PSYCHOMETRIKA—VOL. 56, NO. 4, 601–609
DECEMBER 1991

A GENERAL SOLUTION FOR A CLASS OF WEAKLY CONSTRAINED LINEAR REGRESSION PROBLEMS

JOS M. F. TEN BERGE

UNIVERSITY OF GRONINGEN

This paper contains a globally optimal solution for a class of functions composed of a linear regression function and a penalty function for the sum of squared regression weights. Global optimality is obtained from inequalities rather than from partial derivatives of a Lagrangian function. Applications arise in multidimensional scaling of symmetric or rectangular matrices of squared distances, in Procrustes analysis, and in ridge regression analysis. The similarity of existing solutions for these applications is explained by considering them as special cases of the general class of functions addressed.

Key words: squared distance scaling, Procrustes analysis, ridge regression.

Consider the problem of minimizing, for fixed $\alpha \geq 0$ and fixed δ, the function

$$\mathrm{f}(\mathbf{t}) = \|F\mathbf{t} - \boldsymbol{\phi}\|^2 + \alpha(\mathbf{t}'\mathbf{t} - \delta)^2, \tag{1}$$

where F is a given $n \times m$ $(m \leq n)$ matrix, $\boldsymbol{\phi}$ is a given n-vector, and $\mathbf{t}$ is an m-vector. This function consists of the sum of a linear regression function and a penalty function for the deviation of the squared length of $\mathbf{t}$ from δ. Although there is no constraint on $\mathbf{t}$ in (1), the function is called "weakly constrained," because the penalty function prohibits a relatively large departure of $\mathbf{t}'\mathbf{t}$ from δ.

The purpose of the present paper is to solve the general problem of minimizing $\mathrm{f}(\mathbf{t})$, and to demonstrate its applications in a variety of contexts. Specifically, it will be shown that certain problems in multidimensional scaling can be handled in terms of minimizing $\mathrm{f}(\mathbf{t})$ for certain values of α and δ. If, on the other hand, we let α tend to infinity for $\delta \geq 0$, then the penalty function must be zero if $\mathrm{f}(\mathbf{t})$ is to be minimized, and the problem in (1) can be interpreted as the constrained least squares regression problem of minimizing $\|F\mathbf{t} - \boldsymbol{\phi}\|^2$ subject to $\mathbf{t}'\mathbf{t} = \delta$. Applications of this problem can be found in Procrustes analysis and ridge regression analysis.

Although (1) is the conceptually most relevant starting point for this paper, even fuller generality can be obtained by redefining the problem. That is, let

$$F = VC^{1/2}U' \tag{2}$$

be a singular value decomposition of F, with $V'V = U'U = UU' = I_m$, and C diagonal with diagonal elements $c_1 \geq c_2 \geq \cdots \geq c_m \geq 0$. Then (1) can be written equivalently as the problem of minimizing

$$\mathrm{g}(\mathbf{w}) = \boldsymbol{\phi}'\boldsymbol{\phi} - 2\mathbf{w}'\mathbf{x} + \mathbf{w}'C\mathbf{w} + \alpha(\mathbf{w}'\mathbf{w} - \delta)^2, \tag{3}$$

where $\mathbf{w} \equiv U'\mathbf{t}$ and $\mathbf{x} \equiv U'F'\boldsymbol{\phi}$. Below, the minimum of (1) will be determined through the minimum of (3). The solution for the minimum of (3), to be given below, does not require C to be nonnegative definite. Therefore, this solution can be applied to the

The author is obliged to Henk Kiers and Willem Heiser for helpful comments.

Requests for reprints should be sent to Jos M. F. ten Berge, Vakgroep Psychologie, Rijksuniversiteit Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.

0033-3123/91/1200-90214$00.75/0

problem in (3) regardless of nonnegativity of C. The case where $\alpha = 0$ has a well-known closed-form solution. Accordingly, only the case $\alpha > 0$ will be treated in the sequel.

