Damage Detection and Localization in Pipeline Using Sparse Estimation of Ultrasonic Guided Waves Signals

Mahjoub El Mountassir*,, Gilles Mourot**, Slah Yaacoubi*, Didier Maquin****

** Institut de soudure, Plateforme RDI CND, 4 Bvd Henri Becquerel 57970 Yutz France (m.elmountassir, s.yaacoubi)@isgroupe.com*
*** Centre de Recherche en Automatique de Nancy, CNRS, Université de Lorraine, 2 Avenue de la Forêt de Haye, TSA 60604, 54 518 Vandoeuvre-lès-Nancy, France (gilles.mourot, didier.maquin)@univ-lorraine.fr*

Abstract: This paper focuses on the development of a method for damage detection and localization in pipeline structures. These structures are subject to variation of environmental and operational conditions (EOCs) which have an impact on the collected signals. Since damage detection is generally based on comparison between the reference signals and the current signals acquired from the structure, the effects of EOCs will give rise to false alarm. This issue is addressed by selecting from the database of reference signals those with similar or very close EOCs. Such an operation can be performed by calculating a sparse estimation of the current signal. The estimation error is used as an indication of the presence of damage. Actually, a damage signal will be characterized by a high estimation error compared to that of a healthy signal. The position of the detected damage is obtained by calculating the estimation error on a sliding window over the damaged signal. This method was tested on signals collected on a small scale pipeline placed in laboratory conditions. Results have shown that the created damage was successfully detected and localized.

Keywords: Sparse estimation, damage detection and localization, structural health monitoring, environmental and operational conditions.

1. INTRODUCTION

Pipelines are very critical structures, especially those used for transporting oil, gas and other chemical substances. To ensure better working conditions of these structures, they must be continuously monitored. Structural health monitoring (SHM) systems were proposed to tackle this issue (Farrar and Worden, 2007). They aim at detecting, localizing and estimating the degree of severity of damage in the structure. Damage detection is performed by comparing the reference signals obtained from the healthy state and the current signal. For pipeline structure, the signals are generally acquired using Ultrasonic Guided Waves (UGW) technique (Lowe et al., 1998). UGW are stress waves which propagate through a medium and are guided by the structure boundary. These waves can travel in all directions ensuring hence a volumetric coverage and they interact with the structure discontinuities (damage, welds, structure ends, etc.).

The task of comparison between the reference signal and the current signal is not easy to achieve because the healthy state of the structure could vary due to the changes in the environmental and operational conditions (EOCs) (e.g. temperature, humidity, vibration loads, etc.) (Sohn, 2007). The effects of these EOCs could be similar to those produced by damage. This would result in false warnings. The differentiation between the aforementioned types of changes is a challenging task. Some methods were proposed to compensate the effects of these EOCs namely BSS (Baseline Signal Stretch) and OBS (Optimal Baseline Selection) (Croxford et al. 2010). However, damage detection with these methods is based on a simple subtraction between the reference signal and current signal which is not reliable. Data-driven methods were also proposed to tackle this issue. For example, Eybpoosh et al. (2017) have proposed a supervised method based on sparse representation of UGW signals which can discriminate between damage and variation in EOCs. However, this method requires the use of signals from damage state which are not usually available. Besides, the fact of having reference signals with limited range of EOCs in the training step does not guarantee the robustness of the proposed method in the case of abrupt variation in EOCS. Liu et al. (2015) have proposed an unsupervised damage detection method using Singular Value Decomposition (SVD) of the matrix of collected signals. It was developed on the premise that the effect of damage and the effect of EOCs will be represented in different singular vectors. In this case, damage is detected by observing a jump of the mean in the right singular vectors. But the question here is in which singular vector the jump can be observed and how can we automatically detect this jump. Also, this can be only done in the case where the following hypotheses are fulfilled. Firstly, it supposes that damage occurs abruptly, which is not always true because in real cases, damage may develop progressively during a long period of time. Secondly, EOCs should be constantly changing so a jump cannot be observed in their associated singular vectors, otherwise it will be considered as indication of presence of damage. In the present case, temperature changes, for example, could have a clear trend, so the latter hypothesis might be not verified.

In this study, since signals from healthy state are the only available information which can be provided in the training step, the proposed method for damage detection is based on a novelty detection technique. To deal with variation in EOCs, we consider a learning method in non-stationary environment. This method consists in estimating the current signal using only few signals among the reference database. Physically, if the current signal is from a healthy state it will be estimated using few reference signals with similar or very close EOCs, the others being discarded. As all identification methods, the proposed approach is based on the assumption that the database of reference signal contains sufficient variation of EOCs. Otherwise, an update of the reference database is necessary to ensure viable damage detection. Such development is not implemented in the present study. To localize the damage, we suggest applying the sparse estimation on a sliding window over the damaged signal. This is motivated by the fact that when dealing with UGW the effect of damage is only local on the signal.

In the next section, the proposed methodology for damage detection and localization in pipeline structure is presented. In section 3, the procedure for collecting the database of healthy and damaged signals is explained. Also, the preprocessing of these signals is exposed. Section 4 is devoted to discussions on the obtained results. Finally, section 5 concludes the paper.

2. METHODOLOGY

2.1 Sparse estimation

The idea behind using sparse estimation of the current signal is that an undamaged signal should be well estimated by reference signals which have similar EOCs. Initially, an estimation of the current signal $\boldsymbol{x} \in \mathbb{R}^m$ by the matrix of reference signals $C \in \mathbb{R}^{m\times n}$ could be provided by minimizing the quadric error given below:

$$\mathrm{J}(\boldsymbol{\theta}) = \|C\boldsymbol{\theta} - \boldsymbol{x}\|_2^2 \tag{1}$$

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{J}(\boldsymbol{\theta}) \tag{2}$$

where $\boldsymbol{\theta} \in \mathbb{R}^n$ is the vector of regression coefficients, m and n are the number of samples and the number of reference signals respectively.

