A Neural Stiefel Learning based on Geodesics Revisited

Yasunori Nishimori
Neuroscience Research Institute
AIST Central 2, Umezono 1-1-1, Tsukuba
Japan 305-8568

Abstract: - In this paper we present an unsupervised learning algorithm of neural networks with p inputs and m outputs whose weight vectors have orthonormal constraints. In this setting the learning algorithm can be regarded as optimization posed on the Stiefel manifold, and we generalize the natural gradient method to this case based on geodesics. By exploiting its geometric property as a quotient space: homogeneous space, the previous result [11] for the case of the orthogonal group can be used to derive the algorithm. Relevant as well as possible applications of the geometry of homogeneous spaces are also suggested.

Key-Words: - natural gradient method, Riemannian metric, geodesic, Lie group, Stiefel manifold, Riemannian submersion, shape theory, unsupervised neural learning

1 Introduction

Recently researchers in neural networks and machine learning have been aware of the importance of considering geometric structures intrinsically hidden in data sets or the parameter sets they need to tune for solving learning problems. In neural networks community this trend is triggered by the seminal work by Amari [1] in information geometry and the natural gradient method. On the other hand, researchers in control theory and numerical analysis have been working on the gradient and the Hamiltonian flows on manifolds [2], and these several communities seem to have converged a common set of manifolds: homogeneous spaces. Among others, examples we most frequently encounter are the orthogonal group and its generalizations: like the Stiefel, the Grassmann and the generalized flag manifold. Indeed many matrix-factorization, component-analysis problems can be formulated as optimization of some cost functions on these manifolds, such as PCA, SVD, ICA, minor component analysis, multidimensional ICA, principle subspace analysis and others.

Some researchers have developed optimization methods on these manifolds for this decade; we mention here a historical note on the previous contributions to the problem, based on geodesics. Geometric properties of geodesics on homogeneous spaces have been investigated by mathematicians since many years ago, and yet it is relatively recent they have been implemented by computers. In the early nineties Smith proposed to use geodesics for generalizing standard iter-

ative optimization methods like the gradient descent, the conjugate gradient, and the Newton method to the cases problems are posed on general manifolds [13], however, his paper stressed theoretical aspects related to convergence and explicit forms of updating rules by using the matrix components of the points on the Stiefel, Grassmann manifolds had not been proposed until the study by Edelman [3]. The present author, stimulated by Amari's natural gradient method, proposed a learning algorithm via geodesic flows on the orthogonal and the Stiefel manifold independently of Edelman [11]. Though our mehod is based on the same geodesics on the Stiefel manifold, however, the implementation: the updating rule, is totally different; we fully take advantage of the fibration-structure of the manifold, while Edelman's one is based on a direct solution of a variational problem(See Theorem 2.1 and Corollary 2.2 in [3].) Moreover, as far as I know, the method was applied to ICA for the first time in [11], and it was successfully applied to a non-negative generalization of ICA as well by Plumbley [12]. Also closely related Fiori's algorithm based on a different framework: rigid body dynamics, which yields comparable performance with ours, supports the effectiveness and the naturalness of our method.

Despite its importance beyond optimization, the geometry of homogeneous spaces is still not well known among neural networks community, so in this paper we give a detailed explanation of the geodesic-based learning algorithm on the Stiefel manifold, which is only briefly sketched in [11].

2 Learning via geodesic flows

In this paper we consider the geometry of the following set of matrices,

$$St(p,m,\mathbb{R}) = \{W \in M(p,m) \mid W^tW = I_p\}, \quad (1)$$

where $M(p,m)$ denotes the set of $p \times m$ matrices, I_p p-dimensional identity matrix, and we assume $p \geq m$[1]. By using the column vectors $w_1,w_2,...,w_m \in \mathbb{R}^p$ of W, the relation reduces to $\frac{m(m+1)}{2}$ orthogonal and normalization constraints, $(w_i\ ,\ w_j) = \delta_{ij}$, where δ_{ij} denotes the Kronecker delta. These equations define a submanifold in $p \times m$-dimensional Euclidean space $\mathbb{R}^{p \times m}$ and this manifold is termed the Stiefel manifold; its particular case $(p = m)$ is called the orthogonal group $O(p)$.

