Efficient Iterative Soft Detection Based on Polynomial Approximation for Massive MIMO

Feng Wang[1], Chuan Zhang[1], Xiao Liang[1], Zhizheng Wu[1], Shugong Xu[2], and Xiaohu You[1]
[1]National Mobile Communications Research Laboratory, Southeast University, Nanjing, China
[2]Intel Collaborative Research Institutes on Mobile Networking and Computing, Intel Labs, Shanghai, China
Email: [1]{wfeng, chzhang, 213112009, jasonwu, xhyu}@seu.edu.cn, [2]shugong.xu@intel.com

***Abstract*—In massive multiple-input multiple-output (MIMO) systems, linear minimum mean square error (MMSE) detection is near-optimal but involves large dimensional matrix inversion, which results in high complexity. To this end, *Neumann* series expansion (NSE) approximation, which avoids the direct computation of the matrix inversion, is recently investigated due to its low implementation complexity. Unfortunately, the complexity reduction can only be achieved well when the required number of the NSE terms L is small. To solve this problem, we proposed an iterative NSE (INSE) algorithm for MMSE detection at a manageable complexity even for large L. An approximation method based on NSE is proposed to compute the log-likelihood ratios (LLRs) for channel decoders. Both analytical and numerical results have demonstrated that, the overall complexity of the proposed soft-output MMSE-INSE algorithm is significantly reduced compared with the conventional NSE method and the *Cholesky* decomposition method, while keeping similar detection performance.**

***Index Terms*—Massive MIMO, MMSE detection, approximate matrix inversion, *Neumann* series, soft-output decision.**

I. Introduction

Massive multiple-input multiple-output (MIMO) has been considered as one of the key techniques for the 5th generation (5G) wireless systems [1], [2]. However, it entails an unprecedented number of antennas, which results in a big challenge because of its significantly increasing complexity [3]. One of the critical challenges is the low-complexity uplink signal detection. Usually, linear signal detection algorithms such as zero-forcing (ZE) detection and minimum mean square error (MMSE) detection are considered because of their near-optimal performance for massive MIMO systems [1].

Linear signal detections are always matrix-inversion intensive. For massive MIMO systems, exact matrix inversion based on QR-*Gram Schmidt* [4], *Gauss-Jordan* [5], or *Cholesky* decomposition [6] would entail prohibitive computational complexity of $\mathcal{O}(K^3)$ (K is the number of users) [4]–[6], which comes from the involved complicated data-flow and massive operations of multiplication, division, involution, and evolution. To this end, matrix inversion based on *Neumann* series expansion (NSE) approximation, which equals a sum of powers (multiplications) of matrices, is recently proposed for both uplink massive MIMO detection [7] and downlink pre-coding [8]. Compared to those exact matrix inversion approaches, this method is much more hardware friendly because of its simpler data flow and fewer atomic arithmetic units. Moreover, its inherent parallelization makes it favorable for high-speed communications. However, this complexity advantage can be achieved only when the number of the NSE terms L is small [7], [8].

By exploiting the NSE method for uplink massive MIMO systems, we propose a low-complexity soft-output MMSE detection algorithm based the conventional NSE method. First, a low-complexity iterative NSE-based (INSE) algorithm, which can avoid the direct inversion of the corresponding matrix, is proposed. To compute the log-likelihood ratios (LLRs) for channel decoder [9], a low-complexity approximation method, which is also based on NSE method with a negligible performance loss, is also given. Comparisons have shown that the proposed approach achieves a good compromise of performance and complexity even for large L.

The remainder of this paper is organized as follows. Section II gives the corresponding preliminaries. Section III presents the proposed INSE detection algorithm. Complexity comparison and numerical comparison are given in Section IV and Section V, respectively. Section VI concludes the entire paper.

Notation: In this paper, lowercase and upper bold face letters stand for column vectors and matrices, respectively. The operations $(\cdot)^T$, $(\cdot)^H$, and $\mathbb{E}\{\cdot\}$ denote transpose, conjugate transpose, and expectation, respectively. The entry in the i^{th} row and j^{th} column of $\mathbf{A}$ is a_{ij}. The k^{th} entry of vector $\mathbf{a}$ is a_k. $\mathbf{I}_K$ and $\mathbf{0}_K$ refer to the $K \times K$ identity matrix and zero matrix, respectively. The computational complexity is denoted in term of required number of complex-valued multiplications.

