Nonparametric Independence Screening

Yang Feng

Columbia University

With Prof. Jianqing Fan and Rui Song

2011 Winter Workshop on High Dimensional Inference, UFL
Outline

1. Introduction
2. Nonparametric Independence Screening (NIS)
3. INIS Algorithms
4. Sure Screening Properties
5. Numerical Studies
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

UFL Nonparametric Independence Screening
Impact of Dimensionality

- Computational cost
- Stability
- Estimation accuracy: ★ noise accumulation ★ spurious corr

**Key Idea:** Large-scale screening + moderate-scale searching.
Sure independence screening: By using correlation ranking
\[ r_i = |\text{corr}(X_i, Y)| \] (Fan and Lv, 2008),

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

★ Limitations: ■ Linear models. ■ Joint normality.

\[ Y = \sum_{j \in M_*} \beta_j X_j + \epsilon \]
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:

\[ r_i = \left| \text{corr}(X_i, X_i^2, \ldots, X_i^k, Y) \right| \]

(Hall and Miller, 09)

Other methods: Data-tilting; (Hall, Titterington & Xue, 09);
Marginal LR (Fan, Samworth & Wu, 09); MPLE (Zhao & Li, 09);
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:
\[ r_i = |\text{corr}((X_i, X_i^2, \ldots, X_i^k), Y)| \] (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?

   \[ Y = \sum_{j \in \mathcal{M}_*} m_j(x_j) + \varepsilon. \]

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?

   \[ Y = \sum_{j \in M_*} m_j(x_j) + \varepsilon. \]

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
Additive model: (Stone, 1985)

\[ Y = \sum_{j \in \mathcal{M}_*} m_j(X_j) + \varepsilon \]

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

B-spline basis: \( \psi_j = (\psi_1(X_j), \cdots, \psi_{dn}(X_j))^T \)

Marginal regressions: with \( f_{nj}(x) = \sum_{k=1}^{dn} \beta_{jk} \psi_{jk}(x) \)

\[ \min_{f_{nj} \in S_n} \mathbb{P}_n \left( Y - f_{nj}(X_j) \right)^2 = \min_{\beta_j \in \mathbb{R}^{dn}} \mathbb{P}_n \left( Y - \psi_j^T \beta_j \right)^2, \]

Challenges: ■ growing \( d_n; \) ■ Non-Gaussian random matrices
Additive model: (Stone, 1985)

\[ Y = \sum_{j \in \mathcal{M}_*} m_j(X_j) + \varepsilon \]

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

**B-spline basis:** \( \psi_j \equiv \psi_j(X_j) = (\psi_1(X_j), \ldots, \psi_{d_n}(X_j))^T \)

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

\[
\min_{f_{nj} \in S_n} \mathbb{P}_n \left( Y - f_{nj}(X_j) \right)^2 = \min_{\beta_j \in \mathbb{R}^{d_n}} \mathbb{P}_n \left( Y - \psi_j^T \beta_j \right)^2,
\]

**Challenges:** □ growing \( d_n \); □ Non-Gaussian random matrices
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 \( f_{nj}(x) = \sum_{k=1}^{d_n} \beta_{jk} \psi_{jk}(x) \)

\[
\min_{f_{nj} \in \mathcal{S}_n} \mathbb{P}_n \left( Y - f_{nj}(X_j) \right)^2 = \min_{\beta_j \in \mathbb{R}^{dn}} \mathbb{P}_n \left( Y - \psi_j^T \beta_j \right)^2,
\]

**Challenges:** □ growing \( d_n \); □ Non-Gaussian random matrices
Select variables according to the marginal utilities.

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

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

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.

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

- **Marginal RSS**: $\widetilde{M}_{\gamma} = \{j : u_j \leq \gamma\}$, 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 $u_j = \mathbb{P}_n(Y^2 - \hat{f}_{nj}^2)$.
Principled SIS Zhao and Li (09) proposed using upper $\alpha$ (Control of FDR) quantile of marginal utilities for decoupled response and covariate (PSIS).

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

Remark: We can take $a^*_n$ based on one permutation.
**Principled SIS** Zhao and Li (09) proposed using upper $\alpha$ (Control of FDR) quantile of marginal utilities for decoupled response and covariate (PSIS).

