

Automated Denoising Technique for Random Input Signals using Empirical Mode Decomposition (EMD)-Stabilization Diagram

M. Danial A. Hasan^{1,2*}, *Z. A. B. Ahmad*², *M. Salman Leong*¹ and *L. M. Hee*¹

¹Institute of Noise and Vibration, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

² School of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Abstract. The present paper deals with the novel approach of filtering technique using hybrid of empirical mode decomposition technique with stabilization diagram, that autonomously implemented within Matlab. Noise or unwanted signal is always present in the data and a bad signal-to-noise can cause a severe error in modal parameter extraction. With the recent developments of automated procedures without user interaction for the operational modal analysis (OMA), the corrupted input signals turn out to be a big issue in obtaining reliable results of automated modal parameter identification. The appearance of noise or unwanted modes due to environmental effects could affect the actual structural modes selection. There is a significant issue regarding “noise” (or spurious) modes and eliminating them from the raw signal remains to be solved and requires a lot of interaction with an expert user. In the parametric modal analysis, oversizing of a modal model is usually performed to minimize the bias on modal estimates by getting all physical modes in the frequency range of interest and help to obtain a good model fit to the data. However, this will introduce noise modes. Thus, authors take advantage of tools, such as the stabilization diagram, to illustrate, and decide, if a mode is physical or not. This selection is not a trivial task, but it may be difficult and time consuming depending on the quality of data, the performance of the estimator and the experience of the user. Since the extensive interaction between tools and user is inappropriate for monitoring purposes, image clustering tool is introduced to separate the physical poles from the others with short response time and low computational efforts compared to the available clustering algorithm. Meanwhile, Empirical mode decomposition (EMD) is then introduced to break down a signal into various components without leaving the time domain and purposely used for filtering. These are a great combination as well as an effective procedure in producing a good input signal that free from unwanted modes which can cause disruptive decision making for the actual modes selection.

1 Introduction

Noise or unwanted signals are always present in the data and a bad signal-to-noise can cause a severe error in modal parameter extraction. With the recent developments of automated procedures without user interaction for the operational modal analysis (OMA), the corrupted input signals turn out to be a big issue in obtaining reliable results of automated modal parameter identification. The appearance of noise or unwanted modes due to environmental effects could affect the selection of the actual structural modes.

Several examples of filtering technique performed in OMA have been reported in the technical literature and presented herein. Yang et al. proposed using a band-pass filter based on Hilbert-Huang transform (HHT) in modal identification method via measured free vibration time histories

[1,2]. Yu and Ren proposed an EMD based SSI modal identification method, which reduced the postprocessing work in SSI algorithm (picking the stable poles) and using a band-pass filter to treat the closely spaced modes [3]. Among the methods based on conventional signal processing, the so-called time domain filtering method was a tracking procedure based on the application of a band-pass filter to the system response in order to separate the individual modes in the spectrum [4].

Abovementioned literature indicates that the researchers often used classical filtering technique based on Fourier transforms, particularly band-pass filter rather than low pass and high pass filter in order to remove unwanted signal and also for clear modal parameter extraction. It is possible to remove unwanted from its raw signal, but more attention should be paid to ensure all the necessary information

* Corresponding author: muhd_danial200@yahoo.com

from this signal are not affected. Therefore, it is still worthy to investigate this problem.

In the parametric modal analysis, oversizing of a modal model is usually performed to minimize the bias on modal estimates by getting all physical modes in the frequency range of interest and help to obtain a good model fit to the data [5]. However, this will introduce noise modes. There is a significant issue regarding “noise” (or spurious) modes and eliminating them from the raw signal remains to be solved and requires a lot of interaction with an expert user.

Thus, this paper will focus on automated OMA that can ensure an effective identification of physical modes and removing unwanted signal with short response time and low computational efforts.

