Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Drawing and Proximity Analysis

Lisha Chen and Andreas Buja

Yale University and University of Pennsylvania

May 29, 2008

Abstract

In the past decade there has been a resurgence of interest in nonlinear dimension reduction. Among new proposals are "Local Linear Embedding" (LLE, Roweis and Saul 2000), "Isomap" (Tenenbaum et al. 2000) and Kernel PCA (KPCA, Schölkopf, Smola and Müller 1998), which all construct global low-dimensional embeddings from local affine or metric information. We introduce a competing method called "Local Multidimensional Scaling" (LMDS). Like LLE, Isomap and KPCA, LMDS constructs its global embedding from local information, but it uses instead a combination of MDS and "force-directed" graph drawing. We apply the force paradigm to create localized versions of MDS stress functions with a tuning parameter to adjust the strength of non-local repulsive forces.

We solve the problem of tuning parameter selection with a meta-criterion that measures how well the sets of K-nearest neighbors agree between the data

and the embedding. Tuned LMDS seems to be able to outperform MDS, PCA, LLE, Isomap and KPCA, as illustrated with two well-known image datasets. The meta-criterion can also be used in a pointwise version as a diagnostic tool for measuring the local adequacy of embeddings and thereby detect local problems in dimension reductions.

Key words and phrases: MDS, Local Linear Embedding, LLE, Isomap, Principal Components, PCA, Energy Functions, Force-Directed Layout, Cluster Analysis, Unsupervised Learning

1 INTRODUCTION

Dimension reduction is an essential tool for visualizing high-dimensional data. High dimensionality is one of two possible aspects of largeness of data, meaning that the data have a large number of variables as opposed to cases. High-dimensional data have arisen naturally as one has moved from the analysis of single images or single signals to the analysis of databases of images and signals, so that images and signals are treated as cases and pixel intensities or amplitudes as the variables. The correlations between nearby pixels or time points lend plausibility to intrinsic low dimensionality of the collections of images and signals, and hence to the effectiveness of dimension reduction.

The most common dimension reduction methods are principal component analysis (PCA) and multidimensional scaling (MDS). PCA finds linear combinations of the variables to capture the most variation in multivariate data, while multidimensional scaling (MDS) aims to preserve proximity/distance between pairs of cases. Although widely used, these methods fail to flatten curved, intrinsically low-dimensional manifolds. The (artificial) standard example is the

well-known "Swiss Roll", a two-dimensional spiraling manifold that can be flattened, but from which a successful method needs to eliminate the dimensions taken up by curvature. This cannot be achieved with PCA and MDS as both attempt to preserve global structure.

One of the newer methods capable of flattening manifolds, called "local linear embedding" or *LLE* (Roweis and Saul 2000), is a novel idea: it attempts to *preserve local affine structure* by representing each data point as an approximate affine mixture of its neighbor points and constructing a point scatter in low dimensions that preserves as best as possible the affine mixture coefficients from high-dimensional data space, using an elegant eigenproblem.

A second new method, called "isometric feature mapping" or *Isomap* (Tenenbaum et al. 2000), builds on classical MDS but measures large distances in terms of hops along short distances. That is, *Isomap is classical MDS where large distances have been replaced by estimates of intrinsic geodesic distances.* The use of shortest path lengths as MDS inputs is due to Kruskal and Seery (1980) in graph drawing (see also Kamada and Kawai 1989, Gansner et al. 2004). Isomap's novelty is to use the idea for nonlinear dimension reduction.

A third new method and historically the oldest, called "kernel PCA" or *KPCA* (Schölkopf, Smola and Müller 1998), is also classical MDS but based on a localizing transformation of the inner product data from high-dimensional space. Localization can be achieved with a Gaussian Kernel transformation such as $\langle \mathbf{y}_i, \mathbf{y}_j \rangle = \exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2/(2\sigma^2))$; the smaller σ, the more localized is the inner product compared to the Euclidean metric.

LMDS, proposed here, derives from MDS by restricting the stress function to pairs of points with small distances. Thus LMDS shares with LLE, Isomap and KPCA what we may call "localization." Whereas Isomap completes the

"local graph" with shortest path lengths, LMDS stabilizes the stress function by introducing repulsion between points with large distances. The idea is borrowed from "force-directed energy functions" used in graph drawing, an important specialty in scientific visualization (Di Battista et al., 1999; Kaufmann and Wagner, 2001; Brandes 2001; Noack 2003; Michailidis and de Leeuw 2001).

Localization has an unhappy history in MDS. Removing large distances from the stress function has been tried many times since Kruskal (1964a, 1964b), but the hope that small dissimilarities add up to globally meaningful optimal configurations was dashed by Graef and Spence (1979): Their simulations showed that removal of the smallest third of dissimilarities was benign, while removal of the largest third had calamitous effects by reducing optimal configurations to mere jumbles. Thus the stability of optimal MDS configurations stems from the large dissimilarities, and localized MDS did not appear to be a viable approach.

Isomap's way out of the localization problem — completion of the local graph with shortest path lengths — has drawbacks: shortest paths tend to zig-zag and accumulate noise in estimates of intrinsic geodesic distances. In view of MDS' reliance on large distances, Isomap may be driven mostly by the large but noisy shortest-path imputations while the local distances play only a minor role. LMDS' solution to the localization problem has drawbacks, too, in that it may exhibit systematic distortions. This may be a classical trade-off between variance and bias: Isomap suffers from more variance, LLE and LMDS from more bias. LMDS, on the other hand, has a tuning parameter for controlling the balance between attractive and repulsive forces, permitting a range of embeddings from noisy with little bias to crisp with more bias.

While LMDS' tuning parameter provides flexibility, it also creates a selection problem: how does one know in practice which among several configurations is

most faithful to the underlying high-dimensional structure? This question can be answered with measures of faithfulness separate from the stress functions. We have proposed one such family of measures, called "Local Continuity" or "LC" meta-criteria and defined as the average size of the overlap of K-nearest neighborhoods in the high-dimensional data and the low-dimensional configuration (Chen 2006). These measures turn out to be practically useful for selecting good configurations. Independently, Akkucuk and Carroll (2006) developed similar measures for comparing the performance of different methods. We show how such measures can be employed as part of data analytic methodology 1) for choosing tuning parameters such as strength of the repulsive force and neighborhood size, and 2) as the basis of diagnostic plots that show how faithfully *each point* is embedded in a configuration.

To further motivate LC meta-criteria, we draw an analogy between dimension reduction and classification: In classification, the measures of interest are misclassification rates, yet classifiers are constructed as minimizers of smooth surrogate criteria such as logistic loss. Similarly, in dimension reduction, the measures of interest are the LC meta-criteria, yet configurations are constructed as minimizers of smooth stress functions. Like misclassfication rates, LC meta-criteria are not smooth and statistically unstable, yet of primary interest.

We conclude by noting that LMDS inherits the generality of MDS: The input used from the data is a matrix of distances or dissimilarities $D_{i,j}$, and for this reason the method applies wherever distances or dissimilarities arise: In dimension reduction $D_{i,j} = \|\mathbf{y}_i - \mathbf{y}_j\|$ for high-dimensional $\mathbf{y}_i$; in graph drawing $D_{i,j}$ are minimum path lengths within a graph; in proximity analysis $D_{i,j}$ are observed judgments of pairwise dissimilarity. LMDS applies in all cases.

Terminology: Straddling the areas of dimension reduction, graph drawing

and proximity analysis, we adopt terminology from all three. For $\{\mathbf{x}_i\}$ we use "configuration" from proximity analysis, and also "embedding" and "graph layout." For $D_{i,j}$ we use "dissimilarity" from proximity analysis, and also "target distance," meaning distances between high-dimensional feature vectors in dimension reduction and shortest-path-length distances in graph drawing.

Background on MDS: Two types of MDS must be distinguished: 1) *"Classical" Torgerson-Gower inner-product scaling* transforms dissimilarity data to inner-product data and extracts reduced dimensions from an eigendecomposition. 2) *Kruskal-Shepard distance scaling* approximates dissimilarity data directly with distances from configurations in reduced dimensions, in the simplest case by minimizing a residual sum of squares. Classical scaling applied to high-dimensional Euclidean distances is equivalent to PCA on the underlying multivariate data. It is hierarchical in the sense that, for example, a reduction to 3-D consists of the reduction to 2-D plus one more dimension. Distance scaling is not hierarchical, but it usually approximates dissimilarities better in a given reduced dimension than classical scaling. Isomap and KPCA are descendants of classical scaling; LMDS is a descendant of distance scaling. — A further distinction between metric and nonmetric MDS is irrelevant here as we restrict ourselves to the metric case. For more background see, for example, Borg and Groenen (2005), Buja and Swayne (2002) and Buja et al. (2008).

Further literature: An early pioneer in non-linear dimension reduction is Shepard and Carroll's (1966) PARAMAP method (Akkacuk and Carroll 2006). Various forms of model-based proximity analysis were proposed by Ramsay (1977, 1982), MacKay and Zinnes (1986), and Oh and Raftery (2001). Related to PCA are Hastie and Stuetzle's (1989) principal curves and surfaces. In a similar class are coordinatization approaches such as Zhang and Zha (2005) and

Brand (2005). A hybrid of classical and distance scaling are the semi-definite programming (SDP) approaches by Lu, Keles, Wright and Wahba (2005) and Weinberger, Sha, Zhu and Saul (2006) who fit full-rank Gram matrices K to local proximities via ${D_{i,j}}^2 \approx K_{i,i} + K_{j,j} - 2K_{i,j}$ and extract hierarchical embeddings by decomposing K. Related to KPCA with Gaussian kernels are Laplacian Eigenmaps (Belkin and Niyogi, 2003) and Diffusion Maps (Coifman, Lafon, Lee, Maggioni, Nadler, Warner and Zucker, 2005). Hessian Eigenmaps by Donoho and Grimes (2003) make the stronger assumption of local isometry to parts of a Euclidean parameter space.

