

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ije

Constrained extended Kalman filter design and application for on-line state estimation of high-order polymer electrolyte membrane fuel cell systems

Lukas Böhler^{a,*}, Daniel Ritzberger^b, Christoph Hametner^{a,c}, Stefan Jakubek^a

^a Institute of Mechanics and Mechatronics, Control and Process Automation, TU Wien, Austria

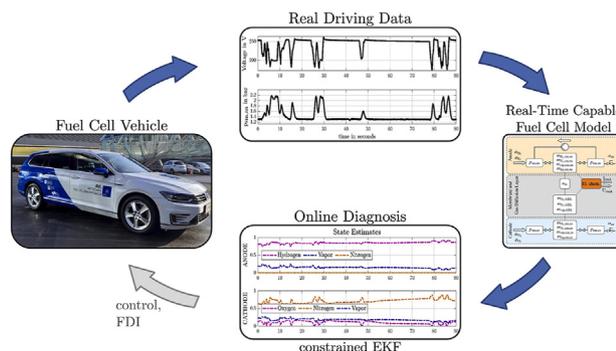
^b AVL List GmbH, Graz, Austria

^c Christian Doppler Laboratory for Innovative Control and Monitoring of Automotive Powertrain Systems, TU Wien, Austria

HIGHLIGHTS

- An alternative EKF formulation utilizing successive linearization is proposed.
- Robustness of the state estimation is improved by introducing inequality constraints.
- The constrained EKF is nine times faster than the comparable constrained UKF.
- The proposed concept is validated using drive-cycle measurements of a real FC stack.
- Estimation results during shut-down and start-up of the stack are presented.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 10 November 2020

Received in revised form

19 January 2021

Accepted 3 March 2021

Available online 17 April 2021

ABSTRACT

In this paper an alternative approach to extended Kalman filtering (EKF) for polymer electrolyte membrane fuel cell (FC) systems is proposed. The goal is to obtain robust real-time capable state estimations of a high-order FC model for observer applications mixed with control or fault detection. The introduced formulation resolves dependencies on operating conditions by successive linearization and constraints, allowing to run the nonlinear FC model at significantly lower sampling rates than with standard approaches. The proposed method provides state estimates for challenging operating conditions such as shut-down and start-up of the fuel cell for which the unconstrained EKF fails. A detailed

* Corresponding author.

E-mail address: lukas.boehler@tuwien.ac.at (L. Böhler).

<https://doi.org/10.1016/j.ijhydene.2021.03.014>

0360-3199/© 2021 The Authors. Published by Elsevier Ltd on behalf of Hydrogen Energy Publications LLC. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords:
PEMFC
Kalman filter
Successive linearization
Observer
Constrained estimation

comparison with the unscented Kalman filter shows that the proposed EKF reconstructs the outputs equally accurate but nine times faster. An application to measured data from an FC powered passenger car is presented, yielding state estimates of a real FC system, which are validated based on the applied model.

© 2021 The Authors. Published by Elsevier Ltd on behalf of Hydrogen Energy Publications LLC. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

The polymer electrolyte membrane fuel cell (PEMFC) is currently the focus of various research topics. Its promising application as a power source and the potential sustainability result in developments in many different directions. From improving hardware and control to enhancing operating strategies, including life cycle assessments and fault diagnosis, everything related to the fuel cell is subject to research. It is always of interest for scientists and engineers to investigate in detail the inner states of the fuel cell, which is also addressed in this paper. While most of the relevant states can be measured in test settings, this is no longer possible for real applications. Therefore, model-based observer methods are required to allow mathematical insight into the fuel cell to monitor and maintain operation. For that matter, a set of methods is introduced in this work which allows the application of the EKF for high-order FC models.

A part of this research recently focused on fault diagnosis to improve the operational safety, durability and reliability of FC systems in general. The necessary measurements to monitor the states of the fuel cell are, however, not always available or accessible. In control and in fault detection and isolation (FDI) applications alike, observers are utilized to estimate the missing states and further to generate residuals, which are processed to detect distinct fault scenarios [1]. Most observer approaches require representative models that describe the effects of interest sufficiently. Due to the complex electrochemical processes involved in PEM fuel cells, their respective models are often highly nonlinear. Standard observer and control theory is for the bigger part based on linear approaches though. The significance of linear approaches is however limited to a certain vicinity. Another challenge is real-time applicability, which can be the decisive factor for the choice of the considered method, which, however, varies from application to application. The goal is therefore to provide an alternative formulation of the EKF which can be beneficially combined with control or fault detection tasks.

Nonlinear observation for different PEMFC applications is addressed in literature in various ways. One possible solution is to adapt linear methods to nonlinear systems, for example nonlinear Luenberger or Kalman formulations. An approach using a nonlinear robust Luenberger method for a PEMFC system is presented in Ref. [2] for example, whereas [3] realized a Luenberger approach with linear parameter varying

models. The extended Kalman filter is a simple adaption of the Kalman filter (KF) to nonlinear systems based on updating the linear system matrices. The EKF is however rarely encountered in applications related to PEMFC models, but examples for systems with few states are provided in Ref. [4] for state of health diagnosis, in Ref. [5] for model predictive control or in Ref. [6], but for monitoring purposes of a molten carbonate fuel cell system. A more common application of a nonlinear KF for PEMFC models is the unscented Kalman filter (UKF) [7]. The application of the UKF does not require Jacobian matrices like the EKF and can be favorable in comparison [8]. Application examples for fault diagnosis related to fuel cells are presented in Refs. [9–11], health prognosis in Ref. [12] and for control in Refs. [5,13].

Fault detection and isolation or fault reconstruction for PEM fuel cell systems is addressed specifically by sliding-mode observer (SMO) approaches in numerous studies [14–18]. They offer robust concepts for nonlinear systems for the generation of residuals or parameter estimation. The design of SMO requires distinct system properties but can become intransparent for large sliding manifolds. The moving horizon estimation (MHE) approach evolved as a mixture of model predictive control (MPC) and optimal observation [19]. MHE advantageously incorporates constraints and past measurements in the estimation but becomes computationally more expensive with increased horizon lengths [20]. Other observation methods are for example realized based on multiple linear lumped models merged with fuzzy methods, as presented in Ref. [21] for fault clustering or similarly in Ref. [22] for a bank of observers. Fuzzy approaches also provide reasonable results, as long as the LTI models cover the nonlinearities well enough, which can however lead to a vast number of local submodels for highly nonlinear systems. A more detailed overview of various approaches for hydrogen fuel cell monitoring and prediction methods including FDI and parameter estimation is given in Ref. [23].

Each of the addressed methods has specific advantages and disadvantages that determine their distinct applicability. Some approaches are motivated around the issue, that linearizing high-order PEMFC models in order to obtain Jacobian matrices is difficult [10,11,24] or simply not necessary [25]. Others motivate their choices based on the drawbacks associated with EKFs for example [7,26], especially to motivate either the UKF [11] or the MHE [20]. Apart from the ability to reconstruct states from measurements in general, however, FDI and control tasks are considered with real-time restrictions and subject to physical constraints, therefore

rendering some methods meaningless for online estimation. This paper specifically addresses the issues that arise in the application of the EKF for high-order PEMFC models and resolves them in a realizable and efficient manner. The focus is to introduce a set of methods that allows a robust and fast application of the EKF with constraints.

