Revision of Manuscript Number: IEEE-TPWRS-01391-2021, version 1   February 19, 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:
The authors are requested to revise and resubmit their paper after addressing reviewer feedback. Of particular importance is that reviewers point out the model leveraged for analysis is overly simplistic.

Response: We are indebted to the AE and the reviewers for their help to improve the manuscript.

In addition to the itemized responses to the reviewers' comments and the corresponding revisions marked in blue in the manuscript, we have made the following additional fixes and improvements (also marked in blue in the revised manuscript):
(i) In equation (8), we corrected a typo in its right hand side: the second summand runs over pairs (i, j) such that $i < j$.
(ii)

Regarding the usage of simple model, please see our response to comment **R3.1** and **1.3, sub-comment (1)**. In summary, the current state-of-the-art in power system literature for this problem is ref. [45] with second order dynamics + single machine infinite bus ($n = 1$ in our manuscript's notation). Given this context, our manuscript enables a new kind of computation that can handle moderately large n with second order sample path dynamics. This computation is online and approximates neither the statistics nor the dynamics. We clarify two important points here:
(i) The proposed proximal recursion framework on the space of probability measures is not conceptually limited by the second order dynamics, and we expect follow up research to design custom proximal recursions for higher order synchronous machine models. In this sense, the present paper should be seen as the stepping stone laying the foundation for handling more complex dynamical models. The authors have been completely up front by explicitly mentioning this in the Conclusions section. We have also reiterated the same in this rebuttal.
(ii) Deriving proximal recursion requires understanding the generalized gradient flow structure over probability measures. Our research disposition is that one cannot bypass understanding the generalized gradient flow structure for the simpler second order sample path dynamics before attempting the same

for more complex sample path models. In fact, a significant contribution of our article, in the authors' opinion, is that it distills out what exactly is technically challenging and what are the peripheral issues.

AE.2. Comments by AE:
Authors are also encouraged to revisit their overview of prior art.

Response: Following the suggestions, the revised manuscript has expanded the overview of prior art. Specifically, we have included two new references [22], [23] on stochastic differential algebraic equations (SDAEs) and cited them inline in Sec. I.1, second paragraph, as detailed in our response to the comment **R3.2** in this document. This includes the reference suggested in comment **R1.3**. We have also included the reference suggested in comment **R4.4**; this is reference [28] in the revised manuscript. Its inline citation is in Sec. I.1, last paragraph, first sentence.

AE.3. Comments by AE:
Presentation of technical points and results can also be improved.

Response: We have made several updates to clarify the technical issues and to improve the presentation.

Following the comment **R4.3** to improve the presentation, in the revised manuscript, we have expanded the last two paragraphs of Sec. III-A with a new reference [50]. Please see our response to that comment below for details.

As detailed in the response to comment **R3.4**, in the revised manuscript, we have enlarged Figures 5 and 6, and expanded the plot in Fig. 6. There we also explained why those plot styles are used.

Furthermore, Fig. 7 in the revised manuscript has also been expanded to include the univariate angular velocity transient marginal plots for all the five generators of the IEEE 14 bus system; previous version had this plot only for the generator 4 (buss 6).

New plots

Response to Reviewer #1:

R1.1. Comment by reviewer:
The original contribution of the article is not clear. Indeed, what is original new from this article, not the summary from existing techniques?

Response: The original contribution of this article is to design an algorithm for *online nonparametric* computation of the joint state PDFs subject to stochastic uncertainties (initial conditions, parameters, process noise and combinations thereof) in power system dynamics.

The qualifier "online" means that it is not a variant of Monte Carlo where samples are first propagated and then a function approximation is performed offline to estimate the joint state PDF at a snapshot. The qualifier "nonparametric" means that neither the statistics nor the dynamics are approximated *a priori*. This is novel because the existing power systems literature makes at least one of the aforesaid approximations. The state-of-the-art in power system literature is discussed in Sec I.1. More importantly, Sec. I.2 details why the state-of-the-art is what it is. We specifically explain the contribution of this article in Sec. I.3 titled "Contributions of this paper".

R1.2. Comments by reviewer:
The structure of the article should be highly improved. A large portion of the contexts is simply repeating the well-known power system classical model, kroon reduction, and the FPF equations. Only from section IV, page 6, I can see some of the proposed weighted points methods.

