Identifying Contingencies in Power Systems using Time-Domain Dynamics

Eric Hans Lee and David Bindel

Abstract—In this paper, we introduce a framework for identifying contingencies in Power Systems using modal information from dynamic data. Assuming there are high-fidelity sensors such as PMUs placed throughout the system, not necessarily on every bus, providing voltage and angle readings, our framework is able to accurately and quickly identify the contingency in question. Our framework uses the differential-algebraic representation of a power system to match modal information to model data. In this light, we formulate the contingency identification problem as calculation of a set of loss functions fitting modal information against an a-priori dictionary of Power System models. Unlike most other methods in the literature, ours takes into account the full dynamic behavior of time-domain sensor data. We present promising experimental results on the IEEE 39, 57, and 145 bus systems.

Index Terms—Contingency Identification, Topology Estimation, Power Systems

I. INTRODUCTION

Real-time identification of contingencies, or failure elements, in a power system is an important network monitoring function and smart grid application. Failure to properly detect contingencies in a power system may lead to incorrect topology estimates or dangerous control actions. Due to sampling rates far too low to recover any nontrivial frequency-domain information, supervisory control and data acquisition (SCADA) systems traditionally used to observe power systems are mainly used for steady-state analysis. Phasor Measurement Units (PMUs) are a newer sensor providing higher resolution time-domain data amenable to frequency-domain analysis. PMUs measure the angles and voltages of the bus that they are located on; there also exists a body of work on detecting angles and voltages of PMUs on neighboring buses.

Current PMU deployments do not provide complete observability and are often sparsely spread throughout the grid. In this light, developing fast and robust methods using PMU data to perform real-time contingency analysis is an important research problem.

II. PRIOR AND RELATED WORK

Identifying topology changes and contingencies using PMU data in a power system and have been looked at widely in the power systems community. In [11], the authors use changes in phase angle to diagnose line failures, and they demonstrate success both on a 37 bus model as well as a 7000 bus real world example. In both [4] and [12], the problem of line failures is first reformulated as an additional set of power injections rather than changes to the system topology itself. This problem of identifying pairs of injections is solved as a mixed-integer programming problem. This formulation is elegant as it does not require the explicit enumeration of all possible contingencies. In [10], the authors take a different approach and perform system identification/state estimation to diagnose a line failure. Steady state is assumed, but this method does not seem scalable and as the authors note, is not as robust to noise as other methods. In [9], the authors (one of whom is an author on this paper) adopt a filtering procedure to quickly narrow down candidate contingencies, including line failures, load trips, and substation reconfigurations. They test on the 56 Bus, 118 Bus, and Polish Systems.

All bodies of work cited previously ignore transient oscillation in favor of pure steady-state information. While the steady state is convenient to use, the dynamics of the system also contain valuable information about the system itself. Considering dynamic data allows one to better leverage information provided by PMUs. Indeed, the steady state of a system is not unique; there exist very simple degenerate examples where the steady state is preserved despite changes in the model.

Analyzing transient dynamics of a system is popular in a wide variety of engineering applications, including vibration-based condition monitoring, model validation, and related areas. Using modal data in particular is popular in civil engineering, where a number of different metrics comparing modal data to analytical models have been developed, see [1], [7]. Our framework generally falls into this set of approaches.

III. MODELING POWER SYSTEMS

A. Variables

These variables and associated dimension sizes will be used consistently for the rest of the paper.

\[
\begin{align*}
    x &\in \mathbb{R}^m = \text{algebraic variables} \\
    y &\in \mathbb{R}^n = \text{differential variables} \\
    z &\in \mathbb{R}^l = \text{aggregated algebraic and differential variables} \\
    f &: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^m = \text{differential equations} \\
    g &: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^m = \text{algebraic equations} \\
    J &\in \mathbb{R}^{l \times l} = \text{DAE Jacobian Matrix} \\
    E &\in \mathbb{R}^{l \times l} = \text{DAE Isolation Matrix} \\
    H &\in \mathbb{R}^{k \times l} = \text{PMU Matrix} \\
    (\lambda_i, v_i) &\text{-i-th Eigenpair solving } J v_i = \lambda_i E v_i \\
    h(t) &\in \mathbb{R}^k = \text{PMU time domain data}
\end{align*}
\]

B. Power System Dynamics

A power system is represented by a system of differential algebraic equations (DAE)

\[
\begin{align*}
    \Omega &= \left\{ 
    \begin{array}{ll}
    x' &= f(x, y) \\
    0 &= g(x, y)
    \end{array}
    \right. \\
    (1)
\end{align*}
\]
\( f \) and \( g \) are differential and algebraic portions of the DAE. \( x \) and \( y \) are the associated algebraic and differential variables, respectively. We also have initial conditions \( x(t_0) = x_0 \) and \( y(t_0) = y_0 \). General DAE theory states that close to steady state, this nonlinear DAE is closely approximated by the linear DAE [8]

\[
Ez' = Jz
\]

