多元响应函数回归

Ruiyan Luo

乔治亚州立大学公共卫生学院流行病学与生物统计学部，140 Decatur Street，亚特兰大，GA 30303 rluo@gsu.edu

Xin Qi

乔治亚州立大学数学与统计学系，25 Park Place，亚特兰大，GA 30303-3083 xqi3@gsu.edu

arXiv:2602.02860v1 [stat.ME] 2 Feb 2026

摘要

我们考虑具有多元响应和函数型预测变量的函数回归模型。与分别拟合各个响应变量相比，利用响应变量之间的相关性可以提高估计和预测精度。综合利用函数型预测变量与多元响应中的信息，我们在总体层面识别出用于预测的系数函数的最优分解。随后，我们分别针对少量和大量 p 函数型预测变量的情形，提出估计该分解并拟合回归模型的方法。对于大量 p，我们提出一种同时光滑稀疏惩罚，既能进行曲线选择，又能提高估计和预测精度。我们给出了当样本量和函数型预测变量个数均趋于无穷时的渐近结果。我们的方法可应用于包含数千个函数型预测变量的模型，并已在R 包 FRegSigCom中实现。

关键词：函数型线性回归；函数型偏最小二乘；多元响应；最优分解；同时光滑稀疏惩罚；数千个函数型预测变量。

短标题：具有多元响应的函数型回归

1 引言

人们对根据一组共同的预测变量曲线同时预测多个量越来越感兴趣。例如，人们希望根据玉米样本的近红外（NIR）光谱预测多种营养成分的浓度。我们考虑以下具有多元响应和函数型预测变量的线性回归模型（多响应标量对函数回归，MSOF）：

$$\mathbf{Y} = \boldsymbol{\mu} + \sum_{j=1}^{p} \int_0^1 X_j(t)\mathbf{b}_j(t)dt + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \int \mathbb{B}(t)^{\mathrm{T}}\mathbf{X}(t)dt + \boldsymbol{\varepsilon}, \tag{1.1}$$

其中 $\mathbf{Y} = (Y_1, \ldots, Y_m)^{\mathrm{T}}$ 是一个 m 维随机向量，且各 Y_i 之间通常存在相关性。函数型预测变量 $\mathbf{X}(t) = (X_1(t) \ldots, X_p(t))^{\mathrm{T}}$ 是随机函数；不失一般性，我们假设各 $X_j(t)$ 定义在 $[0, 1]$ 上，且均值为零。各 $X_j(t)$ 之间可以存在相关性。系数矩阵 $\mathbb{B}(t) = (\mathbf{b}_1(t) \ldots, \mathbf{b}_p(t))^{\mathrm{T}}$ 为 $p \times m$，其中 $\mathbf{b}_j(t) = (b_{j1}(t) \ldots, b_{jm}(t))^{\mathrm{T}}$ 是 m 维的。函数向量或函数矩阵的积分按元素定义。向量 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^{\mathrm{T}}$ 和 $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_m)^{\mathrm{T}}$ 分别表示截距和随机噪声。我们假设 $\boldsymbol{\varepsilon}$ 均值为零，且与 $\mathbf{X}(t)$ 独立。

针对标量对函数线性回归模型（SOF）这一具有 $m = 1$ 的（1.1）式特例，人们提出了多种方法。我们简要回顾这些方法；更全面的综述可参见 Morris（2015）、Wang 等（2016）以及 Reiss等（2017）。拟合 SOF 模型的一种常见通用方法是使用一组基函数展开系数函数（也可能包括预测变量曲线），并将 SOF 模型转换为含有多个标量预测变量的线性模型。选择基函数有两种常见方法。第一种方法中，基函数的形式不依赖于数据，例如 B样条、P样条和傅里叶基。Ramsay 和 Dalzell（1991）、Marx 和 Eilers（1999）、Cardot 等（2003b）、Ramsay 和 Silverman（2005）、Crambes 等（2009）、Crainiceanu 和 Goldsmith（2010）、Goldsmith等（2011）等采用了这种方法。第二种方法则根据数据自适应地选择基函数，例如函数型主成分回归（FPCR）（Rice 和 Silverman，1991；Silverman，1996；Cardot 等，1999；Yao 等，2005b；Hall 和 Vial，2006；Reiss 和 Ogden，2007；Hall等，2007；Hilgert

等，2013；Li 等，2013；Hosseini-Nasab，2013）、函数偏最小二乘（FPLS）（Preda 和 Saporta，2005；Reiss 和 Ogden，2007；Escabias 等，2007；Kr¨amer 等，2008；Aguilera 等，2010；Delaigle 等，2012；Hilgert 等，2013；Li 等，2013；Hosseini-Nasab，2013），以及函数型切片逆回归（Ferr´e 和 Yao，2003，2005）。FPCR 利用预测变量曲线的协方差函数获取基函数，已得到广泛研究，并衍生出多个变体（Fengler 等，2003；Kneip 和 Utikal，2001；Benko 等，2009；Di等，2009；Crainiceanu 等，2009；Greven 等，2011）。FPLS 利用响应和预测变量两方面的信息，通过最大化响应与函数型预测变量之间的协方差来获得基函数。当 p较大时，人们提出了结合基展开与变量选择的混合方法。Zhu 和 Cox（2009）、Lian（2013）以及 Kong等（2016b）分别对每个函数型预测变量进行 FPCA，并将主成分得分作为新的标量预测变量，通过施加组Lasso 或组SCAD惩罚来拟合模型。Kong 等（2016a）、Swihart 等（2014）和 Collazos等（2016）采用假设检验来选择函数型变量。基于小波的方法（Brown 等，2001；Zhao 等，2012，2015；Reiss等，2015）首先将函数型预测变量转换到小波空间，然后将小波系数作为预测变量，通过施加各种稀疏正则化或先验来拟合模型。除了线性模型外，人们还研究了非线性SOF模型，例如函数型单指标模型（Eilers 等，2009；Ait-Sa¨ıdi 等，2008；Ferraty等，2011）、函数型多指标模型（James 和 Silverman，2005；Chen 等，2011；Ferraty 等，2013）、函数型非参数模型（Ferraty 和 Vieu，2002，2006a；Rachdi 和 Vieu，2007；Shang，2013；Zhang等，2014）等。

对于多变量函数型回归，仅有少数方法被提出，用于联合拟合多元响应的所有坐标。Matsui等人（2008年）使用高斯基函数同时展开预测变量曲线和系数函数，并将函数型线性模型转化为具有多元响应和多个标量预测变量的线性模型，再采用最大似然估计进行拟合。Wang 等人（2017年）假设模型$\mathbf{Y} = \mathbf{f}(X) + \varepsilon$，其中$\mathbf{Y}$为多元响应，$\mathbf{f}$为 $X(t)$的未知多变量函数。首先对多元响应进行函数主成分分析以去相关，然后将每个主成分得分分别作为新的响应，并采用高斯过程回归进行拟合（O’Hagan和Kingman，1978年；Rasmussen和Williams，2006年）。Chaouch 等人（2013年）

考虑具有多元响应和函数型预测变量的非参数模型，并提出一种 L_1-中位数估计方法。

对于多元响应，可以使用各种方法，针对 Y 的每个坐标分别拟合 X(t) 的 SOF回归模型。然而，这种方法忽略了 Y 各坐标之间的相关性，从而降低了预测精度和计算效率。一种考虑 Y 中相关性的方法，是考虑如下形式的 $\mathbb{B}(t)$ 分解：$\sum_{k=1}^{K} \boldsymbol{\gamma}_k(t)\mathbf{v}_k^{\mathrm{T}}$，其中 $\boldsymbol{\gamma}_k(t) = (\gamma_{k1}(t) \ldots, \gamma_{kp}(t))^{\mathrm{T}}$ 是一个 p 维函数向量，$\mathbf{v}_k$ 是一个 m 维向量，而 K 称为成分个数。函数主成分回归和函数偏最小二乘中的估计都属于此类分解。为估计回归函数 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} = \mathrm{E}[\mathrm{Y}|X_1(t) \ldots, X_p(t)] = \boldsymbol{\mu} + \sum_{j=1}^{p} \int X_j(t)\mathbf{b}_j(t)^{\mathrm{T}} dt$并进行预测，我们在总体层面确定此类分解的最优形式，并通过广义特征值问题刻画该分解，从而为实际应用中高效估计该分解和拟合模型提供一种方法。Y 各坐标之间的相关性越强，要获得良好的回归函数估计和预测，所需的成分就越少。因此，我们的方法利用响应中的相关结构实现高效降维。我们针对函数型预测变量较少和较多的模型，分别提出了不同的正则化方法。当 p 较大时，我们提出一种同步稀疏平滑惩罚，通过在拟合模型中选择少量预测变量曲线来提高估计和预测精度。我们给出了 p 和 n 均趋于无穷时的渐近结果。大量模拟研究和两项真实数据分析表明，我们的方法表现良好；我们已将其实现于 `R`软件包 `FRegSigCom`中。

我们首先从具有 $p = 1$ 的模型入手，在第 2 节介绍最优分解和我们的估计方法。第 3 节将该方法扩展到具有 $p > 1$ 的模型，并在第 4 节讨论计算问题。我们分别在第 5 节和第 6 节开展模拟研究和真实数据分析。证明、算法以及其他公式和图表见补充材料。

2 单个函数型预测变量的回归

当 $p = 1$ 时，模型 (1.1) 中的系数矩阵 $\mathbb{B}(t)$ 变为一个 m 维向量，记为 $\mathbf{b}(t)$。此时，模型 (1.1) 变为

$$\mathbf{Y} = \boldsymbol{\mu} + \int_0^1 X(t)\mathbf{b}(t)dt + \boldsymbol{\varepsilon}. \tag{2.1}$$

2.1 t的最优分解

对于模型 (2.1)，我们考虑 $\sum_{k=1}^{K} \gamma_k(t)\mathbf{v}_k$ 这一形式的 $\mathbf{b}(t)$ 的所有分解，其中 $\gamma_k(t)$ 是函数，$\mathbf{v}_k$ 是 m 维向量。我们将确定这种形式下的最优分解，以便估计回归函数 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} = \boldsymbol{\mu} + \int_0^1 X(t)\mathbf{b}(t)dt$，并在总体层面预测响应。注意到，函数主成分回归和函数偏最小二乘中的估计具有这种分解形式；随后，我们将把这一最优分解与函数主成分分析和函数偏最小二乘中的分解进行比较。

令 $\Sigma(s,t) = \mathrm{E}[X(s)X(t)]$ 为 $X(t)$ 的协方差函数，并定义

$$\Gamma(s,t) = \int_0^1 \int_0^1 \Sigma(s,s')\mathbf{b}(s')^{\mathrm{T}}\mathbf{b}(t')\Sigma(t',t)ds'dt'. \tag{2.2}$$

令 $\alpha_1(t)$ 为下式的解

$$\max_{\alpha(t)} \int_0^1 \int_0^1 \alpha(s)\Gamma(s,t)\alpha(t)dsdt, \quad \text{subject to} \int_0^1 \int_0^1 \alpha(s)\Sigma(s,t)\alpha(t)dsdt = 1, \tag{2.3}$$

并令 $\alpha_k(t)$、$1 < k \leq K$ 为下式的解

$$\max_{\alpha(t)} \int_0^1 \int_0^1 \alpha(s)\Gamma(s,t)\alpha(t)dsdt, \quad \text{subject to} \int_0^1 \int_0^1 \alpha(s)\Sigma(s,t)\alpha(t)dsdt = 1$$
$$\text{and} \int_0^1 \int_0^1 \alpha_{k'}(s)\Sigma(s,t)\alpha(t)dsdt = 0, \quad 1 \leq k' \leq k-1. \tag{2.4}$$

式 (2.3) 和 (2.4) 定义了一系列广义特征值问题，其解 $\{\alpha_k(t)\ 1 \leq k \leq K\}$ 称为特征函数。由于 $\Sigma(s,\ t)$ 和 $\Gamma(s,\ t)$ 均为对称半正定核函数，因此 (2.3) 和 (2.4) 的最大值非负，记为 $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_K^2$。我们将 k 个特征函数 $\alpha_k(t)$ 对应的特征值称为 σ_k^2。定义一个 m 维向量 $\mathbf{w}_k = \int_0^1 \int_0^1 \alpha_k(s)\Sigma(s,\ t)\mathbf{b}(t)dsdt$，其中 $k \geq 1$。在下面的定理中，我们将证明 $\sum_{k=1}^{K} \alpha_k(t)\mathbf{w}_k$ 是最优分解。

定理 2.1。令 K_0 表示广义特征值问题 (2.4) 的正特征值总数，
$\sigma_1^2 \geq \cdots \geq \sigma_{K_0}^2 > 0 = \sigma_{K_0+1}^2 \cdots$ 表示按降序排列的特征值。$\alpha_k(t)$ 和 $\mathbf{w}_k$ 的定义同上。

(1). K_0 小于或等于 m，且$(\sigma_k^2, \mathbf{w}_k)$ 也是$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$的协方差矩阵的第 k个特征值—特征向量对，其中 $1 \leq k \leq K_0$。

(2). 向量 $\boldsymbol{\mu} + \int_0^1 X(t)\{\sum_{k=1}^K \alpha_k(t)\mathbf{w}_k\}dt$ 在逼近回归函数$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$时具有最小均方误差。具体而言，我们有

$$\mathrm{E}\left[\left\|\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - \boldsymbol{\mu} - \int_0^1 X(t)\left\{\sum_{k=1}^K \alpha_k(t)\mathbf{w}_k\right\}dt\right\|_2^2\right] = \sum_{k=K+1}^{K_0} \sigma_k^2 \tag{2.5}$$

$$= \min_{\substack{\boldsymbol{\nu}, \gamma_k(t), \mathbf{v}_k, \\ 1 \leq k \leq K}} \left\{\mathrm{E}\left[\left\|\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - \boldsymbol{\nu} - \int_0^1 X(t)\left\{\sum_{k=1}^K \gamma_k(t)\mathbf{v}_k\right\}dt\right\|_2^2\right]\right\},$$

其中，最小值是在所有平方可积函数 $\gamma_k(t)$'以及所有 m维向量 $\boldsymbol{\nu}$ 和$\mathbf{v}_k$'上取得的。

(3). 对于一个新观测$(X_{\mathrm{new}}(t), \mathbf{Y}_{\mathrm{new}})$，由 $\boldsymbol{\mu} + \int_0^1 X_{\mathrm{new}}(t)\{\sum_{k=1}^K \alpha_k(t)\mathbf{w}_k\}dt$ 给出的预测响应向量具有如下最小预测误差，

$$\mathrm{E}\left[\left\|\mathbf{Y}_{\mathrm{new}} - \boldsymbol{\mu} - \int_0^1 X_{\mathrm{new}}(t)\left\{\sum_{k=1}^K \alpha_k(t)\mathbf{w}_k\right\}dt\right\|_2^2\right] = \sum_{k=K+1}^{K_0} \sigma_k^2 + \mathrm{E}[\|\boldsymbol{\varepsilon}_{\mathrm{new}}\|_2^2] \tag{2.6}$$

$$= \min_{\substack{\boldsymbol{\nu}, \gamma_k(t), \mathbf{v}_k, \\ 1 \leq k \leq K}} \left\{\mathrm{E}\left[\left\|\mathbf{Y}_{\mathrm{new}} - \boldsymbol{\nu} - \int_0^1 X_{\mathrm{new}}(t)\left\{\sum_{k=1}^K \gamma_k(t)\mathbf{v}_k\right\}dt\right\|_2^2\right]\right\},$$

其中，$\boldsymbol{\varepsilon}_{\mathrm{new}}$ 是$\mathbf{Y}_{\mathrm{new}}$ 中的噪声向量，且与 $X_{\mathrm{new}}(t)$ 及原始变量相互独立 $(X(t), \mathbf{Y})$.

定理 2.1 表明，给定 K，$\boldsymbol{\mu} + \int_0^1 X(t)\left\{\sum_{k=1}^K \alpha_k(t)\mathbf{w}_k\right\}dt$ 在近似 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 时的优良程度可由 $\sum_{k>K} \sigma_k^2$ 衡量。因此，$\sigma_1^2 \geq \sigma_2^2 \geq \cdots$ 衰减得越快，为 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 提供充分近似所需的分量数 K 就越少。σ_k^2 的衰减速率受 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 中各坐标之间相关性的影响。一般而言，$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 中各坐标之间的相关性越强，σ_k^2 下降得越快。作为示例，我们取 $m = 20$，并考虑 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 的协方差矩阵的两种情形。情形 *1*：协方差矩阵的第 (i, j) 个元素为 $\rho^{|i-j|}$;

情形 *2*：协方差矩阵的所有对角元素均等于 1，其余元素均为 ρ，其中在两种情形下，$-1 \leq \rho \leq 1$。$|\rho|$ 的值越大，表明 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 各坐标之间的相关性越强。我们分别针对两种情形，在图1中绘制了相对近似误差 $\sum_{k=K+1}^{m} \sigma_k^2 / \sum_{k=1}^{m} \sigma_k^2$ 随 K 变化的曲线，结果表明，在两种情形下，$|\rho|$ 的值越大，相对近似误差衰减得越快。

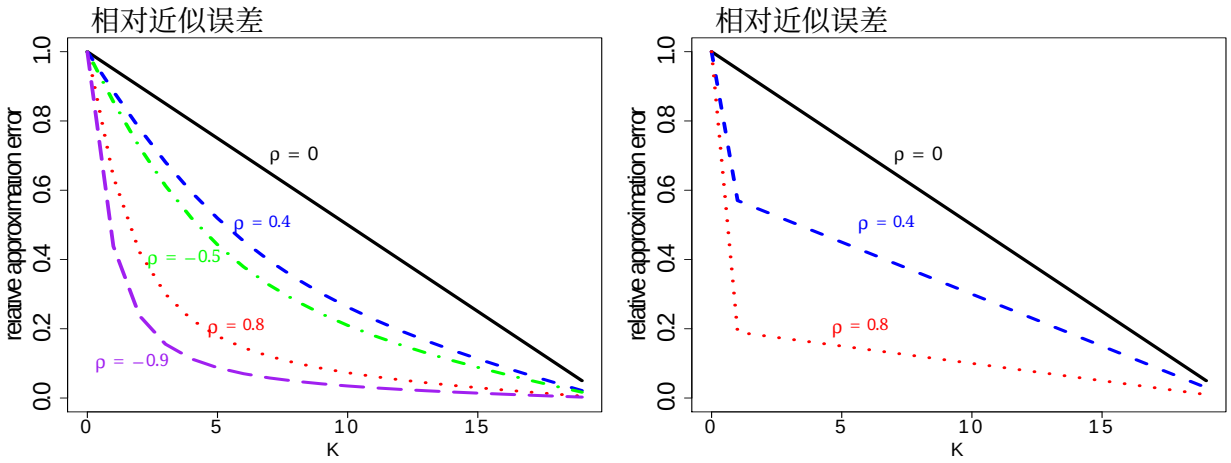

图1：b(t)的最优分解的相对近似误差随分量数 K 变化的两个示例，其中考虑了由ρ 表示的不同相关性水平。左图对应于$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 的协方差矩阵，其第(i, j)个元素等于 $\rho^{|i-j|}$；右图对应于对角线元素等于1且非对角线元素等于 ρ的协方差矩阵。

假设预测变量曲线的期望为零，则式 (2.6) 中的平方预测误差也等于预测向量所有坐标的方差之和。式 (2.6) 中的第一个等式表明，平方预测误差由两部分组成：$\sum_{k=K+1}^{K_0} \sigma_k^2$ 和 $\mathrm{E}[\|\boldsymbol{\varepsilon}_{\mathrm{new}}\|_2^2]$。第二部分是噪声向量 $\boldsymbol{\varepsilon}_{\mathrm{new}}$ 的方差，它独立于 $X_{\mathrm{new}}(t)$ 和原始数据 ($X(t$, $\mathbf{Y}$)。基于 $X_{\mathrm{new}}(t)$ 和 ($X(t$, $\mathbf{Y}$) 进行预测时，这部分无法减小。第一部分 $\sum_{k=K+1}^{K_0} \sigma_k^2$ 源于分解 $\sum_{k=1}^{K_0} \alpha_k(t)\mathbf{w}_k$ 在前 K 项之后的截断。增加分量的数量 K 可以减小这部分。然而，在实际应用中，我们必须估计 $\alpha_k(t)$ 和 $\mathbf{w}_k$。因此，除了截断误差和随机误差，实际中的预测误差还包括 $\alpha_k(t)$ 和$\mathbf{w}_k$的估计误差。尽管截断误差会减小，但较大的 K 会导致需要估计更多函数，从而带来额外的估计误差。在实际应用中，我们必须选择合适的 K，以平衡

截断导致的预测误差与估计导致的预测误差。

两种常用的降维方法——函数主成分回归（FPCR）和函数偏最小二乘（FPLS）——也会生成形如 $\sum_{k=1}^{K}\gamma_k(t)\mathbf{v}_k$ for b(*t*) 的分解。在 FPCR 中，$\gamma_k(t)$ 是预测变量曲线协方差的特征函数。FPLS 将偏最小二乘的思想扩展为构造与响应协方差最大的正交分量。偏最小二乘有多种变体（Wold，1973，1975；ter Braak 和 de Jong，1998；Boulesteix 和 Strimmer，2006），因此也提出了不同版本的 FPLS （Goutis 和 Fearn，1996；Reiss 和 Ogden，2007；Kr¨amer 等，2008；Kr¨amer 和 Sugiyama，2011；Delaigle 等，2012）。例如，软件 R中实现的 FPLS 基于 Kr¨amer等 （2008）以及 Kr¨amer 和 Sugiyama（2011）的研究，通过寻找不相关的潜变量，使$\mathbf{Y}$ 与潜变量之间的协方差平方最大化，同时令载荷向量的长度等于 1。在总体层面，这种 FPLS 求解

$$\max_{\alpha(t)} \quad \int_0^1\int_0^1 \alpha(s)\Gamma(s,t)\alpha(t)dsdt, \tag{2.7}$$
$$\text{subject to} \quad \|\alpha\|_{L^2}=1, \quad \text{and} \int_0^1\int_0^1 \alpha_{k'}(s)\Sigma(s,t)\alpha(t)dsdt=0, \quad 1\le k'\le k-1.$$

问题（2. 4）和（2.7）共享一个目标函数，可表示为 s

$$\int_0^1\int_0^1 \alpha(s)\Gamma(s,t)\alpha(t)dsdt=\mathrm{E}\left[\int_0^1 X(t)\alpha_k(t)dt(\mathbf{Y}-\boldsymbol{\mu})\right]^{\mathrm{T}}\mathrm{E}\left[\int_0^1 X(t)\alpha_k(t)dt(\mathbf{Y}-\boldsymbol{\mu})\right], \tag{2.8}$$

与每个响应坐标的协方差平方和。共同约束 $0=\int_0^1\int_0^1\alpha_{k'}(s)\Sigma(s,t)\alpha(t)dsdt=\mathrm{Cov}(\int_0^1 X(t)\alpha_k(t)dt,\int_0^1 X(t)\alpha_{k'}(t)dt)$意味着，对于 $k\neq k'$，$\int_0^1 X(t)\alpha_k(t)dt$ 与 $\int_0^1 X(t)\alpha_{k'}(t)dt$ 不相关。但式（2.7）中的函数偏最小二乘要求 α_k 具有单位长度，而式（2.7）的区域 $\{\alpha(t):\|\alpha\|_{L^2}=1\}$ 是 $L^2[0,1]$中的单位球面；而式（2.4）要求$\int_0^1 X(t)\alpha_k(t)dt$ 具有单位方差，因为

$$\mathrm{Var}\left(\int_0^1 X(t)\alpha_k(t)dt\right)=\mathrm{E}\left[\left\{\int_0^1 X(t)\alpha_k(t)dt\right\}^2\right]=\int_0^1\int_0^1\alpha(s)\Sigma(s,t)\alpha(t)dsdt=1,$$

且区域 $\{\alpha(t):\int_0^1\int_0^1\alpha(s)\Sigma(s,t)\alpha(t)dsdt=1\}$ 是一个椭球。因此，(2.4) 和 (2.7) 的解通常不同。在图2中，我们给出一个示例来说明最优分解与函数主成分回归和函数偏最小二乘所对应分解之间的差异。设 $X(t)$为 $[0,1],$ 上的布朗运动，其协方差函数为 $\Sigma(t',t)=\min(t',t)$。设

$\mathbf{b}(s) = (b_1(s)\ldots, b_5(s))^{\mathrm{T}}$ 是一个五维向量，其中 $b_k(s) = \sin(k\pi t) + 2\sin((k+1)\pi t)$，适用于 $1 \leq k \leq 5$。我们在图2中绘出了分解所用的前三个分量函数，以及三种分解方法下相对于$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 的近似误差随分量数量 K 变化的情况。图中显示，我们的分量与函数主成分回归和函数偏最小二乘中的分量有很大不同。对于每个 K，最优分解都能使$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$的近似误差最小。当 $K = 1$时，最优分解的近似误差不到另外两种方法中任一种的一半。当 $K = 5$ （响应的维度）时，最优分解的近似误差为零，而另外两种方法仍有正的误差。事实上，定理 2.1 表明，当 K 不超过响应的维度时，最优分解能够对$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 实现精确近似。这一事实为我们在实践中选择最优 K提供了一个上界。

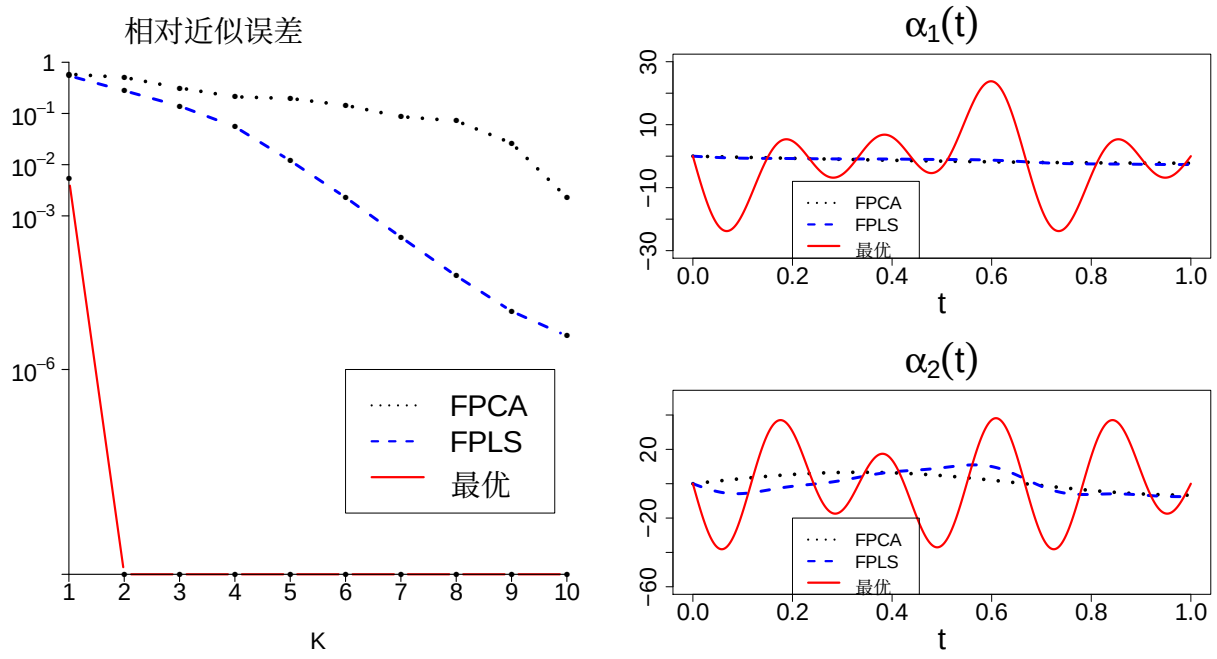

图2：比较通过三种方法得到的分量 $\alpha_k(t)$：函数主成分分析、函数偏最小二乘和我们的最优分解。为便于比较，所有分量均经过缩放，使得$\int_0^1 \int_0^1 \alpha(s)\Sigma(s,\ t)\alpha(t)dsdt = 1$。左上：第一个分量 $\alpha_1(t)$。右上：第二个分量。左下：第三个分量。右下：三种方法的相对近似误差$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$ 与分量数量 K 之间的关系。

2.2 最优分解的估计

设 $\{\mathbf{Y}_\ell = (Y_{\ell 1}, \ldots, Y_{\ell m})^{\mathrm{T}}, X_\ell(t) : 1 \le \ell \le n\}$ 是来自模型(2.1)的 n 独立样本，并且 $\boldsymbol{\varepsilon}_\ell$ 表示 $\mathbf{Y}_\ell$ 中的噪声。令 $\mathbf{Y}$ 和 $X(t)$ 表示样本均值，并令 $\Sigma(s,t) = \sum_{\ell=1}^{n} \{X_\ell(s) - X(s)\} \{X_\ell(t) - X(t)\}/n$ 表示预测变量的样本协方差函数。令 $\widehat{\Gamma}(s,t) = \left[\sum_{l=1}^{n} \{X_l(s) - X(s)\} \{\mathbf{Y}_l - \mathbf{Y}\}^{\mathrm{T}}\right] \left[\sum_{\ell'=1}^{n} \{\mathbf{Y}_{\ell'} - \mathbf{Y}\} \{X_{\ell'}(t) - X(t)\}\right] / n^2$，为说明 $\widehat{\Gamma}(s,t)$ 是 $\Gamma(s,t)$ 的一个估计量，将 $\mathbf{Y}_l = \boldsymbol{\mu} + \int_0^1 X_l(t)\mathbf{b}(t)dt + \varepsilon_\ell$ 代入 $\Gamma(s,t)$ 的定义，可得

$$\begin{aligned}\widehat{\Gamma}(s,t) = &\int_0^1 \int_0^1 \widehat{\Sigma}(s,s')\mathbf{b}(s')^{\mathrm{T}}\mathbf{b}(t')\widehat{\Sigma}(t',t)ds'dt' \\ &+ \frac{2}{n}\sum_{\ell'=1}^{n}\{\boldsymbol{\varepsilon}_{\ell'} - \overline{\boldsymbol{\varepsilon}}\}\left\{X_{\ell'}(t) - \overline{X}(t)\right\}\int_0^1 \widehat{\Sigma}(s,s')\mathbf{b}(s')^{\mathrm{T}}ds' \\ &+ \frac{1}{n^2}\left[\sum_{l=1}^{n}\left\{X_l(s) - \overline{X}(s)\right\}\{\boldsymbol{\varepsilon}_l - \overline{\boldsymbol{\varepsilon}}\}^{\mathrm{T}}\right]\left[\sum_{\ell'=1}^{n}\{\boldsymbol{\varepsilon}_{\ell'} - \overline{\boldsymbol{\varepsilon}}\}\left\{X_{\ell'}(t) - \overline{X}(t)\right\}\right],\end{aligned}$$

其中，最后两项由于大数定律以及 $X_l(s)$ 与 $\boldsymbol{\varepsilon}_l$ 相互独立，依概率收敛于零，$n \to \infty$。因此，$\Gamma(s,t)$ 是 $\Gamma(s,t)$ 的一致估计量。由于广义特征值问题（2.4）等价于以下瑞利商最大化问题，

$$\max_{\alpha(t)} \frac{\int_0^1 \int_0^1 \alpha(s)\Gamma(s,t)\alpha(t)dsdt}{\int_0^1 \int_0^1 \alpha(s)\Sigma(s,t)\alpha(t)dsdt}, \quad \text{subject to} \int_0^1 \int_0^1 \alpha(s)\Sigma(s,t)\alpha(t)dsdt = 1$$
$$\text{and} \int_0^1 \int_0^1 \alpha_{k'}(s)\Sigma(s,t)\alpha(t)dsdt = 0, \quad 1 \le k' \le k-1,$$

我们提出通过求解以下问题，在最优分解中得到 $\alpha_k(t)$ 的估计值 $\widehat{\alpha}_k(t)$

$$\max_{\alpha(t)} \frac{\int_0^1 \int_0^1 \alpha(s)\widehat{\Gamma}(s,t)\alpha(t)dsdt}{\int_0^1 \int_0^1 \alpha(s)\widehat{\Sigma}(s,t)\alpha(t)dsdt + P(\alpha)}, \quad \text{subject to} \int_0^1 \int_0^1 \alpha(s)\widehat{\Sigma}(s,t)\alpha(t)dsdt = 1$$
$$\text{and} \int_0^1 \int_0^1 \widehat{\alpha}_{k'}(s)\widehat{\Sigma}(s,t)\alpha(t)dsdt = 0, \quad 1 \le k' \le k-1, \tag{2.9}$$

其中，$P(\alpha) = \tau\left(\|\alpha\|_{L^2}^2 + \eta\|\alpha''\|_{L^2}^2\right)$，$\alpha''(t)$ 表示 $\alpha(t)$ 的二阶导数，而 $\|\cdot\|_{L^2}$ 表示函数的 L^2 范数。在函数型数据分析中，人们常采用式（2.9）中的这种惩罚方式 $P(\alpha)$。例如，在正则化函数主成分分析（Ramsay 和 Silverman（2005）第 9.3.1 节）中，惩罚目标函数

$\int\int \xi(s)\widehat{\Sigma}(s,t)\xi(t)dsdt/(\|\xi\|_{L^2}^2 + \lambda\|\xi''\|_{L^2}^2)$，是在所有光滑函数 $\xi(s)$ 上最大化的。在目标函数的分母中加入惩罚项后，对于较粗糙且 $\|\xi''\|_{L^2}^2$ 值较大的 $\xi(s)$，目标函数的值会较小。因此，该目标函数的最大化解是一个光滑函数； λ 越大，最大化解就越光滑。在我们的惩罚项 $P(\alpha)$ 中， $\|\alpha''\|_{L^2}^2$ 这一项是光滑性惩罚项，发挥着与正则化函数主成分分析中相同的作用。另一方面，由于 $\Sigma(s,t)$ 是根据有限数量的样本估计得到的，因此它只有有限个正特征值。如果我们只在式（2.9）中加入光滑性惩罚项，式（2.9）的解可能不唯一，而且解 $\widehat{\alpha}_k(s)$ 的范数可能远大于 $\alpha_k(t)$，这与经典线性模型中由多重共线性导致的过拟合问题类似。受岭惩罚的启发（岭惩罚用于缩减估计系数并减轻多重共线性的影响），我们在惩罚项中加入 $\tau\|\alpha\|_{L^2}^2$ ，以控制 α 的范数。参数 η 用于调整惩罚项中 $\|\alpha\|_{L^2}^2$ 与 $\|\alpha''\|_{L^2}^2$ 的相对重要性。

假设我们已得到估计值 $\widehat{\alpha}_k(t)$、 $1 \le k \le K$。为了估计最优分解中的截距向量 $\boldsymbol{\mu}$ 以及向量 $\mathbf{w}_k$，我们定义$T_{\ell,\ k} = \int_0^1 \{X_\ell(t) - X(t)\}\widehat{\alpha}_k(t)dt$，其中 $1 \le \ell \le n$ 和 $1 \le k \le K$，并令 $\widehat{\mathbf{T}}_k = (\widehat{T}_{1,k}, \ldots, \widehat{T}_{n,k})^{\mathrm{T}}$。然后，通过使用以下最小二乘问题，将 $\mathbf{Y}_\ell$ 对新标量预测变量 $\mathbf{T}_k$ 进行回归，即可得到估计值 $\{\widehat{\boldsymbol{\mu}}, \widehat{\mathbf{w}}1, \ldots, \widehat{\mathbf{w}}_K\}$，

$$\min_{\boldsymbol{\nu},\mathbf{v}_k,1\le k\le K} \left\{ \sum_{\ell=1}^{n} \left\| \mathbf{Y}_\ell - \boldsymbol{\nu} - \sum_{k=1}^{K} \widehat{T}_{\ell,k}\mathbf{v}_k \right\|_2^2 \right\}, \tag{2.10}$$

其中，最小值是在所有 m 维向量 $\boldsymbol{\nu}$ 和 $\mathbf{v}_k$ 上取得的。由于$\mathbf{T}_k$ 相互正交，(2.10) 的解可以明确写为 $\widehat{\boldsymbol{\mu}} = \mathbf{Y}$ 和 $\widehat{\mathbf{w}}_k = \sum_{\ell=1}^{n} T_{\ell,\ k}\mathbf{Y}_\ell/n$，其中 $1 \le k \le K$。最后，我们用 $\widehat{\boldsymbol{\mu}}_{\mathbf{Y}|\mathbf{X}} = \widehat{\boldsymbol{\mu}} + \int_0^1 X(t)\{\sum_{k=1}^{K} \widehat{\alpha}_k(t)\widehat{\mathbf{w}}_k\}dt$ 估计 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$。给定函数型预测变量的一个新观测 $X_{\mathrm{new}}(t)$，我们将响应向量预测为$\mathbf{Y}_{\mathrm{pred}} = \widehat{\boldsymbol{\mu}} + \int_0^1 X_{\mathrm{new}}(t)\{\sum_{k=1}^{K} \widehat{\alpha}_k(t)\widehat{\mathbf{w}}_k\}dt$。

3 使用多个函数型预测变量进行回归

对于具有多个函数型预测变量（$p > 1$）的模型（1.1），我们首先给出总体层面系数函数的 $p \times m$ 矩阵 $\mathbb{B}(t)$ 的最优分解。

然后，我们针对两种情况分别估计这一分解：较小的 p和较大的 p。

3.1 $\mathbb{B}(t)$的最优分解

将第 2.1 节中的分解加以推广，对于模型 (1.1) 中的 $p > 1$，我们考虑将 $\mathbb{B}(t)$ 分解为以下形式：$\sum_{k=1}^{K} \boldsymbol{\gamma}_k(t)\mathbf{v}_k^{\mathrm{T}}$，其中 $\boldsymbol{\gamma}_k(t) = (\gamma_{k1}(t) \ldots, \gamma_{kp}(t))^{\mathrm{T}}$ 是一个p维函数向量，而 $\mathbf{v}_k$ 是一个 m维向量。为求得此类形式中的最优分解，我们定义两个 $p \times p$ 核函数矩阵

$$\boldsymbol{\Sigma}(s,t) = \mathrm{E}[\mathbf{X}(s)\mathbf{X}(t)^{\mathrm{T}}], \quad \boldsymbol{\Gamma}(s,t) = \int_0^1 \int_0^1 \boldsymbol{\Sigma}(s,s')\mathbb{B}(s')\mathbb{B}(t')^{\mathrm{T}}\boldsymbol{\Sigma}(t',t)ds'dt', \tag{3.1}$$

它们分别推广了第 2.1 节中的标量函数 $\Sigma(s,\ t)$ 和 $\Gamma(s,\ t)$。这里，$\boldsymbol{\Sigma}(s,\ t)$ 是 $\mathbf{X}(t)$ 的协方差函数矩阵。令 $\boldsymbol{\alpha}_k(t) = (\alpha_{k1}(t) \ldots, \alpha_{kp}(t))^{\mathrm{T}}$,$k \geq 1$ 表示以下广义特征值问题的特征函数：

$$\max_{\boldsymbol{\alpha}(t)} \int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}}\boldsymbol{\Gamma}(s,t)\boldsymbol{\alpha}(t)dsdt, \quad \text{subject to} \int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}}\boldsymbol{\Sigma}(s,t)\boldsymbol{\alpha}(t)dsdt = 1$$
$$\text{and} \int_0^1 \int_0^1 \boldsymbol{\alpha}_{k'}(s)^{\mathrm{T}}\boldsymbol{\Sigma}(s,t)\boldsymbol{\alpha}(t)dsdt = 0, \quad 1 \leq k' \leq k-1, \tag{3.2}$$

并定义用于 $k \geq 1$ 的 m 维向量 $\mathbf{w}_k = \int_0^1 \int_0^1 \mathbb{B}(t)^{\mathrm{T}}\boldsymbol{\Sigma}(s,\ t)\boldsymbol{\alpha}_k(s)dsdt$。与定理 2.1 中一样，我们可以证明，在逼近回归函数和进行预测时，$\sum_{k=1}^{K} \boldsymbol{\alpha}_k(t)\mathbf{w}_k^{\mathrm{T}}$ 是 $\mathbb{B}(t)$ 的最优分解。

3.2 最优分解的估计

设 $\mathbf{Y}_\ell = (Y_{\ell 1}, \ldots, Y_{\ell m})^{\mathrm{T}}$ 和 $\mathbf{X}_\ell(t) = (X_{\ell 1}(t) \ldots, X_{\ell p}(t))^{\mathrm{T}}$, $1 \leq \ell \leq n$是来自模型（1.1）的 n 独立样本。我们通过以下方式估计矩阵 $\boldsymbol{\Sigma}(s,t)$ 和 $\boldsymbol{\Gamma}(s,t)$：

$$\widehat{\boldsymbol{\Sigma}}(s,t) = \frac{1}{n}\sum_{\ell=1}^{n} \left\{\mathbf{X}_\ell(s) - \overline{\mathbf{X}}(s)\right\}\left\{\mathbf{X}_\ell(t) - \overline{\mathbf{X}}(t)\right\}^{\mathrm{T}}, \tag{3.3}$$
$$\widehat{\boldsymbol{\Gamma}}(s,t) = \frac{1}{n^2}\left[\sum_{l=1}^{n} \left\{\mathbf{X}_l(s) - \overline{\mathbf{X}}(s)\right\}\left\{\mathbf{Y}_l - \overline{\mathbf{Y}}\right\}^{\mathrm{T}}\right]\left[\sum_{\ell=1}^{n}\left\{\mathbf{Y}_\ell - \overline{\mathbf{Y}}\right\}\left\{\mathbf{X}_\ell(t) - \overline{\mathbf{X}}(t)\right\}^{\mathrm{T}}\right]$$

分别地，其中 $\mathbf{X}(t) \hat{=} \sum_{\ell=1}^{n} \mathbf{X}_\ell(t)/n$ 和 $\mathbf{Y} = \sum_{\ell=1}^{m} \mathbf{Y}_\ell/n$。我们的估计量 $\widehat{\boldsymbol{\alpha}}_k(t)$ 是以下惩罚广义特征值问题的解，

$$\max_{\boldsymbol{\alpha}(t)} \frac{\int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}} \widehat{\boldsymbol{\Gamma}}(s,t) \boldsymbol{\alpha}(t) ds dt}{\int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}(s,t) \boldsymbol{\alpha}(t) ds dt + P(\boldsymbol{\alpha})}, \quad \text{subject to} \int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}(s,t) \boldsymbol{\alpha}(t) ds dt = 1$$
$$\text{and} \int_0^1 \int_0^1 \widehat{\boldsymbol{\alpha}}_{k'}(s)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}(s,t) \boldsymbol{\alpha}(t) ds dt = 0, \quad 1 \le k' \le k-1, \tag{3.4}$$

其中，$P(\boldsymbol{\alpha})$ 是施加于 $\boldsymbol{\alpha}(s) = (\alpha_1(s) \ldots, \alpha_p(s))^{\mathrm{T}}$ 上的惩罚项。我们分别给出以下两种情况下 $P(\boldsymbol{\alpha})$ 的形式。

情况 *1*：p 相对较小。我们考虑以下惩罚项，它是对式（2.9）中惩罚项的直接推广：

$$P(\boldsymbol{\alpha}) = \tau \sum_{j=1}^{p} \|\alpha_j\|_\eta^2, \quad \text{where} \quad \|\alpha_j\|_\eta^2 = \|\alpha_j\|_{L^2}^2 + \eta \|\alpha_j''\|_{L^2}^2. \tag{3.5}$$

情形 2：p相对较大。我们提出一种惩罚项，同时控制$\widehat{\ }$的平滑性和 $\boldsymbol{\alpha}(t)$的稀疏性。在第 2.2 节中，我们已经证明，当 $p = 1$时，$\boldsymbol{\Lambda}(s,t)$ 和$\sim\boldsymbol{\Sigma}(s,t)$ 分别是 $\boldsymbol{\Lambda}(s,t)$ 和 $\boldsymbol{\Sigma}(s,t)$ 的一致估计量。该结果可以推广至固定 $p > 1$ 的情形，此时 $n \to \infty$。然而，实际中存在这样的情况：预测变量曲线的数量 p可能远大于样本量 n。例如，在神经科学中，可以获得大脑中数千个体素的时间序列曲线。在这种情况下，更适合考虑当 $p, n \to \infty$时的渐近性质。当 $p, n \to \infty$时，尽管 $\boldsymbol{\Lambda}(s,t)$ 和$\sim\boldsymbol{\Sigma}(s,t)$ 的每个分量分别收敛到 $\boldsymbol{\Lambda}(s,t)$ 和 $\boldsymbol{\Sigma}(s,t)$ 的对应分量，但对于一般的 $\boldsymbol{\alpha}(t)$，$\int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}} \widehat{\boldsymbol{\Lambda}}(s,t) \boldsymbol{\alpha}(t) ds dt$ 和 $\int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}(s,t) \boldsymbol{\alpha}(t) ds dt$ 未必分别收敛到$\int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}} \boldsymbol{\Lambda}(s,t) \boldsymbol{\alpha}(t) ds dt$ 和 $\int_0^1 \int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}} \boldsymbol{\Sigma}(s,t) \boldsymbol{\alpha}(t) ds dt$。受同步脑电图-功能性磁共振成像数据的启发（其中脑电图特征仅与少数体素的时间序列曲线相关），我们将研究范围限制在“稀疏”的 $\boldsymbol{\alpha}(t)$。对于标量向量，衡量稀疏性的一种方法是其 l_1 范数；稀疏意味着其大多数坐标的绝对值都很小，其中也包括大多数坐标为零的情形。我们将这一概念推广到函数向量。如果大多数坐标函数 $\alpha_j(t)$的 L^2 范数都很小，我们就称 $\boldsymbol{\alpha}(t) = (\alpha_1(t) \ldots, \alpha_p(t))^{\mathrm{T}}$ 是稀疏的。对于 $\boldsymbol{\alpha}(t)$，我们

定义两个范数

$$\|\boldsymbol{\alpha}\|_{L^2,1} = \sum_{j=1}^{p} \|\alpha_j\|_{L^2} \quad \text{and} \quad \|\boldsymbol{\alpha}\|_{L^2,2} = \sqrt{\sum_{j=1}^{p} \|\alpha_j\|_{L^2}^2}, \tag{3.6}$$

其中，$\|\boldsymbol{\alpha}\|_{L^2,\,2}$ 是常见的 L^2 型范数。 L^1 型范数 $\|\boldsymbol{\alpha}\|_{L^2,\,1}$ 可用于衡量 $\boldsymbol{\alpha}(t)$ 的稀疏性。举例来说，我们考虑 $\boldsymbol{\alpha}(t)$ 的三种情况，其中 $p = 1000$: (1).$\|\alpha_j\|_{L^2}^2 = 0.001$; (2). $\|\alpha_j\|_{L^2}^2 = c_1/j$; (3).$\|\alpha_j\|_{L^2}^2 = c_2/j^2$，其中 $1 \le j \le 1000$，且 c_1和 c_2 为常数，使得三种情况下的 L^2 范数 $\|\boldsymbol{\alpha}\|_{L^2,\,2}$ 均为 1。根据图 3 中绘制的三种情况下 $\|\alpha_j\|_{L^2}^2$ 与 j 的关系可以看出，第三种情况下 $\|\alpha_j\|_{L^2}$ 衰减最快，且在三种情况中最稀疏；第二种情况比第一种情况更稀疏。这三种情况下 $\|\boldsymbol{\alpha}\|_{L^2,\,1}$ 的值分别为 31.62、22.59 和 5.83，这表明了这些向量的稀疏程度。论证

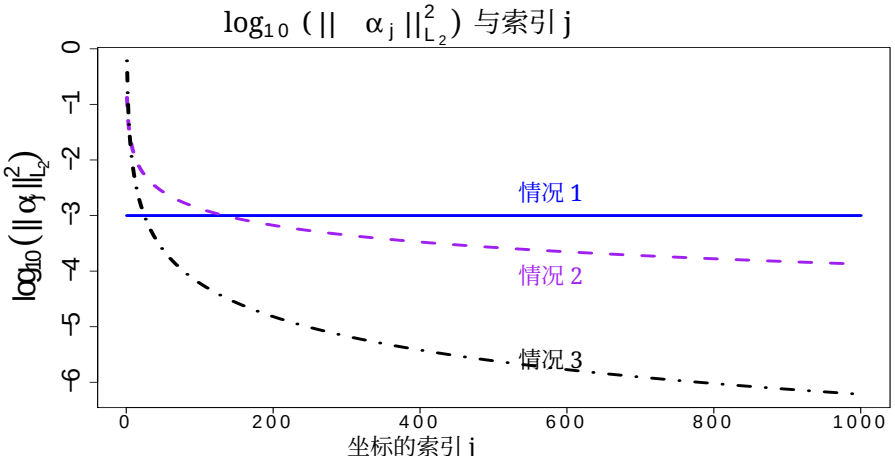

图 3：三个情形下 $\log_{10}(\|\alpha_j\|_{L^2})$ 与 j 的关系图，分别对应 $\|\alpha_j\|_{L^2}^2 = 0.001$、$c_1/j$、$c_2/j^2$; c_1 和 c_2 为常数，且 $\|\boldsymbol{\alpha}\|_{L^2,\,2} = 1$。

第 3.3 节中渐近结果的证明表明，若 $\|\boldsymbol{\alpha}\|_{L^2,1}$ 相对于 $\sqrt{n}/\log\, p$, $\int_0^1\int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}}\widehat{\boldsymbol{\Lambda}}(s,t)\boldsymbol{\alpha}(t)$ $dsdt$ 较小，且 $\int_0^1\int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}(s,t)\boldsymbol{\alpha}(t)dsdt$ 分别收敛到 $\int_0^1\int_0^1 \boldsymbol{\alpha}(s)^{\mathrm{T}}\boldsymbol{\Lambda}(s,t)\boldsymbol{\alpha}(t)dsdt$ 和 $\int_0^1\int_0^1 \boldsymbol{\alpha}(s)$ $^{\mathrm{T}}\boldsymbol{\Sigma}(s,t)\boldsymbol{\alpha}(t)dsdt$。另一方面，若我们假设 $\mathbb{B}(t)$ 是稀疏的（例如， $\mathbb{B}(t)$ 中的大多数 $\mathbf{b}_j(t)$ 为零），则可以证明 $\mathbb{B}(t)$ 的最优分解中的 $\boldsymbol{\alpha}_k(t)$ 具有较小的 $\|\cdot\|_{L^2,1}$ 范数，其中 $\mathbb{B}(t)$ 的稀疏性将在随后的第 3.3 节中定义。因此，我们首先考虑施加 Lasso 型惩罚 $\tau\|\boldsymbol{\alpha}\|_{L^2,1}^2$，以控制 $\boldsymbol{\alpha}(t)$ 的稀疏性并确保一致性

我们估计量的。然而，由于 $\tau\|\boldsymbol{\alpha}\|_{L^2,1}^2$ 是$\boldsymbol{\alpha}$的凸函数而非严格凸函数，从而导致求解优化问题时出现计算困难，我们将$\tau\|\boldsymbol{\alpha}\|_{L^2,1}^2$ 改为严格凸的弹性网型惩罚 $\tau\left[(1-\lambda)\|\boldsymbol{\alpha}\|_{L^2,}^2\right.$ $\left._2+\lambda\|\boldsymbol{\alpha}\|_{L^2,1}^2\right]$。此外，我们通过将$\|\alpha_j\|_{L^2}^2$ 替换为 $\|\alpha_j\|_\eta^2=\|\alpha_j\|_{L^2}^2+\eta\|\alpha_j''\|_{L^2}^2$ 来控制 $\boldsymbol{\alpha}(t)$ 中坐标函数的光滑性。总之，我们提出如下同时光滑稀疏惩罚

$$P(\boldsymbol{\alpha})=\tau\left\{(1-\lambda)\left(\sum_{j=1}^p\|\alpha_j\|_\eta^2\right)+\lambda\left(\sum_{j=1}^p\|\alpha_j\|_\eta\right)^2\right\},\tag{3.7}$$

其中 $\eta\geq 0$控制光滑程度，$0\leq\lambda\leq 1$ 调节惩罚项中 L_1 部分$\sum_{j=1}^p\|\alpha_j\|_\eta$与 L_2 部分 $\sum_{j=1}^p\|\alpha_j\|_\eta^2$的比例，$\tau\geq 0$ 控制整个惩罚项的强度。τ 和 λ 类似于高维数据分析中弹性网惩罚（Zou and Hastie，2005）的调优参数。

在实践中，p是否较大需视具体情况而定。在我们的实证研究中，我们发现当 $p\leq 5$ 时，两种惩罚项在预测性能上没有显著差异，但采用光滑惩罚（3.5）的方法比另一种快得多。当 $p>5$时，在模拟和真实数据分析中，使用同时光滑稀疏惩罚（3.7）可观察到预测精度的显著提升。

在这两种情况下，按照第 2.2 节所述，在获得估计值 $\widehat{\boldsymbol{\alpha}}_k(t)$、$1\leq k\leq K$ 后，我们定义新的标量预测变量：$\widehat{\mathrm{T}}_k=\left(\int_0^1\{\mathrm{X}_1(t)-\mathrm{X}(t)\}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(t)dt,\cdots\cdots,\int_0^1\{\mathrm{X}_n(t)-\mathrm{X}(t)\}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(t)dt\right)^{\mathrm{T}}$，并使用与第 2 节相同的方法获得估计值 $\widehat{\boldsymbol{\mu}}$ 和 $\widehat{\mathrm{w}}_k$'s。最后，我们通过 $\widehat{\boldsymbol{\mu}}_{\mathbf{Y}|\mathbf{X}}=\widehat{\boldsymbol{\mu}}+\sum_{k=1}^K\left[\int_0^1\{\mathrm{X}(t)-\mathrm{X}(t)\}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(t)dt\right]\widehat{\mathrm{w}}_k$ 估计 $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$。给定函数型预测变量的新观测 $\mathrm{X}_{\mathrm{new}}(t)$，我们将响应向量预测为

$$\mathbf{Y}_{\mathrm{pred}}=\widehat{\boldsymbol{\mu}}+\sum_{k=1}^K\left[\int_0^1\{\mathbf{X}_{\mathrm{new}}(t)-\overline{\mathbf{X}}(t)\}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_k(t)dt\right]\widehat{\mathbf{w}}_k.\tag{3.8}$$

3.3 渐近结果

我们考虑样本量 n 和预测曲线数量 p均趋于无穷、但响应向量 $\mathbf{Y}$ 的维度 m 保持固定时的渐近结果。令

$$\mathbb{Y}=[\mathbf{Y}_1,\ldots,\mathbf{Y}_n]^{\mathrm{T}},\quad\mathbb{X}(t)=[\mathbf{X}_1(t),\ldots,\mathbf{X}_n(t)]^{\mathrm{T}},\quad\mathbb{E}=[\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n]^{\mathrm{T}}$$

分别表示 $n \times m$ 响应矩阵、 $n \times p$ 函数型预测变量矩阵和 $n \times m$ 噪声矩阵。模型 (1.1) 的样本版本可写为

$$\mathbb{Y} = \mathbb{F} + \mathbb{E}, \quad \text{where} \quad \mathbb{F} = \mathbf{1}_n \boldsymbol{\mu}^{\mathrm{T}} + \int_0^1 \mathbb{X}(t)\mathbb{B}(t)dt \tag{3.9}$$

是回归函数的数据矩阵， $\mathbf{1}_n$ 是一个 n维的全 1 向量。我们将 $\mathbb{F}$估计为

$$\widehat{\mathbb{F}} = \mathbf{1}_n \widehat{\boldsymbol{\mu}}^{\mathrm{T}} + \sum_{k=1}^{K} \left[\int_0^1 \left\{ \mathbb{X}(t) - \mathbf{1}_n \overline{\mathbf{X}}(t)^{\mathrm{T}} \right\} \widehat{\boldsymbol{\alpha}}_k(t) dt \right] \widehat{\mathbf{w}}_k^{\mathrm{T}}. \tag{3.10}$$

我们将给出估计误差 $\|\widehat{\mathbb{F}} - \mathbb{F}\|_F$的收敛速率，其中 $\|\cdot\|_F$ 为弗罗贝尼乌斯范数。此外，对于一个新观测（$\mathbf{X}_{\mathrm{new}}(t)$，$\mathbf{Y}_{\mathrm{new}}$），我们将给出预测误差 $\|\mathbf{Y}_{\mathrm{pred}} - \mathbf{Y}_{\mathrm{new}}\|_2$的收敛速率，其中$\mathbf{Y}_{\mathrm{pred}}$ 由式（3.8）给出。我们提出两个正则条件。

（C1）。对于任意整数 $l \geq 2$,

$$\mathrm{E}[\|\boldsymbol{\varepsilon}\|_2^{2l}] \leq l! 2^{l-1} (\sigma_\epsilon^2)^l, \quad \max_{0 \leq t \leq 1} \max_{1 \leq j \leq p} \mathrm{E}[X_j(t)^{2l}] \leq l! 2^{l-1},$$

其中 σ_ϵ^2 是 $\boldsymbol{\varepsilon}$所有坐标的方差之和。

（C1）中的第一个不等式对任意多元正态分布均成立，且 σ_ϵ^2 是噪声水平的一种度量。（C1）中的第二个不等式在以下情况下成立：若 $X_j(t)$是一个高斯过程，则对任意 $1 \leq j \leq p$ 和 $0 \leq t \leq 1$，其方差均不大于 1。如果高斯过程 $X_j(t)$在某些时间点的方差大于 1，我们可以对 $X_j(t)$进行缩放，使其方差不大于 1，并相应地缩放 $\mathbb{B}(t)$以保持模型不变；对于缩放后的 $X_j(t)$和 $\mathbb{B}(t)$，我们的主要结果仍然成立。对于系数矩阵$\mathbb{B}(t) = [\mathbf{b}_1(t), \ldots, \mathbf{b}_p(t)]^{\mathrm{T}}$，我们定义一种 L_1 型范数

$$\|\mathbb{B}\|_{L^2,1} = \sum_{j=1}^{p} \|\mathbf{b}_j\|_{L^2,2} = \sum_{j=1}^{p} \sqrt{\sum_{i=1}^{m} \|b_{ji}\|_{L^2}^2}, \tag{3.11}$$

用于衡量 $\{\mathbf{b}_1(t), \ldots, \mathbf{b}_p(t)\}$ 的稀疏性。

(C2). $\lim_{n\to\infty} \sqrt{\frac{\text{对数}}{n}}^{p} = 0$和 $\lim_{n\to\infty} \left(\sigma_\epsilon^2 + \|\mathbb{B}\|_{L^2,1}^2\right) \sqrt{\frac{\text{对数}\, p}{n}} = 0.$

该条件意味着 p 可以远大于 n，但其增长速度应慢于e^n。噪声水平 σ_ϵ^2 与系数函数的稀疏性$\|\mathbb{B}\|_{L^2,1}^2$的增长速度均应慢于 $\sqrt{n}/\log\ p$。

根据定理 2.1，我们只需要考虑少数几个 σ_k^2 相对较大的分量。令

$$K = \max\left\{k : \sigma_k^2 \geq \delta\left(\sigma_\epsilon^2 + \|\mathbb{B}\|_{L^2,1}^2\right)\sqrt{\frac{\log p}{n}}\right\} \tag{3.12}$$

其中 δ 是不依赖于 n 的常数。

定理 3.1。假设条件（*C1*）和（*C2*）成立，且 K 根据(3.12)选取，其中 δ 足够大。假设调优参数满足

$$\tau = M\sqrt{\frac{\log p}{n}}, \quad \lambda_0 \leq \lambda \leq 1, \quad 0 \leq \eta \leq \min_{1\leq k\leq K}\{\|\boldsymbol{\alpha}_k\|_{L^2,1}^2/\|\boldsymbol{\alpha}_k''\|_{L^2,1}^2\}, \tag{3.13}$$

其中 $0 < \lambda_0 < 1$ 是常数，M 是足够大的常数。λ_0 和 M 均不依赖于 n。假设存在一个不依赖于 n 的常数 $c_1 > 0$，使得

$$\frac{\sigma_k^2 - \sigma_{k+1}^2}{\sigma_k^2} \geq c_1, \quad \textit{for all } n \textit{ and } 1 \leq k \leq K. \tag{3.14}$$

那么，对于任意 $n \geq 1$，存在一个事件 Ω_n，满足 $P(\Omega_n) \geq 1 - 12p^{-1}$，使得在 Ω_n 中，我们有

$$\frac{1}{n}\|\widehat{\mathbb{F}} - \mathbb{F}\|_F^2 \leq D_1\left(\sigma_\epsilon^2 + \|\mathbb{B}\|_{L^2,1}^2\right)\sqrt{\frac{\log p}{n}}, \tag{3.15}$$

$$\mathrm{E}\left[\|\mathbf{Y}_{\mathrm{pred}} - \mathbf{Y}_{\mathrm{new}}\|_2^2|\mathbb{X}, \mathbb{Y}\right] \leq D_2\left(\sigma_\epsilon^2 + \|\mathbb{B}\|_{L^2,1}^2\right)\sqrt{\frac{\log p}{n}}, \tag{3.16}$$

其中 D_1 和 D_2 是不依赖于 n的常数。

常数 λ_0 可以取 0 到 1 之间的任意值。因此，只要调优参数 λ 远离零，其取值便可任意选择，且不会影响 (3.15) 和 (3.16) 中上界的收敛速率。这些收敛速率表明，较低的噪声水平和更稀疏的函数系数矩阵能够提高本方法的估计和预测精度。

4 计算问题

为求解优化问题 (3.4)，我们使用 B样条基展开表示 $\boldsymbol{\alpha}(t)$ 中的每个函数，并将 (3.4) 转化为针对这些基展开所对应的拼接系数向量的广义特征值问题。由于光滑性惩罚项的作用，只要基函数的个数足够多，其数量便不会影响估计。求解 (3.4) 的细节见补充材料的 S.2 节。与函数偏最小二乘和函数主成分分析一样，选择适当的分量个数 K 十分重要。函数主成分分析中已有多种选择 K 的方法，例如假设检验（Cardot 等，2003a年；Horv´ath 和 Kokoszka，2012；Su 等，2017）、自助法（Hall 和 Vial，2006年）、信息准则（Yao等，2005a；Li 等，2013）、交叉验证（Hosseini-Nasab，2013；Kr¨amer等，2008年；Goldsmith 等，2011）以及广义交叉验证（Cardot 等，2003b）。选择调优参数的方法包括限制性最大似然（Wood，2011）、广义交叉验证（Wahba，1990；Ruppert等，2003；Reiss 和 Ogden，2007）、交叉验证（Kr¨amer 等，2008年；Goldsmith 等，2011）等。本文基于交叉验证过程，同时选择最优的分量个数 K_{opt} 和最优调优参数，并在 `R` 包`FRegSigCom`中实现了该过程。具体细节见下一节。

4.1 分量数量与调优参数的选择

按照第3.2节，我们分别考虑 p较小与 p较大这两种情形。情形*1*： p 较小，且惩罚项 (3.5)被使用。有两个调优参数， τ 和 η。我们从集合 $\{10^{-9}, 10^{-6}, 10^{-3}, 10^{1}, 10^{3}\}$ 中选择 τ，并从 $\{10^{-7}, 10^{-5}, 10^{-3}, 10^{-1}, 10\}$中选择 η。最优对$(\tau_{\text{opt}}, \eta_{\text{opt}})$连同最优成分个数 K_{opt} ，通过交叉验证过程从25对 $\{(\tau_i, \eta_j)$： $1 \leq i \leq 5, 1 \leq j \leq 5\}$中选出，其中 τ_i与 η_j分别是 τ 和 η的两个候选集中的取值。为了给出 K_{opt}的范围，我们注意到定理 2.1建议选择σ_k^2 取值相对较大的分量，该值可由优化问题(3.4)的最大值 $\widehat{\sigma}_k^2$ 估计。此外，根据定理 2.1，成分个数不能

超过响应向量的维度 m。因此，我们提出 K_{opt} 的上界为

$$K_{upper} = \min\left\{k > 1 : \frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_1^2 + \cdots + \widehat{\sigma}_k^2} \leq 0.001\right\} \bigwedge m, \tag{4.1}$$

其中 $a \bigwedge b$ 表示 a 和 b中的较小值。最优分量数 K_{opt} 从 $\{1, \ldots, K_{upper}\}$中选取，同时从以下算法中选择（τ_{opt}， η_{opt}）。

步骤 *1.* 对于每一对（τ_i， η_j）、 $1 \leq i \leq 5$ 和 $1 \leq j \leq 5$，我们使用整个数据集和（4.1）计算上界 $K_{upper}^{(i,j)}$。

步骤 *2.* 我们将整个数据集随机分成五个大小大致相同的子集。对于每个 $v \in \{1, \ldots, 5\}$，令（$\mathbb{Y}^{(v)}$， $\mathbb{X}^{(v)}(t)$）表示第 v 个子集，该子集用作第 v 个验证集，其余所有观测则作为第 v 个训练集。对于每个 $v \in \{1, \ldots, 5\}$ 以及每一对候选调优参数（ τ_i， η_j ），

(a). 我们使用第 v 个训练集计算所有 $k \in \{1, \ldots, K_{upper}^{(i,j)}\}$对应的 $\widehat{\boldsymbol{\alpha}}_k^{(v,i,j)}(t)$、 $\widehat{\boldsymbol{\mu}}^{(v,i,j)}$ 和 $\mathbf{w}_k^{(v,i,j)}$。

(b). 对于每个 $k \in \{1, \ldots, K_{upper}^{(i,j)}\}$，我们计算

$$\widehat{\mathbb{B}}^{(v,i,j,k)}(t) = \widehat{\boldsymbol{\alpha}}_1^{(v,i,j)}(t)\widehat{\mathbf{w}}_1^{(v,i,j)T} + \cdots + \widehat{\boldsymbol{\alpha}}_k^{(v,i,j)}(t)\widehat{\mathbf{w}}_k^{(v,i,j)T}$$

以及第 v 个验证数据集的预测响应 $\mathbb{Y}^{(v,i,j,k)}{}_{\text{预测}} = \mathbf{1}_{n_v}\widehat{\boldsymbol{\mu}}^{(v,i,j)T} + \int_0^1 \mathbb{X}^{(v)}(t)\widehat{\mathbb{B}}^{(v,i,j,k)}(t)dt$，其中 n_v 是第 v 个验证集中的观测数量。然后我们计算验证误差 $e_{v,\ i,\ j,\ k} = \|\mathbb{Y}^{(v,i,j,k)}{}_{\text{预测}} - \mathbb{Y}^{(v)}\|_F^2$。

对所有 $1 \leq v \leq 5$、 $1 \leq i \leq 5$ 和 $1 \leq j \leq 5$，我们重复步骤 (a)-(b)。

步骤 *3.* 对于所有 i， j， k，我们计算总验证误差 $e_{i,\ j,\ k}^{(total)} = \sum_{v=1}^5 e_{v,\ i,\ j,\ k}$。令 $e_{i_0\ j_0\ ,\ ,\ k_0}^{(total)} = \min\{e_{i,j,k}^{(total)} : \ 1 \leq i \leq 5, \ 1 \leq j \leq 5, \ 1 \leq k \leq \widehat{K}_{upper}^{(i,j)}\}$。我们的最优调优参数为 $\tau_{\text{opt}} = \tau_{i_0}$、 $\eta_{\text{opt}} = \eta_{j_0}$和 $K_{\text{opt}} = k_0$。

情况 *2*： p 较大，并使用惩罚项(3.7) 。我们采用与上文相同的交叉验证流程，但惩罚项 (3.7) 中有三个调优参数： τ、 λ 和 η。通常，人们会在三维网格中选取这三个调优参数，以探索它们的各种组合。然而，由于惩罚项 (3.7) 涉及 L^1 类型的范数，并且可能应用于包含数百或数千条预测变量曲线的数据，因此它会复杂得多，耗时也更长

。为了提高算法在 p较大时的计算效率，我们减少了三元参数的候选值数量。在惩罚项 (3.7) 中，$\left(\sum_{j=1}^{p}\|\alpha_j\|_\eta\right)^2$ 这一项在曲线选择中起主要作用，并由 $\tau\lambda$调节。因此，我们不再在二维网格中探索 (τ，λ) 的所有可能组合，而是从集合 {(10^{-1}，0.1)，(1，0.2)，(10，0.3)，(100，0.4)} 中选取，其乘积 $\tau\lambda$ 的取值范围较广。平滑度调节参数 η 从 $\{10^{-6}, 10^{-3}, 1\}$中选取。因此，最优三元组 (τ_{opt}，λ_{opt}，η_{opt}) 从 $4 \times 3 = 12$个三元值中选出。

5 模拟研究

我们分别针对 $p = 1$ 和 $p > 1$ 进行模拟研究。对于每次模拟，截距 $\boldsymbol{\mu}$ 是一个 m维向量，其所有坐标均为 1；噪声向量$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_m)$ 中，ε_k是来自同一分布$N(0, \sigma^2)$的独立随机变量。我们考虑三种噪声水平：$\sigma = 0.01$，0。1，1，以及三种维度取值：m,$m = 1$，5，10。由于基于小波的方法会进行快速小波变换，而该变换要求观测时间点的数量为 2 的幂，因此在所有模拟中，预测曲线均在 $[0, 1]$ 上的 64 个等距点处观测。

5.1 单预测变量曲线的模拟

我们将我们的方法（$new.Sm$，在 R 包 FRegSigCom中实现）与以下方法进行比较。惩罚函数回归（pfr，Goldsmith 等（2011））使用主成分基展开样本预测变量曲线，并使用样条基展开系数函数，在 R 包 refund中实现。基于基函数表示的函数回归（$fbasis$，Ramsay 和 Silverman（2005）以及 Febrero-Bande 等（2012））使用样条基同时展开预测变量曲线和系数函数，在 R 包 fda.usc中实现。基于主成分分析的函数回归（$fpcr$，Cardot等（1999年））使用主成分基同时展开预测变量曲线和未知系数函数，在 R 包 fda.usc中实现。标量响应的函数型惩罚偏最小二乘回归（$fpls$，

Preda 和 Saporta（2005）以及 Kr¨amer 等（2008年））使用 FPLS基同时展开预测变量曲线和未知系数函数，并在 R包 `fda.usc`中实现。使用非参数核估计的函数回归（*fnp*, Ferraty 和 Vieu（2006年b））在 R 包 `fda.usc`中实现。用于多元标量对函数回归的高斯过程回归（*gp-msof*，Wang 等（2017））先对多元响应进行主成分分析，然后使用高斯过程回归为每个主成分得分拟合非参数模型（O'Hagan 和 Kingman，1978；Rasmussen 和 Williams，2006年）。*gp-msof* 的代码由 Wang等（2017）的作者提供。两种基于小波的方法，$wLasso$ （Zhao等，2012）和 $wwLasso$ （Zhao 等，2015），首先使用快速小波变换将预测变量曲线变换为小波系数向量，然后对小波系数应用 Lasso 或加权Lasso。我们使用 R 包 `wavethresh` 进行快速小波变换，并使用 R 包`glmnet` 进行 Lasso 和加权Lasso。在 *new.Sm*、*pfr* 和 *fbasis*中，对于所有模拟和真实数据分析，B 样条基函数的个数被选为相同且等于30。对于 *fbasis*、*fpcr* 和 *fpls*，我们从集合 $\{10^{-10}, 10^{-9}, 10^{-8}, \ldots, 1, 10, \ldots, 10^{9}, 10^{10}\}$中选择平滑调谐参数，其性能优于默认设置。在 $wLasso$ 中，我们在函数 `cv.glmnet` 中选择选项 `family="mgaussian"` ，以用 Lasso 联合拟合多元响应。然而，在 $wwLasso$中，如果选择选项 `family="mgaussian"`，预测误差将大幅增加，因此我们对响应的每个坐标分别进行拟合。

我们进行两组模拟，采用不同方式生成预测变量曲线和系数函数。

模拟*1.* $X(t)$由协方差函数为$\exp\{-[30(t-t')]^2\}$的高斯过程生成。系数$\mathbf{b}(t) = (b_1(t) \ldots, b_m(t))^{\mathrm{T}}$由 $b_k(t) = \sin(3\pi t + k) + \cos(3\pi t + k), 1 \leq k \leq m$给出。

模拟 2。 $X(t)$由基展开生成： $X(t) = \sum_{k=1}^{20} [V_{k1} \sin(2k\pi t) + V_{k2} \cos(2k\pi t)]$, 其中，对于 $1 \leq k \leq 20$ 和 $j = 1$， $V_{kj} \sim N(0, 1/k^{1/2})$，且二者相互独立。系数 $\mathbf{b}(t) = \mathbb{M}\boldsymbol{\beta}(t)$，其中 $\mathbb{M}$ 是一个 $m \times 3$ 矩阵，每个元素均独立地由 (0, 1) 上的均匀分布生成； $\boldsymbol{\beta}(t) = (\beta_1(t)$ $\beta_2(t)$ $\beta_3(t))^{\mathrm{T}}$，其中 $\beta_1(t) = \cos(2\pi t)$， $\beta_2(t) = 2t^2$ ，且 $\beta_1(t) = 1/(1 + t)$。

这两个模拟中 $X(t)$ 的样本曲线绘于补充材料的图 S.1。我们根据模型 $\mathbf{Y} = \boldsymbol{\mu} + c\int_0^1 X(t)\mathbf{b}(t)dt + \boldsymbol{\varepsilon}$ 生成数据，其中 c 是一个缩放常数，使得当噪声水平为 $\sigma = 1$ 时，信噪比等于 1。对于每种设置，我们重复以下过程 100 次。每次重复中，我们生成一个大小为 100 的训练数据集和一个大小为 500 的测试数据集 $\{X_{\text{test},\ell}(t), \mathbf{Y}_{\text{test},\ell} : 1 \le \ell \le 500\}$。我们使用训练集选择调优参数并拟合模型，然后将拟合后的模型应用于测试数据，以计算均方预测误差（MSPE）

$$\text{MSPE} = \frac{1}{500}\frac{\sum_{\ell=1}^{500} \|\widehat{\mathbf{Y}}_\ell^{\text{pred}} - \mathbf{F}_{\text{test},\ell}\|_2^2}{m},$$

对于每种方法，其中 $\widehat{\mathbf{Y}}_\ell^{\text{pred}}$ 是 $X_{\text{test},\ell}(t)$ 和 $\mathbf{F}_{\text{test},\ell} = \boldsymbol{\mu} + \int_0^1 X_{\text{test},\ell}(t)\mathbf{b}(t)dt$ 的预测响应向量。表 1 汇总了模拟 1 和模拟 2 中 100 次重复所得均方预测误差的平均值和标准差，结果呈现出一些有趣的规律。(1) 当噪声水平相对较低时（$\sigma = 0.01, 0.1$），我们的方法相较于其他方法具有显著优势，尤其是在 m 相对较大时。例如，当 $m = 10$ 和 $\sigma = 0.01$ 时，在模拟 1 中，其他方法的均方预测误差平均值至少是我们方法的 24 倍；在模拟 2 中则为 6.3 倍。

(2) 当噪声水平相对较高时（$\sigma = 1$），对于 $m = 1$，我们的方法、*pfr* 和 *fbasis* 的均方预测误差相近，且在两次模拟中均低于其他方法；当 $m > 1$ 时，我们的方法具有最佳预测性能。

(3). 在两次模拟的所有噪声水平下，随着 m 增大，我们方法的均方预测误差都会降低。在模拟 2 中，对于任意 m，响应的各个坐标具有相同的分布，且该分布也与 $m = 1$ 的分布相同。因此，当 $m > 1$ 时，如果分别对响应的每个坐标拟合模型，均方预测误差应与 $m = 1$ 的相同。通过考虑响应中的相关结构，当 m 较大时，我们的方法提高了预测准确度。在模拟 2 中，对于 $\sigma = 0.1$，我们的方法所选分量个数的平均值在 $m = 5$ 时为 3.34，在 $m = 10$ 时为 3.26。与 $m = 5$ 的情况相比，尽管 $m = 10$ 时坐标更多，我们的方法选择的分量却更少。因此，与 $m = 5$ 相比，当 $m = 10$ 时，我们的方法进行更积极的降维，从而提高了预测精度。我们也可以用类似的方式解释模拟 1 中的这一规律。

表 1：模拟 1 和模拟 2 中均方预测误差的平均值及标准差（括号内为标准差）。

模拟1						
σ	m	*new.Sm*	pfr	*fbasis*	$fpcr$	$fpls$
0.01	1	0.28(0.11)$\times 10^{-4}$	5.3(3.8)$\times 10^{-4}$	3.6(0.9)$\times 10^{-4}$	0.34(0.16)	0.002(0.0003)
	5	0.14(0.04)$\times 10^{-4}$	4.2(1.6)$\times 10^{-4}$	2.3(0.4)$\times 10^{-4}$	0.30(0.10)	0.001(0.0002)
	10	0.09(0.02)$\times 10^{-4}$	4.0(1.4)$\times 10^{-4}$	2.2(0.4)$\times 10^{-4}$	0.31(0.11)	0.001(0.0002)
0.1	1	1.8(0.7)$\times 10^{-3}$	4.9(0.8)$\times 10^{-3}$	4.0(1.0)$\times 10^{-3}$	0.37(0.16)	2.8(0.6)$\times 10^{-3}$
	5	0.9(0.3)$\times 10^{-3}$	3.4(0.4)$\times 10^{-3}$	3.0(0.4)$\times 10^{-3}$	0.30(0.11)	2.2(0.3)$\times 10^{-3}$
	10	0.6(0.1)$\times 10^{-3}$	3.4(0.3)$\times 10^{-3}$	3.0(0.3)$\times 10^{-3}$	0.31(0.12)	2.1(0.3)$\times 10^{-3}$
1	1	0.10(0.05)	0.09(0.04)	0.10(0.05)	0.49(0.20)	0.23(0.08)
	5	0.07(0.03)	0.08(0.02)	0.10(0.03)	0.43(0.11)	0.20(0.04)
	10	0.05(0.02)	0.09(0.02)	0.10(0.02)	0.46(0.11)	0.21(0.04)
σ	m	fnp	*gp-msof*	*wLasso*	*wwLasso*	
0.01	1	0.90(0.04)	0.99(0.03)	0.002(0.0003)	0.002(0.0004)	
	5	0.89(0.03)	1.02(0.02)	0.002(0.0002)	0.002(0.0003)	
	10	0.90(0.03)	1.02(0.02)	0.002(0.0002)	0.001(0.0002)	
0.1	1	0.91(0.04)	1.02(0.02)	6.3(1.9)$\times 10^{-3}$	5.1(1.3)$\times 10^{-3}$	
	5	0.90(0.03)	1.02(0.02)	4.1(0.8)$\times 10^{-3}$	4.7(0.5)$\times 10^{-3}$	
	10	0.90(0.03)	1.02(0.02)	3.6(0.5)$\times 10^{-3}$	4.6(0.4)$\times 10^{-3}$	
1	1	0.92(0.05)	1.00(0.03)	0.39(0.12)	0.32(0.09)	
	5	0.92(0.03)	1.02(0.02)	0.25(0.05)	0.31(0.04)	
	10	0.92(0.03)	1.02(0.02)	0.21(0.04)	0.31(0.02)	
模拟 2						
σ	m	*new.Sm*	pfr	*fbasis*	$fpcr$	$fpls$
0.01	1	1.8(0.7)$\times 10^{-5}$	2.3(2.7)$\times 10^{-3}$	6.8(2.7)$\times 10^{-5}$	0.13(0.08)	0.014(0.011)
	5	1.4(0.3)$\times 10^{-5}$	1.9(0.5)$\times 10^{-3}$	6.3(0.9)$\times 10^{-5}$	0.12(0.05)	0.011(0.002)
	10	1.0(0.2)$\times 10^{-5}$	1.8(0.4)$\times 10^{-3}$	6.3(0.6)$\times 10^{-5}$	0.11(0.05)	0.011(0.002)
0.1	1	1.3(0.6)$\times 10^{-3}$	6.9(0.4)$\times 10^{-3}$	3.7(1.2)$\times 10^{-3}$	0.13(0.07)	0.015(0.01)
	5	0.9(0.2)$\times 10^{-3}$	5.8(0.5)$\times 10^{-3}$	3.5(0.5)$\times 10^{-3}$	0.12(0.05)	0.013(0.003)
	10	0.8(0.2)$\times 10^{-3}$	5.8(0.8)$\times 10^{-3}$	3.6(0.4)$\times 10^{-3}$	0.12(0.06)	0.013(0.002)
1	1	0.06(0.04)	0.06(0.03)	0.08(0.05)	0.59(0.24)	0.15(0.05)
	5	0.05(0.02)	0.06(0.02)	0.08(0.03)	0.59(0.13)	0.15(0.04)
	10	0.04(0.01)	0.06(0.01)	0.08(0.02)	0.59(0.11)	0.16(0.03)
σ	m	fnp	*gp-msof*	*wLasso*	*wwLasso*	
0.01	1	0.91(0.04)	1.01(0.03)	0.0014(0.0003)	0.0015(0.0003)	
	5	0.90(0.03)	1.01(0.02)	0.0014(0.0002)	0.0015(0.0003)	
	10	0.90(0.03)	1.00(0.02)	0.0015(0.0003)	0.0016(0.0003)	
0.1	1	0.90(0.04)	1.00(0.03)	4.7(1.4)$\times 10^{-3}$	4.4(1.4)$\times 10^{-3}$	
	5	0.90(0.04)	1.01(0.03)	3.3(0.7)$\times 10^{-3}$	4.0(0.6)$\times 10^{-3}$	
	10	0.90(0.03)	1.02(0.01)	3.1(0.5)$\times 10^{-3}$	4.2(0.4)$\times 10^{-3}$	
1	1	0.91(0.04)	1.01(0.02)	0.29(0.06)	0.22(0.05)	
	5	0.91(0.03)	1.02(0.03)	0.16(0.04)	0.20(0.03)	
	10	0.92(0.04)	1.02(0.03)	0.14(0.03)	0.21(0.02)	

(4). 我们的方法、函数偏最小二乘 和 函数主成分分析 都考虑了系数函数的分解。函数主成分回归 仅使用 $X(t)$构造分量函数，而忽略了$\mathbf{Y}$中的信息。函数偏最小二乘 和我们的方法都利用了 $X(t)$与$\mathbf{Y}$中的信息，但我们的方法针对预测估计最优分解。因此，在这些模拟中，我们的方法优于函数偏最小二乘，而后者优于函数主成分分析。

(5). 当 $\sigma = 0.1$时，两种基于小波的方法与pfr 和$fbasis$的表现相近。随着 m增大，$wLasso$ 的预测误差会降低，因为它对多元响应进行联合拟合，并考虑了响应中的相关结构。

(6). 这些模拟中的所有数据均由线性模型生成。但fnp 是一种非参数方法，并非专为线性模型设计。fnp 较高的均方预测误差可能是由于其复杂性所致。$gp\text{-}msof$ 也使用非参数方法，情况同样如此。

5.2 多个函数型预测变量曲线的模拟

当 p 适中或较大时，模拟 1 和模拟 2 中的许多方法在计算上并不可行。我们将方法 $new.SmSp$（采用 (3.7) 中的同时光滑稀疏惩罚）和 $new.Sm$ （采用 (3.5) 中的仅光滑惩罚）与 $wLasso$ 和$wwLasso$进行比较。此外，我们还考虑了另外两种混合方法：$FPCA+gSCAD$ （Lian，2013；Kong $et\ al.$，2016b）和$FPCA+gLasso$ （Zhu 和 Cox，2009）。这两种方法分别使用 `R` 软件包 `fdapace` 对每个函数型预测变量进行函数主成分分析，然后将主成分得分作为新的标量预测变量，并使用 `R` 软件包 `grpreg` 拟合惩罚回归模型，分别施加组SCAD或组Lasso惩罚。函数主成分分析中的成分个数根据解释 99% 变异的标准确定。但对于较大的 p（例如 1000），这两种基于函数主成分分析的方法在计算上并不可行。

我们通过以下两种不同的方式生成预测变量曲线和系数函数。

模拟 3。 我们取 $p = 50$。

(1). $\mathbf{X}(t) = (X_1(t) \ldots, X_{50}(t))^{\mathrm{T}}$ 由基展开生成

$$X_j(t) = \sum_{k=1}^{20} [V_{k1}^{(j)} \sin(2k\pi t) + V_{k2}^{(j)} \cos(2k\pi t)], \quad 1 \le j \le p,$$

其中，$\mathbf{V}_{ki} = (V_{(1)ki}, \ldots, V_{(50)ki})^{\mathrm{T}}$ 服从均值为零的多元正态分布，其协方差矩阵的对角元素等于$1/\sqrt{k}$ ，非对角元素等于$\rho/\sqrt{k}$，对于 $1 \le k \le 2\theta$ 和 $i = 1,2$。$\mathbf{V}_{ki}$彼此独立。值 $0 \le \rho \le 1$ 控制着 $X_j(t)$之间的相关性。 ρ越大，相关性越强。我们考虑两个值 $\rho = 0.2, 0.7$。

(2). 系数 $\mathbb{B}(t) = [\boldsymbol{\beta}_1(t)\ \boldsymbol{\beta}_2(t)\ \boldsymbol{\beta}_3(t)]\,\mathbb{M}$ 是一个 $p \times m$ 矩阵，其中 $\mathbb{M}$ 是一个 $3 \times m$ 矩阵，每个元素均独立地从 (0, 1) 上的均匀分布中生成。对于 $1 \le k \le 3$， $\boldsymbol{\beta}_k(t) = (\beta_{1k}(t) \ldots, \beta_{pk}(t))^{\mathrm{T}}$ 是一个 p维向量，其中 $\beta_{1k}(t) = t^k$， $\beta_{2k}(t) = \cos(k\pi t)$， $\beta_{3k}(t) = 1/(t + k)$， $\beta_{4k}(t) = \log(t + k)$， $\beta_{5k}(t) = e^{-kt^2}$，且当$j > 5$时， $\beta_{jk}(t) = 0$。因此，只有前五条预测变量曲线具有非零系数，并被视为真实特征。

模拟*4*。 我们取 $p = 1000$。

(1). 为了生成 $\mathbf{X}(t) = (X_1(t) \ldots, X_{1000}(t))^{\mathrm{T}}$，我们首先从协方差函数为 $\exp\{-[30(t - t')]^2\}$, $1 \le j \le 1000 + Lag$ 的高斯过程中生成相互独立的曲线 $W_j(t)$’，并令 $X_j(t) = [W_{j+1}(t) + \cdots + W_{j+Lag}(t)]/\sqrt{Lag}$，其中 Lag 是一个用于控制 $X_j(t)$’之间相关性的正整数。 Lag 越大，$X_j(t)$’之间的相关性越强。我们取 $Lag = 2$，5。

(2). 系数 $\mathbb{B}(t) = [\boldsymbol{\beta}_1(t)\ \boldsymbol{\beta}_2(t)\ \boldsymbol{\beta}_3(t)]\,\mathbb{M}$是一个 $p \times m$矩阵，其中 $\mathbb{M}$的定义与模拟 3 中相同。对于 $1 \le k \le 3$， $\boldsymbol{\beta}_k(t) = (\beta_{1k}(t) \ldots, \beta_{pk}(t))^{\mathrm{T}}$ 是一个 p维向量，其中$\beta_{1,\,k}(t) = 2(t + 1)e^{-kt}$, $\beta_{11,\,k}(t) = \sin(k\pi t/2)/\sqrt{1 + t}$， $\beta_{21,\,k}(t) = 3\sinh(-t)/k + t^2$， $\beta_{31,\,k}(t) = 0.5(1 + t)^k \cos(k\pi t)$， $\beta_{41,\,k}(t) = \tan(t)/(1 + kt^2)$，而对于所有其他 j， $\beta_{jk}(t) = 0$。因此，只有五条预测变量曲线 $X_1(t)$ $X_{11}(t)$ $X_{21}(t)$ $X_{31}(t)$ $X_{41}(t)$的系数非零，并将它们视为真实特征。

我们从模型 $\mathbf{Y}_\ell = \mathbf{F}_\ell + \boldsymbol{\varepsilon}_\ell$ 生成数据，其中 $\mathbf{F}_\ell = \boldsymbol{\mu} + c\int_0^1 \mathbf{X}_\ell(t)^{\mathrm{T}}\mathbb{B}(t)dt$, $1 \le \ell \le 100$，并且 c 是一个缩放因子，使得当噪声水平为 $\sigma = 1$ 时，信噪比等于 1。我们按照第 5.1 节中的类似步骤进行模拟，共重复实验 100 次，每次重复实验的训练数据大小为 100，测试数据大小为 500。100 次重复实验所得均方预测误差的平均值和标准差汇总于表 2。在两项模拟中，我们的方法 $new.SmSp$ 的均方预测误差最低。与不使用稀疏惩罚的方法相比，引入同时光滑稀疏惩罚显著提升了我们方法的性能

与不使用稀疏惩罚相比。随着 m 增大，我们的方法 $new.SmSp$ 的均方预测误差会降低，尤其是在 $\sigma = 1$ 时。

此外，我们还计算了每种方法在曲线选择中的灵敏度和特异度。令 $\widehat{\mathbb{B}}(t) = \left[\widehat{\mathbf{b}}_1(t)\ldots,\widehat{\mathbf{b}}_p(t)\right]^{\mathrm{T}}$ 表示系数 $\mathbb{B}(t) = [\mathbf{b}_1(t)\ldots,\mathbf{b}_p(t)]^{\mathrm{T}}$ 的估计值，其中 $\mathbf{b}_j(t)$ 是第 j 个预测变量曲线 $1 \le j \le p$ 的系数。灵敏度和特异度分别定义为

$$\text{sensitivity} = \frac{\text{the number of } j\text{'s with both } \mathbf{b}_j(t) \neq \mathbf{0} \text{ and } \widehat{\mathbf{b}}_j(t) \neq \mathbf{0}}{\text{the number of } j\text{'s with } \mathbf{b}_j(t) \neq \mathbf{0}}$$

and

$$\text{specificity} = \frac{\text{the number of } j\text{'s with both } \mathbf{b}_j(t) = \mathbf{0} \text{ and } \widehat{\mathbf{b}}_j(t) = \mathbf{0}}{\text{the number of } j\text{'s with } \mathbf{b}_j(t) = \mathbf{0}}.$$

补充材料中的表 S.1 和 S.2 总结了各方法的灵敏度和特异度。方法 $new.Sm$ 的灵敏度为 1，特异度为 0，因为它不进行变量选择，而是在模型拟合中纳入所有变量。我们的方法 $new.SmSp$ 具有最高的灵敏度，其特异度低于基于小波的方法或基于函数主成分的方法。这意味着 $new.SmSp$选择的曲线比其他方法更多，其中既包括真实的特征曲线，也包括非特征曲线。

6 应用

我们考虑两个数据集：具有单条预测变量曲线的玉米数据，以及具有大量预测变量曲线的同步脑电图-功能性磁共振成像数据。

6.1 玉米数据

该数据集包含 80 个玉米样本，可从 Eigenvector Research 获取（http://www.eigenvector.com/data/Corn/index.html）。每个玉米样本的水分、油、蛋白质和淀粉含量，以及其近红外光谱曲线均经过测量；光谱波长范围为 1100 nm 至 2498 nm，间隔为 2 nm。

表 2：模拟 3 和 4 中均方预测误差的平均值（和标准差）。

模拟 3								
ρ	σ	m	*new.SmSp*	*new.Sm*	*wLasso*	*wwLasso*	函数主成分分析+*gSCAD*	函数主成分分析+*gLasso*
0.2	0.01	1	0.050(0.026)	0.349(0.165)	0.375(0.156)	0.186(0.073)	0.605(0.135)	0.628(0.121)
		5	0.049(0.015)	0.364(0.093)	0.308(0.094)	0.184(0.051)	0.602(0.099)	0.625(0.096)
		10	0.049(0.013)	0.360(0.075)	0.300(0.088)	0.181(0.046)	0.599(0.083)	0.624(0.085)
	0.1	1	0.064(0.030)	0.363(0.174)	0.379(0.156)	0.198(0.088)	0.598(0.133)	0.628(0.125)
		5	0.057(0.017)	0.360(0.082)	0.316(0.092)	0.195(0.048)	0.610(0.098)	0.637(0.097)
		10	0.053(0.013)	0.364(0.084)	0.317(0.086)	0.197(0.042)	0.611(0.088)	0.628(0.093)
	1	1	0.536(0.200)	0.883(0.151)	0.893(0.155)	0.724(0.205)	0.952(0.108)	0.957(0.114)
		5	0.267(0.068)	0.687(0.104)	0.667(0.125)	0.679(0.108)	0.940(0.080)	0.940(0.082)
		10	0.228(0.045)	0.593(0.093)	0.603(0.109)	0.677(0.104)	0.941(0.060)	0.943(0.061)
0.7	0.01	1	0.051(0.022)	0.174(0.076)	0.240(0.096)	0.113(0.052)	0.360(0.103)	0.349(0.088)
		5	0.049(0.013)	0.180(0.043)	0.204(0.056)	0.106(0.028)	0.359(0.065)	0.347(0.056)
		10	0.047(0.011)	0.182(0.038)	0.207(0.045)	0.107(0.028)	0.357(0.059)	0.349(0.053)
	0.1	1	0.059(0.023)	0.191(0.081)	0.256(0.100)	0.125(0.054)	0.361(0.089)	0.357(0.086)
		5	0.054(0.013)	0.188(0.046)	0.216(0.051)	0.121(0.033)	0.368(0.059)	0.356(0.054)
		10	0.051(0.012)	0.184(0.035)	0.213(0.044)	0.121(0.030)	0.363(0.055)	0.354(0.051)
	1	1	0.436(0.153)	0.537(0.172)	0.670(0.192)	0.574(0.193)	0.733(0.142)	0.736(0.151)
		5	0.276(0.064)	0.445(0.072)	0.435(0.085)	0.537(0.098)	0.709(0.075)	0.707(0.072)
		10	0.230(0.042)	0.377(0.059)	0.385(0.063)	0.522(0.077)	0.699(0.061)	0.695(0.059)

模拟4						
Lag	σ	m	*new.SmSp*	*new.Sm*	*wLasso*	*wwLasso*
2	0.01	1	0.187(0.109)	0.973(0.043)	0.465(0.143)	0.378(0.138)
		5	0.174(0.066)	0.968(0.040)	0.394(0.098)	0.351(0.097)
		10	0.148(0.042)	0.971(0.040)	0.372(0.073)	0.350(0.077)
	0.1	1	0.207(0.120)	0.976(0.040)	0.465(0.154)	0.382(0.144)
		5	0.163(0.062)	0.971(0.041)	0.399(0.117)	0.376(0.110)
		10	0.171(0.052)	0.974(0.040)	0.388(0.091)	0.370(0.091)
	1	1	0.657(0.244)	1.03(0.06)	0.884（0.160）	0.847（0.184）
		5	0.328(0.086)	0.997(0.043)	0.632(0.131)	0.785(0.124)
		10	0.274(0.063)	0.987(0.041)	0.572(0.095)	0.775(0.119)
5	0.01	1	0.181(0.132)	0.953(0.052)	0.459(0.184)	0.363(0.168)
		5	0.154(0.050)	0.843(0.052)	0.376(0.088)	0.330(0.089)
		10	0.149(0.053)	0.944(0.050)	0.362(0.099)	0.318(0.094)
	0.1	1	0.180(0.106)	0.945(0.052)	0.437(0.155)	0.348(0.138)
		5	0.164(0.059)	0.947(0.053)	0.392(0.100)	0.353(0.098)
		10	0.168(0.058)	0.954(0.055)	0.379(0.088)	0.352(0.088)
	1	1	0.645(0.220)	1.02(0.054)	0.901(0.137)	0.848(0.166)
		5	0.334(0.112)	0.987(0.056)	0.635(0.143)	0.772(0.130)
		10	0.259(0.051)	0.958(0.054)	0.553(0.090)	0.737(0.114)

人们感兴趣的是根据近红外光谱曲线预测玉米中四种成分的含量。因此，我们考虑以近红外光谱曲线为唯一函数型预测变量、以这四种成分为多元响应的函数线性回归模型。图4展示了全部80条预测变量曲线及其中心化版本。四个响应变量均经过中心化和缩放，使其方差为1；它们的直方图见补充材料中的图S.2。

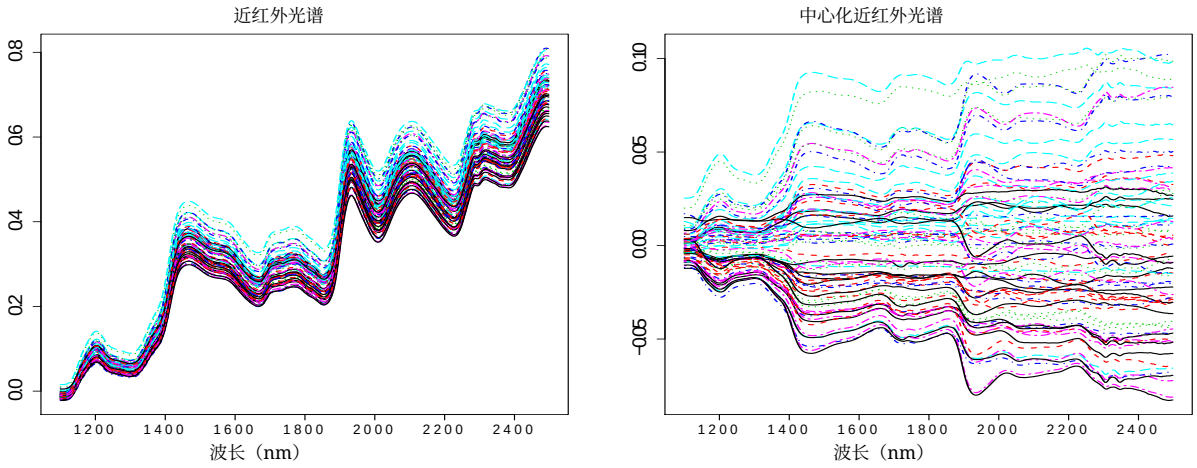

图4：玉米数据中80个样本的原始近红外光谱曲线及其中心化版本。

我们首先将第 5.1 节中用于模拟 1 和模拟 2 的所有方法应用于该数据集，并比较它们的预测性能。我们进行 100 次迭代，在每次迭代中，将总共 80 个观测随机划分为包含 60 个观测的训练集和包含 20 个观测的测试集。除均方预测误差外，我们还在每次重复中使用以下公式计算 R^2 ，以评估每种方法的拟合优度，

$$R^2 = 1 - \frac{\sum_{\ell=1}^{60} \|\widehat{\mathbf{Y}}_{\text{fitted},\ell} - \mathbf{Y}_{\text{train},\ell}\|_2^2}{\sum_{\ell=1}^{60} \|\mathbf{Y}_{\text{train},\ell} - \overline{\mathbf{Y}}_{\text{train}}\|_2^2}, \tag{6.1}$$

其中，四维向量$\mathbf{Y}_{\text{train},\ell}$ 是训练集中响应的第 ℓ个观测值，而$\mathbf{Y}_{\text{fitted},\ell}$ 是相应的拟合值，$1 \leq \ell \leq 60$。在 100 次重复试验中，均方预测误差的平均值（和标准差）为：我们的方法为 0.18（0.03），*pfr*为 0.32（0.08），*fbasis*为 0.43（0.07），*fpcr*为 0.34（0.05），*fpls*为 0.22（0.03），*fnp*为 0.87（0.14），*gp-msof*为 1.01（0.14），*wLasso* 为 0.31（0.06），*wwLasso*为 0.91（0.13）。在 100 次重复试验中，R^2的平均值（和标准差）为：我们的方法为 91%（0.8%），

pfr为 78%（0.7%），$fbasis$为 66%（1%），$fpcr$为 76%（2%），$fpls$为 87%（1%），fnp为 40%（5%），$gp\text{-}msof$为 10%（6%），$wLasso$ 为 78%（3%），$wwLasso$为 15%（3%）。

给定一条新的预测变量曲线，除了给出预测值之外，提供响应的预测区间以评估预测的不确定性也很有意义。使用所提出的两步法，我们无法给出系数函数及其渐近方差矩阵估计值的显式公式。因此，我们建议使用自助法计算预测区间，这也能考虑调优参数选择所带来的变化。作为示例，我们将除前四个观测之外的所有观测用作训练数据，并将前四个观测用作测试数据。利用这些训练数据，我们将计算自助预测区间，并估计前四条预测变量曲线对应的均方预测误差的期望。具体而言，我们重复以下过程 1000 次。每次迭代中，我们从训练数据中有放回地抽取一个与训练数据大小相同的自助样本。我们使用该自助样本和所提出的方法拟合模型，然后计算测试数据中每个观测对应的预测响应。经过 1000 次重复后，我们得到每个响应坐标的 1000 个预测值。以这些预测值的 2.5% 和 97.5% 百分位数作为下限和上限，即可得到相应响应坐标的 95% 自助预测区间。我们在补充材料的图 S.3 和表 S.3 中分别展示了前四个观测对应的预测变量曲线和自助预测区间。前四个观测的均方预测误差期望估计值分别为 0.13、0.24、0.04 和 0.09。

最后，将我们的方法应用于全部 80 个观测后，我们得到一个由四个分量构成的模型。$\widehat{\sigma}_k^2$ of σ_k^2, $1 \leq k \leq 4$ 的四个估计量之比为 0.51 : 0.34 : 0.14 : 0.01。我们在图5中绘制了四个分量对应函数 $\alpha_k(t)$ 的估计量 $\widehat{\alpha}_k(t)$'。

6.2 同步 EEG-fMRI 数据

我们考虑来自“听觉与视觉奇异刺激脑电图-功能性磁共振成像”项目的同步脑电图-功能性磁共振成像数据集（$https://openfmri.org$，访问编号 ds000116）。该数据集包含 102 次观测，分别记录在 102 次扫描中，每次持续 340 秒。在每次

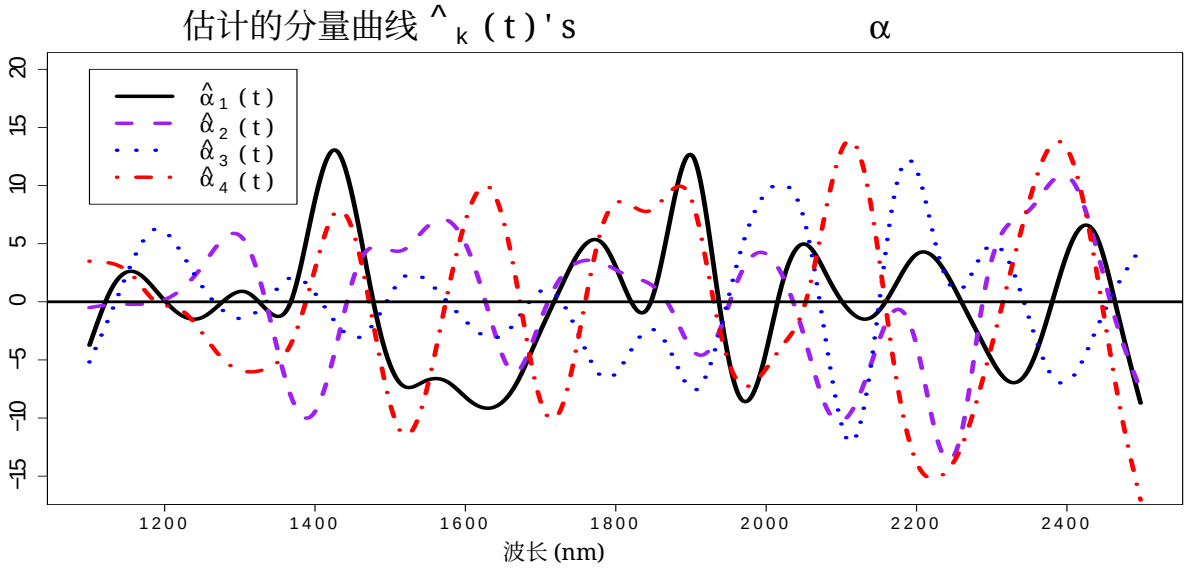

图5：玉米数据的估计值 $\widehat{\alpha}_k(t)$、 $1 \leq k \leq 4$。

扫描中，共呈现 125 个视觉（或听觉）刺激，每个刺激持续 200 毫秒，随后间隔 $2 \sim 3$ 秒的试次间间隔。关于实验背景和详情，请参阅 Walz 等人（ 2013 ）。每次观测中，都会同时记录受试者的功能性磁共振成像时间序列和脑电图时间序列。我们将从脑电图数据中提取多变量响应，并从 fMRI 数据中提取函数型预测变量。

脑电图数据以1,000 Hz的采样率记录，并已预处理以去除梯度伪影，且重参考至34通道电极空间（Walz 等，2013年）。因此，在每次观测中，脑电图数据包含34条时间序列，每条序列有340,000个观测时间点。已知不同频带的脑电功率谱波动与不同的脑活动相关。例如，伽马频段（25Hz至100Hz之间）被认为与视觉（Gray 等，1989年；Engel 等，1991年；Fries 等，2001年；Womelsdorf 等，2006年）、听觉（Brosch 等，2002年；Edwards 等，2005）及其他感觉处理有关。对于每次观测，我们使用“Chronux”（冷泉港实验室的Mitra 实验室，2014）计算伽马功率谱；Chronux是一个广受欢迎的软件包，以Matlab 库的形式实现，用于神经数据分析。将34个通道的34条脑电图时间序列合并，为每次观测计算一条伽马功率谱。两次观测的伽马功率谱示例

绘制于补充材料图 S.4的上方。这些频谱曲线中有许多峰，且这些峰的数量和位置可能因不同观测而异。我们识别出十个共同的峰，并将其绘制在补充材料图 S.4的下方。我们使用图 S.4所示、长度为0.2 HZ的区间内每个峰的积分来度量其幅度。这十个峰的十个幅度值构成了我们的十维多元响应。由于幅度值呈高度偏态分布，我们使用函数$\sqrt{\ln(1+u)}$对响应进行变换，并对变换后的值进行缩放，使十维响应向量的每个坐标的样本方差为1。我们在补充材料图 S.5中绘制了变换后响应的十个坐标的直方图。

fMRI 数据包含 170 张每隔 2 秒获取的三维图像。每张 fMRI 图像被划分为 $64 \times 64 \times 32 = 131{,}072$ 个体素。因此，在 102 个观测中的每一个观测里，fMRI 数据都可视为由 131,072 条时间序列曲线组成。我们以两种不同的方式处理这些曲线，并考虑两种不同的函数型回归模型。

(1). 使用感兴趣区域（*ROI*）曲线。

我们使用 Matlab SPM12 中的 WFU PickAtlas Tools（Lancaster等，1997、2000；Maldjian等，2003），根据人脑的 TD 布罗德曼区（布罗德曼，1909年），将大脑体素划分为 68 个感兴趣区域。我们对每个感兴趣区域内所有体素的时间序列曲线取平均，并将 $p = 68$ 平均曲线用作函数型预测变量。

由于 68 接近模拟 3 中函数型预测变量的数量，我们将模拟 3 中使用的所有方法应用于该模型。我们进行了 100 次迭代。每次迭代中，我们将 102 个观测随机划分为包含 80 个观测的训练集和包含 22 个观测的测试集。我们计算均方预测误差和 R^2，并将结果汇总于表 3。我们的方法在有无稀疏惩罚时的平均均方预测误差非常接近，且远低于其他方法。在 68 个感兴趣区域中，有 33 个区域在全部 100 次迭代中都被我们的方法选中（见图 S.6，以及补充材料中的表 S.4）。这些区域的活动可能与十个峰处的频谱波动有关。例如，布罗德曼区（BA）7 与视运动协调有关，并参与空间物体定位；布罗德曼区 8 包含额叶眼区，并参与复杂动作的规划；

布罗德曼17区和18区分别是处理视觉信息的初级和次级视觉皮层；视束是视觉系统的一部分；外侧背核则是丘脑中视觉通路的中继中心。

表 3：第一个模型中同步脑电图-功能性磁共振成像数据的均方预测误差和 R^2 的平均值（及标准差）。

	new.SpSm	*new.Sm*	*wLasso*	*wwLasso* 函数主成分分析	*+gSCAD* 函数主成分分析	*+gLasso*
MSPE	0.29(0.09)	0.29 (0.07)	1.02(0.16)	1.02(0.16)	1.12(1.08)	1.19(1.3)
R^2	0.86(0.03)	0.88(0.01)	0.02(0.03)	0.02(0.02)	0.03(0.03)	0.04(0.04)

（2）使用体素曲线。

在第二个模型中，我们直接使用体素时间序列曲线。通过补充材料 S.3.3 节中描述的预筛选程序，我们选出 $p = 3559$ 个体素，并将它们的时间序列曲线作为预测曲线。图 S.7 展示了脑图像中 3559 个体素的位置，补充材料 S.3.3 节的图 S.8 则绘制了一个样本的 3559 条预测曲线中的 200 条。

使用3559个体素的功能性磁共振成像时间序列曲线作为函数型预测变量，我们采用模拟4中的方法拟合模型。根据使用感兴趣区域曲线的研究所采用的100次迭代，我们的方法$new.SmSp$、$wLasso$ 和$wwLasso$的均方预测误差均值（标准差）分别为0.176（0.029）、1.023（0.175）和1.044（0.186）。与使用68个感兴趣区域的模型相比，使用 $p = 3559$ 个体素时，我们方法的预测误差大幅降低。这表明，对每个感兴趣区域中的体素时间序列取平均会造成信息损失。在3559个体素中，有49个被选中的频率超过90%，另有97个被选中的频率超过80%。这些被选中的体素较为稀疏，未呈现出有结构的聚类。改进体素选择的一种可能方法是在同步稀疏平滑惩罚（3.7）中加入组惩罚，其中的组指感兴趣区域信息。未来将对此进行探讨。

参考文献

Aguilera, A. M., Escabias, M., Preda, C. 和 Saporta, G.（2010年）使用基展开估计函数型偏最小二乘回归：化学计量学数据中的应用。化学计量学与智能实验室系统，**104**，289–305。

Ait-Sa¨ıdi, A., Ferraty, F., Kassa, R. 和 Vieu, P.（2008年）单函数指标模型中的交叉验证估计。*Statistics*，**42**，475–494。

Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R. 等（2014）SPM12 手册。惠康神经影像中心，伦敦（英国）。

Benko, M., H¨ardle, W., Kneip, A. 等（2009）共同函数型主成分。统计学年刊，**37**，1–34。

Boulesteix, A.-L. 和 Strimmer, K.（2006年）偏最小二乘：分析高维基因组数据的多用途工具。《生物信息学简报》，**8**，32–44。

Brodmann, K. (1909年)大脑皮层比较定位学说：按其原理并基于细胞结构加以阐述。Barth。

Brosch, M.、Budinger, E.和Scheich, H.（2002年）灵长类听觉皮层中的刺激相关伽马振荡。*Journal of Neurophysiology*，**87**，2715–2725。

Brown, P.J.、Fearn, T.和Vannucci, M.（2001年）曲线上的贝叶斯小波回归及其在光谱校准问题中的应用。美国统计学会会刊，**96**，398–408。

Cardot, H.、Ferraty, F.、Mas, A.和Sarda, P.（2003a年）函数型线性模型中的假设检验。《斯堪的纳维亚统计学杂志》，**30**，241–255。

Cardot, H.、Ferraty, F. 和 Sarda, P.（1999年）函数型线性模型。*Statistics and Proba-bility Letters*，**45**，11–22。

Cardot, H.、Ferraty, F. 和 Sarda, P.（2003b）函数型线性模型的样条估计量。*Statistica Sinica*，**13**，571–592。

Chaouch, M., La¨ıb, N. 等（2013）基于函数协变量的非参数多元 $l\ \{1\}$-中位数回归估计。*Electronic Journal of Statistics*，**7**，1553–1586。

Chen, D., Hall, P., M¨uller, H.-G. 等（2011）带有非参数连接函数的单指标和多指标函数型回归模型。统计学年刊，**39**，1720–1747。

Collazos, J. A., Dias, R. 和 Zambom, A. Z.（2016）函数型回归模型的一致变量选择。*Journal of Multivariate Analysis*，**146**，63–71。

Crainiceanu, C. M. 和 Goldsmith, A. J.（2010）使用 WinBUGS 进行贝叶斯函数型数据分析。*Journal of Statistical Software*，**32**。

Crainiceanu, C. M., Staicu, A.-M. 和 Di, C.-Z.（2009）广义多层函数型回归。美国统计学会会刊，**104**，1550–1561。

Crambes, C., Kneip, A., Sarda, P. 等(2009) 函数型线性回归的平滑样条估计量。*The Annals of Statistics*, **37**, 35–72。

Delaigle, A., Hall, P. 等 (2012) 偏最小二乘应用于函数型数据的方法与理论。*The Annals of Statistics*, **40**, 322–352。

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. 和 Punjabi, N. M. (2009) 多层次函数型主成分分析。*The Annals of Applied Statistics*, **3**,458。

Edwards, E., Soltani, M., Deouell, L.Y., Berger, M.S.和Knight, R. T.(2005)对直接记录自人类皮层、对偏差听觉刺激产生响应的高伽马活动。*Journal of Neurophysiology*,**94**, 4269–4280。

Eilers, P. H.、Li, B. 和 Marx, B. D.（2009）使用单指标信号回归进行多元校准。化学计量学与智能实验室系统，**96**，196–202。

Engel, A. K., K¨onig, P., Kreiter, A.K.and Singer, W.(1991)猫视觉皮层中振荡性神经元反应的半球间同步。*Science*，**252**，1177–1179。

Escabias, M., Aguilera, A. M. and Valderrama, M. J. (2007) 函数型偏最小二乘逻辑回归模型。计算统计与数据分析，**51**，4891–4902。Febrero-Bande, M., de la Fuente, M. O. 等(2012) 函数型数据分析中的统计计算：R 包 fda.usc。*Journal of Statistical Software*，**51**，1–28。Fengler, M. R., H¨ardle, W. K. and Villa, C. (2003) 隐含波动率的动态：共同主成分方法。衍生品研究评论，**6**，179–202。Ferraty, F., Goia, A., Salinelli, E. and Vieu, P. (2013) 函数型投影寻踪回归。*Test*，**22**，293–320。Ferraty, F., Park, J. and Vieu, P. (2011) 函数型单指标模型的估计。载于函数型数据分析及相关主题的最新进展，111–116。施普林格。Ferraty, F. and Vieu, P. (2002) 函数型非参数模型及其在光谱测定数据中的应用。*Computational Statistics*，**17**，545–564。Ferraty, F. and Vieu, P. (2006a) 非参数函数型数据分析：理论与实践。施普林格。Ferraty, F. and Vieu, P. (2006b) 非参数函数型数据分析：理论与实践。施普林格科学与商业媒体。Ferr´e, L. and Yao, A.-F. (2003) 函数型切片逆回归分析。*Statistics*，**37**，475–488。Ferr´e, L. and Yao, A.-F. (2005) 平滑函数型逆回归。*Statistica Sinica*，665–683。Fries, P., Reynolds, J. H., Rorie, A. E. and Desimone, R. (2001) 选择性视觉注意对神经元振荡同步的调节。*Science*，**291**，1560–1563。Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B. and Reich, D. (2011) 惩罚函数回归。计算与图形统计学杂志，**20**。

Goutis, C. 和 Fearn, T.（1996）基于平滑因子的偏最小二乘回归。美国统计学会会刊，**91**，627–632。

Gray, C. M., K¨onig, P., Engel, A. K., Singer, W. 等（1989）猫视觉皮层中的振荡响应表现出柱间同步，这种同步反映了全局刺激属性。《自然》，**338**，334–337。Greven, S., Crainiceanu, C., Caffo, B. 和 Reich, D.（2011）纵向函数主成分分析。载于函数型数据分析及相关主题的最新进展，149–154。施普林格。Hall, P., Horowitz, J. L.等（2007）函数型线性回归的方法学与收敛速率。统计学年刊，**35**，70–91。Hall, P. 和 Vial, C.（2006）评估函数型数据的有限维性。《皇家统计学会会刊》*B*辑（统计方法学），**68**，689–705。Hilgert, N., Mas, A., Verzelen, N. 等（2013）函数型线性模型的极小极大自适应检验。统计学年刊，**41**，838–869。Horv´ath, L. 和 Kokoszka, P.（2012）函数型数据推断及其应用，第200卷。施普林格科学与商业媒体。Hosseini-Nasab, M.（2013）函数型线性回归中的交叉验证近似。《统计计算与模拟杂志》，**83**，1429–1439。James, G. M. 和 Silverman, B. W.（2005）函数自适应模型估计。美国统计学会会刊，**100**，565–576。Kneip, A. 和 Utikal, K. J.（2001）利用函数主成分分析对密度族进行推断。美国统计学会会刊，**96**，519–542。Kong, D., Staicu, A.-M. 和 Maity, A.（2016a）函数线性模型的经典检验。《非参数统计杂志》，**28**，813–838。Kong, D., Xue, K., Yao, F. 和 Zhang, H. H.（2016b）高维部分函数线性回归。《生物统计学》，**103**，147–159。

Kr¨amer, N., Boulesteix, A.-L. 和 Tutz, G.（2008）带应用的惩罚偏最小二乘：B样条变换与函数型数据。化学计量学与智能实验室系统，94，60–69。Kr¨amer, N. 和 Sugiyama, M.（2011）偏最小二乘回归的自由度。美国统计学会会刊，106，697–705。Lancaster, J., Summerln, J., Rainey, L., Freitas, C. 和 Fox, P.（1997）Talairach 守护程序：用于 Talairach 图谱标签的数据库服务器。神经影像，5，S633。Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A. 和 Fox, P. T.（2000）用于功能性脑成像的自动 Talairach 图谱标签。人脑图谱，10，120–131。Li, Y., Wang, N. 和 Carroll, R. J.（2013）函数型数据中主成分个数的选择。美国统计学会会刊，108，1284–1294。

Lian, H. (2013) 多重函数型回归的收缩估计与变量选择。统计学研究，51–74。

Maldjian, J.A., Laurienti, P. J., Kraft, R.A. 和 Burdette, J.H. (2003) 一种基于神经解剖学和细胞构筑图谱的自动化方法，用于查询 fMRI 数据集。神经影像，19，1233–1239。

Marx, B. D. 和 Eilers, P. H.（1999年）采样信号与曲线的广义线性回归：一种P样条方法。《技术计量学》，41，1–13。Matsui, H.、Araki, Y. 和 Konishi, S.（2008年）函数型数据的多变量回归建模。《数据科学杂志》，6，313–331。冷泉港实验室的 Mitra 实验室（2014）Chronux 分析软件。http://chronux.org/。Morris, J. S.（2015）函数型回归。《统计学及其应用年度评论》，2，321–359。

O'Hagan, A. 和 Kingman, J. F. C.（1978）曲线拟合与用于预测的最优设计。皇家统计学会杂志. *B*辑（方法学），1–42。

Preda, C. 与 Saporta, G.（2005）随机过程上的偏最小二乘回归。计算统计与数据分析，48，149–158。

Rachdi, M. 与 Vieu, P.（2007）函数型数据的非参数回归：自动平滑参数选择。统计规划与推断杂志，**137**，2784–2801。

Ramsay, J. O. 与 Dalzell, C.（1991）函数型数据分析的若干工具。皇家统计学会杂志。*B*辑（方法论），539–572。

Ramsay, J. O. 与 Silverman, B. W.（2005）函数型数据分析。第*2*版。 纽约：施普林格。

Rasmussen, C. E. 与 Williams, C. K.（2006年）机器学习中的高斯过程。2006年。麻省理工学院出版社，剑桥，马萨诸塞州，美国，**38**，715–719。

Reiss, P. T., Goldsmith, J., Shang, H. L. 和 Ogden, R. T.（2017）函数型回归方法。国际统计评论，**85**，228–249。

Reiss, P. T., Huo, L., Zhao, Y., Kelly 和 Ogden, R. T.（2015）精神病学神经影像中的小波域回归与预测推断。*The Annals of Applied Statistics*，**9**，1076。

Reiss, P. T. 和 Ogden, R. T.（2007）函数型主成分回归与函数型偏最小二乘。美国统计学会会刊，**102**，984–996。

Rice, J. A. 和 Silverman, B. W.（1991）当数据为曲线时，对均值和协方差结构进行非参数估计。*J. Roy. Statist. Soc. Ser. B*，**53**，233–243。

Ruppert, D., Wand, M. P. 和 Carroll, R. J. (2003) 半参数回归。剑桥统计与概率数学丛书 12。剑桥：剑桥大学出版社。数学评论 *(MathSciNet)*：*MR1998720*。

Shang, H. L. (2013) 针对误差密度未知的非参数函数回归模型的贝叶斯带宽估计。计算统计与数据分析，**67**，185– 198。Silverman, B. W. (1996) 通过选择范数进行平滑函数主成分分析。统计学年刊，**24**，1–24。Su, Y.-R., Di, C.-Z. 和 Hsu, L. (2017) 函数线性模型中的假设检验。生物统计学，**73**，551–561。Swihart, B. J., Goldsmith 和 Crainiceanu, C. M. (2014) 函数型线性模型中函数效应的约束似然比检验。《技术计量学》，**56**，483–493。ter Braak, C. J. 和 de Jong, S. (1998) 偏最小二乘回归的目标函数。化学计量学杂志，**12**，41–54。Wahba, G. (1990) 观测数据的样条模型。工业与应用数学学会。Walz, J. M., Goldman, R. I., Carapezza, M., Muraskin, J., Brown, T. R. 和 Sajda, P. (2013) 同步脑电图-功能性磁共振成像揭示超模态皮层注意网络与脑干之间耦合的时间演变。神经科学杂志，**33**，19212–19222。Wang, B., Chen, T. 和 Xu, A. (2017) 带有函数协变量和多元响应的高斯过程回归。化学计量学与智能实验室系统，**163**，1–6。Wang, J.-L., Chiou, J.-M. 和 M¨uller, H.-G. (2016) 函数型数据分析。《统计学及其应用年度评论》，**3**，257–295。Wold, H. (1973) 非线性迭代偏最小二乘 (nipals) 建模：当前的一些进展。载于多元分析–*III*，383–407。爱思唯尔。Wold, H. (1975) 带潜变量的路径模型：NIPALS 方法。载于定量社会学，307–357。爱思唯尔。Womelsdorf, T., Fries, P., Mitra, P. P. 和 Desimone, R. (2006) 视觉皮层中的伽马波段同步可预测变化检测的速度。《自然》，**439**，733–736。

Wood, S. N. (2011) 半参数广义线性模型的快速稳定限制性最大似然和边际似然估计。《皇家统计学会会刊》*B*辑（统计方法学），**73**，3–36。

Yao, F., M¨uller, H. G. 和 Wang, J.-L. (2005a) 稀疏纵向数据的函数型数据分析。美国统计学会会刊，**100**，577–590。

Yao, F., M¨uller, H.-G. 和 Wang, J.-L. (2005b) 纵向数据的函数型线性回归分析。统计学年刊，**33**，2873–2903。

Zhang, X., King, M. L. 和 Shang, H. L. (2014) 灵活误差密度非参数回归模型中带宽估计的抽样算法。计算统计与数据分析，**78**，218–234。

Zhao, Y., Chen, H. 和 Ogden, R. T.（2015）函数型线性回归中的基于小波的加权套索与筛选方法。计算与图形统计学杂志，**24**，655–675。

Zhao, Y., Ogden, R. T. 和 Reiss, P. T.（2012）函数型线性回归中的基于小波的套索。计算与图形统计学杂志，**21**，600–617。

Zhu, H. 和 Cox, D. D.（2009）利用荧光光谱进行宫颈癌前病变诊断的带曲线选择的函数型广义线性模型。讲义专著系列，173–189。

Zou, H. 和 Hastie, T.（2005）通过弹性网进行正则化和变量选择。*Journal of Royal Statistical Society: Series B (Statistical Methodology)*，**67**，301–320。