Towards Completely-Data-Driven Functional Estimation

Xiaoming Huo

School of Industrial & Systems Engineering
Georgia Institute of Technology, Atlanta

Banff International Research Station
Tuesday, December 13, 2011

Jointly work with Z. Yang & H. Xie & Y. Lu
Outline

1. Background
2. Data Driven Method
3. Theoretical Consideration: Rate of Convergence
4. Asymptotic Optimality of the Generalized Cross Validation
5. Fast Computation
6. Simulations
Outline

1 Background

2 Data Driven Method

3 Theoretical Consideration: Rate of Convergence

4 Asymptotic Optimality of the Generalized Cross Validation

5 Fast Computation

6 Simulations
Formulation

- Observe \((X_i, y_i), X_i \in \mathbb{R}^d, p \geq 1, y_i \in \mathbb{R}, i = 1, 2, \ldots, n\)

- Relation

  \[ y_i = f(X_i) + \varepsilon_i, \]

  or

  \[ y_i = f(X_{i1}, X_{i2}, \ldots, X_{id}) + \varepsilon_i, \]

  where \(\varepsilon_i\)'s are i.i.d. and satisfy some conditions

- Objective: estimating \(f\)

- Desideratum: \(f \in \mathcal{F}\)
Formulation

- Observe \((X_i, y_i), X_i \in \mathbb{R}^d, p \geq 1, y_i \in \mathbb{R}, i = 1, 2, \ldots, n\)
- Relation
  \[ y_i = f(X_i) + \varepsilon_i, \]
  or
  \[ y_i = f(X_{i1}, X_{i2}, \ldots, X_{id}) + \varepsilon_i, \]
  where \(\varepsilon_i\)'s are i.i.d. and satisfy some conditions
- Objective: estimating \(f\)
- Desideratum: \(f \in \mathcal{F}\)

xiaoming@isye.gatech.edu (Georgia Tech)

Estimation on Irregular Domains

Birs 4 / 38
Observe \((X_i, y_i), X_i \in \mathbb{R}^d, p \geq 1, y_i \in \mathbb{R}, i = 1, 2, \ldots, n\)

Relation

\[ y_i = f(X_i) + \varepsilon_i, \]

or

\[ y_i = f(X_{i1}, X_{i2}, \ldots, X_{id}) + \varepsilon_i, \]

where \(\varepsilon_i\)'s are i.i.d. and satisfy some conditions

Objective: estimating \(f\)

Desideratum: \(f \in \mathcal{F}\)
Observe $(X_i, y_i), X_i \in \mathbb{R}^d, p \geq 1, y_i \in \mathbb{R}, i = 1, 2, \ldots, n$

Relation

$$y_i = f(X_i) + \varepsilon_i,$$

or

$$y_i = f(X_{i1}, X_{i2}, \ldots, X_{id}) + \varepsilon_i,$$

where $\varepsilon_i$'s are i.i.d. and satisfy some conditions

Objective: estimating $f$

Desideratum: $f \in \mathcal{F}$
Existing Methods

- Linear regression, $\mathcal{F} = \text{a finite-dimensional linear subspace}$
- Kernel
- Splines
- Wavelets
- ...
- $\mathcal{F}$ is a Sobolev space

$$\mathcal{F} = H^m(\Omega) = \left\{ f : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{Z}_+^d, |\alpha| \leq m \right\}$$

where $\Omega \in \mathbb{R}^d$ is the domain of the function, for $\alpha \in \mathbb{Z}_+^d$, $\alpha = (\alpha_1, \alpha_2, ..., \alpha_d)$ and $|\alpha| = \sum_{i=1}^d \alpha_i$, we define the partial derivative

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}$$
Existing Methods

- Linear regression, $\mathcal{F} = \text{a finite-dimensional linear subspace}$
- Kernel
  - Splines
  - Wavelets
  - ...
- $\mathcal{F}$ is a Sobolev space

$$
\mathcal{F} = H^m(\Omega) = \{ f : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{Z}_+^d, |\alpha| \leq m \}
$$

where $\Omega \in \mathbb{R}^d$ is the domain of the function, for $\alpha \in \mathbb{Z}_+^d$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ and $|\alpha| = \sum_{i=1}^{d} \alpha_i$, we define the partial derivative

$$
D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}
$$
Existing Methods

- Linear regression, $\mathcal{F} = a$ finite-dimensional linear subspace
- Kernel
- Splines
- Wavelets
- ...
- $\mathcal{F}$ is a Sobolev space

$$\mathcal{F} = H^m(\Omega) = \left\{ f : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{Z}^d_+, |\alpha| \leq m \right\}$$

where $\Omega \in \mathbb{R}^d$ is the domain of the function, for $\alpha \in \mathbb{Z}^d_+$, 
$\alpha = (\alpha_1, \alpha_2, ..., \alpha_d)$ and $|\alpha| = \sum_{i=1}^d \alpha_i$, we define the partial derivative

$$D^\alpha f = \frac{\partial |\alpha| f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}$$
Existing Methods

- Linear regression, $\mathcal{F} = \text{a finite-dimensional linear subspace}$
- Kernel
- Splines
- Wavelets
- ...
- $\mathcal{F}$ is a Sobolev space

$$\mathcal{F} = H^m(\Omega) = \left\{ f : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{Z}_+^d, |\alpha| \leq m \right\}$$

where $\Omega \subseteq \mathbb{R}^d$ is the domain of the function, for $\alpha \in \mathbb{Z}_+^d$, $\alpha = (\alpha_1, \alpha_2, ..., \alpha_d)$ and $|\alpha| = \sum_{i=1}^d \alpha_i$, we define the partial derivative

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}$$
Existing Methods

- Linear regression, $\mathcal{F} = \text{a finite-dimensional linear subspace}$
- Kernel
- Splines
- Wavelets
- ...

