Model-Free Learning of Optimal Ergodic Policies in Wireless Systems

Dionysios S. Kalogerias, *Member, IEEE*, Mark Eisen, *Member, IEEE*
George J. Pappas, *Fellow, IEEE*, and Alejandro Ribeiro

***Abstract*—Learning optimal resource allocation policies in wireless systems can be effectively achieved by formulating finite dimensional constrained programs which depend on system configuration, as well as the adopted learning parameterization. The interest here is in cases where system models are unavailable, prompting methods that probe the wireless system with candidate policies, and then use observed performance to determine better policies. This generic procedure is difficult because of the need to cull accurate gradient estimates out of these limited system queries. This paper constructs and exploits smoothed surrogates of constrained ergodic resource allocation problems, the gradients of the former being representable exactly as averages of finite differences that can be obtained through limited system probing. Leveraging this unique property, we develop a new model-free primal-dual algorithm for learning optimal ergodic resource allocations, while we rigorously analyze the relationships between original policy search problems and their surrogates, in both primal and dual domains. First, we show that both primal and dual domain surrogates are uniformly consistent approximations of their corresponding original finite dimensional counterparts. Upon further assuming the use of near-universal policy parameterizations, we also develop explicit bounds on the gap between optimal values of initial, infinite dimensional resource allocation problems, and dual values of their parameterized smoothed surrogates. In fact, we show that this duality gap decreases at a linear rate relative to smoothing and universality parameters. Thus, it can be made arbitrarily small at will, also justifying our proposed primal-dual algorithmic recipe. Numerical simulations confirm the effectiveness of our approach.**

***Index Terms*—Wireless Systems, Stochastic Resource Allocation, Zeroth-order Optimization, Constrained Nonconvex Optimization, Deep Learning, Lagrangian Duality, Strong Duality.**

I. Introduction and Problem Formulation

We investigate optimal wireless communication systems operating over realizations of random fading channels $\mathbf{H} \in \mathcal{H} \subseteq \mathbf{R}^{N_{\mathbf{H}}}$ with distribution $\mathcal{M}_{\mathbf{H}}$. Resources such as transmission power and channel access are allocated to jointly maximize the service levels of one or multiple users, in a certain sense. Due to randomness of $\mathbf{H}$, a reasonable objective is to optimize quality of service in an ergodic regime, i.e., by averaging all possible instantaneous service levels relative to the fading distribution $\mathcal{M}_{\mathbf{H}}$. Then, optimal wireless system design may be abstracted to a stylized base resource allocation problem of the form [1]

$$\begin{array}{ll} \underset{\mathbf{x},\mathbf{p}(\cdot)}{\text{maximize}} & g^o(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \le \mathbb{E}\{\mathbf{f}(\mathbf{p}(\mathbf{H}),\mathbf{H})\} \\ & \mathbf{g}(\mathbf{x}) \ge \mathbf{0},\ (\mathbf{x},\mathbf{p}) \in \mathcal{X}\times\mathcal{P} \end{array}; \tag{1}$$

D. S. Kalogerias, G. J. Pappas, and A. Ribeiro are with the Dept. of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA (e-mail: {dionysis, pappasg, aribeiro}@seas.upenn.edu). M. Eisen is with Intel Corporation, Hillsboro, OR 97124, USA (e-mail: mark.eisen@intel.com).
This work is supported by the Intel Science and Technology Center for Wireless Autonomous Systems and NSF under Grant CPS 1837253.

in (1), the policy $\mathbf{p} : \mathcal{H} \to \mathbf{R}^{N_R}$ maps fading states $\mathbf{H}$ to N_R resource allocation decisions $\mathbf{p}(\mathbf{H})$, the function $\mathbf{f} : \mathbf{R}^{N_R} \times \mathcal{H} \to \mathbf{R}^{N_S}$ maps decisions and fading values to N_S *instantaneous* service level metrics, the average of which bounds the ergodic metrics $\mathbf{x} \in \mathbf{R}^{N_S}$, whose worth we evaluate through the utilities $g^o : \mathbf{R}^{N_S} \to \mathbf{R}$ and $\mathbf{g} : \mathbf{R}^{N_S} \to \mathbf{R}^{N_{\mathbf{g}}}$. Ergodic performances are further restricted to the set $\mathcal{X} \subseteq \mathbf{R}^{N_S}$ and resource allocations are further restricted to the set $\mathcal{P}$, the latter inducing *pointwise* constraints on each individual value $\mathbf{p}(\mathbf{H})$ of every candidate policy $\mathbf{p}$ [1], for each fading realization $\mathbf{H}$.

Problem (1) conveniently abstracts several resource allocation tasks of practical importance. Typical examples of service level metrics are system/network capacities, rates, and signal-to-noise ratios. Also, the objective scalar utility function g^o can be chosen as a weighted or proportional fairness combination of ergodic service levels, whereas the vector utility $\mathbf{g}$ may be designed to impose explicit user demand and fairness constraints. In fact, it is relatively straightforward to see that particular cases of (1) appear naturally in, e.g., point-to-point channels [1], interference channels [1]–[4], wireless networking [1], [5], [6], as well as multiple access [7], [8], random access [9], [10] and frequency division multiplexing [11]–[13]. Less obvious application areas where resource allocation tasks can also be formulated as particular cases of (1) include MIMO systems [14], [15], beamforming [16]–[18], caching [19], and wireless control [20]–[22]. For some concrete and concise examples explicitly relating (1) to practical wireless system models, also see [1], [23], [24].

Although problems in [1]–[22] have their own difficulties, they all share three challenges which are well-described by (1): *Dimensionality*, *lack of convexity*, and *model availbaility*. Indeed, when $\mathcal{H}$ is an infinite set –as in most applications– finding an optimal or near-optimal solution to (1) requires direct policy search, which is a rather obscure and complicated task. Further, while the utilities g^o and $\mathbf{g}$ and the feasible set $\mathcal{X}$ are often known design choices and can be made concave or convex as needed, this is not the case with the distribution $\mathcal{M}_{\mathbf{H}}$, the service metric $\mathbf{f}$, or the set $\mathcal{P}$. These entities depend on propagation physics, as well as models of interference and multiple access management. Most often, such models are either inaccurate or unavailable, especially in complex networking settings, whereas in most existing models the form of $\mathbf{f}$ and $\mathcal{P}$ render (1) nonconvex [1].

Lack of convexity is an inherent challenge and it is accepted that we settle for locally optimal solutions, heuristics, or relaxations. To some extent, the same counts for dimensionality and model availability. However, the recent advent of machine learning for wireless communications [23]–[39] has dawned the realization that both these challenges can be ameliorated with the incorporation of learning parametrizations [23], [24]. To see why this is true, introduce a parameterization $\boldsymbol{\phi}:\mathcal{H}\times\mathbf{R}^{N_\phi}\rightarrow\mathbf{R}^{N_R}$, and restrict resource allocations as $\mathbf{p}(\cdot)\equiv\boldsymbol{\phi}(\cdot,\boldsymbol{\theta}),\boldsymbol{\theta}\in\mathbf{R}^{N_\phi}$. Then, the base problem (1) may be relaxed as

$$\begin{array}{rl}\underset{\mathbf{x},\boldsymbol{\theta}}{\text{maximize}} & g^o(\mathbf{x})\\ \text{subject to} & \mathbf{x}\le\mathbb{E}\{\mathbf{f}(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}),\mathbf{H})\}\\ & \mathbf{g}(\mathbf{x})\ge\mathbf{0},\ (\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta\end{array}, \tag{2}$$

where $\Theta\subseteq\left\{\boldsymbol{\theta}\in\mathbf{R}^{N_\phi}\middle|\boldsymbol{\phi}(\cdot,\boldsymbol{\theta})\in\mathcal{P}\right\}$ is a nonempty and closed parameter space. Through the parametrization $\boldsymbol{\phi}$, also known as a Policy Function Approximation (PFA) [40], problem (2) serves as a finite dimensional surrogate for the infinite dimensional problem (1) [23]. Solving such a surrogate incurs some inevitable loss of optimality. Nevertheless, this issue can be mitigated by exploiting well-known parametric function classes with universal or near-universal approximation properties such as Radial Basis Functions (RBFs) [41], Reproducing Kernel Hilbert Spaces (RKHSs) [42] and Deep Neural Networks (DNNs) [43].

While it is clear that (2) replaces infinite dimensional search by finite dimensional optimization, it is not obvious how (2) can circumvent the need for accurate models. This is addressed in [23], which builds on the observation that the PFA formulation (2) represents a scalarization of a multi-objective statistical learning problem. In fact, each entry of $\mathbf{x}$ 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 greedy reinforcement learning problem [40], [44]–[46], in which $\mathbf{H}$ and $\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta})$ correspond to the state and control actions, respectively. In that sense, it is not only that we can reformulate optimal allocation of resources in wireless systems as a learning problem, but that learning resource allocations is *inherently* a learning problem. This observation led to a primal-dual training method for finding an optimal solution to (2) in [23], which relies on stochastic approximation [47], [48], and attains model-free operation borrowing randomization ideas from policy gradient methods in reinforcement learning [46].

Although the primal-dual learning algorithm of [23] has been shown to work well in some examples, including large scale networks with proper parameterizations [24], issues associated with model-free operation are not addressed. As is the case with policy gradient, the algorithm of [23] requires use of randomized policies. We know that these are inefficient as compared with deterministic policies, but we lack understanding of the loss of optimality associated with specific randomization choices. The main contribution of this paper is to put forth a *principled approach* for solving the PFA (2) via model-free training. We do so by avoiding the use of randomized policies altogether, and instead relying on appropriately constructed, smoothed surrogates to (2), which enable exact zeroth-order gradient representation [49]. This approach not only yields a new, efficient *and* technically grounded model-free training algorithm, but also enables detailed analysis, quantifying the relation of *both* problems (1) *and* (2) to the smoothed surrogate corresponding to the latter, in both primal and dual domains. Specifically, our contributions are as follows.

The Primal Smoothed Surrogate (Section III). We introduce a new *smoothed surrogate* to the *constrained* parameterized problem (2), for which we establish consistency, as well as explicit approximation rates. Our construction leverages recent results on function approximation via Gaussian convolution [49], and ensures that both the objective *and* constraints of the proposed smoothed surrogate approximate those of (2) *uniformly* in their feasible sets, under mild regularity conditions (Lemmata 3 and 4). The quality of the approximation is controlled by user-prescribed, nonnegative smoothing parameters μ_S and μ_R, each associated with the decision variables $\mathbf{x}$ and $\boldsymbol{\theta}$ of (2), respectively. The proposed surrogate exhibits rather desirable properties. First, as either of the smoothing parameters decreases, the corresponding approximation errors shrink, and *at a linear rate*. Second, all smoothed approximations involved are always differentiable, and their gradients may be represented *exactly* as averages of finite differences, which are uniformly stable relative to both μ_S and μ_R. Consequently, such approximations can be exploited to define *zeroth-order stochastic quasi-gradients* of the objective and all constraints of (2), with consistent and predictable behavior. Third, it is possible to establish simple and easily satisfiable conditions *on* (2), which ensure well-definiteness and consistency of the smoothed surrogate, as well as feasibility *within* the feasible sets of both (2) and (1) (Theorems 6 and 7).

The Dual Smoothed Surrogate (Section IV-A). We analyze the dual of our smoothed surrogate *as a smoothed approximation to the dual of* (2). We establish explicit upper and lower bounds on the difference of the respective dual optimal values, with both bounds being *linearly decreasing relative to both* μ_S *and* μ_R (Theorem 12). This result is of independent interest, because it confirms that Gaussian smoothing can be effectively leveraged in the dual domain for the design of general zeroth-order (model-free) methods, applicable to *constrained* programs and, more broadly, problems of the saddle point type.

Duality Gap of Smoothed Surrogates (Section IV-B). Assuming an ϵ-universal policy parameterization, we take [23] strictly one step further by completely characterizing the duality gap between the optimal value of the variational problem (1) and the dual value of the proposed smoothed surrogate. Specifically, we show that the aforementioned duality gap is, in absolute value, of the order of $\mathcal{O}(\mu_S+\mu_R+\epsilon)$ (Theorem 16). If $\mu_S\equiv\mu_R\equiv0$, our duality result recovers exactly that developed earlier in [23], whereas, for $\mu_S>0$ and $\mu_R>0$, it explicitly quantifies the combined effects of *both* policy parameterization *and* smoothing on approximating the optimal value of the original problem (1) via surrogate dualization.

Model-Free Learning (Section V). We develop a new randomized zeroth-order primal-dual algorithm for tackling (2), which exploits the stochastic zeroth-order gradient representation of our proposed smoothed approximations, and fits the desired model-free setting *by construction*. Our primal-dual algorithm is similar to that proposed in [23], *but with a couple of twists*; it takes advantage of our sensitivity and duality analyses and, compared to the policy gradient approach of [23], it requires no policy randomization, and it operates exclusively on probing g^o, $\mathbf{g}$ and the *composition* $\mathbf{f}(\boldsymbol{\phi}(\mathbf{H},\cdot),\mathbf{H})$, *without* the need of computing the gradient of the parametric representation $\boldsymbol{\phi}(\mathbf{H},\cdot)$. Further, the proposed algorithm converges to a stationary point of the dualized smoothed surrogate, the latter satisfying our duality gap guarantees; its optimal value can be made *arbitrarily close* to the optimal value of the original resource allocation problem (1) *at* *will*, by properly selecting smoothing parameters μ_S and μ_R, as well as an ϵ-universal parameterization $\boldsymbol{\phi}$.

Our contributions are also supported by indicative numerical simulations (Section VI), justifying our approach and confirming our theoretical findings. Indeed, our simulations demonstrate near-ideal performance of the proposed model-free method, as compared to both strictly optimal solutions and state-of-the-art heuristics, both relying on availability of explicit system models.

In the analysis that follows, we assume that the feasible set of (1) is nonempty, that $\mathbb{E}\{\mathbf{f}(\mathbf{p}(\mathbf{H}),\mathbf{H})\}$ exists and is finite for every $\mathbf{p}\in\mathcal{P}$, and that the optimal value of (1), $\mathscr{P}^*\in(-\infty,\infty]$, is attained for at least one feasible decision; thus, $\mathscr{P}^*<\infty$. Similar to (1), we assume that (2) has at least one feasible point, as well. Then, if $\mathscr{P}^*_{\boldsymbol{\phi}}\in(-\infty,\infty]$ denotes the optimal value of (2), it follows that $\mathscr{P}^*_{\boldsymbol{\phi}}\le\mathscr{P}^*$, implying that $\mathscr{P}^*_{\boldsymbol{\phi}}<\infty$. For simplicity, we also assume that $\mathscr{P}^*_{\boldsymbol{\phi}}$ is attained within the feasible set of (2).

II. Smoothing via Gaussian Convolution

This section introduces Gaussian smoothing and its properties, and follows closely the corresponding treatment in [49].

Let $f:\mathbf{R}^N\to\mathbf{R}$ be Borel. Also, for any random element $\boldsymbol{U}:\Omega\to\mathbf{R}^N$ following the standard Gaussian measure on $\mathbf{R}^N$, hereafter denoted as $\boldsymbol{U}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_N)$, and for $\mu\ge0$, consider another Borel function $f_\mu:\mathbf{R}^N\to\mathbf{R}$, defined, for every $\mathbf{x}\in\mathbf{R}^N$, as

$$f_\mu(\mathbf{x})\triangleq\mathbb{E}\{f(\mathbf{x}+\mu\boldsymbol{U})\}\equiv\int f(\mathbf{x}+\mu\boldsymbol{u})\mathcal{N}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u},\qquad(3)$$

with $\mathcal{N}:\mathbf{R}^N\to\mathbf{R}$ being the standard Gaussian density, i.e.,

$$\mathcal{N}(\boldsymbol{u})\triangleq(2\pi)^{-N/2}\exp\left(-\|\boldsymbol{u}\|_2^2/2\right),\quad\boldsymbol{u}\in\mathbf{R}^N,\qquad(4)$$

provided that the involved integral is well-defined. For every $\mu>0$, f_μ may be easily shown to be a convolution of the original function f with the Gaussian density on $\mathbf{R}^N$ with mean zero and covariance equal to $\mu^2\boldsymbol{I}_N$. Indeed, for every $\mathbf{x}\in\mathbf{R}^N$, and via a simple change of variables, it is true that

$$\begin{aligned}f_\mu(\mathbf{x})&=\int f(\boldsymbol{u})\,\mu^{-N}\mathcal{N}\left(\frac{\mathbf{x}-\boldsymbol{u}}{\mu}\right)\mathrm{d}\boldsymbol{u}\\&\equiv\left(f*\left[\mu^{-N}\mathcal{N}\left((\cdot)\,\mu^{-1}\right)\right]\right)(\mathbf{x}).\end{aligned}\qquad(5)$$

Therefore, the smoothed function f_μ may be seen as the output of a linear filter whose impulse response is the standard gaussian pulse, taking f as its input.

In many cases, f_μ turns out to be everywhere differentiable on $\mathbf{R}^N$, *even if f is not*, whereas the gradient of f_μ admits a *zeroth-order representation*. In particular, such is the case of all Lipschitz functions on $\mathbf{R}^N$ [49], as the next result suggests.

