051018版

推定問題の基礎

確率・統計の基礎から正則化まで

- 確率変数・確率分布の基礎
- 最尤原理と最尤推定
- 最小二乗法とその解の特性
- 正則化とミニマムノルム解
- L1ノルムの解と，解のスパースネスについての議論

参考文献：「ベイズ信号処理」（共立出版）
「統計的信号処理」（共立出版）

確率変数，確率分布，期待値，分散

- それが取る各値に対しそれぞれ確率が与えられている変数を確率変数と呼ぶ.
- 確率変数 x が連続値をとり，

$$P(a \le x \le b) = \int_a^b p(x)dx$$

なる$p(x)$が存在するとき，xは連続型の確率分布を持つといわれる．$p(x)$はxの確率密度分布と呼ばれる.

- 確率変数の期待値と分散

期待値: $E(x) = \int_{-\infty}^{\infty} xp(x)dx = \mu$　　分散: $V(x) = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = E[(x-\mu)^2]$

- 多（2）変数の確率分布: $P(a \le x \le b, c \le y \le d) = \int_c^d \int_a^b p(x,y)dxdy$
- 周辺化：同時確率分布から一方の確率変数を積分消去してもう一方のみの確率分布とすること

例: $g(x) = \int_{-\infty}^{\infty} p(x,y)dy$, および $h(y) = \int_{-\infty}^{\infty} p(x,y)dx$

共分散

確率変数 x と y に対して以下を共分散と呼ぶ.

$$Cov(x,y) = E[(x-\mu_x)(y-\mu_y)], \quad \text{ここで}, \mu_x = E(x),\ \mu_y = E(y)$$

以下が成立:

$$V(x+y) = V(x) + V(y) + 2Cov(x,y)$$
$$Cov(x,y) = E(xy) - E(x)E(y)$$

確率変数の独立:

確率変数 x と y に対し, $p(x,y) = p_1(x)p_2(y)$となるとき, x と y は独立である.

確率変数 x と y が独立な場合以下が成立.

$$Cov(x,y) = 0$$
$$V(x+y) = V(x) + V(y)$$
$$E(xy) = E(x)E(y)$$

N個の確率変数 $x_1,\ldots,x_N$ に対し，以下が成立．

$$E(x_1+\cdots+x_N)=E(x_1)+\cdots+E(x_N)$$

さらに，$x_1,\ldots,x_N$ が独立なら，

$$V(x_1+\cdots+x_N)=V(x_1)+\cdots+V(x_N)$$

N個の確率変数 $x_1,\ldots,x_N$ が独立で同一な分布，

$$E(x_1)=E(x_2)=\cdots=E(x_N)=\mu,\ V(x_1)=V(x_2)\cdots=V(x_N)=\sigma^2$$

に従う場合，

$$E(x_1+\cdots+x_N)=N\mu \text{ および } V(x_1+\cdots+x_N)=N\sigma^2 \text{ である．}$$

確率変数 $x_1,\ldots,x_N$ の算術平均：

$$\overline{x}=\frac{x_1+\cdots+x_N}{N}$$

に対し，

$$E(\overline{x})=\mu,\qquad V(\overline{x})=\frac{\sigma^2}{N} \text{ が成立．}$$

独立で同一な分布（independently, identically distributed)—I.I.D.と略す場合がある．

正規分布

確率密度: $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$

$E(x) = \int_{-\infty}^{\infty} xp(x)dx = \mu$, および $V(x) = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = \sigma^2$ は簡単に確認できる.

簡略的な表記法: $p(x) = N(x \mid \mu, \sigma^2)$

確率変数 x が,期待値μ, 分散σ^2の正規分布する. $\Rightarrow$ $x \sim N(x \mid \mu, \sigma^2)$

正規分布における重要で便利な性質

1) $y = ax + b$なる確率変数 y に対し, $x \sim N(x \mid \mu, \sigma^2)$ なら $y \sim N(y \mid a\mu + b, a^2\sigma^2)$

2) $x \sim N(x \mid \mu_1, \sigma_1^2)$, $y \sim N(y \mid \mu_2, \sigma_2^2)$で x と y が独立なら

$$z = x + y \sim N(z \mid \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

N個の独立な確率変数 $x_1, x_2, \ldots, x_N$ に対し，

（1）$x_1 \sim N(x_1 \mid \mu_1, {\sigma_1}^2), x_2 \sim N(x_2 \mid \mu_2, {\sigma_2}^2), \cdots, x_N \sim N(x_N \mid \mu_N, {\sigma_N}^2)$なら，

$$z = x_1 + x_2 + \cdots + x_N \sim N(z \mid \mu_1 + \mu_2 + \cdots + \mu_N, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_N^2)$$

（2）$x_1, x_2, \cdots, x_N$ が独立で同一な分布 $x_j \sim N(x_j \mid \mu, \sigma^2)$ をする場合，

$$z = x_1 + x_2 + \cdots + x_N \sim N(z \mid N\mu, N\sigma^2)$$

（3）$x_1, x_2, \cdots, x_N$が独立で同一な分布（正規分布に限らず）をする場合，

$$z = x_1 + x_2 + \cdots + x_N \xrightarrow[N \to \infty]{} N(z \mid N\mu, N\sigma^2)$$ （中心極限定理）

- 実際の観測においてデータに重畳しているノイズは，独立な複数の不規則信号の総和と考えることができる．
- 中心極限定理はデータに重畳しているノイズの確率分布に便宜的に正規分布を仮定する事の妥当性の（ある程度の）根拠となる．

ベクトル型確率変数

確率変数$x_1,\ldots,x_N$に対して，列ベクトル：$\boldsymbol{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$を用いて確率変数$x_1,\ldots,x_N$全体を表す．

これをベクトル型確率変数と呼ぶ．

平均ベクトル：
$$E(\boldsymbol{x})=\begin{bmatrix} E(x_1) \\ \vdots \\ E(x_N) \end{bmatrix}=\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}=\boldsymbol{\mu}$$

共分散行列：
$$E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T]=\begin{bmatrix} E[(x_1-\mu_1)^2] & E[(x_1-\mu_1)(x_2-\mu_2)] & \cdots & E[(x_1-\mu_1)(x_N-\mu_N)] \\ E[(x_2-\mu_2)(x_1-\mu_1)] & \cdot & \cdot & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ E[(x_N-\mu_N)(x_1-\mu_1)] & \cdot & \cdots & E[(x_N-\mu_N)^2] \end{bmatrix}=\boldsymbol{\Sigma}$$

確率変数$x_1,\ldots,x_N$が独立で同一な分布をする場合，$Cov(x_i,x_j)=0,\ V(x_j)=\sigma^2$ であるので，

$$\boldsymbol{\Sigma}=\begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}=\sigma^2\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}=\sigma^2\boldsymbol{I}$$

多次元正規分布

1次元（スカラー）正規分布：$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$

多次元（ベクトル型）正規分布：$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\left|\boldsymbol{\Sigma}\right|^{1/2}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]$

（N はベクトル $\boldsymbol{x}$ のサイズ）

確率変数$x_1, x_2, \ldots, x_N$が独立で同一な分布（I.I.D)をする場合，

$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ であるので，$\left|\boldsymbol{\Sigma}\right| = (\sigma^2)^N$ より，

$$
\begin{aligned}
p(\boldsymbol{x}) &= \frac{1}{(2\pi)^{N/2}(\sigma^2)^{N/2}} \exp[-\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})] \\
&= \frac{1}{(2\pi)^{N/2}(\sigma^2)^{N/2}} \exp[-\frac{1}{2\sigma^2}\sum_{j=1}^{N}(x_j-\mu_j)^2] \\
&= \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2\sigma^2}(x_j-\mu_j)^2]
\end{aligned}
$$

多次元正規分布は独立な1次元正規分布の積として表される.

