Improved Homptopy Technique for MIMO signal Detection

Elahe Ghaderi
Laboratory of Wireless Communication and Signal Processing (WCSP), Faculty of Electrical Engineering, Sahand University of Technology
Tabriz, Iran. Email: elaheghaderi37@gmail.com

* Masoumeh Azghani
Laboratory of Wireless Communication and Signal Processing (WCSP), Faculty of Electrical Engineering, Sahand University of Technology
Tabriz, Iran. Email: mazghani@sut.ac.ir

2022 30th International Conference on Electrical Engineering (ICEE) | 978-1-6654-8087-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICEE55646.2022.9827309

***Abstract*—In this paper, an improved Homotopy algorithm has been presented for signal detection in a MIMO system. The channel matrix is trained as a learnable matrix which results in better signal update. Furthermore, a smooth map is adopted to increase the signal detection performance. The nonlinear denoiser is used to remove the noise more efficiently. Also, the hyperbolic cosine loss function has been adopted to smoothen the results. The simulations show that the proposed method is superior to its counterparts in the bit error rate performance.**

***Index Terms*—Deep neural network, MIMO signal detection, Deep learning, Signal detection.**

I. INTRODUCTION

The massive multiple input multiple output (MIMO) systems can increase the spectral efficiency and serve many users simultaneously [1], [2] [3]. One crucial task in a massive MIMO system is the signal detection which can lead to computationally complex problems due to the large number of antennas used in the base station [1]. Therefore, developing efficient signal detection algorithms with tolerable complexity is a challenge which would be addressed in this paper. In general, the nonlinear Maximum Likelihood (ML) detector with sub-optimal performance and linear detectors such as Minimum Mean Squared Error (MMSE), Zero Forcing (ZF) are used for signal detection [4] which suffer from the complexity of matrix inversion operation. Therefore, the Neumann series expansion (NSE) detectors have been introduced to reduce the computational complexity of the matrix inversion [5]. Currently, new signal detection algorithms have been introduced in the literature based on deep learning [6]. Detection network (DetNet) is One of the neural network-based algorithms for MIMO signal detection. The DetNet is an iterative method based on gradient descent algorithm and is comparable in performance to detectors such as Approximate Message Passing (AMP) and Semi-Definite Relaxation (SDR) which takes less time to run but requires high complexity due to the network structure [6], [4] . The OAMP-Net algorithm has been designed in the case of time-varying channels and requires less running time [7]. In the MMNet algorithm [7], in the training process, the parameters are optimized for the realization of each channel matrix. In comparison with the retraining networks like OAMP-Net , MMNet does not require the inverse calculation of the channel matrix.

All the three mentioned algorithms have a denoiser for the additive Gaussian noise, but they all are complex due to the parallel training operations [7].

The OAMP-Net2 algorithm [6] has fewer trainable parameters in the network with improved convergence. Another technique based on deep learning is the Sphere Decoder (SD) [6] in which the radius of the sphere is trained in the network and gives almost the same performance as the ML detector . In all the deep learning based signal detection algorithms, network parameters are trained to achieve optimal performance [8], [6] and the networks with fewer trainable parameters are more preferred. The Homotopy algorithm is a technique based on deep learning to solve the ML problems using the gradient descent scheme [9].

In this paper, a modified Homotopy algorithm has been developed to improve the error performance at the same time of reducing the computational complexity of the network. We have added suitable trainable parameters such as a softness regulator in the signal projection. Furthermore, to eliminate the effect of the outlier data, the hyperbolic cosine function is adopted as the loss function. In addition, in order to provide a smoother estimation of the original signal, a hyperbolic tangent nonlinear denoiser is used for different noise levels. The simulation results indicate that the proposed scheme outperforms its counterparts in the bit error rate and the average loss.

The rest of the paper is organized as follows: The system model is defined in section II. The Homotopy scheme is reviewed in Section III. The proposed algorithm is investigated in Section IV. The simulation results are given in Section V and Section VI concludes the article.

