Generalized Non-metric Multidimensional Scaling

Sameer Agarwal[*]
Computer Science & Engineering
University of Washington
Seattle, WA 98105

Josh Wills[†]
Sony Pictures Imageworks
Culver City, CA 90232

Lawrence Cayton
Compute Science & Engineering
University of California, San Diego
La Jolla, CA 98105

Gert Lanckriet
Electrical & Computer Engineering
University of California, San Diego
La Jolla, CA 98105

David Kriegman
Computer Science & Engineering
University of California, San Diego
La Jolla, CA 98105

Serge Belongie
Computer Science & Engineering
University of California, San Diego
La Jolla, CA 98105

Abstract

We consider the non-metric multidimensional scaling problem: given a set of dissimilarities Δ, find an embedding whose inter-point Euclidean distances have the same ordering as Δ. In this paper, we look at a generalization of this problem in which only a set of order relations of the form $d_{ij} < d_{kl}$ are provided. Unlike the original problem, these order relations can be contradictory and need not be specified for all pairs of dissimilarities. We argue that this setting is more natural in some experimental settings and propose an algorithm based on convex optimization techniques to optimally solve this problem. We apply this algorithm to human subject data from a psychophysics experiment concerning how reflectance properties are perceived. We also look at the standard NMDS problem, where a dissimilarity matrix Δ is provided as input, and show that we can always find an order-respecting embedding of Δ.

1 Introduction

Multidimensional scaling (MDS) refers to the general task of assigning Euclidean coordinates to a set of objects such that a given set of dissimilarity, similarity or ordinal relations between the points are obeyed. This assignment of coordinates is also known as a Euclidean embedding. Multidimensional scaling algorithms fall into two broad classes: non-metric algorithms, which find an embedding respecting only the relative ordering of the input dissimilarities; and metric algorithms, which seek an embedding with inter-point distances closely matching the input dissimilarities. We focus on the non-metric MDS problem in this paper.

Non-metric MDS has been used extensively in the psychometrics and psychophysics communities to embed similarity and dissimilarity ratings derived from a variety of sources. Metric MDS is not appropriate in many of these applications since the magnitude of the input dissimilarities is unreliable, too difficult to measure, or simply unavailable. As a concrete example, suppose we wish to asses the perceptual similarity of some visual stimuli. We could ask human subjects to rate the similarity of these objects on a scale from one to a hundred, then embed these similarity ratings to visualize them. Unfortunately, different users will likely use different internal scales to asses similarity. Moreover, because of drift effects, a subject will sometimes rate a single stimulus differently depending on the order in which the stimuli where presented. Thus, the actual numbers that the users give are typically not reliable; however, the relative ordering of them will be fairly consistent [7]. Because of the inconsistency of user ratings, non-metric MDS is more appropriate in this setting than metric MDS.

Let us formalize the standard non-metric MDS problem.

Problem 1 (Shepard-Kruskal Scaling)**.** *Given a symmetric zero diagonal matrix* $\Delta = [\delta_{ij}]$, *find* $X = [\mathbf{x}_i] \in \mathbb{R}^{d\times n}$ *such that*

$$\forall\ i,j,k,l \quad \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 < \|\mathbf{x}_k - \mathbf{x}_l\|_2^2 \iff \delta_{ij} < \delta_{kl}. \tag{1}$$

This problem was first considered by Shepard [13, 14], but it was Kruskal who posed the problem as an opti-

[*]Part of this work was done while the author was a student at University of California, San Diego

[†]Part of this work was done while the author was a student at University of California, San Diego

mization problem and introduced an alternating minimization procedure for solving it [9, 8, 3, 2]. We review the Shepard-Kruskal algorithm in the next section.

A curiosity of the the Shepard-Kruskal formulation of non-metric MDS is that it actually requires *magnitudes* as input, even though NMDS concerns only ordinal information. Indeed, one of the motivations for NMDS is to avoid the difficulties associated with collecting accurate magnitude information. Another quirk of the Shepard-Kruskal formulation is that it requires *all* order comparisons—one cannot be agnostic about the ordering of any pair of dissimilarities. Collecting all order comparisons may be difficult or impossible in some experiments: We report a reflectance study in which only a small number of comparisons are available. These two issues lead us to a more general formulation of the NMDS problem.