推定

N 回の繰り返し観測で，観測データ $x_1, x_2, \ldots, x_N$ を得るとし，これら観測データを用いて未知量 θ の推定結果 $\hat{\theta}$ を得るとする．

観測データ $x_1, x_2, \ldots, x_N$ は確率変数であるので，これらから求まる $\hat{\theta}$ も確率変数である．したがって，$x_1, x_2, \ldots, x_N$ の値に応じて確率的な値を取る．

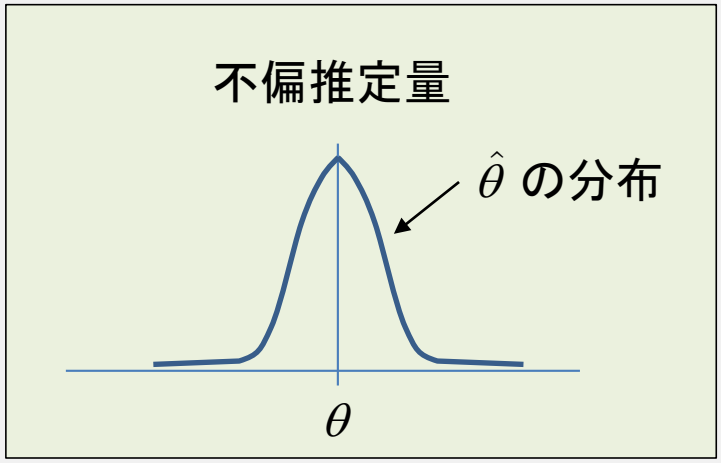
推定の行ない方(推定量)の良さを表す指標

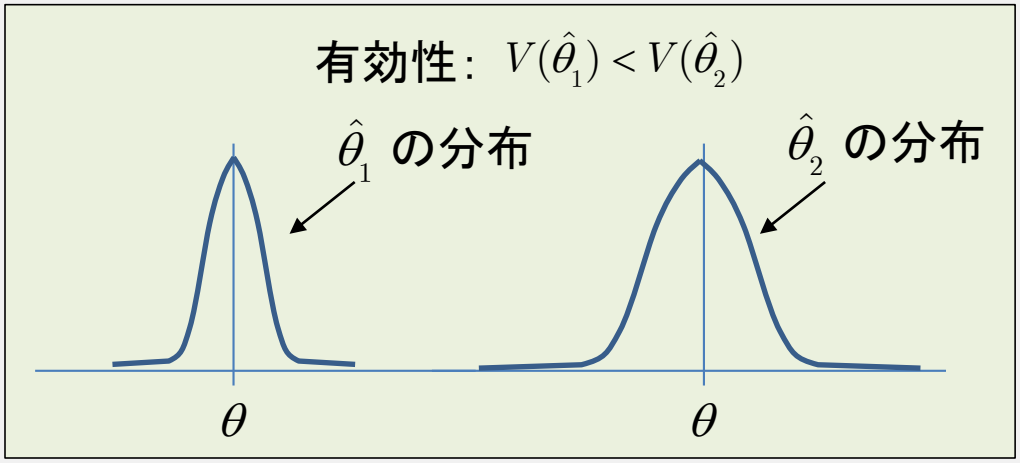
（1）不偏性：$E(\hat{\theta}) = \theta$

（2）有効性：$V(\hat{\theta}_1) < V(\hat{\theta}_2)$

ある推定の行ない方で得た解：$\hat{\theta}_1$

別の推定の行ない方で得た解：$\hat{\theta}_2$





最尤推定法

N 回の繰り返し観測で未知量 θ を推定する

最尤原理

N回の繰り返し観測で，観測値$x_1, x_2, \ldots, x_N$が得られた．

⇓

得られた観測結果$x_1, x_2, \ldots, x_N$は「確率最大のものが実現した．」

すなわち「最も起こりやすいことが起きた．」結果と考える．

N回の繰り返し観測，確率変数$x_1, \ldots, x_N$は独立で，$x_j \sim p(x_j)$とすれば，

観測結果$x_1, \ldots, x_N$が実現する確率：

$$p(x_1, x_2, \ldots, x_N) = \prod_{j=1}^{N} p(x_j)$$

確率分布は未知量θをパラメータとして含む．

尤度関数（likelihood function）：$L(\theta) = \prod_{j=1}^{N} p(x_j)$

最尤推定法：$\hat{\theta} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \log L(\theta)$

対数尤度関数（log-likelihood function）

N回の繰り返し観測を用いた最も単純な観測モデル：

$$
\begin{aligned}
x_1 &= \theta + \varepsilon_1 \\
x_2 &= \theta + \varepsilon_2 \\
&\vdots \\
x_N &= \theta + \varepsilon_N
\end{aligned}
$$

$$
\varepsilon_j \sim N(\varepsilon_j \mid 0, \sigma^2)
$$
$$
\Downarrow
$$
$$
x_j \sim N(x_j \mid \theta, \sigma^2)
$$

尤度関数：$L(\theta) = \prod_{j=1}^{N} p(x_j) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(x_j - \theta)^2}{2\sigma^2}]$

⇩

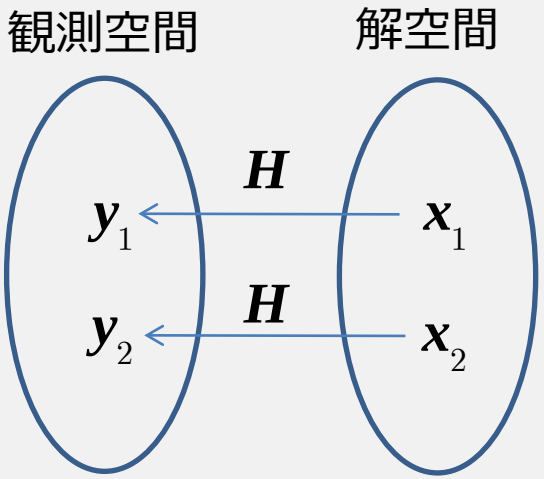
最尤推定解：$\hat{\theta} = \arg\max_{\theta} \log \prod_{j=1}^{N} p(x_j) = \arg\max_{\theta} \left[\sum_{j=1}^{N} -\frac{(x_j - \theta)^2}{2\sigma^2} \right] + C$

⇩

$\frac{\partial}{\partial\theta} \left[\sum_{j=1}^{N} -\frac{(x_j - \theta)^2}{2\sigma^2} + C \right] = 0$ とおいて，最尤推定解 $\hat{\theta} = \frac{1}{N} \sum_{j=1}^{N} x_j$ を得る．

線形離散モデル

未知量 : $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$, 観測データ : $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}$ に



線形な関係 :

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

が存在する. これに加法的にノイズ $\boldsymbol{\varepsilon}$ が重畳するとして

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

を観測データのモデル（線形離散モデル）と呼ぶ.

「未知量 $\boldsymbol{x}$ が原因で観測結果$\boldsymbol{y}$を生じたと解釈することもできる．」

原因 $\boldsymbol{x}$ を与えて結果 $\boldsymbol{y}$ を推定する ⟶ 順（方向）問題 （forward problem)
(行列$\boldsymbol{H}$ を推定する問題に等しい．)

結果 $\boldsymbol{y}$ を与えて原因 $\boldsymbol{x}$ を推定する ⟶ 逆（方向）問題 （inverse problem)

$M > N \Rightarrow$ 逆問題は優決定（over-determined),
$M < N \Rightarrow$ 逆問題は劣決定 (under-determined)と呼ばれる.

