An approach to operational modal analysis using the expectation maximization algorithm

F. Javier Cara [a,*], Jaime Carpio [a], Jesús Juan [a], Enrique Alarcón [b]

[a] Laboratory of Statistics, ETS Ingenieros Industriales, Universidad Politécnica de Madrid, José Gutiérrez Abascal, 2. 28006 Madrid, Spain
[b] Department of Structural Mechanics, ETS Ingenieros Industriales, Universidad Politécnica de Madrid, José Gutiérrez Abascal, 2. 28006 Madrid, Spain

ABSTRACT

This paper presents the Expectation Maximization algorithm (EM) applied to operational modal analysis of structures. The EM algorithm is a general-purpose method for maximum likelihood estimation (MLE) that in this work is used to estimate state space models. As it is well known, the MLE enjoys some optimal properties from a statistical point of view, which make it very attractive in practice. However, the EM algorithm has two main drawbacks: its slow convergence and the dependence of the solution on the initial values used. This paper proposes two different strategies to choose initial values for the EM algorithm when used for operational modal analysis: to begin with the parameters estimated by Stochastic Subspace Identification method (SSI) and to start using random points.

The effectiveness of the proposed identification method has been evaluated through numerical simulation and measured vibration data in the context of a benchmark problem. Modal parameters (natural frequencies, damping ratios and mode shapes) of the benchmark structure have been estimated using SSI and the EM algorithm. On the whole, the results show that the application of the EM algorithm starting from the solution given by SSI is very useful to identify the vibration modes of a structure, discarding the spurious modes that appear in high order models and discovering other hidden modes. Similar results are obtained using random starting values, although this strategy allows us to analyze the solution of several starting points what overcome the dependence on the initial values used.

1. Introduction

System identification can be described, in broad terms, as the construction of a dynamic model for the system based on experimental measurements. In the context of vibrating engineering structures, the system refers to structures such as buildings and bridges, and identification mostly involves the determination of modal parameters (the natural frequencies, damping ratios and mode shapes). Modal parameters are computed from the identified system model so a good identification method influences the quality and the quantity of modal parameters that can be estimated. This explains the increasing interest in accurate system identification methods for modal analysis.

Classic identification algorithms are based on input–output relationships where a known input is applied to a system and then the system response is recorded (called Experimental Modal Analysis). The difficulty concerning the application of this procedure to civil structures is mainly due to the impossibility of conducting experiments on real structures under controlled conditions. Consequently, the tests usually have to be performed in situ under operating conditions, and in this case the measurement of the dynamic forces is generally not possible. This concern led to the idea of using the ambient vibrations, which arise due to environmental factors such as wind and traffic under the operational conditions of the structure, and for this reason is called Operational Modal Analysis.

In this paper, we investigate the application of the Expectation Maximization (EM) algorithm to operational modal analysis of structures. The EM algorithm is a well-known tool for the computation of maximum likelihood estimates. It consists on an E-step (evaluating a conditional Expectation) and later on a M-step (which Maximizes the previous expectation) in an iterative loop until the convergence is reached.

The name EM was given by Dempster et al. [1], who presented the general formulation of the EM algorithm, established its basic properties and provided many examples and applications. However, the EM algorithm applied to the mathematical model written in the state space form was mainly developed by Shumway and Stoffer [2,3].

A review of the applications of the EM algorithm in many different contexts can be found in the monographs of Little and Rubin [4] and McLachlan and Krishnam [5]. The method is mainly used in a variety of incomplete-data problems arising in standard statistical situations such as linear models, contingency tables and loglinear models, random effects models and general variance-components models, time series and stochastic processes, among others. But these algorithms have also been profitably used in engineering, psychometry, econometrics, epidemiology, genetics, astronomy, etc. However, as far as we know, the application of the EM algorithm to maximum likelihood estimation of modal parameters in structures was not used before the present study.

The characteristics of the EM algorithm are well documented (see for instance [5]). It leads in general to simple equations, has the property of increasing the loglikelihood at each iteration until convergence and it derives sensible parameter estimates, and consequently it is a popular tool to derive maximum likelihood estimation. However, the EM algorithm is known to converge slowly in some situations. This important aspect has received much attention recently and many algorithms have been proposed to speed up the convergence of EM algorithm while preserving its simplicity (see [5, Chapter 4]).

The second important drawback of EM algorithm is that its solution can highly depend on its starting position and, consequently, produce sub-optimal maximum likelihood estimates. We propose in this paper two strategies designed to overcome this limitation:

1. The first strategy is to apply the Stochastic Subspace Identification (SSI) method and the EM algorithm jointly (this can be done if the starting point of the EM algorithm is the solution given by SSI). The reason is evident: we think that the solution of SSI is near to the global maximum, so the EM algorithm will convergence to it and not to a local maximum.
2. The second strategy consists in generating random starting points for the EM algorithm in an easy and fast way. Each random starting point provides a set of estimated modes: if a mode is present at several starting points, we propose it as a mode of the system.

The paper is organized as follows. In the next section, stochastic state space models of vibrating structures are presented, as well as the relationship between its parameters and the modal parameters. A complete description of the EM algorithm for state space models is presented in Section 3. In Section 4 we propose different alternatives to build starting values for the EM algorithm. Next, the results obtained from the analysis of both the simulated and the experimental data from a steel frame used as a benchmark structure proposed by the ASCE Task Group on Health Monitoring [6] are discussed. The conclusions are summarized in the last section.

2. Modal analysis in a state space model

2.1. State space equations

A vibrating structure can be represented by a discrete-time stochastic state-space model given as (see Appendix A)

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{1a}$$

$$y_t = Cx_t + Du_t + v_t, \tag{1b}$$

where t denotes the time instant, of a total number N, measured with constant sampling time Δt; $y_t \in \mathbb{R}^{n_o}$ is the measured output vector; $u_t \in \mathbb{R}^{n_i}$ is the measured input vector; $x_t \in \mathbb{R}^{n_s}$ is the state vector; n_o, n_i and n_s are the number of outputs, inputs and the order of the state vector, respectively. $A \in \mathbb{R}^{n_s \times n_s}$ is the transition state matrix describing the dynamics of the system; $B \in \mathbb{R}^{n_s \times n_i}$ is the input matrix; $C \in \mathbb{R}^{n_o \times n_s}$ is the output matrix, which is describing how the internal state is transferred to the output measurements y_t; $D \in \mathbb{R}^{n_o \times n_i}$ is the direct transmission matrix. The noise vectors comprise unmeasurable vector signals; $w_t \in \mathbb{R}^{n_s}$ is the process noise due to disturbances and modelling discrepancies, while $v_t \in \mathbb{R}^{n_o}$ is the measurement noise due to sensor inaccuracy. Both are assumed to be zero-mean, white noise sequences with covariance matrices Q and R, respectively.

In the case of ambient vibration testing, only the responses of a structure are measured, while the input sequence u_t remains unmeasured. Thus, Eq. (1) results in a purely stochastic system:

$$x_{t+1} = Ax_t + w_t, \tag{2a}$$

$$y_t = Cx_t + v_t. \tag{2b}$$

However, the white noise assumptions of these noise terms cannot be omitted and if the input contains also some dominant frequency components in addition to white noise, these frequency components cannot be separated from the eigenfrequencies of the system and they will appear as poles of the state matrix A.

2.2. System identification and modal analysis

The system identification problem investigated here can be defined as the determination of the corresponding system matrices A, C, Q and R (up to within a similarity transformation) using the output measurements $\{y_1, y_2, \ldots, y_N\}$ available for N time steps.

The natural frequencies and modal damping ratios can be retrieved from the eigenvalues of A, and the mode shapes can be evaluated using the corresponding eigenvectors and the output matrix C.

The eigenvalues of A come in complex conjugate pairs and each pair represents one physical vibration mode. Assuming proportional damping, the second order modes are uncoupled and the jth eigenvalue of A has the form:

$$\lambda_j = \exp[(-\zeta_j\omega_j \pm i\omega_j\sqrt{1-\zeta_j^2})\Delta t], \tag{3}$$

where ω_j is the natural frequency, ζ_j is damping ratio, and Δt is the time step. Therefore

$$\omega_j = \frac{|\ln(\lambda_j)|}{\Delta t}, \tag{4}$$

$$\zeta_j = \frac{-\text{Real}[\ln(\lambda_j)]}{\omega_j \Delta t}. \tag{5}$$

The jth mode shape $\upsilon_j \in \mathbb{R}^{n_o}$ evaluated at sensor locations can be obtained using the following expression:

$$\upsilon_j = C\psi_j, \tag{6}$$

where ψ_j is the complex eigenvector of A corresponding to the eigenvalue λ_j.

3. Maximum likelihood estimation using the EM algorithm

In this section we present the proposed identification method for estimating the parameters of the stochastic state space model given by Eq. (2), that is, A, C, Q and R. This method maximize the likelihood applying the iterative Expectation Maximization algorithm.

3.1. Evaluation of the likelihood function for Gaussian state-space models

Given N measurements of the outputs $Y_N = \{y_1, y_2, \ldots, y_N\}$, the likelihood is computed using the innovations $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_N\}$, which are defined by Eq. (B.6). The innovations are independent Gaussian random vectors, $\epsilon_t \rightsquigarrow N(0, \Sigma_t)$, with covariance matrix Σ_t given by Eq. (B.7). Thus, ignoring a constant, the logarithm of the likelihood may be written as

$$l_{Y_N}(\theta) = -\frac{1}{2}\sum_{t=1}^{N}(\ln|\Sigma_t(\theta)| + \epsilon_t(\theta)^T \Sigma_t(\theta)^{-1}\epsilon_t(\theta)), \tag{7}$$

where it has been emphasized the dependence of the innovations on the vector θ. This vector represents the unknown parameters of the model (2) assuming the initial state is normal, $x_0 \rightsquigarrow N(\mu_0, \Sigma_0)$:

$$\theta \stackrel{def}{=} (A, C, Q, R, \mu_0, \Sigma_0).$$

A wide range of numerical search algorithms are available for maximising the loglikelihood (7), and many of these are based on Newton–Raphson's algorithm. In addition to Newton–Raphson, Shumway and Stoffer [3] presented a conceptually simpler estimation procedure based on the Expectation Maximization algorithm. The EM algorithm is simple to apply since at each iteration the optimal solution for the unknown parameters can be obtained from explicit formulas.

