Structural identification based on optimally weighted modal residuals

Konstantinos Christodoulou, Costas Papadimitriou*

Department of Mechanical and Industrial Engineering, University of Thessaly, Pedion Areos, 38334 Volos, Greece

Received 2 June 2005; received in revised form 23 May 2006; accepted 25 May 2006

Abstract

The structural parameter estimation problem based on measured modal data is often formulated as a weighted least-squares problem in which modal residuals measuring the fit between experimental and model predicted modal properties are build up into a single weighted residuals metric using weighting factors. Standard optimisation techniques are then used to find the optimal values of the structural parameters that minimise the weighted residuals metric. Due to model error and measurement noise, the results of the optimisation are affected by the values assumed for the weighting factors. In this work, the parameter estimation problem is first formulated as a multi-objective identification problem for which all Pareto optimal structural parameter values are obtained, corresponding to all possible values of the weights. A Bayesian statistical framework is then used to rationally select the optimal values of the weights based on the measured modal data. It is shown that the optimal weight value for a group of modal properties is asymptotically, for large number of measured data, inversely proportional to the optimal value of the residuals of the modal group. A computationally efficient algorithm is proposed for simultaneously obtaining the optimal weight values and the corresponding optimal values of the structural parameters. The proposed framework is illustrated using simulated data from a multi-dof spring–mass chain structure. In particular, compared to conventional parameter estimation techniques that are based on pre-selected values of the weights, it is demonstrated that the optimal parameter values estimated by the proposed methodology are insensitive to large model errors or bad measured modal data.

Keywords: Structural dynamics; Identification; Pareto optima; Least-squares estimation; Bayesian analysis

1. Introduction

The problem of identifying the best model from a parameterized class of structural models (e.g. a class of finite element models) or, equivalently, identifying the best values of the structural parameter set using dynamic data has received much attention over the years because of its importance in structural model updating, structural health monitoring and structural control. Comprehensive reviews of structural parameter identification methods can be found in [1–3]. The estimate of the parameter values is sensitive to uncertainties that are due to limitations of the mathematical models used to represent the behaviour of the real structure,
the presence of measurement error in the data, and insufficient excitation and response bandwidth. Structural identification and finite element model updating methodologies [e.g. 4–11] are often based on modal data since these data are readily obtained from well-established experimental structural dynamics techniques based on either forced [12,13] or ambient vibration tests [14–17]. The optimal structural models resulting from such methods can be used for response and reliability predictions [18,19], structural health monitoring [5] and control.

Structural model parameter estimation problems based on measured modal data are often formulated as weighted least-squares problems in which modal metrics, measuring the residuals between measured and model predicted modal data, are build up into a single metric formed as a weighted average of the multiple modal metrics. Standard optimisation techniques are then used to find the optimal values of the parameters that minimise the single metric representing an overall measure of fit between measured and model predicted data. The results of the optimisation are affected by the values assumed for the weighting factors. The choice of the weighting factors depends on the model adequacy and the uncertainty in the available measured data, which are not known a priori.

In the work by Haralampidis et al. [20,21] the parameter identification problem was formulated in a multi-objective context that allows the simultaneous minimisation of the multiple modal metrics, eliminating the need for using arbitrary weighting factors for weighting the relative importance of each metric in the overall measure of fit. In contrast to the conventional weighted least-squares fit between measured and model predicted modal data, the multi-objective parameter estimation methodology provides multiple Pareto optimal structural models consistent with the data in the sense that the fit each model provides in a group of measured modal properties cannot be improved without deteriorating the fit in at least one other modal group. These multiple Pareto optimal structural models are due to modelling and measurement errors. It has been demonstrated in [20,21] that the Pareto optimal structural models and the corresponding predictions of structural response and reliability may vary considerably.

Each optimal model in the Pareto set can alternatively be obtained by solving the weighted least-squares problem for a particular choice of the weight values. The whole Pareto optimal set could be estimated by varying the weight values from 0 to 1, excluding the case for which all weight values are simultaneously equal to zero. The final task of rationally selecting a unique set of weight values on which to base the estimation of the optimal structural model using the weighted least-squares method or, equivalently, the task of selecting an optimal structural model among all Pareto optimal models, is subjective and is usually left to the experience of the structural analysts.

In this work, the problem of rationally estimating the optimal values of the weights or, equivalently, selecting the most probable structural model among the Pareto optimal models utilising the available measured data is addressed. Thus, the selection of the optimal structural model is based on weight values that are estimated based on the data, avoiding an arbitrary a priori selection of these weight values. For this, a Bayesian statistical framework [22–26] for structural model parameter identification is extended to identify the values of the weights. Using Bayes theorem, the probability distribution of the weight values based on the data is formulated as a probability integral over the structural model parameters. The best values of the weights are selected as the ones that maximise the probability distribution of the weights. An asymptotic approximation, valid for large number of data, is proposed to approximate the resulting multidimensional probability integrals. An algorithm for obtaining the best values of the weights and the corresponding optimal structural models is proposed. Alternatively, using the stationarity condition at the maximum, and applying available asymptotic approximations, also valid for large number of data, to approximate the resulting multidimensional integrals, it is shown that the optimal weight value for a modal group is inversely proportional to the optimal value of the residuals, measuring the fit between the measured and model predicted data, for the corresponding modal group. A computationally efficient algorithm for simultaneously estimating these optimal weight values and the corresponding optimal structural model parameters is proposed. It is finally shown that the optimal values of the structural parameters corresponding to the optimally weighted modal residuals are obtained by minimising the sum of the logarithm of the modal residuals.

The presentation in this work is organised as follows. In Section 2, the structural identification problem using modal residuals is first formulated as a multi-objective optimisation problem and then as an equivalent single-objective optimisation with the objective formed as a weighted average of the multiple objectives using
weighting factors. For this, the modal properties are assigned into groups and each objective involved in the optimisation measures the residuals of the difference between the experimental and the model predicted modal properties involved in a modal group. In Section 3, a Bayesian statistical system identification framework for structural parameter estimation is reviewed and then extended to address the problem of estimating the probability distribution and the optimal values of the weights based on the measured data. In Section 4, asymptotic approximations are used to approximate the resulting probability integrals describing the probability distribution of the weight values. Two algorithms are proposed for estimating the optimal values of the weights and the corresponding optimal structural model based on the measured modal data. In Section 5, theoretical and computational issues are illustrated by applying the methodology to a chain-like spring mass structure using simulated measured modal data containing model error and measurement noise. The conclusions are summarised in Section 6.

2. Structural parameter estimation based on modal residuals

Let $D = \{ \hat{\omega}_r^{(k)}, \hat{\phi}_r^{(k)} \} \in \mathbb{R}^{N_0}, r = 1, \ldots, m, k = 1, \ldots, N_D \}$ be the measured modal data from a structure, consisting of modal frequencies $\hat{\omega}_r^{(k)}$ and mode shape components $\hat{\phi}_r^{(k)}$ at $N_0$ measured dofs, where $m$ is the number of observed modes and $N_D$ is the number of modal data sets available. Consider a parameterized class of linear structural models used to model the dynamic behaviour of the structure and let $\theta \in \mathbb{R}^{N_D}$, be the set of free structural model parameters that need to be assigned values in order to choose a particular model from the model class. This study addresses the problem of estimating the optimal values of the parameter set $\theta$ using the measured modal data. The problem of identifying the optimal model class from a set of alternative model classes incorporating increasing order of structural model complexity, with different number of degrees of freedom (dof), is not considered in this study. However, it can be addressed by extending (e.g. [27]) the proposed Bayesian statistical framework. Let $\{ \omega_r(\theta), \phi_r(\theta) \} \in \mathbb{R}^{N_d}, r = 1, \ldots, m$, where $N_d$ is the number of model dof, be the predictions of the modal frequencies and mode shapes obtained for a particular value of the parameter set $\theta$ by solving the eigenvalue problem corresponding to the model mass and stiffness matrices $M(\theta)$ and $K(\theta)$, respectively.

The objective in a modal-based structural identification methodology is to estimate the values of the parameter set $\theta$ so that the modal data $\{ \omega_r(\theta), \phi_r(\theta), r = 1, \ldots, m \}$ predicted by the linear class of models best matches, in some sense, the experimentally obtained modal data in $D$. For this, the measured modal properties are grouped into $n$ groups. Each group contains one or more modal properties. For the $i$th group, a norm $J_i(\theta)$ is introduced to measure the residuals of the difference between the measured values of the modal properties involved in the group and the corresponding modal values predicted from the model class for a particular value of the parameter set $\theta$. This difference is due to modelling and measurement errors, always present in structural identification problems.