The essential steps in achieving an efficient EKF formulation require analytical preprocessing of the Jacobian matrices and normalization of the state-space system. With means of successive linearization (SL), as applied in Ref. [27] and in Ref. [28] for model predictive control, the system is updated and discretized in every time step. This provides the matrices required for the Kalman filter related updates and the evaluation of the nonlinear model simultaneously. Another essential requirement for the estimation of the various states of the fuel cell is, that their values remain within physically reasonable bounds. For example, masses and concentrations must not be negative at any time, because nonphysical states can cause infeasible system updates. This work provides a possible approach to implement constraints for the state estimation [29], which is not based on the common clipping method [30]. The presented constrained EKF (CEKF) is applied to a highly nonlinear PEMFC model from literature [31]. The CEKF is validated in a synthetic simulation in the first step in order to show that it provides reliable state estimates. An additional shut-down and start-up sequence is conducted to demonstrate the robustness of the observer under extreme operating conditions and highlights the necessity for constraints. In the last step, the CEKF is validated based on measured data, which shows the ability to reconstruct measurements and provides insight into the states of the fuel cell, such as the anode nitrogen concentration or liquid water formation. The obtained results are benchmarked with an UKF or an unconstrained EKF when meaningful and their performances are compared.

This paper is structured as follows: First, the underlying FC model is shortly introduced and the significant features for this work are described. After that, the applied methods for the incorporation of the observer design with constraints are presented. Then, the results of the derived methods are illustrated and discussed for the available model and measurements. The paper concludes with remarks on the presented content.

Model based observer design

This section briefly introduces the underlying model used for observer design, followed by the extended Kalman filter formulation, successive linearization and the incorporation of the constraints. The resulting equations are presented in a way, such that they can be directly interpreted as a quadratic programming problem.

Fuel cell model

The nonlinear model required for observation is based on the comprehensive, high-order model presented in Ref. [31]. It includes gaseous and liquid species, pressures and electrochemical relations for the voltage. The model is derived based

on a zero-dimensional approach and is therefore expressed by a set of ordinary differential equations. An overview of the structure is provided in Fig. 1.

The model is separated into a supply manifold (sm), exit manifold (em) and a center manifold (cm), of which the latter comprises the lumped volumes of the anode (an) and the cathode (ca). The additional gas diffusion layer (GDL) and the catalyst layer (CL) allow the description of diffusion of species through the cell and are part of the electrochemical equations of the voltage. Various effects like nitrogen crossover and generation of liquid product water are considered in the model, therefore providing the desired description of the states which are typically not measurable. The inputs u of size n_u , the states x of size n_x and the outputs y of size n_y of the model are given by

$$u = \begin{bmatrix} \dot{m}_{H_2} \\ \dot{m}_{O_2} \\ \phi_{H_2} \\ \phi_{O_2} \\ I_{stack} \\ T_{stack} \\ \alpha_{ca} \\ \alpha_{an} \end{bmatrix}, \quad x = \begin{bmatrix} p_{sm,ca} \\ m_{O_2,cm,ca} \\ m_{N_2,cm,ca} \\ m_{vap,cm,ca} \\ m_{liq,cm,ca} \\ p_{em,ca} \\ p_{sm,an} \\ m_{H_2,cm,an} \\ m_{N_2,cm,an} \\ m_{vap,cm,an} \\ m_{liq,cm,an} \\ p_{em,an} \\ a_m \\ m_{O_2,GDL} \\ m_{N_2,GDL} \\ m_{vap,GDL} \end{bmatrix} \quad \text{and } y = \begin{bmatrix} U_{stack} \\ p_{sm,an} \\ p_{sm,ca} \end{bmatrix}.$$

The index ‘vap’ is the abbreviation for vaporized water and thus ‘liq’ refers to water in its liquid state. The index ‘cm’ can be omitted in the following since the gas diffusion layer is labeled as such and a clear distinction is provided. The input vector u comprises the hydrogen and air mass flows \dot{m}_{H_2} and

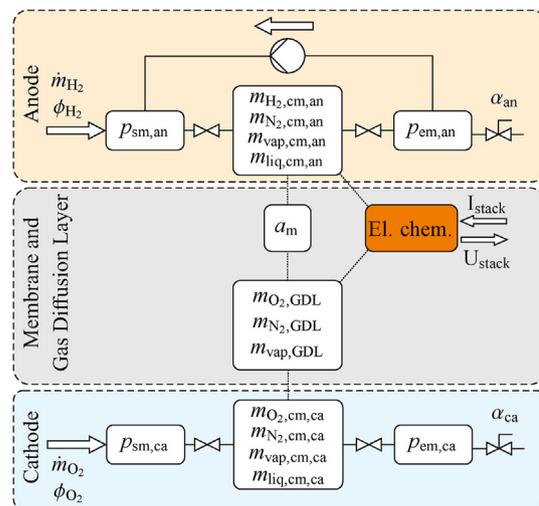


Fig. 1 – Schematic structure of the fuel cell stack model with the states and inputs of concern [31]. The model comprises the anode and cathode sides, as well as a gas diffusion layer. The auxiliary volumes of the supply, center and exit manifold further distinguish the considered submodels of the stack.

\dot{m}_{O_2} respectively, their corresponding relative humidities φ , the stack current I_{stack} and the average stack temperature T_{stack} . The valve positions α_{ca} and α_{an} describe the degree of opening of the back-pressure and the purge valve, respectively. The state vector is comprised of the manifold pressures, the species in the anode, cathode and the GDL. The remaining state a_m represents the dynamic water activity of the membrane at the catalyst layer.

Because the model comprises numerous states and complex electrochemical relations in the output function, a detailed explanation of the model exceeds the scope of this work. For more insight to this model see Refs. [31–33]. Some specific aspects which require the introduction of constraints are however briefly addressed in the result section.

Unconstrained EKF

In order to highlight the incorporation of successive linearization into the extended Kalman filter, the formulation of the EKF is recapitulated first. The discrete system model is given as

$$\begin{aligned} \mathbf{x}_k &= \mathbf{F}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{v}_k \end{aligned} \quad (1)$$

where k is the time step, $\mathbf{F}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1})$ is the solution of the first order differential equations of the model, $\mathbf{h}(\mathbf{x}_k, \mathbf{u}_k)$ is the nonlinear function mapping the states and inputs to the outputs, \mathbf{w}_k is the process noise and \mathbf{v}_k is the measurement noise, both assumed to be uncorrelated and white. It is further assumed, that the covariances are known and given by

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_w), \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_v), \quad \mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{Q}_0)$$

with \mathbf{Q}_w , \mathbf{R}_v and $\mathbf{Q}_0 > \mathbf{0}$. Based on the Kalman filter formulation given in Ref. [34], the vector of predicted state estimates $\hat{\mathbf{x}}_k^-$ and the predicted covariance matrix \mathbf{P}_k^- are

$$\hat{\mathbf{x}}_k^- = \mathbf{F}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) \quad (2)$$

$$\mathbf{P}_k^- = \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_w \quad (3)$$

with the initial conditions $\hat{\mathbf{x}}_0 = \mathbf{x}_0$ and $\mathbf{P}_0 = \mathbf{Q}_0$ and $\hat{\mathbf{x}}_{k-1}$ as the previous state estimate. The system matrix \mathbf{A}_{k-1} is obtained from the analytical derivation of the differential equations of the model, which is elaborated in the next section. The vector of updated or a posteriori state estimates $\hat{\mathbf{x}}_k$ is given by

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{L}_k (\mathbf{y}_k - \hat{\mathbf{y}}_k^-), \quad (4)$$

where \mathbf{L}_k is the Kalman gain matrix and the predicted output estimate $\hat{\mathbf{y}}_k^-$ is obtained from

$$\hat{\mathbf{y}}_k^- = \mathbf{h}(\hat{\mathbf{x}}_k^-, \mathbf{u}_k). \quad (5)$$

Based on $\boldsymbol{\varepsilon}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$, the estimation error covariance matrix update is

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{L}_k \mathbf{C}_k) \mathbf{P}_k^- (\mathbf{I} - \mathbf{L}_k \mathbf{C}_k)^T + \mathbf{L}_k \mathbf{R}_v \mathbf{L}_k^T \quad (6)$$

where \mathbf{I} is the identity matrix of according size and \mathbf{C}_k is the output matrix of the modeled system, which is also elaborated

in the next section. The optimal gain \mathbf{L}_k in the unconstrained case is obtained by minimizing the trace of \mathbf{P}_k as:

$$\mathbf{L}_k = \arg \min(\text{Tr}[\mathbf{P}_k]) \quad (7)$$

Reformulating the expression resulting from Eq. (7) yields

$$\mathbf{L}_k = \underbrace{\mathbf{P}_k^- \mathbf{C}_k^T}_{\mathbf{f}_k} \underbrace{(\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}_v)^{-1}}_{\mathbf{H}_k^T} \quad (8)$$

and therefore the desired Kalman gain. After introducing the matrices \mathbf{H}_k and \mathbf{f}_k , transposing Eq. (8) yields a more compact expression

$$\mathbf{L}_k^T = \mathbf{H}_k^{-1} \mathbf{f}_k^T \quad (9)$$

which is equal to the unconstrained solution acquired from solving the quadratic program stated by \mathbf{H}_k and \mathbf{f}_k as

$$\mathbf{L}_k^T = \min(\mathbf{H}_k, -\mathbf{f}_k^T) \quad (10)$$

or more precisely:

$$\text{minimize : } \frac{1}{2} \mathbf{l}_k^T \mathbf{Q}_k \mathbf{l}_k + \mathbf{E}_k^T \mathbf{l}_k \quad (11)$$

where \mathbf{l}_k is the column-major vector representation of \mathbf{L}_k^T , \mathbf{Q}_k is a symmetric Kronecker tensor product of the unity matrix $\mathbf{I}^{n_x \times n_x}$ and \mathbf{H}_k and \mathbf{E}_k^T is the column-major vector representation of $-\mathbf{f}_k^T$. The well-known formulation in Eq. (11) shows that the extension with state constraints for the quadratic programming problem is straightforward.

Successive linearization

One of the major challenges for the application of the EKF for fuel cell models is to obtain the updated system matrices or Jacobians [10,11]. In general, the Jacobians can be approximated numerically or derived analytically from the system of equations. Both approaches have their benefits, but the analytical solution is always more accurate and can be pre-processed as a function. This is exploited for successive linearization, making the analytical derivatives a necessity and an advantage at the same time. A positive side-effect of SL is, that the linear, updated system matrices are available for any consecutive method, which can be beneficial for control or other observation methods.

The following defines the successive system updates such that they can be incorporated directly into the presented equations for the EKF. Given the nonlinear function $\mathbf{g}(\mathbf{x}, \mathbf{u})$ that represents the first order differential equations of the modeled system as

$$\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{u}), \quad (12)$$

the required analytical derivatives are obtained as the Jacobians from

$$\mathbf{A}_{c,k} = \left. \frac{\partial \mathbf{g}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_k, \mathbf{u}_k} \quad (13)$$

$$\mathbf{B}_{c,k} = \left. \frac{\partial \mathbf{g}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} \right|_{\hat{\mathbf{x}}_k, \mathbf{u}_k} \quad (14)$$

$$C_{c,k} = \left. \frac{\partial h(\mathbf{x}, \mathbf{u})}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_k, \mathbf{u}_k} \quad (15)$$

$$D_{c,k} = \left. \frac{\partial h(\mathbf{x}, \mathbf{u})}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_k, \mathbf{u}_k} \quad (16)$$

where $A_{c,k}$, $B_{c,k}$, $C_{c,k}$ and $D_{c,k}$ represent the resulting continuous system matrices, indicated by the index c and evaluated at the correct time step k . Based on the matrices $A_{c,k}$ and $B_{c,k}$, the evolution of the linear system is expressed as shown in Ref. [27] by

$$\dot{\mathbf{x}}_L = A_{c,k}(\mathbf{x}_L - \hat{\mathbf{x}}_k) + B_{c,k}(\mathbf{u}_L - \mathbf{u}_k) + \dot{\mathbf{x}}_k \quad (17)$$

with $\dot{\mathbf{x}}_k = g(\hat{\mathbf{x}}_k, \mathbf{u}_k)$. Rearranging this equation yields

$$\dot{\mathbf{x}}_L = A_{c,k}\mathbf{x}_L + B_{c,k}\mathbf{u}_L + \mathbf{K}_{c,k}, \quad (18)$$

where $\mathbf{K}_{c,k}$ is the column vector of the collected expressions of Eq. (17) that are constant for k . The required discrete representation of Eq. (18) is obtained based on a matrix exponential. By defining the matrix \mathcal{B}_c which is created by extending $B_{c,k}$ [35] as

$$\mathcal{B}_c = [B_{c,k} \mathbf{K}_{c,k}] \quad (19)$$

and by further defining the symmetric matrix \mathcal{A}_c as

$$\mathcal{A}_c = \begin{pmatrix} A_{c,k} & \mathcal{B}_c \\ 0 & 0 \end{pmatrix}, \quad (20)$$

the exponential expression that yields the desired discrete matrices is given by

$$\mathbf{G}_k = e^{-\mathcal{A}_c T_s} \quad (21)$$

where T_s is the considered sampling interval. Approximating Eq. (21) with a Taylor series expansion yields the discrete matrices of interest. The resulting zero-order-hold updates for A_k , B_k and K_k have to be extracted from \mathbf{G}_k according to

$$\mathbf{G}_k = \begin{pmatrix} A_k & B_k & K_k \\ 0 & I & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (22)$$

Equation (2) for the predicted state estimate $\hat{\mathbf{x}}_k^-$ is therefore obtained as

$$\hat{\mathbf{x}}_k^- = F(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) = A_{k-1}\hat{\mathbf{x}}_{k-1} + B_{k-1}\mathbf{u}_{k-1} + K_{k-1} \quad (23)$$

which is the successive and offset free evolution of the nonlinear system in a discrete, linear fashion. Also the matrix A_{k-1} is then available for the predicted error covariance P_k^- of Eq. (3). Since $C_k = C_{c,k}$ and $D_k = D_{c,k}$, evaluating the Jacobian matrix of Eq. (15) directly yields C_k for Eq. (6) and evaluating D_k is not necessary, if there are no follow-up applications that require this matrix. Due to the conducted successive linearization, the EKF almost becomes the linear Kalman filter for each time step. In contrast to the linearized Kalman filter however, the matrices $C_{c,k}$ and $D_{c,k}$ of the extended Kalman

filter are evolved around the predicted state estimates $\hat{\mathbf{x}}_k^-$ instead of the updated estimate $\hat{\mathbf{x}}_k$ [30,34].

