Multidimensional Scaling with City-Block Distances

Julius Žilinskas and Nerijus Galiauskas

Vilnius University Institute of Mathematics and Informatics,
Akademijos 4, LT-08663 Vilnius, Lithuania
julius.zilinskas@mii.vu.lt
http://www.mii.vu.lt

Abstract. Multidimensional scaling is a technique for exploratory analysis of multidimensional data. The essential part of the technique is minimization of a function with unfavorable properties like multimodality, invariants and non-differentiability. Recently various two-level optimization algorithms for multidimensional scaling with city-block distances have been proposed exploiting piecewise quadratic structure of the least squares objective function with such distances. A problem of combinatorial optimization is tackled at the upper level, and convex quadratic programming problems are tackled at the lower level. In this paper we discuss a new reformulation of the problem where lower level quadratic programming problems seem more suited for two-level optimization.

Keywords: Multidimensional scaling, City-block distances, Multilevel optimization, Global optimization

1 Introduction

Multidimensional scaling (MDS) is a technique for exploratory analysis of multidimensional data widely usable in different applications [1–3]. A set of points in an m-dimensional embedding space is considered as an image of the set of n objects. Coordinates of points $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, should be found whose inter-point distances fit the given pairwise dissimilarities δ_{ij}, $i, j = 1, \ldots, n$. The points can be found minimizing a fit criterion, e.g. the most frequently used least squares Stress function:

$$S(\mathbf{x}) = \sum_{i<j}^{n} \left(d_r\left(\mathbf{x}_i, \mathbf{x}_j\right) - \delta_{ij}\right)^2, \; d_r(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^{m} |x_{ik} - x_{jk}|^r\right)^{1/r},$$

where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$. $d_r(\mathbf{x}_i, \mathbf{x}_j)$ denotes the Minkowski distance between the points $\mathbf{x}_i$ and $\mathbf{x}_j$. The most frequently used distances are Euclidean ($r = 2$), but multidimensional scaling with other Minkowski distances in the embedding space can be even more informative [4]. Here we consider the city-block distances ($r = 1$).

Stress function normally has many local minima. It is invariant with respect to translation and rotation or mirroring. Positiveness of distances at a local minimum point imply differentiability of Stress function [5, 6] with the Minkowski distances except city-block: Stress with the city-block distances can be non-differentiable even at a minimum point [7]. However Stress with the city-block distances is piecewise quadratic, and such a structure can be exploited for tailoring of ad hoc global optimization algorithms. We refer to [3] for review on optimization algorithms for city-block MDS.

2 Two-Level Optimization for Multidimensional Scaling with City-Block Distances

Let us start by describing a reformulation of the Stress function similar to one presented in [3]. Stress with city-block distances $d_1(\mathbf{x}_i, \mathbf{x}_j)$ can be redefined as

$$S(\mathbf{x}) = \sum_{i<j}^{n} \left(\sum_{k=1}^{m} |x_{ik} - x_{jk}| - \delta_{ij} \right)^2 .$$

Let $A(\mathbf{P})$ denotes a set such that

$$A(\mathbf{P}) = \{\mathbf{x} |\ \ x_{ik} \le x_{jk} \text{ for } p_{ki} < p_{kj},\ i, j = 1, \ldots, n,\ k = 1, \ldots, m\} ,$$

where $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_m)$, $\mathbf{p}_k = (p_{k1}, p_{k2}, \ldots, p_{kn})$ is a permutation of $1, \ldots, n$; $k = 1, \ldots, m$. For $\mathbf{x} \in A(\mathbf{P})$, Stress function can be defined as

$$S(\mathbf{x}) = \sum_{i<j}^{n} \left(\sum_{k=1}^{m} (x_{ik} - x_{jk})\, z_{kij} - \delta_{ij} \right)^2 ,\ z_{kij} = \begin{cases} 1, \, p_{ki} > p_{kj}, \\ -1, \, p_{ki} < p_{kj}. \end{cases} .$$

Since function $S(\mathbf{x})$ is quadratic over polyhedron $\mathbf{x} \in A(\mathbf{P})$ the problem

$$\min_{\mathbf{x} \in A(\mathbf{P})} S(\mathbf{x})$$

is a quadratic programming problem.

Taking into account the structure of the minimization problem a two-level minimization algorithm can be applied [7]: to solve a combinatorial problem at the upper level and to solve a quadratic programming problem at the lower level:

$$\min_{\mathbf{P}} S(\mathbf{P}),$$

$$\text{s.t. } S(\mathbf{P}) = \min_{\mathbf{x} \in A(\mathbf{P})} S(\mathbf{x}) \sim$$

$$\sim \min \left(-\mathbf{c_P}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{Q_P} \mathbf{x} \right),\ \text{s.t. } \mathbf{E}\mathbf{x} = \mathbf{0},\ \mathbf{A_P}\mathbf{x} \ge \mathbf{0}. \tag{1}$$

For the lower level problem a standard quadratic programming method can be applied. The upper level combinatorial problem can be solved using different algorithms. Small problems can be solved by the explicit enumeration. A branch-and-bound algorithm for the upper level combinatorial problem is proposed in [8] and its parallel version in [9]. Evolutionary algorithms seem perspective for larger problems. A two-level minimization method for the two-dimensional projection space was proposed in [7], where the upper level combinatorial problem is tackled by an evolutionary search. A generalized method for an arbitrary dimensionality of the projection space is developed and experimentally compared with other approaches in [10]. A multimodal evolutionary algorithm is proposed in [11].

