HAL open science

Diagnosis of PEMFC by using data-driven parity space strategy

Zhongliang Li, Rachid Outbib, Daniel Hissel, Stefan Giurgea

To cite this version:

Zhongliang Li, Rachid Outbib, Daniel Hissel, Stefan Giurgea. Diagnosis of PEMFC by using data-driven parity space strategy. 2014 European Control Conference (ECC), Jun 2014, Strasbourg, France. pp.1268-1273, 10.1109/ECC.2014.6862527 . hal-02476458

HAL Id: hal-02476458

https://hal.science/hal-02476458v1

Submitted on 12 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Autorisation HAL

HAL Authorization

Diagnosis of PEMFC by using data-driven parity space strategy

Zhongliang Li, Rachid Outbib, Daniel Hissel, and Stefan Giurgea

Abstract— **In this paper, a data-driven strategy is proposed for PEMFC (polymer electrolyte membrane fuel cell) diagnosis. In the strategy, *parity space* is directly identified from normal process data without modeling. With the identified *parity space*, a group residuals can be generated and evaluated to achieve fault detection. In addition, a multi-class SVM (support vector machine) is adopted to realize fault isolation. Experiments of a 40-cell stack are dedicated to highlight the approach.**

I. INTRODUCTION

Increasing environment and resource issues motivate the development and commercialization of fuel cell technologies. PEMFC is one of the most promising fuel cells, especially for mobile applications. Fault diagnosis, considered as a crucial factor of PEMFC system operations, is currently receiving considerably increasing attention.

Due to the cost and the requirement of system integration, hardware redundancy is rarely considered as a practical diagnosis solution. In the literature, model-based (analytical model) diagnosis is an intuitional way to realize the aim of fault detection and isolation (FDI) [1]. If a process model is available, a large number of standard methods are available for the FDI design. However, modeling a PEMFC stack is a rather difficult task, identification of fuel cell inner parameters is difficult to obtain [2]. Besides, some accurate models require considerable processing time, and are not always suitable for diagnosis in real time.

Some data-driven methods have also been proposed for PEMFC diagnosis [3] [4] [5] [6]. These methods deal with diagnosis from three directions.

- In the first direction, fault diagnosis is recognized as a classification problem, historical data is used to train the artificial intelligent classifiers, through which measured data can be classified to different conditions. Some signal processing methods are often adopted to process the raw data to extract some features for classification [4]. Since data in various classes are needed, these methods usually suffer from lacking the robustness against unseen data. Besides, some signal processing methods fail to maintain efficiency in a high dynamic process.
- In the second direction, the process model of PEMFC stack is imitated by a topology of artificial intelligent (AI) models, such as neuro-networks [3]. As conventional model-based operation, fault diagnosis is then realized via a residual evaluation and inference procedures. Although the AI models can offer good ability of modeling the complex process of PEMFC system, the residual generation and evaluation of AI models are not so well established as the conventional model. Additionally, as AI models belongs to "black-box" models, the difficulty of explaining the causal relationship among variables exists.
- In the third direction, multivariate analysis techniques, such as PCA (principal component analysis), are used for PEMFC diagnosis [6]. Although multivariate analysis methods are thought to be simple and efficiency to realize FDI, they seems only efficient in dealing with the diagnosis problems in steady state [7].

It should be noticed that there are advantages and disadvantages existing in both model-based and data-driven diagnosis. Hence, a proper combination of them seems to be attractive. Recently, authors of [8] proposed a FDI design scheme, in which *parity vectors* can be identified from the process data with the aid of SIM (subspace identification method), and then used for a model-based fault diagnosis design. The attractive advantages of these methods are that, the well-established model-based fault diagnosis tools can be adopted; the sophisticated modeling work can be avoided.

Motivated by the aforementioned approach, our study is dedicated to apply this new data-driven FDI design scheme to PEMFC diagnosis problem. Moreover, some contributions are made to complete and improve the diagnosis operation:

- An orthogonal projection approach is employed to the procedure of *parity vectors* identification.
- Fault isolation is realized by carrying out a multi-class SVM (support vector machine) in residual space.

The rest of the paper is organized as follows: In section 2, the diagnosis strategy is expounded, including presentations of *parity space* based diagnosis, *parity vectors* identification, and fault isolation in residual space. In section 3, . The diagnostic results are given in section 4, Finally, the conclusion and future work are summarized in section 5.

II. DESCRIPTION OF APPROACH

A. Outline

The approach contains off-line and on-line operations. In the off-line part, *parity space* is identified from normal process data, and the multi-class SVM is trained with the

Z. Li and R. Outbib are with the LSIS Laboratory, UMR CNRS 6168, University of Aix-Marseille, 13397 Marseille Cedex 20, France (e-mail: zhongliang.li@lsis.org; rachid.outbib@lsis.org). Z.li is also with FEMTO-ST/ENERGY (UMR CNRS 6174) and FCLAB, Belfort 90010 Cedex, France

D. Hissel is with the University of Franche-Comté (UFC), with FEMTO-ST/ENERGY (UMR CNRS 6174) and FCLAB, Belfort 90010 Cedex, France (e-mail: daniel.hissel@univ-fcomte.fr).

