Strategic Generation Investment in Energy Markets: A Multiparametric Programming Approach

Sina Taheri *Student Member, IEEE*, Vassilis Kekatos, *Senior Member, IEEE*, and Harsha Veeramachaneni

Abstract—An investor has to carefully select the location and size of new generation units it intends to build, since adding capacity in a market affects the profit from units this investor may already own. To capture this closed-loop characteristic, strategic investment (SI) of generation can be posed as a bilevel optimization. By analytically studying a small market, we first show that its objective function can be non-convex and discontinuous. Realizing that existing mixed-integer problem formulations become impractical for larger markets and number of instances, this work put forth two SI solvers: a grid search to handle setups where the candidate investment locations are few, and a stochastic gradient descent approach for otherwise. Both solvers leverage powerful results of multiparametric programming (MPP), each in a unique way. The grid search entails finding the primal/dual solutions for a large number of optimal power flow (OPF) problems, which nonetheless can be efficiently computed several at once thanks to the properties of MPP. The same properties facilitate the rapid calculation of gradients in a mini-batch fashion, thus accelerating the implementation of a stochastic (sub)-gradient descent search. Tests on the IEEE 30- and 118-bus systems using real-world data corroborate the advantages of the novel solvers.

Index Terms—Mathematical programming with equilibrium constraints; bilevel programming; locational marginal prices; stochastic gradient descent; strategic investment; multiparametric programming.

I. INTRODUCTION

SUPPOSE an investor intends to build one or more power plants to participate in an electricity market. The investor may already own units bidding in the same market. The new units to be built by the investor will add generation capacity to the market, which depending on transmission congestion and load demand, may alter electricity prices and generation schedules in a way so that the investor's total financial gain from existing and new units is lowered. The goal of the investor is to find the optimal location, size, and possibly timing of the new generation units to maximize its total profit. This task of *strategic investment (SI)* is challenging for three reasons. First, the variables involved in SI, namely the generation schedules and prices, are not known beforehand but are computed as the solutions of an optimization problem, a linearized optimal power flow (OPF). Second, an investment can change the market outcome (prices and generation schedules), rendering SI a complex closed-loop problem. Third, increasing uncertainties introduced by renewable generation, loads, fuel prices, and bids from rival generators, call for stochastic methods thus further increasing complexity for SI.

A promising method to handle the closed-loop complication of the SI task is posing it as a bilevel optimization [1]. The inner level involves the OPF that clears the market and decides generation schedules and prices given generation capacities. The outer level aims to maximize the market profit for dispatching new and existing units minus the investment cost for the new units. Nonetheless, this bilevel formulation calls for complex complementarity methods [2]; see [3] for a comprehensive survey. In the case of market OPFs with linear constraints, complementarity methods convert the bilevel problem into a single-level optimization upon replacing the inner problem with its Karush-Kuhn-Tucker (KKT) conditions. While complementarity methods promise globally optimal investment decisions, they entail computationally prohibitive mixed-integer programs. Such models may not scale gracefully in large power networks, and may not be able to consider a sufficient number of scenarios. Alternatively, works like [4], [5], [6] use scenario-based alternating direction method of multipliers or progressive hedging to decompose the related mixed-integer programs, at the expense of losing global optimality.

Alternatively, the SI task could be dealt with by solving the OPF clearing the market for each possible combination of system scenarios and investment options. However, this process is also challenging due to the sheer number of OPFs that need to be solved, calling for efficient OPF solvers for a large number of market scenarios. Interestingly, the OPF problem under the linearized grid model, the so-called DC-OPF, can be viewed as an instance of multiparametric programming (MPP), where loads, generation capacities, and bids are considered as its parameters. According to MPP and as explained in Section III-B, the space of the parameter vector can be partitioned into polytopes termed *critical regions* for which the primal and dual solutions can be identified as affine functions of the problem parameters [7], [8]. The boundaries of these regions as well as the associated affine functions depend on which constraints are active at optimality, and hence, the DC-OPF needs to be solved only once per critical region. This latter property facilitates solving a large number of DC-OPFs with relatively small computational burden.

MPP has been utilized before in power systems operations. The notion of congestion patterns in energy markets identified by [9] pertains exactly to the critical regions of MPP. The

same regions also give rise to the active sets learned in [10]. References [11] combined MPP with importance sampling over critical regions to compute the probability distribution of locational marginal prices (LMPs). The polytopic description of critical regions allows [12] to train a support vector machine classifier and estimate LMPs given loads. Reference [13] utilizes MPP and proposes a critical region exploration algorithm to solve a security-constrained economic dispatch. In the context of distribution grids, references [14] and [15] leverage MPP to handle efficiently a large number of distribution OPF instances, and thus expedite probabilistic hosting capacity analysis. However, none of the previous works engages MPP to deal with the complex bilevel setup involved in SI.[1]

The contribution of this work is fourfold: *c1)* Study SI analytically for a simple power network to demonstrate the challenges involved; *c2)* Extend existing MPP claims to the OPF problem used to clear electricity markets; *c3)* Develop an algorithm to compute efficiently the primal/dual outcomes of hundreds of OPF scenarios at a time. The algorithm can accelerate by an order of magnitude a brute-force grid search to cope with the SI task when the number of investment locations and scenarios are relatively small; and *c4)* Devise a stochastic gradient descent (SGD) scheme to address directly the outer layer of the SI task, especially when multiple investment locations are considered. By uniquely exploiting MPP properties, this SGD scheme calculates gradients in a highly scalable mini-batch fashion.

The rest of this work is organized as follows. Section II models electricity markets, poses the SI problem, and showcases its difficulty via a simple example. Section III-A reviews the solution to SI based on mathematical programming with equilibrium constraints (MPEC) pursued in [2], while Section III-B extends MPP to the problem at hand. Section IV extends properties of MPP to develop an efficient grid search and a stochastic gradient descent approach. The two algorithms are numerically tested in Section V.

Regarding *notation*, column vectors (matrices) are denoted by lowercase (uppercase) boldface letters; calligraphic symbols are reserved for sets and/or mappings. The n-th element of $\mathbf{x}$ is denoted by x_n, and the (n,m)-th entry of $\mathbf{X}$ by X_{nm}. Symbol $\mathbf{1}$ denotes the all-one vector. Inequalities between vectors, such as $\mathbf{x} \geq \mathbf{y}$, apply entry-wise.

II. STRATEGIC INVESTMENT IN ELECTRICITY MARKETS

A. Modeling Electricity Markets

Suppose the energy market operates over a system with N buses and L transmission lines. In a wholesale electricity market, the independent system operator (ISO) calculates the generation schedule and electricity prices upon solving a linear or quadratic program to minimize the total generation cost subject to power balance and line flow constraints. From the viewpoint of a strategic investor, one can identify three types of generators [2]: existing units owned by rival entities; existing units owned by the investor; and new units to be built by

---

[1]In fact, during the final review round and a year after the preprint version of this work was publicized, a related work using MPP for dealing with a tri-level transmission/distribution expansion planning task was published [16].

the investor. The power schedules corresponding to three unit types are denoted respectively by $(\mathbf{p}_r, \mathbf{p}_e, \mathbf{p}_n)$. For notational brevity, suppose all unit types exist at all buses with possibly zero capacities. The market is cleared by the DC-OPF:

$$\min_{\mathbf{p}_r, \mathbf{p}_e, \mathbf{p}_n} \quad f_r(\mathbf{p}_r) + f_e(\mathbf{p}_e) + f_n(\mathbf{p}_n) \tag{1a}$$

$$\text{s.to} \quad \mathbf{1}^\top(\mathbf{p}_r + \mathbf{p}_e + \mathbf{p}_n - \boldsymbol{\ell}) = 0 \quad : \lambda_0 \tag{1b}$$

$$-\bar{\mathbf{f}} \leq \mathbf{S}(\mathbf{p}_r + \mathbf{p}_e + \mathbf{p}_n - \boldsymbol{\ell}) \leq \bar{\mathbf{f}} \quad : \underline{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}} \tag{1c}$$

$$\mathbf{0} \leq \mathbf{p}_r \leq \bar{\mathbf{p}}_r \quad : \underline{\boldsymbol{\gamma}}_r, \overline{\boldsymbol{\gamma}}_r \tag{1d}$$

$$\mathbf{0} \leq \mathbf{p}_e \leq \bar{\mathbf{p}}_e \quad : \underline{\boldsymbol{\gamma}}_e, \overline{\boldsymbol{\gamma}}_e \tag{1e}$$

$$\mathbf{0} \leq \mathbf{p}_n \leq \bar{\mathbf{p}}_n \quad : \underline{\boldsymbol{\gamma}}_n, \overline{\boldsymbol{\gamma}}_n \tag{1f}$$

where $\mathbf{S}$ is the power transfer distribution factor matrix (PTDF) [17]. Function $f_r(\mathbf{p}_r) := \frac{1}{2}\mathbf{p}_r^\top \mathbf{H}_r \mathbf{p}_r + \mathbf{c}_r^\top \mathbf{p}_r$ models the generation cost for rival units. The diagonal matrix $\mathbf{H}_r$ and vector $\mathbf{c}_r$ contain positive values [18]. The generation costs for existing and new units f_e and f_n are defined similarly. Constraint (1b) ensures power balance with $\boldsymbol{\ell}$ being the vector of nodal load demands. Constraint (1c) enforces given line flow limits $\bar{\mathbf{f}}$. Constraints (1e)–(1d) impose capacity limits $(\bar{\mathbf{p}}_r, \bar{\mathbf{p}}_e, \bar{\mathbf{p}}_n)$ on generation schedules. Dual variables are shown in the right-hand side of the constraints in (1).

To account for renewable generation, vector $\bar{\mathbf{p}}_r$ is the *available* capacity of rival units. It can be modeled as $\bar{\mathbf{p}}_r = \boldsymbol{\alpha}_r \odot \hat{\mathbf{p}}_r$, where $\hat{\mathbf{p}}_r$ is the vector of installed capacities, $\boldsymbol{\alpha}_r$ the vector of capacity factors, and symbol $\odot$ denotes entry-wise multiplication between two vectors. For non-renewable generators, the corresponding entry of $\boldsymbol{\alpha}_r$ is unity, whereas for renewable generators it changes with time to capture the available wind energy as a percentage of the maximum capacity. We similarly define vectors $(\boldsymbol{\alpha}_e, \hat{\mathbf{p}}_e)$ for existing units, and $(\boldsymbol{\alpha}_n, \mathbf{x} = \hat{\mathbf{p}}_n)$ for new units. We will be using $\mathbf{x}$ instead of $\hat{\mathbf{p}}_n$ to emphasize that the capacity of new units is the ultimate optimization variable for the SI task at hand. In other words, the investor would eventually build new generation capacities $\mathbf{x}$.

The ISO solves (1) every hour to find the optimal schedules $(\mathbf{p}_r, \mathbf{p}_e, \mathbf{p}_n)$ and computes the locational marginal prices (LMPs) of electricity across buses as; see e.g., [17], [19] for details

$$\boldsymbol{\pi} = -\lambda_0 \mathbf{1} + \mathbf{S}^\top(\underline{\boldsymbol{\mu}} - \overline{\boldsymbol{\mu}}). \tag{2}$$

We have slightly abused notation and used the same symbols with (1) to denote the optimizers of the problem. We next present SI adapting the formulation of [2].

It should be emphasized that the energy market model adopted here is a simplification of the actual clearing process. Energy markets are increasingly complex mechanisms. Under the typical setup of two-stage markets (day-ahead and real-time), the minimum generation capacity may not be zero at times, unit commitment (UC) decisions can keep generators out of the economic dispatch, while ramping and UC constraints call for multi-period dispatches that may dispatch generators in an out-of-merit fashion and use complex make-whole or convex-hull pricing and uplift payments. The situation becomes even more complicated with the option for generators to engage into bilateral agreements, the simultaneous participation in energy and reserves markets, or even

considering the security-constrained renditions of DC-OPF the ISO will be solving to dispatch generators. In addition to the complexity of the market clearing process per se, an investor may oftentimes have to deal with only partial knowledge on market parameters, such as the exact transmission network and line maintenance schedules, as well as the bidding strategies and new investments by competitors. Granted the complexity and uncertainty involved, the proposed formulation serves as a first-order approximation of the actual clearing process, as pursued in prior literature too [2].

B. Problem Formulation

Strategic investment in electricity markets can be viewed as a minimization problem where the objective is the cost for investing in the new units minus the expected revenue obtained from the market through the new units and the existing own units. Investment costs are being normalized across the expected lifetime of the new units. The strategic investment study obviously relies on the market conditions anticipated over the time period of interest. Such conditions may include upgrades in transmission network and competitor's portfolios, as well as changes in operational costs. This investment cost is generally modeled as a known linear function $\mathbf{k}^\top \mathbf{x}$ of the generation capacity. The revenue is made up by the payment received from the ISO (generation schedule times LMP) minus the true generation cost

$$f(\mathbf{x}) := \mathbf{k}^\top \mathbf{x} - \mathbb{E}\left[\boldsymbol{\pi}^\top (\mathbf{p}_e + \mathbf{p}_n) - g_e(\mathbf{p}_e) - g_n(\mathbf{p}_n) \right] \quad (3)$$

where the expectation $\mathbb{E}[\cdot]$ applies over the involved uncertainties. Note f involves the actual cost of generation $g_e(\mathbf{p}_e) + g_n(\mathbf{p}_n)$ rather than the bid $f_e(\mathbf{p}_e) + f_n(\mathbf{p}_n)$ submitted to the market. This is because the market bid can be sometimes larger than the actual generation cost for some or all $(\mathbf{p}_e, \mathbf{p}_n)$ [2]. The expectation in (3) is applied over all random quantities, such as the demand vector $\boldsymbol{\ell}$, the scaling factors $(\boldsymbol{\alpha}_r, \boldsymbol{\alpha}_e, \boldsymbol{\alpha}_n)$ for renewable generation, and possible changes in bids.

The task of strategic investment can be now stated as

$$\min_{\mathbf{x} \in \mathcal{X}} \quad f(\mathbf{x}) \quad (4a)$$

$$\text{s.to} \quad \{\boldsymbol{\pi}, \mathbf{p}_e, \mathbf{p}_n\} \text{ being solutions of (1).} \quad (4b)$$

In addition to constraint (4b), the investment variable $\mathbf{x}$ should also belong to the set of investment options $\mathcal{X} := \{\mathbf{x} : \underline{\mathbf{x}} \leq \mathbf{x} \leq \overline{\mathbf{x}}, \ \boldsymbol{\Delta}\mathbf{x} \leq \boldsymbol{\delta}\}$. Constraint $\boldsymbol{\Delta}\mathbf{x} \leq \boldsymbol{\delta}$ could model an upper bound on the total MW capacity installed or the total number of wind turbines purchased. In this case, matrix $\boldsymbol{\Delta}$ degenerates to the all-one vector $\boldsymbol{\Delta} = \mathbf{1}^\top$ and vector $\boldsymbol{\delta}$ to a scalar x_{total}, where x_{total} is the total capacity to be installed. Investments $\mathbf{x}$ may also be restricted to take discrete values. As in [2], we further postulate two assumptions on the problem setup.

Assumption 1. *The transmission network topology captured by $(\mathbf{S}, \overline{\mathbf{f}})$ is known and remains constant.*

Assumption 2. *The problem parameters $(\boldsymbol{\ell}, \overline{\mathbf{p}}_r, \overline{\mathbf{p}}_e, \overline{\mathbf{f}})$ are such that the DC-OPF of (1) is feasible for all $\mathbf{x} \in \mathcal{X}$.*

Assumption 3. *Bidding and generation capacity parameters for rival units $(\mathbf{H}_r, \mathbf{c}_r, \overline{\mathbf{p}}_r)$ are known to the investor.*
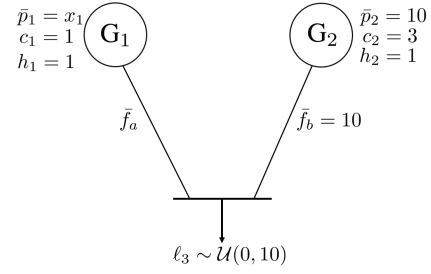

Fig. 1. A 3-bus power system showcasing the challenges in solving (4).

According to Assumption 2, the power system can be dispatched without the new units. Even under these assumptions, problem (4) is challenging due to three reasons: *i)* Constraint (4b) is expressed as an optimization problem itself; *ii)* The products between primal and dual variables inside the expectation in (3) are non-convex functions; and *iii)* Evaluating the expectation in $f(\mathbf{x})$ may be prohibitive. The investment task of (4) will be termed the *outer problem* and the DC-OPF of (1) given $\mathbf{x}$ as the *inner problem* of SI. We next elucidate the difficulty of SI through a small example for which the optimal SI decisions can be computed in closed form.

C. A 3-bus Example

Strategic generation investment can be challenging due to non-convexity and discontinuities. To elucidate on that, consider the 3-bus power system of Fig. 1. An investor considers building a generator on bus 1. A rival generator is located at bus 2 having capacity $\overline{p}_2 = 10$ pu. Bus 3 hosts a load ℓ_3 whose value is modeled as a random variable uniformly distributed within $[0, 10]$. The limit for line $b = (2,3)$ remains fixed at $\overline{f}_b = 10$, whereas the limit $\overline{f}_a$ for line $a = (1,3)$ is treated as a parameter and by varying its value, we will study its effect on SI. The investor does not own any existing generator, so that $f_e(\mathbf{p}_e) = g_e(\mathbf{p}_e) = 0$. Let us assume quadratic bidding functions $f_n(\mathbf{p}_n) = p_1^2 + p_1$ and $f_r(\mathbf{p}_r) = p_2^2 + 3p_2$; investment cost $k_1 = 1$; and $g_n(\mathbf{p}_n) = f_n(\mathbf{p}_n)$. We next compute the SI objective in (4a) by evaluating the expectation involved in (3).

We would like to compute the revenue for the investor

$$f(x_1) = k_1 x_1 + \mathbb{E}_{\ell_3}[g_n(p_1) - \pi_1 p_1]$$
$$= x_1 + \mathbb{E}_{\ell_3}[p_1^2 + p_1(1 - \pi_1)].$$

Nonetheless, the optimal primal/dual solutions (p_1, π_1) of (1) depend on the values of the problem parameters $(x_1, \overline{f}_a, \ell_3)$. As it will be explained later in more detail [cf. Figure 3], we can identify four solution cases for (p_1, π_1) depending on how x_1 and $\overline{f}_a$ compare. We present the detailed derivation for $f(x_1)$ for only one of these four cases for which $x_1 \leq \overline{f}_a \leq 1$. then the solution to inner problem can be expressed parametrically in (ℓ_3, x_1) as

$$(p_1, \pi_1) = \begin{cases} (\ell_3, 2\ell_3 + 1), & \ell_3 \leq x_1 \\ (x_1, 2\ell_3 - 2x_1 + 3), & \ell_3 > x_1 \end{cases}.$$

Since ℓ_3 is uniformly distributed, we can compute the investment cost under the case of $x_1 \leq \overline{f}_a \leq 1$ as

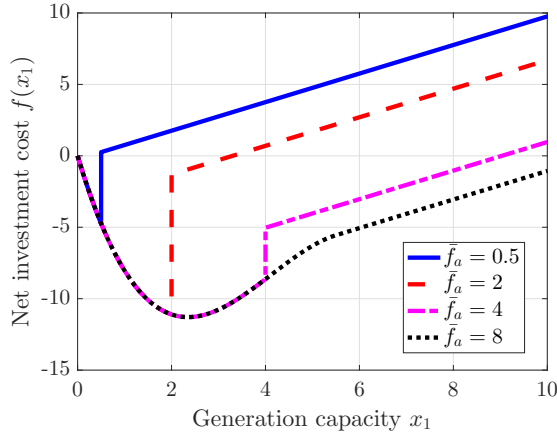$$f(x_1) = x_1 - \frac{1}{10} \int_0^{x_1} \ell_3^2 \, d\ell_3$$

Fig. 2. The net investment cost for the system of Figure 1 over the capacity range of $x_1 \in [0, 10]$ and for different capacities $\bar{f}_a$ of line $a = (1, 3)$.

$$+ \frac{1}{10} \int_{x_1}^{10} (3x_1^2 - 2x_1 - 2\ell_3 x_1) \, \mathrm{d}\ell_3$$

$$= -\frac{7}{30} x_1^3 + 3.2 x_1^2 - 11 x_1 \tag{5}$$

The optimal solutions and the value of $f(x_1)$ can be computed similarly for other combinations of the problem parameters $(x_1, \bar{f}_a, \ell_3)$. We next present the final outcome of this analysis, which involves three major cases depending on the value of $\bar{f}_a$ with 2-3 subcases each. In particular, if $\bar{f}_a \leq 1$

$$f(x_1) = \begin{cases} -\frac{7}{30} x_1^3 + 3.2 x_1^2 - 11 x_1, & x_1 \leq \bar{f}_a \\ x_1 - \frac{19 \bar{f}_a^2}{20} - \frac{1}{240}, & x_1 > \bar{f}_a \end{cases}.$$

