# 悪条件問題における正則化パラメータの決定法

北川高嗣（愛媛大・理）

## 1. 悪条件問題の所在

近年，悪条件問題の重要性が広く認められる様になり，欧米を中心として，その研究もさかんになりつつある。悪条件問題に関する国際的シンポジウムも，Ake Björck（Linköping Univ. Sweden 1977），M.J. Nashed（Delaware 1979）等により開催されている。

その主要な問題の Source は以下の二つに大別されると思われる。

### 1.1). 不適切問題（Ill-posed Problems）

（又は Incorrectly-posed Problems, Improperly-posed Problems.）

この問題の class は，問題の解がデータの連続性を反映しない，即ち線型作用素方程式 $Kf=g$ （$K\in[X,Y]$ $X,Y$: Hilbert, $f\in X$ $g\in Y$）の operator $K$ の逆作用素 $K^{-1}$ が不連続である事によって特徴づけられる。数学的諸問題のうち，この class に入るものの代表例として以下のものがあげられる。

a）関数の微分

b）解析関数の解析接続（Analytic continuation of a function）

c）逆熱伝導問題（The backward heat equation）

d）ディリクレ問題（Dirichlet type boundary value problem）

e）ラプラス逆変換（Inversion of the Laplace Transform）

これらの問題は，全て第一種フレドホルム型積分方程式

$$K(f)(s)=\int_a^b \tilde{K}(s,t)\,f(s)\,ds = g(t) \qquad (1)$$

に帰着させる事ができる。たとえば a)関数 $g(x)$ の n 階微分を求める事は，第一種積分方程式

$$\int_0^t \frac{1}{(n-1)!}(t-s)^{n-1} f(s)\,ds = g(t) \qquad \left( (1)で\ \tilde{K}(s,t)=\frac{(t-s)^{n-1}}{(n-1)!}\ としたもの \right)$$

を解く事に等しい。

### 1.2) 統計的諸問題

(1.1) を数値的に解く場合の本質的な問題点は，作用素 K の inverse が非有界である為，離散化を精密に行なえば行なう程，解くべき線型方程式が数値的に不安定になってしまうところにある。いわば，主要な問題点は，作用素 K を離散化した

行列が悪条件になる点にある。これに対し、

2a) 多項式によるデータのあてはめ (Data fitting.)
2b) スプラインによるデータの平滑化 (Smoothing.)
2c) 自己回帰モデル (Regression problem)

等の統計的諸問題に於いては、問題自体の感度がそれ程高くなくとも、データに含まれる誤差が大きいと、単に正規方程式を解いただけでは満足な結果は得られない。この場合は、データの誤差が大きい事で、解くべき線型方程式を精密に解こうとすればする程（具体的には、最小二乗法に於いて残差を小さくしようとすればする程）本来求めようとするモデルには、程遠いものを解として選んでしまう事になる。(たとえば2a)に於いて多項式の次数を高くすれば、残差はいくらでも小さくできる。)

2. 悪条件問題の定義

上記の2つの悪条件問題のクラスを見る事により、悪条件問題の難しさは、
i) 線型系の感度、ii) 入力の摂動の大きさ、の2つに大きく依存することがわかる。つまり、悪条件問題は、線型系の感度が高ければ高い程、入力の摂動（データに含まれる誤差）が大きければ大きい程、解きにくくなる。

今、離散化された線型系を統一的に

$$X\beta = y + \varepsilon \quad (X \in R^{m\times n}, \varepsilon, y \in R^m, \beta \in R^n) \qquad (2)$$

と表す。$\varepsilon$は入力の摂動項とし、$E\varepsilon^T\varepsilon = \sigma^2$, $E\varepsilon = 0$（即ち平均0、標準偏差$\sigma$）とする。又$X$の摂動はないものと仮定する。問題のときにくさを定量化する。
まずi)の線型系$X$の感度は、$\kappa(X)$で表わされる。

$$\kappa(X) = \|X\|\,\|X^+\|.$$

（$\|\cdot\|$：ユークリッドノルム、$X^+$は$X$のMoore-Penrose一般化逆行列。）

又入力の摂動の大きさは$\sigma$であるから、出力の摂動は$\sigma\times\kappa(X)$で推定できる。
今、使用する計算機の浮動小数点系を$F(\beta, s)$とする（$\beta$進$s$桁）。(2)の最小二乗解を直接法（Q-R、修正グラム-シュミット法等）で解きうる為には、

$$\kappa(X)\times\max(\sigma, \beta^{-s}\|X\|) \leq \|X\|$$

が満たされなければならない。$\|X\|\cdot\beta^{-s}$は丸めの誤差レベル（計算機に依存）である。
本稿では上記の条件が満たされない様な系を悪条件問題として定義し議論の対象とする。即ち、以下の条件(3)が、系が悪条件である事の条件であるとする。

