LOCAL CONDITIONS FOR CRITICAL AND PRINCIPAL MANIFOLDS

Umut Ozertem, Deniz Erdogmus

CSEE Department, Oregon Health and Science University, Portland, Oregon 97239-3098

ABSTRACT

Principal manifolds are essential underlying structures that manifest canonical solutions for significant problems such as data denoising and dimensionality reduction. The traditional definition of self-consistent manifolds rely on a least-squares construction error approach that utilizes semi-global expectations across hyperplanes orthogonal to the solution. This definition creates various practical difficulties for algorithmic solutions to identify such manifolds, besides the theoretical shortcoming that self-intersecting or nonsmooth manifolds are not acceptable in this framework. We present local conditions for critical and principal manifolds by introducing the concept of subspace local maxima. The conditions generalize the two conditions that characterize stationary points of a function to stationary surfaces. The proposed framework yields a unique set of principal points which can be partitioned into principal curves and manifolds of any intrinsic dimensionality. A subspace-constrained fixed-point algorithm is proposed to determine the principal graph.

Index Terms— Principal curves, unsupervised learning, manifold learning, dimensionality reduction, feature extraction, denoising

1. INTRODUCTION

For a jointly Gaussian random vector, principal components analysis (PCA) yields the optimal projection hyperplanes which satisfy least-squares reconstruction error and maximum likelihood projection properties simultaneously. Traditional extensions of the concept to identifying principal manifolds of one or more intrinsic dimensions have typically exploited the least-squares reconstruction error aspect in conjunction with various smoothness and self consistency constraints for the nonlinear solutions. This yields satisfactory results if the high dimensional data can be modelled as a Gaussian or other unimodal symmetric perturbation orthogonal to a smooth regular manifold. These solutions can be determined locally or globally in terms of parametric or nonparametric models [1, 2, 3].

The self-consistent principal curves and surfaces as defined by Hastie [4] and studied by various researchers [5, 6, 7, 8] have been the foundation of most existing techniques due to the traditional appeal of 2^{nd}order statistical optimality criteria. Algorithm convergence proofs are difficult to construct due to the nature of the definition which leads to intersecting orthogonal subspaces of the curve for different points and the smoothness requirement might impose artificial constraints on the solution for certain data distributions as we will illustrate later. Algorithms to find principal curves inspired by Hastie's definition include Tibshirani's mixture-model expectation maximization approach [5], Sandilya and Kulkarni's bounded curvature approach [6], Kegl and colleagues' piecewise-linear approach [7], and Stanford and Raftery's outlier robust algorithm [8].

There are two contributions of this paper: (*i*) following our earlier work [9], we present the concept of critical surfaces of intrinsic dimension $d \leq n$, where n is the dimensionality of the space in which the data is embedded; (*ii*) a subspace-constrained mean-shift algorithm to determine the principal curves for a KDE model. The principal surfaces with desirable best approximation properties reside as subsets in these sets. The definition yields a unique principal set for a given density model; therefore, constraints on smoothness can be and must be imposed on the density model rather than the manifolds. This has advantages and disadvantages; since the regularization problem is transferred from curve/manifold generation to density modelling, extensive existing know-how on density model fitting can be exploited, however at the full dimensionality of the data rather than the intrinsic dimensionality of the desired manifold.

This work is supported by NSF grants ECS-0524835, and ECS-0622239.

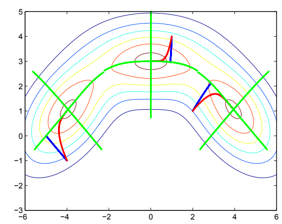

Fig. 1. A 2-dimensional pdf of a Gaussian mixture with 3 components is presented. The trajectories of gradient-ascent (red), local covariance-eigenvector that yields the proposed principal curve projections (blue), and the principal tree (green) are shown.
.

