# Mahalanobis-based Adaptive Nonlinear Dimension Reduction

Djamila Aouada
*SnT Centre, University of Luxembourg*
*Luxembourg*
*djamila.aouada@uni.lu*

Yuliy Baryshnikov
*Alcatel-Lucent, Bell Labs.*
*Murray Hill, NJ*
*ymb@research.bell-labs.com*

Hamid Krim
*ECE Dept., NCSU*
*Raleigh, NC*
*ahk@ncsu.edu*

***Abstract*—We define a new adaptive embedding approach for data dimension reduction applications. Our technique entails a local learning of the manifold of the initial data, with the objective of defining local distance metrics that take into account the different correlations between the data points. We choose to illustrate the properties of our work on the isomap algorithm. We show through multiple simulations that the new adaptive version of isomap is more robust to noise than the original non-adaptive one.**

## I. Introduction

In recent years, data sizes have drastically increased. As a result, there has been a great research focus on improving and defining effective dimension reduction techniques. These efforts are extremely relevant if not crucial to data storage, visualisation, and analysis applications. The objective behind learning and reducing the dimension of data is to eliminate any redundant information while still preserving the intrinsic and underlying structure. One may think of this problem as an attempt to find all the variables that may be combined into fewer variables without destroying the interactions between the data points. To formulate this, we proceed as follows. Given a data point sample of $N$ points $X_i$, $i = 1, \cdots, N$, from a $d$-dimensional smooth manifold $\mathcal{M}$, where $\mathcal{M} \subset \mathbb{R}^n$ and $d < n$, in a dimension reduction problem, we aim at finding $\{Y_i\}$, the image of $\{X_i\}$ by a homeomorphism $f(\cdot)$ such that:

$$\begin{aligned} f: \mathcal{M} \subset \mathbb{R}^n &\rightarrow \mathbb{R}^d \\ X_i &\mapsto f(X_i) = Y_i, \text{ for } i = 1, \cdots, N. \end{aligned} \quad (1)$$

We may distinguish two classes of dimension reduction approaches. The first class includes all the classical methods, or linear methods, such as Principal Component Analysis (PCA) [1], and Multidimensional Scaling (MDS) [2]. In contrast, the second class corresponds to non-linear techniques [3], also referred to as manifold learning methods. There are about four widely known manifold learning techniques; Locally Linear Embedding (LLE) [4], Laplacian eigenmap [5], Hessian eigenmap [6], and isomap [7]. Manifold learning algorithms always assume the observed cloud of data points as part of a smooth manifold. Thus, to proceed with the analysis of this data, we start by constructing a graph connecting all the data points and preserving the structure of the manifold. One usually defines an $\epsilon$-ball neighbourhood of fixed radius around each data point to carry out the analysis. All these techniques have shown very successful results in ideal conditions; nevertheless, there is very limited work in addressing the effect of noise and the choice of the neighbourhood size. Very simple experiments may show how crucial it is to take these considerations into account. Our goal in this paper is to address the noise problem, and propose a way to develop a new manifold learning technique, with a built-in robustness to noise. The key idea is to replace the arbitrary choice of an approximate Euclidean distance, and to instead use a locally adaptive distance. To achieve that, we propose to account for sample data points' correlations in defining their neighbourhood.

The remainder of the paper is organised as follows: In Section II, we discuss the classical isomap algorithm which, in contrast to our proposed technique, is non-adaptive. In Section III, we describe our proposed adaptive method. We evaluate the benefit of an adaptive isomap in Section IV, using the residual covariance as a performance measure.

## II. Non-linear manifold learning

Many of the existing manifold learning techniques show successful results on some well chosen data; and they all share a limiting failure when in presence of more challenging data sets. The difficulties are often due to the intrinsic topological and geometric structure of the manifold with quick variations in their curvature and non-convex boundaries [8]. Additional difficulties result from the properties of the real data such as the sample distribution and the nature and level of the prevailing noise. To address existing limitations and to further improve the embedding results, and extend the applicability of the current manifold learning techniques, we propose to account for these overlooked characteristics. The idea is to progressively adapt to the data at hand in tracing the local connectivity between the point samples. Some recent efforts have explored adaptive manifold learning by specifically focusing on two parameters: the intrinsic dimensionality of the data, and the size of the neighbourhood. Wang *et al.* propose in [8] a method to adaptively select the neighbourhood size. They base their technique on determining the alignment space of local tangents. Costa *et al.* on the other hand define an intrinsic dimensionality using

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.187



IEEE computer society

Table I
NON-ADAPTIVE ISOMAP ALGORITHM.

