An innovations approach to fault diagnosis in linear time-varying descriptor systems*

Abdouramane Moussa Ali[1, 2] and Qinghua Zhang[3]

Abstract— In this paper fault diagnosis is studied for linear time varying descriptor systems, the discrete time counterpart of dynamic systems described by differential-algebraic equations. The Kalman filter for descriptor systems is first revisited by completing existing results about its properties that are essential for the purpose of fault diagnosis. Based on the analysis of the effects of the considered actuator and sensor faults on the innovation of the Kalman filter, it is shown that the considered fault diagnosis problem in linear time varying descriptor systems is equivalent to a classical linear regression problem formulated by appropriately filtering the input-output data. Following this result, algorithms for fault diagnosis through maximum likelihood estimation are then proposed. A numerical example is presented to illustrate the proposed method.

I. INTRODUCTION

Many modern engineering systems can be modeled by explicit ordinary differential equations (ODE) of the following form

$$\dot{x}(t) = f(x(t), u(t)) \tag{1}$$

where $x(t)$ and $u(t)$ represent respectively the (vectorial) state and input of the system, $\dot{x}(t)$ denotes the time derivative of $x(t)$, and $f(\cdot, \cdot)$ is some function characterizing the system. Such state space equations have a long-term mathematical history, and a large number of analytical and numerical tools have been developed for their study.

However, in some cases such an explicit state space model for the considered system is not available. Instead the system may be described by *implicit* differential equations, known as differential-algebraic equations (DAE), of the form

$$F(\dot{x}(t), x(t), u(t)) = 0 \tag{2}$$

where $F(\cdot, \cdot, \cdot)$ is some vector-valued function. If $\dot{x}(t)$ can be solved for from (2), then the DAE can be converted to an ODE, but this operation is not always possible. It is thus necessary to study DAE systems in some situations.

After linearization along a trajectory and discretization in time, the original nonlinear DAE system is approximately described by an *implicit* discrete time state space equation

$$E_k x(k) = A_{k-1} x(k-1) + B_{k-1} u(k-1) + \mu(k-1) \tag{3}$$

where x, u and μ are respectively the discrete time state, the input and the modeling errors indexed by $k = 0, 1, 2, ...,$

and E_k, A_k and B_k are time varying matrices of appropriate sizes. The matrix E_k may not be of full column rank.

Systems governed by (3) are known as *descriptor systems*. This is a general and convenient framework for studying DAE systems which can be appropriately linearized. Some descriptor systems can be simply regarded as implicitly written state-space equations, this is the case of systems in which the matrix E_k has full column rank. In principle, the theory developed for (explicit) state-space equations can be applied to such systems. However, even in this case, it is preferable to study them in the framework of descriptor systems, if E_k is possibly ill-conditioned.

Descriptor systems have attracted considerable attentions in recent decades where great efforts where made to investigate descriptor system theory and applications ([1], [2], [3], [4], [5]). Fault diagnosis (detection and identification of fault) is rarely tackled in the descriptor case, in contrast to the case of systems with classical explicit state space representations, where the theory is relatively well-established ([6], [7], [8], [9], [10], [11] and references therein). For descriptor system fault diagnosis, most studies are on observer-based approaches in the time invariant case ([12], [13], [14], [15], [16]).

In the present paper, fault diagnosis will be studied for the general class of time-varying discrete time linear descriptor systems in the presence of actuator faults and sensor faults, which will be modeled as linearly parametrized additive terms in the descriptor system state equation (3) and in the output equation. The main contribution of this paper is to establish the relationship between the innovation of the *descriptor system* Kalman filter and the considered parametrically modeled faults, which will constitute the basis for fault parameter estimation.

The paper is organized as follows. In Section II, the descriptor system Kalman filter is first revisited from a new point of view, which is essential for the new results presented later in this section for charactering the innovation of this Kalman filter. Section III first establish the relationship between the innovation of the descriptor system Kalman filter and the monitored faults, based on which fault diagnosis algorithms through fault parameter estimation are then proposed. In section IV, a simulation example is used to illustrate the effectiveness of the proposed approach.

II. KALMAN FILTER FOR DESCRIPTOR SYSTEMS

The Kalman filter for descriptor systems has been studied in [1] from the point of view of maximum likelihood estimation. In order to better characterize this Kalman filter,

in this section let us revisit it from another point of view: the estimation of the probability distribution of the state vector conditioned by the past observations.

