最小二乗法とベイズ推定

中野 慎也

「データ科学: 理論から実用へ」

7 September 2023

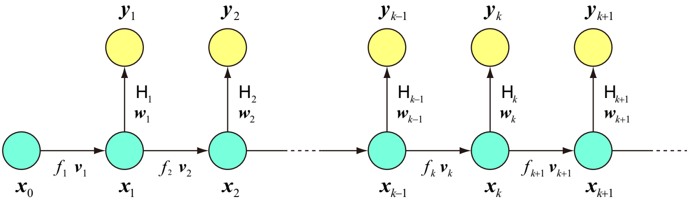
まず時間発展を考慮しない場合について

- データ同化でやりたいことは，

$$\boldsymbol{x}_k = \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{v}_k,$$
$$\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{w}_k$$

の 2 種類の関係式が与えられ，更に各時刻の観測データ $\boldsymbol{y}_k$ が与えられたという条件の下で，各時刻の $\boldsymbol{x}_k$ を推定することであった．

- しかし，まずはシステムモデル $\boldsymbol{x}_k = \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{v}_k$ は無視し，各時間ステップで，観測 $\boldsymbol{y}$ が与えられた下で $\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{w}_k$ という観測モデルに基づいて $\boldsymbol{x}$ を推定するという問題を考える．



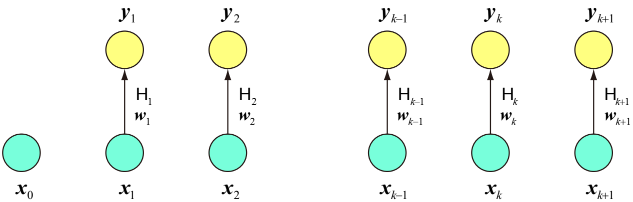
まず時間発展を考慮しない場合について

- データ同化でやりたいことは，

$$\boldsymbol{x}_k = \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{v}_k,$$
$$\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{w}_k$$

  の 2 種類の関係式が与えられ，更に各時刻の観測データ $\boldsymbol{y}_k$ が与えられたという条件の下で，各時刻の $\boldsymbol{x}_k$ を推定することであった．

- しかし，まずはシステムモデル $\boldsymbol{x}_k = \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{v}_k$ は無視し，各時間ステップで，観測 $\boldsymbol{y}$ が与えられた下で $\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{w}_k$ という観測モデルに基づいて $\boldsymbol{x}$ を推定するという問題を考える．



最小二乗法

- まずは，$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{w}$ において，$\|\boldsymbol{w}\|^2$ をなるべく小さくするという方針を考える．それはつまり，

$$\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$$

  を最小化する問題で，所謂最小二乗法である．
- データ数 n が未知変数の数 m よりも多い場合 $(n > m)$ には,「基本的に」これで x が決まる．

最小二乗法

$$\Psi_L = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m} H_{ij} x_j \right)^2 \tag{1}$$

とおくと，

$$\frac{\partial \Psi_L}{\partial x_k} = -2 \sum_{i=1}^{n} H_{ik} \left(y_i - \sum_{j=1}^{m} H_{ij} x_j \right) = 0 \tag{2}$$

なので，

$$-\mathbf{H}^\top \boldsymbol{y} + \mathbf{H}^\top \mathbf{H}\boldsymbol{x} = 0. \tag{3}$$

したがって，$\mathbf{H}^\top\mathbf{H}$ が正則 (逆行列を持つ) なら

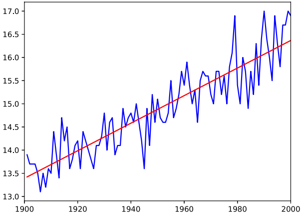
$$\hat{\boldsymbol{x}} = (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top \boldsymbol{y}. \tag{4}$$

最小二乗法の例 (直線あてはめ)

変数 z の列 $\{z_1, z_2, \ldots, z_n\}$ に対して，変数 y に関するデータ $\{y_1, y_2, \ldots, y_n\}$ が得られているとき，y を

$$y = a_0 + a_1 z + \varepsilon \tag{5}$$

という式で説明する．



$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, \quad \boldsymbol{w} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

とおくと，先程の $\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{w}$ の形になる．

最小二乗法の例 (直線あてはめ)

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, \quad \boldsymbol{w} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

に対して，推定式 $\hat{\boldsymbol{x}} = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \boldsymbol{y}$ を適用する．

$\bar{z} = \sum z_i / n$, $\bar{y} = \sum y_i / n$ として，

$$\mathbf{H}^\top \mathbf{H} = \begin{pmatrix} n & \sum z_i \\ \sum z_i & \sum z_i^2 \end{pmatrix} = \begin{pmatrix} n & n\bar{z} \\ n\bar{z} & \sum z_i^2 \end{pmatrix},$$

$$\mathbf{H}^\top \boldsymbol{y} = \begin{pmatrix} \sum y_i \\ \sum z_i y_i \end{pmatrix} = \begin{pmatrix} n\bar{y} \\ \sum z_i y_i \end{pmatrix}.$$

