

State Estimation Model Reduction Through the Manifold Boundary Approximation Method

Vanja G. Švenda¹, Student Member, IEEE, Mark K. Transtrum², Benjamin L. Francis,
Andrija T. Sarić³, Member, IEEE, and Aleksandar M. Stanković⁴, Fellow, IEEE

Abstract—This paper presents a procedure for estimating the systems state when considerable Information and Communication Technology (ICT) component outages occur, leaving entire system areas unobservable. For this task, a novel method for analyzing system observability is proposed based on the Manifold Boundary Approximation Method (MBAM). By utilizing information geometry, MBAM analyzes boundaries of models in data space, thus detecting unidentifiable system parameters and states based on available data. This approach extends local, matrix-based methods to a global perspective, making it capable of detecting both structurally unidentifiable parameters as well as practically unidentifiable parameters (i.e., identifiable with low accuracy). Beyond partitioning identifiable/unidentifiable states, MBAM also reduces the model to remove reference to the unidentifiable state variables. To test this procedure, cyber-physical system (CPS) simulation environments are constructed by co-simulating the physical and cyber system layers.

Index Terms—State estimation, model reduction, communication outages, manifold boundary approximation method, information geometry.

I. INTRODUCTION

WITH increasing demands from their customers and regulators, electric utilities have started to focus on the CPS concept, with potent communication systems based on ICT, being one of its main features. Following this growing tendency, work to analyze and improve the reliability and usage of system ICT has emerged. For instance, a survey on available CPS test-beds and the development of a novel one for reliability studies are given in [1], [2], respectively. Additionally, analysis of the coexisting cyber and physical layers, along with the

concept formulation of cyber-physical reliability are examined through various approaches in [3]–[6].

This paper tackles a specific cyber-physical interconnection predicament—estimating the system state when significant ICT component outages occur, leaving whole system areas unobservable. Conventionally, this problem is approached by defining system areas that are observable with regards to available ICT and focusing state estimation (SE) specifically on them [7]. Various methods have been proposed, most of which focus on matrices derived through the SE problem, for instance finding null pivots through the Gauss elimination of the gain matrix [8], [9], or the Jacobian matrix [10], transferring rows to columns and vice versa in the Jacobian matrix [11], performing back substitutions on the triangular factors of a gain matrix in a non-iterative manner [12], using *a priori* information which is assigned a degree of confidence under the form of a covariance matrix [13], detecting near-zero eigenvalues through the eigenvalue decomposition of the Fisher information matrix, calculated through the Jacobian matrix [14], analyzing the measurement redundancy level through null space calculations of the Jacobian matrix [15], [16], factorizing the Gram matrix, which is calculated through the Jacobian matrix [17], observability analysis based on the triangular factorization of the Jacobian matrix [18], and finally, an algorithm based on a greedy strategy combined with a Gram-Schmidt orthonormalization of the Jacobian matrix [19].

The work presented here goes beyond local, matrix-based approaches by using a global analysis based on information geometry that additionally reduces the SE model while respecting available ICT information. For such a task, a novel system observability analysis model is proposed, based on the MBAM [20]. The MBAM utilizes information geometry where the SE model is given a geometric interpretation as a manifold of possible predictions in data space, known as the model manifold [21].

The geometric approach has several advantages. First, it extends local, matrix-based approaches to a global sensitivity analysis, making it capable of detecting both structurally unidentifiable parameters as well as practically unidentifiable parameters (i.e., parameters that are in principle identifiable but with very low accuracy). It also provides a natural way of reducing the SE model. Unlike other methods in which unidentifiable states are fixed to last-known values, MBAM-reduced models take unidentifiable parameters to extreme values. Limiting cases fundamentally change the computational structure of the

Manuscript received October 18, 2020; revised March 14, 2021; accepted June 6, 2021. Date of publication June 22, 2021; date of current version December 23, 2021. This work was supported by NSF under Grant ECCS-1710944, in part by the CURENT Engineering Research Center of the National Science Foundation and the Department of Energy under NSF Award Number EEC-1041877, and in part by the ONR under Grant N00014-16-1-3028. Paper no. TPWRS-01745-2020. (Corresponding author: Aleksandar Stanković.)

Vanja G. Švenda and Aleksandar M. Stanković are with the Department of Electrical Engineering and Computer Science, Tufts University, Medford, MA 02155 USA. (e-mail: vanja.svenda@tufts.edu; astankov@ece.tufts.edu).

Mark K. Transtrum and Benjamin L. Francis are with the Brigham Young University, Department of Physics and Astronomy, Provo, UT 84602 USA (e-mail: mktranstrum@byu.edu; zor_quatre@yahoo.com).

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

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TPWRS.2021.3091547>.

Digital Object Identifier 10.1109/TPWRS.2021.3091547

model, so that there is no reference to unidentifiable elements. Thus, the computational complexity of the model is commensurate with the information complexity of the data, allowing us to extract the most information possible from the available measurements [22].

MBAM was first proposed for the problem of parameter identifiability in complex, nonlinear models [20]. Recently, authors have investigated using this methodology to reduce dynamic electric system models, for instance, synchronous generators [21], [23], and transient stability models [24].

Here, MBAM is utilized for a different goal—reducing SE models for unidentifiable states and line parameters with regards to (un)available ICT. System line parameters present crucial input information to any SE model, and as such, have been the subject of various SE-based research. For instance, this led to the development of a joint state and line parameter tracking model [25], and a robust SE model that can withstand line parameter errors [26]. In this paper, however, unidentifiable line parameters are additionally reduced using MBAM. Once an ICT outage occurs, we use MBAM first to define unidentifiable states and line parameters for unobservable system areas, and then to reduce the system model. Thus, the estimation problem is reformulated to be commensurate with the available ICT. By removing irrelevant state and line parameters, the reduced model is observable with regards to available ICT, extracting the most from available measurements and deriving the best possible state estimates over the largest system area.

Unlike other methods in which unidentifiable states are fixed to last-known values, MBAM-reduced models take unidentifiable parameters to extreme values. Limiting cases simplify the computational structure of the model. When voltages are taken to zero, for example, they no longer appear in the model; however, if they were fixed to a non-zero value the value would still contribute to the computational structure of each evaluation of the model. For more details on the underlying mechanisms, the reader is referred to [20].

To summarize, the main contribution of this paper is developing a novel observability analysis model that:

- Detects and differentiates structurally and practically unidentifiable parameters.
- Performs global observability analysis.
- Reduces the SE model with regards to unobservable states.
- Reduces the system model with regards to unobservable branch parameters.

The remainder of this paper is organized as follows: Section II formulates the problem, along with the overall system observability model. Section III explains the general idea of MBAM, while its application on SE is given in Section IV. The proposed algorithm and numerical results are given in Sections V and VI, respectively. A list of references follows conclusions in Section VII.

