DUAL DOMAIN LEARNING OF OPTIMAL RESOURCE ALLOCATIONS IN WIRELESS SYSTEMS

Mark Eisen[*] *Clark Zhang*[*] *Luiz F. O. Chamon*[*]
Daniel D. Lee[†] *Alejandro Ribeiro*[*]

[*]Department of Electrical and Systems Engineering, University of Pennsylvania
[†]Department of Electrical and Computer Engineering, Cornell Tech

ABSTRACT

We consider the problem of finding optimal resource allocations subject to system constraints in a generic class of problems in wireless communications. These problems are inherently challenging due to functional optimization and potential non-convexities. However, these problems can be observed to take the form of a regression problem, although one in which the statistical loss function appears as a constraint. This motivates the use of machine learning model parameterizations. To apply gradient-based solution algorithms that do not require model knowledge, we convert the constrained optimization problem to an unconstrained one using Lagrangian duality. Despite the non-convexity in the problem, we formally show that the sub-optimality of the dual domain problem is small when the learning parameterization is sufficiently dense. We then present a primal-dual learning algorithm that looks for solutions to the dual problem using model-free gradient estimates. In a numerical simulation, we demonstrate the near-optimality of the proposed model-free algorithm using a neural network parametrization for a capacity maximization problem.

Index Terms— wireless communications, leaning, resource allocation, neural networks, duality gap

1. INTRODUCTION

In the design of optimal wireless systems, the goal is to allocate available resources—such as power, bandwidth, data rates, etc.—to meet system requirements and optimize with respect to some utility. Because the wireless channel is subject to random fading, we are generally interested in optimizing with respect to long term or average performance. This is due to the fact that that instantaneous channel conditions vary too quickly and unpredictably for users to observe instantaneous performance, and thus average performance is a more useful measure. Problems of this form range from the simple power allocation in wireless fading channels, to the optimization of frequency division multiplexing [1], beamforming [2, 3], random access [4, 5], and wireless control systems [6–8].

While these problems may be easily formulated as a mathematical optimization problem, finding solutions to these problems is far from straightforward. Because we search for resource allocation *functions*, the dimensionality of these problems is a significant challenge. In addition, performance functions of interest are often, in practice, non-convex. However, much work in this area has focused on solving this problem in the dual domain. A key property that enables this solution is the lack of duality gap, which allows dual operation without loss of optimality. This has been shown to hold for a generic class of wireless resource allocation problems under mild technical conditions despite the presence of a non-convex constraint [9]. This motivates methods that seek to find solutions in the dual domain, which make the problem easier to solve, though not necessarily completely tractable without resorting to heuristics [6–8, 10–14]. All such methods will additionally require model knowledge, which is not always available in practice.

The fundamental challenge in finding solutions to resource allocation problems has motivated the use of machine learning techniques. In some cases, it may be possible to obtain a finite collection of optimal solutions given a set of sampled channel conditions, which allows for the application of traditional supervised learning methods to find a more generic solution. Such cases include those in which a training set can be constructed using exact solutions [15, 16] or constructed using heuristic-based approximations [17–19]. An alternative to supervised learning is instead to train the learning model directly with respect to the utility of interest, as is often done in, e.g., reinforcement learning problems [20]. This learning approach has been taken in simpler *unconstrained* problems in wireless optimization [21–23], where the reinforcement learning solution framework can be applied directly. However, in the more generic class of wireless problems that are constrained—due to balancing of capacity, power consumption, channel access, and interference—there is no clear way in which to solve such problems without the need of a training set.

In this paper, we consider a learning-based approach in which we parameterize the resource allocation function directly within the constrained optimization problem (Section 2). To apply model-free gradient-based machine learning algorithms to this problem, we must convert the constrained problem to an unconstrained problem. We propose the use of Lagrangian duality to derive an unconstrained problem in the dual domain. We demonstrate formally that, despite non-convexities, the duality gap of learning problems in wireless optimization is small if the learning parametrization is nearly universal (Section 3). This set of nearly universal parameterizations include reproducing kernel Hilbert spaces and, most famously, deep neural networks. We introduce a model-free learning algorithm in which gradients are estimated by sampling the model functions and wireless channel (Section 3.1). We conclude with numerical experiments on a standard wireless resource allocation problem, in which we demonstrate the near-optimal performance of the proposed dual learning approach using deep neural networks (Section 4).

