Manifold Reconstruction in Dimension Reduction Problem*

Bernstein A. V.[1,2,3]***, Burnaev E. V.***[1,3,4]***, Erofeev P. D.***[1,3,4]
a.bernstein@cpt-ran.ru, evgeny.burnaev@datadvance.net, pavel.erofeev@datadvance.net
[1]IITP, Moscow, Russia; [2]ISA, Moscow, Russia; [3]Premolab, MIPT, Dolgoprudny; [4]Datadvance, Moscow

In this paper manifold embedding and reconstruction procedures are considered in the scope of unsupervised dimension reduction problem. Standard approaches (Isomap, LLE, LTSA, etc.) are compared to newly proposed Grassman-Stiefel Eigenmaps (GSE) algorithm. It turned out that GSE provides best manifold reconstruction abilities on test problems.

Восстановление многообразий в задачах снижения размерности*

Бернштейн А.В., Бурнаев Е.В., Ерофеев П.Д.
Москва, Институт Проблем Передачи Информации РАН; Москва, DATADVANCE; Долгопрудный, PreMoLab

В данной статье рассмотрены различные подходы к построению вложений и восстановлений многообразий в задачах снижения размерности без учителя. Проведено сравнение стандартных подходов (Isomap, LLE, LTSA и др.) с новым методом на основе спектрального вложения Грассманна-Штифеля (GSE). В проведенных экспериментах GSE показал наилучшее качество восстановления многообразий.

Tremendous amounts of data are becoming usual for modern data analysis. This refers not only to the rapidly growing sample sizes but also and sometimes only to the dimensionality of the data. Modern data analysis techniques generally have difficulties in handling high-dimensional data. It is a challenge for data analysis because of the inherent sparsity of the data in high dimensions.

A first step toward addressing this challenge is the use of dimension reduction techniques. There are lots of them offered by the modern analysis (see e.g. [8, 4, 6, 7, 9]). But newly arising problems bring specific demands on algorithms' abilities. One of them is not only to obtain the lower-dimensional representation of the data but also to be able to handle new (the so called out-of-sample) points and reconstruct initial representation from reduced-dimensional. This may be helpful in problems like generation of similar objects or optimization in lower dimensional space. In the present paper several approaches to unsupervised dimension reduction problem are compared on benchmark test problems. Both manifold embedding and reconstruction are examined.

Dimensionality Reduction

Let $\mathbf{X}_N = \{x_i \in \mathbb{R}^p\}_{i=1}^N$ be the given set of points. The main purpose of a dimensionality reduction procedure is to find some lower-dimensional representation of the data at hand, $\mathbf{Y}_N = \{y_i \in \mathbb{R}^d\}_{i=1}^N, d < p$, to discover underlying properties or improve abilities of other data mining techniques which use the data $\mathbf{X}_N$. Dimension reduction can be done in several ways. In this paper we consider such dimension reduction algorithms that preserve local topology of the dataset.

*The authors were partially supported by Laboratory of Structural Methods for Predictive Modeling and Optimization, MIPT, RF government grant, ag. 11.G34.31.0073.

Manifolds. Consider the curve shown in figure 1. Note that the curve is in $\mathbb{R}^3$, yet it has zero volume, and in fact zero area. The extrinsic dimensionality – three – is somewhat misleading since the curve can be parameterized by a single variable. One way of formalizing this intuition is via the idea of a manifold: the curve is a one-dimensional manifold because it locally "looks like" a copy of $\mathbb{R}^1$. Let us quickly review some basic terminology from geometry and topology in order to crystallize this notion of dimensionality.

