# A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data

Jie Peng* & Debashis Paul

Department of Statistics, University of California, Davis, CA 95616

* Correspondence author: email: jie@wald.ucdavis.edu

**Abstract**

In this paper, we consider the problem of estimating the eigenvalues and eigenfunctions of the covariance kernel (i.e., the *functional principal components*) from sparse and irregularly observed longitudinal data. We approach this problem through a maximum likelihood method assuming that the covariance kernel is smooth and finite dimensional. We exploit the smoothness of the eigenfunctions to reduce dimensionality by restricting them to a lower dimensional space of smooth functions. The estimation scheme is developed based on a Newton-Raphson procedure using the fact that the basis coefficients representing the eigenfunctions lie on a *Stiefel manifold*. We also address the selection of the right number of basis functions, as well as that of the dimension of the covariance kernel by a second order approximation to the leave-one-curve-out cross-validation score that is computationally very efficient. The effectiveness of our procedure is demonstrated by simulation studies and an application to a CD4 counts data set. In the simulation studies, our method performs

well on both estimation and model selection. It also outperforms two existing approaches: one based on a local polynomial smoothing of the empirical covariances, and another using an EM algorithm.

**Keywords :** longitudinal data, covariance kernel, functional principal components, Stiefel manifold, Newton-Raphson algorithm, cross-validation.

# 1 Introduction

In recent years there have been numerous works on data that may be considered as noisy curves. When the individual observations can be regarded as measurements on an interval, the data thus obtained can be classified as functional data. For analysis of data arising in various fields, such as longitudinal data analysis, chemometrics, econometrics, etc. [Ferraty and Vieu (2006)], the functional data analysis viewpoint is becoming increasingly popular. Depending on how the individual curves are measured, one can think of two different scenarios - (i) when the individual curves are measured on a dense grid; and (ii) when the measurements are observed on an irregular, and typically sparse set of points on an interval. The first situation usually arises when the data are recorded by some automated instrument, e.g. in chemometrics, where the curves represent the spectra of certain chemical substances. The second scenario is more typical in longitudinal studies where the individual curves could represent the level of concentration of some substance, and the measurements on the subjects may be taken only at irregular time points.

In these settings, when the goal of analysis is either data compression, model building or studying covariate effects, one may want to extract information about the mean, variability, correlation structure, etc. In the first scenario, i.e., data on a regular grid, as long as the individual curves are smooth, the measurement noise level is low,

and the grid is dense enough, one can essentially treat the data to be on a continuum, and employ techniques similar to the ones used in classical multivariate analysis. However, the irregular nature of data in the second scenario, and the associated measurement noise require a different treatment.

The main goal of this paper is the estimation of the functional principal components from sparse, irregularly, observed functional data (scenario (ii)). The eigenfunctions give a nice basis for representing functional data, and hence are very useful in problems related to model building and prediction for functional data [see e.g. Cardot, Ferraty and Sarda (1999), Hall and Horowitz (2007), Cai and Hall (2006)]. Ramsay and Silverman (2005) and Ferraty and Vieu (2006) give an extensive survey of the applications of *functional principal components analysis* (FPCA).

The focus throughout this paper thus is in the estimation of covariance kernel of the underlying process. Covariance is a positive semidefinite operator. The space of covariance operators is a nonlinear manifold. Thus, from statistical as well as aesthetic point of view, it is important that any estimator of the covariance is also positive semidefinite. Moreover, Smith (2005) gives a compelling argument in favor of utilizing the intrinsic geometry of the parameter space in the context of estimating covariance matrix in a multivariate Gaussian setting. He obtains Cramér-Rao bounds for the risk, that are described in terms of intrinsic gradient and Hessian of the log-likelihood function. This work brings out important features of the estimators that are not obtained through the usual Euclidean viewpoint. It also provides a strong motivation for a likelihood-based approach that respects the intrinsic geometry of the parameter space. In this paper, we shall adopt a *restricted maximum likelihood* approach and explicitly utilize the intrinsic geometry of the parameter space when fitting the maximum likelihood estimator.

Now we shall give an outline of the model for the sparse functional data. Suppose

that we observe $n$ independent realizations of an $L^2$-stochastic process $\{X(t) : t \in [0,1]\}$ at a sequence of points on the interval $[0,1]$ (or, more generally, on an interval $[a,b]$), with additive measurement noise. That is, the observed data $\{Y_{ij} : 1 \leq j \leq m_i; 1 \leq i \leq n\}$ can be modeled as :

$$Y_{ij} = X_i(T_{ij}) + \sigma\varepsilon_{ij}, \tag{1}$$

where $\{\varepsilon_{ij}\}$ are i.i.d. with mean 0 and variance 1. Since $X(t)$ is an $L^2$ stochastic process, by Mercer's theorem [cf. Ash (1972)] there exists a positive semi-definite kernel $C(\cdot,\cdot)$ such that $Cov(X(s),X(t)) = C(s,t)$ and each $X_i(t)$ has the following a.s. representation in terms of the eigenfunctions of the kernel $C(\cdot,\cdot)$ :

$$X_i(t) = \mu(t) + \sum_{\nu=1}^{\infty} \sqrt{\lambda_\nu}\psi_\nu(t)\xi_{i\nu}, \tag{2}$$

where $\mu(\cdot) = \mathbb{E}(X(\cdot))$ is the mean function; $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ are the eigenvalues of $C(\cdot,\cdot)$; $\psi_\nu(\cdot)$ are the corresponding orthonormal eigenfunctions; and the random variables $\{\xi_{i\nu} : \nu \geq 1\}$, for each $i$, are uncorrelated with zero mean and unit variance. In the observed data model (1), we assume that $\{T_{ij} : j = 1,\ldots,m_i\}$ are randomly sampled from a continuous distribution. In the problems we shall be interested in, the number of measurements $m_i$ is typically small.

Our estimation procedure is based on the assumption that the covariance kernel $C$ is of finite rank, say $r$; and the representation of the eigenfunctions $\{\psi_\nu\}_{\nu=1}^r$ in a known, finite, basis of smooth functions. This results in an orthogonality constraint on the matrix of basis coefficients, say $B$, as described in Section 2. Specifically, the matrix $B$ lies in a *Stiefel manifold*, that is the space of real valued matrices with orthonormal columns. Our estimation procedure involves maximization of the log-likelihood under the working assumption of normality, satisfying the orthonormality

constraint on $B$. To implement this, we employ a Newton-Raphson algorithm based on the work by Edelman, Arias and Smith (1998) for optimization on a *Stiefel manifold*, that utilizes its intrinsic Riemannian geometric structure. As a remark, the procedure we proposed here is intrinsically nonlinear. Linear estimation procedures for covariance, that assume the basis representation framework, have been studied by various authors including Cardot (2000), Rice and Wu (2001), and Besse, Cardot and Ferraty (1997).

At this point, we would like to mention the main contributions of this paper. The approach based on utilizing the intrinsic geometry of the parameter space in the context of covariance estimation using a maximum likelihood approach is new. The resulting estimation procedure can handle different regimes of sparsity of data efficiently. Its implementation is computationally challenging in the current context because of the irregular nature of the measurements. The resulting estimator is very accurate, based on the simulation studies we conducted. The geometric viewpoint has further important implications. Selection of an appropriate model in the context of longitudinal data is of great importance, and devising a computationally practical, yet effective, model selection procedure remains a challenge [cf. Marron *et al.* (2004), p. 620]. It is well-known that the computational cost of the usual leave-one-curve-out cross-validation score (for selecting the dimension of the basis used in the representation) is prohibitive. We utilize the geometry of the parameter space to derive an approximation of the CV score that is computationally very efficient. This is another main contribution of this paper. Finally, our approach involves analysis of the gradient and Hessian of the log-likelihood. A detailed asymptotic analysis of any estimation procedure based on the likelihood necessarily involves understanding of these objects and the geometry of the parameter space. The work presented here serves as a first step in this direction.

Before ending this section, we give a brief overview of the existing literature on FPCA. The idea of maximizing the restricted likelihood is in the same framework as that studied by James, Hastie and Sugar (2000), who propose an EM algorithm to maximize the log-likelihood. However, there are important differences between the proposed approach and the EM approach. First, the EM algorithm results in an estimator not necessarily satisfying the orthonormality constraints, that is, being outside the parameter space, which is corrected through an eigen-decomposition. But nevertheless, this can lead to an inefficient estimator. Secondly, the EM algorithm does not set the intrinsic gradient of the log-likelihood to zero. Therefore it does not utilize the redundancy in the optimization problem induced by the orthonormality constraints. This could also result in a loss of efficiency in estimation since the value of the objective function may stabilize even though the optimal parameter value in the restricted space may not have been attained. On the other hand, our approach addresses the problem of finding the optimal parameter value more directly. Moreover, the approximate CV score proposed in Section 4 rely heavily on the gradient of the objective function being zero at the estimator, which is not satisfied by the EM algorithm, but is one property of our proposed estimator.

We already mentioned the basis representation approach. Another approach to FPCA is through kernel smoothing. In this approach, the $i$-th curve is pre-smoothed by taking weighted average of $\{Y_{ij}\}_{j=1}^{m_i}$'s where the weights are evaluations of a kernel centered at the time points $\{T_{ij}\}_{j=1}^{m_i}$. Unfortunately, when the number of measurements is small, this procedure results in a highly biased estimate of the covariance kernel as demonstrated by Yao, Müller and Wang (2005). These authors thus propose to estimate the covariance by local polynomial smoothing of the empirical covariances at observed pairs of time points $\{(T_{ij}, T_{ij'}) : i = 1, \cdots, n, 1 \leq j \neq j' \leq m_i\}$. Hall, Müller and Wang (2006) prove optimality of this procedure under rather weak assumptions

on the process for optimal choice of bandwidths. Their work clearly separates the problem of estimating the covariance and its eigenfunctions, and identifies the latter as a one dimensional nonparametric function estimation problem. In spite of its nice asymptotic properties, there are some aspects of the local polynomial method that are somewhat unsatisfactory. First, it does not ensure a positive semi-definite estimate of the population covariance kernel. A common practice to fix that is through projection, however this can lead to an inefficient estimator. Secondly, this procedure sometimes results in a negative estimate of the error variance $\sigma^2$. In contrast, the proposed procedure gives positive semi-definite estimate as well as positive estimate of $\sigma^2$.

The novelty of our work is the explicit utilization of the intrinsic geometry of the parameter space, which results in more efficient estimators. Moreover, this enables an efficient approximation of the cross validation score. As far as we know, an efficient cross validation based model selection procedure has not been discovered for most of the existing procedures in this field, including the two approaches we mentioned above. Simulation studies presented in Section 5 indicate a significant improvement of the proposed method over both EM (James *et al.* (2000)) and local polynomial (Yao *et al.*(2005)) approaches, as well as a satisfactory performance in model selection based on the approximate CV score derived in Section 4. We also want to emphasize that, the estimation procedure presented in this paper should be regarded as a demonstration of the usefulness of the geometric viewpoint while tackling a complex statistical problem. Even though our focus throughout this paper remains on solving the problem described above, the tools developed here can be easily extended to other situations that involve matrix-valued parameters with orthonormality constraints. Two such examples are discussed in Section 7.

The rest of the paper is organized as follows. In Section 2, we describe the re-

stricted maximum likelihood framework. In Section 3, we give an outline of the Newton-Raphson algorithm for optimization on Stiefel manifolds. In Section 4, we derive an approximation to the leave-one-curve-out cross-validation score. Section 5 is devoted to detailed simulation studies and the comparison and discussion of the performance of various procedures. In Section 6, the proposed procedure is illustrated through an application to a CD4 counts data set. In Section 7, we discuss some possible extensions and future works. Technical details are given in the appendices. Tables, Figures and supplementary material are attached at the end.

# 2 Restricted MLE framework

We first describe the basis representation framework. Under some weak conditions on the stochastic processes (like $L^2$-differentiability of certain order, see, e.g. Ash (1972)), the eigenfunctions have some degree of smoothness. This assumption has been used in various studies, including Boente and Fraiman (2000), Cardot (2000), James *et al.* (2000), Yao *et al.* (2005, 2006), and Hall *et al.* (2006). Smoothness of the eigenfunctions means that they can be well approximated in some stable basis for smooth function classes, e.g. the B-spline basis [Chui (1987)]. If in addition, in model (2), we assume that $\lambda_\nu = 0$ for $\nu > r$, for some $r \geq 1$, then we can choose a finite set of linearly independent, $L^2$ functions $\{\phi_1(\cdot), \ldots, \phi_M(\cdot)\}$ with $M \geq r$, such that eigenfunctions can be modeled as $\psi_\nu(\cdot) = \sum_{k=1}^{M} b_{k\nu}\phi_k(\cdot)$ for $\nu = 1, \ldots, r$. Then, for every $t$,

$$\boldsymbol{\psi}(t)^T := (\psi_1(t), \ldots, \psi_r(t)) = (\phi_1(t), \ldots, \phi_M(t))B \tag{3}$$

for an $M \times r$ matrix $B = ((b_{k\nu}))$ that satisfies the constraint

$$B^T(\int \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^T dt)B = \int \boldsymbol{\psi}(t)\boldsymbol{\psi}(t)^T dt = I_r, \tag{4}$$

where $\boldsymbol{\phi}(\cdot) = (\phi_1(\cdot), \ldots, \phi_M(\cdot))^T$. Since the $M \times M$ matrix $\int \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^T dt$ is known and nonsingular, without loss of generality, hereafter we assume $B^T B = I_r$, by orthonormalizing $\{\phi_1(\cdot), \ldots, \phi_M(\cdot)\}$.

Here, we are assuming a reduced rank model for the covariance kernel as in James *et al.* (2000). This model can be motivated as follows. Suppose that the covariance kernel $C(s,t)$ of the underlying process has the infinite Karhunen-Loéve expansion:

$$C(s,t) = \sum_{k=1}^{\infty} \lambda_k \psi_k(s)\psi_k(t), \tag{5}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\sum_{k=1}^{\infty} \lambda_k < \infty$, and $\{\psi_k\}_{k=1}^{\infty}$ forms a complete orthonormal basis for $L^2[0,1]$. The condition $\sum_{k=1}^{\infty} \lambda_k < \infty$ implies that $\lambda_k \to 0$ as $k \to \infty$. Also, the orthonormality of the eigenfunctions $\{\psi_k\}$ implies that $\psi_k$ typically gets more and more "wiggly" as $k$ increases, at least for most reasonable processes with smooth covariance kernel. Therefore, modeling the full covariance kernel remains a challenge. However, one can truncate the series on the RHS of (5) at some finite $r \geq 1$ to get the *projected covariance kernel*

$$C^r_{proj}(s,t) = \sum_{k=1}^{r} \lambda_k \psi_k(s)\psi_k(t). \tag{6}$$

Note that $\| C - C^r_{proj} \|^2_F = \sum_{k=r+1}^{\infty} \lambda_k^2$. Thus, as long as the eigenvalues decay to zero fast, even with a relatively small $r$, the approximation $C^r_{proj}$ only results in a small bias. This motivates the choice of a finite rank model as described above. Furthermore, the restriction to reduced rank model helps in modeling the eigenfunctions as well. If $\psi_k$ for larger $k$ are more wiggly, it takes a lot more basis functions to represent them well. On the other hand, for a model with $r$ relatively small, we can get good approximations to the eigenfunctions with a moderate number of basis functions.

Of course, in practice one could encounter situations for which the projected kernel

$C^r_{proj}$ is not a good approximation for any small value of $r$. This will for example happen if the eigenvalues are decaying slowly. Then in the modeling step one needs to choose large $r$. However under such situations, there is an intrinsic instability of the estimates of the eigenfunctions, as it is well known that, the estimation error grows inversely with the gap between successive eigenvalues. Moreover, it is harder to choose the sufficient rank $r$ by a model selection procedure if it is too large. Appropriate statistical methods to deal with such data still need to be developed.

Finally, it is worthwhile to point out some advantages of the reduced rank formulation over the mixed effects model by Rice and Wu (2000), as also noted by James *et al.* (2000). Notice that, in the unconstrained mixed effects model, one needs to model the covariance kernel using a full-rank representation. Thus if one uses $M$ basis functions to represent it, there are $M(M+1)/2$ basis coefficients of the covariance kernel that need to be estimated. When the observations are sparse, this could lead to an over-parametrization, and it will result in highly variable estimates. Furthermore, if one uses a maximum likelihood approach, the over-parametrization would cause a very rough likelihood surface, with multiple local maxima. Therefore, restricting the rank of the covariance kernel can also be viewed as a form of regularization of the likelihood.

If one assumes Gaussianity of the processes, i.e., $\xi_{i\nu} \overset{i.i.d.}{\sim} N(0,1)$ and $\varepsilon_{ij} \overset{i.i.d.}{\sim} N(0,1)$, and they are independent, then under the assumption (3), the negative log-likelihood of the data, conditional on $\{(m_i, \{T_{ij}\}_{j=1}^{m_i})\}_{i=1}^n$ is given by

$$\begin{aligned} -\log L(B, \Lambda, \sigma^2) &= const. + \frac{1}{2}\sum_{i=1}^{n} Tr[(\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i)^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)^T] \\ &\quad + \frac{1}{2}\sum_{i=1}^{n} \log|\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i|, \end{aligned} \tag{7}$$

where $\Lambda$ is the $r \times r$ diagonal matrix of non-zero eigenvalues of $C(\cdot,\cdot)$, $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$

and $\boldsymbol{\mu}_i = (\mu(T_{i1}), \ldots, \mu(T_{im_i}))^T$ are $m_i \times 1$ vectors, and $\Phi_i = [\boldsymbol{\phi}(T_{i1}) : \ldots : \boldsymbol{\phi}(T_{im_i})]$ is an $M \times m_i$ matrix. One can immediately see that the difficulty with the maximum likelihood approach mainly lies in the irregularity of the form of the objective function (7), and the fact that the parameter $B$ has orthonormal constraints (4). Moreover, this is a non-convex optimization problem with respect to the parameters.

We propose to directly minimize (7) subject to (4) by a Newton-Raphson algorithm on the Stiefel manifold, whose general form has been developed in Edelman *et al.* (1998). The proposed estimator is

$$(\widehat{B}, \widehat{\Lambda}, \widehat{\sigma}^2) = \arg \min_{B \in \mathcal{S}_{M,r}, (\Lambda, \sigma^2) \in \Theta} -\log L(B, \Lambda, \sigma^2),$$

where $\Theta = \mathbb{R}_+^{r+1}$, and $\mathcal{S}_{M,r} := \{A \in \mathbb{R}^{M \times r} : A^T A = I_r\}$ is the Stiefel manifold of $M \times r$ real-valued matrices (with $r \leq M$) with orthonormal columns. The Newton-Raphson procedure involves computation of the intrinsic gradient and Hessian of the objective function, and on convergence, it sets the gradient to zero. Thus the proposed estimator solves the score equation:

$$\nabla_{(B,\Lambda,\sigma^2)} \log L(B, \Lambda, \sigma^2) = 0.$$

We shall discuss the details of this algorithm and its implementation in Section 3.

It is important to note that, one does not need to assume Gaussianity in order to carry out the proposed estimation as well as model selection using the approximated CV score derived in Section 4. This is because (7) is a *bona fide* loss function. Thus the Gaussianity should be viewed as a working assumption which gives the form of the loss function. It is assumed throughout that (7) is differentiable with respect to the eigenvalues and eigenfunctions. This in turn depends on the assumption that all the nonzero eigenvalues of the covariance kernel are distinct, since multiplicity of

eigenvalues results in the covariance kernel being non-differentiable with respect to both eigenvalues and eigenfunctions. It is also worth pointing out that the M-step of the EM algorithm in James *et al.* (2000) does not utilize the orthonormality constraint on $B$. This restriction can be imposed, and the minimization of the corresponding objective function can be carried out in a similar fashion as in the proposed method. This may lead to an improvement in the performance of the EM estimates.

# 3 Newton-Raphson algorithm

In this section, we describe the Newton-Raphson algorithm for minimising the loss function (7). In a seminal paper, Edelman *et al.* (1999) derive Newton-Raphson and conjugate gradient algorithms for optimising functions on Stiefel and Grassman manifolds. As their counterparts in the Euclidean space, these algorithms aim to set the gradient of the objective function (viewed as a function on the manifold) to zero. The algorithms involve the following steps : (i) update the *tangent vector* at the current parameter value; (ii) move along the geodesic in the direction of the recently updated tangent vector to a new point on the manifold.

In our setting, the objective is to minimise the loss function (7). For notational simplicity, drop the irrelevant constants and re-write (7) as

$$F(B, \Lambda, \sigma^2) := \sum_{i=1}^{n} \left[ F_i^1(B, \Lambda, \sigma^2) + F_i^2(B, \Lambda, \sigma^2) \right], \tag{8}$$

where

$$F_i^1(B, \Lambda, \sigma^2) = Tr[P_i^{-1} \widetilde{\mathbf{Y}}_i \widetilde{\mathbf{Y}}_i^T], \;\; F_i^2(B, \Lambda, \sigma^2) = \log |P_i|, \tag{9}$$

with $P_i = \sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i$, $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \boldsymbol{\mu}_i$, and $\Phi_i$ as defined in Section 2. Here we treat $\boldsymbol{\mu}_i$ as known, since we propose to estimate it separately. The parameter spaces for $\Lambda$ and $\sigma^2$ are positive cones in Euclidean spaces and hence convex. The parameter

space for $B$ is $\mathcal{S}_{M,r}$, the Stiefel manifold of $M \times r$ matrices with orthonormal columns.

We adopt a two-step procedure for updating the parameters. Each Newton-Raphson updating step is broken into two parts - (a) an update of $(\Lambda, \sigma^2)$, keeping $B$ at the current value; and (b) an update of $B$, setting $(\Lambda, \sigma^2)$ at the recently updated value. Thus, the algorithm proceeds by starting at an initial estimate and then cycling through these two steps iteratively till convergence.

For now, assume that the orthonormal basis functions $\{\phi_k\}$ and dimensions $M$ and $r$ ($M \geq r$) are given. The choice of these objects will be discussed later. Since $\lambda_k > 0$ for all $k = 1, \ldots, r$ and $\sigma^2 > 0$, it is convenient to define $\boldsymbol{\zeta} = \log(\Lambda)$, i.e. $\zeta_k = \log \lambda_k$, and $\tau = \log \sigma^2$, and treat $F$ as a function of $\boldsymbol{\zeta}$ and $\tau$. Note that $\zeta_k$, $\tau$ can vary freely over $\mathbb{R}$. Then the Newton-Raphson step for updating $(\Lambda, \sigma^2)$ (or equivalently $(\boldsymbol{\zeta}, \tau)$) is straightforward. In the rest of the paper, we treat $\boldsymbol{\zeta}$ interchangeably as an $r \times r$ matrix and as a $1 \times r$ vector.

We then give an outline of the Newton-Raphson step for updating $B$. This involves finding the *intrinsic* gradient and Hessian of $F$, while treating $\Lambda$ and $\sigma^2$ as fixed. The key point is the fact that the gradient is a vector field acting on the tangent space of the manifold $\mathcal{S}_{M,r}$, and the Hessian is a bilinear operator acting on the same tangent space. Some facts about the Stiefel manifold, the tangent space, and its canonical metric, that are essential to describing and implementing the algorithm, are given in Appendix A. Based on the notations used there, we outline the Newton-Raphson algorithm for minimising an arbitrary function $F(B)$ where $B \in \mathcal{S}_{M,r}$. For more details, see Edelman et al. (1999). In the following, we use $\mathcal{M}$ to denote $\mathcal{S}_{M,r}$. Let $\Delta$ denote an element of the tangent space of $\mathcal{S}_{M,r}$ at the current value $B$, denoted by $\Delta \in \mathcal{T}_B\mathcal{M}$. It represents the direction in the tangent space in which a Newton-Raphson step moves from the current $B$. Let $F_B$ denote the usual Euclidean gradient, i.e., $F_B = ((\frac{\partial F}{\partial b_{kl}}))$. For any $\Delta \in \mathcal{T}_B\mathcal{M}$, $F_{BB}(\Delta)$ is defined to be the element of $\mathcal{T}_B\mathcal{M}$

satisfying

$$\langle F_{BB}(\Delta), X\rangle_c = \frac{\partial^2}{\partial s \partial t} F(B + s\Delta + tX)\,|_{s,t=0}\ , \quad \text{for all} \quad X \in \mathcal{T}_B\mathcal{M},$$

where $\langle , \rangle_c$ denotes the canonical metric on the Stiefel manifold $\mathcal{M}$. Also, let $H_F$ denote the Hessian operator acting on the tangent space $\mathcal{T}_B\mathcal{M}$.

**Outline of Newton-Raphson algorithm on $\mathcal{S}_{M,r}$:** Given $B \in \mathcal{S}_{M,r}$,

1. compute the intrinsic gradient $\nabla F\,|_B$ of $F$ at $B$, given by $\nabla F\,|_B = G := F_B - BF_B^T B$;

2. compute the tangent vector $\Delta := -H_F^{-1}(G)$ at $B$, by solving the linear system

$$F_{BB}(\Delta) - B\ \mathrm{skew}(F_B^T\Delta) - \ \mathrm{skew}(\Delta F_B^T)B - \frac{1}{2}\Pi\Delta B^T F_B = -G, \qquad (10)$$

$$B^T\Delta + \Delta^T B \ = \ 0, \qquad (11)$$

   where $\Pi = I - BB^T$, and $\mathrm{skew}(X) := (X - X^T)/2$;

3. move from $B$ in the direction $\Delta$ to $B(1)$ along the geodesics $B(t) = BM(t) + QN(t)$, where

   (i) $QR = (I - BB^T)\Delta$ is the QR-decomposition, so that $Q$ is $M \times r$ with orthonormal columns, and $R$ is $r \times r$, upper triangular;

   (ii) $A = B^T\Delta$, and

$$\begin{bmatrix} M(t) \\ N(t) \end{bmatrix} = \exp\left\{ t \begin{bmatrix} A & -R^T \\ R & 0 \end{bmatrix} \right\} \begin{bmatrix} I_r \\ 0 \end{bmatrix};$$

   Note that the matrix within exponent is a skew-symmetric matrix and so

the exponential of that can be calculated using the singular value decomposition.

4. set $B = B(1)$, and repeat until convergence. This means that the sup-norm of the gradient $G$ is less than a pre-specified tolerance level.

In the calculation of $F_B$ and $F_{BB}(\Delta)$ for $F$ defined by (8), complications associated with the inversion of $m_i \times m_i$ matrices $P_i$ arise, since $m_i$'s could vary from sample to sample. We avoid this by a suitable utilisation of matrix inversion formulae that reduce the problem to computing inverses of $r \times r$ matrices $Q_i$ instead (Appendix B). Therefore the proposed procedure can also efficiently handle the case of relatively dense measurements, where $m_i$ could be much larger than $r$. The formulae of these quantities are given in equations (24) - (28) in Appendix B. In order to update the tangent vector $\Delta$, in step 2 of the algorithm, we need to solve the system of equations given by (10) and (11). These are matrix equations and we propose to solve them via vectorisation [cf. Muirhead (1982)]. This step requires a considerable amount of computational effort since it involves the inversion of an $Mr \times Mr$ matrix.

In order to apply the Newton-Raphson algorithm, we need to choose a suitable basis for representing the eigenfunctions. In the simulation studies presented in Section 5, we have used the (orthonormalised) cubic B-spline basis, with equally spaced knots [Green and Silverman (1994), p. 157]. It is well known that B-splines provide a flexible, localised and stable basis for a wide class of smooth functions and are very easy to compute [Chui (1987), de Boor (1978)]. Different choices of basis functions are certainly possible, and can be implemented without changing the structure of the algorithm. Besides the choice of the basis, the number of basis functions $M$ and the dimension of the process $r$ need to be specified. We treat the determination of these two numbers as a model selection problem and discuss this in Section 4.

Given a basis $\{\phi_k(\cdot)\}$ and fixed $M, r$ with $M \geq r$, an initial estimate of $\Lambda$ and $B$ can

be obtained by projecting an initial estimate of the covariance kernel $\widehat{C}(\cdot,\cdot)$ onto the basis functions: $\{\phi_1(\cdot),\cdots,\phi_M(\cdot)\}$, and then performing an eigen-decomposition. In the simulation studies, the local polynomial method and the EM algorithm discussed in Section 1 are used to obtain initial estimates of the covariance kernel, as well as that of the noise variance $\sigma^2$. The dependence of the proposed method on the initial estimates is discussed in Section 5.

# 4 Approximate cross validation score

One of the key questions pertaining to nonparametric function estimation is the issue of model selection. This, in our context means selecting $r$, the number of nonzero eigenvalues, and the basis for representing the eigenfunctions. Once we have a scheme for choosing the basis, the second part of the problem boils down to selecting $M$, the number of basis functions. Various methods for dealing with this include standard criteria like AIC, BIC, multi-fold cross-validation and leave-one-curve-out cross-validation.

In this paper, we propose to choose $(M,\ r)$ based on the criterion of minimising an approximation of the leave-one-curve-out cross-validation score

$$CV := \sum_{i=1}^{n} \ell_i(\mathbf{Y}_i, \mathbf{T}_i, \widehat{\Psi}^{(-i)}), \tag{12}$$

where $\mathbf{T}_i = (T_{i1},\ldots,T_{im_i})$. Here $\Psi = (B,\tau,\boldsymbol{\zeta})$, and $\widehat{\Psi}^{(-i)}$ is the estimate of $\Psi$ based on the data excluding curve $i$. In this paper

$$\ell_i(\mathbf{Y}_i, \mathbf{T}_i, \Psi) = F_i^1 + F_i^2$$

[cf. (8) and (9)]. Note that in the Gaussian setting, $\ell_i$ is proportional to the negative

log-likelihood (up to an additive constant) of the $i$-th curve. Therefore in that setting, CV defined through (12) is the *empirical predictive Kullback-Leibler risk*. As indicated in Section 1, the computational cost to get $CV$ is prohibitive. This necessitates the use of efficient approximations. Our method of approximation, which parallels the approach taken by Burman (1990) in the context of fitting generalized additive models, is based on the following observations. The Newton-Raphson estimate $\widehat{\Psi}$, which is based on the whole data, satisfies the equation

$$\sum_{i=1}^{n} \nabla \ell_i(\mathbf{Y}_i, \mathbf{T}_i, \widehat{\Psi}) = 0. \tag{13}$$

Also, for each $i$, the corresponding estimate $\widehat{\Psi}^{(-i)}$ satisfies the equation

$$\sum_{j \neq i} \nabla \ell_j(\mathbf{Y}_j, \mathbf{T}_j, \widehat{\Psi}^{(-i)}) = 0. \tag{14}$$

Here $\{\ell_i\}_{i=1}^n$ are viewed as functions on the product space $\widetilde{\mathcal{M}} = \mathcal{M} \times \mathbb{R}^{r+1}$, where $\mathcal{M}$ is the Stiefel manifold with the canonical metric, to be denoted by $g \equiv \langle\ ,\ \rangle_c$. The parameter space $\mathbb{R}^{r+1}$ refers to $\{(\tau, \underset{r\times r}{\boldsymbol{\zeta}}) : \tau \in \mathbb{R}, \zeta_k \in \mathbb{R},\ k = 1, \ldots, r\}$, with Euclidean metric. $\nabla \ell_i$ denotes the gradient of $\ell_i$ viewed as a vector field on the product manifold.

The main idea for our approximation scheme is the observation that for each $i = 1, \ldots, n$, the "leave curve $i$ out" estimate $\widehat{\Psi}^{(-i)}$ is a perturbation of the estimate $\widehat{\Psi}$ based on the whole data. Thus, one can expand the left hand side of (14) around $\widehat{\Psi}$ to obtain an approximation of $\widehat{\Psi}^{(-i)}$. Then we shall use this approximation to get a second order approximation to the cross validation score given by (12).

We introduce some notations first. Let $\delta_\tau^i = \widehat{\tau}^{(-i)} - \widehat{\tau}$, $\boldsymbol{\delta}_\zeta^i = \widehat{\boldsymbol{\zeta}}^{(-i)} - \widehat{\boldsymbol{\zeta}}$ (a $1 \times r$ vector), and $\Delta_i = \dot{\gamma}(0) \in \mathcal{T}_{\widehat{B}}\mathcal{M}$, with $\gamma(t)$ a geodesic on $(\mathcal{M}, g)$ starting at $\gamma(0) = \widehat{B}$, and ending at $\gamma(1) = \widehat{B}^{(-i)}$. Note that, $\Delta_i$ is an element of the tangent space at $\widehat{B}$. Here-

after, we shall use $\ell_j(\widehat{\Psi})$, and $\ell_j(\widehat{\Psi}^{(-i)})$ to denote $\ell_j(\mathbf{Y}_j, \mathbf{T}_j, \widehat{\Psi})$ and $\ell_j(\mathbf{Y}_j, \mathbf{T}_j, \widehat{\Psi}^{(-i)})$, respectively, for $1 \le i, j \le n$. Let $\nabla_B \ell_i$ and $\nabla_B^2 \ell_i$ denote gradient and Hessian of $\ell_i$ with respect to $B$, and $\nabla_{(\tau,\zeta)} \ell_i$ and $\nabla^2_{(\tau,\zeta)} \ell_i$ denote gradient and Hessian of $\ell_i$ with respect to $(\tau, \boldsymbol{\zeta})$. Since the parameter $(\tau, \boldsymbol{\zeta})$ lies in an Euclidean space, $\nabla_{(\tau,\zeta)} \ell_i$ is an $(r+1) \times 1$ vector and $\nabla^2_{(\tau,\zeta)} \ell_i$ is an $(r+1) \times (r+1)$ matrix. As mentioned before, $\nabla_B \ell_i$ is a tangent vector and $\nabla_B^2 \ell_i$ is a bilinear operator on the tangent space $\mathcal{T}_B \mathcal{M}$ of the Stiefel manifold at the point $B$. The Hessian $\nabla^2 \ell_i$ with respect to $\Psi = (B, \tau, \boldsymbol{\zeta})$ can be approximated by

$$\widetilde{\nabla}^2 \ell_i = \begin{bmatrix} \nabla_B^2 \ell_i & 0 \\ 0 & \nabla^2_{(\tau,\zeta)} \ell_i \end{bmatrix},$$

by ignoring the mixed-derivative terms $\nabla_{(\tau,\zeta)}(\nabla_B \ell_i)$ and $\nabla_B(\nabla_{(\tau,\zeta)} \ell_i)$. This approximation simplifies the calculation considerably and allows us to treat the terms involving approximation of $\widehat{B}^{(-i)}$ (keeping $(\tau, \boldsymbol{\zeta})$ fixed at $(\widehat{\tau}, \widehat{\boldsymbol{\zeta}})$) and that of $(\widehat{\tau}^{(-i)}, \widehat{\boldsymbol{\zeta}}^{(-i)})$ (keeping $B$ fixed at $\widehat{B}$) separately. Thus, a second order Taylor expansion of the CV score around $\widehat{\Psi}$ becomes

$$\begin{aligned} CV &:= \sum_{i=1}^n \ell_i(\widehat{\Psi}^{(-i)}) \\ &\approx \sum_{i=1}^n \ell_i(\widehat{\Psi}) + \Big[\sum_{i=1}^n \langle \nabla_{(\tau,\zeta)} \ell_i(\widehat{\Psi}), (\delta_\tau^i, \boldsymbol{\delta}_\zeta^i)^T \rangle + \frac{1}{2} \sum_{i=1}^n \langle [\nabla^2_{(\tau,\zeta)} \ell_i(\widehat{\Psi})](\delta_\tau^i, \boldsymbol{\delta}_\zeta^i)^T, (\delta_\tau^i, \boldsymbol{\delta}_\zeta^i)^T \rangle \Big] \\ &\quad + \Big[\sum_{i=1}^n \langle \nabla_B \ell_i(\widehat{\Psi}), \Delta_i \rangle_c + \frac{1}{2} \sum_{i=1}^n \nabla_B^2 \ell_i(\widehat{\Psi})(\Delta_i, \Delta_i) \Big]. \end{aligned} \tag{15}$$

In order to get first order approximations to the second and third terms in (15), we shall use equations (13) and (14). These equations separate into two sets of equations involving the gradients $\nabla_{(\tau,\zeta)} \ell_i$ and $\nabla_B \ell_i$, respectively. The treatment of the former does not require any extra concept beyond regular matrix algebra, whereas the treatment of the latter requires Riemannian geometric concepts. However, in terms

of the final form of the approximation, both expressions are very similar. Denote the Hessian operator of $\sum_j \ell_j$ with respect to $B$ and $(\tau, \boldsymbol{\zeta})$ by $\mathbf{H}_B$ and $\mathbf{H}_{(\tau,\zeta)}$, respectively. Then our final approximation to the CV score is given by

$$\begin{aligned}
\widetilde{CV} \;:=\; & \sum_{i=1}^{n} \ell_i(\widehat{\Psi}) + \sum_{i=1}^{n} \langle \nabla_{(\tau,\zeta)} \ell_i(\widehat{\Psi}), [\mathbf{H}_{(\tau,\zeta)}(\widehat{\Psi})]^{-1} \nabla_{(\tau,\zeta)} \ell_i(\widehat{\Psi}) \rangle \\
& + \sum_{i=1}^{n} \langle \nabla_B \ell_i(\widehat{\Psi}), [\mathbf{H}_B(\widehat{\Psi})]^{-1} \nabla_B \ell_i(\widehat{\Psi}) \rangle_c \\
& + \frac{3}{2} \sum_{i=1}^{n} \langle \nabla^2_{(\tau,\zeta)} \ell_i(\widehat{\Psi}) [\mathbf{H}_{(\tau,\zeta)}(\widehat{\Psi})]^{-1} \nabla_{(\tau,\zeta)} \ell_i(\widehat{\Psi}), [\mathbf{H}_{(\tau,\zeta)}(\widehat{\Psi})]^{-1} \nabla_{(\tau,\zeta)} \ell_i(\widehat{\Psi}) \rangle \\
& + \frac{3}{2} \sum_{i=1}^{n} \nabla^2_B \ell_i(\widehat{\Psi}) ([\mathbf{H}_B(\widehat{\Psi})]^{-1} \nabla_B \ell_i(\widehat{\Psi}), [\mathbf{H}_B(\widehat{\Psi})]^{-1} \nabla_B \ell_i(\widehat{\Psi})). \qquad (16)
\end{aligned}$$

The details of this derivation are given in Appendix C. Observe that, in order to obtain the estimate $\widehat{\Psi}$ using the Newton-Raphson algorithm, we need to compute the objects $\nabla_B \ell_i$, $\nabla_{(\tau,\zeta)} \ell_i$, $\nabla^2_B \ell_i$, $\nabla^2_{(\tau,\zeta)} \ell_i$, $\mathbf{H}_B$, and $\mathbf{H}_{(\tau,\zeta)}$ at each step. Indeed, since the Newton-Raphson procedure aims to solve (13), whenever the procedure converges, we immediately have these objects evaluated at $\widehat{\Psi}$. Therefore, the additional computational cost for computing $\widetilde{CV}$ is a negligible fraction of the cost of obtaining the estimate $\widehat{\Psi}$. This provides huge computational advantage in comparison to the usual leave-one-curve-out CV score approach. We shall discuss the effectiveness of $\widetilde{CV}$ in model selection in Section 5.

**Remark :** It is worth noting that this approximation approach can be extended to other settings, for example in nonparametric regression problems. In that context, the approximation $\widetilde{CV}$ is different from the usual GCV (*generalized cross validation*) score. Indeed, in the regression setting, GCV score is obtained by performing a first order approximation to the usual leave-one-out CV score. In contrast, our method relies on a second order approximation.

# 5 Simulation

In this section, we conduct two simulation studies. The first study is focussed on the estimation accuracy of the proposed method (henceforth, `Newton`) and comparing it with two existing procedures: the local polynomial method (henceforth, `loc`) [Yao *et al.* (2005)], and the EM algorithm (henceforth, `EM`) [James *et al.* (2000)]. The second study aims to illustrate the usefulness of the model selection approach described in Section 4. All data are generated under model (1) with Gaussian principal component scores $\{\xi_{i\nu}\}$. For all settings, $\mu(t) \equiv 0$, and its estimate $\widehat{\mu}(t)$, obtained by a local linear smoothing, is subtracted from the observations before estimating the other model parameters. The number of measurements $m_i$ are i.i.d. $\sim$ uniform$\{2, \cdots, 10\}$; the measurement points for the ith subject $\{T_{ij} : j = 1, \cdots, m_i\}$ are i.i.d. $\sim$ uniform$[0, 1]$. For `Newton`, cubic $B$-splines with equally spaced knots are used as basis functions. `loc` and `EM` are used to obtain two different sets of initial estimates. The resulting estimates by `Newton` are therefore denoted by `New.loc` and `New.EM`, respectively. For `EM`, only initial value of $\sigma$ is needed. Since the result is rather robust to this choice [James *et al.* (2000)], it is set to be one. $B$-splines are used as basis functions; for some cases natural splines are also considered. To make a distinction, we use `EM.ns` to denote EM algorithm that uses natural splines, and `New.EM.ns` to denote its corresponding `Newton` method. For `loc`, bandwidths are selected by the `h.select()` function in the `R` package `sm`, with `method="cv"`. Due to the limitation of space, we only report simulation results in detail for selected cases which we believe are representative. More results are given as supplementary material (attached at the end of this paper).

In the first study, data are generated under three different settings (`easy`, `practical` and `challenging`) with 100 independent replicates for each combination of parameters. The simulation scheme is summarised in Table 1. As can be seen, different

Table 1: Simulation Settings. Shown are the parameters used in the first simulation study: nonzero eigenvalues ($\lambda_\nu$); basis for eigenfunctions ($\psi_\nu$); error variances ($\sigma^2$); error distributions ($\mathcal{L}(\varepsilon)$); sample sizes ($n$).

| name | $\lambda_\nu$ | $\psi_\nu$ | $\sigma^2$ | $\mathcal{L}(\varepsilon)$ | sample size $n$ |
|---|---|---|---|---|---|
| `easy` | $(1:3)^{-0.6}$ | $<5$ B-spline functions$>$ | 1/16, 1/8 | $N(0,1)$, $t_4$, exp(1) | 100, 200, 500 |
| `practical` | $(1:5)^{-0.6}$ | $<10$ B-spline functions$>$ | 1/16, 1/8 | $N(0,1)$, $t_4$, exp(1) | 300, 500, 1000 |
| `challenging` | $(1:3)^{-0.6}$ | $<3$ spike functions$>$ | 1/16, 1/8 | $N(0,1)$ | 300, 500, 1000 |

sample sizes, error variances and error distributions are considered. For the `easy` and `practical` cases (Figures 1 and 2, respectively), eigenfunctions are represented by the cubic B-splines with $M = 5$ and $M = 10$ equally spaces knots, respectively. For the `challenging` case (Figure 3), the eigenfunctions are represented by three "spike" functions and they can not be represented exactly by cubic B-splines.

In the first study, the true $r$ is used by all three methods. Note that, the estimation of covariance kernel by `loc` does not rely on either $M$ or $r$. For a given $r$, the first $r$ eigenfunctions and eigenvalues of the estimated covariance $\widehat{C}(\cdot,\cdot)$ (using the optimal choice of bandwidth) are used. For `Newton` and `EM`, a number of different values of $M$, including the truth, are used to fit the model. For the `challenging` case, the "true" $M$ means the $M$ resulting in least biased projection of the eigenfunctions onto the B-spline basis, which is 30. The selection of $(M, r)$ is discussed in the second study. For `Newton`, we report the number of converged replicates (cf. Section 3) for each combination of parameters and for each $M$ (Table 6). As we shall see, lack of convergence of `Newton` is primarily caused by poor initial estimates. Therefore, it is fair to compare all three methods on the converged replicates only. The performance of these three methods (based on converged replicates only) is summarised in Tables 2 to 5. For the estimation of eigenfunctions, mean integrated squared error (MISE) is used as a measure of accuracy. We also report the standard deviations of the integrated squared errors. To evaluate the estimation of eigenvalues and error variance, mean squared error (MSE) is used as the measure. Since these quantities

are of different orders of magnitude, for the ease of comparison, the MSEs are divided by the square of the corresponding true values.

As can be seen from Tables 2 to 5, the MISE/MSE corresponding to `Newton` (`New.loc`, `New.EM` and `New.EM.ns`) shows a good risk behaviour under the true $M$. The results under a larger $M$ are comparable to that under the true $M$. As expected, the performance under an inadequate $M$ is much worse, which reflects the lack of fit. To give a visual illustration, in Figures 1 to 3, we plot the point-wise average of estimated eigenfunctions by `New.EM` over all converged replicates, as well as the point-wise 0.95 and 0.05 quantiles, all under the true $M$ ($M = 30$ for the `challenging` case). As can be seen from these figures, the average is very close to the truth, meaning only small biases, except for the `challenging` case where $\psi_1$ is not estimated as accurately mainly due to the intrinsic bias in the B-spline representation. The width between two quantiles is fairly narrow meaning small variations, except for occasional large variances at the boundaries.

In comparison with `loc` and `EM`, `Newton` generally performs better in terms of MISE for eigenfunctions under an adequate $M$ ($\geq$ truth). The reduction in MISE varies from 30% to as high as 95% compared to `loc`; and 10% to around 65% compared to `EM` (except for the first eigenfunction of the `challenging` case) (Tables 2 to 5, where the reduction is always for `Newton` compared to its initial estimate). Moreover, comparison of Table 3 with Table 4 shows greater improvement by `Newton` with larger sample sizes. As is evident from the tables, there is also a big improvement of `New.loc` over `loc` in estimation of eigenvalues when $M$ is adequately large. The reduction in MSEs varies from 30% to as high as 90% with the exception for the last two eigenvalues of the `challenging` case with $n = 500, M = 30$, where only a little improvement is observed. In the `practical` case, there is also an improvement by `New.EM` over `EM`, although the reduction in MSEs is much less compared to the improvement over `loc`.

Moreover, under the `easy` and `challenging` cases, small percentages of increase in MSEs of eigenvalues by `New.EM` compared to `EM` are sometimes observed. In terms of estimating the error variance $\sigma^2$, `Newton` is much better than both `loc` and `EM` in most of cases as long as $M$ is adequate. One problem with `loc` is that, it gives highly variable, and sometimes even negative, estimate of $\sigma^2$. For example, for the easy case with $n = 200$, 56 out of 100 replicates give a negative estimate of $\sigma^2$ and for all the simulations we have done, at least around 20% replicates result in a negative estimate (see numbers reported in Tables 10-28 in the supplementary material). This fact is also reflected by the larger MSE of $\widehat{\sigma}^2$ using `New.loc` than using `New.EM`.

We observe that `New.loc` often suffers from lack of convergence. This phenomenon is more pronounced for the two higher dimensional cases: `practical` and `challenging` (Table 6). This is mainly due to the poor initial estimates by `loc`. For example, for the `practical` case with $n = 500$ and $M = 10$ (Table 3), the MISE of the first eigenfunction by `loc` is 0.434, while that by `EM` is only 0.054. However, among the converged replicates, the performance of `New.loc` is not much worse than that of `New.EM`, especially for the leading eigenfunctions. In the above case, the MISEs of the first eigenfunction by `New.loc` and `New.EM` are 0.035 and 0.036, respectively. `New.loc` does tend to give less accurate and more variable estimates for eigenfunctions corresponding to smaller eigenvalues. It is also noteworthy that, for the `practical` case, `EM.ns` does not work very well compared to `EM` at the true $M$ ($M = 10$), but its performance is much better for a larger $M$ (e.g., $M = 20$). This is because the actual eigenfunctions are represented by cubic B-splines, thus the use of a natural spline basis could result in a significant bias. However, among the converged replicates, the performance of `New.EM` and `New.EM.ns` is rather comparable. The main difference lies in the number of converged replicates. For $n = 500, M = 10$, there are 93 replicates converging under `New.EM`, but only 60 replicates converging under `New.EM.ns` (Table

6). In contrast, in the `challenging` case, the difference between `EM` and `EM.ns` is smaller, since now the biases resulting from representing the eigenfunctions in the cubic B-spline basis and that in the natural spline basis are more similar. We also study the impact of increasing the error variance, as well as different error distributions (see supplementary material: Tables 13-17, 20, 22-24 and 27 for detailed results). These simulations show that all three methods are quite robust with respect to these two aspects.

In summary, we observe satisfactory performance of `Newton` in terms of estimation, as well as improvements of `Newton` over the two alternative methods, especially over `loc`. These improvements pertain to both average and standard deviation of the measures of accuracy, and they increase with the sample size. We also want to point out that, for the Newton-Raphson algorithm, good initial estimates are important mainly for the convergence of the procedure. As long as the estimates converge, the difference in performance is not very large for the estimation of eigenfunctions and eigenvalues. We will discuss possible ways to improve convergence at the end of this section.

As mentioned in Section 1, we shall use the approximate cross validation score defined through (16) as the criterion for selecting $M$ and $r$ for the `Newton` procedure. As can be seen from Table 6, where $r$ is fixed at the true value, as long as `Newton` converges for the correct model (i.e., true $M$), it is selected almost all the time. Moreover, a model with inadequate $M$ is not selected unless it is the only model under which `Newton` converges. In order to study the selection of $M$ and $r$ simultaneously, we conduct the second simulation study, in which there are three leading eigenvalues $(1, 0.66, 0.52)$, and a fourth eigenvalue which is comparable to the error variance ($\lambda_4 = 0.07$). Additionally, there are 6 smaller eigenvalues $(9.47 \times 10^{-3}, 1.28 \times 10^{-3}, 1.74 \times 10^{-4}, 2.35 \times 10^{-5}, 3.18 \times 10^{-6}, 4.30 \times 10^{-7})$. Thus we refer $r = 4$ as the adequate

dimension. The corresponding orthonormal eigenfunctions are represented in a cubic $B$-spline basis with $M = 10$ equally spaced knots. Data are generated with sample size $n = 500$ and Gaussian noises with $\sigma^2 = 1/16$. This setting is referred as the `hybrid` case. We fit models with $M = 10, 15, 20, 25$, and $r = 2, \ldots, 7$. In this setting, our aim is to get an idea about the typical sizes (meaning $(M, r)$) of models selected. At the same time, we want to see, whenever a larger than adequate $r$ (i.e., $r = 4$) is selected, whether small eigenvalues are estimated to be small. This is important because if one uses this estimation procedure for reducing dimensionality of the data, for example by projecting the data onto the selected eigen-basis, then "spurious" components should not have large weights in that representation. Moreover, even if a larger $r$ is selected, as long as the small or zero eigenvalues are estimated to be small, the result is not going to be too misleading, in that, people can always choose a smaller model based on the *fraction of explained variation (FEV)*.

In Table 7, for both `New.EM` and `New.loc`, there is a big drop in the number of converged replicates from $r = 5$ to $r = 6$ and even bigger drop from $r = 6$ to $r = 7$. Now the lack of convergence is a reflection of a combination of poor initial estimates and larger than adequate $r$. The latter is actually a safeguard against selecting unnecessarily large models. Note that, under large $r$, the system under true parameters is going to be (nearly) singular. In the case of `New.loc`, both factors apply whenever there is lack of convergence. In the case of `New.EM`, the second factor is predominant. We find that, for `New.EM`, $M = 10$ and $r = 5$ or 6 are the preferred models by the proposed approximate CV score; however, for `New.loc`, $M = 10$ and $r = 3$ or 4 are the ones selected most often. The latter is mainly due to lack of convergence of `New.loc` for $r \geq 4$. Therefore, we will focus on the results of `New.EM` hereafter. We observe that, for models with $r = 5$ and $r = 6$, the small eigenvalues (the fifth one and/or the sixth one) are estimated to be reasonably small by `New.EM`

(data not shown). We then use the standard procedure of FEV on the selected model to further prune down the value of r: for every model $(M^*, r^*)$ selected by the CV criterion, we choose the smallest index $\overline{r}$ for which the ratio $\sum_{\nu=1}^{\overline{r}} \widehat{\lambda}_\nu / \sum_{\nu=1}^{r*} \widehat{\lambda}_\nu$ exceeds a certain threshold $\kappa$. In this study, we consider $\kappa = 0.995, 0.99, 0.95$. The results of the model selection using this additional FEV criterion are reported in Table 7 in the parentheses. As can be seen, under $\kappa = 0.995$, the most frequently selected models become $M = 10$ and $r = 4$ or $r = 5$ (the first number in the parentheses). If we set $\kappa = 0.99$, the most frequently selected models become $M = 10$ and $r = 4$ (the second number in the parentheses). Under $\kappa = 0.95$, the preferred model becomes $M = 10$ and $r = 3$. Note that, in `hybrid` case, the first three eigenvalues are dominant and compared to the error variance $\sigma^2$, the first four eigenvalues are not negligible. Therefore, the additional FEV criterion gives very reasonable model selection results. This is another indicator that in the models selected by the approximate CV criterion, the small eigenvalues indeed are estimated to be small.

In summary, the approximate CV score (16) is very effective in selecting the correct $M$– the number of basis functions needed to represent the eigenfunctions. It has the tendency to select slightly larger than necessary $r$. However, in those selected models, the `Newton` estimates of the small or zero eigenvalues are quite small. Therefore, the model selection results are not going to be very misleading and an additional FEV criterion can be applied to select a smaller model (in terms of $r$).

Finally, we want to discuss some practical aspects of the proposed method. It is noted that, the $Mr \times Mr$ linear system (10) and (11) is sometimes nearly singular, causing `Newton` to terminate without the gradient converging to zero (i.e., fail to converge). This phenomenon is most obvious for `New.loc` due to the poor initial estimates. The system becomes more stable as sample size $n$ becomes larger, as demonstrated by comparing the `practical` case with $n = 500$ to $n = 1000$ in

Table 6. Combining the `Newton` results with different initial estimates, for example `New.loc` and `New.EM` (replace one by another if the first one fails to converge), can improve convergence, and consequently the model selection results (cf. `combine` and `combine.ns` in Table 6). In addition, it is well known that the initial steps of a Newton-Raphson algorithm are typically too large [Boyd and Vandenberghe (2004)]. To avoid this, we suggest to use smaller step sizes in the beginning. That is, in the Newton-Raphson algorithm, instead of updating $B$ by $B(1)$, we update $B$ by $B(\alpha)$ for some $0 < \alpha < 1$. We have already incorporated this in our implementation. All codes for the simulation studies are written in `R` language and running under `R` version 2.4.0 on a machine with Pentium Duo core, CPU 3.20 GHz and 3.50 GB RAM. The code for `EM` is kindly provided by professor James at USC via personal communication. The computational cost is summarised in Table 8 for two settings. Note that, since `Newton` needs an initial estimate, the computational cost reported there is the additional time cost. As can be seen, `Newton` (together with obtaining initial estimates and calculating the approximate CV score) requires around 2.5 times as much effort as `EM`, and sometimes more compared to `loc`. A more efficient implementation of the estimation procedure is currently being pursued.

# 6 Application

As an application of our method to a real problem, we analyse the data on *CD4+ cell number count* collected as part of the Multicenter AIDS Cohort Study (MACS) [Kaslow *et al.* (1987)]. The data is from Diggle, Heagerty, Liang and Zeger (2002), and is downloadable at `http://www.maths.lancs.ac.uk/~diggle/lda/Datasets/lda.dat`. It consists of 2376 measurements of CD4+ cell counts against time since seroconversion (time when HIV becomes detectable which is used as zero on the time line) for 369 infected men enrolled in the study. Five patients, for whom there was only

one measurement, were removed from our analysis. For the rest 364 subjects, the number of measurements varies between 2 and 12, with a median of 6 and a standard deviation of 2.66. The time span of the study is about 8.5 years (covering about three years before seroconversion and 5.5 years after that). The goal of our analysis is to understand the variability of CD4 counts as a function of time since seroconversion. We expect that this will provide useful insights into the dynamics of the process. This data set has been analysed by many other authors using various approaches, including varying coefficient models [Fan and Zhang (2000), Wu and Chiang (2000)], functional principal component approach [Yao *et al.* (2005)] and parametric random effects models [Diggle *et al.* (2002)].

In our analysis, four methods: `EM`, `New.EM`, `loc` and `New.loc` are used. (The cubic B-spline basis with equally spaced knots are used for `EM` and `Newton`). Several different models, with $M$ taking values $5, 10, 15, 20$, and $r$ taking values $2, \ldots, 6$ are considered. The approximate cross-validation criterion $\widetilde{CV}$ is used for model selection. The model with $M = 10, r = 4$ results in the smallest score and thus is selected. Figure 4 shows the estimated eigenfunctions under the selected model. The estimates of the error variance and eigenvalues are given in Table 9. Under the selected model, `New.EM` and `EM` result in quite similar estimates for both eigenvalues and eigenfunctions. On the other hand, the estimates of `loc` are very different. For `loc`, $\widehat{\lambda}_1$ is much larger compared to that of $\widehat{\lambda}_2$, whereas in the case of `New.EM` and `EM`, they are of the same order. Since `New.loc` fails to converge under the selected model, its estimates are not reliable and thus not reported here. Moreover, based on our experience with the simulation studies, this might be an indicator that the corresponding results by `loc` are not altogether reliable either. The estimated error variance is about $38,000$ by `New.EM` and the results of `New.EM` suggest that there are 4 non-negligible eigenvalues, two of which are large, and the other two are relatively small.

Next, we give an interpretation of the shape of the mean function and that of the eigenfunctions. The estimated mean function is shown as the first panel of Figure 4, together with the optimal bandwidth by `h.select()` function in R package `sm`. The shape of the mean function reflects the fact that with the progression of the disease, the CD4+ cell count tends to decrease. The eigenfunctions capture the fluctuation of individual trajectories around the mean function. The first eigenfunction is rather flat compared to the other three eigenfunctions (Figure 4, panel two). This means that it mainly captures the baseline variability in the CD4+ cell count from one subject to another. This is consistent with the random effects model proposed in Diggle *et al.* (2002) (page 108-113). It is also noticeable that the second eigenfunction has a shape similar to that of the mean function (Figure 4, panel four). The shapes of the first two eigenfunctions, and the fact that their corresponding eigenvalues are relatively large, seem to indicate that a simple linear dynamical model, with random initial conditions, may be employed in studying the dynamics of CD4+ cell count. This observation is also consistent with the implication by the time-lagged graphs used in Diggle *et al.* (2002) (Fig. 3.13, p. 47). The third and fourth eigenvalues are comparable in magnitude to the error variance, and the corresponding eigenfunctions have somewhat similar shapes. They correspond to the contrast in variability between early and late stages of the disease. Of course, there are a number of measured and unmeasured covariates that are very likely to influence the dynamics of this process. Thus a more elaborate model that incorporates covariate effects should give a better interpretation of the eigenfunctions corresponding to the smaller eigenvalues, and that of the individual trajectories.

# 7 Discussion

In this paper, we presented a method that utilizes the intrinsic geometry of the parameter space explicitly to obtain the estimate in a non-regular problem, that of estimating eigenfunctions and eigenvalues of the covariance kernel when the data are only observed at sparse and irregular time points. We did comparative studies with two other estimation procedures by James *et al.* (2000) and Yao *et al.* (2005). We presented a model selection approach based on the minimization of an approximate cross-validation score with respect to the model parameters. Based on our simulation studies, we have found that the proposed geometric approach works well for both estimation and model selection. Moreover, its performance is in general better than that of the other two methods. We also looked at a real-data example to see how our method captures the variability in the data. In the following, we briefly sketch some on-going work relating to the problem studied in this paper.

There are a few aspects of the problem that can be investigated further. One is the asymptotic behaviour of our estimates. Asymptotic results have been established under very general conditions for the local polynomial method in Hall *et al.* (2006). Based on the numerical comparisons, it is expected that `Newton` with either `loc` or `EM` as initial estimate, should have at least as good a risk behavior as that of the local polynomial method. A closely related problem is the estimation of the eigenvalues and eigenvectors of covariance matrix for high dimensional multivariate Gaussian data, under the rank-restricted assumption. It is known that, in this case the usual PCA estimates, i.e., eigenvalues and eigenvectors of the sample covariance matrix, are the MLE's for their population counterparts [Muirhead (1982)]. In Paul (2005), it has been shown that under the above setting, risk of the PCA estimators of eigenvectors, measured under the squared error loss, achieves the optimal nonparametric rate when the dimension-to-sample size ratio converges to zero. Works currently being

pursued by the present authors indicate that the PCA (i.e. restricted ML) estimators should also achieve the asymptotically optimal risk. This is through an efficient score representation of the PCA estimator that utilizes the intrinsic geometry of the parameter space. We also proved consistency, and obtain rates of convergence of the proposed estimator for functional principal components, in a regime of relatively dense measurements. We are currently working on extending these results to sparse measurements case. The key components of the asymptotic analysis of our estimator are : (i) analysis of the expected loss function (for the Gaussian model, this is the Kullback-Leibler discrepancy); (ii) study of the Hessian of the loss function (*intrinsic information operator* in the Gaussian case). The essential difficulty of the analysis in the current context lies in the fact that the measurements are sparsely distributed in time. Regarding the rate of convergence, we do not expect the distribution of noise to play any significant role and the existence of enough moments should suffice.

Finally, we want to point out that, there are many statistical problems with (part of) the parameters having orthornormality constraints. Some examples include, extension of the current framework to spatio-temporal data, inclusion of covariate effects in FPCA, and problems involving orthonormality as natural identifiability constraints. As long as we have (i) explicit form and smoothness of the loss function; (ii) the ability to compute the intrinsic gradient and Hessian of the loss function, we can adopt a similar approach, for both estimation and model selection. Here we briefly discuss two examples which are closely related to the problem studied in this paper.

The first example relates to an alternative to the *restricted maximum likelihood* approach pursued in this paper. This involves representing the eigenfunctions in a sufficiently rich class of basis functions (i.e., $M$ large), and then adding a roughness penalty, e.g. $\kappa \sum_{\nu=1}^{r} \int (\psi_\nu''(t))^2 dt$, for some $\kappa > 0$, to the negative log-likelihood/loss function to control the degree of smoothness of the eigenfunctions [cf. Green and

Silverman (1994)]. If we use the expansion $\psi_\nu(t) = \sum_{k=1}^{M} b_{k\nu}\phi_k(t)$, for a known set of orthonormal functions $\{\phi_1, \ldots, \phi_M\}$, then the roughness penalty can be expressed as $\kappa\ Tr(B^T\mathbf{R}B)$, where $\mathbf{R}$ is the $M \times M$ matrix given by $\mathbf{R} = \int(\Phi''(t))(\Phi''(t))^T dt$, with $\Phi(\cdot) = (\phi_1(\cdot), \ldots, \phi_M(\cdot))^T$. Thus, the penalized log-likelihood is still a function of $B$ (and of $\Lambda$ and $\sigma^2$), where $B \in \mathcal{S}_{M,r}$. Straightforward algebra shows that the corresponding *penalized maximum likelihood estimate* can be obtained by simple modifications of the proposed procedure.

Another problem relates to incorporation of covariate effects in the analysis of longitudinal data. For example, Cardot (2006) studies a model where the covariance of $X(\cdot)$ conditioning on a covariate $W$ has the following expansion : $C^w(s,t) := \mathrm{Cov}(X(s), X(t)|W = w) = \sum_{\nu\geq 1} \lambda_\nu(w)\psi_\nu(s,w)\psi_\nu(t,w)$. He proposes a kernel-based nonparametric approach for estimating the eigenvalues and eigenfunctions (now dependent on $w$). In practice this method would require dense measurements. A modification of our method can easily handle the case, even for sparse measurements, when the eigenvalues are considered to be simple parametric functions of $w$, and eigenfunctions do not depend on $w$. For example, one model is $\lambda_\nu(w) := \alpha_\nu e^{w^T\beta_\nu}$, $\nu = 1, \ldots, r$, for some parameters $\beta_1, \ldots, \beta_r$, assuming that $\alpha_\nu = 0$ and $\beta_\nu = 0$ for $\nu > r$. This model captures the variability in amplitude of the eigenfunctions in the individual sample curves as a function of the covariate. In this setting we can express the conditional likelihood of the data $\{(\{Y_{ij}\}_{j=1}^{m_i}, W_i) : i = 1, \ldots, n\}$ explicitly. Its maximization under the restriction that the estimated eigenfunctions represented by a set of smooth basis functions can be carried out by a modification of the procedure proposed in this paper.

## Acknowledgements

The authors thank G. James for providing the code for the EM approach. We thank D. Temple Lang for helping in implementing the proposed procedure in R and P. Burman for insightful discussions.

## References

Ash, R. B. (1972) *Real Analysis and Probability*, Academic Press.

Besse, P., Cardot, H. and Ferraty, F. (1997) Simultaneous nonparametric regression of unbalanced longitudinal data. *Computational Statistics and Data Analysis* **24**, 255-270.

Boente, G. and Fraiman, R. (2000) Kernel-based functional principal components analysis. *Statistics and Probability Letters* **48**, 335-345.

Boyd, S. and Vandenberghe, L. (2004) *Convex Optimization*. Cambridge University Press.

Burman, P. (1990) Estimation of generalized additive models. *Journal of Multivariate Analysis* **32**, 230-255.

Cai, T. and Hall, P. (2006) Prediction in functional linear regression. *Annals of Statistics* **34**, 2159-2179.

Cardot, H., Ferraty F. and Sarda P. (1999) Functional Linear Model. *Statistics and Probability Letters* **45**, 11-22.

Cardot, H. (2000) Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics* **12**, 503-538.

Cardot, H. (2006) Conditional functional principal components analysis. *Scandina-*

*vian Journal of Statistics* **33**, 317-335.

Chui, C. (1987) *Multivariate Splines*. SIAM.

de Boor, C. (1978) *A Practical Guide to Splines*. Springer-Verlag, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1-38.

Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002) *Analysis of Longitudinal Data, 2nd. Edition*. Oxford University Press.

Edelman, A., Arias, T. A. and Smith, S. T. (1999) The geometry of algorithms with orthogonality constraints, *SIAM Journal on Matrix Analysis and Applications* **20**, 303-353.

Fan, J. and Zhang, J. T. (2000) Two-step estimation of functional linear models with applications to longitudinal data. *Journal of Royal Statistical Society, Series B* **62**, 303-322.

Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis : Theory and Practice*. Springer.

Green, P. J. and Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models : A Roughness Penalty Approach*. Chapman & Hall/CRC.

Hall, P. and Horowitz, J. L. (2007) Methodology and convergence rates for functional linear regression.
`http://www.faculty.econ.northwestern.edu/faculty/horowitz/papers/hhor-final.pdf`

Hall, P., Müller, H.-G. and Wang, J.-L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics* **34**, 1493-1517.

James, G. M., Hastie, T. J. and Sugar, C. A. (2000) Principal component models for sparse functional data. *Biometrika*, **87**, 587-602.

Kaslow R. A., Ostrow D. G., Detels R., Phair J. P., Polk B. F., Rinaldo C. R. (1987) The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology* **126**(2), 310-318.

Lee, J. M. (1997) *Riemannian Manifolds: An Introduction to Curvature*, Springer.

Marron, S. J., Müller, H.-G., Rice, J., Wang, J.-L., Wang, N. and Wang, Y. (2004) Discussion of nonparametric and semiparametric regression. *Statistica Sinica* **14**, 615-629.

Meyer, K. and Kirkpatrick, M. (2005) Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. *Genetic Selection Evolution* **37**, 1-30.

Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*, John Wiley & Sons.

Paul, D. (2005) *Nonparametric Estimation of Principal Components.* Ph.D. Thesis, *Stanford University.*

Rice, J. A. and Wu, C. O. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253-259.

Ramsay, J. and Silverman, B. W. (2005) *Functional Data Analysis, 2nd Edition.* Springer.

Smith, S. T. (2005) Covariance, subspace and intrinsic Cramér-Rao bounds. *IEEE Transactions on Signal Processing* **53**, 1610-1630.

Wu, C. and Chiang, C. (2000) Kernel smoothing on varying coefficient models with longitudinal dependent variables. *Statistica Sinica* **10**, 433-456.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577-590.

Yao, F., Müller, H.-G. and Wang, J.-L. (2006) Functional linear regression for longitudinal data. *Annals of Statistics* **33**, 2873-2903.

# Appendix A : Review of some Riemannian geometric concepts

Let $(\mathcal{M}, g)$ be a smooth manifold with Riemannian metric $g$. We shall denote the tangent space of $\mathcal{M}$ at $p \in \mathcal{M}$ by $\mathcal{T}_p\mathcal{M}$. We shall first give some basic definitions related to the work we present in this article. A good reference is Lee (1997).

## Gradient and Hessian of a function

- **Gradient :** Let $f : \mathcal{M} \to \mathbb{R}$ be a smooth function. Then $\nabla f$, the *gradient* of $f$, is a *vector field* on $\mathcal{M}$ defined by the following:

  for any $X \in \mathcal{T}\mathcal{M}$, (i.e., a vector field on $\mathcal{M}$), $\langle \nabla f, X \rangle_g = X(f)$, where $X(f)$ is the *directional derivative* of $f$ w.r.t. $X$ : $X(f)\,|_p = \frac{df(\gamma(t))}{dt}\,|_{t=0}$ for any differentiable curve $\gamma$ on $\mathcal{M}$ with $\gamma(0) = p$, $\dot{\gamma}(0) = X(p)$.

  Note that $X(f) : p \to X(f)\,|_p$ is a function that maps $\mathcal{M}$ to $\mathbb{R}$.

- **Covariant derivative :** (also known as *Riemannian connection*) : Let $X, Y \in \mathcal{T}\mathcal{M}$ be two vector fields on $\mathcal{M}$. Then the vector field $\overline{\nabla}_Y X \in \mathcal{T}\mathcal{M}$ is called the *covariant derivative of $X$ in the direction of $Y$* if the operator $\overline{\nabla}$ satisfies the following properties:

(a) (Bi-linearity) : For $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$\overline{\nabla}_Y(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 \overline{\nabla}_Y X_1 + \lambda_2 \overline{\nabla}_Y X_2$$

and

$$\overline{\nabla}_{\lambda_1 Y_1 + \lambda_2 Y_2} X = \lambda_1 \overline{\nabla}_{Y_1} X + \lambda_2 \overline{\nabla}_{Y_2} X.$$

(b) (Leibniz) : for a smooth function $f : \mathcal{M} \to \mathbb{R}$ ,

$$\overline{\nabla}_Y(f \cdot X) = Y(f) \cdot X + f \cdot \overline{\nabla}_Y X.$$

(c) (Preserving metric) : for $X, Y, Z \in \mathcal{TM}$,

$$Z(\langle X, Y \rangle_g) = \langle \overline{\nabla}_Z X, Y \rangle_g + \langle X, \overline{\nabla}_Z Y \rangle_g.$$

(d) (Symmetry) : $\overline{\nabla}_X Y - \overline{\nabla}_Y X = [X, Y]$ where $[X, Y] := X(Y) - Y(X) \in \mathcal{TM}$, i.e., for a smooth $f : \mathcal{M} \to \mathbb{R}$, $[X, Y](f) = X(Y(f)) - Y(X(f))$.

- **Hessian operator:** For a smooth function $f : \mathcal{M} \to \mathbb{R}$, $H_f : \mathcal{TM} \times \mathcal{TM} \to \mathbb{R}$ is the bi-linear form defined as

$$H_f(Y, X) = \langle \overline{\nabla}_Y(\nabla f), X \rangle_g, \ X, Y \in \mathcal{TM}.$$

Note that, by definition, $H_f$ is bi-linear and symmetric (i.e., $H_f(Y, X) = H_f(X, Y)$). For national simplicity, sometimes we also write $\overline{\nabla}_Y(\nabla f)$ as $H_f(Y)$. $H_f$ can be calculated in the following manner. Note that, for a smooth curve $\gamma(\cdot)$ on the

manifold $\mathcal{M}$, with $X = \dot{\gamma}(t)$, it follows that

$$\begin{aligned}\frac{d^2 f(\gamma(t))}{dt^2} = \frac{d}{dt}\left(\frac{df(\gamma(t))}{dt}\right) &= \frac{d}{dt}(\langle \nabla f, X\rangle_g \big|_{\gamma(t)}) \\ &= X(\langle \nabla f, X\rangle_g) = \langle \overline{\nabla}_X(\nabla f), X\rangle_g + \langle \nabla f, \overline{\nabla}_X X\rangle_g,\end{aligned}$$

where the last step follows by applying the *Leibniz rule* for the covariant derivative. Since $\gamma(\cdot)$ is a geodesic if and only if $\overline{\nabla}_X X = 0$ (self-parallel), this implies that, for such a $\gamma(\cdot)$,

$$\frac{d^2 f(\gamma(t))}{dt^2} = \langle \overline{\nabla}_X(\nabla f), X\rangle_g = H_f(X, X).$$

From this, we can derive the Hessian of $f$:

$$H_f(X, Y) = \frac{1}{2}(H_f(X + Y, X + Y) - H_f(X, X) - H_f(Y, Y)).$$

- **Inverse of Hessian :** For $X \in \mathcal{TM}$, and a smooth function $f : \mathcal{M} \to \mathbb{R}$, $H_f^{-1}(X) \in \mathcal{TM}$ is defined as the vector field satisfying: for $\forall\ \Delta \in \mathcal{TM}$,

$$H_f(H_f^{-1}(X), \Delta) = \langle X, \Delta\rangle_g.$$

To understand the definition of $H_f^{-1}$, note that if $H : \mathcal{TM} \times \mathcal{TM} \to \mathbb{R}$ is bilinear and $\langle \cdot, \cdot\rangle_g$ is an inner product on $\mathcal{TM}$ (i.e., a Riemannian metric on $\mathcal{M}$), then by the *Riesz representation theorem*, $\exists!\ A : \mathcal{TM} \to \mathcal{TM}$, that is 1-1 and linear such that $H(X, Y) = \langle A(X), Y\rangle_g$. This implies that $H(A^{-1}(X), Y) = \langle X, Y\rangle_g$, so that $H_f^{-1} = A^{-1}$ when $H = H_f$.

## Some facts about Stiefel manifold

The manifold $\mathcal{M} = \{B \in \mathbb{R}^{M\times r} : B^TB = I_r\}$ is known as the *Steifel manifold* in $\mathbb{R}^{M\times r}$. Here we present some basic facts about this manifold which are necessary for implementing the proposed method. A more detailed description is given in Edelman *et al.* (1999).

- **Tangent space :** $\mathcal{T}_B\mathcal{M} = \{\Delta \in \mathbb{R}^{M\times r} : B^T\Delta$ is skew-symmetric $\}$.
- **Canonical metric :** For $\Delta_1, \Delta_2 \in \mathcal{T}_B\mathcal{M}$ with $B \in \mathcal{M}$, the *canonical metric* (a Riemannian metric on $\mathcal{M}$) is defined as

$$\langle \Delta_1, \Delta_2 \rangle_c = Tr(\Delta_1^T(I - \frac{1}{2}BB^T)\Delta_2).$$

- **Gradient :** For a smooth function $f : \mathcal{M} \to \mathbb{R}$,

$$\nabla f\,|_B = f_B - Bf_B^TB,$$

  where $f_B$ is the usual Euclidean gradient of $f$ defined through $(f_B)_{ij} = \frac{\partial f}{\partial B_{ij}}$.

- **Hessian operator :** (derived from the geodesic equation): For $\Delta_1, \Delta_2 \in \mathcal{T}_B\mathcal{M}$,

$$H_f(\Delta_1, \Delta_2)\,|_B = f_{BB}(\Delta_1, \Delta_2) + \frac{1}{2}Tr\left[(f_B^T\Delta_1B^T + B^T\Delta_1f_B^T)\Delta_2\right] - \frac{1}{2}Tr\left[(B^Tf_B + f_B^TB)\Delta_1^T\Pi\Delta_2\right],$$

  where $\Pi = I - BB^T$.

- **Inverse of Hessian :** For $\Delta, G \in \mathcal{T}_B\mathcal{M}$, the equation $\Delta = H_f^{-1}(G)$ means that $\Delta$ is the solution of

$$f_{BB}(\Delta) - B\ \text{skew}(f_B^T\Delta) - \ \text{skew}(\Delta f_B^T)B - \frac{1}{2}\Pi\Delta B^Tf_B = G,$$

subject to the condition that $B^T\Delta$ is skew-symmetric, i.e., $B^T\Delta + \Delta^T B = 0$, where $f_{BB}(\Delta) \in \mathcal{T}_B\mathcal{M}$ such that

$$\langle f_{BB}(\Delta), X\rangle_c = f_{BB}(\Delta, X) = Tr(\Delta^T f_{BB} X) \qquad \forall\ X \in \mathcal{T}_B\mathcal{M}.$$

This implies that $f_{BB}(\Delta) = H(\Delta) - BH^T(\Delta)B$, where $H(\Delta) = f_{BB}^T\Delta$. Here $\text{skew}(X) = \frac{1}{2}(X - X^T)$.

## Exponential of skew-symmetric matrices

Let $X = -X^T$ be a $p \times p$ matrix. Want to compute $\exp(tX) := \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k$ for $t \in \mathbb{R}$, where $X^0 = I$. Let the SVD of $X$ be given by $X = UDV^T$, where $U^TU = V^TV = I_p$, and $D$ is diagonal. So, $X^2 = XX = -XX^T = -UDV^TVDU^T = -UD^2U^T$. This also shows that all the eigenvalues of $X$ are purely imaginary. Using the facts that $D^0 = I_p$; $X^{2k} = (X^2)^k = (-1)^k(UD^2U^T)^k = (-1)^kUD^{2k}U^T$; and $X^{2k+1} = (-1)^kUD^{2k}U^TUDV^T = (-1)^kUD^{2k+1}V^T$, we have

$$\begin{aligned}\exp(tX) &= U\left[\sum_{k=0}^{\infty}\frac{(-t)^k}{(2k)!}D^{2k}\right]U^T + U\left[\sum_{k=0}^{\infty}\frac{(-t)^k}{(2k+1)!}D^{2k+1}\right]V^T \\ &= U\cos(tD)U^T + U\sin(tD)V^T,\end{aligned}$$

where $\cos(tD) = diag((\cos(td_{jj}))_{j=1}^p)$ and $\sin(tD) = diag((\sin(td_{jj}))_{j=1}^p)$, if $D = diag((d_{jj})_{j=1}^p)$.

## Vectorization of matrix equations

A general form of the equation in the $M \times r$ matrix $\Delta$ is given by

$$L = A\Delta + \Delta K + C\Delta D + E\Delta^T F,$$

where $L$ is $M \times r$, $A$ is $M \times M$, $K$ is $r \times r$, $C$ is $M \times M$, $D$ is $r \times r$, $E$ is $M \times r$, and $F$ is $M \times r$. Vectorization of this equation using the vec operation means that $\text{vec}(L)$ is given by

$$\begin{aligned} & \text{vec}(A\Delta) + \text{vec}(\Delta K) + \text{vec}(C\Delta D) + \text{vec}(E\Delta^T F) \\ = \quad & \left[(I_r \otimes A) + (K^T \otimes I_M) + (D^T \otimes C) + (F^T \otimes E)P_{M,r}\right] \text{vec}(\Delta), \end{aligned} \tag{17}$$

where, $\otimes$ denotes the Kronecker product, and we have used the following properties of the vec operator (Muirhead (1982)): (i) $\text{vec}(KXC) = (C^T \otimes K)\text{vec}(X)$; (ii) $\text{vec}(X^T) = P_{m,n}\text{vec}(X)$. Here $X$ is $m \times n$, $K$ is $r \times m$, $C$ is $n \times s$, and $P_{m,n}$ is an appropriate $mn \times mn$ permutation matrix.

# Appendix B : Gradients and Hessians with respect to $B$

We use the following lemmas (cf. Muirhead (1982)) repeatedly in our computations in this subsection.

**Lemma 1 :** *Let $P = I_p + AE$ where $A$ is $p \times q$, $E$ is $q \times p$. Then*

$$\det(P) = |I_p + AE| = |I_q + EA|.$$

**Lemma 2 :** *Let $A$ be $p \times p$ and $E$ be $q \times q$, both nonsingular. If $P = A + CED$, for*

*any $p \times q$ matrix $C$ and any $q \times p$ matrix $D$, then*

$$P^{-1} = (A + CED)^{-1} = A^{-1}[A - CQ^{-1}D]A^{-1}, \quad \textit{where}, \quad Q = E^{-1} + DA^{-1}C$$

*is $q \times q$.*

## Application to the likelihood setting

Let $P_i = \sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i$ (is an $m_i \times m_i$ matrix), where $\Phi_i$ is $M \times m_i$, $B$ is $M \times r$ and $\Lambda$ is $r \times r$ matrices. Then, by *Lemma 1*,

$$|P_i| = \sigma^{2m_i}|I_r + \sigma^{-2}\Lambda B^T \Phi_i \Phi_i^T B| = \sigma^{2(m_i - r)}|\Lambda||\sigma^2 \Lambda^{-1} + B^T \Phi_i \Phi_i^T B| = \sigma^{2(m_i - r)}|\Lambda||Q_i|, \tag{18}$$

where

$$Q_i = \sigma^2 \Lambda^{-1} + B^T \Phi_i \Phi_i^T B$$

is an $r \times r$ positive definite matrix. Also, by *Lemma 2*

$$P_i^{-1} = \sigma^{-2} I_{m_i} - \sigma^{-4} \Phi_i^T B(\Lambda^{-1} + \sigma^{-2} B^T \Phi_i \Phi_i^T B)^{-1} B^T \Phi_i = \sigma^{-2} \left[ I_{m_i} - \Phi_i^T B Q_i^{-1} B^T \Phi_i \right]. \tag{19}$$

In our problem, we consider the $i$-th term in the expression for the log-likelihood and recall that $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \boldsymbol{\mu}_i$. Then

$$\begin{aligned} F_i^1 &= Tr[(\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i)^{-1} \widetilde{\mathbf{Y}}_i \widetilde{\mathbf{Y}}_i^T] = Tr(P_i^{-1} \widetilde{\mathbf{Y}}_i \widetilde{\mathbf{Y}}_i^T) \\ F_i^2 &= \log|\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i| = \log|P_i|. \end{aligned} \tag{20}$$

For simplifying notations we shall drop the subscript $i$ from these functions. We view $F^1 = F^1(B)$ as a function of $B$. Since $F^1 = Tr(P_i^{-1}\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T)$, using (19) we have

$$\begin{aligned} F^1(B) = Tr(P_i^{-1}\widetilde{Y}_i\widetilde{\mathbf{Y}}_i^T) &= \sigma^{-2}Tr(\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T) - \sigma^{-2}Tr(\Phi_i^T B Q_i^{-1} B^T \Phi_i \widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T) \\ &= \sigma^{-2}Tr(\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T) - \sigma^{-2}Tr(B Q_i^{-1} B^T \Phi_i \widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T \Phi_i^T). \quad (21) \end{aligned}$$

Similarly,

$$F^2 = F^2(B) = \log|P_i| = \log(\sigma^{2(m_i-r)}|\Lambda|) + \log|Q_i|. \quad (22)$$

## Gradient of $F^1$

Let $B(t) = B + t\Delta$. Then

$$\frac{dQ_i(t)}{dt}\,|_{t=0} = \Delta^T \Phi_i \Phi_i^T B + B^T \Phi_i \Phi_i^T \Delta,$$

so that

$$\begin{aligned} \langle F_B^1, \Delta\rangle &= \frac{dF^1(B(t))}{dt}\,|_{t=0} \\ &= -\sigma^{-2}Tr\left[(\Delta Q_i^{-1}B^T + BQ_i^{-1}\Delta^T - BQ_i^{-1}\frac{dQ_i}{dt}\,|_{t=0}\, Q_i^{-1}B^T)\,\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T\right] \\ &= -2\sigma^{-2}Tr\left[(\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T BQ_i^{-1} - \Phi_i\Phi_i^T BQ_i^{-1}B^T\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T BQ_i^{-1})\Delta^T\right] \quad (23) \end{aligned}$$

Thus the Euclidean gradient of $F^1$ is

$$\begin{aligned} F_B^1 &= -2\sigma^{-2}\left[\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T BQ_i^{-1} - \Phi_i\Phi_i^T BQ_i^{-1}B^T\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T BQ_i^{-1}\right] \\ &= 2\sigma^{-2}\left[\Phi_i\Phi_i^T BQ_i^{-1}B^T - I_M\right]\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T BQ_i^{-1}, \end{aligned}$$

and the gradient of $F^1$ with respect to $B$ is

$$\nabla F^1 = F^1_B - B(F^1_B)^T B. \tag{24}$$

## Hessian of $F^1$

Let $B(t,s) = B + t\Delta + sX$. Then using (23),

$$\begin{aligned}
F^1_{BB}(\Delta, X) &= \langle F^1_{BB}(\Delta), X\rangle_c \\
&= \langle H^1_{BB}(\Delta), X\rangle = \frac{\partial}{\partial t}\frac{\partial}{\partial s} F^1(B(t,s))\,|_{s,t=0} \\
&= 2\sigma^{-2} Tr\left[\frac{\partial}{\partial t}(\Phi_i\Phi_i^T B(t,0)Q_i(t)^{-1}B(t,0)^T - I_M)\,|_{t=0}\ \Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T B Q_i^{-1}X^T\right] \\
&\quad + 2\sigma^{-2} Tr\left[(\Phi_i\Phi_i^T B Q_i^{-1}B^T - I_M)\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T\frac{\partial}{\partial t}(B(t,0)Q_i(t)^{-1})\,|_{t=0}\ X^T\right].
\end{aligned}$$

Note that

$$\frac{\partial}{\partial t}(B(t,0)Q_i(t)^{-1})\,|_{t=0} = \Delta Q_i^{-1} - BQ_i^{-1}(\Delta^T\Phi_i\Phi_i^T B + B^T\Phi_i\Phi_i^T\Delta)Q_i^{-1},$$

and

$$\frac{\partial}{\partial t}(B(t,0)Q_i(t)^{-1}B(t,0)^T)\,|_{t=0} = \Delta Q_i^{-1}B^T + BQ_i^{-1}\Delta^T - BQ_i^{-1}(\Delta^T\Phi_i\Phi_i^T B + B^T\Phi_i\Phi_i^T\Delta)Q_i^{-1}B^T.$$

Thus $H^1_{BB}(\Delta)$ is given by,

$$\begin{aligned}
&H^1_{BB}(\Delta) \\
=\ & 2\sigma^{-2}\Phi_i\Phi_i^T\left[\Delta Q_i^{-1}B^T + BQ_i^{-1}\Delta^T - BQ_i^{-1}(\Delta^T\Phi_i\Phi_i^T B + B^T\Phi_i\Phi_i^T\Delta)Q_i^{-1}B^T\right]\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T B Q_i^{-1} \\
&+ 2\sigma^{-2}\left[(\Phi_i\Phi_i^T B Q_i^{-1}B^T - I_M)\Phi_i\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\Phi_i^T(\Delta Q_i^{-1} - BQ_i^{-1}(\Delta^T\Phi_i\Phi_i^T B + B^T\Phi_i\Phi_i^T\Delta)Q_i^{-1})\right],
\end{aligned}$$

and

$$F^1_{BB}(\Delta) = H^1_{BB}(\Delta) - B(H^1_{BB}(\Delta))^T B. \tag{25}$$

## Gradient of $F^2$

Let $B(t) = B + t\Delta$. Then

$$\begin{aligned}
\langle F^2_B, \Delta\rangle = \frac{dF^2(B(t))}{dt}\mid_{t=0} &= Tr\left(Q_i^{-1}\frac{dQ_i}{dt}\mid_{t=0}\right) \\
&= Tr(Q_i^{-1}(\Delta^T \Phi_i \Phi_i^T B + B^T \Phi_i \Phi_i^T \Delta)) \\
&= 2Tr(Q_i^{-1} B^T \Phi_i \Phi_i^T \Delta). \qquad (26)
\end{aligned}$$

Thus

$$\nabla F^2 = F^2_B - B(F^2_B)^T B, \quad \text{where} \quad F^2_B = 2\,\Phi_i \Phi_i^T B Q_i^{-1}. \tag{27}$$

## Hessian of $F^2$

Let $B(t,s) = B + t\Delta + sX$. Then using (26),

$$\begin{aligned}
F^2_{BB}(\Delta, X) &= \langle F^2_{BB}(\Delta), X\rangle_c \\
&= \langle H^2_{BB}(\Delta), X\rangle = \frac{\partial}{\partial t}\frac{\partial}{\partial s} F^2(B(t,s))\mid_{s,t=0} \\
&= \frac{\partial}{\partial t}[2Tr(Q_i(t)^{-1} B(t,0)^T \Phi_i \Phi_i^T X)]\mid_{t=0} \\
&= 2Tr\left[(-Q_i^{-1}\frac{dQ_i(t)}{dt}\mid_{t=0} Q_i^{-1} B^T + Q_i^{-1}\Delta^T)\,\Phi_i \Phi_i^T X\right] \\
&= 2Tr\left[(-Q_i^{-1}(\Delta^T \Phi_i \Phi_i^T B + B^T \Phi_i \Phi_i^T \Delta) Q_i^{-1} B^T + Q_i^{-1}\Delta^T)\,\Phi_i \Phi_i^T X\right].
\end{aligned}$$

From this $H^2_{BB}(\Delta)$ is

$$\begin{aligned} H^2_{BB}(\Delta) &= 2\left[-Q_i^{-1}(\Delta^T\Phi_i\Phi_i^T B + B^T\Phi_i\Phi_i^T\Delta)Q_i^{-1}B^T + Q_i^{-1}\Delta^T)\Phi_i\Phi_i^T\right]^T \\ &= 2\Phi_i\Phi_i^T\left[\Delta - BQ_i^{-1}(\Delta^T\Phi_i\Phi_i^T B + B^T\Phi_i\Phi_i^T\Delta)\right]Q_i^{-1}, \end{aligned}$$

and

$$F^2_{BB}(\Delta) = H^2_{BB}(\Delta) - B(H^2_{BB}(\Delta))^T B, \tag{28}$$

# Appendix C : Derivation of $\widetilde{CV}$ (16)

For now, in (14), considering only the part corresponding to the gradient w.r.t. $B$ and expanding it around $\widehat{\Psi}$, while approximating $(\widehat{\tau}^{(-i)}, \widehat{\boldsymbol{\zeta}}^{(-i)})$ by $(\widehat{\tau}, \widehat{\boldsymbol{\zeta}})$, we have (for notational simplicity, write $\ell_j(\widehat{B})$ to denote $\ell_j(\widehat{\Psi})$)

$$0 = \sum_{j\neq i}\nabla_B\ell_j(\widehat{\Psi}^{(-i)}) \approx \sum_{j\neq i}\nabla_B\ell_j(\widehat{B}) + \sum_{j\neq i}\overline{\nabla}_{\Delta_i}(\nabla_B\ell_j(\widehat{B})), \tag{29}$$

where $\overline{\nabla}_{\Delta_i}(\nabla_B\ell_j)$ is the *covariant derivative* of $\nabla_B\ell_j$ in the direction of $\Delta_i$. Now, substituting (13) in (29), we get

$$0 \approx -\nabla_B\ell_i(\widehat{B}) + \overline{\nabla}_{\Delta_i}[\sum_{j\neq i}\nabla_B\ell_j(\widehat{B})]. \tag{30}$$

Then for any $X \in \mathcal{T}_{\widehat{B}}\mathcal{M}$,

$$\langle\overline{\nabla}_{\Delta_i}(\sum_{j\neq i}\nabla_B\ell_j(\widehat{B})), X\rangle_c = [\sum_{j\neq i}\nabla^2_B\ell_j(\widehat{B})](\Delta_i, X) \approx \langle\nabla_B\ell_i(\widehat{B}), X\rangle_c.$$

Thus by the definition of the Hessian inverse operator,

$$\Delta_i \approx [\sum_{j\neq i} \nabla_B^2 \ell_j(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})).$$

This, together with (30), leads to the approximation of $\Delta_i$,

$$\begin{aligned}
\Delta_i &\approx [\sum_{j\neq i} \nabla_B^2 \ell_j(\widehat{B})]^{-1}\nabla_B \ell_i(\widehat{B}) = [\sum_{j} \nabla_B^2 \ell_j(\widehat{B}) - \nabla_B^2 \ell_i(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})) \\
&\approx \left[I + [\sum_{j} \nabla_B^2 \ell_j(\widehat{B})]^{-1}\nabla_B^2 \ell_i(\widehat{B})\right] [\sum_{j} \nabla_B^2 \ell_j(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})) \\
&= \left[I + [\mathbf{H}_B(\widehat{B})]^{-1}\nabla_B^2 \ell_i(\widehat{B})\right] [\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})), \qquad (31)
\end{aligned}$$

where $\mathbf{H}_B = \sum_j \nabla_B^2 \ell_j$. Note that the last approximation is because, for linear operators $A$ and $C$ such that $A$ is invertible and $\| A^{-1}C \|$ is small, we have

$$(A - C)^{-1} = (A(I - A^{-1}C))^{-1} = (I - A^{-1}C)^{-1}A^{-1} \approx (I + A^{-1}C)A^{-1}.$$

The interpretation of the linear operator $[\mathbf{H}_B(\widehat{B})]^{-1}\nabla_B^2 \ell_i(\widehat{B})$ in (31) is as follows : for $\gamma \in \mathcal{T}_{\widehat{B}}\mathcal{M}$,

$$\left[[\mathbf{H}_B(\widehat{B})]^{-1}\nabla_B^2(\ell_i(\widehat{B}))\right](\gamma) = [\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B^2(\ell_i(\widehat{B}))(\gamma)).$$

Now, for $X \in \mathcal{T}_{\widehat{B}}\mathcal{M}$, by definition of Hessian,

$$\begin{aligned}
&\langle \sum_{j=1}^{n} \overline{\nabla}_{\Delta_i}(\nabla_B \ell_j(\widehat{B})), X\rangle_c = [\nabla_B^2(\sum_{j=1}^{n} \ell_j(\widehat{B}))](\Delta_i, X) \\
\approx\ & \mathbf{H}_B(\widehat{B})\left([\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})) + [(\mathbf{H}_B(\widehat{B})]^{-1}\nabla_B^2 \ell_i(\widehat{B})[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})), X\right) \\
=\ & \langle \nabla_B \ell_i(\widehat{B}), X\rangle_c + \langle \nabla_B^2(\ell_i(\widehat{B}))[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})), X\rangle_c, \qquad (32)
\end{aligned}$$

where, by definition,

$$\nabla_B^2(\ell_i(\widehat{B}))(\gamma) = \overline{\nabla}_\gamma(\nabla_B \ell_i(\widehat{B}))$$

for $\gamma \in \mathcal{T}_{\widehat{B}}\mathcal{M}$. In the first approximation of (32), we have used the approximation (31), and the last step follows from the definition of Hessian inverse and linearity of the Hessian. From (29) we also have,

$$\overline{\nabla}_{\Delta_i}(\nabla_B \ell_i(\widehat{B})) \approx -\nabla_B \ell_i(\widehat{B}) + \sum_{j=1}^{n} \overline{\nabla}_{\Delta_i}(\nabla_B \ell_j(\widehat{B})). \tag{33}$$

Substituting (32) in (33), we then have the approximation

$$\nabla_B^2 \ell_i(\widehat{B})(\Delta_i) = \overline{\nabla}_{\Delta_i}(\nabla_B \ell_i(\widehat{B})) \;\approx\; \nabla_B^2 \ell_i(\widehat{B})[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})). \tag{34}$$

Using (31) and (34), and ignoring terms higher than the second order, we have the approximation

$$\begin{aligned}
& \sum_{i=1}^{n} \langle \nabla_B \ell_i(\widehat{B}), \Delta_i \rangle_c + \frac{1}{2} \sum_{i=1}^{n} \nabla_B^2 \ell_i(\widehat{B})(\Delta_i, \Delta_i) \\
\approx\; & \left[ \sum_{i=1}^{n} \langle \nabla_B \ell_i(\widehat{B}), [\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})) \rangle_c + \sum_{i=1}^{n} \langle \nabla_B \ell_i(\widehat{B}), [\mathbf{H}_B(\widehat{B})]^{-1} \nabla_B^2 \ell_i(\widehat{B})[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})) \rangle_c \right] \\
& + \frac{1}{2} \sum_{i=1}^{n} \langle [\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})), \nabla_B^2 \ell_i(\widehat{B})[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B \ell_i(\widehat{B})) \rangle_c \\
=\; & \sum_{i=1}^{n} \langle \nabla_B \ell_i(\widehat{B}), [\mathbf{H}_B(\widehat{B})]^{-1} \nabla_B \ell_i(\widehat{B}) \rangle_c + \frac{3}{2} \sum_{i=1}^{n} \nabla_B^2 \ell_i(\widehat{B})([\mathbf{H}_B(\widehat{B})]^{-1} \nabla_B \ell_i(\widehat{B}), [\mathbf{H}_B(\widehat{B})]^{-1} \nabla_B \ell_i(\widehat{B})).
\end{aligned} \tag{35}$$

Here we give brief justifications for the steps in (35). The first approximation follows from the definition of Hessian, and the approximation of $\Delta_i$ by (31). The last equation follows from: (i) by definition of Hessian, applied to $\nabla_B^2 \ell_i(\widehat{B})$, the term on the third

line equals

$$\frac{1}{2}\sum_{i=1}^{n}\nabla_B^2\ell_i(\widehat{B})([\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B\ell_i(\widehat{B})),[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B\ell_i(\widehat{B})));$$

and (ii) the second term on the second line equals the same term as above, except for the factor $\frac{1}{2}$, by definition of Hessian$^{-1}$, now applied to $[\mathbf{H}_B(\widehat{B})]^{-1}$:

$$\begin{aligned}
& \langle\nabla_B\ell_i(\widehat{B}),[\mathbf{H}_B(\widehat{B})]^{-1}\nabla_B^2\ell_i(\widehat{B})[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B\ell_i(\widehat{B}))\rangle_c \\
= \;& \mathbf{H}_B(\widehat{B})([\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B\ell_i(\widehat{B})),[\mathbf{H}_B(\widehat{B})]^{-1}\nabla_B^2\ell_i(\widehat{B})[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B\ell_i(\widehat{B}))) \\
= \;& \langle[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B\ell_i(\widehat{B})),\nabla_B^2\ell_i(\widehat{B})[\mathbf{H}_B(\widehat{B})]^{-1}(\nabla_B\ell_i(\widehat{B}))\rangle_c.
\end{aligned}$$

Using very similar (but conceptually much simpler) arguments, we also have the second order approximation

$$\begin{aligned}
& \sum_{i=1}^{n}\langle\nabla_{(\tau,\zeta)}\ell_i(\widehat{\tau},\widehat{\boldsymbol{\zeta}}),(\delta_\tau^i,\boldsymbol{\delta}_\zeta^i)^T\rangle+\frac{1}{2}\sum_{i=1}^{n}\langle[\nabla_{(\tau,\zeta)}^2\ell_i(\widehat{\tau},\widehat{\boldsymbol{\zeta}})](\delta_\tau^i,\boldsymbol{\delta}_\zeta^i)^T,(\delta_\tau^i,\boldsymbol{\delta}_\zeta^i)^T\rangle \\
\approx \;& \sum_{i=1}^{n}\langle\nabla_{(\tau,\zeta)}\ell_i(\widehat{\tau},\widehat{\boldsymbol{\zeta}}),[\mathbf{H}_{(\tau,\zeta)}(\widehat{\tau},\widehat{\boldsymbol{\zeta}})]^{-1}\nabla_{(\tau,\zeta)}\ell_i(\widehat{\tau},\widehat{\boldsymbol{\zeta}})\rangle \\
+ \;& \frac{3}{2}\sum_{i=1}^{n}\langle\nabla_{(\tau,\zeta)}^2\ell_i(\widehat{\tau},\widehat{\boldsymbol{\zeta}})[\mathbf{H}_{(\tau,\zeta)}(\widehat{\tau},\widehat{\boldsymbol{\zeta}})]^{-1}\nabla_{(\tau,\zeta)}\ell_i(\widehat{\tau},\widehat{\boldsymbol{\zeta}}),[\mathbf{H}_{(\tau,\zeta)}(\widehat{\tau},\widehat{\boldsymbol{\zeta}})]^{-1}\nabla_{(\tau,\zeta)}\ell_i(\widehat{\tau},\widehat{\boldsymbol{\zeta}})\rangle. \quad (36)
\end{aligned}$$

Combining (15), (35) and (36), we have the approximate CV score given by (16).

# Appendix D : Gradients and Hessians with respect to $\boldsymbol{\zeta}$ and $\tau$

Define $H_i = \Phi_i^T B$, $i = 1,\ldots,n$. In the following we shall use $\exp(\boldsymbol{\zeta})$ and $\exp(\tau)$to denote the $r\times r$ diagonal matrix $\Lambda$ and $\sigma^2$, respectively. Then, as defined in Appendix

A,

$$P_i = \sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i = \sigma^2 I_{m_i} + H_i \Lambda H_i^T = e^\tau I_{m_i} + H_i \exp(\boldsymbol{\zeta}) H_i^T,$$

$$Q_i = \sigma^2 \Lambda^{-1} + B^T \Phi_i \Phi_i^T B = \sigma^2 \Lambda^{-1} + H_i^T H_i = e^\tau \exp(-\boldsymbol{\zeta}) + H_i^T H_i.$$

and re-writing (19):

$$P_i^{-1} = \sigma^{-2} I_{m_i} - \sigma^{-4} H_i (\Lambda^{-1} + \sigma^{-2} H_i^T H_i)^{-1} H_i^T = e^{-\tau} \left[ I_{m_i} - H_i Q_i^{-1} H_i^T \right].$$

For future uses, we calculate the following derivatives.

$$\frac{\partial P_i}{\partial \tau} = \frac{\partial}{\partial \tau} [e^\tau I_{m_i} + H_i \exp(\boldsymbol{\zeta}) H_i^T] = e^\tau I_{m_i}. \tag{37}$$

Let $H_{ik} = \Phi_i^T B_k$, $k = 1, \ldots, r$ where $B_k$ is the $k$-th column of $B$. We shall use the following fact

$$\frac{\partial P_i}{\partial \zeta_k} = \frac{\partial}{\partial \zeta_k} [e^\tau I_{m_i} + \sum_{k=1}^{r} e^{\zeta_k} H_{ik} H_{ik}^T] = e^{\zeta_k} H_{ik} H_{ik}^T. \tag{38}$$

In the following, we shall drop the subscript $i$ from the functions $F_i^1$ and $F_i^2$, and treat the latter as functions of $(\tau, \zeta)$.

## Gradient of $F^1$ and $F^2$

By direct computations we have,

$$\begin{aligned} \frac{\partial F^1}{\partial \tau} &= \frac{\partial}{\partial \tau} Tr[P_i^{-1} \widetilde{\mathbf{Y}}_i \widetilde{\mathbf{Y}}_i^T] = -Tr \left[ P_i^{-1} \left( \frac{\partial P_i}{\partial \tau} \right) P_i^{-1} \widetilde{\mathbf{Y}}_i \widetilde{\mathbf{Y}}_i^T \right] \\ &= -e^\tau Tr[P_i^{-2} \widetilde{\mathbf{Y}}_i \widetilde{\mathbf{Y}}_i^T] = -e^\tau \widetilde{\mathbf{Y}}_i^T P_i^{-2} \widetilde{\mathbf{Y}}_i, \quad \text{(by (37))}, \end{aligned} \tag{39}$$

and

$$\begin{aligned}\frac{\partial F^1}{\partial \zeta_k} &= \frac{\partial}{\partial \zeta_k} Tr[P_i^{-1}\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T] = -Tr\left[P_i^{-1}\left(\frac{\partial P_i}{\partial \zeta_k}\right)P_i^{-1}\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T\right] \\ &= -e^{\zeta_k}Tr[P_i^{-1}H_{ik}H_{ik}^TP_i^{-1}\widetilde{\mathbf{Y}}_i\widetilde{\mathbf{Y}}_i^T] = -e^{\zeta_k}(H_{ik}^TP_i^{-1}\widetilde{\mathbf{Y}}_i)^2, \quad \text{(by (38))}. \quad (40)\end{aligned}$$

Also,

$$\begin{aligned}\frac{\partial F^2}{\partial \tau} = \frac{\partial}{\partial \tau}\log|P_i| &= Tr\left[P_i^{-1}\left(\frac{\partial P_i}{\partial \tau}\right)\right] = e^{\tau}Tr(P_i^{-1}), \quad \text{(by (37))}, \quad (41) \\ \frac{\partial F^2}{\partial \zeta_k} = \frac{\partial}{\partial \zeta_k}\log|P_i| &= Tr\left[P_i^{-1}\left(\frac{\partial P_i}{\partial \zeta_k}\right)\right] = e^{\zeta_k}H_{ik}^TP_i^{-1}H_{ik}, \quad \text{(by (38))}. \quad (42)\end{aligned}$$

## Hessian of $F^1$

From (39),

$$\begin{aligned}\frac{\partial^2 F^1}{\partial \tau^2} &= \frac{\partial}{\partial \tau}\left[-e^{\tau}\widetilde{\mathbf{Y}}_i^TP_i^{-2}\widetilde{\mathbf{Y}}_i\right] \\ &= -e^{\tau}\widetilde{\mathbf{Y}}_iP_i^{-2}\widetilde{\mathbf{Y}}_i + e^{\tau}\widetilde{\mathbf{Y}}_i^TP_i^{-1}\left(\frac{\partial P_i}{\partial \tau}\right)P_i^{-2}\widetilde{\mathbf{Y}}_i + \widetilde{\mathbf{Y}}_i^TP_i^{-2}\left(\frac{\partial P_i}{\partial \tau}\right)P_i^{-1}\widetilde{\mathbf{Y}}_i \\ &= e^{\tau}\widetilde{\mathbf{Y}}_i^T[2e^{\tau}P_i^{-3} - P_i^{-2}]\widetilde{\mathbf{Y}}_i, \quad \text{(by (37))}. \quad (43)\end{aligned}$$

From (40),

$$\begin{aligned}\frac{\partial^2 F^1}{\partial \tau \partial \zeta_k} &= \frac{\partial}{\partial \tau}\left[-e^{\zeta_k}(H_{ik}^TP_i^{-1}\widetilde{\mathbf{Y}}_i)^2\right] \\ &= 2e^{\zeta_k}(\widetilde{\mathbf{Y}}_i^TP_i^{-1}H_{ik})\left[H_{ik}^TP_i^{-1}\left(\frac{\partial P_i}{\partial \tau}\right)P_i^{-1}\widetilde{\mathbf{Y}}_i\right] \\ &= 2e^{\zeta_k+\tau}\widetilde{\mathbf{Y}}_i^TP_i^{-1}H_{ik}H_{ik}^TP_i^{-2}\widetilde{\mathbf{Y}}_i, \quad \text{(by (37))}. \quad (44)\end{aligned}$$

Again using (40), and denoting by $\delta_{kl}$ the indicator of $\{k = l\}$,

$$\begin{aligned}
\frac{\partial^2 F^1}{\partial\zeta_l\partial\zeta_k} &= \frac{\partial}{\partial\zeta_l}\left[-e^{\zeta_k}(H_{ik}^T P_i^{-1}\widetilde{\mathbf{Y}}_i)^2\right] \\
&= -\delta_{kl}e^{\zeta_k}(H_{ik}^T P_i^{-1}\widetilde{\mathbf{Y}}_i)^2 + 2e^{\zeta_k}(\widetilde{\mathbf{Y}}_i^T P_i^{-1}H_{ik})\left[H_{ik}^T P_i^{-1}\left(\frac{\partial P_i}{\partial\zeta_l}\right)P_i^{-1}\widetilde{\mathbf{Y}}_i\right] \\
&= -\delta_{kl}e^{\zeta_k}(H_{ik}^T P_i^{-1}\widetilde{\mathbf{Y}}_i)^2 + 2e^{\zeta_k+\zeta_l}\widetilde{\mathbf{Y}}_i^T P_i^{-1}H_{ik}H_{ik}^T P_i^{-1}H_{il}H_{il}^T P_i^{-1}\widetilde{\mathbf{Y}}_i, \quad \text{(by (38))} \\
&= \begin{cases} 2e^{\zeta_k+\zeta_l}(H_{ik}^T P_i^{-1}\widetilde{\mathbf{Y}}_i)(H_{il}^T P_i^{-1}\widetilde{\mathbf{Y}}_i)(H_{ik}^T P_i^{-1}H_{il}) & \text{if } k \neq l \\ e^{\zeta_k}(H_{ik}^T P_i^{-1}\widetilde{\mathbf{Y}}_i)^2\left[2e^{\zeta_k}(H_{ik}^T P_i^{-1}H_{ik}) - 1\right] & \text{if } k = l. \end{cases}
\end{aligned} \tag{45}$$

## Hessian of $F^2$

From (41),

$$\frac{\partial^2 F^1}{\partial\tau^2} = \frac{\partial}{\partial\tau}\left[e^\tau Tr(P_i^{-1})\right] = e^\tau Tr(P_i^{-1}) - e^\tau Tr\left[P_i^{-1}\left(\frac{\partial P_i}{\partial\tau}\right)P_i^{-1}\right] = e^\tau[Tr(P_i^{-1}) - e^\tau Tr(P_i^{-2})]. \tag{46}$$

From (42),

$$\frac{\partial^2 F^2}{\partial\tau\partial\zeta_k} = \frac{\partial}{\partial\tau}[e^{\zeta_k}H_{ik}^T P_i^{-1}H_{ik}] = -e^{\zeta_k}H_{ik}^T P_i^{-1}\left(\frac{\partial P_i}{\partial\tau}\right)P_i^{-1}H_{ik} = -e^{\zeta_k+\tau}H_{ik}^T P_i^{-2}H_{ik}. \tag{47}$$

Finally,

$$\begin{aligned}
\frac{\partial^2 F^2}{\partial\zeta_l\zeta_k} &= \frac{\partial}{\partial\zeta_l}\left[e^{\zeta_k}H_{ik}^T P_i^{-1}H_{ik}\right] \\
&= \delta_{kl}e^{\zeta_k}H_{ik}^T P_i^{-1}H_{ik} - e^{\zeta_k}H_{ik}^T P_i^{-1}\left(\frac{\partial P_i}{\partial\zeta_l}\right)P_i^{-1}H_{ik} \\
&= \delta_{kl}e^{\zeta_k}H_{ik}^T P_i^{-1}H_{ik} - e^{\zeta_k+\zeta_l}(H_{ik}^T P_i^{-1}H_{il})^2, \quad \text{(by (38))} \\
&= \begin{cases} -e^{\zeta_k+\zeta_l}(H_{ik}^T P_i^{-1}H_{il})^2 & \text{if } k \neq l \\ e^{\zeta_k}H_{ik}P_i^{-1}H_{ik}[1 - e^{\zeta_k}H_{ik}^T P_i^{-1}H_{ik}] & \text{if } k = l. \end{cases}
\end{aligned} \tag{48}$$

# Tables

Table 2: `Easy`, $n = 200$, $\sigma^2 = 1/16$, Gaussian noise

| $\mathbf{M = 5}$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| `New.loc` | MISE/MSE | 0.060 | 0.214 | 0.217 | 1.36 | 1.07 | 3.75 | 88.23 |
| | (Sd) | (0.066) | (0.273) | (0.337) | | | | |
| | Reduction (%) | 81.8 | 64.3 | 65.4 | 73.8 | 65.8 | 36.0 | |
| `loc` | MISE/MSE | 0.329 | 0.599 | 0.628 | 5.19 | 3.13 | 5.86 | 261.82 |
| | (Sd) | (0.333) | (0.577) | (0.617) | | | | |
| `New.EM` | MISE/MSE | 0.058 | 0.191 | 0.169 | 1.27 | 1.04 | 1.65 | 0.96 |
| | (Sd) | (0.061) | (0.244) | (0.238) | | | | |
| | Reduction (%) | 14.7 | 26.3 | 29.5 | 2.3 | 2.8 | 5.7 | |
| `EM` | MISE/MSE | 0.068 | 0.259 | 0.240 | 1.30 | 1.07 | 1.75 | 1.17 |
| | (Sd) | (0.073) | (0.335) | (0.338) | | | | |
| `New.EM.ns` | MISE/MSE | 0.057 | 0.191 | 0.168 | 1.28 | 1.04 | 1.65 | 0.99 |
| | (Sd) | (0.062) | (0.244) | (0.238) | | | | |
| | Reduction (%) | 32.1 | 29.0 | 33.3 | 8.8 | -3.0 | 14.1 | |
| `EM.ns` | MISE/MSE | 0.084 | 0.269 | 0.252 | 1.40 | 1.01 | 1.92 | 2.83 |
| | (Sd) | (0.081) | (0.365) | (0.367) | | | | |
| $M = 4$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\sigma^2$ |
| `New.loc` | MISE/MSE | 0.352 | 0.379 | 0.543 | 9.51 | 0.91 | 4.62 | 348.92 |
| | (Sd) | (0.131) | (0.417) | (0.469) | | | | |
| | Reduction (%) | -10.7 | 34.3 | 13.3 | -118.6 | 75.7 | 23.6 | |
| `loc` | MISE/MSE | 0.318 | 0.577 | 0.626 | 4.35 | 3.75 | 6.05 | 238.40 |
| | (Sd) | (0.325) | (0.569) | (0.603) | | | | |
| `New.EM` | MISE/MSE | 0.355 | 0.388 | 0.492 | 9.39 | 0.89 | 1.53 | 98.66 |
| | (Sd) | (0.127) | (0.433) | (0.401) | | | | |
| | Reduction (%) | -22.4 | 16.0 | 3.3 | -97.7 | 36.9 | 33.5 | |
| `EM` | MISE/MSE | 0.290 | 0.462 | 0.509 | 4.75 | 1.41 | 2.30 | 100.61 |
| | (Sd) | (0.145) | (0.506) | (0.493) | | | | |
| `New.EM.ns` | MISE/MSE | 0.353 | 0.386 | 0.491 | 9.35 | 0.90 | 1.55 | 100.47 |
| | (Sd) | (0.126) | (0.434) | (0.402) | | | | |
| | Reduction (%) | -113.9 | -22.9 | -73.5 | -382.0 | 13.5 | 33.8 | |
| `EM.ns` | MISE/MSE | 0.165 | 0.314 | 0.283 | 1.94 | 1.04 | 2.34 | 9.49 |
| | (Sd) | (0.150) | (0.381) | (0.373) | | | | |
| $M = 9$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\sigma^2$ |
| `New.loc` | MISE/MSE | 0.092 | 0.247 | 0.239 | 1.30 | 1.74 | 3.21 | 745.76 |
| | (Sd) | (0.222) | (0.331) | (0.365) | | | | |
| | Reduction (%) | 71.2 | 58.7 | 61.3 | 73.3 | 48.8 | 32.0 | |
| `loc` | MISE/MSE | 0.319 | 0.598 | 0.617 | 4.87 | 3.40 | 4.72 | 226.58 |
| | (Sd) | (0.329) | (0.569) | (0.607) | | | | |
| `New.EM` | MISE/MSE | 0.065 | 0.192 | 0.169 | 1.20 | 1.00 | 1.66 | 0.77 |
| | (Sd) | (0.064) | (0.241) | (0.232) | | | | |
| | Reduction (%) | 20.7 | 25.0 | 29.0 | 8.4 | 2.0 | 10.3 | |
| `EM` | MISE/MSE | 0.082 | 0.256 | 0.238 | 1.31 | 1.02 | 1.85 | 2.81 |
| | (Sd) | (0.082) | (0.328) | (0.332) | | | | |
| `New.EM.ns` | MISE/MSE | 0.064 | 0.202 | 0.179 | 1.20 | 1.00 | 1.65 | 0.79 |
| | (Sd) | (0.064) | (0.257) | (0.250) | | | | |
| | Reduction (%) | 14.7 | 12.9 | 16.4 | 9.8 | 0.0 | -1.2 | |
| `EM.ns` | MISE/MSE | 0.075 | 0.232 | 0.214 | 1.33 | 1.00 | 1.63 | 2.93 |
| | (Sd) | (0.077) | (0.299) | (0.296) | | | | |

Table 3: `Practical`, $n = 500$, $\sigma^2 = 1/16$, Gaussian noise

| **M = 10** | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| `New.loc` | MISE/MSE | 0.035 | 0.195 | 0.463 | 0.556 | 0.343 | 0.69 | 0.54 | 0.62 | 0.54 | 0.68 |
| | (Sd) | (0.025) | (0.347) | (0.532) | (0.531) | (0.404) | | | | | |
| | Reduction (%) | 91.9 | 81.6 | 59.5 | 54.1 | 69.6 | 88.6 | 80.0 | 74.2 | 85.5 | 90.5 |
| `loc` | MISE/MSE | 0.434 | 1.059 | 1.143 | 1.211 | 1.127 | 6.04 | 2.70 | 2.40 | 3.73 | 7.19 |
| | (Sd) | (0.387) | (0.502) | (0.514) | (0.536) | (0.523) | | | | | |
| `New.EM` | MISE/MSE | 0.036 | 0.172 | 0.396 | 0.498 | 0.332 | 0.62 | 0.54 | 0.52 | 0.43 | 0.80 |
| | (Sd) | (0.031) | (0.288) | (0.432) | (0.509) | (0.420) | | | | | |
| | Reduction (%) | 33.3 | 25.5 | 31.5 | 20.4 | 17.2 | 20.5 | 0.0 | 5.5 | 18.9 | 27.3 |
| `EM` | MISE/MSE | 0.054 | 0.231 | 0.578 | 0.626 | 0.401 | 0.78 | 0.54 | 0.55 | 0.53 | 1.10 |
| | (Sd) | (0.046) | (0.263) | (0.469) | (0.517) | (0.463) | | | | | |
| `New.EM.ns` | MISE/MSE | 0.038 | 0.210 | 0.446 | 0.498 | 0.353 | 0.56 | 0.58 | 0.54 | 0.43 | 0.75 |
| | (Sd) | (0.035) | (0.336) | (0.437) | (0.476) | (0.439) | | | | | |
| | Reduction (%) | 96.9 | 85.3 | 66.6 | 50.2 | 77.8 | 73.3 | 58.6 | 15.6 | 27.1 | 88.5 |
| `EM.ns` | MISE/MSE | 1.226 | 1.432 | 1.336 | 1.000 | 1.593 | 2.10 | 1.40 | 0.64 | 0.59 | 6.51 |
| | (Sd) | (0.355) | (0.272) | (0.380) | (0.456) | (0.294) | | | | | |
| $M = 5$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
| `New.loc` | MISE/MSE | 1.612 | 1.799 | 1.386 | 1.835 | 1.524 | 2.07 | 6.37 | 12.65 | 69.84 | 92.64 |
| | (Sd) | (0.036) | (0.163) | (0.189) | (0.104) | (0.260) | | | | | |
| | Reduction (%) | -277.5 | -81.4 | -16.7 | -45.9 | -28.2 | 64.6 | -108 | -448 | -1738 | -1023 |
| `loc` | MISE/MSE | 0.427 | 0.992 | 1.186 | 1.258 | 1.189 | 5.84 | 3.06 | 2.31 | 3.80 | 8.25 |
| | (Sd) | (0.386) | (0.486) | (0.473) | (0.540) | (0.519) | | | | | |
| `New.EM` | MISE/MSE | 1.615 | 1.808 | 1.381 | 1.840 | 1.442 | 1.81 | 5.64 | 10.09 | 62.83 | 84.27 |
| | (Sd) | (0.041) | (0.144) | (0.190) | (0.119) | (0.204) | | | | | |
| | Reduction (%) | 0.2 | -0.4 | 0.8 | 0.1 | 1.1 | 8.6 | 8.9 | 2.9 | 1.3 | 1.8 |
| `EM` | MISE/MSE | 1.618 | 1.800 | 1.392 | 1.842 | 1.458 | 1.98 | 6.19 | 10.39 | 63.66 | 85.84 |
| | (Sd) | (0.042) | (0.152) | (0.198) | (0.119) | (0.203) | | | | | |
| `New.EM.ns` | MISE/MSE | 1.615 | 1.815 | 1.368 | 1.801 | 1.549 | 1.90 | 5.32 | 10.43 | 66.94 | 93.17 |
| | (Sd) | (0.045) | (0.137) | (0.191) | (0.127) | (0.242) | | | | | |
| | Reduction (%) | -9.7 | -0.2 | -0.7 | -1.1 | 5.1 | 35.2 | 13.5 | -2.5 | -6.8 | -14.1 |
| `EM.ns` | MISE/MSE | 1.472 | 1.811 | 1.358 | 1.781 | 1.633 | 2.93 | 6.15 | 10.18 | 62.65 | 81.63 |
| | (Sd) | ( 0.175) | (0.145) | (0.257) | (0.205) | (0.210) | | | | | |
| $M = 20$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
| `New.loc` | MISE/MSE | 0.029 | 0.204 | 0.790 | 0.732 | 0.629 | 0.65 | 0.70 | 1.67 | 0.58 | 0.27 |
| | (Sd) | (0.020) | (0.336) | (0.677) | (0.539) | (0.668) | | | | | |
| | Reduction (%) | 94.3 | 82.0 | 38.6 | 32.0 | 43.7 | 90.5 | 77.6 | 59.7 | 84.9 | 95.4 |
| `loc` | MISE/MSE | 0.508 | 1.134 | 1.292 | 1.077 | 1.117 | 6.86 | 3.12 | 4.14 | 3.85 | 5.91 |
| | (Sd) | (0.407) | (0.473) | (0.443) | (0.721) | (0.660) | | | | | |
| `New.EM` | MISE/MSE | 0.044 | 0.172 | 0.486 | 0.610 | 0.406 | 0.62 | 0.54 | 0.58 | 0.52 | 0.95 |
| | (Sd) | (0.042) | (0.251) | (0.481) | (0.542) | (0.439) | | | | | |
| | Reduction (%) | 38.9 | 34.1 | 37.8 | 26.8 | 29.6 | 23.5 | 6.9 | 25.6 | 24.6 | 34.0 |
| `EM` | MISE/MSE | 0.072 | 0.261 | 0.781 | 0.833 | 0.577 | 0.81 | 0.58 | 0.78 | 0.69 | 1.44 |
| | (Sd) | (0.075) | (0.229) | (0.531) | (0.553) | (0.539) | | | | | |
| `New.EM.ns` | MISE/MSE | 0.043 | 0.153 | 0.455 | 0.610 | 0.366 | 0.61 | 0.54 | 0.55 | 0.52 | 0.99 |
| | (Sd) | (0.043) | (0.170) | (0.459) | (0.567) | (0.396) | | | | | |
| | Reduction (%) | 49.4 | 31.7 | 28.7 | 13.7 | 12.0 | 30.7 | 11.5 | 24.7 | 18.8 | 46.8 |
| `EM.ns` | MISE/MSE | 0.085 | 0.224 | 0.638 | 0.707 | 0.416 | 0.88 | 0.61 | 0.73 | 0.64 | 1.86 |
| | (Sd) | (0.130) | (0.247) | (0.487) | (0.510) | (0.428) | | | | | |

Table 4: `Practical`, $n = 1000$, $\sigma^2 = 1/16$, Gaussian noise

| **M = 10** | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| `New.loc` | MISE/MSE | 0.015 | 0.067 | 0.169 | 0.228 | 0.146 | 0.26 | 0.34 | 0.25 | 0.35 | 0.41 |
| | (Sd) | (0.014) | (0.134) | (0.226) | (0.259) | (0.203) | | | | | |
| | Reduction (%) | 93.3 | 91.8 | 84.2 | 78.0 | 84.3 | 94.6 | 82.7 | 80.8 | 82.0 | 91.3 |
| `loc` | MISE/MSE | 0.224 | 0.813 | 1.069 | 1.035 | 0.930 | 4.85 | 1.96 | 1.30 | 1.94 | 4.72 |
| | (Sd) | (0.137) | (0.523) | (0.466) | (0.580) | (0.541) | | | | | |
| `New.EM` | MISE/MSE | 0.016 | 0.063 | 0.145 | 0.232 | 0.172 | 0.25 | 0.29 | 0.24 | 0.33 | 0.41 |
| | (Sd) | (0.014) | (0.122) | (0.193) | (0.337) | (0.307) | | | | | |
| | Reduction (%) | 51.5 | 56.3 | 53.1 | 37.3 | 31.2 | 53.7 | 9.4 | 27.3 | 17.5 | 43.1 |
| `EM` | MISE/MSE | 0.033 | 0.144 | 0.309 | 0.370 | 0.250 | 0.54 | 0.32 | 0.33 | 0.40 | 0.72 |
| | (Sd) | (0.039) | (0.204) | (0.330) | (0.416) | (0.359) | | | | | |
| `New.EM.ns` | MISE/MSE | 0.042 | 0.114 | 0.214 | 0.206 | 0.203 | 0.34 | 0.72 | 0.41 | 0.37 | 4.47 |
| | (Sd) | (0.170) | (0.265) | (0.341) | (0.223) | (0.399) | | | | | |
| | Reduction (%) | 96.8 | 92.1 | 84.4 | 79.3 | 87.4 | 80.9 | -7.5 | -64.0 | 30.2 | 23.6 |
| `EM.ns` | MISE/MSE | 1.302 | 1.434 | 1.374 | 0.994 | 1.610 | 1.78 | 0.67 | 0.25 | 0.53 | 5.85 |
| | (Sd) | (0.393) | (0.270) | (0.347) | (0.444) | (0.289) | | | | | |
| $M = 5$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
| `New.loc` | MISE/MSE | 1.613 | 1.888 | 1.393 | 1.825 | 1.483 | 1.02 | 4.69 | 10.74 | 70.95 | 89.69 |
| | (Sd) | (0.028) | (0.099) | (0.143) | (0.099) | (0.260) | | | | | |
| | Reduction (%) | -527.6 | -138.4 | -31.5 | -75.5 | -55.9 | 78.0 | -137 | -567 | -3279 | -1648 |
| `loc` | MISE/MSE | 0.257 | 0.792 | 1.059 | 1.040 | 0.951 | 4.63 | 1.98 | 1.61 | 2.10 | 5.13 |
| | (Sd) | (0.264) | (0.498) | (0.508) | (0.610) | (0.539) | | | | | |
| `New.EM` | MISE/MSE | 1.614 | 1.887 | 1.392 | 1.852 | 1.405 | 0.93 | 4.55 | 8.53 | 64.31 | 79.77 |
| | (Sd) | (0.021) | (0.068) | (0.201) | (0.251) | (0.135) | | | | | |
| | Reduction (%) | -0.7 | -0.9 | -0.2 | 0.2 | 1.1 | 21.2 | 10.6 | 0.7 | 0.0 | 1.1 |
| `EM` | MISE/MSE | 1.603 | 1.870 | 1.389 | 1.855 | 1.421 | 1.18 | 5.09 | 8.59 | 64.32 | 80.67 |
| | (Sd) | (0.030) | (0.142) | (0.394) | (0.393) | (0.180) | | | | | |
| `New.EM.ns` | MISE/MSE | 1.612 | 1.891 | 1.388 | 1.817 | 1.529 | 1.01 | 4.93 | 8.88 | 69.14 | 94.67 |
| | (Sd) | (0.028) | (0.085) | (0.136) | (0.093) | (0.198) | | | | | |
| | Reduction (%) | -10.4 | -1.7 | 0.0 | -4.3 | 3.7 | 52.6 | 15.7 | -0.8 | -7.6 | -13.8 |
| `EM.ns` | MISE/MSE | 1.460 | 1.859 | 1.388 | 1.742 | 1.588 | 2.13 | 5.85 | 8.81 | 64.28 | 83.20 |
| | (Sd) | (0.203) | (0.210) | (0.221) | (0.241) | (0.212) | | | | | |
| $M = 20$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
| `New.loc` | MISE/MSE | 0.017 | 0.057 | 0.180 | 0.273 | 0.154 | 0.19 | 0.20 | 0.14 | 0.34 | 0.30 |
| | (Sd) | (0.012) | (0.055) | (0.307) | (0.484) | (0.361) | | | | | |
| | Reduction (%) | 92.7 | 93.2 | 83.8 | 77.0 | 82.7 | 96.1 | 90.4 | 88.0 | 77.5 | 92.2 |
| `loc` | MISE/MSE | 0.232 | 0.843 | 1.111 | 1.185 | 0.890 | 4.88 | 2.09 | 1.17 | 1.51 | 3.85 |
| | (Sd) | (0.139) | (0.565) | (0.507) | (0.603) | (0.503) | | | | | |
| `New.EM` | MISE/MSE | 0.019 | 0.068 | 0.196 | 0.263 | 0.157 | 0.26 | 0.32 | 0.25 | 0.31 | 0.46 |
| | (Sd) | (0.016) | (0.104) | (0.285) | (0.360) | (0.253) | | | | | |
| | Reduction (%) | 48.6 | 60.5 | 64.3 | 54.4 | 44.9 | 46.9 | -6.7 | 49.0 | 29.5 | 20.7 |
| `EM` | MISE/MSE | 0.037 | 0.172 | 0.549 | 0.577 | 0.285 | 0.49 | 0.30 | 0.49 | 0.44 | 0.58 |
| | (Sd) | (0.029) | (0.189) | (0.489) | (0.502) | (0.316) | | | | | |
| `New.EM.ns` | MISE/MSE | 0.019 | 0.069 | 0.203 | 0.276 | 0.160 | 0.26 | 0.28 | 0.27 | 0.33 | 0.43 |
| | (Sd) | (0.015) | (0.104) | (0.288) | (0.371) | (0.256) | | | | | |
| | Reduction (%) | 52.5 | 50.0 | 46.3 | 35.5 | 27.9 | 48.0 | -7.7 | 37.2 | 19.5 | 48.2 |
| `EM.ns` | MISE/MSE | 0.040 | 0.138 | 0.378 | 0.428 | 0.222 | 0.50 | 0.26 | 0.43 | 0.41 | 0.83 |
| | (Sd) | (0.025) | (0.139) | (0.363) | (0.454) | (0.336) | | | | | |

Table 5: `Challenging`, $n = 500$, $\sigma^2 = 1/16$, Gaussian noise

| **M = 30** | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| `New.loc` | MISE/MSE | 0.162 | 0.196 | 0.135 | 2.67 | 2.06 | 2.27 | 2103.76 |
| | (Sd) | (0.254) | (0.330) | (0.285) | | | | |
| | Reduction (%) | 57.5 | 62.1 | 60.2 | 37.5 | 6.4 | 0.9 | |
| `loc` | MISE/MSE | 0.381 | 0.517 | 0.339 | 4.27 | 2.20 | 2.29 | 2469.33 |
| | (Sd) | (0.307) | (0.453) | (0.373) | | | | |
| `New.EM` | MISE/MSE | 0.124 | 0.130 | 0.082 | 1.43 | 0.53 | 0.67 | 20.82 |
| | (Sd) | (0.080) | (0.147) | (0.113) | | | | |
| | Reduction (%) | -0.8 | 16.7 | 22.6 | -3.6 | 7.0 | 23.0 | |
| `EM` | MISE/MSE | 0.123 | 0.156 | 0.106 | 1.38 | 0.57 | 0.87 | 37.47 |
| | (Sd) | (0.081) | (0.166) | (0.144) | | | | |
| `New.EM.ns` | MISE/MSE | 0.121 | 0.132 | 0.089 | 1.51 | 0.52 | 0.61 | 20.69 |
| | (Sd) | (0.060) | (0.164) | (0.138) | | | | |
| | Reduction (%) | 31.3 | 21.4 | 15.2 | 28.4 | 7.1 | 20.8 | |
| `EM.ns` | MISE/MSE | 0.176 | 0.168 | 0.105 | 2.11 | 0.56 | 0.77 | 86.72 |
| | (Sd) | (0.109) | (0.152) | (0.110) | | | | |
| $M = 20$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\sigma^2$ |
| `New.loc` | MISE/MSE | 0.315 | 0.244 | 0.149 | 3.01 | 2.13 | 3.17 | 1526.18 |
| | (Sd) | (0.261) | (0.288) | (0.317) | | | | |
| | Reduction (%) | 26.2 | 51.9 | 50.7 | 36.6 | -7.0 | -21.0 | |
| `loc` | MISE/MSE | 0.427 | 0.507 | 0.302 | 4.75 | 1.99 | 2.62 | 2281.71 |
| | (Sd) | (0.323) | (0.401) | (0.323) | | | | |
| `New.EM` | MISE/MSE | 0.286 | 0.215 | 0.106 | 2.39 | 0.73 | 0.64 | 196.92 |
| | (Sd) | (0.148) | (0.190) | (0.097) | | | | |
| | Reduction (%) | 1.0 | 7.3 | -8.2 | 14.9 | -19.7 | 23.8 | |
| `EM` | MISE/MSE | 0.289 | 0.232 | 0.098 | 2.81 | 0.61 | 0.84 | 262.13 |
| | (Sd) | (0.155) | (0.203) | (0.096) | | | | |
| `New.EM.ns` | MISE/MSE | 0.287 | 0.209 | 0.100 | 2.10 | 0.78 | 0.75 | 206.30 |
| | (Sd) | (0.148) | (0.173) | (0.071) | | | | |
| | Reduction (%) | 57.2 | 64.6 | 32.0 | 40.3 | 16.1 | 23.5 | |
| `EM.ns` | MISE/MSE | 0.670 | 0.590 | 0.147 | 3.52 | 0.93 | 0.98 | 769.97 |
| | (Sd) | (0.424) | (0.487) | (0.113) | | | | |
| $M = 35$ | | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\sigma^2$ |
| `New.loc` | MISE/MSE | 0.178 | 0.191 | 0.117 | 1.99 | 0.70 | 0.53 | 50.82 |
| | (Sd) | (0.124) | (0.235) | (0.182) | | | | |
| | Reduction (%) | 54.9 | 62.7 | 66.5 | 57.4 | 68.5 | 76.9 | |
| `loc` | MISE/MSE | 0.395 | 0.512 | 0.349 | 4.67 | 2.22 | 2.29 | 2724.98 |
| | (Sd) | (0.274) | (0.428) | (0.353) | | | | |
| `New.EM` | MISE/MSE | 0.178 | 0.189 | 0.114 | 1.85 | 0.54 | 0.63 | 47.87 |
| | (Sd) | (0.125) | (0.224) | (0.162) | | | | |
| | Reduction (%) | -3.5 | 11.3 | 18.0 | -7.6 | -3.8 | 24.1 | |
| `EM` | MISE/MSE | 0.172 | 0.213 | 0.139 | 1.72 | 0.52 | 0.83 | 72.12 |
| | (Sd) | (0.126) | (0.258) | (0.215) | | | | |
| `New.EM.ns` | MISE/MSE | 0.170 | 0.142 | 0.080 | 1.51 | 0.73 | 0.60 | 49.68 |
| | (Sd) | (0.107) | (0.154) | (0.095) | | | | |
| | Reduction (%) | -21.4 | 4.7 | 20.0 | -28.0 | -4.3 | 30.2 | |
| `EM.ns` | MISE/MSE | 0.140 | 0.149 | 0.100 | 1.18 | 0.70 | 0.86 | 61.67 |
| | (Sd) | (0.073) | (0.224) | (0.190) | | | | |

Table 6: Model selection : `Easy` ($n = 200$), `Practical` ($n = 500, 1000$), and `Challenging` ($n = 500$), $\sigma^2 = 1/16$, Gaussian noise

| Model | Method | Number of converged replicates | | | | Frequency of models selected | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
| | `New.loc` | 91 | 82 | 86 | 82 | 3 | 80 | 1 | 9 |
| | `New.EM` | 99 | 99 | 99 | 98 | 0 | 99 | 0 | 0 |
| `Easy` | `Combine`[1] | 100 | 100 | 100 | 100 | 0 | 100 | 0 | 0 |
| ($n = 200$) | `New.EM.ns` | 99 | 99 | 99 | 99 | 0 | 99 | 0 | 0 |
| | `Combine.ns`[2] | 99 | 99 | 99 | 99 | 0 | 99 | 0 | 0 |
| | | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
| | `New.loc` | 53 | 46 | 21 | 7 | 25 | 46 | 8 | 1 |
| | `New.EM` | 96 | 93 | 94 | 88 | 0 | 93 | 2 | 5 |
| `Practical` | `Combine` | 100 | 97 | 95 | 88 | 2 | 97 | 1 | 2 |
| ($n = 500$) | `New.EM.ns` | 72 | 60 | 78 | 86 | 3 | 60 | 17 | 20 |
| | `Combine.ns` | 90 | 82 | 81 | 87 | 2 | 82 | 7 | 9 |
| | | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
| | `New.loc` | 53 | 77 | 43 | 28 | 7 | 77 | 4 | 5 |
| | `New.EM` | 98 | 97 | 98 | 92 | 0 | 97 | 0 | 3 |
| `Practical` | `Combine` | 99 | 100 | 99 | 96 | 0 | 100 | 0 | 0 |
| ($n = 1000$) | `New.EM.ns` | 83 | 62 | 84 | 91 | 0 | 59 | 6 | 35 |
| | `Combine.ns` | 93 | 94 | 89 | 93 | 0 | 91 | 1 | 8 |
| | | $M = 20$ | $M = 25$ | $\mathbf{M = 30}$ | $M = 35$ | $M = 20$ | $M = 25$ | $\mathbf{M = 30}$ | $M = 35$ |
| | `New.loc` | 71 | 71 | 58 | 46 | 5 | 16 | 56 | 16 |
| | `New.EM` | 93 | 91 | 79 | 70 | 1 | 7 | 79 | 11 |
| `Challenging` | `Combine` | 95 | 96 | 88 | 79 | 1 | 4 | 88 | 6 |
| ($n = 500$) | `New.EM.ns` | 65 | 76 | 62 | 62 | 2 | 10 | 62 | 25 |
| | `Combine.ns` | 94 | 92 | 89 | 77 | 2 | 2 | 87 | 9 |

[1] combine the results of `New.loc` with `New.EM`: replace one by another if the first one fails to converge; [2] combine the results of `New.loc` with `New.EM.ns`.