- $\mathcal{F}$ is a Sobolev space

$$\mathcal{F} = H^m(\Omega) = \left\{ f : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{Z}_+^d, |\alpha| \leq m \right\}$$

where $\Omega \in \mathbb{R}^d$ is the domain of the function, for $\alpha \in \mathbb{Z}_+^d$, $\alpha = (\alpha_1, \alpha_2, ..., \alpha_d)$ and $|\alpha| = \sum_{i=1}^d \alpha_i$, we define the partial derivative

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}$$
Existing Methods

- Linear regression, $\mathcal{F} = \text{a finite-dimensional linear subspace}$
- Kernel
- Splines
- Wavelets
- ...
- $\mathcal{F}$ is a Sobolev space

$$\mathcal{F} = H^m(\Omega) = \left\{ f : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{Z}_+^d, |\alpha| \leq m \right\}$$

where $\Omega \in \mathbb{R}^d$ is the domain of the function, for $\alpha \in \mathbb{Z}_+^d$, $\alpha = (\alpha_1, \alpha_2, ..., \alpha_d)$ and $|\alpha| = \sum_{i=1}^d \alpha_i$, we define the partial derivative

$$D^\alpha f = \frac{\partial |\alpha| f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}$$
Penalization estimation approach (equivalently, Lagrange multiplier):

\[
\min_{f \in \mathcal{F}} G(f) + \lambda \cdot R(f)
\]

where

- **G(f)**: goodness of fit, e.g., \( G(f) = \sum_{i=1}^{n} [y_i - f(X_i)]^2 \)
- **R(f)**: regularity of \( f \), e.g., \( R(f) = \int_{\Omega} \sum_{i,j=1}^{p} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 \)
- **\( \Omega \)**: domain of \( f \), e.g., \( \Omega = \mathcal{R}^d \)
Penalization estimation approach (equivalently, Lagrange multiplier):

$$\min_{f \in \mathcal{F}} G(f) + \lambda \cdot R(f)$$

where

- $G(f)$: goodness of fit, e.g., $G(f) = \sum_{i=1}^{n} [y_i - f(X_i)]^2$
- $R(f)$: regularity of $f$, e.g., $R(f) = \int_{\Omega} \sum_{i,j=1}^{p} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2$
- $\Omega$: domain of $f$, e.g., $\Omega = \mathbb{R}^d$
Penalization estimation approach (equivalently, Lagrange multiplier):

\[
\min_{f \in \mathcal{F}} G(f) + \lambda \cdot R(f)
\]

where

- \( G(f) \): goodness of fit, e.g., \( G(f) = \sum_{i=1}^{n} [y_i - f(X_i)]^2 \)
- \( R(f) \): regularity of \( f \), e.g., \( R(f) = \int_{\Omega} \sum_{i,j=1}^{p} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 \)

\( \Omega \): domain of \( f \), e.g., \( \Omega = \mathbb{R}^d \)
Penalization estimation approach (equivalently, Lagrange multiplier):

$$\min_{f \in \mathcal{F}} G(f) + \lambda \cdot R(f)$$

where

- $G(f)$: goodness of fit, e.g., $G(f) = \sum_{i=1}^{n} [y_i - f(X_i)]^2$
- $R(f)$: regularity of $f$, e.g., $R(f) = \int_{\Omega} \sum_{i,j=1}^{p} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2$
- $\Omega$: domain of $f$, e.g., $\Omega = \mathbb{R}^d$
More on the Existing Approach

- Define a regularity functional. Recall

\[ R(f) = \int_\Omega \| H f \|_F^2 \]

- Essence: Need to show that solving problem

\[
\min_{f \in \mathcal{F}} R(f)
\]

subject to \( f(X_i) = y'_i \)

ends up with a quadratic form:

\[ R(f) = f^T M f \]

where \( f = (f(X_1), f(X_2), \ldots, f(X_n))^T \)

- If \( \Omega \) is irregular, determining analytically the gram matrix \( M \) can be very difficult.
More on the Existing Approach

- Define a regularity functional. Recall
  \[ R(f) = \int_{\Omega} \|Hf\|_F^2 \]

- Essence: Need to show that solving problem
  \[
  \min_{f \in \mathcal{F}} R(f) \\
  \text{subject to } f(X_i) = y'_i
  \]

ends up with a quadratic form:

\[ R(f) = f^T M f \]

where \( f = (f(X_1), f(X_2), \ldots, f(X_n))^T \)

- If \( \Omega \) is irregular, determining analytically the gram matrix \( M \) can be very difficult.
More on the Existing Approach

- Define a regularity functional. Recall

\[ R(f) = \int_{\Omega} \|Hf\|^2_F \]

- Essence: Need to show that solving problem

\[
\min_{f \in \mathcal{F}} R(f)
\]

subject to \( f(X_i) = y_i' \)

ends up with a quadratic form:

\[ R(f) = f^T M f \]

where \( f = (f(X_1), f(X_2), \ldots, f(X_n))^T \)

- If \( \Omega \) is irregular, determining analytically the gram matrix \( M \) can be very difficult.
Example of Irregular Domain

(a) Horseshoe.  

(b) Letter R.
Outline

1. Background
2. Data Driven Method
3. Theoretical Consideration: Rate of Convergence
4. Asymptotic Optimality of the Generalized Cross Validation
5. Fast Computation
6. Simulations
An unbiased alternative:

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} [y_i - f(X_i)]^2 + \lambda \sum_{i=1}^{n} \| \mathcal{H} f(X_i) \|^2_F.$$ 

Let $V_i, 1 \leq i \leq k$, denote the $k$ nearest neighbors of $V_0$. Let $\bar{V} = \frac{1}{k+1} \sum_{i=0}^{k} V_i$, i.e., $\bar{V}$ is the average. Taylor expansion:

$$f(V_i) \approx f(\bar{V}) + (V_i - \bar{V})^T \mathcal{J} f(\bar{V}) + \frac{1}{2} (V_i - \bar{V})^T \mathcal{H} f(\bar{V}) (V_i - \bar{V}),$$

$i = 0, 1, \cdots, n$,

where $f(\bar{V})$ is the functional value, $\mathcal{J} f(\bar{V})$ is the Jacobian, and $\mathcal{H} f(\bar{V})$ is the hessian matrix. Note we have $\mathcal{J} f(\bar{V}) \in \mathbb{R}^d$ and $\mathcal{H} f(\bar{V}) \in \mathbb{R}^{d \times d}$. 
An unbiased alternative:

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} [y_i - f(X_i)]^2 + \lambda \sum_{i=1}^{n} \| \mathcal{H} f(X_i) \|^2_F.$$ 