最小二乗法の例 (直線あてはめ)

$$
\begin{aligned}
\hat{\boldsymbol{x}} = (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\boldsymbol{y} &= \frac{1}{n\sum z_i^2 - n^2\bar{z}^2}\begin{pmatrix} \sum z_i^2 & -n\bar{z} \\ -n\bar{z} & n \end{pmatrix}\begin{pmatrix} n\bar{y} \\ \sum z_i y_i \end{pmatrix} \\
&= \frac{1}{n^2 C_{zz}}\begin{pmatrix} n\bar{y}\sum z_i^2 - n\bar{z}\sum z_i y_i \\ n\sum z_i y_i - n^2\bar{z}\bar{y} \end{pmatrix} \\
&= \frac{1}{n^2 C_{zz}}\begin{pmatrix} n\bar{y}\sum z_i^2 - n^2\bar{z}^2\bar{y} - n\bar{z}(\sum z_i y_i - n\bar{z}\bar{y}) \\ n^2 C_{zy} \end{pmatrix} \\
&= \frac{1}{C_{zz}}\begin{pmatrix} \bar{y}C_{zz} - \bar{z}C_{zy} \\ C_{zy} \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{z}(C_{zy}/C_{zz}) \\ C_{zy}/C_{zz} \end{pmatrix}
\end{aligned} \tag{6}
$$

但し，

$$
C_{zz} = \frac{1}{n}\sum z_i^2 - \bar{z}^2, \qquad C_{zy} = \frac{1}{n}\sum z_i y_i - \bar{z}\bar{y}.
$$

特異値分解による方法

$\Psi_L = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$ を最小化する x は，特異値分解を使って得ることもできる．

まずは，$\mathbf{H}$ を以下のように特異値分解する:

$$\mathbf{H} = \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\top = \sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^\top. \tag{7}$$

また，$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$, $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ はそれぞれ x, $\boldsymbol{y}$ が定義されるベクトル空間の正規直交基底なので，

$$\boldsymbol{x} = \sum_{i=1}^{m} \alpha_i \boldsymbol{v}_i, \qquad \boldsymbol{y} = \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i \tag{8}$$

と書くことができる．但し，各 i に対して $\alpha_i = \boldsymbol{v}_i^\top \boldsymbol{x}$, $\beta_i = \boldsymbol{u}_i^\top \boldsymbol{y}$．

これを使って，$\Psi_L = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$ を計算する．

特異値分解による方法

計算してみると，

$$
\begin{aligned}
\Psi_L = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 &= \left\| \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i - \left(\sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^\mathsf{T} \right) \left(\sum_{i=1}^{m} \alpha_i \boldsymbol{v}_i \right) \right\|^2 \\
&= \left\| \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i - \sum_{i=1}^{r} \lambda_i \alpha_i \boldsymbol{u}_i \right\|^2 \\
&= \left\| \sum_{i=1}^{r} (\beta_i - \lambda_i \alpha_i) \boldsymbol{u}_i + \sum_{i=r+1}^{n} \beta_i \boldsymbol{u}_i \right\|^2 \\
&= \sum_{i=1}^{r} (\beta_i - \lambda_i \alpha_i)^2 + \sum_{i=r+1}^{n} \beta_i^2 .
\end{aligned}
\tag{9}
$$

$r = m (= \dim \boldsymbol{x})$ の場合，$\alpha_i = \beta_i / \lambda_i \ (1 \le i \le r)$ とすればよく，

$$
\hat{\boldsymbol{x}} = \mathrm{argmin}_{\boldsymbol{x}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 = \sum_{i=1}^{m} \frac{\beta_i}{\lambda_i} \boldsymbol{v}_i . \tag{10}
$$

特異値分解による方法

先に得られた推定値 $\hat{x} = \left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{H}^\top\boldsymbol{y}$ は，$r = m \leq n$ のとき

$$\left(\mathbf{H}^\top\mathbf{H}\right)^{-1} = \left(\mathbf{V\Lambda U}^\top\mathbf{U\Lambda V}^\top\right)^{-1} = \left(\mathbf{V\Lambda}^2\mathbf{V}^\top\right)^{-1} = \left(\mathbf{V\Lambda}^{-2}\mathbf{V}^\top\right) = \sum_{i=1}^{m}\frac{1}{\lambda_i^2}\boldsymbol{v}_i\boldsymbol{v}_i^\top \tag{11}$$

となり，

$$\mathbf{H}^\top\boldsymbol{y} = \left(\sum_{i=1}^{m}\lambda_i\boldsymbol{v}_i\boldsymbol{u}_i^\top\right)\left(\sum_{j=1}^{m}\beta_j\boldsymbol{u}_i\right) = \sum_{i=1}^{m}\lambda_i\beta_i\boldsymbol{v}_i \tag{12}$$

となることから，

$$\left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{H}^\top\boldsymbol{y} = \left(\sum_{i=1}^{m}\frac{1}{\lambda_i^2}\boldsymbol{v}_i\boldsymbol{v}_i^\top\right)\left(\sum_{i=1}^{m}\lambda_i\beta_i\boldsymbol{v}_i\right) = \sum_{i=1}^{m}\frac{\beta_i}{\lambda_i}\boldsymbol{v}_i. \tag{13}$$

つまり，特異値分解で求めた式と等価であることが分かる．
なお，$\mathbf{H}^\top\mathbf{H}$ が逆行列を持つという条件は，$r = m$ という条件と対応している．

ムーア・ペンローズ一般逆行列

$$\hat{\boldsymbol{x}} = (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\boldsymbol{y}. \tag{14}$$

の $(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}$ の部分は，$\mathbf{H}$ のムーア・ペンローズ一般逆行列になっている．

