Manifold Learning, Explanations and Eigenflows

Marina Meilă
Yu-chia Chen, Samson Koelle, Hanyu Zhang and Ioannis Kevrekidis

University of Washington
mmp@stat.washington.edu

Non-Linear and High Dimensional Inference Workshop
Geometry and Statistics in Data Sciences
Institut Henri Poincaré
October 4, 2022
Mathematics

Mathematical models

Laws of nature

Sciences
Mathematics

Mathematical concepts:
Parameters,
Scalar functions,
Manifolds,
Vector fields,
Topology,
k-Laplacians

Mathematical models
Laws of nature

Sciences

Data

Scientific concepts

Machine learning
Data science
Outline

Manifold coordinates with Scientific meaning

Machine Learning 1-Laplacians, topology, vector fields
  1-Laplacian $\Delta_1(M)$ estimation from samples
  Analysis of vector fields – Helmholtz-Hodge decomposition
  Harmonic Embedding Spectral Decomposition Algorithm
  Spectral Shortest Homologous Loop Detection
Outline

Manifold coordinates with Scientific meaning

Machine Learning 1-Laplacians, topology, vector fields
1-Laplacian $\Delta_1(M)$ estimation from samples
Analysis of vector fields – Helmholtz-Hodge decomposition
Harmonic Embedding Spectral Decomposition Algorithm
Spectral Shortest Homologous Loop Detection
Motivation – understanding data from a Molecular Dynamics simulation

ethanol

original data

preprocessed
Motivation – understanding data from a Molecular Dynamics simulation

ethanol

original data

after manifold learning

preprocessed
Motivation – understanding data from a Molecular Dynamics simulation

- 2 rotation angles (torsions) describe this manifold
- Can we discover these features automatically? Can we select these angles from a larger set of features with physical meaning?
Idea: Replace data driven coordinates with selected torsions

- **Scientist**: proposes a dictionary $G$ with all variables of interest
- **ML algorithm**: outputs embedding $\phi$
- **ManifoldLasso**: finds new coordinates in $G$ “equivalent” with $\phi$ ← our algorithm

▶ Explanation
  ▶ = find manifold coordinates from among scientific variables of interest
  ▶ in the language of the domain
Idea Replace data driven coordinates with selected torsions

- **Scientist**: proposes a dictionary $G$ with all variables of interest
- **ML algorithm**: outputs embedding $\phi$,
- **MANIFOLDLASSO**: finds new coordinates in $G$ “equivalent” with $\phi$ ← our algorithm

▶ Explanation

▶ = find manifold coordinates from among scientific variables of interest
▶ in the language of the domain
Idea  Replace data driven coordinates with selected torsions

- **Scientist**: proposes a dictionary $G$ with all variables of interest
- **ML algorithm**: outputs embedding $\phi$,
- **ManifoldLasso**: finds new coordinates in $G$ “equivalent” with $\phi$ ← our algorithm

Explanation

- = find manifold coordinates from among scientific variables of interest
- in the language of the domain
**Idea**  Replace data driven coordinates with selected torsions

- **Scientist**: proposes a dictionary $\mathcal{G}$ with all variables of interest
- **ML algorithm**: outputs embedding $\phi$,
- **ManifoldLasso**: finds new coordinates in $\mathcal{G}$ “equivalent” with $\phi$ ← our algorithm

**Explanation**

- = find manifold coordinates from among scientific variables of interest
- in the language of the domain
Idea: Sparse regression in function space

\[ \phi = h \circ g_s \]

manifold coordinates functions from \( \mathcal{G} \) (new coordinates)

Challenges

▶ sparse, non-linear regression problem
▶ ML coordinates \( \phi \) defined up to diffeomorphism
▶ hence, \( h \) cannot take parametric form
▶ we cannot choose a basis for \( h \)
▶ will not \( \phi_k \) depends on single \( g_j \)
▶ will not assume \( \phi \) isometric

Functional (Group) Lasso

▶ optimize

\[
\min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_i - X_i \beta_i \|^2_2 + \lambda \sum_j \| \beta_j \|, \quad (\text{MANIFOLD LASSO})
\]

▶ support \( S \) of \( \beta \) selects \( g_{j_1}, \ldots, g_{j_s} \) from \( \mathcal{G} \)

\[ D\phi = DhDg_s \]