If $1 < \bar{f}_a \leq 5.5$, the cost becomes

$$f(x_1) = \begin{cases} -\frac{7 x_1^3}{30} + 3.2 x_1^2 - 11 x_1, & x_1 \leq 1 \\ \frac{-8 x_1^3 + 99 x_1^2 - 333 x_1 + 1}{30}, & 1 < x_1 \leq \bar{f}_a \\ x_1 + \frac{4 \bar{f}_a^3 - 33 \bar{f}_a^2 + 1}{30}, & x_1 > \bar{f}_a \end{cases}$$

Finally, if $\bar{f}_a > 5.5$, the cost takes the form

$$f(x_1) = \begin{cases} -\frac{7 x_1^3}{30} + 3.2 x_1^2 - 11 x_1, & x_1 \leq 1 \\ \frac{-8 x_1^3 + 99 x_1^2 - 333 x_1 + 1}{30}, & 1 < x_1 \leq 5.5 \\ x_1 - \frac{1327}{120}, & x_1 > 5.5 \end{cases}.$$

Figure 2 plots the investment cost $f(x_1)$ for different values of $\bar{f}_a$. The key outcome of this analysis is that $f(x_1)$ is not only non-convex, but also that depending on the value, it can also be discontinuous. Specifically, our analysis has shown that $f(x_1)$ becomes discontinuous when $\bar{f}_a < 5.5$ with the discontinuity occurring at $x_1 = \bar{f}_a$. Thus, the SI problem may have a non-convex non-smooth objective function.

III. MPEC AND MPP

This sections extends results from mathematical programming with equilibrium constraints (MPEC) and multiparametric programming (MPP) to the considered SI setup.

A. Mathematical Programming with Equilibrium Constraints

MPEC is used in economics, where decisions taken by an investor affect the outcome of a market. MPEC provides a means for tackling bilevel optimization programs, such as the one in (4). Reference [2] posed (4) as an MPEC. For later reference and building on [2], we present an MPEC formulation for markets with quadratic bidding costs. Through this process, we provide the MPEC alternative to the novel SI solvers to be developed in Section IV. The MPEC model is built in three steps.

First, the bilinear term $\boldsymbol{\pi}^\top (\mathbf{p}_e + \mathbf{p}_n)$ in (3) is replaced by a quadratic function of the variables of the inner problem

$$\boldsymbol{\pi}^\top (\mathbf{p}_e + \mathbf{p}_n) = -\mathbf{p}_r^\top \mathbf{H}_r \mathbf{p}_r - \mathbf{c}_r^\top \mathbf{p}_r - (\overline{\boldsymbol{\mu}} + \underline{\boldsymbol{\mu}})^\top \overline{\mathbf{f}} - \overline{\boldsymbol{\gamma}}_r^\top \overline{\mathbf{p}}_r$$
$$- (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}})^\top \mathbf{S} \boldsymbol{\ell} - \lambda_0 \mathbf{1}^\top \boldsymbol{\ell}. \tag{6}$$

This follows from strong duality of the inner problem and after some algebraic manipulations; the proof is deferred to the appendix. For linear bidding functions, a similar result can be obtained by setting $\mathbf{H}_r = \mathbf{0}$ [2].

As a second step, the expectation in (3) is surrogated by a sample average over T instances indexed by t as [2]

$$\hat{f}(\mathbf{x}) = \mathbf{k}^\top \mathbf{x} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\pi}_t^\top (\mathbf{p}_{e,t} + \mathbf{p}_{n,t}) - g_e(\mathbf{p}_{e,t}) - g_n(\mathbf{p}_{n,t})$$

so the SI problem can be approximated as

$$\min_{\mathbf{x} \in \mathcal{X}} \hat{f}(\mathbf{x}) \tag{7}$$

$$\text{s.to } \{\boldsymbol{\pi}_t, \mathbf{p}_{e,t}, \mathbf{p}_{n,t}\} \text{ being solutions of (1) for } t = 1 : T.$$

The third step replaces constraint (4b) by the KKT conditions for the inner problem. Primal/dual feasibility and Lagrangian optimality yield a set of linear equality and inequality constraints on the primal/dual variables of the inner problem. Complementary slackness conditions entail non-convex products between primal and dual variables. These products can be formulated via the big-M method [2]. For example, the complementary slackness condition $x \cdot \lambda = 0$ for a primal constraint $x \geq 0$ and the related Lagrange multiplier $\lambda \geq 0$ can be written as

$$0 \leq x \leq \phi M \qquad 0 \leq \lambda \leq (1 - \phi)M \tag{8}$$

where M is a large constant and ϕ is an auxiliary binary variable. A set of constraints similar to (8) has to be enforced for each constraint of the inner problem and every market instance t. Following the aforesaid three steps, the bilevel problem in (4) can be reformulated as a mixed-integer linear or quadratic program (MILP/MIQP), depending on whether functions $\{f_r, f_e, f_n, g_e, g_n\}$ are linear or quadratic.

The MPEC method of [2] finds the global minimum of (4) upon approximating the ensemble average in (3) with the sample average over the T instances in (7). Nonetheless, the resultant mixed-integer model may not scale favorably for large T or large networks. Moreover, finding proper values for M is challenging, since it is hard to upper bound dual variables *a-priori*. To avoid computationally taxing mixed-integer models, we develop solvers of (4) leveraging neat results from MPP, which are outlined next.

B. Multiparametric Programming (MPP)

Multiparametric programming studies the solutions of an optimization problem as a function of its parameters [7]. The main idea of using MPP for SI is to handle the inner problem in (7) as a parametric problem. The inner problem in (7) involves solving T DC-OPFs, one for each instance. Each instance t could be dealing with different load demands, renewable generation conditions, available unit capacities, and generation costs. The inner problem also depends on the investment $\mathbf{x}$. For each instance [indexed by t in (7)] and a particular investment value $\mathbf{x}$, we arrive at a different *scenario* of the DC-OPF indexed by s. Problem (1) has to be solved over a scenario-dependent parameter vector $\boldsymbol{\theta}_s$; more details on the entries of $\boldsymbol{\theta}_s$ to be provided later in (15). Let us focus on the DC-OPF for a particular $\boldsymbol{\theta}_s$. To unclutter notation, we drop the subscript s. To cast (1) as a parametric problem, consider a minimization dependent on $\boldsymbol{\theta}$ as

$$\min_{\mathbf{p}} \quad \frac{1}{2}\mathbf{p}^\top \mathbf{H}\mathbf{p} + (\mathbf{C}\boldsymbol{\theta} + \mathbf{d})^\top \mathbf{p} \tag{9a}$$

$$\text{s.to} \quad \mathbf{A}\mathbf{p} \leq \mathbf{E}\boldsymbol{\theta} + \mathbf{b} \qquad\qquad : \boldsymbol{\lambda} \tag{9b}$$

$$\qquad\quad \mathbf{B}\mathbf{p} = \mathbf{F}\boldsymbol{\theta} + \mathbf{y}. \qquad\qquad : \boldsymbol{\mu}. \tag{9c}$$

If $\mathbf{H} = \mathbf{0}$, problem (9) is a multiparametric linear program (MPLP). If $\mathbf{H} \succeq \mathbf{0}$, problem (9) is a multiparametric convex quadratic program (MPQP). Different from existing results on MPLP/MPQPs [7], [11], problem (9) has $\boldsymbol{\theta}$ appearing not only in the constraints, but also in the cost. We next reproduce MPP analysis to accommodate this setup.

Let Θ be the set of all $\boldsymbol{\theta}$'s for which (9) is feasible. According to MPP theory [7], set Θ can be partitioned into distinct regions, termed *critical regions*, with three interesting properties: *p1)* Each region is described as a polytope in Θ; *p2)* Within each region, the same subset of inequality constraints become *active*, i.e., are satisfied with equality; and *p3)* Within each region, the primal/dual solutions of (9) can be expressed as affine functions of $\boldsymbol{\theta}$. These affine functions have been derived in [14]. They are reviewed next for completeness and to introduce the needed quantities.

Assume (9) is solved for $\boldsymbol{\theta}_o \in \Theta$ and let $(\mathbf{p}_o; \boldsymbol{\lambda}_o, \boldsymbol{\mu}_o)$ be the obtained optimal primal/dual solutions. Let also $\tilde{\mathbf{A}}$ be the submatrix obtained from $\mathbf{A}$ upon selecting the rows corresponding to the active constraints in (9b). The remaining rows of $\mathbf{A}$ related to *inactive constraints* (constraints satisfied with strict inequality) constitute matrix $\bar{\mathbf{A}}_o$. Similar partitions yield $(\tilde{\mathbf{E}}, \tilde{\mathbf{b}}, \tilde{\boldsymbol{\lambda}})$ and $(\bar{\mathbf{E}}, \bar{\mathbf{b}}, \bar{\boldsymbol{\lambda}})$. It is further assumed that matrix

$$\mathbf{K} := [\tilde{\mathbf{A}}^\top \ \ \mathbf{B}^\top]^\top \tag{10}$$

is full row-rank. This condition is known as *linear independence constraint qualification (LICQ)*. Although LICQ cannot be guaranteed before solving (9) for a $\boldsymbol{\theta}$, it occurs in the majority of our tests in Section V. We next consider separately the cases of $\mathbf{H} \succ \mathbf{0}$ and $\mathbf{H} = \mathbf{0}$ in (9).

For $\mathbf{H} \succ \mathbf{0}$ (strictly convex MPQP) and under LICQ, the primal/dual solutions of (9) can be obtained as [14]

$$\begin{bmatrix} \mathbf{p}_o \\ \tilde{\boldsymbol{\lambda}}_o \\ \boldsymbol{\mu}_o \end{bmatrix} = \mathbf{M}\boldsymbol{\theta}_o + \mathbf{r} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \mathbf{M}_3 \end{bmatrix}\boldsymbol{\theta}_o + \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} \tag{11}$$

where

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \mathbf{M}_3 \end{bmatrix} := \begin{bmatrix} \mathbf{H} & \tilde{\mathbf{A}}^\top & \mathbf{B}^\top \\ \tilde{\mathbf{A}} & \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{C} \\ \tilde{\mathbf{E}} \\ \mathbf{F} \end{bmatrix} \tag{12a}$$

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} := \begin{bmatrix} \mathbf{H} & \tilde{\mathbf{A}}^\top & \mathbf{B}^\top \\ \tilde{\mathbf{A}} & \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{b}} \\ \mathbf{y} \end{bmatrix}. \tag{12b}$$

The matrix inverse in (12) exists, since its determinant equals $\det(\mathbf{H})\det(-\mathbf{K}\mathbf{H}^{-1}\mathbf{K}^\top) < 0$ from Schur's complement.

