Power Grid Reliability Estimation via Adaptive Importance Sampling

Aleksander Lukashevich, *Student Member, IEEE*, and Yury Maximov, *Member, IEEE*

arXiv:2105.08753v1 [math.OC] 18 May 2021

***Abstract*—Electricity production currently generates approximately 25% of greenhouse gas emissions in the USA. Thus, increasing the amount of renewable energy is a key step to carbon neutrality. However, integrating a large amount of fluctuating renewable generation is a significant challenge for power grid operating and planning. Grid reliability, i.e. an ability to meet operational constraints under power fluctuations, is probably the most important of them. In this paper, we propose computationally efficient and accurate methods to estimate the probability of failure, i.e. reliability constraints violation, under a known distribution of renewable energy generation. To this end, we investigate an importance sampling approach, a flexible extension of Monte-Carlo methods, which adaptively changes the sampling distribution to generate more samples near the reliability boundary. The approach allows to estimate failure probability in real-time based only on a few dozens of random samples, compared to thousands required by the plain Monte-Carlo. Our study focuses on high voltage direct current power transmission grids with linear reliability constraints on power injections and line currents. We propose a novel theoretically justified physics-informed adaptive importance sampling algorithm and compare its performance to state-of-the-art methods on multiple IEEE power grid test cases.**

***Index Terms*—power system security, power system control, power system faults, sampling methods, fluctuations**

I. Introduction

Carbon-free electricity generation is one of the most vital global challenges for the next decades. Because of their ecological and economic benefits, renewable energy sources, such as wind, hydro, and solar power generation, become more demanded, accessible, and widely used in modern power grids [1], [2]. For instance, California's renewable portfolio standard currently requires 33% of retail electricity sales to come from renewable resources, and will require 60% by 2030, and 100% by 2045 [3]. However, renewable energy generation is highly volatile and brings significant uncertainty to power systems. This also gives rise to many challenges for power system operators trying to integrate renewables into power grids [4], [5]. In particular, power systems operational policies and reliability assessment must be verified over various additional uncertainties, including increased variation in power generation and disturbances.

Aleksander Lukashevich is with the Center for Energy Science and Technology, Skolkovo Institute of Science and Technology, Moscow, Russia (e-mail: Aleksandr.Lukashevich@skoltech.ru)

Yury Maximov is with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA (e-mail: yury@lanl.gov)

Various algorithms have been developed so far for ensuring grid reliability. Some of them are based on machine learning, and utilize historical data about weather, renewables' generation, and grid operating parameters to estimate the risk of failure and influence of uncertainty due to whether changes [6], [7]. Requiring large datasets and high data collecting time to make an accurate prediction, machine learning methods become impractical for real-time operation if a large disturbance, contingency, or a sudden operational policy change precedes the reliability assessment. Another class of algorithms is based on analytical approximation of the failure probability [8]. In this approach, the risk of interest is upper-bounded by an integral of an appropriate function that admits analytical or numerical computation. However, even in the simplest case of linear reliability constraints and Gaussian fluctuations of renewables, existing approaches tend to overestimate the risk. Moreover, for a sufficiently rare event, the risk overestimation for these algorithms can be infinitely high [8] which compromises their practical efficiency.

Finally, algorithms based on sampling values of power generated by renewables and approximating the grid failure probability by its empirical counterpart often provide a valuable alternative for accessing the reliability posture of a power system. Monte-Carlo (MC), hybrid and Markov Chain MC have been earlier applied to risk-based reliability assessment of transmission power grids [9]–[12]. In [13], the authors exploited Monte Carlo simulation for estimating failure probability and interpreted the risk by classifying it into low, medium and high risk operating points. Variations of the load-flow solution due to renewables fluctuation, nodal and line parameters uncertainty were considered in [9], [11]. An inverse problem of wind turbine controls to meet reliability margins with high-probability is discussed in [10]. A comprehensive survey of sampling-based methods for power systems reliability assessment is given in [12]. Unfortunately, these algorithms explore the space of fluctuations uniformly, which dramatically reduces their performance in understanding and evaluating the effect of a rare event such as a severe disturbance.

Importance sampling is a valuable alternative to Monte-Carlo sampling, which allows adjusting distribution for generating more samples in the areas of interest, e.g., close to the reliability boundary. Pmvnorm [14] is one of the most efficient importance sampling algorithms in general, but its performance is often limited for rare events probability estimation [15], which is of the utmost importance to power systems study. ALOE [15] is another efficient method designed especially for computing a rare event probability. However, it

does not fully respect the geometry of reliability constraints. It thus requires a large number of samples to estimate the risk of failure, especially for large power grids and multi-line failures. Finally, a convex optimization-based algorithm for adaptive importance sampling from exponential families was proposed in [16]. At each step, the algorithm adjusts the distribution parameters so that the sampler's variance is minimized. Unfortunately, the distribution of output power that leads to a failure is far from the exponential family, limiting the algorithm's efficiency in power systems.

This paper proposes an adaptive importance sampling method to efficiently estimate the risk of reliability constraints violation. We present an importance sampling algorithm that uses physical information to generate a mixture of distributions to sample from and then uses convex optimization to iteratively adjust the weights of the mixture. Our algorithm substantially improves static weights assignment of ALOE [15] when reliability constraints are highly correlated. The approach allows to address the risk estimation problem in real-time even for large power grids with a small failure probability. We theoretically analyze the accuracy and complexity of our algorithm for the case of Gaussian power fluctuations from renewables; however, the technique is not limited to the Gaussian case. Finally, we evaluate the performance of our sampling methods over multiple real and synthetic test cases and compare it to the state-of-the-art.

The paper is organized as follows. In Section II we present the failure probability estimation problem and introduce notation used in the paper. We outline the importance sampling algorithm and present its theoretical analysis in Section III. Empirical study and comparison to the state-of-the-art are given in Section IV. In Section V we conclude with a brief summary and discussion on possible applications of our results.

II. Background and Problem Setup

Being a popular load flow model, the higher-voltage direct current (DC) model remains simple for the analysis because of linear relations between power injections and phase angles. Let $G = (V, E)$ be a power grid graph with a set of buses V, $|V| = n$ and a set of lines E, $|E| = m$. Let $p \in \mathbb{R}^n$ and $\theta \in \mathbb{R}^n$ be vectors of power injections and phase angles respectively. The power system is balanced, e.g., the sum of all power injections equals zero $\sum_{i\in V} p_i = 0$. To avoid ambiguity let s be the slack bus and $\theta_s = 0$. Let $B \in \mathbb{R}^{n\times n}$ be an admittance matrix of the system, $p = B\theta$. The components B_{ij} are such that $B_{ij} \neq 0$ if there is a line between buses i and j and for any node $B_{ii} = -\sum_{j\neq i} B_{ij}$, e.g., B is a Laplacian matrix. Let $B^\dagger$ be the pseudo-inverse of B, $\theta = B^\dagger p$. The DC power flow equations, generation and reliability constraints are then

$$p = B\theta \tag{1a}$$

$$\underline{p}_i \leq p_i \leq \overline{p}_i, i \in V \text{ and } |\theta_i - \theta_j| \leq \bar{\theta}_{ij},\ (i,j) \in E \tag{1b}$$

Reliability constraints (1b) define a polytope P in the space of power injections, $P = \{p : Wp \leq b\}$, so that the reliability constraints are violated if and only if power injections $p \notin P$.

To derive an explicit expression of matrix W we consider the incidence matrix A, such that for any buses i and j with $i < j$ connected by an edge k, $A_{ki} = 1$, $A_{kj} = -1$ and all other elements in row k are equal to zero. Then the phase angle constraints are $AB^\dagger p \leq \bar{\theta}$, $-AB^\dagger p \leq \bar{\theta}$. Finally, as the slack bus balances the system, $p_s = -\sum_{i\neq s} p_i$ let $C \in \mathbb{R}^{n\times n}$ be a symmetric matrix such that for any non-slack buses i and j, and the slack bus s, $C_{ii} = 1$, $C_{ij} = 0$, $C_{ss} = 0$, and $C_{is} = -1$. In other words, Cp is a vector of grid power injections expressed in terms of non-slack injections only, since the slack bus power injection is fully determined by the other ones.

Finally, from Eqs. (1) the following system of inequalities defines the reliability polytope, $P = \{p : Wp \leq b\}$,

$$(AB^\dagger C, -AB^\dagger C, C, -C)^\top p \leq (\bar{\theta}, \bar{\theta}, \overline{p}, \underline{p})^\top, \tag{2}$$

where $W = (AB^\dagger C, -AB^\dagger C, C, -C)^\top$ and $b = (\bar{\theta}, \bar{\theta}, \overline{p}, \underline{p})^\top$. Let $J = 2m+2n$ be a number of constraints, e.g., rows in matrix W, then the reliability polytope P is $\left\{p : \bigcap_{i=1}^J p^\top \omega^i \leq b_i\right\}$.

Stochastic uncertainty in renewable generation and power consumption imposes a question of power grid reliability, e.g., estimating a probability that at least one of the reliability constraints is violated. Namely, we consider Gaussian fluctuations of power injections p with known mean μ and covariance Σ and aim at computing a failure probability Π:

$$\Pi = \mathbb{P}(p \notin P) = \int_{\mathbb{R}^n} \upsilon(p) 1[p \notin P] dp,\ p \sim \mathcal{N}(\mu, \Sigma), \tag{3}$$

where $\mathbb{P}$ is a probability taken w.r.t. the normal pdf $\upsilon(p)$ of p.

Notice, that the probability Π does not have an analytic expression, is computationally intractable, and even hard to approximate [15]–[18]. In practice, a union bound is often used to upper bound Π. Let Π_i be a probability of a single event, e.g., $p^\top \omega^i > b_i$. It has an explicit expression for the Gaussian distribution, and by union bound inequality $\sum_{i\leq J} \Pi_i \geq \Pi \geq \max_{i\leq J} \Pi_i$; however, the bounds are loose when dealing with correlated failures which is often the case for power systems.

To refine the failure estimate and take into account simultaneous violation of multiple constraints, we propose an importance sampling procedure that allows to count the average number of constraints N violated at the same time and, thus, improve the failure probability estimation to Π/N instead of Π. It is meaningful for large power grids where multiple events are likely to happen synchronously.

Table I summarizes paper's notation. We use lower indices for elements of vectors and matrices, lower-case letters for probability density functions (pdfs), and upper-case letters for cumulative distribution functions (cdfs). When it does not lead to confusion, we use $\mathbb{P}$, $\mathbb{E}$, and $\mathbb{V}$ to denote probability, expectation, and variance without explicitly mentioning a distribution.

TABLE I
PAPER NOTATION.

E	set of lines, $\|E\| = m$	$\upsilon(p)$	nominal distribution pdf
V	set of buses, $\|V\| = n$	υ_D	parametric distribution pdf
B	$n \times n$ admittance matrix	x	mixture distribution parameters, $x \in X \subseteq \mathbb{R}^J$
p_i	power injection		
$\underline{p}_i$	lower generation limit	D_i	$p \sim \mathcal{N}(\mu,\Sigma)$ s.t. $p^\top\omega^i > b_i$
$\overline{p}_i$	upper generation limit	N	number of samples
θ_i	phase angle	$\mathbb{P}$, $\mathbb{E}$	probability, expectation
θ_{ij}	phase angle difference	$\mathbb{V}$, KL	Variance, KL-divergence
$\bar{\theta}_{ij}$	angle difference limits	$\mathcal{N}(\mu,\Sigma)$	Gaussian distribution with mean μ and covariance Σ
I_n	$n \times n$ identity matrix		
J	number of constraints	Φ	$\mathcal{N}(0,1)$ distribution cdf
P	reliability set, $p: Wp \leq b$	$U(0,1)$	uniform $(0,1)$ distribution
ω^i	rows of matrix W, $i \leq J$	Π	failure probability, $p \notin P$

III. Failure Probability Estimation

A. A Single Constraint Case

We start with estimating the probability of fluctuating power injections to cause a failure of an individual constraint, e.g., $\Pi_i = \mathbb{P}(p^\top\omega^i \geq b_i)$ for some i, $1 \leq i \leq J$. In the case of a Gaussian distribution, $p \sim \mathcal{N}(\mu, \Sigma)$ there is a closed form expression for it:

$$\begin{aligned}\Pi_i &= \mathbb{P}_{p\sim\mathcal{N}(\mu,\Sigma)}(p^\top\omega^i \geq b_i) = \mathbb{P}((p-\mu)^\top\omega^i \geq b_i - \mu^\top\omega^i)\\ &= \mathbb{P}_{\tilde{p}\sim\mathcal{N}(0,I_n)}\left((\Sigma^{1/2}\omega^i)^\top\tilde{p} \geq b_i - \mu^\top\omega^i\right)\\ &= \mathbb{P}_{\tilde{p}\sim\mathcal{N}(0,I_n)}\left(\tilde{p}^\top\bar{\omega}_i \geq (b_i - \mu^\top\omega^i)/\|\Sigma^{1/2}\omega^i\|_2\right)\\ &= \Phi((b_i - \mu^\top\omega^i)/\|\Sigma^{1/2}\omega^i\|_2),\end{aligned}$$

where Φ is the standard Gaussian distribution cdf, $\tilde{p} = \Sigma^{-1/2}(p-\mu)$, $\bar{\omega}_i = (\Sigma^{1/2}\omega^i)^\top/\|\Sigma^{1/2}\omega^i\|_2$, and $1 \leq i \leq J$. Π_i are the probabilities of $p^\top\omega^i \geq b_i$, so that $\Pi = \mathbb{P}(\exists i : p^\top\omega^i \geq b_i) \leq \sum_{i\leq J}\mathbb{P}(p^\top\omega^i \geq b_i) = \sum_{i\leq J}\Pi_i$.