$$\|X^+\|\times\max(\sigma, \beta^{-s}\|X\|) \geq 1 \qquad (3)$$

「以上まとめて、
<定義>
線型系 $X\beta = y+\varepsilon$ ($X\in R^{m\times n}$, $\varepsilon, y\in R^m$, $\beta\in R^n$) を$F(\beta, s)$上で解く時、

$$\|X^+\|\times\max(\sigma, \beta^{-s}\|X\|) \geq 1$$

（$X^+$：一般化逆行列、$\sigma = E\varepsilon^T\varepsilon$（$\varepsilon$の標準偏差）、$\|\cdot\|$：ユークリッドノルム）

であれば、線型系は悪条件であるという。」

3) 数値解法と問題点，

2)で定義した悪条件問題，$X\beta = y + \varepsilon$ に対し一般の最小二乗法による解，即ち，$\min_{\beta} \| X\beta - (y+\varepsilon) \|^2$ を求める事は無意味である。何故なら，残差を0に近づけようとする試みが，入力データの摂動 $\varepsilon$ を，$X$ の不安定性 $\kappa(X)$ によって増幅させる事により 解 $\beta$ を大きくかき乱すからである。

この為に，最小二乗法は，何らかの意味で修正されなければならない。修正された最小二乗法として以下のものを挙げる事ができる。

1) 最小自乗最短解．(Minimal Norm Least Squares Solution.)

本来は，$X$ が Singular で解 $\beta$ に一意性がない時，$\beta$ のノルムを最小にするものを解として選ぶものである。即ち $\bar{y} = y + \varepsilon$ として，

$$\min_{\beta} \| X\beta - \bar{y} \|^2 \qquad (X \in R^{m\times n} \quad \text{rank}\, X < \min(m,n)\text{，}\beta\text{に一意性はない。})$$

$$\min_{\beta} \|\beta\| \qquad (\text{この条件により}\ \beta\ \text{は一意となる。})$$

一意解 $\beta$ は $\beta = X^{+}\bar{y}$ ($X^{+}$ は一般化逆行列) で与えられる。実際には，$X$ の0特異値に対応する特異ベクトル成分を0としたものを解としたものになる。

2) Truncated Singular value.

1)の M.N.L.S.S 解は，あくまでも $X$ の張る全ての特異ベクトル成分で解を構成する。それに対し，解をかき乱す可能性のある微小特異値に対応する特異ベクトル成分を強制的に0とおき，意味のある特異値に対応する特異ベクトル成分のみで解を構成しようとするのがこの手法である。この場合，0とおく特異ベクトル成分の選択が問題となる。$X$ の特異値を $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ とし対応する右特異ベクトルを $u_1, u_2, \cdots, u_n$ とすれば，この場合の解 $\beta$ は，($\bar{\beta} = V^{T}\beta$　$V$ は $X$ の右特異行列)

$$\min_{\beta} \| X\beta - \bar{y} \|^2 ,\quad \min_{\beta} \|\beta\| \qquad \bar{\beta} \in \text{span}\,[u_1, u_2, \cdots, u_r] \quad (r < n)\text{．となる。}$$

3) Damped least squares. (Modified least squares, Regularization)

この手法は，統計的問題に対しては，Modified least squares，Ill-posed Problems に対しては regularization (正則化法) と呼ばれているが，制約付き最小自乗問題 (Constrained least squares problems) として定式化できる。

$$\min_{\beta} \| X\beta - \bar{y} \|^2 \quad \text{subject to} \quad \|\beta\|^2 \leq \gamma^2 \ (\text{制約条件})$$

Lagrange の未定乗数法により以下の同値な式に変形される。($\lambda$ は未定乗数)

$$\min_{\beta} \{ \| X\beta - \bar{y} \|^2 + \lambda \|\beta\|^2 \} \iff (X^{T}X\beta + \lambda I)\beta = X^{T}\bar{y}$$

この場合問題になるのは，正則化パラメータ (Regularization parameter) $\lambda$ の選択である。

4) (Modified) Regularization.

