

# 情報量基準 ABIC によるデータの当てはめの制約条件付き問題への適用

村田陸<sup>†</sup> 田中亮平<sup>†</sup> 坂本真貴人<sup>†</sup> 藤井昭宏<sup>†</sup> 田中輝雄<sup>†</sup>

「誤差」を含んだ実測等のデータからその「構造」を抽出するデータあてはめ(Data Fitting)の方法として、離散点上の値をパラメタとして表現した近似関数 $f$ をベイズ型赤池情報量基準(Akaike's Bayesian Information Criterion)を用いて評価し、推定する方法がある。この方法では、推定すべきデータの「構造」に対して何らかの事前知識がわかっているとき、これを $f$ に取り込むことが比較的容易である。 $f$ に関する事前知識を制約条件として表現し、制約条件付き最適化問題として解く。本論文では、この方法を実装し、変化の急な関数の近似で発生する Undershoot の削除や、物理実験の実測データの関数形状の成型などへの適用例を示した。

## 1. はじめに

「誤差」を含んだ実測、調査データなどからそのもの「構造」を抽出するのがデータあてはめ (Data Fitting) である。データのあてはめのひとつの方法として、離散点上の値をパラメタとして表現した近似関数  $f$  をベイズ型赤池情報量基準 ABIC (Akaike's Bayesian Information Criterion)[1]を用いて評価し推定する[2]。この方法では、以下の評価式を最小化する  $f$  を最良近似関数とする。

$$\| \text{データ } \mathbf{y} \text{ と } f \text{ の距離} \|^2 + \alpha^2 \| f \text{ の滑らかさ} \|^2$$

ここでスカラー  $\alpha$  は、第一項と第二項のバランス調整量であり、 $\alpha$  が大きいほど滑らかさが強く(直線に近づく)、 $\alpha$  が小さいとデータ  $\mathbf{y}$  により追従するようになる。ABIC を用いてこの  $\alpha$  を調整し、そのバランスを決定する。離散点  $x_j$  上の値  $f_j$  で表わされ、滑らかさのみを仮定する近似関数  $f$  を離散スプライン(d-Spline)と呼ぶ。

通常、推定すべき「構造」に対して何らかの事前知識として、単調性、非負性などがわかっていたとしても、これを通常の近似関数に組み込むことは容易ではない。逆に、単調性や非負性などの条件を満たす近似関数を用いると、それにより副作用として不自然性を近似関数に与え、もとの「構造」に対する近似を悪くしてしまう可能性がある。d-Spline はこのような副作用をおさえ、単調性、非負性などの事前情報を制約条件として組み込み、制約条件付き最適化問題として解く。本論文では、d-Spline を用いての実装を行い、急激な変化をする関数を近似するときに見れる Undershoot の減少を取り除くことや、物理実験における関数形状の成型などの問題への適用例を示す。

## 2. ABIC を用いたデータあてはめ

推定すべき近似関数 d-Spline  $f$  を離散点  $x_j$  ( $1 \leq j \leq n$ ) 上の値  $f_j = f(x_j)$  とする。ここで  $t$  は転置をあらわす。

$$\mathbf{f} = (f_1, f_2, f_3, \dots, f_j, \dots, f_n)^t$$

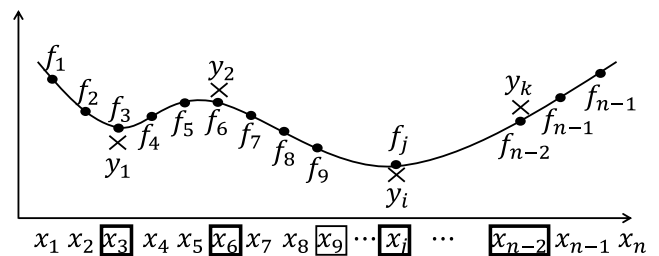


図1 データ  $\mathbf{y}$  と d-Spline  $f$  の概形  
Fig.1 Shape of d-Spline fitting data  $\mathbf{y}$

いま  $N$  個のデータ  $y_i$  ( $1 \leq i \leq N$ ) が得られているとする。

$$\mathbf{y} = (y_1, y_2, y_3, \dots, y_i, \dots, y_N)^t$$

一般に、 $N < n$  である。ここで、 $n$  は  $N$  に対して十分大きく取る。実測されたデータである  $y_i$  が持つ誤差は  $N(0, \sigma^2)$  に独立にしたがうと仮定すると、 $\mathbf{y}$  の確率分布は以下のよう表現できる。

$$p(\mathbf{y} | f, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - E\mathbf{f}\|^2\right) \quad (1)$$