Let \( V_i, 1 \leq i \leq k \), denote the \( k \) nearest neighbors of \( V_0 \). Let \( \bar{V} = \frac{1}{k+1} \sum_{i=0}^{k} V_i \), i.e., \( \bar{V} \) is the average. Taylor expansion:

$$f(V_i) \approx f(\bar{V}) + (V_i - \bar{V})^T \mathcal{J} f(\bar{V}) + \frac{1}{2} (V_i - \bar{V})^T \mathcal{H} f(\bar{V}) (V_i - \bar{V}),$$

\( i = 0, 1, \cdots, n \),

where \( f(\bar{V}) \) is the functional value, \( \mathcal{J} f(\bar{V}) \) is the Jacobian, and \( \mathcal{H} f(\bar{V}) \) is the hessian matrix. Note we have \( \mathcal{J} f(\bar{V}) \in \mathbb{R}^d \) and \( \mathcal{H} f(\bar{V}) \in \mathbb{R}^{d \times d} \).
Rewrite as a linear system:
\[ f^* \approx 1_{k+1} \cdot c + V \cdot J + \frac{1}{2} C \cdot H, \]

A partial implementation of QR-decomposition
\[
\begin{bmatrix}
1_{k+1} & V & \frac{1}{2} C
\end{bmatrix} =
\begin{bmatrix}
Q_1 & Q_2
\end{bmatrix}
\begin{bmatrix}
R_{11} & R_{12} \\
0 & I_{(d^2+d)/2}
\end{bmatrix},
\]

where columns of \( Q_1 \in \mathbb{R}^{(k+1) \times (d+1)} \) are orthonormal
\((Q_1^T Q_1 = I_{d+1})\), and columns of \( Q_2 \in \mathbb{R}^{(k+1) \times \frac{d^2+d}{2}} \) are orthogonal to
the columns of \( Q_1 \) (i.e., \( Q_2^T Q_1 = 0 \)).
Rewrite as a linear system:

\[ f^* \approx \mathbf{1}_{k+1} \cdot c + \mathbf{V} \cdot \mathbf{J} + \frac{1}{2} \mathbf{C} \cdot \mathbf{H}, \]

A partial implementation of QR-decomposition

\[
\begin{bmatrix}
\mathbf{1}_{k+1} & \mathbf{V} & \frac{1}{2} \mathbf{C}
\end{bmatrix} =
\begin{bmatrix}
\mathbf{Q}_1 & \mathbf{Q}_2
\end{bmatrix}
\begin{bmatrix}
\mathbf{R}_{11} & \mathbf{R}_{12} \\
\mathbf{0} & \mathbf{I}_{(d^2+d)/2}
\end{bmatrix},
\]

where columns of \( \mathbf{Q}_1 \in \mathbb{R}^{(k+1) \times (d+1)} \) are orthonormal \( (\mathbf{Q}_1^T \mathbf{Q}_1 = \mathbf{I}_{d+1}) \), and columns of \( \mathbf{Q}_2 \in \mathbb{R}^{(k+1) \times \frac{d^2+d}{2}} \) are orthogonal to the columns of \( \mathbf{Q}_1 \) (i.e., \( \mathbf{Q}_2^T \mathbf{Q}_1 = \mathbf{0} \)).
More on Derivation

- we have

\[
Q_2^T f^* = \begin{pmatrix} 0 & Q_2^T Q_2 \end{pmatrix} \left[ \begin{array}{cc} R_{11} & R_{12} \\ 0 & I_{(p^2+p)/2} \end{array} \right] \begin{pmatrix} \begin{bmatrix} c \\ J \\ H \end{bmatrix} \end{pmatrix} = Q_2^T Q_2 H.
\]

- least-squares estimator of $H$

\[
\hat{H} = (Q_2^T Q_2)^+ Q_2^T f^*,
\]

where $(\cdot)^+$ denotes a pseudo-inverse of a matrix.

- Frobenius norm of the hessian matrix at a point:

\[
\|\hat{H} f(X_i)\|_F^2 = \|\hat{H}\|_2^2 = \hat{H}^T \hat{H} \\
= (f^*)^T Q_2 (Q_2^T Q_2)^+ (Q_2^T Q_2)^+ Q_2^T f^*.
\]
we have

\[ Q_2^T f^* = \begin{pmatrix} 0 & Q_2^T Q_2 \end{pmatrix} \begin{bmatrix} R_{11} & R_{12} \\ 0 & I_{(p^2+p)/2} \end{bmatrix} \begin{bmatrix} c \\ J \\ H \end{bmatrix} = Q_2^T Q_2 H. \]

least-squares estimator of \( H \)

\[ \hat{H} = (Q_2^T Q_2)^+ Q_2^T f^*, \]

where \((\cdot)^+\) denotes a pseudo-inverse of a matrix.

Frobenius norm of the hessian matrix at a point:

\[ \| \hat{H} f(X_i) \|_F^2 = \| \hat{H} \|_2^2 = \hat{H}^T \hat{H} = (f^*)^T Q_2 (Q_2^T Q_2)^+ (Q_2^T Q_2)^+ Q_2^T f^*. \]
we have

\[ Q_2^T f^* = (Q_2^T Q_2) \begin{bmatrix} R_{11} & R_{12} \\ 0 & I_{(p^2+p)/2} \end{bmatrix} \begin{bmatrix} c \\ J \\ H \end{bmatrix} = Q_2^T Q_2 H. \]

least-squares estimator of \( H \)

\[ \hat{H} = (Q_2^T Q_2)^+ Q_2^T f^*, \]

where \((\cdot)^+\) denotes a pseudo-inverse of a matrix.

Frobenius norm of the hessian matrix at a point:

\[ \| \hat{H} f(X_i) \|_F^2 = \| \hat{H} \|_2^2 = \hat{H}^T \hat{H} = (f^*)^T Q_2 (Q_2^T Q_2)^+ (Q_2^T Q_2)^+ Q_2^T f^*. \]
A Quadratic Form

- Denote

\[ K_i = Q_2(Q_2^T Q_2)^+(Q_2^T Q_2)^+ Q_2^T. \]

- We have

\[ \sum_{i=1}^{n} \|\hat{H}f(X_i)\|_F^2 = \sum_{i=1}^{n} (f^T S_i^T K_i S_i f). \]

- Let \( M = (S_1^T, \cdots, S_n^T)\text{diag}\{K_1, K_2, \cdots, K_n\} \begin{pmatrix} S_1 \\ \vdots \\ S_n \end{pmatrix}, \) we have

\[ \sum_{i=1}^{n} \|\hat{H}f(X_i)\|_F^2 = f^T M f, \quad (1) \]

which is a quadratic function of \( f. \)
A Quadratic Form

- Denote
  \[ K_i = Q_2(Q_2^T Q_2)^+ (Q_2^T Q_2)^+ Q_2^T. \]

- We have
  \[ \sum_{i=1}^{n} \| \hat{H} f(X_i) \|_F^2 = \sum_{i=1}^{n} (f^T S_i^T K_i S_i f). \]

- Let \( M = (S_1^T, \cdots, S_n^T) \text{diag}\{K_1, K_2, \cdots, K_n\} \begin{pmatrix} S_1 \\ \vdots \\ S_n \end{pmatrix} \), we have
  \[ \sum_{i=1}^{n} \| \hat{H} f(X_i) \|_F^2 = f^T M f, \quad (1) \]

  which is a quadratic function of \( f \).
A Quadratic Form

Denote

\[ K_i = Q_2(Q_2^T Q_2)^+ (Q_2^T Q_2)^+ Q_2^T. \]

We have

\[ \sum_{i=1}^{n} \| \hat{H}f(X_i) \|_F^2 = \sum_{i=1}^{n} (f^T S_i^T K_i S_i f). \]

Let \( M = (S_1^T, \cdots, S_n^T) \text{diag}\{K_1, K_2, \cdots, K_n\} \) \( \begin{pmatrix} S_1 \\ \vdots \\ S_n \end{pmatrix} \), we have

\[ \sum_{i=1}^{n} \| \hat{H}f(X_i) \|_F^2 = f^T M f, \] (1)

which is a quadratic function of \( f \).
Problem becomes

$$\min_{f} \| Y - f \|^2_2 + \lambda f^T M f,$$

Close form solution:

$$\hat{f} = (I_n + \lambda \cdot M)^{-1} \cdot Y.$$

Close Form Solution

Problem becomes

$$\min_{f} \| Y - f \|_2^2 + \lambda f^T M f,$$

Close form solution:

$$\hat{f} = (I_n + \lambda \cdot M)^{-1} \cdot Y.$$
Problem becomes
\[
\min_f \|Y - f\|_2^2 + \lambda f^T M f,
\]
Close form solution:
\[
\hat{f} = (I_n + \lambda \cdot M)^{-1} \cdot Y.
\]

Outline

1. Background
2. Data Driven Method
3. Theoretical Consideration: Rate of Convergence
4. Asymptotic Optimality of the Generalized Cross Validation
5. Fast Computation
6. Simulations
Rate of convergence: How fast does $\frac{1}{n} \| \hat{f}_n - f \|_2^2$ go to zero?

Stone (1982) showed that $O(n^{-\frac{2m}{2m+d}})$ is the optimal rate of convergence for nonparametric regression for $d$-dimensional input, while the up to order $m$ partial derivatives of the underlying function $f$ are in $L^2(\Omega)$.

Note we have $m = 2$.

Note: we must have $2m > d$. 
Rate of Convergence

Rate of convergence: How fast does $\frac{1}{n} \| \hat{f}_n - f \|_2^2$ go to zero?

Stone (1982) showed that $O(n^{-\frac{2m}{2m+d}})$ is the optimal rate of convergence for nonparametric regression for $d$-dimensional input, while the up to order $m$ partial derivatives of the underlying function $f$ are in $L^2(\Omega)$.

Note we have $m = 2$.

Note: we must have $2m > d$. 
Rate of Convergence

- Rate of convergence: How fast does $\frac{1}{n} \| \hat{f}_n - f \|_2^2$ go to zero?

- Stone (1982) showed that $O(n^{-\frac{2m}{2m+d}})$ is the optimal rate of convergence for nonparametric regression for $d$-dimensional input, while the up to order $m$ partial derivatives of the underlying function $f$ are in $L^2(\Omega)$.

- Note we have $m = 2$.

  - Note: we must have $2m > d$. 

Rate of convergence: How fast does $\frac{1}{n} ||\hat{f}_n - f||_2^2$ go to zero?

Stone (1982) showed that $O(n^{-\frac{2m}{2m+d}})$ is the optimal rate of convergence for nonparametric regression for $d$-dimensional input, while the up to order $m$ partial derivatives of the underlying function $f$ are in $L^2(\Omega)$.

- Note we have $m = 2$.
- Note: we must have $2m > d$. 
For $0 \leq \ell \leq m$, a semi-inner-product in $W^m_2(\Omega)$ is defined by

$$
\langle f, g \rangle_{\Omega, \ell} = \int_{\Omega} \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} (D^\alpha f)(D^\alpha g)dx,
$$

which gives rise to the related semi-norm

$$
|f|^2_{\Omega, \ell} = \int_{\Omega} \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} |D^\alpha f|^2 dx.
$$

With $T = \{X_i\}_{i=1}^n$, we can also give a discrete version of the aforementioned semi-norm as

$$
|f|^2_{T, \ell} = \frac{1}{n} \sum_{i=1}^n \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} |D^\alpha f(X_i)|^2.
$$

Specially, $|f|^2_{\Omega, 0} = \int_\Omega f(x)^2 dx$ and $|f|^2_{T, 0} = \frac{1}{n} \sum_{i=1}^n f(X_i)^2$. 
For $0 \leq \ell \leq m$, a semi-inner-product in $W^{m}_2(\Omega)$ is defined by

$$\langle f, g \rangle_{\Omega, \ell} = \int_{\Omega} \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} (D^\alpha f)(D^\alpha g) dx,$$

(2)

which gives rise to the related semi-norm

$$|f|^{2}_{\Omega, \ell} = \int_{\Omega} \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} |D^\alpha f|^2 dx.$$  

(3)

With $T = \{X_i\}_{i=1}^n$, we can also give a discrete version of the aforementioned semi-norm as

$$|f|^{2}_{T, \ell} = \frac{1}{n} \sum_{i=1}^{n} \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} |D^\alpha f(X_i)|^2.$$  

(4)

Specially, $|f|^{2}_{\Omega, 0} = \int_{\Omega} f(x)^2 dx$ and $|f|^{2}_{T, 0} = \frac{1}{n} \sum_{i=1}^{n} f(X_i)^2$. 