Consider the problem of estimating the state x of the system described for $k \geq 1$ by

$$E_k x(k) = A_{k-1}x(k-1) + B_{k-1}u(k-1) + \mu(k-1) \quad (4)$$

based on the measurement vector y

$$y(k) = C_k x(k) + D_{k-1}u(k-1) + \nu(k-1) \quad (5)$$

where $x(k) \in \mathbb{R}^n$, $u(k) \in \mathbb{R}^m$ is the input vector (possibly including other known variables), $y(k) \in \mathbb{R}^p$, E_{k+1}, A_k, B_k, C_k, D_k are known real matrices of compatible sizes, $\mu(k)$ and $\nu(k)$ represent modeling/measurement errors and are modeled as mutually independent and zero-mean white Gaussian noises with covariance matrices Q_k and S_k respectively.

We shall make the following assumptions on system (4)-(5).

A1 : The initial state $x(0)$ is Gaussian distributed with known mean $\bar{x}_0$ and variance matrix P_0, and is independent of μ and ν.

A2 : For all $k \geq 1$, the matrix $H_k = \begin{bmatrix} E_k \\ C_k \end{bmatrix}$ is supposed to be full column rank.

A3 : For all $k \geq 0$, the variance matrices Q_k and S_k are positive definite.

From (4)-(5), we obtain the following equation

$$\mathcal{Y}_{0:k} = \mathcal{H}_k X_{0:k} + \mathcal{V}_{0:k} \quad (6)$$

where

$$\mathcal{Y}_{0:k} = \begin{bmatrix} B_0 u(0) \\ \vdots \\ B_{k-1}u(k-1) \\ \bar{x}_0 \\ y(1) - D_0 u(0) \\ \vdots \\ y(k) - D_{k-1}u(k-1) \end{bmatrix}, \quad X_{0:k} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(k) \end{bmatrix}$$

$$\mathcal{V}_{0:k} = \begin{bmatrix} -\mu(0) \\ \vdots \\ -\mu(k-1) \\ -\tilde{x}(0) \\ \nu(0) \\ \vdots \\ \nu(k-1) \end{bmatrix}, \quad \mathcal{H}_k = \begin{bmatrix} -A_0 & E_1 & & \\ & \ddots & & \ddots \\ & & -A_{k-1} & E_k \\ I & & & \\ & C_1 & & \\ & & \ddots & \\ & & & C_k \end{bmatrix}$$

$\tilde{x}(0) \sim \mathcal{N}(0, P_0)$ is a Gaussian vector independent of μ and ν (due to Assumption A1).

Due to the linearity of (6) and the Gaussian distributions of the noises involved in (6), $x(k)$ is also Gaussian distributed. It is thus sufficient to compute the mathematical expectation of $x(k)$ conditioned by the observations up to time k and the corresponding covariance matrix to estimate the distribution of $x(k)$ conditioned by the observations up to time k.

Let the conditional expectation of $x(k)$ be denoted by

$$\hat{x}(k) \triangleq \mathbf{E}[x(k)|Y_{0:k}] \quad (7)$$

where

$$Y_{0:k} = \begin{bmatrix} \bar{x}_0^T & y(1)^T & \cdots & y(k)^T \end{bmatrix}^T \quad (8)$$

and $\mathbf{E}(\cdot)$ denotes the mathematical expectation.

Remark 1: Notice that the evaluation of this conditional expectation depends on the input sequence up to the time $k-1$, which are not explicitly expressed in (7) for the compactness of the notation.

Let $p_{X,Y}, p_{X|Y}$ and p_Y denote respectively the joint, conditional and marginal probability density functions of some random variables vectors X and Y. It follows from equation (6) and the Gaussian distributions of $\mathcal{V}_{0:k}$ that the joint probability density function $p_{X_{0:k},Y_{0:k}}(X_{0:k}, Y_{0:k})$ is, up to a normalization factor, equal to

$$\exp\left(-\frac{1}{2}(\mathcal{Y}_{0:k} - \mathcal{H}_k X_{0:k})\mathcal{R}_{k-1}^{-1}(\mathcal{Y}_{0:k} - \mathcal{H}_k X_{0:k})\right) \quad (9)$$

where $\mathcal{R}_{k-1}$ is the covariance matrix of $\mathcal{V}_{0:k}$.

The conditional density function marginalized for $x(k)$ is given by

$$p_{x(k)|Y_{0:k}}(x(k)|Y_{0:k}) = \int_{\mathbb{R}^{kn}} p_{X_{0:k}|Y_{0:k}}(X_{0:k}|Y_{0:k})dX_{0:k-1} \quad (10)$$

where

$$p_{X_{0:k}|Y_{0:k}}(X_{0:k}|Y_{0:k}) = \frac{p_{X_{0:k},Y_{0:k}}(X_{0:k}, Y_{0:k})}{p_{Y_{0:k}}(Y_{0:k})} \quad (11)$$

$$p_{Y_{0:k}}(Y_{0:k}) = \int_{\mathbb{R}^{(k+1)n}} p_{X_{0:k},Y_{0:k}}(X_{0:k}, Y_{0:k})dX_{0:k}. \quad (12)$$

For any given value of $Y_{0:k}$, the conditional density function $p_{x(k)|Y_{0:k}}(x(k)|Y_{0:k})$ is Gaussian, which is maximized by $\hat{x}(k)$ as defined in (7).

