ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Topology-Preserving Prototype Learning on Riemannian Manifolds

Lucas Schwarz[1], Magda Psenickova[2], Thomas Villmann[2,3] and Florian Röhrbein[1] *

1 - University of Technology Chemnitz, Dep. of Computer Science
Chemnitz - Germany
2 - University of Applied Sciences Mittweida, Dep. of Mathematics
Mittweida - Germany
3 - Technical University Bergakademie Freiberg, Dep. of Mathematics
Freiberg - Germany

Abstract. Learning prototypes in an unsupervised manner that respects the data density and topology is crucial for tasks such as clustering, representation learning, and visualization of high-dimensional datasets. In this paper, we propose a generalization of the Neural Gas algorithm to Riemannian manifolds, leveraging geodesic distances for prototype adaptation. The approach additionally generates a prototype neighborhood structure, enabling faithful approximation of both geometry and topology of data distributed on Riemannian manifolds. We demonstrate its effectiveness on real-world datasets from manifolds such as $SO(n)$, S^n_{++} and $Gr(n,k)$ and compare our approach to Riemannian versions of other related methods such as K-Means, K-Medoids and a Riemannian Self-Organizing Map.

1 Introduction

Riemannian geometries arise naturally in many real-world datasets, such as in EEG covariance matrices or object poses represented as homogeneous transformation matrices in a robotics context [1]. These datasets obey geometric constraints that are often lost when embedded in Euclidean space, as is common in many machine learning methods [2]. Respecting the intrinsic geometry is therefore essential for tasks like clustering or representation learning. Euclidean assumptions and standard distance metrics may fail on curved manifolds, where notions such as distance, mean, and gradient must be redefined in a Riemannian framework. Although prior work has explored unsupervised prototype learning on manifolds [3, 4, 5], the problem of generating a condensed representation of the topology of the data remains underexplored. This topology could be used to infer admissible transitions between discrete prototypes under support of the data, such as in switching neural states in an EEG context, for example. We therefore propose a generalization of the Neural Gas algorithm [6] to Riemannian manifolds, capable of learning accurate prototypes and their topology via stochastic gradient descent on an adjusted energy function.

*L. S. is supported by BMFTR (Project Nr. 16SV9458), M. P. is supported by the European Social Fund (ESF, 100715238)

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2 Preliminaries

The paper is organized as follows: First we introduce basic concepts of Riemannian geometry, followed by a description of the Riemannian Neural Gas in the following section. Afterwards, we consider the formal properties in the next section. The last section presents experimental results on synthetic and real-world datasets, comparing our approach to related methods.

2.1 Riemannian Geometry

A manifold $\mathcal{M}$ of dimension n is a Hausdorff topological space locally homeomorphic to an open subset of $\mathbb{R}^n$. A Riemannian manifold is a pair $(\mathcal{M}, g)$, where $g(\cdot,\cdot) : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$ defines a smoothly varying inner product on each tangent space $T_x\mathcal{M}$ given for $x \in \mathcal{M}$. The arc length of a piecewise smooth curve $\gamma(t) : [0,1] \to \mathcal{M}$ is given by $L(\gamma) = \int_0^1 \sqrt{g(\gamma'(t),\gamma'(t))}\, dt$. A curve minimizing $L(\gamma)$ is denoted as a geodesic, and its length defines the Riemannian geodesic distance $d_{\mathcal{M}}$. Furthermore, in iterative algorithms, we must ensure that new iterates $x^{t+1} \in \mathcal{M}$ remain on the manifold. To achieve this, a retraction map $R_x : T_x\mathcal{M} \to \mathcal{M}$ is used to move along a tangent direction while staying on $\mathcal{M}$. The canonical choice is the exponential map Exp_x: for a curve $\gamma(t) \in \mathcal{M}$ with initial velocity $\gamma'(0) = v_x \in T_x\mathcal{M}$, $\mathrm{Exp}_x(v_x) = \gamma(1)$ gives the endpoint of the geodesic at $t = 1$. Its inverse, $\mathrm{Log}_x = \mathrm{Exp}_x^{-1} : \mathcal{M} \to T_x\mathcal{M}$, maps a point in the manifold back to the tangent space. Additionally, the injectivity radius $\mathrm{inj}(x)$ is the supremum of all radii around $x \in \mathcal{M}$ for which Exp_x is a diffeomorphism onto its image.

