Convex Optimization Project Formulation - MMSE Based MIMO Channel Estimator Via Primal-Dual Optimization Mehthod with Neural Network

312510199 Jing-Hao, Li and 312510197 Zhao-Jie, Luo
Advisor: Professor Carrson C. Fung
May. 8, 2024

Introduction

Consider a communication system comprising a single transmitter and a single receiver, communicating with each other through an interference channel with Gaussian noise. To improve efficiency and reliability, both the transmitter and the receiver are equipped with multiple antennas, which it's a MIMO (Multiple Input Multiple Output) system.

The key idea behind MIMO is to exploit the spatial diversity or multiplexing of the wireless channel by transmitting data streams over multiple antennas. This allows for increased data rates, improved link reliability, and better resistance to fading and interference. By using multiple antennas at both ends of the communication link, MIMO systems can achieve spatial multiplexing, diversity, or beamforming, depending on the specific configuration and application requirements.

MIMO technology typically offers numerous advantages, but it may also introduce some challenges and issues, including heightened multipath fading and interference due to multiple antennas.

Channel estimation plays a crucial role in communication systems. It's main purpose is to determine the paths through which signals propagate from the transmitter to the receiver and to understand the signal attenuation and distortion along these paths. The accuracy of channel estimation directly impacts the quality of signal decoding at the receiver, making it an essential step in many communication systems.

In many cases, communication signals are affected by factors such as noise, multipath effects, and interference during transmission, all of which can lead to signal attenuation and distortion. By performing channel estimation, we can better understand these effects and take appropriate compensation measures to ensure the reliable transmission of signals. Channel estimation can also be used to optimize signal transmission and reception strategies, thereby improving system performance and efficiency. Therefore, channel estimation is a crucial component in modern communication systems.

We are highly interested in identifying the channel coefficients, as they play a crucial role in effectively managing the interference channel. By employing methods within the precoder like zero foring, match filter, or MMSE methods. Our aim is to mitigate the interference introduced by the channel and improve the SNR, thereby enhancing the overall performance and reliability of the communication system.

Problem setup

Assume that the communication signals are affected by additive white Gaussian noise, multipath effects, and interference during transmission, consequently, the system model can be represented in the simple linear form:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W} \tag{1}$$

Where $\mathbf{X} \in \mathbb{C}^{n_T \times T}$ represents the transmitted pilot signal, $\mathbf{W} \in \mathbb{C}^{n_R \times T}$ represents the additive white Gaussian noise, $\mathbf{Y} \in \mathbb{C}^{n_R \times T}$ represents the received signal, and $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ represents the channel coefficients that we aim to estimate. Here, n_T and n_R are the numbers of transmitted and recieved antennas, respectively. The parameter T denotes the multiple of n_T, and a onefold is commonly used.

Here we let our transmitted signal $\mathbf{X}$ to be pilot signals, which refers to a special predetermined signal transmitted to the receiver. It is typically used for estimating the channel's state or parameters in wireless communication, therefore, the structure of pilot signal often designed to be easily recognizable and processed in signal processing.

The pilot signal $\mathbf{X}$ is commonly set to be an identity matrix, this approach ensures that only one antenna transmits at any given time, thereby avoiding interference caused by other antennas transmitting simultaneously. Sometimes, we transmit the identity matrix multiple times so that each antenna can be estimated multiple times. Therefore, we concatenate the identity matrix for matrix $\mathbf{X}$. However, if the channel distribution remains constant over time, we set the multiplier to 1.

Problem Formulation

"MMSE introduction"

Under the assumption of additive Gaussian noise, the Minimum Mean Square Error (MMSE) estimator is a commonly used and effective approach. Let our estimation of vectorized channel $\mathbf{h}$ denoted as $\hat{\mathbf{h}}$, then the Bayesian mean square error of the estimator is

$$\mathbf{BMSE}(\hat{\mathbf{h}}) = \mathbb{E}_{\mathbf{y},\mathbf{h}}\left[\left\|\mathbf{h} - \hat{\mathbf{h}}\right\|_2^2\right] \tag{2}$$

Therefore, the problem of the MMSE estimator can be formulated as:

$$\hat{\mathbf{h}} = \arg\min_{\hat{\mathbf{h}}} \mathbf{BMSE}(\hat{\mathbf{h}}) \tag{3}$$

The closed form solution of this MMSE problem is $\hat{\mathbf{h}} = E_{\mathbf{h}|\mathbf{y}}\left[\mathbf{h}|\mathbf{y}\right]$. However, it requires knowledge about the conditional probability $p(\mathbf{h}|\mathbf{y})$ of $\mathbf{h}$ under observations $\mathbf{y}$, which may be unknown and/or difficult to obtain. We can't simply obtain the closed form optimal solution.

We need to consider another approach to finding the solution, the equivalent problem of equation (3) is

$$\min_{\hat{\mathbf{h}}} \mathbb{E}_{\mathbf{y},\mathbf{h}}\left[\left\|\mathbf{h} - \hat{\mathbf{h}}\right\|_2^2\right] \tag{4}$$

Proposed Approach

"primal dual method introduction"

In order to use primal dual method to iteratively approach the optimal solution, we reformulate (4) to its epigraph form:

$$\begin{aligned} \min_{t,\hat{\mathbf{h}}} \quad & t \\ \text{s.t.} \quad & \mathbb{E}_{\mathbf{y},\mathbf{h}}\left[\left\|\mathbf{h} - \hat{\mathbf{h}}\right\|_2^2\right] \leq t \end{aligned} \tag{5}$$

Then the Lagrangian function of (5) can be written as

$$\mathcal{L}\left(\hat{\mathbf{h}}, t, \lambda\right) = t + \lambda\left(\mathbb{E}_{\mathbf{y},\mathbf{h}}\left[\left\|\mathbf{h} - \hat{\mathbf{h}}\right\|_2^2\right] - t\right) \tag{6}$$

In the MMSE estimator, $\hat{\mathbf{h}}$ is a nonlinear function of y, but its exact form is unknown. Also, since $\mathbf{h} \in \mathcal{H}$, and in turn $\mathbf{h} \in \mathcal{H}$, where $\mathcal{H}$ can be interpreted as a set that contains samples of h (and $\hat{\mathbf{h}}$) from certain unknown distribution or certain (unknown/complicated) channel models. Herein, a neural network (or graph convolution neural network) may be used to parameterize $\hat{\mathbf{h}}$ so that $\hat{\mathbf{h}} = \phi(\mathbf{h}; \boldsymbol{\theta})$, with $\boldsymbol{\theta}$ denoting the parameters of the neural network. Then the Lagrangian function can also be written as

$$\mathcal{L}\left(\hat{\mathbf{h}}, t, \lambda\right) = t + \lambda\left(\mathbb{E}_{\mathbf{y},\mathbf{h}}\left[\|\mathbf{h} - \phi(\mathbf{y}; \boldsymbol{\theta})\|_2^2\right] - t\right) \tag{7}$$

The dual function can then be written as

$$g(\lambda) = \min_{t,\boldsymbol{\theta}}\ t + \lambda\left(\mathbb{E}_{\mathbf{y},\mathbf{h}}\left[\|\mathbf{h} - \phi(\mathbf{y}; \boldsymbol{\theta})\|_2^2\right] - t\right) \tag{8}$$

It is uncertain whether or not the duality gap equals zero.

However, the stationary point of $\mathcal{L}\left(\hat{\mathbf{h}}, t, \lambda\right)$ can be found via the KKT conditions by solving for the primal and dual variables alternately using gradient descent and ascent, respectively.

First, according to the stationarity property of KKT conditions: $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(\mathbf{x}^*) = \mathbf{0}$, take the gradient of Lagrangian (7) with respect to the primal and dual variable $\boldsymbol{\theta}, t, \lambda$, respectively.

$$\boldsymbol{\theta}^* = \lambda \nabla_{\boldsymbol{\theta}} \mathbb{E}\left[\|\mathbf{h} - \phi(\mathbf{y}; \boldsymbol{\theta})\|_2^2\right] \tag{9}$$

$$t^* = 1 - \lambda \tag{10}$$

$$\lambda^* = \left[\mathbb{E}\left[\|\mathbf{h} - \phi(\mathbf{y}; \boldsymbol{\theta})\|_2^2\right] - t\right]_+ \tag{11}$$

Then we can update the primal and dual variable using gradient descent or ascent iteratively

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \alpha_{\boldsymbol{\theta},k} \lambda_k \nabla_{\boldsymbol{\theta}_k} \mathbb{E}\left[\|\mathbf{h} - \phi(\mathbf{y}; \boldsymbol{\theta}_k)\|_2^2\right] \tag{12}$$

$$t_{k+1} = t_k - \alpha_{t,k}(1 - \lambda_k) \tag{13}$$

$$\lambda_{k+1} = \left[\lambda_k + \alpha_{\lambda,k}\left(\mathbb{E}\left[\|\mathbf{h} - \phi(\mathbf{y}; \boldsymbol{\theta}_{k+1})\|_2^2\right] - t_{k+1}\right)\right]_+ \tag{14}$$

where $\alpha_{\boldsymbol{\theta},k},\ \alpha_{t,k},\ \lambda_k$ are the step sizes for their respective update equations. According to [3], $\nabla_{\boldsymbol{\theta}} \mathbb{E}\left[\|\mathbf{h} - \phi(\mathbf{y}; \boldsymbol{\theta}_k)\|_2^2\right]$ can be computed using finite-difference gradients or policy gradient. It was claimed in [3] that the finite-difference method can be computational expensive especially when $\boldsymbol{\theta}$ is large, hence, it was suggested that the policy gradient method should be used. Also, [4] suggested sampling the distribution in $\mathbb{E}_{\mathbf{y},\mathbf{h}}$ so that the expectation does not have to be computed in (12).

Then we can iteratively update the primal and dual variables with policy gradient to obtain the solution.

References

[1] C. C. Fung, and D. Ivakhnenkov, ”Model-Driven Neural Network Based MIMO Channel Estimator”.

[2] M. Eisen and A. Ribeiro, “Large scale wireless power allocation with graph neural networks,” *Proc. of the 2019 IEEE 20th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1-5, 2019.

[3] M. Eisen, C. Zhang, L.F.O. Chamon, D.D. Lee and A. Ribeiro, “Learning optimal power allocations in wireless systems,” *IEEE Trans. on Signal Processing*, vol. 67(10), pp. 2775-2790, May 2019.

[4] M. Eisen and A. Ribeiro, “Optimal wireless resource allocation with random edge graph neural networks,” *IEEE Trans. on Signal Processing*, vol. 68, pp. 2977-2991, 2020.

[5] N. NaderiAlizadeh, M. Eisen and A. Ribeiro, “State-Augmented learnable algorithms for resource management in wireless networks,” *IEEE Trans. on Signal Processing*, vol. 70, pp. 5898-5912, Dec. 2022.

[6] N. NaderiAlizadeh, M. Eisen and A. Ribeiro, “Learning resilient radio resource management policies with graph neural networks,” *IEEE Trans. on Signal Processing*, volo. 71, pp. 995-1009, Mar. 2023.