3.2. Expectation Maximization algorithm

The basic idea is that if the states, $X_N=\{x_0,x_1,x_2,\ldots,x_N\}$, could be observed in addition to the observations, $Y_N=\{y_1,y_2,\ldots,y_N\}$, then the complete data could be considered, with the joint density:

$$f_\theta(X_t,Y_t)=f_\theta(X_t|Y_t)\cdot f_\theta(Y_t).$$

The likelihood is computed as

$$\prod_{t=1}^{N} f_\theta(X_t,Y_t)=\prod_{t=1}^{N} f_\theta(X_t|Y_t)\cdot\prod_{t=1}^{N} f_\theta(Y_t),$$

$$L_{X_N,Y_N}(\theta)=L_{X_N|Y_N}(\theta)\cdot L_{Y_N}(\theta).$$

And the logarithm of the likelihood

$$\log L_{X_N,Y_N}(\theta)=\log L_{X_N|Y_N}(\theta)+\log L_{Y_N}(\theta),$$

$$l_{X_N,Y_N}(\theta)=l_{X_N|Y_N}(\theta)+l_{Y_N}(\theta).$$

But $l_{X_N,Y_N}(\theta)$ and $l_{X_N|Y_N}(\theta)$ are functions of the states X_N, which are unknowns: the method proposes to replace them with theirs expected values. So given a value for the parameters at step j, and calling them θ_j, it is defined

$$G(\theta|\theta_j)=E[l_{X_N,Y_N}(\theta)|Y_N,\theta_j],$$

$$H(\theta|\theta_j)=E[l_{X_N|Y_N}(\theta)|Y_N,\theta_j].$$

Thus

$$l_{Y_N}(\theta)=G(\theta|\theta_j)-H(\theta|\theta_j), \tag{8}$$

because of $E[l_{Y_N}(\theta)|Y_N,\theta_j]=l_{Y_N}(\theta)$. From (8), we have that

$$l_{Y_N}(\theta_{j+1})-l_{Y_N}(\theta_j)=(G(\theta_{j+1}|\theta_j)-G(\theta_j|\theta_j))-(H(\theta_{j+1}|\theta_j)-H(\theta_j|\theta_j)).$$

But taking into account Jensen's inequality and the concavity of the logarithm function,

$$H(\theta_{j+1}|\theta_j)-H(\theta_j|\theta_j)\leq 0.$$

So if we develop a procedure which verifies

$$G(\theta_{j+1}|\theta_j)\geq G(\theta_j|\theta_j),$$

then automatically it is achieved

$$l_{Y_N}(\theta_{j+1})-l_{Y_N}(\theta_j)\geq 0$$

and we will reach a maximum for Eq. (7), too. The Expectation Maximization algorithm just provides an iterative method for maximize successively the conditional expectation $G(\theta|\theta_j)$, which is equivalent to maximize the likelihood $l_{Y_N}(\theta)$ (7). Each iteration of the EM algorithm consists in the following two steps, repeated until θ_{j+1} and θ_j are closed each other:

1. The first step (E step) is to compute $G(\theta|\theta_j)=\mathrm{E}[l_{X_N,Y_N}(\theta)|Y_N,\theta_j]$.
2. The second step (M step) consists in maximizing $G(\theta|\theta_j)$ obtaining θ_{j+1}.

3.2.1. Expectation step

Property 1. *The complete likelihood $L_{X_N,Y_N}(\theta)$ is computed taking into account that*

$$x_0 \rightsquigarrow N(\mu_0,\Sigma_0),$$

$$w_t=x_{t+1}-Ax_t,\quad w_t\rightsquigarrow N(0,Q),$$

$$v_t=y_t-Cx_t,\quad v_t\rightsquigarrow N(0,R),$$

with probability density functions

$$f_{\mu_0,\Sigma_0}(x_0)=\frac{1}{(2\pi)^{n_s/2}|\Sigma_0|^{1/2}}\exp\left(-\frac{1}{2}(x_0-\mu_0)^T\Sigma_0^{-1}(x_0-\mu_0)\right),$$

$$f_{A,Q}(w_t)=\frac{1}{(2\pi)^{n_s/2}|Q|^{1/2}}\exp\left(-\frac{1}{2}(x_t-Ax_{t-1})^TQ^{-1}(x_t-Ax_{t-1})\right),$$

$$f_{C,R}(v_t) = \frac{1}{(2\pi)^{n_o/2}|R|^{1/2}} \exp\left(-\frac{1}{2}(y_t - Cx_t)^T R^{-1}(y_t - Cx_t)\right),$$

and it is defined by

$$L_{X_N,Y_N}(\theta) = f_{\mu_0,\Sigma_0}(x_0) \prod_{t=1}^{N} f_{A,Q}(X_N) \prod_{t=1}^{N} f_{C,R}(X_N, Y_N). \tag{9}$$

So, the log-likelihood $l_{X_N,Y_N}(\theta) = \log L_{X_N,Y_N}(\theta)$ *can be written as a sum of three uncoupled functions*

$$l_{X_N,Y_N}(\theta) = -\tfrac{1}{2}[l_1(\mu_0, \Sigma_0) + l_2(A, Q) + l_3(C, R))],$$

where ignoring constants are

$$l_1(\mu_0, \Sigma_0) = \log|\mathbf{\Sigma}_0| + (x_0 - \mu_0)^T \mathbf{\Sigma}_0^{-1}(x_0 - \mu_0), \tag{10}$$

$$l_2(A, Q) = N \log|Q| + \sum_{t=1}^{N} (x_t - Ax_{t-1})^T Q^{-1}(x_t - Ax_{t-1}), \tag{11}$$

$$l_3(C, R) = N \log|R| + \sum_{t=1}^{N} (y_t - Cx_t)^T R^{-1}(y_t - Cx_t). \tag{12}$$

The function $G(\theta|\theta_j)$ is the conditional expectation of the sum of Eqs. (10)–(12), and it depends on the parameters $\theta = (A, C, Q, R, \mu_0, \Sigma_0)$.

Property 2. *Given the value of the parameters* θ_j *at iteration* j, *we introduce the following notation to define the conditional expectations*:

$$x_t^N = \mathrm{E}[x_t|Y_N, \theta_j], \tag{13}$$

$$P_t^N = \mathrm{E}[(x_t - x_t^N)(x_t - x_t^N)^T|Y_N, \theta_j], \tag{14}$$

$$P_{t,t-1}^N = \mathrm{E}[(x_t - x_t^N)(x_{t-1} - x_{t-1}^N)^T|Y_N, \theta_j], \tag{15}$$

which can be evaluated with the well known Kalman filtering and smoothing shown in Appendix B (*Properties* 6–8). *From the previous conditional expectation, it is possible to compute* $G(\theta|\theta_j)$ *as follows*:

$$G(\theta|\theta_j) = \mathrm{E}[l_{X_N,Y_N}(\theta)|Y_N, \theta_j] = \mathrm{E}[l_1(\mu_0, \Sigma_0)|Y_N, \theta_j] + \mathrm{E}[l_2(A, Q)|Y_N, \theta_j] + \mathrm{E}[l_3(C, R)|Y_N, \theta_j],$$

with

$$\mathrm{E}[l_1(\mu_0, \Sigma_0)|Y_N, \theta_j] = \ln|\Sigma_0| + \mathrm{tr}(\Sigma_0^{-1}[P_0^N + (x_0^N - \mu_0)(x_0^N - \mu_0)^T]), \tag{16}$$

$$\mathrm{E}[l_2(A, Q)|Y_N, \theta_j] = N \log|Q| + \mathrm{tr}(Q^{-1}[S_{xx} - S_{xb}A^T - AS_{bx} + AS_{bb}A^T]), \tag{17}$$

$$\mathrm{E}[l_3(C, R)|Y_N, \theta_j] = N \log|R| + \mathrm{tr}(R^{-1}[S_{yy} - S_{yx}C^T - CS_{xy} + CS_{xx}C^T]), \tag{18}$$

where

$$S_{xx} = \sum_{t=1}^{N} (P_t^N + x_t^N (x_t^N)^T), \quad S_{bb} = \sum_{t=1}^{N} (P_{t-1}^N + x_{t-1}^N (x_{t-1}^N)^T), \quad S_{yy} = \sum_{t=1}^{N} (y_t y_t^T), \tag{19}$$

$$S_{xb} = \sum_{t=1}^{N} (P_{t,t-1}^N + x_t^N (x_{t-1}^N)^T), \quad S_{bx} = S_{xb}^T, \tag{20}$$

$$S_{yx} = \sum_{t=1}^{N} (y_t (x_t^N)^T), \quad S_{xy} = S_{yx}^T. \tag{21}$$

and $\mathrm{tr}(A)$ *is the trace of matrix A.*

3.2.2. Maximization Step

Maximizing $G(\theta|\theta_j)$ with respect to the parameters θ at iteration j, constitutes the M-step. This is the strong point of the EM algorithm because the maximum values, obtained equating to zero the corresponding derivatives of the expectations (16)–(18), are obtained from explicit formulas.

Property 3. *The maximum of* $\mathrm{E}[l_1(\mu_0,\Sigma_0)|Y_N,\theta_j]$ *is attained at*

$$\hat{\mu}_0 = x_0^N, \tag{22}$$

$$\hat{\Sigma}_0 = P_0^N. \tag{23}$$

Property 4. *The maximum of* $\mathrm{E}[l_2(A,Q)|Y_N,\theta_j]$ *is attained at*

$$\hat{A} = S_{xb}S_{bb}^{-1}, \tag{24}$$

$$\hat{Q} = \frac{1}{N}(S_{xx}-S_{xb}\hat{A}^T-\hat{A}S_{bx}+\hat{A}S_{bb}\hat{A}^T). \tag{25}$$

Property 5. *The maximum of* $\mathrm{E}[l_3(C,R)|Y_N,\theta_j]$ *is attained at*

$$\hat{C} = S_{yx}S_{xx}^{-1}, \tag{26}$$

$$\hat{R} = \frac{1}{N}(S_{yy}-S_{yx}\hat{C}^T-\hat{C}S_{xy}+\hat{C}S_{xx}\hat{C}^T). \tag{27}$$

3.3. Overall procedure

The overall method can be summarized as an iterative procedure as follows:

- Initialize the procedure ($j=0$) selecting starting values for the parameters $\theta_0=(A_0,C_0,Q_0,R_0,\mu_0,\Sigma_0)$ and a stop tolerance δ.
- Repeat

  1. $j=j+1$.
  2. Perform the E-Step. Apply the Kalman filter (Properties 6–8 of Appendix B) to obtain the expected values x_t^N, P_t^N, and $P_{t,t-1}^N$ with θ_{j-1} as data. Use them to compute the matrices S_{xb}, S_{bb} and S_{xx} given by (19)–(21).
  3. Perform the M-Step. Update the parameters $\theta_j=(\hat{A},\hat{C},\hat{Q},\hat{R},\hat{\mu}_0,\hat{\Sigma}_0)$ using (22)–(27).
  4. Compute the incomplete-data likelihood, $l_{Y_N}(\theta_j)$ with (7).

  Until $|l_{Y_N}(\theta_j)-l_{Y_N}(\theta_{j-1})|/|l_{Y_N}(\theta_{j-1})|<\delta$.

4. Choosing starting values for the EM algorithm

The initial step for the EM algorithm is to choose a starting value for the parameters $\theta_0=(A_0,C_0,Q_0,R_0,\mu_0,\Sigma_0)$. This is a crucial step because, like in other iterative procedures, depending on the starting point, a local maximum can be reached instead of the global one (see for example [7,8]). In this section we present the approaches that we have considered:

1. *Using other estimation methods*: A natural choice is to begin with estimates obtained by other identification methods. There are several techniques to realize system identification (see for example [9]), but among them, we have chosen the Stochastic Subspace Identification method because it is a well known method of structural system identification. This method is based on the solution of the stochastic realization problem [10] and identifies state space models from (input and) output data by applying robust numerical techniques such as QR factorization, SVD and least squares. A complete overview of data-driven subspace identification (both deterministic and stochastic) is provided in [11].
   As we shall discuss in the numerical simulations, the likelihood of the solution given by SSI can be improved using the EM algorithm. So the SSI method provides a good starting point for the EM, or from another point of view, the EM can be used to refine the solution given by SSI. This is not new, and for example, Ljung [9, Section 7.3] proposes improving the quality of SSI by using the estimated model as an initial estimate for the *prediction error method* (PEM). Indeed, MATLAB uses the algorithm N4SID (which is a variant of SSI) as initialization method for the PEM.
   However, using only a starting point could lead to a sub-optimal solution: if the solution given by SSI method is near to a local maximum of the likelihood function, the EM algorithm may be not able to leave this area.
2. *Random initialization*: This is probably the most employed way of initializing the EM algorithm and for this reason we have considered this strategy too. Matrices A,C,Q,R are built using random numbers. Especial attention must be paid to matrices Q and R: they are symmetrical definite positive matrices.
   Random initialization can be a good strategy for the EM algorithm in low-dimensional problems. However, we have found some difficulties when using completely random starting values in the state space models of high order, as is our case, leading to unstable system and inaccurate solutions in some cases. Moreover, with random values we do not take

advantage of the available information of the problem: measured outputs, frequencies rank analyzed, typical values of the damping ratios, and so on.
For this reason, we propose in this paper a more sophisticated strategy to build random starting values that overcomes these cited problems. Our technique use random frequencies, random damping ratios and random mode shapes as the unique random parameters required to build the initial values. The rest of the section is devoted to define how to build this random initial values for the EM algorithm.

4.1. *Procedure to build random starting values*

The complete procedure to build random initial values for the parameters $\theta_0 = (\tilde{A}_0, \tilde{C}_0, \tilde{Q}_0, \tilde{R}_0, \tilde{\mu}_0^0, \tilde{\Sigma}_0^0)$ is outlined here ($(\tilde{\bullet})$ means random values or generated from random values).

1. Given the order n_s for the state space model (2), generate $n_d = n_s/2$ random values for the natural frequencies $\tilde{\omega}_j$, n_d random values for the damping ratios $\tilde{\zeta}_j$, and a random full matrix $\tilde{\Phi} \in \mathbb{R}^{n_d \times n_d}$ with columns linearly independent (if the natural frequencies are all distinct, the mode shapes are linearly independent [12]). To ensure the generated random starting point lies inside the unit circle, admissible values for natural frequencies go from 0 to Nyquist frequency, and for damping ratios are comprised between 0 and 1. In essence, we are defining an imaginary structure with modal parameters equal to $\tilde{\omega}_j$, $\tilde{\zeta}_j$ and $\tilde{\Phi}$.
2. Using the above random values, build the matrices A and C for the state space model (2) . One possibility is to use the expressions included in Appendix A, Eqs. (A.16) and (A.20):
$$\tilde{A} = \exp\left(\Delta t \begin{bmatrix} 0 & I \\ -\tilde{\Omega}^2 & -2\tilde{\Omega}\tilde{Z} \end{bmatrix}\right), \quad \tilde{C} = C_a \tilde{\Phi}[-\tilde{\Omega}^2 \quad -2\tilde{\Omega}\tilde{Z}], \tag{28}$$
where $\tilde{\Omega}$ and $\tilde{Z}$ are diagonal matrices formed by the random natural frequencies and the random damping ratios, $\tilde{\Omega}_{ij} = \tilde{\omega}_j \delta_{ij}$ and $\tilde{Z}_{ij} = \tilde{\zeta}_j \delta_{ij}$ (δ_{ij} being the Kronecker delta). C_a is the measurement location matrix corresponding to the acceleration responses of the structural system; we take $C_a = I^{n_o \times n_d}$ (a $n_o \times n_d$ matrix with ones on the diagonal and zeros elsewhere).
3. Matrices R and Q of the state space model (2) can be estimated after compute the discrete states associated with matrices $\tilde{A}$ and $\tilde{C}$ given by (28). This states are the modal coordinates (see Eq. (A.11)), which can be computed as
$$\tilde{z}_t = \begin{bmatrix} \tilde{\Phi}^{-1} & 0 \\ 0 & \tilde{\Phi}^{-1} \end{bmatrix} \begin{bmatrix} I_0 q_t \\ I_0 \dot{q}_t \end{bmatrix}, \tag{29}$$
where q_t and $\dot{q}_t$ are the discrete displacement and velocity, respectively, and can be obtained from the measured accelerations y_t using numerical integration (Simpson's rule, for example). I_0 is a $n_d \times n_o$ matrix with ones on the diagonal and zeros elsewhere, and it is used for matrix order compatibility; we take $I_0 = C_a^T$.

   (a) The value of R is obtained computing the covariance matrix of the noise vector calculated from the *observation equation* (2b)
$$y_t = Cz_t + v_t \Rightarrow \tilde{v}_t = y_t - \tilde{C}\tilde{z}_t \Rightarrow \tilde{R} = \frac{1}{N}\sum_{t=1}^{N} \tilde{v}_t \tilde{v}_t^T. \tag{30}$$

   (b) In the same way, the value of Q is obtained using the noise calculated from the *state equation* (2a)
$$z_{t+1} = Az_t + w_t \Rightarrow \tilde{w}_t = \tilde{z}_{t+1} - \tilde{A}\tilde{z}_t \Rightarrow \tilde{Q} = \frac{1}{N}\sum_{t=1}^{N} \tilde{w}_t \tilde{w}_t^T. \tag{31}$$

4. For numerical reasons, it is preferred that the elements of matrix Q were close to one. Therefore, we use the state transformation $\tilde{z}_t = T\tilde{\alpha}_t$, being T a diagonal matrix with diagonal elements equal to the standard deviation s_i of the ith component of the noise $\tilde{w}_t$. Taking into account this last transformation, the random initial values for the EM algorithm are computed by
$$\tilde{A}_0 = T\tilde{A}T^{-1}, \quad \tilde{C}_0 = \tilde{C}T^{-1}, \quad \tilde{Q}_0 = T\tilde{Q}T^T, \quad \tilde{R}_0 = \tilde{R}. \tag{32}$$

5. $\tilde{\mu}_0^0$, $\tilde{\Sigma}_0^0$ will be zero. The steady solution of the dynamic system is taken as initial condition.

Other random initial values can be used. However, most of tests we have made led to singular matrices in some steps of the iterations what cause the algorithm was interrupted unexpectedly (especially when we used experimental data). This means that not any matrix A, C, Q and R can be used to start the algorithm because they are all interrelated by means of Eq. (2b).

The procedure explained above is easily implemented with low computational cost and gets very good results.

4.2. Proposed method for operational modal analysis with the EM

In essence, we have two different procedures to perform Operational Modal Analysis of structures with the EM algorithm. The first one is the combination of the SSI method and the EM algorithm. This is probably the simplest approach and can be summarized as follows:

1. Generate the initial matrices for EM algorithm using the SSI method.
2. Run the EM algorithm until convergence.
3. Compute the modal parameters from this solution.

In the presence of multiple local minima, using only one starting point for the EM algorithm can lead to sub-optimal solutions because the solution can depends on the initial point (this is a drawback of all iterative methods, and of course, of the EM too). Therefore, the EM solution would be conditioned by SSI solution. This difficulty can be overcome by generating different starting values with the procedure described in Section 4.1. The problem arising now is that we need to choose a final solution between all the solutions obtained with each initial position. A first option is to choose the solution with the highest log-likelihood (Eq. (7)). A similar strategy has been used for the EM algorithm in mixture models [7] and can be described as follows:

1. Generate p initial positions (Section 4.1).
2. Run the EM algorithm at each initial position with a fixed number of iterations.
3. Select the solution providing best likelihood among the p trials, say θ^* (in the final comparison we recommended to include the solution provided by SSI+EM).
4. Compute the modal parameters from this solution.

The solution obtained this way will have the highest likelihood but not always is the best in the sense of estimated modal parameters. Moreover, due to EM algorithm always increase the likelihood at each iteration, the final likelihoods computed for each initial value are quite similar. For these reasons we propose an alternative strategy:

1. Generate p initial positions (Section 4.1).
2. Run the EM algorithm at each initial position (so we have p different solutions).
3. Compute the modal parameters from each solution and select those which are more repeated.

The latter is not a trivial step. We recommend using a similar approach to that used in the stabilization diagrams (see for example [13]) to decide if two parameters correspond to the same mode, that is

$$\frac{|\omega_{pi}-\omega_{qj}|}{\omega_{pi}} \le \varepsilon_\omega, \tag{33}$$

$$|\zeta_{pi}-\zeta_{qj}| \le \varepsilon_\zeta, \tag{34}$$

$$1-MAC(\upsilon_{pi},\upsilon_{qj}) \le \varepsilon_{MAC}, \tag{35}$$

where ε_ω, ε_ζ, ε_{MAC} are tolerance limits to decide if mode i estimated from initial point p is the same that mode j estimated from initial point q. MAC (modal assurance criterion) shows the degree of correlation between two vectors and it is computed as

$$MAC(\upsilon_1,\upsilon_2) = \frac{|\upsilon_1^H \upsilon_2|^2}{(\upsilon_1^H \upsilon_1)(\upsilon_2^H \upsilon_2)}.$$

where $(\bullet)^H$ means Hermitian operator.

Actually ε_ζ is defined in relative terms in literature (in the same way that ε_ω), but it is well known that the damping is estimated with greater variability than the natural frequencies, so we prefer to use ε_ζ in absolute terms. This means that if two modes identified from different starting point have similar natural frequencies and similar mode shapes, but with damping ratios differing in, for instance, 0.02, we think they correspond to the same mode.

