(ハ) Generalized Cross Validation (GCV 法)

観測方程式 $\mathbf{d} = G\mathbf{s} + \mathbf{n}$

に対し、第k成分を取り除いた方程式

$$\underset{\mathbb{R}^{N-1}}{\mathbf{d}^k} = \underset{\mathbb{R}^{(N-1)\times M}}{G^k}\,\mathbf{s} + \mathbf{n}^k$$

を考える.

Tikhonov 正則化解は

$$\hat{s}^k_\alpha = G^{+k}_\alpha\,\mathbf{d}^k$$

$$G^{+k}_\alpha = \left((G^k)^T G^k + \alpha^2 I\right)^{-1}(G^k)^T$$

となる. これに G を作用させた第k成分

$$\underset{\mathbb{R}^N}{(G\,\mathbf{s}^k_\alpha)_k}$$

は. 第k成分を除いて得られた推定解 $\mathbf{s}^k_\alpha$ から予測されるデータの第k成分といえる.

そこで、

$$d = \begin{pmatrix} d_1 \\ \vdots \\ d_k \\ \vdots \\ d_N \end{pmatrix}$$

$$J(\alpha) = \frac{1}{N} \sum_{k=1}^{N} \left(d_k - \left(G \hat{s}_\alpha^k \right)_k \right)^2$$

を最小化するαを正則化パラメータとする。

ただし、これだと G^{+k} を $k=1 \cdots N$ において計算する必要がある。

そこで、

$$J(\alpha) \simeq \frac{\frac{1}{N} \| d - G \hat{s}_\alpha \|^2}{\left(\frac{1}{N} \mathrm{Tr}(I - G G_\alpha^+) \right)^2}$$

と近似し、G_α^+ の計算を（あるαに対して）1回だけで済ます方法をGCV法とよぶ。

⑪ One-leave-out Lemma

$$\tilde{d} = \begin{pmatrix} d_1 \\ \vdots \\ d_{k-1} \\ (G \hat{s}^k \ldots \end{pmatrix}$$

第k成分だけ予測値に変える

$$\tilde{d} = \begin{pmatrix} \vdots \\ (G\hat{s}_\alpha^k)_k \\ d_{k+1} \\ \vdots \\ \vdots \\ d_N \end{pmatrix}$$

とするとき $\hat{s}_\alpha^k = G_\alpha^{+} \tilde{d}$

∵

$$s_\alpha^k = \underset{s}{\operatorname{argmin}} \left\{ \| d^k - G^k s \|^2 + \alpha^2 \| s \|^2 \right\} \quad (P)$$

$d^k \in \mathbb{R}^{N-1}$

第k成分を追加

$$= \underset{s}{\operatorname{argmin}} \left\{ \| \tilde{d} - G s \|^2 + \alpha^2 \| s \|^2 \right\}$$

第k成分は $(G\hat{s}_\alpha^k)_k$

$s = \hat{s}_\alpha^k$ に対しては 第k成分は $(G\hat{s}_\alpha^k)_k$

引くと0になるので、等式(P)は成り立つ。

したがって

$$\hat{s}_\alpha^k = G^{+} \tilde{d}$$

∎

Cross Validationの評価関数から $G_\alpha^{\dagger k}$, $\hat{s}_\alpha^k$ を除去

$$J(\alpha) = \frac{1}{N}\sum_{k=1}^{N} \| d_k - (G\,\hat{s}_\alpha^k)_k \|^2$$

ここで、

$$\frac{\left(\text{kを除かないときの予測誤差}\right)}{\left(\text{kを除いたときの予測誤差}\right)} = \frac{d_k - (G\,\hat{s}_\alpha)_k}{d_k - (G\,s_\alpha^k)_k}$$

$$= \frac{d_k - \tilde{d}_k + \tilde{d}_k - (G\,\hat{s}_\alpha)_k}{d_k - \tilde{d}_k}$$

$$= 1 + \frac{\tilde{d}_k - (G\,\hat{s}_\alpha)_k}{d_k - \tilde{d}_k}$$

$$= 1 + \frac{(G\,\hat{s}_\alpha^k)_k - (G\,\hat{s}_\alpha)_k}{d_k - \tilde{d}_k}$$