2.2 Neural Gas

The Neural Gas (NG, [6]) is an unsupervised learning algorithm to approximate the probability distribution of data by means of a set of prototypes. In particular, for given data vectors $X \subseteq \mathbb{R}^n$ obtained according to the data density $P(x)$ and equipped with the Euclidean metric, the NG adapts the prototypes $W = \{w_1, ..., w_N\} \in \mathbb{R}^n$ using stochastic gradient descent according to

$$\Delta w_i \propto h_\lambda(k_i(x, W)) \cdot (x - w_i) \qquad i = 1, 2, ..., N \tag{1}$$

such that the energy function

$$E = \frac{1}{2 \cdot C_\lambda} \sum_{i=1}^{N} \int_X P(x) h_\lambda(k_i(x, W)) d(x, w_i)^2 dx \tag{2}$$

is minimized. The corresponding winning rank $k_i(x, W)$ of a prototype w_i with respect to a sample x is given as $k_i(x, W) = \#\{l : d_l^2 < d_i^2\} = \sum_l \theta(d_i^2 - d_l^2)$ where the Heaviside function $\theta(\cdot)$ determines implicitly a neighborhood structure in W between the prototypes in dependence on x using the Euclidean distance $d_i := d(x, w_i)$. The function $h_\lambda(x, W) = \exp(-k_i(x, W)/\lambda)$ determines the neighborhood cooperativity during the prototype learning adjusted by the neighborhood range $\lambda > 0$. Usually, NG starts with large λ-values for rapid rough prototype adaptation in the beginning and is slowly decreased during

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

learning to allow faithful fine-tuning. The positive quantity C_λ in (2) is constant for given (fixed) λ. For trained NG, a topological structure within the prototypes set W can be imposed using the Voronoï cells

$$V_i = \{x \in X : d(x, w_i) \leq d(x, w_j),\ \forall j\} \tag{3}$$

which are restricted to the data X. The corresponding adjacency graph $G_W = (W, E_W)$ for the prototypes W and edges E_W is determined by

$$(w_i, w_j) \in E_W \iff V_i \cap V_j \neq \emptyset \tag{4}$$

augmenting the NG to the Topology Representing Network (TRN) [7]. For the TRN it is required that the prototype set W is dense in X, i.e. if w_i and w_j are the closest and second-closest prototypes to x, respectively, the triangle $\triangle(x, w_i, w_j)$ is fully contained in X.

3 The Riemannian Neural Gas Algorithm

To enable the determination of prototypes from data distributed on a Riemannian manifold, we augment the NG in two ways: First, we replace the standard Euclidean distance in the energy function (2) with the geodesic distance $d_{\mathcal{M}}$ and secondly, we modify the prototype update rule, such that an update is performed by moving the prototypes a small step along the geodesics $\gamma(t)$ connecting x and each w_i, ensuring consistency with the manifold's local geometry. For this purpose, we compute $v_i = \epsilon \cdot h_\lambda(k_i(x, W)) \cdot \mathrm{Log}_{w_i}(x)$ in the tangent space at w_i and map back to the manifold via $w_i^{new} \leftarrow \mathrm{Exp}_{w_i}(v_i)$. Additionally, for each prototype, we restrict the updates s.t. $K_i \subset \mathcal{M}$, where K_i are geodesically convex neighborhoods around each w_i. This ensures the existence of a unique minimizing geodesic for any $x \in K_i$. Using these observations, the energy function of the NG can be adapted for the Riemannian metric, which yields

