Minimization of the mapping error using coordinate descent

Gintautas Dzemyda
Institute of Mathematics and Informatics
Akademijos St. 4
Vilnius 08663, Lithuania
Dzemyda@ktl.mii.lt

Jolita Bernataviciene
Institute of Mathematics and Informatics
Akademijos St. 4
Vilnius 08663, Lithuania
JolitaB@ktl.mii.lt

Olga Kurasova
Institute of Mathematics and Informatics
Akademijos St. 4
Vilnius 08663, Lithuania
Kurasova@ktl.mii.lt

Virginijus Marcinkevicius
Institute of Mathematics and Informatics
Akademijos St. 4
Vilnius 08663, Lithuania
VirgisM@ktl.mii.lt

ABSTRACT

Visualization harnesses the perceptual capabilities of humans to provide the visual insight into data. Structure preserving projection methods can be used for multidimensional data visualization. The goal of this paper is to suggest and examine the projection error minimization strategies that would allow getting a better and less distorted projection. The classic algorithm for Sammon's projection and two new its modifications are examined. All the algorithms are oriented to minimize the projection error because even a slight reduction in the projection error changes the distribution of points on a plane essentially. The conclusions are made on the results of experiments on artificial and real data sets.

Keywords

Multidimensional data, visualization, Sammon's mapping, mapping error, coordinate descent

1. INTRODUCTION

Objects from the real world are frequently described by an array of parameters (multidimensional data). Data understanding is a difficult task, especially when it refers to a complex phenomenon that is described by many parameters. One of the ways in analyzing data is visualization. It involves the constructing of a graphical interface that enables to understand complex data. Visualization is also used to display the properties of data that have a complex relation – possibly patterns not obtainable by the current computation methods.

In this paper, we discuss visualization of multidimensional data by using structure preserving projection methods. These methods are based on the idea that the multidimensional data points can be projected on a lower dimensional space so that the structural properties of the data are preserved as faithfully as possible. Examples of such techniques are principal component analysis [Tay03a], multidimensional scaling [Kas97a], [Bor97a], Sammon's mapping [Sam69a], and others.

Conference proceedings ISBN 80-903100-9-5
WSCG'2005, January 31-February 4, 2005
Plzen, Czech Republic.

This paper deals with Sammon's mapping. Sammon's mapping comes from the area of multidimensional scaling. The only difference between both methods is that the errors in distance preservation are normalized with the distance in the original space. Because of the normalization, the preservation of small distances will be emphasized [Kas97a]. The analysis of relative performance of the different algorithms in reducing the dimensionality of multidimensional vectors, starting from the paper by Biswas [Bis81a], indicates Sammon's projection to be still one of the best methods of this class (see also [Fle97a]) and finds new applications (see, e.g. [Dze01a]). When visualizing the multidimensional data using the nonlinear projection, the projection errors are inevitable. The goal of this paper is to suggest and examine the projection error minimization strategies that would allow getting a better and less distorted projection.

2. STRATEGIES FOR PROJECTION ERROR MINIMIZATION

Let us have s objects (points, vectors) $X_i = (x_1,...,x_n)$, $i = 1,...,s$ from an n-dimensional space R^n. The aim of this method is to find *s* points in an *m*-dimensional space ($m < n$, usually *m*=2) $Y_i = (y_1, ..., y_m) \in R^m$, $i = 1, ..., s$ so that the corresponding distances of *m*-dimensional points approximate the original ones as well as possible. Sammon's mapping [Sam69a] is one of such methods. It tries to minimize the projection error:

$$E_s = \frac{1}{\sum\limits_{\substack{i,j=1 \\ i<j}}^{s} d_{ij}^*} \sum_{\substack{i,j=1 \\ i<j}}^{s} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*} \quad (1)$$

Here d_{ij}^* is the distance between two *n*-dimensional points; d_{ij} is the distance between two points in the two-dimensional space. Even a slight reduction in E_s changes the distribution of points on a plane essentially. This proves the necessity to make every effort for minimizing the distortion of projection E_s. In this paper, three strategies for projection error minimization are investigated: (1) classical Sammon's algorithm [Sam69a] (**S1**); (2) applying the Seidel coordinate descent for Sammon's method (**S2**); (3) applying the noise for **S2** method (**S3**).

Classical Sammon's algorithm (S1). In this method, the coordinates y_{ik}, $i = 1, ..., s,$ $k = 1, 2$ of two-dimensional vectors $Y_i = (y_{i1}, y_{i2})$ are computed by the iteration formula:

$$y_{ik}(m'+1) = y_{ik}(m') - a \left. \frac{\partial E_s(m')}{\partial y_{ik}(m')} \middle/ \frac{\partial^2 E_s(m')}{\partial y_{ik}^2(m')} \right.$$

Here m' denotes the iteration order number; a is a step length, also called a „magic factor", because the obtained projection error depends on it.

One iteration of the algorithm contains calculations, where both components of all the points Y_i, $i = 1, ..., s$ are recalculated. These components are recalculated taking into account the coordinates of Y_i, $i = 1, ..., s$, obtained in the previous iteration.

Seidel-type coordinate descent for Sammon's mapping (S2). Seidel-type coordinate descent method is used for solving linear equation systems and in optimization [Kar03a]. We suggest applying coordinate descent for Sammon's mapping. The coordinates of two-dimensional vectors Y_i are recalculated, taking in to consideration not only the coordinates, obtained in the previous iteration, as in classical Sammon's algorithm, but also the new coordinates, obtained in the current iteration: the coordinates $y_{jk}(m'+1)$, if $j = 1, ..., i-1$, and $y_{jk}(m')$, if $j = i+1, ..., s$.

The coordinate descent method with noise (S3). It has been noted that the 1st derivatives of E_s are smooth enough in algorithm **S1** (Fig. 1), but the 2nd derivatives are alternating (Fig. 2). Iris data set is analyzed. The 1st and 2nd derivatives of E_s were measured in each iteration, when new coordinates of the 4th data point were determined. Numerous experiments allowed us to see that the disperse of two-dimensional points from the line (initialization rule of two dimensional points) increases with an increase in fluctuation of the 2nd derivatives. We decided to apply artificial fluctuation to the 2nd derivatives (to apply noise) in algorithm **S2.** In such a way a new algorithm **S3** has been created.

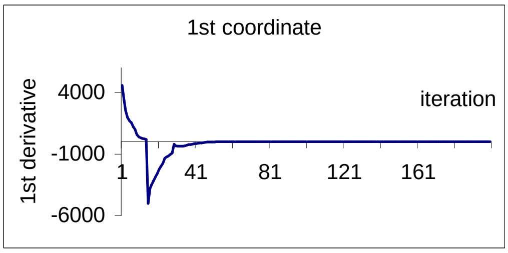


Figure 1. Fluctuation of the 1st derivative in S1.

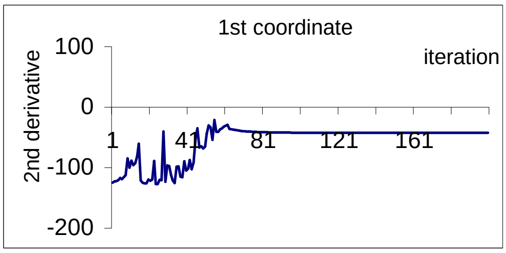


Figure 2. Fluctuation of the 2nd derivative in S1.

The first idea was to add random noise to the second derivative as follows:

$$\frac{\partial^2 E_s}{\partial y_{ik}^2} \Leftarrow \frac{\partial^2 E_s}{\partial y_{ik}^2} + \xi$$

Here x is a random number. However, the problems arose to schedule the level of noise depending on the value of the second derivative and the order number of the current iteration. Therefore, a more effective way has been found to define noise by some heuristic rule:

$$\frac{\partial^2 E_s}{\partial y_{ik}^2} \Leftarrow \frac{\partial^2 E_s}{\partial y_{ik}^2}(1 - e^{-\lambda m'})\left|\sin(\beta m')\right|, \text{ for } m' < \frac{t}{n}.$$

Here l, b are some constants, selected experimentally, t is the total number of iterations, m' is the order number of the current iteration.

3. RESULTS OF THE ANALYSIS

Experiments were carried out with real and artificial data. The dependence of E_s on different factors is investigated on: the computing time, the number of iterations, and the value of "magic factor" a .

Data for analysis. Artificial data sets:

1. Uniformly distributed data: 100 10-dimensional points generated at random in the interval [-1; 1];
2. Uniformly distributed data: 500 points generated at random, as No1.
3. Clustered data: ten 10-dimensional points are generated at random; in the area of each point, nine 10-dimensional points are generated by normal distribution.

The experiments have been repeated for 100 times with different sets of 10-dimensional vectors generated as given above (data sets No1-No3). The average results have been calculated.

Real data:

- The classical Fisher iris data set [Fis36a].
- The Wood data set [Dra66a].
- The HBK data set [Haw84a].

Dependence of the projection error on computing time

The advantages of algorithm **S3** in comparison with algorithm **S1**, **S2** have been shown by analyzing data No2 ($\alpha = 0.25$). A lower projection error and its faster convergence to optimal value have been obtained using algorithm **S3**. In order to get a lower value of E_S, it suffices to perform less iterations, i.e. the computing time is saved (Fig. 3).

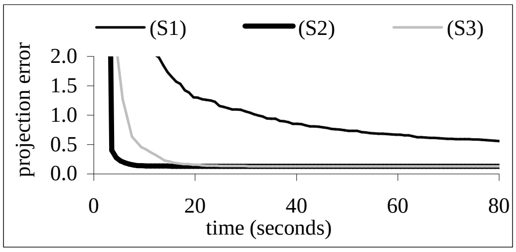


Figure 3. Dependence of the projection error on time.

Difference between projection errors in the whole iteration process of algorithms **S2** and **S3** is rather small, but in most cases, the final results obtained by **S3** are better (see Table 1)**.**

Dependence of the projection error on the "magic factor"

While examining the dependence of projection error on the value of a , artificial data No1 are analyzed with different values of a (0.1; 0.11;…;1.45; 1.5). In Fig. 4, the mean value of the projection error is presented. It is shown that dependence on the "magic factor" a is less using the Seidel-type coordinate descent method (**S2**). Applying noise (**S3**) does not influence the dependence on a .

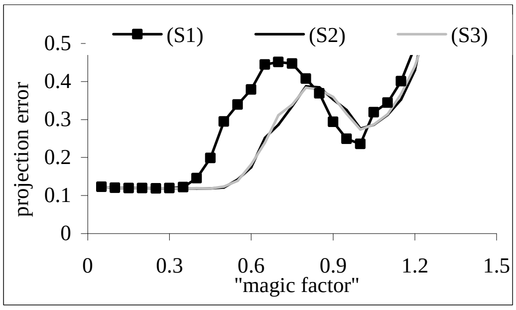


Figure 4. Dependence of the projection error on the value of "magic factor".

Strategies of E_S minimization			**Data**
(S1)	**(S2)**	**(S3)**	
0.1209452	0.1201307	0.1203316	artificial (No1)
0.0710200	0.0711317	0.0695346	clusters
0.0058476	0.0045259	0.0040088	Iris
0.0243263	0.0257550	0.0256691	Wood
0.0112111	0.0113962	0.0049657	HBK

Table 1. Projection errors

Analysis of mappings

The minimal projection errors have been found for each real data set. After examining the artificial data No1, the average values of minimal projection errors have been calculated over 100 experiments ($\alpha = 0.25$) (Table 1). The largest distortion of projection has been obtained using by algorithm **S1** in all the cases. The smallest projection error has been obtained using algorithm **S3** in most cases.

In Figures 5-6, the mappings of the real data are presented. They have been obtained using algorithms **S1** and **S3** that gives the smallest projection error with the real data (see Table 1). The figures show that, if a smaller projection error is obtained, the preserving data structure is more precise. When analyzing the iris data, three flower types are separated more precisely by using algorithm **S3** (see Fig. 5b); when investigating the HBK data, classical Sammon's mapping (**S1**) is able to separate point groups (see Fig. 6a), but all the three groups are separated more exactly when algorithm **S3** is used (see Fig. 6b).

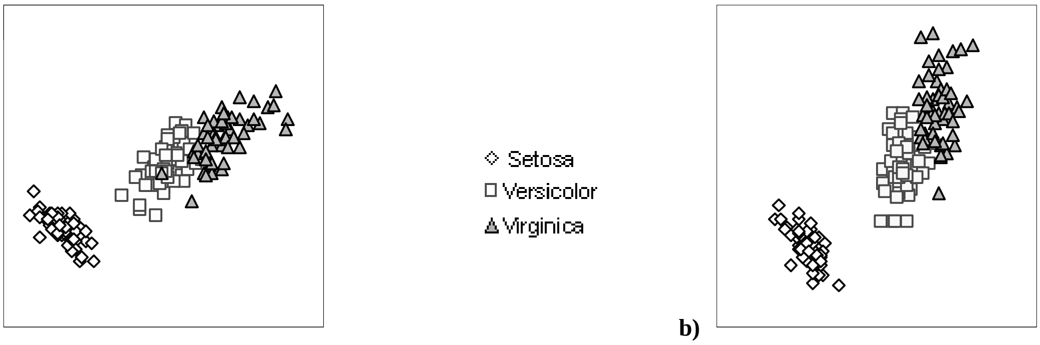
4. CONCLUSIONS

In this paper, the new opportunities for minimizing the multidimensional data projection error E_S (1) are suggested and examined experimentally. Classical Sammon's mapping **S1** is compared with two new algorithms **S2** and **S3**.

Smaller projection errors are usually obtained by using algorithms **S2** and **S3** compared with the errors by classical Sammon's algorithm **S1**. Another advantage of the new algorithms is that small errors are obtained after a smaller number of iterations and sooner. Therefore, the projections of multidimensional data on a plane are more faithfully; lower dependence of the projection quality on the value of a is obtained.

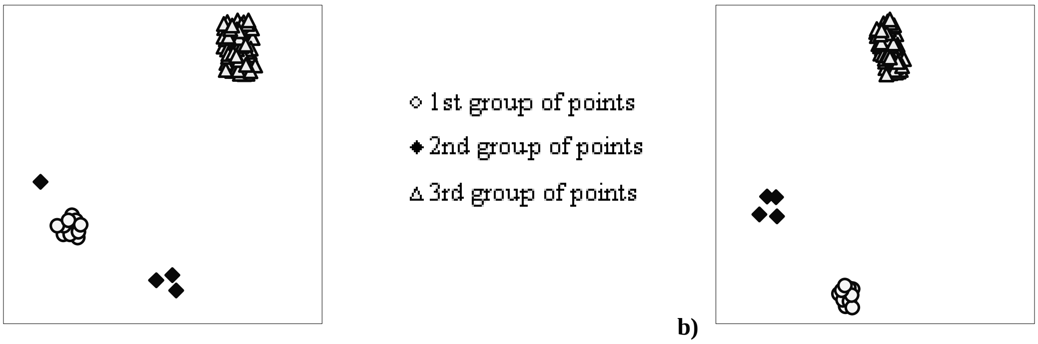
Applying noise to the 2nd derivatives in first iterations of the projection error minimization process (algorithm **S3**) speeds up the moving of the initial two-dimensional points from the line. This improves the quality of visualization, because the smaller projection errors are obtained.

Finally, the discovered new ways for minimizing the projection error allow a better perception of the Sammon-type projection and make a basis for further research.



a) **b)**

Figure 5. Projections of the iris data: a) S1 ($E_s = 0.0059$); b) S3 ($E_s = 0.0040$).



a) **b)**

Figure 6. Projections of the HBK data: a) S1 ($E_s = 0.0112$); b) S3 ($E_s = 0.0050$).

5. REFERENCES

[Bor97a] Borg, I., and Groenen, P. Modern Multidimensional Scaling: Theory and Applications, Springer, New York, 1997.

[Bis81a] Biswas, G., Jain, A.K., and Dubes, R.C. Evaluation of projection algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 3(6), pp. 701-708, 1981.

[Dra66a] Draper, N. R., and Smith, H. Applied Regression Analysis. John Wiley and Sons, New York, 1966.

[Dze01a] Dzemyda, G. Visualization of a set of parameters characterized by their correlation matrix. Computational Statistics and Data Analysis, Vol. 36(10), pp. 15-30, 2001.

[Fis36a] Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, No.7, pp.179-188, 1936.

[Fle97a] Flexer, A. Limitations of self-organizing maps for vectors quantization and multidimensional scaling. In M.C.Mozer, M.I.Jordan and T.Petsche(eds.), Advances in Neural Information Processing Systems 9, Cambridge, MA: MIT Press/Bradford Books, pp.445-451, 1997.

[Haw84a] Hawkins, D. M., Bradu, D., and Kass, G. V. Location of several outliers in multiple regression data using elemental sets. Technometrics, No.26, pp.197–208, 1984.

[Kar03a] Karbowski, A. Direct method of hierarchical nonlinear optimization - reassessment after 30 years. Proceedings of III International Conf. on Decision Support for Telecommunications and Information Society (DSTIS 2003), Warsaw, pp.15-30, 2003.

[Kas97a] Kaski, S. Data Exploration Using Self-Organizing Maps, PhD thesis, Helsinki University of Technology, Department of Computer Science and Engineering, 1997, http://www.cis.hut.fi/~sami/thesis/

[Sam69a] Sammon, J. W. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers, No.18, pp.401-409, 1969.

[Tay03a] Taylor, P. Statistical methods. In M. Berthold and D. J. Hand (eds.), Intelligent Data Analysis: an Introduction, Springer-Verlag, pp. 69-129, 2003.