Proceedings of the 8th International Conference on Structural Dynamics, EURODYN 2011
Leuven, Belgium, 4-6 July 2011
G. De Roeck, G. Degrande, G. Lombaert, G. Müller (eds.)
ISBN 978-90-760-1931-4

Operational Modal Analysis using Expectation Maximization Algorithm

F.J. Cara[1], J. Carpio[1], J. Juan[1], E. Alarcon[2]
[1]Department of Organization Engineering, Business Administration and Statistics,
E.T.S.I. Industriales U.P.M., Jose Gutierrez Abascal 2, 28080 Madrid, Spain
[2]Department of Structural Mechanics and Industrial Constructions
E.T.S.I. Industriales U.P.M., Jose Gutierrez Abascal 2, 28080 Madrid, Spain
email: fjcara@etsii.upm.es, jaime.carpio@upm.es, jjuan@etsii.upm.es, enrique.alarcon@upm.es

ABSTRACT: This paper presents a time-domain stochastic system identification method based on Maximum Likelihood Estimation and the Expectation Maximization algorithm. The effectiveness of this structural identification method is evaluated through numerical simulation in the context of the ASCE benchmark problem on structural health monitoring. Modal parameters (eigenfrequencies, damping ratios and mode shapes) of the benchmark structure have been estimated applying the proposed identification method to a set of 100 simulated cases. The numerical results show that the proposed method estimates all the modal parameters reasonably well in the presence of 30% measurement noise even. Finally, advantages and disadvantages of the method have been discussed.

KEY WORDS: System identification in structures, State Space models, Kalman filter, Operational Modal analysis, Maximum likelihood estimation, Expectation Maximization algorithm.

1 INTRODUCTION

The application of system identification to vibrating structures consist in identifying a modal model (eigenfrequencies, damping ratios and mode shapes) from vibration data. Classically, a measurable input is applied to the system and the output is measured. From these experimental data, a system model can be obtained by a variety of parameter estimation methods, and it is known as experimental modal analysis. However, cases exist where it is practically impossible to measure the excitation and the outputs are the only information that is passed to the system identification algorithms. In these cases the deterministic knowledge of the input is replaced by the assumption that the input is a realization of a stochastic process (white noise), and it is known as stochastic system identification (the terms output-only modal analysis and operational modal analysis are used as well).

Parametric structural identification methods involve the use of mathematical models to represent structural system behaviour in either time or frequency domain. The benefits of using parametric models for structural identification include their direct relationship with physically meaningful quantities such as stiffness and mass, improved accuracy and resolution, and their suitability for analysis, prediction, fault diagnosis and control.

Popular time domain parametric models used for structural identification purposes include: ARX models, ARMAX models, state space models, etc. Many identification algorithms are available to estimate the parameters of such parametric models, e.g. prediction error method (PEM), least squares estimation (LSE), maximum likelihood estimation (MLE), eigensystem realization algorithm (ERA) and stochastic subspace identification method (SSI).

This paper presents a time-domain stochastic system identification method based on Maximum Likelihood Estimation (MLE) and the Expectation Maximization (EM) algorithm. The results of this structural identification method are evaluated through numerical simulation in the context of the ASCE benchmark problem on structural health monitoring [1]. The modal parameters (eigenfrequencies, damping ratios and mode shapes) of the benchmark structure (see Figure 1) have been estimated using the proposed MLE+EM method and then have been compared to the exact values.

2 STATE-SPACE MODEL

2.1 Stochastic state-space equations

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

$$M\ddot{z}(t) + C_\zeta \dot{z}(t) + Kz(t) = Ju(t) \tag{1}$$

where $M, C_\zeta, K \in \mathbb{R}^{n_d \times n_d}$ 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; $z(t)$ is the n_d-dimensional displacement response vector; dot denotes taking derivatives with respect to time.

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

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

where

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

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

$$y(t) = C_d(t)z(t) + C_v\dot{z}(t) + C_a\ddot{z}(t) \tag{4}$$

where $C_d, C_v, C_a \in \mathbb{R}^{n_o \times n_d}$ are the measurement location matrices corresponding to the displacement, velocity and acceleration responses of the structural system, respectively. We can rewrite the output vector into the continuous state space form,

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

where

$$C_c = [C_d - C_a M^{-1}K \quad C_v - C_a M^{-1}C_\zeta] \quad D_c = C_a M^{-1}J \tag{6}$$

In this work, only accelerations are considered, so

$$C_c = C_a[-M^{-1}K \quad -M^{-1}C_\zeta] \tag{7}$$

Equations 2 and 5 define the state space equation in continuous time:

$$\dot{x}(t) = A_c x(t) + B_c u(t) \tag{8a}$$
$$y(t) = C_c x(t) + D_c u(t) \tag{8b}$$

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 the output measurements $y(t)$;
$D_c \in \mathbb{R}^{n_o \times n_i}$ is the *direct transmission matrix*;
Equation (8a) is known as the *State Equation* and equation (8b) is known as the *Observation Equation*.

