Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Sound and Vibration 278 (2004) 923–947
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Optimal sensor placement methodology for parametric identification of structural systems

C. Papadimitriou*

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

Received 19 May 2003; accepted 21 October 2003

Abstract

Theoretical and computational issues arising in the selection of the optimal sensor configuration for parameter estimation in structural dynamics are addressed. The information entropy, measuring the uncertainty in the system parameters, is used as the performance measure of a sensor configuration. A useful asymptotic approximation for the information entropy, valid for a large number of measured data, is derived. The asymptotic estimate is then used to rigorously justify that selections of the optimal sensor configuration can be based solely on a nominal structural model, ignoring the time history details of the measured data which are not available in the experimental design stage. It is further shown that the lower and upper bounds of the information entropy are decreasing functions of the number of sensors. Based on this result, two algorithms are proposed for constructing effective sensor configurations that are superior, in terms of computational efficiency and accuracy, to the sensor configurations provided by genetic algorithms. The theoretical developments and the effectiveness of the proposed algorithms are illustrated by designing the optimal configuration for a 10-degree-of-freedom (d.o.f.) chain-like spring–mass model and a 240-d.o.f. three-dimensional truss structure.

1. Introduction

The problem of parametric identification of structural models using measured dynamic data has received much attention over the years because of its importance in structural model updating, structural health monitoring and structural control. The estimate of the parameter values involves 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

*Tel.: +30-24210-74006; fax: +30-24210-74012.
E-mail address: costasp@uth.gr (C. Papadimitriou).

doi:10.1016/j.jsv.2003.10.063

excitation and response bandwidth. In particular, the quality of information that can be extracted from the data for estimating the model parameters depends on the number and location of sensors in the structure. The objective in an experimental design is to make cost-effective selection of the optimal number and location of sensors such that the resulting measured data are most informative about the condition of the structure.

Previous work addressing the issue of optimally locating a given number of sensors in a structure for modal and/or finite element model parameter estimation has been carried out by several investigators [1–9]. In particular, information theory based approaches (e.g. Refs. [2–4,10–13]) have been developed to provide rational solutions to several issues encountered in the problem of selecting the optimal sensor configuration. In Refs. [2–4] the optimal sensor configuration is taken as the one that maximizes some norm (determinant or trace) of the Fisher information matrix (FIM). Refs. [12,13] treat the case of large model uncertainties expected in model updating. The optimal sensor configuration is chosen as the one that minimizes the expected Bayesian loss function involving the trace of the inverse of the FIM for each model.

Papadimitriou et al. [14] introduced the information entropy norm [15] as the measure that best corresponds to the objective of structural testing which is to minimize the uncertainty in the model parameter estimates. Specifically, the optimal sensor configuration is selected as the one that minimizes the information entropy measure since it gives a direct measure of this uncertainty. In particular, this information entropy-based measure resolved the issue related to the arbitrariness in selecting an appropriate norm for the Fisher information matrix in previous approaches based on the Fisher information matrix. It was shown that the information entropy depends on the determinant of the Fisher information entropy and not the trace.

An important advantage of the information entropy measure is that it allows us to make comparisons between sensor configurations involving a different number of sensors in each configuration [14,16]. The information entropy is particularly useful for trading-off cost of instrumentation with information gained from additional sensors about the state of the structure, thus making cost-effective decisions regarding optimal instrumentation. Furthermore, it has been used to design the optimal characteristics of the excitation (e.g., amplitude and frequency content) useful in the identification of linear and non-linear models [17].

Computational issues arising in the search of the optimal sensor configuration have also been addressed in the literature. The problem of finding the optimal sensor configuration is formulated as a discrete minimization problem. An exhaustive search of the optimal sensor configuration is computationally prohibitive even for structures with relatively small number of degrees of freedom (d.o.f.). Kammer and Yao (see Ref. [2]) proposed an effective iterative algorithm for sensor placement in the case of estimating modal parameters. Starting with FIM computed for all model d.o.f.'s, the sensors resulting in the lowest reduction in the determinant of the FIM are sequentially removed from the structure until the desired number of sensors is reached. Genetic algorithms (GA) have also been proposed as an effective alternative [18–20] to the previous heuristic algorithm which is not guaranteed to give the optimal solution. GAs are well suited for approximately solving the resulting discrete optimization problem by exploring an infinitesimal fraction of the total number of possible sensor configurations. Finally, Udwadia [3] demonstrated that using the trace of the FIM as the performance index is computationally very attractive since the solution of the underlined discrete optimization problem is straightforward.

The objective of the present study is twofold. Firstly, a rigorous formulation for the design of sensor configuration for parameter estimation in dynamic systems is presented based on the information entropy measure. Specifically, the estimation of the information entropy is re-visited and a useful asymptotic approximation, valid for large number of data, is derived. This asymptotic estimate is used to justify that the selection of the optimal sensor configuration can be based solely on a nominal model such as a finite element model, ignoring the time-history details of the measured data that are not available in the initial stage of experimental design. Moreover, analysis shows that the lower and upper bounds of the information entropy for a fixed number of sensors, corresponding, respectively, to the optimal and worst sensor configuration, is a decreasing function of the number of sensors. These bounds are important in evaluating the effectiveness of a sensor configuration and the need for relocating or increasing the number of sensors in the structure.

Secondly, computational issues related to the estimation of the optimal sensor locations are addressed. Exploiting the theoretical results derived in this study, two computationally efficient algorithms are proposed for constructing sensor configurations that correspond to information entropy values very close to lower or upper bounds of the information entropy. The computational efficiency and effectiveness of the proposed algorithms are verified by designing the sensor configuration for a 10 d.o.f. chain-like spring–mass model and a 240 d.o.f. three-dimensional truss structure both excited by an impulse hammer. In particular, numerical results indicate that the proposed algorithms provide sensor configurations that can be extremely good approximations of the optimal sensor configuration. Moreover, it is demonstrated that the predictions from these heuristic algorithms outperform those provided by GAs.

The presentation in this work is organized as follows. The main results of a Bayesian statistical framework for structural identification [21], needed in the formulation of the sensor placement problem, are reviewed in Section 2. The formulation for the information entropy and its asymptotic approximation is presented in Section 3. The problem of finding the optimal sensor configuration is formulated in Section 4 as a discrete minimization problem. In Section 5 useful properties of the information entropy related to the dependence of its lower and upper bounds on the number of sensors are presented. Section 6 deals with computational issues and algorithms for providing good estimates for the optimal sensor configuration with minimal computational effort. Applications illustrating theoretical developments and the effectiveness of the proposed algorithms are given in Section 7. The conclusions of this work are summarized in Section 8.

2. Statistical framework for structural identification

Consider a parameterized class of structural models (e.g., a class of finite element models or modal models) chosen to describe the input–output behaviour of a structure. Let $\boldsymbol{\theta} \in R^{N_\theta}$ be the vector of parameters in the model class (e.g., stiffness parameters, modal parameters, etc.). The Bayesian statistical system identification methodology developed by Beck and Katafygiotis [21] is adopted to estimate the values of the parameter set $\boldsymbol{\theta}$ and their associated uncertainties using the information provided from dynamic test data. For completeness, the main results of the identification methodology needed in the analysis of the information entropy are briefly reviewed.

Let $D = \{\mathbf{y}(m), \mathbf{z}(m), m = 1, \ldots, N\}$ be the measured sampled time histories data, where $\mathbf{y}(m) \in R^{N_0}$ and $\mathbf{z}(m) \in R^{N_I}$ refer to output and input data, respectively, N_0 is the number of observed d.o.f. of the structural model, N_I is the number of input d.o.f,'s, m denotes the time index at time $m\Delta t$, Δt is the sampling interval, and N is the number of sampled data. Given the input time histories $\mathbf{z}(m)$, $m = 1, \ldots, N$, let $\mathbf{x}(m; \boldsymbol{\theta}) \in R^{N_d}$, $m = 1, \ldots, N$ be the sampled response time histories computed at all N_d model d.o.f.'s from a particular structural model that corresponds to a specific value $\boldsymbol{\theta}$ of the model parameters. The measured response and the model response predictions satisfy the equation

$$\mathbf{y}(m) = \mathbf{L}_0\mathbf{x}(m; \boldsymbol{\theta}) + \mathbf{L}_0\mathbf{n}(m; \boldsymbol{\theta}), \tag{1}$$

where $\mathbf{n}(m; \boldsymbol{\theta})$ is the model prediction error that is due to modelling error and measurement noise. The matrix $\mathbf{L}_0 \in R^{N_0 \times N_d}$ is the observation matrix comprised of zeros and ones and maps the model d.o.f.'s to the measured d.o.f.'s. Introducing, for convenience, the sensor configuration vector $\boldsymbol{\delta} \in R^{N_d}$ with elements $\delta_j = 1$ if the jth model d.o.f. is observed and $\delta_j = 0$ if the jth model d.o.f. is not observed, it can be readily shown that $\mathbf{L}_0^{\mathrm{T}}\mathbf{L}_0 = \mathrm{diag}(\boldsymbol{\delta})$.

According to the Bayesian system identification methodology [21], the uncertainties in the values of the parameters $\boldsymbol{\theta}$ are quantified by probability density functions (PDF) that are obtained using the dynamic test data D and the probability model for the prediction error $\mathbf{n}(m; \boldsymbol{\theta})$. Specifically, modelling the components of the prediction error $\mathbf{n}(m; \boldsymbol{\theta})$ by independent Gaussian PDFs with zero mean and variance σ^2, and applying the Bayes' theorem, the updating PDF $p(\boldsymbol{\theta}, \sigma \mid D)$ of the set of structural model and prediction error parameters $(\boldsymbol{\theta}, \sigma)$ given the measured data D takes the form

$$p(\boldsymbol{\theta}, \sigma \mid D) = \tilde{c}\frac{1}{(\sqrt{2\pi}\sigma)^{NN_0}} \exp\left[\frac{NN_0}{2\sigma^2} J(\boldsymbol{\theta}; D)\right] \pi(\boldsymbol{\theta}, \sigma), \tag{2}$$

where

$$J(\boldsymbol{\theta}; D) = \frac{1}{NN_0} \sum_{m=1}^{N} \|\mathbf{y}(m) - \mathbf{L}_0\mathbf{x}(m; \boldsymbol{\theta})\|^2 \tag{3}$$

represents the measure of fit between the measured and the model response time histories, $\|\cdot\|$ is the usual Euclidian norm, $\pi(\boldsymbol{\theta}, \sigma) = \pi_\theta(\boldsymbol{\theta})\pi_\sigma(\sigma)$ is the prior distribution for the parameter set $(\boldsymbol{\theta}, \sigma)$, and $\tilde{c} = \tilde{c}(D)$ is a normalizing constant chosen such that the PDF in Eq. (2) integrates to one.