Lemma 1. (Properties of f_μ [49]) *Let $\boldsymbol{U}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_N)$, and consider any globally Lipschitz function $f:\mathbf{R}^N\to\mathbf{R}$. Then, for any $\mathcal{F}\subseteq\mathbf{R}^N$, the following statements are true:*

- *For every $\mu\ge0$, f_μ is well-defined and finite on $\mathcal{F}$, and*

$$\sup_{\mathbf{x}\in\mathcal{F}}|f_\mu(\mathbf{x})-f(\mathbf{x})|\le\mu L\sqrt{N}.\qquad(6)$$

- *If f is convex on $\mathbf{R}^N$, so is f_μ, and $f_\mu\ge f$ on $\mathcal{F}$.*
- *For every $\mu>0$, f_μ is differentiable on $\mathcal{F}$, and its gradient $\nabla f_\mu:\mathbf{R}^N\to\mathbf{R}^N$ admits the representation*

$$\nabla f_\mu(\mathbf{x})\equiv\mathbb{E}\left\{\frac{f(\mathbf{x}+\mu\boldsymbol{U})-f(\mathbf{x})}{\mu}\boldsymbol{U}\right\},\qquad(7)$$

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

$$\sup_{\mathbf{x}\in\mathcal{F}}\mathbb{E}\left\{\left\|\frac{f(\mathbf{x}+\mu\boldsymbol{U})-f(\mathbf{x})}{\mu}\boldsymbol{U}\right\|_2^2\right\}\le L^2(N+4)^2.\qquad(8)$$

Lemma 1 will be key to the results presented in this paper, as discussed in detail as follows.

III. Smoothed Constrained Program Surrogates

In this section, we introduce a new, *smoothed surrogate* of the *whole* constrained program (2), as promised in Section I, leveraging the results of Section II. We also introduce conditions under which this smoothed surrogate is well-defined, and establish various of its properties, as well as its structural relation to (2). The power of the proposed surrogate is in that it provides a technically grounded means for dealing with (2) in the model-free setting, i.e., when the functions g^o, $\mathbf{g}$ and $\mathbf{f}$ are apriori unknown, and may be only observed through *limited probing*.

A. Surrogate Construction

Let $\mu_S\ge0$, $\mu_R\ge0$, and consider random elements $\boldsymbol{U}_S\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_{N_S})$ and $\boldsymbol{U}_R\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_{N_\phi})$, the latter taken independent of $\mathbf{H}$. Driven by the results of Section II, we define smoothed versions of g^o, $\mathbf{g}$ and $\mathbb{E}\{\mathbf{f}(\boldsymbol{\phi}(\mathbf{H},\cdot),\mathbf{H})\}\triangleq\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\cdot)$ as

$$g^o_{\mu_S}(\mathbf{x})\triangleq\mathbb{E}\{g^o(\mathbf{x}+\mu_S\boldsymbol{U}_S)\},\quad\mathbf{x}\in\mathcal{X},\qquad(9)$$

$$\mathbf{g}_{\mu_S}(\mathbf{x})\triangleq\mathbb{E}\{\mathbf{g}(\mathbf{x}+\mu_S\boldsymbol{U}_S)\},\quad\mathbf{x}\in\mathcal{X}\quad\text{and}\qquad(10)$$

$$\overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})\triangleq\mathbb{E}\{\mathbf{f}(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}+\mu_R\boldsymbol{U}_R),\mathbf{H})\},\quad\boldsymbol{\theta}\in\Theta,\qquad(11)$$

where, at this point, we arbitrarily assume that the involved expectations are well-defined and finite on $\mathcal{X}$ and Θ. We will return to those issues shortly. Then, we may formulate a (hopefully) smoothed version of problem (2) as

$$\begin{array}{ll}\underset{\mathbf{x},\boldsymbol{\theta}}{\text{maximize}} & g^o_{\mu_S}(\mathbf{x})\\ \text{subject to} & \mathbf{x}+\mathtt{S}(\mu_R)\le\overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})\\ & \mathbf{g}_{\mu_S}(\mathbf{x})\ge\mathbf{0},\ (\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta\end{array},\qquad(12)$$

where $\mathtt{S}:\mathbf{R}_+\to\mathbf{R}_+^{N_S}$ is a nonnegative *feasibility slack*, with properties to be determined. Formulation of the smoothed surrogate (12) is well motivated due to the fact that, whenever the objective g^o and all entries of the constraint vector functions $\mathbf{g}$ and $\mathbf{f}(\boldsymbol{\phi}(\mathbf{H},\cdot),\mathbf{H})$ are sufficiently well-behaved, such that Lemma 1 appropriately applies, the smoothed functions $g^o_{\mu_S}$, $\mathbf{g}_{\mu_S}$ and $\overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}$ are differentiable, and the respective gradients may be represented as averages of suitably defined finite differences. This is particularly important in developing effective and predictable methods for solving problem (2) in the model-free setting: Finite differences are by construction based on function evaluations only. Thus, the surrogate (12) constitutes a natural *zeroth-order proxy* for dealing with the original parameterized problem (2).

However, before focusing on how to use (12) in order to solve (2), we have to make sure that (12) is a well-defined *and* feasible problem, and also reveal its fundamental connection to (2). These tasks are the subject of the rest of this section.

B. Smoothing $g^o, \mathbf{g}$ and $\overline{\mathbf{f}}^{\phi}$

Our treatment will require imposing appropriate structure on the functions involved in (2), as we now discuss in detail. Hereafter, the i-th entries of $\mathbf{g}$ (resp. $\mathbf{g}_{\mu_S}$) and $\mathbf{f}$ (resp. $\overline{\mathbf{f}}^{\phi}_{\mu_R}$) will be denoted as g^i ($g^i_{\mu_S}$), $i \in \mathbb{N}^+_{N_\mathbf{g}}$ and f^i ($\overline{f}^{\phi,i}_{\mu_R}$), $i \in \mathbb{N}^+_{N_S}$, respectively.

Assumption 1. *The following conditions are satisfied:*
C1 *For every $i \in \{o, \mathbb{N}^+_{N_\mathbf{g}}\}$, g^i is L^i_g-Lipschitz on $\mathbf{R}^{N_S}$.*
C2 *For every $i \in \mathbb{N}^+_{N_S}$, there is $L^i_f < \infty$, such that*

$$\begin{aligned}\|f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}_1),\mathbf{H}) - f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}_2),\mathbf{H})\|_{\mathcal{L}_2}&\\ \le L^i_f\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2, \forall (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\in \mathbf{R}^{N_\phi}\times\mathbf{R}^{N_\phi}.&\end{aligned} \tag{13}$$

Condition **C2** of Assumption 1 has the following consequences on the behavior of $\mathbb{E}\{\mathbf{f}(\boldsymbol{\phi}(\mathbf{H},\cdot),\mathbf{H})\} \equiv \overline{\mathbf{f}}^{\phi}(\cdot)$.

Proposition 2. (Properties of $\overline{\mathbf{f}}^{\phi}$) *Suppose that condition* **C2** *of Assumption 1 is in effect. Then, for every $i \in \mathbb{N}^+_{N_S}$, $\overline{f}^{\phi,i}$ is L^i_f-Lipschitz on $\mathbf{R}^{N_\phi}$. Additionally, it is true that*

$$\begin{aligned}&\mathbb{E}\{|f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}+\boldsymbol{u}),\mathbf{H})|\}\\ &\qquad\le L^i_f\|\boldsymbol{u}\|_2 + \mathbb{E}\{|f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}),\mathbf{H})|\},\end{aligned} \tag{14}$$

for all $(\boldsymbol{\theta},\boldsymbol{u}) \in \Theta \times \mathbf{R}^{N_\phi}$, and for all $i \in \mathbb{N}^+_{N_S}$.

Proof of Proposition 2: The first part of the result follows immediately from condition **C2**, by the nested structure of $\mathcal{L}_p$-spaces, and Jensen. The second part follows via an application of the triangle inequality. ■

Assumption 1 and Proposition 2 may be further exploited to establish well-definiteness and basic properties of $g^o_{\mu_S}$, $\mathbf{g}_{\mu_S}$ and $\overline{\mathbf{f}}^{\phi}_{\mu_R}$. To this end, for $\mathbf{x} \in \mathcal{X}$, $\mu_S > 0$ and for every $i \in \{o, \mathbb{N}^+_{N_\mathbf{g}}\}$, let us define finite differences

$$\Delta^i_g(\mathbf{x},\mu_S,\boldsymbol{U}_S) \triangleq \frac{g^i(\mathbf{x}+\mu_S\boldsymbol{U}_S) - g^i(\mathbf{x})}{\mu_S}. \tag{15}$$

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

$$\begin{aligned}&\Delta^i_f(\boldsymbol{\theta},\mu_R,\boldsymbol{U}_R,\mathbf{H})\\ &\quad\triangleq \frac{f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}+\mu_R\boldsymbol{U}_R),\mathbf{H}) - f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}),\mathbf{H})}{\mu_R}.\end{aligned} \tag{16}$$

The relevant results now follow.

Lemma 3. (Existence & Properties of $g^o_{\mu_S}$ and $\mathbf{g}_{\mu_S}$) *Suppose that Assumption 1 is in effect. Then, for every $i \in \{o, \mathbb{N}^+_{N_\mathbf{g}}\}$ and for every $\mu_S > 0$, each $g^i_{\mu_S}$ is a well-defined, finite, concave and everywhere differentiable underestimator of g^i on $\mathcal{X}$, such that*

$$\sup_{\mathbf{x}\in\mathcal{X}} |g^i_{\mu_S}(\mathbf{x}) - g^i(\mathbf{x})| \le \mu_S L^i_g\sqrt{N_S}, \tag{17}$$

$$\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}\{\|\Delta^i_g(\mathbf{x},\mu_S,\boldsymbol{U}_S)\boldsymbol{U}_S\|_2^2\} \le (L^i_g)^2(N_S+4)^2 \tag{18}$$

$$\textit{and}\quad \mathbb{E}\{\Delta^i_g(\mathbf{x},\mu_S,\boldsymbol{U}_S)\boldsymbol{U}_S\} \equiv \nabla g^i_{\mu_S}(\mathbf{x}), \tag{19}$$

for all $\mathbf{x} \in \mathcal{X}$.

Proof of Lemma 3: Trivial, see Lemma 1 (Section II). ■

Lemma 4. (Existence & Properties of $\overline{\mathbf{f}}^{\phi}_{\mu_R}$) *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 and differentiable everywhere on Θ, such that*

$$\sup_{\boldsymbol{\theta}\in\Theta} |\overline{f}^{\phi,i}_{\mu_R}(\boldsymbol{\theta}) - \overline{f}^{\phi,i}(\boldsymbol{\theta})| \le \mu_R L^i_f\sqrt{N_\phi}, \tag{20}$$

$$\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\{\|\Delta^i_f(\boldsymbol{\theta},\mu_R,\boldsymbol{U}_R,\mathbf{H})\boldsymbol{U}_R\|_2^2\} \le (L^i_f)^2(N_\phi+4)^2 \tag{21}$$

$$\textit{and}\quad \mathbb{E}\{\Delta^i_f(\boldsymbol{\theta},\mu_R,\boldsymbol{U}_R,\mathbf{H})\boldsymbol{U}_R\} \equiv \nabla\overline{f}^{\phi,i}_{\mu_R}(\boldsymbol{\theta}), \tag{22}$$

for all $\boldsymbol{\theta} \in \Theta$.

Proof of Lemma 4: Fix $i \in \mathbb{N}^+_{N_S}$, and consider the function $\overline{f}^{\phi,i}(\cdot) \equiv \mathbb{E}\{f^i(\boldsymbol{\phi}(\mathbf{H},\cdot),\mathbf{H})\}$, which, by Proposition 2, is L^i_f-Lipschitz on $\mathbf{R}^{N_\phi}$. Then, Lemma 1 implies that, for every $\mu_R \ge 0$,

$$\sup_{\boldsymbol{\theta}\in\Theta} |\mathbb{E}\{\overline{f}^{\phi,i}(\boldsymbol{\theta}+\mu_R\boldsymbol{U}_R)\} - \overline{f}^{\phi,i}(\boldsymbol{\theta})| \le \mu_R L^i_f\sqrt{N_\phi}. \tag{23}$$

Note that we are not done yet, since $\mathbb{E}\{\overline{f}^{\phi,i}((\cdot)+\mu_R\boldsymbol{U}_R)\}$ involves an iterated expectation, and *not* an expectation relative to the joint distribution of $\boldsymbol{U}_R$ and $\mathbf{H}$. However, again by Proposition 2, it follows that, for every $\boldsymbol{\theta} \in \Theta$,

$$\begin{aligned}&\int \mathbb{E}\{|f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}+\mu_R\boldsymbol{u}),\mathbf{H})|\}\mathcal{P}_{\boldsymbol{U}_R}(\mathrm{d}\boldsymbol{u})\\ &\quad\le \mu_R L^i_f\int \|\boldsymbol{u}\|_2\mathcal{P}_{\boldsymbol{U}_R}(\mathrm{d}\boldsymbol{u}) + \mathbb{E}\{|f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}),\mathbf{H})|\}\\ &\quad\le \mu_R L^i_f\sqrt{N_\phi} + \mathbb{E}\{|f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}),\mathbf{H})|\} < \infty.\end{aligned} \tag{24}$$

Then, Fubini's Theorem (Corollary 2.6.5 and Theorem 2.6.6 in [50]) implies that $\overline{f}^{\phi,i}_{\mu_R}(\cdot) \equiv \mathbb{E}\{f^i(\boldsymbol{\phi}(\mathbf{H},(\cdot)+\mu_R\boldsymbol{U}_R),\mathbf{H})\}$ is finite on Θ, and that

$$\begin{aligned}&\mathbb{E}\{\overline{f}^{\phi,i}(\boldsymbol{\theta}+\mu_R\boldsymbol{U}_R)\}\\ &\quad\equiv \int \mathbb{E}\{f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}+\mu_R\boldsymbol{u}),\mathbf{H})\}\mathcal{P}_{\boldsymbol{U}_R}(\mathrm{d}\boldsymbol{u})\\ &\quad\equiv \int f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}+\mu_R\boldsymbol{u}),\mathbf{H})\left[\mathcal{P}_{\mathbf{H}}\times\mathcal{P}_{\boldsymbol{U}_R}\right](\mathrm{d}\left[\boldsymbol{u},\mathbf{H}\right])\\ &\quad\equiv \overline{f}^{\phi,i}_{\mu_R}(\boldsymbol{\theta}),\quad \forall\boldsymbol{\theta}\in\Theta,\end{aligned} \tag{25}$$

where $\mathbf{H}$ and $\boldsymbol{U}_R$ are statistically independent by assumption; now we are done. Next, differentiability of $\overline{f}^{\phi,i}_{\mu_R}$, as well as the form of its gradient also follow from Lemma 1 on $\overline{f}^{\phi,i}$ and, again, (25). Finally, to verify (21), we may write (due to condition **C2**)

$$\begin{aligned}&\mathbb{E}\{\|\Delta^i_f(\boldsymbol{\theta},\mu_R,\boldsymbol{U}_R,\mathbf{H})\boldsymbol{U}_R\|_2^2\}\\ &\quad\equiv \frac{1}{\mu_R^2}\mathbb{E}\{\mathbb{E}\{|f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}+\mu_R\boldsymbol{U}_R),\mathbf{H})\\ &\qquad\qquad - f^i(\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}),\mathbf{H})|^2|\boldsymbol{U}_R\}\|\boldsymbol{U}_R\|_2^2\}\\ &\quad\le \frac{1}{\mu_R^2}\mathbb{E}\{(L^i_f)^2\|\mu_R\boldsymbol{U}_R\|_2^2\|\boldsymbol{U}_R\|_2^2\}\\ &\quad\le (L^i_f)^2(N_\phi+4)^2,\end{aligned} \tag{26}$$

as required. The proof is complete. ■

Remark 5. We would like to mention that a weaker version of Lemma 3 holds if we weaken condition **C2** of Assumption

1, by replacing the $\mathcal{L}_2$-norm with an $\mathcal{L}_1$-norm. In this case, the ℓ_2-norm-squared inside the expectation of (21) would be replaced by a mere ℓ_2-norm; essentially, only boundedness of $\|\Delta_f^i(\boldsymbol{\theta}, \mu_R, \boldsymbol{U}_R, \mathbf{H})\boldsymbol{U}_R\|_2$ in $\mathcal{L}_1$ would be guaranteed, instead of boundedness in $\mathcal{L}_2$. The main reason why (21) (and therefore condition **C2**) is desirable is that it crucially affects the behavior of gradient-based algorithms for solving problems such as (12), considered later in this work. ∎

C. Surrogate Feasibility

We are now in place to investigate conditions ensuring feasibility of the smoothed surrogate (12). In particular, we will be interested in conditions which ensure feasibility of (12), *but are on the original parameterized problem* (2). This is important from a practical point of view, since the exact form of (12) will be, in most cases, unknown. On the other hand, feasibility of (2) should be somehow guaranteed apriori, *even in the model-free setting*; indeed, *both* resource allocation problems (1) and (2) are initially proposed by the wireless engineer, who is the one responsible for formulating meaningful resource allocation tasks.

