4. Bayesian Parameter Estimation in Structural Dynamics Using Response Time Histories

4.1 Introduction

The purpose is to estimate the uncertainties in the values of structural model parameters using measured response time histories. The model class used to represent structural behavior can be linear or non linear. The Bayesian framework is used for the estimation of the structural model parameters along with the associated uncertainties. For this, the uncertainty is quantified by probability distributions measuring the plausibility of each possible model in the model class introduced to represent the behavior of the structure. Prediction errors, measuring the fit between the measured and the model predicted response time histories, are modeled by Gaussian distributions. Two cases are then considered. In the first case, the prediction errors of response time history at different time instants are assumed to be independent Gaussian variables with equal variances for all sampling data of a response time history. The prediction errors between different responses are assumed to be independent. A special case, for which the prediction errors between different responses assumed to be fully correlated, is also considered. In the second case, the prediction error of a response time history at different time instances is quantified by an Autoregressive (AR) model and the Bayesian estimation is applied for identifying the optimal structural model and the optimal AR prediction error model, along with the associated uncertainties in these models. Equation Section 4

4.2 Bayesian Parameter Estimation Utilizing Response Time Histories

Let $D=\{\hat{x}_j(k\Delta t)\in R^{N_0},\ j=1,\cdots,N_0,\ k=1,\cdots,N_D\}$ be the measured response time histories data from a structure, consisting of response data (acceleration, velocity or displacement) at N_0 measured DOFs, where N_D is the number of the sampled data using a sampling rate Δt.

Consider a parameterized class of linear structural models M used to model the dynamic behaviour of the structure and let $\boldsymbol{\theta} \in R^{N_\theta}$ be the set of free structural model parameters to be identified using the measured response time histories. Let also $\{x_j(k\Delta t) \in R^{N_d},\ j = 1,\cdots,N_d,\ k = 1,\cdots,N_D\}$, where N_d is the number of model degrees of freedom (DOF), be the predictions of the response time histories obtained from a model corresponding to a particular value of the parameter set $\boldsymbol{\theta}$. For linear structures, this is done by solving the equation of motion

$$M(\boldsymbol{\theta})\ddot{\boldsymbol{q}}(t) + C(\boldsymbol{\theta})\dot{\boldsymbol{q}}(t) + K(\boldsymbol{\theta})\boldsymbol{q}(t) = \boldsymbol{f}(t) \tag{4.1}$$

where $M(\boldsymbol{\theta}), C(\boldsymbol{\theta})$ and $K(\boldsymbol{\theta}) \in \mathbb{R}^{N_d \times N_d}$ are the global model mass, damping and stiffness matrices respectively, $\boldsymbol{q}(t) \in \mathbb{R}^{N_d}$ is the displacement vector of the model DOFs, $\boldsymbol{f}(t) \in \mathbb{R}^{N_d}$ is the vector of forces at the model DOFs, N_d is the number of model DOFs. For nonlinear structures, this is done by solving the equation of motion

$$M(\boldsymbol{\theta})\ddot{\boldsymbol{q}}(t) + F(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), \boldsymbol{\theta}) = \boldsymbol{f}(t) \tag{4.2}$$

where $F(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), \boldsymbol{\theta})$ is the nonlinear restoring force that depends on the displacement and velocity response time history and the structural model parameter set $\boldsymbol{\theta}$.

The Bayesian approach uses probability distributions to quantify the plausibility of each possible value of the model parameters $\boldsymbol{\theta}$. Using Bayes' theorem, the updated (posterior) probability distribution $p(\boldsymbol{\theta} \mid D, \boldsymbol{\sigma}, \mathsf{M})$ of the model parameters $\boldsymbol{\theta}$ based on the inclusion of the measured data D, the modeling assumptions M and the value of a parameter set $\boldsymbol{\sigma}$, is obtained as follows:

$$p(\boldsymbol{\theta} \mid D, \boldsymbol{\sigma}, \mathsf{M}) = c\ p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}, \mathsf{M})\ p(\boldsymbol{\theta} \mid \boldsymbol{\sigma}, M) \tag{4.3}$$

