or $\mathbf{\Theta} = (1/(N+K))\mathbf{I}_K$, the complexity of (8) is $O(K^2)$ for $L = 1$ and $O(K^3)$ for $L \geq 2$. However, when $\mathbf{\Theta} = \mathbf{Z}_1$, the complexity is $O(K^3)$ for $L \geq 1$. As a relatively large number of iterations (e.g., $L \geq 4$) are usually needed to avoid too much performance loss, the complexity of the three approaches is comparable. Note that, for different $\mathbf{\Theta}$ and $L \geq 2$, the complexity of the Neumann series is $O(K^3)$ which is comparable to that of the exact matrix inversion [10]. However, the Neumann series only contains matrix addition and matrix multiplication which is strongly preferred over matrix inversion in hardware since it does not require any divisions [13].

3.3. Discussion. Compared with the new joint Newton iteration and Neumann series method, conjugate gradient method

[8] and GS method [9] converge faster, but they approximate $\mathbf{W}^{-1}\mathbf{s}$ instead of $\mathbf{W}^{-1}$, as stated in Remark 1. However, $\mathbf{W}^{-1}$ is preferred in the computation of β and other related calculations, such as the sum rate computation of ZF precoding [9] and fast matrix inversion updates [15]. In comparison, the new joint Newton iteration and Neumann series method approximates $\mathbf{W}^{-1}$ directly and thus can significantly reduce the total complexity in those calculations. Furthermore, it only involves matrix multiplication and matrix addition, while conjugate gradient method contains many divisions which are more difficult for hardware implementation. Richardson method [7] was applied in minimum mean square error (MMSE) signal detection, but the relaxation parameter of the method still remains unknown for ZF precoding. In other words, Richardson method [7] may not converge for ZF precoding. By contrast, the new joint Newton iteration and Neumann series method will have a very high probability of convergence if (15) is satisfied.

4. Numerical Results

The bit error rate (BER) performances of different algorithms are evaluated in order to compare their convergence rates. For simplicity, algorithm 1, algorithm 2, and algorithm 3 represent the Neumann series with $\mathbf{\Theta} = \mathbf{D}^{-1}$, $\mathbf{\Theta} = \mathbf{Z}_1$, and $\mathbf{\Theta} = (1/(N + K))\mathbf{I}_K$, respectively. Moreover, ZF precoding with exact matrix inversion of $\mathbf{W}$ is also included as the benchmark. Newton iteration method usually needs complex calculation for initial estimation, which limits its range of application [10]. Therefore, Newton iteration method is not involved in comparison. The typical downlink massive MIMO configuration with $N \times K = 256 \times 32$ is considered and 64 QAM is employed as the modulation scheme.

Figures 2 and 3 show the BER performance comparison between the Neumann series with different $\mathbf{\Theta}$. It is obvious in Figure 2 that the BER performance of both the algorithms improves with the iteration number L. However, for a given number of iterations, algorithm 2 achieves a better BER performance than algorithm 1. Therefore, algorithm 2 has a faster convergence rate than algorithm 1. Moreover, when the iteration number is relatively large (e.g., $L = 4$), the BER of algorithm 2 is close to that of the exact matrix inversion, while algorithm 1 still suffers a great performance loss in comparison. In Figure 3, the advantage of algorithm 2 in convergence rate becomes more obvious. Note that the BER performance of algorithm 3 is even worse when $L = 4$ than when $L = 3$, which implies that algorithm 3 may not converge when $\alpha = 8$. Similarly, in Figure 4, Richardson method shows signs of nonconvergence, which indicates that the relaxation parameter configuration in [7] may not suit ZF precoding. From Figures 2, 3, and 4, it can be concluded that algorithm 2 has the fastest convergence rate.

