A ZEROTH-ORDER LEARNING ALGORITHM FOR ERGODIC OPTIMIZATION OF WIRELESS SYSTEMS WITH NO MODELS AND NO GRADIENTS

Dionysios S. Kalogerias[†]*, Mark Eisen*[⋆]*, George J. Pappas*[†]*, and Alejandro Ribeiro*[†]

[†]Department of Electrical & Systems Engineering,
University of Pennsylvania, Philadelphia, PA 19104, USA
[⋆]Intel Corporation, Hillsboro, OR 97124, USA

ABSTRACT

Optimal resource allocation in real-world wireless systems is rather challenging, not only due to the unavailability of accurate statistical channel models, but also because expressions of maximal or achievable information rates are most often unknown, or not adequately precise. Under a modular stochastic functional optimization framework, we propose a new zeroth-order stochastic primal-dual algorithm for completely data-driven, model-free and gradient-free learning of optimal resource allocation policies for ergodic network optimization. Our contribution relies on Gaussian smoothing of the corresponding constrained policy search problem, and on the representation power of universal policy parameterizations, such as Deep Neural Networks (DNNs). Indeed, our simulations demonstrate that DNN-based policies produced by the proposed primal-dual method attain near-ideal performance, based exclusively on limited channel probing, completely bypassing the need for gradient computations, and at the absence of channel or information rate models.

Index Terms— Wireless Systems, Resource Allocation, Zeroth-order Optimization, Reinforcement Learning, Deep Learning.

1. INTRODUCTION

We investigate the general problem of optimally designing a wireless communication system operating over realizations of a collection of $N_{\boldsymbol{H}}$ random *wireless fading channels* $\boldsymbol{H} \in \mathcal{H} \subseteq \mathbb{R}_{+}^{N_{\boldsymbol{H}}}$ by *adaptively* allocating resources, such as user transmission power or channel access. In particular, we consider *ergodic resource allocation problems* (hereafter, *base problem*) of the generic form [1]

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} \quad & g^{o}(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \leq \mathbb{E}\{\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\}\,(\in \mathbb{R}) \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0},\ (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X} \times \mathcal{P} \end{aligned} \quad , \tag{1}$$

where $g^{o} : \mathbb{R}^{N_S} \to \mathbb{R}$ and $\boldsymbol{g} : \mathbb{R}^{N_S} \to \mathbb{R}^{N_{\boldsymbol{g}}}$ are fixed utility functions, $\boldsymbol{p} : \mathcal{H} \to \mathbb{R}^{N_R}$ is the resource allocation policy, and $\boldsymbol{f} : \mathbb{R}^{N_R} \times \mathcal{H} \to \mathbb{R}^{N_S}$ is an instantaneous performance or service level metric, measuring the efficiency of a policy $\boldsymbol{p}$ at each channel realization $\boldsymbol{H}$ in $\mathcal{H}$. Also, the set of admissible policies $\mathcal{P}$ is an appropriately chosen, nonempty subset of Borel measurable functions from $\mathcal{H}$ to $\mathbb{R}^{N_R}$, and $\emptyset \neq \mathcal{X} \subseteq \mathbb{R}^{N_S}$.

Indeed, problem (1) is surprisingly modular and fits many useful scenarios in wireless communications and networking [2, 3, 4, 5, 6, 7, 8, 9, 10]. However, it is rather challenging to tackle. First, (1) exhibits a *variational character*, i.e., it requires *direct policy search*. Second, even though it is safe to assume that the utilities g^{o} and $\boldsymbol{g}$ are concave and that $\mathcal{X}$ is convex, in most existing wireless system models $\boldsymbol{f}$ is naturally nonconcave, rendering (1) nonconvex [1].

Due to the complex structure of problem (1), there has been a large body of literature on the design of optimization techniques for obtaining potentially suboptimal, but effective solutions. These include dual domain algorithms [1, 11, 12, 13], certain powerful heuristics [14, 15], and also data-driven methods based on learning formulations of the problem [16, 17, 18, 19, 20, 21]. Common characteristic of all aforementioned approaches is that they require an accurate model of the wireless system at hand, that is, explicit knowledge of g^{o}, $\boldsymbol{g}$ and, most importantly, $\boldsymbol{f}$. However, this constitutes a serious practical limitation, because, in addition to the unavailability of accurate statistical channel models, explicit expressions of maximal or achievable information rates (i.e., $\boldsymbol{f}$) are most often unknown, or imprecise, even for modestly complex wireless systems.

As recently discussed in [9, 10], a particularly effective way to facilitate policy search in the stochastic variational program (1) is to *restrict* the class of admissible policies as

$$\boldsymbol{p}(\cdot) \equiv \boldsymbol{\phi}(\cdot,\boldsymbol{\theta}) \text{ on } \mathcal{H}, \text{ with } \boldsymbol{\phi} : \mathcal{H} \times \mathbb{R}^{N_{\boldsymbol{\phi}}} \to \mathbb{R}^{N_R}, \tag{2}$$

where the constraint $\boldsymbol{\theta} \in \Theta \subseteq \left\{\boldsymbol{\theta} \in \mathbb{R}^{N_{\boldsymbol{\phi}}} \middle| \boldsymbol{\phi}(\cdot,\boldsymbol{\theta}) \in \mathcal{P}\right\}$ is additionally enforced, for some nonempty and closed parameter space Θ. The $\boldsymbol{\theta}$-parameterized decision rule $\boldsymbol{\phi}$, also known as a Policy Function Approximation (PFA) [22], allows us to relax (1) by formulating the *finite dimensional surrogate* [9]

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{\theta}}{\text{maximize}} \quad & g^{o}(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \leq \mathbb{E}\{\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H},\boldsymbol{\theta}),\boldsymbol{H})\} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0},\ (\boldsymbol{x},\boldsymbol{\theta}) \in \mathcal{X} \times \Theta \end{aligned} \quad , \tag{3}$$

where maximization is over pairs $(\boldsymbol{x},\boldsymbol{\theta}) \in \mathbb{R}^{N_S} \times \mathbb{R}^{N_{\boldsymbol{\phi}}}$. Note that the feasible set of (1) contains that of (3), by construction. Problem (3) resembles a *multi-objective learning problem scalarized by* g^{o}, where each entry of $\boldsymbol{x}$ is associated with an expected reward, with the difference of the two formulating a stochastic constraint. Each expected reward has the form of the objective of a trivial single-stage reinforcement learning problem [22, 23], where $\boldsymbol{H}$ and $\boldsymbol{\phi}(\boldsymbol{H},\boldsymbol{\theta})$ correspond to the state and the control, respectively.