線形最小二乗法

モデル：$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$ の基で，観測データ $\boldsymbol{y}$ から未知な量 $\boldsymbol{x}$ を推定することを考える．

前提： $\boldsymbol{H}$ は既知である．

最尤推定を行う

ノイズ確率分布： $p(\boldsymbol{\varepsilon}) = N(\boldsymbol{\varepsilon} \mid 0, \sigma^2 \boldsymbol{I})$を仮定すれば，

観測データの確率分布：

$$p(\boldsymbol{y}) = N(\boldsymbol{y} \mid \boldsymbol{H}\boldsymbol{x}, \sigma^2 \boldsymbol{I}) = \frac{1}{(2\pi)^{M/2}(\sigma^2)^{M/2}} \exp[-\frac{1}{2\sigma^2}\left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2]$$

対数尤度関数： $\log L(\boldsymbol{x}) = \log p(\boldsymbol{y}) = -\frac{1}{2\sigma^2}\left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2 + C$

線形最小二乗法（２）

対数尤度関数を最大とする $\boldsymbol{x}$ は，以下の F を最小とする $\boldsymbol{x}$ に等しい.

$$F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2$$

この F を最小二乗のコスト関数と呼ぶ.

最尤推定解： $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} F$ である.

最小二乗のコスト関数 F を最小にすることで未知量を求める手法を最小自乗法と呼ぶ.

最小二乗のコスト関数の解釈：

$$F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2$$

推定した $\boldsymbol{x}$ のデータ $\boldsymbol{y}$ への一致度（適合度）と解釈できる.

F を最小にする解はデータに最もよく合った（データを最もよく「説明する」）解である.

最小二乗解の導出

コスト関数: $F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2 = (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}) = \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}$

$\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y} = \boldsymbol{H}^T\boldsymbol{y}$, $\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{y}^T\boldsymbol{H}\boldsymbol{x} = \boldsymbol{H}^T\boldsymbol{y}$, $\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x} = 2\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}$ を考慮すれば,

$\frac{\partial}{\partial \boldsymbol{x}} F = 2\left(-\boldsymbol{H}^T\boldsymbol{y} + \boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}\right) = 0$ となり,

したがって，以下の最小二乗解を得る.

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{y}$$

データにある重み行列を掛けて未知量を推定する方法を線形推定法と総称する.

最小二乗解は重み行列： $\boldsymbol{W}^T = \left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T$ を用いた線形推定法である.

繰り返し計測モデルでの最小二乗解

$$
\begin{aligned}
y_1 &= \theta + \varepsilon_1 \\
y_2 &= \theta + \varepsilon_2 \\
&\vdots \\
y_M &= \theta + \varepsilon_M
\end{aligned}
\quad \Rightarrow \quad
\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \theta + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{bmatrix}
\quad \Rightarrow \quad
\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}, \ \boldsymbol{H} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}
$$

したがって,

$$
\hat{\boldsymbol{x}} = (\boldsymbol{H}^T \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{y} = \left(\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \frac{1}{M} \sum_{j=1}^{M} y_j
$$

$\boldsymbol{H}^T$ $\boldsymbol{H}$ $\boldsymbol{H}^T$

線形推定解と不偏性

線形推定量： $\hat{\boldsymbol{x}} = \boldsymbol{W}^T \boldsymbol{y} = \boldsymbol{W}^T (\boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}) = \boldsymbol{W}^T \boldsymbol{H}\boldsymbol{x} + \boldsymbol{W}^T \boldsymbol{\varepsilon}$

ノイズの影響

不偏推定量となる条件： $\boldsymbol{W}^T \boldsymbol{H} = \boldsymbol{I}$

線形不偏推定量： $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{W}^T \boldsymbol{\varepsilon}$

線形不偏推定量においては推定結果に含まれる誤差は観測データに含まれるノイズの影響のみである．SN比が大きい極限で推定解は真の解に等しくなる．

最小二乗解の場合

$\boldsymbol{W}^T \boldsymbol{H} = (\boldsymbol{H}^T \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{H} = \boldsymbol{I}$　したがって，　$\hat{\boldsymbol{x}} = \boldsymbol{x} + (\boldsymbol{H}^T \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{\varepsilon}$

ノイズの影響

線形不偏推定量であり，ノイズゼロの極限で推定解は真の解に一致する

最小二乗解の分散

$$\boldsymbol{\Sigma}_x = E\left[(\hat{\boldsymbol{x}}-\boldsymbol{x})(\hat{\boldsymbol{x}}-\boldsymbol{x})^T\right] = E\left[(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{\varepsilon}[(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{\varepsilon}]^T\right]$$

$$= E\left[(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\right] = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1} = \sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1}$$

ここで，$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\boldsymbol{I}$ を用いた.

$(\boldsymbol{H}^T\boldsymbol{H})^{-1}$:ノイズゲインと呼ばれる.

最小二乗解が線形不偏推定量の中で最も分散の小さな推定量ー有効推定量であることを示すことができる．最小二乗解はBest linear unbiased estimator (BLUE)であると言われる.

最小自乗法はすばらしい方法のように思えるが??

すばらしい方法かどうかは順方向行列 $\boldsymbol{H}$ の特性による.

最小二乗解のノイズ耐性

順方向行列$\boldsymbol{H}$の特異値分解（SVD）　over-determined$(M > N)$の場合

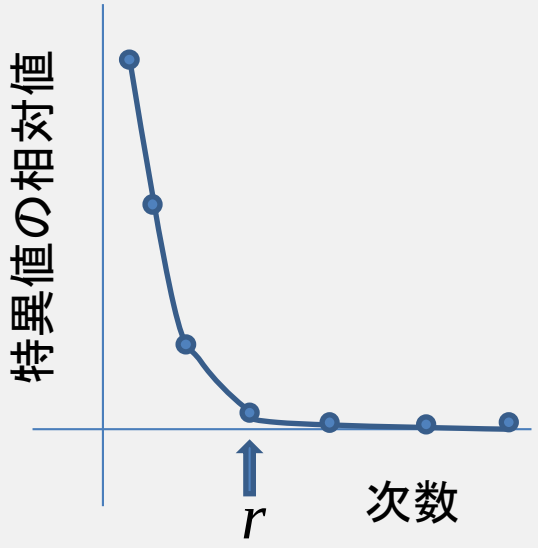
$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_M \end{bmatrix} \begin{bmatrix} \gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_N \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_N^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_N \end{bmatrix} \begin{bmatrix} \gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_N \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_N^T \end{bmatrix} = \sum_{j=1}^{N} \gamma_j \boldsymbol{u}_j \boldsymbol{v}_j^T$$

特異値$\gamma_1, \ldots, \gamma_N$（$\geq 0$）は大きさの順に番号付けされているとする.

ある次数以上の特異値が非常に小さく，ゼロに近くなることがしばしば起る.

つまり，特異値を大きさの順に並べると，

r 次数以上の特異値が非常に小さく，ゼロに近くなることが起る



最小二乗解のノイズ耐性ー続き

ある次数から先の特異値が非常に小さくなる場合に最小二乗解にどんな影響が出るであろうか.

$\boldsymbol{H}$ の特異値と特異値ベクトルを用いて最小二乗解を記述する.

