A Bayesian framework for optimal experimental design in structural dynamics

Costas Argyris[1], Costas Papadimitriou[*1]

[1]Department of Mechanical Engineering, University of Thessaly, Volos, Greece

ABSTRACT

A Bayesian framework for optimal experimental design in structural dynamics is presented. The optimal design is based on an expected utility function that measures the value of the information arising from alternative experimental designs and takes into account the uncertainties in model parameters and model prediction error. The evaluation of the expected utility function requires a large number of structural model simulations. Asymptotic techniques are used to simplify the expected utility functions under small model prediction error uncertainties, providing insight into the optimal design and drastically reducing the computation effort involved in the evaluation of the multi-dimensional integrals that arise. The framework is demonstrated using the design of sensors for modal identification and is applied to the design of a small number of reference sensors for experiments involving multiple sensor configuration setups accomplished with reference and moving sensors. In contrast to previous formulations, the Bayesian optimal experimental design overcomes the problem of the ill-conditioned Fisher information matrix for small number of reference sensors by exploiting the information in the prior distribution.

KEYWORDS: Bayesian inference, information entropy, relative entropy, Kullback-Leibler divergence, structural dynamics.

1. INTRODUCTION

Bayesian optimal experimental design formulations for model parameter estimation have been developed that are based on utility functions that measure the value of the information contained in the data. The optimal design is then obtained by maximizing the expected utility function over the model and prediction error uncertainties with respect to the design variables. A rational choice of the utility function is the information gain or relative entropy or the Kullback-Leibler divergence between the posterior and the prior distribution of model parameters. The estimation of expected information gain involves a nested sampling procedure [1]. The optimization of the utility function involves a moderate to large number of expected utility function evaluations. Thus, the computational effort becomes excessive even for relatively small models. Surrogate techniques have been introduced to reduce the computational effort [1].

In this work, we use asymptotic expansions and small prediction error uncertainties to simplify the formulation and the number of system analyses. For linear models the expected utility function depends on the information entropy of the posterior distribution of the model parameters which can be formulated as a function of the determinant of the sum of the Fisher information matrix and the Hessian of the prior. For nonlinear models and small prediction error uncertainties it is shown that the expected utility function can also be formulated as the robust information entropy which is a weighted integral of the information entropy of the posterior distribution of the model parameters evaluated at control points in the support of the prior PDF of the model parameters, weighted by the prior distribution of these parameters. This result has been independently derived in [2-4] based on information entropy concepts. The results show that the experimental design is based on a scalar measure of the sensitivities of the output quantities of interest predicted by the model at the corresponding measured locations with respect to the model parameters.

The theoretical developments and the effectiveness of the proposed algorithms are applicable to optimal sensor placement problems encountered in structural dynamics, including the optimal sensor placement for finite element model updating,

[*] Corresponding author: costasp@uth.gr

damage detection, as well as model predictions of output quantities. In addition, the proposed optimal experimental design framework can also be applied for optimizing the excitation characteristics, such as frequency content and amplitude, of linear and nonlinear structures in order to enhance the value of information to be obtained from the experimental designs for model parameter estimation and predictions. In this work, in light of the new results, the optimal sensor placement problem for estimating the modes in structural dynamics application is revisited. It is demonstrated that the information entropy and thus the optimal sensor placement depends on the sum of the Fisher information matrix and the Hessian of the prior, instead of only the Fisher information entropy proposed in previous formulations [4,5]. As a result, it is demonstrated that the optimal sensor placement problem can be carried out also for number of sensors less than the number of modes with the aid of the information contained in the prior PDF of the model parameters. The optimal sensor placement design is demonstrated by designing the optimal sensor location of a small number of reference sensors in multiple sensor placement setups using reference and moving (roving) sensors.

