⑩ 正則化

… グリーン関数がローパス（離散では条件数が大きい）の場合、正則化する

↓

ノイズの高周波部分をおさえる

正則化1

⑪ 打切特異値分解

Truncated SVD

推定解：$\hat{S} = G^{+}d = \sum_{k=1}^{R} \frac{v_k^T d}{\lambda_k} u_k$

における小さい特異値による除算をやめ、$T<R$ として、打ち切る

$$\hat{S}_T = G_T^{+} d = \sum_{k=1}^{T} \frac{v_k^T d}{\lambda_k} u_k$$

- 打切次数 T の決定

解のなめらかさ ― 正確さのトレードオフ

推定誤差

$$\frac{|\hat{S}|}{|S|} = \frac{|G^{+}n|}{|G^{+}d|} \leq \frac{\lambda_1}{\lambda_T} \boxed{\frac{|n|}{|d|}} = C$$

SN比

誤差の上限

SN比と誤差の上限がわかれば λ_T を

決定できる

$$\underbrace{\sum_{k=1}^{T} \frac{\sigma^2}{\lambda_k^2}}_{\text{Perturbation error}} + \underbrace{\sum_{k=T+1}^{R} \frac{d_k^2}{\lambda_k^2}}_{\text{Regularization error}}$$

(ノイズが逆問題をとくことによって増幅された値)

正則化2

Tikhonov Regularization

$$\min. \left(|d - Gs|^2, \ |s|^2 \right)$$

$$\downarrow$$

$$\min. \ J(s) = |d - Gs|^2 + \alpha |s|^2$$

$$\downarrow$$

$$(G^T G + \alpha^2 I) s = d^T s$$

ここで、$G = V \Lambda U^T$ (SVD) とすると.

$$G^T G + \alpha^2 I = (V \Lambda U^T)^T (V \Lambda U^T) + \alpha^2 U U^T$$

$$= U \Lambda^T \Lambda U + \alpha^2 U U^T$$

$$= U \begin{pmatrix} \lambda_1^2 + \alpha^2 & & \\ & \ddots & \\ & & \lambda_n^2 + \alpha^2 \end{pmatrix} U^T$$

$$= \begin{pmatrix} \lambda_k & & \\ & \alpha^2 & \\ & & \ddots \\ & & & \alpha^2 \end{pmatrix}$$

対角成分をα^2でもちあげ、

→ 特異値 λ_k を大きくみせる

(ii) Tikhonov の逆行列

$$G_\alpha^+ = (G^T G + \alpha I^2)^{-1} G^T$$

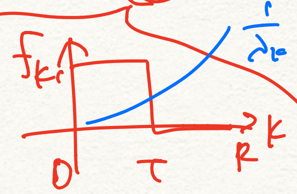
$$= \left(\sum_{k=1}^{M} (\lambda_k^2 + \alpha^2) u_k u_k^T \right)^{-1} \sum_{k=1}^{R} \lambda_k u_k v_k^T$$

$$= \sum_{k=1}^{R} \frac{\lambda_k^2}{\lambda_k^2 + \alpha^2} \frac{u_k}{\lambda_k} v_k^T$$

(iii) TSVD と Tikhonov の違い

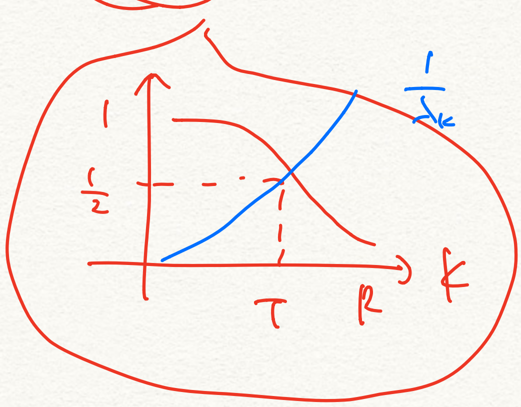
TSVD

$$G_T^+ = \sum_{k=1}^{T} \frac{1}{\lambda_k} u_k v_k^T = \sum_{k=1}^{R} f_k \frac{1}{\lambda_k} u_k v_k^T$$



Tikhonov

$$G_\alpha^+ = \sum_{k=1}^{R} \frac{\lambda_k^2}{\lambda_k^2 + \alpha^2} \frac{1}{\lambda_k} u_k v_k^T$$



正則化3

(iii) 拘束条件付き LS

$$\boxed{\begin{aligned} &\text{min.} \quad |s|^2 \\ &\text{s.t.} \quad |d - Gs|^2 = \sigma^2 \end{aligned}}$$

↑ ノイズレベルが分かっている場合

$$L(s, \alpha) = |s|^2 + \frac{1}{\alpha^2}\left(|d - Gs|^2 - \sigma^2\right)$$

$$\downarrow$$

$$\hat{s} = (G^T G + \alpha^2 I)^{-1} G^T d$$

Tikhonovと一緒だが，

$$|d - G(G^T G + \alpha^2 I)^{-1} G^T d|^2 = \sigma^2$$

から α を決められる.

