「データ科学: 理論から実用へ演習」

中野 慎也

7 September 2023

特異値分解

任意の $m \times n$ 行列 $\mathbf{A}$ (今回，要素はすべて実数とする) は以下のように分解できる:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\top}. \tag{1}$$

ここで $\mathbf{U}$ は $m \times m$ の，$\mathbf{V}$ は $n \times n$ の直交行列であり，また，

$$\boldsymbol{\Lambda} = \begin{pmatrix} \begin{matrix} \lambda_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \lambda_r \end{matrix} & \mathbf{O}_{r\times(n-r)} \\ \mathbf{O}_{(m-r)\times r} & \mathbf{O}_{(m-r)\times(n-r)} \end{pmatrix}. \tag{2}$$

$\mathbf{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m]$, $\mathbf{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ と縦ベクトルに分解すると，$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ は，それぞれ m 次元，n 次元ベクトル空間の正規直交基底をなす．

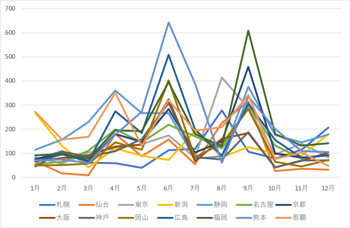
- $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ を特異ベクトルと呼ぶ．また，$\{\lambda_1, \ldots, \lambda_r\}$ を特異値と呼ぶ．

特異値分解の応用例

以下のデータに基づく行列を特異値分解する．

	Jan.	$\cdots$	Dec.
札幌の降水量	$x_{1,1}$	$\cdots$	$x_{12,1}$
仙台の降水量	$x_{1,2}$	$\cdots$	$x_{12,2}$
東京の降水量	$x_{1,3}$	$\cdots$	$x_{12,3}$
新潟の降水量	$x_{1,4}$	$\cdots$	$x_{12,4}$
静岡の降水量	$x_{1,5}$	$\cdots$	$x_{12,5}$
名古屋の降水量	$x_{1,6}$	$\cdots$	$x_{12,6}$
京都の降水量	$x_{1,7}$	$\cdots$	$x_{12,7}$
大阪の降水量	$x_{1,8}$	$\cdots$	$x_{12,8}$
神戸の降水量	$x_{1,9}$	$\cdots$	$x_{12,9}$
岡山の降水量	$x_{1,10}$	$\cdots$	$x_{12,10}$
広島の降水量	$x_{1,11}$	$\cdots$	$x_{12,11}$
福岡の降水量	$x_{1,12}$	$\cdots$	$x_{12,12}$
熊本の降水量	$x_{1,13}$	$\cdots$	$x_{12,13}$
那覇の降水量	$x_{1,14}$	$\cdots$	$x_{12,14}$

データ行列の例 (2016 年の各月の降水量)



$$\boldsymbol{x}_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,m} \end{pmatrix} \quad (i = 1, \dots, n),$$

$$\boldsymbol{\mu} = \sum_i \boldsymbol{x}_i / n \text{ として，}$$

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu} & \cdots & \boldsymbol{x}_n - \boldsymbol{\mu} \end{pmatrix}.$$

特異値分解の応用例

データ行列 $\mathbf{X}$ の特異値分解

$$\mathbf{X} = \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\top$$

を考える．$\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\top$ を成分で書くと，

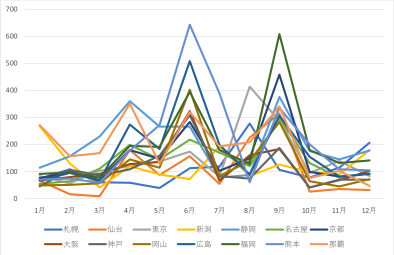
$$\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\top = \begin{pmatrix} \lambda_1 v_{1,1} & \cdots & \lambda_1 v_{1,n} \\ \vdots & \ddots & \vdots \\ \lambda_m v_{m,1} & \cdots & \lambda_m v_{m,n} \end{pmatrix}. \tag{3}$$