2. BAYESIAN PARAMETER ESTIMATION

The Bayesian framework for the estimation of the parameters of a structural dynamics model based on experimental data is first outlined and the results are used in the optimal experimental design formulation. Consider a model of a structure and let $\underline{\theta} \in R^{N_\theta}$ be the vector of model parameters to be estimated using a set of measured data $\underline{y} \equiv \underline{y}(\underline{\delta}) \in R^N$ of output quantities that depend on design variables $\underline{\delta}$. Let $\underline{g}(\underline{\theta};\underline{\delta}) \in R^N$ be the vector of the values of the same output quantities predicted by a structural model for specific values of the parameter set $\underline{\theta}$. The design variables can be related to the location of sensors placed in a structure or to the characteristics of the excitation. For design variables related to sensor locations, the location vector $\underline{\delta}$ contains the coordinates of the sensors with respect to a coordinate system. For design variables related to excitation characteristics, the vector $\underline{\delta}$ contain the values of the parameters of the excitation and the location of the excitation in the structure.

The following prediction error equation is introduced

$$\underline{y} = \underline{g}(\underline{\theta};\underline{\delta}) + \underline{e} \tag{1}$$

where $\underline{e}$ is the additive prediction error term due to model and measurement error. The prediction error is modeled as a Gaussian vector, whose mean value is equal to zero and its covariance is equal to $\Sigma(\underline{\sigma}) \in R^{N\times N}$, where $\underline{\sigma}$ contains the parameters that define the correlation structure of $\Sigma(\underline{\sigma})$. Applying the Bayesian theorem, the posterior probability density function (PDF) of the model parameters $\underline{\theta}$, given the measured data or outcome $\underline{y}$, is given by

$$p(\underline{\theta} \mid \underline{y},\underline{\sigma},\underline{\delta}) = c\frac{1}{(2\pi)^{N/2}\left[\det \Sigma(\underline{\sigma})\right]^{1/2}} \exp\left[-\frac{1}{2} J(\underline{\theta};\underline{y},\underline{\sigma},\underline{\delta})\right] \pi(\underline{\theta}) \tag{2}$$

where

$$J(\underline{\theta};\underline{y},\underline{\sigma},\underline{\delta}) = [\underline{y} - \underline{g}(\underline{\theta};\underline{\delta})]^T \Sigma^{-1}(\underline{\sigma})[\underline{y} - \underline{g}(\underline{\theta};\underline{\delta})] \tag{3}$$

expresses the deviation between the measured and model predicted quantities. The PDF $\pi(\underline{\theta})$ is the prior distribution for $\underline{\theta}$, and c is a normalization constant guaranteeing that the posterior PDF $p(\underline{\theta} \mid \underline{y},\underline{\sigma},\underline{\delta})$ integrates to one.

For sufficiently large number of data, the posterior PDF can be asymptotically approximated by a Gaussian PDF $N(\underline{\theta};\underline{\hat{\theta}},Q^{-1})$ centered at the most probable value $\underline{\hat{\theta}} \equiv \underline{\hat{\theta}}(\underline{y};\underline{\sigma},\underline{\delta})$ that is obtained by minimizing the function $J(\underline{\theta};\underline{y},\underline{\sigma},\underline{\delta})$, i.e.

$$\underline{\hat{\theta}}(\underline{y};\underline{\sigma},\underline{\delta}) = \arg\min_{\underline{\theta}}[J(\underline{\theta};\underline{y},\underline{\sigma},\underline{\delta}) - \ln \pi(\underline{\theta})] \tag{4}$$

and covariance matrix $C = C(\hat{\underline{\theta}}; \underline{y}, \underline{\sigma}, \underline{\delta})$ equal to the inverse of the Hessian of $-\ln p(\underline{\theta} \mid \underline{y}, \underline{\sigma}, \underline{\delta})$, given by

$$C^{-1}(\underline{\theta}; \underline{y}, \underline{\sigma}, \underline{\delta}) = \underline{\nabla}_{\theta} \underline{\nabla}_{\theta}^{T} [-\ln p(\underline{\theta} \mid \underline{y}, \underline{\sigma}, \underline{\delta})] = Q(\underline{\theta}; \underline{y}, \underline{\sigma}, \underline{\delta}) + Q_{\pi}(\underline{\theta}) \tag{5}$$

And evaluated at the most probable value $\hat{\underline{\theta}}(\underline{y}; \underline{\sigma}, \underline{\delta})$, where

$$Q(\underline{y}, \underline{\sigma}, \underline{\delta}) = \frac{1}{2} \underline{\nabla}_{\theta} \underline{\nabla}_{\theta}^{T} J(\underline{\theta}; \underline{y}, \underline{\sigma}, \underline{\delta}) \tag{6}$$

is the Hessian of $J(\underline{\theta}; \underline{y}, \underline{\sigma}, \underline{\delta})$ and $Q_{\pi}(\underline{\theta}) = \underline{\nabla}_{\theta} \underline{\nabla}_{\theta}^{T} [-\ln \pi(\underline{\theta})]$ is the Hessian of the prior PDF. Both the most probable value and the covariance or Hessian depend on the data $\underline{y}$, the values of the prediction error parameter set $\underline{\sigma}$ and the design variables $\underline{\delta}$.

