STOCHASTIC OPTIMIZATION OF POWER SYSTEMS WITH RISK CONSTRAINTS AND SPARSELY DISTRIBUTED STORAGE

Juan Andrés Bazerque (member).

Department of Electrical Engineering
Universidad de la República (UdelaR)
Montevideo, Uruguay
jbazerque@fing.edu.uy

ABSTRACT

The power grid is experiencing a profound transformation in recent years with the proliferation of renewable sources and the advance of bulk storage technologies. These changes have been anticipated in the literature of signal processing and control, where several new techniques for the optimal operation and design of the greed are reported. The present paper builds on the state of the art, proposing a novel stochastic approximation algorithm for optimizing the network under risk constraints. The method is capable of processing massive amounts of data, learning the distributions of the random generation and demand, and adapting to seasonal changes and system evolution. In addition, a parsimonious storage design is achieved by introducing a sparsifying penalty to the problem of siting and sizing. Numerical examples show that the optimization algorithm succeeds in guaranteeing that prescribed voltage limits are satisfied with an outage probability that stands below theoretical bounds.

***Index Terms*—** Optimal power flow, Stochastic approximation, Risk constraints.

1. INTRODUCTION

The ongoing transformation of the power grid poses new technical challenges for its optimal operation and design [1]. New forms of randomness are introduced by the uncertainty of renewable sources. The proliferating photovoltaic and wind farms are distributed geographically to capture the best wind and sun profiles, possibly out of reach of high-capacity power lines. Weaker lines translate into tighter optimization constraints, which call for sharper models of the grid. Additional research efforts are needed for the recent incorporation of short-term distributed storage systems, e.g., batteries, hydrogen cells, etc [2], [3]. By shifting energy across time, these new storage technologies relax the instantaneous constraint that balances generation and demand, thus alleviating the stress of the system during peak hours. On the other hand, modeling storage systems for power system optimization introduces time-evolving dynamics, which increases complexity.

Different models have been proposed for the AC power flow through the grid, trading-off accuracy for simplicity and tractability [4]. The simplest so-called DC power flow model is linear, but imposes the strongest assumptions, including constant voltage amplitudes and lossless lines. On the other side, the full AC model does not make these assumptions but it is nonconvex, although it admits convex relaxations that have been proved exact under milder assumptions [5]. In between these two models, linear alternatives have been proposed which contemplate voltage variation, losses, and reactive power [6].

As in [7], we adopt these linear models as constraints in the problem of minimizing the aggregated generation cost of a power system, and we extend the results for transmission networks with a slack bus. The randomness form renewables and demand is handled by setting risk constraints [7]. Such constraints can be seen as convex relaxations of probabilistic constraints [8], or as a way to control how large are the constraint violations when they occur [9].

The resulting risk-constrained optimization problem is tackled by a tailored projected dual stochastic approximation algorithm (SAA). A number of reasons make SAAs an attractive choice for the optimization of power networks [14]. They implicitly learn from data the underlying probability distributions of loads and sources, adjusting the uncertainty of forecasts when these are available. The recursive nature of SAAs allows one to incorporate one sample at a time, offering a practical alternative to Montecarlo averaging methods which stall when massive data is to be processed. Another reason to use SAAs resides on its online and adaptive processing, that generalizes the celebrated linear mean squares filer to nonlinear setups. With a constant stepsize SAAs can adapt to seasonal changes or to the evolving transformation of the grid.

Apart from a stochastic approximation treatment to the risk constraints, the paper introduces a sparsely distributed storage design. A fully distributed storage design can be obtained by setting storage units (SUs) at all buses and considering their capacity limits as optimization variables [10]. Several small SUs are introduced, whose installation and maintenance costs result impractical [16]. In order to avoid these residual SUs, the sparse design is proposed as as a half-way alternative to centralized bulk storage concentrated in a single node. The design capitalizes existing techniques in sparse signal process to select the most relevant locations for a scattered storage system to be installed at a reduced subset of nodes.

The specific time dependence of the storage variables has been exploited in [12] using techniques of dynamic programming and optimal control, under a linear-quadratic model that yields closed-form solutions for the backward induction updates [13]. Extending these techniques to our generalized setup is not addressed here, but is an opportunity for future research.

The presentation in this paper is incremental. A preliminary optimization approach is developed in section II, incorporating the risk constraints in section III, and storage design in section IV. The SAA is built in section V, to end with numerical examples and conclusions in sections VI and VII.