Exploiting the representation power of *(near-)universal* policy parameterizations, such as Deep Neural Networks (DNNs), in conjunction with (3), a *model-free* primal-dual algorithm was recently proposed in [9]. This method approximates the gradients of g^{o} and $\boldsymbol{g}$ via stochastic finite differences, and advocates a policy gradient approach for estimating the composite gradient of $\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H},\cdot),\boldsymbol{H})$, based only on probing $\boldsymbol{f}$; no model of $\boldsymbol{f}$ or its gradient is required.

However, although the algorithm of [9] has shown promising empirical performance, it presents practical limitations. First, it imposes the use of potentially redundant, *randomized* resource allocation policies, essentially forcing an unnecessary enlargement of the

This work is supported in part by the National Science Foundation (NSF) under grant NSF CPS 1837253, and in part by the Intel Science and Technology Center for Wireless Autonomous Systems.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:21:37 UTC from IEEE Xplore. Restrictions apply.

feasible set of (1). Such policies are undesirable, since they induce extra variability in system performance. Second, the algorithm of [9] requires evaluation of the gradient of the parameterization $\boldsymbol{\phi}(\boldsymbol{H},\cdot)$, which can be challenging in practice, for instance when $\boldsymbol{\phi}(\boldsymbol{H},\cdot)$ is a DNN with a very large number of parameters, nonsmooth activations, or with nonstandard architecture; typical cases are recurrent neural networks, or networks with even more exotic structures.

In this work, we propose a new, simple, model-free *and* gradient-free stochastic primal-dual algorithm for tackling (3), which exhibits excellent empirical performance, but *without* the disadvantages of the method of [9]. Our development is based on leveraging Gaussian smoothing to define a new smoothed surrogate to the parameterized problem (3), whose objective *and* constraints are uniformly close to those of (3), in a simple, user-controlled manner. Compared with [9], the proposed method requires *no policy randomization*, and results in deterministic power allocation policies, which exhibit consistent behavior and are straightforward to implement. Additionally, our algorithm operates exclusively on probing g^o, $\boldsymbol{g}$ and the *composition* $\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H},\cdot),\boldsymbol{H})$, *without* the need of evaluating the gradient of the parameterization $\boldsymbol{\phi}(\boldsymbol{H},\cdot)$. Therefore, the proposed algorithm can be implemented in exactly the same way regardless of the choice of $\boldsymbol{\phi}$, and without additional computational requirements. The effectiveness of our approach is corroborated via numerical simulations, which demonstrate that DNN-based policies produced by the proposed primal-dual method indeed attain near-ideal performance, as compared with state-of-the-art, model-based techniques, such as the well-known and powerful WMMSE heuristic [14].

Note: Proofs to all technical results presented in the paper are omitted, and will be included in a subsequent journal submission.

2. GAUSSIAN SMOOTHING & ITS PROPERTIES

This section introduces Gaussian smoothing and its properties, and follows the corresponding treatment in [24]. Let $f:\mathbb{R}^N\to\mathbb{R}$ be Borel. Also, for any $\mathbb{R}^N$-valued random element $\boldsymbol{U}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_N)$, and for $\mu\geq 0$, consider another Borel function $f_\mu:\mathbb{R}^N\to\mathbb{R}$, defined, for $\boldsymbol{x}\in\mathbb{R}^N$, as $f_\mu(\boldsymbol{x})\triangleq\mathbb{E}\{f(\boldsymbol{x}+\mu\boldsymbol{U})\}$, provided that the expectation is well-defined.

In many cases, f_μ turns out to be everywhere differentiable on $\mathbb{R}^N$, *even if f is not*, whereas the gradient of f_μ admits a *zeroth-order representation*. Actually, such is the case of all Lipschitz functions on $\mathbb{R}^N$, as the next result suggests.

Lemma 1. (Properties of f_μ [24]) *Let $\boldsymbol{U}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_N)$, and consider any globally L-Lipschitz function $f:\mathbb{R}^N\to\mathbb{R}$. Then, for any $\mathcal{F}\subseteq\mathbb{R}^N$, the following statements are true:*

S1 *For every $\mu\geq 0$, f_μ is well-defined and finite on $\mathcal{F}$, and*

$$\sup_{\boldsymbol{x}\in\mathcal{F}}|f_\mu(\boldsymbol{x})-f(\boldsymbol{x})|\leq\mu L\sqrt{N}. \tag{4}$$

S2 *If f is convex on $\mathbb{R}^N$, so is f_μ, and $f_\mu\geq f$ on $\mathcal{F}$.*

S3 *For every $\mu>0$, f_μ is differentiable on $\mathcal{F}$, and its gradient $\nabla f_\mu:\mathbb{R}^N\to\mathbb{R}^N$ admits the zeroth-order representation*

$$\nabla f_\mu(\boldsymbol{x})\equiv\mathbb{E}\left\{\frac{f(\boldsymbol{x}+\mu\boldsymbol{U})-f(\boldsymbol{x})}{\mu}\boldsymbol{U}\right\}, \tag{5}$$

for all $\boldsymbol{x}\in\mathcal{F}$. Further, it is true that

$$\sup_{\boldsymbol{x}\in\mathcal{F}}\mathbb{E}\left\{\left\|\frac{f(\boldsymbol{x}+\mu\boldsymbol{U})-f(\boldsymbol{x})}{\mu}\boldsymbol{U}\right\|_2^2\right\}\leq L^2(N+4)^2. \tag{6}$$