$$O(p) = \{M \in M(p) \mid M^tM = I_p\} \quad (2)$$

We assume column vectors w_i, $(i = 1\ ,\ ..\ ,\ m)$ represent the weight vectors of a one layer linear neural network with p inputs, and m outputs throughout this paper. Before describing the geometry of $St(p\ ,\ m\ ,\ \mathbb{R})$ we review here the previous result [11] needed for getting an learning algorithm on the Stiefel manifold. We generalized the ordinary natural gradient method, and its discretization (3), (4) to the case where W_n is constrained to $O(p)$ (5), (6).

$$\frac{dW(t)}{dt} = -\mu \text{grad}\ f(W(t)) \quad (3)$$

$$W_{n+1} = W_n - \mu \text{grad}\ f(W_n) \quad (4)$$

$$W_{n+1} = W_n \exp(-\mu W_n^t \text{grad}\ f(W_n)) \quad (5)$$

$$= W_n \exp \eta \left\{ \nabla f(W_n)^t W_n - W_n^t \nabla f(W_n) \right\}, \quad (6)$$

where $\eta = \frac{1}{2}\mu$ is a learning constant, f is a cost function, $\nabla f(W_n)$ denotes an ordinary Euclidean gradient $(= \frac{\partial f(W_n)}{\partial (W_n)_{ij}})$, and grad $f(W_n)$ the natural gradient [1] with respect to the bi-invariant Riemannian metric: $g_{O(p)}(X,Y) = \text{tr}X^tY$, for all $X,Y \in T_{W_n}O(p)$,

[1] We use $p \times m$ matrices instead of $m \times p$ ones [11] $(p \geq m)$ for easily compared to the other authors' formulation.

and all $W_n \in O(p)$. Since we usually discretize (3) as (4), which corresponds to approximating the integral curve of the gradient flow (3) by a short straight line, the approximation to the integral curve by a geodesic (5) is geometrically very appealing, because a geodesic is a counterpart of a straight line in Riemannian manifolds. The final expression (6) exhibits the geodesic which emanates from W_n pointing to $-\mu \mathrm{grad}\, f(W_n) \in T_{W_n} O(p)$ as a velocity vector.

3 Neural Stiefel Learning

The Stiefel manifold belongs to a family of manifolds: called homogeneous spaces. Since homogeneous spaces is a very useful concept for analyzing various matrix-factorization, component-analysis problems, we present here its geometric property in detail, and based on it derive a geodesic-based neural learning algorithm on the manifold. First we describe the quotient space structure of the Stiefel manifold. Because the structure is rather simple, geodesics of the Stiefel manifold can be expressed by the orthogonal group acting on them.

Note the p-dimensional orthogonal group $G = O(p)$ acts the Stiefel manifold as follows.

$$(M,W) \mapsto MW \text{ (matrix multiplication)}, \quad (7)$$

where $M \in O(p)$, $W \in St(p\,, m\,, \mathbb{R})$. For every given two points $W_0\,, W_i \in St(p\,, m\,, \mathbb{R})$, there exists an element $M_i \in H$ such that $W_i = M_i W_o$. The action of G on the Stiefel manifold is called transitive if it satisfies the above condition. Therefore starting from a given point $W_0 \in St(p, m, \mathbb{R})$, we can reach any point W_i by the G-action. This means G-orbit $G(W_o)$ of a given point W_0 coincides with the whole Stiefel manifold, where $G(W_o) = \{W \in St(p,m,\mathbb{R}) | W = MW_o, M \in O(p)\}$. Because of this surjectivity we can represent every element of $St(p\,, m\,, \mathbb{R})$ using some orthogonal matrix.

$$\text{Correspondence: } M \mapsto W \quad (8)$$

Actually this correspondence is many to one, and the redundancy is described by so called the isotropy subgroup H_{W_o} of G. H_{Wo} is a set of matrices which does not change W_o by the above multiplication.

$$H = \{N \in O(p) \mid NW_o = W_o\} \quad (9)$$

We assume hereafter

$$W_o = \begin{pmatrix} I_m \\ o_{(p-m,m)} \end{pmatrix} = (\ e_1 \quad ,.., \quad e_m\ )$$
$$\text{where } e_i = (0\,, ..\,, \overset{i}{\overset{\vee}{1}}\,, ..0)^t \in \mathbb{R}^p. \quad (10)$$

