SECOND DERIVATIVES OF *STRESS*,
WITH APPLICATIONS

JAN DE LEEUW

ABSTRACT. Meet the abstract. This is the abstract.

1. INTRODUCTION

In (metric, Euclidean) multidimensional scaling (MDS) we minimize *stress*, given by

$$\sigma(X) = \sum_{1\leq i<j\leq n}\sum w_{ij}(\delta_{ij} - d_{ij}(X))^2 \tag{1}$$

over the $n \times p$ *configurations*. In (1) the δ_{ij} are known non-negative *dissimilarities* and the w_{ij} are known non-negative *weights*. We assume, without loss of generality, that the dissimilarities are normalized as $\sum_{1\leq i<j\leq n}\sum w_{ij}\delta_{ij}^2 = 1$.

For computational and analytical reasons it is convenient to use matrix notation to reformulate the MDS loss function. Remember that e_i is the unit vector with element i equal to one and all others elements equal to zero. Let $A_{ij} \stackrel{\Delta}{=} (e_i - e_j)(e_i - e_j)'$ and

$$A_{ij}^{\oplus p} \stackrel{\Delta}{=} \underbrace{A_{ij} \oplus \cdots \oplus A_{ij}}_{p \text{ times}}.$$

Thus $\overline{A}_{ij}$ is block-diagonal, with all p diagonal blocks equal to A_{ij}.

Date: Friday 25th February, 2011 — 14h 51min — Typeset in KEPLER.
Key words and phrases. Template, LaTeX.

Now

$$d_{ij}^2(X) = (e_i - e_j)'XX'(e_i - e_j) = \mathbf{tr}\ X'A_{ij}X = x'\overline{A}_{ij}x,$$

where $x \stackrel{\Delta}{=} \mathbf{vec}(X)$. From now on, we will also use the notation $d_{ij}(x)$.

Let $V \stackrel{\Delta}{=} \sum\sum_{1\leq i<j\leq n} w_{ij}A_{ij}$ and $\overline{V} \stackrel{\Delta}{=} \underbrace{V \oplus \cdots \oplus V}_{p \text{ times}}$. Then

$$\eta^2(x) \stackrel{\Delta}{=} \sum\sum_{1\leq i<j\leq n} w_{ij}d_{ij}^2(x) = x'\overline{V}x.$$

Note that if $w_{ij} = 1$ for all $1 \leq i < j \leq n$ then $V = nI - J$, where J has all elements equal to $+1$.

Finally, let $B(x) \stackrel{\Delta}{=} \sum\sum_{1\leq i<j\leq n} w_{ij}\left(\frac{\delta_{ij}}{d_{ij}(x)}\right)A_{ij}$, and $\overline{B}(x) \stackrel{\Delta}{=} \underbrace{B(x) \oplus \cdots \oplus B(x)}_{p \text{ times}}$. Then

$$\rho(x) \stackrel{\Delta}{=} \sum\sum_{1\leq i<j\leq n} w_{ij}\delta_{ij}d_{ij}(x) = x'\overline{B}(x)x,$$

and thus $\sigma(x) = 1 - 2\rho(x) + \eta^2(x)$.

2. Derivatives

If $d_{ij}(x) > 0$ then the first and second partials are

$$\mathcal{D}d_{ij}(x) = \frac{1}{d_{ij}(x)}\overline{A}_{ij}x,$$

and

$$\mathcal{D}^{(2)}d_{ij}(x) = \frac{1}{d_{ij}(x)}\left\{\overline{A}_{ij} - \frac{\overline{A}_{ij}xx'\overline{A}_{ij}}{x'\overline{A}_{ij}x}\right\}.$$

Of course

$$\mathcal{D}d_{ij}^2(x) = 2\overline{A}_{ij}x,$$

and

$$\mathcal{D}^{(2)}d_{ij}^2(x) = 2\overline{A}_{ij}.$$

Thus

$$\mathcal{D}\sigma(X) = 2(\overline{V} - \overline{B}(x))x,$$

and

$$\mathcal{D}^{(2)}\sigma(X) = 2(\overline{V} - \overline{H}(x)),$$

where

$$\overline{H}(x) = \sum\sum_{1\leq i<j\leq n} w_{ij}\left(\frac{\delta_{ij}}{d_{ij}(x)}\right)\left\{\overline{A}_{ij} - \frac{\overline{A}_{ij}xx'\overline{A}_{ij}}{x'\overline{A}_{ij}x}\right\}.$$

Note that $\overline{H}(x)$ is positive semi-definite, and $\overline{H}(x)x = 0$. Also, at a local minimum, $\overline{H}(x) \lesssim \overline{V}$ in the Loewner sense, i.e. $\overline{V} - \overline{H}(x)$ is positive semi-definite.

3. Applications

3.1. **SMACOF.** The SMACOF algorithm [De Leeuw, 1977; De Leeuw and Heiser, 1980] in this notation computes updates by $x^{(k+1)} = F(x^{(k)})$, where

$$F(x) = \overline{V}^{+}\overline{B}(x)x$$

is the *Guttman Transform* of x, and $\overline{V}^{+}$ is the Moore-Penrose inverse of $\overline{V}$. If $w_{ij} = 1$ for all $1 \leq i < j \leq n$ then $F(x) = \frac{1}{n}\overline{B}(x)x$.

It follows that $\mathcal{D}F(x) = V^{+}\overline{H}(x)$, and thus the convergence rate of the SMACOF algorithm is the largest eigenvalue of $V^{+}\overline{H}(x)$.

3.2. **Newton's Method.** If we write out the updates computed by the standard Newton-Raphson method we find [De Leeuw, 1993]

$$x^{(k+1)} = (I - \overline{V}^{+}\overline{H}(x^{(k)}))^{-1}F(x^{(k)}).$$

More generally we can define a regularized version by defining

$$x^{(k+1)} = (I - \lambda\overline{V}^{+}\overline{H}(x^{(k)}))^{-1}F(x^{(k)}),$$

with $0 \leq \lambda \leq 1$. This is Newton's method for $\lambda = 1$ and it is SMACOF for $\lambda = 0$. Changing λ allows us to move between a globally linearly convergent to a locally quadratically convergent iteration.

3.3. **Sensitivity Analysis.** If we have found a vector $\hat{x}$ where $\mathcal{D}\sigma(\hat{x}) = 0$ then in a neighborhood of that configuration we have

$$\sigma(x) \approx \sigma(\hat{x}) + (x - \hat{x})'[\overline{V} - \overline{H}(\hat{x})](x - \hat{x}).$$

This can be used to draw "sensitivity regions" around configuration points at a local minimum $\hat{x}$. We show how to do this for point i. Suppose δ is a p-element vector with perturbations, and $\hat{x}_i(\delta) = \hat{x} + \mathbf{vec}(e_i\delta')$. Then for each $K \geq 1$ we can draw the concentric ellipsoids

$$\hat{x}_i + \{\delta \mid (\hat{x}_i(\delta) - \hat{x})'[\overline{V} - \overline{H}(\hat{x})](\hat{x}_i(\delta) - \hat{x}) = K\sigma(\hat{x})\}.$$

3.4. **Inverse MDS.** De Leeuw and Groenen [1997]

References

J. De Leeuw. Applications of Convex Analysis to Multidimensional Scaling. In J.R. Barra, F. Brodeau, G. Romier, and B. Van Cutsem, editors, *Recent developments in statistics*, pages 133–145, Amsterdam, The Netherlands, 1977. North Holland Publishing Company.

J. De Leeuw. Fitting Distances by Least Squares. Preprint Series 130, UCLA Department of Statistics, 1993. URL http://preprints.stat.ucla.edu/download.php?paper=130.

J. De Leeuw and P.J.F. Groenen. Inverse Multidimensional Scaling. *Journal of Classification*, 14(3–21), 1997.

J. De Leeuw and W. J. Heiser. Multidimensional Scaling with Restrictions on the Configuration. In P.R. Krishnaiah, editor, *Multivariate Analysis, Volume V*, pages 501–522, Amsterdam, The Netherlands, 1980. North Holland Publishing Company.

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu