

有限要素近似による曲面回帰

弘前大学大学院理工学研究科 嶋中稔人 (Narihito Shimanaka)

陳小君 (Xiaojun Chen)

Graduate School of Science and Technology,
Hirosaki University

概要

本論文では有限要素基底関数による曲面回帰, およびその応用として交通量の推定について論ずる。

薄板スプライン (Thin Plate Spline) は滑らかな曲面の生成に適しているが, 連立一次方程式の行列の大きさは測定点の数に依存する。一方, 薄板スプライン有限要素法 (Thin Plate Spline Finite Element Method) は測定点の数ではなく基底関数の数に依存する。よって行列の大きさを操作することができるため, 一般に大量のデータに対して用いられる。また, もっとも簡単な基底関数, すなわち区分的線形基底関数を用いているので個々の行列計算も複雑ではなく, 扱いやすいという特徴を持つ。薄板スプライン有限要素法の理論と行列の解法, そして実際の例に対する数値計算を紹介する。

次に曲面回帰による交通量の推定について論ずる。この研究は青森県における橋の最適管理を目的としている。青森県では, 高度経済成長時代後期の 1970 年以降に建設が集中して行われたため, 今後大量に更新時代が到来することが予測される。また, 塩害や凍害が発生しやすい地域であり, 劣化や亀裂が発生している。だが, 青森県の財政状況を考えるとこのような橋をすべて修復することはできない。そこで, 橋を修復する優先順位が必要となる。優先度を決定する要因は多様であるが, ここでは重要な要因のひとつである交通量という視点から考える。

青森県津軽半島の道路, または橋で測定された座標と交通量のデータから橋の最適管理を目的に津軽半島全域における交通量を推定することを試みる。

1 有限要素法による曲面回帰

1.1 薄板スプライン有限要素法

$\Omega \subset \mathbb{R}^2$ を有界な単連結開領域とする。 $H^1(\Omega), H^2(\Omega)$ をソボレフ空間とする。

標準の曲面における薄板スプラインとは汎関数

$$M_\alpha(f) = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 + \alpha \int_{\Omega} \{f_{x_1 x_1}(x)^2 + 2f_{x_1 x_2}(x)^2 + f_{x_2 x_2}(x)^2\} dx \quad (1)$$

を最小にする関数 $f_\alpha \in H^2(\Omega)$ を考えることである。ここで $x = (x_1, x_2)^T$ を変数, $x_i = (x_{i,1}, x_{i,2}) \in \Omega$ を測定点, $y_i \in \mathbb{R}, i = 1, \dots, n$ とする。 n 個の測定点はすべて異なり, 同一直線上にないものとする。 α は滑らかさのパラメーターである。

有限要素基底関数で近似するが, 2 階微分の計算をするには多くの計算を必要とする。そこで問題の複雑さを軽減させるためにもっとも簡単な基底関数, すなわち区分的線形基底関数を使う。ここでは,

$$f(x) \approx u_0(x) + c_0 + c_1 x_1 + c_2 x_2 \quad u_0 \in H^1(\Omega) \quad (2)$$

で近似することを考える。

$$\min J_\alpha(u_0, u_1, u_2, c) \quad (3)$$

$$\text{s.t. } (\nabla u_0, \nabla v) = (u, \nabla v) \quad \forall v \in H^1(\Omega) \quad (4)$$

$$\int_\Omega u_0 dx = 0 \quad (5)$$

ここで, $u_0, u_1, u_2 \in H^1(\Omega)$, $c \in \mathbb{R}^3$

$$J_\alpha(u_0, u_1, u_2, c) = \frac{1}{n} (Pu_0 + Xc - y)^T (Pu_0 + Xc - y) + \alpha(|u_1|_1^2 + |u_2|_1^2)$$

$$u = (u_1, u_2)^T, \quad c = (c_0, c_1, c_2)^T \in \mathbb{R}^3,$$

$$Pu_0 = [u_0(x_1), \dots, u_0(x_n)]^T, \quad y = [y_1 \cdots y_n]^T \in \mathbb{R}^n$$

$$X = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix}^T \in \mathbb{R}^{n \times 3}, \quad |u|_1 = \left[\int_\Omega \{(u_{x_1})^2 + (u_{x_2})^2\} dx \right]^{1/2}$$

(4) は u が ∇u_0 の近似であることを, (5) は u_0 の平均値が 0 であることを示している。

次に $u_j(x) = \sum_{i=1}^m u_{j,i} b_i(x)$ で置き換え, 離散化する。ここで, $b_1(x), \dots, b_m(x)$ は区分的線形基底関数である。 \mathbb{R}^{3m+3} における制約条件をもつ最適化問題

$$\min \hat{J}_\alpha(u_0, u_1, u_2, c) \quad (6)$$

$$\text{s.t. } Au_0 = B_1 u_1 + B_2 u_2 \quad (7)$$

を考えればよい。ここで,

$$\hat{J}_\alpha(u_0, u_1, u_2, c) = \frac{1}{n} \|Nu_0 + Xc - y\|_2^2 + \alpha(u_1^T Au_1 + u_2^T Au_2),$$

$$[A]_{i,j} = \int_\Omega \{(b_i)_{x_1} (b_j)_{x_1} + (b_i)_{x_2} (b_j)_{x_2}\} dx \in \mathbb{R}^{m \times m},$$

$$[B_k]_{i,j} = \int_\Omega (b_i)_{x_k} b_j dx \in \mathbb{R}^{m \times m}, \quad k = 1, 2,$$

$$[N]_{i,j} = b_j(x_i) \in \mathbb{R}^{n \times m}, \quad u_j = (u_{j,1}, \dots, u_{j,m})$$

