Rank-Adaptive A-MMSE: Lightweight Linear Channel Estimation

TaeJun Ha, Jeonghun Park*

Yonsei Univ.

tjha@yonsei.ac.kr, *jhpark@yonsei.ac.kr

랭크 적응형 A-MMSE: 경량 선형 채널 추정

하태준, 박정훈*

연세대학교.

Abstract

In Orthogonal Frequency Division Multiplexing (OFDM) systems, accurate channel estimation is essential for reliable communication, particularly in high-mobility and frequency-selective fading environments. While Deep Neural Networks (DNNs) have demonstrated superior performance over classical methods, they typically incur high computational costs due to extensive non-linear operations. To address this, we introduce the Rank-Adaptive Attention-Aided MMSE (RA-A-MMSE) estimator. Building on the A-MMSE framework which learns an optimal linear filter via an Attention Transformer, the proposed RA-A-MMSE further incorporates a rank-adaptation mechanism to exploit the low-rank structure of the channel estimation filter. This approach allows for a flexible trade-off between estimation accuracy and computational complexity.

Ⅰ. Introduction

Accurate channel state information (CSI) is essential for reliable demodulation in Orthogonal Frequency Division Multiplexing (OFDM) systems [1]. This is particularly critical in modern wireless systems, such as 5G and 6G [2], where channels exhibit high variability and frequency selectivity due to high user mobility and millimeter-wave (mmWave) propagation. While the Least Square (LS) method is simple, it is highly vulnerable to noise. Conversely, the Minimum Mean-Squared Error (MMSE) estimator is statistically optimal but requires prior knowledge of second-order channel statistics and involves high computational complexity due to matrix inversions.

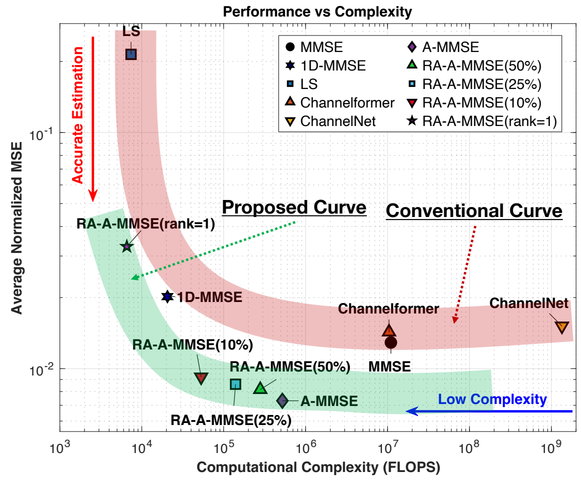
To address these limitations, Deep Neural Network (DNN)-based approaches have been introduced. However, most existing DNN methods suffer from high inference complexity due to extensive non-linear operations. In our previous work, we proposed the Attention-Aided MMSE (A-MMSE), which leverages an Attention Transformer to learn an optimal linear MMSE filter from data. A-MMSE performs estimation via a single linear operation during inference, eliminating non-linear activations.

Despite its efficiency, the A-MMSE filter matrix size $(NM \times L)$ grows with the number of subcarriers and antennas, posing memory and computational challenges for resource-constrained devices8. To overcome this, we propose the Rank-Adaptive A-MMSE (RA-A-MMSE). This method dynamically adjusts the rank of the estimation filter, significantly reducing computational overhead while maintaining high estimation accuracy.

Ⅱ. Method

1. Linear Estimation Structure of A-MMSE
   The core of the A-MMSE approach is to replace complex non-linear neural networks with a learned linear filter. The estimated channel vector $\hat{\mathbf{H}}$ is obtained through a simple matrix-vector multiplication between the learned filter matrix $\mathbf{W}_{\text{A-MMSE}} \in \mathbb{C}^{NM\times L}$ and the received pilot vector $\mathbf{Y}_p \in \mathbb{C}^{L}$:
   $$\mathrm{vec}(\hat{\mathbf{H}}) = \mathbf{W}_{\text{A-MMSE}}\mathbf{Y}_p$$
   This linear structure decouples the complex learning process (training) from the estimation process (inference).
2. Rank-Adaptive Module
   The RA-A-MMSE exploits the inherent low-rank property of the optimal MMSE filter. Instead of storing and computing the full-rank matrix $\mathbf{W}_{\text{A-MMSE}}$, we decompose it into two smaller rank-r matrices, $A \in \mathbb{C}^{NM\times r}$ and $B \in \mathbb{C}^{L\times r}$, where $r \ll \min(NM, L)$ [3].
3. Learning Mechanism
   To learn these low-rank matrices, we introduce a Rank-Adaptive (RA) module during training. The network learns trainable projection matrices $\mathbf{U}_r$ and $\mathbf{V}_r$ to constrain the A-MMSE filter:
   $$\mathbf{W}_{\text{RA-A-MMSE}} = \mathbf{W}_{\text{A-MMSE}}\,\mathbf{U}_r\,\mathbf{V}_r^{\mathrm{T}}$$

This formulation ensures that the rank of the resulting filter is bounded by r. By optimizing this structure end-to-end, the RA-A-MMSE identifies the most significant subspace for channel estimation, effectively filtering out noise and reducing the number of parameters required.



Ⅲ. Simulation Results

In this section, we analyze the computational complexity of the proposed RA-A-MMSE in terms of floating-point operations (FLOPs) required for inference and evaluate its trade-off with channel estimation accuracy (NMSE).

1. Computational Complexity Analysis (FLOPs)
   We quantified the inference complexity by counting the number of real-valued arithmetic operations. The quantitative analysis reveals that the computational cost of the proposed method is drastically lower than that of conventional approaches.
   - Conventional Methods: Deep learning-based methods such as ChannelNet [4] and Channelformer [5] require approximately 1.35×10^9 and 10×10^6 FLOPs, respectively, due to their reliance on extensive non-linear operations and complex architectural layers. Similarly, the SP-based MMSE estimator requires about 11×10^6 FLOPs due to matrix inversions.
   - Proposed RA-A-MMSE: In contrast, the RA-A-MMSE requires only linear matrix multiplications. Its complexity is proportional to the rank r, approximated as $8608r$ FLOPs (for $L = 36$). Even with a rank of $r = 12$, the RA-A-MMSE reduces the computational cost by approximately 35% compared to the full-rank A-MMSE.

2. NMSE vs. FLOPs Trade-off
   The trade-off between average NMSE and computational complexity demonstrates that the RA-A-MMSE establishes a new Pareto frontier in efficiency:
   - Drastic Reduction in Complexity: The RA-A-MMSE reduces computational complexity by approximately 95.1% compared to the average FLOPs of MMSE and Channelformer. Specifically, it requires only 1.5% of the computational resources demanded by Channelformer or MMSE (a 98.5% reduction)
   - Superior Accuracy: Despite this massive reduction in complexity, the RA-A-MMSE achieves approximately 36% to 46.3% lower average NMSE compared to the baselines.

Consequently, the RA-A-MMSE successfully pushes the boundaries of the performance-complexity trade-off, achieving a new level of estimation accuracy at a computational cost that is orders of magnitude lower than state-of-the-art DNN-based estimators.

Ⅳ. Conclusion

In this paper, we presented the RA-A-MMSE, a computationally efficient extension of the A-MMSE framework for OFDM channel estimation. By leveraging the Attention Transformer to learn a linear filter and applying a rank-adaptive mechanism, the proposed method achieves a superior trade-off between performance and complexity.

Our findings confirm that the RA-A-MMSE is capable of delivering powerful estimation performance with a fraction of the computational cost required by state-of-the-art DNN-based estimators. Consequently, the RA-A-MMSE offers a practical and scalable solution for next-generation wireless systems, enabling high-performance channel estimation even on hardware with severe resource constraints.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communication Technology Planning & Evaluation (IITP) under 6G Cloud Research and Education Open Hub (IITP-2025-RS-2024-00428780) grant funded by the Korea government (MSIT).

REFERENCES

[1]Hyunsoo Cheon and Daesik Hong, "Effect of channel estimation error in OFDM-based WLAN," in *IEEE Communications Letters*, vol. 6, no. 5, pp. 190-192, May 2002, doi: 10.1109/4234.1001660.

[2] P. Guan *et al.*, "5G Field Trials: OFDM-Based Waveforms and Mixed Numerologies," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1234-1243, June 2017, doi: 10.1109/JSAC.2017.2687718.

[3] R. A. Horn and C.R. Johnson, Matrix Analysis, 2nd ed. Cambridge University Press, 2012

[4] M. Soltani, V. Pourahmadi, A. Mirzaei and H. Sheikhzadeh, "Deep Learning-Based Channel Estimation," in *IEEE Communications Letters*, vol. 23, no. 4, pp. 652-655, April 2019, doi: 10.1109/LCOMM.2019.2898944.

[5] D. Luan and J. S. Thompson, "Channelformer: Attention Based Neural Solution for Wireless Channel Estimation and Effective Online Training," in *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6562-6577, Oct. 2023, doi: 10.1109/TWC.2023.3244484.