本文書の著作権は株式会社シグナルアナリシスに帰属します.

スパースベイズ推定

株式会社　シグナルアナリシス　関原謙介

1　データモデル

本当は $x_1, x_2, \ldots, x_N$ を観測したいのだがこれらは直接観測できず，代わりに一群の観測データ $y_1, y_2, \ldots, y_M$ が得られる状況を考えよう．知りたい量 $x_1, x_2, \ldots, x_N$ と観測データ $y_1, y_2, \ldots, y_M$ を，次の列ベクトル $\boldsymbol{x}$ と $\boldsymbol{y}$ で表現する．すなわち，

$$\text{未知量:}\quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad \text{観測データ:}\quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} \tag{1}$$

と定義する．この $\boldsymbol{x}$ は未知量ベクトルあるいは解ベクトル，$\boldsymbol{y}$ は観測ベクトルあるいはデータベクトルと呼ばれる．ベクトル $\boldsymbol{x}$ の集合は解空間 (solution space)，ベクトル $\boldsymbol{y}$ の集合は観測空間 (observation space) と呼ばれることもある．

ベクトル $\boldsymbol{x}$ とベクトル $\boldsymbol{y}$ は線形な関係，

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} \tag{2}$$

で結ばれているとする．ここで $\boldsymbol{H}$ は未知量 $\boldsymbol{x}$ と観測結果 $\boldsymbol{y}$ を結びつける $M \times N$ の行列である．本書では未知量 $\boldsymbol{x}$ と観測データ $\boldsymbol{y}$ は共に実数であり，$\boldsymbol{H}$ も実数行列であるとする．さらに，観測結果には，信号成分に加法的にノイズ $\boldsymbol{\varepsilon}$ が重畳すると仮定すれば，観測データのモデルは

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} \tag{3}$$

となる．ノイズベクトル $\boldsymbol{\varepsilon}$ は M 次元の列ベクトルで，

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} \tag{4}$$

であり，j 番目の要素 ε_j は j 番目の観測データ y_j に重畳するノイズを表す．式 (3) で表される観測モデルを線形離散モデルと呼ぶ．

2　スパースベイズ推定：確率モデル

線形離散モデル

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

を仮定し，観測データ $\boldsymbol{y}$ から未知な量 $\boldsymbol{x}$ を求める問題を引き続き考えよう．事前分布として，$\boldsymbol{x}$ の各要素に対し独立で同一な正規分布，

$$x_j \sim \mathcal{N}(x_j|0, \alpha^{-1}) \tag{5}$$

を仮定して，この事前分布のもとで L_2 ノルム正則化ミニマムノルム解が導かれる．それでは，この事前分布を少しだけ変えて，分散 (あるいは精度) が各要素 x_j に固有の値を取る，すなわち，

$$x_j \sim \mathcal{N}(x_j|0, \alpha_j^{-1}) \tag{6}$$

としたらどのような解が得られるであろうか．実は，L_2 ノルム正則化ミニマムノルム解とは全く異なるスパースな解が求まるのである．事前分布として式 (6) を用いた $\boldsymbol{x}$ の推定法はスパースベイズ推定と呼ばれる．以下にこのスパースベイズ推定を説明する．

精度行列を用いて議論を進めることにして，事前確率分布とデータ尤度が，

$$p(\boldsymbol{x}|\boldsymbol{\Phi}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Phi}^{-1}) \tag{7}$$

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{H}\boldsymbol{x}, \beta^{-1}\boldsymbol{I}) \tag{8}$$

の確率モデルに従うとする．説明を簡単にするため，ノイズの精度行列は $\beta\boldsymbol{I}$ として，精度 β は既知であるとする．また，$\boldsymbol{\Phi}$ が事前分布の精度行列で，対角行列：

$$\boldsymbol{\Phi} = \begin{bmatrix} \alpha_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \alpha_N \end{bmatrix}$$

と仮定する．したがって，

$$p(\boldsymbol{x}|\boldsymbol{\Phi}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Phi}^{-1}) = \prod_{j=1}^{N} \mathcal{N}(x_j|0, \alpha_j^{-1}) \tag{9}$$

と表記することもできる．ここで，各 α_j がみな等しい，つまり，$\alpha_1 = \cdots = \alpha_N$ とすれば，L_2 正則化ミニマムノルム推定に等しくなる．ここで，各 α_j が皆等しいとの仮定を置かないと，ミニマムノルム解とは全く異なるスパースな解が求まる．

3　推定の定式化

まず，ベクトル $\boldsymbol{\alpha}$ を $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$ と定義して導入する．ここで，$\boldsymbol{\Phi} = \mathrm{diag}(\boldsymbol{\alpha})$ である．ここで，$\mathrm{diag}(\cdot)$ は括弧内に記載されたベクトルを対角成分に持つ行列を意味する．式 (9) は，あらためて，

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\alpha})^{-1}) = \prod_{j=1}^{N} \mathcal{N}(x_j|0, \alpha_j^{-1}) \tag{10}$$