5. Numerical examples

We have evaluated the performance of the EM method for Operational Modal Analysis using the data provided by the ASCE benchmark problem for structural health monitoring [6]. This benchmark studies consist in Phase I and Phase II, and

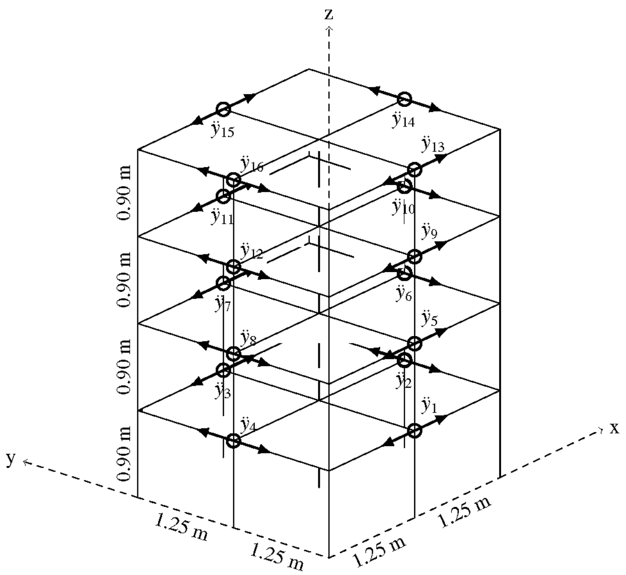

Fig. 1. Model of the benchmark structure and location of the 16 measured nodes.

each phase is divided into simulated and experimental problems. The benchmark structure is a four-story, two-bay by two-bay steel-frame scale model structure built in the Earthquake Engineering Research Laboratory at the University of British Columbia, Canada (Fig. 1). The January 2004 issue of the Journal of Engineering Mechanics contains the results of six different studies of the Phase I simulated benchmark problems, together with a definition and overview paper [14].

We have selected this example because it has analytical and experimental part, Matlab subroutines for simulations as well as experimental data are available on the internet [6], and results can be compared with those of other researchers.

5.1. Simulated data

A MATLAB-based finite element analysis code obtained from the IASC-ASCE SHM Task Group web site [6] has been used to simulate the dynamic response of the prototype structure. Two finite-element models based on the actual test structure were developed by the Task Group to generate the simulated structural response data: a 12DOF shear building model and a more complex 120DOF model. We have used the former because of its simplicity (there are three DOF per floor): the floors move as rigid bodies, with translation in the x and y directions and rotation θ about the center column. The natural frequencies and the mode shapes of this model are given in Fig. 2.

In this work we have used the horizontal acceleration of 16 nodes of the structure. These nodes are located at the center of each side of the structure, two in the x and two in the y directions per floor (called $\ddot{y}_1, \ddot{y}_2, \ldots, \ddot{y}_{16}$ in Fig. 1).

The structure response has been generated by exciting the model with broadband ambient inputs applied in the x and y directions, using a sampling frequency equal to 1000 Hz and 20 s of total duration. The 1% modal damping has been assigned to each mode. Finally, the observed value is the sum of the structure response and Gaussian noise. The root mean square (RMS) of this added noise is equal to the 30% of the largest structure response RMS.

5.1.1. Results obtained using one simulated time history response

First, we are going to perform modal identification using one simulated time history response. Fig. 3 shows the power spectral density function (Welch method) for this simulated case and for eight nodes of the structure: four in x direction (nodes 1, 5, 9 and 13 in Fig. 1) and four in y direction (nodes 4, 8, 12 and 16). The theoretical natural frequencies are also plotted in dashed lines, and it is observed that most of this theoretical frequencies are coincident with a maximum of the spectrum, although not all peaks are present in all channels.

The first value to determine in a real problem is the order of the state space model, n_s. There are different techniques to evaluate this parameter (for example, the number of singular values different to zero of the projection matrix in SSI [11], or the Akaike information criterion AIC for time series [15]), but in structures, where the number of modes is huge, probably the best way to estimate n_s is by inspection in stabilization diagrams.

In simulated structures the theoretical state space order equals two times the number of modes that are present in the data; we know there are 12 modes in the simulated case, so the theoretical order of the state space model is 24. For this reason, we decided to analyze the behavior of the EM under the exact order, and we have not chosen it by mean of the stabilization diagram.

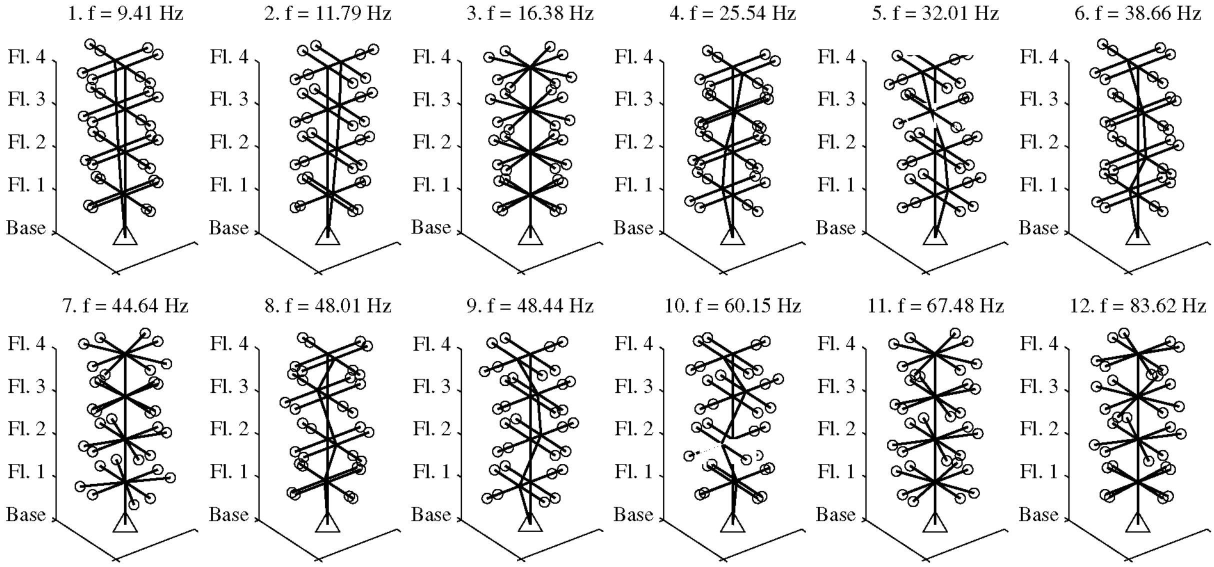

Fig. 2. Natural frequencies and mode shapes corresponding to matrices $M,K \in \mathbb{R}^{12\times 12}$ (the circles are the points where measurements are recorded).

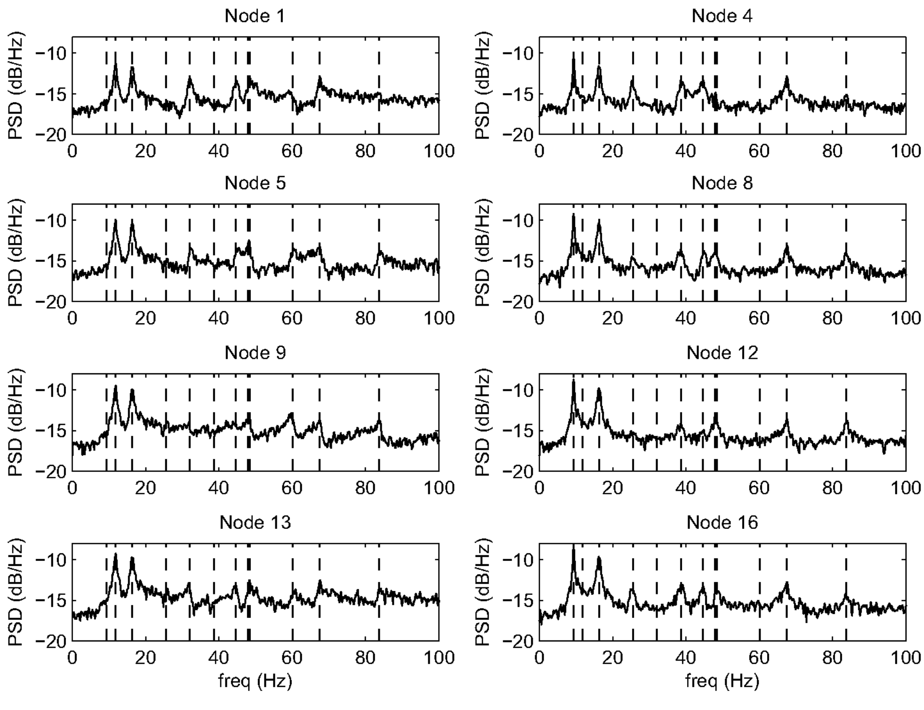

Fig. 3. Power spectral density of accelerations in different nodes of the structure. The analytical natural frequencies are also plot in dashed lines.

Modal parameters have been obtained using three different approaches:

1. SSI method: We have used the SSI-DATA algorithms detailed in [11].
2. EM method using SSI to obtain the initial value (EM1 in the following). We have applied the EM method as described in Section 3.3. The value we have used for the initial parameters $\theta_0 = (A_0,C_0,Q_0,R_0,\mu_0^0,\Sigma_0^0)$ has been: matrices (A,C,Q,R) estimated by SSI method, and zero matrices for (μ_0^0,Σ_0^0).
3. EM method using random starting values (EM2 in the following). In this case, the initial matrices $\theta_0 = (A_0,C_0,Q_0,R_0,\mu_0^0,\Sigma_0^0)$ have been generated as is described in Section 4.1.

Table 1
Modal parameters identified from one simulated time history responses using SSI method and EM method. In SSI and EM1, values in *light gray* do not verify (36)–(38). In EM2, n is the number of times the parameter has been identified over 100 starting random values.

Mode	ω_{th} (Hz)	ζ_{th} (%)	SSI			EM1 = EM + SSI			EM2 = EM + random			
			ω_{id} (Hz)	ζ_{id} (%)	MAC	ω_{id} (Hz)	ζ_{id} (%)	MAC	ω_{id} (Hz)	ζ_{id} (%)	MAC	n
1	9.41	1.00	9.40	0.70	0.99	9.45	1.45	0.99	9.49	2.54	0.99	72
2	11.79	1.00	11.78	1.31	0.99	11.94	3.35	0.99	11.90	3.51	0.99	68
3	16.38	1.00	16.32	1.06	1.00	16.34	1.55	1.00	16.34	1.81	0.99	74
4	25.54	1.00	–	–	–	–	–	–	25.60	2.08	0.96	44
5	32.01	1.00	–	–	–	32.10	1.25	0.97	32.11	1.66	0.95	96
6	38.66	1.00	37.95	*54.93*	0.56	38.68	1.52	0.99	38.80	1.71	0.96	74
7	44.64	1.00	44.55	1.16	0.99	44.68	0.94	0.99	44.70	1.35	0.99	100
8	48.01	1.00	49.84	*71.17*	0.08	48.09	0.89	*0.66*	–	–	–	–
9	48.44	1.00	48.42	0.99	0.98	48.48	1.06	0.99	48.41	1.20	0.94	82
10	60.15	1.00	59.86	1.82	*0.87*	60.06	0.88	0.99	60.11	1.16	0.99	72
11	67.48	1.00	67.50	0.88	0.99	67.41	0.79	0.99	67.39	1.17	0.99	100
12	83.62	1.00	84.13	0.90	*0.46*	83.59	0.64	0.99	83.84	0.97	0.99	100

The results of the three methods for the simulated case are shown in Table 1, where theoretical modal parameters are denoted by subscript *th*, and identified parameters with subscript *id*. We have considered that a parameter has been identified if it is verified at once that

$$\frac{|\omega_{th}-\omega_{id}|}{\omega_{th}} \le \varepsilon_\omega = 0.02, \tag{36}$$

$$|\zeta_{th}-\zeta_{id}| \le \varepsilon_\zeta = 0.03, \tag{37}$$

$$1-MAC(v_{th},v_{id}) \le \varepsilon_{MAC} = 0.10. \tag{38}$$

The behavior of the EM1 method is best seen in Fig. 4(a). In this figure we have plotted at each iteration of the EM1 algorithm the eigenvalues of the transition matrix A (related with the modal parameters ω_j and ζ_j) in the complex plane. The circles ($\circ$) correspond to the starting points which are the eigenvalues of matrix A obtained with SSI method, dots ($\cdot$) correspond to the intermediate solutions in the iterative procedure and crosses ($+$) are the final solution of the EM algorithm. Theoretical eigenvalues (near but inside unit circle) are also represented in the figure as squares ($\square$). To simplify the figure, only the upper part of the unit circle is plotted.