The optimal regression coefficients $\widehat{\boldsymbol{\theta}}$ provide an estimation of the current signal by all the reference signals. But, the current signal is measured in specific EOCs while the reference signals are generally acquired in a wide range of EOCs. Hence, such estimation could give rise to an overfitting. This will jeopardize damage detectability because the regression model will tend to minimize the estimation error for a damage signal which should be very high compared to that of a reference signal. Besides, the coefficients should be positive to avoid the compensation between the reference signals. To overcome this issue, sparsity on the regression coefficients must be defined. It can be obtained by adding a regularisation term to equation (2) as following:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\mathrm{J}(\boldsymbol{\theta}) + \lambda\,\|\boldsymbol{\theta}\|_1\right) \quad \text{subject to } \boldsymbol{\theta} \geq \boldsymbol{0} \tag{3}$$

This optimization problem can be solved using the lasso method with non-negative constraint (Efron et al., 2004). The optimal solution will select from the reference signals a subset which will be used to estimate the current signal and assign zero to the others. The tuning parameter λ controls the power of regularization and it must be chosen in the training step. Generally, it is very difficult to determine a value of this parameter. It requires the use of cross validation technique. In this context, some studies have shown that in the case of positively correlated signals, least squares method with non-negative constraint is an efficient regularization technique (Meinshausen, 2013; Slawski et al. 2013). In other words, we can say that the method has a self-regularization property which means that it automatically generates a regularization term. Thus, there is no need to determine the tuning parameter λ. The condition of positively correlated variable is generally true for UGW signals. Mathematically, this condition is fulfilled if all the entries of the covariance matrix S of C are strictly positive:

$$\min_{i,j}\left(S\,(i,j)\right) \geq \sigma > 0 \tag{4}$$

the covariance matrix S being calculated by:

$$S = \frac{1}{n}C^T C \tag{5}$$

It must be noted here that the farther the parameter σ is to zero the higher the self-regularizing effect.

The non-negative least squares (NNLS) problem is defined as:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{J}(\boldsymbol{\theta}) \quad \text{subject to } \boldsymbol{\theta} \geq 0 \tag{6}$$

To get an optimal solution of this problem, the Karush-Kuhn-Tucker (KKT) conditions must be satisfied. These conditions are defined as follows (Bertsekas, 1999):

$$\widehat{\boldsymbol{\theta}} \geq 0 \tag{7}$$

$$\nabla\mathrm{J}(\widehat{\boldsymbol{\theta}}) \geq 0 \tag{8}$$

$$\nabla\mathrm{J}(\widehat{\boldsymbol{\theta}})^T\widehat{\boldsymbol{\theta}} = 0 \tag{9}$$

where ∇ denotes the gradient of $\mathrm{J}(\widehat{\boldsymbol{\theta}})$. Its expression is given by:

$$\nabla\mathrm{J}(\widehat{\boldsymbol{\theta}}) = 2C^T\left(C\widehat{\boldsymbol{\theta}} - \boldsymbol{x}\right) \tag{10}$$

The constrained minimization problem described in equation (6) can be solved using different numerical approaches. The most common ones are: active set method, projected quasi Newton approach, principal block pivoting method and interior point method (Chen and Plemmons, 2009). A comparison between these approaches is beyond the scope of this study. However, we have chosen to use the active set method because a recursive version of this method can be easily implemented. This will be used later for the purpose of localization of damage. Active set method divides the constraints into active and passive ones. Actually, active constraint refers to zero regression coefficient, otherwise it is considered as passive. At each iteration, the algorithm calculates the solution by the least squares method on the passive set. Then, it tests if the new regression coefficients satisfy the KKT conditions. Afterthat, it updates the set of active coefficients until a final set is found. Lawson and Hanson (1995) have proposed an algorithm to solve NNLS problem which is an active set method. Let us denote by C^P the matrix associated with the constraints in the passive set P. Similarly, let $\boldsymbol{s}$ be a vector of same length as $\boldsymbol{\theta}$. Let $\boldsymbol{s^P}$ denote the subvector with indexes from P and $\boldsymbol{s^R}$ the subvector with indexes from the active constraints R. The algorithm is described hereafter:

Algorithm 1: Non Negative Least Squares

Inputs: C, $\boldsymbol{x}$ and ϵ a real value fixing the tolerance for stopping criterion.

Initialization: $P = \emptyset$, $R = \{1,2,\dots,n\}$, $\widehat{\boldsymbol{\theta}} = 0$, $\boldsymbol{w} = \frac{1}{2}\nabla \mathrm{J}(\widehat{\boldsymbol{\theta}})$

While: $R \neq \emptyset$ and $(max_{i\in R}(w_i) > \epsilon)$ **do**

$j = \arg max_{i\in R}(w_i)$

Include the index j in P and remove it from R

$\boldsymbol{s^P} = [(C^P)^T C^P]^{-1}(C^P)^T \boldsymbol{x},\ \boldsymbol{s^R} = 0$

While: $\min(\boldsymbol{s^P}) \leq 0$ **do**

$\alpha = min_{i\in P}[\frac{\widehat{\theta}_i}{(\widehat{\theta}_i - s_i)}]$ where $s_i \leq 0$

$\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} + \alpha(\boldsymbol{s} - \widehat{\boldsymbol{\theta}})$

Move to R all indices in P such that $\widehat{\boldsymbol{\theta}}_j = 0$

$\boldsymbol{s^P} = [(C^P)^T C^P]^{-1}(C^P)^T \boldsymbol{x}, \boldsymbol{s^R} = 0$

end while

$\widehat{\boldsymbol{\theta}} = \boldsymbol{s}, \boldsymbol{w} = \frac{1}{2}\nabla \mathrm{J}(\widehat{\boldsymbol{\theta}})$

end while

It was proved that this algorithm converges within a finite number of iterations. However, an exact number of these iterations cannot be known in advance and it could be very high.

2.2 Damage indicators

To detect the presence of damage, an indicator must be derived from the sparse estimation of the current signal. If the signal is damaged, we expect that the estimation error will be very high compared to that of an undamaged signal. Hence, a damage indicator could be chosen as the quadratic estimation error $\mathrm{J}(\widehat{\boldsymbol{\theta}})$ given by the solution of NNLS problem described in equation (6):

The sparsity of the regression coefficients (Sr) could be also used as damage indicator (Eybpoosh et al., 2016). It is defined as the ratio of the number of zero coefficients in $\widehat{\boldsymbol{\theta}}$ to the total number of reference signals:

$$Sr = \frac{n - \mathrm{card}(P)}{n} \times 100 \tag{11}$$

where $\mathrm{card}(P)$ denotes the cardinal of the final set of the passive constraints P. In fact, a damaged signal will be characterized by a very small number of passive coefficients. Hence, the value of Sr will converge to unity. While, for a reference signal, the value of Sr will be theoretically a little far from unity. Furthermore, this sparsity will increases as the damage size increases. Thus, the damage severity can be assessed using this factor.

A joint damage indicator I that account for both the quadratic estimation error and the sparsity ratio could be established to maximize the chances for damage detectability as proposed in (Boracchi et al., 2014). It is defined by the vector:

$$I = (\mathrm{J}(\widehat{\boldsymbol{\theta}}), Sr) \tag{12}$$

To ensure automatic damage detection, a threshold must be defined. For this purpose, in the case of one dimensional damage indicator (i.e. $\mathrm{J}(\widehat{\boldsymbol{\theta}})$ or Sr), an empirical distribution for damage indicators, calculated from reference signals, has to be established. Afterthat, the threshold can be chosen by fixing the confidence limit of the distribution. The current signal is considered as damage if its damage indicator exceeds the value of the specified confidence limit. When using a bivariate indicator as defined in equation (12), the signals will be represented in a two-dimensional space. In this case, the threshold can be chosen by defining a metric to measure how close the damage indicator vector of the current signal to those of the reference signals. This metric can be calculated using multivariate statistical tools such as the Mahalanobis Square Distance (Johnson and Wichern, 2014).

2.3 Damage localization using recursive NNLS

To get the position of damage, we suggest calculating the sparse estimation of the damaged signal on a sliding window over the matrix of reference signals. The already found regression coefficients, obtained for the entire signal, are no longer suitable. Hence, the solution of NNLS problem must be obtained for every sliding window. Taking into consideration the fact that adding and removing one sample at a time do not change significantly the regression coefficients found in the previous window, it is of interest to apply a recursive version for the solution of the NNLS.

In the case of classical least squares method (unconstrained), a method to update the regression coefficients already exists (Björck, 1996). It is known as Recursive Least Squares (RLS). Since the NNLS solving method requires the use of least squares on the passive set, the RLS could be applied to update NNLS solution. However, the non-negativity of the new regression coefficients is not guaranteed. In other words, the KKT conditions could be violated. Thus, to get an optimal solution of the recursive NNLS problem, these KKT conditions must be verified. Afterthat, a pivoting exchange between the passive set and the active set could be eventually needed (Mosesov, 2014). Under the assumption of minimal set changes, we expect that only a single pivoting exchange is made between the active set and the passive set. Hence, block pivoting is not necessary in this case.

The calculation of the RLS on a sliding window includes two successive steps: updating and downdating (Zhang, 2000). The former adds a new sample to the window and the latter removes a sample which is excluded from the sliding window. Let us denotes by H the window width and by k the index of the sample. The algorithm starts first with a window which comprises samples from k to $k+H-1$. This window corresponds to the matrix of reference signals C_{k+H-1} and the current signal $\boldsymbol{x}_{k+H-1}$. The solution in this case is denoted as $\hat{\boldsymbol{\theta}}_{k+H-1}$.When adding a new sample $(\boldsymbol{c}_{k+H}, x_{k+H})$ where $C_{k+H} = \begin{pmatrix} C_{k+H-1} \\ \boldsymbol{c}_{k+H} \end{pmatrix}$ and $\boldsymbol{x}_{k+H} = \begin{pmatrix} \boldsymbol{x}_{k+H-1} \\ x_{k+H} \end{pmatrix}$, the updating operation can be written as:

$$A_{k+H-1} = (C_{k+H-1}^T C_{k+H-1})^{-1} \tag{13}$$

$$\boldsymbol{\beta}_{k+H} = \frac{A_{k+H-1}\boldsymbol{c}_{k+H}^T}{1 + \boldsymbol{c}_{k+H}A_{k+H-1}\boldsymbol{c}_{k+H}^T} \tag{14}$$

$$\hat{\boldsymbol{\theta}}_{k+H} = \hat{\boldsymbol{\theta}}_{k+H-1} + \boldsymbol{\beta}_{k+H}\left(x_{k+H} - \boldsymbol{c}_{k+\mathrm{H}}\hat{\boldsymbol{\theta}}_{k+H-1}\right) \tag{15}$$

