UDK 519.23

ESTIMATION OF SMOOTH VECTOR FIELDS ON MANIFOLDS BY OPTIMIZATION ON STIEFEL GROUP

E.N. Abramov[a] , Yu.A. Yanovich[b,c,a]

[a]*National Research University Higher School of Economics, Moscow, 101000 Russia*
[b]*Skolkovo Institute of Science and Technology, Moscow, 143026 Russia*
[c]*Kharkevich Institute for Information Transmission Problems,*
Russian Academy of Sciences, Moscow, 127051 Russia

Abstract

Real data are usually characterized by high dimensionality. However, real data obtained from real sources, due to the presence of various dependencies between data points and limitations on their possible values, form, as a rule, form a small part of the high-dimensional space of observations. The most common model is based on the hypothesis that data lie on or near a manifold of a smaller dimension. This assumption is called the manifold hypothesis, and inference and calculations under it are called manifold learning.

Grassmann & Stiefel eigenmaps is a manifold learning algorithm. One of its subproblems has been considered in the paper: estimation of smooth vector fields by optimization on the Stiefel group. A two-step algorithm has been introduced to solve the problem. Numerical experiments with artificial data have been performed.

Keywords: manifold learning, dimensionality reduction, vector field estimation, optimization on Stiefel manifold

Introduction

The concept of Big Data implies not only a large sample size, but also high dimensionality. Thus, real data usually have a very high dimensionality, for example, the dimensionality of digital black and white photos is equal to the number of pixels (up to 10 000 pixels), while brain images obtained per second with functional magnetic resonance imaging have a size about 1.5 million. However, many traditional methods and algorithms become inefficient or simply inoperable for high-dimensional data, and this phenomenon is called the curse of dimensionality [1]. D. Donoho, a famous statistician, said in 2000 at the Conference on Mathematical Challenges of the 21st Century: "We can say with complete confidence that in the coming century high-dimensional data analysis will be a very significant activity, and completely new methods of high-dimensional data analysis will be developed; we just do not know what they are yet" [2].

However, real data obtained from real sources, due to the presence of various dependencies between data points and limitations on their possible values, take, as a rule, a small part of the high-dimensional space of observations. For example, the set of all black and white portraits of human faces with the original dimension of the order of hundreds of thousands has an intrinsic dimension of not more than 100. The consequence of the low intrinsic dimension is the possibility of constructing a low-dimensional parametrization of such data with the minimal loss of information contained in it. Therefore, many algorithms for high-dimensional data processing begin with the problem of dimensionality reduction, which results in low-dimensional descriptions of such data.

The most common model for describing such data is the hypothesis that data lie on or near a manifold $\mathcal{M}$ of a smaller dimension. This assumption is called the manifold hypothesis [3], and inference and calculations under it are called manifold learning [4, 5].

At the moment, many manifold learning algorithms have been developed: IsoMap [6], locally linear embedding [7], local tangent space alignment [8], Laplacian eigenmaps [9, 10], Hessian eigenmaps [11], Grassman & Stiefel eigenmaps (GSE) [12, 13], etc. It is the GSE calculation aspects that are considered in the paper.

1.   Manifold learning. Grassman–Stiefel eigenmaps

1.1.   Manifold learning. Manifold embedding. Manifold learning problem is to find the mapping $h : \mathcal{M} \subset \mathbb{R}^p \to \mathbf{Y} \subset \mathbb{R}^q$ given $\mathbf{X} = (X_1, \ldots, X_N) \in \mathcal{M} \subset \mathbb{R}^p$, such that $\mathbf{Y}$ shows the structure of high-dimensional data, where $\mathcal{M}$ is a manifold with $\dim \mathcal{M} = q$. The points $X_1, \ldots, X_N$ are assumed to be independent identically distributed (iid) random variables from a continuous measure μ with $\mathtt{supp}\,\mu = \mathcal{M}$. In addition, there should exist (and it should be estimated) the reverse map $g \colon \mathbf{Y} \to \mathbb{R}^p$ and the approximation error $\|X - g(h(x))\|$ should be small.