But measurements are taken in discrete time instants, so equations must be expressed in discrete time too. Typical for the sampling of a continuous-time equation is a Zero-Order Hold assumption, which means that the input is piecewise constant over the sampling period, that is

$$\forall t \in [t_k, t_{k+1}) = [k\Delta t, (k+1)\Delta t) \Longrightarrow$$
$$x(t) = x(t_k) = x_k,\ u(t) = u(t_k) = u_k,\ y(t) = y(t_k) = y_k \tag{9}$$

Under this assumption, the continuous time state-space model (8a) and (8b) is converted to the discrete time state-space model:

$$x_{k+1} = Ax_k + Bu_k \tag{10a}$$
$$y_k = Cx_k + Du_k \tag{10b}$$

where x_k is the discrete time state vector containing the sampled displacements and velocities; u_k and y_k are the sampled input and output; A is the discrete state matrix; B is the discrete input matrix; C is the discrete output matrix; D is the discrete direct transmission matrix. They are related to their continuous-time counterparts as (see for instance [2]):

$$A = e^{A_c \Delta t} \tag{11}$$

$$B = (A - I)A_c^{-1}B_c \tag{12}$$

$$C = C_c \tag{13}$$

$$D = D_c \tag{14}$$

Up to now it was assumed that the system was only driven by a deterministic input u_k. However, besides this applied input there might be other inputs that in a more uncontrollable way 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. The most obvious one is noise generated by the sensors, or noise arising from roundoff errors during A/D conversion. 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 (10a) and (10b) including stochastic components, so *stochastic state space model* is obtained:

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{15a}$$
$$y_k = Cx_k + Du_k + v_k \tag{15b}$$

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

$$w_k \rightsquigarrow N(0, Q) \qquad u_k \rightsquigarrow N(0, R) \tag{16}$$

In the case of ambient vibration testing, only the responses of the structure y_k are measured, while the input sequence u_k remains unmeasured. Equations (15a) and (15b) result now in a purely stochastic system:

$$x_{k+1} = Ax_k + w_k \tag{17a}$$
$$y_k = Cx_k + v_k \tag{17b}$$

The input is now implicitly modeled by the noise terms w_k, v_k. However the white noise assumptions of these noise terms cannot be omitted and (16) remain still applicable in equation (17).

2.2 The Kalman filter

Due to the noise present in the stochastic state space Equations (17), it is only possible to predict the response in term of probability. For state space systems, this prediction is accomplished by the construction of the associated Kalman filter.

The Kalman filter is a computational scheme to estimate the states of a given state-space model in a statistically optimal manner. This filter is derived in a framework based on conditional probability theory, and the theoretical foundations used from statistics are complex.

The following notation will be used in all the expressions where the Kalman filter is involved. Given the output data for s time steps $Y_s = \{y_1, y_2, \ldots, y_s\}$, it is defined:

$$x_t^s = \mathrm{E}\left[\mathrm{x_t}|\mathrm{Y_s}\right]$$

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

where $\mathrm{E}[\bullet|\bullet]$ is the conditional expected operator. When $t_1 = t_2 = t$, $P^s_{t_1,t_2}$ will be written P^s_t:

$$P^s_t = \mathrm{E}\left[(\mathrm{x_t} - \mathrm{x^s_t})(\mathrm{x_t} - \mathrm{x^s_t})^{\mathrm{T}}|\mathrm{Y_s}\right] = \mathrm{Var}\left[\mathrm{x_t}|\mathrm{Y_s}\right]$$

Property 1 (The Kalman Filter) For the state space model specified in (17) with initial conditions $x^0_0 = \mu_0$ and $P^0_0 = \Sigma_0$, for $k = 1, 2, \ldots, N$,

$$x^{k-1}_k = Ax^{k-1}_{k-1} \tag{18}$$

$$P^{k-1}_k = AP^{k-1}_{k-1}A^T + Q \tag{19}$$

with

$$x^k_k = x^{k-1}_k + K_k\varepsilon_k \tag{20}$$

$$P^k_k = (I - K_kC)P^{k-1}_k \tag{21}$$

where

$$K_k = P^{k-1}_k C^T \Sigma^{-1}_k \tag{22}$$

$$\varepsilon_k = y_k - \mathrm{E}[\mathrm{y_k}|\mathrm{Y_{k-1}}] = \mathrm{y_k} - \mathrm{Cx^{k-1}_k} \tag{23}$$

$$\Sigma_k = \mathrm{Var}(\varepsilon_\mathrm{k}) = \mathrm{Var}[\mathrm{C}(\mathrm{x_k} - \mathrm{x^{k-1}_k}) + \mathrm{v_k}] = \mathrm{CP^{k-1}_k C^T} + \mathrm{R} \tag{24}$$

K_k is called the Kalman gain and ε_k are the innovations.