3) の制約条件 $\|\beta\|^2 \leq \gamma^2$ をさらに一般化し $\|L\beta\|^2 \leq \gamma$ としたものである。

安定化因子 $\Gamma$ は、通常微分作用素の離散近似である。3) に対応し

$$\min \|X\beta - y\|^2 \quad \text{subject to} \quad \|\Gamma\beta\| \leq \delta^2. \quad \text{Lagrange の未定乗数法により、}$$

$$\min\{\|X\beta - y\|^2 + \lambda^2\|\Gamma\beta\|^2\} \iff (X^T X + \lambda\Gamma^T\Gamma)\beta = X^T y .$$

と定式化される。この場合問題となるのもやはり、正則化パラメータ $\lambda$ の選択である。以下正則化パラメータの決定法を述べる前に、正則化パラメータ決定の定性的な意味について多角的に概観してみる。表中(controller)とは各 case に対し正則化パラメータ(Regularization Parameter)に対応する量の呼び名である。

| Case | large ← | (Controller) | → small |
|---|---|---|---|
| 1) MNLSS (最小二乗最小ノルム解) | 残差が増大する | (数値的0とするしきい値) | 解のノルムが増大する。 |
| 2) Singular Value Truncation | 解の張る空間の次元の減少 | (強制的に0とする特異値及び特異ベクトル) | 条件数の増大。(小さい特異値を使う。) |
| 3) Data Fitting (多項式によるデータのあてはめ) | 数値的に安定(多項式の次数を下げる。) | (あてはめる多項式の次数) | データ点によりよくFitする。(次数を高くする。) |
| 4) 第1種フレドホルム型積分方程式 | もとの積分方程式がひずむ(安定化因子の導入による。) | (正則化法で解くと仮定する正則化パラメーター) | 数値的に不安定となる。(不適切問題であるから) |
| 5) Smoothing (データの平滑化) | Smoothness よりデータが平滑化される。 | (Smoothing parameter) | 残差が減少するが解が振動のはげしいものとなる。 |
| 6) 制約付き最小二乗問題 | 残差が増大し Constraint が減少 ($\|\Gamma\beta\|^2$) する。 | (Lagrange Multiplier) | 制約条件が増大し残差は減少する。 |
| 7) Ridge Regression (回帰モデル) | Bias の増大 | (Ridge parameter) | Variance の増大 |
| 8) 正則化法の誤差解析 | Truncated error (安定化因子導入による理論誤差)の増大 | (Regularization Parameter) | Round off error (データ誤差等による誤差)の拡大 |
| 9) 非線型モデル (この場合線型系はヤコビアン J) | ヤコビアンは安定、Step length は減少する。 | (Levenverg-Marquardt Parameter.) | 収束は速くなる。ヤコビアンは不安定。 |

以下では、Controller を Regularization parameter で代表し正則化パラメータと呼ぶ事にする。

## 4. 正則化パラメータの決定手法.

前節で述べた様に、"悪条件問題の解を求める"という事は、どの様な場合でも結局、"正則化パラメータを決定する"という問題に帰着される。正則化パラメータの決定に関し、2つの有用と思われる情報がある。それらは、

i) 解に関する前知識. 解を$\beta$とした時 $\|\beta\|^2$ $\|I_0\beta\|^2$ 等の量がわかっていれば、それを制約条件として用い、($\|\beta\|^2=\gamma^2$ $\|I_0\beta\|^2=\gamma^2$ とする。)

$$\min_{\beta} \|X\beta-(y+\varepsilon)\|^2 \quad \text{subject to} \quad \|\beta\|^2=\gamma^2 \ \text{or} \ \|I_0\beta\|^2=\gamma^2$$

とすればよい。しかしながら一般に$\|\beta\|, \|I_0\beta\|$の正確な値が入手可能である事は少ない。又正則化パラメータの決定は入力摂動$\varepsilon$に大きく依存する為、解のノルムに関する知識のみからは、うまく解が求まらない事が多い。

ii) 入力データの誤差$\varepsilon$の標準偏差$\sigma$. これは、解$\beta$を最良に選んだ時の残差がわかっている事に対応する。i)と同様に定式化すれば

$$\min_{\beta} \|I_0\beta\|^2 \quad \text{subject to} \quad \|X\beta-\tilde{y}\|^2=\sigma^2$$

となり 解を一意に定める事ができる。しかしながら 一般に与えられたデータから$\sigma$を推定する事(即ち、モデルの構造と誤差を分離する事)はできない。しかも、同じノイズレベルのデータであっても、異なったノイズセットに対する最適の正則化パラメータの値が、大きく変わってしまう事がある。

以上の理由から 上記の2つの量を必要としない事、即ち解に関する$\|\beta\|^2$ or $\|I_0\beta\|^2$の正確な知識、入力誤差の標準偏差に関する正確な情報は 得られない事を仮定する。この二つの条件を満たす正則化パラメータ決定手法には、以下がある。

a) *Akaike's Information Criterion.* (A.I.C) : *Maximum Likelihood.*
b). *Allen's Prediction Sum of Squares* (PRESS) : *Ordinary Cross Validation.*
c) *Generalized Cross Validation.* (G.C.V.)
d) *Our method.* 以下にその概要を示す。

