BAYESIAN METHODS

In many imaging situations—for instance, image recording by film—the observation model is nonlinear of the form

$$\mathbf{v} = f(\mathcal{H}\mathbf{u}) + \boldsymbol{\eta} \tag{8.210}$$

where $f(x)$ is a nonlinear function of x. The a posteriori conditional density given by Bayes' rule

$$p(\mathbf{u}|\mathbf{v}) = \frac{p(\mathbf{v}|\mathbf{u})p(\mathbf{u})}{p(\mathbf{v})} \tag{8.211}$$

is useful in finding different types of estimates of the random vector $\mathbf{u}$ from the observation vector $\mathbf{v}$. The minimum mean square estimate (MMSE) of $\mathbf{u}$ is the mean of this density. The *maximum a posteriori* (MAP) and the *maximum likelihood* (ML) estimates are the modes of $p(\mathbf{u}|\mathbf{v})$ and $p(\mathbf{v}|\mathbf{u})$, respectively. When the observation model is nonlinear, it is difficult to obtain the marginal density $p(\mathbf{v})$ even when $\mathbf{u}$ and $\boldsymbol{\eta}$ are Gaussian. (In the linear case $p(\mathbf{u}|\mathbf{v})$ is easily obtained since it is Gaussian if $\mathbf{u}$ and $\boldsymbol{\eta}$ are). However, the MAP and ML estimates do not require $p(\mathbf{v})$ and are therefore easier to obtain.

Under the assumption of Gaussian statistics for $\mathbf{u}$ and $\boldsymbol{\eta}$, with covariances $\mathcal{R}_u$ and $\mathcal{R}_n$, respectively, the ML and MAP estimates can be shown to be the solution of the following equations:

ML estimate, $\hat{\mathbf{u}}_{ML}$: $\quad \mathcal{H}^T \mathcal{D} \mathcal{R}_n^{-1}[\mathbf{v} - f(\mathcal{H}\hat{\mathbf{u}}_{ML})] = \mathbf{0} \qquad (8.212)$

where

$$\mathcal{D} \triangleq \mathrm{Diag}\left\{\frac{\partial f(x)}{\partial x}\bigg|_{x=\hat{w}_i}\right\} \tag{8.213}$$

and $\hat{w}_i$ are the elements of the vector $\hat{\mathbf{w}} \triangleq \mathcal{H}\hat{\mathbf{u}}_{ML}$.

MAP estimate, $\hat{\mathbf{u}}_{MAP}$: $\quad \hat{\mathbf{u}}_{MAP} = \boldsymbol{\mu}_u + \mathcal{R}_u \mathcal{H}^T \mathcal{D} \mathcal{R}_n^{-1}[\mathbf{v} - f(\mathcal{H}\hat{\mathbf{u}}_{MAP})] \qquad (8.214)$

where $\boldsymbol{\mu}_u$ is the mean of $\mathbf{u}$ and $\mathcal{D}$ is defined in (8.213) but now $\hat{\mathbf{w}} \triangleq \mathcal{H}\hat{\mathbf{u}}_{MAP}$.

Since these equations are nonlinear, an alternative is to maximize the appropriate log densities. For example, a gradient algorithm for $\hat{\mathbf{u}}_{MAP}$ is

$$\hat{\mathbf{u}}_{j+1} = \hat{\mathbf{u}}_j - \alpha_j\{\mathcal{H}^T \mathcal{D}_j \mathcal{R}_n^{-1}[\mathbf{v} - f(\mathcal{H}\hat{\mathbf{u}}_j)] - \mathcal{R}_u^{-1}[\hat{\mathbf{u}}_j - \boldsymbol{\mu}_u]\} \tag{8.215}$$

where $\alpha_j > 0$, and $\mathcal{D}_j$ is evaluated at $\hat{\mathbf{w}}_j \triangleq \mathcal{H}\hat{\mathbf{u}}_j$.

Remarks

If the function $f(x)$ is linear, say $f(x) = x$, and $\mathcal{R}_n = \sigma_n^2 \mathbf{I}$, then $\hat{\mathbf{u}}_{ML}$ reduces to the least squares solution

$$\mathcal{H}^T \mathcal{H}\hat{\mathbf{u}}_{ML} = \mathcal{H}^T \mathbf{v} \tag{8.216}$$

and the MAP estimate reduces to the Wiener filter output for zero mean noise [see (8.87)],

$$\hat{\mathbf{u}}_{MAP} = \boldsymbol{\mu}_u + \mathcal{G}(\mathbf{v} - \boldsymbol{\mu}_v) \tag{8.217}$$

where $\mathcal{G} = (\mathcal{R}_u^{-1} + \mathcal{H}^T \mathcal{R}_n^{-1} \mathcal{H})^{-1} \mathcal{H}^T \mathcal{R}_n^{-1}$.

In practice, $\boldsymbol{\mu}_v$ may be estimated as a local average of $\mathbf{v}$ and $\boldsymbol{\mu}_u \simeq \mathcal{H}^+ f^{-1}(\boldsymbol{\mu}_v)$, where $\mathcal{H}^+$ is the generalized inverse of $\mathcal{H}$.