II. Preliminaries

A. System Model

Consider an uplink massive multiuser MIMO (MU-MIMO) system with a single base station (BS) equipped with N antennas which simultaneously serves K single-antenna user terminals (UTs). The transmitted sequence is first encoded by a convolutional encoder and then mapped to constellation $\mathbb{Q}$. Let $s_k \in \mathbb{C}$ ($k = 1, 2, ..., K$) denotes the encoded symbol of the k^{th} UT with the assumption of $\mathbb{E}\{|s_k|^2\} = E_s = 1$. Then $\mathbf{y} \in \mathbb{C}^{N\times 1}$ denotes the received symbol vector at the BS side. $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ ($k = 1, 2, ..., K$) denotes the channel gain between the k^{th} UT and the BS. Therefore, the received signal can be expressed as

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

978-1-4673-7687-7/15/$31.00 ©2015 IEEE

where $\mathbf{n}$ is the additive white Gaussian noise (AWGN) with zero-mean and variance σ^2.

Now we have $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$. Here $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_K]$ is the $N \times K$ channel responding matrix, whose entries are assumed to be independent identical distributed (i.i.d.) circularly symmetric complex Gaussian with zero-mean and unit variance. $\mathbf{s} = [s_1, s_2, ..., s_K]^T$ denotes the $K \times 1$ encoded transmit symbol vector.

B. Conventional Soft-Output MMSE Decision

It is assumed that the receiver has perfect knowledge of the channel $\mathbf{H}$. The demodulator then uses the received vector $\mathbf{y}$ and $\mathbf{H}$ to calculate the LLR for $b_{m,k}$, i.e., $LLR_{m,k}$, where $b_{m,k}$ denotes the m^{th} bit of symbol s_k. The resulting sequence of LLRs are then used by soft-input *Viterbi* decoder [9].

According to [10], one can first calculate the MMSE estimate of the transmitted symbol $\hat{\mathbf{s}} = [\hat{s}_1, \hat{s}_2, ..., \hat{s}_k]^T$ by

$$\hat{\mathbf{s}} = \mathbf{T}_{\text{MMSE}}\mathbf{y}, \tag{2}$$

where $\mathbf{T}_{\text{MMSE}} = \mathbf{W}^{-1}\mathbf{H}^H$ is the equivalent MMSE filter matrix, and $\mathbf{W}$ is denoted by $\mathbf{W} = \mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}_K$. Then the max-log $LLR_{m,k}$ [11] is computed as:

$$LLR_{m,k} = \frac{c_{kk}}{1 - c_{kk}}\left[\min_{a \in \mathbb{Q}_m^0}\left|\frac{\hat{s}_k}{c_{kk}} - a\right|^2 - \min_{a' \in \mathbb{Q}_m^1}\left|\frac{\hat{s}_k}{c_{kk}} - a'\right|^2\right], \tag{3}$$

where $\mathbb{Q}_m^0$ and $\mathbb{Q}_m^1$ are sub-sets of $\mathbb{Q}$, of which the m^{th} bit equals 0 and 1, respectively. c_{kk} is the k^{th} diagonal entry of matrix $\mathbf{C}$, which is given as follows according to [11]:

$$\mathbf{C} = \mathbf{W}^{-1}(\mathbf{H}^H\mathbf{H}). \tag{4}$$

This soft-output MMSE linear algorithm inevitably involves complicated matrix inversion, i.e., $\mathbf{W}^{-1}$, to compute MMSE estimate $\hat{\mathbf{s}}$ and the prior coefficient c_{kk} [9]. However, the computational complexity of conventional matrix inversion approaches stays $\mathcal{O}(K^3)$, which entails prohibitive complexity and hinders the application of this soft-output MMSE algorithm for massive MIMO systems.

III. Low-Complexity INSE-Based Detection

In this section, we first propose a low-complexity iterative detection algorithm based on conventional NSE method for massive MIMO, which successfully avoids direct matrix inversion. Then, an approximation method to compute c_{kk} in Eq. (3) is proposed, which further lowers the complexity.

A. Detection Based on Conventional NSE Method

For illustration purpose, we summarize the theoretical background in Lemma 1 as follows.

