Nonparametric Independence Screening

Yang Feng

Columbia University

With **Prof. Jianqing Fan and Rui Song**

2011 Winter Workshop on High Dimensional Inference, UFL

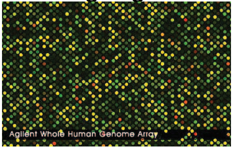
Outline

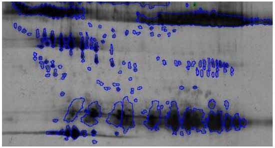
Introduction

Rise of high-dimensionality

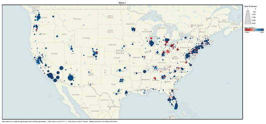
High-dim characterizes many statistical problems:

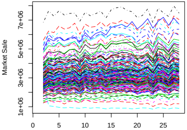
- **Biological science**: disease classification / predicting clinical outcomes using high-throughput data; association studies;





- **Engineering**: Doc or text classification, computer vision.
- **Economics, Finance, Marketing**: sale data collected in many regions.

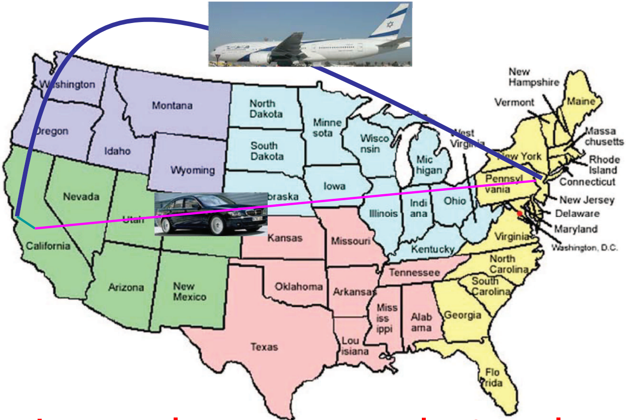




- **Spatial-temporal**: Meteorology; Earth Sciences; Ecology

■ Computational cost ■ Stability

■ Estimation accuracy: ★ noise accumulation ★ spurious corr



Key Idea: **Large-scale** screening + **moderate-scale** searching.

Correlation Screening

Sure independence screening: By using **correlation ranking** $r_i = |\text{corr}(X_i, Y)|$ (Fan and Lv, 2008),

★ reduce dim from $p = O(\exp(n^a))$ to $d = o(n)$

★ Limitations: ■Linear models. ■Joint normality.

$$Y = \sum_{j \in \mathcal{M}_*} \beta_j X_j + \varepsilon$$

Extensions – independence learning

★ Fan and Song (2010) unveil the results in GLIM;
■ remove joint normality; ■ specify capacity of reduction.

★ Even for linear model, marginal regression is **not** necessarily linear.This led to **Generalized corr ranking**:
$r_i = \left|\mathrm{corr}\left((X_i, X_i^2, \cdots, X_i^k), Y\right)\right|$ (*Hall and Miller, 09*)

Other methods: ★**Data-tilting**; (*Hall, Titterington & Xue, 09*); ★**Marginal LR** (*Fan, Samworth & Wu, 09*); ★**MPLE** (*Zhao & Li, 09*);

Extensions – independence learning

★ Fan and Song (2010) unveil the results in GLIM;
■remove joint normality; ■specify capacity of reduction.

★ Even for linear model, marginal regression is **not** necessarily linear.This led to **Generalized corr ranking**:
$r_i = \left|\mathrm{corr}((X_i, X_i^2, \cdots, X_i^k), Y)\right|$ (*Hall and Miller, 09*)

Other methods: ★**Data-tilting**; (*Hall, Titterington & Xue, 09*);
★**Marginal LR** (*Fan, Samworth & Wu, 09*); ★**MPLE** (*Zhao & Li, 09*);

Questions

1. Can we allow sparse high-dim **nonlinear** regression?

$$Y = \sum_{j \in \mathcal{M}_\star} m_j(x_j) + \varepsilon.$$

2. Can we have model selection consistency?
3. Can we have sure screening property? In what capacity?
4. How to choose a thresholding parameter?

Questions

1. Can we allow sparse high-dim **nonlinear** regression?

$$Y = \sum_{j \in \mathcal{M}_\star} m_j(x_j) + \varepsilon.$$

2. Can we have model selection consistency?
3. Can we have sure screening property? In what capacity?
4. How to choose a thresholding parameter?

Nonparametric Independence Screening

Marginal nonparametric regression

Additive model: (*Stone, 1985*)

$$Y = \sum_{j \in \mathcal{M}_\star} m_j(X_j) + \varepsilon$$

but applicable to $Y = f(X_{\mathcal{M}_\star}) + \varepsilon$ with $\|E(Y|X_j)\| > 0, \forall j \in \mathcal{M}_\star$.

B-spline basis: $\Psi_j \equiv \Psi_j(X_j) = (\Psi_1(X_j), \cdots, \Psi_{d_n}(X_j))^T$

Marginal regressions: with $f_{nj}(x) = \sum_{k=1}^{d_n} \beta_{jk} \Psi_{jk}(x)$

$$\min_{f_{nj} \in \mathcal{S}_n} \mathbb{P}_n \Big(Y - f_{nj}(X_j) \Big)^2 = \min_{\beta_j \in \mathbb{R}^{d_n}} \mathbb{P}_n \Big(Y - \Psi_j^T \beta_j \Big)^2,$$

Challenges: ■ growing d_n; ■ Non-Gaussian random matrices

Marginal nonparametric regression

Additive model: (*Stone, 1985*)

$$Y = \sum_{j \in \mathcal{M}_\star} m_j(X_j) + \varepsilon$$

but applicable to $Y = f(X_{\mathcal{M}_\star}) + \varepsilon$ with $\|E(Y|X_j)\| > 0, \forall j \in \mathcal{M}_\star$.

B-spline basis: $\Psi_j \equiv \Psi_j(X_j) = (\Psi_1(X_j), \cdots, \Psi_{d_n}(X_j))^T$

Marginal regressions: with $f_{nj}(x) = \sum_{k=1}^{d_n} \beta_{jk} \Psi_{jk}(x)$

$$\min_{f_{nj} \in \mathcal{S}_n} \mathbb{P}_n \left(Y - f_{nj}(X_j) \right)^2 = \min_{\beta_j \in \mathbb{R}^{d_n}} \mathbb{P}_n \left(Y - \Psi_j^T \beta_j \right)^2,$$

Challenges: ■ growing d_n; ■ Non-Gaussian random matrices

Marginal nonparametric regression

Additive model: (*Stone, 1985*)

$$Y = \sum_{j \in \mathcal{M}_\star} m_j(X_j) + \varepsilon$$

but applicable to $Y = f(X_{\mathcal{M}_\star}) + \varepsilon$ with $\|E(Y|X_j)\| > 0, \forall j \in \mathcal{M}_\star$.

B-spline basis: $\Psi_j \equiv \Psi_j(X_j) = (\Psi_1(X_j), \cdots, \Psi_{d_n}(X_j))^T$

Marginal regressions: with $f_{nj}(x) = \sum_{k=1}^{d_n} \beta_{jk} \Psi_{jk}(x)$

$$\min_{f_{nj} \in \mathcal{S}_n} \mathbb{P}_n \left(Y - f_{nj}(X_j) \right)^2 = \min_{\beta_j \in \mathbb{R}^{d_n}} \mathbb{P}_n \left(Y - \Psi_j^T \beta_j \right)^2,$$

Challenges: ■ growing d_n; ■ Non-Gaussian random matrices

Nonparametric Independence Screening (NIS)

- Select variables according to the marginal utilities.

1. **Marginal Magnitude**: $\widehat{\mathcal{M}}_{\nu_n} = \{j : \|\hat{f}_{nj}\|_n^2 \geq \nu_n\}$, where $\|\hat{f}_{nj}\|_n^2 = n^{-1}\sum_{i=1}^n \hat{f}_{nj}(X_{ij})^2$.
2. **Marginal RSS**: $\widehat{\mathcal{M}}_{\gamma_n} = \{j : u_j \leq \gamma_n\}$, with $u_j = \min_{\beta_j} \mathbb{P}_n(Y - \Psi_j^T \beta_j)^2$ is RSS of marginal fit.

- They are equivalent, since $u_j = \mathbb{P}_n(Y^2 - \hat{f}_{nj}^2)$.

Nonparametric Independence Screening (NIS)

- Select variables according to the marginal utilities.

1. **Marginal Magnitude**: $\widehat{\mathcal{M}}_{\nu_n} = \{j : \|\hat{f}_{nj}\|_n^2 \geq \nu_n\}$, where $\|\hat{f}_{nj}\|_n^2 = n^{-1}\sum_{i=1}^n \hat{f}_{nj}(X_{ij})^2$.

2. **Marginal RSS**: $\widehat{\mathcal{M}}_{\gamma_n} = \{j : u_j \leq \gamma_n\}$, with $u_j = \min_{\beta_j} \mathbb{P}_n(Y - \Psi_j^T \beta_j)^2$ is RSS of marginal fit.

- They are equivalent, since $u_j = \mathbb{P}_n(Y^2 - \hat{f}_{nj}^2)$.

Choice of thresholding parameter

Principled SIS Zhao and Li (09) proposed using upper α (Control of FDR) quantile of marginal utilities for decoupled response and covariate (PSIS).

- Obtain the decoupled synthetic data $\{(\mathbf{X}_{\pi(i)}, Y_i)\}_{i=1}^n$ —Marginal distributions are untouched;
- Compute $a_n^* = \max_j \|\hat{f}_{nj}^*\|_n^2$;
- Choose the top α-quantile of a_n^* as ν_n.

Remark: We can take a_n^* based on one permutation.

Choice of thresholding parameter

Principled SIS Zhao and Li (09) proposed using upper α (Control of FDR) quantile of marginal utilities for decoupled response and covariate (PSIS).

- Obtain the decoupled synthetic data $\{(\mathbf{X}_{\pi(i)}, Y_i)\}_{i=1}^n$ —Marginal distributions are untouched;
- Compute $a_n^* = \max_j \|\hat{f}_{nj}^*\|_n^2$;
- Choose the top α-quantile of a_n^* as ν_n.

Remark: We can take a_n^* based on one permutation.

Potential Drawbacks (Fan & Lv, 2008)

- **False Negative**: What if X_4 marginally uncorrelated with Y, but jointly correlated with Y?

$$Y = X_1 + X_2 + X_3 + \boldsymbol{\beta_4 X_4} + \varepsilon \quad \text{s.t.} \quad \operatorname{cov}(Y, X_4) = 0.$$

- **False Positive**: What if $X_2, \cdots, X_{99}$ highly correlated with an important X_1, but weakly correlated with Y conditionally?

$$Y = X_1 + \mathbf{0.2} X_{100} + \varepsilon$$

Potential Drawbacks (Fan & Lv, 2008)

- **False Negative**: What if X_4 marginally uncorrelated with Y, but jointly correlated with Y?

$$Y = X_1 + X_2 + X_3 + \beta_4 \mathbf{X_4} + \varepsilon \quad \text{s.t.} \quad \text{cov}(Y, X_4) = 0.$$

- **False Positive**: What if $X_2, \cdots, X_{99}$ highly correlated with an important X_1, but weakly correlated with Y conditionally?

$$Y = X_1 + \mathbf{0.2} X_{100} + \varepsilon$$

INIS Algorithm

INIS Algorithm

1. **Large-scale screening**: Apply NIS to pick a set $\mathcal{A}_1$.
2. **Moderate-scale selection**: Employ a penalized method such as penGAM (*Meier et. al, 09*) or SpAM (*Ravikumar et al. (09)*) to select a subset $\mathcal{M}_1$

Conditional screening

3. **Large-scale conditional screening**: Rank features according to additional **conditional** contribution:

$$L_j = \min_{\{f_{ni}, i \in \mathcal{M}_1\}, f_{nj}} \mathbb{P}_n \Big(Y - \sum_{\mathbf{i} \in \mathcal{M}_1} f_{ni}(X_i) - f_{nj}(X_j) \Big)^2 .$$

Pick a set $\mathcal{A}_2$ according to $\{L_j, j \in \mathcal{M}_1^c\}$.

Iterations and Remarks

4. **Moderate-scale selection**: As in Step 2, among variables in $\mathcal{M}_1 \cup \mathcal{A}_2$, use a penalized method to select a set $\mathcal{M}_2$ **—Allow deletion**.

5. We iterate Steps 3-4 until convergence.

Greedy-INIS: Restricting size of set $\mathcal{A}_j$ to be 1.

- Extremely fast to compute.
- Very effective when covariates are **highly correlated**.

Iterations and Remarks

4. **Moderate-scale selection**: As in Step 2, among variables in $\mathcal{M}_1 \cup \mathcal{A}_2$, use a penalized method to select a set $\mathcal{M}_2$ **—Allow deletion**.

5. We iterate Steps 3-4 until convergence.

Greedy-INIS: Restricting size of set $\mathcal{A}_j$ to be 1.

- Extremely fast to compute.
- Very effective when covariates are **highly correlated**.

Iterations and Remarks

4. **Moderate-scale selection**: As in Step 2, among variables in $\mathcal{M}_1 \cup \mathcal{A}_2$, use a penalized method to select a set $\mathcal{M}_2$ **—Allow deletion**.

5. We iterate Steps 3-4 until convergence.

Greedy-INIS: Restricting size of set $\mathcal{A}_j$ to be 1.

■Extremely fast to compute.

■Very effective when covariates are **highly correlated**.

Sure Screening Properties

Theoretical basis: Minimum signal

True model: $\mathcal{M}_\star = \{j : Em_j(X_j)^2 > 0\}$ or
$Y \perp \mathbf{X}_{\mathcal{M}_\star^c}$ given $\mathbf{X}_{\mathcal{M}_\star}$.

Assumption: ■ $f_j = E(Y|X_j) \in C^d$;
■ $\|f_j\| \geq c_1 \sqrt{d_n} n^{-\kappa}$, $j \in \mathcal{M}_\star$, $\kappa < \frac{d}{2d+1}$.

Spline approximation: Let f_{nj} be the spline approximation of f_j.

Lemma 1: If $d_n^{-2d-1} \leq c_1(1-\xi)n^{-2\kappa}/C_1$ for some $\xi \in (0,1)$, then

$$\min_{j \in \mathcal{M}_\star} \|f_{nj}\|^2 \geq c_1 \xi d_n n^{-2\kappa}.$$

Theoretical basis: Minimum signal

True model: $\mathcal{M}_\star = \{j : Em_j(X_j)^2 > 0\}$ or $Y \perp \mathbf{X}_{\mathcal{M}_\star^c}$ given $\mathbf{X}_{\mathcal{M}_\star}$.

Assumption: ■ $f_j = E(Y|X_j) \in C^d$;
■ $\|f_j\| \geq c_1\sqrt{d_n}n^{-\kappa}$, $j \in \mathcal{M}_\star$, $\kappa < \frac{d}{2d+1}$.

Spline approximation: Let f_{nj} be the spline approximation of f_j.

Lemma 1: If $d_n^{-2d-1} \leq c_1(1-\xi)n^{-2\kappa}/C_1$ for some $\xi \in (0,1)$, then

$$\min_{j\in\mathcal{M}_\star}\|f_{nj}\|^2 \geq c_1\xi d_n n^{-2\kappa}.$$

Theorem 1: ▸ Conditions A–F ▸ Idea of the proof

(i) $P\left(\max_{1\le j\le p_n}\left|\|\hat{f}_{nj}\|_n^2 - \|f_{nj}\|^2\right| \ge c_2 d_n n^{-2\kappa}\right) \to 1$, exp fast.

(ii) If $\nu_n = c_5 d_n n^{-2\kappa}$,then $\mathbf{P}(\mathcal{M}_\star \subset \widehat{\mathcal{M}}_{\nu_n}) \to \mathbf{1}$ exponentially fast.

■No conditions needed on the covariance for the SS (Sure Screening) property!

■Can handle the NP-dimensionality:

$$\log p_n = o(n^{1-4\kappa} d_n^{-3} + n d_n^{-3}).$$

★(Variance) $d_n = o(n^{1/3})$. ★(Bias) $d_n \ge B_4 n^{2\kappa/(2d+1)}$.

Nonparametric Independence Screening

Theorem 1: Conditions A–F | Idea of the proof

(i) $P\left(\max_{1\le j\le p_n}\left|\|\hat{f}_{nj}\|_n^2 - \|f_{nj}\|^2\right| \ge c_2 d_n n^{-2\kappa}\right) \to 1$, exp fast.

(ii) If $\nu_n = c_5 d_n n^{-2\kappa}$,then $\mathbf{P}(\mathcal{M}_\star \subset \widehat{\mathcal{M}}_{\nu_n}) \to \mathbf{1}$ exponentially fast.

■No conditions needed on the covariance for the SS (Sure Screening) property!

■Can handle the NP-dimensionality:

$$\log p_n = o(n^{1-4\kappa} d_n^{-3} + n d_n^{-3}).$$

★(Variance) $d_n = o(n^{1/3})$. ★(Bias) $d_n \ge B_4 n^{2\kappa/(2d+1)}$.

Controlling False Selection Rate

Only sure screening: **not insightful**! e.g. select all variables.

Ideal Case: A gap between active variables and inactive variables:

$$\max_{j \notin \mathcal{M}_\star} \|f_{nj}\|^2 = o(d_n n^{-2\kappa}),$$

We have **model selection consistency**:

$$P(\widehat{\mathcal{M}}_{\nu_n} = \mathcal{M}_\star) = 1 - o(1).$$

Controlling False Selection Rate

Only sure screening: **not insightful**! e.g. select all variables.

Ideal Case: A gap between active variables and inactive variables:

$$\max_{j \notin \mathcal{M}_\star} \|f_{nj}\|^2 = o(d_n n^{-2\kappa}),$$

We have **model selection consistency**:

$$P(\widehat{\mathcal{M}}_{\nu_n} = \mathcal{M}_\star) = 1 - o(1).$$

Controlling False Selection Rate

Theorem 2: Conditions A–F | Idea of the proof

For any $\nu_n = c_5 d_n n^{-2\kappa}$,

$$\mathbf{P}[|\widehat{\mathcal{M}}_{\nu_n}| \le \mathbf{O}\{\mathbf{n}^{2\kappa}\lambda_{\max}(\Sigma)\}] \to \mathbf{1},$$

where $\Sigma = E\Psi\Psi^T$ and $\Psi = (\Psi_1, \cdots, \Psi_{p_n})^T$.

$$\textbf{False Selection Rate} = 1 - \frac{s_n}{|\widehat{\mathcal{M}}_{\nu_n}|},$$

Having SS property, the smaller this upper bound, the better.

Controlling False Selection Rate

Theorem 2: Conditions A–F | Idea of the proof

For any $\nu_n = c_5 d_n n^{-2\kappa}$,

$$\mathbf{P}[|\widehat{\mathcal{M}}_{\nu_n}| \leq \mathbf{O}\{\mathbf{n}^{2\kappa}\lambda_{\max}(\Sigma)\}] \to \mathbf{1},$$

where $\Sigma = E\Psi\Psi^T$ and $\Psi = (\Psi_1, \cdots, \Psi_{p_n})^T$.

$$\textbf{False Selection Rate} = 1 - \frac{s_n}{|\widehat{\mathcal{M}}_{\nu_n}|},$$

Having SS property, the smaller this upper bound, the better.

Remarks on Theorem 2

- ★ **The number of selected variables** is related to the covariance matrix of **basis** via the operator norm.

- ★ When the covariates are independent, the matrix Σ is **block diagonal** with j-th block Σ_j. Then $\lambda_{\max}(\Sigma) = O(d_n^{-1})$.

- ★ When $\lambda_{\max}(\Sigma) = O(n^{\tau})$, the selected model size with the sure screening property is only of **polynomial order**.

Remarks on Theorem 2

- ★ **The number of selected variables** is related to the covariance matrix of **basis** via the operator norm.
- ★ When the covariates are independent, the matrix Σ is **block diagonal** with j-th block Σ_j. Then $\lambda_{\max}(\Sigma) = O(d_n^{-1})$.
- ★ When $\lambda_{\max}(\Sigma) = O(n^{\tau})$, the selected model size with the sure screening property is only of **polynomial order**.

Simulation Studies

Simulation Settings and Comparison Criteria

- $p = 1000, n = 400$, 100 repititions for each setting.
- **MMS** required to have the Sure Screening property (Ex 1-2).
- **Measures**: **TP**, **FP** and **PE** for each method (Ex 3-6).
- **Design of experiments:**
  - Ex 1: consistency condition for LASSO fails.
  - Ex 2: marginal projection is **nonlinear**.
  - Ex 3-4: varying correlations.
  - Ex 5: hidden signature variable.
  - Ex 6: Varying d_n **and SNR** (Signal-to-Noise Ratio).

Simulation Settings and Comparison Criteria

- $p = 1000, n = 400$, 100 repititions for each setting.
- **MMS** required to have the Sure Screening property (Ex 1-2).
- **Measures**: **TP**, **FP** and **PE** for each method (Ex 3-6).
- **Design of experiments:**
  - Ex 1: consistency condition for LASSO fails.
  - Ex 2: marginal projection is **nonlinear**.
  - Ex 3-4: varying correlations.
  - Ex 5: hidden signature variable.
  - Ex 6: Varying d_n **and SNR** (Signal-to-Noise Ratio).

Simulation Settings and Comparison Criteria

- $p = 1000, n = 400$, 100 repititions for each setting.
- **MMS** required to have the Sure Screening property (Ex 1-2).
- **Measures**: **TP**, **FP** and **PE** for each method (Ex 3-6).
- **Design of experiments:**
  - Ex 1: consistency condition for LASSO fails.
  - Ex 2: marginal projection is **nonlinear**.
  - Ex 3-4: varying correlations.
  - Ex 5: hidden signature variable.
  - Ex 6: Varying d_n **and SNR** (Signal-to-Noise Ratio).

Comparison of Minimum Model Size

Ex 1: Fan and Song (09): $Y = \beta^{*T}\mathbf{X} + \varepsilon$, $\varepsilon \sim N(0,3)$,
$\beta^* = (1, -1, \cdots)^T$. $\{X_k\}_{k=1}^{p-50} \sim_{i.i.d.} N(0,1)$, and

$$X_k = \sum_{j=1}^{s} X_j(-1)^{j+1}/5 + \sqrt{1 - \frac{s}{25}}\varepsilon_k, \quad \{\varepsilon_k\} \sim_{i.i.d} N(0,1). \quad k \geq p-49$$

Model	NIS	PenGAM	SIS
Ex 1 ($s = 3$)	3(0)	3(0)	3(0)
Ex 1 ($s = 6$)	56(0)	1000(0)	56(0)
Ex 1 ($s = 12$)	66(7)	1000(0)	62(1)
Ex 1 ($s = 24$)	269(134)	1000(0)	109(43)

- When $s > 5$, consistency fails for LASSO;
- Price for NIS is small.

Comparison of Minimum Model Size

Ex 1: Fan and Song (09): $Y = \beta^{*T}\mathbf{X} + \varepsilon$, $\varepsilon \sim N(0,3)$, $\beta^* = (1,-1,\cdots)^T$. $\{X_k\}_{k=1}^{p-50} \sim_{i.i.d.} N(0,1)$, and

$$X_k = \sum_{j=1}^{s} X_j(-1)^{j+1}/5 + \sqrt{1-\frac{s}{25}}\varepsilon_k, \quad \{\varepsilon_k\} \sim_{i.i.d} N(0,1). \quad k \geq p-49$$

Model	NIS	PenGAM	SIS
Ex 1 ($s=3$)	3(0)	3(0)	3(0)
Ex 1 ($s=6$)	56(0)	**1000**(0)	56(0)
Ex 1 ($s=12$)	66(7)	**1000**(0)	62(1)
Ex 1 ($s=24$)	269(134)	**1000**(0)	109(43)

- When $s > 5$, consistency fails for LASSO;
- Price for NIS is small.

Needs of nonlinear methods

Ex 2: $Y = X_1 + X_2 + X_3 + \varepsilon$, $\varepsilon \sim N(0,3)$.
$\{X_k\}_{k\neq 2}$ are i.i.d N(0,1) and $X_2 = -\frac{1}{3}X_1^3 + \tilde{\varepsilon}$, $\tilde{\varepsilon} \sim N(0,1)$.

Model	NIS	PenGAM	SIS
Ex 2	3(0)	3(0)	360(361)

- $E(Y|X_1)$ and $E(Y|X_2)$ are nonlinear.
- Linear method fails.

Two nonparametric models

Notation: $f_1(x) = x, \quad f_2(x) = (2x-1)^2, \quad f_3(x) = \frac{\sin(2\pi x)}{2-\sin(2\pi x)}$

$$f_4(x) = 0.1\sin(2\pi x) + 0.2\cos(2\pi x) + 0.3\sin(2\pi x)^2 + 0.4\cos(2\pi x)^3 + 0.5\sin(2\pi x)^3.$$

Covariates: $X_j \sim N(0,1)$ with equi-correlation ρ.

Ex 3: *(Meier et al., 09)* $Y = 5f_1(X_1) + 3f_2(X_2) + 4f_3(X_3) + 6f_4(X_4) + \sqrt{1.74}\varepsilon$

Ex 4: (Meier et al., 09)

$$Y = f_1(X_1) + f_2(X_2) + f_3(X_3) + f_4(X_4) + 1.5f_1(X_5) + 1.5f_2(X_6) + 1.5f_3(X_7) + 1.5f_4(X_8) + 2f_1(X_9) + 2f_2(X_{10}) + 2f_3(X_{11}) + 2f_4(X_{12}) + \sqrt{0.52}\varepsilon.$$

Two nonparametric models

Notation: $f_1(x) = x, \quad f_2(x) = (2x-1)^2, \quad f_3(x) = \frac{\sin(2\pi x)}{2-\sin(2\pi x)}$

$$\begin{aligned} f_4(x) &= 0.1\sin(2\pi x) + 0.2\cos(2\pi x) + 0.3\sin(2\pi x)^2 \\ &\quad + 0.4\cos(2\pi x)^3 + 0.5\sin(2\pi x)^3. \end{aligned}$$

Covariates: $X_j \sim N(0,1)$ with equi-correlation ρ.

Ex 3: (*Meier et al., 09*) $Y = 5f_1(X_1) + 3f_2(X_2) + 4f_3(X_3) + 6f_4(X_4) + \sqrt{1.74}\varepsilon$

Ex 4: (Meier et al., 09)

$$\begin{aligned} Y &= f_1(X_1) + f_2(X_2) + f_3(X_3) + f_4(X_4) + 1.5f_1(X_5) + 1.5f_2(X_6) + 1.5f_3(X_7) \\ &\quad + 1.5f_4(X_8) + 2f_1(X_9) + 2f_2(X_{10}) + 2f_3(X_{11}) + 2f_4(X_{12}) + \sqrt{0.52}\varepsilon. \end{aligned}$$

Comparison of Model Selection and Estimation

Model	Method	TP	FP	PE	Time
Ex 3 $\rho = 0$	INIS	**4.00**(0.00)	**2.23**(2.99)	**2.99**(0.31)	23.33(7.29)
	g-INIS	**4.00**(0.00)	**0.67**(0.75)	**2.92**(0.30)	33.06(7.21)
	penGAM	**4.00**(0.00)	**29.43**(18.28)	**3.30**(0.42)	**236.32**(5.18)
	ISIS	**3.03**(0.00)	**29.97**(0.00)	**15.92**(1.60)	15.98(4.65)
Ex 3 $\rho = .5$	INIS	3.99(0.00)	16.36(5.97)	2.99(0.39)	82.91(30.92)
	g-INIS	4.00(0.00)	0.98(1.49)	2.61(0.26)	35.90(13.10)
	penGAM	4.00(0.00)	40.33(26.12)	2.97(0.29)	282.98(23.77)
	ISIS	3.01(0.00)	29.99(0.00)	12.91(1.31)	19.36(5.29)
Ex 4 $\rho = 0$	INIS	11.98(0.00)	3.56(2.24)	0.97(0.13)	72.03(26.25)
	g-INIS	12.00(0.00)	0.73(0.75)	0.91(0.10)	155.61(21.61)
	penGAM	11.98(0.00)	81.44(23.51)	1.27(0.14)	281.88(15.04)
	ISIS	7.95(0.75)	25.05(0.75)	4.69(0.38)	17.56(4.92)
Ex 4 $\rho = .5$	INIS	10.03(1.49)	15.46(1.49)	**1.04**(0.16)	139.41(49.16)
	g-INIS	10.78(0.75)	1.08(1.49)	**0.87**(0.11)	137.91(44.06)
	penGAM	10.57(0.75)	65.56(24.63)	**1.12**(0.13)	357.39(43.21)
	ISIS	6.59(0.75)	26.41(0.75)	**4.28**(0.48)	20.39(4.94)

Hidden Signature Variables

Ex 5: (Fan et al., 09) $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$, where $\varepsilon \sim N(0,1)$. $\beta_1 = \beta_2 = \beta_3 = 4$ and $\beta_4 = -\mathbf{6\sqrt{2}}$.
Hidden signature variable: X_4, indep $\mathbf{X}^T\beta^\star$ and Y.

Method	TP	FP	PE	Time
INIS	3.99(0.00)	21.84(0.00)	1.61(0.19)	106.56(35.16)
g-INIS	4.00(0.00)	1.02(1.49)	1.18(0.13)	42.82(9.88)
penGAM	3.03(0.00)	212.46(8.96)	3.11(0.40)	2802.11(764.97)
ISIS	4.00(0.00)	27.23(0.00)	1.23(0.14)	21.46(7.07)

Ex 5: (Fan et al., 09) $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$, where $\varepsilon \sim N(0,1)$. $\beta_1 = \beta_2 = \beta_3 = 4$ and $\beta_4 = -6\sqrt{2}$.
Hidden signature variable: X_4, indep $\mathbf{X}^T \beta^\star$ and Y.

Method	TP	FP	PE	Time
INIS	3.99(0.00)	21.84(0.00)	1.61(0.19)	106.56(35.16)
g-INIS	4.00(0.00)	1.02(1.49)	1.18(0.13)	42.82(9.88)
penGAM	3.03(0.00)	212.46(8.96)	3.11(0.40)	2802.11(764.97)
ISIS	4.00(0.00)	27.23(0.00)	1.23(0.14)	21.46(7.07)

Varying number of spline terms d_n

Ex 6: $Y = 3g_1(X_1) + 3g_2(X_2) + 2g_3(X_3) + 2g_4(X_4) + C\sqrt{3.38}\varepsilon$.
$C^2 = 2, 1, 0.5, 0.25$ such that $SNR = 0.5, 1, 2, 4$.

Table: **SNR=0.5**,$\rho = 0$

d_n	Method	TP	FP	PE	Time
2	INIS	3.96(0.00)	2.28(1.49)	7.74(0.79)	16.09(5.32
	penGAM	4.00(0.00)	27.85(16.98)	8.07(0.92)	354.46(31.48
4	INIS	3.93(0.00)	2.29(1.68)	7.90(0.81)	21.68(8.95
	penGAM	3.99(0.00)	25.61(13.62)	8.21(0.84)	421.17(35.71
8	INIS	3.81(0.00)	2.59(2.24)	8.16(1.08)	33.10(15.79
	penGAM	3.95(0.00)	34.59(20.34)	8.49(0.82)	484.17(179.70
16	INIS	**3.38**(0.75)	**2.02**(1.49)	**8.60**(1.13)	42.69(20.13
	penGAM	**3.74**(0.00)	**33.48**(23.88)	**9.04**(0.93)	685.97(267.43

Varying spline terms d_n

Example 6:

Table: $\rho = 0$, **SNR=4.0**

d_n	Method	TP	FP	PE	Time
2	INIS	4.00(0.00)	2.06(2.24)	1.19(0.13)	17.74(6.42
	penGAM	4.00(0.00)	28.57(14.37)	1.27(0.15)	213.43(12.09
4	INIS	4.00(0.00)	2.33(1.49)	1.09(0.10)	23.28(9.37
	penGAM	4.00(0.00)	30.75(17.35)	1.18(0.14)	300.69(12.21
8	INIS	4.00(0.00)	2.88(2.24)	1.02(0.12)	39.21(19.17
	penGAM	4.00(0.00)	40.51(17.54)	1.14(0.14)	340.06(11.49
16	INIS	4.00(0.00)	**1.72**(1.49)	**1.10**(0.12)	49.79(25.78
	penGAM	4.00(0.00)	**45.77**(19.03)	**1.33**(0.16)	481.19(141.51

Example 6:

Table: $\rho = 0$, $d_n = 2$

SNR	Method	TP	FP	PE	Time
0.5	INIS	3.96(0.00)	2.28(1.49)	7.74(0.79)	16.09(5.32)
	penGAM	4.00(0.00)	27.85(16.98)	8.07(0.92)	354.46(31.48)
1.0	INIS	4.00(0.00)	2.16(2.24)	3.98(0.34)	16.03(5.74)
	penGAM	4.00(0.00)	26.51(14.18)	4.20(0.46)	284.85(20.30)
2.0	INIS	4.00(0.00)	2.03(2.24)	2.12(0.17)	15.92(5.42)
	penGAM	4.00(0.00)	25.89(13.06)	2.25(0.24)	235.69(13.32)
4.0	INIS	4.00(0.00)	**2.06**(2.24)	**1.19**(0.13)	17.74(6.42)
	penGAM	4.00(0.00)	**28.57**(14.37)	**1.27**(0.15)	213.43(12.09)

Example 6:

Table: $\rho = 0$, $d_n = 16$

SNR	Method	TP	FP	PE	Time
0.5	INIS	3.38(0.75)	2.02(1.49)	8.60(1.13)	42.69(20.13)
	penGAM	3.74(0.00)	33.48(23.88)	9.04(0.93)	685.97(267.4)
1.0	INIS	4.00(0.00)	1.80(1.49)	4.26(0.45)	46.81(21.47)
	penGAM	4.00(0.00)	38.60(19.78)	4.80(0.57)	595.87(197.1)
2.0	INIS	4.00(0.00)	1.77(1.49)	2.17(0.25)	48.40(24.65)
	penGAM	4.00(0.00)	42.58(16.60)	2.54(0.30)	540.89(165.4)
4.0	INIS	4.00(0.00)	**1.72**(1.49)	**1.10**(0.12)	49.79(25.78)
	penGAM	4.00(0.00)	**45.77**(19.03)	**1.33**(0.16)	481.19(141.5)

An analysis of Affymetrix GeneChip Rat Array

Affymetrix GeneChip Rat Genome 230 2.0 Array

Purpose: Find the genes related to gene TRIM32, which causes Bardet-Biedl syndrome.

- $n = 120$ male rats (tissue from eyes), and $p = 18975$.
- Following Huang et al. (09), focus on 2000 probe sets, highly correlated w/ TRIM32.
- Three methods: INIS-penGAM, INIS-penGAM (p=2000) and penGAM (p=2000).
- Divide data into training (100) and testing (20).
- Repeat the experiment 100 times to test the stability.

Affymetrix GeneChip Rat Genome 230 2.0 Array

Purpose: Find the genes related to gene TRIM32, which causes Bardet-Biedl syndrome.

- $n = 120$ male rats (tissue from eyes), and $p = 18975$.
- Following Huang et al. (09), focus on 2000 probe sets, highly correlated w/ TRIM32.
- Three methods: INIS-penGAM, INIS-penGAM (p=2000) and penGAM (p=2000).
- Divide data into training (100) and testing (20).
- Repeat the experiment 100 times to test the stability.

Estimation Results

- 8 probes by **INIS-penGAM** with RSS 0.24:
  1371755_at, **1372928_at**, **1373534_at**, **1373944_at**, **1374669_at**, **1376686_at**, **1376747_at**, **1377880_at**.
- 8 probes by **INIS-penGAM** (p=2000) with RSS 0.26:
  1376686_at, **1376747_at**, 1378590_at, **1373534_at**, **1377880_at**, **1372928_at**, **1374669_at**, **1373944_at**.
- 32 probes by penGAM with RSS 0.1.

Repeation: 100 times

Method	Model Size	PE
INIS	7.73(0.00)	0.47(0.13)
INIS ($p = 2000$)	7.68(0.75)	0.44(0.15)
penGAM ($p = 2000$)	26.71(14.93)	0.48(0.16)

Estimation Results

- 8 probes by **INIS-penGAM** with RSS 0.24:
  1371755_at, **1372928_at**, **1373534_at**, **1373944_at**, **1374669_at**, **1376686_at**, **1376747_at**, **1377880_at**.

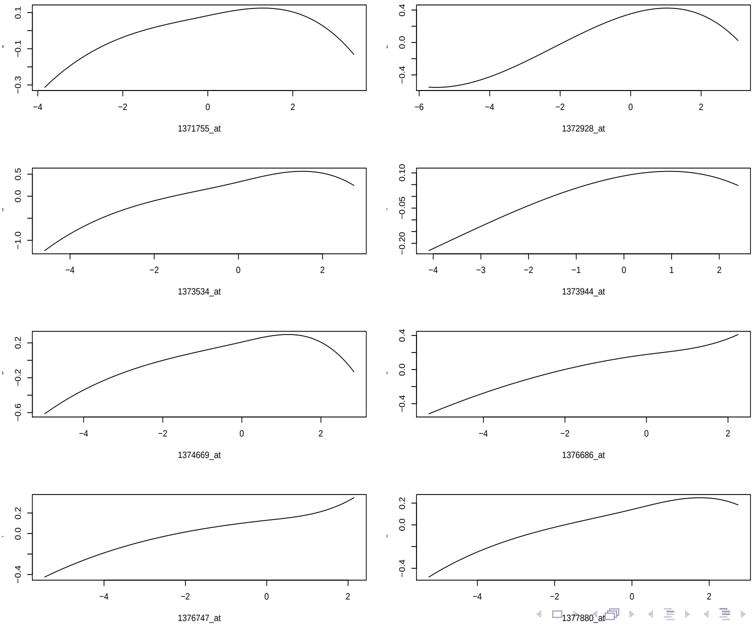
- 8 probes by **INIS-penGAM** (p=2000) with RSS 0.26:
  1376686_at, **1376747_at**, 1378590_at, **1373534_at**, **1377880_at**, **1372928_at**, **1374669_at**, **1373944_at**.

- 32 probes by penGAM with RSS 0.1.

Repeation: 100 times

Method	Model Size	PE
INIS	7.73(0.00)	0.47(0.13)
INIS ($p = 2000$)	7.68(0.75)	0.44(0.15)
penGAM ($p = 2000$)	26.71(14.93)	0.48(0.16)

Estimated functions by INIS-penGAM



Summary

1. Proposed NIS, allowing **nonlinear** relations and **non-Gaussian** covariates.
2. Proposed INIS, discovering hidden signature variables and conditional correlations.
3. Established SIS with controlled model size.
4. Numerical experiment show favorable results, fast computation, and utilities.

Summary

1. Proposed NIS, allowing **nonlinear** relations and **non-Gaussian** covariates.
2. Proposed INIS, discovering hidden signature variables and conditional correlations.
3. Established SIS with controlled model size.
4. Numerical experiment show favorable results, fast computation, and utilities.

Summary

1. Proposed NIS, allowing **nonlinear** relations and **non-Gaussian** covariates.
2. Proposed INIS, discovering hidden signature variables and conditional correlations.
3. Established SIS with controlled model size.
4. Numerical experiment show favorable results, fast computation, and utilities.

Summary

1. Proposed NIS, allowing **nonlinear** relations and **non-Gaussian** covariates.
2. Proposed INIS, discovering hidden signature variables and conditional correlations.
3. Established SIS with controlled model size.
4. Numerical experiment show favorable results, fast computation, and utilities.

Acknowledgement

Thank  *You*

Assumptions

A. The marginal projections $\{f_j\}_{j=1}^{p}$ belong to

$$\mathcal{F} = \left\{ f(\cdot) : \left| f^{(r)}(s) - f^{(r)}(t) \right| \le K|s-t|^{\alpha}, \text{ for } s,t \in [a,b] \right\}.$$

B. The marginal density g_j satisfies $K_1 \le g_j(X_j) \le K_2$.

C. $\min_{j \in \mathcal{M}_\star} E\{f_j(X_j)^2\} \ge c_1 d_n n^{-2\kappa}$.

D. $\|m\|_\infty < B_1$.

E. $\{\varepsilon_i\}_{i=1}^{n}$ are i.i.d. with $E[\exp(B_2|\varepsilon_i|)|\mathbf{X}_i] < B_3$.

F. $d_n^{-2d-1} \le c_1(1-\xi)n^{-2\kappa}/C_1$.

Back to Theorem 1 Back to Theorem 2

Three Facts

Under conditions A and B, when $l \geq d$, we have

1. There exists a positive constant C_1 such that (Stone, 85)

$$\|f_j - f_{nj}\|^2 \leq C_1 d_n^{-2d}.$$

2. There exists a positive constant C_2 such that (Stone 85, Huang et.al. 09)

$$E\Psi_{jk}^2(X_{ij}) \leq C_2 d_n^{-1}.$$

3. There exist some positive constants D_1 and D_2 such that (Zhou et al. 98)

$$D_1 d_n^{-1} \leq \lambda_{\min}(E\Psi_j\Psi_j^T) \leq \lambda_{\max}(E\Psi_j\Psi_j^T) \leq D_2 d_n^{-1}.$$

Recall that

$$\|\hat{f}_{nj}\|_n^2 = (\mathbb{P}_n\Psi_j Y)^T(\mathbb{P}_n\Psi_j\Psi_j^T)^{-1}\mathbb{P}_n\Psi_j Y,$$

and

$$\|f_{nj}\|^2 = (E\Psi_j Y)^T(E\Psi_j\Psi_j^T)^{-1}E\Psi_j Y.$$

Then we can decompose the difference into three parts and bound them separately.

- Difference between $\mathbb{P}_n\Psi_j Y$ and $E\Psi_j Y$
- Difference between $\mathbb{P}_n\Psi_j\Psi_j^T$ and $E\Psi_j\Psi_j^T$

Key tool: Bernstein inequalities

Back to Theorem 1

Main idea in the proof of Theorem 1 (Cont')

Lemma 1 Under Conditions A, B and D, for any $\delta > 0$, there exist some positive constants c_6 and c_7 such that

$$P(|(\mathbb{P}_n - E)\Psi_{jk} Y| \geq \delta n^{-1}) \leq 4\exp(-\delta^2/2(c_6 n d_n^{-1} + c_7\delta)),$$

for $k = 1, \cdots, d_n$, $j = 1, \cdots, p$.

Lemma 2 Under Conditions A and B, for any $\delta > 0$,

$$P(|\lambda_{\min}(\mathbb{P}_n \Psi_j \Psi_j^T) - \lambda_{\min}(E\Psi_j \Psi_j^T)| \geq d_n \delta / n)$$
$$\leq \quad 2d_n^2 \exp\left\{-\frac{1}{2}\frac{\delta^2}{C_2 n d_n^{-1} + \delta/3}\right\}.$$

Back to Theorem 1

Main idea in the proof of Theorem 2

The key idea of the proof is to show:

$$\|E\Psi Y\|^2 = O(\lambda_{\max}(\Sigma)).$$

If so, by definition and $\|\Psi_{jk}\|_\infty \le 1$, we have

$$\sum_{j=1}^{p_n} \|f_{nj}\|^2 \le \max_{1\le j\le p_n} \lambda_{\max}\{(E\Psi_j\Psi_j^T)^{-1}\}\|E\Psi Y\|^2 = O(d_n\lambda_{\max}(\Sigma)).$$

Then the number of $\{j : \|f_{nj}\|^2 > \varepsilon d_n n^{-2\kappa}\}$ can not exceed $O(n^{2\kappa}\lambda_{\max}(\Sigma))$ for any $\varepsilon > 0$.

Back to Theorem 2