Assuming that the manifold $\mathcal{M}$ can be covered by single coordinate chart $\mathbb{B}$, we can define coordinates $(b_1, \ldots, b_q)$ that represent our manifold. Thus, we can write in $\mathbb{R}^p$: $X = f(b)$, $X' = f(b')$ and for two close points X, X' consider the linear approximation:

$$X' = J_f(b)(b' - b) + X,$$

where $J_f(b)$ is the Jacobian matrix. Thus, we reformulate our problem as follows:

$$\mathbb{E}\|(X' - X) - J_f(b)(b' - b)\|^2 \to \min_{\{b\},\{J_f\}} \qquad (1)$$

for all close points (this notion will be defined further). GSE separates this problem into two parts: the first task is to estimate the matrix $J_f = J_f(f^{-1}(X))$, the second one is to estimate coordinates b for a given J_f. The latter problem is ordinary regression, while the former one is of interest.

1.2.   Local description. Neighborhoods construction. The Jacobian matrix $J_f(\mathbf{b}) = (\partial f_i / \partial b_j)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, defines a tangent space $T_X(\mathcal{M})$ at point $X = f(b)$, where $T_X(\mathcal{M})$ is a q-dimensional linear subspace. The tangent spaces can be considered as elements of the Grassmann manifold $\mathtt{Grass}(p, q)$ consisting of all q-dimensional linear subspaces in $\mathbb{R}^p$.

Every neighborhood of each point can be approximated by the tangent space (1) (linear subspace of $\mathbb{R}^p$) as $\mathcal{M}$ is smooth. So, we should define the "linear" neighborhood. For each sample point $X_n \in \mathbf{X}$, local neighborhood $U(X_n) = \{X_1, \ldots, X_{k(n)}\}$ consisting of k-near sample points is constructed. Typical examples of $U(X_n)$ are $U(X_n, \varepsilon)$, the neighborhood of X_n consisting of sample points that from ε-ball centered at X_n, and $U(X_n, k)$ consisting of k-nearest to the X_n sample points; in our calculations, the latter one is used. These neighborhoods determine "Euclidean" kernels $K(X_i, X_j) = I(X_j \in U(X_i))$, where $I(A)$ is the indicator of event A, i.e., $I(A) = 1$ if A occurred, and $I(A) = 0$ otherwise.

Then, defining these linear subspaces is the standard manifold learning problem, which is solved by principal component analysis (PCA). Thus, at each point we search the directions of maximum variation, defining orthogonal vectors. These vectors are joined in matrix $Q_{PCA} \in \mathbb{R}^{p \times q}$, which is considered to be a projector from the q-dimensional tangent space into $\mathbb{R}^p$. It is worth noting that Q_{PCA} are elements of the Stiefel manifold $\mathtt{Stief}(p, q)$ – the set of all q-frames in $\mathbb{R}^p$: $Q_{PCA}^T Q_{PCA} = I_q$.

1.3. Global description. We construct q-dimensional subspaces at each point X_i as an approximation of the tangent bundle of a differentiable manifold. However, our goal is to set a coordinate system on a manifold. Obviously, the PCA algorithm does not provide coordinated frames. PCA estimation consistency and its rates are obtained, for example, in [14, 15]. Thus, these frames should be rearranged in some tanged fields $H(X_i)$ and $\texttt{Span}(Q_{PCA}(X_i)) = \texttt{Span}(H(X_i))$, where $\texttt{Span}(A)$ is a linear hull of columns $A_1, \ldots, A_m$ of the matrix $A = (A_1, \ldots, A_m)$. Let $v \in \mathbb{R}^{q \times q}$ be some non-degenerate transformation in the tangent. Then, the frame Q_{PCA} is transformed into the vector field $Q_{PCA} \cdot v$ from the same tangent space. Generally, for each point we write:

$$H_i = Q_{PCA}(X_i) \cdot v_i \tag{2}$$

where $v_i = v(X_i)$ is a non-degenerate transformation of q-basis into q-dimensional subspace, $H_i = H(X_i)$ is a vector field at point X_i.