Lemma 1 (Neumann series expansion [12]): For a $K \times K$ non-singular matrix $\mathbf{P}$ satisfying $\lim_{i\to\infty} \mathbf{P}^i = \mathbf{0}_K$, $(\mathbf{I}_K - \mathbf{P})$ is also non-singular and its inverse is given by

$$(\mathbf{I}_K - \mathbf{P})^{-1} = \sum_{i=0}^{\infty} \mathbf{P}^i. \tag{5}$$

For uplink massive MIMO, the channel matrix $\mathbf{H}$ can be assumed as column asymptotically orthogonal [1], which guarantees that $\mathbf{G} = \mathbf{H}^H\mathbf{H}$ (and also $\mathbf{W} = \mathbf{G} + \sigma^2\mathbf{I}_K$) is *Hermitian* positive definite. By applying the Lemma 1, we rewrite $\mathbf{W}$ using NSE as

$$\mathbf{W}^{-1} = \sum_{l=0}^{\infty}(\mathbf{I}_K - \mathbf{X}^{-1}\mathbf{W})^l\mathbf{X}^{-1}, \tag{6}$$

where $\mathbf{X}^{-1}$ is an arbitrary matrix satisfying $\lim_{l\to\infty}(\mathbf{I}_K - \mathbf{X}^{-1}\mathbf{W})^l = \mathbf{0}_K$. Keeping only the first L terms in Eq. (6), we obtain the L-term approximation as

$$\widehat{\mathbf{W}}_L^{-1} = \sum_{l=1}^{L}(\mathbf{I} - \mathbf{X}^{-1}\mathbf{W})^{l-1}\mathbf{X}^{-1}. \tag{7}$$

So the L-term approximation of MMSE filter matrix is $\widehat{\mathbf{T}}_{\text{MMSE},L} = \widehat{\mathbf{W}}_L^{-1}\mathbf{H}^H$. The conventional approximation of MMSE estimate based on NSE method is given by

$$\hat{\mathbf{s}}_L = \widehat{\mathbf{T}}_{\text{MMSE},L}\mathbf{y}, \tag{8}$$

where $\hat{\mathbf{s}}_L = \hat{\mathbf{s}}(L \to \infty)$ is the exact estimate.

In this paper, we initiate $\mathbf{X} = \mathbf{D}$, where $\mathbf{D}$ is the diagonal of $\mathbf{W}$. For massive MIMO systems, since $\mathbf{W}$ tends to be a diagonally dominant matrix [1], the initiation of $\mathbf{X} = \mathbf{D}$ is a good choice to balance the performance and complexity [7], [8]. For $L = 1$, we have $\widehat{\mathbf{W}}_1^{-1} = \mathbf{D}^{-1}$, where the MMSE estimation coincides to a re-scaled matched-filter (MF) equalizer. The inverse of $\mathbf{D}$ only requires K reciprocal operations. For $L = 2$, $\widehat{\mathbf{W}}_2^{-1} = \mathbf{D}^{-1} + (\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W})\mathbf{D}^{-1}$. Since $\mathbf{D}$ is a diagonal matrix, the computation of $\widehat{\mathbf{W}}_2^{-1}$ only entails $\mathcal{O}(K^2)$ complex-valued multiplications, in contrary to $\mathcal{O}(K^3)$ for the exact matrix inversion algorithms. However, for $L \geq 3$, the conventional NSE approximation method still suffers from a higher complexity of $\mathcal{O}(K^3)$ because of the inevitable matrix-matrix multiplications.

B. Proposed Low-Complexity MMSE-INSE Detection

In order to reduce the computational complexity for $L \geq 3$, matrix-matrix multiplications should be avoid. Since matrix-vector multiplication only has a complexity of $\mathcal{O}(K^2)$, we would like to decompose the sum of matrix powers in Eq. (7) into sum of matrix-vector multiplications. For the proposed low-complexity MMSE-INSE detection algorithm, we summarise its main points in Theorem 1.

Theorem 1: For uplink massive MIMO systems, if BS has the knowledge of channel responding matrix $\mathbf{H}$, variance of thermal noise σ^2, and received symbol $\mathbf{y}$, the approximated MMSE estimate is given by

$$\hat{\mathbf{s}}_{(l)} = \mathbf{\Theta}\hat{\mathbf{s}}_{(l-1)} + \hat{\mathbf{s}}_{(1)},\ l > 1, \tag{9}$$

where $\mathbf{\Theta} = (\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W})$ denotes the matrix multiplication coefficient, l is the number of iterations, and $\hat{\mathbf{s}}_{(1)} = \mathbf{D}^{-1}(\mathbf{H}^H\mathbf{y})$ is the initial solution. For $l \to \infty$, $\hat{\mathbf{s}}_{(l)}$ becomes the exact MMSE estimate.