これを用いると，

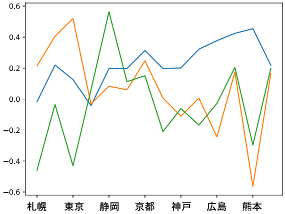
$$\begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{r} \lambda_j v_{j,1} \boldsymbol{u}_j & \cdots & \sum_{j=1}^{r} \lambda_j v_{j,n} \boldsymbol{u}_j \end{pmatrix}. \tag{4}$$

つまり，特異値分解により，各 $\boldsymbol{x}_i$ が共通の基底 $\{\boldsymbol{u}_j\}$ による直交基底展開が得られる．

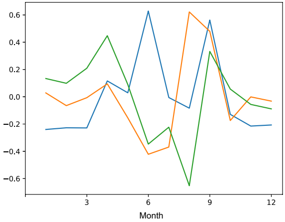
特異値分解の応用例



各月の降水量 (2016)



主成分ベクトル $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$



主成分の変動 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$

特異値分解の応用例

$$(\boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_n) = \left(\sum_{j=1}^{r} \lambda_j v_{j,1} \boldsymbol{u}_j \quad \cdots \quad \sum_{j=1}^{r} \lambda_j v_{j,n} \boldsymbol{u}_j \right)$$

において，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ なので，ある閾値 ε を与えたとき，$\lambda_L > \varepsilon \geq \lambda_{L+1} (\geq \lambda_r)$ なる L をとることができる．このとき

$$(\tilde{\boldsymbol{x}}_1 \quad \cdots \quad \tilde{\boldsymbol{x}}_n) = \left(\sum_{j=1}^{L} \lambda_j v_{j,1} \boldsymbol{u}_j \quad \cdots \quad \sum_{j=1}^{L} \lambda_j v_{j,n} \boldsymbol{u}_j \right) \tag{5}$$

のように $L+1$ 番目以降の成分を落とすと，$\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n$ は L 個の基底ベクトルによる $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ の近似となる．実際，

$$\|\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i\|^2 = \left\| \sum_{j=L+1}^{r} \lambda_j v_{j,i} \boldsymbol{u}_j \right\|^2 = \sum_{j=L+1}^{r} (\lambda_j v_{j,i})^2 < \sum_{j=L+1}^{r} \varepsilon^2 v_{j,i}^2 < \varepsilon^2 \sum_{j=1}^{n} v_{j,i}^2 = \varepsilon^2$$

が言える．(最後の等号は $\mathbf{V}$ が直交行列であることから言える．)
この操作は主成分分析と呼ばれる手法に対応している．

主成分分析について

主成分分析は，多次元データ $\boldsymbol{x}^{(1)}$, …, $\boldsymbol{x}^{(n)}$ が与えられたとき，長さ 1 のベクトル $\boldsymbol{a}$ ($\|\boldsymbol{a}\|^2 = 1$) を，各 $\boldsymbol{x}^{(i)}$ との (平均 $\boldsymbol{\mu}$ を引いた上での) 内積

$$z^{(i)} = \boldsymbol{a}^\mathsf{T}(\boldsymbol{x}^{(i)} - \boldsymbol{\mu}) \quad (i = 1, \dots, n) \tag{6}$$

で得られる $\{z^{(i)}\}$ の分散が最大になるように選ぶというのが本来の定義である．

$$\boldsymbol{d} = \begin{pmatrix} z^{(1)} \\ \vdots \\ z^{(n)} \end{pmatrix} \tag{7}$$

とおくと，その $\{z^{(i)}\}$ の分散は以下のように書ける．

$$P_z = \frac{1}{n-1}\boldsymbol{d}^\mathsf{T}\boldsymbol{d}. \tag{8}$$

これを最大化する $\boldsymbol{a}$ を探せばよい．

主成分分析について

平均を引いたデータ行列

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}^{(1)} - \boldsymbol{\mu} & \cdots & \boldsymbol{x}^{(n)} - \boldsymbol{\mu} \end{pmatrix} \tag{9}$$

を考えると，$\boldsymbol{d}^\top = \boldsymbol{a}^\top \mathbf{X}$ となる．

X の特異値分解

$$\mathbf{X} = \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{V}}^\top$$

を考えると，

$$\boldsymbol{d}^\top = \boldsymbol{a}^\top \mathbf{X} = \boldsymbol{a}^\top \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{V}}^\top$$

なので，

$$P_z = \frac{1}{n-1} \boldsymbol{d}^\top \boldsymbol{d} = \frac{1}{n-1} \boldsymbol{a}^\top \check{\mathbf{U}} \check{\boldsymbol{\Lambda}}^2 \check{\mathbf{U}}^\top \boldsymbol{a}. \tag{10}$$

主成分分析について

$$\boldsymbol{a} = \mathbf{U}\boldsymbol{b} = \sum_{j=1}^{m} b_j \boldsymbol{u}_j$$

とおくと，$\|\boldsymbol{a}\|^2 = 1$ なので $\|\boldsymbol{b}\|^2 = 1$．

これを用いると，

$$P_z = \frac{1}{n-1}\boldsymbol{a}^\top \check{\mathbf{U}}\check{\boldsymbol{\Lambda}}^2\check{\mathbf{U}}^\top \boldsymbol{a} = \frac{1}{n-1}\boldsymbol{b}^\top \boldsymbol{\Lambda}^2 \boldsymbol{b} = \sum_{j=1}^{r} b_j^2 \lambda_j^2. \tag{11}$$

$\lambda_1 \geq \cdots \geq \lambda_r > 0$ とすると，$\|\boldsymbol{b}\|^2 = 1$ の下で P_z を最大化するには

$$b_1 = 1, \qquad b_j = 0 \quad (j \neq 1) \tag{12}$$

とすればよいことが分かる．したがって，$\boldsymbol{a} = \boldsymbol{u}_1$ のとき，P_z が最大となる．

主成分分析について

以上の議論から，

$$z^{(i)} = \boldsymbol{a}^\mathsf{T}(\boldsymbol{x}^{(i)} - \boldsymbol{\mu}) \quad (i = 1, \ldots, n)$$

の形で与えられる $\{z^{(i)}\}$ の分散 P_z を最大にするには，$\boldsymbol{a} = \boldsymbol{u}_1$ とすればよい．

これは，$\{\boldsymbol{x}^{(i)}\}$ が $\boldsymbol{u}_1$ の方向に大きくばらついていることを意味する．この $\boldsymbol{u}_1$ を第 1 主成分と呼ぶ．

$\boldsymbol{u}_1$ の次にばらつきの大きい方向を調べるのは，先程の議論を繰り返せばよい．すなわち，

$$\boldsymbol{a} = \mathbf{U}\boldsymbol{b} = \sum_{j=1}^{m} b_j \boldsymbol{u}_j, \qquad\qquad P_z = \sum_{j=1}^{r} b_j^2 \lambda_j^2$$

と，$\lambda_1 \geq \cdots \geq \lambda_r > 0$ から，$\boldsymbol{u}_1$ に直交で $\|\boldsymbol{a}\|^2 = 1$ という条件の下で P_z を最大化するベクトル $\boldsymbol{a}$ は，$\boldsymbol{a} = \boldsymbol{u}_2$ となる．$\boldsymbol{u}_2$ を第 2 主成分と呼ぶ．
その次にばらつきの大きい方向 (第 3 主成分，第 4 主成分, $\ldots$) も，同様にして順次得られる．

主成分分析について

- 今，
$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}^{(1)} - \boldsymbol{\mu} & \cdots & \boldsymbol{x}^{(n)} - \boldsymbol{\mu} \end{pmatrix}$$
の特異値分解
$$\mathbf{X} = \check{\mathbf{U}}\check{\boldsymbol{\Lambda}}\check{\mathbf{V}}^\top = (\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_r) \begin{pmatrix} \lambda_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \lambda_r \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1^\top \\ \vdots \\ \boldsymbol{v}_r^\top \end{pmatrix}$$
から，主成分分析における各主成分が得られた．
- なお，通常の主成分分析では，$\{x^{(i)}\}$ の分散共分散行列
$$\mathbf{P} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\boldsymbol{x}^{(i)} - \boldsymbol{\mu}\right) \left(\boldsymbol{x}^{(i)} - \boldsymbol{\mu}\right)^\top = \frac{1}{n-1} \mathbf{X}\mathbf{X}^\top \tag{13}$$
の固有値分解を行う．

主成分分析について

- 実際には，$\mathbf{X}$ の特異値分解

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$$

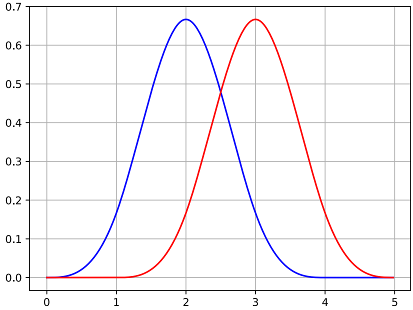
を用いると，分散共分散行列は，

$$\mathbf{P} = \frac{1}{n-1}\mathbf{X}\mathbf{X}^\top = \frac{1}{n-1}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top\mathbf{V}\mathbf{\Lambda}\mathbf{U}^\top = \frac{1}{n-1}\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^\top \tag{14}$$

と変形できる．これは $\mathbf{P}$ の固有値分解となっており，$\mathbf{U}$ を構成する縦ベクトル $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ は，$\mathbf{P}$ の固有ベクトルとなる．したがって，特異値分解を使っても，固有値分解を使っても同じ $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots$ が得られるので，どちらでも問題ない．

スプライン回帰

非線型関数 (直線でない関数) を近似するために，よく用いられる方法として，スプライン曲線の当てはめがある．B スプライン関数 (B-spline) という基底関数の重ね合わせで表す方法が扱いやすい．



3 次の B-spline

B-spline

まず B-spline を導入しておく．

d 次 B-spline 基底関数 $b_{j,d}(z)$ を定義するために，まず，K 個の節点 $z_0 \le z_1 \le \cdots \le z_{K-1}$ を与えておく．
節点が与えられると，0 次の B-spline は $j = 0, \ldots, K-2$ に対して

$$b_{0,j}(z) = 1 \quad (\text{if } z_j \le z < z_{j+1})$$
$$b_{0,j}(z) = 0 \quad (\text{otherwise})$$

となる．1 次以降の B-spline は，以下の漸化式で与えられる:

$$b_{d,j}(z) = \frac{z - z_j}{z_{d+j} - z_j} b_{d-1,j}(z) + \frac{z_{d+j+1} - z}{z_{d+j+1} - z_{j+1}} b_{d-1,j+1}(z).$$

B-spline

回帰に使う場合，節点を等間隔に配置することが多い．節点の幅を h として，節点を $0, h, 2h, \cdots, (K-1)h$ のように配置すると，先程の漸化式は

$$b_{d,j}(z) = \frac{z - jh}{dh} b_{d-1,j}(z) + \frac{(d+j+1)h - z}{dh} b_{d-1,j+1}(z).$$

これを使って，$b_{d,0}(z)$ を求めると，1 次については

$$\begin{aligned} b_{1,0}(z) &= 0 \quad (z < 0,\ z \geq 2h) \\ b_{1,0}(z) &= z/h \quad (0 \leq z < h) \\ b_{1,0}(z) &= 2 - z/h \quad (h \leq z < 2h) \end{aligned}$$

$b_{1,1}(z),\ b_{1,2}(z),\ \ldots,\ b_{1,K-3}$ については

$$b_{1,j}(z) = b_{1,0}(z - jh)$$

から得られる．

B-spline

2 次の $b_{2,0}(z)$ は，

$$
\begin{aligned}
b_{2,0}(z) &= 0 \quad (z < 0,\ \ z \geq 3h) \\
b_{2,0}(z) &= \frac{z^2}{2h^2} \quad (0 \leq z < h) \\
b_{2,0}(z) &= -\left(\frac{z}{h}\right)^2 + \frac{3z}{h} - \frac{3}{2} \quad (h \leq z < 2h) \\
b_{2,0}(z) &= \frac{(3h-z)^2}{2h^2} \quad (2h \leq z < 3h)
\end{aligned}
$$

