ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Nonlinear data projection on a sphere with controlled trade-off between trustworthiness and continuity

Victor Onclinx[1,2], Vincent Wertz [1,2] and Michel Verleysen[1] *

1- Université catholique de Louvain - Machine Learning Group
Place du Levant, 3, 1348 Louvain-la-Neuve - Belgium

2- Université catholique de Louvain - Department of Applied Mathematics
Avenue Georges Lemaître, 4, 1348 Louvain-la-Neuve - Belgium

Abstract. This paper presents a nonlinear method aimed to project data on a non-Euclidean manifold, when their structure is too complex to be embedded in an Euclidean space. The method optimizes a pairwise distance criterion that implements a control between trustworthiness and continuity that respectively represent the risks of flattening and tearing the projection. The method is illustrated to project data on a sphere, but can be extended to other manifolds such as the torus and the cylinder.

1 Introduction

Nowadays, industrial domains have to deal with high dimensional data. To visualize them, projection methods try to minimize the loss of information between original and projected data points. The quality of the projection relies on its ability to preserve the pairwise distances, the neighbourhood and also on the choice of a suitable manifold. Indeed, if the data are too complex, the Euclidean space could be unadapted. Many state-of-the-art projection methods [1] try to preserve either pairwise distances [2, 3, 4] or neighbourhoods [5, 6]. Most of the methods project data on an Euclidean space. However, assuming that original data lie on an unknown manifold, their global, possibly non-Euclidean, topology contains an important part of the information: for example, when the manifold intercepts itself. This is widely known in the context of neighbourhood based methods like SOM [7, 8, 9, 10] where it is common to project on a sphere or on a torus. This paper presents a distance based projection method on a non-Euclidean manifold. The method is detailed for the case of the sphere but the approach can be adapted to other manifolds like the cylinder or the torus.

Moreover the projection methods have to deal with a trade-off between trustworthiness and continuity [11], which means respectively avoiding flattening and tearing the projection. The method presented in this paper controls the trade-off explicitly through the pairwise distance criterion developped in Section 2.

Because the sphere is a nonlinear submanifold embedded in the Euclidean space, the projection process is based on differential geometry theory. In Section

*V. Onclinx is funded by a grant from the Belgium F.R.I.A. Parts of this work present 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).

3, we show how the gradient descent can be adapted to ensure the convergence of the optimization, using the theory of optimization on manifolds [12, 13]. Preliminary results are shown in Section 4 where improvements of the projection on the sphere are assessed trough the trustworthiness and continuity criteria.

2 Projection criterion

The projection of a cylinder on the Euclidean space $\mathbb{R}^2$ illustrates the trade-off between trustworthiness and continuity that dimensionality reduction methods have to deal with. As a first attempt, the cylinder can be flattened such that two opposite generatrices are projected one onto the other. Such a projection is not trustworthy because two close projected data points cannot be trusted to be close in the original space. Conversely, the cylinder can be ripped up along one generatrix and then be unfolded such that some close original data points go away from each other; in this case the projection is not continuous anymore. These types of behaviour led to the idea of two intuitive quality measures [11] that count points that are close in one space (original or projected) and faraway in the other one. However, these quality measures are difficult to optimize directly because they are not continuous. To bypass this problem, most projection methods try to preserve distances by minimizing a weighted distance criterion.

Without weighting, the idea is that distances δ_{ij} between data points i and j in the projected space must remain close to the corresponding original distances D_{ij}. Minimizing the criterion $\sum_{i=1}^{N-1}\sum_{j>i}^{N}(D_{ij}-\delta_{ij})^2$, where N is the number of data points to be projected, does not always lead to good results as large distances (in one of the two spaces) influence too much the criterion. Dividing each term of the sum by the distance D_{ij} between the original data points favours the continuity of the projection. Indeed, the smaller is the original distance, the largest is the weight in $(D_{ij}-\delta_{ij})^2/D_{ij}$, which results in more importance given to tears (small D_{ij} and large δ_{ij}). A tearing error term can thus be expressed:

$$\mathit{Tearing\ error} \equiv \sum_{i=1}^{N-1}\sum_{j>i}^{N}\frac{(D_{ij}-\delta_{ij})^2}{D_{ij}}.$$

Conversely, dividing each term by the distance δ_{ij} in the projection space leads to the flattening error that measures the trustworthiness:

$$\mathit{Flattening\ error} \equiv \sum_{i=1}^{N-1}\sum_{j>i}^{N}\frac{(D_{ij}-\delta_{ij})^2}{\delta_{ij}}.$$

Indeed if two projected data points are close whereas the corresponding original data points are distant, the flattening error will increase.