$$E = \frac{1}{2 \cdot C_\lambda} \sum_{i=1}^{N} \int_{K_i \cap X} P(x) h_\lambda(k_i(x, W)) d_{\mathcal{M}}(x, w_i)^2 \, d\mu(x) \tag{5}$$

where $d\mu(x) = d\,\mathrm{vol}_g(x)$ is the Riemannian volume constituting an n-form [8] associated with the Riemannian metric g on $(\mathcal{M}, g)$. The corresponding Riemannian-like prototype update is obtained as

$$w_i^{new} \leftarrow \mathrm{Exp}_{w_i}(\epsilon \cdot h_\lambda(k_i(x, W)) \cdot \mathrm{Log}_{w_i}(x)) \tag{6}$$

using the exponential map and its inverse [9]. Further, $\epsilon > 0$ is a sufficiently small learning rate, such that w_i^{new} remains in the injectivity radius of w_i, hence $w_i^{new} \in K_i$ remains valid.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

In practice, we can bound the tangent-space adaptation steps v_i by

$$v_i \leftarrow v_i \cdot \min\left(1, \frac{\tau \cdot \text{inj}(w_i)}{||v_i||_{w_i}}\right) \tag{7}$$

if $||v_i||_{w_i} > \tau \cdot \text{inj}(w_i)$ with $\tau \in (0,1)$ representing a safety margin factor and $||v_i||_{w_i} := \sqrt{g_{w_i}(v_i, v_i)}$. This adaptation preserves the update direction but scales down the magnitude. We denote this approach as Riemannian NG (RNG). The corresponding Riemannian TRN (RTRN) is obtained by replacing the distances $d(x, w_j)$ in (3) with their Riemannian counterparts $d_{\mathcal{M}}(x, w_j)$.

4 Formal Aspects

We highlight two formal aspects of the Riemannian Neural Gas (RNG) algorithm, namely that (i) its update rule (6) corresponds to a stochastic gradient descent step on the energy function (5) using geodesic distances, and (ii) the TRN topology construction method can be adapted accordingly.

(i) For RNG we require that the manifold $(\mathcal{M}, g)$ has a C^2 differentiable structure and is connected. Then, the update (6) is obtained in complete analogy to the proof in [6] for standard NG using the relation

$$\nabla_{w_i} \frac{1}{2} d_{\mathcal{M}}(w_i, x)^2 = -\text{Log}_{w_i}(x) \tag{8}$$

Here, the first variation formula [8] is applied, which states that for any $\gamma(t)$ with $\gamma(0) = w_i$ and $\dot{\gamma}(0) = v \in T_{w_i}\mathcal{M}$, the directional derivative $D(d_{\mathcal{M}}(w_i, x)^2)[v]$ into direction v satisfies $D(d_{\mathcal{M}}(w_i, x)^2)[v] = g_{w_i}(-2\ \text{Log}_{w_i}(x), v)$. Both $\text{Log}_{w_i}(x)$ and $\text{Exp}_{w_i}(x)$ are uniquely defined and smooth under the made restrictions regarding the geodesically convex local neighborhoods K_i, which is essential.

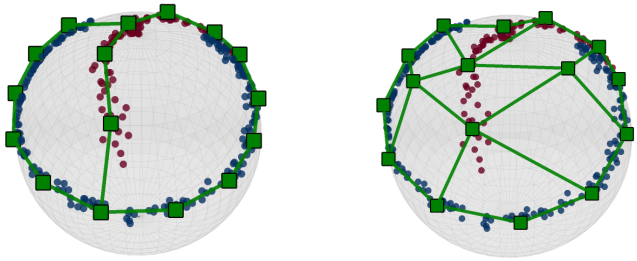
(ii) The TRN property for RNG is verified again in complete analogy to the standard TRN by replacing the distances in (3) with their Riemannian counterparts $d_{\mathcal{M}}(x, w_i)$ and using the metric properties.