S. Giurgea is with the University of Technology Belfort-Montbéliard (UTBM) and FCLAB, Belfort 90010 Cedex, France (e-mail: stefan.giurgea@utbm.fr).

faulty data. In the on-line diagnostic stage, the fault detection and isolation procedures for realtime data can be realized with the obtained *parity space* and trained multi-class SVM. Notice that, the faulty data for training multi-class SVM is usually filtered from normal data with the aid of the *parity space*.

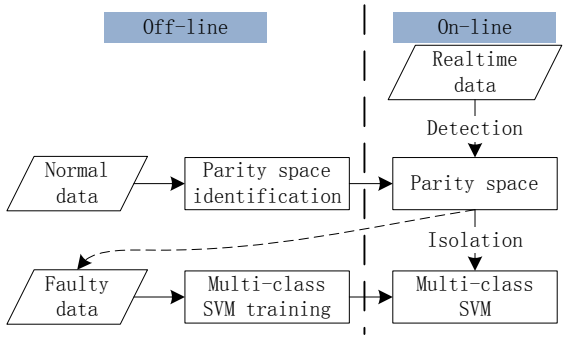

Fig. 1. Flow chart of the diagnosis strategy

B. Process description

Suppose that a process data set including process input and output records is available. These data can be represented by

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k) + \boldsymbol{w}(k) \\ \boldsymbol{y}(k) &= \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{D}\boldsymbol{u}(k) + \boldsymbol{v}(k) \end{aligned} \tag{1}$$

where $\boldsymbol{u}(k) \in \mathbb{R}^l$, $\boldsymbol{y}(k) \in \mathbb{R}^m$, and $\boldsymbol{x}(k) \in \mathbb{R}^n$ are input, output and state variables. $\boldsymbol{w}(k) \in \mathbb{R}^l$, $\boldsymbol{v}(k) \in \mathbb{R}^m$ are zero-mean, normal distributed white noises.

C. Parity space diagnostic approach

Various model-based fault diagnosis methods have been developed, among them the parity relation-based schemes have received much attention. The design of parity relation based residual generator can be achieved in a straightforward manner.

From (1), it can be derived that

$$\boldsymbol{y}_s(k) = \boldsymbol{\Gamma}_s \boldsymbol{x}(k-s) + \boldsymbol{H}_{s,u}\boldsymbol{u}_s(k) + \boldsymbol{H}_{s,w}\boldsymbol{w}_s(k) + \boldsymbol{v}_s(k) \tag{2}$$

where

$$\boldsymbol{y}_s(k) = \begin{bmatrix} \boldsymbol{y}(k-s) \\ \boldsymbol{y}(k-s+1) \\ \vdots \\ \boldsymbol{y}(k) \end{bmatrix}$$

s is the order of the *parity space*, $\boldsymbol{u}_s(k)$, $\boldsymbol{w}_s(k)$, and $\boldsymbol{v}_s(k)$ are defined similarly. $\boldsymbol{\Gamma}_s$, $\boldsymbol{H}_{s,u}$, and $\boldsymbol{H}_{s,w}$ are defined

$$\boldsymbol{\Gamma}_s = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^s \end{bmatrix} \in \mathbb{R}^{m(s+1)\times n}$$

$$\boldsymbol{H}_{s,u} = \begin{bmatrix} \boldsymbol{D} & 0 & \dots & 0 \\ \boldsymbol{C}\boldsymbol{B} & \boldsymbol{D} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{C}\boldsymbol{A}^{s-1}\boldsymbol{B} & \dots & \boldsymbol{C}\boldsymbol{B} & \boldsymbol{D} \end{bmatrix} \in \mathbb{R}^{m(s+1)\times l(s+1)}$$

$$\boldsymbol{H}_{s,w} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \boldsymbol{C} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{C}\boldsymbol{A}^{s-1} & \dots & \boldsymbol{C} & 0 \end{bmatrix} \in \mathbb{R}^{m(s+1)\times n(s+1)}$$

Assume that $(\boldsymbol{C}, \boldsymbol{A})$ is observable, for $s > n$, $rank(\boldsymbol{\Gamma}_s) = n$. This ensures that there exists at least a row vector $\boldsymbol{\alpha}_s$:

$$\boldsymbol{\alpha}_s \boldsymbol{\Gamma}_s = 0 \tag{3}$$

Vectors satisfying (3) are called *parity vectors*, the set of which,

$$\boldsymbol{P}_s = \{\boldsymbol{\alpha}_s | \boldsymbol{\alpha}_s \boldsymbol{\Gamma}_s = 0\} \tag{4}$$

is called the *parity space*.

