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 one-dimensional parameter case, the most probable value or the best estimate $\hat{\underline{x}}$ 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

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

1.1 Special Case of Two Parameters

For demonstration purposes, consider first the special case of two parameters, i.e. $n = 2$. The best estimates of the model parameters are obtained by simultaneously solving the following system of two equations

$$\left.\frac{\partial L}{\partial x_i}\right|_{x=\hat{x}} = 0, \qquad i = 1, 2 \tag{3}$$

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

The local behavior of the posterior PDF about $\hat{\underline{x}}$ is obtained by the Taylor series expansion of the function $L(\underline{x}) = L(x_1, x_2)$ about $\hat{x} = (\hat{x}_1, \hat{x}_2)^T$, given by

$$\begin{aligned} L(x_1, x_2) = L(\hat{x}_1, \hat{x}_2) &+ \left.\frac{\partial L}{\partial x_1}\right|_{x=\hat{x}} (x_1 - \hat{x}_1) + \left.\frac{\partial L}{\partial x_2}\right|_{x=\hat{x}} (x_2 - \hat{x}_2) + \frac{1}{2} \left.\frac{\partial^2 L}{\partial x_1^2}\right|_{x=\hat{x}} (x_1 - \hat{x}_1)^2 \\ &+ \left.\frac{\partial^2 L}{\partial x_1 \partial x_2}\right|_{x=\hat{x}} (x_1 - \hat{x}_1)(x_2 - \hat{x}_2) + \frac{1}{2} \left.\frac{\partial^2 L}{\partial x_2^2}\right|_{x=\hat{x}} (x_2 - \hat{x}_2)^2 + \cdots \end{aligned}$$

Using the fact that we expand around the minimum of $L(\underline{x})$, the linear terms in the Taylor series expansion are zero because of (3). Introducing the Hessian matrix $H(\underline{x})$ of the function $L(\underline{x})$ by the form

$$H(\underline{x}) = \begin{bmatrix} \dfrac{\partial^2 L}{\partial x_1^2} & \dfrac{\partial^2 L}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 L}{\partial x_1 \partial x_2} & \dfrac{\partial^2 L}{\partial x_2^2} \end{bmatrix}$$

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

$$L(x_1, x_2) = L(\hat{x}_1, \hat{x}_2) + \frac{1}{2}(\underline{x} - \underline{\hat{x}})^T H(\underline{\hat{x}})(\underline{x} - \underline{\hat{x}}) + \cdots$$

or equivalently

$$L(x_1, x_2) = L(\hat{x}_1, \hat{x}_2) + \frac{1}{2} Q(\underline{x}) + \cdots \tag{4}$$

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

$$Q(\underline{x}) = (\underline{x} - \underline{\hat{x}})^T H(\underline{\hat{x}})(\underline{x} - \underline{\hat{x}}) \tag{5}$$

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 $\underline{\hat{x}}$ is determined by the contour curves of function $Q(\underline{x})$ which, by making use of (2), are exactly the same as the contour curves of the posterior PDF

$$\begin{aligned} p(\underline{x} \mid D, I) &= \exp[-L(\underline{x})] \\ &\propto \exp\left[-\frac{1}{2} Q(\underline{x})\right] \end{aligned} \tag{6}$$

First, we know from linear algebra that the condition for $\underline{\hat{x}}$ to be a minimum of $L(\underline{x})$ is that the Hessian of $L(\underline{x})$ is positive definite or, equivalently, that the quadratic form $Q(\underline{x})$ is positive for any $\underline{x} - \underline{\hat{x}} \neq (0,0)^T$. The points $\underline{x}$ in the parameter space that belong to the contour curve of $Q(\underline{x})$ corresponding to an energy level $\kappa > 0$, have coordinates that satisfy the equation

$$Q(\underline{x}) = (\underline{x} - \underline{\hat{x}})^T H(\underline{\hat{x}})(\underline{x} - \underline{\hat{x}}) = \kappa \tag{7}$$