Response: We thank the reviewer for this comment. This paper is *not* about novel implementation of an existing idea. The idea itself is new and consequently needs to be contextualized and developed in the paper. The discussion toward the end of Sec. II and that in Sec III are, in our opinion, more fundamental contribution than the implementation in Sec. IV. Those parts of the paper really mark the innovation since they explain why someone, even if conversant in infinite dimensional gradient flow literature, cannot directly apply existing techniques to this problem. Then Sec. III-C and III-D explain how to circumvent the specific difficulties arising in this problem. A reader jumping to Sec. IV assuming that everything before that is known, is completely mistaken. Notice that we clearly laid out the structure of this paper (which section does what) in the paragraph right before Sec. II starts.

R1.3. Comments by reviewer:
I have major doubts about the model used in the article for the following reasons.
(1) The paper claim the proposed method can propagate stochastic uncertainties in realistic power networks. However, only a classical model is used, which is not realistic at all.
(2) Why do active power and reactive power from load have no randomness at all. This is too unrealistic.
(3) The diffusion term of this article is only considered in the frequency. The process noise in the rotor angle is fully ignored. This is also unreasonable.
To improve your work, I highly suggested the authors follow [R1], which is missing in your literature but can help the authors better know Power system SDE. Also, it is suggested to introduce the OU process to model the randomness in the load, which cannot be ignored in practice.
R1. A systematic method to model power systems as stochastic differential-algebraic equations, F. Milano, R. Zarate-Minano, IEEE Transactions on Power Systems, 28(4), 4537-4544.

Response: Our responses to the sub-comments below.

(1) **Regarding usage of classical model:** Whether a model is relevant depends on the problem context. As pointed out in the last paragraph of Sec. II-B, the current state-of-the-art in power system literature for the problem we are addressing is ref. [45] with second order dynamics + single machine infinite bus ($n = 1$

in our manuscript's notation). Given this context, our manuscript enables a new kind of computation that can handle moderately large n with second order sample path dynamics. This computation is online and approximates neither the statistics nor the dynamics. We clarify two important points here:

[1a] The proposed proximal recursion framework on the space of probability measures is not conceptually limited by the second order dynamics, and we expect follow up research to design custom proximal recursions for higher order synchronous machine models. In this sense, the present paper should be seen as the stepping stone laying the foundation for handling more complex dynamical models. The authors have been completely up front by explicitly mentioning this in the Conclusions section.

[1b] Deriving proximal recursion requires understanding the generalized gradient flow structure over probability measures. Our research disposition is that one cannot bypass understanding the generalized gradient flow structure for the simpler second order sample path dynamics before attempting the same for more complex sample path models. In fact, a significant contribution of our article, in the authors' opinion, is that it distills out what exactly is technically challenging and what are the peripheral issues.

(2) **Regarding randomness in active and reactive power:** TBD

(3) **Regarding the diffusion term:** The reason why the diffusion term is present in the RHS of (1b) but not in the RHS of (1a) is because (1) is the state space representation of the second order Langevin dynamics (6). For the Langevin dynamics (6), its corresponding Itô SDE has to be (1). This is simply re-writing of a second order differential equation as an equivalent system of first order differential equations. There is no extra assumption here. Indeed, it is mathematically incorrect to add a noise term in the RHS of (1a) while claiming consistency with (6).

Notice that the absence of additive noise in the RHS of (1a) does *not* imply that the rotor angles are deterministic trajectories. The reason why the rotor angles are stochastic processes is because the process noise in (1b) affects the rotor angle sample paths via dynamical nonlinear coupling with the angular velocities, i.e., these are second order stochastic effects. In short, the noise in rotor angles is *not* ignored; in fact the rotor angles are nonlinearly colored (i.e., non-white) noisy sample paths.

This reviewer may also have misunderstood in thinking that a model that has additive noise in the RHS of *all* SDE components is somehow more general than the case where the additive noise terms are absent from some but not all components of an SDE RHS. In reality, the complexity is in the other way around. The SDEs of the form (1) are the so-called "degenerate Itô SDEs" which are much more difficult to analyze than the non-degenerate versions (noise in the RHS of all SDE components). For instance, this degeneracy makes the joint PDF evolution equation (10) a *kinetic* Fokker-Planck PDE, which is different in structure than the standard Fokker-Planck a.k.a. Kolmogorov's forward PDE. The gradient flow structure and the associated measure-valued proximal recursions for the kinetic Fokker-Planck PDE are less understood than the same for the standard Fokker-Planck PDE. This is exactly why the methodological contribution of this paper is important, even if one completely ignores the algorithmic contribution.