2. OPTIMAL RESOURCE ALLOCATION

Let $\mathbf{h} \in \mathcal{H} \subseteq \mathbb{R}_+^n$ be a random variable representing a collection of n stationary wireless fading channels drawn according to the probability distribution $m(\mathbf{h})$. We define resource allocation function $\mathbf{p}(\mathbf{h}) \in \mathbb{R}^m$ and a performance function $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^u$. Because channel states $\mathbf{h}$ vary quickly and randomly, we are interested in the performance over an ergodic average $\mathbf{x} = \mathbb{E}\big[\mathbf{f}\big(\mathbf{p}(\mathbf{h}), \mathbf{h}\big)\big] \in \mathbb{R}^u$. Optimal wireless design involves finding the instantaneous resource allocation $\mathbf{p}(\mathbf{h})$ that optimizes a utility $g_0 : \mathbb{R}^u \to \mathbb{R}$ of the average performance $\mathbf{x}$, subject to

Supported by ARL DCIST CRA W911NF-17-2-0181 and Intel Science and Technology Center for Wireless Autonomous Systems.

constraints represented by the vector function $\mathbf{g} : \mathbb{R}^u \to \mathbb{R}^r$. We further consider the convex sets $\mathcal{X} \subseteq \mathbb{R}^u$ and $\mathcal{P} \subseteq \mathcal{M}$, where $\mathcal{M}$ is the set of functions integrable with respect to $m(\mathbf{h})$, that define feasible points for $\mathbf{x}$ and $\mathbf{p}(\mathbf{h})$, respectively. The optimal resource allocation problem in wireless communication systems is then formulated as

$$\begin{aligned} P^* := \max_{\mathbf{p}(\mathbf{h}),\mathbf{x}} \; & g_0(\mathbf{x}), \\ \text{s.t.} \quad & \mathbf{x} \leq \mathbb{E}\big[\mathbf{f}\big(\mathbf{p}(\mathbf{h}),\mathbf{h}\big)\big], \\ & \mathbf{g}(\mathbf{x}) \geq \mathbf{0},\ \mathbf{x} \in \mathcal{X},\ \mathbf{p} \in \mathcal{P}. \end{aligned} \tag{1}$$

Observe we relax the definition $\mathbf{x} = \mathbb{E}\big[\mathbf{f}\big(\mathbf{p}(\mathbf{h}),\mathbf{h}\big)\big]$ to an inequality constraint in (1). The functions $g_0(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are assumed to be concave and the set $\mathcal{X}$ is assumed to be convex. However, the function $\mathbf{f}(\cdot,\mathbf{h})$ is not assumed convex or concave and the set $\mathcal{P}$ is not assumed to be convex either. The performance functions often appearing in wireless systems are indeed often non-convex, such as a capacity function over an interference channel [9]. As is common in practical situations, we do not have complete knowledge of the functions $g_0, \mathbf{g}$, and $\mathbf{f}$, but can only sample that at given operating points.

Although we may use the formulation in (1) to characterize the optimal resource allocation in various wireless communication problems, the optimization problem itself is generally very challenging to solve. Because we are optimizing over a set of infinite dimensional variables $\mathbf{p}(\mathbf{h})$ and over a non-convex feasible set, finding optimal solutions may be very hard if not intractable. In select cases of the optimal resource allocation problem, heuristic methods have been developed that may find approximate solutions. Alternately, there are a number of optimization methods that exploit that fact that, despite the non-convexity of the constraint in (1), it is known to exhibit null duality gap when the channel distributions $m(\mathbf{h})$ are non-atomic [9]. This property allows for solutions to be found in the Lagrangian dual problem. Both the heuristic and standard dual methods, however, are only available for certain instances of (1) and, more importantly, require either significant domain or model knowledge to develop.

In this work, we make a key observation about the structure of the optimal resource allocation problem, namely that it takes the form of a learning problem. Indeed, the instantaneous allocation function $\mathbf{p}(\mathbf{h})$ can be viewed as a regression function which predicts optimal resource allocations $\mathbf{p}$ given a set of data $\mathbf{h}$. Because we optimize over the *expected* performance over the statistics of the data—in this case the fading channel conditions—this takes the form of a regression problem common in machine learning. A fundamental difference, however, is that the regression loss function $\mathbb{E}\big[\mathbf{f}\big(\mathbf{p}(\mathbf{h}),\mathbf{h}\big)\big]$ appears as a constraint in (1), rather than the objective as is commonly seen in traditional learning problems.

