ORTHOGONAL DISTANCE FITTING OF PRECISION FREE-FORM SURFACES BASED ON l_1 NORM

XIANGCHAO ZHANG, XIANGQIAN JIANG
Centre for Precision Technologies, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK

PAUL J. SCOTT
Taylor Hobson Ltd, PO Box 36, 2 New Star Road, Leicester LE4 9JQ, UK

Precision free-form surfaces are widely used in advanced optical and mechanical devices. In order to evaluate the form quality of a free-form surface, it is required to fit the measurement data with the design template and compare the relative deviation between them. A common approach is to minimize the sum of squared differences in the z direction. Its solution is not robust enough and may be biased due to outliers. This paper presents a fitting algorithm which employs the sum of orthogonal distances to evaluate the goodness of fit. The orthogonal projection points are updated simultaneously with the shape and motion parameters. Additionally, the l_1 norm is adopted to improve the robustness of the solution. The Monte-Carlo simulation demonstrated that the bias in the fitted intrinsic characteristics of this method is much smaller than the traditional algebraic fitting, whereas the fitted motion parameters have no distinct difference.

1 Introduction

Compared with traditional industrial elements which have simple geometrical shapes, free-form components occupy superior optical and aerodynamic properties, and are crucial to the development of optical and mechanical devices [1]. Surface form plays an essential role in the characteristics of these components. To assess the form quality of a free-form surface, a nominal template is required. The form error is obtained from the relative deviation between the measurement data and the template. However, usually the two surfaces do not exactly lie in the same coordinate system, thus the best *fitting* (or *matching*) needs to be established between them under some error criterion, e.g. least squares,

$$E=\sum_{i=1}^{N}[z_i'-f(x_i',y_i')]^2 \tag{1}$$

where $[x_i',y_i',z_i']^T=\mathbf{p}'_i$, $i=1,2,\cdots N$ are the data points after transformation and $z=f(x,y)$ is the nominal template function. Sometimes the template is only supplied as an analytical function, and some intrinsic characteristics (shape parameters) need to be specified as well.

In the above equation, only the deviation in the z direction is considered, which is called *algebraic fitting* [2]. This approach is commonly adopted because of its ease of implementation. However, its definition of error distances

does not coincide with measurement guidelines and the estimated fitting parameters will be biased, especially when there exist errors in the explanatory variables [2-4]. Consequently researchers have developed the *orthogonal distance fitting* (ODF, also termed *geometric fitting*) method. This technique intends to minimize the sum of squared orthogonal distances from the measurement points to the nominal surface. It can effectively overcome the bias problem of the algebraic fitting. But for free-form surfaces, the orthogonal distances are not straightforward to obtain and the computational cost may increase dramatically. This paper tries to solve this problem.

2 The Orthogonal Distance Fitting Algorithm

Suppose the explicit function of a template is given as $z=f(x,y;\mathbf{a})$. We aim to determine the intrinsic characteristics **a**, an optimal rotation matrix **R** and a translation vector **t** to minimize the sum of squared orthogonal distances from all the measurement points to the template,

$$\min_{\mathbf{a},\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \left\| \mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i \right\|^2 = \min_{\mathbf{a},\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \left\| \mathbf{p}'_i - \mathbf{q}_i \right\|^2 \tag{2}$$

Here $\mathbf{p}'_i = \mathbf{R}\mathbf{p}_i + \mathbf{t} = [x'_i, y'_i, z'_i]^T$ is an arbitrary measurement point after motion and $\mathbf{q}_i$ is the closest point on the template corresponding to $\mathbf{p}'_i$.