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

2.3 The innovation form representation

An alternative model for stochastic systems that is more suitable for some applications is the so-called *innovation form representation*. It is obtained applying the the Kalman filter to the stochastic state-space model (17):

$$x^{k-1}_k = Ax^{k-1}_{k-1} = A\left(x^{k-2}_{k-1} + K_{k-1}\varepsilon_{k-1}\right) = Ax^{k-2}_{k-1} + AK_{k-1}\varepsilon_{k-1} \tag{25}$$

so

$$x^k_{k+1} = Ax^{k-1}_k + AK_k\varepsilon_k \tag{26a}$$

$$y_k = Cx^{k-1}_k + \varepsilon_k \tag{26b}$$

where K_k is the Kalman gain (22) and ε_k are the innovations (23).

Under stationary conditions (see for instance [4]),

$$\lim_{k\to\infty} P^{k-1}_k = P > 0 \tag{27}$$

and the innovation form representation converges to the following time-invariant state-space system:

$$x^k_{k+1} = Ax^{k-1}_k + K\varepsilon_k \tag{28a}$$

$$y_k = Cx^{k-1}_k + \varepsilon_k \tag{28b}$$

where

$$P = APA^T + Q - APC^T(CPC^T + R)^{-1}(APC^T)^T \tag{29}$$

$$K = APC^T(CPC^T + R)^{-1} \tag{30}$$

Equation 29 is called a discrete algebraic Riccati equation (DARE) and it is a steady-state version of Equation 19.

These stationary conditions are satisfied in linear time invariant (LTI) systems. The state space equations considered in this work for operational modal analysis have been (17) and (28).

2.4 System identification and modal analysis in a state-space model

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 low and proportional damping, the second order modes are uncoupled and the jth eigenvalue of A has the form

$$\lambda_j = \exp\left(\left(-\zeta_j\omega_j \pm i\omega_j\sqrt{1-\zeta^2_j}\right)\Delta t\right) \tag{31}$$

where ω_j are the natural frequencies, ξ_j are damping ratios, and Δt is the time step. Natural frequencies ω_j and the damping ratios ξ_j are given by

$$\omega_j = \frac{\left|\ln(\lambda_j)\right|}{\Delta t} \tag{32}$$

$$\zeta_j = \frac{-Real\left[\ln(\lambda_j)\right]}{\omega_j\Delta t} \tag{33}$$

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

$$\phi_j = C\psi_j \tag{34}$$

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

3 MAXIMUM LIKELIHOOD METHOD WITH EM ALGORITHM

In this section is presented the algorithm for estimating the parameters of the stochastic state space model given by Equation (17), which is based on the maximum likelihood method. This method try to maximize the likelihood applying the iterative expectation maximization algorithm (EM).

3.1 Maximum likelihood Estimation

Given N measurements of the outputs $Y_N = \{y_1, y_2, \ldots, y_N\}$, one way to compute the likelihood is using the innovations $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$, defined by Equation 23. The innovations are independent Gaussian random vectors, $\varepsilon_k \rightsquigarrow N(0, \Sigma_k)$, with covariance matrix Σ_k given by Equation 24. Thus, ignoring a constant, the logarithm of the likelihood computed from the innovations may be written as:

$$l_{Y_N}(\theta) = -\frac{1}{2}\sum_{t=1}^{N}\left(\ln|\Sigma_k(\theta)| + \varepsilon_k(\theta)^T\Sigma_k(\theta)^{-1}\varepsilon_k(\theta)\right) \tag{35}$$

where it has been emphasized the dependence of the innovations on the vector θ, which represent the unknown parameters of the

model (17) under the assumption that 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 (35), 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

In this section it is outlined the basis of the method, but a more complete description can be found in [3] and [5]. The basic idea is that if the states could be observed $X_N = \{x_0, x_1, x_2, ..., x_N\}$, in addition to the observations, $Y_N = \{y_1, y_2, \ldots, y_N\}$, then the complete data could be considered. The logarithm of the likelihood of the complete data can be expressed as

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

But $l_{X_N,Y_N}(\theta)$ and $l_{X_N|Y_N}(\theta)$ are function of the unknown states X_N, so they are replaced with its expected values. Given a value for the parameter θ at step j it is defined

$$Q(\theta|\theta_j) = \mathrm{E}[\mathrm{l}_{\mathrm{X_N,Y_N}}(\theta)|\mathrm{Y_N}, \theta_\mathrm{j}]$$
$$R(\theta|\theta_j) = \mathrm{E}[\mathrm{l}_{\mathrm{X_N|Y_N}}(\theta)|\mathrm{Y_N}, \theta_\mathrm{j}]$$
$$S(\theta|\theta_j) = \mathrm{E}[\mathrm{l}_{\mathrm{Y_N}}(\theta)|\mathrm{Y_N}, \theta_\mathrm{j}]$$

Thus

$$S(\theta_j|\theta_j) = Q(\theta_j|\theta_j) - R(\theta_j|\theta_j)$$
$$S(\theta_{j+1}|\theta_j) = Q(\theta_{j+1}|\theta_j) - R(\theta_{j+1}|\theta_j)$$

Subtracting both equations

$$S(\theta_{j+1}|\theta_j) - S(\theta_j|\theta_j) =$$
$$\left[Q(\theta_{j+1}|\theta_j) - Q(\theta_j|\theta_j)\right] - \left[R(\theta_{j+1}|\theta_j) - R(\theta_j|\theta_j)\right]$$

It can be probed that

$$R(\theta_{j+1}|\theta_j) - R(\theta_j|\theta_j) \leq 0 \quad \forall j = 1, 2, \ldots$$

So if we develop a procedure which verifies

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

then automatically it is verified

$$S(\theta_{j+1}|\theta_j) - S(\theta_j|\theta_j) \geq 0$$

and we have a maximum for $l_{Y_N}(\theta)$ (Equation (35)).

In conclusion, the Expectation Maximization algorithm provides an iterative method for finding the maximum likelihood estimators of θ by successively maximizing the conditional expectation of the complete likelihood.

Each iteration of the EM algorithm consists of two steps:
1. The first step (E step) is to compute $Q(\theta|Y_N, \theta_j) = \mathrm{E}[l_{X_N,Y_N}(\theta)|Y_N, \theta_j]$.
2. The second step (M step) consists on maximizing $Q(\theta|Y_N, \theta_j)$, what is equivalent to maximize the likelihood $l_{Y_N}(\theta)$ (Equation (35)).

3.2.1 Computation of the complete likelihood $l_{X_N,Y_N}(\theta)$

The complete likelihood $l_{X_N,Y_N}(\theta)$ is computed taking into account

$$x_0 \rightsquigarrow N(\mu_0, \Sigma_0)$$
$$w_k = x_{k+1} - Ax_k, \quad w_k \rightsquigarrow N(0, Q)$$
$$v_k = y_k - Cx_k, \quad v_k \rightsquigarrow N(0, R)$$

So, the log-likelihood can be written as a sum of three uncoupled functions

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

where, ignoring constants

$$l_1(\mu_0, \Sigma_0) = \ln|\Sigma_0| + (x_0 - \mu_0)^T \Sigma_0^{-1}(x_0 - \mu_0) \tag{36}$$

$$l_2(A, Q) = N \ln|Q| + \sum_{k=1}^{N} (x_k - Ax_{k-1})^T Q^{-1}(x_k - Ax_{k-1}) \tag{37}$$

$$l_3(C, R) = N \log|R| + \sum_{k=1}^{N} (y_k - Cx_k)^T R^{-1}(y_k - Cx_k) \tag{38}$$

3.2.2 Two useful properties

The following properties are used in the Expectation step, so they are presented here. The demonstration of the these properties can be found in [3].

Property 2 (The Kalman Smoother) For the state space model specified in (17) with initial conditions x_N^N and P_N^N obtained via Property 1, for $k = N, N-1, \ldots, 1$,

$$x_{k-1}^N = x_{k-1}^{k-1} + J_{k-1}\left(x_k^N - x_k^{k-1}\right) \tag{39}$$

$$P_{k-1}^N = P_{k-1}^{k-1} + J_{k-1}\left(P_k^N - P_k^{k-1}\right) J_{k-1}^T \tag{40}$$

where

$$J_{k-1} = P_{k-1}^{k-1} A^T \left[P_k^{k-1}\right]^{-1} \tag{41}$$

Property 3 (The Lag-One Covariance Smoother) For the state space model specified in (17), with K_k, J_k $(k = 1, 2, \ldots, N)$, and P_N^N obtained from Properties 1 and 2, with initial condition

$$P_{N,N-1}^N = (I - K_N C) A P_{N-1}^{N-1} \tag{42}$$

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

$$P_{k-1.k-2}^N = P_{k-1}^{k-1} J_{k-2}^T + J_{k-1}\left(P_{k,k-1}^N - A P_{k-1}^{k-1}\right) J_{k-2}^T \tag{43}$$

3.2.3 Expectation Step

The function $Q(\theta|Y_N, \theta_j)$ is the conditional expectation of the sum of the Equations (36)-(38), and it depends on the parameters $\theta = (A, C, Q, R, \mu_0, \Sigma_0)$.

Theorem 1: Given the value of the parameters θ for iteration j, Properties 2 and 3 can be used to obtain the desired conditional expectations as smoothers:

$$x_k^N = \mathrm{E}[x_k|Y_N, \theta_j] \tag{44}$$

$$P_k^N = \mathrm{E}[(x_k - x_k^N)(x_k - x_k^N)^T|Y_N, \theta_j] \tag{45}$$

$$P^N_{k,k-1} = \mathrm{E}[(x_k - x^N_k)(x_{k-1} - x^N_{k-1})^T | Y_N, \theta_j] \quad (46)$$

and from them it is possible to compute $Q(\theta|Y_N,\theta_j)$ as follows

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

with

$$\mathrm{E}[l_1(\mu_0,\Sigma_0)|Y_N,\theta_j] =$$
$$= \ln|\Sigma_0| + \mathrm{tr}\left(\Sigma_0^{-1}\left[P^N_0 + (x^N_0 - \mu_0)(x^N_0 - \mu_0)^T\right]\right) \quad (47)$$

$$\mathrm{E}[l_2(A,Q)|Y_N,\theta_j] =$$
$$= N\log|Q| + \mathrm{tr}\left(Q^{-1}\left[S_{xx} - S_{xb}A^T - AS_{bx} + AS_{bb}A^T\right]\right) \quad (48)$$

$$\mathrm{E}[l_3(C,R)|Y_N,\theta_j] =$$
$$= N\log|R| + \mathrm{tr}\left(R^{-1}\left[S_{yy} - S_{yx}C^T - CS_{xy} + CS_{xx}C^T\right]\right) \quad (49)$$

where it has been used

$$S_{xx} = \sum_{k=1}^{N}\left(P^N_k + x^N_k(x^N_k)^T\right) \quad (50)$$

$$S_{xb} = \sum_{k=1}^{N}\left(P^N_{k,k-1} + x^N_k(x^N_{k-1})^T\right), \quad S_{bx} = S^T_{xb} \quad (51)$$

$$S_{bb} = \sum_{k=1}^{N}\left(P^N_{k-1} + x^N_{k-1}(x^N_{k-1})^T\right) \quad (52)$$

$$S_{yy} = \sum_{k=1}^{N}\left(y_k y_k^T\right) \quad (53)$$

$$S_{yx} = \sum_{k=1}^{N}\left(y_k(x^N_k)^T\right), \quad S_{xy} = S^T_{yx} \quad (54)$$

3.2.4 Maximization Step

Maximizing $Q(\theta|Y_N,\theta_j)$ with respect of the parameters θ, at iteration j, constitutes the M-step. This is the strong point of the EM algorithm because the maximum values are obtained from explicit formulas.

Theorem 2: The maximum of $\mathrm{E}[l_1(\mu_0,\Sigma_0)|Y_N,\theta_j]$ (47) is attained at

$$\hat{\mu}_0 = x^N_0 \quad (55)$$

$$\hat{\Sigma}_0 = P^N_0 \quad (56)$$

Theorem 3: The maximum of $\mathrm{E}[l_2(A,Q)|Y_N,\theta_j]$ (48) is attained at

$$\hat{A} = S_{xb}S^{-1}_{bb} \quad (57)$$

$$\hat{Q} = \frac{1}{N}\left(S_{xx} - S_{xb}\hat{A}^T - \hat{A}S_{bx} + \hat{A}S_{bb}\hat{A}^T\right) \quad (58)$$

Theorem 4: The maximum of $\mathrm{E}[l_3(C,R)|Y_N,\theta_j]$ (49) is attained at

$$\hat{C} = S_{yx}S^{-1}_{xx} \quad (59)$$

$$\hat{R} = \frac{1}{N}\left(S_{yy} - S_{yx}\hat{C}^T - \hat{C}S_{xy} + \hat{C}S_{xx}\hat{C}^T\right) \quad (60)$$

The above properties can be easily obtained equating to zero the corresponding derivatives.

3.3 EM procedure

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

1. Initialize the procedure by selecting starting values for the parameters
$$\theta_0 = (A_0, C_0, Q_0, R_0, \mu_0, \Sigma_0)$$
2. Start iteration j ($j = 1,2,\ldots$).
3. Use Property 1 to compute the innovations (Equation 23) and the incomplete-data likelihood, $l_{Y_N}(\theta_{(j-1)})$ (Equation 35).
4. Perform the E-Step.
   - Use Properties 1, 2 y 3 to obtain the smoothed values x^N_k, P^N_k, and $P^N_{k,k-1}$, for $k = 1,2,\ldots,N$, using the parameters $\theta_{(j-1)}$.
   - Use the smoothed values to calculate S_{xb}, S_{bb}, S_{xx} given in (50)-(52).
5. Perform the M-Step.
   - Update the parameters $\theta_j = (A_j, C_j, Q_j, R_j, \mu_j, \Sigma_j)$ using (55)-(60).
6. Repeat Steps 2-5 to convergence. Two options can be considered in the algorithm:
   - Perform a predefined number of iterations j_{max}.
   - Stop when the values of $l_{Y_N}(\theta_j)$ differs from $l_{Y_N}(\theta_{(j-1)})$ by some predetermined, but small amount δ[1].

$$\frac{\left|l_{Y_N}(\theta_j) - l_{Y_N}\left(\theta_{j-1}\right)\right|}{\left|l_{Y_N}\left(\theta_{j-1}\right)\right|} < \delta \quad (61)$$

4 STARTING VALUES FOR THE PARAMETERS

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.

In this section we proposed a procedure to build starting values for the parameters to initialize the EM algorithm.

4.1 Mathematical preliminaries

The state of a system is not unique. Given a LTI system (8), we can transform the state $x(t)$ into $z(t)$ as follows:

$$x(t) = Tz(t) \quad (62)$$

Replacing this condition into (8) and pre-multiplying by T^{-1}

$$z(t) = A_{c0}z(t) + T^{-1}Bu(t)$$
$$y(t) = C_{c0}z(t) + Du(t)$$

where

$$A_{c0} = T^{-1}A_cT \qquad C_{c0} = C_cT \quad (64)$$

This state representation yields the same dynamic relation between $u(t)$ and $y(t)$, that is, the same input-output behaviour that (8). One option is to choose

$$T = \begin{bmatrix}\Phi & 0\\ 0 & \Phi\end{bmatrix} \quad (65)$$

[1] In our computations we have used $\delta = 10^{-5}$

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

$$\Phi^T M\Phi = M_m$$
$$\Phi^T K\Phi = K_m$$

In the modal analysis theory M_m and K_m are called *modal mass* and *modal stiffness* matrices and both are diagonal. Using (65) A_{c0} becomes

$$A_{c0} = \begin{bmatrix} 0 & I \\ -\Phi^{-1}M^{-1}K\Phi & -\Phi^{-1}M^{-1}C_\zeta\Phi \end{bmatrix} \quad (66)$$

$$C_{c0} = C_a\Phi[-\Phi^{-1}M^{-1}K\Phi \quad -\Phi^{-1}M^{-1}C_\zeta\Phi] \quad (67)$$

Using the orthogonal properties and matrix inverse properties

$$\Phi^{-1}M^{-1}K\Phi = \Phi^{-1}M^{-1}\left(\Phi^T\right)^{-1}\Phi^T K\Phi =$$
$$= (M\Phi)^{-1}\left(\Phi^T\right)^{-1}\Phi^T K\Phi = \left(\Phi^T M\Phi\right)^{-1}\Phi^T K\Phi =$$
$$= M_m^{-1}K_m = \Omega_0^2$$

where Ω_0 is a diagonal matrix formed with the natural frequencies. In matrix form

$$\Omega_0 = \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \omega_{n_d} \end{bmatrix} \quad (68)$$

Applying the same procedure to the other component of the matrix

$$\Phi^{-1}M^{-1}C_\zeta\Phi = 2\Omega_0 Z$$

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

$$Z = \begin{bmatrix} \zeta_1 & 0 & \dots & 0 \\ 0 & \zeta_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \zeta_{n_d} \end{bmatrix} \quad (69)$$

Hence, substituting the above formulas into A_{c0} and C_{c0} results

$$A_{c0} = \begin{bmatrix} 0 & I \\ -\Omega_0^2 & -2\Omega_0 Z \end{bmatrix} \quad (70)$$

$$C_{c0} = C_a\Phi[-\Omega_0^2 \quad -2\Omega_0 Z] \quad (71)$$

The discrete expression of matrix A_{c0} and C_{c0} is obtained by means of equation (11) and (13).

$$A_0 = e^{A_{c0}\Delta t} \qquad C_0 = C_{c0} \quad (72)$$

At end, matrices Q_0 and R_0 can be not chosen arbitrarily because they depend on matrix A_0 and C_0. From (8a), taking variances

$$Var(x_{k+1}) = A^T Var(x_k)A + Var(w_k) \Rightarrow \Sigma = A^T\Sigma A + Q \quad (73)$$

$$Var(y_k) = C^T Var(x_k)C + Var(v_k) \Rightarrow L = C^T\Sigma C + R \quad (74)$$

4.2 Procedure for constructing the starting values

The complete procedure to build the initial values for the parameters $\theta_0 = (A_0, C_0, Q_0, R_0, \mu_0, \Sigma_0)$ is outlined here.

1. Given the order for the state space model, n_s, and given $n_d = n_s/2$ values for the natural frecuencies and for the damping ratios, build Ω_0 and Z matrices using (68) and (69).
2. Build A_0 using (70) and (72).
3. Build C_0 using (71) and (72). Take $\Phi = I \in \mathrm{R}^{n_d}$.
4. Compute R_0.
   - Use $y_k \in \mathrm{R}^{n_o \times N}$ to generate v_k by mean of $v_k = G_k y_k$, $k = 1,2,\dots,N$, where G_k is a random number between 0 and 1.
   - $R_0 = Var(v_k)$.
5. Compute Q_0.
   - $L = Var(y_k)$ and $\Sigma = (C_0^T)^{-1}LC_0^{-1} - R_0$
   - $Q_0 = \Sigma - A_0^T\Sigma A_0$.
6. μ_0 and Σ_0 are composed by zeros.

5 PROPOSED IDENTIFICATION METHOD USING THE EM ALGORITM

In this section is presented the proposed identification method based on the maximum likelihood estimation method for estimating the parameters of the stochastic state space model given by Equation (17). This procedure uses the iterative expectation maximization algorithm (EM) to maximize the likelihood given by (35).

The proposed identification method is defined by a seven-steps algorithm. We assume that the order of the state space equations, n_s, have been previously identified.

1. Start the step $i = 1,2,\dots,n$
2. Generate, using Montecarlo, $n_d = n_s/2$ samples of natural frequencies and damping ratios.
3. Build the initial values for the parameters by mean of the procedure indicated in section 4.2.

$$\theta_{0i} = (A_{0i}, C_{0i}, Q_{0i}, R_{0i}, \mu_{0i}, \Sigma_{0i})$$

4. Using the starting point θ_{0i} apply the EM algorithm (section 3.3) a predefined number of times and compute the log-likelihood $l_{Y_N}(\theta_{0i})$ (equation 35).
5. Repeat step (2) to (5) and select the parameters with higher likelihood θ_{max}.
6. Using these parameters as the starting point, $\theta_0 = \theta_{max}$, apply the EM algorithm (section 3.3) until the convergence is reached (equation 61).

6 NUMERICAL SIMULATIONS

The effectiveness of the proposed identification method has been evaluated using the data provided by the ASCE benchmark problem for structural health monitoring [6]. 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 [1].

A MATLAB-based finite element analysis code obtained from the IASC-ASCE SHM Task Group web site [6] has

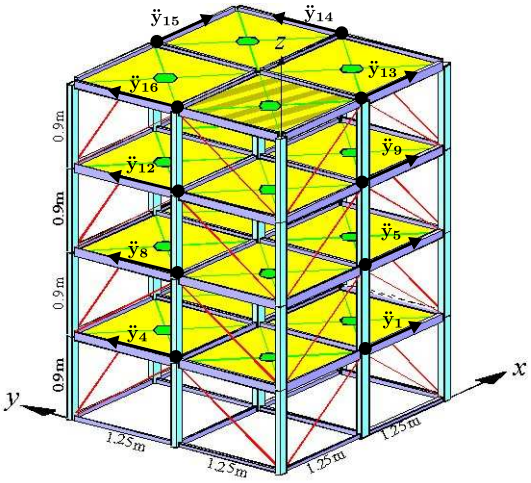

Figure 1. Diagram of the analytical model of the benchmark structure and location of the 16 measured nodes.

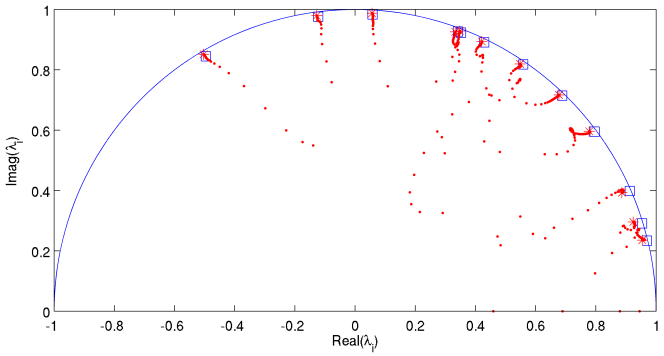

Figure 3. Theoretical eigenvalues (□) and eigenvalues of matrix A identified using EM (∗). The points represent the evolution of the eigenvalues of matrix A with respect to the iterations.

been used to calculate 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. In the first model (which has been used in this paper), the floors move as rigid bodies, with translation in the x and y directions and rotation θ about the center column. Thus, there are 3 DOF per floor. The columns and floor beams are modelled as linear elastic Euler-Bernoulli beams, and the braces as axial bars.

In this work we have used the horizontal acceleration of 16 nodes of the structure as input data for the identification subroutines. These "accelerometers" 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). Each acceleration record is the sum of the structure response and a sensor noise vector, the elements of which are Gaussian processes with root mean square (RMS) 30% of the largest RMS of the acceleration response (typically one of the roof accelerations). The structure response has been computed applying different forces in x and y directions (both white noise). 1% modal damping is assumed in each mode. The length of all the simulations has been 20 seconds, and the time step considered in each one has been $\Delta t = 0.001s$.

We have simulated 100 time history responses in order to evaluate the performance of the method. In Figure 2 it is plotted the natural frequencies identified from each simulation together with the exact values. Table 1 shows the average values of modal frequencies, damping ratios and mode shapes[2] obtained using the proposed identification method (obviously $n_s = 24$, twice the number of modes). The corresponding values of the standard deviation error (std) are shown in this table too.

Furthermore it is presented in detail the results obtained with one of these 100 time history responses. The eigenvalues of the matrix $A \in \mathbb{R}^{24\times24}$ identified using the proposed identification method has been plotted in Fig. 3 together with theoretical eigenvalues. This figure shows a good estimation of the 12 eigenvalues. The evolution with the iterations from the starting values to the final EM parameters are also shown in that figure.

7 CONCLUSIONS

The first aim of this paper was to present a time-domain stochastic system identification method based on Maximum Likelihood Estimation and Expectation Maximization algorithm. The effectiveness of the proposed method has been evaluated through a numerical simulation study in the context of the ASCE benchmark problem. The numerical results show that the proposed method identifies eigenfrequencies, damping ratios and mode shapes reasonably well in the presence of 30% measurement noises even. In this simulated analysis, where the estimates can be compared with the exact solution, the proposed method has proved to be precise.

Advantages of the proposed structural identification method can be summarized as follows:

1. The method is based on maximum likelihood, that implies minimum variance estimates;
2. EM is a computational simpler estimation procedure than other optimization algorithms;
3. The method estimate all the parameters, and this estimates are accurate.

On the contrary, the main disadvantage of the method is the time needed until convergence is reached: first, because the EM algorithm is an iterative procedure, and second, because the initial point is located by means of a Montecarlo procedure.

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.

[2]Mode shapes are evaluated by mean of the modal assurance criterion (MAC) between estimated mode shape and theoretical mode shape (calculated from K and M matrices). MAC is computed as

$$MAC(\phi_1, \phi_2) = \frac{|\phi_1^* \phi_2|^2}{(\phi_1^* \phi_1)(\phi_2^* \phi_2)}$$

where $(\bullet)^*$ = complex conjugate transpose. Therefore, MAC is a scalar between 0 and 1 ans shows the degree of which two vectors are correlated.

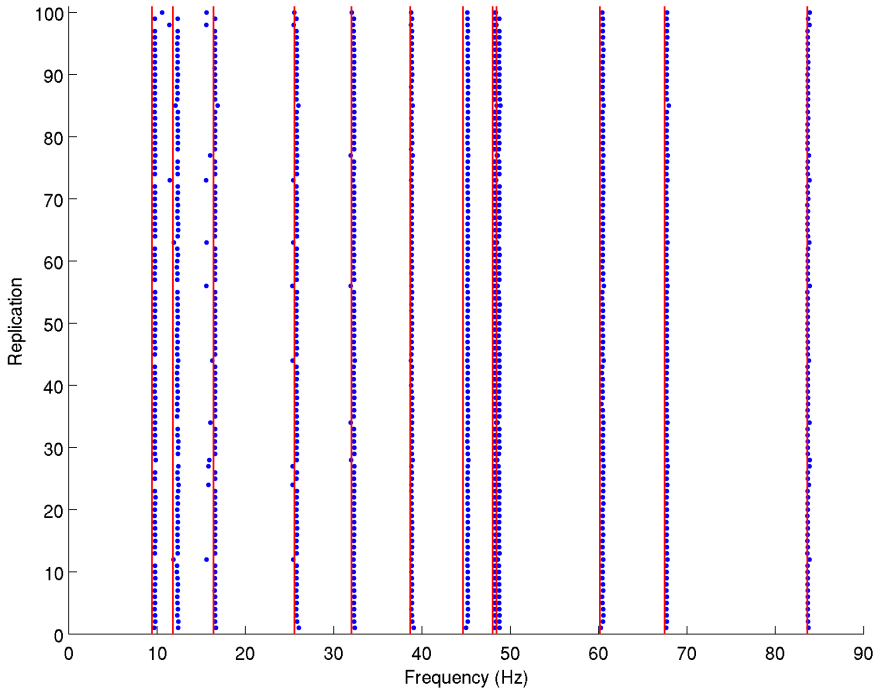

Figure 2. Identified natural frequencies of the full set of 100 simulations (•). The exact values of the frequencies are also plotted (continuous line).

	frequency (Hz)			damping ratio (%)			mode shape (mac)		
mode	exact	mean	std	exact	mean	std	mean	std	n
1	9.41	9.75	8.82	1.00	8.75	0.85	0.775	25.55	89
2	11.79	12.26	23.12	1.00	10.57	1.00	0.969	8.05	96
3	16.38	16.46	26.92	1.00	4.70	0.55	0.994	0.72	100
4	25.54	25.76	12.58	1.00	1.70	0.51	0.505	20.71	97
5	32.01	32.27	7.98	1.00	1.36	0.33	0.920	2.54	100
6	38.66	38.83	4.10	1.00	1.10	0.30	0.892	15.50	96
7	44.64	45.16	2.81	1.00	1.14	0.27	0.895	0.93	100
8	48.01	48.35	19.77	1.00	1.09	0.38	0.279	13.60	100
9	48.44	48.55	21.78	1.00	1.15	0.27	0.822	6.85	100
10	60.15	60.46	4.94	1.00	0.87	0.13	0.934	1.62	100
11	67.48	67.71	4.29	1.00	0.87	0.16	0.993	0.30	100
12	83.62	83.69	6.58	1.00	0.47	0.01	0.975	1.32	100

Table 1. Resulting modal parameters for EM method. std is the standard deviation error ($\times 10^{-2}$) and n is the number of simulations where the parameters have been identified.

REFERENCES

[1] E. A. Johnson, H. F. Lam, L. S. Katafygiotis, J. L. Beck., *Phase I IASC - ASCE structural health monitoring benchmark problem using simulated data*, Journal of Engineering Mechanics, 130(1), 315. 2004.

[2] Jer-Nan Juang, *Applied System Identification* Prentice Hall PTR. 1994.

[3] R.H. Shumway and D.S. Stoffer, *Time series analysis and its applications*, Springer, 2006.

[4] M. Verhaegen, V. Verdult, *Filtering and System Identification. A least squares approach* Cambridge University Press. 2007.

[5] F.J. Cara et all, *Maximum likelihood estimation of modal parameters in structures using the Expectation Maximization algorithm*, Proceedings of the Tenth International Conference on Computational Structures Technology, Civil-Comp Press, 2010.

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