This insight motivates a learning based solution to (1), in particular one in which we replace the generic allocation function $\mathbf{p}(\mathbf{h})$ with a given model $\boldsymbol{\phi}$ parameterized by $\boldsymbol{\theta} \in \mathbb{R}^q$, i.e.

$$\mathbf{p}(\mathbf{h}) = \boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta}). \tag{2}$$

If we now define the set $\Theta := \{\boldsymbol{\theta} \mid \boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta}) \in \mathcal{P}\}$, the optimization problem in (1) becomes one in which the optimization is over $\mathbf{x}$ and $\boldsymbol{\theta}$

$$\begin{aligned} P^*_{\boldsymbol{\phi}} := \max_{\boldsymbol{\theta},\mathbf{x}} \; & g_0(\mathbf{x}), \\ \text{s.t.} \quad & \mathbf{x} \leq \mathbb{E}\big[\mathbf{f}\big(\boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta}),\mathbf{h}\big)\big], \\ & \mathbf{g}(\mathbf{x}) \geq \mathbf{0},\ \mathbf{x} \in \mathcal{X},\ \boldsymbol{\theta} \in \Theta. \end{aligned} \tag{3}$$

In (3), we eliminate the challenge of functional learning by optimizing instead over a finite dimensional variable $\boldsymbol{\theta}$ (in addition to $\mathbf{x}$). Because we restrict our search space to a smaller set of functions that can be parameterized by $\boldsymbol{\phi}$, $P^*_{\boldsymbol{\phi}}$ will by construction by sub-optimal with respect to P^*. Naturally, we should expect that the magnitude of this sub-optimality be related to the degree to which functions of the form $\boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta})$ can approximate the optimal function $\mathbf{p}^*(\mathbf{h})$. In this work, we focus our attention on a widely-used class of parameterizations we define as *near-universal*, which are able to model any function in $\mathcal{P}$ to within a stated accuracy. We present this formally in the following definition.

Definition 1 *A parameterization $\boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta})$ is an ϵ-universal parameterization of functions in $\mathcal{P}$ if, for some $\epsilon > 0$, there exists for any $\mathbf{p} \in \mathcal{P}$ a parameter $\boldsymbol{\theta} \in \Theta$ such that*

$$\mathbb{E}\left\|\mathbf{p}(\mathbf{h}) - \boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta})\right\|_\infty \leq \epsilon. \tag{4}$$

A number of popular machine learning models are known to exhibit the universality property in Definition 1, such as radial basis function networks (RBFNs) [24], reproducing kernel Hilbert spaces (RKHS) [25], and the widely explored deep neural networks (DNNs) [26].

Despite the dimensionality reduction we achieve through the learning parameterization in (3), the problem remains unsolvable in its current form due the existence of constraints. Standard machine learning problems are unconstrained problems, and so the algorithms used to solve them cannot be applied directly to the more complex structure of the machine learning problem that appears in the wireless resource allocation problem. A naive approach to remove the constraints is to add additional terms to the objective function that penalize violation of the constraints. As the resulting problem will seek a solution to strikes some balance between maximizing the utility $\mathbf{g}(\mathbf{x})$ and avoiding constraint violation, this generic penalty approach will not accurately characterize the optimal solution $P^*_{\boldsymbol{\phi}}$. To address this issue, we consider the techniques of Lagrangian duality to derive an appropriate unconstrained problem.

3. DUALITY OF LEARNING PROBLEM

Lagrangian duality provides a common approach towards converting a constrained optimization problem into an unconstrained optimization problem. To derive the so-called dual problem, introduce the nonnegative dual variables $\boldsymbol{\lambda} \in \mathbb{R}^p_+$ and $\boldsymbol{\mu} \in \mathbb{R}^r_+$, respectively associated with the constraints $\mathbf{x} \leq \mathbb{E}\big[\mathbf{f}\big(\boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta}),\mathbf{h}\big)\big]$ and $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. We first form a Lagrangian function of (3) as sum of objective and constraint values weighted by their respective dual variables

$$\begin{aligned} \mathcal{L}_{\boldsymbol{\phi}}(\boldsymbol{\theta},\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) := \; & g_0(\mathbf{x}) + \boldsymbol{\mu}^T\mathbf{g}(\mathbf{x}) \\ & + \boldsymbol{\lambda}^T\Big(\mathbb{E}\big[\mathbf{f}\big(\boldsymbol{\phi}(\mathbf{h},\boldsymbol{\theta}),\mathbf{h}\big)\big] - \mathbf{x}\Big). \end{aligned} \tag{5}$$

The dual function $D_{\boldsymbol{\phi}}(\boldsymbol{\lambda},\boldsymbol{\mu})$ is then a function of the dual variables and is defined as

$$D_{\boldsymbol{\phi}}(\boldsymbol{\lambda},\boldsymbol{\mu}) := \max_{\boldsymbol{\theta}\in\Theta,\mathbf{x}\in\mathcal{X}} \mathcal{L}_{\boldsymbol{\phi}}(\boldsymbol{\theta},\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}). \tag{6}$$

It is widely known that if $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{\mu} \geq \mathbf{0}$ we have $D_{\boldsymbol{\phi}}(\boldsymbol{\lambda},\boldsymbol{\mu}) \geq P^*_{\boldsymbol{\phi}}$– see, e.g. [27]. This motivates definition of the dual problem in which we search for the dual variables that minimize $D_{\boldsymbol{\phi}}(\boldsymbol{\lambda},\boldsymbol{\mu})$, i.e.