Many features are noticeable in the figure:

- Most of the eigenvalues obtained by SSI method are very closed to the theoretical ones: 1, 2, 3, 7, 9, 10, 11 and 12. In these cases, the EM iterations practically do not move the values. In the figure we can not appreciate these changes because the circles, the points and the crosses are inside the squares.
- The SSI method fails to identify modes 4, 5, 6 and 8 in this simulation (note that eigenvalues 8 and 9 are almost identical in the figure). Instead, the SSI method proposes the complex poles A, B, and the real poles C, D, E and F.
- The EM iterations correct the failed SSI eigenvalues A and B towards the theoretical values 6 and 8. It also joins two real poles in a complex pole that grows from the real axis towards mode 5.
- Mode 4 is not identified by either method.

Looking at Table 1 and Fig. 4(a) we can conclude that the EM algorithm improves the solution given by the SSI method. Not only the eigenvalues, even the eigenvectors are also improved: for example, the eigenvalue of mode 12 has been well estimated using SSI, but not so its eigenvector (MAC=0.46); however, the eigenvector identified with EM1 method is very close to the theoretical one (MAC=0.99). Similar behavior can be also seen for modes 6, 8 and 10.

For the same simulated time history response we have generated 100 different random starting points within the unit circle by mean of the procedure described in Section 4.1. The results obtained from these random starting points using the EM2 method have been included in Table 1. We have also included the number of times each mode has been identified over the total of the 100 starting points, called n.

Fig. 4(b) shows the convergence of eigenvalues of matrix A to the theoretical eigenvalues for one of the 100 random starting points. In this case, only the mode 8 remains unidentified and instead has been obtained the eigenvalue a. This mode is hard to find, and in fact, in none of the 100 starting points has been identified (see Table 1).

Comparing Fig. 4(a) and (b), we see how the solution obtained by the Expectation Maximization algorithm method depends on the chosen starting point.

Fig. 4(c) shows the likelihood of the SSI solution (the circles ($\circ$) in Fig. 4(a)), and how the EM algorithm increase the likelihood with the iterations. The likelihood of the solution estimated by EM1 method (the crosses ($+$) in Fig. 4(a)) is indicated as EM1 in Fig. 4(c).

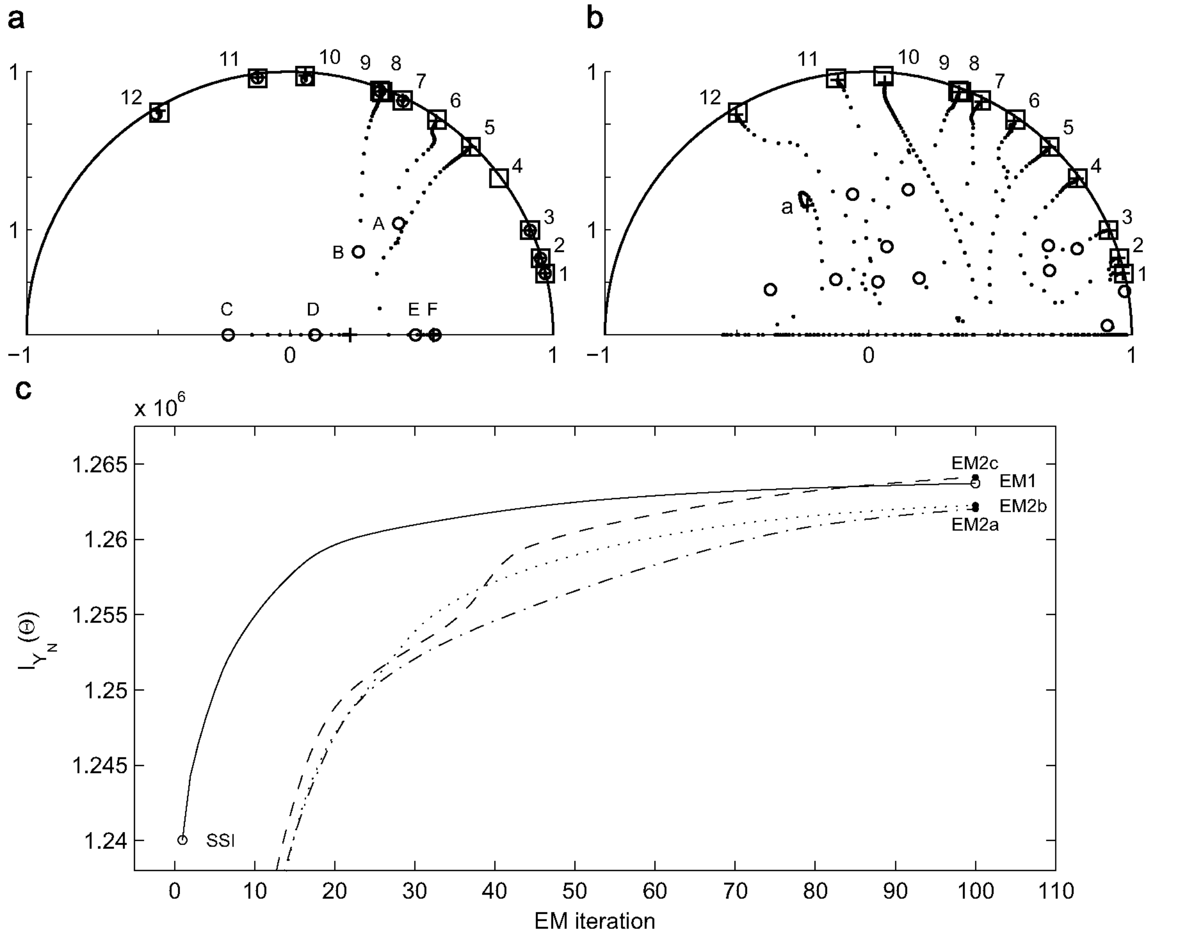

Fig. 4. (a) Theoretical eigenvalues (□) and evolution from eigenvalues of matrix A identified using SSI (○) and eigenvalues of matrix A identified using EM (+). (b) Evolution from random starting eigenvalues (○) to EM eigenvalues (+). (c) Likelihood of SSI, EM1 and EM2 estimates shown in (a) and (b). The likelihood of two additional random starting values has been also plotted.

On the other hand, the likelihood of the solution estimated by EM2 method (the crosses (+) in Fig. 4(b)) is called EM2a in Fig. 4(c). Two additional random starting points (from the total of 100 studied) have been also included, points EM2b and EM2c. From the point of view of the MLE method, we should choose the solution with higher likelihood, that is, the EM2c. However, we think we are wasting the information provided by the remaining 99 starting values. For this reason, we proposed as modal parameters those that are present at different starting points (the so-called EM2 method).

5.1.2. *Results obtained using 100 simulated time history response*

We are going now to analyze 100 simulated cases. Summarizing the results of these 100 simulations is not easy, but it would give us a clear comparison of the differences of the three methods. We have chosen the box-plot to present simultaneously the estimated value for the modal parameters (Fig. 5).

The box-plot is a graphical display based on the order statistics summaries of median and quartiles. A box is drawn from the first quartile to the third. The distributional center is indicated by a line at the median, within each box. The result spread or variation is shown by the box's height, which is the interquartile range. The plot is completed by a line at each side of the box that shown the "acceptable" values according to interquartile range and finally, the outliers which are extreme or discordant estimations (plotted as crosses in the figure).

The upper row of Fig. 5 shows the box-plots for the modal parameters estimated with the SSI algorithm, the second row corresponds to the EM1 method, and the third row corresponds to the EM2 method. By columns, the first one presents the box-plots for the natural frequencies, the second one shows the results for the damping ratio and the last column is reserved for the MAC value.

Considering first the SSI method, the estimated frequencies of modes 1, 2, 3, 11 and 12 are very similar and are close to the actual value (note that the difference is close to zero). The estimated frequencies for modes 4, 5, 8 and 10, however, present a huge variability. The frequencies of the modes 6, 7 and 9 show intermediate variability. This behavior is also

