Using Directional Curvatures to Visualize Folding Patterns of the GTM Projection Manifolds

Peter Tiňo, Ian Nabney, and Yi Sun

Neural Computing Research Group, Aston University,
Aston Triangle, Birmingham B4 7ET
United Kingdom
{tinop,nabneyit,suny}@aston.ac.uk
http://www.ncrg.aston.ac.uk/

Abstract. In data visualization, characterizing local geometric properties of non-linear projection manifolds provides the user with valuable additional information that can influence further steps in the data analysis. We take advantage of the smooth character of GTM projection manifold and analytically calculate its local directional curvatures. Curvature plots are useful for detecting regions where geometry is distorted, for changing the amount of regularization in non-linear projection manifolds, and for choosing regions of interest when constructing detailed lower-level visualization plots.

1 Introduction

Most visualization algorithms project points from a high-dimensional data space onto a two-dimensional projection space. Loosely speaking, algorithms like self-organizing maps (SOM) [7] or the Generalized Topographic Mapping (GTM) [6], a probabilistic reformulation of the SOM, identify the computer screen with a two-dimensional "rubber sheet" that is injected into the high-dimensional data space. The sheet is supposed to "cover" the cloud of data points by locally stretching, contracting and curving. The visualization plot is obtained by first projecting the data points onto the rubber sheet and then letting the rubber sheet relax to its original form of the computer screen. We refer to the injected (possibly curved and stretched) two-dimensional rubber sheet in the data space as the *projection manifold*.

Besides the visualization plot itself, the user is often interested in additional information about the structure of the projection manifold in the high-dimensional data space. For example, local magnification factors describe how small regions on the computer screen are stretched or compressed when mapped to the data space. Magnification factors can be used for detecting (on the visualization plot) separate clusters in the data space (see [4]).

SOM discretizes the "rubber sheet" into a grid of "nodes" and represents the projection manifold only by a grid of nodes mapped to the data space (code-book vectors). On the other hand, GTM forms a *smooth* two-dimensional projection manifold. This allows us to use techniques of differential geometry to analytically describe (local) geometric properties *anywhere* on the manifold. Local magnification factors of GTM models were *analytically* computed by Bishop, Svensén and Williams in [4] [5]. In SOM, the magnification factors can only be approximated by distances between the code-book vectors.

Magnification factors represent the extent to which the areas are magnified on projection to the data space. However, when injecting a two-dimensional rubber sheet into a high dimensional data space, the projection manifold may form complicated folds that cannot be detected by using magnification factors alone. To provide the user with a tool for monitoring the amount of folding and neighborhood preservation in the projection manifold, we need second-order quantities, such as local curvatures. Neighborhood preservation and folding issues in the context of SOM were studied e.g. in [2] [11]. Such studies present largely heuristic techniques for computing higher-order geometric properties of the SOM projection manifold based on a discrete grid of code-book vectors (representing the manifold) in the data space. In contrast, as shown in this paper, the smooth nature of the GTM projection manifold allows us to analytically compute directional curvatures in any point on the manifold.

2 Generative Topographic Mapping

The Generative Topographic Mapping (GTM) belongs to a family of latent space models that model a probability distribution in the (observable) data space by means of latent, or hidden variables. For the purposes of data visualization, we identify the visualization space (i.e. the "rubber sheet") with the latent space. The latent space is usually a bounded subset of the two-dimensional Euclidean space, such as the (two-dimensional) interval $[-1, 1] \times [-1, 1]$.

Consider an L-dimensional latent space $\mathcal{H} \subset \Re^L$ and represent points in $\mathcal{H}$ as column vectors $\mathbf{x} = (x_1, x_2, ..., x_L)^T$. We allow $\mathcal{H}$ to be covered by an array of K latent space centres $\mathbf{x}_i \in \mathcal{H}$, $i = 1, 2, ..., K$.