実際，$\mathbf{H}$ の特異値分解を

$$\mathbf{H} = \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\mathsf{T}$$

として，計算すると，

$$\begin{aligned}(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T} &= (\check{\mathbf{V}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T}\check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\mathsf{T})^{-1}\check{\mathbf{V}}\check{\mathbf{\Lambda}}\check{\mathbf{T}}^\mathsf{T} = (\check{\mathbf{V}}\check{\mathbf{\Lambda}}^2\check{\mathbf{V}}^\mathsf{T})^{-1}\check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\mathsf{T} \\ &= \check{\mathbf{V}}\check{\mathbf{\Lambda}}^{-2}\check{\mathbf{V}}^\mathsf{T}\check{\mathbf{V}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T} = \check{\mathbf{V}}\check{\mathbf{\Lambda}}^{-2}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T} = \check{\mathbf{V}}\check{\mathbf{\Lambda}}^{-1}\check{\mathbf{U}}^\mathsf{T}\end{aligned} \tag{15}$$

特異値分解による方法

- 特異値分解から得られた結果

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{r} \frac{\beta_i}{\lambda_i} \boldsymbol{v}_i \tag{16}$$

より，λ_i が小さい場合，$\boldsymbol{x}$ の推定値は β_i に大きく影響されることも分かる．

- β_i に誤差がある場合，その誤差の $\hat{x}$ への影響が大きく増幅されるため，安定した解が得られない．
- $\hat{x}$ を安定させるため，

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{l} \frac{\beta_i}{\lambda_i} \boldsymbol{v}_i$$

$$(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l > \varepsilon \geq \lambda_{l+1} \geq \cdots \lambda_r > 0),$$

のように特異値の小さい成分を落とす場合もある．

$r < m \le n$ の場合

$r < m \le n$ の場合，$\alpha_{r+1} \dots \alpha_m$ は $\Psi_L = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$ に影響せず，

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{r} \frac{\beta_i}{\lambda_i} \boldsymbol{v}_i + \sum_{i=r+1}^{m} \alpha_i \boldsymbol{v}_i .$$

の形であればよい．

この式から以下のことが分かる:

- $\mathbf{H}$ を介した観測 $\boldsymbol{y}$ から，$\boldsymbol{x}$ のうち $\{\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_m\}$ の成分に関する情報を得ることはできない．
- $\{\alpha_{r+1}, \dots, \alpha_m\}$ が決まらないため，Ψ_L を最小化する解は一意には求まらない．

ノルム最小解

$r < m \leq n$ の場合の式

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{r} \frac{\beta_i}{\lambda_i} \boldsymbol{v}_i + \sum_{i=r+1}^{m} \alpha_i \boldsymbol{v}_i.$$

において，$\alpha_i = 0\ (i > r)$ と置いた

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{r} \frac{\beta_i}{\lambda_i} \boldsymbol{v}_i \tag{17}$$

は，$\Psi_L = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2$ を最小化する $\boldsymbol{x}$ のうち，ノルムが最小となる解を与え，ノルム最小解と呼ばれる．

$r \leq n < m$ の場合もノルム最小解

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{r} \frac{\beta_i}{\lambda_i} \boldsymbol{v}_i \tag{18}$$

が考えられ，Ψ_L を最小にする $\boldsymbol{x}$ のうち，ノルムを最小とする解を与える．

予想を取り入れる

- データ同化の扱う問題で，観測データの量が，推定したい変数 x の次元よりも多いという状況は，滅多にない．
- また，データの量が多かったとしても，行列 H の構造によっては，いい推定にはならない場合がある．
- データの情報が十分ではない場合，x は大体このあたりだろうと事前に予想をつけておき，その予想を第一推定値を $\bar{x}_b$ として，

$$\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \xi^2\|\boldsymbol{x} - \bar{\boldsymbol{x}}_b\|^2 \rightarrow \text{minimum}$$

  となるように $\hat{x}$ を求めるという方法が有効な場合がある．
- $\bar{x}_b$ には，例えば前のステップからシミュレーションを走らせた結果を使ってもよいし，過去のデータの平均値のような経験的な値を使ってもよい．

拘束付き最小二乗法

$$\Psi_C = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \xi^2\|\boldsymbol{x} - \bar{\boldsymbol{x}}_b\|^2 = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m} H_{ij}x_j\right)^2 + \sum_{k=1}^{m}\xi^2(x_k - x_{b,k})^2 \tag{19}$$

とおくと，

$$\frac{\partial\Psi_C}{\partial x_k} = -2\sum_{i=1}^{n} H_{ik}\left(y_i - \sum_{j=1}^{m} H_{ij}x_j\right) + 2\xi^2(x_i - x_{b,i}) = 0$$

なので，これをまとめて書くと，

$$\begin{aligned}-\mathbf{H}^\mathsf{T}\boldsymbol{y} + \mathbf{H}^\mathsf{T}\mathbf{H}\boldsymbol{x} + \xi^2(\boldsymbol{x} - \bar{\boldsymbol{x}}_b) &= -\mathbf{H}^\mathsf{T}\boldsymbol{y} + \mathbf{H}^\mathsf{T}\mathbf{H}\bar{\boldsymbol{x}}_b + \mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x} - \bar{\boldsymbol{x}}_b) + \xi^2(\boldsymbol{x} - \bar{\boldsymbol{x}}_b)\\ &= -\mathbf{H}^\mathsf{T}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b) + (\mathbf{H}^\mathsf{T}\mathbf{H} + \xi^2\mathbf{I})(\boldsymbol{x} - \bar{\boldsymbol{x}}_b) = 0.\end{aligned}$$