Problem 2 (Paired Comparisons)**.** *Given a set $\mathcal{S}$ of quadruples, find $X \in \mathbb{R}^{d\times n}$ such that*

$$(i,j,k,l) \in \mathcal{S} \iff \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 < \|\mathbf{x}_k - \mathbf{x}_l\|_2^2 \qquad (2)$$

This problem subsumes Problem 1 since we can take the order information in a dissimilarity matrix and convert it to paired comparisons.

We begin by reviewing the Shepard-Kruskal problem and algorithm in Section 2. Next, we derive a convex program for Problem 2 in Section 3. In Section 4 we apply the novel algorithm to subject data on surface reflectance and use it to construct a perceptual space for a reflectance. We conclude with a discussion.

2 The Shepard-Kruskal Problem and Algorithm

In this section, we consider the Shepard-Kruskal formulation of NMDS. We first review the the classical Shepard-Kruskal scaling (SKS) algorithm, which has been applied extensively.

The SKS algorithm is based on minimizing the stress-1 functional

$$\sigma_1(X) = \min_{\theta} \frac{\sum_{ij}\left(\|\mathbf{x}_i - \mathbf{x}_j\|_2 - \theta\left(\delta_{ij}\right)\right)^2}{\sum_{ij}\|\mathbf{x}_i - \mathbf{x}_j\|_2}, \qquad (3)$$

where $\theta(\cdot)$ is an arbitrary monotonic function of its arguments. The cost function is minimized with respect to the embedding, X. Shepard's routine for minimizing σ_1 involves iterating two alternating steps: fixing θ and minimizing over X, then fitting θ to the current embedding X. X is fit to a particular transformation of the dissimilarities using gradient descent. Unfortunately, $\sigma_1(X)$ is not convex in X, so this minimization step can easily get caught in local minima– especially considering that the minimization is over a large number of parameters (the number of points times their dimensionality). Moreover, the proper dimensionality of X must be guessed *a priori* and can have a substantial effect on the resulting embedding.

In the second step, θ is aligned with the current X with isotonic regression. Note that since σ_1 is a minimum over all monotonic transformations of δ, only the ordering of the the dissimilarities effects the value of the cost function; any magnitude information is swept away by the monotonic transformation.

The construction of the SKS problem naturally leads one to an interesting theoretical question: *which dissimilarity matrices can be embedded perfectly—i.e., with no order violations?* The following lemma shows that, perhaps surprisingly, *any* dissimilarity can be perfectly embedded and give an explicit construction to do so. This lemma reveals that the Shepard-Kruskal representation—the dissimilarity matrix—is significantly weaker than the paired comparisons representation since it is easy to construct an input in paired comparisons form that cannot be embedded into Euclidean space perfectly (or even embedded into a metric space). Put differently, there are many instances of the paired comparisons problem that cannot be handled by an algorithm for SKS.

Lemma 1. *Given a symmetric matrix Δ with zero diagonal, there exists a matrix X such that*

$$\forall\ i,j,k,l \quad \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 < \|\mathbf{x}_k - \mathbf{x}_l\|_2^2 \iff \delta_{ij} < \delta_{kl}. \qquad (4)$$

Proof. Let $V = I - \frac{1}{n}\mathbf{1}\mathbf{1}^n$ be the centering matrix. Then define $K = -\frac{1}{2}V\Delta V$. By construction, K is a symmetric matrix and the following relation between the entries of Δ and K holds

$$\delta_{ij} = k_{ii} - 2k_{ij} + k_{jj}. \qquad (5)$$

This equality is demonstrated in the standard derivation of classical MDS [2]. Let $\lambda_{\min}(K)$ denote the algebraically smallest eigenvalue of K. It is easy to show that $K' = K - \min(\lambda_{\min}(K), 0)I$ is positive semidefinite. Further, decomposing $K' = X^\top X$, we have

$$\begin{aligned}\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 =& k'_{ii} - 2k'_{ij} + k'_{jj} \\ =& k_{ii} - \min(\lambda_{\min}(K), 0) - 2k_{ij} + k_{jj} \\ & - \min(\lambda_{\min}(K), 0) \\ =& \delta_{ij} - 2\min\left(\lambda_{\min}\left(-\frac{1}{2}V\Delta V\right), 0\right). \end{aligned} \qquad (6)$$

Thus, X is a Euclidean embedding of Δ such that all pairwise distances are a constant shift of the input dissimilarities, and thus all relative comparisons are preserved. □

The lemma gives a very fast, simple algorithm for NMDS. In fact, this algorithm has been suggested for *metric* MDS [3]. However, that this construction can be used to find perfect non-metric embeddings does not appear to be known. The algorithm is equivalent to adding a constant $(-\min(\lambda_{\min}, 0))$ to all of the dissimilarities—in other words, we can convert any dissimilarity matrix into a squared Euclidean distance matrix by simply increasing all dissimilarities by a constant. In practice, this method tends to push all dissimilarities to nearly the same value, and results in high-dimensional solutions. Though this lemma is an interesting theoretical point, the actual construction is not a particularly practical one.

3 Paired Comparisons

In this section we present a novel algorithm for learning a low rank embedding from a collection of paired comparisons. Our method is related to the recent work on distance metric learning [12, 16, 19, 20].

Let $\mathcal{S}$ be a set of 4-tuples (i, j, k, l). We hope to find an embedding X such that

$$\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \leq \|\mathbf{x}_k - \mathbf{x}_l\|_2^2 \quad \forall\ (i, j, k, l) \in \mathcal{S}. \tag{7}$$

The set $\mathcal{S}$ can have repetitions and inconsistencies. We now derive a semidefinite program for finding such an embedding X. Our program will find a Gram matrix $K = X^\top X$, which we will then decompose to recover X.

We can rewrite the distances in terms of the Gram matrix:

$$\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = k_{ii} - 2k_{ij} + k_{jj}. \tag{8}$$

Then, we can express the paired comparisons in $\mathcal{S}$ as linear inequalities on the Gram matrix:

$$k_{ii} - 2k_{ij} + k_{jj} < k_{kk} - 2k_{kl} + k_{ll} \quad \forall\ (i, j, k, l) \in \mathcal{S}.$$

Our aim is to find a positive semidefinite matrix $K \succeq 0$, which satisfies these inequality constraints.

The set of inequality constraints above are not sufficient to determine K uniquely. For example, if X satisfies the set of paired comparisons, so will any scaling, translation or rotation of X. The scale must be fixed to avoid the trivial embedding and we fix the translation ambiguity to avoid numerical instability issues. The translation ambiguity is eliminated by demanding that the embedding be centered at the origin, i.e.,

$$\sum_b \mathbf{x}_b = 0, \tag{9}$$

which can be shown to be equivalent to

$$\sum_{ab} k_{ab} = 0. \tag{10}$$

This is a linear equation of the entries of the matrix K.

Handling the scale ambiguity is a bit more complicated. To prevent the embedding from collapsing into the origin, we constrain the scale of the embedding from below. We will demand that for a relative comparison to be valid the two distances should be different by at least 1 unit distance, giving the constraint

$$k_{kk} - 2k_{kl} + k_{ll} \geq k_{ii} - 2k_{ij} + k_{jj} + 1. \tag{11}$$

Note that the choice of the constant 1 was essentially arbitrary; any positive constant will work. Collecting the constraints that we have imposed on our Gram matrix thus far, we get the following semidefinite feasibility problem

$$\begin{gathered} k_{kk} - 2k_{kl} + k_{ll} \geq k_{ii} - 2k_{ij} + k_{jj} + 1 \quad \forall (i, j, k, l) \in \mathcal{S} \\ \sum_{ab} k_{ab} = 0, \quad K \succeq 0. \end{gathered} \tag{12}$$

We can find a solution to this feasibility program only if there is an embedding satisfying all of the constraints. In general, it may not be possible to satisfy all of the constraints. We thus introduce a slack variable ξ_{ijkl} for each inequality constraint to allow for inequality violations. Our objective is to minimize the total amount of slack, leading to the following semidefinite program:

$$\begin{aligned} \min_{K, \xi_{ijkl}} \quad & \sum_{(i,j,k,l) \in \mathcal{S}} \xi_{ijkl} \\ \text{subject to} \quad & k_{kk} - 2k_{kl} + k_{ll} - k_{ii} + 2k_{ij} - k_{jj} \geq 1 - \xi_{ijkl} \\ & \sum_{ab} k_{ab} = 0, \quad \xi_{ijkl} \geq 0, \quad K \succeq 0. \end{aligned} \tag{13}$$

Once we find the optimal K for this program, we can recover the coordinates of X by spectrally decomposing K into $U \Lambda U^\top$ and setting $X = U \Lambda^{1/2}$.

In many instances a user may desire a *low-dimensional* embedding. Towards this goal, we add a regularizer to the above optimization problem that trades off the embedding complexity with the fitting error (the total slack):

$$\min_{K,\xi_{ijkl}} \quad \sum_{(i,j,k,l)\in\mathcal{S}} \xi_{ijkl} + \lambda\,\mathrm{rank}(K)$$
$$\text{subject to} \quad k_{kk} - 2k_{kl} + k_{ll} - k_{ii} + 2k_{ij} - k_{jj} \geq 1 - \xi_{ijkl}$$
$$\sum_{ab} k_{ab} = 0, \quad K \succeq 0. \tag{14}$$

This λ is a positive scalar that controls the tradeoff between the violations and the rank of the matrix, i.e., the complexity of our model. Unfortunately, the rank of a matrix is a non-convex function and minimizing the rank of a symmetric positive semidefinite matrix subject to linear inequality constraints is NP-Hard [5]. We thus relax the rank function to its convex envelope, the trace. This relaxation is standard in the convex programming literature. Using this relaxation has the additional benefit of constraining the scale of the embedding from above. We are left finally with the following semidefinite program.

$$\min_{K,\xi_{ijkl}} \quad \sum_{(i,j,k,l)\in\mathcal{S}} \xi_{ijkl} + \lambda\,\mathrm{Trace}(K)$$
$$\text{subject to} \quad k_{kk} - 2k_{kl} + k_{ll} - k_{ii} + 2k_{ij} - k_{jj} \geq 1 - \xi_{ijkl}$$
$$\sum_{ab} k_{ab} = 0, \quad K \succeq 0. \tag{GNMDS}$$

This program can be solved using a general purpose semidefinite programming package. We obtained the best results using Sedumi [15]. Unfortunately, general purpose semidefinite programming solvers scale poorly with the problem size. To handle larger problem sizes, we implemented a first order alternating projections algorithm [4, 6]. Though this method converges much more slowly than interior-point based methods for moderately sized problems, it scales to large problem sizes and has minimal memory requirements.

4 Experiments

4.1 A Toy Experiment

As a simple illustration and comparison of the performance of our algorithm, we consider the 10 cities dataset which consist of the pairwise distance between 10 cities in the United States [2]. Figure 1 plots the result of running Classical MDS on this dataset in red. The 10×10 distance matrix was used to generate the set of 44 paired comparisons by first sorting the entries of the distance matrix and comparing adjacent entries in the sorted list. The result of running the GNMDS algorithm on this dataset is plotting in blue. The two embeddings were aligned upto a translation and global scale before plotting. We note that while the two embeddings are not identical, and this is expected given

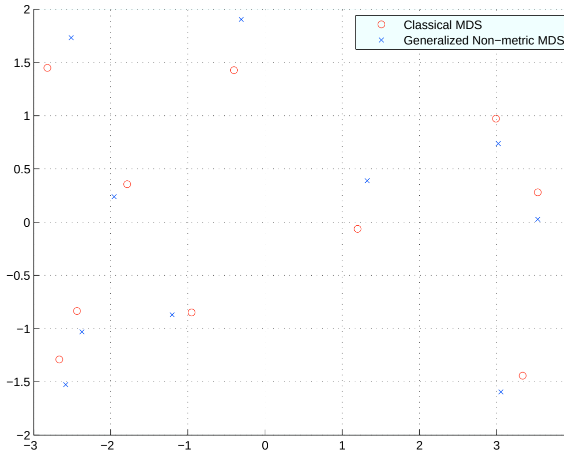


Figure 1: Comparing Classical multidimensional scaling and Generalized non-metric multidimensional scaling. We plot the result of running classical MDS on the 10 cities dataset in red and the result of running the GNMDS algorithm on the set of all pair comparisons generated from the same dataset in blue.

the lack of metric information, they are quite close and capture all the major qualitative features of the embedding produced by the Classical MDS algorithm.

4.2 A Perceptual Space for Reflectance

The primary motivation for the development of the generalized non-metric multidimensional scaling (GNMDS) algorithm was the analysis of how humans perceive the reflection of light from a surface. The Bidirectional Reflectance Distribution Function (BRDF) is a mathematical description of how a surface reflects light. For every incident direction it describes the angular distribution of reflected light [11]. The shape of these distributions determine whether a material appears rough and matte or shiny and metallic.

To understand how humans perceive the reflection of light from different kinds of surfaces we wish to construct a space/embedding for these BRDFs in which the pairwise Euclidean distances between BRDFs are indicative of the *perceptual* distance between them.

To this end, we conducted the following psycho-physics experiment. Each participant in the experiment was shown a series of triplets of rendered images. Each image consisted of the Stanford Bunny [17] rendered under constant illumination and viewing direction. The only difference was material/BRDF used to describe how the surface of the bunny reflects light. For each triplet the subjects were asked to indicate whether the center image was more similar to the image on the left or to the image on the right (Figure 2(a) shows a screen-shot from one such test). The use of triplets

(a) (b)

Figure 2: (a) Screen capture from the distance comparison test. (b) Six of the 55 images used in our psychophysics study.

in this manner is a special case of Problem 2, where $j = l$.

A total of 55 BRDFs from the MIT/MERL BRDF database [10] were used. In this study we restrict our attention to the achromatic aspects of the BRDF, also known as gloss (examples of some of these BRDFs appear in Figure 2(b)). While monochromatic, they have widely varying gloss properties. The BRDFs used include metals, paints, fabrics, minerals, synthetics, and organic materials. Because there are over $78,000$ possible triplets, only a randomly sampled subset of comparisons could be performed. In our study 75 subjects performed 200 comparisons each for a total of 15,000 comparisons (there were a small number of repeated comparisons). The triplets were chosen at random for each subject[1].

The data collected in the above experiment cannot be analyzed using Shepard-Kruskal Scaling. The set of observations do not define a total order on the set of distances/similarities. Further it contains contradictions and repetitions which the Shepard-Kruskal formulation is unable to handle. In the following we describe the result of analyzing this data using the GNMDS algorithm.

4.2.1 Perceptual Properties of the Embedding

We begin by considering the embedding produced by running the GNMDS algorithm on the data from the psychophysics experiment described in the previous section.

Figure 3(a) shows the 2-D embedding with cropped windows of the BRDF images displayed in the locations of the BRDF in the new space. Notice the clustering of the BRDFs into two distinct clumps and the similarity amongst the images corresponding to them in each clump. There are also two pronounced trends in the embedding, a vertical trend with the darker BRDFs at the top gradually getting brighter with the brightest BRDFs at the bottom. Horizontally the embedding can be roughly broken up into two clusters: the primarily diffuse BRDFs and those that have a strong glossy or specular component.

The American Society for Testing and Materials defines six dimensions for the perception of gloss. Figure 3(b)-(d) show plots of three of the ASTM gloss dimensions in our embedding space. The position of each circle corresponds to one of the BRDFs in the embedding space and the diameter corresponds to the measurement of the BRDF in the ASTM gloss dimension. Figure 3(b) shows the measurements of each BRDF for contrast gloss (a measure of diffuseness), Figure 3(c) shows the measurements for specular gloss at 20° (a measure of specularity), and Figure 3(d) shows the measurements for haze (a more robust measure of specularity less prone to measurement error).

Notice the strong correlation between the spatial layout of the embedding and the radii of the circles.

4.2.2 Quantitative Analysis

In the last section we considered the qualitative aspects of the embedding. In this section we present a quantitative analysis. Given a set of subject responses to paired comparisons on the same 55 BRDFs, we measure the error of an embedding as the average number of paired comparisons that are violated if we use the pairwise distance between BRDFs in the embedded space as our estimate of the distance between them.

We ran our MDS algorithm on the data for varying values of λ between 0 and 300, and performed 10-fold cross-validation for each value of λ. Figure 4(a) plots the training (Red) and testing error (Green) as a function of the parameter λ. Figure 4(b) plots the average rank of the embedding as a function of λ. As expected, the rank of the embedding goes down as λ is increased and the gap between the training and testing errors is reduced.

The embedding with the lowest measured cross-

[1] For additional details about the experiment we refer the reader to [1].

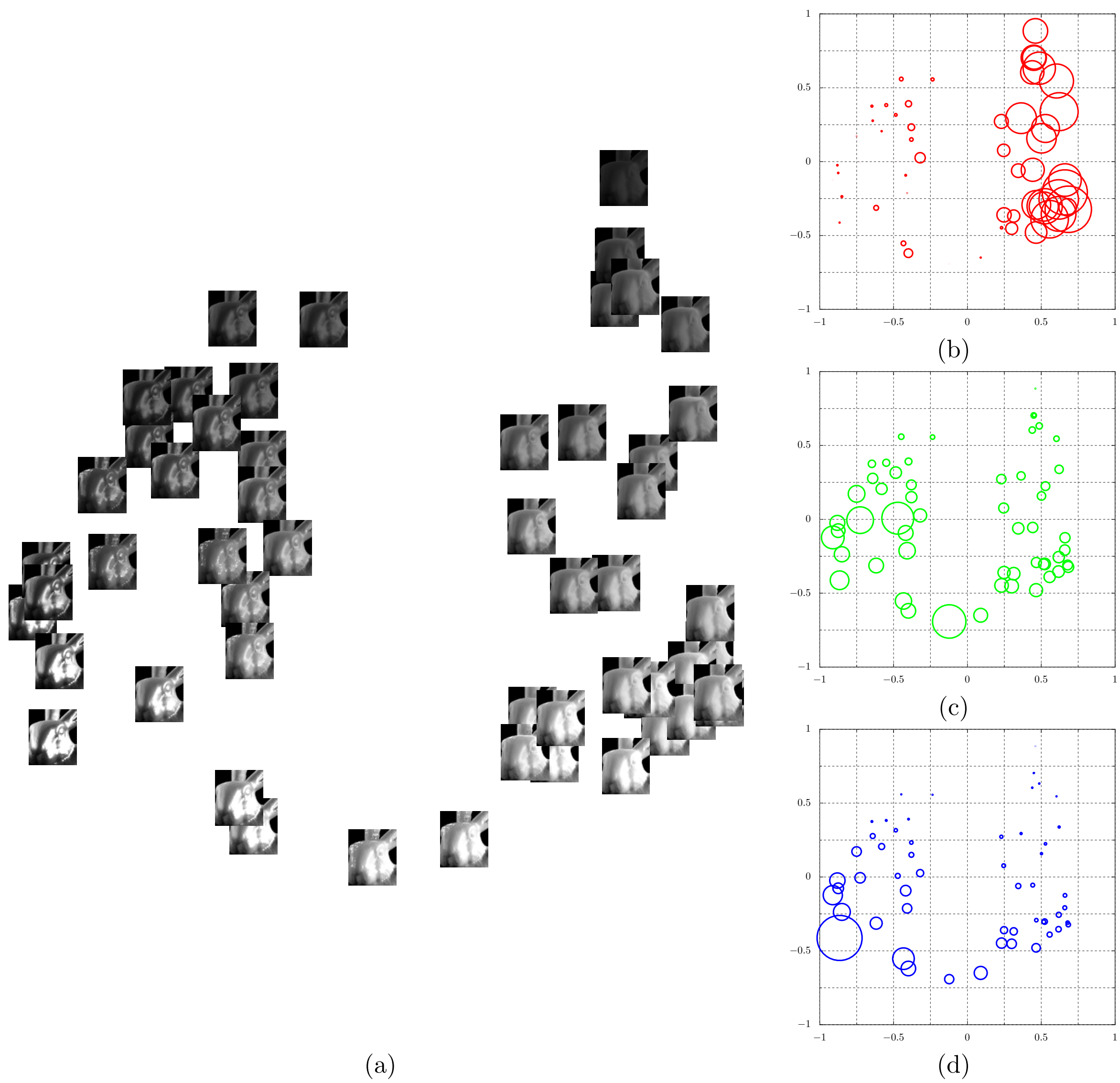
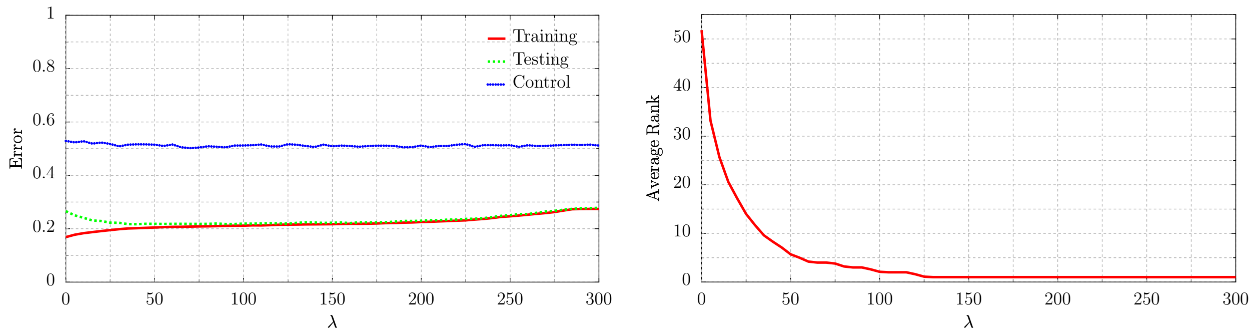


Figure 3: Perceptual Embedding. (a) The 2-D embedding with cropped windows of the BRDF images displayed in the locations of the BRDF in the new space. (b-d): Contrast Gloss (b), Specular Gloss (c) and Haze (d) values shown for BRDFs in the embedding. The diameter of the circles corresponds to the value for each property.



(a) (b)

Figure 4: Cross Validation and Rank. (a) Training (red) and Testing (green) error curves for varying choices of the regularization parameter λ for our MDS algorithm. Testing error (blue) for the randomized control set. (b) Average rank as a function of the regularization parameter.

validation error has a training error of 21.9% and a test error of 21.3%. The embedding has over 95% of the variance contained in the first two dimensions. Truncating the embedding at two dimensions increased the test error insignificantly by 0.5%. This embedding is a significant improvement in terms of test error as well as the complexity of the embedding obtained with $\lambda = 0$.

We found that two different subjects, on average, disagree on about 17% of the orderings. Since the data is derived from these subjects, 17% is a natural lower bound on the amount of error achievable. On the other hand, random guessing leads to an error of about 50%. Thus the embedding that GNMDS finds is very nearly the best possible in terms of error.

As another comparison, suppose we calculate the Euclidean distance between the BRDF vectors describing each image. How well do these distances match up with the user orderings? Perhaps surprisingly, these distances have 37.5% error—*i.e.* they do not match up with user orderings very closely. The reason for this is, differences between BRDFs that maybe algebraically interesting, e.g. the exact angular position of the specular lobe, they may not have any perceptual meaning. Thus, the 21.3% error obtained using our embedding is significantly better than using the L_2 norm between BRDFs and is only a few percent worse than the best possible performance on this dataset.

4.2.3 Stability

Despite the large number of subject responses, the embedding consists of only 55 points corresponding to the 55 BRDFS. One potential concern is over-fitting the embedding to the 55 BRDFs used; perhaps a small change in the set of BRDFs could affect the embedding substantially.

To check for this possibility, we measured the stability of the embedding with respect to deletions. We constructed 55 different embeddings corresponding to leaving the response data corresponding to one of the BRDFs out at a time. Each of the embeddings produced in this manner was then aligned via a similarity transformation to its corresponding 54 points in the final embedding reported above, and the average squared distortion was measured [18]. Note that paired comparisons are invariant to similarity transformations. To establish a scale for these errors, the average distance between pairs of points in the global embedding was calculated. The root mean squared distortion was 2.7e-2 while the average distance between points in the global embedding was 8.7e-1. This is an error of only 3%, indicating a high degree of stability across embeddings.

5 Discussion

We considered a generalized formulation of non-metric multidimensional scaling which subsumes the Kruskal-Shepard formulation. We presented an algorithm based on convex optimization techniques for finding solutions to the generalized problem. Finally, we used this algorithm to analyze human subject data from a psychophysics experiment designed to measure the human perception of material reflection.

There are a couple of directions for future research. The first is the problem of *out-of-sample extensions*. Is it possible to add an object to an embedding using some paired comparisons without re-running the entire semidefinite program? A different avenue of research concerns the issue of sampling pairwise comparisons. Is there an active sampling strategy that can be used to find a good embedding? Such an approach will be critical to the practicality of applying GNMDS in large psychophysical studies. Finally, we are interested in the development of an efficient, special-purpose large-scale solver for the semidefinite program that exploits knowledge of the problem structure.

Acknowledgement

Sameer Agarwal was supported by NSF grant EIA-0321235, the University of Washington Animation Research Labs, the Washington Research Foundation, Adobe, and Microsoft. This work was partially supported by an NSF IGERT Grant (Vision and Learning in Humans and Machines, #DGE-0333451), NSF CAREER #0448615 and The Alfred P. Sloan Research Fellowship.

References

[1] Anonymous. Toward a perceptual space for reflectance. (in review).

[2] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer Verlag, 2005.

[3] T.F. Cox and M.A. Cox. *Multidimensional Scaling*. Chapman & Hall/CRC, 2000.

[4] R. L. Dykstra. An algorithm for restricted least-squares regression. *Journal of the American Statistical Association*, 78:837–842, 1983.

[5] M. Fazel, H. Hindi, and S. Boyd. Rank minimization and applications in system theory. In *Proc. of Americal Control Conf.*, 2004.

[6] S. P. Han. A successive projection method. *Mathematical Programming*, 40:1–14, 1988.

[7] M. Kendall and Kean Dickinson Gibbons. *Rank Correlation Methods*. Oxford University Press, 1990.

[8] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27, 1964.

[9] J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29:115–129, 1964.

[10] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillian. A data-driven reflectance model. In *SIGGRAPH '03*, pages 759–769, 2003.

[11] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. Monograph 161, National Bureau of Standards (US), October 1977.

[12] Matthew Schultz and Thorsten Joachims. Learning a distance metric from relative comparisons. In *NIPS*, 2003.

[13] R.N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27(2):125–140, 1962.

[14] R.N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27:219–246, 1962.

[15] J.F. Sturm. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11-12:625–653, 1999.

[16] I. Tsang and J. Kwok. Distance metric learning with kernels. In *Proc. Int Conf Art Neural Nets, 2003.*, 2003.

[17] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *SIGGRAPH '94*, pages 311–318, 1994.

[18] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *PAMI*, 13(4):376–380, April 1991.

[19] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS 2005*. MIT Press, 2006.

[20] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. In *NIPS 2002*, pages 505–512. MIT Press, 2003.