### 4.a) Akaike's Information Criterion (A.I.C.)

Maximum Likelihood Estimator (最尤推定法)に基づく推定法であり、

$$AIC = (-2)\log \text{maximum likelihood} + 2n$$ により定義される。

ここに 線型システムは $y=X\beta+\varepsilon$ ($X\in R^{m\times n}$ $y,\varepsilon\in R^m$ $\beta\in R^n$) とし、正則化パラメータ$\lambda$は、次式に於ける$\lambda$である。(以下の手法についても同様)

$$\min_{\beta}\left\{\frac{1}{m}\|X\beta-y\| + \lambda\|\beta\|^2\right\} \iff (X^TX+n\lambda I)\beta = X^Ty$$

さらに $A(\lambda)=X(X^TX+m\lambda I)^{-1}X^T$, $\hat{\beta}(\lambda)=(X^TX+m\lambda I)^{-1}X^Ty$ (上式の解)

とおけば $A(\lambda)y=X\hat{\beta}(\lambda)$ であり maximum likelihood は、$\frac{1}{m}\|(I-A)y\|$ となる。

一般化された Maximum Likelihood Estimate $M(\lambda)$は、

$$M(\lambda)=\frac{1}{m}\frac{y^T(I-A(\lambda))y}{[\mathrm{Det}(I-A(\lambda))]^{1/m}}$$

で与えられ、正則化パラメータは、$M(\lambda)$を最小とするものを選ぶ。

4.b). Allen's Prediction Sum of Squares (PRESS)

この手法は，(Ordinary) Cross Validation とも呼ばれる。この手法の基本的な考え方は，「与えられたデータ点の一部をとり除き，(通常一点，) 残りのデータ点を使って，解$\beta$を求めそれがとり除かれた点を良く推定するように Bias (即ち $\lambda$) を決定する。」というものである。即ち$k$番目のデータを除いた系を $\tilde{X}^k$, $\tilde{y}^k$ 解を $\beta^k(\lambda)$ とすれば，

$$\beta^k(\lambda) = (\tilde{X}^{k^T}\tilde{X}^k + \lambda I)^{-1}\tilde{y}^k.$$

この $\beta^k(\lambda)$ を使って $y_k$ に対する残差，$([X\beta^k(\lambda)]_k - y_k)^2$ を求める。これを全てのデータ点 $y_i$ $i=1,\cdots,m$ について行ない，その総和が最小となる様な$\lambda$を，最適の$\lambda$として採用する。つまり PRESS の汎関数を $P(\lambda)$ とすれば，

$$P(\lambda) = \frac{1}{m}\sum_{k=1}^{m}([X\beta^k(\lambda)]_k - y_k)^2$$

で表わされる。$P(\lambda)$ は，$B(\lambda) = diag(1/(1-a_{jj}(\lambda)))$ とすれば，

$$P(\lambda) = \frac{1}{m}\|B(\lambda)(I-A(\lambda))y\|^2 \quad \text{where } A(\lambda) = X(X^TX + m\lambda I)^{-1}X^T$$

とも表わされる。

4.c) Generalized Cross Validation (G.C.V. by Grace Wahba.)

4.b) の Cross Validation は，良い推定値を与えられる事で知られているが，大きな問題点がある。それは，Rotation invariant でない (即ち直交変換によって不変でない。) 事である。というのは，$\lambda$推定の為の汎関数 $\Phi(\beta(\lambda))$ を導入した場合，

$$\min_{\beta(\lambda)}\{\|X\beta - y\|^2 + \lambda\|\beta\|^2\} \qquad \min_{\lambda}\Phi(\beta(\lambda)) \qquad (4)$$

と2つの汎関数を同時に最小化する解 $\beta(\lambda)$ を求めなければならない。その為 Regularization の汎関数を直交変換により，何らかの意味で標準化しておかないと，計算量は，許され難い程に大きなものとなってしまう。

この要請から考案された手法が Generalized Cross Validation である。即ち G.C.V. は，PRESS (Ordinary Cross Validation) の Rotation Invariant version である。この一般化は，X の特異値分解，とフーリエマトリックス W によって実現される。

$$X = UDV^T \qquad [W]_{jk} = \frac{1}{\sqrt{m}}e^{2\pi ijk/m} \text{ により } \quad y = X\beta + \varepsilon \text{ を}$$

$$\tilde{y} = WU^Ty = WDV^T\beta + WU^T\varepsilon \equiv \tilde{X}\beta + \tilde{\varepsilon} \quad (\text{i.e. } \tilde{X} = WDV^T,\ \tilde{\varepsilon} = WU^T\varepsilon)$$

これにより PRESS $P(\lambda)$ は G.C.V. $V(\lambda)$ に変換される。

$$V(\lambda) = \frac{1}{m}\|(I-A(\lambda))y\|^2 \Big/ \left[\tfrac{1}{m}\,\mathrm{Trace}\,(I-A(\lambda))\right]^2. \qquad (5)$$