3. SMOOTHED CONSTRAINED SURROGATES

Let $\mu_S\geq 0$ and $\mu_R\geq 0$, and consider random elements $\boldsymbol{U}_S\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{N_S})$ and $\boldsymbol{U}_R\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{N_{\boldsymbol{\phi}}})$, the latter taken independent of $\boldsymbol{H}$. Driven by the results of Section 2, we define smoothed versions of g^o, $\boldsymbol{g}$ and $\mathbb{E}\{\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H},\cdot),\boldsymbol{H})\}\triangleq\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}$ as

$$g^o_{\mu_S}(\boldsymbol{x})\triangleq\mathbb{E}\{g^o(\boldsymbol{x}+\mu_S\boldsymbol{U}_S)\},\quad \boldsymbol{x}\in\mathcal{X}, \tag{7}$$

$$\boldsymbol{g}_{\mu_S}(\boldsymbol{x})\triangleq\mathbb{E}\{\boldsymbol{g}(\boldsymbol{x}+\mu_S\boldsymbol{U}_S)\},\quad \boldsymbol{x}\in\mathcal{X}\quad\text{and} \tag{8}$$

$$\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})\triangleq\mathbb{E}\{\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H},\boldsymbol{\theta}+\mu_R\boldsymbol{U}_R),\boldsymbol{H})\},\quad \boldsymbol{\theta}\in\Theta, \tag{9}$$

where, temporarily, we arbitrarily assume that the involved expectations are well-defined and finite on $\mathcal{X}$ and Θ. Then, we may formulate a (hopefully) smoothed version of problem (3) as

$$\begin{aligned}\underset{\boldsymbol{x},\boldsymbol{\theta}}{\text{maximize}}\quad & g^o_{\mu_S}(\boldsymbol{x})\\ \text{subject to}\quad & \boldsymbol{x}+\mathtt{S}(\mu_R)\leq\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})\\ & \boldsymbol{g}_{\mu_S}(\boldsymbol{x})\geq\boldsymbol{0},\ (\boldsymbol{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta\end{aligned}, \tag{10}$$

where $\mathtt{S}:\mathbb{R}_+\to\mathbb{R}_+^{N_S}$ is a nonnegative *feasibility slack*, with properties to be determined. Formulation of the *smoothed surrogate* (10) is well-motivated. This is because (10) constitutes a natural *zeroth-order proxy* for dealing with (3) in the model-free setting, i.e., when the functions g^o, $\boldsymbol{g}$ and $\boldsymbol{f}$ are unknown, and may be only observed through *limited probing*. Indeed, whenever the objective g^o and all entries of the vector functions $\boldsymbol{g}$ and $\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H},\cdot),\boldsymbol{H})$ are sufficiently well-behaved, Lemma 1 would imply that the smoothed functions $g^o_{\mu_S}$, $\boldsymbol{g}_{\mu_S}$ and $\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}$ are differentiable, and that the respective gradients may be represented as averages of suitably defined finite differences. To ensure that this is actually the case, we impose minimal appropriate structure on g^o, $\boldsymbol{g}$ and $\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H},\cdot),\boldsymbol{H})$, as follows. Hereafter, the i-th entries of $\boldsymbol{g}$ ($\boldsymbol{g}_{\mu_S}$) and $\boldsymbol{f}$ ($\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}$) are denoted as g^i ($g^i_{\mu_S}$), $i\in\mathbb{N}^+_{N_{\boldsymbol{g}}}$ and f^i ($\overline{f}^{\boldsymbol{\phi},i}_{\mu_R}$), $i\in\mathbb{N}^+_{N_S}$, respectively.

Assumption 1. *The following conditions are satisfied:*

C1 *For every $i\in\{o,\mathbb{N}^+_{N_{\boldsymbol{g}}}\}$, g^i is L^i_g-Lipschitz on $\mathbb{R}^{N_S}$.*

C2 *For every $i\in\mathbb{N}^+_{N_S}$, there is $L^i_f<\infty$, such that*

$$\begin{aligned}&\|f^i(\boldsymbol{\phi}(\boldsymbol{H},\boldsymbol{\theta}_1),\boldsymbol{H})-f^i(\boldsymbol{\phi}(\boldsymbol{H},\boldsymbol{\theta}_2),\boldsymbol{H})\|_{\mathcal{L}_2}\\ &\quad\leq L^i_f\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2,\ \forall(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\in\mathbb{R}^{N_{\boldsymbol{\phi}}}\times\mathbb{R}^{N_{\boldsymbol{\phi}}}.\end{aligned} \tag{11}$$

Next, for $\boldsymbol{x}\in\mathcal{X}$, $\mu_S>0$ and for $i\in\{o,\mathbb{N}^+_{N_{\boldsymbol{g}}}\}$, let us define

$$\Delta^i_g(\boldsymbol{x},\mu_S,\boldsymbol{U}_S)\triangleq\mu_S^{-1}(g^i(\boldsymbol{x}+\mu_S\boldsymbol{U}_S)-g^i(\boldsymbol{x})). \tag{12}$$

Similarly, for $\boldsymbol{\theta}\in\Theta$, $\mu_R>0$ and for $i\in\mathbb{N}^+_{N_S}$, define

$$\begin{aligned}&\Delta^i_f(\boldsymbol{\theta},\mu_R,\boldsymbol{U}_R,\boldsymbol{H})\\ &\triangleq\mu_R^{-1}(f^i(\boldsymbol{\phi}(\boldsymbol{H},\boldsymbol{\theta}+\mu_R\boldsymbol{U}_R),\boldsymbol{H})-f^i(\boldsymbol{\phi}(\boldsymbol{H},\boldsymbol{\theta}),\boldsymbol{H})).\end{aligned} \tag{13}$$

Assumption 1 may be exploited to establish well-definiteness and basic properties of $g^o_{\mu_S}$, $\boldsymbol{g}_{\mu_S}$ and $\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}$, and their gradients.