(2)

Where \( E, J \), and \( z \) are defined as

\[
z = \begin{bmatrix} x \\ y \end{bmatrix}, \quad J = \begin{bmatrix} F_x & F_y \\ G_x & G_y \end{bmatrix}, \quad E = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}
\]

(3)

\( F_x \) is the partial derivatives of \( f \) in terms of \( x \), \( F_y \) is the partial derivatives of \( f \) in terms of \( y \), \( G_x \) is the partial derivatives of \( g \) in terms of \( x \), \( G_y \) is the partial derivatives of \( g \) in terms of \( y \). Hence, \( J \) is the Jacobian of the DAE System. We call \( E \) the Isolation Matrix as it isolates the differential variables.

C. Analytic Solutions to DAEs

There are a few different ways of characterizing (equivalent) analytical solutions to equation (2) according to [5]. The analytical formula we use has the form

\[
z(t) = \sum_{i=1}^{k} c_i e^{\lambda_i t} v_i = \sum_{i=1}^{k} e^{\lambda_i t} d_i
\]

(4)

Where \( c_i \) are some constants determined by the initial condition and \((\lambda_i, v_i)\) are the non-infinite eigenpairs of the generalized eigenvalue problem

\[
Jv_i = \lambda_i Ev_i
\]

(5)

To make the eigenvalue problem well-posed, \( v_i \) are restricted to have unit length. We can fold in our constants into our eigenvectors i.e. rewrite our solution into a slightly more compact form with \( d_i = c_i v_i \). Assuming our system is stable, the eigenvalues \( \lambda_i \) will have real part less than 0 i.e. \( re(\lambda_i) \leq 0 \).

This form of the analytic solution says the solution to any DAE can be decomposed into terms \( e^{\lambda_i t} d_i \). Because our system is stable, each \( e^{\lambda_i t} d_i \) represents a damped sinusoid with mode \( d_i \) and frequency \( \lambda_i \). One can view \( d_i \) and frequency \( \lambda_i \) as particular “identifiers” or “features” of our system determined by \( J \).

D. PMUs

Phasor Measurements read angles and voltage magnitudes of the buses that they are placed on and output the total signal \( h(t) \in \mathbb{R}^k \). Assuming that the angles and voltages of buses are modeled as algebraic variables, \( h(t) \) can be modeled mathematically as

\[
h(t) = Hz(t) = \sum_{i=1}^{k} e^{\lambda_i t} H d_i
\]

(6)

Note that \( H \), known as the sensing matrix, simply picks out a subset of \( z(t) \), since angle and voltage are a subset of the algebraic variables \( x \). Because \( e^{\lambda_i t} \) is simply a scalar, we can push \( H \) onto \( d_i \), \( H \) will consequently pick out the entries of each mode \( d_i \) associated with the indices of angles and voltages read by PMUs on the system. We denote the partial modes of the system induced by PMU configuration \( H \) as \( u_i = Hv_i \), and rewrite equation (6) as

\[
h(t) = Hz(t) = \sum_{i=1}^{k} e^{\lambda_i t} u_i
\]

(7)

This particular analytic solution says that PMU readings can be decomposed into terms \( e^{\lambda_i t} u_i \). We call the pair \((\lambda_i, u_i)\) a partial eigenpair, which may be viewed as a particular “features” of our system determined by \( J \) and PMU configuration \( H \).

E. Steady State

The steady state of a system corresponds to an eigenpair of equation (5) with \( \lambda_0 = 0 \). To be more precise, the steady state is characterized by the solution to

\[
Jv_0 = 0
\]

Which is the solution to the standard linear power flow problem. Assuming a single zero eigenpair \((\lambda_0, v_0)\), plugging in \( \lambda_0 = 0 \) into equation (4) yields

\[
e^{\lambda_0 t} v_0 = v_0
\]

That is to say, the eigenpair of value zero shifts the analytical solution by a constant factor. This constant factor shift is in fact the to steady-state value, see [3]. This characterization of a power system’s steady state motivates a simple degenerate example in which the steady state of a power system is unchanged despite a perturbation \( \Delta J \) to the model. If \( \Delta J \) orthogonal to \( v_0 \) i.e. \( \Delta J v_0 = 0 \), then

\[
(\Delta J + \Delta J) v_0 = 0
\]

(8)

so that the steady state spanned by the vector \( v_0 \) is unchanged despite a change \( \Delta J \) not necessarily small. However, \( \Delta J \) will invariable change some other nonzero eigenvectors of (5). Nonzero eigenvectors are associated with dynamic behavior rather than steady state behavior. In this light, while diagnose this particular change in the model using steady state information is consequently impossible, it is possible using dynamic information. As an addendum, practical-minded readers may find this simple example unsatisfing, as the chances of a model mismatch being nearly orthogonal to the steady state is very low. However, this example illustrates a straightforward edge case of a more general class of eigenvector sensitivity analyses —the chances of the steady-state being less sensitive to changes in the model than dynamic data is not at all unlikely.

