# Stochastic Uncertainty Propagation in Power System Dynamics using Measure-valued Proximal Recursions

Abhishek Halder, *Senior Member, IEEE,* Kenneth F. Caluya, Pegah Ojaghi, and Xinbo Geng, *Member, IEEE*

***Abstract*—We present a proximal algorithm that performs a variational recursion on the space of joint probability measures to propagate the stochastic uncertainties in power system dynamics over high dimensional state space. The proposed algorithm takes advantage of the exact nonlinearity structures in the trajectory-level dynamics of the networked power systems, and is nonparametric. Lifting the dynamics to the space of probability measures allows us to design a scalable algorithm that obviates gridding the underlying high dimensional state space which is computationally prohibitive. The proximal recursion implements a generalized infinite dimensional gradient flow, and evolves probability weighted scattered point clouds. We clarify the theoretical nuances and algorithmic details specific for the power system nonlinearities, and provide illustrative numerical examples.**

***Index Terms*—Uncertainty propagation, power system dynamics, kinetic Fokker-Planck equation, proximal operator.**

## I. Introduction

STOCHASTIC variabilities in power grid have increased significantly in recent years both in the generation side (e.g., due to growing penetration of renewables) as well as in the load side (e.g., due to widespread adoption of plug-in electric vehicles). Several studies [4], [7]–[10] have reported that even small stochastic effects can significantly alter the assessment of transient stability, or the performance of automatic generation control. However, the lack of a rigorous yet scalable stochastic computational framework continues to impede [1] our ability to perform transient analysis involving time varying joint probability density functions (PDFs) over the states of a large power system network. In this paper, we present a new algorithm to address this computational need.

Given a networked power system, one can envisage at least three types of uncertainties affecting the dynamics: initial condition uncertainties in the state variables (e.g., rotor phase angles and angular velocities), parametric uncertainties (e.g., inertia and damping coefficients of the generators, reactance associated with different transmission lines), and stochastic forcing (e.g., intermittencies in renewable power generation, load and ambient temperature fluctuations). Fig. 1 depicts a representative scenario. In addition, one could consider uncertainties due to random change in transmission topology resulting from unexpected outage, and uncertainties due to unmodeled dynamics. Given a statistical description of these uncertainties, our approach is to directly solve the *macroscopic flow* of the joint PDFs governing the probabilistic evolution of the state as summarized in Fig. 2.

A. Halder and K.F. Caluya are with the Department of Applied Mathematics, University of California, Santa Cruz, CA 95064, USA, {ahalder,kcaluya}@ucsc.edu.

P. Ojaghi is with the Department of Computer Science and Engineering, University of California, Santa Cruz, CA 95064, USA, pojaghi@ucsc.edu.

X. Geng is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850, USA, xg72@cornell.edu.

This research was partially supported by the NSF award 1923278.

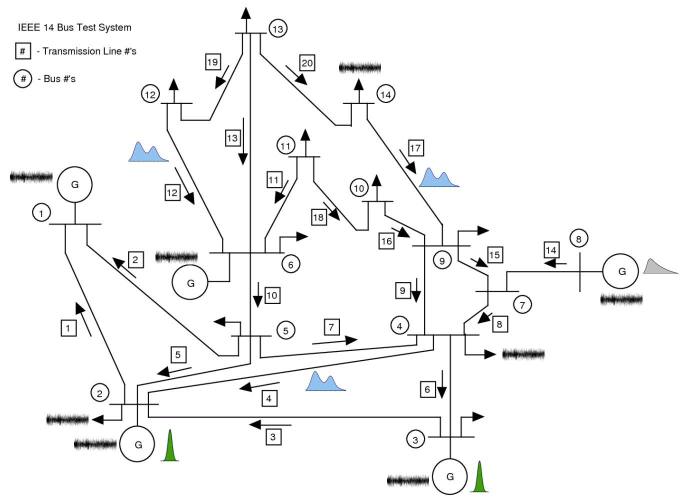


Fig. 1: A schematic of the IEEE 14 bus test system with stochastic uncertainties. The Uncertainty sources may include stochastic forcing and parametric uncertainties at some generators, random variabilities at some loads, and parametric uncertainties along some transmission lines. For depiction purposes, we indicated the parametric uncertainties as PDFs, and stochastic forcing as intermittent signals.

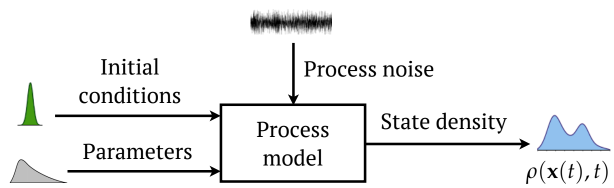


Fig. 2: Block diagram for joint state PDF propagation.

*1) Related works:* Even though the need for quantifying uncertainties in power systems simulations has been long-recognized [2], [3], early studies were limited to statistical reliability assessment. Dynamic simulations with stochastic uncertainties for purposes such as transient stability analysis have been investigated via *Monte Carlo simulations* [4]–[7].

As is well known, the Monte Carlo techniques are easy to apply but the computational cost scales exponentially with the number of dimensions, thus making it prohibitive for realistic power systems dynamic simulation. As an alternative, *probabilistic small signal analysis* [10]–[13] have appeared in the power systems literature, albeit at the expense of the additional assumption that the random perturbations remain “small”. *Polynomial chaos and related stochastic collocation methods* [14], [15] can do away with the “small stochastic perturbation” assumption but due to the finite-dimensional approximation of the probability space, computational performance degrades if the long-term statistics are desired. Furthermore, to cope with the stochasticity, these techniques require simulating a higher-dimensional system than the dimension of the physical state space, which further limits the scalability for nonlinear simulation. More recently, approximation methods such as *stochastic averaging* [16], [17] based on certain energy function [18]–[20] have appeared. In [21], an algorithm for *propagating first few statistical moments* was proposed. However, this leads to moment closure problems since the dimensions of the time-varying sufficient statistics associated with the corresponding transient joint state PDFs are not known in general.

In a different vein, *deterministic bounded uncertainty models* for power flow simulations have been used [22]–[25] for set-valued analysis. These, however, require approximating the underlying nonlinear differential algebraic equations (DAEs) appearing in power system dynamic simulation, and thus lead to conservative analysis. For example, the method in [25] requires converting the nonlinear DAEs to linear DAEs in such a way that guarantees set-valued over-approximation of the reachable sets. Likewise, the convex optimization-based bounded uncertainty propagation methods in [26], [27] require second order approximation of the power flow state variables as function of the uncertainties.

*2) Technical challenges:* Several technical challenges need to be overcome to achieve scalable computation enabling the prediction of the joint PDFs over a time horizon of interest. *First*, the trajectory level dynamic models for power systems are inherently nonlinear, which do not preserve Gaussianity, thereby requiring nonparametric prediction of the joint PDFs. *Second*, the joint PDFs for realistic power systems dynamic simulation must evolve over a high dimensional state space, i.e., the joint PDF at any given time has high dimensional support. This necessitates spatial discretization-free algorithms since standard function approximation or interpolation approaches would, in general, be met with "curse-of-dimensionality" [28]. The numerical challenges aside, one cannot theoretically guarantee to find a suite of basis functions for the manifold of nonparametric PDFs. *Third*, prediction based on first few statistical moments is challenging since it is not possible to *a priori* guarantee a fixed or even finite dimensional sufficient statistic. For instance, propagating only the mean and covariance could be misleading when the underlying joint PDFs are multi-modal.

*3) Contributions of this paper:* Our main contribution is to demonstrate that by harnessing recent developments in generalized gradient flows [42] and proximal algorithms [29], it is possible to perform *nonparametric* propagation of the joint state PDFs subject to the stochastic nonlinear dynamics of a networked power system, and that the computation can be performed in a gridless and scalable manner. In doing so, the paper also makes theoretical contribution by introducing a change of coordinates that transforms the joint PDF evolution equations in way that is amenable for the aforesaid generalized proximal recursions.

Our technical approach juxtaposes with the existing power systems literature discussed above, in that we approximate neither the dynamical nonlinearity nor the statistics. Instead of treating the exact nonlinearity as bane, we exploit the geometry induced by the power system's dynamical nonlinearity over the manifold of time-varying joint state PDFs, thereby enabling a new proximal algorithm to compute the transient joint PDFs.

*4) Notation and organization:* The set of natural numbers, real numbers and complex numbers are denoted as $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$, respectively. The symbol $\nabla_{\boldsymbol{x}}$ denotes the Euclidean gradient operator with respect to (w.r.t.) the vector $\boldsymbol{x}$. Thus, $\nabla_{\boldsymbol{x}}\cdot$ stands for the divergence, and $\Delta_{\boldsymbol{x}}$ stands for the Laplacian w.r.t. vector $\boldsymbol{x}$. We use $\langle\cdot,\cdot\rangle$ to denote the standard Euclidean inner product. The real and imaginary parts of a complex number $z$ are denoted via $\Re(z)$ and $\Im(z)$, respectively. We use the superscript $*$ to denote the complex conjugate. The uniform probability distribution over an interval $[a,b]$ is denoted as $\mathrm{Unif}\left([a,b]\right)$. Likewise, the $n$ dimensional uniform probability distribution over $[a,b]^n$ is denoted as $\mathrm{Unif}\left([a,b]^n\right)$. The symbol $\propto$ denotes proportionality, $|\cdot|$ denotes the absolute magnitude, $\|\cdot\|_2$ denotes the standard Euclidean 2-norm, $\otimes$ denotes the Kronecker product, $\odot$ and $\oslash$ respectively denote the elementwise (Hadamard) product and division, $\det(\cdot)$ stands for the determinant, and the subscript $_\sharp$ denotes the pushforward of a PDF via a map. The $n\times n$ identity and zero matrices are denoted as $\boldsymbol{I}_n$ and $\boldsymbol{0}_{n\times n}$, respectively.

The rest of this paper is structured as follows. Section II details the power system dynamics models at the microscopic or trajectory level (Sec. II-A) as well as at the macroscopic or statistical ensemble level (Sec. II-B). The proposed idea of realizing the flow of the joint state PDFs subject to the macroscopic power system dynamics via infinite dimensional proximal recursions, is explained in Section III. Section IV elucidates the corresponding proximal algorithm that enables the computation of the transient joint state PDFs via weighted point cloud evolution. Numerical simulations illustrating the proposed framework are reported in Section V. Section VI concludes the paper.

