Hitotsubashi Journal of Commerce and Management 31 (1996) pp. 1-11.

ON ASYMPTOTICS OF LOCAL PRINCIPAL COMPONENT ANALYSIS

SHINGO OUE

Abstract

Assume data are randomly distributed on or near a smooth submanifold of $\mathbb{R}^d$. When applied to a subset of data in a neighborhood, principal component analysis (PCA) gives an estimate of the tangent and normal spaces of the underlying manifold. Asymptotic properties of the estimate is surveyed in connection with variations of data and curvatures of the manifold. A dimension estimate based on the work of Waternaux (1975, 1976) is also considered.

1. *Introduction*

Principal component analysis (PCA) is concerned with explaining the variation of data by presenting a new orthonormal basis whose vectors represent the directions of maximum variability. Let $c_1 \geq c_2 \geq \cdots \geq c_d$ be the eigenvalues of the sample covariance matrix and let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d$ be the corresponding eigenvectors, respectively. If the last $d-k$ eigenvalues are all small for $k \in \{1, \ldots, d-1\}$, then the linear subspace spanned by the first k eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is parallel to the "best" affine subspace approximation of the data. By analogy, PCA restricted to data points in a small neighborhood is considered as a procedure to estimate the tangent and normal spaces for the neighborhood: the first k eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ span a tangent space estimate and the last $d-k$ eigenvectors $\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_d$ span a normal space estimate.

An explicit use of local PCA in connection with differentiable manifolds has appeared in Hoppe, DeRose, Duchamp, McDonald, and Stuetzle (1992) where they consider surface reconstruction in $\mathbb{R}^3$ from randomly distributed points on or near the surface. The method they have developed requires an estimate of the normal vector for which the last eigenvector of local PCA is used. Hoppe et al, however, do not inquire further into the property of the last eigenvector as an estimate of the normal vector. In this paper, I would like to clarify how variations in data, curvatures of the underlying manifold, and size of the neighborhood relate to the performance of the estimates for the general case of k-dimensional manifold in $\mathbb{R}^d$ $(1 \leq k < d)$. I also consider the estimation of the dimension k when it is unknown. The proposed method is based on the work of Waternaux (1975, 1976).

2. *Covariance Matrix of Manifold-valued Data*

We assume that data are randomly distributed on a smooth manifold, possibly with noise. The eigensystem of the sample covariance matrix is computed from data points within a small neighborhood. When the size of the neighborhood is fixed, it follows from the multivariate central limit theorem that the sample covariance matrix is a $\sqrt{N}$-consistent estimator of the population covariance matrix where N is the number of observations in the neighborhood. Both the eigenvalues and eigenvectors of the sample covariance matrix are also known as $\sqrt{N}$-consistent estimators of their population counterparts (Waternaux 1975). Our particular interest is in how well the eigenvectors approximate the tangent and normal vectors of the underlying manifold.

The following notations are used throughout this paper:

Let A be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. The same symbol A is used for both the transformation and the corresponding $m \times n$ matrix with respect to the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. The distinction should be clear from the context.

Suppose B is a $n_1 \times n_2 \times n_3$ array of real numbers. The 3-way array B can be thought of as $n_2 \times n_3$ matrix whose $(i,j)^{\text{th}}$ component is a n_1 vector $B_{\cdot ij}$, or as n_1 vector whose i^{th} component is a $n_2 \times n_3$ matrix $B_{i\cdot\cdot}$. For the case $n_1 = n$ and $n_2 = n_3 = m$, we define $\boldsymbol{w}^\top[B]\boldsymbol{w}$ ($\boldsymbol{w} \in \mathbb{R}^m$) to be the vector in $\mathbb{R}^n$ given by

$$(\boldsymbol{w}^\top B_{1\cdot\cdot}\boldsymbol{w}, \ldots, \boldsymbol{w}^\top B_{n\cdot\cdot}\boldsymbol{w})^\top = \sum_{i,j=1}^{m} w_i w_j B_{\cdot ij}$$

The symbol $\mathbf{0}_n$ stands for the zero vector in $\mathbb{R}^n$ and $0_{n\times m}$ stands for the $n \times m$ matrix of zeros.

2. 1 Differentiable Submanifolds in $\mathbb{R}^d$

Let M be a k-dimensional imbedded C^m-submanifold of $\mathbb{R}^d$ with $1 \le k < d$; that is, for each $\boldsymbol{x}_0 \in M$, there exist an open subset $\mathcal{N}$ of $\mathbb{R}^d$ and a function φ such that

1. $\boldsymbol{x}_0 \in \mathcal{N}$;
2. φ is a bijection of class C^m from $\mathcal{W}$ onto $M \cap \mathcal{N}$ where $\mathcal{W}$ is an open subset of $\mathbb{R}^k$;
3. The rank of $\varphi'(\boldsymbol{w})$ is k for every $\boldsymbol{w} \in \mathcal{W}$;
4. φ^{-1} is continuous (with respect to the relative topology of $M \cap \mathcal{N}$).

