Using Fourier Methods for Estimating Central Subspace and Central Mean Subspace

Michael Y. Zhu

Department of Statistics
Purdue University
(joint with Peng Zeng)

Outline

1. Introduction
2. Definitions and existing methods
3. Fourier method for estimating central mean subspace (FMM)
4. Fourier method for estimating central subspace (FMC)
5. Estimate and its asymptotic normality
6. Choosing tuning parameters
7. Simulation examples
8. Conclusion and future work

Introduction

- Response $Y \in \mathbb{R}$ and predictor $X \in \mathbb{R}^p$

$$Y = g(X, \varepsilon)$$

  where $X \perp\!\!\!\perp \varepsilon$, and $E[\varepsilon] = 0$.

- High-dimensional nonparametric regression
- Curse of dimensionality

Structural Nonparametric Approach

Assume some structures on regression function in order to mitigate the curse of dimensionality.

- Projection pursuit regression (Friedman and Stuetzle 1981)

$$Y = g_1(\beta_1^\tau X) + g_2(\beta_2^\tau X) + \cdots + g_M(\beta_M^\tau X) + \varepsilon$$

- Single index model

$$Y = g(\beta^\tau X) + \varepsilon$$

- Multiple index model $(1 < k < p)$

$$Y = g(\beta_1^\tau X, \beta_2^\tau X, \ldots, \beta_k^\tau X) + \varepsilon$$

Regression Graphics Approach

Identify a subspace of X that can *completely* capture the information of X about Y *without estimating the regression function* of Y on X

$\Downarrow$

Visualization

$\Downarrow$

Parametric or Nonparametric Regression of Y on X in the subspace

Three Equivalent Models

regression (Li 1991)	$Y = g(B^\tau X, \varepsilon)$
conditional independence (Cook 1994, 1998)	$Y \perp\!\!\!\perp X \mid B^\tau X$
conditional distribution	$F(Y \mid X) = F(Y \mid B^\tau X)$

where $B = (\beta_1, \ldots, \beta_k)$ is a $p \times k$ matrix, $k < p$,
$g(\cdot)$ is an unknown function, $X \perp\!\!\!\perp \varepsilon$, $E[\varepsilon] = 0$
$F(\cdot \mid \cdot)$ is conditional distribution.

$\mathcal{S}(B)$, space spanned by the columns of B, is called a *dimension reduction subspace* (DRS).

Central Subspace

- DRS is not unique. For any two spaces $\mathcal{S} \subseteq \mathcal{S}'$,

$$\mathcal{S} \text{ is a DRS } \Rightarrow \mathcal{S}' \text{ is a DRS}$$

- *Central subspace* (CS), denoted by $\mathcal{S}_{Y|X}$, is a DRS such that

$$\mathcal{S}_{Y|X} = \bigcap_{\text{all DRS}} \mathcal{S}$$

- $\mathcal{S}_{Y|X}$ is unique and the minimal DRS if it exists.
- Let $Z = \Sigma^{-1/2}(X - E[X])$, where $\Sigma = cov(X)$, then

$$\mathcal{S}_{Y|X} = \Sigma^{-1/2}\mathcal{S}_{Y|Z}$$

- In the following, we assume $E[X] = 0$ and $cov[X] = I_p$.

Example

Suppose $X = (X_1, \ldots, X_5)^\tau \in \mathbb{R}^5$. Consider model

$$\begin{aligned} Y &= X_1 + (X_1 + X_3)^2 + \varepsilon\, X_4 \\ &= g(\beta_1^\tau X, \beta_2^\tau X) + \varepsilon\, h(\beta_3^\tau X) \end{aligned}$$

where $\varepsilon \perp\!\!\!\perp X$ and $E[\varepsilon] = 0$,

$$\begin{aligned} \beta_1 &= (1, 0, 0, 0, 0)^\tau \\ \beta_2 &= (1, 0, 1, 0, 0)^\tau \\ \beta_3 &= (0, 0, 0, 1, 0)^\tau \end{aligned}$$

Therefore,

$$\begin{aligned} \mathcal{S}_{Y|X} &= span\{\beta_1, \beta_2, \beta_3\} \\ \mathcal{S}_{E[Y|X]} &= span\{\beta_1, \beta_2\} \end{aligned}$$

Central Mean Subspace

regression (Li 1992)	$E[Y \mid X] = h(B^\tau X)$
conditional independence (Cook and Li 2002)	$Y \perp\!\!\!\perp E[Y \mid X] \mid B^\tau X$

where B is a $p \times k$, $k < p$, $h(\cdot)$ is an unknown function.

$\mathcal{S}(B)$ is a *mean dimension reduction subspace* (MDRS).

Central mean subspace (CMS), denoted by $\mathcal{S}_{E[Y|X]}$, is a MDRS such that

$$\mathcal{S}_{E[Y|X]} = \bigcap_{\text{all MDRS}} \mathcal{S}$$

Note that $\mathcal{S}_{E[Y|X]} \subseteq \mathcal{S}_{Y|X}$.

A Common Approach for Estimating CS/CMS

- Population version:
  - Find a matrix M depending on X and Y, such that

$$\mathcal{S}(M) \subseteq \mathcal{S}_{Y|X} \quad (\text{or } \mathcal{S}_{E[Y|X]})$$

- Sample version:

  Given a sample (x_i, y_i), $i = 1, 2, \dots, n$,

  1. find an estimate $\hat{M}$ of M.
  2. perform spectral decomposition of $\hat{M}$.
  3. estimate CS (or CMS) by the space spanned by the eigenvectors corresponding to the k largest eigenvalues.

Existing Methods for Estimating CS

- Sliced inverse regression (SIR; Li 1991). Under linear condition,

$$E[X \mid Y] \in \mathcal{S}_{Y|X}, \qquad M_{\mathrm{SIR}} = cov[E[X \mid Y]].$$

$$\text{linear condition:} \quad E[X \mid B^{\tau} X] = P_B X$$

- Sliced average variance estimate (SAVE; Cook and Weisberg 1991). Under linear condition and constant variance condition,

$$M_{\mathrm{SAVE}} = E[(I_p - cov[X \mid Y])^2].$$

$$\text{constant variance condition:} \quad cov[X \mid B^{\tau} X] = I_p - P_B.$$

Existing Methods for Estimating CMS

- Principal Hessian direction (pHd; Li 1992). Under linear condition and constant variance condition,

$$M_{\text{pHd}} = E[(Y - E[Y])XX^{\tau}].$$

  Note that $M_{\text{pHd}} = E\big(\frac{\partial^2}{\partial X \partial^{\tau} X} E[Y \mid X]\big)$ when X is normal.

- Iterative Hessian transformation (IHT; Cook and Li 2002). Under linear condition and constant variance condition,

$$M_{\text{IHT}} = (\beta, M_{\text{pHd}}\beta, M_{\text{pHd}}^2\beta, \ldots, M_{\text{pHd}}^{p-1}\beta)$$

  where $\beta = E[YX]$.

Some Methods from Structural Nonparametric Approach

Need to estimate $\frac{\partial}{\partial X}E[Y \mid X]$.

$$\frac{\partial}{\partial X}E[Y \mid X] = B\frac{\partial}{\partial(B^\tau X)}E[Y \mid B^\tau X] \in \mathcal{S}_{E[Y|X]}$$

Hristache, Juditsky, Polzehl and Spokoiny (2001)

Local linear approximation of regression function at point X_0,

$$E[Y \mid X] \approx a + d^\tau(X - X_0).$$

Xia, Tong, Li, and Zhu (2002)

Local linear approximation of regression function at point X_0,

$$E[Y \mid X] \approx a + b^\tau B^\tau(X - X_0).$$

Comparison between Two Approaches

	regression graphics	structural nonparam.
estimate regression function	No	Yes
distributional assumption	Yes	No
explicit estimate	Yes	No
need iterative algorithm	No	Yes
targeted subspace	CS/CMS	CMS

Limitation of Existing Methods

Limitation of existing methods (Regression Graphics Approach)

- Need distributional assumptions on X
- No theoretical result on when the entire CS/CMS can be recovered

Next: population versions of FMM and FMC.

Assume all the density functions exist and are differentiable. The model is

$$f_{Y|X}(y \mid x) = f_{Y|B^\tau X}(y \mid B^\tau x) = h(y; B^\tau x)$$

where f is the conditional density function.

Estimate CMS

Heuristics: Find some vectors that belong to CMS, and let them span the whole CMS.

- $m(X) = E[Y \mid X]$ is a function of $U = B^\tau X$, then

$$\frac{\partial}{\partial X} m(X) = B \frac{\partial}{\partial U} m(U) \in \mathcal{S}_{E[Y|X]}$$

- For any $\omega \in \mathbb{R}^p$,

$$\begin{aligned} \psi(\omega) &= \int \exp\{\imath\, \omega^\tau X\} \left(\tfrac{\partial}{\partial X} m(X)\right) f_X(X)\, dX \\ &= -E_{(X,Y)}\left[Y(\imath\omega + G(X)) \exp\{\imath\, \omega^\tau X\}\right] \quad \in \mathcal{S}_{E[Y|X]} \end{aligned}$$

where f_X is the density function of X, and $G(X) = \frac{\partial}{\partial X} \log f_X(X)$.

FMM

Because $\psi(\omega) \in \mathcal{S}_{E[Y|X]}$, then

$$\mathcal{S}(\psi(\omega)\bar{\psi}(\omega)^\tau) \subseteq \mathcal{S}_{E[Y|X]}, \quad \text{for any } \omega \in \mathbb{R}^p$$

Proposition 1. *Define matrix*

$$M_{FMM} = \int \psi(\omega)\bar{\psi}(\omega)^\tau K(\omega)\, d\omega$$

where $K(\omega)$ is a positive weight function on $\mathbb{R}^p$. Then M_{FMM} is non-negative definite, and

$$\mathcal{S}(M_{FMM}) = \mathcal{S}_{E[Y|X]}$$

FMM When $K(\omega)$ is Gaussian Function

- Suppose $K(\omega) = (2\pi\sigma_\omega^2)^{-p/2} \exp\{-\|\omega\|^2/2\sigma_\omega^2\}$, then

$$M_{\mathrm{FMM}} = E\left[Y_1 Y_2\, e^{-\frac{\sigma_\omega^2}{2}\|X_{12}\|^2}\left[\sigma_\omega^2 I_p + (G(X_1) - \sigma_w^2 X_{12})(G(X_2) + \sigma_w^2 X_{12})^\tau\right]\right]$$

  where $X_{12} = X_1 - X_2$.

- σ_ω^2 is different from bandwidth in kernel estimation.
  - Proposition 1 is valid for any σ_ω^2.
- Other weight function can also be used.

Connection Between FMM and ADE

Consider single index model

$$m(X) = E[Y \mid X] = g(\beta^\tau X)$$

The average derivative estimate (ADE; Härdle and Stoker 1989) is

$$E[\tfrac{\partial}{\partial X} m(X)] = -E[YG(X)] \propto \beta$$

- $\psi(\omega)$ is a generalization of ADE.

$$\psi(\omega) = E[\exp\{\imath\, \omega^\tau X\} \left(\tfrac{\partial}{\partial X} m(X)\right)]$$

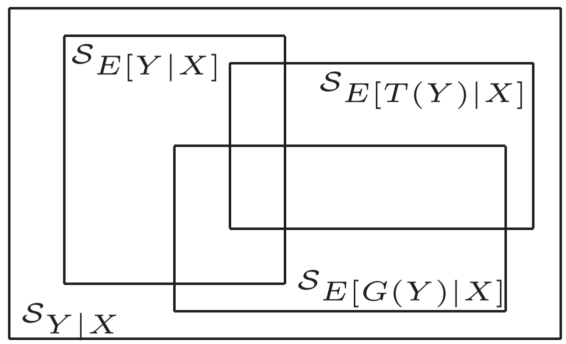
- When $K(\omega)$ is point mass on origin ($\sigma_w^2 = 0$),

$$M_{\text{FMM}} = E\big[YG(X)\big]E\big[YG(X)\big]^\tau$$

In a model with multiple directions, *FMM can estimate multiple directions when $\sigma_\omega^2 \neq 0$, while ADE can only estimate single direction.*

Heuristics on Estimating CS

Relationship between the central subspace and the central mean subspace. $T(Y)$ and $G(Y)$ are two transformations of Y.



It is possible to recover $\mathcal{S}_{Y|X}$ by all possible central mean subspaces.

$$\mathcal{S}_{Y|X} = \sum_{\text{all possible } T} \mathcal{S}_{E[T(Y)|X]}$$

Represent CS by CMSs

- Define transformation,

$$T_t(Y) = \exp\{\imath\, tY\} = \cos(tY) + \imath \sin(tY), \quad t \in \mathbb{R}.$$

- $E[T_t(Y) \mid X]$ is the Fourier transform (characteristic function) of $f_{Y|X}$.

$$m_t(X) = E[T_t(Y) \mid X] = \int \exp\{\imath\, tY\}\, f_{Y|X}(Y \mid X)\, dY.$$

Proposition 2. *CS can be represented as the sum of a family of CMSs.*

$$\mathcal{S}_{Y|X} = \sum_{t\in\mathbb{R}} \mathcal{S}_{E[T_t(Y)|X]} = \sum_{t\in\mathbb{R}} \mathcal{S}_{E[\cos(tY)|X]} + \sum_{t\in\mathbb{R}} \mathcal{S}_{E[\sin(tY)|X]}$$

FMC

Define

$$\begin{aligned}\phi(t,\omega) &= \int \exp\{\imath\,\omega^\tau X\}\left(\tfrac{\partial}{\partial X}\, m_t(X)\right) f_X(X)\, dX \\ &= -E_{(X,Y)}\left[(\imath\omega + \tfrac{\partial}{\partial X}\log f_X(X))\,\exp\{\imath tY + \imath\,\omega^\tau X\}\right] \in \mathcal{S}_{E[Y|X]}\end{aligned}$$

It is obtained by substituting Y by $\exp\{\imath tY\}$ in $\psi(\omega)$.

Proposition 3. *Define matrix*

$$M_{FMC} = \iint \phi(t,\omega)\bar{\phi}(t,\omega)^\tau K(\omega)k(t)\, d\omega dt$$

where $K(\omega)$ and $k(t)$ are positive weight functions. Then M_{FMC} is non-negative definite, and

$$\mathcal{S}(M_{FMC}) = \mathcal{S}_{Y|X}$$

FMC when $K(\omega)$ and $k(t)$ are Gaussian Functions

- When $k(t) = (2\pi\sigma_t^2)^{-1/2}\exp\{-t^2/2\sigma_t^2\}$, and $K(\omega)$ as before,

$$M_{\text{FMC}} = E\left[a_{12}\left[\sigma_\omega^2 I_p + (G(X_1) - \sigma_w^2 X_{12})(G(X_2) + \sigma_w^2 X_{12})^\tau\right]\right]$$

  where $a_{12} = \exp\{-\sigma_t^2(Y_1 - Y_2)^2/2 - \sigma_\omega^2\|X_1 - X_2\|^2/2\}$.

- σ_ω^2 and σ_t^2 are different from bandwidth in kernel estimation.
  - Proposition 3 is valid for any σ_ω^2 and σ_t^2.
- Other weight function can also be used.

Connection between FMC and SIR

- When $K(\omega)$ is point mass on origin ($\sigma_\omega^2 = 0$),

$$M_{\mathrm{FMC}} = \int E[e^{\imath tY} E[G(X) \mid Y]] E[e^{-\imath tY} E[G(X) \mid Y]]^\tau \, k(t) \, dt$$

  Therefore,

$$\begin{aligned}\mathcal{S}(M_{\mathrm{FMC}}) &= span\{E[e^{\imath tY} E[G(X) \mid Y]], t \in \mathbb{R}\} \\ &= span\{E[G(X) \mid Y], Y \in supp(Y)\}\end{aligned}$$

- When X is normal, then $G(X) = -X$,

$$\mathcal{S}(M_{\mathrm{FMC}}) = span\{E[X \mid Y], Y \in supp(Y)\} = \mathcal{S}(M_{\mathrm{SIR}})$$

- *FMC degenerates to SIR when $\sigma_\omega^2 = 0$ and X is normal.*

Summary of Population Versions

In population version, FMC/FMM can

- work for X with almost arbitrary distribution
- recover the *entire* central subspace (central mean subspace)

Next: sample version:

Given a sample (x_i, y_i), $i = 1, 2, \ldots, n$,

- estimate $\hat{M}_{\text{FMC}}$
- asymptotic property

The discussion of FMM is similar and is omitted.

Estimation of $M_{\mathbf{FMC}}$

Note that

$$M_{\text{FMC}} = E\left[a_{12}\left[\sigma_\omega^2 I_p + (G(X_1) - \sigma_w^2 X_{12})(G(X_2) + \sigma_w^2 X_{12})^\tau\right]\right]$$

where $a_{12} = \exp\{-\sigma_t^2 (Y_1 - Y_2)^2/2 - \sigma_\omega^2 \|X_1 - X_2\|^2/2\}$. It can be estimated by sample average, and the only unknown component is

$$G(x_i) = \frac{\partial}{\partial X} \log f_X(x_i) = \frac{\frac{\partial}{\partial X} f_X(x_i)}{f_X(x_i)}$$

How to estimate $G(x_i)$?

- If the density function is known, $G(x_i)$ can be directly calculated.
- If X belongs to a family of distributions indexed by parameter θ, $G(x_i)$ is a function of θ. It can be estimated by substituting θ by its MLE.
- If there is no information about the density function, kernel density estimator can be used to estimate $G(x_i)$.

When X Follows Normal Distribution

- Suppose X is normally distributed with $E[X] = 0$ and $cov[X] = I_p$.

$$G(x_i) = -x_i$$

- We can have an estimates of M_{FMC}.

$$\boxed{\hat{M}_{\text{FMC}} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \left[\sigma_\omega^2 I_p + (x_i + \sigma_w^2 x_{ij})(x_j - \sigma_w^2 x_{ij})^\tau \right]}$$

where $a_{ij} = \exp\{-\sigma_t^2 y_{ij}^2/2 - \sigma_\omega^2 x_{ij}^\tau x_{ij}/2\}$, $x_{ij} = x_i - x_j$ and $y_{ij} = y_i - y_j$.

Asymptotic Results

By Hoeffding decomposition,

$$\hat{M}_{FMC} = M_{FMC} + \frac{1}{n}\sum_{i=1}^{n}\big(A(x_i, y_i) + A(x_i, y_i)^\tau - 2M_{\text{FMC}}\big) + o_p(n^{-1/2})$$

where

$$A(y_i, x_i) = E\big[a_{12}\left[\sigma_\omega^2 I_p + (x_1 + \sigma_w^2 x_{12})(x_2 - \sigma_w^2 x_{12})^\tau \mid (x_1, y_1) = (x_i, y_i)\right]$$

Proposition 4. *When $n \to \infty$,*

$$\sqrt{n}\,(vec(\hat{M}_{FMC})) - vec(M_{FMC})) \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

where Σ is a positive definite matrix.

General Cases

- Estimate $G(x_i)$ by plugging in kernel estimate

$$\hat{G}(x_i) = \frac{\frac{\partial}{\partial X}\hat{f}_h(x_i)}{\hat{f}_h(x_i)}$$

where

$$\hat{f}_h(x_i) = \frac{1}{nh^p}\sum_{\ell=1}^{n} W\left(\frac{x_i - x_\ell}{h}\right)$$

$$\frac{\partial}{\partial X}\hat{f}_h(x_i) = \frac{1}{nh^{p+1}}\sum_{\ell=1}^{n} W'\left(\frac{x_i - x_\ell}{h}\right)$$

and $W(\cdot)$ is a kernel function, h is the bandwidth.

Estimate in General Cases

We have estimates

$$\hat{M}_{\text{FMC}} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \left[\sigma_\omega^2 I_p + (\hat{G}(x_i) - \sigma_w^2 x_{ij})(\hat{G}(x_j) + \sigma_w^2 x_{ij})^\tau \right] \hat{I}_i \hat{I}_j$$

where $a_{ij} = \exp\{-\sigma_t^2 y_{ij}^2/2 - \sigma_\omega^2 x_{ij}^\tau x_{ij}/2\}$, $x_{ij} = x_i - x_j$, and $y_{ij} = y_i - y_j$, $\hat{I}_i = I_{[\hat{f}_h(x_i) > b]}$, $I_{[\cdot]}$ is an indicator function, and b is a threshold.

The technique of using $I_{[\cdot]}$ is called *trimming*. Its purpose is to rule out the cases that the estimated densities are extremely small.

Asymptotic Results

Proposition 5. *Under some regularity conditions, if* $f_X(x)$ *has partial derivatives up to order* $r \geq p+2$, *and*

(1) $n \to \infty$, $h \to 0$, $b \to 0$ *and* $b^{-1}h \to 0$;

(2) for some $\varepsilon > 0$, $b^4 n^{1-\varepsilon} h^{2p+2} \to \infty$;

(3) $nh^{2r-2} \to 0$

then

$$\sqrt{n}\,(vec(\hat{M}_{FMC}) - vec(M_{FMC})) \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

where Σ *is a positive definite matrix.*

Sketch of Proof

- When b is small enough, the effect of trimming is negligible.
- Linearization.

$$\hat{G}(x_i) = \frac{\frac{\partial}{\partial X}\hat{f}_h(x_i)}{\hat{f}(x_i)} = G(x_i) + \frac{\frac{\partial}{\partial X}\hat{f}_h(x_i)}{f(x_i)} - \frac{\hat{f}_h(x_i)}{f(x_i)}G(x_i) + o_p(n^{-1/2})$$

  Then there is no kernel estimator in the denominators.

- Substitute $\hat{G}(x_i)$ by its approximation in $\hat{M}$, and write it as a U-statistic.
- Asymptotic normality is obtained by Hoeffding decomposition and central limit theorem.

Intuition for $\sqrt{n}$ Convergence Rate

- Analog to semi-parametric model.
  - central subspace (central mean subspace) is essentially the parametric part in a semi-parametric model.
- Under some conditions, $\sqrt{n}$ convergence rate is possible for the estimate of an integral functional of density and its derivatives

$$\int T(X, f(X), f'(X), \dots, f^{(v)}(X))\, dX$$

  See Bickel and Ritov (1988), Birge and Massart (1995).
- Although individual $\hat{G}(x_i)$ cannot achieve $\sqrt{n}$ convergence rate, their average can converge more quickly. Same phenomenon was observed in Härdle and Stoker (1989).

Generic Estimation Procedure

Suppose we have observations (x_i, y_i), $i = 1, \ldots, n$.

1. Specify parameters: k, σ_ω^2, σ_t^2, and h, if applicable.
2. Standardize data by $\tilde{x}_i = \hat{\Sigma}^{-1/2}(x_i - \bar{x})$ and $\tilde{y}_i = (y_i - \bar{y})/s_y$
3. Calculate an estimate $\hat{M}$ of M_{FMC} (or M_{FMM}) using data $(\tilde{x}_i, \tilde{y}_i)$.
4. Perform spectral decomposition of $\hat{M}$. The eigenvalues are $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$, and their corresponding eigenvectors are $\hat{\gamma}_1, \ldots \hat{\gamma}_p$.
5. Estimate $\mathcal{S}_{Y|X}$ (or $\mathcal{S}_{E[Y|X]}$) by $\hat{\mathcal{S}} = span\{\hat{\Sigma}^{-1/2}\hat{\gamma}_1, \ldots, \hat{\Sigma}^{-1/2}\hat{\gamma}_k\}$.

Discrepancy Measure

- If A and B are two matrices, $P_A = A(A^\tau A)^- A^\tau$ and $P_B = B(B^\tau B)^- B^\tau$ are projection matrices, then

$$r^2 = \frac{1}{k} trP_A P_B$$

where $k = rank(A) = rank(B)$, and r is called the trace correlation.

- Define

$$D(\hat{\mathcal{S}}, \mathcal{S}) = 1 - r$$

- $0 \leq D(\hat{\mathcal{S}}, \mathcal{S}) \leq 1, \quad D(\hat{\mathcal{S}}, \mathcal{S}) = 0$, if $\hat{\mathcal{S}} = \mathcal{S}, \quad D(\hat{\mathcal{S}}, \mathcal{S}) = 1$, if $\hat{\mathcal{S}} \perp \mathcal{S}$.

Choose Tuning Parameters

- k: assume known.
- $\sigma_\omega^2, \sigma_t^2$: recommend $\sigma_\omega^2 = 0.1$ and $\sigma_t^2 = 1.0$.
- h: use optimal bandwidth given by kernel density estimator.

or

- Choose the parameters that minimize $D(\hat{\mathcal{S}}, \mathcal{S})$.
- If $\mathcal{S}$ is unknown, bootstrap procedure can be used (Ye and Weiss 2003).

Example 1

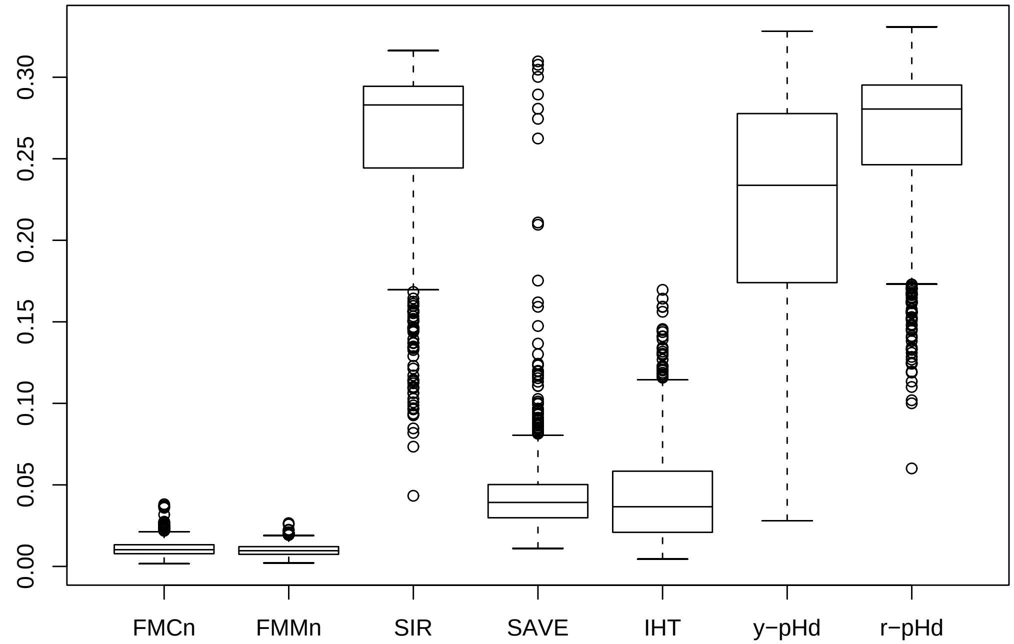
Consider model

$$Y = 2(\beta_1^\tau X) + (\beta_2^\tau X)^2 + \varepsilon$$

where $X \sim N(0, I_{10})$. $\varepsilon \perp\!\!\!\perp X$ and $\varepsilon \sim N(0,1)$. Sample size is 500.

$$\beta_1 = (1,1,1,0,0,0,0,0,0,0)^\tau$$
$$\beta_2 = (0,0,0,0,0,0,0,1,1,1)^\tau$$

1000 replicates are used to construct boxplots.



FMCn: $\sigma_\omega^2 = 0.1$, $\sigma_t^2 = 1.0$. FMMn: $\sigma_\omega^2 = 0.1$.

SIR, SAVE: $H = 5$.

Example 2: Heteroscedastic Model

Consider model

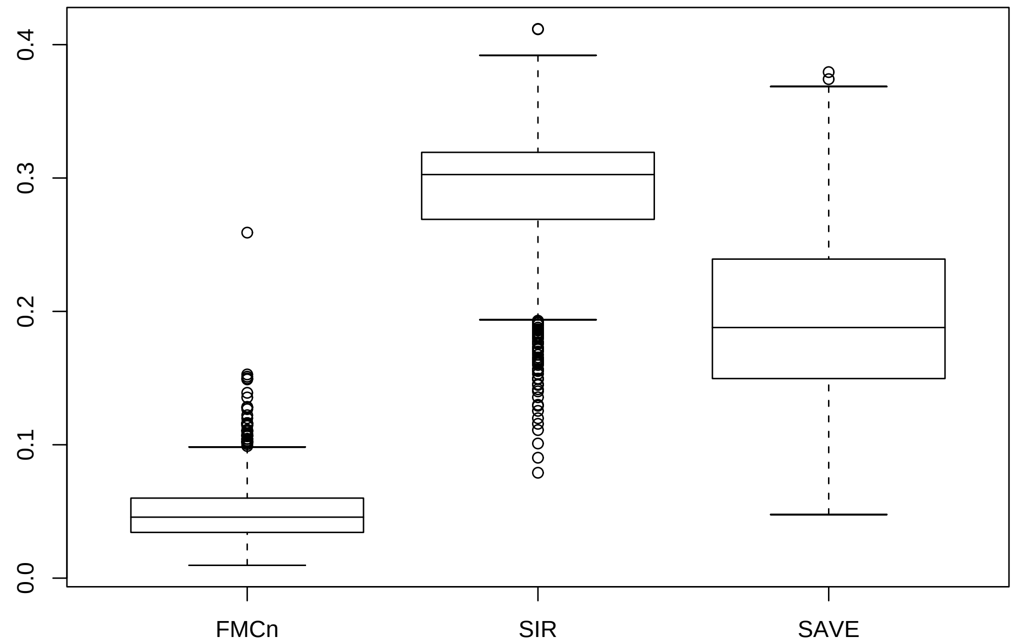
$$Y = (\beta_1^\tau X) + 4(\beta_2^\tau X)\,\varepsilon$$

where $X \sim N(0, I_{10})$. $\varepsilon \perp\!\!\!\perp X$ and $\varepsilon \sim N(0,1)$. Sample size is 500.

$$\beta_1 = (1,1,1,0,0,0,0,0,0,0)^\tau$$

$$\beta_2 = (0,0,0,0,0,0,0,1,1,1)^\tau$$

1000 replicates are used to construct boxplots.



FMCn: $\sigma_\omega^2 = 0.1$, $\sigma_t^2 = 1.0$. SIR, SAVE: $H = 5$.

Example 3: Y is Discreet

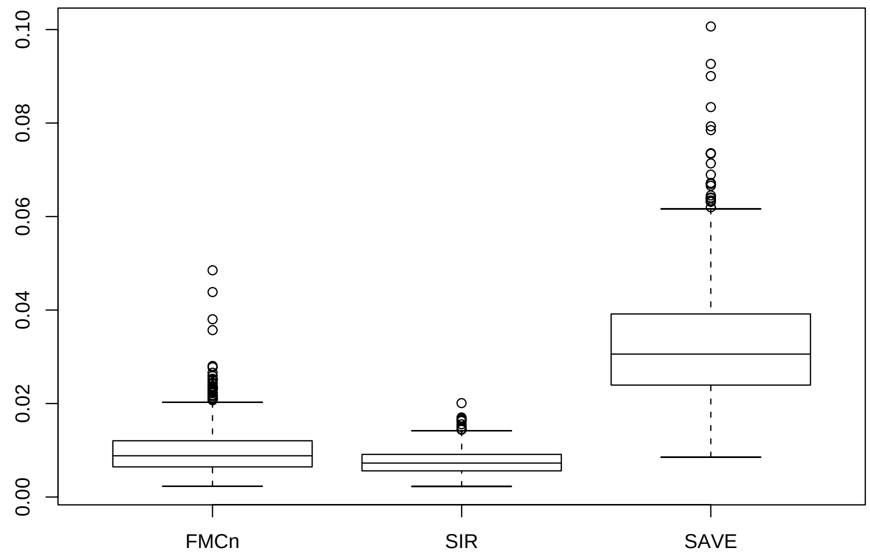
Consider model

$$Y = I_{[\beta_1^\tau X + 0.2\varepsilon_1 > 1]} + 2I_{[\beta_2^\tau X + 0.2\varepsilon_2 > 0]}$$

where $X \sim N(0, I_{10})$. $\varepsilon \perp\!\!\!\perp X$ and $\varepsilon \sim N(0,1)$. $I_{[\cdot]}$ is indicator function, and $Y = 0, 1, 2, 3$. Sample size is 500.

$$\beta_1 = (1,1,1,0,0,0,0,0,0,0)^\tau$$
$$\beta_2 = (0,0,0,0,0,0,0,1,1,1)^\tau$$

1000 replicates are used to construct boxplots.



FMCn: $\sigma_\omega^2 = 0.1$, $\sigma_t^2 = 1.0$. SIR, SAVE: $H = 5$.

Example 4: Mixture Normal

Consider model

$$Y = \frac{\beta_1^\tau X}{3 + (2 + \beta_2^\tau X)^2} + 0.1\varepsilon$$

where $X \sim 0.5N(-d\beta_3, I_{10}) + 0.5N(d\beta_3, I_{10})$, $\varepsilon \sim N(0,1)$ and $\varepsilon \perp\!\!\!\perp X$.
Sample size is 500.

$$\beta_1 = (1,1,1,0,0,0,0,0,0,0)^\tau$$
$$\beta_2 = (0,0,0,0,0,0,0,1,1,1)^\tau$$
$$\beta_3 = (0,0,0,0,1,1,0,0,0,0)^\tau$$

Let $d = 0, 1, 2, \ldots, 10$.

1000 replicates are used to compare the performance of different methods.

Mean and standard deviation of 1000 replicates for each case

	mean of $D(\hat{\mathcal{S}}, \mathcal{S})$'s				std. dev. of $D(\hat{\mathcal{S}}, \mathcal{S})$'s			
d	0	3	7	10	0	3	7	10
SIR (s)	0.025	0.040	0.079	0.094	0.012	0.042	0.077	0.082
SAVE (a)	0.147	0.589	0.656	0.643	0.048	0.308	0.316	0.318
IHT (i)	0.068	0.150	0.240	0.253	0.028	0.071	0.065	0.061
y-pHd (y)	0.083	0.274	0.375	0.394	0.039	0.125	0.142	0.143
r-pHd (r)	0.025	0.104	0.263	0.290	0.010	0.087	0.122	0.126
FMC-n (c)	0.020	0.018	0.018	0.018	0.009	0.009	0.009	0.008
FMM-n (m)	0.031	0.028	0.028	0.028	0.015	0.014	0.013	0.015
FMC-k (k)	0.088	0.129	0.185	0.191	0.063	0.083	0.082	0.082
FMM-k (q)	0.022	0.020	0.020	0.020	0.013	0.012	0.014	0.013

SIR, SAVE: $H = 5$. FMCn: $\sigma_\omega^2 = 0.1$, $\sigma_t^2 = 1.0$. FMMn: $\sigma_\omega^2 = 0.1$.

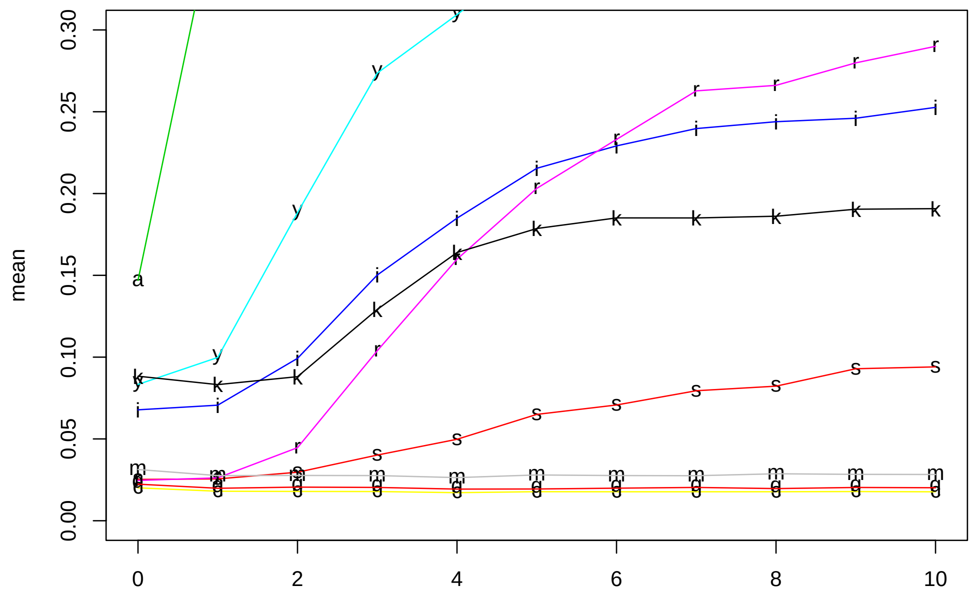
FMCk: $\sigma_\omega^2 = 0.1$, $\sigma_t^2 = 1.0$, $h = 1.0$. FMMk: $\sigma_\omega^2 = 0.1$, $h = 1.0$.

The best three:

FMC-n (c),

FMM-k (q),

FMM-n (m)

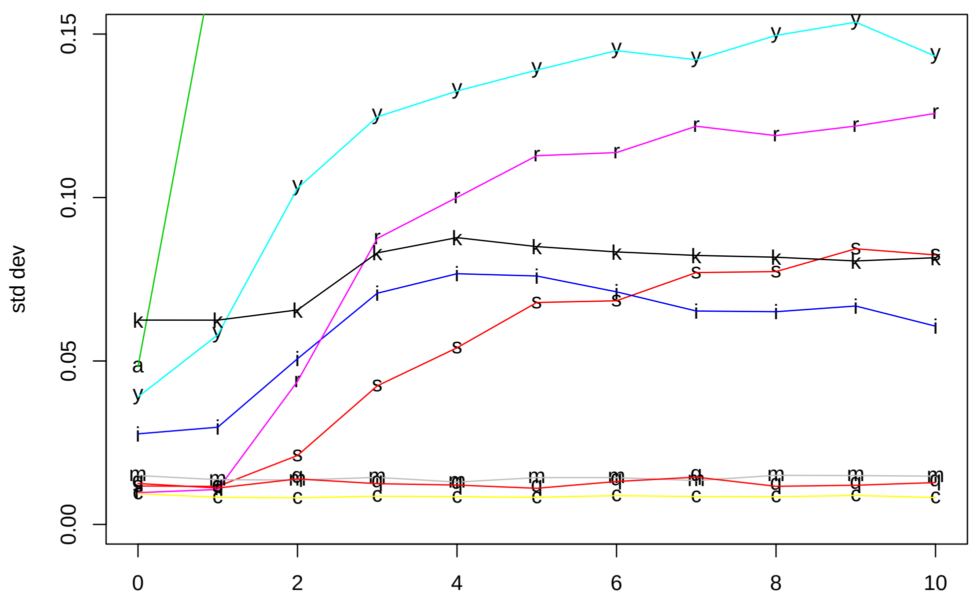


The best three:

FMC-n (c),

FMM-k (q),

FMM-n (m)



Conclusion

- FMC (FMM) can recover the *entire* central subspace (central mean subspace).
- Represent central subspace as the sum of a family of central mean subspaces.
- Implement FMC/FMM under normal distribution and general case.
- Establish the connection of FMC/FMM with other existing methods.

Further Research Direction

- Further study statistical properties of FMM/FMC, including the optimal selection of k, h, σ_ω^2 and σ_t^2.
- Categorical or ordinal predictor variables.
- Dependent data and time series data.
- Robustness.
  - when the model is not true?
  - when the distributional assumption is not true?
- Theoretical comparison with the existing methods.