Condition assessment of pressurized tubular structure using wave data

Student name: Najoua HAMDAOUI (CIVL)

Student ID:00149674

nhamdaoui@connect.ust.hk

I. Introduction

Pressurized tubular structures transporting fluids like oil and water can suffer from serious damages and defects, leading to economic losses, environmental harm, and risks to the public [1]. For example, oil pipeline defects in the USA have caused billions of dollars in damages, hundreds of fatalities, and thousands of injuries. Water pipe failures also have negative effects, causing water wastage, flooding, and disruptions to traffic and urban centers [2][3]. The direct and indirect costs of pipe bursts can be significant. In Asia, the amount of lost water from urban distribution networks has doubled in less than 10 years, costing billions of dollars annually [4]. To address these issues, we aim to use Principal Component Analysis (PCA) and Maximum Likelihood Estimate (MLE) techniques [7] to classify damaged conduits and determine the location and size of damage. The project involves collecting data from mathematical models and experiments, applying PCA and MLE to synthetic and real data, and drawing conclusions about their effectiveness in assessing pressurized tubular structures' health.

Key words: PCA, SVD, Resolution, MLE, signal processing, damage detection, non-convex optimization

II. Terminology

Classification: in this project classification is the process by which we decide whether a damage exists or not given the data. For this purpose, PCA will be used.

Localization: is the process of finding the location of the damage. This will be accomplished through MLE approach.

Sizing: is the step in which we estimate the size of the damage using MLE.

Condition assessment: involves classification followed by localization and finally sizing.

III. Damage classification using PCA

The classification is performed using data collected at the experimental facility situated at Beacon Hill (BH), Kowloon, Hong Kong. The facility is composed of four high-density polyethylene (HDPE) pipelines with a diameter of 150 mm (NS180), connected through a junction as illustrated in Figure 1. The viscoelastic parameters of the pipeline are provided in [1]. The pipeline is almost horizontal in elevation and about 0.5m above the ground, supported by extrusion bars spaced at 1.5 to 2m internally. Incoming hydraulic pressure is regulated by an upstream pressure-reducing valve (PRV), maintaining the pressure head within the range of 0.5 to 6 bar. The upstream section of the pipeline, stretching from the PRV to branch junction J1, measures 65.6m in length. Subsequent pipelines extend J1 to downstream valves with lengths of 79.23m, 72.43m, and 52.61m, respectively. The flow is approximately 9 $L.s^{-1}$, while the simulated leak flow is around 3 $L.s^{-1}$. The distance of leak locations is taken from the upstream of each pipeline section. During the experiment, the gate valve GV3 at junction J2, and the ball valves BV2 and BV6 remained closed. The source of the transient wave is the downstream ball valve of pipeline 4, BV5, and the time of closure of the valve is 50 to 70 milliseconds. The measurement point M42 is at the upstream of BV5 denoted by a yellow square in Figure 1. The data matrix is structured such that its rows correspond to the different 18 scenarios (e.g., intact, leak at L11), and the columns represent the pressure response samples at each measurement locations (M11, M12, M21, M22, M31, M32, M41, M42). The sampling frequency for data collection is set to 1000Hz, and the transient response is acquired for 5s. Therefore, there are 5000 samples of the pressure response (number of columns in the data matrix). The steady state has been subtracted from the transient response for all sensors.

