# Expressivity of congruence-based architectures for DNNs on positive-definite matrices

Antonin Oswald
ICTEAM, UCLouvain
antonin.oswald@uclouvain.be

Symmetric and positive-definite (SPD) matrices arise naturally in a wide range of scientific and engineering applications [3, 4], where they encode essential second-order structures such as correlations between signals. The ubiquity of SPD matrices as data representations has motivated the design of machine learning algorithms specifically dedicated to this type of data. Recent years have seen the extension of various machine learning models to SPD matrices, including deep learning approaches [1, 2]. A classical way to extract features from SPD matrices is through congruence-like transformations of the form $X \mapsto WXW^\top$ where $W^\top \in \mathbb{R}^{n\times p}$ has full column rank [2]. These transformations are alternated with nonlinear activations functions, that operate on the matrix eigenvalues. The resulting feature extraction block is then followed by a classifier, as illustrated on Figure 1. The numerical stability of the training procedure is ensured by enforcing (semi)-orthogonality on the weight matrices of the model.

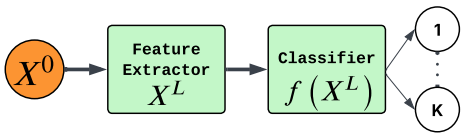


Figure 1: Deep neural network.

**Network description:** Let $\phi : \mathrm{SPD}(d) \to \mathbb{R}^K$ be a neural network mapping SPD-valued inputs to class scores. Let $X^0 \in \mathrm{SPD}(d)$. Let $g(\cdot)$ be a matrix function, and let $L$ denote the depth of the network. The network propagates the signal according to

$$\begin{aligned} X^l &= g\Big(W^l X^{l-1} W^{l\top}\Big), \qquad l = 1, \dots, L, \\ \phi(X^0) &= f(X^L), \end{aligned} \tag{1}$$

where $\{W^l \in \mathbb{R}^{d^l \times d^{l-1}}\}_{l=1}^L$ are the weight matrices of the feature extractor, $f : \mathrm{SPD}(d_L) \to \mathbb{R}^K$ is a classifier, and $g$ is an primary matrix activation function.

We explore here the role of depth in architecture (1). We highlight that the (commonly used) (semi-)orthogonality constraint on the weight matrices impairs model expressivity for these layers, as for some choices of activation functions, the resulting architecture collapses to a one-hidden-layer equivalent network. This lack of expressivity results from a drop of diversity in the spectrum of congruence-like layers, when the weight matrices are restricted to be (semi-)orthogonal, and is a direct consequence of Poincaré's separation theorem. As a second step, we discuss the choice of the final classifier: we compare several celebrated Riemannian classifiers, and discuss their compatibility with the feature map extraction mechanism implemented by congruence-like layers.

**Joint work with:** Estelle Massart (ICTEAM, UCLouvain, estelle.massart@uclouvain.be). This work was supported by the *GRAVIT-AI* Concerted Research Action (ARC) of the Fédération Wallonie-Bruxelles.

## References

[1] Daniel Brooks, Olivier Schwander, Frederic Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for SPD neural networks. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[2] Zhiwu Huang and Luc Van Gool. A Riemannian network for SPD matrix learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 2036–2042. AAAI Press, 2017.

[3] Estelle M. Massart and Sylvain Chevallier. Inductive means and sequences applied to online classification of EEG. In *3rd International Conference on Geometric Science of Information (GSI)*, Paris, France, November 2017. Springer.

[4] Diego Tosato, Michela Farenzena, Mauro Spera, Vittorio Murino, and Marco Cristani. Multi-class classification on Riemannian manifolds for video surveillance. pages 378–391, 09 2010.