While defining the distance between two vectors by the Euclidean norm $\|.\|_2$, it is natural to set the distance between two vector fields by the Frobenius matrix norm

$$\|H_i - H_j\|_F = \sqrt{\texttt{Tr}\left((H_i - H_j)^T (H_i - H_j)\right)}. \tag{3}$$

Then, to measure the closeness between points (and, because of differentiability, also between tangent spaces), we define the function $K(X_i, X_j) = K_{ij}$. Finally, we face the next optimization problem:

$$\Delta(\mathbf{H}_n) = \sum_{i=1}^{n} K_{ij} \|H_i - H_j\|_F^2 \to \min_{\{v_i\}_{i=1}^n} \tag{4}$$

$$\text{subject to } \forall\, i = 1, \ldots, n$$

$$\texttt{rank}(v_i) = q,$$

$$\texttt{Span}H_i = \texttt{Span}Q_{PCA}(X_i),$$

$$H_i^T H_i = I_q.$$

We have $\{H_i\}_{i=1}^n \subset \texttt{Stief}(p,q)$ as $\texttt{Span}H_i = \texttt{Span}Q_{PCA}(X_i)$ and $H_i^T H_i = I_q$ for $i = 1, \ldots, n$. The set of $\{H_i\}_{i=1}^n$ could be explicitly extended $H(X)$ for all $X \in \mathcal{M}$ [12]. So, the solution of (4) with its extension is a smooth vector fields estimation on manifold by optimization on the Stiefel manifold (group). The problem (4) was introduced in [16], but there was only zero-order heuristic solution. Furthermore, our problem was solved in [17] by the direct iterative procedure. In this paper, we introduce a theoretically motivated one.

The sample $\mathbf{X}$ should be "dense" on $\mathcal{M}$ to get a good global vector field estimation. Such a property is guaranteed with high probability for good enough manifolds [15, 18].

2. Algorithm

Orthogonal matrix $v_i \in O(q)$ rotates the q-frame Q_{PCA} in $\mathbb{R}^p$, such that it finally defines the same subspace. Thus, we get the following variation of functional (4):

$$\Delta = \sum_{i=1}^{n} K_{ij} \|Q_i v_i - Q_j v_j\|_F^2 \to \min_{\{v_i\}, i=1\ldots n} \tag{5}$$

$$s.t.\ v_i^T v_i = I_q,$$

where $Q_i = Q_{PCA}(X_i)$. This is an optimization problem with orthogonal constraints: optimization on Lie group $O(q)$. Nevertheless, the implementation of such optimization

is impossible, because $O(q)$ has two connected components (simply put matrices with determinant $+1$ and -1). Thus, optimization is feasible on special orthogonal group, i.e., proper rotation of q-frame. Yet, such optimization with our PCA-frames is degenerate, because for two infinitely close points X_i, X_j the distance $\|H_i - H_j\|_F$ in (4) will be non-zero. To solve this problem, we must obtain an algorithm defining self-orientation of bases on manifold and, then, defining the orientation on manifold. Let us remind here that the manifold is orientable if there exists the set of maps (U_α, ϕ_α), such that the determinant of the Jacobian matrix $\phi_\alpha \phi_\beta^{-1}$ is positive for each α, β; $U_\alpha \bigcap U_\beta \neq \emptyset$.

Proposition 1. *Let the manifold $\mathcal{M}$ be orientable. $Q_i, Q_j \in \mathbb{R}^{p \times q}$ are the matrices defining q-orthogonal unit vectors in the p-dimensional space for some close points X_i, X_j Then, the condition $\det(Q_i^T Q_j) > 0$ for every close points X_i, X_j sets an oriented atlas of manifold $\mathcal{M}$.*

Condition $\det(Q_i^T Q_j) > 0$ is easy to check for constructed frames (matrices $Q_i^T Q_j$ are $\mathbb{R}^{q \times q}$, q no more than 10). Via this fact, we can make an inductive algorithm defining the orientation on the manifold. Let $A^k = \{X_{i_1}, \ldots, X_{i_k}\}$ be the set of k points, where the right orientation is defined. In the first step, we take random point X_{i_1} and set the right orientation at this point Q_{i_1}; $A^1 = \{X_{i_1}\}$. In the k-th step, we get the set of points $A^k = \{X_{i_1}, \ldots, X_{i_k}\}$. Then, we find the closest point among $\mathbf{X} \backslash A^k$ to this set ($X_{i_{k+1}} = \arg \min\limits_{Z \in \mathbf{X} \backslash A^k} \|Z - Y\|_2$); $Y \in A^k$. Following that, we check the condition $Q_{i_{k+1}}^T Q_j > 0$, where $X_j \in A^k$ is the closest point to $X_{i_{k+1}}$. If $\det(Q_{i_{k+1}}^T Q_j) < 0$, then two columns in $Q_{i_{k+1}}$ should be reversed, otherwise $Q_{i_{k+1}}$ gives the right orientation, in both cases $A^{k+1} = A^k \bigcup \{X_{i_{k+1}}\}$. Finally, we get the whole set of points $(X_1, \ldots, X_N)$ and the set of oriented frames $(Q_1, \ldots, Q_N)$.

By transforming functional (5), consider $Q_i^T Q_i = I_q$, we get the following optimization problem:

$$\Delta = -\sum_{ij}^n K_{ij} \cdot \mathtt{Tr}\left(v_i^T (Q_i^T Q_j) v_j\right) \to \min_{\{v_i\}} \tag{6}$$

A brief scheme of the algorithm can be presented as follows:

A. constructing graph K_{ij}; $K_{ij} = 1$ if X_j is among k-closest points of X_i, otherwise $K_{ij} = 0$;

B. constructing frames Q_{PCA} by PCA at each point of sample;

C. initialization of frames Q_{PCA} – defining orientation of manifold;

D. solving optimization problem (6) by Trust Region Method [19].

3. Numerical experiments

The half of two- and four-dimensional spheres $\mathbb{S}^2_{x_1 > 0} \subset \mathbb{R}^3$, $\mathbb{S}^4_{x_1 > 0} \subset \mathbb{R}^5$ and cylinder $\mathbb{S}^1 \times [0, 1]$ are considered as manifold $\mathcal{M}$. All samples are iid uniformly distributed.

One can see that, as a result of optimization, initialization converges to the ideal value of -2 fast and vector fields are visually smooth. The $\widetilde{\Delta}$ values without optimization are higher and the vector fields are not smooth.

Sample generation and steps 1–3 of the algorithm are implemented in R with the standard packages. Step 4 of optimization is performed in Matlab using the Manopt package for optimization on manifolds [20]. The derivatives and the Hessian matrix were theoretically calculated for the functional (6) and were used as a part of input
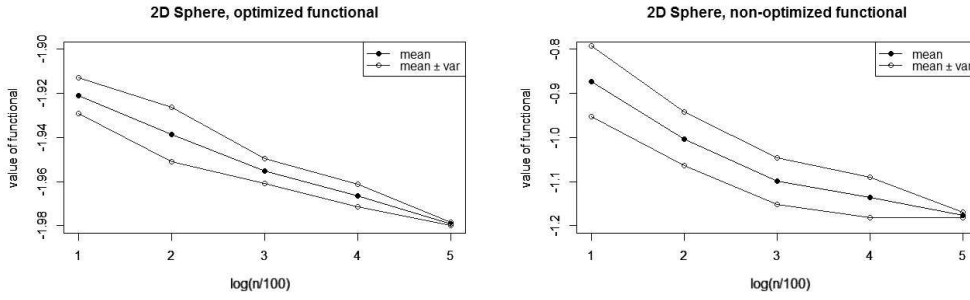
Fig. 1. $\widetilde{\Delta}$ as a function of $\log_{10}(n/100)$ for the 2D sphere
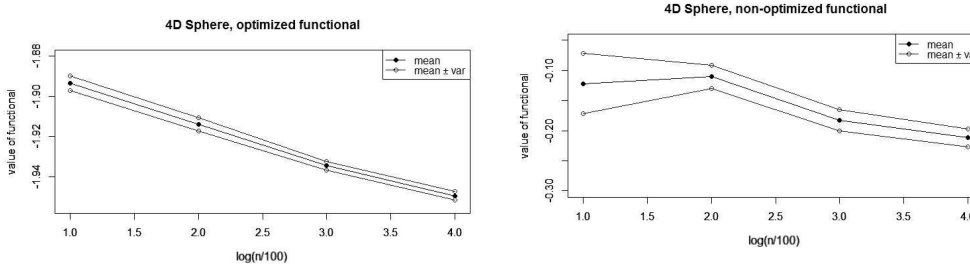