IV. OUR WORK

As mentioned before, the development of this paper’s work was motivated by the lack of power system literature revolving around dynamic data. From a numerics point of view, we
wanted our method to be simple and fast to compute. This meant working with mainly with the Jacobian rather than the full nonlinear model.

Reiterating the key point of sections III-C and III-D, we assume the output $h(t)$ of a PMU sensor network to consist of a sum of damped sinusoids

$$h(t) = \sum_{i=1}^{k} e^{\lambda_i t} u_i$$

Viewing the partial eigenpairs

$$\{ (\lambda_i, u_i) \}_{i=1}^{p}$$

as "features" of the time-domain output, we have developed a simple measure of similarity between these features and a given analytical model of a power system, which we call the I-score. A small I-score indicates high similarity while a large I-score indicates low similarity. That is to say, a model with a smaller I-score is more likely to have generated $h(t)$ than a model with a larger I-score. Possessing an a-priori set of candidate models for any particular contingency, one may then use the I-score to isolate the contingency model in question by calculating the I-score of all candidate models and finding the minimum.

To be slightly more specific, the I-score matches the partial eigenpairs in $h(t)$ to an analytical model by aggregating a set of loss functions. Each loss function is a simple least-squares calculation, which is simple to set up and quick to solve. However, before we get into the details, we first take a slight detour by covering important material regarding the fitting of damped sinusoids to a time-domain signal. This is a necessary step when extracting $u_i$ and $\lambda_i$ from $h(t)$, and thus useful to cover for readers looking to implement our work for themselves.

V. FITTING DAMPED SINUSOIDS

The problem of fitting a set of sinusoids to a multivariate signal $h(t)$ is a broad problem in its own right. Methods for doing so usually take as input a time-domain signal $h(t)$ and some set of metadata (such as the size of the model to fit, covariance of error present, etc), and output a set of partial eigenpairs $\{ (\lambda_i, u_i) \}_{i=1}^{p}$. The partial modes $u_i$ will have some norm $w_i = \| u_i \|_2$ depending upon the scaling and base units of the model in question. We detail a few methods for fitting damped sinusoids below —they are split into two general classes.

A. Matrix Pencil Methods

Matrix Pencil methods generally operate by reducing the calculating of partial eigenpairs to a generalized eigenvalue problem. The most basic Matrix Pencil Method for a Single Input, Single Output (SISO) system is Prony’s method. Most of the following methods are some generalization of Prony’s Method

- Generalized Robust Prony
- Filter Diagonalization

• ESPIRIT

Empirically, we found that Prony yielded oftentimes inaccurate results in our experiments. We did not look more closely at the other methods listed.

B. Subspace and Autoregression Methods

These methods are more expensive but more accurate than Matrix Pencil methods, and generally don’t admit one overarching formulation.

- Dynamic Mode Decomposition
- System Identification (N4SID, SSARX, MOESP, CV, etc)

Dynamic Mode Decomposition as applied to PMU data has been looked at by [2] and [13]. Empirically, we found System Identification to yield the best results in our experiments, with Dynamic Mode Decomposition also doing quite well.

VI. FORMAL PROBLEM FORMULATION

We assume we have some correct prior state estimate, an accurate power system model, and a set of PMUs spread out across the system. We start with a common set-up in inverse problems; assume we have a finite dictionary of linearized power system models

$$\mathcal{D} = \{ \Omega_1, \Omega_2, \ldots, \Omega_q \}$$

and a signal $h(t)$ containing transient dynamics originating from one particular model in $\mathcal{D}$. We want to determine which model $\Omega_\ell$ “best explains” the signal $h(t)$. That is to say, we want to diagnose some $\Omega_\ell$ most likely outputting signal $h(t)$, for some mathematically rigorous definition of “most likely”. We want our method to be fast and accurate; the naive solution of simulating explicit dynamics for each model in $\mathcal{D}$ and matching simulated dynamics to $h(t)$ is both slow and highly sensitive to error.

Our method is the following: we extract a set of partial eigenpairs $\{ (\lambda_i, u_i) \}_{i=1}^{p}$ from $h(t)$ and fit them against each model in $\mathcal{D}$ by calculating some loss function. The model yielding the smallest set of loss function values is the model we diagnose as the correct one.

