LAAT: Locally Aligned Ant Technique for discovering multiple faint low dimensional structures of varying density

Abolfazl Taghribi, Kerstin Bunte, Rory Smith, Jihye Shin, Michele Mastropietro, Reynier F. Peletier, and Peter Tiňo

---

Our method is neither a manifold learning, nor a clustering technique, but rather a preprocessing strategy that potentially improves the results of such methods when they are applied subsequently on the LAAT output. We incorporated the following text in the manuscript introduction to make this more clear to the reader:

> Therefore the proposed method is neither a traditional clustering nor manifold learning method. It is rather a denoising strategy that improves the result of such techniques in multiple scenarios: 1) significant presence of outliers, 2) multiple elongated clusters or nonlinear manifolds, and 3) structures with varying noise within and across.

As explained in the box above, LAAT can denoise a dataset with multiple nonlinear faint structures and different dimensions, that are embedded in a significant noisy background, as often encountered in astronomical applications. Manifold learning techniques such as LLE [1], Isomap [2], or Hessian eigenmaps [3] cannot discover the embeddings of multiple manifolds, and they are typically sensitive to noise, even if it is only situated close to manifold. Some other manifold learning algorithms such as SMCE [4], LRNE [5], or hierarchical GTM [6] can extract multiple manifolds. However, as explained in quite recent papers (published in 2019 and 2022), methods such as LRNE [5] and the novel abstract GTM (AGTM) [7], can, again, only handle a relatively low amount of noise along the manifold. As the number of noise and outliers increases the output results become more and more unreliable. Therefore, we conclude that developing strategies to extract multiple manifold-like structures is still a very active research field, that recognizes the presence of noise as a major problem. The proposed LAAT for noise removal alleviates this problem.

We designed an experiment to show the deficiencies of the conventional manifold learning techniques in scenarios that our proposed technique excels in. In Figure 1 we illustrate the result of applying SMCE, AGTM, and LRNE on the two-arm dataset that is introduced in our paper. For each technique we experimented with several parameter settings as recommended in the respective papers proposing them. Nonetheless, neither of these techniques can uncover the underlying embeddings in the presence of a large number of outliers. Note also, that both SMCE and LRNE assume the dimension of the underlying embeddings is known and provided as an input (dim = 2). As a result, they force all points in the dataset including the noise and outliers to construct two-dimensional manifolds that do not exist. AGTM, on the other hand, extends the embedding graph (green lines) to the whole dataset as presented in Figure 1(d), which indicates, that this method also cannot find the manifolds buried in a large amounts of noise. Note, that all these approaches are state-of-the-art techniques for multiple manifold learning, that just struggle with large amounts of background noise. The reason to bring this comparison is to illustrate the type of scenarios that LAAT is designed to solve. It can therefore assist other techniques to find the embedding of manifolds in the resulting less noisy medium.

Other algorithms such as MBMS [8] and SAF [9] are diffusion strategies introduced to decrease the noise by moving noise points towards the closest structure. Thus, they enhance clusters and manifolds, which can, subsequently, improve the output of both clustering and manifold learning algorithms. Nonetheless, they also typically cannot remove a large number of outliers far away or recover many manifolds from different underlying dimensions in a noisy medium. Figure 2 demonstrates the unsatisfactory result of SAF and MBMS

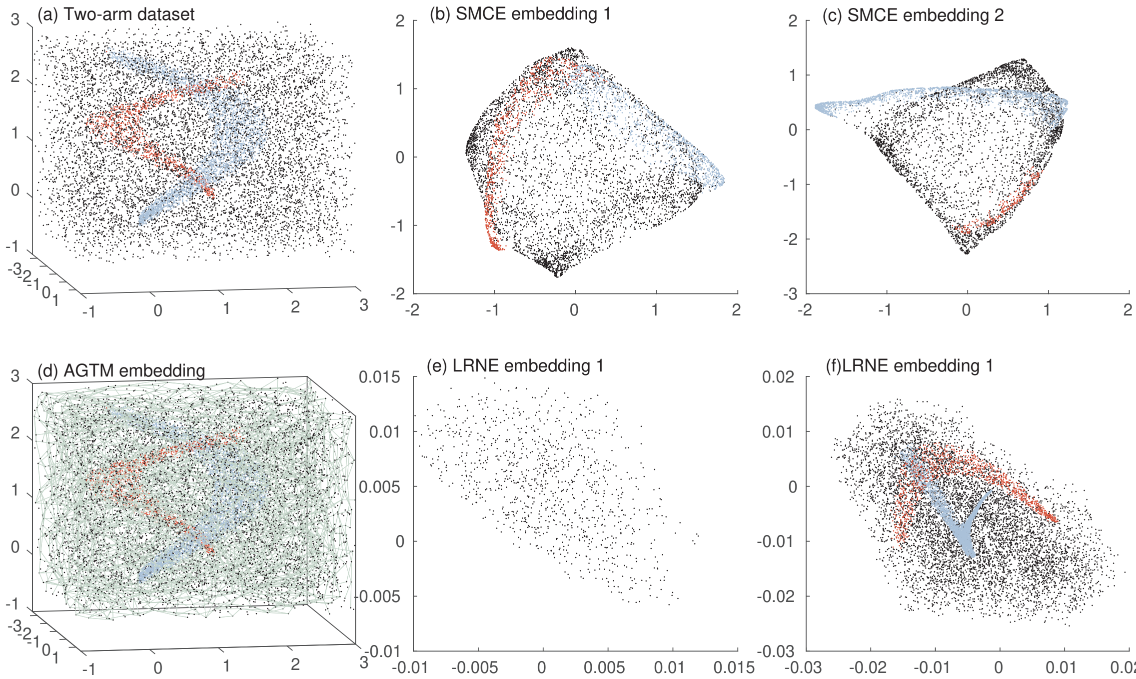

Figure 1: The original two-arm data set (a) and results of SMCE (b and c), AGTM (d), and LRNE (e and f).