$$D^*_{\boldsymbol{\phi}} := \min_{\boldsymbol{\lambda},\boldsymbol{\mu}\geq\mathbf{0}} D_{\boldsymbol{\phi}}(\boldsymbol{\lambda},\boldsymbol{\mu}). \tag{7}$$

The resulting dual problem is then an unconstrained optimization problem of finite dimensional variables, or at least only constrained by simple set constraints. This resulting problem in (7) can then be viewed as a standard unconstrained statistical learning problem of both the dual variables $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ and, indirectly, the primal variables $\boldsymbol{\theta}$ and $\mathbf{x}$. Thus, standard gradient based machine learning algorithms may be used directly on (7). Among such methods include well known *model-free* methods which search for solutions to (7) without requiring knowledge of the functions present in (3). We develop such a learning algorithm in Section 3.1 of this paper. Before proceeding, however, we look at the central question of how well the dual problem in (7) represents the constrained problem on interest in (1).

It is known that convex optimization problems exhibit an important quality known as strong duality, which refers to the fact that the minimum dual value is equivalent to the optimal primal value [27]. However, this property is far from given in the optimal resource allocation problem in (3) due to the non-convexities. In order to properly motivate learning in the dual domain for this problem, it is necessary to characterize this gap in optimalities.

In this paper, we demonstrate a key result in which, despite the non-convexities in (3), the structure of the problem and the choice of parametrization function $\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta})$ induces a dual problem whose optimal is close to that of (1). In particular, this result builds off of the strong duality property the original problem in (1) and the nature of near universal functions in the sense of Definition 1. In proving this result we need to introduce some restrictions to the problem formulation that we state as assumptions next.

Assumption 1 *The probability distribution $m(\mathbf{h})$ is nonatomic in $\mathcal{H}$. I.e., for any set $\mathcal{E} \subseteq \mathcal{H}$ of nonzero probability there exists a nonzero probability strict subset $\mathcal{E}' \subset \mathcal{E}$ of lower probability, $0 < \mathbb{E}_{\mathbf{h}}(\mathbb{I}(\mathcal{E}')) < \mathbb{E}_{\mathbf{h}}(\mathbb{I}(\mathcal{E}))$.*

Assumption 2 *Slater's condition hold for the unparameterized problem in* (1) *and for the parametrized problem in* (3)*. In particular, there exists variables $\mathbf{x}_0$ and $\mathbf{p}_0(\mathbf{h})$ and a strictly positive scalar constant $s > 0$ such that*

$$\mathbb{E}\big[\mathbf{f}\big(\mathbf{p}_0(\mathbf{h}), \mathbf{h}\big)\big] - \mathbf{x}_0 \geq s\mathbf{1}. \tag{8}$$

Assumption 3 *The objective utility function $g_0(\mathbf{x})$ is monotonically non-decreasing in each component. I.e., for any $\mathbf{x} \leq \mathbf{x}'$ it holds $g_0(\mathbf{x}) \leq g_0(\mathbf{x}')$.*

Assumption 4 *The expected performance function $\mathbb{E}\left[\mathbf{f}\big(\mathbf{p}(\mathbf{h}), \mathbf{h}\big)\right]$ is expectation-wise Lipschitz on $\mathbf{p}(\mathbf{h})$ for all fading realizations $\mathbf{h}$. Specifically, for any pair of resource allocations $\mathbf{p}_1(\mathbf{h}) \in \mathcal{P}$ and $\mathbf{p}_2(\mathbf{h}) \in \mathcal{P}$ there is a constant L such that*

$$\mathbb{E}\|\mathbf{f}(\mathbf{p}_1(\mathbf{h}), \mathbf{h}) - \mathbf{f}(\mathbf{p}_2(\mathbf{h}), \mathbf{h})\|_\infty \leq L\mathbb{E}\|\mathbf{p}_1(\mathbf{h}) - \mathbf{p}_2(\mathbf{h})\|_1. \tag{9}$$

Assumption 1 states that there are no points of strictly positive probability in the distributions $m(\mathbf{h})$, or in other words that the fading state $\mathbf{h}$ take values in a dense set with a non-atomic probability density. Assumption 2 simply states that there exist points that are strictly feasible in (1), which is a reasonable assumption in practice. Assumption 3 is restricts utilities $g_0(\mathbf{x})$ to those that are monotonically non-decreasing. Assumption 4 is a Lipschitz continuity restriction on each of the dimensions of the *expectation* of the constraint function $\mathbf{f}$, which we point out this is weaker than general Lipschitz continuity and satisfied by many wireless resource allocation problems, including those with discrete allocation variables.

Assumptions 1-3 collectively imply that the duality gap of the original unparameterized problem in (1) is known to be null – see [9] for details on this result. Given further satisfaction of Assumption 4 and using a parametrization that is nearly universal in the sense of Definition 1, such as DNNs, we show that the parametrization gap, or sub-optimality, $|D^*_{\boldsymbol{\phi}} - P^*|$ that results from the transformation of (1) to the parameterized dual problem in (7) is small as we formally state next.

Theorem 1 *Consider the parameterized resource allocation problem in* (3) *and its Lagrangian dual in* (7) *in which the parametrization $\boldsymbol{\phi}$ is ϵ-universal in the sense of Definition 1. If Assumptions 1 - 4 hold, then the dual value $D^*_{\boldsymbol{\phi}}$ is bounded by*

$$P^* - \|\boldsymbol{\lambda}^*\|_1 L\epsilon \;\leq\; D^*_{\boldsymbol{\phi}} \;\leq\; P^*, \tag{10}$$

where the multiplier norm $\|\boldsymbol{\lambda}^*\|_1$ can be bounded as

$$\|\boldsymbol{\lambda}^*\|_1 \leq \frac{P^* - g_0(\mathbf{x}_0)}{s} < \infty, \tag{11}$$

in which $\mathbf{x}_0$ is the strictly feasible point of Assumption 2.

The proof for this result can be found in [28]. Theorem 1 establishes an upper and lower bound on the optimal dual value in (7) relative to the optimal primal of the original problem in (1) and the accuracy ϵ that characterized the near-universality of function parameterization $\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta})$. Deep neural networks, in particular, fit well into this framework, as they are known to universally approximate any function for any arbitrary ϵ given sufficiently length of the layers [26]. The optimal dual value is not greater than P^* and, more importantly, not worse than a bias on the order of ϵ. These bounds then formalize the notion that we may perform algorithms to solve the unconstrained, parameterized dual problem in (7) in place of the more complex, constrained problem in (3), for which no algorithms are immediately apparent.

3.1. Learning algorithm

To solve the unconstrained dual problem in problem (7), we propose a model-free, *primal-dual* optimization algorithm. A primal-dual method is one in which we iteratively perform gradient-based updates on both the primal $\boldsymbol{\theta}$ and $\mathbf{x}$, and dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ of the Lagrangian function in (5) to find a local stationary point. Consider that we over iterations indexed by k, we successively update values of the primal and dual variables until we reach a point of convergence. At each index k of the primal-dual method, we update the current iterates by adding the corresponding partial gradients of the Lagrangian in (5), i.e. $\nabla_{\boldsymbol{\theta}}\mathcal{L}, \nabla_{\mathbf{x}}\mathcal{L}$, and projecting to the corresponding feasible set. However, because we do not assume that we have explicit forms available for the functions or channel distributions in (1), we cannot evaluate such partial gradients directly. As is commonly done in model-free machine learning methods, such as the case in, e.g., reinforcement learning, we may perform stochastic updates using model-free estimates of the gradients, denoted as $\widehat{\nabla g_0}(\mathbf{x})$, $\widehat{\nabla \mathbf{g}}(\mathbf{x})$ and $\widehat{\nabla_{\boldsymbol{\theta}}}\mathbb{E}_{\mathbf{h}}\mathbf{f}(\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta}))$. Such estimates can be constructed by sampling the functions $g_0, \mathbf{f}, \mathbf{g}$ at or near the location of the current iterates. We may then estimate the partial gradients using the observed outputs, which we denote with hats as in, e.g. $\hat{\mathbf{g}}(\mathbf{x}_k)$. A model-free gradient update is performed on the primal variables as

$$\boldsymbol{\theta}_{k+1} = \mathsf{P}_\Theta\left[\boldsymbol{\theta}_k + \gamma_{\boldsymbol{\theta},k}\widehat{\nabla_{\boldsymbol{\theta}}}\mathbb{E}_{\mathbf{h}}\mathbf{f}(\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta}_k), \mathbf{h})\boldsymbol{\lambda}_k\right], \tag{12}$$

$$\mathbf{x}_{k+1} = \mathsf{P}_{\mathcal{X}}\left[\mathbf{x}_k + \gamma_{\mathbf{x},k}(\widehat{\nabla g_0}(\mathbf{x}) + \widehat{\nabla \mathbf{g}}(\mathbf{x}_k)\boldsymbol{\mu}_k - \mathbf{x}_k)\right]. \tag{13}$$

where we introduce $\gamma_{\boldsymbol{\theta},k}, \gamma_{\mathbf{x},k} > 0$ as scalar step sizes. We likewise update the dual iterates $\boldsymbol{\lambda}_k, \boldsymbol{\mu}_k$ at time k by performing the model-free gradient updates