We start with an illustration to motivate the reader. The definitions provided below lead to a generalization of the usual two conditions for stationary points into two conditions that a point $\mathbf{x} \in \mathbb{R}^n$ needs to satisfy to be an element of the d-dimensional principal surface of a data distribution $p(\mathbf{x})$. For instance, the local conditions that state *a point is a local maximum iff the gradient is equal to zero and all of the eigenvalues of the Hessian are negative* generalizes to *a point is on the d-dimensional principal surface iff the gradient is in the span of d eigenvectors of the Hessian and all remaining (n-d) eigenvectors have negative eigenvalues*. The general conditions contain criteria for local maxima as special case for 0-dimensional principal set and one can prove a deflation property for subsequent dimensionalities of critical sets (includes minor, saddle, and principal surfaces). Furthermore, we will also observe that the conditions for a point being an element of the principal curve reduce to: (*i*) the gradient is an eigenvector of the Hessian, (*ii*) the eigenvalues of the remaining eigenvectors are negative. The subspace-constrained density maximization algorithm, inspired by mean-shift, achieves convergence to a fixed point that satisfies the conditions for being in a local principal surface that is consistent with the definition by performing the mean-shift update after projecting it to the span of the local eigenvectors of the Hessian at the current point on the iteration trajectory. While regular mean-shift trajectories converge to the local maxima of a pdf as an EM update, the subspace-constrained mean-shift updates converge to a point in the principal surface of desired dimensionality (e.g., principal curve along a ridge of the pdf). A comparison of such trajectories following local Hessian eigenvectors to obtain the principal curve and local gradient directions to obtain the local maxima are presented in Figure 1 on a Gaussian mixture.

2. CRITICAL AND PRINCIPAL SETS

The proposed development of the local theory of principal curves and manifolds is inspired by the geometry of curved surfaces embedded in higher dimensional Euclidean spaces. Similarities with differential geometric concepts regarding principal curves of m-dimensional Riemannian manifolds embedded in n-dimensional Euclidean spaces exist. In the context of probabilistic data models, we interpret the pdf of an n-dimensional random vector as an n-dimensional manifold and generalize the concepts of critical and principal lines.

We assume that a pdf $p(\mathbf{x})$ for a random vector $\mathbf{x} \in \Re^n$ is provided (either known or estimated from data). Furthermore, we assume that $p(\mathbf{x}) > 0$ for all $\mathbf{x}$, is continuous, and at least twice differentiable. Let $\mathbf{g}(\mathbf{x})$ be the transpose of the local gradient of this pdf, and $\mathbf{H}(\mathbf{x})$ be the local Hessian of this pdf evaluated at $\mathbf{x}$. Also let $\{(\lambda_1(\mathbf{x}), q_1(\mathbf{x})), \ldots, (\lambda_n(\mathbf{x}), q_n(\mathbf{x}))\}$ be the eigenvalue-eigenvector pairs of $\mathbf{H}(\mathbf{x})$, sorted such that $\lambda_1(\mathbf{x}) \geq \lambda_2(\mathbf{x}) \geq \ldots \geq \lambda_n(\mathbf{x})$.

Definition 2.1. A point $\mathbf{x}$ is an element of the d-dimensional critical set, denoted by C^d iff $\mathbf{g}(\mathbf{x})$ is orthogonal to at least (n-d) eigenvectors of $\mathbf{H}(\mathbf{x})$, where orthogonality of two vectors means null Euclidean inner product.

Lemma 2.1. Critical points of $p(\mathbf{x})$ constitute C^0.

This lemma illustrates that for $d = 0$ the proposed definition of critical sets reduces to the traditional definition of critical points.

Lemma 2.2. $C^d \subset C^{d+1}$.

A natural consequence of this lemma by induction is: $C^0 \subset \ldots \subset C^n$, where $C^n = \Re^n$. This property of critical sets makes it possible, in theory, to utilize deflation or inflation procedures to construct these substructures. Note that, by definition, C^d is a union of $\binom{n}{d}$ submanifolds, which intersect each other at various submanifolds of critical sets with lower dimensionalities.

Definition 2.2. A point $\mathbf{x} \in C^d$ but $\mathbf{x} \notin C^{d-1}$ is called a regular point of C^d. A point $\mathbf{x} \in C^{d-1}$ is called an irregular point C^d.

Lemma 2.3. If $\mathbf{x}$ is a regular point C^d, then there exists an index set $I \subset \{1, \ldots, n\}$ with cardinality $|I| = (n-d)$ such that $\langle \mathbf{g}(\mathbf{x}), \mathbf{q}_i(\mathbf{x}) \rangle = 0$ iff $i \in I$. If $\mathbf{x}$ is an irregular point of C^d, then $|I| > (n-d)$.