2 Automated filtering technique

The automatic filtering technique in OMA based on parametric methods consists of the following steps:

- (1) Measure the responses of the structure and estimate the modal parameters using a high model order, n modes.
- (2) Construct Stabilization diagram by estimating poles with an increasing model order and illustrate, and decide, either mode is physical or computational modes
- (3) Classify the n modes in physical and computational modes using a clustering algorithm.
- (4) The unstable mode is filtered out involving empirical mode decomposition (EMD) method and left only the useful signal.

The acceleration time series of the first data set from the Heritage Court Tower building has been processed by the automatic versions of the SSI-COV. The frequency range of 0 to 5 Hz already contains a considerable number of resonant frequencies.

2.1. Stochastic subspace technique (SSI)

The Stochastic subspace identification (SSI) has been well-known over the past decade, mainly due to user-friendly executions [6].

There are two main types of SSI, that is, the correlation driven, COV-SSI and the data-driven, DATA-SSI. The DATA-SSI is an approach where the estimation of the correlation functions is integrated into the identification algorithm. The COV-SSI, however, is very alike to the ERA [7] and identifies a stochastic state space model from output-only data [5].

This paper only concerns on correlation driven SSI and details derivation about this method are described below.

The first step of this method is to calculate the output correlations (Equation 1). $[R_i]$ denotes the unbiased estimate of the correlation matrix at time lag i based on a finite number of data:

$$[R_i] = \frac{1}{N-i} [Y_{(1:N-i)}][Y_{(i:N)}]^T \quad (1)$$

Where $[Y_{(1:N-i)}]$ is the data matrix Y with the last block rows i removed and $[Y_{(i:N)}]^T$ is the transpose data matrix with the first block rows i removed. Therefore each $[R_i]$ matrix gets dimensions $l \times l$. The estimated correlations at different time lags (Equation 2) are then gathered into a matrix called the block Toeplitz matrix:

$$[T_{1|i}] = \begin{bmatrix} [R_i] & [R_{i-1}] & \cdots & [R_1] \\ [R_{i+1}] & [R_i] & \cdots & [R_i] \\ \vdots & \vdots & \ddots & \vdots \\ [R_{2i-1}] & [R_{2i-2}] & \cdots & [R_i] \end{bmatrix} \quad (2)$$

The Toeplitz matrix contains $i \times i$ $[R_i]$ matrices and is therefore of dimensions $li \times li$. For the identification of the modal parameters of a system of order n , the Toeplitz matrix needs to be $n \times n$. Therefore, the following need to be true for the number of block rows i :

$$li \geq n, \quad i_{min} = x \frac{n_{max}}{l} \quad (3)$$

Where n is the system order. Normally, the system order is unknown and the following (Equation 3) should be fulfilled. For complicated structures the number of block rows, i should be higher than this minimum criterion for better results. The magnitude x of this relies on the problem and should be chosen as an input by the user. In this case, the magnitude of block rows, x was set equal to 2, meanwhile, for maximum system order, n was defined as 50 modes. This value needs need to be selected wisely. If the maximum order is chosen to be smaller than the correct system order you will not get any correct results. If you, however, assume this value too high you will get too many nonphysical modes and it will be harder to derive which modes that are the actual correct physical modes, in addition, the computational time greatly increases.

Then the singular value decomposition (SVD) of the block Toeplitz is calculated. It gives the unitary

matrices $[U]$ and $[V]$, with dimensions $li*n$ and $n*li$ respectively. And the positive singular values (which equals its rank) in descending order in the diagonal matrix $[\Sigma]$ [8].

$$[T_{1|i}] = [U_1][\Sigma_1][V_1]^T \quad (4)$$

To extract the dynamic response, you need to obtain the state matrix $[A]$. This is done for each order from 1 to n_{max} . The first thing needed is to find the observability matrix $[O_i]$ and the reversed controllability matrix $[\Gamma_i]$ which is found by the factorization of $[T_{1|i}]$.

$$[T_{1|i}] = [O_i][\Gamma_i] \quad (5)$$

Where the observability matrix $[O_i]$ with dimensions $li*n$ is given by:

$$[O_i] = \begin{bmatrix} [C] \\ [C][A] \\ \vdots \\ [C][A]^{i-1} \end{bmatrix} \quad (6)$$

And the reversed controllability matrix $[\Gamma_i]$ with dimensions $n*li$ is given by:

$$[\Gamma_i] = [[A]^{i-1}[G] \quad \dots \quad [A][G] \quad [G]] \quad (7)$$

The SVD of $[T_{1|i}]$ is already computed in Equation 4 and we can use the result to find $[O_i]$ and $[\Gamma_i]$. By splitting the SVD into two parts and using the identity matrix $[I]$:

$$[O_i] = [U_1][\Sigma_1]^{1/2}[I]^T \quad (8)$$

$$[\Gamma_i] = [I]^{-1}[\Sigma_1]^{1/2}[V_1]^T \quad (9)$$

Now that we have obtained $[O_i]$ and $[\Gamma_i]$ we can find the output influence matrix $[C]$ and the next state-output covariance matrix $[G]$. $[C]$ is obtained from the first l rows of $[O_i]$ as seen in Equation 6. And in the same way $[G]$ is obtained from the last l columns of $[\Gamma_i]$ as seen in Equation 7. The normal Toeplitz matrix, which yields:

$$[T_{2|i+1}] = [O_i][A][\Gamma_i] \quad (10)$$

Introducing Equation 8 and Equation 9 into Equation 10 and then solving for $[A]$ gives us the $n*n$ state matrix:

$$[A] = [\Sigma_1]^{-1/2}[U_1]^T[T_{2|i+1}][V_1][\Sigma_1]^{-1/2} \quad (11)$$

Solving the eigenvalue problem for $[A]$ yields the diagonal matrix $[M]$ which contains the eigenvalues μ and the eigenvectors $[\Psi]$ [9]:

$$[A] = [\Psi][M][\Psi]^{-1} \quad (12)$$

The mode shapes of the system $[\Phi]$ are obtained from $[\Psi]$ and $[C]$:

$$[\Phi] = [C][\Psi] \quad (13)$$

The rest of the modal parameters are obtained from the eigenvalues μ . They are found in the diagonal matrix $[M]$. For each mode, μ_m are found, they are obtained in discrete time and need to be converted to continuous time which yields:

$$\lambda_m = \frac{\ln(\mu_m)}{\Delta t} \quad (14)$$

And we obtain the complex λ_m which contains the continuous time eigenvalues of each mode for the current order. It can be used to find the natural frequencies (ω_n), damped modal frequencies (ω_d) and the damping ratio (ζ) for the r -th mode as shown in Equation 15, Equation 16 and Equation 17:

$$\omega_{n,r} = |\lambda_{m,r}| \quad (15)$$

$$\omega_{d,r} = \text{Im}(\lambda_{m,r}) \quad (16)$$

$$\zeta_r = -\frac{\text{Re}(\lambda_{m,r})}{|\lambda_{m,r}|} \quad (17)$$

When the modal parameters are found for the current order, the process of finding the state matrix and the modal parameters are repeated for each order up until n_{max} . Then all the parameters are plotted in a stabilization diagram, which will be discussed in the next section.

2.2 Stabilization Diagram

When the modal parameters are attained, the stabilization diagram can be constructed. This tool is one of the most common methods to distinguish physical poles from others, which is performed by estimating poles with an increasing model order. Since

the model of the system is often oversized, thus, the plot will contain noise modes and mathematical modes. The noise modes are caused by physical reasons, while the mathematical modes are generated to ensure the mathematical description of the measured data. Theoretically, the physical poles should stabilize and can be found by an alignment of stable poles, whereas the computational or mathematical poles are scattered, showing the criterion of the unstable poles. This is based on the comparison of the poles associated with a given model order with those attained from a one-order lower model [5].

The natural frequencies and damping ratio of poles from two orders are compared [10]:

$$\frac{|f(n-1) - f(n)|}{f(n-1)} < x \quad (18)$$

$$\frac{|\zeta(n-1) - \zeta(n)|}{\zeta(n-1)} < y \quad (19)$$

Only the poles that fulfill a stabilization criterion defined by the user (x and y) are labeled as stable respected the following limits for variations between models of consecutive orders: natural frequency variation $< 1\%$ [11]; modal damping ratio variation $< 5\%$. These limits proved to permit the clear distinction of vertical alignments of stable modes that represent the modes of vibration.

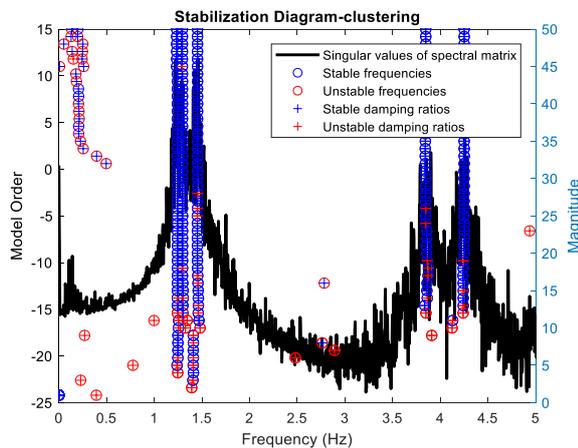


Fig. 1. Singular values of the spectral matrix with stabilization diagram for first data sets of the HCT building case.

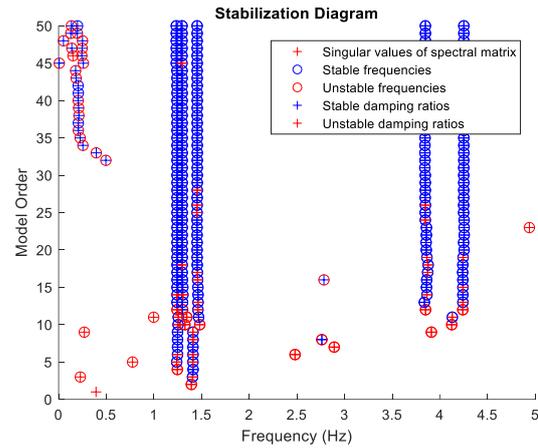


Fig. 2. Stabilization diagram for first data sets of the HCT building case.

2.3 Image Clustering

The poles of the stabilization diagrams that represent the same physical modes were easily grouped by using image clustering. The process of image clustering will be explained in this section.

2.3.1 Input image

The stabilization diagram was cut and displayed separately into every frequency according to 0.01 interval, $5\text{Hz} / 0.01 = 500$ images. So, every image represents the frequency of 0.01 Hz.

2.3.2 Image feature extraction

In this study, six standardized image features in Matlab software were used to extract the image features of each stabilization diagrams. These features specifically represent the characteristics of each parameter (natural frequencies, damping ratios) for different conditions either stable or unstable. Table 1 summarises all the image features that were extracted in this study and their characteristic values.

Table 1. Image Features.

No	Feature	Characteristic value
1	Features from FAST	Corner Points
2	Minimum Eigenvalues	Corner Points
3	Harris Stephens	Corner Points
4	Speeded-Up Robust Features	Blob
5	Binary Robust Invariant Scalable Keypoints	Multiscale Corner

6	Maximally Stable External Regions	Regions
---	-----------------------------------	---------

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (19)$$

This algorithm was configured to group all the poles of the stabilization diagram, stable and not stable, by using image features extraction, as shown in Figure 3.

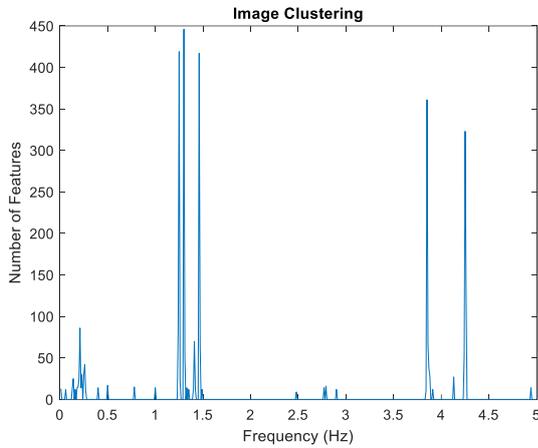


Fig. 3. Image clustering plot by using one of the images features extraction with all the poles of the stabilization diagram that presented in Figure 2.

2.4 Empirical mode decomposition (EMD) method

Empirical mode decomposition (EMD) is an effective approach to decompose a nonstationary signal into various components, known as a series of intrinsic mode functions (IMFs) and a residual, without leaving the time domain [12]. The basic steps of EMD are as follows.

- a) Sifting process: Find all the extremes of $x(t)$. Define the upper and lower envelope of $x(t)$, which can be obtained through connecting all the local maxima and local minima with cubic a spline curve, respectively. Calculate the mean value of the two envelopes. The difference between the $x(t)$ and the mean value is designated as $h1(t)$.
- b) Checking process: Check whether $h1(t)$ satisfies the two conditions of IMF. If $h1(t)$ is an IMF, then take $c1(t) = h1(t)$ as the first IMF of $x(t)$. Otherwise, let $x(t) = h1(t)$ and repeat the sifting process until the first IMF is found out.
- c) Looping process: Define $r1(t) = x(t) - c1(t)$ as the new signal and repeat (a), (b) until $rn(t)$ is smaller than a predetermined threshold. Then, the original signal $x(t)$ can be rewritten as

The acceleration time series signal of the first data set from the Heritage Court Tower building has been breaking down into 13 IMF as shown in Figure 4 below. This figure shows the all the IMFs of the acceleration response signal. The response signals were decomposed into time-domain and of the same length as the original signal allows for varying frequency in time to be preserved, particularly in decreasing order.

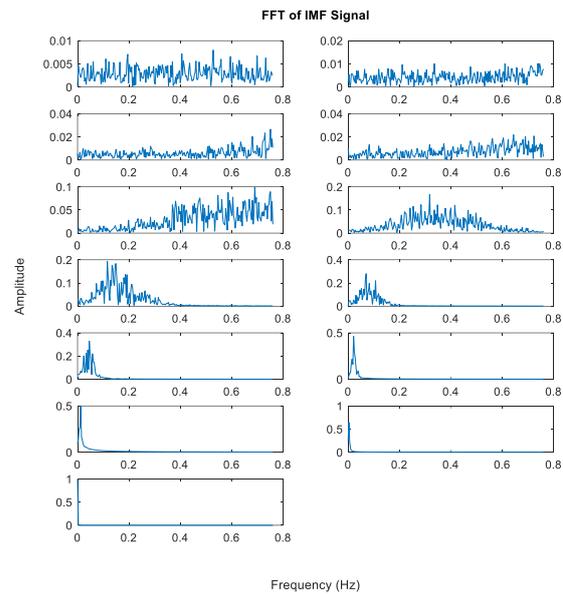


Fig. 4. The decomposition of the acceleration time series signal of the first data set from the Heritage Court Tower building into 13 IMFs.

3 Results and discussion

The selection of the weak modes of the system that autonomously implemented involved Matlab command - *find* and the threshold in order to discriminate the actual modes and left only the weak ones. The approach is based on excluding the dominant modes with a higher value than the threshold. The threshold is determined by the half of maximum peak in image clustering plot. Meanwhile, for the selection of useful IMF and unwanted IMFS are based on the frequency of unwanted mode. The peak of a certain frequency in IMFs can be determined by using peak picking (PP) method. Thus, the useful IMF will reconstruct back to become an input signal.

The result of the automated filtering technique using image clustering and EMD is characterized in Figure 5.

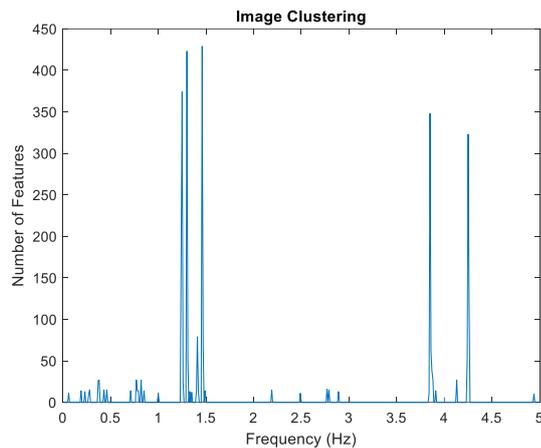


Fig. 5. Image clustering plot by using one of the images features extraction with all the poles of the stabilization diagram after went through a filtering process to remove noise or unwanted mode.