3. OPTIMAL EXPERIMENTAL DESIGN

The objective of an experiment is to maximize the information contain in the data for estimating the parameters of the model. Lindley [6] has introduces the expected utility function $U(\underline{\delta})$ to reflect the usefulness of the experiment in estimating the model parameters. The expected utility function has the form

$$U(\underline{\delta}) = \int_{\Upsilon} \int_{\Theta} u(\underline{\delta}; \underline{\theta}, \underline{y}) \, p(\underline{\theta}, \underline{y} \mid \underline{\delta}) \, d\underline{\theta} \, d\underline{y} \tag{7}$$

where $u(\underline{\delta}; \underline{\theta}, \underline{y})$ is the utility function given a particular value of the model parameter set $\underline{\theta}$ and the outcome $\underline{y}$ from the experiment, $p(\underline{\theta}, \underline{y} \mid \underline{\delta}) = p(\underline{\theta} \mid \underline{y}, \underline{\delta}) \, p(\underline{y} \mid \underline{\delta})$, $p(\underline{\theta} \mid \underline{y}, \underline{\delta})$ is the posterior uncertainty in the model parameters given the outcome $\underline{y}$ and $p(\underline{y} \mid \underline{\delta})$ is the uncertainty in the data. The utility function is chosen to be related to the information gained by the experiment. Based on information theory, it is chosen to be the relative entropy or the Kullback-Leibler divergence between the prior and posterior probability distribution given an outcome $\underline{y}$ obtained from an experimental design $\underline{\delta}$. The expected utility function is an average of the utility function over all possible values of the model parameters as they are inferred from the data, and all the possible outcomes of the experiment. Using the relative entropy in place of the utility function in (7), the expected utility function takes the form

$$\begin{aligned} U(\underline{\delta}) &= \int_{\Upsilon} \int_{\Theta} p(\underline{\theta} \mid \underline{y}, \underline{\delta}) \ln \frac{p(\underline{\theta} \mid \underline{y}, \underline{\delta})}{\pi(\underline{\theta})} \, d\underline{\theta} \, p(\underline{y} \mid \underline{\delta}) \, d\underline{y} \\ &= \int_{\Upsilon} \int_{\Theta} p(\underline{\theta} \mid \underline{y}, \underline{\delta}) \ln p(\underline{\theta} \mid \underline{y}, \underline{\delta}) \, d\underline{\theta} \, p(\underline{y} \mid \underline{\delta}) \, d\underline{y} - \int_{\Theta} \int_{\Upsilon} p(\underline{\theta} \mid \underline{y}, \underline{\delta}) \, p(\underline{y} \mid \underline{\delta}) \, d\underline{y} \ln \pi(\underline{\theta}) \, d\underline{\theta} \end{aligned} \tag{8}$$

where, for prior PDF that it is independent of the design variables, the second term is independent of the design variables and does not affect the optimal values of the design variables. Introducing the information entropy of the posterior PDF of the model parameters

$$H(\underline{y}, \underline{\delta}) = -\int_{\Theta} p(\underline{\theta} \mid \underline{y}, \underline{\delta}) \ln p(\underline{\theta} \mid \underline{y}, \underline{\delta}) \, d\underline{\theta} \tag{9}$$

for a given outcome of the experiment $\underline{y}$ obtained from design $\underline{\delta}$, substituting (9) into (8) and noting that the second integral is a constant, say a, the expected utility function takes the form