Table 7: Model selection : `Hybrid`, $n = 500$, $\sigma^2 = 1/16$, Gaussian noise

| | `New.loc` | | | | `New.EM` | | | |
|---|---|---|---|---|---|---|---|---|
| **I. number of converged replicates (total=100)** | | | | | | | | |
| $r$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ |
| 2 | 75 | 63 | 60 | 54 | 100 | 98 | 99 | 85 |
| 3 | 81 | 78 | 60 | 61 | 99 | 99 | 97 | 96 |
| 4 | 36 | 11 | 3 | 1 | 99 | 94 | 94 | 93 |
| 5 | 6 | 0 | 0 | 0 | 95 | 85 | 60 | 38 |
| 6 | 1 | 0 | 0 | 0 | 61 | 60 | 29 | 8 |
| 7 | 4 | 0 | 0 | 0 | 7 | 22 | 7 | 3 |
| **II. frequencies of models selected** | | | | | | | | |
| $r$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ |
| 1 | 0 (1,1,1) | 0 (0,0,1) | 0 (0,0,0) | 0 (0,0,1) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 2 | 2 (4,6,6) | 2 (3,3,2) | 2 (2,2,2) | 2 (2,2,1) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 3 | 47 (61,63,79) | 2 (1,1,1) | 1 (1,1,1) | 0 (0,0,0) | 0 (0,0,98) | 0 (0,0,1) | 0 (0,0,1) | 0 (0,0,0) |
| 4 | 30 (20,16,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 2 (43,97,0) | 0 (0,1,0) | 0 (1,1,0) | 0 (0,0,0) |
| 5 | 3 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 56 (55,1,0) | 0 (1,0,0) | 1 (0,0,0) | 0 (0,0,0) |
| 6 | 1 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 33 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 7 | 3 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 7 (0,0,0) | 1 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |

Table 8: Mean running time for one replicate in seconds: `Easy` ($n = 200$), `Practical` ($n = 500$), $\sigma^2 = 1/16$, Gaussian noise

| Model | Method | **Mean running time (standard deviation)** * | | | |
|---|---|---|---|---|---|
| | | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
| `Easy` ($n = 200$) | `New.loc` | 14.3 (4.4) | 14.4 (4) | 15.4 (4.9) | 16.7 (5.2) |
| | `loc` | 19 (1.9) | 19 (1.9) | 19 (1.9) | 19 (1.9) |
| | `New.EM` | 14.7 (0.48) | 14.4 (0.52) | 14.8 (0.42) | 16.3 (0.48) |
| | `EM` | 9.8 (0.79) | 9.7 (0.48) | 10.1 (0.32) | 10.7 (1.1) |
| | | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
| `Practical` ($n = 500$) | `New.loc` | 63.8 (27.9) | 80.9 (45.1) | 87.4 (35.8) | 92.7 (31.2) |
| | `loc` | 28.4 (3.4) | 28.4 (3.4) | 28.4 (3.4) | 28.4 (3.4) |
| | `New.EM` | 60.2 (9.5) | 59.4 (3.1) | 70.6 (17.9) | 91.9 (30.2) |
| | `EM` | 54.1 (6.7) | 47.6 (2.2) | 53.7 (6.7) | 61.2 (11.9) |

$*$ for `New.loc` and `New.EM`, this means the additional computational cost after obtaining the initial estimates.

Table 9: `CD4 counts` data: estimated error variance and eigenvalues

| **Model :** $M = 10$, $r = 4$ | $\widehat{\sigma}^2$ | $\widehat{\lambda}_1$ | $\widehat{\lambda}_2$ | $\widehat{\lambda}_3$ | $\widehat{\lambda}_4$ |
|---|---|---|---|---|---|
| `loc` | 42,359.3 | 615,735.6 | 94,188.6 | 47,012.6 | 37,687.1 |
| `New.EM` | 38,411.0 | 473,416.8 | 208,201.4 | 53,253.9 | 24,582.0 |
| `EM` | 38,132.2 | 469,784.3 | 207,961.1 | 54,007.2 | 24,344.5 |

# Figures

Figure 1: True and estimated eigenfunctions for `easy` with $n = 200$, $\sigma^2 = 1/16$, Gaussian noise: true eigenfunctions (`Black`); Point-wise average of estimated eigenfunctions by `New.EM` (`Red`); Point-wise 0.95 and 0.05 quantiles of estimated eigenfunctions by `New.EM` (`Green`)

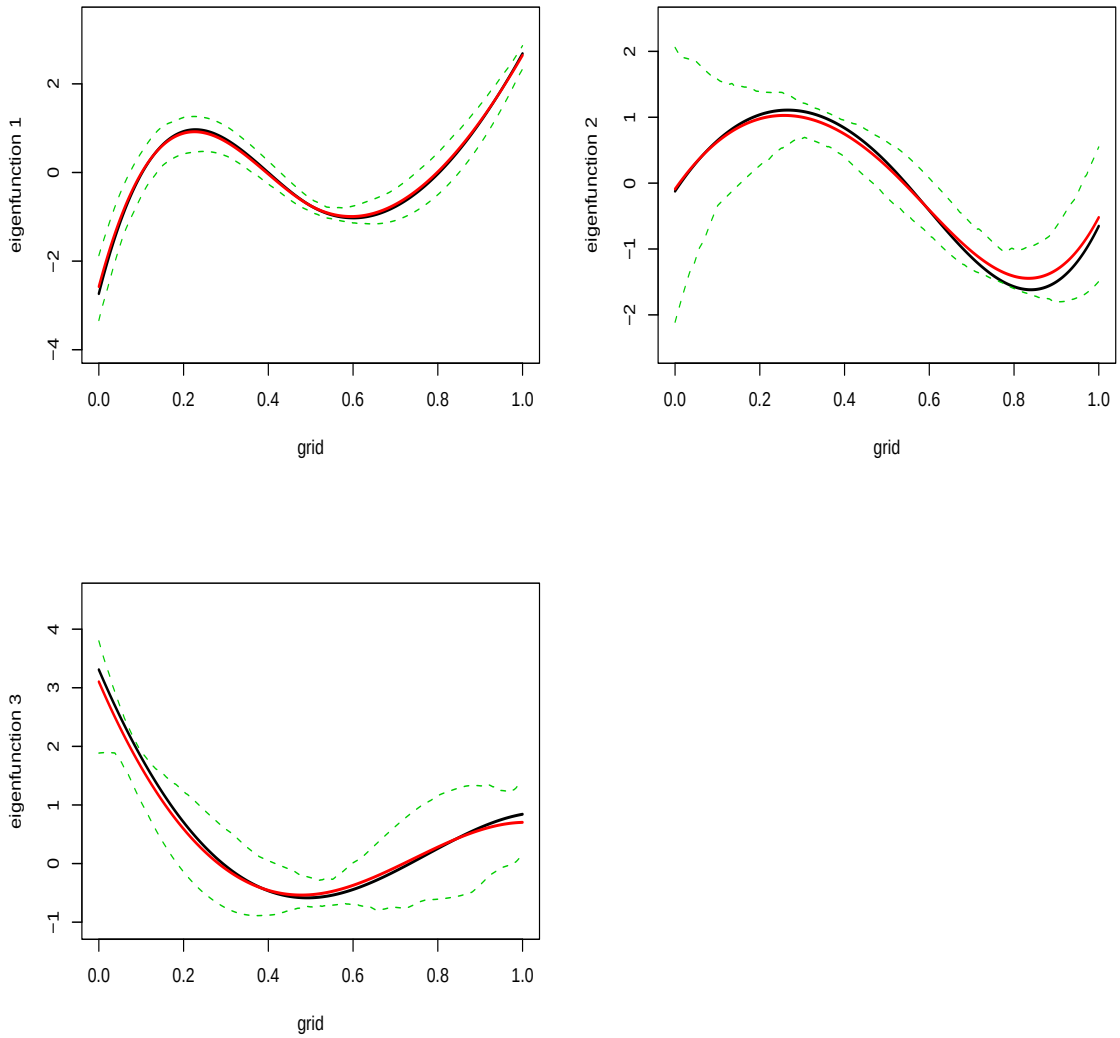


Figure 2: True and estimated eigenfunctions for `practical` with $n = 500$, $\sigma^2 = 1/16$, Gaussian noise: true eigenfunctions (`Black`); Point-wise average of estimated eigenfunctions by `New.EM` (`Red`); Point-wise 0.95 and 0.05 quantiles of estimated eigenfunctions by `New.EM` (`Green`)

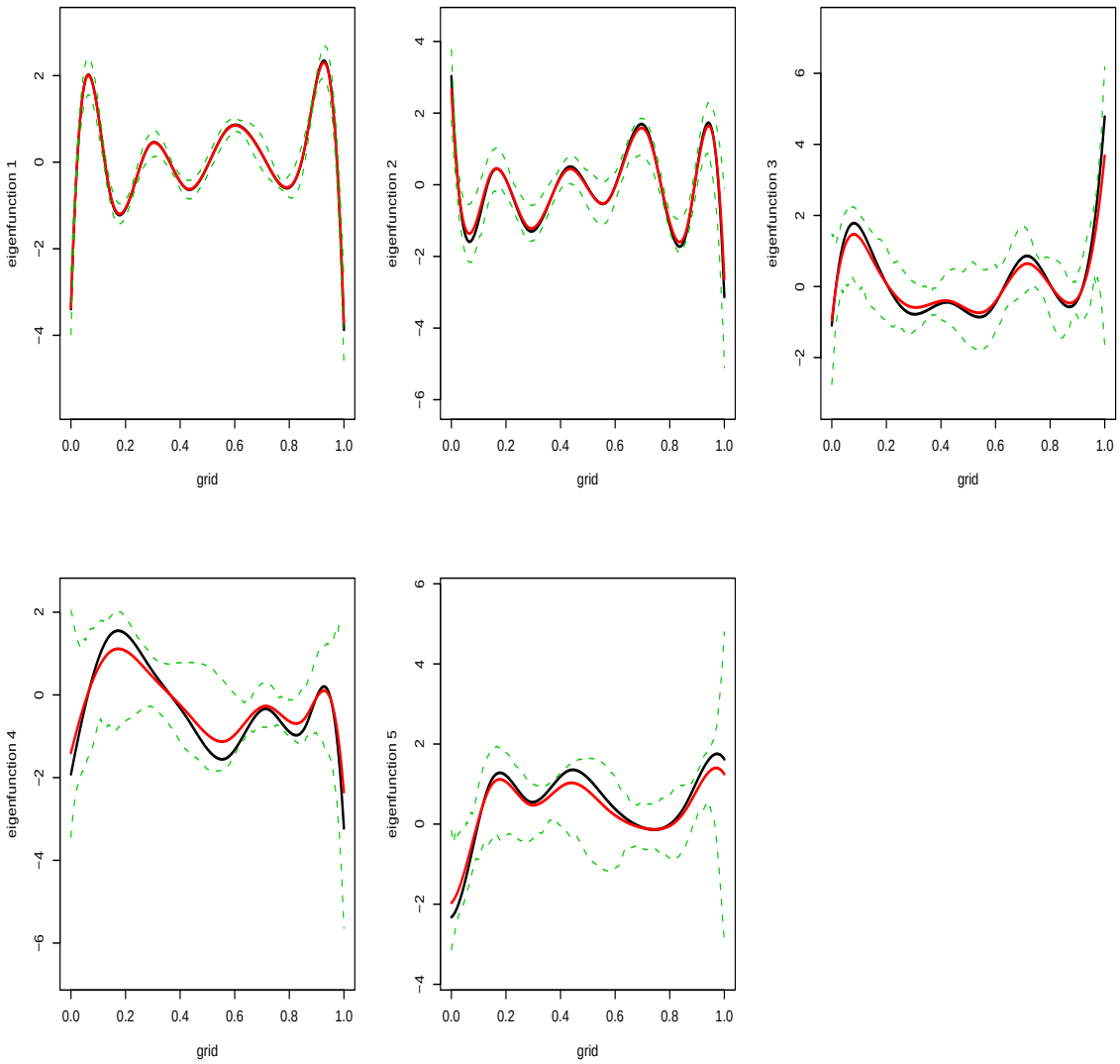


Figure 3: True and estimated eigenfunctions for `challenging` with $n = 500$, $\sigma^2 = 1/16$, Gaussian noise: true eigenfunctions (`Black`); Point-wise average of estimated eigenfunctions by `New.EM` (`Red`); Point-wise 0.95 and 0.05 quantiles of estimated eigenfunctions by `New.EM` (`Green`)

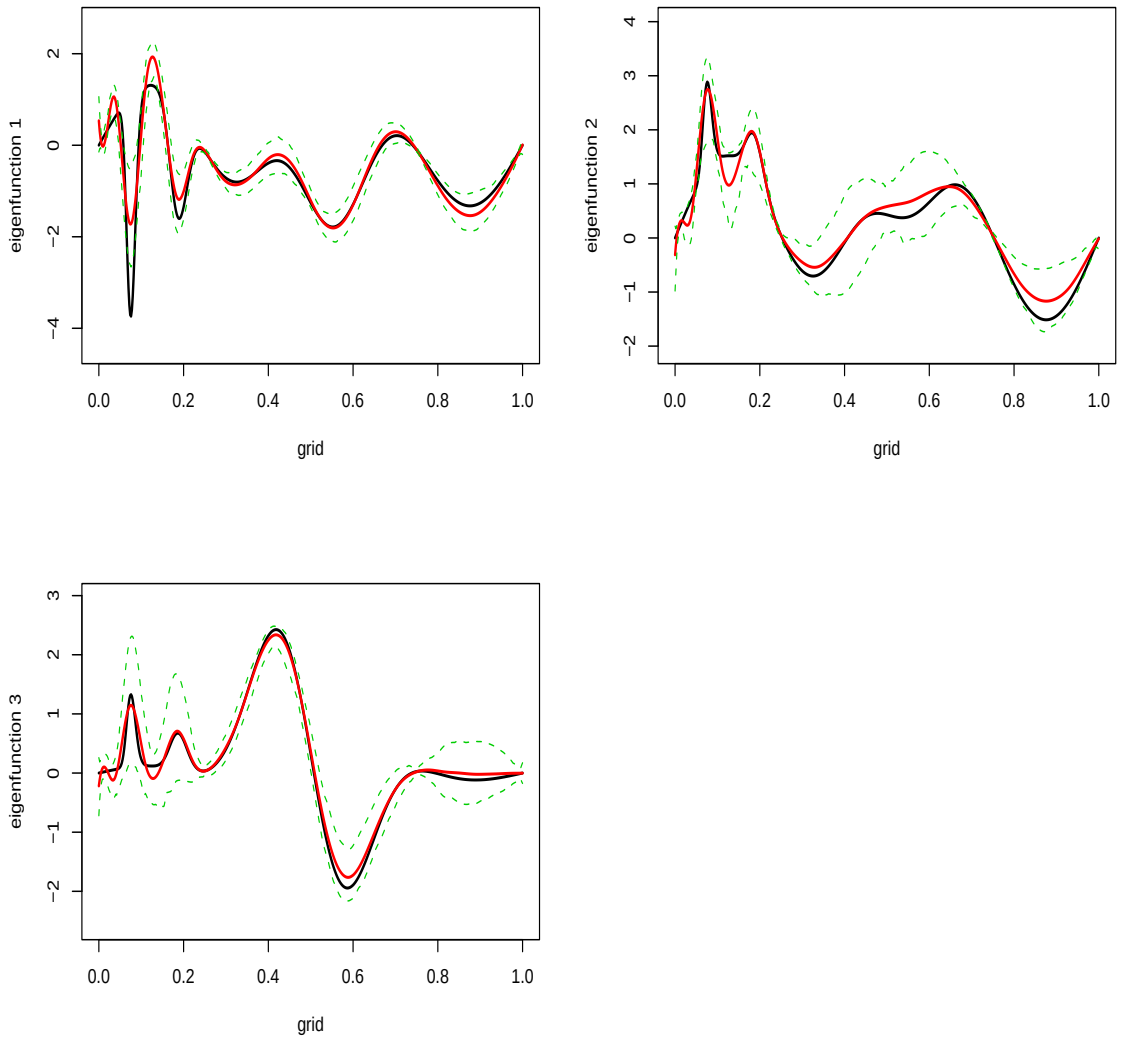
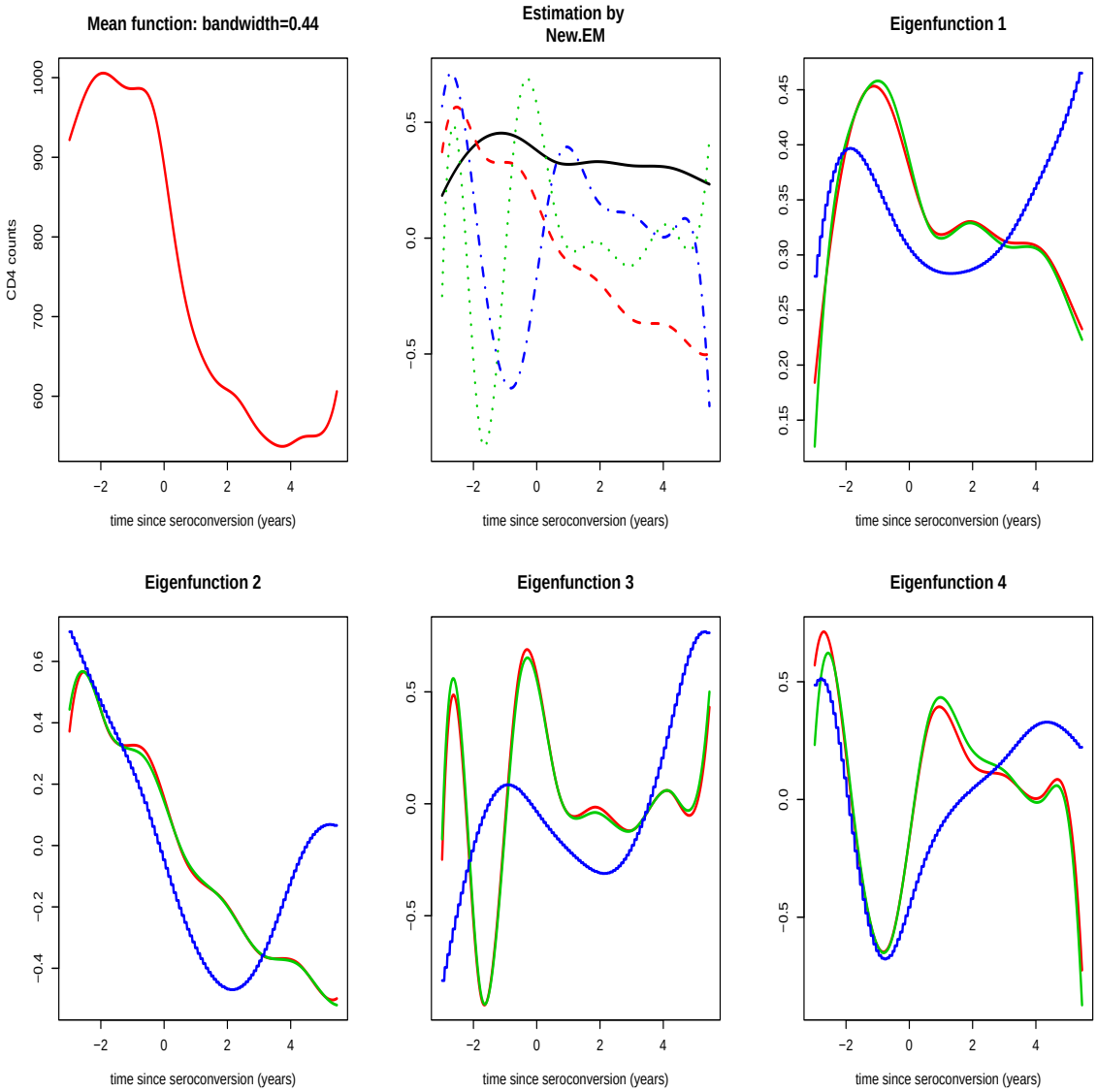


Figure 4: `CD4 counts` data: estimated mean and eigenfunctions. First panel: estimated mean function; Second panel: estimated eigenfunctions by `New.EM`: $\widehat{\psi}_1$=`Black`, $\widehat{\psi}_2$=`Red`, $\widehat{\psi}_3$=`Green`, $\widehat{\psi}_4$=`Blue`; Third to sixth panels: estimated eigenfunctions by `loc` (`Blue`), `New.EM` (`Red`), `EM` (`Green`)



# Supplementary material

# Tables

In the following tables, `New.loc`: `Newton` with `loc` as initial estimate; `New.EM`: `Newton` with `EM` (B-spline basis) as initial estimate; `New.EM.ns`: `Newton` with `EM.ns` (natural spline basis) as initial estimate; `Combine`: result after replacing `New.EM` with `New.loc` if the former fails to converge; `Combine.ns`: result after replacing `New.EM.ns` with `EM.loc` if the former fails to converge; in block (IV), number within parentheses after `loc` is the number of replicates with $\widehat{\sigma}^2 < 0$ by `loc`.

Table 10: `Easy`, $n = 100$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$.

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 91 | 91 | 89 | 72 |
| `New.EM` | 99 | 100 | 100 | 96 |
| `Combine` | 99 | 100 | 100 | 99 |
| `New.EM.ns` | 96 | 100 | 99 | 98 |
| `Combine.ns` | 98 | 100 | 100 | 100 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.397, 0.497, 0.623 | 0.113, 0.342, 0.323 | 0.128, 0.360, 0.375 | 0.143, 0.385, 0.359 |
| (sd) | 0.208, 0.495, 0.529 | 0.124, 0.439, 0.489 | 0.158, 0.410, 0.508 | 0.158, 0.442, 0.479 |
| (% reduction) | 18.1, 33.3, 14.4 | 76.9, 53.7, 55.9 | 74.1, 52.0, 49.7 | 72.3, 46.9, 47.3 |
| `loc` | 0.485, 0.745, 0.728 | 0.489, 0.739, 0.733 | 0.495, 0.750, 0.745 | 0.517, 0.725, 0.681 |
| (sd) | 0.436, 0.569, 0.539 | 0.454, 0.561, 0.537 | 0.442, 0.582, 0.550 | 0.460, 0.543, 0.541 |
| `New.EM` | 0.399, 0.509, 0.625 | 0.136, 0.367, 0.290 | 0.128, 0.374, 0.306 | 0.154, 0.392, 0.308 |
| (sd) | 0.217, 0.509, 0.520 | 0.212, 0.458, 0.418 | 0.206, 0.447, 0.411 | 0.216, 0.427, 0.383 |
| (% reduction) | -14.0, 16.1, 1.7 | 10.5, 20.7, 27.0 | 15.8, 17.4, 23.3 | 11.0, 16.8, 24.1 |
| `EM` | 0.350, 0.607, 0.636 | 0.152, 0.463, 0.397 | 0.152, 0.453, 0.399 | 0.173, 0.471, 0.406 |
| (sd) | 0.281, 0.578, 0.561 | 0.190, 0.501, 0.477 | 0.190, 0.478, 0.465 | 0.198, 0.498, 0.489 |
| `New.EM.ns` | 0.399, 0.499, 0.615 | 0.136, 0.367, 0.290 | 0.128, 0.377, 0.309 | 0.148, 0.389, 0.309 |
| (sd) | 0.217, 0.505, 0.521 | 0.212, 0.458, 0.418 | 0.207, 0.448, 0.412 | 0.215, 0.432, 0.396 |
| (% reduction) | -34.3, 7.4, -40.4 | 29.5, 24.8, 31.1 | 30.4, 20.8, 25.4 | 18.7, 24.3, 29.3 |
| `EM.ns` | 0.297, 0.539, 0.438 | 0.193, 0.488, 0.421 | 0.184, 0.476, 0.414 | 0.182, 0.514, 0.437 |
| (sd) | 0.336, 0.524, 0.471 | 0.251, 0.537, 0.509 | 0.245, 0.534, 0.503 | 0.211, 0.536, 0.528 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 13.11, 1.91, 4.58 | 2.18, 1.72, 5.60 | 3.05, 1.68, 7.75 | 2.18, 1.90, 5.65 |
| (% reduction) | -14.0, 73.1, 59.0 | 81.1, 75.1, 50.6 | 73.2, 77.3, 30.6 | 76.2, 74.3, 47.5 |
| `loc` | 11.50, 7.09, 11.18 | 11.52, 6.91, 11.34 | 11.40, 7.41, 11.17 | 9.15, 7.39, 10.77 |
| `New.EM` | 12.98, 1.87, 2.31 | 2.19, 1.66, 3.47 | 2.30,1.63, 3.63 | 2.07, 1.63, 3.63 |
| (% reduction) | -80.5, 29.2, 16.9 | -9.0, 1.2, -1.8 | -13.9, -1.2, 1.9 | -3.5, -10.1, 9.5 |
| `EM` | 7.19, 2.64, 2.78 | 2.01, 1.68, 3.41 | 2.02, 1.61, 3.70 | 2.00, 1.48, 4.01 |
| `New.EM.ns` | 13.18, 1.87, 2.32 | 2.19, 1.66, 3.47 | 2.32, 1.64, 3.61 | 2.07, 1.62, 3.63 |
| (% reduction) | -453.8, -21.4, 42.7 | -5.3, -4.4, 4.7 | -11.5, -5.8, 0.6 | 4.2, -9.5, 2.9 |
| `EM.ns` | 2.38, 1.54, 4.05 | 2.08, 1.59, 3.64 | 2.08, 1.55, 3.63 | 2.16, 1.48, 3.74 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 189.86 | 46.89 | 305.55 | 439.19 |
| `loc` (52) | 415.81 | 403.10 | 409.70 | 305.32 |
| `New.EM` | 113.02 | 3.98 | 4.27 | 2.51 |
| `EM` | 115.05 | 4.23 | 5.71 | 9.02 |
| `New.EM.ns` | 112.59 | 3.98 | 4.30 | 2.50 |
| `EM.ns` | 15.20 | 6.30 | 5.94 | 7.53 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 1 | 83 | 3 | 7 |
| `New.EM` | 0 | 91 | 1 | 8 |
| `Combine` | 0 | 91 | 1 | 8 |
| `New.EM.ns` | 0 | 91 | 1 | 8 |
| `Combine.ns` | 0 | 91 | 1 | 8 |

Table 11: `Easy`; $n = 200$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 91 | 82 | 86 | 82 |
| `New.EM` | 99 | 99 | 99 | 98 |
| `Combine` | 100 | 100 | 100 | 100 |
| `New.EM.ns` | 99 | 99 | 99 | 99 |
| `Combine.ns` | 99 | 99 | 99 | 99 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.352, 0.379, 0.543 | 0.060, 0.214, 0.217 | 0.057, 0.204, 0.199 | 0.092, 0.247, 0.239 |
| (sd) | 0.131, 0.417, 0.469 | 0.066, 0.273, 0.337 | 0.058, 0.262, 0.303 | 0.222, 0.331, 0.365 |
| (% reduction) | -10.7, 34.3, 13.3 | 81.8, 64.3, 65.4 | 82.6, 65.5, 67.7 | 71.2, 58.7, 61.3 |
| `loc` | 0.318, 0.577, 0.626 | 0.329, 0.599, 0.628 | 0.327, 0.591, 0.616 | 0.319, 0.598, 0.617 |
| (sd) | 0.325, 0.569, 0.603 | 0.333, 0.577, 0.617 | 0.329, 0.574, 0.597 | 0.329, 0.569, 0.607 |
| `New.EM` | 0.355, 0.388, 0.492 | 0.058, 0.191, 0.169 | 0.059, 0.195, 0.171 | 0.065, 0.192, 0.169 |
| (sd) | 0.127, 0.433, 0.401 | 0.061, 0.244, 0.238 | 0.057, 0.249, 0.241 | 0.064, 0.241, 0.232 |
| (% reduction) | -22.4, 16.0, 3.3 | 14.7, 26.3, 29.6 | 16.9, 24.7, 27.8 | 20.7, 25.0, 29.0 |
| `EM` | 0.290, 0.462, 0.509 | 0.068, 0.259, 0.240 | 0.071, 0.259, 0.237 | 0.082, 0.256, 0.238 |
| (sd) | 0.145, 0.506, 0.493 | 0.073, 0.335, 0.338 | 0.070, 0.333, 0.337 | 0.082, 0.328, 0.332 |
| `New.EM.ns` | 0.353, 0.386, 0.491 | 0.057, 0.191, 0.168 | 0.059, 0.195, 0.171 | 0.064, 0.202, 0.179 |
| (sd) | 0.126, 0.434, 0.402 | 0.062, 0.244, 0.238 | 0.057, 0.249, 0.241 | 0.064, 0.257, 0.250 |
| (% reduction) | -113.9, -22.9, -73.5 | 32.1, 29.0, 33.3 | 21.3, 25.9, 31.0 | 14.7, 12.9, 16.4 |
| `EM.ns` | 0.165, 0.314, 0.283 | 0.084, 0.269, 0.252 | 0.075, 0.263, 0.248 | 0.075, 0.232, 0.214 |
| (sd) | 0.150, 0.381, 0.373 | 0.081, 0.365, 0.367 | 0.076, 0.347 0.347 | 0.077, 0.299, 0.296 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 9.51, 0.91, 4.62 | 1.36, 1.07, 3.75 | 1.54, 1.06, 2.42 | 1.30, 1.74, 3.21 |
| (% reduction) | -118.6, 75.7, 23.6 | 73.8, 65.8, 36.0 | 69.1, 64.7, 58.5 | 73.3, 48.8, 32.0 |
| `loc` | 4.35, 3.75 , 6.05 | 5.19, 3.13, 5.86 | 4.98, 3.00, 5.83 | 4.87, 3.40, 4.72 |
| `New.EM` | 9.39, 0.89, 1.53 | 1.27, 1.04, 1.65 | 1.38, 1.03, 1.67 | 1.20, 1.00, 1.66 |
| (% reduction) | -97.7, 36.9, 33.5 | 2.3, 2.8, 5.7 | -4.5, 1.0, 2.9 | 8.4, 2.0, 10.3 |
| `EM` | 4.75, 1.41, 2.30 | 1.30, 1.07, 1.75 | 1.32, 1.04, 1.72 | 1.31, 1.02, 1.85 |
| `New.EM.ns` | 9.35, 0.90, 1.55 | 1.28, 1.04, 1.65 | 1.40, 1.03, 1.67 | 1.20, 1.00, 1.65 |
| (% reduction) | -382.0, 13.5, 33.8 | 8.6, -3.0, 14.1 | -4.5, -4.0, 3.5 | 9.8, 0.0, -1.2 |
| `EM.ns` | 1.94, 1.04, 2.34 | 1.40, 1.01, 1.92 | 1.34, 0.99, 1.73 | 1.33, 1.00, 1.63 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 348.92 | 88.23 | 22.29 | 745.76 |
| `loc` (56) | 238.40 | 261.82 | 203.19 | 226.58 |
| `New.EM` | 98.66 | 0.96 | 1.51 | 0.77 |
| `EM` | 100.61 | 1.17 | 2.38 | 2.81 |
| `New.EM.ns` | 100.47 | 0.99 | 1.57 | 0.79 |
| `EM.ns` | 9.49 | 2.83 | 2.55 | 2.93 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 3 | 80 | 1 | 9 |
| `New.EM` | 0 | 99 | 0 | 0 |
| `Combine` | 0 | 100 | 0 | 0 |
| `New.EM.ns` | 0 | 99 | 0 | 0 |
| `Combine.ns` | 0 | 99 | 0 | 0 |

