Editorial: Special issue on "Geometrical Methods in Neural Networks and Learning"

Simone Fiori[1] and Shun-ichi Amari[2]

[1] *Faculty of Engineering – University of Perugia*
Polo Didattico e Scientifico del Ternano
Via Pentima bassa, 21, I-05100 Terni (Italy, EU)
E-mail: fiori@unipg.it
[2] *RIKEN Brain Science Institute*
Laboratory for Mathematical Neuroscience
Wako-shi Hirosawa 2-1, Saitama 351-0198 (Japan)
E-mail: amari@brain.riken.go.jp

Simply stated, the brain is a remarkable information processor that operates with principles that are completely different from those informing modern computers. Information is represented in the brain by spatio-temporal patterns of excitation spread over a large number of neurons and the brain processes information in parallel through the dynamic interactions of neurons so that efficient and flexible information processing takes place. Also, one of the most interesting features of the brain is that it creates a model of the world by learning and self-organization.

As the design of new systems is getting increasingly inspired by the functions and mechanisms of the brain, these principles have started to shape the development of information technology through a new science that is now commonly referred to as *neurocomputing*. One of the goals of neurocomputing is to establish a new type of information science as an indispensable step toward understanding how the brain works. This is also a necessary step toward creating brain-style information technology. Such trend has been made possible through the art of mathematical neuroscience, whose aim is to elucidate the principles of the brain by establishing mathematical theories on how it works.

Mathematical theories provide insights into how the brain might function as well as powerful analysis tools to test those ideas. Unlike the developmental drivers in computer science, which have traditionally relied exclusively on logic and algebra to produce algorithms and computability, neurocomputing has drawn from a variety of mathematical approaches in its pursuits. Mathematical analysis, and nonlinear analysis in particular, dominates neurocomputing because learning takes place in neurons, which are nonlinear elements of

Preprint submitted to Elsevier Preprint *25 November 2004*

analog computation that rely on parallel dynamics. Neurons also fire stochastically, making probability theory and statistics fundamental means to model the dynamics of learning using stochastic equations.

Geometry, however, while being a part of the mathematical bedrock, has yet to be embraced by neurocomputing. The intrinsic and deep structures of subjects, where invariant properties are sought for, can be intuited through geometry. Invariant structures abound in neural networks: Perhaps that is why geometrical theories have been gradually creeping into mathematical neuroscience. Understanding the underlying geometric structure of a network's parameter space is extremely important to designing systems that can effectively navigate the space while learning. Although modern mathematics is needed in the research of neural networks, and there are some very powerful results and techniques in these geometric methods, these are currently scattered in various sources and research directions.

The present special issues accommodates some general contributions on the usefulness of general algebraic-geometric theories in neurocomputing and machine learning.

The manuscript "Geometrical learning, descriptive geometry, and biomimetic pattern recognition", authored by Shoujue Wang and Jiangliang Lai, proposes a geometrical learning theory from the perspective of high-dimensional descriptive geometry. Geometrical properties of high dimensional structures underlying a set of samples are learned via successive projections from the higher dimension to the lower dimension until 2D Euclidean plane, under guidance of the established properties and theorems in high dimensional descriptive geometry. Specifically, the Authors introduce a novel methodology for learning samples and provide a geometrical learning algorithm that is then applied to biomimetic pattern recognition. Experimental results are presented in the paper to show that the proposed approach outperforms three types of SVMs with either a three degree polynomial kernel or a radial basis function kernel, especially in the cases of high dimensional samples.