Using the total probability theorem, the marginal probability distribution $p(\boldsymbol{\theta} \mid D)$ for the structural model parameters $\boldsymbol{\theta}$ is given by $p(\boldsymbol{\theta} \mid D) = \int p(\boldsymbol{\theta}, \sigma \mid D)\mathrm{d}\sigma$. For a non-informative (uniform) prior distribution $\pi_\sigma(\sigma)$ the integration with respect to σ can be carried out analytically to yield

$$p(\boldsymbol{\theta} \mid D) = c[J(\boldsymbol{\theta}; D)]^{-(NN_0-1)/2}\pi_\theta(\boldsymbol{\theta}), \tag{4}$$

where $c \equiv c(D)$ is a normalizing constant ensuring that the PDF in Eq. (4) integrates to one, that is,

$$\frac{1}{c(D)} = \int [J(\boldsymbol{\theta}, D)]^{-(NN_0-1)/2}\pi_\theta(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}. \tag{5}$$

For a general prior (initial) distribution $\pi_\sigma(\sigma)$, an asymptotic approximation, valid for large number of data ($N \to \infty$), is available [22] in the form (4) with c replaced by $c_0 = c\pi_\sigma(\sqrt{J(\boldsymbol{\theta}; D)})$.

3. Information entropy and its asymptotic approximation

The PDF $p(\boldsymbol{\theta} \mid D)$ specifies the plausibility of each possible value of the structural model parameters. It provides a spread of the uncertainty in the parameter values based on the information contained in the measured data. A unique scalar measure of the uncertainty in the estimate of the structural parameters $\boldsymbol{\theta}$ is provided by the information entropy [15] which is defined by

$$H(D) = \mathrm{E}_{\boldsymbol{\theta}}[-\ln p(\boldsymbol{\theta} \mid D)] = -\int p(\boldsymbol{\theta} \mid D) \ln p(\boldsymbol{\theta} \mid D)\, \mathrm{d}\boldsymbol{\theta}, \tag{6}$$

where $\mathrm{E}_{\boldsymbol{\theta}}$ denotes mathematical expectation with respect to $\boldsymbol{\theta}$. Using the form (4) for the updated PDF $p(\boldsymbol{\theta} \mid D)$, the information entropy takes the simplified form

$$H(D) = \ln \frac{1}{c(D)} + \frac{NN_0 - 1}{2} \mathrm{E}_{\boldsymbol{\theta}}[\ln J(\boldsymbol{\theta}; D)] - \mathrm{E}_{\boldsymbol{\theta}}[\ln \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})] \tag{7}$$

and depends only on the available data D and the sensor configuration $\boldsymbol{\delta}$.

Next, an asymptotic approximation of the information entropy, valid for large number of data ($N \to \infty$), is introduced which will be proved useful in the experimental stage of designing optimal sensor configurations. The asymptotic approximation is obtained by observing that the integrals defining the quantities $1/c(D)$, $\mathrm{E}_{\boldsymbol{\theta}}[\ln J(\boldsymbol{\theta}, D)]$ and $\mathrm{E}_{\boldsymbol{\theta}}[\ln \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})]$ in Eq. (7) can be re-written as Laplace-type integrals and then applying Laplace method of asymptotic expansion [23] to approximate these integrals. Specifically, it can be shown (Appendix A) that for a large number of measured data, i.e., as $N \to \infty$, the following asymptotic results hold for the expressions appearing in Eq. (7):

$$\frac{1}{c(D)} \sim \pi_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \frac{(2\pi)^{N_{\boldsymbol{\theta}}/2} \hat{\sigma}^{-(NN_0-1)}}{\sqrt{\det h(\hat{\boldsymbol{\theta}}, \boldsymbol{\delta})}}, \tag{8}$$

$$\mathrm{E}_{\boldsymbol{\theta}}[\ln J(\boldsymbol{\theta}; D)] \sim \ln \hat{\sigma}^2 \quad \text{and} \quad \mathrm{E}_{\boldsymbol{\theta}}[\ln \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})] \sim \ln \pi_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \tag{9}$$

where $\hat{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\theta}}(D) = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}; D)$ is the optimal value of parameter set $\boldsymbol{\theta}$ that minimizes the measure of fit $J(\boldsymbol{\theta}; D)$ given in Eq. (3), $\hat{\sigma}^2$ is the optimal prediction error given by $\hat{\sigma}^2 = J(\hat{\boldsymbol{\theta}}; D)$, and $\mathbf{h}(\hat{\boldsymbol{\theta}}, \boldsymbol{\delta})$ is an $(N_{\boldsymbol{\theta}} \times N_{\boldsymbol{\theta}})$ positive definite matrix defined, and asymptotically approximated, by

$$\mathbf{h}(\hat{\boldsymbol{\theta}}, \boldsymbol{\delta}) = -[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^{\mathrm{T}} \ln[J(\boldsymbol{\theta}; D)]^{-(NN_0-1)/2}]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \sim \frac{1}{\hat{\sigma}^2} \mathbf{Q}(\hat{\boldsymbol{\theta}}, \boldsymbol{\delta}) \quad \text{as } N \to \infty, \tag{10}$$

while $\boldsymbol{\nabla}_{\boldsymbol{\theta}}^{\mathrm{T}} = [\partial/\partial\theta_1 \cdots \partial/\partial\theta_p]$ is the usual gradient vector with respect to the parameters set $\boldsymbol{\theta}$. The matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$ appearing in Eq. (10) is a positive definite matrix of the form

$$\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta}) = \sum_{j=1}^{N_d} \delta_j \mathrm{P}^{(j)}(\boldsymbol{\theta}) \tag{11}$$

known as the Fisher information matrix [3] and containing the information about the values of the parameters $\boldsymbol{\theta}$ based on the data from all measured positions specified in $\boldsymbol{\delta}$. The matrix $\mathbf{P}^{(j)}(\boldsymbol{\theta})$ is a positive semi-definite matrix given by

$$\mathbf{P}^{(j)}(\boldsymbol{\theta}) = \sum_{m=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{\theta}} x_j(m; \boldsymbol{\theta}) \ \boldsymbol{\nabla}_{\boldsymbol{\theta}}^{\mathrm{T}} x_j(m; \boldsymbol{\theta}) \tag{12}$$

containing the information about the values of the parameters $\boldsymbol{\theta}$ based on the data from one sensor placed at the jth d.o.f.. The matrix $\mathbf{P}^{(j)}(\boldsymbol{\theta})$ depends only on the response of the model at the particular d.o.f. j, while it is independent of the sensor configuration vector $\boldsymbol{\delta}$.

Substituting Eqs. (8) and (9) into Eq. (7) and simplifying, one finally derives that

$$H(D) \sim H(\boldsymbol{\delta}, \hat{\boldsymbol{\theta}}, \hat{\sigma}) = \tfrac{1}{2} N_{\boldsymbol{\theta}}[\ln(2\pi) + \ln \hat{\sigma}^2] - \tfrac{1}{2} \ln[\det \mathbf{Q}(\boldsymbol{\delta}, \hat{\boldsymbol{\theta}})], \tag{13}$$

which implies that the information entropy depends on the sensor configuration vector $\boldsymbol{\delta}$ and the optimal model parameters $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$, while it is independent of the time-history details of the measured data D. The only dependence of the information entropy on the data comes implicitly through the optimal values $\hat{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\theta}}(D)$ and $\hat{\sigma}^2 = J(\hat{\boldsymbol{\theta}}; D)$. The importance of the asymptotic result in the experimental design of sensors and actuators will become evident in the next section.

4. Formulation of optimal sensor configuration problem

In experimental design, it is desirable to design the sensor configuration such that the resulting measured data are most informative about the structural model parameters selected for estimation. The information entropy, introduced in Eq. (6) as the measure of the uncertainty in the system parameters, gives the amount of useful information contained in the measured data. The most informative test data are the ones that give the least uncertainty in the parameter estimates or, equivalently, the ones that minimize the information entropy. Thus, among all sensor configurations, the optimal sensor configuration is selected as the one that minimizes the information entropy. The problem of finding the optimal sensor configuration is formulated as a discrete optimization problem. The objective function is the information entropy, while the discrete variables are related to the number and location of sensors.

It should be emphasized that in the initial stage of designing the experiment the test data are not available. Thus the information entropy defined in Eq. (7) is not specified completely since it depends explicitly on the details contained in the dataset D. In order to further process the information entropy, its explicit dependence on the data D has to be removed. This can be accomplished by considering the limiting case of large number of data ($N \to \infty$), often arising in structural dynamics applications. The resulting asymptotic value of the information entropy, given in Eq. (13), is completely defined by the optimal value $\hat{\boldsymbol{\theta}}$ of the model parameters and the optimal prediction error $\hat{\sigma}^2$ expected for a set of test data, while the time history details of the measured data do not enter explicitly in the formulation.

Moreover, since the data are not available, an estimate of the optimal model parameters $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ cannot be obtained from analysis. Thus, in order to proceed with the design of the optimal sensor configuration, this estimate has to be assumed. In practice, useful designs can be obtained by taking the optimal model parameters $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ to have some nominal values $\boldsymbol{\theta}_0$ and $\hat{\sigma}_0^2$ chosen by the designer to be representative of the system. In this case, the entropy measure in Eq. (13) takes the form (for large N)

$$H(\boldsymbol{\delta}, \boldsymbol{\theta}_0, \sigma_0) = \tfrac{1}{2} N_{\boldsymbol{\theta}}[\ln(2\pi) + \ln \sigma_0^2] - \tfrac{1}{2}\ln[\det \mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta}_0)] \tag{14}$$

and depends on the sensor configuration vector $\boldsymbol{\delta}$, the parametric class of models chosen to describe the behaviour of the structure, and the chosen nominal values of the parameters $\boldsymbol{\theta}_0$ and $\hat{\sigma}_0$ of the models.

The aforementioned analysis provides a rigorous justification of the fact that the optimal sensor design can be based only on a nominal structural model, ignoring the details in the time history of the measured data which are unavailable in the experimental design stage.

5. Dependence of information entropy on number of sensors

In this section useful results are obtained that show the dependence of the information entropy and its lower and upper bounds on a number of sensors. These results will be used in the next section to efficiently construct approximate solutions to optimal sensor location problem.

