

楽しい逆問題 ver 2.0

河原 創

June 17, 2015

- 本文章は 2015/6 月 17 日の統計ゼミ用に作成した文書です。

Contents

1 離散線形逆問題	1
2 特異値分解	2
3 自然一般化逆行列	4
4 自然解の不安定性	8
5 Tikhonov Regularization	9
6 L-curve Criterion	12
7 最後に	14

1 離散線形逆問題

ある種の理論モデルは、モデルパラメタ m をあたえると

$$F(d, m) = 0 \quad (1)$$

のような形式で、観測データ列 d を予言するように問題を構成する事ができる。このような問題系で、観測データ列 d から、モデルパラメタ m を推定する問題を逆問題と呼ぶ。この例は有限個のパラメタを推定するので離散逆問題ともよぶ。

連続モデル関数を推定する逆問題も存在する。代表的な問題の一つに第一種フレドホルム積分型逆問題

$$d(x) = \int G(x, y)m(y)dy \quad (2)$$

がある。ここで $G(x, y)$ は積分核でモデルを記述する。

この第一種フレドホルム積分型逆問題を離散化し、式 (1) の形にすると

$$d - Gm = 0 \quad (3)$$

のように変形することができる。この形式で書かれる逆問題は離散線形逆問題と呼ばれ、逆問題としては最も、基本的な形式となる。ここに G は design matrix とよばれ、モデルをあらわす行列となっている。

実際には、用いる観測データは誤差を含んでいる。観測誤差を含んでいるデータ列を d^{obs} (observation) で表す事にする。逆問題の手法で何らかのモデル推定が得られたとしよう。この推定されたモデルを m^{est} (estimation) と表す。するとこの推定されたモデルと式 (1) から得られる d にたいする解を d^{pre} (prediction) としよう。すなわち d^{pre} は

$$d^{\text{pre}} = Gm^{\text{est}} \quad (4)$$

をみたす。

観測の予測誤差は

$$e \equiv d^{\text{obs}} - d^{\text{pre}} = d^{\text{obs}} - Gm^{\text{est}} \quad (5)$$

で定義される量である。予測誤差は m^{est} の観測データに対する説明力を表していると言えよう。 e はベクトルであるので、なんらかのノルムを定義して指標とする。通常 L2 ノルム、すなわち二乗和の平方根

$$|e| \equiv \sqrt{e^T e} \quad (6)$$

を指標とする。この予測誤差のノルムが小さいほど m^{est} は観測データ d^{obs} を説明していると言える。

2 特異値分解

さて、逆問題の推定値 m^{est} を構成する前に、離散線形逆問題で m と d がどのように関係しているかを調べよう。このためには、Design matrix G を特異値分解して考えると都合が良い。

特異値分解とは $N \times N$ 直交行列 U 、 $M \times M$ 直交行列 V 、 $N \times M$ の対角固有値行列

$$\Lambda = \begin{pmatrix} \Lambda_p & 0 \\ 0 & 0 \end{pmatrix} \quad (7)$$

$$\Lambda_p \equiv \text{diag}(\kappa_1, \dots, \kappa_p) \quad (8)$$

をもちいて、Design matrix を

$$G = U\Lambda V^T \quad (9)$$

と分解することである (特異値 κ_i は普通大きい順から並べる)。

さてここで直交行列の各列ベクトル

$$U = (\mathbf{u}_1, \dots, \mathbf{u}_N) \quad (10)$$

$$V = (\mathbf{v}_1, \dots, \mathbf{v}_M) \quad (11)$$

とかく。特異値分解は固有値の数 p を用いて

$$G = U\Lambda V^T = U_p \Lambda_p V_p^T \quad (12)$$

$$U_p = (\mathbf{u}_1, \dots, \mathbf{u}_p) \quad (13)$$

$$V_p = (\mathbf{v}_1, \dots, \mathbf{v}_p) \quad (14)$$

と変形できる。すなわち、 $d = Gm$ は

$$d = U_p \Lambda_p V_p^T m \quad (15)$$

とも書き直せる¹。

ここで推定したモデル m^{est} を v の線形結合に分解しよう。

$$m^{\text{est}} = \sum_{j=1} c_j v_j \quad (16)$$

で書いて、式 15 に代入してみると、

$$V_p^T m^{\text{est}} = \begin{pmatrix} v_0^T \\ v_1^T \\ \vdots \\ v_p^T \end{pmatrix} \sum_{j=1} c_j v_j = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{pmatrix} \quad (17)$$

となることから (ここで直交行列の性質から $v_i^T v_j = \delta_{ij}$; クロネッカーデルタ、となることを用いている)、

$$d^{\text{pre}} = Gm^{\text{est}} = \sum_{j=0}^p \kappa_j c_j u_j \quad (18)$$

となる。

つまり、 d^{pre} は $j = p + 1$ から $j = M$ の項には全く依存しないことが分かる。そこでこの空間を

$$V_o = (v_{p+1}, \dots, v_M) \quad (19)$$

と定義する。すなわち m^{est} は V_p 、 V_o によって張られる直交する二つの空間 (p 空間、ヌル空間) 内の (直交する) 二つのベクトル $m_p = \sum_{j=1}^p c_j v_j$ 、 $m_o = \sum_{j=p+1}^M c_j v_j$ の和

$$m^{\text{est}} = m_p + m_o \quad (20)$$

に分けることができ、 m_o は全く d^{pre} に影響を与えないことから、モデルからの予測 d^{pre} と観測データ d^{obs} の比較から m_o を推定する事はできない。 V_o が空集合、すなわち $p = M$ である時はデータからモデルを完全に決定できるので優問題 (well-posed problem) とよぶ。 $p \leq N$ でないとならないので、 $N \geq M$ 、すなわちデータ数がモデルパラメタ数より大きい事は優問題である事の必要条件である。しかし、たとえ $N \geq M$ であっても、 $p < M$ で有る場合、もしくは $N \leq M$ であり必然的に $p < M$ である場合は V_o が存在するので、推定できない m_o が存在する。この場合、を劣問題 (ill-posed problem) もしくは混合問題とよぶ。

問題が ill-posed でも絶望する必要は無い。この場合、逆問題の解法は陽に暗に m_o を仮定する事 (prior を与えること) によりモデルを推定する事ができる。ベイズ主義者の場合、prior を与えて問題を解く形式に違和感は無いだろう。ベイズ主義者でない場合も以下のように考えてもらいたい。例えば、 m_o の要素にあまり変動の無いベクトルだった場合、つまり対応する離散化前の関数 $m(x)$ がのっぺりした関数だった場合、

¹ V 、 U は直交行列のため $UU^T = U^T U = VV^T = V^T V = I$ であるが、 U_p ($N \times p$ 行列)、 V_p ($M \times p$ 行列) は、 $p \leq N$ 、 $p \leq M$ であるため、 $V_p^T V_p = U_p^T U_p = I$ ($p \times p$ 行列) であるが、 $V_p V_p^T$ 、 $U_p U_p^T$ は単位行列とは限らない ($p = N$ 、もしくは $p = M$ のときのみ成立) ことに注意。

そういう成分はどうでも必要でないという事もあるだろう。しかし、こういった成分が d^{obs} からは表現できないのかは知っておく必要がある。

次に、観測データのほうを u_i の線形結合

$$d^{\text{obs}} = \sum_{i=1}^N k_i u_i \quad (21)$$

で表してみよう。この場合、

$$e = d^{\text{obs}} - Gm^{\text{est}} \quad (22)$$

を m^{est} を調整する事でゼロベクトルにすることはできるのだろうか？この場合、 d^{obs} の u_{p+1} から u_N の成分は、どのように m^{est} を動かしても引き去る事ができない。なぜなら $q > p$ に対して、 u_q と Gm^{est} の内積をとると、

$$u_q^T Gm^{\text{est}} = u_q^T U_p \Lambda_p V_p^T m^{\text{est}} = 0 \quad (23)$$

となり、 Gm^{est} の u_q の要素は0であるからである。すなわちヌル空間を

$$U_o = (u_{p+1}, \dots, u_N) \quad (24)$$

として、 d^{obs} は U_p 、 U_o によって張られる直交する二つの空間 (p 空間、ヌル空間) 内の (直交する) 二つのベクトル $d_p = \sum_{i=1}^p k_i u_i$ 、 $d_o = \sum_{i=p+1}^N k_i u_i$ の和

$$d^{\text{obs}} = d_p + d_o \quad (25)$$

に分けることができ、どんなモデルパラメタ m^{est} を持ってきても Gm^{est} は d_o の成分をもつことはできない。

3 自然一般化逆行列

もし観測誤差が存在しない場合、明らかに d^{obs} の U_o は空集合となるはずである。しかし、実際には観測誤差のため d_o の成分が d^{obs} に存在する。ということは予測誤差のノルム、式 (6) を最小にするためには d^{obs} の d_p の成分を Gm^{est} で表現できれば良い。このような事を実現する一つの例が自然一般化逆行列である

いま特異値分解により $d = Gm$ は

$$d = U_p \Lambda_p V_p^T m \quad (26)$$

のように書けている訳だが、これを逆行列のようにして

$$\begin{aligned} m^{\text{est}} &= G^{-g} d^{\text{obs}} \\ G^{-g} &\equiv V_p \Lambda_p^{-1} U_p^T \end{aligned} \quad (27)$$

のようにした G^{-g} を自然一般化逆行列 (natural generalized inverse matrix; NGIM) という。NGIM により得られた解は次の性質を持つ