B-spline

3 次の $b_{3,0}(z)$ は，

$$\begin{aligned}
b_{3,0}(z) &= 0 \quad (z < 0,\ z \geq 4h) \\
b_{3,0}(z) &= \frac{z^3}{6h^3} \quad (0 \leq z < h) \\
b_{3,0}(z) &= -\frac{z^3}{2h^3} + \frac{2z^2}{h^2} - \frac{2z}{h} + \frac{2}{3} \quad (h \leq z < 2h) \\
b_{3,0}(z) &= \frac{z^3}{2h^3} - \frac{4z^2}{h^2} + \frac{10z}{h} - \frac{22}{3} \quad (2h \leq z < 3h) \\
b_{3,0}(z) &= \frac{(4h-z)^3}{6h^2} \quad (3h \leq z < 4h)
\end{aligned}$$

B-spline

d 次の B-spline で，区間 $z_0 \leq z < z_{K-1}$ の 1 次元の曲線を表すには，

$$f(z) = \sum_{j=0}^{K-d-2} \beta_j b_{d,j}(z)$$

のように基底関数の重み付き和を用いる．重み $\{\beta_0, \ldots, \beta_{K-d-2}\}$ を推定することで曲線の形状が推定される．

節点を等間隔に配置した場合，節点を $0, h, 2h, \cdots, (K-1)h$ とすると，

$$f(z) = \sum_{j=0}^{K-d-2} \beta_j b_{d,0}(z - jh)$$

のようにも書ける．

B-spline

z と y のデータ $\{(z_1, y_1), (z_2, y_2), \ldots, (z_n, y_n)\}$ に対して，

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} b_{d,0}(z_1) & \cdots & b_{d,K-d-2}(z_1) \\ \vdots & \ddots & \vdots \\ b_{d,0}(z_n) & \cdots & b_{d,K-d-2}(z_n) \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{K-d-2} \end{pmatrix}$$

とおくと，$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{w}$ の形に書けるので，最小二乗法などで，$\boldsymbol{x}$ を推定できる．

拘束付き最小二乗法

- 最小二乗法は，十分に観測データがあれば有効だが，推定したい変数 x の次元に対して観測データの量が少ないという状況も少なくない．
- また，データの量が十分だったとしても，行列 $\mathbf{H}$ 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 $\bar{\boldsymbol{x}}_b$ として，

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

  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．
- x_b には，過去のデータの平均値のような経験的な値が使えればそれを使う．何も情報がない場合は，$\bar{x}_b = \mathbf{0}$ とすることもある．$\bar{x}_b = \mathbf{0}$ の場合は，リッジ回帰と呼ばれるものに相当する．

拘束付き最小二乗法

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

のように成分にばらすと，

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

なので，

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

したがって，以下のようになる．

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

特異値分解の利用

さて，観測できない真の $\boldsymbol{x}$ の値を $\boldsymbol{x}_{\mathrm{true}}$ とし，観測データ $\boldsymbol{y}$ を $\boldsymbol{y} = \mathbf{H}\boldsymbol{x}_{\mathrm{true}} + \boldsymbol{w}$ と書くと，拘束付き最小二乗法による $\boldsymbol{x}$ の推定値は，

$$
\begin{aligned}
\hat{\boldsymbol{x}} &= \bar{\boldsymbol{x}}_b + (\mathbf{H}^\top\mathbf{H} + \xi^2\mathbf{I})^{-1}\mathbf{H}^\top(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b) \\
&= \bar{\boldsymbol{x}}_b + (\mathbf{H}^\top\mathbf{H} + \xi^2\mathbf{I})^{-1}\mathbf{H}^\top\left[\mathbf{H}(\boldsymbol{x}_{\mathrm{true}} - \bar{\boldsymbol{x}}_b) + \boldsymbol{w}\right].
\end{aligned}
$$

このうち，$\varepsilon = (\mathbf{H}^\top\mathbf{H} + \xi^2\mathbf{I})^{-1}\mathbf{H}^\top\boldsymbol{w}$ の部分が観測ノイズの寄与となる．

ここで，行列 $\mathbf{H}$ の階数が $r = m < n$ であったと仮定して，

$$
\mathbf{H} = \sum_{i=1}^{m} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^\top
$$

のように $\mathbf{H}$ を特異値分解する．さらに，$\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ が n 次元ベクトル空間の正規直交基底をなすことに注意し，$\boldsymbol{w}$ を以下のように分解する:

$$
\boldsymbol{w} = \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i.
$$

特異値分解の利用

$\boldsymbol{\varepsilon} = (\mathbf{H}^\top\mathbf{H} + \xi^2\mathbf{I})^{-1}\mathbf{H}^\top\boldsymbol{w}$ に

$$\mathbf{H} = \sum_{i=1}^{m} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^\top, \qquad \boldsymbol{w} = \sum_{i=1}^{n} \beta_i \boldsymbol{u}_i$$

を代入する．$\{\boldsymbol{u}_i\}$, $\{\boldsymbol{v}_i\}$ の直交性から

$$\mathbf{H}^\top\mathbf{H} = \sum_{i=1}^{m} \lambda_i^2 \boldsymbol{v}_i \boldsymbol{v}_i^\top, \qquad \mathbf{I} = \mathbf{V}\mathbf{V}^\top = \sum_{i=1}^{m} \boldsymbol{v}_i \boldsymbol{v}_i^\top$$

などが言えることに注意すると，

$$\begin{aligned}\boldsymbol{\varepsilon} &= \left(\sum_{i=1}^{m} \lambda_i^2 \boldsymbol{v}_i \boldsymbol{v}_i^\top + \xi^2 \sum_{i=1}^{m} \boldsymbol{v}_i \boldsymbol{v}_i^\top\right)^{-1} \left(\sum_{i=1}^{m} \lambda_i \beta_i \boldsymbol{v}_i\right) \\ &= \left(\sum_{i=1}^{m} \frac{\boldsymbol{v}_i \boldsymbol{v}_i^\top}{\lambda_i^2 + \xi^2}\right) \left(\sum_{i=1}^{m} \lambda_i \beta_i \boldsymbol{v}_i\right) = \sum_{i=1}^{m} \frac{\lambda_i \beta_i}{\lambda_i^2 + \xi^2} \boldsymbol{v}_i.\end{aligned}$$

特異値分解の利用

目的関数

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

で $\xi^2 = 0$ とおいた場合，

$$\boldsymbol{\varepsilon} = \sum_{i=1}^{m} \frac{\lambda_i \beta_i}{\lambda_i^2 + \xi^2}\boldsymbol{v}_i = \sum_{i=1}^{m} \frac{\beta_i}{\lambda_i}\boldsymbol{v}_i$$

となる．したがって，$\{\lambda_1, \ldots, \lambda_m\}$ のうち，ある λ_i が 0 に近い非常に小さい値を取るとき，ノイズ $\beta_i \boldsymbol{v}_i$ の寄与が非常に大きくなり，推定が不安定になってしまう．

$\xi^2 > 0$ のときは，

$$\lim_{\lambda_i \to 0} \frac{\lambda_i \beta_i}{\lambda_i^2 + \xi^2} = 0$$

となって，ノイズ $\beta_i \boldsymbol{v}_i$ が推定に寄与しなくなり，推定が安定する．

特異値分解法による逆問題解析

なお，先程は観測ノイズの寄与について議論したが，推定式

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + (\mathbf{H}^\top\mathbf{H} + \xi^2\mathbf{I})^{-1}\mathbf{H}^\top(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b)$$

は，$\mathbf{H}$ を

$$\mathbf{H} = \sum_{i=1}^{m} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^\top,$$

と特異値分解し，$\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}_b$ を

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

と展開すると，以下のようにも書ける:

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

特異値分解法による逆問題解析

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

において $\xi^2 = 0$ とおくと，

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

となり，これが最小二乗法の解を与える．ただ，ある λ_i が非常に小さい値を取る場合，$\hat{x}$ が不安定になるので，以下のように特異値の小さい成分を落とした推定値を使うこともある [e.g., Press et al., Numerical Recipes]:

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}}_b + \sum_{i=1}^{l} \frac{\alpha_i}{\lambda_i} \boldsymbol{v}_i$$

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

ベイズの定理に基づく推定

ここでは，ベイズの定理

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

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，

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

とする．上の式に代入すると，以下のようになる:

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

ベイズの定理に基づく推定

$\boldsymbol{x}$ を推定するには，事後分布

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

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も同じ．)

exp の指数部分に着目すると，目的関数

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

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

ここでは，$\mathbf{P}_b$，$\mathbf{R}$ 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

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

は，ベイズの定理から得た

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

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．

単なる拘束付き最小二乗法と比較すると，$\mathbf{P}_b$ の中にも事前情報を取り入れられるのが利点．

ベイズの定理に基づく推定

J を変形すると

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

但し，

- A は x によらない定数，
- $\hat{\boldsymbol{x}}$ は，$(\mathbf{P}_b^{-1} + \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H})\hat{\boldsymbol{x}} = \mathbf{P}_b^{-1}\bar{\boldsymbol{x}}_b + \mathbf{H}^\top \mathbf{R}^{-1}\boldsymbol{y}$ を満たすベクトル．

したがって，$p(\boldsymbol{x}|\boldsymbol{y}) \propto e^{-J/2}$ は，平均ベクトル $\hat{\boldsymbol{x}}$，分散共分散行列 $(\mathbf{P}_b^{-1} + \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H})^{-1}$ のガウス分布になることが言える．

ベイズの定理に基づく推定

$\hat{\boldsymbol{x}}$ については，

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

より，以下のように求められる:

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

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

となることに注意．

ベイズの定理に基づく推定

分散共分散行列については，

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

とも変形できる．このことは，

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

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，

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

は，

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

を代入すると，

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

と変形できる．

- $\dim \boldsymbol{x} > \dim \boldsymbol{y}$ 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:

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

- 分散共分散行列:

$$
\begin{aligned}
\mathbf{P}_{\mathrm{est}} &= (\mathbf{P}_b^{-1} + \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H})^{-1} \\
&= \mathbf{P}_b - \mathbf{P}_b\mathbf{H}^\top(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_b
\end{aligned}
$$

事前分布の設計

ベイズ推定の式を使用する際には，$\mathbf{P}_b$ や $\mathbf{R}$ の設計が問題になる．

$\mathbf{P}_b = \sigma_b^2\mathbf{I}$, $\mathbf{R} = \sigma_R^2\mathbf{I}$ のような対角行列を使うことも多い．($\mathbf{P}_b$ が対角の場合も，先に述べたように解の安定化に寄与する．)

$\mathbf{P}_b$ については，時空間データの場合，ガウス関数など，グリッド点間の距離の関数で与えることもある．例えば，ベクトル $\boldsymbol{x}$ の要素 x_i が，位置 $\boldsymbol{z}_i$ で取得された何らかの物理量だとすると，

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

のような関数で，$x_i(\boldsymbol{z}_i)$ と $x_j(\boldsymbol{z}_j)$ の共分散を与える．

事前分布の設計

空間的な滑らかさを条件として加えるという考え方もある．

例えば，B-spline で節点を等間隔に配置し，節点を $0, h, 2h, \cdots, (K-1)h$ とした場合，$i = 1, \ldots, K-2$ について，2 階差分の 2 乗

$$\left(\frac{\beta_{i+1} - \beta_i}{2} - \frac{\beta_i - \beta_{i-1}}{2}\right)^2 = \left(\frac{\beta_{i-1} + \beta_{i+1}}{2} - \beta_i\right)^2$$

が小さくなるような条件を課す．端点 $i = 0, K-1$ についても，$(\beta_0 - \beta_1)^2$, $(\beta_{K-1} - \beta_{K-2})^2$ が小さくなるような条件を追加し，観測に近づける条件と合わせて目的関数の形でまとめると，

$$\begin{aligned} J &= (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})^{\mathsf{T}} \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H}\boldsymbol{x}) \\ &\quad + \xi^2 \left[(\beta_0 - \beta_1)^2 + (\beta_{K-1} - \beta_{K-2})^2 + \sum_{i=2}^{K-2} \left(\beta_i - \frac{\beta_{i-1} + \beta_{i+1}}{2} \right)^2 \right]. \end{aligned}$$

事前分布の設計

総和記号の各要素は，

$$\left(\beta_i - \frac{\beta_{i-1}+\beta_{i+1}}{2}\right)^2 = \begin{pmatrix}\beta_{i-1} & \beta_i & \beta_{i+1}\end{pmatrix}\begin{pmatrix}1/4 & -1/2 & 1/4 \\ -1/2 & 1 & -1/2 \\ 1/4 & -1/2 & 1/4\end{pmatrix}\begin{pmatrix}\beta_{i-1} \\ \beta_i \\ \beta_{i+1}\end{pmatrix}$$

$$= \boldsymbol{x}^\top \begin{pmatrix} O & & \cdots & & O \\ & 1/4 & -1/2 & 1/4 & \\ \vdots & -1/2 & 1 & -1/2 & \vdots \\ & 1/4 & -1/2 & 1/4 & \\ O & & \cdots & & O \end{pmatrix} \boldsymbol{x}$$

のように $\boldsymbol{x} = (\beta_1 \quad \cdots \quad \beta_{K-1})$ **の** 2 **次形式の形に書き直すことができる．これを組み合わせると，以下を満たす** (**半正定値**) **対称行列** $\mathbf{A}$ **を構成できる**:

$$(\beta_0 - \beta_1)^2 + (\beta_{K-1} - \beta_{K-2})^2 + \sum_{i=2}^{K-2}\left(\beta_i - \frac{\beta_{i-1}+\beta_{i+1}}{2}\right)^2 = \boldsymbol{x}^\top \mathbf{A} \boldsymbol{x}.$$

事前分布の設計

$\mathbf{A}$ を用いると先程の目的関数は，

$$J = (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})^\top \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H}\boldsymbol{x}) + \xi^2 \boldsymbol{x}^\top \mathbf{A} \boldsymbol{x}.$$

ただ，$\mathbf{A}$ のままでは逆行列を持たないので，微小な対角行列 $\delta^2 \mathbf{I}$ を足して，

$$J = (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})^\top \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H}\boldsymbol{x}) + \boldsymbol{x}^\top (\xi^2 \mathbf{A} + \delta^2 \mathbf{I}) \boldsymbol{x}$$

とした方が扱いやすい場合も多い．ここで，$\mathbf{P}_b = \xi^2 \mathbf{A} + \delta^2 \mathbf{I}$ とおくと，

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

こうして得られた $\mathbf{P}_b$ は事前分布の分散共分散行列と見なすことができる．

事前分布の設計

- $\mathbf{P}_b = \xi^2\mathbf{A} + \delta^2\mathbf{I}$ としたとき，δ は微小量で重要でないとしても，ξ が可変パラメータとして残る．また，$\mathbf{R}$ も可変である．
- 拘束付き最小二乗法の目的関数

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

においても，やはり ξ は可変パラメータである．

こうしたパラメータを決めるには，周辺尤度の最大化という方法がある．

周辺尤度

パラメータをまとめてベクトル $\boldsymbol{\theta}$ で表すと，$\boldsymbol{\theta}$ の周辺尤度は以下のように表される:

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

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

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

とすると，

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\mathbf{H}\bar{\boldsymbol{x}}_b)^\top(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{R})^{-1}(\boldsymbol{y}-\mathbf{H}\bar{\boldsymbol{x}}_b)\right)$$

となる．右辺に $\boldsymbol{\theta}$ が表に出てこないが，P_b や R の各要素などをまとめたものが $\boldsymbol{\theta}$ になっている．