Let $\boldsymbol{\delta}^{(M)}$ denote the sensor configuration involving M sensors. Define also the expression $\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}$ to represent the sensor configuration that is formed from the configuration $\boldsymbol{\delta}^{(M)}$ and L additional sensors placed on the structure as specified by the configuration $\boldsymbol{\delta}^{(L)}$. Then the following proposition is true:

Proposition 1. *The information entropy for a sensor configuration $\boldsymbol{\delta}^{(M)}$ involving M sensors is higher than the information entropy for a sensor configuration $\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}$ involving L additional sensors. That is,*

$$H(\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}, \boldsymbol{\theta}, \sigma) \leqslant H(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta}, \sigma). \tag{15}$$

Proof. Using Eq. (14), it suffices to show that the following inequality holds for two sensor configurations $\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}$ and $\boldsymbol{\delta}^{(M)}$:

$$\det[\mathbf{Q}(\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}, \boldsymbol{\theta})] \geqslant \det[\mathbf{Q}(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta})]. \tag{16}$$

Exploiting the special form (11) and (12) of the matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$ it can be readily shown that the matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$ is symmetric semi-positive definite since for every non-zero vector $\mathbf{y} \in R^{N_{\boldsymbol{\theta}}}$ the quantity

$$\mathbf{y}^{\mathrm{T}}\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})\mathbf{y} = \sum_{i=1}^{N_d} \delta_i \mathbf{y}^{\mathrm{T}}\mathbf{P}^{(i)}(\boldsymbol{\theta})\mathbf{y} = \sum_{i=1}^{N_d} \delta_i \sum_{k=1}^{N} [\mathbf{y}^{\mathrm{T}}\boldsymbol{\nabla}_\theta x_i(k; \boldsymbol{\theta})]^2 \geqslant 0 \tag{17}$$

is always non-negative. Also, using the special form (11) and (12) it is evident that the matrix $\mathbf{Q}(\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}, \boldsymbol{\theta})$ with $L \geqslant 1$ admits the representation

$$\mathbf{Q}(\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}, \boldsymbol{\theta}) = \mathbf{Q}(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta}) + \boldsymbol{\delta}\mathbf{Q}(\boldsymbol{\theta}), \quad L \geqslant 0, \tag{18}$$

where $\delta\mathbf{Q}(\boldsymbol{\theta})$ is also a symmetric semi-positive definite matrix. Substituting Eq. (18) into Eq. (16), it remains to show the validity of the inequality

$$\det[\mathbf{Q}(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta}) + \delta\mathbf{Q}(\boldsymbol{\theta})] \geqslant \det[\mathbf{Q}(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta})], \quad L \geqslant 0. \tag{19}$$

This statement can be shown using the fact that for two symmetric semi-positive definite matrices $\mathbf{A} \in R^{N_\theta \times N_\theta}$ and $\mathbf{B} \in R^{N_\theta \times N_\theta}$ the following is true:

$$\lambda_r[\mathbf{A} + \mathbf{B}] \geqslant \lambda_r[\mathbf{A}] \geqslant 0, \quad r = 1, \ldots, N_\theta, \tag{20}$$

where the symbol $\lambda_r[\mathbf{A}]$ denotes the r eigenvalue of the matrix $\mathbf{A}$. The last inequality can be derived from the application of the minimax theorem for eigenvalues of symmetric matrices. Applying inequality (20) for $\mathbf{A} = \mathbf{Q}(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta})$ and, using the fact that $\det \mathbf{A} = \prod_{r=1}^{N_\theta} \lambda_r[\mathbf{A}]$ for any matrix $\mathbf{A}$, inequality (19) is readily derived. □

Proposition 1 implies that the information entropy reduces as additional sensors are placed in a structure. Given the interpretation of the information entropy as a measure of the uncertainty in the parameter estimates, this should be intuitively expected since adding one or more sensors in the structure will have the effect of providing more information about the system parameters and thus reducing the uncertainty in the parameter estimates.

Let $\boldsymbol{\delta}_{opt}^{(M)}$ and $\boldsymbol{\delta}_{worst}^{(M)}$ denote the optimal and worst sensor configurations for M sensors, respectively. Let also $H_{min}^{(M)} = H(\boldsymbol{\delta}_{opt}^{(M)}, \boldsymbol{\theta}, \sigma)$ and $H_{max}^{(M)} = H(\boldsymbol{\delta}_{worst}^{(M)}, \boldsymbol{\theta}, \sigma)$ be the minimum and maximum information entropy values corresponding to the optimal and worst sensor configurations $\boldsymbol{\delta}_{opt}^{(M)}$ and $\boldsymbol{\delta}_{worst}^{(M)}$, respectively. As a direct consequence of Proposition 1, the following proposition is true.

Proposition 2. *The minimum and maximum information entropies for M sensors are decreasing functions of the number of sensors, M. Mathematically, this could be stated as*

$$H_{min}^{(M+L)} \leqslant H_{min}^{(M)} \tag{21}$$

and

$$H_{max}^{(M+L)} \leqslant H_{max}^{(M)}. \tag{22}$$

This reduction of the information entropy as a function of the number of sensors is expected since increasing the number of sensors has an effect of extracting more information from the data. This trend was observed for the minimum information entropy in the numerical results presented in Ref. [14]. Next, a proof of inequalities (21) and (22) is provided.

Specifically, replacing $\boldsymbol{\delta}^{(M)}$ in Eq. (15) by $\boldsymbol{\delta}_{opt}^{(M)}$ and using the fact that $H_{min}^{(M)} = H(\boldsymbol{\delta}_{opt}^{(M)}, \boldsymbol{\theta}, \sigma)$, Eq. (15) takes the form

$$H(\boldsymbol{\delta}_{opt}^{(M)} + \boldsymbol{\delta}^{(L)}, \boldsymbol{\theta}, \sigma) \leqslant H(\boldsymbol{\delta}_{opt}^{(M)}, \boldsymbol{\theta}, \sigma) \equiv H_{min}^{(M)}. \tag{23}$$

Also, noting that $H_{min}^{(M+L)} \leqslant H(\boldsymbol{\delta}^{(M+L)}, \boldsymbol{\theta}, \sigma)$ for every sensor configuration $\boldsymbol{\delta}^{(M+L)}$ involving $M + L$ sensors and choosing $\boldsymbol{\delta}^{(M+L)} = \boldsymbol{\delta}_{opt}^{(M)} + \boldsymbol{\delta}^{(L)}$, it follows that $H_{min}^{(M+L)} \leqslant H(\boldsymbol{\delta}_{opt}^{(M)} + \boldsymbol{\delta}^{(L)}, \boldsymbol{\theta}, \sigma)$. Statement Eq. (21) follows directly from the last relation and Eq. (23).

The proof of Eq. (22) follows similar arguments to those used for showing the validity of Eq. (21). Specifically, replacing $\boldsymbol{\delta}^{(M)} + \boldsymbol{\delta}^{(L)}$ in Eq. (15) by $\boldsymbol{\delta}_{worst}^{(M+L)}$ and using the fact that $H_{max}^{(M+L)} \equiv H(\boldsymbol{\delta}_{worst}^{(M+L)}, \boldsymbol{\theta}, \sigma)$, Eq. (15) takes the form

$$H_{max}^{(M+L)} \equiv H(\boldsymbol{\delta}_{worst}^{(M+L)}, \boldsymbol{\theta}, \sigma) \leqslant H(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta}, \sigma). \tag{24}$$

Expression (22) follows directly from the fact that $H(\boldsymbol{\delta}^{(M)}, \boldsymbol{\theta}, \sigma) \leqslant H_{max}^{(M)}$ for every sensor configuration $\boldsymbol{\delta}^{(M)}$ involving M sensors.

6. Computational issues

For a structural model with N_d d.o.f.'s, the number of all distinct sensor configurations involving N_0 sensors is

$$N_s = \frac{N_d!}{N_0!(N_d - N_0)!} \tag{25}$$

which for most cases of practical interest can be an extremely large number. Therefore, an exhaustive search over all sensor configurations for the computation of the optimal sensor configuration is extremely time consuming and in most cases prohibitive even for models with a relatively small number of d.o.f.'s. Alternative approximate techniques must be used to solve the discrete optimization and obtain good estimates of sensor configurations that correspond to information entropy values close to the minimum information entropy. Two such techniques are next presented and compared. The first is based on genetic algorithms which are most suitable for solving general discrete optimization problems. The second technique is based on a sequential sensor placement algorithm that exploits the theoretical results stated in Proposition 2.

6.1. Sensor placement using genetic algorithm (GA)

Genetic algorithms are most suitable for solving the resulting discrete optimization problem and providing near optimal solutions. For an introduction to the GA theory the reader is referred to standard textbooks [24,25]. Herein, the optimization is carried out using a simple GA [24] that for N_d d.o.f.'s and N_0 sensors proceeds as follows. The number of variables to be optimized equals the number of sensors to be placed on the structure. The range of values of each variable is determined by the number of d.o.f. of the model. Specifically, each variable takes integer values ranging from 1 to N_d. Thus a string of $n_{bit} = ceil(\log N_d / \log 2)$ bits is used for binary representation of the possible values of each variable. The size of the string is selected to cover the whole range of variation of each variable. For N_0 sensors, the number of bits forming the chromosome in GAs is $n_{bit}N_0$. Thus a possible solution is represented by a string of length $n_{bit}N_0$ bits. A population of possible solutions of a fixed size is initialized randomly in a bitwise fashion. The genetic operators of mutation and the crossover are used to generate the population in the next generation. The probability of mutation p_m and the probability of crossover p_c are kept fixed. The fitness function used for computing the optimal sensor configuration is taken as the value of the information entropy defined in Eq. (14).

It should be noted that for structural models with number of d.o.f.'s N_d less than $2^{n_{bit}}$, where n_{bit} is the number of bits per string, the above representation may give strings with corresponding values greater than the total number of d.o.f.'s of the model. The selection of such strings in the next generation is avoided by assigning them a very large fitness value. Another issue encountered in the applications of GAs is that the binary representation of the variables may yield sensor configurations involving less than N_0 distinct positions. The fitness of such strings is evaluated using only the number of sensors that correspond to distinct positions in the structure, thus neglecting the common sensor positions. Such configuration will involve less than N_0 sensors but they usually have very good fitness values and they should be left in the population. Invalidating these strings by assigning a very large fitness value may decelerate the convergence of the GA.

6.2. Sequential sensor placement (SSP) algorithms

A more systematic and computationally very efficient approach for obtaining a good sensor configuration for a fixed number of N_0 sensors is to use a sequential sensor placement algorithm as follows. The positions of N_0 sensors are computed sequentially by placing one sensor at a time in the structure at a position that results in the highest reduction in information entropy. Specifically, the position of the first sensor is chosen as the one that gives the highest reduction in the information entropy for one sensor. Given the optimal position of the first sensor, the position of the second sensor is chosen as the one that gives the highest reduction in the information entropy computed for two sensors with the position of the first sensor fixed at the optimal one already computed in the first step. Continuing in a similar fashion, given the positions of $(i-1)$ sensors in the structure computed in the previous $(i-1)$ steps, the position of the next ith sensor is obtained as the one that gives the highest reduction in the information entropy for i sensors with the positions of the first $(i-1)$ sensors fixed at the optimal ones already obtained in the previous $(i-1)$ steps. This procedure is continued for up to N_0 sensors.

