Connectivity of Projected High Dimensional Data Charts on One Dimensional Curves

E. Bas[a,*], D. Erdogmus[a]

[a]*Northeastern University, 360 Huntington Ave., 409 Dana Research Center, Boston, MA, 02148*

Abstract

We propose a principal curve tracing algorithm that uses the gradient and the Hessian of a given density estimate. Curve definition requires the local smoothness of data density and is based on the concept of subspace local maxima. Tracing of the curve is handled through the leading eigenvector where fixed-step updates are used. We also propose an image segmentation algorithm based on the original idea and show the effectiveness of the proposed algorithm on a Brainbow dataset. Lastly, we showed a simple approach to define connectivity in complex topologies, by providing a tree representation for the bifurcating synthetic data.

Keywords: topological skeleton, principal curves, connectivity analysis

1. Introduction

Principal curves has been found to have several applications in the areas of signal processing as a dimensionality reduction technique. They are the smooth and self consistent curves that pass through the middle/ridge of the data cloud or probability distribution such that each point on the curve is the conditional mean of the points that orthogonally project onto that point of the curve [1]. Similar to the principal component analysis (PCA) that minimizes the least square construction error, principal curves also satisfy second order statistical optimality conditions by mapping the data to nonlinear manifolds. They provide a nonlinear summary of the data as low dimensional embedded manifolds and they have been studied as a nonlinear dimensionality reduction technique in the literate [2]. Various optimization approaches [1, 3–6] have been proposed to explicitly obtain principal curves.

*Corresponding author

Email addresses: bas@ece.neu.edu (E. Bas), erdogmus@ece.neu.edu (D. Erdogmus)

Principal curves can be used to define local adjacency in the data space and be used for clustering. Local conditions for points on the principal curves have been discussed in [7–9]. Motivated from [9], in our earlier work [10] we presented a method that uses locally defined principal curve definition to trace the data. Our method extracts the topology without solving any optimization problem and provides an implicit segmentation on topological manifolds of the data by linking the adjacent data samples based on local coherency. Given a seed point (or any arbitrary feature vector) and an initial approximate direction, the proposed method highlights the underlying structure in the data having locally similar characteristics with the seed. In order to achieve this, we first find the corresponding locally defined principal curve in the vicinity that is aligned to the initial direction and then trace it through the data space. Complex topologies having bifurcations can also be handled by intra connectivity of piecewise curves with automated initialization and will be discussed later.

2. Principal Curves

Let $\mathbf{x} \in \mathbb{R}^n$ be a random vector with samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ with twice continuously differentiable density/ridge function estimate $f(\mathbf{x})$. Let $\mathbf{g}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$ be the transpose of the local gradient and Hessian of $f(\mathbf{x})$ respectively and $\{(\lambda_1(\mathbf{x}), \mathbf{q}_1(\mathbf{x})), \ldots, (\lambda_n(\mathbf{x}), \mathbf{q}_n(\mathbf{x}))\}$ be the eigenvalue/vector pairs of $\mathbf{H}(\mathbf{x})$ such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. In general, a point, $\mathbf{x}$, is on the principal curve if and only if the local gradient points in the principal direction [7–9] and the eigenvalues corresponding to the eigenvectors that span the normal space that is orthogonal to the principal direction are all negative ($\lambda_2(\mathbf{x}) \ldots \lambda_n(\mathbf{x}) < 0$). Here, the principal direction of a function $\mathbf{x}$ is given as $\mathbf{q}_1(\mathbf{x})$. For instance, without loss of generality, let $S_\perp(\mathbf{x}) = \text{span}\{\mathbf{q_2}(\mathbf{x}), \mathbf{q_3}(\mathbf{x}), \ldots, \mathbf{q_n}(\mathbf{x})\}$ be the normal space spanned by the $n-1$ orthogonal eigenvectors and $S_\parallel(\mathbf{x}) = \text{span}\{\mathbf{q_1}(\mathbf{x})\}$ be the tangent vector at $\mathbf{x}$. Based on the mentioned local conditions, if $\mathbf{x}$ is on the principal curve, then $\mathbf{g}(\mathbf{x})$ is collinear with $S_\parallel(\mathbf{x})$, such that $s(\mathbf{x}) = \frac{\mathbf{g}(\mathbf{x})^T S_\perp(\mathbf{x})\mathbf{g}(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|\|\mathbf{H}(\mathbf{x})\mathbf{g}(\mathbf{x})\|} = 0$.

