A Study of the Stationary Configurations of the SStress Criterion for Metric Multidimensional Scaling

Samuel W. Malone* Michael W. Trosset†

January 30, 2000

Abstract

It is widely believed that both the stress and the sstress criteria for metric multidimensional scaling are plagued by the existence of nonglobal minimizers. At present, there is little theory that enlightens this belief. Trosset and Mathar (1997) established that nonglobal minimizers of the stress criterion can exist, while Glunt, Hayden, and Liu (1991) demonstrated that the distance matrices of all configurations for which the gradient of the sstress criterion vanishes lie on a certain sphere. This report extends existing theory in several directions. Emphasis is placed on the more tractable case of the sstress criterion. Because the stress and sstress criteria must be minimized by numerical optimization, one result that is of immediate practical value is a simple device for improving the quality of the initial configurations from which optimization commences.

Contents

*Department of Mathematics, Duke University, Durham, NC 27708-0586 (e-mail: `swm3@duke.edu`).

†Associate Professor Department of Mathematics, College of William & Mary, P.O. Box 8795, Williamsburg, VA 23187-8795 (e-mail: `trosset@math.wm.edu`); and Adjunct Associate Professor, Department of Computational & Applied Mathematics, Rice University, Houston, TX 77005-1892 (e-mail: `trosset@caam.rice.edu`).

1 Introduction

Multidimensional scaling (MDS) is a general term for techniques that construct configurations of points in a target metric space from information about interpoint distances. In the case of two-way MDS, the information is specified in the form of a *dissimilarity matrix*, i.e. a matrix $\Delta = (\delta_{ij})$ such that $\delta_{ij} \geq 0$, $\delta_{ii} = 0$, and $\delta_{ij} = \delta_{ji}$. Typically, the target metric space is p-dimensional Euclidean space. In most applications, such as the reconstruction of molecular configurations from information about interatomic distances ($p = 3$), the target dimension is small.

For a configuration of points $x_1, \ldots, x_n \in \Re^p$, the $n \times p$ *configuration matrix* X is the matrix whose rows are the x_i'. From X it is easy to compute the Euclidean interpoint distance matrix $D(X) = (d_{ij})$. The goal of metric two-way MDS is to construct a configuration matrix for which the interpoint distances d_{ij} approximate the dissimilarities δ_{ij}.

Two popular ways of measuring the discrepancy between a distance matrix and a dissimilarity matrix are the *stress* and *sstress* criteria. The former is based on the squared errors between the distances and dissimilarities; the latter is based on the squared errors between the squared distances and squared dissimilarities. Let

$$\rho_r(D, \Delta) = \sum_{ij} w_{ij} \left[(d_{ij})^r - (\delta_{ij})^r \right]^2 ,$$

where the w_{ij} are nonnegative weights. Write $D = D(X)$ and let $\sigma_r(X) = \rho_r(D(X))$; then σ_1 is the metric stress criterion and σ_2 is the metric sstress critrion.

In practice, one often sets each $w_{ij} = 1$. This is the only case considered in this report. However, one can use the weights either to accommodate missing data (by setting the appropriate $w_{ij} = 0$) or to weight more reliably measured dissimilarities more heavily.

The remaining sections of this report consider several issues related to the problem of minimizing σ_r, with particular emphasis on the case of $r = 2$. In Section 2 we reprise the example constructed by Trosset and Mathar (1997) to demonstrate that the stress criterion can have a nonglobal minimizer. Somewhat to our surprise, this construction produces a saddle point of the sstress criterion, but not a minimizer. In Section 3 we extend the critical point theorem of Glunt, Hayden, and Liu (1991) from the special case of σ_2 to general σ_r. We also establish several new properties. In particular, we derive an explicit formula for the optimal dilation of a fixed configuration matrix. In Section 4 we investigate the relation between classical MDS and the problem of minimizing the sstress criterion. We focus on the use of the classical solution as an initial configuration from which to begin optimizing σ_2 and consider the extent to which optimally dilating the classical solution improves the initial configuration. Section 5 concludes with some remarks on future research.