Normalization of the model

The FC model introduced in Section 2.1 comprises pressures with magnitudes of 10^5 Pa and masses with magnitudes of 10^{-3} kg and is therefore numerically stiff. This can be addressed by scaling the nonlinear system into according units, but since successive linearization is utilized, the system matrices and vectors are normalized conveniently to ranges between 0 and 1 without altering the underlying model. The according symmetric matrices are $\mathbf{X}_n \in \mathbb{R}^{n_x \times n_x}$ for the states, $\mathbf{u}_n \in \mathbb{R}^{n_u \times n_u}$ for the inputs and $\mathbf{Y}_n \in \mathbb{R}^{n_y \times n_y}$ for the outputs, whose off-diagonal elements are zero. The matrix $A_{c,k}$ of Eq. (13) therefore is replaced by

$$A_{c,k}^n = \mathbf{X}_n^{-1} A_{c,k} \mathbf{X}_n, \quad (24)$$

where n indicates the normalized system. Normalizing Eqs. 14–16 is conducted similar to $A_{c,k}^n$, which can be pre-processed. The state, input and output vectors are transformed to normalized ranges as well, resulting in \mathbf{x}_k^n , \mathbf{u}_k^n and \mathbf{y}_k^n respectively. The index n is omitted in the following for readability but all vectors and matrices in Section 2.5 are normalized.

EKF with constraints

A general problem of any unconstrained state observer is, that the estimates can be projected outside of the physical or modeled bounds. This is especially problematic for masses and concentrations, which cannot be non-negative by definition. The implementation of constraints can prevent undesired estimates and prevent solutions outside of the allowed boundaries. A common approach to realize constraints for the EKF is clipping [30], which means to simply replace estimates with their constraints if they are violated. The predicted state estimates $\hat{\mathbf{x}}_k^-$ are then consistent with these limits, but the covariance matrix P_k is typically not. An overview of different means to realize constraints for Kalman filtering is provided in Ref. [29] and some considerations for their general application to state estimation are discussed in Ref. [26]. Of the available methods, the gain projection method is well applicable for online estimation. Instead of constantly considering the constraints it allows a simple active set formulation similar to MPC or MHE applications that requires a recalculation of the Kalman gain L_k only if constraints are violated. Their activation is followed by a constrained optimization in order to obtain the adjusted gain.

The considered inequality state constraints for the system are expressed by

$$\hat{\mathbf{x}}_k \geq \hat{\mathbf{x}}_{\text{lim}} \quad (25)$$

which requires no further action if the relation is true for all states. If Eq. (25) does not hold, Eq. (7) is additionally subject to the equality constraints Γ given by

$$\Gamma(\mathbf{L}_k) = \widehat{\mathbf{x}}_k^- + \mathbf{L}_k(\mathbf{y}_k - \widehat{\mathbf{y}}_k^-) - \widehat{\mathbf{x}}_{\text{lim}}^- = 0. \quad (26)$$

The constrained problem can be re-formulated as

$$\mathcal{L}(\mathbf{L}_k, \mathbf{\Lambda}) = \text{Tr}[\mathbf{P}_k] + 2\text{Tr}[\Gamma(\mathbf{L}_k)\mathbf{\Lambda}] \quad (27)$$

where $\mathbf{\Lambda}$ are Lagrangian multipliers of according size. The solution to the constrained system is thus obtained from the derivation of

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}_k} = 0 \text{ with } \Gamma(\mathbf{L}_k) = 0. \quad (28)$$

The resulting matrix equation that yields the adjusted gain \mathbf{L}_k is obtained as an extension to Eq. (8) as:

$$\begin{bmatrix} \bar{\mathbf{L}}_k \\ \mathbf{\Lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_k \\ (\widehat{\mathbf{x}}_{\text{lim}}^- - \widehat{\mathbf{x}}_k^-)^T \end{bmatrix} \begin{bmatrix} \mathbf{H}_k & -(y_k - \widehat{y}_k^-) \\ (y_k - \widehat{y}_k^-)^T & 0 \end{bmatrix}^{-1} \quad (29)$$

Because Eq. (29) is a linear matrix equation, re-evaluating only the states for which the constraints have been activated is sufficient, which is favorable in terms of real-time estimation. Alternatively, equations (8) and (25) together pose the constrained convex quadratic program stated by

$$\mathbf{L}_k = \min(\mathbf{H}_k, -\mathbf{f}_k^T) \text{ s.t. : } \widehat{\mathbf{x}}_k^- \geq \widehat{\mathbf{x}}_{\text{lim}}^- \quad (30)$$

which again can be efficiently solved by according commercial software packages after transforming the system into the column-major representation of Eq. (11). The updated state estimates considering the constraints are thus obtained from replacing $\bar{\mathbf{L}}_k$ with \mathbf{L}_k in Eqs. (4) and (6).

Simulation studies

This section presents the results of the state estimation based on the considered FC model obtained with the introduced constrained Kalman filter configuration. The observer is validated in the first step in different settings against the underlying FC model, such that the states are available as a reference for the estimation. This shows the general capability of the proposed methods for state estimation of the high-order PEMFC model. The CEKF is compared to an unconstrained EKF and a constrained UKF (CUKF) for benchmarking. The UKF is designed based on [7] and the constraints are implemented as presented in Ref. [30] via clipping, which has not the same effect on the UKF as on the EKF. Additionally, a simulation-based shut-down and start-up (SDSU) scenario is presented as well, in which the FC model is operated with hydrogen recirculation at the anode. The SDSU-scenario highlights the limits of the EKF and demonstrates the necessity for constraints.

Simulation setup

The basic requirement of the presented CEKF is that the observer is able to reproduce the states of the underlying FC

model correctly. For that matter, two outputs and four states are selected to represent the otherwise rather extensive FC model under consideration. The outputs of interest are the stack voltage U_{stack} and the supply manifold pressure on the anode side $p_{\text{sm,an}}$. The states selected for observation of the model are the liquid water masses $m_{\text{liq,ca}}$ and $m_{\text{liq,an}}$ in the cathode and anode, respectively, the mass of nitrogen in the anode $m_{\text{N}_2,\text{an}}$ and the mass of oxygen in the cathode $m_{\text{O}_2,\text{ca}}$. Especially the oxygen mass allows a detailed insight into the ongoing electrochemical reactions which could lead to severe damage of the stack during shut-down for example. Because these states are more related to FDI applications, the evaluation of the measured data presented at the end of this section provides a more control related selection of the channel species.

The simulations are conducted with a sampling rate of $T_s = 0.01$ s, which is the sampling rate of the available measurements presented in Section 4. This sampling rate further determines the time requirements for observation or control applications for the FC powered car. The lower bounds or inequality constraints $\widehat{\mathbf{x}}_{\text{lim}}^-$ for the CEKF and the CUKF are set to 0, except for the GDL states which are addressed specifically. Mapping the GDL states to the stack voltage U_{stack} is modeled according to the final equation for the voltage presented in Ref. [32]. The output equation states, that if the concentrations of the involved species approach zero, the over-potentials of the reaction kinetics η_{ca} and η_{an} approach negative infinity. For example, if no oxygen is available at the cathode, the limit for η_{ca} becomes:

$$\lim_{c_{\text{O}_2} \rightarrow 0} \eta_{\text{ca}} \rightarrow -\infty \quad (31)$$

In order to avoid these limits, the lower bounds for the GDL masses are set to 1e-6 or 0.001 mg respectively. This allows a reasonable close approximation of η for any missing species and prevents the solution from being undefined, which is paramount for the SDSU scenario. Similar limits can be applied to all masses for that matter. In general, model-based approaches considering these operating conditions require limitations for the concentrations or masses, if these conditions are not excluded from observation.