- Obtain the decoupled synthetic data $\{(X_{\pi(i)}, Y_i)\}_{i=1}^n$ —Marginal distributions are untouched;
- Compute $a_n^* = \max_j \| \hat{f}_{nj}^* \|_2$;
- Choose the top $\alpha$-quantile of $a_n^*$ as $v_n$.

**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 $Y$?

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

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

\[ Y = X_1 + 0.2 X_{100} + \varepsilon \]
Potential Drawbacks (Fan & Lv, 2008)

♦ False Negative: What if $X_4$ marginally uncorrelated with $Y$, but jointly correlated with $Y$?

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

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

$$Y = X_1 + 0.2X_{100} + \varepsilon$$
INIS Algorithm
Large-scale screening: Apply NIS to pick a set $\mathcal{A}_1$.

Moderate-scale selection: Employ a penalized method such as penGAM (Meier et. al, 09) or SpAM (Ravikumar et al. (09)) to select a subset $\mathcal{M}_1$.
Large-scale conditional screening: Rank features according to additional conditional contribution:

\[ L_j = \min_{\{f_{ni}, i \in M_1\}, f_{nj}} \mathbb{P}_n \left( Y - \sum_{i \in M_1} f_{ni}(X_i) - f_{nj}(X_j) \right)^2. \]

Pick a set \( A_2 \) according to \( \{L_j, j \in M_1^c\} \).
Moderate-scale selection: As in Step 2, among variables in $M_1 \cup A_2$, use a penalized method to select a set $M_2$ —Allow deletion.

We iterate Steps 3-4 until convergence.

Greedy-INIS: Restricting size of set $A_j$ to be 1.
- Extremely fast to compute.
- Very effective when covariates are highly correlated.
Moderate-scale selection: As in Step 2, among variables in $M_1 \cup A_2$, use a penalized method to select a set $M_2$—Allow deletion.

We iterate Steps 3-4 until convergence.

Greedy-INIS: Restricting size of set $A_j$ to be 1.
- Extremely fast to compute.
- Very effective when covariates are highly correlated.
Moderate-scale selection: As in Step 2, among variables in $M_1 \cup A_2$, use a penalized method to select a set $M_2$—Allow deletion.

We iterate Steps 3-4 until convergence.

Greedy-INIS: Restricting size of set $A_j$ to be 1.
- Extremely fast to compute.
- Very effective when covariates are highly correlated.
Sure Screening Properties
Theoretical basis: Minimum signal

**True model**: \( \mathcal{M}_* = \{ j : E m_j(X_j)^2 > 0 \} \) or \( Y \perp X_{\mathcal{M}_*^c} \) given \( X_{\mathcal{M}_*} \).

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

**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
\[
\min_{j \in \mathcal{M}_*} \|f_{nj}\|^2 \geq c_1 \xi d_n n^{-2\kappa}.
\]
Theoretical basis: Minimum signal

True model: \( \mathcal{M}_* = \{ j : E m_j(X_j)^2 > 0 \} \) or \( Y \perp X_{\mathcal{M}_*^c} \) given \( X_{\mathcal{M}_*} \).

Assumption: \( \mathbf{f}_j = E(Y|X_j) \in \mathbb{C}^d; \)
\[ \| f_j \| \geq c_1 \sqrt{d_n n^{-\kappa}}, j \in \mathcal{M}_*, \kappa < \frac{d}{2d+1}. \]

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
\[ \min_{j \in \mathcal{M}_*} \| f_{nj} \|^2 \geq c_1 \xi d_n n^{-2\kappa}. \]
**Theorem 1:**

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

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

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

- Can handle the NP-dimensionality:

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

★ (Variance) \( d_n = o(n^{1/3}) \).

★ (Bias) \( d_n \geq B_4 n^{2\kappa}/(2d+1) \).
Theorem 1:

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

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

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

- Can handle the NP-dimensionality:

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

★ (Variance) $d_n = o(n^{1/3})$. ★ (Bias) $d_n \geq B_4 n^{2\kappa}/(2d+1)$. 
Only sure screening: **not insightful**! e.g. select all variables.

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

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

We have **model selection consistency**:

\[
P(\hat{\mathcal{M}}_{vn} = \mathcal{M}_*) = 1 - o(1).
\]
Controlling False Selection Rate

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

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

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

We have **model selection consistency**:

\[
P(\hat{\mathcal{M}}_{\nu_n} = \mathcal{M}_*) = 1 - o(1).
\]
Theorem 2: Conditions A–F

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

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

where $\Sigma = E\Psi\Psi^T$ and $\Psi = (\psi_1, \cdots, \psi_{p_n})^T$.