The article "Nonlinear dimensionality reduction of data manifolds with essential loops", authored by John Aldo Lee and Michel Verleysen, is based on the preliminary observation that numerous methods or algorithms have been designed to solve the problem of non-linear dimensionality reduction (NLDR), however, very few among them are able to embed efficiently 'circular' manifolds like cylinders or tori, which have one or more essential loops. The paper presents a simple and fast procedure that can tear or cut those manifolds, i.e. break their essential loops, in order to make their embedding in a low-dimensional space easier. The key idea introduced in the manuscript is that, starting from the available data points, the tearing procedure represents the underlying manifold by a graph and then builds a maximum sub-graph with no loops. As the procedure works with graphs, it can preprocess data for all NLDR techniques that make use of the same representation. Recent techniques falling in such category are those making use of geodesic distances, such as ISOMAP, geodesic Sammon's mapping and geodesic curvilinear component analysis, or those based on k-ary neighborhoods, like locally linear embedding, Hessian locally linear embedding and laplacean eigenmaps.

The contribution "Geometric preprocessing, geometric feedforward neural networks and Clifford support vector machines for visual learning", authored by Eduardo Bayro-Corrochano, Refugio Vallejo and Nancy Arana-Daniel, aims at showing the design and use of feed-forward neural networks and the Support Vector Machines in the coordinate-free mathematical system of the Clifford geometric algebra. The Authors compare the McCulloch–Pitts neuron and the geometric neuron. An instance of geometric neuron is the conformal neuron which can be used for RBF networks and Support Vector Machines. The paper presents the generalization of the real- and complex-valued multilayer perceptron to the Clifford-valued multilayer perceptron. The paper studies also the Multivector Support Vector Machines which are SVMs for processing multivectors, for which the Authors design kernels involving Clifford products. The resultant kernel resembles a sort of polynomial kernel using a multivector representation. In the context of SVMs, a contribution of the paper is the generalization of the real- and complex-valued Support Vector Machines classifiers over the hyper-complex numbers. This Clifford-valued Support Vector Machine accepts multiple multivector inputs and behaves as a multi-class classifier. For the pre-processing, the Authors introduce a geometric method based on Clifford moments. This method is applied together with geometric MLPs for tasks related to 2D pattern recognition. The experimental part of the manuscript shows applications of Support Vector Machines using the conformal neuron and Clifford kernels. The Authors include applications of the Clifford SVM classifier for nonlinear separable problems.

Differential geometry, or its offshoot information geometry, has been invoked to understand the dynamics of neuromanifolds and learning dynamics. Natural gradient learning is one example of success from this line of investigation. Also, the theory of Lie groups, which are differentiable manifolds with group properties of algebraic operations, have recently gained increasing interest in the neurocomputing community. In studying these groups, it is possible to elucidate other structures in which group operation plays an important role: Grassmann manifolds and Stiefel manifolds are examples of structures that were derived from the Lie group structure and that contribute significantly to linear and nonlinear systems as well as independent component analysis (ICA) in signal processing.

With reference to these topics, the present special issue accommodates some contributions.

The manuscript "Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold", authored by Yasunori Nishimori and Shotaro Akaho, extends the natural gradient method for neural networks to the case in which the weight vectors are constrained to the Stiefel manifold. The proposed method involves integration techniques of the gradient flow that do not violate the manifold constraints, based on geodesics. The Authors formulate the previously proposed natural gradient and geodesics on the manifold exploiting the fact that the Stiefel manifold is a homogeneous space transitively acted upon by the orthogonal group. On the basis of it, the Authors develop a simpler updating rule and a one-parameter family of generalized updating rules. The effectiveness of the proposed methods is validated by experiments on minor subspace analysis and independent component analysis.

The manuscript "Tools for application-driven dimension reduction", authored by Anuj Srivastava and Xiuwen Liu, observes that simplicity and efficiency of linear transformations make them a popular tool for extracting features and reducing dimensions of data before or during statistical analysis. This is relevant in applications involving image compression and reconstruction, discriminant analysis, pattern classification, and image or text retrieval. Linear transformations with natural orthogonality constraints can be represented as elements of the Stiefel and Grassmann manifolds. The Authors advocate that the choice of a transformation for dimension reduction is not standard, but it is dictated by the application at hand and by the data set, and can be formulated as an optimization problem on the above-mentioned manifolds. The authors demonstrate this idea by deriving dimension-reducing transformations in applications such as image-based object recognition and content-based image retrieval.

The paper "Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras", authored by Mark Plumbey, explores the use of geometrical methods to tackle the non-negative independent component analysis problem. The Author concentrates on methods based on the minimization of a cost function over the space of orthogonal matrices. The paper recalls the idea of the Lie group of special orthogonal matrices that it is wished to search over, and explains how this is related to the Lie algebra of skew-symmetric matrices. The Author describes how familiar optimization methods such as steepest-descent and conjugate gradients can be transformed into such Lie-group setting and how the Newton update step has an alternative Fourier version in the special orthogonal group. Finally, the Author introduces the concept of a toral subgroup generated by a particular element of the Lie group or Lie algebra, and explore how this commutative subgroup might be used to simplify searches on the constraint surface.

Over the last decade or so, driven greatly by the work on information geometry, we are seeing the merging of the fields of statistics and geometry applied to neural network, learning and biological modeling. An example of the latter topic is recognized in the development of “tensor network theory of the cerebellum”. From the theoretical side, the functioning of the cerebellum was given a possible mathematical explanation both in its structuro-functional properties (sensorimotor metric tensor of the spacetime manifold) and in its physiological and pathological growth.

Topology and differential forms are among the other ingredients of geometry that have started to influence neurocomputing. Algebraic geometry, among the most abstract of mathematical theories, has also been successfully applied to the study of neural networks to elucidate their learning capabilities since neuromanifolds include algebraic singularities.

The present special issue accommodates the contribution “Algebraic geometry of singular learning machines and symmetry of generalization and training errors”, authored by Sumio Watanabe, which is based on the observation that several hierarchical learning machines, such as neural networks and normal mixtures, are singular learning machines, for which the likelihood function can not be approximated by any quadratic form, therefore the conventional statistical theory does not hold for them. The paper proves the symmetry property of the generalization and training errors based on an algebraic-geometrical method. In particular, a new parameterization is introduced by applying the resolution of singularities. Then, the asymptotic behavior of the likelihood function is clarified based on the empirical process theory. In conclusion, the asymptotic forms of the generalization and training errors are derived. The net result of the proposed study is a mathematical foundation of model selection and hypothesis testing in singular learning machines.

Neural networks are widely used as flexible models for regression and classification applications, but questions remain open about how neural networks can be safely used when training data is limited. Bayesian learning for neural networks shows for instance that Bayesian methods allow complex neural network models to be used without fear of the over-fitting that can occur with traditional neural network learning methods. Also, conventional training methods for multilayer perceptrons can be interpreted in statistical terms as instances of maximum likelihood estimation. In this theoretical setting, the idea is to find a single set of parameters for the network that maximizes the fit to the training data. Insights into the nature of these complex Bayesian models may be gained by a theoretical investigation of the priors over functions that underlie them. Both the theoretical and computational aspects of this work are of wide interest in the neurocomputing and machine learning area, as they can contribute to a better understanding of how Bayesian methods can be applied to complex problems.

The present issue accommodates the article “The geometry of prior selection", authored by Hichem Snoussi, which is devoted to the selection of prior in a Bayesian learning framework. There is an extensive literature on the construction of non-informative priors and the subject seems far from a definite solution. The Author considers this problem with information geometric tools. The differential geometric analysis allows the formulation of the prior selection problem in a general manifold-valued set of probability distributions. In order to construct the prior distribution, the Author proposes a criterion expressing the trade-off between the decision error and the uniformity constraint. The solution has an explicit expression obtained by variational calculus. In addition, the proposed solution has two important invariance properties: Invariance to the dominant measure of the data-space and also invariance to the parameterization of a restricted parametric manifold. The Author shows how the construction of a prior by projection is the best way to take into account the restriction to a particular family of parametric models. For instance, this procedure is applied to autoparallel restricted families. Two practical examples given in the paper illustrate the proposed construction of prior: The first example deals with the learning of a mixture of multivariate Gaussians in a classification perspective. The Author shows in this learning problem how the penalization of likelihood by the proposed prior eliminates the degeneracy occurring when approaching singularity points. The second example concerns the blind source separation problem.

