A Low Complexity Detection Algorithm For Uplink Massive MIMO Systems Based on Alternating Minimization

Anis Elgabli, Ali Elghariani*, Vaneet Aggarwal, and Mark Bell
Purdue University, West Lafayette IN 47907, *University of Tripoli, Libya
Email: aelgabli, vaneet, mrb@purdue.edu and a.elghariani@uot.edu.ly

Abstract—In this paper, we propose an algorithm based on the Alternating Minimization technique to solve the uplink massive MIMO detection problem. The proposed algorithm provides a lower complexity compared to the conventional MMSE detection technique, especially when the number of user equipment (UE) antennas is close to the number of base station (BS) antennas. This improvement is obtained without any matrix inversion. Moreover, the algorithm re-formulates the maximum likelihood (ML) detection problem as a sum of convex functions based on decomposing the received vector into multiple vectors. Each vector represents the contribution of one of the transmitted symbols in the received vector. Alternating Minimization is used to solve the new formulated problem in an iterative manner with a closed form solution update in every iteration. Simulation results demonstrate the efficacy of the proposed algorithm in the uplink massive MIMO setting for both coded and uncoded cases.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) is one of the most promising techniques for the 5th Generation (5G) networks due to its potential for enhancing throughput, spectra efficiency, and energy efficiency [1], [2]. Massive MIMO requires the BS to be equipped with arrays of hundreds of antennas to serve tens of user terminals with single or multi antennas.

Theoretical results of massive MIMO show that linear detectors such as zero forcing (ZF) and minimum mean square error (MMSE) can achieve optimum performance under the favorable propagation conditions [3]. The favorable propagation conditions mean that the number of BS antennas grows very large compared to the number of UE antennas, which leads to the column-vectors of the propagation matrix to be asymptotically orthogonal. As a result of this orthogonality, the ZF and MMSE detectors can be implemented with simple diagonal inversions [4].

The current practical number of the BS antennas in massive MIMO systems is in the order of tens to a hundred. This is far from the theoretical limit that leads to the orthogonality mentioned above [4]. Therefore, the linear detectors still need to perform a matrix inversion for the signal detection of the uplink massive MIMO system, which entails excessive computational complexity [5]. Neumann series expansion, Cholesky decomposition, and successive over-relaxation techniques are proposed in the literature to reduce the complexity of the matrix inversion process in the MMSE detector [6], [7]. These approaches require a lower computational complexity than the exact matrix inversion while delivering near-optimal results only for theoretical massive MIMO configurations. In realistic massive MIMO scenarios the performance of implementing these approximations is far from the exact MMSE performance [6]. In this letter we present a novel formulation of a low complexity iterative algorithm based on Alternating Minimization, referred to as AltMin. This algorithm provides similar bit error rate (BER) performance to the exact matrix inverted MMSE technique, with one order less complexity.

The proposed algorithm reformulates ML detection problem as a sum of convex functions based on decomposing the received vector into multiple vectors. Each vector represents the contribution of one of the transmitted symbols in the received vector. Then, Alternating Minimization is used to solve the new formulated problem in an iterative manner with a closed form solution update in every iteration that does not require any matrix inversion.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Consider the uplink data detection in a multi-user (MU) massive MIMO system with $N_t$ BS antennas and $N_r$ UE antennas. The vector $\mathbf{x} = (x_1, x_2, \ldots, x_{N_r})^T \in \mathbb{C}^{N_r \times 1}$ represents the complex transmitted signal, where $x_k$ is the transmitted symbol for user $k$ with $E|x_k|^2 = 1, \forall i$. Each user transmits symbols over a flat fading channels and the signals are demodulated and sampled at the receiver. The vector $\mathbf{y} = (y_1, y_2, \ldots, y_{N_r})^T \in \mathbb{C}^{N_r \times 1}$ represents the complex received signal, and the channel matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ can be represented as $(h_1, h_2, \ldots, h_{N_r})$, where $h_i = (h_{1,i}, h_{2,i}, \ldots, h_{N_t,i})^T \in \mathbb{C}^{N_r \times 1}$, and $h_{m,n}$ is the complex flat fading channel gain from transmit antenna $n$ to the receive antenna $m$, with $h_{m,n} \sim \mathcal{N}(0,1)$. The system can be modeled as $\mathbf{y} = \mathbf{Hx} + \mathbf{v}$, where the $\mathbf{v} = (v_1, v_2, \ldots, v_{N_r})^T \in \mathbb{C}^{N_r \times 1}$ is the complex additive white Gaussian noise (AWGN) vector whose elements are mutually independent and zero mean with variance $\sigma_v^2$. The corresponding real-valued system model is $\mathbf{y} = \mathbf{Hx} + \mathbf{v}$ [8]. The equivalent ML detection problem of the real model can
be written in the form \( \hat{x} = \arg\min \| y - Hx \|_2^2 \), where 
\( \chi = \frac{1}{M} \{-\sqrt{M} + 1, \ldots, -1, \ldots, \sqrt{M} - 1\} \), 
\( M \) is the QAM constellation size, and \( \frac{1}{M} \) the normalization factor.

