

The out-of-sample problem for classical multidimensional scaling

Michael W. Trosset^{a,*}, Carey E. Priebe^b

^a *Department of Statistics, Indiana University, Bloomington, IN 47405, USA*

^b *Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA*

Received 3 February 2007; received in revised form 23 February 2008; accepted 27 February 2008

Available online 4 March 2008

Abstract

Out-of-sample embedding techniques insert additional points into previously constructed configurations. An out-of-sample extension of classical multidimensional scaling is presented. The out-of-sample extension is formulated as an unconstrained nonlinear least-squares problem. The objective function is a fourth-order polynomial, easily minimized by standard gradient-based methods for numerical optimization. Two examples are presented.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Suppose that one measures the pairwise dissimilarities of n objects, obtaining the $n \times n$ dissimilarity matrix $\Delta = [\delta_{ij}]$. In one application, described in Section 5, the objects are three-dimensional hippocampal images and the dissimilarities are obtained by Large Deformation Diffeomorphic Metric Mapping (LDDMM). In this application, there is no *a priori* feature space—the dissimilarities are measured directly, not computed from feature vectors.

Assume that (some of) the objects are labeled, e.g., by the presence or absence of disease, and that one is interested in the extent to which these classes are discriminated by the specified measure of dissimilarity. One possible approach to this classification problem is the following two-stage procedure:

- (1) Apply classical multidimensional scaling (CMDS) to the measured dissimilarities, thereby embedding *all* of the objects in a (Euclidean) representation space.
- (2) In the representation space so constructed, apply linear discriminant analysis (LDA) to the labeled objects.

Because unlabeled objects can be used in the first (but not the second) stage of this procedure, it is an example of *semisupervised learning*. Of course, one might use other embedding procedures in the first stage and other classification procedures in the second stage. The virtues of this paradigm for semisupervised learning from dissimilarity data are discussed in an accompanying paper (Trosset et al., 2008). The fully supervised case, in which all objects are labeled, was studied by Anderson and Robinson (2003), who proposed tests for group differences based on canonical correlations.

* Corresponding author. Tel.: +1 812 856 7824.

E-mail addresses: mtrosset@indiana.edu (M.W. Trosset), cep@jhu.edu (C.E. Priebe).

Our present concern is with a technical difficulty that arises when one uses the above procedure to classify unlabeled *out-of-sample* objects without including them in the embedding stage, the “exclusive approach” to out-of-sample classification described by Trosset et al. (2008). The difficulty, how to insert the out-of-sample objects into the configuration of points that represents the original objects in Euclidean space, is an out-of-sample embedding problem. As explained below, a proper approach to out-of-sample embedding depends on how the original sample was embedded. For some embedding techniques, but not CMDS, a natural out-of-sample extension is obvious. We present a natural out-of-sample extension of CMDS.

Suppose that, having analyzed a sample of n objects, we obtain k additional out-of-sample objects and measure their dissimilarities, from each other and from each of the original n objects. We desire to apply the classifier constructed from the original n objects to the k new objects. We could, of course, construct a new (and presumably better) classifier from all $n + k$ objects, but that is not our present concern. Rather, we are interested in the performance of the original classifier. This problem can arise in various ways. It arises routinely if one begins with KV labeled objects and attempts to estimate the classifier’s misclassification error rate by V -fold cross-validation. Then, each of V classifiers constructed from $n = K(V - 1)$ training objects must be applied to $k = K$ test objects. The technical challenge that one encounters is the out-of-sample embedding problem, how to embed the k test objects in the representation space constructed from the n training objects by CMDS. Once the k out-of-sample objects have been embedded, it is trivial to apply the original LDA classifier to each of the k new points in the original representation space.

If the objects lie in an accessible feature space, then the out-of-sample problem can be circumvented by learning an embedding function from the feature space to the Euclidean representation space. Such parametric embedding techniques are currently in their infancy; an intriguing example is that of Dr. Lim (Hadsell et al., 2006). In our application, the feature space is not accessible—learning begins with dissimilarities. Furthermore, we argue in Trosset et al. (2008) that there are natural reasons for using CMDS in connection with LDA.

The problem of how to embed new points in relation to a previously specified configuration of fixed points has been considered in various contexts. If the computational expense of embedding a large number of objects is prohibitive, then one may elect to construct an initial configuration by embedding a subset of anchor (or landmark) objects, then individually position the remaining objects with respect to the anchor objects. This technique, variously called the method of standards and landmark MDS, was pioneered by Kruskal and Hart (1966). In sensor network localization, one knows a subset of the pairwise distances between various sensors and the locations of a subset of anchor locations; from this information, one attempts to infer the locations of the remaining sensors. See, for example, Wang et al. (2006) and the references therein.

Gower (1968) added a single point by adding a dimension to the configuration, but doing so results in a different representation space. The method of standards is usually implemented as follows. Suppose that $x_1, \dots, x_n \in \mathcal{R}^d$ were constructed from Δ by minimizing

$$\sigma(x_1, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n (\|x_i - x_j\| - \delta_{ij})^2,$$

the raw stress criterion for metric multidimensional scaling. Letting $A = [a_{ij}]$ denote the augmented $(n + k) \times (n + k)$ dissimilarity matrix, the corresponding out-of-sample embedding of the k new objects is obtained by minimizing

$$\sigma(y_1, \dots, y_k) = 2 \sum_{i=1}^n \sum_{j=1}^k (\|x_i - y_j\| - a_{ij})^2 + \sum_{i=1}^k \sum_{j=1}^k (\|y_i - y_j\| - a_{ij})^2.$$

Of course, one might prefer another error criterion, e.g., the raw sstress criterion, which compares squared distances to squared dissimilarities. Notice that simultaneously embedding k new objects is *not* equivalent to individually embedding the same objects one at a time, as individual embedding does not attempt to approximate dissimilarities between pairs of new objects.

Like many possible error criteria for embedding, the raw s/stress criteria are parametrized by the explicit Cartesian coordinates of the embedded configuration. This property makes it easy to fix the locations of certain points, thereby adapting the error criterion for out-of-sample embedding. In contrast, in CMDS the coordinates of the embedded configuration are implicit—the explicit representation is in terms of pairwise inner products. (In modern parlance, CMDS is a kernel method.) Hence, it is not obvious how to construct an out-of-sample embedding that corresponds to CMDS.

CMDS is often described as a spectral method because the configuration that it constructs has a simple representation in terms of the eigenvalues and eigenvectors of a symmetric positive semidefinite matrix. The out-of-sample extension of CMDS proposed by Bengio et al. (2003) is based on the construction of such eigenmaps. In contrast, we extend the underlying optimization problem that is solved by the eigenmap constructed by CMDS. The resulting optimization problem must be solved numerically, but doing so provides an exact solution to the out-of-sample problem for CMDS.

The special case of $k = 1$, i.e., of adding a single point to an existing configuration, was the only case considered by Anderson and Robinson (2003), who restricted attention to leave-one-out cross-validation. In this case, the optimal out-of-sample embedding can be approximated by a simple formula, and the resulting approximate out-of-sample embedding turns out to be identical to the out-of-sample embedding proposed by Anderson and Robinson (2003).

2. Classical multidimensional scaling

Given $x_1, \dots, x_n \in \mathfrak{R}^d$, let $X = [x_1 | \dots | x_n]^t$ denote the corresponding $n \times d$ configuration matrix. The configuration constructed by CMDS is an eigenmap, i.e., X is the square root of an $n \times n$ inner product matrix, $\bar{B} = XX^t$. The defining feature of CMDS is its choice of the \bar{B} from which X is extracted. This choice relies on a beautiful theorem from classical distance geometry.

By definition, a dissimilarity matrix $\Delta_2 = [\delta_{ij}^2]$ is a Euclidean distance matrix (EDM) if and only if there exist $x_1, \dots, x_n \in \mathfrak{R}^p$ such that $\delta_{ij}^2 = \|x_i - x_j\|^2$. The smallest such p is the embedding dimension of the EDM.