These values will be then used to perform the downdating operation by removing the sample $(\boldsymbol{c}_k, x_k)$. It is expressed as follows:

$$\underline{A}_{k+H-1} = A_{k+H-1} - \frac{A_{k+H-1}\boldsymbol{c}_{k+\mathrm{H}}^T\boldsymbol{c}_{k+\mathrm{H}}A_{k+H-1}}{1 + \boldsymbol{c}_{k+\mathrm{H}}A_{k+H-1}\boldsymbol{c}_{k+\mathrm{H}}^T} \tag{16}$$

$$\underline{\boldsymbol{\beta}}_{k+\mathrm{H}-1} = \frac{\underline{A}_{k+H-1}\boldsymbol{c}_k^T}{1 - \boldsymbol{c}_k\underline{A}_{k+H-1}\boldsymbol{c}_k^T} \tag{17}$$

$$\underline{\hat{\boldsymbol{\theta}}}_{k+\mathrm{H}-1} = \hat{\boldsymbol{\theta}}_{k+H} - \underline{\boldsymbol{\beta}}_{k+\mathrm{H}-1}\left(x_k - \boldsymbol{c}_k\hat{\boldsymbol{\theta}}_{k+H}\right) \tag{18}$$

Finally, the algorithm for recursive NNLS can be described hereafter:

Algorithm 2: Recursive Non Negative Least Squares

Inputs: data C_{k+H-1} and $\boldsymbol{x}_{k+H-1}$

Update $\mathbf{s}^{\mathbf{P}} = \arg\min_{\theta}\|C_{k+H-1}^P\boldsymbol{\theta} - \boldsymbol{x}_{k+H-1}\|_2^2$ using RLS by adding sample $(\boldsymbol{c}_{k+H}, x_{k+H})$ as in equation (15) and removing $(\boldsymbol{c}_k, x_k)$ as in equation (18).

If $(min(\boldsymbol{s}^{\boldsymbol{P}}) \leq 0)$ then remove index $j = arg\,min_i(\,s_i)$ from P and add it to R.

Calculate $\boldsymbol{w} = \frac{1}{2}\nabla\mathrm{J}(\hat{\boldsymbol{\theta}})$, if $(max(\boldsymbol{w}) \geq 0)$ then move index $j = \arg max_i(\,w_i)$ into P

While (KKT conditions are not fulfilled) **do**

Calculate $\boldsymbol{s}^{\boldsymbol{P}} = [(C_{k+H-1}^P)^T C_{k+H-1}^P]^{-1}(C_{k+H-1}^P)^T\boldsymbol{x}_{k+H-1}$

If $(min(\boldsymbol{s}^{\boldsymbol{P}}) \leq 0)$ then remove index $j = arg\,min_i(\,s_i)$ from P and add it to R.

Calculate $\boldsymbol{w} = \frac{1}{2}\nabla\mathrm{J}(\hat{\boldsymbol{\theta}})$, if $(max(\boldsymbol{w}) \geq 0)$ then move index $j = arg\,max_i(\,w_i)$ into P

end while

Notice that the window width (H) must be chosen in order to ensure the invertibility of the matrix $((C_{k+H-1}^P)^T C_{k+H-1}^P)$ which not does represent a difficult issue. However, when this matrix is not invertible, QR decomposition might be used to enforce the stability of the proposed recursive algorithm (Björck, 1996). The QR decomposition could be also used to calculate recursively the vector $\boldsymbol{s}^{\boldsymbol{P}}$ in the case of adding and/or removing a column from the matrix C_{k+H-1}^P which results from the single pivoting exchange made between the active set and the passive set.

3. DATA COLLECTION AND PREPROCESSING

3.1 Data collection

The specimen considered in this study consists of a tube with 6.4 m length. It was placed in laboratory conditions where temperature fluctuates between 19°C and 26°C during the monitoring period. This variation is due to the seasonal temperature changes. The used sensor can excite two separate guided waves modes which are: torsion and flexion at five different frequencies: 14, 18, 24, 30 and 37 kHz. For the sake of brevity, only torsional mode with an excitation frequency of 14 kHz is considered in the present study. Figure 1 illustrates the pipeline configuration.

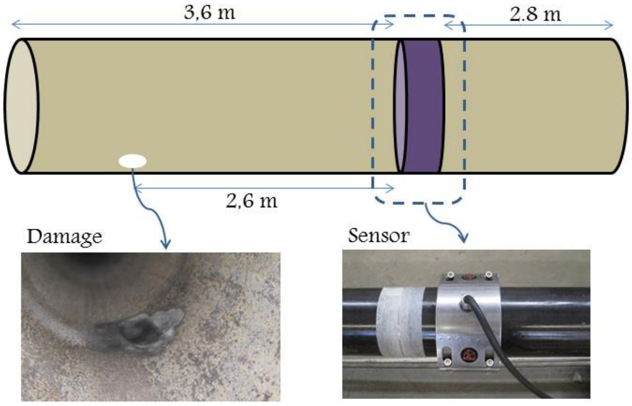

Fig. 1. Pipeline configuration showing the used sensor and the created damage.

The specimen has been monitored during a period of almost 3 months. Each week, multiple measurements were scheduled. At each measurement, five signals were acquired in the morning and at the evening in order to capture temperature changes during the day and to investigate its effects on the collected signals. An example of an acquired signal is illustrated in Figure 2.

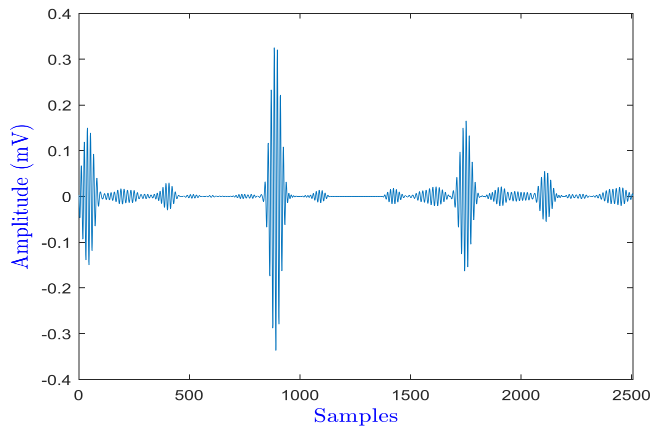

Fig. 2. Example of an ultrasonic guided wave signal excited with torsion mode (frequency: 14 kHz).

The excitation signal has been removed by the acquisition system (portion in the middle of the signal). The three echoes with the highest amplitude represent multiple reflections from the end of the pipe. These echoes have to be removed also from the original signal because they are not useful for damage detection. Besides, their amplitudes are very high, hence they can mask small changes induced by the presence of damage.

Damage was created by removing material from the inside of the pipeline in six increasing steps in order to simulate corrosion growing within the structure. Figure 1 in bottom left shows the defect in the last step. The dimensions and the form of the damage were set randomly as in the case of real corrosion which results from a natural aging of the pipeline.

At the end of the monitoring period, a total of 236 signals were collected where 207 ones were undamaged and 29 signals were acquired from a damage state. Table 1 summarizes information related to the collected database.

Table 1. Characteristics of the collected data

Monitoring period	*3 months*
Reference state	*207 signals*
Damage state	*6 increasing defects (29 signals)*
Temperature	*19 °C → 26 °C*

3.2 Temperature effect

The analysis of the collected signals in the time domain shows that the reference signals underwent drastic changes in both amplitude and phase. Since temperature is the only environmental factor that varied during the measurements, the observed changes in the collected signals are due to the variation of this factor. This can be clearly noticed in Figure 3 which illustrates a zoom over a part of two reference signals acquired at different temperature.

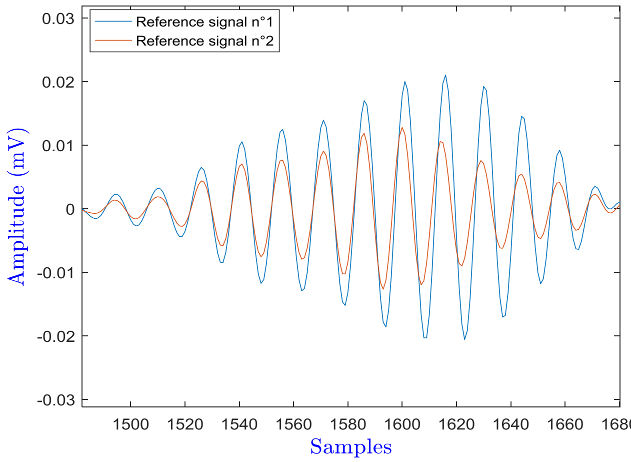

Fig. 3. Zoom over two reference signals acquired at different temperatures.

3.3 Damage effect

Figure 4 illustrates an example of a damage signal and a reference one. Note that, to clearly see the differences between the two signals, a zoom was displayed at the position of damage. This figure shows that the effect of damage is drift in the amplitude with a significant change in the phase. These results seem to be very similar to those observed in the case of temperature variation. However, the changes in the damaged signal are not uniform as in the case of temperature variation. We can clearly observe in the beginning that the two signals are almost identical while at the end, the amplitude of damaged signal drops suddenly and the time arrival of the two signals changes. This confirms the hypothesis that the effect of damage is local on the signal while the temperature variation impacts the entire signal.

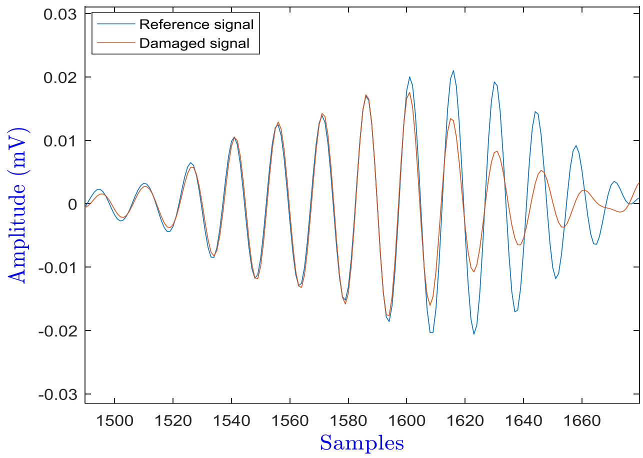

Fig. 4. Zoom at the position of damage for a reference signal and a damaged signal.

Even though, damage could not be detected using classical damage sensitive features proposed by Rizzo et al. (2007). Let us consider for example the RMS as a damage sensitive feature. As a reminder, the RMS of a signal $\boldsymbol{x} = \{x_i\}, \boldsymbol{x} \in \mathbb{R}^n$ is given by:

$$\mathrm{RMS}(\boldsymbol{x}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2} \quad (19)$$

Result in Figure 5 shows that we cannot discriminate between damage and reference signals using the RMS. The variation of temperature is responsible of the great dispersion of RMS values. The dispersion is mainly associated with the amplitude of the signals since RMS uses only amplitude values at each signal sample.

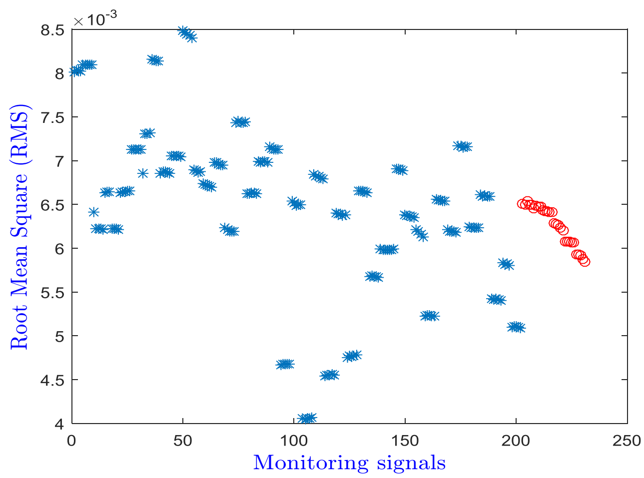

Fig. 5. RMS of the monitoring signals, the damaged signals are coloured in red with a circle marker.

4. RESULTS AND DISCUSSIONS

4.1 Damage detection

One hundred and forty signals were used as a database of reference signals and the others served for the test of the proposed method. Due to the seasonal changes in temperature, the reference signals used for the test could be obtained with different variation of temperature which might be not included in the database of reference signals. Each new signal ($\boldsymbol{x}$) is estimated by sparse model $\widehat{\boldsymbol{\theta}}$ obtained by solving the constrained minimization problem described in equation (6). Here, the sparse estimation is calculated on the entire signal. Let us verify that the reference signals are positively correlated which is the condition of using NNLS rather than the non-negative lasso. In this study, the minimum of the covariance matrix was calculated for the reference signals. It is equal to 0.0414 which is strictly positive. Thus, the condition of positively correlated signals is fulfilled as stated in inequality (4). To confirm this, Principal Component Analysis (PCA) was firstly performed on the matrix of reference signals (Jolliffe, 1986). Afterthat, the correlation coefficient between the first two representative principal components and the reference signals was calculated. The result is presented in Figure 6.This result shows that the reference signals are well represented by the first principal component (the correlation coefficient exceeds 0.88). This can be explained by the fact that there is a strong positive correlation between the reference signals. The variation of the correlation between the reference signals and the second principal component is probably due to the changes in EOCs namely the temperature.

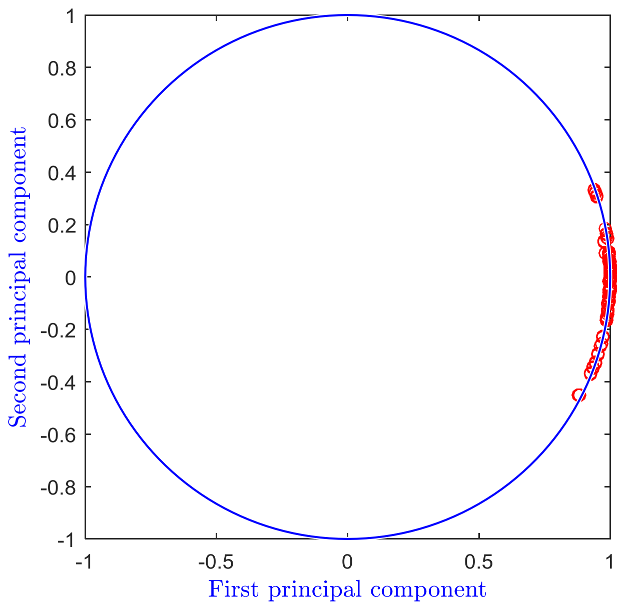

Fig. 6. Correlation coefficients of the reference signals and the first two principal components.

Since the condition of positively correlated signals is now strongly fulfilled, an estimation of a new measured signal can be provided using NNLS. Figure 7 shows an example of an original undamaged signal and its estimation. It can be clearly noticed that the two signals are almost the same which proves that the sparse estimation provides a very good estimation of the undamaged signal.

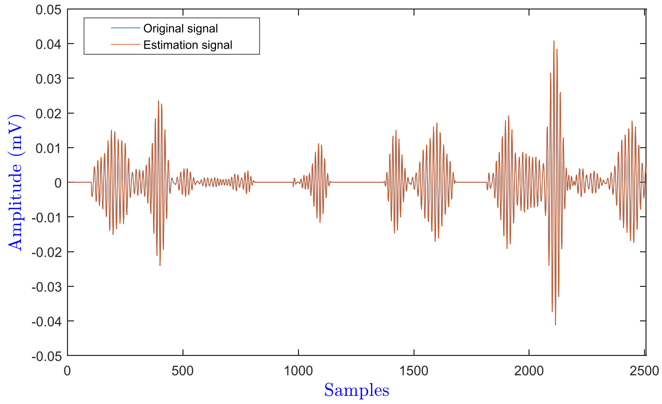

Fig. 7. Original undamaged signal and its estimation.

Figure 8 illustrates the error calculated for the original signal showed in Figure 7. This error presents a random behaviour, except in the zones where the echoes from the end of pipe and excitation signal were deleted. When a defect is present, the error will tend to deviate from the random behaviour to a deterministic one. Also, a damaged signal will be badly estimated by the reference signal; in this case the error will be very high.

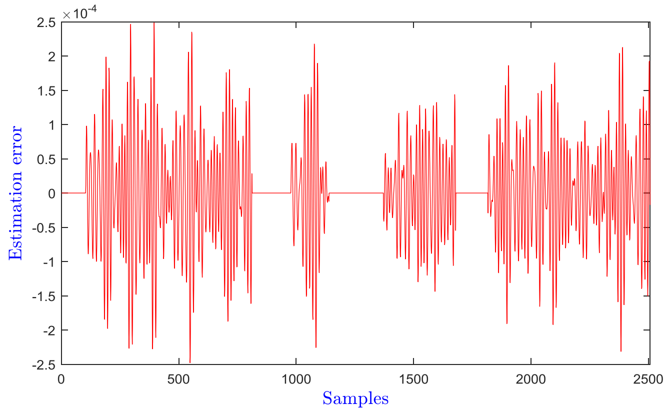

Fig. 8. Estimation error of the reference signal shown in Fig 7.