## II. Models

### A. Sample Path Dynamics

In this work, we consider the coupled stochastic differential equations (SDEs) associated with the networked-reduced power systems model [20, Ch. 7]. Specifically, for a power network with $n$ generators, the stochastic dynamics for the $i$-th generator is given by the Itô SDEs

$$\mathrm{d}\theta_i = \omega_i\,\mathrm{d}t, \tag{1a}$$

$$m_i\,\mathrm{d}\omega_i = \left(P_i - \gamma_i\omega_i - \sum_{j=1}^{n} k_{ij}\sin\left(\theta_i-\theta_j-\varphi_{ij}\right)\right)\mathrm{d}t + \sigma_i\,\mathrm{d}w_i, \tag{1b}$$

where the state variables are the rotor angles $\theta_i\in[0,2\pi)$ and the rotor angular velocities $\omega_i\in\mathbb{R}$, for $i\in\{1,\ldots,n\}$. The stochastic forcing is modeled through the standard Wiener process $w_i(t)$, and the diffusion coefficient $\sigma_i>0$ denotes the intensity of stochastic forcing at the $i$th generator.

**Remark 1.** *We emphasize here that the network-reduced model* (1) *is obtained from the so-called structure preserving power network model [7] after applying the Kron reduction [51], and therefore, has all-to-all connection topology.*

With the $i$th generator, we associate its inertia $m_i > 0$ and damping coefficient $\gamma_i > 0$. The other parameters: the *effective power input* $P_i$, the *phase shift* $\varphi_{ij} \in [0, \frac{\pi}{2})$, and the *coupling coefficients* $k_{ij} \geq 0$, depend on the network reduced admittance matrix $\boldsymbol{Y} \equiv [Y_{ij}]_{i,j=1}^{n} \in \mathbb{C}^{n\times n}$. Specifically,

$$P_i = P_i^{\text{mech}} - P_i^{\text{load}} - |E_i|^2 \Re\left(Y_{ii}\right) + \Re\left(E_i \cdot I_i^*\right), \tag{2a}$$

$$\varphi_{ij} = \begin{cases} -\arctan\left(\dfrac{\Re\left(Y_{ij}\right)}{\Im\left(Y_{ij}\right)}\right), & \text{if } i \neq j, \\ 0, & \text{otherwise,} \end{cases} \tag{2b}$$

$$k_{ij} = \begin{cases} |E_i||E_j||Y_{ij}|, & \text{if } i \neq j, \\ 0, & \text{otherwise,} \end{cases} \tag{2c}$$

where $P_i^{\text{mech}}$ is the mechanical power input, $P_i^{\text{load}}$ is the real load, $E_i$ is the internal voltage, and $I_i \in \mathbb{C}$ is the current for generator $i$.

Suppose that the *unreduced* power network has $n$ generators and $m$ buses. Then the unreduced admittance matrix $\boldsymbol{Y}_{\text{unreduced}} \in \mathbb{C}^{(n+m)\times(n+m)}$ can be partitioned as

$$\boldsymbol{Y}_{\text{unreduced}} = \left[\begin{array}{c|c} \boldsymbol{Y}_{\text{bnd}} & \boldsymbol{Y}_{\text{bnd-int}} \\ \hline \boldsymbol{Y}_{\text{bnd-int}}^{\top} & \boldsymbol{Y}_{\text{int}} \end{array}\right], \tag{3}$$

where $\boldsymbol{Y}_{\text{bnd}} \in \mathbb{C}^{n\times n}$, $\boldsymbol{Y}_{\text{int}} \in \mathbb{C}^{m\times m}$, $\boldsymbol{Y}_{\text{bnd-int}} \in \mathbb{C}^{n\times m}$. The matrix (3) relates the unreduced current vector $I_{\text{unreduced}} \in \mathbb{C}^{n+m}$ with the unreduced voltage vector $E_{\text{unreduced}} \in \mathbb{C}^{n+m}$ via the Kirchhoff equations

$$I_{\text{unreduced}} = \boldsymbol{Y}_{\text{unreduced}} E_{\text{unreduced}}, \tag{4}$$

or equivalently, via its partitioned version associated with the interior and the boundary nodes:

$$\begin{bmatrix} I_{\text{bnd}} \\ I_{\text{int}} \end{bmatrix} = \boldsymbol{Y}_{\text{unreduced}} \begin{bmatrix} E_{\text{bnd}} \\ E_{\text{int}} \end{bmatrix}, \tag{5}$$

where $I_{\text{bnd}}, E_{\text{bnd}} \in \mathbb{C}^n$ and $I_{\text{int}}, E_{\text{int}} \in \mathbb{C}^m$.

The network reduced admittance matrix $\boldsymbol{Y} \in \mathbb{C}^{n\times n}$ in (2) is the Schur complement of (3) w.r.t. the block $\boldsymbol{Y}_{\text{int}}$, i.e.,

$$\boldsymbol{Y} = \boldsymbol{Y}_{\text{unreduced}}/\boldsymbol{Y}_{\text{int}} := \boldsymbol{Y}_{\text{bnd}} - \boldsymbol{Y}_{\text{bnd-int}} \boldsymbol{Y}_{\text{int}}^{-1} \boldsymbol{Y}_{\text{bnd-int}}^{\top}.$$

The network reduced current vector $I \in \mathbb{C}^n$ in (2a) is obtained as

$$I = \boldsymbol{Y} E_{\text{bnd}} - I_{\text{bnd}}.$$

One can view (1) as the noisy version of the second order nonuniform Kuramoto oscillator model [30], [31], given by

$$m_i \ddot{\theta}_i + \gamma_i \dot{\theta}_i = P_i - \sum_{j=1}^{n} k_{ij} \sin\left(\theta_i - \theta_j - \varphi_{ij}\right) + \sigma_i \times \text{stochastic forcing}, \tag{6}$$

where the stochastic forcing is standard Gaussian white noise.

We define the *positive diagonal* matrices

$$\boldsymbol{M} := \operatorname{diag}\left(m_1, \ldots, m_n\right),$$
$$\boldsymbol{\Gamma} := \operatorname{diag}\left(\gamma_1, \ldots, \gamma_n\right),$$
$$\boldsymbol{\Sigma} := \operatorname{diag}\left(\sigma_1, \ldots, \sigma_n\right),$$

and rewrite (1) as a mixed conservative-dissipative SDE in state vector $\boldsymbol{x} := (\boldsymbol{\theta}, \boldsymbol{\omega})^{\top} \in \mathbb{T}^n \times \mathbb{R}^n$ as

$$\begin{pmatrix} \mathrm{d}\boldsymbol{\theta} \\ \mathrm{d}\boldsymbol{\omega} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\omega} \\ -\boldsymbol{M}^{-1}\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega} \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \boldsymbol{0}_{n\times n} \\ \boldsymbol{M}^{-1}\boldsymbol{\Sigma} \end{pmatrix} \mathrm{d}\boldsymbol{w}, \tag{7}$$

where $\boldsymbol{w} \in \mathbb{R}^n$ is the standard vector Wiener process, $\mathbb{T}^n$ denotes the $n$-torus $[0, 2\pi)^n$, and the potential function $V : \mathbb{T}^n \mapsto \mathbb{R}$ is given by

$$V(\boldsymbol{\theta}) := \sum_{i=1}^{n} P_i \theta_i + \sum_{i,j=1}^{n} k_{ij} \left(1 - \cos(\theta_i - \theta_j - \varphi_{ij})\right). \tag{8}$$

The potential (8) has a natural energy function interpretation and can also be motivated by a mechanical mass-spring-damper analogy [32], [33].

### *B. Macroscopic Dynamics*

Given the sample path dynamics (1) or equivalently (7), a prescribed initial joint state PDF

$$\rho_0(\boldsymbol{x}) \equiv \rho(t = t_0, \boldsymbol{\theta}(t_0), \boldsymbol{\omega}(t_0)) \tag{9}$$

denoting initial condition uncertainties at time $t = t_0$, and prescribed parametric uncertainties given by the joint parameter PDF $\rho_{\text{param}}$, the uncertainty propagation problem calls for computing the transient joint state PDFs $\rho(t, \boldsymbol{x}) \equiv \rho(t, \boldsymbol{\theta}, \boldsymbol{\omega})$ for any desired time $t \geq t_0$, which is a nonnegative function supported on the state space $\mathbb{T}^n \times \mathbb{R}^n$ satisfying $\int \rho = 1$ for all $t \geq t_0$.

The corresponding macroscopic dynamics governing the flow of the joint state PDF $\rho(t, \boldsymbol{\theta}, \boldsymbol{\omega})$ is given by a kinetic Fokker-Planck partial differential equation (PDE)

$$\frac{\partial \rho}{\partial t} = -\langle \boldsymbol{\omega}, \nabla_{\boldsymbol{\theta}} \rho \rangle + \nabla_{\boldsymbol{\omega}} \cdot \left(\rho \left(\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega} + \boldsymbol{M}^{-1}\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) + \frac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\top}\boldsymbol{M}^{-1}\nabla_{\boldsymbol{\omega}} \log \rho\right)\right), \tag{10}$$

subject to the initial condition (9) and the joint parameter PDF $\rho_{\text{param}}$. Of course, this subsumes special cases such as when either the initial condition or the parameter vector is deterministic. A direct numerical solution of this PDE initial value problem using conventional discretization (e.g., finite difference) or function approximation techniques will not be scalable in general, as explained in Sec. I-2. In the next Section, we discuss how a measure-valued variational recursion proposed in our recent works [34]–[38] can be employed to address this challenge.