II. PROBLEM FORMULATION

The main goal of this paper is to obtain the best state estimates (voltage phasors at all system buses) based on available ICT results. For this task, the static SE (SSE) model is utilized, which

presents one of the most commonly used models in real-life electric utilities [8], [27]. For the n -th time instant, SSE can be formulated as [28]

$$z(t_n) = h(x(t_n)) + e(t_n) \quad (1)$$

where:

- $z(\cdot)$, $x(\cdot)$ measurement and state vectors, respectively;
- $h(\cdot)$ vector of nonlinear functions relating the measurement and state vectors;
- $e(\cdot)$ measurement error vector.

To solve (1), the weighted least square (WLS) based approach is utilized.

For this work, we assume systems are formulated as CPS, consisting of the following, strongly intertwined parts:

1. *Cyber layer*—primarily used for information exchange. Formed by ICT components, such as measuring instruments, alarms, transmission channels, data centers, etc. [29].
2. *Physical layer*—primarily used for electric power generation, transmission, and consumption. Formed by electric components, such as lines, generators, transformers, etc. [30].
3. *Control center*—the main connection between the two layers as it takes information from the cyber layer, processes it, and gives feedback to the physical layer (e.g., needed actions). In this paper, the energy management system is considered as the control center.

Finally, as ICT component outages are taken into account, an appropriate system observability model is formulated. This is first done for the general case of estimating any system parameters (p), after which the focus shifts specifically to SSE.

A. System Observability

Estimating any system quantities (e.g., bus voltage phasors, line parameters, etc.) through a set of available data (e.g., real-time measurements) can often run into observability issues, due to a lack of adequate data [30]. For instance, observe the problem of estimating parameter vector (p) through a WLS problem:

$$p = \min_p \left\{ \sum_{i=1}^M w_i [h_i(p) - y_i]^2 \right\} \quad (2)$$

where:

$$p = [p_1 \ p_2 \ \cdots \ p_N]^T$$

$$y = [y_1 \ y_2 \ \cdots \ y_M]^T$$

$$w = [w_1 \ w_2 \ \cdots \ w_M]^T$$

$$h = [h_1(p) \ h_2(p) \ \cdots \ h_M(p)]^T$$

parameter vector, corresponding to state vector (x) in the SSE model (1);
data vector, corresponding to measurement vector (z) in the SSE model (1);
vector of data weights;
vector of nonlinear functions relating data to the parameters, corresponding to vector $h(x)$ in the SSE model (1).

Conventionally, the observability of the stated problem is analyzed by examining the Jacobian matrix

$$H(p) = \frac{\partial h(p)}{\partial p}, \quad (3)$$

where vectors h and p are as defined in (2).

If this matrix is singular (i.e., has a non-trivial null-space), observability issues exist [29].

Often however this cannot be observed from such a strict definition. For instance, $H(p)$ usually is not exactly singular, but rather near-singular (has singular values close to 0), an event which would still cause problems for system observability [23]. For example, the error in the solution of (2) may vary much more slowly in some parameter vector (p) directions than in others, leading to ill-conditioned, or unstable parameter estimates in these directions.

To address these challenges, this paper takes a new approach to observability analysis based on the MBAM [20].

III. MANIFOLD BOUNDARY APPROXIMATION METHOD (MBAM)

MBAM is built on information geometry where estimating parameters through a nonlinear WLS problem (2) is given a geometric interpretation. The underlying idea is that the set of possible model predictions form a Riemannian manifold (known as the model manifold) embedded in data space with parameter values as coordinates [21]. In this geometric construction, the parameter estimation problem is equivalent to locating the point on the manifold closest to experimental data; note the WLS function (2) is a distance function. Observability issues are related to manifold boundaries; unidentifiable parameters are those that can be taken to extreme values without incurring a statistically significant error.

In addition to relaxing the requirement of identifying the exact null-space of an ill-conditioned matrix, switching from parameter to data space poses several advantages that have been documented elsewhere [20], [21]. Of particular relevance for this study:

- The model manifold abstracts the model from the data to which it is to be fit, enabling the study of the model itself, independent of any specific experimental data. In comparison, the cost surface for ill-conditioned models is very sensitive to experimental data.
- Differential geometry naturally extends local, matrix-based sensitivity analysis to global analysis through geodesics curves.
- Manifold boundaries are natural reduced models with both fewer parameters and lower computational complexity.
- Two types of unidentifiable parameters may be detected [31]:
 1. structurally unidentifiable—cannot be identified from available data. Note that this corresponds to unobservable states in the state estimation problem [8].
 2. practically unidentifiable—can be identified from available data, but with low accuracy.

For a model manifold embedded in data space, observability quantifies the sensitivity of model predictions to variations in the parameter vector (p). As is the case in matrix-based approaches, the least sensitive parameter direction corresponds to the (near) null-space of the Jacobian matrix. However, information geometry extends this local picture to global sensitivity analysis by moving through the parameter space along a *geodesic curve*.

Geodesics are distance minimizing paths, analogous to straight lines generalized to curved surfaces [23]. They are found as the (numerical) solution to a second-order ordinary differential equation in parameter space (while utilizing quantities from the data space) [21]

$$\begin{aligned} \frac{\partial^2 p_i}{\partial \tau^2} &= \sum_{j,k} \Gamma_{i,j,k} \frac{\partial p_j}{\partial \tau} \cdot \frac{\partial p_k}{\partial \tau}; \quad \Gamma_{i,j,k} \\ &= \sum_{\ell,m} \left(\mathfrak{I}(p)^{-1} \right)_{i,\ell} \frac{\partial z_m}{\partial p_\ell} \cdot \frac{\partial^2 z_m}{\partial p_j \partial p_k} \end{aligned} \quad (4)$$

where:

z_i and p_i are as defined in (1) and (2), respectively;

Γ – Christoffel symbols;

τ – arc length of the geodesic curve as measured on the model manifold;

$\mathfrak{I}(p) = H(p)^T H(p)$ – Fisher information matrix.

To detect unidentifiable parameters, geodesics are calculated with the initial direction that of the smallest change in cost, i.e., the eigenvector of $\mathfrak{I}$ with the smallest eigenvalue.

By moving along a geodesic in the least sensitive direction, we construct the *nonlinear* path in parameter space with minimal change in the cost function. This path quantifies which parameters are least constrained by the available data and thus might be unidentifiable. If the geodesic extends parameters to extreme values (e.g., $\pm\infty$) with finite cost, then those parameters will be unidentifiable at some level of statistical significance [23].

