Optimal Experimental Design for Estimation of Model Parameters

1. Introduction

The purpose in designing an experiment for a system is to optimize certain characteristics of the experimental set up so that measured data provide useful information about the condition of the system. In building mathematical models of a system, a set of measurements should provide the most useful information for selecting a model of a structure from competitive models, as well as identifying with the least uncertainty the parameters involved in the mathematical representation of the model.

Given a system and a parameterized mathematical model representing the behavior of the system, one objective in optimal experimental design is to select the control parameters of the experiments so that the collected measurements are most informative for estimating the parameters of a mathematical model of the system. The control parameters of an experiment may include the excitation characteristics (e.g. frequency content, amplitude, sampling frequency, duration in structural dynamics), type and location of excitation (e.g. impulse, broadband stochastic, harmonic in structural dynamics), type of output sensors (e.g. displacement, acceleration, force), location and number of output sensors, characteristics of output measurements (e.g. monitoring period, sampling frequency).

In this chapter we are particularly interested in answering the following question. Given a system and its parameterized mathematical model, select the optimal sensor configuration (number, type and location of sensors) such that the measured data obtained from the sensor system are most informative for estimating the parameters of a mathematical model of the system. In model parameter estimation, most informative measurements mean the ones that result in the least uncertainty in the estimates of the parameters. Using information entropy as a unique scalar measure of the uncertainty in a parameter set θ, the problem of optimizing the number, type and location of sensors is stated mathematically as a problem of minimizing the information entropy of the posterior distribution of the model parameters given the data, derived using Bayesian inference.

2. Bayesian Optimal Experimental Design Framework

2.1. Bayesian Parameter Estimation

The Bayesian framework for the estimation of the parameters of a model based on experimental data is first outlined and the results are used for the derivation of the optimal sensor locations. Consider a model and let $\underline{\theta} \in R^{N_\theta}$ be the vector of model parameters to be estimated using a set of measured data $\underline{d} \equiv \underline{d}(\underline{\delta}) \in R^N$ of output quantities at locations $\underline{\delta}$. The location vector $\underline{\delta}$ contains the coordinates of the sensors with respect to a coordinate system. Let $\underline{g}(\underline{\theta};\underline{\delta}) \in R^N$ be the vector of the values of the same output quantities predicted by a model for specific values of the parameter set $\underline{\theta}$. The following prediction error equation is introduced

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

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 Σ. Applying the Bayesian theorem, the posterior probability density function (PDF) of $\underline{\theta}$, given the measured data $\underline{d}$, is given by

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

where

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

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 $p(\underline{\theta}|\underline{\sigma},\underline{d},\underline{\delta})$ integrates to one.

2.2. Information Entropy - Asymptotic Approximation

The PDF $p(\underline{\theta}|\underline{\sigma}, \underline{d}, \underline{\delta})$, given by Eq. (2) quantifies the posterior uncertainty in the parameter values $\underline{\theta}$ based on the information contained in the measured data. The information entropy given by the expression [1]

$$h_{\underline{\theta}}(\underline{\delta}; \underline{\sigma}, \underline{d}) = E_{\underline{\theta}}[-\ln p(\underline{\theta}|\underline{\sigma}, \underline{d}, \underline{\delta})] \quad (4a)$$

$$= -\int \ln p(\underline{\theta}|\underline{\sigma}, \underline{d}, \underline{\delta}) \; p(\underline{\theta}|\underline{\sigma}, \underline{d}, \underline{\delta}) \; d\underline{\theta} \quad (4b)$$

is a scalar measure of the uncertainty of the model parameters $\underline{\theta}$. It depends on the location vector $\underline{\delta}$ of the sensors, the correlation structure of the prediction error and the details in the data $\underline{d}$.

The multi-dimensional integral in Eq. (4) is a Laplace-type integral that can be asymptotically approximated [2], for large number of data, by the expression [3]

$$h_{\underline{\theta}}(\underline{\delta}; \underline{\sigma}, \underline{d}) \sim H(\underline{\delta}; \underline{\theta}_0, \underline{\sigma}) = \frac{1}{2} N_\theta \ln(2\pi) - \frac{1}{2} ln \det \left[Q(\underline{\delta}; \underline{\theta}_0, \underline{\sigma}) + Q_\pi(\underline{\theta}_0)\right] \quad (5)$$

where $\underline{\theta}_0$ are the values of $\underline{\theta}$ that minimize $J(\underline{\theta}; \underline{\sigma}, \underline{d}, \underline{\delta})$, $Q(\underline{\delta}; \underline{\theta}, \underline{\sigma})$ is the Fisher information matrix, a semi-positive definite matrix asymptotically given by