Leibnitz Rule

▶ sparse linear regression problem
▶ For every data \( i \)
  ▶ \( Y_i = \text{grad } \phi(\xi_i) \),
  ▶ \( X_i = \text{grad } g_{1:p}(\xi) \)
  ▶ \( \beta_{ij} = \frac{\partial h}{\partial g_j}(\xi_i) \)

▶ Sparse linear system
\[ Y_i = X_i \beta_i \]

▶ Constraint: subset \( S \) is same for all \( i \)
Idea: Sparse regression in function space

\[ \phi = h \circ g_s \]

manifold functions from \( G \) (new coordinates)

Challenges

- sparse, non-linear regression problem
- ML coordinates \( \phi \) defined up to diffeomorphism
- hence, \( h \) cannot take parametric form
- we cannot choose a basis for \( h \)
- will not \( \phi_k \) depends on single \( g_j \)
- will not assume \( \phi \) isometric

Functional (Group) Lasso

- optimize

\[ \min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_i - X_i \beta_i \|^2 + \lambda \sum_j \| \beta_j \|, \quad \text{(MANIFOLD LASSO)} \]

- support \( S \) of \( \beta \) selects \( g_{j_1}, \ldots, j_s \) from \( G \)

\[ D\phi = DhDg_s \]

Leibnitz Rule

- sparse linear regression problem
- For every data \( i \)
  - \( Y_i = \text{grad } \phi(\xi_i) \),
  - \( X_i = \text{grad } g_{1:p}(\xi) \)
  - \( \beta_{ij} = \frac{\partial h}{\partial g_j}(\xi_i) \)

- Sparse linear system
  \( Y_i = X_i \beta_i \)

- Constraint: subset \( S \) is same for all \( i \)
Idea: Sparse regression in function space

$$\phi = h \circ g_S$$

functions from $G$

(new coordinates)

**Challenges**

▸ sparse, non-linear regression problem
▸ ML coordinates $\phi$ defined up to
diffeomorphism
▸ hence, $h$ cannot take parametric form
▸ we cannot choose a basis for $h$
▸ will not $\phi_k$ depends on single $g_j$
▸ will not assume $\phi$ isometric

**Functional (Group) Lasso**

▸ optimize

$$
\min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_i - X_i \beta_i \|^2 + \lambda \sum_j \| \beta_j \|, \quad \text{(MANIFOLD LASSO)}
$$

▸ support $S$ of $\beta$ selects $g_{j_1},...,g_{j_s}$ from $G$
Idea: Sparse regression in function space

\[ \phi = h \circ g_s \]

**Challenges**
- sparse, non-linear regression problem
- ML coordinates \( \phi \) defined up to diffeomorphism
- hence, \( h \) cannot take parametric form
- we cannot choose a basis for \( h \)
- will not \( \phi_k \) depends on single \( g_j \)
- will not assume \( \phi \) isometric

**Functional (Group) Lasso**

- optimize

\[
\min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_i - X_i \beta \|^2 + \lambda \sum_{j} \| \beta_j \|, \quad (\text{MANIFOLD LASSO})
\]

- support \( S \) of \( \beta \) selects \( g_{j_1}, \ldots, g_{j_s} \) from \( G \)

\[ D\phi = DhDg_s \]

**Leibnitz Rule**

\[ D\phi = DhDg_s \]

**sparse linear regression problem**

- For every data \( i \)
  - \( Y_i = \text{grad} \phi(\xi_i) \)
  - \( X_i = \text{grad} g_{1:p}(\xi) \)
  - \( \beta_{ij} = \frac{\partial h}{\partial g_j}(\xi_i) \)

- Sparse linear system
  - \( Y_i = X_i \beta_i \)

- Constraint: subset \( S \) is same for all \( i \)
Idea: Sparse regression in function space

\[ \phi = h \circ g_s \]

**Challenges**
- **sparse**, non-linear regression problem
- ML coordinates \( \phi \) defined up to diffeomorphism
- hence, \( h \) cannot take parametric form
- we cannot choose a basis for \( h \)
- will not \( \phi_k \) depends on single \( g_j \)
- will not assume \( \phi \) isometric

**Functional (Group) Lasso**
- optimize

\[
\min_\beta J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_i - X_i \beta_i \|_2^2 + \lambda \sum_j ||\beta_j||,
\]