Basic operation

First, the basic operation of the fuel cell is investigated in an open-end setting. The FC model is excited by a series of input steps, where the input current I_{stack} is altered over time. At $t = 45$ s the hydrogen mass flow is additionally reduced by 20% of the initial value to further excite the system, whereas the other inputs are kept constant throughout the simulation. Fig. 2 shows the estimation results obtained from the fuel cell model in the considered configuration. Because the purge valve $\alpha_{\text{an}} = 1$ is open, the obtained linearized system matrices \mathbf{A}_k are therefore stable due to open-end operation.

The results of Fig. 2 show, that all considered observers are capable of reconstructing the modeled states for regular operation correctly. The obtained state estimates are almost equal because the constraints are only shortly activated at the beginning and the initial values are close to the actual values. The results of Fig. 2 confirm that the combination of

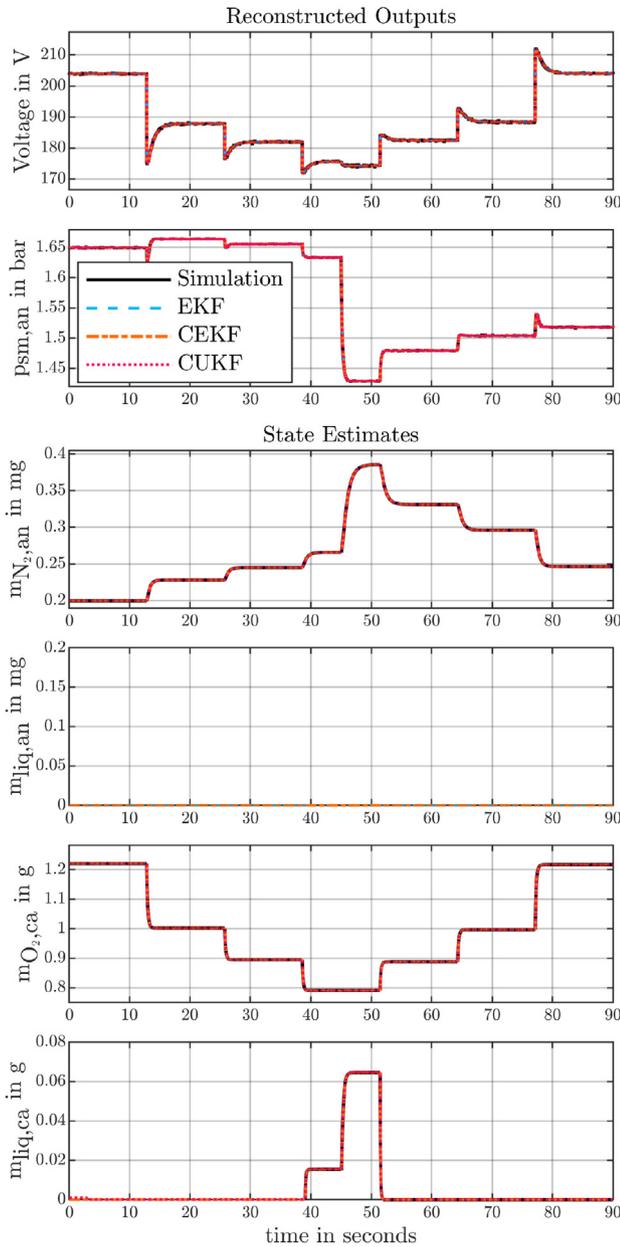


Fig. 2 – Simulation results of the unconstrained EKF, the CEKF and the CUKF compared to simulation results obtained from the nonlinear FC model. This simulation shows, that for regular operation of the stack all observers are capable to reproduce the modeled states correctly.

normalization, successive linearization and discretization of the nonlinear FC model therefore yields the correct estimates.

A decisive reason to apply the CEKF instead of the CUKF for the considered PEMFC model is the required time capability. For the sampling time of $T_s = 0.01$ s, standard discrete approaches as the Runge Kutta or the Euler method are not able to converge to a solution. Due to the stiffness of the considered model, sampling rates have to be decreased up to two magnitudes, which therefore increases the time to obtain a solution by the same factor. Because $2n+1$ system evaluations are required by the sigma points of the UKF, the nonlinear FC model is evaluated disproportionately often compared to a

Table 1 – Comparison of the performance of the considered observers in terms of RMSE of the outputs and the simulation time relative to the unconstrained EKF.

Filter	RMSE U_{stack}	RMSE $p_{sm,an}$	rel. time
EKF	0.508	130.6	1
CEKF	0.508	130.6	1.05
CUKF	0.508	133.1	8.79

standard EKF approach. Table 1 provides a comparison of the execution times of the different filters and the Root Mean Squared Error (RMSE) of the presented outputs. The execution time is given relative to the unconstrained EKF, which is regarded as the benchmark for this matter to eliminate the influence of the hardware as far as possible.

Table 1 shows, that the EKF and the CEKF are faster than the CUKF by a factor of approximately 9. The EKF and the CEKF are almost equally fast in this comparison, because only if inequality (25) is true, the approaches are different from each other. Another perspective for these results is that the EKF requires approximately 13 s to process data of 90 s of length, which relates to 9000 samples due to the discretization defined by T_s , with the considered hardware. This shows, that on average, the real-time requirements can likely be achieved based on the proposed methods. Although the UKF can be considered beneficial in comparison to the EKF [20], the CUKF is not fast enough in the given setting and thus not applicable at the required sampling interval without reducing the complexity of the model which would benefit other approaches as well.

Start-up and shut-down

In order to show that the introduced CEKF is able to respect the introduced constraints and highlight the robustness of the approach, a shut-down and start-up event is simulated and presented in Fig. 3. This sequence is selected because it pushes the EKF to its limits due to the highly nonlinear dynamics and unwanted modeling effects that occur (Section 3.1), although it might have little practical meaning for state observation. The sequence is physically motivated, but occurring negative effects due to oxygen starvation or carbon corrosion are not considered in this simulation. For the SDSU scenario, the FC model is operated in a hydrogen recirculation configuration with $\alpha_{an} = 0$ and pressure setpoints determine the massflows \dot{m}_{H_2} and \dot{m}_{O_2} . The anode valve is opened from time to time ($\alpha_{an} = 1$) according to the applied purging strategy, which is visible in Fig. 3 in the pressure signal of the anode.

