付　　録 A

第 5.5 節に対する補足説明

第 5.5 節では凸（凹）関数の性質を用いてスパースベイズ推定の更新式を導いた．この第 5.5 節の説明では，未知量 $\boldsymbol{x}$ を補助変数として取り扱うなどのため，導出は若干わかりにくいかもしれない．この補足資料では，凸（凹）関数の性質を用いて，更新式の少し異なった（望むべくは少々わかりやすい）導出を行う．

A.1　周辺尤度 (5.20) の自由エネルギーを用いた導出

ここでの説明は，式 (5.20) に示された周辺尤度から出発する．まず，ここでは式 (5.20) を自由エネルギーから導いてみよう．

評価関数 $\mathcal{F}[q,\boldsymbol{\alpha}]$：

$$\mathcal{F}[q,\boldsymbol{\alpha}] = \int d\boldsymbol{x} q(\boldsymbol{x})[\log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha}) - \log q(\boldsymbol{x})] \tag{A.1}$$

は自由エネルギーと呼ばれる．この $\mathcal{F}[q,\boldsymbol{\alpha}]$ は 2 つの量，ハイパーパラメータ $\boldsymbol{\alpha}$ と任意の確率分布 $q(\boldsymbol{x})$ の関数である．第 4.6 節の議論によれば，自由エネルギー $\mathcal{F}[q,\boldsymbol{\alpha}]$ を最大にする確率分布は未知量 $\boldsymbol{x}$ の（$\boldsymbol{\alpha}$ をこの時点での値に固定した）事後分布 $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})$ であることを示している．すなわち，自由エネルギーを $q(\boldsymbol{x})$ について最大にすることは EM アルゴリズム

の E ステップに対応する．E ステップ終了時点での自由エネルギーは

$$
\begin{aligned}
\mathcal{F}[p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}),\boldsymbol{\alpha}] &= \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})[\log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha}) - \log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})] \\
&= \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}) \log \frac{p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha})}{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})} = \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}) \log p(\boldsymbol{y}|\boldsymbol{\alpha}) \\
&= \log p(\boldsymbol{y}|\boldsymbol{\alpha}) \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}) = \log p(\boldsymbol{y}|\boldsymbol{\alpha}) \quad \text{(A.2)}
\end{aligned}
$$

となり，自由エネルギーの値は周辺尤度 $\log p(\boldsymbol{y}|\boldsymbol{\alpha})$ に等しい．したがって，

$$
\begin{aligned}
\log p(\boldsymbol{y}|\boldsymbol{\alpha}) &= \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}) \left[\log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha}) - \log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})\right] \\
&= E\left[\log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha}) - \log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})\right] \\
&= E\left[\log p(\boldsymbol{y}|\boldsymbol{x}) + \log p(\boldsymbol{x}|\boldsymbol{\alpha})\right] + \mathcal{H}[p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})] \quad \text{(A.3)}
\end{aligned}
$$

が成立する．ここで，$E[\cdot]$ は事後分布 $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})$ で平均を取ることを意味し，$\mathcal{H}[p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})]$ は事後分布のエントロピーである．

式 (A.3) に

$$
\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{\alpha}) &= \mathcal{N}(\boldsymbol{x}|\boldsymbol{0},\boldsymbol{\Phi}^{-1}) \\
p(\boldsymbol{y}|\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{y}|\boldsymbol{H}\boldsymbol{x},\beta^{-1}\boldsymbol{I}) \\
p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}) &= \mathcal{N}(\boldsymbol{x}|\bar{\boldsymbol{x}},\boldsymbol{\Gamma}^{-1})
\end{aligned}
$$

を代入して整理する．ただしここで，

$$
\boldsymbol{\Phi} = \mathrm{diag}\left[\alpha_1,\ldots,\alpha_N\right]
$$

である．すると，定数を無視し，$\mathcal{H}[p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})] = -\log|\boldsymbol{\Gamma}|$ を用いて，

$$
\log p(\boldsymbol{y}|\boldsymbol{\alpha}) = \frac{1}{2}\left[|\boldsymbol{\Phi}| + |\beta\boldsymbol{I}| - |\boldsymbol{\Gamma}|\right] - \frac{1}{2}E\left[\beta\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2\right] - \frac{1}{2}E\left[\boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{x}\right] \quad \text{(A.4)}
$$