Consequently, a parity relation based residual generator can be constructed as

$$\begin{aligned} r(k) &= \boldsymbol{\alpha}_s \left(\boldsymbol{y}_s(k) - \boldsymbol{H}_{u,s}(k)\boldsymbol{u}_s(k)\right) \\ &= \boldsymbol{\alpha}_s \left(\boldsymbol{H}_{s,w}\boldsymbol{w}_s(k) + \boldsymbol{v}_s(k)\right) \end{aligned} \tag{5}$$

The value of $r(k)$ should be zero when the disturbances and noises are zero. Notice that, the *parity space* is usually multi-dimensional, and the number of the residuals are equal to the dimensional number of *parity space*. Hence, for a sample $(\boldsymbol{u}_s(k), \boldsymbol{y}_s(k))$, the multiple residuals could be formulated by a *residual vector* $\boldsymbol{r}(k)$.

Followed by *residual evaluation* procedure, fault detection could be realized. In this study, the *residual evaluation* is achieved by using the following decision rule:

$$R = ||\boldsymbol{r}(k)||^2 \begin{cases} \leq Th & \text{Fault free} \\ > Th & \text{Fault} \end{cases} \tag{6}$$

where Th is the pre-defined threshold.

Normally, for the detection design of a given model, the parameters to be designed are the *parity vectors*; whereas for the case that the model is not available (our case), both $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_s\boldsymbol{H}_{u,s}$ are needed.

D. Parity space identification

To present the methodologies conveniently, we define $\boldsymbol{z}(k) = \begin{bmatrix} \boldsymbol{y}(k) \\ \boldsymbol{u}(k) \end{bmatrix}$. In addition, by introducing integers s_p and s_f with $s_p, s_f > n$, the past and future vectors are defined as

$$\boldsymbol{y}_p(k) = \begin{bmatrix} \boldsymbol{y}(k-s_p) \\ \boldsymbol{y}(k-s_p+1) \\ \vdots \\ \boldsymbol{y}(k-1) \end{bmatrix} \in \mathbb{R}^{ms_p} \tag{7}$$

$\boldsymbol{u}_p(k)$, $\boldsymbol{z}_p(k)$, $\boldsymbol{w}_p(k)$ and $\boldsymbol{v}_p(k)$ are defined similarly;

$$\boldsymbol{y}_f(k) = \begin{bmatrix} \boldsymbol{y}(k) \\ \boldsymbol{y}(k+1) \\ \vdots \\ \boldsymbol{y}(k+s_f-1) \end{bmatrix} \in \mathbb{R}^{ms_f} \tag{8}$$

$\boldsymbol{u}_f(k)$, $\boldsymbol{z}_f(k)$, $\boldsymbol{w}_f(k)$ and $\boldsymbol{v}_f(k)$ are defined similarly.

The corresponding matrices are defined as

$$\boldsymbol{Y}(k) = \left[\boldsymbol{y}(k), \boldsymbol{y}(k+1), \ldots, \boldsymbol{y}(k+N-1)\right] \in \mathbb{R}^N \quad (9)$$

$\boldsymbol{U}(k)$, $\boldsymbol{Z}(k)$, $\boldsymbol{W}(k)$ and $\boldsymbol{V}(k)$ are defined similarly;

$$\boldsymbol{Y}_f = \begin{bmatrix} \boldsymbol{Y}(k) \\ \boldsymbol{Y}(k+1) \\ \vdots \\ \boldsymbol{Y}(k+s_f-1) \end{bmatrix} \in \mathbb{R}^{ms_f \times N} \quad (10)$$

$\boldsymbol{U}_f$, $\boldsymbol{Z}_f$, $\boldsymbol{W}_f$ and $\boldsymbol{V}_f$ are defined similarly.

$$\boldsymbol{Y}_p = \begin{bmatrix} \boldsymbol{Y}(k-s_p) \\ \boldsymbol{Y}(k-s_p+1) \\ \vdots \\ \boldsymbol{Y}(k-1) \end{bmatrix} \in \mathbb{R}^{ms_p \times N} \quad (11)$$

$\boldsymbol{U}_p$, $\boldsymbol{Z}_p$, $\boldsymbol{W}_p$ and $\boldsymbol{V}_p$ are defined similarly.

SIM is an approach used for the state space model identification. In [9], the authors proposed a subspace identification approach through an orthogonal projection and subsequent singular value decomposition. The performance of the method seems to be superior to a number of existing SIM algorithms in some aspects [10]. In our study, part of the method is adopted to identify the *parity space*.