Indices $\cdot_{\mathbf{P}}$ in the description of the lower level problem (1) indicate that the coefficients of the quadratic function and inequality constraints depend on the permutations in $\mathbf{P}$. This means that matrices $\mathbf{Q_P}$, $\mathbf{A_P}$, and vector $\mathbf{c_P}$ need to be computed for every lower level problem. If a quadratic programming algorithm factorizes matrix $\mathbf{Q_P}$ and computes the inverse, this should be done each time the lower level problem is solved.

3 Reformulation of Optimization Problem with City-Block Distances

In this paper we present a different formulation where we try to avoid or decrease required computations of coefficients and factorizations/inversions as well as to enable warm-starting for lower level problems. Let us introduce non-negative variables d^+_{ijk} and d^-_{ijk} so that

$$x_{ik} - x_{jk} = d^+_{ijk} - d^-_{ijk},\ i < j \le n,\ k = 1, \ldots, m. \tag{2}$$

If $d^+_{ijk} = 0$ or $d^-_{ijk} = 0$ ($d^+_{ijk} d^-_{ijk} = 0$), $|x_{ik} - x_{jk}| = d^+_{ijk} + d^-_{ijk}$, and

$$S(\mathbf{d}) = \sum_{i<j}^{n} \left(\sum_{k=1}^{m} \left(d^+_{ijk} + d^-_{ijk} \right) - \delta_{ij} \right)^2,$$

where $\mathbf{d} = (d^+_{121}, d^-_{121}, \ldots, d^+_{(n-1)\,n\,m}, d^-_{(n-1)\,n\,m})$.

Now we can formulate the lower level problem as

$$\min \left(-\mathbf{c}^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{Q}\mathbf{d} \right),\ \text{ s.t. } \mathbf{E}\mathbf{d} = \mathbf{0},\ \mathbf{d} \ge \mathbf{0},\ \mathbf{a_P}^T\mathbf{d} = 0. \tag{3}$$

The first system of equality constraints of the lower level problem (3) is the version of the system (2) after reduction avoiding variables $\mathbf{x}$. The system of inequality constraints define non-negativity of variables $\mathbf{d}$. The last equality constraint defines particular lower level problem:

$$a^+_{ijk}, a^-_{ijk} = \begin{cases} 1, 0,\ p_{ki} < p_{kj}, \\ 0, 1,\ p_{ki} > p_{kj}, \end{cases}$$

where $\mathbf{a_P} = (a^+_{121}, a^-_{121}, \ldots, a^+_{(n-1)\,n\,m}, a^-_{(n-1)\,n\,m})$. Since either $a^+_{ijk} = 1$ or $a^-_{ijk} = 1$, the constraint $\mathbf{a_P}^T\mathbf{d} = 0$ ensures that either $d^+_{ijk} = 0$ or $d^-_{ijk} = 0$, therefore $d^+_{ijk} d^-_{ijk} = 0,\ i < j \le n,\ k = 1, \ldots, m$.

We see that in this formulation only $\mathbf{a_P}$ depends on the permutations in $\mathbf{P}$. Therefore, the matrices $\mathbf{Q}$, $\mathbf{E}$ and the vector $\mathbf{c}$ may be computed once in advance. Factorization of matrix $\mathbf{Q}$ and inverse may be also performed once in advance. Moreover new bounds may be built for a branch-and-bound algorithm based on this formulation if relaxing the constraint $d^+_{ijk}d^-_{ijk} = 0$, e.g. when $\mathbf{a_P}$ is chosen with a smaller number of ones. However the number of variables of quadratic programming problems has increased from mn to $mn(n-1)$, but it should be noted that the matrices $\mathbf{Q}$ and $\mathbf{E}$ are sparse while in the formulation described in the previous section the matrix of coefficients is dense.

4 Experimental Investigation

We performed computational experiments with two-level minimization algorithms for multidimensional scaling with the new formulation of lower level quadratic problems. The upper level combinatorial problem is solved using explicit enumeration and branch-and-bound.

The results of experiments are shown in Table 1. The numbers of quadratic programming problems solved (NQPP) and the estimate of the global minimum found (f^*) are shown for the two-level algorithms with explicit enumeration (EE) and branch-and-bound (B&B) at the upper level. The same data sets as in [8] were used. When explicit enumeration is used for the upper level problem, the results correspond to that presented in [8]: the numbers of lower level quadratic programming problems solved and the found minima are the same. Conclusions similar to that presented in [8] can be drawn. The branch-and-bound algorithm behaves in the worst case scenario when highly symmetric data sets of simplices [12] are used with $m = 1$. Branch-and-bound performs much better than the explicit enumeration for cubes and practical data sets even when $m = 1$ and even for simplices when $m > 1$. Comparing the results of branch-and-bound with the new formulation to that of [8] one can see that the numbers of quadratic programming problems solved are a bit smaller.