VII. LOSS FUNCTION

The Loss Function is a measure of how close a model $\Omega_j \in \mathcal{D}$ is to a partial eigenpair $(\lambda_i, u_i)$. Let $\hat{u}_i = \frac{u_i}{\| u_i \|_2}$ i.e. $u_i$ rescaled to have unit norm. One particular family of loss functions seen commonly in Civil and Mechanical Engineering, derived from a correlation score called the Modal Assurance Criterion (MAC) [1]. Our method is a variation of the MAC, albeit heavily modified to support our needs. Given a partial eigenpair $(\lambda_i, u_i)$ our loss function is defined as

$$\mathcal{L}(\lambda_i, u_i, \Omega_j) = \min_{\hat{v}} \| H \hat{v} - \hat{u}_i \|_2^2 + \| (J - \lambda_i E) \hat{v} \|_2^2 \quad (9)$$
\( \mathcal{L} \) attempts to fit the entire eigenvector \( \hat{\mathbf{v}} \) from the partial eigenvector \( \mathbf{u}_i \), and takes the loss to be the residual of the fitting. The first summand forces \( H \hat{\mathbf{v}} \) to approximately equal the partial eigenvector \( \mathbf{u}_i \), and the second summand forces \( \hat{\mathbf{v}} \) to satisfy the generalized eigenvalue problem \( (J_j - \lambda_i E) \hat{\mathbf{v}} = 0 \). More intuitively, the first summand fits to the signal and the second summand fits to the model. We can rewrite as our loss function as in a slightly more compact least-squares form

\[
\mathcal{L}(\lambda_i, u_i, \Omega_j) = \min_{\hat{\mathbf{v}}} \left\| \begin{pmatrix} H & J_j - \lambda_i E \end{pmatrix} \mathbf{v} - \begin{pmatrix} \hat{u}_i \\ 0 \end{pmatrix} \right\|_2^2 \tag{10}
\]

Because \( J, E, \) and \( H \) are all sparse, we may solve (10) with a fast solver such as LSQR [6].

\[ A. \text{ Bounds} \]

One problem with many loss functions is scale; what precise range of numbers constitute a good result and what precise range of numbers constitute a poor result? Luckily, \( \mathcal{L} \) is bounded from below by 0 and above by 1, giving one an intuitive idea of “good” and “poor”.

The lower bound is straightforward; we are minimizing a nonnegative objective function. In particular, the lower bound is achieved if \( \hat{u}_i = H \hat{v}_i \) where \( v_i \) is the generalized eigenvector associated with eigenvalue \( \lambda_i \) of \( J_j \), by noting that setting \( \hat{v} = v_i \) yields \( \mathcal{L}(\lambda_i, u_i, \Omega_j) = 0 \). That is to say, the lower bound is achieved if the empirical mode exactly matches an eigenvector of the analytical model. The upper bound is obtained by plugging in \( \hat{v} = 0 \), yielding the equation

\[
\mathcal{L}(\lambda_i, u_i, \Omega_j) = \min_{\hat{\mathbf{v}}} \left\| \hat{u}_i \right\|_2^2
\]

and because \( \hat{u}_i \) was assumed to have unit norm,

\[
\mathcal{L}(\lambda_i, u_i, \Omega_j) = \min_{\hat{\mathbf{v}}} 1 = 1
\]

A loss function close to 0 indicates that the partial eigenpair \( (\lambda_i, u_i) \) is well match with a model \( \Omega_j \). A loss function close to 1 indicates that a partial eigenpair \( (\lambda_i, u_i) \) does not match to a model \( \Omega_j \).

\[ \text{VIII. AGGREGATING LOSS FUNCTIONS} \]

Given some set of partial eigenpairs \( \{ (\lambda_i, u_i) \}_{i=1}^p \), and for each \( \Omega_j \in \mathcal{D} \), we calculate the loss \( L_{ij} = \mathcal{L}(\lambda_i, u_i, \Omega_j) \). We need a way of boiling down information stored in \( L_{ij} \) \( \forall i, j \) to a correct contingency diagnosis. We suggest one reasonable way; the I-score \( \Omega_j \) of contingency model \( j \) is defined as:

\[
\begin{align*}
L_{ij} &= \mathcal{L}(\lambda_i, u_i, \Omega_j) \\
 w_i &= \left\| u_i \right\|_2 \\
 I_j &= \sum_{i=1}^p w_i L_{ij}
\end{align*}
\tag{11}
\]

The I-score of a model \( \Omega_j \) is simply a weighted linear combination of its loss values for each partial eigenpair, where each weight \( w_i \) is the magnitude of \( u_i \). Naively speaking, one might set \( w_i = 1 \) to equally weigh all loss function calculations for each contingency model. However, this does not account for spurious eigenpairs caused by nonlinearities, small disturbances, or noise. As we expect spurious eigenvectors to be rather small in amplitude when extracted compared to real eigenvectors, we thus set the weights to simply be the norm of the associated partial eigenvector \( u_i \).