This construction of the sensor configuration exploits the theoretical results stated in Proposition 2, that is, the lower bound of the information entropy is a decreasing function of the number of sensors. The sequential sensor placement algorithm will give the optimal sensor configuration only in the case for which the optimal sensor positions for i sensors is a subset of the optimal sensor positions for $(i+1)$ sensors for all i from one to N_0. However, the last argument does not hold in general and the sensor configuration computed by the sequential sensor placement algorithms cannot be guaranteed to be the optimal one. The sensor configurations estimated from the sequential sensor placement algorithm provide information entropy values that are upper bounds of the minimum information entropy. Numerical applications presented next in the application section show that these bounds in most cases examined coincide with, or are very close to, the exact minimum information entropy. Compared to the GA algorithm, the SSP algorithms are preferred since they are found to maintain higher levels of accuracy with less computational effort than that involved in GAs.

For the sake of reference, the aforementioned algorithm is termed the forward sequential sensor placement (FSSP) algorithm. The SSP algorithm can also be used in an inverse order, starting with N_d sensors placed at all d.o.f.'s of the structure and removing successively one sensor at a time from the position that results in the smallest increase in the information entropy. This algorithm is termed as the backward sequential sensor placement (BSSP).

The successive placement of each sensor in the structure using the SSP algorithm requires the optimization of the information entropy with respect to one sensor location. The solution is easily provided using an exhaustive search of the parameter space. Using the forward SSP algorithm, the total number of function evaluations for optimally placing the ith sensor given that $(i-1)$ sensors have already been placed in the structure is equal to $(N_d - i + 1)$, where N_d is the total number of d.o.f.'s. Thus, the total number of function evaluations required for designing the "optimal" sensor configuration for N_0 sensors is $\sum_{i=1}^{N_0}(N_d - i + 1) \leqslant N_0 N_d$. The upper bound $N_0 N_d$ is a good estimate for small number of sensors N_0 relative to N_d. Moreover, the design of sensor configurations from one up to N_d sensors requires a total of $N_d(N_d+1)/2$ function evaluations which is extremely small number compared to the number N_s, given in Eq. (25). Similarly, the total number of function evaluations required by the backward SSP algorithm is $\sum_{i=1}^{N_d-N_0+1}(N_d - i + 1) \leqslant N_d(N_d+1)/2$, where the upper bound $N_d(N_d+1)/2$ is a good estimate for small number of sensors N_0 relative to N_d. For a small number of sensors relative to the number of d.o.f.'s, usually the case in structural dynamics applications, the BSSP algorithm requires $N_d/(2N_0)$ times more computational effort than the FSSP algorithm. Thus, from the computational point of view, the FSSP algorithm should be the preferred algorithm in applications.

6.3. Upper bound of information entropy

The upper bound of the information entropy corresponding to the worst sensor configuration is also useful since when it is compared with the minimum information entropy for the same number of sensors it gives a measure of the reduction that can be achieved by optimizing the sensor configuration. The maximum information entropy and the corresponding worst sensor configuration can be obtained from the aforementioned algorithms by maximizing instead of minimizing the information entropy.

In fact, the GA software used for finding the optimal sensor configuration can also be used with slight modifications to find the worst sensor configuration. One major difference though is that the application of GAs will tend to converge to an invalid solution which involve less than N_0 sensors. To avoid the convergence to such solutions, the fitness of configurations involving less than N_0 sensors is given very small values. Due to the several invalid configurations encountered in the search for the worst configuration, the rate of convergence of GAs when used to find the worst sensor configuration is slower than the one corresponding to the optimal sensor configuration.

Using the FSSP algorithm an approximation to the worst sensor configuration is obtained by placing successively one sensor at a time in the position that results in the smallest decrease in information entropy. Similarly, using the BSSP algorithm, an approximation to the worst sensor configuration is obtained by removing successively one sensor at a time from the position that results in the highest increase in the information entropy value.

7. Applications

The objective of the applications is to explore, illustrate and compare the effectiveness, in terms of accuracy and computational efficiency, of the three methods: the exhaustive search method (exact), the two sequential sensor placement methods and the method based on GAs. The exhaustive search method, which provides exact results against which comparisons should be made, is computationally prohibitive for more than a few number of model d.o.f.'s. Thus, in order to investigate the accuracy of the proposed approximate SSP methods, the methodology is first applied to a 10 d.o.f. model for which exact results for comparison are readily obtained from the exhaustive search method. Then the two approximate SPP methods and the approximate GA method are compared for their computational efficiency and accuracy using a truss model that involves a considerably higher number of 240 d.o.f.'s. In this case, however, the exact results from the exhaustive search method are not available even for as few as four sensors due to the prohibitively large number of computations required.

7.1. Ten-d.o.f. chain-like spring–mass model

In order to facilitate comparisons between the exact exhaustive search method and the two SSP algorithms, the methodology is first applied to a 10 d.o.f. chain-like spring mass model, shown in Fig. 1. The structure is parameterized using 10 parameters, with the ith parameter modelling the ith spring stiffness k_i. The masses are considered to be same for all links in the chain. The nominal structure corresponds to uniform stiffness distribution along the chain. The nominal values of the spring stiffnesses and masses, respectively, are chosen such that the fundamental frequency of the nominal structure is approximately 0.9 Hz. Classical normal modes are assumed with the modal damping fixed at 5% for all modes. The structure is subjected to an impulse excitation of unit magnitude applied on the tenth mass of the model. This impulse excitation can be viewed as simulating the excitation from impact hammer tests. The responses of the nominal model at all model d.o.f., needed in the calculation of the information entropy (see Eqs. (11), (12) and (14)), are readily obtained using modal analysis and deriving analytically the modal responses from the modal equations of motion of the structure.

The information entropy values $\mathbf{h}(\boldsymbol{\delta}, \boldsymbol{\theta}_0, \sigma_0)$ for a sensor configuration vector $\boldsymbol{\delta}$ are used to construct the information entropy index defined by $IEI(\boldsymbol{\delta}) = \exp[(\mathbf{h}(\boldsymbol{\delta}, \boldsymbol{\theta}_0, \sigma_0) - \mathbf{h}(\delta_{ref}, \boldsymbol{\theta}_0, \sigma_0)/N_\theta)]$, where $\mathbf{h}(\delta_{ref}, \boldsymbol{\theta}_0, \sigma_0)$ is the reference information entropy computed for a referenced sensor configuration $\boldsymbol{\delta}_{ref}$. The $IEI(\boldsymbol{\delta})$ is a measure of the uncertainty in the parameter estimates relative to the uncertainty obtained for a referenced sensor configuration. The referenced sensor configuration is selected as the one involving sensors at all d.o.f.'s so that the $IEI(\boldsymbol{\delta})$ values, when compared to one, give the effectiveness of the sensor configuration $\boldsymbol{\delta}$ and the maximum improvement that can be achieved by sensor re-configuration strategies.

The minimum and maximum information entropy index values $IEI(\boldsymbol{\delta})$ as a function of the sensors computed by the exhaustive search method (exact method) and the forward and backward SSP methods are shown in Figs. 2(a), 3(a) and 4(a) for 10, 5 and 2 observable modes, respectively. The corresponding condition numbers for the information matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$ defined in Eq. (11) are

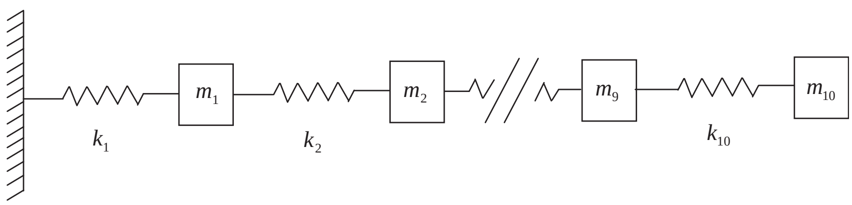


Fig. 1. Ten-d.o.f. chain-like spring–mass model.

also computed using the exhaustive search method and shown in Figs. 2(b), 3(b) and 4(b). It should be noted that the dimension of the information matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$ is 10.

The information entropy estimates predicted by the forward and backward SSP methods are, in most cases, extremely good approximations of the minimum information entropy. Comparing the predictions from the FSSP and the BSSP methods, the results may slightly differ depending on the sensor case considered. This difference is more pronounced in the case of maximum information entropy predictions. An effective use of the FSSP and BSSP algorithms is to combine their predictions as follows. Two predictions of the optimal (respectively worst) sensor configuration, one for each method, are first constructed. Then among these two predictions the one that corresponds to the minimum (respectively maximum) information entropy value is selected as the best one.

It should be mentioned that for a small number of sensors and/or a small number of observable modes, the 10-parameter model is unidentifiable or almost unidentifiable. This gives rise to non-unique solutions encountered in inverse problems for parameter identification [26–29]. The unidentifiability is reflected in the very large values of the condition number of the matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$, defined in Eq. (11). Specifically, the very large values, of the order of 10^{17}, shown in Fig. 4(b) suggest that the model parameters are unidentifiable for the case of 2 observable modes and sensor configurations involving up to 3 sensor. For sensor configurations involving 4 to 6 sensors, the identifiability depends on the sensor locations since the optimal sensor configuration corresponds to a well-conditioned matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$, while the worst sensor configuration corresponds to an ill-conditioned matrix. This case clearly reveals the importance of optimizing the sensor configuration for 4–6 sensors and 2 observable modes in order to avoid configurations that do not provide enough information for obtaining unique solutions to the parameter estimation problem. Finally, it should be observed that the accuracy of the FSSP method deteriorates for the ill-conditioned cases shown in Figs. 3 and 4.

The optimal sensor locations and the corresponding information entropy index values for 1–10 sensors and 10 observable modes are compared in Table 1 for the exact method and the two SSP methods. It is seen that both the SSP algorithms correctly predict the optimal sensor locations for 5–10 sensors. The FSSP algorithm correctly predicts the optimal sensor location for 1 and 4 sensors, while the BSSP algorithm correctly predicts the optimal sensor location for 3 sensors. The combined FSSP and BSSP predictions, shown in boldface in Table 1, coincide in almost all cases with the exact one.

7.2. *Three-dimensional 240-d.o.f. truss structure*

The methodology is next applied to a 240-d.o.f. 20-bay three-dimensional truss structure shown schematically in Fig. 5. For illustration purposes, it is assumed that all vertical and horizontal members of the nominal truss have the same sizes and that the mass of the structure is uniformly lumped at the nodes of the truss. Classical normal modes are assumed with the modal damping fixed at 5% for all modes. The structure is subjected to an impulse excitation of unit magnitude at the top of the truss towards the horizontal direction, shown in Fig. 5.