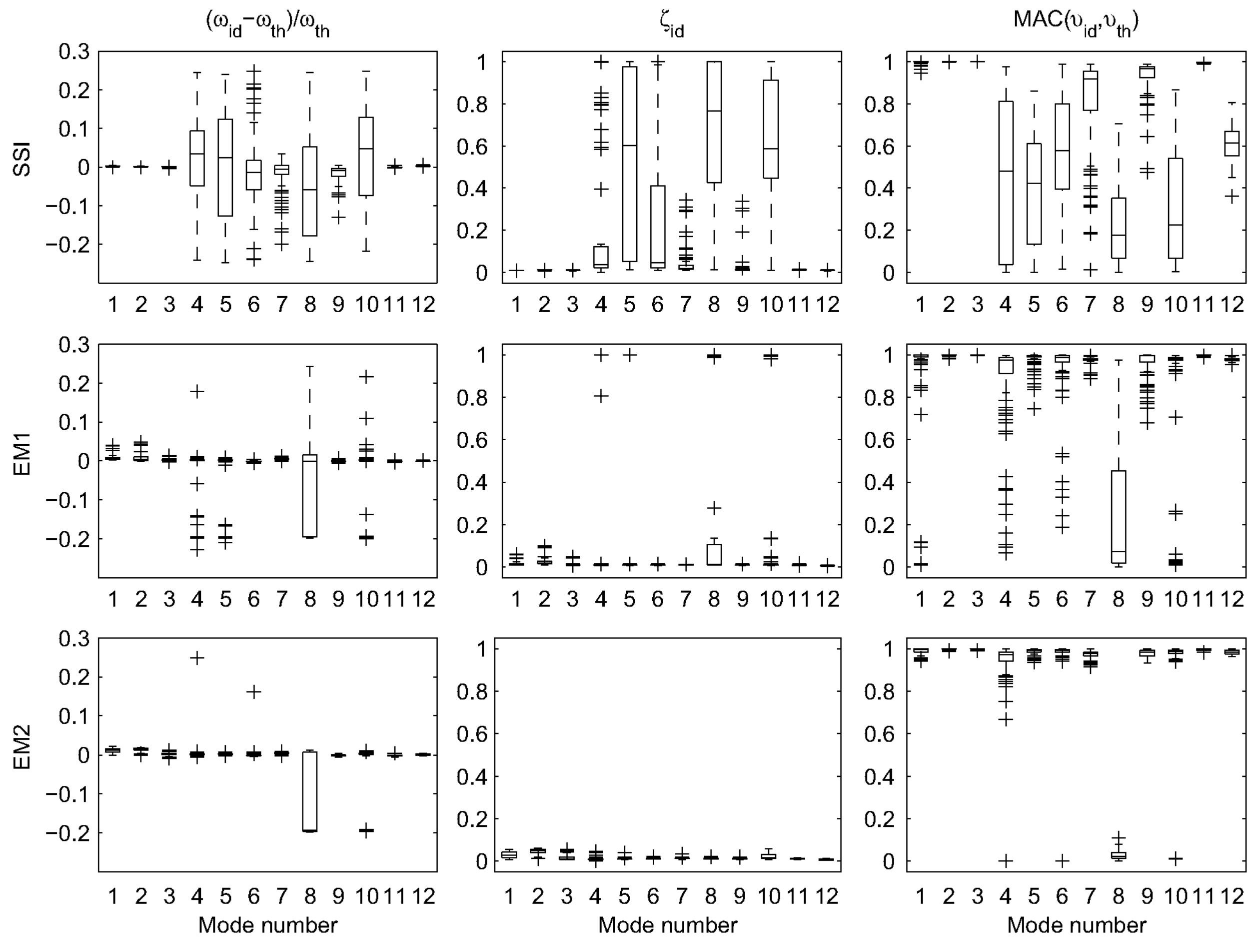

Fig. 5. Box-plots of SSI, EM1 and EM2 results for the 100 simulated cases.

observed in the plot corresponding to the damping ratio and the MAC value. The only difference is that the MAC obtained for mode 12 are about 0.6 and very far from the values that we consider acceptable (greater than 0.9).

Taking into account the results obtained with the EM1 method (second row in Fig. 5), we can notice the improvement achieved in the natural frequencies, damping ratios and MAC values corresponding to all the modes except mode 8. Excluding this mode, most of the estimated frequencies correspond to the theoretical values (although with some outliers), most of damping ratios are close to 0.01, and most of MAC values are close to one.

This improvement is even more evident when using the algorithm EM2, except for mode 8 and for some very few outliers, the results are very close in most of the simulations to the theoretical values.

Fig. 5 shows the whole estimated modal parameters for the 100 simulated responses. From these results we have extracted the modes that verify the criteria (36)–(38). We consider that these modes have been properly identified. Table 2 shows the average values and standard deviations of the modal frequencies, damping ratios and mode shapes. The table also includes the number of simulations, called N (over a total of 100), each mode has been identified. We have observed that:

- Using SSI method, we have identified four modes in the 100% of the simulations (modes 1, 2, 3 and 11) and two modes not so good (modes 7 and 9). The rest of the modes have not been identified. However, the modes estimated with this method have the best mean and standard deviation values of the three methods.
- Using method EM1 results are improved. We can say that all modes have been identified except mode 8. We have observed in the simulations that, in general, parameters well estimated by SSI method are not modified when using these parameters as starting point for the EM algorithm; but sometimes it can happen that, as the likelihood is maximized, some of them deteriorate lightly (as can be seen for modes 1 and 2) while the rest are improved. Usually the estimated values by the EM algorithm present bias estimation with $\omega_{th} < \omega_{id}$ and $\zeta_{th} < \zeta_{id}$, something known when the poles of matrix A are close to the unit circle [3].

Table 2
Modal parameters identified from 100 simulated time history responses using SSI method and EM method. *Std* is the standard value (given in %), and *N* is the number of times the parameter has been identified.

(a) Natural frequencies (Hz)

Mode	Exact	SSI			EM1			EM2		
		Mean	Std	*N*	Mean	Std	*N*	Mean	Std	*N*
1	9.41	9.44	0.03	100	9.48	0.17	89	9.53	0.28	100
2	11.79	11.81	0.03	100	11.86	0.37	94	11.95	0.48	100
3	16.38	16.41	0.05	100	16.42	0.24	100	16.42	0.45	100
4	25.54	25.88	0.47	7	25.62	0.30	73	25.60	0.29	85
5	32.01	–	–	–	32.08	0.37	94	32.08	0.35	100
6	38.66	38.30	1.53	11	38.63	0.36	88	38.63	0.50	100
7	44.64	44.66	2.01	53	44.87	0.57	99	44.89	0.57	100
8	48.01	–	–	–	48.15	0.34	3	–	–	–
9	48.44	48.22	1.54	67	48.42	0.46	90	48.38	0.78	100
10	60.15	–	–	–	60.32	0.50	87	60.35	0.97	96
11	67.48	67.49	0.37	100	67.45	0.53	100	67.42	1.06	100
12	83.62	–	–	–	83.62	0.61	100	83.70	0.90	100

(b) Damping ratios and mode shapes

Mode	ζ (%)							MAC					
	Exact	SSI		EM1		EM2		SSI		EM1		EM2	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	1.00	0.91	0.03	1.41	0.37	2.84	1.31	0.998	0.76	0.995	1.08	0.983	2.25
2	1.00	1.22	0.07	2.13	0.73	4.44	1.27	0.999	0.01	0.999	0.10	0.996	0.98
3	1.00	0.75	0.06	1.07	0.73	1.84	1.54	0.999	0.00	0.999	0.10	0.999	0.21
4	1.00	2.11	0.23	0.89	0.14	1.04	0.47	0.957	1.07	0.972	2.30	0.969	2.26
5	1.00	–	–	1.01	0.12	1.11	0.39	–	–	0.989	1.35	0.987	1.28
6	1.00	1.29	0.23	1.06	0.11	1.22	0.28	0.947	2.98	0.982	1.97	0.983	1.97
7	1.00	1.74	0.41	1.11	0.08	1.31	0.32	0.949	1.96	0.976	1.38	0.971	1.83
8	1.00	–	–	1.22	0.03	–	–	–	–	0.955	2.47	–	–
9	1.00	1.21	0.21	0.99	0.11	1.25	0.28	0.970	1.61	0.984	2.52	0.968	2.62
10	1.00	–	–	1.17	0.46	2.37	1.51	–	–	0.979	1.50	0.983	1.34
11	1.00	1.11	0.08	0.96	0.22	1.26	0.27	0.996	0.18	0.998	0.24	0.994	0.41
12	1.00	–	–	0.60	0.13	0.75	0.20	–	–	0.978	0.75	0.983	0.98

- The method EM2 provides the best results of the three methods considered in this work (in the sense of modal parameters).
- A remarkable fact in the three methods is that significantly larger values of standard deviation are observed for damping ratios than for modal frequencies in most cases. For this reason, we have preferred to take ε_ζ in absolute value while ε_ω is chosen in a relative sense in Eqs. (36) and (37).
- It is important to note that mode 8 is not identified by none of the three methods. We think this is because this mode could not be excited: even if we consider higher state space models, this mode is hard to find.
- In general terms, we have found that the estimates of modal parameters are improved after applying the EM algorithm to the SSI estimates. But sometimes it happens that the estimates of the first three modes become worse, although the likelihood is maximized. In principle, the properties of maximum likelihood estimation ensure that the estimates have asymptotically minimum variance under certain conditions, but this does not necessarily imply that for finite samples this property is maintained. Moreover, the proposed algorithm tries to minimize the errors as a whole, but reducing the variance of the whole can increase the variance of one parameter considered individually. This fact can be corrected in the algorithm, for example, using a weighted maximization of the likelihood, but then the computed solution is not the maximum likelihood solution.

5.2. *Experimental data*

The experimental phase of the benchmark problem was carried out in August 2002: the model structure was instrumented with three uniaxial accelerometers at each story level including the base. Nine configurations were investigated in order to simulate damage within the test structure by removing bracing and loosening beam-column connections (the configuration 1 was the reference – undamaged – case). For each configuration, experimental data were generated by three types of excitation: impacts of a sledge hammer, ambient vibration and electrodynamic shaker on the roof. We have analyzed the time history response corresponding to configuration 1 and ambient vibration, named "shm01a" in the data set obtained at the task group web site [6]. A detailed description of the test structure and experimental procedure can be found on this web site.

The duration of the acceleration data is 300 s with sampling frequency equal to 200 Hz, which have been filtered by a Butterworth high-pass filter with cut-off frequency equal to 0.1 Hz to eliminate the mean and drift. The base accelerations are excluded so the total number of channel is 12.

The stabilization diagram of the experimental data is presented in Fig. 6. This figure has been constructed using SSI method because it is fast and accurate enough to build this type of plots. Modal parameters that belong to two consecutive model orders have been compared according to the criteria expressed in Eqs. (36)–(38).

We have chosen a model order equal to $n_s=30$, because the principal stable modes are included for this order. Table 3 shows the experimental modal frequencies and damping ratios computed applying the same methods that in the simulated phase. If the values estimated with different methods are located in the same row of the table means that they correspond to the same mode, that is, verify the tolerances (36)–(38) between them.

The experimental vibration measurements of this structure have been also studied by various researchers applying different system identification techniques: for example, Ching and Beck [16] applied a two-stage Bayesian approach; and Alicioglu and Lus [17] used the SSI method too. The results of the two analysis are also presented in Table 3 for comparison

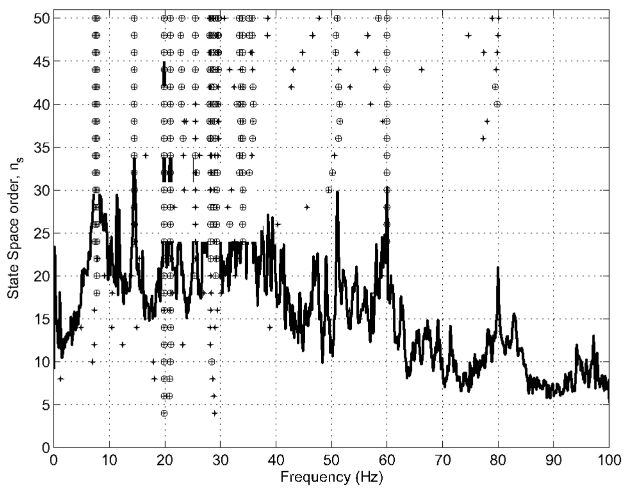

Fig. 6. Stabilization diagram obtained with SSI method. The used symbols are "$\oplus$" for stable pole and "+" for unstable pole. A mode is stable if $\varepsilon_\omega = 0.02$, $\varepsilon_z = 0.03$, $\varepsilon_{MAC} = 0.10$.