The present issue also accommodates the contribution "Lattice Duality: The origin of probability and entropy", authored by Kevin Knuth, which begins with the observation that Bayesian probability theory is an inference calculus that originates from a generalization of inclusion on the Boolean lattice of logical assertions to a degree of inclusion represented by a real number. Dual to this lattice is the distributive lattice of questions constructed from the ordered set of down-sets of assertions, which forms the foundation of the calculus of inquiry - a generalization of information theory. In this paper, the Author introduces this novel perspective on these spaces in which machine learning is performed and discusses the relationship between these results and several proposed generalizations of information theory in the literature.

The interest displayed by the scientific community into research topics related to the geometry of neural networks and of learning machines is also testified by several academic activities, among which we wish to mention:

- The special issue on "Non-Gradient Learning Techniques" of the International Journal of Neural Systems (guest editors A. de Carvalho and S.C. Kremer).
- The Post-NIPS*2000 workshop on "Geometric and Quantum Methods in Learning", organized by S.-i. Amari, A. Assadi and T. Poggio (Colorado, December 2000).
- The workshop "Uncertainty in geometric computations" held in Sheffield,

England, in July 2001, organized by J. Winkler and M. Niranjan (University of Sheffield, UK).
- The special session of the 2002 International Joint Conference on Neural Networks Conference dedicated to “Differential & Computational Geometry in Neural Networks” (session chair: E. Bayro-Corrochano) held in Honolulu, Hawaii (USA), in May.
- The workshop “Information Geometry and its Applications”, held in Pescara (Italy), in July 2002, organized by P. Gibilisco.
- The special session of the 13$^{\text{th}}$ European Symposium on Artificial Neural Networks dedicated to “Dynamical and Numerical Aspects of Neural Computing”, to be held in Bruges (Belgium) in April 2005, organized by M. Atencia.

We believe the time is ripe for establishing geometrical theories of neurocomputing. Many of the scattered topics and theories that have recently emerged have been gathered here, in a special issue on geometrical theories applied to neural networks and learning, to kick off this effort. In particular, Neurocomputing journal dedicates a special issue to the theory and advanced applications of geometric concepts to neural learning and optimization, bringing together contributions well founded in modern mathematics.

While some of the topics presented in the special issue might seem disconnected or fragmented, they are the seeds that will hopefully sow the geometrical theories for neurocomputing. The Readers will have for the first time a collection of approaches including differential geometrical methods for learning, the Lie group learning algorithms, the natural (Riemannian) gradient techniques, learning by weight flows on Stiefel-Grassmann manifolds, the theories for learning on orthogonal group, theories for neurocomputing based on Clifford geometric algebra, the numerical aspects of the solution of the matrix-equations on Lie groups arising in neural learning/optimization and related topics, along with applications of differential geometry to Bayesian learning and reasoning.

As a concluding remark, we wish to express our gratitude to the following people, without whom this special issue could not have been born:

- The journal Editor, Tom Heskes, for enthusiastically approving the special issue proposal and for his constant support during all the phases of its preparation.
- All the Authors of submitted papers for giving us the opportunity to consider their valuable contributions.
- All the Reviewers we got in touch with during the preparation of the special issue, who performed quick and thorough reviews and made an excellent cooperation job.

Some phases of the preparation of this issue were managed when S. Fiori was visiting the Mathematical Neuroscience Laboratory at the Brain Science Institute of RIKEN in Japan (in March 2004 and later on in August 2004). S. Fiori wishes to gratefully thank the Brain Science Institute director, Prof. S.-i. Amari, as well as all the laboratory members, for the warmest and kindest hospitality.

November 25, 2004

Simone Fiori and Shun-ichi Amari