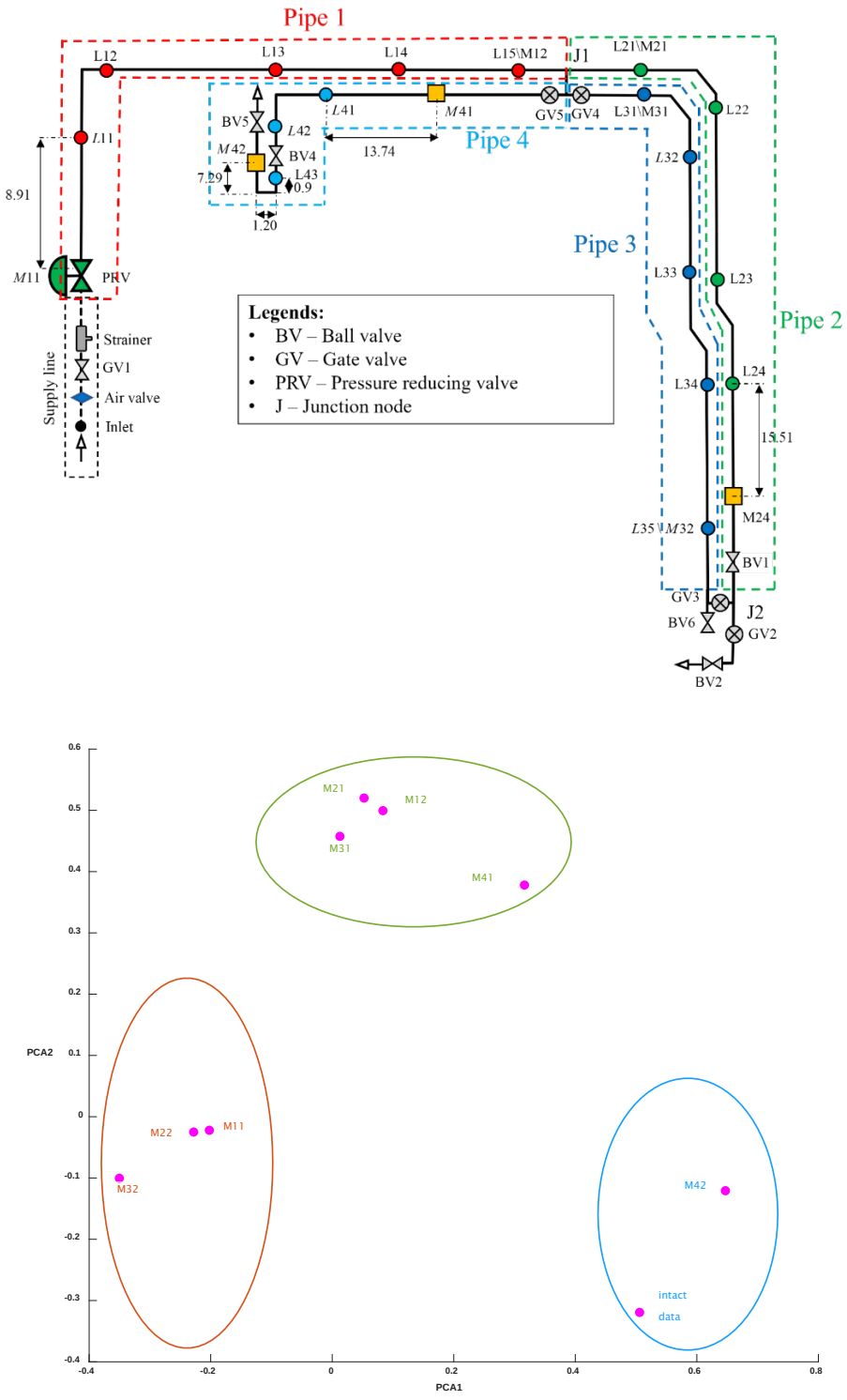

Figure1: *(a) BH facility setup sketch; (b) PCA1 versus PCA2 of a signal data collected at Beacon Hill, Hong Kong*

Using the data corresponding to damage L11 and 8 different sensors, we constructed a matrix A that includes 9 rows and 5001 columns. The first row represents the intact data (no damage) and the remaining 8 rows each represents the pressure signal corresponding to the leakage L11 as shown in figure1-(a).

The PCA results for the first two PCA components are shown in figure1-(b). The figure displays 3 clusters. Cluster 1 represents the projection of the signals measured at sensors M21, M12, M31 and M41 on the singular vectors $\vec{V}_1$and $\vec{V}_2$. We notice that the four sensors give a redundant information because they're located in close proximity to each other. Therefore, to save cost, it may be advisable to limit the number of sensors.

Cluster 2 shows the projection of the signals measured at sensors M22, M11 and M32 on the singular vectors $\vec{V}_1$and $\vec{V}_2$. Again, these sensors are providing nearly redundant information. The last cluster groups the intact data and sensor 42. Sensor 42 is not capturing a significant scattering from the damage because its location is quite remote from the damage and the wave path is complicated with many geometric discontinuities that are preventing reflections from the damage to the measurement station.

Importantly, the PCA is a powerful methodology at distinguishing different signals and will be used in the next section to ascertain if a damage exists or not. This is critical step to perform before embarking on detailed damage localization and sizing.

