

Development of Refined Data-Driven Stochastic Subspace System Identification for Buildings and Bridges

Chia-Ming Chang^(\boxtimes) and Yi-Ji Chuang

National Taiwan University, Taipei 10617, Taiwan changcm@ntu.edu.tw

Abstract. Frequent large-scale earthquakes, climate changes, manmade hazards, and the duration of the service are the possible origins of structural damage in Taiwan. To detect the changing features and damage states of the structures, the demand for understanding the unknown system models of the operating structures has risen. The accuracy of structural health monitoring has become a significant issue. Therefore, four kinds of refined data-driven stochastic subspace system identification (SSI-DATA) methods, namely the mode-by-mode methods, are proposed in this research. Because the mode-by-mode methods only extract a single mode per iteration, the "mode elimination" and "signal reconstruction" steps are added to the traditional SSI-DATA. The mode elimination is realized by removing the singular components that have been exploited in the identified mode. Meanwhile, the signal reconstruction employs a similar approach used in the singular spectrum analysis after the Hankel matrix is regenerated with the removal of identified modes. Moreover, the effective projection operations and modification of the singular value decomposition process are employed in the refined methods. A unified analysis procedure is also introduced to automatically extract all the concerned modes one by one using the methods, while the errors between the reference frequencies and identified frequencies and the calculated frequency resolutions are the criteria for selecting modes. To verify the proposed methods, cases of a simulated eight-story frame and the actual operating bridge structure are studied using the proposed system identification methods. Consequently, the identification results show that the refined methods can yield slightly more accurate modal parameters of the structures. Moreover, the computational time of the second, third, and fourth methods is much less than the traditional SSI-DATA.

Keywords: Structural Health Monitoring · Output-Only System Identification · Mode Elimination · Signal Reconstruction · Data-Driven Stochastic Subspace System Identification

1 Introduction

Taiwan is located in the Pacific seismic zone, which has caused frequent earthquakes and even disasters on the island. Therefore, aside from the aseismic capability of the structures, damage detections and seismic retrofit after the catastrophes are also important issues. In recent years, various technologies have thrived with the ever-changing monitoring instruments and the development of structural health monitoring (SHM) methods, enabling the output responses to be collected in operational structures. These responses can be analyzed to identify structural characteristics or predict the structural behavior subsequently; thus, the accuracy of the system identification results is significant. The system identification also prompts more attention to research on related applications.

System identification (SI) is a methodology to obtain structural modal parameters (i.e., natural frequencies, damping ratios, mode shapes) and the numerical dynamic model of the structure through experiments, which can serve as the inverse problem in structural dynamics analysis. Most available methods require the input and output information; however, under most circumstances, excite known input signals to actual structures is impractical and expensive. Meanwhile, acquiring environmental excitation or ambient vibrations as the input is impossible for input-output system identification methods. Thus, the output-only system identification is the most commonly used method nowadays. Moreover, the stochastic subspace system identification, e.g., the data-driven stochastic subspace system identification (SSI-DATA), has been fully developed and applied in many studies $[1-3]$ $[1-3]$. This identification strategy has also been applied to a number of historical structures (i.e., buildings and bridges) before [\[4](#page-10-2)[–8\]](#page-10-3).

Still, some drawbacks exist in the data-driven stochastic subspace system identification method. For example, this method needs to input several user-defined parameters (e.g., model order, number of block rows in the Hankel matrix, and length of the acquired signals used in the identification) when conducting the SSI method. With the use of the defined parameters, a stabilization diagram, which is a graph of the relationship between the identified modes and the corresponding modal orders of the structural system, can be illustrated. Users can select and confirm the order of the system from this diagram and then extract the selected modal parameters. However, if the parameters are chosen wrongly, spurious modes are generated, leading users to misjudge the modal parameters [\[9\]](#page-10-4). In addition, a fixed number of block rows of the Hankel matrix, which is used in the general SSI calculation, sometimes challenges the identification of higher modes.