The SDSU sequence interrupts regular operation and is initialized by reducing the stack input current I_{stack} to 5A at $t = 15$ s. At $t = 20$ s, the pressure setpoint of the cathode supply manifold is changed to atmospheric together with the setpoint of the anode side, which is set to a slightly higher value. Shortly after the mass flows starts to diminish, the cathode is closed off completely and the mass flow is set to zero. The applied current of 5A completely consumes the remaining oxygen in the cathode. The stack voltage drops at about $t = 37$ s as shown in Fig. 3, which activates the lower bounds

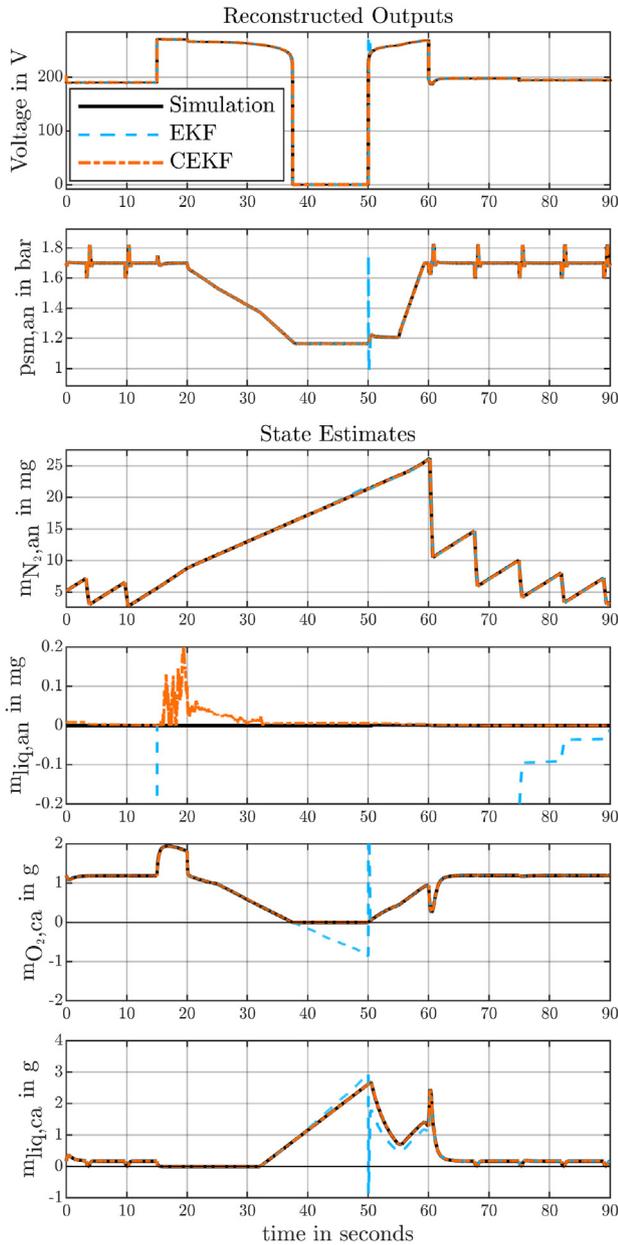


Fig. 3 – Results of the shut-down and start-up scenario for the unconstrained EKF and the CEKF. Once the oxygen in the cathode is depleted, the EKF starts to diverge and is unable to provide reasonable state estimates. Additionally, it is visible that the CEKF respects the desired constraints.

for the GDL states. For the EKF, the stack voltage U_{stack} is set to zero during this period in order to obtain simulation results. This is due to the effects described by Eq. (31), which would cause parts of the output equations to approach negative infinity. It is necessary therefore to limit U_{stack} to zero for the unconstrained EKF to inhibit infeasible solutions during this period or otherwise state estimation cannot continue.

The rapid approach of U_{stack} towards zero represents the part of the fuel cell model where it is close to its limits. This issue is addressed and ultimately prevented by the constraints, especially for the species. The introduced limits for

the GDL states allow the successive linearization to continue without introducing relevant numerical distortions. The performance of the state estimation displayed by the EKF on the other hand is not acceptable, which is the major motivation to implement the constraints. The efficiency of the CEKF is demonstrated, since the estimates are still almost perfect reconstructions of the states of the model. The following start-up part of the SDSU-sequence begins at $t = 50$ s, which is conducted basically equal to the shut-down part but in reverse order. The continued state estimation of the unconstrained EKF for the start-up sequence is able to converge to the true values only after a period of time, yielding nonphysical state

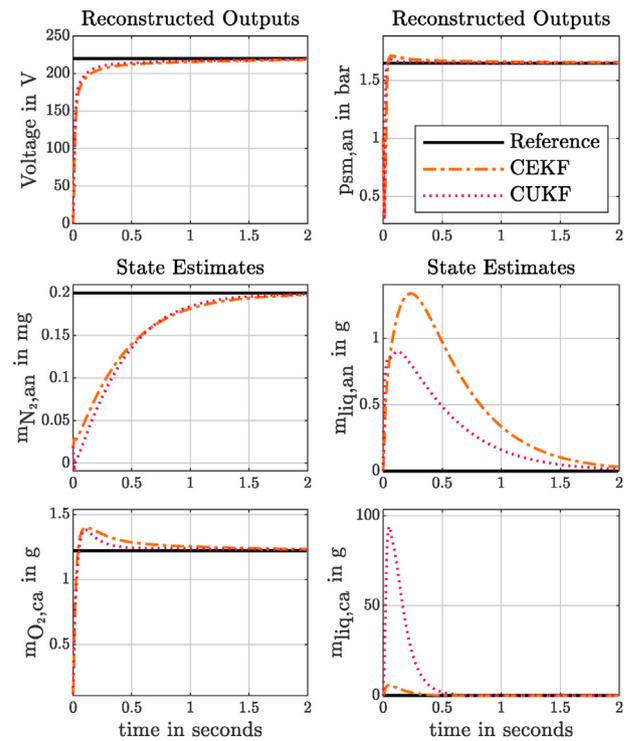


Fig. 4 – Convergence rates of the CEKF compared to the CUKF for the selected signals. The guessed state values for initiation are reduced by 90% of their actual values.



Fig. 5 – KeyTech4Ev demonstrator vehicle developed by AVL List GmbH. Measurements of an actual drive cycle available for the application of the observer have been obtained with this vehicle.

estimates during this period, which highlights the capabilities of the CEKF for a wide range of potential scenarios.

Filter convergence

The performance of the CEKF has been presented based on simulations and the real-time capability for the applied PEMFC model has been addressed in comparison. The results indicate, that the CEKF based on normalization and successive linearization is a suitable approach to handle stiff and extensive models. However, the EKF can fail to converge in general given a poor guess of the initial states [20] and

therefore the convergence rates of the CEKF are analyzed and compared to the constrained UKF.

For this investigation, the initial state values are reduced by 90% of their true values. The responses of the observers are visualized in Fig. 4 for a simulation time of 2 s.

The stack voltage is initialized at $t = 0$ with $I_{\text{stack}} = 0$ and the pressures are set to atmospheric. The results show, that the CEKF is able to approach the true state values equally well as the CUKF. Furthermore, since Fig. 4 can be interpreted as a step response, the different time constants of the system become visible as well. However, by initializing the observer with state values below 90% or more specifically, close to zero, the convergence of the CEKF is not guaranteed. It is therefore not possible to initialize with $\mathbf{x}_0 = 0$, which is also not meaningful for states describing pressure or masses.

