ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Adaptive Locally Aligned Ant Technique for Manifold Detection and Denoising

Felipe Contreras[1,2*], Kerstin Bunte [1] and Reynier Peletier [1]

1 - University of Groningen, Netherlands 2 - Universidad de Valparaíso, Chile
{contreras.sep.felipe;kerstin.bunte}@gmail.com r.f.peletier@rug.nl

Abstract. The detection and extraction of noisy manifolds from data have various applications. In Astronomy, the detection of faint streams and filaments is particularly difficult due to background contamination, which immerses and hides them in noise. The biologically inspired Locally Aligned Ant Technique (LAAT) has been demonstrated as an efficient and flexible algorithm to detect and denoise versatile structures within noisy backgrounds. Our contribution extends LAAT two-fold: (1) introduction of a dynamic local radius, and (2) locally variable pheromone deposition. The former avoids highlighting spurious patterns in noisy regions and allows smaller jumps in areas with strong alignment. The latter increases pheromone deposition in fainter zones. We demonstrate this in 2 datasets.

1 Introduction

Astronomers study evolutionary processes and the history of cosmological interactions by analysing the structures left behind, often employing large N-body simulations [1]. These structures are typically non-linear, ubiquitous, diffuse, of varying size and density, and immersed in large amounts of background noise, for which conventional manifold learning techniques fail [2, 3]. Topology and Delaunay tessellation to find the medial axis are computationally expensive, and their results are sensitive to sampling effects. The recently introduced *Locally Aligned Ant Technique* (LAAT) [4] detects an arbitrary number of diffuse manifolds of different dimensionality and varying density, embedded in large amounts of noise and outliers, inspired by the efficient ant colony algorithm [5, 6]. It uses local alignment information and pheromone dynamics to reinforce faint structures. LAAT extracts relevant points as first step in the *1D Recovery, Extraction and Analysis of Manifolds in noisy environments* (1-DREAM) pipeline[1] [7], that constructs sparse models of each structure, demonstrated for the cosmic web, jellyfish galaxies, and streams [7, 8, 9]. While LAAT is quite robust concerning its parameters, it retains more noise around high-density structures, which can be alleviated by local post-processing [10] that avoids choosing a global threshold. However, the size and faintness of structures that can be detected are sensitive to the radius set by the user. In this contribution, we propose two extensions to the LAAT algorithm, namely a local dynamic radius and dynamic pheromone deposition. The novel strategies allow the deposition of more pheromones in faint structures and retain less noise in high-density areas and backgrounds. Sensitivity analysis is performed on a synthetic data set and finally demonstrated in an astronomical N-body cosmological simulation.

*support by the National Agency for Research and Development scholarship 2020-21200114.
[1] 1-DREAM code publicly available at https://git.lwp.rug.nl/cs.projects/1DREAM

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

2 Methodology

LAAT is an ant colony-based algorithm for the efficient detection and extraction of an arbitrary number of noisy manifolds of different dimensionality and varying density, demonstrated most notably in Astroinformatics [7, 8, 9, 11]. Assume n data points of dimension D $\{\boldsymbol{x}_i \in \mathbb{R}^{\mathrm{D}}\}_{\mathrm{i=1}}^{\mathrm{n}}$ and their neighbourhood $\mathcal{N}_r^i$ with radius r, that is used to compute local eigenvalues $\{\lambda_d\}_{d\leq D}$ and eigenvectors $\{\boldsymbol{v}_d\}_{d\leq D}$ through the principal component analysis (PCA). The transition probability from $\boldsymbol{x}_i$ to one of its neighbours $\boldsymbol{x}_j \in \mathcal{N}_r^i$ in the original LAAT formulation [4] depends on two quantities: 1) the alignment of the jump vector $(\boldsymbol{x}_j - \boldsymbol{x}_i)$ with the eigenvalues $\boldsymbol{v}_d$ computed as angle $|\cos\alpha_d^{(i,j)}|$ weighted by the eigenvalues λ_d. And 2) the biologically inspired normalized pheromone value $\overline{F}^j(t)$ at point $\boldsymbol{x}_j$ at time t accumulated in previous visitations. Both together not only highlight dense regions (as the Markov Chain) but also fainter areas with dominant eigenvectors that indicate manifold structures. While its parameters are quite robust concerning stochastic variations of the data, the global choice of certain hyper-parameters limit the result. For example, a global pheromone threshold can retain more noise in areas of high density or lose fainter structures of low density, which can be tackled by local thresholding [10]. Similarly, a global neighbourhood radius r for the ant's transition probabilities limits the size of the structures that can be found. Therefore, we present LAAT extensions of dynamic radius and pheromone deposition, to improve the detection of manifolds of vastly different sizes and densities.