ここで  $\|\cdot\|$  はユークリッドノルムを表し、また、 $E$  は  $\mathbf{y}$  と d-Spline  $f$  の対応を示す  $N \times n$  の行列である。 $E$  の要素を  $e_{ij}$  とすると、 $f_j$  上に  $y_i$  があるときのみ  $e_{ij} = 1$  であり、それ以外は  $e_{ij} = 0$  となる。 $E$  の形状を図2に示す。

$$E = \begin{bmatrix} 1 & & & 0 \\ & 001 & & \\ & & 01 & \\ 0 & & & \ddots \end{bmatrix}$$

図2  $E$  の形状  
Fig.2 Shape of  $E$

これだけでは、データ数  $N$  より  $f$  のパラメタ数  $n$  の方が大きいので  $f$  は一意には定まらない。そこで  $f$  に関する情報として、「 $f$  は滑らかである」という仮定を置き  $f$  の滑らかさを二階差分  $|f_{j-1} - 2f_j + f_{j+1}|$ , ( $2 \leq j \leq n-1$ )

<sup>†</sup> 工学院大学

で表現し、滑らかさの程度を $\|dDf\|^2$ を用いて測る。ここで、 $D$ は $n-2 \times n$ の行列である。その形状を図3に示す。

$$D = \begin{bmatrix} 1 & -2 & 1 & & 0 \\ & 1 & -2 & 1 & \\ & & 1 & -2 & 1 \\ 0 & & & & \ddots \end{bmatrix}$$

図3  $D$ の形状  
 Fig.3 Shape of  $D$

d-Spline で仮定しているのは、この二階差分による滑らかさだけである。つまり、d-Spline  $f$  は微分可能性までを仮定した連続型モデル（3次スプラインなどのように）とはならない。このことが本報告で用いる近似関数  $f$  を離散スプライン（d-Spline）と呼ぶ理由である。

ここで、ハイパーパラメタ  $d$  を導入し、 $f$  の事前分布として、 $f$  の滑らかさを式(2)と表現する。 $\varphi$  は  $D^t D$  の非零固有値の積である。

$$\pi(f|d) = \left(\frac{1}{\sqrt{2\pi}}\right)^{n-2} d^{n-2} \varphi \exp\left(-\frac{1}{2}\|dDf\|^2\right) \quad (2)$$

ここでこの式(2)を式(1)の事前分布として  $f$  の事後分布はベイズの定理により次の式で表わされる。

$$q(f|y) = \frac{p(y|f, \sigma^2) \pi(f|d)}{\int p(y|f, \sigma^2) \pi(f|d) df}$$

この関数が最大になるような  $f$  を求める。そのためには、ハイパーパラメタ  $\sigma^2, d$  を求める必要がある。これを ABIC 最小の基準で決定する。ABIC を次式に示す。

$$ABIC = -2 \ln L(\sigma^2, d) \quad (3)$$

ここで、 $L(y|\sigma^2, d)$  は  $\sigma^2, d$  による周辺尤度であり、事後分布の正規化項に等しい。

$$L(y|\sigma^2, d) = \int p(y|f, \sigma^2) \pi(f|d) df$$

いま  $\alpha = d\sigma$  とおくと、

$$L(y|\sigma^2, d) = \left(\frac{1}{\sqrt{2\pi}}\right)^{N-2} \left(\frac{1}{\sigma}\right)^{N-2} \alpha^{n-2} \varphi |\det(z_\alpha^t z_\alpha)|^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \|b - z_\alpha f\|^2\right),$$

$$b = \begin{bmatrix} y \\ 0 \end{bmatrix}, z_\alpha = \begin{bmatrix} E \\ \alpha D \end{bmatrix} \quad (4)$$