Let I denote the $n \times n$ identity matrix, let $e = (1, \dots, 1)^t \in \mathfrak{R}^n$, and let $P = I - ee^t/n$. Notice that P is symmetric and idempotent; Pv is the projection of $v \in \mathfrak{R}^n$ into e^\perp , $P\Delta_2P$ is the “double centering” of Δ_2 , and $P\Delta_2Pe = 0$. Then ...

Theorem 1. A dissimilarity matrix Δ_2 is an EDM with embedding dimension p if and only if the symmetric matrix

$$B = \tau(\Delta_2) = -\frac{1}{2}P\Delta_2P$$

is positive semidefinite and has rank p . Furthermore, if $\Delta_2 = [\delta_{ij}^2]$ is an EDM and

$$B = \tau(\Delta_2) = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} [x_1 \ \dots \ x_n],$$

then $\delta_{ij}^2 = \|x_i - x_j\|^2$.

The map τ has been carefully studied by Critchley (1988). Evidently τ maps Euclidean squared distances to Euclidean inner products. But distances do not depend on location, whereas inner products do. Because $Be = 0$, if Δ_2 is an EDM, then $0 = e^tBe = e^tXX^te = [X^te]^t[X^te]$; hence $X^te = 0$. We conclude that τ maps Euclidean squared distances to the Euclidean inner products of an isometric configuration of points whose centroid is the origin. As formulated by Torgerson (1952), this connection is the basis for CMDS.

Now suppose that we want to embed fallible dissimilarities, $\Delta = [\delta_{ij}]$, in \mathfrak{R}^d . If $\Delta_2 = [\delta_{ij}^2]$ is not an EDM with embedding dimension $\leq d$, then $B = \tau(\Delta_2)$ is not positive semidefinite with rank $\leq d$; hence, we cannot factor B to obtain a d -dimensional configuration of points. CMDS circumvents this difficulty by replacing B with \bar{B} , the nearest (in the sense of Frobenius norm) symmetric positive semidefinite matrix with rank $\leq d$. Thus, CMDS is predicated on the least-squares approximation of fallible inner products with Euclidean inner products. Somewhat remarkably, it turns out that a closed-form solution to this nonconvex optimization problem can be obtained by modifying the eigenvalues of $B = \tau(\Delta_2)$. This fact is one of the main attractions of CMDS; however, it is merely a pleasant consequence of a more fundamental formulation. Our solution to the out-of-sample problem makes critical use of the fundamental formulation, stated explicitly by Mardia (1978), but clearly implicit in Torgerson (1952).

3. Out-of-sample extension

For simplicity, suppose that $k = 1$. This is the case that arises in leave-one-out cross-validation. The subsequent extension to $k > 1$, described in Section 5, is straightforward.

Let $\Delta_2 = [\delta_{ij}^2]$ denote the squared dissimilarities of the original n objects. Let $a_2 \in \mathfrak{R}^n$ denote the squared dissimilarities of the new object from the original n objects. Let

$$A_2 = \begin{bmatrix} \Delta_2 & a_2 \\ a_2^t & 0 \end{bmatrix}. \tag{1}$$

Applying CMDS to A_2 does not solve the out-of-sample problem relative to Δ_2 because applying CMDS to Δ_2 approximates inner products that are centered with respect to the centroid of the original n objects, whereas applying CMDS to A_2 approximates inner products that are centered with respect to the centroid of all $n + 1$ objects, thereby changing the representation of the n original objects. Our solution circumvents this difficulty by preserving the original centering.

Given $w \in \mathfrak{R}^m$, we say that $x_1, \dots, x_m \in \mathfrak{R}^d$ is w -centered if and only if $\sum_{j=1}^m w_j x_j = 0$. For w such that $e^t w \neq 0$, let

$$\tau_w(A_2) = -\frac{1}{2} \left(I - \frac{ew^t}{e^t w} \right) A_2 \left(I - \frac{we^t}{e^t w} \right).$$

Notice that τ_e is the τ of Theorem 1. Then...