IV. Condition assessment

IV.1. Classification

At this juncture, we present the mathematical model that is used to generate the wave data. Let $H(t)$ denote the signal model given by:

$$H(t) = \frac{1}{\varepsilon\sqrt{2\pi}} e^{-\frac{1}{2\varepsilon^2}(t-0.05)^2} - S_D \frac{1}{\varepsilon\sqrt{2\pi}} e^{-\frac{\left(t-\frac{2x_D}{C}-0.05\right)^2}{2\varepsilon^2}} \qquad (1)$$

where the first term in the l hand side represents the incident wave; the second term represents the scattered wave by the damage; S_D is the size of the damage; x_D is the location of the damage; C is the wave speed and ε is a measure of the width of the incident wave. Let $n(t)$ be a Gaussian white noise with mean 0 and variance σ^2, then equation (1) becomes [9]:

$$\overrightarrow{D}(t) = \vec{H}(t) + \vec{n}(\mathrm{t}) \qquad (2)$$

Where $\overrightarrow{D}(t)$ is the data sampled at J discrete times. Equation (2) can be rearranged as:

$$\vec{n}(t) = \overrightarrow{D}(t) - \vec{H}(t) = \overrightarrow{D}(t) - \vec{H}_I + S_D\vec{G}(t_D) \qquad (3)$$

where $\vec{H}_I$: Incident wave ($\frac{1}{\varepsilon\sqrt{2\pi}} e^{-\frac{1}{2\varepsilon^2}(t-0.05)^2}$)

$\vec{G}(t_D) = \frac{1}{\varepsilon\sqrt{2\pi}} e^{-\frac{(t-2t_D-0.05)^2}{2\varepsilon^2}}$), with $t_D = \frac{x_D}{C}$ (the time at which the incident wave reaches the damage) (4)

Let $\Delta\vec{H} = \vec{D}(t) - \vec{H}_I$ (5)

To classify whether a signal is present or not in the data, we generated synthetic signals using equation (2) for (i) $\vec{H}(t) = \vec{0}$ (i.e. no signal) and (ii) $\vec{H}(t) \neq \vec{0}$. The parameters used are: $\varepsilon = 0.0001$, $S_D = 0.1$, $x_D = 0.1m$ and $C = 1000m/s$. The resulting data was processed using the PCA and the finding are shown in figure2.

It is evident that the results with signal are well separated from the results with no signals. Below, we take the results with damage signal and perform localization and sizing using MLE.

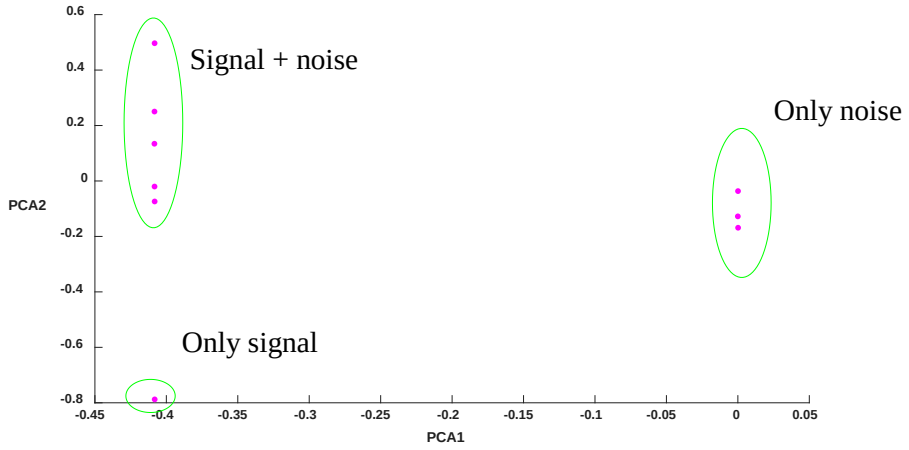

Figure2: *Observing the efficiency of the PCA method in classifying signals with noise, signals without noise, and pure noise using synthetic data generated by equation (2).*

It is evident that PCA1 captures a significant amount of information compared to PCA2. The separation between different classes is more pronounced in PCA1, indicating that it contains more relevant features for classification. This suggests that PCA1 can effectively distinguish between signals, noise, and no noise, while PCA2 may have less discriminative power.