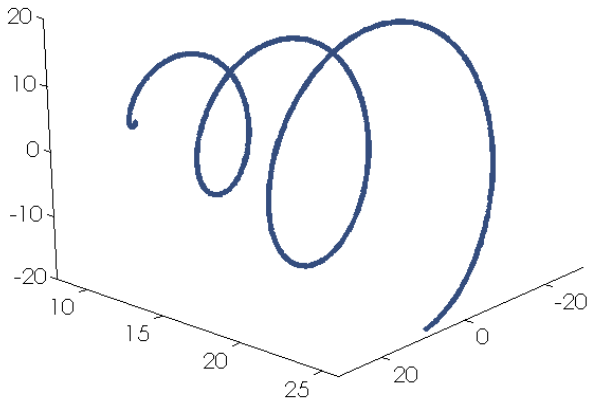

Fig. 1. An example of manifold

Definition 1. A homeomorphism is a continuous function whose inverse is also a continuous function.

Definition 2. A p-dimensional manifold $\mathcal{M}$ is a set that is locally homeomorphic with $\mathbb{R}^d$. That is, for each $x \in \mathcal{M}$, there is an open neighborhood around x, N_x, and a homeomorphism $f : N_x \to \mathbb{R}^d$. These neighborhoods are referred to as coordinate patches, and the map is referred to as a coordinate chart. The image of the coordinate charts is referred to as the parameter space.

The above definition is extremely general. We will be interested only in the case where $\mathcal{M}$ is a subset of $\mathbb{R}^p$, where p is typically much larger than d. In other words, the manifold will lie in a high-dimensional space $\mathbb{R}^p$, but will be *homeomorphic* to a low-dimensional space $\mathbb{R}^d$, with $d < p$.

Additionally, all the algorithms we study in this paper have some smoothness requirements that further constrain the class of manifolds considered.

Definition 3. A smooth (or differentiable) manifold is a manifold such that each coordinate chart is differentiable with a differentiable inverse (i.e., each coordinate chart is a diffeomorphism).

Generally the algorithms in this area are aimed to construct a mapping between train set and its reduced representation. However, some real problems require more information about the embedding.

Manifold Learning. An algorithm may learn an internal model of the data, which can be used to map points, unavailable at training time, into the embedding using a process often called out-of-sample extension. Throughout the paper we consider the following data generation model. Let $\mathcal{Y} \subset \mathbb{R}^d$ be the hidden parameter space and $\mathcal{X} \subset \mathbb{R}^p$ – the data space and the embedding (one-to-one injection)

$$f : \mathcal{Y} \to \mathcal{X},$$

that somehow preserves topological properties of points (e.g. the distances along the manifold).

The task is to construct an embedding $\mathbf{h}(x)$ and reconstruction $\mathbf{g}(y) = \mathbf{h}^{-1}(y)$, where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, given a set of points $\mathbf{X}_N = \{x_i \in \mathbb{R}^p\}_{i=1}^N$.

The reduced-dimensional space is not unique and not obliged to coincide with the real parameter space $\mathcal{Y}$. Suppose that we found some embedding and reconstruction functions $\mathbf{h}$ and $\mathbf{g}$. Consider an invertible function $\mathbf{s} : \mathbb{R}^d \to \mathbb{R}^d$ and a new parametrization $y^* = \mathbf{s}(\mathbf{h}(x))$ for any $x \in \mathcal{X}$ and $x^* = \mathbf{g}(\mathbf{s}^{-1}(y^*))$, where $y^* \in \mathbb{R}^d$. So, the only way to evaluate quality of a method is to calculate the reconstruction error on the independent test set (cf. with the original train set) of points from the same manifold.

Problem Statement. Given a set $\mathbf{X}_N = \{x_i \in \mathbb{R}^p\}_{i=1}^N$ and a value of reduced-dimensional representation d, find an embedding $\hat{\mathbf{h}} : \mathbb{R}^p \to \mathbb{R}^d$ that preserves local structure of the set and a reconstruction function $\hat{\mathbf{g}} : \mathbb{R}^d \to \mathbb{R}^p$ such that $\hat{x}_i = \hat{\mathbf{g}}(\hat{\mathbf{h}}(x_i))$ is close to x_i.