2 The Search for Nonglobal Minimizers

Because most algorithms for minimizing σ_r are designed to find local minimizers, the existence and prevalence of nonglobal minimizers is of considerable importance. Trosset and Mathar (1997) demonstrated the existence of a nonglobal minimizer of stress, but we are not aware that anyone has formally demonstrated the existence of nonglobal minimizers of sstress. In this section, we show that the example constructed by Trosset and Mathar (1997) for stress does not produce a nonglobal minimizer of sstress. Indeed, we have been unable to construct a nonglobal minimizer of sstress.

Following Trosset and Mathar (1997), let

$$X^* = \begin{bmatrix} 0 & 0 \\ x_1 = 1 & 0 \\ x_2 = 1 & x_3 = 1 \\ x_4 = 0 & x_5 = 1 \end{bmatrix}.$$

The configuration X^* represents the unit square, with the vertices labeled counterclockwise from the origin. The matrix of dissimilarities

$$\Delta_1 = D_1\left(X^*\right) = \begin{bmatrix} 0 & 1 & \sqrt{2} & 1 \\ 1 & 0 & 1 & \sqrt{2} \\ \sqrt{2} & 1 & 0 & 1 \\ 1 & \sqrt{2} & 1 & 0 \end{bmatrix}$$

produces a global minimum of $\sigma_1(X^*) = 0$.

We now define an initial configuration

$$X^0 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

by relabelling the upper two vertices of the unit square. Trosset and Mathar (1997) reasoned that the upper vertices would have to cross in order to converge to the global minimizer, but that crossing would increase stress before decreasing it. Indeed, they demonstrated that the configuration

$$\bar{X} = \begin{bmatrix} 0 & 0 \\ \bar{b} & 0 \\ 0 & \bar{c} \\ \bar{b} & \bar{c} \end{bmatrix},$$

where $\bar{b} = (3+\sqrt{3})/6$ and $\bar{c} = \bar{b}\sqrt{2}$, is a nonglobal minimizer of σ_1.

We attempted an analogous construction for sstress. The matrix of squared dissimilarities

$$\Delta_2 = D_2\left(X^*\right) = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix}$$

produces a global minimum of $\sigma_2(X^*) = 0$. We hypothesized the existence of a nonglobal minimizer of the form of $\bar{X}$; hence, for $b, c \in \Re$, let

$$X(b,c) = \begin{bmatrix} 0 & 0 \\ b & 0 \\ 0 & c \\ b & c \end{bmatrix}.$$

Then

$$\begin{aligned} \sigma_2\left(X(b,c)\right) &= \left\|D_2\left(X(b,c)\right) - \Delta_2\right\|_F^2 \\ &= \left\|D_2\left(X(b,c)\right) - D_2\left(X^*\right)\right\|_F^2 \\ &= 4\left(b^2-1\right)^2 + 4\left(c^2-2\right)^2 + 4\left(c^2+b^2-1\right)^2, \end{aligned}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. A nontrivial solution of the stationary equation

$$\nabla \sigma_2 \left(X(b,c)\right) = \left[\begin{array}{c} 16b\left(2b^2 - 2 + c^2\right) \\ 16c\left(2c^2 - 3 + b^2\right) \end{array} \right] = 0$$

is given by $\hat{b} = \sqrt{3}/3$ and $\hat{c} = 2\hat{b}$.

The Hessian matrix of $\sigma_2(X(\hat{b},\hat{c}))$ is positive definite (its eigenvalues are approximately 91.7925 and 14.874), so $\hat{X} = X(\hat{b},\hat{c})$ is a local minimizer of $\sigma_2(X(b,c))$, a function of two real variables. We hoped to find that

$$\hat{X} = \left[\begin{array}{cc} 0 & 0 \\ x_1 = \hat{b} & 0 \\ x_2 = 0 & x_3 = \hat{c} \\ x_4 = \hat{b} & x_5 = \hat{c} \end{array} \right]$$

is also a local minimizer of $\sigma_2(X(x_1,\ldots,x_5))$, a function of five real variables. In fact, $\hat{X}$ *is* a stationary configuration ($\nabla\sigma_2(\hat{X}) = 0$), but it is *not* a local minimizer: the eigenvalues of the Hessian matrix are approximately

$$(48.4053, 37.3879, 10.2383, 1.6043, -6.9689)'.$$

Somewhat surprisingly, the construction of Trosset and Mathar (1997), which produces a nonlocal minimizer of the stress criterion, does not produce a nonlocal minimizer of the sstress criterion. Nor have we discovered another construction that does—the existence of nonglobal minimizers of the sstress criterion remains an open question. The relative ease with which Trosset and Mathar (1997) discovered a nonlocal minimizer of the stress criterion reinforces the popular belief that σ_1 tends to have more nonlocal minimizers than does σ_2.

