Embedding Super-Symmetric Tensors of Higher-Order Similarities of High-Dimensional Data

Praneeth Vepakomma[1] and Ahmed Elgammal[2]

[1] Department of Statistics, Rutgers University, New Brunswick NJ, USA,
praneeth@scarletmail.rutgers.edu
[2] Department of Computer Science, Rutgers University, New Brunswick NJ, USA ,
elgammal@cs.rutgers.edu

Abstract. In this paper we propose an algorithm for non-linear embedding of affinity tensors obtained by measuring higher-order similarities between high-dimensional points. We achieve this by preserving the original triadic similarities using another triadic similarity function obtained by sum of squares of diadic similarities in a low-dimension. We show that this formulation reduces to solving for the non-linear embedding of a graph which has a specific kind of a graph Laplacian. We provide an iterative algorithm for minimizing the loss, and also propose a simple linear-constraint that prevents non-zero solutions for embedding problems unlike the existing variants of quadratic orthonormality constraints used in the literature, that require eigen decompositions to solve for the embedding.

Keywords: Tensors, Laplacians, Embedding

1 Introduction:

Tensors or N-way array or multiway data and its analysis has found extensive applications in Computer Vision, VLSI Design and Synthesis, Statistics, Machine Learning, Neuroimaging, Hyperspectral Data Analysis, Recommendation Systems, Time Evolving Networks, Hypergraph analysis and Linked Data Analysis. A special tensor that often arises in such applications is the affinity tensor that is super-symmetric whose entries are calculated using higher-order tuples of data points. These entries are essentially formed using triadic, teradic or higher-order similarity functions unlike the usual setting of a distance/similarity matrix where pairwise similarities are used. There has been recent research over solving the clustering [1], [3], feature matching[4], embedding Patches to Tensors [6] and classification [2] problems, within this setting where the data is in the form of affinity tensors and hypergraphs. In parallel, there has been considerable research in solving the non-linear embedding problem using dyadic (pairwise) similarities and distance matrices and this area of work has also been referred to as 'Manifold Learning' [7],[8],[9],[10],[11] in the machine learning communities and has primarily focussed on learning with pairwise similarities. In this paper we solve the manifold learning problem, in a setting where the similarities are computed between higher-order (> 2) tuples of points. The organization of the paper is as follows: In the second section we give more details on the tensor embedding problem and our proposed formulation. In the third section, we propose an algorithm to solve the proposed formulation, following which we provide experimental results, a conclusion and a bibliography.

2 Tensor Embedding:

We first begin with defining some basic notation. Let $Y \in R^{n \times d}$ be a matrix of n high-dimensional points of dimension d. We use $d_{ij}^2(M)$ to denote the squared Euclidean distance between the row i and row j

of any real matrix M. Let $f(.)$ be a triadic similarity function (acts on three points at a time) and W be the affinity tensor such that $W_{ijk} = f(Y_{i\cdot}, Y_{j\cdot}, Y_{k\cdot})$. Given an instance W, we work in a setting where the corresponding instances of $f(.)$ and Y that generate the W are unknown. We now would like to solve for a low-dimensional set of points $X \in R^{n\times p}$ where $p \lll d$ for a chosen p and also where the triadic relations in W are preserved using functions of pairwise or diadic similarities between the points of X. Estimating the intrinsic dimension [12],[13],[14],[15],[16] of a high dimensional dataset for generating an optimal embedding has been a closely related problem that has generated extensive research and still remains to be an active research problem. We propose to formulate the objective function for this model as follows using a sum of three squared Euclidean distances of cyclic pairs (dyadic) of rows of the unknown matrix $X \neq 0$:

$$\arg\min_{X\neq 0}\sum_{i,j,k} W_{ijk}^2[d_{ij}^2(X) + d_{jk}^2(X) + d_{ki}^2(X)] \tag{1}$$

