COMPARING PRINCIPAL COMPONENT ANALYSIS

WITH MAXIMUM LIKELIHOOD IN ORDINATION

Henry S. Lynn

New England Research Institutes

9 Galen Street

Watertown, MA 02172, U.S.A.

and

Charles E. McCulloch

Biometrics Unit and Statistics Center

338 Warren Hall

Cornell University, Ithaca, NY 14853, U.S.A.

ABSTRACT

Ordination methods are techniques often used in ecology and other fields to provide a simple arrangement or structure to a multivariate sites by species (or more generally observations by variables) data matrix. Though such techniques are widely used for descriptive statistics, researchers often want to additionally use them to make inferences. Our goal is to elucidate the basic performance of such methods under reasonable models for the ecological context. Maximum likelihood ordination operates by simultaneously estimating the species parameters and latent site parameters in a model approximating the relationship between species and sites. Principal component analysis is shown to be equivalent to maximum likelihood ordination for a linear functional model with independent Gaussian error terms and specific identifiability constraints. The model is extended to allow for interaction between species using an equicorrelated error structure. Conditions for consistency of the estimators from maximum likelihood ordination and principal component analysis are established and related to results on linear functional measurement error models. Finite sample performances are then examined via simulations.

Keywords: principal component analysis; maximum likelihood ordination; functional measurement error models; consistency.

1.   INTRODUCTION

Suppose a researcher samples n different sites and for each site records the abundances (e.g. biomass, percentage, frequency) of m different species of flora/fauna. The resulting samples-by-species data matrix can be examined via ordination. On the one hand, this involves generating low-dimensional summaries of the data, and principal component analysis (PCA) has a long history as such a method (Goodall 1954, Orlóci 1966, Anderson 1971, Gauch 1982). On the other hand, models can be postulated and the species and site parameters estimated using maximum likelihood (ML) or least squares (Gauch *et al.* 1974, Ihm and van Groenewood 1984). Ter Braak and Prentice (1988) pointed out that "PCA provides the ML-solution to the multidimensional linear ordination problem if the errors are independent and normally distributed with constant variance across species and sites". However, these references do not elaborate on inferential aspects and omit discussion of theoretical properties of the estimators. Furthermore, the proposed models assume species act independently of each other, when in fact correlation is expected due to symbiosis or competition effects or it can be induced through site

1

effects. This paper addresses the assumptions for doing inference using PCA. In section 2, we propose a linear functional model that incorporates interaction between species using an equicorrelated covariance structure where equal correlation exists between any two species within a site. This particular correlation structure may not truly reflect how species interact in reality but it serves as an initial approximation, and allows for tractable asymptotic results. Sections 3 and 4 discuss the existence of ML estimators and show how estimates can be computed. PCA is introduced in section 5 and equated with ML estimation for our functional model. Conditions for consistent estimators of the species parameters using both ML and PCA are established in section 6. Section 7 examines the asymptotic results under finite sample sizes, while section 8 uses two small datasets to illustrate how estimators can be constructed. Finally, section 9 pinpoints some similarities and distinctions between ML and PCA, and suggests a possible alternative approach to the ordination problem.

2. MODEL

We assume the response (possibly transformed) for each species to be a linear function of some latent environmental variable with Gaussian additive error. The latent variable can represent a hypothetical gradient, or it can be some environmental characteristic that is too expensive to be measured or can no longer be retrieved such as in studies of fossil records. Formally, let y_{ij} be the abundance of the jth species at site i, x_i the latent environmental variable at site i, a_j and b_j the intercept and slope for species j respectively, and c the baseline species slope, such that

$$y_{ij} = a_j + cx_ib_j + e_{ij} , (i=1,...,n;\ j=1,...,m) , \qquad (1)$$

where $e_i=(e_{i1},...,e_{im})' \sim$ iid $N(\mathbf{0},\mathbf{V})$, $\mathbf{0}$ is a $m \times 1$ vector of zeros, $\mathbf{V}=\sigma_e^2(1-\rho)\mathbf{I} + \rho\sigma_e^2\mathbf{J}$, $\mathbf{I}$ is a $m \times m$ identity matrix, $\mathbf{J}$ is a $m \times m$ matrix of ones, σ_e^2 is the error variance, and ρ is the correlation between species. The x_i's, b_j's, and c are treated as unknown parameters, whereas the a_j's and $\mathbf{V}$ are assumed given.

2

The introduction of the baseline slope c may seem artificial and letting a_j and $\mathbf{V}$ be known may be restrictive, but this is to illustrate how PCA is specifically related to ML estimation in model (1). The objective is to recover the underlying linear model by estimating the b_j's. However, the parameters c, x_i and b_j are only identified as a product, and thus identifiability constraints must be imposed to obtain unique estimates; e.g.

$$\sum_j b_j^2 = k_1 \text{ and } \sum_i x_i^2 = k_2 \text{ for nonzero constants } k_1 \text{ and } k_2.$$

With such a normalization, b_j is identified but only *up to a change in sign*.

Model (1) belongs to the class of linear functional measurement error models since it can be rewritten as

$$y_{i1} = a_1 + \beta'(\mathbf{v}_i - \mathbf{a}^{(1)}) + e_{i1},$$

$$\mathbf{y}_i^{(1)} = \mathbf{v}_i + \mathbf{e}_i^{(1)},$$

where $\mathbf{y}_i^{(1)} = (y_{i2}, \ldots, y_{im})'$, $\mathbf{a}^{(1)} = (a_2, \ldots, a_m)'$, $\mathbf{e}_i^{(1)} = (e_{i2}, \ldots, e_{im})'$, $\beta' = \dfrac{b_1}{m-1}\left(\dfrac{1}{b_2}, \ldots, \dfrac{1}{b_m}\right)$, and

$\mathbf{v}_i = (a_2 + cx_i b_2, \ldots, a_m + cx_i b_m)'$ is the latent variable which is only observed as $\mathbf{y}_i^{(1)}$ because of measurement error. In another context, model (1) is the factor analysis model where the cb_j's are factor loadings and $\mathbf{x} = (x_i, \ldots, x_n)'$ is an unobservable factor but with fixed instead of random components. Both the functional measurement error model and the factor analysis model have been studied extensively especially in the econometrics and psychometrics literature (Gleser 1981; Anderson 1984; Anderson and Ameniya 1988). However most of these results are not readily accessible or directly usable in ordination, because of the different parameterization used in measurement error models and the independent error structure typically assumed in factor analysis models. In the following sections, we provide results that relate to PCA and ML ordination according to model (1).