**Constraint:** subset \( S \) is same for all \( i \)

\[ D\phi = DhDg_s \]

**Leibnitz Rule**
- sparse linear regression problem
- For every data \( i \)
  - \( Y_i = \text{grad} \phi(\xi_i) \),
  - \( X_i = \text{grad} g_{1:p}(\xi) \)
  - \( \beta_{ij} = \frac{\partial h}{\partial g_j}(\xi_i) \)
  - Sparse linear system
    \( Y_i = X_i \beta_i \)

**Sparse linear system**
- \( Y_i = X_i \beta_i \)
Idea: Sparse regression in function space

\[
\phi = h \circ g_S
\]
manifold functions from \( \mathcal{G} \) (new coordinates)

Challenges

- sparse, non-linear regression problem
- ML coordinates \( \phi \) defined up to diffeomorphism
- hence, \( h \) cannot take parametric form
- we cannot choose a basis for \( h \)
- will not \( \phi_k \) depends on single \( g_j \)
- will not assume \( \phi \) isometric

Functional (Group) Lasso

- optimize

\[
\min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_i - X_i \beta \|_2^2 + \lambda \sum_j \| \beta_j \|, \quad \text{(MANIFOLD LASSO)}
\]

- support \( S \) of \( \beta \) selects \( g_{j_1}, \ldots, g_{j_s} \) from \( \mathcal{G} \)

\[
D\phi = DhDg_S
\]
Leibnitz Rule

- sparse linear regression problem
- For every data \( i \)
  - \( Y_i = \text{grad } \phi(\xi_i) \)
  - \( X_i = \text{grad } g_{1:p}(\xi) \)
  - \( \beta_{ij} = \frac{\partial h}{\partial g_j}(\xi_i) \)

- Sparse linear system
  \( Y_i = X_i \beta_i \)

- Constraint: subset \( S \) is same for all \( i \)
Idea: Sparse regression in function space

\[ \phi = h \circ g_{S} \]

manifold functions from \( G \) (new coordinates)

Challenges

▸ sparse, non-linear regression problem
▸ ML coordinates \( \phi \) defined up to diffeomorphism
▸ hence, \( h \) cannot take parametric form
▸ we cannot choose a basis for \( h \)
▸ will not \( \phi_{k} \) depends on single \( g_{j} \)
▸ will not assume \( \phi \) isometric

Functional (Group) Lasso

▸ optimize

\[
\min_{\beta} J_{\lambda}(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_{i} - X_{i}\beta_{i} \|^{2} + \lambda \sum_{j} \| \beta_{j} \|, \quad (\text{MANIFOLD LASSO})
\]

▸ support \( S \) of \( \beta \) selects \( g_{j_1}, \ldots, j_s \) from \( G \)

\[ D\phi = DhDg_{S} \]

Leibnitz Rule

▸ sparse linear regression problem
▸ For every data \( i \)
  ▸ \( Y_{i} = \text{grad} \phi(\xi_{i}) \),
  ▸ \( X_{i} = \text{grad} g_{1:p}(\xi) \)
  ▸ \( \beta_{ij} = \frac{\partial h}{\partial g_{j}}(\xi_{i}) \)
▸ Sparse linear system
  \( Y_{i} = X_{i}\beta_{i} \)
▸ Constraint: subset \( S \) is same for all \( i \)
**ManifoldLasso Algorithm**

**Given** Data $\xi_{1:n}$, $\dim \mathcal{M} = d$, embedding $\phi(\xi_{1:n})$, dictionary $\mathcal{G} = \{g_{1:p}\}$

1. Estimate tangent subspace at $\xi_i$ by (weighted) PCA
2. Project dictionary functions gradients $\nabla g_j$ on tangent subspace, obtain $X_{1:n} \in \mathbb{R}^{d \times p}$
3. **Estimate** gradients of $\phi_{1:k}$, obtain $Y_{1:n} \in \mathbb{R}^{d \times m}$
   By pull-back from embedding space $\phi$
4. Solve $\text{GROUPLASSO}(Y_{1:n}, X_{1:n}, d)$, obtain support $S$

$$\min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} ||Y_i - X_i \beta_i||_2^2 + \lambda \sum_j ||\beta_j||, \quad (\text{ManifoldLasso})$$