In order to plot these fixed-energy contour curves in the two-dimensional parameter space, the following analysis is required. Consider the eigenvalues λ_1 and λ_2 and the eigenvectors $\underline{u}_1$ and $\underline{u}_2$ of the positive definite symmetric matrix $\hat{H} \equiv H(\underline{\hat{x}})$ 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, the eigenvalues are positive i.e. $\lambda_1 > 0$ and $\lambda_2 > 0$, while the eigenvectors $\underline{u}_1$ and $\underline{u}_2$ are orthogonal. Normalize that eigenvectors $\underline{u}_1$ and $\underline{u}_2$ so that they have unit length. These orthogonal unit vectors $\underline{u}_1$ and $\underline{u}_2$ define certain orthogonal directions in the parameter space (x_1, x_2) as shown in

Figure 1. A new coordinate system is introduced, centered at the best estimate $\underline{\hat{x}}$ with unit vectors along the axis of the new system to be the eigenvectors $\underline{u}_1$ and $\underline{u}_2$.

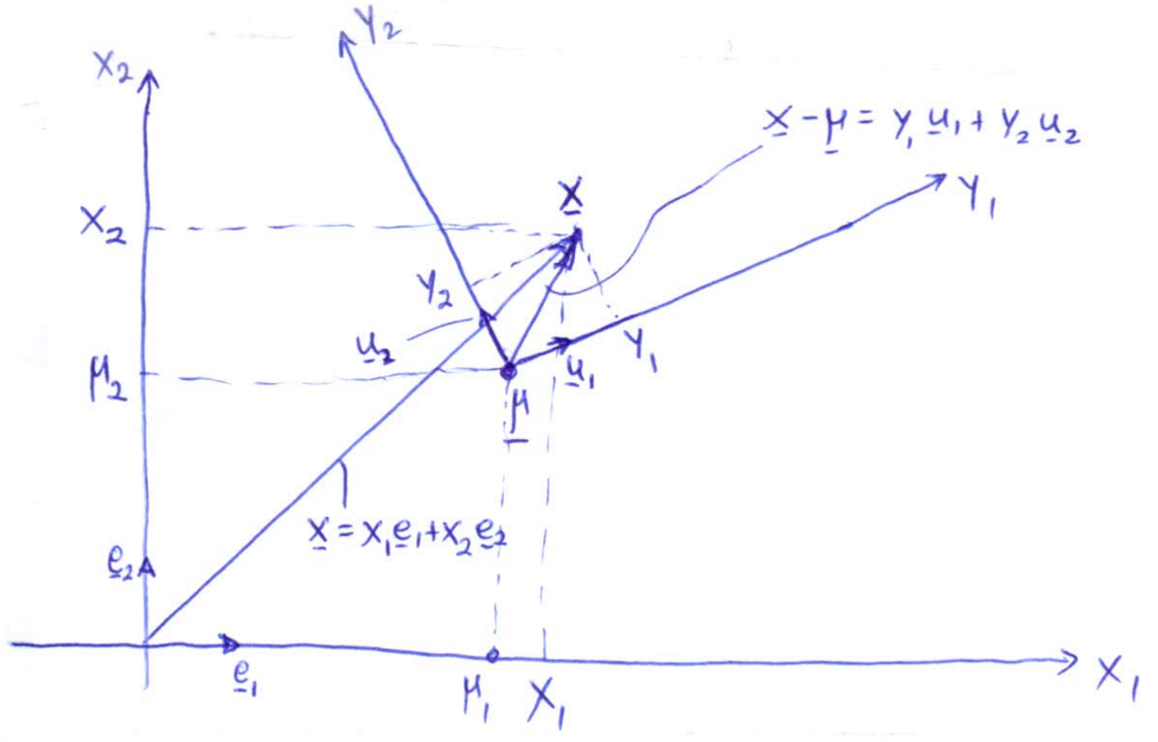

Figure 1: New coordinate system defined by the orthogonal unit vectors $\underline{u}_1$ and $\underline{u}_2$, and the coordinates y_1 and y_2 of the vector $\underline{x} - \underline{\mu}$ with respect to the new coordinate system.