Based iterating (1), we can obtain

$$\boldsymbol{Y}_f = \boldsymbol{\Gamma}_{s_f-1}\boldsymbol{X}(k) + \boldsymbol{H}_{s_f-1,u}\boldsymbol{U}_f + \boldsymbol{H}_{s_f-1,w}\boldsymbol{W}_f + \boldsymbol{V}_f \quad (12)$$

where $\boldsymbol{\Gamma}_{s_f-1}$, $\boldsymbol{H}_{s_f-1,u}$, and $\boldsymbol{H}_{s_f-1,w}$ are defined as $\boldsymbol{\Gamma}_s$, $\boldsymbol{H}_{s,u}$, and $\boldsymbol{H}_{s,w}$ with s replaced by s_f-1. The essential system information is contained in the extended observability matrix $\boldsymbol{\Gamma}_{s_f-1}$.

With equation (12), the parity identification problem can be formulated as follows: Given the fault-free measurement sequences $\{\boldsymbol{u}_k\}$ and $\{\boldsymbol{y}_k\}$ (in the form of $\boldsymbol{U}_p$, $\boldsymbol{U}_f$, $\boldsymbol{Y}_p$ and $\boldsymbol{Y}_f$), we want to estimate the *parity vector* $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_s\boldsymbol{H}_{u,s}$.

We recall the definition of *orthogonal projection*: The orthogonal projection of the row space of $\boldsymbol{A}$ onto the row space of $\boldsymbol{B}$ is denoted by $\boldsymbol{A}/\boldsymbol{B}$ and can be calculated through

$$\boldsymbol{A}/\boldsymbol{B} = \boldsymbol{A}\boldsymbol{B}^{\dagger}\boldsymbol{B} \quad (13)$$

where $\boldsymbol{B}^{\dagger}$ is the pseudo inverse of B, which can be expressed $\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1}$.

From (12), we can deduce

$$\begin{aligned}[\boldsymbol{I}| - \boldsymbol{H}_{s_f-1,u}]\boldsymbol{Z}_f/\boldsymbol{Z}_p = \boldsymbol{\Gamma}_{s_f-1}\boldsymbol{X}(k)/\boldsymbol{Z}_p \\ + \boldsymbol{H}_{s_f-1,w}\boldsymbol{W}_f/\boldsymbol{Z}_p + \boldsymbol{V}_f/\boldsymbol{Z}_p\end{aligned} \quad (14)$$

Since the noise of the future is independent with the past input and output matrix $\boldsymbol{Z}_p$, the last two terms of the right side of (14): $\boldsymbol{H}_{s_f-1,w}\boldsymbol{W}_f/\boldsymbol{Z}_p$ and $\boldsymbol{V}_f/\boldsymbol{Z}_p$ are all zeros, so

$$[\boldsymbol{I}| - \boldsymbol{H}_{s_f-1,u}]\boldsymbol{Z}_f/\boldsymbol{Z}_p = \boldsymbol{\Gamma}_{s_f-1}\boldsymbol{X}(k)/\boldsymbol{Z}_p \quad (15)$$

Now multiplying both sides of (15) by orthogonal column space of $\boldsymbol{\Gamma}_{s_f-1}$, denoted by $\boldsymbol{\Gamma}^{\perp}_{s_f-1}$,

$$(\boldsymbol{\Gamma}^{\perp}_{s_f-1})^T[\boldsymbol{I}| - \boldsymbol{H}_{s_f-1,u}]\boldsymbol{Z}_f/\boldsymbol{Z}_p = 0 \quad (16)$$

Notice that $\boldsymbol{\Gamma}_s$ and $\boldsymbol{\Gamma}_{s_f-1}$ would be equal, if we make $s = s_f - 1$. It is also the case for $\boldsymbol{H}_{s,u}$ and $\boldsymbol{H}_{s_f-1,u}$. Thus, the row vector of $(\boldsymbol{\Gamma}^{\perp}_{s_f-1})^T$ can be seen as the parity vector $\boldsymbol{\alpha}_s$, and the corresponding row vector of $(\boldsymbol{\Gamma}^{\perp}_{s_f-1})^T\boldsymbol{H}_{s_f-1,u}$ can be seen as $\boldsymbol{\alpha}_s\boldsymbol{H}_{u,s}$. Hence, the problem of *parity space* identification is transferred to seeking $(\boldsymbol{\Gamma}^{\perp}_{s_f-1})^T[\boldsymbol{I}| - \boldsymbol{H}_{s_f-1,u}]$, which should equal to the orthogonal column space of $\boldsymbol{Z}_f/\boldsymbol{Z}_p$.