where $p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}, \mathsf{M})$ is the probability of observing the data from a model corresponding to a particular value of the parameter set $\boldsymbol{\theta}$ conditioned on the modeling assumptions M and the value of $\boldsymbol{\sigma}$, $p(\boldsymbol{\theta} \mid \boldsymbol{\sigma}, M)$ is the initial (prior) probability distribution of a model, and c is a normalizing constant selected such that the PDF $p(\boldsymbol{\theta} \mid D, \boldsymbol{\sigma}, \mathsf{M})$ integrates to one. Herein, the modeling assumptions M refer to the structural modeling assumptions as well as those used to derive the probability distributions $p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}, \mathsf{M})$ and the prior $p(\boldsymbol{\theta} \mid \boldsymbol{\sigma}, M)$. The parameter set $\boldsymbol{\sigma}$ contains all parameters that need to be defined in order to completely specify the modeling assumptions M. Measured data are accounted for in the updated estimates through the term $p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}, \mathsf{M})$, while any available prior information is reflected in the term $p(\boldsymbol{\theta} \mid \boldsymbol{\sigma}, M)$. It is usually assumed that $p(\boldsymbol{\theta} \mid \boldsymbol{\sigma}, M) = \pi(\boldsymbol{\theta}) = constant$.

In order to simplify the notation, the dependence of the probability distributions on $\mathcal{M}$ is dropped in the analysis that follows.

The form of $p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}, \mathcal{M}) \equiv p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma})$ is derived by using a probability model for the prediction error vector $\boldsymbol{e}(k) = [e_1(k), \cdots, e_{N_0}(k)]$, $k = 1, \cdots, N_D$, defined as the difference between the measured response time histories involved in D for all measured degrees of freedom (DOFs), $j = 1, \cdots, N_0$, and the corresponding response time history predicted from a particular model that corresponds to a particular value of the parameter set $\boldsymbol{\theta}$.

Specifically, the prediction error $e_j(k)$ between the sampled response measured time histories and the corresponding response time histories predicted from a model that corresponds to a particular value of the parameter set $\boldsymbol{\theta}$, for the jth measured DOF and the kth sampled data, is given by the prediction error equation

$$e_j(k) = \hat{x}_j(k) - x_j(k; \boldsymbol{\theta}) \tag{4.4}$$

where $j = 1, \ldots, N_0$ and $k = 1, \ldots, N_D$.

4.3 Formulation for $p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma})$ Using Independence of the Model Prediction Error

Following the Bayesian methodology proposed, the predictions errors at different time instants are modeled by independent (identically distributed) zero-mean Gaussian vector variables. Specifically, the prediction error $e_j(k)$ for the jth measured DOF is assumed to be a zero mean Gaussian variable, $e_j(k) \sim N(0, \sigma_j^2)$ with variance σ_j^2. The prediction error parameter σ_j, $j = 1, \ldots, N_0$ represents the fractional difference between the measured and the model predicted response at a time instant. The model prediction error is due to modeling error and measurement noise.

In the analysis that follows, the parameter set $\boldsymbol{\sigma}$ is taken to contain the prediction error parameters σ_j, $j = 1, \ldots, N_0$. Given the values of the parameter set $\boldsymbol{\sigma}$, assuming independence of the prediction errors $\boldsymbol{e}(k)$ and using the Gaussian choice for the probability distribution of the prediction errors $e_j(k)$, the probability $p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma})$ of obtaining the data from a model within the class of models $\mathcal{M}$ can be simplified as follows.

First note that

$$p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) = p\left(\left\{\hat{\boldsymbol{x}}(k), k = 1, \ldots, N_D\right\} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}\right) \tag{4.5}$$

$$= p\left(\left\{\hat{x}_1(k), \ldots, \hat{x}_{N_0}(k), k = 1, \ldots, N_D\right\} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}\right) \tag{4.6}$$

From probability theory it is known that:

$$p(x_1,\ldots,x_N \mid y) = \tag{4.7}$$

$$= p(x_N \mid x_1,\ldots,x_{N-1},y)\; p(x_1,\ldots,x_{N-1} \mid y) = \tag{4.8}$$

$$= p(x_N \mid x_1,\ldots,x_{N-1},y)\; p(x_{N-1} \mid x_1,\ldots,x_{N-2},y)\; p(x_1,\ldots,x_{N-2} \mid y) \tag{4.9}$$

$$= \prod_{i=1}^{N} p(x_i \mid x_1,x_2,\ldots,x_{i-1},y) \tag{4.10}$$

Also, it is known that if the variables x_1 and x_2 are independent

$$p(x_1,x_2) = p(x_1)\,p(x_2) \tag{4.11}$$

Using equations (4.10), (4.11) and the independence assumed for the prediction errors between different response time histories, equation (4.6) can equivalent be written as

$$p(D \mid \boldsymbol{\theta},\boldsymbol{\sigma}) = \prod_{j=1}^{N_0} p\left(\left\{\hat{x}_j(k), k = 1,\ldots,N_D\right\} \mid \boldsymbol{\theta},\boldsymbol{\sigma}\right) \tag{4.12}$$

Also, assuming that the prediction errors for a response time history are independent at different time instants, one obtains that the j factor in (4.12) is given by

$$p\left(\hat{x}_j(k), k = 1,\ldots,N_D \mid \boldsymbol{\theta},\boldsymbol{\sigma}\right) = \prod_{k=1}^{N_D} p(\hat{x}_j(k) \mid \boldsymbol{\theta},\boldsymbol{\sigma}) \tag{4.13}$$

From equation (4.12) and (4.13) one derives that

$$p(D \mid \boldsymbol{\theta},\boldsymbol{\sigma}) = \prod_{j=1}^{N_0}\prod_{k=1}^{N_D} p(\hat{x}_j(k) \mid \boldsymbol{\theta},\boldsymbol{\sigma}) \tag{4.14}$$

Assuming that the predictions errors $e_j(k)$ are modeled by zero-mean Gaussian variables so that $e_j(k) \sim N(0,\sigma_j^2)$, the measured time histories $\hat{x}_j(k)$ are also implied to be Gaussian variables, that is, $\hat{x}_j(k) \sim N\left(x_j(k;\boldsymbol{\theta}),\sigma_j^2\right)$, with mean $x_j(k;\boldsymbol{\theta})$ and variance σ_j^2. Therefore, the probability density function (PDF) of $\hat{x}_j(k)$, given the values of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$, is given by

$$p\left(\hat{x}_j(k) \mid \boldsymbol{\theta},\boldsymbol{\sigma}\right) = \frac{1}{\sqrt{2\pi}\ \sigma_j} \exp\left\{-\frac{1}{2\sigma_j^2}\left[\hat{x}_j(k) - x_j(k;\boldsymbol{\theta})\right]^2\right\} \tag{4.15}$$

Substituting equation (4.15) in (4.14), one obtains that

$$p(D \mid \boldsymbol{\theta},\boldsymbol{\sigma}) = \prod_{j=1}^{N_0}\prod_{k=1}^{N_D} \frac{1}{\sqrt{2\pi}\ \sigma_j} \exp\left\{-\frac{1}{2\sigma_j^2}\left[\hat{x}_j(k) - x_j(k;\boldsymbol{\theta})\right]^2\right\} \tag{4.16}$$

Using the property

$$\prod_{i=1}^{N} \exp\{f_i\} = \exp\left\{\sum_{i=1}^{N} f_i\right\} \tag{4.17}$$

equation (4.16) takes the form

$$\begin{aligned} p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) &= \frac{1}{\left(\sqrt{2\pi}\right)^{N_D N_0} \prod_{j=1}^{N_0} \sigma_j^{N_D}} \prod_{j=1}^{N_0} \exp\left\{-\frac{1}{2\sigma_j^2} \sum_{k=1}^{N_D} \left[\hat{x}_j(k) - x_j(k;\boldsymbol{\theta})\right]^2\right\} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^{N_D N_0} \prod_{j=1}^{N_0} \sigma_j^{N_D}} \exp\left\{-\frac{1}{2} \sum_{j=1}^{N_0} \frac{1}{\sigma_j^2} \sum_{k=1}^{N_D} \left[\hat{x}_j(k) - x_j(k;\boldsymbol{\theta})\right]^2\right\} \end{aligned} \tag{4.18}$$