For $\mathbf{H} = \mathbf{0}$ (for which (9) is an MPLP), suppose further that $\mathbf{K}$ is square. This holds if in addition to LICQ, the number of active constraints equals the number of optimization variables. Then, the primal/dual solutions of (9) take again the closed-form expression of (11), but with [14]

$$\mathbf{M}_1 := \mathbf{K}^{-1}\begin{bmatrix} \tilde{\mathbf{E}} \\ \mathbf{F} \end{bmatrix} \quad \text{and} \quad \mathbf{r}_1 := \mathbf{K}^{-1}\begin{bmatrix} \tilde{\mathbf{b}} \\ \mathbf{y} \end{bmatrix} \tag{13a}$$

$$\begin{bmatrix} \mathbf{M}_2 \\ \mathbf{M}_3 \end{bmatrix} := \mathbf{K}^{-\top}\mathbf{C} \quad \text{and} \quad \begin{bmatrix} \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} := \mathbf{K}^{-\top}\mathbf{d}. \tag{13b}$$

One of the interesting claims of MPLP/MPQPs is that for any other $\boldsymbol{\theta} \in \Theta$ yielding rise to the same set of active constraints, the primal/dual solutions are expressed through (11); see e.g., [7], [8]. Contrarily, given a set of constraints, the subset of $\boldsymbol{\theta}$'s activating those constraints can be identified as a polytope $\mathcal{C} \subseteq \Theta$ described as (see [8] for details)

$$\mathcal{C} := \left\{ \boldsymbol{\theta} \in \Theta |\ (\bar{\mathbf{A}}\mathbf{M}_1 - \bar{\mathbf{E}})\,\boldsymbol{\theta} \leq \bar{\mathbf{b}} - \bar{\mathbf{A}}\mathbf{r}_1, \mathbf{M}_2\boldsymbol{\theta} \geq \mathbf{r}_2 \right\}. \tag{14}$$

The quantities $(\mathbf{M}_1, \mathbf{M}_2, \mathbf{r}_1, \mathbf{r}_2)$ are provided by (12) or (13) for MPQP and MPLP, accordingly. The set $\mathcal{C}$ is termed a *critical region* of Θ. We next leverage these MPP properties to cope with (4) in two different ways.

IV. STRATEGIC INVESTMENT VIA MPP

To derive efficient SI solvers, the key idea is to cast the inner problem as an MPP and exploit the rich properties for its solutions. If the bidding functions in (1) are quadratic or (piecewise) affine, then (1) is an instance of (9). The optimization variable $\mathbf{p}$ stacks the variables $(\mathbf{p}_r, \mathbf{p}_e, \mathbf{p}_n)$. The parametric inequalities of (9b) capture the line flow constraints of (1c) and the generation limits of (1d)–(1e). The parametric equalities (9c) relate to the power balance constraint of (1b).

The parameter $\boldsymbol{\theta}$ appearing in (9) consists of three parts. The first part relates to varying generation cost coefficients (bids). Under the assumption that the quadratic component $\frac{1}{2}\mathbf{p}^\top \mathbf{H}\mathbf{p}$ remains invariant across scenarios, these changing costs are modeled by $\mathbf{C}\boldsymbol{\theta} + \mathbf{d}$ in (9a) with $\mathbf{C} = [\mathbf{I} \ \mathbf{0} \ \mathbf{0}]$ and $\mathbf{d} = \mathbf{0}$. The second part of $\boldsymbol{\theta}$ captures the uncertain demand vector $\boldsymbol{\ell}$. The third part captures varying capacities of generation units that are due to scheduled outages or due to variable renewable resources. Moreover, the capacity for new units will be changing while solving SI as solvers will be evaluating (1) for different values of $\mathbf{x}$ seeking the optimal investment. In summary, each parameter vector $\boldsymbol{\theta}_s$ ($\boldsymbol{\theta}$ for simplicity) can be expressed as

$$\boldsymbol{\theta} := [\mathbf{c}^\top \ \boldsymbol{\ell}^\top \ \bar{\mathbf{p}}^\top]^\top \tag{15}$$

where $\mathbf{c} := [\mathbf{c}_r^\top \ \mathbf{c}_e^\top \ \mathbf{c}_r^\top]^\top$; $\bar{\mathbf{p}} := [\bar{\mathbf{p}}_r^\top \ \bar{\mathbf{p}}_e^\top \ (\boldsymbol{\alpha} \odot \mathbf{x})^\top]^\top$; and $\boldsymbol{\alpha} := [\boldsymbol{\alpha}_r^\top \ \boldsymbol{\alpha}_e^\top \ \boldsymbol{\alpha}_r^\top]^\top$. Apart from $\mathbf{x}$, all quantities appearing in (15) should be noted with the instance index t as subscript; we avoided that to simplify notation. Collect all uncertain components of $\boldsymbol{\theta}$ in $\boldsymbol{\omega} := \{\mathbf{c}_r, \mathbf{c}_e, \mathbf{c}_r, \boldsymbol{\ell}, \bar{\mathbf{p}}_r, \bar{\mathbf{p}}_e, \boldsymbol{\alpha}\}$. Evidently from (15), the parameter vector of (9) can be expressed as a mapping $\boldsymbol{\theta} = \mathcal{P}(\boldsymbol{\omega}, \mathbf{x})$ of the uncertain variables $\boldsymbol{\omega}$ and the optimization variable of the outer problem $\mathbf{x}$. Heed that the mapping $\mathcal{P}(\boldsymbol{\omega}, \mathbf{x})$ is not linear in $(\boldsymbol{\omega}, \mathbf{x})$ due to the products $\boldsymbol{\alpha} \odot \mathbf{x}$ in $\bar{\mathbf{p}}$. Nonetheless, the objective and constraint functions of the parametric QP in (9) depend *linearly* on $\boldsymbol{\theta} = \mathcal{P}(\boldsymbol{\omega}, \mathbf{x})$. The matrices $(\mathbf{A}, \mathbf{E}, \mathbf{B}, \mathbf{F})$ and vectors $(\mathbf{b}, \mathbf{y})$ in (9) are straightforward to compute and are not presented here. Having posed (1) as an instance of (9), we next present two methods that leverage the affine mappings of (11)–(13) and the partitioning of (14) to solve (7).

A. An MPP-aided Grid Search (MPP-GS) Scheme

This section exploits the MPP results of Section III-B to solve (1) for a large number of $(\boldsymbol{\omega}, \mathbf{x})$ instances. We can thus evaluate $\hat{f}(\mathbf{x})$ over a grid of $\mathbf{x}$ values efficiently. This grid search approach is preferred when an investor is presented with a single or few possible investment locations. To design our search grid, note that the investment x_m at bus m can be bounded as

$$\overline{x}_m = \sum_{k \sim m} \overline{f}_{m,k} + \max_t\{\ell_{m,t}\} \qquad (16)$$

by the maximum load at bus m plus the sum of capacities for all transmission lines incident to bus m. Symbol $\overline{f}_{m,k}$ denotes the capacity of the line connecting buses m and k, if such line exists. The quantity $\overline{x}_m$ is the maximum power that can be produced at bus m without violating any physical limits. The discretization step over $[0, \overline{x}_m]$ can be chosen based on the type of the power plant. For example, a typical wind turbine is about 2-3 MW, so that multiples of this value are reasonable options for the grid step. When investing at M locations with K_m search values per location m, we get a search grid $\hat{\mathcal{X}} \subseteq \mathcal{X}$ of $K = \prod_{m=1}^M K_m$ points. Since K grows exponentially with M, this approach makes sense only for $M = 1 - 3$ locations.

Given the search grid $\hat{\mathcal{X}}$ and the uncertain parameter set $\Omega := \{\boldsymbol{\omega}_t\}_{t=1}^T$, one can readily form the parameter set $\hat{\Theta}$ using the mapping $\mathcal{P} : \hat{\mathcal{X}} \times \Omega \to \hat{\Theta}$ and $|\hat{\Theta}| = KT$. Here $\hat{\Theta}$ is a finite subset of Θ, over which (9) has to be solved. This slightly abuses notation since in Section III-B symbol Θ denoted the convex set of $\boldsymbol{\theta}$'s rendering (9) feasible. A solution to (7) can be found by solving (1) in its parameterized form of (9) for all KT members of $\hat{\Theta}$, and then evaluating $\hat{f}(\mathbf{x})$ over $\hat{\mathcal{X}}$. For $\hat{f}(\mathbf{x})$ to be a reasonable estimate of $f(\mathbf{x})$ though, a large number T of instances $\boldsymbol{\omega}_t$ needs to be considered, yielding a computationally formidable task even for small K.

Thanks to MPP however, problem (9) needs to be solved for just as many times as the critical regions appearing in $\hat{\Theta}$. To see this, suppose that for a critical region $\mathcal{C}_o \subseteq \Theta$, we have already computed its polytopic description in (14) and the pair $(\mathbf{M}, \mathbf{r})$ parameterizing its primal/dual solutions. Then, for any other $\boldsymbol{\theta}_s \in \hat{\Theta}$ belonging to $\mathcal{C}_o$, we can directly compute its primal/dual solutions from (11) without having to solve (9).

