2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)

Lost Without a Compass: Nonmetric Triangulation and Landmark Multidimensional Scaling

Mark A. Davenport
School of Electrical and Computer Engineering
Georgia Institute of Technology
Email: mdav@gatech.edu

***Abstract*—Suppose that we wish to determine the location of a point $x^* \in \mathbb{R}^k$ by comparing the distance from x^* to various points $x_1, x_2, \dots, x_k \in \mathbb{R}^k$ with known position. In this paper we consider the scenario where we are only provided with (possibly noisy or contradictory) relations of the form $\|x^* - x_i\|_2 < \|x^* - x_j\|_2$. We propose a simple algorithm that uses convex optimization techniques to estimate x^* from such data and we provide simulations demonstrating its effectiveness in practice.**

I. Introduction

In this paper we consider the problem of *triangulation*, i.e., determining the position of a point in Euclidean space based on comparisons to a small set of known *landmark points*. Triangulation[1] arises in a variety of contexts where we would like to estimate the positions of various objects in two- or three-dimensional space, such as sensors in a sensor network [2–4], the positions of hydrogen atoms in nuclear magnetic resonance spectroscopy [5, 6], or simply the location of a single object in space. Triangulation also arises in a variety of contexts where we wish to localize an object in a space with $k > 3$ dimensions. For example, in *multidimensional scaling* (MDS) we wish to determine a configuration of points in a low-dimensional space from a set of pairwise distance measurements [7, 8]. The problem of triangulation arises in this context when we have already determined the configuration of a set of landmark points and wish to add a new point using the already learned configuration [1].

Triangulation is by no means a new problem — indeed, the essential ideas for how to perform this task were laid out by Gemma Frisus in his *"Booklet concerning a way of describing places"* nearly 500 years ago [9]. However, in this paper we consider the case where our observations are *nonmetric*. To be concrete, let $x^* \in \mathbb{R}^k$ denote the true position of the point that we are trying to learn, and let $x_1, x_2, \dots, x_n \in \mathbb{R}^k$ denote the (known) positions of the landmark points. Rather than directly observing the raw distances from x^* to $x_1, x_2, \dots, x_n$, i.e., $\|x^* - x_1\|_2, \|x^* - x_2\|_2, \dots, \|x^* - x_n\|_2$, we instead obtain only paired comparisons of the form $\|x^* - x_i\|_2 < \|x^* - x_j\|_2$ for various pairs (i, j). Our ultimate goal will be to estimate x^* from a set of such inequalities.

[1]Strictly speaking, *triangulation* refers to the process of determining the location of a point by measuring *angles* to the set of landmark points. Our approach will focus on estimating the position using data regarding the *distances* to the landmark points, a process more correctly denoted *trilateration*. However, we will follow the usage of [1] and ignore this subtlety.

Our motivation for considering such nonmetric observations stems from the nature of the data that often arises in many applications of MDS and triangulation. In particular, MDS is often applied in situations where we have a collection of items and hypothesize that it should be possible to embed the items in $\mathbb{R}^k$ in such a way that the distance between points in $\mathbb{R}^k$ corresponds to their "similarity" (with small distances corresponding to similar items), but where "similarity" might be derived from human judgements and difficult to quantify.

As an illustrative example, we will consider one possible architecture for a kind of *collaborative filtering* system [10–12]. In collaborative filtering, we have a set of items that might represent movies, songs, books, or any number of other consumer items, and a large set of users who each provide ratings of some subset of the objects. Our goal is to use the observed ratings to learn a joint model for the items and the users so that we can predict which items a user will prefer. At the core of any approach to this problem is an underlying model of human preference. One popular and effective model of preference is the *ideal point* or *unfolding model* [13], in which it is assumed that we can represent all the items as points in a low-dimensional subspace, and an individual's preferences can be captured by an "ideal point" positioned such that objects closest to this point are the most preferred by the individual.[2] If our observations consisted of accurate measurements of the distance from a user to a particular item, then learning an ideal point model would essentially be a special case of MDS. However, in practice such data is not available. One obstacle in collecting this data is that people are generally not able to reliably distinguish between more than a few categories [18], and so responses are often restricted to a small numerical scale. For example, in many collaborative filtering systems, ratings are restricted to be integers on a scale from 1 to 5, or even simply binary thumbs up/thumbs down or like/dislike responses. In other cases our data may consist of the results of paired comparisons, where a user evaluates a pair of items and indicates whether they are similar or dissimilar, or whether one is preferred to the other. This type of data frequently arises

[2]The ideal point model of preference is subtly different from the model postulating that the underlying "rating matrix" has low-rank, which has received a great deal of attention in recent years in the context of *matrix completion* [14–16]. Both models imply that only a few attributes determine a user's rating, but ideal point models suggest that there is an ideal "concentration level" of each attribute and that "more" of each attribute is not always "better". Empirical studies have suggested that the ideal point model often does a better job of capturing consumer preferences [17].