Geometrically rigorous way of defining connectivity between data samples is to seek geodesic between pairs. Geodesic can be defined as the path that extremize the length function, such that resultant pairwise trajectory between sample pairs is optimum,e.g. having shortest distance, minimum construction error. Consequently, an iterative strategy that trace the principal curve results in such geodesic. Given twice continuously differentiable function and its continuous eigen-

Algorithm 1: Principal Curve Tracing

At iteration t=0 initialize $\mathbf{x}_0$, the direction of the curve γ_0, the tangential and the normal step sizes $\mu_\parallel$ and $\mu_\perp$ respectively.

Correction:

1. Evaluate the gradient, the Hessian, and perform the eigendecomposition of $\mathbf{H}(\mathbf{x}) = \mathbf{V}\mathbf{\Gamma}\mathbf{V}$, where $\mathbf{V}_{1:n}$ are the eigenvectors with corresponding eigenvalues $\mathbf{\Gamma} = \mathbf{diag}\{\lambda_\mathbf{1} \leq \lambda_\mathbf{2} \leq \cdots \leq \lambda_\mathbf{n}\}$.
2. Let $\mathbf{V}_1$, and $\mathbf{V}_{2:n}$ be the eigenvectors that consequently span $S_\parallel(\mathbf{x})$, and $S_\perp(\mathbf{x})$, calculate the normal update such that $\tau(\mathbf{x}) = \mathbf{V_{2:n}}\mathbf{V_{2:n}^T}\mathbf{g}(\mathbf{x})$.
3. If $s(\mathbf{x}) = \frac{\mathbf{g}(\mathbf{x})^T S_\perp(\mathbf{x})\mathbf{g}(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|\|\mathbf{H}(\mathbf{x})\mathbf{g}(\mathbf{x})\|} \neq 0 \quad \mathbf{x}(t) \leftarrow \mathbf{x}(t) + \mu_\perp \tau(\mathbf{x})$ and go back to step 1; else

Tracing:

4. Evaluate the new curve direction using the tangential vector $\gamma_t = sign(\gamma_{t-1}^T \mathbf{q_1}(\mathbf{x}))\mathbf{q_1}(\mathbf{x})$
5. If $p(\mathbf{x}) < thr$ then stop, else $\mathbf{x}(t+1) \leftarrow \mathbf{x}(t) + \mu_\parallel \gamma_t$ and go back to step 1

vector/value pairs, updates constrained to the local tangent space $S_\parallel(\mathbf{x})$ trace the curve without explicitly defining it. Similarly, $S_\perp(\mathbf{x})$ defines the orthogonal projection plane and updates constrained to this plane will converge to the principal curves with proper direction. So an iterative tracing algorithm using correction-update scheme is possible by solving the ordinary differential equations, incorporating the iterations on the normal plane(correction step) and the tangential vector(tracing step) with proper directions. We used Euler integration [11] to project any point to its corresponding principal curve in the correction step and then use a fix point update to propagate on the curve in the tracing step. Alg. 1 outlines the principal curve tracing (PCT) procedure. First three steps corresponds to the projection step. Convergence of the projection iterations is guaranteed by correcting the sign of the direction to have positive inner product with the local gradient $\mathbf{g}(\mathbf{x})$. Once the necessary condition is satisfied ($s(\mathbf{x}) = 0$), a fix length step in the direction pointed by the $\mathbf{q_1}(\mathbf{x})$ traces the curve.