**Output** $S$
**ManifoldLasso Algorithm**

**Given** Data $\xi_{1:n}$, $\dim \mathcal{M} = d$, embedding $\phi(\xi_{1:n})$, dictionary $\mathcal{G} = \{g_{1:p}\}$

1. Estimate tangent subspace at $\xi_i$ by (weighted) PCA
2. Project dictionary functions gradients $\nabla g_j$ on tangent subspace, obtain $X_{1:n} \in \mathbb{R}^{d \times p}$
3. Estimate gradients of $\phi_{1:k}$, obtain $Y_{1:n} \in \mathbb{R}^{d \times m}$
   By pull-back from embedding space $\phi$
4. Solve $\text{GROUPLASSO}(Y_{1:n}, X_{1:n}, d)$, obtain support $S$

$$
\min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} \| Y_i - X_i \beta_i \|_2^2 + \lambda \sum_j \| \beta_j \|, \quad (\text{ManifoldLasso})
$$

Output $S$
**ManifoldLasso Algorithm**

**Given** Data $\xi_{1:n}$, $\dim M = d$, embedding $\phi(\xi_{1:n})$, dictionary $G = \{g_{1:p}\}$

1. Estimate tangent subspace at $\xi_i$ by (weighted) PCA
2. Project dictionary functions gradients $\nabla g_j$ on tangent subspace, obtain $X_{1:n} \in \mathbb{R}^{d \times p}$
3. Estimate gradients of $\phi_{1:k}$, obtain $Y_{1:n} \in \mathbb{R}^{d \times m}$
   By pull-back from embedding space $\phi$
4. Solve $\text{GROUPLASSO}(Y_{1:n}, X_{1:n}, d)$, obtain support $S$

$$\begin{align*}
\min_{\beta} J_\lambda(\beta) &= \frac{1}{2} \sum_{i=1}^{n} \|Y_i - X_i \beta_i\|_2^2 + \lambda \sum_j \|\beta_j\|, \quad (\text{ManifoldLasso})
\end{align*}$$

Output $S$
**ManifoldLasso Algorithm**

**Given** Data \( \xi_{1:n} \), \( \text{dim} \mathcal{M} = d \), embedding \( \phi(\xi_{1:n}) \), dictionary \( \mathcal{G} = \{ g_{1:p} \} \)

1. Estimate tangent subspace at \( \xi_i \) by (weighted) PCA
2. Project dictionary functions gradients \( \nabla g_j \) on tangent subspace, obtain \( X_{1:n} \in \mathbb{R}^{d \times p} \)
3. **Estimate** gradients of \( \phi_{1:k} \), obtain \( Y_{1:n} \in \mathbb{R}^{d \times m} \)
   By pull-back from embedding space \( \phi \)
4. Solve **GroupLasso**(\( Y_{1:n}, X_{1:n}, d \)), obtain support \( S \)

\[
\min_{\beta} J_\lambda(\beta) = \frac{1}{2} \sum_{i=1}^{n} ||Y_i - X_i \beta_i||^2_2 + \lambda \sum_j ||\beta_j||, \quad \text{(ManifoldLasso)}
\]

Output \( S \)
Ethanol MD simulation

regularization paths $\beta_{1:p}$ vs $\lambda$
Theory

- When is $S$ unique? / When can $\mathcal{M}$ be uniquely parametrized by $G$?
  Functional independence conditions on dictionary $G$ and subset $g_{j_1,\ldots,j_s}$

- Basic result
  
  $$f_S = h \circ f_{S'} \text{ on } U \iff \text{rank} \begin{pmatrix} Df_S \\ Df_{S'} \end{pmatrix} = \text{rank} Df_{S'} \text{ on } U$$

- When can GLASSO recover $S$?
  (Simple) Incoherence Conditions

  \[
  \mu = \max_{i=1:n, j \in S, j' \not\in S} \frac{|X_{ji} X_{j'i}|}{\|X_{ji}\| \|X_{j'i}\|} \quad \nu = \frac{1}{\min_{i=1:n} \|X_{iS} X_{iS}\|_2} \quad nd \sigma^2 = \sum_{i,k} \epsilon_{ik}^2
  \]

Theorem If, $\|X_{1:p}\| = 1$, $\mu \nu \sqrt{d} + \frac{\sigma \sqrt{nd}}{\lambda} < 1$ then $\beta_j = 0$ for $j \not\in S$. 
Theory

- When is $S$ unique? / When can $\mathcal{M}$ be uniquely parametrized by $G$?
  Functional independence conditions on dictionary $G$ and subset $g_{j_1, \ldots, j_s}$

- Basic result
  
  $f_S = h \circ f_{S'}$ on $U$ iff

  \[
  \text{rank} \left( \begin{array}{c} Df_S \\ Df_{S'} \end{array} \right) = \text{rank} Df_{S'} \quad \text{on } U
  \]

- When can $\text{GLasso}$ recover $S$?
  (Simple) Incoherence Conditions

  \[
  \mu = \max_{i=1:n, j \in S, j' \not\in S} \frac{|X_{ji}' X_{j'i}'|}{\|X_{ji}\| \|X_{j'i}'\|}, \quad \nu = \frac{1}{\min_{i=1:n} \|X_{iS}'X_{iS}\|_2} \quad nd\sigma^2 = \sum_{i,k} \epsilon_{ik}^2
  \]

  **Theorem** If, $\|X_{1:p}\| = 1$, $\mu \nu \sqrt{d} + \frac{\sigma \sqrt{nd}}{\lambda} < 1$ then $\beta_j = 0$ for $j \not\in S$. 
Theorem 7 (Support recovery) Assume that equation (30) holds, and that \( \sum_{i=1}^{n} \|x_{ij}\|^2 = \gamma_j^2 \) for all \( j = 1 : p \). Let \( \gamma_{\max} = \max_{j \notin S} \gamma_j \), \( \kappa_S = \max_{i=1:n} \max_{j \in S} \frac{\|x_{ij}\|}{\min_{j \in S} \|x_{ij}\|} \). Denote by \( \bar{\beta} \) the solution of (31) for some \( \lambda > 0 \). If \( 1 - (s-1)\mu > 0 \) and

\[
\gamma_{\max} \left( \frac{\mu}{1 - (s-1)\mu} \frac{\kappa_S}{\min_{i=1:n} \min_{j \in S} \|x_{ij}\|} + \frac{\sigma \sqrt{d}}{\lambda \sqrt{n}} \right) \leq 1
\]

then \( \bar{\beta}_{ij} = 0 \) for \( j \notin S \) and all \( i = 1, \ldots n \).

Corollary 8 Assume that equation (31) and condition (37) hold. Let \( \kappa = \frac{\mu}{1 - (s-1)\mu} \frac{\kappa_S}{\min_{i=1:n} \min_{j \in S} \|x_{ij}\|} \) and \( \gamma_S = \|\tilde{X}_S\| \). Denote by \( \hat{\beta} \) the solution to problem (31) for some \( \lambda > 0 \). If (1) \( \lambda = c \frac{\gamma_{\max}\sigma \sqrt{d}}{1 - \kappa \gamma_{\max}} \), \( c > 1 \), and (2) \( \|\beta_j\| > \sigma \sqrt{d}(\gamma_{\max} + \gamma_S) + \lambda(1 + \sqrt{s}) \) for all \( j \in S \), then the support \( S \) is recovered exactly and

\[
\|\hat{\beta}_j - \beta_j^*\| < \sigma \sqrt{d}(\gamma_{\max} + \gamma_S) + \lambda(1 + \sqrt{s}) = \sigma \sqrt{d} \gamma_{\max} \left[ 1 + \frac{\gamma_S}{\gamma_{\max}} + c \frac{1 + \sqrt{s}}{1 - \kappa \gamma_{\max}} \right] \quad \text{for all } j \in S.
\]
**Proposition 2** (after (2)). Let $\mathcal{F}, f_j$ be dictionary and dictionary functions on the $d$–dimensional smooth manifold $\mathcal{M}$. Assume $f_j \in C^\ell$ with $\ell \geq d + 1$. Suppose $S \subset [p]$, and denote by $\text{grad } f_S$ the $\mathbb{R}^{d \times s}$ matrix of concatenated $\text{grad } f_j : f \in S$. Then, if there is a subset $S' \subsetneq S$ such that the following rank condition holds globally:

$$\text{rank} \left( \begin{array}{c} \text{grad } f_S \\ \text{grad } f_{S'} \end{array} \right) = \text{rank } \text{grad } f_{S'}.$$  