である。 A はラプラシアン の有限要素近似行列である。 B_1 と B_2 はそれぞれ, 偏微分 $\partial/\partial x_1$ と $\partial/\partial x_2$ の近似行列である。

目的関数とラグランジュ乗数 $w = (w_1, \dots, w_m)^T$ を制約条件にかけたものを加えた関数, ラグランジュ関数

$$L(u_0, u_1, u_2, c, w) = \hat{J}_\alpha(u_0, u_1, u_2, c) + w^T (Au_0 - B_1 u_1 - B_2 u_2) \quad (8)$$

を考える。それぞれ u_0, u_1, u_2, c, w について偏微分をすると, 以下の線形連立方程式

$$\begin{bmatrix} \alpha A & 0 & 0 & 0 & -B_1^T \\ 0 & \alpha A & 0 & 0 & -B_2^T \\ 0 & 0 & M & F & A \\ 0 & 0 & F^T & E & 0 \\ -B_1 & -B_2 & A & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_0 \\ c \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ n^{-1} N^T y \\ n^{-1} X^T y \\ 0 \end{bmatrix} \quad (9)$$

を得る。ここで、 $M = n^{-1}N^T N \in \mathbb{R}^{m \times m}$, $F = n^{-1}N^T X \in \mathbb{R}^{m \times 3}$, $E = n^{-1}X^T X \in \mathbb{R}^{3 \times 3}$ である。問題(6),(7)は凸計画問題であるので、 $(u_0^*, u_1^*, u_2^*, c^*)$ が問題(6),(7)の解であるための必要十分条件は $w \in \mathbb{R}^m$ が存在して、 $(u_0^*, u_1^*, u_2^*, c^*, w^*)$ が(9)の解であることである。

解の存在性

Neumann 境界値問題では、解に少なくとも付加定数項分の不定性がある。解が一意に存在するために正規化条件

$$u_0(x_1^0, x_2^0) = 0 \quad ((x_1^0, x_2^0) \in \Omega \text{は固定点}) \quad (10)$$

を付け加える。この条件より行列 A は対称正定値行列と考えることができる。(7)より u_0 について

$$u_0 = A^{-1}B_1 u_1 + A^{-1}B_2 u_2 \quad (11)$$

である。また、

$$\begin{aligned} & \frac{1}{n} \|Nu_0 + Xc - y\|_2^2 \\ &= \frac{1}{n} ((Nu_0 + Xc - y)^T, Nu_0 + Xc - y) \\ &= \frac{1}{n} \{ (Nu_0)^T Nu_0 + (Nu_0)^T Xc - (Nu_0)^T y \\ & \quad + (Xc)^T Nu_0 + (Xc)^T Xc - (Xc)^T y - y^T Nu_0 - y^T Xc + y^T y \} \\ &= \frac{1}{n} (u_0^T N^T Nu_0 + 2c^T X^T Nu_0 + c^T X^T Xc - 2y^T Nu_0 - 2y^T Xc + y^T y) \end{aligned}$$

ここで $Q_1 = NA^{-1}B_1$, $Q_2 = NA^{-1}B_2$ とし、(11)を代入すると、

$$\begin{aligned} & \frac{1}{n} \|Nu_0 + Xc - y\|_2^2 \\ &= \frac{1}{n} (u_1^T Q_1^T Q_1 u_1 + 2u_1^T Q_1^T Q_2 u_2 + u_2^T Q_2^T Q_2 u_2 \\ & \quad + 2c^T X^T Q_1 u_1 + 2c^T X^T Q_2 u_2 + c^T X^T Xc \\ & \quad - 2y^T Q_1 u_1 - 2y^T Q_2 u_2 - 2y^T Xc + y^T y) \end{aligned}$$

となる。よって、

$$\begin{aligned} & \frac{1}{n} \|Nu_0 + Xc - y\|_2^2 + \alpha(u_1^T Au_1 + u_2^T Au_2) \\ &= \frac{1}{2} z^T H z + b^T z + \frac{1}{n} y^T y \end{aligned}$$

ここで

$$\begin{aligned} H &= \frac{2}{n} \begin{bmatrix} Q_1^T Q_1 + n\alpha A & Q_1^T Q_2 & Q_1^T X \\ Q_2^T Q_1 & Q_2^T Q_2 + n\alpha A & Q_2^T X \\ X^T Q_1 & X^T Q_2 & X^T X \end{bmatrix} \\ b &= -\frac{2}{n} \begin{bmatrix} y^T Q_1 \\ y^T Q_2 \\ y^T X \end{bmatrix} \quad z = \begin{bmatrix} u_1 \\ u_2 \\ c \end{bmatrix} \end{aligned}$$

と書ける。 H が正定値行列であることを示す。 $Q = (Q_1, Q_2, X)$ とする。

$$H = \frac{2}{n} Q^T Q + \begin{bmatrix} 2\alpha A & 0 & 0 \\ 0 & 2\alpha A & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$t = (t_1, t_2, t_3)^T \in R^{2m+3}$ とする。 $[t_1, t_2]^T \neq 0$ に対して

$$t^T H t = \frac{2}{n} t^T Q^T Q t + 2\alpha t_1^T A t_1 + 2\alpha t_2^T A t_2 > 0$$

(A は正定値行列, $t^T Q^T Q t = \|Qt\|^2 \geq 0$ より)