Note that W is a 3-Tensor, with W_{ijk}^2 being the entry-wise squares. The rensor W is the only known component of this model. The minimization of this model preserves the triadic locality relations within W using a triadic similarity function that is composed of diadic similarities. This formulation was inspired by the concepts of cohomology groups [20] of simplicial complexes in algebraic topology. Now the only required constraint as putforth in (1) is that the matrix X should be constrained to be a non-zero matrix. On rearranging the terms, the proposed model in (1) can now be expressed as

$$\arg\min_{X\neq 0}\sum_{i,j}(\sum_{k=1}^{n} W_{ijk}^2)d_{ij}^2(X) + \sum_{j,k}(\sum_{i=1}^{n} W_{ijk}^2)d_{jk}^2(X) + \sum_{k,i}(\sum_{j=1}^{n} W_{ijk}^2)d_{ki}^2(X) \tag{2}$$

We note that under the special case where W is a matrix (2-array or 2-tensor), the above model can be expressed as the sums of objective functions of three manifold learning problems. Hence, we try to rearrange the above equation inorder to topically make it look so by denoting

$$V_{ij}^{\alpha} = \sum_{k=1}^{n} W_{ijk}^2, V_{jk}^{\beta} = \sum_{i=1}^{n} W_{ijk}^2, V_{ij}^{\gamma} = \sum_{j=1}^{n} W_{ijk}^2 \tag{3}$$

and we have that 1 is exactly the same as below:

$$\arg\min_{X\neq 0}\sum_{ij} V_{ij}^{\alpha} d_{ij}^2(X) + \sum_{jk} V_{jk}^{\beta} d_{jk}^2(X) + \sum_{ki} V_{ki}^{\gamma} d_{ij}^2(X) \tag{4}$$

This can now be expressed using graph Laplacians L_α, L_β and L_γ, with the corresponding degree matrices D^α,D^β, D^γ and adjacency matrices V^α, V^β and V^γ of a graph as follows:

$$L^\alpha = D^\alpha - V^\alpha, L^\beta = D^\beta - V^\beta, L^\gamma = D^\gamma - V^\gamma \tag{5}$$

$$D_{ii}^{\alpha} = -\sum_{j=1}^{n}\sum_{k=1}^{n} W_{ijk}^2, D_{jj}^{\beta} = -\sum_{k=1}^{n}\sum_{i=1}^{n} W_{ijk}^2, D_{kk}^{\gamma} = -\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ijk}^2 \tag{6}$$

Using these matrices our model can now be expressed using the Laplacians as

$$\arg\min_{X\neq 0} TrX^T L^\alpha X + TrX^T L^\beta X + TrX^T L^\gamma X \tag{7}$$

$$\arg\min_{X\neq 0} TrX^T[L^\alpha + L^\beta + L^\gamma]X \tag{8}$$

The above function can be represented as a sum of three trace optimization problems with $L = L^\alpha + L^\beta + L^\gamma$ being a symmetric positive semidefinite matrix as:

$$\arg\min_{X\neq 0}\ \Theta(X) = Tr[X^T LX] \tag{9}$$

3 Optimization Framework:

We build a majorization function over the above model, based on the fact that $[2Diag[L] - L]$ is diagonally dominant. This leads, to the following inequality for any matrix $M_{n\times p}$

$$(X - M)^T[2Diag[L] - L](X - M) \succeq 0 \tag{10}$$

We get the following majorization inequality over our objective function, by separating it from the above equation:

$$Tr(X^T LX) + f(Y) \leq Tr[X^T 2Diag(L)X] - 2Tr[X^T(2Diag(L) - L)M] = g(X, M) \tag{11}$$

which is quadratic in X where,

$$f(M) = Tr(M^T LM) - Tr(M^T 2Diag(L)M)$$

Hence, we achieve the following bound over our objective function:

$$\begin{aligned} Tr(X^T LX) + f(M) &\leq g(X, M), \forall\ X \neq M \\ &= g(X, X),\ X = M \end{aligned}$$

that satisfies the supporting point requirement, and hence g(.) touches the objective function at the current iterate and the following majorization-minimization iteration holds true:

$$X^{t+1} = \arg\min_X g(X, M^t) \text{ and } M^{t+1} = X^t$$

It is important to note that these inequalities occur amongst the presence of additive terms that are independent of X unlike a typical majorization-minimization framework and hence, it is a relaxation.