**B. Problem Formulation**

First, we decompose the received vector \( y \) into a linear combination of vectors so that 
\( y = \sum_{i=1}^{2N_t} y_i \), where \( y_i \) represents the contribution of the \( i \)-th transmitted symbol in the received vector. The element-wise representation of the decomposed received vector is:

\[
\begin{bmatrix}
y^{(1)}_1 \\
y^{(2)}_1 \\
\vdots \\
y^{(2N_r)}_1 \\
y^{(1)}_2 \\
y^{(2)}_2 \\
\vdots \\
y^{(2N_r)}_2 \\
\vdots \\
y^{(1)}_{2N_t} \\
y^{(2)}_{2N_t} \\
\end{bmatrix} = \begin{bmatrix}
y_1^{(1)} \\
y_1^{(2)} \\
\vdots \\
y_1^{(2N_r)} \\
y_2^{(1)} \\
y_2^{(2)} \\
\vdots \\
y_2^{(2N_r)} \\
\vdots \\
y_{2N_t}^{(1)} \\
y_{2N_t}^{(2)} \\
\end{bmatrix}
\]

The \( k \)-th element of \( y \) can be represented as 
\( y^{(k)} = \sum_{i=1}^{2N_t} y_i^{(k)}, k = 1, \ldots, 2N_r \). Let \( h_i \) the \( i \)-th column of the channel matrix \( H \). Now, we relax the non-convexity constraint on the feasible set \( \chi \), and reformulate the ML problem based on the above decomposition as follows:

\[
\min_{x, y} \sum_{i=1}^{2N_t} \| y_i - h_i x_i \|_2^2 \quad \text{subject to} \quad \sum_{i=1}^{2N_t} y_i^{(k)} = y^{(k)}, \forall k = 1, \ldots, 2N_r \quad \text{and} \quad -l \leq x_i \leq l, \forall i = 1, \ldots, 2N_t \tag{1} \]

A. Solving the subproblem \( 4 \)

The Lagrangian function of \( 4 \) can be written as:

\[
\mathcal{L} = \sum_{i=1}^{2N_t} \| y_i - h_i x_i \|_2^2 + \sum_{k=1}^{2N_r} \lambda_k \left( y^{(k)} - \sum_{i=1}^{2N_t} y_i^{(k)} \right) \tag{6}
\]

Therefore, by solving the above Lagrangian function, which is a function of \( \lambda^{(k)} \) and \( y_i \), we get the following closed form expression updates for every element of \( \lambda^{(k)} \) and \( y_i^{(k)} \):

\[
\lambda^{(k)} = C \cdot \frac{1}{N_t} \left( y^{(k)} - \sum_{i=1}^{2N_t} h_i^{(k)} x_i \right), \forall k \tag{7}
\]

\[
y_i^{(k)} = h_i^{(k)} x_i + \lambda^{(k)} / 2, \forall i, k \tag{8}
\]

Note, in the update of \( \lambda^{(k)} \), we introduce the scaling factor, \( C \). We will show in Section III-C that the proposed algorithm is optimal when \( C = 1 \). However, we empirically find that when \( C = N_t \), the number of iterations for convergence drops significantly, as depicted in Section IV-A.

B. Solving the subproblem \( 5 \)

Since the objective function \( 5 \) is separable with respect to every element in the vector \( x \) and no constraint combines the elements of \( x \). The update of the \( i \)-th element in the vector \( x \) (\( x_i \)) reduces to solving the following subproblem.

\[
\min_{x_i} \sum_{k=1}^{2N_r} \| y_i^{(k)} - h_i^{(k)} x_i \|^2 \quad \text{subject to} \quad \text{and} \quad -l \leq x_i \leq l, \forall i = 1, \ldots, 2N_t \tag{9}
\]

The corresponding Lagrangian function of \( 9 \) is:

\[
\mathcal{L} = \sum_{k=1}^{2N_r} (y_i^{(k)} - h_i^{(k)} x_i)^2 + \mu_1^{(i)} (l - x_i) + \mu_2^{(i)} (l + x_i) \tag{10}
\]