5. Conclusions

In this paper, a new joint Newton iteration and Neumann series method has been studied for matrix inversion computation involved in linear precoding techniques. Newton
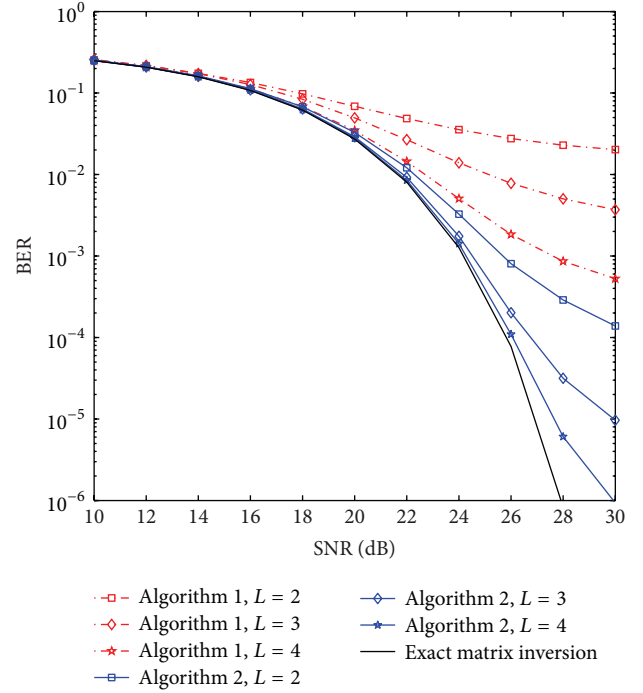

FIGURE 2: BER performance comparison between algorithm 1 and algorithm 2 in $N \times K = 256 \times 32$ massive MIMO system.
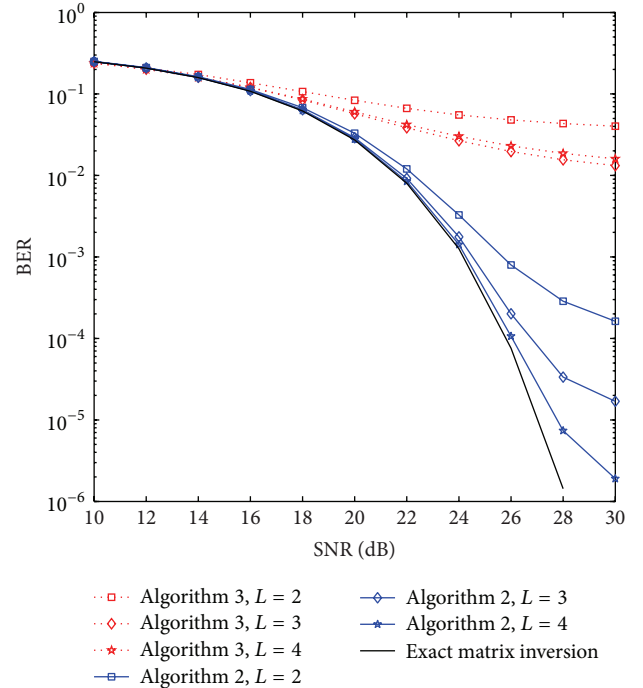

FIGURE 3: BER performance comparison between algorithm 2 and algorithm 3 in $N \times K = 256 \times 32$ massive MIMO system.