3 Properties of Stationary Configurations

Glunt, Hayden, and Liu (1991) established the remarkable result that, for any fixed dissimilarity matrix, all of the interpoint distance matrices generated by stationary configurations of the sstress criterion with unit weights lie on the surface of a sphere. We proceed to extend this result from the special case of $r = 2$ to the general case of $r > 0$. Throughout this section, $\langle\cdot,\cdot\rangle$ will denote the Frobenius inner product and $\|\cdot\|$ will denote the Frobenius norm.

Let $\Delta = (\delta_{ij})$ denote a fixed dissimilarity matrix. Given a configuration matrix X, let $D(X) = (d_{ij})$ denote the corresponding matrix of interpoint Euclidean distances. For $r > 0$, let $\Delta_r = (\delta_{ij}^r)$ and let $D_r(X) = (d_{ij}^r)$. We are interested in stationary configurations for the error criterion

$$\sigma_r(X) = \|D_r(X) - \Delta_r\|^2,$$

i.e. in configuration matrices $\bar{X}$ for which $\nabla\sigma_r(\bar{X}) = 0$.

Theorem 1 *(Generalized Critical Point Theorem)* *If* $\nabla\sigma_r(\bar{X}) = 0$, *then*

(i) $\langle D_r(\bar{X}), \Delta_r\rangle = \|D_r(\bar{X})\|^2$;

(ii) $\|D_r(\bar{X})\|^2 + \|D_r(\bar{X}) - \Delta_r\|^2 = \|\Delta_r\|^2$;

(iii) If $\nabla\sigma_r(\bar{Y}) = 0$, *then* $\sigma_r(\bar{X}) < \sigma_r(\bar{Y})$ *if and only if* $\|D_r(\bar{Y})\| < \|D_r(\bar{X})\|$;

(iv) $\|D_r(\bar{X}) - \Delta_r/2\| = \|\Delta_r/2\|$.

Proof: We introduce an arc in the space of configuration matrices,

$$\begin{aligned} \alpha(t) &= \sigma_r\left(t\bar{X}\right) = \left\| D_r\left(t\bar{X}\right) - \Delta_r \right\|^2 = \left\| t^r D_r\left(\bar{X}\right) - \Delta_r \right\|^2 \\ &= t^{2r} \left\| D_r\left(\bar{X}\right) \right\|^2 - 2t^r \left\langle D_r\left(\bar{X}\right), \Delta_r \right\rangle + \left\| \Delta_r \right\|^2 . \end{aligned}$$

Differentiating, we obtain

$$\alpha'(t) = 2rt^{2r-1} \left\| D_r\left(\bar{X}\right) \right\|^2 - 2rt^{r-1} \left\langle D_r\left(\bar{X}\right), \Delta_r \right\rangle .$$

Because $\nabla\sigma_r(\bar{X}) = 0$, we have $\alpha'(1) = 0$ and therefore

$$\left\langle D_r\left(\bar{X}\right), \Delta_r \right\rangle = \left\| D_r\left(\bar{X}\right) \right\|^2 ,$$

which is (i).

Next, applying (i), we observe that

$$\left\| D_r\left(\bar{X}\right) - \Delta_r \right\|^2 = \left\| D_r\left(\bar{X}\right) \right\|^2 - 2\left\langle D_r\left(\bar{X}\right), \Delta_r \right\rangle + \left\| \Delta_r \right\|^2 = \left\| \Delta_r \right\|^2 - \left\| D_r\left(\bar{X}\right) \right\|^2 .$$

Rearranging terms produces (ii).