It turns out that all that is needed for (12) to be feasible is the existence of at least one *strictly* feasible point for (2). What is more, (12) can be made strictly feasible at will. The relevant result follows right after we define the vectors

$$\boldsymbol{c}_S \triangleq \left[L_g^1 \, \ldots \, L_g^{N_{\mathbf{g}}}\right]^{\boldsymbol{T}} \quad \text{and} \quad \boldsymbol{c}_R \triangleq \left[L_f^1 \, \ldots \, L_f^{N_S}\right]^{\boldsymbol{T}}, \tag{27}$$

and under the following assumption.

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

Theorem 6. (Surrogate Strict Feasibility) *Let Assumptions 1 and 2 be in effect, and suppose that* $(\mathbf{x}^\dagger, \boldsymbol{\theta}^\dagger) \in \mathbf{R}^{N_S} \times \mathbf{R}^{N_\phi}$ *is a strictly feasible point of the parameterized problem* (2)*. Then there exist* $\mu_S^\dagger > 0$ *and* $\mu_R^\dagger > 0$*, possibly dependent on* $(\mathbf{x}^\dagger, \boldsymbol{\theta}^\dagger)$*, such that, for every* $0 \le \mu_S \le \mu_S^\dagger$ *and* $0 \le \mu_R \le \mu_R^\dagger$*, the same point* $(\mathbf{x}^\dagger, \boldsymbol{\theta}^\dagger)$ *is strictly feasible for the smoothed surrogate* (12)*.*

Proof of Theorem 6: Let the point $(\mathbf{x}^\dagger, \boldsymbol{\theta}^\dagger) \in \mathbf{R}^{N_S} \times \mathbf{R}^{N_\phi}$ be strictly feasible for (2), implying that $(\mathbf{x}^\dagger, \boldsymbol{\theta}^\dagger) \in \mathcal{X} \times \Theta$, and

$$\mathbf{g}(\mathbf{x}^\dagger) \ge \mathtt{s}_{\mathbf{g}}^\dagger \quad \text{and} \quad \overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta}^\dagger) - \mathbf{x}^\dagger \ge \mathtt{s}_{\mathbf{f}}^\dagger, \tag{28}$$

for some positive slacks $\mathtt{s}_{\mathbf{g}}^\dagger \in \mathbf{R}_{++}^{N_{\mathbf{g}}}$ and $\mathtt{s}_{\mathbf{f}}^\dagger \in \mathbf{R}_{++}^{N_S}$. Also, from Lemma 3, it follows that, for every $\mu_S \ge 0$ and $\mu_R \ge 0$,

$$\mathbf{g}(\mathbf{x}^\dagger) \le \mathbf{g}_{\mu_S}(\mathbf{x}^\dagger) + \mu_S \boldsymbol{c}_S \sqrt{N_S} \quad \text{and} \tag{29}$$

$$\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta}^\dagger) \le \overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}^\dagger) + \mu_R \boldsymbol{c}_R \sqrt{N_{\boldsymbol{\phi}}}, \tag{30}$$

Consequently, it is true that

$$\mathbf{g}_{\mu_S}(\mathbf{x}^\dagger) \ge \mathtt{s}_{\mathbf{g}}^\dagger - \mu_S \boldsymbol{c}_S \sqrt{N_S} \quad \text{and} \tag{31}$$

$$\overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}^\dagger) - \mathbf{x}^\dagger \ge \mathtt{s}_{\mathbf{f}}^\dagger - \mu_R \boldsymbol{c}_R \sqrt{N_{\boldsymbol{\phi}}}. \tag{32}$$

Therefore, we can find $\mu_S^\dagger > 0$ and $\mu_R^\dagger > 0$ *sufficiently small but strictly positive*, such that, for every $0 \le \mu_S \le \mu_S^\dagger$ and $0 \le \mu_R \le \mu_R^\dagger$, the strict inequalities $\mathtt{s}_{\mathbf{g}}^\dagger - \mu_S \boldsymbol{c}_S \sqrt{N_S} > \mathbf{0}$ and $\mathtt{s}_{\mathbf{f}}^\dagger - \mu_R \boldsymbol{c}_R \sqrt{N_{\boldsymbol{\phi}}} > \mathtt{S}(\mu_R)$ hold. This, of course, implies that $(\mathbf{x}^\dagger, \boldsymbol{\theta}^\dagger)$ is a strictly feasible point for problem (12), for all aforementioned choices of μ_S and μ_R. ∎

Theorem 6 is important, as it confirms the existence of a strictly feasible point for problem (12), *uniformly relative to* μ_S *and* μ_R, the latter being allowed to vary in appropriate sets, whose length is controlled by the feasibility of (2) and the feasibility slack of (12). An evident byproduct of Theorem 6 is that (12) is a feasible and, therefore, meaningful optimization problem.

Another similar question we may ask is how much the constraints of (2) are violated for every feasible solution of (12). In this respect, we may formulate the following result.

Theorem 7. (PFA Constraint Violation) *Let Assumption 1 be in effect. Then, for every* $\mu_R \ge 0$ *such that*

$$\mathtt{S}(\mu_R) - \mu_R \boldsymbol{c}_R \sqrt{N_{\boldsymbol{\phi}}} \ge \mathbf{0}, \tag{33}$$

and for every $\mu_S \ge 0$*, every feasible point of* (12) *is also feasible for* (2)*. Otherwise, if* (33) *fails to hold, then the negative values of its left-hand-side correspond to the respective levels of maximal constraint violation for* (2)*.*

Proof of Theorem 7: Fix qualifying μ_S and μ_R, and let the point $(\mathbf{x}^{\mu_S}, \boldsymbol{\theta}^{\mu_R}) \in \mathbf{R}^{N_S} \times \mathbf{R}^{N_\phi}$ be feasible for problem (12). Then, it is in fact true that $(\mathbf{x}^{\mu_S}, \boldsymbol{\theta}^{\mu_R}) \in \mathcal{X} \times \Theta$, whereas from Lemma 3 it follows that

$$\mathbf{g}(\mathbf{x}^{\mu_S}) \ge \mathbf{g}_{\mu_S}(\mathbf{x}^{\mu_S}) \ge \mathbf{0} \quad \text{and} \tag{34}$$

$$\begin{aligned}\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta}^{\mu_R}) - \mathbf{x}^{\mu_S} + \mu_R \boldsymbol{c}_R \sqrt{N_{\boldsymbol{\phi}}} &\ge \overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}^{\mu_R}) - \mathbf{x}^{\mu_S} \\ &\ge \mathtt{S}(\mu_R).\end{aligned} \tag{35}$$

Rearranging the second inequality, we obtain

$$\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta}^{\mu_R}) - \mathbf{x}^{\mu_S} \ge \mathtt{S}(\mu_R) - \mu_R \boldsymbol{c}_R \sqrt{N_{\boldsymbol{\phi}}}, \tag{36}$$

where the right-hand-sides are independent of the feasible point $(\mathbf{x}^{\mu_S}, \boldsymbol{\theta}^{\mu_R})$. The result now readily follows. ∎

It would be useful to note that if $\mathtt{S}$ is such that condition (33) is satisfied for *all* qualifying μ_R, then feasibility of (2) is ensured *uniformly relative to the choice of* μ_R *(and* μ_S*)*. This means that, whenever a solution to (12) is determined, this solution will automatically satisfy the original resource constraints of the initial parameterized problem (2).

Another important observation is that Theorems 6 and 7 are *not exclusive*; in other words, they can hold simultaneously. Indeed, the former concerns choosing μ_S and μ_R, whereas the latter concerns choosing the slack $\mathtt{S}$, which is a function μ_R, in a way which is compatible with Assumption 2.

As an example, one can set $\mathtt{S}(\mu_R) \equiv \mu_R \boldsymbol{c}_R \sqrt{N_{\boldsymbol{\phi}}}$, where $\mathtt{S}$ readily satisfies Assumption 2. However, this might not be a feasible choice in practice, since the entries of $\boldsymbol{c}_R$ will probably be unknown. Still, Theorem 7 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 $\boldsymbol{C} > \mathbf{0}$, which may chosen experimentally. This last point also highlights the operational importance of Theorem 7.

IV. Lagrangian Duality

A promising approach for dealing with the explicit constraints of either problems (2) or (12) is by exploiting *Lagrangian Duality*, which has been proven essential and undoubtedly important in analyzing and efficiently solving constrained convex optimization problems; see, e.g., [51]–[53]. Note, however, that, since both problems (2) and (12) are typically nonconvex, most standard results in Lagrangian Duality for convex optimization do not apply automatically.

Instead, our treatment will be based on recent results reported in [23], which in turn rely on earlier results reported in [1]. In particular, the purpose of this section is to explicitly link the smoothed surrogate (12) to the parameterized problem (2), and ultimately to the base policy search problem (1), *in the dual domain*, effectively characterizing the respective duality gaps. Our results essentially provide a technically grounded path to dealing with the constrained problem (1) in the model-free setting, through the zeroth-order proxy (12).

To this end, consider the *Lagrangian function* $\mathscr{L}_{\boldsymbol{\phi}}:\mathbf{R}^{N_S}\times\mathbf{R}^{N_{\boldsymbol{\phi}}}\times\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}\to\mathbf{R}$ defined as

$$\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})\triangleq g^o(\mathbf{x})+\langle\boldsymbol{\lambda}_S,\mathbf{g}(\mathbf{x})\rangle+\langle\boldsymbol{\lambda}_R,\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta})-\mathbf{x}\rangle,\quad(37)$$

where $\boldsymbol{\lambda}\equiv(\boldsymbol{\lambda}_S,\boldsymbol{\lambda}_R)\in\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}$ are multipliers associated with the respective constraint of the *primal problem* (2). Then the *dual function* $\mathscr{D}_{\boldsymbol{\phi}}:\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}\to(-\infty,\infty]$ is defined as

$$\mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda})\triangleq\sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta}\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}).\quad(38)$$

Since it is true that $\mathscr{P}^*_{\boldsymbol{\phi}}\leq\mathscr{D}_{\boldsymbol{\phi}}$ *on* $\mathbf{R}^{N_{\mathbf{g}}}_+\times\mathbf{R}^{N_S}_+$, it is most reasonable to consider the *dual problem*

$$\begin{array}{ll}\text{minimize} & \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda})\\ \text{subject to} & \boldsymbol{\lambda}\geq\mathbf{0}\end{array},\quad(39)$$

whose optimal value

$$\mathscr{D}^*_{\boldsymbol{\phi}}\triangleq\inf_{\boldsymbol{\lambda}\geq\mathbf{0}}\mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda})\in(-\infty,\infty]\quad(40)$$

serves as the tightest over-estimate of the optimal value of (2), $\mathscr{P}^*_{\boldsymbol{\phi}}$, when knowing only $\mathscr{D}_{\boldsymbol{\phi}}$.

In the same fashion, for $\mu_S>0$ and $\mu_R>0$, we define the Lagrangian function $\mathscr{L}_{\mu}:\mathcal{X}\times\Theta\times\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}\to\mathbf{R}$ associated with the smoothed surrogate (12) as

$$\begin{aligned}\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})&\triangleq g^o_{\mu_S}(\mathbf{x})+\langle\boldsymbol{\lambda}_S,\mathbf{g}_{\mu_S}(\mathbf{x})\rangle\\&\quad+\langle\boldsymbol{\lambda}_R,\overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})-\mathbf{x}-\mathtt{S}(\mu_R)\rangle,\end{aligned}\quad(41)$$

whereas the dual function $\mathscr{D}_{\boldsymbol{\phi},\mu}:\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}\to(-\infty,\infty]$ and corresponding dual infimal value are

$$\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda})\triangleq\sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta}\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})\quad\text{and}\quad(42)$$

$$\mathscr{D}^*_{\boldsymbol{\phi},\mu}\triangleq\inf_{\boldsymbol{\lambda}\geq\mathbf{0}}\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda})\in(-\infty,\infty],\quad(43)$$

respectively. Note that the both $\mathscr{D}_{\boldsymbol{\phi}}$ and $\mathscr{D}_{\boldsymbol{\phi},\mu}$ are convex on $\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}$, as pointwise suprema of affine functions. In our analysis, we will exploit another basic assumption, as follows.

Assumption 3. *Problem* (2) *is strictly feasible.*

Under Assumption 3, it is true that the base problem (1) is strictly feasible as well; its feasible set contains that of (2).

A. Dual Optimal Values

Our first task will be to explicitly relate the optimal dual values $\mathscr{D}^*_{\boldsymbol{\phi}}$ and $\mathscr{D}^*_{\boldsymbol{\phi},\mu}$. To do so, we develop and exploit the following technical results.

Lemma 8. (Lagrangian Approximation) *Let Assumption 1 be in effect, and for every $\mu_S\geq0$, $\mu_R\geq0$ and $\boldsymbol{\lambda}\geq\mathbf{0}$, define the nonnegative quantities*

$$\begin{aligned}\Gamma^l_\mu(\boldsymbol{\lambda})&\triangleq\mu_SL^o_g\sqrt{N_S}+\mu_S\langle\boldsymbol{\lambda}_S,\boldsymbol{c}_S\rangle\sqrt{N_S}\\&\quad+\mu_R\langle\boldsymbol{\lambda}_R,\boldsymbol{c}_R\rangle\sqrt{N_{\boldsymbol{\phi}}}+\langle\mathtt{S}(\mu_R),\boldsymbol{\lambda}_R\rangle,\ \text{and}\end{aligned}\quad(44)$$

$$\Gamma^r_\mu(\boldsymbol{\lambda})\triangleq\mu_R\langle\boldsymbol{\lambda}_R,\boldsymbol{c}_R\rangle\sqrt{N_{\boldsymbol{\phi}}}-\langle\mathtt{S}(\mu_R),\boldsymbol{\lambda}_R\rangle.\quad(45)$$

Then, for every $(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta$, it is true that

$$-\Gamma^l_\mu(\boldsymbol{\lambda})\leq\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})-\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})\leq\Gamma^r_\mu(\boldsymbol{\lambda}).\quad(46)$$

Proof of Lemma 8: Let $\mu_S\geq0$ and $\mu_R\geq0$. Since Assumption 1 is in effect, Lemma 3 implies that, for every $(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})\in\mathcal{X}\times\Theta\times\mathbf{R}^{N_{\mathbf{g}}}_+\times\mathbf{R}^{N_S}_+$,

$$\begin{aligned}&\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})\\&\equiv g^o_{\mu_S}(\mathbf{x})+\langle\boldsymbol{\lambda}_S,\mathbf{g}_{\mu_S}(\mathbf{x})\rangle+\langle\boldsymbol{\lambda}_R,\overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})-\mathbf{x}-\mathtt{S}(\mu_R)\rangle\\&\leq g^o(\mathbf{x})+\langle\boldsymbol{\lambda}_S,\mathbf{g}(\mathbf{x})\rangle+\langle\boldsymbol{\lambda}_R,\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta})-\mathbf{x}\rangle\\&\quad+\langle\boldsymbol{\lambda}_R,\mu_R\boldsymbol{c}_R\sqrt{N_{\boldsymbol{\phi}}}-\mathtt{S}(\mu_R)\rangle\\&\equiv\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})+\Gamma^r_\mu(\boldsymbol{\lambda}).\end{aligned}\quad(47)$$

By symmetry, a similar argument is possible for $\mathscr{L}_{\boldsymbol{\phi}}$, namely,

$$\begin{aligned}&\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})\\&\equiv g^o(\mathbf{x})+\langle\boldsymbol{\lambda}_S,\mathbf{g}(\mathbf{x})\rangle+\langle\boldsymbol{\lambda}_R,\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta})-\mathbf{x}\rangle\\&\leq g^o_{\mu_S}(\mathbf{x})+\langle\boldsymbol{\lambda}_S,\mathbf{g}_{\mu_S}(\mathbf{x})\rangle+\langle\boldsymbol{\lambda}_R,\overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})-\mathbf{x}-\mathtt{S}(\mu_R)\rangle\\&\quad+\mu_SL^o_g\sqrt{N_S}+\langle\boldsymbol{\lambda}_S,\mu_S\boldsymbol{c}_S\sqrt{N_S}\rangle\\&\qquad+\langle\boldsymbol{\lambda}_R,\mu_R\boldsymbol{c}_R\sqrt{N_{\boldsymbol{\phi}}}\rangle+\langle\boldsymbol{\lambda}_R,\mathtt{S}(\mu_R)\rangle\\&\equiv\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})+\Gamma^l_\mu(\boldsymbol{\lambda}).\end{aligned}\quad(48)$$

Rearranging (47) and (48) gives the result. ∎

Lemma 9. (Dual Functions are Proper & Closed) *As long as $\mathscr{D}^*_{\boldsymbol{\phi}}<\infty$, the dual function $\mathscr{D}_{\boldsymbol{\phi}}$ is proper and closed. If, additionally, Assumption 1 is in effect, then, for every $\mu_S>0$ and $\mu_R>0$, the smoothed dual function $\mathscr{D}_{\boldsymbol{\phi},\mu}$ is also proper and closed.*

Proof of Lemma 9: Since $\mathscr{D}^*_{\boldsymbol{\phi}}<\infty$, there exists a dual feasible point $\boldsymbol{\lambda}^\ddagger\geq\mathbf{0}$ such that $\mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^\ddagger)<\infty$, whereas the fact that $\mathscr{L}_{\boldsymbol{\phi}}$ is real-valued on its domain implies that $\mathscr{D}_{\boldsymbol{\phi}}>-\infty$ on $\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}$. Then $\mathscr{D}_{\boldsymbol{\phi}}$ is proper, by definition (p. 7 in [53]).