$$\boldsymbol{\lambda}_{k+1} = \left[\boldsymbol{\lambda}_k - \gamma_{\boldsymbol{\lambda},k}\left(\hat{\mathbf{f}}(\boldsymbol{\phi}(\hat{\mathbf{h}}_k, \boldsymbol{\theta}_{k+1}), \hat{\mathbf{h}}_k) - \mathbf{x}_{k+1}\right)\right]_+ \tag{14}$$

$$\boldsymbol{\mu}_{k+1} = \left[\boldsymbol{\mu}_k - \gamma_{\boldsymbol{\mu},k}\hat{\mathbf{g}}(\mathbf{x}_{k+1})\right]_+. \tag{15}$$

with associated step sizes $\gamma_{\boldsymbol{\lambda},k}, \gamma_{\boldsymbol{\mu},k} > 0$. The gradient primal-dual updates in (12)-(15) successively move the primal and dual variables towards maximum and minimum points of the Lagrangian function, respectively.

We note that model-free gradient estimation is a well-studied topic in machine learning algorithms. While we not discuss the details of such estimation techniques here, we point out the well-known approaches, namely finite difference methods and the policy gradient method [29]. The former approach is a simplistic approach towards estimating gradients that randomly samples points around the current iterate to construct a finite-difference approximation of a derivative. This method is attractive in its simplicity but may suffer from variance or sampling complexity. The latter method, commonly used in model-free reinforcement learning, is particularly useful in estimating gradient of a policy function, such as $\mathbb{E}\big[\mathbf{f}\big(\mathbf{p}(\mathbf{h}), \mathbf{h}\big)\big]$, by using a known distribution function to approximate the derivative. Both such approaches fit naturally into the model-free algorithm we present in (12)-(15).

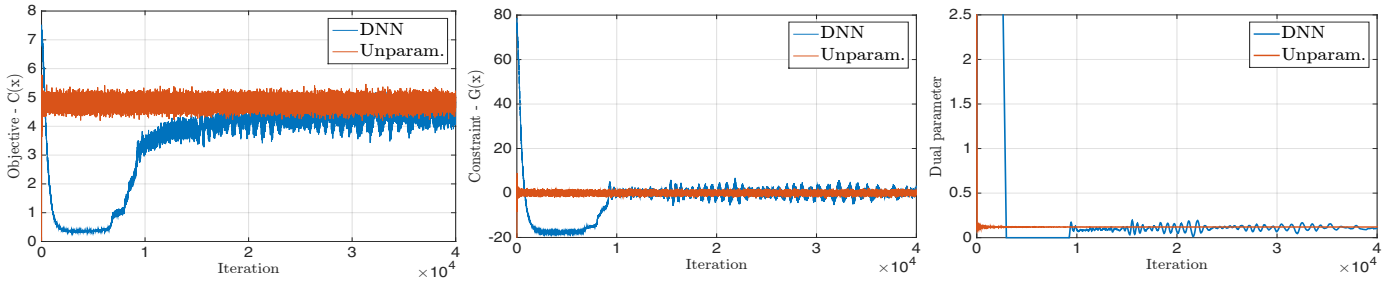

Fig. 1: Convergence of (left) objective function value, (center) constraint value, and (right) dual parameter for simple capacity problem in (16) using proposed DNN method with policy gradients and unparameterized solution. The DNN parameterization obtains near-optimal performance relative to the exact solution.

4. SIMULATION RESULTS

We perform numerical simulations of the proposed dual-learning based approach for learning in wireless systems in a representative problem that takes the form in (1). In particular, consider the problem of maximizing aggregate capacity over a set of dedicated AWGN wireless fading channels. A set of users are each given channels to communicate, and we wish to allocate power between users within a total expected power budget $p_{\max}$. In this case, the capacity over the channel can be modeled as $\log(1+\text{SNR}^i)$, where $\text{SNR}^i := h^i p^i(h^i)/v^i$ is the signal-to-noise ratio experienced by user i and $v^i > 0$ is the noise variance. Likewise, h^i is the fading channel conditions experienced by user i and $p^i(h^i)$ is the power allocation. The capacity function for the ith user is given by $f^i(p^i(h^i), h^i) := \log(1 + h^i p^i(h^i))$. We are interested in maximizing the weighted aggregate throughput across all users, with user i weighted by $w^i \geq 0$. The total capacity problem can be written as

$$P^*_{\boldsymbol{\phi}} := \max_{\boldsymbol{\theta},\mathbf{x}} \sum_{i=1}^{m} w^i x^i \tag{16}$$
$$\text{s.t. } x^i \leq \mathbb{E}_{h^i}\left[\log(1 + h^i \phi^i(h^i, \boldsymbol{\theta})/v^i)\right], \ \forall i$$
$$\mathbb{E}_{\mathbf{h}}\left[\sum_{i=1}^{m} \phi^i(h^i, \boldsymbol{\theta})\right] \leq p_{\max}.$$

Because we do not consider interference in (16), the problem as is may be solved exactly without any learning parametrization using a simple dual stochastic gradient method [30]. Given this, we may use the exact solution as a baseline with which to compare the solution obtained through the dual learning method proposed in this paper, which we stress does not assume any knowledge of the capacity function in (16).

