Uncertainty Propagation in Neural Network Enabled Multi-Channel Optimisation

Chen Li, Schyler C. Sun, Saba Al-Rubaye, Antonios Tsourdous, Weisi Guo IEEE Senior Member

Abstract—Multi-channel optimisation relies on accurate channel state information (CSI) estimation. Error distributions in CSI can propagate through optimisation algorithms to cause undesirable uncertainty in the solution space. The transformation of uncertainty distributions differs between classic heuristic and Neural Network (NN) algorithms. Here, we investigate how CSI uncertainty transforms from an additive Gaussian error in CSI into different power allocation distributions in a multi-channel system. We offer theoretical insight into the uncertainty propagation for both Water-filling (WF) power allocation in comparison to diverse NN algorithms. We use the Kullback–Leibler divergence to quantify uncertainty deviation from the trusted WF algorithm and offer some insight into the role of NN structure and activation functions on the uncertainty divergence, where we found that the activation function choice is more important than the size of the neural network.

Index Terms—machine learning; deep learning; XAI; wireless; mis-estimation, and map to an output $y \in \mathbb{R}^{n \times 1}$ (e.g. power allocation) via a model $f(\cdot)$.

In classic WF, the Lagrangian optimisation produces an iterative solution; and in NNs an approximate non-linear mapping can achieve effective power allocation without iterative search for $\lambda$ (WF level). Direct probability analysis or Bayesian inference can be used to understand the brittleness of classic and heuristic algorithms [12]. Other analytical methods include polynomial chaos expansion, which are more suited to dynamical systems [13]. Currently, black-box NNs cannot explain the essential mapping it performs. There are also legal requirements (e.g. GDPR) for AI to explain its reasoning. As such, there is the need to develop a range of explainable AI (XAI) solutions that attempt to quantify NN mappings [14], [15]. These XAI techniques range from visualising key hidden layer features to localised linear models (LIME) [16], [17].

I. INTRODUCTION

Multi-channel optimisation of radio resources is crucial to current 4G-5G systems and beyond. Traditional optimisation relies on heuristic algorithms which are often formulated as either convex optimisation (e.g. Lagrangian) or non-convex problems (e.g. Genetic algorithm, Mean-field games, Markov Decision Processes, reinforcement learning...etc.). As the scale of the complexity increases, neural networks (NN) [1]-[4] have been proposed to automate and accelerate the mapping between inputs (e.g. CSI, user demand) and output solutions (e.g. transmit power allocation) [5], [6]. One open challenge is the propagation of uncertainty from input to output via an optimisation algorithm. Classic uncertainty quantification (UQ) techniques such as Polynomial Chaos Expansion (PCE) cannot be readily applied due to the complex nature of the algorithms.

A. Uncertainty Propagation in Decision Modules

Here, we start by studying how uncertainty in CSI can propagate through a classic MIMO Water-filling (WF) power allocation algorithm (IEEE 802.xx series, OFDM systems) versus its contemporary NN accelerated versions [7], which offer equivalent accelerated real-time solutions [8], [9]. Other power allocation employ DRL, which adds a learning agent [10], [11]. The essential UQ problem is to quantify the distribution over the output $y$ for: $y = f(x + n; \lambda)$, where inputs $x \in \mathbb{R}^{n \times 1}$ (e.g. channel gains) have a noise $n$ due to

\[
p(n) = f[h(n); \lambda] = \left( \frac{1}{\lambda} - \frac{N_0}{|h(n)|^2} \right)^+, \tag{1}
\]

B. Novelty

In this seminal initial results paper on robustness of NNs for wireless power control, we outline statistical results on how different NN architectures and activation functions transform CSI uncertainty into power allocation solutions. Our novelties in this paper include: (1) deriving a theoretical uncertainty transformation for WF power allocation, (2) provide the statistical results for uncertainty transformation for a range of NNs and measure their KL divergence from the theoretical distribution.

II. SYSTEM SETUP

A. WF Power Allocation

We consider a classical wireless parallel channel power allocation problem comprised of $N$ channels with independent Rayleigh fading characteristics. WF power allocation is well-established and we will not detail it here. Sufficient to say, under the Shannon capacity assumption and a total power budget, the solution form for power in channel $n \in N$ is of:

All authors are with Cranfield University. Acknowledge EC H2020 grant 778305.

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
**B. NN Power Allocation**

We reduce the WF heuristic search time by implementing a multi-layer NN with a number of neurons per layer. The input are the channel states and the output is the Lagrangian multiplier $\lambda$. The Lagrangian parameter in turn gives the power allocation output ($\lambda \rightarrow p(n)$). The implementation parameters are given in Table I. The training data is based on the conventional iterative WF solution, and the training results for different NN architectures is given in Fig. 2.

**C. CSI Uncertainty and Theoretical Transformation**

In both the WF and NN cases, we set up Gaussian distributed additive CSI uncertainty in just one of the multiple channels and examine its impact on the power allocation solution for all the channels. We treat WF as a trusted and reliable benchmark - familiar to engineers and used widely in society.