Lemma 2. (Existence & Properties of $g^o_{\mu_S}$ and $\boldsymbol{g}_{\mu_S}$) *Suppose that Assumption 1 is in effect. Then, for every $i\in\{o,\mathbb{N}^+_{N_{\boldsymbol{g}}}\}$ and for every $\mu_S>0$, each $g^i_{\mu_S}$ is well-defined, finite, concave, differentiable everywhere on $\mathcal{X}$, and*

$$\sup_{\boldsymbol{x}\in\mathcal{X}}|g^i_{\mu_S}(\boldsymbol{x})-g^i(\boldsymbol{x})|\leq\mu_S L^i_g\sqrt{N_S}, \tag{14}$$

$$\sup_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}\{\|\Delta^i_g(\boldsymbol{x},\mu_S,\boldsymbol{U}_S)\boldsymbol{U}_S\|_2^2\}\leq(L^i_g)^2(N_S+4)^2 \tag{15}$$

$$\text{and}\quad \mathbb{E}\{\Delta^i_g(\boldsymbol{x},\mu_S,\boldsymbol{U}_S)\boldsymbol{U}_S\}\equiv\nabla g^i_{\mu_S}(\boldsymbol{x}),\ \forall\boldsymbol{x}\in\mathcal{X}. \tag{16}$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:21:37 UTC from IEEE Xplore. Restrictions apply.

Lemma 3. (Existence & Properties of $\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}$) *Suppose that Assumption 1 is in effect. Then, for every $i \in \mathbb{N}^+_{N_S}$ and for every $\mu_R > 0$, each $f^i_{\mu_R}$ is well-defined, finite, differentiable everywhere on Θ, and*

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\overline{f}^{\boldsymbol{\phi},i}_{\mu_R}(\boldsymbol{\theta}) - \overline{f}^{\boldsymbol{\phi},i}(\boldsymbol{\theta})\right| \leq \mu_R L^i_f \sqrt{N_{\boldsymbol{\phi}}}, \tag{17}$$

$$\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\{\|\Delta^i_f(\boldsymbol{\theta},\mu_R,\boldsymbol{U}_R,\boldsymbol{H})\boldsymbol{U}_R\|_2^2\} \leq (L^i_f)^2(N_{\boldsymbol{\phi}}+4)^2 \tag{18}$$

$$\textit{and} \quad \mathbb{E}\{\Delta^i_f(\boldsymbol{\theta},\mu_R,\boldsymbol{U}_R,\boldsymbol{H})\boldsymbol{U}_R\} \equiv \nabla\overline{f}^{\boldsymbol{\phi},i}_{\mu_R}(\boldsymbol{\theta}),\ \forall \boldsymbol{\theta}\in\Theta. \tag{19}$$

We see that, under Assumption 1, the objective and constraints of (3) are all approximated *uniformly* by those of the smoothed surrogate (10). Further, the gradients of all functions involved in (10) exist and admit *exact* representations as integrated finite differences.

We now investigate conditions ensuring feasibility of (10). In fact, we are interested in conditions which ensure feasibility of (10), *but are on the original parameterized problem* (3). This is useful from a practical point of view, since the exact form of (10) will be, in principle, unknown. The relevant result follows after we define vectors $\boldsymbol{c}_S \triangleq \left[L^1_g \ \ldots \ L^{N_{\boldsymbol{g}}}_g\right]^{\boldsymbol{T}}$ and $\boldsymbol{c}_R \triangleq \left[L^1_f \ \ldots \ L^{N_S}_f\right]^{\boldsymbol{T}}$, and under the following assumption.

Assumption 2. *The feasibility slack $\mathtt{S}_{\boldsymbol{f}}$ is increasing around the origin, and* $\lim_{\mu_R\downarrow 0} \mathtt{S}_{\boldsymbol{f}}(\mu_R) \equiv \mathtt{S}_{\boldsymbol{f}}(0) \equiv \boldsymbol{0}$.

Theorem 1. (Surrogate Strict Feasibility) *Let Assumptions 1 and 2 be in effect, and suppose that $(\boldsymbol{x}^\dagger,\boldsymbol{\theta}^\dagger) \in \mathbb{R}^{N_S}\times\mathbb{R}^{N_{\boldsymbol{\phi}}}$ is a strictly feasible point of problem* (3)*. Then, there exist $\mu^\dagger_S > 0$, $\mu^\dagger_R > 0$, such that, for every $0 \leq \mu_S \leq \mu^\dagger_S$ and $0 \leq \mu_R \leq \mu^\dagger_R$, the same point $(\boldsymbol{x}^\dagger,\boldsymbol{\theta}^\dagger)$ is strictly feasible for the smoothed surrogate* (10)*.*

Theorem 1 is very important, because it confirms the existence of a strictly feasible point for problem (10), *uniformly relative to μ_S and μ_R*. A byproduct of Theorem 1 is that (10) is a feasible, meaningful optimization problem. Another similar question we may ask is how much the constraints of (3) are violated for every feasible solution of (10). In this respect, we formulate the following result.