The I-score may be interpreted as a correlation measure between empirical data and an analytical model. A small I-score implies that each of its summands is likewise small, meaning that each feature of the empirical data is well correlated with the analytical model. Likewise, a large I-score implies one of its summands is large, meaning there exists some feature of the empirical data not correlated with the analytical model.

\[ \text{IX. FORMAL PROCEDURE} \]

Our formal method is quite straightforward and we detail it below.

\[ \text{Algorithm 1: Contingency Identification} \]

1. **Input**: dictionary \( \mathcal{D} \), signal \( \hat{h}(t) \)
2. **Output**: Contingency \( c \)
3. \( \{ (\lambda_i, u_i) \}_{i=0}^k \leftarrow N4SID(\hat{h}(t)) \)
4. for \( \Omega_j \in \mathcal{D} \) do
   5. \( I_j \leftarrow 0 \)
   6. for \( i \in 0 \ldots k \) do
      7. \( w_i \leftarrow \left\| u_i \right\|_2 \)
      8. \( I_j \leftarrow I_j + w_i L(\lambda_i, u_i, \Omega_j) \)
   9. end
10. end
11. \( c \leftarrow \arg \min I_j \)

Note that, because the steady state is also parameterized by an eigenvector of the Jacobian, our method uses strictly more information than any steady state analysis.

\[ \text{X. EXPERIMENTAL VALIDATION} \]

We attempted to diagnose a set of power system contingencies in the form of downed lines and tripped generators. The experiments were set up as follows. Using the IEEE 39, 57, and 145 bus systems, we took the set of contingency models (i.e. elements in the dictionary) to be all possible line failures and generator trips. Contingencies were simulated using the Power System Analysis Tool (PSAT). We ran tests on three different PMU arrangements for each test system.

1) **Single**: A single PMU is placed near the center of the network (topologically speaking) on a bus, arbitrarily picked to have about average connectivity. This corresponds to bus 15 on the 39 bus network; bus 46 on the 57 bus network; and bus 19 on the 145 bus network. This represents the near-worst case deployment for our method.

2) **Sparse**: A few PMUs —logarithmic in the number of buses in the system, are placed arbitrarily around the network. This corresponds to buses 6, 16, and 26 on the 39 bus network; buses 5, 20, 35, and 50 on the 57 bus network; and buses 6, 26, 46, 66, and 86 on the 145 bus network.
(a) I-scores for line failure between buses 1 and 2

(b) I-scores for line failure between buses 28 and 29

Fig. 1: Two sample plots of the I-score distribution for two randomly chosen contingencies, in log scale and sorted in increasing order, for the 57 bus system with the sparse PMU configuration. The correct contingency’s I-score is in blue diamond and incorrect contingencies are labeled with a red. Note that the difference between the I-score of the correct and incorrect contingencies is a few orders of magnitude.

3) **Full**: A PMU is placed at every node in the network

In the case of **Single** and **Sparse** configurations, the systems are not observable. 46 and 88 contingencies were considered for the 39 and 57 bus system, respectively. While we ran each contingency for the 39 and 57 bus systems, we only ran a randomized subset of the IEEE 145 Test System for computational tractability (20 out of 500 plus contingencies). Matlab’s N4SID was used for damped sinusoid fitting, because of the robust implementation and range of options provided. Note that there is an n/a in the full sensor configuration of the 145 bus system; N4SID took longer than an hour to finish, so we did not test the full configuration on the 145 bus system. However, one would naturally expect the results to be equal to or better than the sparse results.

We did not test contingencies that cause convergence failures in the dynamic integrator used to generate time-domain data, which happened once in a 57 bus system test case, as no steady state is achieved. Such a contingency would likely result in system collapse without any control actions taken.

### A. Misdiagnoses

We also note that misdiagnoses provided valuable information; although our method sometimes diagnosed the incorrect contingency, the correct contingency’s I-score was always very close to the misdiagnosis’. Furthermore, we found misdiagnoses to always be topologically similar to the actual contingency that occurred.

We dive into particular misdiagnosis illustrating our point particularly well. Figure X-A contains two images explaining the anatomy of a misdiagnosis. The contingency is question is a downed line connecting buses 9 and 39, with a single PMU located in bus 16. Unfortunately, our method gets the contingency wrong. The top image displays the IEEE 39 bus system with misdiagnosis (line between Buses 9 and 39) circled in red. Note the misdiagnosis was a neighboring line to the correct diagnosis. Note also that the PMU is located on a bus quite far from the location of the contingency.

The bottom image plots I-Scores, sorted in increasing order. The misdiagnosed contingency is leftmost, marked with a red diamond. The correct contingency, right nearby, is marked with a blue diamond; the I-scores of the misdiagnosed and correct contingencies are almost identical. The rest of the incorrect contingencies are marked with red stars and have far larger I-scores than either the misdiagnosed or correct contingencies.