となる。ここで、 $\|b - z_\alpha f\|^2$  は  $\|b - Z_\alpha f\|^2$  の最小二乗

残差である。尤度方程式

$$\frac{\partial}{\partial \sigma^2} L(\sigma^2, \alpha) = 0$$

に  $\sigma^2$  の最適値  $\hat{\sigma}^2$  を導入すると、このような式が得られる。

$$\hat{\sigma}^2 = \frac{1}{N-1} \|b - Z_\alpha \hat{f}\|^2$$

この  $\hat{\sigma}^2$  に関する上記の式を式(4)に代入すれば、次式のように ABIC を  $\alpha$  の関数として扱える。

$$ABIC(\alpha) = \ln(|\det(Z_\alpha^t Z_\alpha)|) - 2(n-2) \ln(\alpha) + (N+2) \ln(\|b - Z_\alpha \hat{f}\|^2) + C \quad (5)$$

式(5)を解くために、まず、 $Z_\alpha = QR$  と QR 分解することで、式(5)の第1項は

$$\ln(|\det(Z_\alpha^t Z_\alpha)|) = \ln(|\det(R^t R)|) = 2 \sum_i \ln(|r_{ii}|)$$

から数値的に計算することができる。また、第2項の最小二乗問題

$$\min_f \|b - Z_\alpha f\|^2 \quad (6)$$

も、QR 分解の結果から、 $Q^t b = Rf$  とし、後退代入によって  $\hat{f}$  を求めることができる。 $Q^t b$  と  $R$  をそれぞれ

$$Q^t b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

と分けると、 $\|Q^t b - R\hat{f}\|^2$  の最小二乗残差は

$$\|b - Z_\alpha \hat{f}\|^2 = \|b_2\|^2$$

と求めることができ、ABIC( $\alpha$ ) を数値的に計算することができる。すなわち、式(5)の ABIC を最小とする  $\alpha$  を求めるために  $\alpha$  を変化させて絞り込む。このとき  $\alpha, f$  を求めるアルゴリズムは以下ようになる。

Step 1	$\alpha$ を設定
Step 2	$Z_\alpha = QR$
Step 3	ABIC 最小ならば 6 へ
Step 4	$\alpha$ を変更
Step 5	2 へ
Step 6	$Q^t b = Rf$ を解く

図4 制約条件なしのアルゴリズム

Fig.4 Algorithm for Data Fitting

実際の計算では、式(5)は凸性の保証がないので大域的にいくつかの点を選んで ABIC( $\alpha$ ) を計算し、ある程度区間を狭めてから黄金分割による区間縮小法を用いる。

### 3. 制約条件付データあてはめ

d-Spline はその性質上推定すべきデータの構造に関して何らかの事前知識が与えられているときに、これを  $f$  に取り込むことが容易である。この事前知識の表現の手段として、データをある点で二つの相に分割できる二相問題 [3], 事前知識を制約条件として表現する制約条件付き問題などが考えられる。本論文ではこの制約条件付き問題について取り扱う。

推定すべきデータの構造に対して、与えられている事前知識を制約条件として表現し、制約条件なしの ABIC を用いた d-Spline によるデータのあてはめに組み込むことを考える。ここで扱う制約条件は、線形関数で表せるとする。たとえば、次のような制約条件が考えられる。

- 正値性  $f_i \geq 0, (1 \leq i \leq n)$ ,
- 単調性  $-f_i + f_{i+1} \geq 0, (1 \leq i \leq n-1)$ ,
- 凸性  $-f_{i-1} + 2f_i - f_{i+1} \geq 0, (2 \leq i \leq n-1)$ ,
- 固定点  $f_i = c_i$ .

$g_i \in R^n (1 \leq i \leq k)$  を線形制約条件を表すベクトルとすると、上記のような制約条件は式(7)となる。ここで、 $k$  は制約条件の数である。

$$g_i^t f \geq c_i, \quad (1 \leq i \leq k) \quad (7)$$

この制約条件の取り込みを二次計画問題(QP)として、式(8)と定式化する。

$$QP \begin{cases} \min_f \frac{1}{2} \|b - Z_\alpha f\|^2 \\ g_i^t f \geq c_i, \quad (1 \leq i \leq k) \end{cases} \quad (8)$$