Equivalently, introducing the function

$$J(\boldsymbol{\theta};\boldsymbol{\sigma}) = \sum_{j=1}^{N_0} \frac{\alpha_i}{\sigma_j^2} J_j(\boldsymbol{\theta}) \tag{4.19}$$

where

$$J_j(\boldsymbol{\theta}) = \frac{1}{N_D} \sum_{k=1}^{N_D} \left[\hat{x}_j(k) - x_j(k;\boldsymbol{\theta})\right]^2 \tag{4.20}$$

and

$$\alpha_i = \frac{1}{N_0} \tag{4.21}$$

equation (4.18) can be written in the form

$$p(D \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) = \frac{1}{\left(\sqrt{2\pi}\right)^{N_D N_0} \rho(\boldsymbol{\sigma})} \exp\left\{-\frac{N_D N_0}{2} J(\boldsymbol{\theta};\boldsymbol{\sigma})\right\} \tag{4.22}$$

Substituting (4.22) into equation (4.3) one readily derives the probability distribution of the structural model parameters in the form

$$p(\boldsymbol{\theta} \mid D, \boldsymbol{\sigma}) = \frac{c \;\; \pi(\boldsymbol{\theta})}{\left(\sqrt{2\pi}\right)^{N_D N_0} \rho(\boldsymbol{\sigma})} \exp\left\{-\frac{N_D N_0}{2\sigma^2} J(\boldsymbol{\theta};\boldsymbol{\sigma})\right\} \tag{4.23}$$

The function $J_j(\boldsymbol{\theta})$ in (4.20) represents the measure of fit between the measured and the model predicted response time history for the jth DOF. The function $J(\boldsymbol{\theta};\boldsymbol{\sigma})$

represents the overall weighted measure of fit between measured and predicted response time histories for all DOFs. The variable $\rho(\boldsymbol{\sigma})$ is a scalar function of the prediction error parameter set $\boldsymbol{\sigma}$ given by

$$\rho(\boldsymbol{\sigma}) = \prod_{j=1}^{N_0} \sigma_j^{N_D} \tag{4.24}$$

The optimal value $\hat{\boldsymbol{\theta}}$ of the model parameter $\boldsymbol{\theta}$ for given $\boldsymbol{\sigma}$, denoted by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{\sigma})$, correspond to the most probable model maximizing the updated PDF $p(\boldsymbol{\theta} \mid D, \boldsymbol{\sigma})$, that is,

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\sigma}) = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid D, \boldsymbol{\sigma}) \tag{4.25}$$

In particular, for a non-informative uniform prior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{\sigma}, M) = = \pi(\boldsymbol{\theta}) = constant$, the optimal values $\hat{\boldsymbol{\theta}}$ of the model parameter $\boldsymbol{\theta}$ correspond to the values that minimize the measure of fit $J(\boldsymbol{\theta};\boldsymbol{\sigma})$ defined in (4.19), that is

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\sigma}) = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta};\boldsymbol{\sigma}) \tag{4.26}$$

It should be noted that the optimal value $\hat{\boldsymbol{\theta}}$ depends on the values of the prediction error parameter set $\boldsymbol{\sigma}$.

4.3.1 Special Case: $\sigma_j = \sigma \ \forall \ j$

Next, the prediction error parameters σ_j, representing the prediction error estimates of the measured time histories involved in D, are assumed to be the same for any measured degree of freedom, that is $\sigma_j = \sigma$ for all $j = 1, \ldots, N_0$. In that case, the measure of fit $J(\boldsymbol{\theta};\boldsymbol{\sigma})$ given in equation (4.19) takes the form

$$J(\boldsymbol{\theta};\boldsymbol{\sigma}) = \frac{1}{\sigma^2} \frac{1}{N_0} \sum_{j=1}^{N_0} J_j(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \overline{J}(\boldsymbol{\theta}) \tag{4.27}$$