2.1 Dynamic radius

We propose a dynamic radius for each data point $\boldsymbol{x}_i$ that is more likely to deposit more pheromone on points that are part of a manifold than noise. Since there is no clear mathematical definition of the astrophysical structures of interest, an objective function cannot be obtained from theoretical considerations. Intuitively an objective function should: **1)** Choose a large radius in areas without preferred alignment, to allow bigger jumps of the ants, which result in quicker traversal and smaller amounts of pheromone deposited. Those areas typically occur in either **(1a)** *background noise* or **(1b)** *isotropic overly dense space*. A bigger radius allows the ants to escape dense attraction zones and reduces the probability of highlighting random small over-densities within the noise. **2)** Choose small and medium radii in **(2a)** *manifold structures*, allowing the ants to spend more time and therefore deliver more pheromone in them. Prefer the smallest radii in **(2b)** *low density areas of the structures* producing a greater balance in the distribution of the pheromone throughout the entire manifold.

Most aims depend only on the alignment of the respective neighbourhood. Except for 2b, which is based on the structure itself, which cannot be determined locally and is left for future work. To tackle (1a-2a) we propose an objective function that considers the robustness of alignment in the presence of a data perturbation. We introduce two values $R_{\min}$ ($R_{\max}$) $\in \mathbb{R}$, representing the minimum (maximum) possible radius and K, a constant chosen by the user to determine the maximum number of radii associated with each point, to steer computa-

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

tion and memory costs. Hence, for every point $\boldsymbol{x}_i$ we associate a set of radii $\mathbf{r}_i = \{\{r_i^1, r_i^2, ..., r_i^{K_i}\} \mid r_i^k < r_i^{k+1}\ \forall k\}$, where $1 \leq K_i \leq K$. Each neighbourhood $\mathcal{N}_{r^k}^i := \mathcal{B}(\boldsymbol{x}_i, r_i^k)$ is associated with a number of neighbours m_i^k induced by radius r_i^k. We impose $m_i^k < m_i^{k+1}\ \forall k$ as we are interested in unique neighbourhoods, and hence K_i can be strictly minor than K. For simplicity we define $m_i^{\min}$ and $m_i^{\max}$ as the number of neighbours in $\mathcal{B}(\boldsymbol{x}_i, R_{\min})$ and $\mathcal{B}(\boldsymbol{x}_i, R_{\max})$ and m_{th} as the minimum number neighbours for a radius to be accepted, to compute PCA properly.We chose m_i^k to increase linearly with the radius:

$$\widehat{m}_i^{\min} = \max\{m_{\text{th}}, m_i^{\min}\}; \quad dt_i = \max\left\{\left(m_i^{\max} - \widehat{m}_i^{\min}\right)/(K-1), 1\right\} \tag{1}$$

$$\boxed{m_i^k = \widehat{m}_i^{\min} + \lfloor dt_i(k-1)\rfloor} \text{ for } 1 \leq k \leq K_i; \quad K_i = \min\{m_i^{\max} - \widehat{m}_i^{\min} + 1, K\}.$$

The neighbourhoods defined by each set of radii $\mathbf{r}_i$ are associated with a set of eigenvectors $\mathcal{V}_i^k = \{\boldsymbol{V}_i^k\}_{k=1}^{K_i}$ and eigenvalues $\Lambda_i^k = \{\lambda_i^{(k,d)}\}_{d=1}^D$, with $\mathbf{V}_i^k = \{\boldsymbol{v}_i^{(k,d)}\}_{d=1}^D$, where $\boldsymbol{v}_i^{(k,d)}$ is the d-th eigenvector of neighbourhood $\mathcal{N}_{r^k}^i$. The alignment preference [4] for moving from $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$ in neighbourhood k is:

$$E_i^{(j,k)} = \sum_{d=1}^{D} \frac{|\cos\alpha_d^{(i,j)}|}{\sum_{d'=1}^{D}|\cos\alpha_{d'}^{(i,j)}|} \cdot \frac{\lambda_i^{(k,d)}}{\sum_{d'=1}^{D}\lambda_i^{(k,d')}} \text{ with } j \leq m_j^k. \tag{2}$$