2. OPTIMAL POWER FLOW

Consider a system with $N+1$ buses modeled as the nodes of a graph, with edges representing the connecting power lines. Node $n=0$ is set as the slack bus, with a fuel-based generator and a load connected to it. All other N buses may connect a load, a renewable generator, a SU, and a fuel-based generator. The following vectors belong to $\mathbb{R}^N$ and represent either optimization variables or data pertaining to each of these N nodes. Entries of $\boldsymbol{\beta}$ represent the capacity limits of SUs at each node, and are treated as constant parameters until section IV. Entries of $\mathbf{x}^\tau$ and $\mathbf{p}_b^\tau$ denote the corresponding stored energy and power levels extracted from storage at time τ. Indexes $\tau=0,\ldots,T-1$ count time intervals of lenght ΔT. The whole length $T\Delta T$ is to be chosen to match the cyclostationary period of demand and renewable processes, typically a week. Vectors $\mathbf{p}_g^\tau$, $\mathbf{p}_l^\tau$, $\mathbf{p}_r^\tau$, $\mathbf{q}_g^\tau$, $\mathbf{q}_l^\tau$, and $\mathbf{q}_r^\tau$ represent the active and reactive power of fuel-generators, loads, and renewable sources, and $\mathbf{p}_n^\tau$, $\mathbf{q}_n^\tau$ stand for the net power injected to the network at each bus. Variables $\boldsymbol{\alpha}^\tau\in[0,1]^N$ describe the fraction of renewable power that is injected to the network after curtailment. Vector $\mathbf{v}^\tau$ collects the voltage *amplitudes* at the nodes. Finally, scalars p_{g0}^τ and q_{g0}^τ represent the power injected by the slack bus to maintain constant voltage. For notational brevity, the variables are collected in $\mathbf{u}^\tau:=(\mathbf{x}^\tau,\mathbf{v}^\tau,\boldsymbol{\alpha}^\tau,\mathbf{p}_b^\tau,\mathbf{p}_g^\tau,\mathbf{p}_n^\tau,\mathbf{q}_r^\tau,\mathbf{q}_g^\tau,\mathbf{q}_n^\tau,p_{g0}^\tau,q_{g0}^\tau)$. Under these definitions, and with the goal of minimizing operational costs, the following optimization problem is set

$$\min_{\{\mathbf{u}^\tau\}_{\tau=0}^{T-1}} \sum_{\tau=0}^{T-1} E\left[c^\tau(\mathbf{u}^\tau)\right] \tag{1}$$

$$\text{s. to: } \mathbf{x}^\tau=\mathbf{x}^{\tau-1}-\mathbf{p}_b^\tau \tag{2}$$

$$\mathbf{0}\le\mathbf{x}^\tau\le\boldsymbol{\beta} \tag{3}$$

$$\mathbf{p}_n^\tau:=\mathbf{p}_r^\tau\odot\boldsymbol{\alpha}^\tau+\mathbf{p}_g^\tau+\mathbf{p}_b^\tau-\mathbf{p}_l^\tau \tag{4}$$

$$\mathbf{q}_n^\tau=\mathbf{q}_r^\tau+\mathbf{q}_g^\tau-\mathbf{q}_l^\tau \tag{5}$$

$$\mathbf{v}_{\min}\le\mathbf{v}^\tau\le\mathbf{v}_{\max} \tag{6}$$

$$\mathbf{v}^\tau:=\mathbf{A}\mathbf{p}_n^\tau+\mathbf{B}\mathbf{q}_n^\tau+\bar{\mathbf{v}}_0 \tag{7}$$

$$p_{g0}^\tau=\bar{p}_0+\mathbf{a}\mathbf{p}_n^\tau-\mathbf{b}\mathbf{q}_n^\tau \tag{8}$$

$$q_{g0}^\tau=\bar{q}_0+\mathbf{a}\mathbf{q}_n^\tau+\mathbf{b}\mathbf{p}_n^\tau \tag{9}$$

$$\mathbf{0}\le\boldsymbol{\alpha}^\tau\le\mathbf{1} \tag{10}$$

$$(\alpha^\tau(n)p_n^\tau(n))^2+(q_r^\tau(n))^2\le(p_r^\tau(n))^2 \tag{11}$$

Constraints (2) and (3) control the charge-discharge dynamics of the SUs and set their capacity limits. Then, (4) and (5) balance the injected power at each bus according to Kirchhoff's node law. The following (6) sets limits for the voltage amplitudes. Equation (7) introduces a linear model for $\mathbf{v}^\tau$ in terms of the injected power levels; see [6] for the definition of constant $\mathbf{A},\mathbf{B}\in\mathbb{R}^{N\times N}$ and $\bar{\mathbf{v}}_0\in\mathbb{R}^N$ in terms of the network admittance matrix, and a discussion of the corresponding approximation errors. Next, (8) and (9) are the power levels that the slack bus needs to inject to the network to maintain constant voltage. More details on (8) and (9) are given in the next paragraph. Constraint (10) makes each entry of $\boldsymbol{\alpha}^\tau$ a fraction. Finally, (11) explains the capacity limits of renewable sources: by shifting the power angle it is possible to inject reactive power $q_r^\tau(n)$ into the grid in detriment of the active power level $\alpha(n)^\tau p_r(n)^\tau$. It is worth noticing that $\boldsymbol{\alpha}^\tau$ and $\mathbf{q}_r^\tau$ are control variables to optimize for, while the available power $\mathbf{p}_r^\tau$ is related to wind flow and solar radiation levels, and thus treated as given data.