とも書ける．ここで，推定すべき未知パラメータは，そもそもの未知量 $\boldsymbol{x}$ と，事前分布に現れるハイパーパラメータ $\boldsymbol{\alpha}$ である．観測データ $\boldsymbol{y}$ に重畳するノイズの分散 β^{-1} も一般的には未知の量であるが，ここでは既知と仮定する．

ここで，真にベイズ的な取り扱いをしようとすれば両方の未知パラメータに対する結合事後分布 $p(\boldsymbol{x}, \boldsymbol{\alpha}|\boldsymbol{y})$ を

$$p(\boldsymbol{x}, \boldsymbol{\alpha}|\boldsymbol{y}) = \frac{p(y|\boldsymbol{x}, \boldsymbol{\alpha})p(\boldsymbol{x}, \boldsymbol{\alpha})}{\iint p(y|\boldsymbol{x}, \boldsymbol{\alpha})p(\boldsymbol{x}, \boldsymbol{\alpha})d\boldsymbol{\alpha}d\boldsymbol{x}} \tag{11}$$

として求めたいのだが，分母の積分が計算できず，結合事後分布 $p(\boldsymbol{x}, \boldsymbol{\alpha}|\boldsymbol{y})$ を求めることができない．したがって，代わりに，

$$p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{y})d\boldsymbol{\alpha} \approx p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{\alpha}}) \tag{12}$$

から，事後分布 $p(\boldsymbol{x}|\boldsymbol{y})$ の近似解 $p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{\alpha}})$ を求め，これを最大とする $\boldsymbol{x}$ の MAP 推定解を求める．

ここで，事前分布と尤度が正規分布（式 (7) および (8)）の場合，事後分布 $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})$ も正規分布，

$$p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{x}|\bar{\boldsymbol{x}}, \boldsymbol{\Gamma}^{-1}) \tag{13}$$

で与えられる．ここで，精度行列と平均は

$$\boldsymbol{\Gamma} = \boldsymbol{\Phi} + \beta \boldsymbol{H}^T \boldsymbol{H} \tag{14}$$

$$\bar{\boldsymbol{x}} = \beta \boldsymbol{\Gamma}^{-1} \boldsymbol{H}^T \boldsymbol{y} \tag{15}$$

で与えられる．したがって，$\widehat{\boldsymbol{\alpha}}$ が求まっていれば，式 (14) および (15) の式において $\boldsymbol{\Phi} = \mathrm{diag}(\widehat{\boldsymbol{\alpha}})$ を用いて計算した事後分布が $p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{\alpha}})$ である．したがって，$p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{\alpha}})$ を最大とする $\boldsymbol{x}$，すなわち，$\boldsymbol{\Phi} = \mathrm{diag}(\widehat{\boldsymbol{\alpha}})$ として，式 (15) で与えられる $\bar{\boldsymbol{x}}$ が未知量 $\boldsymbol{x}$ の MAP（および MMSE）推定解である．

それでは，$\widehat{\boldsymbol{\alpha}}$ はどうやって求めることができるであろうか．もちろん，EM アルゴリズムを用いて求めることができるが，EM アルゴリズムは推定が劣決定 ($M \ll N$) の場合，収束が遅いという問題が知られている．したがって，ここでは周辺尤度 $p(\boldsymbol{y}|\boldsymbol{\alpha})$ を，EM アルゴリズムのように間接的にではなく，直接最大とする方法を説明する．

4 周辺尤度関数の導出

$\boldsymbol{\alpha}$ の推定解 $\widehat{\boldsymbol{\alpha}}$ は周辺尤度 $\log p(\boldsymbol{y}|\boldsymbol{\alpha})$ を最大とする $\boldsymbol{\alpha}$ として求める．まず周辺尤度関数を導出するために，

$$p(\boldsymbol{y}|\boldsymbol{\alpha}) = \int p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\alpha})d\boldsymbol{x} = \int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{\alpha})d\boldsymbol{x}$$

に

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \frac{|\boldsymbol{\Phi}|^{1/2}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{\Phi} \boldsymbol{x}\right] \tag{16}$$

$$p(\boldsymbol{y}|\boldsymbol{x}) = \left(\frac{\beta}{2\pi}\right)^{M/2} \exp\left[-\frac{\beta}{2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2\right] \tag{17}$$

を代入して整理する．すると，

$$p(\boldsymbol{y}|\boldsymbol{\alpha}) = \frac{|\boldsymbol{\Phi}|^{1/2}}{(2\pi)^{N/2}} \left(\frac{\beta}{2\pi}\right)^{M/2} \int \exp\left[-D(\boldsymbol{x})\right] d\boldsymbol{x} \tag{18}$$

を得る．ここで，

$$D(\boldsymbol{x}) = \frac{\beta}{2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{x} \tag{19}$$

である．

上式の $D(\boldsymbol{x})$ を書き直してみると，

$$D(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\left[\boldsymbol{\Phi} + \beta\boldsymbol{H}^T\boldsymbol{H}\right]\boldsymbol{x} - \beta\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y} + \mathcal{C} \tag{20}$$