This case study suggests that I-scores of topologically similar contingencies are sometimes close together; similar contingencies may be indistinguishable to distant sensors. While we have not found a rigorous mathematical reason for why, this phenomena was seen in [9] as well. Indeed, just from our experimental results alone, the correct contingency was always in the top 2 smallest I-scores and any misdiagnosis was always spatially correlated to the correct contingency. In this light, we consider our misdiagnoses a failure due to lack of information rather than any failing in the method itself.

### B. Noise

The standard approach in the literature we reviewed added sensor noise by perturbing the steady-state by independent gaussian noise with a fixed standard deviation (set to be 0.0017 p.u. in terms of voltage angles according to the IEEE standard). Modeling the dynamics of sensor noise is an open research
question without much in the current body of literature from out understanding. Furthermore, the time-domain signal may be corrupted by time-dependent model parameters, such as loads, and possible GPS synchronization issues on the PMUs.

While integrating all these sources of noise into our model and simulation is a worthy cause, the tools and literature simply did not exist at the time of writing for us to do so. As a result, we opted to do the same as other authors, albeit on the dynamics rather than the steady state; we added independent gaussian noise with standard deviation of 0.0017 p.u. to the voltage angles. We opted to not use the voltage magnitudes output by the PMUs, as wave amplitudes p.u. were too small to yield useful information after being corrupted with noise. Finally, we assumed that the PMUs were all perfectly in sync.

The accuracy of our method degraded significantly; while problematic, we note that we did not spend significant effort attempting to more accurately extract modes and frequencies from the noisy time-domain signal. Furthermore, because we only considered angles and not voltages, we are using less data than before. Using standard noise-smoothing techniques, such as a low pass filter, did not yield significantly results.

XI. WEIGHING UNCERTAINTY AND ERRORS IN THE MODEL

Correct modeling of power system parameters such as loads, generator dampening factors, or line impedances is an important research problem in it’s own right. Inaccurate models may lead to faulty analysis of the system in question. This includes our own method —using a dictionary containing inaccurate models is a recipe for disaster. While dealing with arbitrary, unknown sources of uncertainty and error is an unsolved and difficult problem, dealing with structured, known sources of uncertainty and error is not. For example, models of a renewable energy sources are subject to a number of uncertainty factors such as weather. Models of renewables thus might be less accurate than models of more traditional generators. In another example, one might suspect one a load at a particular bus to vary greatly over time compared to the rest of the loads; the static parameters associated with this load may less accurate than the static parameters of other loads in the system.

We would like our method to be robust to known sources of uncertainty and error. We make the reasonable assumption that the number of these sources is small compared to the size of the system. Unfortunately, a localized error in the model will have a global effect on the dynamics, making our task slightly difficult. For instance, a model with only a few incorrect parameters may generate far different dynamics than the correct model’s dynamics.

The same thing may be said of our feature set; a localized error in the model will have a global effect on the mode shapes. Sparse perturbations to the Jacobian matrix cause non-sparse perturbations to the entire system mode. In linear-algebraic terms, perturbing a few entires of a matrix might perturb every entry of a particular eigenvector. We introduce a weighted-norm loss function to solve this problem, which we formally introduce in the following subsection.

A. Formal Analysis

Given a ground truth model defined by the Jacobian and isolation matrix pair $(J, E)$, suppose we have a set of a-priori known parameters in the model are subject to high uncertainty or error. This is equivalent to saying that there is some index set of the Jacobian

$$U = \{(i_1, j_1), (i_2, j_2) \ldots (i_q, j_q)\}$$
For which $J_{i,j} : (i, j) \in U$ is likely to be perturbed by a large amount $\Delta J$ of unknown size. We call the model denoted by the pair $(J + \Delta J, E)$ the given model, as we assume it is the model we have on hand. An eigenvector of the ground truth model is not an eigenvector of the given model.

$$(J - \lambda E)v = 0$$

$$(J + \Delta J - \lambda E)v = \Delta J v$$

This is to be expected, unless $\Delta J$ coincidentally happens to be orthogonal to $v$. To rectify this mismodel, we multiply both sides by a matrix $W$

$$W(J + \Delta J - \lambda E)v = W\Delta J v$$

Because $\Delta J v$ is nonzero along the index set $U$ and zero everywhere else, we can pick $W$ such that

$$\|W\Delta J v\|_2 < \|\Delta J v\|_2$$

In this way, the mismodel $\Delta J$ does not greatly affect the eigenvector property of $v$. That is to say, $v$ is an approximate eigenvector of the model $(WJ + W\Delta J, WE)$.

We can apply this idea to the loss function. Recall $L$ was designed to fit an entire eigenvector from a partially observed one.

$$L = \min_v \left\| \begin{pmatrix} H \\ J_j - \lambda_i E \end{pmatrix} \hat{v} - \begin{pmatrix} \hat{u}_i \\ 0 \end{pmatrix} \right\|_2^2$$

Then we can bake uncertainty into our objective function by changing the $J - \lambda E$ term to a $WJ - \lambda WE$ term.

$$L_W = \min_v \left\| \begin{pmatrix} H \\ W(j_l - \lambda_i E) \end{pmatrix} \hat{v} - \begin{pmatrix} \hat{u}_i \\ 0 \end{pmatrix} \right\|_2^2$$

This is equivalent to changing the standard 2-norm least squares calculation to a $W$-norm least squares calculation induced by the PSD matrix

$$W = \begin{pmatrix} I_k & 0 \\ 0 & WTW \end{pmatrix}$$

$$L_W = \min_\hat{v} \left\| \begin{pmatrix} H \\ J_j - \lambda_i E \end{pmatrix} \hat{v} - \begin{pmatrix} \hat{u}_i \\ 0 \end{pmatrix} \right\|_W$$  \hspace{1cm} (12)

That is to say, rather than performing a regular least-squares fit, we are performing a weighted least-squares fit. What is a reasonable choice of $W$? We suggest setting

$$W = \text{diag}(d_1 \ldots d_k) = \begin{pmatrix} d_1 \\ \vdots \\ d_k \end{pmatrix}$$

with $d_1, d_2 \ldots d_k$ chosen to not only be small but also reflect relative levels of uncertainty. In this way, we will have fulfilled the requirement that $\|W\Delta J v\|_2 < \|\Delta J v\|_2$.

In an extreme case, if we suspect each entry in $U$ to be completely faulty, setting $d_i = d_i = \ldots = d_k$ means $\|W\Delta J v\|_2 = 0$, or equivalently, that we are performing a least squares fit without considering the indices in $U$. However, one guard against doing so; there is a trade-off to be made between weighing for uncertainty and accurate calculation of I-scores. Just because one suspects an index set to be more inaccurate does mean they should discount it.

Using a weighted norm instead of a 2-norm in the loss function thus allows one to have better control over particular modal indices corresponding to a subset of the model with high uncertainty or a-priori known incorrect parameters. To summarize, we have boiled down the problem of compensating for possible sources of inaccuracy in the model stemming from known sources of uncertainty and modeling error to calculating a weighted least-squares fit instead of the standard one.

B. Example and Discussion

In a simple example, we looked a specific case of generator parameter mismodeling among contingencies. Fixing a ground truth model with Jacobian $J$, we then created a dictionary of incorrect models

$$\tilde{D} = \{ \tilde{J}_1, \tilde{J}_2 \ldots \tilde{J}_q \}$$

Where the entries of each $\tilde{J}_i$ has parameters associated with differential parameters of the first and second generators perturbed by a factor of two. Calculating the I-score using $\tilde{D}$ with the standard 2-norm loss function yielded an incorrect identification of the ground-truth model. However, assuming a-priori knowledge of this mismodeling and using an appropriate $W$-norm, we were able to correctly identify the contingency.

![Fig. 5: The sparsity pattern of the 57 bus system’s Jacobian. Nonzero entries are labeled in grey and perturbed parameters associated with generator 1 and 2 are labeled in dark red (Note that for aesthetic purposes, the dark red area has been expanded past the true sparsity pattern of the Jacobian.)](image-url)
Using the basic 2-norm measure yielded poor results. In contrast, using the weighted norm yielded the exact same result one would have gotten without the model error. We note that the perturbation and weighing matrix were selected arbitrarily to illustrate a point rather than a real-world example; we picked a large perturbation and a simple weighing matrix. Picking an optimal weighing matrix balancing a number of different criteria in weighted least squares is an important research problem itself.

XII. DECREASING COMPUTATIONAL COST VIA FILTERING

Because we are looking for the minimum I-score among all possible contingencies, we may save the current minimum I-score obtained so far. Whenever a model’s I-score exceeds the current minimum I-score, we may throw out that model automatically. This straightforward idea gives us the basis for an efficient filtering procedure. Once the current I-score exceeds the minimum, we stop, which incurs significant computational savings. To be more precise, we modify the formal procedure earlier.

A. Formal Procedure with Filtering

<table>
<thead>
<tr>
<th>Algorithm 2: Contingency Identification w/ Filtering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Input: dictionary ( D ), signal ( h(t) )</td>
</tr>
<tr>
<td>2 Output: Contingency ( c )</td>
</tr>
<tr>
<td>3 ( \text{min} \leftarrow \infty )</td>
</tr>
<tr>
<td>4 ( {(\lambda_i, u_i)}_{i=0}^k \leftarrow N4SID(h(t)) )</td>
</tr>
<tr>
<td>5 for ( \Omega_i \in D ) do</td>
</tr>
<tr>
<td>6 ( I_j \leftarrow 0 )</td>
</tr>
<tr>
<td>7 for ( i \in 0 \ldots k ) do</td>
</tr>
<tr>
<td>8 ( w_i \leftarrow |u_i|_2 )</td>
</tr>
<tr>
<td>9 ( I_j \leftarrow I_j + w_iL(\lambda_i, u_i, \Omega_j) )</td>
</tr>
<tr>
<td>10 if ( I_j &gt; \text{min} ) then</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13 if ( I_j &lt; \text{min} ) then</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15 \end</td>
</tr>
<tr>
<td>16 ( c \leftarrow \arg \min I_j )</td>
</tr>
</tbody>
</table>