978-1-4673-3146-3/13/$31.00 ©2013IEEE

when dealing with judgements made by human subjects, since people are typically more accurate and find it easier to make such judgements than to assign numerical scores [19].

In this paper we will restrict our attention to the case where the observations are the results of binary paired comparisons. We will focus on the related problems of nonmetric MDS and nonmetric triangulation. In nonmetric MDS, our goal is to find a configuration of points $x_1, x_2, \ldots, x_n \in \mathbb{R}^k$ that agree with the observed data, which takes the form of a set $\mathcal{S}$ of ordered quadruples such that

$$\|x_i - x_j\|_2 < \|x_k - x_\ell\|_2 \tag{1}$$

for all $(i, j, k, \ell) \in \mathcal{S}$. In nonmetric triangulation, we assume that $x_1, x_2, \ldots, x_n \in \mathbb{R}^k$ are known and we wish to find $x^* \in \mathbb{R}^k$ that agrees with the observed data, which here takes the form of a set $\mathcal{T}$ of ordered pairs such that

$$\|x^* - x_i\|_2 < \|x^* - x_j\|_2, \tag{2}$$

for all $(i, j) \in \mathcal{T}$.

As we will see below, these two problems are closely related. But even ignoring any deeper connections, there is another important link between the problems that lies just below the surface. Specifically, any algorithm for nonmetric triangulation can be exploited to develop efficient algorithms for large-scale nonmetric MDS. This builds on the techniques developed in [1], which showed that a highly effective way to perform MDS on a very large dataset is to select a small number of *landmark points*, apply a traditional MDS algorithm to learn an embedding of these landmark points, and then use triangulation to fill in the rest of the embedding. A key goal of this paper is to develop effective methods for nonmetric triangulation so that this approach can be extended to the nonmetric case.

The remainder of the paper is organized as follows. In Section II we review the approach to nonmetric MDS proposed in [20]. Our approach to nonmetric triangulation is described in Section III. Section IV describes the results of preliminary simulations that demonstrate the effectiveness of this approach in practice, and Section V concludes with a discussion of open questions and future work.

II. Generalized nonmetric MDS

The oldest approaches to nonmetric MDS date back to the work of Shepherd [21, 22] and Kruskal [23, 24] in the early 1960's. These early approaches were based on a simple strategy of essentially trying to learn a set of distances that agreed with the given inequalities by iteratively performing MDS using the estimated distances, and then updating the distances to agree with the given inequalities. While this simple algorithm is easy to implement, it can easily get caught in local minima.

More recently, Agarwal et al. proposed a novel approach to nonmetric MDS based on convex optimization and called *generalized nonmetric MDS* (GNMDS) [20]. Our approach to nonmetric triangulation shares much in common with the approach of GNMDS, and so we will first briefly review the ideas behind the algorithm. To begin, we let X denote the $k \times n$ matrix with columns $x_1, x_2, \ldots, x_n$. This is ultimately the matrix that we would like to recover, but to state the algorithm most concisely, we consider the equivalent problem of recovering the Gram matrix $G = X^T X$. Observe that we can re-express inequality (1) as

$$g_{ii} - 2g_{ij} + g_{jj} < g_{kk} - 2g_{k\ell} + g_{\ell\ell}, \tag{3}$$

where g_{ij} denotes the (i, j)-th element of G. Our goal is to find a Gram matrix G that satisfies these constraints. Of course, as in any MDS problem, G cannot be determined uniquely, since the constraints will be unaffected by any translation or scaling of the dataset. The translation of the embedding is typically fixed to be centered at the origin, which as shown in [20] is equivalent to requiring that

$$\sum_{i,j} g_{ij} = 0.$$

To fix the scale of G and exclude the trivial solution, we enforce a (somewhat arbitrary) margin in (3) by modifying it to be of the form

$$g_{kk} - 2g_{k\ell} + g_{\ell\ell} - g_{ii} + 2g_{ij} - g_{jj} \geq 1.$$

We would now like to find a positive semidefinite matrix G that satisfies these conditions, and will also lead to a low-dimensional embedding. The dimension of the embedding X is exactly the rank of the matrix G, thus, we would like the rank of G to be as small as possible. Of course, finding the matrix of smallest rank that agrees with our constraints is computationally intractable, but it is now well known (see, for example, the recent literature on matrix completion [14–16]) that a good proxy for the rank of a positive semi-definite matrix G is its trace. This leads to the optimization problem of

$$\begin{aligned} \min_{G} \quad & \mathrm{Trace}(G) \\ \text{subject to} \quad & g_{kk} - 2g_{k\ell} + g_{\ell\ell} - g_{ii} + 2g_{ij} - g_{jj} \geq 1 \\ & \sum_{i,j} g_{ij} = 0, \quad G \succeq 0. \end{aligned}$$