To show that $\mathscr{D}_{\boldsymbol{\phi}}$ is also closed, it suffices to observe that it is the pointwise supremum of affine functions, each of which is continuous (thus lower semicontinuous) on the closed set $\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}$, and, therefore, closed (Proposition 1.1.3 in [53]). Then $\mathscr{D}_{\boldsymbol{\phi}}$ must be closed, by ([53], Proposition 1.1.6).

If now Assumption 1 is in effect, then for the same dual feasible point $\boldsymbol{\lambda}^\ddagger$ as above, and for every $(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta$, Lemma 8 implies that

$$\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}^\ddagger)\leq\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}^\ddagger)+\Gamma^r_\mu(\boldsymbol{\lambda}^\ddagger).\quad(49)$$

Therefore, it follows that

$$\begin{aligned}\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}^{\ddagger}) &\equiv \sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}^{\ddagger}) \\ &\le \left[\sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}^{\ddagger})\right] + \Gamma^r_\mu(\boldsymbol{\lambda}^{\ddagger}) \\ &\equiv \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^{\ddagger}) + \Gamma^r_\mu(\boldsymbol{\lambda}^{\ddagger}) < \infty. \end{aligned} \tag{50}$$

As before, $\mathscr{L}_{\boldsymbol{\phi},\mu}$ is real-valued on its domain, thus $\mathscr{D}_{\boldsymbol{\phi},\mu} > -\infty$ everywhere on $\mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S}$, showing that $\mathscr{D}_{\boldsymbol{\phi},\mu}$ is proper. Lastly, closeness of $\mathscr{D}_{\boldsymbol{\phi},\mu}$ follows by the same argument as that for $\mathscr{D}_{\boldsymbol{\phi}}$, above. ■

Lemma 10. (Existence of Dual Optimal Solutions) *Suppose that $\mathscr{D}^*_{\boldsymbol{\phi}} < \infty$, and let Assumption 3 be in effect. Then, the set of dual optimal solutions $\arg\min_{\boldsymbol{\lambda}\ge\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda})$ is nonempty and compact in $\mathbf{R}^{N_{\mathbf{g}}}_+ \times \mathbf{R}^{N_S}_+$. If, additionally, Assumptions 1 and 2 are in effect, then there exist $\mu^{\dagger}_S > 0$ and $\mu^{\dagger}_R > 0$, such that, for every $0 < \mu_S \le \mu^{\dagger}_S$ and $0 < \mu_R \le \mu^{\dagger}_R$, the solution set $\arg\min_{\boldsymbol{\lambda}\ge\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda})$ is a nonempty and compact subset of $\mathbf{R}^{N_{\mathbf{g}}}_+ \times \mathbf{R}^{N_S}_+$, as well.*

Proof of Lemma 10: For any point $(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger}) \in \mathbf{R}^{N_S}\times\mathbf{R}^{N_{\boldsymbol{\phi}}}$, which is strictly feasible for (2), it is true that $\mathbf{g}(\mathbf{x}^{\dagger}) > \mathbf{0}$ and $\overline{\mathbf{f}}^{\boldsymbol{\phi}}(\boldsymbol{\theta}^{\dagger}) - \mathbf{x}^{\dagger} > \mathbf{0}$. Then, for every $\boldsymbol{\lambda}\ge\mathbf{0}$, we have

$$\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger},\boldsymbol{\lambda}) \le \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}), \tag{51}$$

where the left-hand-side is *non-trivially affine in* $\boldsymbol{\lambda}$. Now, consider any sequence $\{\boldsymbol{\lambda}^n\}_{n\in\mathbb{N}} \subseteq \mathbf{R}^{N_{\mathbf{g}}}_+ \times \mathbf{R}^{N_S}_+$, such that $\lim_{n\to\infty}\|\boldsymbol{\lambda}^n\|_2 \equiv \infty$. Since $\mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger},\cdot)$ is a non-trivial affine function *and with positive slope*, we may write

$$\begin{aligned}\liminf_{n\to\infty} \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^n) &\ge \liminf_{n\to\infty} \mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger},\boldsymbol{\lambda}^n) \\ &\equiv \lim_{n\to\infty} \mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger},\boldsymbol{\lambda}^n) \equiv \infty, \end{aligned} \tag{52}$$

which in turn yields

$$\liminf_{n\to\infty} \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^n) \equiv \infty \implies \lim_{n\to\infty} \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^n) \equiv \infty. \tag{53}$$

Noting that $\mathscr{D}_{\boldsymbol{\phi}}$ is proper due to $\mathscr{D}^*_{\boldsymbol{\phi}}$ being finite by assumption and Lemma 9, it is straightforward to show that the *clipped dual function* $\widetilde{\mathscr{D}}_{\boldsymbol{\phi}} : \mathbf{R}^{N_{\mathbf{g}}} \times \mathbf{R}^{N_S} \to (-\infty,\infty]$ defined as

$$\widetilde{\mathscr{D}}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}) \triangleq \begin{cases} \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}), & \text{if } \boldsymbol{\lambda}\ge\mathbf{0} \\ \infty, & \text{otherwise}\end{cases} \tag{54}$$

is proper and *coercive* (p. 119 in [53]). Since, also by Lemma 9, $\mathscr{D}_{\boldsymbol{\phi}}$ is closed, it follows that $\widetilde{\mathscr{D}}_{\boldsymbol{\phi}}$ is closed as well (as a sum of proper closed functions: $\mathscr{D}_{\boldsymbol{\phi}}$ itself and the indicator of $\mathbf{R}^{N_{\mathbf{g}}}_+ \times \mathbf{R}^{N_S}_+$; see ([53], Proposition 1.1.5)). Thus, we may call ([53], Proposition 3.2.1), which ensures that $\arg\min_{\boldsymbol{\lambda}\ge\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda})$ is a nonempty and compact set in $\mathbf{R}^{N_{\mathbf{g}}}_+ \times \mathbf{R}^{N_S}_+$.

Next, whenever Assumptions 1 and 2 are in effect, Theorem 6 ensures the existence of strictly positive numbers $\mu^{\dagger}_S > 0$ and $\mu^{\dagger}_R > 0$, possibly dependent on $(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger})$, such that, for every $0 < \mu_S \le \mu^{\dagger}_S$ and $0 < \mu_R \le \mu^{\dagger}_R$, the particular point $(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger})$ is also strictly feasible for problem (12). Then, with the help of the respective part of Lemma 9, the procedure presented above for (2) may be repeated for (12), for each pair of qualifying μ_S and μ_R. ■

Lemma 11. (Uniform Boundedness of Dual Solution Sets) *Let Assumptions 1, 2 and 3 be in effect, and suppose that $\mathscr{D}^*_{\boldsymbol{\phi}} < \infty$. Then, for the same $\mu^{\dagger}_S > 0$ and $\mu^{\dagger}_R > 0$ of Lemma 10, there is a compact set $\boldsymbol{\Lambda}^{\dagger} \subseteq \mathbf{R}^{N_{\boldsymbol{g}}}_+ \times \mathbf{R}^{N_S}_+$, such that, for every $0 \le \mu_S \le \mu^{\dagger}_S$ and $0 \le \mu_R \le \mu^{\dagger}_R$,*

$$\arg\min_{\boldsymbol{\lambda}\ge\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) \subseteq \boldsymbol{\Lambda}^{\dagger},$$

provided that $\mathtt{S}$ is continuous on $[0,\mu^{\dagger}_R]$. It is then true that

$$\sup_{\mu_S\in[0,\mu^{\dagger}_S],\mu_R\in[0,\mu^{\dagger}_R]} \begin{bmatrix}\boldsymbol{\lambda}^*_S(\mu_S,\mu_R) \\ \boldsymbol{\lambda}^*_R(\mu_S,\mu_R)\end{bmatrix} < \boldsymbol{\infty},$$

where the supremum is taken element-wise.

Proof of Lemma 11: Let us first show that $\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda})$ is jointly (pointwise) continuous in $(\boldsymbol{\lambda},\mu) \in \mathbf{R}^{N_{\boldsymbol{g}}}_+ \times \mathbf{R}^{N_S}_+ \times \mathbf{R}^2_+$. Similar to the conclusions of Lemmata 3 and 4, it is easy to verify that, for every $\mu^1 \ge 0$ and $\mu^2 \ge 0$,

$$\sup_{\mathbf{x}\in\mathcal{X}} |g^i_{\mu^1}(\mathbf{x}) - g^i_{\mu^2}(\mathbf{x})| \le |\mu^1-\mu^2| L^i_g\sqrt{N_S}, i\in\{o,\mathbb{N}^+_{N_{\mathbf{g}}}\} \tag{55}$$

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\overline{f}^{\boldsymbol{\phi},i}_{\mu^1}(\boldsymbol{\theta}) - \overline{f}^{\boldsymbol{\phi},i}_{\mu^2}(\boldsymbol{\theta})\right| \le |\mu^1-\mu^2| L^i_f\sqrt{N_{\boldsymbol{\phi}}}, i\in\mathbb{N}^+_{N_S}. \tag{56}$$

Fix an arbitrary pair $\boldsymbol{\lambda}^o = (\boldsymbol{\lambda}^o_S,\boldsymbol{\lambda}^o_R) \ge \mathbf{0}$ and $\mu^o = (\mu^o_S,\mu^o_R) \ge \mathbf{0}$. For every other choice $(\boldsymbol{\lambda},\mu) \ge \mathbf{0}$, we have the estimates

$$\begin{aligned}&|\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) - \mathscr{D}_{\boldsymbol{\phi},\mu^o}(\boldsymbol{\lambda}_o)| \\ &= |\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) - \mathscr{D}_{\boldsymbol{\phi},\mu^o}(\boldsymbol{\lambda})| + |\mathscr{D}_{\boldsymbol{\phi},\mu^o}(\boldsymbol{\lambda}) - \mathscr{D}_{\boldsymbol{\phi},\mu^o}(\boldsymbol{\lambda}_o)|. \end{aligned} \tag{57}$$

Additionally, we may write

$$\begin{aligned}&|\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) - \mathscr{D}_{\boldsymbol{\phi},\mu^o}(\boldsymbol{\lambda})| \\ &= \left|\sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) - \sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu^o}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})\right| \\ &\le \sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} |\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) - \mathscr{L}_{\boldsymbol{\phi},\mu^o}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})|. \end{aligned} \tag{58}$$

Now, we have

$$\begin{aligned}&|\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) - \mathscr{L}_{\boldsymbol{\phi},\mu^o}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})| \\ &\le |g^o_{\mu_S}(\mathbf{x}) - g^o_{\mu^o_S}(\mathbf{x})| + |\langle\boldsymbol{\lambda}_S, \mathbf{g}_{\mu_S}(\mathbf{x}) - \mathbf{g}_{\mu^o_S}(\mathbf{x})\rangle| \\ &\quad + |\langle\boldsymbol{\lambda}_R, \overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}) - \overline{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu^o_R}(\boldsymbol{\theta})\rangle| + |\langle\boldsymbol{\lambda}_R, \mathtt{S}(\mu_R) - \mathtt{S}(\mu^o_R)\rangle| \\ &\le |\mu_S-\mu^o_S| L^o_g + |\mu_S-\mu^o_S|\langle\boldsymbol{\lambda}_S, \boldsymbol{c}_S\rangle\sqrt{N_S} \\ &\quad + |\mu_R-\mu^o_R|\langle\boldsymbol{\lambda}_R, \boldsymbol{c}_R\rangle\sqrt{N_{\boldsymbol{\phi}}} + \|\boldsymbol{\lambda}_R\|_2\|\mathtt{S}(\mu_R) - \mathtt{S}(\mu^o_R)\|_2. \end{aligned} \tag{59}$$

Together with the continuity of the slack $\mathtt{S}$ and that of the dual function $\mathscr{D}_{\boldsymbol{\phi},\mu}(\cdot)$ for each μ –due to convexity, also note that μ_o is fixed–, we obtain that

$$\lim_{(\boldsymbol{\lambda},\mu)\to(\boldsymbol{\lambda}^o,\mu^o)} |\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) - \mathscr{D}_{\boldsymbol{\phi},\mu_o}(\boldsymbol{\lambda}_o)| = 0, \tag{60}$$

which verifies continuity of $\mathscr{D}_{\boldsymbol{\phi},\cdot}(\bullet)$ on $\mathbf{R}^{N_{\boldsymbol{g}}}_+ \times \mathbf{R}^{N_S}_+ \times \mathbf{R}^2_+$.

Utilizing Lemma 10, we now consider the nonempty and compact-valued solution set multifunction $\boldsymbol{\Lambda}^* : \mathbf{R}^2_+ \rightrightarrows \mathbf{R}^{N_{\boldsymbol{g}}}_+ \times \mathbf{R}^{N_S}_+$ defined for convenience as

$$\boldsymbol{\Lambda}^*(\mu) \triangleq \arg\min_{\boldsymbol{\lambda}\ge\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}), \quad \mu\ge\mathbf{0}, \tag{61}$$

and let $\mathrm{M} \triangleq [0, \mu_S^{\dagger}] \times [0, \mu_R^{\dagger}]$. Let us establish some facts. *First*, it is easy to see that $\mathscr{D}_{\boldsymbol{\phi},\cdot}(\bullet)$ is uniformly bounded from below on $\mathbf{R}_+^{N_{\boldsymbol{g}}} \times \mathbf{R}_+^{N_S} \times \mathrm{M}$. Indeed, because of strict feasibility for every $\mu \in \mathrm{M}$, we may write

$$\begin{aligned}\mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) &= \sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \\ &\geq \mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x}^{\dagger},\boldsymbol{\theta}^{\dagger},\boldsymbol{\lambda}) \\ &\geq g_{\mu_S}^{o}(\mathbf{x}^{\dagger}) \\ &\geq g^{o}(\mathbf{x}^{\dagger}) - \mu_S L_g^o\sqrt{N_S} \\ &\geq g^{o}(\mathbf{x}^{\dagger}) - \mu_S^{\dagger} L_g^o\sqrt{N_S},\end{aligned} \tag{62}$$

which of course implies that $\inf_{\boldsymbol{\lambda}\geq\mathbf{0},\mu\in\mathrm{M}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda})$ is a number. *Second*, we can find a sequence of $\{\mathbb{K}_n\}_n$ of compact subsets *increasing* to $\mathbf{R}_+^{N_{\boldsymbol{g}}} \times \mathbf{R}_+^{N_S} \times \mathrm{M}$, for which

$$\lim_{n\to\infty} \inf_{(\boldsymbol{\lambda},\mu)\in\left[\mathbb{R}_+^{N_{\boldsymbol{g}}}\times\mathbb{R}_+^{N_S}\times\mathrm{M}\right]\setminus\mathbb{K}_n} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) = +\infty. \tag{63}$$

For instance, we can take $\mathbb{K}_n = [0,n]^{N_{\boldsymbol{g}}+N_S} \times \mathrm{M}$, $n \in \mathbb{N}$. Then $\left[\mathbf{R}_+^{N_{\boldsymbol{g}}} \times \mathbf{R}_+^{N_S} \times \mathrm{M}\right]\setminus\mathbb{K}_n$ is empty, and minimizing over the empty set returns $+\infty$. It is then said that $\mathscr{D}_{\boldsymbol{\phi},\cdot}(\bullet)$ satisfies the *moment condition on* $\mathbf{R}_+^{N_{\boldsymbol{g}}} \times \mathbf{R}_+^{N_S} \times \mathrm{M}$ ([54], Section 3). *Third*, it follows that $\mathscr{D}_{\boldsymbol{\phi},\cdot}(\bullet)$ is *inf-compact on* $\mathbf{R}_+^{N_{\boldsymbol{g}}} \times \mathbf{R}_+^{N_S} \times \mathrm{M}$, i.e., the restricted sub-level set multifunction

$$\mathcal{S}(r,\mu) \triangleq \{\boldsymbol{\lambda} \geq \mathbf{0} | \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) \leq r\}, \quad (\mu, r) \in \mathrm{M} \times r \tag{64}$$

is compact-valued. To see this, observe that we can write, for every $\mu \in \mathrm{M}$ and $r \in \mathbb{R}$,

$$\mathcal{S}(r,\mu) = \{\boldsymbol{\lambda} \in \mathbb{R}^N | \widetilde{\mathscr{D}}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) \triangleq \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}) + \mathbb{I}_{\mathbb{R}_+}(\boldsymbol{\lambda}) \leq r\}, \tag{65}$$

and because $\widetilde{\mathscr{D}}_{\boldsymbol{\phi},\mu}(\cdot)$ is coercive for each $\mu \in \mathrm{M}$ (see also proof of Lemma 10), the claim follows. Therefore, from ([54], Theorem 3.1), we obtain that the parameterized optimal dual value

$$\mathscr{D}_{\boldsymbol{\phi}}^*(\mu) = \inf_{\boldsymbol{\lambda}\geq\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}), \quad \mu \in \mathrm{M} \tag{66}$$

is continuous, and that the compact-valued multifunction $\boldsymbol{\Lambda}^*$ is *upper hemicontinuous*. Consequently, by the latter conclusion, the image of the compactum M under $\boldsymbol{\Lambda}^*$ (i.e., the range of $\boldsymbol{\Lambda}^*$ on M), defined as ([55], Section 17.1)

$$\boldsymbol{\Lambda}^{\dagger} \triangleq \bigcup_{\mu\in\mathrm{M}} \boldsymbol{\Lambda}^*(\mu), \tag{67}$$

is also a compact subset of $\mathbf{R}_+^{N_{\boldsymbol{g}}} \times \mathbf{R}_+^{N_S}$ ([55], Lemma 17.8). This further implies that

$$\begin{aligned}\sup_{\mu\in\mathrm{M}} \boldsymbol{\lambda}^*(\mu) &= \sup_{(\mu_S,\mu_R)\in\mathrm{M}} \boldsymbol{\lambda}^*(\mu_S,\mu_R) \\ &= \sup_{(\mu_S,\mu_R)\in\mathrm{M}} \begin{bmatrix}\boldsymbol{\lambda}_S^*(\mu_S,\mu_R) \\ \boldsymbol{\lambda}_R^*(\mu_S,\mu_R)\end{bmatrix} < \boldsymbol{\infty},\end{aligned} \tag{68}$$

where $\boldsymbol{\lambda}^*(\cdot) \in \boldsymbol{\Lambda}^*(\cdot) \subseteq \boldsymbol{\Lambda}^{\dagger}$ is any selection of $\boldsymbol{\Lambda}^*$ on M, and where the supremum is taken in an element-wise manner. ■

Out of Lemmata 8, 9, 10 and 11, the first and the penultimate ones are the most important. In particular, Lemma 8 provides explicit non-symmetric upper and lower bounds for the difference between the Lagrangians of (2) and (12), *which are independent of* $\mathbf{x}$ *and* $\boldsymbol{\theta}$. Therefore, it should be possible to obtain approximation bounds on the respective dual functions, *uniform relative to* $(\mathbf{x},\boldsymbol{\theta})$. On top of this, Lemma 10 verifies the existence of dual optimal solutions for (2) and (12), which could be exploited in conjunction with the aforementioned uniform bounds on the respective dual functions, to provide fundamental upper and lower bounds on the corresponding dual optimal (infimal) values. All this is confirmed and rigorously quantified by the next central result.

