Two-Timescale Stochastic Dispatch of Power Distribution Networks

Luis M. Lopez-Ramos, Student Member, IEEE, Vassilis Kekatos, Member, IEEE, Antonio G. Marques, Senior Member, IEEE, and Georgios B. Giannakis, Fellow, IEEE

Abstract—Smart distribution grids should optimally integrate stochastic renewable resources while effecting voltage regulation. Since some decisions have to be designed in advance, energy management is a multistage problem. For early stages, finding the optimal energy procurement accounting for the variability during real-time operation is a challenging task. The joint dispatch of slow- and fast-timescale controls in a distribution grid is considered here. The substation voltage, the energy exchanged with a main grid, and the generation schedules for small diesel generators have to be decided on a slow timescale; whereas optimal photovoltaic inverter setpoints are found on a more frequent basis. While inverter and looser voltage regulation limits are imposed at all times, tighter bus voltage constraints are enforced on the average or in probability, thus enabling more efficient renewable integration. Upon reformulating the two-stage grid dispatch as a stochastic convex-concave problem, two distribution-free schemes are put forth. An average dispatch algorithm converges provably to the optimal two-stage decisions via a sequence of convex quadratic programs. Its non-convex probabilistic alternative entails solving two slightly different convex problems and is numerically shown to converge. Numerical tests on a real-world distribution feeder verify that both novel schemes yield lower costs over competing alternatives.

Index Terms—Multistage economic dispatch, voltage regulation, stochastic approximation, convex-concave problem.

I. INTRODUCTION

With increasing renewable generation, energy management of power distribution grids is becoming a computationally challenging task. Solar energy from photovoltaic (PV) units can change significantly over one-minute intervals. The power inverters found in PV units can be commanded to curtail active power generation or adjust their power factor within seconds [1], [2]. At a slower timescale, distribution grid operators exchange energy with the main grid hourly or on a 10-minute basis, and may experience cost penalties upon deviating from energy market schedules [3]. Moreover, voltage regulation equipment and small diesel generators potentially installed in microgrids respond at the same slower timescale. This mode of operation calls for multistage dispatch solutions.

Spurred by demand-response programs and the use of PV inverters to accomplish various grid tasks [4], single-stage dispatch schemes for distribution grids have been an active area of research. Power inverters can be controlled using localized rules for voltage regulation, see e.g., [5], [6], [7], [8]. Assuming two-way communication between buses and the utility operator, dispatching a distribution system can be posed as an optimal power flow (OPF) problem. Centralized schemes use nonlinear program solvers [9]; or rely on convex relaxations of the full ac model of balanced [10], [11], or unbalanced grids [12]. Distributed solvers with reduced computational complexity have been devised in [13], [14], [15]. Nevertheless, the efficient and secure operation of distribution grids involves decisions at different timescales. A dynamic programming approach for a two-stage dispatch is suggested in [16]: The taps of voltage regulators are set on a slow timescale and remain fixed for consecutive shorter time slots over which elastic loads are dispatched; yet the flexibility of loads is assumed known a priori. Alternatively, centrally computed OPF decisions can be communicated to buses at a slow timescale, while on a faster timescale, PV power electronics are adjusted to optimally track variations in renewable generation and demand [17], [18], [19]. Relying on approximate grid models, the latter schemes yield a fully localized real-time implementation. However, they presume smooth system transitions and dispatch slow-responding units for a single deterministic fast-timescale scenario.

Multistage dispatching under uncertainty is routinely used in transmission systems and microgrids [20]. Robust approaches find optimal slow-timescale decisions for the worst-case fast-timescale outcome; see [21] and references therein. To avoid the conservativeness of robust schemes, probabilistic approaches postulate a probability density function (pdf) for demand, wind generation, and system contingencies to find day-ahead grid schedules [22], [23]. The risk-limiting dispatch framework adjusts multistage decisions as the variance of the random variables involved decreases while approaching actual time [3]. Decisions can be efficiently calculated upon ignoring network limitations and only for convenient pdfs [24]. The same holds true for a network-constrained risk-limiting dispatch under additional transmission congestion assumptions [25]. As a third alternative, stochastic sample approximation approaches yield optimal slow-timescale decisions using samples drawn from the postulated pdf; see e.g., [26], [27].

Returning to distribution grids, PV inverters could be overloaded sporadically in space and time to accommodate solar fluctuations and prevent overvoltages [28]. Nonetheless, ensuring that overloading occurs sparingly couples decisions across time. The single-stage scheme of [29] finds optimal
PV setpoints while limiting time averages of overloaded quantities. The latter approach has been also adopted in [30] for dispatching a transmission system under a day-ahead/real-time market setup under load shedding.