Regarding the reference [R1]: In the revised manuscript, we have included the ref. [R1] suggested by the reviewer with inline citation in Sec. I.1, toward the end of second paragraph. Please note that the topic of [R1] is different from the topic of our paper. We are interested in how to perform online nonparametric computation of transient joint state PDFs. Our computation is different from Monte Carlo.

Regarding OU process in load: The reviewer might have missed the following key point of our paper. The Kron reduced second order stochastic model implies that in the unreduced sense, the load dynamics was first order noisy Kuramoto, in addition to the synchronous machine dynamics being second order noisy Kuramoto. Therefore, we do have stochastic load dynamics, which is in fact more realistic than OU process model. While OU process assumes linear drift with constant diffusion, whereas our setting considers the load SDEs having drifts with trigonometric nonlinearity and constant diffusion.

R1.4. Comments by reviewer:
There is no comparison of the proposed method compared with other state-of-the-art methods. Therefore, it is hard for the readers to judge the importance of this article.

Response: TBD.

R1.5. Comments by reviewer:
Contributions are over-claimed in this article:
(1) The scalability of the proposed method cannot convince the readers since you are using a too simple model and ignored many important random dimensions from the loads.
(2) The proposed method claims to be nonparametric, however, all the inputs are assumed to be either wiener process or uniform distribution, which are all parametric distributions.
(3) If the "nonparametric" is only referred to the output PDF, I would not think of it as a major contribution since there are a lot of existing techniques that can depict nonparametric PDF.

Response: Our responses to the sub-comments below.

(1) **Regarding simple model and ignoring the randomness in loads**: Please refer to our response to **R1.3, sub-comment (1)** where we explained why this model is relevant for this particular problem, and why this is not limiting to our methodological contribution. In response to **R1.3, last unnumbered sub-comment in bold text: "Regarding OU process in load"**, we explained that we are not ignoring randomness from loads.

(2) **Regarding nonparametric**: We are *not* making any parametric assumptions. Our usage of the word "nonparametric" is standard as in mathematical statistics literature. That the additive noise process is Wiener does not invite any loss of generality since if it was colored, we could always re-cast the same in our SDE form via a linear map. Furthermore, we have zero assumptions on the initial joint state PDF ρ_0 and on the joint PDF of the parameters appearing in the model. As a matter of fact, we don't even need to know the functional form of these joint PDFs. For instance, our framework remains unchanged if we are given the initial joint state and parameter distributional data as numerical table containing samples from these joints *as well as* the joint PDF values evaluated at those samples. The latter might be coming from some measurement-based sensor fusion/Bayesian algorithm. The specific product von Mises and uniform PDFs in Sec. V-A and V-B are only used for generating example data to perform our illustrative numerical simulations. This again underscores how jumping straight to Sec. V without contemplative reading and understanding of the earlier Sections may lead to misleading conclusions.

(3) **Regarding depict nonparametric PDF**: This paper is not about nonparametric *depiction* but about nonparametric *computation* of the transient joint state PDFs. For example, Monte Carlo followed by function approximation via Kernel density estimates and its variants, or other function approximants (e.g., RBF, Gaussian mixture, neural networks) are all parametric methods. We enable computing the value of joint state PDF at a time-varying sample location *online (not as offline post-processing)* without assuming any functional representation of the transient joint which the computed values correspond to. Notice that we are doing this joint computation without gridding the state space, which would be impossible had we followed the traditional scientific computing route: finite element/volume methods, tensor decomposition techniques etc. We believe the proposed computational framework is novel, highly non-trivial and is specifically useful in this power system uncertainty propagation context.

R1.6. Comments by reviewer:
The simulation time is too short to convince the readers. Only 1 second is considered here. Note that the errors can be quickly accumulated in the iteration.

Response: TBD.

R1.7. Comments by reviewer:
It is suggested to count the total computing time, not on each step.

Response: While the depiction of the per-step computational times in Figs. 8 and 9 in the revised manuscript are somewhat unconventional, we explain why we choose to plot that way. To aid clarity, in the revised manuscript, we have added an excerpt of the following at the end of Sec. V-A.

At each fixed physical time $t_k = kh$, $k \in \mathbb{N}$, where h is the fixed step-size, we perform the proximal update $\varrho_{k-1} \mapsto \varrho_k$ via Algorithm 1. This proximal update has no analytical solution and instead requires execution of a contractive fixed point recursion. The physical time t_k is "frozen" (zero-order-hold) during the fixed point recursions underlying this one-step proximal update. To claim that the proposed method is practical, we therefore, need to demonstrate that the computational timescale is faster than the physical timescale, i.e., computational times needed to complete the zero-order-hold fixed point recursions are orders-of-magnitude faster than h, the step-size of physical time. On the other hand, theoretical consistency (see the unnumbered equation after (26), before Sec. IV) requires that h itself should be small. Thus, even if one has a provably convergent algorithm that can perform the one step proximal update, it is not clear *a priori* if the computational time needed to do so can be much faster than h. If not, then the method would be statistically consistent but practically not useful. The Figs. 8 and 9 in the revised manuscript show that the computational time needed to perform the proximal updates, for all t_k, are indeed orders-of-magnitude faster than h.

R1.8. Comments by reviewer:
If the proposed method depends on the Kronecker product, how can it be scaled to a higher dimension?

Response: The proposed method does not entail computation of a Kronecker product. In the manuscript, the Kronecker product with the 2×2 identity matrix appeared in the definition of the change-of-variable (17). Its only purpose is to define that change-of-variable, and thereby derive the PDE in the new variables. The proximal recursion (for the PDE thus derived) detailed in Algorithm 1 does not have any Kronecker product computation.

R1.9. Comments by reviewer:
beta is undefined in equation (15a).

Response: Thanks for the comment. The β in (15a) is the same β in (16), as in part of the same sentence. To avoid any confusion, in the revised manuscript, we have moved the phrase "for some $\beta > 0$" in (15a).

Response to Reviewer #2:

R2.1. Comments by reviewer:
Overall, the paper is well-written and theoretically sound.

Response: Thank you for the kind and encouraging remarks.

R2.2. Comments by reviewer:
The paper's language and presentation make it seem like it is more oriented towards the applied mathematics and control communities. It is highly suggested that the authors revise the exposition of ideas and concepts, putting more emphasis on the power system so that it can be well-understood by the power system experts.

Response: We appreciate the comment. In the revised manuscript, we have made some changes toward this direction. This includes slightly expanded overview of the related power system literature (see the two sentences added in Sec. I.1, second paragraph with two new references [22], [23]), citing recent reference [28] in Sec. I.1, third paragraph. Please note that the literature review in Sec. I.1 is completely based on the power system literature.

Having said that, we believe our technical contribution is indeed interdisciplinary in the sense it is not mere translational research. In other words, we are not simply importing ideas from a different domain and applying them to a power system problem. Instead, we make novel contributions at the (possibly unanticipated) intersection of two areas: measure-valued gradient flows and power system dynamics. The presentation of the paper reflects the same, perhaps rightfully so. Notice in particular that our presentation is bottom-up: we start from power system literature review and mention that the problem addressed in our paper was already recognized as a technical challenge in that literature. We move onto the theoretical ideas and tools much later in the manuscript, as the build up demands.

We expect a significant share of the readers of this work to be power system engineers, but not exclusively so since this contribution will also be of interest to the researchers in control, optimization and applied mathematics. We do not believe this issue is exceptional for our paper since several high impact papers in the IEEE Trans. Power Systems were indeed at the intersection of these areas, and are not only well-read by the researchers from either side but these works have positively contributed toward increasing interdisciplinary research collaboration.

R2.3. Comments by reviewer:
How will the proposed approach be used for the cases in which dependencies among uncertainty sources exist? An explanation and illustration would be helpful.

Response: In our paper, we do not make any assumption on statistical independence of the uncertainties. In fact, all stochastic uncertainties are considered to be "joint", i.e., statistically correlated. Please note that the initial joint PDFs (28) and (30) are taken to be in that form for the sake of concrete example simulations. Their only purpose, in the simulations reported in Sec V, is to generate joint samples, and to evaluate the initial joint PDF values at the joint samples thus generated. If this initial scattered point cloud data is supplied to us only as a numerical table, without specifying any functional form whatsoever, nothing changes in the proposed algorithm. Similarly, the parameters in Table I could instead be correlated random vectors, and all we need are the joint parametric weighted point cloud data. In fact, this is how we expect the proposed computational framework to be applied in practice.

R2.4. Comments by reviewer:
Incorporation of power system dynamics into the proposed uncertainty propagation framework is a bit vague. The proposed study seems not to have utilized any tool capable of simulating power system

dynamics. Could the authors explain the part pertaining to the dynamic simulation in more detail?

Response: In this paper, we are interested in propagating the joint PDF values of the states of a power system dynamic model subject to joint stochastic uncertainties. Our focus is on explicitly computing these transient joint state PDFs at the time-varying state vector samples, which amounts to co-evolving the states and the joint PDF values over time. This is different from Monte Carlo or the traditional density estimation techniques since they approximate the joint PDFs via post-processing, not as an online computation. To the best of our knowledge, there does not exist a power system simulation tool which can do this online computation while making neither statistical nor dynamical approximation.

Please note that in Sec. V, we have used MATPOWER and the Python-library ANDES (with inline citation of refs. [62] and [63] therein) for obtaining some simulation parameters, as explained in the first paragraph of V-A.

R2.5. Comments by reviewer:
With regard to the point above, this reviewer would expect to see uncertainty bounds for the system dynamic response such as the trajectory of the rotor angle. Without such information, the method's usability is difficult to gauge.

Response: TBD.

R2.6. Comments by reviewer:
Instead of using a purely synthetic system, the authors are highly encouraged to choose available test cases such as the IEEE 118-bus system and a larger one, if possible.

Response: Please note that the synthetic system in Sec. V-B is actually simulating a higher dimensional system than the IEEE 118-bus system where the 19 generators would have resulted in the transient joint PDFs with 38 dimensional supports. In comparison, the synthetic system is evolving joint state PDFs having 100 dimensional supports, i.e., in $\mathbb{T}^{50} \times \mathbb{R}^{50}$. Our only intention behind simulating the synthetic system was to show that the computation can handle large dimensional joint state PDF supports. We also point out that the parameter ranges reported in Table I are consistent with the existing literature: refs. [25, Sec. 5], [20], [56], [57]. We hope the reviewer agrees that this simulation is adequate for the intended purpose.

Response to Reviewer #3:

R3.1. Comments by reviewer:
My main concern with this work is that all its derivations are deeply model-dependent and, unfortunately, the model that the author consider is not adequate to represent modern power systems. Moreover, second order model of synchronous machines are adequate only in a time scale of few seconds, so, even if the system were effectively including only conventional power plants, the model would be acceptable only for very short-term transients, i.e., transients for which the propagation of the noise is immaterial. In other words, the proposed model is useless.

Response: The reviewer's point is well taken in the sense the second order model may not be adequate to capture all the complexities of modern power systems, and further research is needed to design/extend nonparametric joint PDF propagation methods such as the one we propose with higher order generator dynamics. Please note that the last sentence of our Conclusion section explicitly stated so.

At the same time, it is important to understand how exactly this paper is making a significant stride w.r.t. the current state-of-the-art even with the second order dynamics. The last paragraph of our Sec. II-B clearly asserts that the state-of-the-art in directly computing the joint PDF remains the ref. [45] published in this journal in 2013, which considered the second order dynamics for the single machine infinite bus case via finite element discretization. In the same paragraph, we also cited two 2018 references [16, 17] published again in this journal, which did not directly propagate the state PDFs and instead computed some proxy for the same. So the key question becomes: if the power system literature has already acknowledged the importance of solving this problem, why is there no progress since [45] for even moderate n and even assuming seemingly simple second order models? What exactly is the fundamental impediment standing in the way of nonparametric computation of the joint state PDF subject to the power system dynamics? Our Sec. I.2 clearly lists the challenges for the same. The crux is that addressing this challenge is not a matter of innovating on a better discretization scheme or innovating on computing architecture. See specifically the last two sentences of our Sec. II-B. The present paper is breaking the stalemate by mathematically innovating on what it really means to solve the underlying PDE, albeit subject to the second order dynamics. As we have stated up front, we expect follow up research to extend the proposed approach for higher order models but we cannot have progress with a complex model unless we first demonstrate how to tackle the simpler second order case – a case that the existing literature cannot tackle for nonparametric online PDF computation.

Last but not the least, we find the reviewer's assertion "In other words, the proposed model is useless" moot for the following reasons.

(i) We do not propose the second order model. We use the second order SDE that albeit simple, is widely used in power system dynamic simulation, as the sample path dynamics. In this paper, we are not interested in simulating sample paths via variants of Monte Carlo. Our interest is in nonparametrically computing the transient joint PDFs subject to the given sample path dynamics in the form of Itô SDE. Our usage of this particular sample path dynamics does not preempt the possibility of designing a variant of the proposed proximal recursion for performing computation of the same nature when the given Itô SDE is of higher order. Indeed, this is what we alluded to in the last sentence of the Conclusions section.

(ii) What we propose is not a model. Transcribing the kinetic Fokker-Planck operator into a generalized gradient flow, and thereby deriving the corresponding proximal recursion, is a purely mathematical construct. There are no modeling assumptions made in that derivation. Our contribution in this paper is to initiate and showcase this new way of computing the transient solutions of the kinetic Fokker-Planck PDE arising from a simple but popular power system dynamic model. Extending this framework to higher order sample path dynamics is mentioned as future work, and is beyond the scope of this particular paper.

R3.2. Comments by reviewer:
The literature is incomplete, many recent works on SDAEs for power system applications are not mentioned.

Response: As per the reviewer's suggestion, in the revised manuscript, we have included two sentences (marked blue) in Sec. I.1, toward the end of second paragraph, mentioning the usage of SDAEs in power system dynamic simulation. Therein, the first sentence contains two newly added references: [22] and [23], both on simulating SDAE sample paths.

R3.3. Comments by reviewer:
There are several typos in the test (especially in the references).

Response: TBD. We corrected the two typos pointed out in comment **R4.8** in this rebuttal document.

R3.4. Comments by reviewer:
Figures 5 and 6 are unreadable.

Response: Thanks for this comment. In the previous version, the strict 10 page limit enforced small figure sizes. In the revised manuscript, Figures 5 and 6 are made larger. Also, the Fig. 6 in the revised manuscript shows the transient (θ_i, ω_i) bivariate marginals for all the five generators in Case II; the previous version showed the same only for the generator 4 (bus 6). We believe these changes have improved clarity.

Regarding the style of plots, we had considered plotting the bivariate marginals as planar contour plots (as in Fig. 1 in this rebuttal document) but decided against it because of the following two reasons.

First, in our opinion, this kind of plot is scientifically misleading as they hide the fact that these bivariate marginal PDFs are not really supported on the plane, or the subsets thereof, but instead have cylindrical supports. Our message that the proposed method explicitly accounts this manifold geometry in the computation, would be missed had we chosen to make the plots as per populist cognitive bias on behalf of the readers. That the planar plots evoke a sense of familiarity cannot be a scientific argument in favor of it.