が成り立つ。また, $[t_1, t_2]^T = 0, t_3 \neq 0$ に対して, n 個の測定点は異なり, 同一直線上にないという仮定から

$$t^T H t = \frac{2}{n} t_3^T X^T X t_3 > 0$$

である。よって H は正定値行列である。問題 (6),(7) は強凸計画問題であるので, 唯一の解をもつ。

1.2 連立一次方程式の解法

今, f を (2) で近似することを考えている。よって, 特に (9) において u_0, c について解きたい。(9) を方程式で再び書き直すと

$$\alpha A u_1 - B_1^T w = 0 \quad (12)$$

$$\alpha A u_2 - B_2^T w = 0 \quad (13)$$

$$F^T u_0 + E c = n^{-1} X^T y \quad (14)$$

$$M u_0 + F c + A w = n^{-1} N^T y \quad (15)$$

$$-B_1 u_1 - B_2 u_2 + A u_0 = 0 \quad (16)$$

である。 u_1, u_2 について (12),(13) よりすぐに

$$u_1 = \alpha^{-1} A^{-1} B_1^T w \quad (17)$$

$$u_2 = \alpha^{-1} A^{-1} B_2^T w \quad (18)$$

とわかる。(16) に (17),(18) を代入して

$$-\alpha^{-1} (B_1 A^{-1} B_1^T + B_2 A^{-1} B_2^T) w + A u_0 = 0 \quad (19)$$

で, $G = B_1 A^{-1} B_1^T + B_2 A^{-1} B_2^T$ とおくと

$$w = \alpha G^{-1} A u_0 \quad (20)$$

を得る。ここで G^{-1} の存在性が問題となるがそれは後に示す。(14) より

$$c = E^{-1} (n^{-1} X^T y - F^T u_0) \quad (21)$$

(15) に (20),(21) を代入して

$$(M + \alpha A G^{-1} A - F E^{-1} F^T) u_0 = n^{-1} (N^T y - F E^{-1} X^T y) \quad (22)$$

$$(N^T (I - X E^{-1} X^T) N + \alpha A G^{-1} A) u_0 = n^{-1} (N^T y - F E^{-1} X^T y) \quad (23)$$

(23) より, u_0 について解くことができる。よって, (21) より c についても解ける。

(23) の左辺の行列において, 最初の部分は半正定値行列である。 G が正定値行列であることは次で示すが, 2 番目の部分は正定値行列であるので, 全体で正定値行列となり, 共役勾配法が適用できる。

G の逆行列の存在性, および正定値行列であることの証明

$[B_1, B_2]^T$ の階数は full column rank とする。 $x^T G x = 0$ となるベクトル $x \in \mathbb{R}^m$ について考えると,

$$\begin{aligned} x^T G x &= x^T (B_1 A^{-1} B_1^T + B_2 A^{-1} B_2^T) x = 0 \\ &\Rightarrow 0 \leq x^T B_1 A^{-1} B_1^T x = -x^T B_2 A^{-1} B_2^T x \leq 0 \\ &\Rightarrow x^T B_1 A^{-1} B_1^T x = x^T B_2 A^{-1} B_2^T x = 0 \end{aligned}$$

B_1^T における零空間を N_1, B_2^T における零空間を N_2 とすると $B_1^T x = 0, B_2^T x = 0$ となる $x \in N_1 \cap N_2$ は $[B_1, B_2]^T$ の階数が full column rank という仮定より, $x = 0$ のみである。よって G に逆行列が存在する。

上記のことより $x \neq 0$ に対して $x^T G x > 0$ が成り立つ。よって, G は正定値行列である。

1.3 数値計算

MATLAB の組み込み関数 `peaks` に対して評価を行った。`peaks` はガウス分布の変換とスケールリングによって得られる 2 変数関数である。その図を図 1 に示す。問題の領域は $\Omega = (-3, 3) \times (-3, 3)$ であるが, 領域を $\Omega = (0, 1) \times (0, 1)$ に正規化している。測定点はランダムに発生させた。図 3 はその測定点の配置図である。データ数と基底関数の数はそれぞれ $n = 4900, m = 2116, \alpha = 10^{-6}$ で計算した。メッシュは一様メッシュとした。図 2 は薄板スプライン有限要素法による曲面回帰の図である。絶対誤差の平均は 0.0866 であった。