From the image clustering plot above we can see a peak at frequency 0.21 Hz had been significantly reached a low point after went through filtering process when compared to image clustering plot in Figure 3, where a shoot up peak clearly seemed at a similar frequency.

Thus, the results of this study indicate that, the implementation of image clustering for identification physical modes and computational modes in stabilization diagram and EMD as a tool for filtering purpose are a great combination as well as an effective procedure in producing a good input signal that free from unwanted modes which can cause disruptive decision making for the actual modes selection. The use of image clustering rather than normal clustering algorithm was capable to reduce processing time and low computational efforts needed. In the case of application of advanced clustering algorithm, the number of parameters had to be set at startup which consumes time for calibrations in order to allow reliable identification of structural modes. Meanwhile, EMD is a suitable method for filtering based on the frequency of peak because it decomposes a signal into various components without leaving the time domain and according to decrease order of frequency. Thus, the unwanted frequency peak can easily be removed out from the original input signal. However, this method has drawbacks in separating closely space modes.

This paper has given an account of and the reasons for the widespread use of stabilization diagram, image clustering and EMD for automated filtering technique that involving random input signal. These findings provide the following insights for future research in term of eliminating noise or unwanted modes from closely spaced modes. Even though, image clustering

was capable to identify closely spaced modes, as appeared in Figure 1 for the all closely spaced modes of the Heritage court building from 1.2 to 1.5 Hz. But, more attention should be paid to select the appropriate tool or improving existing EMD in eliminating this closely spaced noise or unwanted modes.

The authors would like to extend their greatest gratitude to the Institute of Noise and Vibration UTM for funding the current study under the Higher Institution Centre of Excellence (HiCoE) Grant Scheme (R.K130000.7843.4J227). Additional funding for this research came from the UTM Research University Grant (Q.K130000.2543.11H36) and the Fundamental Research Grant Scheme (R.K130000.7840.4F653) from The Ministry of Higher Education, Malaysia.

References

- [1] JN. Yang, Y. Lei, S. Pan, N. Huang, *Earthq. Eng. Struct. Dyn.*, **32**, 1443–1467 (2003)
- [2] JN. Yang, Y. Lei, S. Pan, N. Huang, *Earthq. Eng. Struct. Dyn.*, **32**, 1533–1554 (2003)
- [3] D-J. Yu, W-X. Ren, *Eng. Struct.*, **27**, 1741–1751 (2005)
- [4] H. Guan, VM. Karbhari, CS. Sikorski, *IOMAC proceedings*, **1**, 577–584 (2005)
- [5] C. Rainieri, G. Fabbrocino, *Operational Modal Analysis of Civil Engineering Structures*, (2014)
- [6] R. Bricker, C. Venture, *Introduction to operational modal analysis*, (2015).
- [7] B. Peeters, G. De Roeck, *Mech. Syst. Signal Process.*, **13**, 855–878 (1999)
- [8] ME. Wall, A. Rechtsteiner, LM. Rocha, *A practical approach to microarray data analysis*, 91–109 (2003)
- [9] AK. Chopra, *Dynamics of Structures. Theory and Applications to Earthquake Engineering*, **4**, (2011)
- [10] SA. Schanke, *Operational Modal Analysis of Large Bridges*, (2015)
- [11] F. Magalhães, *Operational modal analysis for testing and monitoring of bridges and special structures*, Thesis (Ph.D), (2010)
- [12] NE. Huang, Z. Shen, SR. Long, MC. Wu, HH. Shih, Q. Zheng, et al., *Proc Math Phys Eng Sci.*, **454**, 903–995 (1998)