

Tracking Electromechanical Oscillations: An Enhanced Maximum-Likelihood Based Approach

Haris M. Khalid, *Member, IEEE*, and Jimmy C.-H. Peng, *Member, IEEE*

Abstract—Lightly damped electromechanical oscillations are major operating concerns if failed to be detected at an early stage. This paper improved the existing extended complex Kalman filter (ECKF) technique of tracking electromechanical oscillations using synchrophasor measurements. The proposed algorithm adopted a distributed architecture for estimating oscillatory parameters from local substations. The novelty lies in handling maximum likelihood (ML) to enhance the convergence property in tracking multiple modes using an expectation maximization (EM) approach. This was achieved by encapsulating the augmented Lagrangian (AL) in the maximization step of the EM algorithm, which utilized a novel ECKF-based smoother (ECKS). Performance evaluations were conducted using IEEE 68 Bus system and recorded synchrophasor measurements collected from the New Zealand grid. Random noise variance test cases were generated to examine the performance of the proposed algorithm. To ensure the robustness to random noisy conditions, the algorithm was tested based on exhaustive Monte Carlo simulations. Comparisons were made with the existing Prony Analysis (PA), Kalman filter (KF), and distributed EM-based FB-KLPF.

Index Terms—Augmented Lagrangian, distributed estimation, maximum-likelihood, oscillations, power system stability, phasor measurement unit (PMU), smoother, synchrophasor.

I. INTRODUCTION

DETECTING electromechanical oscillations in power transmission networks is critical to ensure the reliability of power transfer between regions [1, 2]. Oscillatory modes like inter-area oscillations are prone in stressed systems that lack reactive support. They exhibit a low frequency range of less than 1 Hz, and are difficult to be promptly detected. Consequently, conservative power transfer margins are set in tie-lines to mitigate the occurrences of lightly-damped inter-area oscillations. To better optimize the transmission assets, real-time detection schemes using synchrophasor measurements were proposed.

To date, many oscillation detection schemes were published. They can be classified into recursive and block-processing methods. Recursive methods update the mode estimates at every sampling time instant based on the previous estimates [3]. The first recursive method applied to process ambient data is presented in [4], followed by Kalman filter (KF) approach presented in [5]. On the other hand, block-processing techniques are based on simultaneously processing a set of data within a single sliding window. Each new estimation and its respective calculation is independent to the previous data window. In published block-processing methods, Prony analysis is one of the most widely used technique [6]. A literature review of other monitoring schemes are outlined in [7–12]. Despite many oscillation monitoring methods were published, most of them extract system dynamics based on a single metering location. Moreover, they face a common challenge of tracking

oscillations having similar frequencies under noisy conditions.

The contribution of this paper is to minimize the detection inaccuracies caused by the lack of oscillatory observability, and network perturbations in metering locations. This is accomplished by revamping the state estimation in a distributed architecture. The information of likelihood is then extracted using a proposed scheme named as expected augmented Lagrangian maximization-based on extended complex Kalman smoother (EALM-based ECKS). Compared with its expectation maximization (EM) predecessor [13], the proposed method improves the estimation accuracy under continuous random noise fluctuations. This is done by firstly computing modal parameters collected from local locations in a dynamical system. Subsequently, by considering an observation model for the state variables, maximum-likelihood (ML) estimations are calculated using EALM at maximization step and tackles the random noisy conditions as asymptotic matrix. In addition, extended complex Kalman filter-based smoother (ECKS) is applied at the initial correlation step due to its property of giving an estimated sequence of the unobserved oscillation variable. As a result, the proposed scheme provides a novel and convenient way to enhance the modal estimations at locations that are contaminated by noise and system perturbations.

The paper is organized as follows; a proposed oscillation detection scheme is shown in Section II, followed by the implementation and evaluation of the proposed scheme in Section III, and finally conclusions are drawn in Section IV.

Notations: In this paper, \mathbf{E} is the expectation operator. A symbol $\hat{\cdot}$ over a variable indicates an estimate of that variable e.g. \hat{x} is an estimate of x . The individual entries of a variable like x are denoted by $x(l)$. When any of these variables become a function of time, the time index t appears as a subscript (e.g. x_t, H_t, z_t). The notation x_0^T is used to denote the time sequence (e.g. x_0, x_1, \dots, x_T).

II. THE PROPOSED OSCILLATION DETECTION SCHEME

This section begins with outlining the assumed system model, followed by the state representation of electromechanical oscillations. The EALM-based ECKS algorithm is then built on it for calculating the maximum likelihood. An overview of this section is illustrated in Fig. 1. It summarizes the formulation and equations involved at each step.

A. State Representation Observation Model

Consider a dynamical system, where all sensors are time synchronized and have the same measurement rate in a distributed environment, followed by observations at i -th sensor as:

$$x_{t+1} = f(x_t, v_t), \quad t = 0, 1, \dots, T \quad (1)$$

$$z_t^i = H_t^i x_t + w_t^i, \quad i = 1, \dots, N \quad (2)$$

This work was funded by the Cooperative Agreement between the Masdar Institute of Science and Technology (MI), Abu Dhabi, UAE and the Massachusetts Institute of Technology (MIT), Cambridge, MA.

The authors are with Department of Electrical Engineering and Computer Science, Institute Center for Energy, MIST, Abu Dhabi, UAE. E-mail: mkhalid, jpeng@masdar.ac.ae

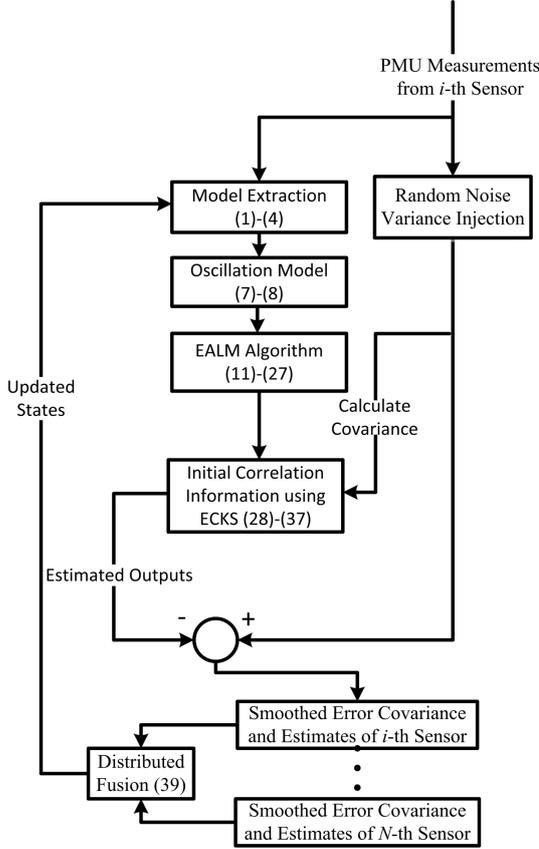


Fig. 1. Framework of the proposed oscillation detection scheme

where $f(\cdot)$ is the known nonlinear function representing the state transition model, $x_0 \in \mathbf{R}^r$ is the initial condition of the oscillation state, r is the oscillation state vector size in the subspace \mathbf{R} . In addition, $v_t \in \mathbf{R}^r$ is the random load fluctuations based on process noise, t is the discrete-time instant, and T is the number of time instants.

Now let the system described in (1) be monitored by a network of N number of sensors, which are basically PMUs installed in high-voltage substations. Therefore, referring to (2), $z_t^i \in \mathbf{R}^{p^i}$ is the observation output of oscillations at i -th sensor, p^i is the number of local simultaneous oscillatory observations made by i -th sensor, $H_t^i \in \mathbf{R}^{p^i \times r}$ is the local observation matrix of i -th sensor, x_t is the state matrix for oscillations, and $w_t^i \in \mathbf{R}^{p^i}$ is the local observation noise.

To integrate a distributed architecture, all local observations from N sensors in the network are integrated synthetically into the master observation output model $z_{M,t} \in \mathbf{R}^{p_M}$. The symbol M denotes the master, and p_M is the dimension of master observation output collected from N number of local i -th sensors. Similar to (2), the master observation model at time-instant t is represented as,

$$z_{M,t} = H_{M,t}x_t + w_{M,t}, \quad (3)$$

The state matrix is x_t , the master observation output vector is $z_{M,t}$, the master observation matrix is $H_{M,t}$, and the master observation noise vector is $w_{M,t}$. They can also be represented

as:

$$z_{M,t} = \begin{bmatrix} z_t^1 \\ \vdots \\ z_t^N \end{bmatrix}, H_{M,t} = \begin{bmatrix} H_t^1 \\ \vdots \\ H_t^N \end{bmatrix}, w_{M,t} = \begin{bmatrix} w_t^1 \\ \vdots \\ w_t^N \end{bmatrix} \quad (4)$$

where N is the number of sensors.

Assumption II.1: It has been initially assumed that the noises w_t and v_t are uncorrelated, and are zero-mean white noise sequence with Gaussian distribution:

$$\mathbf{E}[w_t] = \mathbf{E}[v_t] = \mathbf{E}[w_t v_t^T] = 0, \forall t \quad (5)$$

$$\mathbf{E}[w_g w_h^T] = R_t \delta_{gh}, \mathbf{E}[v_g v_h^T] = Q_t \delta_{gh}, \forall t \quad (6)$$

Note R_t represents the residual covariance, δ_{gh} is a Kronecker delta which is one when variables g and h are the same. Q_t is the process noise correlation factor.

Once the observation model is extracted from the synchrophasor measurements, the corresponding state representation can be formulated in the frequency domain.

B. Electromechanical Oscillation Model Formulation

Assume a signal contains K number of electromechanical oscillations. The observation output signal z_t^i from an i -th sensor at time t can be modeled in the frequency domain as:

$$z_t^i = \sum_{k=1}^K b_k e^{(-\sigma_k + j2\pi f_k)tT_s} + w_t^i, \quad t = 1, 2, \dots, T \quad (7)$$

where b_k is the complex amplitude of k -th mode [14]. σ_k is the damping factor, f_k is the oscillatory frequency, and T_s is the sampling time. For convenience, the term $-\sigma_k + j2\pi f_k$ can be represented in the rectangular form as λ_k . In this paper, the k -th eigenvalue of a particular signal is described by two states denoted as $x_{k,t}$ and $x_{k+1,t}$, respectively. They can also be expressed for an i -th sensor as:

$$x_{k,t}^i = e^{(-\sigma_k + j2\pi f_k)tT_s}, \quad x_{k+1,t}^i = b_{k+1} e^{(-\sigma_{k+1} + j2\pi f_{k+1})tT_s} \quad (8)$$

The term b_k represents the complex amplitude of the k -th mode. Thus, a signal consisting of K number of exponentially damped sinusoids is modeled by $2K$ number of states.

To find the damping factor σ_k and the corresponding oscillatory frequency f_k , the individual oscillation state x_t needs to be determined. Note the direct computation of x_t requires the complete observability of the oscillation observation matrix. This was achieved in [13] by recursively calculating the conditional expectation, and maximizing the possible likelihood of the unobserved latent oscillation variable H_t using the EM algorithm. The distribution of the unobserved latent variable was iteratively approximated. However, EM algorithm only guarantees a convergence to a local maximum of the defined objective function. Moreover, apart from its slower convergence rate due to the data collected from multiple substations, EM is not able to provide estimation to the asymptotic matrix of the maximum likelihood for both variance and covariance respectively. This asymptotic matrix can be defined as a case of random noise with high variance. To achieve the robustness, while optimizing the asymptotic matrix and improving the convergence rate, a conjugate direction method is required. This is achieved by incorporating augmented Lagrangian (AL) at the maximization step of EM algorithm. AL has been chosen as it achieves the similar

estimation accuracy in less number of iterations, while avoiding ill-conditioning. In order to integrate AL into the proposed method, a clear definition of the ML is required in the context of EM.

C. Recursive ML Estimations using EALM-based ECKS

A two-step approach is proposed to recursively calculate ML estimations. They comprise of: 1) Expected augmented Lagrangian-based maximization (EALM), and 2) initial correlation information using extended complex Kalman filter-based smoother (ECKS). The purpose is to individually calculate the maximum-likelihood at each step, while iteratively applying the correlation information of estimation sequence between both steps.

Subsequently, by using the definition of conditional probability, the joint density of observations from oscillations $p_{x_t}(z_1, \dots, z_T)$ for the ideal case can be written as:

$$\begin{aligned} L_{x_t}(z_t) &= \log p_{x_t}(z_1, \dots, z_T) \\ &= \sum_{t=2}^T \log p_{x_t}(z_t|z_{t-1}) + \log p_{x_t}(z_1) \end{aligned} \quad (9)$$

The variable L_{x_t} is the log-likelihood function for oscillation state x_t . In addition, (z_1, \dots, z_T) refers to the measurement sequence z_0^T at time-instant t , p_x is the probability density function of x_t at time instant t . Since log is a strictly increasing function, (9) is equivalent to:

$$\hat{x}_t^{ML} = \arg \max_{x_t} \sum_{t=2}^T \log p_{x_t}(z_t|z_{T-1}) + \log p_{x_t}(z_1) \quad (10)$$

where $\hat{x}_t^{ML} = \arg \max_{x_t} p_{x_t}(z_1, \dots, z_T)$. Although (10) can be solved by knowing the observability of oscillations in H_t matrix, it is not the case here. Therefore, EALM algorithm is required to maximize the log-likelihood function of nearby oscillations.

1) *Expected Augmented Lagrangian-based Maximization Algorithm*: The EALM algorithm has two internal steps: 1) E-step, and 2) ALM-step. The E-step is implemented with respect to the underlying unknown variables conditioned on the observations, thus maximizes the likelihood with respect to the oscillation states. This is done by applying the conditional probability on the observation of unknown variables. As a result, the effects of random load changes and noisy conditions is reduced. In contrast, the ALM-step maximizes the likelihood function with respect to the modal parameters to get a new state estimate. This is achieved by first maximizing using M-step with respect to the oscillation fluctuation parameters, and then further converging to a local minimum using AL algorithm.

The key idea of calculating maximum likelihood using EALM is to consider a joint log-likelihood function containing both the observed variable z_t , and the unobserved latent variable H_t . Note H_t is assumed to be available. Hence,

$$L_{x_t}(H_t, z_t) = \log p_{x_t}(H_t, z_t) \quad (11)$$

According to the definition of conditional probability, the resultant joint density function of (11) can be expressed as:

$$p_{x_t}(H_t|z_t) = \frac{p_{x_t}(H_t, z_t)}{p_{x_t}(z_t)} \quad (12)$$

By combining (9) and (11), we now attempt to solve the function outlined in (10). Firstly, the resultant log-likelihood function becomes:

$$L_{x_t}(z_t) = \log p_{x_t}(z_t) = \log p_{x_t}(H_t, z_t) - \log p_{x_t}(H_t|z_t) \quad (13)$$

The conditional distribution $p_{\hat{x}_t}(H_t|z_t)$ outlined in (12) is integrated to complete the formulation of the E-step. This gives:

$$\begin{aligned} \log p_{x_t}(z_t) &= \sum_{t=2}^T \log p_{x_t}(H_t, z_t) p_{\hat{x}_t}(H_t|z_t) - \sum_{t=2}^T \log p_{x_t}(H_t|z_t) \\ &\quad p_{\hat{x}_t}(H_t|z_t) \\ L_{x_t}(z_t) &= \mathbf{E}_{\hat{x}_t} [\log p_{x_t}(H_t, z_t)|z_t] - \mathbf{E}_{\hat{x}_t} [\log p_{x_t}(H_t|z_t)|z_t] \\ &= W(x_t, \hat{x}_t) - V(x_t, \hat{x}_t) \end{aligned} \quad (14)$$

where $W(x_t, \hat{x}_t) = \mathbf{E}_{\hat{x}_t} [\log p_{x_t}(H_t, z_t)|z_t]$. The term \hat{x}_t denotes the estimated oscillation state x_t at t -th instant. Meanwhile, $V(x_t, \hat{x}_t)$ is assumed for the expectation value of the conditional distribution, i.e. $\mathbf{E}_{\hat{x}_t} [\log p_{x_t}(H_t|z_t)|z_t]$. Note that $\sum_{t=2}^N \log p_{x_t}(z_t) p_{\hat{x}_t}(H_t|z_t) = \log p_{x_t}(z_t)$ is used to derive (14).

The maximum likelihood (ML) is then applied to the E-step in order to determine the convergence of the oscillation estimation. This is done by taking the difference between $L_{x_t}(z_t)$ and $L_{\hat{x}_t}(z_t)$, which is found by substituting (14) into (11) such that:

$$\begin{aligned} L_{x_t}(z_t) - L_{\hat{x}_t}(z_t) &= [W(x_t, \hat{x}_t) - W(\hat{x}_t, \hat{x}_t)] \\ &\quad + [V(\hat{x}_t, \hat{x}_t) - V(x_t, \hat{x}_t)] \end{aligned} \quad (15)$$

In (15), $V(\hat{x}_t, \hat{x}_t) - V(x_t, \hat{x}_t)$ can be solved by Kullback-Leibler (KL) information distance [15]. This results to:

$$\begin{aligned} V(\hat{x}_t, \hat{x}_t) - V(x_t, \hat{x}_t) &= \sum_{t=2}^T \log \left(\frac{p_{x_t}(H_t|z_t)}{p_{\hat{x}_t}(H_t|z_t)} \right) \\ &= p_{\hat{x}_t}(H_t|z_t) \mathbf{E}_{\hat{x}_t} \left[-\log \left(\frac{p_{x_t}(H_t|z_t)}{p_{\hat{x}_t}(H_t|z_t)} \right) |z_t \right] \end{aligned} \quad (16)$$

Note the divergence of the negative algorithm creates an auxiliary convex function, which can be generalized by Jensen's inequality [16]. Therefore, the expectation term of (16) can be expressed as:

$$\begin{aligned} \mathbf{E}_{\hat{x}_t} \left(-\log \frac{p_{x_t}(H_t|z_t)}{p_{\hat{x}_t}(H_t|z_t)} |z_t \right) &\geq -\log \mathbf{E}_{\hat{x}_t} - \hat{x}_t \left(\frac{p_{x_t}(H_t|z_t)}{p_{\hat{x}_t}(H_t|z_t)} |z_t \right) \\ &= -\log \sum_{t=2}^T p_{x_t}(H_t|z_t) = 0 \end{aligned} \quad (17)$$

Based on the inequality constraint outlined in (17), the difference term in (16) will satisfy the following condition:

$$V(\hat{x}_t, \hat{x}_t) - V(x_t, \hat{x}_t) \geq 0 \quad (18)$$

According to the relationships derived from (16) to (18), if (15) is used to choose a new parameter x_t such that $W(x_t, \hat{x}_t) \geq W(\hat{x}_t, \hat{x}_t)$, it will increase the estimation likelihood as,

$$L_{x_t}(z_t) \geq L_{\hat{x}_t}(z_t) \quad (19)$$

It should be noted that the computation of $W(x_t, \hat{x}_t)$ in (14) will maximize the ML with respect to x_t . As a result, the new estimate \hat{x}_{t+1} could be obtained. This completes the E-step derivation.

We will now formulate the ALM-step. Using the definition in (17), and applying the Bayes rule, the augmented Lagrangian-

based maximization can be incorporated in the moments. These moments are calculated using the normal distribution. As a side note, the augmented Lagrangian also defines the co-states and penalty functions with respect to the derived smoother in the initial correlation step. They are explained later in the next subsection, and summarized in (28) to (37). Coming back to the topic on ALM-step, the Karush-Kuhn-Tucker (KKT) first-order optimality necessary conditions is implemented for augmented Lagrangian [17]. It requires $R_{e,t}^i$, K_t , $x_{t|t}^i$, $x_{t|t-1}^i$, $P_{t|t}^i$ and $P_{t|t-1}^i$ to be the solution of the minimization problem based on Lagrangian function $\min_{H_t} f [|H_t^i(l)|^2 |z_t^i(l), x_t^i(l)]$, such that:

$$\mathbf{E}[|H_t^i(l)|^2 |z_t^i(l), x_t^i(l)] = \frac{\sum_{\chi=1}^n |A_\chi|^2 e^{-\frac{|z_t(l)-x_t(l)A_\chi|^2}{\sigma_n^2}}}{\sum_{\chi=1}^n e^{-\frac{|z_t(l)-x_t(l)A_\chi|^2}{\sigma_n^2}}} - a_{1,t} [R_{e,t}^i - 1 - H_t^i P_{t|t}^i H_t^{i*}] - r_t [R_{e,t}^i - 1 - H_t P_{t|t}^i H_t^{i*}]^2 - a_{2,t} [K_t - P_{t|t}^i H_t^{i*} R_{e,t}^{i-1}] - r_t (K_t - P_{t|t}^i H_t^{i*} R_{e,t}^{i-1})^2 - a_{3,t} [x_t^i - (1 - K_t H_t^i) x_{t|t-1}^i - K_t z_t] - r_t [x_t^i - (1 - K_t H_t^i) x_{t|t-1}^i - K_t z_t]^2 - a_{4,t} [x_{t+1}^i - f(x_{t|t}^i)] - r_t [x_{t+1}^i - f(x_{t|t}^i)]^2 - a_{5,t} [P_{t|t-1}^i - F_t (P_{t|t-1}^i - K_t R_{e,t}^i K_t^*) F_t^*] - r_t [P_{t|t-1}^i - F_t (P_{t|t-1}^i - K_t R_{e,t}^i K_t^*) F_t^*]^2 \quad (20)$$

where $a_{1,t}, \dots, a_{5,t}$ are the co-states and r_t is the penalty coefficient. The symbol \top is the transpose operator and σ_n^2 is the noise variance. The variable $x_{t|t}^i$ is the local updated a -posteriori state estimate, and $x_{t|t-1}^i$ is the local predicted a -priori state estimate. Furthermore, $R_{e,t}^i$ denotes the local covariance matrix of the estimation error e_t of the i -th sensor, $P_{t|t-1}^i$ represents the local predicted a -priori estimate covariance matrix, and $P_{t|t}^i$ is the local updated a -posteriori estimate covariance. Finally, F_t is the state transition matrix of the oscillation state x_t . Note that KKT assumed that $H_t^i(l)$ is drawn from $A_\chi = A_1, A_2, \dots, A_n$. It represents the oscillation data from different local substations. According to the augmented Lagrangian (AL) theorem from [17], KKT conditions generalize the method of Lagrange multipliers. They become necessary when the problem is convex, i.e. handling the convergence of M-step with AL and holding Slater's conditions to maintain convexity. Thus, there must exist some positive-Lagrange multipliers and penalty coefficients such that,

$$\frac{\partial M}{\partial R_{e,t}^i} = 0, \frac{\partial M}{\partial K_t} = 0, \frac{\partial M}{\partial x_t^i} = 0, \frac{\partial M}{\partial x_{t|t-1}^i} = 0, \frac{\partial M}{\partial P_{t|t-1}^i} = 0 \quad (21)$$

The partial derivatives outlined in (21) can be individually evaluated as:

$$\frac{\partial L}{\partial R_{e,t}^i} = 2r_t (1 - R_{e,t}^i + H_t^i P_{t|t}^i H_t^{i*} - K_t - \frac{P_{t|t}^i H_t^{i*2}}{R_{e,t}^{i3}} - P_{t|t-1}^i) + F_t (P_{t|t}^i - K_t R_{e,t}^i K_t^* F_t^*) (F_t K_t K_t^* F_t^*) \quad (22)$$

$$\frac{\partial L}{\partial K_t} = -a_{2,t} - a_{3,t} K_t x_{t|t-1}^i z_t - a_{5,t} (F_t K_t K_t^* F_t^*) + 2r_t (-K_t + P_{t|t}^i H_t^* R_{e,t}^{i-1} - x_t^i + x_{t|t-1}^i - P_{t|t-1}^i + F_t (P_{t|t-1}^i - K_t) R_{e,t}^i K_t^* F_t^*) (F_t K_t K_t^* F_t^*) \quad (23)$$

$$\frac{\partial L}{\partial x_t^i} = \frac{|A_\chi|^2 e^{-\frac{|z_t(l)-x_t(l)A_\chi|^2}{\sigma_n^2}}}{(e^{-\frac{|z_t(l)-x_t(l)A_\chi|^2}{\sigma_n^2}})^2} \frac{-2|z_t(l)-x_t(l)A_\chi|\sigma_n^2}{\sigma_n^2} - A_\chi e^{-\frac{|z_t(l)-x_t(l)A_\chi|^2}{\sigma_n^2}} \frac{-2|z_t(l)-x_t(l)A_\chi|}{\sigma_n^3} - a_{3,t} - 2r_t (x_t^i - x_{t|t-1}^i + x_{t|t-1}^i K_t H_t^i - K_t z_t) \quad (24)$$

$$\frac{\partial L}{\partial x_{t|t-1}^i} = -a_{3,t} (1 - K_t H_t^i) - a_{4,t} - 2r_t (x_t^i - x_{t|t-1}^i K_t H_t^i - K_t z_t^i (1 + K_t H_t^i) + x_{t|t-1}^i - f(x_t^i)) \quad (25)$$

$$\frac{\partial L}{\partial P_{t|t-1}^i} = -a_{5,t} - 2r_t (P_{t|t-1}^i - F_t (P_{t|t-1}^i - K_t R_{e,t}^i K_t^*) F_t^*) \quad (26)$$

Here, the co-states a_t are determined by a backward integration of the adjunct state equations yielding:

$$\begin{bmatrix} a_1 \\ \vdots \\ a_5 \end{bmatrix}_{t-1} = -2\tau_s \frac{\partial L_{x_t}}{\partial x_t} - F_t^* a_t - \tau_s \left[\sum_{t=1}^N \nabla_{x_t} \psi_{r_t} s_t(x_t, v_t, \tau_s) \right] \quad (27)$$

where $x_{t+1} = f(x_t, v_t, \tau_s)$, and τ_s is the sampling time such that $\tau_s = x_{t+1} - x_t = \frac{x_T - x_0}{N}$, for $t = 1, 2, \dots, T-1$. Meanwhile, ∇_{x_t} is the gradient of penalty function, and ψ_{r_t} is the penalty function defined for the unobserved latent variable H_t since $\psi_{r_t}(L_{x_t}, a_t) = (L_{x_t} + \frac{r_t}{2} a_t)^* a_t$, $s_t(x_t, v_t, \tau_s) = 0$. Lastly, s_t is the equality constraint parameter. The development of these conditions enable us to derive the iterative formulas to solve the maximization step by adjusting the Lagrange multipliers and penalty functions. This is the core of the ALM step.

However, as shown in the previous subsection and Section I, to achieve an optimum oscillation estimate, an initial estimation step is required.

2) *Initial Correlation Information using ECKF-based Smoother (ECKS)*: Referring to (1) and (2), recall the measurement sequence ($t = 0, 1, \dots, T$) of the input and output measurements H_t^T and z_t^T , respectively. Although ECKF can calculate the ML initial correlation information of a given observation H_t , the aim however is to calculate the initial ML estimate of x_t given the whole sequence H_0^T . To improve on the sub-optimal correlation information provided by ECKF, an ECKF-based smoother is applied to find the initial oscillation state estimate $\hat{x}_{t|T}$. This estimate is superior to that obtained when the final sub-optimal ECKF estimate is extrapolated backwards in time.

The proposed ECKS scheme starts with an initial distribution of the latent variable, and the first observation $p_x(x_1|z_1) = p_x(z_1|x_1)p_x(x_1)$. It also assumed that z_1 has a Gaussian distribution, where $p_x(z_1)$ is approximately $N(\mu_0, \sigma_0^2)$ with mean μ_0 and variance σ_0^2 . For the forward recursion, the initial condition starts from $P_{0|-1} = \text{var}(z_0^T)$, and $x_{0|-1} = 0$. This shows the availability of the a -priori information at the previous instant of time. For $t = 1, \dots, T$, the oscillation state and covariance can be determined by:

$$R_{e,t} = \sigma_n^2 I_{p+r} + H_t^i P_{t|t-1} H_t^{i*} \quad (28)$$

$$e_t^i = z_t^i - H_t^i \hat{x}_{t|t-1}^i \quad (29)$$

$$\hat{x}_{t|t}^i = \hat{x}_{t|t-1}^i + K_t e_t^i \quad (30)$$

$$\hat{x}_{t+1|t}^i = f(\hat{x}_{t|t}^i) \quad (31)$$

$$K_t = \hat{P}_{t|t-1}^i H_t^{i*} [H_t^i \hat{P}_{t|t-1}^i H_t^{i*} + R_{e,t}^i]^{-1} \quad (32)$$

$$\hat{P}_{t|t}^i = \hat{P}_{t|t-1}^i K_t H_t^i \hat{P}_{t|t-1}^i \quad (33)$$

$$\hat{P}_{t|t-1}^i = F_t \hat{P}_{t|t}^i F_t^* + Q_t \quad (34)$$

$$F_t = \frac{\partial f(x_t^i)}{\partial x_t^i} \Big|_{x_t^i = \hat{x}_{t|t}^i} \quad (35)$$

The forward run, i.e. ECKF, is calculated from (28) to (35). To implement the smoother run, i.e. the backward recursion, the sequence of T observations from ECKF is required. This is based on the principle that the smoothed property of the latent variable is the probability at time instant t after a sequence of T observations, i.e. $p(x_t | z_0^T)$. Note initial values of the state and a $-posteriori$ estimate covariance in (36)-(37) are the final values of \hat{x}_t and $P_{t|t}$ in the sequence calculated by ECKF, respectively, i.e. $t = T$. Moreover, assume that $F_t = \frac{\partial f(x_t)}{\partial x_t} \Big|_{x_t = x_t^*}$ in (35) and f_t is linearized around x_t^* . Thus, $x_{t|t-1} = F \hat{x}_{t|t}$.

For $t = T, T-1, \dots, 0$, the smoothed error covariance and states are:

$$P_{t|T}^i = F_t P_{t-1|T}^i F_t^* + \left(K_t H_t^i P_{t-1|T}^i + K_t R_{e,t}^i \right) \left(K_t H_t^i P_{t|T-1}^i + K_t R_{e,t}^i \right)^* \quad (36)$$

$$\hat{x}_{t|T}^i = \hat{x}_{t|t-1}^i + P_{t|T}^i \quad (37)$$

where Q_t is the process noise correlation factor. The term $P_{t|T}^i$ is the local smoothed-updated $a - posteriori$ estimate covariance. Meanwhile, the desired initial estimate is $\hat{x}_{t|T}^i$, which estimates the state at t instants of time while the time sequence T is known.

Based on the formulated ECKS, a considerable amount of storage and latency is required. Additionally, the algorithm needs to wait for all $T+1$ instances before it can execute the backward run to obtain the state estimate. Therefore, one alternate option is to reduce the time size T . Else, the filter can be run in the forward direction only, i.e. run (28)–(35) for both the initial estimation and the EALM iteration with no latency.

Convergence of ECKF: This is proved in the Appendix. ■

Proof of ECKF Smoother: The proof structure follows Appendix of [13]. ■

3) Convergence of the EALM Algorithm:

The convergence of the proposed EALM algorithm can be proved by the following theorem (II.1).

Theorem II.1: In EALM algorithm, the ALM-step can also be formulated such that x_{t+1} is chosen to be any value of sequence $x_0^T \in z_t$ that maximizes $Q(x_0^T, x_t)$:

$$Q(x_{t+1}, x_t) \geq Q(x_0^T, x_t), \forall x_0^T \in z_t \quad (38)$$

where $x_{t+1} = x_t + \psi_{r_t}$, z_t is the parameter space containing the unknown parameters in the postulated form for the moments of the observation output as calculated in (20).

Once the oscillation observations and their respective maximum-likelihood estimates are determined among the local i -th sensors, they are collectively processed using a distributed architecture. Here, the master information filter collects all sensor observations at a central location. Subsequently, the updated covariance matrix and state estimates from the master filter are

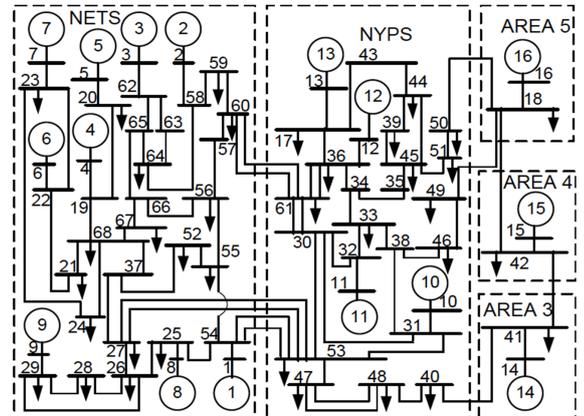


Fig. 2. Single-line diagram of the IEEE-68 Bus system [18]

fed back to all local metering locations. Such approach results in enhancing the modal estimations at those locations that are contaminated by noise and system perturbations. It is adopted here due to the uncertainty developed because of the random load fluctuations. This resultant data fusing architecture is used to quantitatively evaluate the role of synchrophasor measurements collected from each substation, and their overall tracking performance of electromechanical oscillations on the system. Detailed formulation is described in the next subsection.

Proof: This is proved in the Appendix. ■

4) Distributed Filtering Fusion:

According to the derivation of distributed filtering outlined in [13], the master filtering measurement is stated in the information form as:

$$P_{M,t|t}^{-1} \hat{x}_{M,t|t} = P_{M,t|t-1}^{-1} \hat{x}_{M,t|t-1} + H_{M,t}^* R_{M,t}^{-1} z_{M,t} \quad (39)$$

$$P_{M,t|t}^{-1} = P_{M,t|t-1}^{-1} + H_{M,t}^* R_{M,t}^{-1} H_{M,t}$$

The variable $P_{M,t|t}$ is the updated $a - posteriori$ estimate covariance oscilltion matrix. Whereas $P_{M,t|t-1}$ is the updated $a - priori$ estimate covariance matrix of the oscillations.

5) **Convergence of the Distributed fusion:** The proof of the convergence can be seen in Appendix of [13]. ■

III. IMPLEMENTATION AND EVALUATION

The proposed method was exhaustively tested under different network operating conditions. Two of the studies are presented in this paper. Test Case I analyzed synthetic synchrophasor measurements collected from IEEE 68 Bus system, which were simulated in DIGSILENT Power Factory ver.15 [19]. Monte Carlo simulations based on random noise variance were also generated to evaluate the tracking robustness of the proposed method. The proposed method is referenced with two mainstream techniques: 1) Kalman Filter [5], 2) Prony Analysis [6], and 3) EM-based KLPF [13]. Test Case II examined recorded synchrophasor measurements gathered from the New Zealand grid. All detections are computed using voltage phase angles sampled at 50 Hz.

A. Test Case I: IEEE 68-Bus System

The purpose of this study was to examine the monitoring capability of the proposed method under a mix of ringdown and ambient dynamics. The simulated system was based on a reduced order model of the New England and New York interconnection as shown in Fig. 2. There were sixteen synchronous generators, where generator 1 to 13 were each equipped with an exciter (DC1A) and a first order governor providing 4% droop. The remaining three generators were aggregated equivalent models of neighboring networks. To highlight the lightly damped inter-area oscillations, only generator 4 was installed with a power system stabilizer (STAB1). From modal analysis, three dominant electromechanical modes were identified. Their initial values were: 1) an inter-area mode with a frequency of 0.79 Hz and a damping ratio of 1.53%, 2) a local mode with a frequency of 1.08 Hz and a damping ratio of 1.55%, and 3) another local mode with a frequency of 1.12 Hz, with a damping ratio of 1.82%. A 60 second simulation was then conducted of which the network suffered the following two disturbances:

- Line 4-14 opened at 1 second and reclosed at 8 second.
- Load 18 and 51 both increased their active power demand by 10 % over a linearly ramped period of 5 second at 15 second interval.

To create a realistic operating environment, all loads were continuously perturbed with small-magnitude random fluctuations of up to 10 MW every second. Note instantaneous oscillatory parameters would vary slightly based on the system operating conditions. The grid dynamics captured at Bus 10 is shown in Fig. 3. For this test case, synchrophasor measurements captured from Bus 2, 3, 6, 10, 19, 23, 25, 29, 31, 33, 37, 41, 42, 45, 48, and 52 were utilized for EM-based FB-KLPF and the proposed method. Meanwhile, the predecessor Kalman Filter and Prony Analysis evaluated measurements collected from Bus 10, which exhibited dynamics of all three dominant modes.

The averaged estimated oscillatory parameters over a 10 second window size are listed in TABLE I. The comparatively higher estimation errors for all three methods in the first window were due to the transient line event. Overall, all methods were able to monitor the inter-area oscillation accurately as it was distinctly visible throughout the entire simulated timeframe. Since Kalman Filter and Prony Analysis were not originally designed to monitor multiple electromechanical modes of nearby frequencies, they suffered from mode-mixing issue during the entire simulation. The two local oscillations were treated as one mode in the curve fitting step. This was also reflected in the relatively higher MSE in comparison to the other two methods. In contrast, both EM-based FB-KLPF and the proposed method were able to track local oscillations with reasonable accuracy. The proposed method provided more accurate estimations than EM-based FB-KLPF. The statistical properties of the two schemes were also evaluated. In this case, random noise of various variances were applied of which two of the studies are shown in Fig. 4. Note the MSE values were averaged over a 5 second window. Referring to Fig. 4, EM-based FB-KLPF was less accurate under noisy condition. The error fluctuations throughout the monitoring windows indicated the method was less capable of dealing with random noisy conditions. This is due to the linear approximation of the unobserved latent vari-

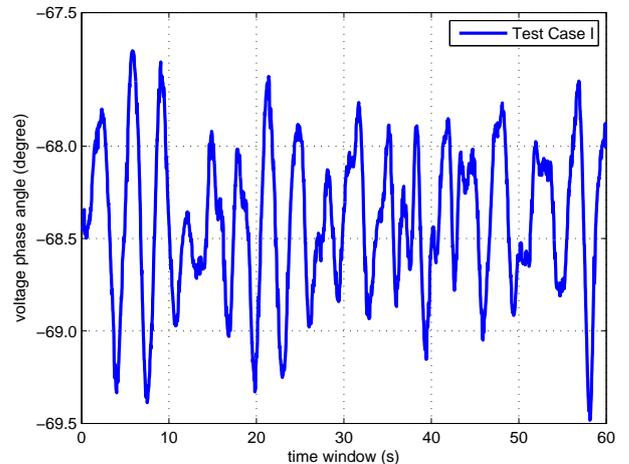


Fig. 3. Test Case I: Voltage phase angle profile of Bus 10

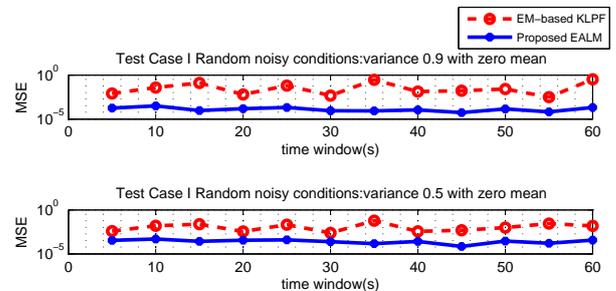


Fig. 4. Statistical tests showing MSE comparison for Test Case I

able. On the other hand, the non-linear nature of the proposed scheme allowed it to be less influenced by random noise. This owes to its property of handling the asymptotic covariance matrix by interpreting the indices of precision of estimation. Estimation errors of the proposed EALM-based ECKS were relatively smaller and constant as shown in Fig. 4. These statistical results also correlates with those in TABLE I, where the MSE errors of the proposed method are always lower than its predecessors; EM-based FB-KLPF and Kalman Filter. To further evaluate the statistical properties of the proposed method, Monte Carlo simulations were also conducted. Simulations were setup based on the procedures outlined in [20]. For this study, the 50 to 60 second window shown in Fig. 3 was used as there were no significant transient events during this period. Therefore, a total of 1200 Monte Carlo runs were performance for EM-based KLPF and the proposed method. Results are illustrated in Fig. 5 and 6. Both methods estimated all electromechanical modes quite accurate in the presence of random noise. However, the population of the proposed method was more concentric to the true values “o” than EM-based KLPF. Note the averaged true values over this time period were computed using non-parametric spectral analysis [20]. Therefore, the proposed EALM-based ECKS scheme demonstrated to improve the tracking resilience under noisy and ambient conditions.

B. Test Case II: New Zealand Network

The backbone of the New Zealand transmission infrastructure is based on 220 kV lines that are interconnected by HVDC links

TABLE I
TEST CASE I – IEEE 68 BUS SYSTEM: DETECTING MULTIPLE
ELECTROMECHANICAL OSCILLATIONS¹

| Time | 0 s–10 s | | | | | | | |
|------|----------------------|----------|----------------------|----------|----------------------|-------|----------------------|-------|
| | ζ_{PA} | f_{PA} | ζ_{KF} | f_{KF} | ζ_D | f_D | ζ_P | f_P |
| | 1.1 | 0.76 | 1.4 | 0.75 | 1.4 | 0.74 | 1.5 | 0.75 |
| | – | – | – | – | 1.4 | 1.01 | 1.5 | 1.03 |
| | 1.6 | 1.10 | 1.5 | 0.95 | 1.7 | 1.13 | 1.8 | 1.14 |
| MSE | 7.3×10^{-2} | | 9.5×10^{-2} | | 5.4×10^{-3} | | 1.4×10^{-3} | |
| Time | 10 s–20 s | | | | | | | |
| | ζ_{PA} | f_{PA} | ζ_{KF} | f_{KF} | ζ_D | f_D | ζ_P | f_P |
| | 1.2 | 0.74 | 1.4 | 0.70 | 1.4 | 0.76 | 1.5 | 0.79 |
| | – | – | – | – | 1.5 | 0.97 | 1.5 | 1.01 |
| | 1.1 | 1.09 | 1.4 | 1.12 | 1.7 | 1.12 | 1.8 | 1.14 |
| MSE | 7.2×10^{-2} | | 4.1×10^{-2} | | 2.9×10^{-3} | | 6.5×10^{-4} | |
| Time | 20 s–30 s | | | | | | | |
| | ζ_{PA} | f_{PA} | ζ_{KF} | f_{KF} | ζ_D | f_D | ζ_P | f_P |
| | 1.5 | 0.75 | 1.5 | 0.79 | 1.2 | 0.77 | 1.5 | 0.78 |
| | – | – | – | – | 1.4 | 0.99 | 1.5 | 1.01 |
| | 1.4 | 1.04 | 1.5 | 1.04 | 1.5 | 1.09 | 1.8 | 1.15 |
| MSE | 7.2×10^{-2} | | 5.4×10^{-2} | | 1.3×10^{-3} | | 4.4×10^{-4} | |
| Time | 30 s–40 s | | | | | | | |
| | ζ_{PA} | f_{PA} | ζ_{KF} | f_{KF} | ζ_D | f_D | ζ_P | f_P |
| | 1.4 | 0.74 | 1.4 | 0.74 | 1.3 | 0.75 | 1.5 | 0.78 |
| | – | – | – | – | 1.4 | 0.99 | 1.5 | 0.96 |
| | 1.7 | 1.03 | 1.5 | 0.96 | 1.8 | 1.15 | 1.8 | 1.16 |
| MSE | 7.3×10^{-2} | | 6.0×10^{-2} | | 8.8×10^{-3} | | 3.7×10^{-4} | |
| Time | 40 s–50 s | | | | | | | |
| | ζ_{PA} | f_{PA} | ζ_{KF} | f_{KF} | ζ_D | f_D | ζ_P | f_P |
| | 1.3 | 0.70 | 1.5 | 0.72 | 1.3 | 0.76 | 1.5 | 0.79 |
| | – | – | – | – | 1.4 | 1.02 | 1.5 | 1.02 |
| | 1.5 | 0.91 | 1.5 | 0.91 | 1.7 | 1.07 | 1.8 | 1.12 |
| MSE | 7.3×10^{-2} | | 6.0×10^{-2} | | 7.0×10^{-3} | | 3.7×10^{-4} | |
| Time | 50 s–60 s | | | | | | | |
| | ζ_{PA} | f_{PA} | ζ_{KF} | f_{KF} | ζ_D | f_D | ζ_P | f_P |
| | 1.0 | 0.74 | 1.4 | 0.71 | 1.3 | 0.78 | 1.5 | 0.80 |
| | – | – | – | – | 1.5 | 1.02 | 1.5 | 1.03 |
| | 1.5 | 1.09 | 1.5 | 1.09 | 1.7 | 1.12 | 1.8 | 1.14 |
| MSE | 7.2×10^{-2} | | 8.7×10^{-3} | | 1.2×10^{-3} | | 5.3×10^{-4} | |

¹In this table, ζ is the damping ratio i.e. $\zeta = \frac{-\sigma}{\sqrt{\sigma^2 + (2\pi f)^2}} \times 100$. The variable f is the frequency in hertz, and MSE is the mean-square error. Meanwhile subscripts PA, KF, D and P are the acronyms for Prony Analysis, Kalman filter, distributed and the proposed scheme, respectively.

between the North and South Islands. Recorded synchrophasor measurements, collected from North Makarewa (NMA) and Twizel (TWZ) substations between 11:14:40 to 11:19:20 on 30 July 2008, were used to evaluate the estimation capability of the proposed algorithm. In contrast to the previous test case, this study focused on the performance under distinct ringdown and ambient conditions. The active power dynamics captured at NMA over 280 second window is illustrated in Fig. 7. Measurements from 11:14:40 to 11:16:32 were considered as ringdown, while measurements from 11:16:32 to 11:19:20 were ambient. Based on non-parametric spectral analysis, four electromechanical oscillations were identified for each situation. The averaged ringdown oscillatory parameters over the 112 second time-frame were:

- Mode 1: A 0.44 Hz frequency with a 6.5% damping ratio.
- Mode 2: A 0.61 Hz frequency with a 6.1% damping ratio.

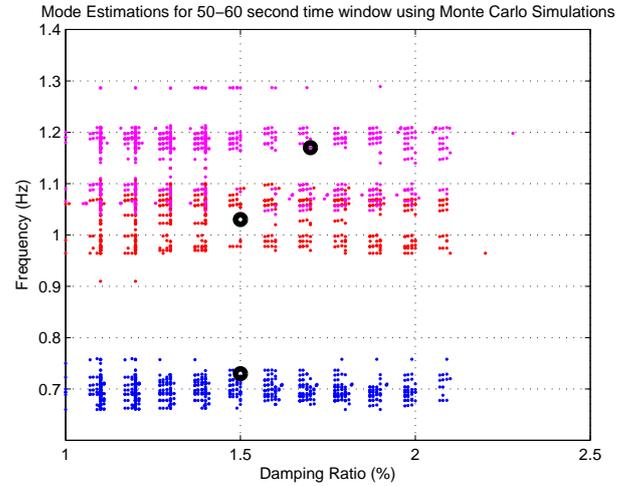


Fig. 5. Test Case I: Mode estimations of EM-based KLPF over 1200 Monte Carlo simulations

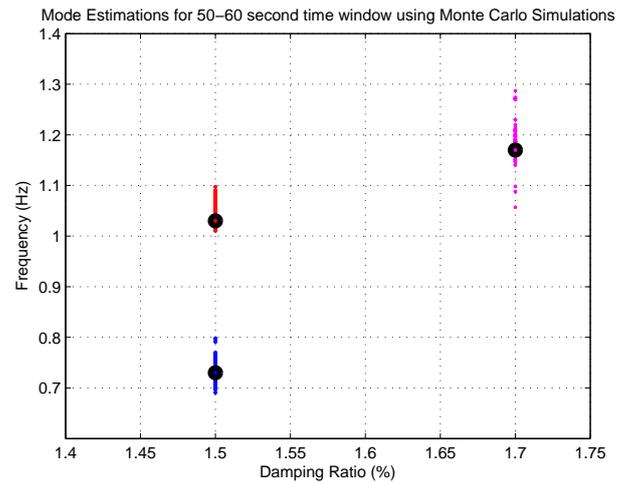


Fig. 6. Test Case I: Mode estimations of EALM-based ECKS over 1200 Monte Carlo simulations

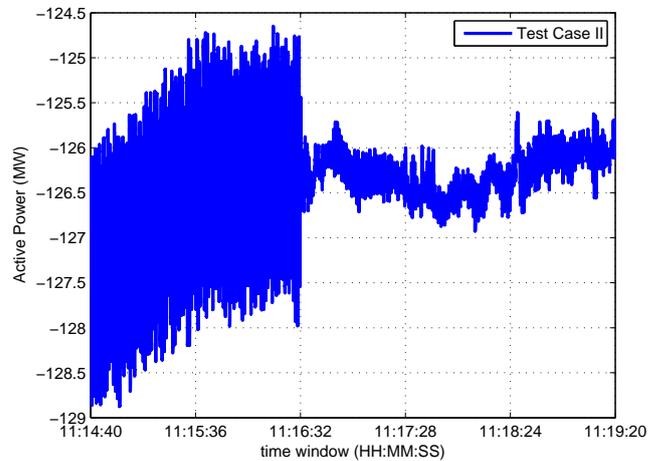


Fig. 7. Test Case II: Recorded active power at North Makarewa (NMA) substation

- Mode 3: A 0.75 Hz frequency with a 5.6% damping ratio.
- Mode 4: A 1.71 Hz frequency with a 2.3% damping ratio.

TABLE II
TEST CASE II – NEW ZEALAND GRID: ESTIMATION RESULTS OF
MULTIPLE ELECTROMECHANICAL OSCILLATIONS

| Time | 11:14:40–11:16:32 | | | | 11:16:32–11:19:20 | | | |
|------|----------------------|-------|----------------------|-------|----------------------|-------|----------------------|-------|
| | ζ_D | f_D | ζ_P | f_P | ζ_D | f_D | ζ_P | f_P |
| | 5.9 | 0.42 | 6.1 | 0.42 | 13.0 | 0.42 | 15.8 | 0.42 |
| | 5.4 | 0.61 | 5.8 | 0.60 | 9.8 | 0.63 | 13.6 | 0.65 |
| | 5.0 | 0.71 | 5.2 | 0.75 | 7.2 | 0.73 | 12.2 | 0.77 |
| | 2.9 | 1.69 | 2.3 | 1.71 | 5.2 | 1.73 | 7.3 | 1.76 |
| MSE | 1.7×10^{-3} | | 1.9×10^{-3} | | 8.9×10^{-2} | | 1.9×10^{-3} | |

Likewise, the averaged values of the ambient condition over the 168 second time-frame were:

- Mode 1: A 0.42 Hz frequency with a 16.0% damping ratio.
- Mode 2: A 0.64 Hz frequency with a 15.5% damping ratio.
- Mode 3: A 0.77 Hz frequency with a 12.1% damping ratio.
- Mode 4: A 1.76 Hz frequency with a 8.2% damping ratio.

Note that instantaneous oscillatory parameters would differ from these values due to continuous changing grid dynamics. The averaged results of ringdown and ambient conditions are listed in TABLE II. Although both methods achieved similar monitoring precisions under the ringdown condition, the more non-linear grid dynamics during the ambient situation caused EM-based KLPF to struggle. This could be observed in the less accurately estimated damping ratios of Mode 2 and 3, which led to an increased MSE. Nevertheless, the dominant Mode 4 could still be precisely monitored. In contrast, the estimation accuracy of the proposed method was maintained during the ambient situation. This demonstrated the proposed EALM-based ECKS is capable of monitoring electromechanical oscillations under both ringdown and ambient dominated conditions.

IV. CONCLUSIONS

In this paper, the tracking accuracy of the existing Kalman filter and the distributed EM-based FB-KLPF was improved by maximizing the estimation convergence in ALM and ECKS steps of the proposed algorithm. These enhancements provide more resistance and robustness against load fluctuations and random noise variance conditions. As a result, the ability to detect oscillations with similar frequencies could be further improved than its predecessors. One observed limitation of the proposed method is the computational complexity, which is increased by calculating the maximum likelihood at two stages. However, this may be resolved by assigning initial steady-state parameters to the algorithm and conduct both computations using high performance computing (HPC).

APPENDIX

1) Convergence of ECKF:

Assumption IV.1: There exists a known positive constant L_0 such that for any norm bounded $x_{1,t}, x_{2,t} \in \mathbf{R}^n$, the following inequality holds:

$$\|f(z_t, x_{1,t}) - f(z_t, x_{2,t})\| \leq L_0 \|x_{1,t} - x_{2,t}\| \quad (40)$$

Assumption IV.2: Assume that

$$A_t = \left. \frac{\partial f(x_t)}{\partial x_t} \right|_{x_t=x_t^*}, B_t = \left. \frac{\partial f(v_t)}{\partial v_t} \right|_{v_t=v_t^*} \quad (41)$$

Note that f_t is linearized around x_t^* and v_t^* , respectively.

Assumption IV.3: The transfer function matrix $H_t[sI - (A_t - K_t H_t)]^{-1} B_t$ is strictly positive real, where $K_t \in \mathbf{R}^{n \times r}$ is chosen such that $A_t - K_t H_t$ is stable.

For a given positive definite matrix $Q_t > 0 \in \mathbf{R}^{n \times n}$, there exists matrices $P_t = P_t^* > 0 \in \mathbf{R}^{n \times n}$ and a scalar R_t such that:

$$(A_t - K_t H_t)^* P_t (A_t - K_t H_t) = -Q_t \quad (42)$$

$$P_t B_t = H_t^* R_t \quad (43)$$

To estimate the state without process and measurement noises, the following equations are derived:

$$\hat{x}_t = A_t \hat{x}_t + g(z_t, x_t) + B_t \xi_t f(z_t, \hat{x}_t) + K_t (z_t - \hat{z}_t) \quad (44)$$

$$\hat{z}_t = H_t \hat{x}_t \quad (45)$$

where $\hat{x}_t \in \mathbf{R}^n$ is the state estimate, and the observation output is $z_t \in \mathbf{R}^r$. Note the pair (A, H) is observable. The non-linear term $g(z_t, x_t)$ depends on z_t and x_t , which are directly available. The term $f(z_t, \hat{x}_t) \in \mathbf{R}^r$ is a nonlinear vector function of z_t and \hat{x}_t . Lastly, $\xi_t \in \mathbf{R}$ is a parameter that changes unexpectedly when a noise occurs. Since it has been assumed that the pair (A, H) is observable, a gain matrix K_t can be selected such that $A_t - K_t H_t$ is a stable matrix. It is defined as:

$$e_{x,t} = x_t - \hat{x}_t, \quad e_{z,t} = z_t - \hat{z}_t \quad (46)$$

Then, the error equations can be given by:

$$e_{x,t+1} = (A_t - K_t H_t) e_{x,t} + B[\xi_t f(u_t, z_t, x_t) - \xi_{H,t} f(z_t, \hat{x}_t)], \quad (47)$$

$$e_{z,t} = H_t e_{x,t} \quad (48)$$

The variable $\xi_{H,t} \in \mathbf{R}$ is a measured parameter of the unexpected noise. Meanwhile, the convergence of the above filter is guaranteed by the following theorem IV.1:

Theorem IV.1: Under the assumption (IV.3), the filter is asymptotically convergent when no noise occurs ($\xi_t = \xi_{H,t}$), i.e. $\lim_{k \rightarrow \infty} e_{z,t} = 0$.

Proof: Consider the following Lyapunov function,

$$V(e_t) = e_{x,t}^* P e_{x,t} \quad (49)$$

where P is the solution of (42). The variable Q is chosen such that $\rho_1 = \lambda_{\min}(Q) - 2\|H\| \cdot |R| \xi_{H,t} L_0 > 0$ is along the trajectory of the noise-free system of (47). Therefore, the corresponding Lyapunov difference along the trajectories e_t is:

$$\begin{aligned} \Delta V &= E\{V(e_{t+1}|e_t, p_t)\} - V(e_t) \\ &= E\{e_{t+1}^* P_t e_{t+1}\} - e_t^* P_t e_t \\ &= (A_e e_t + B_L u_e)^* P_t (A_e e_{x,t} + B_L u_e) - e_{x,t}^* P_t e_{x,t} \\ &= e_t^* [(P_t (A_t - K_t H_t) + (A_t - K_t H_t)^* P_t) \\ &\quad + P_t B \xi_H [f(z_t, x_t) - f(z_t, \hat{x}_t)]] e_t \end{aligned} \quad (50)$$

From assumption (IV.1) and the system described by (42), one can further obtain that:

$$\begin{aligned} \Delta V &\leq -e_{x,t}^*(t) Q e_{x,t}(t) + 2\|e_y(t)\| \cdot |R| \xi_{H,t} L_0 \|e_{x,t}(t)\| \\ &\leq -\rho_1 \|e_t\|^2 < 0 \end{aligned} \quad (51)$$

Thus, $\lim_{k \rightarrow \infty} e_{x,t}(t) = 0$ and $\lim_{k \rightarrow \infty} e_y(t) = 0$. This completes the proof. ■

A. Proof of theorem II.1

Suppose the observed data output z_t has a probability density function and a penalty function ψ_{r_t} for the unobserved latent

variable H_t , i.e. $H_t = f(z_t, \psi_{r_t})$. The objective here is to maximize the likelihood $L_{x_t} = f(z_t, \psi_{r_t})$. First, let $t(x_t|z_t, \psi_{r_t}) = \frac{f(x_t, \psi_{r_t})}{f(z_t, \psi_{r_t})}$ be the conditional density of $X = x_t$ given $Z = z_t$ and ψ_{r_t} . The function $f(x_t, \psi_{r_t})$ represents the probability density function of the random vector X corresponding to the complete-data vector. Similarly, $f(z_t, \psi_{r_t})$ represents the probability density function of the random vector Z containing the unobserved parameters, which corresponds to the observed data output z_t . Based on these assumptions, the complete-data log likelihood can be expressed as:

$$\begin{aligned} \log L_{x_t}(\psi_{r_t}) &= \log f(x_t, \psi_{r_t}) \\ &= \log L_{x_t}(\psi_{r_t}) + \log t(x_t|z_t, \psi_{r_t}) \end{aligned} \quad (52)$$

The expected value on both sides of (52) is applied with respect to the conditional distribution $x_t|z_t$. An estimate $\hat{\psi}_{r_t}$ for ψ_{r_t} is assumed such that:

$$Q(\psi_{r_t}, \hat{\psi}_{r_t}) = \log L(\psi_{r_t}) + H_t(\psi_{r_t}, \hat{\psi}_{r_t}) \quad (53)$$

where $H_t(\psi_{r_t}, \hat{\psi}_{r_t}) = \mathbf{E}_{\hat{\psi}_{r_t}}(\log t(X|z_t, \psi_{r_t})|z_t)$. It follows from (38) and (53) that

$$\begin{aligned} \log L_{x_t}(\hat{\psi}_{r_{t+1}}) - \log L_{x_t}(\hat{\psi}_{r_t}) &= [Q(\hat{\psi}_{r_{t+1}}, \hat{\psi}_{r_t}) \\ &- Q(\hat{\psi}_{r_t}, \hat{\psi}_{r_t})] - [H_t(\hat{\psi}_{r_{t+1}}, \hat{\psi}_{r_t}) - H_t(\hat{\psi}_{r_t}, \hat{\psi}_{r_t})] \end{aligned} \quad (54)$$

where $H_t(\hat{\psi}_{r_{t+1}}, \hat{\psi}_{r_t}) \leq H_t(\hat{\psi}_{r_t}, \hat{\psi}_{r_t})$. From (38) and Jensen's inequality, the first difference on the right-hand side of (54) is non-negative. Thus, the likelihood function is not decreased to $\log L_{x_t}(\hat{\psi}_{r_{t+1}}) \geq \log L_{x_t}(\hat{\psi}_{r_t})$. The convergence of EALM-based ECKS algorithm is $\log L_{x_t}(M(x_t)) \geq \log L(x_t)$. The variable M is a mapping function defined by ψ_{r_t} . This can be represented with equality if and only if:

$$\begin{aligned} Q(M(x_{t+1}), x_t) &= Q(x_t, x_t), \text{ and} \\ t(x_t|z_t, M(x_t)) &= t(x_t|z_t, x_t) \end{aligned} \quad (55)$$

As a result, the likelihood function increases at each iteration of the EALM-based ECKS algorithm with a defined mapping function and penalty coefficients until the condition for equality is satisfied. A fixed point of the iteration is reached.

ACKNOWLEDGMENTS

The authors thank Transpower New Zealand for providing synchrophasor measurements, Professor James L. Kirtley from Massachusetts Institute of Technology and Professor Seddik M. Djouadi from University of Tennessee Knoxville for their insights and suggestions.

REFERENCES

- [1] G. Rogers, Power system oscillations, *Norwell, MA: Kluwer*, 2000.
- [2] P. Kundur, "Power system stability and control," London, U.K.: McGraw-Hill, 1994.
- [3] J. C.-H. Peng and N. K. C. Nair, "Comparative assessment of Kalman filter and Prony methods for power system oscillation monitoring," *IEEE Pow. & Ener. Gen. Meet.*, pp. 1–8, Jul. 2009.
- [4] R. W. Wies, J. W. Pierre and D. J. Trudnowski, "Use of least-mean squares (LMS) adaptive filtering technique for estimating low-frequency electromechanical modes in power systems," *Proc. IEEE Power Eng. Soc. General Meeting*, vol. 2, pp. 1863–1870, 2004
- [5] P. Korba, "Real-time monitoring of electromechanical oscillations in power systems: First findings," *IET, Generation, Transm. Distrib.*, vol. 1 no. 1, pp. 80–88, Jan. 2007.
- [6] J. Hauer, C. Demecure and L. Scharf, "Initial results in Prony analysis of power system response signals," *IEEE Trans. Power Syst.*, vol. 5 no. 1, pp. 80–89, Feb. 1990.

- [7] D. J. Trudnowski, J. Pierre, N. Zhou, J. Hauer, and M. Parashar, "Performance of three mode-meter block-processing algorithms for automated dynamic stability assessment," *IEEE Trans. Power Syst.*, vol. 23 no. 2, pp. 680–690, May 2008.
- [8] N. Kakimoto, M. Sugumi, T. Makino and K. Tomiyama, "Monitoring of interarea oscillation mode by synchronized phasor measurement," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 260–268, Feb 2006.
- [9] A. R. Messina, and V. Vittal, "Nonlinear, non-stationary analysis of inter-area oscillations via Hilbert spectral analysis," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1234–1241, Aug. 2006.
- [10] J. L. Rueda, C. A. Juarez and I. Erlich, "Wavelet-based analysis of power system low-frequency electromechanical oscillations," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 260–268, Aug. 2011.
- [11] J. Turunen, J. Thambirajah, M. Larsson, C. P. Bikash, N. F. Thornhill, L. C. Haarla, W. W. Hung, A. M. Carter and T. Rauhala, "Comparison of three electromechanical oscillation damping estimation methods," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2398–2407, Nov. 2011.
- [12] "Identification of electromechanical modes in power systems", *IEEE Task Force Report*, Special Publication TP462, Jun 2012.
- [13] H. M. Khalid and J. C.-H. Peng, "Improved recursive electromechanical oscillations monitoring scheme: A novel distributed approach," *IEEE Trans. Pow. Syst.*, vol. 30, no. 2, pp. 680–688, Mar. 2015.
- [14] J. C.-H. Peng and N. K. C. Nair, "Enhancing Kalman filter for tracking ringdown electromechanical oscillations," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 1042–1050, May 2012.
- [15] S. Kullback and R. A., Leibler, "On information and sufficiency," *Annals of Mathematical Statistics* vol. 22, no. 1, pp. 79–86, 1951.
- [16] T. Needham, "A visual explanation of Jensen's inequality," *American Mathematical Monthly*, vol. 100, no. 8, pp. 768–771, 1993.
- [17] D. P. Bertsekas, "Constrained optimization and Lagrange multiplier methods," *Athena Scientific, Belmont, Massachusetts*, 1996.
- [18] B. Pal and B. Chaudhuri, "Robust control in power systems," *Springer Science and Business Media, Language Arts and Disciplines - 190 pages*, Jun 2005.
- [19] DiGSILENT, DiGSILENT technical documentation ver. 15, *DiGSILENT GmbH*, 2013.
- [20] "Inter-area oscillations in power systems: A nonlinear and nonstationary perspective," *Springer Power Electronics and Power Systems Series*, Editors: Messina, Arturo Roman, 2009.
- [21] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE* vol. 70, no. 9, pp. 1055, 1982.
- [22] J. G. Proakis and D. G. Manolakis, "Digital signal processing," *Upper Saddle River, NJ: Prentice-Hall*, pp. 910–913, 1996.



Haris M. Khalid (M'13) received his M.S and Ph.D. degrees from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Kingdom of Saudi Arabia, in 2009 and 2012, respectively. He has worked as a research fellow at Distributed Control Research Group at KFUPM. Since 2013, he has been working as a Postdoctoral Researcher with Department of Electrical Engineering and Computer Science at Masdar Institute of Science and Technology (MI) collaborated with MI-MIT Cooperative Program. His research interests are power systems, signal processing, fault diagnostics, filtering, estimation, performance monitoring and battery management systems.



Jimmy C. -H. Peng (S'05-M'12) received the B.E. (Hons.) and Ph.D. degrees from the University of Auckland, Auckland, New Zealand, in 2008 and 2012, respectively. In 2012, he joined the Department of Electrical Engineering and Computer Science at Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates, as an Assistant Professor of Electrical Power Engineering. He was a Visiting Scientist with the Research Laboratory of Electronics (RLE) and a Visiting Assistant Professor with the MIT-MI Cooperative Program at Massachusetts Institute of Technology (MIT), in 2013 and 2014, respectively. His research interests are power system stability, synchrophasor measurements, real-time system identification, and high performance computing.