IV.2. Localization and sizing using MLE

IV.2.1. Mathematical Model of the signal

The probability density function (PDF) of the noise is given by:

$$f\big(\vec{n}(t)\big) = \frac{1}{\sigma^J(2\pi)^{J/2}} e^{-\frac{1}{2\sigma^2}\vec{n}^T(t)\,\vec{n}(t)} \tag{6}$$

where J is the number of pressure samples. Hence maximum likelihood of the noise function is given by:

$$L\big(\sigma^2, \vec{n}(t)\big) = f\big(\vec{n}(t)\big)$$

$$= \frac{1}{\sigma^J(2\pi)^{J/2}} e^{-\frac{1}{2\sigma^2}\vec{n}^T(t)\,\vec{n}(t)}$$

$$= \frac{1}{\sigma^J(2\pi)^{\frac{J}{2}}} e^{-\frac{1}{2\sigma^2}\vec{n}^T(t)\,\vec{n}(t)}$$

$$= \frac{1}{\sigma^J(2\pi)^{\frac{J}{2}}} e^{-\frac{1}{2\sigma^2}\left(\Delta\vec{H}+S_D\vec{G}(t_D)\right)^T\left(\Delta\vec{H}+S_D\vec{G}(t_D)\right)} \tag{7}$$

Maximizing the likelihood is equivalent to minimizing

$$\Phi(S_D, t_D) = \left(\Delta\vec{H} + S_D\vec{G}(t_D)\right)^T \left(\Delta\vec{H} + S_D\vec{G}(t_D)\right)$$

$$= \left(\Delta\vec{H}^T + S_D\vec{G}^T(t_D)\right)\left(\Delta\vec{H} + S_D\vec{G}(t_D)\right)$$

$$= \Delta\vec{H}^T\Delta\vec{H} + S_D\Delta\vec{H}^T\vec{G}(t_D) + S_D\vec{G}^T(t_D)\Delta\vec{H} + S_D^2\vec{G}^T(t_D)\vec{G}(t_D) \tag{8}$$

We seek to find t_D and S_D such that $\Phi(S_D, t_D)$ is minimum. We first require

$$\frac{\partial\Phi(S_D, t_D)}{\partial S_D} = 0 \Longrightarrow 2\Delta\vec{H}^T\vec{G}(t_D) + 2S_D\vec{G}^T(t_D)\vec{G}(t_D) = 0$$

$$\Longrightarrow {S^*}_D = -\frac{\Delta\vec{H}^T\vec{G}(t_D)}{\vec{G}^T(t_D)\vec{G}(t_D)}$$

$$\Longrightarrow {S^*}_D = -\frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|^2} \tag{9}$$

Substituting equation (9) into (8) we get:

$$\Phi(S_D, t_D) = \Delta\vec{H}^T\Delta\vec{H} - \frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|^2}\Delta\vec{H}^T\vec{G}(t_D) - \frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|^2}\vec{G}^T(t_D)\Delta\vec{H} + \left[\frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|}\right]^2 \tag{10}$$

The value of t_D at which $\Phi(S_D, t_D)$ is minimum is unaffected by the constant $\Delta\vec{H}^T\Delta\vec{H}$. Therefore,

$${t^*}_D = argmin\big(\Phi(S_D, t_D)\big)$$

$$= argmin\left(-\frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|^2}\vec{G}^T(t_D)\Delta\vec{H} - \frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|^2}\vec{G}^T(t_D)\Delta\vec{H} + \left[\frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|}\right]^2\right) \quad (11)$$

Now we develop the last term in the above equation as follows:

$$\left[\frac{\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|}\right]^2 = \frac{\Delta\vec{H}^T\vec{G}(t_D)\Delta\vec{H}^T\vec{G}(t_D)}{\left\|\vec{G}(t_D)\right\|^2} = \frac{\Delta\vec{H}^T\vec{G}(t_D)\vec{G}^T(t_D)\Delta\vec{H}}{\left\|\vec{G}(t_D)\right\|^2}$$

Hence,

$$t^*{}_D = argmin\left(-\frac{\Delta\vec{H}^T\vec{G}(t_D)\vec{G}^T(t_D)\Delta\vec{H}}{\left\|\vec{G}(t_D)\right\|^2}\right) = argmax\left(\frac{\Delta\vec{H}^T\vec{G}(t_D)\vec{G}^T(t_D)\Delta\vec{H}}{\left\|\vec{G}(t_D)\right\|^2}\right) \quad (12)$$

where the argument of the maximum function of equation (12) is hereafter denoted by $\boldsymbol{B}$.