4.d) Our Method

我々の提案した手法は正則化法の誤差解析により，丸め誤差（データの摂動による誤差）と打ち切り誤差（正則化パラメータ導入による系Xに入り込む摂動項）の2乗和が最小となるよう正則化パラメータを推定するものである。

正則化パラメータ推定の為の汎関数として $\Lambda_1(\lambda)$, $\Lambda_2(\lambda)$ の2つを導入する。実際のアルゴリズムに於いては，数値微分の安定性，及び最小値を見つける算法の効率のため，変数変換 $\nu = \log_{10}\lambda$ を採用する。$\Lambda_1$, $\Lambda_2$ は Rotation invariant

であり、一つの与えられたλに対し要する計算量は、S.V.D を利用する事により $O(n)$ である。

$$\Lambda_1(\lambda) = \left| \frac{\partial}{\partial\lambda} \|(X^TX+n\lambda I)^{-1}X^Ty\|^2 \right| \qquad (6)$$

$$\Lambda_2(\lambda) = \left\| \frac{\partial}{\partial\lambda}(X^TX+n\lambda I)^{-1}X^Ty \right\| \qquad (7)$$

$\Lambda_1, \Lambda_2$ に関する重要な関係式として以下が挙げられる。
$\bar\beta$を真の解、つまり入力誤差0の系を実数体の上で $\lambda=0$ として解いた解 とし $\|\bar\beta\|$ を有界であると仮定すれば、

$$\left|\frac{\partial}{\partial\lambda}\|\bar\beta-\beta(\lambda)\|^2 - \Lambda_1(\lambda)\right| \simeq O(\Lambda_2(\lambda)) \qquad (8)$$

$$\left|\frac{\partial}{\partial\lambda}\|\bar\beta-\beta(\lambda)\|\right| \le \Lambda_2(\lambda), \quad \Lambda_1(\lambda) \le (\Lambda_2(\lambda))^2。$$

（実際の算法では、(6),(7) に対し変数変換 $\nu = \log_{10}\lambda$ を施す。）

5. ラプラス逆変換（Inversion of Laplace Transform.）

i) ここでは、ラプラス逆変換を例にとり、G.C.V. と我々の手法を比較する。ここで、ある関数のラプラス変換 $\mathcal{L}(f(t))$ が与えられた時、元の関数 $f(t)$ を求める。

$$F(s) = \int_0^\infty e^{-st} f(t)\,dt = \mathcal{L}(f(t)) \quad (9) \quad \left(\Rightarrow \mathcal{L}^{-1}[F(s)] = f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{st}F(s)\,ds \right)$$

一般には、逆変換は、上に示したブロムウィッチ積分により与えられるが、ここでは、$f(t)$を未知関数とした第一種フレドホルム型積分方程式として逆変換を行う。
(9)は、Gauss Laguerre quadrature により離散化される。つまり

$$F(s_i) = \int_0^\infty e^{-s_i t} f(t)\,dt \simeq \sum_{k=1}^{n} w_k e^{t_k(1-s_i)} f(t_k)$$

つまり $X\beta = y+\varepsilon$ に対し

$$[X]_{ik} = w_k e^{t_k(1-s_i)} \qquad \beta_k = f(t_k) \qquad y_i = F(s_i) \qquad \begin{matrix} i=1,\cdots,m \\ k=1,\cdots,n \end{matrix}$$

とすればよい。$\varepsilon$は人為的に誤差を与える。

ii) 数値例（Grace Wahba et al [2] の数値実験例）及び討論。

具体的に ラプラス変換 $F(s)$, 解関数 $f(t)$ $s_i$, $t_k$をそれぞれ以下の様にとる。

$F(s) = \dfrac{1}{(s+1)^2} + \text{Noise}$ $s_i$ : $(0,2)$ の $m$ 等分点,

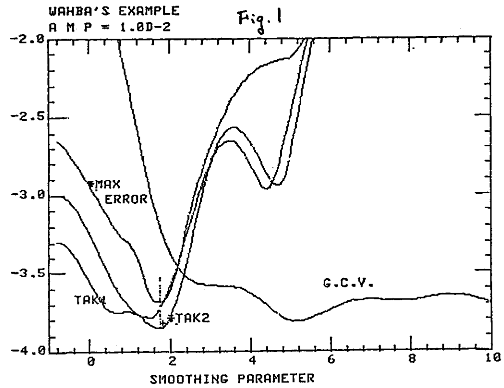
$f(t) = te^{-t}$ $t_k$ Laguerre Polynomial の零点,