Then, the following K.K.T. conditions, which are sufficient and necessary for the optimal solution to the convex optimization problem in \( 9 \) are:

\[
2x_i \sum_{k=1}^{2N_r} h_i^{(k)} - 2 \sum_{k=1}^{2N_r} y_i^{(k)} h_i^{(k)} - \mu_1^{(i)} + \mu_2^{(i)} = 0 \tag{11}
\]

\[
\mu_1^{(i)} (l - x_i) = 0, \quad \mu_2^{(i)} (l + x_i) = 0, \quad \mu_1^{(i)}, \mu_2^{(i)} \geq 0 \tag{12}
\]

To solve \( 11,12 \) for every element in the vector \( x \), among the following \{ \( \mu_1^{(i)}, \mu_2^{(i)}, x_i \) \} choices, we choose the one that minimizes \( 9 \):

\[
\mu_1^{(i)} = 0, \quad \mu_2^{(i)} \neq 0 \Rightarrow x_i = -l \tag{13}
\]

\[
\mu_1^{(i)} \neq 0, \quad \mu_2^{(i)} = 0 \Rightarrow x_i = l \tag{14}
\]

\[
\mu_1^{(i)} \neq 0, \quad \mu_2^{(i)} \neq 0 \Rightarrow x_i \in [-l, l] \tag{15}
\]

III. PROPOSED ALGORITHM

The optimization problem \( 1-3 \) is strictly and jointly convex with respect to \( x \) and \( y_i \) \( \forall i \). Moreover, there is no common constraint that combines both \( x \) and \( y_i \) \( \forall i \). Therefore, in order to efficiently solve this problem, we first decompose it into the following two subproblems:

- Given \( x \), obtain \( y_i \) \( \forall i \) by solving

\[
\arg\min_{y_i} \sum_{i=1}^{2N_t} \| y_i - h_i x_i \|_2^2 \quad \text{subject to} \quad 2 \tag{4}
\]

- Given \( y_i \) \( \forall i \), we obtain \( x \) by solving

\[
\arg\min_{x_i} \sum_{i=1}^{2N_t} \| y_i - h_i x_i \|_2^2 \quad \text{subject to} \quad 3 \tag{5}
\]

Then, we propose AltMin, an iterative algorithm that alternatively solves \( 4 \) for \( y_i \) \( \forall i \) and \( 5 \) for \( x \). Note, the respective KKT conditions of the above two subproblems form the complete set of the KKT conditions for the original problem.
Algorithm 1 Alternating Minimization Algorithm

1: Initialization
2: \( t = 0, x_i = 0 \) for all \( i \)
3: update \( \lambda_{(k)}^{(k)} \) according to equation (7)
4: update \( y_{(k)}^{(k)} \) according to equation (8)
5: \( V(t) = \sum_{i=1}^{N_t} \| y_{i} - h_i x_i \|_2^2 \)
6: \( \delta \) = convergence tolerance
7: \( T = \text{Maximum number of iterations} \)
8: 2: Alternating Minimization:
9: repeat
10: \( t \leftarrow t + 1 \)
11: solve (5) to update \( x_i \) for all \( i \)
12: update \( \lambda_{(k)}^{(k)} \) according to equation (7)
13: update \( y_{(k)}^{(k)} \) according to equation (8)
14: \( V(t) = \sum_{i=1}^{N_t} \| y_{i} - h_i x_i \|_2^2 \)
15: until \( |V(t) - V(t-1)| < \delta \) OR \( t > T \)

Note, we excluded the choice \( \mu_{1}^{(i)} \neq 0 \), and \( \mu_{2}^{(i)} \neq 0 \) since \( x_i \) cannot be equal to \(-l \) and \( l \) at the same time.

To obtain the optimal solution to the proposed optimization problem (1)-(3), AltMin solves (4) for \( y_{i}^{(k)} \) and (5) for \( x_{i}^{(k)} \). To perform the algorithm, we initially set \( x_{i} \) to 0, and solve (4) to obtain the initial \( y_{i}^{(k)} \); with the updated \( y_{i}^{(k)} \), we then solve (5) to update \( x_{i}^{(k)} \) in the next subsection, we will show that Algorithm 1 converges and the pairwise optimal \( x_{i}^{(k)} \) and \( y_{i}^{(k)} \) can be obtained, which is also the optimal solution to the problem in (1)-(3).

C. Optimality of AltMin Algorithm in solving (1)-(3)

In order to show the optimality of the iterative algorithm (when \( C = 1 \), we first give the following lemma.

Lemma 1: Given \( x_{i} \), the optimal \( y_{i}^{(k)} \) for the problem (4) is unique. Similarly, given the \( y_{i}^{(k)} \), the corresponding optimal \( x_{i} \) is unique.