正則化

(iii) L1ノルム最小化

$$\boxed{\min .\quad J(s) = |d - Gs|^2 + \alpha^2 |s|_1}$$

$$s = s^+ - s^- \qquad (s^+ \geq 0,\quad s^- \geq 0)$$

と分解し、

$$J(s^+, s^-) = |Gs|^2 - 2d^T Gs + |d|^2 + \lambda^2 \mathbb{1}_M^T s^+ + \lambda^2 \mathbb{1}_M^T s^-$$

$$(\mathbb{1}_M = (1, \cdots 1)^T \in \mathbb{R}^M)$$

ここで、

$$y = \begin{pmatrix} s \\ s^+ \\ s^- \end{pmatrix} \in \mathbb{R}^{3M}$$

とすると、

$$J(y) = y^T H y + h^T y$$

$$H = \begin{pmatrix} G^T G & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad h = \begin{pmatrix} -2G^T d \\ \lambda^2 \mathbb{1}_M \\ \lambda^2 \mathbb{1}_M \end{pmatrix} \in \mathbb{R}^{3M}$$

これを、拘束条件 $s = s^+ - s^-$

$$\Leftrightarrow \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} y_{M+1} \\ \vdots \\ y_{2M} \end{pmatrix} - \begin{pmatrix} y_{2M+1} \\ \vdots \\ y_{3M} \end{pmatrix}$$

$$y_j \geq 0 \quad (j = M+1 \cdots 3M)$$

のもとで解けばよい。（二次計画問題）

正則化5

Total Variation

値が変化することが稀 = 微分がスパース

$$\boxed{\min . \quad J(s) = |d - Gs|^2 + \alpha^2 |Ds|_1}$$

$$D = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & \cdots & \cdots & 0 & -1 \quad 1 \end{pmatrix}$$

最尤推定法

(11) 正則化を統計的視点を入れると ①

$$\mathbf{d} = G\mathbf{s} + \mathbf{n}$$

に対し、$\mathbf{s}, \mathbf{n}$ は独立な確率変数。

$$p(\mathbf{s}, \mathbf{n}) = p_1(\mathbf{s})\, p_2(\mathbf{n})$$ と仮定すると

$\mathbf{d}$ の pdf は、独立な確率変数の和の分布として

convolution

sがs'だったとして、データがdになるにはノイズが $\mathbf{d} - G\mathbf{s}'$ になっている必要がある

$$p(\mathbf{d}) = \int p_1(\mathbf{s}')\, p_2(\mathbf{d} - G\mathbf{s}')\, d\mathbf{s}'$$

で与えられる。sを固定して条件として与えると

$$p(\mathbf{d}|\mathbf{s}) = \int \delta(\mathbf{s}' - \mathbf{s})\, p_2(\mathbf{d} - G\mathbf{s}')\, d\mathbf{s}'$$

$$= p_2(\mathbf{d} - G\mathbf{s})$$

特に、$\mathbf{n}$ が多次元Gaussian に従う $\mathbf{n} \sim N(0, \Sigma)$ とすれば

$$p(\mathbf{d}|\mathbf{s}) = \frac{1}{(2\pi \det \Sigma)^{\frac{N}{2}}} \exp\left(-(\mathbf{d} - G\mathbf{s})^T \Sigma^{-1} (\mathbf{d} - G\mathbf{s})\right)$$

← likelihood function (sの関数)

データdが得られたらML.

$$\boxed{\max.\ p(d|s)}$$

$$\boxed{\min.\ (d-Gs)^T \Sigma^{-1} (d-Gs)}$$

特に、$\Sigma = \sigma^2 I$のときは

$$\boxed{\begin{gathered}\min.\ (d-Gs)^T(d-Gs)\\ (\text{Least-square})\end{gathered}}$$

MAP推定 = 正則化

統計の視点 ②

$s \to$ データがdになる確率 $p(d|s)$

(likelihood)

ではなく、

$d \to$ sが原因である確率 $p(s|d)$

(a posteriori)

(a posteriori)

を最大化

$$p(s \mid \boldsymbol{d}) = \frac{p(\boldsymbol{d} \mid s)\, p(s)}{p(\boldsymbol{d})}$$

$$\sim \underbrace{p(\boldsymbol{d} \mid s)\, p(s)} \rightarrow \max$$

観測ノイズも s も ガウス分布に従う場合

$$\begin{cases} p(s) \propto e^{-\frac{1}{2\sigma_s^2} s^T s} \\ p(\boldsymbol{d} \mid s) \propto e^{-\frac{1}{2\sigma^2}(\boldsymbol{d} - Gs)^T(\boldsymbol{d} - Gs)} \end{cases}$$