**Step 1:** Construct a weighted graph $G$.

Given a sample of $N$ data points $X_i$, $i = 1, \cdots, N$;
Initialise the graph $G = \{\mathcal{A}, \mathbf{X}\}$, such that:
- $\mathcal{A}$ is the initial $(N \times N)$ adjacency matrix whose elements are set to $\infty$;
- $\mathbf{X} = [X_1, \cdots, X_N]^T$;

Compute $\mathbf{D}_E$, the matrix of Euclidean distances between each two points in $\{X_i\}_{i=1}^N$.
Choose $\epsilon$, the neighbourhood radius.

**for** $i, j \in \{1, \cdots, N\}$ **do**
  **if** $\mathbf{D}_E(i,j) < \epsilon$ **do**
    $\mathcal{A}(i,j) = \mathbf{D}_E(i,j)$;
  **else**
    $\mathcal{A}(i,j) = \infty$;
  **end if**
**end for**

**Step 2:** Compute geodesic distances on $G$.

Let $\mathbf{D}_G$ be the matrix of geodesic distances between each two points in $\{X_i\}_{i=1}^N$.
**do** $\mathbf{D}_G = \mathcal{A}$; (initialisation)
**for** $i, j \in \{1, \cdots, N\}$; $k = 1$; **do**
  **while** $\mathbf{D}_G(i,j) \neq \mathbf{D}_G(i,k) + \mathbf{D}_G(k,j)$ **do**
    **for** $k \in \{1, \cdots, N\}$ **do**
      $\mathbf{D}_G(i,j) = \min\Big(\mathbf{D}_G(i,j), \mathbf{D}_G(i,k) + \mathbf{D}_G(k,j)\Big)$;
    **end for**
  **end while**
**end for**

**Step 3:** Apply MDS on $\mathbf{D}_G$.

$k$-nearest neighbours graphs [9]. In [10], Levina *et al.* adopt a local estimate for the intrinsic dimension at each point. Our present work also focuses on an adaptive embedding of the manifold; we proceed differently and point out an additional characteristic that appears to be as critical as the choice of the embedding dimension or the neighbourhood size. We indeed show in what follows, that the choice of a Euclidean distance is sub-optimal in determining the local connectivity between data points, and therefore introduce a new adaptive distance locally defined for each point.

Our effort builds on existing manifold embedding techniques. We thus start by recalling the preliminary steps of a non-linear manifold learning algorithm. We choose to use the isomap algorithm to illustrate our ideas. This choice is due to the isomap success in numerous embedding problems and its well established properties [7]. The principle and motivations of this work are, nevertheless, extendable to other embedding algorithms. To achieve the dimension reduction, isomap defines a mapping that aims to preserve the geodesic distances on the initial manifold. We may describe isomap as merely an improved version of MDS embedding where the inter-point distance is a geodesic, i.e., restricted to lie on the initial manifold of the data. We detail the different steps of the isomap embedding algorithm in Table I. To practically approximate the intrinsic geodesic distance on a manifold, we need to locally connect each point to its $k$ nearest neighbours, or equivalently to the points within the $\epsilon$-neighbourhood. By so doing, we result in a graph that approximates the real manifold. In Figure 1 (a) we illustrate the approximating graph of a Swiss roll and the corresponding embedding in Figure 1 (b). We show in Figure 1 (c) how severely a connectivity graph may be affected in the presence of noise. This consequently yields an inaccurate embedding of a given manifold as shown in Figure 1 (d).

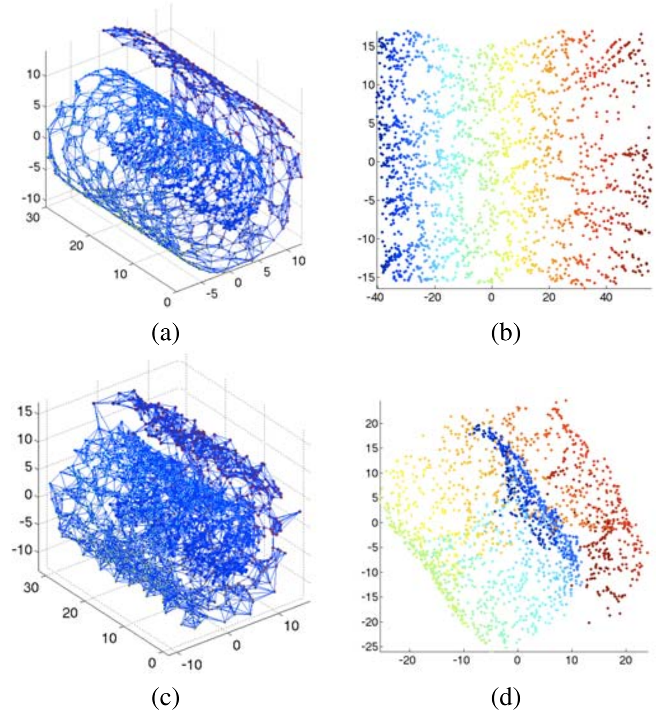

(a) (b) (c) (d)

Figure 1. Failure of isomap in a noisy setting.

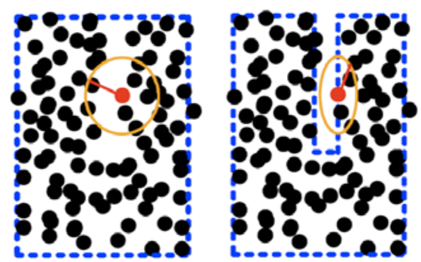


Figure 2. On the left, the structure of the set resulting of a Euclidean neighbourhood is contoured in blue. On the right, a more detailed structure of the same data set when using a Mahalanobis distance.

## III. PROPOSED ADAPTIVE ISOMAP

The graphs illustrated in Figure 1 (b) and (d) are the result of considering a Euclidean neighbourhood. In spite of the isomap good embedding results, it remains very unstable and sensitive to noise, as well as to the choice of the

Table II
DESCRIPTION OF THE LEARNING STEP (NEW ADAPTIVE DISTANCE)

**Step 0:** Compute $\mathbf{D}_M$, the new distance on $\mathcal{M}$.

Choose $\epsilon_1$, the neighbourhood radius for manifold learning; and $\epsilon_2$, the neighbourhood radius for the construction of $G$.
**for** $i \in \{1, \cdots, N\}$ **do**
  $\mathbf{Y}_i = X_i^T$; (initialization)
  **for** $j = 1, \cdots, N$ **do**
    **while** $\mathbf{D}_E(i,j) < \epsilon_1$ **do**
      $\mathbf{Y}_i = [\mathbf{Y}_i; X_j^T]$;
    **end while**
  **end for**
$\sum_i = \text{cov}(\mathbf{Y}_i^T)$,
$\text{cov}(\cdot)$ being the covariance matrix;
**end for**
**for** $i \in \{1, \cdots, N\}$ **do**
  **for** $j \in \{1, \cdots, N\}$ **do**
    $\mathbf{D}_M(i,j) = (X_j - X_i)^T \sum_i^{-1}(X_j - X_i)$;
  **end for**
**end for**
**do** $\epsilon = \epsilon_2$; $\mathbf{D}_E = \mathbf{D}_M$;
**go to** Step 1. (See Table I)

parameter $\epsilon$ and the distance function used prior to applying MDS. Changing the distance from Euclidean to geodesic thus appears to be insufficient to completely preserve the intrinsic geometric structure of the initial manifold $\mathcal{M}$. We herein maintain that the choice of the distance is crucial in constructing good connectivity graphs. Our objective is to define a more appropriate distance that alleviates the effect of noise and ensures accurate graphical approximations. In what follows we provide the intuitive rationale for the choice of a new adaptive distance. We subsequently present a mathematical formulation of new solutions to the embedding problem to result in an improved technique described in Table II. We thus propose to account for the statistical properties of the observed data. Specifically, our technique consists in considering the correlation between each point and the rest of the observed data points, and subsequently exploit this information to connect it to its neighbours. This idea is exactly equivalent to using a *Mahalanobis* distance [11]. To better understand the intuition behind our choice, we illustrate the result of constructing a graph connectivity for the sample points in Figure 2. We note that using a Euclidean distance to determine the neighbours of each data point causes a miss of some details in the structure of the data set. As a result we define a new distance matrix $\mathbf{D}_M$ to replace $\mathbf{D}_E$ in the algorithm described in Table I. Our objective is to define, each time, a distance that is fully dependent on the sample points $\{X_i\}_{i=1}^N$; hence, we re-scale the data coordinates based on their distributions on $\mathcal{M}$ as well as their correlations. Since this technique relies on a learning procedure and directly uses isomap to build on, we refer to it as an adaptive isomap algorithm. We hence use the algorithm of Table I with a learning step, i.e., Step 0, as described in Table II.

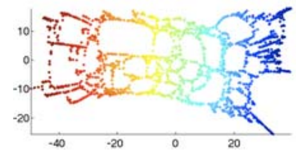
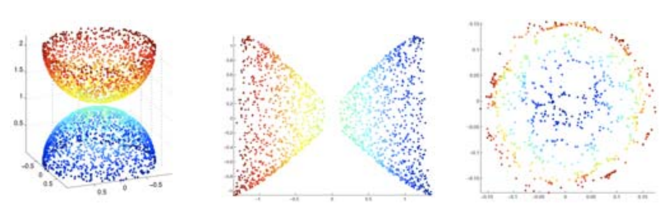



Figure 3. Adaptive embedding of the noisy swiss roll in Figure 1.

## IV. PERFORMANCE COMPARISON

We next treat a slightly more challenging case constituted of two adjacent hemispheres and two parallel sheets. Figure 4 shows the results obtained for the hemispheres. We note that the result of the non-adaptive isomap is not an embedding as multiple points are mapped to the same point. The mapping does not preserve the true structure of the initial manifold. This is due to the connectivity resulting from using a Euclidean neighborhood. Indeed, the two hemispheres end up connected through at least 2 points. We avoid this connection by using a Mahalanobis distance. Figure 4 (c) shows the final mapping resulting from using the proposed adaptive isomap. We only visualise one hemisphere at a time. This separation is the only way of having a true embedding of the data in two dimensions. In this section, we qualitatively and quantitatively compare the performances of the two versions (adaptive and non-adaptive) of isomap embeddings. To that end, we choose the residual variance $\rho$ between the distance matrix for the initial data $\{X_i\}$ and the distance matrix for the final (embedded) data $\{Y_i\}$ to be our performance indicator. We subsequently simulate different classical examples of manifolds to embed in a lower dimensional space. The choice of our examples is such that one may visually inspect and verify the properties as well as the intuition behind each technique. We saw that in the presence of noise, the performance of non-adaptive isomap drastically deteriorates and it only makes sense to evaluate performance changes when we analyse the same noisy data sets using the adaptive isomap. We consider the noisy Swiss roll example for which we determine an embedding as shown in Figure 3. In Figure 5, we experiment the embedding of two parallel sheets. For $\epsilon = 10$, we see in Figure 5 (c) that non-adaptive isomap fails again to define an embedding of the two sheets. The reason is again the



(a) (b) (c)

Figure 4. Embedding two adjacent hemispheres: (a) Hemispheres, (b) the result of a non-adaptive isomap mapping, (c) the result of the proposed adaptive isomap.

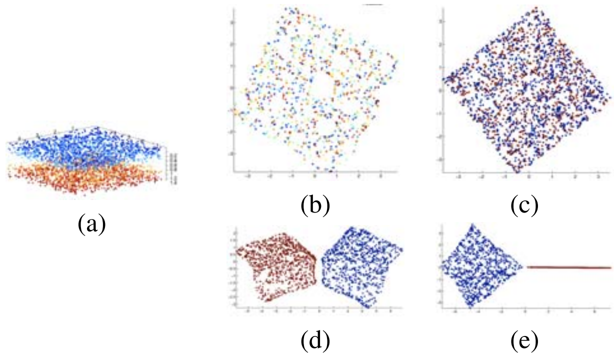



Figure 5. Embedding of two parallel sheets. (a) The initial data. (b) Adaptive isomap with $\epsilon = 10$ and $\epsilon 1 = 55$. Only 50% of the initial points are represented here. (c) Non-adaptive isomap with $\epsilon = 10$. 100% of the initial points are represented and the two sheets are overlapped. Adaptive isomap (d) and non-adaptive isomap (e) with $\epsilon = 15$. 100% of the initial points are represented.

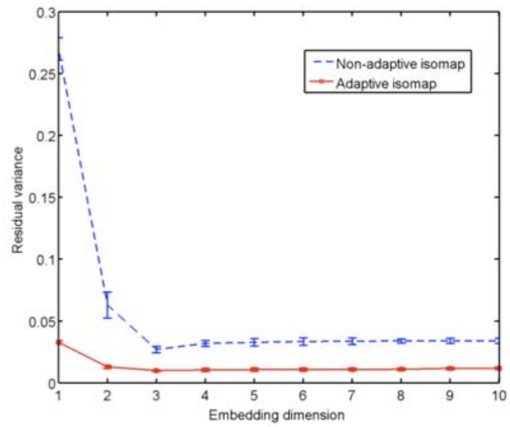



Figure 6. Monte Carlo simulation of the residual variance $\rho$ versus adaptive and non-adaptive isomap dimensionality .

connection that occurs when using a Euclidean distance for a graph construction. The result of the adaptive isomap (Figure 5 (b)) is the disconnection of the two sheets. As they should, they remain two distinct structures and are hence separately embedded. The result of embedding one sheet is shown in Figure 5 (b). Note that it is exactly the sheet itself, but now in 2-dimensions instead of 3-dimensions. For a different value of the neighborhood size $\epsilon > 10$, we find the results in (d) and (e) for the adaptive and non-adaptive isomaps, respectively. We notice the sensitivity of the mapping results to the neighborhood size; we also see that the separation between the two sheets did not happen in both cases; however, in the adaptive case, the two sheets are clearly spread on a plane, while we lose one sheet in the non-adaptive case, as all the points collapse into one line. To further evaluate the effect of noise on our proposed embedding technique, we increase the amount of Gaussian noise to vary between 0% and 8% of the orthogonal distance between the two parallel sheets, the normal distance between two consecutive levels of the swiss roll, and the orthogonal distance between the poles of the two adjacent hemispheres. By way of Monte Carlo simulations on the data in hand, we obtain the results shown in Figure 6. We establish that the adaptive isomap technique consistently outperforms the non-adaptive isomap technique.

## V. Conclusion

We proposed a new adaptive embedding algorithm that first learns from the correlations between the neighbouring points. We showed that combining an adaptive distance with an existing embedding algorithm leads to embedding results with a higher robustness to noise. We illustrated this technique on the isomap algorithm, while it is conceptually compatible with any manifold learning technique that relies on a connectivity graph.

## References

[1] I. T. Joliffe. Principal Component Analysis. Springer-Verlag, New York, 1986.

[2] W. S. Torgerson. Multidimensional Scaling, I: Theory and methods. Psychometrica, 17:401-419, 1952.

[3] J. B. Tanenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319-2323, 2000.

[4] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323-2326, 2000.

[5] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15 (6): 1373-1396, 2003.

[6] D. L. Donoho and C. E. Grimes. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. Proceedings of the National Academy of Arts and Sciences, (100):55915596, 2003.

[7] D. L. Donoho and C. E. Grimes. When does isomap recover natural parameterization of families of articulated images? Technical Report 2002-27, Stanford University 2002.

[8] J. Wang, Z. Zhang and H. Zha, Adaptive manifold learning. NIPS, 2004.

[9] J. Costa, A. Girotra and A. O. Hero, Estimating local intrinsic dimension with k-nearest neighbor graphs. In IEEE Workshop on Statistical Signal Processing (SSP), Bordeaux (2005).

[10] E. Levina and P.J. Bickel, Maximum likelihood estimation of intrinsic dimension. In K.L. Saul, Y. Weiss and I. Bottou, Adances in Neural Information Processing Systems 17. MIT Press, Cambridge, MA (2005) 777-784.

[11] D. M. De Maesschalck and R. D. Jouan-Rimbaud. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems, 50:118.