To parameterize our resource allocation function, we use the popular universal parameterization known as deep neural networks (DNNs). Such models are complex and dense models that, in addition to the theoretical universality, have been observed in practice to accurately model many different types of functions. The architecture consists of of L layers, each of which consisting of a linear operation $\mathbf{W}_l$ followed by a point-wise nonlinear activation function σ_l. Common choices of activation functions σ_l include a sigmoid function, a rectifier function (commonly referred to as ReLu), as well as a smooth approximation to the rectifier known as softplus. For the parameterized resource allocation problem in (3), the policy $\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta})$ can be defined through by an L-layer DNN as

$$\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta}) := \sigma_L(\mathbf{W}_L(\sigma_{L-1}(\mathbf{W}_{L-1}(\ldots(\sigma_1(\mathbf{W}_1\mathbf{h})))))), \tag{17}$$

where $\boldsymbol{\theta} \in \mathbb{R}^q$ contains the entries of $\{\mathbf{W}_l\}_{l=1}^L$.

For the simulations performed, we employ a stochastic policy and implement the policy gradient method approach in [29]. To implement policy gradient for the gradient estimation, we use a truncated Gaussian distribution from which to draw resource allocations. The truncated Gaussian distribution has fixed support on the domain $[0, p_{\max}]$. The output layer of the DNN $\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta}) \in \mathbb{R}^{2m}$ is the set of m means and standard deviations to specify the respective truncated Gaussian distributions, i.e. $\boldsymbol{\phi}(\mathbf{h}, \boldsymbol{\theta}) := [\mu^1; \sigma^1; \mu^2; \sigma^2; \ldots; \mu^m; \sigma^m]$. Furthermore, to represent policies that are bounded on the support interval, the output of the last layer is fed into a scaled sigmoid function such that the mean lies in the area of support and the variance is no more than the square root of the support region. In the following experiments, this interval is [0, 10].

For updating the primal and dual variables, we use a batch size of 32. The primal dual method is performed with an exponentially decaying step size for dual updates and the ADAM optimizer [31] for the DNN parameter update. Both updates start with a learning rate of 0.0005, while random channel conditions are generated with an exponential distribution. For each user, their channel gain h^i is provided as input to a single-input-single-output (SISO) DNN, which outputs a power allocation $p^i(h^i)$. Each DNN is constructed with $L = 2$ hidden layers, of size 8 and 4, respectively, each employing ReLU activation, i.e. $\sigma(\mathbf{z}) = [\mathbf{z}]_+$.

The results of a simple experiment with $m = 20$ users with unit weight and variance $w^i = v^i = 1$ is shown in Figure 1. We observe in the left figure that the the total capacity achieved by the DNN dual learning method converges a close to that obtained by the unparameterized method. Likewise, in the center figure, we plot the value of the constraint function, whose convergence to 0 implies that the solution found by our method is indeed feasible. These results support the conclusion made in Theorem 1, namely that learning in the dual domain incurs only a small loss in optimality with respect the unparameterized problem.

5. CONCLUSION

In this paper, we study the problem of finding optimal resource allocation policies in a wide range of wireless communications problems. We make an observation that the resource allocation problem takes a form of a regression problem common in machine learning, which motivates the use of a learning model parameterization. Due to the appearance of the statistical loss in the constraints, there is no clear way to apply model-free learning methods to find solutions to the optimization problem. We demonstrate, however, that the solution to the unconstrained Lagrangian dual problem is close to the original solution, despite the non-convexities of the performance function and learning model parametrization. This closeness is proportional to the accuracy with which the parameterization can approximate arbitrary functions. We further present a learning algorithm to find solutions to the dual problem using model-free gradient estimates. We demonstrate in a numerical simulation that, by using a neural network parametrization of the allocation function, we find near-optimal solutions in a capacity maximization problem.

6. REFERENCES

[1] Xin Wang and Georgios B Giannakis, "Resource allocation for wireless multiuser ofdm networks," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4359–4372, 2011.

[2] Nikos D Sidiropoulos, Timothy N Davidson, and Zhi-Quan Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Processing*, vol. 54, no. 6-1, pp. 2239–2251, 2006.

[3] Juan-Andrés Bazerque and Georgios B Giannakis, "Distributed scheduling and resource allocation for cognitive ofdma radios," *Mobile Networks and Applications*, vol. 13, no. 5, pp. 452–462, 2008.