The coordinates of $\mathbf{q}_i$ are represented as,

$$\mathbf{q}_i = [x'_i + \xi_i,\ y'_i + \zeta_i,\ f(x'_i + \xi_i, y'_i + \zeta_i; \mathbf{a})]^T$$

and the weighting technique is incorporated. Then the error metric in Equation (2) can be rewritten as,

$$E = \sum_{i=1}^{N} w_i^2 [z'_i - f(x'_i + \xi_i, y'_i + \zeta_i; \mathbf{a})]^2 + \sum_{i=1}^{N} u_i^2 \xi_i^2 + \sum_{i=1}^{N} v_i^2 \zeta_i^2 = \mathbf{G}^T\mathbf{G} + \mathbf{H}^T\mathbf{H} \tag{3}$$

with $\mathbf{G} \in \Re^{N\times 1} : G_i = w_i [z'_i - f(x'_i + \xi_i, y'_i + \zeta_i; \mathbf{a})],\ i = 1, \cdots, N$

and $\mathbf{H} \in \Re^{2N\times 1} : H_j = \begin{cases} -u_{(j+1)/2}\xi_{(j+1)/2}, & j = 1,3,\cdots,2N-1 \\ -v_{j/2}\zeta_{j/2}, & j = 2,4,\cdots,2N \end{cases}$

If regarding $\{\xi_i\}$ and $\{\zeta_i\}$ as unknown variables, we can solve them simultaneously with the intrinsic characteristics $\mathbf{a} \in \Re^{p\times 1}$ and the six motion parameter $\theta_x, \theta_y, \theta_z, t_x, t_y,$ and t_z.

Boggs *et al* [5] proposed a method to solve this nonlinear least squares problem based on the Levenberg-Marquardt algorithm.

Denoting the shape and motion parameters with a vector $\mathbf{m} \in \Re^{(p+6)\times 1}$ and denoting $\{\xi_i\}$ and $\{\zeta_i\}$ with a vector $\boldsymbol{\beta} \in \Re^{2N\times 1}$, then Equation (3) can be solved iteratively by,

$$\begin{bmatrix} \mathbf{J} & \mathbf{V} \\ 0 & \mathbf{D} \\ \lambda^{1/2}\mathbf{S} & 0 \\ 0 & \lambda^{1/2}\mathbf{T} \end{bmatrix} \begin{bmatrix} \delta\mathbf{m} \\ \delta\boldsymbol{\beta} \end{bmatrix} = -\begin{bmatrix} \mathbf{G} \\ \mathbf{H} \\ 0 \\ 0 \end{bmatrix} \tag{4}$$

with $\mathbf{J} \in \Re^{N\times(p+6)}: J_{ij} = \partial g_i / \partial m_j, i = 1,\cdots,N,\ j = 1,\cdots,p+6$,

$\mathbf{V} \in \Re^{N\times 2N}: V_{ij} = \partial g_i / \partial \beta_j = -w_i \partial f_i / \partial \beta_j,\ i = 1,\cdots,N, j = 1,\cdots,2N$.

and $\mathbf{D} \in \Re^{2N\times 2N}: \mathbf{D} = -diag\{u_1, v_1, u_2, v_2, \cdots, u_N, v_N\}$

Here $\lambda > 0$ is a damping factor and $\mathbf{S} \in \Re^{(p+6)\times(p+6)}$ and $\mathbf{T} \in \Re^{2N\times 2N}$ are two scaling matrices. Employing damping terms can greatly improve the numerical stability of the system and guarantee the convergence of its solution. The damping factor λ need to be selected properly according to the specific problem. If the value of λ is too large, the step length of the updates $\delta\boldsymbol{\alpha}$ and $\delta\boldsymbol{\beta}$ will be very small, consequently leading to a rather slow convergence speed; On the other hand, a λ too small cannot significantly affect the system's numerical property. The normal function of Equation (4) is,

$$\begin{bmatrix} \mathbf{J}^T\mathbf{J} + \lambda\mathbf{S}^2 & \mathbf{J}^T\mathbf{V} \\ \mathbf{V}^T\mathbf{J} & \mathbf{V}^T\mathbf{V} + \mathbf{D}^2 + \lambda\mathbf{T}^2 \end{bmatrix} \begin{bmatrix} \delta\mathbf{m} \\ \delta\boldsymbol{\beta} \end{bmatrix} = -\begin{bmatrix} \mathbf{J}^T\mathbf{G} \\ \mathbf{V}^T\mathbf{G} + \mathbf{D}\mathbf{H} \end{bmatrix} \tag{5}$$