$$Q(\underline{\delta}; \underline{\theta}, \underline{\sigma}) = \nabla_{\underline{\theta}} \underline{g}(\underline{\theta}; \underline{\delta})^T \Sigma^{-1}(\underline{\sigma}) \nabla_{\underline{\theta}} \underline{g}(\underline{\theta}; \underline{\delta}) \quad (6)$$

computed at the N locations where the sensors are placed, and $Q_\pi(\underline{\theta}_0) = -\nabla_{\underline{\theta}}^T \nabla_{\underline{\theta}} ln\pi(\underline{\theta})$ evaluated at the value $\underline{\theta}_0$, with $\nabla_{\underline{\theta}} = [\partial/\partial\theta_1, ..., \partial/\partial\theta_{N_\theta}]$, represents the negative of the Hessian of the natural logarithm of the prior distribution of the model parameters.

For uniform prior distribution the term $Q_\pi(\underline{\theta}_0) = 0$ and the optimal sensor placement is based only on the Fisher information matrix. However, for small number of sensors, the matrix $Q(\underline{\delta}; \underline{\theta}, \underline{\sigma})$ can be ill-conditioned and the determinant could tend to zero independent of the location of sensor. Such cases arise from unidentifiability issues due to the insufficient information provided by the data to estimate the number of model parameters involved. Non-informative uniform prior distribution do not provide any information to correct this problem. In has been proposed in [4] to use non-uniform distributions to remove the ill-conditioning in $Q(\underline{\delta}; \underline{\theta}, \underline{\sigma})$ due to the extra information provided by the prior distribution about the uncertainty in the model parameters. For the specific case of a Gaussian prior distribution, the matrix $Q_\pi(\underline{\theta}_0) = \tilde{Q}$, where $\tilde{Q}$ is the inverse of the covariance matrix of the Gaussian distribution and thus it is constant independent of $\underline{\theta}$.

The asymptotic estimate (5) is very useful since it does not explicitly depend on the details of the data $\underline{d}(\underline{\delta})$ which are not available during the experimental design phase of sensor placement. The dependence on the data comes implicitly through the optimal value $\underline{\theta}_0$ of the parameter set $\underline{\theta}$. However, since the data are not available, the optimal value $\underline{\theta}_0$ cannot be estimated. Thus, optimal sensor placement designs are based on assuming a nominal value of the parameter set $\underline{\theta}_0$. Alternatively, the uncertainty in the nominal value can be accounted for as described in Section 2.3.

In addition, from Eq. (6), it can be deduced that the information entropy depends on the derivatives of the output quantities predicted by the model at the sensor locations with respect to the model parameters. The higher the derivatives, the higher the information entropy value. The computation of these derivatives is based on the differentiation of the model with respect to the parameters.

2.3. Robust Information Entropy Formulation

The previous formulation is based on nominal values $\underline{\theta}_0$ assigned to the optimal value of the model parameter set $\underline{\theta}$ and the nominal values $\underline{\sigma}_0$ assigned to the prediction error parameters $\underline{\sigma}$ involved in the covariance $\Sigma(\underline{\sigma})$ of the model prediction error. A robust formulation is next presented which takes into account the uncertainties in the augmented parameter set $\underline{\varphi}_0 = (\underline{\theta}_0, \underline{\sigma}_0)$.

Using Bayes framework, the uncertainty in the nominal values $\underline{\varphi}_0$ of the model and prediction error parameters is quantified by the prior distribution $\pi(\underline{\varphi}_0)$. Papadimitriou et al. [1] have introduced the change of uncertainty or the change of information entropy from the prior to posterior distribution of the model parameters as a measure of the quality of a sensor configuration. The change of information entropy is given as

$$\begin{aligned} \Delta h(\underline{\delta}) &= E_{\underline{\theta},\underline{\varphi}_0}[-lnp(\underline{\theta},\underline{\varphi}_0|\underline{\delta})] - E_{\underline{\varphi}_0}[-lnp(\underline{\varphi}_0)] && \text{(7a)} \\ &= \int H(\underline{\delta};\underline{\varphi}_0)\ \pi(\underline{\varphi}_0)\ d\underline{\varphi}_0 && \text{(7b)} \\ &= \frac{1}{2}N_\theta \ln(2\pi) - \frac{1}{2}\int ln \det\left[Q(\underline{\delta};\underline{\varphi}_0) + Q_\pi(\underline{\varphi}_0)\right] \pi(\underline{\varphi}_0)\ d\underline{\varphi}_0 && \text{(7c)} \end{aligned}$$

which, using (7b), is an integral of the information entropy conditioned on the nominal values of the model parameters, weighted by the prior distribution of the model parameters. The integral (7b) represents the robust information entropy and arises by substituting $p(\underline{\theta}, \underline{\varphi}_0|\underline{\delta}) = p(\underline{\theta}|\underline{\varphi}_0, \underline{\delta})\ \pi(\underline{\varphi}_0)$ into the definition of the information entropy given in (4b) and simplifying. The equality in (7c) arises after substituting the information entropy given by (5) and simplifying. Details are given in [1].

