

Revision of Manuscript Number: TPWRS-01391-2021.R1

August 12, 2022

Journal: IEEE Transactions on Power Systems

Author: Abhishek Halder, Kenneth F. Caluya, Pegah Ojaghi, Xinbo Geng

Response to decision:

The perceptive input from all reviewers and the AE are much appreciated.

Please find below itemized responses to the individual comments addressing the concerns and detailing the revisions. All the edits in the revised manuscript are marked in blue.

Sincerely,
The authors

Response to AE:

AE.1. Comments by AE:

Authors are requested to revise and resubmit the manuscript after addressing all reviewer comments that persist.

Response: We have addressed all the reviewers' comments as detailed itemized responses below. We are thankful to the AE and the reviewers for their helpful comments and efforts contributing to elevate the quality of our manuscript.

Response to Reviewer #1:

R1.1. Comment by reviewer:

No further comments.

Response: Thank you for your time and efforts to help improve the manuscript.

Response to Reviewer #2:

R2.1. Comments by reviewer:

This reviewer appreciates the authors' effort in revising the manuscript based on the feedback in the previous round while clarifying and addressing the remarks with due diligence. The revised version reads well and adequately takes into account several relevant facets. The authors are kindly requested to proofread the article prior to final submission.

Response: Thank you, we have thoroughly proofread the article. We appreciate your comments from previous rounds to help us improve the manuscript.

Response to Reviewer #3:

R3.1. Comments by reviewer:

First and very importantly, the authors are suggested to readdress my previous two questions.

(1) Regarding my previous question about the classical model, I need to reemphasize here that the model is too simplistic that does not sufficiently reflect the responses from the exciter and the governor. Therefore, for a full paper on IEEE transactions on power systems, this model is incomplete, and cannot reflect reality. Also, for such an over-simplistic model, I still doubt the scalability of the method. Unfortunately, this problem is not addressed and improved after the revision.

(2) Regarding my previous questions about the comparison studies with the state-of-the-art method, indeed, I am very disappointed to find that the revised version still has added no comparison study with the state-of-the-art method at all. The ONLY comparison is still with Monte Carlo.

Response: (1) We fully acknowledge the reviewer's comment about the simplicity of our SDE model. In the revised manuscript's Conclusion (last paragraph) and in the previous round's rebuttal document, we had explicitly mentioned about incorporating more complex models in our future work. As we explained in previous round, our contribution is nonparametric joint PDF propagation methodology and even for our simple model, presents significant technical challenge as we explain in Sec. III-A and III-B. We do not believe it is possible to design a proximal recursion for more complex models if one does not understand how to do so in simpler setting, as there is no "one-size-fits-all" proximal algorithm in the manifold of PDFs for arbitrary nonlinear SDEs. In particular, it is impossible for us to design such an algorithm for more complex SDE models by simply revising our paper—this requires completely new line of research and is out-of-scope of the present work.

(2) Our ground truth state-of-the-art needs to be non-parametric. We had explained that Monte Carlo remains the state-of-the-art ground truth since e.g., it is impossible to perform an FEM simulation on a general purpose computer solving the Kolmogorov's forward PDE even for the "simpler" 5 generator model in Sec. V-A as it requires us to grid a 10 dimensional space and 1 dimensional time, which is what the transient joint PDFs are supported on. This is why ref. [45] did a 2D state + 1D time simulation, and the authors are not aware of any further attempts in the power systems literature on solving a higher dimensional Kolmogorov's forward PDE subject to the nonlinear power system dynamics.

R3.2. Comments by reviewer:

Second and most importantly, I disagree with the theory and the claim behind the SDE model used in this paper.1. Here, I am very familiar with the Ito SDE model in power systems that you cited here in your refs [16], [17], [45].

Although the authors claim the model you used here in equation (7) is the same as the ones used in ref[16] and ref[17] and extended from ref[45], I would like to tell the authors: NO, you are not using the same one. 16][17][45] are all right, but the one used in this article is incorrect. Please pay very special attention to [16][17], they only add uncertainties directly on differential equations, namely \dot{w} here, but not from P and Q . This is because transferring the power system model from DAE to Ito-SDE is still an unsolved problem where P and Q is contained in algebraic equations while Ito-SDE only considers differential equations. Also, [45] is one machine system that does not even have a network. That is different from your problem. Here, the authors might argue that you are using the Kron reductions to remove algebraic equations. However, I have to tell the authors the Kron reductions to apply to the SDE model here. It is because the Kron reduction is relying on constant P and Q to derive a constant admittance matrix. However, the P and Q in this manuscript are also assumed to be uncertain, and also claimed in the response letter to be time-varying stochastic processes. Then, its assumptions do not hold. Then, its Ito SDE form is inconsistent. I highly suggest to the authors study the following article carefully and learn more about the basics of Kron reduction.