3.1 A linear constraint for dimensionality reduction to coerce a non-zero solution:

We now propose a linear, non-orthonormal regularizer for nonlinear dimensionality reduction over the quadratic loss function proposed in (7). Our regularizer prevents degenerate solutions, where the rows(or columns) of X coincide thereby preventing $d_{ij}(X)$ from going to zero. Its linearity, makes it easier to practically enforce it due to the quadratic nature of the loss function. Row Unique Matrix: A matrix M is row-unique, if all the rows in the matrix are distinct.

Proposition: For any row-unique matrix $M_{n\times p}$, and for any given Laplacian matrix $L_{n\times n}$, if $Tr(X^T LM) \neq 0$, then there exist at least two rows in $X_{n\times p}$, that are distinct.

Proof. $Tr(X^T LM) = \sum_{i<j} w_{ij}\phi_{ij}(X,M)$ where, $\phi_{ij}(X,M) = \sum_{a=1}^{p}(x_{ia} - x_{ja})(m_{ia} - m_{ja})$ Hence, for a row unique M, there exists at least two rows in X, such that $x_{i.} \neq x_{j.}$ in order to satisfy the inequality on $Tr(X^T LM)$. Note that, $\phi_{ij}(X,X) = d_{ij}^2(X)$

We define our regularizer in its basic form for nonlinear dimensionality reduction as follows:

$$Tr(X^T LM) = \nu \tag{12}$$

where $\nu > 0$ is a user-defined constant. As a result of $g(.)$ being a quadratic majorizer we have

$$\lim_{t\to\infty} \|X_{t+1} - M_t\| \to 0$$

as a result of which, we have the following over $\phi(.)$ in our linear regularizer

$$\lim_{t\to\infty} \phi_{ij}(X_{t+1}, M_t) \to d_{ij}^2(X) \in \Re^+$$

and hence we require that ν be non-negative inorder to simultaneously achieve convergence and enforce regularization. The following is the total loss function, $T(.)$ obtained when the regularizer is combined with our majorizing function $g(.)$, defined in (7) with λ being a positive multiplier over the regularizer:

$$T(X,\lambda) = g(X,M) + \lambda\left[Tr(X^T LM) - \nu\right] \tag{13}$$

The gradient, is given by:

$$\bigtriangledown T(X,\lambda) = 4Diag(L)X - 4Diag(L)M + 2LM + \lambda LM \tag{14}$$

We get the following update, by setting the gradient equal to zero.

$$X_{t+1} = M_t - (0.5 + 0.25\lambda)\,[Diag(L)]^{-1} LM_t \tag{15}$$

and solving for the constraint, we get the follwing update, for the multiplier:

$$\lambda_t = \frac{4Tr[M_t^T LM_t - \nu]}{Tr(M_t^T V Diag(L)^{-1} LM_t)} - 2 \tag{16}$$

$$M_t = X_{t+1} \tag{17}$$

The assignment in (17) is to construct a majorization function at the latest iterate, that is recursively minimized using (15) and (16). Hence, this generates updates that satisfy the following set of inequalities.

$$\Theta(X_t) \leq g(X_t, M_{t-1}) \leq g(X_{t-1}, X_{t-1}) \leq \Theta(X_{t-1})$$

With every iterate, doing better than the previous, it proves the convergence of our updates.

4 Results:

For experimenting our tensor embedding algorithm we use the Abalone (real-life) dataset and the USPS Images of Handwritten Digits (real-life) dataset. We then compute the entries of an instance of the input tensor W_{ijk} using $e^{-\frac{||y_i - y_j||^2 + ||y_j - y_k||^2 + ||y_k - y_i||^2}{\sigma}}$ for experimental purposes though it is indeed possible that these weights could be anything depending on the application. Especially, in computer vision such tensors are directly computed/arise without using any pairwise (dyadic) similarity functions in a directly applicable setting. σ, the bandwidth or the locality parameter is manually chosen just like in Laplacian Eigenmaps.