2.1. Modal grouping schemes

The grouping of the modal properties $\{ \omega_r(\theta), \phi_r(\theta), r = 1, \ldots, m \}$ into $n$ groups and the selection of the measures of fit $J_i(\theta), \ldots, J_n(\theta)$ are usually based on user preference. Specifically, let

$$J_{\omega_r}(\theta) = \frac{1}{N_D} \sum_{k=1}^{N_D} \frac{[\omega_r(\theta) - \hat{\omega}_r^{(k)}]^2}{[\hat{\omega}_r^{(k)}]^2} \quad \text{and} \quad J_{\phi_r}(\theta) = \frac{1}{N_D} \sum_{k=1}^{N_D} \frac{[\phi_r(\theta) - \hat{\phi}_r^{(k)}]^2}{[\hat{\phi}_r^{(k)}]^2},$$

$(1)$

$r = 1, \ldots, m$, be the measures of fit (residuals) between the $N_D$ measured set of modal data and the model predicted modal data for the $r$th modal frequency and mode shape components, respectively, where $\|z\|^2 = z^Tz$ is the usual Euclidian norm and $p_r^{(k)} = \hat{\phi}_r^{(k)T} L_0 \hat{\phi}_r / \| L_0 \hat{\phi}_r \|^2$ is a normalization constant that guarantees that the measured mode shape $\hat{\phi}_r$ at the measured dofs is closest to the model mode shape $\beta_r^{(k)} L_0 \hat{\phi}_r(\theta)$ predicted by the particular value of $\theta$. The matrix $L_0 \in \mathbb{R}^{N_0 \times N_d}$ is an observation matrix comprised of zeros and ones that
maps the \( N_d \) model dofs to the \( N_0 \) observed dofs. Among the various grouping schemes available, the following are considered for illustration purposes.

A grouping scheme A may be defined so that each group contains one modal property, the modal frequency or the mode shape for each mode. In this case, there are \( n = 2m \) measures of fit given by \( J_i(\theta) = J_{\phi_i}(\theta) \) and \( J_{m+i}(\theta) = J_{\phi_i}(\theta), i = 1, \ldots, m \). A special case of grouping is to consider only the first \( m \) groups measuring the fit between the modal frequencies, ignoring the fit in the mode shapes.

More general grouping schemes can be defined by forming \( n \) groups with each group containing a number of modal properties. The modal properties assigned to each group are selected by the user according to their type and the purpose of the analysis. In particular, a grouping scheme B can be defined by grouping the modal properties into two groups as follows. The first group contains all modal frequencies with the measure of fit \( J_1(\theta) \) selected to represent the difference between the measured and the model predicted frequencies for all modes, while the second group contains the mode shape components for all modes with the measure of fit \( J_2(\theta) \) selected to represent the difference between the measured and the model predicted mode shape components for all modes. Specifically, the two measures of fit are given by

\[
J_1(\theta) = \sum_{r=1}^{m} J_{\phi_r}(\theta) \quad \text{and} \quad J_2(\theta) = \sum_{r=1}^{m} J_{\phi_r}(\theta).
\]

2.2. Formulation as multi-objective identification problem

The problem of identifying the model parameter values that give the best fit in all groups of modal properties can be formulated as a multi-objective optimisation problem stated as follows [21]. Find the values of the structural parameter set \( \theta \) that simultaneously minimises the objectives

\[
y = J(\theta) = (J_1(\theta), \ldots, J_n(\theta)),
\]

where \( \theta = (\theta_1, \ldots, \theta_{N_0}) \in \Theta \) is the parameter vector, \( \Theta \) is the parameter space, \( y = (y_1, \ldots, y_n) \in Y \) is the objective vector and \( Y \) is the objective space. For conflicting objectives \( J_1(\theta), \ldots, J_n(\theta) \), there is no single optimal solution, but rather a set of alternative solutions, known as Pareto optimal solutions, that are optimal in the sense that no other solutions in the parameter space are superior to them when all objectives are considered.

Such alternative solutions trade-off the fit in different modal properties. Specifically, using the grouping scheme A, all optimal models that trade-off the fit in various modal frequencies and mode shapes are obtained. Using the grouping scheme B, all optimal models that trade-off the overall fit in modal frequencies with the overall fit in the mode shapes are estimated.

Using multi-objective terminology, the Pareto optimal solutions are the non-dominating vectors in the parameter space \( \Theta \), defined mathematically as follows. A vector \( \theta \in \Theta \) is said to be non-dominated regarding the set \( \Theta \) if and only if there is no vector in \( \Theta \) which dominates \( \theta \). A vector \( \theta \) is said to dominate a vector \( \theta' \) if and only if

\[
J_i(\theta) \leq J_i(\theta') \quad \forall \ i \in \{1, \ldots, n\} \quad \text{and} \quad \exists \ j \in \{1, \ldots, n\} : J_j(\theta) < J_j(\theta').
\]

The set of objective vectors \( y = J(\theta) \) corresponding to the set of Pareto optimal solutions \( \theta \) is called Pareto optimal front. The characteristics of the Pareto solutions are that the modal residuals cannot be improved in any modal group without deteriorating the modal residuals in at least one other modal group.

It should be noted that the multiple Pareto optimal solutions are due to modelling and measurement errors. This can be easily deduced by considering the ideal case for which the model and measurement errors do not exist. In this case, there is a value \( \hat{\theta} \) of the parameter set \( \theta \) for which the model-based modal frequencies and mode shape components match exactly the corresponding measured modal properties. Thus, all objective functions \( J_1(\hat{\theta}), \ldots, J_n(\hat{\theta}) \) take the value of zero and, consequently, the Pareto front consists of a single point at the origin of the objective space.

For given modelling and measurement error, the Pareto optimal structural models may vary considerably in the parameter space. In order to understand the factors that affect the variability of the Pareto optimal
models, Figs. 1a and b depict the Pareto solutions in the objective and the parameter space, respectively, for two objectives and one parameter. The variation of the objective functions $J_1(y)$ and $J_2(y)$ with respect to $y$ are also plotted in Fig. 1b. Let $\theta^{(1)}$ and $\theta^{(2)}$ be the optimal values of the parameter that minimise the objective functions $J_1(\theta)$ and $J_2(\theta)$, respectively. The Pareto points $J_1^{(1)} = J(\theta^{(1)})$ and $J_2^{(2)} = J(\theta^{(2)})$, shown in Fig. 1a, determine the location of the boundaries of the Pareto front in the objective space. The size of the Pareto front is defined as the distance between the points $J_1^{(1)} = J(\theta^{(1)})$ and $J_2^{(2)} = J(\theta^{(2)})$, given by $\Delta J_1$ and $\Delta J_2$, respectively. All values of $\theta$ in the interval $[\theta^{(1)}, \theta^{(2)}]$, shown in Fig. 1b, are Pareto optimal solutions. The distance between the Pareto optimal values $\theta^{(1)}$ and $\theta^{(2)}$ in the parameter space is a measure of the variability of the Pareto optimal structural models.

For given measurement and model errors, that is, for given position $J_1^{(1)}$ and $J_2^{(2)}$ and given size $\Delta J_1$ and $\Delta J_2$ of the Pareto front shown in Fig. 1a, the variability of the Pareto optimal solutions depends on the overall sensitivity of the objective functions or, equivalently, the sensitivity of the modal properties, to model parameter values $\theta$ in the interval $[\theta^{(1)}, \theta^{(2)}]$. The lower the sensitivity, the higher the distance $|\theta^{(2)} - \theta^{(1)}|$, the higher the variability of the Pareto optimal models. This can be readily deduced from Fig. 1b where an increase (respectively decrease) in the distance of $\theta^{(1)}$ and $\theta^{(2)}$ for fixed $J_1^{(1)}$, $J_2^{(2)}$, $\Delta J_1$ and $\Delta J_2$ is only possible if the sensitivity of the objective functions $J_1(\theta)$ and $J_2(\theta)$ or, equivalently, the sensitivity with respect to $\theta$ of the corresponding modal properties involved in groups 1 and 2, is decreased (respectively increased). Similarly, for the same level of the overall sensitivity of the objective functions within the interval $[\theta^{(1)}, \theta^{(2)}]$, an increase in the size of the Pareto front, that is, an increase in $\Delta J_1$ and $\Delta J_2$ holding $J_1^{(1)}$ and $J_2^{(2)}$ fixed, is only possible if the length of the interval $[\theta^{(1)}, \theta^{(2)}]$ is increased, resulting in higher variability of the Pareto optimal models.

The set of Pareto optimal solutions can be obtained using evolutionary algorithms, such as the strength Pareto evolutionary algorithm [28], well-suited to solve the multi-objective optimisation problem. These algorithms process a set of promising solutions simultaneously and therefore are capable of capturing several points along the Pareto front. They are based on an arbitrary initialised population of search points in the parameter space, which by means of selection, mutation and recombination evolves towards better and better regions in the search space. In addition, techniques such as clustering are introduced in the algorithms to uniformly distribute the points along the Pareto front. In the case of one or two parameters, the Pareto set can alternatively be obtained using an exhaustive search over a desired region in the parameter space $\Theta$ and then finding the non-dominating vectors $\theta$ among the whole set of vectors in $\Theta$ that satisfy (3).