We proceed as follows: Section 2 derives LMDS from Kruskal's distance scaling; Section 3 introduces LC meta-criteria, followed by thoughts on population modeling (Section 4) and illustrations with two image datasets (Section 5).

2 LOCAL MULTIDIMENSIONAL SCALING

The goal of MDS is to map objects $i = 1, ..., N$ to configuration points $\mathbf{x}_1, ..., \mathbf{x}_N \in \mathbb{R}^d$ such that the data, given as dissimilarities $D_{i,j}$, are well-approximated by the configuration distances $\|\mathbf{x}_i - \mathbf{x}_j\|$. In "metric distance scaling" one uses measures of lack of fit, called "Stress," between $\{D_{i,j}\}$ and $\{\|\mathbf{x}_i - \mathbf{x}_j\|\}$, which in the simplest case is a residual sum of squares:

$$\mathrm{MDS}^D(\mathbf{x}_1, ..., \mathbf{x}_N) \;=\; \sum_{i,j=1...N} \left(D_{i,j} - \|\mathbf{x}_i - \mathbf{x}_j\|\right)^2 \;. \qquad (1)$$

For localization, let $\mathcal{N}$ be a symmetric set of nearby pairs (i, j), such as a symmetrized K-nearest neighbor (K-NN) graph: $(i, j) \in \mathcal{N}$ if j is among the K nearest neighbors of i, or i is among the K nearest neighbors of j. If $\mathcal{N}$ does not form a connected graph, one may map the connected components separately, or one connects the components by adding connecting pairs to $\mathcal{N}$.

Our initial proposal for localized MDS is to replace the dissimilarities for (i,j) not in $\mathcal{N}$ with a very large value D_∞ but with small weight w:

$$\begin{aligned} \mathrm{LMDS}^D_{\mathcal{N}}(\mathbf{x}_1,...,\mathbf{x}_N) &= \sum_{(i,j)\in\mathcal{N}} \left(D_{i,j} - \|\mathbf{x}_i - \mathbf{x}_j\|\right)^2 && (2) \\ &+ \sum_{(i,j)\notin\mathcal{N}} w \cdot \left(D_\infty - \|\mathbf{x}_i - \mathbf{x}_j\|\right)^2 \ . && (3) \end{aligned}$$

The pairs $(i,j) \in \mathcal{N}$ describe the "local fabric" of a high-dimensional manifold or a graph, whereas the pairs $(i,j) \notin \mathcal{N}$ introduce a bias that should avoid the typical problem of MDS when the large dissimilarities are eliminated from the Stress: crumpling up of the configuration to a jumble, meaning that many distant pairs of points are placed close together. The imputation of a very large distance introduces a pervasive repulsive force throughout the configuration, similar to electric static that makes dry hair fly apart. The imputation of a single large dissimilarity D_∞ with little weight is likely to introduce less noise than the imputation of shortest-path estimates, at the price of some bias. Thus LMDS derives configurations directly from local distances, whereas Isomap derives them indirectly from estimated noisy large distances.

A question is how to choose the weight w relative to the imputed value D_∞. The following argument shows that w should be on the order of $1/D_\infty$. We expand the "repulsion term" (3), discarding functions of the dissimilarities that do not affect the minimization problem:

$$\begin{aligned} \mathrm{LMDS}^D_{\mathcal{N}}(\mathbf{x}_1,...,\mathbf{x}_N) &\sim \sum_{(i,j)\in\mathcal{N}} \left(D_{i,j} - \|\mathbf{x}_i - \mathbf{x}_j\|\right)^2 && (4) \\ &\quad -2wD_\infty \sum_{(i,j)\notin\mathcal{N}} \|\mathbf{x}_i - \mathbf{x}_j\| && (5) \\ &\quad +w \sum_{(i,j)\notin\mathcal{N}} \|\mathbf{x}_i - \mathbf{x}_j\|^2 && (6) \end{aligned}$$

As $D_\infty \to \infty$, we let $w \to 0$ at least on the order of $1/D_\infty$ in order to prevent the term (5) from blowing up. The weight w cannot go to zero faster than

$1/D_\infty$, though, because otherwise both terms (5) and (6) vanish. This leaves $w \sim 1/D_\infty$ as the only non-trivial choice, in which case (6) disappears. We let therefore $D_\infty \to \infty$ subject to $w = t/(2D_\infty)$, where t is a fixed constant, and arrive at the final definition of localized Stress:

$$\text{LMDS}^D_{\mathcal{N}}(\mathbf{x}_1, ..., \mathbf{x}_N) \;=\; \sum_{(i,j)\in N} (D_{i,j} - ||\mathbf{x}_i - \mathbf{x}_j||)^2 \;-\; t \sum_{(i,j)\notin N} ||\mathbf{x}_i - \mathbf{x}_j|| \quad (7)$$

We call the first term "local stress" and the second term "repulsion." A benefit of passing to the limit is the replacement of two parameters, w and D_∞, with a single parameter t. In addition, the two remaining terms in (7) have intuitive meaning: The first term forces $\|\mathbf{x}_i - \mathbf{x}_j\|$ to follow $D_{i,j}$ for $(i,j) \in \mathcal{N}$ and is responsible for preserving the intended local structure as much as possible. The second term contributes repulsion outside $\mathcal{N}$ and is responsible for pushing points away from each other if they are not locally linked.

The relative strength of attractive and repulsive forces is balanced by the parameter t. Selecting it in a data-driven way is the subject of the next section. As it stands, however, t is unsatisfactory because it suffers from a lack of invariance under desirable transformations. This problem can be corrected:

• *Invariance under change of units:* $D_{i,j}$ and t have the same units, hence the units of t can be eliminated, for example, by $t = \text{median}_{\mathcal{N}}(D_{i,j})\, t'$, where the new parameter t' is unit free. Instead of $\text{median}_{\mathcal{N}}(D_{i,j})$ any statistic S that satisfies $S(\{cD_{i,j}\}) = c\, S(\{D_{i,j}\})$ will do.

• *Approximate invariance under change of graph size:* As the graph size $|\mathcal{N}|$ changes, so does the number of summands in (7): $|\mathcal{N}|$ for the local stress and $|\mathcal{N}^C| = N(N-1)/2 - |\mathcal{N}|$ for the repulsion. As $|\mathcal{N}|$ grows, the relative importance of the repulsion diminishes for fixed t. This can be corrected by

reparametrizing t with a factor $|\mathcal{N}|/|\mathcal{N}^C|$)):

$$t \;=\; \frac{|\mathcal{N}|}{|\mathcal{N}^C|} \cdot \text{median}_{\mathcal{N}}(D_{i,j}) \cdot \tau \;,$$

where τ is unit free. — A good strategy for optimization is to start with a large value such as $\tau = 1$ to obtain a stable configuration, and lower τ successively as low as 0.01, using previous configurations as initializations for smaller values of τ. Along the way one collects the quantities proposed in the next section for selecting a value of τ that may be nearly optimal in specific sense.

Notation: We write $\text{LMDS}^D_{\mathcal{N}}$ when we allow a general graph $\mathcal{N}$, but we may write LMDS^D_K when the graph $\mathcal{N}$ is a symmetrized K-NN graph. Depending on the context, we may omit or add arguments, such as LMDS_K or $\text{LMDS}_{K,\tau}$.

3 CRITERIA FOR PARAMETER SELECTION

For automatic selection of tuning parameters and comparison of methods we need measures of faithfulness of configurations separate from the stress functions used for optimizing configurations. One such family of measures was independently developed by Carroll and his students (Akkucuk and Carroll 2006, France and Carroll 2006) and by us (Chen 2006). The idea is to compare for a given case 1) the K'-NN with regard to $D_{i,j}$ in data space, and 2) the K'-NN with regard to $\|\mathbf{x}_i - \mathbf{x}_j\|$ in configuration space. [We use K' in distinction from the K used in the stress function LMDS_K .] A high degree of overlap between the two neighbor sets yields a measure of local faithfulness of the embedding of the given case. By averaging over all cases we obtain a global measure which we call "local continuity" or "LC meta-criterion". The neighborhood size K' is a free parameter, and its choice requires further discussion.

Notation: For case i we form the index set $\mathcal{N}^D_{K'}(i) = \{j_1, ..., j_{K'}\}$ of K'-

NNs with regard to $D_{i,j}$, and $\mathcal{N}^X_{K'}(i) = \{k_1, ..., k_{K'}\}$ of K'-NNs with regard to $\|\mathbf{x}_i - \mathbf{x}_k\|$ (excluding i). The overlap is measured pointwise and globally by

$$N_{K'}(i) \;=\; |\,\mathcal{N}^D_{K'}(i) \;\cap\; \mathcal{N}^X_{K'}(i)\,|\,, \qquad N_{K'} \;=\; \frac{1}{N}\sum_{i=1}^{N} N_{K'}(i)\,. \tag{8}$$

The pointwise criteria $N_{K'}(i)$ lend themselves for diagnostic plots that pinpoint local lack of faithfulness of embeddings.

Both the pointwise $N_{K'}(i)$ and the global $N_{K'}$ are bounded by K', and $N_{K'} = K'$ would imply maximal faithfulness: $N_{K'}(i) = K'$ for all i, meaning perfect identity of K'-nearest neighborhoods in terms of the data $\{D_{i,j}\}$ and the configuration $\{\mathbf{x}_i\}$.