Given the observation sequence $Y_{0:k}$, the conditional density function $p_{X_{0:k}|Y_{0:k}}(X_{0:k}|Y_{0:k})$ is Gaussian, which is maximized by $\mathbf{E}[X_{0:k}|Y_{0:k}]$. The last n components of $\mathbf{E}[X_{0:k}|Y_{0:k}]$ are equal to $\hat{x}(k)$ as defined in (7).

The value $\mathbf{E}[X_{0:k}|Y_{0:k}]$ maximizing $p_{X_{0:k}|Y_{0:k}}(X_{0:k}|Y_{0:k})$ maximizes also $p_{X_{0:k},Y_{0:k}}(X_{0:k}, Y_{0:k})$, as in (11) the denominator is independent of $X_{0:k}$.

It has been shown in [1] (Lemma 2.3 and Lemma 2.4) that the maximization solution of (10) given by $\hat{x}(k)$ defined in (7) is equivalent to the recursive state estimate derived from the following equation

$$Z(k) = H_k x(k) + V(k-1) \quad (13)$$

where

$$Z(k) = \begin{bmatrix} A_{k-1}\hat{x}(k-1) + B_{k-1}u(k-1) \\ y(k) - D_{k-1}u(k-1) \end{bmatrix} \quad (14)$$

$$V(k-1) = \begin{bmatrix} -A_{k-1}\tilde{x}(k-1) - \mu(k-1) \\ \nu(k-1) \end{bmatrix} \quad (15)$$

$\hat{x}(k-1)$ is the estimate of $x(k-1)$ at the previous instant and $\tilde{x}(k-1)$ is the corresponding estimate error with covariance matrix P_{k-1}.

Given $\hat{x}(k-1)$ and $y(k)$, the recursive state estimate $\hat{x}(k)$ derived from (13) maximizes in $x(k)$ the joint probability density function $p_{\hat{x}(k-1),x(k),y(k)}(x(k-1),x(k),y(k))$, which is, up to a normalization factor, equal to

$$\exp\left(-\frac{1}{2}(Z(k)-H_k x(k))R_{k-1}^{-1}(Z(k)-H_k x(k))\right) \quad (16)$$

where R_{k-1} is the covariance matrix of $V(k-1)$ in (13) given by

$$R_{k-1} = \left[\begin{array}{cc} A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1} & 0 \\ 0 & S_{k-1} \end{array}\right] \quad (17)$$

Given $Z(k)$, the maximization in $x(k)$ of (16) leads to

$$\hat{x}(k) = (H_k^T R_{k-1}^{-1} H_k)^{-1} H_k^T R_{k-1}^{-1} Z(k) \quad (18)$$

Note that, the inverse matrices in (16) and (18) are well defined because of Assumptions A2 and A3.

The corresponding estimate error $\tilde{x}(k)$ is given by

$$\tilde{x}(k) = -(H_k^T R_{k-1}^{-1} H_k)^{-1} H_k^T R_{k-1}^{-1} V(k-1) \quad (19)$$

The calculation of the error covariance associated with this estimate is independent of state and is given by

$$P_k = (H_k^T R_{k-1}^{-1} H_k)^{-1}. \quad (20)$$

It has been shown that the distribution of $x(k)$ conditioned by $Y(k)$ is Gaussian, with the conditional mean equal to $\hat{x}(k)$. The covariance matrix of this conditional distribution is equal to that of the estimation error, namely P_k, hence the conditional density function $p_{x(k)|Y(k)}(x(k)|Y(k))$ is fully characterized as $\mathcal{N}(\hat{x}(k), P_k)$.