False Selection Rate $= 1 - \frac{S_n}{|\hat{M}_{\nu_n}|}$

Having SS property, the smaller this upper bound, the better.
Theorem 2: For any $\nu_n = c_5 d_n n^{-2\kappa}$,

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

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

False Selection Rate $= 1 - \frac{S_n}{|\hat{M}_{\nu_n}|}$.

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 $\Sigma$ is **block diagonal** with $j$-th block $\Sigma_j$. Then $\lambda_{\text{max}}(\Sigma) = O(d_n^{-1})$.

★ When $\lambda_{\text{max}}(\Sigma) = O(n^\tau)$, 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 $\Sigma$ is **block diagonal** with $j$-th block $\Sigma_j$. Then $\lambda_{\text{max}}(\Sigma) = O(d_n^{-1})$.

★ When $\lambda_{\text{max}}(\Sigma) = O(n^\tau)$, 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 repetitions 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 repetitions 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 repetitions 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):** \( Y = \beta^* X + \varepsilon, \varepsilon \sim N(0,3), \)
\( \beta^* = (1, -1, \cdots)^T \mathcal{X}_{k}^{p-50} \sim i.i.d. N(0,1), \) and

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

<table>
<thead>
<tr>
<th>Model</th>
<th>NIS</th>
<th>PenGAM</th>
<th>SIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex 1 (s = 3)</td>
<td>3(0)</td>
<td>3(0)</td>
<td>3(0)</td>
</tr>
<tr>
<td>Ex 1 (s = 6)</td>
<td>56(0)</td>
<td>1000(0)</td>
<td>56(0)</td>
</tr>
<tr>
<td>Ex 1 (s = 12)</td>
<td>66(7)</td>
<td>1000(0)</td>
<td>62(1)</td>
</tr>
<tr>
<td>Ex 1 (s = 24)</td>
<td>269(134)</td>
<td>1000(0)</td>
<td>109(43)</td>
</tr>
</tbody>
</table>

- When \( s > 5, \) consistency fails for LASSO;
- Price for NIS is small.
**Comparison of Minimum Model Size**

**Ex 1: Fan and Song (09):**  
\[ Y = \beta^*^T X + \epsilon, \epsilon \sim N(0, 3), \]
\[ \beta^* = (1, -1, \cdots)^T \{X_k\}_{k=1}^{p-50} \sim i.i.d. N(0, 1), \text{ and} \]
\[ X_k = \sum_{j=1}^{s} X_j (-1)^{j+1} / 5 + \sqrt{1 - \frac{s}{25}} \epsilon_k, \quad \{\epsilon_k\} \sim i.i.d N(0, 1). \quad k \geq p - 49 \]

<table>
<thead>
<tr>
<th>Model</th>
<th>NIS</th>
<th>PenGAM</th>
<th>SIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex 1 ((s = 3))</td>
<td>3(0)</td>
<td>3(0)</td>
<td>3(0)</td>
</tr>
<tr>
<td>Ex 1 ((s = 6))</td>
<td>56(0)</td>
<td><strong>1000</strong>(0)</td>
<td>56(0)</td>
</tr>
<tr>
<td>Ex 1 ((s = 12))</td>
<td>66(7)</td>
<td><strong>1000</strong>(0)</td>
<td>62(1)</td>
</tr>
<tr>
<td>Ex 1 ((s = 24))</td>
<td>269(134)</td>
<td><strong>1000</strong>(0)</td>
<td>109(43)</td>
</tr>
</tbody>
</table>

- 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 \( X_2 = -\frac{1}{3} X_1^3 + \tilde{\varepsilon}, \tilde{\varepsilon} \sim N(0,1). \)

<table>
<thead>
<tr>
<th>Model</th>
<th>NIS</th>
<th>PenGAM</th>
<th>SIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex 2</td>
<td>3(0)</td>
<td>3(0)</td>
<td>360(361)</td>
</tr>
</tbody>
</table>

\( E(Y|X_1) \) 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 \( \rho \).