- A. 予測誤差最小である。

どんな m を持ってきても d を変えることはできないことから、予測誤差最小とは、予測誤差が d_p の成分を持たないことである。

$$U_p^T(d^{\text{obs}} - Gm^{\text{est}}) = U_p^T(d^{\text{obs}} - GG^{-g}d^{\text{obs}}) = U_p^T(d^{\text{obs}} - U_p U_p^T d^{\text{obs}}) = 0 \quad (28)$$

であるので、やはり自然一般化逆行列による解 m^{est} は予測誤差最小の解である。

- B. m^{est} は m_o の成分を持たない。

これは

$$V_o^T m^{\text{est}} = 0 \quad (29)$$

であることから、 $m_o = 0$ であることが分かる。このような解を自然解という。

ヌルベクトルは d^{obs} に影響を与えることがないから、NGIM を用いて一般解

$$m^{\text{est}} = G^{-g}d^{\text{obs}} + V_o\alpha \quad (30)$$

の形式が得られる。ここに α はヌルベクトルの各成分である。

先験情報として平均ベクトル \hat{m} を与えた場合

$$m^{\text{est}} = G^{-g}d^{\text{obs}} + (I - G^{-g}G)\hat{m} \quad (31)$$

を用いて、解の推定を行うと平均が反映される。ここに

$$V_p^T(I - G^{-g}G)\hat{m} = (V_p^T - V_p^T V_p V_p^T)\hat{m} = 0 \quad (32)$$

となっているので、 $(I - G^{-g}G)\hat{m}$ の V_p 成分は 0 であり、確かに一般解の一つになっている。また

$$V_o^T(I - G^{-g}G)\hat{m} = V_o^T \hat{m} = V_o^T \hat{m}_0 \quad (33)$$

なので、 $(I - G^{-g}G)\hat{m}$ の V_o 成分は \hat{m} の V_o 成分 \hat{m}_o になっているので、たしかに先験情報 \hat{m} にもっとも距離が近い一般解を選んでいることになる。実はこの式はモデルパラメタの不定性を Gaussian とおいた時の最尤値を与えるが、計算はここでは略する (Menke 5.10 参照)。

練習：Che mapping

さて、逆問題の一つの解法 NGIM まできたので、具体例を導入しよう。図 1 のように、山の向こうに M このタイルでできているタイル壁画がおいてある。我々の位置からではタイル壁画が遠すぎて点にしか見えな。斥候隊がタイル壁画まで到達したが、彼らはその壁画に描かれている人物を知らない。そして彼らは無線機とサーチライトしかもっていない。そこで夜までまって、斥候隊に壁画の様々な場所 N 点にライトを当ててもらおうよう頼んだ。ライトの大きさはタイルよりかなり大きく、山の向こう側から見ている我々は、ライトの当たっている部分の壁外による反射光の和 (観測データ) が我々の持つ測光機を通じてわかるのみである。また、壁画のどのタイルを中心にライトがどれだけ広がっているか (design matrix) は無線機を通じて

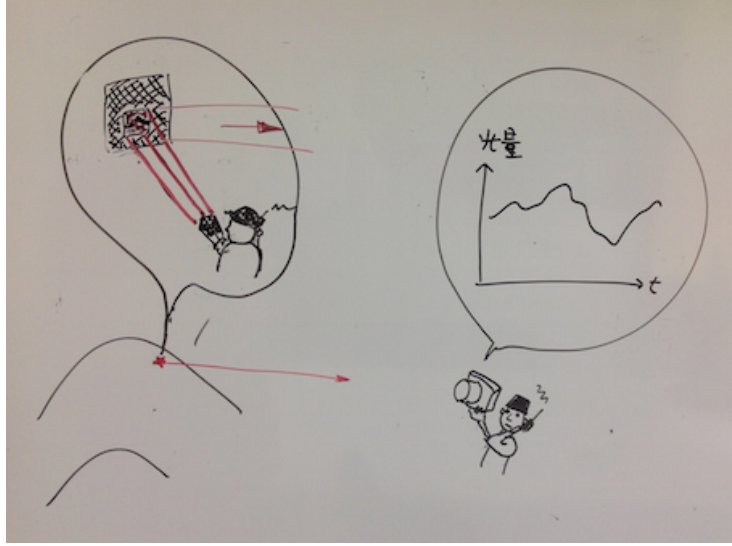


Figure 1: Che mapping.

知らされる。さて、このような観測データと design matrix から元の壁画に誰が描かれているか推定しよう。この問題を Che mapping と呼ぶ。

データ処理は数値計算をもちいるので、この Che mapping のコードを用意した。NGIM で Che mapping をするときは、こんな感じの option を設定する ($l=0.0$ が NGIM を指定する)。

```
./random_light.py -f che.png -n 1000 -l 0.0 -p 0.7 -w 20.0
```

いま壁画の画素数は $M = 2268$ ある。斥候隊にサーチライトを $N = 1000$ カ所、ランダムに当ててもらった ($-n 1000$)。またサーチライトの平均の幅は 20 タイル分ある ($-w 20.0$)。このプログラムでは幅もまちまちにしてあるが、各ショットごとのサーチライトの中心と幅は無線機を通じて正確に伝えられる。観測誤差はないとしている。この場合、NGIM による推定を図 3 に表示した。左が壁画、真ん中が壁画のどこにサーチライトを当てたか、右が NGIM による推定である。斥候隊が壁画の上から 7 割の部分にしかサーチライトの中心を当てなかったため ($-p 0.7$)、下部分では解像度が悪くなっているが、推定自体はできている。これはサーチライトの幅のおかげで、上 7 割より下も少しだけ情報が含まれているからである。

NGIM でも実際には、数値誤差のためすべての特異値は、なんらかの値をもっているなので、実用的には、特異値の列をみて (図 3)、ギャップのある値以下は 0 であると見なして p を決定する。

特異値分解と一般化逆行列の直交ベクトル表現

線形逆問題では u や v による展開が頻繁に出てくるため、通常の行列表記のみだと、今後の計算が表現しづらい。ここでは Hansen 2010 に従って特異値分解をベクトルで表現し直そう。まず、行列 A 、 B を

$$A = (a_1, \dots, a_n) \quad (34)$$

$$B = (b_1, \dots, b_n) \quad (35)$$

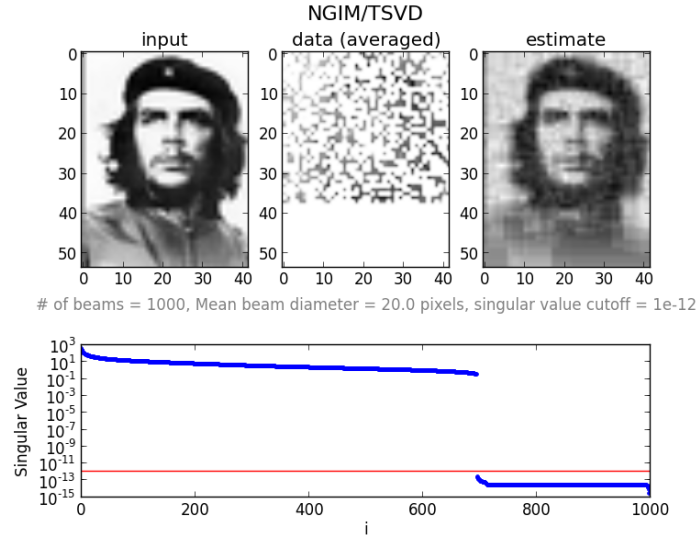


Figure 2: NGIM による Che mapping。観測データに（数値誤差を除く）、ノイズがない場合、予測誤差最小は良い推定値を与える。下パネルは特異値を大きい順に並べてある。

のように表現する時、

$$AB^T = \sum_{i=1}^n \mathbf{a}_i \mathbf{b}_i^T \quad (36)$$

となる。 $\mathbf{a} \mathbf{b}^T$ はダイアド ($\mathbf{a} \otimes \mathbf{b}$) である。また、

$$AB^T \mathbf{c} = \sum_{i=1}^n (\mathbf{a}_i \mathbf{b}_i^T) \mathbf{c} = \sum_{i=1}^n (\mathbf{b}_i^T \mathbf{c}) \mathbf{a}_i \quad (37)$$

である²。 $\mathbf{b}^T \mathbf{c} = \mathbf{b} \cdot \mathbf{c}$ は内積である。

上記のような表記を用いると特異値分解は

$$U \Lambda V^T = \sum_{i=1}^{\min(N,M)} \kappa_i (\mathbf{u}_i \mathbf{v}_i^T) \quad (38)$$

と表現できる。また一般化逆行列は

$$G^{-g} = \sum_{i=1}^p \frac{(\mathbf{v}_i \mathbf{u}_i^T)}{\kappa_i} \quad (39)$$

²($\mathbf{a} \otimes \mathbf{b}) \mathbf{c} = \mathbf{a}(\mathbf{b} \cdot \mathbf{c})$)

もしくは特異値に 0 が無いとすると

$$G^{-g} = \sum_{i=1}^{\min(N,M)} \frac{(v_i u_i^T)}{\kappa_i} \quad (40)$$

となる。自然解は

$$m^{\text{est}} = G^{-g} d = \sum_{i=1}^{\min(N,M)} \frac{(u_i^T d)}{\kappa_i} v_i \quad (41)$$

のように、 v_i の線形結合で表すことができ、要素は $(u_i^T d)/\kappa_i$ となる。

4 自然解の不安定性

2 章ではモデルとデータをそれぞれ

$$m^{\text{est}} = m_p + m_o \quad (42)$$

$$d^{\text{obs}} = d_p + d_o \quad (43)$$

と分けて、お互いに影響を及ぼせない成分を抽出した。しかし、 $i \leq p$ の成分についても、特異値が小さい場合はどうなるだろうか？

$$e = d^{\text{obs}} - d^{\text{pre}} = d^{\text{obs}} - G m^{\text{est}} = \sum_{j=1}^p (k_j - \kappa_j c_j) u_j, \quad (44)$$

の u_i 成分 ($i \leq p$) を m^{est} を調節してゼロにすることを考えよう。この場合、調節するのは c_i ということになる。ここで、もし κ_i が小さいと c_i を大きく変更して d^{obs} の u_i 成分である k_i に合わせないとならない。つまり、特異値が小さい成分に対して、予測誤差最小化をしてしまうと、 d^{obs} のちょっとした変化にたいして大幅にモデルが変化してしまう事になる。これを予測誤差最小である自然解の m^{est} の面からも見てみよう。いま自然解は式 (41) で表されるように

$$m^{\text{est}} = \sum_{i=1}^{\min(N,M)} \nu_i v_i \quad (45)$$

$$\nu_i \equiv \frac{u_i^T d^{\text{obs}}}{\kappa_i} \quad (46)$$

のように、 v_i の線形結合で表した時、要素は ν_i となる。ここで、 κ_i が小さいと $u_i d^{\text{obs}}$ が少し変化しただけで、 m^{est} の v_i 成分は大きく変更受けることとなる。これが逆問題の不安定性であり、overfit などと呼ばれる問題である。一般に小さい κ_i に対応する v_i は高周波成分である事が多く、不安定性が起きると、振幅の大きい高周波の不安定成分が m^{est} に現われる

5 Tikhonov Regularization

一般に不安定性を抑えるための処方を regularization とよび、様々な種類がある。最も簡単なやり方は p を決定する際に、特異値が厳密な 0 で切るのではなく、特異値をある値以上のものだけ採用して p を選ぶというやり方であり、Truncated SVD などと呼ばれている (そもそも数値的な誤差をかंगाえると、実用上は NGIM も Truncated SVD に含まれる)。他に有用な regularization として、Tikhonov regularization がある。これは、特異値を

$$1/\kappa_i \rightarrow \kappa_i/(\kappa_i^2 + \lambda^2) \quad (47)$$

ように dump し、小さい特異値のものを大きいものに入れ替えてから一般化逆行列を構成する。この操作により、推定モデルの不安定性を安定化させる事ができる。すなわち、推定値は

$$\mathbf{m}^{\text{est}} = V \Sigma_{\lambda} U^T \mathbf{d}^{\text{obs}} = \sum_{i=1}^{\min(N,M)} \frac{\kappa_i}{\kappa_i^2 + \lambda^2} (\mathbf{u}_i^T \mathbf{d}) \mathbf{v}_i \quad (48)$$

$$(\Sigma_{\lambda})_{ij} \equiv \frac{\kappa_i}{\kappa_i^2 + \lambda^2} \delta_{ij} \quad (49)$$

$$(50)$$

から求められる。ここに λ は特異値を安定化させるパラメタであり、regularization parameter と呼ばれる。

NGIM と Tikhonov Regularization の比較

図 3 には、観測データにノイズを加えた場合の NGIM と Tikhonov Regularization での Che mapping の結果が示されている。練習コードでは -s でノイズを付加できる。NGIM の場合は、

```
./random_light.py -f che.png -n 2500 -l 0.0 -p 0.7 -s 1.0
```

となる。NGIM の場合、overfit が起きてしまい、モデルの分散が大きくなってしまっている。結果、推定値がランダムノイズ的に見えてしまう。

一方、Tikhonov Regularization で適当な λ を設定して、regularize した場合、それなりに良く推定ができているのがわかる。Tikhonov regularization で Che mapping を行う場合、

```
./random_light.py -f che.png -n 2500 -l 3.0 -p 0.7 -s 1.0
```

のようにすればよい。 λ の値はこの例では 3.0 である (-l 3.0)。ノイズは 1.0 である (-s 1.0)

Truncated SVD による regularization も試してみよう。-lim option で 0 とみなす特異値を指定できる。

```
./random_light.py -f che.png -n 2500 -l 0.0 -p 0.7 -s 1.0 -lim 1.0
```

Tiknov regularization と misfit 関数

Tikhonov regularization は以下の misfit 関数を最小化することと同値である。

$$Q_{\lambda} = |G\mathbf{m} - \mathbf{d}^{\text{obs}}|^2 + \lambda^2 |\mathbf{m}|^2 \quad (51)$$