Consider the following partition

$$[L_k \quad K_k] = P_k H_k^T R_{k-1}^{-1} \quad (21)$$

with $L_k \in \mathbb{R}^{n \times n}$ and $K_k \in \mathbb{R}^{n \times p}$. The dynamics of the state estimation error $\tilde{x}(k)$ can be expressed from (15), (19) and (21), yielding

$$\tilde{x}(0) = x(0) - \bar{x}_0 \quad (22a)$$
$$\tilde{x}(k) = L_k A_{k-1} \tilde{x}(k-1) + [L_k \mu(k-1) - K_k \nu(k-1)] \quad (22b)$$

Finally, the descriptor system Kalman filter is summarized in the following equations

$$\hat{x}(0) = \bar{x}_0 \quad (23a)$$
$$\hat{x}(k) = L_k A_{k-1} \hat{x}(k-1)$$
$$+ [L_k B_{k-1} - K_k D_{k-1}]u(k-1) + K_k y(k) \quad (23b)$$
$$\hat{y}(k) = C_k \hat{x}(k) \quad (23c)$$
$$P_k = L_k A_{k-1} P_{k-1} A_{k-1}^T L_k^T + L_k Q_{k-1} L_k^T + K_k S_{k-1} K_k^T \quad (23d)$$
$$[L_k \quad K_k] = (H_k^T R_{k-1}^{-1} H_k)^{-1} H_k^T R_{k-1}^{-1} \quad (23e)$$
$$R_{k-1} = \left[\begin{array}{cc} A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1} & 0 \\ 0 & S_{k-1} \end{array}\right]. \quad (23f)$$

In filtering theory, the output prediction error defined by

$$\tilde{y}(k) = y(k) - D_{k-1}u(k-1) - C_k \hat{x}(k) \quad (24)$$

is called the innovation. For classical explicit state space systems, the Gaussian distribution and the whiteness property of the innovation sequence is well known. For descriptor systems, this whiteness property was not discussed in [1], nor in any other reference, up to our knowledge. Its extension to linear descriptor systems is summarized in Theorem 1.

Theorem 1: The innovation $\tilde{y}(k)$ of the Kalman filter (23) for the descriptor system (4)-(5) is a centered white Gaussian sequence with the covariance matrix equal to

$$C_k P_k C_k^T + S_{k-1}. \quad (25)$$

The proof, similar to that of [17], is omitted here due to space limitation.

III. INNOVATION BASED FAULT DIAGNOSIS

A. Problem formulation

Consider a descriptor system as expressed in (4)-(5), but now subject to actuator and sensor faults, in the form of

$$E_k x(k) = A_{k-1}x(k-1) + B_{k-1}u(k-1) + \Phi_{k-1}\theta + \mu(k-1)$$
$$y(k) = C_k x(k) + D_{k-1}u(k-1) + \Psi_{k-1}\theta + \nu(k-1) \quad (26)$$

where the terms $\Phi_{k-1}\theta$ and $\Psi_{k-1}\theta$ represent respectively faults possibly affecting the state and the output equations, typically the former for actuator faults, and the latter for sensor faults. The matrices $\Phi_{k-1} \in \mathbb{R}^{n \times q}$ and $\Psi_{k-1} \in \mathbb{R}^{p \times q}$ are assumed known, whereas the *constant* parameter vector $\theta \in \mathbb{R}^q$ is unknown. The fault-free system is characterized by the nominal value $\bar{\theta}$ of the parameter vector. In order to simplify the presentation, the zero nominal value $\bar{\theta} = 0$ is assumed in this paper. It is straightforward to generalize the results to the case of non zero nominal value.

The fault diagnosis approach presented in this paper will be essentially based on the estimation of the fault parameter vector. As the state vector $x(k)$ is also unknown, typically joint state-parameter estimation should be considered in such an approach. A classical solution consists in considering the *augmented system* obtained by concatenating the parameter vector θ into the state vector, so that Kalman filter can be applied to the augmented system. Such a solution has a few drawbacks:

- treating equally $x(k)$ and θ as if they had similar dynamics may make the tuning of the Kalman filter delicate.
- the computational cost of the Kalman filter increases mainly because of the larger covariance matrix of the augmented state estimation error.

The approach presented below decouples to some extent parameter estimation from state estimation.

B. Innovation in the presence of faults

In the considered system modeled by (26), as the value of θ is unknown, it is not possible to apply the Kalman filter to it. Instead, let us apply the Kalman filter to the fault-free system (corresponding to $\theta = 0$), no matter if the considered system is affected by faults or not. The main idea of this paper is to analyze the innovation sequence of this Kalman filter *designed for the fault-free system* for the purpose of

fault diagnosis. If the monitored system is indeed fault-free, then the innovation is a centered white Gaussian sequence, as shown in Theorem 1. However, if the monitored system is affected by some faults ($\theta \neq 0$), the same Kalman filter, which has been designed for the fault-free system, will have an innovation sequence with different properties. The full characterization of the properties of the innovation sequence in the faulty case will be helpful for the diagnosis of faults corresponding to changes in θ. This characterization is the purpose of the following theorem.