Now suppose that $\nabla\sigma_r(\bar{X}) = \nabla\sigma_r(\bar{Y}) = 0$. Applying (ii), we obtain

$$\begin{aligned} \sigma_r\left(\bar{X}\right) - \sigma_r\left(\bar{Y}\right) &= \left\| D_r\left(\bar{X}\right) - \Delta_r \right\|^2 - \left\| D_r\left(\bar{Y}\right) - \Delta_r \right\|^2 \\ &= \left[\left\| \Delta_r \right\|^2 - \left\| D_r\left(\bar{X}\right) \right\|^2 \right] - \left[\left\| \Delta_r \right\|^2 - \left\| D_r\left(\bar{Y}\right) \right\|^2 \right] \\ &= \left\| D_r\left(\bar{Y}\right) \right\|^2 - \left\| D_r\left(\bar{X}\right) \right\|^2 , \end{aligned}$$

from which (iii) follows immediately.

Finally,

$$\left\| D_r\left(\bar{X}\right) - \Delta_r/2 \right\|^2 = \left\| D_r\left(\bar{X}\right) \right\|^2 - \left\langle D_r\left(\bar{X}\right), \Delta_r \right\rangle + \left\| \Delta_r/2 \right\|^2 = \left\| \Delta_r/2 \right\|^2 ,$$

which is (iv). $\square$

Part (iv) of Theorem 1 states that, for any fixed dissimilarity matrix Δ, the interpoint distance matrices of all stationary configurations lie on the same sphere of radius $\|\Delta/2\|$, centered at $\Delta/2$. Theorem 1 does not address the existence of nonglobal minimizers, but part (iii) states that global minimizers are those stationary configurations of maximal norm. This is a quite remarkable fact—one that would seem to have profound implications for global optimization of σ_r.

Our next result was motivated by Theorem 1. Suppose that a configuration matrix X has been proposed as a possible minimizer of σ_r. (For example, suppose that X has been proposed as the initial configuration from which an iterative optimization method will start.) If (i) in Theorem 1 is not satisfied, then X cannot be a minimizer of σ_r. The following theorem states that, by dilating X so that (i) is satisfied, we necessarily decrease σ_r.

Theorem 2 *(Dilation Theorem) Let Δ be a fixed dissimilarity matrix, let X be a fixed configuration matrix for which $\|D(X)\| > 0$, and consider the function $\alpha : \Re \to \Re$ defined by $\alpha(t) = \sigma_r(tX)$. Let*

$$t^* = \left(\frac{\langle D_r(x), \Delta_r \rangle}{\|D_r(X)\|^2} \right)^{1/r} .$$

Then

(i) $D_r(t^*X)$ lies on the sphere described by (iv) in Theorem 1; and

(ii) t^* is a global mimimizer of α.

Proof: To establish (i), we compute

$$\begin{aligned}\|D_r(t^*X)-\Delta_r/2\|^2 &= \|D_r(t^*X)\|^2-\langle D_r(t^*X),\Delta_r\rangle+\|\Delta_r/2\|^2\\ &= (t^*)^{2r}\|D_r(X)\|^2-(t^*)^r\langle D_r(X),\Delta_r\rangle+\|\Delta_r/2\|^2\\ &= \frac{\langle D_r(X),\Delta_r\rangle^2}{\|D_r(X)\|^2}-\frac{\langle D_r(X),\Delta_r\rangle^2}{\|D_r(X)\|^2}+\|\Delta_r/2\|^2\\ &= \|\Delta_r/2\|^2,\end{aligned}$$

which is (iv) in Theorem 1.

To minimize

$$\begin{aligned}\alpha(t) &= \sigma_r(tX)=\|D_r(tX)-\Delta_r\|^2=\|t^rD_r(X)-\Delta_r\|^2\\ &= t^{2r}\|D_r(X)\|^2-2t^r\langle D_r(X),\Delta_r\rangle+\|\Delta_r/2\|^2,\end{aligned}$$

we first note that $\alpha(t)\to\infty$ as $t\to\pm\infty$. Hence, it suffices to consider the stationary points of α, i.e. the values of t at which

$$\alpha'(t)=2rt^{2r-1}\|D_r(X)\|^2-2rt^{r-1}\langle D_r(X),\Delta_r\rangle$$

vanishes. By inspection $\alpha'(t)=0$ if and only if either $t=0$ or

$$t^r=\frac{\langle D_r(x),\Delta_r\rangle}{\|D_r(X)\|^2}=(t^*)^r,$$

and

$$\alpha(t^*)=\frac{\langle D_r(X),\Delta_r\rangle^2}{\|D_r(X)\|^2}-2\frac{\langle D_r(X),\Delta_r\rangle^2}{\|D_r(X)\|^2}+\|\Delta_r\|^2<\|\Delta_r\|^2=\alpha(0).$$

Noting that $\alpha(t)$ depends on t only through t^r, we conclude that t^* is a global minimizer of α. □

Now we specialize to $r=2$. The remainder of this section develops some additional theory that we hope will lead to a better understanding of the stationary configurations of σ_2, the sstress criterion. We require the following lemma.

