Geometrically Motivated Reparameterization for Identifiability Analysis in Power Systems Models

Mark K. Transtrum
Dept. of Physics and Astronomy
Brigham Young University
Provo, Utah, USA
mktranstrum@byu.edu

Benjamin L. Francis
Dept. of Physics and Astronomy
Brigham Young University
Provo, Utah, USA

Clifford C. Youn
Dept. of Electrical and Computer Engineering
Tufts University
Medford, MA, USA

Andrija T. Saric
Dept. of Power, Electronics, and Com. Engineering
University of Novi Sad
Faculty of Technical Sciences, Serbia
asaric@uns.ac.rs

Aleksander M. Stankovic
Dept. of Electrical and Computer Engineering
Tufts University
Medford, MA, USA
astankov@ece.tufts.edu

***Abstract*—This paper describes a geometric approach to parameter identifiability analysis in models of power systems dynamics. When a model of a power system is to be compared with measurements taken at discrete times, it can be interpreted as a mapping from parameter space into a data or prediction space. Generically, model mappings can be interpreted as manifolds with dimensionality equal to the number of structurally identifiable parameters. Empirically it is observed that model mappings often correspond to bounded manifolds. We propose a new definition of practical identifiability based the topological definition of a manifold with boundary. In many ways, our proposed definition extends the properties of structural identifiability. We construct numerical approximations to geodesics on the model manifold and use the results, combined with insights derived from the mathematical form of the equations, to identify combinations of practically identifiable and unidentifiable parameters. We give several examples of application to dynamic power systems models.**

***Index Terms*—power system modeling, system identification, parameter identifiability, computational differential geometry, model reduction, manifold boundary approximation method**

I. Introduction

Dynamical models used in power system analysis and control are facing a number of challenges. These stem from the need to model novel components (e.g., power electronic inverter-connected sources and loads), from the operation governed by potentially volatile markets, and from the need to predict system behavior in atypical scenarios (e.g., in resilience studies). While the size of the models involved (thousands of sources and tens of thousands of nodes) has been addressed mostly via the advances in the computer technology, the issue of model fidelity has lagged behind. Historically, the lack of sensors and of mechanisms to share actual event recordings has hampered efforts to validate models. There exists a strong preference for physics-based models in the power system community, supported both by the tradition and by the important insights gained from such models in the past (e.g., in the case of power system stabilizers). Such models tend to be nonlinear both in terms of parameters and in terms of states, and their validation poses methodological and practical challenges.

On several occasions the power system community has been on the forefront of new developments in engineered dynamical systems, as in the case of applications of trajectory sensitivity [1]–[3] or in the case of subset selection [4]. The industry feels the need for a systematic effort in this direction, as described in [5].

In this paper, we consider the question of whether parameters can be identified from available measurements. When parameters can be inferred from measurements, at least in principle, they are classified as structurally identifiable [6], [7]. Even when all the parameters can be identified in principle, obtaining precise estimates may not be practical. In contrast to structural identifiability, there is no universally accepted definition of practical identifiability [8], [9].

A recently introduced term that describes a class of complex models exhibiting large parameter uncertainty when fit to data is sloppiness [10]–[12]. Sloppy models are closely related to the existence of practically unidentifiable parameters. The premise of this approach is that a model with many parameters is a mapping from a parameter space into a data (prediction) space. A key difficulty in dealing with models of complex systems is the highly anisotropic nature of the mapping between these two spaces. This anisotropy is manifested locally in the wide spread of eigenvalues of the measurement Hessian, and globally as the hierarchy of widths of the corresponding bounded manifold in data space. Thus, the issue is not just a simple over-parametrization in terms of the number of parameters, but is due to the very nature of the models being verified.

This work was supported in part by the NSF under grant number EPCN-1710727

978-1-5386-7138-2/18/$31.00 © 2018 IEEE

In this paper, we review the concept of local, structural identifiability and note that it is equivalent to the topological characterization (i.e., dimensionality) of neighborhoods on the model manifold. A particularly useful locally-calculated object in our study is the Fisher Information Matrix (FIM), or the Hessian of the sensitivities of measurements to model parameters. The rank of the FIM corresponds to the number of locally, structurally identifiable parameters in the model. We give an example of a model with structurally unidentifiable parameters, and observe that the structurally identifiable parameter combinations provide a bridge between the mechanistically-interpretable, bare parameters, and collections of phenomenologically-interpretable parameter combinations.

We next consider the question of practical identifiability. Since sloppy models have manifolds with a hierarchy of widths, their topology is that of a manifold with boundary. We propose a definition of practical identifiability in terms of manifold boundaries and give several examples from dynamic power systems models. We then discuss how geodesics can be used as a tool to identify non-trivial reparameterizations of the model that partition the parameter space into practically identifiable and unidentifiable parameter combinations. Similar to the structurally identifiable combinations, this partition groups mechanistic parameters into their phenomenologically relevant combinations.

II. Structural Identifiability

Dynamic models of power systems are typically written in DAE form:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{z}, \mathbf{p}, t) \tag{1}$$

$$\mathbf{0} = \mathbf{g}(\mathbf{x}, \mathbf{z}, \mathbf{p}, t) \tag{2}$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{z}, \mathbf{p}, t) \tag{3}$$

where $\mathbf{x}$ is the vector of differential state variables, $\mathbf{z}$ are the algebraic variables, $\mathbf{p}$ are the parameters, $\mathbf{y}$ is the system measurement vector, and t is the scalar time variable.

The parameters $\mathbf{p}$ are to be estimated from measurements $\mathbf{y}$ that we assume are made at a discrete set of time points, which we denote as a vector $\mathbf{t} = \{t_1, t_2, \dots\}$. In this case, the total number of measurements is $M = \dim(\mathbf{y}) \times \dim(\mathbf{t})$, and by evaluating the model defined in by Eqs. (1)-(3) at these discrete times, the model makes M independent predictions. The values of these predictions depend on the values of the parameter vector, which suggests interpreting the model, evaluated at the discrete times as a mapping: $\mathcal{M} : \mathbb{R}^N \to \mathbb{R}^M$, where $N = \dim(\mathbf{p})$. We assume throughout that $N < M$. We refer to $\mathcal{M}$ as the *model mapping*.

An important first question is whether it is possible *in principle* to infer all the parameter values from measurements. This question, known as the structural identifiability, is equivalent to asking whether the model mapping is injective. A partial answer can be found, by constructing the Jacobian matrix $J = \partial \mathcal{M} / \partial \mathbf{p}$. In general J will be an $M \times N$ dimensional matrix and the parameters are structurally identifiable (at least in the local sense, as we assume for now) if the rank of $J = N$.

If rank $J < N$, the model has $N' =$ rank J structurally identifiable parameter combinations. In this case, the goal is often to reparameterize the model so that it explicitly depends on only the N' identifiable parameters. In some cases, the reparameterization can be inferred by direct inspection of the mathematical form in Eqs. (1)-(3).

To illustrate, consider a single cage induction machine. Of particular interest are parameters r_S (stator resistance), x_S (stator reactance), r_{R1} (rotor resistance), x_{R1} (rotor reactance) and x_m (magnetizing reactance). The model equations take the form

$$\dot{e}'_d = \Omega_b \sigma e'_q - \frac{1}{T'_0}\left[e'_d - (x_0 - x')i_q\right] \tag{4}$$

$$\dot{e}'_q = \Omega_b \sigma e'_d - \frac{1}{T'_0}\left[e'_q - (x_0 - x')i_d\right] \tag{5}$$

$$v_d = e'_d + r_s i_d - x' i_q \tag{6}$$

$$v_q = e'_q + r_S i_q + x' i_d \tag{7}$$

where

$$x_0 = x_S + x_m \tag{8}$$

$$x' = x_S + \frac{x_{R1} x_m}{x_{R1} + x_m} \tag{9}$$

$$T'_0 = \frac{x_{R1} + x_m}{\Omega_b r_{R1}}. \tag{10}$$

Although the model involves five, mechanistically interpretable parameters, they naturally group into four structurally identifiable parameters: x_S, x_0, x' and T'_0. While the structurally identifiable parameter combinations do not have a direct mechanistic interpretations, they do have phenomenological meaning: x_0 is the open circuit resistance, x' is the transient reactance, and T'_0 is the open circuit time constant.