Perform SVD decomposition of $\boldsymbol{Z}_f/\boldsymbol{Z}_p$ as

$$\boldsymbol{Z}_f/\boldsymbol{Z}_p = \begin{bmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_{z,1} & 0 \\ 0 & \boldsymbol{\Sigma}_{z,2}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T\end{bmatrix} \quad (17)$$

where $\boldsymbol{\Sigma}_{z,2} \approx 0$. In theory, its rank should be $s_f l + n$ [9]. In practice, $\boldsymbol{Z}_f/\boldsymbol{Z}_p$ is not strictly singular and one has to determine its rank. The rank determination is equivalent to the determination of system order n. In [10], the optimal order of the model will be the one which makes the a criteria AIC (Akaike Information Criterion) minimum. Details can be found in this paper. And s_f and s_p is assigned $n+1$ empirically.

With (17), we can easily find the orthogonal column space of $\boldsymbol{Z}_f/\boldsymbol{Z}_p$ is $\boldsymbol{U}_2$, therefore

$$\left[(\boldsymbol{\Gamma}^{\perp}_{s_f-1})^T[\boldsymbol{I}| - \boldsymbol{H}_{s_f-1,u}]\right]^T = \boldsymbol{U}_2\boldsymbol{M} \quad (18)$$

where $\boldsymbol{M} \in \mathbb{R}^{(ms_f-n)\times(ms_f-n)}$ is any constant non-singular matrix.

$$\boldsymbol{U}_2\boldsymbol{M} = \begin{bmatrix}\boldsymbol{U}_{2,y} \\ \boldsymbol{U}_{2,u}\end{bmatrix} \quad (19)$$

From (18) and (19)

$$(\boldsymbol{\Gamma}^{\perp}_{s_f-1})^T = \boldsymbol{U}_{2,y}^T \quad (20)$$

$$-(\boldsymbol{\Gamma}^{\perp}_{s_f-1})^T(\boldsymbol{H}_{s_f-1,u}) = \boldsymbol{U}_{2,u}^T \quad (21)$$

Consequently, the row vectors of $\boldsymbol{U}_{2,y}^T$ and $\boldsymbol{U}_{2,u}^T$ can be considered as $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_s\boldsymbol{H}_{u,s}$ with $s = s_f - 1$. The *parity space* dimension is $ms_f - n$.

Remarks:

- As a whole SIM operation, the system matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ should be extracted from the estimated $\boldsymbol{\Gamma}^{\perp}_{s_f-1}$ and $\boldsymbol{H}_{s_f-1,u}$, although it is not necessary in our case.
- The data for *parity space* identification are collected from routine normal operation rather than a system identification plant test.

E. Fault isolation in residual space

From last section, it is known that we can construct a $ms_f - n$ dimension *parity space*. Thus, a residual space of identical dimension can be generated. It is reasonable and convincing that the magnitudes and signs of residuals differ for different faults. Consequently, fault isolation can be realized by carrying out a classification procedure in residual space, when the data sets of concerning faults are available in hand.

The fault isolation problem can be formulated as a classification problem: suppose that we have a training dataset of N samples of *residual vector* $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\ldots$, $\boldsymbol{r}_N$, which are distributed in C fault classes denoted by ω_1, ω_2, $\ldots$,

ω_C. The sample number in ω_i is N_i. A classifier can be trained based on the training dataset. Through the trained models, a real-time sample $\boldsymbol{x}$ can be classified to a defined class $\omega_i,\ i = 1, \ldots, C$, and thus the fault isolation can be achieved. In this study, a multi-class SVM is adopted to fulfill the fault isolation task.

The basic binary SVM is as Fig. 2 shows, 2-class data. Suppose we have some hyperplane which separates the data. Margin is defined as the sum of distances from the separating hyperplane to the closest points of the two classes. These "closest points" are called support vectors. SVM simply looks for the separating hyperplane with largest margin. The details about how to train and perform a binary SVM can be found in [11].

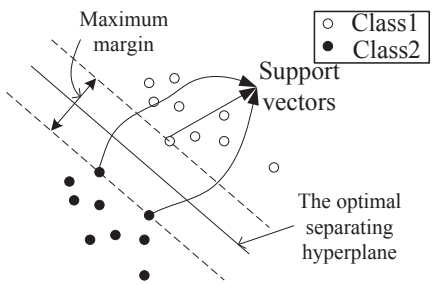

Fig. 2. SVM schematic diagram

Directed Acyclic Graph SVM (DAGSVM) is adopted to achieve multi-class classification problem [12]. In the training stage, all the possible binary classifiers from a training set of classes are constructed. As for performing stage, as Fig. 3 shows, given a test sample, starting at the root node, the binary classification function at a node is evaluated. Then it moves to the node in the next layer from either left or right path (depending on the binary classification result). Then, the binary classification function of the next node is evaluated. After passing $C - 1$ nodes, a leaf node which indicates the predicted class can be reached.