This research aims to develop four refined methods of data-driven stochastic subspace system identification and obtain more accurate modal parameters by orderly extracting the modes. All these methods are developed in accordance with the original structure of SSI-DATA but both "mode elimination" and "signal reconstruction" steps are added to achieve extracting a single mode at a time. The mode elimination is realized by removing the singular components that have been exploited in the identified mode, while the signal reconstruction employs a similar approach used in the singular spectrum analysis after the Hankel matrix is regenerated with the removal of identified modes. Moreover, the effective projection operations and modification of the singular value decomposition process are employed in the refined methods. A unified analysis procedure is also introduced to automatically extract all the modes of interest one by one using the methods, while the errors between the reference frequencies and identified frequencies and the calculated frequency resolutions are the criteria for selecting modes. Finally, the proposed methods are applied to a simulated eight-story frame and the actual operating bridge structure for verification. The identification results show that the refined methods can yield slightly more accurate modal parameters of the structures with reduced computational time.

2 Mode-by-Mode SSI-DATA Methods

Before all proposed methods are introduced, the traditional data-driven stochastic subspace system identification method is briefly reviewed and illustrated in Fig. [1.](#page-2-0) As seen in this figure, the acquired responses from a structure are stacked into a Hankel matrix, and this matrix is divided into past and future submatrices, such as Y_p and Y_f , respectively. Then, the subspace of this Hankel matrix is obtained through the orthogonal matrix projection. By the singular value decomposition, the extended observability matrix can be calculated, and this matrix contains the information about the modal free vibration responses. Using the state-space realization and mode extraction, the modal properties, i.e., natural frequencies, damping ratios, and mode shapes, are derived.

Fig. 1. Flowchart of traditional data-driven stochastic subspace system identification method.

As for the proposed mode-by-mode methods (see Fig. [2\)](#page-3-0), the first step is to modify the construction of the Hankel matrix with a variable matrix size. For the first round of iteration, the Hankel matrix size *L* is defined by

$$
L = \lceil \frac{f_s}{f_{n1}} + 1 \rceil \times n_y \tag{1}
$$

Fig. 2. Architecture of proposed mode-by-mode methods.

where f_s is the sampling rate, f_{n1} is the first mode's frequency, and n_v denotes the number of output channels.

The signals need to be reconstructed before forming a new Hankel matrix per iteration. The signal reconstruction method is similar to the reconstruction stage in the singular spectrum analysis. Note that the signal can be directly reconstructed from the projection matrix **O** with the contributions of the previously extracted mode eliminated. Although the projection matrix **O** has been presented in a different subspace, the reconstructed signals still contain the completely dynamic characteristics of the structure to be identified. This reconstruction method averages the derived signals through the form of a Hankel matrix. Moreover, the projection matrix **O** with the previously extracted mode eliminated can also be directly exploited without the signal reconstruction to obtain the next mode.

When evaluating whether the identified modal frequency is within a reasonable error range, the setting of the error range has a significant relationship with the size of the frequency. Thus, Frequency resolution refers to the minimum interval of two frequencies that can be resolved. Accordingly, the parameter can be used to justify whether the difference between the identified natural frequency and the reference frequency value is within the resolution range. The definition of frequency resolution is presented by

$$
\Delta f = \frac{f_s}{L} \tag{2}
$$

where Δf denotes the frequency resolution.

In this study, four SSI-DATA methods are developed and primarily vary from the projection techniques and the mode elimination methods. In the following, each refined method are individually introduced.

2.1 Mode-by-Mode Method 1

In this method, the main change to the traditional SSI-DATA method is adding the mode elimination. By subtracting the first two components taken from the SVD of the projection matrix, a new projection matrix without the extracted mode, represented by the first two components, is obtained. The formulation of this step is expressed by

$$
\mathbf{O}_r = \mathbf{O} - \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^T
$$
 (3)

where O_r is the projection matrix without the extracted mode, $U_1S_1V_1^T$ is the components from the SVD of **O**. As a result, because the components are in the other subspace of the future part of Hankel matrix Y_f , the new projection matrix O_f can be regarded as the new Y_f .

In the signal reconstruction, the new Y_f serves as the new Hankel matrix to produce a new time-series response data without the extracted mode. Because the new Hankel matrix is smaller than the original one, the time length of the reconstructed signal is shorter than the original output signal.

2.2 Mode-by-Mode Method 2

In this method, the refinement is focused on the matrix projection and mode elimination. When calculating the projection for the projection matrix \bf{O} from the projection of \bf{Y}_p and Y_f from the Hankel matrix, enough singular values should be selected from Y_p to represent the space. Therefore, the more efficient projection in this method is to use Y_f to project on the right singular vector of Y_p and presented by

$$
\mathbf{Y}_p = \mathbf{U}_p \mathbf{S}_p \mathbf{V}_p^T \tag{4}
$$

$$
\mathbf{O} = \mathbf{Y}_f \mathbf{V}_p \tag{5}
$$

where U_p , S_p , and V_p are the left singular matrix, the singular value matrix, and the right singular matrix from the SVD of Y_p , respectively.

In the mode elimination step, the first two singular vectors should also be removed from the projection matrix. However, the new Y_f is not represented directly by the new projection matrix; thus, the new matrix should be converted back to its original space to become the new Y_f . Because only a relatively small amount of information was selected to represent the space of Y_p when performing the projection, the inverse of the right singular vector of Y_p is necessary to convert the projection matrix back to the space of Y_f , such that

$$
\mathbf{O}_r = \mathbf{O} - \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^T \tag{6}
$$

$$
\mathbf{O}_n = \mathbf{O}_r \mathbf{V}_p^{-1} \tag{7}
$$

Hereafter, the new projection matrix \mathbf{O}_n will serve as the new \mathbf{Y}_f for the implementation of the successful signal reconstruction.

2.3 Mode-by-Mode Method 3

The mode-by-mode method 3 further modify the mode elimination, wherein the eliminated components of the projection matrix are utilized in this method. In the previous two methods, the left singular vector used for the formation of the eliminated components is from the SVD of the original projection matrix; however, in this method, the left singular value are obtained from the SVD of the temporary reconstructed projection matrix

 O_i . This reconstructed matrix is framed by the system matrices A_d and C_d from the state-space realization step, i.e. reconstructed by the natural frequency and mode shape identified in the previous step. The above discussion of the temporary reconstructed projection matrix **O***ⁱ* is also performed by

$$
\mathbf{O}_{i} = \begin{bmatrix} \mathbf{C}_{d} \\ \mathbf{C}_{d} \mathbf{A}_{d} \\ \vdots \\ \mathbf{C}_{d} \mathbf{A}_{d} \end{bmatrix} \approx U_{1i} S_{1i} V_{1i}^{T}
$$
(8)

where U_{1i} , S_{1i} , and V_{1i} are the left singular matrix, the singular value matrix, and the right singular matrix from the SVD of **O***i*, respectively.

Afterward, the first 2 singular vectors should be removed from the projection matrix **O**. In this method, because the SVD process provides the best approximation of the matrix, the original projection matrix is not consistent with the actual one. Therefore, the original left singular vector **U**¹ should be multiplied by the reconstructed left singular vector **U**1*ⁱ* and the original projection matrix will be converted into the reconstructed new space. With this, the components that can represent the first two singular vectors are obtained by $\mathbf{U}_{1i}(\mathbf{U}_{1i}^T \mathbf{U}_1) \mathbf{S}_1 \mathbf{V}_1^T$ in

$$
\mathbf{O}_r = \mathbf{O} - \mathbf{U}_{1i}(\mathbf{U}_{1i}^T \mathbf{U}_1) \mathbf{S}_1 \mathbf{V}_1^T
$$
\n(9)

Subtract these components from the original projection matrix **O**, the projection matrix without the extracted mode \mathbf{O}_r is calculated. After that, convert \mathbf{O}_r back to the space to obtain the new projection matrix O_n by Eq. [\(7\)](#page-4-0), which can also represent as the Y_f matrix. Then, this Y_f will also be used as a new Hankel matrix for signal reconstruction.