A General Solution for the Minimum of g(w)

Let q denote the multiplicity of the smallest diagonal element in C and let the function $\psi_s(b)$ be defined as

$$\psi_s(b) = \delta b - \left(\frac{b^2}{4\alpha}\right) - \sum_{i=1}^{s} \frac{x_i^2}{(c_i - b)}, \tag{4}$$

with first derivative

$$\psi_s'(b) = \delta - \left(\frac{b}{2\alpha}\right) - \sum_{i=1}^{s} \frac{x_i^2}{(c_i - b)^2}, \tag{5}$$

for $b \neq c_i$, $i = 1, \ldots, s$, and $s \leq m$. Define $r = m - q$ and partition C and the vectors $\mathbf{w}$ and $\mathbf{x}$ as

$$C = \left(\begin{array}{c|c} C_r & 0 \\ \hline 0 & C_q \end{array}\right); \quad \mathbf{w} = \left(\begin{array}{c} \mathbf{w}_r \\ \hline \mathbf{w}_q \end{array}\right); \quad \mathbf{x} = \left(\begin{array}{c} \mathbf{x}_r \\ \hline \mathbf{x}_q \end{array}\right), \tag{6}$$

where C_r contains $c_1, \ldots, c_r$ on the diagonal, and $\mathbf{w}_r$ and $\mathbf{x}_r$ contain the first r elements of $\mathbf{w}$ and $\mathbf{x}$, respectively. Note that $C_q = c_m I_q$. The solution for the minimum of g(w) reduces to three distinct cases:

Case 1. Not every element of $\mathbf{x}_q$ is zero. In this case g(w) is minimized by

$$\mathbf{w} = (C - b_0 I)^{-1}\mathbf{x}, \tag{7}$$

where b_o is the unique value of $b < c_m$ for which $\psi_m'(b) = 0$. This b_o can be found by a Bolzano search, for instance. Details will be given below.

Case 2. Every element of $\mathbf{x}_q$ is zero, and $\psi_r'(c_m) < 0$. In this case g(w) is minimized by

$$\mathbf{w} = \left(\begin{array}{c} \mathbf{w}_r \\ \hline \mathbf{w}_q \end{array}\right)$$

with $\mathbf{w}_q = 0$ and

$$\mathbf{w}_r = (C_r - b_o I_r)^{-1}\mathbf{x}_r, \tag{8}$$

where b_o is the unique value of $b < c_m$ for which $\psi_r'(b) = 0$. This b_o can again be determined by a Bolzano search.

Case 3. Every element of $\mathbf{x}_q$ is zero, and $\psi_r'(c_m) \geq 0$. In this case, $\mathbf{w}_r$ can again be computed as in (8), but this time with $b_o = c_m$. In addition, $\mathbf{w}_q$ should be determined arbitrarily, subject to the constraint that

$$\mathbf{w}_q'\mathbf{w}_q = \psi_r'(c_m). \tag{9}$$

This yields an infinite number of solutions if $q > 1$. A solution among others is

$$\mathbf{w}' = \left(\frac{x_1}{c_1 - c_m}, \ldots, \frac{x_r}{c_r - c_m}, (\psi_r'(c_m))^{1/2}, 0, \ldots, 0\right). \tag{10}$$

A Proof for Case 1

The proof for Case 1 is organized as follows. First, it will be shown that for every $\mathbf{w}$, $g(\mathbf{w}) \geq \boldsymbol{\phi}'\boldsymbol{\phi} + \psi_m(b)$, where $\boldsymbol{\phi}'\boldsymbol{\phi}$ is a known constant and $\psi_m(b)$ is a known function, see (4), of a real parameter $b < c_m$. Because the inequality holds for every $b < c_m$, it represents a family of lower bounds to $g(\mathbf{w})$.

Next, this family will be narrowed down to its best member, by considering only the $b_0 < c_m$ for which $\psi_m(b)$ is a maximum.

Finally, a vector $\mathbf{w}_0$ will be constructed for which this best lower bound is attained. The very fact that a lower bound is attained is sufficient for a global minimum, and will serve to justify the seemingly pointless developments in retrospect.

The proof starts by noting that, for every $b < c_m$, we have the inequality

$$\|(C - bI)^{1/2}\mathbf{w} - (C - bI)^{-1/2}\mathbf{x}\|^2 \geq 0, \tag{11}$$

and hence

$$\mathbf{w}'C\mathbf{w} - 2\mathbf{w}'\mathbf{x} \geq b\mathbf{w}'\mathbf{w} - \mathbf{x}'(C - bI)^{-1}\mathbf{x}. \tag{12}$$