Equations (8)-(9) are derived using the model in [6] for the vector of complex-valued voltages at the nodes. Then multiply by the network admittance matrix to find the current and power that the slack bus has to inject in order to keep a prescribed complex voltage, e.g, $V_0=1$. The following result is obtained: partitioning the network admittance matrix as in $[y_{00}\;\mathbf{y}^T;\;\mathbf{y}\;\mathbf{Y}]$ it results $\bar{p}_0+j\bar{q}_0=V_0(y_{00}+\mathbf{y}^H(\mathbf{Y}^{-1}\mathbf{y})^*)$ and $\mathbf{a}+j\mathbf{b}=\operatorname{diag}((\mathbf{Y}^{-1}\mathbf{y})^*)$ where $*$ stands for complex conjugate and $(\cdot)^H$ denotes Hermitian transposition.

Regarding (2), efficiency coefficients and power-rate limits for SUs can be incorporated without affecting the structure of the (1)-(11), but they are not included for simplicity. In particular, rate constraints couple the projection operators in section V.

Variables $\mathbf{v}^\tau,\ \mathbf{p}_b^\tau,\ \mathbf{p}_n^\tau,\ \mathbf{q}_n^\tau,\ p_{g0}^\tau$ and q_{g0}^τ are redundant in (1)-(11), so that our next step is to eliminate them keeping only those in a reduced $\tilde{\mathbf{u}}^\tau:=(\mathbf{x}^\tau,\boldsymbol{\alpha}^\tau,\mathbf{p}_g^\tau,\mathbf{q}_r^\tau,\mathbf{q}_g^\tau)$. As detailed next, we eliminate $\mathbf{v}^\tau$ by substituting (2), (4) and (5) in (7) and the result in (6)

$$\begin{aligned}
\mathbf{v}^\tau&=\mathbf{v}(\tilde{\mathbf{u}}^{\tau-1},\tilde{\mathbf{u}}^\tau):=\mathbf{A}\mathbf{p}_n^\tau+\mathbf{B}\mathbf{q}_n^\tau+\bar{\mathbf{v}}_0=\mathbf{A}\mathbf{p}_n^\tau+\mathbf{B}\mathbf{q}_n^\tau+\bar{\mathbf{v}}_0\\
&=\mathbf{A}\left(\mathbf{P}_r^\tau\boldsymbol{\alpha}^\tau+\mathbf{p}_g^\tau+\mathbf{p}_b^\tau-\mathbf{p}_l^\tau\right)+\mathbf{B}\left(\mathbf{q}_r^\tau+\mathbf{q}_g^\tau-\mathbf{q}_l^\tau\right)+\bar{\mathbf{v}}_0\\
&=\mathbf{A}\left(\mathbf{P}_r^\tau\boldsymbol{\alpha}^\tau+\mathbf{p}_g^\tau+\mathbf{x}^{\tau-1}-\mathbf{x}^\tau-\mathbf{p}_l^\tau\right)\\
&\quad+\mathbf{B}\left(\mathbf{q}_r^\tau+\mathbf{q}_g^\tau-\mathbf{q}_l^\tau\right)+\bar{\mathbf{v}}_0
\end{aligned}$$

Similarly we set $p_{g0}^\tau=p(\tilde{\mathbf{u}}^\tau)$ and $q_{g0}^\tau=q(\tilde{\mathbf{u}}^\tau)$ where $p(\tilde{\mathbf{u}}^\tau)$ and $q(\tilde{\mathbf{u}}^\tau)$ result from substituting (4) and (5) in (8) and (9), respectively. With a slight abuse of notation we redefine the cost as $c^\tau(\tilde{\mathbf{u}}^\tau)=c^\tau(\mathbf{u}^\tau)$ and recast (1)-(11) as

$$\min_{\{\mathbf{u}^\tau\}_{\tau=0}^{T-1}} \sum_{\tau=0}^{T-1} E\left[c^\tau(\tilde{\mathbf{u}}^\tau)\right] \tag{12}$$

$$\text{s. to: } \mathbf{0}\le\mathbf{x}^\tau\le\boldsymbol{\beta} \tag{13}$$

$$\mathbf{0}\le\boldsymbol{\alpha}^\tau\le\mathbf{1} \tag{14}$$

$$\bar{\mathbf{v}}_{\mathbf{min}}\le\mathbf{v}(\tilde{\mathbf{u}}^{\tau-1},\tilde{\mathbf{u}}^\tau)-\bar{\mathbf{v}}_{\mathbf{max}}]\le 0 \tag{15}$$

$$\mathbf{g}_r\left(\boldsymbol{\alpha}^\tau,\mathbf{p}_r^\tau,\mathbf{q}_r^\tau\right)\le 0. \tag{16}$$

The vector-valued $\mathbf{g}_r\left(\boldsymbol{\alpha}^\tau,\mathbf{p}_r^\tau,\mathbf{q}_r^\tau\right)\in\mathbb{R}^N$ collects scalar functions $g_r(\alpha,p,q)=(\alpha p)^2+q^2-p^2$, so that $g_r(\alpha^\tau(n),p_r^\tau(n),q_r^\tau(n))\le 0$ substitutes (11).

Inequalities (15) and (16) may be too restrictive since they have to be satisfied for all realizations of the random variables in play. A probabilistic alternative is introduced in [7] and presented in next section, where a prescribed outage probability is admitted in exchange for a lower average cost.

3. RISK CONSTRAINTS

If (15) or (16) are not satisfied, then either the grid will operate without meeting quality standards or the protection devices will activate causing the system outage. None of these situations are desirable, and the optimization scheme should be designed such that they occur with low probability. Prescribing an admissible outage probability ϵ and focusing on (16), we relax the worst case restriction $g_r\le 0$ for its probabilistic counterpart $P(g_r\le 0)\ge 1-\epsilon$ [7]. As depicted in Fig. 1, this inequality is equivalent to $z_{1-\epsilon}\le 0$ where the percentile $z_{1-\epsilon}$ is defined as the value such that $P(g_r\le z_{1-\epsilon})=1-\epsilon$. In order to obtain a convex relaxation of $z_{1-\epsilon}\le 0$ we follow [7] and [8] and resort to the *conditional value at risk* (CVAR). The CVAR is defined as the conditional mean of g_r above the percentile; i.e., $\text{Cvar}_{(1-\epsilon)}(g_r):=E(g_r|g_r>z_{1-\epsilon})$. It readily follows from its definition that $\mathrm{Cvar}_{(1-\epsilon)} > z_{1-\epsilon}$, since values lower than or equal to $z_{1-\epsilon}$ are not averaged. Hence, enforcing $\mathrm{Cvar}_{(1-\epsilon)} \leq 0$ guarantees $z_{1-\epsilon} \leq 0$; see Fig. 1. One reason why CVAR is worth introducing is that it admits the convex equivalent expression $\mathrm{Cvar}_{(1-\epsilon)}(g_r) = \inf_{z\in\mathbb{R}} \left\{\frac{1}{\epsilon}E[[g_r + z]_+ - z\right\}$ [9], where $[\cdot]_+$ stands for the projection onto the nonnegative real numbers. The previous equality implies that if a scalar z exists such that $\frac{1}{\epsilon}E[[g_r + z]_+ - z \leq 0$, then the infimum will be negative or equal to zero, that is, $\mathrm{Cvar}_{(1-\epsilon)} \leq 0$. This also implies that the percentile $z_{1-\epsilon} \leq 0$. Hence, it was established that

$$\frac{1}{\epsilon}E[[g_r + z]_+ - z \leq 0 \Rightarrow P(g_r \leq 0) \geq 1 - \epsilon. \qquad (17)$$

Even though the CVAR was utilized as a convex relaxation for probabilistic constraints, the constraint $\mathrm{Cvar}_{(1-\epsilon)}(g_r) \leq 0$ could have been introduced directly as a way to control the risk of a large deviations from the levels of normal operation [9]. Building on the previous discussion we substitute (17) for the worst-case constraints in (12)-(16) yielding

$$\min_{\{\mathbf{u}^\tau\}_{\tau=0}^{T-1}} \sum_{\tau=0}^{T-1} E\left[c^\tau(\tilde{\mathbf{u}}^\tau)\right] \qquad (18)$$

$$\text{s. to: } \mathbf{0} \leq \mathbf{x}^\tau \leq \boldsymbol{\beta} \qquad (19)$$

$$\mathbf{0} \leq \boldsymbol{\alpha}^\tau \leq \mathbf{1} \qquad (20)$$

$$\frac{1}{\epsilon}E\left[\mathbf{v}(\tilde{\mathbf{u}}^\tau, \tilde{\mathbf{u}}^{\tau-1}) - \bar{\mathbf{v}}_{\mathbf{max}} + \mathbf{z}_M^\tau\right]_+ - \mathbf{z}_M^\tau \leq 0 \qquad (21)$$

$$\frac{1}{\epsilon}E\left[\bar{\mathbf{v}}_{\mathbf{min}} - \mathbf{v}(\tilde{\mathbf{u}}^\tau, \tilde{\mathbf{u}}^{\tau-1}) + \mathbf{z}_m^\tau\right]_+ - \mathbf{z}_m^\tau \leq 0 \qquad (22)$$

$$\frac{1}{\epsilon}E\left[\mathbf{g}_r\left(\boldsymbol{\alpha}^\tau, \mathbf{p}_r^\tau, \mathbf{q}_r^\tau\right) + \mathbf{z}_r^\tau\right]_+ - \mathbf{z}_r^\tau \leq 0 \qquad (23)$$

In the next section we advance to storage design

4. DISTRIBUTED STORAGE

To optimize the siting and sizing of the SUs, $\boldsymbol{\beta}$ is incorporated as optimization variable [11], [10], [12]

$$\min_{\{\mathbf{u}^\tau\}_{\tau=0}^{T-1}, \boldsymbol{\beta}} \sum_{\tau=0}^{T-1} E\left[c^\tau(\tilde{\mathbf{u}}^\tau, \boldsymbol{\beta})\right] + \lambda\|\boldsymbol{\beta}\|_1 \qquad (24)$$

$$\text{s. to: } \mathbf{0} \leq \mathbf{x}^\tau \leq \boldsymbol{\beta} \qquad (25)$$

$$\mathbf{0} \leq \boldsymbol{\alpha}^\tau \leq \mathbf{1} \qquad (26)$$

under the additional constraints (21)-(23), and with $c^\tau(\tilde{\mathbf{u}}^\tau, \boldsymbol{\beta})$ redefined to incorporate the cost of storage.

The extra term in (24) is introduced to control the number of nodes where a SU will be set. Indeed, $\|\boldsymbol{\beta}\|_1$ has been proved to be a convex surrogate for the nonconvex pseudo-norm $\|\boldsymbol{\beta}\|_0$ which counts the number of nonzero entries of $\boldsymbol{\beta}$ [21]. By penalizing $\|\boldsymbol{\beta}\|_0$ we avoid the installation and maintenance costs of small-capacity SUs corresponding to entries $\beta(b) \simeq 0$. Selecting $\lambda = 0$ the solution of (24)-(23) yields a fully distributed design, and increasing λ progressively only the most relevant sites are retained up to a point where storage costs exceed generation costs and storage is discarded as a whole.

Although the expected values in (24) and (21)-(23) could be approximated by sampling averages, that may introduce a prohibitive numerical cost since each constraint would change into a sum over the samples. To avoid such a curse of dimensionality we resort to the recursive method of next section.

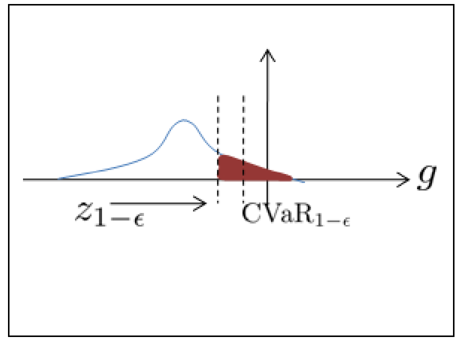

Fig. 1. Conditional value at risk as a relaxation of probabilistic constraints

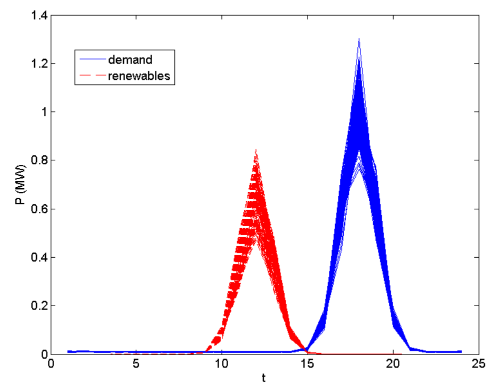

Fig. 2. Random data simulating renewables and demand.

5. PROJECTED STOCHASTIC GRADIENT ALGORITHM

In order to minimize (24), we resort to a dual method in which the risk constraints (21)-(23) are dualized and (25)-(26) are kept in the constraint set. Upon grouping primary variables in $\mathbf{y} = (\{\tilde{\mathbf{u}}^\tau, \mathbf{z}^\tau\}, \boldsymbol{\beta})$ $= (\{\mathbf{x}^\tau, \boldsymbol{\alpha}^\tau, \mathbf{p}_g^\tau, \mathbf{q}_r^\tau, \mathbf{q}_g^\tau, \mathbf{z}_M^\tau, \mathbf{z}_m^\tau, \mathbf{z}_g^\tau, \}, \boldsymbol{\beta})$ we write

$$\begin{aligned} L_e(\mathbf{y}) = \lambda\|\boldsymbol{\beta}\|_1 + \sum_{\tau=0}^{T-1} \Bigg\{ & \tilde{c}^\tau(\tilde{\mathbf{u}}^\tau, \boldsymbol{\beta}) \\ & + (\boldsymbol{\mu}_M^\tau)' \left(\frac{1}{\epsilon}\left[\mathbf{v}(\tilde{\mathbf{u}}^\tau, \tilde{\mathbf{u}}^{\tau-1}) - \bar{\mathbf{v}}_{\mathbf{max}} + \mathbf{z}_M^\tau\right]_+ - \mathbf{z}_M^\tau\right) \\ & + (\boldsymbol{\mu}_m^\tau)' \left(\frac{1}{\epsilon}\left[\bar{\mathbf{v}}_{\mathbf{min}} - \mathbf{v}(\tilde{\mathbf{u}}^\tau, \tilde{\mathbf{u}}^{\tau-1}) + \mathbf{z}_m^\tau\right]_+ - \mathbf{z}_m^\tau\right) \\ & + (\boldsymbol{\mu}_g^\tau)' \left(\frac{1}{\epsilon}\left[\mathbf{g}_r\left(\boldsymbol{\alpha}^\tau, \mathbf{p}_r^\tau, \mathbf{q}_r^\tau\right) + \mathbf{z}_g^\tau\right]_+ - \mathbf{z}_g^\tau\right) \Bigg\} \end{aligned} \qquad (27)$$

Considering the founding theory of Robbins-Monro [22], we wrote a stochastic version of the Lagrangian by dropping the expected values and keeping samples that depend on the realizations of $\mathbf{p}_r^\tau$, $\mathbf{p}_l^\tau$, and $\mathbf{q}_l^\tau$ (recall that these are parameters hidden in functions $c(\cdot)$, $\mathbf{v}(\cdot)$, and $\mathbf{g}_r(\cdot)$). Building on (27) we propose an online stochastic primal-dual projected-subgradient descent iteration.
For each period k of length $T\Delta T$ we collect samples
$\mathbf{w}(k) = \{\mathbf{p}_r^\tau, \mathbf{p}_l^\tau, \mathbf{q}_l^\tau, \tau = kT\Delta T, \ldots, (k+1)T\Delta T - 1\}$. Then we update $\mathbf{y}$ and the multipliers according to (28)-(31), and wait for a next set of data $\mathbf{w}(k+1)$ in order to perform the next update. These

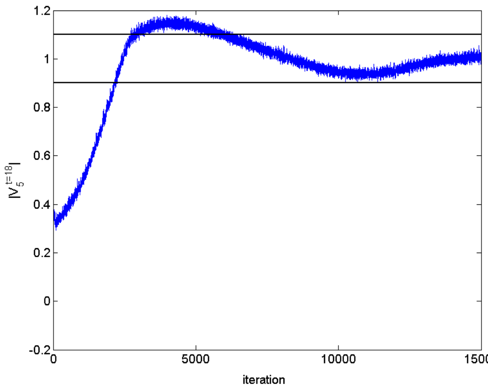

Fig. 3. Evolution of the solution across iterations

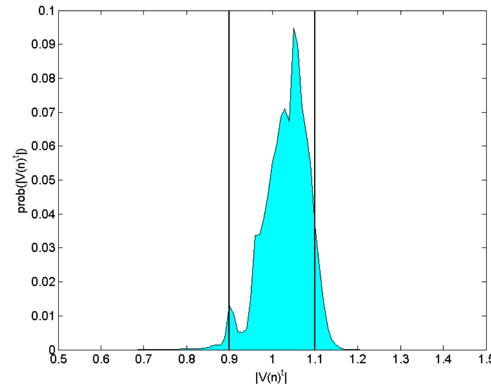

Fig. 4. Histogram of voltage amplitudes

updates are given by

$$\mathbf{y}(k+1) = \Pi_X \left[\mathbf{y}(k) - \gamma \nabla_{\mathbf{y}} L_e(\mathbf{y}(k), \boldsymbol{\mu}(k); \mathbf{w}(k))\right] \quad (28)$$

$$\boldsymbol{\mu}_M^\tau(k+1) = \boldsymbol{\mu}_M^\tau(k) + \gamma \left(\frac{1}{\epsilon} \left[\mathbf{g}_M + \mathbf{z}_M^\tau\right]_+ - \mathbf{z}_M^\tau \right) \quad (29)$$

$$\boldsymbol{\mu}_m^\tau(k+1) = \boldsymbol{\mu}_m^\tau(k) + \gamma \left(\frac{1}{\epsilon} \left[\mathbf{g}_m + \mathbf{z}_M^\tau\right]_+ - \mathbf{z}_m^\tau \right) \quad (30)$$

$$\boldsymbol{\mu}_g^\tau(k+1) = \boldsymbol{\mu}_g^\tau(k) + \gamma \left(\frac{1}{\epsilon} \left[\mathbf{g}_r + \mathbf{z}_g^\tau\right]_+ - \mathbf{z}_g^\tau \right) \quad (31)$$

where $\nabla_{\mathbf{y}}$ stands for the subgradient operator, $\mathbf{g}_M := \mathbf{v}(\mathbf{u}^\tau, \mathbf{u}^{\tau-1}) - \bar{\mathbf{v}}_{\mathbf{max}}$, and $\mathbf{g}_m := \bar{\mathbf{v}}_{\mathbf{min}} - \mathbf{v}(\mathbf{u}^\tau, \mathbf{u}^{\tau-1})$. The operator $\Pi_X(\cdot)$ projects $\mathbf{x}^\tau$, $\mathbf{p}_g^\tau$, $\boldsymbol{\alpha}^\tau$, and $\boldsymbol{\beta}$ into the convex set $X = \{\mathbf{0} \leq \boldsymbol{\alpha}^\tau \leq 1, \mathbf{0} \leq \mathbf{x}^\tau \leq \boldsymbol{\beta}, \mathbf{p_g} \geq \mathbf{0}\}$ which collects the constraints that where not dualized. Projections $\mathbf{p}_g^\tau \geq 0$, $\boldsymbol{\alpha}^\tau \in [0,1]$ and $\mathbf{x}^\tau \geq 0$ are trivial, while $\mathbf{x}^\tau \leq \boldsymbol{\beta}$ are coupled across τ. However, the latter admits the following closed-form solution

1) Project each $\mathbf{x}^\tau$ into $\mathbf{x}^\tau \geq 0$ and sort the inputs $\mathbf{x}^\tau$ and $\boldsymbol{\beta}$ from larger to smaller.
2) For $t = 0, 1, 2, \ldots$, compute the cumulative sample means $\sigma_\tau = (\boldsymbol{\beta} + \sum_{\tau=0}^{t} x^\tau)/(t+1)$, stopping at the first $t^\star$ such that $\sigma_{t^\star} \geq \mathbf{x}_{t_\star+1}$, or set $t^\star = T-1$ otherwise.
3) Set $\boldsymbol{\beta} = \mathbf{x}^\tau = [\sigma_{t^\star}]_+$ for $\tau \leq t^\star$ and $\mathbf{x}^\tau = \mathbf{x}^\tau$ for $\tau > t^\star$.

Convergence of dual SAAs with inequality constraints and projection operators is addressed in [17] and [18]. The dual SAA in (28)-(31) is to be run in two stages. In the first offline design stage it runs on past record data in order to find the optimal storage distribution $\boldsymbol{\beta}$. In a second operational stage $\boldsymbol{\beta}$ is not updated and the iteration is run with a constant step size for adaptive control.

6. NUMERICAL EXAMPLES

The SAA developed in last section is tested in the reduced New England IEEE 39 bus system, with bus $n = 39$ selected as the slack bus. Buses 30-39 are connected to a local fuel-based generator.

The cost is selected as $c(\tilde{\mathbf{u}}^\tau, \boldsymbol{\beta}) = c_b\|\boldsymbol{\beta}\|_1 + c_p\left(\|\mathbf{p}_g^\tau\|_1 + p_{g0}\right)$ $+c_q\left(\|\mathbf{q}_g^\tau\|_1 + |q_{g0}|\right)$, with $c_p = c_q = 1$, and $c_b = 0.01$. Loads are simulated as random processes with time dependent mean and variance. As it is shown in Fig. 2, the mean is constructed as a Gaussian bell with peak at $\tau = 17$, and sampled every hour from

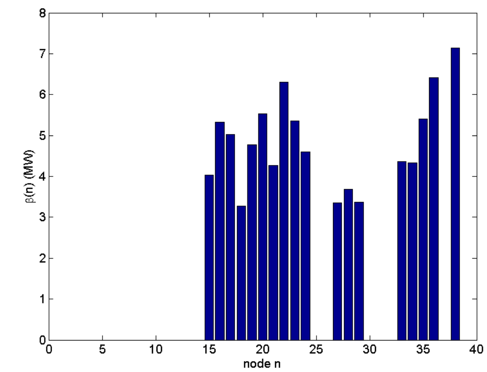

Fig. 5. Sparsely designed storage

$\tau = 0$ to $\tau = T - 1 = 23$. The standard deviation of the additive white Gaussian noise is proportional to the mean at 10%. Every day a new set of 24 samples is acquired to run one step of the SAA in section V. Buses 1, 2, 5, 6, 9, 10, 11, 13, 14, 17, and 19 are fed by renewable sources simulated as random processes with the same structure as the loads but shifted in time with average peak at $\tau = 12$.

Fig. (3) shows the evolution of the solution across SAA iterations when run with a constant stepsize $\gamma = 10^{-3}$. The blue marks depict the voltage of node $n = 5$ and time $\tau = 18$, which enters the prescribed accepted limits after a transient period. After the transient has elapsed, voltages across all n and τ are distributed according to Fig 4, and outage occurs with a frequency of 0.0618, staying below the theoretical bound $\epsilon = 0.1$. Finally Fig. 5 shows that when $\lambda = \frac{1}{T}\sum_{\tau=1}^{T-1} E[\mathbf{p}_l^\tau]$ only 15 SUs are kept in the system, achieving the intended sparse design.