Table 3
Resulting modal parameters for the experimental time history response using SSI, EM1 and EM2 methods. n is the number of times that the parameter has been identified using EM2 from a total of 100. Results obtained in other studies are included.

Mode	SSI		EM1		EM2			Other studies	
	ω (Hz)	ζ (%)	ω (Hz)	ζ (%)	ω (Hz)	ζ (%)	N	Ref. [16]	Ref. [17]
1	7.52	1.04	7.55	1.46	7.55	2.12	36	7.48	7.49
2	7.77	0.60	7.81	1.53	7.83	2.10	85	7.76	7.76
3	14.53	0.44	14.54	0.78	14.90	4.32	48	14.48	14.49
4	19.89	0.01	19.89	0.01	19.89	0.02	100	19.89	19.89
5	21.01	0.04	21.08	0.52	21.23	1.11	100	21.01	21.01
6	23.49	85.42	–	–	–	–	–		22.69
7	25.49	0.09	25.62	0.47	25.88	2.69	67		25.49
8	28.16	2.12	–	–	–	–	–		28.31
9	28.26	0.09	28.37	0.32	28.40	0.49	87		
10	28.87	0.41	–	–	–	–	–		
11	–	–	29.04	2.12	28.96	3.74	37		
12	29.36	0.20	29.15	0.49	29.20	0.78	88		
13	–	–	30.44	4.34	31.50	6.45	26		
14	32.03	1.51	–	–	–	–	–		
15	–	–	32.14	1.85	32.19	2.33	52		
16	–	–	34.03	0.74	–	–	–		
17	34.04	0.11	–	–	33.31	2.45	36		
18	–	–	35.24	8.40	–	–	–		
19	49.51	2.94	–	–	–	–	–		
20	–	–	52.22	3.03	–	–	–		
21	59.99	0.03	59.99	0.02	59.84	0.37	69		
22	–	–	–	–	81.67	1.39	19		

purpose (in this case the frequencies have been placed in the table in ascending order and not following the criteria (36)–(38) because we do not have their corresponding mode shapes and damping ratios).

Comprehensive analyzes of Table 3 and Fig. 6 show the following:

- The lower frequencies (first five modes) have been identified by the three algorithms, and they also appear in the two references mentioned above [16,17]. Moreover, they are very clear modes in the spectrum and the stabilization diagram.
- Modes 7, 9, 12 and 21, which are higher frequency modes, have been identified by the three methods. None of them appears in [16] and two could coincide with the frequencies given in [17]. The stabilization diagram indicates the existence of stable modes in these frequency ranges, although it is difficult to specify more graphic information.
- In total, the three methods agree on the identification in 9 of the 15 modes that the model order ($n_s=30$) allowed us to estimate.
- Modes 6, 8, 10, 14 and 19 have been identified by the SSI method but they are not confirmed by any of the two versions of the EM method. The stabilization diagram is of little assistance. On the other hand, Table 3 shows the existence of close frequency values identified by the EM algorithms, but they differ in the damping or the eigenvector. We have checked the solutions obtained with higher orders (up to order 50) for the SSI method and the problem worsens: the new modes are still not mismatched with those of the EM algorithm and appear many new ones without correspondence.
- There are other three modes that have been identified for EM1 and EM2 jointly, modes 11, 13 and 15. In general terms, the modes estimated with EM2 for a larger number of starting points have been also identified either by one of the other methods or by both at the same time. This is a very valuable property of EM2 method.
- Modes 16, 18 and 20 have been only detected by the EM1 method. Finally, the natural frequency near to 80 Hz has been estimated only by the algorithm EM2. However, we can see in the stabilization diagram (Fig. 6) that this mode exits and it is identified by the SSI method when using higher state space orders.
- The modal analysis of a structure is a complex problem and we should not take the results provided by only one method as the *exact* solution. We can have errors of two types: to include spurious modes and to forget others which really exist. This analysis shows that having multiple solutions is very useful for a better identification of the vibration modes of a structure.

6. Conclusions

The aim of this paper was to present the application of a time-domain stochastic system identification method based on the EM algorithm to operational modal analysis of structures. The EM algorithm is a well-known tool for iterative maximum likelihood estimation which for state space models has a particularly neat form. Due to the EM algorithm is iterative, it is needed a starting value. When dealing with multivariate problems, like state space equations, the construction of a starting value is not trivial. Not any starting value works, and most of them failed in the iterative process. In our opinion, the easiest approach is to begin with the estimates of another identification method. We have chosen SSI because it is very much used in output-only problems, and it is fast and robust. On the other hand, we have developed a procedure to generate feasible random starting points as well.

The proposed method has been evaluated through a numerical and an experimental study in the context of the ASCE benchmark problem. The numerical results show that the proposed method estimates natural frequencies, damping ratios and mode shapes reasonably well in the presence of measurement noises even.

From our experiments, the following comments can be made:

- SSI is a powerful identification method for output only analysis and has numerous advantages. In fact is one of the most currently used method to perform operational modal analysis of structures. However, maximizing the likelihood in the state space model is equivalent to minimizing the error, that is, the state space model fitted by the EM algorithm to the data has a lower error than the SSI one. So we expect that the modal parameters (which are derived from the state space model) have a lower error as well. And the results of the examples seem to support this fact.
- But using a single run of the EM can lead to suboptimal solutions. For this reason we have proposed a procedure for building random initial values for the EM. This is not trivial because we are working with multivariate systems and any value taken at random does not work. Generating different starting points we can pick the *optimal solution* between them (we choose the parameters that appear again and again at different starting points). From the point of view of modal parameters, we have obtain the best results using the EM2 method.
- Although not crucial, the random stating values allow us to use the EM method independently, without recourse to another identification method.

The paper shows that the EM algorithm can be used to compute the maximum likelihood estimate of the modal parameters of a structure using output-only measurements. This is important because the maximum likelihood parameters have well-known statistical properties.

Finally, we would like to finish with some other examples where we have found profitable the use of the EM algorithm:

- In structures, the value of the modal parameters is not known. Applying the SSI method we compute an estimate of the parameters in a least square sense. On the other hand, the estimate given by the EM algorithm is based on the maximum likelihood method. The joint analysis of both solutions, which are based on a sound theoretical basis, is a valuable tool for modal identification.
  For example, the separation between structural modes and spurious modes in stabilization diagrams is not always straightforward: in built structures there are many closely spaced modes, and the stabilization diagrams are not clear at these frequencies. If we apply the EM1 method, all the estimates present simultaneously in the SSI solution and in the EM1 solution can be selected as modal parameters of the structure, because they have been identified by SSI and kept after the EM iterations. So the EM algorithm can help us to find structural modes where the stabilization diagram is not clear.
- The stabilization diagrams built using EM1 are clearer than SSI ones, and the stables poles appear at lower orders.
- We always refine the SSI solution using the EM1 method before *gluing* the modes estimated from different setups of sensors in a structure. The number and quality of modes obtained is improved.
- The solution given by the EM2 method does not include spurious modes, because it is difficult for the same spurious mode to appear at different starting points. But it is very time consuming. We recommend using it when we want more confidence in our results, or for large and complicated structures where some modes cannot be easily extracted.

We think that the strategies we have adopted in this work can be useful when performing an operational modal analysis and can be applied without difficulty.

Acknowledgments

This research was supported by the Ministerio de Educación y Ciencia of Spain under the research project BIA2008-04089 *Vibraciones estructurales inducidas por la actividad humana*. The financial support is gratefully acknowledged.

Appendix A. From second-order structural models to state space models

The equations of motion for an n_d degrees-of-freedom linear, time invariant, viscously damped system subjected to external excitation is expressed as

$$M\ddot{q}(t)+C_z\dot{q}(t)+Kq(t)=Ju(t), \tag{A.1}$$

where $M, C_z, K \in \mathbb{R}^{n_d \times n_d}$ are the mass, damping and stiffness matrices, respectively; $J \in \mathbb{R}^{n_d \times n_i}$ is the excitation influence matrix that relates the n_i-dimensional input vector $u(t)$ to the n_d-dimensional response vector; $q(t)$ is the n_d-dimensional displacement response vector; dot denotes derivatives with respect to time.

By defining the state vector $x(t) = [q(t)\ \dot{q}(t)]^T$, Eq. (A.1) can be converted into the continuous state space form:

$$\dot{x}(t) = A_c x(t) + B_c u(t), \tag{A.2}$$

where

$$A_c = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C_z \end{bmatrix} \quad B_c = \begin{bmatrix} 0 \\ M^{-1}J \end{bmatrix}. \tag{A.3}$$

In practice, only a limited number of measurements are available; therefore, the dimension of the measurement output is less than or equal to the total number of degrees of freedom. The n_o-dimensional output vector $y(t)$ can be expressed as

$$y(t) = C_a \ddot{q}(t), \tag{A.4}$$

where $C_a \in \mathbb{R}^{n_o \times n_d}$ is the measurement location matrix corresponding to the acceleration responses of the structural system, composed by zeros and ones. We can rewrite the output vector into the continuous state space form:

$$y(t) = C_c x(t) + D_c u(t), \tag{A.5}$$

where

$$C_c = C_a[-M^{-1}K\ \ -M^{-1}C_z] \quad D_c = C_a M^{-1} J. \tag{A.6}$$

Eqs. (A.2) and (A.5) define the state space equation in continuous time:

$$\dot{x}(t) = A_c x(t) + B_c u(t), \tag{A.7a}$$

$$y(t) = C_c x(t) + D_c u(t), \tag{A.7b}$$

where $y(t) \in \mathbb{R}^{n_o}$ is the measured *output vector*; $u(t) \in \mathbb{R}^{n_i}$ is the measured *input vector*; $x(t) \in \mathbb{R}^{n_s}$ is the *state vector*; $A_c \in \mathbb{R}^{n_s \times n_s}$ is the *transition state matrix* describing the dynamics of the system; $B_c \in \mathbb{R}^{n_s \times n_i}$ is the *input matrix*; $C_c \in \mathbb{R}^{n_o \times n_s}$ is the *output matrix*, which is describing how the internal state is transferred to the output measurements $y(t)$; $D_c \in \mathbb{R}^{n_o \times n_i}$ is the *direct transmission matrix*;

Eq. (A.7a) is known as the *State Equation* and Eq. (A.7b) is known as the *Observation Equation*.

It is important to say the state vector of the system is not unique and we can transform the state vector $x(t)$ into $z(t)$ by a linear transformation as follows:

$$x(t) = T_1 z(t), \tag{A.8}$$

where T_1 is the transformation matrix. Replacing this condition into (A.7) and pre-multiplying by T_1^{-1}

$$z(t) = A_{c1} z(t) + B_{c1} u(t),$$

$$y(t) = C_{c1} z(t) + D_{c1} u(t),$$

where

$$A_{c1} = T_1^{-1} A_c T_1, \quad C_{c1} = C_c T_1, \quad B_{c1} = T_1^{-1} B_c, \quad D_{c1} = D_c. \tag{A.10}$$

This state representation yields the same dynamic relation between $u(t)$ and $y(t)$, that is, the same input–output behavior that (A.7). An important state vector is the formed with the n_d-dimensional modal coordinates vector, $\eta(t)$

$$q(t) = \Phi \eta(t) \Rightarrow z(t) = \begin{bmatrix} \eta(t) \\ \dot{\eta}(t) \end{bmatrix} \Rightarrow T_1 = \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix}, \tag{A.11}$$

where Φ is the eigenvector matrix of $M^{-1}K$ matrix, which verifies the *orthogonality properties*:

$$\Phi^T M \Phi = M_m,$$

$$\Phi^T K \Phi = K_m.$$

M_m and K_m are called *modal mass* and *modal stiffness* matrices and both are diagonal. Using Eqs. (A.3), (A.6) and (A.11), A_{c1} and C_{c1} become

$$A_{c1} = \begin{bmatrix} 0 & I \\ -\Phi^{-1}M^{-1}K\Phi & -\Phi^{-1}M^{-1}C_z\Phi \end{bmatrix}, \tag{A.12}$$

$$C_{c1} = C_a\Phi[-\Phi^{-1}M^{-1}K\Phi \;\; -\Phi^{-1}M^{-1}C_z\Phi]. \tag{A.13}$$

Taking into account the orthogonal properties and matrix inverse properties

$$\Phi^{-1}M^{-1}K\Phi = \Phi^{-1}M^{-1}(\Phi^T)^{-1}\Phi^T K\Phi = (M\Phi)^{-1}(\Phi^T)^{-1}\Phi^T K\Phi = (\Phi^T M\Phi)^{-1}\Phi^T K\Phi = M_m^{-1}K_m = \Omega^2,$$

where Ω is a diagonal matrix formed with the natural frequencies. In matrix form

$$\Omega = \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \omega_{n_d} \end{bmatrix}. \tag{A.14}$$

Applying the same procedure to the other component of the matrices, and assuming proportional damping:

$$\Phi^{-1}M^{-1}C_z\Phi = 2\Omega Z,$$

where Z is a diagonal matrix formed with the *damping ratios* of each vibrational mode:

$$Z = \begin{bmatrix} \zeta_1 & 0 & \dots & 0 \\ 0 & \zeta_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \zeta_{n_d} \end{bmatrix}. \tag{A.15}$$

Hence, substituting the above formulas into A_{c1} and C_{c1} results

$$A_{c1} = \begin{bmatrix} 0 & I \\ -\Omega^2 & -2\Omega Z \end{bmatrix}, \quad C_{c1} = C_a\Phi[-\Omega^2 \;\; -2\Omega Z], \tag{A.16}$$

$$B_{c1} = \begin{bmatrix} 0 \\ \Phi^{-1}J \end{bmatrix}, \quad D_{c1} = D_c, \tag{A.17}$$

However, the continuous state space formulation is only useful to theoretical considerations because measurements are taken in discrete time instants, so equations must be expressed in the discrete-time state space form. For the sampling of a continuous-time equation we consider the input is piecewise constant over the sampling period of length Δt (Zero-Order Hold assumption), that is

$$\forall t \in [t_k, t_{k+1}) = [k\Delta t, (k+1)\Delta t) \Rightarrow x(t) = x(t_k) = x_k, \quad u(t) = u(t_k) = u_k, \quad y(t) = y(t_k) = y_k \tag{A.18}$$

Under this assumption, the continuous time state-space model (A.7a) and (A.7b) is converted to the discrete time state-space model:

$$x_{k+1} = Ax_k + Bu_k, \tag{A.19a}$$

$$y_k = Cx_k + Du_k. \tag{A.19b}$$

The parameters are related to their continuous-time counterparts as (see for instance [18]):

$$A = e^{A_c\Delta t}, \quad B = (A-I)A_c^{-1}B_c, \quad C = C_c, \quad D = D_c. \tag{A.20}$$

Up to now it has been considered that the system was only subjected to deterministic and known inputs, u_k. However, besides this measured inputs always there are other that, although they are not known, contribute to the system response. This unmeasurable influence is characterized as disturbance or noise. In system identification, system response disturbance might be caused by different phenomena. In any case, noise will always be present in measured data and should be therefore always take into account.

It is necessary to extend the state space model (A.19a) and (A.19b) including stochastic components, so *stochastic state space model* is obtained:

$$x_{k+1} = Ax_k + Bu_k + w_k, \tag{A.21a}$$

$$y_k = Cx_k + Du_k + v_k, \tag{A.21b}$$

where $w_k \in \mathbb{R}^{n_s}$ is the *process noise* due to disturbances and modeling inaccuracies; $v_k \in \mathbb{R}^{n_o}$ is the measurement noise due to sensor inaccuracy. We assume they are both independent and identically distributed, zero-mean normal vectors:

$$w_k \rightsquigarrow N(0,Q) \quad u_k \rightsquigarrow N(0,R). \tag{A.22}$$

Appendix B. Kalman filtering and smoothing

We present here the Kalman filter, which provides the basis for the equations of the algorithm. The Kalman filter is a well-known and used tool but we include it here to complete. We have also included two additional properties that are needed in the equations.

B.1. Notation for the Kalman filter

The following notation has been used in all the expressions derived from the Kalman filter. Given the output data for s time steps $Y_s = \{y_1, y_2, \ldots, y_s\}$, it is defined as follows:

$$x_t^s = E[x_t | Y_s],$$

$$P_{t_1,t_2}^s = E[(x_{t_1} - x_{t_1}^s)(x_{t_2} - x_{t_2}^s)^T | Y_s],$$

where $E[\bullet|\bullet]$ is the conditional expected operator.

When $t_1 = t_2 = t$ it will be written P_t^s:

$$P_t^s = E[(x_t - x_t^s)(x_t - x_t^s)^T | Y_s] = \mathrm{Var}[x_t | Y_s].$$

B.2. Three basic properties

Property 6 (*The Kalman Filter*). *For the state space model specified in* (2) *with initial conditions* $x_0^0 = \mu_0$ *and* $P_0^0 = \Sigma_0$, *for* $t = 1, 2, \ldots, N$,

$$x_t^{t-1} = A x_{t-1}^{t-1}, \tag{B.1}$$

$$P_t^{t-1} = A P_{t-1}^{t-1} A^T + Q, \tag{B.2}$$

with

$$x_t^t = x_t^{t-1} + K_t \epsilon_t, \tag{B.3}$$

$$P_t^t = [I - K_t C] P_t^{t-1}, \tag{B.4}$$

where

$$K_t = P_t^{t-1} C^T \Sigma_t^{-1}, \tag{B.5}$$

$$\epsilon_t = y_t - E[y_t | Y_{t-1}] = y_t - C x_t^{t-1}, \tag{B.6}$$

$$\Sigma_t = \mathrm{Var}[\epsilon_t] = \mathrm{Var}[C(x_t - x_t^{t-1}) + v_t] = C P_t^{t-1} C^T + R, \tag{B.7}$$

K_t *is called the Kalman gain and* ϵ_t *are the innovations.*

Property 7 (*The Kalman Smoother*). *For the state space model specified in* (2) *with initial conditions* x_N^N *and* P_N^N *obtained via Property* 6, *for* $t = N, N-1, \ldots, 1$,

$$x_{t-1}^N = x_{t-1}^{t-1} + J_{t-1}(x_t^N - x_t^{t-1}), \tag{B.8}$$

$$P_{t-1}^N = P_{t-1}^{t-1} + J_{t-1}(P_t^N - P_t^{t-1}) J_{t-1}^T, \tag{B.9}$$

where

$$J_{t-1} = P_{t-1}^{t-1} A^T [P_t^{t-1}]^{-1}. \tag{B.10}$$

Property 8 (*The Lag-One Covariance Smoother*). *For the state space model specified in* (2), *with* K_t, J_t $(t = 1, 2, \ldots, N)$, *and* P_N^N *obtained from Properties* 6 *and* 7, *with initial condition*

$$P_{N,N-1}^N = (I - K_N C) A P_{N-1}^{N-1} \tag{B.11}$$

for $t = N, N-1, \ldots, 2$

$$P_{t-1,t-2}^N = P_{t-1}^{t-1} J_{t-2}^T + J_{t-1}(P_{t,t-1}^N - A P_{t-1}^{t-1}) J_{t-2}^T. \tag{B.12}$$

The demonstration of the above properties can be found in [3].

References

[1] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B (Methodological) 39 (1) (1977) 1–38.

[2] R.H. Shumway, D.S. Stoffer, An approach to time series smoothing and forecasting using the EM algorithm, J. Time Ser. Anal. 3 (4) (1982).

[3] R.H. Shumway, D.S. Stoffer, Time Series Analysis and its Applications, Springer, 2006.

[4] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, second ed., Wiley, New York, 2002.

[5] G.J. McLachlan, T. Krishnam, The EM Algorithm and Extensions, 2nd ed., Wiley, New York, 2008.

[6] IASC-ASCE SHM Task Group. Task Group website (last visit on May 2010): 〈http://bc029049.cityu.edu.hk/asce.shm/wp_home.asp〉.

[7] C. Biernacki, G. Celeux, G. Govaert, Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models, Comput. Stat. Data Anal. 41 (2003) 561–575.

[8] D. Karlis, E. Xekelaki, Choosing initial values for the EM algorithm for finite mixtures, Comput. stat. Data Anal. 41 (2003) 577–590.

[9] L. Ljung, System Identification. Theory for the Users, 2nd ed., PTR Prentice-Hall, Upper Saddle River, NJ, 1999.

[10] H. Akaike, Stochastic theory of minimal realization, IEEE Trans. Autom. Control 16 (6) (1974) 667–674.

[11] P. Van Overschee, B. De Moor, Subspace Identification for Linear Systems. Theory—Implementation—Applications, Kluwer Academic Publishers, 1996.

[12] D.E. Newland, Mechanical Vibrations Analysis and Computation, Dover Publications, 2006.

[13] B. Peeters, System Identification and Damage Detection in Civil Engineering, Ph.D. Thesis, Katholieke Universiteit Leuven, 2000.

[14] E.A. Johnson, H.F. Lam, L.S. Katafygiotis, J.L. Beck, Phase I IASC—ASCE structural health monitoring benchmark problem using simulated data, J. Eng. Mech. 130 (1) (2004) 315.

[15] C.M. Hurvich, C.L. Tsai, Regression and time series model selection in small samples, Biometrika 76 (1989) 297–307.

[16] J. Ching, J.L. Beck, Bayesian analysis of the Phase II IASC-ASCE structural health monitoring experimental benchmark data, J. Eng. Mech. 130 (10) (2004) 1233–1244.

[17] B. Alicioglu, M. Lus, Ambient vibration analysis with subspace methods and automated mode selection: case studies, J. Struct. Eng. ASCE (June) .

[18] J. Juang, Applied System Identification, Prentice Hall PTR, 1994.