Theorem 2: When the system (26) is affected by faults modeled by $\Phi_{k-1}\theta$ and/or $\Psi_{k-1}\theta$, then the innovation sequence $\tilde{y}(k)$ of the Kalman filter (23) designed for the fault-free system and applied to (26) satisfies

$$\tilde{y}(k) = \tilde{y}^0(k) + \Omega_k\theta \qquad (27)$$

where $\tilde{y}^0(k)$ is the innovation sequence that the fault-free system would generate with the same Kalman filter and Ω_k is a matrix generated by linearly filtering Φ_{k-1} and Ψ_{k-1} through the state space equations

$$\Upsilon_0 = 0 \qquad (28a)$$
$$\Upsilon_k = L_k A_{k-1}\Upsilon_{k-1} + L_k\Phi_{k-1} - K_k\Psi_{k-1} \qquad (28b)$$
$$\Omega_k = C_k\Upsilon_k + \Psi_{k-1} \qquad (28c)$$

The proof, omitted here due to space limitation, is obtained by adapting the approach of [18] with the main step analyzing the dynamics of the variable

$$\eta(k) = \tilde{x}(k) - \Upsilon_k\theta \qquad (29)$$

Remark 2: Up to the white Gaussian noise term $\tilde{y}^0(k)$, the only unknown in (27) is the fault parameter vector θ, hence this equation can be viewed as a linear regression with the regression coefficients contained in θ. The original fault diagnosis problem in the descriptor system (26) is then equivalent to the much simpler problem formulated in the linear regression equation (27).

The method adopted in this paper is based on parameter estimation: after the estimation of θ, fault diagnosis can be easily made by comparing the estimated value to the known nominal value, by taking into account the uncertainty related to the white Gaussian noise $\tilde{y}^0(k)$ in (27).

C. Maximum likelihood estimation of the fault parameter vector

In this subsection, the fault parameter vector θ is first estimated based on data collected for time instant $i = 1, \ldots, k$. In principle, this estimation assumes constant θ during the collection of the data. In practice, it is assumed that the value of θ evolves slowly so that it can be considered constant within the data window. In this case fault diagnosis can be made by processing data in batch.

Alternatively, it is assumed that the parameter vector θ is constant most of the time, but jumps from time to time. This case is discussed in the second part of this subsection.

1) Batch estimation: For compact notations, let us define the $kp \times 1$ vector $\tilde{Y}_{1:k}$, the $kp \times q$ matrix $G_{1:k}$ and the $kp \times 1$ vector $\tilde{V}_{1:k}$ as

$$\tilde{Y}_{1:k} = \begin{bmatrix} \tilde{y}(1) \\ \vdots \\ \tilde{y}(k) \end{bmatrix}, \ G_{1:k} = \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \cdots \\ \Omega_k \end{bmatrix}, \ \tilde{V}_{1:k} = \begin{bmatrix} \tilde{y}^0(1) \\ \tilde{y}^0(2) \\ \cdots \\ \tilde{y}^0(k) \end{bmatrix} \qquad (30)$$

Then equation (27) evaluated at different time instants can be compactly written as

$$\tilde{Y}_{1:k} = G_{1:k}\theta + \tilde{V}_{1:k} \qquad (31)$$

According to Theorems 1 and 2, the sub-vectors $\tilde{y}^0(k)$ of $\tilde{V}_{1:k}$ follow each the Gaussian distribution

$$\mathcal{N}\left(0, (C_k P_k C_k^T + S_{k-1})\right)$$

and are independent between different time instants.

Assume that the matrix

$$\Lambda_k = \sum_{i=1}^{k} \Omega_i^T (C_i P_i C_i^T + S_{i-1})^{-1}\Omega_i$$

is invertible (it implies that Φ_k and Ψ_k vary sufficiently), then the maximum likelihood estimation of θ is given by

$$\hat{\theta}(k) = \Lambda_k^{-1}\sum_{i=1}^{k}\Omega_i^T(C_i P_i C_i^T + S_{i-1})^{-1}\tilde{y}(i) \quad (32)$$

One can easily show that the estimation of the parameter vector θ given by (32) is an unbiased minimum-variance estimator. After the computation of the parameter estimate $\hat{\theta}(k)$, its error covariance matrix given by Λ_k^{-1} provides a confidence region of the true parameter vector θ, which then can be used to decide if the parameter estimate $\hat{\theta}(k)$ is significantly different from the nominal parameter value or not.

2) Jumps in the parameter vector: Let us assume that the parameter vector θ may change abruptly from one level to another at some instants, but such changes are sufficiently rare events so that there is at most a single jump of the parameter vector in the considered time window, say from instant 1 to k.

If a jump occurred at some instant j between 1 and k, Theorem 2 can no longer be applied, though the value of θ is kept constant before and after the jump. In this case, Theorem 2 can be replaced by the following result.

Corollary 1: If the parameter vector $\theta = 0$ before the time instant j and then takes another constant value, the innovation sequence satisfies

$$\tilde{y}(k) = \tilde{y}^0(k) + \Omega_k\theta \qquad (33)$$

where $\tilde{y}^0(k)$ is, as before, the fault-free innovation, but Ω_k is a matrix generated by linearly filtering $\tilde{\Phi}_{k-1}$ and $\tilde{\Psi}_{k-1}$ through the state space equations

$$\Upsilon_0 = 0 \qquad (34a)$$
$$\Upsilon_k = L_k A_{k-1}\Upsilon_{k-1} + [L_k\tilde{\Phi}_{k-1} - K_k\tilde{\Psi}_{k-1}] \qquad (34b)$$
$$\Omega_k = C_k\Upsilon_k + \tilde{\Psi}_{k-1} \qquad (34c)$$

with $\tilde{\Phi}_{k-1} = \Phi_{k-1}\mathbb{1}_{\{k \geq j\}}$, $\tilde{\Psi}_{k-1} = \Psi_{k-1}\mathbb{1}_{\{k \geq j\}}$ and $\mathbb{1}_{\{\cdot\}}$ being the indicator function. $\qquad\square$

In practice, the jump time j is usually unknown. To solve this problem, the estimation of θ can be made once for each assumed jump time $j = 1, ..., k$, what is computationally expensive.

D. Recursive estimation of the fault parameter vector

The parameter estimation based on (32) is unreliable if jumps of the parameter vector θ occur within the considered data window, as the constant parameter assumption is then violated. Parameter jumps can be handled with techniques based on variable/time-varying forgetting factors ([19]), covariance matrix re-setting/modification ([20]), and sliding window ([21]). The recursive algorithm proposed below provides a simple solution with a minimum computer memory requirement, and yet with an efficient treatment of parameter jumps.

By using the matrix inversion lemma

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(DA^{-1}B + C^{-1}\right)^{-1}DA^{-1}$$

it easy to deduce from (32) the following recursive form

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \Theta_k(y(k) - \Omega_k\hat{\theta}(k-1)) \qquad (35)$$

with the matrix gain computed from the following equations:

$$\begin{cases} \Theta_k = M_{k-1}\Omega_k^T(C_kP_kC_k^T + S_{k-1} + \Omega_kM_{k-1}\Omega_k^T)^{-1} \\ M_k = M_{k-1} - \Theta_k\Omega_kM_{k-1} \end{cases} \qquad (36)$$

If the parameter estimate $\hat{\theta}(k)$ and the matrix M_k are appropriately initialized, then the recursive estimate $\hat{\theta}(k)$ given by (35) is equivalent to the batch estimate (32). In order to exponentially forget past observations, it is well known that a forgetting factor can be introduced, so that the matrix gain computation is modified as

$$\begin{cases} \Theta_k = M_{k-1}\Omega_k^T \left(\lambda[C_kP_kC_k^T + S_{k-1}] + \Omega_kM_{k-1}\Omega_k^T\right)^{-1} \\ M_k = \frac{1}{\lambda}M_{k-1} - \frac{1}{\lambda}\Theta_k\Omega_kM_{k-1} \end{cases} \qquad (37)$$

where $\lambda \in (0, 1)$ is a forgetting factor.

The inverse of the matrix M_k provides an estimate of the parameter estimation error covariance matrix, which can be used to derive a confidence region of the true parameter vector for the purpose of fault diagnosis.

Jumps in some parameters may cause a large fluctuation in the parameter estimates during some transient interval after the jump, typically impacting the estimates of all the components of θ, though some of them have not changed. To improve the tracking capability of the above algorithm, a detection function is used to determine whether a parameter change has occurred or not. Once a change is detected at some instant j, then some elements of the estimation algorithm are modified as follows. In the computation of Υ_k, a reinitialization is made at instant j ($\Upsilon_j = 0$) and in the Kalman filter the zero nominal parameter is replaced by the estimate $\hat{\theta}(j_c)$ computed after the transient time ($j_c \geq j$) and to considering the detection time j_c as the new initial

time instant in the algorithm. The detection function can be derived from the innovation $\tilde{y}(k)$ as follow

$$\epsilon(k) = \left(\tilde{y}(k) - \Omega_k\hat{\theta}(k-1)\right)^T [C_kP_kC_k^T + S_{k-1}]^{-1} \cdot \left(\tilde{y}(k) - \Omega_k\hat{\theta}(k-1)\right) \qquad (38)$$

to be compared with a detection threshold to be chosen appropriately. As $\tilde{y}(k)$ has Gaussian distribution, $\epsilon(k)$ has chi-square distribution with one degree of freedom, thus the threshold can be determined by looking up the chi-square table following a specified probability of false alarm ([22], [6]).

IV. NUMERICAL EXAMPLE

Consider a nonlinear system described by

$$\begin{cases} F(\dot{x}, x, t) = 0 \\ y = x_5 + \theta_2 \end{cases} \qquad (39)$$

with

$$F(\dot{x}, x, t) = \begin{bmatrix} x_5\dot{x}_1 + \theta_1\alpha(t) + gx_3 \\ x_5\dot{x}_2 + \theta_1\beta(t) + gx_4 - gx_5 \\ \dot{x}_3 - x_1 \\ \dot{x}_4 - x_2 \\ x_3^2 + x_4^2 - x_5^2 \end{bmatrix} \qquad (40)$$

where $g = 10$, θ_2 represents a possible bias on the measurement and

$$\alpha(t) = -2 - 0.1\sin(0.3t) - 0.2\sin(0.5t) \qquad (41)$$
$$\beta(t) = 2 + 0.2\sin(0.3t) - 0.4\sin(0.5t) \qquad (42)$$

For this illustrative example, the nominal parameter vector is given by $\bar{\theta}_1 = 1$ and $\bar{\theta}_2 = 0$.

After discretization in time and by replacing the derivative term by backward Euler approximation, then a linearization along a given trajectory $\mathbf{z}(t) = (0, 0, 0, -l, l)$ where $l > 0$, the continuous nonlinear system in (39) is approximately described by the following implicit discrete time state space equation

$$\begin{cases} E_k\mathbf{x}(k) = A_{k-1}\mathbf{x}(k-1) + \Phi_{k-1}\theta + \mu(k-1) \\ \mathbf{y}(k) = C_k\mathbf{x}(k) + \Psi_{k-1}\theta + \nu(k-1) \end{cases} \qquad (43)$$

where

$$E_k = \begin{bmatrix} \frac{l}{\Delta_k} & 0 & g & 0 & 0 \\ 0 & \frac{l}{\Delta_k} & 0 & g & 0-g \\ -1 & 0 & \frac{1}{\Delta_k} & 0 & 0 \\ 0 & -1 & 0 & \frac{1}{\Delta_k} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$A_{k-1} = \begin{bmatrix} \frac{l}{\Delta_k} & 0 & 0 & 0 & 0 \\ 0 & \frac{l}{\Delta_k} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\Delta_k} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\Delta_k} & 0 \\ 0 & 0 & 0 & 2l & 2l \end{bmatrix}$$

$\Delta_k = t(k) - t(k-1)$ is the sampling period for the discrete time model, $C_k = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$, $\Psi_{k-1} = [0,\ 1]$ and

$$\Psi_{k-1} = [0,\ 1] \text{ and } \Phi_{k-1} = \begin{bmatrix} -\alpha(t_k) & 0 \\ -\beta(t_k) & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The sampling period for the discrete time model is 1s.

The nominal (computed) output is compared with the simulated output in figures 1, where dotted lines represent the true simulated values, and the solid lines represent the nominal values.

The recursively estimated values of θ_1 and θ_2 are respectively illustrated in figure 2, where the dotted lines represent the true simulated parameter values, and the solid lines represent the estimated values.

These results show that, after the transient time corresponding to the unmodeled transient of θ, the parameter estimates follow closely the simulated parameter. It is then possible to detect the simulated faults and estimate their severity.
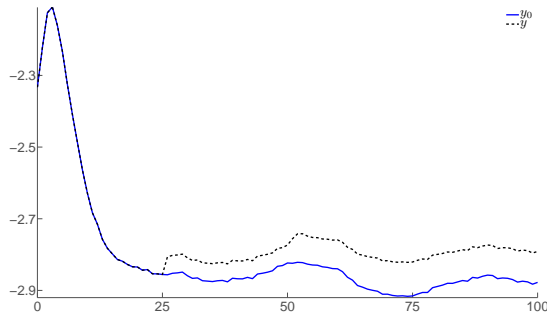

Fig. 1. Graphical representations of the simulated noised output y in dotted line and the fault-free system output y_0 in solid line, over time (k).
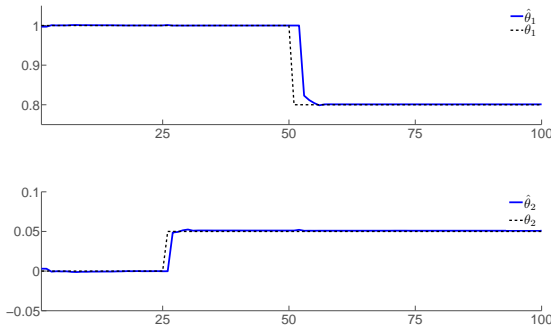

Fig. 2. Graphical representations of the simulated parameters θ_1, θ_2 in dotted line and their estimates $\hat\theta_1$, $\hat\theta_2$ in solid line, over time (k).