Introducing now the matrix of eigenvectors $U = [\underline{u}_1, \underline{u}_2]$ 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 $\underline{u}_1$ and $\underline{u}_2$ constitute a basis of the two-dimensional vector space or, equivalently, any vector $\underline{x}-\underline{\hat{x}} \in R^2$ in Figure 1 can be written in terms of the basis unit vectors $\{\underline{u}_1, \underline{u}_2\}$ in the new coordinate system as

$$\underline{x}-\underline{\hat{x}} = y_1\underline{u}_1 + y_2\underline{u}_2 = [\underline{u}_1 \quad \underline{u}_2]\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = U\underline{y} \tag{8}$$

where $\underline{y}=(y_1, y_2)^T \in R^2$ are the components of the vector $\underline{x}-\underline{\hat{x}}$ with respect to the new coordinate system defined by the orthogonal unit vectors $\{\underline{u}_1, \underline{u}_2\}$.

Substituting $\underline{x}-\underline{\hat{x}} = U\underline{y}$ into the quadratic form (5), one derives the quadratic form $Q(\underline{x})$ in terms of the new coordinates y_1, y_2 of the vector $\underline{x}-\underline{\mu}$ in the new coordinate system as

$$Q(\underline{x}) = \underline{y}^T U^T \hat{H} U \underline{y} = \underline{y}^T \Lambda \underline{y} = \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \lambda_1 y_1^2 + \lambda_2 y_2^2 \tag{9}$$

Consider now the points at the contour curve of the function $Q(\underline{x})$ corresponding to the “energy” level κ, satisfying the equation (7). Using (9), the points on the contour curve can conveniently be written with respect to their coordinates y_1, y_2 in the new system defined by the eigenvector basis as follows

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = \kappa$$

Introducing the variables $\alpha_i = \sqrt{\dfrac{\kappa}{\lambda_i}}$, $i=1,2$, this equation can be re-written in the form

$$\frac{y_1^2}{\alpha_1^2} + \frac{y_2^2}{\alpha_2^2} = 1$$

which represents an ellipse with respect to the new coordinate system (see Figure 2 for a geometric representation of the contour curves), centered at the point $\underline{\hat{x}}$ in the parameter space with principal axis along the directions specified by the eigenvectors $\underline{u}_1$ and $\underline{u}_2$. The sizes of the principal axes of the ellipse are equal to α_1 and α_2. It is clear that the lengths 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 ellipse in the two-dimensional space. It should be noted that the contour curve specifies the spread of the uncertainty in the values of the parameters x_1 and x_2 in the two-dimensional parameter space.