Theorem 2. (PFA Constraint Violation) *Let Assumption 1 be in effect. Then, for every $\mu_R \geq 0$ such that $\mathtt{S}(\mu_R) - \mu_R\boldsymbol{c}_R\sqrt{N_{\boldsymbol{\phi}}} \geq \boldsymbol{0}$, and for every $\mu_S \geq 0$, every feasible point of* (10) *is also feasible for* (3)*. Otherwise, if the stated condition fails to hold, then the negative values of its left-hand-side correspond to the respective levels of maximal constraint violation for* (3)*.*

As an example, one can apparently set $\mathtt{S}(\mu_R) \equiv \mu_R\boldsymbol{c}_R\sqrt{N_{\boldsymbol{\phi}}}$, since $\mathtt{S}$ readily satisfies Assumption 2. However, this might not be a feasible choice in practice, since $\boldsymbol{c}_R$ will probably be unknown. Still, Theorem 2 provides a basic principle for choosing $\mathtt{S}$. For instance, the choice $\mathtt{S}(\mu_R) \equiv \boldsymbol{C}\mu_R\sqrt{N_{\boldsymbol{\phi}}}$ would work fine, for an appropriate constant vector $\boldsymbol{C} > \boldsymbol{0}$, which may chosen experimentally.

4. PRIMAL-DUAL MODEL-FREE LEARNING

We now present a simple and efficient *zeroth-order* randomized primal-dual algorithm for dealing directly with the smoothed surrogate (10) in the model-free setting. The algorithm leverages stochastic approximation to tackle the minimax problem

$$\text{minimize}_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \quad \sup_{(\boldsymbol{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu}(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \ , \tag{20}$$

for every qualifying choice of μ_S and μ_R, where the function $\mathscr{L}_{\boldsymbol{\phi}} : \mathbb{R}^{N_S}\times\mathbb{R}^{N_{\boldsymbol{\phi}}}\times\mathbb{R}^{N_{\boldsymbol{g}}}\times\mathbb{R}^{N_S} \to \mathbb{R}$ is the *Lagrangian* of (10), given by

$$\mathscr{L}_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \triangleq g^o(\boldsymbol{x}) + \langle\boldsymbol{\lambda}_S, \boldsymbol{g}(\boldsymbol{x})\rangle + \left\langle\boldsymbol{\lambda}_R, \overline{\boldsymbol{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta}) - \boldsymbol{x}\right\rangle, \tag{21}$$

Algorithm 1 Model-Free Randomized Primal-Dual Learning

Input: $\boldsymbol{x}^0, \boldsymbol{\theta}^0, \boldsymbol{\lambda}^0_S, \boldsymbol{\lambda}^0_R, \{\gamma^n_{\boldsymbol{x}}, \gamma^n_{\boldsymbol{\theta}}, \gamma^n_{\boldsymbol{\lambda}_S}, \gamma^n_{\boldsymbol{\lambda}_R}\}_{n\in\mathbb{N}}, \mu_S, \mu_R$
Output: $\{\boldsymbol{x}^{n+1}, \boldsymbol{\theta}^{n+1}\}_{n\in\mathbb{N}}$
1: **for** $n = 0, 1, 2, \ldots$ **do**
2: Draw samples $\boldsymbol{U}^{n+1}_S$ and $\boldsymbol{U}^{n+1}_R$.
3: Sample values $g^o(\boldsymbol{x}^n)$, $g^o(\boldsymbol{x}^n + \mu_S\boldsymbol{U}^{n+1}_S)$ and $\boldsymbol{g}(\boldsymbol{x}^n)$, $\boldsymbol{g}(\boldsymbol{x}^n + \mu_S\boldsymbol{U}^{n+1}_S)$, and probe the wireless system to obtain

$$\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H}^{n+1},\boldsymbol{\theta}^n),\boldsymbol{H}^{n+1}) \quad \text{and}$$
$$\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H}^{n+1},\boldsymbol{\theta}^n + \mu_R\boldsymbol{U}^{n+1}_R),\boldsymbol{H}^{n+1}).$$

4: Compute $\boldsymbol{x}^{n+1}$ and $\boldsymbol{\theta}^{n+1}$ from (26) and (27).
5: Sample $\boldsymbol{g}(\boldsymbol{x}^{n+1} + \mu_S\boldsymbol{U}^{n+1}_S)$ and probe the wireless system to obtain $\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H}^{n+1},\boldsymbol{\theta}^{n+1} + \mu_R\boldsymbol{U}^{n+1}_R)$.
6: Compute $\boldsymbol{\lambda}^{n+1}_S$ and $\boldsymbol{\lambda}^{n+1}_R$ from (28) and (29).
7: **end for**

where $\boldsymbol{\lambda} \equiv (\boldsymbol{\lambda}_S, \boldsymbol{\lambda}_R) \in \mathbb{R}^{N_{\boldsymbol{g}}}\times\mathbb{R}^{N_S}$ are multipliers associated with the respective constraint of the *primal problem* (10).

A basic idea for designing a simple primal-dual method for solving (20) is to iteratively ascend in $(\boldsymbol{x},\boldsymbol{\theta}) \in \mathcal{X}\times\Theta$ and descend in $\boldsymbol{\lambda} \equiv (\boldsymbol{\lambda}_S,\boldsymbol{\lambda}_R) \geq \boldsymbol{0}$, in an alternating fashion. If $n \in \mathbb{N}$ denotes an iteration index, and by taking the gradients of $\mathscr{L}_{\boldsymbol{\phi},\mu}$ relative to $\boldsymbol{x}$, $\boldsymbol{\theta}$, $\boldsymbol{\lambda}_S$ and $\boldsymbol{\lambda}_R$, such a procedure may be implemented via the updates

$$\boldsymbol{x}^{n+1} \equiv \Pi_{\mathcal{X}}\left\{\boldsymbol{x}^n + \boldsymbol{\gamma}^n_{\boldsymbol{x}} \circ (\nabla g^o_{\mu_S}(\boldsymbol{x}^n) + \nabla\boldsymbol{g}_{\mu_S}(\boldsymbol{x}^n)\boldsymbol{\lambda}^n_S - \boldsymbol{\lambda}^n_R)\right\}, \tag{22}$$

$$\boldsymbol{\theta}^{n+1} \equiv \Pi_{\Theta}\left\{\boldsymbol{\theta}^n + \boldsymbol{\gamma}^n_{\boldsymbol{\theta}} \circ \nabla\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}^n)\boldsymbol{\lambda}^n_R\right\}, \tag{23}$$