The optimization problem above would find a Gram matrix G that agrees with every paired comparison in $\mathcal{S}$. This approach would likely over-fit the data (and may not even have a feasible solution in the case where $\mathcal{S}$ contains contradictory inequalities), which would lead to a G having unnecessarily high rank. To allow for some of these constraints to be violated, [20] introduces slack variables $\xi_{ijk\ell}$, but also adds an ℓ_1 penalty term to the objective function to ensure that ξ is relatively sparse, i.e., most of the constraints will still be enforced. This leads to the optimization problem of

$$\begin{aligned} \min_{G,\xi} \quad & \mathrm{Trace}(G) + C \sum_{(i,j,k,\ell) \in \mathcal{S}} \xi_{ijk\ell} \\ \text{subject to} \quad & g_{kk} - 2g_{k\ell} + g_{\ell\ell} - g_{ii} + 2g_{ij} - g_{jj} \geq 1 - \xi_{ijk\ell} \\ & \sum_{i,j} g_{ij} = 0, \quad G \succeq 0, \quad \xi_{ijk\ell} \geq 0, \end{aligned}$$

where $C > 0$ controls how strictly the constraints are enforced. The resulting optimization problem is a semidefinite program and, at least for relatively small problems, can be solved using a software package like CVX [25, 26].

III. Nonmetric triangulation

In comparison to nonmetric MDS, the problem of nonmetric triangulation is somewhat simpler. We assume that we already have some "ground truth" in an embedding X of the points $x_1, x_2, \ldots, x_n$ and are given a new point x^* and wish to estimate its location from the comparison data collected in the form of the set $\mathcal{T}$. We first consider the problem of finding a single candidate x^* that agrees with the given constraints. To begin, note that by squaring and expanding both sides of (2), we can re-express the inequality as

$$\|x^*\|_2^2 - 2\langle x^*, x_i\rangle + \|x_i\|_2^2 < \|x^*\|_2^2 - 2\langle x^*, x_j\rangle + \|x_j\|_2^2,$$

which we can rearrange to obtain

$$\langle x^*, x_i - x_j\rangle > \frac{\|x_i\|_2^2 - \|x_j\|_2^2}{2}.$$

This leads to a total of $|\mathcal{T}|$ inequalities. If the inequalities are self-consistent, then we would likely be satisfied to find any possible $x^* \in \mathbb{R}^k$ that agrees with these inequalities. One possible choice would be to simply pick the x^* with smallest ℓ_2-norm by solving the following optimization problem

$$\begin{aligned} \min_{x^*} \quad & \|x^*\|_2^2 \\ \text{subject to} \quad & (x_i - x_j)^T x^* > \frac{\|x_i\|_2^2 - \|x_j\|_2^2}{2}. \end{aligned}$$

Since the $x_1, x_2, \ldots, x_n$ are given and we are directly optimizing over a point $x^* \in \mathbb{R}^k$, this problem is simpler than the nonmetric MDS problem as there is no need for any rank penalty to control the dimension of x^*. Of course, as in the case of the GNMDS algorithm, strictly enforcing all possible constraints will likely over-fit the data and may not be feasible. Thus, to allow for some of these constraints to be violated, we again introduce slack variables ξ_{ij} and add an ℓ_1 penalty to the objective function to ensure that the number of constraint violations is small. This results in the optimization problem of

$$\begin{aligned} \min_{x^*, \xi} \quad & \|x^*\|_2^2 + C \sum_{(i,j)\in\mathcal{T}} \xi_{ij} \\ \text{subject to} \quad & (x_i - x_j)^T x^* \geq \frac{\|x_i\|_2^2 - \|x_j\|_2^2}{2} - \xi_{ij} \\ & \xi_{ij} \geq 0, \end{aligned}$$

where $C > 0$ again controls how strictly the constraints are enforced. This optimization problem is a quadratic program and can be easily solved using a software package like CVX [25, 26]. It is also essentially equivalent in structure to the optimization problem for support vector machines, and so many of the specialized techniques developed in that context could be adapted to provide highly optimized methods for solving this problem.

IV. Experiments

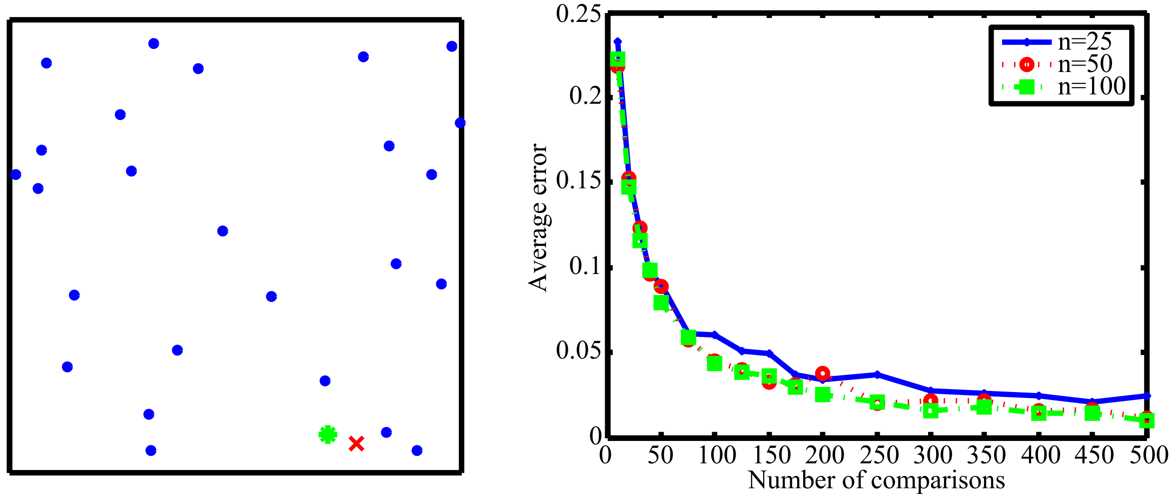
We implemented and tested our algorithm for nonmetric triangulation using CVX [25, 26]. We performed a suite of synthetic simulations to evaluate the performance of the algorithm as a function of several factors, including the impact of the number of landmark points, the number of paired comparisons, and the underlying "noise level" in the comparisons. Specifically, we began by generating n landmark points uniformly at random on the unit square. We then generate a random x^* and conduct a sequence of paired comparisons by choosing pairs (i, j) uniformly at random, and testing whether x^* is closer to x_i or x_j. To examine the impact of noise, rather than directly measuring the distance to x_i or x_j, we instead measured the distance to versions of x_i and x_j that were perturbed by two-dimensional Gaussian noise with $\sigma = 0.1$. The results of a representative example are shown in Fig. 1(a). The blue circles represent the landmark points, the green star represents the true x^*, and the red cross is the result of our algorithm.

In Fig. 1(b) we show the impact of the number of comparisons on the localization error for $n = 25, 50, 100$, setting $C = 10$ in our algorithm. We measure localization error by computing the ℓ_2 distance of the estimated x^* to the true one and averaging over 100 realizations of landmark points and x^*. Not surprisingly, we observe that as the number of comparisons increases, the error goes down, but we also observe that it initially decays very rapidly and quite quickly reaches a relatively low error using far fewer comparisons than the total number possible. The results are broadly similar for a wide range of the parameter C, suggesting that the performance is not particularly sensitive to the choice of C. Our experiments confirm that an increasingly accurate localization is possible by increasing the number of landmark points or the number of comparisons, but these factors differ in importance. In particular, the results in Fig. 1(b) suggest that having a sufficient number of comparisons is relatively more important than having a large number of landmark points. This is fortunate in our context, since it suggests that we can perform the initial localization via GNMDS on only a small sampling of points. Initial simulations combining these ideas are quite promising, but are omitted here due to a lack of space.

V. Conclusion

We have presented preliminary simulations that suggest that our proposed method for nonmetric triangulation can perform well even when confronted with a noisy and incomplete set of paired comparisons. This allows our nonmetric triangulation algorithm to be combined successfully with GNMDS to provide a framework for nonmetric MDS on relatively large datasets. While the process of triangulating a single point using our approach does involve solving a convex optimization problem, which might seem to be rather computationally demanding if we plan to apply this to hundreds or thousands of additional points, this would still be significantly less demanding (both computationally, and also in terms of memory requirements), than scaling GNMDS to handle the full dataset. It would also be interesting to exploit potential connections between the nonmetric MDS and triangulation problems and the recently developed theory of 1-bit matrix completion [27]. By exploiting this connection it might be possible to develop theoretical bounds for these or similar algorithms.

Finally, it would be very interesting to explore connections between the proposed triangulation algorithm and the approaches to active nonmetric MDS and active triangulation in [28, 29]. In this paper we have assumed that the sets $\mathcal{S}$ and $\mathcal{T}$ of paired comparisons are simply given to us and are fixed in advance. In [28, 29] it is demonstrated that by actively and adaptively selecting which comparisons to obtain/use, dramatic improvements are often possible. While these papers primarily deal with a noise-free framework, it should be possible to



(a) *(b)*

Fig. 1. *(a) Example with* $n = 25$ *and* 50 *paired comparisons. (b) Impact of the number of comparisons on localization error for* $n = 25, 50, 100$.

combine the two ideas, leading to potentially more robust versions of the approaches in [28, 29], and potentially allowing for dramatic reductions in the computational complexity of our nonmetric triangulation approach by cleverly removing unnecessary constraints in our optimization problem.

Acknowledgment

The author would like to thank James Roberts and Mark Rusinski for many interesting and enlightening discussions.

References

[1] V. de Silva and T. Tenenbaum, "Sparse multidimensional scaling using landmark points," Stanford University, Tech. Rep., 2004.

[2] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *ACM Trans. Sen. Netw.*, vol. 2, no. 2, pp. 188–220, 2006.

[3] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, no. 2, pp. 215–245, 1995.

[4] A. So and Y. Ye, "Theory of semidefinite programming for sensor network localization," *Math. Programming, Series A and B*, vol. 109, no. 2, pp. 367–384, 2007.

[5] A. Singer, "A remark on global positioning from local distances," *Proc. Natl. Acad. Sci.*, vol. 105, no. 28, pp. 9507–9511, 2008.

[6] A. Singer and M. Cucuringu, "Uniqueness of low-rank matrix completion by rigidity theory," *SIAM J. Matrix Anal. and Appl.*, vol. 31, no. 4, pp. 1621–1641, 2010.

[7] I. Borg and P. Groenen, *Modern multidimensional scaling*. New York, NY: Springer Science, 2010.

[8] T. Cox and M. Cox, *Multidimensional Scaling*. Boca Raton, FL: Chapman & Hall / CRC, 1994.

[9] G. Frisus, "Libellus de locorum describendum ratione," 1533.

[10] D. Goldberg, D. Nichols, B. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Comm. ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[11] N. Srebro, "Learning with matrix factorizations," Ph.D. dissertation, MIT, 2004.

[12] J. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proc. Int. Conf. Machine Learning*, Bonn, Germany, Aug. 2005.

[13] C. Coombs, "Psychological scaling without a unit of measurement," *Psychological Rev.*, vol. 57, pp. 145–158, 1950.

[14] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[15] E. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[16] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[17] B. Dubois, "Ideal point versus attribute models of brand preference: A comparison of predictive validity," *Adv. Consumer Research*, vol. 2, pp. 321–334, 1975.

[18] G. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Rev.*, vol. 63, no. 2, pp. 81–97, 1956.

[19] H. David, *The method of paired comparisons*. London, UK: Charles Griffin & Company Limited, 1963.

[20] S. Agarwal, J. Wills, L. Cayton, G. Lanckriet, D. Kriegman, and S. Belongie, "Generalized non-metric multidimensional scaling," in *Proc. Int. Conf. Art. Intell. Stat. (AISTATS)*, San Juan, Puerto Rico, Mar. 2007.

[21] R. Shepard, "The analysis of proximities: Multidimensional scaling with an unknown distance function. I," *Psychometrika*, vol. 27, pp. 125–140, 1962.

[22] ——, "The analysis of proximities: Multidimensional scaling with an unknown distance function. II," *Psychometrika*, vol. 27, pp. 219–246, 1962.

[23] J. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, pp. 1–27, 1964.

[24] ——, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, pp. 115–129, 1964.

[25] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sep. 2012.

[26] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[27] M. Davenport, Y. Plan, E. van den Berg, and M. Wootters, "1-bit matrix completion," *Arxiv preprint arxiv:1209.3672*, 2012.

[28] K. Jamieson and R. Nowak, "Low-dimensional embedding using adaptively selected ordinal data," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2011.

[29] ——, "Active ranking using pairwise comparisons," *Arxiv preprint arxiv:1109.3701*, 2011.