Then, from the conditions $Ne_i = e_i$, $(i = 1\,, ..\,, m)$, H can be represented as follows.

$$H = \left\{ \begin{pmatrix} I_m & O_{(m,p-m)} \\ O_{(p-m,m)} & U \end{pmatrix} \right.$$
$$\left. \text{where } U \in O(p-m) \right\}, \quad (11)$$

where $U \in O(p-m)$, and $O_{(q,r)}$ denotes $q \times r$ zero matrix. Two orthogonal matrices M_1, M_2 represent the same point on the Stiefel manifold if and only if the first m column vectors coincide, in other words,

$$\exists N \in H\,, \text{ s.t. } M_2 = M_1 N. \quad (12)$$

Namely the Stiefel manifold is a quotient space: $G = O(p)$ divided by the ambiguity arising from the isotropy subgroup $H = O(p-m)$. We usually express this relation as

$$St(p,m,\mathbb{R}) \simeq G/H \simeq O(p)/O(p-m), \quad (13)$$

and we hereafter use a representative orthogonal matrix of the above equivalence class to describe a point on the Stiefel manifold: $\widetilde{W} =$

$(w_1,..,w_m,v_1,..,v_{p-m}) \in O(p)$ represents $W = (w_1\,,..\,,w_m) \in St(p\,,m\,,\mathbb{R})$, where $v_1\,,..\,,v_{p-m}$ form complementary orthonormal frames.

To get a learning algorithm via geodesic flows we need to derive the geodesic emanating from a given point W with a specified velocity, tangent vector V. The previous result for the orthogonal group [11] is directly applicable with the aid of the property of Riemannian submersion, which we explain below.

Firstly the natural projection p from $O(p)$ to $St(p\,,m\,,\mathbb{R})$ is a submersion in the sense that the tangential map: $dp|_{\tilde{W}} : T_{\tilde{W}}O(p) \to T_W St(p\,,m\,,\mathbb{R})$ is surjective, where

$$\begin{array}{rccc} p: & O(p) & \to & St(p\,,m\,,\mathbb{R}) \\ & \cup & & \cup \\ & \tilde{W} & \mapsto & p(\tilde{W}) = W \end{array} \tag{14}$$

Because p is a linear map, $dp|_{\tilde{W}}$ is identical with p.

$$p(W) = W \begin{pmatrix} I_m \\ O_{(p-m,m)} \end{pmatrix} = (w_1,..,w_m) \tag{15}$$

$$dp|_{\tilde{W}}(X) = p(X) = X \begin{pmatrix} I_m \\ O_{(p-m,m)} \end{pmatrix}, \tag{16}$$

where $X \in T_{\tilde{W}}O(p)$. For each submersion p, we can decompose $T_{\tilde{W}}O(p)$ into the vertical subspace $V_{\tilde{W}}$ and the horizontal subspace $H_{\tilde{W}}$. The vertical subspace $V_{\tilde{W}}$ is the subset of $T_{\tilde{W}}O(p)$ defined by the kernel of the tangential linear map.

$$V_{\tilde{W}} = \{v \in T_{\tilde{W}}O(p) | dp|_{\tilde{W}}(v) = O_{p\times m}\} \tag{17}$$

And we define the horizontal subspace $H_{\tilde{W}}$, which is orthogonal to the vertical subspace with respect to the standard Riemannian metric $g_{O(p)}$ on $O(p)$.

$$H_{\tilde{W}} = \big\{u \in T_{\tilde{W}}O(p) | g_{O(p)}\,(u\,,v) = 0 \quad \text{for all } v \in V_{\tilde{W}}\big\} \tag{18}$$

Notice this decomposition corresponds to the following orthogonal direct sum decomposition of the Lie algebra with respect to the killing metric (identical with $g_{O(p)}$).

$$\mathfrak{g} = \mathfrak{h} + \mathfrak{m}, \tag{19}$$

where $\mathfrak{g}$ and $\mathfrak{h}$ denotes the Lie algebras of $O(p)$, and $O(p-m)$ respectively.