Algorithm

Step 1: introduce a damage of size S_D at location x_D

Step 2: use the model to determine the vector of incident wave $\vec{H}_I$

Step 3: collect $\overrightarrow{D}(t) = \vec{H}(t) + \vec{n}$(t)

Step 4: Compute $\Delta\vec{H} = \overrightarrow{D}(t) - \vec{H}_I$

Step 5: Select Δt_D and use equation (4) to evaluate $\vec{G}(l\Delta t_D)$ for $l = 0, \ldots, L$

Step 6: obtain $t^*{}_D$ using equation (12)

Step 7: obtain $S^*{}_D$ using equation (9)

Step 8: go back to step 1 and vary the damage parameters and repeat steps 2 through 7

IV.2.1.1. Non convexity of the optimization problem given by equation (12): case of the single damage

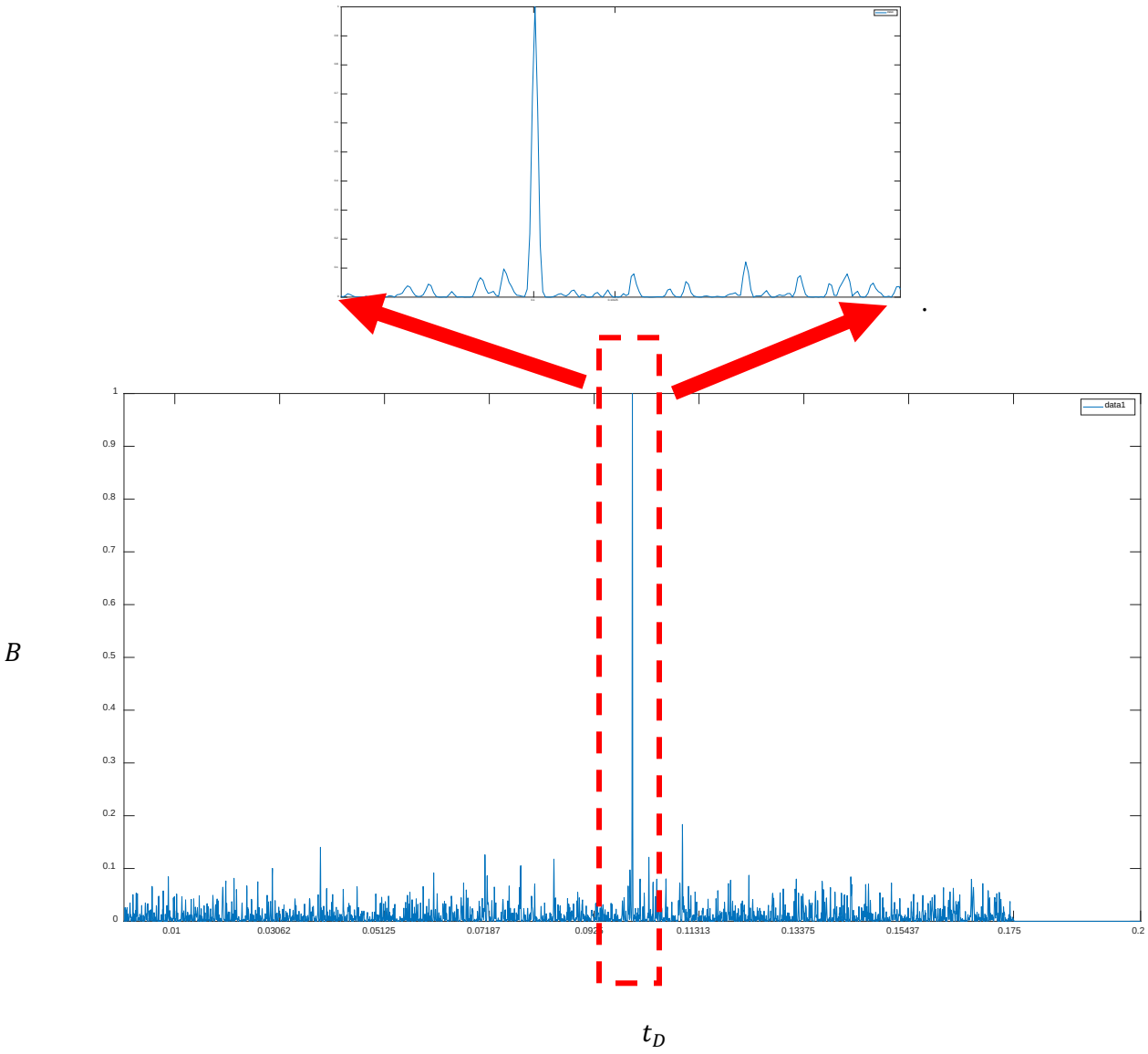

Figure3: *the lower figure represents the objective function B versus the time t_D at which the incident wave reaches the damage. The upper figure illustrates a zoomed version of B near the peak which clearly displays the non-convex nature of the optimization problem given by equation (12).*

The parameters used to generate the results in figure 1 are: $t \in [0,2]s$; $t_D \in [0,0.2]s$; $\varepsilon = 0.0001$; $C = 1000m/s$; the noise is white and Gaussian with mean 0 and $\sigma = 80$ and signal to noise ratio (SNR)= $log(1/\sigma^2\varepsilon^2 2\pi) = 3.4dB$. The power function $\boldsymbol{B}$ versus the time t_D is highly nonlinear and non-

convex (see figure3). Therefore, classical gradient search methods are inefficient and there is no guarantee they can achieve global maximum [6].