Normalization: To enable comparison of $N_{K'}$ across different values of K', we use the fact that $N_{K'}$ has K' as an upper bound and normalize overlap to the $[0, 1]$ interval:

$$M_{K'} \;=\; \frac{1}{K'} N_{K'}\,. \tag{9}$$

An example of a trace $K' \mapsto M_{K'}$ is shown in Figure 3 (top right, upper curve), which illustrates the fact that the trace ascends to $M_{K'} = 1$ for $K' = N - 1$.

Adjusting for random overlap: If there is complete absence of association between the data and the configuration, the local overlap $N_{K'}(i)$ is random and can be modeled by a hypergeometric distribution with K' defectives out of $N - 1$ items and K' draws, and hence $\mathrm{E}[N_{K'}] = K'^2/(N-1)$. This suggests defining *adjusted LC meta-criteria* (Chen 2006):

$$M^{adj}_{K'} \;=\; M_{K'} \;-\; \frac{K'}{N-1}\,. \tag{10}$$

An example of an adjusted trace $K' \mapsto M^{adj}_{K'}$ is also shown in Figure 3 (top right plot, lower curve). — Akkucuk and Carroll (2006) go further by mapping $N_{K'}$ to a z-score under random overlap, but in most applications these z-scores are

extreme because even weak structure results in extreme statistical significance under the null hypothesis of random overlap.

Comparing two configurations: For comparing two configurations, one may plot the trace of differences, $K' \mapsto M^{(1)}_{K'} - M^{(2)}_{K'}$. With differences the issue of random overlap becomes moot. An example is shown in Figure 8 (right hand plot) where LMDS and plain MDS configurations are compared. The idea is that LMDS' localization should produce benefits over plain MDS in terms of $M_{K'}$ for small K'.

Selection of τ: Given K for $\text{LMDS}_{K,\tau}$ and K' for $M_{K'}$, we can optimize the repulsion weight τ with a grid search. This is illustrated in Figure 3 (top left plot) with a trace $\tau \mapsto M_{K'}$ applied to configurations that are optimized for $\text{LMDS}_{K,\tau}$ for $K = K' = 6$. Henceforth, when applied to $\text{LMDS}_{K,\tau}$-optimal configurations, $M_{K'}$ denotes the τ-maximized value for a given K.

Selection of K: There is a question of which $M^{adj}_{K'}$ to use to judge the configurations that minimize LMDS_K. Here are two strategies:

• $K' = K$: For fixed K minimize $\text{LMDS}_{K,\tau}$ for various values of τ; pick the τ whose configuration maximizes M^{adj}_K; repeat for various values of $K = K'$ and plot a trace $K' \mapsto M^{adj}_{K'}$ as, for example, in Figure 3 (upper right, lower curve). Finally, pick that $K = K'$ which maximizes the trace ($K = K' \approx 8$ in the figure).

• K', K decoupled: It is desirable that K is not just $M^{adj}_{K'}$-optimal for $K' = K$, but for a range of values K'. To find out, one may plot traces $K' \mapsto M^{adj}_{K'}$, one for each value of K, as in Figure 3 (bottom right). It is comforting that in this case $K = 8$ dominates uniformly over a range of K' from 4 up to over 20.

Concluding remark: The LC meta-criteria are doubly "non-metric" in the sense that they only use rank information of both $\{D_{i,j}\}$ and $\{\|\mathbf{x}_i - \mathbf{x}_j\|\}$. Equiv-

alently, they are invariant under strictly monotone increasing transformations. They therefore add — at the level of parameter tuning — a non-metric element to LMDS which is otherwise a metric form of MDS.

4 A POPULATION FRAMEWORK

Nonlinear dimension reduction can be approached with statistical modeling if it is thought of as "manifold learning." Zhang and Zha (2005), for example, examine "tangent space alignment" with an theoretical error analysis whereby an assumed target manifold is reconstructed from noisy point observations. With practical experience in mind, we introduce an alternative to manifold learning and its assumption that the data falls near a "warped sheet." We propose instead that the data be modeled by a distribution in high dimensions that describes the data, warts and all, with variation caused by digitization, rounding, uneven sampling density, and so on. In this view the goal is to develop methodology that shows what can be shown in low dimensions and to provide diagnostics that pinpoint problems with the reduction. By avoiding "prejudices" implied by assumed models, the data are allowed to show whatever they have to show, be it manifolds, patchworks of manifolds of different dimensions, clusters, sculpted shapes with protrusions or holes, general uneven density patterns, and so on. For an example where this unprejudiced EDA view is successful, see the Frey face image data (Section 5.2) which exhibit a noisy patchwork of 0-D, 1-D and 2-D submanifolds in the form of clusters, rods between clusters, and webfoot-like structures between rods. This data example also makes it clear that no single dimension reduction may be sufficient to show all that is of interest: very localized methods ($K = 4$) reveal

the underlying video sequence and transitions between clusters, whereas global methods (PCA, MDS) show the extent of the major interpretable clusters and dimensions more realistically. In summary, it is often pragmatic to assume less (a distribution) rather than more (a manifold) and to assign the statistician the task of "flattening" the distribution to lower dimensions as faithfully and usefully as possible.

An immediate benefit of the distribution view is to recognize Isomap as non-robust due to its use of geodesic distances. For a distribution, a geodesic distance is a path with shortest length in the support of the distribution. If a distribution's support is convex, the notion of geodesic distance in the population reduces to chordal distance. For small samples, Isomap may find manifolds that approximate the high-density areas of the distribution, but for increasing sample size the geodesic paths will find shortcuts that cross the low-density areas and asymptotically approximate chordal distances, thus changing the qualitative message of the dimension reduction as $N \to \infty$.

By comparison, the LMDS criterion has a safer behavior under statistical sampling as meaningful limits for $N \to \infty$ exist. Here is the population target:

$$\begin{aligned} \mathrm{LMDS}^{\mathrm{P}}_{\mathcal{N}}(x) = \mathrm{E}\left[\left(\|Y' - Y''\| - \|x(Y') - x(Y'')\| \right)^2 \cdot 1_{[(Y',Y'')\in\mathcal{N}]} \right] \\ - t\,\mathrm{E}\left[\|x(Y') - x(Y'')\| \cdot 1_{[(Y',Y'')\notin\mathcal{N}]} \right], \end{aligned} \tag{11}$$

where Y', Y'' are $iid\,\mathrm{P}(d\mathbf{y})$ on $\mathbb{R}^p$, $\mathcal{N}$ is a symmetric neighborhood definition in $\mathbb{R}^p$, and the "configuration" $\mathbf{y} \mapsto x(\mathbf{y})$ is interpreted as a map from the support of $\mathrm{P}(d\mathbf{y})$ in $\mathbb{R}^p$ to $\mathbb{R}^d$ whose quality is being judged by this criterion. When specialized to empirical measures, $\mathrm{LMDS}^{\mathrm{P}}_{\mathcal{N}}(x)$ becomes $\mathrm{LMDS}^{D}_{\mathcal{N}}(x(\mathbf{y}_1), ..., x(\mathbf{y}_N))$. A full theory would establish when local minima of (11) exist and when those of (7) obtained from data converge to those of (11) for $N \to \infty$. Further the-

ory would describe rates with which $\mathcal{N}$ and t can be shrunk to obtain finer resolution while still achieving consistency.

The LC meta-criteria also have population versions for dimension reduction: For $\mathbf{y}$ in the support of $P(d\mathbf{y})$ (which we assume continuous), let $\mathcal{N}_\alpha^{\mathcal{Y}}(\mathbf{y})$ be the Euclidean neighborhood of $\mathbf{y}$ that contains mass α: $P[\,Y \in \mathcal{N}_\alpha^{\mathcal{Y}}(\mathbf{y})\,] = \alpha$; similarly, in configuration space let $\mathcal{N}_\alpha^{\mathcal{X}}(x(\mathbf{y}))$ be the mass-α neighborhood of $x(\mathbf{y})$: $P[\,x(Y) \in \mathcal{N}_\alpha^{\mathcal{X}}(x(\mathbf{y}))\,] = \alpha$. Then the population version of the LC meta-criterion for parameter α is

$$M_\alpha = \frac{1}{\alpha} P[\,Y'' \in \mathcal{N}_\alpha^{\mathcal{Y}}(Y') \text{ and } x(Y'') \in \mathcal{N}_\alpha^{\mathcal{X}}(x(Y'))\,],$$

where again Y', Y'' are *iid* $P(d\mathbf{y})$. This specializes to $M_{K'}$ (equation (9)) for $K'/N = \alpha$ when $P(d\mathbf{y})$ is an empirical measure (modulo exclusion of the center points, which is asymptotically irrelevant). The quantity M_α is a continuity measure, unlike for example the criterion of Hessian Eigenmaps (Donoho and Grimes 2003) which is a smoothness measure.

A population point of view had proven successful once before in work by Buja, Logan, Reeds and Shepp (1994) which tackled the problem of MDS performance on completely uniformative or "null" data. It was shown that in the limit ($N \to \infty$) null data $\{D_{i,j}\}$ produce non-trivial spherical distributions as configurations and that this effect is also a likely contributor to the "horseshoe effect," the artifactual bending of MDS configurations in the presence of noise.

5 EXAMPLES

In this section we apply LMDS to two sets of facial image data for two reasons: 1) such data are often intrinsically piecewise low-dimensional and hence promising for dimension reduction (see below), and 2) facial image data have been the

primary examples in the recent literature (Roweis and Saul 2000; Tenenbaum et al. 2000). This fact enables us to compare the performance of LMDS with those of competing methods on the same data sets.

Images are two-way arrays of pixels which in the simplest case describe light intensities on a grey scale. By treating each pixel as a variable, an image of size 64×64 pixels becomes a single data point in a 4096-dimensional image space. These dimensions, however, are highly redundant. Two sources of redundancy are the following: 1) In most images a majority of nearby pixel pairs have nearly equal light intensities, which translates to strong correlations between variables. 2) There are often far fewer degrees of freedom when a collection of images shows articulations of the same object. The physical degrees of freedom in the facial image datasets considered below include viewing direction and lighting direction in one case and facial expression in the other case.