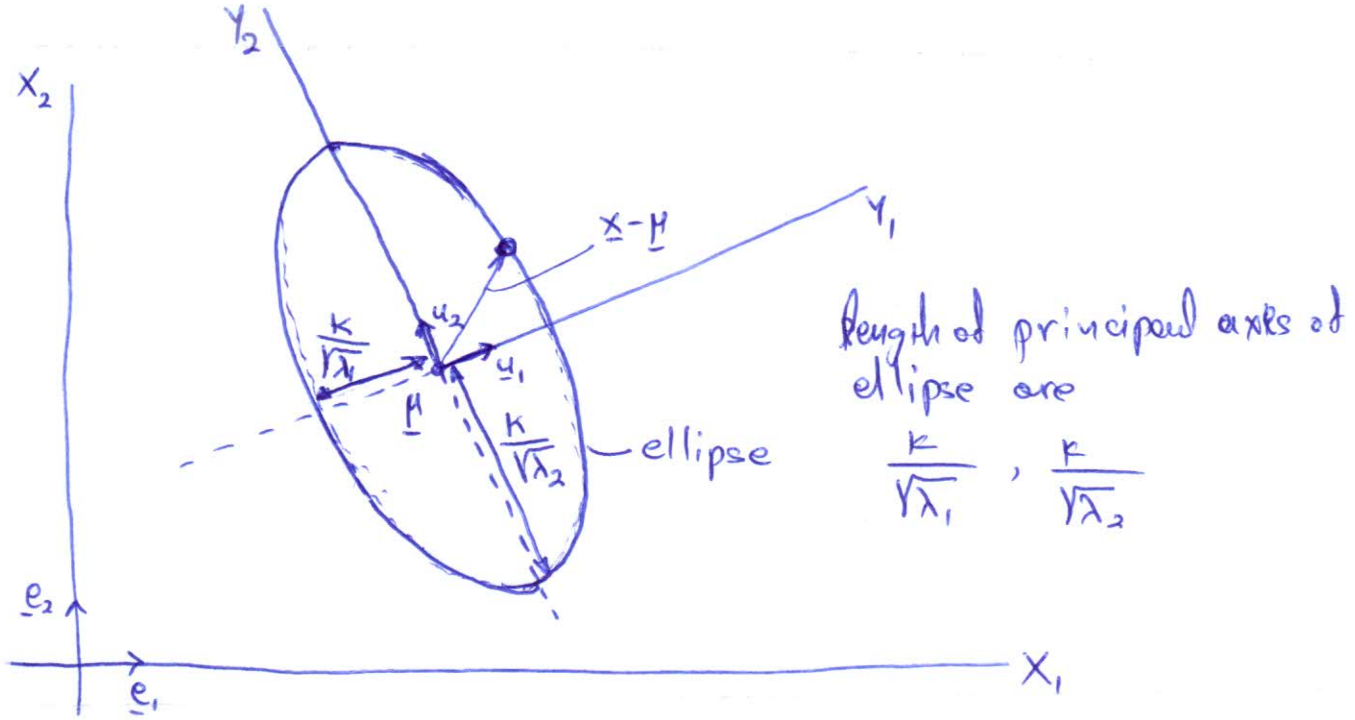

Figure 2: Contour plots $f(x_1, x_2) = c$

<u>Asymptotic Approximation of Posterior PDF</u>: Substituting the Taylor series expansion (4) into the posterior PDF (6) and keeping only up to the quadratic terms in the Taylor expansion, the posterior PDF is approximated by

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

Introducing the covariance matrix

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

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

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

<u>Remark 1</u>: Bayesian Central Limit Theorem

It can be shown that asymptotically, for large number of data, the posterior PDF tends to a the Gaussian distribution (10), centered at its most probable value and with covariance matrix equal to the inverse of the Hessian of the minus the logarithm of the posterior PDF, evaluated at the most probable value. The error of the asymptotic approximation is of order of N^{-1} where N denotes the number of data.

<u>Remark 2</u>: Spread of Uncertainty about the Best Estimate

The spread of the uncertainty in the parameters around the best estimate $\hat{\underline{x}}$ is completely defined by the Hessian matrix $\hat{H} = H(\hat{\underline{x}})$ or equivalently by the covariance matrix $C = H^{-1}(\hat{\underline{x}})$.

<u>Remark 3</u>: Marginal Distribution of x_1 or x_2

If our interest is to compute the uncertainty in the value of x_1 we need to integrate out the value of x_2 using the marginalization theorem

$$p(x_1 \mid D, I) = \int p(x_1, x_2 \mid D, I)\ dx_2 \tag{11}$$

For the two-parameter case and a general posterior PDF, the integral (11) is an one-dimensional integral that for each value of x_1 can be carried out numerically to yield the marginal posterior PDF $p(x_1 \mid D, I)$. Using, however, the asymptotic Gaussian approximation of the joint posterior PDF $p(x_1, x_2 \mid D, I)$ defined in (10), one can readily obtain that the marginal PDF $p(x_1 \mid D, I)$ is also Gaussian distribution with mean $\hat{x}_1$ and variance $C_{11} = H_{11} / \det H$, the (1,1) diagonal component of the covariance matrix C. The best estimate of x_1 is $\hat{x}_1$ and the spread of the uncertainty in the parameter x_1 about the best estimate is defined by $\sqrt{C_{11}}$.

A similar results holds for the marginal distribution of x_2.

However, the estimates $\sqrt{C_{11}}$ and $\sqrt{C_{22}}$ give an incomplete picture of the uncertainties in both x_1 and x_2 since they do not take into account the correlation between the variables x_1 and x_2.

<u>Remark 4</u>: Linear Transformation

Using the linear transformation of variables (8), that is

$$\underline{x} - \hat{\underline{x}} = [\underline{u}_1 \quad \underline{u}_2] \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = U\underline{y}$$

the fact that asymptotically the variables in x are Gaussian and that a linear transformation of Gaussian variables results in Gaussian variables as well, the posterior PDF for the new variables $\underline{y} = U^{-1}(\underline{x} - \hat{\underline{x}})$ are also Gaussian with mean

$$E[\underline{y}] = U^{-1}(E[\underline{x}] - \hat{\underline{x}}) = \underline{0}$$

and covariance matrix

$$E[\underline{y}\underline{y}^T] = U^{-1}E[(\underline{x} - \hat{\underline{x}})(\underline{x} - \hat{\underline{x}})^T]U^{-T} = U^{-1}CU^{-T} = U^{-1}\hat{H}^{-1}U^{-T} = (U^T\hat{H}\ U)^{-1} = \Lambda^{-1}$$

which is diagonal. The new variables in the vector $\underline{y}$ are thus uncorrelated and the posterior PDF of $\underline{y}$ follows a zero-mean Gaussian distribution with diagonal covariance matrix Λ^{-1} and variances inversely proportional to the eigenvalues λ_1 and λ_2 of the Hessian matrix $\hat{H}$. Specifically, the posterior PDF of $\underline{y}$ is given by

$$p(\underline{y} \mid D, I) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{1/(\lambda_1\lambda_2)}} \exp\left[-\frac{1}{2}\underline{y}^T\Lambda\underline{y}\right] = \prod_{k=1}^{2} \frac{1}{\sqrt{2\pi}\sqrt{1/\lambda_k}} \exp\left[-\frac{y_k^2}{2(1/\lambda_k)}\right]$$

The spread of the uncertainty in the parameters y_1 or y_2 along the directions defined by the unit eigenvectors $\underline{u}_1$ and $\underline{u}_2$ are inversely proportional to the square root of the eigenvalues λ_1 and λ_2. The variables $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$ provide the spread of the uncertainties of the variables y_1 and y_2. Moreover, the variables $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$ give a complete picture of the spread of the uncertainties in the parameter space (x_1, x_2), locally around the best estimate $\hat{\underline{x}}$, in the directions specified by the eigenvectors $\underline{u}_1$ and $\underline{u}_2$ of the Hessian matrix $\hat{H} = H(\hat{\underline{x}})$.

APPENDIX: Marginal and Conditional Distributions of Jointly Gaussian Variables

Theorem 1:

Consider a vector $\underline{x} \in R^n$ which has a Gaussian distribution with mean $\underline{\mu} \in R^n$ and covariance matrix $\Sigma \in R^{n \times n}$:

$$f(\underline{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) \right]$$

Let a partition of the random vector $\underline{x} \in R^n$ be

$$\underline{x} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix}$$

where $\underline{x}_1 \in R^{n_1}$ and $\underline{x}_2 \in R^{n_2}$, $n_1 + n_2 = 1$, and let the corresponding partitions of the mean and the covariance matrix be

$$\underline{\mu} = \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

A. Marginal Distributions

The marginal distributions of the random vector $\underline{x}_i$, $i = 1, \ldots, n$ is normal with mean $\underline{\mu}_i$ and covariance matrix Σ_{ii}, that is,

$$f(\underline{x}_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_{ii}|^{n/2}} \exp\left[-\frac{1}{2} (\underline{x}_i - \underline{\mu}_i)^T \Sigma_{ii}^{-1} (\underline{x}_i - \underline{\mu}_i) \right]$$

B. Conditional Distributions

The conditional distribution of $\underline{x}_i$ given $\underline{x}_j$ is normal with mean

$$\underline{\mu}_{i|j} = \underline{\mu}_i + \Sigma_{ij} \Sigma_{jj}^{-1} (\underline{x}_j - \underline{\mu}_j)$$

and covariance matrix

$$\Sigma_{i|j} = \Sigma_{jj} - \Sigma_{ij}^T \Sigma_{ii}^{-1} \Sigma_{ij}$$