$$\mathfrak{g} \simeq \{\text{the set of } p\text{- dim . SSMs}\} \tag{20}$$

$$\mathfrak{h} \simeq \left\{ \begin{pmatrix} O_{(m,m)} & O_{(m,p-m)} \\ O_{(p-m,m)} & C \end{pmatrix} \middle| \; C{:}(p-m)\text{- dim . SSM} \right\} \tag{21}$$

$$\mathfrak{m} \simeq \left\{ \begin{pmatrix} A & -B^t \\ B & O_{(p-m,p-m)} \end{pmatrix} \middle| \; A{:}\ m\text{- dim . SSM, } B{:}\ \text{arbitrary} \right\}; \tag{22}$$

$$\mathrm{tr}\left\{ \begin{pmatrix} O_{(m,m)} & O_{(m,p-m)} \\ O_{(p-m,m)} & C \end{pmatrix}^t \cdot \begin{pmatrix} A & -B^t \\ B & O_{(p-m,p-m)} \end{pmatrix} \right\} = 0 \tag{23}$$

Since $\mathfrak{g}$ is considered as the tangent space of $O(p)$ at I_p, and left multiplication of $\tilde{W}$ is an isometry of $O(p)$, $V_{\tilde{W}}$ and $H_{\tilde{W}}$ can be represented as follows

$$V_{\tilde{W}} = \left\{ \tilde{W} \begin{pmatrix} O_{(m,m)} & O_{(m,p-m)} \\ O_{(p-m,m)} & C \end{pmatrix} \middle| \; C{:}\ (p-m)\text{- dim . SSM} \right\}, \tag{24}$$

$$H_{\tilde{W}} = \left\{ \tilde{W} \begin{pmatrix} A & -B^t \\ B & O_{(p-m,p-m)} \end{pmatrix} \middle| \; A{:}\ m\text{- dim . SSM, } B{:}\ \text{arbitrary} \right\}, \tag{25}$$

where SSM denotes skew symmetric matrix. Using the decomposition defined in a similar fashion for a general manifold, we can define a Riemannian submersion. A submersion $f : M \to N$ is called a Riemannian submersion if

$$df|_p : H|_p \to T_{f(p)}N \tag{26}$$

is an isometry at every points $p \in M$, where $H|_p$ denotes the horizontal subspace of T_pM.

It is a crucial observation that the projection p from $O(p)$ to $St(p\,,\,m\,,\,\mathbb{R})$ becomes a Riemannian submersion. Or it might be better to say that we equip the Stiefel manifold such a metric with which the projection becomes a Riemannian submersion. Such a metric g_{St} is called the normal homogeneous metric [7]. Since p as well as dp is a many-to-one map, we cannot get a unique inverse image of $dp|_{\tilde{W}}$, so for a given $W \in St(p\,,\,m\,,\,\mathbb{R})$ and $X\,,\,Y \in T_WSt(p\,,\,m\,,\,\mathbb{R})$, take any one of the inverse image $\tilde{W} \in p|^{-1}(W)$, $\tilde{X} \in dp^{-1}(X)$, $\tilde{Y} \in dp^{-1}(Y)$, then decompose $\tilde{X}$, $\tilde{Y}$ into the vertical and the horizontal components: $\tilde{X} = \tilde{X}_V + \tilde{X}_H$, $\tilde{Y} = \tilde{Y}_V + \tilde{Y}_H$. Based on this decomposition, the normal homogeneous metric is defined as follows,

$$\begin{aligned} g_{St}(X\,,\,Y) &\equiv g_{O_{(p)}}(\tilde{X}_H\,,\,\tilde{Y}_H) \\ &= \mathrm{tr}\left\{(\tilde{X}_H)^t\tilde{Y}_H\right\}. \end{aligned} \tag{27}$$

It is easy to show this definition does not depend on the way which inverse images $\tilde{W}$, $\tilde{X}$ are picked up. As is shown in (24), (25), every inverse image of X is expressed as

$$\begin{aligned} &\tilde{W}\begin{pmatrix} A & -B \\ B & C \end{pmatrix} \\ &= \tilde{W}\begin{pmatrix} O_{(m,m)} & O_{(m,p-m)} \\ O_{(p-m,m)} & C \end{pmatrix} \\ &\quad + \tilde{W}\begin{pmatrix} A & -B \\ B & O_{(p-m,p-m)} \end{pmatrix}. \end{aligned} \tag{28}$$