II. THE SYSTEM MODEL

Suppose a MIMO system with N antennas in the transmitter and M antennas in the receiver. The transmitted symbol vector is represented by $\overline{\mathbf{x}} \in C^{N\times 1}$. The received vector, $\overline{\mathbf{y}} \in C^{M\times 1}$, is defined as [10]:

$$\overline{\mathbf{y}} = \overline{\overline{\mathbf{H}}}\overline{\mathbf{x}} + \overline{\mathbf{n}} \tag{1}$$

where $\overline{\mathbf{n}} \in C^{M\times 1} \sim CN(0, \sigma^2 \mathbf{I}_M)$ is additive white Gaussian noise vector, and $\overline{\mathbf{H}} \in C^{N\times M}$ is the channel matrix. Separating the real and imaginary parts, the real equivalent form of (1) can be written as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2}$$

and

$$\mathbf{y} = \begin{bmatrix} \Re(\bar{\mathbf{y}}) \\ \Im(\bar{\mathbf{y}}) \end{bmatrix}, \mathbf{H} = \begin{pmatrix} \Re(\bar{\mathbf{H}}) & -\Im(\bar{\mathbf{H}}) \\ \Im(\bar{\mathbf{H}}) & \Re(\bar{\mathbf{H}}) \end{pmatrix}, \quad \mathbf{x} = \begin{bmatrix} \Re(\bar{\mathbf{x}}) \\ \Im(\bar{\mathbf{x}}) \end{bmatrix}, \mathbf{n} = \begin{bmatrix} \Re(\overline{\mathbf{n}}) \\ \Im(\overline{\mathbf{n}}) \end{bmatrix} \tag{3}$$

where $\Re(.)$ and $\Im(.)$ indicates for the real and imaginary parts, respectively.

The maximum likelihood (ML) algorithm is the suboptimal receiver which can be optimal for the equiprobable transmited symbols. The ML detection for the mentioned real system model can be written as [9], [7]:

$$\hat{\mathbf{x}}_{ML} = \arg \min_{\mathbf{x}\in\{-1,1\}^N} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \tag{4}$$

The above problem is NP-hard since it requires exhaustive search among all possible vectors wirh discrete entries of $-1, 1$. One of the algorithms to solve the ML problem is the Homotopy technique which would be illustrated in the following.

III. Deep Homotopy Algorithm for Signal Detection in MIMO

The Homotopy algorithm works based on the gradient extrapolated majorization-minimization (GEMM) which is expressed in the following [9]. In order to solve the discrete minimization problem in (4) with $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$, the GEMM scheme aims to solve the following problem instead:

$$\min_{\mathbf{x}\in[-1,1]^N} f(\mathbf{x}) - \lambda\|\mathbf{x}\|_2^2 \tag{5}$$

This restatement means that By adding the second term to the cost function, the elements of the vector $\mathbf{x}$ are encouraged to get larger values. Furthermore, the constraint guarantees that $\mathbf{x}_i^2 \leq 1$ for $i = 1, \cdots, N$. Therefore, the entries of $\mathbf{x}$ would be obliged to be 1 or -1. As another step, the minimization of the cost function in (5) is conducted with the aid of the majorization minimization technique with the majorant defined as:

$$G(\mathbf{x}) = f(\mathbf{x}) - 2\lambda\langle\overline{\mathbf{x}}, \mathbf{x} - \overline{\mathbf{x}}\rangle - \lambda||\overline{\mathbf{x}}||^2 \tag{6}$$

which is derived from the following inequality:

$$\|\mathbf{x}\|_2^2 \geq \|\overline{\mathbf{x}}\|_2^2 + 2\langle\overline{\mathbf{x}}, \mathbf{x} - \overline{\mathbf{x}}\rangle - \lambda||\overline{\mathbf{x}}||^2 \tag{7}$$

and λ is a relaxation parameter. Appling the gradient extrapolated relation followed by the projected gradient descent step on $G(\mathbf{x})$, the GEMM iteations are obtained as follows:

$$\begin{aligned} &\mathbf{z}^k = \mathbf{x}^k + \alpha_k\left(\mathbf{x}^k - \mathbf{x}^{k-1}\right) \\ &\mathbf{x}^{k+1} = \Pi(\mathbf{z}^k - \beta_k\mathbf{H}^T\mathbf{H}\mathbf{z}^k + \omega_k\mathbf{H}^T\mathbf{y} + \gamma_k\mathbf{x}^k) \end{aligned} \tag{8}$$

where $2\beta_k\lambda_k = \gamma_k$ and Π is the operator which maps its arguments into the set $[-1,1]^N$, $\Pi(\mathbf{x}) = [\Pi(\mathbf{x}_1), ..., \Pi(\mathbf{x}_N)]^T$, defined as:

$$\Pi(x) = \begin{cases} -1, & x \leq -1 \\ x, & -1 \leq x \leq 1 \\ 1, & x \geq 1 \end{cases} \tag{9}$$

β_k is the step size of the k^{th} iteratrion, $\mathbf{z}$ is the extrapolated solution. In the deep learning based Homotopy algorithm [9], the parameters $\beta_k, \omega_k, \alpha_k, \gamma_k$ are obtained through a deep training step while in the original GEMM technique the parameters are updated differently. Also, the loss function of the Homotopy algorithm is the minimum mean squared error defined as follows:

$$L\left(\mathbf{x}_{original}, \mathbf{x}^k\right) = \sum_{k=1}^{K} \log(k)\left\|\mathbf{x}_{original} - \mathbf{x}^k\right\|_2^2 \tag{10}$$

where $\mathbf{x}_{original}$ is the original signal sent from the transmitter and $\mathbf{x}^k$ is the reconstruction of the k^{th} iteration.

To improve the error performance, we upgrade the Homotopy algorithm by controlling the different levels of noise to obtain a better estimation of the transmitted signal. Also, we provide a better loss function to optimize the learnable parameters in a more efficient manner.

IV. The Proposed Method

In this section, the proposed method is illustrated. The goal is to improve the Homotopy algorithm to achieve lower bit error rate.

As the first suggestion, we use the trainable channel matrix derived from the MMNet algorithm [7] so that at each time the network parameters are trained to realize the new channel matrix. Therefore, for 10,000 samples per 1000 epochs, we have a new channel matrix for 10 iterations. As a result, by optimizing the network parameters each time, we achieve a more realistic channel matrix and the network performance is improved in terms of error reduction. This process is done in the first iterative relation of (11) where $\mathbf{\Theta}_k^{(1)}$ is the $N \times M$ channel matrix to be trained and $\mathbf{x}^k$, and $\mathbf{x}^{k-1}$ are the transmitted signals in k^{th},$(k-1)^{th}$ iterations.

As the second modification, the hyperbolic tangent is used to map the signal to $+1$,-1 similar to the TPGNet algorithm [11] and ξ_k is used to control the smoothness of the projection. The hyperbolic tangent is used as a nonlinear denoiser to reduce the complexity of noise removal process in the training phase and make a soft decision of output signal. Therefore, more accurate information can be obtained from the estimated signal. Also, in order to impose the divergence-free property on the algorithm, the parameter Φ_k is continuously trained similar to the OAMPNet-2 technique [7]. This process is shown in the second relation of the (11). Thus, the modified homotopy algorithm is as:

$$\begin{aligned} &\mathbf{v}^k = \mathbf{x}^k + \alpha_k\mathbf{\Theta}_k^{(1)}(\mathbf{x}^k - \mathbf{x}^{k-1}) \\ &\mathbf{z}^{\mathbf{k}} = \Phi_k\left(\tanh\left(\xi_k\mathbf{v}^k\right)\right) \\ &\mathbf{x}^{k+1} = \Pi(\mathbf{z}^k - \beta_k\mathbf{H}^T\mathbf{H}\mathbf{z}^k + \omega_k\mathbf{H}^T\mathbf{y} + \gamma_k\mathbf{x}^k) \end{aligned} \tag{11}$$

Furthermore, the loss function of the proposed method is the hyperbolic cosine logarithm which is adopted to smoothen the overall function and reduce the error resultantly:

$$L\left(\mathbf{x}_{original}, \mathbf{x}^{k}\right) = \sum_{k=1}^{K} \log\cosh(k) \frac{\|\mathbf{x}_{original} - \mathbf{x}^{k}\|_2^2}{\|\mathbf{x}_{original}\|_2^2} \quad (12)$$

The proposed signal detection scheme is provided in Algorithm 1.

Algorithm 1 The proposed algorithm

input:
Receiver vector $\mathbf{y} \in C^{2N\times 1}$ and channel matrix $\mathbf{H} \in C^{2N\times 2M}$
output:
$\mathbf{x} \in C^{2M\times 1}$
Initialization:
$\mathbf{x}_0 \leftarrow 0^{2M\times 1}$
$\mathbf{z}_0 \leftarrow \max(\min(\mathbf{y}, 1), -1)$
The proposed algorithm:
for k=0:L-1 **do**
  $\mathbf{v}^k \leftarrow \mathbf{x}^k + \alpha_k \mathbf{\Theta}_k^{(1)}(\mathbf{x}^k - \mathbf{x}^{k-1})$
  $\mathbf{z^k} \leftarrow \Phi_k\left(\tanh\left(\xi_k \mathbf{v}^k\right)\right)$
$\mathbf{x}^{k+1} \leftarrow \Pi(\mathbf{z}^k - \beta_k \mathbf{H}^T\mathbf{H}\mathbf{z}^k + \beta_k\mathbf{H}^T\mathbf{y} + \omega_k\mathbf{H}^T\mathbf{y} + \gamma_k\mathbf{x}^k)$
end for