B. Finding an Approximate Ordering

This filtering only works well if we can calculate the correct model’s I-score early, as this would allow one to discount more models. Of course, possessing a method guaranteed to calculate the correct contingency’s I-score is contradictory, as that would obviate the need for filtering in the first place. However, we can utilize a computationally cheaper method to determine an approximate ordering. For example, fast contingency identification routine such as [9] or a small time domain least squares fit. Alternatively, an operator might already have a good idea of where the contingency is located geographically, allowing one to refine the evaluation order. Moreover, we make sure calculate the loss functions \( L \) in an order based on the size of the their respective weights —with heavier weighted partial eigenvectors fit first. This increases the growth of I-scores and thus increases the speed at which contingencies may be discounted.

Determining the expected savings is a difficult problem dependent upon the evaluation order, number and magnitude of the partial eigenvectors, and the structure of the contingency itself. TODO: What to say?

C. Experimental Validation

Necessary?

XIII. IMPORTANT OTHER QUESTIONS

There are a wide range of considerations when using dynamic data when considering steady state data.

A. Modeling and Accounting for PMU Noise

Accurately extracting partial eigenpairs from noisy time domain data is of the utmost importance. Currently, there do not seem to be comprehensive models of PMU noise and failure modes as pertaining to dynamic data.

B. Dealing with False Positives

Sometimes, other non-permanent faults may cause transients in the power system, such as a lightning strike. We can distinguish between non-permanent faults and permanent contingencies by simply including the original system model in the dictionary. If the original system model’s I-score is lowest, a non-permanent fault has most likely occurred. Otherwise, a permanent contingency has most likely occurred.

C. How many contingencies to put in the dictionary

While we only looked at N-1 contingencies in our experimental validation, in principle, this method can deal with arbitrary N-k contingencies. One only needs to specify the models in the dictionary. Of course, the dictionary will grow drastically in size (to the power of k for N-k contingencies for line failures alone.), so balancing available computational power with key contingencies is a consideration the user will need to determine.
XIV. CONCLUSION AND FUTURE WORK

We have presented a method for contingency identification by matching PMU data to a power system model. Our method is based off the linearization of a power system’s differential-algebraic equations together with a loss function correlating features of the PMU data (in the form of modes and frequencies) to the Jacobian Matrix of the model. Previous results have already shown success in the analysis of steady-state data coming from PMUs to diagnose contingencies; our work is novel not only because it takes advantage of dynamic data but also because it introduces a general-purpose loss function with applications in other areas of power systems. We have also extended this work to include a filtering procedure for computational efficiency as well as a robust alternative to guard against modeling error and other forms of uncertainty in a system.

Future work includes working with infinite dictionaries. We assume a finite dictionary; many inverse problems have an infinite dictionary whose entries are some linear or nonlinear combinations of a set of parameters. Naively speaking, one could try discretizing over this parameter space, but this is theoretically unsound and expensive. Finally, the code used to generate our performance numbers is located here https://github.com/ericlee0803/PMU-Contingency-Identification for validation purposes.

APPENDIX A
ON DAE NORMAL FORMS

\[ \hat{\Omega} = \begin{cases} \hat{E}z(t)' + \hat{A}z(t) = PBu(t) \\ y(t) = CQz(t) + Du(t) \end{cases} \]

(13)

Where \( Qz(t) = x(t) \), \( N \) is nilpotent of degree \( m \), and \( M \) is invertible. This form is also known as the \( \Omega \), which effectively decouples our DAE. Let’s perform one last re-write to make the decoupling more obvious.

\[ \hat{\Omega} = \begin{cases} z_1(t)' + Mz_1(t) = u_1(t) \\ Nz_2(t)' + z_2(t) = u_2(t) \\ y(t) = CQz(t) \end{cases} \]

(14)

With \( z(t) = \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} \), \( PBu(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} \). So after all this work, we obtain the explicit solution

\[ \begin{cases} z_1(t) = e^{Mt}z_1(0) + \int_0^t e^{(\tau-t)M}u_1(\tau)d\tau \\ z_2(t) = \sum_{i=0}^{\infty} (-N)^i u_2^{(i)}(t) \end{cases} \]

(15)

ACKNOWLEDGMENT

The authors would like to thank Robert Owusu-Mireku and Professor Hsiao-Dong Chiang of Cornell’s ECE department for their help providing data, troubleshooting software, and giving feedback to make this project successful.

REFERENCES