5.1 Example 1: Sculpture Face Data

This dataset includes 698 images of size 64×64 of a sculpture face and was analyzed in the Isomap article (Tenenbaum et al. 2000). The images show the same sculpture face while varying three conditions: left-right pose, up-down pose, and lighting direction, amounting to three angular parameters that characterize an image. One can therefore anticipate that the true intrinsic dimension of the image data is three, and one hopes that nonlinear dimension reduction tools will reveal them in a meaningful way. Because the underlying truth is known, this example serves the same purpose as a simulation.

The Isomap analysis of Tenenbaum et al. (2000) was done with $K = 6$ nearest neighbors and Euclidean distances between images, and we adopt these choices in our LMDS analysis. The 3-D configuration generated by LMDS

is close to a hyperrectangle as Figure 1 shows. We labeled a few points in both views with the corresponding images. On the widest dimension (drawn horizontally) the images show transitions in the left-right pose, on the second widest dimension (vertical axis of the upper view) transitions in the up-down pose, and on the third dimension (vertical axis of the lower view) transitions in the lighting direction.

To compare recovery of the three angular parameters, we shaded the configuration points by dividing the range of each parameter in turn into its three tercile brackets and encoding them in gray scale. If recovery occurs in a configuration, we should see coherence in the distribution of gray tones. In Figure 2 we show shaded 2-D views of 3-D configurations for PCA, MDS, Isomap, LLE and LMDS, with the configurations rotated to best reveal the gray separations. We note that the three tones of gray overlap in some of the plots; in particular the up-down transition is not well-captured by PCA, MDS and LLE. Interesting is also a wrap-around structure in the PCA configuration. The meeting of the extremes (light gray and black) visible in the frame of PCA and 'Left-right Pose' is probably caused by the darkness of images showing the extreme left and extreme right poses. The LLE configuration shows characteristic spikes which we observed in most applications: LLE is generally prone to linear artifacts in its configurations, and LLE may be more difficult to fine-tune for competitive performance than Isomap and LMDS. The Isomap and LMDS configurations in Figure 2 look quite similar and show clear color separations. They are most successful at flattening the nonlinear structure in this data set. Of the two, LMDS shows crisper boundaries in the configuration, which in our experience is a general observation. The fuzziness in the Isomap configurations is consistent with the noisiness of shortest-path length imputations discussed earlier.

We also compared the five methods according to the LC meta-criterion $N_{K'}$ for $K' = 6$, shown in Table 1. We see that Isomap and LMDS generate better configurations by this measure, with LMDS winning out by a small margin.

Method	PCA	MDS	Isomap	LLE	$\text{LMDS}_{K=6}$
$N_{K'=6}$	2.6	3.1	4.5	2.8	5.2
$M_{K'=6}$	0.43	0.52	0.75	0.47	0.87

Table 1: *Sculpture Face Data: LC meta-criteria for* $K' = 6$.

In Figure 3 (top left) we show a trace of $M_{K'}$ ($K' = 6$) as a function of the repulsion weight τ. The shape is typical: a unimodal graph that attains a maximum, in this case of $M_{K'} \approx .87$ near $\tau = 0.005$. The decrease of the meta-criterion to the left of the peak indicates that too weak a repulsion allows the local stress to take over and causes the configuration to degenerate. Thus, repulsion is essential for the stability of optimized configurations.

For comparison purposes, we used the same number of nearest neighbors in LMDS, $K = 6$, as in Tenenbaum et al. (2000). They did not discuss how they chose K; presumably they used trial-and-error to find a useful configuration. We can be more systematic as we can use the LC meta-criteria to choose K.

Using first the simpler selection methodology, we link $K' = K$. We tried a grid between $K' = 4$ and $K' = 650$, with small increments near 4 and larger increments near 650. For each K, we optimized the meta-criterion with $K' = K$ with regard to the repulsion weight τ, and we used this optimized value of $M_{K'}^{(adj)}$ as the criterion for selecting K. [Note: For different K' the highest $M_{K'}^{(adj)}$ is achieved at different values of τ.] The resulting traces are shown in Figure 3 (top right): unexpectedly, there are two local maxima, at

$K' = 8$ and $K' = 16$, the latter quite minor, though. After these peaks, the traces dip before $M_{K'}$ ascends to its maximum (1) and $M_{K'}^{adj}$ descends to its minimum (0) at $K' = N - 1$. Adjustment for random overlap has the desired effect of creating a single absolute maximum for $M_{K'}^{adj}$ at $K = K' = 8$. By this methodology, the Isomap authors' choice $K = 6$ is near optimal.

In the next exercise we unlink K and K' and plot traces $K' \mapsto M_{K'}^{(adj)}$ for a selection of values of K, as in Figure 3 (bottom row). The traces reach their maxima at or near $K' = K$, which, in this data example, lends support to the simpler methodology based on $K' = K$.

Finally, we use the pointwise LC meta-criterion $N_6(i)$ as a diagnostic by gray scale coding the configurations from PCA, Isomap, LLE and LMDS, as shown in Figure 4. The overall gray scale impression of a configuration reflects its average level of $N_6(i)$, namely, N_6. Correspondingly, the lower quality configurations from PCA and LLE contain overall more black, those from Isomap and LMDS more light gray. It appears that Isomap's configuration is of lesser quality in the left and right extremes where we find a greater density of black points. Overall the LMDS configuration has a slight edge.

5.2 Example 2: Frey Face Data

This data set, from the LLE article (Roweis and Saul 2000), includes 1965 images of the face of a single person (by name of Brendan Frey), taken as sequential frames from a piece of video footage. The time ordering of these frames and the expected correlation between frames close in time constitutes a true known structure that one hopes to recover with dimension reduction. The image size is 20×28, hence the image space is 560-dimensional. We include PCA, MDS, Isomap, LLE, as well as KPCA as competitors of LMDS.

For the localized methods, nearest neighbor computations were based on plain Euclidean distances in 560-dimensional image space. With visualization in mind we choose again a reduction to 3-D and show an informative 2-D projection of each 3-D configuration. Figure 5 shows four solutions of LMDS for various values of K, and Figure 6 compares the six methods.

Overall, LMDS configurations with small K show more structure than those of the other methods. All configurations reveal two major clusters (top and bottom in each) corresponding to, respectively, smiling faces (top) and serious faces (bottom). A second feature shared by all is a dimension correponding roughly to the left-right pose. All plots were rotated to line up the serious-smiling division vertically and the left-right pose horizontally.

The aspect in which the competing methods differ the most is in the degree to which the major divisions are further subdivided into subclusters and linked by transitions between them. The PCA and MDS configurations show essentially only the two base dimensions and give little indication of further subdivisions. LLE produces a configuration that is quite structured, but it suffers again from a tendency to spikiness that is most likely artifactual but difficult to prevent. Isomap comes closest to LMDS in regards to subdivisions and transitions, but its noisiness obscures transitions between clusters. The LMDS configurations for small K show transitions between subclusters that can be shown to be real. Our confidence in this belief is backed up by Figure 7 where the same plots show the time order with connecting lines. The LMDS configuration makes it clear that there exist essentially four transitions between smiling faces at the top and serious faces at the bottom. Although LMDS' spikiness could raise suspicions that it suffers from the same problem as the LLE configuration, the connecting lines show that the spikes describe indeed paths

taken by the video footage. The main bias of the LMDS configuration for $K = 4$ is most likely in the extent to which it attempts to separate the serious and smiling faces; the "true" separations are most likely better reflected in the cruder configurations from PCA, MDS and Isomap.

Methods	PCA	MDS	Isomap	LLE	KPCA	$\text{LMDS}_{K=12}$	$\text{LMDS}_{K=4}$
$N_{K'=12}$	3.6	4.8	4.2	3.2	3.7	4.6	5.1
$M_{K'=12}$	.30	.40	.35	.27	.31	.38	.43

Table 2: *Frey Face Data: LC meta-criteria for* $K' = 12$.

Table 2 evaluates the six methods according to LC meta-criteria. Even though we use $K' = 12$, $\text{LMDS}_{K=4}$ is the top performer, whereas $\text{LMDS}_{K=12}$ is slightly dominated by regular MDS. With regard to wealth of structure, MDS is not a serious competitor of $\text{LMDS}_{K=12}$, though. As globally biased as the $\text{LMDS}_{K=4}$ configuration appears, it is the most congenial for the video path, and the meta-criterion $M_{K'=12}$ appropriately singles it out.

We next decouple K and K' and consider $M_{K'}$-traces for LMDS_K configurations for various choices of K, as shown in Figure 8. The unadjusted traces on the left show a very different behavior from those of the sculpture face data in that all traces are ascending and maximization is not possible. Adjustment for random overlap is insufficient as it barely affects the traces in the range of K'-values shown (4 to 150 out of $N - 1 = 1964$). We therefore use more drastic adjustment with MDS as the baseline, shown on the right of Figure 8. The horizontal line (diamonds) at level zero marks MDS. It turns out that configurations based on LMDS_K with $K = 4$ (circles) performs best in terms of $M^{adj}_{K'}$ for K' up to 8, whereas the configurations generated by $K = 8$ (trian-

gles) and $K = 12$ (plus signs) perform badly, as they are uniformly beaten by MDS. Apparently the neighborhood structure for $K' \geq 10$ is best captured by global MDS, but the locally zero- and one-dimensional structures (clusters and transitions) best rendered by $K = 4$, as measured by $M_{K'}^{adj}$ for $K' < 10$.