Combining (3) and (12) yields

$$\begin{aligned} g(\mathbf{w}) &\geq \boldsymbol{\phi}'\boldsymbol{\phi} + b\mathbf{w}'\mathbf{w} - \mathbf{x}'(C - bI)^{-1}\mathbf{x} + \alpha(\mathbf{w}'\mathbf{w} - \delta)^2 \\ &= \boldsymbol{\phi}'\boldsymbol{\phi} + \alpha\left(\mathbf{w}'\mathbf{w} - \delta + \frac{b}{2\alpha}\right)^2 + \delta b - \frac{b^2}{4\alpha} - \mathbf{x}'(C - bI)^{-1}\mathbf{x} \equiv \hat{g}(\mathbf{w}). \end{aligned} \tag{13}$$

Clearly,

$$\hat{g}(\mathbf{w}) \geq \boldsymbol{\phi}'\boldsymbol{\phi} + \delta b - \frac{b^2}{4\alpha} - \mathbf{x}'(C - bI)^{-1}\mathbf{x} = \boldsymbol{\phi}'\boldsymbol{\phi} + \psi_m(b). \tag{14}$$

It follows that $g(\mathbf{w}) \geq \boldsymbol{\phi}'\boldsymbol{\phi} + \psi_m(b)$ for every $b < c_m$. It is sufficient for global minimality of $g(\mathbf{w})$ to have a $\mathbf{w}_o$ and a $b_o < c_m$ for which $g(\mathbf{w}_o) = \boldsymbol{\phi}'\boldsymbol{\phi} + \psi_m(b_o)$. Such a b_o must of course satisfy

$$\psi_m(b_o) = \max_{b < c_m} \psi_m(b). \tag{15}$$

This maximum occurs at a $b_o < c_m$ which satisfies $\psi_m'(b) = 0$; that is,

$$\psi_m'(b) = \delta - \frac{b}{2\alpha} - \mathbf{x}'(C - bI)^{-2}\mathbf{x} = 0. \tag{16}$$

It has been shown by Greenacre and Browne (1986) that, whenever $\mathbf{x}_q'\mathbf{x}_q > 0$, equations like (16) have a unique root $b_o < c_m$. If we take $b = b_o$ and

$$\mathbf{w} = \mathbf{w}_o = (C - b_oI)^{-1}\mathbf{x}, \tag{17}$$

then it is easy to see that both (13) and (14) hold as an equality, and that the global minimum has been attained. This completes the proof for Case 1. □

It should be noted that (7) is not only sufficient, but also necessary for the global minimum in Case 1. Specifically, once it has been established that $\phi'\phi + \psi_m(b_o)$ is a lower bound to g(w) that can be attained, it follows that the inequalities that lead to this lower bound must hold as equalities. If we require (13) and (14) to hold as equalities, then solution (7) follows uniquely. The fact that only $b < c_m$ have been considered in (11) and (12) does not detract from this uniqueness because stationary points of f(t), corresponding to $b > c_m$, cannot be globally optimal, as has been shown by Shapiro, see Greenacre and Browne (1986, p. 241). It can be concluded that (7) is the unique globally minimal solution for Case 1.

It should be noted that Case 1 implies that F in (1) is of full column rank. This can be seen from $F = VC^{1/2}U'$ and $\mathbf{x} = U'F'\phi$. If the smallest singular value $c_m^{1/2}$ is zero, then the last q columns of FU are zero, and hence the last q elements of x are zero, and Case 1 does not apply. The reverse, however, does not follow: We need not have Case 1 if F has full column rank, because ϕ can be orthogonal to the last columns of FU, even if these columns are nonzero. A demonstration of this possibility will be given below (Table 1).

A Proof for Case 2 and Case 3

The proof for Case 2 and Case 3 follows essentially the same logic as the proof for Case 1 given above. That is, a family of lower bounds to g(w) is derived and the best member of this family is shown to be attained for a certain vector w.

Both in Case 2 and Case 3, $\mathbf{x}_q = 0$. Define for $b \le c_m$ the $m \times m$ diagonal matrix

$$D_b = \left(\begin{array}{cccc|c} (c_1 - b) & & & & \\ & (c_2 - b) & & & \\ & & \ddots & & 0 \\ & & & (c_r - b) & \\ \hline & & 0 & & 0 \end{array}\right), \tag{18}$$

with generalized inverse D_b^-, obtained by inverting the r nonzero elements. Clearly, we can consider D_b in partitioned form as

$$D_b = \begin{pmatrix} (C_r - bI_r) & 0 \\ 0 & 0 \end{pmatrix}. \tag{19}$$

For every $b \le c_m$, we have the inequality

$$\|D_b^{1/2}\mathbf{w} - (D_b^-)^{1/2}\mathbf{x}\|^2 \ge 0, \tag{20}$$

and hence

$$\mathbf{w}_r'C_r\mathbf{w}_r - 2\mathbf{w}_r'\mathbf{x}_r \ge b\mathbf{w}_r'\mathbf{w}_r - \mathbf{x}'D_b^-\mathbf{x}. \tag{21}$$