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

**Ex 4:** (Meier et al., 09)

\[
Y = f_1(X_1) + f_2(X_2) + f_3(X_3) + f_4(X_4) + 1.5f_1(X_5) + 1.5f_2(X_6) + 1.5f_3(X_7) + 1.5f_4(X_8) + 2f_1(X_9) + 2f_2(X_{10}) + 2f_3(X_{11}) + 2f_4(X_{12}) + \sqrt{0.52} \epsilon.
\]
Two nonparametric models

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

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

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

**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)

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

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

UFL Nonparametric Independence Screening
**Hidden Signature Variables**

**Ex 5**: (Fan et al., 09) \( Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon, \) where \( \epsilon \sim N(0, 1). \) \( \beta_1 = \beta_2 = \beta_3 = 4 \) and \( \beta_4 = -6\sqrt{2}. \)

**Hidden signature variable**: \( X_4, \) indep \( X^T \beta^* \) and \( Y. \)

<table>
<thead>
<tr>
<th>Method</th>
<th>TP</th>
<th>FP</th>
<th>PE</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>INIS</td>
<td>3.99</td>
<td>21.84</td>
<td>1.61</td>
<td>106.56(35.16)</td>
</tr>
<tr>
<td>g-INIS</td>
<td>4.00</td>
<td>1.02</td>
<td>1.18</td>
<td>42.82(9.88)</td>
</tr>
<tr>
<td>penGAM</td>
<td>3.03</td>
<td>212.46</td>
<td>3.11</td>
<td>2802.11(764.97)</td>
</tr>
<tr>
<td>ISIS</td>
<td>4.00</td>
<td>27.23</td>
<td>1.23</td>
<td>21.46(7.07)</td>
</tr>
</tbody>
</table>

Nonparametric Independence Screening
**Ex 5**: (Fan et al., 09) \( Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \), where \( \epsilon \sim N(0, 1) \). \( \beta_1 = \beta_2 = \beta_3 = 4 \) and \( \beta_4 = -6 \sqrt{2} \).

**Hidden signature variable**: \( X_4 \), indep \( X^T \beta^* \) and \( Y \).

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

UFL  Nonparametric Independence Screening
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: $SNR=0.5, \rho = 0$
### Example 6:

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

<table>
<thead>
<tr>
<th>$d_n$</th>
<th>Method</th>
<th>TP</th>
<th>FP</th>
<th>PE</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>INIS</td>
<td>4.00(0.00)</td>
<td>2.06(2.24)</td>
<td>1.19(0.13)</td>
<td>17.74(6.42)</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>4.00(0.00)</td>
<td>28.57(14.37)</td>
<td>1.27(0.15)</td>
<td>213.43(12.09)</td>
</tr>
<tr>
<td>4</td>
<td>INIS</td>
<td>4.00(0.00)</td>
<td>2.33(1.49)</td>
<td>1.09(0.10)</td>
<td>23.28(9.37)</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>4.00(0.00)</td>
<td>30.75(17.35)</td>
<td>1.18(0.14)</td>
<td>300.69(12.21)</td>
</tr>
<tr>
<td>8</td>
<td>INIS</td>
<td>4.00(0.00)</td>
<td>2.88(2.24)</td>
<td>1.02(0.12)</td>
<td>39.21(19.17)</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>4.00(0.00)</td>
<td>40.51(17.54)</td>
<td>1.14(0.14)</td>
<td>340.06(11.49)</td>
</tr>
<tr>
<td>16</td>
<td>INIS</td>
<td>4.00(0.00)</td>
<td>1.72(1.49)</td>
<td>1.10(0.12)</td>
<td>49.79(25.78)</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>4.00(0.00)</td>
<td>45.77(19.03)</td>
<td>1.33(0.16)</td>
<td>481.19(141.51)</td>
</tr>
</tbody>
</table>
Example 6:

Table: \( \rho = 0, d_n = 2 \)

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

Varying Signal-to-noise Ratios (\( d_n = 2 \))

Nonparametric Independence Screening
### Example 6:

Table: \( \rho = 0, \; d_n = 16 \)

<table>
<thead>
<tr>
<th>SNR</th>
<th>Method</th>
<th>TP</th>
<th>FP</th>
<th>PE</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>INIS</td>
<td>3.38</td>
<td>2.02</td>
<td>8.60</td>
<td>42.69</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>3.74</td>
<td>33.48</td>
<td>9.04</td>
<td>685.97</td>
</tr>
<tr>
<td>1.0</td>
<td>INIS</td>
<td>4.00</td>
<td>1.80</td>
<td>4.26</td>
<td>46.81</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>4.00</td>
<td>38.60</td>
<td>4.80</td>
<td>595.87</td>
</tr>
<tr>
<td>2.0</td>
<td>INIS</td>
<td>4.00</td>
<td>1.77</td>
<td>2.17</td>
<td>48.40</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>4.00</td>
<td>42.58</td>
<td>2.54</td>
<td>540.89</td>
</tr>
<tr>
<td>4.0</td>
<td>INIS</td>
<td>4.00</td>
<td>1.72</td>
<td>1.10</td>
<td>49.79</td>
</tr>
<tr>
<td></td>
<td>penGAM</td>
<td>4.00</td>
<td>45.77</td>
<td>1.33</td>
<td>481.19</td>
</tr>
</tbody>
</table>

UFL Nonparametric Independence Screening
An analysis of Affymetrix GeneChip Rat 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.
**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.

**Repetition**: 100 times

<table>
<thead>
<tr>
<th>Method</th>
<th>Model Size</th>
<th>PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INIS</td>
<td>7.73(0.00)</td>
<td>0.47(0.13)</td>
</tr>
<tr>
<td>INIS (p = 2000)</td>
<td>7.68(0.75)</td>
<td>0.44(0.15)</td>
</tr>
<tr>
<td>penGAM (p = 2000)</td>
<td>26.71(14.93)</td>
<td>0.48(0.16)</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Method</th>
<th>Model Size</th>
<th>PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INIS</td>
<td>7.73(0.00)</td>
<td>0.47(0.13)</td>
</tr>
<tr>
<td>INIS (p = 2000)</td>
<td>7.68(0.75)</td>
<td>0.44(0.15)</td>
</tr>
<tr>
<td>penGAM (p = 2000)</td>
<td>26.71(14.93)</td>
<td>0.48(0.16)</td>
</tr>
</tbody>
</table>

UFL Nonparametric Independence Screening
Estimated functions by INIS-penGAM

UFL Nonparametric Independence Screening
Summary

- Proposed NIS, allowing **nonlinear** relations and **non-Gaussian** covariates.
- Proposed INIS, discovering hidden signature variables and conditional correlations.
- Established SIS with controlled model size.
- 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.

   Numerical experiment show favorable results, fast computation, and utilities.
Proposed NIS, allowing **nonlinear** relations and **non-Gaussian** covariates.

Proposed INIS, discovering hidden signature variables and conditional correlations.

Established SIS with controlled model size.

Numerical experiment show favorable results, fast computation, and utilities.
Thank You
Assumptions

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

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

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

C. \( \min_{j \in M^*} E \{ f_j(X_j)^2 \} \geq c_1 d_n n^{-2\kappa} \).

D. \( \| m \|_\infty < B_1 \).

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

F. \( d_n^{-2d-1} \leq c_1 (1 - \xi) n^{-2\kappa} / C_1 \).
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)
   \[
   D_1 d_n^{-1} \leq \lambda_{\min}(E \Psi_j \Psi_j^T) \leq \lambda_{\max}(E \Psi_j \Psi_j^T) \leq D_2 d_n^{-1}.
   \]
Main idea in the proof of Theorem 1

Recall that

\[ \| \hat{f}_{nj} \|^2_n = (P_n \Psi_j Y)^T (P_n \Psi_j \Psi_j^T)^{-1} P_n \Psi_j Y, \]

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 $P_n \Psi_j Y$ and $E \Psi_j Y$
- Difference between $P_n \Psi_j \Psi_j^T$ and $E \Psi_j \Psi_j^T$

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

$$P\left(\left| (\mathbb{P}_n - E)\psi_{jk} Y \right| \geq \delta n^{-1} \right) \leq 4 \exp\left( -\frac{\delta^2}{2(c_6 nd_n^{-1} + c_7 \delta)} \right),$$

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

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

$$P\left( \left| \lambda_{\min}(\mathbb{P}_n \psi_j \psi_j^T) - \lambda_{\min}(E \psi_j \psi_j^T) \right| \geq d_n \delta / n \right) \leq 2d_n^2 \exp\left\{ -\frac{1}{2} \frac{\delta^2}{C_2 nd_n^{-1} + \delta / 3} \right\}.$$
The key idea of the proof is to show:

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

If so, by definition and \( \| \Psi_{jk} \|_{\infty} \leq 1 \), we have

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

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