2.4 Mode-by-Mode Method 4

This mode-by-mode method 4 further modifies the mode elimination presented in the mode-by-mode method 3. In the mode-by-mode method 3, the temporary reconstructed projection matrix O_i is established by the system matrices A_d and C_d from the statespace realization step; however, this method utilizes the natural frequency f , damping ratio *ξ* and mode shape *φ* extracted in the mode extraction step to reconstruct a new **A***^d* and **C***d*, as presented by

$$
\mathbf{A}_d = e^{\mathbf{A}_c \Delta t} \n\mathbf{C}_d = \left[\phi \ \phi \right]
$$
\n(10)

where

$$
\mathbf{A}_c = \begin{bmatrix} 0 & \mathbf{I} \\ -(2\pi f)^2 & -2\xi(2\pi f) \end{bmatrix}
$$
 (11)

 A_c is the continuous-time system matrix, A_d is the discrete-time system matrix, **I** is an identity matrix, ϕ is an identified mode shape per mode, and Δt is the discrete-time step.

2.5 Implementation of Mode-by-Mode System Identification

To effectively use the above four different mode-by-mode methods, a unified process is developed and briefly described as follows:

- (1) Load data and perform signal preprocessing.
- (2) Use SSI-DATA to extract the preliminary natural frequency results, which will be utilized as reference frequency values for the mode-by-mode procedures.
- (3) Set the Hankel matrix size using Eq. [\(1\)](#page-2-1).
- (4) Determine whether the frequencies of all modes are extracted. Then, confirm whether to perform the following mode-by-mode steps. If all modes have been found, the system identification procedure is completed.
- (5) Calculate the Hankel matrix size (L_{temp}) of each round.
- (6) Perform the mode-by-mode procedures to extract the modal parameters.
- (7) Check whether the results from the current round of mode-by-mode had extracted any natural frequency.
- (8) Find the mode with the minimum error. Check whether the minimum error is lower than 15%.
- (9) Calculate the difference between the identified natural frequency and the reference natural frequency value and check if it is within the preferable frequency resolution range.
- (10) Go back to step 4 to check whether all the modes were identified through the mode-by-mode process.

3 Numerical Study of Eight-Story Steel-Frame Building

An eight-story steel-frame building model is constructed using finite-element analysis. The simulated output response signals are constructed as 3-min acceleration data with a damping ratio $\zeta = 0.01$ and sampling rate $f_s = 250$ Hz.

The stabilization diagrams should be plotted to obtain the reference natural frequencies, and the selected dominant frequencies of the structure are 4.2056, 12.6978, 20.7127, 28.2823, 37.1001, 47.3181, 56.8621, and 65.0942 Hz in order. Therefore, By applying f_s as 250 Hz, f_{n1} as 4.2056 Hz, and n_v as 8 into Eq. [\(1\)](#page-2-1), the Hankel matrix size of this case can be calculated as 512.

By applying the four refined methods to analyze the simulated responses, the overall identification results are compared as follows. First, for the identified natural frequencies, all modes can be extracted with little difference compared with the numerical solutions, where the results from the refined methods had closer values to the actual frequencies. Then, regarding the orders of iterations, all the modes are successfully extracted within eight rounds of mode-by-mode operations. In addition, the extracted orders follow the dominance of modes. To quantify the consistency among the identified mode shapes (see Fig. [3\)](#page-7-0), modal assurance criterion (MAC) values are introduced. For this case, all the resulting MAC values are close to 1, meaning that the identified mode shapes and the numerical solutions are approximately equal. To sum up, the mode-by-mode methods generally demonstrate the slightly better performance of identified natural frequencies and mode shapes than the traditional SSI-DATA.

Fig. 3. Identified mode shapes from proposed mode-by-mode methods, i.e., (a)-(d) present the method 1–4, respectively.

4 Experimental Study of Chillon Viaduct

The Chillon viaducts are the two independent parallel pre-stressed reinforced concrete box girder highways located in Switzerland, connecting the Glion Tunnel and Villeneuve for the southern end and the northern end, respectively. Several instruments are installed to evaluate the viaducts' performance, including 11 accelerometers, 4 strain gauges, 7 temperature, and humidity sensors, as illustrated in Fig. [4.](#page-7-1)

Taking the data collected at 5 p.m. on March 27th, 2017 as an example, the proposed four mode-by-mode methods are applied. The sampling rate of the accelerometer response is 100 Hz. The signals are initially processed to be zero-mean. Additionally, because the frequency range of interest is below 10 Hz, the measured signals are then down-sampled to 25 Hz.

