A Bayesian Approach for Parameter Estimation with Uncertainty for Dynamic Power Systems

Noémi Petra, Cosmin G. Petra, Zheng Zhang, Emil M. Constantinescu, and Mihai Anitescu

***Abstract*—We address the problem of estimating the uncertainty in the solution of power grid inverse problems within the framework of Bayesian inference. We investigate two approaches, an adjoint-based method and a stochastic spectral method. These methods are used to estimate the maximum a posteriori point of the parameters and their variance, which quantifies their uncertainty. Within this framework we estimate several parameters of the dynamic power system, such as generator inertias, which are not quantifiable in steady-state models. We illustrate the performance of these approaches on a 9-bus power grid example and analyze the dependence on measurement frequency, estimation horizon, perturbation size, and measurement noise. We assess the computational efficiency, and discuss the expected performance when these methods are applied to large systems.**

***Index Terms*—Power systems, uncertainty, parameter estimation, inverse problems, Bayesian analysis.**

I. Introduction

Estimating the parameters of a system given noisy measurements is a critical problem in the operation of energy systems. Decisions about the best and safe usage of resources depend critically on knowing the current parameters or states; typically, not all these quantities are instrumented. Therefore, their values are obtained indirectly by reconciliation between the mathematical model of the system and existing measurements by an inverse estimation procedure, such as state estimation. Before the advent of phasor measurement units (PMUs) the phase angle differences in an electrical network were determined primarily indirectly by estimation from SCADA data. While PMU instrumentation can be rapidly installed on many parts of the power grid, thus resulting in their phasor angles with respect to a universal time reference being directly sensed, the ones without such measurements will still need to be inferred indirectly from model and measurements.

Moreover, the advent of renewable and distributed energy generation systems creates additional challenges that need mathematical inversion. The amount, type, and setting of generation may not be known a priori by the operator. Therefore, the parameters of their generator equivalents that need to be used for balancing the load and assessing the dynamical stability will need to be determined from measurements. The dynamical parameters—such as the equivalent inertia of a windfarm—pose particular challenges because they are not observable in steady state [1]. Therefore they are likely to need more frequent data to capture even fast perturbations from whose transients they can be inverted. The rapid deployment of PMU means that such data streams will become rapidly available, and thus such parameters can be obtained, provided that dynamic parameter estimation can be carried out.

In this paper, we focus on dynamical parameter estimation for energy systems. Given the increasing dynamic ranges of the energy systems and the uncertainty due to evolving user behavior and the increased use of distributed generation, we find it important to provide uncertainty estimates for these parameters. In this way the operator can assess the realistic stability range for next-generation energy systems.

In prior work, parameter estimation in power grid models typically has been formulated in the context of aggregated load models [2] and on the simplified swing equations [3]. The load models' parameter estimation are often obtained as a result of least-squares approaches [4]. Generally, derivative-free methods are preferred, which typically lead to minimizations based on genetic algorithms [5]; however, derivative-based least-squares have been introduced by Hiskens et al. [6], [7], [8].

Inertia estimation has gained recent traction in the literature through several studies. The approach is based on inferring individual generator inertia [9], [3], [10] or aggregated inertia [11], [12] from simplified swing equations under the assumption that all relevant dynamic quantities are measured at the generator bus or equivalent. In this study we consider the full differential algebraic system. This is important when only a limited set of observations are available, since the only way to infer parameters at unobserved locations is through a dynamical model.

Since, in an operational environment, one needs to provide an answer in all circumstances, in this work we embrace a Bayesian point of view. In this case, even with very little information we can produce an estimate that at least will encapsulate prior information about the possible ranges of parameters. With more informative data the estimation will approach the real value of the parameters, without changing the inference framework. In this sense the spread of the posterior probability density function (pdf), namely the solution

Cosmin G. Petra, Zheng Zhang, Emil M. Constantinescu, and Mihai Anitescu (e-mail: {petra,zhengzhang,emconsta,anitescu}@mcs.anl.gov) are with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439.

Noemi Petra (e-mail: npetra@ucmerced.edu) is with the Department of Applied Mathematics, School of Natural Sciences, University of California, Merced, CA 95343.

This material was based upon work supported in part by the Office of Science, U.S. Dept. of Energy, Office of Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. N. P. also acknowledges partial funding through the U. S. Dept. of Energy, Office of Workforce Development for Teachers and Scientists, 2015 Visiting Faculty Program. M. A. also acknowledges partial funding from NSF through awards FP061151-01-PR and CNS-1545046. We thank Hong Zhang and Shrirang Abhyankar for their help in setting up the IEEE example and its adjoint implementation.

of the Bayesian inverse problem, will quantify how much information from the data can be used for identifying the parameters. The challenge in solving this Bayesian inverse problem is in computing statistics of the pdf, which is a surface in high dimensions. This is extremely difficult for problems governed by expensive forward models (as is the power grid model) and high-dimensional parameter spaces (as is the case for a large-scale power grid). The difficulty stems from the fact that evaluation of the probability of each point in parameter space requires solution of the forward problem, and many such evaluations may be required to adequately sample the posterior density in high dimensions by conventional Markov-chain Monte Carlo (MCMC) methods. Hence, quantifying the uncertainties in parameters becomes intractable as we increase the grid dimension. Therefore, the approach we take is based on a local Gaussian approximation of the posterior around the maximum a posteriori (MAP) point. This approximation will be accurate when the parameter-to-observable map behaves nearly linearly over the support of the posterior [13].

We present two methods for computing MAP and estimating the parametric uncertainty: (1) an adjoint-based method and (2) a surrogate modeling approach based on polynomial chaos expansions. These methods solve the same problem but have different properties and computational cost. We will use these techniques to estimate the inertias of three generators in an IEEE 9-bus model. The situation models the circumstance where the actual bulk or distributed inertia is not known to the grid operator (as would be the case of a windfarm or other energy resources).

We carry out extensive validation experiments to demonstrate the consistency and accuracy of the methods. Moreover, we use our approach to investigate the effect of important data features on the precision of MAP. These features include the frequency of the measurements and the size of the perturbation. In addition, we compare the behavior of the two methods and discuss their computational efficiency and the expected complexities when applied to larger systems.