Table 12: `Easy`; $n = 500$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 83 | 84 | 87 | 80 |
| `New.EM` | 99 | 100 | 100 | 100 |
| `Combine` | 100 | 100 | 100 | 100 |
| `New.EM.ns` | 97 | 100 | 100 | 100 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.311, 0.358, 0.460 | 0.027, 0.067, 0.089 | 0.024, 0.076, 0.098 | 0.020, 0.069, 0.062 |
| (sd) | 0.097, 0.416, 0.454 | 0.067, 0.088, 0.247 | 0.029, 0.111, 0.255 | 0.021, 0.073, 0.071 |
| (% reduction) | -106.0, 11.6, -10.8 | 82.5, 82.8, 77.8 | 84.4, 81.0, 76.2 | 87.4, 81.3, 84.3 |
| `loc` | 0.151, 0.405, 0.415 | 0.154, 0.389, 0.400 | 0.154, 0.399, 0.411 | 0.159, 0.369, 0.396 |
| (sd) | 0.124, 0.472, 0.415 | 0.129, 0.461, 0.403 | 0.128, 0.463, 0.406 | 0.132, 0.404, 0.395 |
| `New.EM` | 0.318, 0.316, 0.397 | 0.017, 0.070, 0.067 | 0.021, 0.071, 0.067 | 0.019, 0.072, 0.068 |
| (sd) | 0.079, 0.365, 0.352 | 0.017, 0.104, 0.103 | 0.016, 0.106, 0.105 | 0.017, 0.111, 0.110 |
| (% reduction) | -45.9, 12.9, -6.1 | 22.7, 36.4, 38.0 | 16.0, 38.8, 41.7 | 20.8, 37.9, 41.4 |
| `EM` | 0.218, 0.363, 0.374 | 0.022, 0.110, 0.108 | 0.025, 0.116, 0.115 | 0.024, 0.116, 0.116 |
| (sd) | 0.083, 0.442, 0.427 | 0.022, 0.157, 0.160 | 0.021, 0.175, 0.178 | 0.024, 0.224, 0.223 |
| `New.EM.ns` | 0.307, 0.371, 0.438 | 0.018, 0.067, 0.060 | 0.020, 0.065, 0.060 | 0.019, 0.067, 0.060 |
| (sd) | 0.085, 0.417, 0.398 | 0.019, 0.099, 0.098 | 0.018, 0.076, 0.075 | 0.020, 0.071, 0.069 |
| (% reduction) | -338.6, -126.2, -184.4 | 43.8, 48.9, 52.0 | 25.9, 48.8, 50.8 | 24.0, 36.2, 38.8 |
| `EM.ns` | 0.070, 0.164, 0.154 | 0.032, 0.131, 0.125 | 0.027, 0.127, 0.122 | 0.025, 0.105, 0.098 |
| (sd) | 0.046, 0.197, 0.199 | 0.029, 0.166, 0.166 | 0.028, 0.149, 0.148 | 0.030, 0.125, 0.122 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 6.18, 0.69, 4.16 | 0.39 0.51, 2.09 | 0.50, 0.62, 2.40 | 0.41, 0.47, 0.61 |
| (% reduction) | -77.6, 35.5, -60.0 | 85.5, 54.1, 17.4 | 85.0, 41.5, 5.5 | 88.3, 56.1, 73.4 |
| `loc` | 3.48, 1.07, 2.60 | 2.69, 1.11, 2.53 | 3.33, 1.06, 2.54 | 3.51, 1.07, 2.29 |
| `New.EM` | 6.46, 0.53, 1.40 | 0.49, 0.43, 0.70 | 0.62, 0.42, 0.71 | 0.49, 0.42, 0.68 |
| (% reduction) | -231.3, 68.1, 39.9 | 9.3, 2.3, 5.4 | -5.1, 4.5, -1.4 | 15.5, 8.7, 8.1 |
| `EM` | 1.95, 1.66, 2.33 | 0.54, 0.44, 0.74 | 0.59, 0.44, 0.70 | 0.58, 0.46, 0.74 |
| `New.EM.ns` | 6.64, 0.65, 1.85 | 0.39, 0.48, 0.55 | 0.50, 0.48, 0.62 | 0.42, 0.46, 0.59 |
| (% reduction) | -361.1, -35.4, -172.1 | 35.0, 2.0, 16.7 | 3.8, 2.0, 6.1 | 16.0, 6.1, 13.2 |
| `EM.ns` | 1.44, 0.48, 0.68 | 0.60, 0.49, 0.66 | 0.52, 0.49, 0.66 | 0.50, 0.49, 0.68 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 231.06 | 71.81 | 74.39 | 0.24 |
| `loc` (53) | 104.34 | 98.10 | 96.41 | 93.67 |
| `New.EM` | 89.29 | 0.18 | 0.46 | 0.16 |
| `EM` | 91.96 | 0.26 | 0.78 | 0.51 |
| `New.EM.ns` | 89.91 | 0.25 | 0.57 | 0.23 |
| `EM.ns` | 8.17 | 1.21 | 0.69 | 0.56 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 0 | 80 | 1 | 7 |
| `New.EM` | 0 | 99 | 0 | 1 |
| `Combine` | 0 | 99 | 0 | 1 |
| `New.EM.ns` | 0 | 96 | 0 | 4 |
| `Combine.ns` | 0 | 96 | 0 | 4 |

Table 13: `Easy`; $n = 200$; $\sigma^2 = 1/8$; noise distribution : $N(0, 1)$

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 86 | 86 | 86 | 85 |
| `New.EM` | 99 | 99 | 99 | 99 |
| `Combine` | 100 | 100 | 100 | 100 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.345, 0.369, 0.463 | 0.064, 0.186, 0.158 | 0.065, 0.177, 0.170 | 0.074, 0.181, 0.153 |
| (sd) | 0.185, 0.414, 0.400 | 0.069, 0.221, 0.207 | 0.064, 0.201, 0.254 | 0.075, 0.204, 0.186 |
| (% reduction) | -7.8, 34.9, 23.1 | 81.2, 69.0, 73.8 | 78.8, 68.3, 72.4 | 78.1, 67.9, 73.2 |
| `loc` | 0.320, 0.567, 0.602 | 0.340, 0.600, 0.604 | 0.307, 0.559, 0.616 | 0.338, 0.564, 0.570 |
| (sd) | 0.353, 0.580, 0.596 | 0.360, 0.598, 0.599 | 0.306, 0.570, 0.607 | 0.362, 0.570, 0.570 |
| `New.EM` | 0.339, 0.387, 0.481 | 0.063, 0.200, 0.177 | 0.065, 0.203, 0.178 | 0.073, 0.213, 0.190 |
| (sd) | 0.137, 0.423, 0.401 | 0.066, 0.244, 0.236 | 0.063, 0.243, 0.234 | 0.072, 0.255, 0.246 |
| (% reduction) | -18.1, 10.2, -0.6 | 13.7, 16.7, 19.5 | 12.2, 13.2, 16.0 | 15.1, 11.6, 14.8 |
| `EM` | 0.287, 0.431, 0.478 | 0.073, 0.240, 0.220 | 0.074, 0.234, 0.212 | 0.086, 0.241, 0.223 |
| (sd) | 0.150, 0.477, 0.465 | 0.076, 0.290, 0.291 | 0.073, 0.285, 0.285 | 0.086, 0.287, 0.288 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 7.47, 0.94, 2.71 | 1.37, 1.19, 2.02 | 1.53, 1.10, 2.90 | 1.26, 1.10, 2.00 |
| (% reduction) | -58.6, 71.3, 54.3 | 72.3, 66.6, 68.8 | 68.2, 65.5, 50.4 | 74.9, 65.8, 65.0 |
| `loc` | 4.71, 3.28, 5.93 | 4.94, 3.56, 6.47 | 4.81, 3.19, 5.85 | 5.02, 3.22, 5.71 |
| `New.EM` | 7.21, 0.93, 1.44 | 1.36, 1.12, 1.88 | 1.45, 1.11, 1.91 | 1.27, 1.08, 1.82 |
| (% reduction) | -78.5, 25.0, 31.4 | 0.0, 3.4, 15.7 | -5.1, 2.6, 14.3 | 5.9, 4.4, 20.5 |
| `EM` | 4.04, 1.24, 2.10 | 1.36, 1.16, 2.23 | 1.38, 1.14, 2.23 | 1.35, 1.13, 2.29 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 72.11 | 0.53 | 5.70 | 0.49 |
| `loc` (39) | 51.49 | 51.10 | 56.40 | 57.63 |
| `New.EM` | 25.78 | 0.51 | 0.65 | 0.47 |
| `EM` | 26.37 | 0.89 | 1.17 | 1.21 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 2 | 85 | 3 | 1 |
| `New.EM` | 0 | 98 | 1 | 0 |
| `Combine` | 0 | 99 | 1 | 0 |

Table 14: `Easy`; $n = 100$; $\sigma^2 = 1/16$; noise distribution : $\frac{1}{\sqrt{2}}t_4$

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 85 | 79 | 79 | 66 |
| `New.EM` | 100 | 100 | 99 | 99 |
| `Combine` | 100 | 100 | 100 | 99 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.366, 0.613, 0.682 | 0.120, 0.393, 0.404 | 0.129, 0.360, 0.370 | 0.165, 0.400, 0.343 |
| (sd) | 0.200, 0.494, 0.494 | 0.173, 0.449, 0.519 | 0.233, 0.413, 0.494 | 0.252, 0.434, 0.438 |
| (% reduction) | 30.9, 21.5, 9.4 | 75.8, 48.9, 47.3 | 75.1, 52.0, 48.9 | 69.2, 48.9, 56.2 |
| `loc` | 0.530, 0.781, 0.753 | 0.495, 0.769, 0.766 | 0.519, 0.750, 0.724 | 0.536, 0.783, 0.783 |
| (sd) | 0.488, 0.571, 0.588 | 0.458, 0.576, 0.609 | 0.488, 0.553, 0.567 | 0.510, 0.564, 0.590 |
| `New.EM` | 0.359, 0.593, 0.645 | 0.114, 0.399, 0.344 | 0.113, 0.373, 0.326 | 0.120, 0.404, 0.352 |
| (sd) | 0.186, 0.511, 0.480 | 0.154, 0.452, 0.447 | 0.141, 0.416, 0.405 | 0.160, 0.464, 0.442 |
| (% reduction) | -10.8, 10.0, -0.5 | 12.3, 23.1, 24.4 | 22.1, 27.4, 27.6 | 34.1, 25.0, 22.6 |
| `EM` | 0.324, 0.659, 0.642 | 0.130, 0.519, 0.455 | 0.145, 0.514, 0.450 | 0.182, 0.539, 0.455 |
| (sd) | 0.251, 0.530, 0.499 | 0.191, 0.496, 0.482 | 0.213, 0.489, 0.480 | 0.256, 0.517, 0.472 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 12.78, 3.06, 5.37 | 1.80, 2.13, 7.55 | 2.69, 2.90, 6.86 | 2.69, 2.31, 3.08 |
| (% reduction) | -52.0, 56.3, 44.7 | 74.8, 70.0, 35.2 | 67.6, 50.3, 14.7 | 70.1, 62.2, 58.3 |
| `loc` | 8.41, 7.01, 9.71 | 7.15, 7.10, 11.65 | 8.31, 5.83, 8.04 | 9.00, 6.11, 7.39 |
| `New.EM` | 13.29, 2.75, 3.07 | 2.20, 2.14, 4.07 | 2.87, 2.16, 4.00 | 2.51, 2.12, 4.13 |
| (% reduction) | -87.7, 20.5, 22.1 | -8.9, 1.4, 5.1 | -21.6, 2.7, 4.8 | -26.1, 12.0, 3.3 |
| `EM` | 7.08, 3.46, 3.94 | 2.02, 2.17, 4.29 | 2.36, 2.22, 4.20 | 1.99, 2.41, 4.27 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 221.41 | 206.37 | 110.76 | 81.95 |
| `loc` (57) | 244.37 | 254.89 | 242.56 | 250.48 |
| `New.EM` | 99.66 | 3.92 | 4.31 | 3.25 |
| `EM` | 99.05 | 3.91 | 5.29 | 8.91 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 3 | 72 | 6 | 8 |
| `New.EM` | 0 | 92 | 3 | 5 |
| `Combine` | 0 | 92 | 3 | 5 |

Table 15: `Easy`; $n = 200$; $\sigma^2 = 1/16$; noise distribution : $\frac{1}{\sqrt{2}}t_4$

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 91 | 90 | 88 | 81 |
| `New.EM` | 98 | 98 | 99 | 99 |
| `Combine` | 99 | 100 | 100 | 99 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.343, 0.472, 0.587 | 0.047, 0.189, 0.173 | 0.062, 0.195, 0.183 | 0.054, 0.183, 0.163 |
| (sd) | 0.136, 0.431, 0.457 | 0.046, 0.261, 0.264 | 0.112, 0.290, 0.302 | 0.054, 0.225, 0.224 |
| (% reduction) | 12.1, 33.0, 3.0 | 87.9, 72.1, 70.2 | 84.2, 71.5, 68.8 | 86.8, 73.0, 70.6 |
| `loc` | 0.390, 0.704, 0.605 | 0.387, 0.677, 0.581 | 0.393, 0.685, 0.586 | 0.409, 0.677, 0.555 |
| (sd) | 0.418, 0.572, 0.540 | 0.414, 0.535, 0.513 | 0.416, 0.538, 0.518 | 0.425, 0.520, 0.491 |
| `New.EM` | 0.348, 0.473, 0.544 | 0.052, 0.187, 0.168 | 0.053, 0.192, 0.177 | 0.055, 0.187, 0.167 |
| (sd) | 0.128, 0.453, 0.397 | 0.055, 0.259, 0.264 | 0.051, 0.282, 0.290 | 0.057, 0.253, 0.255 |
| (% reduction) | -23.8, 10.9, 0.4 | 17.5, 35.3, 36.6 | 20.9, 34.9, 34.7 | 26.7, 38.3, 39.1 |
| `EM` | 0.281, 0.531, 0.546 | 0.063, 0.289, 0.265 | 0.067, 0.295, 0.271 | 0.075, 0.303, 0.274 |
| (sd) | 0.165, 0.466, 0.446 | 0.064, 0.361, 0.357 | 0.072, 0.378, 0.374 | 0.085, 0.353, 0.350 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 9.77, 1.54, 6.20 | 1.02, 1.17, 1.65 | 1.27, 1.22, 1.96 | 1.13, 1.19, 1.68 |
| (% reduction) | -108.8, 64.3, 24.4 | 78.0, 72.1, 79.1 | 69.5, 53.6, 58.6 | 74.8, 55.3, 67.0 |
| `loc` | 4.68, 4.31, 8.20 | 4.64, 4.20, 7.88 | 4.17, 2.63, 4.73 | 4.49, 2.66, 5.09 |
| `New.EM` | 10.23, 1.56, 1.91 | 0.98, 1.28, 1.54 | 4.17, 2.63, 4.73 | 1.04, 1.28, 1.54 |
| (% reduction) | -137.4, 43.1, 35.0 | 2.0, 8.6, 14.9 | -289.7, -93.4, -135.3 | -1.0, 0.8, 23.8 |
| `EM` | 4.31, 2.74, 2.94 | 1.00, 1.40, 1.81 | 1.07, 1.36, 2.01 | 1.03, 1.29, 2.02 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 513.98 | 2.79 | 80.16 | 2.62 |
| `loc` (47) | 240.66 | 231.44 | 248.53 | 226.98 |
| `New.EM` | 94.38 | 2.69 | 3.11 | 2.48 |
| `EM` | 95.70 | 2.62 | 3.60 | 4.42 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 2 | 88 | 2 | 2 |
| `New.EM` | 0 | 96 | 0 | 3 |
| `Combine` | 0 | 98 | 0 | 2 |

Table 16: `Easy`; $n = 100$; $\sigma^2 = 1/16$; noise distribution : $\text{Exp}(1) - 1$

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 83 | 84 | 84 | 67 |
| `New.EM` | 97 | 100 | 100 | 99 |
| `Combine` | 98 | 100 | 100 | 100 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.403, 0.548, 0.652 | 0.128, 0.358, 0.340 | 0.136, 0.369, 0.412 | 0.150, 0.377, 0.320 |
| (sd) | 0.264, 0.485, 0.500 | 0.133, 0.436, 0.455 | 0.154, 0.423, 0.532 | 0.197, 0.408, 0.405 |
| (% reduction) | 24.8, 19.1, 0.0 | 76.3, 46.4, 47.9 | 74.0, 43.5, 38.3 | 72.9, 44.5, 50.2 |
| `loc` | 0.536, 0.677, 0.652 | 0.541, 0.668, 0.652 | 0.524, 0.653, 0.668 | 0.553, 0.679, 0.642 |
| (sd) | 0.488, 0.513, 0.432 | 0.500, 0.501, 0.442 | 0.487, 0.498, 0.444 | 0.506, 0.528, 0.452 |
| `New.EM` | 0.387, 0.537, 0.606 | 0.147, 0.377, 0.318 | 0.132, 0.348, 0.303 | 0.144, 0.361, 0.305 |
| (sd) | 0.217, 0.467, 0.457 | 0.196, 0.442, 0.412 | 0.150, 0.405, 0.382 | 0.180, 0.411, 0.384 |
| (% reduction) | -13.8, 9.3, 5.2 | 26.9, 18.0, 14.5 | 33.3, 23.0, 19.6 | 33.6, 16.2, 14.1 |
| `EM` | 0.340, 0.592, 0.639 | 0.201, 0.460, 0.372 | 0.198, 0.452, 0.377 | 0.217, 0.431, 0.355 |
| (sd) | 0.241, 0.541, 0.510 | 0.272, 0.477, 0.437 | 0.260, 0.464, 0.431 | 0.287, 0.435, 0.392 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 13.44, 3.50, 5.30 | 2.35, 1.74, 4.30 | 3.43, 2.05, 8.67 | 2.59, 2.15, 4.68 |
| (% reduction) | -59.6, 29.1, 54.1 | 75.6, 65.5, 61.3 | 66.2, 59.5, 18.5 | 71.9, 48.2, 49.9 |
| `loc` | 8.42, 4.94, 11.54 | 9.63, 5.05, 11.12 | 10.15, 5.06, 10.64 | 9.22, 4.15, 9.35 |
| `New.EM` | 14.77, 3.19, 2.57 | 2.27, 1.59, 3.37 | 2.59, 1.66, 3.62 | 2.26, 1.74, 3.59 |
| (% reduction) | -86.3, 20.4, 23.7 | -10.7, 1.9, 3.4 | -18.8, 1.2, 1.6 | -17.1, 3.3, 6.5 |
| `EM` | 7.93, 4.01, 3.37 | 2.05, 1.62, 3.49 | 2.18, 1.68, 3.68 | 1.93, 1.80, 3.84 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 177.13 | 29.61 | 339.90 | 211.38 |
| `loc` (54) | 927.89 | 994.62 | 919.17 | 815.63 |
| `New.EM` | 105.24 | 3.67 | 4.00 | 2.74 |
| `EM` | 105.76 | 4.31 | 6.41 | 7.88 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 2 | 77 | 5 | 6 |
| `New.EM` | 0 | 90 | 5 | 5 |
| `Combine` | 0 | 90 | 5 | 5 |

Table 17: `Easy`; $n = 200$; $\sigma^2 = 1/16$; noise distribution : $\text{Exp}(1) - 1$

| | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 90 | 91 | 87 | 79 |
| `New.EM` | 99 | 100 | 100 | 99 |
| `Combine` | 100 | 100 | 100 | 100 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `New.loc` | 0.337, 0.463, 0.527 | 0.061, 0.248, 0.221 | 0.057, 0.208, 0.184 | 0.073, 0.233, 0.191 |
| (sd) | 0.124, 0.471, 0.443 | 0.060, 0.335, 0.349 | 0.055, 0.274, 0.271 | 0.111, 0.298, 0.275 |
| (% reduction) | 8.4, 25.0, 4.7 | 83.5, 61.2, 61.7 | 84.6, 64.3, 64.3 | 80.3, 59.7, 62.6 |
| `loc` | 0.368, 0.617, 0.553 | 0.369, 0.639, 0.577 | 0.369, 0.583, 0.516 | 0.371, 0.578, 0.511 |
| (sd) | 0.369, 0.533, 0.489 | 0.367, 0.548, 0.511 | 0.369, 0.491, 0.442 | 0.383, 0.508, 0.465 |
| `New.EM` | 0.347, 0.452, 0.525 | 0.071, 0.253, 0.217 | 0.065, 0.240, 0.210 | 0.074, 0.252, 0.214 |
| (sd) | 0.131, 0.445, 0.411 | 0.131, 0.321, 0.303 | 0.100, 0.312, 0.300 | 0.122, 0.306, 0.297 |
| (% reduction) | -19.7, 12.9, 3.5 | 17.4, 24.3, 26.4 | 21.7, 30.2, 31.4 | 26.7, 28.0, 27.7 |
| `EM` | 0.290, 0.519, 0.544 | 0.086, 0.334, 0.295 | 0.083, 0.344, 0.306 | 0.101, 0.350, 0.296 |
| (sd) | 0.155, 0.457, 0.450 | 0.147, 0.414, 0.396 | 0.135, 0.436, 0.420 | 0.154, 0.434, 0.409 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `New.loc` | 11.04, 1.79, 2.89 | 1.34, 1.13, 2.03 | 1.53, 0.91, 1.67 | 2.19, 1.31, 1.82 |
| (% reduction) | -35.8, 37.4, 56.3 | 83.7, 62.7, 70.3 | 82.0, 69.7, 74.3 | 61.4, 49.8, 70.8 |
| `loc` | 8.13, 2.86, 6.61 | 8.21, 3.03, 6.83 | 8.50, 3.00, 6.50 | 5.68, 2.61, 6.23 |
| `New.EM` | 11.38, 1.70, 1.65 | 1.40, 0.95, 1.41 | 1.50, 0.96, 1.60 | 1.41, 1.00, 1.52 |
| (% reduction) | -128.1, 43.5, 46.1 | -12.9, 11.2, 12.4 | -21.0, 11.1, 7.0 | -9.3, 10.7, 15.6 |
| `EM` | 4.99, 3.01, 3.06 | 1.24, 1.07, 1.61 | 1.24, 1.08, 1.72 | 1.29, 1.12, 1.80 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 207.35 | 85.51 | 2.11 | 67.43 |
| `loc`(56) | 219.92 | 219.92 | 235.18 | 249.00 |
| `New.EM` | 97.11 | 1.78 | 2.19 | 1.58 |
| `EM` | 98.76 | 1.92 | 2.87 | 3.10 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 0 | 90 | 1 | 3 |
| `New.EM` | 0 | 99 | 1 | 0 |
| `Combine` | 0 | 99 | 1 | 0 |

Table 18: `Practical`; $n = 300$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 5$ | **M = 10** | $M = 15$ | $M = 20$ |
|---|---|---|---|---|
| | | **I. number of converged replicates** | | |
| New.loc | 60 | 17 | 4 | 0 |
| New.EM | 96 | 98 | 87 | 73 |
| Combine | 99 | 98 | 89 | 73 |
| New.EM.ns | 81 | 48 | 67 | 80 |
| | | **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | |
| New.loc | 1.634, 1.764, 1.400, 1.825, 1.533 | 0.133, 0.298, 0.604, 0.710, 0.613 | 0.075, 0.537, 0.782, 1.017, 0.790 | * * * * * |
| (sd) | 0.060, 0.175, 0.233, 0.107, 0.261 | 0.314, 0.415, 0.626, 0.616, 0.632 | 0.026, 0.743, 0.634, 0.551, 0.718 | * * * * * |
| (% reduction) | -104.8, -34.6, -11.1, -29.9, -9.5 | 83.6, 75.2, 45.8, 46.7, 50.2 | 82.2, 49.1, -5.4, 42.6, 53.8 | * * * * * |
| loc | 0.798, 1.311, 1.260, 1.405, 1.400 | 0.810, 1.203, 1.115, 1.333, 1.231 | 0.421, 1.054, 0.742, 1.772, 1.711 | * * * * * |
| (sd) | 0.538, 0.486, 0.456, 0.454, 0.521 | 0.540, 0.505, 0.457, 0.408, 0.541 | 0.204, 0.295, 0.323, 0.209, 0.319 | * * * * * |
| New.EM | 1.628, 1.778, 1.395, 1.809, 1.489 | 0.066, 0.279, 0.568, 0.688, 0.483 | 0.085, 0.247, 0.615, 0.711, 0.522 | 0.091, 0.288, 0.636, 0.766, 0.581 |
| (sd) | 0.055, 0.168, 0.225, 0.136, 0.227 | 0.057, 0.308, 0.474, 0.514, 0.465 | 0.081, 0.208, 0.492, 0.551, 0.517 | 0.100, 0.313, 0.520, 0.572, 0.538 |
| (% reduction) | 0.1, -0.3, 0.5, 0.1, 0.9 | 28.3, 21.8, 23.2, 13.8, 13.3 | 15.0, 32.0, 29.5, 14.4, 7.1 | 26.6, 25.8, 29.9, 19.4, 14.4 |
| EM | 1.630, 1.772, 1.402, 1.810, 1.503 | 0.092, 0.357, 0.740, 0.798, 0.557 | 0.100, 0.363, 0.872, 0.831, 0.562 | 0.124, 0.388, 0.907, 0.950, 0.679 |
| (sd) | 0.056, 0.172, 0.231, 0.136, 0.229 | 0.098 , 0.313, 0.519, 0.560, 0.504 | 0.106, 0.342, 0.555, 0.591, 0.513 | 0.137, 0.344, 0.564, 0.517, 0.532 |
| New.EM.ns | 1.628, 1.765, 1.396, 1.798, 1.563 | 0.069, 0.253, 0.567, 0.708, 0.472 | 0.087, 0.270, 0.632, 0.737, 0.564 | 0.085, 0.277, 0.638, 0.755, 0.553 |
| (sd) | 0.058, 0.165, 0.226, 0.131, 0.243 | 0.061, 0.336, 0.506, 0.525, 0.408 | 0.083, 0.298, 0.505, 0.572, 0.574 | 0.089, 0.226, 0.490, 0.561, 0.517 |
| (% reduction) | -8.6, 0.9, -2.3, -0.7, 1.1 | 94.6, 82.7, 56.0, 33.8, 70.1 | 42.0, 39.6, 25.3, 22.1, 22.6 | 26.7, 31.6, 29.8, 14.6, 7.5 |
| EM.ns | 1.499, 1.781, 1.365, 1.786, 1.581 | 1.267, 1.459, 1.289, 1.070, 1.579 | 0.150, 0.447, 0.846, 0.946, 0.729 | 0.116, 0.405, 0.909, 0.884, 0.598 |
| (sd) | 0.133, 0.159, 0.281, 0.162, 0.255 | 0.372, 0.277, 0.400, 0.490, 0.332 | 0.121, 0.359, 0.512, 0.596, 0.628 | 0.106, 0.383, 0.541, 0.587, 0.512 |
| | | **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | |
| New.loc | 3.71, 5.74, 15.44, 68.11, 92.74 | 5.85, 0.92, 1.50, 0.75, 7.22 | 1.47, 0.85, 0.15, 1.32, 0.50 | * * * * * |
| (% reduction) | 55.5, -33.5, -305.2, -1051, -765.9 | 23.0, 85.0, 44.9, 83.7, 38.4 | 86.0, 82.7, 95.0, 38.3, 94.1 | * * * * * |
| loc | 8.34, 4.30, 3.81, 5.92, 10.71 | 7.60, 6.13, 2.72, 4.59, 11.73 | 10.51, 4.90, 3.03, 2.14, 8.45 | * * * * * |
| New.EM | 2.99, 6.01, 12.32, 60.87, 86.52 | 1.12, 0.88, 0.82, 0.68, 1.51 | 1.33, 0.94, 1.00, 0.78, 1.99 | 1.33, 0.83, 0.83, 0.84, 2.18 |
| (% reduction) | -2.0, 7.4, 3.2, 1.6, 1.8 | 4.3, 9.3, -18.8, 17.1, 23.0 | -6.4, -4.4, -16.3, -1.3, 31.6 | -4.7, 19.4, 0.0, -2.4, 26.1 |
| EM | 2.93, 6.49, 12.73, 61.85, 88.07 | 1.17, 0.97, 0.69, 0.82, 1.96 | 1.25, 0.90, 0.86, 0.77, 2.91 | 1.27, 1.03, 0.83, 0.82, 2.95 |
| New.EM.ns | 2.97, 6.05, 12.16, 62.79, 93.92 | 0.83, 0.89, 0.75, 0.81, 1.21 | 1.38, 1.09, 1.03, 0.85, 2.20 | 1.35, 0.90, 1.03, 0.90, 2.13 |
| (% reduction) | -6.5, 6.9, 1.9, -3.6, -13.0 | 64.4, 47.6, 40.5, 40.9, 85.0 | 8.6, 6.8, -25.6, 23.4, 10.9 | -7.1, 9.1, -24.1, 15.1, 17.4 |
| EM.ns | 2.79, 6.50, 12.39, 60.63, 83.12 | 2.33, 1.70, 1.26, 1.37, 8.09 | 1.51, 1.17, 0.82, 1.11, 2.47 | 1.26, 0.99, 0.83, 1.06, 2.58 |
| | | **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | |
| New.loc | 39780.24 | 42.54 | 0.25 | * |
| loc (32) | 1522.77 | 793.55 | 4548.91 | * |
| New.EM | 35980.84 | 1.14 | 0.66 | 1.09 |
| EM | 36297.68 | 1.65 | 1.98 | 2.39 |
| New.EM.ns | 36688.90 | 0.78 | 0.72 | 0.99 |
| EM.ns | 33822.99 | 715.94 | 5.59 | 3.23 |
| | | **V. number of times model with $M$ basis functions selected** | | |
| New.loc | 50 | 17 | 2 | 0 |
| New.EM | 0 | 97 | 3 | 0 |
| Combine | 0 | 97 | 3 | 0 |
| New.EM.ns | 5 | 48 | 28 | 17 |
| Combine.ns | 5 | 55 | 23 | 15 |

Table 19: `Practical`; $n = 500$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| New.loc | 53 | 46 | 21 | 7 |
| New.EM | 96 | 93 | 94 | 88 |
| Combine | 100 | 97 | 95 | 88 |
| New.EM.ns | 72 | 60 | 78 | 86 |
| Combine.ns | 90 | 82 | 81 | 87 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| New.loc | 1.612, 1.799, 1.386, 1.835, 1.524 | 0.035, 0.195, 0.463, 0.556, 0.343 | 0.047, 0.169, 0.317, 0.670, 0.552 | 0.029, 0.204, 0.790, 0.732, 0.629 |
| (sd) | 0.036, 0.163, 0.189, 0.104, 0.260 | 0.025, 0.347, 0.532, 0.531, 0.404 | 0.070, 0.262, 0.382, 0.646, 0.620 | 0.020, 0.336, 0.677, 0.539, 0.668 |
| (% reduction) | -277.5, -81.4, -16.9, -45.9, -28.2 | 91.9, 81.6, 59.5, 54.1, 69.6 | 87.1, 83.8, 72.3, 48.7, 51.9 | 94.3, 82.0, 38.9, 32.0, 43.7 |
| loc | 0.427, 0.992, 1.186, 1.258, 1.189 | 0.434, 1.059, 1.143, 1.211, 1.127 | 0.363, 1.044, 1.145, 1.305, 1.148 | 0.508, 1.134, 1.292, 1.077, 1.117 |
| (sd) | 0.386, 0.486, 0.473, 0.540, 0.519 | 0.387, 0.502, 0.514, 0.536, 0.523 | 0.261, 0.535, 0.496, 0.510, 0.587 | 0.407, 0.473, 0.443, 0.721, 0.660 |
| New.EM | 1.615, 1.808, 1.381, 1.840, 1.442 | 0.036, 0.172, 0.396, 0.498, 0.332 | 0.044, 0.154, 0.407, 0.559, 0.387 | 0.044, 0.172, 0.486, 0.610, 0.406 |
| (sd) | 0.041, 0.144, 0.190, 0.119, 0.204 | 0.031, 0.288, 0.432, 0.509, 0.420 | 0.044, 0.241, 0.445, 0.553, 0.473 | 0.042, 0.251, 0.481, 0.542, 0.439 |
| (% reduction) | 0.2, -0.4, 0.8, 0.1, 1.1 | 33.3, 25.5, 31.5, 20.4, 17.2 | 27.9, 32.5, 37.6, 18.5, 2.0 | 38.9, 34.1, 37.8, 26.8, 29.6 |
| EM | 1.618, 1.800, 1.392, 1.842, 1.458 | 0.054, 0.231, 0.578, 0.626, 0.401 | 0.061, 0.228, 0.652, 0.686, 0.395 | 0.072, 0.261, 0.781, 0.833, 0.577 |
| (sd) | 0.042, 0.152, 0.198, 0.119, 0.203 | 0.046, 0.263, 0.469, 0.517, 0.463 | 0.067, 0.263, 0.530, 0.551, 0.444 | 0.075, 0.229, 0.531, 0.553, 0.539 |
| New.EM.ns | 1.615, 1.815, 1.368, 1.801, 1.549 | 0.038, 0.210, 0.446, 0.498, 0.353 | 0.041, 0.140, 0.412, 0.592, 0.400 | 0.043, 0.153, 0.455, 0.610, 0.366 |
| (sd) | 0.045, 0.137, 0.191, 0.127, 0.242 | 0.035, 0.336, 0.437, 0.476, 0.439 | 0.033, 0.170, 0.421, 0.563, 0.488 | 0.043, 0.170, 0.459, 0.567, 0.396 |
| (% reduction) | -9.7, -0.2, -0.7, -1.1, 5.1 | 96.9, 85.3, 66.6, 50.2, 77.8 | 57.3, 53.5, 41.9, 22.9, 28.7 | 49.4, 31.7, 28.7, 13.7, 12.0 |
| EM.ns | 1.472, 1.811, 1.358, 1.781, 1.633 | 1.226, 1.432, 1.336, 1.000, 1.593 | 0.096, 0.301, 0.709, 0.768, 0.561 | 0.085, 0.224, 0.638, 0.707, 0.416 |
| (sd) | 0.175, 0.145, 0.257, 0.205, 0.210 | 0.355, 0.272, 0.380, 0.456, 0.294 | 0.062, 0.234, 0.522, 0.561, 0.554 | 0.130, 0.247, 0.487, 0.510, 0.428 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| New.loc | 2.07, 6.37, 12.65, 69.84, 92.64 | 0.69, 0.54, 0.62, 0.54, 0.68 | 0.86, 0.54, 0.71, 0.60, 0.77 | 0.65, 0.70, 1.67, 0.58, 0.27 |
| (% reduction) | 64.6, -108.2, -447.6, -1738, -1023 | 88.6, 80.0, 74.2, 85.5, 90.5 | 88.4, 78.2, 70.4, 80.5, 90.6 | 90.5, 77.6, 59.7, 84.9, 95.4 |
| loc | 5.84, 3.06, 2.31, 3.80, 8.25 | 6.04, 2.70, 2.40, 3.73, 7.19 | 7.42, 2.48, 2.40, 3.07, 8.19 | 6.86, 3.12, 4.14, 3.85, 5.91 |
| New.EM | 1.81, 5.64, 10.09, 62.83, 84.27 | 0.62, 0.54, 0.52, 0.43, 0.80 | 0.67, 0.59, 0.58, 0.53, 0.92 | 0.62, 0.54, 0.58, 0.52, 0.95 |
| (% reduction) | 8.6, 8.9, 2.9, 1.3, 1.8 | 20.5, 0.0, 5.5, 18.9, 27.3 | 20.2, -20.4, 15.9, 8.6, 32.8 | 23.5, 6.9, 25.6, 24.6, 34.0 |
| EM | 1.98, 6.19, 10.39, 63.66, 85.84 | 0.78, 0.54, 0.55, 0.53, 1.10 | 0.84, 0.49, 0.69, 0.58, 1.37 | 0.81, 0.58, 0.78, 0.69, 1.44 |
| New.EM.ns | 1.90, 5.32, 10.43, 66.94, 93.17 | 0.56, 0.58, 0.54, 0.43, 0.75 | 0.71, 0.58, 0.58, 0.48, 0.91 | 0.61, 0.54, 0.55, 0.52, 0.99 |
| (% reduction) | 35.2, 13.5, -2.5, -6.8, -14.1 | 73.3, 58.6, 15.6, 27.1, 88.5 | 30.4, 9.4, 32.6, 27.3, 18.8 | 30.7, 11.5, 24.7, 18.8, 46.8 |
| EM.ns | 2.93, 6.15, 10.18, 62.65, 81.63 | 2.10, 1.40, 0.64, 0.59, 6.51 | 1.02, 0.64, 0.86, 0.66, 1.12 | 0.88, 0.61, 0.73, 0.64, 1.86 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| New.loc | 39653.64 | 0.39 | 0.36 | 0.73 |
| loc (28) | 499.17 | 1345.32 | 1600.02 | 1872.78 |
| New.EM | 36335.23 | 0.33 | 0.31 | 0.46 |
| EM | 36646.27 | 0.60 | 1.43 | 1.34 |
| New.EM.ns | 37713.38 | 0.33 | 0.33 | 0.46 |
| EM.ns | 34339.84 | 856.30 | 4.24 | 3.70 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| New.loc | 25 | 46 | 8 | 1 |
| New.EM | 0 | 93 | 2 | 5 |
| Combine | 2 | 97 | 1 | 2 |
| New.EM.ns | 3 | 60 | 17 | 20 |
| Combine | 2 | 82 | 7 | 9 |

