Wissenschaftliches Rechnen II/Scientific Computing II

Sommersemester 2016
Prof. Dr. Jochen Garcke
Dipl.-Math. Sebastian Mayer

Exercise sheet 12

To be handed in on **Thursday, 14.07.2016**

Isomap

1 Group exercises

G 1. Consider the manifold $M = \{x \in \mathbb{R}^d : \|x\|_2 = r\}$, i.e., the Euclidean sphere with radius $r > 0$ in $\mathbb{R}^d$. Compute the minimum radius of curvature $r_0(M)$ and the minimum branch separation $s_0(M)$. Moreover, show that Lemma 2.16 holds true with equality, i.e., for all $x, y \in M$ such that $d_M(x, y) < \pi r_0(M)$ we have

$$d_E(x, y) = 2r_0(M) \sin\left(\frac{d_M(x, y)}{2r_0(M)}\right).$$

Solution. Consider the curve $\gamma(t) = r(\cos(t/r), \sin(t/r))$, which is the optimal parametrization. Then $\|\dot{\gamma}\|_2 = 1$ and $\|\ddot{\gamma}\| = 1/r$. Moreover, it is easy to see that $s_0(M) = 2r_0(M)$. Finally,

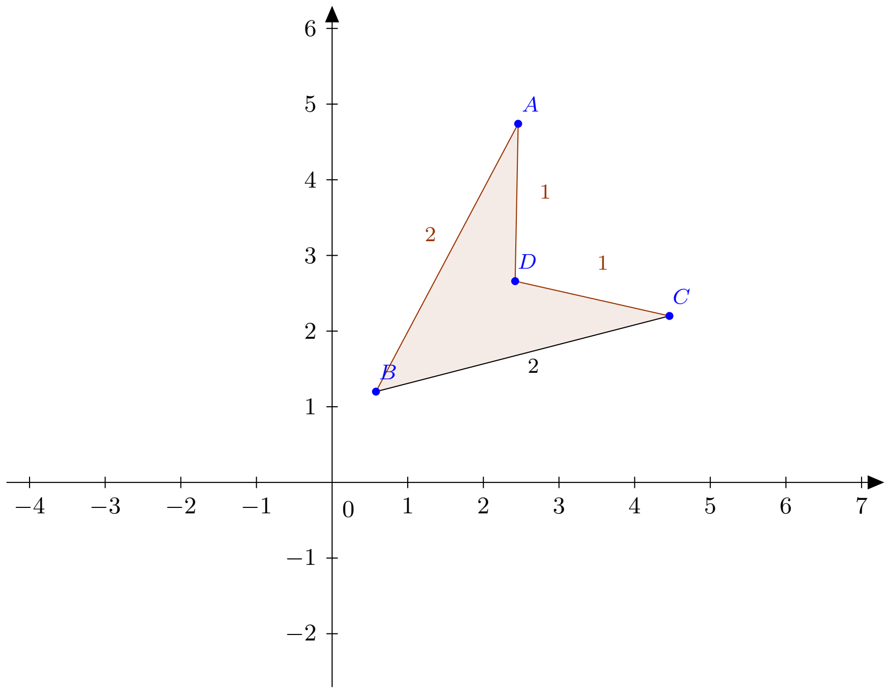
$$\|x - y\|_2^2 = \|\gamma(t) - \gamma(0)\|_2^2 = 2r^2(1 - \cos(t/r)) = 4r^2 \sin^2(t/2r).$$

For the last equality, we have used $\cos(2t/(2r)) = \cos^2(t/(2r)) - \sin^2(t/(2r))$. □

G 2. Construct a graph distance matrix which is not a Euclidean distance matrix.

Solution. The solution is simple. We just choose 4 points in the plane to form a *kite* (dt. *Drachen*) or *dart* (dt. *Pfeil*), see the diagram. The sides become the edges of our graph. Whatever lengths you assume for the sides, the resulting graph distance matrix will never be a Euclidean distance matrix, since

$$\begin{aligned} \overline{BD} &< \overline{AB} + \overline{AD}, \\ \overline{AC} &< \overline{AD} + \overline{CD}. \end{aligned}$$



□

G 3. Let $M \subset \mathbb{R}^d$ be manifold. Assume you have run Isomap with input $x_1, \ldots, x_n \in M$ to obtain p-dimensional embeddings. Write down ready-to-use formulas that compute you the p-dimensional embedding for a new, unseen data point $x \in M$ without rerunning the whole Isomap algorithm.

Solution. In the end, we want to use the same kind of embedding as for CMDS, see Sheet 1, G1. For isomap, this requires some further computation steps since we first have to estimate the geodesic distance between the new point $x \in M$ and the training points $x_1, \ldots, x_n \in M$. A possible procedure to embed the new point $x \in M$ is as follows.

1. **Extending the neighbourhood graph:** Using the same criterion as for the training data (ε-neighbourhood, k-nearest neighbours, adaptive neighbourhood), we add edges (x, x_i) to the edge set E of the neighbourhood graph.

2. **Computing the graph distance vector:** Using Dijkstra's algorithm with source node x, we compute the graph distance vector

$$x_G := \begin{pmatrix} d_G^2(x, x_1) \\ \vdots \\ d_G^2(x, x_n) \end{pmatrix}.$$

3. **Transformation into inner products:** Compute

$$x_S := -\frac{1}{2}\left(x_G - \frac{1}{n} D_G 1_n + \frac{1}{n} x_G^T 1_n 1_n^T + \frac{1}{n^2} 1_n^T D_G 1_n 1_n^T\right)$$

4. **Projection onto p prinicpal components:** Compute

$$\hat{x} := I_{p \times n} \Lambda^{-1/2} V^T x_S.$$

□

2 Homework

H 1. Let M be a compact manifold. Prove the following simplyfied version of Lemma 2.16. For any $\varepsilon > 0$, we have

$$(1-\varepsilon)d_M(x,y) \leq d_E(x,y) \leq d_M(x,y)$$

for all $x, y \in M$ such that $d_M(x,y) < 2r_0(M)\varepsilon$. **Hint:** Consider a unit speed parametrization $\gamma : (0,l) \to M$ with $\gamma(0) = x$, $\gamma(l) = y$ and use the fundamental theorem of calculus to obtain a first order estimate.

(10 Punkte)

H 2. (Isomap and non-EDM graph distance matrices)

a) Let S be a symmetric $n \times n$ matrix with eigenvalue decomposition $U\Lambda U^T$. Let Λ' be Λ with all negative eigenvalues replaced by zero and put $S^+ := U\Lambda' U^T$. Show that S^+ is the solution of

$$\min_{B \in \mathcal{S}_n^+} \|S - B\|_F^2,$$

where $\mathcal{S}_n^+$ is the set of all positive semi-definite $n \times n$ matrices.

b) Based on your insights from a), argue why Isomap can also be used when the graph distance matrix D_G is not a Euclidean distance matrix. How should the algorithm be modified in this case?

(10 Punkte)