where

$$\overline{J}(\boldsymbol{\theta}) = \frac{1}{N_0} \sum_{j=1}^{N_0} J_j(\boldsymbol{\theta}) \tag{4.28}$$

while the function for $\rho(\boldsymbol{\sigma})$, given in equation (4.24), simplifies to

$$\rho(\boldsymbol{\sigma}) = \sigma^{N_D N_0} \tag{4.29}$$

The optimal value $\hat{\boldsymbol{\theta}}$ of the model parameter $\boldsymbol{\theta}$ correspond to the most probable model maximizing the updated PDF $p(\boldsymbol{\theta} \mid D, \boldsymbol{\sigma})$ or equally minimizing $\overline{J}(\boldsymbol{\theta})$, that is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \overline{J}(\boldsymbol{\theta}) \tag{4.30}$$

with its value be independent of the values of the prediction error parameter set $\boldsymbol{\sigma}$ .The optimal value $\hat{\boldsymbol{\sigma}}$ of the prediction error parameter $\boldsymbol{\sigma}$ is given by

$$\boldsymbol{\sigma}^2 = \overline{J}(\hat{\boldsymbol{\theta}}) \tag{4.31}$$

4.4 Bayesian Parameter Estimation Using AR Models for the Model Prediction Error

Here the case where the prediction error of a response time history is quantified by an Autoregressive (AR) model is considered. A Bayesian estimation is presented for identifying the optimal values of the structural models and the parameters of the AR prediction error model, as well as their associated uncertainties.

The prediction error $e_j(k)$ between the sampled measured response time histories and the response time histories predicted by a model that corresponds to a particular value of the parameters $\boldsymbol{\theta}$, for the j th measured DOF, $j = 1,\ldots,N_0$, and the k th sampled data, $k = 1,\ldots,N_D$, is given by the prediction error equation

$$e_j(k) = \hat{x}_j(k) - x_j(k;\boldsymbol{\theta}) \tag{4.32}$$

where $e_j(k)$ are no longer independent but they are related through the structure of an autoregressive (AR) probability model of order p. That is, $e_j(k)$ satisfy the AR equation

$$e_j(k) - \sum_{i=1}^{p} \alpha_i e_i(k-i) = n_j(k) \tag{4.33}$$

The AR prediction errors $n_j(k)$, $k = 1,\ldots,N_D$, corresponding to the j th measured point, are assumed to be independent zero mean Gaussian variables, $n_j(k) \sim N(0,\sigma_j^2)$, with variance σ_j^2 independent of k. The parameters σ_j, $j = 1,\ldots,N_0$ represent the prediction error estimates of the AR model and using the structure of the AR model, they also affect the model error corresponding to the measured time histories involved in D. The model prediction error is due to modeling error and measurement noise.

Using the Bayes theorem to calculate the posterior probability distribution $p(\boldsymbol{\theta} \mid D,\boldsymbol{\alpha},\boldsymbol{\sigma})$ of the model parameters $\boldsymbol{\theta}$ based on the inclusion of the measured data D and given the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$, one has that

$$p(\boldsymbol{\theta} \mid D,\boldsymbol{\alpha},\boldsymbol{\sigma}) = \frac{p(D \mid \boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\sigma})\; p(\boldsymbol{\theta} \mid \boldsymbol{\alpha},\boldsymbol{\sigma})}{p(D \mid \boldsymbol{\alpha},\boldsymbol{\sigma})} \tag{4.34}$$

$p(D \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma})$ is obtained by using information from the data D, the model class M, and the structure of the AR model.

$p(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\sigma})$ is the initial (prior) probability distribution of the model parameters $\boldsymbol{\theta}$. It is usually assumed uniform, and here equal to $\pi(\boldsymbol{\theta})$. It is also assumed that $\boldsymbol{\theta}$ is independent of $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ before the data are obtained.

$p(D \mid \boldsymbol{\alpha}, \boldsymbol{\sigma})$ is a normalizing constant, c^{-1}, selected such that the PDF $p(\boldsymbol{\theta}|D, \boldsymbol{\alpha}, \boldsymbol{\sigma})$ integrates to one.

