# AN IMPROVED TUNNELING FUNCTION FOR FINDING A DECREASING SERIES OF LOCAL MINIMA IN MDS

Patrick Groenen
Willem J. Heiser

Department of Data Theory
University of Leiden

RR-91-06

# AN IMPROVED TUNNELING FUNCTION FOR FINDING A DECREASING SERIES OF LOCAL MINIMA IN MDS

ABSTRACT. The tunneling method is applied to obtain a decreasing series of local minima of the STRESS function. It alternates a local search procedure with the tunneling step in which another configuration is sought of equal STRESS. The tunneling step is perfomed by finding the zero points of the tunneling function. This function is adapted to avoid trivial zero points like rotations of the local minimum configuration or horizon points. A majorization algorithm is derived to find these points. Two examples are given to illustrate the method.

## 1. Introduction.

In this report we study systematic ways for finding a decreasing series of local minima of the MDS STRESS function by using the tunneling method (see Groenen, 1990). This method consists of a two-step procedure: in the first step a local minimum is sought and in the second one another configuration is determined with exactly the same STRESS. The second step is performed by minimization of a particular function, called the tunneling function. In order to find another configuration with the same STRESS, this function must have several characteristics. Some of these characteristics are met by the tunneling function defined by Groenen (1990):

$$\tau_1(\mathbf{X}) = \frac{(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2}{\|\mathbf{X}^* - \mathbf{X}\|^2}$$

where $\mathbf{X}$, of order $n$ by $p$, is the configuration of coordinates, $\mathbf{X}^*$ is the configuration of a local minimum and $\sigma(\mathbf{X})$ is the STRESS of the configuration $\mathbf{X}$ given by

$$\sigma(\mathbf{X}) = \left(\frac{1}{2}\sum_{i,j}^{n} (\delta_{ij} - d_{ij}(\mathbf{X}))^2\right)^{1/2}$$

where for each pair of stimuli $ij$ a non-negative dissimilarity is represented by $\delta_{ij}$ and their Euclidean distance by $d_{ij}(\mathbf{X})$. The first characteristic that the tunneling function clearly exhibits is that it has zero points for configurations with STRESS $\sigma(\mathbf{X}^*)$. Secondly, these zero points are the lowest possible value of the tunneling function. Thirdly, the factor $\|\mathbf{X}^* - \mathbf{X}\|^{-2}$, also called

Acknowledgement
The authors are grateful to Rudolf Mathar for his comments and suggestions on early drafts of this paper.

the *pole* of the tunneling function, ensures elevated values of the tunneling function near the local minimum configuration **X***, so that a zero point at **X*** is presumably excluded. Because the pole is a factor, it does not change the positions of other zero points.

The problem of finding the global minimum of the STRESS function has been replaced by the minimization of the tunneling function. Clearly, if **X*** is a unique global minimum it will be impossible to find a zero point of $\tau_1(\mathbf{X})$. However, in the sequel we assume that **X*** is not a unique global minimum, so that zero points of the tunneling function do exist.

Unfortunately, the tunneling function as defined above has some major defects. Groenen (1990) found that the pole generally is not strong enough to exclude a zero point at **X***. A second defect involves the behavior of the tunneling function under rotations of **X*** and **X**. Since distances do not change under rotation of the configuration, neither does STRESS. However, due to the form chosen for the pole, the tunneling function as defined above does change under rotation. Therefore, mere rotations of **X*** are not excluded as zero points of the tunneling function, and actually do occur as solutions of it. The third problem is the behavior of the tunneling function for very large configurations, or horizon configurations. It turns out that the algorithm may be attracted to the horizon since the tunneling function tends to zero for increasing uniform dilations of **X**.

These problems and some of their mathematical properties are discussed in section 2 and will result in a redefinition of the tunneling function. Section 3 then shows how the redefined function can be minimized to obtain its zero points. Two important methods used for the minimization of the tunneling function are *iterative majorization* and *parametric programming*. In section 4 some first results of the developed method are given.

## 2. Redefining the tunneling function.

In this section we show how to resolve the problems described in the introduction. The first problem involves changes of the tunneling function value when a configuration is rotated. This is inconsistent with the STRESS function, which is invariant under rotation. Therefore we shall adapt the tunneling function so that it also becomes invariant under rotations. The second problem involves the strength of the pole. It is shown analytically that a pole strength parameter must be included. A similar analytical result is obtained for the third problem: the behavior of the tunneling function for large configurations. We present a redefinition of the tunneling function that solves all three of these problems and still exhibits the characteristics initially set out as desirable.

## *2.1. Obtaining rotational invariance of the denominator.*

One basic property of the STRESS function is its invariance under rotation of the solution **X**. This property can be easily understood by realizing that the STRESS function is defined on the distances between points of the configuration, not on the configuration itself. Clearly, distances do not change under rotation, reflection or translation. Therefore it is desirable that this property is kept under control in the tunneling function. Unreported numerical experiments showed that minimizing $\tau_1(\mathbf{X})$ leads to a rotation of the local minimum configuration within a small number of iterations. The behavior of $\tau_1(\mathbf{X})$ can be understood by examining it more closely. The problem is caused by the numerator $(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2$, since the STRESS function is invariant under rotation of **X**. It can be seen quite easily that a rotation of $\mathbf{X}^*$ is a minimum of the tunneling function. Let **R** be a rotation matrix with $\mathbf{R}'\mathbf{R} = \mathbf{R}\mathbf{R}' = \mathbf{I}$. Since $\sigma(\mathbf{X}^*\mathbf{R}) = \sigma(\mathbf{X}^*)$, the numerator of $\tau_1(\mathbf{X}^*\mathbf{R})$ is zero. Furthermore, the denominator is

$$\begin{aligned} \text{tr}(\mathbf{X}^*\mathbf{R} - \mathbf{X}^*)'(\mathbf{X}^*\mathbf{R} - \mathbf{X}^*) &= \text{tr}\, \mathbf{R}'\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{R} + \text{tr}\, \mathbf{X}^{*\prime}\mathbf{X}^* - 2\text{tr}\, \mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{R} \\ &= 2\text{tr}\, \mathbf{X}^{*\prime}\mathbf{X}^* - 2\text{tr}\, \mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{R} \\ &= 2\text{tr}\, \mathbf{X}^*(\mathbf{I} - \mathbf{R})\mathbf{X}^{*\prime}. \end{aligned}$$

The left-hand-side being greater than zero, $\mathbf{I} - \mathbf{R}$ must be positive definite. The latter can be verified by using the Cauchy-Schwartz inequality

$$(\text{tr}\, \mathbf{Y}'\mathbf{Y})^{1/2}\, (\text{tr}\, \mathbf{R}'\mathbf{Y}'\mathbf{Y}\mathbf{R})^{1/2} = (\text{tr}\, \mathbf{Y}'\mathbf{Y})^{1/2}\, (\text{tr}\, \mathbf{Y}'\mathbf{R}'\mathbf{R}\mathbf{Y})^{1/2} = \text{tr}\, \mathbf{Y}'\mathbf{Y} \geq \text{tr}\, \mathbf{Y}'\mathbf{R}\mathbf{Y}$$

for any rotation matrix **R**, with equality if and only if $\mathbf{R} = \mathbf{I}$. Thus any $\mathbf{R} \neq \mathbf{I}$ yields a non-zero denominator. The zero numerator and non-zero denominator of $\tau(\mathbf{X}^*\mathbf{R})$ imply a minimum, but not a desirable one. Therefore, we need to adapt the denominator to remove the possibility of finding rotations of the local minimum configuration.

This can be achieved by applying the very same idea that makes the STRESS function invariant to rotations of the configuration. We simply use a function that measures the difference between the distances of the local minimum configuration and the distances of the current configuration. Or, to put it differently, take the squared norm of the difference between the distance matrix of **X*** and the distance matrix of **X**, i.e.

$$\|D(\mathbf{X}^*) - D(\mathbf{X})\|^2 = \|\mathbf{d}^* - \mathbf{d}\|^2 = \sum_{i<j}^{n} (d_{ij}(\mathbf{X}^*) - d_{ij}(\mathbf{X}))^2 \ . \tag{1}$$

This denominator can be interpreted as the squared distance between the vector of distances of the local minimum configuration and the vector of distances among **X**. In a different context the idea to measure the distance between configurations through their interpoint distances has been studied extensively by Meulman (1986). We use (1) as the denominator of the tunneling function, which leads to the following redefinition of the tunneling function

$$\tau_2(\mathbf{X}) = \frac{(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2}{\|D(\mathbf{X}^*) - D(\mathbf{X})\|^2} \ . \tag{2}$$

Another advantage of this redefinition is that both numerator and denominator are comparable entities. Both STRESS and the current denominator can be viewed as (squared) norms of the difference between two vectors. By using (2) we have obtained invariance of the tunneling function under any rotation **R** of the configuration, which implies that the effectiveness of the pole is extended to all **X** in the neighbourhood of **X*****R**.

### *2.2. The pole strength.*

In this section we discuss the strength of the pole analytically and give a suggestion how to increase the pole strength.

*2.2.1. The need for a stronger pole in the tunneling function.*

Groenen (1990) showed empirically that the pole (denominator of the tunneling function) is not strong enough to cancel out a zero point of the tunneling function at **X***. In this section we give an upper bound for the tunneling function. Furthermore, we shall prove analytically that the tunneling function may tend to zero when **X** approaches **X***.

The following theorem states that the numerator of $\tau_2(\mathbf{X})$ is always smaller than its denominator.

*Theorem 1*:

For each $\mathbf{X}, \mathbf{X}^* \in \mathbb{R}^{np}$ it holds that $(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2 \leq \|D(\mathbf{X}^*) - D(\mathbf{X})\|^2$.

*Proof*:

Let $\boldsymbol{\delta}$, $\mathbf{d}$, $\mathbf{d}^*$ be the $n(n-1)/2$ vectors with lower triangular elements of $\Delta$, D(**X**), D(**X***), so that $\sigma(\mathbf{X}) = \|\boldsymbol{\delta} - \mathbf{d}\|$ and $\|D(\mathbf{X}^*) - D(\mathbf{X})\|^2 = \|\mathbf{d}^* - \mathbf{d}\|^2$. Applying the triangle inequality twice we obtain

$$\|\boldsymbol{\delta} - \mathbf{d}\| \leq \|\boldsymbol{\delta} - \mathbf{d}^*\| + \|\mathbf{d} - \mathbf{d}^*\|,$$
$$\|\boldsymbol{\delta} - \mathbf{d}^*\| \leq \|\boldsymbol{\delta} - \mathbf{d}\| + \|\mathbf{d} - \mathbf{d}^*\|.$$

So $\|\mathbf{d} - \mathbf{d}^*\|$ must be larger than both $\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*)$ and $\sigma(\mathbf{X}^*) - \sigma(\mathbf{X})$, and therefore larger than their maximum too:

$$|\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*)| \leq \|\mathbf{d} - \mathbf{d}^*\|.$$

Squaring both sides gives the desired result. Q.E.D.

The preceding inequality implies that $0 \leq \tau_2(\mathbf{X}) \leq 1$. This upper bound of $\tau_2(\mathbf{X})$ does not tell us whether the pole is strong enough to avoid a minimum at **X***. It merely shows that the maximum value of the tunneling function equals one. Another interesting question is whether $\tau_2(\mathbf{X}) = 1$ can be found. To answer these questions we can look at the behavior of $\tau_2(\mathbf{X})$ as **X** approaches **X***.

Consider the complete space of vectors **d** in $\mathbb{R}^{n(n-1)/2}$. Although the set of vectors of Euclidean distances is a subset of $\mathbb{R}^{n(n-1)/2}$, we can investigate what happens when **d** approaches **d*** from any direction **z**. If **z** has unit norm, this amounts to developing the limit of $\tau_2(\mathbf{d}^*-r\mathbf{z})$ as $r$ tends to 0;

$$\lim_{r\to 0} \tau_2(\mathbf{d}^*-r\mathbf{z}) = \lim_{r\to 0} \frac{(\|\boldsymbol{\delta}-\mathbf{d}^*+r\mathbf{z}\| - \|\boldsymbol{\delta}-\mathbf{d}^*\|)^2}{\|\mathbf{d}^*-\mathbf{d}^*+r\mathbf{z}\|^2}$$

$$= \left| \lim_{r\downarrow 0} \frac{\|\boldsymbol{\delta}-\mathbf{d}^*+r\mathbf{z}\| - \|\boldsymbol{\delta}-\mathbf{d}^*\|}{r} \right|^2 .$$

By substituting $\mathbf{a} = \boldsymbol{\delta}-\mathbf{d}^*$, it can be seen that the latter equation is exactly the square of the directional derivative of the function $f(\mathbf{a}) = \|\mathbf{a}\|$. If **g** is the gradient of f(**a**), the directional derivative of f(**a**) along direction **z** can be expressed as $\mathbf{g}'\mathbf{z}$ (see for example Gill, Murray and Wright, 1981, p. 53). The gradient **g** of f(**a**) is given by $\mathbf{a}/\|\mathbf{a}\|$, and the directional derivative is thus $\mathbf{a}'\mathbf{z}/\|\mathbf{a}\|$. Now we can write

$$\lim_{r\to 0} \tau_2(\mathbf{d}^*-r\mathbf{z}) = \left(\frac{(\boldsymbol{\delta}-\mathbf{d}^*)'\mathbf{z}}{\|\boldsymbol{\delta}-\mathbf{d}^*\|}\right)^2$$

which is between zero, when **z** is orthogonal to $\boldsymbol{\delta}-\mathbf{d}^*$, and one, when **z** is a multiple of $\boldsymbol{\delta}-\mathbf{d}^*$. In this proof we assumed that any **z** could be used from $\mathbb{R}^{n(n-1)/2}$. Clearly, the set of $\mathbf{d}^*-r\mathbf{z}$ that are Euclidean distance vectors is a subset of $\mathbb{R}^{n(n-1)/2}$. Since the proof holds for the whole set it must also hold for a subset. Therefore we conclude that the value of $\tau_2(\mathbf{X})$ may be any value between zero and unity as **X** approaches **X***, though stronger results may exist for the subset.

In this section it has been proven that the tunneling function (2) has values between 0 and 1 and that, unfortunately, near the pole the function also may have values between zero and one, depending from which direction we approach. Therefore we cannot guarantee that the pole is strong enough to avoid a zero point of the tunneling function at the local minimum configuration of the STRESS function. Thus we need to adapt the tunneling function once more to obtain a sufficiently strong pole.

### 2.2.2. *A stronger pole of the tunneling function.*

The previous section showed the desirability for a stronger pole. One obvious way of creating a stronger pole is raising the denominator to some power larger than one, say $\lambda$ (the pole strength parameter). Another way of doing this is to take a root of the numerator. Clearly, both options are possible, because they are monotonic transformations of each other, and such a transformation does not influence the minima or maxima. From a computational point of view, we have a slight preference for the second option. If the denominator is close to zero, raising it to some power causes the denominator to get even closer to zero. Even for moderate sized $\lambda$, underflow can occur easily. Taking a root (that is, raising it to the power $\lambda$ with $0 < \lambda < 1$) of the numerator does not suffer equally from this problem. If the numerator is close to zero and smaller than one, which is the case if $\mathbf{X}$ is close to $\mathbf{X}^*$, the root is a larger value. The overflow problem occurs only for small values of $\lambda$. One drawback might be that this transformation increases the rounding error by the power $1/\lambda$. This remains a matter to investigate.

Having the rotational problem and the pole strength problem eliminated, we redefine the tunneling function by

$$\tau_3(\mathbf{X}) = \frac{(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^{2\lambda}}{\|\mathrm{D}(\mathbf{X}^*) - \mathrm{D}(\mathbf{X})\|^2} \ . \tag{3}$$

In section 3 we shall see that taking the numerator to the power $\lambda$, with $0 < \lambda < 1$, is indeed a good choice for the minimization procedure.

### 2.3. *Attractive horizon.*

It is of importance to see how the tunneling function behaves when the configuration moves to the horizon, i.e. when the coordinates get uniformly large. It is shown here that during minimization of (3) the configuration may be attracted to the horizon due to tunneling

Let us first define what is understood by a large configuration. Given a configuration **X**, not equal to **0**, we say that $c\mathbf{X}$ is large when the uniform dilation factor $c$ is large. The behavior of the tunneling function for large configurations can now be translated in the mathematical problem of finding the limit of $\tau_3(c\mathbf{X})$ as $c$ goes to infinity, i.e.

$$\lim_{c\to\infty} \tau_3(c\mathbf{X}) = \lim_{c\to\infty} \frac{(\|\Delta - \mathrm{D}(c\mathbf{X})\| - \|\Delta - \mathrm{D}(\mathbf{X}^*)\|)^{2\lambda}}{\|\mathrm{D}(\mathbf{X}^*) - \mathrm{D}(c\mathbf{X})\|^2} .$$

The second step in solving this limit is to use the fact that the Euclidean distance function is homogenous, so that $\mathrm{D}(c\mathbf{X}) = c\mathrm{D}(\mathbf{X})$, thus the limit can be written as

$$\lim_{c\to\infty} \tau_3(c\mathbf{X}) = \lim_{c\to\infty} \frac{(\|\Delta - c\mathrm{D}(\mathbf{X})\| - \|\Delta - \mathrm{D}(\mathbf{X}^*)\|)^{2\lambda}}{\|\mathrm{D}(\mathbf{X}^*) - c\mathrm{D}(\mathbf{X})\|^2} .$$

Further, to evaluate the limit it is convenient to substitute $c = \alpha/(1-\alpha)$ and subsequently let $\alpha$ tend to 1. This simplifies the limit into

$$\begin{aligned}
\lim_{c\to\infty} \tau_3(c\mathbf{X}) &= \lim_{\alpha\to 1} \left( \frac{\|\Delta - \alpha/(1-\alpha)\mathrm{D}(\mathbf{X})\| - \|\Delta - \mathrm{D}(\mathbf{X}^*)\|}{\|\mathrm{D}(\mathbf{X}^*) - \alpha/(1-\alpha)\mathrm{D}(\mathbf{X})\|^{1/\lambda}} \right)^{2\lambda} \\
&= \lim_{\alpha\to 1} \left( \frac{(1-\alpha)^{-1}\|(1-\alpha)\Delta - \alpha\mathrm{D}(\mathbf{X})\| - \|\Delta - \mathrm{D}(\mathbf{X}^*)\|}{(1-\alpha)^{-1/\lambda}\|(1-\alpha)\mathrm{D}(\mathbf{X}^*) - \alpha\mathrm{D}(\mathbf{X})\|^{1/\lambda}} \right)^{2\lambda} \\
&= \lim_{\alpha\to 1} \left( \frac{(1-\alpha)^{1/\lambda-1}\|(1-\alpha)\Delta - \alpha\mathrm{D}(\mathbf{X})\|}{\|(1-\alpha)\mathrm{D}(\mathbf{X}^*) - \alpha\mathrm{D}(\mathbf{X})\|^{1/\lambda}} - \frac{(1-\alpha)^{1/\lambda}\|\Delta - \mathrm{D}(\mathbf{X}^*)\|}{\|(1-\alpha)\mathrm{D}(\mathbf{X}^*) - \alpha\mathrm{D}(\mathbf{X})\|^{1/\lambda}} \right)^{2\lambda} \\
&= \lim_{\alpha\to 1} \left( \frac{(1-\alpha)^{1/\lambda-1}\|\mathrm{D}(\mathbf{X})\|}{\|\mathrm{D}(\mathbf{X})\|^{1/\lambda}} - \frac{(1-\alpha)^{1/\lambda}\|\Delta - \mathrm{D}(\mathbf{X}^*)\|}{\|\mathrm{D}(\mathbf{X})\|^{1/\lambda}} \right)^{2\lambda} \\
&= \lim_{\alpha\to 1} \left( (1-\alpha)^{1/\lambda-1}\|\mathrm{D}(\mathbf{X})\|^{1-1/\lambda} - \frac{(1-\alpha)^{1/\lambda}\|\Delta - \mathrm{D}(\mathbf{X}^*)\|}{\|\mathrm{D}(\mathbf{X})\|^{1/\lambda}} \right)^{2\lambda} \\
&= (0-0)^{2\lambda} = 0 .
\end{aligned}$$

The final part of the limit follows from the assumption that $0 < \lambda < 1$, so that $1/\lambda-1$ is greater than zero, $1/\lambda$ is greater than one and consequently $(1-\alpha)^{1/\lambda-1}$ and $(1-\alpha)^{1/\lambda}$ tend to zero for $\alpha$ approaching one. This clearly shows that for large configurations the tunneling function with a strong pole tends to zero. This is an unattractive feature, because the original tunneling problem has been transformed into finding a zero point of the tunneling function. The limit above shows that tunneling function values close to zero can be expected not only near an equal STRESS configuration, but also for large configurations. This may direct our algorithm inadvertently to the horizon.

### 2.4. *The final tunneling function.*

Here we present the final version of the tunneling function. The previous section indicated that a minimization algorithm based on (3) may fail. Now we are looking for a factor that has high values if **X** is close to **X***, also has high or constant nonzero values for large configurations, but retains the zero points at configurations with STRESS equal to the local minimum STRESS value.

Let the numerator of the tunneling function (3) be presented by N(**X**) and the denominator $\|D(\mathbf{X}^*) - D(\mathbf{X})\|^2$ by P(**X**), so the function can be written as N(**X**)/P(**X**). Instead of the factor 1/P(**X**), a factor is sought that is close to unity for large **X**, and yields a large value for **X** close to **X***. One such factor is (1 + 1/P(**X**)). This leads to the final redefinition of the tunneling function,

$$
\begin{aligned}
\tau_4(\mathbf{X}) &= (\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^{2\lambda}\left[1 + \frac{1}{\|D(\mathbf{X}^*) - D(\mathbf{X})\|^2}\right] \\
&= (\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^{2\lambda}\frac{\|D(\mathbf{X}^*) - D(\mathbf{X})\|^2 + 1}{\|D(\mathbf{X}^*) - D(\mathbf{X})\|^2}.
\end{aligned} \tag{4}
$$

This function seems to satisfy all requirements: the zero points (that is, points with equal STRESS) do not change after multiplication, a configuration that is a rotation of **X** yields the same tunneling function value, the pole is strong enough, and for large configurations minimizing $\tau_4(\mathbf{X})$ amounts to minimizing $(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^{2\lambda}$ because $1 + \|D(\mathbf{X}^*) - D(\mathbf{X})\|^{-2}$ tends to one.

In the remaining part of this report we will refer to (4) whenever the term tunneling function is used.

## 3. Minimizing the tunneling function.

In the previous section a tunneling function was developed that meets the requirements set in the introduction. This section discusses a zero finding algorithm of this function. Since the zero points are also the global minima of the tunneling function, we can use a minimization algorithm to find a zero point. The minimization method used here is a combination of parametric programming and iterative majorization. First we show that parametric programming also works when iterative majorization is applied. Then we discuss how a product of two functions can be majorized. Next a majorization inequality is developed for a root of $x$. These results are then applied to the tunneling function and an algorithm is given.

### *3.1. Convergence proof of majorized parametric programming.*

The tunneling function can be considered as a ratio of two functions of **X** and hence leads to a fractional programming problem. An algorithm to minimize such a function was proposed by Dinkelbach (1967), and used by Heiser (1981) and Groenen (1990). However, Dinkelbach's algorithm assumes that at each iteration the absolute minimum over **X** of an auxiliary function $\mathrm{F}(q,\mathbf{X})$ can be obtained. Such an assumption is too strong for our purpose and in this section we develop a more flexible version of the algorithm. More specifically, we prove that the algorithm is still convergent if we find the minimum of an auxiliary function that majorizes $\mathrm{F}(q,\mathbf{X})$.

The tunneling function $\tau_4(\mathbf{X})$ is given by $\mathrm{N}(\mathbf{X})(1+\mathrm{P}(\mathbf{X}))/\mathrm{P}(\mathbf{X})$, or, for ease of notation, $\tau_4(\mathbf{X})$ equals $\mathrm{M}(\mathbf{X})/\mathrm{P}(\mathbf{X})$ where $\mathrm{M}(\mathbf{X}) = \mathrm{N}(\mathbf{X})(1+\mathrm{P}(\mathbf{X}))$. Then Dinkelbach's algorithm amounts to

1. initialize $q^0$
2. $\mathbf{X}^+ \leftarrow \mathrm{argmin}\ \mathrm{F}(q^0,\mathbf{X})$
3. $q^+ \leftarrow \tau_4(\mathbf{X}^+)$
4. If $q^+ < \omega$ then *stop* (with $\omega$ a preset small positive constant)
5. else set $q^0 \leftarrow q^+$ and go to 2

where the auxiliary function $F(q,X)$ is defined by

$$F(q,X) = M(X) - q\,P(X) \tag{5}$$

Since we cannot find the argument that minimizes $F(q^0,X)$ in step 2 analytically, we need a more relaxed procedure that still guarantees convergence. Suppose we have one function $\mu_M(\mathbf{X},\mathbf{Y})$ that majorizes the numerator of $\tau_4(X)$, and another function $\mu_P(\mathbf{X},\mathbf{Y})$ that minorizes the denominator. Thus

$$\mu_M(\mathbf{X},\mathbf{Y}) \;\geq\; M(X) \qquad \text{with } \mu_M(\mathbf{Y},\mathbf{Y}) = M(\mathbf{Y}),$$
$$\mu_P(\mathbf{X},\mathbf{Y}) \;\leq\; P(X) \qquad \text{with } \mu_P(\mathbf{Y},\mathbf{Y}) = P(\mathbf{Y}).$$

Then the parametric function

$$\mu(q,\mathbf{X},\mathbf{Y}) = \mu_M(\mathbf{X},\mathbf{Y}) - q\mu_P(\mathbf{X},\mathbf{Y})$$

majorizes $F(q,X)$ for all $q \geq 0$. We can now state the following theorem, which guarantees that a one-step version of the parametric programming algorithm always improves $\tau_4(X)$, unless $\mathbf{X}$ satisfies the necessary conditions for a minimum.

*Theorem 2*:

Let $\mathbf{X}^+$ minimize $\mu(q,\mathbf{X},\mathbf{Y})$ for $q = \tau_4(\mathbf{Y})$. Then either $\tau_4(\mathbf{X}^+) < \tau_4(\mathbf{Y})$, or $\tau_4(\mathbf{X}^+) = \tau_4(\mathbf{Y})$. In the latter case $\mathbf{X}^+$ is stationary.

*Proof*

From the definition of $\mathbf{X}^+$ and from the properties of a majorizing function we have

$$F(q,\mathbf{X}^+) \leq \min_{\mathbf{X}} \mu(q,\mathbf{X},\mathbf{Y}) < \mu(q,\mathbf{Y},\mathbf{Y}) = F(q,\mathbf{Y})$$

for all $\mathbf{X}^+ \neq \mathbf{Y}$. Now the particular choice $q = \tau_4(\mathbf{Y})$ yields

$$F(\tau_4(\mathbf{Y}),\mathbf{Y}) = M(\mathbf{Y}) - \frac{M(\mathbf{Y})}{P(\mathbf{Y})}P(\mathbf{Y}) = 0,$$

and therefore the inequality becomes

$$M(\mathbf{X}^+) = M(\mathbf{X}^+) - \tau_4(\mathbf{Y})P(\mathbf{X}^+) < 0.$$

Using the equality $\tau_4(\mathbf{X}^+) = \mathrm{M}(\mathbf{X}^+)/\mathrm{P}(\mathbf{X}^+)$ we obtain

$$[\tau_4(\mathbf{X}^+) - \tau_4(\mathbf{Y})]\, \mathrm{P}(\mathbf{X}^+) < 0.$$

Since $\mathrm{P}(\mathbf{X}^+)$ is always non-negative by assumption, we must have $\tau_4(\mathbf{X}^+) < \tau_4(\mathbf{Y})$.

Now suppose $\tau_4(\mathbf{X}^+) = \tau_4(\mathbf{Y})$; this can only happen if

$$\mu(\tau_4(\mathbf{Y}),\mathbf{X}^+,\mathbf{Y}) = \mu(\tau_4(\mathbf{Y}),\mathbf{Y},\mathbf{Y}).$$

Since the majorizing function has a unique minimum, it must be true that $\mathbf{X}^+ = \mathbf{Y}$, and $\mathbf{X}^+$ should satisfy the stationary equation for the majorizing function. Setting the partial derivatives of $\mu(\tau_4(\mathbf{Y}),\mathbf{X},\mathbf{Y})$ equal to zero, we get

$$\frac{\partial \mathrm{M}(\mathbf{X})}{\partial \mathbf{X}} = \frac{\mathrm{M}(\mathbf{Y})}{\mathrm{P}(\mathbf{Y})} \frac{\partial \mathrm{P}(\mathbf{X})}{\partial \mathbf{X}} .$$

For a minimum of $\tau_4(\mathbf{X})$ the stationary equation is

$$\frac{\partial \mathrm{M}(\mathbf{X})}{\partial \mathbf{X}} \mathrm{P}(\mathbf{X}) - \frac{\partial \mathrm{P}(\mathbf{X})}{\partial \mathbf{X}} \mathrm{M}(\mathbf{X}) = \mathbf{0} .$$

When $\mathbf{X}^+ = \mathbf{Y}$ these two conditions are equivalent. Q.E.D.

This result implies that step 2 of Dinkelbach's algorithm may be replaced by $\mathbf{X}^+ \leftarrow \operatorname{argmin} \mu(q,\mathbf{X},\mathbf{Y})$.

The relaxed version of Dinkelbach's algorithm keeps convergence to a local minimum of our ratio function. However, some other nice properties of the original algorithm, like concavity of $\mathrm{F}(q,\mathbf{X}^+)$ in $q$, cannot be proven anymore. The stopping criterion (step 4) must be changed, because it cannot be guaranteed that the algorithm finds a zero point of the tunneling function; it may get stuck in a local minimum of $\tau_4(\mathbf{X})$.

### *3.2. Majorizing the product of two functions.*

One term of the auxiliary function $\mathrm{F}(q,\mathbf{X})$, introduced by parametric programming, is a product of two functions. Here we show how such a product can be majorized by two separate squared terms.

Let the product N(X)(1 + P(X)) be represented here by $x_1x_2$. The inequality

$$\left(\frac{x_1}{y_1} - \frac{x_2}{y_2}\right)^2 \geq 0$$

is always true due to non-negativeness of any square of a real argument. Working out this inequality gives

$$\left(\frac{x_1}{y_1}\right)^2 + \left(\frac{x_2}{y_2}\right)^2 - 2\frac{x_1x_2}{y_1y_2} \geq 0 \quad \text{or}$$

$$x_1x_2 \leq \frac{1}{2}\frac{y_2}{y_1}x_1^2 + \frac{1}{2}\frac{y_1}{y_2}x_2^2 \quad . \tag{6}$$

Furthermore, the inequality becomes a strict equality when $x_1$ equals $y_1$ and $x_2$ equals $y_2$. However, this happens also whenever $x_1$ equals $x_2y_1/y_2$, which can be derived from setting the inequality to zero or, equivalently, setting $x_1/y_1 = x_2/y_2$. Note that this does not complicate the majorization approach for minimization. The only requirement for majorization is that the auxiliary function touches the original function at a supporting point (as 4 does) and is larger or at most equal to the original function.

### *3.3. Majorizing a root of a positive value.*

Above it was pointed out that we must raise the numerator of the tunneling function to the power λ, where λ < 1. It was needed in order to avoid a zero point of the tunneling function at the previous local minimum. Throughout this method, we wish to stay within the majorization framework that guarantees a convergent algorithm. Therefore a majorization inequality is developed for this root. This is done for some distinct values of λ.

The idea for majorizing this root was taken from Heiser (1987), who showed how to majorize an absolute value. He used the inequality

$$\begin{aligned} (|x| - |y|)^2 &= x^2 + y^2 - 2|x||y| \geq 0 \\ |x| &\leq \frac{1}{2}|y| + \frac{1}{2}|y|^{-1}x^2 \end{aligned} \tag{7}$$

where $|x|$ is majorized by a quadratic function of $x$. Since we will use majorization for positive values only –N(**X**) is positive by definition– we assume in the sequel that $x$ is non-negative. Then (7) reduces to $x \leq \frac{1}{2} y + \frac{1}{2} y^{-1} x^2$. This is the main inequality used to find the extension that majorizes a positive root of $x$.

*Theorem 3*:

For all $\lambda \equiv 2^{-\kappa}$, $\kappa \in \{1,2,3,...\}$, $x \in \mathbb{R}^+$ and $y \in \mathbb{R}^+ \backslash \{0\}$ the following inequality holds

$$x^{\lambda} \leq (1 - \lambda)\, y^{\lambda} + \lambda y^{\lambda - 1}\, x \,. \tag{8}$$

*Proof*:

This will be proven by induction. Suppose the theorem is true for $\lambda = a$. Clearly, this assumption is true for $a = 1/2$, since (8) simply reduces to $x^{1/2} \leq \frac{1}{2} y^{1/2} + \frac{1}{2} y^{-1/2} x$ or $(x^{1/2} - y^{1/2})^2 \geq 0$. Let us examine a smaller value of $\lambda$, i.e. $a/2$. For this value we can write an inequality analogous to (7):

$$\begin{aligned} (x^{a/2} - y^{a/2})^2 &= x^a + y^a - 2x^{a/2}\, y^{a/2} \geq 0 \\ x^{a/2} &\leq \tfrac{1}{2} y^{a/2} + \tfrac{1}{2} y^{-a/2}\, x^a \,. \end{aligned} \tag{9}$$

From (8) we know that the last factor, $x^a$, can be substituted by an expression that is always larger or equal to this factor. Therefore, the substitution does not affect the sign of inequality (9). The substitution yields

$$\begin{aligned} x^{a/2} &\leq \tfrac{1}{2} y^{a/2} + \tfrac{1}{2} y^{-a/2}\, ((1-a)\, y^a + a y^{a-1}\, x) \\ &= \tfrac{1}{2} y^{a/2} + \tfrac{1}{2} (1-a) y^{a/2} + \tfrac{1}{2} a\, y^{a/2-1}\, x \\ &= (1-a/2) y^{a/2} + \tfrac{1}{2} a\, y^{a/2-1}\, x \end{aligned} \tag{10}$$

This is the same expression as is given in (8) for $\lambda = a/2$. Q.E.D.

Theorem 3 was proven for integer values of $\kappa$ only. However, a similar majorization inequality can be developed for any real valued $\lambda$ between zero and unity. It is derived in a different way by Hardy, Littlewood and Pólya (1967). The inequality (8) can be applied directly to majorize the numerator of the tunneling function, which is done in the next section.

### 3.4. *The tunneling algorithm.*

The minimization method for the tunneling function (4) consists of two concepts. The first one, parametric programming, is used to minimize a ratio of two functions by means of a help function $F(q,\mathbf{X})$. In section 3.1 it was shown that convergence is retained when an $\mathbf{X}^+$ is found that yields a lower value of the help function. The concept of majorization is used to find such $\mathbf{X}^+$. The numerator of the tunneling function is a product of two functions, that can be majorized by the sum of the squares of each function (see section 3.2). Therefore, we change to minimizing the square root of the tunneling function, so that after majorization we have a sum of the two functions. Note that taking the square root (or any other monotonic transformation) does not change the position of the minimum $\tau_4(\mathbf{X})$ or, in other words, minimizing $\tau_4(\mathbf{X})$ is equivalent to minimizing $\sqrt{\tau_4(\mathbf{X})}$. In the remaining part of this section we change to

$$\sqrt{\tau_4(\mathbf{X})} = \frac{\sqrt{N(\mathbf{X})}\,\sqrt{1 + P(\mathbf{X})}}{\sqrt{P(\mathbf{X})}} = |\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*)|^\lambda \frac{\sqrt{\|D(\mathbf{X}^*)-D(\mathbf{X})\|^2 + 1}}{\|D(\mathbf{X}^*)-D(\mathbf{X})\|}, \tag{11}$$

which results in the help function

$$F(q,\mathbf{X}) = \sqrt{N(\mathbf{X})}\,\sqrt{1 + P(\mathbf{X})} - q\,\sqrt{P(\mathbf{X})}\,, \tag{12}$$

where $q$ equals $\sqrt{\tau_4(\mathbf{Y})}$ if $\mathbf{Y}$ is the previous configuration from the iterative process. In the next sections the various steps are taken that are needed to majorize (12).

#### 3.4.1. *Majorizing* $\sqrt{N(\mathbf{X})}\,\sqrt{1 + P(\mathbf{X})}$.

The outer majorization of the product $\sqrt{N(\mathbf{X})}\,\sqrt{1 + P(\mathbf{X})}$ is given by (6) and can be directly applied

$$\sqrt{N(\mathbf{X})}\,\sqrt{1 + P(\mathbf{X})} \leq \tfrac{1}{2}\sqrt{\frac{1 + P(\mathbf{Y})}{N(\mathbf{Y})}}\,N(\mathbf{X}) + \tfrac{1}{2}\sqrt{\frac{N(\mathbf{Y})}{1 + P(\mathbf{Y})}}\,(1 + P(\mathbf{X}))\ . \tag{13}$$

Further, the functions $N(\mathbf{X})$ and $P(\mathbf{X})$ need to be majorized themselves. Let us focus on the majorization of $N(\mathbf{X})$ first. This function can be expressed by raising $(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2$ to the power $\lambda$, with $0 < \lambda < 1$. From (8) the inequality

$$(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^{2\lambda} \le (1-\lambda)(\sigma(\mathbf{Y})-\sigma(\mathbf{X}^*))^{2\lambda} + \lambda(\sigma(\mathbf{Y})-\sigma(\mathbf{X}^*))^{2(\lambda-1)}(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2$$
$$\le (1-\lambda)\mathrm{N}(\mathbf{Y}) + \lambda\mathrm{N}(\mathbf{Y})^{1-1/\lambda}(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2 \quad (14)$$

is obtained. In Groenen (1990) it was shown that $(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2$ can be majorized by

$$(\sigma(\mathbf{X}) - \sigma(\mathbf{X}^*))^2 \le c_1 + n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2\left[1 - \frac{\sigma(\mathbf{X}^*)}{\sigma(\mathbf{Y})}\right]\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathbf{Y})\mathbf{Y} - 2\frac{n\,\sigma(\mathbf{X}^*)}{\sigma(\mathbf{Y})}\mathrm{tr}\mathbf{X}'\mathbf{Y}$$
$$= c_1 + n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2[1-\phi]\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathbf{Y})\mathbf{Y} - 2\phi n\,\mathrm{tr}\mathbf{X}'\mathbf{Y}. \quad (15)$$

where

$$c_1 = \sigma^2(\mathbf{X}^*) + (1-2\phi) + \phi\Sigma_{ij}\mathrm{d}_{ij}(\mathbf{Y})\delta_{ij} \quad . \quad (16)$$

To simplify the notation, we shall use an indexed $c$ to indicate all terms that are not a function of **X**. Inserting (15) in (14) yields

$$\mathrm{N}(\mathbf{X}) \le (1-\lambda)\mathrm{N}(\mathbf{Y}) + \lambda\mathrm{N}(\mathbf{Y})^{1-1/\lambda}[c_1 + n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2(1-\phi)\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathbf{Y})\mathbf{Y} - 2\phi n\,\mathrm{tr}\mathbf{X}'\mathbf{Y}] \quad . \quad (17)$$

The majorization of $\mathrm{P}(\mathbf{X})$ is analogous to the majorization of the STRESS function itself. We only give the result here and refer to De Leeuw and Heiser (1980) or Groenen (1990) for details. The function $\mathrm{P}(\mathbf{X})$ can be majorized by the inequality

$$\mathrm{P}(\mathbf{X}) \le n\mathrm{tr}\mathbf{X}^{*\prime}\mathbf{X}^* + n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathrm{D}(\mathbf{X}^*),\mathbf{Y})\mathbf{Y} \quad ,$$

where $\mathrm{B}(\mathrm{D}(\mathbf{X}^*),\mathbf{Y})$ has off diagonal elements $b_{ij} = -\mathrm{d}_{ij}(\mathbf{X}^*)/\mathrm{d}_{ij}(\mathbf{Y})$ and diagonal elements $b_{ii} = -\Sigma_{i\neq j}b_{ij}$.

To conclude this section the previous results are substituted in (13) so that $\sqrt{\mathrm{N}(\mathbf{X})}\,\sqrt{1+\mathrm{P}(\mathbf{X})}$ is majorized by a function that has only quadratic and linear terms in **X**. The substitution yields

$$\sqrt{\mathrm{N}(\mathbf{X})}\,\sqrt{1+\mathrm{P}(\mathbf{X})} \le \tfrac{1}{2}\lambda[1+\mathrm{P}(\mathbf{Y})]^{1/2}\mathrm{N}(\mathbf{Y})^{1/2-1/\lambda}[n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2(1-\phi)\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathbf{Y})\mathbf{Y} - 2\phi n\,\mathrm{tr}\mathbf{X}'\mathbf{Y}]$$
$$+ \tfrac{1}{2}\mathrm{N}(\mathbf{Y})^{1/2}[1+\mathrm{P}(\mathbf{Y})]^{-1/2}[n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathrm{D}(\mathbf{X}^*),\mathbf{Y})\mathbf{Y}] + c_2 \quad . \quad (18)$$

In section 3.4.3 (18) is used in the majorization of $\mathrm{F}(q,\mathbf{X})$.

*3.4.2. Majorizing* $-\sqrt{\mathrm{P}(\mathbf{X})}$.

The term $-q\sqrt{\mathrm{P}(\mathbf{X})}$ can be expressed as $-q\|\mathbf{d}^* - \mathbf{d}\|$. Such a negative Euclidean norm can be majorized using the Cauchy-Schwartz inequality, i.e.

$$(\mathbf{d}^* - \mathbf{d})'(\mathbf{d}^* - \mathbf{d}_\mathrm{Y}) \leq \|\mathbf{d}^* - \mathbf{d}\| \, \|\mathbf{d}^* - \mathbf{d}_\mathrm{Y}\| , \tag{19}$$

where $\mathbf{d}_\mathrm{Y}$ is a vector of distances between the points of configuration $\mathbf{Y}$. The inequality becomes an equality if $\mathbf{d}_\mathrm{Y}$ equals $\mathbf{d}$. Dividing both sides of (19) by $\|\mathbf{d}^* - \mathbf{d}_\mathrm{Y}\|$ and multiplying with $-1$ yields

$$\begin{aligned} -\|\mathbf{d}^* - \mathbf{d}\| &\leq -\frac{(\mathbf{d}^* - \mathbf{d})'(\mathbf{d}^* - \mathbf{d}_\mathrm{Y})}{\|\mathbf{d}^* - \mathbf{d}_\mathrm{Y}\|} = \frac{-\mathbf{d}^{*\prime}\mathbf{d}^* + \mathbf{d}^{*\prime}\mathbf{d}_\mathrm{Y} + \mathbf{d}'\mathbf{d}^* - \mathbf{d}'\mathbf{d}_\mathrm{Y}}{\|\mathbf{d}^* - \mathbf{d}_\mathrm{Y}\|} \\ &= \frac{\mathbf{d}'\mathbf{d}^* - \mathbf{d}'\mathbf{d}_\mathrm{Y}}{\|\mathbf{d}^* - \mathbf{d}_\mathrm{Y}\|} + c_3 . \end{aligned} \tag{20}$$

The term $\mathbf{d}'\mathbf{d}^*$, in sum notation $\frac{1}{2}\Sigma_{ij}\mathrm{d}_{ij}(\mathbf{X}^*)\,\mathrm{d}_{ij}(\mathbf{X})$, can itself be majorized. More specifically, $\mathrm{d}_{ij}(\mathbf{X})$ is majorized by $\frac{1}{2}\mathrm{d}_{ij}(\mathbf{Y}) + \frac{1}{2}\mathrm{d}^2_{ij}(\mathbf{X})/\mathrm{d}_{ij}(\mathbf{Y})$, using an inequality derived from (7). Heiser (1991) notes that when $\mathrm{d}_{ij}(\mathbf{Y})$ is zero, this majorizing function does not work. Therefore he majorizes $\mathrm{d}_{ij}(\mathbf{X})$ by $\frac{1}{2}\varepsilon + \frac{1}{2}\mathrm{d}^2_{ij}(\mathbf{X})/\varepsilon$ whenever $\mathrm{d}_{ij}(\mathbf{Y})$ is smaller than a small positive constant $\varepsilon$. In the sequel, we shall use this implicitly when necessary. Multiplying $\mathrm{d}_{ij}(\mathbf{X})$ and its majorizing function by $\mathrm{d}_{ij}(\mathbf{X}^*)$, and summing over all $ij$, yields

$$\begin{aligned} \mathbf{d}'\mathbf{d}^* &\leq \tfrac{1}{4}\Sigma_{ij}\mathrm{d}_{ij}(\mathbf{X}^*)\,\mathrm{d}_{ij}(\mathbf{Y}) + \tfrac{1}{4}\Sigma_{ij}\frac{\mathrm{d}_{ij}(\mathbf{X}^*)}{\mathrm{d}_{ij}(\mathbf{Y})}\mathrm{d}^2_{ij}(\mathbf{X}) \\ &= c_4 + \tfrac{1}{2}\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathrm{D}(\mathbf{X}^*),\mathbf{Y})\mathbf{X} . \end{aligned} \tag{21}$$

The term $-\mathbf{d}'\mathbf{d}_\mathrm{Y}$ can be majorized using the SMACOF theory, where the term $-2\mathbf{d}'\boldsymbol{\delta}$ is majorized by $-2\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathbf{Y})\mathbf{Y}$. Thus simply substituting $\delta_{ij}$ by $\mathrm{d}_{ij}(\mathbf{Y})$ gives the inequality

$$-\mathbf{d}'\mathbf{d}_\mathrm{Y} \leq -\mathrm{tr}\mathbf{X}'\mathrm{B}(\mathrm{D}(\mathbf{Y}),\mathbf{Y})\mathbf{Y} .$$

However, the elements of the matrix $\mathrm{B}(\mathrm{D}(\mathbf{Y}),\mathbf{Y})$ are $b_{ij} = -\mathrm{d}_{ij}(\mathbf{Y})/\mathrm{d}_{ij}(\mathbf{Y}) = -1$. This implies that $\mathrm{B}(\mathrm{D}(\mathbf{Y}),\mathbf{Y})$ is equal to $n\mathbf{I} - \mathbf{11}'$. Since $\mathbf{Y}$ is a centered matrix we have $\mathbf{11}'\mathbf{Y} = \mathbf{0}$, so that the inequality can be simplified even more to

$$-\mathbf{d}'\mathbf{d}_\mathrm{Y} \leq -n\mathrm{tr}\mathbf{X}'\mathbf{Y} . \tag{22}$$

Combining the inequalities (20), (21) and (22) shows that the term $-q\sqrt{P(\mathbf{X})}$ can be majorized:

$$-q\sqrt{P(\mathbf{X})} \qquad \leq \tfrac{1}{2}q\,\frac{\mathrm{tr}\mathbf{X}'B(D(\mathbf{X}^*),\mathbf{Y})\mathbf{X} - 2n\mathrm{tr}\mathbf{X}'\mathbf{Y}}{\|\mathbf{d}^* - \mathbf{d}_{\mathbf{Y}}\|} + c_5\ . \tag{23}$$

The results from this section and the previous one are combined in the next section to present the update formula.

### *3.4.3. The update.*

From section 3.1 it is known that the tunneling function can be minimized iteratively by finding an update that yields a lower value of the auxiliary function $F(q,\mathbf{X})$. The majorization inequalities are used to find such an update. Expressing $F(q,\mathbf{X})$ as $\sqrt{N(\mathbf{X})}\,\sqrt{1+P(\mathbf{X})} - q\sqrt{P(\mathbf{X})}$, where $q$ equals $\sqrt{\tau_4(\mathbf{X})}$, let us combine the majorization results (18) and (23) into

$$\begin{aligned} e\,F(q,\mathbf{X}) \leq\ & \lambda[1+P(\mathbf{Y})]N(\mathbf{Y})^{1-1/\lambda}\ [n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2(1-\phi)\mathrm{tr}\mathbf{X}'B(\mathbf{Y})\mathbf{Y} - 2\phi n\,\mathrm{tr}\mathbf{X}'\mathbf{Y}] + \\ & + N(\mathbf{Y})\ [n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2\mathrm{tr}\mathbf{X}'B(D(\mathbf{X}^*),\mathbf{Y})\mathbf{Y}] + \\ & + \tau_4(\mathbf{Y})\ [\mathrm{tr}\mathbf{X}'B(D(\mathbf{X}^*),\mathbf{Y})\mathbf{X} - 2n\mathrm{tr}\mathbf{X}'\mathbf{Y}] + c_6 \end{aligned} \tag{24}$$

where $e$ is the factor $2\sqrt{N(\mathbf{Y})}\,\sqrt{1+P(\mathbf{Y})}$ introduced here to simplify the notation. Further, let $\alpha$ equal $\lambda[1+P(\mathbf{Y})]N(\mathbf{Y})^{1-1/\lambda}$, $\beta$ equal $N(\mathbf{Y})$ and $\gamma$ equal $\tau_4(\mathbf{Y})$. Then the preceding inequality can be rewritten as

$$\begin{aligned} e\,F(q,\mathbf{X}) \leq\ & \alpha\ [n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2(1-\phi)\mathrm{tr}\mathbf{X}'B(\mathbf{Y})\mathbf{Y} - 2\phi n\,\mathrm{tr}\mathbf{X}'\mathbf{Y}] \\ & + \beta\ [n\mathrm{tr}\mathbf{X}'\mathbf{X} - 2\mathrm{tr}\mathbf{X}'B(D(\mathbf{X}^*),\mathbf{Y})\mathbf{Y}] \\ & + \gamma\ [\mathrm{tr}\mathbf{X}'B(D(\mathbf{X}^*),\mathbf{Y})\mathbf{X} - 2n\mathrm{tr}\mathbf{X}'\mathbf{Y}] + c_6 \\ =\ & \mathrm{tr}\mathbf{X}'[(\alpha+\beta)n\mathbf{I} + \gamma B(D(\mathbf{X}^*),\mathbf{Y})]\mathbf{X} \\ & - 2\mathrm{tr}\mathbf{X}'[B(\alpha(1-\phi)\Delta+\beta D(\mathbf{X}^*),\mathbf{Y})\mathbf{Y} + (\alpha\phi+\gamma)n\mathbf{Y}] + c_6 \\ =\ & \mathrm{tr}\mathbf{X}'\mathbf{V}\mathbf{X} - 2\mathrm{tr}\mathbf{X}'\mathbf{Z} + c_6 \end{aligned} \tag{25}$$

where the matrix

$$\mathbf{V} = (\alpha+\beta)n\mathbf{I} + \gamma B(D(\mathbf{X}^*),\mathbf{Y})\ , \text{ and}$$
$$\mathbf{Z} = B(\alpha(1-\phi)\Delta + \beta D(\mathbf{X}^*),\mathbf{Y})\mathbf{Y} + (\alpha\phi+\gamma)n\mathbf{Y}.$$

Thus F($q$,**X**) is majorized by a quadratic function of **X**. When **V** is positive definite, the minimum of a quadratic function can be found in one step by setting the gradient to zero, i. e.

$$\nabla(\mathrm{tr}\mathbf{X}'\mathbf{V}\mathbf{X} - 2\mathrm{tr}\mathbf{X}'\mathbf{Z} + c_6) = 2\mathbf{V}\mathbf{X} - 2\mathbf{Z} = \mathbf{0}, \quad (26)$$

which implies that $\mathbf{VX} = \mathbf{Z}$ or $\mathbf{X} = \mathbf{V}^{-1}\mathbf{Z}$. Note that **V** is of full rank (it is the sum of the full rank matrix $(\alpha+\beta)n\mathbf{I}$ and the matrix $\gamma\mathrm{B}(\mathrm{D}(\mathbf{X}^*),\mathbf{Y})$ of rank $n-1$) so that its inverse exists. Moreover, because both $(\alpha+\beta)n\mathbf{I}$ and $\gamma\mathrm{B}(\mathrm{D}(\mathbf{X}^*),\mathbf{Y})$ are positive semi-definite, so is their sum. Thus (26) can be used to minimize (25) in one step.

This leads us to the update formula needed in step 2 of the parametric programming algorithm: the update $\mathbf{X}^+$, given by

$$\mathbf{X}^+ = [(\alpha+\beta)n\mathbf{I} + \gamma\mathrm{B}(\mathrm{D}(\mathbf{X}^*),\mathbf{Y})]^{-1}[\mathrm{B}(\alpha(1-\phi)\Delta+\beta\mathrm{D}(\mathbf{X}^*),\mathbf{Y})\mathbf{Y} + (\alpha\phi+\gamma)n\mathbf{Y}] \ , \quad (27)$$

ensures a lower value of $\tau_4(\mathbf{X})$ in every iteration until convergence is attained.

## 4. Empirical results.

After the theoretical considerations in the previous sections, we present here the results of two empirical studies. The examples are not exhaustive, but illustrate the tunneling algorithm. The first example is a constant dissimilarity matrix of four points. The second example originates from Robinson (1951), and is also analyzed by Hubert and Arabie (1986); the data come from the Mani collection of archaelogical deposits.

The first small example, studied extensively by De Leeuw (1988), concerns a 4×4 dissimilarity matrix, with all dissimilarities equal to $1/\sqrt{6}$. He reports three stationary two-dimensional configurations: four points on a line with $\sigma(\mathbf{X}_1) = 0.4082482905$, three points in the corners of an equilateral triangle and a point in the centroid with $\sigma(\mathbf{X}_2) = 0.2588190451$, four points in the corners of a square with $\sigma(\mathbf{X}_3) = 0.1691019787$. (Note that De Leeuw reports the square of the STRESS value, i.e. $\sigma^2(\mathbf{X})$.)

We start the tunneling algorithm from a stationary point defined by a one-dimensional scaling solution, see Figure 2a, which is in fact a saddle point. The first objective for the tunneling algorithm is to find another configuration with STRESS 0.4082482905. The tunneling algorithm used a starting configuration that is a sum of the one-dimensional scaling solution and a random matrix. The latter is needed to increase the rank of the solution from one to two. The pole strength parameter λ was set to 1/4. After 102 iterations a solution was found with the same STRESS, as can be seen in Figure 1 that shows the history of iterations of the tunneling function values $\tau_4(X)$, the subsequent STRESS values $\sigma(X)$, and the values of the pole $P(X)$.

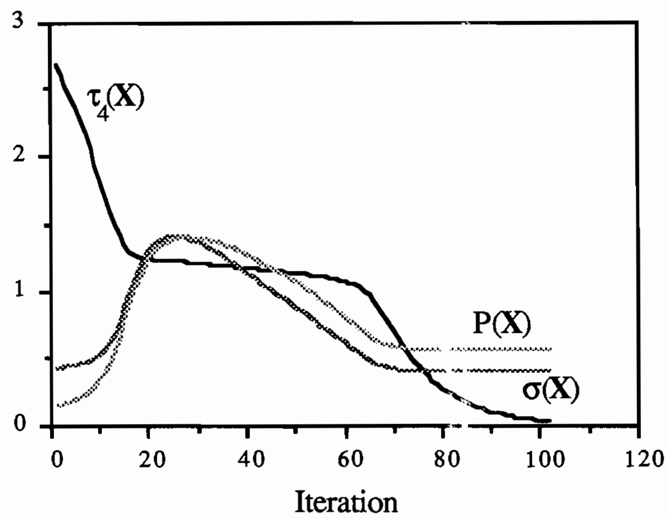


**Fig.1** The sequence of tunneling function values $\tau_4(X)$, STRESS values $\sigma(X)$ and value of the pole $P(X)$ against the iteration number.

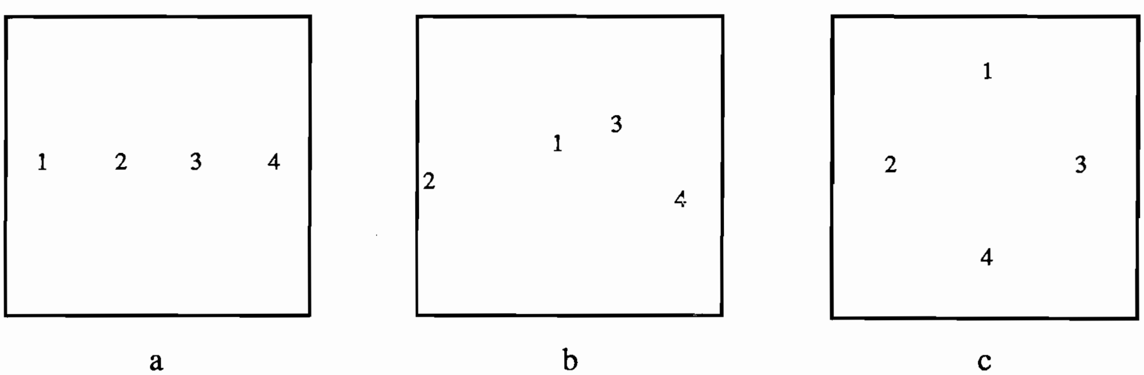


**Fig.1** Three configurations of the constant dissimilarity matrix: a. an unidimensional starting configuration, b. a configuration with the same STRESS as the unidimensional configuration, c. a lower local minimum of the STRESS function.

After the first tunneling step configuration b. of Figure 1 had a STRESS 0.4082483676. This solution differs in the seventh decimal position from the local minimum configuration. The STRESS value of the square configuration (see Figure 1c) is 0.1691019835 which differs only in the eighth decimal place compared with the STRESS reported by De Leeuw.

The second stationary point, an equilateral triangle with centroid (the solution given by Torgerson-Gower scaling), is a local minimum. Starting the tunneling algorithm here yields also the square configuration after one tunneling step and one local optimization step. This leads to several conclusions: a. the tunneling algorithm seems to work for this small example, and b. the square configuration attracts the algorithm strongly enough for both starting configurations.

The second example uses a part of the Mani collection data. The dissimilaties are normalized to have sum of squares $8^2$. A stationary point was found with STRESS 0.88376265 and is shown in Figure 4a. Note that point 6, 7 and 8 are very close to each other. The tunneling algorithm reached after 1162 iterations a configuration with STRESS 0.88376279. Figure 3 displays a summary of the iterations.

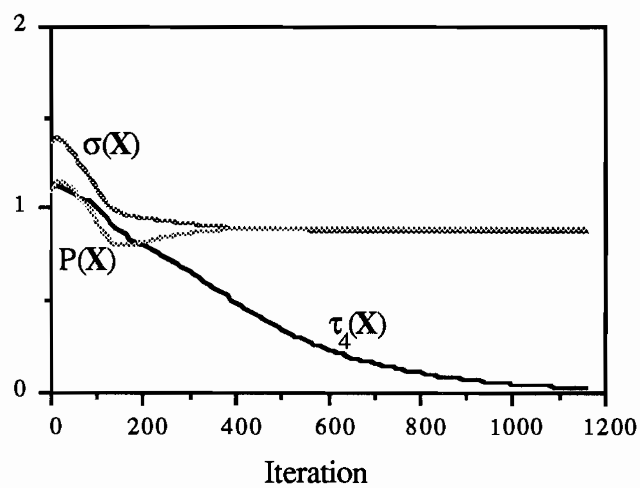


**Fig.3** The sequence of tunneling function values $\tau_4(\mathbf{X})$, STRESS values $\sigma(\mathbf{X})$ and value of the pole $P(\mathbf{X})$ against the iteration number in the first tunneling step on data of the Mani collection.

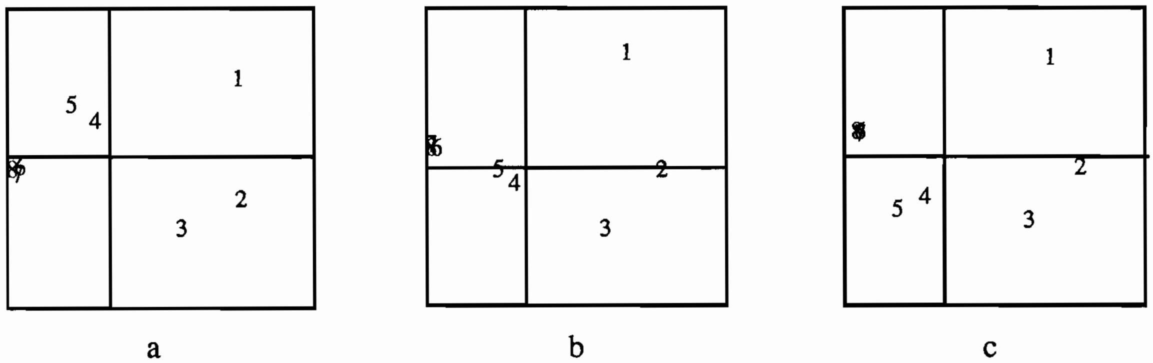


**Fig.4** Three configurations of the Mani collection data: a. stationary starting configuration, b. a configuration with the same STRESS as the starting configuration, c. a new local minimum with lower value of the STRESS function.

After the tunneling step one local search step was performed that yielded the configuration displayed in Figure 4 c with STRESS 0.64722524. The tunneling algorithm was not able to find another configuration with STRESS 0.64722524. The three configurations in Figure 4 show two parts of the configuration: one group with points 1, 2 and 3 and the other group with points 4, 5, 6, 7 and 8. It seems that the second group needed to be reflected along the horizontal axis to obtain lower STRESS.

## 5. Discussion and conclusions.

Our main conclusion is that our tunneling algorithm works, at least for small problems. Several problems have been solved so that the final redefinition of the tunneling function does not lead to trivial solutions. Furthermore, an algorithm is given that minimizes the tunneling function, based on majorization. Finally, two examples are presented that illustrate the tunneling method.

However, several issues need to be addressed in future research. Obviously, a considerable increase of efficiency can be expected from a clever starting configuration in the tunneling step. Furthermore, we need to know how to choose the best value of the pole strength parameter. Also, effective stop criteria for the tunneling algorithm should be developed. Next, the effect of the normalization of the dissimilarities on the tunneling algorithm needs to be considered. Furthermore, the minimization method used for the tunneling function only guarantees a stationary point that is at best a local minimum. Therefore we may expect that for larger dissimilarity matrices the tunneling algorithm can stop at a non-zero value. One way to solve this problem is to introduce extra poles (or a moving pole) as suggested by Levy and Gomez (1985). Further, using majorization as the minimization method implies that majorization inequalities should be found. However, certain functions can be majorized in other ways and the effect of the inequality chosen is not always clear. Finally, it is proven here that parametric programming remains valid when majorization is used. In the algorithm we developed $q$ was updated at every iteration; we may as well perform more inner iterations before updating $q$.

Extensions of the tunneling algorithm can be developed without major problems. It seems fairly straightforward to include weights for each pair of points. We do not know the rate of convergence of the tunneling algorithm yet, but acceleration schemes, like the relaxed update for the SMACOF algorithm proposed by De Leeuw and Heiser (1980), may well be implemented (see Ramsay (1975) for other accalueration devices). Another area of interest is nonmetric scaling, where the dissimilarities are replaced by pseudo-distances $\hat{\mathbf{d}}$ that are monotone with the dissimilarities. Kruskal (1977), De Leeuw (1988) and De Leeuw and Heiser (1980) view the nonmetric scaling STRESS function as a function of $\mathbf{X}$ alone where the problem of finding $\hat{\mathbf{d}}$ is eliminated by innner minimization. Using this interpretation the tunneling algorithm can be applied to the pseudo-distances instead of the dissimilarities to use the tunneling method for nonmetric scaling.

## Appendix A. Notation.

For convenience we summarize here the notation used throughout this report.

| | |
|---|---|
| $n$ | Number of points. |
| $i,j$ | Running index for points, $i,j = 1,\ldots,n$. |
| $p$ | Number of dimensions. |
| $s$ | Running index for dimensions, $s = 1,\ldots,p$. |
| $\mathbf{X}$ | Matrix of coordinates of $n$ points in $p$ dimensions. |
| $\mathbf{X}^*$ | A matrix of coordinates belonging to a local minimum. |
| $\delta_{ij}$ | Dissimilarity between stimuli $i$ and $j$. |
| $\boldsymbol{\delta}$ | Vector of all dissimilarities, of order 1/2 $n(n-1)$. |
| $\Delta$ | Matrix of dissimilarities, sized $n \times n$. |
| $\mathrm{d}_{ij}(\mathbf{X})$ | The Euclidean distance between row $i$ and row $j$ of $\mathbf{X}$. |
| $\mathbf{d}$ | Vector of all Euclidean distances between the rows of $\mathbf{X}$, of order 1/2 $n(n-1)$. |
| $\mathbf{d}^*$ | Vector of all Euclidean distances between the rows of $\mathbf{X}^*$, of order 1/2 $n(n-1)$. |
| $\mathrm{D}(\mathbf{X})$ | Matrix of Euclidean distances between the rows of $\mathbf{X}$. |
| $\mathrm{B}(\mathrm{C};\mathbf{X})$ | Matrix with off diagonal elements $b_{ij} = -c_{ij}/\mathrm{d}_{ij}(\mathbf{X})$ if $\mathrm{d}_{ij}(\mathbf{X}) \neq 0$ and $b_{ij} = 0$ otherwise and diagonal elements $b_{ij} = -\sum_{j \neq i} b_{ij}$. |
| $\mathrm{B}(\mathbf{X})$ | Short form of the matrix $\mathrm{B}(\boldsymbol{\Delta};\mathbf{X})$ |
| $\lVert\mathbf{X}\rVert$ | The Euclidean norm of matrix $\mathbf{X}$; if $\mathbf{X}$ is symmetric $\lVert\mathbf{X}\rVert = \sqrt{\sum_{i<j}^{n} x_{ij}^2}$ and if $\mathbf{X}$ is asymmetric or non-square $\lVert\mathbf{X}\rVert = \sqrt{\sum_{i,j}^{n} x_{ij}^2}$ . |

Two important formulae are given here; the first defines a Euclidean distance, the second defines the STRESS.

$$\mathrm{d}_{ij}(\mathbf{X}) = \sqrt{\sum_{s=1}^{p} (x_{is} - x_{js})^2}$$

$$\sigma(\mathbf{X}) = \sqrt{\frac{1}{2} \sum_{i,j}^{n} (\delta_{ij} - \mathrm{d}_{ij}(\mathbf{X}))^2}$$

**References.**

De Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification, 5*, 163-180.

De Leeuw, J., Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. P.R. Krishnaiah (Ed.), *Multivariate analysis, Vol. V* (pp.501-522). Amsterdam: North-Holland.

Dinkelbach, W. (1967). On nonlinear fractional programming. *Management Science, Vol 13*, 492-498.

Gill, P.E., Murray, W., Wright, M.H. (1981). *Practical optimization*. London: Academic Press.

Groenen, P.J.F. (1990). *The tunneling method applied to multidimensional scaling: progress report 1*. Research report RR 90-03. Leiden: Department of Data Theory.

Hardy, G.H., Littlewood, J.E., Pólya, G. (1967). *Inequalities* (2nd edition). Cambridge: University Press.

Heiser, W.J. (1981). *Unfolding analysis of proximity data*. Doctoral dissertation. Leiden: University of Leiden.

Heiser, W.J. (1987). Correspondence analysis with least absolute residuals. *Computational statistics & Data Analysis, 5*, 337-356.

Heiser, W.J. (1991). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika, 56*, 7-27.

Hubert, L.J., Arabie, P. (1986). Unidimensional scaling and combinatorial optimization. In: J. De Leeuw, W.J. Heiser, J. Meulman and F. Critchley (Eds.), *Multidimensional Data Analysis* (pp.181-196). Leiden: DSWO Press.

Kruskal, J.B. (1977). Multidimensional scaling and other methods for discovering structure. In: K. Enslein, A. Ralston and H.S. Wilf (Ed.), *Statistical methods for digital computers, Vol. III* (pp. 296-339). New York: Wiley.

Levy, A.V., Gomez, S. (1985). The tunneling method applied to global optimization. P.T. Boggs, R.H. Byrd, R.B. Schnabel (Ed.), *Numerical optimization 1984* (pp.213-244).Philadelphia: SIAM

Meulman, J. (1986). *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO press.

Ramsay, J.O. (1975). Solving implicit equations in psychometric data analysis. *Psychometrika, 40*, 337-360.

Robinson, W.S. (1951). A method for chronologically ordering archaeological deposits. *American Antiquity, 16*, 293-301.