ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Principal Curve Tracing

Erhan Bas, Deniz Erdogmus *

Cognitive Systems Laboratory, ECE Department, Northeastern University
Boston, MA 02115 - USA

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.

1 Introduction

Principal curves are the smooth curves that pass through the middle of the data cloud or probability distribution. This definition is first formalized by Hastie and Stuetzle [1], where each point on the curve is the conditional mean of the points that project there. These are called self consistent curves and their shape is inherited from the data. Following this self consistent curves definition, various principal curve algorithms have been defined in the literature, mostly based on Hastie's original definition [2, 3, 4, 5, 6]. In general, most of these algorithms use semi-global least square construction error to find principal curves, but there are also other locally defined principal curve definitions as in [5, 6]. Nevertheless, in all these past works, principal curves have been formulated as nonlinear curves (for linear case they are principal lines) that pass through the middle of the data or the probability distribution, and they provide a nonlinear summary of the data as compact representations on nonlinear manifolds. While in principal component analysis (PCA) this compact representation is given in terms of finding linear manifolds having least-square construction error, principal curves generally satisfy second order statistical optimality criteria by mapping the data to nonlinear manifolds.

Following the formal definition of locally defined principal curves and concept of critical and principal sets discussed in [6, 7], we employed a subspace constrained tracing method. 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, first we 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.

*This work has been partially supported by NSF under grant ECS0622239/ECCS0934506. The authors thank Ryan Draft for providing the Brainbow images used in this paper.

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

2 Principal Curves

Let $\mathbf{x} \in \mathbb{R}^n$ be a random vector with samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, having a given pdf estimate of $p(\mathbf{x})$. Let $\mathbf{g}(\mathbf{x})$ be the transpose of its local gradient, and let $\mathbf{H}(\mathbf{x})$ be the local Hessian of this pdf. Finally, using the gradient and Hessian, define the local covariance as: $\mathbf{\Sigma}^{-1}(\mathbf{x}) = -p^{-1}(\mathbf{x})\mathbf{H}(\mathbf{x}) + p^{-2}\mathbf{g}(\mathbf{x})^T\mathbf{g}(\mathbf{x})$[1] Let $\{(\lambda_1(\mathbf{x}), \mathbf{q}_1(\mathbf{x})), \ldots, (\lambda_n(\mathbf{x}), \mathbf{q}_n(\mathbf{x}))\}$ be the eigenvalue-eigenvector pairs of $\mathbf{\Sigma}^{-1}(\mathbf{x})$, sorted in ascending order: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. In general, a point, $\mathbf{x}$, is on the principal curve iff the local gradient is an eigenvector of the local covariance, where it is orthogonal to the other $n-1$ eigenvectors [6]. For instance, without loss of generality, let $S_\perp(\mathbf{x}) = span\{\mathbf{q_2}(\mathbf{x}), \mathbf{q_3}(\mathbf{x}), \ldots, \mathbf{q_n}(\mathbf{x})\}$ be the normal space spanned by the $n-1$ orthogonal eigenvectors and $S_\parallel(\mathbf{x}) = span\{\mathbf{q_1}(\mathbf{x})\}$ be the tangent vector at $\mathbf{x}$. If a point is on the principal curve, then $\mathbf{g}(\mathbf{x})$ is orthogonal to $S_\perp(\mathbf{x})$. Updates constrained to the $S_\perp(\mathbf{x})$ plane will converge to or diverge from the principal curves depending on the update direction, whereas propagating through tangential vector ($S_\parallel(\mathbf{x})$) will trace the locally defined principal curve at $\mathbf{x}$. So an iterative tracing algorithm using correction-update scheme is possible by incorporating the iterations on the normal plane(correction step) and the tangential vector(update step) with proper directions. Fig. 2 shows the tracing of a perturbed semicircular data given an initial direction (black arrow) and seed location (green circle). Consider a general weighted variable-width kernel density estimate (KDE)[2] obtained from samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ and initial tracing direction γ_0. KDE is given as

