

**Bayesian Inference of Parameters in Power System Dynamic
Model Using Trajectory Sensitivities**

by

Rubinderjit Singh Nagi

B.A.Sc., University of British Columbia, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Electrical & Computer Engineering)

The University of British Columbia
(Vancouver)

March 2021

© Rubinderjit Singh Nagi, 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Bayesian Inference of Parameters in Power System Dynamic Model Using Trajectory Sensitivities

submitted by **Rubinderjit Singh Nagi** in partial fulfillment of the requirements for the degree of **MASTER OF APPLIED SCIENCE** in **Electrical & Computer Engineering**.

Examining Committee:

Dr. Yu Christine Chen, Electrical and Computer Engineering
Supervisor

Dr. Juri Jatskevich, Electrical and Computer Engineering
Supervisory Committee Member

Dr. Ryoza Nagamune, Mechanical Engineering
Supervisory Committee Member

Abstract

We propose an analytically tractable Bayesian method to infer parameters in power system dynamic models from noisy measurements of bus-voltage magnitudes and frequencies as well as active- and reactive-power injections. The proposed method is computationally appealing as it bypasses the large number of system model simulations typically required in sampling-based Bayesian inference. Instead, it relies on analytical linearization of the nonlinear system differential-algebraic-equation model enabled by trajectory sensitivities. Central to the proposed method is the construction of a linearized model with the maximum probability of being (closest to) the actual nonlinear model that gave rise to the measurement data. The linear model together with Gaussian prior leads to a conjugate family where the parameter posterior, model evidence, and their gradients can be computed in closed form, markedly improving scalability for large-scale power systems. We illustrate the effectiveness and key features of the proposed method with numerical case studies for a 3-bus system. Algorithmic scalability is then demonstrated via case studies involving the New England 39-bus test system.

Lay Summary

Power system reliability studies help to ensure the grid is functioning properly at all times. In this regard, time-domain simulations are typically used to assess the system behaviour in many potential what-if scenarios. The simulations require a system model with accurate parameters. Inaccurate parameters can cause discrepancies between the simulated and actual system behaviours, leading to inability to predict potentially undesired or damaging outcomes. The measurements we take in the field are corrupted by noise, which makes it difficult to determine the true values of the parameters. We use Bayes' theorem to compute the probability distribution of parameter values in the model that could have generated the measurement data. By assuming Gaussian prior for parameters, we get closed-form expressions of the parameter posterior, model likelihood, and their gradients. We further use derivatives of the measured outputs with respect to the parameters to find a linear model which most likely resembles the measurement data.

Preface

This thesis is based on the following paper that has been submitted for review:

- R. Nagi, X. Huan, and Y. C. Chen, “Bayesian Inference of Parameters in Power System Dynamic Model Using Trajectory Sensitivities,” (Under review) submitted in January 2021.

As first author of the above mentioned paper, the author of this thesis developed simulations, analyzed approaches and solutions, and verified mathematical models and findings. My supervisor, Dr. Yu Christine Chen provided constructive feedback and comments throughout my master program and has contributed to the development of my paper and thesis. The work was supported by the co-author of the paper, Dr. Xun Huan, who assisted with the mathematical models of the Bayesian framework problem.

Table of Contents

- Abstract iii
- Lay Summary iv
- Preface v
- Table of Contents vi
- List of Tables viii
- List of Figures ix
- Acknowledgments xi
- Dedication xii

- 1 Introduction 1**
 - 1.1 Motivation 1
 - 1.2 Literature Review 3
 - 1.3 Thesis Contribution 4
 - 1.4 Thesis Outline 4

- 2 Pertinent Models 5**
 - 2.1 System DAE Model 5
 - 2.2 Trajectory Sensitivities 7
 - 2.3 Output Approximation Using Trajectory Sensitivities 11

2.4	Measurement Model and Problem Statement	14
3	Bayesian Approach	15
3.1	Inference on λ	15
3.2	Choice of λ^*	17
3.2.1	Evaluation Metric and Problem Formulation	18
3.2.2	Computing the Metric	19
3.2.3	Update Strategy	19
4	Case Studies	26
4.1	3-Bus Test System	26
4.1.1	Choice of λ^*	27
4.1.2	Inference on λ	27
4.2	WECC 9-Bus Test System	31
4.3	New England 39-Bus Test System	34
5	Conclusions	37
	Bibliography	39
A	3-Bus Test System	42
A.1	System Parameters	42
A.2	Model Details	43
B	Bayesian Conjugacy for Linear-Gaussian Models	45
B.1	Linear-Gaussian Posterior Mean and Covariance	45
B.2	Linear-Gaussian Log-Evidence	47
C	Log-Evidence Gradient for Linearized DAE	48
C.1	Log-Evidence Gradient with respect to Nominal Trajectory λ^*	48
C.2	Evaluating $z_{\lambda\lambda_i}^*$	50

List of Tables

Table 2.1	Parameter values for dynamic model of SMIB system.	7
Table 4.1	WECC 9-bus test system: true parameter values λ^{true} and posterior mean $\mu_{\pi, \lambda_{\text{opt}}^*}$	33
Table 4.2	NE 39-bus test system: true parameter values λ^{true} and posterior mean $\mu_{\pi, \lambda_{\text{opt}}^*}$	35
Table A.1	Parameter values for dynamic model of 3-bus system.	42

List of Figures

Figure 2.1	One-line diagram for SMIB test system.	6
Figure 2.2	One-line diagram for 3-bus test system.	9
Figure 2.3	Trajectory sensitivities z_{λ}^* of the system output z with respect to λ around the nominal trajectory.	10
Figure 2.4	Actual and approximate trajectories for SMIB system.	12
Figure 2.5	Actual and approximate trajectories for 3-bus system.	13
Figure 3.1	Prior and posterior of λ	17
Figure 3.2	Optimization trajectory for nominal parameter value λ^*	23
Figure 3.3	Optimization trajectory for λ posterior.	24
Figure 3.4	Pairwise marginal PDFs and marginal PDFs for λ posterior for different number of measurements and noise level.	25
Figure 4.1	3-bus test system: convergence of nominal parameters	28
Figure 4.2	3-bus test system: convergence of gradient vector	29
Figure 4.3	3-bus test system: pairwise marginal PDFs and convergence of objective function.	30
Figure 4.4	3-bus test system: comparison amongst output trajectories.	31
Figure 4.5	One-line diagram for WECC 9-bus test system.	32
Figure 4.6	WECC 9-bus test system: convergence of nominal parameters	32
Figure 4.7	WECC 9-bus test system: convergence of objective function.	33
Figure 4.8	WECC 9-bus test system: pairwise parameter correlation.	33
Figure 4.9	WECC 9-bus test system: standard deviations	34
Figure 4.10	One-line diagram for NE 39-bus test system.	34

Figure 4.11 NE 39-bus test system: convergence of objective function.	35
Figure 4.12 NE 39-bus test system: standard deviations	36
Figure 4.13 NE 39-bus test system: pairwise parameter correlation.	36

Acknowledgments

I am thankful for my wonderful supervisor, Dr. Yu Christine Chen, whose guidance, support, and patience is much appreciated. I am also thankful to Dr. Xun Huan for his insight and ideas, without whom this work would not have been possible.

Special thanks to my senior research group members Abdullah Al-Digs and Bo Chen for their advice, mentoring, and technical support.

Dedication

I dedicate wholeheartedly this work to my *soulmate I.R., mother, father, and sister* whose support and passion helped me to stand strong in challenging situations.

Chapter 1

Introduction

This chapter presents motivation, and related literature for parameter inference in power system and Bayesian framework. Also included in this chapter are the thesis contributions and its structure.

1.1 Motivation

Online monitoring of power system operational reliability relies on a model with accurate parameters for the network, generators, loads, and other components. Discrepancies between the actual system and its model, including erroneous parameter values, have contributed to major cascading failures. For example, following the 1996 Western Interconnection outages, engineers could not reproduce recorded disturbance measurements in simulation with the prevailing models, indicating that previous studies to set system operating limits were likely invalid due to model mismatches [1]. Recognizing the importance of accurate models for operating studies, the North American Electric Reliability Corporation (NERC) requires utilities to validate and calibrate models (and pertinent parameters therein) used in system-level dynamic simulations every ten years [2, 3]. The development of computational tools aimed at model calibration can be greatly facilitated by widespread deployment of phasor measurement units (PMUs). Measurements available from PMUs include magnitude, angle, and frequency of voltage and current phasors. These quantities are typically collected at a very high speed (usually 30 measurements per second), and phasors measured at different locations by different devices are time synchronized [4].

Model calibration generally leverages the formulation and solution of either a parameter *esti-*

mation or *inference* problem. In the former, parameters are tuned so that behaviour predicted by the model best matches the measurement data via, e.g., weighted least-squares estimation [5, 6], gradient-based optimization [7], and optimization incorporating trajectory sensitivities [8]. These approaches typically do not quantify the uncertainty in the resulting parameter estimates affected by, e.g., the quantity and quality of measurement data. Yet, uncertainty quantification can provide highly insightful confidence measures in decision making, so as to avoid unnecessary risk or excessive engineering margins. In light of this, we pursue the latter *inference* approach, which computes the entire probability distribution of parameter values that could have induced the measurement data. In accordance with Bayes' theorem, an initial *prior* distribution is determined before observing any measurements, and it is updated to a *posterior* distribution that represents the uncertainty associated with the inferred model parameters *conditioned* on the measurement data. Under the Bayesian framework (see, e.g., [9, 10]), we develop an analytically tractable method to compute an approximate posterior distribution for model parameters conditioned on voltage-phasor and complex-power measurements obtained from (possibly a subset of) generator buses.

Classical Bayesian inference uses Markov chain Monte Carlo (MCMC) algorithms (see, e.g., [11, 12]) that typically require thousands (or more) repeated simulations of the system model under study. However, since models of large-scale power systems consist of many nonlinear differential-algebraic equations (DAEs) describing generator and load dynamics coupled across an expansive transmission network, MCMC can become computationally impractical, even prohibitive. Our strategy avoids MCMC altogether by linearizing the nonlinear DAE model around a nominal system trajectory (*not* a single operating point) resulting from potentially major disturbances like large-signal load changes or faults. The linearized time-varying model comprises trajectory sensitivities that are analytically derived by differentiating the DAEs with respect to parameters, which are then evaluated along the nominal trajectory induced by a particular choice of parameter values [13]. A question then naturally arises: what are the parameter values that engender the *best* linearized model? To address this, we employ a Bayesian model selection formalism to compute and maximize the model evidence (or equivalently, Bayes factor), i.e., the probability of measurement data given a candidate model [14, 15]. Although the model evidence is generally difficult to estimate [16], we can compute its value analytically in the space of linearized models along with Gaussian conjugate priors. We also derive the analytical gradient of model evidence along the trajectory of linearization,

so as to facilitate gradient-based optimization methods to find the evidence-maximizing linear model. The posteriors for all linear models considered in the procedure, including the Bayesian-optimal one, are Gaussian due to conjugacy, where the mean and covariance can be obtained in closed form.

1.2 Literature Review

We next provide a review of methods for power system parameter inference. Classical online data assimilation can be derived as various approximations of the Bayesian filtering problem (see, e.g., [17] Ch. 2 & 4), and the extended [18], unscented [19], and ensemble [20] Kalman filters have been used to approximate posteriors for static parameters by augmenting the dynamic state vector. However, filter-based methods are typically inefficient for updating static model parameters because the DAE states are also filtered, and done so sequentially over time. Thus, when only static parameters are of interest, a batch inference approach is suitable to find their posterior conditioned simultaneously on all measurement data. For batch inference, MCMC algorithms are commonly used to sample the posterior. MCMC is ergodic and can capture generic distributions with complex correlation structures. However, direct MCMC with large-scale DAE models is generally impractical, especially with higher-order dynamic models or larger parameter spaces. To sidestep this issue, [21–23] pre-build computationally inexpensive surrogate models, via polynomial chaos expansions (PCEs), that replace the DAE model in MCMC. Although these demonstrate orders-of-magnitude speedups compared to MCMC with DAEs, PCEs are difficult to scale up to high-dimensional settings. For instance, a third-order PCE for the 40-parameter example in this thesis would entail solving for 12,341 PCE coefficients and require a large training set of offline DAE simulations. Furthermore, a separate PCE is needed for each scalar element of the entire observation vector. Thus, [22,23] construct distinct PCEs for individual generators while neglecting the transmission network, and [21] demonstrates a case with only three model parameters. A promising alternative also proposed in [21] is a Laplace approximation to the posterior enabled by the adjoint solution to the DAE model, but the method is validated on a relatively small test system with three parameters.

1.3 Thesis Contribution

Our work advances over the previous by considering the full DAE model describing generator dynamics coupled across the transmission network. The explicit inclusion of the network enables parameter inference at buses without sufficient local measurements. We demonstrate this aspect and other key features along with algorithmic scalability with case studies involving a 3-bus system, the Western Electric Coordinating Council (WECC) 3-machine 9-bus, and the New England (NE) 10-machine 39-bus test system. Also distinct from previous work in this domain, we construct an optimal linearized model that maximizes the probability of measurement data amongst candidate linear models. The linear models enable closed-form evaluations of the posterior, model evidence, and their gradients, so the proposed method scales Bayesian inference to higher dimensions than PCEs. Finally, we note that the trajectory sensitivities employed in model linearization are useful in other pertinent problems, such as parametric uncertainty assessment, transient stability analysis, and dynamic security assessment [13, 24].

1.4 Thesis Outline

The remainder of the thesis is organized as follows. Chapter 2 outlines pertinent models for the power system and noisy measurements, and it also describes the inference problem. In Chapter 3, we present the proposed Bayesian inference computation. Chapter 4 provides numerical case studies to demonstrate the effectiveness and scalability of the proposed method. Finally, Chapter 5 concludes the thesis.

Chapter 2

Pertinent Models

In this chapter, we present the power system DAE model and the associated trajectory sensitivities. We further describe the measurement statistical model and state the Bayesian parameter inference problem tackled in this thesis.

2.1 System DAE Model

Consider an AC transmission grid with synchronous generators serving constant-power loads over an interconnected power network. Let $x \in \mathbb{R}^n$ collect generator dynamic state variables, such as generator rotor angular position and speed. Also let $y \in \mathbb{R}^q$ collect stator algebraic state variables of all generators and voltage magnitudes and phase-angles of all buses in the network. Further let $u \in \mathbb{R}^d$ collect generator setpoints and $\lambda \in \mathbb{R}^p$ unknown parameters to be inferred. Then, the system electromechanical behaviour can be described by the following DAE model:

$$\dot{x} = f(x, y, u; \lambda), \quad (2.1)$$

$$0 = g(x, y; \lambda), \quad (2.2)$$

where, for a given λ , $f : \mathbb{R}^{n+q+d} \mapsto \mathbb{R}^n$ collects system dynamic equations, and $g : \mathbb{R}^{n+q} \mapsto \mathbb{R}^q$ collects algebraic constraints [25]. The output $z \in \mathbb{R}^m$ can be mapped from system dynamic and algebraic state variables, as follows:

$$z = h(x, y; \lambda), \quad (2.3)$$

where, for a given λ , $h : \mathbb{R}^{n+q} \mapsto \mathbb{R}^m$. Note that although we consider a static load model for simplicity, dynamic loads can be easily incorporated into the DAE modelling framework in (2.1)–(2.3) [25]. Next, we will illustrate the DAE modelling considerations with the canonical single-machine infinite-bus (SMIB) system depicted in Fig. 2.1.

Example 1 (SMIB: DAE Model). Consider the SMIB system shown in Fig. 2.1, where the synchronous generator is described by the classical model and the infinite bus is modelled as a constant voltage source [25]. Let δ denote the angular position of the generator rotor in electrical radians, and let ω denote the angular velocity of the rotor in per unit (p.u.). Collect dynamic state variables in $x = [\delta, \omega]^T$, algebraic state variables in $y = P^e$, which is the generator active-power output, and the system input is $u = P^m$, which is the generator turbine mechanical power. Also consider measurements in $z = [\omega, P^e]^T$, and parameters $\lambda = [H, D]^T$, where H is the inertia constant, and D is the damping constant. With the above notation in place, the SMIB system can be described by the following DAE model:

$$\dot{\delta} = 120\pi(\omega - \omega_s) =: f_1(x, y, u; \lambda), \quad (2.4)$$

$$\frac{2H}{\omega_s} \dot{\omega} = P^m - D(\omega - \omega_s) - P^e =: f_2(x, y, u; \lambda), \quad (2.5)$$

$$0 = P^e - \frac{EV_\infty}{X} \sin \delta =: g(x, y), \quad (2.6)$$

$$z_1 = \omega =: h_1(x, y), \quad (2.7)$$

$$z_2 = P^e =: h_2(x, y), \quad (2.8)$$

where E , V_∞ , and X are constants [25]. The system parameter values are reported in Table 2.1. The initial steady-state values are $P^m(0) = 1$ p.u., $\omega(0) = \omega_s = 1$ p.u., $P^e(0) = 1$ p.u., $\delta(0) = \sin^{-1} \left(\frac{P^e(0)X}{EV_\infty} \right) = 0.269$ rad. ■

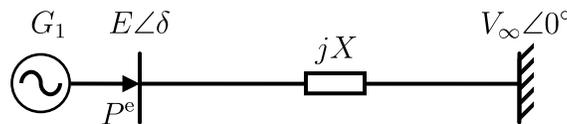


Figure 2.1: One-line diagram for SMIB test system.

Table 2.1: Parameter values for dynamic model of SMIB system shown in Fig. 2.1.

E (p.u.)	V_∞ (p.u.)	ω_s (p.u.)	X (p.u.)	H (s)	D (p.u.)
1	1	1	0.266	1	20

2.2 Trajectory Sensitivities

Suppose the DAE system described in (2.1)–(2.3) evolves from initial conditions $x(0) = x_0$ and $y(0) = y_0$ according to nominal input vector $u = u^*$ and parameter value $\lambda = \lambda^*$, so that system states follow nominal trajectory $(x^*, y^*, u^*; \lambda^*)$. Denote by $x_\lambda^* \in \mathbb{R}^{n \times p}$ ($y_\lambda^* \in \mathbb{R}^{q \times p}$) the linear sensitivities of x (y) with respect to λ around the nominal trajectory. Note that u is assumed to be independent of λ , so $u_\lambda^* \equiv 0$. Differentiation of (2.1)–(2.2) with respect to λ yields the following linear dynamic system for trajectory sensitivities:

$$\dot{x}_\lambda^* = f_x^* x_\lambda^* + f_y^* y_\lambda^* + f_\lambda^*, \quad (2.9)$$

$$0 = g_x^* x_\lambda^* + g_y^* y_\lambda^* + g_\lambda^*, \quad (2.10)$$

where

$$\begin{aligned} f_x^* &= \frac{\partial f}{\partial x}, & f_y^* &= \frac{\partial f}{\partial y}, & f_\lambda^* &= \frac{\partial f}{\partial \lambda}, \\ g_x^* &= \frac{\partial g}{\partial x}, & g_y^* &= \frac{\partial g}{\partial y}, & g_\lambda^* &= \frac{\partial g}{\partial \lambda}, \end{aligned}$$

are, in general, time-varying matrices evaluated along the nominal trajectory $(x^*, y^*, u^*; \lambda^*)$ [8]. Similarly, differentiation of (2.3) with respect to λ yields trajectory sensitivities of output z around the nominal trajectory, denoted by $z_\lambda^* \in \mathbb{R}^{m \times p}$ and given by

$$z_\lambda^* = h_x^* x_\lambda^* + h_y^* y_\lambda^* + h_\lambda^*, \quad (2.11)$$

where

$$h_x^* = \frac{\partial h}{\partial x}, \quad h_y^* = \frac{\partial h}{\partial y}, \quad h_\lambda^* = \frac{\partial h}{\partial \lambda},$$

are evaluated along the nominal trajectory $(x^*, y^*, u^*; \lambda^*)$. We assume that along the nominal trajectory, the power flow Jacobian matrix is invertible. Then we can rearrange (2.10) as

$$y_\lambda^* = -(g_y^*)^{-1}(g_x^* x_\lambda^* + g_\lambda^*). \quad (2.12)$$

Further, substitution of (2.12) into (2.9) and (2.11) results in the following linear time-varying system describing how the trajectory sensitivities evolve along the nominal trajectory:

$$\dot{x}_\lambda^* = A^* x_\lambda^* + B^*, \quad (2.13)$$

$$z_\lambda^* = C^* x_\lambda^* + D^*, \quad (2.14)$$

with initial condition $x_\lambda^*(0) = \mathbb{0}_{n \times p}$, and where time-varying matrices A^* , B^* , C^* , and D^* are given by

$$A^* = f_x^* - f_y^*(g_y^*)^{-1}g_x^*, \quad B^* = f_\lambda^* - f_y^*(g_y^*)^{-1}g_\lambda^*, \quad (2.15)$$

$$C^* = h_x^* - h_y^*(g_y^*)^{-1}g_x^*, \quad D^* = h_\lambda^* - h_y^*(g_y^*)^{-1}g_\lambda^*. \quad (2.16)$$

A simultaneous time-domain simulation of (2.1)–(2.3) and (2.13)–(2.14) yields the output trajectory sensitivities in z_λ^* evaluated along the nominal output trajectory z^* .

Example 2 (SMIB: First-order Trajectory Sensitivities). In this example, we refer to the SMIB system from Example 1 and demonstrate how first-order trajectory sensitivities z_λ^* are computed. We assume that the nominal input $P^{m*}(0) = 1$ p.u. and $P^{m*}(t) = 1.45$ p.u. for $t > 0$ s. We evaluate the following trajectory sensitivities along the nominal trajectory $(x^*, y^*, u^*; \lambda^*)$:

$$\begin{aligned} f_x^* &= \begin{bmatrix} 0 & 120\pi \\ 0 & -\frac{\omega_s D^*}{2H^*} \end{bmatrix}, & f_y^* &= \begin{bmatrix} 0 \\ -\frac{\omega_s}{2H^*} \end{bmatrix}, & f_\lambda^* &= \begin{bmatrix} 0 & 0 \\ -\frac{\omega_s(P^{m*} - D^*(\omega^* - \omega_s) - P^{e*})}{2H^{*2}} & -\frac{\omega_s(\omega^* - \omega_s)}{2H^*} \end{bmatrix}, \\ g_x^* &= \begin{bmatrix} -\frac{EV_\infty}{X} \cos \delta^* & 0 \end{bmatrix}, & g_y^* &= 1, & g_\lambda^* &= \begin{bmatrix} 0 & 0 \end{bmatrix}, \\ h_x^* &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, & h_y^* &= \begin{bmatrix} 0 \\ 1 \end{bmatrix}, & h_\lambda^* &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

With the above in place, matrices in (2.13)–(2.14) evaluate as follows:

$$A^* = \begin{bmatrix} 0 & 120\pi \\ -\frac{\omega_s EV_\infty}{2XH^*} \cos \delta^* & -\frac{\omega_s D^*}{2H^*} \end{bmatrix}, \quad B^* = \begin{bmatrix} 0 & 0 \\ -\frac{\omega_s(P^{m*} - D^*(\omega^* - \omega_s) - P^{e*})}{2H^{*2}} & -\frac{\omega_s(\omega^* - \omega_s)}{2H^*} \end{bmatrix},$$

$$C^* = \begin{bmatrix} 0 & 1 \\ \frac{EV_\infty}{X} \cos \delta^* & 0 \end{bmatrix}, \quad D^* = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

A simultaneous simulation of (2.4)–(2.8) and (2.13)–(2.14) with the above provides the nominal trajectory and trajectory sensitivities evaluated around it. ■

Example 3 (3-Bus System: Trajectory Sensitivities). Here, we consider the 3-bus toy system shown in Fig. 2.2 to further illustrate how the quantity z_λ^* informs how variations in parameter λ affects the output z , especially with respect to couplings between pairs of generators. The system initially operates at steady state, and at $t = 0^+$ s, the active-power load P_3 increases by 0.5 p.u. Simulations of the DAE system in (2.1)–(2.3) are performed in PSAT [26] using a detailed two-axis machine model along with turbine/governor and exciter controls for each generator [25]. Parameter values and the initial steady-state power-flow solution are reported in Appendix A.1. Suppose z collects the terminal voltage magnitude, rotor speed, and active- and reactive-power injections of G_1 and G_2 , i.e., $z = [V^T, \omega^T, P^T, Q^T]^T$ ($m = 8$), where $V = [V_1, V_2]^T$, $\omega = [\omega_1, \omega_2]^T$, $P = [P_1, P_2]^T$, and $Q = [Q_1, Q_2]^T$. Furthermore, we are interested in sensitivities with respect to inertia constant $H = [H_1, H_2]^T$, damping constant $D = [D_1, D_2]^T$, droop constant $R_D = [R_{D1}, R_{D2}]^T$, and governor time constant $T_{CH} = [T_{CH1}, T_{CH2}]^T$ of both generators, so $\lambda = [H^T, D^T, R_D^T, T_{CH}^T]^T$ ($p = 8$).

We evaluate z_λ^* at the parameter values listed in Table A.1. In Fig. 2.3, we plot the time

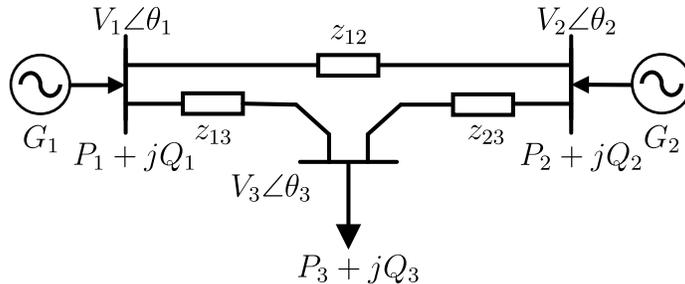


Figure 2.2: One-line diagram for 3-bus test system.

evolution of trajectory sensitivity $z_{i,j}^*$ where $i = 1, \dots, m$, $j = 1, \dots, p$, which corresponds to the entry in matrix z_λ^* at the i th row and j th column. As an example, the plot for P_D^* in Fig. 2.3j consists of the following trajectory sensitivities:

- P_{1,D_1}^* : Sensitivity of output P_1 with respect to parameter D_1 (solid dark blue trace),
- P_{1,D_2}^* : Sensitivity of output P_1 with respect to parameter D_2 (solid dark orange trace),
- P_{2,D_1}^* : Sensitivity of output P_2 with respect to parameter D_1 (dash-dot blue trace),
- P_{2,D_2}^* : Sensitivity of output P_2 with respect to parameter D_2 (dash-dot orange trace).

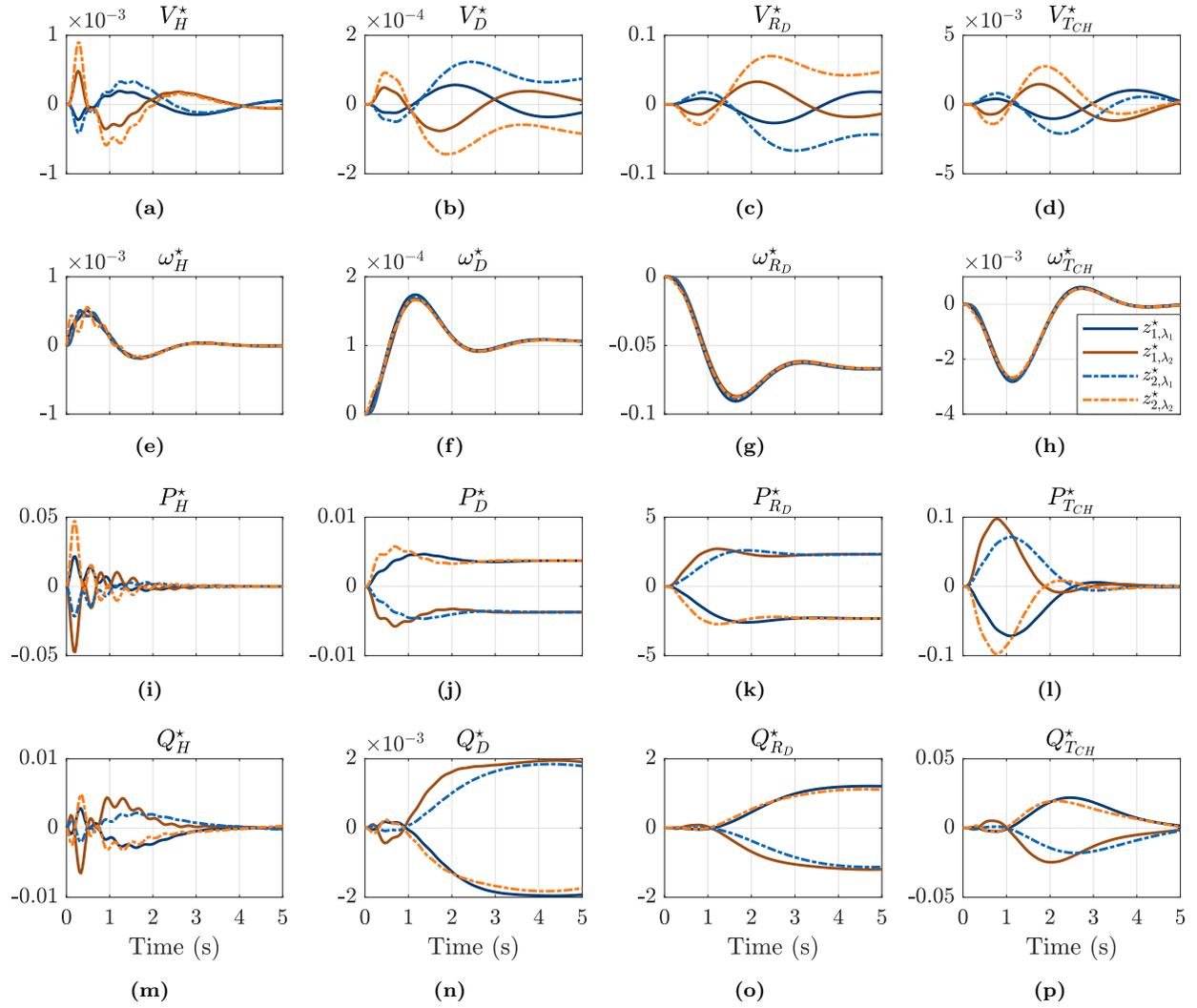


Figure 2.3: Example 3: Trajectory sensitivities z_λ^* of the system output z with respect to λ around the nominal trajectory.

We next focus on outputs ω and P for illustrative purposes. The sensitivities of these outputs with respect to H and T_{CH} decrease as the system approaches steady state, i.e., ω_H^* , $\omega_{T_{CH}}^*$, P_H^* , and $P_{T_{CH}}^*$ approach zero as t grows, as shown in Figs. 2.3e, 2.3h, 2.3i, and 2.3l, respectively. Thus, changing these parameters does not affect the steady-state value of the system output, which is expected as time constants are multiplied by the time derivatives. On the other hand, sensitivities with respect to D and R_D have nonzero values at steady state, and indeed these parameters affect the steady-state values of ω and P as follows:

$$\Delta\omega_i(\infty) = \frac{-\Delta P_3}{D_1 + \frac{1}{R_{D1}} + D_2 + \frac{1}{R_{D2}}}, \quad (2.17)$$

$$\Delta P_i(\infty) = -\Delta\omega_i \left(D_i + \frac{1}{R_{Di}} \right), \quad (2.18)$$

for each generator i . A visual inspection of Figs. 2.3e–h reveals that ω_λ^* consists of similar trajectory sensitivities for each generator and parameter as the buses to which the generators are connected are coupled through a line. Since the load increases, the angular velocity of the rotors, ω , decreases. Positive ω_D^* implies increasing D would cause an increase in ω accordingly, i.e., for this system disturbance, $\Delta\omega_i$ in (2.17) would become smaller when D increases. On the other hand, increasing R_D would lead to a larger drop in ω for the same disturbance. The traces for ω_D^* , $\omega_{R_D}^*$ in Figs. 2.3f, 2.3g and P_D^* , $P_{R_D}^*$ in Figs. 2.3j, 2.3k match the steady-state behaviour in (2.17) and (2.18), respectively.

In Fig. 2.3j, the traces for P_{1,D_1}^* (P_{1,D_2}^*) and P_{2,D_1}^* (P_{2,D_2}^*) have the same magnitude but opposite signs, which implies if D_1 (D_2) is positively correlated with P_1 (P_2), then D_1 (D_2) would be negatively correlated with P_2 (P_1) to balance the generation and load in the system. This can be observed in the trajectory sensitivities of P with respect to H , R_D , and T_{CH} in Figs. 2.3i, 2.3k, and 2.3l, respectively. ■

2.3 Output Approximation Using Trajectory Sensitivities

We can use the trajectory sensitivities obtained in (2.14) to approximate the output of a perturbed system that results from variations in λ around λ^* . To this end, let $z = z^* + \Delta z$, where Δz results from $\Delta\lambda = \lambda - \lambda^*$. Then, assuming that $\Delta\lambda$ is sufficiently small, we can approximate the system

output around the nominal trajectory as follows:

$$z \approx a(\lambda^*)\lambda + b(\lambda^*) =: \tilde{z}(\lambda; \lambda^*), \quad (2.19)$$

where $a(\lambda^*) = z_\lambda^*$ and $b(\lambda^*) = z^* - z_\lambda^*\lambda^*$ are parameterized by the choice of λ^* .

Example 4 (SMIB: Output Approximation). In this example, going back to the SMIB system from Example 1, we approximate the system output using first-order trajectory sensitivities in (2.19). We start at the initial steady state described in Example 1; apply disturbance by increasing P^m from 1 p.u. to 1.45 p.u. at time $t = 0^+$ s and plot the nominal trajectories ω^* and P^{e*} in Fig. 2.4 (solid trace). We decrease H and D by 10%, repeat the above simulation and plot the perturbed trajectories ω and P^e (dash-dot trace) in Fig. 2.4a. We use (2.19) for approximate trajectories $\tilde{\omega}$ and \tilde{P}^e (dashed trace). We then repeat the above, but with a decrease of 20% in parameter values, and plot the results in Fig. 2.4b.

The perturbed trajectory for 20% change in parameter values in Fig. 2.4b is unsurprisingly farther away from the nominal trajectory than the perturbed one with 10% parameter variation in Fig. 2.4. The first-order trajectory sensitivities provide sufficiently accurate approximation of the actual perturbed trajectory with lower parameter variations. We next provide further justification

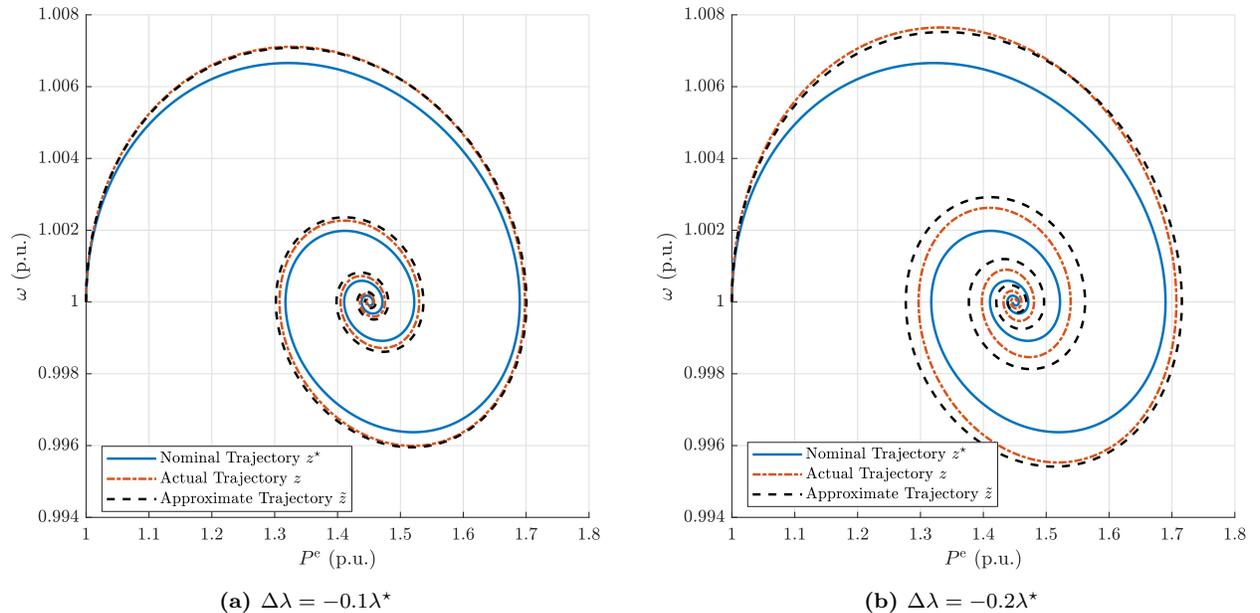


Figure 2.4: Example 4: Actual and approximate trajectories of active-power generation and rotor speed due to parameter variations.

for the accuracy of the first-order approximation using the 3-bus test system from Example 3. ■

Example 5 (3-Bus: Output Approximation). We revisit the 3-bus test system from Example 3 and illustrate the modelling concepts introduced in this chapter. The system initially operates at steady state and then responds to an increase of 10% in the active-power load at bus 3 at time $t = 15$ s. Simulations of the DAE system in (2.1)–(2.3) are performed in PSAT [26] using a detailed two-axis machine model along with turbine/governor and exciter controls for each generator [25]. Parameter values and the initial steady-state power-flow solution are reported in Appendix A.1. Suppose z collects the G_1 terminal voltage magnitude, its rotor speed, and active- and reactive-power injections, i.e., $z = [V_1, \omega_1, P_1, Q_1]^T$. In Fig. 2.5, the nominal trajectory for G_1 terminal voltage V_1^* and reactive power Q_1^* resulting from the load disturbance and nominal parameter values are plotted as the solid blue trace. Next, suppose that, just after time $t = 0$ s, exciter parameters for both generators collected in λ increase by 15%. The red dash-dot trace in Fig. 2.5 depicts the new V_1 and Q_1 trajectory with the perturbed parameter values. For comparison, the black dashed trace represents the trajectory \tilde{V}_1, \tilde{Q}_1 approximated by (2.19) with perturbed parameter values, where z_λ^* is obtained via custom MATLAB code that implements time-domain simulations of (2.13)–(2.14) along the nominal trajectory. The output trajectory approximated by trajectory sensitivities indeed closely matches the actual nonlinear system behaviour. ■

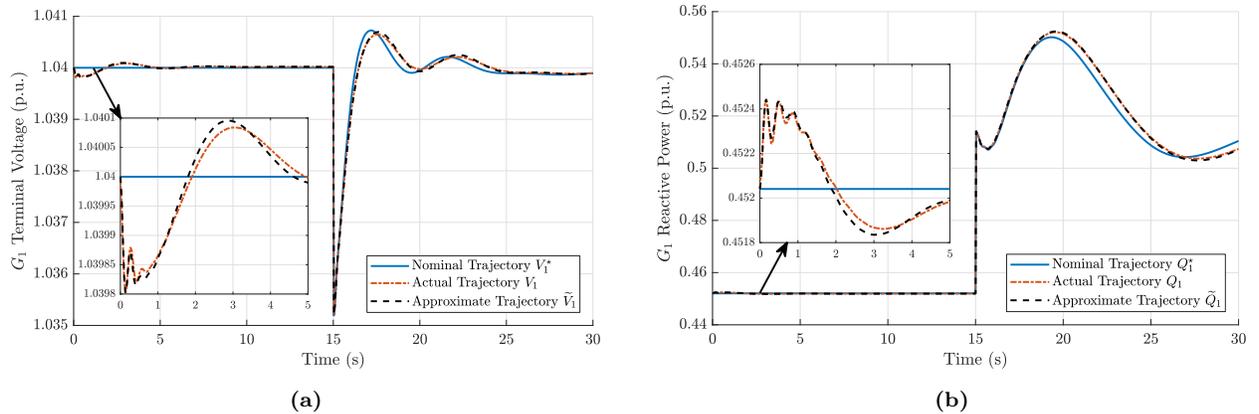


Figure 2.5: Example 5: Actual and approximate trajectories of (a) G_1 terminal voltage magnitude and (b) reactive power due to exciter parameter variations.

2.4 Measurement Model and Problem Statement

In our setting, we assume PMUs provide synchronized measurements of terminal voltage and current phasors, as well as rotor speeds of (possibly a subset of) generators. If generator rotor speed measurements are not readily available, measurements of the voltage phasor frequency at the bus connected to the generator provide sufficient estimates [27]. Let $z_{[k]}$ denote the actual system output (in our case studies, this is supplied by a time-domain simulation of the DAE model furnished with the true parameter values) collected at time $t = k\Delta t$, $k = 0, 1, 2, \dots, M$, $\Delta t > 0$, i.e., $z_{[k]} = z(k\Delta t)$. Also let $\widehat{z}_{[k]}$ denote the measurement of system output at time $t = k\Delta t$. Considering noisy measurements, $\widehat{z}_{[k]}$ can be modelled as

$$\widehat{z}_{[k]} = z_{[k]} + \xi_{[k]} \approx a_{[k]}\lambda + b_{[k]} + \xi_{[k]} =: \widetilde{z}_{[k]} + \xi_{[k]}, \quad (2.20)$$

where $\xi_{[k]} \in \mathbb{R}^m$ denotes additive Gaussian noise associated with PMU measurements, and the approximation is obtained by substituting (2.19). In (2.20), $\widetilde{z}_{[k]}$ is linear with respect to λ , and the linear- and constant-term coefficients $a_{[k]}$ and $b_{[k]}$ depend on the choice of λ^* . Specifically, $a_{[k]}$ and $b_{[k]}$ are constructed with $z_{[k]}^*$ and $z_{\lambda, [k]}^*$, which are obtained by observing discrete data points in a time-domain simulation of (2.1)–(2.3) and (2.13)–(2.14) with nominal parameter value λ^* . Furthermore, the entries in $\xi_{[k]}$ are independent and identically distributed under a joint Gaussian distribution with zero mean and covariance Σ_ξ , i.e., $\xi_{[k]} \sim \mathcal{N}(\mathbf{0}_m, \Sigma_\xi)$, where $\Sigma_\xi \in \mathbb{R}^{m \times m}$ is diagonal with each diagonal entry reflecting the corresponding measurement precision. Finally, we collect, in $\{\widehat{z}_{[k]}\}$, the set of $M + 1$ available measurements of the system output, $\widehat{z}_{[0]}, \dots, \widehat{z}_{[M]}$.

Using the models established in this chapter, we tackle two interrelated but distinct problems. The first is to identify the best λ^* where the approximate linear model in (2.20) most likely resembles the measurement-generating nonlinear system. Second, we infer the parameter λ from measurements $\widehat{z}_{[k]}$, given a linearized model constructed around the nominal output trajectory resulting from a particular choice of λ^* . We approach both problems under a Bayesian framework, where λ^* and λ are treated as random variables, as detailed next in Chapter 3.

Chapter 3

Bayesian Approach

In this chapter, we first describe the solution strategy for the inference of λ given a linearized model constructed with a particular choice of λ^* . We then outline the proposed optimization approach to find the best λ^* choice.

3.1 Inference on λ

The model in (2.20), which is linear in λ , is obtained by linearizing the DAE model along the nominal trajectory induced by a particular choice of λ^* . Treating the unknown parameter λ as a random variable, its prior represents the uncertainty in λ before making any observations through measurements, and the posterior represents the updated uncertainty after observing measurement data collected in $\{\hat{z}_{[k]}\}$. We make the reasonable assumption that the prior of λ is independent of the trajectory of linearization, so the prior probability density function (PDF) $f(\lambda|\lambda^*) = f(\lambda)$ irrespective of λ^* . Further denote the posterior PDF of λ conditioned on measurements by $f(\lambda|\{\hat{z}_{[k]}\}, \lambda^*)$. Then, direct application of Bayes' theorem for conditional probability yields

$$f(\lambda|\{\hat{z}_{[k]}\}, \lambda^*) = \frac{f(\{\hat{z}_{[k]}\}|\lambda, \lambda^*)f(\lambda)}{f(\{\hat{z}_{[k]}\}|\lambda^*)}, \quad (3.1)$$

where $f(\lambda)$ is the prior PDF for λ , $f(\{\hat{z}_{[k]}\}|\lambda, \lambda^*)$ is the likelihood function, and $f(\{\hat{z}_{[k]}\}|\lambda^*)$ is the model evidence (a λ -independent normalization constant for the posterior PDF). Solving the Bayesian inference problem entails characterizing the posterior (e.g., sampling from or calculating its PDF). We can typically evaluate the prior and likelihood, but not the model evidence. However,

we next describe a conjugate formulation to compute the posterior analytically.

We prescribe Gaussian prior $\lambda \sim \mathcal{N}(\mu_\circ, \Sigma_\circ)$ to represent the initial uncertainty in λ . The likelihood then follows from the linearized measurement model in (2.20):

$$f(\{\hat{z}_{[k]}\}|\lambda, \lambda^\star) = \prod_{k=0}^M (2\pi)^{-\frac{m}{2}} |\Sigma_\xi|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(\hat{z}_{[k]} - \tilde{z}_{[k]})^\top \Sigma_\xi^{-1}(\hat{z}_{[k]} - \tilde{z}_{[k]})\right), \quad (3.2)$$

where $\tilde{z}_{[k]}$ represents discrete data points taken from (2.19) as the approximate system output for a given parameter value λ with the linearized model constructed from nominal parameter value λ^\star . Note that we drop the dependence of $\tilde{z}_{[k]}$ on λ and λ^\star in (3.2) to contain notational burden. The combination of linear model together with Gaussian prior and likelihood leads to a conjugate system, where the posterior is also Gaussian:

$$(\lambda|\{\hat{z}_{[k]}\}, \lambda^\star) \sim \mathcal{N}(\mu_{\pi, \lambda^\star}, \Sigma_{\pi, \lambda^\star}), \quad (3.3)$$

with the mean and covariance in closed form given by

$$\mu_{\pi, \lambda^\star} = \Sigma_{\pi, \lambda^\star} \left(\Sigma_\circ^{-1} \mu_\circ + \sum_{k=0}^M a_{[k]}^\top \Sigma_\xi^{-\top} (\hat{z}_{[k]} - b_{[k]}) \right), \quad (3.4)$$

$$\Sigma_{\pi, \lambda^\star} = \left(\Sigma_\circ^{-1} + \sum_{k=0}^M a_{[k]}^\top \Sigma_\xi^{-1} a_{[k]} \right)^{-1}, \quad (3.5)$$

respectively. See Appendix B.1 for the derivation of (3.2), (3.4), and (3.5). Above, the subscript π indicates posterior and the subscript λ^\star reminds us that the mean and covariance of the posterior depend on the choice of λ^\star for linearization.

Example 6 (SMIB: Prior and Posterior of λ). We utilize the SMIB system from Example 1 to demonstrate how the posterior mean and covariance is computed. We measure output signals collected in $z = [\omega, P^e]^\top$ ($m = 2$) and infer the parameters collected in $\lambda = [H, D]^\top$ ($p = 2$). We run a time-domain simulation for $t = 1$ s and record samples every 0.0333 s which provides 30 sets of measurements ($M = 30$). The initial guess for λ^\star is set as $\lambda_{(0)}^\star = 0.9\lambda^{\text{true}}$, where λ^{true} denotes the true measurement-generating parameter values in the nonlinear DAE. We prescribe Gaussian prior $\lambda \sim \mathcal{N}(\mu_\circ, \Sigma_\circ)$ regardless of the value taken by $\lambda_{(\ell)}^\star$, where $\mu_\circ = 1.4\lambda^{\text{true}}$, and $\Sigma_\circ = 0.5^2 \text{diag}(\lambda^{\text{true}})^2$.

We assume Gaussian measurement noise with 0 mean and 0.01% standard deviation for ω and 0.1% for P^e . These values are consistent with measurement uncertainty associated with PMU technology [28,29]. With this, we get

$$\lambda^{\text{true}} = \begin{bmatrix} 1 \\ 20 \end{bmatrix}, \quad \lambda_{(0)}^* = \begin{bmatrix} 0.9 \\ 18 \end{bmatrix}, \quad \mu_o = \begin{bmatrix} 1.4 \\ 28 \end{bmatrix},$$

and the covariance matrices

$$\Sigma_o = \begin{bmatrix} 0.5^2 & 0 \\ 0 & 10^2 \end{bmatrix}, \quad \Sigma_\xi = \begin{bmatrix} 10^{-8} & 0 \\ 0 & 10^{-6} \end{bmatrix}.$$

We then substitute the above values into (3.4) and (3.5) to get the posterior mean and covariance. The prior and posterior mean with one standard deviation for parameters H and D are plotted in Fig. 3.1a and Fig. 3.1b, respectively. The posterior distribution is conditioned on the measurements, and it is associated with much lower uncertainty than the initial prior distribution, indicating the effectiveness of Bayesian inference. ■

3.2 Choice of λ^*

The process of finding the best linearized model requires three main ingredients: i) formulating the optimal model selection problem, including the metric that evaluates the “goodness” of candidate

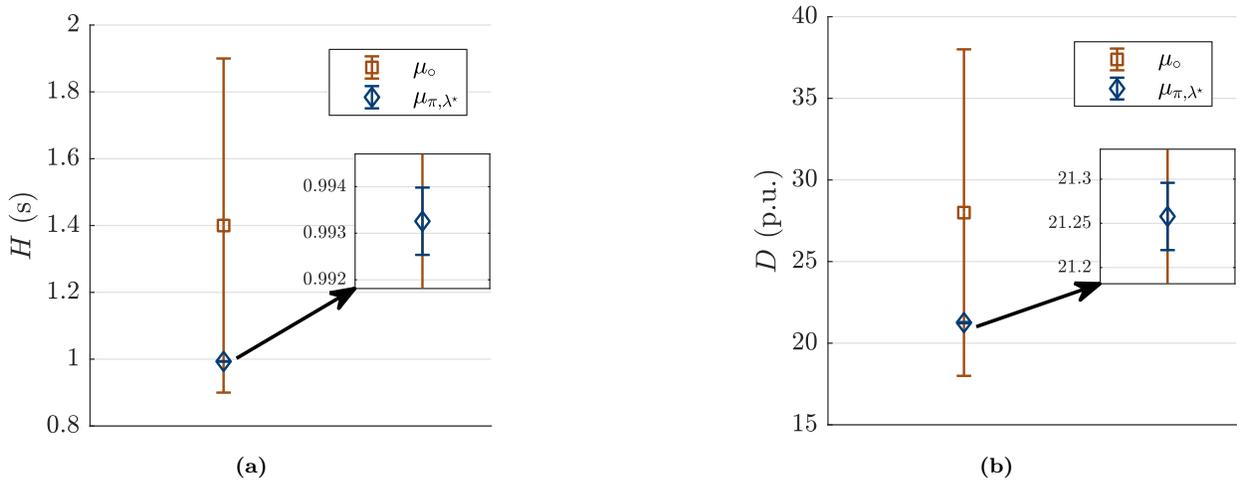


Figure 3.1: Example 6: Prior and posterior of (a) H and (b) D .

models, ii) computing the value of this metric for a given candidate model, and iii) proposing new candidate models within the optimization routine.

3.2.1 Evaluation Metric and Problem Formulation

Continuing under the Bayesian framework, we adopt the methods of Bayesian model selection (or equivalently, Bayes factors) [14, 15]. When considering a finite number of models, application of Bayes' theorem given candidate model Φ_i yields the following model-posterior probability mass function:

$$\Pr(\Phi_i|\{\hat{z}_{[k]}\}) = \frac{f(\{\hat{z}_{[k]}\}|\Phi_i)\Pr(\Phi_i)}{f(\{\hat{z}_{[k]}\})}, \quad (3.6)$$

which is the probability of model Φ_i being (closest to) the true measurement-generating model as supported by, i.e., conditioned on, measurement data in $\{\hat{z}_{[k]}\}$. Since our problem deals with a continuous spectrum of models (parameterized by λ^* that is treated as a continuous random variable), we consider the continuous analogue of (3.6) given by

$$f(\lambda^*|\{\hat{z}_{[k]}\}) = \frac{f(\{\hat{z}_{[k]}\}|\lambda^*)f(\lambda^*)}{f(\{\hat{z}_{[k]}\})}. \quad (3.7)$$

Therefore, we evaluate the “goodness” of different linearized models by comparing their model-posterior $f(\lambda^*|\{\hat{z}_{[k]}\})$ in (3.7).¹ The *best* candidate λ^* thus maximizes this quantity (equivalently, its logarithm), as follows:

$$\lambda_{\text{opt}}^* = \arg \max_{\lambda^*} \ln f(\lambda^*|\{\hat{z}_{[k]}\}). \quad (3.8)$$

In (3.7), $f(\{\hat{z}_{[k]}\})$ is a constant normalization factor that does not depend on λ^* . Furthermore, assuming a uniform model-prior (i.e., $f(\lambda^*)$ remaining constant regardless of the choice of λ^*), (3.8) is equivalent to

$$\lambda_{\text{opt}}^* = \arg \max_{\lambda^*} \ln f(\{\hat{z}_{[k]}\}|\lambda^*). \quad (3.9)$$

In other words, here the maximum *a posteriori* (MAP) λ^* value in (3.8) is identical to the maximum likelihood estimator (MLE) of λ^* in (3.9). We thus seek to solve (3.9).

¹Bayes factor differs slightly by focusing on the (ratios of) model likelihood $f(\{\hat{z}_{[k]}\}|\lambda^*)$ instead of the model-posterior. However, they are equivalent when the model-prior is uniform. We will invoke this shortly.

3.2.2 Computing the Metric

The key to solving (3.9) is recognizing that the model likelihood $f(\{\widehat{z}_{[k]}\}|\lambda^*)$ is precisely the model evidence (i.e., the denominator) in (3.1). The model evidence is generally very challenging to compute and, for this reason, the task of computing it is typically avoided in Bayesian inference. However, we can obtain it in closed form owing to the analytical posterior in (3.3)–(3.5). Rearranging (3.1) and taking the logarithm of the resultant, we get

$$\ln f(\{\widehat{z}_{[k]}\}|\lambda^*) = \ln f(\{\widehat{z}_{[k]}\}|\lambda, \lambda^*) + \ln f(\lambda) - \ln f(\lambda|\{\widehat{z}_{[k]}\}, \lambda^*). \quad (3.10)$$

We then substitute into (3.10) the closed-form expressions for the prior PDF $\lambda \sim \mathcal{N}(\mu_o, \Sigma_o)$ and posterior PDF $(\lambda|\{\widehat{z}_{[k]}\}, \lambda^*) \sim \mathcal{N}(\mu_{\pi, \lambda^*}, \Sigma_{\pi, \lambda^*})$ along with the likelihood function in (3.2). We then arrive at the following analytical closed-form expression for the log-evidence:

$$\begin{aligned} \ln f(\{\widehat{z}_{[k]}\}|\lambda^*) &= -\frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_o| - \frac{1}{2} (\lambda - \mu_o)^T \Sigma_o^{-1} (\lambda - \mu_o) - \frac{m(M+1)}{2} \ln(2\pi) \\ &\quad - \frac{M+1}{2} \ln |\Sigma_\xi| - \frac{1}{2} \sum_{k=0}^M (\widehat{z}_{[k]} - \widetilde{z}_{[k]})^T \Sigma_\xi^{-1} (\widehat{z}_{[k]} - \widetilde{z}_{[k]}) + \frac{p}{2} \ln(2\pi) \\ &\quad + \frac{1}{2} \ln |\Sigma_{\pi, \lambda^*}| + \frac{1}{2} (\lambda - \mu_{\pi, \lambda^*})^T \Sigma_{\pi, \lambda^*}^{-1} (\lambda - \mu_{\pi, \lambda^*}), \end{aligned} \quad (3.11)$$

where $\widetilde{z}_{[k]}$, μ_{π, λ^*} , and Σ_{π, λ^*} depend on the value of λ^* . Note that (3.11) holds for any value of λ .

3.2.3 Update Strategy

Various optimization algorithms (e.g., gradient-based, quasi-Newton, and derivative-free methods) can be employed to iteratively select candidates for λ^* toward the optimizer λ_{opt}^* of (3.9). For example, adopting gradient-ascent leads to the following update formula:

$$\lambda_{(\ell+1)}^* = \lambda_{(\ell)}^* + \gamma_{(\ell)} \nabla_{\lambda^*} \ln f(\{\widehat{z}_{[k]}\}|\lambda^*)|_{\lambda_{(\ell)}^*}, \quad (3.12)$$

where $\gamma_{(\ell)}$ is a learning rate (gradient-ascent step size) and $\nabla_{\lambda^*} \ln f(\{\widehat{z}_{[k]}\}|\lambda^*)|_{\lambda_{(\ell)}^*}$ is the gradient of the objective $\ln f(\{\widehat{z}_{[k]}\}|\lambda^*)$ evaluated at $\lambda_{(\ell)}^*$. A major advantage of the proposed framework is that the objective function in (3.9) and its gradient can be computed in closed form, enabling greater scalability. Particularly, we completely bypass all numerical approximations of the gradient

involving, e.g., finite differences, which would be computationally impractical for high-dimensional λ^* . Next, we discuss details with respect to analytical computation of the gradient and iterative updates of candidate models.

Remark 1 (Analytical Gradient Computation). The i th entry of the gradient vector in (3.12) can be computed analytically by differentiating (3.11) with respect to λ_i^* as

$$\begin{aligned} \frac{\partial}{\partial \lambda_i^*} \ln f(\{\widehat{z}_{[k]}\}|\lambda^*) &= \frac{1}{2} \sum_{k=0}^M \widetilde{z}_{\lambda_i^*, [k]}^T (\Sigma_\xi^{-1} + \Sigma_\xi^{-T}) (\widehat{z}_{[k]} - \widetilde{z}_{[k]}) - \frac{1}{2} \text{Tr} \left(\Sigma_{\pi, \lambda^*} \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} \right) \\ &\quad - (\lambda - \mu_{\pi, \lambda^*})^T \Sigma_{\pi, \lambda^*}^{-1} \frac{\partial \mu_{\pi, \lambda^*}}{\partial \lambda_i^*} + \frac{1}{2} (\lambda - \mu_{\pi, \lambda^*})^T \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} (\lambda - \mu_{\pi, \lambda^*}), \end{aligned} \quad (3.13)$$

where $\widetilde{z}_{\lambda_i^*, [k]}$ is obtained by observing the derivative of (2.19) with respect to λ_i^* at time instant k , given by

$$\widetilde{z}_{\lambda_i^*, [k]} = a_{\lambda_i^*, [k]} \lambda + b_{\lambda_i^*, [k]}. \quad (3.14)$$

with

$$a_{\lambda_i^*, [k]} = z_{\lambda \lambda_i^*, [k]}^*, \quad b_{\lambda_i^*, [k]} = -z_{\lambda \lambda_i^*, [k]}^* \lambda^*. \quad (3.15)$$

The expression for $b_{\lambda_i^*, [k]}$ is derived by applying the product rule and recognizing that $z_{\lambda_i^*, [k]}^* = z_{\lambda, [k]}^* e_i$, where e_i is the i th basis vector. Also, $z_{\lambda \lambda_i^*, [k]}^*$ represents discrete data points of $z_{\lambda \lambda_i^*}$, the partial derivative of z_λ^* with respect to λ_i^* , at time instant k . See Appendix C.2 for details on how to obtain $z_{\lambda \lambda_i^*}^*$. Furthermore, we can differentiate (3.4) and the inverse of (3.5) with respect to λ_i^* to get

$$\frac{\partial \mu_{\pi, \lambda^*}}{\partial \lambda_i^*} = \Sigma_{\pi, \lambda^*} \left(\sum_{k=0}^M \frac{\partial a_{[k]}^T}{\partial \lambda_i^*} \Sigma_\xi^{-T} (\widehat{z}_{[k]} - b_{[k]}) - a_{[k]}^T \Sigma_\xi^{-T} \frac{\partial b_{[k]}}{\partial \lambda_i^*} \right) + \frac{\partial \Sigma_{\pi, \lambda^*}}{\partial \lambda_i^*} \Sigma_{\pi, \lambda^*}^{-1} \mu_{\pi, \lambda^*}, \quad (3.16)$$

$$\frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} = \sum_{k=0}^M \frac{\partial a_{[k]}^T}{\partial \lambda_i^*} \Sigma_\xi^{-1} a_{[k]} + a_{[k]}^T \Sigma_\xi^{-1} \frac{\partial a_{[k]}}{\partial \lambda_i^*}. \quad (3.17)$$

Finally, recognizing that

$$\frac{\partial \Sigma_{\pi, \lambda^*}}{\partial \lambda_i^*} = -\Sigma_{\pi, \lambda^*} \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} \Sigma_{\pi, \lambda^*}, \quad (3.18)$$

we can substitute (3.17) into (3.18) and use the resultant to further simplify (3.16). ■

Remark 2 (Normalized Quantities). In our problem setting, the inferred parameters and associated

gradients may differ by several orders of magnitude. In order to promote numerical stability, we make use of normalized parameters and gradients defined as, respectively,

$$\bar{\lambda}_{(\ell)}^* = \Sigma_o^{-\frac{1}{2}} (\lambda_{(\ell)}^* - \mu_o), \quad (3.19)$$

$$\bar{\nabla}_{\lambda^*} \ln f(\{\hat{z}_{[k]}\} | \lambda^*) |_{\lambda_{(\ell)}^*} = \Sigma_o^{-\frac{1}{2}} \nabla_{\lambda^*} \ln f(\{\hat{z}_{[k]}\} | \lambda^*) |_{\lambda_{(\ell)}^*}, \quad (3.20)$$

in updating λ^* through a given optimization solver, e.g., gradient ascent in (3.12). ■

Example 7 (SMIB: Second-order Trajectory Sensitivities). In this example, we refer to the SMIB system from Example 1 and demonstrate how second-order trajectory sensitivities $z_{\lambda \lambda_i^*, [k]}^*$ in (3.15) are computed at time instant k . We obtain the second-order trajectory sensitivities of $f(x, y, u; \lambda)$ along the nominal trajectory $(x^*, y^*, u^*; \lambda^*)$ by using (C.6), where

$$f_{yx^*}^* = \frac{\partial f_y^*}{\partial x^*}, \quad f_{yy^*}^* = \frac{\partial f_y^*}{\partial y^*}, \quad f_{y\lambda_i^*}^* = \frac{\partial f_y^*}{\partial \lambda_i^*}, \quad (3.21)$$

$$f_{xx^*}^* = \frac{\partial f_x^*}{\partial x^*}, \quad f_{xy^*}^* = \frac{\partial f_x^*}{\partial y^*}, \quad f_{x\lambda_i^*}^* = \frac{\partial f_x^*}{\partial \lambda_i^*}, \quad (3.22)$$

$$f_{\lambda x^*}^* = \frac{\partial f_\lambda^*}{\partial x^*}, \quad f_{\lambda y^*}^* = \frac{\partial f_\lambda^*}{\partial y^*}, \quad f_{\lambda \lambda_i^*}^* = \frac{\partial f_\lambda^*}{\partial \lambda_i^*}, \quad (3.23)$$

and each derivative term above evaluates as

$$\begin{aligned} f_{y\delta^*}^* = f_{y\omega^*}^* = f_{yP^{e*}}^* = f_{yD^*}^* &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & f_{yH^*}^* &= \begin{bmatrix} 0 \\ \frac{\omega_s}{2H^{*2}} \end{bmatrix}, \\ f_{x\delta^*}^* = f_{x\omega^*}^* = f_{xP^{e*}}^* &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, & f_{xH^*}^* &= \begin{bmatrix} 0 & 0 \\ 0 & \frac{\omega_s D^*}{2H^{*2}} \end{bmatrix}, \\ f_{xD^*}^* &= \begin{bmatrix} 0 & 0 \\ 0 & -\frac{\omega_s}{2H^*} \end{bmatrix}, & f_{\lambda\delta^*}^* &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \\ f_{\lambda P^{e*}}^* &= \begin{bmatrix} 0 & 0 \\ \frac{\omega_s}{2H^{*2}} & 0 \end{bmatrix}, & f_{\lambda\omega^*}^* &= \begin{bmatrix} 0 & 0 \\ \frac{\omega_s D^*}{2H^{*2}} & -\frac{\omega_s}{2H^*} \end{bmatrix}, \\ f_{\lambda D^*}^* &= \begin{bmatrix} 0 & 0 \\ \frac{\omega_s(\omega^* - \omega_s)}{2H^{*2}} & 0 \end{bmatrix}, & f_{\lambda H^*}^* &= \begin{bmatrix} 0 & 0 \\ \frac{\omega_s(P^{m*} - D^*(\omega^* - \omega_s) - P^{e*})}{H^{*3}} & \frac{\omega_s(\omega^* - \omega_s)}{2H^{*2}} \end{bmatrix}. \end{aligned}$$

We obtain the following second-order trajectory sensitivities of $g(x, y)$ along the nominal trajectory $(x^*, y^*, u^*; \lambda^*)$ by using (C.7), where

$$g_{xx^*}^* = \frac{\partial g_x^*}{\partial x^*}, \quad g_{xy^*}^* = \frac{\partial g_x^*}{\partial y^*}, \quad g_{x\lambda_i^*}^* = \frac{\partial g_x^*}{\partial \lambda_i^*}, \quad (3.24)$$

$$g_{yx^*}^* = \frac{\partial g_y^*}{\partial x^*}, \quad g_{yy^*}^* = \frac{\partial g_y^*}{\partial y^*}, \quad g_{y\lambda_i^*}^* = \frac{\partial g_y^*}{\partial \lambda_i^*}, \quad (3.25)$$

$$g_{\lambda x^*}^* = \frac{\partial g_\lambda^*}{\partial x^*}, \quad g_{\lambda y^*}^* = \frac{\partial g_\lambda^*}{\partial y^*}, \quad g_{\lambda\lambda_i^*}^* = \frac{\partial g_\lambda^*}{\partial \lambda_i^*}, \quad (3.26)$$

which evaluate as follows:

$$\begin{aligned} g_{x\delta^*}^* &= \left[\frac{EV_\infty}{X} \sin \delta^* \quad 0 \right], & g_{x\omega^*}^* &= g_{xPe^*}^* = g_{xH^*}^* = g_{xD^*}^* = \begin{bmatrix} 0 & 0 \end{bmatrix}, \\ g_{y\delta^*}^* &= g_{y\omega^*}^* = g_{yPe^*}^* = g_{yH^*}^* = g_{yD^*}^* = 0, \\ g_{\lambda\delta^*}^* &= g_{\lambda\omega^*}^* = g_{\lambda Pe^*}^* = g_{\lambda H^*}^* = g_{\lambda D^*}^* = \begin{bmatrix} 0 & 0 \end{bmatrix}. \end{aligned}$$

We obtain the following second-order trajectory sensitivities of $h(x, y)$ along the nominal trajectory $(x^*, y^*, u^*; \lambda^*)$ by using (C.5), where

$$h_{xx^*}^* = \frac{\partial h_x^*}{\partial x^*}, \quad h_{xy^*}^* = \frac{\partial h_x^*}{\partial y^*}, \quad h_{x\lambda_i^*}^* = \frac{\partial h_x^*}{\partial \lambda_i^*}, \quad (3.27)$$

$$h_{yx^*}^* = \frac{\partial h_y^*}{\partial x^*}, \quad h_{yy^*}^* = \frac{\partial h_y^*}{\partial y^*}, \quad h_{y\lambda_i^*}^* = \frac{\partial h_y^*}{\partial \lambda_i^*}, \quad (3.28)$$

$$h_{\lambda x^*}^* = \frac{\partial h_\lambda^*}{\partial x^*}, \quad h_{\lambda y^*}^* = \frac{\partial h_\lambda^*}{\partial y^*}, \quad h_{\lambda\lambda_i^*}^* = \frac{\partial h_\lambda^*}{\partial \lambda_i^*}, \quad (3.29)$$

are evaluated as follows:

$$\begin{aligned} h_{x\delta^*}^* &= h_{x\omega^*}^* = h_{xPe^*}^* = h_{xH^*}^* = h_{xD^*}^* = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \\ h_{y\delta^*}^* &= h_{y\omega^*}^* = h_{yPe^*}^* = h_{yH^*}^* = h_{yD^*}^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\ h_{\lambda\delta^*}^* &= h_{\lambda\omega^*}^* = h_{\lambda Pe^*}^* = h_{\lambda H^*}^* = h_{\lambda D^*}^* = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

We then substitute these derivatives into (C.9)–(C.10) and obtain the second-order sensitivities of the nominal trajectories with respect to λ_i^* collected in $z_{\lambda_i^*}^*$. ■

Example 8 (SMIB: Optimization Algorithm to Compute λ_{opt}^*). Continuing from Example 7 above, we demonstrate the application of the optimization algorithm described in Section 3.2. We further visualize the results. Once we compute μ_{π, λ^*} and Σ_{π, λ^*} using (3.4) and (3.5), respectively, as explained in Example 6, we choose an arbitrary $\lambda = 1.25\lambda^{\text{true}}$ and evaluate the objective function in (3.11) and the gradient in (3.13). MATLAB native `fminunc` function uses the negative objective function from (3.11), the choice of $\bar{\lambda}_{(\ell)}^*$ from (3.19), and the normalized gradient from (3.20) to compute the next $\bar{\lambda}_{(\ell+1)}^*$. We quit iterating once the stopping criterion $\|\lambda_{(\ell+1)}^* - \lambda_{(\ell)}^*\| < 10^{-6}$ is met.

The algorithm converges to the optimal nominal parameter value of λ_{opt}^* after 17 iterations, and in even fewer iterations if the termination criterion is relaxed. In Fig. 3.2a, we plot the trajectory taken by $\lambda_{(\ell)}^*$ to reach the optimal value λ_{opt}^* . Note that we do not expect convergence to true parameter values listed in Table 2.1 due to inherent model discrepancy [30] as the optimization routine explores the space of approximate linearized models whereas the measurement data arise from a nonlinear system. In Fig. 3.2b, we plot the convergence of objective function towards its maximum value. Its value does not change much in the last 5 iterations as we approach λ_{opt}^* . In Fig. 3.2c, we plot the gradient vector approaching zero as we reach the maximum of the objective

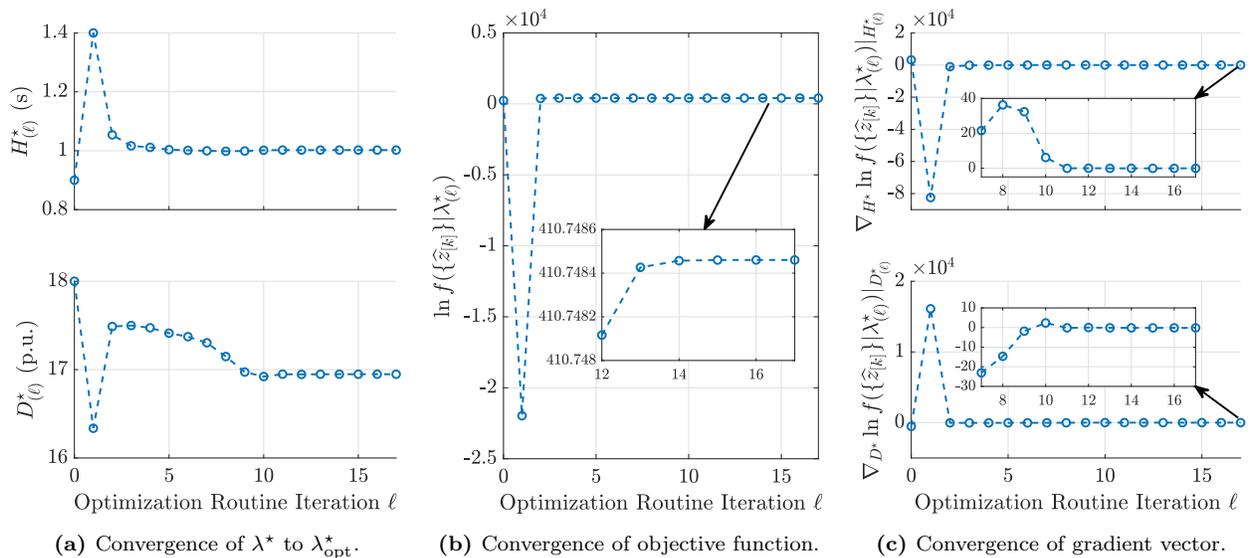


Figure 3.2: Example 8: Optimization trajectory for nominal parameter value λ^* .

function, which is the expected behaviour at an extremum. In Fig. 3.3a, we plot the trajectory taken by μ_{π, λ^*} and its associated Σ_{π, λ^*} . The mean value converges towards the true parameter value λ^{true} . The diagonal insets in Fig. 3.3b show the marginal PDF for λ posterior with $\lambda^* = \lambda_{\text{opt}}^*$. The bottom left inset shows the pairwise marginal PDF for λ posterior with $\lambda^* = \lambda_{\text{opt}}^*$. In this example, the acquired measurements lead us to conclude that the inferred parameters H and D are nearly uncorrelated. In Chapter 4, we will present numerical results that demonstrate various levels of correlation amounts different parameter pairs in a larger test system with higher-order dynamic models. ■

Example 9 (SMIB: Effect of Different Number of Measurements and Noise Level on Posterior). Here, we repeat the simulation setup in Example 8 with different number of measurements M and observe the corresponding effect on parameter inference. We run a time-domain simulation for $t = 1, 2,$ and 4 s to get $M = 30, 60,$ and $120,$ respectively. We run the optimization algorithm for these three cases and show the pairwise marginal and marginal posterior PDFs for λ in Fig. 3.4a. Changing M does not have a significant impact on the posterior mean or covariance as all three posterior marginal PDFs share similar characteristics.

Next, we repeat the simulation setup in Example 8 with different noise levels and observe their effect on parameter inference. Let $\xi_o \sim \mathcal{N}(0_m, \Sigma_\xi)$ denote the reference noise from Example 6. We reduce the standard deviation of noise to half, i.e., $\xi = 0.5\xi_o \sim \mathcal{N}(0_m, 0.25\Sigma_\xi)$ and run the optimization algorithm. Then we double the noise level, i.e., $\xi = 2\xi_o \sim \mathcal{N}(0_m, 4\Sigma_\xi)$ and repeat the

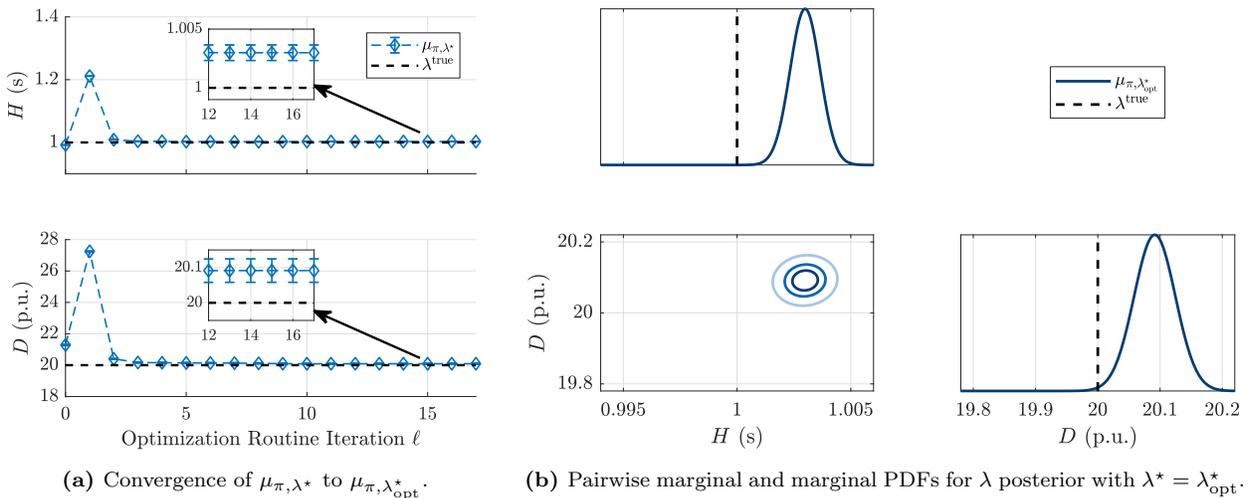


Figure 3.3: Example 8: Optimization trajectory for λ posterior.

process. We plot the pairwise marginal and marginal PDFs for the λ posterior in Fig. 3.4b. The uncertainty associated with the inferred parameters increases as the noise level increases. ■

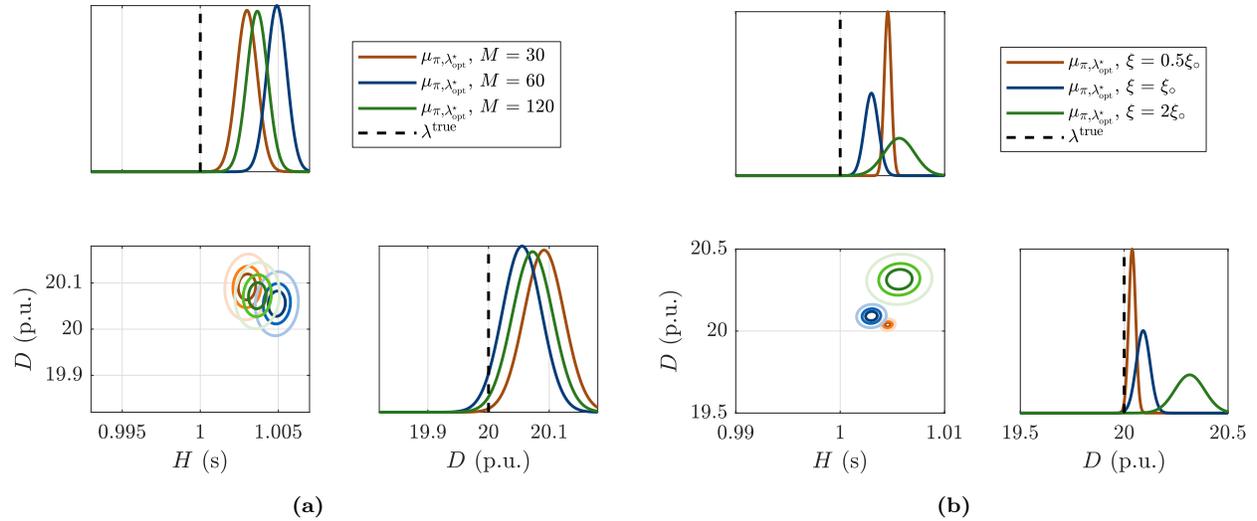


Figure 3.4: Example 9: Pairwise marginal PDFs and marginal PDFs for λ posterior for different (a) number of measurements M and (b) noise level ξ .

Chapter 4

Case Studies

This chapter demonstrates the effectiveness of the proposed Bayesian inference approach along with the modelling framework via numerical case studies involving the 3-bus test system from Example 3, the WECC 9-bus and the NE 39-bus test system (see, e.g., [26]). The 3-bus case study details results of the proposed approach, and the NE case study demonstrates scalability. Time-domain simulations of the DAE model in (2.1)–(2.3) that include dynamics arising from two-axis generators, governors, and exciters are performed using PSAT [26]. Synthetic measurements are collected from the simulation at discrete intervals of $\Delta t = 0.0333$ s, within the capability of current measurement technology [4]. In accordance with [28, 29], we assume that measurements of bus-voltage magnitudes, rotor frequency, and active- and reactive-power injections are subject to additive Gaussian noise with 0.05%, 0.01%, and 0.1% standard deviation, respectively, and all with zero mean.

4.1 3-Bus Test System

For each generator $i = 1, 2$ in the system shown in Fig. 2.2, we infer its inertia constant H_i , damping constant D_i , droop constant R_{Di} , and governor time constant T_{CHi} . Suppose the load at bus 3 increases from 2.35 p.u. to 2.85 p.u. just after time $t = 0$ s and then decreases to 2.1 p.u. at time $t = 4$ s. Measurements are acquired from $t = 0$ s to $t = 8$ s. We assume measurements of generator terminal voltage magnitude V_i , rotor speed ω_i , and active- and reactive-power injections P_i and Q_i are available at: i) bus 1 only with output vector $z = [V_1, \omega_1, P_1, Q_1]^T$ ($m = 4$), and ii) both buses 1 and 2 with output vector $z = [V_1, V_2, \omega_1, \omega_2, P_1, P_2, Q_1, Q_2]^T$ ($m = 8$).

4.1.1 Choice of λ^*

We minimize the negative of the objective in (3.9) with the MATLAB native `fminunc` function, which uses a quasi-Newton method. The initial λ^* is set as $\lambda_{(0)}^* = 0.9\lambda^{\text{true}}$, where λ^{true} denotes the true measurement-generating parameter values in the nonlinear DAE. In each iteration ℓ , the optimization routine provides an updated candidate $\lambda_{(\ell)}^*$, with which we perform time-domain simulation of (2.1)–(2.3), (2.13)–(2.14), and (C.9)–(C.10). The value of the objective is evaluated analytically in each iteration ℓ via (3.11) in conjunction with the posterior mean and covariance in (3.4) and (3.5), respectively, as well as the approximate output \tilde{z} resulting from the simulation. Also, instead of relying on `fminunc` to approximate the gradient vector via numerical finite differences, we compute its value analytically using (3.13) along with pertinent trajectories and trajectory sensitivities provided by the simulation. We then pass the gradient vector to `fminunc` so as to inform the direction for the next update $\lambda_{(\ell+1)}^*$. The optimization routine searches for the optimizer until the stopping criterion $\|\lambda_{(\ell+1)}^* - \lambda_{(\ell)}^*\| < 10^{-6}$, and at this point, we return $\lambda_{\text{opt}}^* = \lambda_{(\ell+1)}^*$. Both measurement scenarios converge within 24 iterations, and in even fewer iterations if the termination criterion is relaxed. Fig. 4.1 shows updates in nominal parameters. Note that we do not expect convergence to true parameter values listed in Table A.1 due to inherent model discrepancy [30] as the optimization routine explores the space of approximate linearized models whereas the measurement data arise from a nonlinear system. Fig. 4.2 shows the analytical gradient values in (3.13), which are close to zero. Although they are non-zero at λ_{opt}^* , the change in parameters is small enough to warrant exiting the optimization algorithm. In the top-right inset of Fig. 4.3b, we plot the convergence of the objective function in (3.9) for the two measurement scenarios. The objective function increases very little, indicating we are close to a maximum.

4.1.2 Inference on λ

We prescribe Gaussian prior $\lambda \sim \mathcal{N}(\mu_{\circ}, \Sigma_{\circ})$ regardless of the value taken by $\lambda_{(\ell)}^*$, where $\mu_{\circ} = 1.5\lambda^{\text{true}}$, $\Sigma_{\circ} = 0.5^2 \text{diag}(\lambda^{\text{true}})^2$. In Fig. 4.3a, we plot pairwise prior marginal PDF contours of λ and the marginal PDF of each parameter to be inferred; there is no correlation between the parameters. The time-domain simulation of (2.1)–(2.3) and (2.13)–(2.14) with each updated candidate $\lambda_{(\ell)}^*$ yields trajectories of $a(\lambda_{(\ell)}^*)$ and $b(\lambda_{(\ell)}^*)$, with which we construct the approximate output $\tilde{z}(\lambda, \lambda_{(\ell)}^*)$ in (2.19). We then take the value of \tilde{z} at each time instant $k = 0, \dots, M$ to compute the

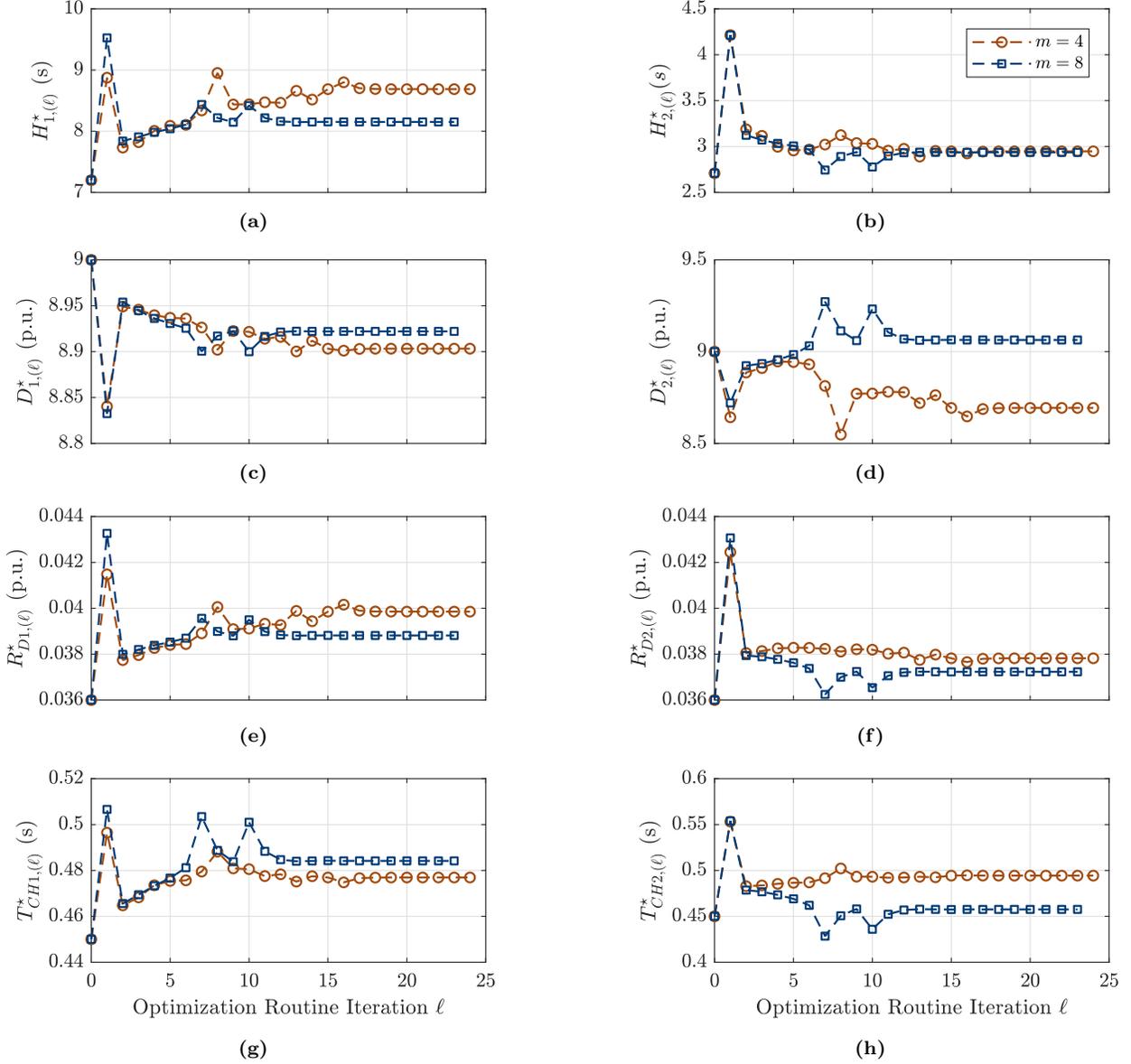


Figure 4.1: 3-bus test system: convergence of λ^* to λ_{opt}^*

mean $\mu_{\pi, \lambda_{(\ell)}^*}$ and covariance $\Sigma_{\pi, \lambda_{(\ell)}^*}$ of the posterior in closed form via (3.4) and (3.5), respectively. In Fig. 4.3b, we plot pairwise posterior marginal PDF contours of λ and the marginal PDF of each inferred parameter, with $\lambda^* = \lambda_{\text{opt}}^*$, for the two measurement scenarios. We observe that the $m = 8$ scenario yields narrower posteriors (i.e., lower uncertainty), which is expected since the measurements for the $m = 8$ scenario contain those for the $m = 4$. The true parameter values are within 3 standard deviations of $\mu_{\pi, \lambda_{\text{opt}}^*}$ for all parameters in both cases. We also observe different degrees of correlation amongst different parameter pairs. For example, $\{R_{D1}, D_1\}$, $\{R_{D2}, D_2\}$,

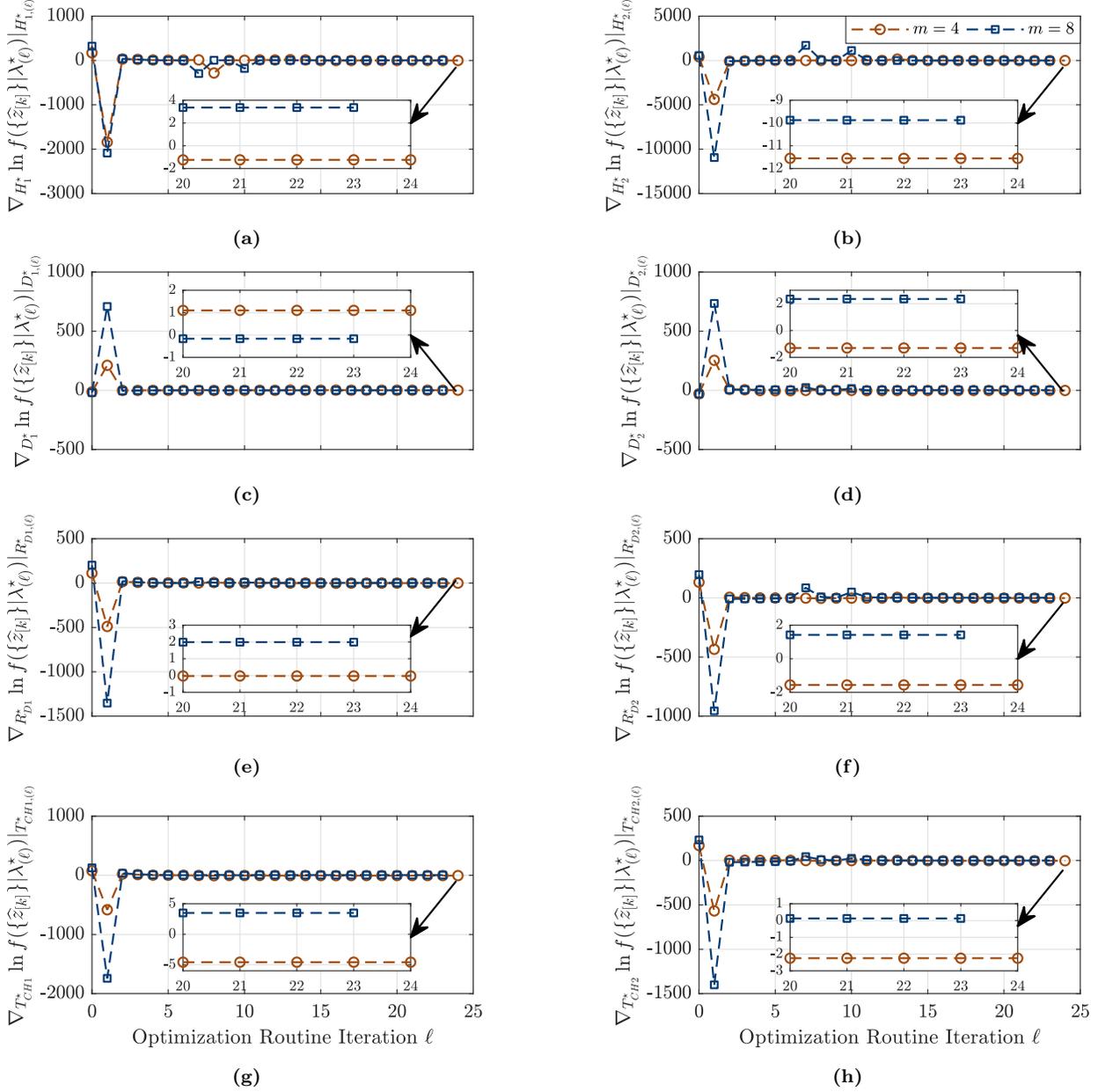
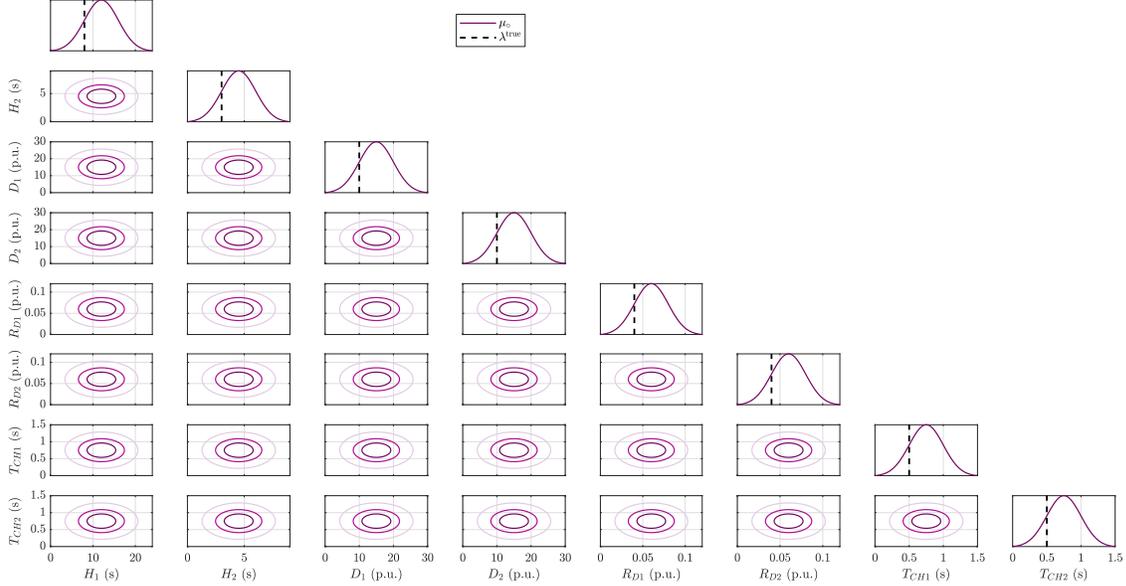
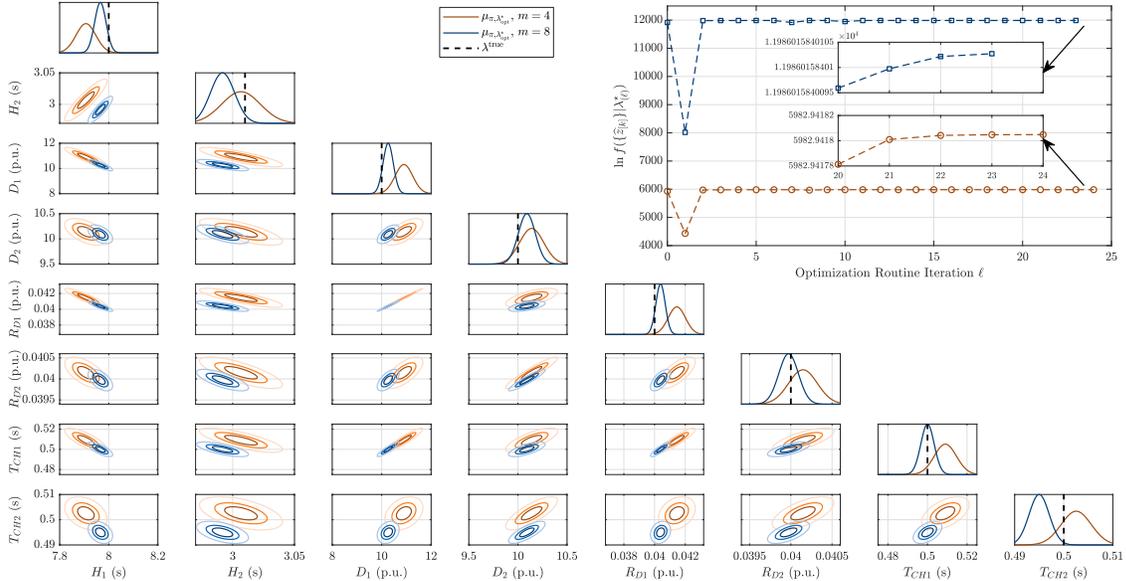


Figure 4.2: 3-bus test system: convergence of gradient vector $\nabla_{\lambda^*} \ln f(\{\hat{z}_{[k]}\} | \lambda^*) |_{\lambda_{(\ell)}^*}$.

$\{T_{CH1}, D_1\}$, and $\{T_{CH1}, R_{D1}\}$ appear to have strong positive correlations; $\{D_1, H_1\}$ and $\{R_{D1}, H_1\}$ carry strong negative correlations; and $\{T_{CH2}, H_1\}$, $\{T_{CH2}, D_1\}$, and $\{T_{CH2}, R_{D1}\}$ are nearly uncorrelated. When compared to the prior standard deviations, the posterior marginal PDFs achieve 2 to 3 orders of magnitude reduction of uncertainty for both the $m = 4$ and $m = 8$ scenarios as a result of inference from the noisy measurements. We also note that some true parameter values are quite close to the centres of the corresponding posterior marginal Gaussian distributions, but



(a) Lower left triangle: pairwise marginal PDFs for λ prior with $\mu_o = 1.5\lambda^{\text{true}}$; Diagonal: marginal PDF for λ prior with $\mu_o = 1.5\lambda^{\text{true}}$.



(b) Orange-coloured traces correspond to measurement scenario i) with $z = [V_1, \omega_1, P_1, Q_1]^T$ ($m = 4$), and blue-coloured traces correspond to scenario ii) with $z = [V_1, V_2, \omega_1, \omega_2, P_1, P_2, Q_1, Q_2]^T$ ($m = 8$). Upper right inset: demonstrating convergence of objective function; Lower left triangle: pairwise marginal PDFs for λ posterior with $\lambda^* = \lambda_{\text{opt}}^*$; Diagonal: marginal PDF for λ posterior with $\lambda^* = \lambda_{\text{opt}}^*$.

Figure 4.3: 3-bus test system: pairwise marginal PDFs and convergence of objective function.

others appear further away. This is unsurprising due to two factors: i) inference is conducted with only a finite number of noisy measurements, and ii) there exists inherent model discrepancy [30,31] with the approximate linearized model.

To further demonstrate the accuracy of the inferred parameter values, we sample λ from the

resulting Gaussian posterior $\mathcal{N}(\mu_{\pi, \lambda_{\text{opt}}^*}, \Sigma_{\pi, \lambda_{\text{opt}}^*})$ and perform time-domain simulation of (2.1)–(2.3) 100 times for the $m = 4$ scenario. We plot the trajectories resulting from each set of parameters sampled from the posterior PDF along with the actual output induced by the true parameter values as well as the recorded noisy measurement data in Fig. 4.4. The simulated output generated from the posterior closely matches the actual output without the noise, providing strong evidence that the inferred parameters are accurate in that they could have induced the measurement data.

Remark 3 (Computational Burden). In each iteration ℓ , with the updated nominal parameter value $\lambda_{(\ell)}^*$, the proposed Bayesian framework performs a *single* time-domain simulation of (2.1)–(2.3) and (2.13)–(2.14) to obtain the approximate linearized model around the nominal trajectory induced by $\lambda_{(\ell)}^*$, with which we compute the posterior mean and covariance in closed form. In this way, we completely bypass MCMC sampling of the nonlinear DAE model or surrogates thereof. ■

4.2 WECC 9-Bus Test System

For each generator $i = 1, 2, 3$ in the WECC 9-bus system shown in Fig. 4.5, we infer parameters H_i , D_i , R_{Di} , and T_{CHi} assuming measurements V_i , ω_i , P_i , and Q_i are available for all generators.

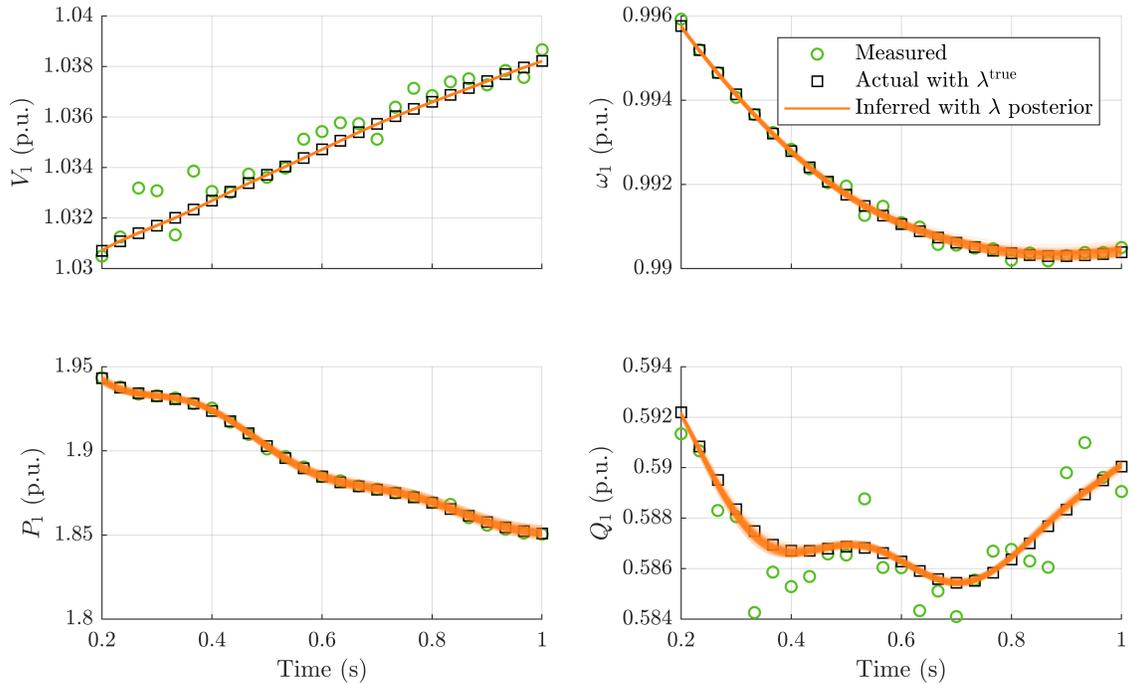


Figure 4.4: 3-bus test system: comparison amongst output trajectories that are obtained from noisy measurements, simulation of nonlinear DAE model with the true parameter values, and simulations of the nonlinear DAE model with parameters sampled from the posterior PDF.

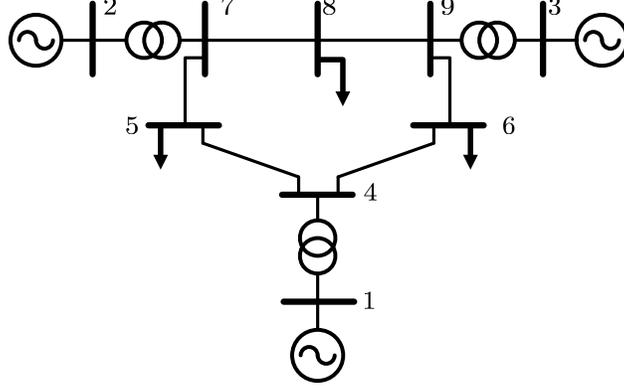


Figure 4.5: One-line diagram for WECC 9-bus test system.

A uniform increase in all loads of 20% is applied just after $t = 0$ s, and synthetic measurements are acquired from $t = 0$ s to $t = 3$ s. We utilize the same procedure as detailed in Section 4.1 to obtain the optimal nominal parameter value of λ_{opt}^* after 52 iterations. Given λ_{opt}^* and Gaussian prior $\mathcal{N}(1.5\lambda^{\text{true}}, 0.5^2 \text{diag}(\lambda^{\text{true}})^2)$, we compute the Gaussian posterior $\mathcal{N}(\mu_{\pi, \lambda_{\text{opt}}^*}, \Sigma_{\pi, \lambda_{\text{opt}}^*})$ via (3.4)–(3.5). In Table 4.1, we report the true measurement-generating parameter values and the mean of the posterior marginal PDF corresponding to each parameter.

Fig. 4.6 shows the nominal parameters λ^* converging to λ_{opt}^* . The sudden change in parameter

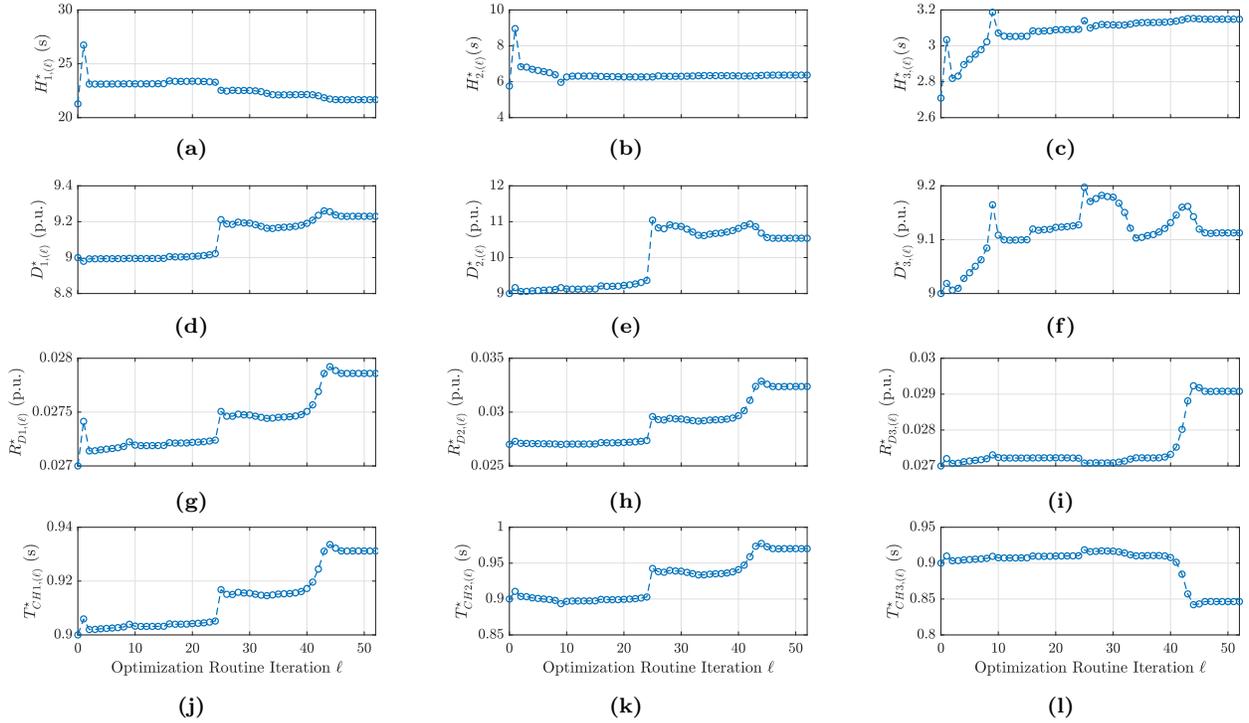


Figure 4.6: WECC 9-bus test system: convergence of λ^* to λ_{opt}^* .

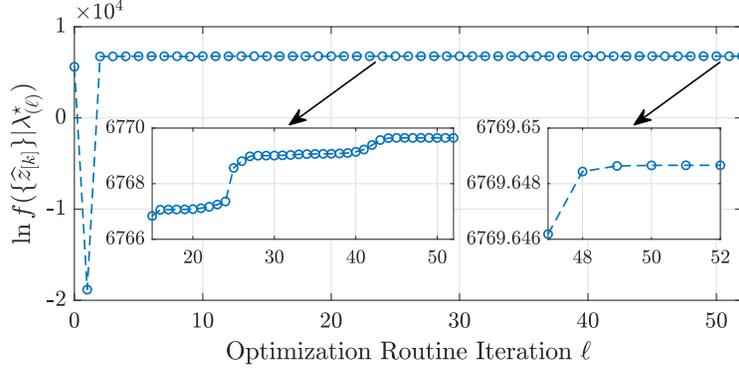


Figure 4.7: WECC 9-bus test system: convergence of objective function.

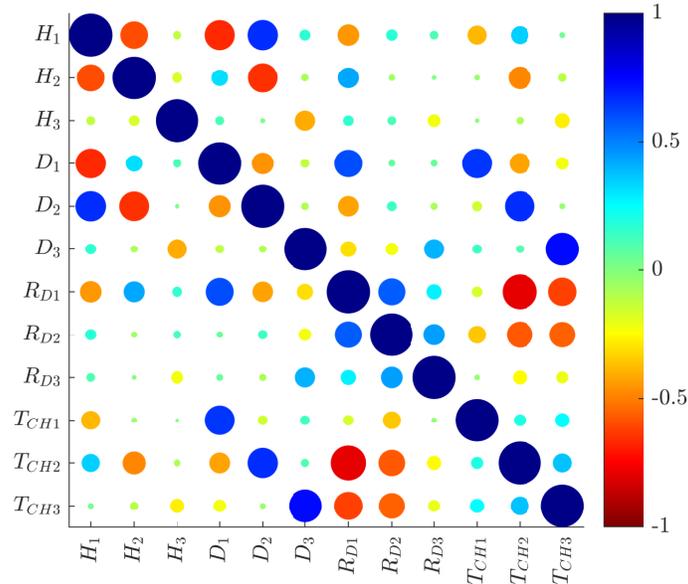


Figure 4.8: WECC 9-bus test system: pairwise parameter correlation.

values correspond to the increase in in the objective function in Fig. 4.7. In Fig. 4.8, we make use of the posterior covariance matrix $\Sigma_{\pi, \lambda^*_{\text{opt}}}$ to visualize the degree of pairwise correlation between different parameters. While most parameter pairs are weakly correlated, we observe strong positive correlation for $\{D_2, H_1\}$, $\{R_{D1}, D_1\}$, and $\{T_{CHi}, D_i\}$ for each generator i ; and strong negative correlation for $\{D_1, H_1\}$, $\{D_2, H_2\}$, $\{T_{CH2}, R_{D1}\}$, and $\{T_{CH3}, R_{D1}\}$. In Fig. 4.9a, we plot a histogram

Table 4.1: WECC 9-bus test system: true parameter values λ^{true} and posterior mean $\mu_{\pi, \lambda^*_{\text{opt}}}$.

Generator i	H_i (s)		D_i (p.u.)		R_{Di} (p.u.)		T_{CHi} (s)	
	λ^{true}	$\mu_{\pi, \lambda^*_{\text{opt}}}$						
1	23.64	23.5706	10	9.7951	0.03	0.0300	1	0.9528
2	6.4	6.3908	10	10.3993	0.03	0.0298	1	1.0437
3	3.01	3.0046	10	10.2817	0.03	0.0305	1	0.9755

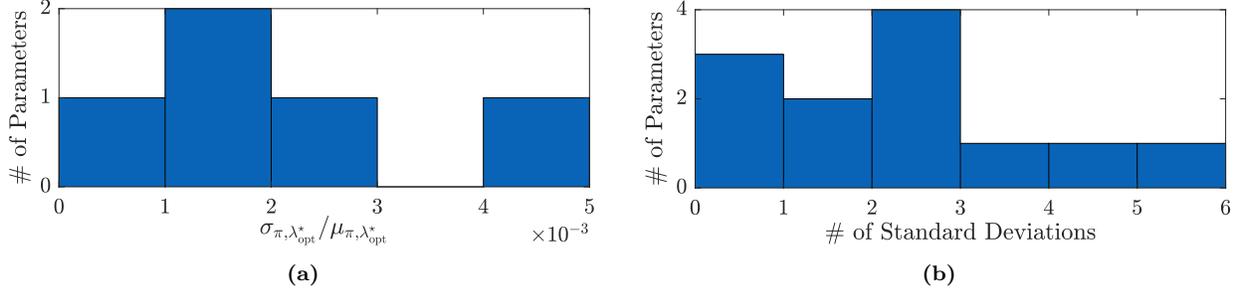


Figure 4.9: WECC 9-bus test system: histograms illustrate the number of parameters out of a total of 12 for which (a) normalized posterior marginal standard deviations and (b) the number of standard deviations between the posterior mean and the respective true parameter values, lie within certain ranges.

of the standard deviations of the posterior marginal PDFs normalized with respect to the corresponding mean values, demonstrating low posterior uncertainty. Also, Fig. 4.9b shows a histogram of the number of standard deviations between the posterior mean and the respective true parameter values. We observe reasonably good agreement between the posterior coverage compared to λ^{true} , again keeping in mind finite noisy observations as well as the linearized model approximation.

4.3 New England 39-Bus Test System

For each generator $i = 1, \dots, 10$ in the NE test system shown in Fig. 4.10, we infer parameters H_i , D_i , R_{Di} , and T_{CHi} assuming measurements V_i , ω_i , P_i , and Q_i are available for all generators. A uniform increase in all loads of 10% is applied just after $t = 0$ s, and synthetic measurements are acquired from $t = 0$ s to $t = 2$ s. We utilize the same procedure as detailed in Section 4.1 to obtain the optimal nominal parameter value of λ_{opt}^* after 11 iterations. Given λ_{opt}^* and Gaussian prior

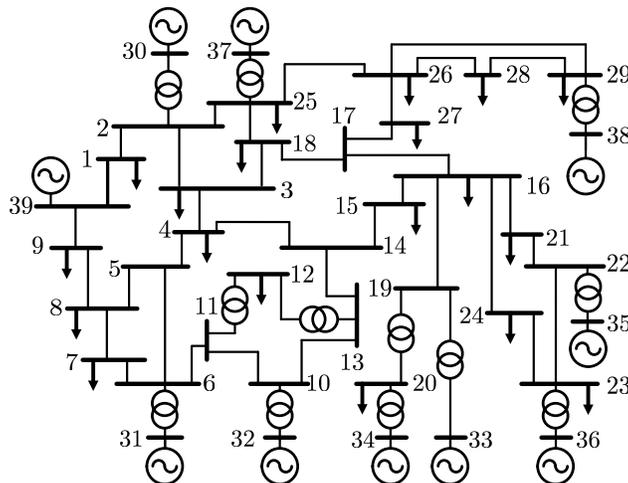


Figure 4.10: One-line diagram for NE 39-bus test system.

Table 4.2: NE 39-bus test system: true parameter values λ^{true} and posterior mean $\mu_{\pi, \lambda_{\text{opt}}^*}$.

Generator i	H_i (s)		D_i (p.u.)		R_{Di} (p.u.)		T_{CHi} (s)	
	λ^{true}	$\mu_{\pi, \lambda_{\text{opt}}^*}$						
1	0.1114	0.11095	10	10.002	0.04	0.039976	0.5	0.49960
2	0.08035	0.07964	10	9.9964	0.04	0.039984	0.5	0.50088
3	0.09495	0.093523	10	10.014	0.04	0.040001	0.5	0.50166
4	0.07585	0.075475	10	10.003	0.04	0.039990	0.5	0.50073
5	0.06895	0.068751	10	9.9825	0.04	0.039936	0.5	0.50175
6	0.0923	0.091737	10	10.013	0.04	0.039985	0.5	0.50198
7	0.07005	0.06902	10	9.9844	0.04	0.040013	0.5	0.49974
8	0.06445	0.064233	10	9.9774	0.04	0.039948	0.5	0.50009
9	0.0915	0.090849	10	9.9861	0.04	0.039931	0.5	0.50065
10	1.3263	1.3255	10	10.045	0.04	0.040058	0.5	0.49558

$\mathcal{N}(1.5\lambda^{\text{true}}, 0.5^2 \text{diag}(\lambda^{\text{true}})^2)$, we compute the Gaussian posterior $\mathcal{N}(\mu_{\pi, \lambda_{\text{opt}}^*}, \Sigma_{\pi, \lambda_{\text{opt}}^*})$ via (3.4)–(3.5). In Table 4.2, we report the true measurement-generating parameter values and the mean of the posterior marginal PDF corresponding to each parameter. Fig. 4.11 shows the objective function increases by a very small value when λ^* converges to λ_{opt}^* .

In Fig. 4.12a, we plot a histogram of the standard deviations of the posterior marginal PDFs normalized with respect to the corresponding mean values, demonstrating low posterior uncertainty. Also, Fig. 4.12b shows a histogram of the number of standard deviations between the posterior mean and the respective true parameter values. We observe reasonably good agreement between the posterior coverage compared to λ^{true} , again keeping in mind finite noisy observations as well as the linearized model approximation. In Fig. 4.13, we make use of the posterior covariance matrix $\Sigma_{\pi, \lambda_{\text{opt}}^*}$ to visualize the degree of pairwise correlation between different parameters. While most parameter pairs are weakly correlated, we observe strong positive correlation for $\{D_i, T_{CHi}\}$

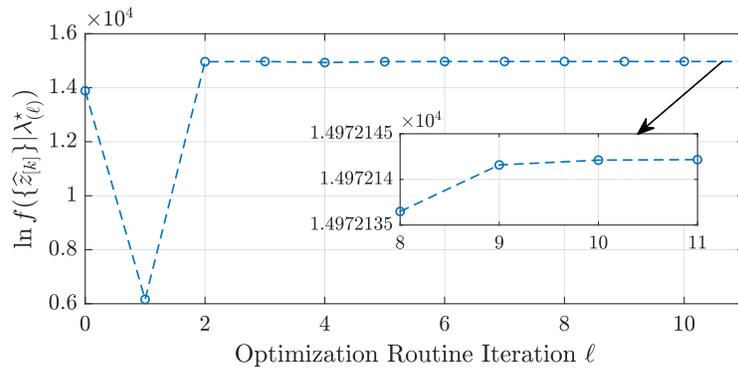


Figure 4.11: NE 39-bus test system: convergence of objective function.

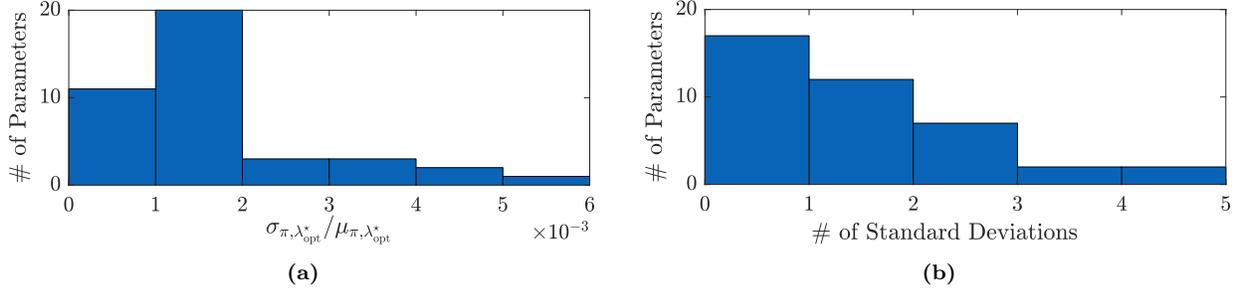


Figure 4.12: NE 39-bus test system: histograms illustrate the number of parameters out of a total of 40 for which (a) normalized posterior marginal standard deviations and (b) the number of standard deviations between the posterior mean and the respective true parameter values, lie within certain ranges.

and $\{D_i, R_{Di}\}$ for each generator i . Strongly correlated parameters are often associated with identifiability challenges (e.g., both parameters can simultaneously increase/decrease and still appear plausible as the measurement-generating setting). In our problem context, it may suggest the need for additional measurement data or observation under different disturbance scenarios.

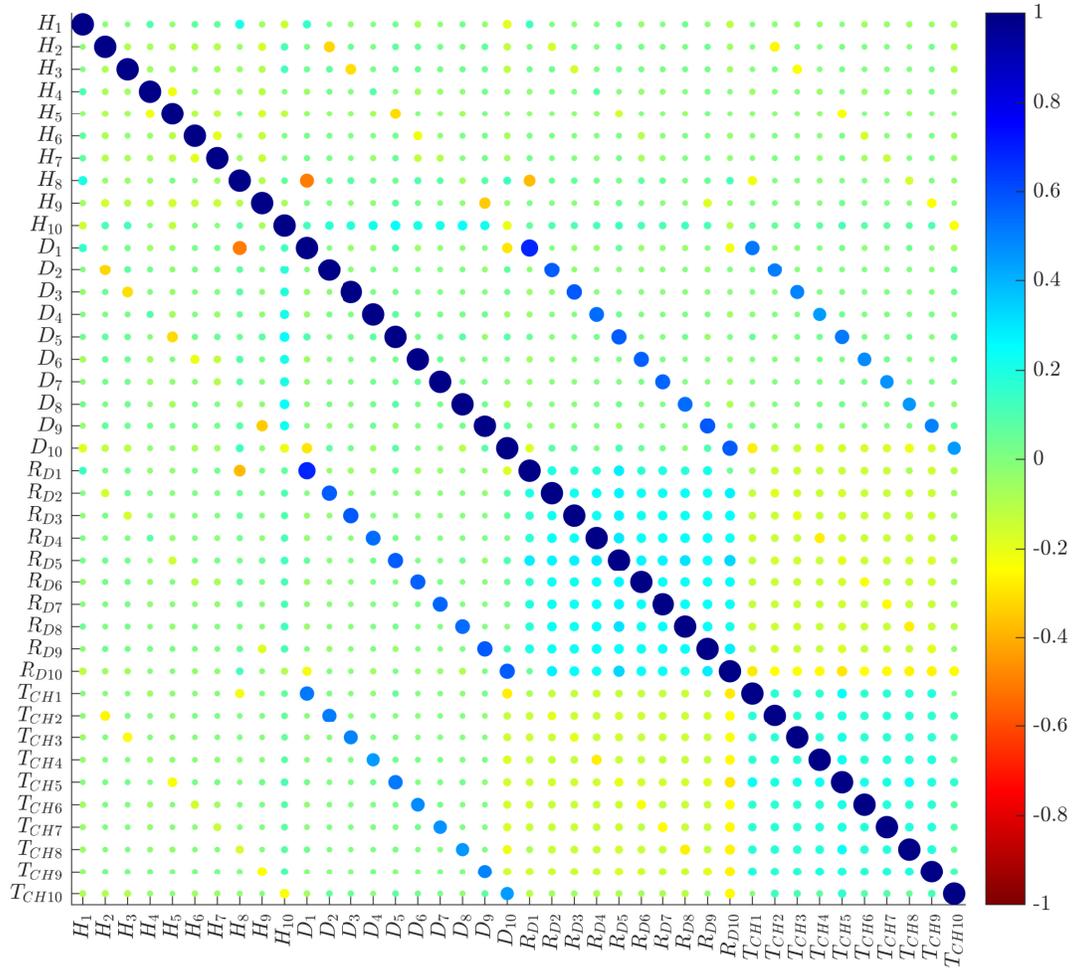


Figure 4.13: NE 39-bus test system: pairwise parameter correlation.

Chapter 5

Conclusions

This thesis proposes a method to find accurate power system parameters for model calibration. In particular, we focus on parameter inference, where we compute the entire probability distribution of parameter values that could have induced the measurement data. The proposed analytically tractable Bayesian framework infers dynamic power system parameters conditioned on noisy measurements obtained at generator terminals. In Chapter 2, we outlined pertinent models for the power system and noisy measurements, and also described the inference problem. In Chapter 3, we found the best linearized model by formulating the optimal model selection problem, computing the value of this metric for a given candidate model, and proposing new candidate models within the optimization routine. The DAE modeling, first- and second-order trajectory sensitivities, and Bayesian concepts were demonstrated with SMIB examples throughout Chapter 2 and Chapter 3. In Chapter 4, we applied the proposed Bayesian framework to two benchmark test systems—the WECC 9-bus 3-machine and the NE 39-bus 10-machine systems—to demonstrate the effectiveness and scalability of the proposed method.

Existing parameter inference methods in the literature, such as MCMC and PCE surrogate modelling, are computationally impractical for large-scale power systems models consisting of many nonlinear DAEs. Our work advances over the previous by considering the full DAE model describing generator dynamics coupled across the transmission network. The explicit inclusion of the network enables parameter inference at buses without sufficient local measurements. We demonstrated this aspect and other key features along with algorithmic scalability with case studies involving a SMIB, a 3-bus system, the WECC 9-bus, and the NE 39-bus test system. The posterior marginal PDFs

achieve 2 to 3 orders of magnitude reduction of uncertainty when compared to the prior standard deviations as a result of inference from the noisy measurements.

We constructed an optimal linearized model that maximizes the probability of measurement data amongst candidate linear models. The linear models enable closed-form evaluations of the posterior, model evidence, and their gradients, so the proposed method scales Bayesian inference to higher dimensions than PCEs. Although the model evidence is generally difficult to estimate, we computed its value analytically in the space of linearized models along with Gaussian conjugate priors. We also derived the analytical gradient of model evidence along the trajectory of linearization, so as to facilitate gradient-based optimization methods to find the evidence-maximizing linear model. The posteriors for all linear models considered in the procedure, including the Bayesian-optimal one, are Gaussian due to conjugacy, where the mean and covariance can be obtained in closed form.

Bibliography

- [1] “Power system model validation,” NERC, Princeton, NJ, USA, Dec. 2010. [Online]. Available: https://www.nerc.com/comm/PC/Model%20Validation%20Working%20Group%20MVWG/MV%20White%20Paper_Final.pdf (→) *page1*
- [2] *Verification of Models and Data for Generator Excitation Control System or Plant Volt/Var Control Functions*, NERC Standard MOD-026-1, Nov. 2014. [Online]. Available: <https://www.nerc.com/pa/Stand/Reliability%20Standards/MOD-026-1.pdf> (→) *page1*
- [3] *Verification of Models and Data for Turbine/Governor and Load Control or Active Power/Frequency Control Functions*, NERC Standard MOD-027-1, Nov. 2014. [Online]. Available: <https://www.nerc.com/pa/Stand/Reliability%20Standards/MOD-027-1.pdf> (→) *page1*
- [4] US DOE & FERC. (2006, Feb.) Steps to establish a real-time transmission monitoring system for transmission owners and operators within the eastern and western interconnections. [Online]. Available: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/final_1839.pdf (→) *pages1, 26*
- [5] K. Bollinger, H. Khalil, L. Li, and W. Norum, “A method for on-line identification of power system model parameters in the presence of noise,” *IEEE Trans. Power App. Syst.*, no. 9, pp. 3105–3111, 1982. (→) *page2*
- [6] S. Guo, S. Norris, and J. Bialek, “Adaptive parameter estimation of power system dynamic model using modal information,” *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2854–2861, 2014. (→) *page2*
- [7] M. Shen, V. Venkatasubramanian, N. Abi-Samra, and D. Sobajic, “A new framework for estimation of generator dynamic parameters,” *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 756–763, 2000. (→) *page2*
- [8] I. A. Hiskens, “Nonlinear dynamic model evaluation from disturbance measurements,” *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 702–710, 2001. (→) *pages2, 7*
- [9] D. S. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*, 2nd ed. New York, NY: Oxford University Press, 2006. (→) *page2*
- [10] U. Von Toussaint, “Bayesian inference in physics,” *Reviews of Modern Physics*, vol. 83, pp. 943–999, 2011. (→) *page2*
- [11] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. New York, NY: Chapman & Hall, 1996. (→) *page2*

- [12] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, Eds., *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011. (→) *page2*
- [13] I. A. Hiskens and M. Pai, “Trajectory sensitivity analysis of hybrid systems,” *IEEE Trans. Circuits Syst. I. Fundam. Theory Appl.*, vol. 47, no. 2, pp. 204–220, 2000. (→) *pages2, 4*
- [14] R. E. Kass and A. E. Raftery, “Bayes Factor,” *Journal of American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995. (→) *pages2, 18*
- [15] L. Wasserman, “Bayesian model selection and model averaging,” *Journal of Mathematical Psychology*, vol. 44, pp. 92–107, 2000. (→) *pages2, 18*
- [16] N. Friel and J. Wyse, “Estimating the evidence - a review,” *Statistica Neerlandica*, vol. 66, no. 3, pp. 288–308, 2012. (→) *page2*
- [17] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation*, in Texts in Applied Mathematics, vol. 62. Cham, Switzerland: Springer International Publishing, 2015. (→) *page3*
- [18] Z. Huang, P. Du, D. Kosterev, and S. Yang, “Generator dynamic model validation and parameter calibration using phasor measurements at the point of connection,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1939–1949, 2013. (→) *page3*
- [19] M. Ariff, B. Pal, and A. K. Singh, “Estimating dynamic model parameters for adaptive protection and control in power system,” *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 829–839, 2014. (→) *page3*
- [20] R. Huang, R. Diao, Y. Li, J. Sanchez-Gasca, Z. Huang, B. Thomas, P. Etingov, S. Kincic, S. Wang, R. Fan *et al.*, “Calibrating parameters of power system stability models using advanced ensemble Kalman filter,” *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2895–2905, 2017. (→) *page3*
- [21] N. Petra, C. G. Petra, Z. Zhang, E. M. Constantinescu, and M. Anitescu, “A Bayesian approach for parameter estimation with uncertainty for dynamic power systems,” *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2735–2743, 2016. (→) *page3*
- [22] Y. Xu, L. Mili, X. Chen, M. Korkali, and L. Min, “A Bayesian approach to real-time dynamic parameter estimation using PMU measurement,” *IEEE Trans. Power Syst.*, 2019. (→) *page3*
- [23] Y. Xu, C. Huang, X. Chen, L. Mili, C. H. Tong, M. Korkali, and L. Min, “Response-surface-based Bayesian inference for power system dynamic parameter estimation,” *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 5899–5909, 2019. (→) *page3*
- [24] M. J. Laufenberg and M. A. Pai, “A new approach to dynamic security assessment using trajectory sensitivities,” *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 953–958, 1998. (→) *page4*
- [25] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998. (→) *pages5, 6, 9, 13*
- [26] F. Milano, “An open source power system analysis toolbox,” *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1199–1206, Aug. 2005. (→) *pages9, 13, 26*

- [27] M. Lewandowski, L. Majka, and A. Świetlicka, “Effective estimation of angular speed of synchronous generator based on stator voltage measurement,” *International Journal of Electrical Power & Energy Systems*, vol. 100, pp. 391–399, 2018. (→) *page14*
- [28] “Phasor measurement unit (PMU) datasheet,” VIZIMAX Inc., Longueuil, QC, Canada, Sept. 2017. [Online]. Available: https://blob.opal-rt.com/medias/L00161_0917.pdf (→) *pages17, 26*
- [29] “Balancing and frequency control,” NERC, Princeton, NJ, USA, Jan 2011. [Online]. Available: https://www.nerc.com/comm/PC/Model%20Validation%20Working%20Group%20MVWG/MV%20White%20Paper_Final.pdf (→) *pages17, 26*
- [30] M. C. Kennedy and A. O’Hagan, “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001. (→) *pages23, 27, 30*
- [31] J. Brynjarsdóttir and A. O’Hagan, “Learning about physical parameters: the importance of model discrepancy,” *Inverse Problems*, vol. 30, no. 11, p. 114007, Nov. 2014. (→) *page30*

Appendix A

3-Bus Test System

A.1 System Parameters

Transmission Lines. The line impedance $z_{12} = 0.01 + j0.085$, $z_{23} = 0.02 + j0.161$, $z_{13} = 0.01 + j0.092$. The shunt admittance $y_{12}^{\text{sh}} = j0.088$, $y_{23}^{\text{sh}} = j0.153$, and $y_{13}^{\text{sh}} = j0.079$.

Initial Steady State. $V_1 \angle \theta_1 = 1.04 \angle 0^\circ$, $V_2 \angle \theta_2 = 1.025 \angle -0.148^\circ$, $V_3 \angle \theta_3 = 0.994 \angle -7.65^\circ$, $S_1 = 1.597 + j0.452$, $S_2 = 0.791 - j0.279$, $S_3 = -2.35 - j0.5$.

Parameters related to the machine, turbine-governor, and excitation system are reported in Table A.1. The saturation function $S_E(E_{fd}) = 0.0039e^{1.555E_{fd}}$ for both generators.

Table A.1: Parameter values for dynamic model of 3-bus system shown in Fig. 2.2.

	Machine Data								
	T'_{d0}	T'_{q0}	H	D	X_d	X_q	X'_d	X'_q	R_s
Bus 1	8.96	0.31	8	10	0.146	0.0969	0.0608	0.0608	0
Bus 2	5.89	0.60	3.01	10	1.3125	1.2578	0.1813	0.1813	0
Unit	s	s	s	p.u.	p.u.	p.u.	p.u.	p.u.	p.u.
	Turbine/Governor Data			Excitation System Data					
	T_{CH}	T_{SV}	R_D	T_A	T_E	T_F	K_A	K_E	K_F
Bus 1	0.5	0	0.04	0.2	0.314	0.35	200	1	0.063
Bus 2	0.5	0	0.04	0.2	0.314	0.35	200	1	0.063
Unit	s	s	p.u.	s	s	s	p.u.	p.u.	p.u.

A.2 Model Details

The dynamic state variables for each machine i are

$$x_i = [E'_{qi} \quad E'_{di} \quad \delta_i \quad \omega_i \quad E_{fdi} \quad R_{fi} \quad V_{Ri} \quad T_{Mi}]^T,$$

and algebraic state variables are

$$y = [I_{d1} \quad I_{d2} \quad I_{q1} \quad I_{q2} \quad V_1 \quad V_2 \quad V_3 \quad \theta_1 \quad \theta_2 \quad \theta_3]^T,$$

and inputs are

$$u = [\omega_s \quad P_{C1} \quad P_{C2} \quad V_{\text{ref}1} \quad V_{\text{ref}2}]^T.$$

The differential equations $f(x, y, u; \lambda)$ for each generator $i = 1, 2$ are

$$\begin{aligned} T'_{d0i} \dot{E}'_{qi} &= -E'_{qi} - (X_{di} - X'_{di})I_{di} + E_{fdi}, \\ T'_{q0i} \dot{E}'_{di} &= -E'_{di} + (X_{qi} - X'_{qi})I_{qi}, \\ \dot{\delta}_i &= 120\pi(\omega_i - \omega_s), \\ \frac{2H_i}{\omega_s} \dot{\omega}_i &= T_{Mi} - (X'_{qi} - X'_{di})I_{di}I_{qi} - E'_{qi}I_{qi} - E'_{di}I_{di} - D_i(\omega_i - \omega_s), \\ T_{Ei} \dot{E}_{fdi} &= -(K_{Ei} + S_E(E_{fdi}))E_{fdi} + V_{Ri}, \\ T_{Fi} \dot{R}_{fi} &= -R_{fi} + \frac{K_{Fi}}{T_{Fi}}E_{fdi}, \\ T_{Ai} \dot{V}_{Ri} &= -V_{Ri} + K_{Ai}R_{fi} - \frac{K_{Ai}K_{Fi}}{T_{Fi}}E_{fdi} + K_{Ai}(V_{\text{ref}i} - V_i), \\ T_{CHi} \dot{T}_{Mi} &= P_{Ci} - T_{Mi} - \frac{1}{R_{Di}}(\omega_i - \omega_s). \end{aligned}$$

The algebraic constraints $g(x, y, u; \lambda)$ include the stator algebraic equations for generator $i = 1, 2$:

$$\begin{aligned} 0 &= E'_{di} - V_i \sin(\delta_i - \theta_i) - R_{si}I_{di} + X'_{qi}I_{qi}, \\ 0 &= E'_{qi} - V_i \cos(\delta_i - \theta_i) - R_{si}I_{qi} - X'_{di}I_{di}, \end{aligned}$$

the network algebraic equations for generator $i = 1, 2$:

$$0 = I_{di}V_i \sin(\delta_i - \theta_i) + I_{qi}V_i \cos(\delta_i - \theta_i) + P_i - \sum_{k=1}^3 V_i V_k Y_{ik} \cos(\theta_i - \theta_k - \alpha_{ik}),$$

$$0 = I_{di}V_i \cos(\delta_i - \theta_i) - I_{qi}V_i \sin(\delta_i - \theta_i) + Q_i - \sum_{k=1}^3 V_i V_k Y_{ik} \sin(\theta_i - \theta_k - \alpha_{ik}),$$

and the network algebraic equations for the load bus $i = 3$:

$$0 = P_i - \sum_{k=1}^3 V_i V_k Y_{ik} \cos(\theta_i - \theta_k - \alpha_{ik}),$$

$$0 = Q_i - \sum_{k=1}^3 V_i V_k Y_{ik} \sin(\theta_i - \theta_k - \alpha_{ik}),$$

where $Y = |Y_{\text{bus}}|$, $\alpha = \angle Y_{\text{bus}}$ and Y_{bus} is the admittance matrix.

Appendix B

Bayesian Conjugacy for Linear-Gaussian Models

In Appendices B.1 and B.2, we simplify the notation and use $a_{[k]}$, $b_{[k]}$, μ_π , and Σ_π , but with the understanding that they all depend on a particular given nominal trajectory λ^* .

B.1 Linear-Gaussian Posterior Mean and Covariance

Consider a linear model calibration problem:

$$\widehat{z}_{[k]} = a_{[k]}\lambda + b_{[k]} + \xi_{[k]}, \quad (\text{B.1})$$

where $\widehat{z}_{[k]} \in \mathbb{R}^m$ is the vector of noisy measurements at observation time j , $\lambda \in \mathbb{R}^p$ is the model parameter, $a_{[k]} \in \mathbb{R}^{m \times p}$ and $b_{[k]} \in \mathbb{R}^m$ result for a given linearized model (i.e., created from a given λ^*), and $\xi_{[k]} \in \mathbb{R}^m$ is an assumed additive Gaussian noise $\xi_{[k]} \sim \mathcal{N}(0_m, \Sigma_\xi)$ with $\Sigma_\xi \in \mathbb{R}^{m \times m}$ being a diagonal noise covariance matrix for the measurements at time k (we assume noise to be independent across components of $\xi_{[k]}$ and across k). When Gaussian prior $\lambda \sim \mathcal{N}(\mu_0, \Sigma_0)$ is employed (here $\mu_0 \in \mathbb{R}^p$ and $\Sigma_0 \in \mathbb{R}^{p \times p}$), then we have a conjugate system where the posterior is also Gaussian:

$$\begin{aligned} f(\lambda | \{\widehat{z}_{[k]}\}) &\propto f(\lambda) f(\{\widehat{z}_{[k]}\} | \lambda) \\ &= (2\pi)^{-\frac{p}{2}} |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\lambda - \mu_0)^T \Sigma_0^{-1} (\lambda - \mu_0) \right\} \end{aligned}$$

$$\begin{aligned}
& \times \prod_{k=0}^M (2\pi)^{-\frac{m}{2}} |\Sigma_\xi|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\widehat{z}_{[k]} - a_{[k]}\lambda - b_{[k]})^T \Sigma_\xi^{-1} (\widehat{z}_{[k]} - a_{[k]}\lambda - b_{[k]}) \right\} \\
& \propto \exp \left\{ -\frac{1}{2} (\lambda - \mu_0)^T \Sigma_0^{-1} (\lambda - \mu_0) - \frac{1}{2} \sum_{k=0}^M (\widehat{z}_{[k]} - a_{[k]}\lambda - b_{[k]})^T \Sigma_\xi^{-1} (\widehat{z}_{[k]} - a_{[k]}\lambda - b_{[k]}) \right\} \\
& = \exp \left\{ -\frac{1}{2} \left[\lambda^T \Sigma_0^{-1} \lambda - 2\lambda^T \Sigma_0^{-1} \mu_0 + \mu_0^T \Sigma_0^{-1} \mu_0 \right. \right. \\
& \quad \left. \left. + \sum_{k=0}^M \left((\widehat{z}_{[k]} - b_{[k]})^T \Sigma_\xi^{-1} (\widehat{z}_{[k]} - b_{[k]}) - 2(\widehat{z}_{[k]} - b_{[k]})^T \Sigma_\xi^{-1} a_{[k]}\lambda + \lambda^T a_{[k]}^T \Sigma_\xi^{-1} a_{[k]}\lambda \right) \right] \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left[\lambda^T \Sigma_0^{-1} \lambda - 2\lambda^T \Sigma_0^{-1} \mu_0 + \sum_{k=0}^M \left(-2(\widehat{z}_{[k]} - b_{[k]})^T \Sigma_\xi^{-1} a_{[k]}\lambda + \lambda^T a_{[k]}^T \Sigma_\xi^{-1} a_{[k]}\lambda \right) \right] \right\} \\
& = \exp \left\{ -\frac{1}{2} \left[\lambda^T \left(\Sigma_0^{-1} + \sum_{k=0}^M a_{[k]}^T \Sigma_\xi^{-1} a_{[k]} \right) \lambda - 2\lambda^T \left(\Sigma_0^{-1} \mu_0 + \sum_{k=0}^M a_{[k]}^T \Sigma_\xi^{-1} (\widehat{z}_{[k]} - b_{[k]}) \right) \right] \right\}, \tag{B.2}
\end{aligned}$$

where the first equality is the product of Gaussian prior and Gaussian likelihood PDFs, the second proportionality omits the constants in front of the exponentials and turns the product of exponentials into sum of its arguments, the second equality then expands out the products, the third proportionality omits the new multiplying constants that resulted from the constant terms inside the exponential, and the last equality factors the λ 's to arrive at a quadratic and linear term (of λ) inside the exponential (note that $(\widehat{z}_{[k]} - b_{[k]})^T \Sigma_\xi^{-1} a_{[k]}\lambda = \lambda^T a_{[k]}^T \Sigma_\xi^{-1} (\widehat{z}_{[k]} - b_{[k]})$ since the transpose of a scalar is equal to itself, and Σ_ξ^{-1} is symmetric). Without having extracted the mean and covariance, we already see this is a Gaussian PDF since it is an exponential of a quadratic.

Knowing the posterior is a Gaussian, we can write out its PDF in terms of its mean μ_π and covariance Σ_π :

$$\begin{aligned}
f(\lambda | \{\widehat{z}_{[k]}\}) &= (2\pi)^{-\frac{p}{2}} |\Sigma_\pi|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\lambda - \mu_\pi)^T \Sigma_\pi^{-1} (\lambda - \mu_\pi) \right] \\
&\propto \exp \left[-\frac{1}{2} (\lambda^T \Sigma_\pi^{-1} \lambda - 2\lambda^T \Sigma_\pi^{-1} \mu_\pi + \mu_\pi^T \Sigma_\pi^{-1} \mu_\pi) \right] \\
&\propto \exp \left[-\frac{1}{2} (\lambda^T \Sigma_\pi^{-1} \lambda - 2\lambda^T \Sigma_\pi^{-1} \mu_\pi) \right]. \tag{B.3}
\end{aligned}$$

Comparing (B.2) and (B.3), matching the quadratic term yields

$$\Sigma_\pi = \left(\Sigma_0^{-1} + \sum_{k=0}^M a_{[k]}^T \Sigma_\xi^{-1} a_{[k]} \right)^{-1}, \quad (\text{B.4})$$

and matching the linear term yields

$$\mu_\pi = \Sigma_\pi \left[\Sigma_0^{-1} \mu_0 + \sum_{k=0}^M a_{[k]}^T \Sigma_\xi^{-T} (\hat{z}_{[k]} - b_{[k]}) \right]. \quad (\text{B.5})$$

B.2 Linear-Gaussian Log-Evidence

The log-evidence is

$$\begin{aligned} \ln f(\{\hat{z}_{[k]}\}) &= \ln \left[\frac{f(\lambda) f(\{\hat{z}_{[k]}\}|\lambda)}{f(\lambda|\{\hat{z}_{[k]}\})} \right] \\ &= \ln f(\lambda) + \ln f(\{\hat{z}_{[k]}\}|\lambda) - \ln f(\lambda|\{\hat{z}_{[k]}\}) \\ &= -\frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_0| - \frac{1}{2} (\lambda - \mu_0)^T \Sigma_0^{-1} (\lambda - \mu_0) - \frac{m(M+1)}{2} \ln(2\pi) \\ &\quad - \frac{M+1}{2} \ln |\Sigma_\xi| - \frac{1}{2} \sum_{k=0}^M (\hat{z}_{[k]} - a_{[k]} \lambda - b_{[k]})^T \Sigma_\xi^{-1} (\hat{z}_{[k]} - a_{[k]} \lambda - b_{[k]}) \\ &\quad + \frac{p}{2} \ln(2\pi) + \frac{1}{2} \ln |\Sigma_\pi| + \frac{1}{2} (\lambda - \mu_\pi)^T \Sigma_\pi^{-1} (\lambda - \mu_\pi). \end{aligned} \quad (\text{B.6})$$

This expression should hold true for any choice of λ ; one natural choice is to use $\lambda = \lambda^*$.

Appendix C

Log-Evidence Gradient for Linearized DAE

In Appendices C.1 and C.2, we make all the λ^* notation explicit.

C.1 Log-Evidence Gradient with respect to Nominal Trajectory λ^*

Consider the gradient operator $\nabla_{\lambda^*} = \left[\frac{\partial}{\partial \lambda_1^*}, \frac{\partial}{\partial \lambda_2^*}, \dots, \frac{\partial}{\partial \lambda_p^*} \right]^T$, the k th component of $\nabla_{\lambda^*} \ln f(\{\widehat{z}_{[k]}\}|\lambda^*)$ is then

$$\begin{aligned}
\frac{\partial}{\partial \lambda_i^*} \ln f(\widehat{z}_{[k]}|\lambda^*) &= \frac{\partial}{\partial \lambda_i^*} \left[-\frac{1}{2} \sum_{k=0}^M (\widehat{z}_{[k]} - \widetilde{z}_{[k]}(\lambda; \lambda^*))^T \Sigma_\xi^{-1} (\widehat{z}_{[k]} - \widetilde{z}_{[k]}(\lambda; \lambda^*)) - \frac{1}{2} \ln |\Sigma_{\pi, \lambda^*}^{-1}| \right. \\
&\quad \left. + \frac{1}{2} (\lambda - \mu_{\pi, \lambda^*})^T \Sigma_{\pi, \lambda^*}^{-1} (\lambda - \mu_{\pi, \lambda^*}) + C \right] \\
&= \frac{1}{2} \sum_{k=0}^M \left(\frac{\partial \widetilde{z}_{[k]}}{\partial \lambda_i^*} \right)^T (\Sigma_\xi^{-1} + \Sigma_\xi^{-T}) (\widehat{z}_{[k]} - \widetilde{z}_{[k]}(\lambda; \lambda^*)) - \frac{1}{2} \text{Tr} \left(\Sigma_{\pi, \lambda^*} \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} \right) \\
&\quad + \frac{1}{2} \lambda^T \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_k^*} \lambda - \lambda^T \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_k^*} \mu_{\pi, \lambda^*} - \lambda^T \Sigma_{\pi, \lambda^*}^{-1} \frac{\partial \mu_{\pi, \lambda^*}}{\partial \lambda_k^*} \\
&\quad + \frac{1}{2} \frac{\partial \mu_{\pi, \lambda^*}^T}{\partial \lambda_k^*} \Sigma_{\pi, \lambda^*}^{-1} \mu_{\pi, \lambda^*} + \frac{1}{2} \mu_{\pi, \lambda^*}^T \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_k^*} \mu_{\pi, \lambda^*} + \frac{1}{2} \mu_{\pi, \lambda^*}^T \Sigma_{\pi, \lambda^*}^{-1} \frac{\partial \mu_{\pi, \lambda^*}}{\partial \lambda_k^*} \\
&= \frac{1}{2} \sum_{k=0}^M \left(\frac{\partial \widetilde{z}_{[k]}}{\partial \lambda_i^*} \right)^T (\Sigma_\xi^{-1} + \Sigma_\xi^{-T}) (\widehat{z}_{[k]} - \widetilde{z}_{[k]}(\lambda; \lambda^*)) - \frac{1}{2} \text{Tr} \left(\Sigma_{\pi, \lambda^*} \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} \right) \\
&\quad + \frac{1}{2} (\lambda - \mu_{\pi, \lambda^*})^T \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_k^*} (\lambda - \mu_{\pi, \lambda^*}) - (\lambda - \mu_{\pi, \lambda^*})^T \Sigma_{\pi, \lambda^*}^{-1} \frac{\partial \mu_{\pi, \lambda^*}}{\partial \lambda_k^*}, \tag{C.1}
\end{aligned}$$

where in the first equality we substituted the log-PDF expressions similar to (B.6) but used $-\frac{1}{2} \ln |\Sigma_{\pi, \lambda^*}| = \frac{1}{2} \ln |\Sigma_{\pi, \lambda^*}^{-1}|$. From here, the needed derivative calculations are the partial derivatives of $\tilde{z}_{[k]}$, $\Sigma_{\pi, \lambda^*}^{-1}$, and μ_{π, λ^*} with respect to λ_i^* : from (3.14),

$$\frac{\partial \tilde{z}_{[k]}}{\partial \lambda_i^*} = \tilde{z}_{\lambda_i^*, [k]} = a_{\lambda_i^*, [k]} \lambda + b_{\lambda_i^*, [k]} = z_{\lambda \lambda_i^*, [k]}^* \lambda - z_{\lambda \lambda_i^*, [k]}^* \lambda^*,$$

from (B.4),

$$\begin{aligned} \Sigma_{\pi, \lambda^*}^{-1} &= \Sigma_0^{-1} + \sum_{k=0}^M a_{[k]}^T \Sigma_{\xi}^{-1} a_{[k]} \\ \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} &= \sum_{k=0}^M \frac{\partial a_{[k]}^T}{\partial \lambda_i^*} \Sigma_{\xi}^{-1} a_{[k]} + a_{[k]}^T \Sigma_{\xi}^{-1} \frac{\partial a_{[k]}}{\partial \lambda_i^*}, \end{aligned} \quad (\text{C.2})$$

and from (B.5)

$$\begin{aligned} \mu_{\pi, \lambda^*} &= \Sigma_{\pi, \lambda^*} \left[\Sigma_0^{-1} \mu_0 + \sum_{k=0}^M a_{[k]}^T \Sigma_{\xi}^{-T} (\hat{z}_{[k]} - b_{[k]}) \right] \\ \frac{\partial \mu_{\pi, \lambda^*}}{\partial \lambda_i^*} &= \frac{\partial \Sigma_{\pi, \lambda^*}}{\partial \lambda_i^*} \left[\Sigma_0^{-1} \mu_0 + \sum_{k=0}^M a_{[k]}^T \Sigma_{\xi}^{-T} (\hat{z}_{[k]} - b_{[k]}) \right] \\ &\quad + \Sigma_{\pi, \lambda^*} \left[\sum_{k=0}^M \left(\frac{\partial a_{[k]}^T}{\partial \lambda_i^*} \Sigma_{\xi}^{-T} (\hat{z}_{[k]} - b_{[k]}) - a_{[k]}^T \Sigma_{\xi}^{-T} \frac{\partial b_{[k]}}{\partial \lambda_i^*} \right) \right] \\ &= \frac{\partial \Sigma_{\pi, \lambda^*}}{\partial \lambda_i^*} \Sigma_{\pi, \lambda^*}^{-1} \mu_{\pi, \lambda^*} \\ &\quad + \Sigma_{\pi, \lambda^*} \left[\sum_{k=0}^M \left(\frac{\partial a_{[k]}^T}{\partial \lambda_i^*} \Sigma_{\xi}^{-T} (\hat{z}_{[k]} - b_{[k]}) - a_{[k]}^T \Sigma_{\xi}^{-T} \frac{\partial b_{[k]}}{\partial \lambda_i^*} \right) \right] \end{aligned} \quad (\text{C.3})$$

where

$$\frac{\partial \Sigma_{\pi, \lambda^*}}{\partial \lambda_i^*} = -\Sigma_{\pi, \lambda^*} \frac{\partial \Sigma_{\pi, \lambda^*}^{-1}}{\partial \lambda_i^*} \Sigma_{\pi, \lambda^*} \quad (\text{C.4})$$

can leverage (C.2).

C.2 Evaluating $z_{\lambda\lambda_i}^*$

In this section, we compute $z_{\lambda\lambda_i}^*$ at each time instant $k = 0, \dots, M$. To obtain the second-order trajectory sensitivity $z_{\lambda\lambda_i}^*$, we will find it useful to define $x_{\lambda\lambda_i}^*$ and $y_{\lambda\lambda_i}^*$ as the partial derivatives of x_λ^* and y_λ^* , respectively, with respect to λ_i^* . Then, applying chain rule to differentiate (2.14) with respect to λ_i^* , we get

$$\begin{aligned} z_{\lambda\lambda_i}^* &= (h_{xx^*}^* x_\lambda^* e_i + h_{xy^*}^* y_\lambda^* e_i + h_{x\lambda_i^*}^*) x_\lambda^* + h_x^* x_{\lambda\lambda_i}^* \\ &\quad + (h_{yx^*}^* x_\lambda^* e_i + h_{yy^*}^* y_\lambda^* e_i + h_{y\lambda_i^*}^*) y_\lambda^* + h_y^* y_{\lambda\lambda_i}^* \\ &\quad + h_{\lambda x^*}^* x_\lambda^* e_i + h_{\lambda y^*}^* y_\lambda^* e_i + h_{\lambda\lambda_i^*}^*, \end{aligned} \quad (\text{C.5})$$

where, in general, the notation h_{ab}^* refers to the partial derivative of gradient vector h_a^* with respect to the vector or scalar variable b , and e_i is an appropriate sized basis vector with 0s in all entries except the i th one with 1. The partial derivatives in (C.5) can all be computed in analytical closed form. Furthermore, we have nominal trajectories x_λ^* and y_λ^* resulting from nominal parameter value λ^* . Thus, to evaluate (C.5), we need only to solve for sensitivity trajectories $x_{\lambda\lambda_i}^*$ and $y_{\lambda\lambda_i}^*$. To do so, we take the derivative of (2.9) and (2.10) with respect to λ_i^* to get

$$\begin{aligned} \dot{x}_{\lambda\lambda_i}^* &= (f_{xx^*}^* x_\lambda^* e_i + f_{xy^*}^* y_\lambda^* e_i + f_{x\lambda_i^*}^*) x_\lambda^* + f_x^* x_{\lambda\lambda_i}^* \\ &\quad + (f_{yx^*}^* x_\lambda^* e_i + f_{yy^*}^* y_\lambda^* e_i + f_{y\lambda_i^*}^*) y_\lambda^* + f_y^* y_{\lambda\lambda_i}^* \\ &\quad + f_{\lambda x^*}^* x_\lambda^* e_i + f_{\lambda y^*}^* y_\lambda^* e_i + f_{\lambda\lambda_i^*}^*, \end{aligned} \quad (\text{C.6})$$

$$\begin{aligned} 0 &= (g_{xx^*}^* x_\lambda^* e_i + g_{xy^*}^* y_\lambda^* e_i + g_{x\lambda_i^*}^*) x_\lambda^* + g_x^* x_{\lambda\lambda_i}^* \\ &\quad + (g_{yx^*}^* x_\lambda^* e_i + g_{yy^*}^* y_\lambda^* e_i + g_{y\lambda_i^*}^*) y_\lambda^* + g_y^* y_{\lambda\lambda_i}^* \\ &\quad + g_{\lambda x^*}^* x_\lambda^* e_i + g_{\lambda y^*}^* y_\lambda^* e_i + g_{\lambda\lambda_i^*}^*, \end{aligned} \quad (\text{C.7})$$

where, in general, the notation f_{ab}^* (g_{ab}^*) refers to the partial derivative of gradient vector f_a^* (g_a^*) with respect to the vector or scalar variable b . We further rearrange (C.7) to get

$$\begin{aligned} y_{\lambda\lambda_i}^* &= -(g_y^*)^{-1} ((g_{xx^*}^* x_\lambda^* e_i + g_{xy^*}^* y_\lambda^* e_i + g_{x\lambda_i^*}^*) x_\lambda^* + g_x^* x_{\lambda\lambda_i}^* \\ &\quad + (g_{yx^*}^* x_\lambda^* e_i + g_{yy^*}^* y_\lambda^* e_i + g_{y\lambda_i^*}^*) y_\lambda^* \\ &\quad + g_{\lambda x^*}^* x_\lambda^* e_i + g_{\lambda y^*}^* y_\lambda^* e_i + g_{\lambda\lambda_i^*}^*), \end{aligned} \quad (\text{C.8})$$

where we have persisted with the assumption that the power-flow Jacobian matrix g_y^* is invertible along the nominal trajectory. Finally, by substituting (2.12) and (C.8) into (C.6) and (C.5), we get the following linear time-varying system describing the evolution of $z_{\lambda\lambda_i}^*$:

$$\dot{x}_{\lambda\lambda_i}^* = F_1 x_{\lambda\lambda_i}^* + F_2 x_{\lambda}^* + F_3, \quad (\text{C.9})$$

$$z_{\lambda\lambda_i}^* = G_1 x_{\lambda\lambda_i}^* + G_2 x_{\lambda}^* + G_3, \quad (\text{C.10})$$

where

$$F_1 = f_x^* - f_y^*(g_y^*)^{-1}g_x^* = A^*, \quad (\text{C.11})$$

$$\begin{aligned} F_2 = & f_{xx^*}^* x_{\lambda}^* e_i + f_{xy^*}^* y_{\lambda}^* e_i + f_{x\lambda_i^*}^* \\ & - (f_{yx^*}^* x_{\lambda}^* e_i + f_{yy^*}^* y_{\lambda}^* e_i + f_{y\lambda_i^*}^*) (g_y^*)^{-1} g_x^* \\ & - f_y^*(g_y^*)^{-1} (g_{xx^*}^* x_{\lambda}^* e_i + g_{xy^*}^* y_{\lambda}^* e_i + g_{x\lambda_i^*}^*) \\ & + f_y^*(g_y^*)^{-1} (g_{yx^*}^* x_{\lambda}^* e_i + g_{yy^*}^* y_{\lambda}^* e_i + g_{y\lambda_i^*}^*) (g_y^*)^{-1} g_x^*, \end{aligned} \quad (\text{C.12})$$

$$\begin{aligned} F_3 = & f_{\lambda x^*}^* x_{\lambda}^* e_i + f_{\lambda y^*}^* y_{\lambda}^* e_i + f_{\lambda\lambda_i^*}^* \\ & - f_y^*(g_y^*)^{-1} (g_{\lambda x^*}^* x_{\lambda}^* e_i + g_{\lambda y^*}^* y_{\lambda}^* e_i + g_{\lambda\lambda_i^*}^*) \\ & - (f_{yx^*}^* x_{\lambda}^* e_i + f_{yy^*}^* y_{\lambda}^* e_i + f_{y\lambda_i^*}^*) (g_y^*)^{-1} g_{\lambda}^* \\ & + f_y^*(g_y^*)^{-1} (g_{yx^*}^* x_{\lambda}^* e_i + g_{yy^*}^* y_{\lambda}^* e_i + g_{y\lambda_i^*}^*) (g_y^*)^{-1} g_{\lambda}^*, \end{aligned} \quad (\text{C.13})$$

$$G_1 = h_x^* - h_y^*(g_y^*)^{-1}g_x^* = C^*, \quad (\text{C.14})$$

$$\begin{aligned} G_2 = & h_{xx^*}^* x_{\lambda}^* e_i + h_{xy^*}^* y_{\lambda}^* e_i + h_{x\lambda_i^*}^* \\ & - (h_{yx^*}^* x_{\lambda}^* e_i + h_{yy^*}^* y_{\lambda}^* e_i + h_{y\lambda_i^*}^*) (g_y^*)^{-1} g_x^* \\ & - h_y^*(g_y^*)^{-1} (g_{xx^*}^* x_{\lambda}^* e_i + g_{xy^*}^* y_{\lambda}^* e_i + g_{x\lambda_i^*}^*) \\ & + h_y^*(g_y^*)^{-1} (g_{yx^*}^* x_{\lambda}^* e_i + g_{yy^*}^* y_{\lambda}^* e_i + g_{y\lambda_i^*}^*) (g_y^*)^{-1} g_x^*, \end{aligned} \quad (\text{C.15})$$

$$\begin{aligned} G_3 = & h_{\lambda x^*}^* x_{\lambda}^* e_i + h_{\lambda y^*}^* y_{\lambda}^* e_i + h_{\lambda\lambda_i^*}^* \\ & - h_y^*(g_y^*)^{-1} (g_{\lambda x^*}^* x_{\lambda}^* e_i + g_{\lambda y^*}^* y_{\lambda}^* e_i + g_{\lambda\lambda_i^*}^*) \\ & - (h_{yx^*}^* x_{\lambda}^* e_i + h_{yy^*}^* y_{\lambda}^* e_i + h_{y\lambda_i^*}^*) (g_y^*)^{-1} g_{\lambda}^* \\ & + h_y^*(g_y^*)^{-1} (g_{yx^*}^* x_{\lambda}^* e_i + g_{yy^*}^* y_{\lambda}^* e_i + g_{y\lambda_i^*}^*) (g_y^*)^{-1} g_{\lambda}^*. \end{aligned} \quad (\text{C.16})$$

Simultaneous time-domain simulation of (2.1)–(2.3), (2.13)–(2.14), and (C.9)–(C.10) yields the nominal trajectories taken by the system outputs collected in z^* , their sensitivities to parameters collected in z_λ^* , as well as their second-order sensitivities with respect to λ_i^* collected in $z_{\lambda\lambda_i}^*$, which is acquired at time instant k to help evaluate (3.14).