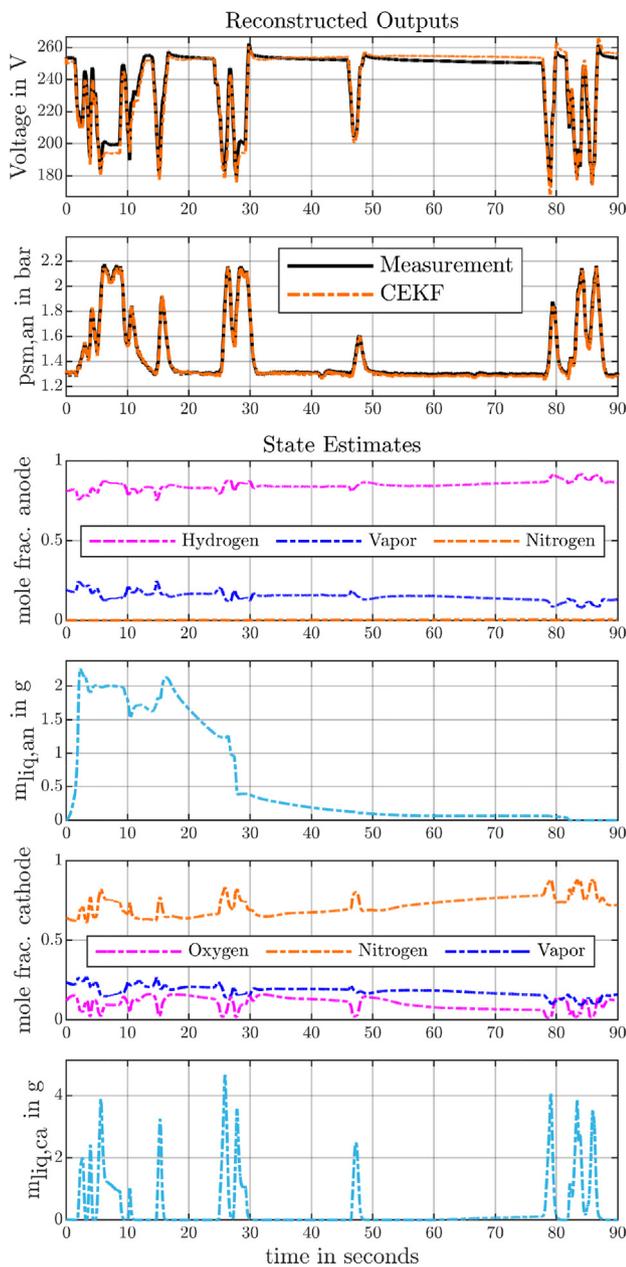


Fig. 6 – Application results of the CEKF to real measurements. The data show a segment of a drive cycle of the demonstrator vehicle. For better insight, the estimated mole fractions of the gaseous species of the anode and cathode side are presented.

Real driving data

The CEKF is applied in a final step to available measured data obtained from a demonstrator vehicle which has been developed in the KeyTech4Ev project [36]. The vehicle is the VW Passat presented in Fig. 5 that has been retrofitted with a fuel cell system.

Measurements of a real drive cycle are available, which are utilized together with the nonlinear FC model in order to observe the internal states of the stack. A more detailed explanation about the vehicle and its specifications is given in Ref. [31].

The results of the application of the CEKF are presented in Fig. 6, where only the measurements of the outputs are available and the states have to be estimated.

The drive cycle data has been obtained in a hydrogen recirculation configuration, similar to the SDSU scenario, where the massflows of hydrogen and oxygen are pressure driven. It has been shown in Sections 3.2 and in 3.3, that the CEKF is able to reproduce the internal states of the model based on the given outputs and that the estimates can be trusted to reproduce the model behavior. For a better insight, the mole fractions of the gaseous elements in the anode and cathode are presented. The obtained estimates are therefore available for fault detection of the real system. Additionally, purging strategies can be designed based on the distribution of the species within the cell in order to prevent nitrogen ramp-up for example.

Conclusions

The introduced successive linearization allows the discrete CEKF to run at sampling rates where Runge Kutta or Euler methods are not able to converge to meaningful solutions for the investigated model. The sampling interval has been reduced by a factor of 100 due to the successive updates based on analytical derivatives, which results in the CEKF being nine times faster compared to the CUKF without the loss of performance. This is demonstrated by the RMSE of the measured stack voltage and supply manifold pressure at the anode. Because the analytical derivatives need to be updated and discretized in every time step, a part of the additional speed

gained by the sampling rate reduction is however lost. The results of the shut-down and start-up scenario highlight that state constraints are necessary and beneficial in order to provide reliable estimation results even under extreme conditions. This is especially visualized by the oxygen concentration during shut-down. The application to measured data shows that the CEKF is able to reconstruct the outputs of the hybrid vehicle and thus provide actual insight into the states of the cell. The introduced EKF is additionally presented in a convenient quadratic programming formulation.

For the given setup, the CEKF based on successive linearization, normalization and discretization is selected because it provides the required mix between capability and reliability. The CUKF had to be dismissed since it could not be operated sufficiently fast for the considered high-order model due to the $2n+1$ system evaluations that are conducted in every iteration. The unconstrained EKF, on the other hand, is not sufficiently reliable for all considered operating conditions, since negative or missing species have to be prevented in the voltage output equation. General issues with convergence rates of the EKF due to poor initial guesses can be dismissed to a certain degree if the CEKF is initialized based on physical meaningful conditions. Based on the conducted investigations, the presented CEKF formulation can provide promising results for high-order fuel cell models.

It is straightforward that by reducing the complexity of the considered model the application of a MHE, UKF or other observation methods may be favorable compared to the EKF. The real-time requirements are different for these tasks and depend on the type of application. FDI approaches for observing long-term degradation of the cell are not expected to face the same real-time requirements as observers applied for controllers. Similarly, the requirements for predictive control strongly depend on the horizon length and the control task. The proposed approach however aims at providing the system matrices for a combined control or fault diagnosis and observation task for the presented sampling rates. The result is a fast and reliable observer approach for the considered high-order FC model that is able to meet the sampling requirements of the application. Validating the inner states based on measurements is the next step for the application of the CEKF, or any other observation method for fuel cell systems for that matter.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been funded by the Austrian Research Promotion Agency (FFG) in the scope of the projects *KeyTech4Ev* (# 855237) and *ThermoSense* (# 871503). The authors would also like to acknowledge and thank the industrial and scientific partners for their support.