最小二乗推定の重み： $(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T = \sum_{j=1}^{N}\frac{1}{\gamma_j}\boldsymbol{v}_j\boldsymbol{u}_j^T$

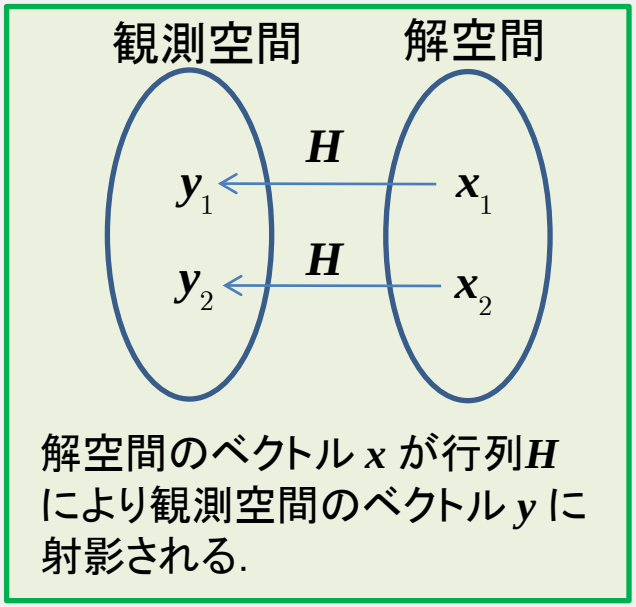
最小二乗解： $\hat{\boldsymbol{x}} = \boldsymbol{x} + (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{\varepsilon} = \boldsymbol{x} + \left[\sum_{j=1}^{N}\frac{1}{\gamma_j}\boldsymbol{v}_j\boldsymbol{u}_j^T\right]\boldsymbol{\varepsilon} = \boldsymbol{x} + \sum_{j=1}^{N}\frac{\left(\boldsymbol{u}_j^T\boldsymbol{\varepsilon}\right)}{\gamma_j}\boldsymbol{v}_j$

ある次数以上の特異値が非常に小さくなる場合，その特異値を含む項はノイズの影響を増幅してしまう.

結果として： $\boldsymbol{x} \ll \sum_{j=1}^{N}\frac{\left(\boldsymbol{u}_j^T\boldsymbol{\varepsilon}\right)}{\gamma_j}\boldsymbol{v}_j$ 最小二乗解はノイズに起因した大きな誤差を含む.

それでは，どんな場合にある次数以上の特異値がゼロに近くなるのか？

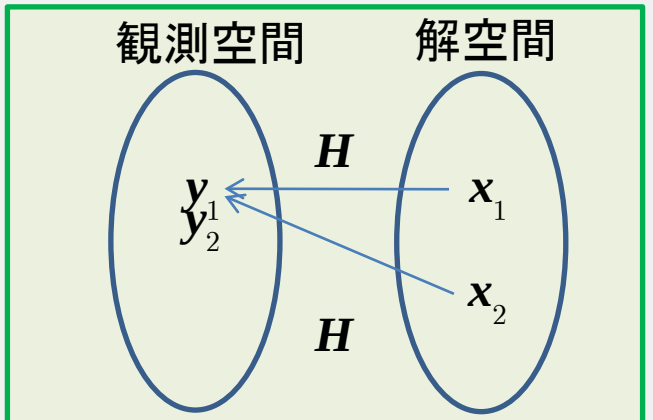
推定に対する更なる考察



解空間のベクトル x が行列 H により観測空間のベクトル y に射影される.

$$y_1 = Hx_1$$

$$y_2 = Hx_2$$



かなり異なる x が似通ったyに射影される.

x_1 と x_2 がかなり異なるにも関わらず, $y_1 \approx y_2$ となる場合を考える.

⇓

y_1 と y_2 のわずかな違いから, x_1 なのか x_2 なのかを推定しなければならない.

⇓

こんな場合，未知量 x をノイズの重畳した観測データ y から推定するのは難しい

⇓

このむずかしさを反映したのが，ある次数以上の特異値がゼロに近くなることである.

それでは，ある次数以上の特異値がゼロに近くなるような，順方向行列の場合に，推定はどのように行えばよいか？

データに含まれる誤差（ノイズ）に対してその影響を受けにくい（robust）な方法が望ましい!

正則化の導入

正則化を用いた最小二乗解

最小二乗解：$F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2$ とおいたF を最小とする $\boldsymbol{x}$ を求める．

⇓

求まる推定解は観測データ $\boldsymbol{y}$ に「最もよく一致する」 $\boldsymbol{x}$ である．

⇓

しかし，$\boldsymbol{y}$ に大きなノイズが重畳している場合，観測データとの一致度をあまり追求しても意味がなく，返って観測データに含まれるノイズの影響を受けてしまう．

$\left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2$ のみでなく別の基準$\phi(\boldsymbol{x})$を導入し，コスト関数を以下のように定義する：

$$F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2 + \xi\phi(\boldsymbol{x})$$

$\phi(\boldsymbol{x})$ ：制約条件．解として望ましい性質を関数に表したもの．

ξ ：正則化定数．正の定数で，第1項（データとの一致度）と第2項（制約条件）のバランスを取る．

$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} F$から解を求めれば，推定解がノイズの影響を受けにくくなることが期待できる．

正則化を用いた最小二乗解（２）

解のノルム $\phi(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ がよく用いられる.

$F = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \xi\|\boldsymbol{x}\|^2$ として，$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} F$ から解を求めれば以下の公式を得る：

$$\hat{\boldsymbol{x}} = (\boldsymbol{H}^T\boldsymbol{H} + \xi\boldsymbol{I})^{-1}\boldsymbol{H}^T\boldsymbol{y}$$

解のノイズ耐性

$$\hat{\boldsymbol{x}} = (\boldsymbol{H}^T\boldsymbol{H} + \xi\boldsymbol{I})^{-1}\boldsymbol{H}^T\boldsymbol{y} = \left[\sum_{j=1}^{N}\frac{\gamma_j^2}{\gamma_j^2+\xi}\boldsymbol{v}_j\boldsymbol{v}_j^T\right]\boldsymbol{x} + \sum_{j=1}^{N}\frac{\gamma_j}{\gamma_j^2+\xi}\left(\boldsymbol{u}_j^T\boldsymbol{\varepsilon}\right)\boldsymbol{v}_j$$

正の定数ξが分母に含まれるため，小さなγ_jを含む項は小さくなり，ノイズ増幅を回避できる.

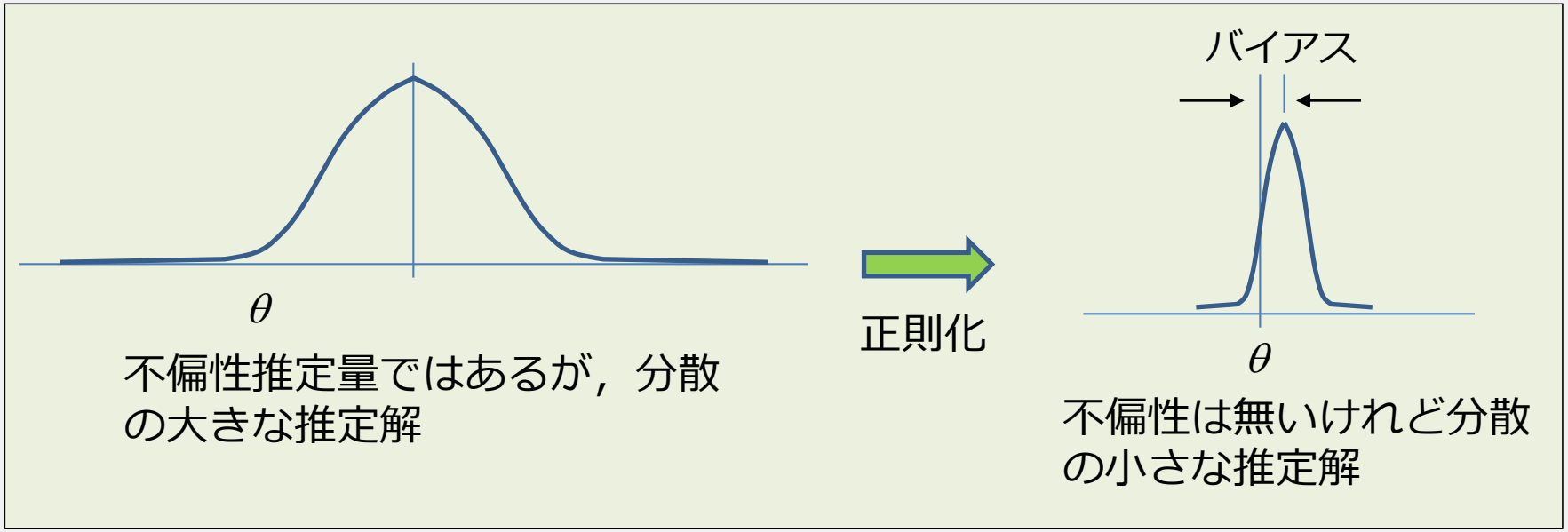
第1項は $\boldsymbol{x}$ とは異なるため，この解は不偏性を持たない．ただし：

$\left[\sum_{j=1}^{N}\frac{\gamma_j^2}{\gamma_j^2+\xi}\boldsymbol{v}_j\boldsymbol{v}_j^T\right]\boldsymbol{x} \xrightarrow[\xi\to 0]{} \left[\sum_{j=1}^{N}\boldsymbol{v}_j\boldsymbol{v}_j^T\right]\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}$ であるので，ξ→0で不偏推定解となる.