Lemma 1 *Suppose that $p(t)=t^4+bt^3+ct^2+dt+e$, a quartic polynomial in $t\in\Re$. Then*

(i) If $p'(0)=0$, then $d=0$;

(ii) If $p'(0)=p'(1)=0$ and $p(0)=p(1)$, then $b=-2$ and $c=1$.

Proof: Differentiating, we obtain

$$p'(t)=4t^3+3bt^2+2ct+d$$

and $p'(0)=d$, hence (i). If, in addition,

$$0=p'(1)=4+3b+2c+d=4+3b+2c \tag{1}$$

and $p(0)=p(1)$, hence,

$$e=p(0)=p(1)=1+b+c+d+e=1+b+c+e$$

and therefore

$$0=1+b+c, \tag{2}$$

then the linear system defined by (1) and (2) has the unique solution $b=-2$ and $c=1$, which is (ii). □

Using Lemma 1, we proceed to investigate the behavior of σ_2 along the line segment that connects two stationary configuration matrices.

Theorem 3 *Let Δ be a fixed dissimilarity matrix and suppose that $\nabla\sigma_2(X) = \nabla\sigma_2(Y) = 0$. Then*

$$\langle D_2(X), H\rangle = \langle D_2(Y), H\rangle = \langle \Delta_2, H\rangle, \tag{3}$$

where

$$H = \frac{1}{2}\left[D_2(X+Y) - D_2(X) - D_2(Y)\right]. \tag{4}$$

Furthermore, if $\sigma_2(X) = \sigma_2(Y)$, then

$$\langle D_2(X) - D_2(Y), D_2(tX + (1-t)Y)\rangle = (2t-1)\left[\|D_2(X)\|^2 - \langle D_2(X), D_2(Y)\rangle\right], \tag{5}$$

for every $t \in [0,1]$, and

$$\left\langle D_2(X), D_2\left(\frac{X+Y}{2}\right)\right\rangle = \left\langle D_2(Y), D_2\left(\frac{X+Y}{2}\right)\right\rangle. \tag{6}$$

Proof: First, define $H = (h_{ij})$ by

$$h_{ij} = \sum_{k=1}^{p} (x_{ik} - x_{jk})(y_{ik} - y_{jk}).$$

Then

$$D_2(tX + (1-t)Y) = t^2 D_2(X) + 2t(1-t)H + (1-t)^2 D_2(Y)$$

and therefore

$$H = \frac{1}{2t(1-t)}\left[D_2(tX + (1-t)Y) - t^2 D_2(X) - (1-t)^2 D_2(Y)\right]. \tag{7}$$

Choosing $t = 1/2$, we obtain (4).

Now we study the quartic polynomial

$$\begin{aligned}
p(t) &= \sigma_2(tX + (1-t)Y) \\
&= \|D_2(tX + (1-t)Y) - \Delta_2\|^2 \\
&= \left\|t^2 D_2(X) + 2t(1-t)H + (1-t)^2 D_2(Y) - \Delta_2\right\|^2 \\
&= t^4\|D_2(X)\|^2 + 4t^2(1-t)^2\|H\|^2 + (1-t)^4\|D_2(Y)\|^2 + \\
&\quad 4t^3(1-t)\langle D_2(X), H\rangle + 2t^2(1-t)^2\langle D_2(X), D_2(Y)\rangle + \\
&\quad 4t(1-t)^3\langle D_2(Y), H\rangle - 2t^2\langle D_2(X), \Delta\rangle - \\
&\quad 4t(1-t)\langle H, \Delta_2\rangle - 2(1-t)^2\langle D_2(Y), \Delta_2\rangle + \|\Delta_2\|^2 \\
&= At^4 + Bt^3 + Ct^2 + Dt + E,
\end{aligned}$$

where

$$\begin{aligned}
A &= 2\langle D_2(X), D_2(Y)\rangle - 4\langle D_2(Y), H\rangle + \|D_2(X)\|^2 + \|D_2(Y)\|^2 - 4\langle D_2(X), H\rangle + 4\|H\|^2, \\
B &= -4\langle D_2(X), D_2(Y)\rangle + 12\langle D_2(Y), H\rangle - 4\|D_2(Y)\|^2 - 8\|H\|^2 + 4\langle D_2(X), H\rangle, \\
C &= -12\langle D_2(Y), H\rangle + 4\langle H, \Delta_2\rangle + 2\langle D_2(X), D_2(Y)\rangle - 2\langle D_2(Y), \Delta_2\rangle + 6\|D_2(Y)\|^2 + \\
&\quad 4\|H\|^2 - 2\langle D_2(X), \Delta_2\rangle, \\
D &= 4\langle D_2(Y), H\rangle - 4\langle H, \Delta_2\rangle - 4\|D_2(Y)\|^2 + 4\langle D_2(Y), \Delta_2\rangle, \\
E &= \|D_2(Y)\|^2 - 2\langle D_2(Y), \Delta_2\rangle + \|\Delta_2\|^2.
\end{aligned}$$

Because

$$A = \|D_2(X) - 2H\|^2 + \|D_2(X) - 2H\|^2 + 2\langle D_2(X), D_2(Y)\rangle > 0$$

and $p'(0) = 0$ because $\nabla\sigma_2(Y) = 0$, we can apply Lemma 1 to conclude that $D = 0$, i.e. that

$$\langle D_2(Y), H\rangle + \langle D_2(Y), \Delta_2\rangle = \langle H, \Delta_2\rangle + \|D_2(Y)\|^2 . \tag{8}$$

Part (i) of Theorem 1 states that $\langle D_2(Y), \Delta_r\rangle = \|D_2(Y)\|^2$; hence, equation (8) simplifies to

$$\langle D_2(Y), H\rangle = \langle \Delta_2, H\rangle . \tag{9}$$

Because X and Y are interchangeable, we also have

$$\langle D_2(X), H\rangle = \langle \Delta_2, H\rangle . \tag{10}$$

Combining equations (10) and (9) gives (3).

Next, applying (3) with H represented as in (7), we see that

$$\begin{aligned} 0 &= 2t(1-t)\langle D_2(X), H\rangle - \langle D_2(Y), H\rangle \\ &= \left\langle D_2(X) - D_2(Y), D_2(tX + (1-t)Y) - t^2 D_2(X) - (1-t)^2 D_2(Y)\right\rangle \\ &= \langle D_2(X) - D_2(Y), D_2(tX + (1-t)Y)\rangle - \\ &\quad t^2\langle D_2(X) - D_2(Y), D_2(X)\rangle - (1-t)^2\langle D_2(X) - D_2(Y), D_2(X)\rangle \end{aligned}$$

Because $\sigma_2(X) = \sigma_2(Y)$, we conclude from part (iii) of Theorem 1 that $\|D_2(X)\| = \|D_2(Y)\|$; hence, that

$$\begin{aligned} \langle D_2(X) - D_2(Y), D_2(tX + (1-t)Y)\rangle &= t^2\|D_2(X)\|^2 - t^2\langle D_2(X), D_2(Y)\rangle + \\ &\quad (1-t)^2\langle D_2(X), D_2(Y)\rangle - (1-t)^2\|D_2(Y)\|^2 \\ &= (2t-1)\|D_2(X)\|^2 - (2t-1)\langle D_2(X), D_2(Y)\rangle , \end{aligned}$$

which is (5).

Finally, if $t = 1/2$ in (5), then

$$\left\langle D_2(X) - D_2(Y), D_2\left(\frac{X+Y}{2}\right)\right\rangle = 0,$$

from which (6) follows immediately. □

4 Approximate Solutions

The problem of minimizing σ_2, the metric sstress problem, must be solved by numerical optimization. In contrast, the classical approach to MDS proposed by Torgerson (1952), leads to an explicit formula for an optimal configuration. For this reason, as discussed by Kearsley, Tapia, and Trosset (1998), the classical solution is often used as an initial configuration from which to begin minimizing the sstress criterion. In this section, we explore the quality of the classical solution as measured by σ_2.

Given n, let e denote the n-vector $(1,\ldots,1)'$. Given an $n\times n$ matrix B, let b denote the n-vector $\mathrm{diag}(B)$. Let κ denote the linear transformation defined by

$$\kappa(B) = be' + eb' - 2B.$$

Given $p \le n$, let $\mathcal{D}_n(p)$ denote the set of matrices that can be realized as the interpoint distances of some $x_1,\ldots,x_n \in \Re^p$ and let $\Omega_n(p)$ denote the set of symmetric positive semidefinite $n\times n$ matrices of rank no greater than p. Then the following result is well-known:

Theorem 4 *$D \in \mathcal{D}_n(p)$ if and only if there exists $B \in \Omega_n(p)$ such that $\kappa(B) = D$.*

The linear transformation κ does not have a unique inverse. In fact, for any $s \in \Re^p$ such that $s'e = 1$, the linear transformation τ_s defined by

$$\tau_s(D) = -\frac{1}{2}(I - es')\,D\,(I - es')$$

is an inverse of κ, where I denotes the $n \times n$ identity matrix. We are interested in the inverse τ_1 obtained by setting $s = e/n$. See Critchley (1988) for a detailed study of the properties of κ and τ_1.

Classical MDS can be defined by the optimization problem

$$\begin{array}{ll} \text{minimize} & \|B - \tau_1(\Delta_2)\|^2 \\ \text{subject to} & B \in \Omega_n(p), \end{array} \tag{11}$$

which is implicit in Torgerson (1952). The objective function was subsequently dubbed the strain criterion.

The following explicit solution to Problem 11 is also well-known:

Theorem 5 *Given Δ, let $\lambda_1 \geq \cdots \geq \lambda_n$ denote the eigenvalues of $B = \tau_1(\Delta_2)$ and let $v_1, \ldots, v_n$ denote the corresponding eigenvectors. Given $p \leq n$, let $\lambda_i^+ = \max(\lambda_i, 0)$ for $i = 1, \ldots, p$. Then*

$$\hat{B} = \sum_{i=1}^{p} \lambda_i^+ v_i v_i'$$

is a global minimizer of Problem 11. Furthermore, if $\hat{X}$ is the $n \times p$ configuration matrix whose ith column is $(\lambda_i^+)^{1/2} v_i$, then

$$\kappa\left(\hat{B}\right) = D_2\left(\hat{X}\right).$$

Because the classical solution, $\hat{X}$, can be computed explicitly, it is often used as the initial configuration from which optimization of the sstress criterion commences. This practice begs the question of how close $\hat{X}$ comes to solving the metric sstress problem.

Let us write

$$\sigma_2\left(\hat{X}\right) = \left\|\kappa\left(\hat{B}\right) - \Delta_2\right\|^2 = \left\|\sum_{i=1}^{p} \lambda_i^+ \kappa\left(v_i v_i'\right) - \Delta_2\right\|^2.$$

Tarazaga and Trosset (1998) observed that a configuration matrix with a smaller sstress value than $\sigma_2(\hat{X})$ can be obtained by replacing the λ_i^+ with free variables $\mu \in \Re^p$, resulting in the objective function

$$f(\mu) = \left\|\sum_{i=1}^{p} \mu_i \kappa\left(v_i v_i'\right) - \Delta_2\right\|^2,$$

and solving the p-variate optimization problem:

$$\begin{array}{ll} \text{minimize} & f(\mu) \\ \text{subject to} & \mu \geq 0, \end{array} \tag{12}$$

This is a fairly easy problem to solve numerically, but it does require an iterative algorithm for bound-constrained optimization.

We can obtain a configuration matrix of intermediate quality by allowing fewer degrees of freedom. Instead of replacing each λ_i^+ with a freely varying μ_i, we write $\mu_i = t\lambda_i^+$ and allow $t \in \Re$ to vary. This results in the objective function

$$g(t) = \left\| \sum_{i=1}^{p} t\lambda_i \kappa\left(v_i v_i'\right) - \Delta_2 \right\|^2 = \left\| \kappa\left(t\hat{B}\right) - \Delta_2 \right\|^2 = \left\| D_2\left(t\hat{X}\right) - \Delta_2 \right\|^2$$

and the optimization problem

$$\begin{array}{ll} \text{minimize} & g(t) \\ \text{subject to} & t \geq 0. \end{array} \tag{13}$$

Recognizing that Problem (13) is the problem of optimally dilating the classical solution, we can apply Theorem 2 to compute its solution explicitly, without recourse to numerical optimization. We leave the problem of quantifying how much solving either Problem (13) or Problem (12) improves on $\hat{X}$ to be determined by future computational experiments.