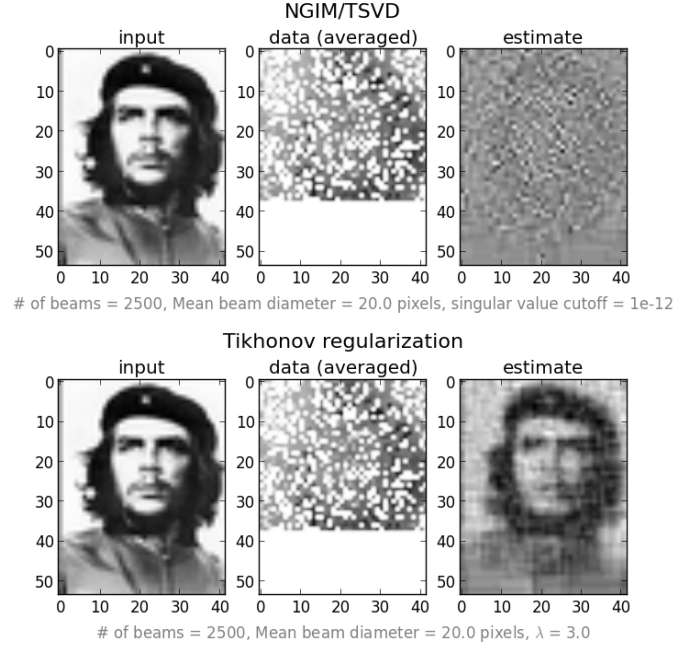


Figure 3: データにノイズが有る場合の NGIM(上) と Tikhonov Regularization(下) による Che mapping。NGIM (予測誤差最小) では、モデルの分散が大きすぎ何が書かれているか分からない。これは高周波成分が乗っているためである。そのため、この図をちょっと遠くから見るとなんとなくみえるようになる (この事実は小谷隆行氏指摘による)。Tikhonov Regularization では人だというのはわかるし、チェゲバラな気がする。

これは以下のように示される。式 (71) は

$$Q_\lambda = |G'm - d'|^2 \quad (52)$$

$$G' \equiv \begin{pmatrix} G \\ \lambda I \end{pmatrix} \quad (53)$$

$$d' \equiv \begin{pmatrix} d^{\text{obs}} \\ 0 \end{pmatrix} \quad (54)$$

のように変形できる。この関数の最小化は、単純に最小二乗解であるから、解を m^{est} とおくと、正規方程式

$$(G')^T [G'm^{\text{est}} - d'] = 0 \quad (55)$$

が成り立つ。すなわち

$$(G^T G + \lambda^2 I) m^{\text{est}} - G^T d^{\text{obs}} = 0 \quad (56)$$

であり、これを \mathbf{m}^{est} について解くと

$$\mathbf{m}^{\text{est}} = (G^T G + \lambda^2 I)^{-1} G^T \mathbf{d}^{\text{obs}} \quad (57)$$

$$= (V \Lambda^T \Lambda V^T + \lambda^2 I V V^T)^{-1} V \Lambda^T U^T \mathbf{d}^{\text{obs}} \quad (58)$$

$$= V(\Lambda^T \Lambda + \lambda^2 I)^{-1} V^T V \Lambda^T U^T \mathbf{d}^{\text{obs}} \quad (59)$$

$$= V(\Lambda^T \Lambda + \lambda^2 I)^{-1} \Lambda^T U^T \mathbf{d}^{\text{obs}} \quad (60)$$

となる。 $V V^T = I$ を利用している。 $(\Lambda^T \Lambda + \lambda^2 I)$ は対角行列であり $\lambda > 0$ である限り逆行列が存在していることに注意。ベクトル形式で書き直すと

$$\mathbf{m}^{\text{est}} = V \Sigma_\lambda U^T \mathbf{d}^{\text{obs}} \quad (61)$$

$$= \sum_{i=1}^{\min(N, M)} \frac{\kappa_i}{\kappa_i^2 + \lambda^2} (\mathbf{v}_i \mathbf{u}_i^T) \mathbf{d}^{\text{obs}} \quad (62)$$

$$= \sum_{i=1}^{\min(N, M)} \frac{\kappa_i}{\kappa_i^2 + \lambda^2} (\mathbf{u}_i^T \mathbf{d}^{\text{obs}}) \mathbf{v}_i \quad (63)$$

となり、確かに Tikhonov regularization となっている。

Model Prior

モデルの prior $\hat{\mathbf{m}}$ を考え、

$$Q_\lambda = |G \mathbf{m} - \mathbf{d}^{\text{obs}}|^2 + \lambda^2 |\mathbf{m} - \hat{\mathbf{m}}|^2 \quad (64)$$

を最小化した場合、Tikhonov Regularization はどうなるだろうか？この場合、

$$Q_\lambda = |G' \mathbf{m} - \mathbf{d}'|^2 \quad (65)$$

$$G' \equiv \begin{pmatrix} G \\ \lambda I \end{pmatrix} \quad (66)$$

$$\mathbf{d}' \equiv \begin{pmatrix} \mathbf{d}^{\text{obs}} \\ \lambda \hat{\mathbf{m}} \end{pmatrix} \quad (67)$$