Jointly dispatching slow- and fast-timescale distribution grid resources under average or probabilistic constraints over fast-timescale decisions is considered here. Our contributions are three-fold. First, using an approximate grid model, the expected cost over a slow control period is minimized while inverter and looser bus voltage constraints are satisfied at all fast-timescale slots. Further, tighter voltage limits are enforced either on the average or in probability across successive fast-timescale slots (see Section II). Two-stage grid dispatch is formulated as a convex-concave optimization (see Section III). Second, adopting a stochastic saddle-point approximation scheme from [31], the provably convergent algorithm of Section IV finds the optimal slow-timescale decisions in the case of average constraints. Third, for the case of non-convex probabilistic constraints, an algorithm solving two similar convex problems for each fast-timescale period is put forth in Section V. Albeit the related expected recourse function enjoys zero-duality gap [32], the overall two-stage dispatch is not convex-concave; and the algorithm’s performance is only numerically validated. Both algorithms require only samples rather than pdfs of loads and solar generation, and involve solving simple convex quadratic programs. Numerical tests in Section VI on a Southern California Edison feeder corroborate the validity of our findings. The paper is concluded in Section VII.

Regarding notation, lower-(upper-)case boldface letters denote column vectors (matrices), with the only exception of the power flow vectors, which are uppercase. Calligraphic letters are used to denote sets. Symbol $\mathbb{I}\{\cdot\}$ denotes transposition, while $\mathbf{0}$ and $\mathbf{1}$ are the all-zeros and all-ones vectors of appropriate dimensions. The indicator function $\mathbb{I}\{\cdot\}$ equals 1 when its argument is true, and 0 otherwise. A diagonal matrix with the entries of vector $\mathbf{x}$ on its main diagonal is denoted by $\text{diag}(\mathbf{x})$. The operator $[\cdot]_{+}$ projects its argument onto the positive orthant. Symbols $\text{Pr}\{\cdot\}$ and $\mathbb{E}\{\cdot\}$ are reserved for probabilities and the expectation operator accordingly.

II. PROBLEM FORMULATION

Consider a distribution grid whose energy needs are procured by distributed renewable generation, distributed conventional (small diesel) generators, and the main grid. The distribution grid operator aims at serving the load at the minimum cost while respecting voltage regulation and distribution network constraints. Energy is exchanged with the main grid at whole-sale electricity prices through the feeder bus. To effectively integrate stochastic renewable generation, the focus here is on short-term grid dispatch. To that end, the distribution grid is operated at two timescales: a slower timescale corresponds to 5- or 10-min real-time energy market intervals, while the power inverters found in PVs are controlled at a faster timescale of say 10-sec intervals. One period of the slower timescale is comprised by $T$ faster time slots indexed by $t = 1, \ldots, T$.

The grid is operated as a radial network with $N + 1$ buses rooted at the substation bus indexed by $n = 0$. The distribution line feeding bus $n$ is also indexed by $n$ for $n = 1, \ldots, N$. Let $p_{n,t}$ and $q_{n,t}$ denote respectively the net active and reactive power injections at bus $n$ and slot $t$; the $N$-dimensional vectors $\mathbf{p}_t$ and $\mathbf{q}_t$ collect the net injections at all buses except for the substation. Diesel generators are dispatched at the slower timescale to generate $\mathbf{p}_t^d$ throughout the subsequent $T$ slots at unit power factor. During slot $t$, PVs can contribute solar generation up to $\mathbf{p}_t^s$ that is modeled as a random process. Smart inverters perform active power curtailment and reactive power compensation by following the setpoints $\mathbf{p}_t^r$ and $\mathbf{q}_t^r$ commanded by the utility operator. Load demands represented by $p_{0,t}^l$ and $q_{0,t}^l$ are modeled as random processes as well. To simplify the exposition, loads ($\mathbf{p}_t^l, \mathbf{q}_t^l$) are assumed inelastic and known at the beginning of slot $t$; even though elastic loads can be incorporated without any essential differences. The operator buys a power block $\mathbf{p}_t^b$ from the main grid at the slow timescale, which can be adjusted to $p_{0,t} := p_{0,t}^b + p_{0,t}^d$ in actual time.

Voltage regulation is effected by controlling (re)active power injections at slot $t$. Let $\mathbf{v}_{n,t}$ denote the squared voltage magnitude at bus $n$ and slot $t$, and $\mathbf{v}_t$ the vector collecting $\{\mathbf{v}_{n,t}\}_{n=1}^N$. The substation voltage $\mathbf{v}_t^0$ is controlled at the slower timescale [16], while voltage magnitudes at all buses must adhere to voltage regulation standards, e.g., ANSI C84.1 and EN50160 in [33], [34]. These standards differentiate between two voltage regions: one for normal operation denoted here by $\mathcal{V}_A$, and one whose use is limited in frequency and duration and will be denoted by $\mathcal{V}_B$. Based on these regions, two alternative dispatching schemes will be formulated. Both schemes constraint voltages within $\mathcal{V}_B$ at all times, while region $\mathcal{V}_A$ is enforced in a stochastic fashion: The first scheme guarantees that voltages lie in $\mathcal{V}_A$ on the average, whereas the second one maintains the probability of under/over-voltage at a specified low value.