The stabilization diagrams are also established to obtain the referenced natural frequencies. The eight dominant frequencies are 1.1684, 1.5298, 1.9018, 3.3072, 3.7449, 4.3460, 6.7055, and 7.3860 Hz. The Hankel matrix size can be calculated by Eq. [\(1\)](#page-2-1), where f_s is 25 Hz, n_1 is 1.1684 Hz, and n_v is 7 in this case, resulting in an *L* of 161.

The system identification results of the Chillon Viaducts are presented in Table [1](#page-8-0) and discussed as follows. For the identified natural frequencies, the differences between the ones obtained from the stabilization diagram of the SSI-DATA and the ones from the refined methods are under 0.20 Hz and acceptable. In addition, the extraction orders of the modes follow the extent of the mode's dominance.

Fig. 4. Illustration of Chillon viaducts: (a) a photo with the selected span highlighted [\[10\]](#page-10-5) and (b) sensor plan for accelerometers [\[11\]](#page-10-6).

As for the identified mode shapes, the MAC values were also calculated to quantify the consistency of the mode shapes. The values of the five selected modes in this case also show MACs over 0.94, except for the fourth mode from the second refined method.

This refined method also has an acceptable MAC value of 0.9189 because it contained the least energy among the modes.

The computational time of each method is also discussed in this case. The durations of the original method and the mode-by-mode method 1 to 4 are 40.6051, 99.718, 18.0253, 17.4154, and 18.4232 s, respectively. As observed, the mode-by-mode method 1 uses the most time to extract the modes, even longer than the original SSI-DATA, in particular for the mode-by-mode method 1 conducting the traditional projection process to extract so many modes. For the mode-by-mode methods 2, 3, and 4, the time these methods need for conducting system identification is only about 18 s, which are preferable as compared to the original method.

		SSI-DATA	method 1	method 2	method 3	method 4
mode 1	frequency (Hz)	1.1684	1.2118	1.2108	1.1660	1.1660
	difference		0.0434	0.0424	0.0024	0.0024
	iteration		3	3	5	5
	MAC value		0.9983	0.9983	0.9982	0.9982
mode 2	frequency (Hz)	1.5298	1.5974	1.6032	1.5080	1.5055
	difference		0.0676	0.0734	0.0218	0.0243
	iteration		5	5	$\overline{4}$	$\overline{4}$
	MAC value		0.9812	0.9813	0.9991	0.9988
mode 3	frequency (Hz)	1.9018	1.8844	1.8845	1.8594	1.8593
	difference		0.0174	0.0173	0.0424	0.0425
	iteration		2	2	3	3
	MAC value		0.9960	0.9960	0.9935	0.9933
mode 4	frequency (Hz)	3.3072	3.5096	3.5065	3.2469	3.3958
	difference		0.2024	0.1993	0.0603	0.0886
	iteration		17	18	16	16
	MAC value		0.9257	0.9189	0.9829	0.9538
mode 5	frequency (Hz)	3.7449	3.6457	3.6455	3.6175	3.6173
	difference		0.0993	0.0994	0.1274	0.1276
	iteration		$\overline{4}$	$\overline{4}$	6	6
	MAC value		0.9950	0.9950	0.9945	0.9945

Table 1. Identification results from proposed mode-by-mode methods for Chillon Viaducts

(*continued*)

		SSI-DATA	method 1	method 2	method 3	method 4
mode 6	frequency (Hz)	4.3460	4.4555	4.4534	4.2955	4.1610
	difference		0.1095	0.1075	0.0505	0.1850
	iteration		12	12	13	9
	MAC value		0.9412	0.9418	0.9809	0.9752
mode 7	frequency (Hz)	6.7055	6.6937	6.6927	6.6939	6.6951
	difference		0.0118	0.0129	0.0116	0.0105
	iteration		3	3	3	3
	MAC value		0.9947	0.9946	0.9949	0.9942
mode 8	frequency (Hz)	7.3860	7.4113	7.4088	7.4088	7.4088
	difference		0.0253	0.0228	0.0228	0.0228
	iteration		1	1	1	$\mathbf{1}$
	MAC value		0.9989	0.9988	0.9988	0.9988