Proof: Please see Appendix A. ∎

Theorem 1 shows that each iteration of Eq. (9) only involves a matrix-vector multiplication and a vector-vector addition, whose complexity is much lower compared with the conventional MMSE detection based on NSE in Eq. (8). Detailed complexity analysis is presented in Section IV.

C. Efficient Approximation Method to Compute LLRs

Although the proposed algorithm can obtain the MMSE estimation $\hat{\mathbf{s}}$ with a low complexity of $\mathcal{O}(K^2)$, the conventional methods calculating c_{kk} still suffer from the complexity of $\mathcal{O}(K^3)$. This is because inverse of $\mathbf{W}$ has to be involved. To this end, a NSE-based approximation method is proposed to calculate c_{kk} with a negligible performance loss.

First, we rewrite Eq. (4) as follows:

$$\mathbf{C} = \mathbf{W}^{-1}(\mathbf{H}^H\mathbf{H}) = \mathbf{I}_K - \sigma^2\mathbf{W}^{-1}. \tag{10}$$

Eq. (10) shows that c_{kk} is a linear function of w'_{kk}:

$$c_{kk} = 1 - \sigma^2 w'_{kk}, \quad k = 1, ..., K, \tag{11}$$

where w'_{kk} and w_{kk} denote the k^{th} diagonal entry of $\mathbf{W}^{-1}$ and $\mathbf{W}$, respectively. For complexity issue, we focus on the approximated w'_{kk}, instead of the approximation of $\mathbf{W}^{-1}$.

According to Eq. (7), $\widehat{\mathbf{W}}_1^{-1} = \mathbf{D}^{-1}$ for $L = 1$. Therefore, we have $w'_{kk} \approx w_{kk}^{-1}$. For $L = 2$, $\widehat{\mathbf{W}}_2^{-1} = \mathbf{D}^{-1} + \mathbf{\Theta}\mathbf{D}^{-1}$. Therefore, w'_{kk} can be approximated by $w'_{kk} \approx w_{kk}^{-1} + \theta_{kk} w_{kk}^{-1}$, where θ_{kk} is the k^{th} diagonal entry of $\mathbf{\Theta}$. For $L = 3$, $\widehat{\mathbf{W}}_3^{-1} = \mathbf{D}^{-1} + \mathbf{\Theta}\mathbf{D}^{-1} + \mathbf{\Theta}^2\mathbf{D}^{-1}$. w'_{kk} can be approximated by

$$w'_{kk} \approx w_{kk}^{-1} + \theta_{kk} w_{kk}^{-1} + \theta'_k \theta_k w_{kk}^{-1}, \tag{12}$$

where θ'_k and θ_k denote the k^{th} row and k^{th} column of $\mathbf{\Theta}$, respectively. However, for $L = 4$, w'_{kk} is approximated by $w_{kk}^{-1} + \theta_{kk} w_{kk}^{-1} + \theta'_k \theta_k w_{kk}^{-1} + \theta'_k \mathbf{\Theta} \theta_k w_{kk}^{-1}$, whose computational complexity is $\mathcal{O}(K^3)$. Therefore, in this paper, for $L \geq 4$, the approximation of Eq. (12) is still employed. Substituting w'_{kk} and $\hat{\mathbf{s}}_L$ into Eq. (3) and (4), we can obtain LLRs finally.

For the massive MIMO systems, the property that $\mathbf{W}$ tends to be diagonally dominant [1], [7], [8] guarantees the accuracy of the proposed approximation method. Comparisons in Section V have verified that the resulting performance loss is negligible. For more details, please refer to Section V.

The entire procedure is summarized in Algorithm 1 and referred to as *low-complexity soft-output MMSE-INSE detection algorithm*.

IV. Computational Complexity Analysis

In this section, the computational complexity of the proposed algorithm is analyzed and compared with conventional algorithms. It should be noted that all detection algorithms mentioned here are soft-output ones and the complexity does not include channel encoder and decoder.

Since both linear MMSE algorithms and the proposed algorithm have to pre-compute $\mathbf{W} = \mathbf{G} + \sigma^2\mathbf{I}_K$ and $\mathbf{y}^{\mathrm{HF}} = \mathbf{H}^H\mathbf{y}$, we only focus on the complexity of the following four distinguishable parts of Algorithm 1 but not the full complexity to make a distinction contrast:

Algorithm 1 Proposed soft-output MMSE-INSE algorithm

input: $\mathbf{H}$, $\mathbf{y}$ and σ^2.