We mention here that (7) has been used in [16], [17] for uncertainty propagation via stochastic averaging approximation where the univariate energy PDF was proposed as a "proxy" for the entire joint PDF. Most relevant to our approach in the power systems literature is the work in [39], which indeed voiced the need for computing the transient joint PDFs but only dealt with the single-machine-infinite-bus case – simplest ($n = 1$) instance of (7). The resulting bivariate Fokker-Planck PDE in [39] was solved via finite element discretization, and revealed rich stochastic dynamics and nontrivial transient stability aspects even in this simple

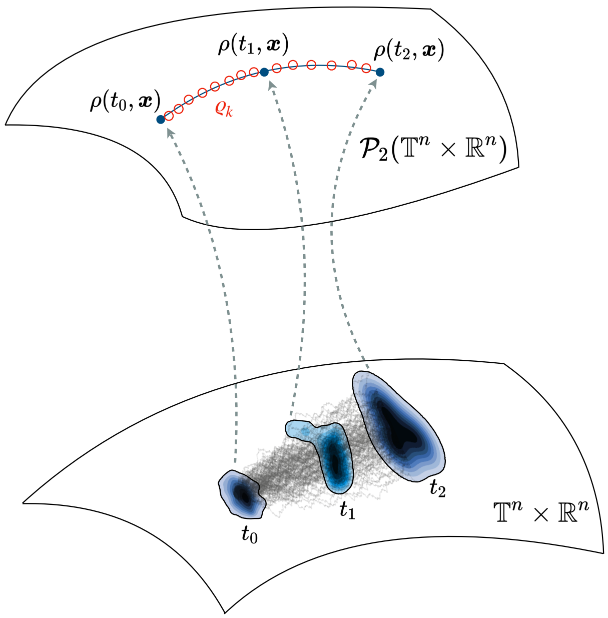


Fig. 3: A schematic of approximating the flow of the joint PDF trajectory $\rho(t, \boldsymbol{x})$ via the sequence $\{\varrho_k\}_{k\in\mathbb{N}}$ generated by a variational recursion over $\mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n)$. The curve $\rho(t, \boldsymbol{x})$ solves a kinetic Fokker-Planck PDE initial value problem. While the joint PDFs are supported over the finite dimensional base manifold $\mathbb{T}^n \times \mathbb{R}^n$, the proximal updates $\{\varrho_k\}_{k\in\mathbb{N}}$ (shown as circled markers with no face-color) evolve over the infinite dimensional manifold $\mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n)$.

case. However, it is unreasonable to expect that a finite element discretization, or in fact any spatial discretization scheme to solve (10) for moderately large $n$ in seconds of computational time, thereby limiting our current ability for realistic power systems simulation with stochastic variability. This calls for fundamentally re-thinking what does it mean to solve the PDE (10) for dynamics (7).

## III. Measure-valued Proximal Recursion

### A. Generalized Gradient Descent

Let $\mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n)$ denote the manifold of joint PDFs supported over the state space $\mathbb{T}^n \times \mathbb{R}^n$, with finite second raw moments. Symbolically,

$$\mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n) := \Big\{\rho : \mathbb{T}^n \times \mathbb{R}^n \mapsto \mathbb{R}_{\geq 0} \mid \int \rho = 1,$$
$$\int \boldsymbol{x}^\top \boldsymbol{x}\, \rho(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} < \infty \text{ for all } \boldsymbol{x} \equiv (\boldsymbol{\theta}, \boldsymbol{\omega})^\top \in \mathbb{T}^n \times \mathbb{R}^n\Big\}. \tag{11}$$

We propose to solve the initial value problem for the PDE (10) by viewing its flow $\rho(t, \boldsymbol{\theta}, \boldsymbol{\omega})$ as the gradient descent of some functional $\Phi : \mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n) \mapsto \mathbb{R}_{\geq 0}$ w.r.t. some distance

$$\mathrm{dist} : \mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n) \times \mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n) \mapsto \mathbb{R}_{\geq 0}.$$

We now explain this idea in detail.

For $k \in \mathbb{N}$, and for some chosen step size $h > 0$, we discretize time as $t_k := kh$, and define the infinite dimensional proximal operator of the functional $h\Phi$ w.r.t. the distance $\mathrm{dist}(\cdot, \cdot)$, given by

$$\mathrm{prox}_{h\Phi}^{\mathrm{dist}}(\varrho_{k-1}) := \underset{\varrho \in \mathcal{P}_2}{\arg\inf}\ \frac{1}{2}\mathrm{dist}^2(\varrho, \varrho_{k-1}) + h\,\Phi(\varrho). \tag{12}$$

Now consider a proximal recursion over the manifold $\mathcal{P}_2$ as

$$\varrho_k = \mathrm{prox}_{h\Phi}^{\mathrm{dist}}(\varrho_{k-1}), \quad k \in \mathbb{N}, \quad \varrho_0(\boldsymbol{x}) := \rho_0(\boldsymbol{x}). \tag{13}$$

Given the PDE (10), we would like to design the functional pair $(\Phi, \mathrm{dist})$ such that the sequence of functions $\{\varrho_k\}_{k\in\mathbb{N}}$ generated by the proximal recursion (13), in the small time step limit, converges to the flow $\rho(t = kh, \boldsymbol{\theta}, \boldsymbol{\omega})$ generated by the PDE initial value problem of interest. In particular,

$$\varrho_k(\boldsymbol{\theta}, \boldsymbol{\omega}) \xrightarrow{h \downarrow 0} \rho(t = kh, \boldsymbol{\theta}, \boldsymbol{\omega}) \text{ in } L^1(\mathbb{T}^n \times \mathbb{R}^n). \tag{14}$$

We remark here that (11), (12), (13), (14) can be written more generally in therms of the joint probability measures instead of PDFs, i.e., even when the underlying measures are not absolutely continuous. Fig. 3 illustrates the idea of approximating the joint PDF trajectory $\rho(t, \boldsymbol{x})$ through the sequence $\{\varrho_k\}_{k\in\mathbb{N}}$ computed from a variational recursion over $\mathcal{P}_2(\mathbb{T}^n \times \mathbb{R}^n)$.

Notice that the proximal recursions given by (12)-(13) define an infinite dimensional gradient descent of the functional $h\Phi$ over $\mathcal{P}_2$ w.r.t. the distance $\mathrm{dist}$. This is reminiscent of the finite dimensional gradient descent, where a gradient flow generated by an ordinary differential equation initial value problem can be recovered as the small time step limit of the sequence of vectors generated by a standard Euclidean proximal recursion; see e.g., [35, Sec. I].

That the flow generated by a Fokker-Planck PDE initial value problem can be recovered from a variational recursion of the form (13) was first proposed in [40], showing that when the drift in the sample path dynamics is a gradient vector field and the diffusion is a scalar multiple of identity matrix, then $\mathrm{dist}(\cdot, \cdot)$ can be taken as the Wasserstein-2 metric arising in the theory of optimal transport [41] with $\Phi(\cdot)$ as the free energy functional. In particular, the functional $\Phi$ serves as a Lyapunov functional in the sense $\frac{\mathrm{d}}{\mathrm{d}t}\Phi < 0$ along the transient solution of the Fokker-Planck PDE initial value problem. This idea has since been generalized to many other types of PDE initial value problems, see e.g., [42], [43].

The algorithmic appeal of the proximal recursion (13) is that it opens up the possibility to compute the solution of the PDE initial value problem via recursive convex minimization. A point cloud-based proximal algorithm was proposed in [34], [35] which was reported to have very fast runtime. Notice that even though the drift in (7) is *not* a gradient vector field, the algorithm in [35, Sec. V.B] constructed a pair $(\Phi, \mathrm{dist})$ such that (13) provably approximates the transient solution of the corresponding kinetic Fokker-Planck PDE with guarantee (14). However, that algorithm cannot be applied to (10) as is. The reasons are explained next.

### B. Statistical Mechanics Perspective

A new difficulty for our SDE (7) is that we have *anisotropic* degenerate diffusion, i.e., the strengths of the noise acting in the last $n$ components of (7) are nonuniform since $\boldsymbol{M}^{-1}\boldsymbol{\Sigma}$ is not identity. This complicates the matter because the construction of the functional $\Phi$ in (13) is usually motivated via free energy considerations utilizing the structure of the *stationary*

PDF $\rho_\infty(\boldsymbol{\theta},\boldsymbol{\omega})$ for (10). The $\rho_\infty$ is, in turn, guaranteed to be a unique *Boltzmann distribution* of the form*

$$\rho_\infty(\boldsymbol{\theta},\boldsymbol{\omega}) = \frac{1}{Z}\exp\left(-\beta H\right), \tag{15a}$$

$$H(\boldsymbol{\theta},\boldsymbol{\omega}) := V(\boldsymbol{\theta}) + \frac{1}{2}\langle\boldsymbol{\omega},\boldsymbol{M}\boldsymbol{\omega}\rangle, \tag{15b}$$

*if and only if* the so-called *Einstein relation* [44], [45] holds:

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top = \beta^{-1}\left(\boldsymbol{\Gamma}+\boldsymbol{\Gamma}^\top\right) \quad \text{for some } \beta > 0. \tag{16}$$

In our case, $\boldsymbol{\Sigma},\boldsymbol{\Gamma}$ are positive diagonal, and (16) is equivalent to the proportionality constraint: $\sigma_i^2 \propto \gamma_i$ for all $i = 1,\dots,n$.

In the power systems context, we cannot relate the damping coefficients $\gamma_i$ with the squared intensities of stochastic forcing $\sigma_i^2$ for the generators. Thus, (16) will not hold in practice, meaning either we cannot guarantee existence-uniqueness for $\rho_\infty$, or even if $\rho_\infty$ exists, it will not be of the form (15). On one hand, this implies that our construction of $\Phi$ may not be guided by free energy considerations. On the other hand, since we are only interested in computing the transient joint PDFs, i.e., *non-equilibrium* statistical mechanics, the lack of a fluctuation-dissipation relation like (16) should not be a fundamental impediment in setting up a recursion such as (13). We next show that a simple change of variable can indeed circumvent this issue.