Proof 1: This lemma can be obtained by verifying the strict convexity of (1)-(3) with respect to \( x_{i} \) given \( y_{i}^{(k)} \), and with respect to \( y_{i}^{(k)} \) given \( x_{i} \).

Given a pair \((y_{i}^{(k)}, x_{i})\), if \( y_{i}^{(k)} \) is the optimal solution to (4) given \( x_{i} \), and \( x_{i} \) is the optimal solution to (5) given \( y_{i}^{(k)} \), it can be said that \( y_{i}^{(k)} \) and \( x_{i} \) are pairwise optimal for (1)-(3). Furthermore, for each subproblem, the K.K.T. conditions form a subset of the K.K.T conditions of (1)-(3), giving the other primal variables, where the two subsets contain no common dual variables. Therefore, if the primal variables are pairwise optimal, the K.K.T. conditions in each corresponding subset are satisfied, all K.K.T. conditions of (1)-(3) are satisfied, i.e., the pairwise optimal solution is also the optimal \( y_{i}^{(k)}, x_{i} \) for (1)-(3).

Lemma 2: \((y_{i}^{(k)}, x_{i})\) is the optimal solution to (1)-(3) if and only if, \( y_{i}^{(k)} \) is optimal to (1)-(3) given \( x_{i} \), and \( x_{i} \) is optimal to (1)-(3) given \( y_{i}^{(k)} \).

Theorem 1: AltMin Algorithm converges, and \( y_{i}^{(k)}, x_{i} \) is optimal to (1)-(3).

Proof 2: Since \( y_{i}^{(k)}, x_{i} \) are updated by successively solving (4) and (5) in every iteration, the obtained objective value is non-increasing over iterations [9]. At the convergence point, \( y_{i}^{(k)}, x_{i} \) are jointly optimal to the subproblems, since otherwise the objective value can be further decreased in the next iteration. Then, by Lemma 2, the obtained \( (y_{i}^{(k)}, x_{i}) \) is the optimal solution to (1)-(3).

D. Complexity Analysis of AltMin

The update of \( y_{i}^{(k)} \) has a complexity of \( O(N_{t}N_{r}) \), while the update of \( x_{i} \) has a complexity of \( O(N_{t}) \). The AltMin Algorithm performs \( T \) iterations between these updates before converging to the optimal solution. Therefore, the overall complexity of AltMin in solving (1)-(3) is in the order of \( O(TN_{t}N_{r}) \), which is notably lower than \( O(N_{t}^2) \) of MMSE for large \( N_{t} \), as it is demonstrated in Section IV.

IV. Numerical Results

Simulation results for coded and uncoded uplink massive MIMO systems in a block flat fading channel is presented. We assume perfect knowledge of the channel state information at the receiver. QPSK modulations is considered for demonstration; however, the proposed algorithm can be extended for higher QAM modulations in a straightforward manner. The performance and computational complexity of the proposed algorithm based on AltMin Algorithm is compared with the linear MMSE detector.

A. Number of AltMin Iterations

In this simulation experiment we examine the number of iterations required by the AltMin algorithm such that the BER performance of both the proposed algorithm and the MMSE technique are equal, for various massive MIMO configurations. We fix convergence tolerance \( \delta \) at \( 10^{-3} \), while the maximum number of iterations \( T \) is changing in a step size of 2. We also set the initial guess of \( x_{i} \) in the AltMin algorithm to zeros, and \( C = N_{t} \) in the update of \( \lambda_{(k)} \) in (7).

Fig. 1 shows BER performance versus maximum number of iterations for \( N_{t} \times N_{r} = 16 \times 128, 32 \times 128, 64 \times 128, \) and \( 128 \times 128 \) configurations, at SNR=12 dB. It can be noticed that the larger the ratio \( N_{r} \) is, the smaller the number of iterations is required by AltMin to reach MMSE performance. For example, 8 iterations is required for \( 16 \times 128 \) configuration, while 15 iterations is required for \( 128 \times 128 \) configuration. This results show that, on the average, to reach MMSE performance the number of iterations, \( T \), is much smaller than \( N_{t} \), which keeps the complexity order of the proposed algorithm around \( O(N_{t}N_{r}) \).

B. BER performance comparison

In this subsection we present the BER performance of both the proposed algorithm and the MMSE technique with respect to SNR at various massive MIMO configuration. For each SNR value, we stop the AltMin algorithm at the iteration number at which its performance matches the MMSE performance based on the results from Fig. 1.