と表すことができる [問題 **5.1**][1]．なお上式では $\boldsymbol{x}$ を含ま無い項は $\mathcal{C}$ で表した．この $D(\boldsymbol{x})$ は平方完成された

$$D(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right) + \xi \tag{21}$$

の形で表すことができる [問題 **5.2**]．ここで，右辺の ξ は $\boldsymbol{x}$ を含まない残りの項を表す．上式の $D(\boldsymbol{x})$ は $\boldsymbol{x} = \bar{\boldsymbol{x}}$ で最小値を取り，その最小値が ξ である．この ξ の値は式 (19) に $\boldsymbol{x} = \bar{\boldsymbol{x}}$ を代入することにより求まり，$\xi = D(\bar{\boldsymbol{x}})$ である．したがって結局，$D(\boldsymbol{x})$ は，

$$D(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right)^T\boldsymbol{\Gamma}^{-1}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right) + D(\bar{\boldsymbol{x}}) \tag{22}$$

と求まる．

この結果を用いて式 (18) 右辺の積分を実行してみよう．まず，

$$\int \exp\left[-D(\boldsymbol{x})\right] d\boldsymbol{x} = \exp\left[-D(\bar{\boldsymbol{x}})\right] \int \exp\left[-\frac{1}{2}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right)\right] d\boldsymbol{x}$$

が成り立つ．上式右辺の積分は，平均 $\bar{\boldsymbol{x}}$ で精度行列 $\boldsymbol{\Gamma}$ である正規分布の全積分を考慮すれば，

$$\int \exp\left[-\frac{1}{2}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right)\right] d\boldsymbol{x} = \frac{(2\pi)^{N/2}}{|\boldsymbol{\Gamma}|^{1/2}}$$

と表されるので，式 (18) にこれらの結果を代入して，

$$p(\boldsymbol{y}|\boldsymbol{\alpha}) = \frac{|\boldsymbol{\Phi}|^{1/2}}{(2\pi)^{N/2}} \left(\frac{\beta}{2\pi}\right)^{M/2} \exp\left[-D(\bar{\boldsymbol{x}})\right] \frac{(2\pi)^{N/2}}{|\boldsymbol{\Gamma}|^{1/2}} \tag{23}$$

を得る．さらに，上式両辺の対数を取って，定数項を無視すれば，

$$\log p(\boldsymbol{y}|\boldsymbol{\alpha}) = -\frac{1}{2}\log|\boldsymbol{\Gamma}| + \frac{1}{2}\log|\boldsymbol{\Phi}| + \frac{M}{2}\log\beta - D(\bar{\boldsymbol{x}}) \tag{24}$$

を得る．

[1]問題番号は「ベイズ信号処理」(共立出版）の問題番号を指す．

ここで，$D(\bar{\boldsymbol{x}})$ をさらに以下のように変形する．まず，

$$\begin{aligned} D(\bar{\boldsymbol{x}}) &= \frac{\beta}{2}\|\boldsymbol{y}-\boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \frac{1}{2}\bar{\boldsymbol{x}}^T\boldsymbol{\Phi}\bar{\boldsymbol{x}} \\ &= \frac{\beta}{2}\left(\boldsymbol{y}^T\boldsymbol{y} - 2\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{y} + \bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{H}\bar{\boldsymbol{x}}\right) + \frac{1}{2}\bar{\boldsymbol{x}}^T\boldsymbol{\Phi}\bar{\boldsymbol{x}} \\ &= \frac{1}{2}\left[\beta\boldsymbol{y}^T\boldsymbol{y} - 2\bar{\boldsymbol{x}}^T\beta\boldsymbol{H}^T\boldsymbol{y} + \bar{\boldsymbol{x}}^T(\boldsymbol{\Phi}+\beta\boldsymbol{H}^T\boldsymbol{H})\bar{\boldsymbol{x}}\right] \\ &= \frac{1}{2}\left[\beta\boldsymbol{y}^T\boldsymbol{y} - 2\bar{\boldsymbol{x}}^T\beta\boldsymbol{H}^T\boldsymbol{y} + \bar{\boldsymbol{x}}^T\boldsymbol{\Gamma}\bar{\boldsymbol{x}}\right] \end{aligned} \tag{25}$$

である．さらに，$\beta\boldsymbol{H}^T\boldsymbol{y} = \boldsymbol{\Gamma}\bar{\boldsymbol{x}}$ の関係を用いれば，最終的に

$$D(\bar{\boldsymbol{x}}) = \frac{1}{2}\boldsymbol{y}^T\left[\beta^{-1}\boldsymbol{I} + \boldsymbol{H}\boldsymbol{\Phi}^{-1}\boldsymbol{H}^T\right]^{-1}\boldsymbol{y} = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y} \tag{26}$$

となることを示すことができる [問題 **5.3**]．ここで，$\boldsymbol{\Sigma}_y$ はモデルデータ共分散行列と呼ばれ

$$\boldsymbol{\Sigma}_y = \beta^{-1}\boldsymbol{I} + \boldsymbol{H}\boldsymbol{\Phi}^{-1}\boldsymbol{H}^T \tag{27}$$