### C. From Anisotropic to Isotropic Degenerate Diffusion

Consider the $2n \times 2n$ matrix

$$\boldsymbol{\Psi} := \boldsymbol{I}_2 \otimes \left(\boldsymbol{M}\boldsymbol{\Sigma}^{-1}\right), \tag{17}$$

and define the invertible linear map

$$\begin{pmatrix}\boldsymbol{\theta}\\ \boldsymbol{\omega}\end{pmatrix} \mapsto \begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix} := \boldsymbol{\Psi}\begin{pmatrix}\boldsymbol{\theta}\\ \boldsymbol{\omega}\end{pmatrix}. \tag{18}$$

Applying Itô's lemma [46, Ch. 4.2] to the map (18), and using (7), we find that the transformed state vector $(\boldsymbol{\xi},\boldsymbol{\eta})^\top$ solves the Itô SDE

$$\begin{pmatrix}\mathrm{d}\boldsymbol{\xi}\\ \mathrm{d}\boldsymbol{\eta}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\eta}\\ -\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi}) - \nabla_{\boldsymbol{\eta}}F\left(\boldsymbol{\eta}\right)\end{pmatrix}\mathrm{d}t + \begin{pmatrix}\boldsymbol{0}_{n\times n}\\ \boldsymbol{I}_n\end{pmatrix}\mathrm{d}\boldsymbol{w}, \tag{19}$$

where the potentials

$$U(\boldsymbol{\xi}) := \sum_{i=1}^{n}\frac{1}{\sigma_i}P_i\xi_i + \sum_{i,j=1}^{n}\frac{m_i}{\sigma_i^2}k_{ij}\left(1-\cos\left(\frac{\sigma_i}{m_i}\xi_i + \frac{\sigma_j}{m_j}\xi_j - \varphi_{ij}\right)\right), \tag{20a}$$

$$F(\boldsymbol{\eta}) := \frac{1}{2}\langle\boldsymbol{\eta},\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta}\rangle. \tag{20b}$$

Notice that (19) is a mixed conservative-dissipative SDE with *isotropic* degenerate diffusion. In particular, the pushforward of the known initial joint PDF (9) via $\boldsymbol{\Psi}$, is given by

$$\tilde{\rho}_0(\boldsymbol{\xi},\boldsymbol{\eta}) := \boldsymbol{\Psi}_\sharp\rho_0 = \frac{\rho_0\left(\boldsymbol{\Psi}^{-1}\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}\right)}{|\det\left(\boldsymbol{\Psi}\right)|} = \frac{\rho_0\left(\boldsymbol{\Sigma}\boldsymbol{M}^{-1}\boldsymbol{\xi},\boldsymbol{\Sigma}\boldsymbol{M}^{-1}\boldsymbol{\eta}\right)}{\left(\displaystyle\prod_{j=1}^{n}m_j/\sigma_j\right)^2}, \tag{21}$$

*here $Z$ is a normalizing constant known as the "partition function".

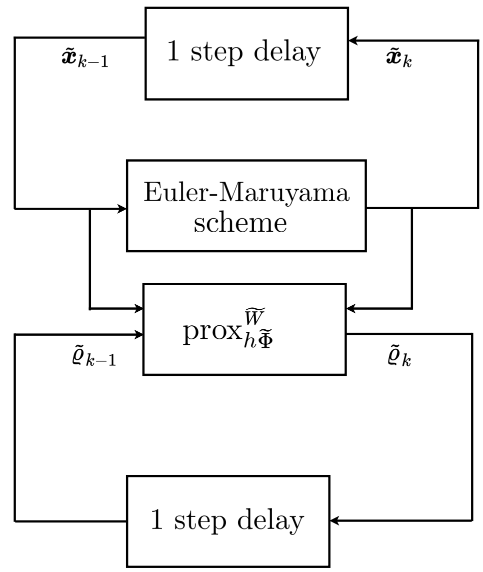


Fig. 4: Schematic of the proposed proximal algorithm for propagating the joint state PDF as probability weighted scattered point cloud $\{\tilde{\boldsymbol{x}}_k^i,\tilde{\varrho}_k^i\}_{i=1}^N$. The states $\{\tilde{\boldsymbol{x}}_k^i\}_{i=1}^N$ are updated by the Euler-Maruyama scheme applied to (7); the corresponding probability weights $\{\tilde{\varrho}_k^i\}_{i=1}^N$ are updated via discrete version of the proximal recursion (26) as detailed in Sec. IV.

where we used the standard properties of the Kronecker product. The transient joint state PDF $\tilde{\rho}(t,\boldsymbol{\xi},\boldsymbol{\eta})$ corresponding to (19) solves the PDE initial value problem

$$\frac{\partial\tilde{\rho}}{\partial t} = -\langle\boldsymbol{\eta},\nabla_{\boldsymbol{\xi}}\tilde{\rho}\rangle + \nabla_{\boldsymbol{\eta}}\cdot\left(\tilde{\rho}\left(\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi}) + \nabla_{\boldsymbol{\eta}}F(\boldsymbol{\eta})\right)\right) + \frac{1}{2}\Delta_{\boldsymbol{\eta}}\tilde{\rho}, \tag{22a}$$

$$\tilde{\rho}(t=t_0,\boldsymbol{\xi},\boldsymbol{\eta}) = \underbrace{\tilde{\rho}_0(\boldsymbol{\xi},\boldsymbol{\eta})}_{\text{from (21)}}. \tag{22b}$$

In other words, (22) is the macroscopic dynamics corresponding to the sample path dynamics (19).

Since (22a) is a kinetic Fokker-Planck PDE with isotropic degenerate diffusion, our strategy is to perform a proximal recursion of the form (13) for (22) in $(\boldsymbol{\xi},\boldsymbol{\eta})$ coordinates, and then to pushforward the resulting joint PDFs via $\boldsymbol{\Psi}^{-1}$ to the original state space. This is what we detail next.

### D. Proximal Update

Looking at (20) and (22a), it is natural to consider an energy functional of the form

$$\Phi(\tilde{\rho}) := \int_{\mathbb{T}^n\times\mathbb{R}^n}\left(\boldsymbol{M}^{-1}\boldsymbol{\Gamma}U(\boldsymbol{\xi}) + F\left(\boldsymbol{\eta}\right) + \frac{1}{2}\log\tilde{\rho}\right)\tilde{\rho}\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta}, \tag{23}$$

which is the sum of a potential energy (expected value of scaled $U$), a weighted kinetic energy (expected value of $F$), and an internal energy (scaled negative entropy, the entropy being $-\int\tilde{\rho}\log\tilde{\rho}$). Indeed, it can be shown that (Appendix A) the functional (23) is a Lyapunov-like functional, i.e., $\Phi$ is decreasing along the solution of (22a).

However, unlike the gradient drift case mentioned in Sec. III-A, it is not possible to express the right hand side of (22a) as the Wasserstein gradient of the functional (23). To see this, recall that the Wasserstein gradient is defined as [42, Ch. 8]

$$\nabla^{\text{Wasserstein}}\Phi := -\nabla\cdot\left(\rho\nabla\frac{\delta\Phi}{\delta\rho}\right),$$

where $\nabla$ denotes the standard Euclidean gradient w.r.t. the vector $(\boldsymbol{\xi},\boldsymbol{\eta})^{\top}$, and $\frac{\delta}{\delta\rho}$ denotes the functional derivative. From (33), (31) and (34a) in Appendix A, it is clear that the vector field $\boldsymbol{v}$ defined in (31) is not equal to the negative of $\nabla\dfrac{\delta\Phi}{\delta\rho}$. Thus, even though the functional (23) decreases along the transient solution of (22a), we cannot interpret the flow generated by (22) as the Wasserstein gradient flow of the functional (23). Thus, in (12), we cannot construct $(\Phi,\mathrm{dist})$ by pairing (23) with the Wasserstein metric.

To set up a variational recursion of the form (12) for (22), we set $(\Phi,\mathrm{dist})\equiv\left(\widetilde{\Phi},\widetilde{W}\right)$ where

$$\widetilde{\Phi}(\tilde{\rho}) := \int_{\mathbb{T}^n\times\mathbb{R}^n}\left(F(\boldsymbol{\eta})+\frac{1}{2}\log\tilde{\rho}\right)\tilde{\rho}\,\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{\eta}, \tag{24}$$

and

$$\widetilde{W}^2\left(\tilde{\varrho},\tilde{\varrho}_{k-1}\right) := \inf_{\pi\in\Pi(\tilde{\varrho},\tilde{\varrho}_{k-1})}\int_{\mathbb{T}^{2n}\times\mathbb{R}^{2n}}\Big\{\|\overline{\boldsymbol{\eta}}-\boldsymbol{\eta}+h\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi})\|_2^2 + 12\left\|\frac{\overline{\boldsymbol{\xi}}-\boldsymbol{\xi}}{h}-\frac{\overline{\boldsymbol{\eta}}+\boldsymbol{\eta}}{2}\right\|_2^2\Big\}\,\mathrm{d}\pi(\boldsymbol{\xi},\boldsymbol{\eta},\overline{\boldsymbol{\xi}},\overline{\boldsymbol{\eta}}). \tag{25}$$

In (25), $\Pi\left(\tilde{\varrho},\tilde{\varrho}_{k-1}\right)$ denotes the collection of all joint PDFs supported on $\mathbb{T}^{2n}\times\mathbb{R}^{2n}$ that have finite second raw moments, with the first marginal $\tilde{\varrho}$, and the second marginal $\tilde{\varrho}_{k-1}$.

That the sequence of functions $\{\tilde{\varrho}_k\}$ for $k\in\mathbb{N}$, generated by the proximal recursion

$$\begin{aligned}\tilde{\varrho}_k &= \mathrm{prox}_{h\widetilde{\Phi}}^{\widetilde{W}}(\tilde{\varrho}_{k-1})\\ &\equiv \underset{\tilde{\rho}\in\mathcal{P}_2}{\arg\inf}\ \frac{1}{2}\widetilde{W}^2\left(\tilde{\varrho},\tilde{\varrho}_{k-1}\right)+h\,\widetilde{\Phi}(\tilde{\rho}),\quad \tilde{\varrho}_0 := \tilde{\rho}_0,\end{aligned} \tag{26}$$

converges to the flow generated by (22), i.e.,

$$\tilde{\varrho}_k(\boldsymbol{\xi},\boldsymbol{\eta}) \xrightarrow{h\downarrow 0} \tilde{\varrho}(t=kh,\boldsymbol{\xi},\boldsymbol{\eta}) \text{ in } L^1\left(\mathbb{T}^n\times\mathbb{R}^n\right),$$

was established in [48]. To numerically perform the recursion (26), we employ the proximal algorithm proposed in [35] with finite number of samples, as explained next.

## IV. Proximal Algorithm

We solve (26) by recursively updating the probability-weighted scattered point clouds $\{\tilde{\boldsymbol{x}}_k^i,\tilde{\varrho}_k^i\}_{i=1}^N$ where

$$\tilde{\boldsymbol{x}}_k^i := \left(\boldsymbol{\xi}_k^i,\boldsymbol{\eta}_k^i\right)^{\top},\quad i=1,\ldots,N,\quad k\in\mathbb{N}.$$

Thus, $\tilde{\varrho}_k^i$ is the joint PDF value obtained from (26) at $\tilde{\boldsymbol{x}}_k^i$, the $i$th (transformed) state sample at the $k$th time step. The high level schematic of the algorithm is shown in Fig. 4.