Model reduction is then performed by formally taking the limit of the extreme parameter values in the original model [20]. These limiting values fundamentally simplify the computational structure of the model by removing unidentifiable elements, as opposed to simplify fixing them to predetermined values. In other words, we approximate the high-dimensional model manifold by its boundary. Evaluating the limit of extreme parameter values in the model removes the reference to the unidentifiable parameters, but also simplifies the computational structure of the model. Finally, to differentiate between structurally and practically unidentifiable states, we rely on the geodesic arc length. While structurally unidentifiable parameters have a geodesic arc length of zero, practically unidentifiable parameters will have a small, non-zero geodesic arc length. Note that this difference is precisely analogous to whether a matrix has an exact null-space or is just ill-conditioned.

A. Illustrative Example

We demonstrate the procedure on a simple example with two parameters and three data points. The WLS-based cost function is set up as in (2) with $h_i = \exp(-p_1 t_i) + \exp(-p_2 t_i)$ sampled at time points $t = (1/2, 1, 2)$.

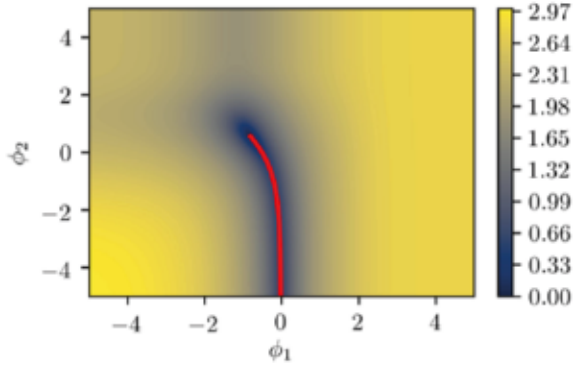
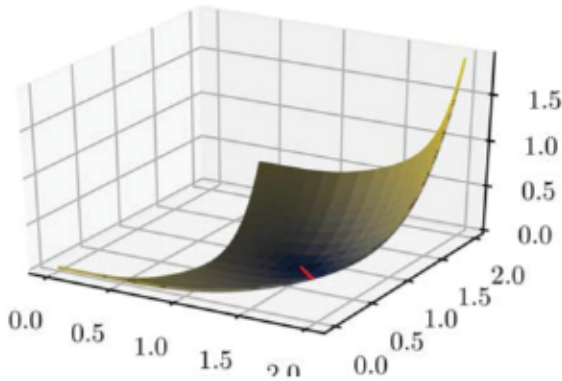
Fig. 1. Contours of $\log(\text{Cost})$ in parameter space.

Fig. 2. Model manifold in data space.

These parameters could represent rates of a generating process with two latent, state-space variables. As such, they should be restricted to non-negative values. Manifold boundaries are a consequence of this physical constraint. There are also boundaries when the exponential function is saturated at infinite parameter values. Finally, this model also exhibits a permutation symmetry; exchanging p_1 and p_2 leaves the predictions unchanged. We, therefore, reparameterize the model using $\varphi_1 = \log \min(p_1, p_2)$ and $\varphi_2 = \log |p_1 - p_2|$. In this way, all manifold boundaries are mapped to infinite values of the φ parameters.

Identifiability issues are first illustrated by considering the contours of the (log) cost surface in parameter space in Fig. 1. Unidentifiable parameters correspond to the long, narrow canyon of low cost that extends from the best fit all the way to $\varphi_2 \rightarrow -\infty$. The geodesic (red) naturally follows this canyon.

A complementary picture is shown in Fig. 2 by considering the manifold of possible predictions in data space. This figure is generated by recognizing that the model is a mapping between a two-dimensional parameter space and a three-dimensional data space. The image of this map for all physically-realizable parameters generates the two-dimensional manifold shown. The images of the cost contours are nearly concentric circles on the model manifold. Contours that extend to infinite parameter values are those that intersect the boundary of the model manifold in data space, making the relationship between parameter identifiability and manifold boundaries explicit. The

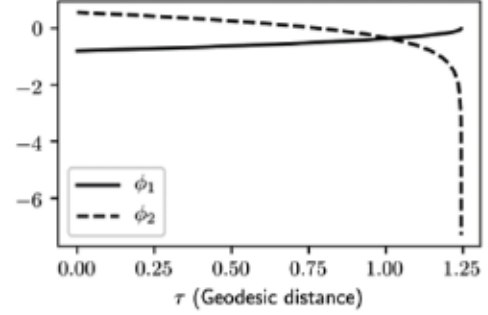


Fig. 3. Parameter values versus geodesic distance.

geodesic curve, infinitely long in parameter space, is the short red curve extending to the nearest boundary on the model manifold.

This two-dimensional example illustrates basic concepts. In realistic models, there are too many dimensions to visualize as we have done here. Instead, we visualize geodesics as in Fig. 3, where we plot the value of each parameter versus the geodesic distance on the model manifold. Parameters that realize extreme values (i.e., become singular) in a finite cost inform us of manifold boundaries.

Finally, reduced models are constructed by analytically evaluating the limit revealed by the geodesic. In this case, we find that $\varphi_2 \rightarrow -\infty$ corresponding to $p_1 \rightarrow p_2$, so that the reduced model becomes $h_i = 2 \exp(-pt)$. This model is both statistically simpler (having only one unknown parameter) as well as computationally simpler (only requiring the evaluation of a single exponential function, or equivalently having a lower-dimensional generating state space).

IV. MBAM-BASED IDENTIFIABILITY ANALYSIS OF THE SSE MODEL

For the specific needs of this paper, an MBAM-based identifiability analysis is adapted to SSE models. The focus is first directed to detecting unidentifiable system states (bus voltage phasors, *Section IV.A*) and then detecting unidentifiable system line parameters (*Section IV.B*). In both cases, general steps for defining the analysis process are (see *Section III.A*):

1. Define the physical constraints for parameters that serve as manifold boundaries needed for MBAM.
2. Apply a change in parameter variables such that the manifold boundaries are mapped to infinite parameter values.
3. Formulate an appropriate cost function in whose parameter space geodesic curves can be calculated.

A. Identifiability Analysis of System States

For this case, parameters correspond to system states (x) [voltage magnitudes (V) and angles (θ)] and data are real-time measurements (z).

For this model, manifold boundaries are a consequence of the physical constraint that voltage magnitudes (V) cannot be negative. Angles (θ), on the other hand, can be both negative and positive; however, it is convenient to impose angular constraints based on the values from the previous time instance—if they were positive (negative) they could not now be negative

(positive). We apply this constraint only for the identifiability analysis, so it does not prevent the estimated states from changing signs at future time instances. Effectively, we check whether the estimated states differ from zero in a statistically significant way and retain only those that do. We enforce these constraints through a change in variables. Let $y_{\text{state}} = [y_V, y_\theta]$ be related to state variables ($x = [V^T \theta^T]^T$)

$$\begin{aligned} V &= +e^{y_V}, \text{ (e.g. } V_1 = +e^{y_{V1}}); \\ \theta &= \pm e^{y_\theta}, \text{ (e.g. } \theta_1 = \pm e^{y_{\theta1}}), \end{aligned} \quad (5)$$

where $y_V = [y_{V1}, y_{V2}, \dots, y_{Vn}]$, $y_\theta = [y_{\theta1}, y_{\theta2}, \dots, y_{\theta n}]$, and n is the number of system buses.