Table 20: `Practical`; $n = 500$; $\sigma^2 = 1/8$; noise distribution : $N(0, 1)$

| | $M = 5$ | **M = 10** | $M = 15$ | $M = 20$ |
|---|---|---|---|---|
| | | **I. number of converged replicates** | | |
| `New.loc` | 51 | 51 | 26 | 6 |
| `New.EM` | 96 | 97 | 98 | 91 |
| `Combine` | 96 | 98 | 99 | 92 |
| | | **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | |
| `New.loc` | 1.612, 1.782, 1.383, 1.807, 1.537 | 0.037, 0.183, 0.495, 0.595, 0.340 | 0.101, 0.189, 0.581, 0.793, 0.537 | 0.057, 0.301, 0.579, 0.711, 0.439 |
| (sd) | 0.039, 0.172, 0.190, 0.115, 0.281 | 0.034, 0.341, 0.506, 0.559, 0.416 | 0.300, 0.261, 0.562, 0.643, 0.542 | 0.049, 0.306, 0.648, 0.678, 0.311 |
| (% reduction) | -241.5, -69.4, -19.5, -49.1, -36.9 | 92.6, 83.9, 56.7, 53.0, 69.0 | 79.7, 82.6, 48.3, 32.0, 52.9 | 82.0, 69.2, 45.4, 47.7, 62.9 |
| `loc` | 0.472, 1.052, 1.157, 1.212, 1.123 | 0.497, 1.134, 1.144, 1.266, 1.098 | 0.498, 1.084, 1.123, 1.166, 1.141 | 0.317, 0.977, 1.061, 1.359, 1.183 |
| (sd) | 0.406, 0.485, 0.459, 0.517, 0.525 | 0.437, 0.540, 0.514, 0.537, 0.532 | 0.340, 0.562, 0.489, 0.497, 0.502 | 0.174, 0.433, 0.596, 0.677, 0.635 |
| `New.EM` | 1.615, 1.800, 1.384, 1.840, 1.445 | 0.040, 0.177, 0.454, 0.580, 0.385 | 0.049, 0.169, 0.529, 0.644, 0.443 | 0.050, 0.180, 0.518, 0.637, 0.420 |
| (sd) | 0.041, 0.152, 0.190, 0.122, 0.208 | 0.035, 0.294, 0.466, 0.549, 0.454 | 0.046, 0.212, 0.512, 0.576, 0.505 | 0.048, 0.234, 0.492, 0.570, 0.424 |
| (% reduction) | 0.2, -0.4, 0.6, 0.1, 1.2 | 27.3, 22.0, 26.7, 13.3, 5.9 | 21.0, 26.8, 23.7, 12.1, -4.7 | 31.5, 28.3, 34.8, 24.2, 29.3 |
| `EM` | 1.618, 1.793, 1.393, 1.842, 1.463 | 0.055, 0.227, 0.619, 0.669, 0.409 | 0.062, 0.231, 0.693, 0.733, 0.423 | 0.073, 0.251, 0.794, 0.840, 0.594 |
| (sd) | 0.042, 0.156, 0.199, 0.122, 0.210 | 0.046, 0.257, 0.510, 0.560, 0.460 | 0.065, 0.238, 0.536, 0.575, 0.442 | 0.076, 0.218, 0.537, 0.555, 0.542 |
| | | **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) × 100** | | |
| `New.loc` | 2.15, 6.52, 14.20, 71.74, 94.02 | 0.49, 0.60, 0.66, 0.51, 1.02 | 1.49, 0.99, 0.68, 0.67, 1.68 | 0.93, 0.69, 1.12, 0.51, 0.37 |
| (% reduction) | 66.2, -125.6, -542.5, -1876, -1124 | 92.4, 80.9, 68.3, 86.6, 87.4 | 79.9, 72.6, 76.4, 83.2, 78.8 | 84.6, 38.9, 56.1, 87.5, 94.2 |
| `loc` | 6.37, 2.89, 2.21, 3.63, 7.68 | 6.43, 3.14, 2.08, 3.82, 8.09 | 7.42, 3.61, 2.88, 4.00, 7.93 | 6.02, 1.13, 2.55, 4.09, 6.42 |
| `New.EM` | 1.85, 5.95, 10.37, 63.46, 84.85 | 0.63, 0.60, 0.61, 0.55, 0.99 | 0.78, 0.69, 0.68, 0.60, 1.15 | 0.77, 0.62, 0.64, 0.61, 1.24 |
| (% reduction) | 11.1, 8.5, 3.0, 1.4, 1.9 | 24.1, 0.0, 15.3, 19.1, 33.1 | 10.3, -15.0, 17.1, 6.3, 27.7 | 14.4, 3.1, 23.8, 16.4, 26.2 |
| `EM` | 2.08, 6.50, 10.69, 64.37, 86.49 | 0.83, 0.60, 0.72, 0.68, 1.48 | 0.87, 0.60, 0.82, 0.64, 1.59 | 0.90, 0.64, 0.84, 0.73, 1.68 |
| | | **IV. normalized mean squared error of estimated error variance ($\sigma^2$) × 100** | | |
| `New.loc` | 10088.01 | 0.22 | 70.69 | 1.31 |
| `loc` (19) | 131.61 | 277.32 | 165.85 | 685.82 |
| `New.EM` | 9176.53 | 0.23 | 0.24 | 0.50 |
| `EM` | 9261.08 | 11.69 | 0.36 | 0.31 |
| | | **V. number of times model with $M$ basis functions selected** | | |
| `New.loc` | 19 | 51 | 9 | 2 |
| `New.EM` | 0 | 97 | 3 | 0 |
| `Combine` | 0 | 98 | 2 | 0 |

Table 21: `Practical`; $n = 1000$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
|---|---|---|---|---|
| | | **I. number of converged replicates** | | |
| New.loc | 53 | 77 | 43 | 28 |
| New.EM | 98 | 97 | 98 | 92 |
| Combine | 99 | 100 | 99 | 96 |
| New.EM.ns | 83 | 62 | 84 | 91 |
| Combine.ns | 93 | 94 | 89 | 93 |
| | | **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | |
| New.loc | 1.613, 1.888, 1.393, 1.825, 1.483 | 0.015, 0.067, 0.169, 0.228, 0.146 | 0.019, 0.062, 0.200, 0.270, 0.152 | 0.017, 0.057, 0.180, 0.273, 0.154 |
| (sd) | 0.028, 0.099, 0.143, 0.099, 0.260 | 0.014, 0.134, 0.226, 0.259, 0.203 | 0.013, 0.051, 0.356, 0.442, 0.309 | 0.012, 0.055, 0.307, 0.484, 0.361 |
| (% reduction) | -527.6, -138.4, -31.5, -75.5, -55.9 | 93.3, 91.8, 84.2, 78.0, 84.3 | 91.2, 92.4, 80.7, 76.0, 85.9 | 92.7, 93.2, 83.8, 77.0, 82.7 |
| loc | 0.257, 0.792, 1.059, 1.040, 0.951 | 0.224, 0.813, 1.069, 1.035, 0.930 | 0.216, 0.815, 1.035, 1.123, 1.081 | 0.232, 0.843, 1.111, 1.185, 0.890 |
| (sd) | 0.264, 0.498, 0.508, 0.610, 0.539 | 0.137, 0.523, 0.466, 0.580, 0.541 | 0.127, 0.499, 0.435, 0.573, 0.558 | 0.139, 0.565, 0.507, 0.603, 0.503 |
| New.EM | 1.614, 1.887, 1.392, 1.852, 1.405 | 0.016, 0.063, 0.145, 0.232, 0.172 | 0.021, 0.068, 0.201, 0.251, 0.135 | 0.019, 0.068, 0.196, 0.263, 0.157 |
| (sd) | 0.026, 0.089, 0.136, 0.095, 0.180 | 0.014, 0.122, 0.193, 0.337, 0.307 | 0.016, 0.126, 0.322, 0.371, 0.240 | 0.016, 0.104, 0.285, 0.360, 0.253 |
| (% reduction) | -0.7, -0.9, -0.2, 0.2, 1.1 | 51.5, 56.2, 53.1, 37.3, 31.2 | 30.0, 52.1, 49.0, 36.1, 25.0 | 48.6, 60.5, 64.3, 54.4, 44.9 |
| EM | 1.603, 1.870, 1.389, 1.855, 1.421 | 0.033, 0.144, 0.309, 0.370, 0.250 | 0.030, 0.142, 0.394, 0.393, 0.180 | 0.037, 0.172, 0.549, 0.577, 0.285 |
| (sd) | 0.133, 0.165, 0.180, 0.095, 0.178 | 0.039, 0.204, 0.330, 0.416, 0.359 | 0.024, 0.245, 0.439, 0.445, 0.294 | 0.029, 0.189, 0.489, 0.502, 0.316 |
| New.EM.ns | 1.612, 1.891, 1.388, 1.817, 1.529 | 0.042, 0.114, 0.214, 0.206, 0.203 | 0.020, 0.074, 0.223, 0.276, 0.144 | 0.019, 0.069, 0.203, 0.276, 0.160 |
| (sd) | 0.028, 0.085, 0.136, 0.093, 0.198 | 0.170, 0.265, 0.341, 0.223, 0.399 | 0.016, 0.135, 0.344, 0.395 0.256 | 0.015, 0.104, 0.288, 0.371, 0.256 |
| (% reduction) | -10.4, -1.7, 0.0, -4.3, 3.7 | 96.8, 92.1, 84.4, 79.3, 87.4 | 71.0, 67.1, 48.0, 50.3, 60.8 | 52.5, 50.0, 46.3, 35.5, 27.9 |
| EM.ns | 1.460, 1.859, 1.388, 1.742, 1.588 | 1.302, 1.434, 1.374, 0.994, 1.610 | 0.069, 0.225, 0.429, 0.555, 0.367 | 0.040, 0.138, 0.378, 0.428, 0.222 |
| (sd) | 0.203, 0.210, 0.221, 0.241, 0.212 | 0.393, 0.270, 0.347, 0.444, 0.289 | 0.037, 0.178, 0.358, 0.520, 0.450 | 0.025, 0.139, 0.363, 0.454, 0.336 |
| | | **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | |
| New.loc | 1.02, 4.69, 10.74, 70.95, 89.69 | 0.26, 0.34, 0.25, 0.35, 0.41 | 0.29, 0.39, 0.30, 0.35, 0.37 | 0.19, 0.20, 0.14, 0.34, 0.30 |
| (% reduction) | 78.0, -136.9, -567.1, -3279, -1648 | 94.6, 82.7, 80.8, 82.0, 91.3 | 94.1, 79.5, 72.5, 78.4, 92.0 | 96.1, 90.4, 88.0, 77.5, 92.2 |
| loc | 4.63, 1.98, 1.61, 2.10, 5.13 | 4.85, 1.96, 1.30, 1.94, 4.72 | 4.88, 1.90, 1.09, 1.62, 4.63 | 4.88, 2.09, 1.17, 1.51, 3.85 |
| New.EM | 0.93, 4.55, 8.53, 64.31, 79.77 | 0.25, 0.29, 0.24, 0.33, 0.41 | 0.22, 0.37, 0.27, 0.32, 0.46 | 0.26, 0.32, 0.25, 0.31, 0.46 |
| (% reduction) | 21.2, 10.6, 0.7, 0.0, 1.1 | 53.7, 9.4, 27.3, 17.5, 43.1 | 53.2, -12.1, 38.6, 15.8, 46.5 | 46.9, -6.7, 49.0, 29.5, 20.7 |
| EM | 1.18, 5.09, 8.59, 64.32, 80.67 | 0.54, 0.32, 0.33, 0.40, 0.72 | 0.47, 0.33, 0.44, 0.38, 0.86 | 0.49, 0.30, 0.49, 0.44, 0.58 |
| New.EM.ns | 1.01, 4.93, 8.88, 69.14, 94.67 | 0.34, 0.72, 0.41, 0.37, 4.47 | 0.24, 0.39, 0.28, 0.31, 0.46 | 0.26, 0.28, 0.27, 0.33, 0.43 |
| (% reduction) | 52.6, 15.7, -0.8, -7.6, -13.8 | 80.9, -7.5, -64.0, 30.2, 23.6 | 64.2, -8.3, 37.8, 29.5, 25.8 | 48.0, -7.7, 37.2, 19.5, 48.2 |
| EM.ns | 2.13, 5.85, 8.81, 64.28, 83.20 | 1.78, 0.67, 0.25, 0.53, 5.85 | 0.67, 0.36, 0.45, 0.44, 0.62 | 0.50, 0.26, 0.43, 0.41, 0.83 |
| | | **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | |
| New.loc | 39226.91 | 0.13 | 0.26 | 0.17 |
| loc (22) | 496.55 | 793.63 | 790.51 | 948.91 |
| New.EM | 36356.12 | 0.15 | 0.23 | 0.14 |
| EM | 36328.60 | 2.04 | 2.26 | 0.66 |
| New.EM.ns | 39170.75 | 125.20 | 0.25 | 0.13 |
| EM.ns | 36399.80 | 862.07 | 3.46 | 0.96 |
| | | **V. number of times model with $M$ basis functions selected** | | |
| New.loc | 7 | 77 | 4 | 5 |
| New.EM | 0 | 97 | 0 | 3 |
| Combine | 0 | 100 | 0 | 0 |
| New.EM.ns | 0 | 59 | 6 | 35 |
| Combine.ns | 0 | 91 | 1 | 8 |

Table 22: `Practical`; $n = 300$; $\sigma^2 = 1/16$; noise distribution : $\frac{1}{\sqrt{2}}t_4$

| | $M = 5$ | **M = 10** | $M = 15$ | $M = 20$ |
|---|---|---|---|---|
| | | **I. number of converged replicates** | | |
| New.loc | 55 | 23 | 0 | 0 |
| New.EM | 99 | 94 | 95 | 66 |
| Combine | 100 | 94 | 95 | 66 |
| | | **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | |
| New.loc | 1.628, 1.750, 1.457, 1.743, 1.422 | 0.073, 0.261, 0.549, 0.639, 0.582 | * * * * * | * * * * * |
| (sd) | 0.045, 0.201, 0.250, 0.145, 0.373 | 0.053, 0.358, 0.468, 0.553, 0.576 | * * * * * | * * * * * |
| (% reduction) | -85.6, -49.7, -24.0, -34.4, -2.2 | 92.6, 79.8, 51.7, 45.2, 56.3 | * * * * * | * * * * * |
| loc | 0.877, 1.169, 1.175, 1.297, 1.392 | 0.985, 1.291, 1.136, 1.167, 1.331 | * * * * * | * * * * * |
| (sd) | 0.579, 0.475, 0.472, 0.490, 0.445 | 0.601 , 0.403, 0.539, 0.417, 0.523 | * * * * * | * * * * * |
| New.EM | 1.628, 1.794, 1.427, 1.795, 1.445 | 0.072, 0.269, 0.570, 0.670, 0.500 | 0.094, 0.252, 0.557, 0.679, 0.537 | 0.098, 0.287, 0.616, 0.810, 0.662 |
| (sd) | 0.045, 0.183, 0.212, 0.121, 0.205 | 0.075, 0.296, 0.429, 0.493, 0.510 | 0.096, 0.273, 0.436, 0.493, 0.527 | 0.105, 0.321, 0.475, 0.582, 0.612 |
| (% reduction) | 0.1, -0.3, 0.3, 0.3, 1.0 | 7.7, 19.0, 14.5, 17.1, 22.1 | 2.1, 16.8, 14.0, 18.3, 17.3 | 31.9, 12.8, 11.1, 12.3, 24.1 |
| EM | 1.630, 1.789, 1.431, 1.800, 1.459 | 0.078, 0.332, 0.667, 0.808, 0.642 | 0.096, 0.303, 0.648, 0.831, 0.649 | 0.144, 0.329, 0.693, 0.924, 0.872 |
| (sd) | 0.046, 0.186, 0.215, 0.122, 0.206 | 0.072, 0.401, 0.503, 0.533, 0.548 | 0.081, 0.318, 0.454, 0.549, 0.557 | 0.167, 0.289, 0.496, 0.589, 0.655 |
| | | **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) × 100** | | |
| New.loc | 2.71, 5.93, 15.25, 71.29, 92.85 | 0.36, 0.56, 0.71, 1.06, 1.60 | * * * * * | * * * * * |
| (% reduction) | 64.0, -29.5, -381.1, -1498, -1007 | 95.7, 87.8, 77.9, 82.7, 85.3 | * * * * * | * * * * * |
| loc | 7.52, 4.58, 3.17, 4.46, 8.39 | 8.31, 4.60, 3.21, 6.13, 10.91 | * * * * * | * * * * * |
| New.EM | 2.87, 5.25, 11.76, 63.00, 87.37 | 0.87, 0.83, 0.74, 0.95, 2.18 | 1.06, 1.08, 0.93, 1.11, 2.41 | 1.00, 1.11, 1.22, 1.11, 2.50 |
| (% reduction) | 1.4, 7.6, 2.6, 1.8, 1.0 | 12.1, -12.2, 12.9, 4.0, 18.0 | 1.9, -36.7, 9.7, 11.9, 17.7 | 25.4, -27.6, -27.1, 14.0, 31.7 |
| EM | 2.91, 5.68, 12.07, 64.16, 88.26 | 0.99, 0.74, 0.85, 0.99, 2.66 | 1.08, 0.79, 1.03, 1.26, 2.93 | 1.34, 0.87, 0.96, 1.29, 3.66 |
| | | **IV. normalized mean squared error of estimated error variance ($\sigma^2$) × 100** | | |
| New.loc | 39350.28 | 7.52 | * | * |
| loc (27) | 1185.29 | 2450.28 | * | * |
| New.EM | 36143.08 | 1.63 | 1.34 | 1.64 |
| EM | 36423.37 | 2.16 | 3.08 | 2.77 |
| | | **V. number of times model with $M$ basis functions selected** | | |
| New.loc | 40 | 23 | 0 | 0 |
| New.EM | 1 | 94 | 5 | 0 |
| Combine | 1 | 94 | 5 | 0 |

Table 23: `Practical`; $n = 500$; $\sigma^2 = 1/16$; noise distribution : $\frac{1}{\sqrt{2}}t_4$

| | $M = 5$ | **M = 10** | $M = 15$ | $M = 20$ |
|---|---|---|---|---|
| | | **I. number of converged replicates** | | |
| New.loc | 44 | 41 | 27 | 6 |
| New.EM | 98 | 95 | 93 | 93 |
| Combine | 99 | 99 | 96 | 93 |
| | | **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | |
| New.loc | 1.621, 1.820, 1.382, 1.789, 1.494 | 0.046, 0.182, 0.290, 0.431, 0.376 | 0.081, 0.144, 0.300, 0.406, 0.425 | 0.113, 0.280, 0.358, 0.537, 0.469 |
| (sd) | 0.040, 0.127, 0.154, 0.141, 0.258 | 0.044, 0.311, 0.313, 0.463, 0.478 | 0.132, 0.176, 0.303, 0.452, 0.583 | 0.086, 0.201, 0.136, 0.660, 0.707 |
| (% reduction) | -224.8, -67.1, -23.8, -50.2, -17.0 | 91.1, 84.1, 74.4, 64.5, 67.4 | 86.0, 86.8, 72.6, 62.7, 61.2 | 68.5, 80.2, 70.3, 45.9, 57.1 |
| loc | 0.499, 1.089, 1.116, 1.191, 1.277 | 0.517, 1.142, 1.134, 1.213, 1.155 | 0.579, 1.095, 1.094, 1.089, 1.094 | 0.359, 1.413, 1.206, 0.992, 1.092 |
| (sd) | 0.417, 0.560, 0.469, 0.480, 0.502 | 0.398, 0.518, 0.499, 0.522, 0.510 | 0.474, 0.569, 0.463, 0.472, 0.592 | 0.270, 0.467, 0.279, 0.568, 0.524 |
| New.EM | 1.622, 1.853, 1.393, 1.825, 1.429 | 0.044, 0.161, 0.345, 0.519, 0.383 | 0.051, 0.165, 0.345, 0.529, 0.405 | 0.052, 0.183, 0.382, 0.496, 0.396 |
| (sd) | 0.039, 0.126, 0.152, 0.114, 0.189 | 0.042, 0.226, 0.398, 0.561, 0.515 | 0.052, 0.242, 0.416, 0.540, 0.515 | 0.055, 0.248, 0.410, 0.491, 0.480 |
| (% reduction) | 0.1, -0.2, 0.4, 0.1, 1.0 | 20.0, 30.0, 29.0, 23.3, 22.3 | 15.0, 19.5, 25.2, 23.8, 20.9 | 32.5, 22.5, 35.7, 37.0, 37.7 |
| EM | 1.624, 1.850, 1.398, 1.827, 1.444 | 0.055, 0.230, 0.486, 0.677, 0.493 | 0.060, 0.205, 0.461, 0.694, 0.512 | 0.077, 0.236, 0.594, 0.787, 0.636 |
| (sd) | 0.040, 0.132, 0.160, 0.113, 0.194 | 0.047, 0.267, 0.472, 0.617, 0.570 | 0.046, 0.214, 0.466, 0.649, 0.624 | 0.092, 0.222, 0.522, 0.591, 0.578 |
| | | **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) × 100** | | |
| New.loc | 1.85, 4.95, 14.38, 71.42, 93.44 | 0.79, 0.59, 0.37, 0.46, 1.08 | 1.76, 1.04, 0.97, 0.33, 3.29 | 1.12, 1.32, 0.49, 0.80, 1.30 |
| (% reduction) | 74.3, -60.2, -588.0, -1775, -957.0 | 88.8, 83.0, 81.7, 89.1, 87.3 | 73.0, 70.8, 56.7, 84.6, 57.0 | 85.0, 71.9, 86.3, 80.8, 84.4 |
| loc | 7.19, 3.09, 2.09, 3.81, 8.84 | 7.05, 3.47, 2.02, 4.21, 8.49 | 6.51, 3.56, 2.24, 2.14, 7.66 | 7.47, 4.70, 3.58, 4.16, 8.31 |
| New.EM | 1.78, 5.01, 9.94, 63.55, 83.96 | 0.66, 0.48, 0.62, 0.49, 1.12 | 0.66, 0.66, 0.70, 0.50, 1.15 | 0.64, 0.58, 0.74, 0.47, 1.13 |
| (% reduction) | 6.3, 9.1, 2.5, 1.2, 1.6 | 23.3, -6.7, 18.4, 15.5, 12.5 | 22.4, -37.5, 13.6, 16.7, 14.8 | 35.4, -20.8, 3.9, 27.7, 30.7 |
| EM | 1.90, 5.51, 10.19, 64.32, 85.34 | 0.86, 0.45, 0.76, 0.58, 1.28 | 0.85, 0.48, 0.81, 0.60, 1.35 | 0.99, 0.48, 0.77, 0.65, 1.63 |
| | | **IV. normalized mean squared error of estimated error variance ($\sigma^2$) × 100** | | |
| New.loc | 40360.97 | 1.40 | 11.16 | 2.62 |
| loc (18) | 894.67 | 1403.41 | 1377.64 | 917.00 |
| New.EM | 36390.17 | 0.90 | 0.92 | 0.95 |
| EM | 36685.30 | 1.17 | 1.94 | 1.79 |
| | | **V. number of times model with $M$ basis functions selected** | | |
| New.loc | 21 | 41 | 11 | 1 |
| New.EM | 0 | 95 | 2 | 3 |
| Combine | 0 | 99 | 0 | 1 |

Table 24: `Practical`; $n = 500$; $\sigma^2 = 1/16$; noise distribution : $\text{Exp}(1) - 1$

| | $M = 5$ | **M = 10** | $M = 15$ | $M = 20$ |
|---|---|---|---|---|
| | | **I. number of converged replicates** | | |
| `New.loc` | 56 | 51 | 20 | 6 |
| `New.EM` | 98 | 95 | 96 | 88 |
| `Combine` | 100 | 95 | 96 | 88 |
| | | **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | |
| `New.loc` | 1.626, 1.839, 1.355, 1.794, 1.501 | 0.042, 0.186, 0.402, 0.438, 0.365 | 0.013, 0.012, 0.004, 0.001, 0.007 | 0.058, 0.272, 0.346, 0.285, 0.168 |
| (sd) | 0.044, 0.142, 0.191, 0.124, 0.244 | 0.063, 0.290, 0.403, 0.358, 0.444 | 0.239, 0.407, 0.435, 0.457, 0.457 | 0.039, 0.178, 0.126, 0.168, 0.132 |
| (% reduction) | -212.1, -73.3, -25.9, -49.9, -21.4 | 92.0, 80.6, 67.9, 58.3, 64.8 | 97.7, 98.8, 99.7, 99.9, 99.3 | 84.6, 62.1, 59.0, 76.1, 81.7 |
| `loc` | 0.521, 1.061, 1.076, 1.197, 1.236 | 0.524, 0.958, 1.251, 1.051, 1.036 | 0.563, 1.038, 1.191, 1.164, 1.075 | 0.376, 0.718, 0.843, 1.190, 0.916 |
| (sd) | 0.403, 0.429, 0.493, 0.497, 0.519 | 0.424, 0.471, 0.451, 0.542, 0.536 | 0.403, 0.514, 0.519, 0.488, 0.509 | 0.196, 0.238, 0.322, 0.418, 0.450 |
| `New.EM` | 1.622, 1.844, 1.388, 1.817, 1.424 | 0.037, 0.150, 0.339, 0.446, 0.317 | 0.047, 0.161, 0.391, 0.491, 0.339 | 0.045, 0.170, 0.393, 0.500, 0.356 |
| (sd) | 0.042, 0.138, 0.180, 0.127, 0.220 | 0.032, 0.151, 0.356, 0.442, 0.379 | 0.042, 0.144, 0.404, 0.471, 0.394 | 0.038, 0.147, 0.398, 0.489, 0.401 |
| (% reduction) | 0.2, -0.2, 0.4, 0.1, 1.0 | 28.8, 39.5, 27.1, 9.5, 15.2 | 11.3, 30.3, 15.7, 4.7, 10.3 | 27.4, 28.0, 32.0, 28.6, 29.8 |
| `EM` | 1.625, 1.840, 1.393, 1.818, 1.439 | 0.052, 0.248, 0.465, 0.493, 0.374 | 0.053, 0.231, 0.464, 0.515, 0.378 | 0.062, 0.236, 0.578, 0.700, 0.507 |
| (sd) | 0.042, 0.144, 0.190, 0.127, 0.213 | 0.042, 0.283, 0.430, 0.465, 0.412 | 0.036, 0.219, 0.395, 0.463, 0.422 | 0.045, 0.214, 0.470, 0.578, 0.528 |
| | | **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | |
| `New.loc` | 1.41, 4.71, 10.06, 74.74, 90.43 | 0.50, 1.05, 0.86, 0.47, 5.97 | 1.34, 2.65, 1.51, 0.75, 4.79 | 0.40, 0.37, 0.68, 0.43, 1.42 |
| (% reduction) | 76.1, -103.0, -449.7, -2434, -945.4 | 90.3, 38.6, 49.4, 79.5, 22.0 | 82.0, 26.4, 21.4, 73.1, 45.0 | 95.6, 91.7, 81.1, 89.4, 77.2 |
| `loc` | 5.91, 2.32, 1.83, 2.95, 8.65 | 5.15, 1.71, 1.70, 2.29, 7.65 | 7.45, 3.60, 1.92, 2.79, 8.71 | 9.05, 4.44, 3.60, 4.05, 6.22 |
| `New.EM` | 1.40, 4.73, 9.23, 65.15, 84.72 | 0.58, 0.42, 0.53, 0.41, 1.11 | 0.61, 0.54, 0.62, 0.44, 1.19 | 0.61, 0.46, 0.58, 0.42, 1.28 |
| (% reduction) | 4.1, 9.7, 2.6, 1.2, 1.3 | 15.9, -16.7, 14.5, 14.6, 25.5 | 11.6, -28.6, 13.9, 10.2, 27.4 | 17.6, 4.2, 14.7, 23.6, 32.3 |
| `EM` | 1.46, 5.24, 9.48, 65.96, 85.81 | 0.69, 0.36, 0.62, 0.48, 1.49 | 0.69, 0.42, 0.72, 0.49, 1.64 | 0.74, 0.48, 0.68, 0.55, 1.89 |
| | | **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | |
| `New.loc` | 37899.82 | 35.64 | 14.17 | 0.75 |
| `loc` (29) | 587.07 | 1692.27 | 2044.38 | 1641.59 |
| `New.EM` | 36544.75 | 0.64 | 0.66 | 0.73 |
| `EM` | 36824.82 | 0.95 | 1.86 | 1.53 |
| | | **V. number of times model with $M$ basis functions selected** | | |
| `New.loc` | 23 | 50 | 6 | 1 |
| `New.EM` | 0 | 95 | 2 | 3 |
| `Combine` | 0 | 95 | 2 | 3 |

Table 25: `Challenging`; $n = 300$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 20$ | $M = 25$ | **M = 30** | $M = 35$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 56 | 51 | 45 | 21 |
| `New.EM` | 92 | 92 | 90 | 64 |
| `Combine` | 96 | 96 | 93 | 72 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `Bias` | 0.169, 0.061, 0.026 | 0.142, 0.045, 0.020 | 0.053, 0.009, 0.005 | 0.084, 0.019, 0.009 |
| `New.loc` | 0.350, 0.296, 0.176 | 0.335, 0.357, 0.211 | 0.147, 0.166, 0.118 | 0.218, 0.224, 0.122 |
| (sd) | 0.209, 0.306, 0.222 | 0.262, 0.383, 0.265 | 0.090, 0.156, 0.116 | 0.225, 0.270, 0.110 |
| (% reduction) | 32.8, 52.2, 60.1 | 40.3, 51.6, 57.5 | 72.4, 74.2, 73.7 | 52.5, 56.6, 63.8 |
| `loc` | 0.521, 0.619, 0.441 | 0.561, 0.738, 0.497 | 0.532, 0.643, 0.449 | 0.459, 0.516, 0.337 |
| (sd) | 0.342, 0.447, 0.440 | 0.401, 0.493, 0.518 | 0.395, 0.511, 0.463 | 0.287, 0.415, 0.376 |
| `New.EM` | 0.374, 0.353, 0.222 | 0.363, 0.380, 0.203 | 0.163, 0.247, 0.189 | 0.216, 0.301, 0.233 |
| (sd) | 0.268, 0.363, 0.267 | 0.280, 0.391, 0.285 | 0.135, 0.322, 0.308 | 0.115, 0.343, 0.328 |
| (% reduction) | -6.6, -10.0, -25.4 | -7.7, -11.1, -12.8 | 5.8, 19.0, 19.2 | -1.4, 2.6, 5.7 |
| `EM` | 0.351, 0.321, 0.177 | 0.337, 0.342, 0.180 | 0.173, 0.305, 0.234 | 0.213, 0.309, 0.247 |
| (sd) | 0.231, 0.304, 0.194 | 0.251, 0.350, 0.214 | 0.157, 0.397, 0.376 | 0.118, 0.325, 0.304 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `Bias` | -0.132, -0.017, -0.007 | -0.116, -0.014, -0.006 | -0.050, -0.004, -0.002 | -0.076, -0.008, -0.003 |
| `New.loc` | 10.19, 1.00, 1.06 | 1.79, 1.10, 0.95 | 1.57, 0.48, 0.89 | 2.20, 0.65, 1.05 |
| (% reduction) | -105.4, 76.0, 72.5 | 61.5, 60.1, 77.7 | 64.2, 79.8, 66.5 | 60.5, 80.1, 79.1 |
| `loc` | 4.96, 4.17, 3.86 | 4.65, 2.76, 4.26 | 4.39, 2.38, 2.66 | 5.57, 3.27, 5.03 |
| `New.EM` | 7.70, 0.97, 1.20 | 1.73, 1.22, 1.16 | 1.34, 0.85, 1.08 | 1.26, 0.80, 1.08 |
| (% reduction) | -206.8, -9.0, -1.7 | 16.4, -19.6, 1.7 | 14.6, 10.5, 18.2 | 19.2, -5.3, 22.9 |
| `EM` | 2.51, 0.89, 1.18 | 2.07, 1.02, 1.18 | 1.57, 0.95, 1.32 | 1.56, 0.76, 1.40 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 140.30 | 119.43 | 16.33 | 38.51 |
| `loc` (39) | 4493.40 | 4887.01 | 4193.35 | 3574.83 |
| `New.EM` | 160.38 | 115.27 | 16.77 | 32.78 |
| `EM` | 196.45 | 175.62 | 51.09 | 71.53 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 56 | 51 | 45 | 21 |
| `New.EM` | 1 | 3 | 90 | 4 |
| `Combine` | 1 | 1 | 93 | 3 |