In the numerical simulations reported in Sec. V, the states $\{\tilde{\boldsymbol{x}}_k^i\}_{i=1}^N$ are updated by the Euler-Maruyama scheme applied to (7). If one wishes so, the Euler-Maruyama scheme in Fig. 4 may be replaced by other SDE integrators, see e.g., [35, Sec. III.B.2, Remark 1].

**Algorithm 1** Proposed proximal algorithm for $\tilde{\boldsymbol{\varrho}}_{k-1}\mapsto\tilde{\boldsymbol{\varrho}}_k$

**procedure** Prox($\boldsymbol{\varrho}_{k-1},\tilde{\boldsymbol{x}}_{k-1},\tilde{\boldsymbol{x}}_k,h,\varepsilon,N,\delta,L$)
  **for** $i=1$ to $N$ **do**
    $\boldsymbol{\zeta}_{k-1}(i)\leftarrow\exp\left(-\left(\boldsymbol{\eta}_{k-1}^i\right)^{\top}\boldsymbol{\eta}_{k-1}^i-1\right)$
    **for** $j=1$ to $N$ **do**
      $\boldsymbol{C}_k(i,j)\leftarrow\|\,\boldsymbol{\eta}_{k-1}^j-\boldsymbol{\eta}_k^i+h\nabla U(\boldsymbol{\xi}_k^i)\,\|_2^2$
        $+12\left\|\dfrac{\boldsymbol{\xi}_{k-1}^j-\boldsymbol{\xi}_k^i}{h}-\dfrac{\boldsymbol{\eta}_{k-1}^j+\boldsymbol{\eta}_k^i}{2}\right\|_2^2$
    **end for**
  **end for**
  $\boldsymbol{\Gamma}_k\leftarrow\exp\left(-\boldsymbol{C}_k/2\varepsilon\right)$ ▷ elementwise exponential
  $\boldsymbol{z}_0\leftarrow\mathrm{rand}_{N\times 1}$ ▷ random vector of size $N\times 1$
  $\boldsymbol{z}\leftarrow\left[\boldsymbol{z}_0,\boldsymbol{0}_{N\times(L-1)}\right]$ ▷ initialize
  $\boldsymbol{y}\leftarrow\left[\tilde{\boldsymbol{\varrho}}_{k-1}\oslash\left(\boldsymbol{\Gamma}_k\boldsymbol{z}_0\right),\boldsymbol{0}_{N\times(L-1)}\right]$ ▷ initialize
  $\ell=1$
  **while** $\ell\leq L$ **do**
    $\boldsymbol{z}(:,\ell+1)\leftarrow\left(\boldsymbol{\zeta}_{k-1}\oslash\left(\boldsymbol{\Gamma}_k^{\top}\boldsymbol{y}(:,\ell)\right)\right)^{\frac{1}{1+2\varepsilon/h}}$
    $\boldsymbol{y}(:,\ell+1)\leftarrow\tilde{\boldsymbol{\varrho}}_{k-1}\oslash\left(\boldsymbol{\Gamma}_k\boldsymbol{z}(:,\ell+1)\right)$
    **if** $\|\boldsymbol{y}(:,\ell+1)-\boldsymbol{y}(:,\ell)\|_2<\delta$ & $\|\boldsymbol{z}(:,\ell+1)-\boldsymbol{z}(:,\ell)\|_2<\delta$ **then** ▷ error within tolerance
      break
    **else**
      $\ell\leftarrow\ell+1$
    **end if**
  **end while**
**return** $\tilde{\boldsymbol{\varrho}}_k\leftarrow\boldsymbol{z}(:,\ell)\odot\left(\boldsymbol{\Gamma}_k^{\top}\boldsymbol{y}(:,\ell)\right)$ ▷ proximal update
**end procedure**

To numerically perform the proximal updates $\{\tilde{\varrho}_{k-1}^i\}_{i=1}^N\mapsto\{\tilde{\varrho}_k^i\}_{i=1}^N$ for $k\in\mathbb{N}$, we implement an instance of the Algorithm 1 in [35]. The algorithm involves a dualization along with an entropic regularization of the variational update (26), and then solving the same using a fixed point recursion that is provably contractive; we refer the interested readers to [35, Sec. V.B] for details. This enables a nonparametric computation of $\tilde{\varrho}_k^i\equiv\tilde{\varrho}_k\left(\boldsymbol{\xi}_k^i,\boldsymbol{\eta}_k^i\right)$ for $i=1,\ldots,N$.

Finally, we transform the proximal updates back to the $\boldsymbol{x}\equiv(\boldsymbol{\theta},\boldsymbol{\omega})^{\top}$ state space via the pushforward $\boldsymbol{\Psi}^{-1}$ as

$$\begin{aligned}\varrho_k\left(\boldsymbol{\theta}_k^i,\boldsymbol{\omega}_k^i\right) &= \boldsymbol{\Psi}_{\sharp}^{-1}\tilde{\varrho}_k\left(\boldsymbol{\xi}_k^i,\boldsymbol{\eta}_k^i\right)\\ &= \left(\prod_{j=1}^{n}m_j/\sigma_j\right)^2\tilde{\varrho}_k\left(\boldsymbol{M}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_k^i,\boldsymbol{M}\boldsymbol{\Sigma}^{-1}\boldsymbol{\omega}_k^i\right),\end{aligned} \tag{27}$$

for all $i=1,\ldots,N$. The $\varrho_k(\boldsymbol{\theta},\boldsymbol{\omega})$ from (27) approximates $\rho(t,\boldsymbol{\theta},\boldsymbol{\omega})$ (the transient solution of (10)) in the sense (14).

For completeness, the algorithm Prox for updating $\{\tilde{\varrho}_{k-1}^i\}_{i=1}^N\mapsto\{\tilde{\varrho}_k^i\}_{i=1}^N$ is outlined in Algorithm 1. As shown in Fig. 4, this algorithm, at a conceptual level, takes the pre and post update state samples

$$\{\tilde{\boldsymbol{x}}_{k-1}^i\}_{i=1}^N\equiv\{\left(\boldsymbol{\xi}_{k-1}^i,\boldsymbol{\eta}_{k-1}^i\right)\}_{i=1}^N,\ \{\tilde{\boldsymbol{x}}_k^i\}_{i=1}^N\equiv\{\left(\boldsymbol{\xi}_k^i,\boldsymbol{\eta}_k^i\right)\}_{i=1}^N,$$

and $\{\tilde{\varrho}_{k-1}^i\}_{i=1}^N$ as inputs, and outputs the proximal updates $\{\tilde{\varrho}_k^i\}_{i=1}^N$. For each $k\in\mathbb{N}$, the updated probability weighted point clouds $\{\tilde{\boldsymbol{x}}_k^i,\tilde{\varrho}_k^i\}_{i=1}^N$ are then brought back to the original state space via $\boldsymbol{\Psi}^{-1}$ as $\{\boldsymbol{x}_k^i,\varrho_k^i\}_{i=1}^N$, as explained earlier.

The Algorithm 1 also needs input parameters $h, \varepsilon, N, \delta, L$. Specifically, $h$ is the time-step size, $\varepsilon$ is an entropic regularization weight internal to the PROX algorithm, and $N$ is the number of samples. The parameters $\delta$ and $L$ codify the numerical tolerance and maximum number of iterations, respectively, for the while loop in Algorithm 1. Its convergence guarantees can be found in [35, Sec. III.C].

## V. NUMERICAL SIMULATIONS

To illustrate the proposed computational framework, we next provide two numerical simulation case studies. In Sec. V-A, we consider the prediction of stochastic states for the IEEE 14 bus system–our main objective being to highlight the interpretability and utility of the proposed method in power systems engineering. In Sec. V-B, we propagate the joint PDFs over the 100 dimensional state space of a synthetic power network with randomly generated parameters–our intent there is to highlight the scalability of the proposed method.

All simulations were performed in MATLAB R2019b on an iMac with 3.4 GHz Quad-Core Intel Core i5 processor and 8 GB memory.

### A. IEEE 14 Bus System

We consider the Kron-reduced dynamics (1) for the IEEE 14 bus system shown in Fig. 1. In this case, we have $n = 5$ nodes which correspond to the buses 1, 2, 3, 6 and 8 in Fig. 1. We obtained the parameters of the 14 bus system from MATPOWER [55]. The calculation of Kron-reduced admittance matrix and current vector in (2) followed Sec. II-A; see also [51]. The parameters $m_i, \gamma_i$ are obtained from the open-source Python-based library ANDES [56].

We randomly generated the noise coefficients $\sigma_i \in [1, 5]$ as

$$(\sigma_1, \ldots, \sigma_5) = (2.4628, 4.9266, 4.8724, 1.4215, 3.8681).$$

We take the initial joint PDF as

$$\begin{aligned}\rho_0 \equiv \rho(t=0, \boldsymbol{\theta}(0), \boldsymbol{\omega}(0)) &= \vartheta_0\left(\boldsymbol{\theta}(0)\right)\Omega_0\left(\boldsymbol{\omega}(0)\right)\\ &= \underbrace{\left(\prod_{i=1}^{n=5} \frac{\exp\left(\kappa_i \cos\left(2\theta_i(0) - \mu_{\theta_i(0)}\right)\right)}{2\pi I_0(\kappa_i)}\right)}_{=:\vartheta_0(\boldsymbol{\theta}(0))}\\ &\quad \times \underbrace{\operatorname{Unif}\left([-0.1, 0.1] \text{ rad/s}\right)}_{=:\Omega_0(\boldsymbol{\omega}(0))},\end{aligned} \tag{28}$$

whose $\boldsymbol{\theta}$ marginal $\vartheta_0\left(\boldsymbol{\theta}(0)\right)$ is a product von Mises PDF† [52], [53] supported over $\mathbb{T}^5$ with mean angles (in rad)

$$(\mu_{\theta_1}(0), \ldots, \mu_{\theta_5}(0)) = (0, 6.1963, 6.0612, 6.0350, 6.0500), \tag{29}$$

and concentration parameter $\kappa_i \geq 0$ set as

$$(\kappa_1, \ldots, \kappa_5) = (5, 6, 7, 4, 5).$$

†the factor 2 inside the cosine ensures the range $[0, 2\pi)$; it can be dispensed if we instead use the range $[-\pi, \pi)$ for the angular variables, as common in the directional statistics literature [54].

| Parameter description | Values |
|---|---|
| nominal frequency | $f_0 = 60$ Hz |
| inertia | $m_i \in \operatorname{Unif}([2, 12])/2\pi f_0$ |
| damping coefficient | $\gamma_i \in \operatorname{Unif}([20, 30])/2\pi f_0$ |
| diffusion coefficient | $\sigma_i \in \operatorname{Unif}([1, 5])$ |
| tangent of phase shift | $\tan\varphi_{ij} \begin{cases} = 0 & \text{for } i = j \\ \in \operatorname{Unif}([0, 0.25]) & \text{for } i \neq j \end{cases}$ |
| effective power input | $P_i \in \operatorname{Unif}([0, 10])$ |
| coupling coefficient | $k_{ij} \begin{cases} = 0 & \text{for } i = j \\ \in \operatorname{Unif}([0.7, 1.2]) & \text{for } i \neq j \end{cases}$ |

TABLE I: Parameters used for the numerical simulation in Sec. V-B. The indices $i, j \in \{1, \ldots, n\}$ where the number of generators $n = 50$ in Sec. V-B.

In (28), $I_0(\cdot)$ denotes the modified Bessel function of the first kind with order zero, given by $I_0(\kappa) = \sum_{r=0}^{\infty} \frac{(\kappa^2/4)^r}{(r!)^2}$, and can be evaluated via MATLAB command `besseli(0,·)`.

The mean angles (29) were obtained from steady state AC power flow solutions. Larger $\kappa_i$ entails a higher concentration around the mean angle $\mu_{\theta_i}(0)$ (setting $\kappa_i = 0$ implies that $\theta_i(0)$ follows uniform distribution over $[0, 2\pi)$).

In (28), the $\boldsymbol{\omega}$ marginal $\Omega_0\left(\boldsymbol{\omega}(0)\right)$ is a uniform PDF over $[-0.1, 0.1]$ rad/s. Thus, the PDF $\rho_0$ in (28) is supported over $\mathbb{T}^5 \times [-0.1, 0.1]$.

With step size $h = 10^{-3}$, we discretized time $t \in [0, 1]$. Using the algorithm detailed in Sec. IV, we propagated $N = 1000$ random samples from (28), and evaluated at (28), to update the scattered weighted point clouds $\{\boldsymbol{x}_k^i, \varrho_k^i\}_{i=1}^N$ over $\mathbb{T}^5 \times \mathbb{R}^5$ at times $t_k := kh$, $k \in \mathbb{N}$. In Algorithm 1, we used the algorithmic parameters $\varepsilon = 0.05$, $\delta = 10^{-3}$, $L = 300$.

Fig. 5 shows the time snapshots of the $(\theta_i, \omega_i)$ bivariate marginal PDFs for generator $i = 1, \ldots, n = 5$, computed using the weighted point clouds $\{\boldsymbol{x}_k^i, \varrho_k^i\}_{i=1}^N$. Notice that in Fig. 5, the horizontal axes of the subfigures have periodic boundary conditions, i.e., $(\theta_i, \omega_i) \in \mathbb{T} \times \mathbb{R}$ (cylinder). Fig. 6 shows the computational times needed to update $\{\boldsymbol{x}_k^i, \varrho_k^i\}_{i=1}^N$ via the algorithm in Sec. IV.

### B. Synthetic Test System

We next consider a power network with $n = 50$ generators, and propagate the transient joint state PDFs supported over the 100 dimensional state space $\mathbb{T}^{50} \times \mathbb{R}^{50}$. We take the initial joint state PDF at $t = 0$ as

$$\begin{aligned}\rho_0 \equiv \rho(t=0, \boldsymbol{\theta}(0), \boldsymbol{\omega}(0)) &= \operatorname{Unif}\left(([0, 2\pi) \text{ rad})^n\right)\\ &\quad \times \operatorname{Unif}\left([-0.1, 0.1] \text{ rad/s}\right),\end{aligned} \tag{30}$$

and following [30, Sec. 5], randomly generate the parameters as in Table I. These parameter ranges are consistent with the same found in [20], [49], [50].

With $N = 2000$ random samples from the initial joint PDF (30), and with the aforesaid randomly generated parameters,

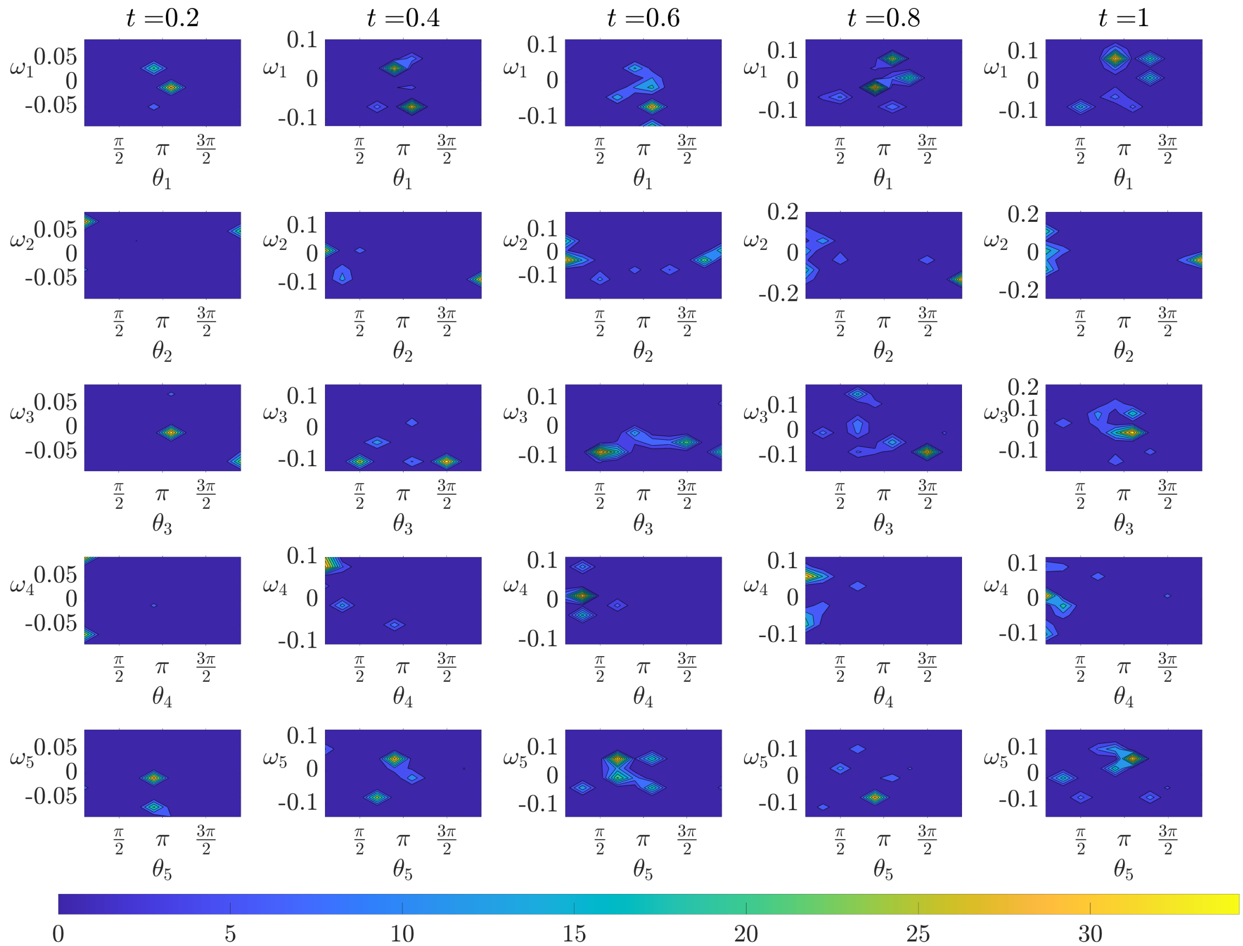


Fig. 5: Time evolution of the $(\theta_i, \omega_i)$ bivariate marginal PDFs of the generator $i = 1, \ldots, 5$, for the Kron-reduced IEEE 14 bus simulation setup described in Sec. V-A. The 5 rows above correspond to the 5 generator nodes; these are buses 1, 2, 3, 6 and 8 in Fig. 1. The columns correspond to the time snapshots. The colorbar at the bottom shows the marginal PDF values. The $\theta_i$ are in rad, the $\omega_i$ are in rad/s.

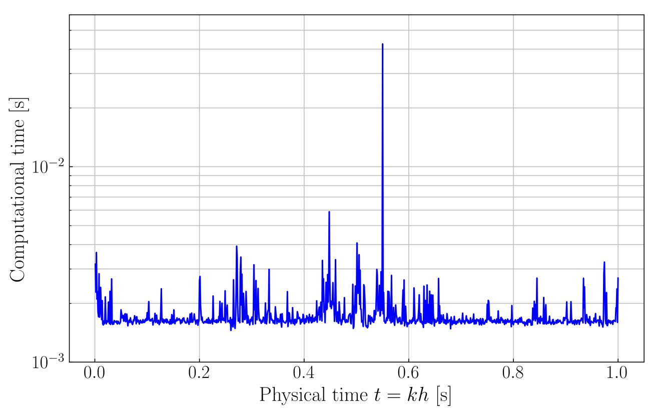


Fig. 6: The computational time for propagating the transient joint state PDFs over $\mathbb{T}^5 \times \mathbb{R}^5$ for the simulation set up in Sec. V-A.