4.4.1 Formulation for $p(D \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma})$

Assuming independence of the measured response time histories $\hat{x}_j(k)$ for different values of the index j, the probability $p(D \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma})$, known also as "the likelihood", of obtaining the data from a model within the class of models M, is given by

$$p(D \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}) = p\left(\left\{\hat{\boldsymbol{x}}(k), k = 1, \ldots, N_D\right\} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}\right) =$$

$$= p\left(\left\{\hat{x}_1(k), \ldots, \hat{x}_{N_0}(k), k = 1, \ldots, N_D\right\} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}\right)$$

$$= \prod_{j=1}^{N_0} p\left(\left\{\hat{x}_j(k), k = 1, \ldots, N_D\right\} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}\right) \tag{4.35}$$

Also, using the fact that

$$p\left(\left\{\hat{x}_j(k), k = 1, \ldots, N_D\right\} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}\right) = \prod_{k=1}^{N_D} p\left(\hat{x}_j(k) \mid \hat{x}_j(k-1), \ldots, \hat{x}_j(1), \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}\right) \tag{4.36}$$

and substituting in (4.35), one derives that

$$p(D \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}) = \prod_{j=1}^{N_0} \prod_{k=1}^{N_D} p\left(\hat{x}_j(k) \mid \hat{x}_j(k-1), \ldots, \hat{x}_j(1), \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}\right) \tag{4.37}$$

Next, the expressions for the factors in (4.37) are developed. Substituting $e_j(k)$ from (4.32) in (4.33), one has

$$\hat{x}_j(k) - x_j(k; \boldsymbol{\theta}) - \sum_{i=1}^{p} \alpha_{ji} \left[\hat{x}_j(k-i) - x_j(k-i; \boldsymbol{\theta})\right] = n_j(k) \tag{4.38}$$

which can be rearranged in the form

$$\hat{x}_j(k) = x_j(k; \boldsymbol{\theta}) + \sum_{i=1}^{p} \alpha_{ji} \left[\hat{x}_j(k-i) - x_j(k-i; \boldsymbol{\theta})\right] + n_j(k) \tag{4.39}$$

Assuming that the predictions errors $n_j(k)$ are modeled by zero-mean Gaussian variables so that $n_j(k) \sim N(0,\sigma_j^2)$, the measured time history $\hat{x}_j(k)$ at time instant $k\Delta t$, given the values of $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ and the measured response time histories at previous time instances $(k-1)\Delta t,\ldots,\Delta t$, are also implied to be Gaussian variables, that is, $\hat{x}_j(k) \sim N\left(m_j(k;\boldsymbol{\theta}),\sigma_j^2\right)$ with mean $m_j(k;\boldsymbol{\theta})$ and variance σ_j^2, where

$$m_j(k;\boldsymbol{\theta}) = x_j(k;\boldsymbol{\theta}) + \sum_{i=1}^{p} \alpha_{ji}\left[\hat{x}_j(k-i) - x_j(k-i;\boldsymbol{\theta})\right] \tag{4.40}$$

Therefore, the probability density function (PDF) of $\hat{x}_j(k)$, given the values of $\boldsymbol{\theta},\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$, is given by

$$p\left(\hat{x}_j(k) \mid \hat{x}_j(k-1),\ldots,\hat{x}_j(1),\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\sigma}\right) = p\left(\hat{x}_j(k) \mid \hat{x}_j(k-1),\ldots,\hat{x}_j(k-i),\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\sigma}\right) =$$

$$\frac{1}{\sqrt{2\pi}\ \sigma_j} \exp\left\{-\frac{1}{2\sigma_j^2}\left[\hat{x}_j(k)-x_j(k;\boldsymbol{\theta})-\sum_{i=1}^{p}\alpha_{ji}\left[\hat{x}_j(k-i)-x_j(k-i;\boldsymbol{\theta})\right]\right]^2\right\} \tag{4.41}$$

Substituting (4.41) into (4.36) one has

$$p\left(\left\{\hat{x}_j(k),k=1,\ldots,N_D\right\}|\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\sigma}\right)=\frac{1}{\left(\sqrt{2\pi}\right)^{N_D}\sigma_j^{N_D}}\exp\left\{-\frac{1}{2\sigma_j^2}\sum_{k=1}^{N_D}\left[\sum_{i=0}^{p}\alpha_{ji}\left[\hat{x}_j(k-i)-x_j(k-i;\boldsymbol{\theta})\right]\right]^2\right\} \tag{4.42}$$

where $\alpha_{j0} = -1$.