Let the data space $\mathcal{D}$ be the D-dimensional Euclidean space $\Re^D$. We define a non-linear transformation $f : \mathcal{H} \to \mathcal{D}$ from the latent space to the data space using a radial basis function network (see e.g. [3]). To this end, we cover the latent space with a set of M fixed non-linear basis functions $\phi_j : \mathcal{H} \to \Re$, $j = 1, 2, ..., M$. As usual in the GTM literature, we choose to work with spherical Gaussian functions of the same width σ, although other choices are possible and require only simple modifications[1]. Usually, the centres of the Gaussian basis functions ϕ_j are positioned in the latent space on a regular grid. Given a point $\mathbf{x} \in \mathcal{H}$ in the latent space, its image under the map f is[2]

$$f(\mathbf{x}) = \mathbf{W}\ \phi(\mathbf{x}), \tag{1}$$

[1] GTM with other choices of basis functions can be easily constructed using NETLAB available from `http://www.ncrg.aston.ac.uk/netlab/`

[2] We assume that the data set has been normalized to zero mean. Equivalently, we could include a constant basis function $\phi_0(\mathbf{x}) = 1$.

where $\mathbf{W}$ is a $D \times M$ matrix of weight parameters and $\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), ..., \phi_M(\mathbf{x}))^T$.

GTM creates a generative probabilistic model in the data space by placing a radially-symmetric Gaussian with zero mean and inverse variance β around images, under f, of the latent space centres $\mathbf{x}_i \in \mathcal{H}$, $i = 1, 2, ..., K$:

$$P(\mathbf{t}|\ \mathbf{x}_i, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left\{-\frac{\beta}{2}\|f(\mathbf{x}_i) - \mathbf{t}\|^2\right\}. \tag{2}$$

Defining a uniform prior over $\mathbf{x}_i$, the density model in the data space provided by the GTM is

$$P(\mathbf{t}|\ \mathbf{W}, \beta) = 1/K \sum_{i=1}^{K} P(\mathbf{t}|\ \mathbf{x}_i, \mathbf{W}, \beta). \tag{3}$$

Given a data set $\zeta = \{\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_N\}$ of independently generated points in the data space, the adjustable parameters $\mathbf{W}$ and β of the model can be fitted to the data by maximum likelihood using the expectation-maximization algorithm [6].

For the purpose of data visualization, we use Bayes' theorem to invert the transformation f from the latent space $\mathcal{H}$ to the data space $\mathcal{D}$. The posterior distribution on $\mathcal{H}$, given a data point $\mathbf{t}_n \in \mathcal{D}$, is a sum of delta functions centered at centres $\mathbf{x}_i$, with coefficients equal to the posterior probability R_{in} that the i-th Gaussian (corresponding to the latent space centre $\mathbf{x}_i$, eq. (2)) generated $\mathbf{t}_n$ [6]. The latent space representation of the point $\mathbf{t}_n$, i.e. the *projection of* $\mathbf{t}_n$, is taken to be the mean $\sum_{i-1}^{K} R_{in}\ \mathbf{x}_i$ of the posterior distribution on $\mathcal{H}$.

The f–image of the latent space $\mathcal{H}$,

$$\Omega = f(\mathcal{H}) = \{f(\mathbf{x}) \in \Re^D |\ \mathbf{x} \in \mathcal{H}\}, \tag{4}$$

forms a *smooth* L-dimensional manifold in the data space. We refer to the manifold Ω as the *projection manifold* of the GTM.

3 Local Directional Curvatures

The idea of directional curvature is explained in figure 1.

Consider a point $\mathbf{x}_0 \in \mathcal{H}$. Let $\mathbf{x}(b)$, $b \in \Re$, be a straight line passing through $\mathbf{x}_0$ along a unit directional vector $\mathbf{h} = (h_1, h_2, ..., h_L)^T$:

$$\mathbf{x}(b) = \mathbf{x}_0 + b\mathbf{h}, \quad b \in \Re. \tag{5}$$

As the parameter b varies, the image of the line $\mathbf{x}(b)$ generates on Ω the curve

$$\mu(b) = f(\mathbf{x}(b)). \tag{6}$$

The tangent to this curve at $f(\mathbf{x}_0) = \mu(0)$ is

$$\begin{aligned} \dot{\mu}(0) &= \left[\frac{\mathrm{d}\ \mu(b)}{\mathrm{d}\ b}\right]_{b=0} = \left[\sum_{r=1}^{L} \frac{\partial f(\mathbf{x})}{\partial x_r}\ \frac{\mathrm{d}\ x_r(b)}{\mathrm{d}\ b}\right]_{\mathbf{x}=\mathbf{x}_0, b=0} \\ &= \sum_{r=1}^{L} \mathbf{\Gamma}_r^{(1)}\ h_r = \mathbf{\Gamma}^{(1)}\ \mathbf{h}, \end{aligned} \tag{7}$$

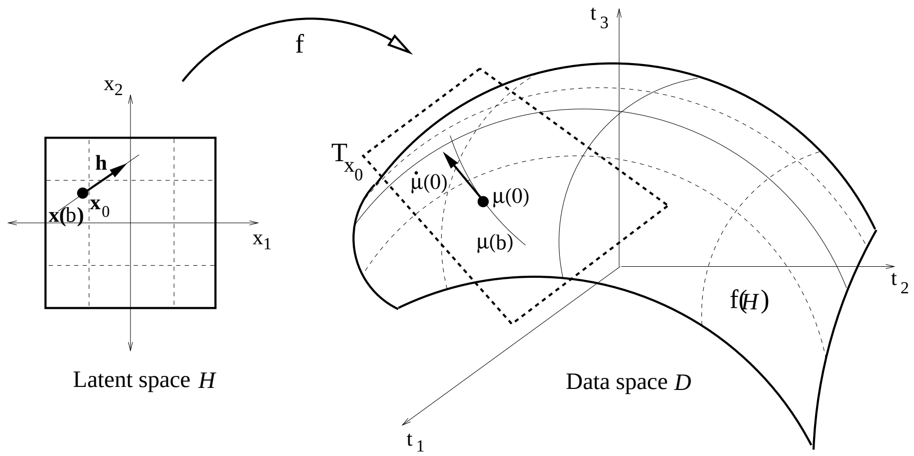


Fig. 1. An explanation of local directional derivative of the visualization manifold. A straight line $\mathbf{x}(b)$ passing through the point $\mathbf{x}_0$ in the latent space $\mathcal{H}$ is mapped via f to the curve $\mu(b) = f(\mathbf{x}(b))$ in the data space $\mathcal{D}$. Curvature of μ at $f(\mathbf{x}_0) = \mu(0)$ is related to the directional curvature of the projection manifold $f(\mathcal{H})$ with respect to the direction $\mathbf{h}$. The tangent vector $\dot{\mu}(0)$ to μ at $\mu(0)$ lies in $\mathbf{T}_{\mathbf{x}_0}$ (dashed rectangle), the tangent plane of the manifold $f(\mathcal{H})$ at $\mu(0)$.

where

$$\mathbf{\Gamma}_r^{(1)} = \mathbf{W} \left(\frac{\partial \phi_1(\mathbf{x}_0)}{\partial x_r}, \frac{\partial \phi_2(\mathbf{x}_0)}{\partial x_r}, ..., \frac{\partial \phi_M(\mathbf{x}_0)}{\partial x_r} \right)^T \tag{8}$$

is a (column) vector of partial derivatives of the GTM map f (at $\mathbf{x}_0 \in \mathcal{H}$) with respect to the r-th latent space variable x_r, and $\mathbf{\Gamma}^{(1)}$ is the $D \times L$ matrix

$$\mathbf{\Gamma}^{(1)} = [\mathbf{\Gamma}_1^{(1)}, \mathbf{\Gamma}_2^{(1)}, ..., \mathbf{\Gamma}_L^{(1)}]. \tag{9}$$

The tangent vector $\dot{\mu}(0)$ to the lifted line $\mu(b)$ is a linear combination of the columns of $\mathbf{\Gamma}^{(1)}$, and so the range of the matrix $\mathbf{\Gamma}^{(1)}$ is the tangent plane $\mathbf{T}_{\mathbf{x}_0}$ of the projection manifold Ω at $f(\mathbf{x}_0) = \mu(0)$.

The second directional derivative [9] of $\mu(b)$ at $\mu(0)$ is

$$\begin{aligned} \ddot{\mu}(0) &= \left[\sum_{s=1}^{L} \frac{\partial}{\partial x_s} \left\{ \sum_{r=1}^{L} \frac{\partial f(\mathbf{x})}{\partial x_r} \; h_r \right\} \frac{\mathrm{d} \; x_s(b)}{\mathrm{d} \; b} \right]_{\mathbf{x}=\mathbf{x}_0, b=0} \\ &= \left[\sum_{r=1}^{L} \sum_{s=1}^{L} \frac{\partial^2 f(\mathbf{x})}{\partial x_r \partial x_s} \; h_r h_s \right]_{\mathbf{x}=\mathbf{x}_0} = \sum_{r=1}^{L} \sum_{s=1}^{L} \mathbf{\Gamma}_{r,s}^{(2)} \; h_r \; h_s, \end{aligned} \tag{10}$$

where $\mathbf{\Gamma}_{r,s}^{(2)}$ is a column vector of second-order partial derivatives of f (at $\mathbf{x}_0 \in \mathcal{H}$) with respect to the r-th and s-th latent space variables,

$$\mathbf{\Gamma}_{r,s}^{(2)} = \mathbf{W} \left(\frac{\partial^2 \phi_1(\mathbf{x}_0)}{\partial x_r \partial x_s}, \frac{\partial^2 \phi_2(\mathbf{x}_0)}{\partial x_r \partial x_s}, ..., \frac{\partial^2 \phi_M(\mathbf{x}_0)}{\partial x_r \partial x_s} \right)^T. \tag{11}$$

We decompose $\ddot{\mu}(0)$ into two orthogonal components, one lying in the tangent space $\mathbf{T}_{\mathbf{x}_0}$, the other lying in its orthogonal complement $\mathbf{T}^{\perp}_{\mathbf{x}_0}$,

$$\ddot{\mu}(0) = \ddot{\mu}^{\|}(0) + \ddot{\mu}^{\perp}(0), \qquad \ddot{\mu}^{\|}(0) \in \mathbf{T}_{\mathbf{x}_0},\ \ddot{\mu}^{\perp}(0) \in \mathbf{T}^{\perp}_{\mathbf{x}_0}. \tag{12}$$

The component $\ddot{\mu}^{\|}(0)$ describes changes in the first-order derivatives due to "varying speed of parameterization", while the direction of the first-order derivatives remains unchanged. Changes in the first-order derivatives that are responsible for curving of the projection manifold Ω are described by the component $\ddot{\mu}^{\perp}(0)$.

Orthogonal projection onto $\mathbf{T}_{\mathbf{x}_0}$ is a linear operator described by the projection matrix $\mathbf{\Pi} = \mathbf{\Gamma}^{(1)} \left(\mathbf{\Gamma}^{(1)}\right)^{+}$, where $\left(\mathbf{\Gamma}^{(1)}\right)^{+}$ is the Moore-Penrose generalized inverse of $\mathbf{\Gamma}^{(1)}$ (see e.g. [8]). So, $\ddot{\mu}^{\perp}(0) = (\mathbf{I} - \mathbf{\Pi})\ \ddot{\mu}(0)$, where $\mathbf{I}$ is the $D \times D$ identity matrix.

The directional curvature at $\mu(0)$ associated with the latent space direction $\mathbf{h}$ is the (Euclidean) norm of the vector $\ddot{\mu}^{\perp}(0)$. It measures the degree to which the visualization manifold Ω (locally) "curves" in the data space $\mathcal{D}$ [1]. It is the embedding curvature of $\Omega \subset \mathcal{D}$ at $f(\mathbf{x}_0)$, evaluated with respect to the latent space direction $\mathbf{h}$.

4 Experiments

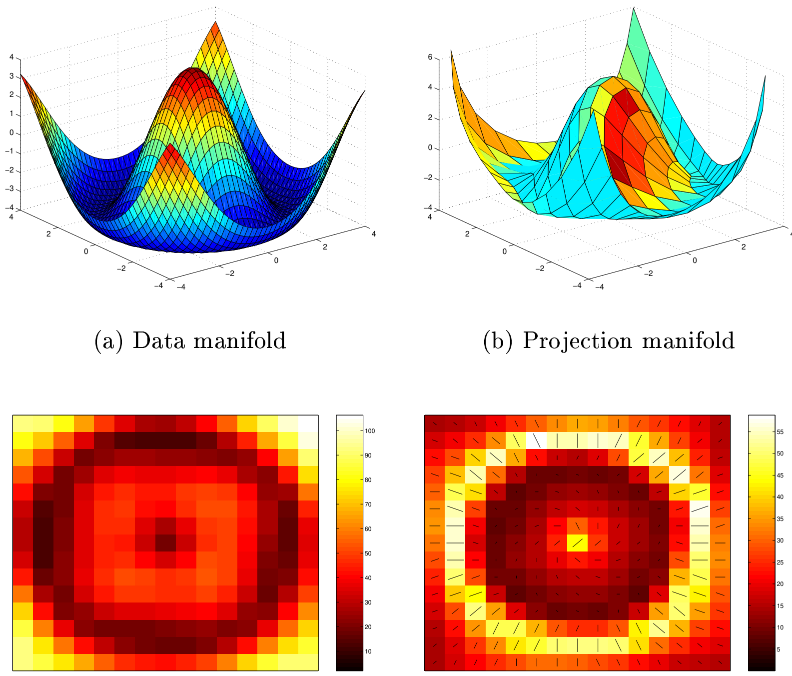
In the experiments reported here, the GTM latent space $\mathcal{H}$ was the square $\mathcal{H} = [-1, 1] \times [-1, 1]$, the latent space centres $\mathbf{x}_i \in \mathcal{H}$ were positioned on a regular 15×15 square grid and there were 16 basis functions ϕ_j centered on a regular 4×4 square grid. The basis functions were spherical Gaussian functions of the same width $\sigma = 0.44$. Magnification factors and directional curvatures were evaluated at each latent space centre $\mathbf{x}_i$.

In the first experiment we randomly generated 2000 points in $\Re^3$ lying on the two-dimensional manifold shown in figure 2a. As expected, after training, the GTM projection manifold shown in figure 2b closely followed the two-dimensional distribution of the data points. Latent space layouts of local magnification factors and directional curvatures are shown in figures 2c and 2d, respectively. The magnification factor at a centre $\mathbf{x}_i$, which is the Jacobian of the GTM map f at $\mathbf{x}_i$ [4], is represented by the degree of shading of the corresponding patch. In the curvature plot, we show for each latent space centre $\mathbf{x}_i$, the direction $\mathbf{h}$ yielding the maximal norm of $\ddot{\mu}^{\perp}(0)$. The length of the direction line and the degree of shading of the corresponding patch are proportional to the maximal norm of $\ddot{\mu}^{\perp}(0)$.

The curvature and expansion patterns in the projection manifold (figure 2b) are clearly reflected in the magnification factor and curvature plots (figures 2c,d). The form of the projection manifold can be approximately guessed on the basis of its local first- and second-order characterizations.

In the second experiment, we trained GTM on an oil flow data set[3]. This 12-dimensional data set arises from a physics-based simulation of a non-invasive

[3] `http://www.ncrg.aston.ac.uk/GTM/3PhaseData.html`



(a) Data manifold (b) Projection manifold

(c) Magnification factors (d) Curvatures

Fig. 2. Toy data experiment.

monitoring system, used to determine the quantity of oil in a multi-phase pipeline containing a mixture of oil, water and gas. The data set consists of 1000 points obtained synthetically by simulating the physical process in the pipe. Points in the data set are classified into three different multi-phase flow configurations, namely *homogeneous*, *annular* and *laminar*.

The top level curvature plot in figure 3b reveals that the two-dimensional projection manifold folded three times in order to "capture" the distribution of points in the 12-dimensional space. Interestingly, the three multi-phase flow configurations seem to be roughly separated by the folds (compare the top level visualization plot in figure 3a with the corresponding curvature plot). We confirmed this hypothesis by constructing three local lower level visualization plots initiated in the regions between the folds. The lower level plots correspond to a mixture of GTMs (see [10]) and are shown in figure 3a. The lower level plots are numbered left-to-right and were initiated in points shown as circles in the top level plot (the number inside each circle indicates the index of the corresponding lower level plot). Curvature plots of the lower level GTMs reveal that, compared to the top level GTM, the lower level projection manifolds are almost flat. We do not show the magnification factor plots here, but there are no significant expansive/contractive tendencies in the lower level projection manifolds. In this example the use of curvature information is crucial to successful modeling.

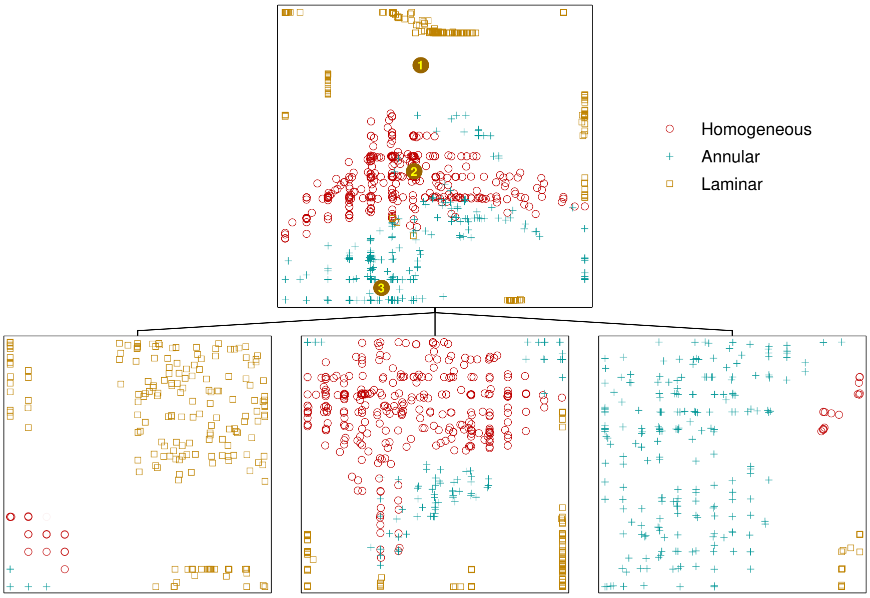
5 Conclusion

Compared to linear projection methods such as PCA, non-linear visualization techniques are more capable of revealing the nature of data distribution in a high-dimensional space. Characterization of local geometric properties of the non-linear projection manifolds provides user with a valuable additional information that can influence further steps in the data analysis.

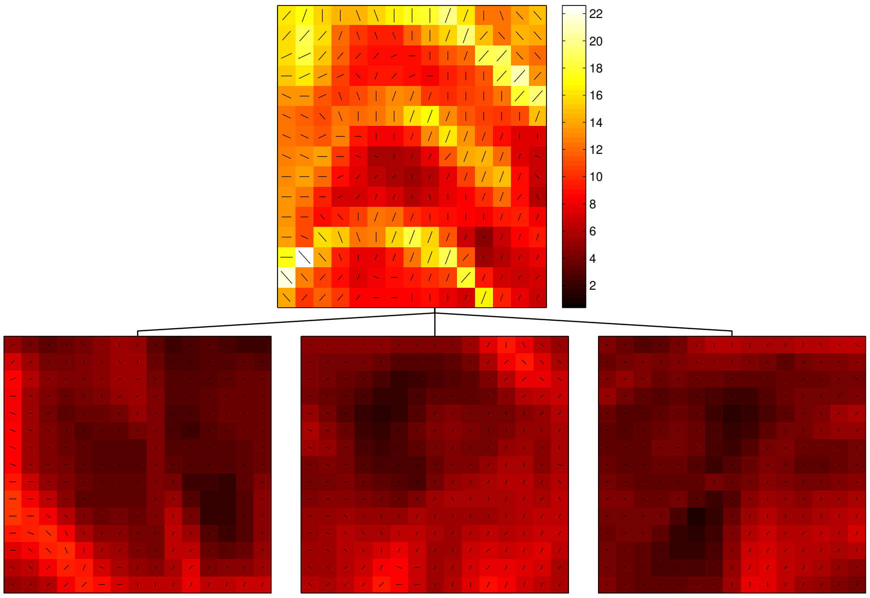
In this paper, we extended the work of Bishop, Svensén and Williams [4] [5] on magnification factors in GTM describing local expansion tendencies of the projection manifold. We analytically calculate and graphically represent local directional curvatures of the GTM manifold. Curvature plots are useful for detecting regions where geometry is distorted, for changing the amount of regularization in non-linear projection manifolds and, as illustrated in our experiment, for choosing regions of interest when constructing detailed lower-level visualization plots.

References

1. Bates, D.M., Watts, D.G.: Relative curvature measures of nonlinearity (with Discussion). J. R. Stat. Soc. B **42** (1980) 1–25
2. Bauer, H.U., Pawelzik, K.: Quantifying the neighborhood preservation of self-organizing feature maps. IEEE Transactions on Neural Networks **3** (1992) 570–579
3. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford, UK (1995)
4. Bishop, C.M., Svensén, M., Williams, C.K.I.: Magnification Factors for the SOM and GTM Algorithms. In: Proceedings 1997 Workshop on Self-Organizing Maps, Helsinki, Finland. (1997)
5. Bishop, C.M., Svensén, M., Williams, C.K.I.: Magnification Factors for the GTM Algorithm. In: Proceedings IEE Fifth International Conference on Artificial Neural Networks. IEE, London (1997) 64–69
6. Bishop, C.M., Svensén, M., Williams, C.K.I.: GTM: The Generative Topographic Mapping. Neural Computation **1** (1998) 215–235
7. Kohonen, T.: The self–organizing map. Proceedings of the IEEE **9** (1990) 1464–1479
8. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge, England (1985)
9. Seber, G.A.F., Wild, C.J.: Nonlinear Regression. John Wiley and Sons, New York, NY (1989)
10. Tiňo, P., Nabney, I.: Constructing localized non-linear projection manifolds in a principled way: hierarchical Generative Topographic Mapping. Technical Report NCRG/2000/011, Neural Computation Research Group, Aston University, UK. (2000)
11. Willmann, T., Der, R., Martinez, T.: A new quantitative measure of topology preservation in Kohonen's feature maps. In: ICNN'94 Proceedings. IEEE Service Center (1994) 645–648



(a) Data projections



(b) Curvatures

Fig. 3. Oil data experiment.