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

where $w(\mathbf{x}_i)$ is the weight and Σ_i is the variable kernel covariance[3] of the Gaussian kernel $G(\mathbf{x_i}) = C_{\Sigma_i} e^{-\frac{1}{2}x^T {\Sigma_i}^{-1} x}$ for the i^{th} data sample $\mathbf{x_i}$. The gradient and the Hessian of the KDE are:

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

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

For $p(\mathbf{x})$ mean-shift (MS) updates are in the form $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{ms}(\mathbf{x})$, where

$$\mathbf{ms}(\mathbf{x}) \quad = \quad (\sum_{i=1}^{N} G_{\Sigma_i}(\mathbf{x} - \mathbf{x}_i){\Sigma_i}^{-1})^{-1} \sum_{i=1}^{N} G_{\Sigma_i}(\mathbf{x} - \mathbf{x}_i){\Sigma_i}^{-1}\mathbf{x}_i \qquad (4)$$

[1]The local covariance is defined in this manner using the second order term in the Taylor series expansion of *log* $p(\mathbf{x})$ in order for principal curve projections to be consistent with PCA projections in the case of a Gaussian density.

[2]KDE is used as an example since it encompasses parametric mixture models as a special case; the method is general for any pdf model.

[3]Assuming Gaussian kernels here for simplicity.

and $\mathbf{ms}(\mathbf{x})$ can be decomposed as $\mathbf{ms}(\mathbf{x}) = \mathbf{ms}_{\|}(\mathbf{x}) + \mathbf{ms}_{\perp}(\mathbf{x})$. Here $\mathbf{ms}_{\perp}(\mathbf{x})$ is the normal component given as $\mathbf{V}_{\perp}\mathbf{V}_{\perp}^T\mathbf{ms}(\mathbf{x})$, where $\mathbf{V}_{\perp} = [\mathbf{q}_2(\mathbf{x}), \ldots, \mathbf{q}_n(\mathbf{x})]$. Similarly, $\mathbf{ms}_{\|}(\mathbf{x})$ is the tangential component given by $\mathbf{q}_1(\mathbf{x})\mathbf{q}_1(\mathbf{x})^T\mathbf{ms}(\mathbf{x})$. Constrained MS iterations on $S_{\perp}(\mathbf{x})$ force $\mathbf{x}$ to converge to the principal curve through fix-point iterations whose convergence is guaranteed. At each iteration, sign of the $\mathbf{ms}_{\|}(\mathbf{x})$ must be corrected with the current tracing direction defined by the previous iteration (γ_{t-1}) and normalized to the step length: $\gamma(\mathbf{x})_t = sign(\gamma(\mathbf{x})_{t-1}^T\mathbf{ms}_{\|}(\mathbf{x}))\mathbf{ms}_{\|}(\mathbf{x})$. Summary of the algorithm is presented in Table 1.

3 Principal Curves for Fiber Tracing in Volumetric Images

Curvilinear objects are very common in images, especially in biomedical imaging (i.e. vessels, neurons). Color values ($\mathbf{c}$) and pixel locations($\mathbf{p}$) form the feature space. In order to reduce computational load, KDE computations are restricted to the K-Nearest Neighbors (KNN) in space denoted by N_p^{KNN}. Weights in KDE can be selected as the pixel intensity values (e.g., sum of RGB color coordinates). Separable bounded-support kernels replace Gaussian kernels $G_{\Sigma_i}(\mathbf{x} - \mathbf{x}_i) \leftarrow \alpha B_{2,\epsilon}(\mathbf{p} - \mathbf{p}_i)G_{\Sigma_i^P}(\mathbf{p} - \mathbf{p}_i)G_{\Sigma_i^C}(\mathbf{c} - \mathbf{c}_i)$ constraining computations to neighboring pixels. Here, $\mathbf{p_i}$ and $\mathbf{c_i}$ are the position and color vector components of $\mathbf{x_i}$. $B_{2,\epsilon}(\mathbf{p} - \mathbf{p}_i)$ is the support ball with L_2 norm radius ϵ, and α is the normalization constant of the kernel. KDE becomes