$$U(\underline{\delta}) + a = -\int_{\Upsilon} H(\underline{y},\underline{\delta})\, p(\underline{y} \mid \underline{\delta})\, d\underline{y} = -\int_{\Theta}\int_{\Upsilon} H(\underline{y},\underline{\delta})\, p(\underline{y} \mid \underline{\theta},\underline{\delta})\, d\underline{y}\, p(\underline{\theta} \mid \underline{\delta})\, d\underline{\theta} = -\int_{\Theta} \overline{H}(\underline{\theta},\underline{\delta})\, \pi(\underline{\theta})\, d\underline{\theta} \tag{10}$$

where $\overline{H}(\underline{\theta},\underline{\delta})$ is a function of the parameters $\underline{\theta}$ given by

$$\overline{H}(\underline{\theta},\underline{\delta}) = \int_{\Upsilon} H(\underline{y},\underline{\delta})\, p(\underline{y} \mid \underline{\theta},\underline{\delta})\, d\underline{y} \tag{11}$$

which is the average information entropy over all outcomes, conditions on the value $\underline{\theta}$ of the model parameters.

The optimal experimental design is formulated as finding the optimal values $\underline{\delta}_{\max}$ of the design variables $\underline{\delta}$ the maximize the utility function $U(\underline{\delta})$, i.e.

$$\underline{\delta}_{opt} = \arg\max_{\underline{\delta}} U(\underline{\delta}) \tag{12}$$

Optimal experimental design problems involving the design the position of sensors often results in multiple local and local solutions. Also the gradient of the objective function with respect to the design variables in most cases of practical interest cannot be evaluated analytically. To avoid premature convergence to a local optimum and the evaluation of sensitivities of the utility function with respect to the design variables, stochastic optimization algorithms can be used to find the optimum. Herein the CMA-ES algorithm [7] is used for solving the optimization problem, requiring only evaluation of the objective function at different values of the design variables.

Using sampling techniques to approximate the integrals involved in the expected utility function, a nested sampling scheme has been developed [1] which can result in excessive computational effort for each evaluation of the expected utility function since a very large number of model runs are required. To reduce the computational effort a number of surrogate techniques have been proposed [1]. Herein, we provide asymptotic estimates of the utility function that reduce the computational burden and provide insight into the optimal design.

4. OPTIMAL EXPERIMENTAL DESIGN BASED ON ASYMPTOTIC APPROXIMATION

Note that from (1) the PDF $p(\underline{y} \mid \underline{\theta},\underline{\delta})$ is Gaussian with respect to $\underline{y}$ with mean $\underline{g}(\underline{\theta},\underline{\delta})$, the predictions from the model given the value of the parameter $\underline{\theta}$ and covariance Σ. Using Laplace asymptotic approximation [8], the integral in (11) can be asymptotically approximated, for small prediction error variance, by

$$\overline{H}(\underline{\theta},\underline{\delta}) = H(\underline{g}(\underline{\theta},\underline{\delta}),\underline{\delta}) \tag{13}$$

Substituting in (10), the expected utility function is given by

$$U(\underline{\delta}) + a = -\int_{\Theta} H(\underline{g}(\underline{\theta},\underline{\delta}),\underline{\delta})\, \pi(\underline{\theta})\, d\underline{\theta} \tag{14}$$

which is an integral of the information entropy of the posterior PDF of the model parameters corresponding to the outcome $\underline{y} = \underline{g}(\underline{\theta};\underline{\delta})$ conditioned on the values of the model parameters over the support of the prior PDF, weighted by the prior distribution of the model parameters. The integral (9) represents the robust information entropy over all possible values of the model parameters quantified by the prior PDF.
Using Monte Carlo simulations or sparse grid techniques [9,10], the integral (14) can be evaluated in the form

$$U(\underline{\delta}) + a \approx -\sum_{j=1}^{n} w_j \overline{H}(\underline{\theta}^{(j)},\underline{\delta}) \sim -\sum_{j=1}^{n} w_j H(\underline{g}(\underline{\theta}^{(j)},\underline{\delta}),\underline{\delta}) \tag{15}$$

where $\underline{\theta}^{(j)}$, $j=1,\ldots,n$, are either the samples drawn from the prior $\pi(\underline{\theta})$ or the sparse grid points in the parameters space, and w_j are weights equal to $w_j = 1/n$ for the Monte Carlo technique or, for sparse grid techniques, their values depend on the sparse grid order and the prior distribution selected [9].
Using the Bayesian asymptotic approximation for the posterior PDF, valid for large number of data, the information entropy for the outcome $\underline{y} = \underline{g}(\underline{\theta}^{(j)};\underline{\delta}) \equiv \underline{g}_j(\underline{\delta})$ is given by

$$H(\underline{g}(\underline{\theta}^{(j)},\underline{\delta}),\underline{\delta}) \equiv H(\underline{g}_j(\underline{\delta}),\underline{\delta}) = \frac{1}{2}N_\theta \ln(2\pi) - \frac{1}{2}\ln\det[\hat{Q}_j(\underline{\sigma},\underline{\delta}) + Q_\pi(\hat{\underline{\theta}}_j)] \tag{16}$$

where $\hat{Q}_j(\underline{\sigma},\underline{\delta}) \equiv Q(\hat{\underline{\theta}}_j;\underline{g}(\hat{\underline{\theta}}_j,\underline{\delta}),\underline{\sigma},\underline{\delta})$ and $\hat{\underline{\theta}}_j = \hat{\underline{\theta}}(\underline{g}(\underline{\theta}^{(j)},\underline{\delta});\underline{\sigma},\underline{\delta})$.

In the special case of a linear model with respect to the model parameters, i.e. $\underline{g}(\underline{\theta};\underline{\delta}) \equiv \underline{a}_0 + A\underline{\theta}$, and Gaussian prior $\pi(\underline{\theta}) = N(\underline{\theta};\underline{\mu},S)$ with mean $\underline{\mu}$ and covariance S, the information entropy in (16) is independent of $\hat{\underline{\theta}}_j$ and is given by

$$H(\underline{g}(\underline{\theta}^{(j)},\underline{\delta}),\underline{\delta}) \equiv H(\underline{g}_j(\underline{\delta}),\underline{\delta}) = \frac{1}{2}N_\theta \ln(2\pi) - \frac{1}{2}\ln\det[A\Sigma^{-1}A + S^{-1}] \tag{17}$$

For the general case, for each point $\underline{\theta}^{(j)}$ in the parameter space, one can use a first-order approximation of the model prediction $\underline{g}(\underline{\theta},\underline{\delta})$ in the neighborhood of each point $\underline{\theta}^{(j)}$ in (15) to obtain for Gaussian prior that

$$\hat{\underline{\theta}}_j = \underline{\theta}^{(j)} + (A_j^T\Sigma A_j + S^{-1})^{-1}S^{-1}(\underline{\mu} - \underline{\theta}^{(j)}) \tag{18}$$

and

$$H(\underline{g}(\underline{\theta}^{(j)},\underline{\delta}),\underline{\delta}) \equiv H(\underline{g}_j(\underline{\delta}),\underline{\delta}) = \frac{1}{2}N_\theta \ln(2\pi) - \frac{1}{2}\ln\det[A_j^T\Sigma^{-1}A_j + S^{-1}] \tag{19}$$

where $A_j = \underline{\nabla}_\theta \underline{g}(\underline{\theta},\underline{\delta})\big|_{\underline{\theta}=\hat{\underline{\theta}}_j}$. In this case the information entropy evaluated at a point $\underline{\theta}^{(j)}$ depends on the sensitivity of the output quantity of interest with respect to the model parameters, evaluated at the neighbor point $\hat{\underline{\theta}}_j$. The analysis holds if $\hat{\underline{\theta}}_j$ is close to $\underline{\theta}^{(j)}$ so that the first order approximation of $\underline{g}(\underline{\theta},\underline{\delta})$ about $\underline{\theta}^{(j)}$ applies. To study the conditions under which the first order approximation can be used and the results (18) and (19) are valid, let $\Sigma = \varepsilon^2\Sigma_0$ and $S = s^2S_0$, denoting that Σ and S are of order of ε^2 and s^2, respectively. Substituting in (18) one derives that

$$\hat{\underline{\theta}}_j = \underline{\theta}^{(j)} + \frac{\varepsilon^2}{s^2}\left[S_0A_j^T\Sigma_0^{-1}A_j + \frac{\varepsilon^2}{s^2}\right]^{-1}(\underline{\mu} - \underline{\theta}^{(j)})$$