したがって，以下のようになる．

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + (\mathbf{H}^\mathsf{T}\mathbf{H} + \xi^2\mathbf{I}_n)^{-1}\mathbf{H}^\mathsf{T}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b) \tag{20}$$

逆行列の存在について

今，行列 $\mathbf{H}^\top\mathbf{H}+\xi^2\mathbf{I}_n$ が逆行列を持つものとして話を進めたが，$\xi \neq 0$ であれば，必ず逆行列は存在する．

一般に，$\mathbf{A}$ を n 次の半正定値対称行列，$\mathbf{B}$ を m 次の正定値対称行列，$\mathbf{H}$ を $m \times n$ 行列としたとき，行列 $\mathbf{H}^\top\mathbf{A}\mathbf{H}+\mathbf{B}$ は正則行列となる．

まず，$\mathbf{A}, \mathbf{B}$ が対称行列であることから

$$(\mathbf{H}^\top\mathbf{A}\mathbf{H}+\mathbf{B})^\top = \mathbf{H}^\top\mathbf{A}\mathbf{H}+\mathbf{B}$$

となり，$\mathbf{H}^\top\mathbf{A}\mathbf{H}+\mathbf{B}$ は対称行列．また，$\mathbf{A}$ が半正定値対称行列で，任意の n 次元ベクトル $\boldsymbol{x}$ に対して $\boldsymbol{x}^\top\mathbf{A}\boldsymbol{x} \geq 0$ が成り立つので，任意の m 次元ベクトル $\boldsymbol{z}$ に対して $\boldsymbol{z}^\top\mathbf{H}^\top\mathbf{A}\mathbf{H}\boldsymbol{z} \geq 0$．

$\mathbf{B}$ が正定値対称行列なので，$\boldsymbol{z}$ が零ベクトルでなければ $\boldsymbol{z}^\top(\mathbf{H}^\top\mathbf{A}\mathbf{H}+\mathbf{B})\boldsymbol{z} > 0$ が成り立ち，$\mathbf{H}^\top\mathbf{A}\mathbf{H}+\mathbf{B}$ は正定値対称行列．したがって，正則である．

特異値分解による方法

次に，事前の予想 $\bar{\boldsymbol{x}}_b$ を用いた推定を特異値分解を使って行ってみる．

そのためには，$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ を用いて $\boldsymbol{x} - \bar{\boldsymbol{x}}_b$ を

$$\boldsymbol{x} - \bar{\boldsymbol{x}}_b = \sum_{i=1}^{m} \alpha_i' \boldsymbol{v}_i, \tag{21}$$

$\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ を用いて $\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b$ を

$$\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b = \sum_{i=1}^{n} \beta_i' \boldsymbol{u}_i \tag{22}$$

のように分解した上で，

$$\Psi_C = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \xi^2 \|\boldsymbol{x} - \bar{\boldsymbol{x}}_b\|^2 \to \text{minimum}$$

とする $\boldsymbol{x}$ を求める．

特異値分解による方法

$$
\begin{aligned}
\Psi_C &= \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \xi^2 \|\boldsymbol{x} - \bar{\boldsymbol{x}}_b\|^2 \\
&= \left\| \sum_{i=1}^{n} \beta_i' \boldsymbol{u}_i - \left(\sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^\mathsf{T} \right) \left(\sum_{i=1}^{r} \alpha_i' \boldsymbol{v}_i \right) \right\|^2 + \xi^2 \left\| \sum_{i=1}^{m} \alpha_i' \boldsymbol{v}_i \right\|^2 \\
&= \left\| \sum_{i=1}^{n} \beta_i' \boldsymbol{u}_i - \sum_{i=1}^{r} \lambda_i \alpha_i' \boldsymbol{u}_i \right\|^2 + \xi^2 \left\| \sum_{i=1}^{m} \alpha_i' \boldsymbol{v}_i \right\|^2 \\
&= \sum_{i=1}^{r} (\beta_i' - \lambda_i \alpha_i')^2 + \sum_{i=r+1}^{n} \beta_i'^2 + \xi^2 \sum_{i=1}^{m} \alpha_i'^2
\end{aligned}
\tag{23}
$$

$$
\therefore \frac{\partial \Psi_C}{\partial \alpha_i'} = -\sum_{i=1}^{r} 2\lambda_i (\beta_i' - \lambda_i \alpha_i') + \xi^2 \sum_{i=1}^{m} 2\alpha_i' = 0 \tag{24}
$$

したがって，Ψ_C が最小になる α_i' は以下を満たす:

$$
\lambda_i \beta_i' = \lambda_i^2 \alpha_i' + \xi^2 \alpha_i' \quad (i \le r), \quad \alpha_i' = 0 \quad (i > r) \tag{25}
$$

特異値分解による方法

したがって，

$$\alpha'_i = \frac{\lambda_i \beta'_i}{\lambda_i^2 + \xi^2} \quad (i \leq r), \quad \alpha'_i = 0 \quad (i > r). \tag{26}$$