Figure 9 illustrates an example of the estimation error of a damaged signal. It can be noticed that orders of the magnitude of the estimation error for a damaged signal are far from that of a reference signal.

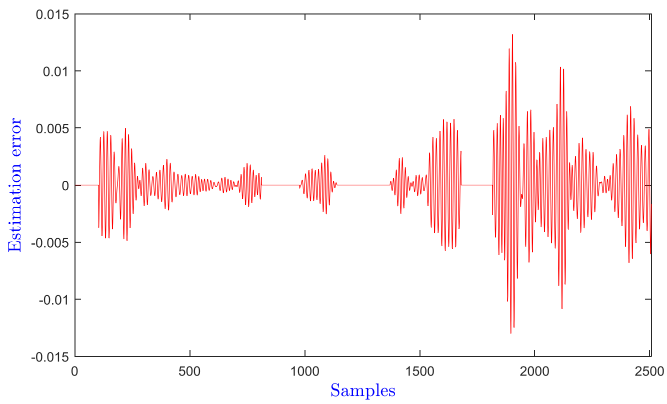

Fig. 9. Estimation error of a damaged signal.

The quadratic estimation error $J(\widehat{\boldsymbol{\theta}})$ was calculated for all test signals including undamaged and damaged signals. The result is shown in Figure 10. Different comments can be drawn from this result:

- Damaged signals are well separated from the undamaged ones. Hence, damage can be detected automatically by defining a threshold.
- The quadratic estimation error $J(\widehat{\boldsymbol{\theta}})$ of the reference signals presents very low variation if we compared them with the result of RMS shown in Figure 7.
- The values of $J(\widehat{\boldsymbol{\theta}})$ increase as the size of damage increases. Thus, $J(\widehat{\boldsymbol{\theta}})$ can be used to assess the severity of damage.

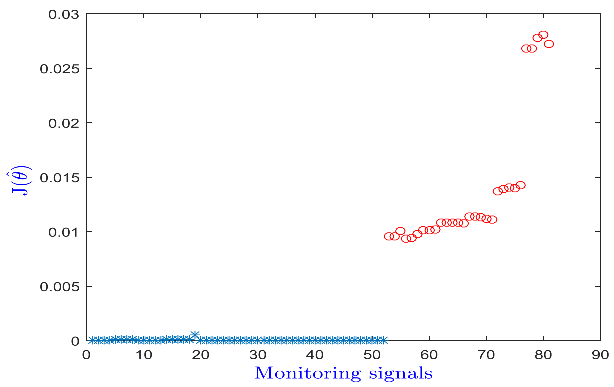

Fig. 10. Quadratic estimation error $J(\widehat{\boldsymbol{\theta}})$ of undamaged and damaged signals.

As described in section 2.2 the sparsity ratio (Sr) can be also used for damage detection. Figure 11 shows the Sr for the monitoring signals including damaged and reference signals. The variation of the Sr for the reference signals is important while the Sr of damaged signals presents a relatively low variation. Besides, the values of Sr increase as the size of damage increases. However, a threshold cannot be defined to ensure automatic damage detection without triggering false alarms. Thus, $J(\widehat{\boldsymbol{\theta}})$ outperforms Sr in terms of damage detectability. In this case, the use of a bivariate indicator which was defined in equation (12) will be useless. The quadratic estimation error $J(\widehat{\boldsymbol{\theta}})$ is sufficient to ensure reliable damage detection.

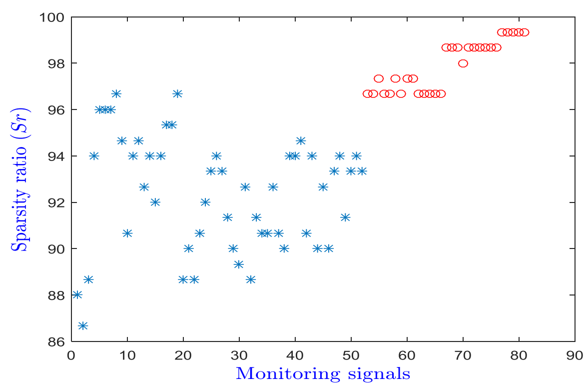

Fig. 11. Sparsity ratio of undamaged and damaged signals.

4.2 Damage localization

To determine the position of damage, the idea which is described in section 2.3, is to apply the sparse estimation on a sliding window over the signal (i.e. samples). Thus, by moving the window through the signal, each part of it will be examined separately. In this case, the window width may influence the final result. It can be determined based on the desired precision of localization. In the present study, it was set initially at forty samples. Note that larger window width could be tested to investigate its influence on the result of localization. This window was moved by one sample per step. At each step, we calculate the quadratic estimation error $J(\widehat{\boldsymbol{\theta}})$ using recursive NNLS algorithm. At the end, we obtain a value of $J(\widehat{\boldsymbol{\theta}})$ at each moving window over the damaged signal.

Figure 12 shows the found position of the created defect using the proposed method for damage localization. To avoid interpretation of false damage position, the damage signal was truncated between the first two arrivals of the end of pipe echoes. This Figure shows that the position of damage which corresponds to the maximum of $J(\widehat{\boldsymbol{\theta}})$ is 3 m. The real position of damage, as shown in Figure 1, is 2.6 m. Thus, we have an error of localization of 0.4 m. This error can be explained by the fact that the window width (H) induces an error of localization. This error can be calculated by knowing the propagation velocity (V) of the waves in the pipeline and the sampling frequency (F). Knowing that this error has been minimized by affecting the value of $J(\widehat{\boldsymbol{\theta}})$ to the position of the half of the window. The error of localization E is given as follows:

$$E = \frac{V}{F} \times \left(\frac{H}{2}\right) \tag{20}$$

The propagation velocity of the waves in the pipeline is equal to 3200 m/s. The sampling frequency is fixed by the acquisition system at 195 kHz. Finally, the position of damage D is:

$$D = 3 \pm 0.41\, m \tag{21}$$

This result is coherent with the value of the real position of damage indicated in Figure 1.