5 Experiments

We evaluate the performance of the Riemannian Neural Gas on synthetic and real-world datasets, comparing it to Riemannian variants of related algorithms: Self-Organizing Map (RSOM) [4], Principal Surface (RPS) [10], K-Means [3], K-Medoids [5] and Competitive Learning Riemannian Quantization (CLRQ) [11]. The first dataset is the *Two Moons* toy dataset from Scikit-Learn, projected onto $\mathbb{S}^2$, with one moon having double the density of data points. The second dataset, *Monash*, comes from robotics and contains stable object grasp poses [1] represented as 4x4 homogeneous transformation matrices. We use the first four classes and extract 3×3 rotation matrices describing grasp orientations, which are elements of $SO(3)$. The third dataset, from the *BCI IV 2a* competition, consists of covariance matrices from EEG signals which are elements of the manifold of symmetric positive definite matrices $\mathcal{S}^{25}_{++}$. For the Grassmannian

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

manifold $Gr(10, 220)$, we follow [12] to extract 20 ten-dimensional subspaces from 220 hyperspectral bands for the 12 largest classes of the Indian Pines dataset [13]. We set the number of prototypes or discretization points to 16, 16, 16, and 48 prototypes for the *Moons*, *Monash*, *BCI IV*, and *Pines* datasets respectively and train for 50 epochs. The hyperparameters were selected according to the respective author's recommendations. The performance is evaluated using the mean and standard deviation of quantization error (QE) and silhouette score (SIL) over five runs, calculated using the geodesic distances. We also measure the topographic error (TE) [14] to assess topology preservation for the RNG, RSOM and RPS. A visualization of an exemplary learning result for 16 prototypes on the projected *Moons* dataset is shown in Fig. 1. As summarized in Tab. 1, the Riemannian K-Means provides a robust baseline as evidenced through QE and SIL, often achieving the second-best results. Yet, the RNG learns prototypes that accurately represent the data in terms of the QE and better preserve the topology compared to the RSOM or RPS across different manifolds and datasets, while also achieving competitive clustering performance in terms of SIL.



(a) (b)

Fig. 1: $\mathbb{S}^2$ Learning results for RNG (a) and RSOM (b).

Table 1: Experimental results over five runs. Best bold, second best underlined.

Dataset & Manifold		Algorithm					
		R - NG	R - SOM	R - PS	R - K-Means	R - K-Medoids	CLRQ
Moons $\mathbb{S}^n$	QE	**0.11 ± 0.01**	0.26 ± 0.01	0.48 ± 0.06	0.12 ± 0.01	0.13 ± 0.01	0.16 ± 0.02
	SIL	**0.49 ± 0.01**	0.38 ± 0.01	0.36 ± 0.04	0.46 ± 0.02	0.45 ± 0.02	0.41 ± 0.04
	TE	**0.01 ± 0.01**	0.05 ± 0.02	0.23 ± 0.10	-	-	-
Monash SO(n)	QE	**0.26 ± 0.01**	0.44 ± 0.01	0.33 ± 0.02	0.27 ± 0.01	0.31 ± 0.01	0.33 ± 0.03
	SIL	**0.31 ± 0.02**	0.11 ± 0.01	0.28 ± 0.03	0.30 ± 0.03	0.21 ± 0.03	0.25 ± 0.06
	TE	**0.13 ± 0.03**	0.36 ± 0.01	0.45 ± 0.07	-	-	-
BCI IV $\mathcal{S}^n_{++}$	QE	**2.76 ± 0.01**	3.59 ± 0.01	4.57 ± 0.09	2.83 ± 0.06	3.47 ± 0.02	3.24 ± 0.06
	SIL	**0.22 ± 0.01**	0.13 ± 0.01	0.01 ± 0.02	0.19 ± 0.01	0.18 ± 0.01	0.17 ± 0.02
	TE	**0.18 ± 0.04**	0.31 ± 0.01	0.26 ± 0.01	-	-	-
Pines $\mathrm{Gr}(k, n)$	QE	**1.33 ± 0.01**	6.98 ± 0.71	2.25 ± 0.11	1.53 ± 0.02	1.93 ± 0.01	1.52 ± 0.01
	SIL	0.02 ± 0.01	0.01 ± 0.02	0.01 ± 0.01	**0.08 ± 0.01**	0.07 ± 0.01	0.05 ± 0.01
	TE	**0.41 ± 0.01**	0.90 ± 0.01	0.87 ± 0.06	-	-	-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

