

# Multi-dimensional data scaling – dynamical cascade approach

Milan Jovovic, Geoffrey Fox  
Indiana University

## ABSTRACT

*In this report a multi-dimensional data scaling approach is proposed in data mining and knowledge discovery applications. We derive the method based on an analogy to the physical computation of signal distortion. A dynamical cascade computation diagrams result from the statistical physics model computation in the free energy decomposition. We assess the scale invariance of various data sets, such as with the image motion sequences, and with the high dimensional chemical data sets. Theoretical model of error propagation is given by the numerical computational schemes. Statistical mapping of the data is analyzed through dynamical cascades, as a way of approaching its coding and control data structure. We show how it correlates by segmenting set of chemical compounds observations in a high dimensional property space. The proposed algorithm, also, is suitable for the implementation in parallel computer architectures. An example implementation on the multicore processors is given in the end of this report.*

## 1. INTRODUCTION

A multifractal model formalism is derived in the “Thalweg ARC.” project report [12], to explain the decomposition of image sequences into the singular data sets. The partition function describes the probabilistic model of data clusters and is analyzed as a multifractal measure in the method. Singularity analysis of computational maps of clustering vectors is derived to describe the computational means of decomposing the image information into different singular sets. We show also that the propagation of information in image sequences is governed by the scale-space wave equation, therefore enabling us to treat singular frequencies of data clusters in an unified way, both in space and in time.

Contextual information of the spatial coherency of data is used in the segmentation process in the hierarchical scale computation of feature vectors. The spatial segmentation of images is performed while using the Green's function, parameterized with the scale parameter, as the integration function in the segmentation process. The scale information is evaluated by conjoining the two parameters: the scale parameter  $\beta$  of the signal distortion, and the spatial scale parameter  $r$ . A larger extent of spatial integration of the motion information is used on a larger scale, while it

becomes effectively more local in space as we decrease the scale of segmentation.

Distinct singular features are segmented on a certain scale and the least singular feature become segmented in two spatial windows with the Laplacian system regularity constraints, in the hierarchical scale computation. Accordingly, the reconstruction formula is derived based on the Laplacian system of the diffusion of the residual information from the most singular sets. This gives us an effective way of compressing and progressive coding of the information in image sequences. The binary tree data structure of the clustering parameters is suitable in the coding schemes that use the hierarchical structure of the binary images of the spatial distribution of cluster windows, along with the feature vectors and residual image information that make up for the point feature vector estimation.

We give here a derivation of the computational scheme for a 2-dimensional case, like in image sequences. We then consider a dynamical coupling and the energy exchange between 3 clusters computed. Corresponding statistical maps are analyzed w.r.t. the dimensionality of the eigenvalue decomposition of the clusters' covariances.

The results are shown for the chemical compounds in the 155 properties dimensional data set. Projections along the most singular components are computed in 1 and 2 dimensional statistical maps.

## 2. METHOD

### Variational data distributions

We define a cluster of data points here with its computed cluster vector representative  $y$ , and the selected group window of computation,  $W$ . Let  $d(x, y)$  denotes a distortion measure introduced to a data point  $x$  by the representation  $y$ . The distortion energy, or variance  $V$  of a cluster is defined by:

$$V = \int_W d(x, y)P(x).$$

It can be shown [8] that the probability density function that maximize/minimize the entropy:

$$\max/ \min \left\{ \int_W P(x) \log P(x) \right\},$$

subject to:  $V = \int_W d(x, y)P(x)$  and  $\int_W P(x) = 1$ ,

is the Gibbs distribution:  $P_-(x) = \frac{e^{-\beta_- d(x, y)}}{Z_-} = \frac{e^{-\beta_- d(x, y)}}{\int_W e^{-\beta_- d(x, y)}}$ ,

and  $P_+(x) = \frac{e^{+\beta_+ d(x, y)}}{Z_+} = \frac{e^{+\beta_+ d(x, y)}}{\int_W e^{+\beta_+ d(x, y)}}$ , respectively, where  $Z_-$

and  $Z_+$  are the partition functions, and  $\beta_-$  and  $\beta_+$  are the corresponding Lagrange's multipliers.

## Covariance differentiability and a scale-space computing approach

The nonlinear dynamics of clustering, in this work is derived from the model of "free energy", originally used in statistical physics to model different complex systems. The free energy describes the state of a cluster for a given parameter  $\beta$ ,

$$F(\beta_{+/-}) = -\frac{1}{\beta_{+/-}} \log_r Z_{+/-}.$$

The parameter  $\beta$  is inversely proportional to temperature ( $\beta=1/T$ ), in physical analogy. The equilibrium states are computed to minimize the energy exchange among clusters for a given spatial distribution of the clusters.

The distortion measure, applied in the algorithm, is chosen to be the linear constraint equation on the motion vector  $\vec{v}$ , also known as the extended optical flow constraint equation:

$$d = z^2 = (I_t + \nabla I \vec{v})^2 + I \text{div}(\vec{v}),$$

which provides the mass conservation principle [1]. In this work the coherency of data is estimated with its Green's function, to control the smoothness of the optical flow adaptively in the scale.

The constrained equation of motion for a coupled pair of clusters is given by:

$$\begin{aligned} \dot{\vec{v}}_1 &= \mp \frac{\partial F_1}{\partial \vec{v}_1} \pm \beta \frac{\partial F_2}{\partial \vec{v}_2} \\ \dot{\vec{v}}_2 &= \mp \frac{\partial F_2}{\partial \vec{v}_2} \pm \beta \frac{\partial F_1}{\partial \vec{v}_1} \\ \beta' &= \pm \frac{\partial F}{\partial \beta} \end{aligned}$$

where the upper sign corresponds to the cooling part and lower to the melting part of the algorithm.

This system of equations can be analyzed by the series expansion of the system's free energies:

$$\begin{aligned} \Delta F &= [\Delta_1 \quad \Delta_2] \begin{bmatrix} F_1 \mp \beta F_2 \\ F_2 \mp \beta F_1 \end{bmatrix} \\ &= \left( \frac{\partial F_1}{\partial \vec{v}_1} \delta \vec{v}_1 + \frac{\partial F_2}{\partial \vec{v}_2} \delta \vec{v}_2 \right) + \left( \frac{\partial F_1}{\partial \beta_1} \delta \beta_1 + \frac{\partial F_2}{\partial \beta_2} \delta \beta_2 \right) \\ &+ \frac{1}{2} (\delta \vec{v}_1 \quad \delta \vec{v}_2) \begin{bmatrix} \frac{\partial^2 F_1}{\partial \vec{v}_1^2} & \mp \beta \frac{\partial^2 F_2}{\partial \vec{v}_2^2} \\ \mp \beta \frac{\partial^2 F_1}{\partial \vec{v}_1^2} & \frac{\partial^2 F_2}{\partial \vec{v}_2^2} \end{bmatrix} (\delta \vec{v}_1 \quad \delta \vec{v}_2)^T \\ &+ \frac{1}{2} \left( \frac{\partial^2 F_1}{\partial \beta_1^2} \delta \beta_1^2 + \frac{\partial^2 F_2}{\partial \beta_2^2} \delta \beta_2^2 \right) \mp (F_1 \Delta_2 \beta + F_2 \Delta_1 \beta) + \mathcal{O}^*(2). \end{aligned}$$

This gives an update formula for the parameter  $\beta$ :

$$\begin{aligned} \Delta \beta_- &= + \sqrt{\frac{\delta \vec{v}_1 \frac{\partial^2 F_1}{\partial \vec{v}_1^2} \delta \vec{v}_1^T - \beta \delta \vec{v}_1 \frac{\partial^2 F_2}{\partial \vec{v}_2^2} \delta \vec{v}_2^T}{1 + \frac{1}{2} \frac{\partial^2 F_1}{\partial \beta^2}}} \\ &+ \sqrt{\frac{\delta \vec{v}_2 \frac{\partial^2 F_2}{\partial \vec{v}_2^2} \delta \vec{v}_2^T - \beta \delta \vec{v}_1 \frac{\partial^2 F_1}{\partial \vec{v}_1^2} \delta \vec{v}_1^T}{1 + \frac{1}{2} \frac{\partial^2 F_2}{\partial \beta^2}}}, \end{aligned}$$

for the cooling part of the algorithm, and

$$\begin{aligned} \Delta \beta_+ &= - \sqrt{\frac{\delta \vec{v}_1 \frac{\partial^2 F_1}{\partial \vec{v}_1^2} \delta \vec{v}_1^T + \beta \delta \vec{v}_1 \frac{\partial^2 F_2}{\partial \vec{v}_2^2} \delta \vec{v}_2^T}{1 + \frac{1}{2} \frac{\partial^2 F_1}{\partial \beta^2}}} \\ &- \sqrt{\frac{\delta \vec{v}_2 \frac{\partial^2 F_2}{\partial \vec{v}_2^2} \delta \vec{v}_2^T + \beta \delta \vec{v}_1 \frac{\partial^2 F_1}{\partial \vec{v}_1^2} \delta \vec{v}_1^T}{1 + \frac{1}{2} \frac{\partial^2 F_2}{\partial \beta^2}}}, \end{aligned}$$

for the melting part.

Note that this way we keep the integral:

$$\int_s dU = \int_s \frac{\partial F}{\partial \beta} d\vec{v} + \frac{\partial V}{\partial \vec{v}} d\beta = 0, \quad (7)$$

The above system of equations results in the determinant of the map:

$$\mathbf{D} = \Lambda^2 + \left( \frac{\partial^2 F_1}{\partial \bar{v}_1^2} + \frac{\partial^2 F_2}{\partial \bar{v}_2^2} \right) \Lambda + (1 - \beta^2) \frac{\partial^2 F_1}{\partial \bar{v}_1^2} \cdot \frac{\partial^2 F_2}{\partial \bar{v}_2^2}.$$

For  $0 \leq \beta < 1$  the eigenvalues of the determinant of the map have negative values if the Hessians of the free energies,  $F_1$  and  $F_2$ , are positive definite, what gives a condition of numerical stability of the coupled system's equations.

## Scale-space pathways: resonance computing

Singularity of data clusters is evaluated by the means of the scalability of the maps of feature vector representatives in scale space. For a given maximal value of the distortion energy the minimal number of singularity manifolds is obtained in the hierarchical scale computation.

For a given cluster, the partition function  $Z(\beta, r)$  describes the distribution of the data points with respect to the cluster vector representative. The scale information is evaluated by conjoining the two parameters: the signal energy distortion scale parameter  $\beta$ , and the spatial scale parameter  $r$ , which equals to the number of data points inside the spatial window of computation  $W$ . For a given point distortion measure of the signal,  $z^2 = d(\bar{x}, \bar{y})$ , the partition function is written by:

$$Z(\beta, r) = \int_W r^{\text{sign}^* \beta z^2}$$

and, a data point belongs to the cluster in probability, with the probability density function:

$$P = \frac{r^{\text{sign}^* \beta z^2}}{Z}$$

The partition function can be conveniently written as:

$$Z(\beta) = r^{\text{sign}^* \beta F} = r^{\text{sign}^* (\beta V - 1)} = r^{\int_0^{\beta} \text{sign}^* V(\beta) d\beta}.$$

This function is a multifractal measure, giving a way of decomposing the signal into feature vector clusters ordered by the singularity exponents,  $H = 2\beta V < 1$ , and singular frequency factors,  $F/V < 1$ , as written in:

$$Z = r^{-\frac{1}{2}H\left(\frac{F}{V}\right)}.$$

The nonlinear dynamics of clustering is governed by the two energy functions. For a given cluster its free energy, and the distortion energy is defined by:

$$F(\bar{v}, \beta) = -\frac{1}{\beta} \log_r Z, \quad V = \int_W d(\bar{x}, \bar{y}) P.$$

We relate the mechanism of the multifractal decomposition of the signal to the stability analysis of the map.

The stability condition of this map is given by the relation:  $2\beta V < 1$ . We limit the singularity exponent of the clusters with the maximal value  $H$ , by splitting that cluster in two for which the condition is reached:  $2\beta_c V = H$ , at the critical value of the scale parameter  $\beta_c$ .

Interesting points of observation become the “signal energy levels”:

$$F - V = \text{const.}, \quad (\text{A.1})$$

and its propagation in time. The time derivative of this equation becomes:

$$\begin{aligned} & \left[ \left( \frac{F}{V} - 1 \right) \frac{\partial}{\partial \beta} \log V + \frac{\partial}{\partial \beta} \left( \frac{F}{V} \right) \right] \dot{\beta} \\ & = - \left[ \left( \frac{F}{V} - 1 \right) \frac{\partial}{\partial \bar{v}} \log V + \frac{\partial}{\partial \bar{v}} \left( \frac{F}{V} \right) \right] \dot{\bar{v}} \end{aligned} \quad (\text{A.2})$$

The Green's formula:

$$\int_S \frac{\partial F}{\partial \beta} d\bar{v} + \frac{\partial V}{\partial \bar{v}} d\beta = 0,$$

as well as, the accompanying wave equation:

$$\frac{\partial^2 F}{\partial \beta^2} = \nabla^2 V,$$

are used in the computational scheme for the estimation of the clustering parameters, and are still part of the ongoing research.

At the “signal scale equilibrium”  $\frac{\partial F}{\partial \beta} = 0$ , an isolated cluster

can be modeled by the equations:

$$\frac{\partial V}{\partial \beta} = \beta \nabla^2 V,$$

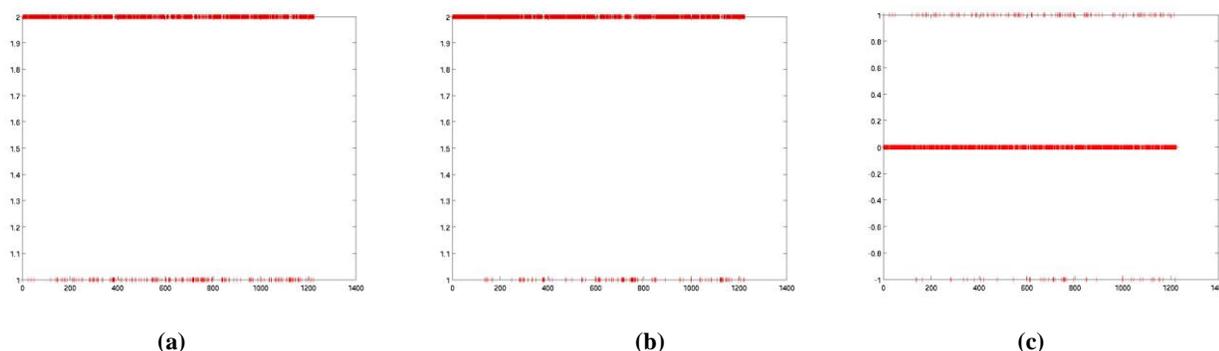
and,

$$\int_S \frac{\partial V}{\partial \bar{v}} d\beta = \int_S \frac{\|\nabla V\|^2}{\beta \nabla^2 V} d\bar{v} = 0.$$

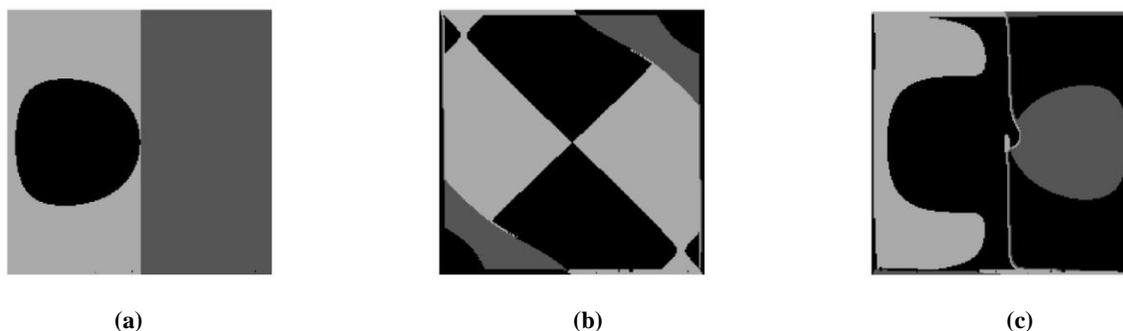
The Green's function gives a model of spatial coherency of information for a data cluster:

$$G(\beta, \bar{v}) = \frac{\|\nabla V\|^2}{\beta \nabla^2 V},$$

and is applied locally in the segmentation process.



**Figure 1:** Clustering of chemical data: 1225 observations in 155 dimensional property space. Statistical maps projections along the most singular components: (a) 1-dimensional, and (b) 2-dimensional. The observations are labeled with ‘1’ and ‘2’ to indicate a membership to the corresponding clusters. The difference in the membership labels in (a) and (b) is shown in figure 1.(c).



**Figure 1:** Clustering of the expanding ball image sequence. 3 clusters statistical maps projections along the singular components with: (a) divergence, (b) rotor, and, (c) divergence and rotor point segmentation.

### 3. RESULTS

A dataset of 1236 compounds with 155 real-valued descriptors is used as a test case. The observations on the solubility of the compounds are analyzed in the segmentation algorithm. In the 1-dimensional case, the 1236 compounds are segmented in two groups with:  $N_1 = 226$ ,  $N_2 = 1010$  groups’ members. Corresponding clusters’ variances are:  $V_1 = 25 \times 10^8$ , and  $V_2 = 30 \times 10^8$ , respectively. The resulting distribution of the data is shown in Figure 1(a).

In the 2-dimensional case we computed spatial map out of the 2 most singular components of the data. 1225 compounds were segmented in two groups with:  $N_1 = 158$ ,  $N_2 = 1067$  members. And the resulting variances of the groups decreased as:  $V_1 = 19 \times 10^8$ , and  $V_2 = 28 \times 10^8$ , respectively. The resulting clusters’ membership distribution of the data is shown in Figure 1(b). The difference in clustering as in (a) and (b) is shown in Figure 1.(c).

A test image sequence of an expanding ball is used as an example pattern for 2-dimensional signal decomposition. The statistical maps of the clustered 3 image segments are shown in Figure 2. Differential operators: divergence, rotor, and a combination of the two are used for clustering data. The resulting membership distributions of data points are shown on Figures 2.(a-c) for these operators, respectively.

### 4. CONCLUDING REMARKS

We have described a new method for multi-dimensional data scaling, in data mining and knowledge discovery applications. The hierarchical scale decomposition into different singular sets is obtained by evaluating scale-space scalability of singular features for the clusters of data points. The information content is evaluated by the scale-space frequencies of corresponding data.

The Green's function expresses spatial coherency of data clusters and is used as an integration function for the segmenting data by applying local operators. In the method proposed, a decomposition in harmonic data sets is achieved by hierarchical scale computation. We use dynamical cascade diagrams to form a knowledge base discovered in the data, by this method.

We have run the algorithm only by applying the solubility as an additional property value in the data set. The improvements can be achieved by running the differential version segmentation on the solubility property. The associated structure diagrams can be used to relate different subsets of property values, and therefore making a knowledge base suitable for coding and control purposes.

We intend to investigate this tool in data mining applications – its coding and control structure implementation. The optimization of the space-energy exchange step along with the refinement of our numerical schemes is a part of our ongoing research work, as well.

## ACKNOWLEDGEMENT

This work has been performed in the Pervasive Technology Labs' - Community Grids Lab, at the Indiana University.

## 5. APENDIX

---

**Main:** decomposition – basic algorithm:  $K_{max} = 3$

1. **Divide data** evenly:  $N/N_p$  to each core,  $i(p)$  and  $c = 1$ , indexed  
Start with:  $Couple = 0$ .

2. **For**  $K \leq K_{max}$  // Recursive procedure

3. **Do gradual descent until resonate**

*Covar\_project* (); //Compute parameters:  $3D + 2*d_{cov}$  singular components

*Cls\_map* ( $\Delta\beta_+$ ,  $\Delta\beta_-$ ,  $\Delta Space$ ); // Adjust with Green's functions

**For all** the clusters:

*Cls\_equilib* ( $Sign = \pm 1, Couple$ ); // Equilibrate clusters

**For all** the clusters:

*Cls\_covar* ( $Sign = \pm 1$ ); // Covariance computing

*Cls\_svmdecomp* ( $Sign = \pm 1$ ); // Singular values decomp.

### 4. Test stability

**IF** the most spatially coherent cluster *unstable*

Split in 2 clusters,  $K++$

**ELSE IF** a cluster *empty*

Merge in 1,  $K--$

**ELSE** *Resonance*

---



---

*Cls\_equilib* ( $Sign, Coupling$ )

**Do gradual descent until converge for all** the clusters with partial sums:

1. **Parallel\_Sum\_1** ( $p, \bar{y}_c$ )

**For all** the clusters  $c$ :

$$\rho_i^c = e^{-\frac{Sign}{T}(\bar{x}_i - \bar{y}_c)^2}$$

$$Z_c(p) = \sum_{i(p)} \rho_i^c,$$

$$\text{and } \bar{g}_c(p) = \sum_{i(p)} (\bar{y}_c - \bar{x}_i) \rho_i^c$$

2. **Sequential\_update\_1** ()

$$\bar{y}_c = \bar{y}_c - Sign * \frac{\sum \bar{g}_c(p)}{\sum_p Z_c(p)} + Sign * Coupling \left( \frac{1}{T}, \bar{y}_{couple} \right)$$


---



---

*Cls\_covar* ( $Sign$ );

1. **Parallel\_Sum\_2** ( $p, \bar{y}_c, CC^c$ )

**For all** the clusters  $c$ , and data dimension  $d_{cov}$ :

$$\rho_i^c = e^{-\text{Sign}^* \frac{1}{T} (\bar{x}_i - \bar{y}_c)^2}$$

$$CC_{d_1, d_2}^c(p) = \sum_{i(p)} (y_c^{d_1} - x_i^{d_1}) (y_c^{d_2} - x_i^{d_2}) \rho_i^c$$

$$Z(p) = \sum_k \sum_{i(p)} \rho_i^k$$

## 2. Sequential\_update\_2 ( )

$$CC^c = \frac{\sum_p CC^c(p)}{\sum_p Z(p)}$$

## 6. REFERENCES

- [1] D.Bereziat, and J-P.Berroy, "Motion Estimation on Meteorological Infrared Data using Total Brightness Invariance Hypothesis," *Environmental Modeling System*, 2000.
- [2] M.Pierce, G.C.Fox, et al. "[Smart Mining of Drug Discovery Information: 1. A web service and workflow infrastructure](#)," 2006.
- [3] R.Courant, and D.Hilbert, *Methods of Mathematical Physics*, John Wiley & Sons, New York, 1962.
- [4] U.Grenander, and M.Miller, *Pattern Theory: From Representation to Inference*, Oxford University Press, 2007.
- [5] F.Hoppensteadt, and Elzhikevich, "Neural Computations by Networks and Oscillators," *IJCNN*, vol. 4, p. 4041, 200.
- [6] B.K.P.Horn, and B.G.Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
- [7] E.T.Jaynes, "Information theory and statistical mechanics," in *Papers on probability, statistics and statistical physics* (R.D. Rosenkrantz, ed.), *Kulwer Academic Publishers*, 1989.
- [8] M.Jovovic, "A Markov random fields model for describing unhomogeneous textures: generalized random stereograms," *IEEE Workshop Proceedings on Visualization and Machine Vision, and Biomedical Image Analysis*, Seattle, 1994.
- [9] M.Jovovic, "Image segmentation for feature selection from motion and photometric information by clustering," *SPIE Symposium on Visual Information Processing V*, Orlando, 1996.
- [10] M.Jovovic, "A multiscale processing of data streams," (in English) *Informacione Tehnologije IV*, Zabljak, Yugoslavia 1999.
- [11] M.Jovovic, S.Jonic, D.Popovic, "Automatic synthesis of synergies for control of reaching - hierarchical clustering," *Med. Eng. and Physics*, vol. 21/5, pp. 325-337, 1999.
- [12] M.Jovovic, "Space-Color Quantization of Multispectral Images in Hierarchy of Scales," *ICIP01*, vol. I, pp. 914-917, Thessaloniki, Greece, 2001.
- [13] M.Jovovic, H.Yahia, I.Herlin, "Hierarchical scale decomposition of images - singular feature analysis," *Technical report*, INRIA, 2003.
- [14] M.Jovovic, "Texture segmentation by 2-layers recurrent dynamics," *XXVIIIth International Symposium in COMPUTATIONAL NEUROSCIENCE*, Montreal. 2006.
- [15] K.Rose, E.Gurewitz, and G.C.Fox, "A Deterministic Annealing Approach to Clustering," *Pattern Recog. Lett.* vol. 11, pp. 589-594, 1990.
- [16] A.Turiel, and A. del Pozo, "Reconstructing Images From Their Most Singular Fractal Manifold," *IEEE Trans. Image Processing*, vol. 11/4, pp. 345-350, 2002.
- [17] S.Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Springer-Verlag, New York, 1990