Theorem 12. (Dual Value Approximation) *Let Assumptions 1, 2 and 3 be in effect, and suppose that* $\mathscr{D}_{\boldsymbol{\phi}}^* < \infty$. *Then there exist* $\mu_S^{\dagger} > 0$ *and* $\mu_R^{\dagger} > 0$, *such that, for every* $0 \leq \mu_S \leq \mu_S^{\dagger}$ *and* $0 \leq \mu_R \leq \mu_R^{\dagger}$,

$$\boxed{-\Gamma_\mu^l(\boldsymbol{\lambda}_{\boldsymbol{\phi}}^{\dagger}) \leq -\Gamma_\mu^l(\boldsymbol{\lambda}_{\boldsymbol{\phi},\mu}^*) \leq \mathscr{D}_{\boldsymbol{\phi},\mu}^* - \mathscr{D}_{\boldsymbol{\phi}}^* \leq \Gamma_\mu^r(\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*).} \tag{69}$$

where $\boldsymbol{\lambda}_{\boldsymbol{\phi},\mu}^* \in \arg\min_{\boldsymbol{\lambda}\geq\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda})$, $\mathscr{D}_{\boldsymbol{\phi},0}^* \equiv \mathscr{D}_{\boldsymbol{\phi}}^*$ *and* $\boldsymbol{\lambda}_{\boldsymbol{\phi},0}^* \equiv \boldsymbol{\lambda}_{\boldsymbol{\phi}}^*$, *identically, and where* $\boldsymbol{\lambda}_{\boldsymbol{\phi}}^{\dagger} \geq \mathbf{0}$ *is a constant independent of* μ_S *and* μ_R, *which is finite if* $\mathtt{S}$ *is continuous on* $[0, \mu_R^{\dagger}]$. *Further, if* $\mathtt{S}(\mu_R) \equiv \boldsymbol{C}\mu_R\sqrt{N_{\boldsymbol{\phi}}}, \boldsymbol{C} \geq \mathbf{0}$, *then there always exist finite constants* $\Sigma_S^l \geq 0$, $\Sigma_R^l \geq 0$ *and* $\Sigma_R^r \in \mathbf{R}$, *problem dependent but independent of* μ_S *and* μ_R, *such that*

$$\boxed{-(\mu_S\Sigma_S^l + \mu_R\Sigma_R^l) \leq \mathscr{D}_{\boldsymbol{\phi},\mu}^* - \mathscr{D}_{\boldsymbol{\phi}}^* \leq \mu_R\Sigma_R^r.} \tag{70}$$

In particular, Σ_S^l, Σ_R^l *and* Σ^r *are defined as*

$$\Sigma_S^l \triangleq \left(L_g^o + \langle\boldsymbol{\lambda}_{\boldsymbol{\phi},S}^{\dagger}, \mathbf{c}_S\rangle\right)\sqrt{N_S}, \tag{71}$$

$$\Sigma_R^l \triangleq \langle\boldsymbol{\lambda}_{\boldsymbol{\phi},R}^{\dagger}, \mathbf{c}_R + \boldsymbol{C}\rangle\sqrt{N_{\boldsymbol{\phi}}} \quad \textit{and} \tag{72}$$

$$\Sigma_R^r \triangleq \langle\boldsymbol{\lambda}_{\boldsymbol{\phi},R}^*, \mathbf{c}_R - \boldsymbol{C}\rangle\sqrt{N_{\boldsymbol{\phi}}}. \tag{73}$$

Lastly, whenever $\mathtt{S}(\mu_R) \equiv \boldsymbol{C}\mu_R\sqrt{N_{\boldsymbol{\phi}}}$ *with* $\boldsymbol{C} \geq \mathbf{c}_R$, *then the right-hand-sides of (69) and (70) are nonpositive, and may be replaced by zero.*

Proof of Theorem 12: Under the assumptions of the theorem, Lemma 10 ensures that the dual optimal solution set $\arg\min_{\boldsymbol{\lambda}\geq\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda})$ is nonempty, for all $0 \leq \mu_S \leq \mu_S^{\dagger}$ and $0 \leq \mu_R \leq \mu_R^{\dagger}$. Consequently, there exist optimal multipliers $\boldsymbol{\lambda}_{\boldsymbol{\phi},\mu}^* \geq \mathbf{0}$, such that

$$-\infty < \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}_{\boldsymbol{\phi},\mu}^*) \equiv \mathscr{D}_{\boldsymbol{\phi},\mu}^* \equiv \inf_{\boldsymbol{\lambda}\geq\mathbf{0}} \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}), \tag{74}$$

for all allowable values of μ_S and μ_R. Therefore, invoking Lemma 8, we may carefully write

$$\begin{aligned}\mathscr{D}_{\boldsymbol{\phi},\mu}^* &\leq \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*) \\ &\equiv \sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*) \\ &\leq \left[\sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*)\right] + \Gamma_\mu^r(\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*) \\ &\equiv \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*) + \Gamma_\mu^r(\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*) \\ &\equiv \mathscr{D}_{\boldsymbol{\phi}}^* + \Gamma_\mu^r(\boldsymbol{\lambda}_{\boldsymbol{\phi}}^*).\end{aligned} \tag{75}$$

By symmetry, we also have

$$\begin{aligned}\mathscr{D}_{\boldsymbol{\phi}}^* &\leq \mathscr{D}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}_{\boldsymbol{\phi},\mu}^*) \\ &\equiv \sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi}}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\phi},\mu}^*)\end{aligned}$$

$$\begin{aligned}&\leq \left[\sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu})\right] + \Gamma^l_\mu(\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu}) \\ &\equiv \mathscr{D}_{\boldsymbol{\phi},\mu}(\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu}) + \Gamma^l_\mu(\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu}) \\ &\equiv \mathscr{D}^*_{\boldsymbol{\phi},\mu} + \Gamma^l_\mu(\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu}). \end{aligned} \tag{76}$$

Rearranging (75) and (76), we obtain the last three inequalities of (69) (left-to-right), as in the statement of the theorem.

Now, recalling that $\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu} \equiv (\boldsymbol{\lambda}^*_{\boldsymbol{\phi},S}(\mu_S,\mu_R), \boldsymbol{\lambda}^*_{\boldsymbol{\phi},R}(\mu_S,\mu_R))$, we may invoke Lemma 11 to define *maximal multipliers* (here the supremum is taken *elementwise* on the involved vectors)

$$\boldsymbol{\lambda}^\dagger_{\boldsymbol{\phi},S} \triangleq \sup_{\mu_S\in[0,\mu_S^\dagger],\mu_R\in[0,\mu_R^\dagger]} \boldsymbol{\lambda}^*_{\boldsymbol{\phi},S}(\mu_S,\mu_R), \quad \text{and} \tag{77}$$

$$\boldsymbol{\lambda}^\dagger_{\boldsymbol{\phi},R} \triangleq \sup_{\mu_S\in[0,\mu_S^\dagger],\mu_R\in[0,\mu_R^\dagger]} \boldsymbol{\lambda}^*_{\boldsymbol{\phi},R}(\mu_S,\mu_R). \tag{78}$$

Using these definitions, it follows that, for every $0 \leq \mu_S \leq \mu_S^\dagger$ and $0 \leq \mu_R \leq \mu_R^\dagger$,

$$\Gamma^l_\mu(\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu}) \leq \Gamma^l_\mu(\boldsymbol{\lambda}^\dagger_{\boldsymbol{\phi}}), \tag{79}$$

verifying the left inequality of (69).

When $\mathtt{S}(\mu_R) \equiv \boldsymbol{C}\mu_R\sqrt{N_{\boldsymbol{\phi}}}, \boldsymbol{C} \geq \mathbf{0}$, the rest of the claims stated in the theorem follow by noting that, for every $\boldsymbol{\lambda} \geq \mathbf{0}$,

$$\begin{aligned}\Gamma^l_\mu(\boldsymbol{\lambda}) &\triangleq \mu_S\big(L^o_g + \langle \boldsymbol{\lambda}_S, \boldsymbol{c}_S\rangle\big)\sqrt{N_S} \\ &\quad + \mu_R\big(\langle \boldsymbol{\lambda}_R, \boldsymbol{c}_R\rangle + \langle \boldsymbol{C}, \boldsymbol{\lambda}_R\rangle\big)\sqrt{N_{\boldsymbol{\phi}}} \quad \text{and} \\ \Gamma^r_\mu(\boldsymbol{\lambda}) &\triangleq \mu_R\big(\langle \boldsymbol{\lambda}_R, \boldsymbol{c}_R\rangle - \langle \boldsymbol{C}, \boldsymbol{\lambda}_R\rangle\big)\sqrt{N_{\boldsymbol{\phi}}},\end{aligned} \tag{80}$$

where both functions $\Gamma^l_{(\cdot)}$ and $\Gamma^r_{(\cdot)}$ are nondecreasing in both μ_S and μ_R. If, additionally, $\boldsymbol{C} \geq \boldsymbol{c}_R$, then $\Gamma^r_{(\cdot)} \leq 0$. The proof is complete. ■

B. Approximate Strong Duality

After explicitly relating the dual optimal values $\mathscr{D}^*_{\boldsymbol{\phi}}$ and $\mathscr{D}^*_{\boldsymbol{\phi},\mu}$, our second task will be to relate $\mathscr{D}^*_{\boldsymbol{\phi},\mu}$ to the optimal value of the base problem (1). In particular, we would like to characterize the *duality gap between the primal problem* (1) *and the dual to problem* (12). Note that we are *not* interested in characterizing the duality gap of (12); to the best of our knowledge, this constitutes a nontrivial problem.

Following [23], we exploit the notion of a ϵ-universal policy parameterization. This allows us to characterize the *intermediate* duality gap between the optimal value of (1), and the optimal value of the dualization of (12) [23].

Definition 13. (ϵ-Universality) Fix $\epsilon \geq 0$, choose a parameterization $\boldsymbol{\phi}: \mathcal{H}\times\mathbf{R}^{N_{\boldsymbol{\phi}}} \to \mathbf{R}^{N_R}$, and let $\Theta \subseteq \mathbf{R}^{N_{\boldsymbol{\phi}}}$ be any parameter subspace. A class of admissible policies

$$\mathcal{P}^{\boldsymbol{\phi}}_\Theta \triangleq \{\mathbf{p}\in\mathcal{P} \,|\, \mathbf{p}(\cdot) \equiv \boldsymbol{\phi}(\cdot,\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta\} \subseteq \mathcal{P} \tag{81}$$

is called ϵ*-universal in* $\mathcal{P}$ if and only if, for every $\mathbf{p}\in\mathcal{P}$, there exists $\boldsymbol{\phi}(\cdot,\boldsymbol{\theta}\equiv\boldsymbol{\theta}(\epsilon,\mathbf{p})) \in \mathcal{P}^{\boldsymbol{\phi}}_\Theta$, such that

$$\mathbb{E}\{\|\mathbf{p}(\mathbf{H}) - \boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta})\|_\infty\} < \epsilon. \tag{82}$$

Remark 14. Note that, when $\Theta \subseteq \big\{\boldsymbol{\theta}\in\mathbf{R}^{N_{\boldsymbol{\phi}}} \big| \boldsymbol{\phi}(\cdot,\boldsymbol{\theta})\in\mathcal{P}\big\}$, as assumed in Section I, it follows that $\mathcal{P}^{\boldsymbol{\phi}}_\Theta \equiv \{\boldsymbol{\phi}(\cdot,\boldsymbol{\theta}) | \boldsymbol{\theta}\in\Theta\}$. Then, trivially, $\boldsymbol{\phi}(\cdot,\boldsymbol{\theta})\in\mathcal{P}^{\boldsymbol{\phi}}_\Theta$, for all $\boldsymbol{\theta}\in\Theta$, and ϵ-universality of $\mathcal{P}^{\boldsymbol{\phi}}_\Theta$ (or $\boldsymbol{\phi}$, simply) in $\mathcal{P}$ is ensured as long as, for every admissible policy $\mathbf{p}\in\mathcal{P}$, there is at least one $\boldsymbol{\theta}\in\Theta$ satisfying (82). ■

Additionally, also as in [23], we will impose the following additional structural assumptions.

Assumption 4. *The Borel pushforward $\mathcal{M}_{\mathbf{H}}: \mathscr{B}(\mathcal{H}) \to [0,1]$ is nonatomic: For every Borel set $\mathcal{E}\in\mathscr{B}(\mathcal{H})$ such that $\mathcal{M}_{\mathbf{H}}(\mathcal{E}) > 0$, there exists another Borel set $\mathcal{E}^o \in \mathscr{B}(\mathcal{H})$ satisfying $\mathcal{M}_{\mathbf{H}}(\mathcal{E}) > \mathcal{M}_{\mathbf{H}}(\mathcal{E}^o) > 0$.*

Assumption 5. *For every pair $(\mathbf{x},\mathbf{x}')\in\mathcal{X}\times\mathcal{X}$ such that $\mathbf{x}\leq\mathbf{x}'$, it is true that $g^o(\mathbf{x}) \leq g^o(\mathbf{x}')$ and $\mathbf{g}(\mathbf{x}) \leq \mathbf{g}(\mathbf{x}')$.*

Assumption 6. *There exists a number $L^{\mathbf{f}}_{\mathcal{P}} < \infty$, such that, for every pair $(\mathbf{p},\mathbf{p}')\in\mathcal{P}\times\mathcal{P}$, it is true that*

$$\begin{aligned}&\|\mathbb{E}\{\mathbf{f}(\mathbf{p}(\mathbf{H}),\mathbf{H})\} - \mathbb{E}\{\mathbf{f}(\mathbf{p}'(\mathbf{H}),\mathbf{H})\}\|_\infty \\ &\qquad \leq L^{\mathbf{f}}_{\mathcal{P}}\mathbb{E}\{\|\mathbf{p}(\mathbf{H}) - \mathbf{p}'(\mathbf{H})\|_\infty\}.\end{aligned} \tag{83}$$

As clearly explained in ([23], Section III.A), Assumptions 4, 5 and 6 are reasonable and are fulfilled by most practically significant wireless resource allocation problems. We thus do not further comment.

Under Assumptions 3, 4, 5 and 6, an important result was presented in [23], which characterizes the duality gap between the base problem (1) and the parameterized surrogate (2), leveraging the notion of ϵ-universality of Definition 13. For completeness, we also report this result here, as follows.

Theorem 15. (PFA Duality Gap [23]) *Let Assumptions 3, 4, 5 and 6 be in effect, and suppose that, for some $\epsilon\geq 0$, $\boldsymbol{\phi}$ is ϵ-universal in $\mathcal{P}$. Then it is true that $\mathscr{D}^*_{\boldsymbol{\phi}} < \infty$ and, further,*

$$-\|\boldsymbol{\lambda}^*\|L^{\mathbf{f}}_{\mathcal{P}}\epsilon \leq \mathscr{D}^*_{\boldsymbol{\phi}} - \mathscr{P}^* \leq 0, \tag{84}$$

where $\boldsymbol{\lambda}^* \in \arg\min_{\boldsymbol{\lambda}\geq\mathbf{0}} \mathscr{D}(\boldsymbol{\lambda}) \neq \emptyset$, and $\mathscr{D}: \mathbf{R}^{N_{\mathbf{g}}}\times\mathbf{R}^{N_S} \to (-\infty,\infty]$ denotes the dual function of the base problem (1).

We now combine Theorem 15 with Theorem 12 developed in Section IV-A, resulting in the main result of this paper. The proof is elementary, and thus omitted.

Theorem 16. (Smoothed PFA Duality Gap) *Let Assumptions 1, 2, 3, 4, 5 and 6 be in effect, and suppose that, for some $\epsilon\geq 0$, $\boldsymbol{\phi}$ is ϵ-universal in $\mathcal{P}$. Then, by the definitions of Lemma 8 and Theorem 12, there exist $\mu_S^\dagger > 0$ and $\mu_R^\dagger > 0$, such that, for every $0\leq\mu_S\leq\mu_S^\dagger$ and $0\leq\mu_R\leq\mu_R^\dagger$,*

$$\boxed{-\big(\Gamma^l_\mu(\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu}) + \|\boldsymbol{\lambda}^*\|L^{\mathbf{f}}_{\mathcal{P}}\epsilon\big) \leq \mathscr{D}^*_{\boldsymbol{\phi},\mu} - \mathscr{P}^* \leq \Gamma^r_\mu(\boldsymbol{\lambda}^*_{\boldsymbol{\phi}}),} \tag{85}$$

with $\Gamma^l_\mu(\boldsymbol{\lambda}^*_{\boldsymbol{\phi},\mu}) \leq \Gamma^l_\mu(\boldsymbol{\lambda}^\dagger_{\boldsymbol{\phi}})$. Further, if $\mathtt{S}(\mu_R)\equiv\boldsymbol{C}\mu_R\sqrt{N_{\boldsymbol{\phi}}}$, $\boldsymbol{C}\geq\mathbf{0}$, it is true that

$$\boxed{-\big(\mu_S\Sigma^l_S + \mu_R\Sigma^l_R + \|\boldsymbol{\lambda}^*\|L^{\mathbf{f}}_{\mathcal{P}}\epsilon\big) \leq \mathscr{D}^*_{\boldsymbol{\phi},\mu} - \mathscr{P}^* \leq \mu_R\Sigma^r_R.} \tag{86}$$

Lastly, whenever $\mathtt{S}(\mu_R)\equiv\boldsymbol{C}\mu_R\sqrt{N_{\boldsymbol{\phi}}}$ with $\boldsymbol{C}\geq\boldsymbol{c}_R$, then the right-hand-sides of (85) *and* (86) *are nonpositive, and may be replaced by zero.*

We may also state a trivial corollary to Theorem 16, masking all its technicalities, which sometimes might be unnecessary in more qualitative arguments.

Algorithm 1 Model-Free Randomized Primal-Dual Learning

Input: $\mathbf{x}^0$, $\boldsymbol{\theta}^0$, $\boldsymbol{\lambda}_S^0$, $\boldsymbol{\lambda}_R^0$, $\{\gamma_{\mathbf{x}}^n, \gamma_{\boldsymbol{\theta}}^n, \gamma_{\boldsymbol{\lambda}_S}^n, \gamma_{\boldsymbol{\lambda}_R}^n\}_{n\in\mathbb{N}}$, μ_S, μ_R

Output: $\{\mathbf{x}^{n+1}, \boldsymbol{\theta}^{n+1}\}_{n\in\mathbb{N}}$

for $n = 0, 1, 2, \ldots$ **do**

Draw samples $\boldsymbol{U}_S^{n+1}$ and $\boldsymbol{U}_R^{n+1}$.

Sample values

$$g^o(\mathbf{x}^n),\, g^o(\mathbf{x}^n + \mu_S \boldsymbol{U}_S^{n+1}) \quad \text{and} \quad \mathbf{g}(\mathbf{x}^n),\, \mathbf{g}(\mathbf{x}^n + \mu_S \boldsymbol{U}_S^{n+1}),$$

and probe the wireless system to obtain

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

Compute $\mathbf{x}^{n+1}$ and $\boldsymbol{\theta}^{n+1}$ from (98) and (99).

Sample $\mathbf{g}(\mathbf{x}^{n+1} + \mu_S \boldsymbol{U}_S^{n+1})$ and probe the wireless system to obtain $\mathbf{f}(\boldsymbol{\phi}(\mathbf{H}^{n+1}, \boldsymbol{\theta}^{n+1} + \mu_R \boldsymbol{U}_R^{n+1})$.

Compute $\boldsymbol{\lambda}_S^{n+1}$ and $\boldsymbol{\lambda}_R^{n+1}$ from (100) and (101).

end for

Corollary 17. (Smoothed PFA Duality Gap | Simplified) *Let Assumptions 1, 2, 3, 4, 5 and 6 be in effect and suppose that, for some $\epsilon \geq 0$, $\boldsymbol{\phi}$ is ϵ-universal in $\mathcal{P}$. Further, choose $\mathtt{S}(\mu_R) \equiv \boldsymbol{C}\mu_R\sqrt{N_{\boldsymbol{\phi}}}, \boldsymbol{C} \geq \mathbf{0}$. Then, it is true that*

$$\boxed{\left|\mathscr{D}^*_{\boldsymbol{\phi},\mu} - \mathscr{P}^*\right| \equiv \mathcal{O}(\mu_S + \mu_R + \epsilon),} \tag{87}$$

as $(\mu_S, \mu_R, \epsilon) \downarrow \mathbf{0}$. If, further, $\boldsymbol{C} \geq \boldsymbol{c}_R$, (87) may be improved as

$$\boxed{0 \leq \mathscr{P}^* - \mathscr{D}^*_{\boldsymbol{\phi},\mu} \equiv \mathcal{O}(\mu_S + \mu_R + \epsilon),} \tag{88}$$

as $(\mu_S, \mu_R, \epsilon) \downarrow \mathbf{0}$.

Theorem 16 and Corollary 17 explicitly quantify the gap between dual optimal value of the smoothed surrogate (12) and the (primal) optimal value of the constrained variational problem (1). What is more, the gap can be made *arbitrarily small at will*, and scales *linearly* relative to the near-universality precision ϵ, *and* the smoothing parameters μ_S and μ_R.

The importance of Theorem 16 and Corollary 17 is *twofold*. First, similarly to [23] and together with Theorems 6 and 7, Theorem 16 and Corollary 17 provide solid technical evidence justifying the dualization of (12) as a proxy for obtaining the optimal value of (1), *and* a corresponding optimal solution. This very useful per se, since the dual problem embeds the constraints of (12) in its objective, via the Lagrangian formulation.

Second, *and most importantly*, solving for (12) in the dual domain can be performed in a *gradient-free fashion*, using *only* evaluations of the functions present in *both* the objective and constraints of (12), as we discuss next. This makes optimal wireless resource allocation in the model-free setting possible, within a *non-heuristic* and *predictable* framework.

V. 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 (12) in the model-free setting. The algorithm is *non-heuristic* and *derived* from first principles, and uses stochastic approximation to tackle the minimax problem

$$\begin{array}{ll} \text{minimize} & \sup_{(\mathbf{x},\boldsymbol{\theta})\in\mathcal{X}\times\Theta} \mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \\ \text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}, \tag{89}$$

for every qualifying choice of μ_S and μ_R.

Apparently, a basic primal-dual method for solving (89) may be easily derived by taking gradients relative to all of its variables, and then performing alternating gradient steps in appropriate directions. Specifically, for every $(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{X} \times \Theta$ and for every $\boldsymbol{\lambda} \equiv (\boldsymbol{\lambda}_S, \boldsymbol{\lambda}_R) \geq \mathbf{0}$, the gradients of $\mathscr{L}_{\boldsymbol{\phi},\mu}$ with respect to each of its arguments may be readily expressed as

$$\nabla_{\mathbf{x}}\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \equiv \nabla g^o_{\mu_S}(\mathbf{x}) + \nabla \mathbf{g}_{\mu_S}(\mathbf{x})\boldsymbol{\lambda}_S - \boldsymbol{\lambda}_R, \tag{90}$$

$$\nabla_{\boldsymbol{\theta}}\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \equiv \nabla \bar{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta})\boldsymbol{\lambda}_R, \tag{91}$$

$$\nabla_{\boldsymbol{\lambda}_S}\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \equiv \mathbf{g}_{\mu_S}(\mathbf{x}) \quad \text{and} \tag{92}$$

$$\nabla_{\boldsymbol{\lambda}_R}\mathscr{L}_{\boldsymbol{\phi},\mu}(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \equiv \bar{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}) - \mathbf{x} - \mathtt{S}(\mu_R). \tag{93}$$

Then, the idea is to iteratively ascend in $(\mathbf{x}, \boldsymbol{\theta})$ and descend in $\boldsymbol{\lambda}$, in an alternating fashion. If $n \in \mathbb{N}$ denotes an iteration index, this implies the updates

$$\mathbf{x}^{n+1} \equiv \Pi_{\mathcal{X}}\{\mathbf{x}^n + \boldsymbol{\gamma}^n_{\mathbf{x}} \circ (\nabla g^o_{\mu_S}(\mathbf{x}^n) + \nabla \mathbf{g}_{\mu_S}(\mathbf{x}^n)\boldsymbol{\lambda}^n_S - \boldsymbol{\lambda}^n_R)\}, \tag{94}$$

$$\boldsymbol{\theta}^{n+1} \equiv \Pi_{\Theta}\{\boldsymbol{\theta}^n + \boldsymbol{\gamma}^n_{\boldsymbol{\theta}} \circ \nabla \bar{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}^n)\boldsymbol{\lambda}^n_R\}, \tag{95}$$

$$\boldsymbol{\lambda}^{n+1}_S \equiv (\boldsymbol{\lambda}^n_S - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \circ \mathbf{g}_{\mu_S}(\mathbf{x}^{n+1}))_+ \quad \text{and} \tag{96}$$

$$\boldsymbol{\lambda}^{n+1}_R \equiv \left(\boldsymbol{\lambda}^n_R - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \circ \left(\bar{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}(\boldsymbol{\theta}^{n+1}) - \mathbf{x}^{n+1} - \mathtt{S}(\mu_R)\right)\right)_+, \tag{97}$$

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

We may observe that the algorithm described by (94), (95), (96) and (97) is in general *not implementable*. This is because *neither* the functions $g^o_{\mu_S}$, $\mathbf{g}_{\mu_S}$ and $\bar{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}$ (in general), *nor* their gradients are explicitly known apriori, *for any possible values of* $\mu_S > 0$ and $\mu_R > 0$.

Nevertheless, by Lemma 3 from Section II (assuming that the respective assumptions are satisfied), *all* three functions $g^o_{\mu_S}$, $\mathbf{g}_{\mu_S}$, $\bar{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}$, *and* their derivatives are given by *known expectation functions*. What is more, all these expectation functions depend exclusively on zeroth-order information, that is, on evaluations of $g^o_{\mu_S}$, $\mathbf{g}_{\mu_S}$ and $\bar{\mathbf{f}}^{\boldsymbol{\phi}}_{\mu_R}$, *only*. Therefore, given standard Gaussian *iid* sequences $\{\boldsymbol{U}^n_S\}_{n\in\mathbb{N}^+}$, $\{\boldsymbol{U}^n_S\}_{n\in\mathbb{N}^+}$, together with a channel state observable sequence $\{\mathbf{H}^n\}_{n\in\mathbb{N}^+}$, all mutually independent of each other, a (zeroth-order) stochastic gradient version of the algorithm consisting of (94), (95), (96) and (97) may be readily formulated by replacing all involved expectations as

$$\mathbf{x}^{n+1} \equiv \Pi_{\mathcal{X}}\{\mathbf{x}^n + \boldsymbol{\gamma}^n_{\mathbf{x}} \circ (\Delta^o_g(\mathbf{x}^n, \mu_S, \boldsymbol{U}^{n+1}_S)\boldsymbol{U}^{n+1}_S + \langle \boldsymbol{\Delta}_{\mathbf{g}}(\mathbf{x}^n, \mu_S, \boldsymbol{U}^{n+1}_S), \boldsymbol{\lambda}^n_S\rangle \boldsymbol{U}^{n+1}_S - \boldsymbol{\lambda}^n_R)\}, \tag{98}$$

$$\boldsymbol{\theta}^{n+1} \equiv \Pi_{\Theta}\{\boldsymbol{\theta}^n + \boldsymbol{\gamma}^n_{\boldsymbol{\theta}} \circ \langle \boldsymbol{\Delta}_{\mathbf{f}}(\boldsymbol{\theta}^n, \mu_R, \boldsymbol{U}^{n+1}_R, \mathbf{H}^{n+1}), \boldsymbol{\lambda}^n_R\rangle \boldsymbol{U}^{n+1}_R\}, \tag{99}$$

$$\boldsymbol{\lambda}^{n+1}_S \equiv (\boldsymbol{\lambda}^n_S - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \circ \mathbf{g}(\mathbf{x}^{n+1} + \mu_S \boldsymbol{U}^{n+1}_S))_+ \quad \text{and} \tag{100}$$

$$\boldsymbol{\lambda}^{n+1}_R \equiv \left(\boldsymbol{\lambda}^n_R - \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_R} \circ \left(\mathbf{f}(\boldsymbol{\phi}(\mathbf{H}^{n+1}, \boldsymbol{\theta}^{n+1} + \mu_R \boldsymbol{U}^{n+1}_R), \mathbf{H}^{n+1}) - \mathbf{x}^{n+1} - \mathtt{S}(\mu_R)\right)\right)_+, \tag{101}$$

where, dropping dependencies, the vectors of finite differences $\mathbf{\Delta_g} \in \mathbf{R}^{N_\mathbf{g}}$ and $\mathbf{\Delta_g} \in \mathbf{R}^{N_S}$ are defined as

$$\mathbf{\Delta_g} \triangleq \left[\Delta_g^1 \,\ldots\, \Delta_g^{N_\mathbf{g}}\right]^{\boldsymbol{T}} \text{ and } \mathbf{\Delta_f} \triangleq \left[\Delta_f^1 \,\ldots\, \Delta_f^{N_S}\right]^{\boldsymbol{T}}, \tag{102}$$

respectively. A complete description of the proposed model-free primal-dual method is provided in Algorithm 1. We may readily observe that the algorithm requires exactly *three system probes per user (i.e., of dimension N_S), per iteration*.

Two key differences between Algorithm 1 and the primal-dual method presented in [23] are the presence of the feasibility slack S, which follows from our analysis, as well as the interesting fact that, due to our explicit formulation of the smoothed surrogate (12), the dual updates (100) and (101) are naturally *randomized*, in addition to the primal updates (98) and (99). Most importantly, while the model-free method of [23] relies on policy gradient updates, Algorithm 1 completely bypasses the need for introducing randomized policies into the learning procedure. This also makes Algorithm 1 straightforward to implement, as computation of the gradient of $\boldsymbol{\phi}$ is not required; in fact, Algorithm 1 may be executed as described for any admissible choice of $\boldsymbol{\phi}$, without additional computational requirements.

Additionally, although we do not discuss such settings in detail, we would like to note that the proposed primal-dual approach should work for certain resource allocation programs considered over nonstationary channels, as long as the associated channel process is *ergodic in a sufficiently wide sense*. A typical example would be Markovian Rayleigh fading channels, where small scale fading follows a stable order-1 linear autoregression through time (see, e.g., the model adopted in [34], [35]). Since the distribution of the channel may be time-varying in such cases, expectations in (1) would be with respect to a *limit channel state distribution* (e.g., an invariant measure of an ergodic Markov process), say $\mathcal{M}_{\mathbf{H}}^{\infty}$, and Algorithm 1 would train a deterministic policy $\boldsymbol{\phi}$ in order to maximize the quality of service of the wireless system under consideration *in the long term*, that is, relative to $\mathcal{M}_{\mathbf{H}}^{\infty}$. Of course, if the channel state process is *iid*, then it is true that $\mathcal{M}_{\mathbf{H}}^{\infty} \equiv \mathcal{M}_{\mathbf{H}}$, and the resource allocation task reduces to what we have *implicitly* considered so far.

Lastly, it would be valuable to compare our proposed policy training approach with recent work on (deep) Q-Learning-based schemes for resource allocation in wireless systems, which also consider channel state Markovianity by construction [34]–[38]. There are two important advantages of Algorithm 1, as compared to state-of-the-art (deep) Q-Learning-based schemes, e.g., those recently proposed in [35], [36], [38]. First, (deep) Q-Learning (still) suffers from multiple curses of dimensionality, especially for problems with large or uncountable action spaces; this is in particular the case in many resource allocation problems arising in wireless communications and networking, since the involved (physical-layer) channel state is real-valued. This important issue may be overcome either heuristically, or by appropriate formulation of the resource allocation problem to begin with [34]–[36]. Note that this issue does not occur within our approach, since the PFA $\boldsymbol{\phi}(\cdot, \boldsymbol{\theta})$ is $\mathbf{R}^{N_\phi}$-valued *and* defined directly as a function of the channel state $\mathbf{H}$ (lying in the Euclidean subspace $\mathcal{H}$), which constitutes the *information primitive* of the base problem (1). In a sense, Algorithm 1 stands as a model-free *actor-only* analog to (gradient-based and more complex) actor-critic schemes [34].