Response: The reviewer may have misunderstood our computational framework. We are considering P, Q to be uncertain as in epistemic/parametric uncertainty, but not as in aleatoric uncertainty (i.e., not as in considering P, Q to have their own SDEs). This is definitely a simplifying assumption, as the reviewer aptly points out and we find the reviewer's suggested reference very insightful. Please note that the suggested reference is published on IEEE Xplore on 23rd June, 2022, and we were not aware of this work as this manuscript has been under review since August, 2021. Our future work on more sophisticated dynamics models will definitely use the suggested framework. We are indebted to the reviewer for this suggestion.

R3.3. Comments by reviewer:

The simulation results are weird. Just a simple check at fig 8. How a power system dynamic simulation with line 13 tripped can has no oscillations at all within the first 1 second? If so, it has no practical meaning at all.

Response: That is not exactly what Fig. 8 says since it is plotting the ensemble mean and box-whisker plots but not the individual 1000 sample paths. In fact, the whiskers show significant spread of probability mass indicating large excursions made by the stochastic sample paths. Please note that this Fig. 8 was included by the authors in the first round of revision in response to the specific request made by the reviewer's comment R2.5 in that round. Furthermore, in the revised manuscript's p. 8, right column, we specifically discuss how Fig. 7(c) depict the post-fault marginals having larger spread or dispersion of the probability mass. There, even though the means stay around zero, the spreading of probability mass implies larger probability of having sample paths taking excursions away from zero, compared to the nominal Case I. In the same paragraph, we also point out how similar information can be gleaned by comparing the fourth row in Fig. 5 with the same in Fig. 6.

R3.4. Comments by reviewer:

Third, my engineering motivation for me is not clear. The authors reclaim multiple times about the joint pdf for the output system state. I understand that we consider joint pdf in the model inputs and understand its practical meaning. However, among the current literature, to my best knowledge, I saw no previous efforts in obtaining the "joint" pdf of outputs. They only seed for marginal pdf for output states. So an important question is what motivates us to get joint pdf, instead of individual pdf, for model output? For which exactly applications, you must know the joint pdf of the output states? What is its practical meaning and why it is important?

Response: The reviewer is spot on: the existing power system dynamics literature mostly do not compute the joint PDFs, and directly try to approximate statistics of derived quantities such as marginal PDFs or moments. The fact that our proposed methodology enables joint computation, is a feature not a bug. Here is why:

(i) Even if one is only interested in univariate or bivariate marginals, their computation is more accurate if the joint is first computed and then the marginals are obtained from the joint PDF by numerical integration (following the definition of the corresponding marginal) as we have done in Figs. 5, 6, 7. This is because the coupling in the nonlinear power system dynamics makes the states correlated in a non-trivial manner, and this correlation evolves over time, which can only be resolved accurately by directly computing the joint PDFs. Please note that joint PDFs are information-theoretically complete i.e., encode all the statistical information such as the statistical moments and marginals. In fact, unlike the joint PDF generator given by the forward Kolmogorov operator, one cannot derive a PDE for the marginal generator for general nonlinear Itô SDEs. For application-oriented references focusing on the importance and usage of the joint PDF prediction subject to nonlinear state dynamics, we suggest the following references:

S. Haddad, A. Halder, B. Singh, “Density-based stochastic reachability computation for occupancy prediction in automated driving”, *IEEE Transactions on Control Systems Technology*, 2022.

A. Halder, R. Bhattacharya, “Dispersion analysis in hypersonic flight during planetary entry using stochastic Liouville equation”, *Journal of Guidance, Control, and Dynamics*, 34(2), 459-474, 2011.

(ii) Related to the above: given the joint PDF, the marginals are unique. However, the converse is false, i.e., looking at the marginals one cannot reconstruct the complete statistics in the state space. In fact, there are infinitely many joints consistent with the given marginals. This means that every time some different statistics such as higher order moments or other integral functionals of the joint are desired, in the absence of the availability of transient joint PDFs, separate algorithms need to be designed in a case-by-case manner.

(iii) The transient joint PDFs are also useful if one is interested in model (in)validation or controller verification in the probabilistic sense, see e.g.,

A. Halder, R. Bhattacharya, “Probabilistic model validation for uncertain nonlinear systems”, *Automatica*, 50(8), 2038–2050, 2014.

A. Halder, K. Lee, R. Bhattacharya, “Optimal transport approach for probabilistic robustness analysis of F-16 controllers”, *Journal of Guidance, Control, and Dynamics*, 38(10), 1935–1946, 2015.