iteration method was employed to choose the initial value for the Neumann series. A high probability convergence condition was derived to ensure the convergence of the new method. Simulation results were provided to illustrate that the new matrix inversion computing method converges faster than some old ones.
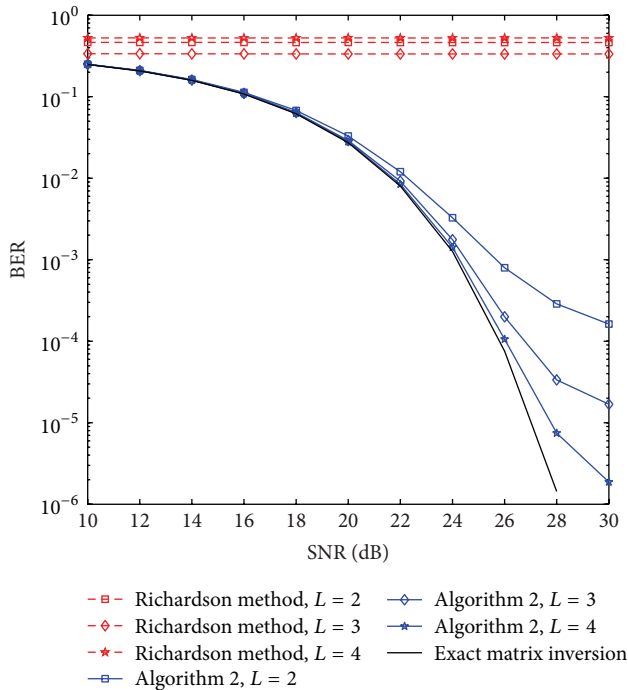
Figure 4: BER performance comparison between algorithm 2 and Richardson method in $N \times K = 256 \times 32$ massive MIMO system.
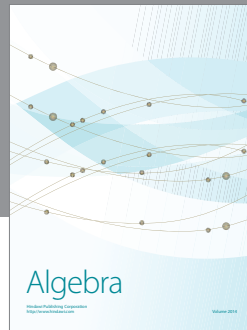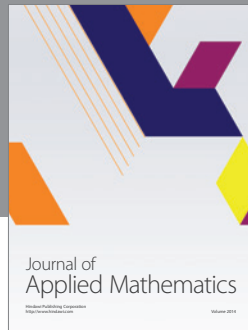
Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] X. Qin, Z. Yan, and G. He, "A near-optimal detection scheme based on joint steepest descent and Jacobi method for uplink massive MIMO systems," *IEEE Communications Letters*, vol. 20, no. 2, pp. 276–279, 2016.

[2] J. G. Andrews, S. Buzzi, W. Choi et al., "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] J. Thompson, X. Ge, H.-C. Wu et al., "5G wireless communication systems: prospects and challenges part 2," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 24–26, 2014.

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE Journal on Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

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

[6] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?"

[7] X. Gao, L. Dai, Y. Ma, and Z. Wang, "Low-complexity near-optimal signal detection for uplink large-scale MIMO systems," *Electronics Letters*, vol. 50, no. 18, pp. 1326–1328, 2014.

[8] B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, "Conjugate gradient-based soft-output detection and precoding in massive MIMO systems," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '14)*, pp. 3696–3701, Austin, Tex, USA, December 2014.

[9] X. Gao, L. Dai, J. Zhang, and S. Han, "Capacity-approaching linear precoding with low-complexity for large-scale MIMO systems," in *Proceedings of the IEEE International Conference on Communications (ICC '15)*, pp. 1577–1582, London, UK, June 2015.

[10] C. Tang, C. Liu, L. Yuan, and Z. Xing, "High precision low complexity matrix inversion based on newton iteration for data detection in the massive MIMO," *IEEE Communications Letters*, vol. 20, no. 3, pp. 490–493, 2016.

[11] 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 *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '13)*, pp. 2710–2715, Shanghai, China, April 2013.

[12] 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 Journal on Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 916–929, 2014.

[13] F. Rusek, D. Persson, B. K. Lau et al., "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[14] D. Zhu, B. Li, and P. Liang, "On the matrix inversion approximation based on Neumann series in massive MIMO systems," in *Proceedings of the IEEE International Conference on Communications (ICC '15)*, pp. 1763–1769, London, UK, June 2015.

[15] F. Rosario, F. A. Monteiro, and A. Rodrigues, "Fast matrix inversion updates for massive MIMO detection and precoding," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 75–79, 2016.

IEEE Journal on Selected Areas in Communications, vol. 31, no. 2, pp. 160–171, 2013.