6 DISCUSSION AND CONCLUSION

This article makes three contributions: (1) It introduces a novel version of multidimensional scaling, called LMDS, that lends itself to locally faithful non-linear dimension reduction and as such competes successfully with recent proposals such as "local linear embedding" (LLE, Roweis and Saul 2000) and "isometric feature mapping" (Isomap, Tenenbaum et al. 2000). (2) This article proposes a solution to the problem of selecting tuning parameters. (3) It finally proposes a diagnostics tool for detecting local flaws in embeddings.

LMDS also applies to graph drawing problems and to proximity analysis. Inspired by energy functions used in graph drawing, LMDS uses graph-internal attraction forces to faithfully render local proximities and graph-external repulsion forces to stabilize the configurations. Novel is 1) the derivation of this particular repulsion and 2) the fact that we subject it to tuning.

The tuning problem is solved with "local continuity" or LC meta-criteria that measure K'-NN agreement in the data and in the configurations. (A version of it was independently proposed by Akkucuk and Carroll (2006) for comparing different dimension reduction methods.) In addition, we are able to use LC meta-criteria for tuning the degree of localization, that is, the size of neighborhoods in the LMDS stress function.

Exploratory tools such as LMDS require diagnostics to detect local flaws in

embeddings. We provide methodology to this end with a pointwise version of the LC meta-criteria. Further diagnostics for stability and multiplicity of embeddings with subsampling and perturbation are described by Buja and Swayne (2002) and Buja et al. (2008). Diagnostics for establishing local dimensionality can be based on "prosections" proposed by Furnas and Buja (1994).

The Achilles heal of methods considered here is the complete reliance on distance data or dissimililarities, which holds both for the LMDS fitting criterion and the LC meta-criteria. As methods cannot be better than their inputs, future research should address ways to choose distances/dissimilarities in a problem-specific manner. Such efforts could blend with similar needs in "kernelizing" regression methods such as SVMs which essentially replace predictor spaces with similarity measures.

REFERENCES

Akkucuk, U. and Carroll, J.D., 2006, PARAMAP vs. Isomap: A Comparison of Two Nonlinear Mapping Algorithms, *Journal of Classification*, **23** (2), 221-254.

Belkin, M., and Niyogi, P., 2003, Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, **15** (6), 1373-1396.

Borg, I., and Groenen, P. 2005, *Modern Multidimensional Scaling: Theory and Applications*, New York: Springer-Verlag.

Brand, M., 2005, Nonrigid Embeddings for Dimensionality Reduction. *European Conference on Machine Learning (ECML)*, proceedings, in *Springer Lecture Notes in Computer Science*, **3720**, 47-59.

Brandes, U., 2001, Drawing on Physical Analogies, in: *Drawing Graphs*, Kaufmann, M. and Wagner, D., eds., Berlin:Springer, 71-86.

Buja, A., Swayne, D.F., Littman, M., Hofmann, H., and Chen, L., 2008, Data Visualization with Multidimensional Scaling, *Journal of Computational and Graphical Statistics*, **17** (2), 1-29.

Buja, A., and Swayne, D.F., 2002, Visualization Methodology for Multidimensional Scaling, *Journal of Classification*, **19**, 7-43.

Buja, A., Logan, B.F, Reeds, J.R., Shepp, L.A. 1994, Inequalities and Positive-Definite Functions Arising from a Problem in Multidimensional Scaling, *The Annals of Statistics*, **22**, 406-438.

Chen, L., 2006, Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis, Ph.d. Thesis, University of Pennsylvania.

R.R.Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner and S.W. Zucker (2005), Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data, *Proceedings of the National Academy of Sciences*, **102** (21), 7426-7431.

Davidson, R. and Harel, D., 1996, Drawing graphs nicely using simulated annealing, *ACM Transactions on Graphics*, **15**(4), 301-331.

Di Battista, G., Eades, P., Tamassia, R., Tollis, I. G., 1999, Graph Drawing: Algorithms for the Visualization of Graphs. Prentice Hall.

Donoho, D.L., and Grimes, C., 2003, Hessian Eigenmaps: Locally Linear Embedding Techniques For High-Dimensional Data, *Proc. of National Academy of Sciences*, **100** (10), 5591-5596.

Eades P., 1984, A heuristic for graph drawing, In *Congressus Numerantium*, **42**, 149-160.

France, S. L. and Carroll, J. D., 2006, Development of an agreement metric based upon the RAND index for the evaluation of dimensionality reduction techniques, with applications to mapping customer data, Preprint, Rutgers University, NJ.

Fruchterman, T. M. J. and Reingold, E. M., 1991, Graph drawing by force-directed placement, *Software-Practice and Experience*, **21(11)**, 1129-1164.

Furnas, G.W., and Buja, A., 1994, Prosection Views: Dimensional Inference through Sections and Projections, *J. of Computational and Graphical Statistics*, 3, 323-385.

Gansner, E., Koren, Y., and North, S., 2004, Graph Drawing by Stress Majorization, *Graph Drawing*, 239-250.

Graef J., and Spence, I., 1979, Using Distance Information in the Design of Large Multidimensional Scaling Experiments, *Psychological Bulletin*, **86**, 60-66.

Hastie, T. and Stuetzle, W. 1989, Principal curves. *J. of the American Statistical Association* **84**, 502-516.

Kaufmann, M. and Wagner, D. (eds.), 2001, Drawing Graphs. Springer: Berlin.

Kamada, T., and Kawai, S., 1989, An algorithm for drawing general undirected graphs. *Information Processing Letters*, **31**, 7-15.

Kruskal, J. B., 1964a, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika*, **29**, 1-27.

Kruskal, J. B., 1964b, Nonmetric multidimensional scaling: a numerical method, *Psychometrika*, **29**, 115-129.

Kruskal, J. B. and Seery, J. B., 1980, Designing Network Diagrams, Technical Memorandum, Bell Laboratories, Murray Hill, NJ.

Lu, F., Keles, S., Wright, S. J., and Wahba, G., 2005, Framework for kernel regularization with application to protein clustering, *PNAS* **102**, (35), 12332-12337 (http://www.pnas.org/cgi/content/full/102/35/12332)

MacKay, D. B. and Zinnes, J. L. 1986, A Probabilistic Model for the Multidimensional Scaling of Proximity and Preference Data, *Marketing Science*, **5**, 325-344.

Michailidis, G. and de Leeuw, J., 2001, Data visualization through graph drawing, *Computation. Stat.*, **16**, 435-50.

Noack, A., 2003, Energy models for drawing clustered small-world graphs, Computer Science Report 07/2003, Brandenburg Technical University at Cottbus, Germany.

Oh, M. and Raftery, A. E., 2001, Baysian Multidimensional Scaling and choice of dimension, *J. of the American Statistical Association*, Vol. 96, No. 455, pp. 1031-1044.

Ramsay, J. O., 1977, Maximum likelihood estimation in Multidimensional Scaling, *Psychometrika* Vol. 42, No. 2, 241-266

Ramsay, J. O., 1982, Some Statistical Approaches to Multidimensional Scaling Data (with discussion), *J. of the Royal Statistical Society*, Series A, Vol. 145, No. 3, 285-312.

Roweis S. T. and Saul, L. K., 2000, Nonlinear dimensionality oreduction by local linear embedding, *Science*, **290**, 2323-2326.

Sammon, J., 1969, A Non-Linear Mapping for Data Structure Analysis. *IEEE Transactions on Computers*, **C-18**(5), 401-409.

Saul, L. K. and Roweis, S. T., 2003, Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds, *J. of Machine Learning*, **4**, 119-155.

Shepard, R. N. and Carroll, J. D. 1966, Parametric representation of nonlinear data structures. In: P. R. Krishnaiah (Ed.), Multivariate Analysis (pp. 561-592). New York: Academic Press.

Schölkopf, B., Smola, A. J., and Müller, K.-R., 1998, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, **10**, 1299-1319.

Tenenbaum, J. B., Silva, V. d., and Langford, J. C., 2000, A global geometric framework for nonlinear dimensionality reduction, *Science*, **290**, 2319-2323.

Torgerson, W.S., 1952, Psychometrika, **17**, 401-419.

Weinberger, K. Q., Sha, F., Zhu, Q., and Saul, L. K., 2006, Graph Laplacian Regularization for Large-Scale Semidefinite Programming. *NIPS 2006*,1489-1496.

Zhang, Z., and Zha, H., 2005, Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Scientific Computing*, **26**, 313-338.