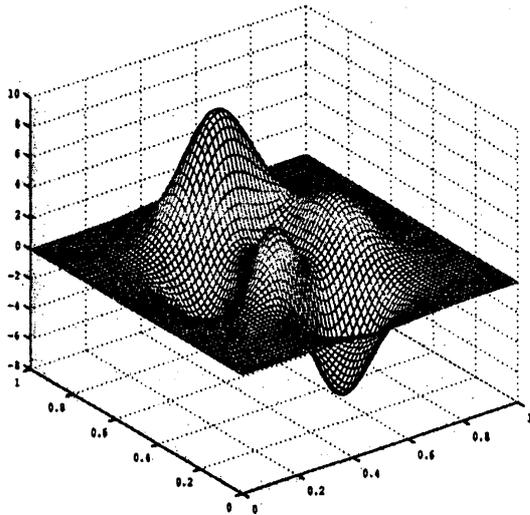


図 1: 厳密解

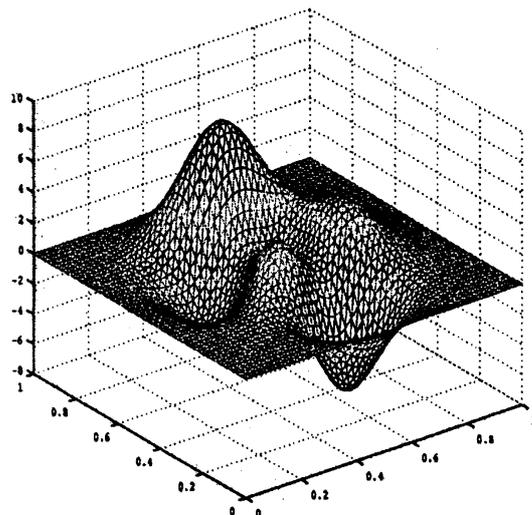


図 2: 薄板スプライン有限要素法

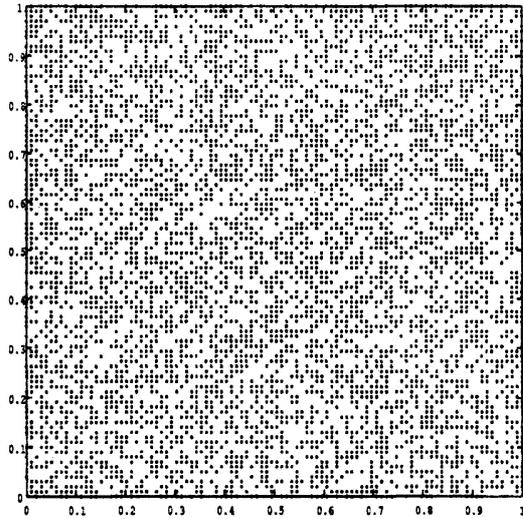


図 3: 測定点

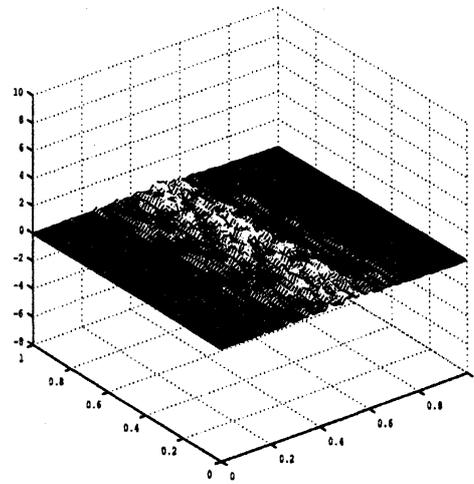


図 4: 絶対誤差

2 交通量推定への応用

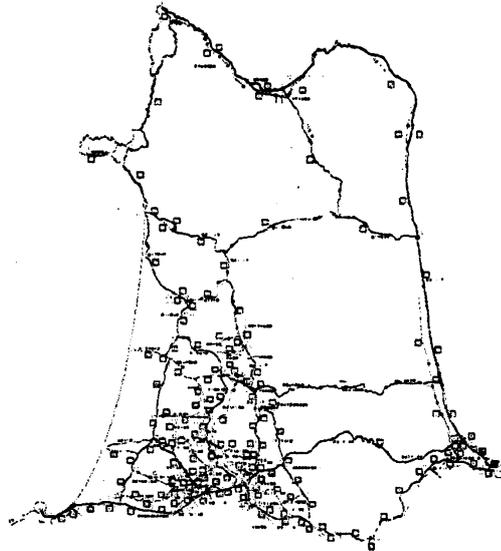


図 5: 青森県津軽半島

2.1 曲面回帰による交通量の推定

図 5 は青森県津軽半島の地図である。この地図上の交差点と橋の座標、および平日の 24 時間交通量平均がわかっている。ただし、交通量の調査地点の座標は得られなかったため、調査地点にもっとも近い交差点、もしくは橋の座標に交通量を当ててデータとした。この既知のデータ 79 個の測定点と交通量を表 1 に示す。交通量については青森県土木部「道路交通センサス」より、交差点、橋

の座標は株式会社キタコンのデータを利用した。

表 1: 交通量と測定点

通し 番号	交通量	X(m,North)	Y(m,East)	通し 番号	交通量	X(m,North)	Y(m,East)
1	25140	84464.082	-20380.223	41	2321	121101.395	-23560.232
2	23573	91894.279	-14849.68	42	5042	94291.319	-12715.908
3	43392	91695.382	-8836.742	43	2049	91466.591	-31696.921
4	16250	84401.183	-20799.319	44	673	92787.367	-29201.133
5	18593	86550.553	-30304.321	45	2981	92161.375	-29156.233
6	12469	88489.52	-32680.686	46	3758	89201.016	-36636.756
7	12483	86952.548	-39935.522	47	3897	89949.006	-33928.984
8	11170	86990.356	-40341.808	48	3096	101670.243	-35328.77
9	4270	86601.653	-51957.699	49	16082	91596.239	-8346.76
10	5419	87415.141	-50334.015	50	520	104657.501	-32932.394
11	10209	92381.785	-9139.159	51	820	125168.516	-48611.733
12	5959	94906.737	-11552.114	52	6427	92371.104	-39373.921
13	5009	101031.852	-13595.793	53	9595	90019.15	-36831.459
14	819	116284.139	-15936.056	54	5095	90364.7	-38120.741
15	854	109502.334	-38463.237	55	3358	90330.201	-33052.793
16	2956	130737.638	-28703.638	56	11133	86609.553	-28346.541
17	16868	92725.808	-12024.972	57	6649	89475.157	-32055.063
18	8164	97656.867	-13579.102	58	3423	87760.29	-32063.081
19	1873	123715.871	-16936.056	59	3129	97284.804	-37558.047
20	2146	131486.318	-31196.157	60	879	91715.086	-36264.515
21	2956	132155.019	-32504.599	61	4873	90103.504	-32975.794
22	12266	88610.025	-32697.497	62	1508	95258.032	-39571.426
23	7170	92933.793	-31776.557	63	2600	104091.019	-34670.011
24	3573	106689.368	-33724.23	64	2317	97249.204	-36974.353
25	5480	111697.003	-34720.376	65	883	93359.459	-43521.185
26	3740	116869.531	-39813.723	66	1134	88403.228	-38126.341
27	3613	116939.006	-39996.536	67	2581	89912.707	-39917.122
28	4439	120065.516	-42693.125	68	1281	88180.631	-42261.698
29	443	125256.812	-44028.481	69	4439	109565.713	-38205.33
30	917	134231.922	-34282.489	70	1760	98655.115	-29668.394
31	6719	100363.86	-32364.674	71	1708	101697.459	-34927.936
32	1331	102558.639	-39029.116	72	820	103834.013	-40699.614
33	1100	100174.463	-30024.374	73	4123	92810.566	-13021.499
34	3817	101081.651	-33821.685	74	5417	89585.875	-36714.885
35	4838	88946.38	-45163.759	75	5417	88951.52	-38483.337
36	1802	110498.557	-40199.531	76	4216	86513.254	-18265.545
37	2577	117680.476	-21079.621	77	16851	91828.002	-9539.959
38	1975	103470.359	-40400.472	78	847	131797.416	-32982.641
39	1087	114592.263	-34904.274	79	11041	132452.622	-33659.065
40	2370	118331.211	-23497.591				