Now suppose that $\Delta = (\delta_{ij})$ is actually a distance matrix, i.e. that $\Delta \in \mathcal{D}_n(n-1)$, and write

$$D\left(\hat{X}\right) = \hat{D} = \left(\hat{d}_{ij}\right).$$

Then it is well-known—see Meulman (1992) for discussion— that

$$\hat{d}_{ij}^2 \leq \delta_{ij}^2.$$

Because

$$g_r(t) = \left\| t\hat{D}_r - \Delta_r \right\|^2 = \sum_{ij} \left(t\hat{d}_{ij}^r - \delta_{ij}^r \right)^2,$$

we have the following result:

Theorem 6 *Let Δ be a fixed dissimilarity matrix. Given p, let $\hat{X}$ denote the configuration matrix defined in Theorem 5. Given r, let t^* denote the optimal dilation of $\hat{X}$. If $\Delta \in \mathcal{D}_n(n-1)$, then $t^* \geq 1$, with equality if and only if $\Delta \in \mathcal{D}_n(p)$.*

5 Discussion

This study is a work in progress and our results raise as many questions as they answer. We conclude by cataloguing some of the concerns that we plan to address in future work.

1. Do nonglobal minimizers of the sstress criterion exist?

   In Section 2, we demonstrated that the example used by Trosset and Mathar (1997) to construct a nonglobal minimizer of the stress criterion does not lead to a local minimizer of the sstress criterion. This surprised us. Our inability to easily discover nonglobal minimizers underscores the question of whether or not they exist. While we continue to search for an example of a nonglobal minimizer, we will also attempt to exploit the quartic structure of σ_2 to characterize its stationary configurations. Although we have not yet fully explored its consequences, Theorem 3 is a first step in that direction.

2. How much does optimal dilation improve on the classical solution?

   Theorem 2 is a powerful result. Through an explicit formula, it provides a simple way of improving most suboptimal configurations with respect to any σ_r. We would like to know how much optimal dilation improves on the classical solution—the canonical configuration from which numerical optimization of σ_r is commenced—with respect to (i) the value of σ_r and (ii) the probability of converging to a global minimizer of σ_r. We expect to address these issues empirically, by extensive computational experimentation, but we have some hope of a theoretical analysis in the comparatively tractable case of σ_2.

3. How can Theorem 1 be exploited for global optimization?

   It is apparent from Theorem 1 that the stationary configurations of σ_r have considerable structure. Furthermore, part (iii) of Theorem 1 suggests a strategy for globally optimizing σ_r. Because we now understand that these observations apply to any σ_r, not just the case of $r = 2$ addressed by Glunt, Hayden, and Liu (1991), the potential gains from exploiting them are greatly increased.

Acknowledgements

This research was conducted during the summer of 1999 as part of the Research Experiences for Undergraduates program in the Department of Mathematics at the College of William & Mary. The first author was supported by NSF Grant DMS-96-19577; the second author was supported by NSF Grants DMS-9622749 and DMS-96-19577.

References

Critchley, F. (1988). On certain linear mappings between inner-product and squared-distance matrices. *Linear Algebra and Its Applications*, 105:91–107.

Glunt, W., Hayden, T. L., and Liu, W.-M. (1991). The embedding problem for predistance matrices. *Bulletin of Mathematical Biology*, 53:769–796.

Kearsley, A. J., Tapia, R. A., and Trosset, M. W. (1998). The solution of the metric STRESS and STRESS problems in multidimensional scaling using Newton's method. *Computational Statistics*, 13(3):369–396.

Meulman, J. J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika*, 57:539–565.

Tarazaga, P. and Trosset, M. W. (1998). An approximate solution to the metric SSTRESS problem in multidimensional scaling. *Computing Science and Statistics*, 30(1).

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 17:401–419.

Trosset, M. W. and Mathar, R. (1997). On the existence of nonglobal minimizers of the STRESS criterion for metric multidimensional scaling. In *1997 Proceedings of the Statistical Computing Section*, pages 158–162. American Statistical Association.