xiaoming@isye.gatech.edu (Georgia Tech)
For $0 \leq \ell \leq m$, a semi-inner-product in $\mathcal{W}_2^m(\Omega)$ is defined by

$$\langle f, g \rangle_{\Omega, \ell} = \int_{\Omega} \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} (D^\alpha f)(D^\alpha g) dx,$$

which gives rise to the related semi-norm

$$|f|_{\Omega, \ell}^2 = \int_{\Omega} \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} |D^\alpha f|^2 dx.$$  

With $T = \{X_i\}_{i=1}^n$, we can also give a discrete version of the aforementioned semi-norm as

$$|f|_{T, \ell}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{|\alpha| = \ell} \frac{\ell!}{\alpha!} |D^\alpha f(X_i)|^2.$$  

Specially, $|f|_{\Omega, 0}^2 = \int_{\Omega} f(x)^2 dx$ and $|f|_{T, 0}^2 = \frac{1}{n} \sum_{i=1}^n f(X_i)^2$. 

xiaoming@isye.gatech.edu (Georgia Tech)
Preparation: Ideal Quadratic form & Sampling Property

- **Ideal quadratic form.** For \( \ell = m \) in (4), we define \( E_{T,m} \) as the matrix representing the quadratic form

\[
|f|_{T,m}^2 = \frac{1}{n} f^T E_{T,m} f
\]  

where \( f = (f(X_1), \cdots, f(X_n))^T \) is the vector of function values at the knots of \( T = \{X_i\}_{i=1}^n \).

- **Sampling property.** For the set of sampling points \( T = \{X_i\}_{i=1}^n \) in domain \( \Omega \), we assume that there exists a constant \( B_0 > 0 \) such that

\[
\frac{\delta_{\text{max}}}{\delta_{\text{min}}} \leq B_0,
\]

where \( \delta_{\text{max}} = \sup_{X \in \Omega} \inf_{X_i \in T} \|X - X_i\| \), and \( \delta_{\text{min}} = \min_{j \neq i} \|X_j - X_i\| \).
**Ideal quadratic form.** For \( \ell = m \) in (4), we define \( E_{T,m} \) as the matrix representing the quadratic form

\[
\|f\|_{T,m}^2 = \frac{1}{n} f^T E_{T,m} f
\]

where \( f = (f(X_1), \cdots, f(X_n))^T \) is the vector of function values at the knots of \( T = \{X_i\}_{i=1}^n \).

**Sampling property.** For the set of sampling points \( T = \{X_i\}_{i=1}^n \) in domain \( \Omega \), we assume that there exists a constant \( B_0 > 0 \) such that

\[
\frac{\delta_{\text{max}}}{\delta_{\text{min}}} \leq B_0,
\]

where \( \delta_{\text{max}} = \sup_{X \in \Omega} \inf_{X_i \in T} \|X - X_i\| \), and \( \delta_{\text{min}} = \min_{j \neq i} \|X_j - X_i\|. \)
A **Lipschitz domain** (or domain with Lipschitz boundary) is a set in Euclidean space whose boundary is sufficiently regular in the sense that it can be thought of as locally being the graph of a Lipschitz continuous function.

Ω be an open set of \( \mathbb{R}^d \) satisfying a **uniform cone condition**: there exist a radius \( r > 0 \) and an angle \( \theta \in (0, \pi/2) \) such that for any \( X \in \Omega \) a unit vector \( \zeta(X) \in \mathbb{R}^d \) exists such that the cone

\[
C(X, \zeta(X), r, \theta) = \{ X + ts : s \in \mathbb{R}^d, \|s\| = 1, \zeta(X)^T s \geq \cos \theta, 0 \leq t \leq r \}
\]

is entirely contained in \( \Omega \).

\( U_2^m(\Omega) = \{ f \in W_2^m(\Omega) \mid B|f|_{\Omega, m}^2 \leq |f|_{T, m}^2 \leq \overline{B}|f|_{\Omega, m}^2 \} \) be a class of functions with bilaterally bounded constraint on their \( m \)th-order derivatives, where the constants \( B, \overline{B} > 0 \) do not depend on functions \( f \).
A **Lipschitz domain** (or domain with Lipschitz boundary) is a set in Euclidean space whose boundary is sufficiently regular in the sense that it can be thought of as locally being the graph of a Lipschitz continuous function.

\( \Omega \) be an open set of \( \mathbb{R}^d \) satisfying a **uniform cone condition**: there exist a radius \( r > 0 \) and an angle \( \theta \in (0, \pi/2) \) such that for any \( X \in \Omega \) a unit vector \( \zeta(X) \in \mathbb{R}^d \) exists such that the cone

\[
C(X, \zeta(X), r, \theta) = \{ X + ts : s \in \mathbb{R}^d, \|s\| = 1, \zeta(X)^T s \geq \cos \theta, 0 \leq t \leq r \}
\]  

is entirely contained in \( \Omega \).

\( U^m_2(\Omega) = \{ f \in W^m_2(\Omega) \mid B|f|^2_{\Omega,m} \leq |f|^2_{T,m} \leq \overline{B}|f|^2_{\Omega,m} \} \) be a class of functions with bilaterally bounded constraint on their \( m \)th-order derivatives, where the constants \( B, \overline{B} > 0 \) do not depend on functions \( f \).
A **Lipschitz domain** (or domain with Lipschitz boundary) is a set in Euclidean space whose boundary is sufficiently regular in the sense that it can be thought of as locally being the graph of a Lipschitz continuous function.

\( \Omega \) be an open set of \( \mathbb{R}^d \) satisfying a **uniform cone condition**: there exist a radius \( r > 0 \) and an angle \( \theta \in (0, \pi/2) \) such that for any \( X \in \Omega \) a unit vector \( \zeta(X) \in \mathbb{R}^d \) exists such that the cone

\[
C(X, \zeta(X), r, \theta) = \{ X + ts : s \in \mathbb{R}^d, ||s|| = 1, \zeta(X)^T s \geq \cos \theta, 0 \leq t \leq r \}
\]

(7) is entirely contained in \( \Omega \).

\[
U_m^2(\Omega) = \{ f \in W_2^m(\Omega) | B|f|_{\Omega,m}^2 \leq |f|_{T,m}^2 \leq \overline{B}|f|_{\Omega,m}^2 \}
\]

be a class of functions with bilaterally bounded constraint on their \( m \)th-order derivatives, where the constants \( B, \overline{B} > 0 \) do not depend on functions \( f \).
A few theorems, Step 1

- **Bound the eigenvalues of the ideal quadratic form.** Let $e_1 \leq \cdots \leq e_n$ be the eigenvalues of $E_{T,m}$ in ascending order.

