Vesic Library

hpgavin@duke.edu
08/15/2003 10:26 AM

To: lsc-requests@notes.duke.edu
Subject: Please fill LSC/DOSS request for...

This data was submitted on: Friday, August 15, 2003 at 10:26:47

ID NUMBER:
771822661

NAME:
Henri Gavin

PHONE:
660-5201

CLIENT_EMAIL:
hpgavin@duke.edu

STATUS:
Faculty

CALL NUMBER:

N.S. GFA
D0215 9472

Article

TITLE:
Siam Review

VOL:
36

YEAR:

Photo Copy

NUMBER:

PAGES:
258-264

AUTHOR:
Y. Nievergelt

TITLE OF ARTICLE:
Total least squares: state of the art regression in numerical analysis

NO LONGER:
december 2003

DELIVER_TO:
Vesic Library

STATUS:
Faculty

NOTES:

SIAM REVIEW
Vol. 36, No. 2, pp. 258–264, June 1994

CLASSROOM NOTES

EDITED BY JACK W. MACKI

This section contains brief notes which are essentially self-contained applications of mathematics that can be used in the classroom. New applications are preferred, but exemplary applications not well known or readily available are accepted.

Both "modern" and "classical" applications are welcome, especially modern applications to current real world problems.

Notes should be submitted to Jack W. Macki, Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1.

TOTAL LEAST SQUARES: STATE-OF-THE-ART REGRESSION IN NUMERICAL ANALYSIS*

YVES NIEVERGELT†

Abstract. Total least squares regression (TLS) fits a line to data where errors may occur in both the dependent and independent variables. In higher dimensions, TLS fits a hyperplane to such data. The elementary algorithm presented here fits readily in a first course in numerical linear algebra.

Key words. regression, total least squares, singular value decomposition

AMS subject classifications. 15A18, 65F20, 65U05

1. Total least squares fits in numerical and applied analysis. Many applied quantitative problems, especially in engineering and in the sciences, involve fitting a curve or a surface to data points [15]. The most common solution, through the normal equations for least squares regression, appears with complete derivations in textbooks at many levels: business mathematics [16], calculus [12], linear algebra [13], numerical analysis [9], [11], probability and statistics [8], and applied mathematics [14]. For practical computations less sensitive to errors, solutions through orthogonal factorizations of matrices also appear in many texts in numerical analysis [9], [11]. Yet students often question the least squares regression's goal of minimizing the sum of squared discrepancies of only one of the variables but not of the others. Indeed, in many applications, all coordinates of the data may suffer from errors [15]. Fortunately, a method exists to allow for such errors in all coordinates: total least squares regression (TLS), as developed by Golub and Van Loan [6] and then perfected by Van Huffel [15]. The method appears promising "to interpret chemical analyses of minerals in metamorphic rocks" in geology [4], and for signal processing, system identification, and control system design, and analysis in electrical engineering [10]. Unfortunately, though useful to practitioners and known to numerical analysts, the concept of total least squares may still be unfamiliar to most mathematicians; for instance, it does not appear in the recent expository survey [3]. Moreover, despite its usefulness and its simplicity, such a method has not yet appeared in numerical analysis and applied mathematics texts, even though it would fit very nicely there as an application of the singular value decomposition, as in [9], [11], and [14]. Therefore, by introducing students to total least squares, the present note may complement

*Received by the editors March 22, 1993; accepted for publication September 13, 1993. This research was supported by National Science Foundation grant DUE-9255539, and was conducted while the author was visiting Massachusetts Institute of Technology, Cambridge, Massachusetts.

†Department of Mathematics, MS-32, Eastern Washington University, Cheney, Washington 99004-2415.

the usual courses in numerical analysis and applied mathematics, or serve as a transition from such courses to a more advanced and specialized course, for example [15].

2. The line of total least squares passes through the centroid. In the plane, regression by total least squares determines a line that minimizes the sum of the squared distances from that line to given data points in the plane. To this end, recall that the distance d from a point (p, q) to a straight line with equation $rx + sy - c = 0$ admits the following formula, which results from high school geometry [1] or calculus [12]:

$$d^2 = \frac{(rp + sq - c)^2}{r^2 + s^2}.$$

Moreover, if the line passes through a point $\vec{\mathbf{x}}_0 = (x_0, y_0)$, then $c = rx_0 + sy_0$, and the equation becomes $r(x - x_0) + s(y - y_0) = 0$. Thus, with given data points $\vec{\mathbf{x}}_1 = (x_1, y_1), \ldots, \vec{\mathbf{x}}_n = (x_n, y_n)$, the problem of total least squares consists of determining a line, defined by a point (x_0, y_0) on the line and by a normal vector (r, s), which minimizes the sum D of the squared distances:

$$D(r, s, \vec{\mathbf{x}}_0; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) = \sum_{i=1}^{n} \frac{(r(x_i - x_0) + s(y_i - y_0))^2}{r^2 + s^2}. \tag{1}$$

A first simplification results from realizing that such a line must pass through the centroid of the data, $\overline{\vec{\mathbf{x}}} = (\overline{x}, \overline{y})$, defined by

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \overline{y} := \frac{1}{n}\sum_{i=1}^{n} y_i.$$

LEMMA. *For every normal vector $(r, s) \in \mathbb{R}^2\backslash\{(0, 0)\}$, for every point $(x_0, y_0) \in \mathbb{R}^2$, and for every set of data points $(x_1, y_1), \ldots, (x_n, y_n)$, the following inequality holds:*

$$D(r, s, \vec{\mathbf{x}}_0; \vec{\mathbf{x}}_1, \ldots \vec{\mathbf{x}}_n) \geq D(r, s, \overline{\vec{\mathbf{x}}}; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n), \tag{2}$$

with equality if, but only if, $(x_0, y_0) = (\overline{x}, \overline{y})$. Consequently, the line of total least squares must pass through the centroid $(\overline{x}, \overline{y})$.

Proof. Firstly, consider the vector $\vec{\mathbf{w}} = (w_1, \ldots, w_n)$ defined by $w_i := r(x_i - x_0) + s(y_i - y_0)$ for $i \in \{1, \ldots, n\}$, and, with $\| \ \|_2$ denoting the Euclidean norm, observe that

$$D(r, s, \vec{\mathbf{x}}_0; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) = \|\vec{\mathbf{w}}\|_2^2/(r^2 + s^2).$$

Secondly, also consider the vector $\vec{\mathbf{z}} = (z_1, \ldots, z_n)$ defined by $z_i := r(x_i - \overline{x}) + s(y_i - \overline{y})$ for $i \in \{1, \ldots, n\}$, so that

$$D(r, s, \overline{\vec{\mathbf{x}}}; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) = \|\vec{\mathbf{z}}\|_2^2/(r^2 + s^2).$$

Moreover, let $\vec{\mathbf{1}} := (1, \ldots, 1) \in \mathbb{R}^n$, with all its coordinates equal to 1; hence, with

$$h := r(\overline{x} - x_0) + s(\overline{y} - y_0),$$

notice that

$$\vec{\mathbf{w}} = \vec{\mathbf{z}} + h\vec{\mathbf{1}}.$$

Furthermore, $\vec{\mathbf{z}} \perp \vec{\mathbf{1}}$; indeed, with $\langle , \rangle$ denoting the dot product,

$$\langle \vec{\mathbf{z}}, \vec{\mathbf{1}} \rangle = \sum_{i=1}^{n} z_i 1 = r\left(\sum_{i=1}^{n} x_i - n\overline{x}\right) + s\left(\sum_{i=1}^{n} y_i - n\overline{y}\right) = 0r + 0s = 0.$$

Finally, the orthogonality $\vec{\mathbf{z}} \perp \vec{\mathbf{1}}$, the relation $\vec{\mathbf{w}} = \vec{\mathbf{z}} + h\vec{\mathbf{1}}$, and the Pythagorean t give

$$\begin{aligned} D(r, s, \vec{\mathbf{x}}_0; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) &= \frac{1}{r^2 + s^2} \|\vec{\mathbf{w}}\|_2^2 \\ &= \frac{1}{r^2 + s^2} \left(\|\vec{\mathbf{z}}\|_2^2 + h^2 \|\vec{\mathbf{1}}\|_2^2 \right) \\ &= D(r, s, \overline{\vec{\mathbf{x}}}; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) + \frac{h^2 n}{r^2 + s^2} \\ &\geq D(r, s, \overline{\vec{\mathbf{x}}}; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n). \end{aligned}$$

Thus, $D(r, s, \vec{\mathbf{x}}_0; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) \geq D(r, s, \overline{\vec{\mathbf{x}}}; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n)$, with equality if, but only if, $(\overline{x}$ (x_0, y_0). □

3. Total least squares minimizes the norm of a linear map. To find the normal v (r, s), or the slope $m = -r/s$, that minimizes the discrepancy D, rewrite D as the sq Euclidean norm of the product of the unit vector $\vec{\mathbf{t}} := (r^2 + s^2)^{-1/2}(r, s)$ and the m $M \in \mathbb{M}_{n \times 2}(\mathbb{R})$ defined by

$$M := \begin{pmatrix} x_1 - \overline{x} & y_1 - \overline{y} \\ \vdots & \vdots \\ x_n - \overline{x} & y_n - \overline{y} \end{pmatrix}.$$

PROPOSITION. *The sum of squared distances D reaches its minimum at the unit nor vector $\vec{\mathbf{t}} := (r^2 + s^2)^{-1/2}(r, s)$, where the map $\vec{\mathbf{t}} \mapsto \|M\vec{\mathbf{t}}\|_2$ reaches its minimum on the sphere $S^1 = \{\vec{\mathbf{t}} \in \mathbf{R}^2 : \|\vec{\mathbf{t}}\|_2 = 1\}$.*

Proof.

$$\begin{aligned} D(r, s, \overline{\vec{\mathbf{x}}}; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) &= \sum_{i=1}^{n} \frac{(r(x_i - \overline{x}) + s(y_i - \overline{y}))^2}{r^2 + s^2} \\ &= \left\| \begin{pmatrix} x_1 - \overline{x} & y_1 - \overline{y} \\ \vdots & \vdots \\ x_n - \overline{x} & y_n - \overline{y} \end{pmatrix} \frac{1}{\sqrt{r^2 + s^2}} \begin{pmatrix} r \\ s \end{pmatrix} \right\|_2^2 \\ &= \left\| M \vec{\mathbf{t}} \right\|_2^2. \quad \square \end{aligned}$$

4. The line of total least squares corresponds to the smallest singular value. Tl preceding proposition shows where D reaches its minimum, and the following theorem revea how to compute that minimum.

THEOREM. *The sum of squared distances D reaches its minimum at every unit eigenvectc $\vec{\mathbf{t}} := (r^2 + s^2)^{-1/2}(r, s)$ corresponding to the smallest eigenvalue of the matrix $M^T M$.*

Proof. Because $M^T M$ is real, symmetric, and positive semidefinite, a basis $U = (\vec{\mathbf{u}}_1, \vec{\mathbf{u}}_2$ of $\mathbb{R}^2$ exists, consisting of orthonormal eigenvectors of $M^T M$ with eigenvalues $\sigma_1^2 \geq \sigma_2^2$ These are the same eigenvectors and eigenvalues as in the singular value decomposition (SVD of M. Since the matrix U is orthogonal, it preserves Euclidean norms, and, consequently,

$$\begin{aligned}
\|M\vec{\mathbf{t}}\|_2^2 &= \vec{\mathbf{t}}M^TM\vec{\mathbf{t}}\\
&= \vec{\mathbf{t}}^T U \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} U^T\vec{\mathbf{t}}\\
&= (U^T\vec{\mathbf{t}})^T \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} (U^T\vec{\mathbf{t}})\\
&= \sigma_1^2((U^T\vec{\mathbf{t}})_1)^2 + \sigma_2^2((U^T\vec{\mathbf{t}})_2)^2.
\end{aligned}$$

However, $\|U^T\vec{\mathbf{t}}\|_2 = \|\vec{\mathbf{t}}\|_2 = 1$, and, consequently, $\vec{\mathbf{t}} \mapsto \|M\vec{\mathbf{t}}\|_2$ reaches its minimum on the unit sphere where $\langle \vec{\mathbf{u}}_1, \vec{\mathbf{t}}\rangle = (U^T\vec{\mathbf{t}})_1 = 0$, that is, where $\vec{\mathbf{t}} = \vec{\mathbf{u}}_2$, the eigenvector with the smallest eigenvalue. □

Example. Consider the three data points (1, 2), (2, 6), and (6, 1). Calculate $(\overline{x}, \overline{y})$:

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i = (1+2+6)/3 = 3, \qquad \overline{y} := \frac{1}{n}\sum_{i=1}^{n} y_i = (2+6+1)/3 = 3.$$

Subtract the centroid from each data point, and form the matrix M:

$$M = \begin{pmatrix} 1-3 & 2-3 \\ 2-3 & 6-3 \\ 6-3 & 1-3 \end{pmatrix} = \begin{pmatrix} -2 & -1 \\ -1 & 3 \\ 3 & -2 \end{pmatrix}.$$

Compute the eigenvalues of the symmetric matrix M^TM:

$$M^TM = \begin{pmatrix} 14 & -7 \\ -7 & 14 \end{pmatrix}.$$

For instance, among many algorithms, Jacobi's method yields an orthogonal matrix Ω with the eigenvectors of M^TM in its columns, and the eigenvalues on the diagonal of $\Omega^T(M^TM)\Omega$:

$$\Omega = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \qquad \Omega^T(M^TM)\Omega = \begin{pmatrix} 7 & 0 \\ 0 & 21 \end{pmatrix}.$$

Thus, the smaller eigenvalue, 7, corresponds to the eigenvector $(r, s) = (1/\sqrt{2})(1, 1)$. Consequently, the line of total least squares has equation

$$r(x-\overline{x}) + s(y-\overline{y}) = 0, \qquad (1/\sqrt{2})(x-3) + (1/\sqrt{2})(y-3) = 0,$$

whence algebra yields any equivalent form, for example, $y = -x + 6$.

In contrast, the line of ordinary least squares, which minimizes $G(m, b) := \sum_{i=1}^{n}(mx_i + b - y_i)^2$, has equation $y = -1/2x + 9/2$, while the line that minimizes $H(m, b) := \sum_{i=1}^{n}(my_i + b - x_i)^2$ has equation $x = -1/2y + 9/2$, so that $y = -2x + 9$. For comparison, Fig. 1 displays all three regression lines.

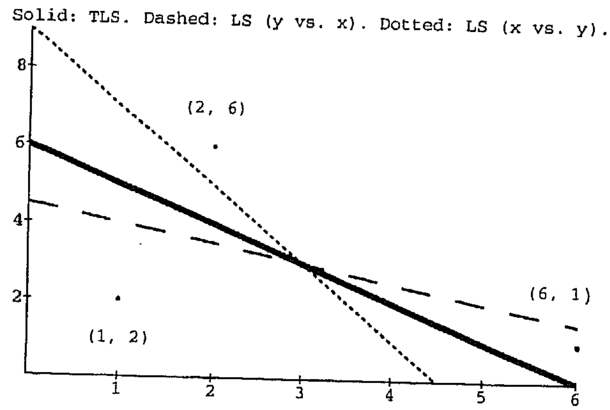

FIG. 1. *Illustration of the lines of total least squares and ordinary least squares.*

Exercise. Determine the line of total least squares and the lines of ordinary least squar for the data

$$(1,3), \quad (3,1), \quad (4,5), \quad (5,7), \quad (7,4).$$

5. The total least squares algorithm generalizes to higher dimensions. The precedir sections have introduced the concept of total least squares and an algorithm that generaliz mutatis mutandis to higher dimensions. To n data points in $\mathbb{R}^m$,

$$\vec{\mathbf{x}}_1 = (x_{1,1}, \ldots, x_{1,m}), \ldots, \vec{\mathbf{x}}_n = (x_{n,1}, \ldots, x_{n,m}),$$

the method of total least squares fits a hyperplane passing through a point $\vec{\mathbf{x}}_0$ with a norm vector $\vec{\mathbf{r}} = (r_1, \ldots, r_m)$ and equation

$$\langle \vec{\mathbf{x}} - \vec{\mathbf{x}}_0, \vec{\mathbf{r}} \rangle = 0,$$

which minimizes the sum D of the squared distances from the data points to that hyperplane

$$D(\vec{\mathbf{r}}, \vec{\mathbf{x}}_0; \vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n) = \sum_{i=1}^{n} \frac{\langle \vec{\mathbf{x}}_i - \vec{\mathbf{x}}_0, \vec{\mathbf{r}} \rangle^2}{\|\vec{\mathbf{r}}\|_2^2}.$$

For a geometric derivation of the formula for the distance from a point to a plane, see, fo instance, [5]. As in the plane, a similar lemma with a similar proof shows that a hyperplane o total least squares must pass through the centroid $\overline{\vec{\mathbf{x}}}$ defined by $\overline{x_j} := (1/n)\sum_{i=1}^{n} x_{i,j}$. Then as in the plane, a similar argument shows that the normal vector $\vec{\mathbf{r}}$ must be an eigenvecto corresponding to the smallest eigenvalue of the matrix $M^T M$, that is, the smallest singula value of M, with $\vec{\mathbf{x}}_i - \overline{\vec{\mathbf{x}}}$ in the ith row:

$$M := \begin{pmatrix} \vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}} \\ \vdots \\ \vec{\mathbf{x}}_n - \overline{\vec{\mathbf{x}}} \end{pmatrix} = \begin{pmatrix} x_{1,1} - \overline{x_1} & \cdots & x_{1,m} - \overline{x_m} \\ \vdots & \ddots & \vdots \\ x_{n,1} - \overline{x_1} & \cdots & x_{n,m} - \overline{x_m} \end{pmatrix}.$$

The observations just made lead to the following algorithm.

ALGORITHM.

Data. n points $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n$ in $\mathbb{R}^m$.

Step 1. Calculate the centroid, $\overline{\vec{\mathbf{x}}} := (1/n)\sum_{i=1}^{n} \vec{\mathbf{x}}_i$.

Step 2. Form the matrix $M \in \mathbb{M}_{n\times m}(\mathbb{R})$ with $\vec{\mathbf{x}}_i - \overline{\vec{\mathbf{x}}}$ in the ith row.

Step 3. Compute the smallest singular value of M, that is, the smallest eigenvalue of $M^T M$, and one eigenvector $\vec{\mathbf{r}}$ in the corresponding eigenspace of $M^T M$, for instance, with Jacobi's method.

Result. The hyperplane with equation $\langle \vec{\mathbf{x}} - \overline{\vec{\mathbf{x}}}, \vec{\mathbf{r}} \rangle = 0$ minimizes the sum D of squared distances from that hyperplane to the data points.

Remark. Of all numerical methods to estimate the smallest singular value, Jacobi's method appears to provide the best accuracy [2]. Moreover, especially at such an introductory level, the proof of the convergence of Jacobi's method [7], [11] is much more elementary than that of other methods, for instance, the QR algorithm of Francis [7], [11].

Extensions. Under some circumstances, the problem of total least squares may have multiple solutions or no solution. Multiple solutions arise if the smallest singular value has a multiplicity k greater than one. In such a situation, an infinite family of hyperplanes exists, indexed by the $(k-1)$-dimensional projective space $\mathbb{P}(\mathbb{R}^k)$, with each hyperplane giving the same minimum value for D. Such multiple solutions may mean that a linear model might not be adequate for the data at hand. In contrast, additional requirements, for instance, the selection of one of the variables as the "dependent" variable, may preclude the existence of any solution. For example, in the plane, the search for a line of total least squares with an equation of the form $y = mx + b$, which has normal vector $(r, s) = (m, -1)$, may fail if only one line of total least squares exists but has a horizontal normal vector $(r, s) = (r, 0)$. Such considerations, however, form the object of more advanced treatments, for instance, [15].

Solution to the exercise. The line of total least squares has equation $x - y = 0$, whereas the lines of ordinary least squares have equations $y = 9/20x + 11/5 = 0.45x + 2.2$ for y vs. x and $x = 0.45y + 2.2$, or $y = 20/9x - 44/9$, for x vs. y.

REFERENCES

[1] LARRY E. ASKINS, *Distance from a point to a line*, The Mathematics Teacher, 80 (1987), p. 608.

[2] J. DEMMEL AND K. VESELIĆ, *Jacobi's method is more accurate than* QR, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1204–1245.

[3] RICHARD J. DUFFIN, *Lines of best fit by graphics and the Wald line*, Amer. Math. Monthly, 98 (1991), pp. 835–839.

[4] GEORGE W. FISHER, *Review of reference* [15], in Math. Computation, 59 (1992), pp. 724–725.

[5] ABDUS SATTAR GAZDAR, *Distance from a point to a plane with a variation on the Pythagorean theorem*, College Math. J., 23 (1992), pp. 410–412.

[6] GENE H. GOLUB AND CHARLES F. VAN LOAN, *Matrix Computations*, 2nd ed., John Hopkins University Press, Baltimore, MD, 1989.

[7] GÜNTHER HÄMMERLIN AND KARL-HEINZ HOFFMAN, *Numerical Mathematics*, Springer-Verlag, New York, 1991.

[8] ROBERT V. HOGG AND ALLEN T. CRAIG, *Introduction to Mathematical Statistics*, 4th ed., Macmillan, New York, 1978.

[9] DAVID KINCAID AND WARD CHENEY, *Numerical Analysis: Mathematics of Scientific Computing*, Brooks/Cole, Pacific Grove, CA, 1991.

[10] BART DE MOOR, *Structured Total Least Squares and L_2 Approximation Problems*, Linear Algebra Appl., 188-189 (1993), pp. 163–205.

[11] JOSEF STOER AND ROLAND BULIRSCH, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1983.

[12] GILBERT STRANG, *Calculus*, Wellesley-Cambridge Press, Wellesley, MA, 1991.

[13] ———, *Linear Algebra and its Applications*, 3rd ed., Harcourt Brace Jovanovich, San Diego, CA, 1988.

[14] ———, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, Wellesley, MA, 1986.

[15] SABINE VAN HUFFEL AND JOOS VANDERWALLE, *The Total Least Squares Problem: Computational Aspects and Analysis*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1991.

[16] ANDRÉ L. YANDL, *Introduction to Mathematical Analysis for Business and Economics*, Brooks/Cole, Pacific Grove, CA, 1990.