Solution of problem (12)

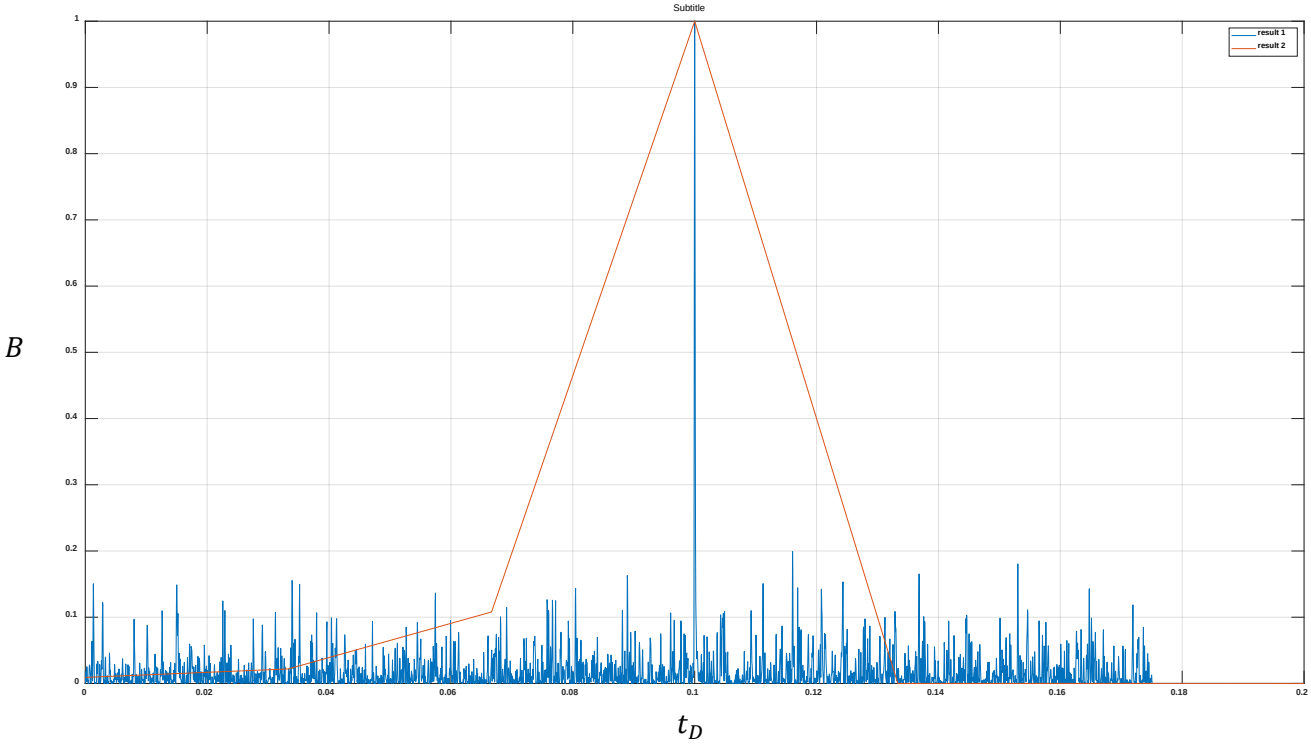

Figure4: ***B*** *versus* t_D *for the case* $P = 5000$. *The red figure shows B versus* t_D *for the case* $P = 7$.

In this project, we search for global maximum by discretizing the time t_D into P steps. To elucidate the effect of discretization, we consider P in the range 7 to 5000. For brevity, we show the results for $P = 7$ and $P = 5000$ only. These results are shown in figure4. It is clear that the coarse discretization cuts off all the side lobes and provides very poor resolution. Using the diffraction limit principle we obtain ($\sim 0.0355s$ which translates $35.5m$). As a result, two or more damages within a distance of 35.5m cannot be individually distinguished. On the other hand when $P = 5000$, a high resolution is achieved ($\sim 0.0002s$ which translates $0.2m$). Effectively the $P = 5000$ gives a resolution that about 180 times better than $P = 7$. Both discretization lead to an accurate estimates of the actual location and size of the damage although the $P = 5000$ produces slightly better size estimate. However, we stress that the coarse

discretization has completely wiped out all the side lobes. Therefore, secondary damages (with $\sim 20\%$ of the power of the main lobe or smaller) cannot be identified by the coarse discretization.