**Theorem**

Let $\Omega$ be an open bounded Lipschitz domain satisfying the uniform cone condition, and the sample points $\{X_j\}_{j=1}^n$ fulfill the assumption of (6). Then there exist constants $C_3, C_4 > 0$ such that

$$C_3\rho_j \leq e_j \leq C_4\rho_j,$$

where $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_n$ are the first $n$ eigenvalues of the variational eigenvalue problem

$$\langle \phi, \psi \rangle_{\Omega,m} = \rho \langle \phi, \psi \rangle_{\Omega,0}, \quad \forall \psi \in W^m_2(\Omega).$$
A few theorems, Step 1

- **Bound the eigenvalues of the ideal quadratic form.** Let \( e_1 \leq \cdots \leq e_n \) be the eigenvalues of \( E_{T,m} \) in ascending order.

**Theorem**

Let \( \Omega \) be an open bounded Lipschitz domain satisfying the uniform cone condition, and the sample points \( \{X_j\}_{j=1}^{n} \) fulfill the assumption of (6). Then there exist constants \( C_3, C_4 > 0 \) such that

\[
C_3 \rho_j \leq e_j \leq C_4 \rho_j,
\]

where \( \rho_1 \leq \rho_2 \leq \cdots \leq \rho_n \) are the first \( n \) eigenvalues of the variational eigenvalue problem

\[
\langle \phi, \psi \rangle_{\Omega,m} = \rho \langle \phi, \psi \rangle_{\Omega,0}, \quad \forall \psi \in W_2^m(\Omega).
\]
Step 2: Quote a Known Result

Using a known rate from functional analysis

Theorem

Let $\Omega$ be an open bounded Lipschitz domain satisfying the uniform cone condition, and \( \{e_1 \leq \cdots \leq e_n\} \) the eigenvalues of $E_{T,m}$ in ascending order. Then there exist constants $C_5, C_6 > 0$ such that for $m(d) = \frac{(d+m-1)!}{d!(m-1)!} < j \leq n$ we have

\[
C_5 j^{\frac{2m}{d}} \leq e_j \leq C_6 j^{\frac{2m}{d}}. \tag{8}
\]
Using a known rate from functional analysis

**Theorem**

Let $\Omega$ be an open bounded Lipschitz domain satisfying the uniform cone condition, and $\{e_1 \leq \cdots \leq e_n\}$ the eigenvalues of $E_{T,m}$ in ascending order. Then there exist constants $C_5, C_6 > 0$ such that for $m(d) = \frac{(d+m-1)!}{d!(m-1)!} < j \leq n$ we have

$$C_5 j^{\frac{2m}{d}} \leq e_j \leq C_6 j^{\frac{2m}{d}}. \quad (8)$$
Step 3: Rates on Eigenvalues

- Property of eigenvalues.

**Theorem**

Let \( \mu_1 \leq \cdots \leq \mu_n \) be the eigenvalues of the matrix \( \mathbf{M} \) in (1). There exist constants \( C_7, C_8 > 0 \) such that for \( m^{(d)} < j \leq n \) we have

\[
C_7 j \frac{2m}{d} \leq \mu_j \leq C_8 j \frac{2m}{d}.
\]
Step 3: Rates on Eigenvalues

- Property of eigenvalues.

Theorem

Let $\mu_1 \leq \cdots \leq \mu_n$ be the eigenvalues of the matrix $\mathbf{M}$ in (1). There exist constants $C_7, C_8 > 0$ such that for $m^{(d)} < j \leq n$ we have

$$C_7 j \frac{2m}{d} \leq \mu_j \leq C_8 j \frac{2m}{d}.$$
Step 4: Rate of Convergence

Asymptotic rate of convergence

Theorem

Let $\hat{f}_n(\lambda) = A_n(\lambda)y = (I_n + \lambda M)^{-1}y$ be the CDS estimator from the multivariate model with the order $m > d/2$ and denote $r_n(\lambda) = n^{-1}\|\hat{f}_n(\lambda) - f\|^2$. If $n \to \infty$ and $\lambda \sim n^{-2m/(2m+d)}$ is chosen, then

$$E[r_n(\lambda)] = O(n^{-\frac{2m}{2m+d}}).$$

The above matches the optimal rate in Stone (1982).
Step 4: Rate of Convergence

- Asymptotic rate of convergence

**Theorem**

Let \( \hat{f}_n(\lambda) = A_n(\lambda)y = (I_n + \lambda M)^{-1}y \) be the CDS estimator from the multivariate model with the order \( m > d/2 \) and denote \( r_n(\lambda) = n^{-1}\|\hat{f}_n(\lambda) - f\|^2 \). If \( n \to \infty \) and \( \lambda \sim n^{-2m/(2m+d)} \) is chosen, then

\[
E[r_n(\lambda)] = O(n^{-\frac{2m}{2m+d}}).
\]

- The above matches the optimal rate in Stone (1982).
Step 4: Rate of Convergence

Asymptotic rate of convergence

**Theorem**

Let \( \hat{f}_n(\lambda) = A_n(\lambda)y = (I_n + \lambda M)^{-1}y \) be the CDS estimator from the multivariate model with the order \( m > d/2 \) and denote \( r_n(\lambda) = n^{-1}\|\hat{f}_n(\lambda) - f\|^2. \) If \( n \to \infty \) and \( \lambda \sim n^{-2m/(2m+d)} \) is chosen, then

\[
E[r_n(\lambda)] = O(n^{-\frac{2m}{2m+d}}).
\]

The above matches the optimal rate in Stone (1982).
Outline

1. Background
2. Data Driven Method
3. Theoretical Consideration: Rate of Convergence
4. Asymptotic Optimality of the Generalized Cross Validation
5. Fast Computation
6. Simulations
Asymptotic Optimality