3. EXISTENCE OF MLE

It is well known that the existence of maximum likelihood estimators (MLEs) for functional measurement error models hinges on the structure of the covariance matrix. If $\mathbf{V}$ were unknown and totally arbitrary, the likelihood would be unbounded and no

maximum is attainable (Anderson and Rubin 1956; Kendall and Stuart 1979; Gleser 1981). The log-likelihood for model (1) is

$$\log L = -\frac{mn}{2}\log(2\pi) - \frac{n}{2}\log|\mathbf{V}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i - cx_i\mathbf{b})'\mathbf{V}^{-1}(\mathbf{z}_i - cx_i\mathbf{b}), \tag{2}$$

where $\mathbf{b}=(b_1,...,b_m)'$, $\mathbf{z}_i=(z_{i1},...,z_{im})'$, and we write $z_{ij}=y_{ij}-a_j$ for notation convenience. Now,

$$|\mathbf{V}| = \sigma_e^{2m}|(1-\rho)\mathbf{I} + \rho\mathbf{J}| = \sigma_e^{2m}(1-\rho)^{m-1}(1 + (m-1)\rho),$$

and

$$\mathbf{V}^{-1} = \frac{1}{\sigma_e^2(1-\rho)}(\mathbf{I} - \frac{\rho}{1-\rho+m\rho}\mathbf{J}).$$

Therefore logL equals

$$-\frac{mn}{2}\log(2\pi) - \frac{n}{2}(m\log\sigma_e^2 + (m-1)\log(1-\rho) + \log(1 + (m-1)\rho)) -$$

$$\frac{1}{2\sigma_e^2(1-\rho)}\sum_{i=1}^{n}tr((\mathbf{I} - \frac{\rho}{1-\rho+m\rho}\mathbf{J})(\mathbf{z}_i - cx_i\mathbf{b})(\mathbf{z}_i - cx_i\mathbf{b})').$$

We first argue that the sum in the above expression must be positive. It takes its minimum value (as a function of ρ) when $\rho=1$, in which case it is equal to sum of squared deviations of the elements of $\mathbf{z}_i - cx_i\mathbf{b}$. By the identifiability constraints this sum must therefore be positive. Next, as a function σ_e^2 for fixed ρ, it takes its unique maximum in the interior of $(0,\infty)$. Finally, that profile likelihood as a function of ρ also takes a unique, finite maximum on the interior of its range, $(\frac{-1}{m-1},1)$. The likelihood is thus bounded away from positive infinity. Hence, for an equicorrelated $\mathbf{V}$ the existence of MLEs in (1) is guaranteed, although for other covariance structures this must be checked individually. In ML ordination, the focus is on estimating the mean structure and likewise we assume $\mathbf{V}$ to be given. This avoids the issue of nonexistence of MLEs and will serve to illustrate the connection between ML estimation and PCA.

4. MLE AS EIGENVECTOR SOLUTION

With $\mathbf{V}$ known, the likelihood equations for the parameters c, $\mathbf{x}$, and $\mathbf{b}$ are respectively

$$c(\mathbf{b'V^{-1}b})(\mathbf{x'x})=\mathbf{x'ZV^{-1}b}, \tag{3a}$$

$$c\mathbf{x}(\mathbf{b'V^{-1}b})=\mathbf{ZV^{-1}b}, \text{ and} \tag{3b}$$

4

$$c\mathbf{b}(\mathbf{x'x})=\mathbf{Z'x}, \tag{3c}$$

where $\mathbf{Z}=\{z_{ij}\}$ is the $n \times m$ matrix with entries $\{y_{ij}-a_j\}$. Equation (3a) can be derived from either (3b) or (3c) and hence is redundant. According to (3b) and (3c), the MLE of x_i is obtained by a weighted regression of the rows of $\mathbf{Z}$ on $\mathbf{b}$, whereas the MLE of b_j is obtained by regressing the columns of $\mathbf{Z}$ on $\mathbf{x}$. This suggests that instead of jointly maximizing the likelihood (2), one can find the ML estimates by iterating between the individual likelihood equations (3b) and (3c) until convergence. This algorithm is an example of the Gauss-Seidel-Newton (GSN) method (Ortega and Rheinboldt 1970; Thisted 1988). Equations (3b) and (3c) can also be combined to give

$$\mathbf{Z'ZV^{-1}b}=c^2\mathbf{b} \text{ and} \tag{4}$$

$$\mathbf{ZV^{-1}Z'x}=c^2\mathbf{x}, \tag{5}$$

using the normalizations $\mathbf{b'V^{-1}b}=1$ and $\mathbf{x'x}=1$ as identifiability constraints. This shows that c^2 is an eigenvalue, and $\mathbf{x}$ and $\mathbf{b}$ are eigenvectors. ML estimates can then be computed using standard algorithms for calculating eigenvectors without having to perform joint maximization via Newton-Raphson. In fact, the following proposition holds. (The proofs of all the propositions and corollaries stated in the text are given in the Appendix.)

Proposition 1. Let model (1) hold. The MLEs of $\mathbf{x}$ and $\mathbf{b}$ are then the dominant eigenvectors (i.e. those associated with the largest eigenvalue) of (4) and (5).

5. PRINCIPAL COMPONENT ANALYSIS

Introduced to ecologists by Goodall (1954), PCA provides an orthogonal least squares approximation to the data via a singular value decomposition (Greenacre 1984). The popular method is species-centered PCA (Orlóci 1966), where the mean of each species has been subtracted from the columns of the data matrix. Formally, if $\mathbf{W} \equiv y_{ij} - \bar{y}_{.j}$ has rank r, the singular value decomposition of $\mathbf{W}$ gives

$$\mathbf{W} = \sum_{k=1}^{r} \lambda_k \mathbf{p}_k \mathbf{q}_k' \text{ , such that} \tag{6}$$

$$\lambda_k \mathbf{q}_k = \mathbf{W}'\mathbf{p}_k \text{ and } \lambda_k \mathbf{p}_k = \mathbf{W}\mathbf{q}_k , \tag{7}$$

$$\mathbf{W}\mathbf{W}'\mathbf{p}_k = \lambda_k^2 \mathbf{p}_k , \tag{8}$$

$$\mathbf{W}'\mathbf{W}\mathbf{q}_k = \lambda_k^2 \mathbf{q}_k , \tag{9}$$

$\mathbf{p}_k'\mathbf{p}_l = \mathbf{q}_k'\mathbf{q}_l = \delta_{kl}$, where δ_{kl} is Kronecker's delta.

This definition encompasses the usual formulation of PCA, which finds orthonormal vectors $\mathbf{q}_k$ that maximizes the norm of $\mathbf{W}\mathbf{q}_k$, yielding (9) and with the λ_k^2's being the maxima attained for each $\mathbf{q}_k$. Since the columns of $\mathbf{W}$ sum to zero, (9) also shows that $\mathbf{q}_k$ is an eigenvector of the sample covariance matrix.