So that $\delta\mathbf{m} = -(\mathbf{J}^T\mathbf{M}\mathbf{J} + \lambda\mathbf{S}^2)^{-1}\mathbf{J}^T\mathbf{M}(\mathbf{G} - \mathbf{V}\mathbf{E}^{-1}\mathbf{D}\mathbf{H})$ (6)

and $\delta\boldsymbol{\beta} = -\mathbf{E}^{-1}[\mathbf{V}^T\mathbf{M}(\mathbf{G} + \mathbf{J}\delta\mathbf{m} - \mathbf{V}\mathbf{E}^{-1}\mathbf{D}\mathbf{H}) + \mathbf{D}\mathbf{H}]$ (7)

In the equation $\mathbf{M} = \mathbf{I} - \mathbf{V}(\mathbf{V}^T\mathbf{V} + \mathbf{E}^2)^{-1}\mathbf{V}^T$. In fact, it is a diagonal matrix

$$\mathbf{M} = diag\left\{ 1 \Big/ \left[1 + \left(w_i \frac{\partial f_i}{\partial \xi_i} \right)^2 \Big/ E_{2i-1} + \left(w_i \frac{\partial f_i}{\partial \zeta_i} \right)^2 \Big/ E_{2i} \right] \right\},\ i = 1,\cdots,N,$$

and $\mathbf{E} = \mathbf{D}^2 + \lambda\mathbf{T}^2$ is diagonal as well.

3 Improving the Robustness of the Solution Based on the l_1 Norm

The least squares approach is commonly adopted as in Equation (2) for its ease of calculation. It is unbiased when the error is normally distributed [6]. However, measurement outliers and manufacturing defects can make the result distorted, or even break down.

To improve the robustness of the fitting results, various improvements have been proposed [7]. Among these methods, the l_1 norm pays less attention to the wild points and concentrates on the vast majority of the data points. But it has discontinuous derivatives and consequently difficult to solve. Hunter and Lange [8] proposed an algorithm based on the Majorize-Minimize theory. A continuous surrogate function is adopted to approximate the initial l_1 norm objective function.

For the l_1 norm fitting problem,

$$E = \sum_{i=1}^{N} \left\| \mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i \right\| = \sum_{i=1}^{N} r_i \tag{8}$$

A small perturbation $\varepsilon > 0$ is introduced,

$$E^{\varepsilon} = \sum_{i=1}^{N} \rho^{\varepsilon}(r_i) = \sum_{i=1}^{N} \left[r_i - 2\varepsilon \ln(\varepsilon + |r_i|) \right] \tag{9}$$

It turns out that for a given residual $r^{(k)}$ at the k-th iteration, $\rho^{\varepsilon}(r)$ is majorized by $\psi^{\varepsilon}(r|r^{(k)}) = \dfrac{r^2}{\varepsilon + |r^{(k)}|} + c$, i.e. $\begin{cases} \psi^{\varepsilon}(r^{(k)}|r^{(k)}) = \rho^{\varepsilon}(r^{(k)}) \\ \psi^{\varepsilon}(r|r^{(k)}) \geq \rho^{\varepsilon}(r) \text{ for all } r \end{cases}$, thus the new objective function will be

$$E^{\varepsilon} = \sum_{i=1}^{N} \psi^{\varepsilon}(r_i) = \sum_{i=1}^{N} \left[\frac{r_i^2}{\varepsilon + |r_i^{(k)}|} + c_i \right] \tag{10}$$

It can be proved that $\left| E^{\varepsilon} - E \right| \leq -2\varepsilon N \ln \varepsilon$, hence the constant ε can be determined according to the overall error threshold $\tau = -2\varepsilon N \ln \varepsilon$.

Obviously the constants $\{c_i\}$ do not affect the solution. Therefore, the l_1 norm optimization becomes a reweighted least squares problem with $w_i = 1/\left(\varepsilon + |r_i^{(k)}|\right)$. Now recall the weights of the ODF algorithm. We set them as $\begin{cases} u_i = aw_i \\ v_i = bw_i \end{cases}, i = 1, 2, \cdots, N$. Here $a, b \geq 0$ are utilized to measure the relative influence of the lateral and vertical deviations onto the error metric. In practice if there is noise in all the x, y, and z coordinates, we set $a, b > 0$. On the contrary, if only z coordinates contain errors, we set $a = b = 0$, i.e. algebraic fitting will be adopted.

4 Numerical Example and Discussion

A biconic surface is adopted for numerical simulation [9],

$$z = \frac{c_x x^2 + c_y y^2}{1 + \sqrt{1 - (1 + k_x) c_x^2 x^2 - (1 + k_y) c_y^2 y^2}} \tag{11}$$

In the equation, the intrinsic characteristics are set to be $c_x = 2/9\ \text{mm}^{-1}$, $c_y = 2/7\ \text{mm}^{-1}$, $k_x = -0.3$, and $k_y = 0.3$. 100×100 points are sampled within the area $-2.5\ \text{mm} \leq x, y \leq 2.5\ \text{mm}$ as measurement data and transformed with $\theta_x = -2°, \theta_y = 4°, \theta_z = 3°$ and $\mathbf{t} = [-1.0\ \text{mm}, -2.0\ \text{mm}, 1.5\ \text{mm}]^{\mathrm{T}}$. Gaussian noise of $N(0, (0.5\ \mu\text{m})^2)$ is introduced onto the x and y directions and $N(0, (0.1\ \mu\text{m})^2)$ onto the z direction. 200 points are randomly sampled and Gaussian error of $N(0, (50\ \mu\text{m})^2)$ is added on them as outliers. Defects are also involved as shown in Figure 2.

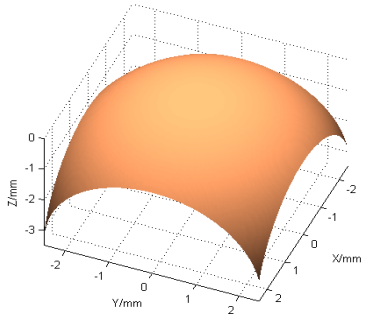

Figure 1. Biconic surface

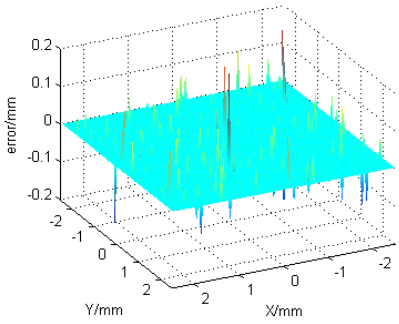

Figure 2. Defects and noise

Here we compare the l_1 norm geometric fitting, l_2 norm geometric fitting, and l_1 norm algebraic fitting. These programs are coded in MATLAB R2007a and run on a Pentium 4 PC with 3.0GHz, 2.00GB RAM. In order to make the fitting results more reliable, the Monte-Carlo simulation is employed and the fitting procedure is run 500 times. The average values of the 500 fitting errors are adopted to measure the bias of each method; and the standard deviations (σ) to assess the uncertainty. The uncertainty of the three algorithms turns out to be not significantly different, with l_2 norm ODF a little poorer. The reason is that the uncertainty of the fitted parameters is mainly determined by the amplitude of the introduced random noise. The biases of the six intrinsic characteristics and six motion parameters are listed in Table 1.

Table 1. Biases of three fitting methods

Method	l_1 Norm ODF	l_2 Norm ODF	l_1 Algebraic fitting
c_x	-2.85×10^{-6} mm^{-1}	-6.04×10^{-5} mm^{-1}	-1.17×10^{-4} mm^{-1}
c_y	1.75×10^{-5} mm^{-1}	1.36×10^{-4} mm^{-1}	-1.13×10^{-4} mm^{-1}
k_x	4.20×10^{-4}	3.62×10^{-3}	1.73×10^{-3}
k_y	-3.28×10^{-4}	-4.13×10^{-3}	1.50×10^{-3}
θ_x	-0.0736 °	0.116 °	-0.231 °
θ_y	-0.166 °	-0.711 °	-0.0755 °
θ_z	1.22×10^{-3} °	2.09×10^{-2} °	7.69×10^{-4} °
t_x	11.6 μm	49.1 μm	5.33 μm
t_y	-3.66 μm	5.34 μm	-11.5 μm
t_z	-0.0210 μm	-0.157 μm	0.108 μm
Running time	1.570 sec	1.539 sec	0.253 sec

It can be seen that the biases of the intrinsic characteristics of the l_1 norm geometric fitting are at least one order smaller than the other two methods. It demonstrates that ODF technique can effectively overcome the bias problem and the l_1 norm error metric can greatly improve the system stability and robustness against outliers and defects.

The geometric fitting shows no significant superiority over algebraic fitting on motion parameters. In fact, it holds true for most smooth surfaces. Therefore, when the shape of the surface is completely fixed, the motion parameters can be calculated using the conventional algebraic fitting algorithm, which is relatively simpler and quicker to implement.

5 Conclusions

Orthogonal distance fitting shows obvious superiority on accuracy and unbiasness over algebraic fitting methods, especially when the surface is highly curved and both its lateral and vertical coordinates contain errors. The error metric of ODF is consistent with the standard error definition. Numerical simulation shows the fitting accuracy of the intrinsic characteristics can be greatly improved. Moreover, real measurement data may be polluted by outliers, missing data or defects, consequently causing serious error in the least-squared fitted parameters. In this case, the l_1 norm regression can be adopted to improve the robustness of the system. On the other hand, if the specific distribution of the noise has been known beforehand, we will try to transform the corresponding objective function into an appropriate reweighted least squares problem, so that the Boggs algorithm can still be employed.

References

[1] X Jiang, P J Scott, D J Whitehouse et al. Paradigm shifts in surface metrology II. The current shift. Proc Royal Soc. A. 2007; 463: 2071-2099

[2] S J Ahn, W Rauh et al. Least-squares orthogonal distances fitting of circle sphere, ellipse, hyperbola and parabola. Patt. Recog. 2001; 34: 2283-2303

[3] DIN 32880-1, Coordinate Metrology; Geometrical Fundamental Principles, Terms and Definitions, German Standard, Beuth Verlag, Berlin, 1986

[4] W Sun. Precision Measurement and Characterisation of Spherical and Aspheric Surfaces. Ph.D Thesis, Univ. of Southampton, UK. May 2007

[5] P T Boggs, R H Byrd and R B Schabel. A stable and efficient algorithm for nonlinear orthogonal distance regression. SIAM J Stat Comput. 1987; 8(6): 1052-1078

[6] R M Barker, M G Cox, A B Forbes and P M Harris. Discrete modelling and experimental data analysis. Version 2. NPL Report. 2004

[7] W J J Rey. Introduction to Robust and Quasi-Robust Statistical Methods. Springer-Verlag. 1983

[8] D R Hunter and K Lange. Quantile regression via an MM Algorithm. Journal of Computational and Graphical Statistics. 2000; 9(1): 60-77

[9] ZEMAX Optical Design Program, User's Guide, April 2006