Lemma 2.4. Let $\mathbf{x}$ be a regular point of C^d and I be an index set with cardinality $|I| = (n-d)$ and such that $\langle \mathbf{g}(\mathbf{x}), \mathbf{q}_i(\mathbf{x}) \rangle = 0$ iff $i \in I$. The tangent subspace of C^d at $\mathbf{x}$ is $C^d_{\parallel}(\mathbf{x}) = span\{\mathbf{q}_i(\mathbf{x}) | i \notin I\}$ and the normal subspace of C^d at $\mathbf{x}$ is $C^d_{\perp}(\mathbf{x}) = \Re^n - C^d_{\parallel}(\mathbf{x}) = span\{\mathbf{q}_i(\mathbf{x}) | i \in I\}$.

So far, we have defined the critical sets as unions of submanifolds of the pdf manifold and we have characterized the tangent and normal Euclidean subspaces to these submanifolds at every point. However, we have not characterized the critical manifolds as locally maximum, minimum, or saddle. This characterization has to utilize the sign of the eigenvalues of the local Hessian and will lead to the definition of locally maximal principal sets as the canonical solution for dimensionality reduction in a maximum likelihood manner.

Theorem 2.1.(Subspace Stationarity) Let $\mathbf{x}$ be a regular point of C^d and I be an index set with cardinality $|I| = (n-d)$ such that $\langle \mathbf{g}(\mathbf{x}), \mathbf{q}_i(\mathbf{x}) \rangle = 0$ iff $i \in I$. The following statements hold:

1. local maximum in $C^d_{\perp}(\mathbf{x})$ iff $\lambda_i(\mathbf{x}) < 0 \; \forall i \in I$.
2. local minimum in $C^d_{\perp}(\mathbf{x})$ iff $\lambda_i(\mathbf{x}) > 0 \; \forall i \in I$.
3. saddle point in $C^d_{\perp}(\mathbf{x})$ iff $\exists \lambda_i(\mathbf{x}) < 0$ and $\exists \lambda_i(\mathbf{x}) > 0$; $i \in I$.

Relaxing the inequalities to include zero-eigenvalues would result in platitude points as usual. However, for the purposes of the following definitions, inequalities are kept strict.

Definition 2.3. A point $\mathbf{x}$ is an element of the: (1) principal set P^d iff $\mathbf{x}$ is a regular local maximum point in $C^d_{\perp}(\mathbf{x})$; (2) minor set M^d iff $\mathbf{x}$ is a regular local minimum point in $C^d_{\perp}(\mathbf{x})$; (3) Saddle set S^d iff $\mathbf{x}$ is a regular saddle point in $C^d_{\perp}(\mathbf{x})$.

Lemma 2.5. (P^d, M^d, S^d) is a partition of C^d.

Lemma 2.6. (1) $\mathbf{x} \in P^0$ iff $\mathbf{x}$ is a local maximum of $p(\mathbf{x})$. (2) $\mathbf{x} \in M^0$ iff $\mathbf{x}$ is a local minimum of $p(\mathbf{x})$. (3) $\mathbf{x} \in S^0$ iff $\mathbf{x}$ is a saddle point of $p(\mathbf{x})$.

This lemma states that the modes of a pdf (now called principal points), form the 0-dimensional principal set. Note that the modes of a pdf provide a natural clustering solution for data. In fact, the widely used mean-shift algorithm [11] utilizes this property to arrive at a nonparametric clustering solution for a given data set in a manner similar to the Morse-Smale decomposition of the vector space.

Lemma 2.7. $P^d \subset P^{d+1}$. $M^d \subset M^{d+1}$.

The importance of this property is that the principal and minor sets can be, in principal, constructed by a procedure similar to deflation. One can determine the peaks P^0 and the pits M^0 of a pdf $p(\mathbf{x})$ and then trace out P^1 and M^1 by following the eigenvectors of the local Hessian via a suitable differential equation with P^0 and M^0 initial conditions. The same could be done for each element of P^1 and M^1 as initial conditions to suitable differential equations to determine P^2 and M^2, etc. The procedure outlined above, in general, requires numerical integration of nonlinear differential equations to identify the lines of curvature, which provide a natural local coordinate frame on the manifold. However, this process might be unsuitable for learning applications where computational complexity is a constraint. We have experimented with this approach, employing Runge-Kutta order-4 as the integration technique, to determine P^1. The error accumulation, especially at high curvature points of P^1, prevent accurate determination of P^1 by brute force numerical integration using a fixed integration step size and this method is not preferred for practical purposes. Still, these lines of curvature provide a natural optimal nonlinear projection scheme.

Definition 2.4. A point $\mathbf{x} \in (P^d, M^d)$ but $\mathbf{x} \notin (P^{d-1}, M^{d-1})$ is called a regular point of (P^d, M^d). A point $\mathbf{x} \in (P^d, M^d)$ and $\mathbf{x} \in (P^{d-1}, M^{d-1})$ is called an irregular point of (P^d, M^d).

Lemma 2.8. Let $\mathbf{x}$ be regular point of (P^d, M^d) and I be an index set with cardinality $|I| = (n-d)$ such that $\langle \mathbf{g}(\mathbf{x}), \mathbf{q}_i(\mathbf{x}) \rangle = 0$ iff $i \in I$. The tangent subspace to (P^d, M^d) at $\mathbf{x}$ is $(P^d_{\parallel}(\mathbf{x}), M^d_{\parallel}(\mathbf{x})) = span\{\mathbf{q}_i | i \notin I\}$ and the normal subspace of (P^d, M^d) at x is $(P^d_{\perp}(\mathbf{x}), M^d_{\perp}(\mathbf{x})) = span\{\mathbf{q}_i | i \in I\}$.

So far we have achieved the following: (1) a self-consistent local definition of critical, principal, minor, and saddle sets of a pdf is presented and the relationships between them are established, (2) the concept of critical nets is generalized to encompass manifolds with dimensionality higher than one, (3) a unifying framework between maximum likelihood clustering, curve and surface fitting, and manifold learning using deflation. Theorem 2.1 establishes generalized conditions for a point being in a critical (stationary) submanifold utilizing local gradient and Hessian spectral information, of which the usual stationary point conditions remain as special cases. The definitions demonstrate that in general a globally smooth and maximally likely dimensionality reduction manifold that *passes through* the data is not feasible. As will be illustrated with examples below, the principal sets for optimal dimensionality reduction might form irregular, self-intersecting manifolds in the global scheme, although their local existence and uniqueness is guaranteed by the theorem.

Definition 2.5. Let $\mathbf{x}$ be a point in P^d (M^d). Let $\langle \mathbf{g}(\mathbf{x}), \mathbf{q}_i(\mathbf{x})\rangle = 0$ for $i \in I = \{1, \ldots, n\} - I^c$, where I is some index set with cardinality d. Define the local covariance matrix of $p(\mathbf{x})$ to be $\mathbf{\Sigma^{-1}}(\mathbf{x}) = -\mathbf{p^{-1}}(\mathbf{x})\mathbf{H}(\mathbf{x}) + \mathbf{p^{-2}}\mathbf{g}(\mathbf{x})^{\mathbf{T}}\mathbf{g}(\mathbf{x})]$ and assume that its eigendecomposition is $\{\gamma_i(\mathbf{x}), \mathbf{v}_i(\mathbf{x})\}$ for $i \in \{1, \ldots, n\}$. Assume that the eigenvalues with index $i \in I^C$ of the local covariance matrix satisfy the following: $\gamma_1 > \ldots > \gamma_m > 0 > \gamma_{(m+1)} > \ldots > \gamma_{n-d}$. Then, the local ranking of principal directions at $\mathbf{x}$ from principal to minor follow the same ranking of indices.

For the special case of Gaussian distributions, the local covariance defined above becomes constant over the data space and equal to the data covariance. Thus, the local principal directions are aligned with the global principal directions and following these directions starting from any point, takes one to the corresponding subsurface of C^{d-1}. Proofs of the theorems are omitted due to restricted space and will be included in a future journal publication[1].