A. Grid modeling

To capture voltage and network limitations, the distribution grid is captured by the approximate linear distribution flow (LDF) model, which is briefly reviewed next [35]. Let $\mathbf{r}$ and $\mathbf{x}$ be accordingly the vectors of line resistances and reactances across lines. Define also the branch-bus incidence matrix $\mathbf{A} \in \mathbb{R}^{N \times (N+1)}$ whose $(i,j)$-th entry is

$$
\mathbf{A}_{ij} = \begin{cases} +1 & \text{if } j = 1 \text{ is the source bus of line } i \\ -1 & \text{if } j = 1 \text{ is the destination bus of line } i \\ 0 & \text{otherwise.} \end{cases}
$$

Partition $\mathbf{A}$ into its first column and the reduced branch-bus incidence matrix $\tilde{\mathbf{A}}$ as $\mathbf{A} = [\mathbf{a}_0 \ \tilde{\mathbf{A}}]$. Ignoring line losses, the LDF model asserts that the vectors of active and reactive line power flows at time $t$ can be approximated by

$$
\mathbf{p}_t = F^\top \mathbf{p}_t^l \text{ and } \mathbf{q}_t = F^\top \mathbf{q}_t^l
$$

where $F := A^{-1}$. Moreover, the squared voltage magnitudes can be expressed as [35], [36], [37]

$$
\mathbf{v}_t = 2R\mathbf{p}_t + 2X\mathbf{q}_t + \mathbf{v}_t^0 \mathbf{1}
$$

Here $R$ and $X$ are the diagonal matrices of line resistances and reactances, respectively, $\mathbf{1}$ is a column vector of ones, and $\mathbf{v}_t^0$ is the initial condition of voltage magnitudes.
where \( R := F d g ( r ) F^T \) and \( X := F d g ( x ) F^T \). Let us define the voltage regulation regions

\[
V_A := \{ v : \| v \|_2 \leq \tau_A \}
\]

\[
V_B := \{ v : \| v \|_2 \leq \tau_B \}
\]

(4a) (4b)

with \( \tau_B \geq \tau_A \) and \( \tau_B \leq \tau_A \). Compliance with \( V_A \) can be imposed either on the average as \( \mathbb{E}_t [v_i] \in V_A \), or in probability as \( \Pr \{ v_i \in V_A \} \geq 1 - \alpha \) for some small \( \alpha \). Either way, safe grid operation requires that \( v_i \in V_B \) at all times \( t \). The random processes involved (demand and renewable generation) can be assumed ergodic, i.e., their time averages converge to their ensemble averages. For this reason, voltage constraints pertaining to \( V_A \) will be referred to as ergodic.

According to (2), if \( f_n \) is the \( n \)-th column of \( F \), the squared power flow on line \( n \) can be written as \( P^2_{n,t} = p^T_n f_n f_n^T p_t \) and \( Q^2_{n,t} = q^T_n f_n f_n^T q_t \). Imposing the upper limit \( S_n \) on the apparent power flow on line \( n \) is thus expressed as the convex quadratic constraint

\[
P^T_n f_n f_n^T p_t + q^T_n f_n f_n^T q_t \leq S^2_n.
\]

Assuming voltage magnitudes to be close to unity, active power losses can be approximated as \( \sum_{n=1}^{N} r_n (P^2_{n,t} + Q^2_{n,t}) \) [38], which from (2)–(3), can be equivalently expressed as \( P^T_t d g ( r ) P_t + Q^T_t d g ( r ) Q_t = p^T_t R p_t + q^T_t R q_t \). Thus, the active power injection at the substation is approximately

\[
p_{0,t} = -1^T p_t + p^T_t R p_t + q^T_t R q_t
\]

(6)

Regarding smart inverters, the injection from the inverter located on bus \( n \) at slot \( t \) should belong to the feasible set

\[
\Omega_n ( t ) := \{ ( p^r_{n,t}, q^r_{n,t} ) : 0 \leq p^r_{n,t} \leq p^r_{n,t} , \quad | q^r_{n,t} | \leq \phi_n p^r_{n,t} , \quad \langle ( p^r_{n,t} )^2 + ( q^r_{n,t} )^2 \rangle \leq \sigma^2_n \}
\]

(7a) (7b) (7c)

that is random and time-variant due to the variability of \( p^r_{n,t} \). Constraint (7a) limits the active power generation according to the available solar power; constraint (7b) enforces the lower limit \( \cos(\arctan(\phi_n)) \) on the power factor (lagging or leading); and (7c) limits the inverter apparent power.