$$= 1 + \frac{(G\,G_\alpha^+\,\tilde{d}l)_k - (G\,G_\alpha^+\,dl)_k}{d_k - \tilde{d}_k}$$

$$= 1 + \frac{\left(GG_\alpha^\dagger\left(\begin{pmatrix}\vdots\\ d_{k-1}\\ \tilde{d}_k\\ d_{k+1}\\ \vdots\\ d_N\end{pmatrix} - \begin{pmatrix}\vdots\\ d_k\\ \vdots\\ d_N\end{pmatrix}\right)\right)_k}{d_k - \tilde{d}_k}$$

$$= 1 + \frac{\left(GG_\alpha^\dagger\begin{pmatrix}0\\ \vdots\\ 0\\ \tilde{d}_k - d_k\\ 0\\ \vdots\\ 0\end{pmatrix}\right)_k}{d_k - \tilde{d}_k}$$

$$= 1 + \frac{(GG_\alpha^\dagger)_{kk}(\tilde{d}_k - d_k)}{d_k - \tilde{d}_k}$$

$$= 1 - (GG_\alpha^\dagger)_{kk}$$

よって、

$$d_k - (G\hat{s}_\alpha^k)_k = \frac{d_k - (G\hat{s}_\alpha)_k}{1 - (GG_\alpha^\dagger)_{kk}}$$

となり、したがって、

$$J(\alpha) = \frac{1}{N}\sum_{k=1}^{N}\frac{(d_k - (G\hat{s}_\alpha)_k)^2}{(1-(GG_\alpha^\dagger)_{kk})^2}$$

K=1~Nの平均で近似

$$\simeq \frac{1}{N}\sum^{N}\frac{(d_k - (G\hat{s}_\alpha)_k)^2}{\quad}$$

$$= \frac{1}{N} \sum_{k=1} \quad \left(\frac{1}{N} \sum_{j=1}^{N} (1 - G G_\alpha^\dagger)_{jj} \right)^2$$

$$= \frac{\frac{1}{N} \sum_{k=1}^{N} \| d - G \hat{s}_\alpha \|^2}{\left(\frac{1}{N} \mathrm{Tr} (I - G G_\alpha^\dagger) \right)^2}$$

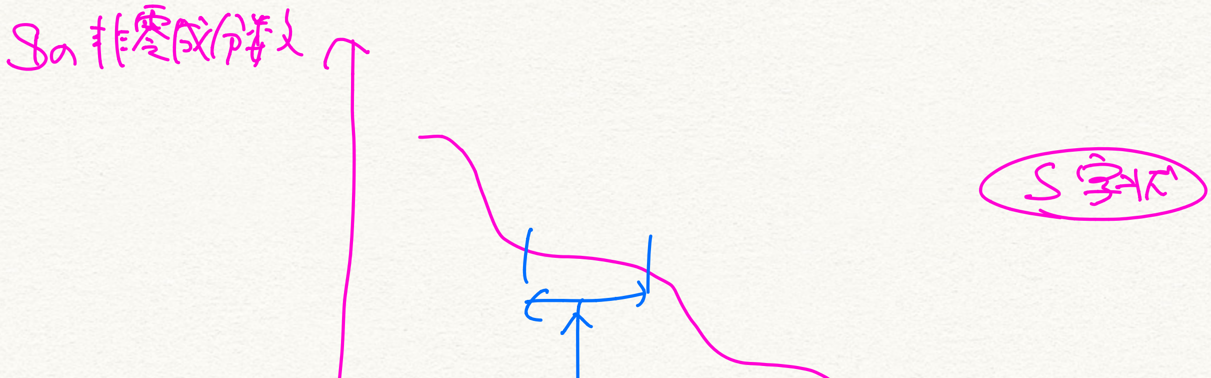
結局，Tikhonov 正則化解 $\hat{s}_\alpha$ の自乗誤差 $\| d - G \hat{s}_\alpha \|^2$ を $(\mathrm{Tr}(I - G G_\alpha^\dagger))^2$ で補正したものを最小化する α が，(近似的に) Cross Validation 誤差を最小化する正則化パラメータである．