Table 1. (*continued*)

5 Conclusions

In this research, four kinds of refined data-driven stochastic subspace identification methods were proposed to enhance system identification result accuracy. The major concluding remarks of this research are listed as follows:

- (1) The four refined SSI-DATA methods were applicable to both simulation structures and real-world building and bridge structures of buildings and bridges. For example, the proposed formulation of the Hankel matrix size could sufficiently include the information of the modes of interest from the structures. In addition, the designed analysis procedure could successfully and automatically extract all the concerned modes one by one.
- (2) Because the numerical solution was given for the numerical studies, the identification results from the traditional and refined methods could be compared with the actual values. The results of the mode-by-mode methods were actually slightly more accurate than the original SSI-DATA, and the performance of the second refined method was the best among the methods.
- (3) For the experimental studies, because the true identification results were unknown, the results of the refined methods were simply compared to those of the traditional methods. The identified frequencies and the calculated MAC values had minor differences between the original methods and the refined methods, while the computational time of the refined methods 2, 3, and 4 was much less than the traditional methods, which were often increased by order of magnitude.
- (4) Because the second method took the least computational time and its identified modal parameters were relatively more stably accurate than the other ones, this method is the preferable one in the comparison of the overall identification results and performances of the four refined methods. In contrast, the first refined method

required the most time during iterations, and this method might not be able to extract all the concerned modes or obtain relatively more satisfactory results in the cases, indicating that this method was less effective among all refined methods. The identification results and iteration times were similar to each other for the third and the fourth method, and they could also provide acceptable results.

References

- 1. Van Overschee, P., De Moor, B.: Subspace algorithm for the stochastic identification problem. In: $30th$ IEEE Conference on Decision and Control, pp. 1321–1326 (1991)
- 2. Peeters, B., De Roeck, G.: Reference-based stochastic subspace identification for output-only modal analysis. Mech. Syst. Signal Process. **13**(6), 855–878 (1999)
- 3. Liu, Y.C., Loh, C.H., Ni, Y.Q.: Stochastic subspace identification for output-only modal analysis: application to super high-rise tower under abnormal loading condition. Earthquake Eng. Struct. Dynam. **42**(4), 477–498 (2013)
- 4. Shimpi, V., Sivasubramanian, M.V., Singh, S.B.: System identification of heritage structures through AVT and OMA: a review. Struct. Durabl. Health Monit. **13**(1), 1–40 (2019)
- 5. Gentile, C., Gallino, N.: Condition assessment and dynamic system identification of a historic suspension footbridge. Struct. Control. Health Monit. **15**(3), 369–388 (2008)
- 6. Zonno, G., Aguilar, R., Boroschek, R., Lourenço, P.B.: Automated long-term dynamic monitoring using hierarchical clustering and adaptive modal tracking: validation and applications. J. Civ. Struct. Heal. Monit. **8**(5), 791–808 (2018). <https://doi.org/10.1007/s13349-018-0306-3>
- 7. Ercan, E.: Assessing the impact of retrofitting on structural safety in historical buildings via ambient vibration tests. Constr. Build. Mater. **164**, 337–349 (2018)
- 8. Elyamani, A., Roca Fabregat, P.: A review on the study of historical structures using integrated investigation activities for seismic safety assessment Part I: dynamic investigation. Sci. Cult. **4**(1), 1–27 (2018)
- 9. Peeters, B., De Roeck, G.: Stochastic system identification for operational modal analysis: a review. J. Dyn. Syst. Meas. Contr. **123**(4), 659–667 (2001)
- 10. Brühwiler, E., Bastien Masse, M.: Strengthening the Chillon viaducts deck slabs with reinforced UHPFRC. In: IABSE Conference Geneva 2015 - Structural Engineering: Providing Solutions to Global Challenges, pp. 1171–1178, IABSE (2015)
- 11. Martín-Sanz, H., Tatsis, K., Dertimanis, V.K., Avendaño-Valencia, L.D., Brühwiler, E., Chatzi, E.: Monitoring of the UHPFRC strengthened Chillon viaduct under environmental and operational variability. Struct. Infrastruct. Eng. **16**(1), 138–168 (2020)