The total error f measures a compromise between the trustworthiness and the continuity, more specifically between the flattening and the tearing errors, through a user-defined control parameter $\lambda \in [0,1]$:

$$f \equiv \sum_{i=1}^{N-1} \sum_{j>i}^{N} \lambda \frac{(D_{ij} - \delta_{ij})^2}{D_{ij}} + (1-\lambda) \frac{(D_{ij} - \delta_{ij})^2}{\delta_{ij}}. \quad (1)$$

3 Optimization procedure

We now show how the criterion (1) proposed in the previous section can be optimized on a specific non-Euclidean manifold, in our case the sphere. As a first attempt, we could define $\mathbf{y}_i$ in the spherical coordinate space and optimize (1) in a direct way by taking its gradient according to the spherical coordinates. Unfortunately, the search space should have to be restricted to $\{(\phi, \theta) \in [0, 2\pi[\times[\frac{-\pi}{2}, \frac{\pi}{2}]\}$. Moreover, in this space, there exist singularities in the corresponding north and south poles; for example, the line segment $\theta = \frac{\pi}{2}$ for $\phi \in [0, 2\pi[$ corresponds to a single point on the sphere, the north pole.

In order to circumvent these difficulties, the minimization problem is reformulated as an optimization problem in the Euclidean space $\mathbb{R}^3$, with the additive constraint that the projected data points must remain on an embedded manifold - in our case a sphere. To search for a minimum of f, a gradient descent procedure is used. In order to add the manifold constraint, the algorithm is adapted to the optimization on manifold principle [12] that relies on differential geometry.

Firstly, at iteration k, the algorithm tries to determine a search direction. The gradient $\nabla f(\mathbf{y}_1(k), ..., \mathbf{y}_N(k), \nu(k))$ of the criterion is evaluated, where ν, discussed later, represents the parameters of the manifold $\mathcal{M}$. The aim of the gradient direction is to approximate locally the cost function. Because of the spherical constraints, the gradient must stay close to the manifold $\mathcal{M}$. The projection of ∇f on the tangent space $T_{\mathbf{y}(k)}\mathcal{M}$ is then used:

$$\nabla' f(\mathbf{y}_1(k), ..., \mathbf{y}_N(k), \nu(k)) = Proj_{T_{\mathbf{y}(k)}\mathcal{M}}(\nabla f).$$

Secondly, the algorithm updates the current locations $\mathbf{y}(k)$. Through this second search direction $\nabla' f$, a new set of locations $\mathbf{y}'(k+1)$ is evaluated on the tangent space $T_{\mathbf{y}(k)}\mathcal{M}$ using a step size α:

$$\mathbf{y}'(k+1) = \mathbf{y}(k) - \alpha \nabla' f.$$

Third, in order to satisfy the manifold constraint, another set $\mathbf{y}(k+1)$ is evaluated by a *Retraction* step [12] which is a kind of deterministic projection of the tangent space on the manifold.

Finally, in order to provide a sufficient decrease of the cost function, the step size α must satisfy the Armijo condition [12]. Let $\sigma \in [0,1]$; if the condition $f(\mathbf{y}(k)) - f(\mathbf{y}(k+1)) \geq \sigma\alpha ||\nabla' f||^2$ is satisfied, $\mathbf{y}(k+1)$ is accepted; otherwise, the step size α is decreased. When $\sigma = 0$, the condition checks the decrease of the cost function. But if $\sigma \neq 0$, the decrease of the cost function is compared to the expected decrease of its first order approximation in the less powerful

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

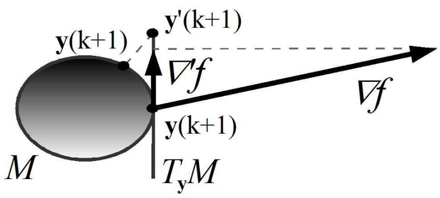


Fig. 1: Illustration of an optimization iterate

direction $-\sigma\nabla' f$. The different steps are repeated until convergence and are illustrated in Fig. 1.