Choose the parameter $\lambda$ via the Generalized Cross Validation

Let $\hat{f}_n(\lambda) = A_n(\lambda)y = (I_n + \lambda M)^{-1}y$ be the estimator of CDS model with the order $m$ and denote $r_n(\lambda) = n^{-1}||\hat{f}_n(\lambda) - f||^2$. The asymptotic optimality of GCV is defined as

$$\frac{r_n(\hat{\lambda}_G)}{\inf_{\lambda \in \mathbb{R}^+} r_n(\lambda)} \overset{p}{\rightarrow} 1$$  \quad (9)

where $\overset{p}{\rightarrow}$ means the convergence in probability.
Asymptotic Optimality

- Choose the parameter $\lambda$ via the Generalized Cross Validation.
- Let $\hat{f}_n(\lambda) = A_n(\lambda)y = (I_n + \lambda M)^{-1}y$ be the estimator of CDS model with the order $m$ and denote $r_n(\lambda) = n^{-1}\|\hat{f}_n(\lambda) - f\|^2$. The asymptotic optimality of GCV is defined as

$$\frac{r_n(\hat{\lambda}_G)}{\inf_{\lambda \in \mathbb{R}^+} r_n(\lambda)} \xrightarrow{p} 1$$

(9)

where $\xrightarrow{p}$ means the convergence in probability.
Three Conditions

(A.1) \( \inf_{\lambda \in \mathbb{R}_+} nE[r_n(\lambda)] \to \infty. \)

(A.2) There exists a sequence \( \{\lambda_n\} \) such that \( r_n(\lambda_n) \to^p 0 \) (the convergence in probability).

(A.3) Let \( 0 \leq \kappa_1 \leq \cdots \leq \kappa_n \) be the eigenvalues of \( K_n(\lambda) = \lambda M \). For any \( \ell \) such that \( \frac{\ell}{n} \to 0 \), then \( \frac{(n^{-1} \sum_{i=\ell+1}^{n} \kappa_i^{-1})^2}{n^{-1} \sum_{i=\ell+1}^{n} \kappa_i^{-2}} \to 0 \) as \( n \to \infty \).
Three Conditions

\[(A.1)\] \[\inf_{\lambda \in \mathbb{R}_+} nE[r_n(\lambda)] \to \infty.\]

\[(A.2)\] There exists a sequence \(\{\lambda_n\}\) such that \(r_n(\lambda_n) \to p 0\) (the convergence in probability).

\[(A.3)\] Let \(0 \leq \kappa_1 \leq \cdots \leq \kappa_n\) be the eigenvalues of \(K_n(\lambda) = \lambda M\). For any \(\ell\) such that \(n^{-1} \sum_{i=\ell+1}^n \kappa_i^{-1} \to 0\), then
\[\frac{\left( n^{-1} \sum_{i=\ell+1}^n \kappa_i^{-1}\right)^2}{n^{-1} \sum_{i=\ell+1}^n \kappa_i^{-2}} \to 0\] as \(n \to \infty\).
Three Conditions

(A.1) \( \inf_{\lambda \in \mathbb{R}_+} nE[r_n(\lambda)] \to \infty. \)

(A.2) There exists a sequence \( \{\lambda_n\} \) such that \( r_n(\lambda_n) \to_p 0 \) (the convergence in probability).

(A.3) Let \( 0 \leq \kappa_1 \leq \cdots \leq \kappa_n \) be the eigenvalues of \( K_n(\lambda) = \lambda M \). For any \( \ell \) such that \( \frac{\ell}{n} \to 0 \), then \( \frac{n^{-1} \sum_{i=\ell+1}^n \kappa_i^{-1}}{n^{-1} \sum_{i=\ell+1}^n \kappa_i^{-2}} \to 0 \) as \( n \to \infty \).
Asymptotic Optimality of GCV

Formal result

Theorem

Under conditions (A.1), (A.2) and (A.3), \( \hat{f}_n(\hat{\lambda}_G) \) is asymptotically optimal, where \( \hat{\lambda}_G \) is the GCV choice.
Asymptotic Optimality of GCV

Formal result

**Theorem**

Under conditions (A.1), (A.2) and (A.3), $\hat{f}_n(\hat{\lambda}_G)$ is asymptotically optimal, where $\hat{\lambda}_G$ is the GCV choice.
Outline

1 Background

2 Data Driven Method

3 Theoretical Consideration: Rate of Convergence

4 Asymptotic Optimality of the Generalized Cross Validation

5 Fast Computation

6 Simulations
A Permutation Approach

- Sparsity of design matrix $\mathbf{M}$; Recall $\hat{\mathbf{f}} = (\mathbf{I}_n + \lambda \cdot \mathbf{M})^{-1} \cdot \mathbf{Y}$
- The number of nonzeros of $\mathbf{M}$ is strictly less than $(k + 1)^2 n$
- Roughly $3kn$ as shown in numerical experiments
- $\mathbf{M}$ can be permuted to a band matrix by the symmetric reverse Cuthill-McKee ordering (1969) with $O(k \log(k)n)$ complexity
- See an example on the next page...
Sparsity of design matrix $M$; Recall $\hat{f} = (I_n + \lambda \cdot M)^{-1} \cdot Y$

- The number of nonzeros of $M$ is strictly less than $(k + 1)^2 n$
- Roughly $3kn$ as shown in numerical experiments
- $M$ can be permutated to a band matrix by the symmetric reverse Cuthill-McKee ordering (1969) with $O(k \log(k)n)$ complexity
- See an example on the next page...
A Permutation Approach

- Sparsity of design matrix $\mathbf{M}$; Recall $\hat{\mathbf{f}} = (\mathbf{I}_n + \lambda \cdot \mathbf{M})^{-1} \cdot \mathbf{Y}$
- The number of nonzeros of $\mathbf{M}$ is strictly less than $(k + 1)^2 n$
- Roughly $3kn$ as shown in numerical experiments
- $\mathbf{M}$ can be permuted to a band matrix by the symmetric reverse Cuthill-Mckee ordering (1969) with $O(k \log(k)n)$ complexity
- See an example on the next page...
A Permutation Approach