Finally, the cost function can be defined as

$$y_V, y_\theta = \min_{y_V, y_\theta} \left\{ \sum_{i=1}^M w_i [h_i(+e^{y_V}, \pm e^{y_\theta}) - z_i]^2 \right\}, \quad (6)$$

where w_i, h_i, z_i are as defined in (1) and (2).

MBAM can thus be applied to model (6) for detecting unidentifiable states through the following pseudo-algorithm:

- Based on the given system form y_{state} (5) and the initial (empty) set of unidentifiable states x_{unid} .
- Test if any i -th component of y_{state} reach a limiting value ($\pm\infty$) for a finite geodesic arc length τ , (see Section III). If not the procedure is finished, else continue.
- Denote state x_i (5) as unidentifiable and add to set x_{unid} .
- Reduce the model by taking the limiting values for the unidentifiable state.
- Repeat the process on the now reduced model.

Through this procedure, the set of unidentifiable states x_{unid} is derived corresponding to the unobservable system area. A reduced model is simultaneously constructed that removes unidentifiable states.

B. Identifiability Analysis of System Line Parameters

For this case, parameters and data defined for MBAM correspond to system line parameters and real-time measurements (z), respectively. While various line parameters could be chosen, here attention is given to line admittance

$$y = g + jb \quad (7)$$

where g and b are line conductance and susceptance, respectively.

Next, note that admittance values are generally known in real-life systems [32]. From this, the ratio between g and b is known, and the unknown piece is parameterized as

$$y = kb + jb \quad (8)$$

where $k = g/b$.

Manifold boundaries are consequences of the physical constraints that susceptances (b) cannot be positive [31]. Again, we implement these constraints through the change of variables $y_b = [y_{b1}, y_{b2}, \dots, y_{bn}]$, where

$$b = -e^{y_b} \text{ (e.g. } b_1 = -e^{y_{b1}}) \quad (9)$$

Finally, the cost function can be defined as

$$y_b = \min_{y_b} \left\{ \sum_{i=1}^M w_i [h_i^{\text{line}}(-e^{y_b}) - z_i]^2 \right\} \quad (10)$$

where $h_i^{\text{line}}(\cdot)$ is a nonlinear function relating the i -th measurement to the vector (b). Note that for certain measurement i , $h_i^{\text{line}}(\cdot)$ and $h_i(\cdot)$ will be formed the same with the only difference of what is the variable input—(b) or (x).

MBAM can thus be applied to model (10) for detecting unidentifiable line parameters, through the following pseudo-algorithm:

- Based on the given system form y_b (9).
- Test if any i -th component of y_b will reach a limiting value ($\pm\infty$) for a finite geodesic arc length τ , as explained in Section III. If not the procedure is finished, else continue.
- Denote y_{bi} as unidentifiable and reduce the model by taking the appropriate limiting case for the corresponding line parameter. Note two distinct cases:
 1. y_{bi} has reached the boundary value of $-\infty$. From (8) and (9) it can be concluded that b_i , and thus the line admittance, would go to 0. Further, this would mean that the corresponding impedance value goes to $+\infty$, making the corresponding line equivalent to an open circuit [30].
 2. y_{bi} has reached the boundary value of $+\infty$. From (8) and (9) it can be concluded that b_i , and thus the line admittance, would go to $-\infty$. Further, this would mean that the corresponding impedance value goes to 0, making the corresponding line equivalent to a short circuit [30].
- Repeat the process on the now reduced model.

Through this procedure, the physical layer of the system is reduced (see Section II). Finally, note that this system reduction does not require estimating state or line parameter values nor formulating any matrices (see Section II).

With regards to line parameters, our algorithm detects unidentifiable ones based on available measurements and last-known (i.e., previous time instance) state values. For given this data, the cost function (10) can be formulated. After that, MBAM is applied to (10) by utilizing geodesic curves to detect unidentifiable line parameters. In that case, estimating the line parameter values is not required. This can also be observed in the proposed algorithm in Section V.

V. PROPOSED ALGORITHM

The proposed algorithm consists of the following steps:

Step 1: State estimation and ICT outage monitoring

Estimate the state through SSE with a predefined rate, as explained in Section II. While this is done, track ICT component outages by monitoring measurement refresh times (time between two consecutive measurements received) [29]

$$\begin{aligned} \text{No outage : } t_{i,j}^{\text{ref}} &\leq \mu_i + \Delta t_{i,j}^{\text{break}} \\ \text{Outage detected : } t_{i,j}^{\text{ref}} &> \mu_i + \Delta t_{i,j}^{\text{break}} \end{aligned} \quad (11)$$

where:

- i observed ICT component;
- j certain refresh time observation;
- μ_i expected refresh time value;
- Δt^{break} calculated based on the false rejection error possibility (3 % in this paper).

It is important to note that (11) is not restricted to ICT component outages, but can rather detect other sources of missing information. One such being bad data which can cause measurements to be disregarded for the predefined Δt^{break} . The reader is referred to [29] for more details. Once outages occur, go to Step 2.

Step 2: Observability analysis of system states

Once ICT component's outage(s) occur, check for unobservable areas with regards to structurally unidentifiable system states, per Section IV.A:

- if no such area(s) exist, go back to Step 1; or
- if such area(s) exist, split the system into observable and unobservable areas with regards to system states, and continue with the following step.

Step 3: Observability analysis of system line parameters

In areas unobservable with regards to system states, detect unidentifiable line parameters, and re-parametrize corresponding lines as short or open circuit, per Section IV.B.

Step 4: SE and ICT outage monitoring

Detect and remove practically unidentifiable states in the now reduced model, per Section IV.A.

Once the SSE model is reduced, and practically unidentifiable states denoted, go back to Step 1.

Through Steps 2-4, the SSE model is iteratively reduced to only require system information that can be inferred from the available data.

Step 2 and Step 3 detect unobservable states and unidentifiable line parameters, respectively, and reduce the system accordingly. Only after that is the state estimated.