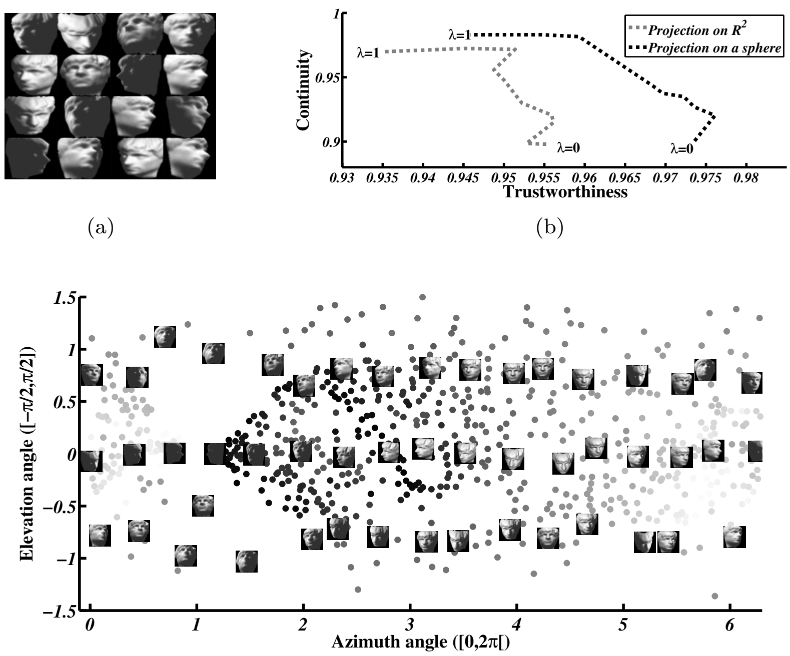
The general methodology of the optimization on a manifold is now applied to our problem of minimizing the criterion f defined by (1) on the sphere; the expressions of the manifold $\mathcal{M}$ and the tangent space $T_{\mathbf{y}}\mathcal{M}$ must therefore be detailed. Because the most appropriate radius R of the sphere cannot be determined *a priori* (remind that distances on the sphere must correspond, according to (1), to the distances in the original space, whose scale is fixed), it is considered as another variable. This leads to the following definitions:

$$\begin{aligned}
\mathcal{M} &\equiv \{(\mathbf{y}_1, ..., \mathbf{y}_N, R) \in S^3 \times ... \times S^3 \times \mathbb{R}^+ | \mathbf{y}_i^T \mathbf{y}_i - R^2 = 0, 1 \leq i \leq N\}, \\
T_{\mathbf{y}}\mathcal{M} &\equiv \{(\mathbf{u}_1, ..., \mathbf{u}_N, u_R) \in \mathbb{R}^3 \times ... \times \mathbb{R}^3 \times \mathbb{R} | \mathbf{y}_i^T \mathbf{u}_i - R u_R = 0, 1 \leq i \leq N\}.
\end{aligned}$$

Concerning the distance δ_{ij} in the projected space, it is naturally defined by $\delta_{ij} \equiv Rarccos\frac{\mathbf{y}_i'\mathbf{y}_j}{R^2}$. In the original high-dimensional space, a graph is constructed between the data; the distances D_{ij} are evaluated by the shortest path [3, 4] as an approximation to the geodesic distances on the original and unknown manifold. As for the evaluation of the gradient, ∇f is the vector of the partial derivatives with respect to the variables $\mathbf{y}_i$ and R.

4 Results

In order to evaluate the performance of the projection method, an experiment is performed on the widely known data base of virtual face pictures [1, 14]. This database contains 698 pictures of 64×64 pixels of faces taken from different angles and lighting. The dimension of the original space is thus 4096 while its intrinsic dimension is 3 because the pictures can be totaly described by the elevation and the azimuth angles of the camera and by the lighting. Samples of these pictures are presented in Fig. 2(a). The distance D_{ij} between the data is evaluated by the shortest path in the graph built between the pictures with 15 neighbours [3, 4]. The faces are projected both on $\mathbb{R}^2$ and on the sphere, according to criterion (1), in order to show the advantages of a projection on a sphere instead of on $\mathbb{R}^2$. Different values of λ are used to evaluate the trustworthiness and the continuity, as defined in [11]. The closer to 1 these measures are, the most trustworthy or continuous the projection is. As shown in Fig. 2(b), the results corresponding to the projection on a sphere are closer to $(1, 1)$ than those corresponding to a projection on $\mathbb{R}^2$. To visualize the projected data, they are represented in the spherical coordinate space in Fig. 2(c).