Fig. 2. $\widetilde{\Delta}$ as a function of $\log_{10}(n/100)$ for the 4D sphere
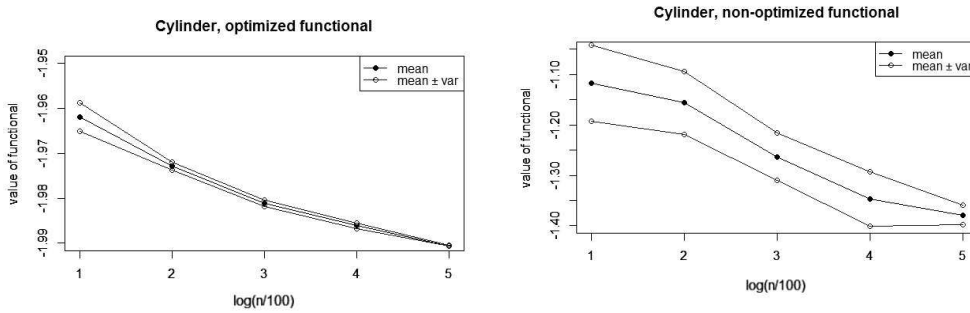

Fig. 3. $\widetilde{\Delta}$ as a function of $\log_{10}(n/100)$ for the cylinder

for manopt. The functional (6) was scaled to present results for different sample sizes, dimensions and neighbor numbers k: $\widetilde{\Delta} = 2(Nqk)^{-1}\Delta$. For two close points X_i, X_j, the function is $q^{-1}\mathtt{Tr}\left(H_i^T H_j\right) \to 1$. Therefore, functional $\widetilde{\Delta}$ should be close to -2. The value of functional $\widetilde{\Delta}$ as a function of $\log_{10}(N/100)$ is calculated for sample sizes $N = 100 \cdot 2^m$, $m = 0, \ldots, 4$ for the cylinder and two-dimensional sphere and $m = 0, \ldots, 3$ for the four-dimensional sphere. For each sample size, 10 samples were calculated. In Figs. 1, 2 and 3, the mean value of functional (6) and its mean $+/-$ standard deviation are shown.

In Figs. 4 and 5, the optimized vector fields (a and c) and vector fields without optimization (b and d) for the 2D sphere and cylinder are indicated.

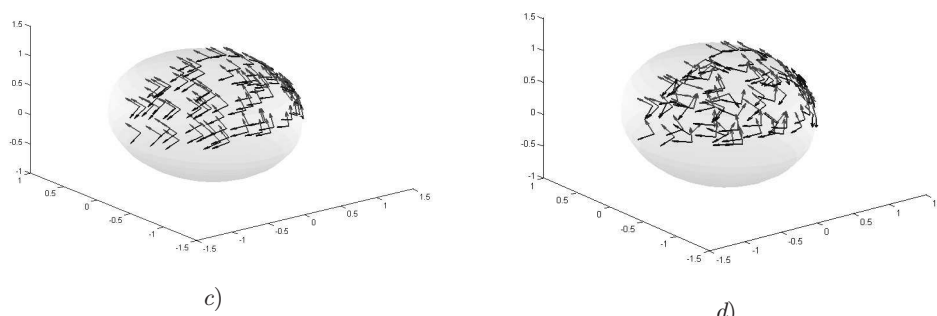c)   d)

Fig. 4. Vector fields on sphere
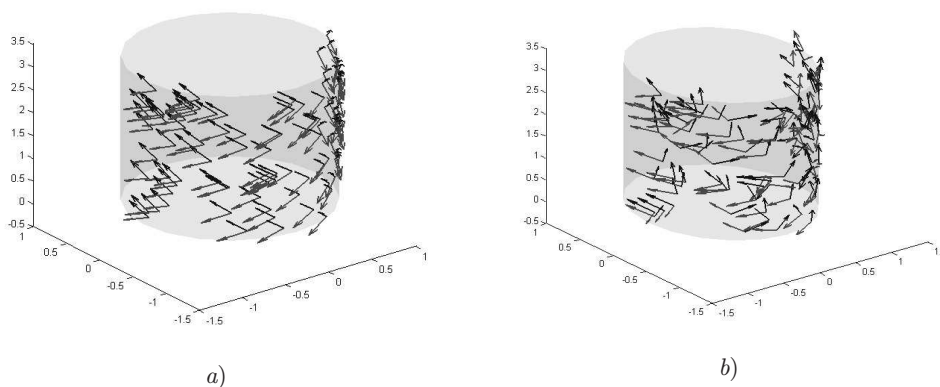

a)   b)

Fig. 5. Vector fields on cylinder

Conclusions

The optimization aspects of vector field alignment on an unknown manifold were considered. It is an important part of the Grassmann & Stiefel eigenmaps algorithm for the manifold learning problem.

The algorithm for vector field alignment was introduced. The algorithm consists of two steps: bases initialization at the same oriented atlas and Stiefel group optimization. The numerical experiments with artificial data were performed. The proposed algorithm shows good results.

References

1. Bellman R.E. *Dynamic Programming*. Princeton, Princeton Univ. Press, 1957. 339 p.

2. Donoho D.L. High-dimensional data analysis: The curses and blessings of dimensionality. *Proc. AMS Conf. on Math Challenges of 21st Century*, 2000, pp. 1–33.

3. Seung H.S., Lee D.D. Cognition. The manifold ways of perception. *Science*, 2000, vol. 290, no. 5500, pp. 2268–2269. doi: 10.1126/science.290.5500.2268.

4. Huo X., Ni X.S., Smith A.K. A survey of manifold-based learning methods. In: Liao T.W., Triantaphyllou E. (Eds.)*Recent Advances in Data Mining of Enterprise Data*. Singapore, World Sci., 2007, pp. 691–745. doi: 10.1142/9789812779861_0015.

5. Ma Y., Fu Y. *Manifold Learning Theory and Applications*. London, CRC Press, 2011. 314 p.

6. Tenenbaum J.B., de Silva V., Langford J. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, vol. 290, no. 5500, pp. 2319–2323. doi: 10.1126/science.290.5500.2319.

7. Roweis S.T., Saul L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, vol. 290, no. 5500, pp. 2323–2326. doi: 10.1126/science.290.5500.2323.

8. Zhang Z., Zha H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM J. Sci. Comput.*, 2004, vol. 26, no. 1, pp. 313–338. doi: 10.1137/S1064827502419154.

9. Belkin M., Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *J. Neural Comput.*, 2003, vol. 15, no. 6, pp. 1373–1396. doi: 10.1162/089976603321780317.

10. Belkin M., Niyogi P. Convergence of Laplacian eigenmaps. *Adv. Neural Inf. Process. Syst.*, 2007, vol. 19, pp. 129–136.

11. Donoho D.L., Grimes C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 10, pp. 5591–5596. doi: 10.1073/pnas.1031596100.

12. Bernstein A., Kuleshov A.P. Manifold learning: Generalizing ability and tangent proximity. *Int. J. Software Inf.*, 2013, vol. 7, no. 3, pp. 359–390.

13. Bernstein A., Kuleshov A., Yanovich Y. Manifold learning in regression tasks. In: Gammerman A., Vovk V., Papadopoulos H. (Eds.) *Statistical Learning and Data Sciences. SLDS 2015. Lecture Notes in Computer Science*. Vol. 9047. Cham, Springer, 2015, pp. 414–423. doi: 10.1007/978-3-319-17091-6_36. 2015.

14. Pelletier B. Non-parametric regression estimation on closed Riemannian. *J. Nonparametric Stat.*, 2006, vol. 18, no. 1, pp. 57–67. doi: 10.1080/10485250500504828.

15. Niyogi P., Smale S., Weinberger S. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 2008, vol. 39, no. 1, pp. 419–441. doi: 10.1007/s00454-008-9053-2.

16. Bernstein A.V., Kuleshov A.P., Yanovich Yu.A. Locally isometric and conformal parameterization of image manifold. *Proc. 8th Int. Conf. on Machine Vision (ICMV 2015)*, 2015, vol. 9875, art. 987507, pp. 1–7, doi: 10.1117/12.2228741.