7. CONCLUSIONS

The optimal grid operation and design was modeled as a stochastic optimization problem with risk constraints and storage dynamics. The setup includes a linear model that relates voltages and injected power levels, with the inclusion of a slack bus. A semi-distributed storage layout is designed by adding the storage unit capacities as optimization variables and imposing a sparsifying penalty. A stochastic dual algorithm is derived and simulated, satisfying the constraints within an outage probability that stays below theoretical bounds.

8. REFERENCES

[1] G. B. Giannakis, V. Kekatos, N. Gatsis, S.J. Kim, H. Zhu, and B. F. Wollenberg, “Monitoring and optimization for power grids: A signal processing perspective,” *IEEE Signal Processing Magazine,* vol. 30, no. 5, pp. 107-128, 2013.

[2] B. Dunn, H. Kamath, and J.-M. Tarascon, “Electrical energy storage for the grid: a battery of choices,” *Science* vol. 334, no. 6058, pp. 928-935, 2011.

[3] O. Palizban and K. Kauhaniemi, “Energy storage systems in modern grids - matrix of technologies and applications,” *Journal of Energy Storage,* pp. 248-259, 2016.

[4] S. H. Low “Convex relaxation of optimal power flowPart I: Formulations and equivalence,” *IEEE Transactions on Control of Network Systems,* vol 1, no.1 pp. 15-27, 2014.

[5] S.H. Low, “Convex relaxation of optimal power flowPart II: Exactness,” *IEEE Transactions on Control of Network Systems,* vol 1, no. 2 pp. 177-189, 2014.

[6] S. Bolognani, and S. Zampieri, “On the existence and linear approximation of the power flow solution in power distribution networks,” *IEEE Transactions on Power Systems,*vol. 31, no. 1, pp. 163-172, 2016 .

[7] E. Dall’Anese, K. Baker, and T. Summers, “Chance-constrained AC optimal power flow for distribution systems with renewables,” *IEEE Transactions on Power Systems,* vol. 32, no.5, pp. 3427 - 3438, 2017.

[8] A. Nemirovski, and A. Shapiro, “Convex approximations of chance constrained programs,”*SIAM J. on Optimization,*vol. 17, no. 4, pp. 969-996, 2006.

[9] R. T. Rockafellar, and S. Uryasev “Optimization of conditional value-at-risk,”*Journal of risk*, vol. 2, pp. 21-42, 2000.

[10] Baker, Hug, Li, “Energy Storage Sizing Taking into Account Forecast Uncertainties and Receding Horizon Operation.” *IEEE Transactions on Sustainable Energy,* vol. 8, no. 1, pp. 331-340, 2017.

[11] S. Bose, D. Gayme, O. Topcu, M. Chandy, “Optimal placement of energy storage in the grid”, *IEEE Conference on Decision and Control,* Maui, HI, 2012.

[12] J. Qin and R. Rajagopal, “Dynamic Programming Solution to Distributed Storage Operation and Design,” *IEEE Power and Energy Society General Meeting,* pp. 1-5, 2013.

[13] D. P.Bertsekas, *Dynamic programming and optimal control,* Vol. 1, 3th Ed., Belmont, MA: Athena Scientific, 2005.

[14] T. Chen, A. G. Marqus, and G. B. Giannakis, “DGLB: Distributed Stochastic Geographical Load Balancing with Incentive Payment,”*IEEE Transactions on Parallel and Distributed Systems,* vol. 28, no. 7, pp. 1866-1880, 2017.

[15] M. S. Whittingham, “History, evolution, and future status of energy storage,” *Proceedings of the IEEE 100, Special Centennial Issue,* pp. 1518-1534, 2012.

[16] S. van der Linden, “Bulk energy storage potential in the USA, current developments and future prospects,” *Energy,* vol. 31, no. 15, pp. 3446-3457, 2006.

[17] H. J. Kushner and E. Sanvicente “Stochastic approximation of constrained systems with system and constraint noise,” *Automatica,* vol. 11, no. 4, pp.375-380, 1975.

[18] S. S. Ram, A. Nedic, and V. Veeravalli, “Distributed stochastic subgradient projection algorithms for convex optimization,” *Journal of optimization theory and applications,* vol. 147, no. 3, pp.516-45, 2010.

[19] B. Efron, T. Hastie, I. M. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics,* vol. 32, pp. 407499, 2004.

[20] H. Zou, “The adaptive lasso and its oracle properties,” *J. Amer. Stat. Assoc.,* vol. 101, pp. 14181429, 2006.

[21] E. J. Candès and Y. Plan, “Near-ideal model selection by 1 minimization,” *Annals of Statistics,* vol. 37, no. 5A, pp. 2145-2177, 2009.

[22] H. Robbins, S. Monro “A stochastic approximation method,” *Annals of Math. Statistics*, vol. 22, pp. 400407, 1951.