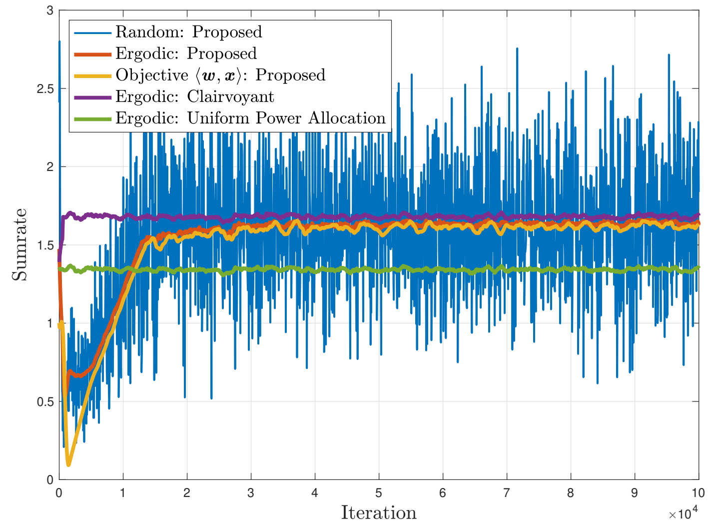

Figure 1: Sumrates (in *nats per unit of time*) achieved by the proposed method, the clairvoyant policy, as well as a uniform power allocation policy, in the case of a simple AWGN channel.

Second, and specifically in regard to *deep* Q-Learning-based resource allocation, training Deep Q-Networks (DQNs) constitutes an often computationally intensive and complicated procedure. In fact, in order to achieve good performance without utilizing a prohibitively large DQN, *problem-specific* feature selection is commonly employed; this is achieved by incorporating model information into the quantities selected as inputs (i.e., the features) to the resulting DQN. Again, guided feature selection is performed heuristically by trial-and-error; see, for instance, [35], [38]. As a result, such state-of-the-art deep Q-Learning schemes are inherently model-based, since knowledge about the underlying system model is explicitly built into the training process of the corresponding DQN. We should note though that complex feature selection in existing works often advocates *distributed execution* of the resulting allocation schemes. Nevertheless, our approach supports this desirable feature as well, via appropriate choice of the adopted policy parameterization. Overall, here we end up with a simple and effective *fully model-free policy training algorithm*, and with an equally simple implementation. Further, Algorithm 1 applies to a large variety of resource allocation tasks (*other* than those in [34]–[38]), while also being technically grounded.

Further core operational issues related to distributed or agent-based implementation, incomplete or delayed channel state observations, etc., are not considered herein, but can be excellent topics for future work.

VI. Numerical Simulations & Discussion

We now numerically confirm and discuss the efficacy of the proposed primal-dual algorithm (Algorithm 1) by application on two basic wireless models, namely, a classical Additive White Gaussian Noise (AWGN) channel, as well as a Multiple Access Interference (MAI) channel. Also, in all simulations presented in this section, the parameterization $\boldsymbol{\phi}$ is appropriately selected from the well-known ϵ-universal class of fully connected, feed-forward DNNs, with ReLU hidden layers and sigmoid output layers, similar to the setting considered in [23].

For the AWGN channel case, we consider a simple multiuser networking scenario where each user is given a dedicated channel

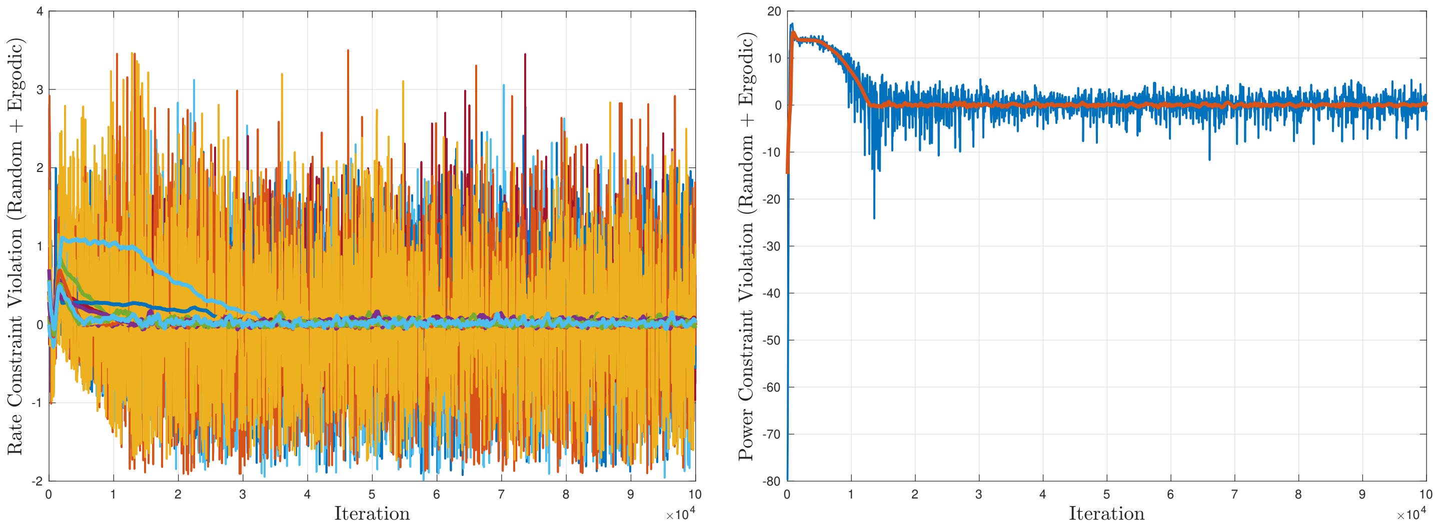

Figure 2: Rate (left) and power (right) constraint violation exhibited by the proposed method, both random and ergodic (smoother curves in both figures), in the case of a simple AWGN channel.

to communicate, with no channel interference. We wish to allocate power between users in order to maximize the weighted sumrate of the network, within a total expected power budget p_{max}, provided as a specification. Given fixed and given user priority weights $w^i \geq 0, i \in \mathbb{N}^+_{N_S}$, selected, without loss of generality, such that $\sum_i w^i \equiv 1$, optimal power allocation may be achieved by solving the stochastic program

$$\begin{array}{ll} \underset{x^i,\boldsymbol{\theta}^i, i\in\mathbb{N}^+_{N_S}}{\text{maximize}} & \sum\limits_{i\in\mathbb{N}^+_{N_S}} w^i x^i \\ \text{subject to} & x^i \leq \mathbb{E}\left\{\log\left(1+\dfrac{H^i\phi^i(H^i,\boldsymbol{\theta}^i)}{\nu^i}\right)\right\} \\ & \mathbb{E}\left\{\sum\limits_{i\in\mathbb{N}^+_{N_S}} \phi^i(H^i,\boldsymbol{\theta}^i)\right\} \leq p_{max} \\ & (x^i,\boldsymbol{\theta}^i) \in \mathbf{R}_+ \times \mathbf{R}^{N_{\phi^i}}, \ \forall i \in \mathbb{N}^+_{N_S} \end{array}, \tag{103}$$

where $H^i \geq 0$ and $\nu^i > 0$ are the fading power and noise variance experienced by the i-th user, and each parameterization $\phi^i : \mathbf{R}_+ \times \mathbf{R}^{N_{\phi^i}} \rightarrow [0, p_{max}]$ is chosen as a DNN with single input, two hidden layers with eight and four neurons, respectively, and a single output, for all $i \in \mathbb{N}^+_{N_S}$. The rest of details in regard to the architecture of each of the involved DNNs follows the discussion above. The reason for choosing N_S uncoupled DNNs, one for each user, comes from the structure of the globally optimal solution to the most general, unparameterized version of problem (103) (mapping to (1)), which, for this simple networking setting, may be efficiently determined [7]. Of course, this solution results in an ultimate benchmark upper bound of the sumrate achieved by any feasible ϵ-universal resource allocation policy, at the expense of assuming complete knowledge of the true information theoretic description of the communication system; for this reason, we hereafter fairly refer to this solution as *clairvoyant*.

By defining vectors

$$\boldsymbol{w} \triangleq [w^1 \ \ldots \ w^{N_S}]^{\boldsymbol{T}}, \tag{104}$$

$$\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}) \triangleq \left[\phi^1(H^1,\boldsymbol{\theta}^1) \ \ldots \ \phi^{N_S}(H^{N_S},\boldsymbol{\theta}^{N_S})\right]^{\boldsymbol{T}} \text{ and} \tag{105}$$

$$\boldsymbol{\nu} \triangleq \left[\nu^1 \ \ldots \ \nu^{N_S}\right]^{\boldsymbol{T}}, \tag{106}$$

problem (103) may be reexpressed in the canonical form (2) as

$$\begin{array}{ll} \underset{\mathbf{x},\boldsymbol{\theta}}{\text{maximize}} & \langle \boldsymbol{w}, \mathbf{x} \rangle \\ \text{subject to} & \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix} \leq \mathbb{E}\begin{bmatrix} \log(\mathbf{1}+\mathbf{H}\circ\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta})\oslash\boldsymbol{\nu}) \\ p_{max} - \mathbf{1}^{\boldsymbol{T}}\boldsymbol{\phi}(\mathbf{H},\boldsymbol{\theta}) \end{bmatrix}, \\ & (\mathbf{x},\boldsymbol{\theta}) \in \mathbf{R}^{N_S}_+ \times \mathbf{R}^{N_{\boldsymbol{\phi}}} \end{array} \tag{107}$$

where $\log(\cdot)$ and "$\oslash$" denote the operations of entrywise logarithm and division, respectively. Therefore, Algorithm 1 is applicable to problem (107) and, in turn, (103), by considering the corresponding smoothed surrogate based on (12).

Since the objective of (103) is usually perfectly known to the wireless engineer, we may set $\mu_S \equiv 0$. In other words, in the corresponding smoothed surrogate (cf. (12)), Gaussian smoothing is applied only to the constraints of problem (107).

Remark 18. Note that, in (103), H^i denotes the *square* of the fading channel experienced by user i. Nevertheless, since the square function is one-to-one and onto on the nonnegative reals, taking directly the square of the involved channels is virtually consistent with our setting established in Section I. ■

To assess the performance of the proposed Algorithm 1 on problem (107), we assume $N_S \equiv 10$ users, set $p_{max} \equiv 20$, and consider a randomly generated weight vector $\boldsymbol{w}$. We also assume that $\nu^i \equiv 1$, and that 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 10^5 iterations, with initial values $\mathbf{x}^0 \equiv \mathbf{1}, \boldsymbol{\theta}^0 \equiv \mathbf{0}$ and $\boldsymbol{\lambda}^0_S \equiv \mathbf{1}_{N_S+1}$, constant stepsizes $\boldsymbol{\gamma}^n_{\mathbf{x}} \equiv 0.001\mathbf{1}, \boldsymbol{\gamma}^n_{\boldsymbol{\theta}} \equiv 0.0008\mathbf{1}, \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \equiv \left[0.008\mathbf{1}^{\boldsymbol{T}}_{N_S} \ 0.0001\right]^{\boldsymbol{T}}$, for all $n \in \mathbb{N}$, null feasibility slack $\mathtt{S} \equiv \mathbf{0}$, and with the smoothing parameter set as $\mu_R \equiv 10^{-9}$.

Remark 19. Our seemingly trivial choice for the feasibility slack ($\mathtt{S} \equiv \mathbf{0}$) is justified because, as long as μ_R is small enough with

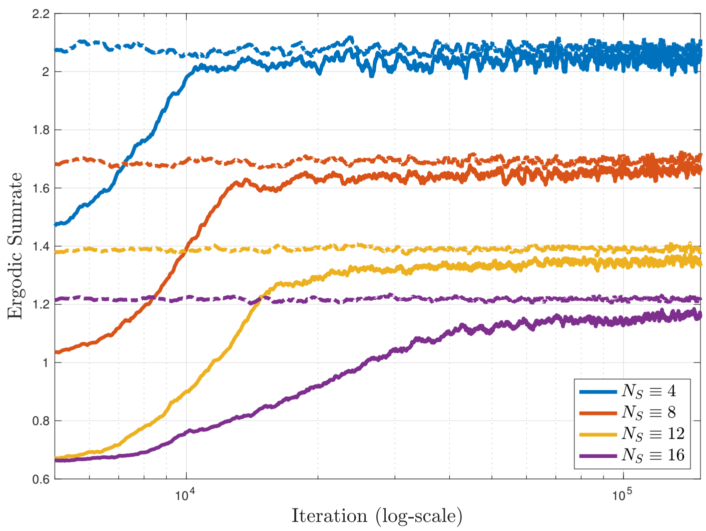

Figure 3: Ergodic sumrates (smoothened) achieved by the proposed method for different values of N_S, in the case of a simple AWGN channel.

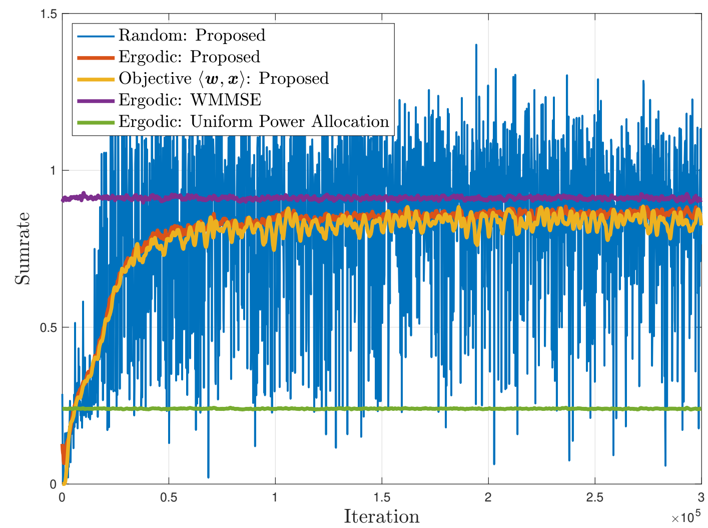

Figure 4: Sumrates (*nats per unit of time*) achieved by the proposed method, the WMMSE policy, and a deterministic uniform power allocation policy, in the case of a MAI channel.

$\mu_R \ll \sqrt{N_{\boldsymbol{\phi}}}$ (true in all our simulations as, *at most*, $N_{\boldsymbol{\phi}} \approx 1000$), any reasonable value for the constant vector $\mathbf{C}$, even if different from $\boldsymbol{c}_R$, would result in tiny, practically unnoticeable constraint violation, as asserted by Theorem 7. ■

Fig. 1 shows the evolution of the sequence of objective values $\{\langle \boldsymbol{w}, \mathbf{x}^n \rangle\}_{n\in\mathbb{N}}$, the instantaneous sumrate sequence $\{\langle \boldsymbol{w}, \log(\mathbf{1} + \mathbf{H}^n \circ \boldsymbol{\phi}(\mathbf{H}^n, \boldsymbol{\theta}^n) \oslash \boldsymbol{\nu})\rangle\}_{n\in\mathbb{N}}$, as well as an approximation of the ergodic sumrate sequence $\{\langle \boldsymbol{w}, \mathbb{E}\{\log(\mathbf{1} + \mathbf{H} \circ \boldsymbol{\phi}(\mathbf{H}, \boldsymbol{\theta}^n) \oslash \boldsymbol{\nu})\}\rangle\}_{n\in\mathbb{N}}$. The ergodic performance of Algorithm 1, expressed by the latter estimate, is also compared with the ergodic performance of the unparameterized, globally optimal power allocation policy solving (1) (the clairvoyant), as well as that of a deterministic uniform power allocation policy across users. All ergodic estimates were computed via simple moving average smoothing of the respective process realizations.

Fig. 1 readily demonstrates that the values of the objective of (107) match the values of the estimated ergodic sumrate, as both obtained from Algorithm 1. At the same time, the ergodic sumrate obtained from Algorithm 1 converges remarkably close to that achieved by the clairvoyant policy, which assumes full knowledge of the model describing the wireless system. Therefore, in this case, the proposed zeroth-order primal-dual method attains actually near-optimal system performance.

Fig. 2 shows similar type estimates (instantaneous and ergodic) concerning violation of the rate and power constraints of problem (107), during execution of Algorithm 1 (positive values indicate constraint violation). We observe that all constraints are active (i.e., met with equality) on average, which confirms that the proposed primal-dual-method indeed converges to feasible power allocation policies, while achieving maximal ergodic rates on a per user basis, as desired. We emphasize that, contrary to the clairvoyant solution, such good performance of Algorithm 1 is achieved without the availability of a baseline model of the wireless system, and at the absence of gradient information of information rate functions, as well as DNN parameterizations.

Our discussion regarding the AWGN channel case is concluded by discussing Fig. 3, which shows smoothened paths generated by the proposed algorithm for different values of N_S (i.e., the number of users in the network), each contrasted with the respective maximal ergodic sumrate achieved by the strictly optimal, clairvoyant policy. The total numbers of decision variables ($N_S + N_{\boldsymbol{\phi}} \equiv N_S(1 + N_{\boldsymbol{\phi}^1})$) for each of the cases considered in Fig. 1 are 228, 456, 684 and 912, respectively. We observe that, while the rate of convergence gets slower as N_S increases (as expected), the performance achieved by our method is consistently very close to that obtained by the strictly optimal, clairvoyant policy. This also provides empirical evidence that the proposed primal-dual method is correct.