17. Kachan O., Yanovich Y., Abramov E. Vector fields alignment on manifolds via contraction mappings. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2018, vol. 160, no. 2, pp. 300–308.

18. Yanovich Yu. Asymptotic properties of local sampling on manifold. *J. Math. Stat.*, 2016, vol. 12, no. 3, pp. 157–175. doi: 10.3844/jmssp.2016.157.175.

19. Absil P.A., Mahony R., Sepulchre R. *Optimization Algorithms on Matrix Manifolds*. Princeton, Princeton Univ. Press, 2007. 240 p.

20. Boumal N., Mishra B., Absil P.-A., Sepulchre R. Manopt, a Matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.*, 2014, vol. 15, no. 1, pp. 1455–1459.

Abramov Evgeny Nikolayevich, Graduate Student of the Faculty of Computer Science

National Research University "Higher School of Economics"

ul. Myasnitskaya, 20, Moscow, 101000 Russia

E-mail: *petzchner@gmail.com*

Yanovich Yury Alexandrovich, Candidate of Physical and Mathematical Sciences, Researcher of the Center for Computational and Data-Intensive Science and Engineering; Researcher of the Intelligent Data Analysis and Predictive Modeling Laboratory; Lecturer of the Faculty of Computer Science

Skolkovo Institute of Science and Technology

ul. Nobelya, 3, Territory of the Innovation Center "Skolkovo", Moscow, 143026 Russia

Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences

Bolshoy Karetny pereulok, 19, str. 1, Moscow, 127051 Russia

National Research University "Higher School of Economics"

ul. Myasnitskaya, 20, Moscow, 101000 Russia

E-mail: *yury.yanovich@iitp.ru*

УДК 519.23

Оценивание гладких векторных полей на многообразии с помощью оптимизации на специальной ортогональной группе

Е.Н. Абрамов[1] , *Ю.А. Янович*[2,3,1]

[1]*Национальный исследовательский университет «Высшая школа экономики»,
г. Москва, 101000, Россия*
[2]*Сколковский институт науки и технологий, г. Москва, 143026, Россия*
[3]*Институт проблем передачи информации Харкевича РАН, г. Москва, 127051, Россия*

Аннотация

Зачастую данные, полученные из реальных источников, имеют высокую размерность. Однако часто, в силу возможного наличия зависимости между параметрами, данные занимают лишь малую часть высокоразмерного пространства. Самая общая модель описания таких закономерностей – предположение о том, что данные лежат на или около многообразия меньшей размерности. Такое предположение называется гипотезой многообразия, область применения такой гипотезы – обучение на многообразии.

Вложения Грассмана–Штифеля – один из алгоритмов обучения на многообразии, вариация которого представлена в работе: оценивание гладких векторных полей на многообразии с помощью оптимизации на специальной ортогональной группе. Представлен алгоритм для решения задачи, проведены численные эксперименты на искусственных данных.

Ключевые слова: обучение на многообразии, снижение размерности, оценивание векторных полей, оптимизация на ортогональной группе

Абрамов Евгений Николаевич, студент факультета компьютерных наук

Национальный исследовательский университет «Высшая школа экономики»
    ул. Мясницкая, д. 20, г. Москва, 101000, Россия
E-mail: *petzchner@gmail.com*

Янович Юрий Александрович, кандидат физико-математических наук, научный сотрудник Центра по научным и инженерным вычислительным технологиям для задач с большими массивами данных; научный сотрудник лаборатории интеллектуального анализа данных и предсказательного моделирования; старший преподаватель факультета компьютерных наук

Сколковский институт науки и технологий
    ул. Нобеля, д. 3, Территория Инновационного Центра «Сколково», г. Москва, 143026,
Россия

Институт проблем передачи информации им. А.А. Харкевича РАН
    Большой Каретный переулок, д. 19, стр. 1, г. Москва, 127051, Россия

Национальный исследовательский университет «Высшая школа экономики»
    ул. Мясницкая, д. 20, г. Москва, 101000, Россия
E-mail: *yury.yanovich@iitp.ru*