3. Estimation of the density/ridge function

In order to estimate the twice continuously differentiable density/ridge function, $f(\mathbf{x})$, of data samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, we employed well known nonparametric weighted variable-width kernel density estimate (wKDE) technique [12]. KDE is used as an example since it encompasses parametric mixture models as a special case; however as we will present in Sec. 5 method is general for any pdf model and can also be used with parametric models such as GMM [13]. Weighted KDE can be written as

$$p(\mathbf{x}) = \sum_{i=1}^{N} w(\mathbf{x}_i) G_{\mathbf{\Sigma}_i}(\mathbf{x} - \mathbf{x}_i) \tag{1}$$

where $w(\mathbf{x}_i)$ is the weight and Σ_i is the variable kernel covariance of the Gaussian kernel $\mathbf{G}(\mathbf{x_i}) = C_{\mathbf{\Sigma}_i} e^{-\frac{1}{2}\mathbf{x}_i^T {\Sigma_i}^{-1} \mathbf{x}_i}$ for the i^{th} data sample $\mathbf{x_i}$.

There are various techniques to estimate the optimal kernel size from the data. In our calculations, we used leave-one-out maximum likelihood procedure to estimate the global scaling factor for the local covariance [12]. The gradient and the Hessian of the KDE are:

$$\mathbf{g}(\mathbf{x}) = -\sum_{i=1}^{N} w(\mathbf{x}_i) G_{\Sigma_i}(\mathbf{x}-\mathbf{x}_i) {\Sigma_i}^{-1}(\mathbf{x}-\mathbf{x}_i) \tag{2}$$

$$\mathbf{H}(\mathbf{x}) = \sum_{i=1}^{N} w(\mathbf{x}_i) G_{\Sigma_i}(\mathbf{x}-\mathbf{x}_i)({\Sigma_i}^{-1}(\mathbf{x}-\mathbf{x}_i)(\mathbf{x}-\mathbf{x}_i)^T {\Sigma_i}^{-1} - {\Sigma_i}^{-1}) \tag{3}$$

Note that, selection of the step sizes effect the computation time and the location of the projected samples on the curve, but does not have any influence on the construction error. Since we project the sample, $\mathbf{x}$, to its underlying principal curve at every iteration t, reconstruction error with respect to ridge function is ideally zero. However, this does not guarantee the zero reconstruction error with respect to the original curve. This brings up the question "*How well the ridge of KDE approximates the original curve?*". In fact, this question can be interpreted by the robustness and accuracy analysis of the KDE which has been well studied in the literature [14]. In general density estimation utilizing variable width kernels with sufficient number of samples reported to perform well [12]. In order to estimate the KDE with a given accuracy, the number of the samples increases exponentially with the number of dimensions [15]- known as "The curse of dimensionality". Since our approach utilizes Hessian, we reported 100 Monte-Carlo simulation of the construction error (in terms of the mean square error (MSE) between the projected samples and the original curve) by varying the number of samples and the dimensions used in KDE in Fig.1. Here, N data points are sampled from a 1-dimensional manifold (semi-circle) in n-dimensional space. Data is perturbed with a spherical Gaussian noise ($\mathbf{0}$-mean $0.05\mathbf{I}$-std, $\mathbf{I} \in \mathbb{R}^{nxn}$: identity matrix) and depicted in Fig.1(a) for $n = 2$ and $N = 200$. Initial direction ,γ_0, and the location, $\mathbf{x}_0$ are represented with a black arrow and a green dot, respectively. In the experiments, we fixed the normal and tangential step sizes to $\mu_{\perp,\parallel} = 0.2$ and uniform $w_i = 1/N$ kernel weights are selected. Estimated KDE for the particular choice is visualized in Fig.1(b). The traced curve in red color is overlayed on the underlying curve in green and displayed in Fig.1(c). We selected 20 distinct N

instances uniformly between 100 and 1000 in logarithmic scale. We repeated the calculations for $n = 2, 4, 8, 16, 32, 64, 100$. Fig.1(d) shows the traced ridge accuracy in terms of MSE with respect N for selected n values. As a consequence of the degeneracy in the estimation, accuracy decreases with the increase in the dimension.