Formulating the parameter identification problem as a multi-objective minimisation problem, the need for using arbitrary weighting factors for weighting the relative importance of the residuals $J_i(\theta)$ of a modal group to an overall weighted residuals metric is eliminated. An advantage of the multi-objective identification methodology is that all admissible solutions in the parameter space are obtained. However, this is a time consuming task, it requires that multi-objective optimisation algorithms are available and that the number of objectives remain small in order to limit the number of solutions required to fully represent the multidimensional Pareto front.
2.3. Formulation as a weighted least-squares problem

The parameter estimation problem is traditionally solved by minimising the single objective

\[ J(\bar{\theta}; w) = \sum_{i=1}^{n} w_i J_i(\bar{\theta}) \]

(4)

formed from the multiple objectives \( J_i(\bar{\theta}) \) using the weighting factors \( w_i, i=1, \ldots, n \). The results of the identification depend on the weight values used. The weight values depend on the adequacy of the model class used to represent structural behaviour and the accuracy with which the measured modal data are obtained. However, the choice of weight values is arbitrary since the modelling error and the uncertainty in the measured data are usually not known a priori.

The single objective is computationally attractive since conventional minimisation algorithms can be applied to solve the problem. It can be readily shown that the optimal solution to problem (4) is one of the Pareto optimal solutions. Thus, solving a series of single objective optimisation problems of type (4) and varying the values of the weights \( w_i \) from 0 to 1, excluding the case for which the values of all weights are simultaneously equal to zero, Pareto optimal solutions are alternatively obtained. However, an important property that a multi-objective optimisation algorithm should have is to generate uniformly distributed points along the Pareto front so that the entire Pareto shape and not parts of it are adequately represented. Solving the series of weighted single-objective optimisation problems by uniformly varying the values of the weights, often results in cluster of points in parts of the Pareto front that fail to provide an adequate representation of the entire Pareto shape. Thus, alternative algorithms for solving directly the multi-objective optimisation problem and generating uniformly spread points along the entire Pareto front are preferred. Among various available algorithms, the strength Pareto evolutionary algorithm [28] is well-suited for such problems [21]. For the case of one or two parameters in \( \bar{\theta} \), the exhaustive search method described in Section 2.2 is computationally more effective in obtaining the Pareto optimal solutions.

The objective of this study is to address the problem of rationally estimating the optimal value of the weighting parameter set \( w = (w_1, \ldots, w_n) \) or, equivalently, estimating the most probable structural model among the Pareto optimal models, utilising the measured data. In the following section, a Bayesian approach for structural identification [22] is extended to simultaneously estimate the optimal values of the weighting set \( w \) and the corresponding optimal values of the structural model parameter set \( \bar{\theta} \).

3. Estimation of weight values using Bayesian analysis

3.1. Probability distribution of structural parameter values utilising modal data

The Bayesian approach to structural identification [22,23] uses probability distributions to quantify the plausibility of each possible value of the model parameters \( \bar{\theta} \). Using Bayes’ theorem, the updated (posterior) probability distribution \( p(\bar{\theta} | \bar{\sigma}, \mathcal{M}) \) of the model parameters \( \bar{\theta} \) based on the inclusion of the measured data \( D \), the modelling assumptions \( \mathcal{M} \) and the value of a parameter set \( \bar{\sigma} \), is obtained as follows:

\[ p(\bar{\theta} | \bar{\sigma}, \mathcal{M}) = \frac{cp(D | \bar{\theta}, \bar{\sigma}, \mathcal{M})p(\bar{\theta} | \bar{\sigma}, \mathcal{M})}{p(D | \bar{\sigma}, \mathcal{M})}, \]

(5)

where \( p(D | \bar{\theta}, \bar{\sigma}, \mathcal{M}) \) is the probability of observing the data from a model corresponding to a particular value of the parameter set \( \bar{\theta} \) conditioned on the modelling assumptions \( \mathcal{M} \) and the value of \( \bar{\sigma} \). \( p(\bar{\theta} | \bar{\sigma}, \mathcal{M}) \) is the initial (prior) probability distribution of a model, and \( c \) is a normalising constant selected such that the PDF \( p(D | \bar{\theta}, \bar{\sigma}, \mathcal{M}) \) integrates to one. Herein, the modelling assumptions \( \mathcal{M} \) refer to the structural modelling assumptions as well as those used to derive the probability distributions \( p(D | \bar{\theta}, \bar{\sigma}, \mathcal{M}) \) and the prior \( p(\bar{\theta} | \bar{\sigma}, \mathcal{M}) \). The parameter set \( \bar{\sigma} \) contains all parameters that need to be defined in order to completely specify the modelling assumptions \( \mathcal{M} \). Measured data are accounted for in the updated estimates through the term \( p(D | \bar{\theta}, \bar{\sigma}, \mathcal{M}) \), while any available prior information is reflected in the term \( p(\bar{\theta} | \bar{\sigma}, \mathcal{M}) \). In order to simplify the notation, the dependence of the probability distributions on \( \mathcal{M} \) is dropped in the analysis that follows.

The form of \( p(D | \bar{\theta}, \bar{\sigma}, \mathcal{M}) \equiv p(D | \bar{\theta}, \bar{\sigma}) \) is derived by using a probability model for the prediction error vector \( \xi^{(k)} = [\xi_1^{(k)}, \ldots, \xi_m^{(k)}], k = 1, \ldots, N_D \), defined as the difference between the measured modal quantities involved
in $D$ for all $m$ modes and the corresponding modal quantities predicted from a particular model within the model class. Specifically, the prediction error $e^{(k)} = [e^{(k)}_o, e^{(k)}_r]$ is given separately for the modal frequencies and the mode shapes by the prediction error equations:

$$\hat{o}_r = o_r(\bar{\theta}) + e^{(k)}_o,$$  \hfill (6)

$$\hat{\phi}_r = \beta_i L_0 \phi_r(\bar{\theta}) + e^{(k)}_r,$$  \hfill (7)

where $e^{(k)}_o$ and $e^{(k)}_r \in \mathbb{R}^{N_o}$ are, respectively, the prediction errors for the modal frequency and mode shape components of the $r$th mode.

Following the Bayesian methodology proposed by Beck and Katafygiotis [22], the predictions errors are modelled by zero-mean Gaussian vector variables. Specifically, the prediction error $e^{(k)}_o$ for the $r$th modal frequency is assumed to be a zero-mean Gaussian variable, $e^{(k)}_o \sim N(0, \sigma_{e_o}^2 \hat{o}_r^2)$, with standard deviation $\sigma_{e_o} \hat{o}_r^2$. The prediction error parameter $\sigma_{e_o}$ represents the fractional difference between the measured and the model predicted frequency of the $r$th mode. The prediction error for the $r$th truncated mode shape vector $e^{(k)}_r \in \mathbb{R}^{N_o}$ is also assumed to be zero-mean Gaussian vector, $e^{(k)}_r \sim N(0, C_r^{(k)})$, with covariance matrix $C_r^{(k)}$, where $N(\mu, \Sigma)$ denotes the multidimensional normal distribution with mean $\mu$ and covariance matrix $\Sigma$. In the analysis that follows, a diagonal covariance matrix $C_r^{(k)}$ is assumed with diagonal elements $\sigma_{\phi_r}^2 \| \hat{\phi}_r \|_N^2$, where $\| \cdot \|$ is the usual Euclidian norm. The prediction error parameter $\sigma_{\phi_r}$ represents the difference between the measured and the model predicted component of the $r$th mode shape relative to an average value $\| \hat{\phi}_r \|_N$ of the mode shape components. The parameters $\sigma_{e_o}$ and $\sigma_{\phi_r}$, $r = 1, \ldots, m$, represent the prediction error estimates of the measured modal frequencies and mode shapes involved in $D$.

In the analysis that follows, the parameter set $\sigma$, introduced in Eq. (5), is taken to contain the parameters $\sigma_r = \sigma_{e_o}$ and $\sigma_{m+r} = \sigma_{\phi_r}$, $r = 1, \ldots, m$. Given the values of the parameter set $\sigma$, assuming independence of the prediction errors in $e^{(k)}$, and using the Gaussian choice for the probability distribution of the prediction errors $e^{(k)}_o$ and $e^{(k)}_r$, the probability $p(D|\bar{\theta}; \sigma)$ of observing the data $D$ is readily obtained in the form (e.g. [5,27])

$$p(D|\bar{\theta}; \sigma) = \frac{1}{b(\sqrt{2\pi})^{NN_D} n_{p(\sigma)}} \exp \left[-\frac{NN_D}{2} J_D(\bar{\theta}; \sigma)\right],$$

where

$$J_D(\bar{\theta}; \sigma) = \sum_{i=1}^n \frac{\gamma_i}{\sigma_i} J_i(\bar{\theta})$$

with $J_i(\bar{\theta}) = J_{o_i}(\bar{\theta})$, $J_{m+i}(\bar{\theta}) = J_{\phi_i}(\bar{\theta})$, $i = 1, \ldots, m$, $n = 2m$ represents the weighted measure of fit between the measured modal data and modal data predicted by a particular model within the selected model class

$$p(\sigma) = \prod_{i=1}^n (\sigma_i)^{\gamma_i NN_D}$$

is a function of the prediction error parameters $\sigma$, $N = m(N_0 + 1)$ is the number of measured data per modal set, $\gamma_i = 1/N$ and $a_{m+i} = N_0/N$, $i = 1, \ldots, m$, satisfying $\sum_{i=1}^m \gamma_i = 1$, represent the number of data contained in each modal group in relation to the total number $N$ of data in a modal set, and $b = \prod_{r=1}^m \prod_{k=1}^{N_0} \sigma_{e_o}^2 \| \hat{\phi}_r \|_N^{N_0}$ is a constant.