Algorithm 1 MPP-aided Grid Search (MPP-GS)

Input: Set of OPF scenarios $\hat{\Theta} = \{\boldsymbol{\theta}_s\}_{s=1}^{KT}$
Output: OPF solutions $\{\boldsymbol{\pi}_s, \mathbf{p}_{e,s}, \mathbf{p}_{n,s}\}_{s=1}^{KT}$ to (1) via (9) for all $\boldsymbol{\theta}_s \in \hat{\Theta}$
1: **while** $\hat{\Theta} \neq \emptyset$ **do**
2:      Randomly select $\boldsymbol{\theta}_o \in \hat{\Theta}$ and $\hat{\Theta} \leftarrow \hat{\Theta} \setminus \boldsymbol{\theta}_o$
3:      Solve (9) for $\boldsymbol{\theta}_o$ to find its primal/dual solutions and active constraints
4:      Record $(\mathbf{p}_{e,o}, \mathbf{p}_{n,o}, \boldsymbol{\pi}_o)$
5:      **if** matrix $\mathbf{K}$ of (10) is full row-rank, **then**
6:          Compute region's parameters $(\mathbf{M}, \mathbf{r})$ from (11)
7:          Compute region's polytope $\mathcal{C}$ from (14)
8:          **for** all $\boldsymbol{\theta}_s \in \hat{\Theta}$ **do**
9:              **if** $\boldsymbol{\theta}_s \in \mathcal{C}$ [satisfying (14)], **then**
10:             Compute OPF solution as $\mathbf{p}_s = \mathbf{M}_1\boldsymbol{\theta}_s + \mathbf{r}_1$, $\boldsymbol{\lambda}_s = \mathbf{M}_2\boldsymbol{\theta}_s + \mathbf{r}_2$, and $\boldsymbol{\mu}_s = \mathbf{M}_3\boldsymbol{\theta}_s + \mathbf{r}_3$ and compute $\boldsymbol{\pi}_s$ using (2)
11:             Record $(\mathbf{p}_{e,s}, \mathbf{p}_{n,s}, \boldsymbol{\pi}_s)$
12:             $\hat{\Theta} \leftarrow \hat{\Theta} \setminus \boldsymbol{\theta}_s$
13:          **end if**
14:          **end for**
15:      **end if**
16: **end while**
17: Evaluate $\hat{f}(\mathbf{x})$ over $\hat{\mathcal{X}}$ and find the minimizing $\mathbf{x}$

This procedure, termed *MPP-based Grid Search (MPP-GS)*, is formalized as Algorithm 1 and its steps are explained next.

MPP-GS selects a $\boldsymbol{\theta}_o$ from $\hat{\Theta}$ at step 2. At step 3, it solves (9) for $\boldsymbol{\theta}_o$. If the related $\mathbf{K}$ is of full row rank, the algorithm constructs a description for the visited critical region (steps 6-7). It further scans the remaining dataset $\hat{\Theta}$ to find other $\boldsymbol{\theta}_s$'s belonging to this region (step 8); computes their solution in closed form (steps 10-11); and removes these $\boldsymbol{\theta}_s$'s from $\hat{\Theta}$ (step 12). The process continues until $\hat{\Theta}$ becomes empty.

MPP-GS explores a critical region only when $\mathbf{K}$ is of full row rank (step 5). Albeit such cases could be handled [7], [8], they involve methods of high complexity. Instead, when we come across such an instance of (9), we only record its primal/dual solutions. During the tests of Section V, these instances appear infrequently. Vectors $\boldsymbol{\theta}_s$ are visited in an arbitrary rather than sequential fashion, by randomly sampling from $\hat{\Theta}$ (step 2). In this way, we increase the chances of exploring more popular critical regions early on. It is hence more likely to handle a larger number of $\boldsymbol{\theta}_s$'s earlier, so that $\hat{\Theta}$ shrinks faster and step 9 is run on progressively much fewer $\boldsymbol{\theta}_s$'s. To cope with (7) for larger K or T, we next pursue an MPP-aided stochastic gradient descent approach.

B. MPP-aided Stochastic Gradient Descent (MPP-SGD)

The objective $\hat{f}(\mathbf{x})$ of (7) involves a summation over a large number T of instances $\boldsymbol{\omega}_t$. Rather than finding the costly gradient of $\hat{f}(\mathbf{x})$, we adopt stochastic approximation and update $\mathbf{x}$ by taking each time a descent step over the gradient for only one of the summands of $\hat{f}(\mathbf{x})$. Define the summand of $\hat{f}(\mathbf{x})$ related to instance $\boldsymbol{\omega}_t$ as

$$f_t(\mathbf{x}) := \mathbf{k}^\top \mathbf{x} - \boldsymbol{\pi}_t^\top (\mathbf{p}_{e,t} + \mathbf{p}_{n,t}) + g_e(\mathbf{p}_{e,t}) + g_n(\mathbf{p}_{n,t}). \quad (17)$$
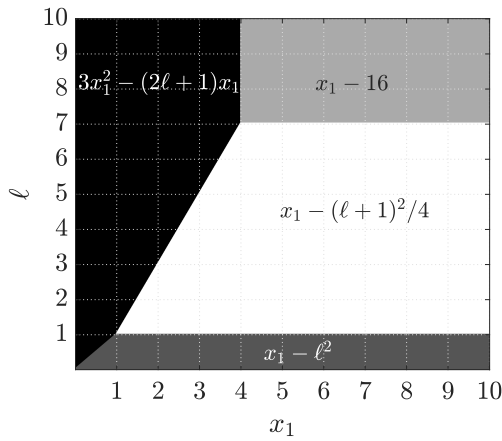
Fig. 3. The critical regions of the inner problem in Figure 2 for $\bar{f}_a = 4$. The dark gray region corresponds to p_2 at its lower limit ($p_2 = 0$); the black region to p_1 at its upper limit $p_1 = x_1$; the light gray region to line $(1,3)$ being congested; and the white region to no active constraints.

Recall that the dispatches $(\mathbf{p}_{e,t}, \mathbf{p}_{n,t})$ and prices $\boldsymbol{\pi}_t$ are all functions of $\mathbf{x}$, since they are outcomes of (1) given $\mathbf{x}$.

Apparently $\nabla_{\mathbf{x}}(\mathbf{k}^\top \mathbf{x}) = \mathbf{k}$. To study the differentiability of the remaining terms of f_t, assume for now that $\boldsymbol{\theta} = \mathcal{P}(\boldsymbol{\omega}_t, \mathbf{x})$ is strictly inside a critical region $\mathcal{C}_o \subseteq \Theta$. According to (11), the optimal dispatch vectors $\mathbf{p}_{e,t}$ and $\mathbf{p}_{n,t}$ are affine in $\boldsymbol{\theta}$ and hence, affine in $\mathbf{x}$ for a particular $\boldsymbol{\omega}_t$. to (2), optimal prices $\boldsymbol{\pi}_t$ are affine in $(\boldsymbol{\lambda}, \boldsymbol{\mu})$. Since $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ are affine functions of $\boldsymbol{\theta}$ from (11), the prices $\boldsymbol{\pi}_t$ are affine in $\boldsymbol{\theta}$ as well. Consequently, the revenue term $\boldsymbol{\pi}_t^\top (\mathbf{p}_{e,t} + \mathbf{p}_{n,t})$ is quadratic in $\mathbf{x}$ and its gradient takes the form

$$\nabla_{\mathbf{x}} \left[\boldsymbol{\pi}_t^\top (\mathbf{p}_{e,t} + \mathbf{p}_{n,t}) \right] = \mathbf{Q}_o \boldsymbol{\theta} + \mathbf{q}_o. \qquad (18)$$

The parameters $(\mathbf{Q}_o, \mathbf{q}_o)$ can be computed using (11). Heed these parameters remain *constant* within each critical region of Θ, that is for all pairs $(\boldsymbol{\omega}_t, \mathbf{x})$ for which $\boldsymbol{\theta} = \mathcal{P}(\boldsymbol{\omega}_t, \mathbf{x}) \in \mathcal{C}_o$. As in Section IV-A, the uncertain parameters $\boldsymbol{\omega}_t$ are drawn from a finite set of instances. On the contrary, the investment variable $\mathbf{x}$ is drawn now from a continuous set.

Regarding the term $g_e(\mathbf{p}_{e,t}) + g_n(\mathbf{p}_{n,t})$, its gradient with respect to $\mathbf{x}$ can be computed using the chain rule, since functions (g_e, g_n) are known (quadratic or affine) and $(\mathbf{p}_{e,t}, \mathbf{p}_{n,t})$ are affine functions of $\boldsymbol{\theta}$ and consequently $\mathbf{x}$. Consider now the case where $\boldsymbol{\theta} = \mathcal{P}(\boldsymbol{\omega}_t, \mathbf{x})$ is on a boundary between critical regions. Then, functions $\boldsymbol{\pi}_t^\top (\mathbf{p}_{e,t} + \mathbf{p}_{n,t})$, $g_e(\mathbf{p}_{e,t})$, and $g_n(\mathbf{p}_{n,t})$ may not be differentiable or even continuous with respect to $\mathbf{x}$. Take for example the 3-bus example of Figure 2: Figure 3 shows its critical regions over $\boldsymbol{\theta}$. It also displays the functional form of $f_t(x_1)$ per region. Evidently, function $f_t(x_1)$ is differentiable within each region, but not on their boundaries. Nonetheless, these boundaries are zero-probability events over Θ. Being a stochastic algorithm, the probability of coming across such $\boldsymbol{\theta}$'s during the SGD iterations is zero.

Instead of updating $\mathbf{x}$ for one $\boldsymbol{\omega}_t$ at a time, we exploit results from MPP and derive a mini-batch rendition to get improved algorithmic convergence at a minimal increase in computational complexity. The novel idea is to exploit MPP regions and efficiently compute gradients with respect to $\mathbf{x}$ not

---

Algorithm 2 MPP-Stochastic Gradient Descent (MPP-SGD)

Input: Ω, initialization $\mathbf{x}_o$, tolerance τ, and step size η
Output: Optimal investment $\mathbf{x}^*$