Dimension Reduction Techniques

Isomap. Isomap [8] is a combination of the Floyd-Warshall algorithm for finding shortest paths in a weighted graph with classic Multidimensional Scaling. Classic Multidimensional Scaling (MDS) takes a matrix of pair-wise distances between all points, and estimates a position for each point. So, Isomap initially calculates pair-wise distances only between neighboring points. And then Floyd-Warshall algorithm is employed to compute pair-wise distances between all points. This effectively estimates the full matrix of pair-wise geodesic distances between points. Isomap then uses classic MDS to compute the reduced-dimensional positions of all points.

Landmark-Isomap is a variant of this algorithm that uses landmarks to increase speed, at the cost of some accuracy.

Linear Embedding. Locally-Linear Embedding (LLE) [6] was developed at approximately the same time as Isomap. It has several advantages over Isomap, including faster optimization due to advantage of sparse matrix algorithms, and better results for many problems. LLE begins by finding a set of the nearest neighbors of each point. It then computes a set of weights for each point that best describe the point as a linear combination of its neighbors. Finally, the algorithm uses an eigenvector-based optimization technique to find the low-dimensional embedding of points, such that each point is still described with the same linear combination of its neighbors. LLE tends to handle non-uniform sample densities poorly.

Out-of-sample extension and manifold reconstruction for LLE are presented in [1] and [3] correspondingly.

Several improvements of the algorithm are proposed in literature. Some of them are described below.

Hessian LLE. Like LLE, Hessian LLE [4] also uses nearest neighbors for determining local geometry. But in contrast to original LLE it estimates Hessians, $H_f(x)$, of the embedding and calculates injection minimizing average Hessian norm

$$\int \left\| H_f(x) \right\|_F^2 \mathrm{d}x.$$

LLE tends to yield results of a much higher quality than LLE. Unfortunately, it has a very costly computational complexity, so it is not well-suited for heavily-sampled manifolds.

Conformal Eigenmaps. Conformal Eigenmaps [7] are based on the observation that local nonlinear techniques for dimensionality reduction (which computationally are usually reduced to calculation of eigenvectors corresponding to the highest eigenvalues for some matrix) do not employ information on the geometry of the data manifold that is contained in discarded eigenvectors that correspond to relatively small eigenvalues. So, Conformal Eigenmaps initially perform LLE (or alternatively, another local nonlinear technique for dimensionality reduction) to reduce the high-dimensional data to a dataset of dimensionality d_t , where $d < d_t < p$. Then Conformal Eigenmaps use the resulting intermediate solution in order to con-

struct a d-dimensional embedding that is maximally angle-preserving (i.e., conformal).

Local Tangent Space Alignment. LTSA [9] is based on the intuition that when a manifold is correctly unfolded, all of the tangent hyperplanes to the manifold will become aligned. It begins by computing the k-nearest neighbors of every point. Then it computes the tangent space at every point by computing first d principal components in each local neighborhood. Finally, LTSA algorithm finds an embedding that aligns the tangent spaces.

The manifold reconstruction procedure for LTSA is also described in the original paper.

Grassman-Stiefel Eigenmaps

Recently proposed local algorithm GSE [2] unlike other similar algorithms does not utilize the concept of nearest neighbors while using points' neighborhoods. This implies that the manifold should be contained in an ε-tube and enough points should lie in the ε-neighborhood for each data point. Before the actual description of the algorithm we have some definitions to start with.

Definition 4. The (compact) Stiefel manifold $S_{n,p}$ is the set of all p-tuples $(x_1, \dots, x_p)$ of orthonormal vectors in $\mathbb{R}^n$, i.e.

$$S_{n,p} = \{X \in \mathbb{R}^{n\times p} : X^T X = I_p\}.$$

Definition 5. A Grassmannian (Grassmann manifold) is a space $G(d, V)$ which parameterizes all linear subspaces of a vector space V of a given dimension d. The Grassmanians are compact, topological manifolds.