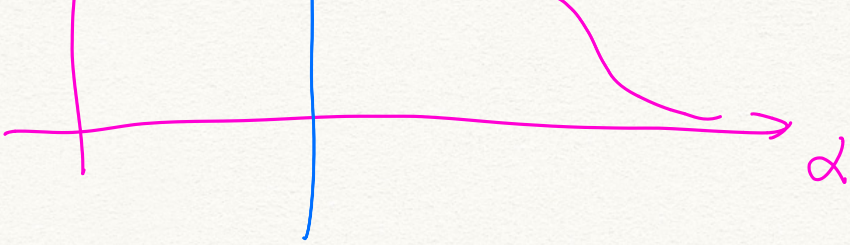
S - curve 法

L1ノルム最小化

$$\boxed{\min. \ \| d - G s \|^2 + \alpha^2 \| s \|_1}$$

の α を決定する．





幅広い区間のαに対して
非零成分数があまり変化しない領域
この区間からαを選ぶ.

Ⅲ 非線形解法の基礎

$$d(x) = \int s(x')\, g(x-x')\, dx' + n(x)$$

ソース $s(x)$ をパラメトリックにする

$$\Rightarrow \quad \mathbf{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_M \end{pmatrix}$$ を未知数として.

評価関数

$$J(\mathbf{s}) = \| \mathbf{d} - \mathbf{g}(\mathbf{s}) \|^2$$

を非線形最適化により, 最小化.

例

$$s(x) = \sum_{k=1}^{M} \underbrace{q_k}_{\text{点ソース強度}} \delta(x - \underbrace{x_m}_{\text{位置}})$$

このときは,

$$d(x) = \sum_{k=1}^{M} g_k \, g(x - x_m)$$

$\uparrow$

gが非線形でもx_mが既知だと$g(x-x_m)$はただの値になるが，x_mが未知だと非線形になってしまう

最急降下法

$$s_{j+1} = s_j - \alpha_j \nabla J(s_j)$$

更新幅αは，

$$\boxed{\min.\quad J(\alpha_j) = J(s_j - \alpha \nabla J(s_j))}$$

で決める.(Curryの規則)

$\downarrow$

$$\frac{\partial}{\partial \alpha} J(s_j - \alpha \nabla J(s_j)) = \nabla J(s_j - \alpha \nabla J(s_j)) \cdot (-\nabla J(s_j)) = 0$$

$$\Longleftrightarrow \quad \nabla J(s_{j+1}) \cdot \nabla J(s_j) = 0$$

すなわち，探索方向は前回と直交する方向となる!

(ジグザグ)

$g(s) = Gs$ では、

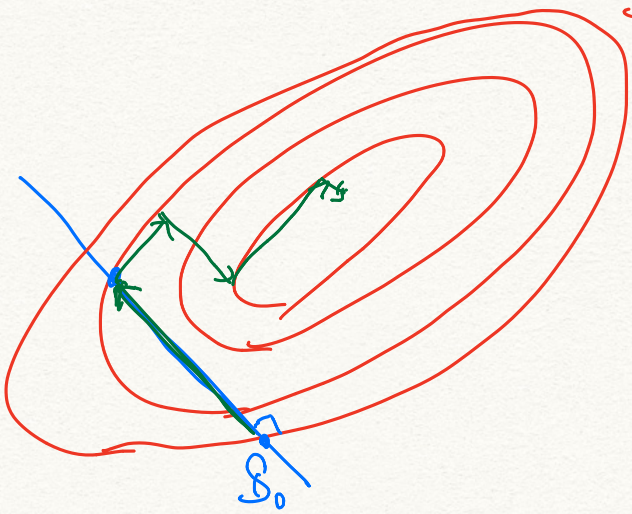
$J(s) = \| d - Gs \|^2$

$= (0次) + (1次) + s^T \boxed{G^T G} s$

特異値の値の差が大きい

Gの条件数が大きいと、J(s)の等高線は細長い楕円

J=一定



(iii) Gauss - Newton法

$g(s)$ を s_j まわりで Taylor展開.

$$g(s) \simeq g(s_j) + \underbrace{\left(\frac{\partial g}{\partial s}\right)^T}_{=ヤコビアンG} (s - s_j) + \underbrace{\frac{\partial^2 g}{\partial s \partial s^T}}_{ヘッシアン(3階テンソル)H} \times_1 (s - s_j) \times_2 (s - s_j)$$

モード積

$$G = \begin{pmatrix} \frac{\partial g_1}{\partial s_1} & \cdots & \frac{\partial g_N}{\partial s_1} \\ \vdots & & \vdots \\ \frac{\partial g_1}{\partial s_M} & \cdots & \frac{\partial g_N}{\partial s_M} \end{pmatrix}^T = \begin{pmatrix} \frac{\partial g_1}{\partial s_1} & \cdots & \frac{\partial g_1}{\partial s_M} \\ \vdots & & \vdots \\ \frac{\partial g_N}{\partial s_1} & \cdots & \frac{\partial g_N}{\partial s_M} \end{pmatrix}$$

して、評価関数に代入

$$J(s) = \| d - g(s_j) - G(s-s_j) - H x_1(s-s_j) x_2(s-s_j) \|^2$$

$$= (0次項) - (s-s_j)^T G^T (d - g(s_j)) + \underbrace{(s-s_j)^T G^T G (s-s_j)}_{②}$$

$$\underbrace{- 2(d - g(s_j)) \cdot (H x_1(s-s_j) x_2(s-s_j))}_{①} + (高次項)$$

$\downarrow$ 予測誤差 $\| d - g(s_j) \|$ は十分小さい として、② に対して ① はムシ

$$\simeq (0次項) - (s-s_j)^T G^T (d - g(s_j)) + (s-s_j)^T G^T G (s-s_j)$$

$\downarrow$ min.

$$s - s_j = (G^T G)^{-1} G^T (d - g(s_j))$$

よって

$$s_{j+1} = s_j + (G^T G)^{-1} G^T (d - g(s_j))$$

- ヤコビアン $G = \left(\frac{\partial g}{\partial s}\right)^T$ の条件が悪いとき → 正則化

$$s_{j+1} = s_j + (G^T G + \alpha^2 I)^{-1} G^T (d - g(s_j))$$

Levenberg-Marquardt 法

α^2 小さい $\longrightarrow$ Newton 法　収束速い

α^2 大きい $\longrightarrow$ 最急降下法　収束遅い

例

収束解を s^* とし，s のまわりで，

$$dl = G(s - s^*) + n \qquad \left(G = \left(\frac{\partial g}{\partial s}\right)^T (s^*) \right)$$

と展開できるものとする

$$n \sim N(0,\ \sigma^2 I)$$

と仮定する．このとき，推定誤差分散は，

$$E\left[(s - s^*)(s - s^*)^T\right] = \sigma^2 (G^T G)^{-1}$$

となる．

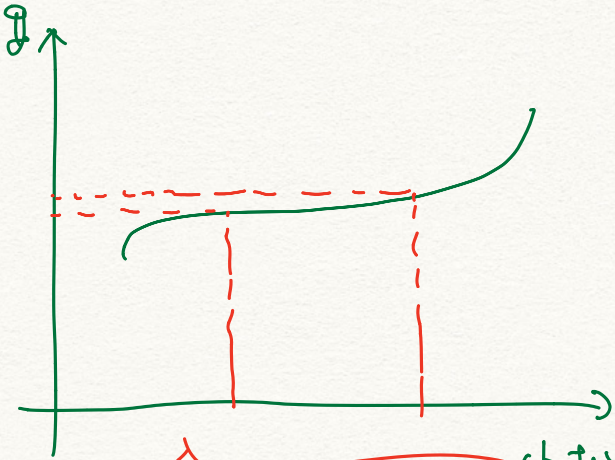
σ^2 が未知の場合は

$$\hat{\sigma}^2 = \frac{1}{N-M} \| dl - g(s^*) \|^2$$

により得られる推定値を用いる．

- ヤコビアンの条件が悪い

$\longrightarrow$ $\left(\frac{\partial g}{\partial s}\right)^T$ が小さな特異値をもつ



得られるデータの差は小さくても
実際の値は大きく異なる