Number of iterations and the computation time depends on the the particular choice of the selected step sizes. Fig.1(e) shows the required time for $N = 200$ sample points with different $\mu_\perp$ and $\mu_\parallel$ selection. Similarly, Fig. 1(f-g) displays the total number of iterations during the projection and tracing steps respectively. In general number of iterations and the computation time is inversely proportional to the selected step sizes as expected. Deviation from the ridge increases as we increase the $\mu_\parallel$ and in consequence number of iterations to converge to the local principal curve increases. Although we fixed step sizes in our experiments, an adaptive strategy that captures the local deviation and adjust the step length would be ideal. Moreover, in the presence of multiple principal curves, this choice becomes more critical. For example, a large $\mu_\parallel$ at a point with high curvature might result in another nearby principal curve, and result in inaccurate tracing. High correlation between the process time and the overall total number of iterations in the projection step indicates the projection procedure as the current bottleneck in the computations. Moreover, total number of tracing updates is independent of the $\mu_\perp$. In fact, this implies that the algorithm decouples the correction procedure from tracing.

4. Principal Curves for Fiber Tracing in Volumetric Images

We tested the described technique principal curve tracing method on color images of neural tissue as seen in Fig.2. Our goal is to extract the topology of such curvilinear structures that will help neuroscientists to make quantitative analyze based on the morphology/centerline information ($d = 1$). Due to the high volume of data, in order to reduce the computations we replaced Gaussian kernels $G_{\Sigma_i}(\mathbf{x} - \mathbf{x}_i) \leftarrow \alpha B_{2,\epsilon}(\mathbf{p} - \mathbf{p}_i) G_{\mathbf{\Sigma}_{i,p}}(\mathbf{p} - \mathbf{p}_i)$ with bounded-support kernels constraining computations to neighboring pixels. Here, $\mathbf{p_i}$ is the position of the voxel in 3D ($n = 3$). $B_{2,\epsilon}(\mathbf{p} - \mathbf{p}_i)$ is the support ball with L_2 norm radius ϵ, and α is the normalization constant of the kernel. Since the traced structures have high intensity profile along the centerline of the curvilinear structure, an adaptive weighting scheme that emphasis the luminance and the spatial continuity in the appearance model is utilized. Weight of the i^{th} sample is selected as $w_i = I(\mathbf{p}_i) G_{\mathbf{\Sigma}_{i,c}}(\mathbf{c} - \mathbf{c}_i)$,