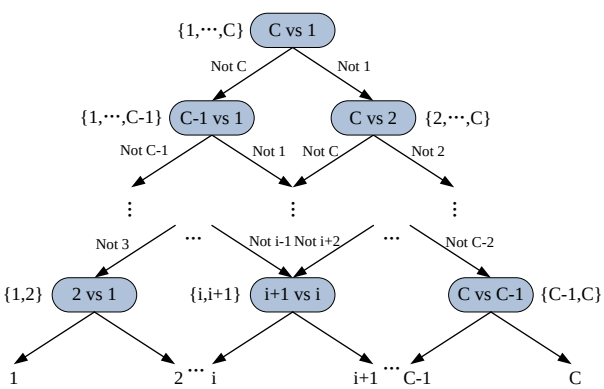

Fig. 3. DAGSVM structure

III. APPLICATION OF PEMFC DIAGNOSIS

A. PEMFC system description

A number of fuel cells are connected in series as a fuel cell stack. Three ancillary circuits, other than fuel cell stack, are usually added to the system: hydrogen circuit, air circuit, and temperature circuit. In hydrogen circuit and air circuit, the flow rates and pressures of gases can be regulated respectively. The temperature of fuel cell stack can be regulated by temperature circuit. In addition, some humidify instruments are usually added to the air circuit and hydrogen circuit to guarantee a proper humidity.

In our lab, a 5 kW PEMFC testbench is set to test a 40-cell PEMFC stack. Table I summarize the nominal operating conditions of the investigated stack. In this testbench, many physical parameters impacting stack performances can be controlled and measured. Stack temperature (measured by a thermocouple placed at the cooling circuit outlet), gas flow rates, fluid hygrometry rates, and load current can be set. Inlet and outlet flow rates of air and hydrogen, inlet and outlet pressures of air and hydrogen, stack temperatures, current, stack voltage and single cell voltages can be monitored. Additional details about the test bench and the test protocol have been previously published in [13].

TABLE I
NOMINAL CONDITIONS OF 40-CELL PEMFC STACK

Physical variables	Values
Stoichiometry H_2	1.5
Stoichiometry Air	2
Pressure of H_2 in entrance	150 kPa
Pressure of Air in entrance	150 kPa
Temperature	80 °C
Anode relative humidity	50%
Cathode relative humidity	50%
Current	110 A
Nominal stack voltage	28 V
Nominal power	3080 W

B. Experiments and data acquisition

The experiments of normal state and various fault states were carried out in our test bench. The input and output data were measured and saved with the sample frequency of 1 Hz. The normal state includes the processes of normal operating (load current of nominal value and the values in safe range), system starting and stopping. The fault-free data is used for identification of the *parity space*. Four representative faults were induced intentionally. Table II summarizes the four faults. Among them, in the F_1 and F_4 experiments, the faults were emulated in between the normal operation. Because these two faults can cause irreversible damage to the fuel cell stack, the normal state was recovered after a short faulty time during the experiment; while in the F_2 and F_3 experiments, stoichiometry value was set respectively at $200\% \sim 250\%$ and 70% of the nominal value, and the fault states could be maintained in the stable states.

Each faulty experiment was repeated several times. The data from fault-free state were used to identify the *parity vectors*. The data from faulty experiments were firstly used to verify the *parity space* based fault detection, and then used for training and testing of DAGSVM based fault isolation.

In order to carry out the approach, 8 variables, including pressures of input air and hydrogen ($P_{in,air}$, P_{in,H_2}), flow

TABLE II

TABLE OF VARIOUS CONCERNED DEFAULTS OF 40-CELL STACK

Fault location	Description	Abbreviation
Electric circuit	Short circuit	F_1
Air circuit	High air stoichiometry	F_2
	Low air stoichiometry	F_3
Cooling circuit	Stop cooling water	F_4

rates of input air and hydrogen ($D_{in,air}$, D_{in,H_2}), relative humidity of input air and hydrogen ($RH_{in,air}$, RH_{in,H_2}), current (I), and stack temperature (T_{fc}), are selected as input variables. 2 output variables are stack voltage (V_{st}), and pressure drop of input air and output air (ΔP_{air}). Hence, $l = 8$ and $m = 2$.

IV. RESULTS AND DISCUSSION

A. Parity space identification

The AIC values are shown in Fig. 4 for cases of system order from 1 to 15. So the system order n is considered to be 2 in the study. s_f and s_p were assigned 3 accordingly. The dimension *parity space* is $ms_f - n = 4$. By using aforementioned *parity space* identification procedure, 4 *parity vectors* were obtained. The residuals (denoted by r_1, r_2, r_3, r_4 in the figures) corresponding to these *parity vectors* were calculated using (5). The residuals and the corresponding R in normal state are shown in Fig. 5. It can be seen that the residuals in normal state are mostly near to zero. Since data from two normal experiments were jointed together to identify the *parity space*, a strike can be observed in the joining together of the two data sets (at about 5000 s).