$$\begin{aligned} p(\mathbf{x}) &= \sum_{i=1}^{N} w_i \alpha B_{2,\epsilon}(\mathbf{p} - \mathbf{p}_i)G_{\Sigma_i^P}(\mathbf{p} - \mathbf{p}_i)G_{\Sigma_i^C}(\mathbf{c} - \mathbf{c}_i) \\ &= \sum_{x_i \in B_{2,\epsilon}(\mathbf{p})} w_i \alpha G_{\Sigma_i^P}(\mathbf{p} - \mathbf{p}_i)G_{\Sigma_i^C}(\mathbf{c} - \mathbf{c}_i) \end{aligned} \quad (5)$$

Letting $\beta_i(\mathbf{x}) = \alpha w_i G_{\Sigma_i^P}(\mathbf{p} - \mathbf{p}_i)G_{\Sigma_i^C}(\mathbf{c} - \mathbf{c}_i)$, the gradient and Hessian are

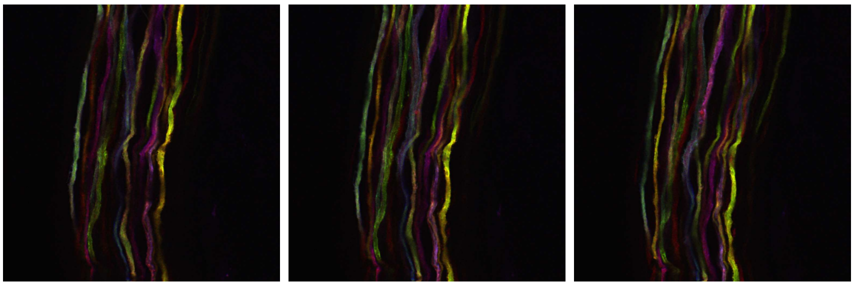
$$\mathbf{g}(\mathbf{x}) = -\sum_{x_i \in B_{2,\epsilon}(\mathbf{p})} \beta_i(\mathbf{x}) \begin{bmatrix} \Sigma_{ip}^{-1} & 0 \\ 0 & \Sigma_{ic}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{p_i} \\ \mathbf{c_i} \end{bmatrix} \quad (6)$$

$$\mathbf{H}(\mathbf{x}) = \sum_{x_i \in B_{2,\epsilon}(\mathbf{p})} \beta_i(\mathbf{x}) \begin{bmatrix} H_{pp} & H_{pc} \\ H_{pc}^T & H_{cc} \end{bmatrix} \quad (7)$$

where $H_{pp} = \Sigma_{ip}^{-1}(\mathbf{p} - \mathbf{p}_i)(\mathbf{p} - \mathbf{p}_i)^T\Sigma_{ip}^{-1} - \Sigma_{ip}^{-1}$, $H_{pc} = H_{cp}^T = \Sigma_{ip}^{-1}(\mathbf{p} - \mathbf{p}_i)(\mathbf{c} - \mathbf{c}_i)^T\Sigma_{ic}^{-1}$, and $H_{cc} = \Sigma_{ic}^{-1}(\mathbf{c} - \mathbf{c}_i)(\mathbf{c} - \mathbf{c}_i)^T\Sigma_{ic}^{-1} - \Sigma_{ic}^{-1}$. Furthermore, in order to obtain continuous tracing over voxels, tracing iterations can be restricted to immediate spatially neighboring voxels.

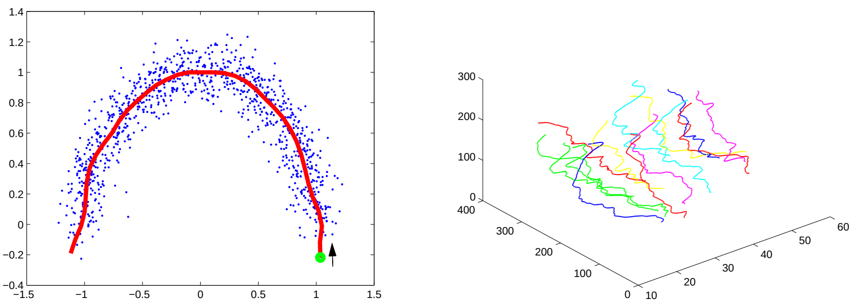
4 Experimental Results

In this section we present results obtained on synthetic and real datasets. For simplicity, isotropic fixed bandwidth kernels are employed. Since principal curves



(a) $7^{th}Slice$ (b) $8^{th}Slice$ (c) $9^{th}Slice$

Fig. 1: 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.