$$\boldsymbol{\lambda}^{n+1}_S \equiv (\boldsymbol{\lambda}^n_S - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \circ \boldsymbol{g}_{\mu_S}(\boldsymbol{x}^{n+1}))_+ \quad \text{and} \tag{24}$$

$$\boldsymbol{\lambda}^{n+1}_R \equiv \left(\boldsymbol{\lambda}^n_R - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \circ \left(\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}^{n+1}) - \boldsymbol{x}^{n+1} - \mathtt{S}(\mu_R)\right)\right)_+, \tag{25}$$

respectively, where "$\circ$" denotes the operation of the Hadamard product, $\{\boldsymbol{\gamma}^n_{\boldsymbol{x}}, \boldsymbol{\gamma}^n_{\boldsymbol{\theta}}, \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S}, \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_R}\}_{n\in\mathbb{N}}$ are nonnegative vector stepsizes, $\nabla\boldsymbol{g}_{\mu_S} : \mathbb{R}^{N_S} \to \mathbb{R}^{N_S}\times\mathbb{R}^{N_{\boldsymbol{g}}}$ and $\nabla\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R} : \mathbb{R}^{N_{\boldsymbol{\phi}}} \to \mathbb{R}^{N_{\boldsymbol{\phi}}}\times\mathbb{R}^{N_S}$ are corresponding Jacobians and, for every nonempty closed set $\mathcal{A} \subseteq \mathbb{R}^N$, $\Pi_{\mathcal{A}} : \mathbb{R}^N \to \mathcal{A}$ is the usual Euclidean projection.

However, the algorithm described by (22)-(25) is in general *not implementable*, because *neither* the functions $g^o_{\mu_S}$, $\boldsymbol{g}_{\mu_S}$ and $\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}$, *nor* their gradients are explicitly known, *for any possible values of* $\mu_S > 0$ and $\mu_R > 0$. Nevertheless, by Lemma 2 (provided the respective assumptions are satisfied), *all* functions $g^o_{\mu_S}$, $\boldsymbol{g}_{\mu_S}$, $\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}$, *and* their derivatives are given by *known expectation functions*. What is more, all these expectation functions depend exclusively on zeroth-order information, i.e., on evaluations of $g^o_{\mu_S}$, $\boldsymbol{g}_{\mu_S}$ and $\overline{\boldsymbol{f}}^{\boldsymbol{\phi}}_{\mu_R}$, *only*. Then, a stochastic gradient version of the algorithm consisting by (22)-(25) may be formulated by replacing all expectations as

$$\begin{aligned}\boldsymbol{x}^{n+1} \equiv \Pi_{\mathcal{X}}\big\{&\boldsymbol{x}^n + \boldsymbol{\gamma}^n_{\boldsymbol{x}} \circ (\Delta^o_g(\boldsymbol{x}^n,\mu_S,\boldsymbol{U}^{n+1}_S)\boldsymbol{U}^{n+1}_S \\ &+ \langle\boldsymbol{\Delta}_{\boldsymbol{g}}(\boldsymbol{x}^n,\mu_S,\boldsymbol{U}^{n+1}_S),\boldsymbol{\lambda}^n_S\rangle\boldsymbol{U}^{n+1}_S - \boldsymbol{\lambda}^n_R)\big\},\end{aligned} \tag{26}$$

$$\begin{aligned}\boldsymbol{\theta}^{n+1} \equiv \Pi_{\Theta}\big\{&\boldsymbol{\theta}^n \\ &+ \boldsymbol{\gamma}^n_{\boldsymbol{\theta}} \circ \langle\boldsymbol{\Delta}_{\boldsymbol{f}}(\boldsymbol{\theta}^n,\mu_R,\boldsymbol{U}^{n+1}_R,\boldsymbol{H}^{n+1}),\boldsymbol{\lambda}^n_R\rangle\boldsymbol{U}^{n+1}_R\big\},\end{aligned} \tag{27}$$

$$\boldsymbol{\lambda}^{n+1}_S \equiv (\boldsymbol{\lambda}^n_S - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \circ \boldsymbol{g}(\boldsymbol{x}^{n+1} + \mu_S\boldsymbol{U}^{n+1}_S))_+ \quad \text{and} \tag{28}$$

$$\begin{aligned}\boldsymbol{\lambda}^{n+1}_R \equiv \big(&\boldsymbol{\lambda}^n_R - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_R} \circ \big(\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{H}^{n+1},\boldsymbol{\theta}^{n+1} + \mu_R\boldsymbol{U}^{n+1}_R),\boldsymbol{H}^{n+1}) \\ &- \boldsymbol{x}^{n+1} - \mathtt{S}(\mu_R)\big)\big)_+,\end{aligned} \tag{29}$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:21:37 UTC from IEEE Xplore. Restrictions apply.

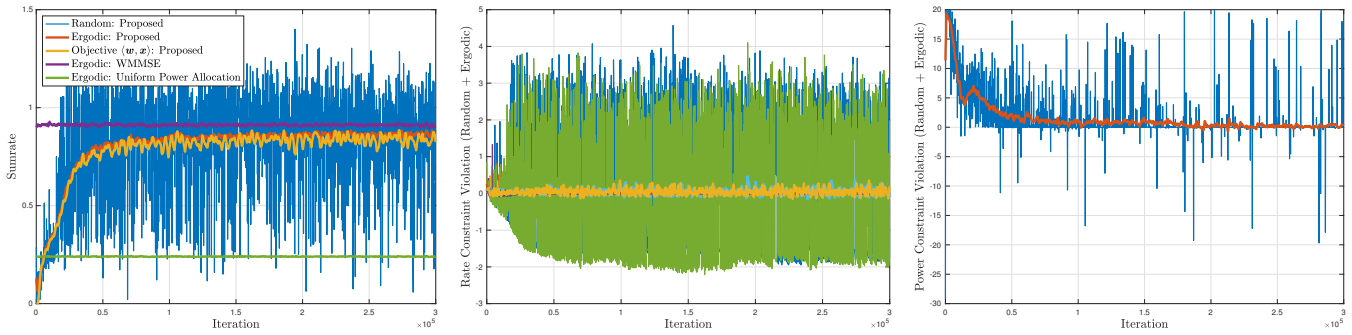


Fig. 1: Rate (left) and power (right) constraint violation exhibited by the proposed method, both random and ergodic, in the case of a MAI channel.