を得る．さらに，

$$
\begin{aligned}
&E\left[\beta\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2\right]+E\left[\boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{x}\right] \\
&\qquad =\beta\left[\boldsymbol{y}^T\boldsymbol{y}-2\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{y}\right]+E\left[\boldsymbol{x}^T(\beta\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{\Phi})\boldsymbol{x}\right] \qquad \text{(A.5)}
\end{aligned}
$$

であり，ここで，

$$
\begin{aligned}
E\left[\boldsymbol{x}^T(\beta\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{\Phi})\boldsymbol{x}\right] &= E\left[\boldsymbol{x}^T\boldsymbol{\Gamma}\boldsymbol{x}\right]=E\left[\mathrm{tr}\left(\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{\Gamma}\right)\right] \\
&= E\left[\mathrm{tr}\left(\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{\Gamma}\right)\right]=\mathrm{tr}\left[E\left(\boldsymbol{x}\boldsymbol{x}^T\right)\boldsymbol{\Gamma}\right] \\
&= \mathrm{tr}\left[\left(\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T+\boldsymbol{\Gamma}^{-1}\right)\boldsymbol{\Gamma}\right]=\bar{\boldsymbol{x}}^T(\beta\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{\Phi})\bar{\boldsymbol{x}} \qquad \text{(A.6)}
\end{aligned}
$$

であるので，結局，

$$
\begin{aligned}
&E\left[\beta\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2\right]+E\left[\boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{x}\right] \\
&\qquad =\beta\left[\boldsymbol{y}^T\boldsymbol{y}-2\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{y}+\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{H}\bar{\boldsymbol{x}}\right]+\bar{\boldsymbol{x}}^T\boldsymbol{\Phi}\bar{\boldsymbol{x}} \\
&\qquad =\beta\|\boldsymbol{y}-\boldsymbol{H}\bar{\boldsymbol{x}}\|^2+\bar{\boldsymbol{x}}^T\boldsymbol{\Phi}\bar{\boldsymbol{x}} \qquad \text{(A.7)}
\end{aligned}
$$

を得る．したがって，式 (5.26)：

$$
\log|\boldsymbol{\Sigma}_y|=-\left(|\boldsymbol{\Phi}|+|\beta\boldsymbol{I}|-|\boldsymbol{\Gamma}|\right)
$$

を用いれば結局，周辺尤度として式 (5.20)：

$$
\log p(\boldsymbol{y}|\boldsymbol{\alpha})=-\frac{1}{2}\log|\boldsymbol{\Sigma}_y|-\frac{1}{2}\left[\beta\|\boldsymbol{y}-\boldsymbol{H}\bar{\boldsymbol{x}}\|^2+\bar{\boldsymbol{x}}^T\boldsymbol{\Phi}\bar{\boldsymbol{x}}\right]
$$

を得る．ここで，

$$
\boldsymbol{\Sigma}_y=\beta^{-1}\boldsymbol{I}+\boldsymbol{H}\boldsymbol{\Upsilon}\boldsymbol{H}^T
$$

である．

A.2 凹関数の性質を用いたコスト関数の導出

周辺尤度の式 (5.20) からスタートするのであるが，精度 α_j に対応した分散を ν_j（$\nu_j = \alpha_j^{-1}$）とし，列ベクトル $\boldsymbol{\nu}$ を $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_N]$ と定義する．凹関数の性質を用いるため，精度ベクトル $\boldsymbol{\alpha}$ の代わりに分散ベクトル $\boldsymbol{\nu}$ を，精度行列 $\boldsymbol{\Phi}$ の代わりに共分散行列 $\boldsymbol{\Upsilon}$（$\boldsymbol{\Upsilon} = \boldsymbol{\Phi}^{-1}$）を用いる．したがって，式 (5.20) の周辺尤度は

$$\log p(\boldsymbol{y}|\boldsymbol{\nu}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_y| - \frac{1}{2}\left[\beta\|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \bar{\boldsymbol{x}}^T\boldsymbol{\Upsilon}^{-1}\bar{\boldsymbol{x}}\right] \tag{A.8}$$