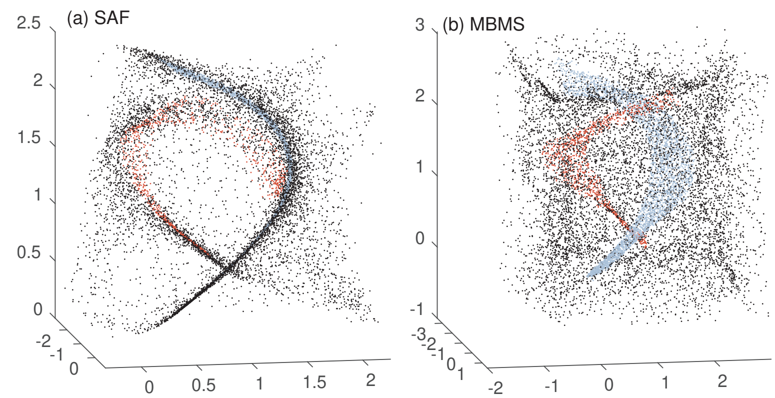

Figure 2: Two-arms result for SAF (a) and MBMS (b), which both move noise points towards the manifolds.

on our two-arms dataset in panels a and b, respectively. We experimented with different parameter settings to tune the algorithms, but since the background noise is not aligned with the manifolds these techniques have difficulties to remove it.

However, if we remove the outliers from the dataset in advance using LAAT, and subsequently apply MBMS or SAF to the remaining relevant point cloud, we can retrieve the clean manifolds. These cleaned manifolds can then also be successfully modeled with AGTM, where AGTM used to fail on its own. Figure 3 exemplifies how the LAAT strategy extracts the manifolds and therefore enables the subsequent methods to further clean and model them.

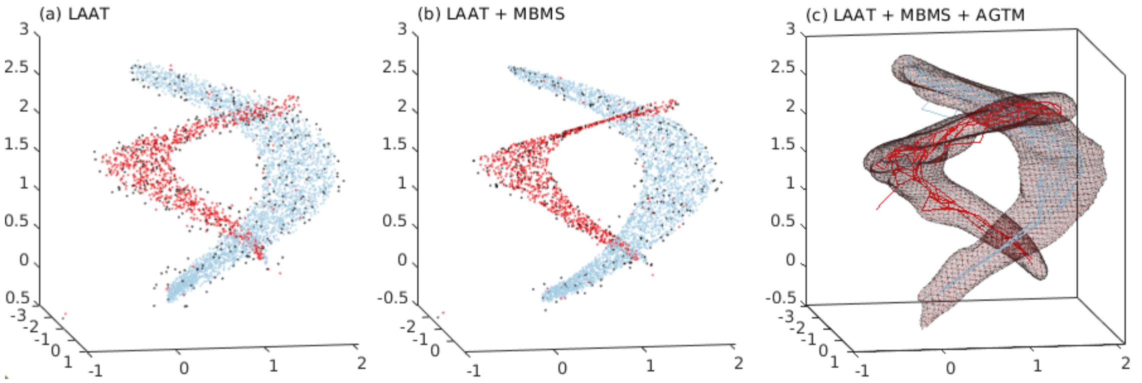

Figure 3: LAAT (a) extracts the manifolds and enables successful application of MBMS (b) and AGTM (c).

References

[1] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323–2326, Dec. 2000. Publisher: American Association for the Advancement of Science Section: Report.

[2] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, pp. 2319–2323, Dec. 2000. Publisher: American Association for the Advancement of Science.

[3] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 5591–5596, May 2003. ISBN: 9781031596106 Publisher: National Academy of Sciences Section: Physical Sciences.

[4] E. Elhamifar and R. Vidal, "Sparse Manifold Clustering and Embedding," in *Advances in Neural Information Processing Systems*, vol. 24, Curran Associates, Inc., 2011.

[5] A. M. Saranathan and M. Parente, "On Clustering and Embedding Mixture Manifolds Using a Low Rank Neighborhood Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 3890–3903, June 2019. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

[6] P. Tino and I. Nabney, "Hierarchical GTM: Constructing Localized Nonlinear Projection Manifolds in a Principled Way," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 639–656, May 2002.

[7] M. Canducci, P. Tiño, and M. Mastropietro, "Probabilistic modelling of general noisy multi-manifold data sets," *Artificial Intelligence*, vol. 302, p. 103579, Jan. 2022.

[8] W. Wang and M. A. Carreira-Perpiñán, "Manifold blurring mean shift algorithms for manifold denoising," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1759–1766, June 2010. ISSN: 1063-6919.

[9] S. Wu, P. Bertholet, H. Huang, D. Cohen-Or, M. Gong, and M. Zwicker, "Structure-Aware Data Consolidation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 2529–2537, Oct. 2018.