- Sparsity of design matrix $\mathbf{M}$; Recall $\hat{\mathbf{f}} = (\mathbf{I}_n + \lambda \cdot \mathbf{M})^{-1} \cdot \mathbf{Y}$
- The number of nonzeros of $\mathbf{M}$ is strictly less than $(k + 1)^2 n$
- Roughly $3kn$ as shown in numerical experiments
- $\mathbf{M}$ can be permuted to a band matrix by the symmetric reverse Cuthill-McKee ordering (1969) with $O(k \log(k)n)$ complexity

See an example on the next page...
A Permutation Approach

- Sparsity of design matrix $\mathbf{M}$; Recall $\hat{f} = (\mathbf{I}_n + \lambda \cdot \mathbf{M})^{-1} \cdot \mathbf{Y}$
- The number of nonzeros of $\mathbf{M}$ is strictly less than $(k + 1)^2 n$
- Roughly $3kn$ as shown in numerical experiments
- $\mathbf{M}$ can be permuted to a band matrix by the symmetric reverse Cuthill-McKee ordering (1969) with $O(k \log(k) n)$ complexity
- See an example on the next page...
Results of Reordering

Estimation on Irregular Domains
Complexity After Reordering

- $p$ is the bandwidth of reordered matrix $M$
- A band $LDL^T$ decomposition procedure, plus other steps, can find all eigenvalues with $pn^2$ complexity. We observe that $p \approx O((kn)^{0.5})$. So the overall complexity is $O(n^{2.5})$.
- The above so far is an empirical observation.
Complexity After Reordering

- $p$ is the bandwidth of reordered matrix $M$
- A band $LDL^T$ decomposition procedure, plus other steps, can find all eigenvalues with $pn^2$ complexity. We observe that $p \approx O((kn)^{0.5})$.
  So the overall complexity is $O(n^{2.5})$.
- The above so far is an empirical observation
Complexity After Reordering

- $p$ is the bandwidth of reordered matrix $\mathbf{M}$
- A band $LDL^T$ decomposition procedure, plus other steps, can find all eigenvalues with $pn^2$ complexity. We observe that $p \approx O((kn)^{0.5})$. So the overall complexity is $O(n^{2.5})$.
- The above so far is an empirical observation
Outline

1. Background
2. Data Driven Method
3. Theoretical Consideration: Rate of Convergence
4. Asymptotic Optimality of the Generalized Cross Validation
5. Fast Computation
6. Simulations
Two Major Competitors

- CDS (our method): completely-data-drive smoothing,
- Soap film (Wood et al. JRSSB 2008)
  - based on penalty function
    \[ R(f) = \int_{\Omega} \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right)^2 \, dx \, dy \]
- TPS: thin-plate splines
  - with \( \Omega = \mathbb{R}^d \), you have \( \phi(x) = \|z - x_i\|^2 \log \|z - x_i\| \) in 2-D as basis functions

xiaoming@isye.gatech.edu (Georgia Tech)
Two Major Competitors

- CDS (our method): completely-data-drive smoothing,
- Soap film (Wood et al. JRSSB 2008)
  - based on penalty function

\[ R(f) = \int_{\Omega} \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right)^2 \, dx \, dy \]

- TPS: thin-plate splines
  - with \( \Omega = \mathbb{R}^d \), you have \( \phi(z) = \|z - x_i\|^2 \log \|z - x_i\| \) in 2-D as basis functions
Two Major Competitors

- **CDS (our method):** completely-data-drive smoothing,
- **Soap film** (Wood *et al.* JRSSB 2008)
  - based on penalty function

\[
R(f) = \int_{\Omega} \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right)^2 dxdy
\]

- **TPS:** thin-plate splines
  - with \( \Omega = \mathbb{R}^d \), you have \( \phi(x) = ||z - x||^2 \log ||z - x|| \) in 2-D as basis functions
CDS (our method): completely-data-drive smoothing,
Soap film (Wood et al. JRSSB 2008)
  based on penalty function

\[ R(f) = \int_{\Omega} \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right)^2 \, dx \, dy \]

TPS: thin-plate splines
  with \( \Omega = \mathbb{R}^d \), you have \( \phi(z) = \|z - x_i\|^2 \log \|z - x_i\| \) in 2-D as basis functions
Two Major Competitors

- CDS (our method): completely-data-drive smoothing,
- Soap film (Wood et al. JRSSB 2008)
  - based on penalty function

\[ R(f) = \int_\Omega \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right)^2 dxdy \]

- TPS: thin-plate splines
  - with \( \Omega = \mathbb{R}^d \), you have \( \phi(z) = \|z - x_i\|^2 \log \|z - x_i\| \) in 2-D as basis functions
Recall the Irregular Domains

(a) Horseshoe.

(b) Letter R.
Figure: The first, second, and third rows are for $n = 1000, 2000, 5000$, respectively. From left to right the noise is dominated by $\sigma = 0.1, 1, 10$. 
### Regular Domain $[0, 1]^2$

<table>
<thead>
<tr>
<th></th>
<th>cds</th>
<th>soap</th>
<th>tps</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.060</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.070</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.080</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.090</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```plaintext
xiaoming@isye.gatech.edu (Georgia Tech)
```

---

**Estimation on Irregular Domains**

**Birs**

36 / 38
Figure: The RMSE is scaled by $\log_{10}$. 
The idea of using local estimates to replace an analytical penalty function—*complete data driven approach*—seems appealing.

1. Numerical experiments demonstrate promise
2. Theoretical justification is provided

Acknowledgment: Partially supported by NSF.

Thank You!
The idea of using local estimates to replace an analytical penalty function—*complete data driven approach*—seems appealing.

1. Numerical experiments demonstrate promise
2. Theoretical justification is provided

Acknowledgment: Partially supported by NSF.

Thank You!
The idea of using local estimates to replace an analytical penalty function—*complete data driven approach*—seems appealing.

1. Numerical experiments demonstrate promise
2. Theoretical justification is provided

Acknowledgment: Partially supported by NSF.

Thank You!
The idea of using local estimates to replace an analytical penalty function—*complete data driven approach*—seems appealing.

1. Numerical experiments demonstrate promise
2. Theoretical justification is provided

Acknowledgment: Partially supported by NSF.

Thank You!
The idea of using local estimates to replace an analytical penalty function—*complete data driven approach*—seems appealing.

1. Numerical experiments demonstrate promise
2. Theoretical justification is provided

Acknowledgment: Partially supported by NSF.

Thank You!