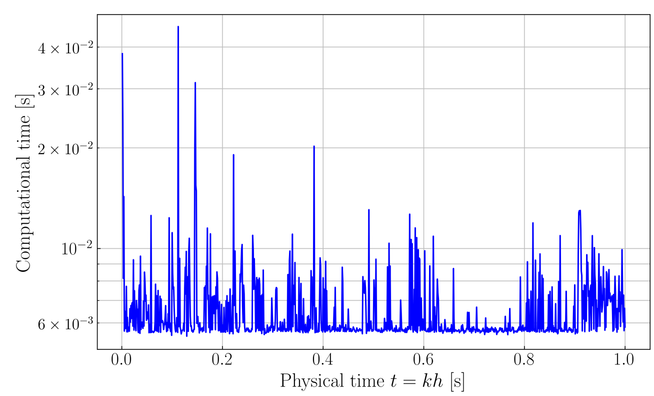


Fig. 7: The computational time for propagating the transient joint state PDFs over $\mathbb{T}^{50} \times \mathbb{R}^{50}$ for the simulation set up in Sec. V-B.

we employed the procedure detailed in Sec. IV for propagating $\{\boldsymbol{x}_k^i, \varrho_k^i\}_{i=1}^N$. As in Sec. V-A, we used $h = 10^{-3}$, $\varepsilon = 0.05$, $\delta = 10^{-3}$, $L = 300$ in Algorithm 1.

Fig. 7 highlights that for computing high dimensional transient joint state PDFs as in here, the proposed variational framework enjoys remarkably fast computational time. Notice that constructing histograms or other conventional density estimators for the joint state PDF by making a grid over the 100 dimensional state space, is computationally prohibitive.

To depict the numerical accuracy, Fig. 8 plots the (time-varying) relative error between the empirical (i.e., Monte Carlo estimate) mean vector $\boldsymbol{\mu}_k^{\text{MC}} \in \mathbb{T}^{50} \times \mathbb{R}^{50}$, and the "proximal mean" vector $\boldsymbol{\mu}_k^{\text{Prox}} \in \mathbb{T}^{50} \times \mathbb{R}^{50}$. The latter was computed using the proximal updates $\{\boldsymbol{x}_k^i, \varrho_k^i\}_{i=1}^N$. Fig. 8 shows that the time varying statistics from these two computation match very well at all times.

## VI. Conclusion

The conclusion goes here.

## Appendix A
## Showing $\Phi(\tilde{\rho})$ is Decreasing along the Flow Generated by (22)

Let $\tilde{\rho}(t,\boldsymbol{\xi},\boldsymbol{\eta})$ be the transient solution of (22). Define

$$\boldsymbol{\ell}_{\boldsymbol{\xi}} := \frac{1}{2}\nabla_{\boldsymbol{\xi}}\log\tilde{\rho}, \quad \boldsymbol{\ell}_{\boldsymbol{\eta}} := \frac{1}{2}\nabla_{\boldsymbol{\eta}}\log\tilde{\rho},$$

and rewrite (22a) as

$$\frac{\partial\tilde{\rho}}{\partial t} = -\nabla\cdot(\tilde{\rho}\boldsymbol{v})$$

where $\nabla$ is the $2n\times 1$ Euclidean gradient w.r.t. the vector $(\boldsymbol{\xi},\boldsymbol{\eta})^\top$, and the vector field

$$\boldsymbol{v} := \begin{pmatrix}\boldsymbol{\eta}\\ -\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi}) - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta} - \boldsymbol{\ell}_{\boldsymbol{\eta}}\end{pmatrix}. \tag{31}$$

Now consider the time derivative of the functional (23) along (22a). Following [47, Proposition 1],

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi = \int_{\mathbb{T}^n\times\mathbb{R}^n}\left\langle\nabla\frac{\delta\Phi}{\delta\tilde{\rho}},\tilde{\rho}\boldsymbol{v}\right\rangle\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{\eta}, \tag{32}$$

wherein the functional derivative

$$\frac{\delta\Phi}{\delta\tilde{\rho}} = \boldsymbol{M}^{-1}\boldsymbol{\Gamma}U(\boldsymbol{\xi}) + F(\boldsymbol{\eta}) + \frac{1}{2} + \frac{1}{2}\log\tilde{\rho}. \tag{33}$$

Recalling the explicit form of $F(\boldsymbol{\eta})$ from (20b), and using (33), we simplify (32) as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Phi &= \int_{\mathbb{T}^n\times\mathbb{R}^n}\left\langle\begin{pmatrix}\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi}) + \boldsymbol{\ell}_{\boldsymbol{\xi}}\\ \boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta} + \boldsymbol{\ell}_{\boldsymbol{\eta}}\end{pmatrix},\begin{pmatrix}\boldsymbol{\eta}\\ -\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi}) - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta} - \boldsymbol{\ell}_{\boldsymbol{\eta}}\end{pmatrix}\right\rangle\tilde{\rho}(t,\boldsymbol{\xi},\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta} &\text{(34a)}\\ &= \int_{\mathbb{T}^n\times\mathbb{R}^n}\Big\{-\|\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta}\|_2^2 - \|\boldsymbol{\ell}_{\boldsymbol{\eta}}\|_2^2 - 2\langle\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta},\boldsymbol{\ell}_{\boldsymbol{\eta}}\rangle + \langle\boldsymbol{\ell}_{\boldsymbol{\xi}},\boldsymbol{\eta}\rangle - \langle\boldsymbol{\ell}_{\boldsymbol{\eta}},\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi})\rangle\Big\}\tilde{\rho}(t,\boldsymbol{\xi},\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta} &\text{(34b)}\\ &= -\int_{\mathbb{T}^n\times\mathbb{R}^n}\|\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta} + \boldsymbol{\ell}_{\boldsymbol{\eta}}\|_2^2\,\tilde{\rho}(t,\boldsymbol{\xi},\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\eta}, &\text{(34c)}\end{aligned}$$

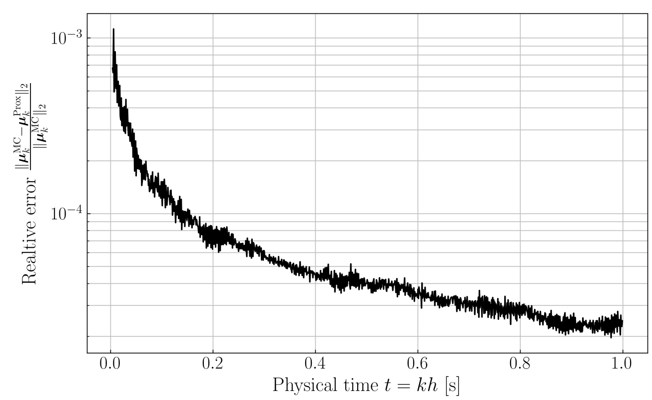


Fig. 8: The relative error between the empirical mean (i.e., Monte Carlo estimate) $\boldsymbol{\mu}_k^{\mathrm{MC}}\in\mathbb{T}^{50}\times\mathbb{R}^{50}$, and the mean $\boldsymbol{\mu}_k^{\mathrm{Prox}}\in\mathbb{T}^{50}\times\mathbb{R}^{50}$ obtained using the proximal updates of the joint state PDFs, for the simulation set up in Sec. V-B.

where (34c) follows from the fact that the integrals involving $\langle\boldsymbol{\ell}_{\boldsymbol{\xi}},\boldsymbol{\eta}\rangle$ and $\langle\boldsymbol{\ell}_{\boldsymbol{\eta}},\nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi})\rangle$ in (34b) are zero, see [47, Appendix D.3].

From (34), it follows that $\frac{\mathrm{d}}{\mathrm{d}t}\Phi\leq 0$, i.e., $\Phi$ is non-increasing along the flow generated by (22). Furthermore, (34c) tells us that $\frac{\mathrm{d}}{\mathrm{d}t}\Phi = 0$ if and only if

$$\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta} + \boldsymbol{\ell}_{\boldsymbol{\eta}} = 0,$$

i.e., $\Phi$ is stationary if and only if $\tilde{\rho}$ has $\boldsymbol{\eta}$ marginal $\int\tilde{\rho}\,\mathrm{d}\boldsymbol{\xi}\propto\exp\left(-\boldsymbol{\eta}^\top\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta}\right)$. Following the steps as in [47, Appendix E.1], this is achieved at the stationary solution of (22a) with $\boldsymbol{\xi}$ marginal $\int\tilde{\rho}\,\mathrm{d}\boldsymbol{\eta}\propto\exp\left(-2U(\boldsymbol{\xi})\right)$. In other words, the stationary solution of (22a) must be of the form

$$\tilde{\rho}_\infty(\boldsymbol{\xi},\boldsymbol{\eta})\propto\exp\left(-2U(\boldsymbol{\xi}) + \boldsymbol{\eta}^\top\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta}\right),$$

and is exactly where $\frac{\mathrm{d}}{\mathrm{d}t}\Phi = 0$. Therefore, $\frac{\mathrm{d}}{\mathrm{d}t}\Phi < 0$ along the transient solution of (22a), and $= 0$ at the stationary solution of (22a).

## References


[1] M. Schwalbe, *Mathematical Sciences Research Challenges for the Next-generation Electric Grid: Summary of a Workshop*, National Academies Press, 2015.

[2] R.N. Allan, R. Billinton, A.M. Breipohl, and C.H. Grigg, "Bibliography On the Application of Probability Methods in Power Systems Reliability Evaluation 1987–1991", *IEEE Transactions on Power Systems*, Vol. 9, pp. 41–49, 1994.

[3] ——, "Bibliography On the Application of Probability Methods in Power Systems Reliability Evaluation 1992–1996", *IEEE Transactions on Power Systems*, Vol. 14, pp. 51–57, 1999.

[4] K.J. Timko, A. Bose, and P.M. Anderson, "Monte Carlo Simulation of Power System Stability", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-102, No. 10, pp. 3453–3459, 1983.

[5] Z.Y. Dong, J.H. Zhao, and D.J. Hill, "Numerical Simulation for Stochastic Transient Stability Assessment", *IEEE Transactions on Power Systems*, Vol. 27, No. 4, pp. 1741–1749, 2012.

[6] M. Perninge, F. Lindskog, and L. Soder, "Importance Sampling of Injected Powers for Electric Power System Security Analysis", *IEEE Transactions on Power Systems*, Vol. 27, No. 1, pp. 3–11, 2012.

[7] T. Odun-Ayo, and M.L. Crow, "Structure-preserved Power System Transient Stability using Stochastic Energy Functions", *IEEE Transactions on Power Systems*, Vol. 27, No. 3, pp. 1450–1458, 2012.

[8] G. Ghanavati, P.D.H. Hines, and T.I. Lakoba, "Identifying Useful Statistical Indicators of Proximity to Instability in Stochastic Power Systems", *IEEE Transactions on Power Systems*, Vol. 31, No. 2, pp. 1360–1368, 2016.

[9] D. Apostolopoulou, A.D. Domínguez-García, P.W. Sauer, "An Assessment of the Impact of Uncertainty on Automatic Generation Control Systems", *IEEE Transactions on Power Systems*, Vol. 31, No. 4, pp. 2657–2665, 2016.

[10] C.O. Nwankpa, S.M. Shahidehpour, and Z. Schuss, "A Stochastic Approach to Small Disturbance Stability Analysis", *IEEE Transactions on Power Systems*, Vol. 7, No. 4, pp. 1519–1528, 1992.

[11] J.L. Rueda, D.G. Colomé, and I. Erlich, "Assessment and Enhancement of Small Signal Stability Considering Uncertainties", *IEEE Transactions on Power Systems*, Vol. 24, No. 1, pp. 198–207, 2009.

[12] H. Huang, C. Chung, K.W. Chan, and H. Chen, "Quasi-Monte Carlo Based Probabilistic Small Signal Stability Analysis for Power Systems with Plug-in Electric Vehicle and Wind Power Integration", *IEEE Transactions on Power Systems*, Vol. 28, No. 3, pp. 3335–3343, 2013.

[13] S.V. Dhople, Y.C. Chen, L. DeVille, A.D. Domínguez-García, "Analysis of Power System Dynamics Subject to Stochastic Power Injections", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 60, No. 12, pp. 3341–3353, 2013.

[14] J.R. Hockenberry, and B.C. Lesieutre, "Evaluation of Uncertainty in Dynamic Simulations of Power System Models: The Probabilistic Collocation Method", *IEEE Transactions on Power Systems*, Vol. 28, No. 3, pp. 3335–3343, 2013.

[15] Y. Xu, L. Mili, A. Sandu, M.R. von Spakovsky, and J. Zhao, "Propagating Uncertainty in Power System Dynamic Simulations using Polynomial Chaos", *IEEE Transactions on Power Systems*, Vol. 34, No. 1, pp. 338–348, 2019.

[16] P. Ju, H. Li, C. Gan, Y. Liu, Y. Yu, and Y. Liu, "Analytical Assessment for Transient Stability Under Stochastic Continuous Disturbances", *IEEE Transactions on Power Systems*, Vol. 33, No. 2, pp. 2004–2014, 2018.

[17] P. Ju, H. Li, X. Pan, C. Gan, Y. Liu, and Y. Liu, "Stochastic Dynamic Analysis for Power Systems Under Uncertain Variability", *IEEE Transactions on Power Systems*, Vol. 33, No. 4, pp. 3789–3799, 2018.

[18] A. Pai, *Energy Function Analysis for Power System Stability*, Springer Science & Business Media, 1989.

[19] H-D. Chang, C-C. Chu, and G. Cauley, "Direct Stability Analysis of Electric Power Systems using Energy Functions: Theory, Applications, and Perspective", *Proceedings of the IEEE*, Vol. 83, No. 11, pp. 1497–1529, 1995.

[20] P.W. Sauer, and M.A. Pai, *Power System Dynamics and Stability*, Prentice-Hall, 1998.

[21] D.A. Maldonado, M. Schanen, and M. Anitescu, "Uncertainty Propagation in Power System Dynamics with the Method of Moments", *2018 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, 2018. doi: 10.1109/PESGM.2018.8586023

[22] A. Dimitrovski, and K. Tomsovic, "Boundary Load Flow Solutions", *IEEE Transactions on Power Systems*, Vol. 19, No. 1, pp. 348–355, 2004.

[23] I.A. Hiskens, and J. Alseddiqui, "Sensitivity, Approximation, and Uncertainty in Power System Dynamic Simulation", *IEEE Transactions on Power Systems*, Vol. 21, No. 4, pp. 1808–1820, 2006.

[24] Y.C. Chen, and A.D. Dominguez-Garcia, "A Method to Study the Effect of Renewable Resource Variability on Power System Dynamics", *IEEE Transactions on Power Systems*, Vol. 27, No. 4, pp. 1978–1989, 2012.

[25] M. Althoff, "Formal and Compositional Analysis of Power Systems using Reachable Sets", *IEEE Transactions on Power Systems*, Vol. 29, No. 5, pp. 2270–2280, 2014.

[26] H. Choi, P.J. Seiler, and S.V. Dhople, "Propagating Uncertainty in Power-System DAE Models with Semidefinite Programming", *IEEE Transactions on Power Systems*, Vol. 32, No. 4, pp. 3146–3156, 2017.

[27] H. Choi, P.J. Seiler, and S.V. Dhople, "Propagating Uncertainty in Power Flow With the Alternating Direction Method of Multipliers", *IEEE Transactions on Power Systems*, Vol. 33, No. 4, pp. 4124–4133, 2018.

[28] R.E. Bellman, *Dynamic Programming*, Courier Dover Publications, 1957.

[29] N. Parikh, and S. Boyd, "Proximal Algorithms", *Foundations and Trends in Optimization*, Vol. 1, No. 3, pp. 127–239, 2014.

[30] F. Dörfler, and F. Bullo, "Synchronization and Transient Stability in Power Networks and Nonuniform Kuramoto Oscillators", *SIAM Journal on Control and Optimization*, Vol. 50, No. 3, pp. 1616–1642, 2012.

[31] F.A. Rodrigues, T.K. Peron, P. Ji, and J. Kurths, "The Kuramoto Model in Complex Networks", *Physics Reports*, Vol. 610, pp. 1–98, 2016.

[32] F. Dörfler, M. Chertkov, and F. Bullo, "Synchronization in Complex Oscillator Networks and Smart Grids", *Proceedings of the National Academy of Sciences*, Vol. 100, No. 6, pp. 2005–2010, 2013.

[33] T. Ishizaki, A. Chakrabortty, and J-I. Imura. "Graph-Theoretic Analysis of Power Systems", *Proceedings of the IEEE*, Vol. 106, No. 5, pp. 931–952, 2018.

[34] K.F. Caluya, and A. Halder, "Proximal Recursion for Solving the Fokker-Planck Equation", *Proceedings of the 2019 American Control Conference*, pp. 4098–4103, 2019.

[35] K.F. Caluya, and A. Halder, "Gradient Flow Algorithms for Density Propagation in Stochastic Systems", *IEEE Transactions on Automatic Control*, Vol. 65, No. 10, pp. 3991–4004, 2019.

[36] A. Halder, K.F. Caluya, B. Travacca, and S.J. Moura, "Hopfield Neural Network Flow: A Geometric Viewpoint", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 11, pp. 4869–4880, 2020.

[37] K.F. Caluya, and A. Halder, "Wasserstein Proximal Algorithms for the Schrödinger Bridge Problem: Density Control with Nonlinear Drift", *IEEE Transactions on Automatic Control*, 2021.

[38] K.F. Caluya, and A. Halder,"Reflected Schrödinger Bridge: Density Control with Path Constraints", *American Control Conference*, 2021.

[39] K. Wang, and M.L. Crow, "The Fokker-Planck Equation for Power System Stability Probability Density Function Evolution", *IEEE Transactions on Power Systems*, Vol. 28, No. 3, pp. 2994–3001, 2013.

[40] R. Jordan, D. Kinderlehrer, and F. Otto, "The Variational Formulation of the Fokker–Planck Equation", *SIAM journal on mathematical analysis*, Vol. 29, pp. 1–17, 1998.

[41] C. Villani, *Topics in optimal transportation*, No. 58, American Mathematical Society, 2003.

[42] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*, Springer Science & Business Media, 2008.

[43] F. Santambrogio, "{Euclidean, Metric, and Wasserstein} Gradient Flows: An Overview", *Bulletin of Mathematical Sciences*, Vol. 7, No. 1, pp. 87–154, 2017.

[44] D.B. Hernandez, and M. Pavon, "Equilibrium Description of A Particle System in A Heat Bath", *Acta Applicandae Mathematica*, Vol. 14, No. 3, pp. 239–258, 1989.

[45] Y. Chen, T.T. Georgiou, and M. Pavon, "Fast Cooling for A System of Stochastic Oscillators", *Journal of Mathematical Physics*, Vol. 56, No. 11, pp. 113302, 2015.

[46] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, *Springer Science & Business Media*, 2013.

[47] W. Krichene, K.F. Caluya, and A. Halder, "Global Convergence of Second-order Dynamics in Two-layer Neural Networks", *arXiv preprint arXiv:2007.06852*, 2020.

[48] M.H. Duong, M.A. Peletier, and J. Zimmer, "Conservative-dissipative Approximation Schemes for A Generalized Kramers Equation", *Mathematical Methods in the Applied Sciences*, Vol. 37, No. 16, pp. 2517–2540, 2014.

[49] P.M. Anderson, and A.A. Fouad, *Power System Control and Stability*, Iowa State University Press, Ames, IA, 1977.

[50] P. Kundur, *Power System Stability and Control*, McGraw-Hill, New York, 1994.

[51] F. Dorfler, and F. Bullo, "Kron Reduction of Graphs with Applications to Electrical Networks", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 60, No. 1, pp. 150–163, 2012.

[52] K.V. Mardia, G. Hughes, C.C. Taylor, H. Singh, "A Multivariate von Mises Distribution with Applications to Bioinformatics", *Canadian Journal of Statistics*, Vol. 36, No. 1, pp. 99–109, 2008.

[53] K.V. Mardia, and J. Voss, "Some Fundamental Properties of a Multivariate von Mises Distribution", *Communications in Statistics - Theory and Methods*, Vol. 43, No. 6, pp. 1132–1144, 2014.

[54] K.V. Mardia, and P.E. Jupp, *Directional Statistics*, Wiley, 1st edition, 1999.

[55] R.D. Zimmerman, C.E. Murillo-Sánchez, and R.J. Thomas, "MATPOWER: Steady-state Operations, Planning, and Analysis Tools for Power Systems Research and Education", *IEEE Transactions on Power Systems*, Vol. 26, No. 1, pp. 12–19, 2010.

[56] H. Cui, F. Li, and K. Tomsovic, "Hybrid Symbolic-Numeric Framework for Power System Modeling and Analysis," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1373-1384, 2021.