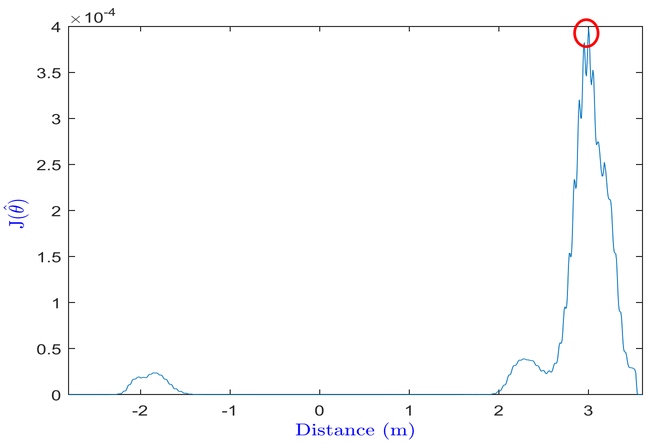

Fig. 12. Quadratic estimation error $\mathbf{J}(\widehat{\boldsymbol{\theta}})$ of a damaged signal showing the position damage.

Finally, it is worth noting that the result of localisation could be optimized by assigning the position of damage to a value of $J(\widehat{\boldsymbol{\theta}})$ where we estimate that the error is significant and not necessarily to the maximum of $J(\widehat{\boldsymbol{\theta}})$.

5. CONCLUSION

In this paper, a method for damage detection and localization in pipeline structures was proposed. It is based on sparse estimation of the measured signals by the reference signals. A simplified form of this estimation using the non-negative least squares is investigated. It is based on the fact that the acquired UGW signals are highly correlated. The sparsity helps to enhance damage detectability because a damaged signal will have a high estimation error compared to that of a healthy signal. Besides, it can face the problem of variation in EOCs provided that the database of reference signals contains large variations of these EOCs.

The detection of defect was ensured by calculating the quadratic estimation error $J(\widehat{\boldsymbol{\theta}})$ on the entire current signal. While the localization of damage was established by implementing a recursive version of the sparse estimation on a sliding window over the damaged signal.

As a perspective of this work, an update of the database of reference signals could be considered in the case where these signals present limited variation in EOCS. This can be achieved by adding to this database new healthy signals with unknown variation in EOCs. Also, the proposed method has to be validated on operational pipeline serves in different EOCs.

6. REFERENCES

Bertsekas, D.P. (1999). *Nonlinear programming* (p. 794). Belmont: Athena scientific.

Björck, Å. (1996). *Numerical methods for least squares problems*. Society for Industrial and Applied Mathematics.

Boracchi, G., Carrera, D., & Wohlberg, B. (2014). Novelty detection in images by sparse representations. IEEE Symposium on Intelligent Embedded Systems, 47-54, Orlando, FL, USA.

Chen, D., & Plemmons, R.J. (2009). Nonnegativity constraints in numerical analysis. Symposium on the birth of numerical analysis, 109-140, Leuwen, Belgium.

Croxford, A. J., Moll, J., Wilcox, P.D., & Michaels, J.E. (2010). Efficient temperature compensation strategies for guided wave structural health monitoring. *Ultrasonics*, 50(4), 517-528.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, 32(2), 407-499.

Eybpoosh, M., Bergés, M., & Noh, H.Y. (2016). Sparse representation of ultrasonic guided waves for robust damage detection in pipelines under varying environmental and operational conditions. *Structural Control and Health Monitoring*, 23(2), 369-391.

Eybpoosh, M., Bergés, M., & Noh, H.Y. (2017). An energy-based sparse representation of ultrasonic guided-waves for online damage detection of pipelines under varying environmental and operational conditions. *Mechanical Systems and Signal Processing*, 82, 260-278.

Farrar, C.R., & Worden, K. (2007). An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1851), 303-315.

Johnson, R.A., & Wichern, D.W. (2014). *Applied multivariate statistical analysis* (Vol. 4). New Jersey: Prentice-Hall.

Jolliffe, I.T. (1986). *Principal Component Analysis and Factor Analysis* (pp. 115-128). Springer New York.

Lawson, C.L., & Hanson, R.J. (1995). *Solving least squares problems*. Society for Industrial and Applied Mathematics.

Liu, C., Harley, J.B., Bergés, M., Greve, D.W., & Oppenheim, I.J. (2015). Robust ultrasonic damage detection under complex environmental conditions using singular value decomposition. *Ultrasonics*, 58, 75-86.

Lowe, M.J., Alleyne, D.N., & Cawley, P. (1998). Defect detection in pipes using guided waves. *Ultrasonics*, 36(1-5), 147-154.

Meinshausen, N. (2013). Sign-constrained least squares estimation for high-dimensional regression. *Electronic Journal of Statistics*, 7, 1607-1631.

Mosesov, A. (2014). Adaptive Non-negative Least Squares with Applications to Non-Negative Matrix Factorization. Master dissertation, University of Minnesota.

Rizzo, P., Sorrivi, E., di Scalea, F.L., & Viola, E. (2007). Wavelet-based outlier analysis for guided wave structural monitoring: application to multi-wire strands. *Journal of Sound and Vibration*, 307(1), 52-68.

Slawski, M., & Hein, M. (2013). Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. *Electronic Journal of Statistics*, 7, 3004-3056.

Sohn, H. (2007). Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1851), 539-560.

Zhang, Q. (2000). Some implementation aspects of sliding window least squares algorithms. IFAC Proceedings Volumes, 33(15), 763-768.