where, dropping dependencies, $\boldsymbol{\Delta_g} \in \mathbb{R}^{N_g}$ and $\boldsymbol{\Delta_g} \in \mathbb{R}^{N_S}$ are defined as $\boldsymbol{\Delta_g} \triangleq \left[\Delta_g^1 \ \dots \ \Delta_g^{N_g}\right]^{\boldsymbol{T}}$ and $\boldsymbol{\Delta_f} \triangleq \left[\Delta_f^1 \ \dots \ \Delta_f^{N_S}\right]^{\boldsymbol{T}}$, respectively. A complete description of the proposed model-free method is provided in Algorithm 1. We see that the algorithm requires exactly *three system probes per user, per iteration*.

Key differences between Algorithm 1 and the primal-dual method presented in [9] are the presence of the feasibility slack $\mathtt{S}$, which follows from our analysis, as well as the fact that, due to our explicit formulation of the smoothed surrogate (10), the dual updates (28) and (29) are naturally *randomized*, in addition to the primal updates (26) and (27). Additionally, [9] relies on a policy gradient type approach for achieving model-free learning for the algorithm developed therein, whereas Algorithm 1 completely bypasses the need for introducing randomized policies into the learning procedure. This also makes Algorithm 1 rather straightforward to implement, as evaluation of the gradient of $\boldsymbol{\phi}$ is *not* required.

5. NUMERICAL SIMULATIONS

The efficacy of Algorithm 1 is now numerically confirmed by application on a Multiple Access Interference (MAI) channel model, where N_S transmitters simultaneously communicate with a central node, for instance, a common receiver, or a base station. The signal transmitted by each user creates interference to the signals transmitted by all other users in the network. We would like to optimally allocate power between users in order to maximize the weighted sumrate of the network, within a total expected power specification p_{max}. To do that, we consider the stochastic program

$$\begin{aligned} \underset{x^i, i\in\mathbb{N}_{N_S}^+, \boldsymbol{\theta}}{\text{maximize}} \ & \sum_{i\in\mathbb{N}_{N_S}^+} w^i x^i \\ \text{subject to } & x^i \leq \mathbb{E}\left\{\log\left(1+\frac{H^i\phi^i(\boldsymbol{H},\boldsymbol{\theta})}{\nu^i+\sum_{i\neq j} H^j\phi^j(\boldsymbol{H},\boldsymbol{\theta})}\right)\right\} \\ & \mathbb{E}\left\{\sum_{i\in\mathbb{N}_{N_S}^+}\phi^i(\boldsymbol{H},\boldsymbol{\theta})\right\} \leq p_{max},\ x^i \geq 0, i\in\mathbb{N}_{N_S}^+ \end{aligned} \quad , \quad (30)$$

where $\phi^i : \mathbb{R}_+^{N_S} \times \mathbb{R}^{N_{\boldsymbol{\phi}}} \to [0, p_{max}]$, $i \in \mathbb{N}_{N_S}^+$ is an element of the vector sigmoid output layer of a DNN taking as input the *squared* fading channel vector $\boldsymbol{H}$, and having two ReLU hidden layers with thirty-two and sixteen neurons, respectively, as in [9]. We note that problem (30) may be easily reexpressed in the canonical form of (3).

In our simulations, we assume $N_S \equiv 5$ users, a power budget $p_{max} \equiv 20$, and a randomly generated weight vector $\boldsymbol{w}$. We also let $\nu^i \equiv 1$, and H^i is exponentially distributed with parameter $\lambda \equiv 1/2$, modeling the square of a unit variance Rayleigh fading channel state, for all $i \in \mathbb{N}_{N_S}^+$. Then, we execute Algorithm 1 for $3{\cdot}10^5$ iterations, with initial values $\boldsymbol{x}^0 \equiv \boldsymbol{\theta}^0 \equiv \mathbf{0}$, $\boldsymbol{\lambda}_S^0 \equiv \mathbf{1}_{N_S+1}$, constant step-sizes $\boldsymbol{\gamma}_{\boldsymbol{x}}^n \equiv 0.0008\mathbf{1}, \boldsymbol{\gamma}_{\boldsymbol{\theta}}^n \equiv 0.0005\mathbf{1}, \boldsymbol{\gamma}_{\boldsymbol{\lambda}_S}^n \equiv \left[0.0051_{N_S}^{\boldsymbol{T}}\ 0.0001\right]^{\boldsymbol{T}}$, for all $n \in \mathbb{N}$, null feasibility slack $\mathtt{S} \equiv \mathbf{0}$, and with $\mu_R \equiv 10^{-9}$. Since the objective of (30) is usually perfectly known to the wireless engineer, we may set $\mu_S \equiv 0$ in this example.

Fig. 1(left) shows the evolution of the sequence of objective values $\{\langle \boldsymbol{w}, \boldsymbol{x}^n\rangle\}_{n\in\mathbb{N}}$, the instantaneous sumrate sequence obtained from Algorithm 1, as well as an approximation of the corresponding ergodic sumrate. Since the solution to the variational version of (30) is unavailable (cf. (1)), we compare the ergodic performance of Algorithm 1 with that of the powerful WMMSE policy [14], which however assumes *complete knowledge* of the information theoretic model of the wireless system. For reference, Fig. 1(left) also shows the ergodic performance achieved by a uniform power allocation policy across users. All ergodic estimates were computed via moving average smoothing of the respective process realizations.