The measure (7c) can be extended to include uncertainties in model-related quantities, that are not included in the parameters to be inferred. It is straightforward to show that the change of information entropy is given by (7c), provided that the vector $\underline{\varphi}_0$ is augmented to also include these model-related quantities. Also, it should be noted that the result (5) is a special case of (7c) for the case where $\underline{\varphi}_0$ is known deterministic quantity.

Using Monte Carlo simulations or sparse grid techniques [5, 6], the integral in (7c) can be approximated by

$$\int ln \det \left[Q(\underline{\delta}; \underline{\varphi}_0) + Q_\pi(\underline{\varphi}_0)\right] \pi(\underline{\varphi}_0)\ d\underline{\varphi}_0 \simeq \sum_{j=1}^{n} w_j ln \det \left[Q(\underline{\delta}; \underline{\varphi}_0^{(j)}) + Q_\pi(\underline{\varphi}_0^{(j)})\right] \tag{8}$$

where $\underline{\varphi}_0^{(j)}$, $j = 1, \ldots, n$, are either the samples drawn from the prior $\pi(\underline{\varphi}_0)$ 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 their values depend on the sparse grid order and the prior distribution selected [5].

2.4. Optimal Sensor Location Methodology

The sensor configuration should be designed in such a way that the measured data are as much informative as possible about the model parameters to be estimated. The information entropy, defined by Eq. (5), measuring the uncertainty in the these 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 the ones that minimize the information entropy or the change of information entropy in (7c). Thus, the sensors should be located at the places that minimize the information entropy. It should be noted that expression (8) requires the sensitivities of the output quantities to be computed at all sample points $\varphi_0^{(j)}$, $j = 1, \ldots, n$.

The problem of finding the optimal sensor configuration is formulated as an optimization problem where the objective function is the information entropy or the change of information entropy in the robust case and the design variables are the locations of sensors. Specifically, the optimal sensor location $\underline{\delta}_{best}$ is given by

$$\underline{\delta}_{best} = \arg\min_{\underline{\delta}} \Delta h(\underline{\delta}) \tag{9}$$

The minimization of (9) is equivalent to the maximization of det $[Q(\underline{\delta};\underline{\theta}_0,\underline{\sigma})+Q_\pi(\underline{\theta}_0)$ which involves the Fisher information matrix (FIM). Given that the FIM $Q(\underline{\delta};\underline{\theta}_0,\underline{\sigma})$ defined in (6) depends on the sensitivity of output quantities, computed from the model at the measured locations, with respect to the parameters, the sensors tend to be placed at locations where the output quantities are most sensitive to parameter changes. This is consistent with intuition since sensors places at locations where the quantities are insensitive to parameter changes do not provide information to estimate the values of the parameters.

A stochastic or deterministic optimization algorithm may be used to find the location $\underline{\delta}$ of the sensors that minimizes $H(\underline{\delta};\underline{\theta}_0,\underline{\sigma})$ or $\Delta h(\underline{\delta})$. The deterministic method is usually based on the sensitivity derivatives $\frac{\partial H(\underline{\delta};\underline{\theta}_0,\underline{\sigma})}{\partial \delta_i}$ of the information entropy with respect to the coordinates of the sensor locations and a descent algorithm is used to locate the optimum. Although such an algorithm is quite efficient and converges very fast to the optimal solution, there is a drawback due to the several local optimal encountered in the optimal sensor location problem. Another drawback of the deterministic methods is their complexity, since the methodology to compute the aforementioned gradient components should be formulated.

Alternatively, a stochastic optimization algorithm is employed in order to avoid the entrapment to local minima. Specifically, the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [8] is used. To carry out the optimization in the continuous space of the design parameter $\underline{\delta}$, an interpolation scheme is used to compute the required values of the sensitivities involved in Eqs. (5) and (7c) at locations between the grid nodes of a mesh used to discretize the spatial domain.

2.5. Prediction Error Correlation Model

In order to find the optimal location of the sensors, based on the information entropy framework, the structure of the covariance matrix $\Sigma(\underline{\sigma})$ of the prediction error correlation model should be postulated. The prediction error in (1) is due to measurement and model error. Assuming independence between the two errors, the covariance matrix takes the form $\Sigma = \bar{\Sigma} + \tilde{\Sigma}$, where $\bar{\Sigma}$ and $\tilde{\Sigma}$ are the covariance matrices of the measurement and model errors, respectively. Assuming that the measurement error is independent of the location of sensors, the covariance matrix $\bar{\Sigma}$ takes the form $\bar{\Sigma} = \bar{\sigma}^2 I$, where I is the identity matrix.

A certain degree of spatial correlation is expected for the model error since model predictions at two neighborhood points in the physical space are usually correlated due to the structure of the model. It was demonstrated in [7] that when the optimal sensor placement algorithm is applied on the continuous space for uncorrelated model errors, it tends to place sensors very close to each other, providing almost exactly the same information. This clustering of sensors is due to the wrong assumption of uncorrelated prediction errors and can be avoided when the correlation structure of the model predictions are taken into account. It has been theoretically shown [7] that two or more sensors, within an area in the spatial domain of the size of the correlation length, tend to be placed further apart in order to increase the information provided by the sensors. However, how far these sensors will be placed from each other is also controlled by the gradients of the output QoI with respect to the model parameters $\underline{\theta}$ involved in the definition of the information matrix. Very high derivatives tend to limit to the size of the area affected by the correlation length. Drastic changes in the sensitivities of the QoI to parameter changes that occur between two closely-spaced sensors may justify clustering of such sensors with qualitatively distinct information [3].

A certain degree of correlation should be accounted for the model errors between any two locations depending on the physics of te problem analysed. The correlation is postulated by selecting the correlation $\tilde{\Sigma}_{kl}$ between two sensor locations $\underline{x}_k$ and $\underline{x}_l$ in the physical space as

$$\tilde{\Sigma}_{kl} = \sqrt{\tilde{\Sigma}_{kk}\tilde{\Sigma}_{ll}} R(\eta_{kl}) \tag{10}$$

where $R(\eta_{kl})$ is the spatial correlation structure which is assumed herein to depend on the distance $\eta_{kl} = |\underline{x}_k - \underline{x}_l|$ between the measurements k and l at location $\underline{x}_k$ and $\underline{x}_l$, respectively. The variance $\tilde{\Sigma}_{kk}$ of the prediction error at measured location $\underline{x}_k$ can be taken to be $\tilde{\Sigma}_{kk} = \tilde{\sigma}^2 g_k^2(\underline{\theta})$ to reflect the fact that the standard deviation of the error will depend on the intensity $g_k(\underline{\theta})$ of the prediction of the output QoI in the measured location, where $\tilde{\sigma}$ is the standard deviation of the error normalized with respect to the intensity of the output QoI. For demonstration purposes, the spatial correlation structure

$R(\eta_{kl})$ is selected to be of the exponential form

$$R(\eta_{kl}) = exp\left[-\frac{\eta_{kl}}{\lambda}\right] \tag{11}$$

where λ is a measure of the spatial correlation length. Thus, the parameter set $\underline{\sigma}$ defining the structure of Σ is given by $\underline{\sigma} = (\bar{\sigma}, \tilde{\sigma}, \lambda)$.

The correlation model in the prediction error dominates the information provided by two sensors within the correlation length, causing this information content to increase as a function of the distance of the sensors within the correlation length, avoiding sensor clustering [7].

3. Conclusions

The information entropy (IE) is a rational measure of the Bayesian posterior uncertainty of the model parameters suitable to be used for quantifying the information contained in the data collected from a sensor configuration. Minimizing the IE with respect to the sensor positions provides the least uncertainty in the posterior parameter estimates and thus the optimal sensor configuration with the highest information. The design variables associated with the location of sensors in domains of several problems are defined in a continuous space. The stochastic optimization algorithm CMA-ES is suitable to carry out the minimization of the IE and obtain the global optimum, avoiding premature convergence to several observed local optima. An asymptotic estimate expresses the IE in terms of the sensitivities of the output quantities of interest with respect to the model parameters to be inferred. To avoid information redundancy that arise from sensor clustering, a spatially correlated prediction error model was used. The optimal experimental design is conditioned on nominal model and prediction error model parameters. To account for uncertainties in the nominal values of these model parameters and cover a number of experimental conditions, the robust information entropy is introduced as an integral of the conditional information entropy on these nominal values, weighted by the prior distribution of these parameters postulated in Bayesian analysis.

The largest information gain is obtained for a relatively small number of optimally placed sensor profiles. As the number of sensor profiles increases, the information gain from extra sensor profiles reduces. The optimal number of sensor profiles to be used in an experiment is a trade off between information gain and cost of instrumentation. Furthermore, uncertainties in the nominal model parameters are important since they leads to different sensor profile locations than the ones corresponding to the nominal parameter values. The proposed OSP framework is flexible to handle different spatially correlated modeling errors and it is applicable to more general flows and turbulence models.

4. References

[1] 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.

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

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

[4] K.-V. Yuen, S.-C. Kuok Efficient Bayesian sensor placement algorithm for structural identification: a general approach for multi-type sensory systems, Earthquake Engineering & Structural Dynamics, 44(5):757-774, 2015.

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

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

[7] 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.

[8] 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.