We observe here that by an appropriate reparameterization in terms of identifiable parameter combinations, the mechanistic structure of the model is explicitly connected to the model's phenomenology. We will demonstrate in later sections how reparameterizations in terms of practically identifiable combinations enables a similar bridging between mechanism and phenomenology.

Before considering the practical identifiability problem, we make a brief observation about the mathematical nature of the structural identifiability problem. Assuming that $\mathcal{M}$ is a smooth function of the parameters, the model mapping defines a Riemannian manifold, often called the *model manifold* [13]. The dimensionality of the model manifold is equal to the number of structurally identifiable parameters. Again restricting ourselves to local properties, the manifold dimension is the only topological property of the manifold, i.e., all manifolds of the same dimension are locally isomorphic. Thus, the (local) structurally identifiability problem is equivalent to identifying the local topology of the model manifold. Motivated by this observation, we now give a topological interpretation of practical identifiable parameter combinations.

III. Model Manifolds and Hyper-ribbons

A systematic study of the geometric properties of model manifolds from a large number of diverse fields, including power systems [12], [14], [15], has revealed a remarkable empirical result: model manifolds are often bounded with a hierarchy of widths. Often, it is possible to vary a combination of parameters over their entire physically allowed range (e.g., zero to infinity) and the behaviors of the model will only change by a finite amount. Models with this geometric property are called *sloppy* and are often associated with practically unidentifiable parameters. Thus, when considering non-local topological properties of a model manifolds, the relevant structure is often that of a *manifold with boundary*.

We remind readers that a manifold with boundary is different from a manifold. The latter is a space locally isomorphic to $\mathbb{R}^N$, where N is the manifold dimension. In the context of modeling, the inverse model mapping provides such an isomorphism, so that the dimension is given by the number of structurally identifiable parameters. In contrast, a manifold with boundary is a space locally isomorphic to $\mathbb{R}^{N-1} \times \mathbb{R}_+$, where $\mathbb{R}_+$ is the set of non-negative real numbers. The boundary of the manifold corresponds to the set in which the last coordinate is zero. In other words, for a model manifold with boundary, there exists a reparameterization in which one parameter is non-negative and zero on the boundary.

We propose a topological definition of practically unidentifiable parameters in terms of manifold boundaries: At a given level of statistical confidence, if the confidence region on the model manifold extends to the boundary, then we say that the parameter combination associated with that boundary is practically unidentifiable at that confidence level. The remaining parameters are the practically identifiable combinations. This definition is a natural extension of structural identifiability in several ways. First, it extends the topological interpretation of parameter identifiability. Second, structural unidentifiability occurs in the limit that the width of the model manifold becomes zero. Third, it is reparameterization invariant, but naturally partitions a model's parameters into identifiable and unidentifiable combinations. The identifiable combinations are those that have phenomenological interpretations.

Using this definition, it is straightforward to see that many parameters in dynamic power systems models are practically unidentifiable at some confidence level. We illustrate with the single-cage induction machine in Eqs. (4)-(7). Each of the four structurally identifiable parameters are non-negative, thus, for example, the limit $x_S \to 0$ corresponds to a portion of the boundary of the model manifold. The parameter x_S is only unidentifiable on a portion of the manifold boundary. Setting other parameters to zero corresponds to other regions of the boundary.

Simply considering the case that all the model parameters become zero does not lead to a complete description of the manifold's boundary. Indeed, other regions can be more subtle to find. For example, one can also consider the limit $T_0' \to \infty$. This observation suggests a reparameterization in terms of $\lambda = 1/T_0'$. This limit is a singular limit of the model in which the dynamic variables e_q' and e_d' become algebraic and corresponds to yet another portion of boundary. Furthermore, inspecting Eqs. (8) and (9), we see that physical parameters lead to the restriction $x' < x_0$. Consequently, another portion of the boundary corresponds to the case $x' = x_0$, suggesting the reparameterization $\delta x = x_0 - x'$.

While these examples are motivated by simple considerations, they suggest the possibility of non-trivial combinations of practically identifiable parameters and highlight the need of an algorithm for discovering such combinations. In the next section, we use computational differential geometry on the model manifold to deduce such reparameterizations.

IV. Geodesic Identify practically unidentifiable combinations

We have seen that, with our proposed definition of practical identifiability, it is possible to find potentially identifiable/unidentifiable parameter combinations by inspection. In all of these examples, the identifiable combinations were deduced by considerations of the physically allowed range of the parameters. This raises the question of whether or not these combinations collectively represent the entire boundary of the manifold. For a particular model, do there exist less obvious combinations?

In order to find other regions of the manifold boundary, we use methods of computational differential geometry. Our goal is to computationally explore the geometric structure of the model manifold to find the least identifiable combinations in the model. To accomplish this we numerically construct curves on the model manifold known as geodesics. Geodesics are the analogs of straight lines generalized to curved surfaces. A geodesic curve can be written as $\mathbf{p}(\tau)$, where τ parameterizes the geodesic. Here we give a brief tutorial of how this is done. The process is described in more detail and several examples are given in [16]–[18].

Geodesics are found as the numeric solution to a second order ordinary differential equation in parameter space:

$$\frac{d^2 p}{d\tau^2} = \sum_{j,k,m} \left(I^{-1}\right)^{il} \frac{\partial y_m}{\partial p^l} \frac{\partial^2 y_m}{\partial p^j \partial p^k} \frac{dp^j}{d\tau} \frac{dp^k}{d\tau}. \qquad (11)$$

In Eq. (11) we have dropped the bold-face vector notation for explicit index notation, simplifying the translation to a computer program. Note that the superscript on the parameter vector, e.g., p^l, is an index, not a power (this is a standard notation in differential geometry). The matrix $I = J^T J$ is the Fisher Information Matrix, a symmetric, non-negative matrix summarizing the local structure of the model mapping. In Eq. (11), the parameter τ corresponds to the geodesic arclength on the model manifold.

Eq. (11) makes use of derivatives of the model mapping with respect to the parameters. Care must be taken when evaluating these derivatives. In particular, the matrix I is ill-conditioned for sloppy models, so it is important that derivatives be evaluated with sufficient accuracy to avoid numerical artifacts. We find two approaches that work well. First, derivatives can be estimated using finite differences, although higher-order estimates are often necessary to achieve sufficient accuracy. Alternatively, one can solve the *sensitivity equations*, found by implicitly differentiating Eqs. (1)-(3) with respect to the parameters:

$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{p}} + \frac{\partial \mathbf{f}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{p}} + \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \tag{12}$$

$$\mathbf{0} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{p}} + \frac{\partial \mathbf{g}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{p}} + \frac{\partial \mathbf{g}}{\partial \mathbf{p}} \tag{13}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{p}} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{p}} + \frac{\partial \mathbf{h}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{p}} + \frac{\partial \mathbf{h}}{\partial \mathbf{p}}. \tag{14}$$

These equations are linear in terms of sensitivities, but the matrices involved do vary along a system trajectory. The geodesic equation involves second order sensitivities, which can be derived in a similar way. We omit an explicit formula as the derivation is straightforward and the result is lengthy and not illuminating. Since the explicit expressions for sensitivity equations can be rather complicated, we recommend they be evaluated using automatic differentiation methods [19], [20] rather than deriving explicit analytic expressions for Eqs. (12)-(14).

The second order sensitivities enter Eq. (11) in an interesting way that allows efficient calculations. In particular, it only depends on the combination

$$A_m\left(d\mathbf{p}/d\tau\right) = \sum_{j,k} \frac{\partial^2 y_m}{\partial p^j \partial p^k}\frac{dp^j}{d\tau}\frac{dp^k}{d\tau}. \tag{15}$$

This is a directional second derivative (our notation highlights the explicit dependence on the direction $d\mathbf{p}/d\tau$) and can be calculated much more efficiently than Eq. (15) might suggest. In particular, in terms of finite differences

$$A_m\left(\mathbf{v}\right) \approx \frac{y_m(\mathbf{p}+h\mathbf{v}) - 2y_m(\mathbf{p}) + y_m(\mathbf{p}-h\mathbf{v})}{h^2}. \tag{16}$$

That is, it can be evaluated with a computational cost that is independent of the size of the parameter space. In contrast, the Jacobian matrix costs approximately N function evaluations and the full second derivative matrix costs approximately N^2 function evaluations.

With these considerations, one can find the right hand side of Eq. (11) given Eqs. (1)-(3). The initial value problem corresponding to the geodesic can then be numerically approximated using standard integration algorithms.

Our algorithm for finding the least identifiable parameter combination is as follows. First, we parameterize the model in a way that all the parameters can vary over the entire real line, what we call *unbound coordinates*. If the natural parameterization includes constraints such as positivity (as in T_0' in Eqs. (4) and (5) for example), then we reparameterize the model so that parameters have no such constraints. In this case, we recommend a log-transform: $p = \log T_0'$. Other parameterizations can be similarly constructed. For example, the constraint that $x' < x_0$ can be enforced by introducing a parameter f that satisfies $x' = x_0/(1+e^f)$.

Next, given a model parameterization in unbound coordinates, we take as initial conditions to the geodesic equation the best estimate of the parameter values, $\mathbf{p}_0$ (e.g., the best fit when fit to data) and the least sensitive parameter combination as estimated by the Jacobian matrix. That is, $d\mathbf{p}_0/d\tau$ is the eigenvector of the Fisher Information Matrix I with smallest eigenvalue. With these initial conditions, we then solve Eq. (11) numerically.

Because we have parameterized our model to have unbound parameters, when the geodesic encounters a manifold boundary, some combination of parameters becomes infinite. Mathematically, the solution to Eq. (11) exhibits a singularity at some finite value of τ that we denote by τ^*. We illustrate this in Figure 1, which is the solution of the geodesic equation for a model of a synchronous generator (see [21]). In this case, the geodesic encounters a singularity just before $\tau = 2$, corresponding to the limit that $\log T_{q0}' \to -\infty$. Because we do not know a priori the value of τ^*, we use a simple heuristic to decide when to terminate the geodesic integration. We monitor the value of $\sqrt{\sum_l \left(dp^l/d\tau\right)^2}$, i.e., the norm of the velocity vector in parameter space, and when it has grown by a factor of ten, say, then we terminate the geodesic integration.

From the solution of the geodesic equation, it is possible to identify non-trivial reparameterizations that correspond to practically unidentifiable parameter combinations. The reparameterizations are manifest when multiple parameters take on infinite values in the same geodesic curve. Because the boundary of the manifold is another manifold of one less dimension, there will always be combinations of these infinite parameters that remain finite. This is best demonstrated through examples; we give several in the next section.

V. NONTRIVIAL REPARAMETERIZATIONS

When several parameters simultaneously approach positive or negative infinity when solving the geodesic equation (Eq. (11)), it indicates that the practically unidentifiable parameter combination is a nontrivial combination of bare parameters. We do not give an algorithm for extracting the identifiable and unidentifiable combinations from the geodesic. Doing this successfully is an art that combines insight from the numerical geodesic calculation with the mathematical structure of the model; to illustrate, we here give several examples.

In reference [22], we consider a Doubly-Fed Induction Generator (DFIG) model, appropriate for a wind generator. When solving the geodesic equation in two cases, a singularity in the geodesic was encountered involving three bare parameters. Here we discuss one of these two singularities; the second is nearly identical. In this case, three bare parameters (k_{i1}, k_{p1}, and T_3) simultaneously approached infinity. Inspecting the relevant model equations, we notice that these three parameters all occur in an equation for current:

$$\frac{di_{rq}}{dt} = \frac{1}{T_e}\left[k_{p1}\left(\tau_m^* - \frac{P_g}{\omega_m}\right) + k_{i1}x_1 - i_{rq}\right]. \tag{17}$$

We refer the reader to reference [22] for more information about the physical interpretation of the symbols in this equa-

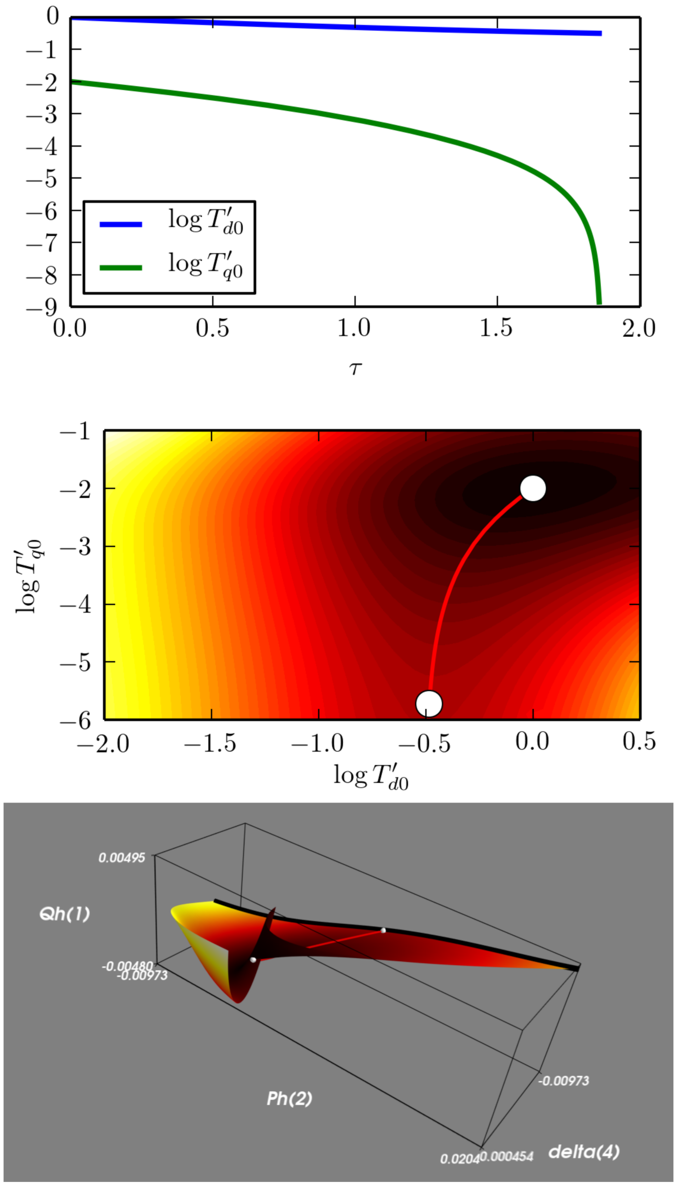

Fig. 1. Solution to the geodesic equation for a model of a synchronous generator [21].

tion. Only the mathematical structure of the equations is relevant for the present discussion. Notice that with a few algebraic manipulations, this equation takes the form:

$$\frac{d}{dt} i_{rq} = \left(\frac{k_{p1}}{T_e}\right)\left(\tau_m^* - \frac{P_g}{\omega_m}\right) + \left(\frac{k_{i1}}{T_e}\right) x_1 - \frac{i_{rq}}{T_e}. \quad (18)$$

The mathematical structure of the model naturally groups the parameters into the combinations $\tilde{k}_{p1} = k_{p1}/T_3$ and $\tilde{k}_{i1} = k_{i1}/T_e$. This suggests that $\tilde{k}_{p1}$ and $\tilde{k}_{i1}$ are identifiable parameter combinations while $\lambda = 1/T_e$ is the unidentifiable parameter that becomes zero at the boundary. This guess is confirmed by inspecting the numerical solution to the geodesic equation. From the approximate geodesic, we calculate $\tilde{k}_{i1}(\tau) = k_{i1}(\tau)/T_e(\tau)$ and observe that while $k_{i1}(\tau)$ and $T_e(\tau)$ each diverge at the boundary, the ratio remains finite. This test is a nontrivial confirmation that our guess is correct. Although all three parameters are singular as the curve approaches τ^*, it is not necessary that all three singularities are $\mathcal{O}(1/\lambda)$. With this confirmation, we reparameterize the model in terms of practically identifiable and unidentifiable parameter combinations.

As a second example, we consider parameters from a single cage induction machine in the context of the WECC load model [23]. A discussion of the complete identifiability analysis of this model is presented in a recent thesis [24]. Solving the geodesic indicates that the parameters x_0, x' and T_0' all approach infinity simultaneously. Rearranging Eq. (4) gives

$$\dot{e}_d' = \Omega_b \sigma e_q' - \left[\frac{e_d'}{T_0'} - \left(\left(\frac{x_0}{T_0'}\right) - \left(\frac{x'}{T_0'}\right)\right) i_q\right], \quad (19)$$

suggesting the combinations $\tilde{x}_0 = x_0/T_0'$ and $\tilde{x}' = x'/T_0$ and the unidentifiable combination $\lambda = 1/T_0'$. Indeed, considering the solution to the geodesic equation we find that $\tilde{x}_0$ and $\tilde{x}'$ remain finite at the boundary and correspond to the identifiable combinations while T_0' is practically unidentifiable.

VI. Conclusions

In this paper we have considered the problem of parameter identifiability from time series measurements in power systems using tools of information geometry. We have reviewed the concept of structural identifiability and noted that it is mathematically equivalent to a topological characterization of the neighborhood of a point on the interior of the model manifold. The dimension of a manifold is the only topological property of neighborhoods of points on a manifold. Using this insight, we have proposed a definition of practical identifiability that is based on nonlocal topological properties of the model manifold. In our definition, a parameter combination is practically unidentifiable if a confidence region intersects the boundary of the model manifold. The two examples of nontrivial reparameterizations considered in section V are very similar in terms of the required mathematical manipulations and resulting combinations. The identifiable combinations correspond to natural groupings of parameters as they appear in the mathematical structure of the model. This is remarkably similar to structurally identifiable combinations introduced in section II. Indeed, this is an indication that our proposed definition of practical identifiability is a natural extension of the structural identifiability.

Identifying the combinations of parameters that are potentially unidentifiable has important consequences for constructing predictive models and interpreting their behavior. By reparameterizing the model in terms of the identifiable and unidentifiable combinations, reduced order models can be constructed by explicitly taking the limit that the unidentifiable combinations become zero. This approach, known as the manifold boundary approximation was first introduced in reference [17] and applied to power systems models in reference [21]. As we have seen here, these limits may correspond to singular limits of the model; other types of approximations for synchronous generators are discussed in reference [21]. The methods scale to larger systems as well, such as the WECC load model in [24]. Analysis of the IEEE 14 bus model is forthcoming. A priori knowledge of the potentially unidentifiable parameter combinations could also be useful for optimal experimental design. As in the case of structural identifiability, the practically identifiable and unidentifiable combinations correspond to groups of parameters that are directly linked to the behaviors of the model. Practically unidentifiable combinations correspond to groups of parameters that can be taken to their extreme values without changing the behavior of the model beyond a given statistical tolerance. In contrast, the practically identifiable combinations are those that must be tuned in order for the model to match a desired behavior. Thus, by construction, the practically identifiable parameters are the combinations of mechanistic parameters that combine to determine a systems-level behavior.

One of the challenges of the current approach is the need to numerically solve the geodesic equation, (11) due to the computational cost of the calculating parameter sensitivities to the necessary accuracy. However, since the identifiable combinations exhibit regular patterns linked to the structure of the model, it may be possible to list the potentially identifiable combinations by inspecting the mathematical structure of the model, perhaps with the aid of a few geodesic calculations. This observation motivates a potential way forward for generalizing to larger models. By combining insights from solving the geodesic equation on small or moderately sized models with mathematical acumen, a catalog of potentially unidentifiable combinations could be constructed for a given model class. This catalog could then be directly applied to larger models within the same model class. An open question that remains is whether such a catalog includes all the potentially unidentifiable parameter combinations. Although beyond the scope of this paper, one approach to this question involves further topological calculations. Each MBAM-reduced model corresponds to a portion of the boundary of the model manifold. The question of completeness is therefore equivalent to asking whether the union of this list of reduced models has any "holes" which is revealed by calculating the Euler characteristic.

Accurate parameter estimates are important for training predictive models. In this paper, we have proposed a new approach to parameter identifiability based on geometric and topological considerations. Our approach groups parameters into practically identifiable and unidentifiable combinations and links to new methods of model reduction while making connection to existing techniques such as singular perturbation.

REFERENCES

[1] J. Sanchez-Gasca, C. Bridenbaugh, C. Bowler, and J. Edmonds, "Trajectory sensitivity based identification of synchronous generator and excitation system parameters," *IEEE Transactions on Power Systems*, vol. 3, no. 4, pp. 1814–1822, 1988.

[2] S. M. Benchluch and J. H. Chow, "A trajectory sensitivity method for the identification of nonlinear excitation system models," *IEEE Transactions on Energy Conversion*, vol. 8, no. 2, pp. 159–164, 1993.

[3] I. A. Hiskens, "Nonlinear dynamic model evaluation from disturbance measurements," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 702–710, 2001.

[4] M. Burth, G. C. Verghese, and M. Vélez-Reyes, "Subset selection for improved parameter estimation in on-line identification of a synchronous generator," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 218–225, 1999.

[5] P. Overholt, D. Kosterev, J. Eto, S. Yang, and B. Lesieutre, "Improving reliability through better models: Using synchrophasor data to validate power plant models," *IEEE Power and Energy Magazine*, vol. 12, no. 3, pp. 44–51, 2014.

[6] R. Bellman and K. J. Åström, "On structural identifiability," *Mathematical biosciences*, vol. 7, no. 3-4, pp. 329–339, 1970.

[7] P. Ju and E. Handschin, "Identifiability of load models [power systems]," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 144, no. 1, pp. 45–49, 1997.

[8] R. Brun, P. Reichert, and H. R. Künsch, "Practical identifiability analysis of large environmental simulation models," *Water Resources Research*, vol. 37, no. 4, pp. 1015–1030, 2001.

[9] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, 2009.

[10] K. S. Brown and J. P. Sethna, "Statistical mechanical approaches to models with many poorly known parameters," *Physical Review E*, vol. 68, no. 2, p. 021904, 2003.

[11] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, "Perspective: Sloppiness and emergent theories in physics, biology, and beyond," *The Journal of chemical physics*, vol. 143, no. 1, p. 07B201_1, 2015.

[12] M. K. Transtrum, A. T. Sarić, and A. M. Stanković, "Information geometry approach to verification of dynamic models in power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 440–450, 2018.

[13] M. K. Transtrum, B. B. Machta, and J. P. Sethna, "Why are nonlinear fits to data so challenging?" *Physical review letters*, vol. 104, no. 6, p. 060201, 2010.

[14] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, "Parameter space compression underlies emergent theories and predictive models," *Science*, vol. 342, no. 6158, pp. 604–607, 2013.

[15] M. K. Transtrum, "Manifold boundaries give" gray-box" approximations of complex models," *arXiv preprint arXiv:1605.08705*, 2016.

[16] M. K. Transtrum, B. B. Machta, and J. P. Sethna, "Geometry of nonlinear least squares with applications to sloppy models and optimization," *Physical Review E*, vol. 83, no. 3, p. 036701, 2011.

[17] M. K. Transtrum and P. Qiu, "Model reduction by manifold boundaries," *Physical review letters*, vol. 113, no. 9, p. 098701, 2014.

[18] ——, "Bridging mechanistic and phenomenological models of complex biological systems," *PLoS computational biology*, vol. 12, no. 5, p. e1004915, 2016.

[19] L. B. Rall, "Automatic differentiation: Techniques and applications," 1981.

[20] H. M. Bücker, G. Corliss, P. Hovland, U. Naumann, and B. Norris, *Automatic differentiation: applications, theory, and implementations*. Springer Science & Business Media, 2006, vol. 50.

[21] M. K. Transtrum, A. T. Sarić, and A. M. Stanković, "Measurement-directed reduction of dynamic models in power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2243–2253, 2017.

[22] A. Saric, M. Transtrum, and A. Stankovic, "Information geometry for model identification and parameter estimation in renewable energy-dfig plant case," *IET Generation, Transmission & Distribution*, 2017.

[23] D. Kosterev, A. Meklin, J. Undrill, B. Lesieutre, W. Price, D. Chassin, R. Bravo, and S. Yang, "Load modeling in power system studies: Wecc progress update," in *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*. IEEE, 2008, pp. 1–8.

[24] C. C. Youn, "Information geometry for model reduction in power systems," Ph.D. dissertation, Tufts University, Medford MA, 5 2018.