V. NUMERICAL RESULTS

In this section, the simulation results are reported. The Homotopy algorithm [9], the MMNet [7], and ZF [4] have been compared with the proposed method in terms of Bit Error Rate (BER). In the simulations, the number of transmitter and receiver antennas are equal and the modulation is QPSK. The MIMO channel matrix is generated from Gaussian ditribution. It is assumed that the channel matrix has been estimated beforehand and the full channel state information (CSI) is available in the receiver side. The simulations are conducted on a PC with Intel(R) Core i7-7500U CPU @ 2.70GHz 2.90 GHz in the TensorFlow environment.

For minimizing the loss function, the Adam optimizer [12] has been applied. We set $M = 40$, $N = 40$, and $L = 30$ which is the total number of layers or iterations in deep neural network in all of these algorithms. Also, in the Homotopy algorithm, the initial values are: $\beta_0 = 0.01$, $\omega_0 = 0.01$, $\alpha_0 = 0.5$, $\gamma_0 = 0.001$. For the proposed method, the initial values are $\Phi_0 = 1$, $\xi_0 = 1.0$ [9]. The training data of size $10K$ is used in the training phase.

Figure 1 shows the Bit Error Rate (BER) versus SNR curves for ZF, MMNet, Homotopy, and the proposed methods for $M = N = 40$.

We employ $10K$ iterations and the training batch size is 250. For the training phase, the SNR is set to $40dB$ and for the testing phase, the SNR has been selected in the range of $0dB$ to $40dB$. Also, the BER versus SNR curves of all the methods for $M = N = 30$ have been depicted in Figure 2.

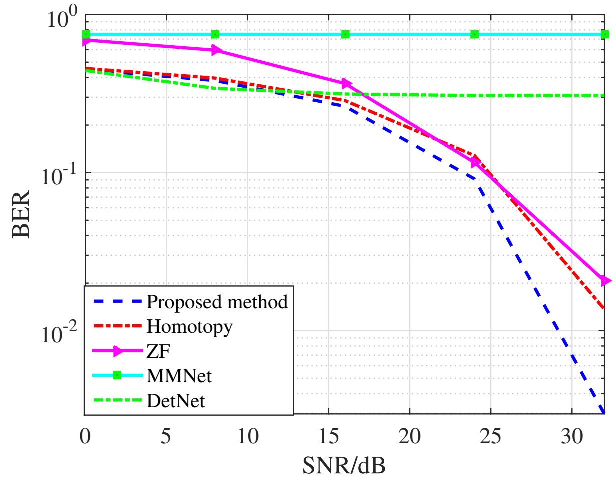

Fig. 1: The BER versus SNR curves of all the methods for $M = N = 40$.

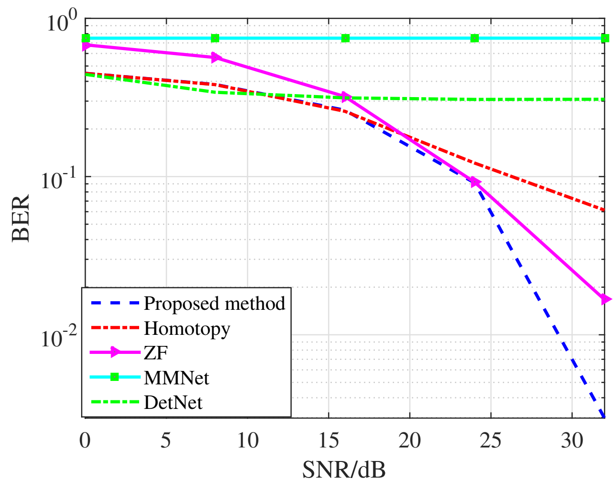

Fig. 2: The BER versus SNR curves of all the methods for $M = N = 30$.

It is observed that the proposed method has acheived lower BER compared to its rivals. Also, it can be seen that the error performance in MMNet algorithm is not good when the number of transmitter and receiver antennas is large.

Figure 3 compares the error performance of the two modes with different number of layers and batch sizes ($L = 30$, $B = 180$ and $L = 35$, $B = 250$) in the Homotopy algorithm and the proposed method.

It can be seen that by simultaneously increasing the number of layers and batch size, the error performance has been reasonably improved in both of the algorithms.