VI. APPLICATION

The proposed algorithm is illustrated on the IEEE 14-bus test system, supplemented with ICT to form a CPS. Physical (electric) layers are modeled in MATLAB (version R2020b), while the cyber (ICT) layers are modeled in NS-2 (version 2.33) [32]. These two layers are then co-simulated using PiccSIM (version 1.15) [33] to formulate a CPS simulation environment. Once this is done, simulations of the proposed algorithm can be executed using the following software: 1) SE and ICT outage monitoring (Step 1) are performed in MATLAB; and 2) observability analysis and system reduction (Steps 2-4) are performed in Julia (version 1.0.5) [34].

Details of the physical and cyber layers are given in [35] and [29], respectively, except for deployed measuring instruments given in Table I. Note that while not all measurement types

TABLE I
AVAILABLE MEASURING INSTRUMENTS

Measurement ID	Measurement type	Instrument type	Sending bus	Receiving bus	Denoted by
1	Pow. injection	RTU	5	/	PQi ₅
2	Pow. injection	RTU	6	/	PQi ₆
3	Pow. injection	RTU	8	/	PQi ₈
4	Pow. injection	RTU	9	/	PQi ₉
5	Pow. injection	RTU	10	/	PQi ₁₀
6	Pow. injection	RTU	11	/	PQi ₁₁
7	Pow. injection	RTU	12	/	PQi ₁₂
8	Pow. injection	RTU	13	/	PQi ₁₃
9	Pow. injection	RTU	14	/	PQi ₁₄
10	Pow. flow	RTU	1	5	PQf ₁₋₅
11	Pow. flow	RTU	2	3	PQf ₂₋₃
12	Pow. flow	RTU	4	3	PQf ₄₋₃
13	Pow. flow	RTU	6	13	PQf ₆₋₁₃
14	Volt. phasor	PMU	1	/	θV ₁
15	Volt. phasor	PMU	2	/	θV ₂
16	Volt. phasor	PMU	3	/	θV ₃
17	Volt. phasor	PMU	3	/	θV ₄

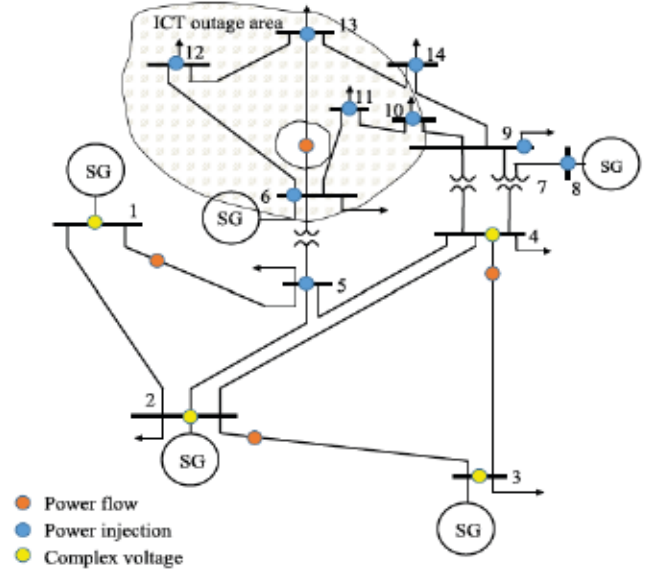


Fig. 4. IEEE 14-bus test system with denoted ICT outages.

(for example, the current measurements) are explored here, the algorithm has no restrictions in this regard.

An example is demonstrated here when significant ICT component outages occur, as demonstrated in Fig. 4. The corresponding area affected by these outages is denoted as "ICT outage area," with the following measurements being affected:

- Bus 6 – measurement # 2, denoted by PQi₆ in Table I.
- Bus 10 – measurement # 5, denoted by PQi₁₀ in Table I.
- Bus 11 – measurement # 6, denoted by PQi₁₁ in Table I.
- Bus 12 – measurement # 7, denoted by PQi₁₂ in Table I.
- Bus 13 – measurement # 8, denoted by PQi₁₃ in Table I.

Finally, note the exception of measurement #13, denoted by PQf₆₋₁₃ in Table I, that is, this measurement remains active after the outages.

The following should be noted about the example:

- Simulations are run for one hour (3600 s), where SSE is executed every 2 s over the observable system area.

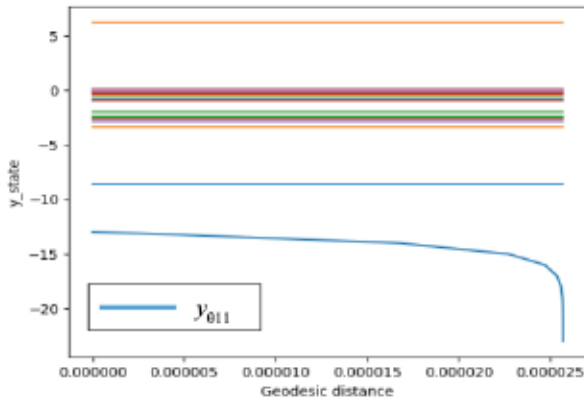


Fig. 5. Geodesics denoting state θ_{11} structurally unidentifiable.

- ICT component outages occur at $t = 1000$ s and remain persistent for the rest of the simulations.
- Measurements are formed as random variables, with power flow solution means and variances as 10^{-3} percentage of their measured values. These values are later denoted in figures as ‘Measured’.
- Measurement weights for SSE are set as the reciprocal value of corresponding measurement variances.
- RTU and PMU sampling rates are 2 s and 0.02 s, respectively. Note that for both measurement types, corresponding measurements are assumed to have been taken at the same moment (same snapshot).
- To quantify the proposed algorithm, results (denoted in tables and figures as ‘SSE’) are compared with measured values.

Utilizing the algorithm provided in Section V, initially the state is estimated through SSE, while ICT outages are monitored. Once outages occur, unobservable areas with regards to structurally unidentifiable system states are detected (Section VI.A), followed by the detection of unidentifiable line parameters (Section VI.B). Finally, the SSE model is reduced and the system state estimation continues (Section VI.C).

A. Detecting Structurally Unidentifiable System States

First, structurally unidentifiable system states are detected, thus formulating system areas unobservable with regards to voltage phasors (Step 2 of the proposed algorithm).

The results of this are shown in Fig. 5. Different lines (different colors) denote all the y_{state} elements, which are defined through (6). Their values are then examined along the geodesic, checking if any will reach a boundary value. It can be observed that $y_{\theta_{11}}$ approaches $-\infty$ at a finite geodesic distance, indicating that the system state θ_{11} is unidentifiable. Note that due to a large number of y_{state} elements, its legend is abbreviated to include only the element of interest— $y_{\theta_{11}}$. The corresponding state is taken to the limit of the extreme parameter value in the original model, in this case, zero. The process is then repeated.

By repeating this procedure, states θ_{12} , V_{11} , θ_{10} , V_{12} , and V_{10} are labeled unidentifiable in the subsequent steps. Thus, the area consisting of buses 10, 11, and 12 is denoted as unobservable

with regards to states. In the model reduction step, these voltages are fixed to zero (buses are grounded), per Section IV.

To validate these results, observability is also analyzed through the method given in [14]:

1. Form the corresponding Jacobian matrix $H(x)$ (Section II.A).
2. Hessian matrix is approximated as $\mathfrak{H}(x) = H(x)^T H(x)$.
3. Eigenvalue decomposition as $\mathfrak{H}(x) = W\Lambda W^T$ is performed, where the following eigenvalues are derived $\text{diag}\{\Lambda\} = [-2.57 \cdot 10^{-13}; -1.33 \cdot 10^{-13}; -9.04 \cdot 10^{-14}; 5.07 \cdot 10^{-16}; 1.73 \cdot 10^{-15}; 1.23 \cdot 10^{-13}; 0.048; 0.057; 0.760; 0.776; 12.195 \dots]$. From this, the first 6 eigenvalues are denoted as close to zero.
4. With regards to the eigenvalues, submatrices W_{22} and W_6 are formed—column partitioning of matrix $W = [W_{22} \ W_6]$.
5. QR decomposition $W_{22}^T P = QR$ is performed, where P is the permutation matrix.
6. P is used to reorder the state vector x as $x = P^T x$.
7. The reordered x is partitioned as $x = [x_{22} \ x_6]$, where x_6 corresponds to the unidentifiable states. For this example, $x_6 = [\theta_{10}, \theta_{11}, \theta_{12}, V_{10}, V_{11}, V_{12}]$. Thus, the same area (buses 10, 11, and 12) is denoted unobservable with regards to states.

B. Detecting Unidentifiable System Line Parameters

Next, unidentifiable line parameters are detected and reduced in the unidentifiable system area defined in Section VI. A (Step 3 of the proposed algorithm).

In the first geodesic, it is observed that y_{b12-13} approaches $-\infty$, thus denoting the admittance of 12-13 unidentifiable. The corresponding line is re-parametrized as an open circuit, y_{b12-13} is fixed to zero, and the process is repeated until all unidentifiable line parameters are removed.

C. Reduced System and State Estimation Results

The fully reduced system is shown in Fig. 6:

- ICT components affected by outages are disregarded.
- Unobservable buses are identified (Section VI.A).
- Lines corresponding to unidentifiable line parameters are re-parametrized (Section VI.B).

Finally, for the given example, illustrative results for the entire simulation period of one hour (3600 s) are given as:

- Average errors for voltage magnitudes and angles calculated by SSE are shown in Table II for the following bus groups:
 - Group # 1—unobservable buses (10, 11, 12);
 - Group # 2—buses neighboring unobservable ones (6, 9, 13);
 - Group # 3—distant buses measured by RTUs (5, 7, 8, 14);
 - Group # 4—distant buses measured by PMUs (1, 2, 3, 4).
- Values are given for the following time instances: 1) before ICT outages (Pre-outage); 2) after ICT outages (Post-outage) and 3) entire time interval.

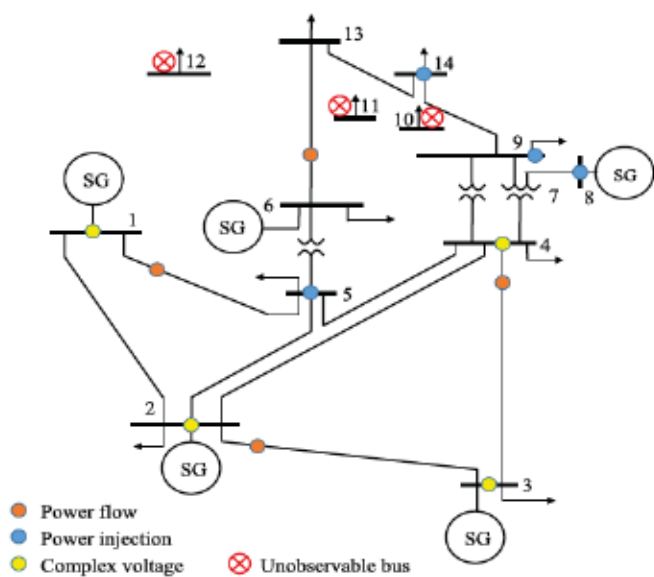


Fig. 6. IEEE 14-bus test system reduced through MBAM.

TABLE II
AVERAGE VOLTAGE ESTIMATION ERRORS ($\cdot 10^{-3}$)

Time interval		Pre-outage	Post-outage	Entire interval
Group # 1	V [p.u.]	2.6	7.1	5.9
	θ [rad]	3.5	5.0	4.5
Group # 2	V [p.u.]	2.8	5.5	4.4
	θ [rad]	3.0	4.0	3.7
Group # 3	V [p.u.]	3.3	3.7	3.5
	θ [rad]	3.6	3.7	3.7
Group # 4	V [p.u.]	0.7	0.8	0.8
	θ [rad]	0.8	0.9	0.9

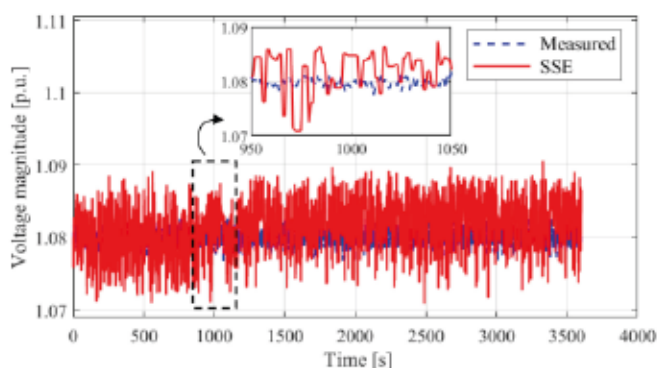


Fig. 7. Voltage magnitude at bus 8.

- Estimated and measured voltage magnitude for bus 8, with the emphasized time interval 950-1050 s, is given in Fig. 7. Note that this bus is from group # 3.
- Estimated and measured voltage magnitude and angle for bus 12, with the emphasized time interval 950-1050 s, are given in Fig. 8. Note that this bus is from group # 1.

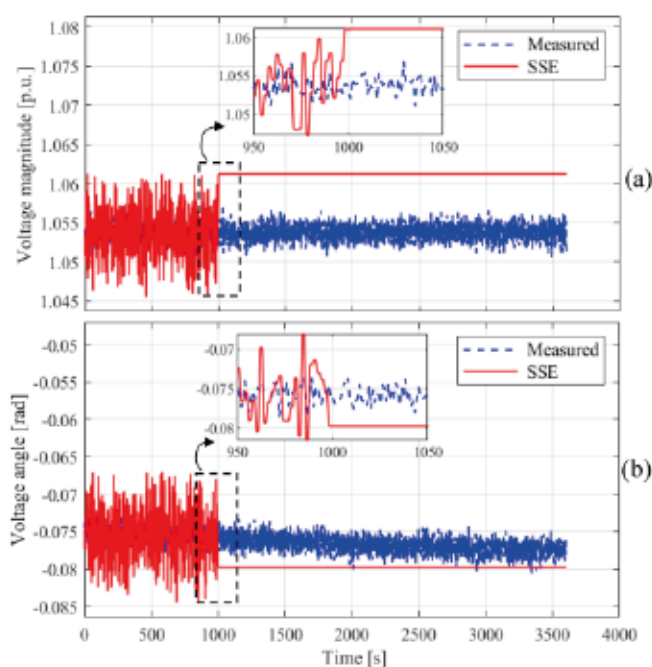


Fig. 8. Voltage (a) magnitude and (b) angle at bus 12.

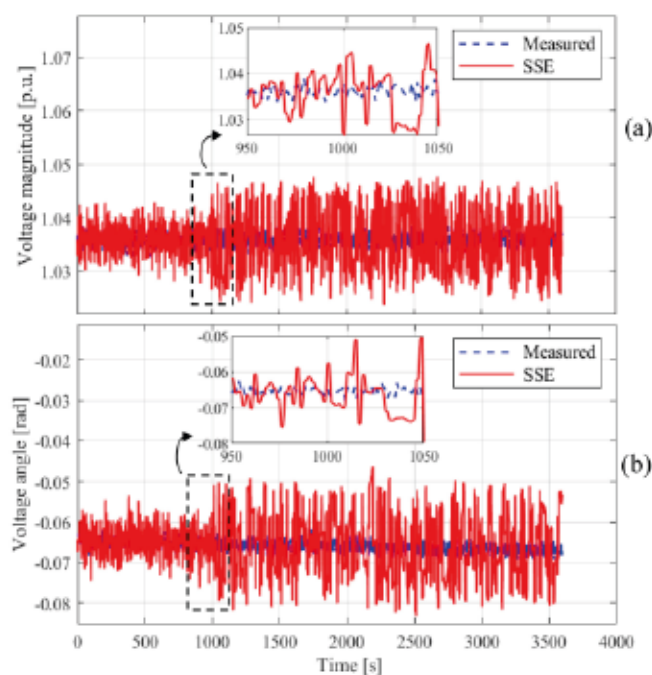


Fig. 9. Voltage (a) magnitude and (b) angle at bus 6.

- Estimated and measured voltage magnitude and angle for bus 6 within the time interval 950–1050 s, are given in Fig. 9. Note that this bus is from group # 2.
- From these results, the following can be concluded:
- Overall there is good state tracking for all buses (Table II)
 - Insignificant result degradation is observed in buses distant from ICT component outages (see Fig. 7 and Table II).
 - For buses that are unobservable after the ICT outage, voltage values can no longer be estimated and their values are

From these results, the following can be concluded:

- Overall there is good state tracking for all buses (Table II)
- Insignificant result degradation is observed in buses distant from ICT component outages (see Fig. 7 and Table II).
- For buses that are unobservable after the ICT outage, voltage values can no longer be estimated and their values are

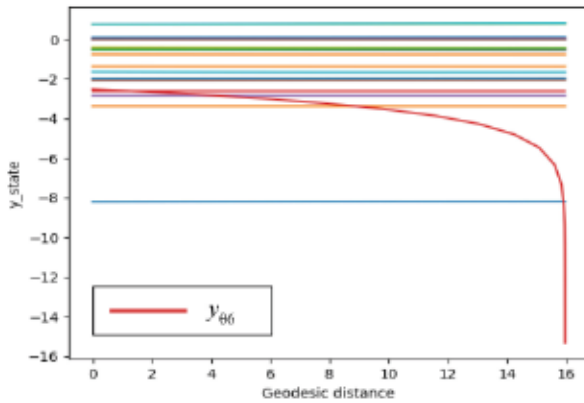


Fig. 10. Geodesics denoting state θ_6 practically unidentifiable.

kept as the last estimated ones (see Fig. 8 and Table II). Note that the decision on what to do with these bus voltages is left to the user—this is one way to deal with them, acceptable here since only slow state changes are observed.

- Result degradation is also observed in buses neighboring to unobservable ones once ICT outages occur. Nevertheless, state tracking for these buses is still acceptable (see Fig. 9 and Table II).

Finally, as explained in Section III, practically unidentifiable states (identifiable with low accuracy) may also be detected by MBAM on the now reduced model. The results of this are shown in Fig. 10, similar to Fig. 5—values of y_{state} elements (denoted by different colors) are analyzed along the geodesic. We find the geodesic takes y_{θ_6} to $-\infty$ in a finite, but the larger distance on the model manifold. We thus denote the system state θ_6 as practically unidentifiable. The model can be further reduced by fixing element y_{θ_6} to zero.

The process is repeated and states θ_{13} , V_{11} , θ_{10} , V_{12} , and V_{10} are also labeled practically unidentifiable. Notice that the practically unidentifiable area consists of buses 6, 9, and 13, i.e., neighboring the structurally unobservable buses. Note that even though the corresponding states may still be estimated (see Table II) their inferred values may have high variance should be used with caution. Depending on the resolution of the actual data, the proposed algorithm will construct an appropriate reduced-order model.

VII. CONCLUSION

This paper proposes a novel method of system observability analysis, which is further capable of reducing the model to remove reference to the unidentifiable state variables. The procedure is motivated by scenarios when considerable ICT component outages occur, which we demonstrate on the IEEE-14 bus test system, supplemented by the corresponding ICT to create a CPS-based simulation environment.

The basis for our method is the MBAM, which analyzes state estimation models as manifolds of potential predictions embedded in data space. Unidentifiable states and parameters correspond to manifold boundaries. The information geometry approach extends local, matrix analysis to global sensitivity

analysis through the calculation of geodesic paths. As such, the method is additionally able to detect practically unidentifiable states and parameters, i.e., those that can be estimated in principle but may have low accuracy or be ill-conditioned. Recent advances in large-scale MBAM give us confidence that the method is relevant for the state estimation in large power systems with thousands of nodes.

Once ICT outages occur, unidentifiable states and parameters are detected, and the state estimation model is reduced. This reduction does more than just fix unidentifiable states and parameters to last-known values, as is common practice in other methods. Rather, by taking limiting cases, we fundamentally change the computational structure of the model so that there is no reference to unidentifiable elements. In this way, the computational complexity of the model is commensurate with the information content of the data, and we extract the most information possible.

REFERENCES

- [1] M. H. Cintuglu, O. A. Mohammed, K. Akkaya, and A. S. Uluagac, "A survey on smart grid cyber-physical system testbeds," *IEEE Commun. Surv. Tut.*, vol. 19, no. 1, pp. 446–464, First quarter 2017.
- [2] H. Lei, and C. Singh, "Developing a benchmark test system for electric power grid cyber-physical reliability studies," *Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Beijing, 2016, pp. 1–5.
- [3] B. Falahati, Y. Fu, and L. Wu, "Reliability assessment of smart grid considering direct cyber-power interdependencies," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1515–1524, Sep. 2012.
- [4] Z. Zhang, W. An, and F. Shao, "Cascading failures on reliability in cyber-physical system," *IEEE Trans. Rel.*, vol. 65, no. 4, pp. 1745–1754, Dec. 2016.
- [5] H. Lei, and C. Singh, "Non-sequential monte carlo simulation for cyber-induced dependent failures in composite power system reliability evaluation," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1064–1072, Mar. 2017.
- [6] K. Marashi, S. S. Sarvestani, and A. R. Hurson, "Consideration of cyber-physical interdependencies in reliability modeling of smart grids," *IEEE Trans. Sustain. Comput.*, vol. 3, no. 2, pp. 73–83, Apr.–Jun. 2018.
- [7] J. Zhao et al., "Roles of dynamic state estimation in power system modeling, monitoring and operation," *IEEE Trans. Power Syst.*, to be published. [Online]. Available: <https://arxiv.org/abs/2005.05380>
- [8] A. Monticelli, and F. F. Wu, "Network observability: Theory," *IEEE Power Engin. Rev.*, vol. PER-5, no. 5, pp. 32–33, May 1985.
- [9] B. Gou and A. Abur, "A direct numerical method for observability analysis," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 625–630, May 2000.
- [10] B. Gou, "Jacobian matrix-based observability analysis for state estimation," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 348–356, Feb. 2006.
- [11] R. E. Pruneda, C. Solares, A. J. Conejo, and E. Castillo, "An efficient algebraic approach to observability analysis in state estimation," *Electric Power System Res.*, vol. 80, no. 3, pp. 277–286, Mar. 2010.
- [12] N. M. Manousakis and G. N. Korres, "Observability analysis and restoration for systems with conventional and phasor measurements," *Int. Trans. Elect. Energy Syst.*, vol. 23, no. 8, pp. 1548–1566, Nov. 2013.
- [13] A. Simões Costa, A. Albuquerque, and D. Bez, "An estimation fusion method for including phasor measurements into power system real-time modeling," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1910–1920, May 2013.
- [14] 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.
- [15] E. Castillo, A. J. Conejo, R. E. Pruneda, and C. Solares, "State estimation observability based on the null space of the measurement jacobian matrix," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1656–1658, Aug. 2005.
- [16] E. Castillo, A. J. Conejo, R. E. Pruneda, and C. Solares, "Observability analysis in state estimation: A unified numerical approach," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 877–886, May 2006.
- [17] M. C. de Almeida, E. N. Asada, and A. V. Garcia, "Power system observability analysis based on gram matrix and minimum norm solution," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1611–1618, Nov. 2008.

- [18] J. B. A. London, S. A. R. Piereti, R. A. S. Benedito, and N. G. Bretas, "Redundancy and observability analysis of conventional and PMU measurements," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1629–1630, Aug. 2009.
- [19] H. Herrero and C. Solares, "A greedy algorithm for observability analysis," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1638–1641, Mar. 2020.
- [20] M. K. Transtrum and P. Qui, "Model reduction by manifold boundaries," *Phys. Rev. Lett.*, vol. 113, no. 9, pp. 098701–1–098701–6, Aug. 2014.
- [21] M. K. Transtrum, A. T. Sarić, and A. M. Stanković, "Information geometry approach to verification of dynamic models in power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 440–450, Jan. 2018.
- [22] H. H. Mattingly, M. K. Transtrum, M. C. Abbott, and B. B. Machta, "Maximizing the information learned from finite data selects a simple model," *Proc. Nat. Acad. Sci.*, vol. 115, no. 8, pp. 1760–1765, Feb. 2018.
- [23] M. K. Transtrum, A. T. Sarić, and A. M. Stanković, "Measurement-directed reduction of dynamic models in power systems," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2243–2253, May 2017.
- [24] B. L. Francis, J. R. Nuttall, M. K. Transtrum, A. T. Sarić, and A. M. Stanković, "Network reduction in transient stability models using partial response matching," *North Amer. Power Symp. (NAPS)*, Wichita, KS, USA, 2019.
- [25] P. Ren, H. Lev-Ari, and A. Abur, "Tracking three-phase untransposed transmission line parameters using synchronized measurements," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4155–4163, Jul. 2018.
- [26] Y. Lin and A. Abur, "Robust state estimation against measurement and network parameter errors," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4751–4759, Sep. 2018.
- [27] V. G. Švenda, A. M. Stanković, A. T. Sarić, and M. K. Transtrum, "Flexible hybrid state estimation for power systems with communication irregularities," *IET Gen., Transm. Distr.*, vol. 14, no. 11, pp. 2111–2119, Jun. 2020.
- [28] A. Abur and A. G. Exposito, *Power System State Estimation – Theory and Implementation*, 1st ed., CRC Press, Boca Raton, FL, pp. 9–22, 2004.
- [29] V. G. Švenda, A. M. Stanković, A. T. Sarić, and M. K. Transtrum, "Probabilistic extension of flexible hybrid state estimation for cyber-physical systems," *Int. J. Elect. Power Energy Syst. (IIEPES)*, vol. 122, pp. 1–23, Nov. 2020.
- [30] J. J. Grainger and W. D. Stevenson, *Power System Analysis*, McGraw-Hill Education, 1994.
- [31] A. Raue *et al.*, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, Jun. 2009.
- [32] K. Fall and K. Varadhan, "The *ns* manual (formerly *ns* notes and documentation)," Nov. 2011. [Online]. Available: https://www.isi.edu/nsnam/ns/doc/ns_doc.pdf
- [33] M. Björkborn, S. Nethi, and T. Kohtamäki, "PiccSIM manual," Aalto Univ., Dec. 2013. [Online]. Available: http://wsn.aalto.fi/en/tools/piccsim/piccsim_manual_1.1.pdf
- [34] *The Julia Language, The Julia Project*, USA, Nov. 2020. [Online]. Available: <https://docs.julialang.org/en/v1/>
- [35] R. Christie, *14 Bus Power Flow Test Case*, Univ. of Washington, Seattle, WA, USA, 1993. [Online]. Available: https://labs.ece.uw.edu/pstca/pf14/pg_tca14bus.htm