V. CONCLUSION

This paper has dealt with fault diagnosis in descriptor systems. We have focused our study on the diagnosis of faults

modeled as abrupt parameter changes in a class of discrete-time descriptor systems. The descriptor system Kalman filter is used to generate an innovation sequence. We have shown that the innovation sequence is white and can be used to estimate possible changes in the parameter vector. The implementation of this fault diagnosis approach requires knowledge of the nominal parameter. The decision for fault diagnosis is based on the time evolution of parameter estimates.

REFERENCES

[1] R. Nikoukhah, A. Willsky, and B. Levy, "Kalman filtering and riccati equations for descriptor systems," *IEEE Trans. on Automatic Control*, vol. 37, no. 9, pp. 1325 –1342, 1992.

[2] M. Darouach and M. Boutayeb, "Design of observers for descriptor systems," *IEEE Trans. on Automatic Control*, vol. 40, no. 7, pp. 1323 –1327, 1995.

[3] D. N. Shields, "Observer design and detection for nonlinear descriptor systems," *Int. J. of Control*, vol. 67, no. 2, pp. 153–168, 1997.

[4] A. Vemuri, M. Polycarpou, and A. Ciric, "Fault diagnosis of differential-algebraic systems," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 31, no. 2, pp. 143 –152, 2001.

[5] D. Koenig and S. Mammar, "Design of proportional-integral observer for unknown input descriptor systems," *IEEE Trans. on Automatic Control*, vol. 47, no. 12, pp. 2057 – 2062, 2002.

[6] M. Basseville and I. Nikiforov, *Detection of abrupt changes: theory and application*. Prentice Hall, 1993.

[7] M. Blanke, M. Kinnaert, J. Lunze, and J. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. Springer, 2003.

[8] S. X. Ding, *Model-Based Fault Diagnosis Techniques - Design Schemes Algorithms and tools*. Springer-Verlag, 2008.

[9] J. Chen and R. J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Boston, USA: Kluwer, 1999.

[10] J. J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. New York, USA: Marcel Dekker, 1998.

[11] R. Isermann, *Fault-Diagnosis System*. Berlin: Springer, 2006.

[12] G. R. Duan, D. Howe, and R. J. Patton, "Robust fault detection in descriptor linear systems via generalized unknown input observers," *Int. J. of Systems Science*, vol. 33, no. 5, pp. 369–377, 2002.

[13] B. Marx, D. Koenig, and D. Georges, "Robust fault-tolerant control for descriptor systems," *IEEE Trans. on Automatic Control*, vol. 49, no. 10, pp. 1869 – 1876, 2004.

[14] M. Darouach, M. Zasadzinski, and H. Hayar, "Reduced-order observer design for descriptor systems with unknown inputs," *IEEE Transactions on Automatic Control,*, vol. 41, no. 7, pp. 1068–1072, 1996.

[15] Z. Gao and H. Wang, "Descriptor observer approaches for multi-variable systems with measurement noises and application in fault detection and diagnosis," *Systems & Control Letters*, vol. 55, no. 4, pp. 304 – 313, 2006.

[16] T. Yeu and S. Kawaji, "Fault detection and isolation for descriptor systems using sliding mode observer," in *Proceedings of the 40th IEEE Conference on Decision and Control*, vol. 1, 2001, pp. 596–597.

[17] T. Kailath, "An innovations approach to least-squares estimation–part i: Linear filtering in additive white noise," *IEEE Transactions on Automatic Control*, vol. 13, no. 6, pp. 646 – 655, 1968.

[18] Q. Zhang and M. Basseville, "Statistical fault detection and isolation for linear time-varying systems," in *16th IFAC Symposium on System Identification (SYSID)*, 2012, pp. 625–630.

[19] L. Ljung and T. Söderström, *Theory and practice of recursive identification*, ser. MIT Press series in signal processing, optimization, and control. MIT Press, 1985.

[20] J. Jiang and R. Cook, "Fast parameter tracking rls algorithm with high noise immunity," *Electronics Letters*, vol. 28, no. 22, pp. 2043–2045, 1992.

[21] B. Choi and Z. Bien, "Sliding-windowed weighted recursive least-squares method for parameter estimation," *Electronics Letters*, vol. 25, no. 20, pp. 1381–1382, 1989.

[22] H. G. Zhang, J. Chen, and H. Zhang, "A recursive identification algorithm for estimating abrupt parameter changes," in *International Conference on Control*, 1991, pp. 220–222 vol.1.