Information Geometry Approach to Verification of Dynamic Models in Power Systems

Mark K. Transtrum, Andrija T. Sarić, and Aleksandar M. Stanković, *Fellow, IEEE*

Abstract—**This paper describes a new class of system identification procedures that are tailored to electric power systems, in particular to synchronous generators (SGs) and other dynamic components. Our procedure builds on computational advances in differential geometry, and offers a new, global characterization of challenges frequently encountered in system identification of electric power systems. The approach also benefits from increasing availability of high-quality measurements. While the proposed procedure is illustrated on SG example in a multimachine benchmark (IEEE 14-bus and real-world 441-bus power systems), it is equally applicable to identification of other system components, such as loads.**

Index Terms—**Computational differential geometry, information geometry, power system stability, system identification.**

I. Introduction

DYNAMIC models of power systems (for example, electromechanical models used in transient analysis) have grown in size to thousands of generators, and tens of thousands of nodes. However, this growth in quantitative terms has largely not been accompanied with improvements in fidelity. Specifically, models have been largely unable to replicate major events like the 2003 blackout in the Eastern interconnection [1] and several such events in the 1990's in the Western interconnection [2]. This is even more of concern, given the relatively widespread presence of sensors that have made detailed recordings during transients. Additional challenges are posed by market-driven operation and by renewable energy sources that lead to new flow patterns.

Attempts to use on-line data to improve dynamical models of key components, especially synchronous generators have a long history in power systems. First efforts to employ general dynamical concepts like trajectory sensitivity go back more than a quarter century [3], [4]. That particular approach has been successfully extended to hybrid systems in [5]. Another influential approach that is based on local information extracted from the measurement Jacobian is described in [6]. Recent experiences with dynamic model validation in industry are described in [7].

In this paper, we aim to demonstrate that the root cause of these issues stems from the very nature of models being verified from available measurements. A recently introduced term [8], [9] that describes a class of complex models exhibiting large parameter uncertainty when fit to data is sloppiness. 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 mapping between the parameter and data 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. The issue is not just a simple over-parametrization in terms of the number of parameters – the observed system behavior actually constraints combinations of the original parameters.

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. Sloppy models are characterized by FIM eigenvalues that are log-linear−roughly evenly spaced over several decades; it is not uncommon for aspect ratios to be much greater than 1000 to 1 [10], [11]. The eigenvalues of the FIM tell us which parameter combinations are well-constrained by the data (stiff directions in parameter space, corresponding to large eigenvalues) and which are not (sloppy directions). Later we illustrate that sloppiness is not a property that is to be eliminated by better modeling–rather, it is intrinsic to many physics-based models of networked systems, and needs to be effectively managed. In particular, useful predictions are possible without precise parameter knowledge. As long as the model predictions depend on the same stiff parameter combinations as the data, the predictions of the model will be constrained in spite of large numbers of poorly determined parameters. In the case of power systems, the origins of sloppiness include component models (see, e.g. [4]), the networked model structure (such as weak couplings of electrically distant components), and the measurement structure (sampling density and repeated measurements [12]).

We illustrate our ideas later using the example of identification of parameters of a synchronous generator (SG) from terminal measurements, assuming local data rates commensurate with PMUs. The basic problem has been addressed many times in the literature, and we only list references of immediate

Manuscript received August 15, 2016; revised January 8, 2017; accepted March 18, 2017. Date of publication April 12, 2017; date of current version December 20, 2017. Paper no. TPWRS-01238-2016. *(Corresponding Author: Aleksandar M. Stanković.)*

M. K. Transtrum is with the Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602 USA (e-mail: mktranstrum@byu.edu).

A. T. Sarić is with the Department for Power, Electronic and Communication Engineering, Faculty of Technical Sciences, University of Novi Sad, Novi Sad 21000, Serbia (e-mail: asaric@uns.ac.rs).

A. M. Stanković is with the Department of Electrical Engineering and Computer Science, Tufts University, Medford, MA 02155 USA (e-mail: astankov@ece.tufts.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2017.2692523

relevance to our development. For example, [5], [6], [13] consider parameter estimation for a single generator and re-casts SG parameter identification in a differential-algebraic equation (DAE) framework. Refs. [14], [15] consider the same overall setup involving PMU-derived measurements. They estimate parameters one-by-one (later up to 3), use the Extended Kalman filter to validate the parameter values, and use the "playback" method to validate the parameter values. In this paper we consider the basic problem (in a multi-machine formulation) and the extension with key control loops closed [automatic voltage regulator (AVR) and power system stabilizer (PSS); note that in these analyses the turbine-governor (T-G) dynamics is neglected as much slower].

Ref. [16] introduced key terms from information geometry and applied them to (relatively) simple power system components [the doubly-fed induction generator (DFIG) is a prototypical wind energy source, and the direct-drive synchronous generator (DDSG) is routinely used for large solar (and some wind) plants], while focusing on comparisons with local characterizations. Ref. [17] considers in detail the reduction process that parallels singular perturbation-derived modeling of SGs when the state count goes from 6 (two dampers) to 4 (*d*-*q*), 3 or 2 ("classical" model). It also considers damping quantification in larger systems. In this paper, we focus on system aspects and implications of information geometry. We discuss the notion of parameter sloppiness, which permeates many models in power systems. We illustrate the consequences on a large model of industrial relevance.

The outline of the paper is as follows: Section II describes the power system model used for identification; Section III describes the parameter estimation from local perspective, while Section IV describes the information geometry, semi-global and global sensitivity based approach to model identification; Section V shows obtained results in a multi-machine benchmark example and Section VI contains our recommendations and conclusions. Appendix provides details for dynamic models described by differential and algebraic equations.

II. Power System Model for Identification

Static and dynamic models of power systems are typically written in DAEs form:

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

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

where $\boldsymbol{x}$ is the vector of (differential) state variables, $\boldsymbol{z}$ are the algebraic variables, $\boldsymbol{p}$ are parameters and t is the (scalar) time variable.

Details of differential ($\boldsymbol{f}$) and algebraic equations ($\boldsymbol{g}$) used in test power systems are given in Appendix.

System measurement vector is assumed to be of the form:

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

The parameters ($\boldsymbol{p}$) are to be estimated from measurements ($\boldsymbol{y}$), and there typically exists some prior information about parameters, often in the form of plausible ranges for each. The least squares optimization formulation of the identification problem is by far the most prevalent in the literature. This formulation is equivalent to assuming the measurement noise is additive and Gaussian, in which case the sum of squares is (apart from an overall additive constant) the negative log likelihood.

The key quantities in the case of least squares identification are parametric sensitivities whose dynamics can be written as [13]:

$$\frac{d}{dt}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}} = \frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{x}} \cdot \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}} + \frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{z}} \cdot \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{p}} + \frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{p}} \tag{4}$$

$$\boldsymbol{0} = \frac{\partial \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{x}} \cdot \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}} + \frac{\partial \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{z}} \cdot \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{p}} + \frac{\partial \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{p}} \tag{5}$$

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}} = \frac{\partial \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{x}} \cdot \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}} + \frac{\partial \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{z}} \cdot \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{p}} + \frac{\partial \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p}, t)}{\partial \boldsymbol{p}}. \tag{6}$$

These equations are linear in terms of sensitivities, but the matrices involved do vary along a system trajectory. Our method below also makes use of the second order sensitivities, which can be derived in a similar way. We use automatic differentiation [18], [19] to simplify the process.

III. Parameter Estimation – Local Perspective

Equation (6) determines $m \times p$-dimensional (m and p are total numbers of available measurements and uncertain parameters, respectively) time series of Jacobian matrices ($\boldsymbol{J_p}(t) = \partial \boldsymbol{h}(t)/\partial \boldsymbol{p}$), first partial derivatives of system measurement vector (3) with respect to the parameter vector ($\boldsymbol{p}$). For small increments, the Hessian matrix is approximately the FIM, $\boldsymbol{H_p}(t) = \boldsymbol{J_p}^T(t)\boldsymbol{J_p}(t)$, which is symmetric and positive semidefinite, so all its eigenvalues are real and non-negative. Quite often $\boldsymbol{H_p}(t)$ is nearly singular, and the nearness to singularity is measured by the condition number $\kappa(\boldsymbol{H_p})$, which is the ratio of the largest $\lambda_{\max}$ to the smallest eigenvalue $\lambda_{\min}$. Above analysis could be equivalently framed in terms of singular values for associated singular vector of $\boldsymbol{J_p}(t)$ [6]. In our simulations, the condition numbers are calculated for column-stacked Hessians corresponding to different time points following the beginning of the transient of interest (clearing of the short circuit in our simulations−see Table II); the eigenvalues do not vary significantly in our examples.

Based on previously published work in power systems [3]–[6], [13], [20], it is to be expected that all parameters cannot be estimated at once. In our paper, the ill-conditioned parameters are detected by the participation factors [21] in eigenvalues of the (approximate) Hessian.

IV. Information Geometry, Semi-Global and Global Sensitivities

Recent advances focusing on data space ($\boldsymbol{y}$) rather than parameter space ($\boldsymbol{p}$) have proven beneficial for understanding the global properties of models and for advancing numerical techniques for exploring them [22]. This approach, usually known as information geometry, combining information theory with differential geometry, is an effective mathematical language for exploring parameterized models [23]. The essence of the approach is the interpretation of a model as a manifold embedded in the space of data ($\boldsymbol{y}$), known as the model manifold. Information geometry offers a useful parameterization−independent perspective on modeling that has led to advances in numerical algorithms, such as improved data fitting routines [22] (which we use in Section V for constructing parameter likelihood profiles) and Markov-Chain Monte Carlo (MCMC) sampling that account for the anisotropic cost surfaces in parameter space [24].

Semi-global methods of sensitivity analysis address the shortcomings of local approaches by sampling parameter space in a finite neighborhood around the best fit. While in this paper we consider a deterministic formulation for parameter identification, we will classify the semi-global methods using a probabilistic framework. Consider a stochastic model $\mathcal{P}(\boldsymbol{y}|\boldsymbol{p})$ for observing a vector of random variables ($\boldsymbol{y}$) given a parameter vector ($\boldsymbol{p}$). It is, according to Bayes' rule, proportional to the product of the prior probability the parameters $\mathcal{P}(\boldsymbol{p})$ and the likelihood function $\mathcal{P}(\boldsymbol{p}|\boldsymbol{y})$. Broadly, semi-global methods fall into two categories: 1) scanning methods and 2) Bayesian methods. Scanning methods sample parameters (sometimes without regard to the data) and look for correlations between locations in parameter space and the model behavior, or value of the cost function at those locations [25]. Bayesian methods sample from the posterior distribution of the parameters given the data and use those samples to make inferences about the sensitivity of the model [24], [26]. We make use of both types of methods in Section V.

Information geometry allows one to go further and characterize the global sensitivities of the dynamic model. Key aspects of the information geometry approach are:

1) There is no information loss in the model manifold, i.e., the manifold contains all information about the model behavior. In contrast, the cost surface in parameter space condenses the high-dimensional quantities making up the prediction and data vectors into a single number, i.e., the cost.
2) Information geometry separates the model, i.e., the manifold embedded in data space, from the data to which it is being fit, i.e., a point in the data space. In contrast, the cost surface in parameter space is a function of and often very sensitive to the data point being fit.
3) The set of points that constitute the model manifold are the same regardless of how the model is parameterized. This is in contrast with local characterizations using FIM that are parametrization-dependent. The parameters are not ignored completely, but act as coordinates on the manifold.
4) Interestingly, when a probabilistic structure is assumed for the manifold in data space, then the Riemannian metric on the model manifold (with distance between models given by the change in measurement residuals) is the FIM [26], [27].

Geodesics are the analogs of straight lines generalized to curved surfaces. We use computational differential geometry to find numerical approximations to geodesic curves on the model manifold to systematically explore the manifold boundaries. The process is described in detail and several examples are given in [22], [27]. Briefly, geodesics are calculated numerically as the solution to a second order ordinary differential equation in parameter space (while utilizing quantities from the data space):

$$\frac{\partial^2 p^i}{\partial \tau^2} = \sum_{j,k} \Gamma^i_{jk} \frac{\partial p^j}{\partial \tau} \cdot \frac{\partial p^k}{\partial \tau}; \Gamma^i_{jk} = \sum_{\ell,m} (\boldsymbol{I}^{-1})^{i\ell} \frac{\partial y_m}{\partial p^\ell} \cdot \frac{\partial^2 y_m}{\partial p^j \partial p^k}, \tag{7}$$

where Γ are the so-called Christoffel symbols [28], which are expressed in terms of the parametric sensitivities in (4)–(6) and $\boldsymbol{I}$ is the FIM. The parameter τ is the arc length of the geodesic curve as measured on the model manifold, i.e., in data space. Notice how the model provides the connection between the parameter space and data space through the Jacobian matrix $\boldsymbol{J}_{\boldsymbol{p}}(t) = \partial \boldsymbol{h}(t)/\partial \boldsymbol{p}$, as calculated in (4)–(6). Further note that the Christoffel symbols involve the second order sensitivities that are found by taking another derivative in (4)–(6). We omit an explicit formula as the derivation is straightforward and the result is lengthy and not illuminating. Furthermore, because we evaluate these sensitivities using automatic differentiation [18], [19], these expressions are not explicitly needed. There is a technical subtlety in the evaluation of (7) that is critical for our approach to be tractable for large models. Because the second derivative of the observation vector is contracted twice with the geodesic velocity vector (i.e., the sums over indices j and k in (7) form two "dot products" with the geodesic velocities and the array of second derivatives), only a directional second derivative is needed, which can be calculated efficiently as in [22], [27].

Solutions to the geodesic (7) are calculated using standard methods for numerically integrating Initial Value Problem. Since geodesics are central to our proposed global analysis, we demonstrate the calculation and interpretation with a simple power systems example. Consider a model of a SG with two unknown parameters (the transient time constants in the d- and q-axes) that makes predictions for generator rotor angle, generator speed, real and reactive powers at several times after some disturbance (we postpone specific details).

The geodesic is found by first selecting initial parameter values and an initial direction in parameter space ($\partial \boldsymbol{p}/\partial \tau$). In this example, we take these to be the "true" parameter values and the eigenvector of the FIM with smallest eigenvalue. (We use quotes to denote that these "true" parameter values are not necessarily the true values used to generate the data; they are the starting point of a geodesic). Our global analysis requires starting from a variety of initial parameter values and directions. Next, we numerically solve the model DAEs (1), (2), the sensitivities (4)–(6), and the second order sensitivities in the direction of $\partial \boldsymbol{p}/\partial \tau$. These quantities are used to construct the Jacobian matrix ($\boldsymbol{J}_{\boldsymbol{p}}(t) = \partial \boldsymbol{h}(t)/\partial \boldsymbol{p}$), the FIM matrix ($\boldsymbol{I} = \boldsymbol{J}_p^T \boldsymbol{J}_p$), and the geodesic acceleration (7). Next, we numerically solve (7), evaluating the model equations and first and second order sensitivities at each step of the integration.

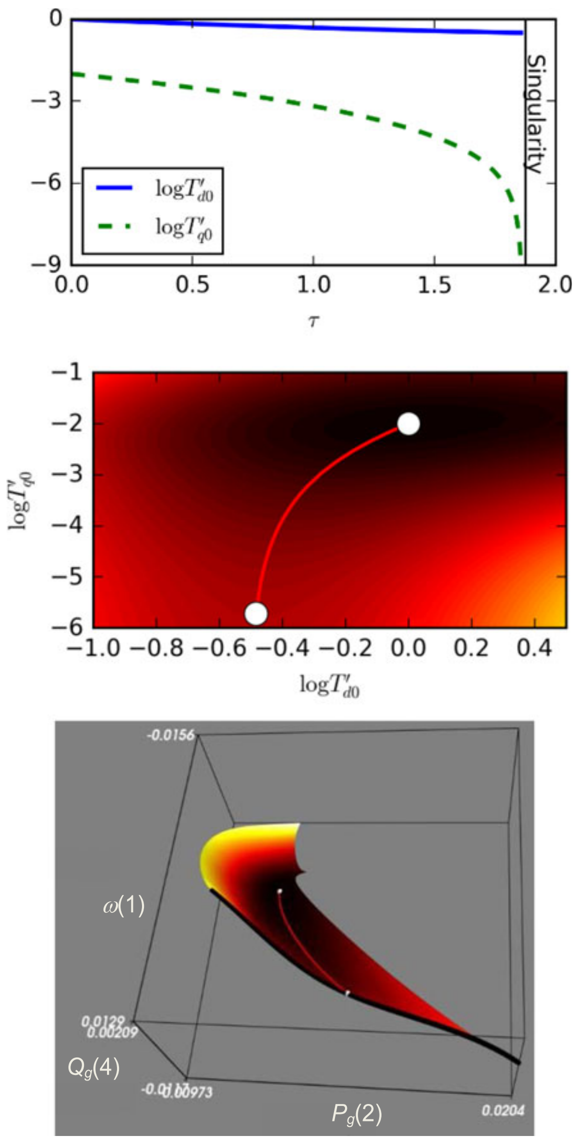

Fig. 1. Solving the geodesic equation (7) generates a sequence of parameter values as a function of τ (top panel) that parameterizes a curve through parameter space (middle panel). This curve corresponds to a path on a model manifold (bottom panel) that is approximately a straight line. This curve encounters a singularity near $\tau = 1.8$ that corresponds to the boundary of the model manifold (black line in bottom panel).

A numerical solution to the geodesic generates a sequence of parameter values as a function of τ (Fig. 1, top panel) that parameterize a curve through parameter space (Fig. 1, middle panel). Colors in Fig. 1, middle panel represent contours of least squares cost measuring deviation of the model behavior from the initial parameter values. Notice how the geodesic naturally curves through parameter space to construct the path of smallest least squares cost. This path can also be interpreted as a path on a manifold (Fig. 1, bottom panel). In this example, the model manifold is a two-dimensional surface (corresponding to the two parameters of the model) embedded in a 1680 dimensional space (1680 = four measurements at 420 time points each). In order to visualize this manifold, we have chosen three axes corresponding to three measurements [generator speed at $t = 1$, $\omega(1)$, real power at $t = 2$, $P_g(2)$, and reactive power at $t = 4$, $Q_g(4)$]. These measurements were selected to make features of the manifold visually clear. Colors on the model manifold match the corresponding parameter values in the middle panel.

Notice that the geodesic is approximately a straight line through the data space (bottom panel), curving only to match the curvature of the model manifold. The length of this curve is proportional to τ. The model manifold has a boundary (black curve, bottom panel). This boundary is manifest as a singularity in the geodesic at a finite τ (near 1.8) that corresponds to the geodesic encountering this boundary. Inspection reveals that this boundary corresponds to the limit $T'_{q0} = 0$. The existence of this boundary is the geometric indication that T'_{q0} is practically unidentifiable from below; it can be taken to its extreme limit without incurring an infinite cost.

By constructing several orthogonal geodesic paths, we identify different cross sections of the model manifold and use the geodesic distance (τ) to measure the width of the manifold. It is empirically observed that the entire manifold is bounded and often highly anisotropic with widths typically forming an exponential hierarchy, reminiscent of the hierarchy of eigenvalues revealed by the local analysis [22]. The boundaries are a feature of the model manifold that are invariant to certain changes in the observation vector and give a topological (i.e., global) description of the model manifold [29]. Furthermore, by calculating many geodesics with a variety of initial conditions, we can identify all of the boundaries and by extension all of the parameter combinations that are susceptible to being unidentifiable.

V. Application

We consider transient stability-related models in the DAE form and we have developed a Matlab-derived simulation environment. Our environment is based on PSAT, which is a suite of freely available Matlab routines well documented in [30], to which we have added our code for evaluation of measurement sensitivities and for computational differential geometry (in Python/Julia). Our Matlab code is fully general in the sense that it allows for a variety of measurements (rotor angle and speed, nodal active and reactive power injections, nodal voltage magnitudes and angles, branch active and reactive flows, and branch current magnitudes). The right-hand sides of the sensitivity equations are found using Julia's "DualNumbers" package for forward automatic differentiation [31]. The differential equations for both the model (1)–(6) and the geodesic (7) are solved using the legacy FORTRAN solver VODE [32]. Our simulations are performed for IEEE 14-bus [30, Fig. 2.4, also see *Appendix D* for detailed input data)] and real-world 441-bus test systems. In part A we deal with the model of a synchronous generator [21], [30], [33]; in part B we consider the system with AVR and PSS loops closed.

A. The Sensitivity for IEEE 14-Bus Test System

The single-line diagram of IEEE 14-bus test system and illustration of analyzed test cases are shown in Fig. 2. The dynamic model of test system is summarized in Table I (detailed

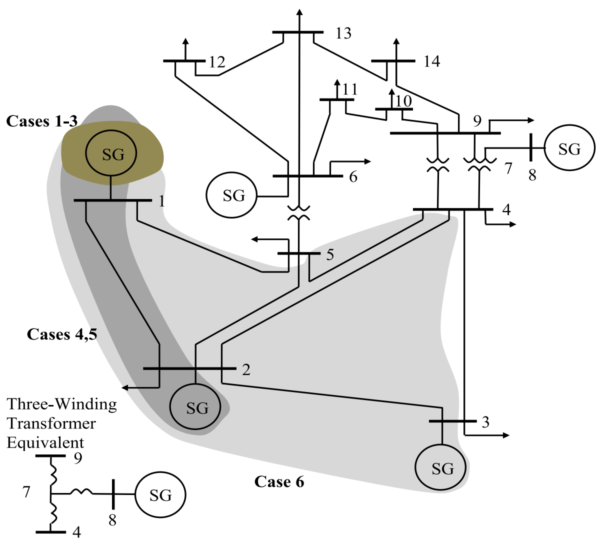

Fig. 2. Single-line diagram of IEEE 14-bus test system with analyzed cases.

TABLE II
EIGENVALUES, PARTICIPATION FACTORS AND CONDITION NUMBERS FOR CHARACTERISTIC SETS OF UNCERTAIN PARAMETER SETS ON SINGLE SG (BUS 1)

Parameters, p_ℓ	Eigenvalues, λ_ℓ	Participation factors, $p_{k\ell}$	$\kappa(\boldsymbol{H_p})$
Case 1: T'_{d0}, T'_{q0}, x_d, x'_d, x_q, x'_q	0.01 0.05 463.99 31510.47 1691525.97 2887301.24	0.41;0.57;0.01;0.00;0.00;0.00 0.25;0.09;0.65;0.00;0.00;0.00 0.30;0.30;0.31;0.00;0.00;0.08 0.03;0.03;0.03;0.00;0.22;0.69 0.00;0.00;0.00;0.89;0.08;0.03 0.00;0.00;0.00;0.11;0.69;0.19	$2.3 \cdot 10^8$
Case 2: x_d, x'_d, x_q, x'_q	164.60 29878.15 1689838.40 2875800.11	0.97;0.00;0.00;0.03 0.03;0.00;0.22;0.75 0.00;0.89;0.08;0.03 0.00;0.11;0.70;0.19	17471
Case 3: x'_d, x_q, x'_q	29052.04 1689534.41 2869080.65	0.00;0.22;0.78 0.89;0.08;0.03 0.11;0.70;0.19	98.8

TABLE I
THE MAIN CHARACTERISTICS OF DYNAMIC MODEL FOR IEEE 14-BUS TEST SYSTEM

Bus	SGs	AVRs	T-Gs	PSSs	Buses	Σ
Model order (state variables)						
1	4 (δ,ω,e'_q,e'_d)	4 (v_{r1},v_{r2},v_f)	3	-		
2	4 (δ,ω,e'_q,e'_d)	4 (v_{r1},v_{r2},v_f)	3	1 (v_1)		
3	3 (δ,ω,e'_q)	4 (v_{r1},v_{r2},v_f)	-	-		
6	6 ($\delta,\omega,e'_q,e'_d,e''_q,e''_d$)	4 (v_{r1},v_{r2},v_f)	-	-		
8	6 ($\delta,\omega,e'_q,e'_d,e''_q,e''_d$)	4 (v_{r1},v_{r2},v_f)	-	-		
Total number of state variables						
	23	20	6	1	-	50
Total number of algebraic variables (variable names)						
	5×4=20 (v_f,P_m,P_g,Q_g)	5×1=5 (v^{ref})	2×1=2	1×1=1 (v_s)	2×14=28	56

dynamic models for SG, AVR and PSS are provided in Appendix; dynamic model for T-G is omitted, because it is neglected in sloppy analysis due to slower dynamics).

For one d-axis and one q-axis SG's model described in Appendix, the state variables of interest are (A1): two mechanical variables (δ and ω) and electrical states (e'_q and e'_d, where for single q-axis model e'_d is omitted), while algebraic variables are (A2): v_f, P_m, P_g and Q_g. Parameters (electrical) of interest are: T'_{d0}, T'_{q0}, x_d, x'_d, x_q and x'_q (for single-axis model T'_{q0} is omitted). Mechanical parameters (H and D) correspond to the much slower dynamics, and are often assumed known, or estimated separately.

Available measurements for the SG are the rotor angle (δ), speed (ω), and the real and reactive powers (P_g and Q_g, respectively), as well as the terminal voltage magnitude (V) and angle (θ). Note that the rotor angle (δ) typically cannot be measured directly, but it can be estimated from local measurements. This is an old engineering problem, with first estimation-type solution proposed by Kinitsky in 1958 [34], which in turn build upon work of electric machine experts from the 1930s. More recent reference [35] consider a transient model for the machine and deploy the full power on microprocessors for online calculations. The most relevant reference for our purposes is [15], as it contains not only a very effective algorithm for estimation of the generator angle [15, eq. (8)], but also quantifies its performance under faults and for significant machine parameter mismatches.

In some special instances, such as a single machine, the algebraic equations [denoted with $\boldsymbol{g}$ in (2)] can be solved in terms of states; the algebraic variables $\boldsymbol{z}$ (1)–(3) still remain (V, θ for network buses and other for generator units). To demonstrate salient features of our method on a model that is relevant and transparent, we focus on the four differential equations on SG example [denoted with $\boldsymbol{f}$ in (1)]. However, in actual power systems (and in the IEEE 14-bus power system considered), there exist additional dynamical components (turbines, AVRs, PSSs etc.), which we consider in the next section, as well as multiple generators and loads. A single unit described by $\boldsymbol{f}$ in (1) would see these other components through variation in the field voltage (v_f), mechanical power (P_m) and in the complex voltage in the point of connection (represented by V and θ). For simplicity, we assume that interface variables for SG in Bus 1 (these are P_m, v_f, V and θ) are functions of time, but independent of the parameters considered. This, of course, is an approximation for a multi-generator system, but it allows direct comparison with numerous references that focus on a single generator.

We start transients in sensitivities following three-phase short circuit in bus 4 in $t = 0.1$ s, which cleared after 150 ms. For example, the transient variations of the voltage magnitudes and angles are approximately 4% and 20 degrees, respectively.

In Table II we present eigenvalues, participation factors and condition numbers for different uncertain parameter sets. The sloppiness of most of the uncertain parameter sets used is clear:

TABLE III
SENSITIVITY OF CONDITION NUMBERS TO THE STRUCTURE OF AVAILABLE MEASUREMENTS (LOCAL AND GLOBAL) FOR *Case 2* IN TABLE II

Local measurements		Global measurements	
Measurements	$\kappa(\boldsymbol{H_p})$	Measurements	$\kappa(\boldsymbol{H_p})$
$P_{g1}, Q_{g1}, V_1, \theta_1$	48198	$P_{g1}, Q_{g1}, V_1, \theta_1, P_{\text{inj}2}, Q_{\text{inj}2}$	958.2
$V_1, \theta_1, P_{\text{inj}1}, Q_{\text{inj}1}{}^*$	$\sim\infty$	$P_{g1}, Q_{g1}, V_1, \theta_1, P_{\text{inj}5}, Q_{\text{inj}5}$	1086.8
$P_{g1}, Q_{g1}, P_{\text{inj}1}, Q_{\text{inj}1}$	$\sim\infty$	$P_{g1}, Q_{g1}, V_1, \theta_1, P_{\text{inj}2}, Q_{\text{inj}2}, P_{\text{inj}5}, Q_{\text{inj}5}$	1086.8
$P_{g1}, Q_{g1}, V_1, \theta_1, P_{\text{inj}1}, Q_{\text{inj}1}$	509.7		

* Injection measurements are obtained from corresponding generation/load injection measurements and branch flow measurements

1) Time constants (T'_{d0} and T'_{q0}−see Appendix) are both ill-conditioned (near-zero eigenvalues are dominantly influenced by these) and cannot be estimated simultaneously from the transient (see *Case 1* in Table II).
2) SG's normal and transient reactances (x_d, x'_d, x_q and x'_q−see Appendix) are better conditioned; however, the overall condition number is high at 17473 (see *Case 2* in Table II). The most challenging SG's reactance is x_d; it influences the smallest eigenvalue dominantly (with participation factor 0.97).
3) SG reactances x'_d, x_q and x'_q (without x_d) are well-conditioned and can be estimated reliably (see *Case 3* in Table II); this is agreement with [33, Chapter 5].
4) Condition number and eigenvalue plots are changing for different uncertain parameter sets (second column in Table III).
5) Transient sensitivity of eigenvalues, whose magnitude variations are below 10%, indicates that above conclusions hold throughout the analyzed time period.
6) Condition number and eigenvalue plots vary negligibly for different short-circuit locations and time durations (see Table III).
7) All presented results are for SG in Bus 1; the conclusions are also valid for remaining generators (in buses 2, 3, 6 and 8 – regardless of different voltage levels and ratings).

In Table III we explore the variation of condition numbers to changes in measurement structure (local measurements on analyzed SG and connection bus, or a combination of local and distant bus measurements):

1) For well-conditioned parameter estimation, the SG's active/reactive power and voltage in connection point measurements are always needed.
2) Measurement of power injections improves the conditioning for parameter estimation.
3) Local injection measurements can be replaced with power flows at remote/adjacent lines. When local injection measurements are available, the global injection measurements do not improve the estimation significantly.

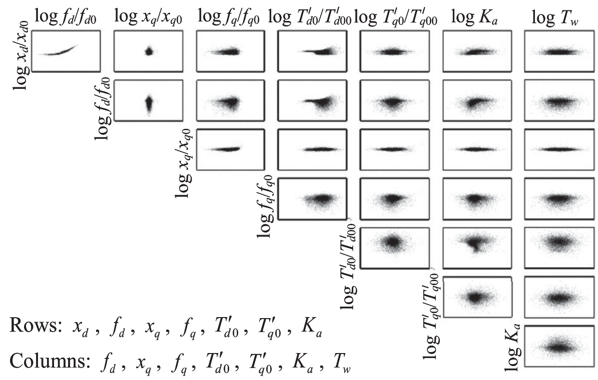

Fig. 3. Projections of the Bayesian posterior sampling for each pair of electrical log-parameters. True values are denoted with a zero subscript.

B. Information Geometry Based Local, Semi-Global and Global Results for IEEE 14-Bus Test System

In this section we consider the SG (in Bus 1) with AVR and PSS (T-G is neglected due to much slower dynamics). We have generated artificial data for a set of "true" parameter values and performed a MCMC sampling [36], [37] of the posterior distribution. To enforce the physical constraints $x_d > x'_d$ and $x_q > x'_q$ [21], we have introduced the positive parameters f_d and f_q, so that $x'_d = x_d/(1+f_d)$ and $x'_q = x_q/(1+f_q)$. The parameters to be validated are: 1) SG: x_d, f_d, x_q, f_q, T'_{d0} and T'_{q0} [30, eq. (15.29)]; 2) AVR: K_a [37, eq. (16.12)], and 3) PSS: T_w [30, eq. (16.38)].

Our results suggest that parameters x_d (first row in Fig. 3) and T'_{d0} (fourth column in Fig. 3) are likely to be difficult to estimate, which is one of the key findings of local analyses reported in the literature [4, eq. (27)]. Observe how few of the clouds are elliptical (for example, the joint distribution of x_d and f_d has a "wing" of acceptable parameter values), indicating that the local sensitivity analysis will be inadequate to capture details of the parameter correlations. Please note that our results are in very good agreement with [3], [4], [6]. Without the generator angle, the cloud for x_d extends significantly to the right, indicating that has become unconstrained from above, but otherwise all the results are qualitatively the same.

In spite of its quantitative limitation, the local analysis captures many qualitative features (e.g., the relative uncertainty of each parameter) that are confirmed by the MCMC analysis. Correlation matrices estimated from a local sensitivity analysis and a semi-global Bayesian sampling in log-parameters are shown in left and right panels in Fig. 4, respectively. The local correlation matrix correctly predicts the relative difficulty of inferring each parameter; however, it is quantitatively inaccurate when compared with the covariance matrix of the global correlation. Furthermore, from the Bayesian sampling, higher-order moments can be estimated that describe the non-quadratic features of the sampling cloud (Fig. 3).

Note the richness of the structure in many of the credible regions. Simply stating accuracy of individual parameters (e.g., x_d was constrained to be within 10% of the true value), or even

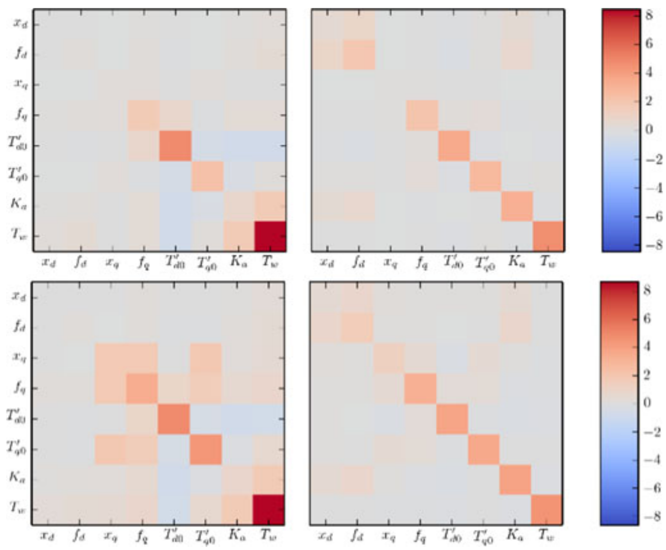

Fig. 4. Correlation matrices estimated from a local sensitivity analysis (left) and a semi-global Bayesian sampling (right) in log-parameters; top row – with angle measurements, bottom row – without angle. The variance of individual parameters is given by the diagonal elements. Note that a variance of 6 in log-parameters corresponds roughly to a relative error of about 90%. The two pairs of matrices have many of the same qualitative features. The local correlation correctly predicts that that T'_{q0}, T'_{d0} and f_q will be difficult to identify. However, there are quantitative differences between each set of two matrices. Furthermore, the covariance matrices do not reflect the rich structure of the credible regions in Fig. 3.

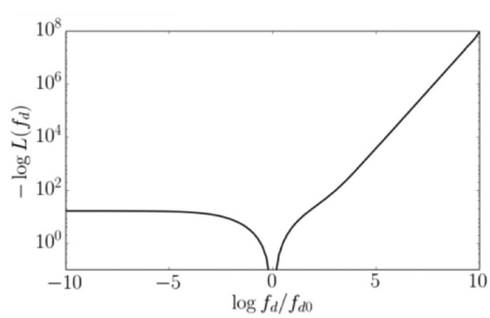

Fig. 5. Non-Bayesian scanning – likelihood variation for f_d. Notice that f_d is unidentifiable from below, but is constrained from above.

parameter correlations fails to convey the structure in these point clouds. Furthermore, because nonlinear effects play an important role in defining the extent of the credible regions, the clouds analogous to those in Fig. 3 are likely to vary depending on the SG's "true" parameter values and the scale of the measurement noise to which they are fit.

Next, we consider a scanning method that is a non-Bayesian alternative to MCMC sampling. We fix one parameter (or more) and fit the model by varying the remaining parameters; we identify the confidence intervals for each parameter by considering what range still yields a good fit (measured by log likelihood). For example, in the case of the parameter f_d (Fig. 5), notice that it can be made very small, but has an upper cut-off. This is confirmed by MCMC sampling (Fig. 3, top left plot). We note that this scanning method has better scaling properties in the case of large systems than MCMC, provided likelihood profiles are calculated for one parameter at a time.

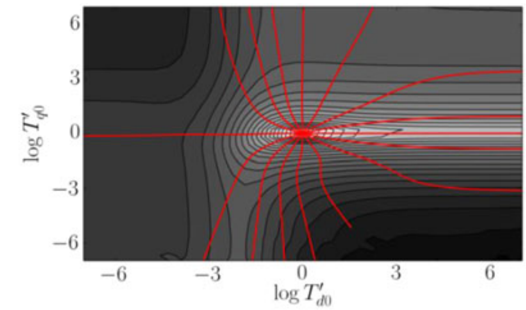

Fig. 6. Global exploration of a two-dimensional cross section of the manifold corresponding T'_{q0} and T'_{d0}. The contours are the joint likelihood profile for the two parameters. The red lines are geodesics on the model manifold. Notice that the geodesics align with the four "canyons" of the cost surface.

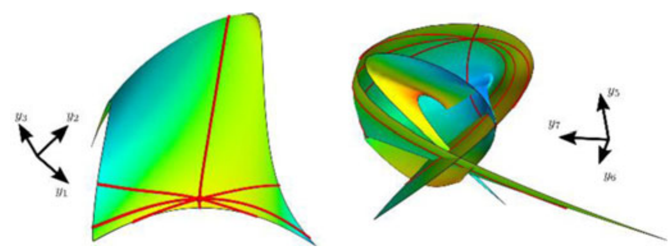

Fig. 7. Data space visualization of the model manifold for the two-dimensional cross section defined by T'_{d0} and T'_{q0} projected onto the first three principal components in data space (left) and the fifth through seventh principal components (right). Color represents the fourth principal component (left) and eighth principal component (right). Red lines are geodesics curves (corresponding to a subset of those in Fig. 6 for clarity). This cross section has four edges and four corners (like a deformed square). The four edges of the model manifold correspond to the four "canyons" in Fig. 6.

In Fig. 7, we consider the joint likelihood profile for the two time-scale parameters T'_{d0} and T'_{q0} (cf. Fig. 3, bottom right panel). The contours correspond to level sets of the negative log-likelihood. Notice that the cost surface has four "canyons". The primary (deepest) canyon extends from the origin to the right (T'_{d0} to infinity). In addition, there are two more shallow canyons running up and down from the origin along the T'_{q0} axis. Finally, there is a fourth, very shallow canyon running to the left from the origin, corresponding to T'_{d0} going to zero.

The red lines in Fig. 6 are geodesics on the model manifold restricted to the two-dimensional cross section spanned by T'_{d0} and T'_{q0}. These curves are highly nonlinear; however, they tend to bend in agreement with the local sensitivity analysis (e.g., aligning the canyons of the likelihood profile). The nonlinearity of the geodesic curves reflects the incompleteness of the local analysis. However, the geodesics naturally "connect" the local analyses to reveal the global structure of the parameter space.

To better understand the relation between the information geometry of the model and parameter identifiability, consider the visualization of the model manifold in data space in Fig. 7.

The cross section has four edges and four corners−like a deformed square. These edges correspond to the limits that T'_{d0} and T'_{q0} become either zero or infinity. The edges are closely related to the four "canyons" in Fig. 6. As one follows the canyons toward extreme parameter values, the canyons slowly rise up, reflecting a decrease in quality of fit, but then

TABLE IV
EIGENVALUES, PARTICIPATION FACTORS AND CONDITION NUMBERS FOR THREE CHARACTERISTIC SETS OF UNCERTAIN PARAMETERS IN AREA (SGS AND NETWORK PARAMETERS)

Uncertain parameters, p_ℓ	Eigenvalues, λ_ℓ	Participation factors, $p_{k\ell}$	Condition number, $\kappa(\boldsymbol{H}_p)$
Case 4:	193.144	0.0066;0.0044;0.0207;0.2330;0.0003;0.7316;0.0035	542.329
x'_{d1}, x_{q1}, x'_{q1},	249.135	0.0010;0.0041;0.0360;0.1273;0.7818;0.0498;0.0000	
	572.138	0.0010;0.0606;0.5016;0.3741;0.0166;0.0439;0.0020	
x'_{d2}, x_{q2}, x'_{q2},	3123.810	0.7436;0.1419;0.0236;0.0516;0.0213;0.0097;0.0083	
	10023.514	0.0089;0.5412;0.2571;0.0515;0.0311;0.0168;0.0933	
x_{1-2}	36575.992	0.1263;0.1691;0.0003;0.0350;0.0229;0.0149;0.6315	
	104747.433	0.1127;0.0787;0.1606;0.1275;0.1259;0.1332;0.2613	
Case 5:	0.550	0.0742;0.0005;0.0000;0.0000;0.7310;0.0024;0.0823;0.1095;0.0000	240917.701
x_{d1}, x'_{d1}, x_{q1},	3.490	0.6579;0.0046;0.0000;0.0008;0.0013;0.0001;0.1481;0.1872;0.0000	
	220.985	0.0109;0.0016;0.0003;0.0001;0.0140;0.0483;0.3889;0.5334;0.0025	
x'_{q1}, x_{d2}, x'_{d2},	317.241	0.0794;0.0028;0.0072;0.0467;0.1023;0.5334;0.2199;0.0061;0.0021	
	595.621	0.0185;0.0011;0.0754;0.6081;0.0188;0.2252;0.0150;0.0367;0.0013	
x_{q2}, x'_{q2},	3257.122	0.0260;0.7735;0.1377;0.0060;0.0056;0.0290;0.0105;0.0035;0.0082	
	10440.035	0.0192;0.0244;0.5768;0.2082;0.0117;0.0331;0.0184;0.0080;0.1003	
x_{1-2}	37428.164	0.0070;0.0980;0.1369;0.0004;0.0095;0.0229;0.0134;0.0073;0.7046	
	132526.931	0.1068;0.0935;0.0656;0.1298;0.1058;0.1056;0.1035;0.1084;0.1810	
Case 6:	1.843	0.0000;0.0000;0.0000;0.0009;0.1618;0.1946;0.0000;0.6427;0.0000;0.0000;0.0000;0.0000;0.0000;0.0000	321361.216
x'_{d1}, x_{q1}, x'_{q1},	111.623	0.0036;0.0003;0.0022;0.1654;0.1825;0.1504;0.0009;0.1974;0.2914;0.0001;0.0025;0.0025;0.0007;0.0000	
	195.371	0.0009;0.0002;0.0001;0.0588;0.4059;0.4704;0.0000;0.0044;0.0528;0.0016;0.0010;0.0002;0.0036;0.0000	
x'_{d2}, x_{q2}, x'_{q2},	267.743	0.0018;0.0000;0.0000;0.4169;0.0910;0.0339;0.0003;0.0054;0.4457;0.0003;0.0012;0.0008;0.0029;0.0000	
	563.144	0.0328;0.0933;0.7077;0.0259;0.0125;0.0081;0.0000;0.0087;0.0206;0.0003;0.0002;0.0896;0.0003;0.0000	
x'_{d3}, x_{q3}, x'_{q3},	700.509	0.2812;0.0011;0.0126;0.1633;0.0403;0.0421;0.0011;0.0389;0.0800;0.0000;0.0012;0.3337;0.0044;0.0000	
	4454.557	0.1990;0.0246;0.0401;0.0268;0.0083;0.0150;0.6080;0.0120;0.0178;0.0057;0.0262;0.0140;0.0019;0.0007	
$x_{1-2}, x_{2-3}, x_{1-5}$,	5118.436	0.2648;0.1957;0.0048;0.0021;0.0002;0.0000;0.2117;0.0001;0.0004;0.0076;0.0099;0.2941;0.0086;0.0000	
x_{2-4}, x_{5-4}	5118.436	0.0000;0.0539;0.0000;0.0114;0.0048;0.0002;0.0000;0.0017;0.0004;0.0350;0.0877;0.0010;0.7977;0.0061	
	11155.955	0.0417;0.3223;0.0710;0.0005;0.0038;0.0001;0.0129;0.0005;0.0000;0.1496;0.3331;0.0629;0.0003;0.0014	
	15138.216	0.0157;0.1060;0.0909;0.0435;0.0120;0.0125;0.0814;0.0132;0.0167;0.0013;0.4222;0.1008;0.0588;0.0250	
	20972.739	0.0113;0.0260;0.0000;0.0010;0.0025;0.0028;0.0031;0.0026;0.0022;0.0000;0.0196;0.0002;0.0340;0.8944	
	39680.315	0.0860;0.1220;0.0009;0.0196;0.0105;0.0053;0.0130;0.0079;0.0076;0.7030;0.0042;0.0044;0.0072;0.0083	
	592223.651	0.0611;0.0546;0.0695;0.0639;0.0639;0.0648;0.0676;0.0643;0.0643;0.0955;0.0910;0.0959;0.0795;0.0641	

TABLE V
MAXIMAL AND AVERAGE ERRORS IN TIME RESPONSES OF ORIGINAL AND SIMPLIFIED MODELS FOR REAL-WORLD TEST SYSTEM

$\frac{1}{T}\sum_{t=1}^{T}\lvert V_t^{Or.} - V_t^{Sl.}\rvert$		$\frac{1}{T}\sum_{t=1}^{T}\lvert\theta_t^{Or.} - \theta_t^{Sl.}\rvert$		$\frac{1}{T}\sum_{t=1}^{T}\lvert P_{g,t}^{Or.} - P_{g,t}^{Sl.}\rvert$		$\frac{1}{T}\sum_{t=1}^{T}\lvert Q_{g,t}^{Or.} - Q_{g,t}^{Sl.}\rvert$	
Max (Bus 433)	Aver.	Max (Bus 429)	Aver.	Max (Bus 15)	Aver.	Max (Bus 7)	Aver.
0.0054	0.0010	0.0751	0.0732	0.0410	0.0011	0.0199	0.0008

plateau. Thus, the parameters can be taken to extreme values (e.g., zero of infinity) with finite cost. Because they correspond to extreme parameter values with finite cost, each canyon can identified with one edge of the model manifold. The manifold boundaries make it possible to have infinite confidence regions. The depth of each canyon at extreme parameter values is the inverse distance to the corresponding edge on the model manifold. Which parameters will have infinite confidence regions will in general depend on the "true" parameter values, the measurement noise in the data, and the level of statistical confidence.

We now consider the boundaries of the six-dimensional manifold corresponding to the six electric parameters in a SG, one of the primary results of this study. Because the manifold is not easily visualized, we report the results of our geodesic analysis. In addition to T'_{d0} and T'_{q0} each being unidentifiable from above or below, we also find that x_d and f_d could each become infinite in a correlated way (i.e., so that $x_d - x'_d$ remains finite), f_d could become infinite (x'_d goes to zero) or zero (x'_d goes to x_d) with similar results for x_q and x'_q.

Significantly, these results give a global characterization of the structure of the model manifold. Unlike local analyses that focus on the relationship between a specific realization of data and parameters, the global analysis characterizes properties of the model that are transferable to other data sets. In particular, it identifies which parameters could potentially be unidentifiable for different measurements or data of different quality.

C. System-Wide Estimation for IEEE 14-Bus Test System

The identification of power system areas ("dynamic equivalents") is a key for application of our method in industrial practice. As an illustration, we consider the case when, in addition to SG's and appropriate control's parameters, the transmission network needs to be identified as well. The elements of transmission network are described by the static branch equivalents (lines and transformers) with uncertain resistances, reactances and susceptances; the influence of reactances to the dynamic behavior is typically considered dominant. We assume a full set of available measurements: SG's output active/reactive powers,

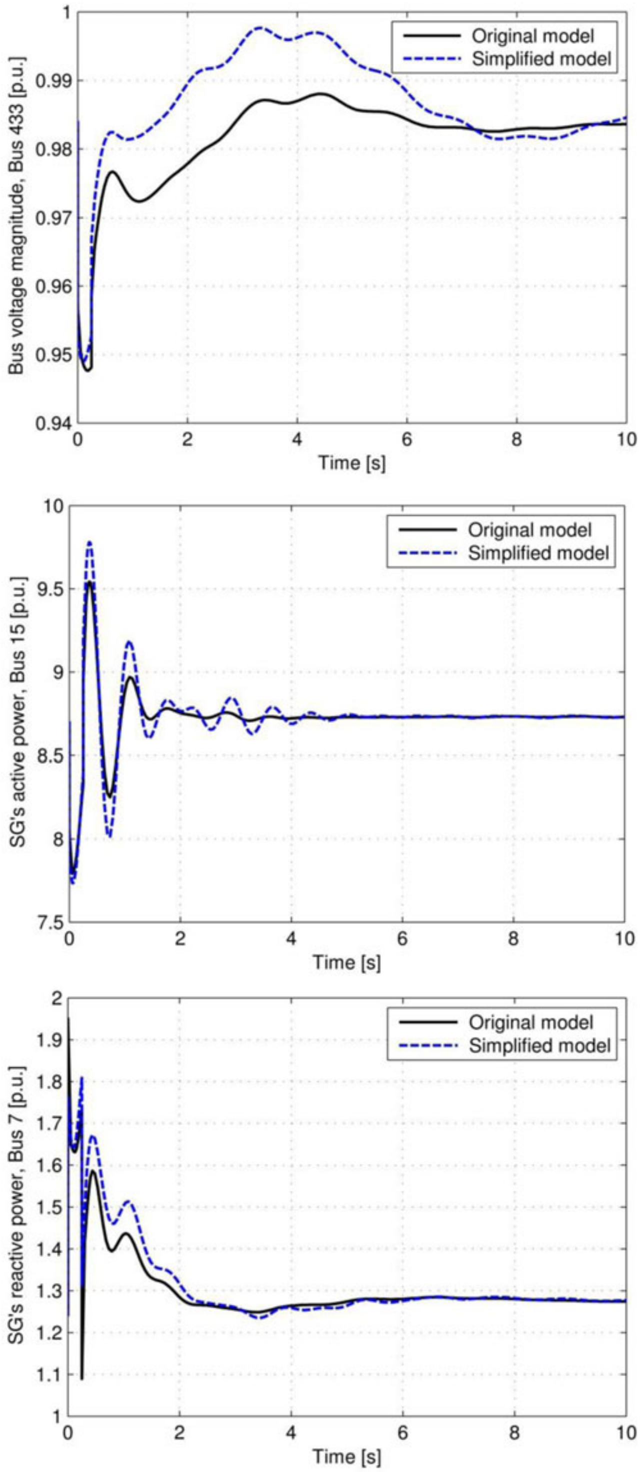

Fig. 8. Time responses of bus voltage and SG's active and reactive powers with maximal errors (from Table V) for original and simplified models for real-world test system.

bus voltage magnitude/angle, bus active/reactive power injections and branch active/reactive power flows. In Table IV we display the eigenvalues of the Hessian matrix, participation factors and condition numbers for three uncertain parameter sets (three SGs and five branches – *Case 6*):

1) The conclusions about quality of parameter identification are aligned with those of Table II.
2) Largest eigenvalues are dominantly influenced by uncertain branch reactances, making them relatively easy to estimate.
3) Quality of parameter identification is worsening with (spatial) increase of uncertain area.

Our strategy for extending the approach to large systems involves the following steps:

1) Determination of key modes to be included in the system model with key states (via participation factors of the system matrix), and corresponding physical components and their vital ("systemic") parameters (via sensitivities of the system matrix to parameters).
2) Measurement structure and model selection to achieve low to moderate sloppiness and the presence of all systemic parameters among the identifiable ones.
3) Nodal local and semi-global analysis for typical transients. For example, in the test system in Fig. 1, the least damped pole pair is at $-0.61 \pm$ j10.89 (thus critical for oscillation damping). The sensitivities of these eigenvalues to parameters of SG in Bus 1 vary widely, and are largest for x'_{q1} (0.92) and x'_{d1} (0.05); thus it makes sense to declare x'_{q1} as a systemic parameter. This parameter is retained in even very low-order models [30], so area or SG identification with typical models will suffice for the critical pole as far as SG in Bus 1 is concerned.

D. Real-World Test System

The main characteristics of original dynamic model for real-world (Electric Power Industry of Serbia, a part of the ENTSO-E interconnection) power system can be summarized as: 441 buses, 655 branches (lines and transformers), 72 SGs (43 of 4-order models and 29 of 6-order models), with AVRs and turbines. The model has total 850 differential and 1314 algebraic variables. Test system is subjected to the three-phase short circuit in Bus 1 in $t = 0.0$ s, which cleared after 250 ms (fault internal impedance is $\underline{Z}_f = (0 + \mathrm{j}0.1)$ p.u.).

In this section we show how insights gained from the sloppiness analysis may be used to simplify the model of a real-life power system, while maintaining remarkable fidelity of the response with very little tuning. Based on conclusions derived in Sections V.B and V.C, related to stiff and sloppy SG's parameters, we assume: 1) all SG's models are 4-order; all SG's time constants are adjusted to typical values: $T'_{d0} = 7$ s and $T'_{q0} = 0.2$ s; all SG's reactances in d-axis are adjusted to typical values: $x_d = 0.3$ p.u. Maximal and average errors in time responses of original and simplified models for real-world test system are reported in Table V. In Fig. 8 we show time responses of bus voltage and SG's active and reactive powers with maximal errors reported in Table V.

Results of this pilot study are encouraging, suggesting that models with low, managed sloppiness can indeed be useful in industrial practice. A case-specific re-parametrization may reduce or even eliminate sloppiness, but it is unlikely to be useful for other stakeholders in the power system enterprise.

VI. Conclusion

This paper outlines a new class of system identification procedures tailored to electric power systems. Our procedure, illustrated on synchronous generator example, builds on computational advances in differential geometry, and offers a new, global characterization of challenges encountered in system identification of electric power systems. From an optimization perspective, we offer domain-specific tools for regularization of ill-posed optimization models that are prevalent in studies of power system dynamics and in static state estimation. One of the major challenges to system identification in large models is method scalability. Somewhat surprisingly, the global methods we advocate scale well with system size, since they exploit the low effective dimensionality of the model manifold in data space.

Appendix

For one d-axis and one q-axis SG model, the differential and algebraic equations respectively can be written as [30, eqs. (15.5), (15.29), (15.7), (15.8) and (15.2)–(15.4), respectively]:

$$\boldsymbol{f} = \begin{cases} \dot{\delta} = \Omega_b(\omega - \omega_s) \\ \dot{\omega} = \dfrac{1}{2H}\left(P_m - P_g - D(\omega - \omega_s)\right) \\ \dot{e}'_q = \dfrac{1}{T'_{d0}}\left(-e'_q - (x_d - x'_d)i_d + v_f\right) \\ \dot{e}'_d = \dfrac{1}{T'_{q0}}\left(-e'_d + (x_q - x'_q)i_q\right) \end{cases} \tag{A1a}$$

$$\boldsymbol{g} = \begin{cases} v_f = v_{f0} \\ P_m = P_{m0} \\ P_g = v_d i_d + v_q i_q \; = V\sin(\delta-\theta)\frac{e'_q - V\cos(\delta-\theta)}{x'_d} \\ \qquad + V\cos(\delta-\theta)\dfrac{V\sin(\delta-\theta) - e'_d}{x'_q} . \\ Q_g = v_q i_d - v_d i_q \; = V\cos(\delta-\theta)\frac{e'_q - V\cos(\delta-\theta)}{x'_d} \\ \qquad - V\sin(\delta-\theta)\dfrac{V\sin(\delta-\theta) - e'_d}{x'_q} \end{cases} \tag{A1b}$$

For one q-axis (3-order) model differential equation for $\dot{e}'_d$ in (A1a) is neglected.

Additional differential equations for two d- and two q-axes (6-order) model (for e''_q and e''_d– see Table I) can be found in [30, eq. (15.9)] or [17, Appendix].

AVR (based on a typical DC-based exciter) model is described as [30, eq. (16.12)]:

$$\boldsymbol{f} = \begin{cases} \dot{v}_{r1} = \dfrac{K_a}{T_a}\left(v^{\mathrm{ref}} - v_m - v_{r2} - \dfrac{K_f}{T_f}v_f - v_{r1}\right) \\ \dot{v}_{r2} = -\dfrac{1}{T_f}\left(\dfrac{K_f}{T_f}v_f - v_{r2}\right) \\ \dot{v}_f = \dfrac{1}{T_e}\left(-v_f(K_e + S_e(v_f) - v_{r1}\right) \end{cases} \tag{A2a}$$

$$\boldsymbol{g} = \left\{ v^{\mathrm{ref}} = v_0^{\mathrm{ref}} \right. . \tag{A2b}$$

PSS model is described as [30, eq. (16.38)]:

$$\boldsymbol{f} = \left\{ \dot{v}_1 = -\frac{1}{T_w}\left(K_\omega \omega + K_p P_g + K_v V + v_1\right) \right. \tag{A3a}$$

$$\boldsymbol{g} = \left\{ v_s = K_\omega \omega + K_p P_g + K_v V + v_1 \right. . \tag{A3b}$$

Details of larger models for SGs, AVRs, T-Gs and PSS can be found in [21], [30], [33].

Bus active/reactive power balance equations in ith bus respectively are:

$$P_{gi} - P_{pi} = P_{\mathrm{net},i} \tag{A4a}$$

$$Q_{gi} - Q_{pi} = Q_{\mathrm{net},i}, \tag{A4b}$$

where:

P_{gi}, Q_{gi} – active and reactive power generations in ith bus, respectively [given in (A1b)];

P_{pi}, Q_{pi} – active and reactive loads in ith bus, respectively;

$P_{\mathrm{net},i}$, $Q_{\mathrm{net},i}$ – active and reactive power network flows in ith bus, respectively:

$$P_{\mathrm{net},i} = \sum_{j=1}^{N}\left[V_i V_j \left(G_{ij}\cos(\theta_i - \theta_j) + B_{ij}\sin(\theta_i - \theta_j)\right)\right] \tag{A5a}$$

$$Q_{\mathrm{net},i} = \sum_{j=1}^{N}\left[V_i V_j \left(G_{ij}\sin(\theta_i - \theta_j) - B_{ij}\cos(\theta_i - \theta_j)\right)\right], \tag{A5b}$$

where G_{ij} and B_{ij} are elements of bus admittance matrix $\underline{\boldsymbol{Y}} = \boldsymbol{G} + j\boldsymbol{B}$.

References

[1] G. Andersson *et al.*, "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1922–1928, Nov. 2005.

[2] D.N. Kosterev, C.W. Taylor, and W.A. Mittelstadt, "Model validation the August 10, 1996 WSCC system outage," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 967–979, Aug. 1999.

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

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

[5] I. A. Hiskens, "Nonlinear dynamic model evaluation from disturbance measurement," *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 702–710, Nov. 2001.

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

[7] P. Overholt, D. Kosterev, J. Eto, S. Young, and B. Lesieutre, "Improving reliability through better models," *IEEE Power Energy Mag.*, vol. 12, no. 3, pp. 44–51, May/Jun. 2014.

[8] K. S. Brown, and J. P. Sethna, "Statistical mechanical approaches to models with many poorly known parameters," *Phys. Rev. E*, vol. 68, no. 2, pp. 21904-1–21904-9, Aug. 2003.

[9] M.K. Transtrum *et al.*, "Perspective: Sloppiness and emergent theories in physics, biology, and beyond," *J. Chem. Phys.*, vol. 143, no. 1, pp. 010901-1–010901-13, Jul. 2015.

[10] K. Brown *et al.*, "The statistical mechanics of complex signaling networks: Nerve growth factor signaling," *Phys. Biol.*, vol. 1, no. 3, pp. 184−195, Sep. 2004.
[11] R. Gutenkunst *et al.*, "Universally sloppy parameter sensitivities in systems biology models," *PLoS Comput. Biol.*, vol. 3, no. 10, pp. 1871−1878, Oct. 2007.
[12] C. Tonsing, J. Timmer, and C. Kreutz, "Cause and cure of sloppiness in ordinary differential equation models," *Phys. Rev. E*, vol. 90, no. 2, 023303-1−023303-16, Aug. 2014.
[13] E. Cari and L.F.C. Alberto, "Parameter estimation of synchronous generators from different types of disturbances," in *Proc. IEEE Power Energy Soc. General Meeting*, 2011, pp. 1−7.
[14] Z. Huang *et al.*, "Generator dynamic model validation and parameter calibration using phasor measurements at the point of connection," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1939−1949, May 2013.
[15] E. Ghahremani and I. Kamwa, "Local and wide-area PMU-based decentralized dynamic state estimation in multi-machine power systems," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 547−562, Feb. 2015.
[16] M. K. Transtrum, A. T. Sarić, and A. M. Stanković;, "Information geometry for model verification in energy systems," in *Proc. 19th Power Syst. Comput. Conf.*, Genoa, Italy, Jun. 20–24, 2016, pp. 1−7.
[17] M.K. Transtrum, A.T. Sarić, and A. M. Stanković, "Measurement-directed reduction of dynamic models in power systems," *IEEE Trans. Power Syst.*, to be published.
[18] L.B. Rall, *Automatic Differentiation: Techniques and Applications*. New York, NY, USA: Springer-Verlag, 1981.
[19] C.H. Bischof, H.M. Bücker, P. Hovland, U. Naumann, and J. Utke, Eds., *Advances in Automatic Differentiation*. Berlin, Germany: Springer-Verlag, 2008.
[20] B.K. Choi and H.-D. Chiang, "Multiple solutions and plateau phenomenon in measurement-based load model development: issues and suggestions," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 824–831, May 2009.
[21] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.
[22] M.K. Transtrum, B.B. Machta, and J.P. Sethna, "Why are nonlinear fits to data so challenging?," *Phys. Rev. Lett.*, vol. 104, no. 6, pp. 060201-1−060201-4, Feb. 2010.
[23] S.-I. Amari, *Differential-Geometrical Methods in Statistics* (Lecture Notes in Statistics). Berlin, Germany: Springer-Verlag, 1985.
[24] M. Girolami and B. Calderhead, "Riemann manifold, Langevin and Hamiltonian Monte Carlo methods," *J. Roy. Statist. Soc.: Series B (Statist. Methodol.)*, vol. 73, no. 2, pp. 123–214, Mar. 2011.
[25] S. Marino, I.B. Hogue, C.J. Ray, and D.E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," *J. Theoretical Biol.*, vol. 254, no. 11, pp. 178–196, Sep. 2008.
[26] S. Chib and E. Greenberg, "Understanding the metropolis Hastings algorithm," *Amer. Statist.*, vol. 49, no. 4, pp. 327−335, Nov. 1995.
[27] M.K. Transtrum and P. Qiu, "Model reduction by manifold boundaries," *Phys. Rev. Lett.*, vol. 113, no. 9, pp. 098701-1–098701-6, Aug. 2014.
[28] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation*. San Francisco, CA, USA: W.H. Freeman, 1973.
[29] M.K. Transtrum, G. Hart, and P. Qiu, "Information topology identifies emergent model classes," [Online]. Available: http://arxiv.org/abs/1409.6203
[30] F. Milano, *Power System Modelling and Scripting*. London, U.K.: Springer, 2010.
[31] DualNumbers.jl, GitHub repository. 2017. [Online]. Available: https://github.com/JuliaDiff/DualNumbers.jl
[32] P.N. Brown, G.D. Byrne, and A.C. Hindmarsh, "VODE: A variable-coefficient ODE solver," *SIAM J. Sci. Statist. Comput.*, vol. 10, no. 5, pp. 1038−1051, Sep. 1989.
[33] P.W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[34] V.A. Kinitsky, "Measurement of rotor displacement angle on synchronous machines," *AIEE Trans. Power App. Syst.*, vol. 77, no. 3, pp. 349−352, Apr. 1958.
[35] G. Valverde, E. Kyriakides, G.T. Heydt, and V. Terzija, "Nonlinear estimation of synchronous machine parameters using operating data," *IEEE Trans. Energy Convers.*, vol. 26, no. 3, pp. 831–839, Sep. 2011.
[36] I. T. Jolliffe, *Principal Component Analysis* (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer-Verlag, 2002.
[37] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The MCMC hammer," *Publication Astronomical Soc. Pacific*, vol. 125, no. 925, pp. 306−312, Mar. 2013.

Mark K. Transtrum received the Ph.D. degree in physics from Cornell University, Ithaca, NY, USA, in 2011. He then studied computational biology as a Postdoctoral Fellow at MD Anderson Cancer Center, Houston, TX, USA. Since 2013, he has been an Assistant Professor of physics and astronomy at Brigham Young University, Provo, UT, USA. His work considers representations of a variety of complex systems including power systems, systems biology, materials science, and neuroscience.

Andrija T. Sarić received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the University of Belgrade, Belgrade, Serbia, in 1988, 1992, and 1997, respectively. He is a Full Professor of electrical engineering at the Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia. His main areas of interest are power system analysis, optimization and planning, as well as application of artificial intelligence methods in these areas.

Aleksandar M. Stanković (F'05) received the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1993. From 1993 to 2010, he was with Northeastern University, Boston. He currently serves as an A.H. Howell Professor at Tufts University, Medford, MA, USA. He has held visiting positions at the United Technologies Research Center (sabbaticals in 2000 and 2007) and at L'Universite de Paris-Sud and Supelec (in 2004). He is a co-editor of a book series on power electronics and power systems for Springer.

Dr. Stanković is an Associate Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS. He has previously served in the IEEE TRANSACTIONS ON SMART GRID in power systems and on control system technology in the same capacity since 1996.