である．

ふたたび式 (24) に戻って $D(\bar{\boldsymbol{x}})$ 以外の部分を計算する．行列式に関して，

$$|\boldsymbol{\Phi}||\beta^{-1}\boldsymbol{I} + \boldsymbol{H}\boldsymbol{\Phi}^{-1}\boldsymbol{H}^T| = |\beta^{-1}\boldsymbol{I}||\boldsymbol{\Phi} + \beta\boldsymbol{H}^T\boldsymbol{H}| \tag{28}$$

が成り立つ．したがって，式 (27) および (14) を代入すれば，

$$|\boldsymbol{\Phi}||\boldsymbol{\Sigma}_y| = |\beta^{-1}\boldsymbol{I}||\boldsymbol{\Gamma}| \tag{29}$$

より，

$$\log|\boldsymbol{\Sigma}_y| = \log|\boldsymbol{\Gamma}| - M\log(\beta) - \log|\boldsymbol{\Phi}| \tag{30}$$

を得る．式 (26) および (30) を式 (24) に代入すれば，結局，周辺尤度 $\log p(\boldsymbol{y}|\boldsymbol{\alpha})$ として

$$\log p(\boldsymbol{y}|\boldsymbol{\alpha}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_y| - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y} \tag{31}$$

を得る．したがって，$\log p(\boldsymbol{y}|\boldsymbol{\alpha})$ を最大とする $\boldsymbol{\alpha}$，あるいは，

$$\mathcal{F}(\boldsymbol{\alpha}) = \log|\boldsymbol{\Sigma}_y| + \boldsymbol{y}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y} \tag{32}$$

として，$\mathcal{F}(\boldsymbol{\alpha})$ を最小とする $\boldsymbol{\alpha}$ として最適推定値 $\hat{\boldsymbol{\alpha}}$ を求める．すなわち，この $\mathcal{F}(\boldsymbol{\alpha})$ が $\boldsymbol{\alpha}$ を推定するためのコスト関数である．

5　周辺尤度 (31) の自由エネルギーを用いた導出

周辺尤度（式 (31)）は自由エネルギーを用いて導出でき，前節に述べたものより若干簡便に導出できる[2]．評価関数 $\mathcal{F}[q,\boldsymbol{\alpha}]$：

$$\mathcal{F}[q,\boldsymbol{\alpha}] = \int d\boldsymbol{x}\, q(\boldsymbol{x})[\log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha}) - \log q(\boldsymbol{x})] \tag{33}$$

[2]自由エネルギーについては「ベイズ信号処理」(共立出版）を参照のこと．

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

$$\begin{aligned}\mathcal{F}[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}), \boldsymbol{\alpha}] &= \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}) [\log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\alpha}) - \log p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})] \\ &= \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}) \log \frac{p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\alpha})}{p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})} = \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}) \log p(\boldsymbol{y}|\boldsymbol{\alpha}) \\ &= \log p(\boldsymbol{y}|\boldsymbol{\alpha}) \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}) = \log p(\boldsymbol{y}|\boldsymbol{\alpha}) \end{aligned} \tag{34}$$

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，

$$\begin{aligned}\log p(\boldsymbol{y}|\boldsymbol{\alpha}) &= \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}) \left[\log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\alpha}) - \log p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})\right] \\ &= E\left[\log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\alpha}) - \log p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})\right] \\ &= E\left[\log p(\boldsymbol{y}|\boldsymbol{x}) + \log p(\boldsymbol{x}|\boldsymbol{\alpha})\right] + \mathcal{H}[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})] \end{aligned} \tag{35}$$

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 (35) 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，

$$\log p(\boldsymbol{y}|\boldsymbol{\alpha}) = \frac{1}{2}\left[|\boldsymbol{\Phi}| + |\beta\boldsymbol{I}| - |\boldsymbol{\Gamma}|\right] - \frac{1}{2}E\left[\beta\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2\right] - \frac{1}{2}E\left[\boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{x}\right] \tag{36}$$

を得る．さらに，

$$E\left[\beta\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2\right] + E\left[\boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{x}\right] = \beta\left[\boldsymbol{y}^T\boldsymbol{y} - 2\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{y}\right] + E\left[\boldsymbol{x}^T(\beta\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{\Phi})\boldsymbol{x}\right] \tag{37}$$

であり，ここで，

$$\begin{aligned} E\left[\boldsymbol{x}^T(\beta\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{\Phi})\boldsymbol{x}\right] &= E\left[\boldsymbol{x}^T\boldsymbol{\Gamma}\boldsymbol{x}\right] = E\left[\mathrm{tr}\left(\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{\Gamma}\right)\right] \\ &= E\left[\mathrm{tr}\left(\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{\Gamma}\right)\right] = \mathrm{tr}\left[E\left(\boldsymbol{x}\boldsymbol{x}^T\right)\boldsymbol{\Gamma}\right] \\ &= \mathrm{tr}\left[\left(\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T + \boldsymbol{\Gamma}^{-1}\right)\boldsymbol{\Gamma}\right] = \bar{\boldsymbol{x}}^T(\beta\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{\Phi})\bar{\boldsymbol{x}} \end{aligned} \tag{38}$$

であるので，結局，

$$E\left[\beta\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2\right]+E\left[\boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{x}\right] \\ =\beta\left[\boldsymbol{y}^T\boldsymbol{y}-2\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{y}+\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{H}\bar{\boldsymbol{x}}\right]+\bar{\boldsymbol{x}}^T\boldsymbol{\Phi}\bar{\boldsymbol{x}} \\ =\beta\|\boldsymbol{y}-\boldsymbol{H}\bar{\boldsymbol{x}}\|^2+\bar{\boldsymbol{x}}^T\boldsymbol{\Phi}\bar{\boldsymbol{x}} \quad (39)$$

を得る．したがって，

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

を式 (36) に代入すれば，結局，周辺尤度として式 (31)：

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

を得る．ここで，

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

である．

6　ハイパーパラメータ $\boldsymbol{\alpha}$ の更新式

ハイパーパラメータ $\boldsymbol{\alpha}$ の推定値 $\widehat{\boldsymbol{\alpha}}$ は，式 (32) に示されるコスト関数 $\mathcal{F}(\boldsymbol{\alpha})$ を用いて，

$$\widehat{\boldsymbol{\alpha}}=\underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\,\mathcal{F}(\boldsymbol{\alpha})$$

として求める．この最小値を計算するため，

$$\frac{\partial\mathcal{F}(\boldsymbol{\alpha})}{\partial\alpha_j}=\frac{\partial}{\partial\alpha_j}\log|\boldsymbol{\Sigma}_y|+\frac{\partial}{\partial\alpha_j}\boldsymbol{y}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y} \quad (40)$$

を計算する．まず，右辺第一項は，再び式 (30) を用いて，

$$\frac{\partial}{\partial\alpha_j}\log|\boldsymbol{\Sigma}_y|=\frac{\partial}{\partial\alpha_j}\left[-M\log\beta-\log|\boldsymbol{\Phi}|+\log|\boldsymbol{\Gamma}|\right]=-\frac{\partial}{\partial\alpha_j}\log|\boldsymbol{\Phi}|+\frac{\partial}{\partial\alpha_j}\log|\boldsymbol{\Gamma}| \quad (41)$$

である．ここで，まず，

$$\frac{\partial}{\partial\alpha_j}\log|\boldsymbol{\Phi}|=\frac{\partial}{\partial\alpha_j}\sum_{j=1}^{N}\log\alpha_j=\alpha_j^{-1} \quad (42)$$

である．また，$\boldsymbol{\Omega}_{j,j}$ を (j,j) 要素のみ 1，他の要素はすべてゼロであるような $N\times N$ の行列として，

$$\frac{\partial}{\partial\alpha_j}\log|\boldsymbol{\Gamma}|=\operatorname{tr}\left[\boldsymbol{\Gamma}^{-1}\frac{\partial}{\partial\alpha_j}\boldsymbol{\Gamma}\right]=\operatorname{tr}\left[\boldsymbol{\Gamma}^{-1}\frac{\partial}{\partial\alpha_j}\left[\boldsymbol{\Phi}+\beta\boldsymbol{H}^T\boldsymbol{H}\right]\right] \\ =\operatorname{tr}\left[\boldsymbol{\Gamma}^{-1}\frac{\partial}{\partial\alpha_j}\boldsymbol{\Phi}\right]=\operatorname{tr}\left[\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}_{j,j}\right]=\Sigma_{j,j} \quad (43)$$

となる．ここで，$\Sigma_{j,j}$ は事後分布の共分散行列 $\boldsymbol{\Sigma}(=\boldsymbol{\Gamma}^{-1})$ の (j,j) 成分である．

次に，式 (40) の右辺第 2 項を微分する．まず，式 (25) および式 (26) を用いれば，

$$\frac{\partial}{\partial \alpha_j} \boldsymbol{y}^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{y} = \frac{\partial}{\partial \alpha_j} \left[\beta \|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \bar{\boldsymbol{x}}^T \boldsymbol{\Phi} \bar{\boldsymbol{x}} \right] = \bar{\boldsymbol{x}}^T \left[\frac{\partial}{\partial \alpha_j} \boldsymbol{\Phi} \right] \bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}^T \boldsymbol{\Omega}_{j,j} \bar{\boldsymbol{x}} = \bar{x}_j^2 \tag{44}$$

したがって，

$$\frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \alpha_j} = \Sigma_{j,j} - \alpha_j^{-1} + \bar{x}_j^2 = 0 \tag{45}$$