このとき、本来は式(7)を  $f$  の事前分布として定義し、ABIC で  $\alpha$  を再度評価し設定しなければならない。しかしながら、これは計算量が膨大となり現実的ではない。そのため、制約条件なしのアルゴリズム (図4参照) で求めた  $\alpha$  に固定して式(8)を解く。ここで式(8)の Karush-Kuhn-Tucker 条件は式(9)となる。式(9)を解くことで式(8)の最適解を求める。

$$\begin{cases} Z_\alpha^t Z_\alpha f - G^t \lambda = Z_\alpha^t b \\ Gf \geq c, \\ \lambda \geq 0, \quad \lambda^t (Gf - c) = 0 \end{cases} \quad (9)$$

$$G = \begin{bmatrix} g_1^t \\ g_2^t \\ \vdots \\ g_k^t \end{bmatrix}$$

ここで、 $G$  は制約条件  $g_i^t$  を列方向に並べた  $k \times n$  行列で、 $\lambda$  はラグランジュ乗数である。式(9)を有効制約戦略[4]を用いて解く。そのアルゴリズムの概略を以下に示す (図5参照)。

Step 1	初期の $\alpha, f$ を制約条件なしで求める
Step 2	$k = 0$
Step 3	$I^{(0)} = \{i   g_i^t f \leq c_i\}$ なる添字集合を生成
Step 4	$I^{(k)}$ 内の添字からなる $G$ を $\tilde{G}$ として生成し、これに対応する $\lambda, c$ を $\tilde{\lambda}, \tilde{c}$ とする
Step 5	$\begin{bmatrix} Z_\alpha^t Z_\alpha & -\tilde{G}^t \\ -\tilde{G} & 0 \end{bmatrix} \begin{bmatrix} f \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} Z_\alpha^t b \\ -\tilde{c} \end{bmatrix}$ を解く
Step 6	$k = k + 1$
Step 7	$I^{(k)} = \{i   g_i^t f \leq c_i\} \cup \{i \in I^{(k-1)}   \tilde{\lambda} \geq 0\}$
Step 8	$I^{(k)} \equiv I^{(k-1)}$ ならば終了 さもなくば 4 へ

図5 制約条件付きのアルゴリズム

Fig.5 Algorithm for Data Fitting with Constrained Optimization

ここで、図5の Step 5 の連立方程式を  $f, \tilde{\lambda}$  について解く方法について説明する。

$$\begin{cases} Z_\alpha^t Z_\alpha f - \tilde{G}^t \tilde{\lambda} = Z_\alpha^t b \\ -Gf = -\tilde{c} \end{cases} \quad (10)$$

$$\begin{cases} Z_\alpha^t Z_\alpha f - \tilde{G}^t \tilde{\lambda} = Z_\alpha^t b \\ -Gf = -\tilde{c} \end{cases} \quad (11)$$

式(10)を式変形すると次式を得る。

$$f = (Z_\alpha^t Z_\alpha)^{-1} Z_\alpha^t b + (Z_\alpha^t Z_\alpha)^{-1} \tilde{G}^t \tilde{\lambda}$$

ここで、 $Z_\alpha = QR$  と分解すると、

$$f = R^{-1} Q^t b + R^{-1} R^{-t} \tilde{G}^t \tilde{\lambda},$$

さらに、 $R^t A = \tilde{G}^t$  とおくと、

$$Rf = Q^t b + A\tilde{\lambda},$$

となる。これを式(11)に代入し式変形すると、

$$-A^t A\tilde{\lambda} = A^t Q^t b - \tilde{c}$$

を得て、 $A = Q'R'$  と分解すれば

$$-R'^t R\lambda = A^t Q^t b - \tilde{c}$$

となる。これをまとめると、

$$\begin{cases} Rf = Q^t b + A\tilde{\lambda} \\ -R'^t R\lambda = A^t Q^t b - \tilde{c} \end{cases}$$

$$Z_\alpha = QR, A = Q'R', R^t A = \tilde{G}^t$$

となる。 $R, R'$  はそれぞれ三角行列となっているため、 $f, \lambda$  を後退代入・前進代入によって求めることができる。 $Z_\alpha$  はすでに制約条件なしのアルゴリズム (図4参照) で QR 分解されているため、必要な計算回数は  $Rx = y$  の形式の方程式が 4 回と QR 分解が 1 回となる。

#### 4. 適用例

この方法をいくつかの関数・実験データに適用した。この計算は以下の環境で行った。

CPU	AMD Phenom 9600@2.3GHz
Memory	4.0GB (DDR3-1333)
OS	Fedora 19
Compiler	gcc 4.7.2

図 6 測定環境

Fig.6 Measurement Environment

##### 4.1 Undershoot を削除する適用例

傾きの変化が急な部分のある関数に対してデータのあてはめを行うとき、しばしば傾きの大きい部分に引っ張られて Undershoot が発生することがある。図 7 では例として、直線を円に沿うように曲げた曲線のデータあてはめに Undershoot が発生している様子を示している。実線は本来の関数を表し、点線は Undershoot が生じたときの曲線を表す。この Undershoot は本来の関数にない「構造」であり、近似として好ましくない。今回例として用いた関数の一次導関数  $\frac{d}{dx}f(x)$  は常に正であり、単調増加性の条件を満たしているので、この近似を単調増加性の制約条件を付加して行うことでこの「構造」を削除することを試みる。

この例では、図 7 のように直線を円に沿うように曲げた曲線に誤差を付加したものをデータとして与え、それに対してデータあてはめを行った。この誤差はそれぞれ独立に  $N(0,0.5)$  に従うものとした。直線の長さ  $l$  に対して円の半径  $r$  を  $r = l/4$  と定め、角度は  $\pi/6$  まで曲げるものとした。この曲線の式を以下に示す。

$$f(x) = \begin{cases} 0 & (0 \leq x \leq \frac{l}{2}) \\ \sqrt{r^2 - x^2} + r & (\frac{l}{2} \leq x \leq \frac{\sqrt{3}}{2}r) \\ \sqrt{3}x - 5 & (\frac{\sqrt{3}}{2}r \leq x \leq l) \end{cases}$$

図 8 にあてはめの結果を示す。図 9 は変化の大きくなる近傍を拡大したもので、図 10 はあてはめの結果の一次差分を表す。横軸に変数  $x$  をとり、縦軸に関数値をとった。  $x$  で表現される点列は誤差を付加したデータ、点線が真の関数  $f(x)$ 、破線が制約条件なしのデータあてはめの結果、実線が制約条件付きのデータあてはめである。図 10 から一次差分が負となっている部分があり、この例におけるデータあてはめでは、変化の大きい部分で本来の関数にない性質である Undershoot が生じていることがわかる。ここで  $f$  に制約条件として単調増加性を付加すると、一次差分が全体で非負となり制約によって Undershoot が取り除かれたことがわかる。

