Manifold based optimization techniques for low-rank distance matrix completion

Bamdev Mishra and Rodolphe Sepulchre
Department of Electrical Engineering and Computer Science
Institut Montefiore
University of Liège
B-4000 Liège
Belgium
Emails: b.mishra@ulg.ac.be and r.sepulchre@ulg.ac.be

1 Introduction

We discuss the problem of low-rank Euclidean distance matrix completion (EDMCP) and its various formulations. We emphasize on the geometrical structure of the problem and develop on this idea to design manifold based algorithms [AMS08]. The rank-constraint makes the problem non-convex and difficult to analyze. Concerning the recovery of the matrix, the sparsity (number of missing entries) also plays an important role. We assume here that the sparsity in the dissimilarity matrix has a standard normal distribution. This removes some pathological cases, for which an exact recovery can never be possible.

2 The Problem

The book [dat05] describes the problem in detail. It also discusses different cost functions that are normally used to solve the distance matrix completion problem. We, however, focus on the following non-convex cost function.

$$\begin{array}{ll} minimize_D & ||H \circ (A-D)||_F^2 \\ such\ that & D \in \mathbb{EDM}^N \\ & rank(D) = r \ (r << N). \end{array} \tag{1}$$

Here A is an incomplete *dissimilarity* matrix and H is a *weight* matrix to take into account the corresponding missing entries in A with the Hadamard product $H \circ A$. D is an Euclidean distance matrix (EDM) [1]. An exact reformulation of the problem (1) on the set of symmetric positive semidefinite matrices (PSD) is the following.

$$\begin{array}{ll} minimize_X & ||H \circ (A-\kappa(X))||_F^2 \\ such\ that & X = X^T \\ & X \in \mathbb{PSD}^N \\ & rank(X) = p \ (p << N). \end{array} \tag{2}$$

$\kappa(X)$ is a linear function relating D and X and so, r and p are related ($r \leq p+2$). Geometrically speaking, X is the Gram matrix of the coordinates of the points under consideration. Problem (2) makes the EDMCP an optimization problem on the low-rank PSD cone.

[1] $D_{ij}=$ represents the Euclidean distance *square*.

3 Optimization points of view 1: EDM cone

The book [BG97] mentions classical multidimensional scaling (CMDS) to solve the problem (1) for the case of no missing entries. We extend this idea into a simple algorithm to adapt to missing entries using soft-thresholding of singular values to finally arrive at a low-rank solution.

4 Optimization point of view 2: PSD cone

The thesis [Jou09] solves a non-convex optimization problem on the low-rank PSD cone. The problem (2) can be looked into from this point view. We solve the problem on the PSD cone using two techniques. One, a trust-region algorithm on the quotient manifold. Other, an iterative hard thresholding of eigenvalues on the PSD cone.

References

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, Princeton, NJ, 2008.

[dat05] J. Dattorro, *Convex optimization & euclidean distance geometry*, Meboo Publishing USA, PO Box 12, Palo Alto, CA 94302, 2005.

[BG97] I. Borg and P. J. Groenen, *Modern multidimensional scaling : theory and applications*, Springer series in statistics, Springer, 1997.

[Jou09] M. Journée, *Geometric algorithms for component analysis with a view to gene expression data analysis*, Ph.D. thesis, University of Liège, Liège, Belgium, 2009.

Acknowledgments

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author(s).