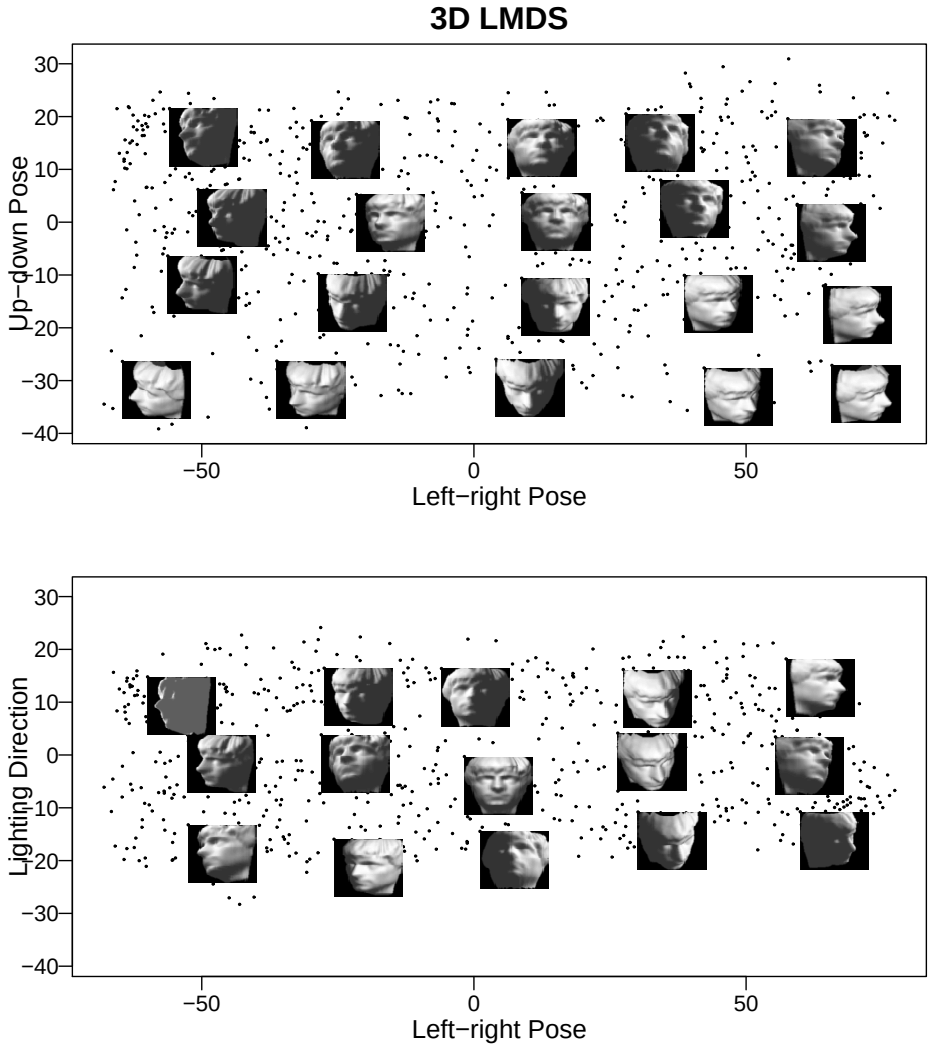

Figure 1: *Sculpture Face Data. Three-D LMDS configuration,* $K = 6$*, optimized* τ*.*

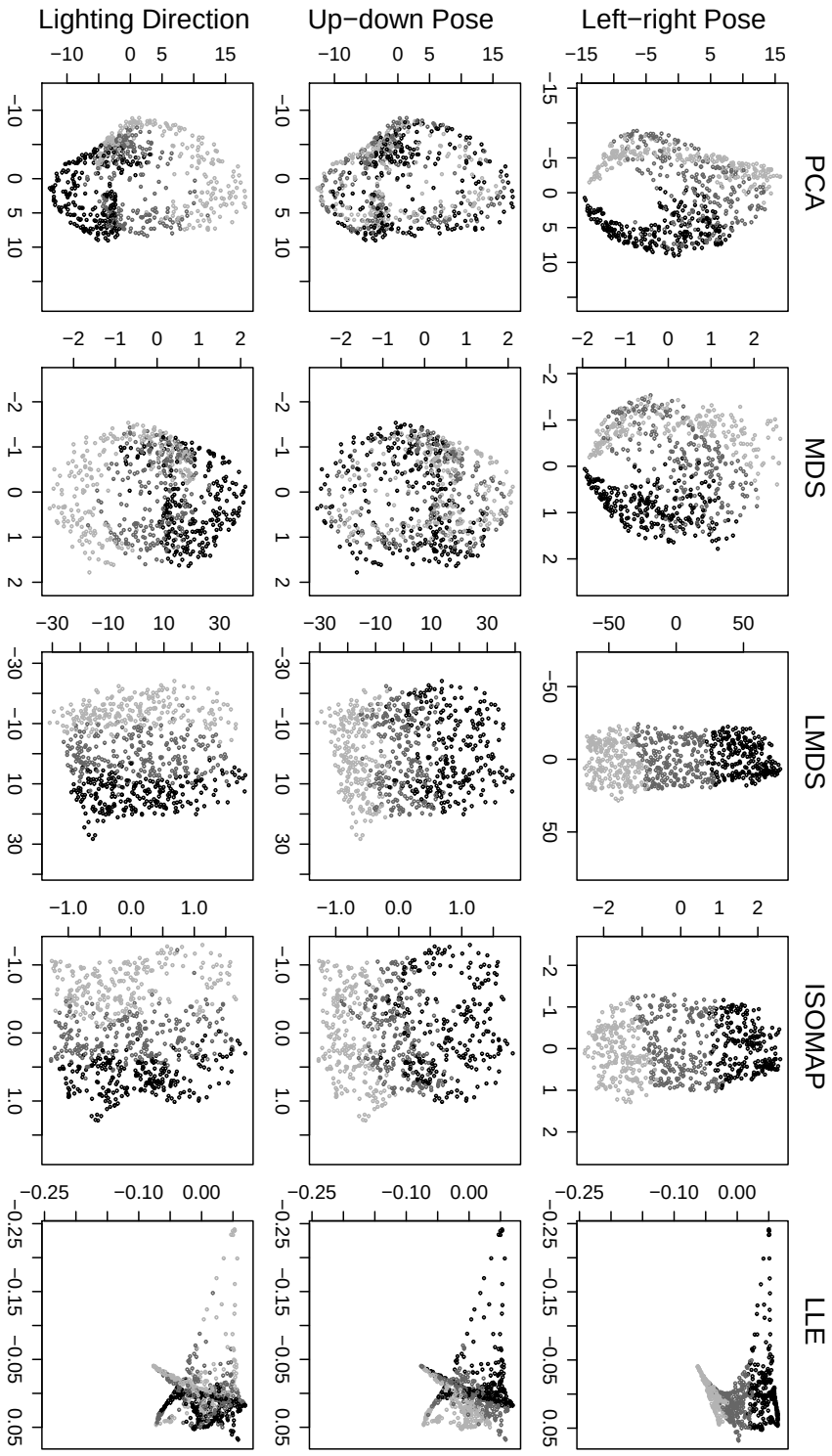

Figure 2: *Sculpture Face Data. Three-D configurations from five methods. The gray tones encode terciles of known parameters indicated in the three column labels, rotated to best separation.*

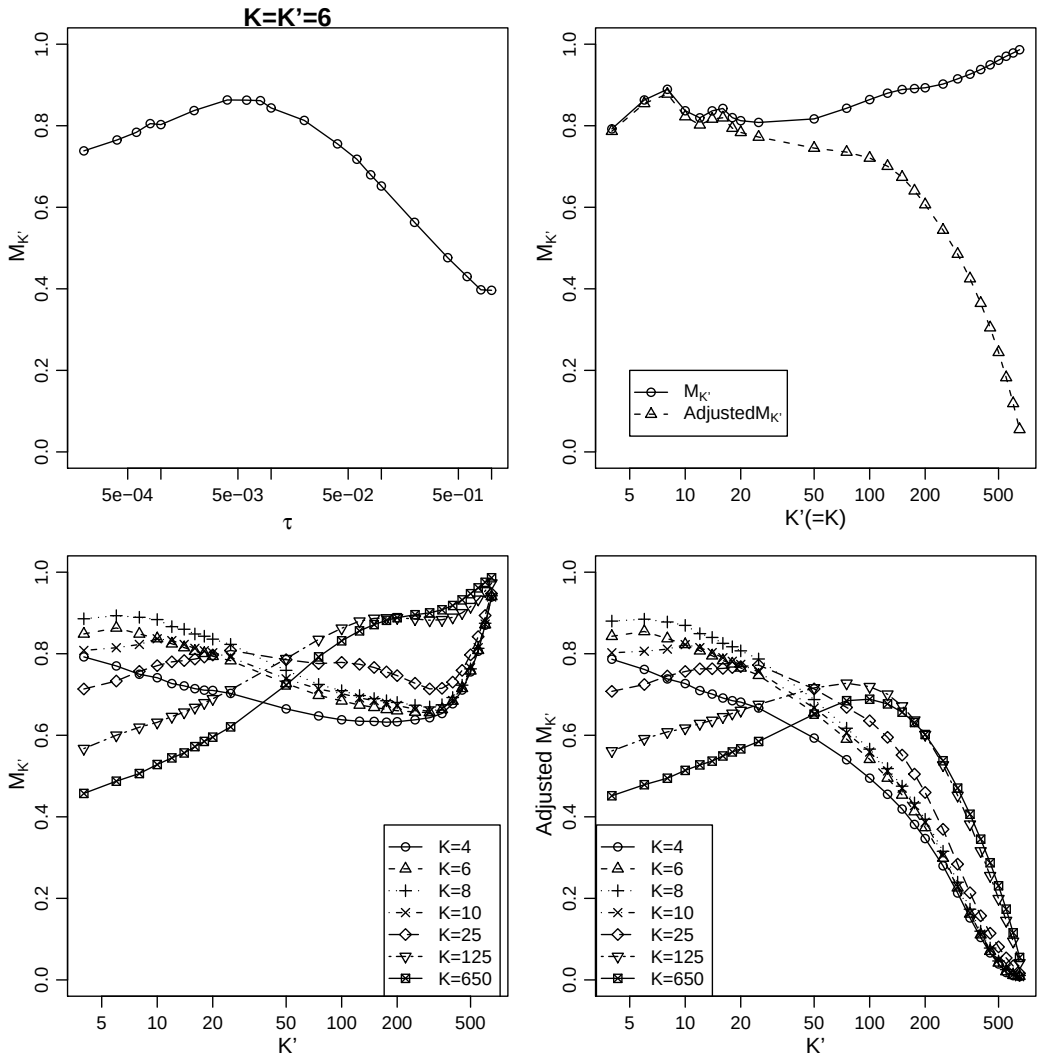

Figure 3: *Sculpture Face Data: Selection of τ and K in the stress function* $\mathrm{LMDS}_{K,\tau}$. **Top Left:** *Trace $\tau \mapsto M_{K'=6}$ for configurations that minimize* $\mathrm{LMDS}_{K=6,\tau}$. *A global maximum is attained near $\tau = .005$.* **Top Right:** *Traces $K' \mapsto M_{K'}$, $M^{adj}_{K'}$ for τ-minimized configurations, constraining $K = K'$ in* LMDS_K. *The adjusted criterion points to a global maximum at $K = K' = 8$.* **Bottom:** *Assessing τ-optimized solutions of* LMDS_K *with traces $K' \mapsto M_{K'}$, $M^{(adj)}_{K'}$ for various values of K: $K = 8$ dominates over a range of K' from 4 to 20.*