1: Set $\mathbf{x}^0 = \mathbf{x}_o$, $\epsilon > \tau$, $k = 0$
2: **while** $\epsilon \geq \tau$ **do**
3:     Define $\Theta^k \leftarrow \{\boldsymbol{\theta}_1^k, \ldots, \boldsymbol{\theta}_T^k\}$ where $\boldsymbol{\theta}_t^k = \mathcal{P}(\boldsymbol{\omega}_t, \mathbf{x}^k)$
4:     Randomly select $\boldsymbol{\theta}_o$ from Θ^k
5:     Solve (9) for $\boldsymbol{\theta}_o$ to find its primal/dual solutions and active constraints
6:     Set $\mathbf{g}^k \leftarrow \mathbf{0}$ and $c^k \leftarrow 0$
7:     **if** matrix $\mathbf{K}$ is full row-rank, **then**
8:         Compute region's parameters $(\mathbf{M}, \mathbf{r})$ from (11) and gradient coefficients $(\mathbf{Q}_o, \mathbf{q}_o)$
9:         Compute region's polytope $\mathcal{C}$ from (14)
10:        **for** all $\boldsymbol{\theta}_t^k \in \Theta^k$ **do**
11:           **if** $\boldsymbol{\theta}_t^k \in \mathcal{C}$, **then**
12:             compute the gradient $\mathbf{g}_t$ and set $\mathbf{g}^k \leftarrow \mathbf{g}^k + \mathbf{g}_t$
13:             set $c^k \leftarrow c^k + 1$
14:           **end if**
15:        **end for**
16:     **end if**
17:     Set $\mathbf{x}^{k+1} \leftarrow \left[\mathbf{x}^k - \frac{\eta}{c^k \sqrt{k}} \mathbf{g}^k \right]_{\mathcal{X}}$
18:     Compute the moving average $\bar{\mathbf{x}}^k \leftarrow \frac{\sum_{i=\lceil k/2 \rceil}^{k} \left(\mathbf{x}^{(i)} / \sqrt{i} \right)}{\sum_{i=\lceil k/2 \rceil}^{k} \sqrt{i}}$
19:     Set $\epsilon \leftarrow \frac{\|\bar{\mathbf{x}}^k - \bar{\mathbf{x}}^{(k-1)}\|}{\|\bar{\mathbf{x}}^k\|}$ and $k \leftarrow k + 1$
20: **end while**
21: Set $\mathbf{x}^* = \bar{\mathbf{x}}^k$

---

only for a single, but multiple $\boldsymbol{\omega}_t$'s at a time. To elaborate, notice that for a particular $\mathbf{x}_o$, all $\boldsymbol{\theta}_t = \mathcal{P}(\boldsymbol{\omega}_t, \mathbf{x}_o)$ that belong to the same critical region share the same gradient coefficients $(\mathbf{Q}_o, \mathbf{q}_o)$ in (18). Hence, all these gradients can be readily computed once this critical region, its parameters $(\mathbf{M}, \mathbf{r})$, and its polytopic description of (14) have been identified.

Our MPP-aided stochastic gradient descent algorithm is tabulated as Algorithm 2. Step 3 constructs a parameter set Θ^k based on the current estimate of the investment vector $\mathbf{x}^k$ and all instances $\boldsymbol{\omega}_t$'s. In steps 4-8, a random $\boldsymbol{\theta}_o$ is drawn from Θ^k and we identify the region it belongs to. Steps 9-14 compute the gradient with respect to $\mathbf{x}$ for all $\boldsymbol{\theta} \in \Theta^k$ and sum them up in $\mathbf{g}^k$. Step 12 counts the members of the said region, so that the average gradient can be computed in step 16. The updates of step 17 are guaranteed to converge to a stationary point [20]. The random draw of step 4 ensures an unbiased exploration of regions in Θ, hence the average gradient per region is an unbiased estimate of the gradient of $\hat{f}(\mathbf{x})$ in (7).

V. NUMERICAL TESTS

Algorithms 1 and 2 were contrasted against the MPEC method of [2] on three systems: the 3-bus system of Figure 2; the IEEE 30-bus system; and the IEEE 118-bus system with line limits estimated from surge impedances per [21]. The investing bus were chosen as $\{1\}$, $\{3, 20\}$, and $\{29, 95\}$ for the three networks, respectively. For the 30- and 118-bus systems, we chose the investor to already own the generator at bus 1. We used hourly bidding and load data from the day-ahead PJM

TABLE I
RESULTS FOR ALGORITHM 1

system	$T \times K$	# critical regions	optimal cost	time [s]
3-bus	876,000	4	−11.31	67
30-bus	876,000	229	−422	203
118-bus	876,000	23,695	−5,290	62,720

TABLE II
OPTIMAL COST ATTAINED BY ALGORITHM 2 [$/H]

system	initialization				
	1	2	3	4	5
3-bus	−11.28	−11.28	−11.28	−11.28	−11.28
30-bus	−422	−419	−417	−421	−420
118-bus	−5,280	−5,274	−5,296	−5,256	−5,275

TABLE III
RUNNING TIME OF ALGORITHM 2 [s]

system	initialization				
	1	2	3	4	5
3-bus	363	323	260	362	483
30-bus	2,246	4,240	4,266	2,000	2,178
118-bus	14,100	9,747	10,743	9,495	9,507

TABLE IV
STATISTICS OF ALGORITHM 2 FOR 30-BUS SYSTEM

	μ	σ
investment cost [$/H]	−419.876	10.839
computation time [s]	2,231	275.7
optimal solution [100 MW]	[0.5686, 0.2095]	[0.0120, 0.0144]

market for 2018 [22]. Since load profiles correspond to areas and there are only 21 of them, profiles were randomly assigned to buses. Each bus load profile was perturbed by adding a uniformly distributed deviation of $\pm 5\%$ independently over time and buses. Load profiles were finally scaled so their annual peak matched the benchmark load. For thermal units, we assumed $f_e(\mathbf{p}_{e,t}) = g_e(\mathbf{p}_{e,t})$, whereas for wind ones we set $f_e(\mathbf{p}_{e,t}) = g_e(\mathbf{p}_{e,t}) = 0$. For wind units, we assumed a cost of $3 \cdot 10^6$ \$/MW for purchase and installation, plus $1 \cdot 10^6$ \$/MW for operation and maintenance over 25 years. Converting that cost to dollars per hour per unit of active power for a base of 100 MVA yielded $k = 1,826.5$ in (3). All tests were performed on an Intel Core i7 @ 3.4 GHz (16 GB RAM) computer, using MATLAB on a single CPU without any parallelization. Problem (9) was solved using the ECOS solver in YALMIP [23], [24]. All times reported are wall-clock times.

The first test explores the effect of the network size and the number of instances T on the scalability and optimality of MPEC. MPEC was implemented in Gurobi v.9.0 [25]. To prevent the MPEC method from stalling for small improvements, the solver's optimality gap was set to 5%. We observed that for smaller optimality gaps the solver took much longer computational times. We also set $M = 10^4$ in (8). Because MPEC could not consider all $8,760$ instances, we ran MPEC for $T \in \{10, 25, 50\}$ instances per network. Figure 4 shows the box plots of $\hat{f}(\mathbf{x})$ achieved by the MPEC solutions, while Figure 5 shows the box plots for the related running times. For the 3- and 30-bus systems, the MPEC solutions seem to be achieving the optimal cost using only 25-50 instances and within reasonable time. For the 118-bus system however, the investment cost varies widely even for 50 instances. This observation suggests that finding a meaningful optimizer requires considering $T > 50$ instances. However, Figure 5 indicates that the running time grows fast with T. It is also worth reporting that running MPEC for the 118-bus system and $T = 100$ took approximately four days.

The second test evaluates the optimality and scalability of our Alg. 1 for the MPP-GS method. For the 3-bus system, we considered a 1-D grid of 100 uniformly-spaced values

for $x_1 \in [0, 10]$ and $T = 8,760$ instances drawn from $\ell \in \mathcal{U}(0, 10)$. For the 30-bus and 118-bus systems, we considered 10 uniformly-spaced values for each investment in the range of $[0, 1]$ and $[0, 10]$ for each location, respectively, resulting in a 2-D grid with 100 points. With $T = 8,760$ instances, this gave a total of $876,000$ DC-OPFs to be solved as reported in Table I. This table also shows the number of critical regions identified; the optimal cost found; and the running times. Figure 7 shows the investment cost achieved by Alg. 1. The MPP-GS was successful in finding an investment cost lower than that of MPEC indicating the advantage of considering the complete scenario set. For the 3- and 30-bus systems, there is also significant computational advantage. The relatively longer time needed for the 118-bus system can be attributed to the large number of critical regions identified that are due to the wider range of $\mathbf{x}$ and larger system.

To analyze the running times of Alg. 1 further, we classified the running time into 3 main components: the YALMIP building time; the solver's time for dealing with each DC-OPF; and the time needed for Steps 5-13 of Alg. 1 that mainly consist of evaluating region membership through equation (14). We also evaluated these times on the 118-bus system for the same number of total scenarios ($KT = 876,000$), but for different grid volumes of $\mathbf{x}$. If $\mathbf{x} \in [\underline{x}_1, \overline{x}_1] \times [\underline{x}_2, \overline{x}_2] \cdots \times [\underline{x}_M, \overline{x}_M]$, then we define the grid volume as $\prod_{m=1}^{M} (\overline{x}_m - \underline{x}_m)$. Table V shows these running times as well as the number of critical regions identified. Since Steps 5-15 of the MPP-GS algorithm seem to be the bottleneck, the computation time can be further reduced by using a parallel implementations of Alg. 1 especially for evaluating (14). For example, just by replacing the for-loop of Step 10 with matrix operations, the checking time can be reduced from $12,217$ [s] to $4,292$ [s] for grid volume of 1. Table V also shows that the larger the volume of the $\mathbf{x}$-grid, the larger the volume of Θ and so more critical regions are identified.

The third test evaluated Alg. 2 and our MPP-SGD method. Figure 8 shows the convergence of the investment decisions for 5 randomly initialized trials. Tables II and III show the optimal cost and running times, respectively. We observe that for all trials, the optimal cost is much lower than that attained
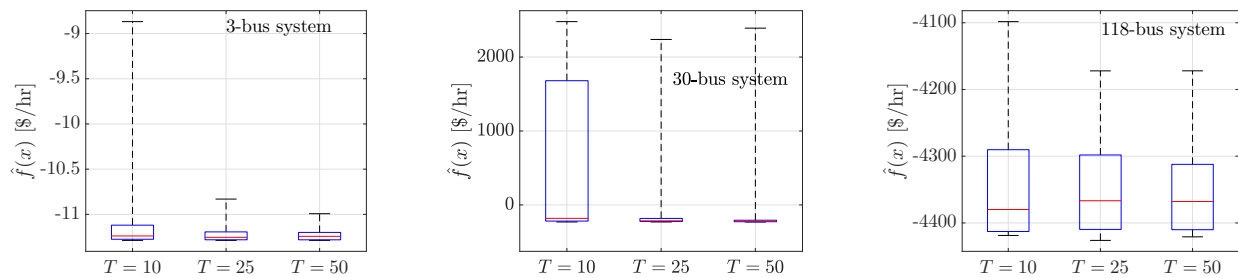
Fig. 4. Optimal investment cost attained by the MPEC method of [2] for the three systems. Although the investment decision $\mathbf{x}$ has been computed using only $T \in \{10, 25, 50\}$ instances due to computational limitations, the cost shown here is computed over all $8,760$ market instances. Box plots are computed over 100 Monte Carlo runs by randomly selecting T instances each time. *Left:* For the 3-bus system, the MPEC solutions attained the true optimal cost of -11.28 using 25-50 instances in 1-5 seconds – the true optimal cost can be found as the stationary point of $f(x_1)$ in Fig. 2 for $\bar{f}_a$. For the two other systems, the true optimal cost is not known. *Center:* For the 30-bus system, the MPEC solution reaches a seemingly minimal cost using 25 instances, even though the maximum value over the Monte Carlo runs varies widely. *Right:* For the 118-bus system, the sample mean costs lie significantly above the sample minimum ones, and using 50 instances is not sufficient to reduce the cost variability.
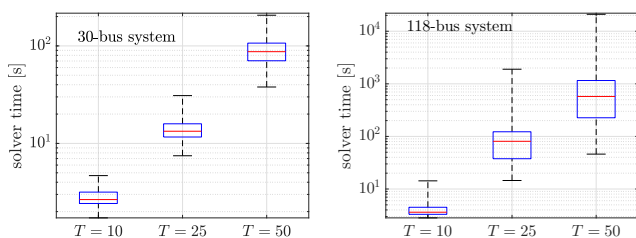

Fig. 5. Box plots of running times for the MPEC method of [2] over 100 Monte Carlo runs by randomly selecting T instances each time.
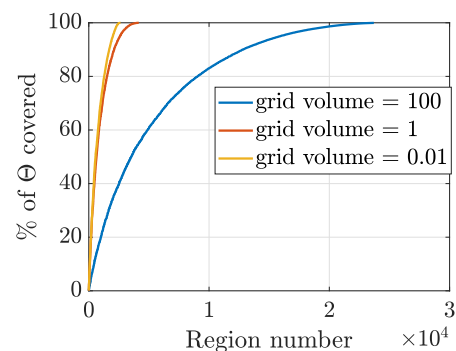

Fig. 6. Cardinalities of critical regions for the test cases of Table V. To show the distribution of cardinalities across critical regions, regions were sorted by decreasing cardinality. We then computed the running of those cardinalities. The vertical axis here shows this running sum as a percentage across all critical regions. For the grid volumes of 0.01 and 1, there were a total of roughly 5,000 critical regions, whereas for the grid volume of 100, there were 24,000 regions.

TABLE V
RUNNING TIMES OF ALGORITHM 1 FOR THE 118-BUS SYSTEM

grid volume	# critical regions	YALMIP [s]	solver [s]	check [s]
0.01	2,636	259	440	10,329
1	4,153	376	640	12,217
100	23,695	2,184	3,681	56,855

by MPEC for $T = 50$ instances, which was -230 for the 30-bus system and $-4,2208$ for the 118-bus one. Compared to Algorithm 1, Algorithm 2 achieves similar optimal costs. The running times of Alg. 2 are longer for the 3- and 30-bus systems, but much lower for the 118-bus system. Even though the MPP-SGD iterates do not converge to the same decisions for all trials, they attain relatively similar investment costs. This agrees with the findings of Fig. 7, where the cost function seems to be relatively flat at the optimum.

To study the variability in the execution time of Algorithm 2, we conducted 100 random initializations of MPP-SGD algorithm for the 30 bus system. Table IV shows the mean and standard deviation of the optimal cost, computation time, and optimal solution of the MPP-SGD algorithm across the 100 tests performed. These results show that MPP-SGD exhibits relatively small standard deviation in all three quantities.

To recapitulate, MPEC can find a globally optimal solution within reasonable time for smaller systems. For larger systems, the complexity involved confines MPEC to relatively few instances T, which may result in subpar investment solutions if those solutions are to be evaluated on realistic scenario sets. Algorithm 1 is able to achieve much lower average costs over

a year-long complete dataset, yet is limited by the number of investment locations and the granularity of the search grid. The latter issues are alleviated by Algorithm 2, which seems to be finding near-optimal decisions by handling the complete dataset in running times that improved relatively with the network size.

VI. CONCLUSIONS

This work has exploited MPP to devise two SI solvers. The grid search algorithm can handle cases where the number of investment locations is small. Although the needed function evaluations constitute an enormous dataset of DC-OPF instances, their exact primal/dual solutions can be computed upon solving only a limited number of these OPFs, thus accelerating the search by 8-12 times. For larger numbers of investment locations, we have devised a stochastic gradient search scheme, which computes the gradient of the SI objective over entire critical regions in an extremely efficient manner. The developed tools facilitate faster and more educated energy market decisions, while the ideas put forth can be proved fruitful for coping more efficiently with transmission expansion planning and contingency analysis.
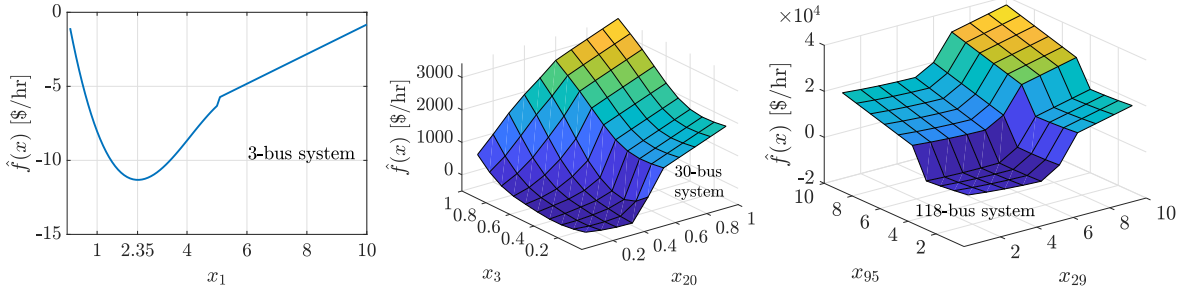
Fig. 7. The value of the investment cost $\hat{f}(\mathbf{x})$ for the 3-bus *(left)*; 30-bus *(middle)*; and 118-bus systems *(right)*.
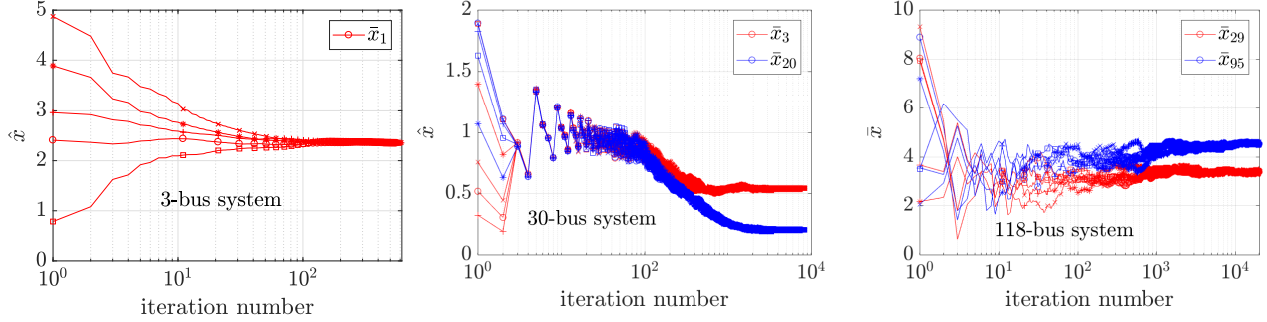

Fig. 8. Convergence of Alg. 2 for 5 random initializations for the 3-bus system *(left)*; the 30-bus system *(center)*; and the 118-bus system (right). For the 3-bus system, the minimizer is unique and the MPP-SGD algorithm converges to the same point for all initializations.

APPENDIX A
MPEC WITH QUADRATIC COST

This adopts the results from [2] to replace the inner quadratic program of (1) by its KKT conditions and the bilinear terms of (3) with linear and quadratic ones. It is presented here since [2] considered the case where (9) is a linear rather than a quadratic program.

Lagrangian optimality conditions imply that

$$\mathbf{H}_r \mathbf{p}_r + \mathbf{c}_r + \lambda_0 \mathbf{1} + \mathbf{S}^\top (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}}) + (\overline{\boldsymbol{\gamma}}_r - \underline{\boldsymbol{\gamma}}_r) = \mathbf{0} \quad (19a)$$

$$\mathbf{H}_e \mathbf{p}_e + \mathbf{c}_e + \lambda_0 \mathbf{1} + \mathbf{S}^\top (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}}) + (\overline{\boldsymbol{\gamma}}_e - \underline{\boldsymbol{\gamma}}_e) = \mathbf{0} \quad (19b)$$

$$\mathbf{H}_n \mathbf{p}_n + \mathbf{c}_n + \lambda_0 \mathbf{1} + \mathbf{S}^\top (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}}) + (\overline{\boldsymbol{\gamma}}_n - \underline{\boldsymbol{\gamma}}_n) = \mathbf{0}. \quad (19c)$$

Since $(\mathbf{H}_r, \mathbf{H}_e, \mathbf{H}_n)$ are invertible, the primal solutions can be expressed as

$$\mathbf{p}_r = -\mathbf{H}_r^{-1} \mathbf{z}_r, \quad \mathbf{p}_e = -\mathbf{H}_e^{-1} \mathbf{z}_e, \quad \mathbf{p}_n = -\mathbf{H}_n^{-1} \mathbf{z}_n \quad (20)$$

where $\mathbf{z}_r := \mathbf{c}_r + \lambda_0 \mathbf{1} + \mathbf{S}^\top (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}}) + (\overline{\boldsymbol{\gamma}}_r - \underline{\boldsymbol{\gamma}}_r)$, while $\mathbf{z}_e$ and $\mathbf{z}_n$ are defined accordingly.