6 Conclusion

In this work, we introduced an extension of the Neural Gas algorithm to datasets on Riemannian manifolds, based on geodesic distances and incremental updates in tangent spaces. This enables the generation of prototypes and their topology in an accurate manner. Furthermore, we provided formal and empirical arguments for the viability of the algorithm. Applications of the approach could be found in representation learning of manifold-valued data and as a method to acquire accurate topologies of latent spaces of deep neural networks, which have been recently interpreted as Riemannian manifolds [2].

References

[1] Andreas Ten Pas and Robert Platt. Using geometry to detect grasp poses in 3d point clouds. In *Robotics Research: Volume 1*, pages 307–324. Springer, 2017.

[2] Sophia Sanborn, Johan Mathe, Mathilde Papillon, Domas Buracas, Hansen Lillemark, Christian Shewmake, Abby Bertics, Xavier Pennec, and Nina Miolane. Beyond euclid: An illustrated guide to modern machine learning with geometric, topological, and algebraic structures. In *2025 Joint Mathematics Meetings (JMM 2025)*. AMS, 2025.

[3] Alvina Goh and René Vidal. Clustering and dimensionality reduction on Riemannian manifolds. In *2008 IEEE Conference on computer vision and pattern recognition*, pages 1–7. IEEE, 2008.

[4] Dongjun Yu, Edwin R Hancock, and William AP Smith. A Riemannian self-organizing map. In *International Conference on Image Analysis and Processing*, pages 229–238. Springer, 2009.

[5] P Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1):S143–S152, 2009.

[6] Thomas M Martinetz, Stanislav G Berkovich, and Klaus J Schulten. 'neural-gas' network for vector quantization and its application to time-series prediction. *IEEE transactions on neural networks*, 4(4):558–569, 1993.

[7] Thomas Martinetz and Klaus Schulten. Topology representing networks. *Neural Networks*, 7(3):507–522, 1994.

[8] John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.

[9] A. Engelsberger, M. Pseničkova, and T. Villmann. Prototype learning for classification on spherical manifolds and its relation to quantum classification approaches. In L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, and J.M. Zurada, editors, *Proceedings of the 24th International Conference on Artificial Intelligence and Soft Computing - ICAISC'2025, Zakopane*, volume Part I of *LNAI*, pages 82–88, Cham, 2026. Springer International Publishing, Switzerland.

[10] Søren Hauberg. Principal curves on Riemannian manifolds. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1915–1921, 2015.

[11] Alice Le Brigant and Stéphane Puechmorel. Quantization and clustering on Riemannian manifolds with an application to air traffic analysis. *Journal of Multivariate Analysis*, 173:685–703, 2019.

[12] Michael Kirby and Chris Peterson. Visualizing data sets on the Grassmannian using self-organizing mappings. In *2017 12th international workshop on self-organizing maps and learning vector quantization, clustering and data visualization (WSOM)*, pages 1–6. IEEE, 2017.

[13] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe. 220 band AVIRIS hyperspectral image data set: June 12, 1992 indian pine test site 3, Sep 2015.

[14] Kimmo Kiviluoto. Topology preservation in self-organizing maps. In *Proc. of International Conf. on Neural Networks (ICNN'96)*, volume 1, pages 294–299. IEEE, 1996.