Real-Time Diagnostics in the EGR System of Diesel Engines

Javad Mohammadpour^{†*}, Karolos M. Grigoriadis[†], Matthew A. Franchek[†], and Benjamin J. Zwissler[‡]

Abstract—The paper presents a real-time approach for parameter identification used for diagnosing different types of faults in the exhaust gas recirculation (EGR) system of Diesel engines. An important benefit of the proposed diagnostics method is its ability to detect and estimate a leak or a restriction in the EGR system. This is achieved by making use of a recursive-least-squares (RLS) method, as well as, a recursive formulation of a robust version of the RLS method, which we refer to as recursive total-least-squares (RTLS) method. The proposed approach of fault detection is successfully applied to diagnose low flow or high flow faults in an engine and is validated using experimental data.

I. INTRODUCTION

Modern engine on-board diagnostics (OBD) systems are mainly based on simple limits or rationality checks of some measured signals and on simple signal-based methods such as the frequency analysis of the engine speed signal [7]. In the future, these methods will most probably not be able to meet strict OBD requirements. Therefore, model-based fault detection methods are promising ways for improving the fault detection task. Analytical process information in the form of mathematical models can be used to evaluate the information from different sensors, whereby dependencies between the different signals can be used. Therefore, source of the faults can be better identified, and an isolation and localization of the faults along with determining the severity of the fault can be achieved.

The present paper focuses on the detection and estimation of the severity of the low flow and high flow faults in the EGR system of Diesel engines. Our fault identification approach is a model-based method implemented in real-time. Analysis of the residual signals generated using the proposed identification scheme provides the detailed information related to the fault type, *i.e.*, low flow or high flow, as well as, the severity of the fault. The proposed method is shown to be capable of discriminating between two different levels of the fault magnitude. Our real-time parameter identification method will be based on the RLS algorithm, as well as, a modified version of the RTLS algorithm developed in this paper.

Recently, there have been several efforts to develop realtime model-based methods for diagnosis of faults occurring in Diesel engines [2], [14], [1], [8]. The authors in [2] develop a steady-state intake airpath diagnostics method. In particular, the work in [2] focuses on the development and adaptation of the steady-state airpath models that track the mass air flow (MAF) sensor output by utilizing a model regressor identification technique called system probing. The authors in [14] consider different fault types including airmass flow sensor fault, intake-manifold pressure sensor fault, air-leakage between the air-mass flow sensor and the cylinders, and the EGR valve stuck in a closed position. The diagnosis system design in [14] follows the framework of structured hypothesis tests. An effective method for improving model-based diagnosis for the airpath of a truck engine is presented in [1]. The latter work builds statistical charts from truck operational data, where quite accurate static models of both the volumetric efficiency and sensors are developed. The diagnostics method of [1] is tested in order to reduce the overall residual scattering. The authors in [8] develop a model-based approach for the fault detection and diagnosis of combustion engines. They divide the engine into three modules: the intake system, the injection system along with the combustion, and the exhaust system. For the intake and injection systems, residuals are generated which are zero under normal operating conditions. In order for generating the residual signals, semi-physical models, identification with local linear neural networks, signal models and filter methods are used.

Many real-world applications require a model of the system to be available in real-time while the system is in operation. Estimating an online model for batches of inputoutput data might be used to address the question of whether or not a failure has occurred, and if so, what type of failure [9], [2]. The online models might also be used to investigate the time variations in system and signal properties. The methods for computing online models are called recursive identification methods. There are some recursive estimation algorithms that are widely used for conducting the parameter estimation to adapt an online model. These include Kalman filter algorithm, forgetting factor algorithm, and unnormalized and normalized gradient algorithms. In the linear regression case, the forgetting factor algorithm is also known as the recursive-least-squares (RLS) algorithm [9]. For the real-time parameter identification of the proposed diagnostics method of this paper, we use RLS and a recently developed modified recursive method [13], which we call RTLS, to reduce the effects of the parametric uncertainty and sensor noise. RTLS is a recursive formulation of total-least-squares (TLS) method. TLS was initially proposed to provide more accurate parameter identification when error exists not only in the sensor measurements but also the data matrices [4], in

This work was supported by On-Board Diagnostics group at Cummins Inc., Columbus, IN

[†]Department of Mechanical Engineering, University of Houston, Houston, TX 77204

[‡]Common Diagnostics/OBD, Cummins Inc., Columbus, IN 47201

^{*}Corresponding author. Tel: 713-743-4811, Fax: 713-743-4503; Email: JMohammadpour@uh.edu



Fig. 1. Simplified schematic view of the airpath in a Diesel engine [12]

which cases the best solution in the *least-squares* sense is not often as good as the best solution in the *eigenvector* sense. In the present paper, the results of the application of the RTLS algorithm are shown to provide significant improvement in the parameter identification step in our diagnostics algorithm.

The rest of the paper is organized as follows. Section II presents the model-based formulation used for the proposed diagnosis technique in the present paper. Section III presents our parameter identification-based approach for low flow and high flow diagnosis of the EGR system. Section IV includes the results that validate our diagnostics algorithm using the data collected from both a Diesel engine test cell and a test truck. Section V concludes the paper.

II. PRELIMINARIES AND PROBLEM STATEMENT

A schematic view of the engine airpath is shown in Fig. 1. As shown in Fig. 1, the air entering the engine is measured by an air mass flow sensor and the corresponding measurement is represented by W_{in} . Then, the air goes through a compressor and a charge air cooler, enters the intake manifold, where it is mixed with the exhaust gases, and flows into the cylinders where the fuel is added. In the exhaust manifold, gas is divided into two parts: one flows through the turbocharger to drive it, and another part flows back to the intake manifold through the EGR passage. The measurements used to generate the residual signals are the EGR mass flow W_{egr} , the boost pressure P_i , the exhaust pressure P_o , and the exhaust manifold temperature T_o .

We use a mathematical model to represent the mass flow past the EGR valve, which is well described by the standard orifice flow equation, representing flow through a restriction [3]

$$W_{egr} = \begin{cases} C_{egr}(\xi_{egr}) \frac{p_o}{\sqrt{RT_o}} \phi(\frac{p_i}{p_o}) & \text{if } p_i < p_o \\\\ 0 & \text{if } p_i = p_o \\\\ C_{egr}(\xi_{egr}) \frac{p_i}{\sqrt{RT_i}} \phi(\frac{p_o}{p_i}) & \text{if } p_i > p_o \end{cases}$$
(1)

where R, p_i , p_o , T_o are the gas constant, downstream pressure, upstream pressure, and upstream temperature, respectively. C_{egr} is the effective flow area of the EGR valve expressed as a function of the normalized EGR valve opening $\xi_{egr} \in [0, 1]$. The pressure ratio correction factor ϕ is calculated as

$$\phi(a) = \begin{cases} \gamma^{1/2} \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} & \text{if } a \le r_c \\ \\ \sqrt{\frac{2\gamma}{\gamma-1} \left(a^{\frac{2}{\gamma}} - a^{1+\frac{1}{\gamma}}\right)} & \text{if } a > r_c \end{cases}$$
(2)

where γ is the ratio of specific heats, and $r_c = (\frac{2}{\gamma+1})^{\frac{\gamma}{\gamma-1}}$ is known as the critical pressure ratio. The EGR valve effective area $C_{egr}(\xi_{egr})$ is developed from the steady-state engine mapping data; however, in the current work, it is represented as a function of the EGR valve opening (percentage) as

$$C_{egr}(\xi_{egr}) = b_1 \xi_{egr} + b_2 \xi_{egr}^2$$

where b_1 and b_2 are scalar coefficients. To incorporate the sensor measurement data in the used model, a parametric model including two parameters a_1 and a_2 is introduced as

$$\Psi(p_i, p_o, T_o, W_{egr}) = a_1 \xi_{egr} + a_2 \xi_{egr}^2 \tag{3}$$

The parameters a_1 and a_2 are to be identified using a realtime parameter identification technique. We will be using changes in the values of the parameters a_1 and a_2 as the indication of the occurrence of a fault, either a leak or a restriction. The percentage of change of these coefficients will be shown to be a representative of the severity of the fault in the EGR system causing a variation of the amount of gas flowing back to the intake system.

III. PROPOSED APPROACH FOR DETECTION AND ESTIMATION OF THE FAULTS

The success of a model-based diagnostic technique is highly dependent on the accuracy of the model structure developed to represent the dynamics of the system. For the Diesel engine system under study, suitable models and mathematical relationships must be used, which extract knowledge about the healthy or faulty system state with the restricted information of only a few sensors. Noting that we are concerned with the types of faults that take place inside the EGR valve, we claim that employing the appropriate static physics-based equations that represent the flow through the EGR passage can address the identification and estimation of the faults not only in steady-state but also transient cycle.

This section presents the proposed fault detection and estimation methodology for the EGR system of Diesel engines. The algorithm is implemented by adapting the coefficients of the equation given in (3) that represents the static behavior of the EGR system in terms of the amount of gas flowing back to the intake manifold. A coefficient error vector is generated by comparing the healthy model parameters and the adapted model parameters. The value (and sign) of the coefficient error vector will be used to indicate whether or not a fault exists and the type of fault occurred. Fault estimation is also performed by analyzing the residual between the healthy model output and the actual sensor output.

There are always issues that may contribute to the possibility of false fault detection, such as model inaccuracy, sensor noise, as well as, the sensor inaccuracy. The approach taken in this work is shown to be less sensitive to model uncertainty and sensor noise, since the RTLS method is employed for the parameter identification purpose instead of the RLS. Note that our parameter identification approach, which is based on a static relationship, is robust against the noise and uncertainties even though the algorithm is implemented over the transient cycle. This is in contrast to the approach taken in e.g. [2], which only considers the fault detection and isolation in steady-state. It is noted that sensor noise could be alternatively attenuated by averaging and filtering the sensor output. To this end, suitable filters may be determined [5], [15] using a similar approach used to derive recursive formulation of the total-least-squares. We note that the normal variability of the system appearing during the work cycle is incorporated into the proposed process by determining a threshold where the magnitude of the coefficient error vector, that should be greater than a precalculated tolerance before occurrence of a fault, is declared.

A. Fault Detection Scheme

Let the healthy system steady-state model coefficients in (3) construct a vector H defined as

$$H = \begin{bmatrix} a_1 & a_2 \end{bmatrix}^T \tag{4}$$

Next, we define the steady-state model coefficient vector F for the adapted model, containing a fault, as

$$F = \begin{bmatrix} a_1^f & a_2^f \end{bmatrix}^T \tag{5}$$

Using (4) and (5), we define the model coefficient error vector E as

$$E = S(H - F) \tag{6}$$

where

$$S = diag(\frac{1}{a_1}, \frac{1}{a_2}) \tag{7}$$

Defining E as in (6) represents a normalized change in the model coefficients. It is, therefore, possible to effectively address models having coefficients that differ significantly in magnitude [2].

Detecting the existence of a fault is accomplished by evaluating the magnitude of E and making sure that this error is not due to the standard variability of the system. The presence of a system fault can then be performed by calculating the residual signal as

$$M(t) = \begin{cases} y(t) - y^{H}(t) & \text{if } ||E||_{2} > \epsilon \\ 0 & \text{otherwise} \end{cases}$$
(8)

where y(t) is the sensor output at the time instant t, $y^{H}(t)$ represents the healthy EGR system model having the structure defined in (4), ϵ is a threshold value that represents normal variability of the system, and $||E||_{2}^{2} =$

 $E^T E$. Defining M as in (8) suggests that a nonzero value for M indicates the presence of a fault.

The degree of variability of a healthy system, represented by ϵ in (8), can be quantified when determining the values a_1 and a_2 in (4). To identify these coefficients for the healthy model, a series of experiments is performed. For the multiple experiments, the mean value of each identified parameter is determined as a healthy model parameter in (4). Due to the availability of the multiple values for each parameter (a_1 and a_2), a standard deviation calculation can be performed, where it can quantify the normal variability of the healthy system. Hence, the parameter ϵ in (8) is determined as

$$\epsilon = \|[\sigma_{a_1}, \sigma_{a_2}]^T\|_2 \tag{9}$$

where σ denotes the standard deviation.

B. Fault Estimation Scheme

Fault estimation is performed by establishing a representative for the fault size. Let us define the following metric which is a percent error in the EGR mass flow predicted by the healthy model

$$\Delta(t) = \frac{M(t)}{y^H(t)} \tag{10}$$

where M(t) is defined in (8). The metric Δ represents the severity of the fault, e.g., how much gas flow is increased due to the leak in the EGR system, or how much gas flow is reduced due to the presence of a restriction in the EGR system.

IV. EXPERIMENTAL VALIDATION

The engine under study in this work is a Cummins Diesel engine equipped with exhaust gas recirculation (EGR) system to reduce the emissions. The data are collected either by running an FTP-75 test cycle in a test cell or from a test truck driving on different routes.

To create the different types of faults, we concentrate on high gas flow (leak) and low gas flow (restriction) faults. A leak is applied by drilling a hole of certain size in the poppet valve. A restriction is applied by placing a tube of a particular size inside the EGR valve to restrict the amount of the gas flowing back to the intake manifold through the EGR passage. It is also worth mentioning that the full closed-loop control, for coordinated control of the EGR valve opening, VGT nozzle position, and intake throttle valve opening, is in effect during all the FTP test cycles we run.

The proposed fault detection and estimation method is applied to the EGR system of the internal combustion engine. We are concerned with detecting internal leak, as opposed to the external leak which is easy to detect [12], as well as, the restriction inside the EGR valve. Using the model described in (3), we determine a pair of coefficients (a_1, a_2) that best represents the data collected from the healthy system. The faults of different magnitudes have been created on different engines, so different pairs of the healthy parameters will be determined as described in the following subsections. As explained earlier in the paper, we employed RLS and RTLS



Fig. 2. The EGR mass flow residual signal generated using the measurement and output of the adapted model

methods for the parameter identification purpose. The details of the RTLS method can be found in [15], [13], [10]. The RTLS algorithm is implemented using a modified version of the early recursive formulation presented in [13]. The modified formulation of the RTLS used in the present work can be found in [10].

A. Detection and Estimation of a Leak in the EGR System

1) Validation of the Method on Engine Test Cell Data: We run the FTP-75 cycle in a test cell to collect data in the healthy condition and in the presence of the leak. The coefficients associated with the model of the system (in both the healthy and the leaky conditions) are identified by employing the RLS algorithm using a forgetting exponential factor $\lambda = 0.999$. Table I lists the coefficients, as well as, the error vector determined using (6). The data listed in the table are the results of several runs and subsequently taking an average of the identified coefficients. Fig. (2) depicts the difference between the output of the adapted model, *i.e.* the output of the RLS algorithm, and the sensor measurements used for the model adaptation purpose. Careful investigation of the numbers listed in Table I indicates that E < 0 implies that there is a leak in the EGR system, and that $||E||_2$ will increase if the size of the leak increases as shown in Table I for leak diameters of 4.3mm and 4.5mm. This is consistent with the physical interpretation of the coefficients in (3).

Once the EGR system is identified to be leaky, it still remains to address the question of *how severe the fault is*. To deal with this question, we need to evaluate the residual signal generated from the difference between the EGR mass flow measurement and the amount of gas that should be flowing through the EGR path if the EGR system was in the healthy condition. Fig. 3 illustrates the residual signals associated with two sizes of the leak. The lower plot is the magnification of the upper one, in a certain time interval, to observe the difference between the residual signal associated with the 4.5mm leak and that of the 4.3mm leak. The figures indicate that the residual generated from the 4.5mm leak is visibly higher than that generated from the 4.3mm leak. Using the generated residual signals, one may determine the



Fig. 3. The EGR mass flow residual signal calculated using actual measurement and output of the adapted healthy system

TABLE I MODEL COEFFICIENTS (a_1, a_2) FOR THE EGR SYSTEM OPERATION IN BOTH THE HEALTHY AND LEAKY CONDITIONS

	Coefficients	Error vector E	$ E _2$
Healthy	(11.37, -7.86)	×	Х
4.3mm Leak	(13.51, -10.32)	(-0.188, -0.313)	0.365
4.5mm Leak	(14.08, -11.05)	(-0.238, -0.405)	0.469

size of the leak (in percent of opening). The interested reader is referred to [10] for details about the procedure.

B. Detection and Estimation of the Restriction in the EGR System

1) Validation of the Method on Engine Test Cell Data: To validate the proposed algorithm of detection and estimation of the restriction, a different engine, including the EGR system, than the one we used for the diagnostics of the leak is employed to collect the data in the healthy condition and in the presence of the restriction inside the EGR valve. The coefficients associated with the model of the EGR system (in both healthy and restricted conditions) are identified using the RTLS method (see [10] for the corresponding descriptions) with a forgetting exponential factor $\lambda = 0.999$. Note that the RLS algorithm was initially used for the parameter identification; however, it did not provide convergence to identify the coefficients a_1 and a_2 due to the

TABLE II





Fig. 4. Results of the parameter identification using the RLS and RTLS algorithms: (a) baseline and (b) restricted condition

presence of parameter uncertainties and sensor noise in the experimental data. To show the superior performance of the RTLS compared to the RLS, Fig. 4 illustrates the profiles of the coefficients a_1 and a_2 for baseline, *i.e.* healthy condition, in the upper plot and the restricted condition created by placing a tube inside the EGR valve in the lower plot. This figure clearly indicates that the RLS algorithm is not able to provide convergence especially for the baseline data. On the other hand, RTLS provides convergence on the baseline data, and similar results hold for the faulty data.

Table II lists the coefficients and the error vector calculated using (6). The numbers presented in this table are the results of several runs with the average of the identified coefficients. Carefully investigating the numbers listed in Table II indicates that E > 0 implies that there is a restriction in the EGR system that causes the gas to flow less than expected, and that the larger the 2-norm of the vector E is, the bigger the size of the restriction inside the EGR valve is.

Next, we use an identified pair of coefficients in the baseline case, using the results of the RTLS algorithm, to measure the right amount of gas that should have flowed through the EGR path. Fig. 5 illustrates the actual measured EGR mass flow in the system including a restriction plate and the EGR mass flow that would have flowed through the EGR path if there was no restriction placed in the valve.

2) Validation of the Method on Data Collected from a *Test Truck:* In this section, we provide the results of using the proposed diagnostics method for detecting the low flow gas in the EGR passage using data collected from a heavy-



Fig. 5. The EGR mass flow in the low flow faulty condition and the amount of gas that should have flowed through the EGR passage

duty Diesel engine in driving conditions. The data are collected from a test truck during a trip that represents a city/highway driving cycle. We show the results of applying our diagnostics algorithm to this set of data that includes the different EGR restriction diameter sizes of 0.8", 0.62", 0.5", 0.36", 0.3", and 0.25", as well as, the baseline system. The diameter sizes were chosen on the basis of descending amounts of relative unrestricted areas of 50%, 30%, 20%, 10%, 7%, and 5%, respectively.

To implement the diagnostics algorithm as described in the previous section, the RTLS algorithm is employed to identify the coefficients a_1 and a_2 in (3). Fig. 6 shows the coefficients a_1 and a_2 versus the percentage of the unrestricted area. The plot illustrates that using the developed methodology the difference between small change in sizes of the restrictions is distinguishable, particularly for higher restrictions. It should be noted that for the identification purpose, the RTLS algorithm does a satisfactory job, and that the RTLS algorithm has been applied only to the part of the data, where the EGR valve is not fully closed.

Fig. 6 verifies the fact that absolute values of both coefficients a_1 and a_2 decrease if the size of the restriction inside the EGR valve increases. The plots may be used to generate a 2-dimensional look-up table that represents the relation between the coefficients a_1 and a_2 identified in real-time and the size of the restriction in the EGR valve. This will provide a quantitative assessment of the size of the restriction that leads to the low flow in the EGR passage.



Fig. 6. Coefficients a_1 and a_2 identified using the RTLS algorithm vs. the different unrestricted areas



Fig. 7. Profiles of the coefficients a_1 and a_2 vs. the sample number for the baseline system

To illustrate the convergence of the RTLS algorithm employed for the parameter identification, we have shown in Fig. 7 and 8 the profiles of the identified coefficients a_1 and a_2 over the course of time for the baseline and the maximum restriction in the EGR system.

V. CONCLUSION

Presented in this paper is a real-time fault detection methodology for the EGR system of Diesel engines. The proposed approach is based on the identification of two parameters in a static relationship obtained from the standard



Fig. 8. Profiles of the coefficients a_1 and a_2 vs. the sample number for 95% restriction

orifice flow equation. For parameter identification purposes, we use the RLS algorithm, as well as, a modified version of the RTLS method which we develop in this paper. Occurrence of the fault, either a leak or a restriction, provides a trend observed in the coefficients of the determined static relationship. An advantage of the proposed diagnostics method is its capability of estimating the magnitude of a fault, that is, the size of a leak or a restriction in the EGR system. The proposed method has been successfully implemented and validated to diagnose low flow or high flow faults in Diesel engines using experimental data.

VI. ACKNOWLEDGEMENT

The first author would like to thank Michael Hagigeorgiou and Taire Avbovbo at the OBD team in Cummins Inc., for providing the data, and Gary Salemme at Cummins Inc., for the useful discussions at the early stage of the project. The financial support under a Cummins Inc., grant is gratefully appreciated.

REFERENCES

- A. Berton, E. Frisk, and M. Nyberg, "Improving Diagnosis Performances on a Truck Engine Making Use of Statistical Charts," in Proc. IFAC Symposium on Advances in Automotive Control, 2004, Salerno, Italy.
- [2] M.A. Franchek, P.J. Buehler, and I. Makki, "Intake Air Path Diagnostics for Internal Combustion Engines," *Journal of Dynamic Systems, Measurement, and Control*, Vol. 129, Issue 2, pp. 32-40, Jan. 2007.
- [3] J.B. Heywood, Internal Combustion Engine Fundamentals, Mc Graw-Hill, New York, 1988.
- [4] S.V. Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, ser. Frontiers in applied mathematics, Philadelphia: Society of Industrial and Applied Mathematics, Vol. 9, 1991.
- [5] B. Hassibi, A.H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation* and Control: A Unified Approach to \mathcal{H}^2 and \mathcal{H}^{∞} Theories, SIAM Studies in Applied Mathematics, New York, 1999.
- [6] P. King, J. Vinsonneau, D. Shields, and K. Burnham, "Improved SI Engine Modeling Techniques with Application to Fault Detection," *in Proc. IEEE International Conference on Control Applications*, pp. 719-724, Glasgow, Scotland, 2002.
- [7] F. Kimmich and R. Isermann, "Model-Based Engine Torque Supervision for Diesel Engines Using the Engine Speed Signal," *in Proc. American Control Conference*, pp. 2002-2007, Anchorage, AK, 2002.
 [8] F. Kimmich, A. Schwarte, and R. Isermann, "Fault Detection for Mod-
- [8] F. Kimmich, A. Schwarte, and R. Isermann, "Fault Detection for Modern Diesel Engines Using Signal- and Process Model-Based Methods," *in Control Engineering Practice*, 13, pp. 189-203, 2005.
- [9] L. Ljung, System Identification Theory For the User, 2nd ed, PTR Prentice Hall, Upper Saddle River, N.J., 1999.
- [10] J. Mohammadpour, Filtering and Fault Tolerant Control of Parameter-Varying Time-Delay Systems and Applications, PhD dissertation, University of Houston, TX, Dec. 2007.
- [11] M. Nyberg and L. Nielsen, "Model-Based Diagnosis for the Air Intake System of the SI Engine," SAE paper 970209, 1997.
- [12] M. Nyberg and A. Perkovic, "Model-Based Diagnosis for the Air Intake System of the SI Engine," SAE paper 980514, 1998.
- [13] C. Phum and T. Ogunfunmi, "On the Recursive Total-Least-Squares," in Proc. ICASSP, pp. 1989-1992, Munich, Germany, 1997.
- [14] M. Nyberg, T. Stutte, and V. Wilhelmi, "Model-Based Diagnosis of the Air Path of an Automotive Diesel Engine," *in Proc. IFAC Workshop on Advances in Automotive Control*, Karlsruhe, Germany, pp. 629-634, 2001.
- [15] A.H. Sayed, B. Hassibi, and T. Kailath, "Inertia Conditions for the Minimization of Quadratic Forms in Indefinite Metric Spaces," in *Operator Theory: Advances and Applications*, I. Gohberg, P. Lancaster, and P.N. Shivakumar, Eds., pp. 309-347, Vol. 87, Birkhauser, 1996.