(a) Circular data (b) Brainbow images

Fig. 2: Tracing of synthetic and real datasets.

pass through cluster modes, MS algorithm [8] is run starting from the provided arbitrary seed and tracing starts from the corresponding mode. Synthetic data is composed of data uniformly distributed along a semicircle ($radius = 1$) and radially perturbed by a Gaussian (0-mean 0.1-std). Sample weights in KDE are equal and the stopping threshold thr is 0.01 with unit step size ($\mu = 1$). Kernel covariance is $\sigma^2\mathbf{I}$ with $\sigma = 0.1$. Fig. 2-a shows the tracing result for this semi-circular dataset. Second, a Brainbow image set composed of 31 confocal microscopy image slices (with z-direction resolution of $64\mu m$) with each slice being 1024×1024 pixels (x & y directions at $11\mu m$ resolution) is used. In the preprocessing step the images are downsampled by 3 in the x & y directions and upsampled by 2 in the z direction yielding a voxel size of $33 \times 33 \times 32\mu m^3$. The resampled image stack is smoothed in 3D using a bilateral filter with Gaussian kernel with diagonal covariance (spatial scale of 5 voxels and RGB-color scale of 0.2). Fig. 1 shows sample slices after smoothing with axons visible in each 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

Table 1: Summary of Principal Curve Tracing Algorithm

1. At iteration t=0 initialize $\mathbf{x}$, the step size μ, and the direction of the curve γ_0.
2. At iteration t evaluate the mean shift update $\mathbf{ms(x(}t\mathbf{))}$ as in Eqn. 4.
3. Evaluate the gradient, the Hessian, and perform the eigendecomposition of $\mathbf{\Sigma^{-1}}(\mathbf{x}) = \mathbf{V\Gamma V}$, where $\mathbf{V}_{1:n}$ are the eigenvectors with corresponding eigenvalues $\mathbf{\Gamma} = \mathbf{diag}\{\lambda_\mathbf{1} \leq \lambda_\mathbf{2} \leq \cdots \leq \lambda_\mathbf{n}\}$.
4. Let $\mathbf{V}_1$, and $\mathbf{V}_{2:n}$ be the eigenvectors that consequently span $S_{\|}(\mathbf{x})$, and $S_{\perp}(\mathbf{x})$, such that $\mathbf{ms}_{\|}(\mathbf{x}) = \mathbf{q_1}(\mathbf{x})\mathbf{q_1}(\mathbf{x})^T\mathbf{ms}(\mathbf{x})$, and $\mathbf{ms}_{\perp}(\mathbf{x}) = \mathbf{V}_{2:n}\mathbf{V}_{2:n}^T\mathbf{ms}_(\mathbf{x})$.
5. Evaluate the new curve direction vector $\gamma_t = sign(\gamma_{t-1}^T\mathbf{ms}_{\|}(\mathbf{x}))\mathbf{ms}_{\|}(\mathbf{x})$
6. If $p(\mathbf{x}) < thr$ then stop, else $\mathbf{x}(t+1) \leftarrow \mathbf{x}(t) + \mathbf{ms}_{\perp}(\mathbf{x}) + \mu\frac{\gamma_t}{\|\gamma_t\|}$

Table 2: Summary of Principal Curve Tracing Algorithm for Image Datasets

1. At iteration t=0 initialize $\mathbf{x}$, and the direction of the curve γ_0.
2. At iteration t evaluate the wKDE mean shift update $\mathbf{ms(x(}t\mathbf{))}$ as in Eqn. 4 using the samples that are in the vicinity.
3. Evaluate the gradient, the Hessian, and perform the eigendecomposition of $\mathbf{\Sigma^{-1}}(\mathbf{x}) = \mathbf{V\Gamma V}$, where $\mathbf{V}_{1:n}$ are the eigenvectors with corresponding eigenvalues $\mathbf{\Gamma} = \mathbf{diag}\{\lambda_\mathbf{1} \leq \lambda_\mathbf{2} \leq \cdots \leq \lambda_\mathbf{n}\}$.
4. Let $\mathbf{V}_1$, and $\mathbf{V}_{2:n}$ be the eigenvectors that consequently span $S_{\|}(\mathbf{x})$, and $S_{\perp}(\mathbf{x})$, such that $\mathbf{ms}_{\|}(\mathbf{x}) = \mathbf{q_1}(\mathbf{x})\mathbf{q_1}(\mathbf{x})^T\mathbf{ms}(\mathbf{x})$, and $\mathbf{ms}_{\perp}(\mathbf{x}) = \mathbf{V}_{2:n}\mathbf{V}_{2:n}^T\mathbf{ms}_(\mathbf{x})$.
5. Evaluate the new curve direction vector $\gamma_t = sign(\gamma_{t-1}^T\mathbf{ms}_{\|}(\mathbf{x}))\mathbf{ms}_{\|}(\mathbf{x})$
6. If $\mathbf{x}$ is outside the image boundary stop, else $\mathbf{x}(t+1) = \arg\min_{x_i \in \mathbf{T_x}}(\gamma_t^T(\mathbf{x_i} - \mathbf{x}))$. Here $\mathbf{T_x}$ is the connected neighborhood of the $\mathbf{x}$ composed of 26 voxels in 3D.

as $\sigma_p^2\mathbf{I}$ and $\sigma_c^2\mathbf{I}$ where $\sigma_p = 5$ and $\sigma_c = 0.3$. Since fiber colors vary along the axis, we employed moving averaging smoothing for the colors to adapt to this spatial color change along each axon. At each step, direction of the principal curve is calculated (γ_t) as described in Tab. 2. The immediate neighbor voxel center that is closest to the direction pointed by γ_t is selected as the next approximate curve sample. In displaying the axon trajectories another moving averaging filter is used to smooth these. Fig. 2-b shows the trajectories of some selected axons.

5 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 towards to the principal curve, fixed step size or fixed-length updates trace the curve along the *center* of the data (if local maximum coincides with mean). 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). Kernel bandwidth selection is well researched for density estimation (e.g. leave-one-out cross validation maximum likelihood), but best density estimation bandwidth might not be the best principal curve estimation bandwidth and procedures specific to the latter should be researched in the future. Local principal set definition solves issues related to bifurcations, loops, and self intersections naturally. In tracing branching curves, one possibility is to initialize another tracing algorithm when a bifurcation is detected (indicated by a circular local Hessian on the curve). Further work is needed to solve practical issues in this area.

References

[1] Hastie Trevor and Stuetzle Werner, Principal Curves, *Journal of the American Statistical Association*,406:502–516,1989.

[2] Robert Tibshirani, Principal Curves Revisited, *Statistics and Computing*, 2:183-190, 1992.

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

[4] Sandilya, S. and Kulkarni, S.R., Principal curves with bounded turn, *IEEE Transactions on Information Theory*,48:2789-2793, 2002.

[5] Einbeck, Jochen and Tutz, Gerhard and Evers, Ludger, Local principal curves, *Statistics and Computing*, 15:301-313, Elsevier, 2005.

[6] Erdogmus, D. and Ozertem, U., Self-Consistent Locally Defined Principal Surfaces, *IEEE International Conference on Acoustics, Speech and Signal Processing*(ICASSP 2007),pages II-549-II-552, April, 2007.

[7] Ozertem, U. and Erdogmus, D., Local Conditions for Critical and Principal Manifolds, *IEEE International Conference on Acoustics, Speech and Signal Processing*, (ICASSP 2008), pages 1893-1896, 2008.

[8] Yizong Cheng, Mean shift, mode seeking, and clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*,volume 17, pages 790-799