Given the values of the prediction error parameters $\sigma$, the optimal value of the model parameter set $\bar{\theta}$ corresponds to the most probable model maximising the updated PDF $p(\bar{\theta}|D, \sigma, \bar{\mu})$ given in Eq. (5). In particular, using Eq. (8) and assuming a non-informative prior distribution $p(\bar{\theta}|\sigma, \bar{\mu}) = \pi_0(\bar{\theta})$, the optimal values $\hat{\bar{\theta}}$ of the model parameters $\theta$ are equivalently obtained by minimising the measure of fit $J_D(\bar{\theta}; \sigma)$ defined
in Eq. (9), i.e.
\[ \hat{\theta}(\sigma) = \arg\min_{\theta} J_D(\theta; \sigma). \] (11)

The notation \( \hat{\theta}(\sigma) \) is used to indicate that the optimal value \( \hat{\theta} \) depends on the value of the prediction error parameter set \( \sigma \).

3.2. Relation between weighting set \( w \) and prediction error parameter set \( \sigma \)

It should be noted that the overall measure of fit \( J_D(\theta; \sigma) \) between the experimental and model predicted modal data is constructed as a weighted sum of the individual measures of fit for each group of modal properties involved in the data set \( D \), with the weights to be inversely proportional to the squares of the prediction error parameters. Comparing \( J_D(\theta; \sigma) \) and \( J(\theta; w) \) given in Eqs. (9) and (4), respectively, it is clear that they are exactly the same for the modal grouping scheme A, provided that the weights \( w_i \) are chosen to be inversely proportional to the prediction error parameters \( \sigma_i^2 \), i.e.
\[ w_i = \frac{x_i}{\sigma_i^2} \quad \text{and} \quad J(\theta; w) = J_D(\theta; \sigma), \] (12)

where the vector notation \( \vec{\sigma}/\sigma^2 = (x_1/\sigma_1^2, \ldots, x_n/\sigma_n^2) \) was introduced in Eq. (12) for convenience. Thus, the problem of estimating the weight values in the weighted residuals metric defined in Eq. (4) is equivalent to the problem of estimating the prediction error parameters in the Bayesian formulation.

The formulation presented is general and applicable to other grouping schemes. Specifically, the formulation for the grouping scheme B, introduced in Section 2.1, is obtained by assuming that the prediction error parameters \( \sigma_{ov} = \sigma_1, r = 1, \ldots, m \), are the same for all the modal frequencies, and that \( \sigma_{ov} = \sigma_2, r = 1, \ldots, m \), are the same for all mode shapes. In this case, \( n = 2 \), the prediction error parameters are \( \vec{\sigma} = (\sigma_1, \sigma_2) \), and the exponents \( \alpha_i \) appearing in Eq. (10) are given by \( x_1 = m/N \) and \( x_2 = mN_0/N \).

3.3. Probability distribution of prediction error parameter values utilising modal data

The Bayesian framework is next extended to rationally estimate the optimal values of the prediction error parameters \( \sigma \) and the weights \( w \) from the available measured data \( D \), taking into account modelling error and measurement noise. Similar to the case of the parameter set \( \theta \), probability distributions are used to quantify the uncertainty in the values of the parameter set \( \sigma \). Using Bayes’ theorem, the posterior probability distribution \( p(\sigma|D) \) of the parameter set \( \sigma \) given the data \( D \) is given by
\[ p(\sigma|D) = \frac{p(D|\sigma)p_{\sigma}(\sigma)}{d}, \] (13)

where \( p(D|\sigma) \) is the probability of observing the data given the values of the parameter set \( \sigma \), \( p_{\sigma}(\sigma) \) is the prior probability distribution of the parameter set \( \sigma \) before the collection of data, while \( d \) is a normalising constant given by \( d = \int p(D|\sigma)p_{\sigma}(\sigma) \sigma \sigma \) so that the probability density function in Eq. (13) integrates to one.

Using the total probability theorem, the quantity \( p(D|\sigma) \) is given by
\[ p(D|\sigma) = \int_{\Theta} p(D|\theta, \sigma)p(\theta) \sigma \sigma, \] (14)

where it is assumed that the prior probability distribution \( p(\theta|\sigma) = \pi(\theta) \) is independent of \( \sigma \). Replacing \( p(D|\theta, \sigma) \) in Eq. (14) by Eq. (8) and then substituting Eq. (14) into Eq. (13), one readily derives that
\[ p(\sigma|D) = c_0^{-1} p_{\sigma}(\sigma) \int \pi(\theta) \exp[-0.5NN J_D(\theta; \sigma)] \theta, \] (15)

where \( c_0^{-1} = db(\sqrt{2\pi})^N N_D \). The optimal value \( \hat{\sigma} \) of the prediction error parameter set \( \sigma \) given the data \( D \) is the one that maximises the function \( p(\sigma|D) \). However, in order to compute the value of the function \( p(\sigma|D) \) for given \( \sigma \), one needs to estimate a multidimensional integral over the parameter space \( \Theta \). This makes the approach computationally very demanding and in most cases inefficient. In Section 4, asymptotic
approximations are introduced to approximate analytically the integrals and, thus, provide a more efficient algorithm for finding the optimal value $\hat{\theta}$ of the parameter set $\theta$.

### 3.4. Estimation of optimal values of structural model parameters

The marginal distribution $p(\theta | D)$ of the structural model parameters $\theta$, taking into account the uncertainty in $\sigma$, is obtained by integrating $p(\theta, \sigma | D)$ over the parameters $\sigma$ as follows:

$$p(\theta | D) = \int_\sigma p(\theta, \sigma | D) \, d\sigma = c_1 \prod_{i=1}^n [J_i(\theta)]^{-0.5} \sigma_i^{-NND-1}$$

(16)

for $\sigma_i NND > 1$, where $c_1$ is a normalising constant selected such that the $p(\theta | D)$ integrates to one. In deriving Eq. (16), Bayes theorem was first used to replace $p(\theta, \sigma | D) = c_2 p(\theta) \pi(\sigma) p(\theta, \sigma)$ in the integrand in Eq. (16), $p(D | \theta, \sigma)$ was then substituted by Eq. (8) and the integration over the parameters $\sigma$ was performed analytically assuming a non-informative prior distribution $p(\theta, \sigma) = \text{constant}$. The optimal value $\hat{\theta}_{\text{opt}}$ of the structural model parameter set $\theta$, taking into account the uncertainty in $\sigma$, is readily obtained by maximising $p(\theta | D)$ in Eq. (16).

### 4. Asymptotic approximations

#### 4.1. Algorithm I—$p(\sigma | D)$ and optimal value of $\sigma$

A well-known asymptotic approximation for the Laplace-type integral is used to give a useful estimate of the integral in Eq. (15). Specifically, for a twice differentiable function $f(\theta)$ and a continuous function $g(\theta)$, the following asymptotic approximation for the Laplace-type integral [29] holds:

$$\int g(\theta) \exp[-\beta^2 f(\theta)] \, d\theta \sim (2\pi)^{N\theta} g(\theta) \exp[-\beta^2 f(\theta)] \sqrt{|H(\theta)|}$$

(17)

as $\beta \to \infty$, where $\theta$ is the value of $\theta$ that globally minimises $f(\theta)$, $H(\theta)$ is the Hessian of the function $\beta^2 f(\theta)$, and $|H|$ denotes the determinant of the matrix $H$.

The integral in Eq. (15) is of the Laplace type (17) with $g(\theta) = \pi(\theta)$, $f(\theta) = J_D(\theta; \sigma)$, $\beta^2 = 0.5 NND$. Therefore, as $NND \to \infty$, i.e. for large number of data, the asymptotic approximation (17) can be applied for the integral in Eq. (15) to yield [30]

$$p(\sigma | D) \sim c_0 \, (2\pi)^{N\theta} \frac{\pi(\hat{\theta}(\sigma)) \exp[-0.5 NND J_D(\hat{\theta}(\sigma); \sigma)]}{\rho(\sigma) \sqrt{|H_D(\hat{\theta}(\sigma); \sigma)|}}$$

(18)

where $\hat{\theta}(\sigma)$ is the value that minimises the function $J_D(\theta; \sigma)$ with respect to $\theta$ for given value of $\sigma$, that is, $\hat{\theta}(\sigma)$ is given by Eq. (11), $H_D(\hat{\theta}(\sigma); \sigma)$ is the Hessian of the function $0.5 NND J_D(\theta; \sigma)$ evaluated at $\hat{\theta}(\sigma)$.