1: Compute $\mathbf{G} = \mathbf{H}^H\mathbf{H}$, $\mathbf{W} = \mathbf{G} + \sigma^2\mathbf{I}_K$ and $\mathbf{y}^{\mathrm{MF}} = \mathbf{H}^H\mathbf{y}$.
2: Compute $\mathbf{D}^{-1} = \mathrm{diag}(w_{11}^{-1}, w_{22}^{-1}, ..., w_{KK}^{-1})$ and $\hat{\mathbf{s}}_{(1)} = \mathbf{D}^{-1}\mathbf{y}^{\mathrm{MF}}$.
3: Compute the multiplication coefficient $\mathbf{\Theta} = \mathbf{I}_K - \mathbf{D}^{-1}\mathbf{W}$.
4: **For** l=2,...,L, do $\hat{\mathbf{s}}_{(l)} = \mathbf{\Theta}\hat{\mathbf{s}}_{(l-1)} + \hat{\mathbf{s}}_{(1)}$.
5: For $k = 1, ..., K$, compute the w'_{kk} according to Section III-C and c_{kk} as $c_{kk} = 1 - \sigma^2 w'_{kk}$.

Output: $\hat{\mathbf{s}} = \hat{\mathbf{s}}_{(L)}$ and $[c_{11}, c_{22}, ..., c_{KK}]$.

---

1) To compute $\hat{\mathbf{s}}_{(1)}$: Since $\mathbf{D}$ is a $K \times K$ diagonal matrix, its inverse can be achieved by a reciprocal unit. In this paper, we estimate its computational complexity in terms of complex-valued multiplications. Note that $\mathbf{y}^{\mathrm{MF}}$ is a $K \times 1$ column vector, the complexity for computing $\hat{\mathbf{s}}_{(1)}$ is $\mathcal{O}(2K)$.

2) To compute $\mathbf{\Theta}$: The multiplication of a $K \times K$ diagonal matrix $\mathbf{D}^{-1}$, and a $K \times K$ matrix $\mathbf{W}$, has the computational complexity of $\mathcal{O}(K^2)$. Since the matrix $\mathbf{\Theta}$ is *Hermitian*, only the lower-triangle part of multiplication array should be processed [13]. So the total complexity is $\mathcal{O}(1/2K^2)$.

3) To compute $\hat{\mathbf{s}}_{(L)}$: Each iteration requires a multiplication of a $K \times K$ matrix and a $K \times 1$ vector. The computation of $\hat{\mathbf{s}}_{(L)}$ $(L \geq 2)$ has the complexity of $\mathcal{O}\left((L-1)K^2\right)$.

4) To compute c_{kk}: According to Section III-C, the approximation of w_{kk} is always given by Eq. (12). Therefore, the complexity of calculating c_{kk} (for $k = 1, ...K$) is $\mathcal{O}(K^2 + 3K)$.

To sum up, the overall complexity for the proposed *low-complexity soft-output MMSE-INSE detection algorithm* is $\mathcal{O}\left((L + 1/2)K^2 + 5K\right)$. The comparison with the conventional NSE algorithm and the *Cholesky* decomposition based algorithm is listed in Table I. For massive MIMO, since K is usually large, complexity of the proposed algorithm reduces from $\mathcal{O}(K^3)$ to $\mathcal{O}(LK^2)$ (L is usually relatively small). However, the conventional NSE method and *Cholesky* decomposition-based method suffer from complexities of $\mathcal{O}\left((L-2)K^3\right)$ $(L \geq 3)$ [7] and $\mathcal{O}(5/6K^3)$ [12], respectively. Illustrated in Fig. 1, the proposed algorithm significantly reduces the computational complexity, especially when the ratio of N/K grows large.

V. Numerical Simulation

The numerical results of BER performance against the signal-to-noise (SNR) are provided to compare the proposed soft-out MMSE-INSE algorithm with the *Cholesky* decomposition approach and the conventional NSE-based approach. The simulation is carried out with a block *Rayleigh* fading channel with 96 symbols of each block, where each symbol is modulated with 16-QAM scheme. Here the convolutional code with a rate $\approx 3/4$ in [14] and *Viterbi* decoder are employed at the transmitter side and the receiver side, respectively. Furthermore, the SNR is defined by NE_s/σ^2, which corresponds to the average SNR at each BS antenna. Again, all algorithms mentioned here are soft-output ones.

TABLE I
COMPUTATIONAL COMPLEXITY FOR DIFFERENT APPROACHES.

Different Designs	Computational Complexity	
MMSE-INSE	part 1)	$2K$
	part 2)	$1/2K^2$
	part 3)	$(L-1)K^2$
	part 4)	K^2+3K
	total ($L \geq 2$)	$(L+1/2)K^2+5K$
MMSE-NSE [7]	$L=2$	$3K^2$
	$L \geq 3$	$(L-2)K^3+3K^2$
MMSE-*Cholesky* [12]	$5/6K^3+3/4K^2+4/3K$	

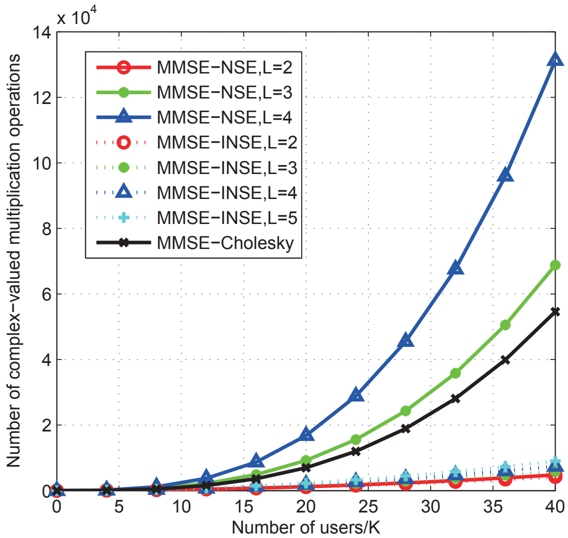

Fig. 1. Complexity complexity comparison against the number of users in terms of the number of complex-valued multiplication operations.

Fig. 2 illustrates the coded BER performances of the proposed MMSE-INSE algorithm, the MMSE-*Cholesky* algorithm, and the conventional MMSE-NSE algorithm for $N = 70$, $K = 12$, and $L = 3$, 5, and 7. According to Fig. 2, we observe that when the number of iterations grows, the BER performance of both approximation algorithms can achieve comparable performance as MMSE-*Cholesky* algorithm. For BER$\approx 0.3 \times 10^{-4}$, the difference between the conventional MMSE-*Cholesky* algorithm and both approximation methods with $L = 7$ is within 0.3 dB. Moreover, for $L \leq 3$, performances of both approximation methods are the same. For $L \geq 4$, difference of the BER performances between those two approximation methods is negligible. Notably, according to Table I, for $L \geq 3$ the proposed algorithm demonstrates a low complexity of $\mathcal{O}\left((L+1)K^2\right)$ with a fixed hardware architecture [15], whereas the convention MMSE-NSE algorithm entails a higher complexity of $\mathcal{O}\left((L-2)K^3\right)$ (as shown in Fig. 1) and has to be carefully re-implemented every time when the value of L changes.

The overall BER performances of all the approaches illustrated in Fig. 3 is much better than those illustrated in Fig. 2 because of the more *favorable propagation environment* for massive MIMO systems (see more details in [1], [7]). In this paper, the more *favorable propagation environment* means larger ratio of N/K, which results in a better condition of $\mathbf{W}$ for the initialization of $\mathbf{X} = \mathbf{D}$ in Eq. (7). Thus, in a more *favorable propagation environment* ($N/K \geq 8$), the number of iterations L required by the proposed MMSE-INSE method, can be smaller compared to lower ratio conditions. On the contrary, the proposed algorithm can easily offer a larger-term approximation to accommodate the less *favorable propagation environment* thanks to its fixed hardware architecture and much lower processing latency compared to the conventional

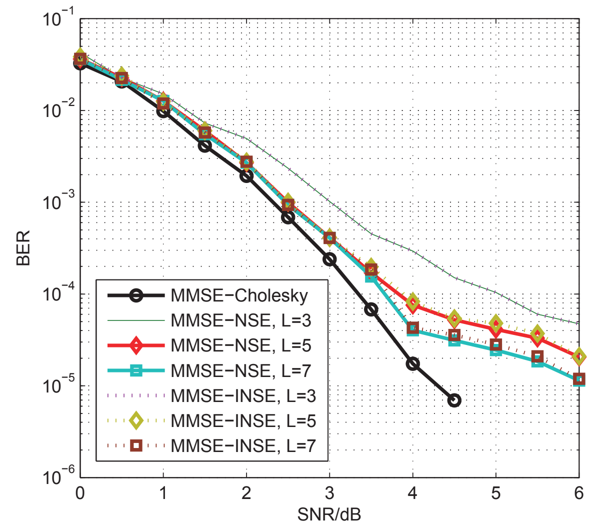

Fig. 2. Coded BER performance comparison between the proposed MMSE-INSE algorithm and the MMSE-NSE algorithm, for $N = 70$, $K = 12$.

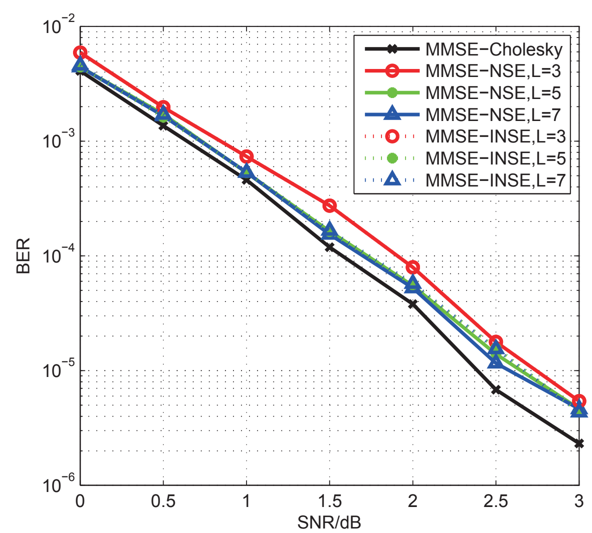

Fig. 3. Coded BER performance comparison between the proposed MMSE-INSE algorithm and the MMSE-NSE algorithm, for $N = 96$, $K = 12$.

MMSE-NSE detection algorithm.

VI. Conclusion

In this paper, we propose a low-complexity soft-output MMSE-INSE algorithm by fully exploiting the characteristics of NSE method for the uplink massive MIMO systems. Analysis shows that the proposed algorithm can successfully reduce the computational complexity from $\mathcal{O}(K^3)$ to $\mathcal{O}(K^2)$, compared to the conventional MMSE-NSE algorithms. Simulation results have verified that the proposed algorithm can achieve similar performance compared to conventional MMSE-NSE method and the classical MMSE-*Cholesky* algorithm with a small number of iterations. Moreover, the proposed algorithm has demonstrated its implementation advantage in freely changing value of L to achieve the performance/complexity trade-off for various propagation environments. Since the proposed algorithm has shown the great potential of being hardware-friendly, more detailed studies of the hardware architecture and implementation will be investigated in our further work soon.

Acknowledgement

This work is partially supported by NSFC under grants 61501116 and 61221002, International Science & Technology Cooperation Program of China under grant 2014DFA11640, Huawei HIRP Flagship Program under grant YB201504, Jiangsu Provincial NSF under grant BK20140636, Intel Collaborative Research Institute for MNC, the Fundamental Research Funds for the Central Universities under grants 3204004202, 3204004102, and 3204005101, the Research Fund of National Mobile Communications Research Laboratory, Southeast University under grant 2014B02, and the Project Sponsored by the SRF for the Returned Overseas Chinese Scholars of State Education Ministry.

Appendix

To prove Theorem 1, first of all, we reform the NSE of Eq. (7) into a iterative architecture. Considering the property of Eq. (7), we further rewrite Eq. (7) into iterative form as

$$\widehat{\mathbf{W}}_{(l)}^{-1} = \mathbf{\Theta}\widehat{\mathbf{W}}_{(l-1)}^{-1} + \mathbf{D}^{-1}, \tag{13}$$

where $\widehat{\mathbf{W}}_{(1)}^{-1} = \mathbf{D}^{-1}$, and $l(\geq 2)$ is the number of iteration.

Substituting Eq. (13) into Eq. (2), we have the approximation MMSE estimate as

$$\begin{aligned}\hat{s}_{(l)} &= \mathbf{T}_{\text{MMSE},(l)}\mathbf{y} = \widehat{\mathbf{W}}_{(l)}^{-1}\mathbf{H}^H\mathbf{y} \\ &= \left(\mathbf{\Theta}\widehat{\mathbf{W}}_{(l-1)}^{-1} + \mathbf{D}^{-1}\right)\mathbf{H}^H\mathbf{y}.\end{aligned} \tag{14}$$

For the calculation of Eq. (14), we begin to compute the prior information of detection, i.e., the matched-filter output of $\mathbf{y}$, $\mathbf{y}^{\text{MF}} = \mathbf{H}^H\mathbf{y}$, and the initial estimated MMSE symbol vector $\hat{\mathbf{s}}_{(1)} = \mathbf{X}^{-1}\mathbf{y}^{\text{MF}}$, instead of directly calculating the approximation of $\mathbf{W}^{-1}$. To guarantee correctness, the following Lemma 2 is applied in each related operations.