Evaluating the Lagrangian function at the optimal primal solutions of (20) provides the dual function

$$-\frac{1}{2} \left[\mathbf{z}_r^\top \mathbf{H}_r^{-1} \mathbf{z}_r + \mathbf{z}_e^\top \mathbf{H}_e^{-1} \mathbf{z}_e + \mathbf{z}_n^\top \mathbf{H}_n^{-1} \mathbf{z}_n \right] + y - \overline{\boldsymbol{\gamma}}_e^\top \bar{\mathbf{p}}_e - \overline{\boldsymbol{\gamma}}_n^\top \bar{\mathbf{p}}_n$$

where $y := -\left[\lambda_0 \mathbf{1} + \mathbf{S}^\top (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}}) \right]^\top \boldsymbol{\ell} - \overline{\boldsymbol{\gamma}}_r^\top \bar{\mathbf{p}}_r - (\overline{\boldsymbol{\mu}} + \underline{\boldsymbol{\mu}})^\top \bar{\mathbf{f}}$. Due to strong duality, the optimal value of the dual function equals the optimal value of the primal objective

$$\frac{1}{2} \mathbf{p}_r^\top \mathbf{H}_r \mathbf{p}_r + \mathbf{c}_r^\top \mathbf{p}_r + \frac{1}{2} \mathbf{p}_e^\top \mathbf{H}_e \mathbf{p}_e + \mathbf{c}_e^\top \mathbf{p}_e + \frac{1}{2} \mathbf{p}_n^\top \mathbf{H}_n \mathbf{p}_n.$$

Equating the optimal primal and dual functions provides

$$\overline{\boldsymbol{\gamma}}_e^\top \bar{\mathbf{p}}_e + \overline{\boldsymbol{\gamma}}_n^\top \bar{\mathbf{p}}_n = -\frac{1}{2} \left[\mathbf{z}_r^\top \mathbf{H}_r^{-1} \mathbf{z}_r + \mathbf{z}_e^\top \mathbf{H}_e^{-1} \mathbf{z}_e + \mathbf{z}_n^\top \mathbf{H}_n^{-1} \mathbf{z}_n \right]$$
$$- \frac{1}{2} \left[\mathbf{p}_r^\top \mathbf{H}_r \mathbf{p}_r + \mathbf{p}_e^\top \mathbf{H}_e \mathbf{p}_e + \mathbf{p}_n^\top \mathbf{H}_n \mathbf{p}_n \right]$$
$$- \mathbf{c}_r^\top \mathbf{p}_r - \mathbf{c}_e^\top \mathbf{p}_e - \mathbf{c}_n^\top \mathbf{p}_n + y. \quad (21)$$

From (20), it follows that $\mathbf{p}_r^\top \mathbf{H}_r \mathbf{p}_r = \mathbf{z}_r^\top \mathbf{H}_r^{-1} \mathbf{z}_r$ and similarly for the subscripts e and n. Then, (21) simplifies as

$$\overline{\boldsymbol{\gamma}}_e^\top \bar{\mathbf{p}}_e + \overline{\boldsymbol{\gamma}}_n^\top \bar{\mathbf{p}}_n = -\mathbf{p}_r^\top \mathbf{H}_r \mathbf{p}_r - \mathbf{p}_e^\top \mathbf{H}_e \mathbf{p}_e - \mathbf{p}_n^\top \mathbf{H}_n \mathbf{p}_n$$
$$- \mathbf{c}_r^\top \mathbf{p}_r - \mathbf{c}_e^\top \mathbf{p}_e - \mathbf{c}_n^\top \mathbf{p}_n + y. \quad (22)$$

Combining the definition of electricity prices in (2) with (19) provides

$$\boldsymbol{\pi} = \mathbf{H}_e \mathbf{p}_e + \mathbf{c}_e + (\overline{\boldsymbol{\gamma}}_e - \underline{\boldsymbol{\gamma}}_e) \quad (23a)$$
$$= \mathbf{H}_n \mathbf{p}_n + \mathbf{c}_n + (\overline{\boldsymbol{\gamma}}_n - \underline{\boldsymbol{\gamma}}_n) \quad (23b)$$

Pre-multiplying (23a) by $\mathbf{p}_e^\top$ and (23b) by $\mathbf{p}_n^\top$, and summing up gives

$$\boldsymbol{\pi}^\top (\mathbf{p}_e + \mathbf{p}_n) = \mathbf{p}_e^\top \mathbf{H}_e \mathbf{p}_e + \mathbf{c}_e^\top \mathbf{p}_e + \overline{\boldsymbol{\gamma}}_e^\top \mathbf{p}_e - \underline{\boldsymbol{\gamma}}_e^\top \mathbf{p}_e$$
$$+ \mathbf{p}_n^\top \mathbf{H}_n \mathbf{p}_n + \mathbf{c}_n^\top \mathbf{p}_n + \overline{\boldsymbol{\gamma}}_n^\top \mathbf{p}_n - \underline{\boldsymbol{\gamma}}_n^\top \mathbf{p}_n. \quad (24)$$

Thanks to complementary slackness $\underline{\boldsymbol{\gamma}}_n^\top \mathbf{p}_n = \underline{\boldsymbol{\gamma}}_e^\top \mathbf{p}_e = 0$. Then, from (22), we finally get that

$$\boldsymbol{\pi}^\top (\mathbf{p}_e + \mathbf{p}_n) = -\mathbf{p}_r^\top \mathbf{H}_r \mathbf{p}_r - \mathbf{c}_r^\top \mathbf{p}_r + y. \quad (25)$$

REFERENCES

[1] D. Pozo, E. Sauma, and J. Contreras, "Basic theoretical foundations and insights on bilevel models and their applications to power systems," *Annals of Operations Research*, vol. 254, no. 1, pp. 303–334", Jul. 2017.

[2] S. J. Kazempour, A. J. Conejo, and C. Ruiz, "Strategic generation investment using a complementarity approach," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 940–948, May 2011.

[3] A. J. Conejo, L. B. Morales, S. J. Kazempour, and A. S. Siddiqui, *Investment in Electricity Generation and Transmission: Decision Making under Uncertainty*, 1st ed. Switzerland: Springer, 2016.

[4] Y. Liu, R. Sioshansi, and A. J. Conejo, "Multistage stochastic investment planning with multiscale representation of uncertainties and decisions," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 781–791, Jan. 2018.

[5] F. D. Munoz and J.-P. Watson, "A scalable solution framework for stochastic transmission and generation planning problems," *Computational Management Science*, vol. 12, no. 4, pp. 491–518, Oct 2015.

[6] V. Dvorkin, J. Kazempour, L. Baringo, and P. Pinson, "A consensus-ADMM approach for strategic generation investment in electricity markets," in *Proc. IEEE Conf. on Decision and Control*, Miami Beach, FL, USA, Dec. 2018, pp. 780–785.

[7] F. Borrelli, A. Bemporad, and M. Morari, "Geometric algorithm for multiparametric linear programming," *Journal of Optimization Theory and Applications*, vol. 118, no. 3, pp. 515–540, Sep. 2003.

[8] P. Tondel, T. A. Johansen, and A. Bemporad, "An algorithm for multi-parametric quadratic programming and explicit MPC solutions," *Automatica*, vol. 39, no. 3, pp. 489 – 497, 2003.

[9] Q. Zhou, L. Tesfatsion, and C. Liu, "Short-term congestion forecasting in wholesale power markets," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2185–2196, Nov. 2011.

[10] Y. Ng, S. Misra, L. A. Roald, and S. Backhaus, "Statistical learning for DC optimal power flow," in *Proc. Power Systems Computation Conference*, Dublin, Ireland, Jun. 2018, pp. 1–7.

[11] Y. Ji, R. J. Thomas, and L. Tong, "Probabilistic forecast of real-time LMP via multiparametric programming," in *International Conference on System Sciences*, Kauai, HI, USA, Jan. 2015, pp. 2549–2556.

[12] X. Geng and L. Xie, "Learning the LMP-load coupling from data: A support-vector machine based approach," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1127–1138, Mar. 2017.

[13] A. N. Madavan, S. Bose, Y. Guo, and L. Tong, "Risk-sensitive security-constrained economic dispatch via critical region exploration," in *Proc. IEEE PES General Meeting*, Atlanta, GA, USA, Aug. 2019.

[14] S. Taheri, M. Jalali, V. Kekatos, and L. Tong, "Fast probabilistic hosting capacity analysis for active distribution systems," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2000–2012, May 2021.

[15] Y. Ji, R. J. Thomas, and L. Tong, "Probabilistic forecasting of real-time LMP and network congestion," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 831–841, Mar. 2017.

[16] M. A. El-Meligy, M. Sharaf, and A. T. Soliman, "A coordinated scheme for transmission and distribution expansion planning: A tri-level approach," *Electric Power Systems Research*, vol. 196, p. 107274, 2021.

[17] V. Kekatos, G. B. Giannakis, and R. Baldick, "Online energy price matrix factorization for power grid topology tracking," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1239–1248, May 2016.

[18] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. New York, NY: Wiley & Sons, 1996.

[19] M. Garcia and R. Baldick, "Approximating economic dispatch by linearizing transmission losses," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1009–1022, Mar. 2020.

[20] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.

[21] J. D. Glover, T. J. Overbye, and M. S. Sarma, *Power system analysis and design*, 6th ed. Boston, MA: Cengage Learning, 2017.

[22] PJM enhanced data management tool. [Online]. Available: https://www.pjm.com/markets-and-operations/etools/data-miner-2.aspx

[23] A. Domahidi, E. Chu, and S. Boyd, "ECOS: An SOCP solver for embedded systems," in *European Control Conference (ECC)*, Zurich, Switzerland, Jul. 2013, pp. 3071–3076.

[24] J. Lofberg, "A toolbox for modeling and optimization in MATLAB," in *Proc. of the CACSD Conf.*, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip/

[25] L. Gurobi Optimization, "Gurobi optimizer reference manual," 2020. [Online]. Available: http://www.gurobi.com

Sina Taheri (S'19) received the B.S. degree from Sharif University of Technology, Tehran, Iran, in 2016; and the M.Sc. degree from Virginia Tech, Blacksburg, VA, USA, in 2019; both in electrical engineering. He is currently pursuing a Ph.D. degree at Virginia Tech. His research interests are focused on the application of optimization and machine learning to develop algorithmic solutions for monitoring and operation of smart power systems.

Vassilis Kekatos (SM'16) is an Associate Professor with the Bradley Dept. of ECE at Virginia Tech. He obtained his Diploma, M.Sc., and Ph.D. in computer science and engineering from the Univ. of Patras, Greece, in 2001, 2003, and 2007, respectively. He is a recipient of the NSF Career Award in 2018 and the Marie Curie Fellowship. He has been a research associate with the ECE Dept. at the Univ. of Minnesota, where he received the postdoctoral career development award (honorable mention). During 2014, he stayed with the Univ. of Texas at Austin and the Ohio State Univ. as a visiting researcher. His research focus is on optimization and learning for future energy systems. He is currently serving in the editorial board of the IEEE Trans. on Smart Grid.

Sriharsha Veeramachaneni is a researcher at NextEra Analytics Inc., a subsidiary of NextEra Energy Inc., specializing in statistics, machine learning and optimization. He holds a Ph.D. in computer engineering from the Rensselaer Polytechnic Institute.