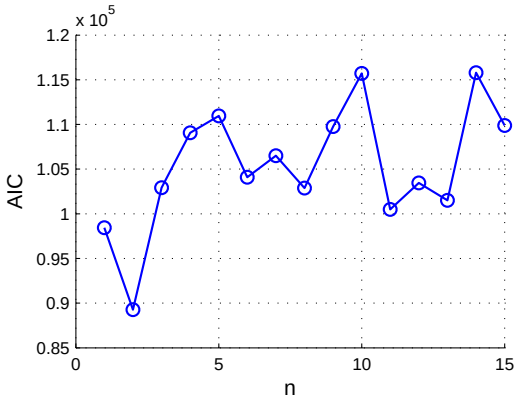

Fig. 4. Values of AIC for different system order

B. Fault detection

In the fault detection procedure, (6) was used to decide whether the faults happened. Here the threshold Th was set to 1 to insure 99.9% of normal data is justified in to the normal state. The residuals and R in faulty experiments are shown in Fig. 6 to 9. In Fig. 6 and 9, the faults F_1 and F_4 occurred in a session of the experiments. For faults F_2 and F_3, the faults could be maintained in the stable states. In Fig. 7, the F_2 was emerged from the beginning and aggravated after some time (about 1300s). In Fig. 8, the F_3 was brought after a period of normal operating.

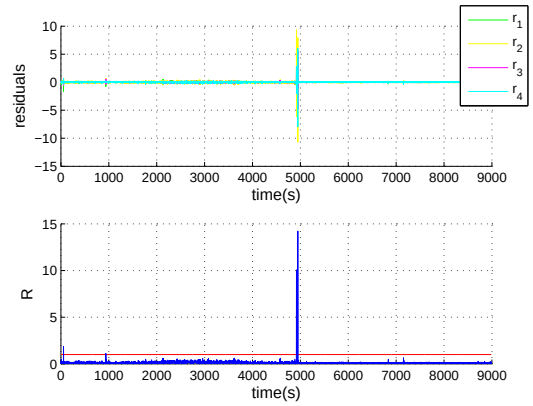

Fig. 5. Residuals and R in normal state

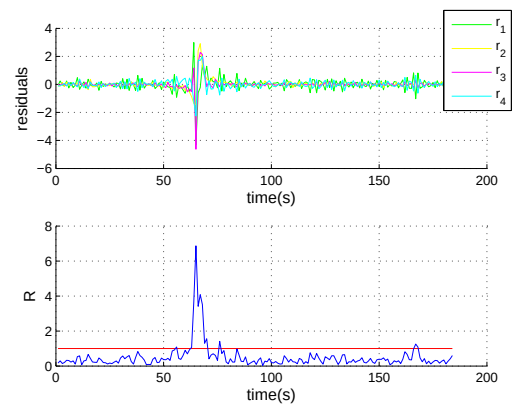

Fig. 6. Residuals and R in the experiment of F_1 fault

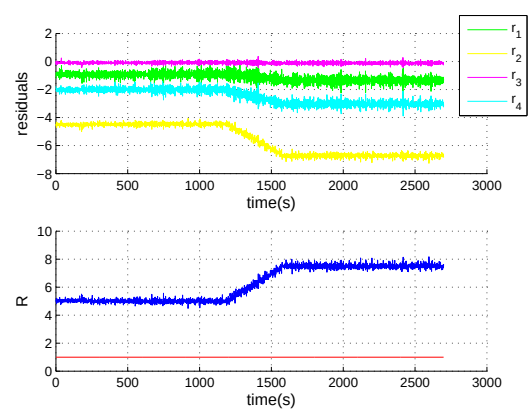

Fig. 7. Residuals and R in the experiment of F_2 fault

C. Fault isolation

After faults are detected, fault isolation should be taken into account. The samples, whose R values are greater than