これらのデータ点から薄板スプライン有限要素法による曲面回帰によって青森県津軽半島全体の交通量を推定することを試みる。

問題の領域 Ω はデータ点を作る多角形領域、つまりデータ番号 1, 49, 19, 30, 51, 9, 4 を結んだ多角形領域とするが、 $(0, 1) \times (0, 1)$ におさまるような多角形領域に正規化して考える。

2.1.1 メッシュについて

図6はデータ点同士を結んで作ったメッシュである。しかし、極端な鋭角や鈍角の三角形要素が多く、適したメッシュとは言えない。そこで、次のように改良した。領域を覆うような格子点を追加し、領域からはみ出す部分と領域の点でも格子点を打つ間隔より境界に近い距離の点は削除する。それだけでは、境界の部分での潰れた三角形要素は避けられないので、境界にも節点を追加する。極力一様メッシュのような二等辺三角形の要素に分割したいという目的であるが、境界に節点を打つ間隔はその二等辺三角形の斜辺とした。図7は改良後のメッシュである。実際の計算ではもっと細かいメッシュを用いる。

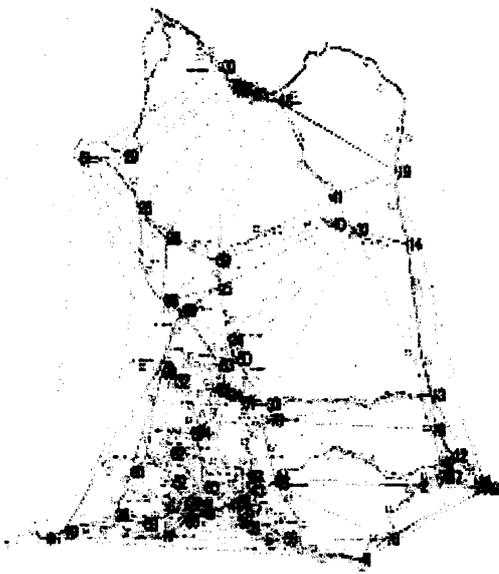


図6: データ点のみを結んだメッシュ
(数字はデータの通し番号)

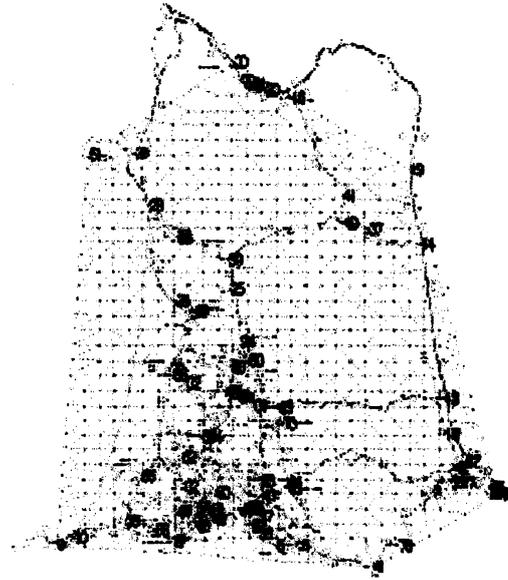


図7: 改良されたメッシュ
($m = 839$)

2.2 誤差評価と考察

まず、データ点79個すべてを使って計算を行う。今回は交通量であるので、負の値があるときは0とする。

表2は $n = 79, m = 4834, \alpha = 10^{-11}$ のときの各データ点に対する推定値と絶対誤差, 相対誤差である。

表 2: 誤差評価 ($n = 79, m = 4834, \alpha = 10^{-11}$ のとき)