For our purpose, it is sufficient to assume that M is a C^3-submanifold.

The *tangent* space $T(\boldsymbol{x}_0)$ of M at $\boldsymbol{x}_0$ can be identified as a k-dimensional linear subspace of $\mathbb{R}^d$ given by

$$T(\boldsymbol{x}_0) = \{\varphi'(\boldsymbol{w}_0)\boldsymbol{z} \ : \ \boldsymbol{z} \in \mathbb{R}^k\}$$

where $\boldsymbol{w}_0$ is the unique point satisfying $\boldsymbol{x}_0 = \varphi(\boldsymbol{w}_0)$. It can be shown that $T(\boldsymbol{x}_0)$ is, in fact, independent of the particular parameterization φ. We also denote the orthogonal complement of $T(\boldsymbol{x}_0)$, called the *normal space*, by $T(\boldsymbol{x}_0)^\perp$. For more general treatments of differentiable manifolds, we refer the reader to, e.g., Boothby (1986).

Let

$$\varphi''(\boldsymbol{w}_0) = \left[\left. \frac{\partial^2 \varphi_h}{\partial w_i \partial w_j} \right|_{\boldsymbol{w}=\boldsymbol{w}_0} \right]_{hij}$$

be the $d \times k \times k$ array of the second derivative of φ at $\boldsymbol{w}_0$. For each $\boldsymbol{w} \in \mathbb{R}^k$, let $\eta_{\boldsymbol{w}}(t) = \varphi(\boldsymbol{w}_0 + t\boldsymbol{w})$ be the curve on $M \cap \mathcal{N}$ going through $\boldsymbol{x}_0$ in the direction $\boldsymbol{w}$. The vector

$$
\begin{aligned}
\ddot{\eta}_{\boldsymbol{w}} &= \left.\frac{d^2\eta_{\boldsymbol{w}}}{dt^2}\right|_{t=0} = \boldsymbol{w}^\top[\varphi''(\boldsymbol{w}_0)]\boldsymbol{w} \\
&= \sum_{i,j=1}^{k} w_i w_j \left(\left.\frac{\partial^2\varphi}{\partial w_i \partial w_j}\right|_{\boldsymbol{w}=\boldsymbol{w}_0} \right) = (\boldsymbol{w}^\top \varphi_1''(\boldsymbol{w}_0)\boldsymbol{w}, \ldots, \boldsymbol{w}^\top \varphi_d''(\boldsymbol{w}_0)\boldsymbol{w})^\top
\end{aligned}
$$

is called the acceleration vector, following Bates and Watts (1980). If we imagine a point moving along the curve $\eta_{\boldsymbol{w}}(t)$, $\ddot{\eta}_{\boldsymbol{w}}$ is the instantaneous acceleration when $t = 0$.

The acceleration $\ddot{\eta}_{\boldsymbol{w}}$ can be uniquely decomposed as $\ddot{\eta}_{\boldsymbol{w}} = \ddot{\eta}_{\boldsymbol{w}}^N + \ddot{\eta}_{\boldsymbol{w}}^T$ where $\ddot{\eta}_{\boldsymbol{w}}^T \in T(\boldsymbol{x}_0)$ and $\ddot{\eta}_{\boldsymbol{w}}^N \in T(\boldsymbol{x}_0)^\perp$. The quantity

$$
K_{\boldsymbol{w}} = \frac{\|\ddot{\eta}_{\boldsymbol{w}}^N\|}{\|\varphi'(\boldsymbol{w}_0)\boldsymbol{w}\|^2} \tag{1}
$$

is called the ***intrinsic curvature in the direction*** $\boldsymbol{w}$, which is, in fact, the inverse of the radius of the circle which best approximates the curve $\eta_{\boldsymbol{w}}$ at $t = 0$. The curvatures are also invariant under re-parameterization.

2. 2 Eigensystem of the Covariance Matrix

Let $\mathcal{N}_0$ be a bounded open subset of $\mathbb{R}^d$ satisfying $\boldsymbol{x}_0 \in \mathcal{N}_0 \subset \mathcal{N}$ and let $\delta \in (0, \infty)$ be the diameter of $\mathcal{N}_0$ defined by

$$
\delta = \sup\{\|\boldsymbol{x} - \boldsymbol{x}'\| : \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{N}_0\}.
$$

This is the neighborhood on which we compute covariance matrices. Since $\varphi : \mathcal{W} \to M \cap \mathcal{N}$ is a homeomorphism, $\mathcal{W}_0 = \varphi^{-1}(M \cap \mathcal{N}_0)$ is also a bounded open subset of $\mathcal{W}$.