正則化とは解の普遍性を犠牲にして，解のノイズ耐性を向上させる技術であり，
正則化定数 ξ は，解のノイズ耐性と解のバイアスのトレードをコントロールする.

不偏性と有効性の観点から見た正則化

- 正則化とは，不偏性を持つが分散の大きな推定解を，不偏は持たないが分散の小さな推定量へ変換する技術.



$F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2 + \xi\phi(\boldsymbol{x})$ における正則化定数 ξ はどのように決めるか

- 正則化定数ξは観測データに含まれるノイズの量に依存する．ノイズが小さければ ξを小さくして観測データとの一致度を優先させることが合理的であるし，ノイズが大きければ観測データとの一致度にこだわっても意味がないため，大きな ξを用いることがノイズ増幅率の点からも合理的である．

- 多くの場合 ξの値は経験的に決められている．つまり，ξのいろいろな値で解を計算してみて最も妥当な解が得られるξを採用する．実用的にはこのやり方で特に問題がない場合が多い．

- ξの理論的な決め方についてはベイズ推定を用いた方法を後に紹介する．

劣決定系における推定解の求め方:

線形離散モデル

未知量: $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$, 観測データ: $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}$ に

線形な関係:

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

が存在する. これに加法的にノイズ$\boldsymbol{\varepsilon}$が重畳するとして

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

を観測データのモデル(線形離散モデル)と呼ぶ.

$M > N \Rightarrow$ 逆問題は優決定 (over-determined)

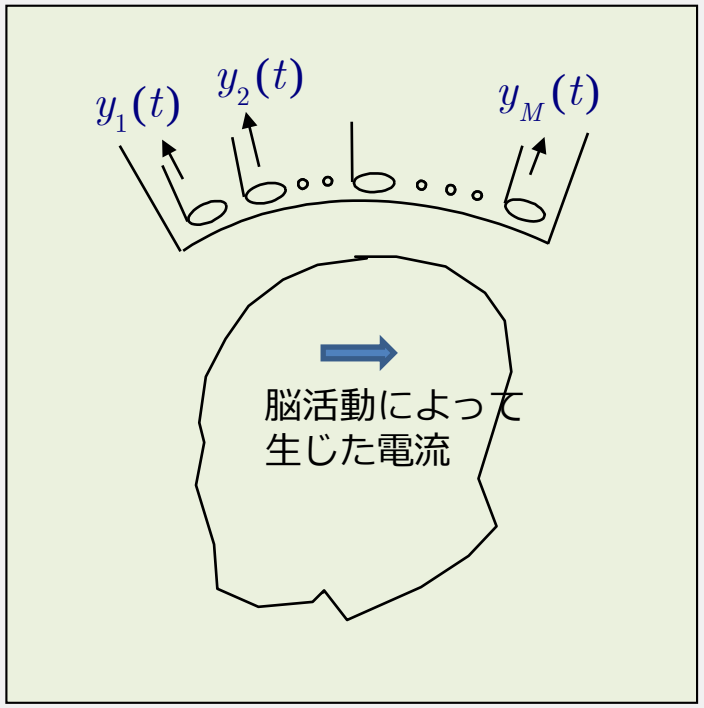
$M < N \Rightarrow$ 逆問題は劣決定 (under-determined)

生体磁気再構成問題は劣決定な推定問題として定式化される

データベクトル

$$\boldsymbol{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_M(t) \end{bmatrix}$$

M: センサー数

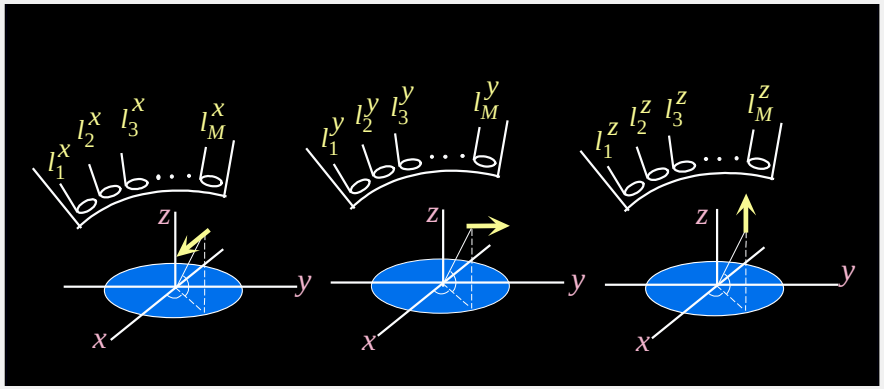


いかにして，データベクトル $\boldsymbol{y}(t)$ と，脳活動（厳密には脳活動によって生じた電流と）の関連を表現するか？

ソースと呼ぶ

ソース電流と計測データの関係を導く

センサーリードフィールドの導入



位置 $\boldsymbol{r}$ に 単位強度の電流源があり，

x 方向を向いている場合のセンサー出力：

$$l_1^x, l_2^x, \ldots, l_M^x$$

y 方向を向いている場合のセンサー出力：

$$l_1^y, l_2^y, \ldots, l_M^y$$

z 方向を向いている場合のセンサー出力：

$$l_1^z, l_2^z, \ldots, l_M^z$$

リードフィールド行列

以下の（M× 3）の行列はリードフィールド行列と呼ばれ，センサーアレイの位置 $\boldsymbol{r}$ における感度を表す．

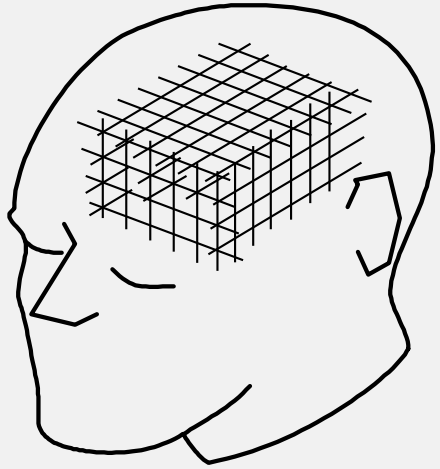
$$\boldsymbol{L}(\boldsymbol{r}) = \begin{bmatrix} l_1^x(\boldsymbol{r}) & l_1^y(\boldsymbol{r}) & l_1^z(\boldsymbol{r}) \\ \vdots & \vdots & \vdots \\ l_j^x(\boldsymbol{r}) & l_j^y(\boldsymbol{r}) & l_j^z(\boldsymbol{r}) \\ \vdots & \vdots & \vdots \\ l_M^x(\boldsymbol{r}) & l_M^y(\boldsymbol{r}) & l_M^z(\boldsymbol{r}) \end{bmatrix}$$

センサー・リードフィールドを用いてセンサーデータとソースの関係を表す

ソースが領域 Ω に連続的に分布している場合

（ Ωをソーススペースと呼ぶ）

$$\boldsymbol{y} = \int_{\Omega} \boldsymbol{L}(\boldsymbol{r})\boldsymbol{s}(\boldsymbol{r})d\boldsymbol{r}$$



ボクセルの導入：

$$\boldsymbol{y} = \int \boldsymbol{L}(\boldsymbol{r})\boldsymbol{s}(\boldsymbol{r})d\boldsymbol{r} = \sum_{j=1}^{N} \boldsymbol{L}(\boldsymbol{r}_j)\boldsymbol{s}(\boldsymbol{r}_j)$$

$$= \underbrace{\left[\boldsymbol{L}(\boldsymbol{r}_1), \quad \cdots, \quad \boldsymbol{L}(\boldsymbol{r}_N)\right]}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} \boldsymbol{s}(\boldsymbol{r}_1) \\ \vdots \\ \boldsymbol{s}(\boldsymbol{r}_N) \end{bmatrix}}_{\boldsymbol{x}} = \boldsymbol{H}\boldsymbol{x}$$

$\boldsymbol{H} = \left[\boldsymbol{L}(\boldsymbol{r}_1), \quad \cdots, \quad \boldsymbol{L}(\boldsymbol{r}_N)\right]$ ボクセルリードフィールド行列と呼ばれ，
センサー数 × 3 (ボクセル数) の行列である．

劣決定系における推定解の求め方：

劣決定系の場合，$\left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2 = 0$ となる，つまり，$\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}$ を満たす無数の$\boldsymbol{x}$ が存在する．

⇓

観測データとの一致度という要請だけでは解を決めることができない．

⇓

観測データとの一致度以外に解が持つ望ましい性質を組み込む．

⇓

$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$ を満たす$\boldsymbol{x}$ のなかで，最も望ましい性質を持つものを最適解とする．

望ましい性質として解のノルム最小を再び用いれば

$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\| \boldsymbol{x} \right\|^2$, subject to $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$ から $\boldsymbol{x}$ を求める．

ミニマムノルムの解： $\hat{\boldsymbol{x}} = \boldsymbol{H}^T \left(\boldsymbol{H}\boldsymbol{H}^T \right)^{-1} \boldsymbol{y}$ が得られる．

（最小二乗解：$\hat{\boldsymbol{x}} = \left(\boldsymbol{H}^T \boldsymbol{H} \right)^{-1} \boldsymbol{H}^T \boldsymbol{y}$）

ミニマムノルム解の性質

$$\hat{\boldsymbol{x}} = \boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{H}^T)^{-1}(\boldsymbol{H}\boldsymbol{x}+\boldsymbol{\varepsilon}) = \underbrace{\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{H}^T)^{-1}\boldsymbol{H}\boldsymbol{x}}_{\neq \boldsymbol{x}} + \boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{H}^T)^{-1}\boldsymbol{\varepsilon}$$

ミニマムノルム解は不偏性を持たない.

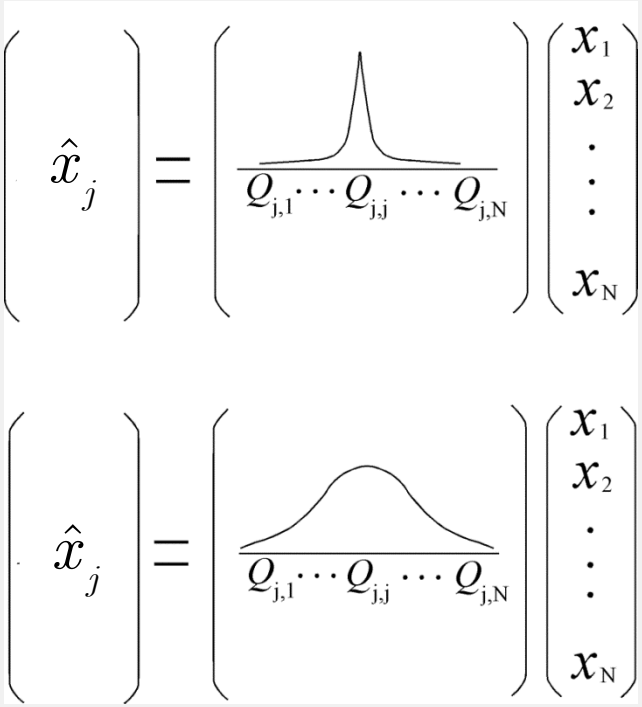
分解能行列

$\boldsymbol{Q} = \boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{H}^T)^{-1}\boldsymbol{H}$ は解の「良さ」を予測する.

（ノイズ項を無視して）

$$\hat{x}_j = \sum_{k=1}^{N} Q_{j,k} x_k$$

と表すことができる. $Q_{j,k}$がなるべくシャープなピークを構成するなら推定解$\hat{x}_j$は真の値x_jに近いものになる.



分解能行列の良さは位置バイアスを用いて定量化できる. →non-adaptive filterのスライド

データのノイズを考慮した場合のミニマムノルム解

ミニマムノルム解 : $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\|\boldsymbol{x}\right\|^2$ subject to $\underbrace{\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}}_{\Uparrow}$ から解を求める.

観測データに一致する $\boldsymbol{x}$

$\boldsymbol{y}$ にノイズが重畳している場合，データとの一致度をあまり追求しても意味がない.

$$\Downarrow$$

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\|\boldsymbol{x}\right\|^2 \quad \text{subject to} \quad \left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2 \le d \text{ から解を求める.}$$

$$\Downarrow$$

最適解は制約条件の境界に存在することが多く，以下とほとんど等価である.

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\|\boldsymbol{x}\right\|^2 \quad \text{subject to} \quad \left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2 = d$$

$$\Downarrow$$

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} F : \quad F = \xi\left\|\boldsymbol{x}\right\|^2 + \left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2$$

$$\Downarrow$$

$$\hat{\boldsymbol{x}} = \boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{H}^T + \xi\boldsymbol{I}\right)^{-1}\boldsymbol{y}$$（正則化ミニマムノルム解と呼ばれる）

ノルム最小の基準で解を求めることの意味

データとの一致度がある範囲内である解が複数ある場合，なるべく「小さな」解を選ぶという考え方

「オッカムのかみそり」と言われる

ウィキペディアより

オッカムの剃刀（オッカムのかみそり、Occam‘s razor）とは、「ある事柄を説明するためには、必要以上に多くを仮定するべきでない」とする指針。もともとスコラ哲学にあり、14世紀の哲学者・神学者のオッカムが多用したことで有名になった。様々なバリエーションがあるが、20世紀にはその妥当性を巡って科学界で議論が生じた。「剃刀」という言葉は、説明に不要な存在を切り落とすことを比喩しており、そのためオッカムの剃刀は思考節約の原理や思考節約の法則、思考経済の法則とも呼ばれる。またケチの原理と呼ばれることもある。

ある事柄を説明するのに複数の説明がある場合に，最も簡単な説明を選ぶ

重み付きノルムの解

$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x}$ subject to $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$ から解を求める.

最適推定解: $\hat{\boldsymbol{x}} = \boldsymbol{W}^{-1}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{W}^{-1}\boldsymbol{H}^T)^{-1}\boldsymbol{y}$

例: 滑らかさ最大の解

$x_1, \ldots, x_N$が画素のように空間的に隣り合った解とすれば, 解の滑らかさは

$$D = (x_1 - x_2)^2 + (x_2 - x_3)^2 + \cdots + (x_{N-1} - x_N)^2$$

で与えられる.

$$\begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \\ \vdots \\ x_{N-1} - x_N \end{bmatrix}$$

左辺の$(N-1) \times N$の行列を$\boldsymbol{A}$とおいて

$D = \|\boldsymbol{A}\boldsymbol{x}\|^2 = (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}$ であるので, $\boldsymbol{W} = \boldsymbol{A}^T\boldsymbol{A}$ とすれば,
解の滑らかさは重み付きノルム $\boldsymbol{x}^T\boldsymbol{W}\boldsymbol{x}$ で表されることが理解できる.