[4] Yichuan Hu and Alejandro Ribeiro, "Adaptive distributed algorithms for optimal random access channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2703–2715, 2011.

[5] Yichuan Hu and Alejandro Ribeiro, "Optimal wireless networks based on local channel state information," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4913–4929, 2012.

[6] Konstantinos Gatsis, Miroslav Pajic, Alejandro Ribeiro, and George J. Pappas, "Opportunistic control over shared wireless channels," *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3140–3155, December 2015.

[7] Konstantinos Gatsis, Alejandro Ribeiro, and George J. Pappas, "Random access design for wireless control systems," *Automatica*, 2018, To appear. Available on Arxiv.

[8] Mark Eisen, Konstantinos Gatsis, George J Pappas, and Alejandro Ribeiro, "Learning in wireless control systems over non-stationary channels," *arXiv preprint arXiv:1803.01078*, 2018.

[9] Alejandro Ribeiro, "Optimal resource allocation in wireless communication and networking," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, pp. 272, 2012.

[10] Junshan Zhang and Dong Zheng, "A stochastic primal-dual algorithm for joint flow control and mac design in multi-hop wireless networks," in *Information Sciences and Systems, 2006 40th Annual Conference on*. IEEE, 2006, pp. 339–344.

[11] Xin Wang, Tianyi Chen, Xiaojing Chen, Xiaolin Zhou, and Georgios B Giannakis, "Dynamic resource allocation for smart-grid powered mimo downlink transmissions," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3354–3365, 2016.

[12] Xiaojun Liu, Edwin K. P. Chong, and Ness B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2053–2064, 2001.

[13] Atilla Eryilmaz and R Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *IEEE/ACM Transactions on Networking (TON)*, vol. 15, no. 6, pp. 1333–1344, 2007.

[14] Vassilis Ntranos, Nicholas D Sidiropoulos, and Leandros Tassiulas, "On multicast beamforming for minimum outage," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, 2009.

[15] Nariman Farsad and Andrea Goldsmith, "Detection algorithms for communication systems using deep learning," *arXiv preprint arXiv:1705.08044*, 2017.

[16] Haoran Sun, Ziping Zhao, Xiao Fu, and Mingyi Hong, "Limited feedback double directional massive MIMO channel estimation: From low-rank modeling to deep learning," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications*, 2018, p. (to appear).

[17] Haoran Sun, Xiangyi Chen, Qingjiang Shi, Mingyi Hong, Xiao Fu, and Nikos D Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," *arXiv preprint arXiv:1705.09412*, 2017.

[18] Lei Lei, Lei You, Gaoyang Dai, Thang Xuan Vu, Di Yuan, and Symeon Chatzinotas, "A deep learning approach for optimizing content delivering in cache-enabled hetnet," in *Wireless Communication Systems (ISWCS), 2017 International Symposium on*. IEEE, 2017, pp. 449–453.

[19] Woongsup Lee, Minhoe Kim, and Dong-Ho Cho, "Deep power control: Transmit power control scheme based on convolutional neural network," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1276–1279, 2018.

[20] Richard S Sutton, Andrew G Barto, et al., *Reinforcement learning: An introduction*, MIT press, 1998.

[21] Paul de Kerret, David Gesbert, and Maurizio Filippone, "Decentralized deep scheduling for interference channels," *arXiv preprint arXiv:1711.00625*, 2017.

[22] Timothy J O'Shea, Tugba Erpek, and T Charles Clancy, "Physical layer deep learning of encodings for the mimo fading channel," in *Communication, Control, and Computing (Allerton), 2017 55th Annual Allerton Conference on*. IEEE, 2017, pp. 76–80.

[23] Timothy O?Shea and Jakob Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[24] Jooyoung Park and Irwin W Sandberg, "Universal approximation using radial-basis-function networks," *Neural computation*, vol. 3, no. 2, pp. 246–257, 1991.

[25] Bharath Sriperumbudur, Kenji Fukumizu, and Gert Lanckriet, "On the relation between universality, characteristic kernels and rkhs embedding of measures," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 773–780.

[26] Kurt Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.

[27] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[28] Mark Eisen, Clark Zhang, Luiz FO Chamon, Daniel D Lee, and Alejandro Ribeiro, "Learning optimal resource allocations in wireless systems," *arXiv preprint arXiv:1807.08088*, 2018.

[29] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour, "Policy gradient methods for reinforcement learning with function approximation," in *Advances in neural information processing systems*, 2000, pp. 1057–1063.

[30] Xin Wang and Na Gao, "Stochastic resource allocation over fading multiple access and broadcast channels," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2382–2391, 2010.

[31] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.