We assume that the data $\boldsymbol{X}_i$, $i = 1, \ldots, N$ are *iid* and that each $\boldsymbol{X}_i$ has the form

$$
\boldsymbol{X}_i = \boldsymbol{Y}_i + \boldsymbol{\varepsilon}_i
$$

where $\boldsymbol{Y}_i$ and $\boldsymbol{\varepsilon}_i$ are independent random vectors in $\mathbb{R}^d$. The distribution P of $\boldsymbol{Y}_i$ is defined on the Borel field of M and assumed to satisfy $P(M) = 1$. Also, the noise component $\boldsymbol{\varepsilon}_i$ is assumed to have the mean $\mathbf{0}_d$ and covariance $\tau^2 I$ where $\tau \geq 0$ is sufficiently small compared with δ. Since we restrict our analysis to the neighborhood $\mathcal{N}_0$, let us normalize P conditioned on $\mathcal{N}_0$ by

$$
P_0 = P/P(\mathcal{N}_0 \cap M).
$$

Since φ is a homeomorphism, φ^{-1} induces a probability measure $P_0\varphi$ on $\mathcal{W}_0$ which is defined by $P_0\varphi(A) = P_0(\varphi(A))$ for any Borel set A of $\mathcal{W}_0$. The covariance of $\boldsymbol{X}_i$ is given by

$$
\Sigma + I\tau^2 \tag{2}
$$

where

$$
\Sigma = E[(\boldsymbol{Y} - E\boldsymbol{Y})(\boldsymbol{Y} - E\boldsymbol{Y})^\top] \tag{3}
$$

and $\boldsymbol{Y} \sim P_0$.

Let $\sigma_1 \geq \cdots \geq \sigma_d$ be the eigenvalues of Σ. It follows from (2) that the eigenvalues of the covariance matrix for the noisy data $\{\boldsymbol{X}_i\}$ are $\sigma_j + \tau^2$, $j = 1, \ldots, d$ and the eigenvectors are the same as those for the noiseless data $\{\boldsymbol{Y}_i\}$. If τ^2 is very large and

$$\sigma_1 + \tau^2 \approx \sigma_2 + \tau^2 \approx \cdots \approx \sigma_d + \tau^2,$$

the eigenvectors for $\{\boldsymbol{X}_i\}$ are almost indeterminate. Geometrically, it implies that the noise component $\boldsymbol{\varepsilon}_i$ inflates the low-dimensional structure toward a ball in $\mathbb{R}^d$.

3. *Second-order Approximation of Eigensystem*

Let

$$\varphi'(\boldsymbol{w}_0) = QR \tag{4}$$

where Q is a $d \times d$ orthogonal matrix and R is a $d \times k$ right triangular matrix, i.e.,

$$R = \begin{bmatrix} R_1 \\ \mathbf{0}_{(d-k)\times k} \end{bmatrix} \quad \text{and} \quad R_1 = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ 0 & r_{22} & \cdots & r_{2k} \\ & & \cdots & \\ 0 & 0 & \cdots & r_{kk} \end{bmatrix}.$$

The RHS of (4) is the QR decomposition of $\varphi'(\boldsymbol{w}_0)$. Since $\varphi'(\boldsymbol{w}_0)$ is of rank k, the triangular matrix R_1 is nonsingular and, hence, $r_{ii} \neq 0$, $i = 1, \ldots, k$. For later purposes, let $Q = [Q_1 \mid Q_2]$ where Q_1 is the $d \times k$ matrix of the first k columns of Q.

Now, we introduce another local parameterization around $\boldsymbol{x}_0 \in M$. Let $f : \boldsymbol{w} \mapsto \boldsymbol{v}$ be the affine transformation on $\mathcal{W}$ given by

$$\boldsymbol{v} = f(\boldsymbol{w}) = R_1(\boldsymbol{w} - \boldsymbol{w}_0).$$

Since R_1 is nonsingular, f maps $\mathcal{W}$ onto a nonempty open subset $\mathcal{V}$ of $\mathbb{R}^k$ and $\mathcal{W}_0$ onto a bounded open subset $\mathcal{V}_0 \subset \mathcal{V}$ such that $\mathbf{0}_k \in \mathcal{V}_0$. Obviously, f is a diffeomorphism between $\mathcal{W}$ and $\mathcal{V}$ so that we can define a new parameterization $\phi = \varphi \circ f^{-1}$. Then, by chain rule,

$$\phi'(\mathbf{0}_k) = \varphi'(\boldsymbol{w}_0) \left(\left. \frac{d(f^{-1})}{d\boldsymbol{v}} \right|_{\boldsymbol{v}=\mathbf{0}_k} \right) = QRR_1^{-1} = Q \begin{pmatrix} I_k \\ \mathbf{0}_{(d-k)\times k} \end{pmatrix}. \tag{5}$$

Let δ_ϕ be the diameter of $\mathcal{V}_0$. Also, Let $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ be points in $\mathcal{V}_0$ and let $\boldsymbol{x}_i = \phi(\boldsymbol{v}_i)$ for $i = 1, 2$. Since

$$\boldsymbol{x}_i = \phi'(\mathbf{0}_k)\boldsymbol{v}_i + O(\delta_\phi^2)$$

and Q is orthogonal,

$$\begin{aligned} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| &= \left\| Q \begin{pmatrix} \boldsymbol{v}_1 - \boldsymbol{v}_2 \\ \mathbf{0}_{d-k} \end{pmatrix} \right\| + O(\delta_\phi^{3/2}) \\ &= \|\boldsymbol{v}_1 - \boldsymbol{v}_2\| + O(\delta_\phi^{3/2}). \end{aligned}$$

This implies that

$$\delta/\delta_\phi = 1 + o(1)$$

and the diameters δ and δ_ϕ are asymptotically equivalent.