Assuming that the prior distributions $\pi(\theta)$ and $\pi(\sigma)$ are non-informative uniform distributions over the domain of variation of the structural parameter set $\theta$ and the prediction error parameter set $\sigma$, respectively, one may set $\pi(\sigma) = c_0 = \text{constant}$ and $\pi(\theta) = c_\theta = \text{constant}$ in Eq. (18). The optimal value $\hat{\sigma}$ of the prediction error parameter set $\sigma$ is the one that minimises $p(\sigma | D)$ in Eq. (18) or, equivalently, the one that minimises the function

$$G_D(\sigma) = -\ln p(\sigma | D) + \kappa = -NND \left[ J_D(\hat{\theta}(\sigma); \sigma) + \ln \rho(\sigma) + 2 \ln |H_D(\hat{\theta}(\sigma); \sigma)| \right]$$

(19)

where $\kappa = \ln(c_0 c_\theta) + N\theta \ln(2\pi)$ is a constant that does not affect the selection of the optimal $\hat{\sigma}$. The optimisation of the function in Eq. (19) can readily be carried out numerically using any available algorithm for optimising a non-linear function of several variables. It should be noted that in evaluating the objective function $G_D(\sigma)$ for given value of $\sigma$, an internal minimisation problem is involved for finding $\hat{\theta}(\sigma)$ from Eq. (11).
4.2. Algorithm II—optimal values of $\sigma$, $w$ and $\theta$

Alternatively, applying the stationarity conditions in Eq. (13) with respect to the elements of $\sigma$, one can derive the conditions that the optimal $\hat{\sigma}$ should satisfy. Specifically, the stationarity conditions for $p(\sigma|D)$ yield the following system of equations for $\sigma$ which are satisfied for the optimal $\hat{\sigma}$:

$$\frac{\partial p(\sigma|D)}{\partial \sigma_i} = 0$$

for $i = 1, \ldots, n$. Using Eq. (13), then substituting $p(D|\sigma)$ from Eq. (14), and finally assuming a non-informative constant prior distribution $\pi_0(\sigma) = c_\sigma$, conditions (20) can be written in the form

$$\int \frac{\partial p(D|\theta, \sigma)}{\partial \sigma_i} \pi(\theta) d \theta = 0.$$  \hfill (21)

Substituting $p(D|\theta, \sigma)$ from Eq. (8), differentiating $p(D|\theta, \sigma)$ with respect to $\sigma_i$ and rearranging the terms, one obtains the following result for the optimal $\hat{\sigma}$:

$$\int \exp(-0.5NN_D J_D(\hat{\theta}; \hat{\sigma})) d \theta = \frac{1}{\sigma_i^2} \int J_i(\theta) \exp(-0.5NN_D J_D(\hat{\theta}; \hat{\sigma})) d \theta.$$  \hfill (22)

The integrals in the left- and right-hand sides of Eq. (22) are simplified by noting that they are of the Laplace type (17). Specifically, the integral in the left-hand side of Eq. (22) is of the type (17) with $g(\theta) = 1, f(\theta) = J_D(\hat{\theta}; \sigma)$ and $\beta^2 = 0.5NN_D$, while the integral in right-hand side of Eq. (22) is of the type (17) with $g(\theta) = J_i(\theta)$, $f(\theta) = J_D(\hat{\theta}; \sigma)$ and $\beta^2 = 0.5NN_D$. Thus, both asymptotic approximations correspond to the same optimal value $\hat{\theta}(\sigma)$ given by Eq. (11) and the same Hessian $H_D(\hat{\theta}(\sigma); \sigma)$.

Applying Laplace method of asymptotic approximation [29], valid for large number of data, $NN_D \to \infty$, to the integrals in the right- and left-hand side of Eq. (22), one readily derives that the optimal values $\hat{\sigma}$ of the parameter set $\sigma$ satisfies the set of equations

$$\sigma_i^2 = J_i(\hat{\theta}(\sigma)), \quad i = 1, \ldots, 2m,$$  \hfill (23)

where $\hat{\theta}(\sigma)$ is given by Eq. (11). The result (23) indicates that the optimal value $\sigma_i^2 = J_i(\hat{\theta}(\sigma))$ of the prediction error variance $\sigma_i^2$ for the $i$th modal group involved in the data is the optimal residual value between the data involved in the $i$th group and the prediction from the optimal model $\hat{\theta}_{\text{opt}} = \hat{\theta}(\sigma)$.

The derivation of Eq. (23) is based on a non-informative (uniform) prior distribution for $\pi_0(\sigma) = c_\sigma$. The same result can alternatively be obtained using the maximum likelihood approach. The Bayesian methodology, however, generalises the result for non-uniform prior distribution $\pi_0(\sigma)$. In this case, following a similar analysis, one readily derives the set of equations $\sigma_i^2(1 + (\sigma_i/\pi_0(\sigma)\hat{\sigma}_i)/\sigma_i)) = J_i(\hat{\theta}(\sigma)), \quad i = 1, \ldots, 2m$ which, assuming slowly varying distributions $\pi_0(\sigma)$ with respect to $\sigma$ at the optimal $\hat{\sigma}$ and/or small prediction errors $\hat{\sigma}$, this set of equations reduces to the set (23).

Using Eqs. (12) and (23), the optimal weights in Eq. (4) are given by $\hat{w}_i = \sigma_i/J_i(\hat{\theta}_{\text{opt}})$, while the optimal value $\hat{\theta}_{\text{opt}} = \hat{\theta}(\sigma)$ which minimises $J_D(\hat{\theta}; \sigma)$, also minimises the optimally weighted residuals $J(\hat{\theta}; \hat{\sigma})$ in Eq. (4), that is, $\hat{\theta}_{\text{opt}} = \hat{\theta}(\hat{\sigma})$, where $\hat{\theta}(w) = \arg \min_{\theta} J(\hat{\theta}; w)$.

It should be noted that the optimal values $\hat{\sigma}$ and $\hat{\sigma}_{\text{opt}} = \hat{\sigma}(\hat{\sigma})$ are obtained by simultaneously solving Eqs. (23) and (11) with respect to $\sigma$ and $\theta$. Equivalently, the optimal values $\hat{w}$ and $\hat{\theta}_{\text{opt}} = \hat{\theta}(\hat{\sigma})$ are obtained by simultaneously solving $w_i = \sigma_i/J_i(\hat{\theta}(\hat{\sigma}))$, $i = 1, \ldots, 2m$ and $\hat{\theta}(w) = \arg \min_{\theta} J(\hat{\theta}; w)$ with respect to $w$ and $\theta$. It
can be shown that both problems are equivalent to the problem of finding $\hat{\theta}_{\text{opt}}$ that minimises the objective function

$$I(\theta) = \sum_{i=1}^{n} x_i \ln J_i(\theta)$$

(24)

with respect to the parameter set $\theta$ and then computing the optimal $\hat{\sigma}$ from $\hat{\sigma}^2 = J_i(\hat{\theta}_{\text{opt}})$ or the optimal $\hat{w}$ from $\hat{w}_i = x_i / J_i(\hat{\theta}_{\text{opt}})$. This can be readily verified by noting that the stationarity conditions $\nabla_{\theta} I(\theta) |_{\theta = \hat{\theta}_{\text{opt}}} = 0$ for the objective function $I(\theta)$, where $\nabla_{\theta} I(\theta)$ is the gradient vector with respect to $\theta$, are exactly the same as the stationarity conditions $\nabla J_D(\theta; \sigma) |_{\theta = \hat{\theta}} = 0$ for Eq. (11) with $\sigma^2$ replaced by the optimal $\hat{\sigma}^2 = J_i(\hat{\theta}_{\text{opt}})$ computed by Eq. (23).

It should be pointed out that the objective function (24) does not require information from the weights $w$ or the prediction error parameters $\sigma$. The optimisation of $I(\theta)$ with respect to $\theta$, referred to as Algorithm II, can readily be carried out numerically using any available algorithm for optimising a non-linear function of several variables. The optimisation of $I(\theta)$ is computationally much more efficient than the equivalent problem of solving simultaneously the set of Eqs. (23) and the optimisation problem (11).

Logarithmic estimators similar to Eq. (24) have been proposed in Refs. [31–33] for estimating modal parameters based on frequency response data using statistical approaches. These logarithmic estimators are formed as the sum of the logarithm of the difference between measured and model predicted quantities, instead of the sum of the logarithmic values between the measured and the model predicted parameters based on frequency response data using statistical approaches. These logarithmic estimators require the solution of an inner optimisation problem (11) for minimising the measure of error parameters $\sigma$ or the optimal $\hat{\sigma}^2$ for all $i = 1, \ldots, n$. This result also holds asymptotically for large number of data since in this case $0.5(z_i N_N D - 1) \rightarrow 0.5 z_i N_N D$ in Eq. (16) and the relation between $p(\theta | D)$ and $I(\theta)$ becomes

$$p(\theta | D) \sim c_1 \exp[-0.5 N_N D I(\theta)].$$

(25)

5. Numerical examples

Numerical examples are given next to illustrate the theoretical issues. Consider a structure represented by an $N_d$-dof chain-like spring–mass model, shown in Fig. 2, with one end of the chain fixed and the other end free. The model could be thought of representing the behaviour of an $N_d$-floor shear-type building. The nominal model of the structure is assumed to have mass and stiffness properties that are uniformly distributed along the chain, i.e. $k_i = k_0$ and $m_i = m_0$, $i = 1, \ldots, N_d$. The values of $k_0$ and $m_0$ are selected such that the lowest modal frequency of the structure is equal to 1.0 Hz.

Fig. 2. $N_d$-dof spring–mass model.
For demonstration purposes, measured modal data \( \hat{\omega}_i^{(k)} \) and \( \hat{\phi}_i^{(k)} \) are simulated by computing the modal frequencies \( \omega_{r,nom}^{(k)} \) and mode shape components \( \phi_{r,nom}^{(k)} \) from the nominal model and then adding Gaussian noise to simulate the effects of measurement noise and modelling error. This added noise is simulated for the \( r \)th modal frequency and mode shape from the normal distributions \( N(\mu_{\omega r}, \sigma_{\omega r}^2, \omega_{r,nom}^{(k)}) \) and \( N(0, \sigma_{\phi r}^2, I||\phi_{r,nom}^{(k)}|N_0) \), respectively, where \( I \) is the identity matrix. The magnitude of the model error and measurement noise is controlled by the values of the mean \( \mu_{\omega r} \) and the standard deviation parameters \( \sigma_{\omega r} \) and \( \sigma_{\phi r} \). Multiple sets of measured modal data are simulated by repeating the previous process using the same nominal model with different samples of the Gaussian noise.

5.1. Two dof model

Results are first presented for an \( N_d = 2 \) dof model. The class of models \( \mathcal{H} \) used to represent the behaviour of the structure has uniform mass distribution, i.e. \( m_1 = m_2 = m_0 \), and stiffness distribution based on the parameterization \( k_1 = k_2 = \theta k_0 \), where the single parameter \( \theta \) is used to scale the spring stiffness constants. For illustration purposes, results are presented for the modal grouping scheme A and for measured modal frequencies only \( (n = m = 2) \). This case involves two objectives \( J_1(\theta) = J_{\omega_1}(\theta) \) and \( J_2(\theta) = J_{\omega_2}(\theta) \), allowing one to graphically demonstrate the Pareto front and the features of the proposed methodology. Results are presented for model error levels corresponding to \( \mu_{\omega 1} = 0 \), \( \mu_{\omega 2} = 5\% \), \( \varepsilon_{\omega 1} = 1\% \), and for four values of \( \varepsilon_{\omega 2} = 0.5\% \), \( 1\% \), \( 3\% \) and \( 5\% \). The values of \( \mu_{\omega 1} = 0 \) and \( \mu_{\omega 2} = 2 \) are purposely chosen to simulate model error such that no model in the model class \( \mathcal{H} \) can exactly fit the measured data for both modes simultaneously. The number of modal data sets is taken to be \( N_d = 2 \).

The optimal values \( \hat{\theta}_{opt} = \hat{\theta}(\hat{\sigma}) \) of the model parameter and the corresponding optimal values \( \hat{\sigma} \) of the prediction error parameters computed using the Algorithms I and II are given in Table 1 for various values of the model error \( \varepsilon_{\omega 2} \). According to Algorithm II (see Eq. (23)), the optimal residual errors \( \hat{\sigma}_i = |J_{\omega i}(\hat{\theta}_{opt})| \) equal the optimal value of the prediction error parameters \( \hat{\sigma}_i \), that is, \( \hat{\sigma}_i = \hat{\sigma} \). This relation is asymptotically correct for the results of Algorithm I. The normalised PDF \( p(\theta|D) \) of the structural model parameter \( \theta \) computed using Eq. (25) is shown in Fig. 3. In the same figure, the Pareto optimal parameter values computed using an exhaustive search method are also shown (horizontal lines). For example, the Pareto optimal values for \( \varepsilon_{\omega 2} = 0.5\% \) computed for all \( \sigma_i \in [0, \infty) \), \( i = 1, \ldots, 2 \), vary in the range \( \hat{\theta}(\bar{\sigma}) \in [1.036, 1.10] \). The corresponding Pareto front in the objective space \( (J_1, J_2) \) is shown in Figs. 4a and b for the cases of \( \varepsilon_{\omega 2} = 0.5\% \) and \( \varepsilon_{\omega 2} = 3\% \), respectively. Each point \( (J_1, J_2) \) along the Pareto front corresponds to a particular value of the prediction error parameters \( \bar{\sigma} = (\sigma_1, \sigma_2) \) or, due to Eq. (12), to a particular value of the weights \( w = (w_1, w_2) \) in Eq. (4). The residual errors \( \hat{\sigma}_i = \{J_{\omega i}(\hat{\theta}(\bar{\sigma}))\}^{1/2} \) that each Pareto optimal model \( \hat{\theta}(\bar{\sigma}) \) provides to the measured modal frequencies differs as one moves along the Pareto front. The PDF \( p(\sigma_1, \sigma_2|D) \) of the prediction error parameter values along the Pareto front is readily computed using Eq. (18) and is drawn in Figs. 4a and b.

<table>
<thead>
<tr>
<th>Model error</th>
<th>Proposed methodology</th>
<th>Minimisation of ( J(\hat{\theta}; w) ) for ( w = \frac{1}{2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varepsilon_{\omega 2} ) (%)</td>
<td>Optimal models</td>
<td>Optimal residual errors</td>
</tr>
<tr>
<td></td>
<td>Alg I</td>
<td>Alg II</td>
</tr>
<tr>
<td>0.5</td>
<td>1.10*</td>
<td>1.10*</td>
</tr>
<tr>
<td>1</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>3</td>
<td>1.04*</td>
<td>1.04*</td>
</tr>
<tr>
<td>5</td>
<td>1.04</td>
<td>1.04</td>
</tr>
</tbody>
</table>

*Global.
along the Pareto front. The peak values (global and local optima) of \( p(\sigma_1, \sigma_2|D) \) along the Pareto front occur at the most probable values \( \hat{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2) \) (see Table 1) of the prediction error parameters computed from Algorithm I and correspond to the points PP-Alg I shown along the Pareto front. The most probable Pareto optimal points PP-Alg II, also shown in these Figures, correspond to the optimal prediction error parameters \( \hat{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2) \) (see Table 1) computed from Algorithm II using Eq. (23).

It is observed from the results in Table 1 and in Fig. 4 that the estimates from the two Algorithms I and II are almost identical. Furthermore, for the cases \( \varepsilon_{w2} = 0.5\% \) and \( \varepsilon_{w2} = 1\% \), the \( p(\theta|D) \) in Fig. 3 and the results in Table 1 reveal that there are two locally most probable optima models at \( \hat{\theta}_1 = 1.04 \) and \( \hat{\theta}_2 = 1.10 \), close to the edge points of the Pareto optimal points. For \( \varepsilon_{w2} = 0.5\% \), the optimum at \( \hat{\theta}_2 = 1.10 \) is the global optimum \((\hat{\theta}_{\text{opt}} \equiv \hat{\theta}_2)\), while the local optimum at \( \hat{\theta}_1 = 1.04 \) corresponds to slightly lower probability \( p(\hat{\theta}_1|D) = 0.913p(\hat{\theta}_2|D) \), signifying that the models in the two different regions in the parameter space fit the data almost equally well. Both optimal structural models correspond to two different Pareto points (points PP-Alg I or PP-Alg II) along the Pareto front that are almost equally probable as seen by the most probable values of the PDF \( p(\sigma_1, \sigma_2|D) \) in Fig. 4a. For \( \varepsilon_{w2} = 1\% \) the optimum at \( \hat{\theta}_1 = 1.04 \) is the global optimum \((\hat{\theta}_{\text{opt}} \equiv \hat{\theta}_1)\) and dominates the other local optimum at \( \hat{\theta}_2 = 1.10 \) since \( p(\hat{\theta}_2|D) = 0.28p(\hat{\theta}_1|D) \). For the cases \( \varepsilon_{w2} = 3\% \) and \( \varepsilon_{w2} = 5\% \), there is only one optimum at \( \hat{\theta}_{\text{opt}} = 1.04 \) that corresponds to a single most probable Pareto point (point PP-Alg I or PP-Alg II) along the Pareto front shown in Fig. 4b for the case \( \varepsilon_{w2} = 3\% \).

For all cases considered, the Pareto optimal models at \( \hat{\theta}_1 = 1.04 \) and \( \hat{\theta}_2 = 1.10 \) differ by the fit they provide to the two modal frequencies. Specifically, the Pareto optimal model \( \hat{\theta}_1 = 1.04 \) provides a very good fit to the first modal frequency with small residual error \( \hat{\varepsilon}_1[J_{o_1}(\hat{\theta}_{\text{opt}})]^{1/2} = \hat{\sigma}_1 = 0.19\% \), while the residual error for the second modal frequency is relatively high corresponding to \( \hat{\varepsilon}_2 = [J_{o_2}(\hat{\theta}_{\text{opt}})]^{1/2} = \hat{\sigma}_2 > 2.55\% \). In contrast, the Pareto optimal model \( \hat{\theta}_2 = 1.10 \) provides a very good fit to the second modal frequency with small residual error \( \hat{\varepsilon}_2 = [J_{o_2}(\hat{\theta}_{\text{opt}})]^{1/2} = \hat{\sigma}_2 < 0.36\% \), while the residual error for the first modal frequency is relatively high corresponding to \( \hat{\varepsilon}_1 = [J_{o_1}(\hat{\theta}_{\text{opt}})]^{1/2} = \hat{\sigma}_1 > 2.87\% \).

From the results in Table 1 for all four values of \( \varepsilon_{w2} \) it is worth noting that, among the structural models \( \hat{\theta}_1 = 1.04 \) and \( \hat{\theta}_2 = 1.10 \), the most probable structural model \( \hat{\theta}_{\text{opt}} \) is chosen as the one that provides a better fit (smallest residual error) to the modal frequency that contains the least scatter in the measured value of that frequency (smallest value of \( \varepsilon_{w2} \)). Specifically, for \( \varepsilon_{w2} = 1\%, 3\% \) and \( 5\% \), the optimal model \( \hat{\theta}_{\text{opt}} = \hat{\theta}_1 \) provides the best fit to the first modal frequency \( \omega_1 \) since \( \varepsilon_{w2} = 1\% \leq \varepsilon_{w2} \). For \( \varepsilon_{w2} = 0.5\% \), the optimal model \( \hat{\theta}_{\text{opt}} = \hat{\theta}_2 \) provides the best fit to the second modal frequency \( \omega_2 \) since \( \varepsilon_{w2} > \varepsilon_{w1} = 1\% \).
For comparison purposes, the optimal value $\hat{\theta}_{w=1}$ and the corresponding probability $p(\hat{\theta}_{w=1}|D)$ of the structural model parameter $\hat{\theta}_{w=1}$ computed from conventional approaches for fixed value $w = 1$ (or $\sigma^2 = 0$), are also reported in Table 1 and shown with a circle in the parameter and objective spaces in Figs. 3 and 4 (see point PP-$\sigma = 1$), respectively. Note that the optimal value $\hat{\theta}_{w=1} \approx 1.07$ and it differs from the most probable values $\hat{\theta}_{\text{opt}} = \hat{\theta}(\hat{\sigma}) = 1.04$ or 1.10 computed at the optimal $\hat{\sigma}$. From the results in Fig. 3, the probability of the optimal value $\hat{\theta}_{w=1} \approx 1.07$ as predicted by the proposed methodology is small compared to the probability of $\hat{\theta}_{\text{opt}}$. Similarly, as reported in Table 1, the Pareto optimal points for $w = 1$ correspond to residual errors ranging from $\hat{\varepsilon}_1 = [J_{\sigma_1}(\hat{\theta}_{w=1})]^{1/2} = 1.49\%$ to $0.94\%$ for the first modal frequency and $\hat{\varepsilon}_2 = [J_{\sigma_2}(\hat{\theta}_{w=1})]^{1/2} = 1.53\%$ to $2.05\%$ for the second modal frequency. Moreover, as shown in Figs. 4a and b, the probability of the Pareto optimal points (PP-$\sigma = 1$) corresponding to $w = 1$ is small compared to the probability of Pareto points (PP-Alg I or PP-Alg II) corresponding to the most probable Pareto optimal points $\hat{\sigma}$.

The optimal model $\hat{\theta}_{w=1}$ assuming equal weight values $w = 1$ or, equivalently, fully correlated prediction error parameters with $\sigma^2 = \bar{z}$, tends to trade off the fit to both modal frequencies, giving no clear preference to
the fit in one of the two modal frequencies. Specifically, compared to the most probable structural model 
\( \hat{\theta}(\tilde{\theta}) = 1.04 \) for \( \varepsilon_{\omega_2} = 1\% \), the Pareto optimal structural model \( \hat{\theta}_{w=1} = 1.07 \) improves the fit in the second mode from 2.84% to 1.50% in the expense of deteriorating the fit in the first mode from 0.19% to 1.42%. As it will be seen in the next example, this trade-off in the fit of various modal properties can make the estimation of the optimal structural model \( \hat{\theta}_{w=1} \) very sensitive to large model errors and measurement noise.

5.2. Ten dof model

Results are next presented for an \( N_d = 10 \) dof model. Three parametric model classes \( \mathcal{M}_1, \mathcal{M}_2, \) and \( \mathcal{M}_3 \) are considered, involving one, two and five parameters, respectively. All three model classes are selected to have uniform mass distribution, \( m_i = m_0, i = 1, \ldots, N_d \). The first model class \( \mathcal{M}_1 \) involves one parameter \( \theta \) based on the uniform parameterization \( k_i = \theta k_0, i = 1, \ldots, N_d \), where the single parameter \( \theta \) is used to scale all \( N_d = 10 \) spring stiffness constants assumed fully correlated. The second model class \( \mathcal{M}_2 \) involves two parameters \( \theta_1 \) and \( \theta_2 \) based on the parameterization \( k_i = \theta_1 k_0, i = 1, \ldots, 5 \) and \( k_i = \theta_2 k_0, i = 6, \ldots, 10 \). The third model class \( \mathcal{M}_3 \) involves five parameters \( \theta_i, i = 1, \ldots, 5 \), based on the parameterization \( k_{2(i-1)+j} = \theta_i k_0, j = 1, 2 \).

For illustration purposes, it is assumed that measurements are available for the lowest four modal frequencies and mode shapes \( (m = 4) \). Results will be presented for the grouping scheme A and for a sensor configuration involving five sensors placed at dofs \( 2, 4, 6, 8 \) and \( 10 \). Thus, the number of objectives is \( n = 2m = 8 \). In this case, the Pareto front belongs to an 8-dimensional space and so it cannot be depicted graphically. Measured data are generated from the nominal model and for model error levels corresponding to values of \( \mu_{\omega_r} = 0 \) and \( \varepsilon_{\omega_r} = 1\% \) for \( r = 1, 2 \) and \( 3 \), while for \( r = 4, \mu_{\omega_4} = 20\% \) and \( \varepsilon_{\omega_4} = 1\% \). Also, for the mode shape components, \( \varepsilon_{\phi_r} = 5\% \) for \( r = 1, \ldots, 4 \). These values are purposely chosen so that no model in the model classes \( \mathcal{M}_1, \mathcal{M}_2 \) and \( \mathcal{M}_3 \) can exactly fit the measured data for all modes simultaneously.

Also, the high discrepancy between the simulated value in the fourth modal frequency and the modal frequency obtained from the nominal model corresponding to \( \hat{\theta}_{\text{nom}} = \tilde{\theta} \) is purposely introduced to simulate significant model error that could be due to false structural modelling assumptions, or to simulate bad measured data for this frequency that could be due to false measurements. This allows us to study the behaviour of the proposed parameter estimation framework in the presence of significant modelling error or bad measured data in one of the modal properties. It should be noted that the nominal model \( (\hat{\theta}_{\text{nom}} = \tilde{\theta}) \) is one of the models in the three model classes that can simultaneously fit the lowest three measured frequencies to within an error of \( 1\% \), maintaining a high error of the order of \( 20\% \) in the fourth modal frequency. Any departure of the values of the parameter set \( \hat{\theta} \) from the nominal values in order to improve the fit in the fourth modal frequency will result in the deterioration of the fit in the lowest three modal frequencies.

---

**Fig. 5.** The PDF \( p(\theta|D) \) of the model parameters.
For the case of model class $M_1$ and for $ND = 2$ data sets, the PDF $p(\theta | D)$ of the structural model parameter $\theta$ is shown in Fig. 5. For the case of model class $M_2$ and for $ND = 2$, the contour plots of the $p(\theta | D)$ in the two-dimensional parameter space is shown in Fig. 6. It is observed in these figures that a highly non-linear, non-convex, objective function is obtained which involves multiple local/global optima. The Pareto optimal values of the structural parameters obtained using an exhaustive search in the parameter space are also shown in Fig. 5 (PP points along the horizontal line) and Fig. 6 (PP points). A large variability of Pareto optimal parameter values is observed which trade-off the fit in the $n = 2m = 8$ objectives involved. As expected, the optimal points (global and local ones) $\hat{\theta}_{opt} = \hat{\theta}(\hat{s})$ maximising $p(\theta | D)$ belong to the set of Pareto optimal points as shown in Figs. 5 and 6.

It is worth pointing out that conventional gradient-based local optimisation algorithms lack reliability in dealing with the multiple local/global optima observed, since convergence to the global optimum is not guaranteed. Evolution strategies [34] are more appropriate and effective to use in such cases. Evolution strategies are random search algorithms that explore better the parameter space for detecting the neighbourhood of the global optimum, avoiding premature convergence to a local optimum. A disadvantage of evolution strategies is their slow convergence since they do not exploit the gradient information. In this work, a hybrid optimisation algorithm is used that exploits the advantages of evolution strategies and gradient-based methods. Specifically, an evolution strategy is used to explore the parameter space and detect the neighbourhood of the global optimum. Then the method switches to a gradient-based algorithm starting with the best estimate obtained from the evolution strategy and using gradient information to accelerate convergence to the global optimum.