Algorithm 1 is an instance of the inverse transorm method [19] which allow to sample $p \sim \mathcal{N}(\mu,\Sigma)$ s.t. $p^\top\omega^i > b_i$. We refer this distribution as D_i, and its pdf is $\upsilon(p)/\Pi_i$ if $p^\top\omega^i > b_i$ and 0 otherwise. Notice, that sample $p \sim \mathcal{N}(\mu,\Sigma)$ can be obtained with the plain MC from $\mathcal{N}(\mu,\Sigma)$, but it requires on average $1/\Pi_i$ trials instead of just one for Algorithm 1.

Algorithm 1 Sampling $p \sim \mathcal{N}(\mu,\Sigma)$ conditioned on $p^\top\omega^i \geq b_i$

Input: Mean μ, covariance Σ, and a constraint $p^\top\omega^i \leq b_i$.
Output: $p \sim \mathcal{N}(\mu,\Sigma)$ s.t. $p^\top\omega^i \geq b_i$

1: Sample $z \sim \mathcal{N}(0, I_n)$ and $u \sim U(0,1)$
2: Compute $y = \Phi^{-1}(\Phi(\tau) + u(1-\Phi(\tau)))$
3: Set $\phi = \bar{\phi}y + (I_n - \bar{\phi}\bar{\phi}^\top)z$, with $\bar{\phi} = \Sigma^{1/2}\omega^i/\|\Sigma^{1/2}\omega^i\|_2$
4: **return** $p = \Sigma^{1/2}(\phi + \mu)$

B. Multiple Constraints Case

The case of multiple constraints is more involved. Indeed, there is no analytical formula for a failure probability and, moreover, its exact computation is intractable [18]. Monte-Carlo sampling, $p \sim \mathcal{N}(\mu,\Sigma)$ is inefficient in estimating the failure probability, especially if it is small. Indeed, it requires on average $O(1/\Pi)$ samples to get at least one of the outside the reliability polytope, $p \notin P$.

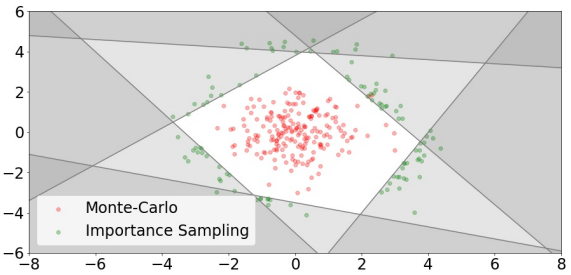


Fig. 1. White area stands for generations which do not lead to a failure. A set of generations leading to at least one constraint violation is in grey. Two or more reliability constraints are not satisfied in dark grey area. Samples from a nominal distribution and the constructed mixture marked in red and green resp.

The importance sampling idea is to change the distribution one samples from and assign a weight to each sample to account for the change:

$$\begin{aligned}\Pi &= \mathbb{P}(p \notin P) = \int_{\mathbb{R}^n} f(p)\upsilon(p)dp\\ &= \int_{\mathbb{R}^n}\frac{f(p)\upsilon(p)}{\upsilon_D(p,x)}\upsilon_D(p,x)dp \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(p^i)\upsilon(p^i)}{\upsilon_D(p^i,x)}, p^i \sim \upsilon_D(p,x),\end{aligned}$$

where we refer to $\upsilon(p)$ as *nominal* distribution, and $\upsilon_D(p,x)$ as *synthetic* distribution with parameter x, and $f(p) = 1[p \notin P]$.

A natural extension of importance sampling with a single linear constraint to the case of multiple linear constraints is to sample from a mixture distribution:

$$D = \sum_{i\leq J} x_i D_i, \text{ with } \sum_{i\leq J} x_i = 1, x_i \geq 0,\ 1 \leq i \leq J, \quad (4)$$

where D_i is $\mathcal{N}(\mu,\Sigma)$ conditioned on $p^\top\omega^i > b_i$. The sampling algorithm consists of two steps. First, we choose a distribution D_i with probability x_i. Second, we sample $p \sim D_i$, i.e. $p \sim \mathcal{N}(\mu,\Sigma)$ given $p^\top\omega^i > b_i$, according to Algorithm 1.

Probability density function $\upsilon_D(p,x)$ of D is given by

$$\upsilon_D(p,x) = \begin{cases}0, & p \in P,\\ \sum_{i\leq J} x_i\upsilon_i(p)1[p^\top\omega^i > b_i], & p \notin P,\end{cases}$$

where $1[\cdot]$ is an indicator of an event, and

$$\upsilon_i(p) = \frac{\upsilon(p)}{\Phi((b_i - \mu^\top\omega^i)/\|\Sigma^{1/2}\omega^i\|_2)}1[p^\top\omega^i > b_i],$$

In contrast to the classical Monte-Carlo, which explores the uncertainty space uniformly according to the nominal distribution, importance sampling from parametric distribution $\upsilon_D(p,x)$ yields samples only from the area of interest, i.e., $p \notin P$. More specifically, Monte-Carlo generates many samples from the true distribution of power injections to estimate failure probability, while the proposed approach only samples power injections that lead to a failure and adjusts their weights. Figure 1 illustrates the difference.

Having a distribution mixture $\{D_i\}_{i=1}^J$we are looking for the weights $\{x_i\}_{i=1}^J$ to approximate a distribution D, $p \sim \mathcal{N}(\mu, \Sigma)$ s.t. $p \notin P$, in the optimal way. Note that, any positive $\{x_i\}_{i=1}^J$ weights lead to an unbiased estimate

$$\hat{\Pi} = \frac{1}{N}\sum_{i=1}^{N} \frac{\upsilon(p^i)}{\upsilon_D(p^i, x^i)}, \quad p^i \sim D^{x^i} \tag{5}$$

of the probability Π. Here x^i is the vector of mixture weights for i^{th} sample. Indeed, by linearity of the expectation $\mathbb{E}_{\upsilon_D(p,x)}\hat{\Pi} = \frac{1}{n}\sum_{i\leq n}\int \frac{\upsilon(p^i)1[p^i\notin P]}{\upsilon_D(p^i,x)}\upsilon_D(p^i,x)dp = \Pi$.

Despite being unbiased for any x with positive components, variance of the estimate (5) highly depends on the choice of x. In [15] the authors suggested to take $x_i \propto \Pi_i$. While it leads to a consistent estimate, the estimator's variance is still high, especially when violation of multiple constraints is likely to happen in the system [15]. In practice, it leads to high sample complexity of the estimator which compromises its real time application. In the next subsection, we significantly improve the sampler's efficiency by using convex optimization to find the optimal combination of the mixture distribution weights x.

C. Convexity of Importance Sampling Variance

We will measure the effectiveness of our estimator by its mean squared error, which is equal to the variance since the estimator is unbiased. The importance sampler variance is

$$\begin{aligned}\mathbb{V}_{\upsilon_D(p,x)}\left(\frac{\upsilon(p)f(p)}{\upsilon_D(p,x)}\right) &= \int_{p\in\mathbb{R}^n}\left(\frac{f(p)\upsilon(p)}{\upsilon_D(p,x)} - \Pi\right)^2 \upsilon_D(p,x)dp \\ &= \int_{p\in\mathbb{R}^n}\frac{f^2(p)\upsilon^2(p)}{\upsilon_D(p,x)}dp - \Pi^2 =: V(x),\end{aligned}$$

where $f(p) = 1[p \notin P]$.

The optimal synthetic distribution can be chosen to minimize the variance, $\upsilon_*(p) = f(p)\upsilon(p)/\Pi$, and thus provide a better approximation to the integral. Notice, that for $\upsilon_D = \upsilon_*$ the variance $V(x) = 0$ and attains its minimum. However, if $\upsilon_*(p)$ does not belong to the parametric family $\{\upsilon_D(p,x)\}_x$, we are looking for the best approximation of $\upsilon_*(p)$ within it, i.e. a minimum possible value of $V(x)$.

Figure 2 illustrates our approach. Starting from an initial weight assignment, $x_i^1 \propto \Pi_i$, at each iteration N of the algorithm we sample $p^k \sim \upsilon_D(\cdot, x^k)$ and compute x^{k+1} to minimize the variance of the estimate. We also update an empirical estimate to the probability

$$\hat{\Pi} = \frac{1}{k}\sum_{t=1}^{k}\frac{\upsilon(p^t)f(p^t)}{\upsilon_D(p^t,x^t)},\ p^t \sim \upsilon_D(\cdot,x^t), t\leq k \tag{6}$$

and update parameters x based on the value of $V(x^N)$ and its gradient. Before discussing the update strategy for the parameters x, we outline some important properties of the variance $V(x)$ and empirical estimate (6).

Theorem 1 implies convexity of the variance minimization problem

$$\min_x V(x),\ \text{s.t.}\ \sum_{i=1}^{J} x_i = 1, x_i \geq 0, 1 \leq i \leq J \tag{7}$$

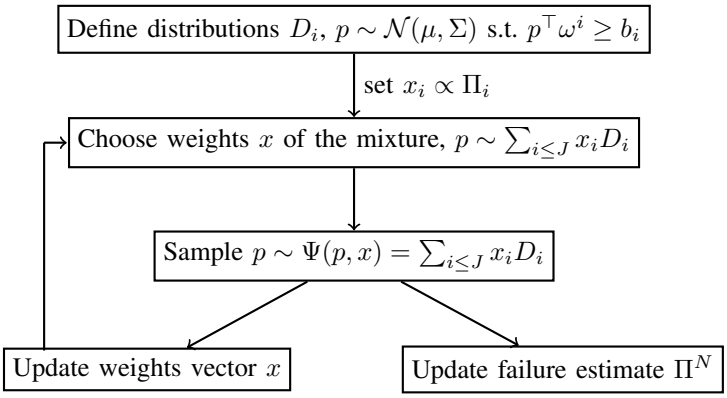


Fig. 2. Scheme of the proposed algorithm. First, we initialize distributions D_i based on a given set of constraints. Second, we initialize the weights of a mixture distribution to sample $x_i \propto \Pi_i$. Then, with each new sample from $\sum_{i\leq J} x_i D_i$ we update the weight vector x and a failure estimate Π_n. See Eq. (9) for details.

in x for the mixture distribution $\upsilon_D(p,x) = \sum_{i=1}^J x_i\upsilon_i(p)$. To improve numerical stability one may also add constraints $x_i \geq \varepsilon > 0$ that guarantee that the variance is bounded.

Theorem 1. *Optimization problem* (7) *is convex in* x. *Moreover,*

$$\nabla V(x) = \mathbb{E}_{p\sim\upsilon_D(\cdot,x)}\left[-\frac{f^2(p)\upsilon^2(p)}{\upsilon_D^2(p,x)}(\upsilon_1(p),\ldots,\upsilon_J(p))^\top\right],$$

for any $x > 0$.

Proof. The sub-integral expression, $f^2(p)\upsilon^2(p)/\upsilon_D(p,x)$, is convex for any x, which implies the integral's convexity. Indeed, the Hessian of the sub-integral expression is non-negative for any x

$$\nabla_x^2\left[f^2(p)\upsilon^2(p)/\upsilon_D(p,x)\right] = 2\frac{f^2(p)\upsilon^2(p)}{\upsilon_D^3(p,x)}h^\top h \succeq 0,$$

where $h = (\upsilon_1(p),\ldots,\upsilon_J(p)) = \nabla_x\upsilon_D(p,x)$. Finally, by the dominated convergence theorem as $\mathbb{E}_{p\sim\upsilon_D(\cdot,x)}\nabla_x\left[f^2(p)\upsilon^2(p)/\upsilon_D(p,x)\right]$ is finite for every $x > 0$, one can exchange the order of differentiation and integration and

$$\begin{aligned}\mathbb{E}_{p\sim\upsilon_D(\cdot,x)}\left[-\frac{f^2(p)\upsilon^2(p)}{\upsilon_D^2(p,x)}h^\top\right] &= \mathbb{E}_{p\sim\upsilon_D(\cdot,x)}\nabla_x\left[\frac{f^2(p)\upsilon^2(p)}{\upsilon_D(p,x)}\right] \\ &= \nabla_x\mathbb{E}_{p\sim\upsilon_D(\cdot,x)}\left[\frac{f^2(p)\upsilon^2(p)}{\upsilon_D(p,x)}\right] = \nabla V(x),\end{aligned}$$

which concludes the proof of the theorem. □

Theorem 2. $\hat{\Pi}$ *is an unbiased estimate of* Π *if for all* k, $1 \leq k \leq N$, $x_k > 0$ *and* x_k *is independent of* x^j *and* p^j *for* $N \geq j > k \geq 1$.

Proof. Let $f(p) = 1[p \notin P]$. By the law of total expectation

$$\mathbb{E}\,\hat{\Pi} = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[\mathbb{E}_{p^k\sim\upsilon_D(p,x^k)}\left[\frac{\upsilon(p^k)f(p^k)}{\upsilon_D(p^k,x^k)}\middle| x^k\right]\right] = \sum_{i=1}^{n}\frac{\Pi}{n} = \Pi,$$

as $\mathbb{E}_{p^k}\left[\frac{\upsilon(p^k)f(p^k)}{\upsilon_D(p^k,x^k)}\middle|x^k\right] = \int_{\mathbb{R}^n}\frac{\upsilon(p^k)1[p^k\notin P]}{\upsilon_D(p^k,x^k)}\upsilon_D(p^k,x^k)dp^k = \Pi$. □

According to Theorem 2, the importance sampling estimate is unbiased. Theorem 3 bounds the variance of $\hat{\Pi}$.

Theorem 3. *Variance of $\hat{\Pi}$ equals $N^{-2}\sum_{k=1}^N V(x^k)$ if for all k and j, $1\le k<j\le N$, $x_k>0$ and x_k is independent of x^j and p^j.*

Proof. Let $f(p)=1[p\notin P]$. As $\mathbb{V}(\hat{\Pi})=\mathbb{E}(\hat{\Pi}-\Pi)^2$ one has

$$\begin{aligned}\mathbb{V} &= \frac{1}{N^2}\sum_{k=1}^N\mathbb{E}\left(\frac{\upsilon(p^k)f(p^k)}{\upsilon_D(p^k,x^i)}-\Pi\right)^2\\ &+\frac{2}{N^2}\sum_{k<j}\mathbb{E}\left(\frac{\upsilon(p^k)f(p^k)}{\upsilon_D(p^k,x^k)}-\Pi\right)\left(\frac{\upsilon(p^j)f(p^j)}{\upsilon_D(p^j,x^j)}-\Pi\right)\\ &=\frac{1}{N^2}\sum_{k=1}^N\mathbb{E}\left[\mathbb{E}\left[\left(\frac{\upsilon(p^k)f(p^k)}{\upsilon_D(p^k,x^k)}-\Pi\right)|x^k\right]\right]\\ &+\frac{2}{N^2}\sum_{k<j}\mathbb{E}\mathbb{E}\left[\left(\frac{\upsilon(p^k)f(p^i)}{\upsilon_D(p^i,x^k)}-\Pi\right)\left(\frac{\upsilon(p^j)f(p^j)}{\upsilon_D(p^j,x^j)}-\Pi\right)|x^k\right],\end{aligned}$$

where the latter is equal to $N^{-2}\sum_{k=1}^N V(x^k)$. □

In the next section, we present a numerical method that guarantees convergence of $V(x^N)$ to the optimal value V^* with an additive error $O(1/\sqrt{N})$.

D. Numerical Method

In this section we focus on efficient numerical methods for minimizing variance which, therefore, accelerate convergence of the importance sampling procedure. The mirror descent [20] is known for its efficiency for simplex-constrained minimization problems. Its particular advantage compared to the stochastic gradient descent [16] and other optimization algorithms is only a logarithmic dependence on the problem dimension.

The mirror descent update for solving

$$\min_{x\ge 0,\sum_{i\le J}x_i=1} V(x) \tag{8}$$

is an iterative modification of a point x_k according to

$$x_{k+1}=\underset{\substack{x\ge 0\\ \sum_{i\le J}x_i=1}}{\arg\min}\left\{\eta^k\nabla V(x^k)^\top(x-x^k)+D_\omega(x,x^k)\right\}, \tag{9}$$

where $\eta^k>0$ is a step-size, and $D_\omega(x,x^k)$ is the Bregman divergence which is defined for any strongly convex and smooth (distance generating) function ω as

$$D_\omega(x,x^k)=\omega(x)-\{\omega(x^k)+\nabla\omega(x^k)^\top(x^k-x)\}.$$

So as the distance generating function is strongly convex and smooth in x, so is the Bregman divergence. When $\omega(x)=\|x\|_2^2/2$, mirror descent step is the same as in the gradient descent method, $x^{k+1}=x_k-\eta^k\nabla V(x^k)$. However, the negative entropy, $\omega(x)=-\sum_{i=1}^n x_i\log x_i$, is known to be the optimal choice for simplex constrained optimization. Solving Eq. (9) in x leads to an update

$$x_i^{k+1}=x_i^k\frac{\exp(-\eta^k(\nabla V(x^k))_i)}{\sum_{i=1}^J x_k\exp(-\eta^k(\nabla V(x^k))_i)},\eta^k>0$$

for $k\ge 1$ and $1\le i\le J$.

Finally, upon minimizing stochastic objective $V(x)$, the expectation of the gradient is inaccessible, so one substitutes $\nabla V(x)$ with a stochastic gradient that comes from the uncertainty realization p,

$$g(x,p)=-\frac{f^2(p)\upsilon^2(p)}{\upsilon_D^2(p,x)}h^\top,\ \mathbb{E}_{p\sim\upsilon_D}g(x,p)=\nabla V(x),$$

where $h=(\upsilon_1(p),\dots,\upsilon_k(p))$. Finally

$$\upsilon_D(p,x)/\upsilon(p)=\sum_{i\le d}x_i 1[p^\top\omega^i\ge b_i],$$

and $f(p)=1[p\notin P]=1$ for any p sampled from υ_D. Thus

$$x_i^{k+1}=x_i^k\frac{\exp\left(\frac{\eta^k\upsilon(p)1[p^\top\omega^i>b_i]}{\sum_{i\le J}x_i1[p^\top\omega^i\ge b_i]}\right)}{\sum_{i=1}^J x_i^k\exp\left(\frac{\eta^k\upsilon(p)1[p^\top\omega^i>b_i]}{\sum_{i\le d}x_i1[p^\top\omega^i\ge b_i]}\right)},1\le i\le J \tag{10}$$

Theorem 4 is a restatement of [21, Theorem 4.1.] which establishes the convergence rate of the mirror descent algorithm.

Theorem 4. *For any function $V(x)$ that is M-Lipschitz in ℓ_1 norm, i.e. $\|V(x)-V(y)\|_\infty\le M\|x-y\|_1\forall x,y$, a constant step-size policy $\eta^k=\eta\le 1/M$, and a sequence $\{x^k\}_{k\ge1}$ generated by* (9) *with $\omega(x)=\sum_{i\le J}x_i\log x_i$, one has*

$$\frac{1}{N}\sum_{i=1}^N(V(x^i)-V^*)\le\frac{\log J+(M^2+\sigma^2)N\eta^2}{N\eta},$$

where $\mathbb{E}_{p\sim\upsilon_D}\|g(p,x)-\nabla V(x)\|_\infty^2\le\sigma^2$, and V_* is the optimal value of Problem (8)*.*

In our study, function $V(x)$ is M-Lipschitz for $x\ge\varepsilon$ with

$$M\le\|\nabla V(x)\|_\infty\le\int_{\mathbb{R}^n}\frac{f^2(p)\upsilon^2(p)}{\upsilon_D(p,x)^2}\upsilon(p)dp\le\Pi/\varepsilon,$$

and

$$\mathbb{E}_{p\sim\upsilon_D}\|g(p,x)-\nabla V(x)\|_\infty^2=\mathbb{E}\left\|\frac{f^2(p)\upsilon^3(p)}{\upsilon^2(p,x)}h-\nabla V(x)\right\|_\infty^2\le$$
$$2\mathbb{E}\left\|\frac{f^2(p)\upsilon^3(p)}{\upsilon^2(p,x)}h\right\|_\infty^2+2\mathbb{E}\|\nabla V(x)\|_\infty^2\le 4\Pi^2/\varepsilon^2.$$

To this end, according to Theorem 4 the optimal choice of $\eta=\varepsilon\Pi^{-1}\sqrt{\log J/(5N)}\le\varepsilon\min_{i\le d}\Pi_i^{-1}\sqrt{\log J/(5N)}$, which yields almost dimension independent convergence rate stated in Theorem 5.

Theorem 5. *Mirror descent with stochastic update 10 and a step-size policy $\eta^k=\eta\varepsilon\Pi^{-1}\sqrt{(\log J)/N}$, $\eta\sqrt{N/\log J}\le 1$ yields*

$$\mathbb{V}_{\upsilon_D}(\hat{\Pi})=\frac{1}{N}\sum_{k=1}^N V(x^k)<\frac{V^*}{N}+\frac{\Pi\sqrt{\log J}}{\varepsilon\eta N^{3/2}}+\frac{5\eta\Pi\sqrt{\log J}}{\varepsilon N^{3/2}}\eta,$$

where V^* is the optimal value of Problem (8).

Compared to the earlier results of [15], the rate of convergence depends as $O(\sqrt{\log J})$ on the dimension J, while earlier results [15] claim linear at dependence. Thus, our result provides a substantial acceleration for large-scale problems.

IV. Empirical Study

A. Algorithms and implementation details

We compare performance of importance samplers over real and simulated test cases whose dimensions vary from several dozens to several thousands variables. We limit the empirical setting to considering Gaussian distributions and linear constraints only.

a) Compared Algorithms: In this study, we have compared Monte-Carlo Sampling, ALOE [15], pmvnorm [14] and mirror descent for variance minimization (MD-Var). We have also applied the algorithms to the same setting with KL-divergence [19] between the generated distribution υ_D and the optimal distribution υ_D^* as a measure of estimator's quality (instead of variance $V(x)$). This similarly leads to a convex optimization problem similar to [22]. The former and the latter are the proposed methods.

b) Implementation details: We have used Python 3.8.5. and PandaPower 2.2.2 [23] on MacBook Pro (2.4GHz, 8-Core Intel i9, 64 GB RAM). In the experiments computational time for each of the cases for MD-Var method have not exceeded two minutes, which makes the solution applicable for the operational practice. Our code is publicly available on Github[1].

B. Test cases and numerical results

We evaluate our algorithms on multiple real (power grids) and simulated test cases. We estimate the probability of system failure, i.e. the probability that at least one of the realibility constraints fails. Assuming Gaussian fluctuations of output power of renewables, the probability equals to the Gaussian volume of the reliability polytope's complement $\mathbb{R}^n \setminus P$, as it was shown earlier. First, we conduct our experiments on the regular polytope, then we consider degenerate polytope. The latter is merely two parallel planes, one of them has a number of slightly shivered duplicates. This test assesses the stability of the algorithms and ability to handle joint geometry of the problem. Finally, we apply the proposed algorithms to various power grids.

a) Regular polytope: We consider a regular 2 dimensional polytope with J faces ($J \geq 3$) centered at zero, $P = \{p \in \mathbb{R}^2 : \omega_j^\top x \leq \tau, 1 \leq j \leq J\}$, where $\omega_j = (\sin(2\pi j/J), \cos(2\pi j/J))$. We assume $p \sim \mathcal{N}(0, I_2)$, where I_2 is 2×2 identity matrix. The probability $p \notin P$ rapidly converges to $\exp(-\tau^2/2)$ as $J \to \infty$ [15]. Figure 3 compares performance of MC, ALOE [15], mirror descent (Section III-D) minimizing variance (MD-Var) and KL-divergence (MD-KL) and pmvnorm [14] methods for $\tau = 6$ and $J = 360$. Figure 3 shows the histogram of $\hat{\Pi}/\Pi$ for 100 runs of the algorithms on a sample size of 1000.The MD-Var method

[1]https://github.com/vjugor1/adaptive_importance_sampling_power_grids

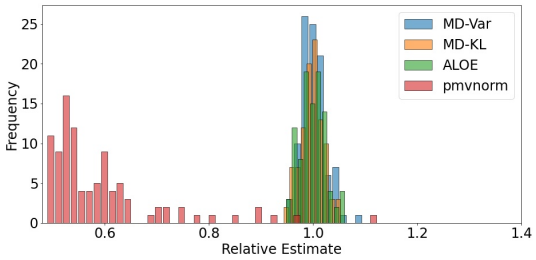


Fig. 3. Importance sampling methods performance on a 2-dimensional regular polytope with 360 faces and failure probability $\Phi(-6)$.

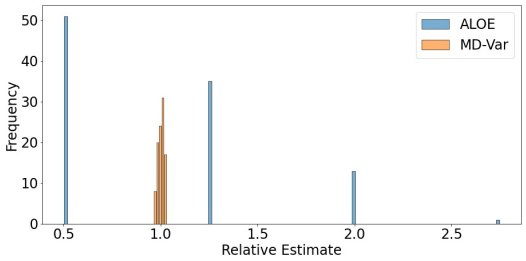


Fig. 4. Importance sampling methods performance on a 2-dimensional degenerate polytope with 1500 faces and failure probability $2\Phi(-1)$.

demonstrates a slightly better performance then ALOE, while pmvnorm tends to significantly underestimate the probability of $p \notin P$. Monte-Carlo sampling from the nominal distribution failed to generate any event $p \notin P$ in 10^6 tries, and estimated a failure probability Π as zero.

b) Degenerate polytope: Although ALOE is one of the best choices for a regular polytope, the algorithm does not take into account the joint geometry of various hyperplanes. As the second example we consider a degenerate polytope with $J = 1500$ faces, where $\omega^1 = (0, 1)$ and $\omega^j = (\xi, -1-\xi)$, $2 \leq j \leq J$. We take $\xi \sim \mathcal{U}[-\varepsilon, \varepsilon]$ for small $\varepsilon = 10^{-6}$. Note that ω^j for $j \geq 2$ are almost identical. Hence probability Π is quite close to $2\Phi(-\tau)$. In this experiment, ALOE puts a lot of efforts on sampling points in the area $\{p : (0, -1)^\top p \geq \tau\}$, while the set $\{p : \omega_1^\top p \geq \tau\}$ remains unexplored which leads to a higher variance of the sampler and a less efficient method compared to the proposed optimization approach. Figure 4 illustrates the performance of ALOE in this case.

c) Power Grid Cases: In this subsection we consider real-world polytopes corresponding to DC power grids (IEEE test cases) and Gaussian power injections. We ran the algorithms on all the test cases accessible through PandaPower [23]. There were 27 cases with the number of buses varying from 4 to 9241. The proposed methods (MD-Var and MD-KL) took less than two minutes of computational time on a personal laptop for each of them.

Table II shows the minimal number of samples that are required by the algorithms to achieve

$$\Pi/2 \leq \hat{\Pi} \pm s(\hat{\Pi}) \leq 3\Pi/2, \tag{11}$$

where $s(\hat{\Pi})$ is the empirical standard deviation of the estimate. This ensures that not only the estimated value, but also its confidence interval is contained in $(\Pi/2, 3\Pi/2)$ and that the

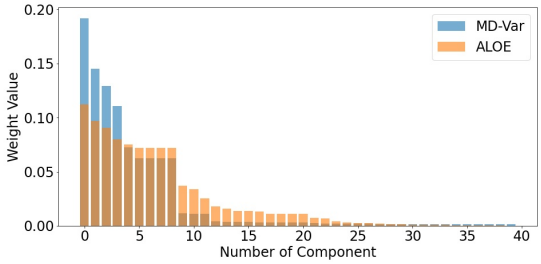


Fig. 5. Weights of mixture distribution (4) assigned by ALOE and MD-Var. Iceland 118 case with maximum phase angle difference $\pi/3$ and a standard deviation of power injection on generators equal to 0.1 of the average value.

sum of the empirical estimate and its standard deviation are close to the true probability.

TABLE II
NUMBER OF SAMPLES TO SATISFY INEQ. (11) FOR ICELAND118

Bound $\bar{\theta}_{ij}$, failure prob. Π	MC	ALOE	Pmvnorm	MD-Var
$\lvert\theta_{ij}\rvert \leq \pi/8$, Π = 1.2e-01	6.4e+02	3.7e+02	3.2e+02	4.1e+02
$\lvert\theta_{ij}\rvert \leq \pi/7$, Π = 3.0e-02	5.1e+04	4.1e+02	1.1e+03	3.5e+02
$\lvert\theta_{ij}\rvert \leq \pi/6$, Π = 2.5e-03	6.2e+06	4.5e+02	6.3e+03	3.9e+02
$\lvert\theta_{ij}\rvert \leq \pi/5$, Π = 2.6e-05	8.9e+10	3.3e+02	1.4e+04	2.1e+02

Table III shows failure probability estimates and their standard deviations for the algorithms based on $N = 200$ samples on various PandaPower [23] power grids. In all the presented cases except for the Iceland grid, we set the standard deviations of output powers of generators to 0.25 of their average values. For the Iceland test case, we use 0.1 instead of 0.25. Pmvnorm [14] did not terminate on Polish 3120sp case after an hour of computations which we indicated as N/A. All other methods terminate in less than a minute. The proposed algorithms reduce variance and are more computationally efficient than the state-of-the-art ALOE and pmvnorm. Fig. 5 shows a substantial change in hyperplane weight assignment made by MD-Var.

V. CONCLUSION

Importance sampling can be a useful tool for real-time reliability assessment in direct current power grids. We proposed an algorithm that, first, constructs a physics-informed mixture distribution for importance sampling, and, second, utilizes convex optimization to adjust the weights of the mixture. The method outperforms state-of-the-art algorithms in accuracy and efficiency of reliability assessment. We hope that this approach can be further used for optimization and control in power grids.

TABLE III
FAILURE PROBABILITY ESTIMATION FOR POWER GRIDS.

Estimate, $\hat{\Pi}$	$\bar{\theta}$	Π	ALOE	MD-Var	MD-KL	pmvnorm
IEEE 30						
$\hat{\Pi}\times$ 1e+15	$\pi/4$	8.2	8.2 $\pm$ 0.9	8.2 $\pm$ 0.0	8.2 $\pm$ 0.0	8.2 $\pm$ 1.2
$\hat{\Pi}\times$ 1e+06	$\pi/6$	5.8	5.8 $\pm$ 0.7	5.8 $\pm$ 0.0	5.8 $\pm$ 0.0	5.8 $\pm$ 1.0
$\hat{\Pi}\times$ 1e+04	$\pi/7$	2.9	2.9 $\pm$ 0.3	2.9 $\pm$ 0.0	2.9 $\pm$ 0.0	2.9 $\pm$ 0.5
$\hat{\Pi}\times$ 1e+03	$\pi/8$	3.1	3.1 $\pm$ 0.4	3.1 $\pm$ 0.0	3.1 $\pm$ 0.1	3.1 $\pm$ 0.4
IEEE 57						
$\hat{\Pi}\times$ 1e+03	$\pi/2$	8.8	9.1 $\pm$ 0.8	8.7 $\pm$ 0.0	8.8 $\pm$ 0.0	8.9 $\pm$ 1.2
$\hat{\Pi}\times$ 1e+02	$\pi/3$	8.4	8.5 $\pm$ 1.1	8.4 $\pm$ 0.1	8.3 $\pm$ 0.5	9.0 $\pm$ 0.9
Iceland 118						
$\hat{\Pi}\times$ 1e+09	$\pi/2$	6.2	6.2 $\pm$ 0.1	6.1 $\pm$ 0.0	6.1 $\pm$ 0.0	5.7 $\pm$ 2.6
$\hat{\Pi}\times$ 1e+04	$\pi/3$	2.8	3.0 $\pm$0.0	2.9 $\pm$ 0.0	2.9 $\pm$ 0.0	2.8 $\pm$ 1.4
$\hat{\Pi}\times$ 1e+02	$\pi/4$	1.1	1.1 $\pm$ 0.2	1.1 $\pm$ 0.0	1.1 $\pm$ 0.0	1.1 $\pm$ 0.2
$\hat{\Pi}\times$ 1e+01	$\pi/6$	1.4	1.4 $\pm$ 0.2	1.4 $\pm$ 0.0	1.3 $\pm$ 0.0	1.4 $\pm$ 0.1
Illinois 200						
$\hat{\Pi}\times$ 1e+12	$\pi/4$	7.9	7.9 $\pm$ 0.9	7.9 $\pm$ 0.0	7.9 $\pm$ 0.0	7.9 $\pm$ 3.3
$\hat{\Pi}\times$ 1e+04	$\pi/6$	1.1	1.1 $\pm$ 0.1	1.1 $\pm$ 0.0	1.1 $\pm$ 0.0	1.1 $\pm$ 0.3
$\hat{\Pi}\times$ 1e+03	$\pi/7$	2.3	2.3 $\pm$ 0.2	2.3 $\pm$ 0.0	2.3 $\pm$ 0.0	2.3 $\pm$ 0.3
$\hat{\Pi}\times$ 1e+02	$\pi/8$	1.5	1.5 $\pm$ 0.1	1.5 $\pm$ 0.0	1.5 $\pm$ 0.0	1.5 $\pm$ 0.1
Polish 3120sp						
$\hat{\Pi}\times$ 1e+13	$\pi/2$	3.7	3.7 $\pm$ 0.4	3.7 $\pm$ 0.0	3.7 $\pm$ 0.0	N/A
$\hat{\Pi}\times$ 1e+04	$\pi/3$	1.2	1.2 $\pm$ 0.1	1.2 $\pm$ 0.0	1.2 $\pm$ 0.0	N/A
$\hat{\Pi}\times$ 1e+02	$\pi/4$	3.4	3.4 $\pm$ 0.5	3.4 $\pm$ 0.3	3.4 $\pm$ 0.6	N/A

REFERENCES

[1] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," *Energy Strategy Reviews*, vol. 24, pp. 38–50, 2019.

[2] A. Harjanne and J. M. Korhonen, "Abandoning the concept of renewable energy," *Energy policy*, vol. 127, pp. 330–340, 2019.

[3] K. S. Golden, "Senate bill 1078: The renewable portfolio standard-california asserts its renewable energy leadership," *Ecology LQ*, vol. 30, p. 693, 2003.

[4] K. Schmietendorf, J. Peinke, and O. Kamps, "The impact of turbulent renewable energy production on power grid stability and quality," *The European Physical Journal B*, vol. 90, no. 11, pp. 1–6, 2017.

[5] X. Liang, "Emerging power quality challenges due to integration of renewable energy sources," *IEEE Transactions on Industry Applications*, vol. 53, no. 2, pp. 855–866, 2016.

[6] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2017.

[7] F. Xiao, J. D. McCalley, Y. Ou, J. Adams, and S. Myers, "Contingency probability estimation using weather and geographical data for on-line security assessment," in *2006 International Conference on Probabilistic Methods Applied to Power Systems*, 2006, pp. 1–7.

[8] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2007.

[9] C.-L. Su, "Probabilistic load-flow computation using point estimate method," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1843–1851, 2005.

[10] E. Vittal, M. O'Malley, and A. Keane, "A steady-state voltage stability analysis of power systems with high penetrations of wind," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 433–442, 2009.

[11] H. Yu, C. Chung, K. Wong, H. Lee, and J. Zhang, "Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 661–667, 2009.

[12] P. Chen, Z. Chen, and B. Bak-Jensen, "Probabilistic load flow: A review," in *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, 2008, pp. 1586–1591.

[13] A. M. M. Marsadek and Z. Norpiah, "Risk of static security assessment of a power system using non-sequential monte carlo simulation," *Journal of Applied Sciences, 11: 300-307*, no. 11, pp. 300–307, 2011.

[14] A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, B. Bornkamp, M. Maechler, T. Hothorn, and M. T. Hothorn, "Package 'mvtnorm'," *Journal of Computational and Graphical Statistics*, vol. 11, pp. 950–971, 2020.

[15] A. B. Owen, Y. Maximov, and M. Chertkov, "Importance sampling the union of rare events with an application to power systems analysis," *Electronic Journal of Statistics*, vol. 13, no. 1, pp. 231–254, 2019.

[16] E. K. Ryu and S. P. Boyd, "Adaptive importance sampling via stochastic convex programming," *arXiv preprint arXiv:1412.4845*, 2014.

[17] O. Cappé, R. Douc, A. Guillin, J.-M. Marin, and C. P. Robert, "Adaptive importance sampling in general mixture classes," *Statistics and Computing*, vol. 18, no. 4, pp. 447–459, 2008.

[18] L. G. Khachiyan, "The problem of calculating the volume of a polyhedron is enumerably hard," *Russian Mathematical Surveys*, vol. 44, no. 3, p. 199, 1989.

[19] P. L'Ecuyer and A. B. Owen, *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Springer, 2009.

[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] G. Lan, *First-order and Stochastic Optimization Methods for Machine Learning*. Springer, 2020.
[22] R. Y. Rubinstein and D. P. Kroese, *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*. Springer Science & Business Media, 2013.
[23] L. Thurner, A. Scheidler, F. Schäfer, J. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, "Pandapower — an open-source python tool for convenient modeling, analysis, and optimization of electric power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, Nov 2018.