II. Problem Formulation

Assume that we have measurements of a dynamical system that can be modeled by an additive Gaussian noise model

$$\boldsymbol{d} = \boldsymbol{f}(\boldsymbol{m}) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{noise}}), \tag{1}$$

where $\boldsymbol{\Gamma}_{\text{noise}} \in \mathbb{R}^{q\times q}$ is the measurement noise covariance matrix and $\boldsymbol{f}(\cdot)$ a (generally nonlinear) operator mapping model parameters $\boldsymbol{m}$ to observations $\boldsymbol{d}$. Here, evaluation of this *parameter-to-observable* map $\boldsymbol{f}(\boldsymbol{m})$ requires solution of a differential-algebraic system (DAE) that models the dynamics of a power grid (followed by extraction of the DAE solution at observation points):

$$\dot{\boldsymbol{x}} = h(t, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{m}), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{2a}$$

$$\boldsymbol{0} = g(t, \boldsymbol{x}, \boldsymbol{y}), \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0. \tag{2b}$$

Here $\boldsymbol{x}$ represents the dynamic state variables (e.g., rotor angle, generator speed), $\boldsymbol{y}$ the static algebraic variables (e.g., bus voltages and line currents), $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ are the initial states,

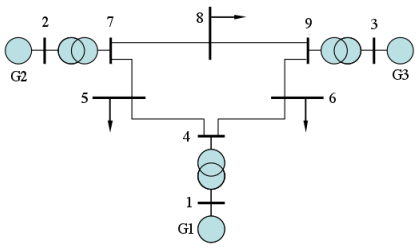

Fig. 1. IEEE 9-bus test case system. Here the buses 1, 2 and 3 are generator buses and 5, 6, and 8 load buses.

t represents time, and $\boldsymbol{m}$ the model parameters. The right-hand side h in (2a) is in general a nonlinear function that models the dynamics of the system, and g in (2b) is a set of algebraic equations modeling the passive network of the power system. For the IEEE 9-bus power grid model problem, as illustrated in Figure 1, for each generator we have seven differential (i.e., $\boldsymbol{x} \in \mathbb{R}^{21}$) and two algebraic equations, and for each network node two additional algebraic equations (i.e., $\boldsymbol{y} \in \mathbb{R}^{24}$) [14]. The inference parameter $\boldsymbol{m}$ we consider in this paper is the inertia of each generator, and thus $\boldsymbol{m} \in \mathbb{R}^3$. In realistic applications, the initial state $\boldsymbol{x}_0$ may not be known either, and it would have to be inferred from data. However, since our focus is on understanding the reconstructability of parameters that cannot be determined from steady-state measurements, such as inertias, we assume that the initial conditions $\boldsymbol{x}_0$ are known. Nevertheless, initial conditions can also be considered uncertain; and the framework introduced herein naturally extends to such cases.

The Bayesian formulation poses the parameter estimation problem as a problem of statistical inference over parameter space. The solution of the resulting Bayesian inverse problem is a *posterior probability density function (pdf)*. Bayes' Theorem states the posterior pdf explicitly as

$$\pi_{\text{post}}(\boldsymbol{m}) \propto \pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{m})\ \pi_{\text{prior}}(\boldsymbol{m}).$$

For the likelihood model, $\pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{m})$, we assume that the measured quantities are the bus voltages from a disturbance. We note that here we measure the voltage at all buses in the IEEE 9-bus power grid; however, as is the case for all other state space models [15], our framework can be used to experiment with various measurement scenarios (e.g., measurements at a subset of buses) at various time intervals and measurements of different quantities. Furthermore, since the noise $\boldsymbol{\eta}$ is independent of $\boldsymbol{m}$, thus $\boldsymbol{d}|\boldsymbol{m} \sim \mathcal{N}(\boldsymbol{f}(\boldsymbol{m}), \boldsymbol{\Gamma}_{\text{noise}})$, the likelihood is given by

$$\pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{m}) \propto \exp\Big(-\|\boldsymbol{f}(\boldsymbol{m}) - \boldsymbol{d})\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}}\Big). \tag{3}$$

The noise covariance, $\boldsymbol{\Gamma}_{\text{noise}}$, can be obtained by offline studies of the measurement setup. If the measurements are from PMUs, one can reasonably assume that the measurement noise is independent between sensors and white noise in time for one of them (on the time scale of interest, which is between 0.03 and 30 s). The variance then can be computed from the precision rating of the instrument.

The Bayesian prior, $\pi_{\text{prior}}(\boldsymbol{m})$, on the other hand, requires quantification of the existing information about the parameters.

Considerable literature exists in the area of eliciting priors, but it certainly requires an intimate analysis of the system at hand [16]. For example, for a windfarm, one can use historical logs or a simulation-based model to create a statistical model of the active inertia at a given time of the year, conditional on ambient conditions, or use information from similar windfarms. Here we use a Gaussian prior, a common choice for Bayesian inverse problems [17]. We use a prior with large variance and diagonal covariance because of the lack of a priori information about the parameters.

Restating Bayes' theorem with Gaussian noise and prior, the posteriori density function of $\boldsymbol{m}$ is described as [13], [17]

$$\pi_{\text{post}}(\boldsymbol{m}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{f}(\boldsymbol{m})-\boldsymbol{d})\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}}-\frac{1}{2}\|\boldsymbol{m}-\boldsymbol{m}_{\text{pr}}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{pr}}}\right), \quad (4)$$

where $\boldsymbol{m}_{\text{pr}}$ and $\boldsymbol{\Gamma}_{\text{pr}} \in \mathbb{R}^{n\times n}$ are the mean and covariance matrix of the prior distribution $\pi_{\text{prior}}(\boldsymbol{m})$, respectively.

Despite the choice of Gaussian prior and noise probability distributions, the posterior probability distribution need not be Gaussian, because of the nonlinearity of $\boldsymbol{f}(\boldsymbol{m})$ [13], [17]. Here we make a quadratic approximation of the negative log of the posterior (4), resulting in a Gaussian approximation $\mathcal{N}(\boldsymbol{m}_{\text{MAP}}, \boldsymbol{\Gamma}_{\text{post}})$ of the posterior. The mean of this posterior approximation, $\boldsymbol{m}_{\text{MAP}}$, is the MAP point obtained by minimizing the negative log posterior:

$$\boldsymbol{m}_{\text{MAP}} = \arg\min_{\boldsymbol{m}} \mathcal{J}(\boldsymbol{m}) := -\log \pi_{\text{post}}(\boldsymbol{m}). \quad (5)$$

The posterior covariance matrix $\boldsymbol{\Gamma}_{\text{post}}$ is then obtained by computing the inverse of the Hessian of $\mathcal{J}$ at $\boldsymbol{m}$ [13], [17].

III. Solution Methods

We present two methods for solving the inverse problem: the adjoint-based method and the stochastic spectral method. In Section IV we will illustrate the circumstances in which one approach will be favored over the other.

A. Adjoint-based method

We first introduce a numerical discretization of the forward problem. Then we detail the adjoint method in Section III-A2 for computing the gradients required when solving (5).

1) The forward problem: We represent (2) compactly by

$$M\dot{\boldsymbol{u}} = F(t, \boldsymbol{u}; \boldsymbol{m}), \ \boldsymbol{u}(0) = [x(0), y(0)]^T, \quad (6)$$

where $\boldsymbol{u} := (\boldsymbol{x}, \boldsymbol{y})$ denotes the state variables, $F = [h(\cdot), g(\cdot)]^T$, and M is the DAE mass matrix, which is block identity for $\boldsymbol{x}$ variables and zero in the rest. Note that M should not be confused with the parametric inertia $\boldsymbol{m}$. Equation (6) is discretized by using a time-stepping method. For instance, a trapezoidal-rule discretization leads to $M\boldsymbol{u}_{k+1} = M\boldsymbol{u}_k + \frac{\Delta t}{2}\left(F(t_k, \boldsymbol{u}_k; \boldsymbol{m}) + F(t_{k+1}, \boldsymbol{u}_{k+1}; \boldsymbol{m})\right)$, where $\Delta t = t_{k+1} - t_k$. With fixed $\boldsymbol{u}(0)$, each choice of the parameters $\boldsymbol{m}$ will generate a new trajectory.

2) The adjoint problem and gradient computation: To facilitate the gradient computation needed to solve (5), we use a Lagrangian approach that augments $\mathcal{J}$ with additional terms consisting of the forward DAE problem (2). Using the discrete adjoint approach [18], [19] in PETSc [20], we obtain the following discrete adjoint equations (these expressions are simply stated here; their derivation is presented in [19] and readers wanting more details on procedures for finding gradients of cost functionals governed by differential equations may consult [21], [22]):

$$\begin{aligned} M^T\lambda^* &= \lambda_{k+1} + \frac{\Delta t}{2}\left(F^T_{\boldsymbol{u}}(\boldsymbol{u}_{k+1})\lambda^* + r^T_{\boldsymbol{u}}(t_{k+1}, \boldsymbol{u}_{k+1})\right), \\ \lambda_k &= M^T\lambda^* + \frac{\Delta t}{2}\left(F^T_{\boldsymbol{u}}(\boldsymbol{u}_k)\lambda^* + r^T_{\boldsymbol{u}}(t_k, \boldsymbol{u}_k)\right) \quad (7) \\ \mu_k &= \mu_{k+1} + \frac{\Delta t}{2}\left(F^T_{\boldsymbol{m}}(\boldsymbol{u}_{k+1}) + F^T_{\boldsymbol{m}}(\boldsymbol{u}_k)\right)\lambda^* + \\ &\quad \frac{\Delta t}{2}\left(r^T_{\boldsymbol{m}}(t_{k+1}, \boldsymbol{u}_{k+1}) + r^T_{\boldsymbol{m}}(t_k, \boldsymbol{u}_k)\right), \end{aligned}$$

with $k = N-1, \ldots, 0$, $\lambda_N = \boldsymbol{0}$, $\mu_N = \boldsymbol{0}$, where $r = -\log(\pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{m}))$, and the gradients are defined by $F_{\boldsymbol{u}} = \frac{\partial F}{\partial \boldsymbol{u}}$, $F_{\boldsymbol{m}} = \frac{\partial F}{\partial \boldsymbol{m}}$, $r_{\boldsymbol{u}} = \frac{\partial r}{\partial \boldsymbol{u}}$, and $r_{\boldsymbol{m}} = \frac{\partial r}{\partial \boldsymbol{m}}$.

The gradient of $\mathcal{J}$ with respect to $\boldsymbol{m}$ can be found by enforcing that the derivative of the Lagrangian with respect to $\boldsymbol{u}$ and the adjoint variables (λ, μ) vanish. This is given by $\nabla_{\boldsymbol{m}}\mathcal{J}(\boldsymbol{m}) = \mu_0 - \boldsymbol{\Gamma}^{-1}_{\text{pr}}(\boldsymbol{m} - \boldsymbol{m}_{\text{pr}})$.

The iterative procedures of a gradient computation are as follows. First, given a parameter sample $\boldsymbol{m}$, DAE (6) is solved for the forward solution $\boldsymbol{u}$. The solution $\boldsymbol{u}$ is stored or checkpointed and further used to evaluate the data misfit term r in (7). The adjoint equation is then solved (backward in time) to obtain the adjoint solution (λ, μ). Both the forward and adjoint solutions, along with the current parameter $\boldsymbol{m}$, are used to evaluate $\nabla_{\boldsymbol{m}}\mathcal{J}(\boldsymbol{m})$. Thus, a gradient computation requires two (forward and adjoint) DAE solves. We note that while in this paper we estimate the inertia parameter, the framework is applicable to estimate other parameters in the system. In particular, with the adjoint-based method, the only expression that will change is the gradient, the adjoint problem remains the same.

3) Numerical solution to posterior minimization: The optimization problem (5) is solved with the bounded limited-memory variable-metric quasi-Newton method for nonlinear minimization implemented in TAO [23]. The method maintains a secant approximation to the Hessian (curvature) from a limited number of previous evaluations of $\mathcal{J}(\boldsymbol{m})$ and $\nabla_{\boldsymbol{m}}\mathcal{J}(\boldsymbol{m})$ and uses this approximation to compute the quasi-Newton search direction. This approach achieves the asymptotic superlinear convergence characteristic of Newton method, but without evaluating second-order derivatives [24]. The numerical estimation procedure starts with an initial guess for $\boldsymbol{m}$ and iteratively updates this parameter by performing a Moré-Thuente search [25] along the quasi-Newton direction. During this search a couple of evaluations of $\mathcal{J}$ may be needed in order to ensure sufficient decrease, which is necessary to reach local minimizers from arbitrary remote starting points. The process stops when $\|\nabla_{\boldsymbol{m}}\mathcal{J}(\boldsymbol{m})\|$ is small, which indicates that $\boldsymbol{m}$ is a local minimizer.

B. Stochastic spectral method

For the second method, the DAE (6) is simulated at a small number of samples to build a surrogate model. Then, the obtained surrogate model (instead of the forward solver) is used in the subsequent optimization to estimate the parameters. This method is particularly useful when the dimension of the parameter space is small and the forward solver has a large state-space dimension, because it saves on the number of forward dynamic simulations. Our example has 3 parameters and 21 state variables, so it belongs to this category.

Given the prior density function of $\boldsymbol{m}$, a set of polynomial chaos basis functions $\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{m})\}_{|\boldsymbol{\alpha}|=0}^{p}$ are specified. Here $\boldsymbol{\alpha} \in \mathbb{N}^n$ is an index vector, $|\boldsymbol{\alpha}|$ denotes the ℓ_1 norm, and positive integer p is the highest order of the basis functions. These basis functions are orthornormal to each other, namely $\int_{\mathbb{R}^n} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{m})\Psi_{\boldsymbol{\beta}}(\boldsymbol{m})\pi_{\text{prior}}(\boldsymbol{m})d\boldsymbol{m} = \delta_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. Then, $\boldsymbol{f}(\boldsymbol{m})$ is approximated by a truncated generalized polynomial chaos expansion $\boldsymbol{f}(\boldsymbol{m}) \approx \hat{\boldsymbol{f}}(\boldsymbol{m}) = \sum_{|\boldsymbol{\alpha}|\le p} \boldsymbol{c}_{\boldsymbol{\alpha}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{m})$ with $\boldsymbol{c}_{\boldsymbol{\alpha}}$ defined as $\boldsymbol{c}_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^n} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{m})\boldsymbol{f}(\boldsymbol{m})\pi_{\text{prior}}(\boldsymbol{m})d\boldsymbol{m}$. The total number of basis functions is $K = (p+n)!/(p!\,n!)$.

In our implementations, $\boldsymbol{c}_{\boldsymbol{\alpha}}$ are computed in two ways. The first choice is to employ projection-based stochastic collocation [26], [27], [28]. Let $\{\boldsymbol{m}_i, w_i\}_{i=1}^{N}$ be a set of quadrature points and weights corresponding to a numerical integration rule in the parameter space. Then we have $\boldsymbol{c}_{\boldsymbol{\alpha}} \approx \sum_{i=1}^{N} w_i\Psi_{\boldsymbol{\alpha}}(\boldsymbol{m}_i)\boldsymbol{f}(\boldsymbol{m}_i)$. Popular methods for choosing the quadrature points include tensor-product rules and sparse-grid methods [29]. The former needs $(p+1)^n$ samples to simulate the dynamic power systems, whereas the latter needs fewer samples by using nested grid samples. The second way is to use an interpolation method such as stochastic testing [30]. Specifically, K samples are selected, and the $\boldsymbol{c}_{\boldsymbol{\alpha}}$'s are obtained by solving a linear equation. In [30], a set of samples are generated by a quadrature rule (such as a tensor product Gauss-quadrature method); then K dominant samples $\{\boldsymbol{m}_j\}_{j=1}^{K}$ are subselected such that the matrix V (with its jth row being made of $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{m}_j)$, $|\boldsymbol{\alpha}| \le p$) is well conditioned.

With a pth-order polynomial chaos expansion for $\boldsymbol{f}(\boldsymbol{m})$, the negative log posterior now becomes a non-negative $2p$th-order polynomial function. We first write it as a combination of polynomial chaos basis function by stochastic collocation, then convert it to the summation of monomials: $-\log\pi_{\text{post}}(\boldsymbol{m}) \approx \sum_{|\boldsymbol{\alpha}|=0}^{2p} q_{\boldsymbol{\alpha}}\boldsymbol{m}^{\boldsymbol{\alpha}}$, with $\boldsymbol{m}^{\boldsymbol{\alpha}} = m_1^{\alpha_1}m_2^{\alpha_2}\dots m_n^{\alpha_n}$. With this surrogate model, (5) is simplified to $\boldsymbol{m}_{\text{MAP}} = \arg\min_{\boldsymbol{m}} \sum_{|\boldsymbol{\alpha}|=0}^{2p} q_{\boldsymbol{\alpha}}\boldsymbol{m}^{\boldsymbol{\alpha}}$. This nonconvex optimization can be solved locally with gradient-based methods as in Section III-A or globally with specialized polynomial optimization solvers such as GloptiPoly [31], [32], [33], when the parameter dimension is low.

IV. Numerical Results

In this section we evaluate how well does the MAP estimate approach the parameters used to generate the data and how good of an error indicator is the posterior variance.

A load disturbance at $t = 0.1\,\text{s}$, constant for $0.2\,\text{s}$, is inserted in the 9-bus model to provoke a transient. Its value during the switching action, L, is what characterizes this disturbance. We use as $\boldsymbol{\Gamma}_{\text{pr}}$ a diagonal matrix with diagonal entries $[5.76, 0.36, 0.09]$. The prior mean and the "truth" inertia values are $\boldsymbol{m}_{\text{pr}} = [24.00, 6.00, 3.10]$ and $\boldsymbol{m}_{\text{true}} = [23.64, 6.4, 3.01]$, respectively. We carry out the forward simulation of the DAE (2) and we create synthetic voltage measurements at all 9 buses. Here, we consider the case of independent observations; hence $\boldsymbol{\Gamma}_{\text{noise}}$ is a diagonal matrix, with diagonal entries for all computations, unless otherwise specified, 10^{-4}. The resulting voltage amplitudes, phases, and synthetic measurements are depicted in Figure 2. This synthetic data is then used in the Bayesian framework encapsulated in (5). We aim to quantify the estimation error as a function of L and the frequency of the observations, which we assume consist of time series of the voltages at all 9 buses, mimicking PMU data streams.

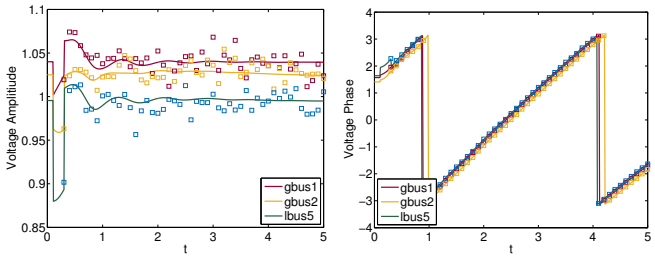

Fig. 2. Bus voltage amplitude (left) and phase (right) for the load buses 1 and 2 and generator bus 5 (red, yellow and green solid lines, respectively) based on the "truth" inertia parameter $\boldsymbol{m}_{\text{true}} = [23.64, 6.40, 3.01]$. The squares show the corresponding synthetic observations with 1% noise.

A. Computational Setup

The IEEE 9-bus example is implemented by using PETSc and is available as a part of the PETSc distribution. For future, larger, examples, the setup has the advantage of having intrinsic parallel capabilities [20], [34]. The forward and adjoint problems needed by TAO for the numerical minimization of the posterior, as described in Section III-A, are set up and solved by using the PETSc time-stepping library for DAEs.

Both methods used in this paper will produce MAP estimates and their variance for the inertia parameters $\boldsymbol{m}$ (a common notation for them in power engineering literature is H [14]). We compare the MAP estimate to the truth using a relative error metric:

$$Err = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{m}(i) - \boldsymbol{m}_{\text{true}}(i))^2/\boldsymbol{m}_{\text{true}}(i)^2}. \qquad (8)$$

To facilitate interpretation we use a similar relative metric for the standard deviation. Specifically, we use the following formula to normalize the square root of the trace of the Hessian inverse (τ):

$$\tau = \sqrt{\sum_{i=1}^{n}\boldsymbol{\Gamma}_{\text{post}}(i,i)/\boldsymbol{m}_{\text{true}}^2(i)}. \qquad (9)$$

Another statistic of interest is the positioning of the real parameter in relation to the distribution. This is not completely captured by the variance, because there could be significant bias in the estimation. To this end, we compute the cumulative

normal scores (CNS) p for the *actual values* in relation to the Gaussian approximation of the distribution:

$$p_i = \mathbf{cdf}\Big[(\boldsymbol{m}(i) - \boldsymbol{m}_{\text{true}}(i))/\sqrt{\boldsymbol{\Gamma}_{\text{post}}(i,i)}\Big]. \qquad (10)$$

Here $\mathbf{cdf}$ is the cumulative density of the standard normal random variable. CNS are between 0 and 1 and indicate how likely is that the real parameters are drawn from the aposteriori distribution, with the distinction that values very close to either 0 or 1 are considered *unlikely*.

To determine whether our analysis had a good outcome, we use the following considerations. If the estimation procedure is successful, the error Err should be small by engineering standards (a few percentages or less). If the stochastic model is a good depiction of reality, then τ should be mostly larger than Err but comparable. This reflects the fact that we are uncertain about the parameter used in the estimation (as opposed to the deterministic case); but when the data is informative, the standard deviations should be comparable to the error (though exact relational statements are difficult). A measure of successful representation of the uncertainty analysis and validation of the statistical approach is that the standard confidence values contain the real parameter. That is, the CNS of the real parameters should be away from 0 and 1 (for example, in the $[0.05, 0.95]$ range) but not clustered at 0.5, which would indicate an excessively conservative variance.

B. Results

1) Dependence on experimental design parameters: Our approach has two experimental design parameters: the length of the time horizon over which the estimation is carried out and the frequency of the data. We now present the behavior of Err, τ, and the CNS values as a function of the various choices of these parameters.

Table I shows the estimation results for different estimation horizons t_f and data frequencies Δ_t^{obs}, shown by the first column in (a) and (b), respectively. The second, third, and forth columns (m_i, i = 1, 2, 3) indicate the inverse solution, that is, the MAP point obtained with the adjoint-based and the surrogate-based methods, separated by "/"; the fifth column (#iter) indicates the number of iterations taken by the adjoint-based method to converge. The sixth and seventh columns (τ and Err) show the standard deviation normalized by the "truth" inertia parameter (as given in (9)) for the two methods and the deterministic error computed with the adjoint-based method), respectively. The last three columns show the p values computed with the adjoint-based method by using (10). For these simulations the forward problem time step was $\Delta_t = 0.01$, the load parameter (at load bus 5) was 5.5, and the iterations were terminated when the norm of the gradient fell below 10^{-6}.

For the experiment of varying the observation frequency, Δt^{obs}, we see in Table I(b), that the deterministic error Err never gets above 2%. On the other hand, the dependence of the error on the observation frequency seems weak. Concerning statistical parameters, the scaled standard deviation τ is 5% or better and the ratio of τ/Err is always less than 4.5. This means that τ will be a good indicator of the size of the error. For specific estimates, we examine the CNS values in Table I(b). In light of their definition (10), CNS are indicators of how well the posterior distribution captures the real value of the inertia that generated the data. If CNS has a value p, that is an indication that two-sided confidence intervals of level at least $|1 - 2p|$ would contain the real value, a desirable outcome. If the confidence level is chosen to be 90%, a succesful confidence interval – one that would contain the true value – would occur if the CNS score p satisfies $0.05 \leq p \leq 0.95$. From table Table I(b) we see that the CNS values are comfortably within $[0.05, 0.95]$ except for the most frequent case, $\Delta_t^{obs} = 0.01$, where the p values of the first two inertias are in the $[0.005, 0.995]$ range. This implies that, other than this case where only the 99% confidence intervals would include the real inertias, the 90% confidence intervals would indeed include the real value for all other experiments. We conclude that, in our experimental setup range, the error does not depend significantly on the frequency of the observations. But the posterior standard deviation does, as we see through the change in τ, and it subsequently reduces the uncertainty in the confidence interval for the real inertia as the observations become more frequent. We also note that standard deviations are conservative measures in the sense that τ is on the order of the error Err or larger. In terms of making specific inference statements, the CNS scores correctly bound all 15 99% confidence intervals and all but two of the 90% ones.

The experiment of increasing the observation horizon is done at a data frequency of 20 Hz ($\Delta_t^{obs} = 0.05$), which is within the sampling rates (30 per second for 0.033 Hz) supported by PMUs. The results are displayed in Table I(a). We observe that both the scaled error Err and the scaled standard deviation τ are decreasing with the increase of the time horizon. Moreover, the value of Err is always less than 4% and we have that $0.25 \leq Err/\tau \leq 1$, which indicates that τ is a conservative but accurate estimator of error. The CNS values are always in the interval of $[0.05, 0.95]$, which indicates that the 90% confidence interval will always contain the true value of the inertias. To summarize, we observe that the error improves with increased observation horizon, that the standard deviation evolution captures this improvement, and that the Bayesian 90% confidence intervals always contain the true inertia value.

Therefore, when having the choice of more frequent observations or longer estimation intervals, the latter appears to be more beneficial to the quality of the estimation once we are in range of 10 measurements per second or better. In both cases, however, the confidence intervals shrink with more information and generally capture the uncertainty of the real values.

2) Dependence on the nature of the perturbation: Having established in Section IV-B1 that the Bayesian posterior standard deviation is a good indicator of the parameter error, we estimate its behavior with the size of the load perturbation L. We note that if there were no perturbation, the system would be in steady state, and its inertias would thus not be observable. We thus anticipate that a larger perturbation would result in better estimation properties and thus lower posterior variances.

TABLE I
A STUDY OF THE EFFECT OF THE TIME HORIZON AND MEASUREMENT FREQUENCY ON THE ABILITY TO RECOVER THE INERTIA PARAMETER FOR THE POWER GRID INVERSE PROBLEM.

t_f	m_1	m_2	m_3	#iter	τ	Err	p_1	p_2	p_3
	(a) Δ_t = 0.01, Δ_t^{obs} = 0.05								
5.0	23.60 / 23.60	6.35 / 6.37	3.02 / 3.00	15	1.59e-02 / 1.79e-02	5.17e-03	0.1679	0.1484	0.5804
3.0	23.79 / 23.81	6.39 / 6.41	3.06 / 3.05	9	1.85e-02 / 2.04e-02	1.01e-02	0.9659	0.4163	0.8470
1.0	23.56 / 23.55	6.32 / 6.32	3.06 / 3.06	14	3.60e-02 / 3.66e-02	1.30e-02	0.4019	0.2384	0.7423
0.8	23.67 / 23.63	6.54 / 6.53	2.95 / 2.95	11	5.81e-02 / 5.80e-02	1.76e-02	0.5123	0.7314	0.2295
0.6	22.45 / 22.45	6.14 / 6.13	3.01 / 3.00	10	9.43e-02 / 9.29e-02	3.74e-02	0.1924	0.2337	0.4892
Δ_t^{obs}	(b) t_f = 1, Δ_t = 0.01								
0.01	23.25 / 23.23	6.29 / 6.28	2.97 / 2.97	12	1.65e-02 / 1.67e-02	1.56e-02	0.0078	0.0195	0.1641
0.02	23.81 / 23.76	6.50 / 6.49	3.00 / 2.99	12	2.34e-02 / 2.36e-02	9.79e-03	0.7635	0.8897	0.4218
0.05	23.56 / 23.55	6.32 / 6.32	3.06 / 3.06	14	3.60e-02 / 3.66e-02	1.30e-02	0.4019	0.2384	0.7423
0.10	22.91 / 23.87	6.42 / 6.43	3.06 / 3.04	13	4.82e-02 / 4.89e-02	1.11e-02	0.7332	0.5607	0.6522
0.35	23.53 / 23.52	6.23 / 6.21	2.98 / 2.98	11	9.04e-02 / 9.08e-02	1.69e-02	0.4297	0.2452	0.4412

TABLE II
COMPUTATIONAL COST FOR COMPUTING THE MAP POINT USING THE ADJOINT-BASED METHOD VERSUS AN APPROXIMATE GRADIENT-BASED METHOD. THE COST IS MEASURED IN NUMBER OF DAE SOLVES (FOR THE ADJOINT-BASED METHOD WE COUNT THE PAIR OF FORWARD AND ADJOINT SOLVES).

Load	Noise	# Iter	# Solves	Time (s)
4.25	0.01	11 / 11	14 / 100	49 / 1148
4.25	0.1	6 / 10	8 / 100	27 / 1037
7.00	0.01	9 / 10	10 / 100	33 / 1571
7.00	0.1	6 / 10	7 / 100	25 / 909

TABLE III
THE EFFECT OF THE NUMBER OF OBSERVATION BUSES ON THE ABILITY TO RECOVER THE INERTIA PARAMETER FOR THE POWER GRID INVERSE PROBLEM.

excluded buses	m_1	m_2	m_3	#iter
none	23.56	6.32	3.06	14
1, 4	23.82	6.44	3.13	9
1, 4, 2, 7	23.86	6.56	3.03	10
all except 9	23.45	6.29	3.02	9

In Figure 3 we show a surface plot of the trace of the Gaussianized posterior covariance (the sum of the parameter variances) for several noise and load values (left) and the "whiskers boxplot" of the prior and posterior mean and variances for L and σ_m values of (4.25, 0.01) and (7,0.1), respectively. These figures show that, as anticipated, the variance increases as the noise increases and the perturbation decreases, which indicates that the deterministic error will have a similar behavior. The computational cost for computing MAP points (measured in number of forward and adjoint solves) is shown in Table II, which indicates that the optimization effort is unaffected by the values of the perturbation parameters.

3) Dependence on incomplete state observations: We continue with a study on the effect of the number of observation buses on the ability to recover the inertia parameter for the power grid inverse problem. To this end we look at four cases: i) the case where we observe at all buses (we note that this case is identical with the results shown in Table I (a) with $t_f = 1$, i.e., third row); ii) the case where we remove observations from the first generator (G1), i.e., remove buses 1 and 4; iii) the case where we remove observations from the first (G1) and second (G2) generators, i.e., remove buses 1, 4, 2, and 7; and iv) the case when we keep only bus 9. The results are shown in Table III. These show that even with one bus, we can estimate the parameters reasonably well.

4) Dependence on the prior: We have also performed numerical studies to understand the effect of the prior mean and covariance on the ability to recover the inertia parameter for the power grid inverse problem. Our results show that for the first case, listed in Table I (b), we recover the same MAP point even with a drastic modification of the prior mean, i.e., up to 50% deviation from the truth, and a prior covariance with diagonal $[51.84, 3.24, 0.86]$. These results show that the estimation framework is robust with respect to the choice of the prior.

C. Computational analysis

We now discuss the computational cost for the two methods presented in this study. The adjoint-based method requires the value of the full nonlinear model and its gradient for each iteration. Additional iterations may be required in the line-search procedure. The number of forward and adjoint solves for selected cases is listed in Table II. For these simulations we used $t_f = 2$ s, Δ_t = 0.01 s, and Δ_t^{obs} = 0.1 s. The iterations for these simulations were terminated when the norm of the gradient fell below 10^{-6}. For completeness, in this table we also show the computational cost of a commonly applied approximate gradient-based method (i.e., finite difference approximation). The results show that the approximate gradient-based method failed to satisfy the convergence criteria of the optimization after 100 forward solves, which indicates lack of robustness and accuracy. In comparison, the adjoint-based method converged with considerably fewer model evaluations and much faster. While the errors in the estimation with the two methods are of same order, namely 10^{-2}, the difference in speed clearly indicates that the adjoint-based method should be used whenever adjoints can be computed.

To compare the adjoint-based method cost with the stochastic spectral method, we need to account for the cost of computing the adjoint, which is roughly the same as in the forward run. In addition to the computational time, however, the stochastic spectral element method has the advantage of working *without sensitivity information*. Given the considerable amount of legacy software for which adjoints would be

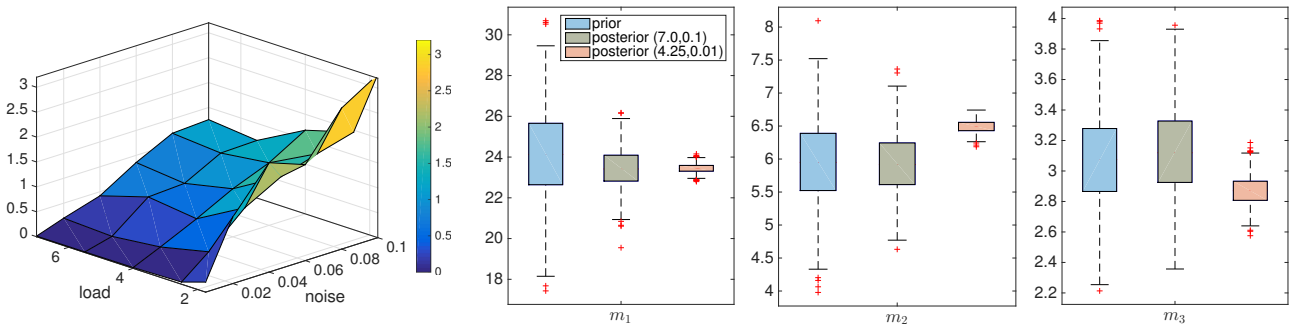

Fig. 3. Left: Surface plot of the trace of the Gaussianized posterior covariance as a function of noise and load for $t_f = 2$ s. Right: A "whiskers boxplot" of the prior and posterior mean and variances for load and noise values (7,0.1) and (4.25, 0.01) for the three inertia parameters. The central mark is the median, the edges of the box are the 25th and 75th percentiles and the "whiskers" extend to the most extreme data points.

labor-intensive to implement, this could confer it a practical advantage. Moreover, the variance can be naturally estimated with no additional cost, whereas the adjoint-based approach would need either finite differences or second-order adjoints to compute the covariance.

In Table IV we show the costs of constructing surrogate models using different approaches. In Table V we show the MAP results using different orders of surrogate models constructed by different methods. Clearly, the accuracy is significantly improved when we increase the order of polynomial chaos expansion from 1 to 2, but the improvement is marginal when we use third-order polynomial chaos expansions. From these tables we see that we can obtain good-quality estimates of the parameters and their variance using only 10 forward runs (degree 2). This is less intensive than the adjoint-based method by a factor of about 2.

1) Challenges as we increase the number of parameters and the complexity of the problem: The adjoint-based method has two major requirements: (1) code differentiation, that is, the computation and implementation of derivatives such as the ones in (7), and (2) storing the forward trajectory through checkpoints. Because only a few thousands of states need to be stored if the time scales remain the same, even for interconnect size examples this is unlikely to become a problem even on a desktop. On the one hand point (1) is a significant undertaking, although HPC tools such as PETSc increasingly provide support for it natively. Large-scale simulations, up to 22K buses are demonstrated in PETSc [35]. On the other hand, the cost of the adjoint-based method is independent of the number of parameters, and parallel implementations are also possible. In a recent study, we showed that gradient computations on IEEE 118-bus system incurs an insignificant cost for computing the gradient [19].

The stochastic spectral method proved to be robust in our experimental setting, requiring few model evaluations to construct a viable surrogate. In addition, all calculations can be trivially parallelized, and a variance estimator is intrinsic. Moreover, once a surrogate is obtained, one need not to regenerate it if the model and setting do not change. As discussed above, however, when the parameter dimensionality n is large, the number of simulation samples required can be very large, leading to an extremely high computational cost. Arguably one can obtain efficiently a high-dimensional surrogate model by using some advanced techniques, such as compressed sensing [36], tensor recovery [37], and proper generalized decomposition [38]; but these techniques may still be inefficient for extremely high-dimensional cases (e.g., when $n > 1000$).

While a definitive comparison between the two approaches is difficult to make in general because of the multiple features of the target problems, for a small number of parameters and lack of sensitivity information, the stochastic spectral element approach would be a strong candidate for a solution. In our case, it did produce good estimates a factor of 2 faster than the adjoint-based approach for the proper choice of degree and construction method, which may be difficult to guarantee a priori.

2) Considerations about deployment: While this is only an initial study, a practical implementation is worth considering. In such cases the initial state and the load would need to be inverted as well. Because these are classical analyses, a tiered approach is possible, where they are estimated separately. One can, of course, create a unified estimation approach with hybrid data sources; a mix of PMU and other data, such as SCADA, may need to be considered. While the performance of the method would need itself to be re-evaluated, this can be done in the Bayesian framework described in Section II. As described, the method assumes that we have a way to identify "micro-transients" suitable to trigger dynamical estimation. This can be done for PMU data. The method can also be modified to support any type of perturbation, as well as in a "rolling horizon" approach, where it is not triggered but used continuously. This can be done, for example, by restarting the estimation with the prior covariance being the posterior one from the previous estimation interval. We anticipate that as long as the perturbations show enough dynamic range so that the method can excite transients that are informative about the inertias, similar behaviors and performance can be expected. A more significant concern is the ability to compute the estimate in real time. We note that forward simulations for power grid transients using PETSc on interconnect-sized networks have been run faster than real time with less than 16 cores [39]. Therefore, for a few dynamic parameters to invert with uncertainty, the stochastic spectral element method could

TABLE IV
TOTAL NUMBER OF FORWARD SIMULATIONS TO CONSTRUCT THE SURROGATE MODELS.

polyn. order	Total number of forward simulations.		
	stoch. testing	tensor prod.	sparse grid
1	4	8	7
2	10	27	19
3	20	64	39

TABLE V
MAP RESULTS USING DIFFERENT ORDER OF GPC EXPANSIONS.

surrogate model	gPC order	m_1	m_2	m_3
stoch. testing	1	22.818	6.745	2.248
	2	23.600	6.372	3.000
	3	23.611	6.351	3.021
SC w/ tensor prod.	1	23.751	6.420	2.973
	2	23.585	6.374	2.991
	3	23.618	6.347	3.026
SC w/ sparse grid	1	23.962	6.322	3.156
	2	23.584	6.375	2.990
	3	23.617	6.361	3.016

in principle work "out of the box". For a large number of dynamic parameters to invert, the issue is whether the optimization can be fast enough. Certainly a promising direction is the usage of a rolling horizon approach in conjunction with inexact optimization.

V. CONCLUSIONS

We have presented a Bayesian framework for parameter estimation with uncertainty focused on the estimation of dynamic parameters of energy systems. This investigation is prompted by the rapid expansion of PMU sensors and the increased usage of renewable generation whose inertia features may change in time and may not be known to the stakeholder that must ensure transient stability operation of the system. For such systems, inertia cannot be assumed known and must thus be estimated together with its uncertainty. Because inertia has no impact on steady-state features of the system, it needs transient scenarios under which to be estimated.

We have proposed two methods to compute the MAP estimates and their variances: an adjoint-based method and a stochastic spectral method. The former has the advantage that it can compute gradients of the log-likelihood function in a time that is a constant factor of the one of the forward simulation irrespective of the number of parameters considered. This method was implemented in PETSc. The latter has the benefit of needing no sensitivity capabilities, of employing only forward simulations, and of providing an intrinsic estimate of the variance. It is suitable for the case of a limited number of parameters. We have demonstrated these methods on a 9-bus example case that is available for download [40]. The three parameters to be estimated were the generator inertias. For this example we have generated synthetic data of transient behavior by perturbing the load and adding measurement noise that we have used to assess the behavior of our approaches. When applying our method we have found that estimation time horizons of 1 s or more and data frequency of at least 10 samples per second were sufficient for the error to be less than 2%, the posterior variance to be a good estimate of the error, and some of the standard confidence intervals to cover the real parameter (with the 3 standard deviations ones always containing the real parameters). We have also observed that, as expected, the error and posterior variance decrease with increased system perturbation and decreased measurement error. The computational effort was on the order of 10 forward simulations for the stochastic spectral method and 30 forward simulations for the adjoint-based method. For usage in larger systems under real time constraints, and under realistic data streams and use cases, further work may be necessary. Nevertheless, for the small parameter case the state of technology is such that, with the use of parallel computing, the stochastic spectral method may already provide sufficient capabilities.

REFERENCES

[1] Y. Zhang, J. Bank, Y.-H. Wan, E. Muljadi, and D. Corbus, "Synchrophasor measurement-based wind plant inertia estimation," in *Green Technologies Conference, 2013 IEEE*. IEEE, 2013, pp. 494–499.

[2] V. Knyazkin, C. Canizares, L. H. Söder *et al.*, "On the parameter estimation and modeling of aggregate power system loads," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1023–1031, 2004.

[3] M. A. M. Ariff, B. C. Pal, and A. K. Singh, "Estimating dynamic model parameters for adaptive protection and control in power system," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 829–839, 2015.

[4] B.-K. Choi, H.-D. Chiang, Y. Li, H. Li, Y.-T. Chen, D.-H. Huang, and M. G. Lauby, "Measurement-based dynamic load models: derivation, comparison, and validation," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1276–1283, 2006.

[5] H. Bai, P. Zhang, and V. Ajjarapu, "A novel parameter identification approach via hybrid learning for aggregate load modeling," *Power Systems, IEEE Transactions on*, vol. 24, no. 3, pp. 1145–1154, 2009.

[6] I. A. Hiskens and A. Koeman, "Power system parameter estimation," *Journal of Electrical and Electronics Engineering Australia*, vol. 19, pp. 1–8, 1999.

[7] I. Hiskens, "Inverse problems in power systems," *Bulk Power System Dynamics and Control V*, pp. 180–195, 2001.

[8] I. Hiskens *et al.*, "Power system modeling for inverse problems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 3, pp. 539–551, 2004.

[9] P. Wall and V. Terzija, "Simultaneous estimation of the time of disturbance and inertia in power systems," *Power Delivery, IEEE Transactions on*, vol. 29, no. 4, pp. 2018–2031, 2014.

[10] S. Guo, S. Norris, and J. Bialek, "Adaptive parameter estimation of power system dynamic model using modal information," *Power Systems, IEEE Transactions on*, vol. 29, no. 6, pp. 2854–2861, 2014.

[11] X. Cao, B. Stephen, I. Abdulhadi, C. Booth, and G. M. Burt, "Switching Markov Gaussian models for dynamic power system inertia estimation," *IEEE Transactions on Power Systems*, 2015.

[12] P. M. Ashton, C. S. Saunders, G. A. Taylor, A. M. Carter, and M. E. Bradley, "Inertia estimation of the GB power system using synchrophasor measurements," *Power Systems, IEEE Transactions on*, vol. 30, no. 2, pp. 701–709, 2015.

[13] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM, 2005.

[14] P. W. Sauer and M. A. Pai, *Power system dynamics and stability*. Prentice Hall, 1998.

[15] J. L. Crassidis and J. L. Junkins, *Optimal estimation of dynamic systems*. CRC press, 2011.

[16] J. E. Oakley and A. O'hagan, "Uncertainty in prior elicitations: a non–parametric approach," *Biometrika*, vol. 94, pp. 427–441, 2007.

[17] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, ser. Applied Mathematical Sciences. New York: Springer-Verlag, 2005, vol. 160.

[18] W. Hager, "Runge-Kutta methods in optimal control and the transformed adjoint system," *Numerische Mathematik*, vol. 87, no. 2, pp. 247–282, 2000.

[19] H. Zhang, S. Abhyankar, E. Constantinescu, and M. Anitescu, "Discrete sensitivity analysis of power system dynamics," *IEEE Transactions on Circuits and Systems I: Regular Papers, submitted*, 2016.

[20] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, B. F. Smith, S. Zampini, and H. Zhang, "PETSc Web page," 2015. [Online]. Available: http://www.mcs.anl.gov/petsc

[21] Y. Cao, S. Li, and L. Petzold, "Adjoint sensitivity analysis for differential-algebraic equations: algorithms and software," *Journal of computational and applied mathematics*, vol. 149, no. 1, pp. 171–191, 2002.

[22] N. Petra and G. Stadler, "Model variational inverse problems governed by partial differential equations," The Institute for Computational Engineering and Sciences, The University of Texas at Austin, Tech. Rep. 11-05, 2011.

[23] T. Munson, J. Sarich, S. Wild, S. Benson, and L. C. McInnes, "TAO 2.0 users manual," Mathematics and Computer Science Division, Argonne National Laboratory, Tech. Rep. ANL/MCS-TM-322, 2012.

[24] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[25] J. J. Moré and D. J. Thuente, "Line search algorithms with guaranteed sufficient decrease," *ACM Trans. Math. Softw.*, vol. 20, no. 3, pp. 286–307, Sep. 1994.

[26] D. Xiu and J. S. Hesthaven, "High-order collocation methods for differential equations with random inputs," *SIAM J. Sci. Comp.*, vol. 27, no. 3, pp. 1118–1139, Mar 2005.

[27] I. Babuška, F. Nobile, and R. Tempone, "A stochastic collocation method for elliptic partial differential equations with random input data," *SIAM J. Numer. Anal.*, vol. 45, no. 3, pp. 1005–1034, Mar 2007.

[28] F. Nobile, R. Tempone, and C. G. Webster, "A sparse grid stochastic collocation method for partial differential equations with random input data," *SIAM J. Numer. Anal.*, vol. 46, no. 5, pp. 2309–2345, May 2008.

[29] T. Gerstner and M. Griebel, "Numerical integration using sparse grids," *Numer. Algor.*, vol. 18, pp. 209–232, Mar. 1998.

[30] Z. Zhang, T. A. El-Moselhy, I. M. Elfadel, and L. Daniel, "Stochastic testing method for transistor-level uncertainty quantification based on generalized polynomial chaos," *IEEE Trans. CAD of Integr. Circuits and Syst.*, vol. 32, no. 10, pp. 1533–1545, Oct 2013.

[31] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optimization*, vol. 11, no. 3, p. 796-817, 2001.

[32] ——, "An explicit equivalent positive semidefinite program for 0-1 nonlinear programs," *SIAM J. Optimization*, vol. 12, no. 3, pp. 756–769, 2001.

[33] D. Henries and J. B. Lasserre, "Gloptipoly: Global optimization over polynomials with Matlab and SeDuMi," *ACM Trans. Mathematical Software*, vol. 29, no. 2, p. 165-194, 2003.

[34] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, B. F. Smith, S. Zampini, and H. Zhang, "PETSc users manual," Argonne National Laboratory, Tech. Rep. ANL-95/11 - Revision 3.6, 2015. [Online]. Available: http://www.mcs.anl.gov/petsc

[35] S. Abhyankar, E. Constantinescu, B. Smith, A. Flueck, and D. Maldonado, "Acceleration of dynamic simulations using parallel Newton-GMRES-Schwarz methods," *Transactions on Smart Grid Special Issue on High Performance Computing (HPC) Applications for a More Resilient and Efficient Power Grid, Submitted*, 2016.

[36] X. Yang and G. E. Karniadakis, "Reweighted l_1 minimization method for stochastic elliptic differential equations," *J. Comp. Phys.*, vol. 248, no. 1, pp. 87–108, Sept. 2013.

[37] Z. Zhang, H. D. Nguyen, K. Turitsyn, and L. Daniel, "Probabilistic power flow computation via low-rank and sparse tensor recovery," *arXiv preprint: arXiv:1508.02489*, pp. 1–8, Aug 2015.

[38] A. Nouy, "Proper generalized decomposition and separated representations for the numerical solution of high dimensional stochastic problems," *Arch. Comp. Meth. Eng.*, vol. 27, no. 4, pp. 403–434, Dec 2010.

[39] S. Abhyankar, B. Smith, and E. Constantinescu, "Evaluation of overlapping restricted additive Schwarz preconditioning for parallel solution of very large power flow problems," in *Proceedings of the 3rd International Workshop on High Performance Computing, Networking and Analytics for the Power Grid*, ser. HiPCNA-PG '13. ACM, 2013, pp. 5:1–5:8.

[40] [Online]. Available: https://github.com/Argonne-National-Laboratory/PowerSystemsEstimation.git

Noemi Petra is an assistant professor in the Applied Mathematics department in the School of Natural Sciences at the University of California, Merced. She earned her Ph.D. degree in applied mathematics from the University of Maryland, Baltimore County in 2010. Prior to joining UC Merced, Noemi was the recipient of the ICES Postdoctoral Fellowship in the Institute for Computational Engineering & Sciences at the University of Texas at Austin. Her research interests include inverse problems, uncertainty quantification and optimal experimental design.

Cosmin G. Petra is an assistant computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory. He received the B.Sc. degree in mathematics and computer science from "Babeş-Bolyai" University, Romania, and the M.S. and Ph.D. degrees in applied mathematics from the University of Maryland, Baltimore County. His research interests lie at the intersection of numerical optimization, operations research and high-performance scientific computing with a focus on the optimization of complex energy systems under uncertainty.

Zheng Zhang received his Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA. He is currently a postdoc associate with the Mathematics and Computer Science Division at Argonne National Laboratory. His research interests include high-dimensional uncertainty quantification and data analysis for nanoscale devices and systems, energy systems, and biomedical applications. He received the 2015 Doctoral Dissertation Seminar Award from the Microsystems Technology Laboratory of MIT, and the 2014 Best Paper Award from IEEE Transactions on CAD of Integrated Circuits and Systems. His industrial research experiences include Coventor Inc. and Maxim-IC.

Emil M. Constantinescu received his Ph.D. degree in computer science from Virginia Tech, Blacksburg, in 2008. He is currently a computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory and he is on the editorial board of SIAM Journal on Scientific Computing. His research interests include numerical analysis of time-stepping algorithms and their applications to energy systems.

Mihai Anitescu is a senior computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory, a professor in the Department of Statistics at the University of Chicago, and a Senior Fellow of the Computation Institute at the University of Chicago. He obtained his engineer diploma (electrical engineering) from the Polytechnic University of Bucharest in 1992 and his Ph.D. in applied mathematical and computational sciences from the University of Iowa in 1997. He specializes in the areas of numerical optimization, computational science, numerical analysis and uncertainty quantification. He co-authored more than 100 peer-reviewed papers in scholarly journals, book chapters, and conference proceedings, and he is on the editorial board of Mathematical Programming A and B, SIAM Journal on Optimization, SIAM Journal on Scientific Computing, and SIAM/ASA Journal in Uncertainty Quantification; and he is a senior editor for Optimization Methods and Software.