解の大きさを表すのに，他のやり方は??

ベクトル: $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$ のノルム

ベクトルの「大きさ」を表す“なんらか”の量

L_2-ノルム: $\left\|\boldsymbol{x}\right\|_2 = \sqrt{\sum_{j=1}^{N} x_j^2}$ （代表的）

L_∞-ノルム: $\left\|\boldsymbol{x}\right\|_\infty = \max(x_1, \ldots, x_N)$

L_1-ノルム: $\left\|\boldsymbol{x}\right\|_1 = \sum_{j=1}^{N} \left|x_j\right|$

L_p-ノルム: $\left\|\boldsymbol{x}\right\|_p = \sqrt[p]{\sum_{j=1}^{N} \left|x_j\right|^p}$

L_0-ノルム: $\left\|\boldsymbol{x}\right\|_0 = \sum_{j=1}^{N} I(x_j)$ where $I(x_j) = \begin{cases} 1 & x_j \neq 0 \\ 0 & x_j = 0 \end{cases}$

L_1 -ノルム正則化ミニマムノルム解

L_2 -ノルム正則化

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_{j=1}^{N} x_j^2 \quad \text{subject to} \quad \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2 \le d \text{ から解を求める.}$$

L_1 -ノルム正則化

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_{j=1}^{N} \left| x_j \right| \quad \text{subject to} \quad \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\|^2 \le d \text{ から解を求める.}$$

全く同じ理屈で最適化問題のコスト関数は以下のようになる.

$$\hat{\boldsymbol{x}} = \arg\min F : \quad F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\| + \xi \sum_{j=1}^{N} | x_j |$$

この最適化問題はクローズドフォームの解を持たない. 数値計算解を求める.

L_1 -ノルム正則化ミニマムノルム解はスパースな解を与える.

$\boldsymbol{x}$ の要素 $x_1, \ldots, x_N$ で，ほとんどのものがゼロ，わずかなものがノンゼロとなる解

スパース制約

なるべくスパースな解（ほとんどのボクセルがゼロ，わずかなボクセルがノンゼロの値を持つ解）を与えるような制約

機械学習やcompressed sensingなどの分野で用いられている.

スパースな解を与える最も直截的な方法：

L_0 -ノルム正則化ミニマムノルム解

ベクトル $\boldsymbol{x}$ のL_0ノルム: $\left\|\boldsymbol{x}\right\|_0 = \sum_{j=1}^{N} I(x_j)$

ここで, $I(x_j) = \begin{cases} 1 & x_j \neq 0 \\ 0 & x_j = 0 \end{cases}$

L_0 -ノルム正則化

$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_{j=1}^{N} I(x_j)$ subject to $\left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2 \leq d$ から解を求める.

ゼロでない要素の数

この解を数値計算で求めようとしても，非常に時間がかかる.
NP-hard問題と呼ばれるものになる.

L_p-正則化の解：コスト関数の比較

コスト関数の一般形

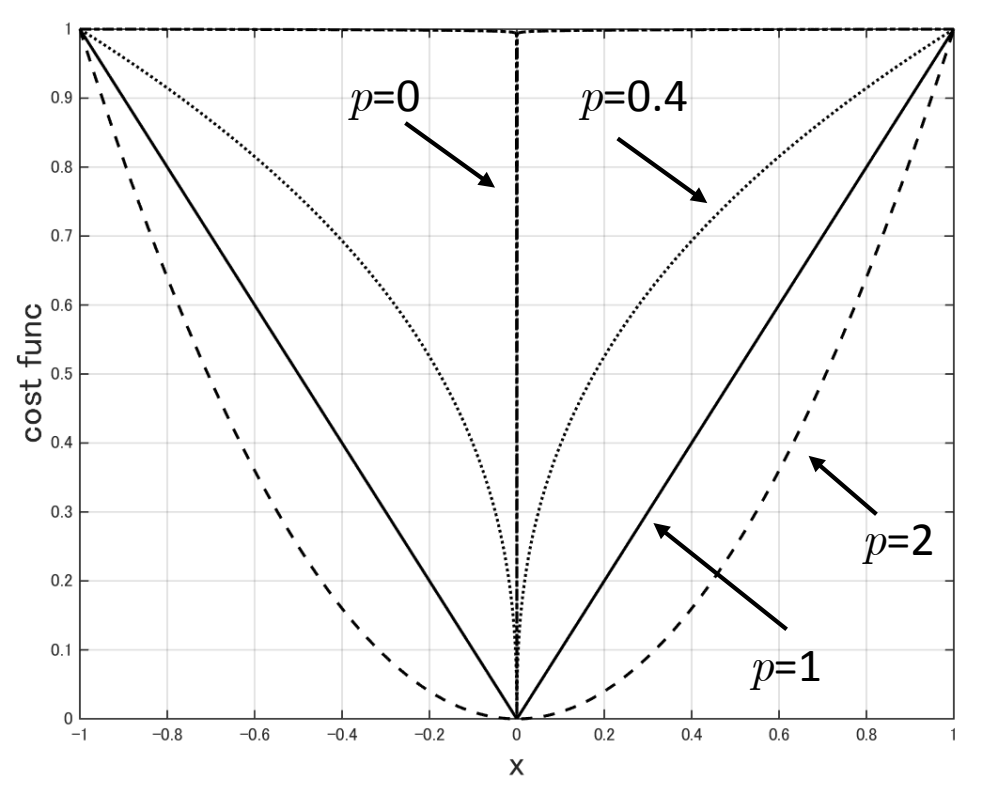
$$\hat{\boldsymbol{x}} = \arg\min F : \quad F = \left\| \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \right\| + \xi\phi(\boldsymbol{x})$$

L_0-正則化: $\phi(\boldsymbol{x}) = \sum_{j=1}^{N} I(x_j)$

L_1-正則化: $\phi(\boldsymbol{x}) = \sum_{j=1}^{N} \left| x_j \right|$

L_2-正則化: $\phi(\boldsymbol{x}) = \sum_{j=1}^{N} x_j^2$

L_p-正則化: $\phi(\boldsymbol{x}) = \sqrt[p]{\sum_{j=1}^{N} \left| x_j \right|^p}$



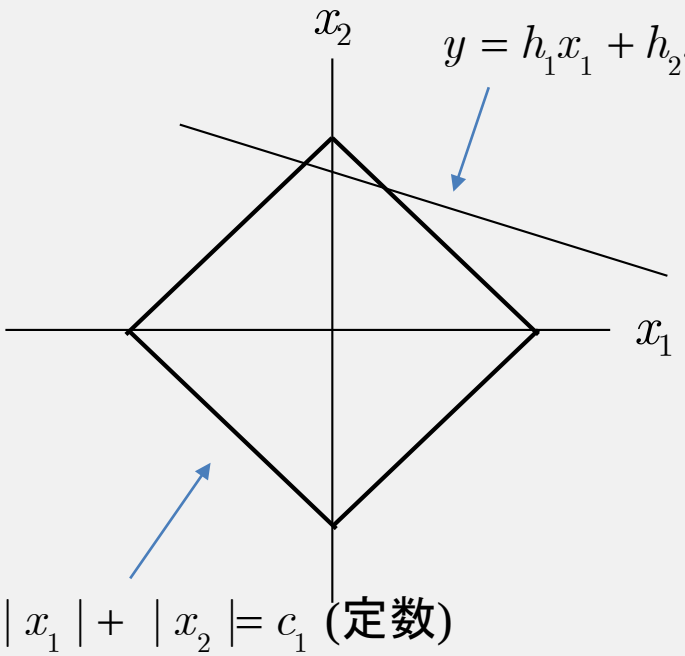
$p > 1$ ではスパースな解は生じないが，
$p \leq 1$ ではスパースな解となる．

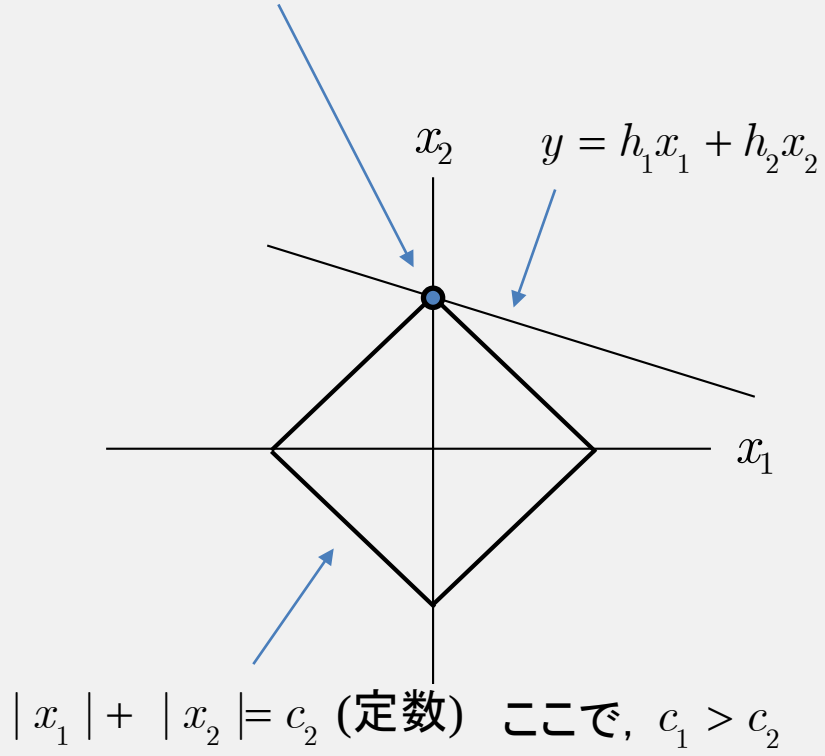
L_1ノルム制約はL_0ノルム制約の近似になっていて，実現可能な計算時間でスパースな解を与えることができる

2次元問題で考えてみる：L_1-正則化の解

L_1-正則化：　$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}(|x_1| + |x_2|)$ subject to $y = h_1 x_1 + h_2 x_2$

$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}(|x_1| + |x_2|)$ subject to $y = h_1 x_1 + h_2 x_2$ の解



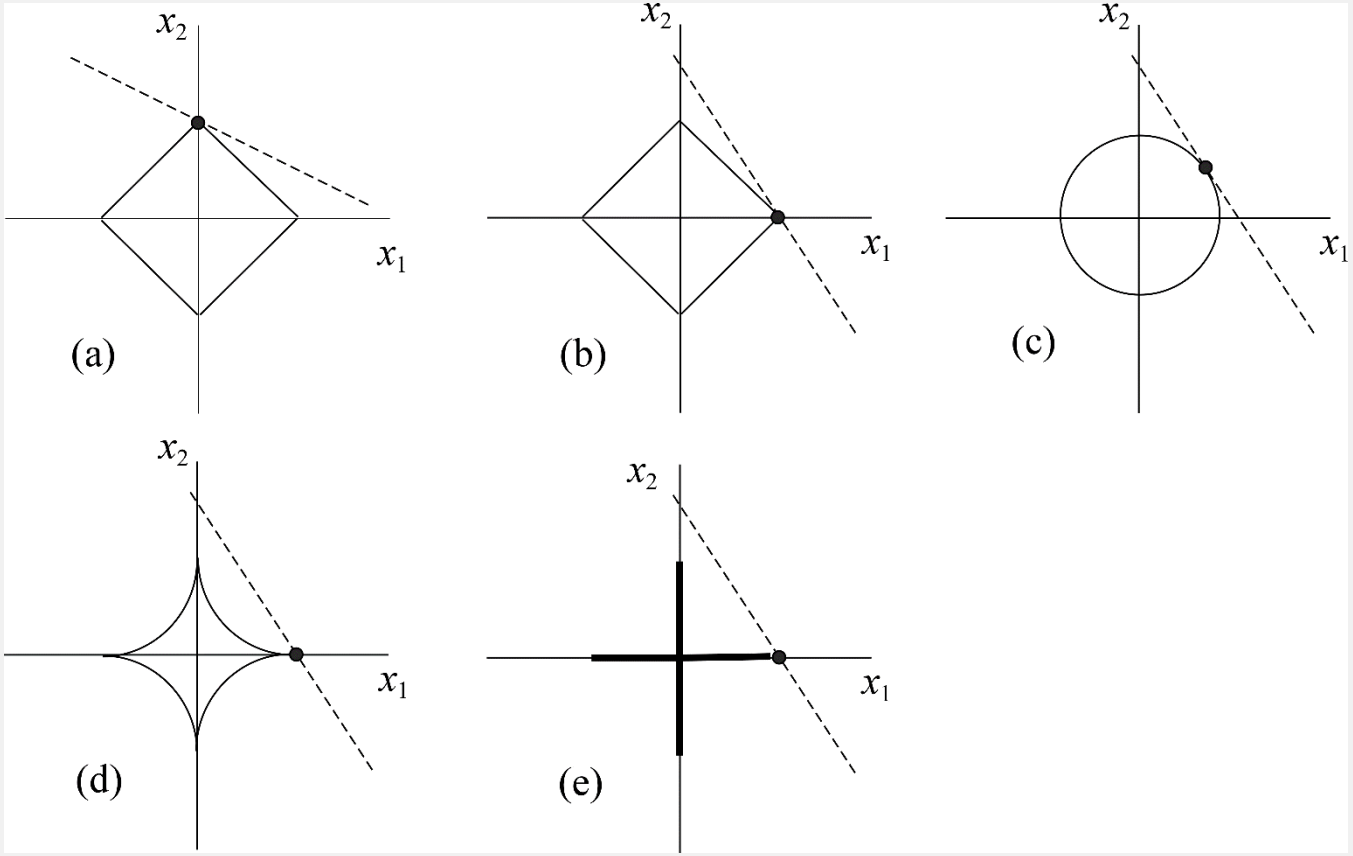


この例では $x_1 = 0,\ x_2 \neq 0$ である解が得られる.

2次元問題で考えてみる：他の制約条件の場合

L_1-正則化: $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}(|x_1| + |x_2|)$ subject to $y = h_1 x_1 + h_2 x_2$

L_2-正則化: $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}(x_1^2 + x_2^2)$ subject to $y = h_1 x_1 + h_2 x_2$



L_2-正則化の場合のみx_1, x_2ともにノンゼロの解（ノンースパースな解）が得られる.

L_1 -ノルム正則化ミニマムノルム解のMEGへの応用

Robert Tibshirani, “Regression shrinkage and selection via the lasso”, Journal of the Royal Statistical Society. Series B (Methodological), p.267--288, 1996.

MEGへの応用

1. Brian D. Jeffs, “Maximally sparse constrained optimization for signal processing applications”, Ph.D. thesis, University of Southern California,1989.

2. K. Uutela, M. Hamalainen, E. Somersalo, “Visualization of Magnetoencephalographic data using minimum current estimate, NeuroImage, Vol.10, P. 173-180, 1999.

3. K. Matsuura, and Y. Okabe, “Multiple current-dipole distribution reconstructed by modified selective minimum-norm method”, Proceedings of Biomag 96, P.290—293, 2000.

ソースの向き，タイムコースなどの推定が不正確になる等の問題点があり，現在ではあまり話題にならない.