$\lambda_{r+1} = \cdots = \lambda_m = 0$ とおくと，$i > r$ の場合も含めて，

$$\alpha'_i = \frac{\lambda_i \beta'_i}{\lambda_i^2 + \xi^2} \tag{27}$$

と書け，Ψ_C を最小とする $\hat{\boldsymbol{x}}$ は

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + \sum_{i=1}^{m} \frac{\lambda_i \beta'_i}{\lambda_i^2 + \xi^2} \boldsymbol{v}_i. \tag{28}$$

特異値分解による方法

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + \sum_{i=1}^{m} \frac{\lambda_i \beta_i'}{\lambda_i^2 + \xi^2} \boldsymbol{v}_i$$

において，

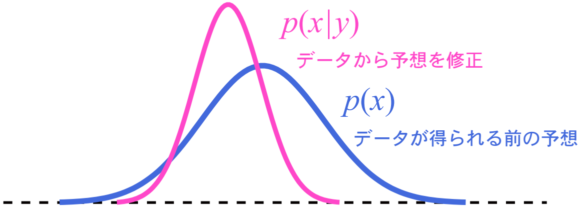
- $\lambda_i^2 \gg \xi^2$ となる項は，最小二乗解とほぼ同じとなり，
- $\lambda_i^2 \ll \xi^2$ となる項は，ほぼ 0 になる．

したがって，拘束条件を付けた場合は，λ_i^2 が小さい項が無視され，解を安定化させる効果が得られる．

ベイズの定理

予想をつけておいた上で，実データを参照して修正するという考え方は，ベイズの定理を用いて表現することもできる．

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\,d\boldsymbol{x}}. \tag{29}$$



ベイズの定理

- ある事象 A が起こったとき，考えられる原因を $B_1, B_2, \ldots, B_n$ とする．
- 原因 B_j が与えられたとき，結果が A となる確率 $P(A|B_i)$ はわかっているとする．
- $B_1, \ldots, B_n$ は互いに排反 (同時には起こり得ない) で，かつ $B_1, \ldots, B_n$ で考えられるすべての場合が考慮されているとする．
- このとき，原因が B_i である確率は

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}. \tag{30}$$

(事前情報 $p(B_j)$ が与えられれば，観測 A から原因を推定可能．)

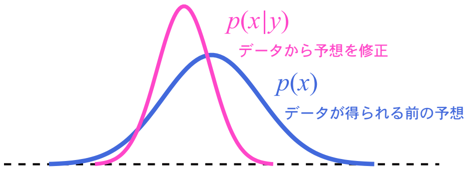
- 連続量の場合に以下の形になる．

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\,d\boldsymbol{x}} \tag{31}$$

ベイズの定理

ベイズの定理 $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\,d\boldsymbol{x}}$ において，

- $p(\boldsymbol{x})$ は，$\boldsymbol{x}$ に関する事前の予想を表す．(事前分布と呼ぶ)．
- $p(\boldsymbol{y}|\boldsymbol{x})$ は，$\boldsymbol{x}$ がこの値だったら，このくらいの確率で $\boldsymbol{y}$ が観測されるということを表す．(また，観測への $\boldsymbol{x}$ の当てはまり具合を表している面もある．)
- $p(\boldsymbol{x}|\boldsymbol{y})$ は，与えられた観測 $\boldsymbol{y}$ の下で $\boldsymbol{x}$ を推定したもの．(事後分布と呼ぶ．)



ベイズの定理に基づく推定

ここでは，ベイズの定理

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\,d\boldsymbol{x}} \tag{32}$$

において，$p(\boldsymbol{x})$, $p(\boldsymbol{y}|\boldsymbol{x})$ がガウス分布に従うと仮定して，

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m|\mathbf{P}_b|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b)^\mathsf{T}\mathbf{P}_b^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b)\right), \tag{33}$$

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{R}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})^\mathsf{T}\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})\right), \tag{34}$$

とする．上の式に代入すると，以下が満たされることが分かる:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b)^\mathsf{T}\mathbf{P}_b^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b) - \frac{1}{2}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})^\mathsf{T}\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})\right). \tag{35}$$

ベイズの定理に基づく推定

x を推定するには，事後確率

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b)^\top \mathbf{P}_b^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b) - \frac{1}{2}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})^\top \mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})\right)$$

の平均，あるいは事後確率を最大化する x を求めるのが，よく使われる手段．
(ガウス分布を仮定するとどちらでも同じ結果になる．)

$p(\boldsymbol{x}|\boldsymbol{y})$ の exp の指数部分を $-\Psi_B/2$ とおくと，

$$\Psi_B = (\boldsymbol{x}-\bar{\boldsymbol{x}}_b)^\top \mathbf{P}_b^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b) + (\boldsymbol{y}-\mathbf{H}\boldsymbol{x})^\top \mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x}) \tag{36}$$

を最小にする x が，$p(\boldsymbol{x}|\boldsymbol{y})$ を最大にする x となる．

以下，しばらくの間，$\mathbf{P}_b$，$\mathbf{R}$ は正定値行列 (0 固有値を持たない) とする．

$*$ 分散共分散行列が 0 固有値を持つ場合，逆行列が存在しなくなって都合が悪い．

ベイズの定理に基づく推定

拘束付き最小二乗法の目的関数

$$\Psi_C = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 + \xi^2 \|\boldsymbol{x} - \bar{\boldsymbol{x}}_b\|^2$$