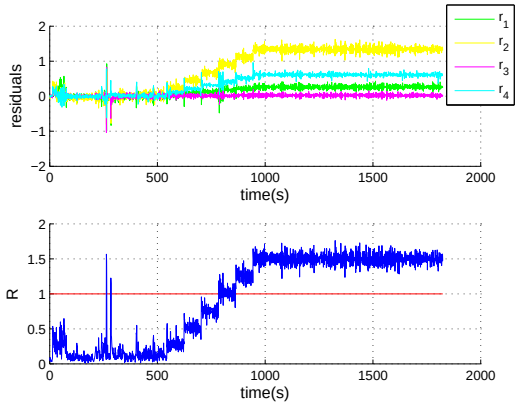

Fig. 8. Residuals and R in the experiment of F_3 fault

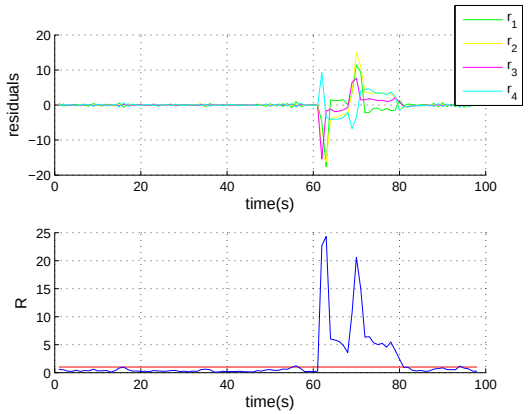

Fig. 9. Residuals and R in the experiment of F_4 fault

threshold, were selected from various faulty data sets. Since the faults were deliberately caused, the faulty data can be labeled into four fault classes. The labeled samples were used to train and test the DAGSVM. For each fault case, the data from one corresponding experiment were used for training, while that from other experiments were used for testing. On the basis of our results, the fault isolation accuracy is 100% for both training data and testing data, which verify the good fault isolation capability of DAGSVM classification.

V. CONCLUSION AND PERSPECTIVE

A data-driven strategy is proposed to address PEMFC diagnosis problems. With the aid of an orthogonal projection SIM approach, *parity space* can be identified from normal process data. With the *parity vectors*, fault detection can be realized. DAGSVM is used for the following fault isolation procedure. The diagnostic results of a 40-cell stack show that the 4 concerned faults can be detected and isolated with a high accuracy.

In this study, the *parity space* is generated from a LTI (linear time invariant) state space model. how to extend the approach in nonlinear case, for instance, the utilization for linear parameter varying (LPV) model is still in study.

ACKNOWLEDGMENT

This Ph.D. work is a contribution to the ANR DIAPASON2 project (fuel cell diagnosis methods for vehicle and stationary applications 2nd phase).

REFERENCES

[1] T. Escobet, D. Feroldi, S. D. Lira, V. Puig, J. Quevedo, J. Riera, and M. Serra, "Model-based fault diagnosis in PEM fuel cell systems," *Journal of Power Sources*, vol. 192, pp. 216–223, 2009.

[2] D. Hissel, S. Member, D. Candusso, and F. Harel, "Fuzzy-Clustering Durability Diagnosis of Polymer Electrolyte Fuel Cells Dedicated to Transportation Applications," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2414–2420, 2007.

[3] N. Yousfi Steiner, D. Hissel, P. Moçotéguy, and D. Candusso, "Diagnosis of polymer electrolyte fuel cells failure modes (flooding & drying out) by neural networks modeling," *International Journal of Hydrogen Energy*, vol. 36, no. 4, pp. 3067–3075, Feb. 2011.

[4] N. Yousfi Steiner, D. Hissel, P. Moçotéguy, and D. Candusso, "Non intrusive diagnosis of polymer electrolyte fuel cells by wavelet packet transform," *International Journal of Hydrogen Energy*, vol. 36, no. 1, pp. 740–746, Jan. 2011.

[5] Z. Li, R. Outbib, D. Hissel, and S. Giurgea, "Online Diagnosis of PEMFC by Analyzing Individual Cell Voltages," in *2013 European Control Conference (ECC)*, Zurich, Switzerland, July 17-19 2013.

[6] J. Hua, J. Li, M. Ouyang, L. Lu, and L. Xu, "Proton exchange membrane fuel cell system diagnosis based on the multivariate statistical method," *International Journal of Hydrogen Energy*, pp. 1–10, 2011.

[7] S. X. Ding, "Data-driven design of model-based fault diagnosis systems," in *IFAC ADCHEM*, Singapore, July 10-13 2012.

[8] S. X. Ding, P. Zhang, A. Naik, E. Ding, and B. Huang, "Subspace method aided data-driven design of fault detection and isolation systems," *Journal of Process Control*, vol. 19, no. 9, pp. 1496–1510, Oct. 2009.

[9] B. Huang, S. X. Ding, and S. Qin, "Closed-loop subspace identification: an orthogonal projection approach," *Journal of Process Control*, vol. 15, no. 1, pp. 53–66, Feb. 2005.

[10] J. Wang and S. J. Qin, "Closed-loop subspace identification using the parity space," *Automatica*, vol. 42, no. 2, pp. 315 – 320, 2006.

[11] J. C. Platt, "Sequential Minimal Optimization : A Fast Algorithm for Training Support Vector Machines," *Technical Report MSR-TR-98-14, Microsoft Research*, pp. 1–21, 1998.

[12] J. C. Platt, N. Cristianini, and J. Shawe-taylor, "Large margin dags for multiclass classification," in *Advances in Neural Information Processing Systems*. MIT Press, 2000, pp. 547–553.

[13] D. Candusso, A. De Bernardinis, M.-C. Péra, F. Harel, X. François, D. Hissel, G. Coquery, and J.-M. Kauffmann, "Fuel cell operation under degraded working modes and study of diode by-pass circuit dedicated to multi-stack association," *Energy Conversion and Management*, vol. 49, no. 4, pp. 880–895, Apr. 2008.