It follows from the definition of the tangent space that

$$T(\boldsymbol{x}_0) = \phi'(\mathbf{0}_k)\mathrm{I\!R}^k = Q\mathcal{L} \tag{6}$$

where

$$\mathcal{L} = \mathrm{I\!R}^k \times \{\mathbf{0}_{d-k}\} = \{\boldsymbol{x} \in \mathrm{I\!R}^d : x_{k+1} = \cdots = x_d = 0\}.$$

Since the tangent space $T(\boldsymbol{x}_0)$ is a k-dimensional linear subspace of $\mathrm{I\!R}^d$, it can be obtained by rotating the linear subspace $\mathcal{L}$. The equation (6) gives an orthogonal matrix which represents the rotation, though $\det Q$ can be -1 because a basis in $T(\boldsymbol{x}_0)$ can have a different orientation from the canonical basis in $\mathcal{L}$. The representation of $T(\boldsymbol{x}_0)$ which appears on the RHS of (6) suggests an advantage of the coordinates with respect to the column vectors of Q. That is, for any vector of $\mathrm{I\!R}^d$, the first k coordinates correspond to the tangent space, and the last $d-k$ coordinates correspond to the normal space under the Q-coordinate system. The result can be directly translated into the original coordinate system through the orthogonal transformation Q. It is also suggested that the columns of Q_1 (or Q_2) form an orthonormal basis of $T(\boldsymbol{x}_0)$ (or $T(\boldsymbol{x}_0)^\perp$).

Suppose $\boldsymbol{V}$ is a random vector in $\mathrm{I\!R}^k$ such that $\boldsymbol{Y} = \phi(\boldsymbol{V})$ and $\boldsymbol{Y} \sim P_0$. We assume that the covariance of $\boldsymbol{V}$ is positive definite. Now, let us look at the second order expansion of ϕ at $\mathbf{0}_k$,

$$\phi(\boldsymbol{V}) = \phi(\mathbf{0}_k) + \phi'(\mathbf{0}_k)\boldsymbol{V} + \boldsymbol{V}^\top[\phi''(\mathbf{0}_k)]\boldsymbol{V} + O(\|\boldsymbol{V}\|^3) \tag{7}$$

Using the QR decomposition (5), (7) can be expressed as

$$\phi(\boldsymbol{V}) = \phi(\mathbf{0}_k) + Q\left[\begin{pmatrix} \boldsymbol{V} \\ \mathbf{0}_{d-k} \end{pmatrix} + \begin{pmatrix} \boldsymbol{V}^\top[A_1]\boldsymbol{V} \\ \boldsymbol{V}^\top[A_2]\boldsymbol{V} \end{pmatrix}\right] + O(\|\boldsymbol{V}\|^3) \tag{8}$$

where $\binom{A_1}{A_2}$ is the representation of $\phi''(\mathbf{0}_k)$ in the Q-coordinate system, A_1 is the $k \times k \times k$ array of the tangent components, A_2 is the $(d-k) \times k \times k$ array of the normal components and

$$\boldsymbol{V}^\top[A_1]\boldsymbol{V} = (\boldsymbol{V}^\top[\phi''(\mathbf{0}_k)]\boldsymbol{V})^T, \quad \boldsymbol{V}^\top[A_2]\boldsymbol{V} = (\boldsymbol{V}^\top[\phi''(\mathbf{0}_k)]\boldsymbol{V})^N.$$

The $(h,i,j)^{\text{th}}$ element of A_1 is given by

$$\boldsymbol{q}_h^\top \left(\partial^2\phi/\partial v_i \partial v_j\big|_{\boldsymbol{v}=\mathbf{0}_k}\right)$$

where $\boldsymbol{q}_h$ is the hth column of Q_1. From the equation (8), we see that, at low order, elements of the normal space enter the covariance matrix through $\boldsymbol{V}^\top[A_2]\boldsymbol{V}$.

Let us define the following centered random vectors

$$\begin{aligned}
\boldsymbol{V}^{(0)} &= \boldsymbol{V} - E\boldsymbol{V}, \\
\boldsymbol{V}^{(1)} &= \boldsymbol{V}^\top[A_1]\boldsymbol{V} - E\left[\boldsymbol{V}^\top[A_1]\boldsymbol{V}\right], \\
\boldsymbol{V}^{(2)} &= \boldsymbol{V}^\top[A_2]\boldsymbol{V} - E\left[\boldsymbol{V}^\top[A_2]\boldsymbol{V}\right]
\end{aligned}$$

The vectors $\boldsymbol{V}^{(0)}$, $\boldsymbol{V}^{(1)}$, and $\boldsymbol{V}^{(2)}$ can be seen as the first order variation in the tangent space, the second order variation in the tangent space, and the second order variation in the normal space, respectively.

Now, using the second-order approximation (8), the covariance matrix Σ can be expanded as

$$\Sigma = Q(\Sigma_1 + \delta^2\Sigma_2)Q^\top + O(\delta^4)$$

where

$$\Sigma_1 = \begin{bmatrix} \mathrm{Cov}(\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)}) & 0_{k\times(d-k)} \\ 0_{(d-k)\times k} & 0_{(d-k)\times(d-k)} \end{bmatrix}, \tag{9}$$

$$\Sigma_2 = \delta^{-2}\begin{bmatrix} 0_{k\times k} & C \\ C^\top & \mathrm{Cov}(\boldsymbol{V}^{(2)}) \end{bmatrix}, \tag{10}$$

and

$$C = E[(\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)})\boldsymbol{V}^{(2)\top}].$$

Note that, after taking the expectation, the remainder term is expressed with respect to δ since $\|\boldsymbol{V}\| \leq \delta_\phi$ and δ_ϕ is asymptotically equivalent to δ.

Let $\sigma_1^1 \geq \cdots \geq \sigma_d^1$ be the eigenvalues of Σ_1 and let $\boldsymbol{b}_1^1, \ldots, \boldsymbol{b}_d^1$ be the corresponding eigenvectors, respectively. Also, let $\sigma_1 \geq \cdots \geq \sigma_d$ be the eigenvalues of Σ. The eigenvectors of Σ_1 are supposed to be normalized so that their norms are equal to unity. It follows from (9) that

$$\sigma_{k+1}^1 = \sigma_{k+2}^1 = \ldots = \sigma_d^1 = 0 \tag{11}$$

and

$$\boldsymbol{b}_i^1 \in \mathcal{L} \text{ for } i = 1, \ldots, k, \quad \boldsymbol{b}_i^1 \in \mathcal{L}^\perp \text{ for } i = k+1, \ldots, d \tag{12}$$

Since $\boldsymbol{b}_{k+1}^1, \ldots, \boldsymbol{b}_d^1$ which span $\mathcal{L}^\perp$ have the same eigenvalue 0, we can replace them by any orthonormal set of vectors spanning $\mathcal{L}^\perp$. Our choice is to set $\boldsymbol{b}_i^1 = \boldsymbol{e}_i$, $i = k+1, \ldots, d$.

Let $i \in \{1, \ldots, k\}$. We assume that the multiplicity of σ_i^1 is one so that we can use an expansion of the i^{th} eigenvector of Σ around $\boldsymbol{b}_i^1$ with respect to δ. The multiplicity of the other eigenvalues is arbitrary in the expansions. We also have to rescale the covariance matrix with respect to δ, since the size of eigenvalues σ_i^1 and σ_i also depends on δ. For that purpose, let us make a few additional assumptions. First, the neighborhood $\mathcal{N}_0$ is taken to be the $(\delta/2)$-ball $\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta/2\}$ so that the neighborhood shrinks symmetrically in every direction as δ becomes smaller. Secondly, we assume that M is orientable and P_0 has a positive density g with respect to a volume element Ω of M, i.e.,

$$P_0(A) = \int_A g\Omega$$

(Boothby 1986). This induces a positive density g_0 with respect to the Lebesgue measure on $\mathcal{W}_0$. By applying the mean value theorem to g_0, the variance of any *one*-dimensional marginal distribution of $P_0\phi$ (and, therefore, of P_0) is of order δ^2 as $\delta \to 0$. With these assumptions and the simplicity of σ_i^1,

$$\sigma_i^* = \delta^{-2}\sigma_i^1 > 0, \quad \sigma_i^* = O(1) \text{ as } \delta \to 0$$

Then, using a perturbation argument (Wilkinson 1965, pp.66-70), for sufficiently small δ,

$$\begin{aligned} \boldsymbol{b}_i &= \boldsymbol{b}_i^1 + \delta^2 \left[\sum_{j \neq i} \frac{\beta_{ji} \boldsymbol{b}_j^1}{\sigma_i^* - \sigma_j^*} \right] + O(\delta^3), \\ \delta^{-2}\sigma_i &= \sigma_i^* + \delta^2 \beta_{ii} + \delta^4 \left[\sum_{j \neq i} \frac{\beta_{ji}\beta_{ij}}{\sigma_i^* - \sigma_j^*} \right] + O(\delta^4) \end{aligned}$$

where

$$\beta_{ji} = \boldsymbol{b}_j^{1\top} (\delta^{-2}\Sigma_2) \boldsymbol{b}_i^1 = \beta_{ij}$$

and $\boldsymbol{b}_i$ is an unnormalized eigenvector corresponding to σ_i. It follows from (10) that

$$\beta_{ji} = \begin{cases} 0 & \text{for } j \in \{1, \dots, k\} \\ \frac{1}{\sigma_i^*} E[\delta^{-4} \boldsymbol{b}_i^{1k\top} (\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)}) V_j^{(2)}] & \text{for } j \in \{k+1, \dots, d\} \end{cases}$$

where $\boldsymbol{b}_i^{1k} \in \mathbb{R}^k$ is the vector of the first k components of $\boldsymbol{b}_i^1$. Since $\boldsymbol{b}_i^1 \in \mathcal{L}$, $\boldsymbol{b}_i^{1k}$ is also a unit vector. Then, using (11),

$$\boldsymbol{b}_i = \boldsymbol{b}_i^1 + \frac{\delta}{\sigma_i^*} \begin{pmatrix} \boldsymbol{0}_k \\ E[\delta^{-3}(\boldsymbol{b}_i^{1k\top} \boldsymbol{V}^{(0)}) \boldsymbol{V}^{(2)}] \end{pmatrix} + O(\delta^2) \tag{13}$$

$$\delta^{-2}\sigma_i = \sigma_i^* + \frac{\delta^3}{\sigma_i^*} \sum_{l=1}^{d-k} E[\delta^{-3}(\boldsymbol{b}_i^{1k\top} \boldsymbol{V}^{(0)}) V_l^{(2)}] + O(\delta^4) \tag{14}$$

The first order term in the expansion (13) is a vector in $\mathcal{L}^\perp$. For $l = 1, \dots, d-k$, the $(k+l)^{\text{th}}$ coordinate of the first order term is proportional to

1. the inverse of the eigenvalue σ_i^*,
2. the covariance between the $\boldsymbol{q}_l$-coordinate variation $V_l^{(2)}$ in the normal space and the projection of the variation $\boldsymbol{V}^{(0)}$ onto the $\boldsymbol{b}_i^{1k}$-direction in the tangent space.

Let $\boldsymbol{t}$ be a unit vector in $\mathbb{R}^k$. When the manifold M is fairly flat in the direction of $Q \begin{pmatrix} \boldsymbol{t} \\ \boldsymbol{0}_{d-k} \end{pmatrix}$ near $\boldsymbol{x}_0$, the linear order term along the direction, $\boldsymbol{t}^\top \boldsymbol{V}^{(0)}$ in the Q_1-coordinate system, is likely to be large compared with the second order terms $\boldsymbol{t}^\top \boldsymbol{V}^{(1)}$ and $V_m^{(2)}$, $m = 1, \dots, d-k$. Since $\boldsymbol{V}^{(0)}$ is bounded, this implies that the covariance between $\boldsymbol{t}^\top \boldsymbol{V}^{(0)}$ and $V_m^{(2)}$ tends to be small. Although the variances of $\boldsymbol{V}^{(0)}$, $\boldsymbol{V}^{(1)}$ and $\boldsymbol{V}^{(2)}$ depend on their probability distributions, the idea suggests the influential role played by the underlying geometry of the support $M \cap \mathcal{N}_0$. Since $\sigma_1^*, \dots, \sigma_k^*$ are the eigenvalues of $\delta^{-2}\mathrm{Cov}(\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)})$, $\boldsymbol{b}_i^1$ with large $i \in \{1, \dots, k\}$ corresponds to the direction where the variation of $\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)}$ is relatively small and, hence, M is less linear. It follows that the eigenvectors $Q\boldsymbol{b}_i$ of Σ with large $i \in \{1, \dots, k\}$ is more likely to be contaminated since

1. σ_i^* is smaller than σ_j^* for $j < i$,

2. $\boldsymbol{b}_i^1$ corresponds to the direction where M is less linear.

Therefore, the tangent space estimate spanned by $Q\boldsymbol{b}_1, \ldots, Q\boldsymbol{b}_k$ is actually dragged to the manifold M in the directions where M is relatively nonlinear.

The expansion (14) shows a similar result for the eigenvalue: the third order coefficient is proportional to both σ_i^{*-1} and the sum of covariances $\sum_{l=1}^{d-k} E[(\boldsymbol{b}_i^{1k\top}\boldsymbol{V}^{(0)})V_l^{(2)}]$.

In the rest of this section, we derive a few inequalities which relate the contamination from the normal space to the curvatures of M. First, the orthogonality of the matrix Q implies that

$$\delta^{-2}\|\boldsymbol{V}^{(2)}\| \leq \frac{\|\boldsymbol{V}^{(2)}\|}{\|\boldsymbol{V}\|^2} = \frac{\|\boldsymbol{V}^{(2)}\|}{\|Q^\top\boldsymbol{V}\|^2} = K_{\boldsymbol{V}},$$

where $K_{\boldsymbol{V}}$ is the intrinsic curvature in the direction of $\boldsymbol{V}$ (1) and satisfies

$$K_{\boldsymbol{V}} \leq \sup_{\|\boldsymbol{v}\|=1} K_{\boldsymbol{v}} = K < \infty$$

