Deep Learning High-Gain Observer for Power System Dynamic State Estimation

Irenge Baguma Raoul[1], Tarek Ahmed-Ali[1,2], Benatia Mohamed-Amin[1], Nkwawo Homère[3] and Jean-Marie Janik[2]

***Abstract*— This paper introduces a novel approach for designing dynamic state estimators for electric power systems based on observer theory. The proposed method integrates deep learning techniques to bypass the analytical computation of Jacobian matrices and their inverses, which becomes particularly intractable in the presence of highly nonlinear output functions. An autoencoder-based architecture is employed to simultaneously learn the coordinate transformation and perform state reconstruction. Consequently, a new class of discrete-time high-gain observers is developed, offering a flexible and computationally efficient framework for dynamic state estimation in electric power systems.**

I. INTRODUCTION

Modern power systems are undergoing rapid evolution, driven by the large-scale integration of renewable energy sources, the development of microgrids, and increasing load variability. This paradigm shift renders the long-standing quasi-steady-state operating assumption obsolete for power system analysis and supervision. In this context, traditional Supervision Control and Data Acquisition (SCADA) systems reach their limits due to their low data acquisition rates and lack of time synchronization [1].

The advent of Phasor Measurement Units (PMUs) has overcome these limitations by providing synchronized, high-frequency measurements. This technological advancement has fostered the development of Dynamic State Estimation (DSE), which has become a central tool for real-time analysis and monitoring of power networks. Among existing methods, Kalman filter–based approaches and their variants have dominated both the literature and industrial applications, owing to their maturity and proven effectiveness [2].

Nevertheless, nonlinear observers constitute a promising alternative, particularly when the system exhibits strong nonlinearities a situation frequently encountered in power systems. Several classes of observers, including H_∞ observers [2], [3], sliding-mode observers [1], Gain-Scheduled Nonlinear Observer [4], and unknown-input observer [5], have demonstrated their relevance in DSE applications. More recently, continuous–discrete high-gain observers have been successfully applied in [6], showing exponential convergence of the estimation error and strong robustness with respect to disturbances and measurement noise.

Despite these advances, one of the main scientific bottlenecks limiting the large-scale adoption of nonlinear observers for dynamic state estimation in power networks lies in the need to analytically compute Jacobian matrices and their inverses, which often arise during the observer design process. This task becomes particularly challenging when the output expression is highly nonlinear, as in the case of active or reactive power derived from a fourth-order synchronous generator model.

In this context, Péralez and Nadri recently proposed in [7], [8] the use of deep neural networks to numerically approximate coordinate transformations in the design of KKL (Kazantzis-Kravaris-Luenberger) observers, thereby avoiding analytically intensive and complex computations. Inspired by these works, the present paper proposes a hybrid approach combining a discrete high-gain observer with deep learning techniques for dynamic state estimation in power systems with highly nonlinear outputs. This contribution extends our previous work presented in [6], in which the Jacobian structure was sufficiently simple to allow direct analytical computation. Here, the use of active power as the measured output leads to a significantly more complex Jacobian, making its analytical derivation difficult, if not impractical.

Accordingly, the main contributions of this paper are summarized as follows: (1) The development of a hybrid architecture combining a high-gain observer and neural networks for dynamic state estimation in power systems, tailored to sampled PMU measurements; (2) The formulation of a simplified observer design methodology that enables learning Jacobian transformations and their inverses—an essential step in high-gain observer synthesis thereby circumventing the complexity of analytical computations; (3) The validation of the proposed approach on a fourth-order synchronous generator model connected to an infinite bus, using active power as the measured output, along with an assessment of robustness to measurement noise.

The remainder of this paper is organized as follows. Section II presents the studied model and its generalization. Section III details the neural-network-based high-gain observer design methodology. Section IV provides simulation results. Finally, Section V concludes the paper and outlines future research directions.

[1] CESI-LINEACT UR 7527, 80 Avenue Edmund Halley, Saint Etienne du Rouvray, 76800, Rouen, France `rirengebaguma@cesi.fr`

[2]UNICAEN, UR – Laboratoire des systèmes d'énergie et microélectronique (LSEM), 6, Boulevard du Maréchal Juin, 14050 CAEN cedex 04, Caen , France

[3]Université Sorbonne Paris Nord, UR – Laboratoire des systèmes d'énergie et microélectronique (LSEM), 6, Boulevard du Maréchal Juin, 14050 CAEN cedex 04, Caen , France

II. Studied Model and Generalization

As in [6], we consider an observer designed to estimate the states of a fourth-order synchronous generator model connected to an infinite bus, where the measured output corresponds to the active power. The studied model is expressed as follows:

$$\begin{cases} \dot{x}_1 = \omega_0 x_2, \\ \dot{x}_2 = \dfrac{1}{J}\Big[T_m - \dfrac{V_t}{x'_d} x_3 \sin(x_1) \\ \qquad - \dfrac{V_t^2}{2}\Big(\dfrac{1}{x_q} - \dfrac{1}{x'_d}\Big)\sin(2x_1) - Dx_2\Big] \\ \dot{x}_3 = \dfrac{1}{T'_{do}}\left[E_{fd} - x_3 - (x_d - x'_d)\frac{x_3 - V_t\cos(x_1)}{x'_d}\right], \\ \dot{x}_4 = \dfrac{1}{T'_{qo}}\left[-x_4 + (x_q - x'_q)\frac{V_t\sin(x_1)}{x_q}\right], \\ y = \dfrac{V_t}{x'_d} x_3 \sin(x_1) + \dfrac{V_t^2}{2}\left(\dfrac{1}{x_q} - \dfrac{1}{x'_d}\right)\sin(2x_1) \end{cases} \tag{1}$$

The state and input vectors are defined as:

$$x = \begin{bmatrix} \delta \\ \Delta\omega \\ e'_q \\ e'_d \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \quad u = \begin{bmatrix} V_t \\ T_m \\ E_{fd} \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}. \tag{2}$$

where δ is the rotor angle of the generator,$\Delta\omega$ is the rotor speed deviation,$\omega_0 = 2\pi f_0$ is the synchronous speed of the rotor, e'_q and e'_d are the transient voltages along the q and d axes,T_e is the electromagnetic torque (or electrical output power),T_m is the mechanical torque applied to the generator,E_{fd} is the internal excitation voltage,D is the damping coefficient, x_d and x_q are the synchronous reactances along the d and q axes, J is the rotor inertia constant, x'_d and x'_q are the transient reactances along the d and q axes (in per unit),T'_{do} and T'_{qo} are the open-circuit time constants.

A. Problem Formulation

We rewrite relation (1) in a general form:

$$\begin{cases} \dot{x} = f(x,u), \\ y = h(x), \end{cases} \tag{3}$$

where $x \in \mathbb{R}^{d_x}$ denotes the state to be estimated, $y \in \mathbb{R}^{d_y}$ is the measured output, and $u \in \mathbb{R}^{d_u}$ is a known control input. Although the functions f and h are assumed to be unknown during online operation (data-driven approach), this paper relies on their mathematical models defined in (1) to generate, offline, training data in the form of batches of output trajectories $Y_{1:K}$, as well as the associated state trajectories $X_{1:K}$, sampled with a time step μ_k. It is worth noting that such data could also be obtained from real measurements. Moreover, it should be emphasized that the functions f and h are not directly used within the observer architecture.

The objective is to design a discrete high-gain observer capable of reconstructing the system state x in (3) without explicitly computing the Jacobian matrix or its inverse during the design phase.

III. Design of the High-Gain Observer Based Neural-Network

The first step in the observer design consists of performing a change of variables that maps the original state-space model into a new coordinate system. Starting from model (3), we consider the following coordinate transformation:

$$\begin{cases} \dot{\xi}_1(t) = \xi_2(t), \\ \dot{\xi}_2(t) = \xi_3(t), \\ \vdots \\ \dot{\xi}_{n-1}(t) = \xi_n(t), \\ \dot{\xi}_n(t) = \tilde{f}(\xi, u, \dot{u}, \ldots, u^{(n-r)}), \\ y(t) = \xi_1(t), \end{cases} \tag{4}$$

where r denotes the relative degree of the system.

In this new coordinate system, the state-space model can be expressed compactly as follows:

$$\begin{cases} \dot{\xi}(t) = A\xi(t) + B\tilde{f}(\xi, u, \dot{u}, \ldots, u^{(n-r)}), \\ y(t) = C\xi(t), \end{cases} \tag{5}$$

where the matrices A, B, and C are defined as

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

.

A. Continuous High-Gain Observer

Since the system states remain bounded, the estimated states $\hat{x}$ are also bounded. Consequently, the nonlinear function f can be regarded as a bounded perturbation. It follows that the estimation error converges to a bounded ball, whose radius can be reduced by increasing the parameter θ. In this context, Hassan Khalil showed in [10] that the high-gain observer suited to systems of the form (5) can be simplified as follows.

$$\begin{cases} \dot{z}(t) = Az(t) - \theta\Delta^{-1}L\big(z_1(t) - y(t)\big), \\ \dot{\hat{x}}(t) = M^{-1}\dot{z}(t), \end{cases} \tag{6}$$

where z denotes an estimate of ξ, L is the observer gain chosen such that $\tilde{A} = A - LC$ is Hurwitz, and M denotes the Jacobian matrix of the coordinate transformation. The scaling matrix Δ is defined as $\Delta = \mathrm{diag}\big(I_n,\ \theta I_n,\ \ldots,\ \theta^{d_z-1}I_n\big)$ with $\theta \geq 1$, a design parameter that determines the convergence speed of the observer. It guarantees the convergence of the error within a bound of order $1/\theta$.

B. Discrete High-Gain Observer

Given the discrete nature of the data obtained from PMUs and the use of neural networks for observer design, system (3) is discretized using the Euler method in order to obtain a formulation compatible with the estimation approach. Let μ denote the sampling time. The Euler method allows one to approximate the continuous dynamics $\dot{x}(t) = f(x(t), u(t))$ according to the following discretization scheme:

$$x_{k+1} = x_k + \mu f(x_k, u_k)$$

We thus obtain the following discrete-time model:

$$\begin{cases} x_{k+1} &= g(x_k, u_k), \\ y_k &= h(x_k), \end{cases} \tag{7}$$

where $k \in \mathbb{N}$ denotes the discrete-time index.

The observer best suited to the system described by (7) is obtained by applying Euler discretization to the continuous-time observer given in (6). This leads to the following discrete-time high-gain observer:

$$\begin{cases} z_{k+1} &= \bar{A} z_k + \bar{B} y_k, \\ \hat{x}_k &= T^{-1}(z_k), \end{cases} \tag{8}$$

where $\bar{A} = I + \mu\left(A - \theta\Delta^{-1}LC\right), \qquad \bar{B} = \mu\theta\Delta^{-1}L.$

Here, $I \in \mathbb{R}^{d_z \times d_z}$ denotes the identity matrix, T^{-1} represents the inverse coordinate transformation mapping z_k to the estimated state $\hat{x}_k$, and μ denotes the sampling period.

The structure of relation (8) allows us to draw inspiration from the recent works of Péralez and Nadri [7], [8] to design a discrete high-gain observer based on deep learning using neural networks.

C. Design of the Discrete High-Gain Observer Based on Deep Learning

The chosen neural network architecture is that of an autoencoder. The encoder performs the coordinate transformation from the state-space model to the latent space, while the associated decoder ensures the reconstruction of the states. However, this reconstruction relies on a switching strategy between two types of observers, both based on model (8), in order to balance measurement noise robustness with convergence speed.

1) Coordinate Transformation: The coordinate transformation consists of identifying the mapping T that allows transitioning from the state-space model to a latent space consistent with the observer dynamics. Consequently, the desired latent space, defined as $z = T(x)$, must satisfy the observer dynamics given by equation (8). The following loss function is introduced to enforce this constraint on the model:

$$L_{\text{dyn}} = \|T(x_{k+1}) - z_{k+1}\|, \tag{9}$$

2) State Reconstruction: This step consists of identifying the transformation T^{-1} that maps the latent space—constrained to follow the observer dynamics—back to the estimated state space. This transformation ensures reconstruction of the original states by learning the inverse of the Jacobian matrix. Following the approaches proposed in [7], [8], we adopt a multi-observer strategy based on a switching principle to learn this transformation. The associated reconstruction loss is then expressed as:

$$L_{\text{recon}} = \|x_k - T^{-1}(z_k)\|, \tag{10}$$

where $\|\cdot\|$ denotes the Mean Squared Error (MSE).

Thus, combining the two loss terms L_{dyn} and L_{recon} enables the model to simultaneously learn a linear representation of the system dynamics and an accurate reconstruction of the original state.

3) Learning Method for T^{-1}: The learning of this transformation, which avoids the explicit computation of the often complex inverse of the Jacobian matrix, is based on the multi-observer principle. It combines the advantages of a transient observer and an asymptotic observer through a switching approach. The strategy consists of identifying an optimal transition between these two methods based on their observed performance. A switching variable $m \in \{1, 2\}$ then leads to the following computational structure of the observer, applied to a sequence of observations Y_k:

$$z_0 = \Gamma^{(m)}, \tag{11}$$

$$z_{k+1} = \bar{A} z_k + \bar{B} y_k, \tag{12}$$

$$\hat{x}_k = \psi^{(m)}(z_k) \tag{13}$$

The functions $\Gamma^{(m)}$ and $\psi^{(m)}$ depend on the desired system behavior: asymptotic for $m = 1$ and transient for $m = 2$. The function $\Gamma^{(m)}$ represents an initialization function for the latent state, defined based on the information available at the time of initialization.

a) Asymptotic Observer: The structure of this observer is shown in Figure 1. The latent space is initialized according to

$$z_0 = \Gamma^{(1)} \triangleq T(x_0)$$

so that the dynamics associated with $\bar{A}$ do not need to be particularly fast. During execution, at time k, the latent state can be re-initialized from the estimate provided by the transient observer. The transformation $\psi(1)$, responsible for reconstructing the original state x, is defined as the inverse transformation, i.e.,

$$\psi^{(1)} \equiv T^{-1}.$$

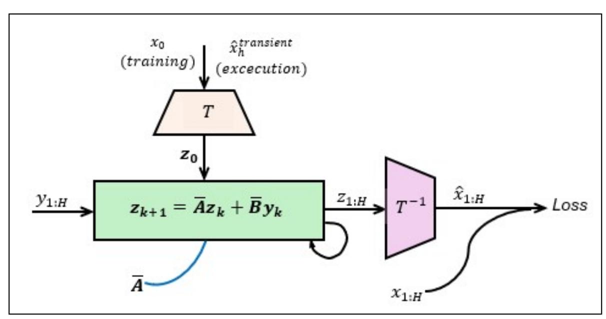


Fig. 1: Structure of the Asymptotic Observer in [8]

The complete asymptotic model is trained end-to-end by jointly optimizing the two components $\{T, T^{-1}\}$, while the matrix $\bar{A}^{(1)}$ is considered constant during training.

b) Transient Observer: To improve the transient behavior of the observer, the variable z is initialized as close as possible to $T(x_0)$ by exploiting the information available at runtime. To this end, the measured output y_0 is encoded into the latent space through a parametrized function Γ_α, such that

$$z_0 = \Gamma^{(2)} \triangleq \Gamma_\alpha(y_0).$$

The transient variant also learns an extension $\psi^{(2)} \triangleq T^*$, which is intended to mitigate the peaking phenomenon resulting from the fact that z is initialized solely from the measurement y_0.

The complete transient model is trained end-to-end by jointly optimizing the components $\{\Gamma^{(2)}, T^*\}$.

The selected neural network architecture is designed to learn the transformation $T^* : Z \to X$ such that

$$T^*(z) = T^{-1}(z), \quad \forall z \in Z \cap T(\chi).$$

As in the case of the asymptotic observer, the matrix $\bar{A}$ is assumed to be constant, since it is fixed by the observer design and is not learned. It is identical to the one used for the asymptotic observer.

The architecture of the transient observer is illustrated in Fig. 2.

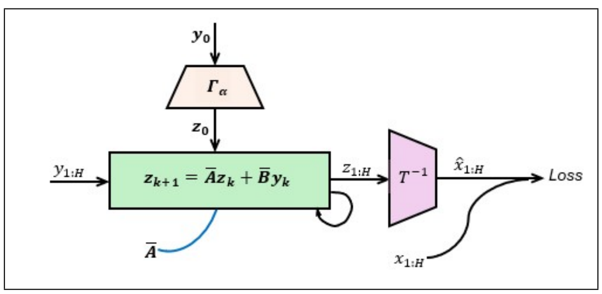


Fig. 2: Structure of the Transient Observer in [8]

4) Switching Approach: The switching between the transient observer and the asymptotic observer is performed dynamically according to internal performance criteria. At each time instant t_k, internal errors are evaluated by computing two scalar quantities $v_k^{(1)}$ and $v_k^{(2)}$ from the correction residues $\varepsilon_k^{(m)}$, and the observer corresponding to the smallest $v_k^{(m)}$ is selected. Based on the work in [11],the monitoring variable $v^{(m)} \in \mathbb{R}_{\geq 0}$ is defined by the following dynamics:

$$\dot{v}^{(m)} = -\sigma v^{(m)} + \left(\varepsilon^{(m)}\right)^\top \Lambda \varepsilon^{(m)} \tag{14}$$

where $\sigma \in (0, \gamma)$ is a design parameter. By applying the Euler discretization method with a sampling time μ, the following discrete-time representation is obtained:

$$v_{k+1}^{(m)} = \alpha v_k^{(m)} + \left(\varepsilon_k^{(m)}\right)^\top \Lambda_1 \varepsilon_k^{(m)} \tag{15}$$

where $\alpha = 1 - \mu\sigma$ and $\Lambda_1 = \mu\Lambda$.

In particular, if Λ_1 is chosen as the identity matrix, the above equation (15) simplifies to

$$v_{k+1}^{(m)} = \alpha v_k^{(m)} + \left\|\varepsilon_k^{(m)}\right\|^2. \tag{16}$$

where $0 \leq \alpha \leq 1$ is a forgetting factor, and $\varepsilon_k^{(m)}$ denotes a consistency measure(real measure) between two state reconstructions. The first reconstruction is obtained from the actual measurement y_k, given by: $\hat{x}(y_k) = T^{-1}\left(\bar{A}z_k + \bar{B}y_k\right)$,while the second is obtained from the output associated with the estimated state $\hat{y}_k = y_{\text{obs},k}$, defined as: $\hat{x}(y_{\text{obs},k}) = T^{-1}\left(\bar{A}z_k + \bar{B}y_{\text{obs},k}\right)$. where $y_{\text{obs},k} \triangleq h(\hat{x}_k)$ is the predicted measurement generated by the observer from its own estimated state.

Accordingly, $\varepsilon_k^{(m)}$ is defined as:

$$\begin{aligned} \varepsilon_k^{(m)} &= \hat{x}(y_k) - \hat{x}(y_{\text{obs},k}) \\ &= T^{-1}\left(\bar{A}z_k + \bar{B}y_k\right) - T^{-1}\left(\bar{A}z_k + \bar{B}\hat{y}_k\right). \end{aligned} \tag{17}$$

Thus, if $\varepsilon_k^{(m)} \approx 0$, one can conclude that the dynamics $z \mapsto \hat{x}$ are stable and coherent with respect to the input y.

a) Switching Mechanism: The hybrid observer selects, at each time instant, the observer exhibiting the smallest performance measure. It is defined as follows:

$$\hat{x}_k = \begin{cases} \hat{x}_k^{(1)}, & \text{if } v_k^{(1)} \leq v_k^{(2)}, \\ \hat{x}_k^{(2)}, & \text{otherwise}, \end{cases} \tag{18}$$

This strategy guarantees an optimal trade-off between convergence speed and robustness by adapting in real time to system conditions (noise, nonlinearities, etc.). It exploits the fast convergence of the transient mode during the initial phase, and the stability of the asymptotic mode in the long term.

b) Training: Model training is carried out by minimizing the loss functions (10) and (9) over a large dataset $\mathscr{D} = \{(x_k, x_{k+1})\}$, automatically generated from the discrete-time model of the considered system. The following algorithm illustrates the training procedure for learning T and T^{-1}.

IV. Simulation Results

Let us now apply the proposed observer to the selected model. We begin by discretizing the differential equation (1) using the explicit Euler method. We thus obtain the following discrete-time model:

$$\begin{cases} x_{1,k+1} = x_{1,k} + \mu\left(\omega_0 x_{2,k}\right) \\ x_{2,k+1} = x_{2,k} + \dfrac{\mu}{J}\Big[T_m - \dfrac{V_t}{x'_d} x_{3,k}\sin(x_{1,k}) \\ \qquad - \dfrac{V_t^2}{2}\left(\dfrac{1}{x_q} - \dfrac{1}{x'_d}\right)\sin(2x_{1,k}) - Dx_{2,k}\Big] \\ x_{3,k+1} = x_{3,k} + \dfrac{\mu}{T'_{do}}\Big[E_{fd} - x_{3,k} \\ \qquad - (x_d - x'_d)\dfrac{x_{3,k} - V_t\cos(x_{1,k})}{x'_d}\Big] \\ y_k = \dfrac{V_t}{x'_d} x_{3,k}\sin(x_{1,k}) + \dfrac{V_t^2}{2}\left(\dfrac{1}{x_q} - \dfrac{1}{x'_d}\right)\sin(2x_{1,k}) \end{cases} \tag{19}$$

State x_4 does not appear in (19) because it is not observable from the output $y = P_t$. Indeed, by computing the successive Lie derivatives of the output, it can be verified

Algorithm 1 Learning T and T^{-1}

Initialize: Choose latent dimension z_{dim}, output dimension y_{dim}, and state dimension x_{dim}.
Sample a dataset $\mathscr{D}$ of transitions (x_k, x_{k+1}).
Choose observer parameters $\theta \geq 1$ and sampling period $\mu > 0$.
Choose the matrices A, B, C, and L.
Construct the matrices $\Delta, \bar{A}$, and $\bar{B}$.
Initialize neural network parameters of T at random.
for $i = 1$ **to** $\text{iter}_{\max}$ **do**
    (a) Compute latent encodings $z_k = T(x_k)$.
    (b) Propagate latent state: $z_{k+1} = \bar{A} z_k + \bar{B} y_k$.
    (c) Compute dynamics loss L_{dyn} Eq.(9).
    (d) Update encoder parameters using SGD to minimize L_{dyn}.
end for
Initialize neural network parameters of T^{-1} at random.
for $i = 1$ **to** $\text{iter}_{\max}$ **do**
    (a) Compute reconstruction $\hat{x}_k = \psi^{(m)}(z_k)$.
    (b) Compute reconstruction loss L_{recon} (Eq. 10).
    (c) Update decoder parameters using SGD (Stochastic Gradient Descent) to minimize L_{recon}.
end for

that the state variable x_4 does not appear in any of these derivatives, thereby establishing its unobservability from the measured output.

Nevertheless, it was shown in [6] that the dynamics of x_4 are stable, with a nominal pole equal to $-\frac{1}{T'_{qo}} < 0$, independent of the estimated states. The system is therefore detectable, which justifies the observer design based on the observable third-order subsystem (x_1, x_2, x_3) without loss of generality.

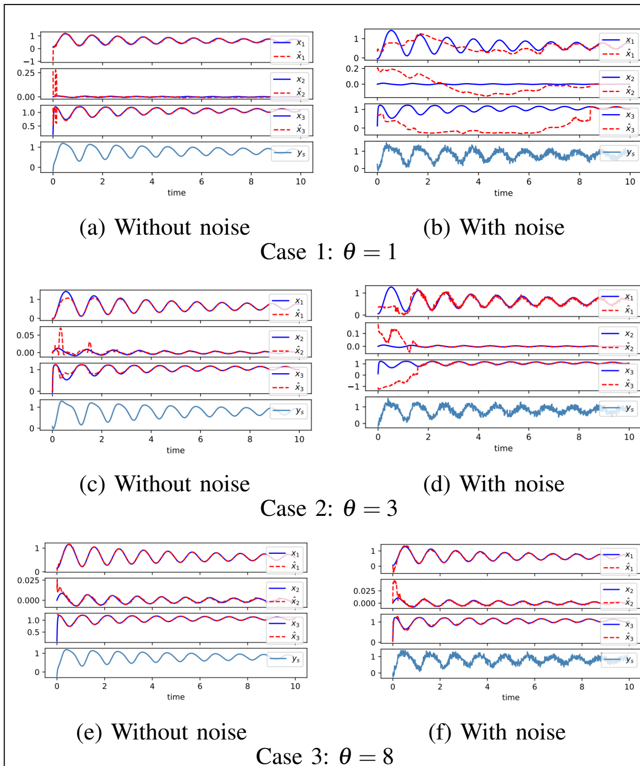
To evaluate the performance of the designed observer, its ability to reconstruct the system states was tested for different values of θ, both in the presence and absence of measurement noise. The matrices A, B, and C obtained after the coordinate transformation are given by:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}.$$

One of the key properties of high-gain observers is that these matrices remain invariant, regardless of the considered output. Moreover, the gain matrix L is chosen such that the matrix $\tilde{A} = A - LC$ is Hurwitz, using the LQR(Linear Quadratic Regulator) method. Thus, $L = \begin{bmatrix} 17.3 & 22.2 & 11.6 \end{bmatrix}^\top$, for $Q = \text{diag}(300, 90, 90)$. By contrast, the matrix $\bar{B}$ has dimensions $d_z \times d_y$.

The neural network was trained over 150 iterations (epochs), with a simulation horizon ranging from 0 to 10 seconds. The encoder T and decoder T^{-1} architectures are identical, each based on a multilayer perceptron (MLP) structure composed of three hidden layers with 400 neurons each, together with input and output layers of three neurons each. The activation function used is the hyperbolic tangent (tanh). Training is performed using the Adam optimizer with a learning rate of 10^{-3},mini-batch size of 10 and the adopted loss is the Mean Squared Error (MSE). The training dataset consists of 200 000 state transitions (x_k, x_{k+1}) generated from System (19), split into 80% training and 20% validation. Simulations were carried out on an Intel Core i5-1235U processor with 16 GB of RAM.

The numerical values of the generator parameters used in this study were taken from [9]. The results obtained for different values of θ are presented in Fig. 3.



(a) Without noise (b) With noise
Case 1: $\theta = 1$
(c) Without noise (d) With noise
Case 2: $\theta = 3$
(e) Without noise (f) With noise
Case 3: $\theta = 8$

Fig. 3: Simulation results for various θ values with the system output as active power.

These results demonstrate the ability of the designed observer to reconstruct the dynamic states of the studied system, both in the presence and absence of measurement noise. They thus confirm the relevance of the adopted approach for learning the transformations T and T^{-1}. It also emerges that the parameter θ has a strong influence on the observer's convergence speed: higher values of θ lead to faster convergence, whereas lower values result in slower convergence. This effect is particularly pronounced in the presence of measurement noise. However, depending on the sampling period, θ must not exceed certain values, otherwise convergence may be lost. In particular, for a sampling period of 0.01 s, this parameter should not exceed the value 8. To go beyond this limit, the sampling period must be reduced.

Furthermore, it can be observed that the presence of noise does not prevent the observer from reconstructing the system states. The estimated states remain very close to their true values, which confirms the ability of the multi-observer approach to mitigate the influence of noise on the reconstructed states. The results presented in Fig. 3 correspond to a noise variance of 0.01, *i.e.*, a signal-to-noise ratio of approximately 12.5% relative to the measured signal, without preventing the observer from accurately reconstructing the states.

1) Performance Evaluation of the Proposed Observer Under Measurement Noise: To quantitatively assess the proposed estimator, it is compared against the Extended Kalman Filter (EKF), which remains the dominant baseline in power system dynamic state estimation (DSE). The reported results are averaged over 15 simulation runs with randomly selected initial conditions within $[-0.2,\ 0,\ -0.2]^T \leq x_0 \leq [0.2,\ 0,\ 0.2]^T$. The EKF is implemented using a process noise covariance matrix $Q = \mathrm{diag}(10^{-4}, 10^{-4}, 10^{-4})$, and a measurement noise covariance matrix $R = 10^{-4}$. The comparison metrics include the global RMSE over the three estimated states and the convergence time, defined as the time required for the estimation error to decrease below 5% of the nominal state range. These results indicate that, in the noise-free and low-noise cases, the HGO exhibits better performance than the EKF in terms of convergence speed. However, this performance degrades as the noise level increases, highlighting the sensitivity of the system to noise. Moreover, the EKF provides better results in terms of RMSE. Nevertheless, the estimation error of the HGO remains bounded for all tested noise levels, which confirms the stability of the proposed multi-observer approach.

It is also observed that the HGO exhibits, in most tested scenarios, a lower or comparable execution time compared to the EKF. Although this difference remains moderate, it mainly reflects the fact that the EKF requires recomputing the Jacobian matrix at each time step.

TABLE I: Performance Comparison of State Estimators

Scenario	Method	Global RMSE	Conv. Time [s]	Exec. Time [s]
Noise-free (std = 0)	EKF	4.77×10^{-6}	0.328	0.0719
	HGO	3.72×10^{-4}	0.317	0.0748
Low noise (std = 0.01)	EKF	4.600×10^{-3}	9.990	0.0717
	HGO	1.420×10^{-2}	9.646	0.0678
Medium noise (std = 0.05)	EKF	2.32×10^{-2}	9.990	0.0717
	HGO	9.74×10^{-2}	9.990	0.0674
High noise (std = 0.1)	EKF	4.7900×10^{-2}	9.990	0.0784
	HGO	1.7700×10^{-1}	9.990	0.085

V. CONCLUSIONS

This paper has been devoted to the design of a neural-network-based high-gain observer aimed at avoiding the explicit computation of the Jacobian matrix and its inverse, which are generally required in observer design and are particularly difficult to derive analytically for highly nonlinear outputs. The proposed observer, applied to a fourth-order synchronous generator connected to an infinite bus with the active power as the output, provides satisfactory results. It therefore represents a further step toward overcoming some of the limitations that hinder the use of observers as dynamic state estimators in electric power systems.

Moreover, although the proposed observer partially mitigates the effect of noise, it remains sensitive to its increase, which leads to performance degradation when the noise level becomes high. In addition, the discretization imposes constraints on the sampling step that can be used for the system. Future work will therefore focus on improving robustness to noise as well as optimizing the discretization strategy.

REFERENCES

[1] G. Rinaldi, P. P. Menon, C. Edwards, and A. Ferrara, "Sliding mode based dynamic state estimation for synchronous generators in power systems," *IEEE Control Systems Letters*, vol. 2, no. 4, pp. 785–790, Oct. 2018.

[2] S. A. Nugroho, A. F. Taha, N. Gatsis, and J. Zhao, "Observers for differential algebraic equation models of power networks: Jointly estimating dynamic and algebraic states," *IEEE Trans. Control Netw. Syst.*, vol. 9, no. 3, pp. 1531–1543, Sept. 2022.

[3] M. Nadeem, S. A. Nugroho, and A. F. Taha, "Dynamic state estimation of nonlinear differential algebraic equation models of power networks," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2539–2552, May 2023.

[4] J. Chang, G. N. Taranto, and J. H. Chow, "Dynamic state estimation in power system using a gain-scheduled nonlinear observer," in *Proc. Int. Conf. Control Appl.*, 1995, pp. 221–226.

[5] H. H. Alhelou, M. E. H. Golshan, and N. D. Hatziargyriou, "Deterministic dynamic state estimation-based optimal LFC for interconnected power systems using unknown input observer," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1582–1592, Mar. 2020.

[6] I. B. Raoul, N. Homere, and T. Ahmed-Ali, "Sampled-data high gain observers for dynamic state estimation in power system," *Electr. Power Syst. Res.*, vol. 242, p. 111491, May 2025.

[7] J. Peralez and M. Nadri, "Deep learning-based luenberger observer design for discrete-time nonlinear systems," in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, Dec. 2021, pp. 4370–4375.

[8] J. Peralez and M. Nadri, "Deep model-free KKL observer: A switching approach," in *Proc. 6th Annu. Learn. Dyn. Control Conf. (L4DC)*, vol. 245, Jul. 2024, pp. 929–940.

[9] E. Ghahremani and I. Kamwa, "Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2556–2566, Nov. 2011.

[10] H. K. Khalil and J. W. Grizzle, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.

[11] E. Petri, R. Postoyan, D. Astolfi, D. Nešić, and V. Andrieu, "Hybrid Multi-Observer for Improving Estimation Performance," *IEEE Transactions on Automatic Control*, vol. 70, no. 5, pp. 3256–3271, May 2025, doi: 10.1109/TAC.2024.3500792.