は，ベイズの定理から得た

$$\Psi_B = (\boldsymbol{x} - \bar{\boldsymbol{x}}_b)^\mathsf{T} \mathbf{P}_b^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_b) + (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})^\mathsf{T} \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})$$

において，$\mathbf{P}_b = \xi^2 \mathbf{I}$ かつ $\mathbf{R} = \mathbf{I}$ とした場合と等価．

Ψ_B は Ψ_C の一般になっており，Ψ_B の方が予測の確度，観測の確度の情報を柔軟に取り入れられるのが利点．

ベイズの定理に基づく推定

例えば，$\boldsymbol{x}$ の i 番目の要素 x_i に関する予想の不確かさの幅を ζ_i，$\boldsymbol{y}$ の i 番目の要素 y_i の不確かさの幅を σ_i で表し，$\boldsymbol{x}$, $\boldsymbol{y}$ の異なる要素の間に事前に想定される相関がないものとすれば，

$$\mathbf{P}_b = \begin{pmatrix} \zeta_1^2 & & & \mathbf{O} \\ & \zeta_2^2 & & \\ & & \ddots & \\ \mathbf{O} & & & \zeta_m^2 \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} \sigma_1^2 & & & \mathbf{O} \\ & \sigma_2^2 & & \\ & & \ddots & \\ \mathbf{O} & & & \sigma_m^2 \end{pmatrix} \tag{37}$$

のように $\mathbf{P}_b$, $\mathbf{R}$ はそれぞれ対角行列で書ける．このとき Ψ_B は

$$\Psi_B = \sum_{i=1}^{m} \zeta_i^{-2}(x_i - \bar{x}_{b,i})^2 + \sum_{i=1}^{n} \sigma_i^{-2} \left(y_i - \sum_{j=1}^{m} H_{ij} x_{b,j} \right)^2 . \tag{38}$$

$\boldsymbol{x}$, $\boldsymbol{y}$ の各要素が不確かさの逆数で重みづけされており，不確かさの小さい (信用できる) 情報ほど大きな重みがついている．したがって，Ψ_B を最小化すると，信用できる情報をより重視した推定値が得られる．

ベイズの定理に基づく推定

$$\Psi_B = (\boldsymbol{x} - \bar{\boldsymbol{x}}_b)^\mathsf{T} \mathbf{P}_b^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_b) + (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})^\mathsf{T} \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})$$

は，以下の形に変形できる:

$$\Psi_B = (\boldsymbol{x} - \hat{\boldsymbol{x}})^\mathsf{T} (\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T} \mathbf{R}^{-1} \mathbf{H}) (\boldsymbol{x} - \hat{\boldsymbol{x}}) + A. \tag{39}$$

但し，

- A は x によらない定数，
- $\hat{x}$ は，

$$(\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T} \mathbf{R}^{-1} \mathbf{H}) \hat{\boldsymbol{x}} = \mathbf{P}_b^{-1} \bar{\boldsymbol{x}}_b + \mathbf{H}^\mathsf{T} \mathbf{R}^{-1} \boldsymbol{y}$$

を満たすベクトル．

ベイズの定理に基づく推定

$\hat{\boldsymbol{x}}$ については，

$$\begin{aligned}(\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})\hat{\boldsymbol{x}} &= \mathbf{P}_b^{-1}\bar{\boldsymbol{x}}_b + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\boldsymbol{y} \\ &= (\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})\bar{\boldsymbol{x}}_b + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b)\end{aligned} \tag{40}$$

より，以下のように求められる:

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + (\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b). \tag{41}$$

但し，分散共分散行列 $\mathbf{P}_b$，$\mathbf{R}$ を正定値対称行列としたので，

- $\mathbf{P}_b^{-1}$，$\mathbf{R}^{-1}$ も正定値対称行列で，
- $\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}$ も正定値対称行列

となり，$\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}$ が逆行列を持つことに注意．

ベイズの定理に基づく推定

$$
\begin{aligned}
-\frac{\Psi_B}{2} &= -\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b)^\mathsf{T}\mathbf{P}_b^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_b) - \frac{1}{2}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})^\mathsf{T}\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x}) \\
&= -\frac{1}{2}(\boldsymbol{x}-\hat{\boldsymbol{x}})^\mathsf{T}(\mathbf{P}_b^{-1}+\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})(\boldsymbol{x}-\hat{\boldsymbol{x}}) - A.
\end{aligned}
$$

から，$\hat{\mathrm{P}} = (\mathbf{P}_b^{-1}+\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})^{-1}$ とおくと，事後分布 $p(\boldsymbol{x}|\boldsymbol{y})$ は，

$$
p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}-\hat{\boldsymbol{x}})^\mathsf{T}\hat{\mathbf{P}}^{-1}(\boldsymbol{x}-\hat{\boldsymbol{x}})\right) \tag{42}
$$

を満たすことがわかる．

したがって，$p(\boldsymbol{x}|\boldsymbol{y})$ は，平均ベクトル $\hat{\boldsymbol{x}}$，分散共分散行列 $\hat{\mathrm{P}}$ のガウス分布になる．

ベイズの定理に基づく推定

分散共分散行列については，

$$\hat{\mathbf{P}} = (\mathbf{P}_b^{-1} + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H})^{-1} = \mathbf{P}_b - \mathbf{P}_b \mathbf{H}^\top (\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1} \mathbf{H}\mathbf{P}_b \tag{43}$$