Second, the planar contour marginal plots require additional mental efforts from the readers to correctly parse the statistical information conveyed. For instance, consider the fourth column, second row subplot in Fig. 1 in this rebuttal document. It is upon the reader to parse that the two high probability regions on the left and right of that subplot are actually referring to the same angular high probability region around $\theta_2 = 0$ rad, thanks to the periodic boundary condition on the angular variable. Our philosophy of technical communication is that the onus is on the authors, not on the readers, to clarify the scientific information as correctly as possible. And yes, we are aware that many papers in the power system literature make plots in the rectangular domain ignoring the manifold nature of the support, which in our experience, continue to add pain to the graduate students and beginning researchers trying to decipher the statistical meaning conveyed in those plots. We thus prefer to make plots that depict the variables as they are.
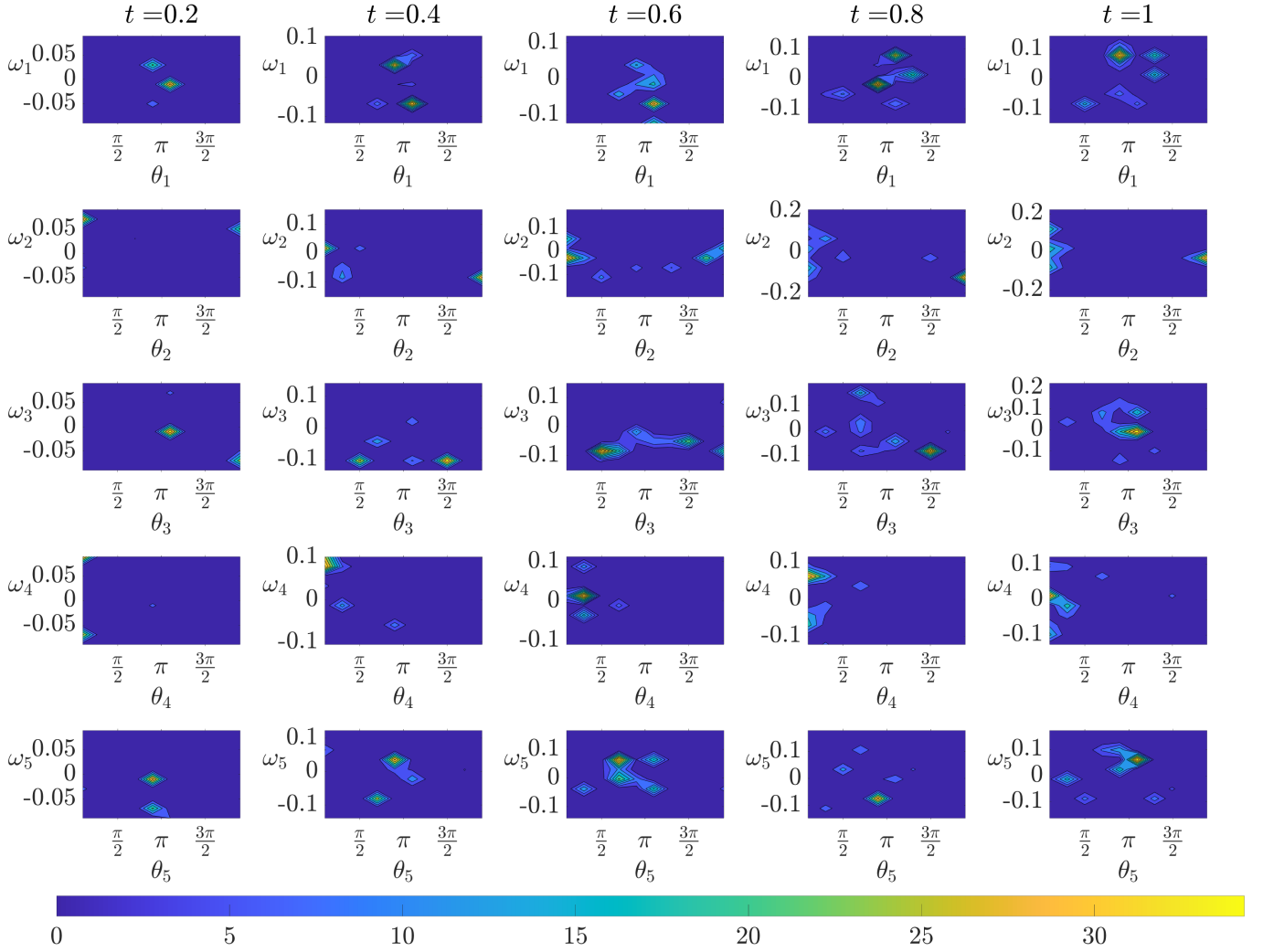
Fig. 1: Example (θ_i, ω_i) bivariate marginals, for the IEEE 14 bus system's generators $i = 1, \ldots, 5$, shown as planar contour plots. The columns correspond to time snapshots; the rows correspond to the five generators. The colorbar corresponds to the value of the bivariate marginals. The simulation set up is described in Sec. V-A of the manuscript.

Response to Reviewer #4:

R4.1. Comments by reviewer:
The main contribution of this paper lies in applying results on the governing partial differential equation (PDE) for the joint PDF to avoid discretization of the state space, thereby circumventing computational challenges. The reviewer appreciates that this paper is rigorous in development and thinks it deals with an important problem.

Response: We appreciate the positive remark.

R4.2. Comments by reviewer:
The key contribution of this paper is that the proposed method can reduce the computation burden when solving for the transient PDF. However, no comparison is provided with existing methods. The authors should provide comparisons to the exiting approaches, for example, based on finite element discretization or other numerical methods. The test example could be performed using the 5-generator system or larger systems. Besides, referring to Figs. 8 and 9, the total computation time should be reported for a longer horizon.

Response: TBD.