Ecologists typically use PCA as a data reduction technique, with the output summarized in a biplot (Gabriel 1971). The first two terms in (6) are selected and the coordinates $(\lambda_1 p_{1i}, \lambda_2 p_{2i})$ $(i=1,...,n)$ and (q_{1j}, q_{2j}) $(j=1,...,m)$ or their rescaled versions are then plotted together for a visual appraisal of the relationship between species and sites. This application is mainly descriptive and does not explicitly require a model for the data. However, PCA can also be viewed as multivariate linear regression with latent predictors. Choose a rank one approximation to $\mathbf{W}$ by selecting only the first term in (6), then

$$\lambda \mathbf{q} = \mathbf{W}'\mathbf{p} \text{ and } \lambda \mathbf{p} = \mathbf{W}\mathbf{q}, \text{ yielding} \tag{10}$$

$$\mathbf{W}'\mathbf{W}\mathbf{q} = \lambda^2 \mathbf{q} , \text{ and} \tag{11}$$

$$\mathbf{W}\mathbf{W}'\mathbf{p} = \lambda^2 \mathbf{p}. \tag{12}$$

Observe that (11) and (12) have the same form as the likelihood equations (4) and (5) but with $\mathbf{V}=\mathbf{I}$. Now since

$$w_{ij} \equiv y_{ij} - \bar{y}_{\cdot j} = c(x_i - \bar{x})b_j + (e_{ij} - \bar{e}_{\cdot j}), \tag{13}$$

and $(e_{i1} - \bar{e}_{\cdot 1}, ..., e_{im} - \bar{e}_{\cdot m})' \sim$ iid $N(\mathbf{0}, (1-n^{-1})\mathbf{V})$, the likelihood for (13) has the same form as that for model (1). Therefore, solving for λ, p_i, and q_j in (10) is equivalent to finding the MLEs of c, $x_i - \bar{x}$, and b_j respectively in (1) but assuming $\rho=0$. It should also be noted from this definition of PCA that the species intercepts and covariance matrix cannot be estimated, which is why the a_j's and $\mathbf{V}$ are treated as known in order to establish the equivalence of PCA with ML estimation.

6. CONSISTENCY

The previous section showed that PCA is equivalent to ML estimation of the species and sites parameters in (1) when $\mathbf{V}=\mathbf{I}$, but the number of site parameters increases with the sample size n. This leads to the infinite incidental parameters problem (Neyman and Scott 1948), and causes the ML approach to give inconsistent estimators unless some 'population' assumption is imposed on the x_i's.

6.1 Consistency of the MLE

Proposition 2. Let model (1) hold. If $\lim_{n\to\infty}\dfrac{\Sigma x_i^2}{n}=h$, where $h\in(0,\infty)$, and λ^*_{max}, the maximum eigenvalue of $\mathbf{Z'ZV}^{-1}$ is simple, then the MLE of $\mathbf{b}$ consistently estimates $s\mathbf{b}$ for some nonzero constant s.

The condition that $\lim_{n\to\infty}\dfrac{\Sigma x_i^2}{n}$ be positive and finite is in fact the univariate version of Gleser's (1981) assumption C for establishing consistency for estimators in errors in variables regression models. Anderson and Taylor (1975) also required a similar condition for proving strong consistency of the least squares estimator when the number of independent variables increases with sample size. Intuitively, this means that we regard x_i as coming from some population with finite variance, which is reminiscent of Kiefer and Wolfowitz's (1956) approach.

6.2 Consistency of the PCA estimator

For the PCA estimator, i.e. the dominant eigenvector of $\mathbf{W'W}$, additional conditions on the eigenvalues are required for consistency.

Proposition 3. Let model (1) hold and let $\hat{\mathbf{b}}$ be the PCA estimator, i.e., the eigenvector associated with eigenvalue λ_{max}. Suppose $\lim_{n\to\infty}\dfrac{\Sigma x_i^2}{n}$ is positive and finite, and λ_{max} is simple, then the following are true:

i) If $\mathbf{b'1}=0$ and $m\rho\sigma_e^2 < c^2\sigma_x^2(\mathbf{b'b})$, where $\mathbf{1}$ is the $m\times1$ unit vector and $\sigma_x^2 \equiv \dfrac{\Sigma(x_i - \bar{x})^2}{n}$,

then $\hat{\mathbf{b}}$ consistently estimates $s\mathbf{b}$ for some nonzero constant s.

ii) If $\mathbf{b'1}=0$ and $m\rho\sigma_e^2 > c^2\sigma_x^2(\mathbf{b'b})$, then $\hat{\mathbf{b}}$ converges to $s\mathbf{1}$ for some nonzero constant s and is inconsistent.

iii) $\gamma\hat{\mathbf{b}} - \rho\sigma_e^2(\mathbf{1'}\hat{\mathbf{b}})\mathbf{1}$, where $\gamma = \lambda_{\max}\hat{\mathbf{b}}'\hat{\mathbf{b}} - \sigma_e^2(1 - \rho)$, consistently estimates $s\mathbf{b}$ for some nonzero constant s.

In ordination, sometimes only the ordering of the species is important. It therefore suffices to estimate $\mathbf{b}$ up to a location and scale change, and the PCA estimator consistently estimates the ordering of the species under the following conditions.

Corollary. Let model (1) hold. If $\lim\limits_{n\to\infty}\dfrac{\Sigma x_i^2}{n}$ is positive and finite, and $\lambda_{\max}$ is simple, then the PCA estimator consistently estimates $(s_1\mathbf{1} + s_2\mathbf{b})$ for nonzero constants $\{s_1,s_2\}$.

Proposition 3 and its corollary are in fact special cases of a more general result which shows what particular structure the covariance matrix needs to possess in order for the PCA estimator to be consistent.

Proposition 4. Define $w_{ij} = c(x_i - \bar{x})b_j + \varepsilon_{ij}$, where $\varepsilon_i \sim$ iid $N(\mathbf{0},\mathbf{V}_\varepsilon)$. When $\lim\limits_{n\to\infty}\dfrac{\Sigma x_i^2}{n}$ is positive and finite, and $\lambda_{\max}$ of $\mathbf{W'W}$ is simple, the following conditions hold:

i) If $\mathbf{V}_\varepsilon\mathbf{b} = s\mathbf{b}$ for some nonzero constant s, and $c^2\sigma_x^2(\mathbf{b'b}) + s$ is the dominant eigenvalue of $c^2\sigma_x^2\mathbf{bb'} + \mathbf{V}_\varepsilon$, the asymptotic matrix of $\mathbf{W'W}$, then the PCA estimator is consistent up to a scale change.

ii) If the equation

$$(c^2\sigma_x^2\mathbf{bb'} + \mathbf{V}_\varepsilon)(s_1\mathbf{1} + s_2\mathbf{b}) = \lambda(s_1\mathbf{1} + s_2\mathbf{b}).$$

has a nonzero solution set $\{s_1,s_2\}$ and λ is the dominant eigenvalue, then the PCA estimator is consistent up to a location and scale change.

8

So far we have only dealt with a special kind of covariance matrix which may not hold true in real examples. In order to establish consistency under a general covariance matrix, an adjusted PCA estimator can be derived using an estimate of the covariance matrix. Unfortunately such an estimate needs to be obtained from either a separate study or a confirmatory sample.

Proposition 5. Let $\hat{\mathbf{V}}_\varepsilon$ be an estimator of $\mathbf{V}_\varepsilon$. When $\lim\limits_{n\to\infty}\dfrac{\Sigma x_i^2}{n}$ is positive and finite, and $\lambda_{\max}$ of $\mathbf{W'W}\hat{\mathbf{V}}_\varepsilon^{-1}$ is simple, the dominant eigenvector of $\mathbf{W'W}\hat{\mathbf{V}}_\varepsilon^{-1}$ consistently estimates $\mathbf{b}$ up to a scale change.

The previous results on consistency are derived assuming that the number of species m remains fixed but the number of sites n increases to infinity. This scenario seems appropriate in most ecological studies, but in other applications where m and n can both increase to infinity, consistency can be attained without requiring any condition on the x_i's (Haberman 1977; Portnoy 1988). Nevertheless, the question still remains as to how large m and n need to be before accurate estimates can be attained.

7.   SIMULATIONS

We compared the finite sample performance of the PCA estimator of $\mathbf{b}$ (i.e. the MLE assuming independence) with the MLE assuming equicorrelated species. The latter represents the ideal (albeit unrealistic) case when the covariance structure is known. Data were generated according to model (1), with x_i simulated from a random uniform (0,1) generator. (Without loss of generality, $\mathbf{x}$ was normalized to have zero mean.) The number of species, m, was fixed at 5, and 4 levels of ρ (.8,.2,–.1,–.2) and 2 levels of n (30,200) were chosen. Each simulation used 1000 replicates. Tables 1a and 1b contrast how the behavior of the estimates of $\mathbf{b}$ depend on the value of $\sum x_i^2$.

In Table 1a with $\sum x_i^2$ set equal to n, the PCA estimates behaved according to Proposition 3. When $m\rho\sigma_e^2 > c^2\sigma_x^2\sum b_j^2$ the estimates converged to a constant vector in

9

accordance with the inconsistency result (part (ii)). When $c^2\sigma_x^2\sum b_j^2$ was larger they converged to a **b**, and this will be the case whenever $\rho<0$. (Note that absolute values of the estimates were taken since estimates are only determined up to a sign-change.) The mean square errors for the ML estimates were smaller than those for the PCA estimates when $\rho>0$, but were larger when $\rho<0$. This suggests that naively assuming independence may have its advantages under special circumstances.

In Table 1b, $\sum x_i^2$ was fixed at 4 and $\sum b_j^2$ was increased such that the variability in the responses was comparable to that in Table 1a. As expected, the mean square errors of both the PCA and the ML estimates no longer diminish with sample size, in agreement with the inconsistency results. However, the behavior of the PCA estimates are still predicted by the magnitudes of $m\rho\sigma_e^2$ and $c^2\sigma_x^2\sum b_j^2$. When $\rho>0$, $c^2\sigma_x^2\sum b_j^2$ can be less than $m\rho\sigma_e^2$ for large sample sizes, and the PCA estimates converged to a constant vector. As in Table 1a, the mean square errors of the PCA estimates were larger if $\rho>0$ but were smaller if $\rho<0$.

For the general case when the species slope parameters do not have mean zero, we compared the corrected PCA estimates constructed using Proposition 3(iii) with the ML estimates. ML performed better in terms of bias and mean square error, but the advantage diminished with weaker correlations and larger sample sizes (Table 1c). This is to be expected since PCA corresponds to ML estimation under independence, and both methods are asymptotically equivalent under the conditions of Propositions 2 and 3.

Table 2 compares PCA and ML in terms of a summary statistic which is location and scale invariant, and thus without having to specify the identifiability constraints. The particular statistic chosen was

$$\frac{1}{(\hat{b}_{[1]}-\hat{b}_{[m]})^2}\sum_{j=1}^{m}(\hat{b}_j-\hat{b}_m)^2\,,$$

where $\hat{b}_{[1]}$ and $\hat{b}_{[m]}$ are respectively the minimum and maximum elements of the estimate of **b**. The biases of the PCA estimates approached zero with increasing sample size regardless of the correlation. The standard deviations of the ML estimates were smaller,

10

although the differences were minimal except for large correlations. This suggests that PCA may be reasonable for ordering the species slopes under model (1), especially for the case where ρ and σ_e^2 are unknown and we expect the variability of the ML estimates to increase due to their estimation.

8.   EXAMPLES

We illustrate the construction of the estimator proposed in Proposition 3(iii) using data from the national records for men track races in 55 countries (Belcham and Hymans 1984), which was analyzed by Dawkins (1989) using PCA and the biplot. For our purposes, only the completion times (in min./km) of the 1.5 km, 5 km, 10 km, and marathon races are analyzed. The assumption of fixed site/country effects seems tenable since the countries may be of interests in themselves, and we hypothesize that the completion times vary linearly with a latent athletic ability variable. Furthermore, we assume an equicorrelated error structure among these four long-distances races, with an estimated correlation of 0.93 and variance of 0.11. This was arrived at by using 100 bootstrap samples of size 20 taken from the original 55 countries.

$$\mathbf{W'W} = \begin{pmatrix} 0.58 & 0.84 & 0.95 & 1.06 \\ & 1.39 & 1.52 & 1.76 \\ & & 1.76 & 2.01 \\ & & & 2.58 \end{pmatrix}, \text{ and}$$

the PCA estimator is given by $\hat{\mathbf{b}}' = (.290\ .470\ .534\ .640)$ with $\lambda_{max} = .112$. The rates of change for the four races are then given by $(.160\ .141\ .135\ .124)$. This gives an ordination in agreement with the distances of the races and the spacings suggest that the 5 km and 10 km races are most similar whereas the 1.5 km race is more distinct from the other three races.

Our second example is taken from an agricultural experiment of 40 plots designed to test whether intercropping diminishes weed productivity (Mohler and Liebman 1987). The authors performed a PCA on the log transformed and standardized weed composition

11

and concluded that there is a latent gradient which runs from barley to intercrop to pea to unplanted controls with the least productivity in barley plots and the most in the unplanted controls. We split the 40 plots into two random samples, one of which was used to estimate the covariance matrix and the other was used for the PCA. The estimated correlation matrix of the 3 most abundant weed species, *Brassica kaber*, *Ambrosia artemisiifolia* and *Agropyron repens*, was

$$
\hat{\mathbf{V}}_\varepsilon = \begin{pmatrix} 1.0 & -.11 & .56 \\ & 1.0 & -.47 \\ & & 1.0 \end{pmatrix}, \text{ and } \mathbf{W'W} = \begin{pmatrix} 1.0 & -.12 & .44 \\ & 1.0 & -.40 \\ & & 1.0 \end{pmatrix}.
$$

The predicted relative growth rates of the three weed species are then obtained by the dominant eigenvector of $\mathbf{W'W}\hat{\mathbf{V}}_\varepsilon^{-1}$ according to Proposition 5, which gives an ordination of the three species of $(-.357 \ .236 \ .379)'$.

9. SUMMARY AND DISCUSSION

ML ordination seeks to uncover the underlying relationship between species and environmental data by alternately estimating the species and site parameters for a specific model. We formally equated ML ordination with PCA in model (1) for independent errors. However, several points should be highlighted. Firstly, ML allows for the estimation of the species intercepts a_j and certain covariance structures, but PCA does not. Secondly, the sole purpose of including the baseline slope c in (1) was to show its relation to the dominant eigenvalue in PCA. In fact, ML estimation is simpler in (1) without c, and one of the identifiability constraints can also be removed. Thirdly, ML estimation and likewise PCA may not even be desirable because of the incidental parameters problem which can lead to inconsistent estimators. Lastly, estimates of $\mathbf{b}$ are not unique because of its identifiability constraint, which implies that asymptotic results can at best establish consistency up to a scale change unless the identifiability constraint is known a priori.

For an equicorrelated error structure, the PCA estimator was shown to be consistent up to a scale change when $\Sigma b_j = 0$ and $c^2 \sigma_x^2 \Sigma b_j^2 > m\rho\sigma_e^2$, but consistent up to a location and

scale change without the conditions. For a general covariance matrix, conditions for the PCA estimator to be consistent were derived, and when an external estimate of the covariance matrix is available, the adjusted PCA estimator is consistent up to a scale change. For finite samples, simulations indicated that the mean square errors of the PCA estimates were larger than those of the ML estimates for positive correlations but can be smaller for negative correlations. Therefore, using PCA may be advantageous when there is reason to believe that species share a negative equicorrelated covariance structure. When the results were summarized in terms of a location and scale invariant statistic, the ML estimates tended to have slightly smaller mean square errors, although in general the performances of PCA and ML were similar. This suggests that PCA may be robust for estimating the ranks of the species slopes in a linear model with equicorrelated errors. Together, these asymptotic and finite sample results strengthen the appeal of PCA for the linear model.

The proofs of consistency depend on the 'population' assumption of the site parameters, which is a consequence of treating the x_i's as fixed effects. This has been the traditional approach in ML ordination, yet in cases where the sites are a random sample of ecological environments it may be reasonable to treat the x_i's as random effects and thereby remove the problem of incidental parameters. The social science literature contains many such examples, and their methods may possibly be adapted for use in ordination. Likewise, the results here also carry over to other disciplines where the response varies linearly with some latent fixed effect. Finally, although the linear model may not be realistic in certain situations it serves as an approximation over small ranges of the environmental variable for nonlinear models. Research has also begun on the properties of estimators from PCA and correspondence analysis in unimodal models.

ACKNOWLEDGEMENT

Table 1a. Absolute values and mean square errors of ML and PCA estimates.

$\sum x_i^2 = n$, $\sum b_j^2 = 4$, $\sum b_j = 0$, $\sigma_e^2 = 2$, $c=1$, $a_j = 0$, and $\mathbf{b'} = (-0.853 \quad -0.689 \quad -0.361 \quad 0.295 \quad 1.607)$

ρ	n	ML $|\hat{b}_j|$	ML mse$(\hat{b}_j)$	PCA $|\hat{b}_j|$	PCA mse$(\hat{b}_j)$	$c^2 \sigma_x^2 \sum b_j^2$	$m\rho\sigma_e^2$
0.8	30	0.817	0.065	0.854	0.065		
		0.660	0.060	0.850	0.072		
		0.368	0.049	0.850	0.261	4	8
		0.321	0.042	0.847	0.323		
		1.545	0.061	0.884	0.698		
	200	0.851	0.010	0.891	0.010		
		0.689	0.010	0.892	0.048		
		0.362	0.010	0.892	0.284	4	8
		0.291	0.011	0.888	0.353		
		1.592	0.009	0.883	0.552		
0.2	30	0.822	0.073	0.796	0.123		
		0.646	0.078	0.648	0.114		
		0.386	0.062	0.435	0.093	4	2
		0.318	0.048	0.388	0.084		
		1.534	0.051	1.440	0.164		
	200	0.850	0.012	0.844	0.028		
		0.690	0.013	0.686	0.027		
		0.361	0.013	0.363	0.026	4	2
		0.299	0.013	0.299	0.025		
		1.589	0.007	1.570	0.024		
−0.1	30	0.816	0.080	0.818	0.076		
		0.654	0.087	0.654	0.083		
		0.392	0.065	0.389	0.062	4	−1
		0.349	0.062	0.345	0.059		
		1.520	0.043	1.526	0.037		
	200	0.844	0.012	0.845	0.011		
		0.679	0.013	0.680	0.012		
		0.361	0.013	0.362	0.013	4	−1
		0.296	0.014	0.296	0.014		
		1.597	0.005	1.598	0.004		
−0.2	30	0.793	0.085	0.794	0.081		
		0.662	0.089	0.667	0.086		
		0.384	0.065	0.382	0.064	4	−2
		0.352	0.060	0.346	0.056		
		1.529	0.038	1.533	0.034		
	200	0.845	0.013	0.845	0.013		
		0.689	0.016	0.689	0.015		
		0.353	0.015	0.353	0.015	4	−2
		0.292	0.015	0.292	0.014		
		1.593	0.004	1.594	0.004		

Table 1b. Absolute values and mean square errors of ML and PCA estimates.

$\sum x_i^2=4$, $\sum b_j^2=100$, $\sum b_j=0$, $\sigma_e^2=2$, $c=1$, $a_j=0$, and $\mathbf{b'}=(-4.263\ \ -3.443\ \ -1.804\ \ 1.476\ \ 8.034)$

ρ	n	ML $|\hat{b}_j|$	ML $mse(\hat{b}_j)$	PCA $|\hat{b}_j|$	PCA $mse(\hat{b}_j)$	$c^2\sigma_x^2\sum b_j^2$	$mp\sigma_e^2$
0.8	30	4.208	0.453	4.064	2.033		
		3.407	0.493	3.314	1.967		
		0.771	0.482	1.928	1.431	13	8
		1.466	0.448	1.683	1.232		
		7.944	0.427	7.544	2.729		
	200	4.220	0.547	4.469	0.106		
		3.414	0.528	4.468	1.098		
		1.808	0.525	4.466	7.118	2	8
		1.469	0.536	4.469	8.989		
		7.905	0.512	4.450	13.016		
0.2	30	4.232	0.468	4.227	0.546		
		3.425	0.525	3.421	0.604		
		1.798	0.509	1.801	0.571	13	2
		1.432	0.487	1.434	0.538		
		7.925	0.290	7.907	0.374		
	200	4.149	0.786	3.849	4.160		
		3.361	0.810	3.546	3.459		
		1.832	0.772	3.034	3.993	2	2
		1.492	0.727	2.983	4.609		
		7.898	0.471	5.828	13.162		
−0.1	30	4.221	0.467	4.222	0.457		
		3.446	0.503	3.447	0.493		
		1.779	0.533	1.778	0.526	13	−1
		1.479	0.534	1.479	0.527		
		7.920	0.193	7.923	0.183		
	200	4.207	0.824	4.213	0.748		
		3.357	0.922	3.362	0.840		
		1.784	0.856	1.780	0.818	2	−1
		1.512	0.736	1.503	0.704		
		7.868	0.358	7.885	0.272		
−0.2	30	4.232	0.463	4.233	0.458		
		3.401	0.498	3.402	0.490		
		1.784	0.559	1.785	0.552	13	−2
		1.489	0.522	1.489	0.513		
		7.934	0.147	7.935	0.140		
	200	4.157	0.819	4.163	0.776		
		3.373	0.956	3.372	0.902		
		1.823	0.922	1.822	0.900	2	−2
		1.566	0.852	1.556	0.847		
		7.858	0.325	7.869	0.273		

Table 1c. Absolute biases and mean square errors of ML and adjusted PCA estimates.

$\sum x_i^2 = n$, $\sum b_j^2 = 4$, $\sigma_e^2 = 1$, $c = 1$, $a_j = 0$, and b'=(0.156 0.467 0.778 1.090 1.401)

ρ	n	ML $|\text{bias}(\hat{b}_j)|$	ML $\text{mse}(\hat{b}_j)$	PCA $|\text{bias}(\hat{b}_j)|$	PCA $\text{mse}(\hat{b}_j)$
0.8	30	.037	.018	.133	.084
		.021	.021	.040	.039
		.019	.011	.065	.041
		.012	.007	.044	.023
		.000	.008	.004	.017
	300	.004	.003	.005	.006
		.002	.002	.008	.004
		.002	.001	.003	.002
		.001	.000	.001	.001
		.000	.001	.001	.002
0.4	30	.056	.024	.071	.034
		.016	.031	.032	.039
		.012	.029	.028	.037
		.019	.020	.025	.025
		.023	.017	.018	.020
	300	.003	.004	.005	.005
		.004	.003	.007	.004
		.002	.002	.004	.003
		.002	.002	.001	.002
		.000	.002	.001	.002
0.2	30	.044	.021	.044	.022
		.006	.037	.012	.038
		.021	.032	.024	.033
		.020	.026	.021	.027
		.024	.020	.020	.020
	300	.001	.004	.002	.004
		.000	.003	.001	.004
		.003	.003	.002	.003
		.002	.003	.002	.003
		.002	.002	.002	.002
−0.2	30	.039	.021	.048	.024
		.001	.034	.002	.036
		.014	.033	.013	.034
		.022	.030	.024	.033
		.029	.020	.033	.022
	300	.000	.004	.000	.004
		.002	.004	.003	.004
		.001	.003	.002	.004
		.003	.003	.003	.003
		.002	.002	.002	.002

Table 2. **Absolute biases and standard deviations of ML and PCA estimates of a location and scale invariant statistic.**

$\sum x_i^2 = n$, $\sigma_e^2 = 2$, $c = 1$, $a_j = 0$, and $\mathbf{b}' = (0.054\ 0.108\ 0.217\ 0.433\ 0.866)$.

		ML		PCA	
ρ	n	|Bias|	Std	|Bias|	Std
0.8	10	.086	.266	.154	.356
	20	.035	.213	.074	.257
	40	.012	.160	.031	.197
	80	.001	.128	.009	.153
	160	.000	.090	.002	.113
0.2	10	.310	.438	.317	.442
	20	.199	.357	.202	.356
	40	.078	.272	.082	.272
	80	.044	.217	.044	.222
	160	.019	.173	.021	.174
−0.1	10	.391	.475	.396	.485
	20	.220	.403	.218	.402
	40	.118	.317	.119	.317
	80	.064	.249	.065	.252
	160	.020	.186	.021	.187
−0.2	10	.354	.456	.370	.487
	20	.212	.373	.231	.380
	40	.109	.307	.123	.309
	80	.062	.251	.070	.256
	160	.029	.190	.034	.192

REFERENCES

Anderson, A. J. B. (1971), "Ordination methods in ecology," *Journal of Ecology*, 59, 713-726.

Anderson, T. W. (1984), "Estimating linear statistical relationships," *The Annals of statistics*, 5, 1-45.

Anderson, T. W. and Ameniya, Y. (1988), "The asymptotic normal distribution of estimators in factor analysis under general conditions," *The Annals of statistics*, 16, 759-771.

Anderson, T. W. and Rubin, H. (1956), "Statistical inference in factor analysis," *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, 5, 111-150.

Anderson, T. W. and Taylor, J. B. (1976), "Strong consistency of least squares estimates in normal linear regression," *The Annals of statistics*, 4, 788-790.

Belcham, P. and Hymans, R., eds. (1994), *IAAF/ATFS Track and Field Statistics Handbook for 1994 Los Angeles Olympic Games*, London: International Amateur Athletic Foundation.

Dawkins, B. (1989), "Multivariate analysis of national track records," *The American Statistician*, 43, 110-115.

Gabriel, K. R. (1971), "The biplot graphic display of matrices with application to principal components analysis," *Biometrika*, 58, 453-467.

Gauch, H. G., Chase, G. B. and Whittaker, R. H. (1974), "Ordinations of vegetation samples by Gaussian species distributions," *Ecology*, 55, 1382-1390.

Gauch, H. G. (1982), *Multivariate Analysis in Community Ecology*, Cambridge, England: Cambridge University Press.

Gleser, L. J. (1981), "Estimation in a multivariate 'errors-in-variables' regression model, Large sample results," *The Annals of Statistics*, 9, 24-44.

Goodall, D. W. (1954), "Objective methods for the classification of vegetation. III. A essay in the use of factor analysis," *Australian Journal of Botany*, 1, 39-63.

Greenacre, M. J. (1984), *Theory and Application of Correspondnece Analysis*, London: Academic Press.

Haberman, S. J. (1977), "Maximum likelihood estimation in exponential response models," *The Annals of Statistics*, 5, 815-841.

Ihm, P. and van Groenewoud, H. (1984), "Correspondence analysis and Gaussian ordination," *COMPSTAT lectures*, 3, 5-60.

Kendall, M. G. and Stuart, A. (1979), *The Advanced Theory of Statistics*, vol. 2, 4th edition, New York: Macmillan.

Kiefer, J. and Wolfowitz, J. (1956), "Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters," *Annals of Mathematical Statistics*, 27, 887-906.

Mohler, C. L. and Liebman, M. (1987), "Weed productivity and composition in sole crops and intercrops of barley and field pea," *Journal of Applied Ecology*, 24, 685-699.

Neyman, J. and Scott, E. L. (1948), "Consistent estimators based on partially consistent observations," *Econometrica*, 16, 1-32.

Orlóci, L. (1966), "Geometric models in ecology. I. The theory and application of some ordination methods," *Journal of Ecology*, 54, 193-215.

Ortega, J. and Rheinboldt, W. (1970), *Iterative solution of nonlinear equations in several variables*, New York: Academic Press.

Ortega, J. M. (1972), *Numerical Analysis. A Second Course*, New York: Academic Press.

Portnoy, S. (1988), "Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity," *The Annals of Statistics*, 16, 356-366.

Ter Braak, C. J. F. and Prentice, I. C. (1988), "A theory of gradient analysis," *Advances of Ecological Research*, 18, 271-317.

Thisted, R. A. (1988), *Elements of Statistical Computing*, London: Chapman and Hall.

Proof of Proposition 1. From (2) the part of the log-likelihood containing the unknown parameters is proportional to

$$\sum_{i=1}^{n} tr((\mathbf{z}_i - c\mathbf{x}_i\mathbf{b})'\mathbf{V}^{-1}(\mathbf{z}_i - c\mathbf{x}_i\mathbf{b}))$$

$$= \sum_{i=1}^{n} tr(\mathbf{V}^{-1}(\mathbf{z}_i - c\mathbf{x}_i\mathbf{b})(\mathbf{z}_i - c\mathbf{x}_i\mathbf{b})') = tr(\mathbf{V}^{-1}(\mathbf{Z} - c\mathbf{xb}')'(\mathbf{Z} - c\mathbf{xb}'))$$

$$= tr(\mathbf{V}^{-1}\mathbf{Z}'\mathbf{Z}) + tr(c^2\mathbf{b}'\mathbf{V}^{-1}\mathbf{bx}'\mathbf{x}) - 2tr(c\mathbf{V}^{-1}\mathbf{bx}'\mathbf{Z}).$$

Assume without loss of generality that $\mathbf{b}'\mathbf{V}^{-1}\mathbf{b}=1$ and $\mathbf{x}'\mathbf{x}=1$, then $tr(c^2\mathbf{b}'\mathbf{V}^{-1}\mathbf{bx}'\mathbf{x})=c^2$ and from (3b)

$$2tr(c\mathbf{V}^{-1}\mathbf{bx}'\mathbf{Z})=2ctr(\mathbf{Z}\mathbf{V}^{-1}\mathbf{bx}')=2ctr(c\mathbf{xx}')=2c^2.$$

Therefore, the log-likelihood attains its maximum when c^2 is the largest, which implies that the dominant eigenvectors of (4) and (5) are the MLEs.

In order to establish Proposition 2 we make use of the following lemma.

Lemma 1 (Ortega 1972). An eigenvalue is a continuous function of the matrix elements. If it is also simple (i.e. has multiplicity one) then its associated eigenvector is a continuous function of the matrix elements.

Proof of Proposition 2. Recall that the MLE is the dominant eigenvector of (4). Under model (1), the (j,k)th element $(j,k=1,\dots,m)$ of $\mathbf{Z}'\mathbf{Z}$ is

$$\sum_i z_{ij}z_{ik} = \sum_i (cb_j x_i + e_{ij})(cb_k x_i + e_{ik})$$

$$= c^2 b_j b_k \Sigma x_i^2 + cb_j \Sigma x_i e_{ij} + cb_k \Sigma x_i e_{ik} + \Sigma e_{ij}e_{ik}.$$

According to the iid Weak Law of Large Numbers,

$$\frac{1}{n}\sum_i e_{ij}e_{ik} \rightarrow \rho\sigma_e^2 \text{ for } j\neq k, \text{ and } \frac{1}{n}\sum_i e_{ik}^2 \rightarrow \sigma_e^2.$$

Furthermore, when $\lim_{n\to\infty}\frac{1}{n}\sum_i x_i^2 = h$, for $h\in(0,\infty)$,

$$\lim_{n\to\infty}\frac{1}{n^2} Var(\sum_i x_i e_{ik}) = \lim_{n\to\infty}\frac{1}{n^2}\sum_i x_i^2 \sigma_e^2 = \lim_{n\to\infty}\frac{h\sigma_e^2}{n} = 0,$$

and thus $\lim\limits_{n\to\infty} \dfrac{\Sigma x_i e_{ik}}{n} \to 0$. Define $\mu_{2x} \equiv \dfrac{\Sigma x_i^2}{n}$. It follows that

$$n^{-1}\mathbf{Z'ZV^{-1}} \to \Sigma, \text{ where } \Sigma = (\sigma_e^2(1-\rho)\mathbf{I} + \rho\sigma_e^2\mathbf{J} + c^2\mu_{2x}\mathbf{bb'})\mathbf{V^{-1}}.$$

Observe that

$$\Sigma = (\sigma_e^2(1-\rho)\mathbf{I} + \rho\sigma_e^2\mathbf{J} + c^2\mu_{2x}\mathbf{bb'})\frac{1}{\sigma_e^2(1-\rho)}(\mathbf{I} - \frac{\rho}{1-\rho+m\rho}\mathbf{J})$$

$$= (\mathbf{I} + k\mathbf{bb'} - \frac{k(\mathbf{b'1})\rho}{(1-\rho+m\rho)}\mathbf{b1'}), \text{ where } k = \frac{c^2\mu_{2x}}{\sigma_e^2(1-\rho)}.$$

The eigenvectors of Σ are equivalently those defined by the equation

$$(k\mathbf{bb'} - \frac{k(\mathbf{b'1})\rho}{(1-\rho+m\rho)}\mathbf{b1'})\mathbf{u} = (\lambda-1)\mathbf{u},$$

where λ is an eigenvalue of Σ and $\mathbf{u}$ an eigenvector. The matrix in parentheses on the left hand side of the equation has rank 1, and it is easily verified that $\mathbf{b}$ is an eigenvector with eigenvalue $k(\mathbf{b'b} - \dfrac{\rho(\mathbf{b' 1})^2}{1-\rho+m\rho})$. The other $m-1$ eigenvalues all equal zero. This implies that Σ has eigenvalues

$$1 + k(\mathbf{b'b} - \frac{\rho(\mathbf{b' 1})^2}{1-\rho+m\rho}), \text{ and } 1 \text{ with multiplicity } m-1.$$

Furthermore, since

$$k(\mathbf{b'b} - \frac{\rho(\mathbf{b' 1})^2}{1-(m-1)\rho}) > 0 \text{ within the range of } \rho,$$

$\mathbf{b}$ is the dominant eigenvector. Let $\hat{\mathbf{b}}_{mle}$ be the dominant eigenvector of $\mathbf{Z'ZV^{-1}}$ with eigenvalue λ_{max}^*. It follows by Lemma 1 that since $n^{-1}\mathbf{Z'ZV^{-1}} \to \Sigma$, $n^{-1}\lambda_{max}^*$ converges to the largest eigenvalue of Σ, and $n^{-1}\hat{\mathbf{b}}_{mle}$ converges to its dominant eigenvector; which is $s\mathbf{b}$ for some nonzero constant s.