The ratio ε^2/s^2 is a measure of how close the optimum $\hat{\underline{\theta}}_j$ is to the point $\underline{\theta}^{(j)}$ and can be used as a direct measure of the validity of the first-order approximation of $\underline{g}(\underline{\theta},\underline{\delta})$ about $\underline{\theta}^{(j)}$. For $\varepsilon^2/s^2 << 1$ the following approximation holds $\hat{\underline{\theta}}_j \approx \underline{\theta}^{(j)}$. This approximation is true for small prediction error variance in relation to the variance of the prior distribution. Thus, the information entropy evaluated at a point $\underline{\theta}^{(j)}$ depends on the sensitivity of the output QoI with respect to the model parameters, evaluated at $\underline{\theta}^{(j)}$, the sparse grid point or the sample points in an MC estimate. It should be noted that the

estimate (15) with $H(\underline{g}(\underline{\theta}^{(j)}, \underline{\delta}), \underline{\delta})$ given by (19) and $A_j \approx \underline{\nabla}_{\theta} \underline{g}(\underline{\theta}, \underline{\delta})\Big|_{\underline{\theta}=\underline{\theta}^{(j)}}$ is equivalent to the asymptotic estimate given in [3,11] based on the change of information entropy between the prior and posterior distribution, yielding the robust information entropy measure using the prior uncertainties.

5. IMPLEMENTATION IS STRUCTURAL DYNAMICS

The optimal experimental design methodology is implemented in structural dynamics for estimating the optimal sensor locations for modal identification. Specifically, the problem of estimating the modal coordinate vector $\underline{\xi} \in R^m$ ($m \leq N_d$) encountered in modal identification is re-visited. Using modal analysis, the response vector $\underline{u} \equiv \underline{g}(\underline{\theta}) \in R^N$ is given with respect to the parameter set $\underline{\theta}$ in the form $\underline{u} = \Phi \underline{\theta}$, where $\Phi \in R^{N \times m}$ is the mode shape matrix for m contributing modes. The objective is to design the sensor configuration that provides the most information in order to estimate the modal coordinate vector $\underline{\xi}$ which constitute the parameter set $\underline{\theta} \equiv \underline{\xi}$. Noting that $\underline{\nabla}_{\theta} \underline{g}(\underline{\theta}) = \Phi$ and substituting into (19), the expected utility function takes the form

$$U(\underline{\delta}) = \frac{1}{2} \ln \det[\hat{Q}(\underline{\sigma}, \underline{\delta}) + S^{-1}] - a - \frac{1}{2} N_{\theta} \ln(2\pi) \tag{20}$$

where

$$\hat{Q}(\underline{\sigma}, \underline{\delta}) = \{L_{\underline{\delta}} \Phi\}^T \{L_{\underline{\delta}} \Sigma(\underline{\sigma}) L_{\underline{\delta}}^T\}^{-1} \{L_{\underline{\delta}} \Phi\} \tag{21}$$

which is independent of the model parameters. In addition, the optimal sensor locations are independent of the excitation used. The information matrix $\hat{Q}(\underline{\sigma}, \underline{\delta})$ in (21) has exactly the same form as the one proposed in [4] for designing the optimal sensor location. So, with the same assumed prediction error correlation matrix Σ, the optimal sensor location from the proposed methodology and the effective independence are based on exactly the same information matrix Q given in (20).

Based on the form of (20), a non-singular FIM matrix $\hat{Q}(\underline{\sigma}, \underline{\delta})$ is obtained if the number of sensors N is at least equal to the number of contributing modes m ($N_{\theta} = m$). For $N < m$, the matrix $Q(L, \Sigma)$ in (20) is singular and for uniform prior PDF the determinant of the FIM will be zero for any sensor configuration. Thus, for $N_0 < m$ the optimal sensor location problem cannot be performed for uniform prior PDF. This means that the information content in the measured data and the prior is not sufficient to estimate all the parameters simultaneously. The problem is critical for the FSSP algorithm where one starts with no sensors placed on the structure and sequentially adds one sensor at a time on the structure. Previous formulations on the subject based only on the Fisher information matrix [4,5] failed to address this problem. Moreover, forward sequential sensor placement algorithms could not be used due to the singular $\hat{Q}(\underline{\sigma}, \underline{\delta})$. The Bayesian optimal experimental design formulation yields a nonsingular matrix, provided that the prior is non-uniform distribution. The non-uniform prior yields a Hessian that is added to the Fisher information matrix $\hat{Q}(\underline{\sigma}, \underline{\delta})$ and makes the combined matrix $\hat{Q}(\underline{\sigma}, \underline{\delta}) + S^{-1}$ non-singular. Assuming without loss of generality, an isotropic covariance matrix $S = (1/\rho) I$, the utility function is given by