3. SUBSPACE-CONSTRAINED MEAN-SHIFT TO FIND PRINCIPAL CURVES

A natural consequence of Theorem 2.1 is that a point is on a critical curve iff the local gradient is an eigenvector of the local Hessian, since the gradient has to be orthogonal to the other $n-1$ eigenvectors. Furthermore, for this point to be in the principal curve, the corresponding $n-1$ eigenvalues must be negative. Under the assumption of a KDE, a modification of the mean-shift algorithm by constraining the fixed-point iterations to the directions of local eigenvectors at the current point in the trajectory leads to an update that converges to the principal curves and not to the local maxima. The algorithm could be modified to converge to the d-dimensional principal manifold P^d in a conceptually trivial manner; however, computational requirements would increase combinatorially with d. Consider $\{\mathbf{x}_i\}_{i=1}^N$ where $\mathbf{x}_i \in \Re^n$. The KDE of this data set (using Gaussian kernels for illustration) is given as

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

where Σ_i is the kernel covariance for $\mathbf{x}_i$; $G_{\Sigma_i}(\mathbf{y}) = C_{\Sigma_i} e^{-\mathbf{y}^T \Sigma_i^{-1}\mathbf{y}/2}$. For simplicity, isotropic fixed bandwidth kernels may also be employed. The gradient and the Hessian of the KDE are

$$\begin{array}{c} \mathbf{g}(\mathbf{x}) = -N^{-1}\sum_{i=1}^{N} c_i \mathbf{u}_i \\ \mathbf{H}(\mathbf{x}) = N^{-1}\sum_{i=1}^{N} c_i(\mathbf{u}_i\mathbf{u}_i^T - \Sigma_i^{-1}) \\ \text{where} \quad \mathbf{u}_i = \Sigma_i^{-1}(\mathbf{x} - \mathbf{x}_i) \quad \text{and} \quad c_i = G_{\Sigma_i}(\mathbf{x} - \mathbf{x}_i) \end{array} \quad (2)$$

Let $\{(\lambda_1(\mathbf{x}), q_1(\mathbf{x})), \ldots, (\lambda_n(\mathbf{x}), q_n(\mathbf{x}))\}$ be the eigenvalue-eigenvector pairs of $\mathbf{\Sigma^{-1}}(\mathbf{x})$ as defined in Definition 2.5 and the mean-shift update emerging from (2) be

$$\mathbf{x} \leftarrow \mathbf{m}(\mathbf{x}) = (\textstyle\sum_{i=1}^{N} c_i \Sigma_i^{-1})^{-1} \sum_{i=1}^{N} c_i \Sigma_i^{-1}\mathbf{x}_i \quad (3)$$

At $\mathbf{x}$, the subspace mean-shift update is performed in two steps: $\tilde{\mathbf{x}}_k = (\mathbf{q}_k\mathbf{q}_k^T\mathbf{m}(\mathbf{x}))$, $k = 1, \ldots, n$ and $\mathbf{x} \leftarrow \arg\max_{\{\tilde{\mathbf{x}_k}\}} p(\tilde{\mathbf{x}}_k)$. Stopping criterion checks for $\|\mathbf{H}(\mathbf{x})\mathbf{g}(\mathbf{x}) - \widehat{\lambda}(\mathbf{x})\mathbf{g}(\mathbf{x})\| < threshold$, where $\widehat{\lambda}(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T\mathbf{H}(\mathbf{x})\mathbf{g}(\mathbf{x})/\mathbf{g}(\mathbf{x})^T\mathbf{g}(\mathbf{x})$. The iterations can be initialized to each data sample as in mean-shift clustering or at an arbitrary location and iterated until convergence to the principal curve. The generalized version of this algorithm that converges to the d-dimensional principal set is presented in Table 1. The threshold in step 6 checks if the gradient is in the span of only d eigenvectors of the local covariance whose other $(n-d)$ eigenvalues are negative by construction of the fixed-point ascent.

[1] Proofs will be temporarily available for ICASSP reviewers at `www.csee.ogi.edu/~ozertemu/icassp2008Proofs.pdf`

Table 1. Summary of Subspace Gaussian Mean-Shift to find P^d

1. Initialize the trajectories to a mesh or data points and $t = 0$.
2. For every trajectory evaluate $\mathbf{m}(\mathbf{x}(t))$ as in (3).
3. Evaluate the gradient, the Hessian, and perform the eigendecomposition of $\mathbf{\Sigma^{-1}}(\mathbf{x}(t)) = \mathbf{V\Gamma V}$.
4. Let $\mathbf{V_k} = [\mathbf{v}_{k_1} \ldots \mathbf{v}_{k_d}]$ be a particular d-subset of the eigenvectors determined by an index vector $\mathbf{k}$ that spans all $\binom{n}{d}$ subsets.
5. $\tilde{\mathbf{x}}_\mathbf{k} = \mathbf{V_k}\mathbf{V_k}^T\mathbf{m}(\mathbf{x})$, $\tilde{\mathbf{x}}_\mathbf{k}^* \leftarrow \arg\max_{\{\tilde{\mathbf{x}}_\mathbf{k}\}} p(\tilde{\mathbf{x}}_\mathbf{k})$.
6. If $\|\mathbf{g}(\mathbf{x}) - \mathbf{V_k^*}\mathbf{V_k^*}^T\mathbf{g}(\mathbf{x})\| < threshold$ then *stop*, else $\mathbf{x}(t+1) \leftarrow \tilde{\mathbf{x}}_\mathbf{k}^*$. Here $\mathbf{V_k^*}$ denotes the combination of eigenvectors that lead to $\tilde{\mathbf{x}}_\mathbf{k}^*$.
7. Convergence is not achieved. Increment t and go to step 2.

4. EXPERIMENTAL RESULTS

Loops, Self intersections and Bifurcations: A principal set of intrinsic dimension one could contain closed loops, self intersections, and bifurcations. This example is constructed to illustrate these possibilities in a synthetic *hangman* distribution (Figure 2, left). Traditional principal curve fitting approaches require explicit consideration of the occurrences of such irregularities, since they are specifically designed to fit *smooth* curves. On the other hand, recent manifold learning algorithms would possibly obtain similar looking results, however claiming that their outputs are actually samples of the principal curve in either traditional least-squares or current local likelihood subspace maximum perspectives would not be rigorous.

As the definition provides two simple local conditions to check to decide whether a point is in P^1 or not, one could run the algorithm in Table 1 from as many initial points as needed to populate P^1 to a desired density. For the hangman data, a KDE using isotropic fixed bandwidth Gaussian kernels is used, selecting the bandwidth by maximizing the leave-one-out cross-validation log-likelihood measure. The corresponding KDE and the resulting samples of P^1 upon iterating the subspace mean-shift algorithm in Table 1 from each data point are shown in Figure 2 (right).

Fractal Distributions: An interesting realization is that distributions with fractal structure are theoretically possible. Consequently, the corresponding critical and principal sets would have fractal structures. In a finite sample setting, clearly the small-scale components of these objects would not be reliably estimated, however, the proposed definition and the algorithm would determine the underlying fractal principal sets satisfactorily. To illustrate, we have created a *noisy tree* data set by perturbing the location of each pixel in a binary tree image by a Gaussian that has a covariance proportional the local k-nearest neighbor (KNN) covariance. The global scale of the perturbation is selected by maximizing the leave-one-out cross-validation measure as before, and the Gaussians used to perturb the data are also used to construct a variable bandwidth KDE.

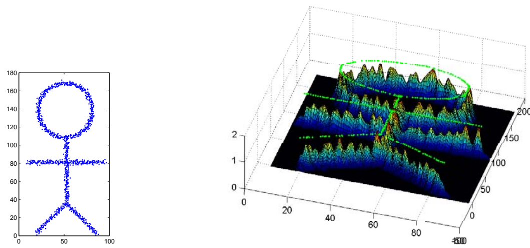

Fig. 2. The hangman dataset (left); the kernel density estimate of the dataset and corresponding principal set (right).
.

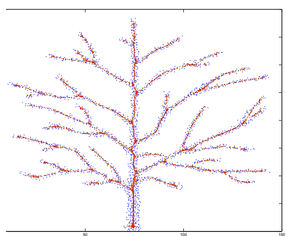

Fig. 3. The tree dataset; along with its principal set

Employing the subspace mean-shift algorithm in Table 1 to trajectories initialized to the noisy tree samples reveals samples from the *principle tree* of this data set as seen in Figure 3. Closer inspection indicates that all samples do not converge to the dominant branches of the principal tree as one would perceive through visual inspection and global information about the data distribution. Deterministic annealing of the global scale of the kernel bandwidths (keeping the local KNN-covariance portions fixed) could lead to samples from the principal tree in this case, or in general, from the more dominant (higher probability density) portions of P^1.

