1 Parameter Estimation: Multi-Dimensional Case

Consider the case of several uncertain parameters $\underline{X} = (X_1, \ldots, X_n) \in R^n$ of a model. Bayes theorem is used to make inference about the values of these parameters based on a set of data D and the background information I. Specifically the posterior distribution of the model parameters is given by

$$p(\underline{x} \mid D, I) = \frac{p(D \mid \underline{x}, I)\ \ p(\underline{x} \mid I)}{p(D \mid I)} \tag{1}$$

which completely quantifies the uncertainties in the values $\underline{x} = (x_1, \ldots, x_n)$ of the model parameters. Similar to the two-dimensional parameter case, the most probable value or the best estimate $\underline{\hat{x}}$ of the values of the model parameters is the one that maximizes the posterior PDF $p(\underline{x} \mid D, I)$ or, equivalently, minimizes the function

$$L(\underline{x}) = -\log[\, p(\underline{x} \mid D, I)] \tag{2}$$

Note that the posterior PDF can be written in terms of the function $L(\underline{x})$ in the form

$$p(\underline{x} \mid D, I) = \exp[-L(\underline{x})] \tag{3}$$

1.1 General Case of Several Parameters

Consider now the general case of n parameters. The best estimates of the model parameters are obtained by simultaneously solving the following system of two equations

$$\underline{\nabla}^T L(\underline{x})\Big|_{\underline{x}=\underline{\hat{x}}} = \underline{0} \tag{4}$$

and ensure that the solution $\underline{\hat{x}}$ corresponds to a minimum of $L(\underline{x})$. The uncertainty in the values of the parameters are obtained by considering the spread of the n-dimensional posterior PDF about the best estimate $\underline{\hat{x}}$.

The local behavior of the posterior PDF about $\underline{\hat{x}}$ is obtained by the Taylor series expansion of the function $L(\underline{x})$ about $\underline{\hat{x}}$, given by

$$L(\underline{x}) = L(\underline{\hat{x}}) + \nabla^T L(\underline{x})\Big|_{\underline{x}=\underline{\hat{x}}} (\underline{x} - \underline{\hat{x}}) + \frac{1}{2}(\underline{x} - \underline{\hat{x}})^T \underline{\nabla}\underline{\nabla}^T L(\underline{x})\Big|_{\underline{x}=\underline{\hat{x}}} (\underline{x} - \underline{\hat{x}}) + \cdots$$

Using the fact that we expand around the minimum value $\underline{\hat{x}}$ of $L(\underline{x})$, the linear terms in the Taylor series expansion are zero because of (4). Introducing the Hessian matrix $H(\underline{x})$ of the function $L(\underline{x})$ by the form

$$H(\underline{x}) = \underline{\nabla}\underline{\nabla}^T L(\underline{x})$$

the Taylor series expansion of $L(\underline{x})$ takes the form

$$L(\underline{x}) = L(\hat{\underline{x}}) + \frac{1}{2}(\underline{x} - \hat{\underline{x}})^T H(\hat{\underline{x}})(\underline{x} - \hat{\underline{x}}) + \cdots$$

or equivalently

$$L(\underline{x}) = L(\hat{\underline{x}}) + \frac{1}{2}Q(\underline{x}) + \cdots \tag{5}$$

where $Q(\underline{x})$ takes the quadratic form

$$Q(\underline{x}) = (\underline{x} - \hat{\underline{x}})^T H(\hat{\underline{x}})(\underline{x} - \hat{\underline{x}}) \tag{6}$$

Note that at the neighbor of the best estimate, the terms of the order of three or higher in the Taylor series expansion of $L(\underline{x})$ can be neglected and the behavior of the function $L(\underline{x})$ locally is specified by the behavior of the quadratic form $Q(\underline{x})$. Specifically the spread of uncertainty around the best estimate $\hat{\underline{x}}$ is determined by the contour curves of function $Q(\underline{x})$. Using the fact that $\hat{\underline{x}}$ is the minimum of $L(\underline{x})$, then the Hessian of $L(\underline{x})$ is positive definite or, equivalently, that the quadratic form $Q(\underline{x})$ is positive for any $\underline{x} - \hat{\underline{x}} \neq (0,\ldots,0)^T$. The points $\underline{x}$ in the parameter space that belong to the contour curve of $Q(\underline{x})$ corresponding to an energy level $\kappa^2 > 0$, have coordinates that satisfy the equation

$$Q(\underline{x}) = (\underline{x} - \hat{\underline{x}})^T H(\hat{\underline{x}})(\underline{x} - \hat{\underline{x}}) = \kappa^2$$