(iv) If one is interested in nonlinear estimation of the states using noisy sensor data and known stochastic process dynamics, then the predicted transient joint state PDF serves as the prior on which the Bayesian update is to be made to compute the posterior. Directly propagating the prior joint PDFs can alleviate well-known “particle degeneracy” issue suffered by sequential Monte Carlo-based methods such as the particle filter; see e.g.,

P. Dutta, A. Halder, R. Bhattacharya, “Nonlinear estimation with Perron-Frobenius operator and Karhunen-Loeve expansion”, *IEEE Transactions on Aerospace and Electronic Systems*, 51(4), 3210–3225, 2015.

R3.5. Comments by reviewer:

Also, the authors should put a very clear position about some of the references that you are making use to argue with the reviewer. For example, [45] is a 10-year ago paper. Why do you claim it to the state-of-the-art? How can this argument convince the reviewer?

Response: The reason why [45] is the state-of-the-art is because that is the only reference to our knowledge, which attempted to directly and non-parametrically solve the Kolmogorov’s forward PDE subject to the power system dynamics, albeit in even simpler setting (single-machine-infinite-bus, i.e., independent variables being two dimensional state and one dimensional time) than ours. The technical topic of our interest is *not* about what SDE/SDAE model is being used for the sample path dynamics, it is about how to solve the transient joint PDFs by directly solving the generator of the corresponding stochastic process while approximating neither the dynamics nor the statistics. In our previous round’s rebuttal’s response to comment **R3.1**, we clearly explained that there is good reason why no progress were made on this specific technical issue—which is that finite element/volume or Monte Carlo based function approximation schemes cannot be scaled to the type of examples we consider (e.g., 100 dimensional state space in Sec. V-B). This is not a matter of doing research to find intelligent discretization schemes. Instead, it requires to re-think what does it mathematically mean to solve the generator, and fundamentally break away from “solve PDE as a PDE” philosophy. This is what motivated our contribution.

R3.6. Comments by reviewer:

Regarding Ref [23]F. Milano, and R. Zarate-Minano, “A systematic method to model power systems as

stochastic differential-algebraic equations”, IEEE transactions on Power Systems, Vol. 28, No. 4, pp. 4537–4544, 2013. I mentioned this article in the first-round review, not for the purpose of a simple cite, but trying to help you to improve the model and make your work realistic. However, a suggestion is ignored. This is indeed a pity. Besides, please do not argue again that this is a different work because this ref is well-known in modeling realistic stochastic power system dynamics, which should be the foundation of your model.

Response: We have no doubt that ref. [23] is important and well-known for modeling realistic stochastic power system dynamics. However, please notice that:

(i) Our present work is not about modeling stochastic power system dynamics.

(ii) Our citation of [23] in pg. 2, left column, top paragraph, last two sentences, is not mere perfunctory. For the readers’ understanding, we clarified that both refs. [22] and [23] simulate the sample paths and do not directly propagate the joint state PDFs. The latter is what this manuscript is about.

(iii) We have no doubt that more sophisticated models as in [23] should be investigated in future for designing the proximal recursion algorithms such as the current manuscript. However, as we discuss in detail in Sec. III-A and Sec. III-B, deriving the proximal recursions is not a matter of simply swapping one SDE/SDAE model with another. Exploiting the sub-Riemannian-like geometry induced by the base SDE/SDAE model on the ground (finite dimensional) manifold, requires a significant technical tour-de-force (see e.g., refs. [39, 48]) for the specific model under consideration unless the base manifold dynamics has a gradient or mixed conservative-dissipative drift with isotropic diffusion—none of which occurs in stochastic power system dynamics even for the simple model we considered. The proximal updates are essentially trying to construct a “steepest-descent-like” scheme on the manifold of probability measures (which is not even a vector space). Despite being a highly active research topic for the last two decades in the PDE and geometry community, there still is no systematic mathematical way to derive the distance and functional pair as in (27) to rigorously establish that the associated PDE is a steepest descent of which functional measured w.r.t. which distance. This situation is significantly different from the finite dimensional Euclidean or Bregman proximal algorithms as in refs. [32, 34]. The current results in the PDF setting have focused on deriving proximal schemes in case-by-case basis. Therefore, we believe our contribution, though restricted to a simple model, is a significant first stepping stone necessary to consider more realistic models such as in [23], and should also be of interest to mathematicians outside the power systems community.

R3.7. Comments by reviewer:

In general, I need to claim here that math is fine in itself, but only publishable when connected to or interpreted correctly in a convincing engineering application.

Response: Thank you. We have done our best to clarify these connections and interpretations. We hope you approve.

Response to Reviewer #4:

R4.1. Comments by reviewer:

The authors have successfully addressed most of the reviewer's previous concerns. However, the reviewer still has some questions, mostly regarding the numerical simulations.

Response: Thank you. Please find our itemized responses to the questions related to numerical simulations.