とも変形できる．このことは，

$$\begin{aligned}
&(\mathbf{P}_b^{-1} + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}) \left[\mathbf{P}_b - \mathbf{P}_b \mathbf{H}^\top (\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1} \mathbf{H}\mathbf{P}_b\right] \\
&\quad = \mathbf{I} + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}\mathbf{P}_b - \mathbf{H}^\top (\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1} \mathbf{H}\mathbf{P}_b \\
&\qquad - \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}\mathbf{P}_b\mathbf{H}^\top (\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1} \mathbf{H}\mathbf{P}_b \\
&\quad = \mathbf{I} + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}\mathbf{P}_b - \mathbf{H}^\top \mathbf{R}^{-1} (\mathbf{R} + \mathbf{H}\mathbf{P}_b\mathbf{H}^\top)(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1} \mathbf{H}\mathbf{P}_b = \mathbf{I}
\end{aligned} \tag{44}$$

となることから確認できる．

ベイズの定理に基づく推定

また，平均ベクトルの式，

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + (\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b). \tag{45}$$

は，

$$(\mathbf{P}_b^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})^{-1} = \mathbf{P}_b - \mathbf{P}_b\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_b \tag{46}$$

を代入すると，以下のように変形できる:

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \bar{\boldsymbol{x}}_b + \left[\mathbf{P}_b - \mathbf{P}_b\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_b\right]\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b) \\
&= \bar{\boldsymbol{x}}_b + \mathbf{P}_b\mathbf{H}^\mathsf{T}\left[\mathbf{I} - (\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T}\right]\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b) \\
&= \bar{\boldsymbol{x}}_b + \mathbf{P}_b\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\left[(\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T} + \mathbf{R}) - \mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T}\right]\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b) \\
&= \bar{\boldsymbol{x}}_b + \mathbf{P}_b\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b).
\end{aligned}$$

- $\dim \boldsymbol{x} > \dim \boldsymbol{y}$ の場合には，こちらの方が便利である．
- 後で出てくるカルマンフィルタでもこの式を使う．

ベイズの定理に基づく推定 (まとめ)

$p(\boldsymbol{x}) = \mathcal{N}(\bar{\boldsymbol{x}}_b, \mathbf{P}_b)$, $p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\mathbf{H}\boldsymbol{x}, \mathbf{R})$ のとき，事後分布 $p(\boldsymbol{x}|\boldsymbol{y})$ もガウス分布となり，平均ベクトル，分散共分散行列は以下のようになる．

- 平均ベクトル:

$$\begin{aligned} \hat{\boldsymbol{x}} &= \bar{\boldsymbol{x}}_b + (\mathbf{P}_b^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b) \\ &= \bar{\boldsymbol{x}}_b + \mathbf{P}_b\mathbf{H}^\top(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b), \end{aligned} \tag{47}$$

- 分散共分散行列:

$$\begin{aligned} \hat{\mathbf{P}} &= (\mathbf{P}_b^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H})^{-1} \\ &= \mathbf{P}_b - \mathbf{P}_b\mathbf{H}^\top(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_b \end{aligned} \tag{48}$$

最適内挿法

- ベイズ推定の式

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + \mathbf{P}_b\mathbf{H}^\top(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b),$$
$$\hat{\mathbf{P}} = \mathbf{P}_b - \mathbf{P}_b\mathbf{H}^\top(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_b$$

  は，データ同化では最適内挿法と呼ばれる手法に相当する．
- 実際に使用する際には，$\mathbf{P}_b$ や $\mathbf{R}$ の設計が問題になる．
- $\mathbf{P}_b$ については，グリッド点間の距離の関数

$$\sigma^2 \exp\left[-\frac{\|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2}{2\ell^2}\right] \tag{49}$$

  のようなものを使うこともある．

補足

なお，途中で使った，

$$(\mathbf{P}^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1} = \mathbf{P}\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}, \tag{50a}$$

$$(\mathbf{P}^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H})^{-1} = \mathbf{P} - \mathbf{P}\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P} \tag{50b}$$

という公式は，逆行列補題，Sherman-Morrison-Woodbury 公式などと呼ばれ，このあと，何度か出てくる．

P_b が 0 固有値を持つ場合

$\boldsymbol{x}$ の事前分布の分散共分散行列 $\mathbf{P}_b$ が 0 固有値を持つ場合，$\mathbf{P}_b = \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T}$ のように固有値分解し，

$$\boldsymbol{x} = \bar{\boldsymbol{x}}_b + \check{\mathbf{U}}\boldsymbol{z} \tag{51}$$

とすれば，ガウス分布 $\mathcal{N}(\mathbf{0}, \check{\mathbf{\Lambda}})$ に従う確率変数 z を使って x を表現することができる．

(ここでは，$\check{\mathbf{\Lambda}}$ が正定値 (正則) になるように 0 の固有値に対応する成分を削った，

$$\mathbf{P}_b = \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \end{pmatrix} \begin{pmatrix} \lambda_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \lambda_r \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \end{pmatrix}^\mathsf{T} \tag{52}$$

の方を使う．)