Substituting (4.42) into equation (4.37) results in

$$p(D|\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\sigma})=\frac{1}{\left(\sqrt{2\pi}\right)^{N_D N_0}\prod_{j=1}^{N_0}\sigma_j^{N_D}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{N_0}\frac{1}{\sigma_j^2}\sum_{k=1}^{N_D}\left[\sum_{i=0}^{p}\alpha_{ji}\left[\hat{x}_j(k-i)-x_j(k-i;\boldsymbol{\theta})\right]\right]^2\right\} \tag{4.43}$$

Equivalently, introducing the function

$$J(\boldsymbol{\theta};\boldsymbol{\alpha},\boldsymbol{\sigma}) = \sum_{j=1}^{N_0}\frac{\alpha_i}{\sigma_j^2}J_j(\boldsymbol{\theta};\boldsymbol{\alpha}) \tag{4.44}$$

where

$$J_j(\boldsymbol{\theta};\boldsymbol{\alpha})=\frac{1}{N_D}\sum_{k=1}^{N_D}\left[\sum_{i=0}^{p}\alpha_{ji}\left[\hat{x}_j(k-i)-x_j(k-i;\boldsymbol{\theta})\right]\right]^2 \tag{4.45}$$

and $\alpha_i = 1/N_0$, the probability density function $p(D \mid \boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\sigma})$ in (4.43) takes the compact form

$$p(D \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}) = \frac{1}{\left(\sqrt{2\pi}\right)^{N_D N_0} \rho(\boldsymbol{\sigma})} \exp\left\{ -\frac{N_D N_0}{2} J(\boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\sigma}) \right\} \tag{4.46}$$

The function $J_j(\boldsymbol{\theta}; \boldsymbol{\alpha})$ represents the measure of fit between the measured and the model predicted response time history for the jth DOF. The function $J(\boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\sigma})$ represents the overall weighted measure of fit between measured and predicted response time histories for all DOFs. The variable $\rho(\boldsymbol{\sigma})$ is the scalar function defined in (4.24).

The optimal value $\hat{\boldsymbol{\theta}}$ of the model parameter $\boldsymbol{\theta}$ for given $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$, denoted by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{\alpha}, \boldsymbol{\sigma})$, correspond to the most probable model maximizing the updated PDF $p(\boldsymbol{\theta} \mid D, \boldsymbol{\alpha}, \boldsymbol{\sigma})$, that is

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\alpha}, \boldsymbol{\sigma}) = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid D, \boldsymbol{\alpha}, \boldsymbol{\sigma}) \tag{4.47}$$

In particular for non-informative uniform prior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{\sigma}, \mathsf{M}) = \pi(\boldsymbol{\theta}) = = constant$, the optimal value $\hat{\boldsymbol{\theta}}$ of the model parameter $\boldsymbol{\theta}$ correspond to the value that minimize the measure of fit $J(\boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\sigma})$ defined in (4.44). Hence, it is given by

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\alpha}, \boldsymbol{\sigma}) = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\sigma}) \tag{4.48}$$

It should be noted that the optimal value $\hat{\boldsymbol{\theta}}$ of the model parameters $\boldsymbol{\theta}$ depends on the values of prediction error parameter set $\boldsymbol{\sigma}$ and the parameter set $\boldsymbol{\alpha}$ defining the structure of the AR model. This dependence is explicitly shown by writing $\hat{\boldsymbol{\theta}}(\boldsymbol{\alpha}, \boldsymbol{\sigma})$.