$$U(\underline{\delta}) = \frac{1}{2} \sum_{k=1}^{N_{\theta}} \log(\lambda_k + \rho) \tag{22}$$

where λ_k, $k = 1, \ldots, N_{\theta}$, are the eigenvalues of the Fisher information matrix. Note that when one takes the limit when $\rho \to 0$, the optimal experimental design of equivalent to the one proposed in Papadimitriou and Lombaert [4] where the non-zero eigenvalues were considered in the sum in (22), which was equivalent of maximizing the product of the non-zero eigenvalues in the FIM, instead of maximizing the product of all eigenvalues. This procedure allows to systematically place the sensors optimally in the structure even for the unidentifiable case that arises for a small number of sensors. This procedure also considerably improves the FSSP estimates for $N \geq m$.

6. APPLICATIONS

To demonstrate the use of the optimal sensor placement for modal identification in the case of singular Fisher information matrix, consider the case of selecting the reference sensors in a multiple sensor configuration set up experiment conducted with limited number of sensors in order to obtain the modal frequencies and assemble the modeshapes from multiple setups. It is important in this case that the reference sensor contain the maximum possible information for all modes that are planned to be identified. Wrong locations of the reference sensor may degrade the modal information for one or more modes, degrading the accuracy of the corresponding assembled modeshapes since such accuracy is based on the information contained in the reference sensors. The number of reference sensors is in most cases smaller than the number of modes to be identified. The methodology presented is convenient to optimize the location of sensors in a structure.

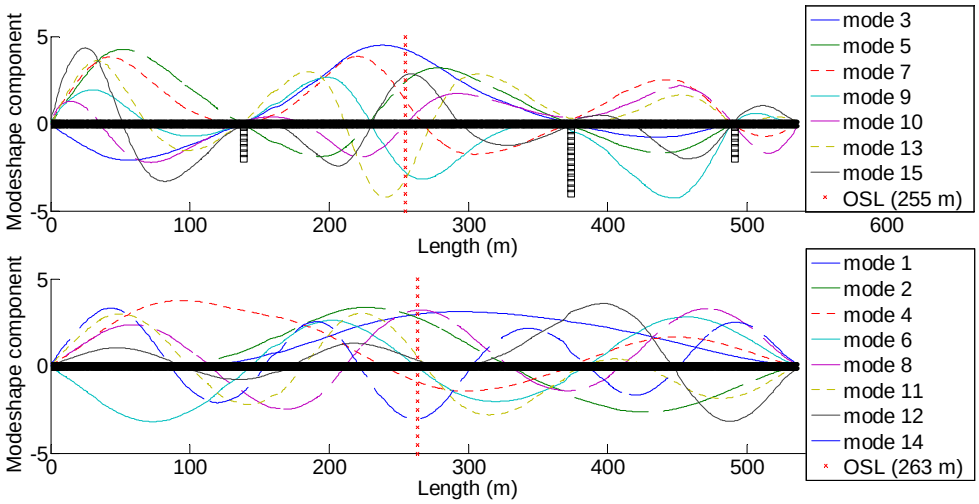

Fig. 1: Optimal location of the reference sensors in multiple sensor configuration setups for identifying the lowest 7 vertical modes (top) and the lowest 8 transverse modes (bottom) of the bridge. The corresponding modes of the bridge are also shown, drawn along the sidewalk curve using the finite element model