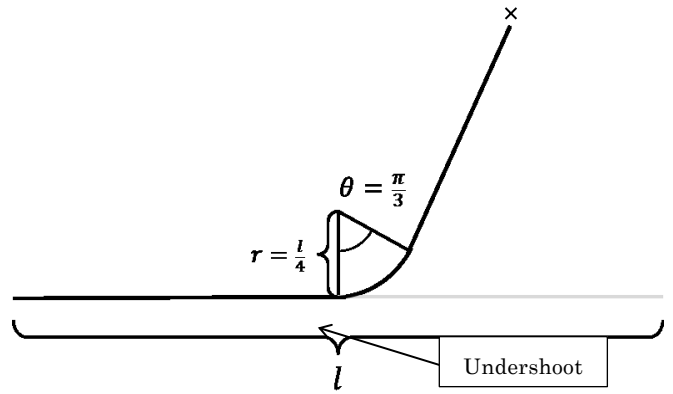


図 7 円に沿って曲げた曲線

Fig.7 Shape of a Curve Fitting to Circle

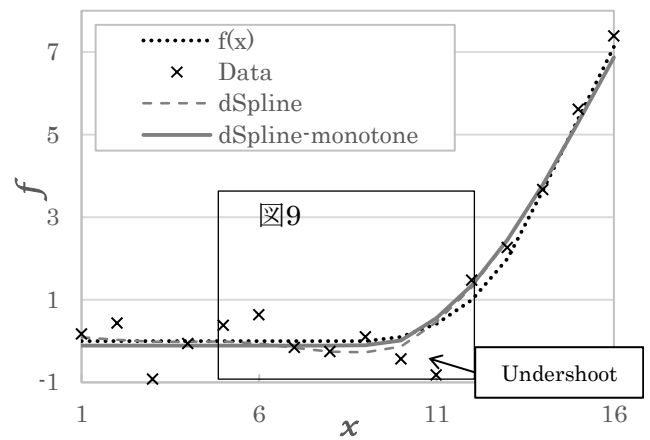


図 8 円に沿った曲線に対するデータあてはめ

Fig.8 Data Fitting for a Curve Fitting to Circle

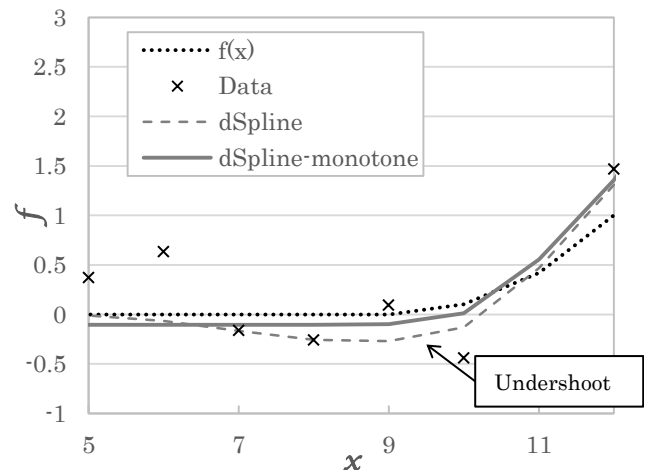


図 9 図 8 の拡大図

Fig.9 Enlargement of Fig.8

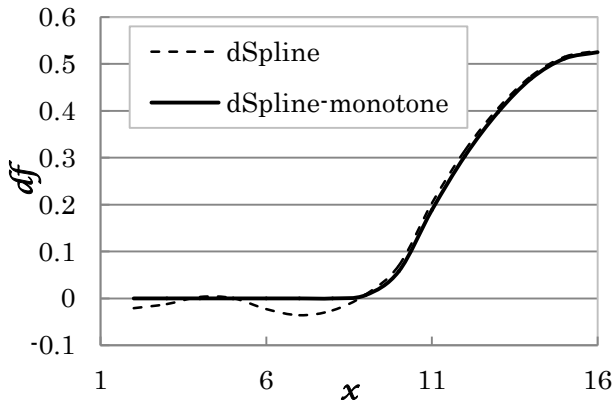


図 10 データあてはめ結果の一次差分  
 Fig.10 Primary Difference of d-Spline

4.2 実問題(Almel-Chromel 熱電対実験データ)への適用例

実問題への適用例として, Almel-chromel 熱電対の錫の冷却過程での起電力のデータに対して d-Spline によるあてはめを行った[5]. この実験で重要なのは, 錫の凝固点における熱起電力を推定することである. このデータでは冷却過程で過冷却が起きており, データ中に大きな落ち込みが発生している. これを取り除き, 凝固点における熱起電力を推定するために単調減少性を付加してデータあてはめを行った. データあてはめの結果を図 11 に示す. 横軸に経過時間[s]をとり, 縦軸には熱電対の熱起電力[mV]をとった. ×で表現される点列が与えられた実測データであり, 点線が制約条件なしのデータあてはめ結果, 実線が単調減少性を付加したデータあてはめ結果となっている. グラフから単調減少性が付加されたことで, データの落ち込みが取り除かれているのがわかる. 本論文で使用したアルゴリズムでは, 制約条件の領域を徐々に拡大して計算を行う. 図 11 で示すように, loop1 では初期の制約条件無しでのデータあてはめの結果のうち, 制約条件を満たさない部分に制約条件を付加し, 計算を行う. loop2 では loop1 の範囲と loop1 で新しく制約条件を満たさなくなった範囲を新しい制約条件の範囲として再び計算する. この反復を近似関数全体が制約条件を満たすまで続ける. このデータにおいては loop10 まで反復計算を行った. この挙動が図 5 中の Step4-8 の反復構造に対応している.

単調性のみを付加したとき, 広範囲が直線で近似され, 近似関数全体がデータの落ち込みに引き付けられる現象を生じた. これは近似として好ましくない. この場合は落ち込みの直後の点などに固定点の条件を付加することで, 制約条件領域の拡大を固定点で制御することができる. 図 12 に 図 11 を拡大し, 固定点条件を付加したデータあてはめ結果を実線で追加した. 図 11 から, 固定点条件を付加したものでは固定点までは直線で近似され, それ以降はデータに追従していることがわかる. この場合直線部分の関数値は固定点の値となる.

固定点の位置を推定する方法の一つの案として, ABIC

による再評価が考えられる. 各  $f_j, 1 \leq j \leq n$  を固定点とにおいて, 最も ABIC を小さくするような固定点の位置を決定する. 2章の式(4)より, ABIC は,  $\alpha$  および  $\|b - Z_\alpha f\|^2$  によって決定するため, 実質的には固定点の中で  $\|b - Z_\alpha f\|^2$  を最小にするものを選ぶこととなる. 今回のケースでは, 点(410,9.39)が固定点として推定された. この計算は各  $f$  の点を固定点として計算するため, 計算は  $n$  回行われる. そのため計算時間は  $n$  に比例して大幅に増大する.