Table 26: `Challenging`; $n = 500$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 20$ | $M = 25$ | **M = 30** | $M = 35$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 71 | 71 | 58 | 46 |
| `New.EM` | 93 | 91 | 79 | 70 |
| `Combine` | 95 | 96 | 88 | 79 |
| `New.EM.ns` | 65 | 76 | 62 | 62 |
| `Combine.ns` | 94 | 92 | 89 | 77 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `Bias` | 0.169, 0.061, 0.026 | 0.142, 0.045, 0.020 | 0.053, 0.009, 0.005 | 0.084, 0.019, 0.009 |
| `New.loc` | 0.315, 0.244, 0.149 | 0.269, 0.232, 0.100 | 0.162, 0.196, 0.135 | 0.178, 0.191, 0.117 |
| (sd) | 0.261, 0.288, 0.317 | 0.160, 0.205, 0.107 | 0.254, 0.330, 0.285 | 0.124, 0.235, 0.182 |
| (% reduction) | 26.2, 51.9, 50.7 | 38.9, 55.5, 67.8 | 57.5, 62.1, 60.2 | 54.9, 62.7, 66.5 |
| `loc` | 0.427, 0.507, 0.302 | 0.440, 0.521, 0.311 | 0.381, 0.517, 0.339 | 0.395, 0.512, 0.349 |
| (sd) | 0.323, 0.401, 0.323 | 0.342, 0.445, 0.343 | 0.307, 0.453, 0.373 | 0.274, 0.428, 0.353 |
| `New.EM` | 0.286, 0.215, 0.106 | 0.261, 0.215, 0.101 | 0.124, 0.130, 0.082 | 0.178, 0.189, 0.114 |
| (sd) | 0.148, 0.190, 0.097 | 0.147, 0.191, 0.101 | 0.080, 0.147, 0.113 | 0.125, 0.224, 0.162 |
| (% reduction) | 1.0, 7.3, -8.2 | -0.8, 2.7, 0.0 | -0.8, 16.7, 22.6 | -3.5, 11.3, 18.0 |
| `EM` | 0.289, 0.232, 0.098 | 0.259, 0.221, 0.101 | 0.123, 0.156, 0.106 | 0.172, 0.213, 0.139 |
| (sd) | 0.155, 0.203, 0.096 | 0.149, 0.188, 0.101 | 0.081, 0.166, 0.144 | 0.126, 0.258, 0.215 |
| `New.EM.ns` | 0.287, 0.209, 0.100 | 0.271, 0.224, 0.094 | 0.121, 0.132, 0.089 | 0.170, 0.142, 0.080 |
| (sd) | 0.148, 0.173, 0.071 | 0.159, 0.197, 0.089 | 0.060, 0.164, 0.138 | 0.107, 0.154, 0.095 |
| (% reduction) | 57.2, 64.6, 32.0 | -68.3, -7.2, 32.9 | 31.2, 21.4, 15.2 | -21.4, 4.7, 20.0 |
| `EM.ns` | 0.670, 0.590, 0.147 | 0.161, 0.209, 0.140 | 0.176, 0.168, 0.105 | 0.140, 0.149, 0.100 |
| (sd) | 0.424, 0.487, 0.113 | 0.092, 0.311, 0.285 | 0.109, 0.152, 0.110 | 0.073, 0.224, 0.190 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `Bias` | -0.132, -0.017, -0.007 | -0.116, -0.014, -0.006 | -0.050, -0.004, -0.002 | -0.076, -0.008, -0.003 |
| `New.loc` | 3.01, 2.13, 3.17 | 1.94, 0.94, 0.75 | 2.67, 2.06, 2.27 | 1.99, 0.70, 0.53 |
| (% reduction) | 36.6, -7.0, -21.0 | 59.4, 61.6, 72.0 | 37.5, 6.4, 0.9 | 57.4, 68.5, 76.9 |
| `loc` | 4.75, 1.99, 2.62 | 4.78, 2.45, 2.68 | 4.27, 2.20, 2.29 | 4.67, 2.22, 2.29 |
| `New.EM` | 2.39, 0.73, 0.64 | 1.85, 0.99, 0.68 | 1.43, 0.53, 0.67 | 1.85, 0.54, 0.63 |
| (% reduction) | 14.9, -19.7, 23.8 | 15.1, -28.6, 25.3 | -3.6, 7.0, 23.0 | -7.6, -3.8, 24.1 |
| `EM` | 2.81, 0.61, 0.84 | 2.18, 0.77, 0.91 | 1.38, 0.57, 0.87 | 1.72, 0.52, 0.83 |
| `New.EM.ns` | 2.10, 0.78, 0.75 | 1.87, 0.81, 0.78 | 1.51, 0.52, 0.61 | 1.51, 0.73, 0.60 |
| (% reduction) | 40.3, 16.1, 23.5 | -6.3, -50.0, 21.2 | 28.4, 7.1, 20.8 | -28.0, -4.3, 30.2 |
| `EM.ns` | 3.52, 0.93, 0.98 | 1.76, 0.54, 0.99 | 2.11, 0.56, 0.77 | 1.18, 0.70, 0.86 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 1526.18 | 136.76 | 2103.76 | 50.82 |
| `loc` (38) | 2281.71 | 2355.92 | 2469.33 | 2724.98 |
| `New.EM` | 196.92 | 138.32 | 20.82 | 47.87 |
| `EM` | 262.13 | 232.15 | 37.47 | 72.12 |
| `New.EM.ns` | 206.30 | 132.51 | 20.69 | 49.68 |
| `EM.ns` | 769.97 | 67.81 | 86.72 | 61.67 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 5 | 16 | 56 | 16 |
| `New.EM` | 1 | 7 | 79 | 11 |
| `Combine` | 1 | 4 | 88 | 6 |
| `New.EM.ns` | 2 | 10 | 62 | 25 |
| `Combine.ns` | 2 | 2 | 87 | 9 |

Table 27: `Challenging`; $n = 500$; $\sigma^2 = 1/8$; noise distribution : $N(0, 1)$

| | $M = 20$ | $M = 25$ | **M = 30** | $M = 35$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 76 | 60 | 54 | 39 |
| `New.EM` | 90 | 81 | 79 | 82 |
| `Combine` | 96 | 86 | 87 | 85 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `Bias` | 0.169, 0.061, 0.026 | 0.142, 0.045, 0.020 | 0.053, 0.009, 0.005 | 0.084, 0.019, 0.009 |
| `New.loc` | 0.307, 0.242, 0.115 | 0.268, 0.222, 0.093 | 0.116, 0.134, 0.097 | 0.173, 0.183, 0.122 |
| (sd) | 0.231, 0.282, 0.199 | 0.161, 0.207, 0.072 | 0.056, 0.190, 0.163 | 0.117, 0.239, 0.177 |
| (% reduction) | 28.4, 53.1, 65.2 | 38.7, 51.9, 66.1 | 72.4, 73.3, 68.0 | 54.8, 62.7, 65.3 |
| `loc` | 0.429, 0.516, 0.330 | 0.437, 0.462, 0.274 | 0.420, 0.501, 0.303 | 0.383, 0.491, 0.352 |
| (sd) | 0.349, 0.441, 0.320 | 0.367, 0.395, 0.219 | 0.331, 0.440, 0.333 | 0.281, 0.429, 0.327 |
| `New.EM` | 0.297, 0.222, 0.100 | 0.265, 0.213, 0.098 | 0.125, 0.128, 0.080 | 0.173, 0.170, 0.106 |
| (sd) | 0.184, 0.223, 0.092 | 0.158, 0.195, 0.090 | 0.078, 0.132, 0.091 | 0.124, 0.205, 0.136 |
| (% reduction) | -4.6, 0.9, -5.3 | -3.5, 0.0, 5.8 | 10.1, 18.5, 15.8 | 1.1, 10.1, 9.4 |
| `EM` | 0.284, 0.224, 0.095 | 0.256, 0.213, 0.104 | 0.139, 0.157, 0.095 | 0.175, 0.189, 0.117 |
| (sd) | 0.152, 0.203, 0.097 | 0.155, 0.190, 0.101 | 0.115, 0.160, 0.106 | 0.138, 0.226, 0.157 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `Bias` | -0.132, -0.017, -0.007 | -0.116, -0.014, -0.006 | -0.050, -0.004, -0.002 | -0.076, -0.008, -0.003 |
| `New.loc` | 2.89, 1.62, 1.60 | 1.76, 0.87, 0.83 | 1.25, 0.46, 0.70 | 1.77, 0.63, 0.60 |
| (% reduction) | 35.5, 16.1, 36.8 | 60.5, 59.9, 70.7 | 67.8, 77.7, 70.8 | 61.4, 72.6, 76.5 |
| `loc` | 4.48, 1.93, 2.53 | 4.46, 2.17, 2.83 | 3.88, 2.06, 2.40 | 4.58, 2.30, 2.55 |
| `New.EM` | 2.29, 0.77, 0.67 | 1.85, 0.99, 0.68 | 1.26, 0.55, 0.74 | 1.53, 0.64, 0.65 |
| (% reduction) | 10.9, -35.1, 6.9 | 10.6, -32.0, 16.0 | 6.0, 11.3, 19.6 | -2.7, -10.3, 27.8 |
| `EM` | 2.57, 0.57, 0.72 | 2.07, 0.75, 0.81 | 1.34, 0.62, 0.92 | 1.49, 0.58, 0.90 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 298.90 | 32.53 | 4.48 | 9.86 |
| `loc` (25) | 494.28 | 555.93 | 378.26 | 615.58 |
| `New.EM` | 45.81 | 32.15 | 4.06 | 9.60 |
| `EM` | 49.56 | 45.12 | 22.77 | 26.25 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 9 | 14 | 54 | 14 |
| `New.EM` | 1 | 2 | 79 | 15 |
| `Combine` | 2 | 2 | 87 | 9 |

Table 28: `Challenging`; $n = 1000$; $\sigma^2 = 1/16$; noise distribution : $N(0, 1)$

| | $M = 20$ | $M = 25$ | **M = 30** | $M = 35$ |
|---|---|---|---|---|
| **I. number of converged replicates** | | | | |
| `New.loc` | 76 | 59 | 49 | 45 |
| `New.EM` | 95 | 71 | 79 | 88 |
| `Combine` | 95 | 81 | 90 | 93 |
| **II. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | | |
| `Bias` | 0.169, 0.061, 0.026 | 0.142, 0.045, 0.020 | 0.053, 0.009, 0.005 | 0.084, 0.019, 0.009 |
| `New.loc` | 0.237, 0.148, 0.057 | 0.223, 0.151, 0.053 | 0.099, 0.071, 0.035 | 0.149, 0.115, 0.053 |
| (sd) | 0.116, 0.144, 0.042 | 0.120, 0.130, 0.039 | 0.048, 0.062, 0.041 | 0.071, 0.092, 0.052 |
| (% reduction) | 21.3, 51.8, 69.4 | 29.0, 55.1, 73.8 | 66.8, 73.6, 78.1 | 55.5, 65.8, 67.7 |
| `loc` | 0.301, 0.307, 0.186 | 0.314, 0.336, 0.202 | 0.298, 0.269, 0.160 | 0.335, 0.336, 0.164 |
| (sd) | 0.202, 0.293, 0.218 | 0.219, 0.324, 0.252 | 0.151, 0.201, 0.137 | 0.243, 0.297, 0.149 |
| `New.EM` | 0.243, 0.158, 0.058 | 0.218, 0.149, 0.057 | 0.104, 0.092, 0.047 | 0.146, 0.109, 0.051 |
| (sd) | 0.112, 0.142, 0.041 | 0.106, 0.119, 0.039 | 0.046, 0.078, 0.057 | 0.063, 0.088, 0.045 |
| (% reduction) | -2.5, 8.1, 13.4 | -2.8, 9.1, 26.9 | 7.1, 14.8, 24.2 | -5.0, 4.4, 21.5 |
| `EM` | 0.237, 0.172, 0.067 | 0.212, 0.164, 0.078 | 0.112, 0.108, 0.062 | 0.139, 0.114, 0.065 |
| (sd) | 0.093, 0.186, 0.155 | 0.087, 0.192, 0.179 | 0.061, 0.103, 0.089 | 0.055, 0.097, 0.076 |
| **III. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | | |
| `Bias` | -0.132, -0.017, -0.007 | -0.116, -0.014, -0.006 | -0.050, -0.004, -0.002 | -0.076, -0.008, -0.003 |
| `New.loc` | 2.45, 0.47, 0.41 | 2.06, 0.77, 0.34 | 1.28, 0.32, 0.34 | 1.87, 0.37, 0.41 |
| (% reduction) | 47.5, 59.8, 77.1 | 54.7, 38.9, 79.4 | 74.1, 74.8, 82.2 | 64.2, 74.1, 80.1 |
| `loc` | 4.67, 1.17, 1.79 | 4.55, 1.26, 1.65 | 4.94, 1.27, 1.91 | 5.22, 1.43, 2.06 |
| `New.EM` | 2.51, 0.43, 0.40 | 2.14, 0.70, 0.51 | 1.28, 0.27, 0.34 | 1.78, 0.34, 0.35 |
| (% reduction) | 11.3, -13.2, 13.0 | 6.1, -12.9, 8.9 | 5.9, 34.1, 34.6 | -18.7, 12.8, 30.0 |
| `EM` | 2.83, 0.38, 0.46 | 2.28, 0.62, 0.56 | 1.36, 0.41, 0.52 | 1.50, 0.39, 0.50 |
| **IV. normalized mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | | |
| `New.loc` | 236.51 | 165.16 | 27.42 | 60.40 |
| `loc` (33) | 968.17 | 905.22 | 860.51 | 872.33 |
| `New.EM` | 236.00 | 163.41 | 26.73 | 60.70 |
| `EM` | 245.43 | 210.32 | 52.84 | 77.08 |
| **V. number of times model with $M$ basis functions selected** | | | | |
| `New.loc` | 6 | 18 | 49 | 18 |
| `New.EM` | 2 | 2 | 79 | 15 |
| `Combine` | 1 | 2 | 90 | 5 |

Table 29: Model selection : `Easy` ($n = 200$), `Practical` ($n = 500, 1000$), and `Challenging` ($n = 500$), $\sigma^2 = 1/16$, $N(0,1)$ noise

| Model | Method | Number of converged replicates | | | | Frequency of models selected | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
| | `New.loc` | 91 | 82 | 86 | 82 | 3 | 80 | 1 | 9 |
| | `New.EM` | 99 | 99 | 99 | 98 | 0 | 99 | 0 | 0 |
| `Easy` | `Combine` | 100 | 100 | 100 | 100 | 0 | 100 | 0 | 0 |
| ($n = 200$) | `New.EM.ns` | 99 | 99 | 99 | 99 | 0 | 99 | 0 | 0 |
| | `Combine.ns` | 99 | 99 | 99 | 99 | 0 | 99 | 0 | 0 |
| | | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
| | `New.loc` | 53 | 46 | 21 | 7 | 25 | 46 | 8 | 1 |
| | `New.EM` | 96 | 93 | 94 | 88 | 0 | 93 | 2 | 5 |
| `Practical` | `Combine` | 100 | 97 | 95 | 88 | 2 | 97 | 1 | 2 |
| ($n = 500$) | `New.EM.ns` | 72 | 60 | 78 | 86 | 3 | 60 | 17 | 20 |
| | `Combine.ns` | 90 | 82 | 81 | 87 | 2 | 82 | 7 | 9 |
| | | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
| | `New.loc` | 53 | 77 | 43 | 28 | 7 | 77 | 4 | 5 |
| | `New.EM` | 98 | 97 | 98 | 92 | 0 | 97 | 0 | 3 |
| `Practical` | `Combine` | 99 | 100 | 99 | 96 | 0 | 100 | 0 | 0 |
| ($n = 1000$) | `New.EM.ns` | 83 | 62 | 84 | 91 | 0 | 59 | 6 | 35 |
| | `Combine.ns` | 93 | 94 | 89 | 93 | 0 | 91 | 1 | 8 |
| | | $M = 20$ | $M = 25$ | $\mathbf{M = 30}$ | $M = 35$ | $M = 20$ | $M = 25$ | $\mathbf{M = 30}$ | $M = 35$ |
| | `New.loc` | 71 | 71 | 58 | 46 | 5 | 16 | 56 | 16 |
| | `New.EM` | 93 | 91 | 79 | 70 | 1 | 7 | 79 | 11 |
| `Challenging` | `Combine` | 95 | 96 | 88 | 79 | 1 | 4 | 88 | 6 |
| | `New.EM.ns` | 65 | 76 | 62 | 62 | 2 | 10 | 62 | 25 |
| | `Combine.ns` | 94 | 92 | 89 | 77 | 2 | 2 | 87 | 9 |

Table 30: Model selection : `Hybrid`, $n = 300$, $\sigma^2 = 1/16$, $N(0,1)$ noise

| | `New.loc` | | | | `New.EM` | | | |
|---|---|---|---|---|---|---|---|---|
| **I. number of converged replicates (total=100)** | | | | | | | | |
| $r$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ |
| 2 | 76 | 62 | 43 | 29 | 99 | 98 | 94 | 90 |
| 3 | 65 | 42 | 30 | 26 | 97 | 99 | 95 | 94 |
| 4 | 19 | 7 | 0 | 0 | 97 | 89 | 87 | 79 |
| 5 | 8 | 0 | 0 | 0 | 92 | 78 | 32 | 18 |
| 6 | 0 | 0 | 0 | 0 | 70 | 44 | 13 | 4 |
| 7 | 2 | 0 | 0 | 0 | 11 | 12 | 6 | 0 |
| **II. frequencies of models selected** | | | | | | | | |
| $r$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ | $M = 25$ |
| 1 | 0 (2,2,2) | 0 (0,0,0) | 0 (0,0,0) | 0 (1,1,1) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 2 | 8 (8,8,9) | 7 (8,8,8) | 4 (4,4,4) | 1 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 3 | 46 (52,54,62) | 2 (1,2,2) | 1 (1,1,1) | 0 (0,0,0) | 0 (0,0,98) | 0 (0,0,1) | 0 (0,0,0) | 0 (0,0,0) |
| 4 | 12 (11,9,0) | 1 (1,0,0) | 0 (0,0,0) | 0 (0,0,0) | 2 (32,90,1) | 1 (1,1,0) | 0 (0,0,0) | 0 (0,0,0) |
| 5 | 5 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 36 (64,9,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 6 | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 50 (2,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 7 | 2 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 11 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |

Table 31: Model selection : `Hybrid`, $n = 500$, $\sigma^2 = 1/16$, $N(0,1)$ noise

| | `New.loc` | | | | `New.EM` | | | |
|---|---|---|---|---|---|---|---|---|
| **I. number of converged replicates (total=100)** | | | | | | | | |
| $r$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ |
| 2 | 75 | 63 | 60 | 54 | 100 | 98 | 99 | 85 |
| 3 | 81 | 78 | 60 | 61 | 99 | 99 | 97 | 96 |
| 4 | 36 | 11 | 3 | 1 | 99 | 94 | 94 | 93 |
| 5 | 6 | 0 | 0 | 0 | 95 | 85 | 60 | 38 |
| 6 | 1 | 0 | 0 | 0 | 61 | 60 | 29 | 8 |
| 7 | 4 | 0 | 0 | 0 | 7 | 22 | 7 | 3 |
| **II. frequencies of models selected** | | | | | | | | |
| $r$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ |
| 1 | 0 (1,1,1) | 0 (0,0,1) | 0 (0,0,0) | 0 (0,0,1) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 2 | 2 (4,6,6) | 2 (3,3,2) | 2 (2,2,2) | 2 (2,2,1) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 3 | 47 (61,63,79) | 2 (1,1,1) | 1 (1,1,1) | 0 (0,0,0) | 0 (0,0,98) | 0 (0,0,1) | 0 (0,0,1) | 0 (0,0,0) |
| 4 | 30 (20,16,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 2 (43,97,0) | 0 (0,1,0) | 0 (1,1,0) | 0 (0,0,0) |
| 5 | 3 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 56 (55,1,0) | 0 (1,0,0) | 1 (0,0,0) | 0 (0,0,0) |
| 6 | 1 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 33 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 7 | 3 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 7 (0,0,0) | 1 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |

Table 32: Model selection : `Hybrid`, $n = 1000$, $\sigma^2 = 1/16$, $N(0,1)$ noise

| | `New.loc` | | | | `New.EM` | | | |
|---|---|---|---|---|---|---|---|---|
| **I. number of converged replicates (total=50)** | | | | | | | | |
| $r$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ |
| 2 | 41 | 39 | 34 | 31 | 48 | 49 | 49 | 46 |
| 3 | 42 | 37 | 35 | 35 | 49 | 49 | 45 | 43 |
| 4 | 26 | 13 | 8 | 11 | 49 | 45 | 43 | 46 |
| 5 | 9 | 3 | 0 | 0 | 49 | 45 | 44 | 32 |
| 6 | 5 | 0 | 0 | 0 | 29 | 32 | 24 | 12 |
| 7 | 1 | 0 | 0 | 0 | 6 | 11 | 9 | 3 |
| **II. frequencies of models selected** | | | | | | | | |
| $r$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ | **M = 10** | $M = 15$ | $M = 20$ | $M = 25$ |
| 1 | 0 (0,0,0) | 0 (0,0,1) | 0 (0,0,1) | 0 (0,0,1) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 2 | 0 (1,1,1) | 1 (1,1,0) | 1 (1,1,0) | 1 (1,1,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |
| 3 | 11 (17,21,41) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,1) | 0 (0,0,49) | 0 (0,0,0) | 0 (0,0,1) | 0 (0,0,0) |
| 4 | 19 (22,20,0) | 0 (0,0,0) | 0 (0,0,0) | 1 (1,1,0) | 0 (27,49,0) | 0 (0,0,0) | 0 (0,1,0) | 0 (0,0,0) |
| 5 | 7 (2,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 26 (22,0,0) | 0 (0,0,0) | 0 (1,0,0) | 0 (0,0,0) |
| 6 | 5 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 18 (0,0,0) | 0 (0,0,0) | 1 (0,0,0) | 0 (0,0,0) |
| 7 | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 5 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) | 0 (0,0,0) |

Table 33: Risk for most frequently selected models: `Hybrid`, $\sigma^2 = 1/16$, $N(0,1)$ noise (based on *all converged replicates*)

| | | | |
|---|---|---|---|
| **I. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | |
| | | **M = 10**, $r = 5$ | **M = 10**, $r = 6$ |
| $n = 300$ | `New.EM` | 0.042, 0.138, 0.118, 0.023, 0.513 | 0.042, 0.153, 0.136, 0.022, 0.474, 1.248 |
| | (sd) | 0.054, 0.168, 0.157, 0.013, 0.528 | 0.054, 0.195, 0.185, 0.011, 0.532, 0.508 |
| | (% reduction) | 0.0, 34.3, 38.9, 11.5, 6.4 | 2.3, 36.5, 39.6, 18.5, 17.0, -8.6 |
| | `EM` | 0.042, 0.210, 0.193, 0.026, 0.548 | 0.043, 0.241, 0.225, 0.027, 0.571, 1.149 |
| | (sd) | 0.048, 0.291, 0.289, 0.016, 0.465 | 0.047, 0.312, 0.310, 0.016, 0.507, 0.497 |
| $n = 500$ | `New.EM` | 0.025, 0.066, 0.053, 0.012, 0.233 | 0.022, 0.076, 0.065, 0.014, 0.215, 1.098 |
| | (sd) | 0.027, 0.091, 0.088, 0.008, 0.331 | 0.027, 0.109, 0.105, 0.009, 0.231, 0.531 |
| | (% reduction) | 3.8, 28.3, 32.9, 14.3, 30.7 | 8.3, 30.9, 33.0, 6.7, 44.0, 2.5 |
| | `EM` | 0.026, 0.092, 0.079, 0.014, 0.336 | 0.024, 0.110, 0.097, 0.015, 0.384, 1.126 |
| | (sd) | 0.027, 0.138, 0.133, 0.008, 0.339 | 0.028, 0.169, 0.165, 0.009, 0.459, 0.538 |
| $n = 1000$ | `New.EM` | 0.010, 0.022, 0.020, 0.006, 0.140 | 0.009, 0.023, 0.021, 0.006, 0.099, 1.066 |
| | (sd) | 0.009, 0.023, 0.021, 0.003, 0.293 | 0.007, 0.025, 0.025, 0.004, 0.142, 0.513 |
| | (% reduction) | 37.5, 38.9, 28.6, 0.0, 4.1 | 30.8, 39.5, 32.3, 14.3, 34.0, -25.3 |
| | `EM` | 0.016, 0.036, 0.028, 0.006, 0.146 | 0.013, 0.038, 0.031, 0.007, 0.150, 0.851 |
| | (sd) | 0.015, 0.032, 0.030, 0.004, 0.175 | 0.012, 0.039, 0.037, 0.005, 0.237, 0.540 |
| **II. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) × 100** | | | |
| | | **M = 10**, $r = 5$ | **M = 10**, $r = 6$ |
| $n = 300$ | `New.EM` | 0.83, 0.69, 0.62, 1.80, 27.32 | 0.80, 0.80, 0.62, 1.93, 20.50, 1212.58 |
| | (% reduction) | -5.1, 16.9, 4.6, 2.2, 53.6 | -1.3, 14.9, 1.6, -2.1, 50.1, 12.1 |
| | `EM` | 0.79, 0.83, 0.65, 1.84, 58.90 | 0.79, 0.94, 0.63, 1.89, 41.08, 1379.04 |
| $n = 500$ | `New.EM` | 0.37, 0.49, 0.40, 1.12, 7.56 | 0.32, 0.52, 0.49, 1.21, 6.30, 540.12 |
| | (% reduction) | 0.0, 19.7, -8.1, 4.3, 47.0 | 5.9, 26.8, -4.3, 0.8, 54.5, 33.2 |
| | `EM` | 0.37, 0.61, 0.37, 1.17, 14.26 | 0.34, 0.71, 0.47, 1.22, 13.86, 808.37 |
| $n = 1000$ | `New.EM` | 0.26, 0.26, 0.20, 0.64, 5.90 | 0.23, 0.22, 0.16, 0.61, 1.62, 516.18 |
| | (% reduction) | 7.1, 42.2, -5.3, 1.5, 9.8 | 23.3, 53.2, -6.7, -1.7, 55.6, 6.6 |
| | `EM` | 0.28, 0.45, 0.19, 0.65, 6.54 | 0.30, 0.47, 0.15, 0.60, 3.65, 552.56 |
| **III. mean squared error of estimated error variance ($\sigma^2$) × 100** | | | |
| | | **M = 10**, $r = 5$ | **M = 10**, $r = 6$ |
| $n = 300$ | `New.EM` | 0.602 | 0.601 |
| | `EM` | 0.613 | 0.438 |
| $n = 500$ | `New.EM` | 0.321 | 0.283 |
| | `EM` | 0.514 | 0.256 |
| $n = 1000$ | `New.EM` | 0.223 | 0.138 |
| | `EM` | 0.366 | 0.160 |

Table 34: Risk for most frequently selected models: `Hybrid`, $\sigma^2 = 1/16$, $N(0,1)$ noise (based on *selected replicates only*)

| | | | |
|---|---|---|---|
| **I. mean integrated squared error of estimated eigenfunctions ($\psi_\nu$)** | | | |
| | | **M = 10**, $r = 5$ | **M = 10**, $r = 6$ |
| $n = 300$ | `New.EM` | 0.044, 0.125, 0.103, 0.022, 0.406 | 0.041, 0.140, 0.122, 0.023, 0.495, 1.212 |
| | (sd) | 0.057, 0.149, 0.118, 0.013, 0.430 | 0.055, 0.165, 0.163, 0.011, 0.572, 0.529 |
| | (% reduction) | 2.2, 36.2, 42.1, 12.0, -4.1 | -5.1, 36.4, 40.8, 14.8, 14.8, -4.3 |
| | `EM` | 0.045, 0.196, 0.178, 0.025, 0.390 | 0.039, 0.220, 0.206, 0.027, 0.581, 1.162 |
| | (sd) | 0.052, 0.289, 0.273, 0.014, 0.350 | 0.045, 0.278, 0.283, 0.016, 0.515, 0.505 |
| $n = 500$ | `New.EM` | 0.024, 0.055, 0.043, 0.012, 0.164 | 0.025, 0.091, 0.077, 0.013, 0.216, 1.118 |
| | (sd) | 0.024, 0.056, 0.054, 0.007, 0.147 | 0.032, 0.134, 0.131, 0.008, 0.178, 0.543 |
| | (% reduction) | 0.0, 33.7, 40.3, 7.7, 37.2 | 7.4, 26.0, 28.7, 7.1, 53.3, -1.7 |
| | `EM` | 0.024, 0.083, 0.072, 0.013, 0.261 | 0.027, 0.123, 0.108, 0.014, 0.463, 1.099 |
| | (sd) | 0.023, 0.107, 0.105, 0.74, 0.246 | 0.033, 0.195, 0.189, 0.008, 0.524, 0.540 |
| $n = 1000$ | `New.EM` | 0.011, 0.021, 0.017, 0.005, 0.084 | 0.008, 0.020, 0.019, 0.007, 0.086, 1.010 |
| | (sd) | 0.011, 0.019, 0.016, 0.002, 0.119 | 0.007, 0.027, 0.025, 0.005, 0.075, 0.522 |
| | (% reduction) | 42.1, 40.0, 32.0, 0.0, 22.9 | 38.5, 41.2, 26.9, 12.5, 25.2, -21.7 |
| | `EM` | 0.019, 0.035, 0.025, 0.005, 0.109 | 0.013, 0.034, 0.026, 0.008, 0.115, 0.830 |
| | (sd) | 0.015, 0.025, 0.022, 0.003, 0.126 | 0.014, 0.041, 0.037, 0.005, 0.100, 0.561 |
| **II. normalized mean squared error of estimated eigenvalues ($\lambda_\nu$) $\times$ 100** | | | |
| | | **M = 10**, $r = 5$ | **M = 10**, $r = 6$ |
| $n = 300$ | `New.EM` | 0.67, 0.72, 0.58, 1.91, 16.15 | 0.93, 0.90, 0.64, 2.00, 23.40, 1340.23 |
| | (% reduction) | -8.1, -16.1, 3.3, -3.2, 80.8 | -3.3, 18.2, 0.0, 0.5, 30.2, 11.0 |
| | `EM` | 0.62, 0.62, 0.60, 1.85, 84.11 | 0.90, 1.10, 0.64, 2.01, 33.52, 1505.25 |
| $n = 500$ | `New.EM` | 0.36, 0.48, 0.30, 0.85, 5.85 | 0.27, 0.31, 0.60, 1.41, 5.50, 833.66 |
| | (% reduction) | 2.7, 27.3, 0.0, -1.2, 43.9 | 10.0, 44.6, -9.1, 2.8, 67.8, 29.7 |
| | `EM` | 0.37, 0.66, 0.30, 0.84, 10.42 | 0.30, 0.56, 0.55, 1.45, 17.08, 1186.57 |
| $n = 1000$ | `New.EM` | 0.36, 0.30, 0.24, 0.73, 5.98 | 0.19, 0.24, 0.20, 0.55, 1.49, 480.45 |
| | (% reduction) | 0.0, 40.0, -4.3, 6.4, 5.2 | 24.0, 45.5, 0.0, 1.8, 38.7, 7.0 |
| | `EM` | 0.36, 0.50, 0.23, 0.78, 6.31 | 0.25, 0.44, 0.20, 0.56, 2.43, 516.39 |
| **III. mean squared error of estimated error variance ($\sigma^2$) $\times$ 100** | | | |
| | | **M = 10**, $r = 5$ | **M = 10**, $r = 6$ |
| $n = 300$ | `New.EM` | 0.353 | 0.673 |
| | `EM` | 0.462 | 0.462 |
| $n = 500$ | `New.EM` | 0.238 | 0.345 |
| | `EM` | 0.396 | 0.275 |
| $n = 1000$ | `New.EM` | 0.196 | 0.114 |
| | `EM` | 0.325 | 0.142 |

Table 35: Mean running time for one replicate in seconds: `Easy` ($n = 200$), `Practical` ($n = 500$), $\sigma^2 = 1/16$, Gaussian noise

| Model | Method | **Mean running time (standard deviation)** * | | | |
|---|---|---|---|---|---|
| | | $M = 4$ | $\mathbf{M = 5}$ | $M = 6$ | $M = 9$ |
| | `New.loc` | 14.3 (4.4) | 14.4 (4) | 15.4 (4.9) | 16.7 (5.2) |
| `Easy` | `loc` | 19(1.9) | 19(1.9) | 19(1.9) | 19(1.9) |
| ($n = 200$) | `New.EM` | 14.7(0.48) | 14.4 (0.52) | 14.8 (0.42) | 16.3 (0.48) |
| | `EM` | 9.8 (0.79) | 9.7 (0.48) | 10.1 (0.32) | 10.7 (1.1) |
| | | $M = 5$ | $\mathbf{M = 10}$ | $M = 15$ | $M = 20$ |
| | `New.loc` | 63.8 (27.9) | 80.9 (45.1) | 87.4 (35.8) | 92.7 (31.2) |
| `Practical` | `loc` | 28.4 (3.4) | 28.4 (3.4) | 28.4 (3.4) | 28.4 (3.4) |
| ($n = 500$) | `New.EM` | 60.2 (9.5) | 59.4 (3.1) | 70.6 (17.9) | 91.9 (30.2) |
| | `EM` | 54.1 (6.7) | 47.6 (2.2) | 53.7 (6.7) | 61.2 (11.9) |

$*$ for `New.loc` and `New.EM`, this means the additional computational cost after obtaining the initial estimates.

Table 36: `CD4 counts` data: estimated error variance and eigenvalues

| **Model :** $M = 10$, $r = 4$ | $\widehat{\sigma}^2$ | $\widehat{\lambda}_1$ | $\widehat{\lambda}_2$ | $\widehat{\lambda}_3$ | $\widehat{\lambda}_4$ |
|---|---|---|---|---|---|
| `loc` | 42,359.3 | 615,735.6 | 94,188.6 | 47,012.6 | 37,687.1 |
| `New.EM` | 38,411.0 | 473,416.8 | 208,201.4 | 53,253.9 | 24,582.0 |
| `EM` | 38,132.2 | 469,784.3 | 207,961.1 | 54,007.2 | 24,344.5 |