Fig. 2 shows that the BER of the proposed algorithm is upper bounded by that of the MMSE with the exact matrix inversion. For example, at higher ratio between \( N_{t} \) and \( N_{r} \), such as \( 16 \times 128 \) or \( 32 \times 128 \), the performance of the proposed
algorithm and MMSE are the same. As the ratio becomes closer to 1, the BER of the proposed algorithm becomes slightly lower than that of MMSE.

C. Turbo Coded BER Performance

The turbo coded BER performance of the proposed algorithm compared to MMSE is shown in Fig. 3 using coded QPSK modulation. In this simulation, all the above massive MIMO configurations are examined with rate-1/2 turbo encoder and decoder of 10 iterations. ±1 output valued vector from all detectors is fed as an input to the BCJR-based turbo decoder. In Fig. 3, AltMin based algorithm performs similar to the MMSE detector for 16 × 128, and slightly better than MMSE for 32 × 128. As the number of uplink antennas increases, the coded AltMin based algorithm outperforms coded MMSE, for example the improvement for the case of 128 × 128 at $10^{-3}$ coded BER is about 1 dB compared to only 0.2 dB improvement in the case of 64 × 128.

D. Computational Complexity Analysis

In this subsection we compare the computational complexity of the proposed algorithm with the MMSE technique in terms of the number of multiplication operations, as depicted in Table I. The comparison is based on the same SNR of 12 dB and the same BER performance. More specifically, the number of iterations of AltMin is taken based on the results of Fig. 1 at which the BER of the two techniques coincides.

From Table I it can be observed that at small $N_t$, such as 16, MMSE outperforms the proposed algorithm by approximately a factor of 3. While for a large $N_t$, such as 128, the proposed algorithm shows superior computational reduction by a factor of $\frac{1}{3}$ as compared to MMSE. Note, although our algorithm requires more computations than MMSE for small $N_t$, it does not exhibit any matrix inversion, which is more advantageous in terms of hardware implementations.

| $N_t$ | 16 | 32 | 64 | 128 |
|-------|----|----|----|-----|
| MMSE  | 0.057 | 0.311 | 2.195 | 16.97 |
| AltMin | 0.204 | 0.409 | 1.409 | 2.818 |
| AltMin/MMSE | 3.57 | 1.31 | 0.64 | 0.166 |

V. CONCLUSION

In this letter, we proposed an iterative low complexity algorithm based on Alternating Minimization approach. This algorithm is a good alternative for the MMSE technique in Massive MIMO applications where the number of BS antennas is not very large compared to the number of user equipment antennas. We show that the proposed algorithm avoids complicated matrix inversion by solving the reformulated ML problem in an iterative manner, in which each iteration performs a simple computation based on a closed form solution. The results reveal that the algorithm can provide lower computational complexity as compared to the exact matrix inverted MMSE with similar BER performance in both coded and uncoded cases.

REFERENCES

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” IEEE Communications Magazine, vol. 52, no. 2, pp. 74–80, 2014.
[2] T. Van Chien and E. Björnson, “Massive mimo communications,” in 5G Mobile Communications. Springer, 2017, pp. 77–116.
[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive mimo: Benefits and challenges,” IEEE journal of selected topics in signal processing, vol. 8, no. 5, pp. 742–758, 2014.
[4] 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,” Signal Processing Magazine, IEEE, vol. 30, no. 1, pp. 40–60, 2013.
[5] C. Zhang, X. Liang, Z. Wu, F. Wang, S. Zhang, Z. Zhang, and X. You, “On the low-complexity, hardware-friendly tridiagonal matrix inversion for correlated massive mimo systems,” arXiv preprint arXiv:1802.10444, 2018.
[6] X. Gao, L. Dai, Y. Hu, Z. Wang, and Z. Wang, “Matrix inversion-less signal detection using sor method for uplink large-scale mimo systems,” in Global Communications Conference (GLOBECOM), 2014 IEEE. IEEE, 2014, pp. 3291–3295.
[7] B. Yin, M. Wu, C. Shu, J. R. Cavallaro, and C. Dick, “Implementation trade-offs for linear detection in large-scale mimo systems,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2013, pp. 2679–2683.
[8] A. Elgabli, A. Elgharibani, A. O. Al-Abbas, and M. Bell, “Two-stage lasso admn signal detection algorithm for large scale mimo,” in Signals, Systems, and Computers, 2017 51st Asilomar Conference on. IEEE, 2017, pp. 1660–1664.
[9] Z. Wang, V. Aggarwal, and X. Wang, “Optimal energy-bandwidth allocation for energy-harvesting networks in multiuser fading channels,” IEEE Journal on Selected Areas in Communications, vol. 34, no. 5, pp. 1565–1577, 2016.