Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(7):302-309

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Scattered data fitting using least squares with interpolation method

Tianhe Zhou

Department of Mathematics, Zhejiang Sci-Tech University, Hangzhou, Zhejiang, China

---

ABSTRACT

Scattered data fitting is a big issue in numerical analysis. In many applications, some of the data are contaminated by noise and some are not. It is not appropriate to interpolate the noisy data, and the traditional least squares method may lose accuracy at the points which are not contaminated. In this paper, we present least squares with interpolation method to solve this problem. The existence and uniqueness of its solution are proved and an error bound is derived. Some numerical examples are also presented to demonstrate the effectiveness of our method.

Key words: Bivariate splines, least squares method, minimal energy method, least squares with interpolation.

INTRODUCTION

Scattered data fitting is a big issue in numerical analysis. There are lots of literatures about this topic (e.g.,[6, 8, 10, 11, 16]). If the scattered data are contaminated by noise, we can use least squares method (e.g., [13, 9,15]) or other methods (e.g., [1, 2, 3, 12, 17]). If the scattered data are not contaminated, we can use interpolation methods (e.g., [4, 5, 7, 13, 14]). But in some circumstances, some of the data are contaminated, and others are not. It is not appropriate to interpolate the noisy data, and traditional least squares method may lose accuracy at the points which are not contaminated. To overcome this difficulty, we propose a new approach contained with two methods called least squares with interpolation.

Let $V=\{v_i=(x_i,y_i)\}_{i=1}^{M+N}$ be a set of points lying in a domain $\Omega\subset\mathbf{R}^2$ with polygonal boundary and $\{f_i, i=1,\cdots,M+N\}$ be a corresponding set of real numbers. The data are assumed to be contaminated by noise

$$f_i=f(x_i,y_i)+\epsilon_i$$

where f belongs to the standard Sobolev space $W_\infty^2(\Omega)$ and ϵ_i are noisy terms. In some circumstances, the data can be divided into two groups by the size of noise. Suppose $D^0=\{x_i^0,y_i^0,f_i^0\}_{i=1}^{M}$ be a group of data with $\epsilon_i=0$ and $D^1=\{x_i^1,y_i^1,f_i^1\}_{i=1}^{N}$ be another group of data with $\epsilon_i\neq 0$. We first construct a surface which minimize

$$L(s_L):=\sum_{i=1}^{M+N}|s_L(v_i)-f_i|^2=\sum_{i=1}^{M}|s_L(x_i^0,y_i^0)-f_i^0|^2+\sum_{i=1}^{N}|s_L(x_i^1,y_i^1)-f_i^1|^2 \qquad (1.1)$$

where $S_d^r(\Delta_L)$ is a space of splines of degree d and smoothness r defined in next section. Next we modify some of the data in two groups. Let $\bar{D}^0=\{x_i^0,y_i^0,z_i^0\}_{i=1}^{M}$ and $\bar{D}^1=\{x_i^1,y_i^1,z_i^1\}_{i=1}^{N}$ be two new groups where

$z_i^0 := f_i^0 - s_L(x_i^0, y_i^0)$ and $z_i^1 := 0$. Define a set of all spline in $S_d^r(\Delta_I)$ that interpolate the data at the points of V by

$$I_f := \{s \in S_d^r(\Delta_I) : s(x_i^0, y_i^0) = z_i^0, s(x_j^1, y_j^1) = z_j^1, i = 1, \ldots, M, j = 1, \ldots, N\}. \quad (1.2)$$

where Δ_I is a new triangulation of data locations. After that, we construct a spline function $s_I \in S_d^r(\Delta_I)$ such that

$$E(s_I) = \min\{E(s) : s \in I_f\} \quad (1.3)$$

where $E(s)$ is a energy function defined by

$$E(s) := \sum_{T \in \Delta_I} \int_T [s_{xx}^2 + 2s_{xy}^2 + s_{yy}^2].$$

Finally, we get a new spline function by

$$s_F = s_L + s_I. \quad (1.4)$$

From the above, we can see that it need to construct two splines with two triangulations in our approach. One is s_L for least squares fit, another one is s_I for data interpolation. We construct the spline s_L base on triangulation Δ_L and the s_I base on Δ_I.