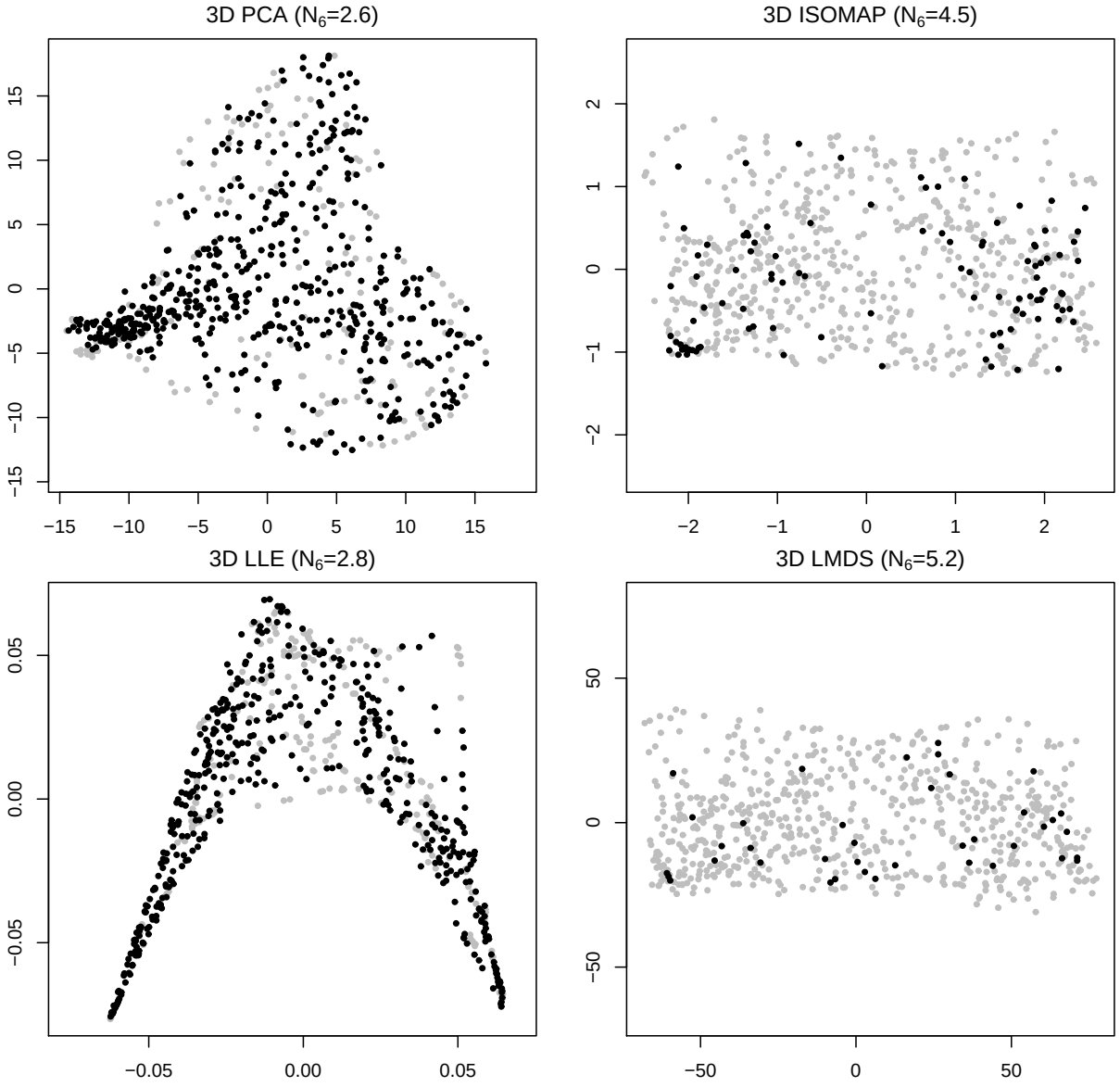

Figure 4: *Sculpture Face Data, pointwise diagnostics for 2-D views of 3-D configurations from four methods. The gray scales code $N_6(i)$ as follows: overlap 0 to 3 $\sim$ black, overlap 4 to 6 $\sim$ gray. A comparison of the density of black points reveals the local dominance of Isomap and LMDS over PCA and LLE, as well as an edge of LMDS over Isomap on these data.*

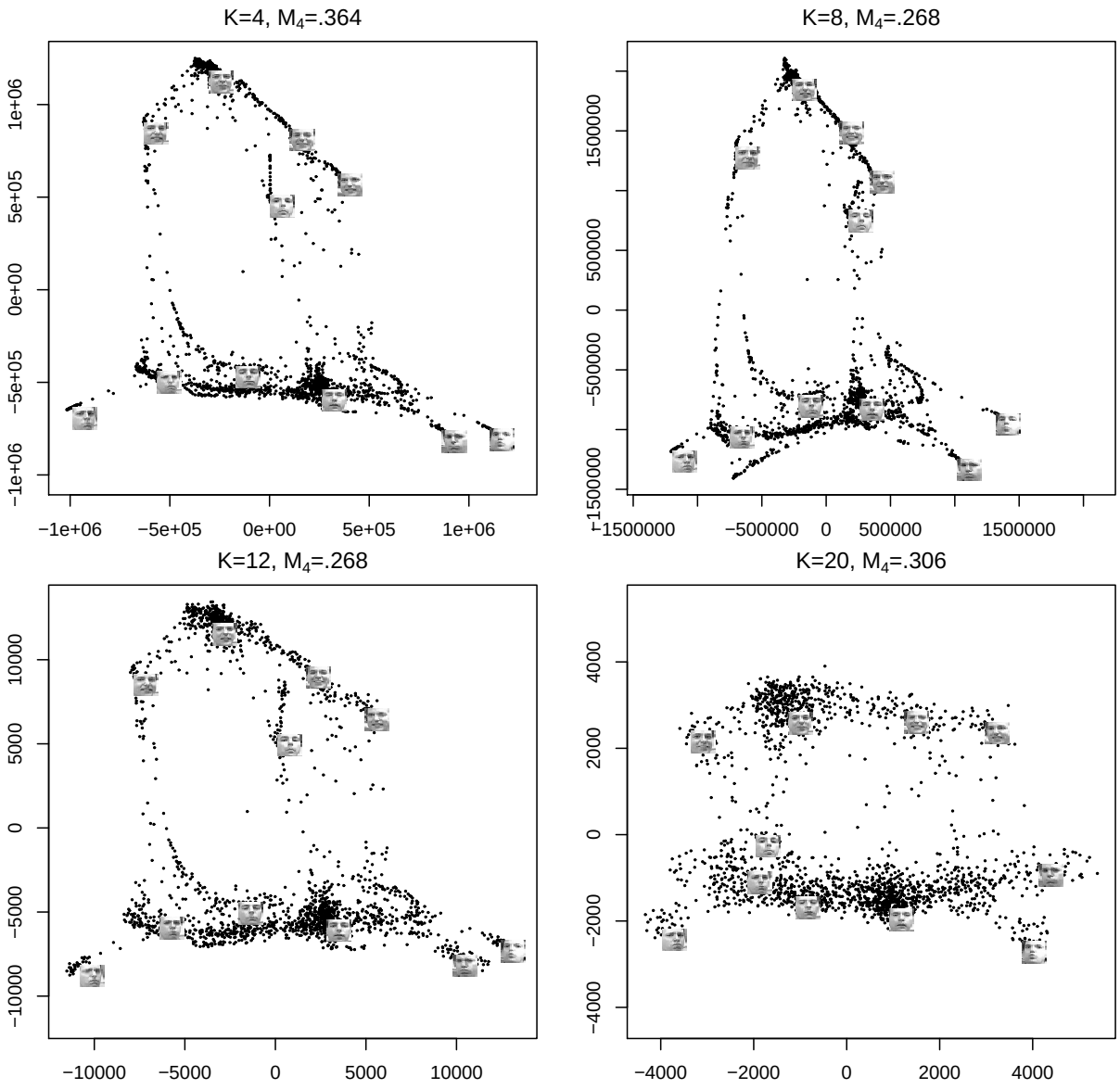

Figure 5: *Frey Face Data. 2-D views of 3-D configurations from LMDS with four choices of* K.

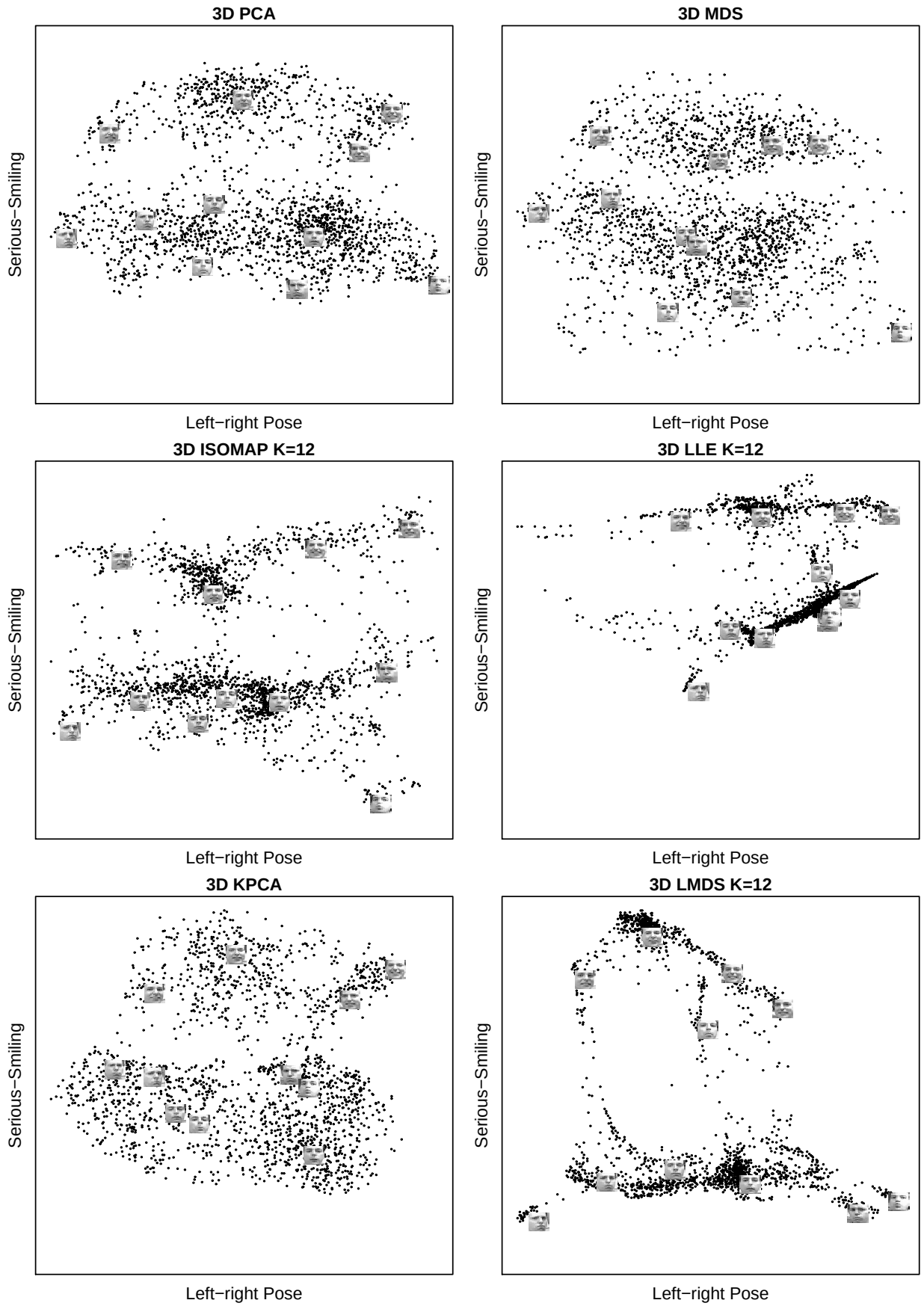

Figure 6: *Frey Face Data. 2-D views of 3-D configurations, comparing PCA, MDS, Isomap, LLE, KPCA and LMDS.*

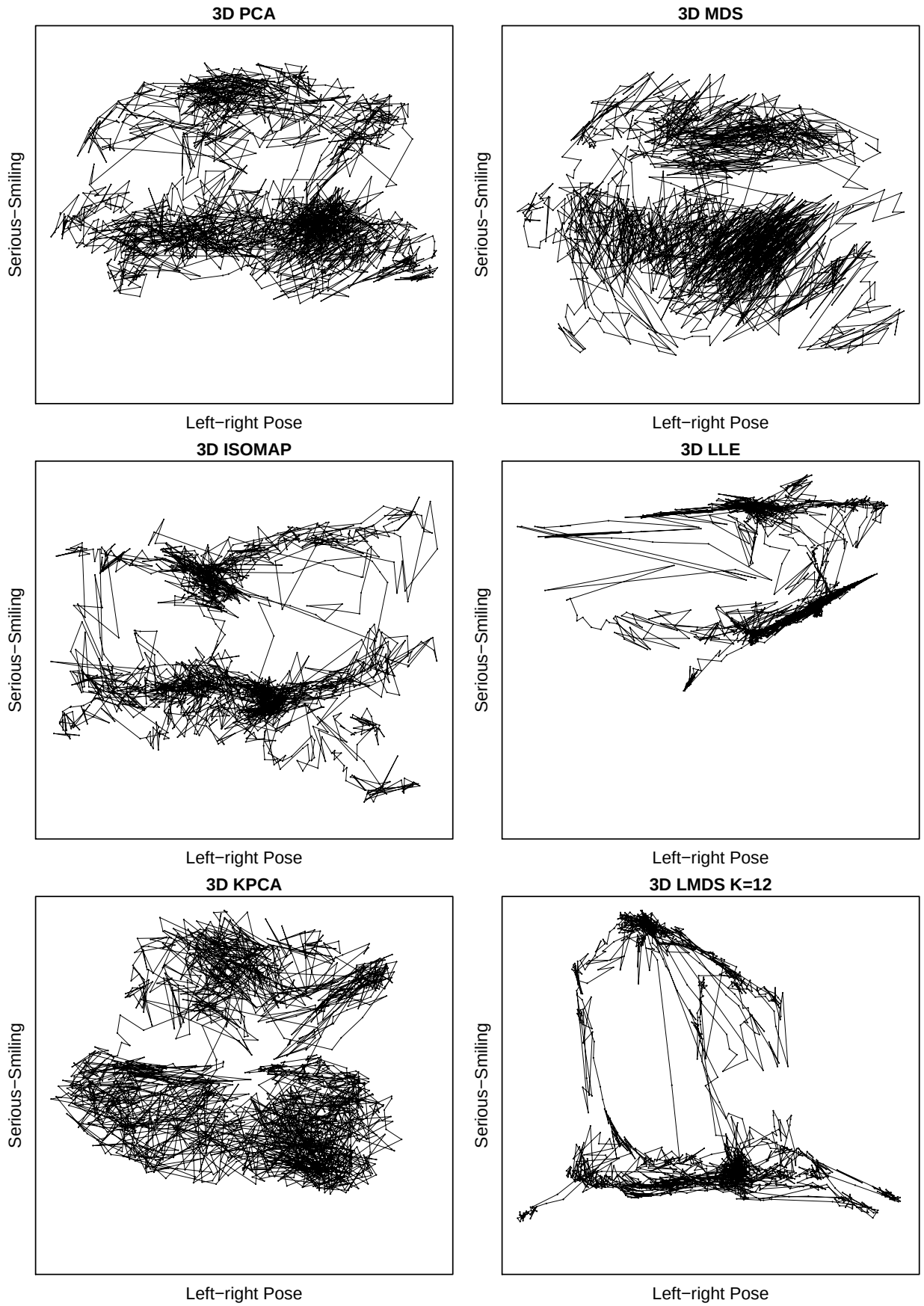

Figure 7: *Frey Face Data. The six configurations (from Figure 6) with connecting lines that reflect the time order of the video footage.*

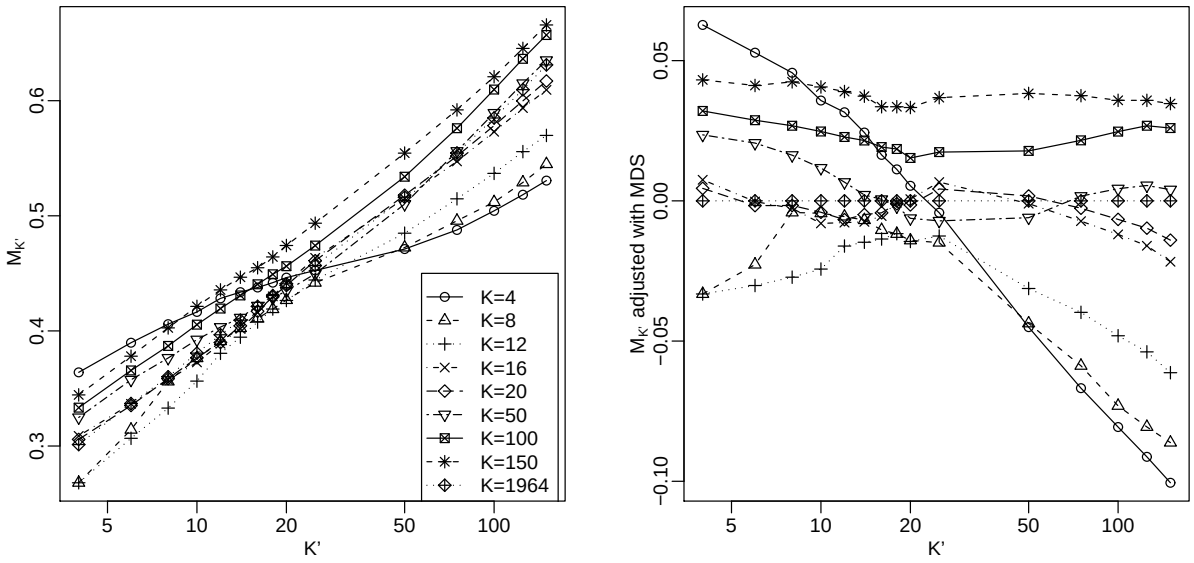

Figure 8: *Frey Face Data. Traces of the Meta-Criterion, $K' \mapsto M_{K'}$, for various choices of the neighborhood size K in the local stress* LMDS_K. *Left: unadjusted fractional overlap; right: adjusted with MDS as the baseline. MDS is LMDS with K=1964, which by definition has the profile $M^{adj}_{K'} \equiv 0$ in the right hand frame.*