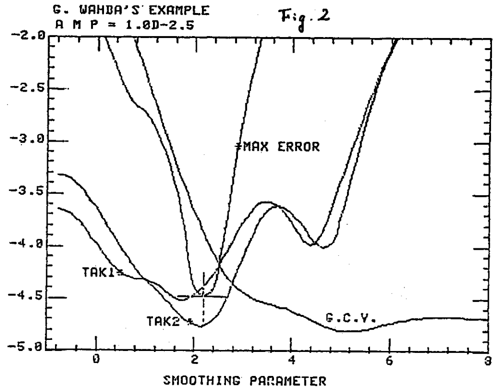
Noise $\varepsilon$ ; $E\varepsilon = 0$ $E\varepsilon^T\varepsilon = \sigma^2$（標準偏差）（$\sigma$=AMP と表示する。）

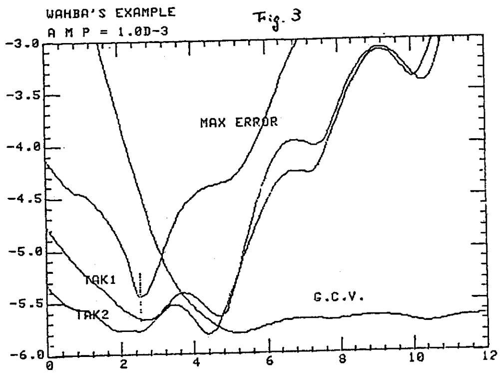
$n=10$ $m=10\sim16$ とする。 AMP $= 10^{-2}, \dots, 10^{-6}$

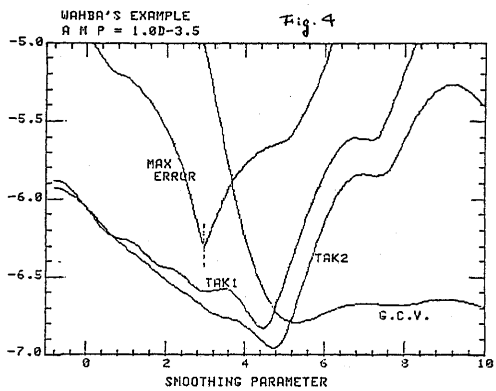
この系の条件数 $\kappa(X) = \|X^+\|\|X\|$ は約 $10^{10}$ であり、(3)式の意味に於いて十分悪条件となっている。図中のインデックスキイは以下の通り。

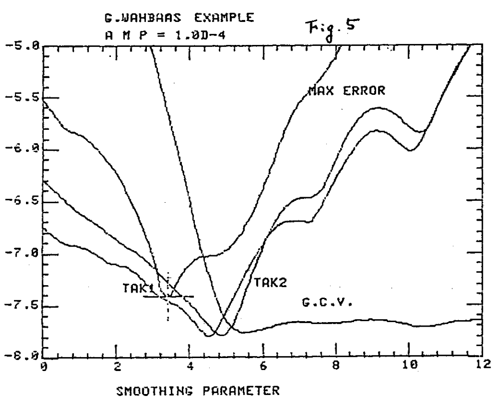
*Key to the figures;
'AMP'→ Standard deviation σ of the noise in the data y.
'x - axis'→ $-\log_{10}$(Smoothing parameter λ).
'y - axis'→ $\log_{10}$(G.C.V. V(λ)).
'G.C.V.'→ Generalized Cross Validation computed by (5).
'Max Error'→ $\|\bar\beta - \beta(\lambda)\|_\infty$ whose minimizer is to be estimated by a functional.
'Tak1,Tak2'→ Our functional $\Lambda_1(\lambda)$, $\Lambda_2(\lambda)$ defined by (6) and (7) respectively.



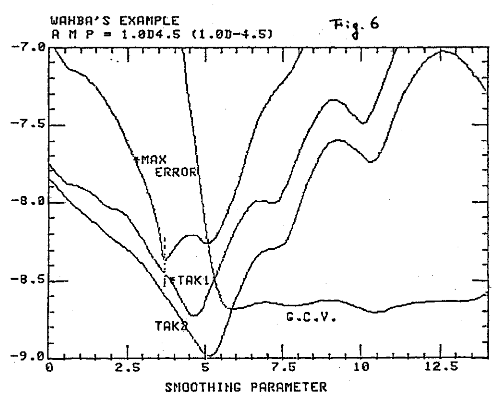
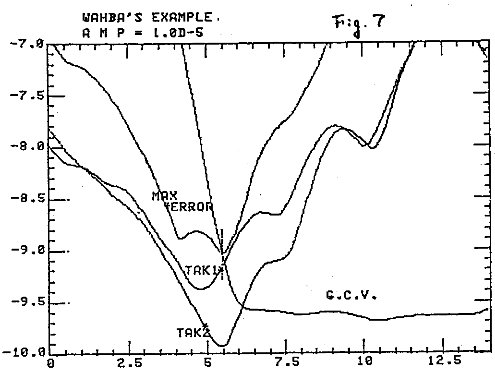
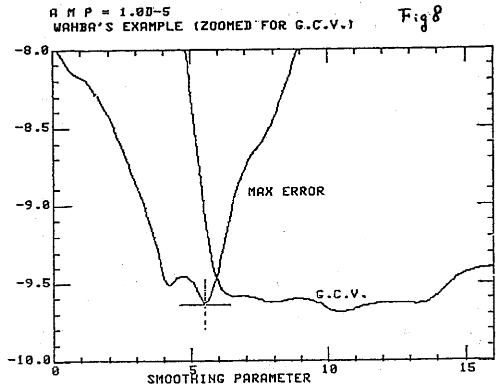
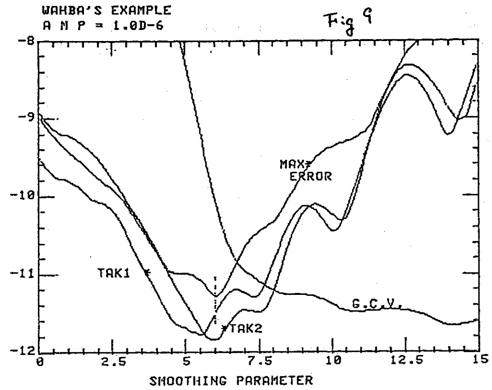




Fig.1〜2は、AMP = $10^{-2}$, $10^{-2.5}$と、データ誤差が大きい場合を示している。この様な場合、解の誤差（図中のmax error）は、データの振動が大きい為、一般に鋭い極値を持つ事になる。我々の手法では、解のλに対する微係数を推定する（式(8)）ので鋭い極値をよく推定している。我々の手法では、一般に、正則化パラメータの推定が高精度を要する場合（ピークが鋭く、少しの推定の差が誤差に大きく影響する場合）によりよい推定を与える。G.C.Vは、その定義から、残差に近いふるまいをする（PRESSの定義）為、それほどはっきりと極値は現われない。この傾向は、ノイズレベルが小さくなればなる程はっきりしてくる。Fig.3〜4では比較的G.C.Vによる推定もよいが、Fig.5〜6では、最小値の同定が難しくなっている。



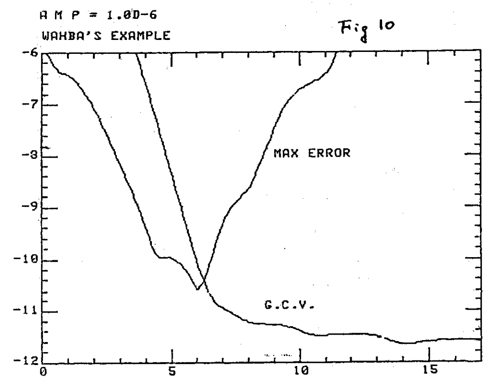
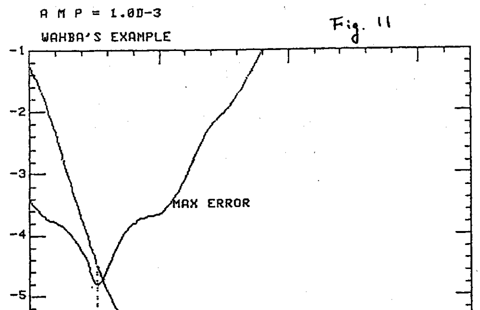
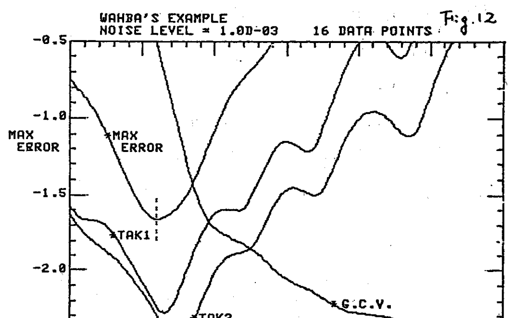
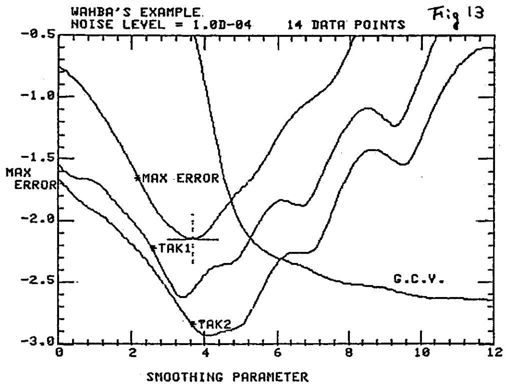