The optimal values $\hat{\theta}(\hat{s})$ (global and local optima) of the structural model parameters and the corresponding optimal values $\hat{s}_i = [U(\hat{\theta}_{opt})]^{1/2} = \hat{s}_i$, $i = 1, \ldots, 8$, of the residual errors obtained using the Algorithm II are given in Table 2 for model classes $M_1$ and $M_2$ and for two cases of datasets $ND = 2$ and 1000. The case of $ND = 1000$ is presented in order to show the asymptotic behaviour of the proposed methodology at the
limiting case of large number of data. For \( N_D = 2 \) and for the model class \( \mathcal{M}_1 \), the \( p(\theta|D) \) has one global optimum at \( \hat{\theta}_{opt} = 0.986 \) and two local optima at \( \hat{\theta}_{2} = 0.998 \) and \( \hat{\theta}_{3} = 1.45 \). The local optimum at \( \hat{\theta}_{2} = 0.998 \) is almost equally probable to the global optimum at \( \hat{\theta}_{opt} = 0.986 \). Both these optima give small residual errors \( \hat{\epsilon}_i = [J_i(\hat{\theta}_{opt})]^{1/2} = \hat{\sigma}_i , i = 1, 2, 3 \), of the order of 1\% for the lowest three modal frequencies and a very high residual error \( \hat{\epsilon}_4 = [J_i(\hat{\theta}_{opt})]^{1/2} = \hat{\sigma}_4 \geq 17.1\% \) for the fourth modal frequency. That is, these optimal structural models \( \hat{\theta}_{opt} \) and \( \hat{\theta}_{2} \) provide a very good fit to the lowest three modal frequencies. In contrast, the local optimal model corresponding to \( \hat{\theta}_3 = 1.45 \), fits the fourth frequency very well within an error of 1\%, while it provides a bad fit to the lowest three modal frequencies corresponding to residuals \( \hat{\epsilon}_i \), \( i = 1, 2, 3 \), of the order of 20\%. As shown in Fig. 5, the Bayesian methodology gives a very small probability to the optimal model \( \hat{\theta}_3 = 1.45 \), compared to the probability of the most probable model at \( \hat{\theta}_{opt} = 0.986 \) or the next most probable model at \( \hat{\theta}_{2} = 0.998 \).

Increasing the number of data sets to values higher than \( N_D = 2 \), one of the two global/local optimal models in the region close to \( \theta = 1 \) disappear. For large number of datasets (\( N_D = 1000 \)), the global optimum occurs at \( \hat{\theta}_{opt} = 1.00 \) with its value coinciding with the nominal value \( \theta_{nom} = 1 \) of the structural parameter, while the next local optimum at \( \hat{\theta}_{2} = 1.44 \) has several orders of magnitude less probability than the global one (see results in Table 2). The optimal prediction errors \( \hat{\sigma} \) converge to the values of approximately 1\% for the lowest three modal frequencies, 20\% for the fourth modal frequency and 4.5\% for the mode shape components, which are close to the noise levels used to simulate the model and measurement error.

It should be noted that the results for the model class \( \mathcal{M}_2 \) shown in the last four columns in Table 2 lead to conclusions that are qualitatively similar to the ones drawn for the model class \( \mathcal{M}_1 \).

Results obtained using the proposed Algorithms I and II based on the asymptotic approximations are next compared. The most probable (global optimal) values \( \hat{\theta}_{opt} = \hat{\theta}(\hat{\sigma}) \) of the structural model parameters, the corresponding optimal values of the residuals \( \hat{\epsilon}_i = [J_i(\hat{\theta}_{opt})]^{1/2} = \hat{\sigma}_i , i = 1, \ldots, 8 \), and the \( p(\hat{\theta}_{opt}|D) \) are given in Table 3 for the model class \( \mathcal{M}_3 \) involving five structural parameters, for \( N_D = 2 \) and 1000 number of data sets. The differences in the optimal estimates of the prediction error parameters and the structural model parameters obtained by the two Algorithms I and II for \( N_D = 2 \) are small. As the number of data increases, due to increasing the number of datasets from \( N_D = 2 \) to 1000, the estimates of optimal models \( \hat{\theta}(\hat{\sigma}) \) from the two algorithms converge to the same value. Such convergence was also observed by increasing the number of sensors or increasing the number of measured modes. This tendency is consistent with the range of validity of
the asymptotic approximations that tend to be asymptotically correct for higher number of data. Also, as the number of data increases, the corresponding values of the optimal prediction errors are of the order of the simulated model and measurement errors used to generate the measured data. This is clearly demonstrated for the case of $N_D = 1000$ for which the optimal prediction errors tend to approximately 1% for the lowest three modal frequencies, 16.6% for the fourth modal frequency and 4.5% for the mode shapes, which are of the order of the noise levels used to simulate the Gaussian model and measurement error in the measured data.

For comparison purposes, results from conventional algorithms based on pre-specified weight values are also computed. The optimal value $\hat{\theta}_{w=1}$ of the structural model parameters, the corresponding prediction errors $\tilde{\varepsilon}_i = [J_i(\hat{\theta}_{w=1})]^{1/2}$ and the probability $p(\hat{\theta}_{w=1}|D)$ computed by assuming equal weights values $w = 1$ or fully correlated prediction error parameters ($\sigma_y = \bar{z}$), are also reported in Fig. 5 and Table 3 (see columns 5 and 8). The optimal model $\hat{\theta}_{w=1}$, is also one of the Pareto optimal points but it differs from the most probable value $\hat{\theta}_{opt} = \hat{\theta}(\hat{\sigma})$ or the other local optima. As predicted by the proposed Bayesian methodology, the probability of the optimal value $\hat{\theta}_{w=1}$ is very small compared to the probability of $\hat{\theta}_{opt} = \hat{\theta}(\hat{\sigma})$ based on the optimal prediction errors $\hat{\sigma}$.

From the results in Table 3 for $N_D = 2$, it can be noted that compared to the most probable optimal structural model $\hat{\theta}_{opt}$, the optimal model $\hat{\theta}_{w=1}$ improves the fit in the fourth modal frequency from $\hat{\varepsilon}_4 = 17\%$ to $\hat{\varepsilon}_4 = 12.9\%$, in the expense of deteriorating the fit for the lowest three modal frequencies from $\hat{\varepsilon}_i$ values ($i = 1, 2, 3$) of the order of 1% to $\hat{\varepsilon}_i$ values of the order of 3.12–4.5%, much higher than the errors used to simulate the measured data for the lowest three modal frequencies. Also, the proposed methodology consistently gives the optimal model parameters to be very close to the nominal ones ($\hat{\theta}_{nom} = 1$) with largest error $\|\hat{\theta}_{opt} - \hat{\theta}_{nom}\|_\infty = 0.029$, while the parameter estimation based on equal weights ($w = 1$) gives optimal models that differ significantly from the nominal model used to generate the measured data, with largest error $\|\hat{\theta}_{w=1} - \hat{\theta}_{nom}\|_\infty = 0.158$. For the limiting case of large number of data ($N_D = 1000$), the error $\|\hat{\theta}_{opt} - \hat{\theta}_{nom}\|_\infty = 0.004$ tends to zero, while the error $\|\hat{\theta}_{w=1} - \hat{\theta}_{nom}\|_\infty = 0.148$ remains at the same high level as before.

Comparisons clearly demonstrate that the optimal model assuming equal weights ($w = 1$), or fully correlated prediction error parameters, is selected to trade off the fit to all four measured modal frequencies, giving no preference to the fit in a specific group of modal properties. As a result, large modelling and
measurement errors in the fourth modal frequency drive the predictions from the method and significantly deteriorate the fit in the lowest three modal frequencies. In contrast, the proposed Bayesian methodology takes into account the levels of measurement noise and model error, providing optimal models that are relatively insensitive to large model errors or bad data contained in the measurements, maintaining a very good fit in the lowest three modal frequencies and ignoring the fit in the fourth modal frequency.

6. Conclusions

The problem of selecting the optimal values of the weights in weighted least squares approaches for structural parameter estimation based on modal residuals in structural dynamics is addressed. The problem is formulated as a multi-objective identification problem for which all Pareto optimal structural parameter values corresponding to all possible values of the weights are obtained. Based on a Bayesian statistical framework and the measured modal data, two algorithms are proposed for estimating simultaneously the optimal values of the weights and the optimal values of the structural model parameters. Numerical results indicate that the estimates from the two algorithms are close, while they converge to the same value as the number of modal data increases. In particular, it is shown that the optimal weight value for a modal group is inversely proportional to the optimal residuals between the measured and the model predicted properties involved in the modal group. The estimate is asymptotically correct for relatively large number of modal data. Finally, it is shown that the optimal values of the structural parameters corresponding to the optimally weighted modal residuals are obtained by minimising the sum of the logarithm of the modal residuals given by Eq. (24). Hybrid algorithms based on evolution strategies and gradient methods are well-suited optimisation tools for solving the resulting non-convex optimisation problem and identifying the global optimum from multiple local ones. Compared to conventional parameter estimation techniques based on pre-specified weight values, the proposed approach provides optimal models that are insensitive to large model errors or bad data contained in the measurements. In future work, the applicability of the proposed optimally weighted modal residuals methodology for model updating remains to be explored with real measurement data.

The methodology can readily be applied to alternative modal residual metrics proposed in the literature (e.g. [5–7,10]) to measure the fit between experimental and model predicted modal data. Also, it can be extended to identify the structural parameters of linear and non-linear models using measured acceleration time histories instead of modal properties.

Acknowledgements

This research was partially funded by the Greek Ministry of Education and the European Community Fund within the Hrakleitos programme framework under Grant MIS: 88730. Additional funding was provided by the Greek Secretariat of Research and Technology and the European Community Fund, through the EPAN programme framework under Grant DP15. Both supports are gratefully acknowledged.

References