We propose a robustness criterium to determine the probabilities for each local radius r_i^k to be chosen for transition. Concretely, we define two ways to drop-out a percentage from each neighbourhood $\mathcal{N}_{r^k}^i$ and measure the difference of the eigenvectors before and after. **a) Major drop-out** removes neighbours with the highest preference values $E_i^{(j,k)}$ and **b) random drop-out** removes with equal probability. Subsequently, we use the Grassmann distance d_G (or Bhattacharyya, Hellinger) to measure the difference between the subspace spanned by the eigenvectors $\mathcal{V}_i^k$ before, and $\widetilde{\mathcal{V}}_i^k$ after neighbourhood perturbation:

$$d_G(\mathcal{V}_i^k, \widetilde{\mathcal{V}}_i^k) = \sum_{d=1}^{D} \arccos\left[\frac{|\langle \mathbf{v}_i^{(k,d)}, \widetilde{\mathbf{v}}_i^{(k,d)}\rangle|}{\|\mathbf{v}_i^{(k,d)}\| \cdot \|\widetilde{\mathbf{v}}_i^{(k,d)}\|}\right]; \quad 1 \leq k \leq K_i\ , \tag{3}$$

with $0 \leq d_G|_i^k = \frac{2}{\pi D} d_G(\mathcal{V}_i^k, \widetilde{\mathcal{V}}_i^k) \leq 1$. We furthermore tested two radii selection probabilities. **Parallel preference** P_{p} favours radii which preserve the original neighbourhood alignment information, and **orthogonal preference** P_{o} increases the probability for radii that change the alignment:

$$P_{\text{p}}(r_i^k) = \frac{e^{-\log(K_i)\left(d_G|_i^k\right)^2}}{\sum_{k'=1}^{K_i} e^{-\log(K_i)\left(d_G|_i^{k'}\right)^2}}; \quad P_{\text{o}}(r_i^k) = \frac{e^{-\log(K_i)\left(1-d_G|_i^k\right)^2}}{\sum_{k'=1}^{K_i} e^{-\log(K_i)\left(1-d_G|_i^{k'}\right)^2}}\ , \tag{4}$$

for $1 \leq k \leq K_i$. One could use local factors similar to the perplexity of tSNE [12], which requires non-linear problem solving. As computationally cheap alternative we use $\log(K_i)$ that increases (decreases) the certainty of choice at points that have more (fewer) radii to choose from. Intuitively, background noise regions tend to have fewer radii than dense ones, inducing a more uniform choice.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

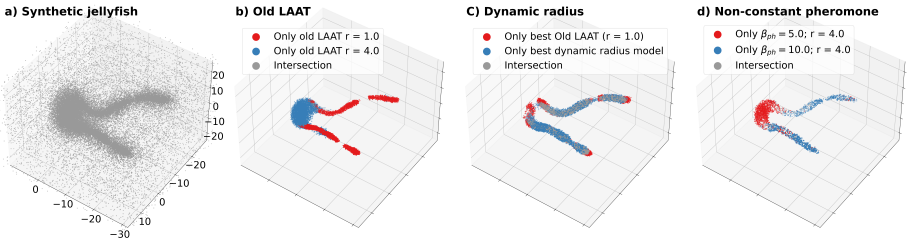

Fig. 1: a) Synthetic jellyfish. b) 14% of points highlighted by old LAAT. c) 22.5% of high pheromone points comparing the best LAAT and best dynamic radius. d) 15% of highlighted points for dynamic pheromone deposition with different β_{ph}.

2.2 Dynamic pheromone deposition

Like the Markov Chain, LAAT visits high-density regions more often, and hence accumulates more pheromone there. This effect is further amplified, since the ants are also attracted by pheromones, enabling them to reinforce weak signals. However, this can also trap them in dense attractors, as often seen in astronomical data in the form of Galaxies, Globular Clusters, and nodes of the cosmic web. To alleviate this effect we vary the pheromone deposited on visitation based on the transition preference $E_i^{(j,k)}$ from $\boldsymbol{x}_i$ to neighbour $\boldsymbol{x}_j$:

$$F^j(t) = e^{-\beta_{\text{ph}}\left(1-E_i^{(j,k)}\right)} \quad \text{with } 1 \leq j \leq m_i^k \ , \tag{5}$$

and m_i^k being the number of neighbours in neighbourhood $\mathcal{N}_{r^k}^i$, and the adjustable parameter β_{ph} that determines the importance given to aligned points.