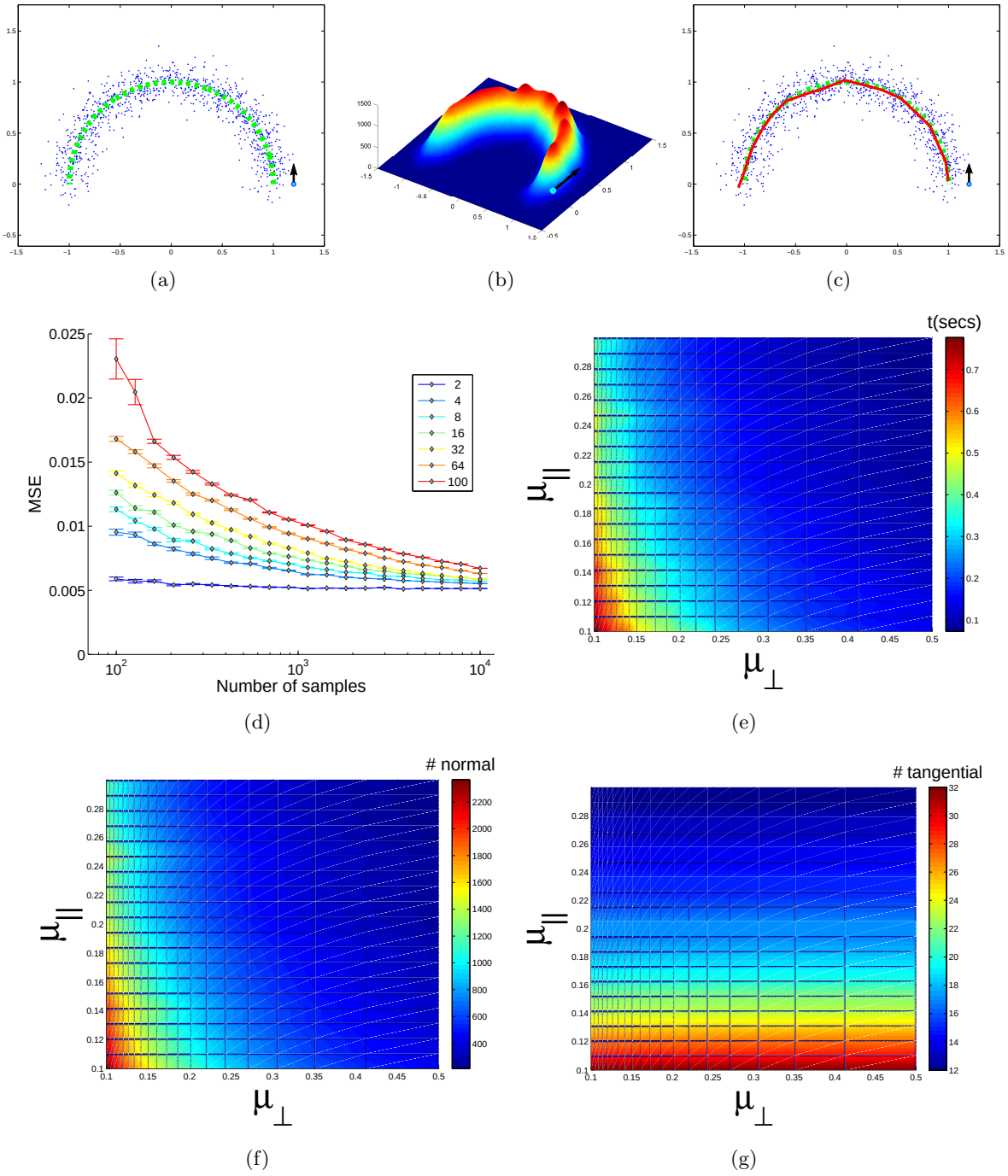
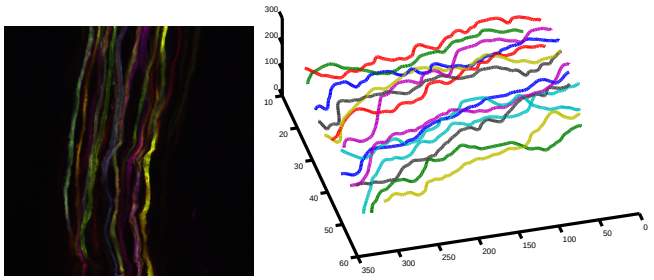


Figure 1: Top row: Circular data. Black arrow and green dot show the initial direction ,γ_0, and the location, $\mathbf{x}_0$, respectively. (a) Underlying curve. (b) Estimated density. (c) PCT result. Bottom row: Algorithm performance. (d) MSE vs number of samples for 100 Monte-Carlo simulation. (e) Elapsed time for varying $\mu_\perp$ and $\mu_{||}$. (f) Total number of iterations during the orthogonal projection steps for different $\mu_\perp$ and $\mu_{||}$ selection. (g) Similarly, total number of iterations in tracing steps.



(a) (b)

Figure 2: (a)Axial (xy-plane) slices of a sample brainbow image stack showing a bundle of axons of motor neurons connecting from the spinal cord to a muscle. (b)Tracing result.

where $I(\mathbf{p}_i)$ is the intensity of the voxel. In fact proposed weighting scheme has connections to the bilateral filter [16] and has shown to be having superior outlier robustness compared to simple intensity based appearance models. Moreover, in order to reduce the computations at an arbitrary point $\mathbf{p}$, support of the nearest grid location, $B_{2,\epsilon}(\mathbf{p}_i)$, is used to make the calculations such that $i = \arg_j \min \|\mathbf{p}_j - \mathbf{p}\|, \forall j = 1 \dots N$. In this way, we reemploy precalculated nearest neighbors and avoid recalculation of pairwise distance at every correction iteration.

Image dataset composed of 62 confocal microscopy image slices (with z-direction resolution of $32\mu m$) with each slice being 341×341 pixels (x & y directions at $33\mu m$ resolution) is used. Fig. 2 shows a sample slice. We used an ϵ-ball having a radius sufficient to cover $N_x = 250$ neighbors (approximately radius of 4 voxels). Kernel covariances Σ_i^P and Σ_i^C are selected as $\sigma_p^2\mathbf{I}$ and $\sigma_c^2\mathbf{I}$ where $\sigma_p = 5$ and $\sigma_c = 0.3$. At each step, direction of the principal curve is calculated (γ_t) as described in Alg. 1. The immediate neighbor voxel center that is closest to the direction pointed by γ_t is selected as the next approximate curve sample. Fig. 2(b) shows the trajectories of some selected axons.

5. Connectivity in Tree and Graph Structures

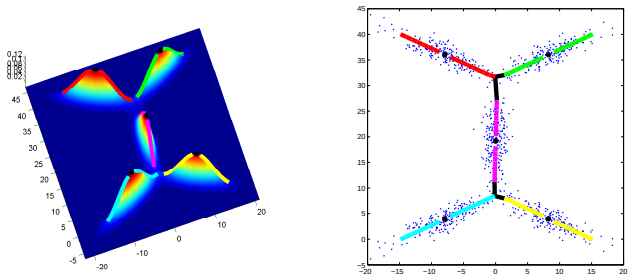
In most cases data samples do not lie on a smooth varying manifold and dataset includes bifurcations and self intersections resulting in complex topologies. However, it is easy to define topological manifolds in a rigorous way by constructing a tree/graph structure that represents the data distribution in the locally. This section describes a method to construct a complete connectivity map from learned piecewise curve manifolds. In order to show the generic applicability of the

method, this time we used a parametric model and generated samples from a 5-component Gaussian mixture. Density, gradient and Hessian are estimated for the Gaussian Mixture Model (GMM) and mixture parameters are obtained using the Expectation-Maximization algorithm [17]. Figure 3-a shows a 5-component Gaussian Mixtures (GM) that is spatially connected. Given that the principal curves pass through density modes, and assuming that the GMM modes are well approximated by the component means/modes we assumed that local density peaks overlap with the mixture modes $(\theta_i, i = 1, 2, \ldots, 5)$, where modes are represented with black dots in Fig. 3, two tracing algorithms (as described in Alg. 2) are initialized in opposite directions to each other at every mixture mode θ_i. Initial tracing directions $(S_{\|}(\theta_{\mathbf{i}}), -S_{\|}(\theta_{\mathbf{i}}))$ are estimated from the local covariance as in the KDE case and termination of each trace is controlled by a threshold. To make the process more presentable and prevent possible ill-posed locations, we used a relatively high threshold value resulting in more separated mixtures. Each trace is labeled with a different color in the figure.

Although we have compact representations for individual piecewise curves, in order to form a complete connectivity, one also needs to define intra connections between these curves. Theoretically, if the stopping threshold applied to the pdf is reduced sufficiently, the curves will eventually connect to each other by the virtue of principal curve continuity. One approximate method is to define a distance measure between curves and construct a pairwise distance matrix between curves. The distance between curves is defined as the shortest Euclidean distance between the pairwise curve samples. Minimum Spanning Tree (MST) algorithm is used in our experiments to analyze the connectivity of the obtained reduced distance graph. Fig. 3-b shows the final constructed tree on the GM pdf estimate. Black lines represent matched samples across piecewise curves. Although piecewise curves are lines since the underlying distribution is a GMM with some symmetry, the method is still valid for nonlinear structures as demonstrated in Sec. 2.

6. Discussion and Conclusion

In this paper, we presented a curve tracing algorithm that uses locally defined critical set definitions. Proposed method uses the gradient and the Hessian of the density estimate to calculate the principal curves as the underlying structures. While iterations on the constrained normal space pull samples towards the principal curve, fixed step size or fixed-length updates trace the curve



(a) (b)

Figure 3: Connectivity of piecewise curves.(a)Tracing of GMM.(b)Constructed tree overlayed with pdf.

along the *center* of the data (if local maximum coincides with local conditional mean). Termination of traces is related to density estimate at that location. Selection of the step size and kernel bandwidths are manual and proper selection should be determined by data geometry. Depending on the curvature, large step sizes might result in irrecoverable errors (one possible case in Brainbow analysis is two fibers with similar colors and high curvatures getting very close to each other - this might result in the traced curve to jump from one axon to the other).

We also describe a methodology to construct complete connectivity from traced piecewise curves for complex topologies. Based on our knowledge there is no rigorous global coordinate unwrapping technique that can be directly applicable to tree/graph data distribution structures, and the method described here can be classified as a topological skeletonisation process. Currently, it is still too early to mention complete charts for the whole data, but our method locally trace the data on one dimensional manifolds, then globally connect piecewise manifolds based on a distance measure.

In this work, we mainly focussed on building tree-structured one-dimensional data representation to describe the complex connectivity in the data topology. We assumed that samples from the traced curves in a very close proximity lie on the same tangential plane (in fact on the same line) around the intersection points and we simply used Euclidean distance to connect the piecewise curves while employing MST. Effectively, nonlinear manifolds are connected with linear measures around zero gradient locations as a practical approximation. Future work will entail extracting any-dimensional manifold graphs from data samples for dimension reduction.

References

[1] T. Hastie, W. Stuetzle, Principal curves, Journal of the American Statistical Association 84 (406) (1989) 502–516.

[2] A. Gorban, B. Kgl, D. Wunsch, A. Zinovyev, Principal manifolds for data visualization and dimension reduction, Springer Publishing Company, Incorporated, 2007.

[3] R. Tibshirani, Principal curves revisited, Statistics and Computing 2 (1992) 183–190.

[4] B. Kegl, A. Krzyzak, T. Linder, K. Zeger, Learning and design of principal curves, Pattern Analysis and Machine Intelligence, IEEE Transactions on 22 (3) (2000) 281–297.

[5] J. Einbeck, G. Tutz, L. Evers, Local principal curves, Statistics and Computing 15 (4) (2005) 301–313.

[6] I. Cleju, P. Franti, X. Wu, Clustering by principal curve with tree structure, in: International Symposium on Signals, Circuits and Systems, Vol. 2, 2005.

[7] A. Gray, E. Abbena, S. Salamon, Modern differential geometry of curves and surfaces with Mathematica, Chapman & Hall/CRC, 2006.

[8] D. Eberly, Ridges in image and data analysis, Kluwer Academic Pub, 1996.

[9] U. Ozertem, Locally Defined Principal Curves and Surfaces, Ph.D. thesis, Oregon Health & Science University, Department of Science & Engineering (2008).

[10] E. Bas, D. Erdogmus, Principal curve tracing, in: European Symposium on Artificial Neural Networks, 2010, pp. 405–410.

[11] U. Ascher, L. Petzold, Computer methods for ordinary differential equations and differential-algebraic equations, Society for Industrial Mathematics, 1998.

[12] B. Silverman, Density estimation for statistics and data analysis, Chapman & Hall/CRC, 1998.

[13] R. Duda, P. Hart, D. Stork, Pattern classification, John Willey & Sons, 2001.

[14] M. Wand, M. Jones, Comparison of smoothing parameterizations in bivariate kernel density estimation, Journal of the American Statistical Association 88 (422) (1993) 520–528.

[15] R. Bellman, R. Kalaba, On adaptive control processes, Automatic Control, IRE Transactions on 4 (2) (1959) 1–9.

[16] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Computer Vision, 1998. Sixth International Conference on, IEEE, 2002, pp. 839–846.

[17] A. Dempster, N. Laird, D. Rubin, et al., Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society. Series B (Methodological) 39 (1) (1977) 1–38.