Lemma 2: For arbitrary $\mathbf{A} \in \mathbb{C}^{P\times Q}$, $\mathbf{B} \in \mathbb{C}^{Q\times K}$, $\mathbf{c} \in \mathbb{C}^{K\times 1}$, we have $(\mathbf{AB})\mathbf{c} = \mathbf{A}(\mathbf{Bc})$.

Proof: Let $\mathbf{A} = [a_{ij}]_{P\times Q}$, $\mathbf{B} = [b_{ij}]_{Q\times K}$, $\mathbf{c} = [c_i]_{K\times 1}$, $\mathbf{AB} = [d_{ij}]_{P\times K}$, $\mathbf{Bc} = [e_i]_{Q\times 1}$, $(\mathbf{AB})\mathbf{c} = [f_i]_{P\times 1}$, $\mathbf{A}(\mathbf{Bc}) = [g_i]_{P\times 1}$. According to the definition of matrix multiplication, we have $d_{ij} = \sum_{n=1}^{Q} a_{in}b_{nj}$, and $e_i = \sum_{n=1}^{K} b_{in}c_n$. Then, we have

$$f_i = \sum_{n=1}^{K} d_{in}c_n = \sum_{n=1}^{K}\sum_{l=1}^{Q} a_{il}b_{ln}c_n, \tag{15}$$

$$g_i = \sum_{l=1}^{Q} a_{il}e_l = \sum_{n=1}^{K}\sum_{l=1}^{Q} a_{il}b_{ln}c_n. \tag{16}$$

Thus, for arbitrary entry f_i of $(\mathbf{AB})\mathbf{c}$ and g_i of $\mathbf{A}(\mathbf{Bc})$, we have $f_i = g_i$. Finally, we obtain $(\mathbf{AB})\mathbf{c} = \mathbf{A}(\mathbf{Bc})$. ■

Note that $\hat{\mathbf{s}}_{(l)} = \widehat{\mathbf{W}}_{(l)}^{-1}\mathbf{y}^{\text{MF}}$ is the l^{th} iterative output, the estimate of transmitted symbol based MMSE-INSE method is calculated by Eq. (9). By apply Lemma 1, the $\hat{\mathbf{s}}_{(l)}$ converges to the exact estimate $\hat{\mathbf{s}}$ when l grows large.

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 Commun. Lett.*, vol. 52, no. 2, pp. 3590–3600, Nov. 2014.

[2] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, 2014.

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

[4] C. K. Singh, S. H. Prasad, and P. T. Balsara, "VLSI architecture for matrix inversion using modified Gram-Schmidt based QR decomposition," in *Proc. VLSI Design*, 2007, pp. 836–841.

[5] J. Arias-García, R. P. Jacobi, C. H. Llanos, and M. Ayala-Rincon, "A suitable FPGA implementation of floating-point matrix inversion based on Gauss-Jordan elimination," in *Proc. Southern Conference on Programmable Logic (SPL)*, 2011, pp. 263–268.

[6] C. Studer, S. Fateh, and D. Seethaler, "ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1754–1765, 2011.

[7] 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 Commun. Lett.*, vol. 40, no. 7, pp. 1566–1577, Nov. 2014.

[8] H. Prabhu, O. Edfors, J. Rodrigues, L. Liu, and F. Rusek, "Hardware efficient approximative matrix inversion for linear pre-coding in massive MIMO," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2014, pp. 1700–1703.

[9] S. G. Wilson, *Digital modulation and coding*. Prentice-Hall, Inc., 1995.

[10] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge university press, 2003.

[11] D. Seethaler, G. Matz, and F. Hlawatsch, "An efficient MMSE-based demodulator for MIMO bit-interleaved coded modulation," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 4, 2004, pp. 2455–2459.

[12] G. Stewart, *Matrix Algorithms: Volume 1, Basic Decompositions*. Cambridge University Press, 1998, vol. 1.

[13] M. Wu, B. Yin, A. Vosoughi, C. Studer, J. R. Cavallaro, and C. Dick, "Approximate matrix inversion for high-throughput data detection in the large-scale MIMO uplink," *IEEE Commun. Lett.*, vol. 8, no. 5, pp. 2155–2158, Nov. 2013.

[14] C. Berrou and A. Glavieux, "Turbo codes," *Encyclopedia of Telecommunications*, 2003.

[15] 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. of IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, accepted.