(a) (b)

(c)

Fig. 2: (a) Sample of face data base; (b) comparison between the projections on $\mathbb{R}^2$ and on the sphere with the trustworthiness and the continuity measure; (c) Representation of the projected face data on the sphere for $\lambda = 0.7$ in the spherical coordinate space (the color varies as the azimuthal angle of the camera does). Only a few data are represented as faces to increase the readability of the figure and to show the smoothness of the projection.

5 Conclusion

This paper describes a nonlinear data projection method able to project data on a non-Euclidean manifold (the sphere is taken as an example) by preserving the pairwise distances. It combines different interpretations of the information underlying the distribution of the data. Firstly, with a pairwise distance criterion, the proximity among data can be visualized and quantified. Secondly, the method controls the trade-off between trustworthiness and continuity by the introduction of flattening and tearing errors in the criterion. Finally, to embed the global topology of the high dimensional manifold, the method builds the mapping on a non-Euclidean manifold using a dedicated optimization method.

The paper presents the projection on a sphere; however the methodology can be easily generalized to other manifolds like the torus or the cylinder, through the use of the generic optimization on manifold methodology. Future work will include the development of the projections on other manifolds, including the automatic determination of the most suitable one in a family of manifolds. First ideas in this direction may be found in [15].

Acknowledgements

The authors thank Prof. Pierre-Antoine Absil for his suggestions on the optimization on manifolds.

References

[1] J.A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction.* Springer Science+Business Media, LLC, 233 Spring Street New York, NY, 10013, USA, 2007.

[2] J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[3] J.A. Lee, A. Lendasse, N. Donckers, and M. Verleysen. A robust nonlinear projection method. In M. Verleysen, editor, *Proceedings of ESANN 2000, 8th European Symposium on Artificial Neural Networks*, pages 13–20, Bruges, Belgium, April 2000. D-Facto public.

[4] P. Demartines and J. Hérault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks*, 8(1):148–154, January 1997.

[5] J.A. Lee and M. Verleysen. Nonlinear projection with the isotop method. In J.R. Dorronsoro, editor, *LNCS 2415: Artificial Neural Networks, Proceedings of International Conference on Artificial Neural Networks 2002*, pages 933–938, Madrid, Spain, August 2002. ICANN, Springer-Verlag.

[6] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[7] H. Ritter. Self-organizing maps on non-euclidean spaces. In S. Oja and E. Kaski, editors, *Kohonen Maps*, pages 97–110. Elsevier, Amsterdam, 1999.

[8] J.X. Li. Visualization of high-dimensional data with relational perspective map. *Information Visualization*, 3(1):49–59, 2004.

[9] H. Nishio, Md. Altaf-Ul-Amin, K. Kurokawa, K. Minato, and S. Kanaya. Spherical som with arbitrary number of neurons and measure of suitability. In *5th Workshop On Self-Organizing Maps*, pages 323–330, Paris, France, September 2005.

[10] Y. Wu and M. Takatsuka. Spherical self-organizing map using efficient indexed geodesic data structure. *Neural Network*, 19(6):900–910, 2006.

[11] J. Venna and S. Kaski. Neighborhood preservation in nonlinear projection methods: An experimental study. In G. Dorffnet, H. Bischof, and K. Hornik, editors, *Proceedings of the International Conference on Artificial Neural Networks*, pages 485–491, Vienna, Austria, August 2001. Springer-Verlag.

[12] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, Princeton, NJ, January 2008.

[13] D. P. Bertsekas. *Nonlinear programming.* Athena Scientific, Belmont, Massachusetts, 1995.

[14] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Sci. Comput.*, 26(1):313–338, 2005.

[15] J.A. Lee and M. Verleysen. Nonlinear dimensionality reduction of data manifolds with essential loops. *Neurocomputing*, 67:29–53, 2005.