よって、事後確率は

$$p(\boldsymbol{d} \mid s)\, p(s) \propto \underbrace{e^{-\frac{1}{2\sigma^2}(\boldsymbol{d} - Gs)^T(\boldsymbol{d} - Gs)}\, e^{-\frac{1}{2\sigma_s^2} s^T s}}$$

$$\Downarrow$$

$$\boxed{\min_{s}\ \frac{1}{2\sigma^2}(\mathbb{d}-Gs)^T(\mathbb{d}-Gs)+\frac{1}{2\sigma_s^2}s^Ts}$$

$$\Downarrow$$

$$\hat{s}=\left(G^TG+\frac{\sigma^2}{\sigma_s^2}I\right)^{-1}G^T\mathbb{d}$$

(3) 正則化パラメータの選び方

- L-Curve 法

Tikhonov 正則化解

$$\hat{s}_\alpha=(G^TG+\alpha^2 I)^{-1}G^T\mathbb{d}$$

$$=\sum_{k=1}^{R}\frac{\lambda_k d_k}{\lambda_k^2+\alpha^2}u_k\quad(d_k=v_k^T\mathbb{d})$$

$$|\hat{s}_\alpha|^2=\sum_{k=1}^{R}\left(\frac{\lambda_k d_k}{\lambda_k^2+\alpha^2}\right)^2$$

$$|\mathbb{d}-G\hat{s}_\alpha|^2=\left|\sum_{k=1}^{N}d_k v_k-\sum_{k=1}^{R}\frac{\lambda_k^2 d_k}{\lambda_k^2+\alpha^2}v_k\right|^2$$

$$=\left|\sum_{k=1}^{N}\frac{\alpha^2 d_k}{\lambda_k^2+\alpha^2}v_k\right|^2$$

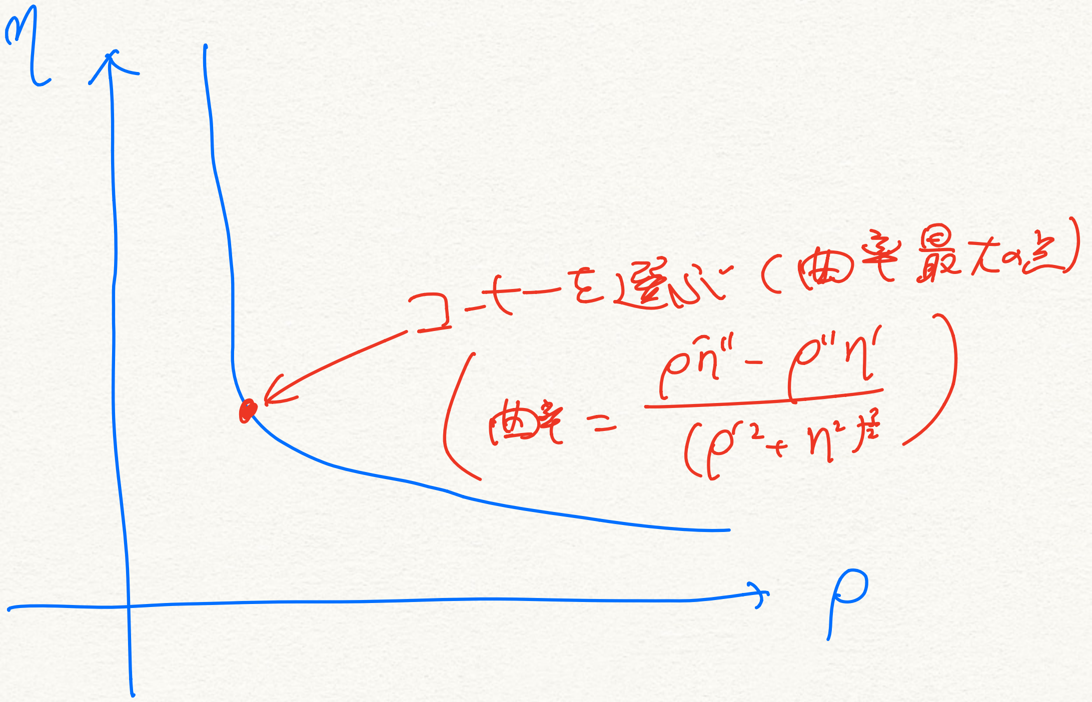
$(k=R+1$ 以降は $\lambda_k=0)$

$$= \sum_{k=1}^{N} \left(\frac{\alpha^2 d_k}{\lambda_k^2 + \alpha^2} \right)^2$$

ここで、次元を合わせるために

$$\begin{cases} \rho \equiv |d - G\hat{s}_\alpha|^2 \\ \eta \equiv \lambda_1^2 |\hat{s}_\alpha|^2 \end{cases}$$

とおくと、αを決めるごとに(ρ, η)の値が定まる。そこで、(ρ, η)平面上にαを変化させながら点をプロットしていく。



$$\frac{\partial \eta}{\partial \alpha^2} = -\lambda_1^2 \sum_{k=1}^{R} \frac{2\lambda_k^2 d_k^2}{(\lambda_k^2 + \alpha^2)^3}$$

$$\frac{\partial \rho}{\partial \alpha^2} = \cdots\cdots = -\frac{\partial \eta}{\partial \alpha^2}\frac{\alpha^2}{\lambda_1^2}$$

$$\frac{\partial \eta}{\partial \rho} = -\frac{\lambda_1^2}{\alpha^2}$$