Proof of Proposition 3. Under model (13), the (j,k)th element of $\mathbf{W'W}$ is

$$\sum_i w_{ij}w_{ik} = c^2 b_j b_k \Sigma(x_i^2 - \bar{x}) + cb_j\Sigma(x_i - \bar{x})\varepsilon_{ik} + cb_k\Sigma(x_i - \bar{x})\varepsilon_{ij} + \Sigma\varepsilon_{ij}\varepsilon_{ik},$$

where $\varepsilon_{ij} = e_{ij} - \bar{e}_{.j}$. Now

$$\frac{1}{n}\sum_i \varepsilon_{ij}\varepsilon_{ik} \to \rho\sigma_e^2 \text{ for } j \neq k, \text{ and } \frac{1}{n}\sum_i \varepsilon_{ik}^2 \to \sigma_e^2.$$

When $\lim\limits_{n\to\infty}\dfrac{\Sigma x_i^2}{n}\in(0,\infty)$, $\bar{x}$ is finite and therefore $\lim\limits_{n\to\infty}\dfrac{1}{n}\sum\limits_i(x_i-\bar{x})\varepsilon_{ik}\to0$. With $\sigma_x^2\equiv\dfrac{\Sigma(x_i-\bar{x})^2}{n}$, it follows that

$$n^{-1}\mathbf{W'W}\to\Sigma,\ \text{where}\ \Sigma=\sigma_e^2(1-\rho)\mathbf{I}+\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'}.$$

The eigenvectors of Σ are equivalently those defined by

$$(\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'})\mathbf{u}=(\lambda-\sigma_e^2(1-\rho))\mathbf{u}.$$

i) The matrix $\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'}$ has rank 2. When $\mathbf{b'1}=0$, it has eigenvectors $\mathbf{1}$ and $\mathbf{b}$ with eigenvalues $\rho\sigma_e^2(\mathbf{1'1})$ and $c^2\sigma_x^2\mathbf{b'b}$ respectively. The other $m-2$ eigenvalues all equal zero. This implies that Σ has eigenvalues

$$m\rho\sigma_e^2+(1-\rho)\sigma_e^2,\ c^2\sigma_x^2(\mathbf{b'b})+(1-\rho)\sigma_e^2,\ \text{and}\ (1-\rho)\sigma_e^2\ \text{with multiplicity}\ m-2.$$

It follows by Lemma 1 that when $c^2\sigma_x^2(\mathbf{b'b})>m\rho\sigma_e^2$, $n^{-1}\hat{\mathbf{b}}$ converges to $s\mathbf{b}$ for some nonzero constant s.

ii) Otherwise, $\mathbf{1}$ is the dominant eigenvector and $n^{-1}\hat{\mathbf{b}}$ converges to $s\mathbf{1}$ for some nonzero constant s.

iii) Using a spectral decomposition on Σ, we have

$$\sigma_e^2(1-\rho)\mathbf{I}+\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'}=\sum_{k=1}^{m}\lambda_k\mathbf{u}_k\mathbf{u}_k'.$$

Since the $\mathbf{u}_k$'s are orthogonal to each other, it follows algebraically that

$$(\lambda_1\mathbf{u}_1'\mathbf{u}_1-\sigma_e^2(1-\rho))\mathbf{u}_1-\rho\sigma_e^2(\mathbf{1'u}_1)\mathbf{1}=c^2\sigma_x^2(\mathbf{b'u}_1)\mathbf{b}.$$

By Lemma 1, $n^{-1}\hat{\mathbf{b}}$ converges to $\mathbf{u}_1$ and the result follows.

Proof of Corollary. According to Proposition 3, $n^{-1}\hat{\mathbf{b}}$ converges to the dominant eigenvector of Σ, which is also the dominant eigenvector of $\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'}$. Now since

$$(\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'})\mathbf{u}=\rho\sigma_e^2(\mathbf{1'u})\mathbf{1}+c^2\sigma_x^2(\mathbf{b'u})\mathbf{b},$$

all nontrivial eigenvectors of $\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'}$ are of the form $(s_1\mathbf{1}+s_2\mathbf{b})$ for nonzero constants s_1 and s_2. Observe that

$$(\rho\sigma_e^2\mathbf{J}+c^2\sigma_x^2\mathbf{bb'})(s_1\mathbf{1}+s_2\mathbf{b})=\rho\sigma_e^2(ms_1+s_2(\mathbf{b'1}))\mathbf{1}+c^2\sigma_x^2(ms_1+s_2(\mathbf{b'b}))\mathbf{b}.$$

This implies that $(s_1\mathbf{1}+s_2\mathbf{b})$ is an eigenvector whenever

$$\rho\sigma_e^2(ms_1 + s_2(\mathbf{b'1})) = \lambda s_1 \text{ and } c^2\sigma_x^2(ms_1 + s_2(\mathbf{b'b})) = \lambda s_2,$$

which solving for λ gives

$$mc^2\sigma_x^2 s_1^2 + s_1 s_2(m\rho\sigma_e^2 - c^2\sigma_x^2(\mathbf{b'b})) + \rho\sigma_e^2(\mathbf{b'1})s_2^2 = 0.$$

This conic equation always has a nonzero solution set $\{s_1, s_2\}$, and thus the PCA estimator is consistent up to a location and scale change.

Proof of Proposition 4. Under the conditions of Proposition 4 and following similar arguments as in Proposition 3,

$$n^{-1}\mathbf{W'W} \rightarrow c^2\sigma_x^2\mathbf{bb'} + \mathbf{V}_\varepsilon.$$

Observe that if $\mathbf{u}$ is an eigenvector of $c^2\sigma_x^2\mathbf{bb'} + \mathbf{V}_\varepsilon$, then $\mathbf{u}$ is proportional to

$$c^2\sigma_x^2(\mathbf{b'u})\mathbf{b} + \mathbf{V}_\varepsilon\mathbf{u}.$$

Conditions (i) and (ii) then follow by applying Lemma 1.

Proof of Proposition 5. Under the conditions of Proposition 5,

$$n^{-1}\mathbf{W'W}\hat{\mathbf{V}}_\varepsilon^{-1} \rightarrow \Sigma_\varepsilon \equiv c^2\sigma_x^2\mathbf{bb'V}_\varepsilon^{-1} + \mathbf{V}_\varepsilon\mathbf{V}_\varepsilon^{-1}.$$

The rest of the proof is similar to that for Proposition 2. Observe first that Σ_ε and $\mathbf{bb'V}_\varepsilon^{-1}$ share the same eigenvectors. Now $\mathbf{bb'V}_\varepsilon^{-1}$ has rank 1 and $\mathbf{b}$ is its eigenvector with eigenvalue $c^2\sigma_x^2(\mathbf{b'V}_\varepsilon^{-1}\mathbf{b})$. It follows that Σ_ε has eigenvalues

$$c^2\sigma_x^2(\mathbf{b'V}_\varepsilon^{-1}\mathbf{b})+1, \text{ and } 1 \text{ with multiplicity } m-1,$$

with $\mathbf{b}$ as the dominant eigenvector since $c^2\sigma_x^2(\mathbf{b'V}_\varepsilon^{-1}\mathbf{b})$ is always positive. Therefore, by Lemma 1, the dominant eigenvector of $\mathbf{W'W}\hat{\mathbf{V}}_\varepsilon^{-1}$ consistently estimates $\mathbf{b}$ up to a scale change.

24