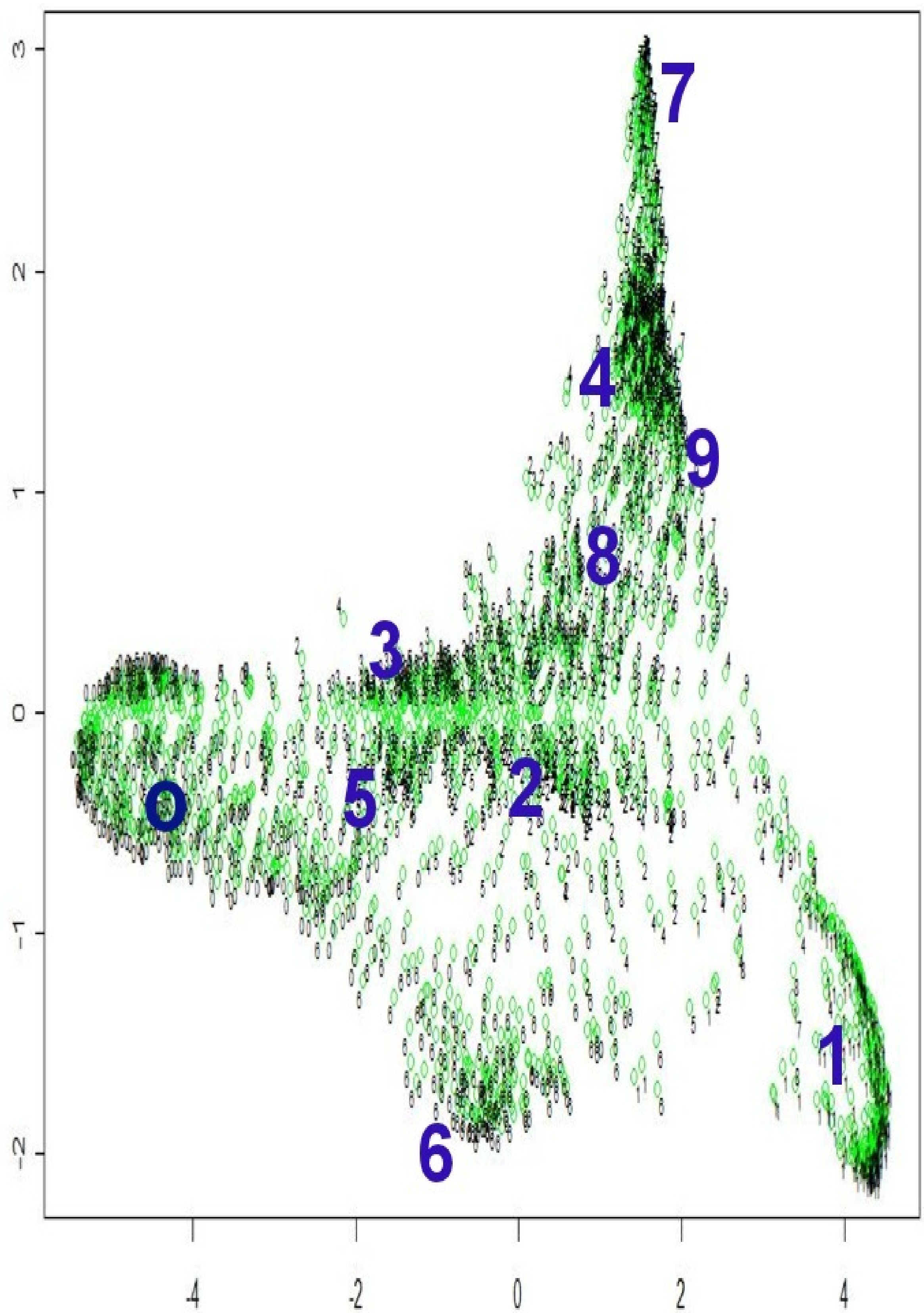

Fig. 1: Learned Embedding of all handwritten images in the USPS (United States Postal Service) Dataset with the blue labels indicating the corresponding image patterns

4.1 Abalone Dataset:

This is a dataset traditionally used in testing supervised learning algorithms and is availabke on the UCI-ML repository. The aim of our experiment is to learn a tensor embedding in three dimensions, where we try to separate infant Abalones from adult Abalones. We drop the variables of gender, number of rings and height of the abalone from the dataset and we use the variables of shell weight, diameter, whole weight, shucked weight and viscera weight in our experiment. In Fig 2. we show the results of the embedding of Abalones, and we then color the infant (red circles) and adult abalones (green circles). Fig 3. shows the effect of the choice of the ν parameter in the linear constraint on data depth [17],[18],[19] which is a robust measure of scatter. As seen, the scatter initially increases with greater values of ν and then begins to start stabilizing at a point.

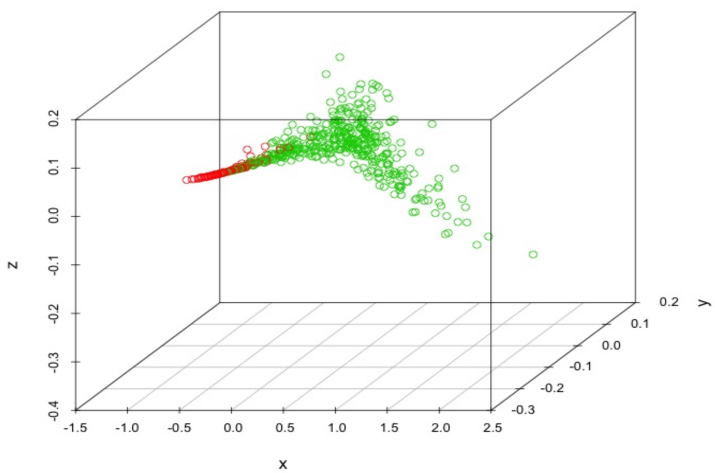

Fig. 2: Learned embedding over the Abalone Dataset, with the Red circles representing infant abalones and Green circles representing adult abalones

4.2 USPS Dataset:

The usps dataset of handwritten images was gathered at the Center of Excellence in Document Analysis and Recognition (CEDAR) at SUNY Buffalo, as part of a project sponsored by the US Postal Service. This dataset has traditionally been used in a splitting of 7291 cases for training and 2007 cases for demonstrating learning algorithms. In Fig 1. we show the results of the embedding of the USPS dataset. We provide manual labels in blue to point to the groups of corresponding digits observed in the embedding.

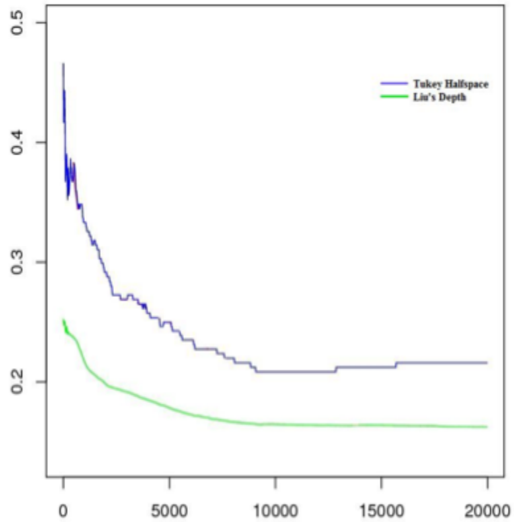

Fig. 3: Effect of ν in our algorithm on Data Depth Measured using J.Tukey's (blue line) and R.Liu's (green line) Robust Data Depth Functions

5 Conclusion:

We introduce a method for non-linear embedding of tensors with triadic affinities. Previous work on non-linear embedding has mainly focussed on two-way affinities. We give a distance-geometric formulation of tensor embedding and provide an iterative algorithm to solve for the embedding. Our method is naturally applicable in the areas of Computer Vision and hypergraph data analysis where non-linear embedding problems and triadic similarity data often occur together. We finally show results of our algorithm on two real-life datasets where we obtain reasonable embeddings that preserve the neighborhood relations within the affinities.