であるので，結局，

$$\widehat{\alpha}_j^{-1} = \Sigma_{j,j} + \bar{x}_j^2 \tag{46}$$

を得る．実は，上式は EM アルゴリズムによる更新式に等しい [**問題 5.5**]．

一方，式 (45) は

$$1 - \alpha_j \Sigma_{j,j} = \alpha_j \bar{x}_j^2 \tag{47}$$

と変形できる．一方，式 (14) は

$$\boldsymbol{I} - \boldsymbol{\Gamma}^{-1} \boldsymbol{\Phi} = \beta \boldsymbol{\Gamma}^{-1} \boldsymbol{H}^T \boldsymbol{H}$$

と変形できる．ここで，上式左辺の (j,j) 成分は $1 - \alpha_j \Sigma_{j,j}$ に等しいことに注目すれば，式 (47) は

$$\widehat{\alpha}_j = \frac{\left[\beta \boldsymbol{\Gamma}^{-1} \boldsymbol{H}^T \boldsymbol{H} \right]_{j,j}}{\bar{x}_j^2} \tag{48}$$

と表記することができる．ここで，上式右辺の $[\,\cdot\,]_{j,j}$ は，括弧内の行列の (j,j) 成分を取ることを意味する．式 (48) に示す更新式は MacKay によって提案されたもので，EM アルゴリズムによる更新式 (46) よりも，$M \ll N$ なる劣決定の状況において，収束が早いことが知られている．

式 (48) の右辺において，$\boldsymbol{\Gamma}$ にも α_j は含まれているのに，この α_j を無視しているように感じられるため，読者はここで妙に思うかもしれない．しかし，式 (46) において，$\Sigma_{j,j}$ も $\bar{x}_j^2$ も α_j に依存していて，単に，その依存性が明示的に示されていないだけである．したがって，式 (46) と (48) との違いは，式 (46) においては，$\Sigma_{j,j}$ と $\bar{x}_j^2$ を古い α_j の値を用いて計算し，式 (46) にしたがって α_j の値を更新するのに対して，式 (48) においては，$\boldsymbol{\Gamma}$ と $\bar{x}_j^2$ を古い α_j の値を用いて計算し，α_j の値を更新することである．上の議論から示唆されるように，更新式の導出にはいくつかの任意性があり，結果として得られた更新式のどれがより優れたものであるかを理論的にあらかじめ予測するのは難しく，実際の問題に適用してみて初めて分かるのが普通である．式 (48) に示す MacKay の更新式はこのように heuristic（発見的）に見いだされ，提案されたものである．式 (46) に示す EM アルゴリズムの更新式と異なり，更新ごとに周辺尤度を増加させる理論的な保障はない．

式 (48) に示すように，$\boldsymbol{\alpha}$ の更新には事後分布のパラメータ $\boldsymbol{\Gamma}$ と $\bar{\boldsymbol{x}}$ が必要である．したがって，このアルゴリズムはやはり再帰的なものとなる．まず，$\boldsymbol{\alpha}$ に適当な初期値を仮定して，事後分布のパラメータを式 (14) および (15) から計算する．これは EM アルゴリズムの E ステップと全く同じである．このステップから得られた事後分布のパラメータを用いて式 (48) よりハイパーパラメータ $\boldsymbol{\alpha}$ を更新する．このステップは更新式が異なっているが，EM アルゴリズムの M ステップに相当する．上記のステップを収束条件が満たされるまで繰り返す．収束の判定は，$\widehat{\boldsymbol{\Phi}} = \mathrm{diag}(\widehat{\boldsymbol{\alpha}})$ として，コスト関数 $\mathcal{F}(\widehat{\boldsymbol{\alpha}})$ を計算し，更新を繰り返してもほとんど減少しなくなったときに計算を打ち切る．

7　凹関数の性質を用いた更新式の導出

周辺尤度の式 (31) を増加させる別の更新式を導こう．この導出では，凹関数の性質を用いるため，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．精度ベクトル $\boldsymbol{\alpha}$ の代わりにこの分散ベクトル $\boldsymbol{\nu}$ 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 (31) の周辺尤度は

$$\log p(\boldsymbol{y}|\boldsymbol{\nu}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_y| - \frac{1}{2}\left[\beta\|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \bar{\boldsymbol{x}}^T\boldsymbol{\Upsilon}^{-1}\bar{\boldsymbol{x}}\right] \tag{49}$$

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，

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

が常に成立する[3]．したがって，

$$\log p(\boldsymbol{y}|\boldsymbol{\nu}) \geq -\frac{1}{2}(\boldsymbol{z}^T\boldsymbol{\nu} - z_0) - \frac{1}{2}\left[\beta\|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \bar{\boldsymbol{x}}^T\boldsymbol{\Upsilon}^{-1}\bar{\boldsymbol{x}}\right] \tag{50}$$

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．

補助変数 $\boldsymbol{z}$ の更新値 $\widehat{\boldsymbol{z}}$ は，

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

から求まり，「ベイズ信号処理」5.5.2 項に示すように

$$\widehat{z}_j = \boldsymbol{h}_j^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{h}_j \tag{51}$$

となる．

$\boldsymbol{\nu}$ の更新解は

$$\widehat{\boldsymbol{\nu}} = \operatorname*{argmax}_{\boldsymbol{\nu}} \widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z}) \tag{52}$$

として求めることができる．すなわち，$\boldsymbol{\nu}$ の更新値 $\widehat{\boldsymbol{\nu}}$ は

$$\operatorname*{argmax}_{\boldsymbol{\nu}} \widetilde{\mathcal{F}}(\boldsymbol{\nu}, \boldsymbol{z}) = \operatorname*{argmin}_{\boldsymbol{\nu}} \left[\bar{\boldsymbol{x}}^T\boldsymbol{\Upsilon}^{-1}\bar{\boldsymbol{x}} + \boldsymbol{z}^T\boldsymbol{\nu}\right] = \operatorname*{argmin}_{\boldsymbol{\nu}} \sum_{j=1}^{N}\left[z_j\nu_j + \frac{\bar{x}_j^2}{\nu_j}\right] \tag{53}$$

から求まる．したがって，

$$\frac{\partial}{\partial\nu_j}\sum_{j=1}^{N}\left[z_j\nu_j + \frac{\bar{x}_j^2}{\nu_j}\right] = z_j - \frac{\bar{x}_j^2}{\nu_j^2} = 0 \tag{54}$$

[3]説明については「ベイズ信号処理」（共立出版）を参照のこと．

とおいて，式 (51) を用いることにより

$$\widehat{\nu}_j = \frac{|\bar{x}_j|}{\sqrt{z_j}} = \frac{|\bar{x}_j|}{\sqrt{\boldsymbol{h}_j^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{h}_j}} \tag{55}$$

を得る．$\bar{\boldsymbol{x}}$ は事後分布の平均であり，式 (14)，(15) から求まる．

式 (55) に示すように，$\boldsymbol{\nu}$ の更新には事後分布のパラメータ $\bar{\boldsymbol{x}}$ が必要である．したがって，このアルゴリズムはやはり再帰的なものとなる．まず，$\boldsymbol{\nu}$ に適当な初期値を仮定して，事後分布のパラメータを式 (14) および (15) から計算する．これは EM アルゴリズムの E ステップと全く同じである．このステップから得られた事後分布のパラメータを用いて式 (55) よりハイパーパラメータ $\boldsymbol{\nu}$ を更新する．このステップは EM アルゴリズムの M ステップに相当する．上記のステップを収束条件が満たされるまで繰り返す．収束の判定はコスト関数 $\mathcal{F}(\widehat{\boldsymbol{\nu}})$ を計算し，更新を繰り返してもほとんど減少しなくなったときに計算を打ち切る．

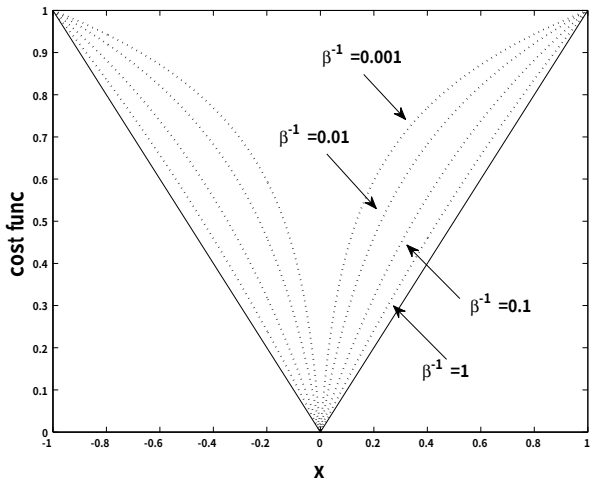

図 1: 式 (64) に示すコスト関数 $\varphi(x)$ の計算結果．$\varphi(\boldsymbol{x})$ のプロットを，4 種類の β^{-1} の値について点線で示す．また，比較のため L_1 ノルム解の制約条件 $|x|$ を実線で示す．それぞれのコスト関数は，$|x| = 1$ で値 1 を取るよう規格化して示す．

8　どうしてスパースな解が得られるのかについての考察

線形離散モデル $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \varepsilon$ もとで，事前確率分布を以下の正規分布，

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{j=1}^{N} \mathcal{N}(x_j|0, \alpha^{-1}) \tag{56}$$

を仮定すると，ベイズ推定解 $\bar{\boldsymbol{x}}$ はスパースな解，すなわち，$\bar{x}_1, \ldots, \bar{x}_N$ のなかで，ほとんどがのものがゼロに非常に近く，少数が明らかにノンゼロな値を持つという解が得られる．本節では，どうしてスパースな解が得られるのかについてコスト関数を用いて考察してみよう．

ここ精度 $\boldsymbol{\alpha}$ ではなく分散 $\boldsymbol{\nu}$ を用いて議論しよう[4]．コスト関数を変数 $\boldsymbol{\nu}$ を用いてもう 1 度表記してみ

[4] $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ および $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ であり，$\nu_j = 1/\alpha_j$ である．

ると，

$$\mathcal{F}(\boldsymbol{\nu}) = \log|\boldsymbol{\Sigma}_y| + \boldsymbol{y}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y}: \quad \boldsymbol{\Sigma}_y = \beta^{-1}\boldsymbol{I} + \boldsymbol{H}\boldsymbol{\Upsilon}\boldsymbol{H}^T \tag{57}$$

ただし，$\boldsymbol{\Upsilon} = \mathrm{diag}(\boldsymbol{\nu})$ である．したがって，$\boldsymbol{\nu}$ の更新値は

$$\widehat{\boldsymbol{\nu}} = \underset{\boldsymbol{\nu}}{\mathrm{argmin}} \left(\log|\boldsymbol{\Sigma}_y| + \boldsymbol{y}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y}\right)$$

と表すことができる．ここで，上式右辺の第 2 項は

$$\boldsymbol{y}^T\boldsymbol{\Sigma}_y\boldsymbol{y} = \min_{\boldsymbol{x}} \left[\beta\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \boldsymbol{x}^T\boldsymbol{\Upsilon}^{-1}\boldsymbol{x}\right] \tag{58}$$

と表すことができる．[**問題 5.6**]．したがって，これらの式を組み合わせると，$\boldsymbol{x}$ を求めるためのコスト関数として，

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\mathrm{argmin}}\, \mathcal{F}: \quad \mathcal{F} = \beta\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \phi(\boldsymbol{x}) \tag{59}$$

を導くことができる．ここで，制約条件 $\phi(\boldsymbol{x})$ は

$$\phi(\boldsymbol{x}) = \min_{\boldsymbol{\nu}} \left(\boldsymbol{x}^T\boldsymbol{\Upsilon}^{-1}\boldsymbol{x} + \log|\boldsymbol{\Sigma}_y|\right) = \min_{\boldsymbol{\nu}} \left(\sum_{j=1}^{N} \frac{x_j^2}{\nu_j} + \log|\boldsymbol{\Sigma}_y|\right) \tag{60}$$

と与えられる．

ここで，なぜスパースな解が得られるかを考察するため，この制約条件 $\phi(\boldsymbol{x})$ の特徴を調べてみよう．と言っても，この制約条件の式には $\boldsymbol{\nu}$ を含む $\log|\boldsymbol{\Sigma}_y|$ の項が存在するため，この $\phi(\boldsymbol{x})$ を計算するのは簡単ではない．そこで，少し話を簡単にするため，行列 $\boldsymbol{H}$ の列ベクトルが正規直交系を成す，すなわち，$\boldsymbol{h}_i^T\boldsymbol{h}_j = I_{i,j}$ が成り立つと仮定してみよう．この場合

$$\log|\boldsymbol{\Sigma}_y| = \sum_{j=1}^{N} \log(\beta^{-1} + \nu_j)$$

であるので，式 (60) を用いれば，

$$\phi(\boldsymbol{x}) = \min_{\boldsymbol{\nu}} \sum_{j=1}^{N} \left(\frac{x_j^2}{\nu_j} + \log(\beta^{-1} + \nu_j)\right) \tag{61}$$

を得る．したがって，

$$\phi(\boldsymbol{x}) = \sum_{j=1}^{N} \varphi(x_j) \tag{62}$$

と表すことができ，ここで，

$$\varphi(x_j) = \min_{\nu_j} \left(\frac{x_j^2}{\nu_j} + \log(\beta^{-1} + \nu_j)\right) \tag{63}$$

である．この $\varphi(x_j)$ は

$$\varphi(x_j) = \frac{2|x_j|}{\sqrt{x_j^2 + 4\beta^{-1}} + |x_j|} + \log\left(\beta^{-1} + \frac{x_j^2}{2} + \frac{1}{2}|x_j|\sqrt{x_j^2 + 4\beta^{-1}}\right) \tag{64}$$

と表すことができる [**問題 5.7**].

式 (64) に示すコスト関数 $\varphi(x)$ の計算結果を図 1 に示す．同図に $\varphi(\boldsymbol{x})$ のプロットを，$\beta^{-1} = 0.001$，$\beta^{-1} = 0.01$，$\beta^{-1} = 0.1$ および $\beta^{-1} = 1$ の 4 種類の場合について，点線で示す．比較のため，L_1 ノルムの場合の制約条件 $|x|$ のプロットを実線で示す．図 1 に示されるように，$\varphi(x)$ は L_1 ノルムにおける制約条件 $|x|$ に非常に類似した形をしており，スパースな解を生じることが理解できよう．さらに，$\varphi(x)$ が，β^{-1} すなわちノイズの分散に依存することも見て取れる．すなわち，ノイズ分散が小さい場合には，$\varphi(x)$ は，L_1 ノルム制約より，さらに急峻となり，L_0 ノルム制約により近づく．すなわち，スパースベイズ推定のコスト関数における制約条件は，固定のものではなく，ノイズの大きさによって急峻さを調節し，ノイズが小さな時には，よりスパースネスを強調した L_0 ノルム制約に近いものになる．反対にノイズが大きい場合には，$\varphi(x)$ は L_1 ノルム制約に近いものとなり，したがって，スパースネスはより穏やかなものに調節されることになる．