Theorem 2. For any $w \in e^\perp$, the $m \times m$ dissimilarity matrix A_2 is an EDM with embedding dimension p if and only if there exists a w -centered spanning set of \mathfrak{R}^p , $\{y_1, \dots, y_m\}$, for which

$$\tau_w(A_2) = [y_i^t y_j].$$

In Theorem 2, the special case of $w = e_n$ is due to Schoenberg (1935) and was independently discovered by Young and Householder (1938). Torgerson (1952, 1958) popularized $w = e$, the special case of Theorem 1. The general case of $w \in e^\perp$ is due to Gower (1982, 1985).

Let $e = (1, \dots, 1)^t \in \mathfrak{R}^n$ and let $f = (e^t, 1)^t \in \mathfrak{R}^{n+1}$. Applying CMDS to Δ_2 entails approximating the fallible inner products $\tau_e(\Delta_2)$, i.e., inner products computed with respect to the centroid of the original n objects. Applying CMDS to A_2 entails approximating the fallible inner products $\tau_f(\Delta_2)$, i.e., inner products computed with respect to the centroid of all $n + 1$ objects. The out-of-sample problem requires us to maintain the original set of inner products. We do so by setting $w = (e^t, 0)$ and approximating the fallible inner products

$$B = \tau_w(A_2) = \begin{bmatrix} \tau_e(\Delta_2) & b \\ b^t & \beta \end{bmatrix}$$

with $x_1, \dots, x_n \in \mathfrak{R}^d$ fixed, resulting in the nonlinear optimization problem

$$\min_{y \in \mathfrak{R}^t} \left\| B - \begin{bmatrix} x_1^t \\ \vdots \\ x_n^t \\ y^t \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n & y \end{bmatrix} \right\|^2 = \min_{y \in \mathfrak{R}^t} 2 \sum_{i=1}^n (b_i - x_i^t y)^2 + (\beta - y^t y)^2. \tag{2}$$

Global solutions of (2) are exact solutions of the out-of-sample problem for CMDS with $k = 1$.

If the term $(\beta - y^t y)^2$ is dropped from the objective function in (2), then the function that remains is convex, with stationary equation $X^t X y = X^t b$. Assuming that X has full rank (otherwise, a smaller d will suffice), $X^t X$ is invertible and $\hat{y} = (X^t X)^{-1} X^t b$, the unique solution of the stationary equation, approximates the optimal out-of-sample embedding defined by (2). The approximate out-of-sample embedding \hat{y} was previously proposed by Anderson and Robinson (2003) for reasons that differ from ours. Notice that their Eq. (7) computes the components of our b .

4. Example 1: Simulated dissimilarity data

We illustrate the potentially dramatic discrepancy between the out-of-sample embedding of a new object and computing an entirely new embedding of all $n + 1$ objects with a simple example. Consider the two-dimensional EDM

$$\Delta_2 = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix},$$

which CMDS embeds (exactly) by placing $n = 4$ points at the vertices of a unit square. Suppose that we wish to embed a new object whose squared dissimilarities from the original objects are $a_2 = (6.2, 4.5, 6.3, 4.5)^t$.

The augmented squared dissimilarity matrix

$$A_2 = \begin{bmatrix} \Delta_2 & a_2 \\ a_2^t & 0 \end{bmatrix}$$

is not an EDM. Applying CMDS to A_2 produces a configuration, X , whose squared interpoint distances are

$$D_2(X) = \begin{bmatrix} 0 & 0.634879 & 0.000839 & 0.634879 & 5.692261 \\ 0.634879 & 0 & 0.656991 & 2.000000 & 4.574696 \\ 0.000839 & 0.656991 & 0 & 0.656991 & 5.831299 \\ 0.634879 & 2.000000 & 0.656991 & 0 & 4.574696 \\ 5.692261 & 4.574696 & 5.831299 & 4.574696 & 0 \end{bmatrix}.$$

This construction does not insert the fifth object into the original representation, as is evident from the fact that the original $n = 4$ objects are no longer positioned at the vertices of a unit square.

In contrast, Y , the configuration that solves the out-of-sample problem for $d = 2$, has squared interpoint distances

$$D_2(Y) = \begin{bmatrix} 0 & 1 & 2 & 1 & 1.602608 \\ 1 & 0 & 1 & 2 & 4.392919 \\ 2 & 1 & 0 & 1 & 7.183230 \\ 1 & 2 & 1 & 0 & 4.392919 \\ 1.602608 & 4.392919 & 7.183230 & 4.392919 & 0 \end{bmatrix}.$$

This solution was obtained by numerical optimization of (2).¹ A number of initial y were tried, each leading to the same solution. This embedding does preserve the representation of the original $n = 4$ objects as the vertices of a unit square.

In constructing X , each of the $n + 1 = 5$ points is free to vary. In constructing Y , only the fifth point is free to vary. It is not surprising that X and Y differ, but the extent to which the relative locations of the fifth point differ is eye-opening. For example, the Euclidean distance between y and x_1 is $\sqrt{5.692261} \doteq 2.386$ in X , whereas it is only $\sqrt{1.602608} \doteq 1.266$ in Y . Because it is able to reposition the original points, CMDS of A_2 results in a better approximation of the squared dissimilarities in a_2 than is possible if the original points are fixed. Thus, assessments of how the original embedding performs on a new point can easily mislead if a new embedding is recomputed using all $n + 1$ points.

5. Example 2: Hippocampal dissimilarity data

Finally, we demonstrate out-of-sample CMDS on dissimilarities derived from 101 elderly subjects. Miller et al. (submitted for publication) attempted to differentiate demented and normal subjects on the basis of hippocampal shape. Asymmetric dissimilarities were obtained by

¹ The R functions used to obtain the results reported in Sections 4 and 5 are available from the first author.

- (1) Scanning individual whole brain structure using high-resolution T1-weighted structural MRI (magnetic resonance imaging).
- (2) Segmenting the scans using FreeSurfer (see <http://surfer.nmr.mgh.harvard.edu/fswiki> for documentation and citations).
- (3) Measuring asymmetric pairwise dissimilarity by large deformation diffeomorphic metric mapping (LDDMM), as described by Beg et al. (2005).

These data were obtained through the Biomedical Informatics Research Network, described at <http://www.rbirn.net>. Step (1) was performed at Washington University, step (2) at the Martinos Center at Massachusetts General Hospital, and step (3) at the Center for Imaging Science at Johns Hopkins University. Step (3) was performed separately for both left and right hippocampi.

The resulting asymmetric dissimilarities can be combined in various ways, to potentially different effect on the performance of the subsequent classifier. Here, to illustrate out-of-sample embedding, we symmetrized the left and right dissimilarities by averaging, then summed left and right to obtain a single dissimilarity for each pair of subjects. In Trosset et al. (2008), we illustrate a more elaborate approach that retains separate left and right (symmetrized) dissimilarities.

The data described above were not collected concurrently. Initially, the brain scans of $n = 45$ subjects were available for LDDMM and subsequent analysis. These data were used to train a classifier that was subsequently tested on the other $k = 56$ subjects. If classification necessitates embedding by CMDS, as in Trosset et al. (2008), then out-of-sample embedding of the $k = 56$ test subjects is necessary in order to fairly evaluate the classifier's performance. Here, we demonstrate the out-of-sample embedding and compare it to the embedding obtained by applying CMDS to the entire data set.

As in Section 3, let Δ_2 denote the squared dissimilarities of the original n objects. In analogy to (1), let A_2 denote the squared dissimilarities of all $n + k$ objects. Let $e = (1, \dots, 1)^t \in \mathfrak{R}^n$, let $f = (1, \dots, 1)^t \in \mathfrak{R}^{n+k}$, and let $w = (e^t, 0, \dots, 0)^t \in \mathfrak{R}^{n+k}$. Then, factoring $\tau_e(\Delta_2)$ gives the CMDS embedding of the original n objects and factoring $\tau_f(A_2)$ gives the CMDS embedding of all $n + k$ objects.

Let $X = [x_1 \cdots x_n]^t$ denote the $n \times d$ configuration matrix obtained by factoring $\tau_e(\Delta_2)$. Let $Y = [y_1 \cdots y_k]^t$ denote a $k \times d$ configuration matrix that is free to vary. Then the out-of-sample embedding is obtained by computing

$$B = \tau_w(A_2) = \begin{bmatrix} \tau_e(\Delta_2) & B_{xy} \\ B_{xy}^t & B_{yy} \end{bmatrix},$$

then choosing Y to minimize

$$\left\| B - \begin{bmatrix} X \\ Y \end{bmatrix} \begin{bmatrix} X^t & Y^t \end{bmatrix} \right\|^2 = 2 \|B_{xy} - XY^t\|^2 + \|B_{yy} - YY^t\|^2. \quad (3)$$

For $d = 2$, the CMDS embedding of all 101 subjects is displayed in Fig. 1. The bias between the initial $n = 45$ subjects and the subsequent $k = 56$ subjects is striking.² This bias complicates the problem of training a classifier on the former sample and testing it on the latter. This difficulty, however, which is addressed in Miller et al. (submitted for publication), is not germane to our present concern with out-of-sample embedding.

Fig. 2 displays the out-of-sample embedding for $d = 2$. In this representation, the initial $n = 45$ subjects were embedded by CMDS, then the subsequent $k = 56$ subjects were embedded relative to the initial subjects by minimizing (3). It is evident that this configuration is not the configuration constructed by embedding all 101 subjects by CMDS. In particular, the configuration of the initial $n = 45$ subjects (filled dots) in Fig. 2 is the configuration constructed by embedding only those subjects by CMDS. That configuration is not preserved in Fig. 1.

6. Discussion

If, as in Anderson and Robinson (2003) and Trosset et al. (2008), a classifier is constructed from an embedded representation of the data, then evaluation of the classifier's performance requires the ability to embed new data in

² In fact, the segmentation methodology used in step (2) had been modified in the interim.

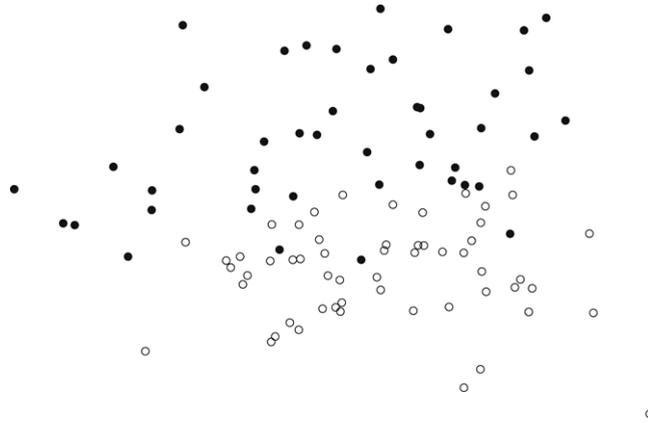


Fig. 1. Two-dimensional embedding of all 101 subjects by CMDS. The initial $n = 45$ subjects are denoted by filled dots; the subsequent $k = 56$ subjects are denoted by unfilled dots.

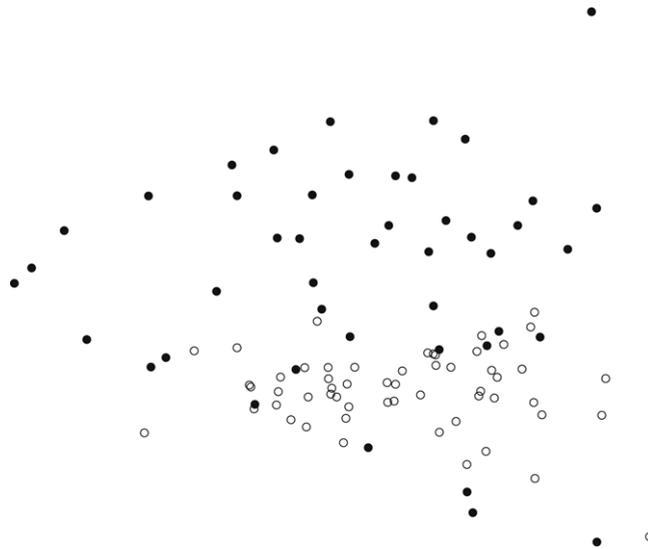


Fig. 2. Two-dimensional embedding by out-of-sample CMDS. First, the initial $n = 45$ subjects (filled dots) were embedded by CMDS; then the subsequent $k = 56$ subjects (unfilled dots) were embedded relative to the initial subjects.

the same Euclidean representation. This is the out-of-sample problem for multidimensional scaling. For embedding methods that are parametrized by the Cartesian coordinates of the points in the Euclidean representation, it is straightforward to fix some points and vary others in relation to them. We have demonstrated how to accomplish the same feat in the case of CMDS, which is parametrized by inner products rather than by Cartesian coordinates.

The key to the out-of-sample extension of CMDS is conceiving of CMDS not as a spectral technique, but as a least-squares technique that happens to have a spectral solution. This perspective leads to our formulation of out-of-sample CMDS as an unconstrained nonlinear least-squares problem. This problem does not have a spectral solution, but the objective function is a fourth-order polynomial that is easily minimized by standard gradient-based methods for numerical optimization. In our experience to date, nonglobal minimizers have not been an undue burden.

Acknowledgments

This research was supported by the Center for Imaging Science at Johns Hopkins University and by grants from the Office of Naval Research. We are grateful to Michael I. Miller and Wendy L. Martinez. Youngser Park facilitated the calculations in Section 5. Brent Castle discovered an error in the calculations reported in an earlier version of Section 4.

References

- Anderson, M.-J., Robinson, J., 2003. Generalized discriminant analysis based on distances. *Australian & New Zealand Journal of Statistics* 45, 301–318.
- Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mapping via geodesic flows of diffeomorphisms. *International Journal of Computer Vision* 61, 139–157.
- Bengio, Y., Paiement, J. F., Vincent, P., 2003. Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering. Technical Report 1238. Département d'Informatique et Recherche Opérationnelle, Université de Montréal, Montréal, Québec, Canada, July 2003.
- Critchley, F., 1988. On certain linear mappings between inner-product and squared-distance matrices. *Linear Algebra and Its Applications* 105, 91–107.
- Gower, J.C., 1968. Adding a point to vector diagrams in multivariate analysis. *Biometrika* 55, 582–585.
- Gower, J.C., 1982. Euclidean distance geometry. *Mathematical Scientist* 7, 1–14.
- Gower, J.C., 1985. Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra and Its Applications* 67, 81–97.
- Hadsell, R., Chopra, S., LeCun, Y., 2006. Dimensionality reduction by learning an invariant mapping. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society. Available at: <http://yann.lecun.com/exdb/publis/>.
- Kruskal, J.B., Hart, R.E., 1966. A geometric interpretation of diagnostic data from a digital machine: Based on a study of the Morris, Illinois electronic central office. *Bell System Technical Journal* 45, 1299–1338.
- Mardia, K.V., 1978. Some properties of classical multi-dimensional scaling. *Communications in Statistics—Theory and Methods* A7, 1233–1241.
- Miller, M.I., et al. (Morphometry BIRN), 2008. Collaborative computational anatomy: an MRI morphometry study of the human brain via diffeomorphic metric mapping. *Human Brain Mapping* (submitted for publication).
- Schoenberg, I.J., 1935. Remarks to Maurice Fréchet's article *Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicable sur l'espace de Hilbert*. *Annals of Mathematics* 38, 724–732.
- Torgerson, W.S., 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* 17, 401–419.
- Torgerson, W.-S., 1958. *Theory and Methods of Scaling*. John Wiley & Sons, New York.
- Trosset, M.W., Priebe, C.E., Park, Y., Miller, M.I., 2008. Semisupervised learning from dissimilarity data. *Computational Statistics and Data Analysis*.
- Wang, Z., Zheng, S., Boyd, S., Ye, Y., 2006. Further relaxations of the SDP approach to sensor network localization. Available at: <http://www.stanford.edu/yyye+>, September 2006.
- Young, G., Householder, A.S., 1938. Discussion of a set of points in terms of their mutual distances. *Psychometrika* 3, 19–22.