3 Experiments and Discussion

In this section, we perform a parameter sensitivity analysis of the new dynamic radius and pheromone deposition using a synthetic jellyfish galaxy as introduced in [7]. Furthermore, we demonstrate the improvements on a cube of 40^3 Mpc^3/h selected from a Dark Matter-only N-body cosmological simulation with $\approx 2.7{\cdot}10^5$ particles. We always use 100 epochs and $K = 100$ radii. In the first (second) experiment we use 5^3 ants (7^3 ants), with 2500 (12000) steps and radii in [1,4] (0.05 to 1.5 Mpc/h). Other parameters were set to default.

The synthetic jellyfish: consists of two branches and a head, immersed in Gaussian noise, each part being composed of 10^4 data points, shown left in Figure 1. For the systematic test, we compare the minimum amount of high pheromone points highlighted by the different strategies and parameters, that are necessary to retain the structures without losing relevant information in Table 1. We consider a structure recovered if the subsequent application of the 1DREAM [7] pipeline is able to model the full backbone. The synthetic jellyfish exhibits two difficulties, namely the head has no alignment and the filaments vary in density. The old LAAT accumulates pheromone unevenly in the structures, favouring dense parts over the faint filament areas and the effect exacerbates with increasing radius. Panel b) of Figure 1 shows that a small radius allows the ants to spend more time in filamentary structures, fulfilling aim (2a), while bigger radii accumulate higher amounts in dense regions, useful to achieve the aim (1a

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Table 1: Minimun data necessary to recover the full jellyfish structure

1) Fixed radius with variable pheromone delivered				
Strategy \Radius size	1.0	2.0	3.0	4.0
Old LAAT	29.0%	32.5%	36.5%	40.0%
$\beta_{\text{ph}} = 1.0$	27.0%	31.5%	35.0%	37.5%
$\beta_{\text{ph}} = 2.5$	27.0%	32.0%	33.5%	36.0%
$\beta_{\text{ph}} = 5.0$	27.0%	26.5%	30.0%	27.5%
$\beta_{\text{ph}} = 7.5$	30.0%	28.0%	21.5%	**14.5%**
$\beta_{\text{ph}} = 10.0$	29.0%	27.0%	24.5%	20.5%
2) Dynamic radius				
Strategy \Drop-up percentage		15.0%	30.0%	45.0%
Parallel + Major drop-out		39.0%	49.0%	49.0%
Orthogonal + Major drop-out		27.5%	**22.5%**	31.5%
Parallel + Random drop-out		38.0%	42.0%	39.5%
Orthogonal + Random sort		41.0%	38.0%	37.0%
3) Best case, Orthogonal + Major sort at 30% of drop-out				
$\beta_{\text{ph}} = 1.0$	$\beta_{\text{ph}} = 2.5$	$\beta_{\text{ph}} = 5.0$	$\beta_{\text{ph}} = 7.5$	$\beta_{\text{ph}} = 10.0$
22.5%	23.5%	25.0%	28.0%	30.0%

and 1b), but not all at once when using a fixed radius. The best model with dynamic radius to achieve all these aims uses orthogonal preference with 30% Major drop-out, as shown in panel c). It favours small radii in areas with strong alignment information, and bigger ones where the eigenvectors change a lot with perturbation, which is the case in dense regions and background noise. The parallel preference has the opposite effect. The major drop-out removes key points, which is a stricter measure of robustness than random drop-out, which produces better results. In contrast to the dynamic radius the dynamic pheromone is intuitive to understand and model. It deposits more pheromone in areas with strong alignment than areas without, which counteracts the accumulation in dense regions to even it with fainter structured ones. Increasing β increases this effect as shown in Table 1 and Figure 1d). In contrast to the dynamic radius this is a global criterium and the result depends highly on the interplay between β and the radius, making it quite sensitive and difficult to set on its own. However, combined with the dynamic radius it becomes less sensitive and produces better results than almost every setting of the old LAAT.

The Cosmic web data: demonstrates the behaviour of the new LAAT in an astronomical application. Panel b) and c) Figure 2 show 50% of the data highlighted with the old and new LAAT. The old version needs a fixed radius of 1.5 Mpc/h, since smaller ones have many neighbourhoods underflow. This is avoided with the dynamic strategy and hence increases user-friendliness. Furthermore, the distinct union plot in panel d) shows that old LAAT keeps more noise around the high-density nodes and misses a lot of faint filaments as compared to the new strategy, that surpasses the old one substantially, detecting filamentous structures more effectively and better defined, requiring less information.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

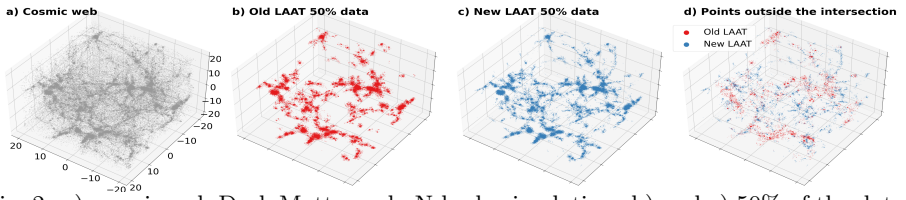

Fig. 2: a) cosmic web Dark Matter-only N-body simulation. b) and c) 50% of the data highlighted with old LAAT with 1.5 Mpc/h radius, and the best new LAAT model. d) the distinct union shows new LAAT retains fainter filaments.

4 Conclusions and future work

This paper presents improvements to the Locally Aligned Ant Technique (LAAT) for the detection and noise removal of multiple diffuse manifolds in the presence of large amounts of noise and outliers. The dynamic radius and pheromone extension are more robust and user-friendly, highlighting structures more clearly, and needing fewer points for it. In future work, we will replace high dense areas with sparse models in the ant algorithm to avoid local attractors even further.

References

[1] A. Knebe, F. R. Pearce, H. Lux, Y. Ascasibar, P. Behroozi, J. Casado, and et. al. Structure finding in cosmological simulations: the state of affairs. *Mon. Not. R. Astron. Soc.*, 435(2):1618–1658, 2013.

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

[3] M. Hein and M. Maier. Manifold Denoising. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in NeurRIPS 19*, pages 561–568. MIT Press, 2007.

[4] A. Taghribi, K. Bunte, R. Smith, J. Shin, M. Mastropietro, R. F. Peletier, and P. Tino. LAAT: Locally aligned ant technique for discovering multiple faint low dimensional structures of varying density. *IEEE Trans. Knowl. Data Eng.*, 2022.

[5] M. Dorigo, M. Birattari, and T. Stutzle. Ant colony optimization. *IEEE Comput. Intell. Mag.*, 1(4):28–39, 2006.

[6] S. Sengupta, S. Basak, and R. A. Peters. Particle swarm optimization: A survey of historical and recent developments with hybridization perspectives. *Mach. Learn. Knowl. Extr.*, 1(1):157–191, 2018.

[7] M. Canducci, P. Awad, A. Taghribi, M. Mohammadi, M. Mastropietro, S. De Rijcke, R. F. Peletier, R. Smith, K. Bunte, and P. Tino. 1-DREAM: 1d recovery, extraction and analysis of manifolds in noisy environments. *Astron. Comput.*, page 100658, 2022.

[8] P. Awad, R. Peletier, M. Canducci, R. Smith, A. Taghribi, M. Mohammadi, J. Shin, P. Tino, and K. Bunte. Swarm intelligence-based extraction and manifold crawling along the large-scale structure. *Mon. Not. R. Astron. Soc.*, 520(3):4517–4539, 02 2023.

[9] M. A. Raj, P. Awad, R. F. Peletier, R. Smith, U. Kuchner, R. van de Weygaert, N. I. Libeskind, M. Canducci, P. Tino, and K. Bunte. The large-scale structure around the fornax-eridanus complex. *Astron. Astrophys.*, 690:A92, 7 2024.

[10] F. Contreras, R. Peletier, and K. Bunte. Improved the locally aligned ant technique (laat) strategy to recover manifolds embedded in strong noise. In *31st ESANN*, 2023.

[11] P. Awad, L. S. Ting, D. Erkal, P. F. Reynier, K. Bunte, and et al. s^5: New insights from deep spectroscopic observations of the tidal tails of the globular clusters NGC 1261 and NGC 1904. *Astron. Astrophys.*, 11 2024.

[12] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *J. Mach. Learn. Res.*, 9(11), 2008.