B. Operation costs

If PV owners are compensated at price \( \pi \) for the active power surplus they inject into the distribution grid, the related utility cost at slot \( t \) is \( C_{PV}(p_t^r) := \pi^T [p^r_t - p^r_0]_+ \) with \( [x]_+ := \max \{ 0, x \} \) applied entrywise on vector \( x \). The diesel generation cost is represented by \( C_D(p_t^d) \). Regarding energy transactions with the main grid, the power block \( p_0^d \) bought in advance is charged at a fixed and known price \( \beta \). Deviating from \( p_0^d \) by \( p_t^d \) at slot \( t \) is charged at

\[
C^t(\delta_t^d) := \gamma_b [p_0^d, p_t^d]_+ - \gamma_s [p_0^d]_+ \]

for known prices \( \gamma_b, \gamma_s \). To avoid arbitrage, it is assumed that \( 0 < \gamma_s < \beta < \gamma_b \); see e.g., [3], [25]. Then, the deviation charge can also be expressed as \( C^t(\delta_t^d) = \max \{ \gamma_b p_0^d, \gamma_s p_0^d \} + \) which is certainly convex [21].

C. Optimal grid dispatch

Depending on the way compliance with voltage regulation region \( V_A \) is enforced, two grid dispatch formulations are developed next. Commencing with the average dispatch, the optimal grid operation is posed as

\[
P_a^* := \min \{ C_D(p_t^d) + \beta p_0^d + E_t [C^t(\delta^d_{0,t}) + C_{PV}(p_t^r)] \}
\]

s.t.:

\[
q_t = q^r_t - d_t, \quad p_{0,t} = p^d_0 + p^d_{0,t}, \quad p_{0,t} \geq -1^T p_t + p^T_t R p_t + q^T_t R q_t \quad \forall n \in N
\]

\[
q_p^T f_n f_n^T p_t + q^T_n f_n f_n^T q_t \leq S^2_n, \quad \forall n \in N
\]

(9a) (9b) (9c) (9d) (9e)

\[
p^d \leq p^d \leq \eta^d
\]

\[
(p^r_{n,t}, q^r_{n,t}) \in \Omega_n ( t ), \quad \forall n \in N
\]

\[
\nu_0 \leq \nu^a_0 \leq \nu_0
\]

\[
v_t = 2 R p_t + 2 X q_t + \nu^0_1
\]

\[
v_t \in V_B
\]

(9f) (9g) (9h) (9i) (9j) (9k)

The slow-timescale variables \( \{ \nu^a_0, p_0^d, p^d \} \) are set in advance, and remain fixed throughout the \( T \) subsequent control slots over which the fast-timescale variables \( \{ p_t, q_t, v_t, p^r_t, q^r_t, p_{0,t}, p^d_{0,t}, T \}_{t=1}^T \) are implemented. The latter variables depend on the randomness of slot \( t \) as well as slow-timescale decisions.

Alternatively to (9), the optimal grid operation can be expressed by the probabilistic dispatch that is identical to (9) with the exception that (9l) is replaced by the probabilistic constraint

\[
Pr \{ v_t \notin V_A \} \leq \alpha
\]

(10)

for some small parameter \( \alpha > 0 \), say \( \alpha = 0.05 \). The optimal cost for the probabilistic dispatch will be denoted by \( P_p^* \).

The objective function in (9a) involves the cost of energy dispatched at the slow timescale plus the average fast-timescale energy management cost. Nodal (re)active power balance is ensured via (9b)–(9c). Constraint (9e) accounts for the active power losses upon relaxing the quadratic equation in (6) to a convex inequality without loss of optimality. Constraint (9f) limits the apparent power flow at each line \( n \) based on (5). Constraints (9i)–(9l) are voltage regulation constraints: In detail, (9i) relates squared voltage magnitudes to power injections [cf. (3)]; (9j) constrains the substation bus voltage; and (9k) constrains voltages in \( V_B \). While (9l) maintains the average voltage magnitudes in \( V_A \), its alternative in (10) limits the probability of voltage magnitudes being outside \( V_A \).

III. Problem Analysis

To facilitate algorithmic developments, the problem in (9) is expressed in a more compact form next. Collect the slow-timescale variables in vector \( z^T := [\nu^a_0, p_0^d, p^d] \); the fast-timescale variables at slot \( t \) in \( y^T_t := [p_t, q_t, v_t, p^r_t, q^r_t, p_{0,t}, p^d_{0,t}] \); and the random variables involved at slot \( t \) in \( \xi^T_t := [p^r_t, p^d_t, q^r_t] \).
The constraints of (9) can be then classified into four groups:

(i) Constraints involving fast-timescale variables only, such as (9c), (9f), (9h), and (9k), that will be abstracted as \( y_t \in Y_t \).

(ii) Constraints (9g) and (9i) that involve slow-timescale variables only, and they will be denoted as \( z \in Z \).

(iii) The linear constraints (9b), (9d), and (9j), coupling slow- and fast-timescale variables as well as random variables. These constraints are collectively expressed as \( Kz + By_t = H\xi_t \) for appropriate matrices \( K, B, \) and \( H \).

(iv) The ergodic constraints (9l) and (10) depend on the voltage sequence \( \{v_t\} \), hence coupling decisions across time. A substantial difference between (9l) and (10) is that the latter is a non-convex constraint.

Based on the above grouping, the two dispatch problems can be compactly rewritten as

\[
P_{(a,p)}^{\ast} := \min_{z \in Z} f(z) + \mathbb{E}_t [g_t(y_t)]
\]

s.to: \( y_t \in Y_t \quad \forall t \)

\[
Kz + By_t = H\xi_t
\]

where \( f(z) := C_D(p^d) + \beta_p^q + g_t(y_t) := C^f(\nu^t) + C_{PV}(p^t) \). For the average dispatch, the optimal cost in (11) is \( P_a^\ast \) and the function in (11e) is \( h(y_t) = \mathbb{E}_t - t_A \mathbf{1}, z_A \mathbf{1} - v_t \).

For the probabilistic dispatch, the optimal cost is \( P_p^\ast \) and the function in (11e) is \( h(y_t) = \mathbb{E}_t - t_A \mathbf{1} - \alpha \).

The optimal values for the slow-timescale variables \( z \) have to be decided in advance. Once the optimal \( z \) has been found, it remains fixed over the slow-timescale interval. The fast-timescale decisions \( y_t(z) \) for slot \( t \) depend on the slow-timescale variable \( z \), while the subscript \( t \) indicates their dependence on the realization \( \xi_t \). Both the average and the probabilistic dispatch are stochastic programming problems with recourse [3]. Their costs can be decomposed as \( P_{(a,p)}^\ast = \min_{z \in Z} f(z) + G_{(a,p)}(z) \), where the so termed expected recourse function is defined as

\[
G_{(a,p)}(z) := \min_{y_t \in Y_t} \mathbb{E}_t [g_t(y_t)]
\]

s.to: \( Kz + By_t = H\xi_t \quad \forall t \)

\[
\mathbb{E}_t [h(y_t)] \leq 0
\]

with \( \{y_t(z)\} \) being one of its minimizers. The ensuing two sections solve the average and the probabilistic dispatches.

### IV. AVERAGE DISPATCH ALGORITHM

This section tackles problem (11) with the ergodic constraint in (11e), for which \( h(y_t) = \mathbb{E}_t - t_A \mathbf{1}, z_A \mathbf{1} - v_t \). Although convex, this problem is challenging due to the coupling across \( \{y_t\} \) and between fast- and slow-timescale variables. A dual decomposition approach is adopted to resolve the coupling across \( \{y_t\} \). The partial Lagrangian function for (12) is

\[
L_a(\{y_t\}, \nu) := \mathbb{E}_t [g_t(y_t) + \nu^T h(y_t)].
\]

The corresponding dual function is

\[
D_a(\nu; z) := \min_{y_t \in Y_t} L_a(\{y_t\}, \nu)
\]

s.to: \( Kz + By_t = H\xi_t \quad \forall t \).

Observe that the minimization in (14) is separable over the realizations \( \{\xi_t\} \). Precisely, the optimal fast-timescale variable for fixed \((\nu, z)\) and for a specific realization \( \xi_t \) can be found by solving:

\[
y_t^\ast(\nu, z) \in \arg \min_{y_t \in Y_t} g_t(y_t) + \nu^T h(y_t)
\]

s.to: \( Kz + By_t = H\xi_t \quad \forall t \).

For future reference, let us also define \( \lambda_t^\ast(\nu, z) \) as the optimal Lagrange multiplier associated with (15). If \( \nu \) is partitioned as \( \nu = [\nu, \nu] \) with \( \nu \) corresponding to constraint \( \mathbb{E}_t [v_t] \geq t_A \mathbf{1} \) and \( \nu \) to \( E_t [v_t] \leq t_A \mathbf{1} \), then (15) simplifies to

\[
y_t^\ast(\nu, z) \in \arg\min \{C_t^f(\nu^t) + C_{PV}(p_t^f) + (\nu - \nu)^T v_t \}
\]

s.to: \( (9b) - (9f), (9h), (9j) \)

over \( \{p_t, q_t, v_t, p_t^f, q_t^f, p_{0,t}, p_{0,t}^f\} \)

and can be solved as a convex quadratic program. Given the optimal pair \((\nu^\ast, z^\ast)\), the optimal fast-timescale variables \( y_t \) can be thus found for any \( \xi_t \).

Back to finding the optimal primal and dual slow-timescale variables, observe that the dual problem associated to (14) is

\[
\nu^\ast := \arg \max_{\nu \geq 0} D_a(\nu; z).
\]

Duality theory asserts that (17) is a convex problem. Moreover, assuming a strictly feasible point exists for (12), strong duality implies that \( G_a(\nu) = D_a(\nu^\ast; z) \). Due to the latter, the original problem in (11) can be transformed to:

\[
\min_{x \in Z} f(x) + g_a(z) = \min_{x \in Z} \{f(x) + \max_{\nu \geq 0} D_a(\nu; z)\}
\]

\[
= \min_{x \in Z} \tilde{f}_a(\nu, z)
\]

where the auxiliary function \( \tilde{f}_a \) is defined as:

\[
\tilde{f}_a(\nu, z) := f(x) + D_a(\nu; z).
\]

Being a dual function, \( D_a(\nu; z) \) is a concave function of \( \nu \). At the same time, \( D_a(\nu; z) \) is a perturbation function with respect to \( z \); and hence, it is a convex function of \( z \). Recall that \( f(z) \) is a convex function of \( z \) too. Therefore, the auxiliary function \( \tilde{f}_a(\nu, z) \) is convex in \( z \) and concave in \( \nu \). Because of the randomness of \( \{\xi_t\} \), function \( D_a(\nu; z) \) in (14) is stochastic. Consequently, problem (18b) is a stochastic convex-concave saddle point problem [39], [31].

To solve (18b), we rely on the stochastic saddle-point approximation method of [31]. The method involves the subgradient of \( \tilde{f}_a \) wrt \( z \) and its supergradient wrt \( \nu \). Upon viewing \( D_a(\nu; z) \) in (14) as a perturbation function of \( z \), the subgradient of \( \tilde{f}_a \) with respect to \( z \) is [39]

\[
\partial_z \tilde{f}_a = \partial_z f(z) + K^T \mathbb{E}_t [\lambda_t^\ast(\nu, z)].
\]

By definition of the dual function, the supergradient of \( \tilde{f}_a \) with respect to \( \nu \) is

\[
\partial_\nu \tilde{f}_a = (\nu) = \mathbb{E}_t [h(y_t^\ast(\nu, z))].
\]

The stochastic saddle point approximation method of [31] involves primal-dual subgradient iterates with the expectations...
in (20)–(21) being replaced by their instantaneous estimates based on a single realization $\xi_k$. Precisely, the method involves the iterates over $k$:

$$\nu^{k+1} := \nu^k + \delta \mu_k h(y_k^*(\nu^k, z_k))$$  \hspace{1cm} (22a)

$$z^{k+1} := [z^k - \delta e_k (\partial_z f(z^k) + \mathbf{K} \Lambda^*_k(\nu^k, z^k))]_Z$$  \hspace{1cm} (22b)

where the operator $[\cdot]_Z$ projects its argument onto $Z$; and vectors $\mu_k = \mu_0/\sqrt{k}$ and $e_k = e_0/\sqrt{k}$ collect respectively the primal and dual step sizes for positive $\mu_0$ and $e_0$. At every iteration $k$, the method draws a realization $\xi_k$ and solves (15) for the tuple $(\xi_k, \nu^k, z^k)$ to acquire $(y_k^*(\nu^k, z^k), \Lambda^*_k(\nu^k, z^k))$ and perform the primal-dual updates in (22). The method finally outputs the sliding averages of the upates as:

$$\bar{z}_k := \frac{\sum_{i=[k/2]}^k z^i}{\sum_{i=[k/2]}^k 1/\sqrt{t}}$$  \hspace{1cm} (23a)

$$\bar{\nu}_k := \frac{\sum_{i=[k/2]}^k \nu^i}{\sum_{i=[k/2]}^k 1/\sqrt{t}}$$  \hspace{1cm} (23b)

The proposed scheme converges to the value $\bar{f}_z(\nu^*, z^*)$ obtained at a saddle point $(\nu^*, z^*)$ asymptotically in the number of iterations $k$ [31, Sec. 3.1].

Upon convergence of the iterates in (23), the slow-timescale variables $z^*$ have been derived together with the optimal Lagrange multiplier $\nu^*$ related to constraint (12c). The grid operator can implement $\nu^*$, and the fast-timescale decisions $y^*_t$ for a realization $\xi_t$ can be found by solving (16). The average dispatch algorithm (ADA) is summarized as Alg. 1.

V. PROBABILISTIC DISPATCH ALGORITHM

The probabilistic version of problem (11) is considered next. Here, the ergodic constraint (11e) reads $h(y_t) = I \{ v_t \notin V_A \} - \alpha$. Although the non-convexity of the probabilistic constraint does not affect the optimality while solving (12), optimality cannot be claimed when solving for (11). A heuristic solution is detailed next by adapting the solution of Sec. IV. Optimality cannot be claimed when solving (12), the constraint does not affect the optimality while solving (12), however, the solution to (24) can be found by solving a pair of slightly different convex problems. The first problem is

$${y_t}^* (v, z) \in \arg\min_{y_t \in Y} g_t(y_t) + \nu I \{ v_t \notin V_A \}$$  \hspace{1cm} (24)

s.t.: $Kz + By_t = H\xi_t$.

Because the indicator function takes only the values {0, 1} however, the solution to (24) can be found by solving a pair of slightly different convex problems. The first problem is

$${y_t}_B (z) \in \arg\min_{y_t \in Y} g_t(y_t)$$  \hspace{1cm} (25a)

s.t.: $Kz + By_t = H\xi_t$.  \hspace{1cm} (25b)

Let $\Lambda^*_B(z)$ being the optimal Lagrange multiplier associated with (25b). The second problem is obtained from (25) with the extra constraint that $v_t \in V_A$:

$${y_t}_A (z) \in \arg\min_{y_t \in Y} g_t(y_t)$$  \hspace{1cm} (26a)

s.t.: $Kz + By_t = H\xi_t$.  \hspace{1cm} (26b)

Let also $\Lambda^*_A(z)$ being the optimal Lagrange multiplier associated with (26b). Observe that neither problem (25) nor (26) depend on $\nu$, while their complexity is similar to that of the problem in (15). Having solved (25)–(26), problem (24) can be neatly tackled by identifying two cases:

(c1) If $g_t(y_t^*_A(z)) > g_t(y_t^*_B(z)) + \nu$, then $y_t^*_B(z)$ is a minimizer of (24) as well and voltages are allowed to lie outside $V_A$. In this case, set $y_t^*_B(v, z) := y_t^*_B(z)$ and $\Lambda^*_B(v, z) := \Lambda^*_B(z)$. This case includes instances where problem (26b) is infeasible for which $g_t(y_t^*_A(z)) = \infty$.

(c2) If $g_t(y_t^*_A(z)) \leq g_t(y_t^*_B(z)) + \nu$, then $y_t^*_A(z)$ minimizes (24) too and voltages lie within $V_A$. In this case, set $y_t^*_B(v, z) := y_t^*_A(z)$ and $\Lambda^*_B(v, z) := \Lambda^*_A(z)$.

Case (c2) covers also instances where $v_t^*_B(z)$ happens to lie in $V_A$. In these particular instances, $y_t^*_B(z)$ serves as a minimizer of (26b) too. Then, it follows that $g_t(y_t^*_A(z)) = g_t(y_t^*_B(z)) \leq g_t(y_t^*_B(z)) + \nu$ for $\nu \geq 0$. Hence, if $v_t^*_B(z) \in V_A$, there is no need to solve problem (26b).
To find the optimal slow-timescale variables under the probabilistic dispatch, the stochastic primal-dual iterations of Sec. IV are adapted here as

\[ \nu^{k+1} := [\nu^k + \mu_k(\{V_A^{\nu}(z^k) \notin \mathcal{V}_A\} - \alpha)]_+, \quad (27a) \]

\[ z^{k+1} := [z^k - d(g(\epsilon_k))\partial_z f(z^k) + K^T \lambda^*_k(\nu^k, z^k)]_+. \quad (27b) \]

The probabilistic dispatch algorithm (PDA) is tabulated as Alg. 2. Because function \( G_A(z) \) is not necessarily convex, the iterates in (27) are not guaranteed to converge to a minimizer of (11). The practical performance of PDA in finding \( z^* \) is numerically validated in Sec. VI.

VI. NUMERICAL TESTS

The proposed grid dispatches were tested on a 56-bus Southern California Edison (SCE) distribution feeder [11]. 5-MW PVs were added on buses 44 and 50; both with 6-MVA inverters enabling power factors as low as 0.83 (leading or lagging) at full solar generation. The prices for the energy exchange with the main grid were \( \beta \) ing or lagging) at full solar generation. The prices for the energy exchange with the main grid were

\[ \beta = 45 \$/MWh, \quad \gamma_p = 19 \$/MWh. \]

Diesel generators with capacity \( P^d_{\text{nom}} = 0.5 \text{MW} \) were sited on buses 10, 18, 21, 30, 36, 43, 51, and 55. The cost of diesel generation was

\[ C_D(p^d) = \sum_{n=1}^{N} (30p^d_{n} + 15(p^d_{n})^2) \$/h \]

with \( p^d \) expressed in MW. Apparent power flows were limited to 7 MVA. The voltage operation limits were set to \( V_{A_{\text{min}}} = 0.98^2, \quad V_{A_{\text{max}}} = 1.02^2 \), \( V_{B_{\text{min}}} = 0.97^2, \) and \( V_{B_{\text{max}}} = 1.03^2 \), expressed in pu with respect to a voltage base of 12 kV. Active and reactive nodal loads were Gaussian distributed with a mean value equal to the nominal load of the SCE benchmark, and a standard deviation of 0.2 times the nominal load. The solar energy generated at each PV was uniformly distributed between 0.5 and 1 times the actual power PV rating.