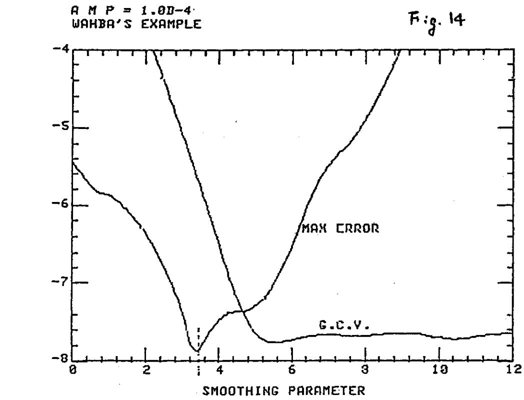

Fig 7〜10. は、比較的ノイズレベルが小さい($10^{-5}$〜$10^{-6}$)場合の数値結果である。ノイズレベルがこの程度になってくると線型系の ill-conditioning が問題となってくる。一般に、悪条件問題は、解の誤差が大きくとも、いくらでも残差は小さくできる。Fig 7.〜10. では、特にG.C.V.が残差に近いふる舞いをしている様に見うけられる。文献[1]によれば、G.C.V.ばかりでなく、AI.C., PRESS もノイズレベルの小さい場合は、正則化パラメータの推定は易しい(極値が鋭くない。)にも拘らず、推定の精度がよくない。我々の手法は、ノイズレベルが丸めの誤差レベルに達するまで安定して良い推定値を与える。(図中に示したのは、AMP=$10^{-6}$迄。)

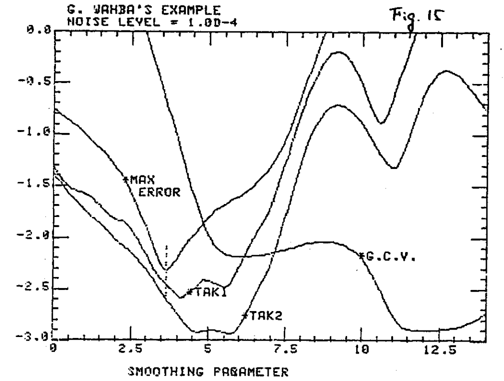
Fig.11 は、Fig.3 (AMP=$10^{-3}$) の G.C.V. をとり出したもので比較的良い推定値を与えている。これに対し Fig12. は AMP は等しいが異なったデータセット (16点) を与えたも

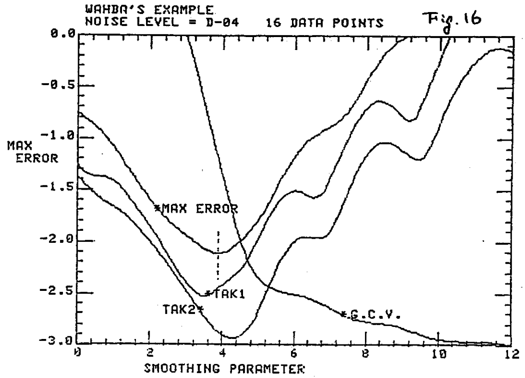





のである。Fig.13～16も同様にノイズレベルは全て$10^{-4}$である。G.C.V.はFig.14(Fig.5に対応)では良い推定値を与えているが、異なったデータセット(Fig.13,16)、σは同じであるが異ったノイズのセット(Fig.15)を与えると大きく変化してしまう。これは特にデータ点数が少ない場合顕著に見られる。これは、[2]の著者自ら認めるG.C.V.の欠点である。我々の手法では、汎関数の形は変化しているが(これは、異なるデータセット或いはノイズセットを取ったためXの特異値が変化している事に対応する。)どの場合も安定して良い推定値を与えている(Fig.13,15,16)。

又、G.C.V.、我々の手法とも定義された汎関数はconvexでない。この為汎関数の最小値を見つける算法が問題となるが、我々の手法では、変数変換 $\nu=\log_{10}\lambda$ の下で、定義された汎関数は、Xの特異値を$\mu_i$とした時 $\lambda \fallingdotseq \mu_i^2$ のところで局所的な極大値をとる事がわかっているので効率よく最小とする正則化パラメータを見つける事ができる。

以上まとめて、我々の手法は、G.C.V.と比較し、以下の点で優れていると考えられる。(Operator Equationsに対して。) 1) 与えられたデータ(又はノイズ)セットの変化に対する安定性、2) 悪条件に対する頑強さ、3) 定義された汎関数の最小化算法、4) ノイズレベルに応じたパラメータ推定の信頼性。

## 6) References

[1] T.Kitagawa,T.Torii,I.Ninomiya, 'Stabilization of linear system by SVD and Regularization.' Proceeding of IPSJ conference (1980), pp. 595-596. ( In Japanese.)

[2] G.H.Golub,M.Heath,G.Wahba, 'Generalized cross validation as a method for choosing a good ridge parameter.' Technometrics,21 (1979), pp. 215-223.