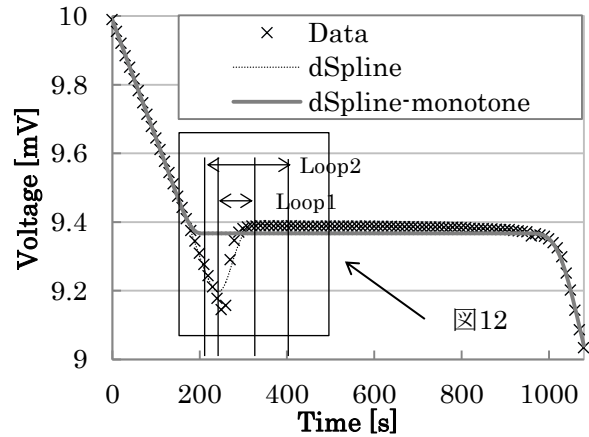


図 11 almel-chromel 熱電対の起電力データに対するデータあてはめ

Fig.11 Data Fitting for Almel-Chromel Thermocouple Data

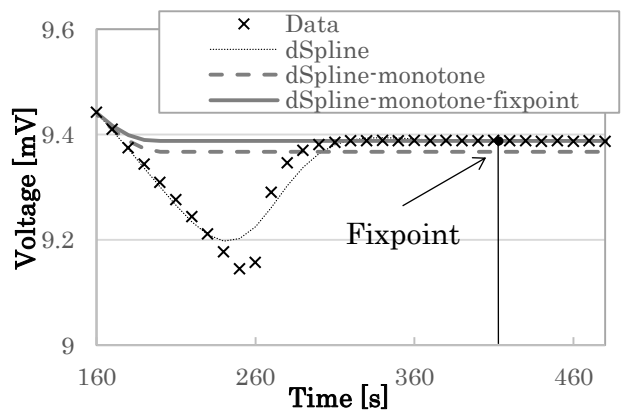


図 12 図 11 の拡大図

Fig.12 Enlargement of Fig.11

## 5. おわりに

本論文では、ABICを用いたd-Splineによるデータあてはめの方法と、それに対して制約条件を付加して解くアルゴリズムを実装した。実際に、単調性を満たす急激な変化のある関数に、この方法でデータあてはめを行い、Undershootなどの好ましくない「構造」を削除する例を示し、また実問題への適用例としてAlmel-Chromel熱電対の冷却過程のデータに対してデータの落ち込みを削除する例を示した。この例では、データの落ち込みによって近似関数が引き付けられ、データへの追従性が悪くなる。これに対する方法として、固定点条件を適切な位置に追加することで、より適当なあてはめ結果が得られることを示した。しかし、この固定点の適切な位置の推定方法は現在検討中である。

今後の課題として、更なる適用先の拡大や、固定点位置を推定する方法の開発、計算速度や精度から二次計画問題の別の解法の検討、本手法の別の応用である二相問題との組み合わせなどが考えられる。

## 謝辞

本論文で利用した物理実験データは工学院大学基礎・教養教育部門教授 渡部隆史氏より頂きました。

## 参考文献

- [1] Akaike, H., Likelihood and Bayes procedure, In Bayesian Statistics, J.M.Bernardo, M.H.DeGroot, D.V.Lindley and A.F.M.Smith, eds, University Press, Valencia, Spain, pp.143-166(1980).
- [2] 田中輝雄, 田辺國士, バイズの方法によるデータのあてはめ, 京都大学数理解析研究所講義録, vol.483, no.5, pp.86-111 (1983).
- [3] 田中亮平, 村田陸, 坂本真貴人, 藤井昭宏, 田中輝雄, ABICを用いたデータあてはめの二相問題への適用, 情報処理学会第76回全国大会, no.4K-1 (2014)(掲載予定).
- [4] 今野浩, 山下浩, 非線形計画法, 日科技連出版, pp.58,263(1978).
- [5] 本河光博, 三浦登, 物理学実験講座2 基礎技術II 一実験環境技術一, pp.59, 丸善出版(1999).