Fig. 1(left) readily demonstrates that the values of the objective of (30) match the values of the estimated ergodic sumrate, as both obtained from Algorithm 1. Perhaps more importantly, the ergodic sumrate converges *remarkably close* to that achieved by model-based WMMSE. Fig. 1(center, right) shows both instantaneous and similarly obtained ergodic estimates of the rate and power constraint violation of the decisions produced by Algorithm 1. We observe that all constraints are active on average, thus feasible for (30), as required. Therefore, the proposed zeroth-order primal-dual-method indeed produces feasible *and* near-state-of-the-art deterministic power allocation policies without knowledge of a system model, and in absence of gradient information.

6. CONCLUSIONS

We have proposed a new zeroth-order stochastic primal-dual algorithm for data-driven learning of optimal ergodic policies in wireless systems, free of *both* models *and* gradients. Our development relied on leveraging Gaussian smoothing and replacing the original constrained policy search problem by a suitable, technically grounded, smoothed surrogate, as well as on the representation power of universal policy parameterizations, in particular DNNs. Our simulations demonstrated that deterministic, DNN-based policies produced by the proposed method attain excellent performance as compared with state-of-the-art, model-based techniques, while based exclusively on limited channel probing, and bypassing the need for gradient computations, as well as channel or information rate models.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:21:37 UTC from IEEE Xplore. Restrictions apply.

7. REFERENCES

[1] A. Ribeiro, "Optimal Resource Allocation in Wireless Communication and Networking," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, pp. 272, 2012.

[2] N. Sidiropoulos, T. Davidson, and Zhi-Quan Luo, "Transmit Beamforming for Physical-Layer Multicasting," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, jun 2006.

[3] X. Wang and N. Gao, "Stochastic Resource Allocation Over Fading Multiple Access and Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2382–2391, may 2010.

[4] D. Wang, Z. Li, and X. Wang, "Joint Optimal Subcarrier and Power Allocation for Wireless Cooperative Networks Over OFDM Fading Channels," in *IEEE Transactions on Vehicular Technology*, jan 2012, vol. 61, pp. 249–257.

[5] T. He, X. Wang, and W. Ni, "Optimal Chunk-Based Resource Allocation for OFDMA Systems With Multiple BER Requirements," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4292–4301, nov 2014.

[6] K. Gatsis, M. Pajic, A. Ribeiro, and G. J. Pappas, "Opportunistic Control Over Shared Wireless Channels," *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3140–3155, dec 2015.

[7] J. Rubio, O. Muñoz-Medina, A. Pascual-Iserte, and J. Vidal, "Stochastic Resource Allocation with a Backhaul Constraint for the Uplink," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*. sep 2016, pp. 1–6, IEEE.

[8] H. Yu and M. J. Neely, "Dynamic Transmit Covariance Design in MIMO Fading Systems with Unknown Channel Distributions and Inaccurate Channel State Information," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3996–4008, jun 2017.

[9] M. Eisen, C. Zhang, L. F. Chamon, D. D. Lee, and A. Ribeiro, "Learning Optimal Resource Allocations in Wireless Systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2775–2790, may 2019.

[10] M. Eisen and A. Ribeiro, "Optimal Wireless Resource Allocation with Random Edge Graph Neural Networks," *arXiv preprint, arXiv:1909.01865*, sep 2019.

[11] J. Zhang and D. Zheng, "A Stochastic Primal-Dual Algorithm for Joint Flow Control and MAC Design in Multi-hop Wireless Networks," in *2006 40th Annual Conference on Information Sciences and Systems*. mar 2006, pp. 339–344, IEEE.

[12] Wei Yu and R. Lui, "Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310–1322, jul 2006.

[13] X. Wang, T. Chen, X. Chen, X. Zhou, and G. 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, dec 2016.

[14] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, sep 2011.

[15] C. S. Chen, K. W. Shum, and C. W. Sung, "Round-Robin Power Control for the Weighted Sumrate Maximisation of Wireless Networks over Multiple Interfering Links," *European Transactions on Telecommunications*, vol. 22, no. 8, pp. 458–470, dec 2011.

[16] Z. Xu, Y. Wang, J. Tang, J. Wang, and M. C. Gursoy, "A Deep Reinforcement Learning-based Framework for Power-Efficient Resource Allocation in Cloud RANs," in *2017 IEEE International Conference on Communications (ICC)*. may 2017, pp. 1–6, IEEE.

[17] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to Optimize: Training Deep Neural Networks for Interference Management," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5438–5453, oct 2018.

[18] W. Lee, M. Kim, and D.-H. Cho, "Deep Power Control: Transmit Power Control Scheme Based on Convolutional Neural Network," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1276–1279, jun 2018.

[19] T. Van Chien, E. Björnson, and E. G. Larsson, "Sum Spectral Efficiency Maximization in Massive MIMO Systems: Benefits from Deep Learning," *arXiv preprint, arXiv:1903.08163*, mar 2019.

[20] W. Cui, K. Shen, and W. Yu, "Spatial Deep Learning for Wireless Scheduling," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1248–1261, dec 2019.

[21] F. Meng, P. Chen, L. Wu, and J. Cheng, "Power Allocation in Multi-User Cellular Networks: Deep Reinforcement Learning Approaches," *arXiv preprint, arXiv:1901.07159*, jan 2019.

[22] W. B. Powell, "A Unified Framework for Stochastic Optimization," *European Journal of Operational Research*, vol. 275, no. 3, pp. 795–821, jun 2019.

[23] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, Wiley-Interscience, 2nd editio edition, 2011.

[24] Y. Nesterov and V. Spokoiny, "Random Gradient-Free Minimization of Convex Functions," *Foundations of Computational Mathematics*, vol. 17, no. 2, pp. 527–566, apr 2017.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:21:37 UTC from IEEE Xplore. Restrictions apply.