REFERENCES

- [1] Patton RJ, Chen J. Observer-based fault detection and isolation: robustness and applications. *Contr Eng Pract* 1997;5(5):671–82.
- [2] Olteanu SC, Aitouche A, Belkoura L, Nakrachi A. Robust state estimation of PEM fuel cell systems. 2013.
- [3] De Lira S, Puig V, Quevedo J, Husar A. LPV observer design for PEM fuel cell system: application to fault detection. *J Power Sources* 2011;196(9):4298–305. <https://doi.org/10.1016/j.jpowsour.2010.11.084>. ISSN 3787753.
- [4] Bressel M, Hilairat M, Hissel D, Ould Bouamama B. Extended kalman filter for prognostic of proton exchange membrane fuel cell. *Appl Energy* 2016;164:220–7. <https://doi.org/10.1016/j.apenergy.2015.11.071>. ISSN 3062619.
- [5] Hahnel C, Cloppenborg A, Horn J. Offset-free nonlinear model predictive control of electrical power of a PEM fuel cell system using an Extended Kalman Filter. In: 24th Mediterranean Conference on Control and Automation, MED 2016; 2016. p. 106–11. <https://doi.org/10.1109/MED.2016.7535896>.
- [6] Grötsch M, Gundermann M, Mangold M, Kienle A, Sundmacher K. Development and experimental investigation of an extended Kalman filter for an industrial molten carbonate fuel cell system. *J Process Contr* 2006;16(9):985–92.
- [7] Uhlmann JK, Julier SJ. A new extension of the kalman filter to nonlinear systems, *Signal processing, sensor fusion, and target recognition VI*, 3068; 1997. p. 182–94.
- [8] Kandepu R, Foss B, Imsland L. Applying the unscented Kalman filter for nonlinear state estimation. *J Process Contr* 2008;18:753–68. <https://doi.org/10.1016/j.jprocont.2007.11.004>.
- [9] Zhang X, Pisu P. An unscented Kalman filter based on-line diagnostic approach for PEM fuel cell flooding. *Int J Prognostics Health Manag* 2014;5(1):1–18. ISSN 21532648.
- [10] Yuan H, Dai H, Wei X, Ming P. A novel model-based internal state observer of a fuel cell system for electric vehicles using improved Kalman filter approach. *Appl Energy* 2020;268(January):115009. <https://doi.org/10.1016/j.apenergy.2020.115009>. ISSN 0306–2619.
- [11] Vepa R. Adaptive state estimation of a PEM fuel cell. *IEEE Trans Energy Convers* 2012;27(2):457–67. <https://doi.org/10.1109/TEC.2012.2190073>. ISSN 8858969.
- [12] Chen K, Laghrouche S, Djerdir A. Fuel cell health prognosis using Unscented Kalman Filter: postal fuel cell electric vehicles case study. *Int J Hydrogen Energy* 2019;44(3):1930–9. <https://doi.org/10.1016/j.ijhydene.2018.11.100>.
- [13] Schultze M, Horn J. State estimation for pem fuel cell systems with time delay by an unscented kalman filter and predictor strategy. In: 21st mediterranean conference on control and automation. IEEE; 2013. p. 104–12.
- [14] Laghrouche S, Liu J, Ahmed FS, Harmouche M, Wack M. Adaptive second-order sliding mode observer-based fault reconstruction. *IEEE Trans Contr Syst Technol* 2014;23(3):1098–109.
- [15] Luna J, Husar A, Serra M. Nonlinear distributed parameter observer design for fuel cell systems. *Int J Hydrogen Energy* 2015;40(34):11322–32.
- [16] Luna J, Usai E, Husar A, Serra M. Nonlinear observation in fuel cell systems: a comparison between disturbance estimation and High-Order Sliding-Mode techniques. *Int J Hydrogen Energy* 2016;41(43):19737–48. <https://doi.org/10.1016/j.ijhydene.2016.06.041>. ISSN 3603199.
- [17] Liu J, Laghrouche S, Ahmed F-S, Wack M. PEM fuel cell air-feed system observer design for automotive applications: an adaptive numerical differentiation approach. *Int J Hydrogen Energy* 2014;39(30):17210–21.

- [18] Liu J, Luo W, Yang X, Wu L. Robust model-based fault diagnosis for PEM fuel cell air-feed system. *IEEE Trans Ind Electron* 2016;63(5):3261–70. <https://doi.org/10.1109/TIE.2016.2535118>. ISSN 2780046.
- [19] Rao CV, Rawlings JB, Mayne DQ. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Trans Automat Contr* 2003;48(2):246–58. <https://doi.org/10.1109/TAC.2002.808470>. ISSN 189286.
- [20] Haseltine EL, Rawlings JB. Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Ind Eng Chem Res* 2005;44(8):2451–60. <https://doi.org/10.1021/ie034308l>. ISSN 8885885.
- [21] Buchholz M, Pecheur G, Niemeier J, Krebs V. Fault detection and isolation for PEM fuel cell stacks using fuzzy clusters. In: 2007 European control conference, vol. 2007. ECC; 2007. p. 971–7. <https://doi.org/10.23919/ecc.2007.7068720>.
- [22] Rotondo D, Fernandez-Canti RM, Tornil-Sin S, Blesa J, Puig V. Robust fault diagnosis of proton exchange membrane fuel cells using a Takagi-Sugeno interval observer approach. *Int J Hydrogen Energy* 2016;41(4):2875–86. <https://doi.org/10.1016/j.ijhydene.2015.12.071>. ISSN 3603199.
- [23] Lin R-H, Xi X-N, Wang P-N, Wu B-D, Tian S-M. Review on hydrogen fuel cell condition monitoring and prediction methods. *Int J Hydrogen Energy* 2019;44(11):5488–98.
- [24] Tumuluri U. Nonlinear state estimation in polymer electrolyte membrane fuel cells. Ph.D. thesis. Cleveland State University; 2008.
- [25] Pilloni A, Pisano A, Usai E. Observer-based air excess ratio control of a PEM fuel cell system via high-order sliding mode. *IEEE Trans Ind Electron* 2015;62(8):5236–46.
- [26] Rawlings JB, Bakshi BR. Particle filtering and moving horizon estimation. *Comput Chem Eng* 2006;30(10–12):1529–41. <https://doi.org/10.1016/j.compchemeng.2006.05.031>. ISSN 981354.
- [27] Zhakatayev A, Rakhim B, Adiyatov O, Baimyshev A, Varol HA. Successive linearization based model predictive control of variable stiffness actuated robots. In: 2017 IEEE international conference on advanced intelligent mechatronics (AIM). IEEE; 2017. p. 1774–9.
- [28] Pannocchia G. Offset-free tracking MPC : a tutorial review and comparison of different formulations. In: 2015 European control conference (ECC), vol. 527; 2015. p. 527–32. <https://doi.org/10.1109/ECC.2015.7330597> (3).
- [29] Simon D. Kalman filtering with state constraints : a survey of linear and nonlinear algorithms. January 2009. <https://doi.org/10.1049/iet-cta.2009.0032>.
- [30] Kolås S, Foss BA, Schei TS. Constrained nonlinear state estimation based on the UKF approach. *Comput Chem Eng* 2009;33(8):1386–401. <https://doi.org/10.1016/j.compchemeng.2009.01.012>. ISSN 981354.
- [31] Ritzberger D, Hametner C, Jakubek S. A real-time dynamic fuel cell system simulation for model-based diagnostics and control: validation on real driving data. *Energies* 2020;13(12):3148. <https://doi.org/10.3390/en13123148>.
- [32] Kravos A, Ritzberger D, Tavcar G, Ritzberger D, Hametner C, Jakubek S, Katrasnik T. Thermodynamically consistent reduced dimensionality electrochemical model for proton exchange membrane fuel cell performance modelling and control. *J Power Sources* 2020;454. ISSN 3787753. <https://doi.org/10.1016/j.jpowsour.2020.227930>.
- [33] Kravos A, Ritzberger D, Hametner C, Jakubek S, Katrasnik T. Methodology for efficient parametrisation of electrochemical PEMFC model for virtual observers: model based optimal design of experiments supported by parameter sensitivity analysis. *Int J Hydrogen Energy* 2021;46:13832–44. <https://doi.org/10.1016/j.ijhydene.2020.10.146>.
- [34] Simon D. Optimal state estimation: Kalman, Hinf, and nonlinear approaches. 2006. <https://doi.org/10.1002/0470045345>. ISBN 9780470045343.
- [35] Bamimore A, Taiwo O, King R. Comparison of two nonlinear model predictive control methods and implementation on a laboratory three tank system. *Proc IEEE Conf Decis Contr* 2011:5242–7. <https://doi.org/10.1109/CDC.2011.6160244>. ISSN 1912216.
- [36] Keytech4ev project homepage. 2017. <http://www.iesta.at/keytech4ev>. [Accessed 31 August 2020].