R4.3. Comments by reviewer:
Regarding the presentation, the reviewer suggests making the paper more self-contained. Some discussion in this paper highly relies on the authors' previous work, which can be improved by either adding a summarization or trimming out the materials.

Following the same vein, there are some unnecessary information and technical points provided in this paper that may be distracting. For example, on page 3, why would the authors mention the Kuramoto oscillator model? On page 4, is the detailed discussion regarding Wasserstein-2 metric necessary? Similarly, on page 5, the discussion about the natural energy functional and how the gradient drift case does not apply also seems too detailed.

Response: We have followed the reviewer's suggestion to improve the presentation. Specifically, in the revised manuscript's Sec. III-A, last two paragraphs summarize the technical relation of this paper w.r.t. the authors' previous work. In particular, we have added three more sentences with a new reference. To improve readability, have also broken what was a single paragraph in the previous version, into two separate paragraphs.

We appreciate the reviewer's point that the sentence containing equation (6) and the mention of Kuramoto model can be removed without affecting the rest of the development. However, we prefer to keep it since it adds didactic value, and many engineers may feel comfortable by the explicit mention that (1) is simply a rewriting of (6). Furthermore, an astute reader, encountering that sentence will realize that the scope of the technical contribution of this paper is broad because the algorithmic framework developed here will be applicable to several other non-power system applications (mentioned in the two cited references in that sentence) where the second order Kuramoto model is used. Since this is just a matter of one sentence, we will prefer to keep it. We hope you approve.

On the other hand, the detailed discussion regarding Wasserstein gradient flow and why that standard development cannot be applied in this case, is in fact, the heart of this paper. Please note that this is far from a recipe paper where a numerical algorithm is proposed backed by some theory and simulations. The key contribution of this paper, instead, is to help distill out the core technical difficulty that hinders simply importing standard techniques from the optimal transport literature. We believe this clarification of what really is the technical crux of the matter–versus what are peripheral issues–should be illuminating to not just power systems researchers, but also to applied mathematicians underscoring how the standard theory and algorithms for measure-valued gradient flows need to be adapted in engineering applications.

R4.4. Comments by reviewer:
Along with [25], there are recent developments that over-approximates the reachable set for nonlinear systems under uncertain initial conditions, parameters, and external disturbances: "Reach-Set Estimation for DAE Systems under Uncertainty and Disturbances Using Trajectory Sensitivity and Logarithmic Norm".

Response: Thanks for providing that interesting recent reference. We have cited the same in the revised manuscript.

R4.5. Comments by reviewer:
On page 7, the authors stated about "To account the pertinent geometry of $\mathbb{T}^n \times \mathbb{R}^n$". What exactly does this mean? Do the authors indicate that in (28) the two groups of state values are independent? Or do the authors mean the specific choices of the product von Mises PDF and the uniform distribution? If so, what are the connections?

Response: We simply mean that the joint PDF should be supported on $\mathbb{T}^n \times \mathbb{R}^n$, or a submanifold thereof. For instance, one cannot assume the initial joint PDF to be a $2n$ dimensional multivariate normal since that is supported over $\mathbb{R}^{2n}$, not over $\mathbb{T}^n \times \mathbb{R}^n$. It is, however, perfectly fine if some or all of the $2n$ components of the initial state vector are statistically correlated. As long as the support is $\mathbb{T}^n \times \mathbb{R}^n$, there is no issue if the angular variables are correlated with the Euclidean (i.e., angular velocity) variables.

R4.6. Comments by reviewer:
On page 5, is there a mistake in "random samples from (28), and evaluated at (28)"?

Response: That was not a typo. Indeed, equation (28) was used in two ways: it was used to generate initial state samples, and then the initial joint PDF values at those samples were evaluated via the functional form of (28), i.e., as in function evaluation. This is needed in our case because unlike standard Monte Carlo, we co-evolve the state samples and joint PDF values at those state samples. To clarify this further, in the revised manuscript, we have slightly expanded that sentence.

R4.7. Comments by reviewer:
As a validation, it would be good to also provide information of higher-order moments along with the mean given in Fig. 10.

Response: TBD.

R4.8. Comments by reviewer:
There are some typos in this paper:
Page 4: in therms of
Page 5: in the right-hand side of the arrow in the expression following "converges to the flow generated by (22)", it should be plain ρ

Response: Corrected both. Thanks for pointing them out.