5. CONCLUSIONS

The evolution of learning regression models, principal surfaces, nonlinear principal components, most recently manifold learning, and the numerous applications these solutions play key roles in demonstrate that identifying principal manifolds is a fundamental problem of machine learning. The widely acknowledged and exploited definition of self-consistent principle surfaces based on the least-squares reconstruction error principle suffer from the use of semi-global expectations, leading to the requirement of partitioning the data before averages can be evaluated. For global identification, various piecewise local approximations are employed, and various heuristics for regularization are developed for specific applications. More recent propositions on manifold learning are either motivated by simplicity of convex optimization or the linear updates of various diffusion processes; in either case, the geometrical relevance of the resulting *manifold* to the desired principal manifolds, which are expected to exhibit certain invariance properties, are not clearly addressed.

In this paper, we proposed a generalization of the two simple local conditions on the gradient and the Hessian spectrum of a multidimensional pdf to determine whether a point is a critical/stationary point and specify whether it is a local maximum, minimum, or saddle. This generalization allows us to characterize the elements of ridges (principal curves) and valleys (minor curves), as well as higher dimensional critical (principal and minor) manifolds by simply checking two local conditions, provided that a density model is given. An interesting consequence of the proposed definition is a simple condition to check whether a point is in the one-dimensional critical set or not is that the gradient at this point must be an invariant vector of the Hessian (satisfied for critical points since gradient is zero, eigenvector otherwise). Local principality and minority can be assessed by observing the eigenvalue signs of the orthogonal subspace.

Although the proposition characterizes principal structures using only local first and second order information in a manner consistent with geometrical insights, it also yields challenging computational problems to be solved in future work: (1) globally principal smooth curves are a myth, one can only characterize principality on segments of the principal sets locally and patch selected segments under a global criterion to determine the global principle sets, which might have loops, bifurcations, and other irregularities; (2) ideal nonlinear projection of a data point to a lower dimensional manifold is given by the trajectory of a differential equation that follows the lines of curvature and fast projection and reconstruction algorithms are necessary for practical use; (3) regularization of the manifolds are implicitly handled by the model selection procedure used in density estimation, therefore the need for reliable regularization procedures in high dimensional spaces is evident.

6. REFERENCES

[1] S. Roweis, L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol 290, no. 5500, pp.2323–2326, 2000.

[2] B. Schlkopf, A. J. Smola, and K. R. Mller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 12991319, 1998.

[3] M. Belkin, P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," Advances in Neural Information Processing Systems, Cambridge, MA, MIT Press, 2002.

[4] T. Hastie, W. Stuetzle, "Principal Curves," *Jour. Am. Statistical Assoc.*, vol. 84, no. 406, pp. 502-516, 1989.

[5] R. Tibshirani, "Principal Curves Revisited," *Statistics and Computation*, vol. 2, pp. 183-190, 1992.

[6] S. Sandilya, S. R. Kulkarni, "Principal Curves with Bounded Turn," *IEEE Trans. on Information Theory*, vol. 48, no. 10, pp. 2789-2793, 2002.

[7] B. Kegl, A. Kryzak, T. Linder, K. Zeger, "Learning and Design of Principal Curves," *IEEE Trans. on PAMI*, vol. 22, no. 3, pp. 281-297, 2000.

[8] D. C. Stanford, A. E. Raftery, "Finding Curvilinear Features in Spatial Point Patterns: Principal Curve Clustering with Noise," *IEEE Trans. on PAMI*, vol. 22, no. 6, 601-609, 2000.

[9] D. Erdogmus, U. Ozertem, "Self-Consistent Locally Defined Principal Curves," *ICASSP 2007*, vol. 2, II-549 - II-552, 2007.

[10] B. Kegl, A. Kryzak, "Piecewise Linear Skeletonization Using Principal Curves," *IEEE Trans. on PAMI*, vol. 24, no. 1, 59-74, 2002.

[11] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Trans. on PAMI*, vol. 17, no. 8, pp. 790-799, 1995.