と表される．

ここで，$\log|\boldsymbol{\Sigma}_y|$ は $\boldsymbol{\nu}$ に関して凹関数であり，補助変数 $\boldsymbol{z}$ を用いて，式 (5.41)，すなわち，

$$\boldsymbol{z}^T\boldsymbol{\nu} - z_0 \geq \log|\boldsymbol{\Sigma}_y|$$

が常に成立する．したがって，

$$\log p(\boldsymbol{y}|\boldsymbol{\nu}) \geq -\frac{1}{2}(\boldsymbol{z}^T\boldsymbol{\nu} - z_0) - \frac{1}{2}\left[\beta\|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \bar{\boldsymbol{x}}^T\boldsymbol{\Upsilon}^{-1}\bar{\boldsymbol{x}}\right] \tag{A.9}$$

であり，補助的コスト関数 $\widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z})$ を

$$\widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z}) = -\frac{1}{2}(\boldsymbol{z}^T\boldsymbol{\nu} - z_0) - \frac{1}{2}\left[\beta\|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \bar{\boldsymbol{x}}^T\boldsymbol{\Upsilon}^{-1}\bar{\boldsymbol{x}}\right]$$

と定義すれば，必ず，

$$\log p(\boldsymbol{y}|\boldsymbol{\nu}) \geq \widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z})$$

が成立するため，$\widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z})$ を増加させる $\boldsymbol{\nu}$ は周辺尤度 $\log p(\boldsymbol{y}|\boldsymbol{\nu})$ を増加させる．したがって，$\boldsymbol{\nu}$ の更新解は

$$\widehat{\boldsymbol{\nu}} = \underset{\boldsymbol{\nu}}{\operatorname{argmax}}\, \widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z}) \tag{A.10}$$

として求めることができる．

A.3 更新式の導出

ここで，補助変数 $\boldsymbol{z}$ の更新値 $\widehat{\boldsymbol{z}}$ は，式 (5.46) すなわち

$$\widehat{\boldsymbol{z}} = \frac{\partial}{\partial \boldsymbol{\nu}} \log |\boldsymbol{\Sigma}_y|$$

から求まり，5.5.2 項で示したように

$$\widehat{z}_j = \boldsymbol{h}_j^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{h}_j \tag{A.11}$$

となる．上式は式 (5.49) である．$\boldsymbol{\nu}$ の更新値 $\widehat{\boldsymbol{\nu}}$ は

$$\begin{aligned}\underset{\boldsymbol{\nu}}{\operatorname{argmax}}\, \widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z}) &= \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \left[\bar{\boldsymbol{x}}^T \boldsymbol{\Upsilon}^{-1} \bar{\boldsymbol{x}} + \boldsymbol{z}^T \boldsymbol{\nu} \right] \\ &= \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \sum_{j=1}^{N} \left[z_j \nu_j + \frac{\bar{x}_j^2}{\nu_j} \right] \end{aligned} \tag{A.12}$$

から求まる．ここで，

$$\frac{\partial}{\partial \nu_j} \sum_{j=1}^{N} \left[z_j \nu_j + \frac{\bar{x}_j^2}{\nu_j} \right] = z_j - \frac{\bar{x}_j^2}{\nu_j^2} = 0 \tag{A.13}$$

とおいて，式 (A.11) を用いることにより

$$\widehat{\nu}_j = \frac{|\bar{x}_j|}{\sqrt{z_j}} = \frac{|\bar{x}_j|}{\sqrt{\boldsymbol{h}_j^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{h}_j}} \tag{A.14}$$

を得る．上式が式 (5.53) である．$\bar{\boldsymbol{x}}$ は事後分布の平均であり，式 (5.10), (5.11) を用いて更新する．

A.4 L_2 ノルム正則化解に対する正則化パラメータ更新式の導出

以上の考え方を用いれば，L_2 正則化ミニマムノルム解の正則化パラメータに対する更新式を，凹関数の性質を用いて導くこともできる．確率モデ

ルは，全ボクセルに共通の分散 ν を用いて，

$$p(\boldsymbol{x}|\nu)=\mathcal{N}(\boldsymbol{x}|\boldsymbol{0},\nu\boldsymbol{I})$$
$$p(\boldsymbol{y}|\boldsymbol{x})=\mathcal{N}(\boldsymbol{y}|\boldsymbol{H}\boldsymbol{x},\beta^{-1}\boldsymbol{I})$$
$$p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha})=\mathcal{N}(\boldsymbol{x}|\bar{\boldsymbol{x}},\boldsymbol{\Gamma}^{-1})$$

であるので，周辺尤度は

$$\log p(\boldsymbol{y}|\nu)=-\frac{1}{2}\log|\boldsymbol{\Sigma}_y|-\frac{1}{2}\left[\beta\|\boldsymbol{y}-\boldsymbol{H}\bar{\boldsymbol{x}}\|^2+\nu^{-1}\|\bar{\boldsymbol{x}}\|^2\right] \tag{A.15}$$

を得る．ここで，

$$\boldsymbol{\Sigma}_y=\beta^{-1}\boldsymbol{I}+\nu\boldsymbol{H}\boldsymbol{H}^T \tag{A.16}$$

である．スカラー補助変数 z を用いて，

$$z\nu-z_0\geq\log|\boldsymbol{\Sigma}_y|=\log|\beta^{-1}\boldsymbol{I}+\nu\boldsymbol{H}\boldsymbol{H}^T|$$

が成立するので，補助的コスト関数を

$$\widetilde{\mathcal{F}}(\nu,z)=-\frac{1}{2}(z\nu-z_0)-\frac{1}{2}\left[\beta\|\boldsymbol{y}-\boldsymbol{H}\bar{\boldsymbol{x}}\|^2+\nu^{-1}\|\bar{\boldsymbol{x}}\|^2\right]$$

と定義すれば，必ず，

$$\log p(\boldsymbol{y}|\nu)\geq\widetilde{\mathcal{F}}(\nu,z)$$

が成立するため，ボクセル分散 ν の更新解は

$$\widehat{\nu}=\underset{\nu}{\mathrm{argmax}}\,\widetilde{\mathcal{F}}(\nu,z) \tag{A.17}$$

から求めることができる．

補助変数 z の更新値 $\widehat{z}$ は式 (5.46) すなわち

$$\widehat{z}=\frac{\partial}{\partial\nu}\log|\boldsymbol{\Sigma}_y|$$

から求まる．ここで，

$$\frac{\partial}{\partial\nu}\log|\boldsymbol{\Sigma}_y|=\mathrm{tr}\left[\boldsymbol{\Sigma}_y^{-1}\frac{\partial}{\partial\nu}\boldsymbol{\Sigma}_y\right] \tag{A.18}$$

であり,

$$\frac{\partial}{\partial \nu}\boldsymbol{\Sigma}_y = \frac{\partial}{\partial \nu}\left[\beta^{-1}\boldsymbol{I} + \nu \boldsymbol{H}\boldsymbol{H}^T\right] = \boldsymbol{H}\boldsymbol{H}^T \tag{A.19}$$

であるので，結局,

$$\widehat{z} = \mathrm{tr}\left[\boldsymbol{\Sigma}_y^{-1}\boldsymbol{H}\boldsymbol{H}^T\right] = \mathrm{tr}\left(\boldsymbol{H}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{H}\right) \tag{A.20}$$

を得る．

次に，ν の更新値 $\widehat{\nu}$ は,

$$\widehat{\nu} = \underset{\nu}{\mathrm{argmax}}\,\widetilde{\mathcal{F}}(\nu, z) = \underset{\nu}{\mathrm{argmin}}\left[z\nu + \nu^{-1}\|\bar{\boldsymbol{x}}\|^2\right] \tag{A.21}$$

から求めることができる．

$$\frac{\partial}{\partial \nu}\left[z\nu + \nu^{-1}\|\bar{\boldsymbol{x}}\|^2\right] = z - \nu^{-2}\|\bar{\boldsymbol{x}}\|^2 = 0$$

より，式 (A.20) を用い

$$\widehat{\nu} = \sqrt{\frac{\|\bar{\boldsymbol{x}}\|^2}{z}} = \sqrt{\frac{\|\bar{\boldsymbol{x}}\|^2}{\mathrm{tr}\left[\boldsymbol{H}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{H}\right]}} \tag{A.22}$$

を得る．すなわち，ν を上式を用いて更新し，β は式 (4.33) を用いて更新することにより，L_2 ノルム正則化解を正則化パラメータを更新しながら求めることができる．