5 Conclusions

A new formulation of optimization problems for multidimensional scaling with city-block distances is proposed. Two-level algorithms have been built with explicit enumeration or branch-and-bound at the upper level and convex quadratic programming at the lower level. Experiments with geometrical and empirical data sets have been performed. The experimental investigation revealed that the numbers of quadratic programming problems solved are a bit smaller for the new formulation.

Acknowledgments. This research was funded by a grant (No. MIP-063/2012) from the Research Council of Lithuania.

Table 1. Results of experimental investigation

		EE		B&B		B&B [8]	
n	m	NQPP	f^*	NQPP	f^*	NQPP	f^*
				unit simplices			
3	1	3	0.00	3	0.00	3	0.00
4	1	12	0.3651	12	0.3651	14	0.3651
5	1	60	0.4140	71	0.4140	73	0.4140
6	1	360	0.4554	430	0.4554	432	0.4554
7	1	2520	0.4745	2949	0.4745	2951	0.4745
8	1	20160	0.4917	23108	0.4917	23110	0.4917
9	1	181440	0.5018	204538	0.5018	204549	0.5018
3	2	6	0.00	6	0.00	6	0.00
4	2	78	0.00	78	0.00	73	0.00
5	2	1830	0.00	942	0.00	662	0.00
6	2	64980	0.1869	15963	0.1869	16076	0.1869
				standard simplices			
3	1	3	0.3333	3	0.3333	3	0.3333
4	1	12	0.4082	12	0.4082	14	0.4082
5	1	60	0.4472	71	0.4472	73	0.4472
6	1	360	0.4714	430	0.4714	432	0.4714
7	1	2520	0.4879	2949	0.4879	2951	0.4879
8	1	20160	0.5000	23108	0.5000	23110	0.5000
9	1	181440	0.5092	204547	0.5092	204549	0.5092
3	2	6	0.00	6	0.00	6	0.00
4	2	78	0.00	78	0.00	63	0.00
5	2	1830	0.1907	1317	0.1907	1322	0.1907
6	2	64980	0.2309	27322	0.2309	27255	0.2309
				cubes			
4	1	12	0.4082	12	0.4082	14	0.4082
8	1	20160	0.4787	11114	0.4787	11260	0.4787
4	2	78	0.00	78	0.00	73	0.00
				ruusk			
8	1	20160	0.2975	643	0.2975	665	0.2975
8	2			81139	0.1096	82617	0.1096
				hwa12			
9	1	181440	0.0107	2167	0.0107	2217	0.0107
				cola			
10	1			59599	0.3642	60077	0.3642
				uhlen			
12	1			36251	0.2112	36559	0.2112
				hwa21			
12	1			70583	0.1790	71748	0.1790

References

1. Borg, I., Groenen, P.J.F.: Modern Multidimensional Scaling: Theory and Applications. 2nd edn. Springer, New York (2005)
2. Cox, T.F., Cox, M.A.A.: Multidimensional Scaling. 2nd edn. Chapman & Hall/CRC, Boca Raton (2001)
3. Dzemyda, G., Kurasova, O., Žilinskas, J.: Multidimensional Data Visualization: Methods and Applications. Springer, New York (2013) doi:10.1007/978-1-4419-0236-8.
4. Arabie, P.: Was Euclid an unnecessarily sophisticated psychologist? Psychometrika **56**(4) (1991) 567–587 doi:10.1007/BF02294491.
5. de Leeuw, J.: Differentiability of Kruskal's stress at a local minimum. Psychometrika **49**(1) (1984) 111–113 doi:10.1007/BF02294209.
6. Groenen, P.J.F., Mathar, R., Heiser, W.J.: The majorization approach to multidimensional scaling for Minkowski distances. Journal of Classification **12**(1) (1995) 3–19 doi:10.1007/BF01202265.
7. Žilinskas, A., Žilinskas, J.: Two level minimization in multidimensional scaling. Journal of Global Optimization **38**(4) (2007) 581–596 doi:10.1007/s10898-006-9097-x.
8. Žilinskas, A., Žilinskas, J.: Branch and bound algorithm for multidimensional scaling with city-block metric. Journal of Global Optimization **43**(2-3) (2009) 357–372 doi:10.1007/s10898-008-9306-x.
9. Žilinskas, J.: Parallel branch and bound for multidimensional scaling with city-block distances. Journal of Global Optimization **54**(2) (2012) 261–274 doi:10.1007/s10898-008-9624-7.
10. Žilinskas, A., Žilinskas, J.: A hybrid method for multidimensional scaling using city-block distances. Mathematical Methods of Operations Research **68**(3) (2008) 429–443 doi:10.1007/s00186-008-0238-5.
11. Redondo, J., Ortigosa, P., Žilinskas, J.: Multimodal evolutionary algorithm for multidimensional scaling with city-block distances. Informatica **23**(4) (2012) 601–620
12. Žilinskas, J.: Reducing of search space of multidimensional scaling problems with data exposing symmetries. Information Technology and Control **36**(4) (2007) 377–382