In this paper, we derive the following error bound for this approach,

$$\| f - s_F \|_{\infty,\Omega} \le (2C_1 |\Delta_L|^2 + C_2 |\Delta_I|^2) | f |_{2,\infty,\Omega}, \quad (1.5)$$

where $| f |_{2,\infty,\Omega}$ denotes the maximum norm of the second order derivatives of f over Ω and $\| f \|_{\infty,\Omega}$ is the standard infinity norm of f over Ω. Here $|\Delta_L|$ and $|\Delta_I|$ are the maximum of the diameters of the triangles in Δ_L and Δ_I and all the constants in the error bound will be introduced in following sections.

The paper is organized as follows. In Section 2 we review some well-known Bernstein-Bezier notation. The result on error bounds of this approach is derived in Section 3 together with a discussion on existence and uniqueness. In Section 4 we present some numerical examples for our method.

PRELIMINARIES

Given a triangulation Δ and integers $0 \le r < d$, we write

$$S_d^r(\Delta) := \{s \in C^r(\Omega) : s|_T \in P_d, \text{for all } T \in \Delta\}$$

for the usual space of splines of degree d and smoothness r, where P_d is the $\binom{d+2}{2}$ dimensional space of bivariate polynomials of degree d. Throughout the paper we shall make extensive use of the well-known Bernstein-Bezier representation of splines. For each triangle $T = \langle v_1, v_2, v_3 \rangle$ in Δ with vertices $v_1, v_2, v_3,$ the corresponding polynomial piece $s|_T$ is written in the form

$$s|_T = \sum_{i+j+k=d} c_{ijk}^T B_{ijk}^T,$$

where B_{ijk}^T are the Bernstein-Bezier polynomials of degree d associated with T. In particular, if $(\lambda_1, \lambda_2, \lambda_3)$

are the barycentric coordinates of any point $u \in \mathbf{R}^2$ relative to the triangle T, then

$$B_{ijk}^{T}(u) := \frac{d!}{i!\,j!\,k!}\lambda_1^i \lambda_2^j \lambda_3^k, i+j+k=d.$$

As usual, we associate the Bernstein-Bezier coefficients $\{c_{ijk}^T\}_{i+j+k=d}$ with the domain points $\{\xi_{ijk}^T := (iv_1 + jv_2 + kv_3)/d\}_{i+j+k=d}$.

Definition 1. Let $\beta < \infty$. A triangulation $\triangle$ is said to be β-quasi-uniform provided that $|\triangle| \le \beta \rho_{\triangle}$, where $\rho_{\triangle}$ is the minimum of the radii of the incircles of triangles of $\triangle$.

We say the data locations are evenly distributed over each triangle of $\triangle$, if there are enough data points for every triangle such that

$$\| P \|_{\infty,T} \le C\Big(\sum_{v_i \in T} P(v_i)^2\Big)^{1/2}$$

for any polynomial P and some constant C.

ERROR ANALYSIS

In this section, we mainly discuss the error bounds for the spline function created by this approach. First, we show the existence and uniqueness of this spline. Note that s_L is a solution of least squares method and s_I is a solution of minimal energy interpolation.

Theorem 3.1 (cf. [15]) Suppose all the data locations are evenly distributed over each triangle of $\triangle_L$. Then there exists a unique $s_L \in S_d^r(\triangle_L)$ satisfying (1.1).

Theorem 3.2 (cf. [7]) There exists a unique solution $s_I \in S_d^r(\triangle_I)$ which satisfies (1.3).

Note that $s_F = s_L + s_I$, we can get

Theorem 3.3 There exists a unique solution s_F of this approach.

Next we show the error bound for the spline function. Recall from [10], we can get

$$\| f - s_L \|_{\infty,\Omega} \le C_1 |\triangle_L|^2 | f |_{2,\infty,\Omega} . \tag{3.1}$$

Theorem 3.4 Suppose that two triangulations $\triangle_L$ and $\triangle_I$ are β-quasi-uniform and $f \in W_\infty^2(\Omega)$. Then there exists two constants C_1 and C_2 depending on d, β, C and the smallest angel in each triangulation such that the spline s_F defined in (1.4) satisfies
$\| f - s_F \|_{\infty,\Omega} \le (2C_1 |\triangle_L^2 + C_2 |\triangle_I|^2) | f |_{2,\infty,\Omega}$.

Proof: Since $s_F = s_L + s_I$, we can get

$$\| f - s_F \|_{\infty,\Omega} = \| f - s_L - s_I \|_{\infty,\Omega} \le \| f - s_L \|_{\infty,\Omega} + \| s_I \|_{\infty,\Omega}. \qquad (3.2)$$

Note that s_I is a minimal energy spline that interpolate the data $\bar{D}^0$ and $\bar{D}^1$. We can assume that the data comes from a function $z \in W_\infty^2(\Omega)$. Recall from [11], we can get

$$\| s_I - z \|_{\infty,\Omega} \le C_2 |\triangle_I|^2 | f |_{2,\infty,\Omega} . \qquad (3.3)$$

Since $z_i^0 := f_i^0 - s_L(x_i^0, y_i^0)$ and $z_i^1 := 0$, we have

$$\| z \|_{\infty,\Omega} \le \| f - s_L \|_{\infty,\Omega}. \qquad (3.4)$$

Then by (3.3) and (3.4), we have

$$\| s_I \|_{\infty,\Omega} = \| s_I - z + z \|_{\infty,\Omega} \le \| s_I - z \|_{\infty,\Omega} + \| z \|_{\infty,\Omega} \le C_2 |\triangle_I|^2 | f |_{2,\infty,\Omega} + \| f - s_L \|_{\infty,\Omega}. \qquad (3.5)$$

Therefore, by substituting (3.1) and (3.5) into (3.2), we have

$$\| f - s_F \|_{\infty,\Omega} \le (2C_1 |\triangle_L|^2 + C_2 |\triangle_I|^2) | f |_{2,\infty,\Omega} .$$

This completes the proof.

NUMERICAL EXAMPLE

In this section, we present a numerical example to demonstrate the performance of our approach.

Example 4.1 Let $D^0 = \{(x_i^0, y_i^0, f(x_i^0, y_i^0))\}$ be a scattered data set which is marked with a triangle without contaminated and $D^1 = \{(x_i^1, y_i^1, f(x_i^1, y_i^1) + \epsilon)\}$ be a contaminated scattered data set which is marked with a circle in Fig. 1, where $f(x, y)$ is a test function and ϵ_i is a random number between -0.01 to 0.01. We use the following test functions

$$\begin{aligned} f_1(x, y) &= 2x^4 + 5y^4, \\ f_2(x, y) &= \sin(2(x - y)), \\ f_3(x, y) &= 0.75\exp(-0.25(9x-2)^2 - 0.25(9y-2)^2) \\ &\quad +0.75\exp(-(9x+1)^2/49 - (9y+1)/10) \\ &\quad +0.5\exp(-0.25(9x-7)^2 - 0.25(9y-3)^2) \\ &\quad -0.2\exp(-(9x-4)^2 - (9y-7)^2). \end{aligned}$$

to evaluate the scattered data set. First we use the spline space $S_5^1(\triangle_L)$ to find the the fitting surfaces s_L, where $\triangle_L$ is the triangulation given in Fig. 1. Next we triangulate all data locations to get a new triangulation $\triangle_I$ as shown in Fig. 2. We modify some of the data in two groups. Let $\bar{D}^0 = \{x_i^0, y_i^0, z_i^0\}_{i=1}^M$ and $\bar{D}^1 = \{x_i^1, y_i^1, z_i^1\}_{i=1}^N$ be two new groups where $z_i^0 := f(x_i^0, y_i^0) - s_L(x_i^0, y_i^0)$ and $z_i^1 := 0$. Then we construct a minimal energy interpolatory spline $s_I \in S_5^1(\triangle_I)$. Finally, we can get our solution by $s_F = s_L + s_I$. Moreover, we compared the maximum errors against the exact function between our approach the traditional least squares method and weighted least squares method (cf. [15]) in Table 1. Here in weighted least squares method, the weights are set to be $\omega_i = 1$ and $\omega_i = 0.01$ for the exact data and contaminated data, respectively. The maximum errors are measured using 101×101 equally-spaced points over $[0,1] \times [0,1]$.

Table 1: **Approximation errors.**

Test functions	Our method	Traditional l.s. method	Weighted l.s. method
$f_1(x,y)$	0.0342	0.4383	0.0965
$f_2(x,y)$	0.0474	0.4778	0.0407
$f_3(x,y)$	0.0441	0.2988	0.0216

Discussion: From Table 1, we can see that the error bounds of our approach is better than that of the traditional least squares method. And compare to weighted least squares method, it is hard to say which method is better. But there is a one big difference between our method and weighted least squares method. That is the surface created by our method can pass through the exact data.

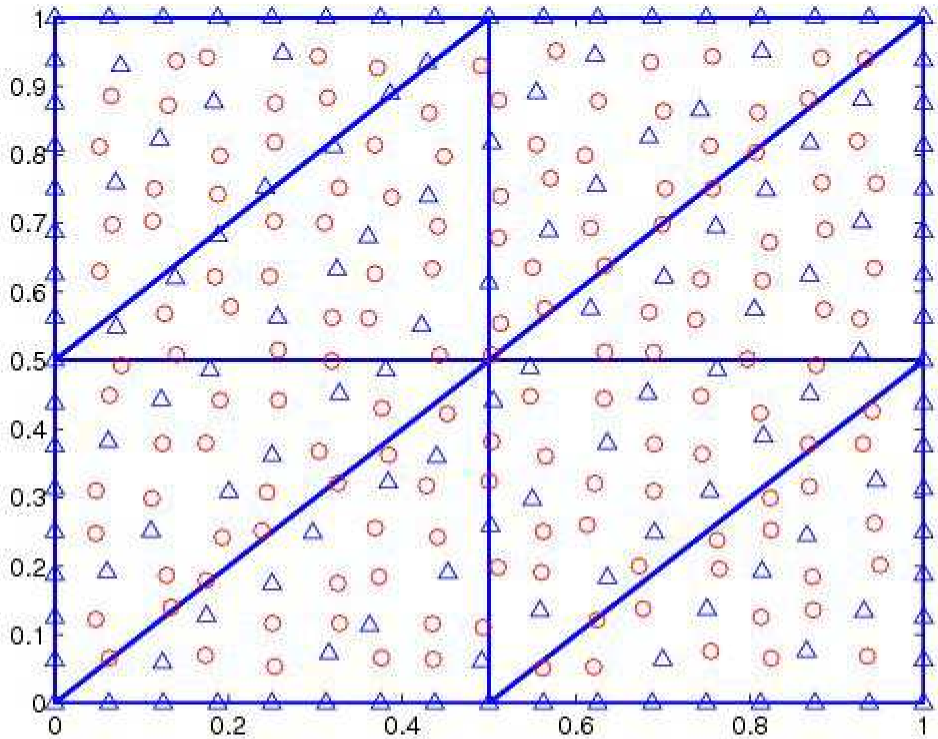

Figure 1. Scattered data and triangulation $\triangle_L$

Below, we present an example to illustrate an application of our method.

Example 4.2 We consider the shape design of a car. For example, the hood (a part of a car). A set of data are given in Fig. 3 which are used to construct the hood. The data can be divided into two groups. One is marked with triangle which means the shape must go through the data (For instance, the edge of the hood). That means we need to interpolate these data. The other is marked with cycle which need to be least squares fit. We first use the spline space $S_5^1(\triangle_L)$ to find the the fitting surfaces s_L, where $\triangle_L$ is the triangulation given in Fig. 3. Next we modify two groups of data from our method, then we construct a minimal energy interpolatory spline $s_I \in S_5^1(\triangle_I)$, where $\triangle_I$ is a new triangulation of all data locations given in Fig. 4. Finally, we draw the shape of the hood $s_F = s_L + s_I$ as shown in Fig. 5.

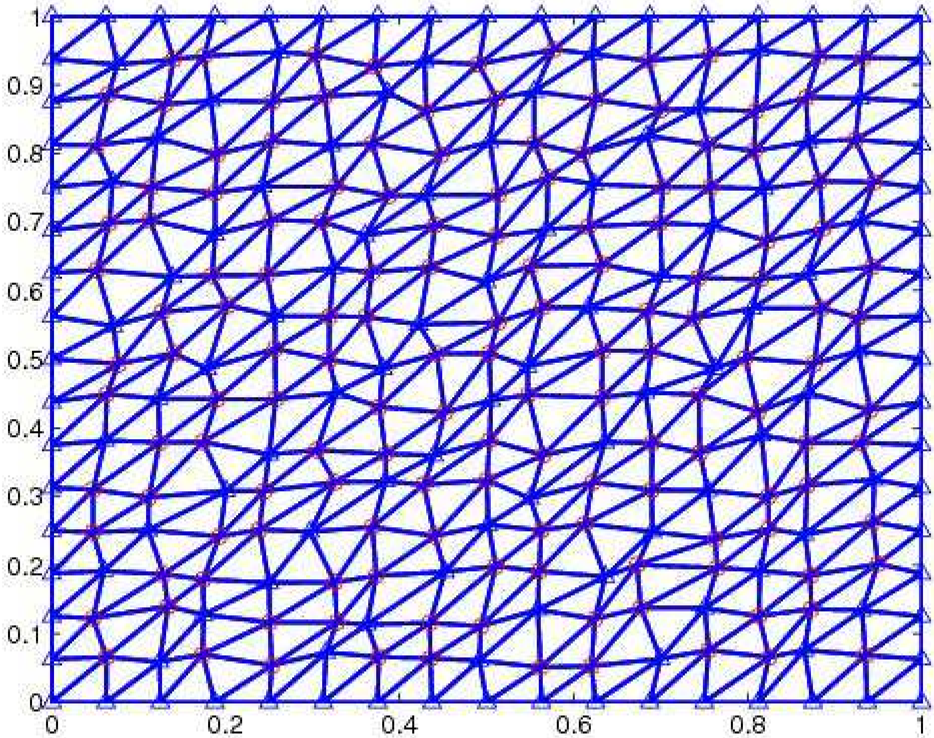

Figure 2. Triangulation Δ_I of data locations

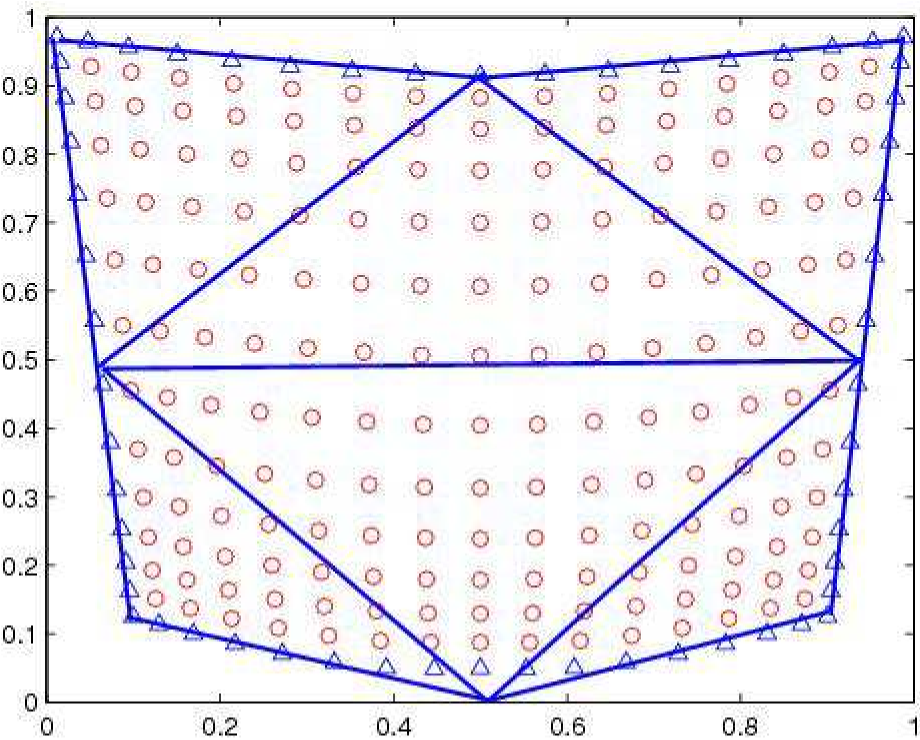

Figure 3. Data locations and triangulation Δ_L

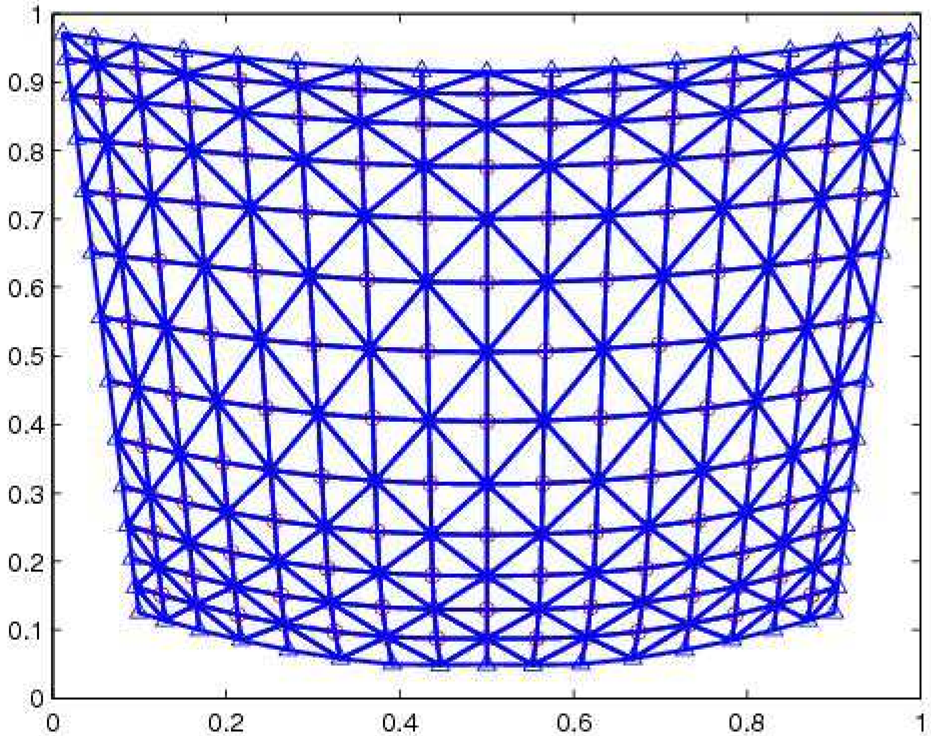

Figure 4. Triangulation $\square_I$ of data locations

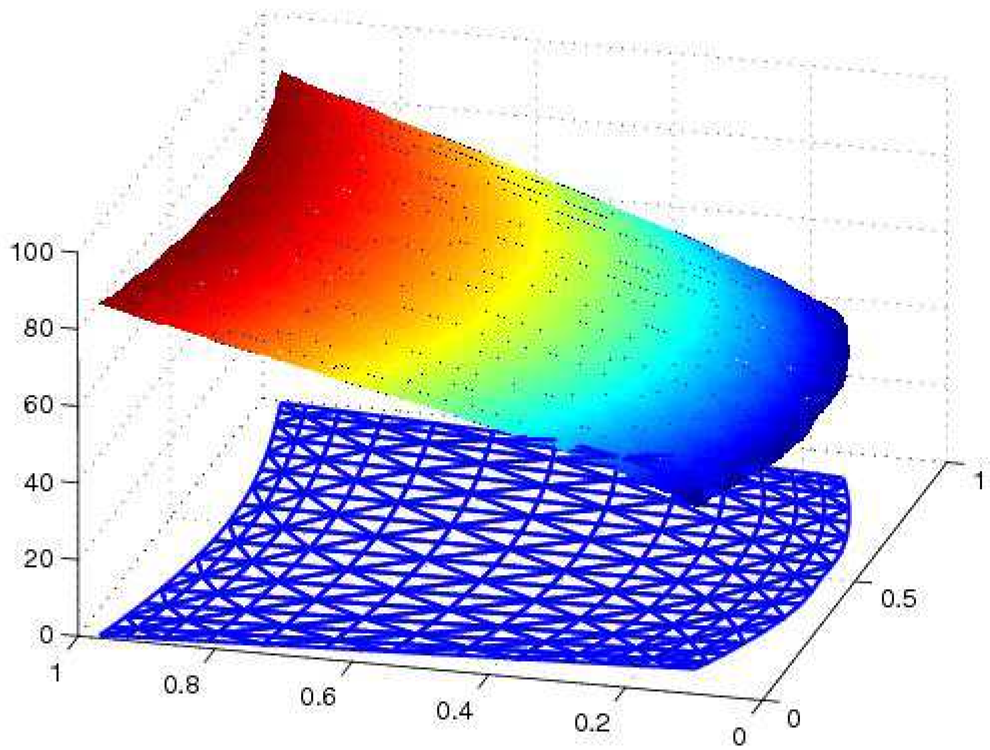

Figure 5. The shape of the hood

Acknowledgments
This work is supported by National Natural Science Foundation of China under Grant No.11201429.

REFERENCES

[1] D. L. Ragozin. *J. Approx. Theory.* v33, pp. 335-355, **1983**.
[2] M. C. Lopez de Silanes, R. Arcangeli,. *Sur la convergence des Dm--splines d'ajustement pour des données exactes ou bruitées, Revista Matemática de la Universidad Complutence de Madrid.* v4, pp 279-284, **1991**.
[3] R.Arcangeli, B.Ycart, *Almost sure convergence of smoothing Dm-spline functions. Numer. Math.* v66, pp.281-294, **1993**.
[4] M. J. Lai. *Comput. Aided Geom. Design.* v. 13, pp. 81-88, **1996**.
[5] M. J. Lai, L. L. Schumaker, *SIAM J. Numer. Anal.* v.34, pp.905-921, **1997**.
[6] M. J. Lai, L. L. Schumaker, *On the approximation power of bivariate splines, Adv. in Comp. Math.* v.9, pp251-279, **1998**.
[7] K. Farmer, M. J. Lai. *Scattered data interpolation by C2 quintic splines using energy minimization, Approximation Theory IX*, pp. 47-54, **1998**.
[8] O. Davydov,L. L. Schumaker, *On stable local bases for bivariate polynomial spline spaces, Constr. Approx.* v 18, pp 87-116, **2001**.
[9] M. von Golitschek, L. L. Schumaker,. *Bounds on projections onto bivariate polynomial spline spaces with stable local bases, Constr. Approx*, v.18, pp 241-254, **2002**.
[10] M. von Golitschek, M. J. Lai, L. L. Schumaker, , *Numer. Math*, pp.315-331, **2002**.
[11] R. Arcangeli, M. C. Lopez de Silanes, J. J. Torrens, *Multidimensional minimizing splines, Kluwer Academic Publishers,* **2004**.
[12] G. Awanou, M. J. Lai, P. Wenston, *Wavelets and Splines*, pp 24-74, **2006**.
[13] T. Zhou, D. Han, M. J. Lai *Appl. Numer. Math.* v 58, pp 646-659, **2008**.
[14] T. Zhou, D. Han. *J. Comput. Appl. Math.* v 217, pp 56-63, **2008**.
[15] M. J. Lai, L. L. Schumaker, *Spline Functions on Triangulations, United Kingdom at the University Press, Cambridge.* **2007**.
[16] A. Kouibia, M. Pasadas. *J. Comput. Appl. Math.* v 211, pp 213-222, **2008**.