An eight-parameter model is considered with the eight parameters modelling the stiffness of the eight diagonals at the first (lower) and second bay of the truss shown with broken lines in Fig. 5(b). The parameterization represents the case for which the truss is to be monitored for local

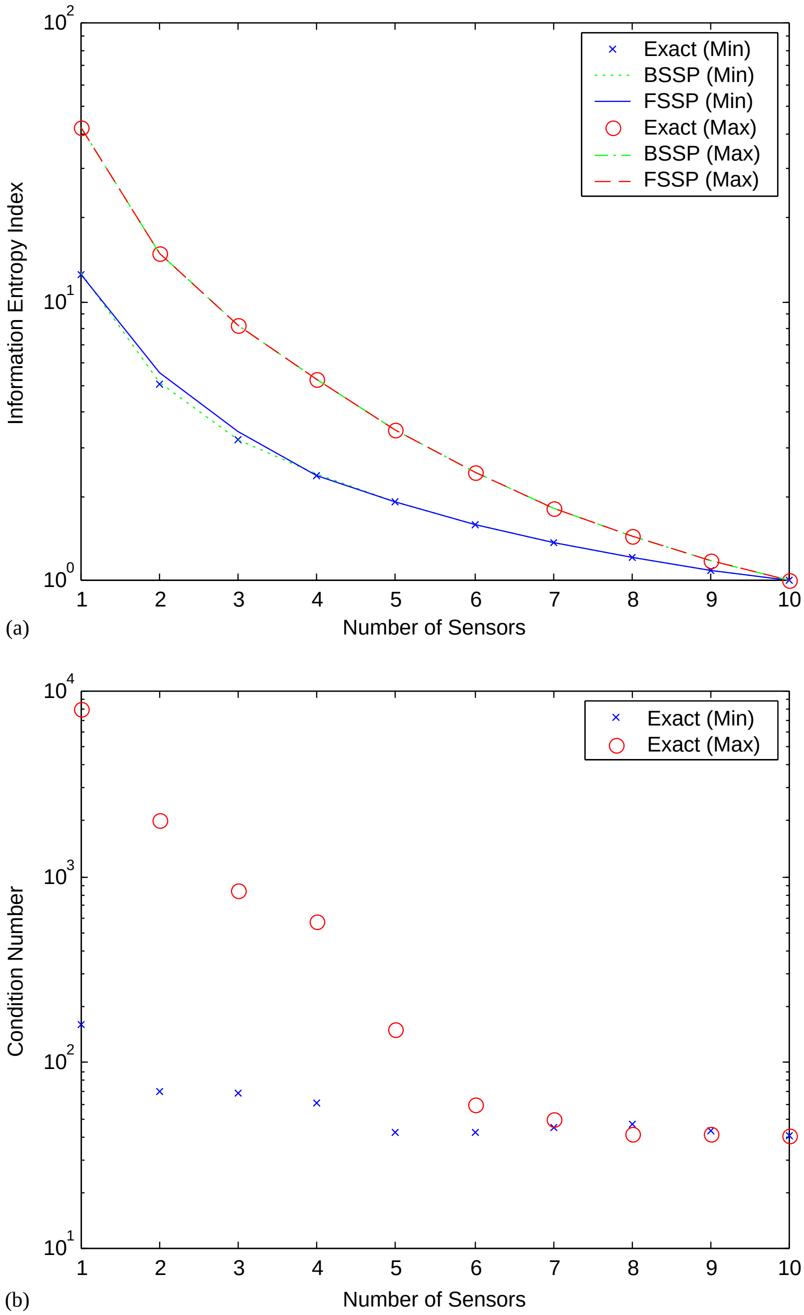
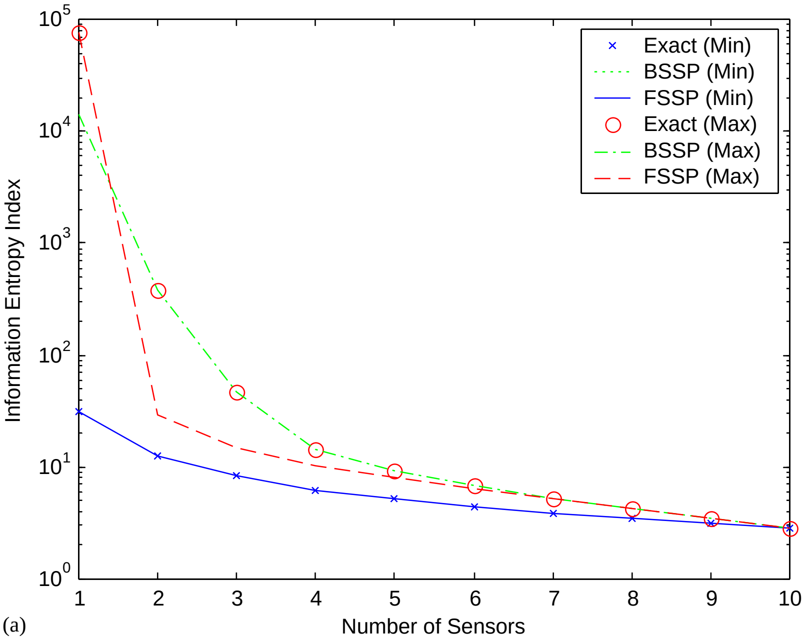
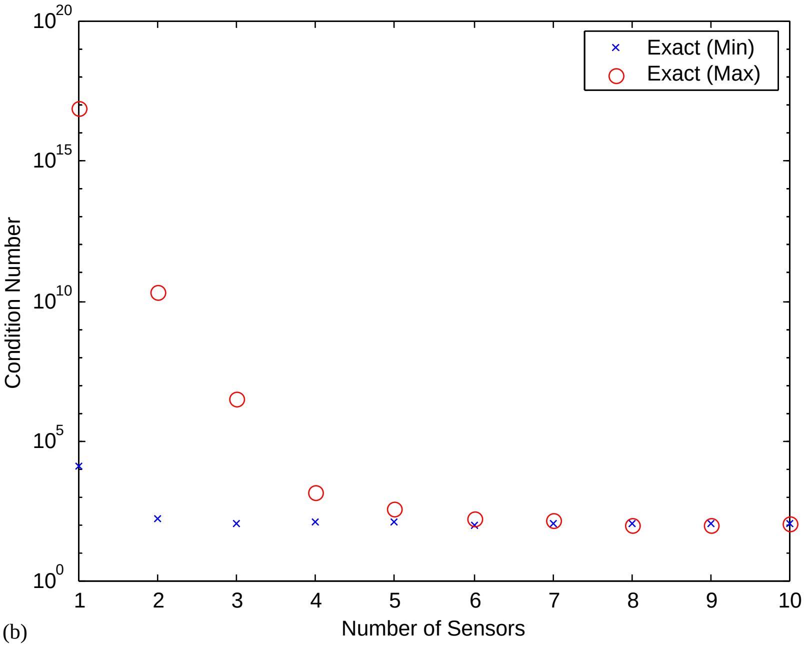


Fig. 2. (a) Minimum and maximum information entropy index values $IEI(\boldsymbol{\delta})$ for 10 observable modes; (b) Condition number of matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$.

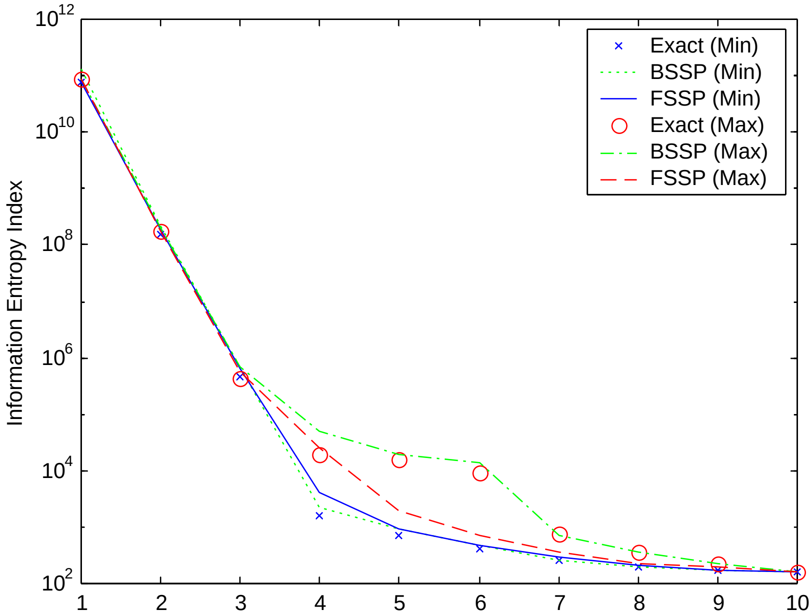


(a)

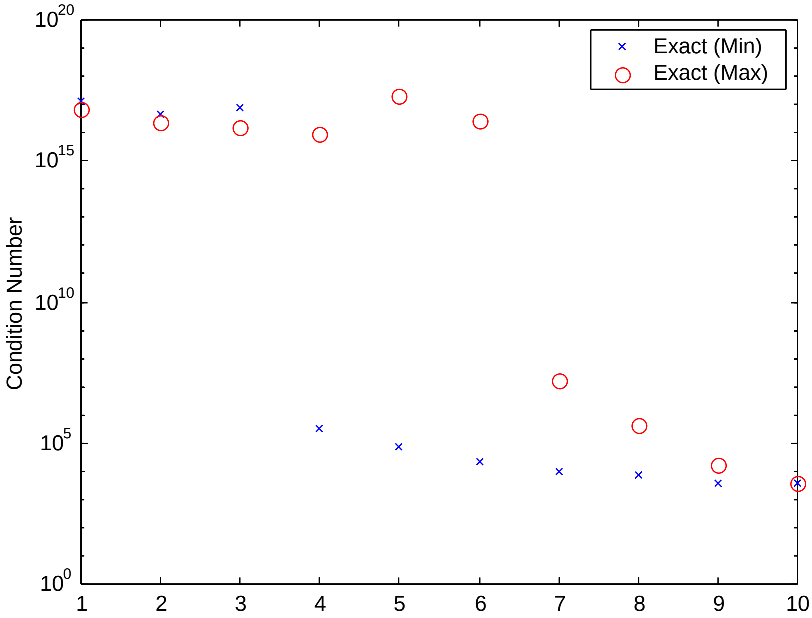


(b)

Fig. 3. (a) Minimum and maximum information entropy index values $IEI(\boldsymbol{\delta})$ for 5 observable modes; (b) Condition number of matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$.



(a)



(b)

Fig. 4. (a) Minimum and maximum information entropy index values $IEI(\boldsymbol{\delta})$ for 2 observable modes; (b) Condition number of matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$.

Table 1
Comparison between the exhaustive search (exact), FSSP and BSSP predictions of optimal sensor locations and corresponding information entropy index values for 1–10 sensors

EXACT Method

Number of sensors	1	2	3	4	5	6	7	8	9	10	*IEI*
1										■	23.34
2					■				■		7.48
3				■			■		■		4.21
4				■			■		■	■	2.97
5		■		■			■	■	■		2.25
6		■		■			■	■	■	■	1.78
7		■		■		■	■	■	■	■	1.47
8		■		■	■	■	■	■	■	■	1.25
9		■	■	■	■	■	■	■	■	■	1.11
10	■	■	■	■	■	■	■	■	■	■	1.00

FSSP Method

Number of sensors	1	2	3	4	5	6	7	8	9	10	*IEI*
1										■	**23.34**
2							■			■	8.48
3				■			■			■	4.60
4				■			■		■	■	**2.97**
5		■		■			■		■	■	**2.25**
6		■		■			■	■	■	■	**1.78**
7		■		■		■	■	■	■	■	**1.47**
8		■		■	■	■	■	■	■	■	**1.25**
9		■	■	■	■	■	■	■	■	■	**1.11**
10	■	■	■	■	■	■	■	■	■	■	**1.00**

BSSP Method

Number of sensors	1	2	3	4	5	6	7	8	9	10	*IEI*
1									■		23.63
2				■					■		**7.67**
3				■			■		■		**4.21**
4		■		■			■		■		3.00
5		■		■			■	■	■		**2.25**
6		■		■			■	■	■	■	**1.78**
7		■		■		■	■	■	■	■	**1.47**
8		■		■	■	■	■	■	■	■	**1.25**
9		■	■	■	■	■	■	■	■	■	**1.11**
10	■	■	■	■	■	■	■	■	■	■	**1.00**

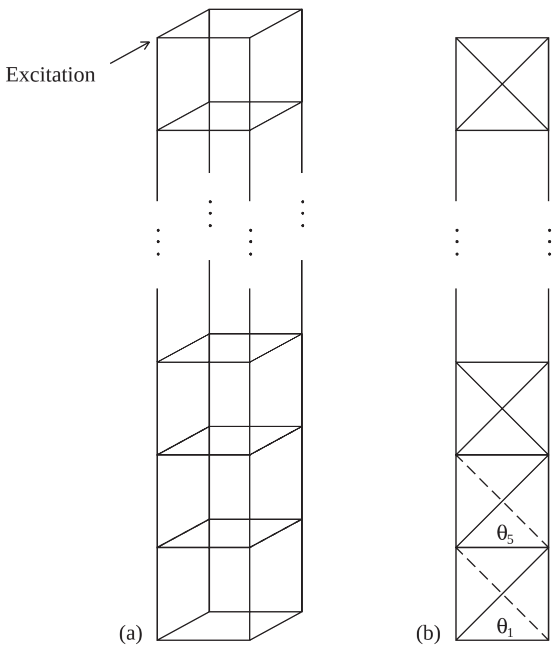


Fig. 5. Two hundred and forty-d.o.f. 20-bay three-dimensional truss; (a) Schematic diagram (diagonal members not shown), (b) Face view of truss with diagonal members shown.

damage at the lowest two bays, caused by environmental effects or severe loads. The dimension of the information matrix $\mathbf{Q}(\boldsymbol{\delta}, \boldsymbol{\theta})$ is 8 in this case.

The objective is to explore and compare the effectiveness, in terms of computational efficiency and accuracy, of the approximate methods: the two SSP methods and the method based on GAs. The number of sensor configurations required to be searched for optimality using the exhaustive

Table 2
Number of function evaluations using the exhaustive search (exact) and the SSP methods

Method	Number of sensors								
	2	4	8	12	16	20	25	30	50
Exact	2.8×10^4	1.3×10^8	2.4×10^{14}	5.7×10^{19}	3.4×10^{24}	7.3×10^{28}	5.6×10^{33}	1.4×10^{38}	1.3×10^{52}
FSSP	479	954	1892	2814	3720	4610	5700	6765	10775
BSSP	28919	28914	28892	28854	28800	28730	28620	28485	27695

search method is given in Table 2 as a function of the number of sensors in a configuration. The number of function evaluations required to construct the optimal sensor configuration for 2, 4, 8, 12, 16, 20, 25, 30 and 50 sensors using the FSSP and BSSP methods are also shown in the table. It is clear that an exhaustive search is computationally very expensive even for the case of a few sensors, while it is prohibitive for more than 4 sensors. In contrast, the SSP algorithms are computationally effective alternatives for constructing approximations to the optimal or worst sensor configurations. Comparing the computational effort of the FSSP and BSSP methods, the FSSP method is preferred since it requires one to two orders of magnitude less effort than the BSSP method. More specifically, the results in Table 2 suggests that the computational effort in FSSP method is 80 and 2.4 times less than that in BSSP method for 2 and 50 sensors, respectively.

The accuracy of the sensor configurations provided by the SSP algorithms is evaluated by comparing the corresponding information entropy values with those computed by GAs. For this, optimal and worst sensor configurations are computed for 1 to 240 sensors using the SSP algorithm and for 2, 4, 8, 12, 16, 20, 25, 30 and 50 sensors using GAs. In all runs the GA results are computed using the following choices: probability of mutation $p_m = 0.01$, probability of crossover $p_c = 0.9$, population size 20 and number of generation 2000. This choice corresponds to 40,000 function evaluations and should be compared to the function evaluations given in Table 2 for the SSP methods. Results from the exhaustive search method (exact method) have also been obtained for a very limited number of 2 sensors.

Figs. 6 and 7 show the variation of the corresponding predictions of the lower and upper bounds of the information entropy index as a function of the number of sensors placed at their optimal and worst locations, respectively. In order to study the effect of observable modes on the optimal sensor configuration, results in Figs. 6 and 7 correspond to 20 and 10 observable modes, respectively. For comparison purposes the reference value $\mathbf{h}(\boldsymbol{\delta}_{ref}, \boldsymbol{\theta}_0, \sigma_0)$ used in the definition of the information entropy index is chosen to correspond to the optimal sensor configuration case for which all two hundred and forty (240) d.o.f.'s of the structure are instrumented with sensors, and 20 modes of the structure are observable.

The BSSP and FSSP methods give approximately the same predictions for the minimum information entropy for almost all cases considered. The performance of the two methods for predicting the maximum information entropy values depends on the number of sensors and observable modes. The combined results provided by the SSP methods are in all cases better than the ones provided by the GAs. Considering also the higher computation effort involved in the GA estimates, it becomes clear that the SSP methods are considerably more effective than the GA method. In addition, the application of the SSP methods for 1 and N_d sensors provide predictions

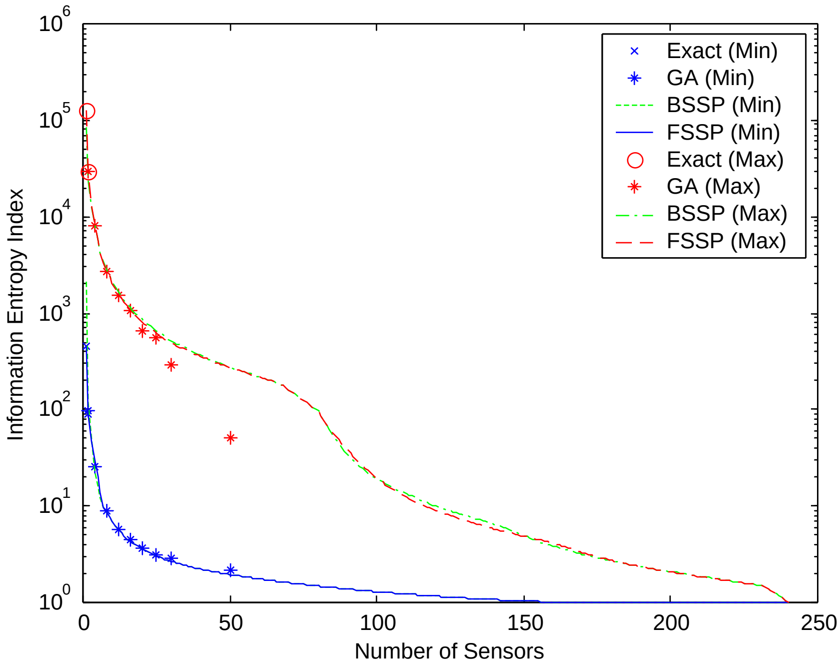


Fig. 6. Minimum and maximum information entropy index values $IEI(\boldsymbol{\delta})$ for 20 observable modes.

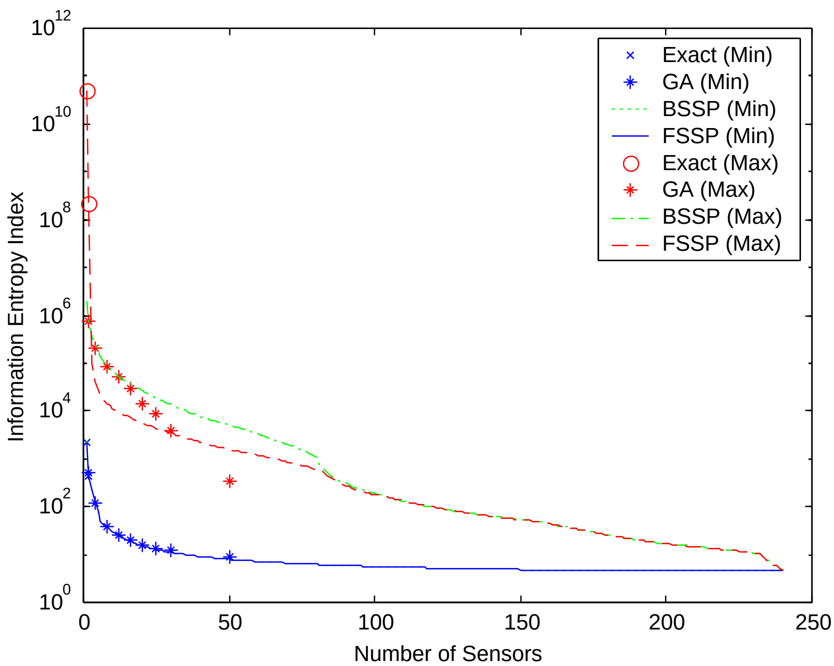


Fig. 7. Minimum and maximum information entropy index values $IEI(\boldsymbol{\delta})$ for 10 observable modes.

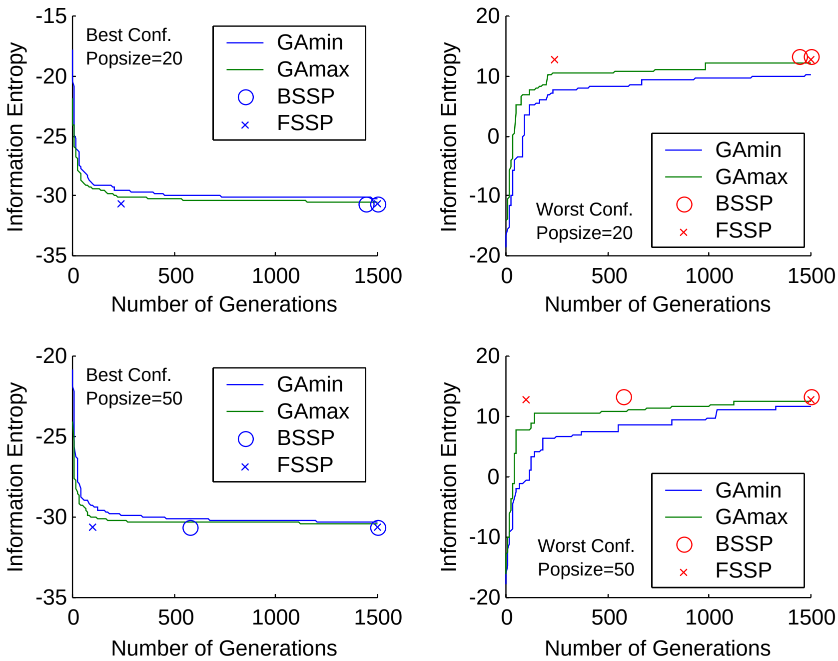


Fig. 8. Information entropy values corresponding to the optimal and worst sensor configurations computed for 20 sensors as a function of the number of generations.

of the upper and lower bounds of the information entropy for sensor configurations involving any number of sensors ranging from 1 to N_d, while GAs provide predictions for an optimal configuration involving a fixed number of sensors only. Finally, comparing the predictions of the SSP methods with the exact ones available for the case of one and two sensors, it is observed that the FSSP method gives better predictions of the optimal and worst sensor configurations than the predictions from the BSSP method.

To further illustrate the rate of convergence and effectiveness of GAs, the optimal and worst information entropy values computed for 20 sensors using GAs is plotted as a function of the number of generations and for various population sizes in Fig. 8. It should be noted that the curves identified as "GAmin" and "GAmax" in these figures correspond to the minimum and maximum values of the information entropy obtained from GAs using five independent runs. For comparison purposes, the information entropy values predicted by the FSSP and BSSP algorithms are also shown in these figures at generation number 1500. In order to make direct comparisons of computational effort between methods, the SSP predictions are also marked at intermediate generation numbers corresponding to function evaluations for the GAs approximately equal to the function evaluations required from either FSSP or BSSP method. The predictions from the two SSP methods are very close. Thus, the FSSP method is more effective since it requires one order of magnitude less computational effort than the BSSP method. It is also clearly shown in these figures that for the same computational effort, the SSP methods always give superior estimates than the GA estimate. It is noted, however, that as the number of generations increases, the GA estimate converges closer to the combined estimate provided by the SSP algorithms. In all cases examined, the predictions from the SSP algorithms outperform those obtained from the GAs for a very large number of generations.

It is worth observing in Figs. 6, 7 and 2–4 that the minimum or maximum information entropy is a decreasing function of the number of sensors placed in the structure at the optimal or worst positions, respectively. This is consistent with the theoretical result stated in Proposition 2. For the truss structure in Figs. 6 and 7, a drastic reduction of the minimum information entropy is observed for the first 10–20 sensors, while the rate of reduction for the next sensors is slower.

Another observation that is worth mentioning is related to the difference between the upper and lower bounds of the information entropy index computed at the optimal and worst sensor locations. This difference is a direct measure of the maximum improvement that can be achieved by optimizing the sensor locations in the structure. This difference is seen in Figs. 6, 7 and 2–4 to decrease monotonically with increasing the number of sensors, which suggests that the fewer the number of sensors, the more effective an optimal sensor placement technique will be.

Also, one can conclude that a given number of sensors placed at their optimal locations may yield much better information than a higher number of sensors arbitrarily placed in the structure. For example, it is seen in Fig. 6 that 8 and 20 sensors placed at their optimal locations may yield better information than 120 and 160 sensors, respectively, arbitrarily placed in the structure. Thus optimizing the sensors in the structure is highly desirable and can result in significant reduction in the cost of the instrumentation.

The optimal sensor locations computed using the BSSP method are presented in Fig. 9 for up to 16 sensors and the case of 20 observable modes. The results from the FSSP method, although they can be obtained with approximately one order of magnitude less effort, are less accurate than the BSSP results for 4–6 sensors, while they coincide with the BSSP results for 7–16 sensors. For the optimal sensor configuration, the first four sensors (sensors 1–4) are placed at the first bay at

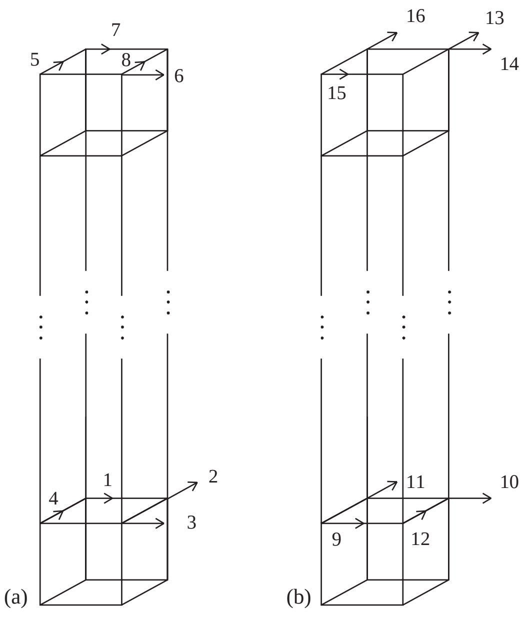


Fig. 9. Optimal sensor locations based on the SSP method for (a) first 8 sensors, (b) next 8 sensors.

d.o.f.'s pointing along the two horizontal directions, while the next four sensors (sensors 5–8) are placed at the top bay at d.o.f.'s pointing along the horizontal direction. The next four sensors (sensors 9–12) are placed at the first bay, while the next four sensors (sensors 13–16) are placed at the top bay at d.o.f.'s pointing along the two horizontal directions. Finally, the worst sensor configuration computed for up to 80 sensors using the SSP methods is the one for which all sensors are monitoring d.o.f.'s pointing along the vertical direction.

8. Conclusions

A rigorous formulation of the optimal sensor placement problem for structural identification was presented based on the information entropy measure of parameter uncertainty. An asymptotic estimate, valid for large number of data, was derived and used to justify that the sensor placement design can be based solely on a nominal model, ignoring the details in the measured data. The analysis also showed that the lower and upper bounds of the information entropy values for a fixed number of sensors, corresponding respectively to the optimal and worst sensor configuration, is a decreasing function of the number of sensors.

Based on the analysis, two heuristic sensor placement algorithms, the BSSP and FSSP, were proposed for constructing predictions of the optimal and worst sensor configurations. The computations involved in the SSP algorithm are an infinitesimal fraction of the ones involved in the exhaustive search method and can be done in realistic time, independently of the number of sensors and the number of model d.o.f.'s. Genetic algorithms, well-suited for solving the resulting discrete optimization problem, were also used to provide an estimate of the optimal sensor location for a fixed number of sensors. The effectiveness of the proposed algorithms was evaluated based on two example applications involving stiffness-related parameter identification in structural dynamics. It was found that for essentially the same accuracy, the GA algorithm requires significantly more computational effort than the heuristic SSP algorithms. In almost all cases considered, the estimate from the GA algorithm did not improve the estimate provided by the SSP algorithms. Thus, although the SSP algorithms are not guaranteed to give the optimal solution, they were found to be effective and computationally attractive alternatives to the GAs. In particular, SSP algorithms provide with minimal computational effort the variation of the lower and upper bounds of the information entropy as a function of the number of sensors. Such bounds are useful in evaluating the effectiveness of a sensor configuration as well as in guiding the cost-effective selection of the number of sensors, trading-off information provided from extra sensors with cost of instrumentation.

Acknowledgements

This research was funded by the Greek General Secretariat of Research and Technology and the European Union under grant PENED 99ED580. Additional funding was provided by the Greek Earthquake Planning and Protection Organization. This support is gratefully acknowledged.

Appendix A. Proof of asymptotic approximations (8) and (9)

A well-known asymptotic approximation for the Laplace type integral [23] is used in the proof of asymptotic estimates (8) and (9). Specifically, let $f(\boldsymbol{\theta})$ be a twice differentiable function of $\boldsymbol{\theta}$, $g(\boldsymbol{\theta})$ be a continuous function of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ be the value of $\boldsymbol{\theta}$ that globally maximizes $f(\boldsymbol{\theta})$. Then the following asymptotic approximation of the Laplace-type integral holds:

$$\int g(\boldsymbol{\theta}) \exp[\beta^2 f(\boldsymbol{\theta})]\mathrm{d}\boldsymbol{\theta} \sim (2\pi)^{N_\theta} \frac{g(\hat{\boldsymbol{\theta}}) \exp[\beta^2 f(\hat{\boldsymbol{\theta}})]}{\sqrt{\det \mathbf{h}(\hat{\boldsymbol{\theta}})}} \quad \text{as } \beta \to \infty, \tag{A.1}$$

where $\mathbf{h}(\boldsymbol{\theta})$ is the Hessian of the function $\beta^2 f(\boldsymbol{\theta})$.

The asymptotic approximation (8) is shown first. For this, the integral in Eq. (5) is rewritten in the form

$$\frac{1}{c(D)} = \int \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \exp[-(NN_0 - 1)/2 \ln J(\boldsymbol{\theta}; D)]\mathrm{d}\boldsymbol{\theta} \tag{A.2}$$

which is of the type of Eq. (A.1) with $g(\boldsymbol{\theta}) = \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, $f(\boldsymbol{\theta}) = -\ln J(\boldsymbol{\theta}; D)$ and $\beta^2 = (NN_0 - 1)/2$. Therefore, as $N \to \infty$, i.e., for large number of data, the asymptotic approximation (A.1) can be applied for integral (A.2) to yield

$$\frac{1}{c(D)} \sim (2\pi)^{N_\theta/2} \frac{\pi_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \exp[-(NN_0 - 1)/2 \ln J(\hat{\boldsymbol{\theta}}; D)]}{\sqrt{\det \mathbf{h}(\hat{\boldsymbol{\theta}}, D)}}, \tag{A.3}$$

where $\mathbf{h}(\boldsymbol{\theta}, D)$ is the Hessian of the function $\beta^2 f(\boldsymbol{\theta}) = -(NN_0 - 1)/2 \ln J(\hat{\boldsymbol{\theta}}; D)$ with the (i, l) element given by

$$h_{il}(\boldsymbol{\theta}, D) = -(NN_0 - 1)/2 \frac{\partial^2[\ln J(\boldsymbol{\theta}; D)]}{\partial\theta_i \partial\theta_l}. \tag{A.4}$$

Substituting the form of $J(\boldsymbol{\theta}; D)$ from Eq. (3), carrying out the second order differentiation, simplifying, using the relation $\mathbf{L}_0^{\mathrm{T}}\mathbf{L}_0 = \mathrm{diag}(\boldsymbol{\delta})$, and evaluating the resulting expression at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, one readily obtains that

$$h_{il}(\hat{\boldsymbol{\theta}}, D) = \frac{(NN_0 - 1)}{J(\hat{\boldsymbol{\theta}}; D)} \left[-\frac{2}{J(\hat{\boldsymbol{\theta}}; D)} \frac{\partial J(\hat{\boldsymbol{\theta}}; D)}{\partial \theta_l} A_N + B_N + \frac{1}{NN_0} \sum_{k=1}^{N} \sum_{j=1}^{N_d} \delta_j \frac{\partial x_j(k; \hat{\boldsymbol{\theta}})}{\partial \theta_i} \frac{\partial x_j(k; \hat{\boldsymbol{\theta}})}{\partial \theta_l} \right], \tag{A.5}$$

where

$$A_N = \frac{1}{NN_0} \sum_{k=1}^{N} \sum_{j=1}^{N_d} \delta_j n_j(k; \hat{\boldsymbol{\theta}}) \frac{\partial x_j(k; \hat{\boldsymbol{\theta}})}{\partial \theta_i}, \quad B_N = \frac{1}{NN_0} \sum_{k=1}^{N} \sum_{j=1}^{N_d} \delta_j n_j(k; \hat{\boldsymbol{\theta}}) \frac{\partial^2 x_j(k; \hat{\boldsymbol{\theta}})}{\partial \theta_i \partial \theta_l}. \tag{A.6}$$

The values of the quantities $n_j(k, \hat{\boldsymbol{\theta}})$, defined in Eq. (1), are samples of independent, identically distributed Gaussian random variables. Thus, using the law of large numbers, as $N \to \infty$ the coefficients $A_N \to 0$ and $B_N \to 0$. Therefore, neglecting their contribution in Eq. (A.5) and using the fact that $\hat{\sigma}^2 = J(\hat{\boldsymbol{\theta}}; D)$, the elements $h_{il}(\hat{\boldsymbol{\theta}}, D)$ of the Hessian matrix simplify to

$$h_{il}(\hat{\boldsymbol{\theta}}, D) \sim \frac{1}{\hat{\sigma}^2} Q_{il}(\hat{\boldsymbol{\theta}}, \boldsymbol{\delta}) = \frac{1}{\hat{\sigma}^2} \sum_{k=1}^{N} \sum_{j=1}^{N_d} \delta_j \frac{\partial x_j(k; \hat{\boldsymbol{\theta}})}{\partial \theta_i} \frac{\partial x_j(k; \hat{\boldsymbol{\theta}})}{\partial \theta_l}. \tag{A.7}$$

Finally, the matrix $\mathbf{Q}(\boldsymbol{\theta}, \boldsymbol{\delta})$ formed by the elements $Q_{il}(\boldsymbol{\theta}, \boldsymbol{\delta})$ takes the compact form given in Eqs. (10)–(12). Also, substituting Eq. (A.7) into Eq. (A.3) and using the fact that $\hat{\sigma}^2 = J(\hat{\boldsymbol{\theta}}; D)$, the asymptotic estimate (A.3) takes the final form given in Eq. (8).

The asymptotic approximations (9) are shown next. Using Eq. (4), the integral in $\mathrm{E}_{\boldsymbol{\theta}}[\ln J(\boldsymbol{\theta}; D)]$ can be written in the form

$$\mathrm{E}_{\boldsymbol{\theta}}[\ln J(\boldsymbol{\theta}; D)] = c \int \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \ln[J(\boldsymbol{\theta}; D)] \exp[-(NN_0 - 1)/2 \ln J(\boldsymbol{\theta}; D)] \mathrm{d}\boldsymbol{\theta}, \tag{A.8}$$

Applying the asymptotic result (A.1) with $g(\boldsymbol{\theta}) = \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \ln J(\boldsymbol{\theta}; D)$, $f(\boldsymbol{\theta}) = -\ln J(\boldsymbol{\theta}; D)$ and $\beta^2 = (NN_0 - 1)/2 \to \infty$ as $N \to \infty$, the integral in Eq. (A.8) is asymptotically approximated by

$$\mathrm{E}_{\boldsymbol{\theta}}[\ln J(\boldsymbol{\theta}; D)] \sim c(2\pi)^{N_\theta/2} \pi_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \ln J(\hat{\boldsymbol{\theta}}; D) \frac{\exp[-(NN_0 - 1)/2 \ln J(\hat{\boldsymbol{\theta}}; D)]}{\sqrt{\det \mathbf{h}(\hat{\boldsymbol{\theta}}, D)}}. \tag{A.9}$$

Substituting from Eq. (A.3) the asymptotic result derived for $1/c(D)$, the expression in Eq. (A.9) simplifies to

$$\mathrm{E}_{\boldsymbol{\theta}}[\ln J(\boldsymbol{\theta}; D)] \sim \ln[J(\hat{\boldsymbol{\theta}}; D)] \tag{A.10}$$

which, using the relation $\hat{\sigma}^2 = J(\hat{\boldsymbol{\theta}}; D)$, is exactly the same as the one given in Eq. (9).

Following a similar analysis, the second asymptotic approximation in Eq. (9) for the term $\mathrm{E}_{\boldsymbol{\theta}}[\ln \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})]$ is readily obtained.

References

[1] P. Shah, F.E. Udwadia, A methodology for optimal sensor locations for identification of dynamic systems, *Journal of Applied Mechanics* 45 (1978) 188–196.

[2] D.C. Kammer, Sensor placements for on orbit modal identification and correlation of large space structures, *Journal of Guidance, Control and Dynamics* 14 (1991) 251–259.

[3] F.E. Udwadia, Methodology for optimal sensor locations for parameter identification in dynamic systems, *American Society of Civil Engineers, Journal of Engineering Mechanics* 120 (2) (1994) 368–390.

[4] P.H. Kirkegaard, R. Brincker, On the optimal locations of sensors for parametric identification of linear structural systems, *Mechanical Systems and Signal Processing* 8 (1994) 639–647.

[5] J.E.T. Penny, M.I. Friswell, S.D. Garvey, Automatic choice of measurement location for dynamic testing, *American Institute of Aeronautics and Astronautics Journal* 32 (1994) 407–414.

[6] F.M. Hemez, C. Farhat, An energy based optimum sensor placement criterion and its application to structural damage detection, in: *12th International Modal Analysis Conference (IMAC)*, Society of Experimental Mechanics, Honolulu, 1994, pp. 1568–1575.

[7] R.G. Cobb, B.S. Liebst, Sensor location prioritisation and structural damage localization using minimal sensor information, *American Institute of Aeronautics and Astronautics Journal* 35 (2) (1996) 369–374.

[8] M. Reynier, H. Abou-Kandil, Sensors location for updating problems, *Mechanical Systems and Signal Processing* 13 (2) (1999) 297–314.

[9] Z.Y. Shi, S.S. Law, L.M. Zhang, Optimum sensor placement for structural damage detection, *American Society of Civil Engineers, Journal of Engineering Mechanics* 126 (11) (2000) 1173–1179.

[10] Z. Kotulski, Optimisation of sensors' location in a stochastic extrapolation problem, *Journal of Sound and Vibration* 138 (3) (1990) 351–363.

[11] K. Sobczyk, Theoretic information approach to identification and signal processing, in: *Proceedings of the IFIP Conference on Reliability and Optimisation of Structural Systems*, Aalborg, Denmark, 1987.

[12] E. Heredia-Zavoni, R. Montes-Iturrizaga, L. Esteva, Optimal instrumentation of structures on flexible base for system identification, *Earthquake Engineering and Structural Dynamics* 28 (12) (1999) 1471–1482.

[13] E. Heredia-Zavoni, L. Esteva, Optimal instrumentation of uncertain structural systems subject to earthquake motions, *Earthquake Engineering and Structural Dynamics* 27 (4) (1998) 343–362.

[14] C. Papadimitriou, J.L. Beck, S.K. Au, Entropy-based optimal sensor location for structural model updating, *Journal of Vibration and Control* 6 (5) (2000) 781–800.

[15] E.T. Jaynes, Where do we stand on maximum entropy? in: R.D Levine, M. Tribus (Eds.), *The Maximum Entropy Formalism*, MIT Press, Cambridge, MA, 1978.

[16] K.V. Yuen, L.S. Katafygiotis, C. Papadimitriou, N.C. Mickleborough, Optimal sensor placement methodology for identification with unmeasured excitation, *Journal of Dynamic Systems, Measurement and Control* 123 (4) (2001) 677–686.

[17] P. Metallidis, G. Verros, S. Natsiavas, C. Papadimitriou, Identification, fault detection and optimal sensor location in vehicle suspensions, *Journal of Vibration and Control* 9 (3–4) (2003) 337–359.

[18] H. Bedrossian, S.F. Masri, Optimal placement of sensors and shakers for modal identification, in: P.D. Spanos, G. Deodatis (Eds.), *Computational Stochastic Mechanics*, Millpress, Rotterdam, 2003, pp. 53–57.

[19] L. Yao, W.A. Sethares, D.C. Kammer, Sensor placement for on orbit modal identification via a genetic algorithm, *American Institute of Aeronautics and Astronautics Journal* 31 (1993) 1167–1169.

[20] K. Worden, A.P. Burrows, Optimal sensor placement for fault detection, *Engineering Structures* 23 (2001) 885–901.

[21] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties—Bayesian statistical framework, *American Society of Civil Engineers, Journal of Engineering Mechanics* 124 (4) (1998) 455–461.

[22] L.S. Katafygiotis, C. Papadimitriou, H.F. Lam, A probabilistic approach to structural model updating, *International Journal of Soil Dynamics and Earthquake Engineering* 17 (7–8) (1998) 495–507.

[23] N. Bleistein, R. Handelsman, *Asymptotic Expansions for Integrals*, Dover Publications, New York, 1986.

[24] D.E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison-Wesley, Reading, MA, 1989.

[25] Z. Michalewich, *Genetic Algorithms+Data Structures=Evolution Programs*, Springer, New York, 1999.

[26] F.E. Udwadia, D.K. Sharma, Some uniqueness results related to building structural identification, *SIAM Journal of Applied Mathematics* 34 (1) (1978) 104–151.

[27] A. Berman, Non-unique structural system identification, in: *Proceedings of the Seventh International Modal Analysis Conference*, Bethel, CT, 1989, pp. 355–359.

[28] L.S. Katafygiotis, J.L. Beck, Updating models and their uncertainties—model identifiability, *American Society of Civil Engineers, Journal of Engineering Mechanics* 124 (4) (1998) 463–467.

[29] L.S. Katafygiotis, H.F. Lam, C. Papadimitriou, Treatment of unidentifiability in structural model updating, *Advances in Structural Engineering* 3 (1) (2000) 19–39.