At higher SNR, by increasing the number of network layers in the proposed method from 20 to 30, the error performance is greatly improved at the expense of complexity increase.

Figure 4 shows the average run time over a batch versus SNR for various batch sizes in the test phase.

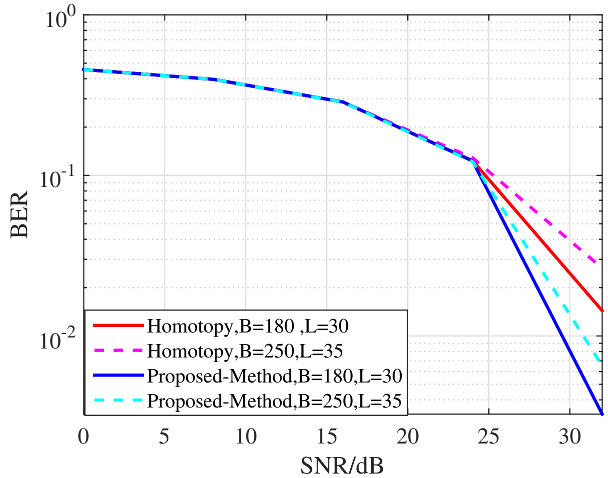

Fig. 3: The BER versus SNR curves for the proposed method and the Homotopy algorithm with different number of layers, $L = 30$,$L = 35$, and different number of batch sizes, $B = 180$, $B = 250$.

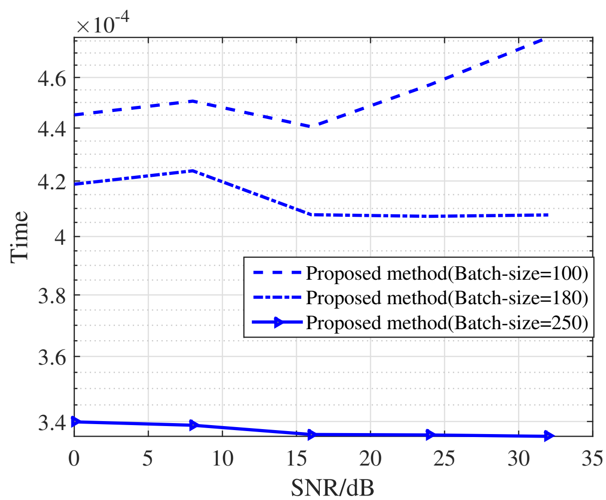

Fig. 4: The average run time over a batch versus SNR in the test phase of the proposed method.

By increasing the batch size, the average run time over a batch size in the test phase is reduced.

VI. Conclusion

In this paper, a new method based on the Homotopy algorithm is suggested. A more realistic channel training relation has been added to the Homotopy algorithm. Also, the hyperbolic tangent map has been adopted to smoothen the mapping of the signals. Some more trainable parameters are added to provide more efficient iterative relations. The loss function has also been modified which is minimized using the Adam optimizer. The simulation results indicate that the proposed method offers more accurate signal detection compared to its counterparts.

References

[1] J. Sun, Y. Zhang, J. Xue, and Z. Xu, "Learning to search for mimo detection," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7571–7584, 2020.

[2] J. Tachibana and T. Ohtsuki, "Learning and analysis of damping factor in massive mimo detection using bp algorithm with node selection," *IEEE Access*, vol. 8, pp. 96 859–96 866, 2020.

[3] M. Kim and D. Park, "Learnable mimo detection networks based on inexact admm," *IEEE Transactions on Wireless Communications*, 2020.

[4] Y. Wei, M.-M. Zhao, M. Hong, M.-J. Zhao, and M. Lei, "Learned conjugate gradient descent network for massive mimo detection," *IEEE Transactions on Signal Processing*, vol. 68, pp. 6336–6349, 2020.

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

[6] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for mimo detection," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1702–1715, 2020.

[7] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive mimo," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5635–5648, 2020.

[8] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[9] M. Shao and W.-K. Ma, "Binary mimo detection via homotopy optimization and its deep adaptation," *IEEE Transactions on Signal Processing*, vol. 69, pp. 781–796, 2020.

[10] P. Zheng, Y. Zeng, Z. Liu, and Y. Gong, "Deep learning based trainable approximate message passing for massive mimo detection," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[11] S. Takabe, M. Imanishi, T. Wadayama, R. Hayakawa, and K. Hayashi, "Trainable projected gradient detector for massive overloaded mimo channels: Data-driven tuning approach," *IEEE Access*, vol. 7, pp. 93 326–93 338, 2019.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.