のように変形できるから、正規方程式は

$$(G^T G + \lambda^2) \mathbf{m}^{\text{est}} - (G^T \mathbf{d}^{\text{obs}} + \lambda^2 \hat{\mathbf{m}}) = (G^T G + \lambda^2) (\mathbf{m}^{\text{est}} - \hat{\mathbf{m}}) - G^T (\mathbf{d}^{\text{obs}} - G \hat{\mathbf{m}}) = 0 \quad (68)$$

となる。これは式 (56) において、 $\mathbf{m}^{\text{est}} \rightarrow \mathbf{m}^{\text{est}} - \hat{\mathbf{m}}$ 、 $\mathbf{d}^{\text{obs}} \rightarrow \mathbf{d}^{\text{obs}} - G \hat{\mathbf{m}}$ と置き換えたものと同じなので、結局

$$\mathbf{m}^{\text{est}} = V \Sigma_\lambda U^T (\mathbf{d}^{\text{obs}} - G \hat{\mathbf{m}}) + \hat{\mathbf{m}} \quad (69)$$

$$= \sum_{i=1}^{\min(N, M)} \frac{\kappa_i}{\kappa_i^2 + \lambda^2} [\mathbf{u}_i^T (\mathbf{d}^{\text{obs}} - G \hat{\mathbf{m}})] \mathbf{v}_i + \hat{\mathbf{m}} \quad (70)$$

となる。

Maximum a Posterior (MAP) と Tikhonov Regularization

最小化 (71) は、 d の誤差が独立に 1、また m の prior として、平均 0 で variance が λ^{-2} の独立な (covariance matrix が対角行列である) Gaussian を仮定したときの、posterior likelihood の最大化に対応している。ここでもう少し一般的に

$$Q_\lambda = \sum_{i=1}^N \frac{|d_i^{\text{obs}} - G_{ij}m_j|^2}{\sigma_i^2} + \lambda^2 |\mathbf{m} - \hat{\mathbf{m}}|^2 \quad (71)$$

の最小化を考えると、この misfit 関数の最小化は、 d の誤差が独立に σ_i 、 m の prior として、平均 \hat{m} で variance が λ^{-2} の独立な Gaussian を仮定したときの、posterior likelihood の最大化に対応している。

Tarantola 2005 に従って、これを説明しよう。モデルの posterior distribution は、モデルの prior distribution を $\rho_M(\mathbf{m})$ 、データの distribution を $\rho_D(\mathbf{d})$ とすると、

$$p_M(\mathbf{m}) \propto \rho_M(\mathbf{m})\rho_D(\mathbf{d}) \quad (72)$$

のように書ける。ここで、モデルの prior として Gaussian

$$\rho_M(\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^M \det C_M}} \exp \left[-\frac{1}{2} (\mathbf{m} - \hat{\mathbf{m}})^T C_M^{-1} (\mathbf{m} - \hat{\mathbf{m}}) \right], \quad (73)$$

を仮定する (ここに C_M はモデルの covariance)。またデータの distribution も Gaussian を仮定する。

$$\rho_D(\mathbf{d}) = \frac{1}{\sqrt{(2\pi)^N \det C_D}} \exp \left[-\frac{1}{2} \boldsymbol{\varepsilon}^T C_D^{-1} \boldsymbol{\varepsilon} \right]. \quad (74)$$

すると、

$$p_M(\mathbf{m}) \propto \exp \left\{ -\frac{1}{2} [\boldsymbol{\varepsilon}^T C_D^{-1} \boldsymbol{\varepsilon} + (\mathbf{m} - \hat{\mathbf{m}})^T C_M^{-1} (\mathbf{m} - \hat{\mathbf{m}})] \right\}. \quad (75)$$

となるため、Posterior probability 最大値を探すことは、

$$Q = \boldsymbol{\varepsilon}^T C_D^{-1} \boldsymbol{\varepsilon} + (\mathbf{m} - \hat{\mathbf{m}})^T C_M^{-1} (\mathbf{m} - \hat{\mathbf{m}}). \quad (76)$$

を最小化する事と等しい。ここでデータもモデルも独立ガウシアンを仮定すると、 $(C_D)_{ij} = \sigma_i^2 \delta_{ij}$ 、 $(C_M)_{ij} = \lambda^{-2} \delta_{ij}$ となり、すなわち

$$Q_\lambda = \sum \frac{|d_i - G\mathbf{m}|^2}{\sigma_i^2} + \lambda^2 |\mathbf{m} - \hat{\mathbf{m}}|^2. \quad (77)$$

を最小化する事が、posterior likelihood の最大化に対応する。

6 L-curve Criterion

さて Tikhonov regularization において、 λ をどのように決めるべきであろうか？この一つの基準として有効なのが L-curve criterion と呼ばれるものである (Hansen 2010)。L-curve とは、モデルのノルム $\xi \equiv |\mathbf{m}^{\text{est}} - \hat{\mathbf{m}}|^2$

と予測誤差 $\rho \equiv |G\mathbf{m}^{\text{est}} - \mathbf{d}|^2$ を、 λ をパラメタとしてプロットしたものである。図 4(左) は、L curve の一例である。この例では λ を 0.01 から 10.0 まで動かしてプロットしている。

小さい λ では、モデルの分散が大きく、すなわちモデルのノルムも大きい値である (図で左上)。 λ を大きくしていくとモデルのノルムは下がっていくが、あるところで下がり方が鈍り、予測誤差の急激な増大が始まる (右下方向)。この中間点 (図で赤点のあるあたり) が選ぶべき λ であろう。

これは言い換えると、対数空間で $\log \rho - \log \xi$ での曲率

$$c(\lambda) \equiv -2 \frac{(\log \rho)'(\log \xi)'' - (\log \rho)''(\log \xi)'}{[(\log \rho)'^2 + (\log \xi)'^2]^{3/2}}, \quad (78)$$

が最大になるところが最適値であるといえる。ここに ' は λ による微分を表す。ここで

$$\xi' = -\frac{4}{\lambda} \sum_{i=1}^M [1 - w_i(\lambda)] w_i(\lambda)^2 \frac{\mathbf{u}^T(\mathbf{d} - G\hat{\mathbf{m}})}{\kappa_i^2} \quad (79)$$

$$\rho' = -\lambda^2 \xi' \quad (80)$$

$$w_i(\lambda) \equiv \frac{\kappa_i^2}{\kappa_i^2 + \lambda^2}. \quad (81)$$

を利用すると、

$$c(\lambda) = -2 \frac{\xi \rho}{\xi'} \frac{(\lambda^2 \xi' \rho + 2\lambda \xi \rho' + \lambda^4 \xi \xi')}{(\lambda^4 \xi^2 + \rho^2)^{3/2}}, \quad (82)$$

となる。 $c(\lambda)$ が最大になる λ を探せばよい。図 4(右) は曲率 $c(\lambda)$ を示している。曲率の最大をとれば、たしかに L-curve の角にあたる場所を検出できることがわかる。

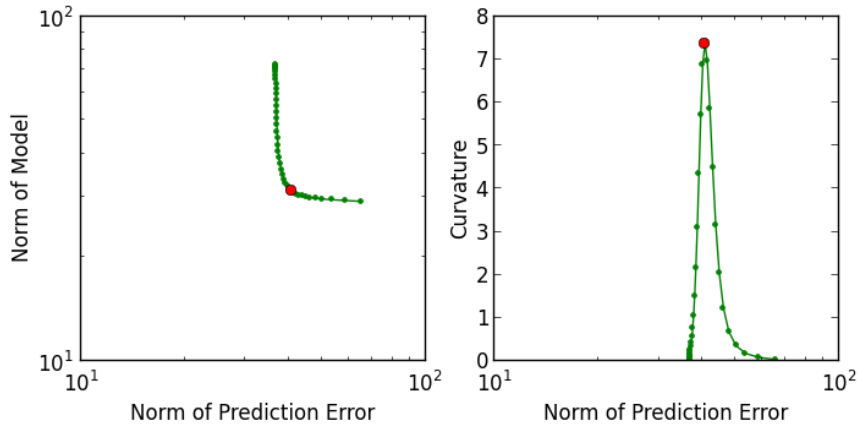


Figure 4: L-curve (左) と曲率 (右)。曲率が最大の赤い点に対応する λ が L-curve criterion のえらぶ λ である。

附属プログラムで L-curve criterion を用いるには以下のように行う。

```
./random_light.py -f che.png -n 2500 -L 0.01 100.0 -p 0.7 -s 1.0
```

この場合、 $\lambda = 0.01$ から $\lambda = 100.0$ まで探索している。

7 最後に

パラメタが縮退するからといって、あきらめる必要はない。適当な先験情報を与えることで、モデルの主要な特徴を再構築することができる。以下の参考文献は逆問題 / フィット等の理論的背景や様々なテクニックが紹介されている。

- Menke (1989) *Geophysical Data Analysis: Discrete Inverse Theory* : 逆問題の基本的知識が集約されている。日本語訳が古今書院からでている。
- Tarantola (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation* : Bayesian 的な観点から、逆問題の解法を整理している。どちらかというと理論的理解が主。
- Hansen (2010) *Discrete Inverse Problems: Insight and Algorithms* : Tikhonov regularization、特に regularization parameter の選び方の一つである L-curve criterion について詳しく書かれている。応用が容易なように書かれている。
- Lawson & Hanson (1974,1995) *Solving least squares problems* : 線形不等式拘束条件のルーチン等、実際の様々な先験条件のもとでの解法がされている。1995 年の改訂版のほうがたぶんよい。テキスト中のコードは netlib で手に入れることができる。