The effectiveness is illustrated by designing the optimal location of the reference sensors for the Metsovo bridge. A description of the bridge and the finite element model is given in [12]. The sensors are located along the pedestrian walkway in the two sides of the bridge deck. The problem is formulated separately for the vertical and the transverse modes. One reference sensor is located optimally in the sidewalk so that the measurement contain the maximum possible information for the lowest seven vertical modes of the bridge. The optimal location of a second reference sensor is designed so that the lowest eight transverse modes are identified. The number of design variables for each optimization problem is one, while the number of optimization problems solved are two. In both cases the Fisher information matrix $\hat{Q}(\underline{\sigma}, \underline{\delta})$ in (21) and a number of eigenvalues λ_k in (22) are zero. To successfully proceed with the design, a non-uniform prior distribution has to be used so that the information from the prior makes the matrix $\hat{Q}(\underline{\sigma}, \underline{\delta}) + S^{-1}$ in (21) non-singular. Herein the optimal design is based on a isotropic Gaussian prior with value ρ in (22) selected to be much smaller than the lowest non-zero eigenvalue λ_k. The optimization is performed using CMA-ES [7] over a continuous curve along the two sidewalks of the bridge deck. The results for the optimal sensor placement of the reference sensors are shown in Fig. 1, for the lowest seven vertical and eight transverse modes. It is seen that the vertical reference sensor is optimally located at a distance 255m from the left end of the deck, while the transverse sensor is optimally located at a distance of 263m from the left end.

7. CONCLUSION

Asymptotic approximations, valid for sufficiently large number of data and small prediction error, are introduced to simplify the expected information gain used for optimal experimental design. The utility is based on relative entropy between posterior and prior PDF in Bayesian inference. The asymptotic estimate of the expected utility function depends on the sensitivity of

the output quantity of interest predicted by the model to the model parameter, evaluated at control points in the support of the prior PDF. These control points are the sparse points in a sparse grid technique or the sample points in a MC technique used to estimate the robust information entropy, an integral of the information entropy over the support of the prior PDF of the model parameters. The asymptotic estimate based on the expected information gain is equivalent to the one obtained using the concept of robust information entropy [2,11] for optimal experimental design. Application of the proposed Bayesian optimal experimental design to the optimal sensor placement for modal identification results in optimal design even for small number of sensors in relation to the number of contributing modes, avoiding the problem of singular Fisher information matrix encountered in previous approaches. The non-singularity arises from the information contained in the prior which is taken into account in a Bayesian experimental design approach. The method was successfully used to design the reference sensors for use in multiple configuration setups with reference and roving sensors in order to accurately assemble the modeshapes of a R/C bridge.

ACKNOWLEDEGMENT

This research has been implemented under the “ARISTEIA” Action of the ”Operational Programme Education and Lifelong Learning” and was co-funded by the European Social Fund (ESF) and Greek National Resources.

REFERENCES

[1] X. Huan, Y. M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, Journal of Computational Physics, 232(1):288-317, 2013.

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

[3] C. Papadimitriou, Optimal sensor placement methodology for parametric identification of structural systems, Journal of Sound and Vibration, 278(4):923-947, 2004.

[4] C. Papadimitriou, G. Lombaert, The effect of prediction error correlation on optimal sensor placement in structural dynamics, Mechanical Systems and Signal Processing, 28:105-127, 2012.

[5] D.C. Kammer, Sensor placements for on orbit modal identification and correlation of large space structures, Journal of Guidance, Control and Dynamics, 14:251-259, 1991.

[6] D.V. Lindley, On a measure of the information provided by an experiment, The Annals of Mathematical Statistics 27 (1956) 986–1005.

[7] N. Hansen, S.D. Muller, P. Koumoutsakos, Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evolutionary Computation, 11(1):1-18, 2003

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

[9] T. Gerstner, M. Griebel, Numerical integration using sparse grids, Numerical Algorithms, 18:209-232, 1998.

[10] H.J. Bungartz, M. Griebel, Sparse grids, Acta numerica, 13:147-269, 2004.

[11] D.I. Papadimitriou, C. Papadimitriou, Optimal sensor placement for the estimation of turbulence model parameters in CFD, International Journal for Uncertainty Quantification, 5:1, 2015.

[12] C. Papadimitriou, D.C. Papadioti, D.C., Component mode synthesis techniques for finite element model updating, Computers and Structures, 126:15-28, 2013