It follows that

$$\tilde{X}_H = \tilde{W}\begin{pmatrix} A_1 & -B_1^t \\ B_1 & O_{(p-m,p-m)} \end{pmatrix}, \tag{29}$$

and together with the assumption $p(\tilde{X}_H) = X$, we get

$$\tilde{W}\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = X \Leftrightarrow \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \tilde{W}^tX. \tag{30}$$

Namely once we choose an inverse image $\tilde{W}$ of p, the horizontal lift of the tangent vector is uniquely determined.

$$\begin{aligned} &g_{St}(X\,,\,Y) \\ &= \mathrm{tr}\left[\left\{\tilde{W}\begin{pmatrix} A_1 & -B_1^t \\ B_1 & O_{(p-m,p-m)} \end{pmatrix}\right\}^t \cdot \right. \\ &\qquad \left.\left\{\tilde{W}\begin{pmatrix} A_2 & -B_2^t \\ B_2 & O_{(p-m,p-m)} \end{pmatrix}\right\}\right] \\ &= \mathrm{tr}\left\{\begin{pmatrix} A_1 & -B_1^t \\ B_1 & O_{(p-m,p-m)} \end{pmatrix}^t \cdot \right. \\ &\qquad \left.\tilde{W}^t\tilde{W}\begin{pmatrix} A_2 & -B_2^t \\ B_2 & O_{(p-m,p-m)} \end{pmatrix}\right\} \\ &= \mathrm{tr}(A_1^tA_2 + B_1^tB_2^t + B_1B_2^t) \\ &= \mathrm{tr}\left\{\begin{pmatrix} A_1 \\ B_1 \end{pmatrix}^t\begin{pmatrix} A_2 \\ B_2 \end{pmatrix}\right\} + \mathrm{tr}\left(B_1B_2^t\right) \\ &= \mathrm{tr}\left(X^tY\right) + \mathrm{tr}\left(B_1B_2^t\right), \end{aligned} \tag{31}$$

where $X\,,\,Y \in T_pSt(p\,,\,m\,,\,\mathbb{R})$.

The procedure to get a geodesic which starts from $W \in St(p\,,\,m\,,\,\mathbb{R})$ pointing to the direction $V \in T_WSt(p\,,\,m\,,\,\mathbb{R})$ goes as follows. First embed the W to $O(p)$ by adding $(p-m)$ orthogonal column vectors $(v_1\,,\,..\,,\,v_{p-m})$,

$$W \to \tilde{W} = (w_1\,,\,..\,,\,w_m\,,\,v_1\,,\,..\,,\,v_{p-m}), \tag{32}$$

then lift the velocity vector V to the horizontal vector subspace of $T_{\tilde{W}}St(p\,,\,m\,,\,\mathbb{R})$. $V \mapsto \tilde{V}_H$ based on (29). Recall, by the previous result [11], we obtained the geodesic on $O(p)$ starting from $\tilde{W}$ with $\tilde{V}_H$ as follows.

$$\tilde{c}(t) = \tilde{W}\exp(tW^t\tilde{V}_H). \tag{33}$$

According to a beautiful property of Riemannian submersion [7], the geodesic on the homogeneous space is described by projecting

this lifted geodesic on the orthogonal group $O(p)$ to $St(p\,,m\,,\mathbb{R})$ again. Thus we get the final expression of the geodesic

$$\tilde{c}(t) = \tilde{W}\exp(t\tilde{W}^t\tilde{V}_H)\begin{pmatrix} I_m \\ O_{(p-m,m)} \end{pmatrix}, \quad (34)$$

and this yields the learning algorithm via geodesic flows on the Stiefel manifold:

$$W_{n+1} = \tilde{W}_n\exp(-\eta\tilde{W}_n^t\tilde{V}_H)\begin{pmatrix} I_m \\ O_{(p-m,m)} \end{pmatrix}; \quad (35)$$

$$\tilde{V}_H = \mathrm{grad}f(W_n) = \frac{1}{2}\left\{\nabla f(W_n) - W_n\nabla f(W_n)^t W_n\right\}, \quad (36)$$

where f is the cost function on $St(p\,,m\,,\mathbb{R})$ to be minimized, η is a learning constant.

4 Connections of homogeneous spaces to other problems

The geometry of homogeneous spaces and the orbit method give deep insights not only to optimization but also into computer vision and pattern recognition. We suggest some connections here. First example is statistical theory of shape due to Kendall [8]. He showed that a configuration of triangle landmark points in the plane can be regarded as a point on homogeneous space like the complex projective space. Secondly, in the plenary lecture in ICM 2002 [9], Mumford discussed possible applications of geodesics on some homogeneous spaces arising from the infinite dimensional diffeomorphism group, to mathematical theory of shape. Thirdly, Fukumizu [6] showed some convergence property of multilayer perceptorn is heavily affected by the geometric structure of the neural manifold. The groups acting on the neural manifold are not Lie groups, instead finite groups such as the permutation group, therefore singularities appear in the manifold after devided by the groups unlike homogeneous spaces. Moreover, in Murase's eigenspace method [10] and learning a manifold structure in high dimensionl data sets e.g. [15], transformations cannot be described by Lie groups as we used in this paper, and yet the manner they try to grasp a set of some configurations, (human faces or words in documents or whatever) divided by redundancies, as a manifold is very close in the spirit to homogeneous spaces, and so one of the most ambitious future plan shall be to make a more subtle homogeneous space theory directly applicable to represent configurations of various patterns in the real world.

5 Conclusion

In this paper we described a generalization of the natural gradient method to the case parameters are constrained to the Stiefel manifold based on the geometry of homogeneous space. It is a very natural method from a geometrical point of view and therefore has nice properties such as equivariance when it is applied to ICA, also several numerical simulations have validated its effectiveness [11], [12]. However, from a computational complexity point of view our method is still demanding because of the computation of the matrix exponential. Through circumventing this difficulty, recent discretization and integration methods of differential equations on manifolds may shed a new light on our method. Other direction of future work among others shall be to generalize the method to the case posed on the generalized flag manifold

for a possible application to multidimensional ICA. Historically matrix-factorization problems have been closely related to so called the orbit method in Lie group theory, and the generalized flag manifold plays a key role there, however, we still have not observed a geometric learning algorithm on the manifold in the scientific literature.

References:

[1] S. Amari, Natural gradient works efficiently in Learning *Neural Computation*, **10**, pp.251-276, 1998.

[2] R.W. Brockett, Differential Geometry and the Design of Gradient Algorithms, *Proceedings of Symposium in Pure Mathemtics* **54** (1), pp.69-92, 1993.

[3] A. Edelman, T.A. Arias, and S.T. Smith, The Geometry of algorithms with orthogonality constraints, *SIAM Journal on Matrix Analysis and Applications*, **20**(2), pp.303-353, 1998.

[4] S. Fiori, A theory for learning by weight flow on Stiefel-Grassman manifold, *Neural Computation*, **13**(7), pp.521-531, 2001.

[5] S. Fiori, A theory for learning based on rigid bodies dynamics, *IEEE Transactions on Neural Networks*, **13**(3), pp. 521-531, 2002.

[6] K. Fukumizu, Geometry of neural networks: Natural gradient for learning, *Handbook of Biological Physics*, **4**: *Neuro-informatics and Neural Modelling*, pp.731-769. Elsevier 2001.

[7] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, Springer Verlag, 1990.

[8] D. G. Kendall, Shape manifolds, procrustean metrics and complex projective spaces, *Bull. London Math. Soc.*, **16**, pp. 81–121, 1984.

[9] D. Mumford, Pattern theory: The mathematics of perception, *Proceedings of ICM 2002*, **1**, pp. 81–121, 2002.

[10] H. Murase, S K. Nayar, Visual learning and recognition of 3-D objects from appearance, *International Journal of Computer Vision*, **14**, pp.5-24, 1995.

[11] Y. Nishimori, Learning Algorithm for Independent Componetnt Analysis by Geodesic Flows on Orthogonal Group, *Proceedings of International Joint Conference on Neural Networks (IJCNN1999)*, **2**, pp.1625-1647, 1999.

[12] M. D. Plumbley, Algorithms for nonnegative independent component analysis. *IEEE Transactions on Neural Networks*, **14**(3), pp.534-543, 2003.

[13] S. T. Smith, Optimization Technique on Riemannian Manifolds, *Field Institute Communications,* **3**, pp.113-136, 1994.

[14] Y.B. Suris, *The Problem of Integrable Discretization : Hamiltonian Approach* , Birkhäuser, 2003.

[15] J. B. Tenenbaum, V. De Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* **290** (5500), 2000.