---

**TABLE I**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>2.4GHz 3GPP Micro, Rayleigh</td>
</tr>
<tr>
<td>Wireless System</td>
<td>5 OFDMA channels, 40W budget</td>
</tr>
<tr>
<td>Iterative WF Monte-Carlo Loops</td>
<td>$s^4$ Gaussian in Channel 1</td>
</tr>
<tr>
<td>CSI Noise</td>
<td>$s^5$</td>
</tr>
<tr>
<td>Neurons per Layer</td>
<td>5-15</td>
</tr>
<tr>
<td>Activation Function</td>
<td>Sigmoid ($s$) or ReLU ($p$)</td>
</tr>
<tr>
<td>Training</td>
<td>Random division, $\leq 1e3$ epochs</td>
</tr>
</tbody>
</table>
The theoretical transformation via WF solution in Eq.(1) can be found as follows. We perturb one of the multiple channels with a Gaussian noise due to CSI mis-estimation: channel fading is $x \sim N(\mu, \sigma^2)$, where $\mu$ is the accurate channel fading value and $\sigma^2$ is the input uncertainty variance due to CSI mis-estimation. We then find the uncertainty cascade for the perturbed channel under the split conditions of:

1) $x < \sqrt{N_0/\lambda}$, $y = 0$,
2) $x \geq \sqrt{N_0/\lambda}$, $y = \frac{1}{\lambda} \left( 1 - N_0 x^2 \right)$.

Using the standard probability Jacobian transformation approach, we arrive at:

$$P(y = 0) = P(x_1 < \sqrt{N_0/\lambda})$$  
$$f_Y(y) = \frac{(b-a)^{3/2}}{2b\sigma\sqrt{2\pi}} \exp\left( -\frac{(b-a-y-\mu)^2}{2\sigma^2} \right)$$

where $a, b$ are numerical values based on $\lambda, N_0$ for a given set of fading channels.

Given WF’s theoretical uncertainty in Eq.(2) and that we trust its transparent nature, we now compare how the NN’s uncertainty differs.

III. UNCERTAINTY PROPAGATION RESULTS

A. Power Allocation Uncertainty

We demonstrate that the PDF and CDF results for uncertainty transformation from CSI mis-estimation into power solutions’ uncertainty for both WF and NNs in Fig. 1. We can observe that compared to the theoretical WF distribution, the NN solutions have two attributes. If no post-hoc $(\cdot)^+$ operator is used, then the NN produces negative power allocation solutions of diverse nature. If the operator is used (as is in WF), then solutions are similar (see Fig. 1d-ii/-iii). In the other channels of Fig. 1c, we can see the power distribution from both WF and an example NN. It is clear that the impact on other channels can also be large and should be investigated in future.

**TABLE II**

<table>
<thead>
<tr>
<th>NN Module (Xz-Ys)</th>
<th>KL Div.</th>
<th>NN Module (Xz)</th>
<th>KL Div.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN 10s-10s</td>
<td>0.009</td>
<td>NN 15s</td>
<td>0.013</td>
</tr>
<tr>
<td>NN 10p-10s</td>
<td>0.031</td>
<td>NN 15p</td>
<td>0.012</td>
</tr>
<tr>
<td>NN 5p-5s</td>
<td>0.028</td>
<td>NN 10s</td>
<td>0.01</td>
</tr>
<tr>
<td>NN 5s-5s</td>
<td>0.037</td>
<td>NN 10p</td>
<td>0.13</td>
</tr>
</tbody>
</table>

B. KL Divergence in Different NN Architectures

Using WF heuristic as a benchmark (novel theoretical bound derived in Eq. (2)), we examine the uncertainty transformations’ KL divergence in Table II. The NN architectures experimented are labeled as Xz-Ys, where X stands for the number of neurons in the first layer, Y stands for the number of neurons in the second layer (if present), and z stands for the activation function in the first layer (s is sigmoid, y is ReLU), and the second layer activation is always sigmoid (s).

We show the results in descending architecture complexity from a multi-layer 10-10 to a single layer 10 structure. In general, the more sophisticated architectures offer a lower KL divergence (as expected), and we also see that the sigmoid function (s) offers a vastly superior performance in all cases (10z-10s, 15s, 10s) and a similar performance in the (5z-5s) case. We can conclude that indeed more sophisticated activation functions offer a more robust performance (by being closer to the original WF solution), whilst accelerating the algorithm speed by avoiding the search for $\lambda$. 
Multi-channel optimisation relies on accurate channel state information (CSI) estimation. Error distributions in CSI can propagate through optimisation algorithms to cause undesirable uncertainty in the solution space. The transformation of uncertainty distributions differs between classic heuristic and Neural Network (NN) algorithms. Here, we examined uncertainty propagation in both classic heuristic Water-Filling (WF) power allocation and different Neural Network (NN) accelerated implementations. We derived a theoretical uncertainty distribution for WF and used KL divergence to measure the difference between different NN architectures against WF. Generally the activation function choice is more important than the size of the neural network, which will inform the design priority of future NNs.

Our future work will aim to quantify the reason for this finding via other supporting XAI features such as LIME [14].

REFERENCES

2020-06-30

Uncertainty propagation in neural network enabled multi-channel optimisation

Li, Chen

IEEE


https://doi.org/10.1109/VTC2020-Spring48590.2020.9128702

Downloaded from Cranfield Library Services E-Repository