References

1. D Zhou ., J Huang ., B Scholkopf : Beyond Pairwise Classication and Clustering Using Hypergraphs, Proceedings of the Neural Information Processing Systems (2005)
2. A Shashua., R Zass., T Hazan: Multi-way clustering using super-symmetric non-negative tensor factorization, European Conference on Computer Vision, ECCV, (2006)
3. Sameer Agarwal , Jongwoo Lim , Lihi Zelnik-Manor , Pietro Perona , David Kriegman , Serge Belongie: Beyond Pairwise Clustering, Computer Vision and Pattern Recognition, CVPR, (2005)
4. Cheng ZQ, Chen Y, Martin RR, Lai YK, Wang A: SuperMatching: Feature Matching Using Supersymmetric Geometric Constraints, IEEE Transactions on Visualization and Computer Graphics, Feb 13, (2013)
5. Ronald R. Coifman, Stphane Lafon: Diffusion maps, Applied and Computational Harmonic Analysis, Vol. 21, 5-30 (2006)
6. Moshe Salhov, Guy Wolf, Amir Averbuch: Patch-to-tensor embedding, Applied and Computational Harmonic Analysis, Vol. 33, 182-203 (2012)
7. Mikhail Belkin, Partha Niyogi: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, Neural Computation, Vol. 15, No. 6, 1373-1396, June (2003)

8. D.L. Donoho, C. Grimes: Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data Proceedings of National Academy of Sciences, (PNAS)- 100, 55915596, (2003)
9. S.T. Roweis, L.K. Saul: Nonlinear dimensionality reduction by locally linear embedding Science, 290, 23232326, (2000).
10. R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, S. Zucker: Geometric diffusions as a tool for harmonic analysis and structure definition of data, Proceedings of National Academy of Sciences-(PNAS), (2004)
11. J. B. Tenenbaum, V. de Silva and J. C. Langford: A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290 (5500), 22: 2319-2323, (2000)
12. Elizaveta Levina, Peter J. Bickel: Maximum Likelihood Estimation of Intrinsic Dimension, Proceedings of Neural Information Processing Systems, NIPS, (2005)
13. B. Kegl: Intrinsic dimension estimation using packing numbers, Proceedings of Neural Information Processing Systems, NIPS (2002)
14. Maxim Raginsky, Svetlana Lazebnik: Estimation of Intrinsic Dimensionality Using High-Rate Vector Quantization, Proceedings of Neural Information Processing Systems, NIPS (2005)
15. Kumar Sricharan, Raviv Raich, Alfred O. Hero III: Optimized Intrinsic Dimension Estimator using Nearest Neighbor Graphs, International Conference on Acoustics, Speech, and Signal Processing, ICASSP, (2010)
16. Kevin M. Carter, Raviv Raich, Alfred O. Hero III: On Local Intrinsic Dimension Estimation and Its Applications IEEE Transactions on Signam Processing, Vol. 58, NO. 2, February, (2010)
17. John Tukey: Mathematics and picturing data. In Proceedings of the 1975 International Congress of Mathematics 2 523 531. Z.(1975)
18. Regina Y. Liu: On a Notion of Data Depth Based on Random Simplices, Annals of Statistics (1990)
19. Regina Y. Liu, Multivariate spacings based on data depth: I. Construction of nonparametric multivariate tolerance regions, Annals of Statistics (2008)
20. Joseph J. Rotman: An Introduction to Algebraic Topology (Graduate Texts in Mathematics), Springer