通し 番号	推定値	絶対 誤差	相対 誤差	通し 番号	推定値	絶対 誤差	相対 誤差
1	25140	0.23204	9.23×10^{-6}	41	2321	0.0088855	3.83×10^{-6}
2	23573	0.18243	7.74×10^{-6}	42	5042	0.043434	8.61×10^{-6}
3	43391	0.61329	1.41×10^{-5}	43	2049	0.049591	2.42×10^{-5}
4	16250	0.14642	9.01×10^{-6}	44	673.14	0.13788	0.00020488
5	18593	0.19718	1.06×10^{-5}	45	2980.9	0.12245	4.11×10^{-5}
6	12469	0.3904	3.13×10^{-5}	46	3758.2	0.17764	4.73×10^{-5}
7	12483	0.08202	6.57×10^{-6}	47	3897.1	0.078552	2.02×10^{-5}
8	11170	0.032078	2.87×10^{-6}	48	3095.8	0.19847	6.41×10^{-5}
9	4270	0.0015713	3.68×10^{-7}	49	16082	0.01269	7.89×10^{-7}
10	5419	0.0031103	5.74×10^{-7}	50	520.07	0.073476	0.0001413
11	10210	1.1248	0.00011017	51	820	0.0010913	1.33×10^{-6}
12	5958.9	0.11414	1.92×10^{-5}	52	6426.9	0.079166	1.23×10^{-5}
13	5009	0.0062085	1.24×10^{-5}	53	9594.2	0.84051	8.76×10^{-5}
14	819	0.0035279	4.31×10^{-5}	54	5095.3	0.29454	5.78×10^{-5}
15	854.6	0.60066	0.00070335	55	3358.2	0.17012	5.07×10^{-5}
16	2956	0.0061235	2.07×10^{-5}	56	11133	0.07845	7.05×10^{-5}
17	16868	0.14926	8.85×10^{-5}	57	6649.1	0.059739	8.98×10^{-5}
18	8164	0.019298	2.36×10^{-5}	58	3423.5	0.53961	0.00015764
19	1873	0.0019495	1.04×10^{-5}	59	3129	0.045541	1.46×10^{-5}
20	2146	0.014678	6.84×10^{-5}	60	879.17	0.17061	0.0001941
21	2956	0.011608	3.93×10^{-5}	61	4872.9	0.090998	1.87×10^{-5}
22	12266	0.23427	1.91×10^{-5}	62	1508	0.020435	1.36×10^{-5}
23	7169.9	0.085383	1.19×10^{-5}	63	2600	0.025916	9.97×10^{-5}
24	3573	0.0025392	7.11×10^{-7}	64	2317	0.036342	1.57×10^{-5}
25	5480	0.0079734	1.46×10^{-5}	65	883	0.00045747	5.18×10^{-7}
26	3739.9	0.064864	1.73×10^{-5}	66	1134.4	0.38255	0.00033735
27	3613.1	0.064481	1.78×10^{-5}	67	2581.1	0.092637	3.59×10^{-5}
28	4439	0.013612	3.07×10^{-5}	68	1281	0.043049	3.36×10^{-5}
29	443	0.0044537	1.01×10^{-5}	69	4438.4	0.58023	0.00013071
30	917.17	0.16642	0.00018148	70	1760	0.0016334	9.28×10^{-7}
31	6719	0.044648	6.65×10^{-5}	71	1708.2	0.21748	0.00012733
32	1331	0.03376	2.54×10^{-5}	72	820.07	0.069922	8.53×10^{-5}
33	1100	0.02612	2.37×10^{-5}	73	4123.5	0.47897	0.00011617
34	3817	0.010174	2.67×10^{-5}	74	5417.3	0.34515	6.37×10^{-5}
35	4838	0.018353	3.79×10^{-5}	75	5416.5	0.49779	9.19×10^{-5}
36	1802	0.039794	2.21×10^{-5}	76	4216.1	0.095324	2.26×10^{-5}
37	2577	0.0040524	1.57×10^{-5}	77	16850	0.50819	3.02×10^{-5}
38	1974.9	0.10264	5.20×10^{-5}	78	847.28	0.27565	0.00032544
39	1087	0.0065792	6.05×10^{-5}	79	11041	0.42109	3.81×10^{-5}
40	2370	0.0042464	1.79×10^{-5}				

データ点における誤差は少ないようである。今回、測定点を節点として使っているが、節点における値は基底関数の性質より、ひとつの基底関数で決定される。区分的線形基底関数のひとつの側面(平面)で複数のデータを回帰するより、近似が容易であるためである。

表 3,4,5 には 79 個のデータを使用したある m に対するそれぞれ α を変更したときの絶対誤差, 相対誤差の最大値, 最小値, 平均を示す。

表 3: 誤差評価 ($n = 79, m = 1267$ のとき)

	絶対誤差			相対誤差		
	最大値	最小値	平均	最大値	最小値	平均
$\alpha = 10^{-6}$	11261	9.0710	1735.3	4.7264	0.0018	0.5983
$\alpha = 10^{-8}$	1439.6	0.1243	148.9607	0.5052	2.9110×10^{-5}	0.0448
$\alpha = 10^{-11}$	2.5740	5.3948×10^{-4}	0.2059	7.6119×10^{-4}	3.0571×10^{-7}	5.9962×10^{-5}

表 4: 誤差評価 ($n = 79, m = 2749$ のとき)

	絶対誤差			相対誤差		
	最大値	最小値	平均	最大値	最小値	平均
$\alpha = 10^{-6}$	11928	11.1120	1672.1	4.5399	0.0023	0.5391
$\alpha = 10^{-8}$	738.4438	0.4497	113.5197	0.4245	2.4011×10^{-4}	0.0376
$\alpha = 10^{-11}$	1.0607	0.0010	0.1492	4.9267×10^{-4}	3.4022×10^{-7}	4.3824×10^{-5}

表 5: 誤差評価 ($n = 79, m = 4834$ のとき)

	絶対誤差			相対誤差		
	最大値	最小値	平均	最大値	最小値	平均
$\alpha = 10^{-6}$	10545	7.0079	1677.1	3.6429	0.0014	0.5088
$\alpha = 10^{-8}$	787.4305	0.1901	126.7223	0.5285	7.3750×10^{-5}	0.0404
$\alpha = 10^{-11}$	1.1248	4.5747×10^{-4}	0.1549	7.0335×10^{-4}	3.6799×10^{-7}	5.0763×10^{-5}

データ点の部分ではなかなかよい近似であると思われる。今度はデータ点ではないところの誤差について調べたい。そのために、79個のデータのうち約半分の39個を取り除いて誤差評価をする。ある一定の範囲に集中して取り除いた場合、推定が困難になるため、ランダムに削除するリストを生成して行う。ここでは、データ番号 3,5,8,10,11,12,13,17,18,20,22,24,25, 26, 27, 32,34,36,37,38,39,40,41,42,45,47,54,56, 57,59,61,62,63,64,69,70,74,75,77 を削除し、評価している。表 6 にその結果を示す。

表 6: 誤差評価 ($n = 40, m = 4834, \alpha = 10^{-11}$ のとき)

通し番号	推定値	絶対誤差	相対誤差	通し番号	推定値	絶対誤差	相対誤差
1	25140	0.1413	5.62×10^{-6}	41	2747.4	426.4	0.18371
2	23573	0.054069	2.29×10^{-6}	42	670.85	4371.1	0.86695
3	11541	31851	0.73404	43	2049	0.024349	1.19×10^{-6}
4	16250	0.10664	6.56×10^{-6}	44	673.01	0.0086573	1.29×10^{-5}
5	0	18593	1	45	1510.9	1470.1	0.49316
6	12469	0.18047	1.45×10^{-6}	46	3758.1	0.14822	3.94×10^{-6}
7	12483	0.042193	3.38×10^{-6}	47	7852.5	3955.5	1.015
8	14544	3374.3	0.30208	48	3095.9	0.10205	3.30×10^{-6}
9	4270	0.0013078	3.06×10^{-7}	49	16082	0.010504	6.53×10^{-7}
10	7258.2	1839.2	0.3394	50	520.02	0.015648	3.01×10^{-5}
11	14413	4204.2	0.41181	51	820	0.00016589	2.02×10^{-7}
12	1810.9	4148.1	0.6961	52	6427	0.0047501	7.39×10^{-7}
13	6814.6	1805.6	0.36046	53	9594.8	0.24832	2.59×10^{-5}
14	819	0.00058633	7.16×10^{-7}	54	8123.8	3028.8	0.59446
15	854	0.0034283	4.01×10^{-6}	55	3358.1	0.062182	1.85×10^{-5}
16	2956	0.0043173	1.46×10^{-6}	56	0	11133	1
17	2989.2	13879	0.82279	57	7320.5	671.55	0.101
18	5055.7	3108.3	0.38073	58	3423.1	0.11201	3.27×10^{-6}

19	1873	0.00093298	4.98×10^{-7}	59	2438.7	690.29	0.22061
20	213.55	1932.5	0.90049	60	879.07	0.074986	8.53×10^{-5}
21	2956	0.028895	9.77×10^{-6}	61	3865.6	1007.4	0.20673
22	12370	104.14	0.0084901	62	2614.7	1106.7	0.73391
23	7170	0.033647	4.69×10^{-6}	63	0	2600	1
24	361.1	3211.9	0.89894	64	2467.4	150.4	0.064911
25	882.44	4597.6	0.83897	65	883	0.0027562	3.12×10^{-6}
26	2324	1416	0.37862	66	1134.1	0.051436	4.54×10^{-5}
27	2128.2	1484.8	0.41096	67	2581	0.013672	5.30×10^{-6}
28	4439	0.0041646	9.38×10^{-7}	68	1281	0.024863	1.94×10^{-5}
29	443	0.00048767	1.10×10^{-6}	69	931.26	3507.7	0.79021
30	917.08	0.081424	8.88×10^{-5}	70	2189.1	429.06	0.24379
31	6719	0.022657	3.37×10^{-6}	71	1708.1	0.10933	6.40×10^{-6}
32	1100.1	230.93	0.1735	72	820	0.00138	1.68×10^{-6}
33	1100	0.012705	1.16×10^{-5}	73	4123.1	0.058624	1.42×10^{-5}
34	3251.7	565.26	0.14809	74	7642.2	2225.2	0.41078
35	4838	0.0024736	5.11×10^{-7}	75	3121.7	2295.3	0.42371
36	232.06	1569.9	0.87122	76	4216	0.032155	7.63×10^{-6}
37	2644.7	67.701	0.026271	77	17086	234.56	0.01392
38	1329.7	645.32	0.32674	78	847.15	0.14589	0.00017224
39	808.84	278.16	0.2559	79	11041	0.19765	1.79×10^{-5}
40	3075.2	705.24	0.29757				

すべてのデータを使用したとき同様、データ点を使用したところでの近似はよいが、そうでないところには、誤差の大きな点が存在する。誤差が大きくなる理由として、とても近くで急速に交通量が増加しているデータ点を削除した場合、その誤差は大きくなるようである。

表 7,8,9 には 40 個のデータを使用したある m に対するそれぞれ α を変更したときの絶対誤差、相対誤差の最大値、最小値、平均を示す。

表 7: 誤差評価 ($n = 40, m = 1267$ のとき)

	絶対誤差			相対誤差		
	最大値	最小値	平均	最大値	最小値	平均
$\alpha = 10^{-6}$	28990	3.1167	2408.6	3.4568	0.0015	0.5447
$\alpha = 10^{-8}$	28463	1.1749	1942.6	2.4635	4.7262×10^{-4}	0.3221
$\alpha = 10^{-11}$	33248	0.0017	2046.1	2.8731	4.0557×10^{-7}	0.3252

表 8: 誤差評価 ($n = 40, m = 2749$ のとき)

	絶対誤差			相対誤差		
	最大値	最小値	平均	最大値	最小値	平均
$\alpha = 10^{-6}$	32175	0.5173	2306.7	3.3470	6.3083×10^{-4}	0.4929
$\alpha = 10^{-8}$	34248	0.0949	2050.4	2.7624	1.1572×10^{-4}	0.3511
$\alpha = 10^{-11}$	31676	5.4877×10^{-4}	1838.7	1.6767	2.9299×10^{-7}	0.2933

表 9: 誤差評価 ($n = 40, m = 4834$ のとき)