R4.2. Comments by reviewer:

It is stated that "The Fig 9 shows that the computational time needed to perform the proximal updates, for all t_k , are indeed orders-of-magnitude faster than h ." This statement doesn't seem to hold, since $h = 10^{-3}$ and the computational time lies within the range of around 10^{-3} to 10^{-2} . How does the computational time needed to perform the proximal updates relate to time step h ? If they are not strongly related, does that mean the step size h is constrained to be greater than that computational time?

Response: The reviewer is right, we should have been more careful in phrasing that sentence. We meant to say that the computational time needed to perform a proximal update is statistically comparable (in the big O sense) to h . The reason why the precise computational time is probabilistic is twofold: random sampling, and randomness in computational latency, power and memory footprint. Fig. 9 shows that $h = 10^{-3}$ and the computational time needed to perform a single proximal update is $O(10^{-3})$ with high probability in the sense most time-steps incur $O(10^{-3})$ time. In Fig. 9, while instances when computational times are $O(10^{-2})$ do occur for Case I, they are rare. In the present revision, we have now rephrased the corresponding text (highlighted in blue) toward the end of Sec. V-A.

The proximal updates involve certain block coordinate ascent recursions in the dual co-ordinates \mathbf{y}, \mathbf{z} ; see lines 13-21 in Algorithm 1. This recursion is guaranteed to be contractive (hence existence, uniqueness and convergence are guaranteed) in the product orthant cone: $(\mathbf{y}, \mathbf{z}) \in \mathbb{R}_{>0}^N \times \mathbb{R}_{>0}^N$, see ref. [41, Sec. III-B, III-C] and especially Corollary 4 in that reference. We have cited that reference in this manuscript's Sec. IV. While h appears as parameter in this recursion (lines 5 and 14 in Algorithm 1), we do not have analytical relation between the computational time and h . We did not constrain or adapt h in Algorithm 1. We simply set a small value $h = 10^{-3}$ to promote statistical consistency (see the unnumbered equation after (27)). Depending on the random sampling, dynamics and computational platform, it is possible for the single-step proximal update computational time to fall below h (see e.g., ref. [41, Fig. 9] which used the same $h = 10^{-3}$). The standard desktop computer/processor we used for our simulation is reported in Sec. V, first paragraph, last sentence.

R4.3. Comments by reviewer:

In Fig. 10, could the authors please explain why the computational time of Case I is less than that of Case II? The computational cost for the 5-generator system is roughly 4 times that of Case II. Could the authors comment if such a scale makes sense intuitively considering the state dimension of the two systems?

Response: Thanks for this comment. The state dimensions for both Cases I and II in Fig. 10 are the same, i.e., $5 \times 2 = 10$ dimensions. Other than the difference in dynamics due to line failure in Case II, we do not think there is any specific physical insight to be gleaned on why the computational times for Case II are below Case I; they both use identical set of random initial samples although the Wiener process noise realizations are different.

Regarding the scaling comment, perhaps the reviewer meant to say that the total computational times for the 50 generator (100 state) system are roughly 4 times that of Case II (5 generator, 10 state). The possible explanation of why the scaling is not so bad is because the number of samples for the 50

generator case is 2000, which is twice the number of samples used in Cases I and II. This is made possible by our usage of scattered point clouds with explicit proximal computation of the time-varying probability weights, as opposed to gridding the state space and then approximating the joint PDF via histogram (which would have caused exponential in dimension scaling in the computational cost). Being able to directly evaluate the joint PDF values via proximal computation allows us to bypass construction of function approximation oracle over the high dimensional state space (which in general scales poorly with dimensions), and then to evaluate it. A direct comparison between the two cases is less meaningful due to difference in parameter values, initial joint PDF etc. However, the general trend that the proposed scattered point cloud-type proximal framework, similar to the one proposed here, can incur computational times which scale conveniently with dimensions while resolving the joint PDF well, were reported in our earlier works in different application contexts: ref. [41, Sec. V-B and Fig. 14], ref. [42, Fig. 5].

R4.4. Comments by reviewer:

Both Fig. 12 and Fig. 13 demonstrate a decreasing trend as k increases. Could the authors comment on this and link this to the property of the proposed proximal algorithm? Similarly, why is the trend in Fig. 14? Comparing Figs. 12, 13 and Fig. 14, did the authors omit the initial time instant in Figs. 12, 13?

Response: We appreciate these helpful comments.

(i) As per the reviewer's suggestion, in the current revised version, we have added two sentences (highlighted in blue) at the end of Sec. V explaining the trends in Fig. 12, 13 and 14.

(ii) In the revised version, we have updated Figs. 12 and 13 to include the initial time instance $k = 0$. To be consistent with the computation of the errors for all k , we did the empirical versus ensemble computation in all the three Figs. 12, 13, 14 at $k = 0$ too, even though theoretically these curves are known to be zero at $k = 0$.