Then, by Jensen's inequality,

$$\left\|E[\delta^{-2}\boldsymbol{V}^{(2)}]\right\|^2 = \sum_{l=1}^{d-k}\left|E[\delta^{-2}V_l^{(2)}]\right|^2 \leq \sum_{l=1}^{d-k} E\left|\delta^{-2}V_l^{(2)}\right|^2 = E[\delta^{-4}\|\boldsymbol{V}^{(2)}\|^2] \leq K^2. \tag{15}$$

Since $\|\boldsymbol{b}_i^{1k}\| = 1$, it follows from the Cauchy-Schwarz inequality that

$$|\boldsymbol{b}_i^{1k\top}\boldsymbol{V}^{(0)}| \leq \|\boldsymbol{V}^{(0)}\| \leq \delta. \tag{16}$$

Then, by (15) and (16), the norm of the first order term in (13) is bounded by

$$\frac{\delta}{\sigma_i^*}\left\|\begin{pmatrix} \boldsymbol{0}_k \\ E[\delta^{-3}(\boldsymbol{b}_i^{1k\top}\boldsymbol{V}^{(0)})\boldsymbol{V}^{(2)}] \end{pmatrix}\right\| \leq \frac{K}{\sigma_i^*}\delta. \tag{17}$$

Since

$$\left|\sum_{l=1}^{d-k} V_l^{(2)}\right| \leq \sum_{l=1}^{d-k}|V_l^{(2)}| \leq \sqrt{d-k}\|\boldsymbol{V}^{(2)}\| \leq \sqrt{d-k}\delta^2 K,$$

the absolute value of the third order term in (14) is also bounded by

$$\left|\frac{\delta^3}{\sigma_i^*}\sum_{l=1}^{d-k} E[\delta^{-3}(\boldsymbol{b}_i^{1k\top}\boldsymbol{V}^{(0)})V_l^{(2)}]\right| \leq \frac{\sqrt{d-k}K}{\sigma_i^*}\delta^3. \tag{18}$$

The same argument does not hold for $\sigma_{k+1}, \ldots, \sigma_d$ since they correspond to the multiple eigenvalue 0. However, using the Gerschgorin circle theorem (see, e.g., Noble and Daniel 1977), we can still access their sizes. In fact, for $i \in \{k+1, \ldots, d\}$,

$$\delta^{-2}\sigma_i \leq \|\Sigma_2\|_2 + O(\delta^2) \leq \|\Sigma_2\|_F + O(\delta^2) \tag{19}$$

where $||A||_2$ is the Euclidean (l_2) matrix norm and $||A||_F$ is the Frobenius norm of $A = [a_{ij}]$ given by

$$\begin{aligned} ||A||_2 &= \max_{||v||=1} \frac{||Av||}{||v||}, \\ ||A||_F &= \left(\sum_i \sum_j |a_{ij}|^2 \right)^{1/2}. \end{aligned}$$

Then, the leading term (first order) in the RHS of (19) is bounded by

$$\left(2 \sum_{l=1}^{k} \sum_{m=1}^{d-k} \left| \delta^{-2} E[V_l^{(0)} V_m^{(2)}] \right|^2 \right)^{1/2} \leq \sqrt{2k(d-k)} K \delta. \tag{20}$$

4. *Local Dimension Estimation*

Let γ_i, $i = 1, \ldots, d$, be the eigenvalues of the population covariance matrix of $\boldsymbol{X}_1$. If M is a k-dimensional linear submanifold,

$$\gamma_{k+1} = \gamma_{k+2} = \cdots = \gamma_d = \tau^2.$$

As we have seen in the previous section, γ_i, $i = k+1, \ldots, d$, may be greater than τ^2 for a general submanifold M and the amount of deviations from τ^2 is related to the curvatures of M. (See (20).)

Let κ_r^i be the r^{th} cumulant of the i^{th} component X_{i1} of $\boldsymbol{X}_1$; that is, κ_r^i is the coefficient of $(it)^r/r!$ in the expansion in powers of t of $\log \phi(t)$ where $\phi(t)$ is the characteristic function of X_{i1}. Similarly, the bivariate cumulant κ_{rs}^{ij} of X_{i1} and X_{j1} is defined as the coefficient of

$$\frac{(it_i)^r}{r!} \frac{(it_j)^s}{s!}$$

in the expansion of the joint log-characteristic function $\log \phi(t_i, t_j)$ of X_{i1} and X_{j1}. Let $c_1 \geq c_2 \geq \cdots \geq c_d$ be the eigenvalues of the sample covariance matrix. Waternaux (1975, 1976) has shown that, if $\gamma_{k+1} > \gamma_{k+2} > \cdots > \gamma_d$, the asymptotic distribution of

$$\sqrt{N} \left[\begin{pmatrix} c_{k+1} \\ \vdots \\ c_d \end{pmatrix} - \begin{pmatrix} \gamma_{k+1} \\ \vdots \\ \gamma_d \end{pmatrix} \right] \tag{21}$$

is normal with mean zero and covariance matrix $\Lambda = [\lambda_{ij}]$ where

$$\lambda_{ij} = \begin{cases} 2\gamma_i^2 + \kappa_4^i & \text{for } i = j \\ \kappa_{22}^{ij} & \text{for } i \neq j \end{cases} \tag{22}$$

Anderson (1963) has derived the same asymptotic distribution also for the spherical case $\gamma_{k+1} = \cdots = \gamma_d$ under the assumption of normal population ($\kappa_4^i = \kappa_{22}^{ij} = 0$, $\forall(i,j)$). His result for the spherical case can be directly generalized to nonnormal populations if

$$\begin{aligned} \kappa_{22}^{ij} &= \frac{1}{2}\kappa_4^i \qquad i \neq j \\ \kappa_{1111}^{ijlm} &= 0 \qquad (i,j,l,m) \neq (i,i,i,i) \text{ and } (i,j) \neq (l,m) \end{aligned}$$

Otherwise, the expression for the joint distribution becomes intractable for $d-k>2$ (Waternaux 1975).

According to (21) and (22), $\sqrt{N}(\bar{c}-\bar{\gamma})$ where

$$\bar{c} = \sum_{i=k+1}^{d} c_i/(d-k), \quad \bar{\gamma} = \sum_{i=k+1}^{d} \gamma_i/(d-k)$$

is asymptotically normal with mean zero and variance

$$\frac{2\sum_{i=k+1}^{d}\gamma_i^2 + \sum_{i=k+1}^{d}\kappa_4^i + \sum_{i,j=k+1}^{d}\kappa_{22}^{ij}}{(d-k)^2} \tag{23}$$

By applying the mean value theorem to the positive density g_0 of $P_0\phi$, the cumulants of $\boldsymbol{Y}_1$ can be approximated by those of the uniform distribution on $\mathcal{W}_0$ for which $\kappa_4^i < 0$ and $\kappa_{22}^{ij} = 0$. If the noise component $\boldsymbol{\varepsilon}_i$ is normal or any random vector which has $\kappa_4^i \leq 0$ and $\kappa_{22}^{ij} = 0$, the variance given by (23) is smaller than $2(\sum_{i=k+1}^{d}\gamma_i^2)/(d-k)^2$ for small δ. Let z_α be such that

$$\int_{-\infty}^{z_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1-\alpha.$$

Using the consistent estimator $2(\sum_{i=k+1}^{d} c_i^2)/(d-k)^2$ of $2(\sum_{i=k+1}^{d}\gamma_i^2)/(d-k)^2$ in place of the variance (23), an asymptotic test of level α for the null hypothesis $\bar{\gamma}=0$ is defined to reject the hypothesis if

$$\bar{c} > z_\alpha \left(\frac{2\sum_{i=k+1}^{d} c_i^2}{N(d-k)^2}\right)^{1/2}. \tag{24}$$

In order to estimate the local dimension k of $\mathcal{N}_0$, we specify an admissible upper bound C for τ^2/δ^2. Then, our dimension estimate is defined as the smallest k which satisfies

$$\bar{c} > C\delta^2 + z_\alpha \left(\frac{2\sum_{i=k+1}^{d} c_i^2}{N(d-k)^2}\right)^{1/2}.$$

One direction relating to future research is to connect the test given by (24) with the work of Cutler (1986) and Cutler and Dawson (1989) on a general theory of local dimension.

HITOTSUBASHI UNIVERSITY

REFERENCES

1. Anderson, T. W. (1963) Asymptotic Theory for Principal Component Analysis. *The Annals of Mathematical Statistics*, **34**, 122-148.

2. Bates, D. M. and Watts, D. G. (1980) Relative Curvature Measures of Nonlinearity (with discussion). *Journal of Royal Statistical Society*, B, **42**, 1, 1-25.

3. Boothby, W. M. (1986) *An Introduction to Differentiable Manifolds and Riemannian Geometry. 2nd ed.* Academic Press, San Diego, CA.

4. Cutler, C. D. (1986) The Hausdorff Dimension Distribution of Finite Measures in Euclidean Space. *Canadian Journal of Mathematics*, **38**, 6, 1459-1484.

5. Cutler, C. D. and Dawson, D. A. (1989) Estimation of Dimension for Spatially Distributed Data and Related Limit Theorems. *Journal of Multivariate Analysis*, **28**, 115-148.

6. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W. (1992) Surface Reconstruction from Unorganized Points. *Computer Graphics*, **26**, 2, 71-78.

7. Noble, B. and Daniel, J. W. (1977) *Applied Linear Algebra. 2nd ed.* Prentice-Hall, Englewood Cliffs, NJ.

8. Waternaux, C. M. (1975) *Asymptotic Distributions of the Characteristic Roots and Vectors of the Sample Covariance Matrix in the Nonnormal Case.* Ph.D. Thesis, Yale University.

9. Waternaux, C. M. (1976) Asymptotic Distribution of the Sample Roots for a Nonnormal Population. *Biometrika*, **63**, 3, 639-645.

10. Wilkinson, J. H. (1965) *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford.