

Deep learning-aided high-precision data detection for massive MU-MIMO systems

Sen Yang¹, Zerun Li¹, Jinhui Wei¹, and Zuocheng Xing^{1,*}

¹College of Computer, National University of Defence Technology, 410073 Changsha, China

Abstract. The data detector for future wireless system needs to achieve high throughput and low bit error rate (BER) with low computational complexity. In this paper, we propose a deep neural networks (DNNs) learning aided iterative detection algorithm. We first propose a convex optimization-based method for calculating the efficient detection of iterative soft output data, and then propose a method for adjusting the iteration parameters using the powerful data driven by DNNs, which achieves fast convergence and strong robustness. The results show that the proposed method can achieve the same performance as the known algorithm at a lower computation complexity cost.

1 Introduction

The data detector used in a real multi-user multi-input and multi-output (MIMO) wireless system needs to achieve high throughput and low BER with low computational complexity [1]. In this paper, we propose a deep neural network learning aided iterative detection algorithm. We first propose a convex optimization-based method for calculating the efficient detection of iterative soft output data, and then propose a method for adjusting the iteration parameters using the powerful data driven by DNNs, which achieves fast convergence and strong robustness. The results show that the proposed method can achieve the same performance as the known algorithm at a lower cost.

MU-MIMO wireless system, which is equipping the base station (BS) with hundreds of or thousands of antennas, offers significant improvements in spectral efficiency, coverage, and link reliability compared to small-scale MIMO wireless system [2-6]. MU-MIMO technology plays a major role in next generation of wireless communication systems. However, new challenges arise with the benefits. One of problems is the computational complexity increases due to the great number of antennas.

Although various data detection algorithms have been proposed in recent years. The maximum likelihood (ML) detector is the optimal detector for detecting the minimum joint error probability of all symbols simultaneously. Whereas, its worst-case computational complexity is an NP-hard problem in practical applications. Consequently, People are interested in implementing suboptimal detection algorithms [7]. Several modified algorithms have been proposed, which offer improved computational complexity performance. Such as detectors the minimum mean squared error (MMSE) based algorithms. More advanced

* Corresponding author: zcxing@nudt.edu.cn.

algorithms are based on convex optimization approach, such as approximate message passing (AMP) and semidefinite relaxation (SDR). At present, AMP and SDR provide near optimal accuracy in many practical situations. AMP are simple and inexpensive to implement in practice, but vary in problem settings. SDR are more robust, polynomial complex, but much slower.

In the last decade, with the development of big data, optimization algorithms and stronger computing resources, machine learning has made great progress and attracted wide attention. It has been extended to many fields, such as image recognition [8] and voice processing, etc. It has also exerted a huge influence on the design of wireless communication and signal processing algorithms [9]. Leading algorithms, such as the trainable iterative shrinkage-thresholding algorithm (TISTA) and sparse versions of AMP, are improved by iteratively extending into the network and learning their best parameters.

In this paper, firstly we propose an iterative soft output data detection algorithm with high performance and low computational complexity. Then we used DNNS powerful data-driven tuning capabilities to optimize our previously designed algorithm [10]. Our algorithm is named as learned neighbourhoods-aided alternating direction method of multipliers (LN-ADMM), achieves near optimal error-rate performance in MU-MIMO systems with faster convergence speed. In our study, the convex analysis method ADMM was used to optimize the data detection problem. Further, the algorithm was expanded as a network and one layer was calculated for each iteration to improve the accuracy and convergence speed of the detection signal. The remarkable performance of this algorithm is verified by comparing with several detection algorithms.

2 Preliminaries

We consider the system model of the uplink massive access in large scale MU-MIMO systems. There is one base station (BS) equipped with N receive antennas and K potential user antennas. The mathematical expression of the standard input/output model of MIMO wireless channel is

$$\tilde{y} = \tilde{H}\tilde{x} + \tilde{n} \quad (1)$$

where $\tilde{y} \in \mathbb{C}^N$ is the received vector, $\tilde{H} \in \mathbb{C}^{2N \times 2K}$ is the MIMO channel matrix assumed to remain constant per time plot, $x \in \{\pm 1\}^K$ is an unknown vector of independent and equal probability binary symbols, and $\tilde{n} \in \mathbb{R}^N$ is the nose vector of size N with independent, zero mean Gaussian variables of variance σ^2 .

The optimal detection is essentially same as the maximum likelihood estimation rule for the Gaussian linear vector channel, can be stated as:

$$\tilde{x}^{\text{ML}} = \underset{\tilde{x} \in \{\pm 1\}^{2K}}{\text{argmin}} \|\tilde{y} - \tilde{H}\tilde{x}\|^2 \quad (2)$$

Since solving this problem is known as an NP-hard problem in general, we reformulate the problem in (2) by converting the constraint $x \in \{\pm 1\}^K$ with one-hot function to a convex set $x \in \mathcal{R}_x$, a line segment containing all points of χ on the real axis, using real valued vectors and one-hot mappings as described below:

$$y = Hx + n \quad (3)$$

$$y = \begin{bmatrix} \Re(\tilde{y}) \\ \Im(\tilde{y}) \end{bmatrix}, H = \begin{bmatrix} \Re(\tilde{H}) & -\Im(\tilde{H}) \\ \Im(\tilde{H}) & \Re(\tilde{H}) \end{bmatrix}, x = \begin{bmatrix} \Re(\tilde{x}) \\ \Im(\tilde{x}) \end{bmatrix}, n = \begin{bmatrix} \Re(\tilde{n}) \\ \Im(\tilde{n}) \end{bmatrix} \quad (4)$$

where $y \in \mathbb{R}^{2N}$ is the received vector, $H \in \mathbb{R}^{2N \times 2K}$ is the channel matrix and $x \in \mathbb{S}^{2K}$ where $\mathbb{S} = \mathfrak{R}\{\tilde{\mathfrak{S}}\}$. Here is the regularized least-squares problem:

$$x^{ML} == \underset{\tilde{x} \in \mathfrak{R}_x}{\operatorname{argmin}} \|y - Hx\|^2 \quad (5)$$

To solve the convex problem in (5) without complicated approaches, we propose a new method, and we will show our work in the third and fourth sections.

3 Data detection algorithms based on N-ADMM

We here exploit a valiant of the N-ADMM algorithm to solve (5) approximately [11]. The N-ADMM algorithm can be described by the recursive formula:

$$x^{k+1} = W^{-1}(H^T y + \beta(z^k - \lambda^k)) \quad (6)$$

$$\lambda^{k+1} = \lambda^k + \gamma(x^{k+1} - z^{k+1}) \quad (7)$$

$$z^{k+1} = \operatorname{proj}_{\mathfrak{R}_x}(\hat{x}^{k+1} + \lambda^k) \quad (8)$$

where $W = H^T H + \beta I$ is a constant Gram matrix with the parameter β fix, $k = 1, \dots, K$ and the proximal operator $\operatorname{proj}_{\mathfrak{R}_x}(\cdot)$, shown $\operatorname{proj}_{\mathfrak{R}_x}(\omega) \min\{|\omega|, (M-1)\}$, projects variables to the convex set \mathfrak{R}_x . The step size γ ranges from 0 to 1, which guarantees the fast convergence and improves the error-rate performance with larger values.

The N-ADMM algorithm consists of three computational steps for each iteration. In the gradient descent step (7), a search point moves to the opposite direction to the gradient of the objective function, i.e. In the projection step (8), soft projection based on convex optimization theory is used to project the search points onto the convex set \mathfrak{R}_x . Note that this type of nonlinear projection has been commonly used in several iterative multiuser detection algorithms such as the soft parallel interference canceller [12].

4 Data-driven MIMO detector

In this section, we extend our N-ADMM algorithm into an architecture that we call LN-ADMM. The derivation begins by noting that an efficient MIMO detector should not work with y directly,

$$H^T H x = H^T y + \beta(z^k - \lambda^k) \quad (9)$$

This means that the two main components of the architecture should be $H^T H x$ and $H^T y$. By applying standard non-linearities which are common in deep neural networks [15] and unfolding the iteration, this results in the following architecture:

$$q_k = \hat{x}_{k-1} - \delta_{1k} H^T y + \delta_{2k} H^T H x_{k-1}, v_k = \rho \left(W_{1k} \begin{bmatrix} q_k \\ z_{k-1} \end{bmatrix} + b_{1k} \right), \hat{x}_{oh,k} = W_{2k} z_k + b_{2k}, \\ \hat{x}_k = f_{oh}(\hat{x}_{oh,k}), \hat{z}_k = \beta(W_{3k} z_k + b_{3k}), \hat{x}_0 = 0, \hat{z}_0 = 0 \quad (10)$$

Thus, the final estimate is defined as \hat{x}_L , and we have given the structure of each layer of the LN-ADMM in Fig.1. Training parameters of our architecture network are:

$$\theta = \{W_{1k}, b_{1k}, W_{2k}, b_{2k}, W_{3k}, b_{3k}, \delta_{1k}, \delta_{2k}, \beta\} \quad (11)$$

Cause the gradient disappearance prevents the parameter update we need, and the derivative of the soft projection function (8) is almost zero everywhere. To solving the vanishing gradient phenomenon, we adopt an alternative loss function that considers outputs of each layer. Thus:

$$l(x_{oh}; \hat{x}_{oh}(H, y; \theta)) = \sum_{i=1}^L \log(l) \|x_{oh} - \hat{x}_{oh,i}\|^2. \quad (12)$$

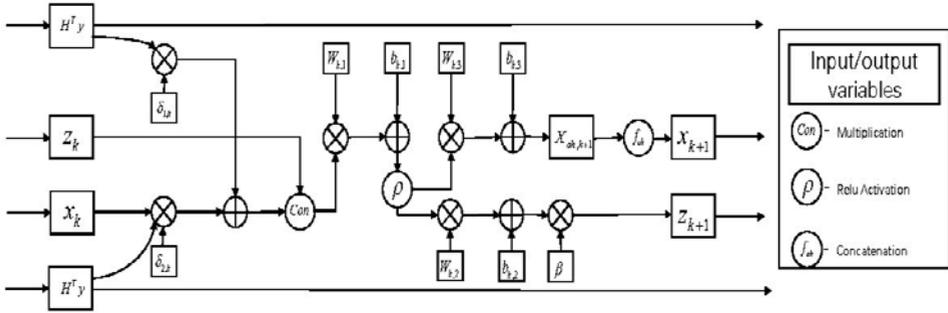


Fig. 1. A flowchart representing a single layer of LN-ADMM. The network is composed out of L layers as such where each layers' output is the previous layers' input.

5 Results and comparisons

In this section, we show the detection performance of LN-ADMM algorithm and compare it to various detectors cover a 64-QAM modulated signal.

5.1 Experimental setup

LN-ADMM detector was implemented by the Python 3.6 environment based TensorFlow library. The configuration of the computer used in the experiment is GPU NVIDIA GerForce RTX 2070 Super and AMD Ryzen 7 3700X 8-Core Processor.

We used batch training with 10240 random data samples at each iteration, and trained the network for 102400 iterations. Each sample data is generated independently. Considering that the variance of gaussian noise is unknown, the SNR is generated randomly with based on a range of $U(SNR_{min}, SNR_{max})$. In this way, we realize the network training under the wide range of SNR. The receive matrices H were randomly generated with i.i.d. $N(0,1)$ elements. Meanwhile, x and n are randomly generated based on their data characteristics.

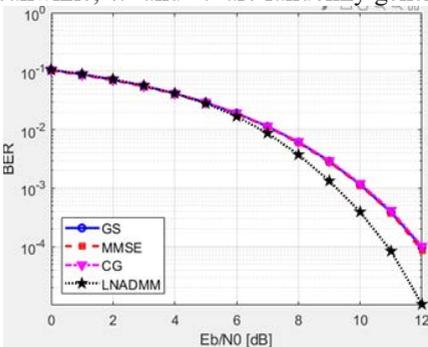


Fig. 2. 64 BS antennas, 16 users antennas and 64-QAM modulated signal.

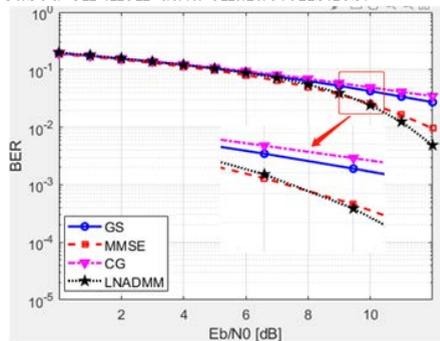


Fig. 3. 128 BS antennas, 64 users antennas and 64-QAM modulated signal.

5.2 BER performance

Fig.2 and Fig.3 illustrates the BER performance for antenna configurations of 64×16 and 128×64 . It is illustrated in Fig.2 and Fig.3 that the proposed LN-ADMM detector can achieve significant error-rate performance (SER) compared to the CG, GS and MMSE detectors. As the number of base station antennas increases, our detector is very competitive compared with other linear algorithms. Moreover, using data sparsity method brings us faster convergence speed and less iteration, equivalently, we realize the decrease of computational complexity.

6 Conclusion

In this paper, we have presented a LN-ADMM detector for large scale MU-MIMO wireless systems. On large constellations, such as 64QAM, our detector was shown to have a better BER performance than other linear MMSE detectors. Moreover, a good balance is formed between low computational complexity and precision. Therefore, for modern wireless communication systems, LN-ADMM provides a highly accurate signal detection performance that can be flexibly adjusted.

References

1. F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, *IEEE Sig. Proc. Mag.* 30, 40 (2013).
2. S. Yang and L. Hanzo, *IEEE Comm. Surv.* 17, 1941 (2015).
3. B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, *IEEE Glob.Commu.Con.* 3696 (2014).
4. T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, in *Proc. National Conf. Comm.* 1 (2012)
5. M. Nagahara, *IEEE S. P. Lett.* 22, 1575 (2015)
6. I. Khan, *IEEE Trans. Veh. Technol.* 67, 9597 (2018)
7. L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, *IEEE J. Sel. Topics S. P.* 8, 742 (2014)
8. G. E. Hinton and R. R. Salakhutdinov, *Sci.* 313,504 (2006)
9. E. Nachmani, Y. Beéry, and D. Burshtein, *54th Annu. Allert. Conf. Comm. Contr. Comput.* 341 (2016)
10. Samuel, Neev, Tzvi Diskin, and Ami Wiesel. *IEEE Trans. S. P.* 67 2554 (2019)
11. Qiu, S., Wang, Y., Chen, L., Xing, Z. *Inter.Conf.Elec. Engi.Inform.* 290 (2019)
12. D. Divsalar, M. K. Simon, and D. Raphaeli, *IEEE Trans. Commun.* 46, 258 (1998).