ADA was run for 40,000 iterations. Step sizes were selected proportional to \( 1/\sqrt{k} \) with initial values \( \epsilon_0^{\nu} = 4 \times 10^{-5}, \) \( \epsilon_0^p = 4 \times 10^{-1}, \) \( \epsilon_0^{\nu} = 6 \times 10^{-3}, \) and \( \mu_0 = 225 \), to account for different dynamic ranges. The iterates for primal and dual variables as well as their corresponding sliding averages are depicted in Fig. 1. Primal and dual slow-timescale variables hover in a small range whose width diminishes with time. Their sliding averages converge asymptotically. The algorithm reaches a practically meaningful solution within 5,000 iterations. Buses 44 and 50 are prone to over-voltages since they host PV generation, thus yielding non-zero dual variables for the average upper voltage constraints.

PDA was tested using the same simulation setup for \( \alpha = 0.05 \) and \( \mu_0 = 1 \). Figure 2 shows the convergence of primal and dual variables, the probability of voltages lying within \( \mathcal{V}_A \), and the grid operational cost. Granted that the probabilistic constraint in (10) applies collectively to all buses, the under/over-voltage probabilities on a per-bus basis is depicted in Fig. 3. The occurrences of overvoltage seems to be shared primarily among buses 40–56 which are neighboring to the PV buses 40 and 55. On the contrary, buses 10–16 being electrically far from both the substation and PVs, experience under-voltage with a small probability.

The effect of the average versus the probabilistic constraint on voltage magnitudes was evaluated next. After slow-timescale variables \( z \) had converged, fast-timescale variables \( y_t \) were calculated for 6,000 instances of \( \xi_t \) using both ADA and PDA. The histograms of the voltage magnitudes on two representative buses are presented in Fig. 4. Under PDA, the average voltage on bus 15 is slightly higher than the average voltage obtained by ADA. In exchange, the instantaneous value of the voltage on bus 15 stays in \( \mathcal{V}_A \) with higher probability. A similar behavior is observed for the overvoltage instances on PV bus 40.

ADA and PDA were finally compared against two alternative schemes. According to the first henceforth called approximate scheme, variables \( z \) are calculated by setting loads and solar generation to their expected values, while \( y_t \) are found by solving either (15) or (24), depending on whether the setting is average or probabilistic. The second deterministic scheme finds \( z \) as the approximate scheme does, and \( y_t \)’s are found by enforcing \( v_t \in \mathcal{V}_A \) at all times. The four dispatches were tested under five scenarios: Scenario 1 is the setup described earlier. Scenario 2 involved the tighter voltage limits \( V_{A_{\text{min}}} = 0.99^2 \) and \( V_{A_{\text{max}}} = 1.01^2 \). Scenarios 3, 4, and 5 were generated by scaling the mean value and the standard deviation for loads of Scenario 1 by 0.5, 1.5, and 2, respectively. Figure 5 shows the expected operation costs for all five scenarios. ADA (PDA) yielded the lowest cost under all scenarios in the average (probabilistic) setting as expected. Under the same operating conditions, the optimal cost for the average setting turned out to be slightly lower than that for the probabilistic one. Moreover, the improvement from the approximate scheme to the optimal one is larger than that from the deterministic scheme to the approximate one.

VII. CONCLUSIONS

By nature of renewable generation, electromechanical component limits, and the manner markets operate, energy management of smart distribution grids involves decisions at slower and faster timescales. Since slow-timescale controls remain fixed over multiple PV operation slots, decisions are coupled across time in a stochastic manner. To accommodate solar energy fluctuations, voltages have been allowed to be sporadically overloaded; hence introducing coupling of fast-timescale variables on the average or in probability. Average voltage constraints have resulted in a stochastic convex-concave problem, whereas non-convex probabilistic constraints have yielded a convex recourse function. Efficient algorithms for finding both slow and fast controls using only random samples have been put forth. Our two novel solvers converge in terms of the primal and dual variables, and have attained lower operational costs compared to deterministic alternatives. Although probabilistic constraints have been applied grid-wise, voltages on individual buses remained within limits. Enforcing probabilistic constraints on a per-bus basis, developing decentralized implementations, and including voltage regulators constitute interesting research directions.

REFERENCES

Fig. 1. Convergence for ADA: a) diesel generation; b) substation voltage $v_0$ (left y-axis) and energy exchange $p_{a0}^d$ (right y-axis); c) dual variables associated with average lower voltage limits for all buses; and d) dual variables associated with average upper voltage limits for all buses. Sliding averages of optimization variables are depicted too.

Fig. 2. Convergence for PDA: a) diesel generation; b) substation voltage $v_0$ (left y-axis) and energy exchange $p_{a0}^d$ (right y-axis); and c) dual variables associated with probabilistic constraint (left y-axis) and probability of under-/over-voltage (right y-axis). Sliding averages of optimization variables are depicted too.


Fig. 3. Per-bus probability of under-/over-voltages.

Fig. 4. Histograms of voltage magnitudes on buses 15 and 40 under ADA (top) and PDA (bottom). Dashed lines show regulation limits \( V_A \) and \( V_B \).

Fig. 5. Performance for ADA, PDA, the approximate, and the deterministic scheme.