Noting that $\mathbf{w}'C\mathbf{w} = \mathbf{w}_r'C_r\mathbf{w}_r + c_m\mathbf{w}_q'\mathbf{w}_q$ and that $\mathbf{w}_r'\mathbf{x}_r = \mathbf{w}'\mathbf{x}$, we have from (21),

$$\mathrm{g}(\mathbf{w}) \ge \phi'\phi + \alpha(\mathbf{w}'\mathbf{w} - \delta)^2 + b\mathbf{w}_r'\mathbf{w}_r + c_m\mathbf{w}_q'\mathbf{w}_q - \mathbf{x}'D_b^-\mathbf{x}. \tag{22}$$

From $b \le c_m$ it follows that

$$bw_r'w_r + c_m w_q'w_q \geq bw_r'w_r + bw_q'w_q = bw'w. \tag{23}$$

Combining (22) and (23) yields

$$g(w) \geq \phi'\phi + \alpha(w'w - \delta)^2 + bw'w - x'D_b^- x$$

$$= \phi'\phi + \alpha\left(w'w - \delta + \frac{b}{2\alpha}\right)^2 + \delta b - \frac{b^2}{4\alpha} - x'D_b^- x \equiv g^*(w), \tag{24}$$

(see (13)). Clearly,

$$g^*(w) \geq \phi'\phi + \delta b - \frac{b^2}{4\alpha} - x'D_b^- x = \phi'\phi + \psi_r(b). \tag{25}$$

It follows that $g(w) \geq \phi'\phi + \psi_r(b)$ for every $b \leq c_m$, when $x_q = 0$. This time we are interested in the $b \leq c_m$ that maximizes $\psi_r(b)$, to obtain equality.

Let t be the largest index ($t = 0, 1, \ldots, r$) such that $x_t \neq 0$ and $x_{t+1} = \cdots = x_r = 0$. There is a unique $b_o < c_t$ (if $t = 0$ then $b_o = 2\alpha\delta$), that maximizes $\psi_r(b)$ and satisfies

$$\psi_r'(b) = \delta - \frac{b}{2\alpha} - x'(D_b^-)^2 x = 0. \tag{26}$$

This b_o satisfies $b < c_m$ if and only if $\psi_r'(c_m) < 0$, as in Case 2. Therefore, in Case 2 we define b_o as the unique root of (26) for $b < c_m$, and we take w_o according to (8). It can be verified that this yields $g(w_o) = \phi'\phi + \psi_r(b_o)$, which implies that the unique global minimum for $g(w)$ has been obtained. Specifically, using (8) we have $w_q = 0$, which implies that (23) holds as an equality, and we have (20) holding as an equality. From (8) and (26) it is clear that

$$w'w = w_r'w_r = \sum_{i=1}^{r} \frac{x_i^2}{(c_i - b_o)^2} = \delta - \frac{b}{2\alpha}, \tag{27}$$

which shows that (25) also holds as an equality. Because every inequality which led to $g(w) \geq \phi'\phi + \psi_r(b)$ holds as an equality, it has been shown that (8) yields the global minimum for $g(w)$ in Case 2. Uniqueness follows as it did in Case 1.

Finally, consider Case 3, where $\psi_r'(c_m) \geq 0$. Here the maximum of $\psi_r(b)$ for $b \leq c_m$ occurs at $b = c_m$. Accordingly, we take $b_o = c_m$ to obtain w_r from (8) and we construct a w_q that satisfies (9). Then we have equality in (20). Although now $w_q \neq 0$, equality in (23) in still guaranteed because $b_o = c_m$. Finally, we have a w_o satisfying

$$w_o'w_o = w_r'w_r + w_q'w_q = \sum_{i=1}^{r} \frac{x_i^2}{(c_i - c_m)^2} + \psi_r'(c_m) = \delta - \frac{b}{2\alpha}, \tag{28}$$

which shows that (25) holds as an equality. It can be concluded that the solution for Case 3 is globally optimal, albeit nonunique. A computational example for Case 3, with $t = 1$, $r = 2$ and rank $(C) = 3$ can be found in Table 1. □

TABLE 1

A Computational Example for Case 3 ($\alpha = 6$; $\delta = \frac{2}{3}$)

$$F = \begin{bmatrix} 4 & 1 & .5 \\ 4 & -1 & -.5 \\ -4 & 1 & -.5 \\ -4 & -1 & .5 \\ 2 & 0 & 0 \\ -2 & 0 & 0 \end{bmatrix}; \quad \phi = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ 1 \end{bmatrix}; \quad C = \begin{bmatrix} 72 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} -4 \\ 0 \\ 0 \end{bmatrix}$$

$(q = 1;\ r = 2;\ t = 1)$

$$\psi_2'(1) = \frac{2}{3} - \frac{1}{12} - \frac{16}{71^2} = \frac{35095}{60492} > 0 \rightarrow \text{Case 3}$$

$$\mathbf{w}_o = \begin{bmatrix} -4/71 \\ 0 \\ \pm\sqrt{\psi_2'(1)} \end{bmatrix}; \quad \mathbf{w}_o'\mathbf{w}_o = \frac{7}{12}; \quad g(\mathbf{w}_o) = 6 + \frac{2}{3} - \frac{1}{24} - \frac{16}{71} = 6.399648.$$

Applications

Unfolding

Greenacre and Browne (1986) suggested an alternating least-squares method for fitting squared distances in multidimensional unfolding. The heart of their method consists of minimizing the squared distance function

$$\mathrm{h}(\mathbf{t}) = \|(\mathbf{d}^2 - \mathbf{h}) - \mathbf{t}'\mathbf{t}\mathbf{1} - 2Y\mathbf{t}\|^2, \tag{29}$$

for given n-vectors $\mathbf{d}^2$ and $\mathbf{h}$ and a given $n \times m$ matrix Y satisfying $\mathbf{1}'Y = 0'$. Although $\mathrm{h}(\mathbf{t})$ is not at all similar to (1), it is easy to redefine $\mathrm{h}(\mathbf{t})$ in a form that does resemble (1). That is, let $\boldsymbol{\phi} \equiv (\mathbf{h} - \mathbf{d}^2)$, $F \equiv 2Y$, and $\delta \equiv -n^{-1}\mathbf{1}'\boldsymbol{\phi}$. Then $\mathrm{h}(\mathbf{t})$ can be written as

$$\mathrm{h}(\mathbf{t}) = \|F\mathbf{t} - \boldsymbol{\phi} - \mathbf{t}'\mathbf{t}\mathbf{1}\|^2 = \|F\mathbf{t} - \boldsymbol{\phi}\|^2 + n(\mathbf{t}'\mathbf{t})^2 + 2(\mathbf{1}'\boldsymbol{\phi})(\mathbf{t}'\mathbf{t})$$
$$= \|F\mathbf{t} - \boldsymbol{\phi}\|^2 + n(\mathbf{t}'\mathbf{t} - \delta)^2 - n\delta^2, \tag{30}$$

which is a special case of (1), with $\alpha = n$, if the constant $n\delta^2$ is ignored. Greenacre and Browne (1986) solved the problem of minimizing $\mathrm{h}(\mathbf{t})$ in Case 1. For all practical purposes this seems sufficient because the chances that other cases occur seems highly remote in their method.

Multidimensional Scaling

A similar application of minimizing $\mathrm{f}(\mathbf{t})$ also occurs in multidimensional scaling, although this application seems to have gone unnoticed. That is, let V be a symmetric

$n \times n$ matrix, with zero diagonal, the elements of which are regarded as approximate squared distances between pairs of points in Euclidean space. Browne (1987) considered the problem of minimizing

$$\mathrm{h}(A) = \|V - \mathbf{h}\mathbf{1}_n' - \mathbf{1}_n\mathbf{h}' + 2AA'\|^2, \tag{31}$$

where A is an $n \times m$ matrix and the vector $\mathbf{h}$ contains the diagonal elements of AA'. Although Browne has considered various iterative methods for minimizing (31), none of these seem to be based on alternating least squares by updating the rows of A iteratively. Such a straightforward procedure arises quite naturally upon decomposition of $\mathrm{h}(A)$ in a function of any row $\mathbf{a}_i'$ of A and a constant term with respect to $\mathbf{a}_i$. For example, if we take $i = 1$ and define A_1 as the $(n - 1) \times m$ matrix obtained by deleting the first row of A, then that part of $\mathrm{h}(A)$ that varies with $\mathbf{a}_1$ can we written as

$$\mathrm{h}_1(\mathbf{a}_1) = (v_{11} - \mathbf{a}_1'\mathbf{a}_1 - \mathbf{a}_1'\mathbf{a}_1 + 2\mathbf{a}_1'\mathbf{a}_1)^2 + 2\|\mathbf{v}_1 - \mathbf{u}_1 - \mathbf{a}_1'\mathbf{a}_1\mathbf{1}_{n-1} + 2A_1\mathbf{a}_1\|^2, \tag{32}$$

where $\mathbf{v}_1$ is the $n - 1$ vector containing the off-diagonal elements of the first column (and row) of V, and $\mathbf{u}_1 = (\mathbf{a}_2'\mathbf{a}_2, \ldots, \mathbf{a}_n'\mathbf{a}_n)'$. Clearly, the first term of $\mathrm{h}_1(\mathbf{a}_1)$ is constant. Upon centering A_1 columnwise (which does not affect the distances), it remains to minimize

$$\tfrac{1}{2}\mathrm{h}_1(\mathbf{a}_1) = \|2A_1\mathbf{a}_1 - \boldsymbol{\phi}\|^2 + (n - 1)(\mathbf{a}_1'\mathbf{a}_1)^2 + 2(\mathbf{a}_1'\mathbf{a}_1)\boldsymbol{\phi}'\mathbf{1}_{n-1}, \tag{33}$$

where $\boldsymbol{\phi} \equiv (\mathbf{u}_1 - \mathbf{v}_1)$. This function can be written as

$$\tfrac{1}{2}\mathrm{h}_1(\mathbf{a}_1) = \|F\mathbf{a}_1 - \boldsymbol{\phi}\|^2 + (n - 1)(\mathbf{a}_1'\mathbf{a}_1 - \delta)^2 - (n - 1)\delta^2, \tag{34}$$

with $F = 2A_1$ and $\delta = -\boldsymbol{\phi}'\mathbf{1}_{n-1}/(n - 1)$. It is obvious that (34) can be considered a special case of (1) with $\alpha = (n - 1)$, ignoring the constant $-(n - 1)\delta^2$. Because every row of A can be optimized, keeping the other rows fixed, an alternating least squares method for minimizing (31) can readily be obtained. It is not claimed that such an algorithm would be more efficient than any of the algorithms considered by Browne. We merely wish to demonstrate the applicability of (1) in the context of least-squares squared distance scaling.

Procrustes analysis

Gower (1984) discussed a predecessor of the Greenacre and Browne method (Greenacre, 1978; see Case 1 above), at a time when the global minimality had not yet been established (a gap to be filled by Shapiro; see Greenacre & Browne, 1986, p. 241), but was merely conjectured (rightly) to be associated with a root of (16) smaller than c_m, the smallest eigenvalue in C. Gower expressed his conjecture as follows: "In problems of this kind it is usual for the smallest root $[b_o]$ to correspond to the smallest residual sum of squares, see e.g. oblique Procrustes analysis" (p. 758). Gower's suspicion of the similarity between the problem he was discussing and oblique Procrustes analysis will now be corroborated by showing that oblique Procrustes analysis is also a special case of (1). That is, if $\delta > 0$ and we let α tend to infinity, then the penalty function $\alpha(\mathbf{t}'\mathbf{t} - \delta)^2$ dominates $\mathrm{f}(\mathbf{t})$, and minimizing $\mathrm{f}(\mathbf{t})$ unconstrained becomes equivalent to minimizing $\|F\mathbf{t} - \boldsymbol{\phi}\|^2$ subject to $\mathbf{t}'\mathbf{t} = \delta$. Such a problem has been considered by Browne (1967) and ten Berge and Nevels (1977), among others, for $\delta = 1$. The ten Berge and Nevels solution is exactly the same as the solution of the present paper, if we choose $\alpha = \infty$ and $\delta = 1$. This shows the applicability of (1) in contexts different from multidimensional scaling, and serves to unify approaches that might seem unrelated. It is important to note that Case 2 and Case 3, which seem irrelevant for vector re-estimation problems, do play a role in oblique Procrustes analysis, where $\boldsymbol{\phi}$ may have structural zero elements; see ten Berge and Nevels (p. 597) for an example.

An Algorithm for Minimizing f(t)

On the basis of (7), (8), and (10), the following algorithm for minimizing f(t) seems appropriate, starting from a known $n \times m$ predictor matrix F, an n-vector $\boldsymbol{\phi}$, and parameters α and δ.

1. Compute U and C from the eigendecomposition $F'F = UCU'$. Evaluate q, the multiplicity of the smallest eigenvalue c_m in C.
2. Compute $\mathbf{x} = U'F'\boldsymbol{\phi}$, and $\mathbf{x}_q$, the vector containing the last q elements of $\mathbf{x}$.

3a. If $\mathbf{x}_q'\mathbf{x}_q > 0$, then find the unique $b_0 < c_m$ for which

$$\psi_m'(b) = \delta - (b/2\alpha) - \sum_{i=1}^{m} \frac{x_i^2}{(c_i - b)^2} = 0. \tag{35}$$

Although Newton's method is often recommended to find such a root, it may wander off to a $b > c_m$ unless it starts close to the root desired. For this reason, it is safer to use a Bolzano search. This method consists of iteratively deleting the left or right-hand half of an interval, depending on whether the derivative is positive or negative at the midpoint, respectively. For $\delta > 0$ it is obvious that $-\mathbf{x}'C^{-1}\mathbf{x} = \psi_m(0) \leq \psi_m(b_0) \leq \delta b_0$, so we start with the interval $[-\delta^{-1}\mathbf{x}'C^{-1}\mathbf{x}, c_m]$. For $\delta = 0$, we start with the interval $[-2(\alpha\mathbf{x}'C^{-1}\mathbf{x})^{1/2}, c_m]$. The derivative to be evaluated at the midpoint is given in (35). After convergence, compute $\mathbf{w}_0$ by (7).

3b. If $\mathbf{x}_q'\mathbf{x}_q = 0$, set $r = m - q$ and evaluate

$$\psi_r'(c_m) = \delta - (c_m/2\alpha) - \sum_{i=1}^{r} \frac{x_i^2}{(c_i - c_m)^2}. \tag{36}$$

If this is nonnegative, compute $\mathbf{w}_0$ by (10). Else, do a Bolzano search to find the unique $b_0 < c_m$ for which

$$\psi_r'(b) = \delta - (b/2\alpha) - \sum_{i=1}^{r} \frac{x_i^2}{(c_i - b)^2} = 0. \tag{37}$$

The Bolzano search may be started with the interval $[-\delta^{-1}\mathbf{x}_r'C_r^-\mathbf{x}_r, c_m]$ if $\delta > 0$, and with $[-2(\alpha\mathbf{x}_r'C_r^-\mathbf{x}_r)^{1/2}, c_m]$ otherwise. After convergence, obtain $\mathbf{w}_0$ by (8).

4. Compute $\mathbf{t}_0 = U\mathbf{w}_0$ as the minimizing $\mathbf{t}$ of (1).

Discussion

The weakly constrained regression problem of this paper has been shown to have a, typically unique, global minimum. We have used a completing-the-squares type approach, rather than calculus, because it yields global minimality rather easily for Case 2 and Case 3. For Case 1 alone, the calculus solution given by Greenacre and Browne (1986) would seem to be more efficient.

The regression problem has been shown to have various applications, and more applications are forthcoming. In addition, our treatment serves to unify apparently different approaches, by allowing the parameter α to tend to infinity.

References

Browne, M. W. (1967). On oblique Procrustes rotation. *Psychometrika*, *32*, 125–132.

Browne, M. W. (1987). The Young-Householder algorithm and the least-squares multidimensional scaling of squared distances. *Journal of Classification*, *4*, 175–190.

Gower, J. C. (1984). Multivariate analysis: Ordination, multidimensional scaling and allied topics. In E. Lloyd (Ed.), *Handbook of applicable mathematics*—volume 6 (pp. 727–781). New York: Wiley.

Greenacre, M. J. (1978). *Some objective methods of graphical display of a data matrix* (Translation of doctoral thesis, Université de Paris, VI). Pretoria: University of South Africa.

Greenacre, M. J., & Browne, M. W. (1986). An efficient alternating least-squares algorithm to perform multidimensional unfolding. *Psychometrika*, *51*, 241–250.

ten Berge, J. M. F., & Nevels, K. (1977). A general solution to Mosier's oblique Procrustes problem. *Psychometrika*, *42*, 593–600.

Manuscript received 2/16/90
Final version received 10/22/90