	絶対誤差			相対誤差		
	最大値	最小値	平均	最大値	最小値	平均
$\alpha = 10^{-6}$	28260	12.3970	2704.0	3.8538	0.0238	0.6108
$\alpha = 10^{-8}$	27934	0.1412	2289.3	3.7821	1.7242×10^{-4}	0.4441
$\alpha = 10^{-11}$	31851	1.6589×10^{-4}	1758.4	1.0150	2.0231×10^{-7}	0.2398

図 10 には $n = 79, m = 4834, \alpha = 10^{-11}$ のときの等高線図を, 図 8,9 にはそのときの曲面を示す。また, 図 13 には $n = 40, m = 4834, \alpha = 10^{-11}$ のときの等高線図を, 図 11,12 にはそのときの曲面を示す。

今回, $\alpha = 10^{-11}$ のとき一番良い結果が得られたようであるが, 値をいろいろ変更してみてもよし悪しである。最適な α は論理的な計算で決定するべきである。

等高線図を見れば, 道路に沿って交通量があるような傾向が極わずかだけ見られるが, 基本的には道路のない地点における交通量は無視する考えである。道路の形状を意識した曲面を生成するにはどのようにすべきか, 今後の課題となる。

計算環境

OS Windows XP
 CPU Pentium 4 (3.0GHz)
 メモリ 1GB
 MATLAB 7.0

ただし, $m = 4834$ のときのそれぞれ n, α に対する計算は

OS Red Hat Linux
 CPU XEON (3.40GHz)
 メモリ 4GB
 MATLAB 7.0

を使用した。

謝辞

最後にデータを提供して下さった津村浩三助教授, 弘前大学理工学研究科安全システム工学専攻の藤田弘昭氏に深く感謝申し上げます。

参考文献

- [1] P.Christen, M.Hegland, S.Roberts and I.Altas, A scalable parallel FEM surface fitting algorithm for data mining, Technical Report TR-CS-01-01, The Australian National University, 2001.
- [2] M.Hegland, S.Roberts and I.Altas, Finite element thin plate splines for surface fitting, in Computational Techniques and Application: CTAC97 (Singapore, 1997), B.Noye, M.Teubner and A.Gill, eds. World Scientific, River Edge, NJ, 289-296.
- [3] S.Roberts, M.Hegland and I.Altas, Approximation of a thin plate spline smoother using continuous piecewise polynomial functions, SIAM J. Numer. Anal. 41(2003), 208-234.

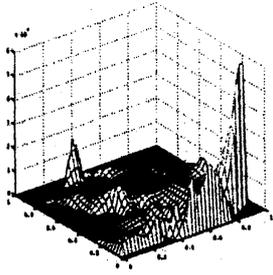


図 8: 図 10 を左下から見た曲面図

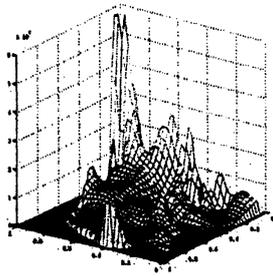


図 9: 図 10 を左上から見た曲面図

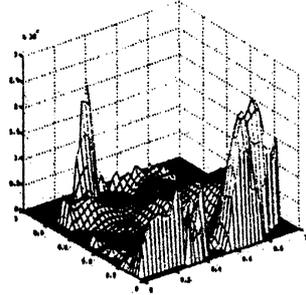


図 11: 図 13 を左下から見た曲面図

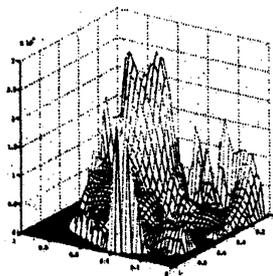


図 12: 図 13 を左上から見た曲面図

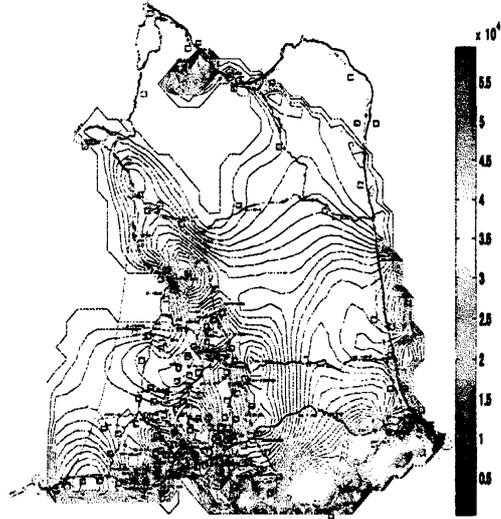


図 10: 等高線図
 $n = 79, m = 4834, \alpha = 10^{-11}$ のとき

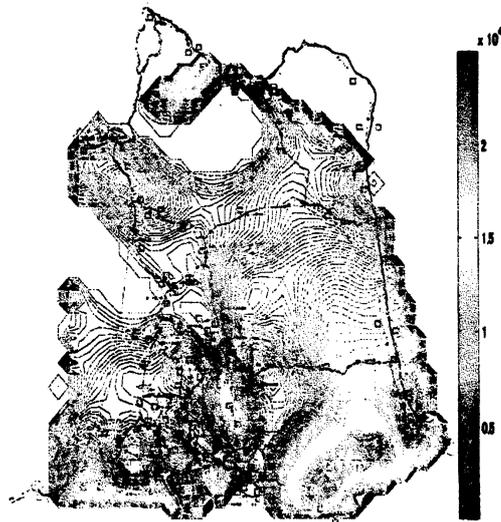


図 13: 等高線図
 $n = 40, m = 4834, \alpha = 10^{-11}$ のとき