IV.2.1.2. Non convexity of the optimization problem given by equation (12): case of multiple damages

The equation (12) remains the same with more than one damage [4][5]

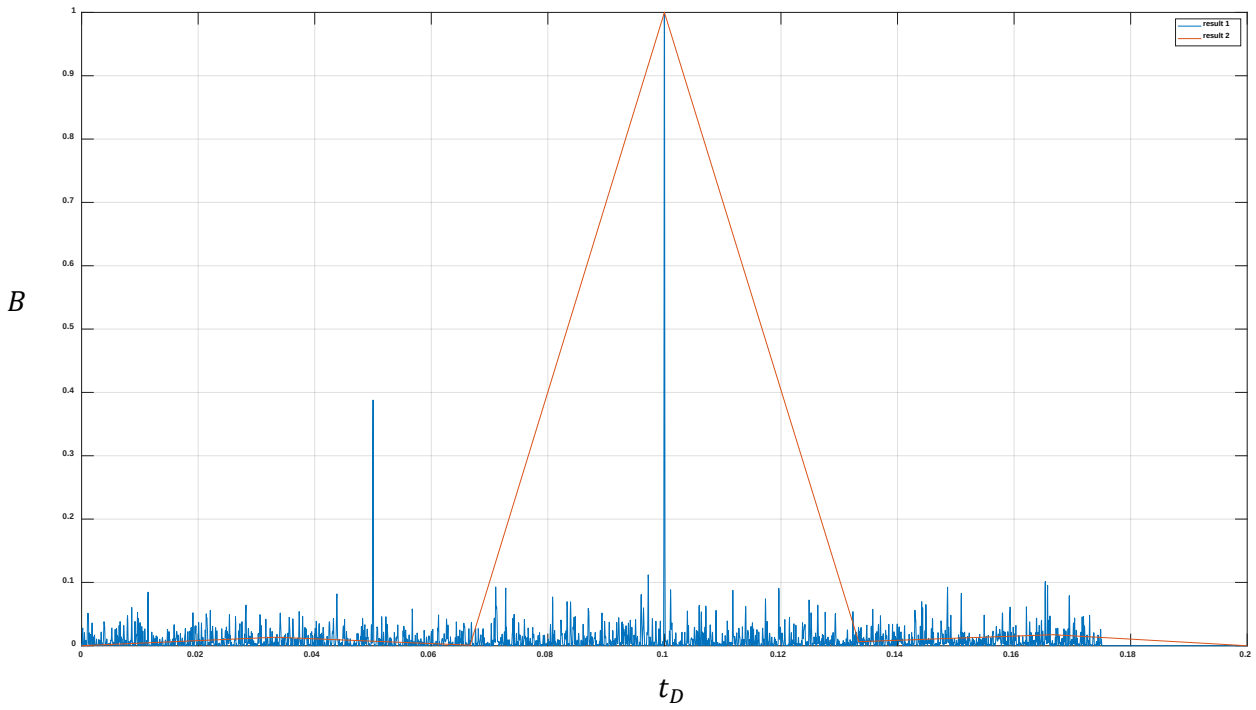

Figure5: *two damages for the case* $P = 5000$*(blue color) and the case* $P = 7$*(red color).*

We have generalized the single case damage to multiple (two in this case) damages. Both damages are identified for the fine discretization because the distance between the two damages (50m) is significantly larger than the resolution limit (0.2m). However, the two damages cannot be resolved by the coarse discretization solution.