The embedding procedure consists of the following steps:

Preprocessing. For each $x_i \in \mathbf{X}_N$

1. Build a set $U_\varepsilon(x_i) = \{x\prime \in \mathbf{X}_N : \|x_i - x\prime\| < \varepsilon\}$ and
$$k_{\varepsilon,\tau}(x_i, x\prime) = \begin{cases} e^{-\tau\|x_i - x\prime\|^2}, & \text{if} \quad x\prime \in U_\varepsilon(x_i), \\ 0, & \text{otherwise,} \end{cases}$$
where $\varepsilon > 0$ and $\tau > 0$ are parameters.
2. Apply *Principal Component Analysis* for the set $U_\varepsilon(x_i)$ in order to determine d eigenvectors of the sample correlation matrix, corresponding to the largest eigenvalues $Q(x_i) = [q_1(x_i), \dots, q_d(x_i)]$ – compact Stiefel manifold.
3. Denote the intersection of the tanget spaces for x_i and $x\prime$ by $S(x_i, x\prime) = Q^T(x_i)Q(x\prime)$,
$$U_\eta(x_i) = \{x\prime \in U_\varepsilon(x_i) : \sqrt{1 - |S(x_i, x\prime)|^2} < \eta\}$$
and
$$k(x_i, x\prime) = k_{\varepsilon,\tau,\eta}(x_i, x\prime) = \\ = \begin{cases} k_{\varepsilon,\tau}(x_i, x\prime)|S(x_i, x\prime)|^2, & \text{if} \quad x\prime \in U_\eta(x_i), \\ 0, & \text{otherwise,} \end{cases}$$
where $\eta > 0$ is a parameter.

Tangent spaces alignment. Let Grassmaniann $H(x_i) = Q(x_i)V(x_i)$ be an approximation for embedding Jacobean and $\text{span}\{H(x_i)\} = \text{span}\{Q(x_i)\}$. Find such $V(x_i), i = 1, \dots, N$ that

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} k(x_i, x_j)\|H(x_i) - H(x_j)\|_F^2 \to \min_{V_i, i=1,\dots,N},$$

where $\|\cdot\|_F$ – Frobenius norm. This minimization could be reduced to eigenvalues problem for explicitly written matrix.

Embedding construction. Solve the system of linear equations to construct the embedding:

$$k(x_j, x_i)(x_j - x_i) = k(x_j, x_i)H(x_i)(y_j - y_i),$$
$$i, j = 1, \dots, N.$$

Out-of-sample embedding is based on the multidimensional nonparametric regression with kernels constructed in the same way as in preprocessing step of training set embedding.

The manifold reconstruction is based on the multidimensional nonparametric regression with kernels constructed in a way that $k(y\prime, y\prime\prime) \approx k(\mathbf{g}(y\prime), \mathbf{g}(y\prime\prime))$ to preserve local geometrical structure.

Embedding and Reconstruction

Some methods (e.g., LLE and LTSA) have explicit solutions, described in literature, for out-of-sample embedding and manifold reconstruction while others were not intended to solve the out-of-sample extension and/or reconstruction problems at all. A general workaround was applied for those manifold learning methods that do not have any explicit out-of-sample extension and/or reconstruction procedures, described in the literature.

In order to perform the out-of-sample extension for a new point x (see the full description in [5]):

1. Find the closest point $x\prime \in \mathbf{X}_N$ to x.
2. Calculate the local embedding matrix:
$$L = (x\prime - \overline{x\prime})^+(y\prime - \overline{y\prime}),$$
where $\overline{x\prime}$ is the mean vector over nearest neighbors of $x\prime$, $\overline{y\prime}$ is the mean vector over nearest neighbors of $y\prime$ and $(\cdot)^+$ represents the matrix pseudoinverse.
3. Calculate the reduced-dimensional representation as
$$y = \overline{y\prime} + (x - \overline{x\prime})L$$
for out-of-sample point x.

The same idea is used for reconstruction.

Numerical Experiments

In numerical experiments algorithms' embedding and manifold reconstruction abilities were tested on artificial datasets, such as 1d spiral in 3d space and Swissroll manifold.