P_b が 0 固有値を持つ場合

ガウス分布 $\mathcal{N}(\mathbf{0}, \check{\mathbf{\Lambda}})$ に従う確率変数 $\boldsymbol{z}$ を使い，$\boldsymbol{x}$ を

$$\boldsymbol{x} = \bar{\boldsymbol{x}}_b + \check{\mathbf{U}}\boldsymbol{z} \tag{53}$$

のように表現したとき，

$$p(\boldsymbol{z}) = \frac{1}{\sqrt{(2\pi)^r}} \exp\left(-\frac{1}{2}\boldsymbol{z}^\top \check{\mathbf{\Lambda}}^{-1}\boldsymbol{z}\right), \tag{54}$$

$$p(\boldsymbol{y}|\boldsymbol{z}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{R}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b - \mathbf{H}\check{\mathbf{U}}\boldsymbol{z})^\top \mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b - \mathbf{H}\check{\mathbf{U}}\boldsymbol{z})\right), \tag{55}$$

の下で，$\boldsymbol{z}$ の事後分布を求めればよい．

$*$ $\mathbf{\Lambda}$ は正則なので，$\boldsymbol{z}$ の確率密度関数が定義できる．

P_b が 0 固有値を持つ場合

$\boldsymbol{z}$ の分散共分散行列 $\check{\boldsymbol{\Lambda}}$ は正定値なので，$\boldsymbol{z}$ の事後分布の平均は，先程の式から直ちに求まる．

$$\hat{\boldsymbol{z}} = (\check{\boldsymbol{\Lambda}}^{-1} + \check{\mathbf{U}}^\intercal \mathbf{H}^\intercal \mathbf{R}^{-1} \mathbf{H} \check{\mathbf{U}})^{-1} \check{\mathbf{U}}^\intercal \mathbf{H}^\intercal \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H} \bar{\boldsymbol{x}}_b). \tag{56}$$

両辺に左から $\check{\mathrm{U}}$ を掛けた上で変形すると，

$$\begin{aligned} \check{\mathbf{U}} \hat{\boldsymbol{z}} &= \check{\mathbf{U}} (\check{\boldsymbol{\Lambda}}^{-1} + \check{\mathbf{U}}^\intercal \mathbf{H}^\intercal \mathbf{R}^{-1} \mathbf{H} \check{\mathbf{U}})^{-1} \check{\mathbf{U}}^\intercal \mathbf{H}^\intercal \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H} \bar{\boldsymbol{x}}_b) \\ &= \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{U}}^\intercal \mathbf{H}^\intercal (\mathbf{H} \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{U}}^\intercal \mathbf{H}^\intercal + \mathbf{R})^{-1} (\boldsymbol{y} - \mathbf{H} \bar{\boldsymbol{x}}_b) \\ &= \mathbf{P}_b \mathbf{H}^\intercal (\mathbf{H} \mathbf{P}_b \mathbf{H}^\intercal + \mathbf{R})^{-1} (\boldsymbol{y} - \mathbf{H} \bar{\boldsymbol{x}}_b) \end{aligned} \tag{57}$$

したがって，

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + \mathbf{P}_b \mathbf{H}^\intercal (\mathbf{H} \mathbf{P}_b \mathbf{H}^\intercal + \mathbf{R})^{-1} (\boldsymbol{y} - \mathbf{H} \bar{\boldsymbol{x}}_b). \tag{58}$$

P_b が 0 固有値を持つ場合

z の事後分布の分散共分散行列も，先程の式から

$$\hat{\mathbf{P}}_z = (\check{\mathbf{\Lambda}}^{-1} - \check{\mathbf{U}}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}\check{\mathbf{U}})^{-1} \tag{59}$$

となるので，両辺の両側から $\check{\mathbf{U}}$ を掛けた上で変形すると，

$$\begin{aligned}
\hat{\mathbf{P}} &= \check{\mathbf{U}}\hat{\mathbf{P}}_z\check{\mathbf{U}}^\mathsf{T} \\
&= \check{\mathbf{U}}(\check{\mathbf{\Lambda}}^{-1} - \check{\mathbf{U}}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}\check{\mathbf{U}})^{-1}\check{\mathbf{U}}^\mathsf{T} \\
&= \check{\mathbf{U}}\left[\check{\mathbf{\Lambda}} - \check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T}\mathbf{H}^\mathsf{T}(\mathbf{H}\check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T}\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\mathbf{H}\check{\mathbf{U}}\check{\mathbf{\Lambda}}\right]\check{\mathbf{U}}^\mathsf{T} \\
&= \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T} - \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T}\mathbf{H}^\mathsf{T}(\mathbf{H}\check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T}\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\mathbf{H}\check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{U}}^\mathsf{T} \\
&= \mathbf{P}_b - \mathbf{P}_b\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}_b\mathbf{H}^\mathsf{T} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_b
\end{aligned} \tag{60}$$

つまり，正則の場合の平均ベクトル，分散共分散行列の式を非正則な場合に適用しても問題ない．

まとめ

- 多変量の問題における最小二乗法とその拡張について述べた．
- データの情報が十分ではない場合，単純な最小二乗法ではうまく行かないため，解に関して事前に予想を付け加え，その予想に基づいて推定を行うという方法がある．
- 事前の予想を付加する方法は，ベイズ推定の枠組みに一般化できる．今回は，正規分布を仮定できる場合に限って説明したが，正規分布以外の事前分布，尤度を考えることも可能である．但し，正規分布を仮定しない場合，解析的に事後分布を得られるとは限らない．