Data fitting with a set of two concentric spheres

HELMUTH SPÄTH*

Abstract. *We consider fitting data points in space by a set of two concentric spheres. This problem ought to occur within computational metrology. A heuristic algorithm is developed and its efficiency is demonstrated by some numerical example.*

Key words: *data fitting, two concentric spheres*

AMS subject classifications: 65D10

Received June 8, 2006 Accepted June 13, 2006

1. Statement of the problem

One standard problem in computational metrology [1, Problem C6] is to find for measured points in space two concentric spheres of minimum difference in radii that contain all the points between them. But assuming all points to have errors it ought to be more adequate to determine two spheres such that a suitable sum of squared errors attains a minimum where simultaneously each point has to be associated with one of the spheres.

A set of two concentric spheres is given by one common center (a, b, c) and two radii r and R with $R > r > 0$. Some parametric representation for the smaller sphere is given by

$$\begin{aligned} x(u,v) &= a + r\cos u \sin v\,, \\ y(u,v) &= b + r\sin u \sin v\,, \\ z(u,v) &= c + r\cos v, \qquad 0 \le u < 2\pi,\ 0 < v \le \pi \end{aligned} \tag{1}$$

and the larger one is obtained when replacing r by R.

2. Smallest distance of some given point to some sphere

To find the minimal squared distance $p_j^2 = p_j^2(u,v)$ of some given point (x_j, y_j, z_j), $j = 1, \ldots, m$, onto the smaller sphere requires to minimize

$$\begin{aligned} p_j^2(u,v) = {} & (a + r\cos u \sin v - x_j)^2 + (b + r\sin u \sin v - y_j)^2 \\ & +(c + r\cos v - z_j)^2 \end{aligned} \tag{2}$$

*Department of Mathematics, University of Oldenburg, 26 111 Oldenburg, Germany, e-mail: spaeth@mathematik.uni-oldenburg.de

with respect to u and v. The corresponding function $q_j^2 = q_j^2(u,v)$ for the larger sphere is received by replacing r by R in (2). The necessary conditions for a minimum of (2) (and similarly for q_j^2) are

$$\frac{1}{2r}\frac{\partial p_j^2}{\partial u} = \sin v(\cos u(y_j - b) - \sin u(x_j - a)) = 0\,, \tag{3}$$

$$\frac{1}{2r}\frac{\partial p_j^2}{\partial v} = \cos v(\cos u(x_j - a) + \sin u(y_j - b)) - \sin v(z_j - c) = 0\,. \tag{4}$$

For (3) it results either

$$\sin v = 0 \qquad \text{(implying } v = \pi \text{ and } p_j^2 \text{ constant)} \tag{5}$$

or

$$\operatorname{tg} u = \frac{y_j - b}{x_j - a}\,. \tag{6}$$

From (4) it results

$$\operatorname{tg} v = \frac{\cos u(x_j - a) + \sin u(y_j - b)}{z_j - c}\,. \tag{7}$$

Both expressions (6) and (7) have two solutions

$$u = \operatorname{atan}\frac{(y_j - b)}{(x_j - a)} \quad \text{and } u := u + \pi\,, \tag{8}$$

$$v = \operatorname{atan}\left(\frac{\cos u(x_j - a) + \sin u(y_j - b)}{z_j - c}\right) \quad \text{and } v := v + \pi, \tag{9}$$

stemming from the fact that e.g.

$$\operatorname{tg} u = \frac{\sin u}{\cos u} = \frac{\sin(u+\pi)}{\cos(u+\pi)} = \frac{-\sin u}{-\cos u}\,,$$

i.e. we have to choose the right signs of $\sin u$ and $\cos u$ to get the absolute minimum. As expressions (6) and (7) do not depend on r, they are the same for q_j^2.

The decision process is as follows [2]. For $j = 1, \ldots, m$

(i) Use (6) to determine u.

(ii) Calculate $suj = \sin u$, $cuj = \cos u$.

(iii) If $(y_j - b) < 0$, set $suj = -suj$.

(iv) If $(x_j - a) < 0$, set $cuj = -cuj$.

(v) Use (7) to determine v.

(vi) Calculate $svj = \sin v$, $cvj = \cos v$. (10)

(vii) If $(cuj(x_j - a) + suj(y_j - b)) < 0$, set $svj = -svj$.

(viii) If $(z_j - c) < 0$, set $cvj = -cvj$.

(ix) Store the resulting right values into arrays, i.e. set $su(j) = suj$, $cu(j) = cuj$, $sv(j) = svj$, $cv(j) = cvj$, and use these values corresponding to the absolute minima of p_j^2 and q_j^2, respectively, when evaluating these expressions.

3. The objective function

In order to set up a reasonable least squares principle we try to find two subsets V and W of $\{1, \dots, m\}$ with $V \cap W = \emptyset$, $V \cup W = \{1, \dots, m\}$, $|V| \leq 4$, $|W| \leq 4$ such that the sum of squared minimal distances

$$S(V, W, a, b, c, r, R) = \sum_{j \in V} \min_{u,v} p_j^2(u, v) + \sum_{j \in W} \min_{u,v} q_j^2(v, v) \tag{11}$$

is minimized. Here V is the set of given points associated with the smaller sphere and W is that one for the larger sphere. Thus we have a mixed combinatorial (V, W) and continuous (a, b, c, r, R) minimization problem without an analytic solution.

4. Some heuristic algorithm

We propose the following iterative method where (beginning with suitable starting values for a, b, c, r, R) alternatively (V, W) and (a, b, c, r, R) are improved such that S is descending. This will not guarantee convergence to some local or even to a global minimum. But numerical experience shows that the method will normally work in the desired sense, i.e. will find a global minimum.

Step 1 [One special way of estimating starting values].
Set $t = 0$ *(iteration counter) and* $S^{(0)} = \infty$. *For an initial center* $(a^{(0)}, b^{(0)}, c^{(0)})$ *we propose the means of all given points, i.e.*

$$a^{(0)} = \overline{x} = \frac{1}{m} \sum_{j=1}^{m} x_j, \quad b^{(0)} = \overline{y} = \frac{1}{m} \sum_{j=1}^{m} y_j, \quad c^{(0)} = \overline{z} = \frac{1}{m} \sum_{j=1}^{m} z_j \, . \tag{12}$$

Then we calculate

$$\begin{aligned} h^2 &= \min_j \left((x_j - a^{(0)})^2 + (y_j - b^{(0)})^2 + (z_j - c^{(0)})^2 \right), \\ H^2 &= \max_j \left((x_j - a^{(0)})^2 + (y_j - b^{(0)})^2 + (z_j - c^{(0)})^2 \right), \end{aligned} \tag{13}$$

i.e. the minimal and the maximal squared distance of all the given points ($j = 1, \dots, m$) *to the defined starting center. If we now take* $r^{(0)} = h$ *and* $R^{(0)} = H$ *as estimates, then all given points would lie between the inner and the outer sphere. To avoid this (because not all points with errors are estimated to be such) we reduce* $R^{(0)}$ *and increase* $r^{(0)}$ *by setting*

$$R^{(0)} = f * H, \quad r^{(0)} = \frac{1}{f} * h \quad \text{with } f < 1 \tag{14}$$

and additionally f *such that* $R^{(0)} > r^{(0)}$.

Step 2 [Determination of $V^{(t)}$ and $W^{(t)}$].
Set $V^{(t)} = W^{(t)} = \emptyset$. *For* $j = 1, \dots, m$ *we use* $su(j)$, $cu(j)$, $sv(j)$, $cv(j)$ *from* (10) *depending on* $(a^{(t)}, b^{(t)}, c^{(t)})$ *to calculate* p_j^2 *and* q_j^2 *via* (2). *If* $p_j^2 < q_j^2$, *then* j *is adjoined to* $V^{(t)}$ *otherwise to* $W^{(t)}$. *Also* $|V| = |V^{(t)}|$, $|W| = |W^{(t)}|$, *and* $S^{(t)} = S(V^{(t)}, W^{(t)}, a^{(t)}, b^{(t)}, c^{(t)}, r^{(t)}, R^{(t)})$ *are calculated.*

Step 3 [Updating of a, b, c, r, R].
For fixed $V = V^{(t)}$ and $W = W^{(t)}$ and corresponding values for $su(j)$, $cu(j)$, $sv(j)$, $cv(j)$, $j = 1, \dots, m$ from Step 2 we will minimize

$$S(V, W, a, b, c, r, R) = \sum_{j \in V} \min p_j^2 + \sum_{j \in W} \min q_j^2 \tag{15}$$

w.r.t. a, b, c, r, R. The necessary conditions accordingly are

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = \frac{\partial S}{\partial r} = \frac{\partial S}{\partial R} = 0$$

and result in the following linear systems of five equations and five unknowns a, b, c, r, R:

$$\begin{pmatrix} m & 0 & 0 & F(V) & F(W) \\ 0 & m & 0 & G(V) & G(W) \\ 0 & 0 & m & H(V) & H(W) \\ F(V) & G(V) & H(V) & |V| & 0 \\ F(W) & G(W) & H(W) & 0 & |W| \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ r \\ R \end{pmatrix} = \begin{pmatrix} m\overline{x} \\ m\overline{y} \\ m\overline{z} \\ E(V) \\ E(W) \end{pmatrix}, \tag{16}$$

where

$$\begin{aligned} F(V) &= \sum_{j \in V} cu(j)\, sv(j)\,, \\ G(V) &= \sum_{j \in V} su(j)\, sv(j)\,, \\ H(V) &= \sum_{j \in V} cv(j)\,, \\ E(V) &= \sum_{j \in V} x_j\, cu(j)\, sv(j)\ + y_j\, su(j)\, sv(j)\ + z_j\, cv(j)\,. \end{aligned} \tag{17}$$

The coefficient matrix in (16) is symmetric. If it is nonsingular (empirically true) there will be a unique solution of (16) and we can proceed.

Step 4. *We set $t := t + 1$, $a^{(t)} = a$, $b^{(t)} = b$, $c^{(t)} = c$, $r^{(t)} = r$, $R^{(t)} = R$ and go back to Step 2 if there was some descent for S or if t is below some given upper bound. Otherwise stop.*

5. Some numerical example

To get test data we started with $(a, b, c) = (1, 2, 3)$, $r = 4$, $R = 5$. Then we generated 8 points

$$\begin{aligned} x_j &= a + r \cos u \sin v \\ y_j &= b + r \sin u \sin v \qquad (j = 1, \dots, 8) \\ z_j &= c + r \cos v \end{aligned}$$

where for each j a new pair of pseudo random numbers (u, v) with u equally distributed in $[0, 2\pi)$ and v equally distributed in $(0, \pi]$ was used. The resulting data points were disturbed by rounding them to the next integers. Similarly for $j = 9, \dots, 16$ the same procedure was done with R instead of r. Putting together these data points we got

x	1	-1	0	-1	2	2	4	5	1	-1	0	-1	3	2	5	6
y	2	1	3	4	2	3	1	1	2	1	3	5	2	3	1	1
z	-1	0	-1	6	-1	-1	6	4	-2	-1	-2	7	-2	-2	6	5

Thus it can be expected that $V = \{1, \dots, 8\}$ and $W = \{9, \dots, 16\}$ should result and also (a, b, c) near $(1, 2, 3)$ and (r, R) near $(4, 5)$. We used for the factor f in Step 1 four values $f = .7, .8, .9, 1$. As the results were identical and also the course of the iterations was very similar, we will give details only for $f = .8$. To characterize V and W we introduce an integer vector p with $p(j) = 1$ if $j \in V$ and $p(j) = 2$ if $j \in W$ $(j = 1, \dots, m = 16)$.

t	p															
0–30	1	1	1	2	1	1	2	2	1	1	1	2	1	1	2	2
31–32	1	1	1	2	1	1	2	1	1	1	1	2	1	1	2	2
33	1	1	1	1	1	1	2	1	1	1	1	2	2	1	2	2
34	1	1	1	1	1	1	2	1	1	1	2	2	2	1	2	2
35–138	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2

Table 1.

t	s	a	b	c	r	R
0	6.5376	1.6875	2.1875	1.3125	2.9263	5.5125
33	3.46524	1.4118	2.1721	2.3017	4.0188	5.2403
34	2.53933	1.4017	2.7472	2.5183	4.1544	5.1793
35	1.56106	1.4035	2.7647	2.6623	4.1878	5.0905
50	.18400	1.4569	2.7644	3.0095	4.1580	5.1365
75	.18354	1.4498	2.7474	3.0113	4.1562	5.1360
100	.18347	1.4470	2.7404	3.0120	4.1555	5.1358
138	.18345	1.4456	2.7369	3.0124	4.1551	5.1357

Table 2.

Table 1 shows the change of the vector p during the iterations. Thus we started with $V = \{1, 2, 3, 5, 6, 9, 10, 11, 13, 14\}$ and $W = \{4, 7, 8, 12, 15, 16\}$ and ended up (as expected) with $V = \{1, \dots, 8\}$ and $W = \{9, \dots, 16\}$. *Table 2* shows for $t = 0$ the initial values after Step 1 and the slow changes after $t = 50$. Other examples showed a similar behaviour. Sometimes the initial V and W were so good that those sets did not change through the iterations on (a, b, c, r, R).

References

[1] V. Srinivasan, *How tall is the pyramid of Cheops ... And other problems in computational metrology*, SIAM News **29**(1996), 8–9.

[2] H. Späth, *Least-squares fitting with spheres*, J. Opt. Theory and Appl. **96**(1998), 191–199.