Consider the eigenvalues λ_i, $i = 1,\ldots,n$, and the corresponding eigenvectors $\underline{u}_i$, $i = 1,\ldots,n$, of the positive definite symmetric matrix $\hat{H} \equiv H(\hat{\underline{x}})$ obtained by solving the eigenvalue problem

$$\hat{H}\underline{u} = \lambda \underline{u}$$

From linear algebra results, it is well known that for a positive definite symmetric matrix $\hat{H}$, the eigenvalues are positive i.e. $\lambda_i > 0$, $i = 1,\ldots,n$, while the eigenvectors $\underline{u}_i$, $i = 1,\ldots,n$, are orthogonal. Normalize that eigenvectors $\underline{u}_i$, $i = 1,\ldots,n$, so that they have unit length. These orthogonal unit vectors define certain orthogonal directions in the parameter space. Introducing now the matrix of eigenvectors $U = [\underline{u}_1,\ldots,\underline{u}_n]$ and invoking known relevant results from linear algebra, one can write the orthogonality conditions:

$$UU^T = U^T U = I$$

$$U^T \hat{H} U = \Lambda$$

where Λ is the diagonal matrix of the eigenvalues of $\hat{H}$. The first condition implies that the matrix of eigenvectors Q is orthogonal. Also, from linear algebra, it is well-known that the orthonormal eigenvectors $\underline{u}_i$, $i = 1,\ldots,n$, constitute a basis of the n-dimensional vector space or, equivalently, any vector $\underline{x} - \hat{\underline{x}} \in R^n$ can be written in terms of the basis of eigenvectors $\{\underline{u}_1,\ldots,\underline{u}_n\}$ as

$$\underline{x} - \underline{\mu} = \sum_{i=1}^{n} y_i \underline{u}_i = U\underline{y} \tag{7}$$

where $\underline{y} = (y_1, \ldots, y_n)^T \in R^n$, which is an alternative representation of the vector $\underline{x} - \underline{\hat{x}}$ using its components $y_1, \ldots, y_n$ with respect to the new orthonormal basis of eigenvectors $\{\underline{u}_1, \ldots, \underline{u}_n\}$.

Substituting (7) into the quadratic form (6), one derives the quadratic form in terms of the components $y_1, \ldots, y_n$ of the vector $\underline{x} - \underline{\hat{x}}$ in the new basis as

$$Q(\underline{x}) = \underline{y}^T U^T H U \underline{y} = \underline{y}^T \Lambda \underline{y} = \sum_{i=1}^{n} \lambda_i y_i^2 \tag{8}$$

It is clear that the symmetric matrix associated with the quadratic form in the new basis $\{\underline{u}_1, \ldots, \underline{u}_n\}$ of the eigenvectors of H is the diagonal matrix Λ of the eigenvalues of H.

Consider now the points at the contour of the function $Q(\underline{x})$ corresponding to an "energy" level κ. Such points in the n-dimensional space satisfy the equation

$$Q(\underline{x}) = \kappa^2$$

or, equivalently, using (8), the components with respect to the eigenvector basis satisfy

$$\sum_{i=1}^{n} \lambda_i y_i^2 = \hat{Q}(\underline{y}) = \kappa^2$$

which can be written in the form

$$\sum_{i=1}^{n} \frac{y_i^2}{\alpha_i^2} = 1 \tag{9}$$

where $\alpha_i = \dfrac{\kappa}{\sqrt{\lambda_i}}$. Equation (9) represents an hyper-ellipse that is centered in the point $\underline{\hat{x}}$ in the parameter space with principal axis along the directions specified by the eigenvectors and size of the principal axis equal to α_i, i.e. the size of the principal axes are inversely proportional to the square root of the eigenvalues. Thus, the eigenvalues and the eigenvectors of the matrix $\hat{H}$ define completely the characteristics of this hyper-ellipse in the n-dimensional space, containing all points with coordinate values satisfying the equation (9). As in the two-parameter case, the contour curve specifies the spread of the uncertainty in the values of the parameters in $\underline{x}$ in the n-dimensional parameter space.

Asymptotic Approximation of Posterior PDF: Substituting the Taylor series expansion (5) into the posterior PDF (3) and keeping only up to the quadratic terms in the Taylor expansion, the posterior PDF is approximated by

$$p(\underline{x} \mid D, I) = \exp[-L(\underline{x})] \propto \exp[-Q(\underline{x})]$$
$$\propto \exp\left[-\frac{1}{2}(\underline{x} - \underline{\hat{x}})^T H(\underline{\hat{x}})(\underline{x} - \underline{\hat{x}})\right]$$

Introducing the covariance matrix

$$C = H^{-1}(\underline{\hat{x}})$$

as the inverse of the Hessian of $L(\underline{x})$ evaluated at the most probable value $\underline{\hat{x}}$ of the model parameters, the posterior PDF is approximated by the multi-variable Gaussian PDF.

$$p(\underline{x} \mid D, I) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{\det C}} \exp\left[-\frac{1}{2} (\underline{x} - \underline{\hat{x}})^T C^{-1} (\underline{x} - \underline{\hat{x}}) \right] \qquad (10)$$

Remarks

1. The Bayesian Central Limit Theorem, outlined for the two-dimensional case in Remark 1, holds also for the n dimensional case. Specifically, the posterior PDF asymptotically approximates the Gaussian multivariate PDF centered at the most probable value $\underline{\hat{x}}$ and with covariance matrix $C = H^{-1}(\underline{\hat{x}})$, given by (10).
2. The spread of the uncertainty in the parameters around the best estimate $\underline{\hat{x}}$ is completely defined by the Hessian matrix $\hat{H} = H(\underline{\hat{x}})$ or equivalently by the covariance matrix $C = H^{-1}(\underline{\hat{x}})$.
3. In order to obtain the marginal distribution of a parameter, say x_i, we need to integrate out the values of the rest of the parameters $\underline{\tilde{x}}_i = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ using the marginalization theorem

   $$p(x_i \mid D, I) = \int p(x_i, \underline{\tilde{x}}_i \mid D, I)\, d\underline{\tilde{x}}_i$$

   However, this is a multi-dimensional integral which cannot be evaluated numerically for more than a few parameters. However, using the asymptotic Gaussian approximation of the joint posterior PDF $p(\underline{x} \mid D, I)$ defined in (10), one can readily obtain that the marginal PDF $p(x_i \mid D, I)$ is also Gaussian distribution with mean $\hat{x}_i$ and variance C_{ii} , the (i,i) diagonal component of the covariance matrix C. The best estimate of x_i is $\hat{x}_i$ and the spread of the uncertainty in the parameter x_i about the best estimate is defined by $\sqrt{C_{ii}}$. It should be emphasized that the estimates $\sqrt{C_{ii}}$ of the uncertainties in each one of the parameters x_i give an incomplete picture of the uncertainties since they do not take into account the correlation between the variables in the vector $\underline{x}$.
4. Using the linear transformation of variables

   $$\underline{x} - \underline{\hat{x}} = U\underline{y}$$

   The fact that asymptotically the variables in $\underline{x}$ are Gaussian and that a linear transformation of Gaussian variables results in Gaussian variables as well, the posterior PDF for the new variables $\underline{y} = U^{-1}(\underline{x} - \underline{\hat{x}})$ are also Gaussian with zero mean and diagonal covariance $E[\underline{y}\underline{y}^T] = \Lambda^{-1}$ (see general proof in Remark 4) . Specifically, the posterior PDF of $\underline{y}$ is given by

$$p(\underline{y} \mid D, I) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{1/(\lambda_1 \ldots \lambda_n)}} \exp\left[-\frac{1}{2}\underline{y}^T \Lambda \underline{y}\right] = \prod_{k=1}^{k} \frac{1}{\sqrt{2\pi}\sqrt{1/\lambda_k}} \exp\left[-\frac{y_k^2}{2(1/\lambda_k)}\right]$$

The spread of the uncertainty in the parameters in y along the directions defined by the unit eigenvectors $\underline{u}_i$, $i = 1, \ldots, n$, are inversely proportional to the square root of the eigenvalues λ_i, $i = 1, \ldots, n$,. The variables $1/\sqrt{\lambda_i}$, $i = 1, \ldots, n$, provide the spread of the uncertainties of the variables y_i, $i = 1, \ldots, n$. Moreover, the variables $1/\sqrt{\lambda_i}$, $i = 1, \ldots, n$, give a complete picture of the spread of the uncertainties in the n-dimensional parameter space, locally around the best estimate $\underline{\hat{x}}$, in the directions specified by the eigenvectors $\underline{u}_i$, $i = 1, \ldots, n$, of the Hessian matrix $\hat{H} = H(\underline{\hat{x}})$.