Although the dependence of the performance of our algorithm relative to N_S and $N_{\boldsymbol{\phi}}$ is not guaranteed to be sublinear (in general), it is possible to make this dependence better, if needed, by exploiting mini-batches of stochastic finite differences, at the trade-off, however, of increasing the number of required system probes [56]. Further, by using more efficient policy parameterizations which exploit (stochastic) network structure, such as Random Edge Graph Neural Networks (REGNNs) [24], it is possible to not only reduce the number of the involved free parameters ($N_{\boldsymbol{\phi}}$ in particular), but also train REGNN-based resource allocation policies on smaller, moderately sized wireless systems, and then successfully extrapolate to large-scale wireless systems; this is a property of REGNNs is known as *transference* [24]. We would also like to mention that detailed convergence rate analysis of zeroth-order primal-dual methods for nonconvex problems is currently a challenging and open research topic [56]; of course, such an analysis is out of the scope of this paper.

Remark 20. Note that, although the adopted statistical channel model is the same for all network links, the ergodic rate estimates in Fig. 2 (left) are different from each other. This is mainly due to two reasons. First, problem (103) does not treat each user the same way. This is because each user i is assigned a random priority weight $w_i > 0$, present in the linear objective of (103). Second, the empirical results that we provide refer to a *single sample path* generated by our method, and not averaged experiment trials. As a result, until convergence, our

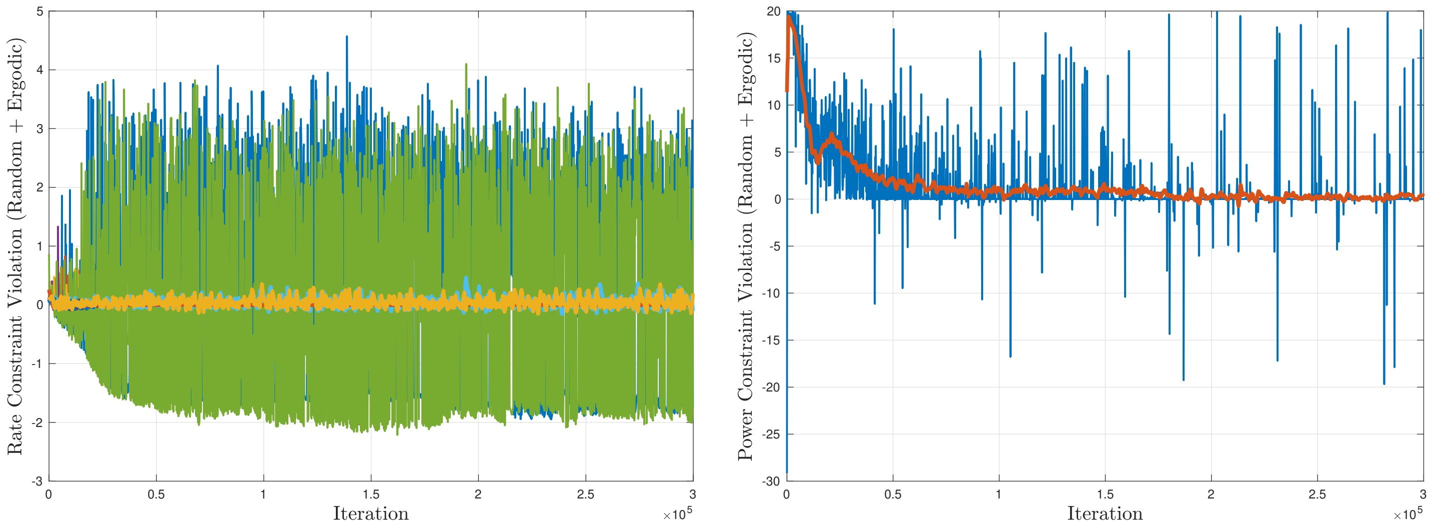

Figure 5: Rate (left) and power (right) constraint violation exhibited by the proposed method, both random and ergodic (smoother curves in both figures), in the case of a MAI channel.

(pseudo-)ergodic estimates, which we obtain via moving average smoothing, will be initially biased, depending on the particular channel realization for every user. Exactly the same observations are true for our second simulation case, as discussed below. ■

Next, we consider the case of a MAI channel, where N_S transmitters simultaneously communicate with a central node, for instance, a common receiver, or a base station. In this case, the signal transmitted by each user creates interference to the signals transmitted by all other users in the network. As before, 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}. Working similarly to the AWGN channel case discussed above, we may formulate the stochastic program

$$\begin{aligned} \underset{x^i, i\in\mathbb{N}^+_{N_S}, \boldsymbol{\theta}}{\text{maximize}} \quad & \sum_{i\in\mathbb{N}^+_{N_S}} w^i x^i \\ \text{subject to} \quad & x^i \le \mathbb{E}\left\{\log\left(1+\frac{H^i\phi^i(\mathbf{H},\boldsymbol{\theta})}{\nu^i+\sum_{i\neq j} H^j\phi^j(\mathbf{H},\boldsymbol{\theta})}\right)\right\}, \\ & \mathbb{E}\left\{\sum_{i\in\mathbb{N}^+_{N_S}} \phi^i(\mathbf{H},\boldsymbol{\theta})\right\} \le p_{max} \\ & (x^i,\boldsymbol{\theta}) \in \mathbf{R}_+ \times \mathbf{R}^{N_\phi},\ \forall i\in\mathbb{N}^+_{N_S} \end{aligned} \tag{108}$$

where each parameterization $\phi^i : \mathbf{R}^{N_S}_+ \times \mathbf{R}^{N_\phi} \to [0, p_{max}]$, $i\in\mathbb{N}^+_{N_S}$ is an element of the output layer of a *single* DNN taking as input the full fading channel vector $\mathbf{H}\in\mathbf{R}^{N_S}_+$, and having two hidden layers with thirty-two and sixteen neurons, respectively. The intuition behind the adopted multiple-input multiple-output DNN architecture lies in the strong coupling among the channels of all users in every rate constraint of (108). The rest of details in regard to the architecture of each of the involved DNNs follows the discussion above.

As before, problem (108) may be reexpressed in the form of (2); however, the details are slightly more technical compared to the case of an AWGN channel, and are omitted for brevity.

In our simulations for this setting, 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 follows the same exponential distribution as before, for all $i \in \mathbb{N}^+_{N_S}$. Then, we execute Algorithm 1 for $3\cdot 10^5$ iterations, with initial values $\mathbf{x}^0 \equiv \boldsymbol{\theta}^0 \equiv \mathbf{0}$ and $\boldsymbol{\lambda}^0_S \equiv \mathbf{1}_{N_S+1}$, constant stepsizes $\boldsymbol{\gamma}^n_{\mathbf{x}} \equiv 0.0008\mathbf{1}, \boldsymbol{\gamma}^n_{\boldsymbol{\theta}} \equiv 0.0005\mathbf{1}, \boldsymbol{\gamma}^n_{\boldsymbol{\lambda}_S} \equiv \left[0.005\mathbf{1}^{\boldsymbol{T}}_{N_S}\ 0.0001\right]^{\boldsymbol{T}}$, for all $n\in\mathbb{N}$, null feasibility slack $\mathtt{S}\equiv\mathbf{0}$, and with the smoothing parameter set as $\mu_R \equiv 10^{-9}$. As before, we set $\mu_S\equiv 0$, that is, the objective of (108) is reasonably assumed known.

Fig. 4 shows the evolution of the sequence of objective values $\{\langle \boldsymbol{w}, \mathbf{x}^n\rangle\}_{n\in\mathbb{N}}$ and, as before, 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 (108) is unavailable mainly due to nonconvexity of the involved rate constraints (cf. (1)), we compare the ergodic performance of Algorithm 1 with that of the well-known WMMSE policy [15], which, in our setting, provides an iterative algorithm for finding an approximate solution, *for each fading realization*, to the *deterministic* sumrate maximization problem

$$\begin{aligned} \underset{p^i, i\in\mathbb{N}^+_{N_S}}{\text{maximize}} \quad & \sum_{i\in\mathbb{N}^+_{N_S}} w^i \log\left(1+\frac{H^i p^i}{\nu^i+\sum_{i\neq j} H^j p^j}\right) \\ \text{subject to} \quad & \sum_{i\in\mathbb{N}^+_{N_S}} p^i \le p_{max} \\ & p^i \ge 0,\ \forall i\in\mathbb{N}^+_{N_S} \end{aligned} . \tag{109}$$

Note that, as the form of problem (109) suggests, the WMMSE heuristic assumes complete knowledge of the information theoretic model of the wireless system. In our simulations, WMMSE is for 50 iterations, for each observed channel realization. For reference, Fig. 4 also shows the ergodic performance achieved by a uniform power allocation policy across users. As before, all ergodic estimates were computed via simple moving average smoothing of the respective process realizations.

Fig. 4 confirms that Algorithm 1 exhibits similar behavior as in the AWGN channel case previously discussed, but for the significantly more complicated resource allocation problem (108). Again, the objective of (108) and the ergodic sumrate obtained from Algorithm 1 match, whereas the latter converges rather close to the ergodic sumrate achieved by WMMSE.

Instantaneous and ergodic estimates of the rate and power constraint violation of the decisions produced by Algorithm 1 for problem (108) are provided in Fig. 5 (again, positive values indicate constraint violation). As in the AWGN channel case, all constraints are active on average, confirming that the proposed zeroth-order primal-dual method produces feasible and near-state-of-the-art power allocation policies without knowledge of a system model and in absence of gradient information, verifying the effectiveness of the method in a model-agnostic setting.

In comparison with Algorithm 1, WMMSE produces decisions which satisfy the constraints of problem (108) for every channel realization, by construction. However, this is achieved at the expense of solving one full optimization problem per channel realization (i.e., at every "iteration" in our graphs). In the literature, the approach taken by WMMSE is well-known as a Cost Function Approximation (CFA) [40].

On the other hand, the proposed model-free primal-dual algorithm produces a policy *by training*, that is, a *known and tractable* function (here, a DNN), which outputs (near-)optimal resource allocation decisions for all possible channel states, *simultaneously*. This policy satisfies the constraints of the postulated stochastic resource allocation problem on average (as it should, of course, based on our formulation). Lastly, we would also like to mention that the percentage or rate of constraint violation (corresponding to positive values in Fig. 5) relative to the number of observed channel states is dependent on both the particular problem setting and the adopted policy parameterization.

VII. Conclusion

We have considered the general problem of learning optimal resource allocation policies in wireless systems, under a model-free, data-driven setting. Starting with a generic variational formulation of the resource allocation problem, and driven by its intractability in most wireless networking scenarios, we focused on parametric policy function approximations. Leveraging classical results on Gaussian smoothing, we first showed that it is possible to crucially simplify gradient evaluation for all utility and service functions involved, by appropriately constructing a finite dimensional, smoothed surrogate to the original variational problem. Then, assuming near-universal policy parameterizations, e.g., Deep Neural Networks (DNNs), we completely characterized the duality gap between the original problem and the dual of the proposed surrogate, establishing linear dependence of this duality gap relative to smoothing and near-universality parameters. In fact, this gap may be made arbitrarily small at will. Motivated by our results, and in conjunction with the special properties of the proposed smoothed surrogate, we also developed a zeroth-order stochastic primal-dual algorithm, enabling completely model-free, data-driven optimal resource allocation for ergodic network optimization. Our simulations show that DNN-based, data-driven policies produced by the proposed primal-dual method attain near-ideal performance, relying exclusively on limited system probing, completely bypassing the need for gradient computations and policy randomization, and at the absence of baseline channel or information rate models.

References

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

[2] 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.

[3] 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.

[4] N. Naderializadeh and A. S. Avestimehr, "ITLinQ: A New Approach for Spectrum Sharing in Device-to-Device Communication Systems," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1139–1151, Jun. 2014.

[5] 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*. IEEE, Mar. 2006, pp. 339–344.

[6] X. Wu, S. Tavildar, S. Shakkottai, T. Richardson, J. Li, R. Laroia, and A. Jovicic, "FlashLinQ: A Synchronous Distributed Scheduler for Peer-to-Peer Ad Hoc Networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 4, pp. 1215–1228, Aug. 2013.

[7] 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.

[8] 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*. IEEE, Sep. 2016, pp. 1–6.

[9] Yichuan Hu and A. Ribeiro, "Adaptive Distributed Algorithms for Optimal Random Access Channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2703–2715, Aug. 2011.

[10] Y. Hu and A. Ribeiro, "Optimal Wireless Networks Based on Local Channel State Information," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4913–4929, Sep. 2012.

[11] X. Wang and G. B. Giannakis, "Resource Allocation for Wireless Multiuser OFDM Networks," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4359–4372, Jul. 2011.

[12] 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*, vol. 61, no. 1, Jan. 2012, pp. 249–257.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] J.-A. Bazerque and G. B. Giannakis, "Distributed Scheduling and Resource Allocation for Cognitive OFDMA Radios," in *2007 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications*. IEEE, Aug. 2007, pp. 343–350.

[18] A. S. Bedi and K. Rajawat, "Asynchronous Incremental Stochastic Dual Descent Algorithm for Network Resource Allocation," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2229–2244, May 2018.

[19] ——, "Network Resource Allocation via Stochastic Subgradient Descent: Convergence Rate," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 2107–2121, May 2018.

[20] 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.

[21] M. Eisen, K. Gatsis, G. J. Pappas, and A. Ribeiro, "Learning in Wireless Control Systems Over Nonstationary Channels," *IEEE Transactions on Signal Processing*, vol. 67, no. 5, pp. 1123–1137, Mar. 2019.

[22] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Random Access Design for Wireless Control Systems," *Automatica*, vol. 91, pp. 1–9, May 2018.
[23] 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.
[24] M. Eisen and A. Ribeiro, "Optimal Wireless Resource Allocation with Random Edge Graph Neural Networks," *arXiv preprint, arXiv:1909.01865*, Sep. 2019.
[25] 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.
[26] D. Xu, X. Che, C. Wu, S. Zhang, S. Xu, and S. Cao, "Energy-Efficient Subchannel and Power Allocation for HetNets Based on Convolutional Neural Network," *arXiv preprint, arXiv:1903.00165*, Mar. 2019.
[27] L. Lei, L. You, G. Dai, T. X. Vu, D. Yuan, and S. Chatzinotas, "A Deep Learning Approach for Optimizing Content Delivering in Cache-Enabled HetNet," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, Aug. 2017, pp. 449–453.
[28] 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.
[29] 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.
[30] P. de Kerret, D. Gesbert, and M. Filippone, "Team Deep Neural Networks for Interference Channels," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, May 2018, pp. 1–6.
[31] F. Liang, C. Shen, W. Yu, and F. Wu, "Towards Optimal Power Control via Ensembling Deep Neural Networks," *arXiv preprint, arXiv:1807.10025*, Jul. 2018.
[32] 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.
[33] 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)*. IEEE, May 2017, pp. 1–6.
[34] 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.
[35] Y. S. Nasir and D. Guo, "Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2239–2250, Oct. 2019.
[36] R. Amiri, M. A. Almasi, J. G. Andrews, and H. Mehrpouyan, "Reinforcement Learning for Self Organization and Power Control of Two-Tier Heterogeneous Networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 3933–3947, Jun. 2019.
[37] L. Liang, H. Ye, G. Yu, and G. Y. Li, "Deep-Learning-Based Wireless Resource Allocation with Application to Vehicular Networks," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 341–356, Feb. 2020.
[38] N. Naderializadeh, J. Sydir, M. Simsek, and H. Nikopour, "Resource Management in Wireless Networks via Multi-Agent Deep Reinforcement Learning," *arXiv preprint, arXiv:2002.06215*, 2020.
[39] H. Lee, S. H. Lee, and T. Q. Quek, "Deep Learning for Distributed Optimization: Applications to Wireless Resource Management," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2251–2266, Oct. 2019.
[40] W. B. Powell, "A Unified Framework for Stochastic Optimization," *European Journal of Operational Research*, vol. 275, no. 3, pp. 795–821, Jun. 2019.
[41] J. Park and I. W. Sandberg, "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, vol. 3, no. 2, pp. 246–257, Jun. 1991.
[42] B. K. Sriperumbudur, K. Fukumizu, and G. R. G. Lanckriet, "On the Relation between Universality, Characteristic Kernels and RKHS Embedding of Measures," in *13th Intern. Conf. Artificial Intelligence and Statistics*, no. 9, Mar. 2010, pp. 773–780.
[43] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
[44] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 2nd ed. Wiley-Interscience, 2011.
[45] D. Bertsekas, *Dynamic Programming & Optimal Control*, 4th ed. Belmont, Massachusetts: Athena Scientific, 2012, vol. II.
[46] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.
[47] H. J. H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.
[48] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley-Interscience, 2003.
[49] 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.
[50] R. B. Ash and C. Doléans-Dade, *Probability and Measure Theory*. Academic Press, 2000.
[51] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[52] A. P. Ruszczyński, *Nonlinear Optimization*. Princeton University Press, 2006.
[53] D. P. Bertsekas, *Convex optimization theory*, 1st ed. Athena Scientific, 2009.
[54] R. Montes-de Oca and E. Lemus-Rodríguez, "An unbounded Berge's minimum theorem with applications to discounted Markov decision processes," *Kybernetika*, vol. 48, no. 2, pp. 268–286, 2012. [Online]. Available: www.kybernetika.cz/content/2012/2/268
[55] C. D. Aliprantis and K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media, 2006.
[56] Z. Wang, K. Balasubramanian, S. Ma, and M. Razaviyayn, "Zeroth-Order Algorithms for Nonconvex Minimax Problems with Improved Complexities," *arXiv preprint, arXiv:2001.07819*, Jan. 2020.