All approaches appeared to be rather efficient in embeddings (provided sufficient sample size is given). While manifold reconstruction became a challenging task for most of the algorithms. A representative results on Swissroll manifold (see figure 3) are presented on figure 2 for different sample sizes. In all the experiments GSE outperforms other approaches in reconstructing initial manifold (thus preserving local structural features).

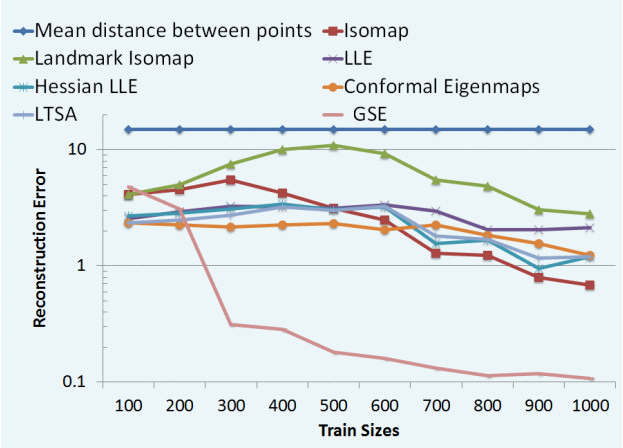

Fig. 2. Reconstruction error for different train sample sizes (Swissroll test case)

Discussion

The newly proposed algorithm, Grassmann-Stiefel Eigenmaps, proved to solve typical unsupervised dimension reduction problems. It also has good manifold reconstruction ability compared to other approaches which allows to solve more general class of problems and to broaden area of GSE algorithm application to supervised dimension reduction (feature extraction), target function optimization and several other data mining problems.

Further research should be aimed on

— automatic estimation of GSE parameters;
— algorithmic optimization to speed up the computations;
— studying the possibility of usage of GSE in other areas of data mining.

References

[1] Y. Bengio, J.-F. Paiement, and P. Vincent. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. In *Advances in Neural Information Processing Systems*, pages 177–184. MIT Press, 2003.

[2] A. Bernstein, E. Burnaev. Dimension Reduction Problem in Predictive Modelling. In Proceedings of *Intellectualization of Information Processing*, 2012.

[3] J. Chen, S.-J. Deng, and X. Huo. Electricity price curve modeling and forecasting by manifold learning, 2008.

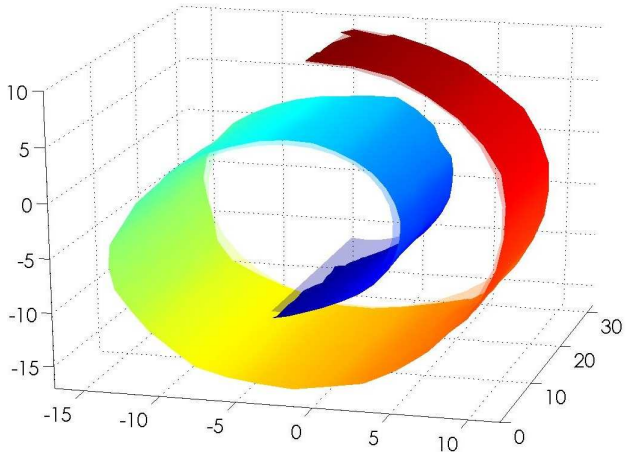

Fig. 3. Swissroll: GSE reconstructed manifold (solid) and original manifold (transparent)

[4] D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10):5591–5596, 2003.

[5] L.J.P. van der Maaten. An Introduction to Dimensionality Reduction Using Matlab. *Report MICC 07-07, Universiteit Maastricht, The Netherlands*, July 2007.

[6] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE*, 290:2323–2326, 2000.

[7] F. Sha and L. K. Saul. Analysis and extension of spectral methods for nonlinear dimensionality reduction. In *In Proceedings of the Twenty Second International Conference on Machine Learning (ICML-05*, pages 785–792, 2005.

[8] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000.

[9] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal of Scientific Computing*, 26:313–338, 2005.