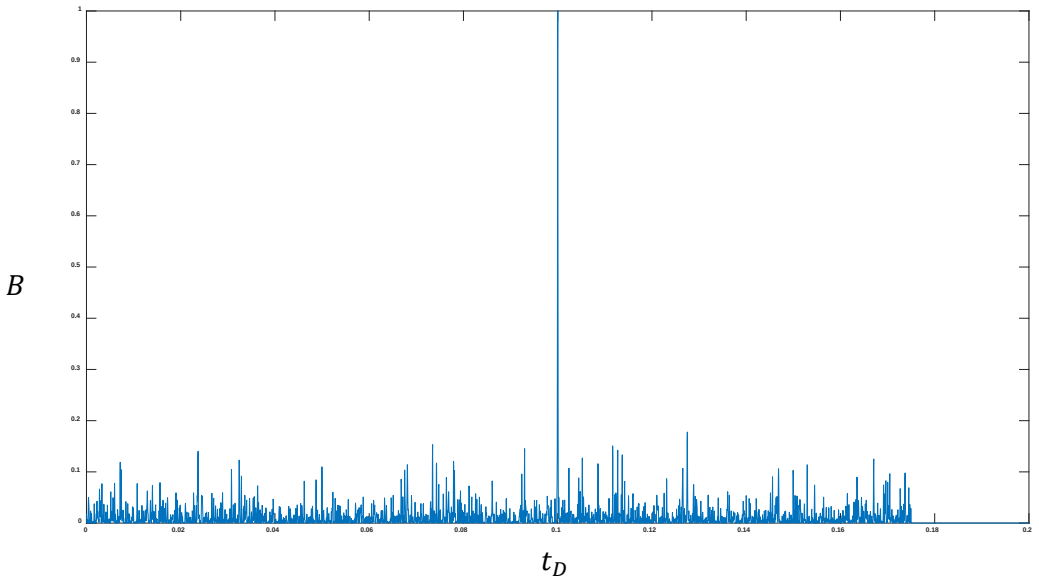

Figure6: *multiple (two) damage case where the distance (0.1m) separating both damages is less than the resolution limit (0.2m).*

Figure6 shows the case of two damages that are $0.1m$ apart using fine discretization. It is clear that the 2 damages cannot be resolved. Put simply the two damages appear as a single damage [4][7]. This is not surprising given that the separation between the damages is smaller than the diffraction limit.

V. Conclusion

The key conclusions from this project are:

1. PCA is a powerful method for clustering and optimizing sensor locations to identify damages within a structure. It captures essential information from the data and reduces the number of features, improving clustering algorithms and identifies redundant sensors.
2. PCA is efficient in classifying damages from non-damages by extracting significant variations in the data.
3. Maximum Likelihood Estimation (MLE) accurately estimates the location and size of damages in structures.
4. The MLE method is a high resolution technique since it achieves the diffraction limit.

Reference

1. Keramat, A., Karney, B.W., Ghidaoui, M.S., and Wang, X. (2021). "Transient-based leak detection in the frequency domain considering fluid-structure interaction and viscoelasticity." *Mechanical Systems and Signal Processing*, 153, 107500. https://doi.org/10.1016/j.ymssp.2020.107500
2. Blåsten, E., Zouari, F., Louati, M., and Ghidaoui, M.S. (2019). "Blockage detection in networks: the area reconstruction method." *Mathematics in Engineering*: Inverse problems in imaging and engineering science, 1(4), 849-880. https://doi.org/10.3934/mine.2019.4.849
3. Wang, X., Lin, J., Keramat, A., Ghidaoui, M.S., Meniconi, S., and Brunone, B. (2019). "Matched-field processing for leak localization in a viscoelastic pipe: an experimental study." *Mechanical Systems and Signal Processing*, 124, 459-478. https://doi.org/10.1016/j.ymssp.2019.02.004
4. Wang, X., Ghidaoui, M.S., and Lin, J. (2019). "Identification of multiple leaks in pipeline III: Experimental results." *Mechanical Systems and Signal Processing*, 130, 395-408. https://doi.org/10.1016/j.ymssp.2019.05.015
5. Wang, X., and Ghidaoui, M. S. (2019). "Identification of multiple leaks in pipeline II: Iterative beamforming and leak number estimation." *Mechanical Systems and Signal Processing*, 119, 346-362. https://doi.org/10.1016/j.ymssp.2018.09.020
6. Wang, X., and Ghidaoui, M.S. (2018). "Pipeline leak detection using the matched-field processing method." *Journal of Hydraulic Engineering*, ASCE, 144(6), 04018030. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001476
7. Wang, X., and Ghidaoui, M.S. (2018). "Identification of multiple leaks in pipeline: Linearized model, maximum likelihood, and super-resolution localization." *Mechanical Systems and Signal Processing*, 107, 529-548. https://doi.org/10.1016/j.ymssp.2018.01.042
8. Yao Yuan (2023). Geometric and Topological Data Reduction: A Mathematical Introduction to Data Science. 232 URL:https://yao-lab.github.io/bookdatasci/
9. J. L. Beck, and L. S. Katafygiotis. (1998). "Updating models and their uncertainties. I: Bayesian statistical framework" Journal of Engineering Mechanics,