Then there exists a function $h$ which is $C^\ell$ almost everywhere in the image of $f_{S'}(\mathcal{M})$ such that $f_S = h \circ f_{S'}$

$$\mu_S = \sup_{\xi \in \mathcal{M}, j \in S, j' \notin S} |X_{\{j\}}^T X_{\{j'\}} \xi|$$  

$$\nu_S = \sup_{\xi \in \mathcal{M}, \alpha \in \mathbb{R}^d, \|\alpha\|_2 = 1} \alpha^T (X_S^T X_S)^{-1} \alpha.$$  

**Proposition 3.** Assume that

1. $\mathcal{M}$ is $d$–dimensional $C^k$ compact manifold with strictly positive reach.
2. Data $\xi$ are sampled from some density $p$ on $\mathcal{M}$ with $p > 0$ all over $\mathcal{M}$.
3. $\xi \in \mathcal{M}$ with probability 1 under $p$.

Let $S$ be the 'true' support, $S(\hat{B})$ be the support selected by TSLASSO, $\mu_S$ and $\nu_S$ be defined by (5) and (6), and further assume

4. $|S| = d$.
5. $Df_S$ has rank $d$ on $\mathcal{M}$.
6. $\mu_S \nu_S d < 1$.

Then if we adopt the tangent space estimation algorithm in (7) with bandwidth choice $h = O((\log n/(n - 1))^d$, with $n \geq ((1 - \mu_S \nu_S d)/2 \nu_S d)^d/(k-1)$ we have

$$\Pr(S(\hat{B}) \subset S) \geq 1 - O \left( \frac{1}{n^\frac{k}{d}} \right).$$
### Experiments

<table>
<thead>
<tr>
<th>Dataset</th>
<th>(n)</th>
<th>(N_a)</th>
<th>(D)</th>
<th>(d)</th>
<th>(\epsilon_N)</th>
<th>(m)</th>
<th>(n')</th>
<th>(p)</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>SwissRoll</td>
<td>10000</td>
<td>NA</td>
<td>51</td>
<td>2</td>
<td>.18</td>
<td>2</td>
<td>100</td>
<td>51</td>
<td>synthetic</td>
</tr>
<tr>
<td>RigidEthanol</td>
<td>10000</td>
<td>9</td>
<td>50</td>
<td>2</td>
<td>3.5</td>
<td>3</td>
<td>100</td>
<td>12</td>
<td>skeleton (G)</td>
</tr>
<tr>
<td>Ethanol</td>
<td>50000</td>
<td>9</td>
<td>50</td>
<td>2</td>
<td>3.5</td>
<td>3</td>
<td>100</td>
<td>12</td>
<td>skeleton (G)</td>
</tr>
<tr>
<td>Malonaldehyde</td>
<td>50000</td>
<td>9</td>
<td>50</td>
<td>2</td>
<td>3.5</td>
<td>3</td>
<td>100</td>
<td>12</td>
<td>skeleton (G)</td>
</tr>
<tr>
<td>Toluene</td>
<td>50000</td>
<td>16</td>
<td>50</td>
<td>1</td>
<td>1.9</td>
<td>2</td>
<td>100</td>
<td>30</td>
<td>skeleton (G)</td>
</tr>
<tr>
<td>Ethanol</td>
<td>50000</td>
<td>9</td>
<td>50</td>
<td>2</td>
<td>3.5</td>
<td>3</td>
<td>100</td>
<td>756</td>
<td>exhaustive (G)</td>
</tr>
<tr>
<td>Malonaldehyde</td>
<td>50000</td>
<td>9</td>
<td>50</td>
<td>2</td>
<td>3.5</td>
<td>3</td>
<td>100</td>
<td>756</td>
<td>exhaustive (G)</td>
</tr>
<tr>
<td>(\phi)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lasso (</td>
</tr>
</tbody>
</table>

\(p\) = dictionary size, \(m\) = embedding dimension, \(n\) = sample size for manifold estimation, \(n'\) = sample size for MANIFOLDLASSO
Two-stage sparse recovery for exhaustive $G$, $p = 756$

Ethanol

Malonaldehyde
Tangent Space Lasso experiments
Summary of **ManifoldLasso/FunctionalLasso**

Technical contribution

- **non-linear** sparse regression in function spaces
- Method to push/pull vectors through mappings $\phi$
- **ManifoldLasso**: regression of data driven coordinates $\phi_{1:m}$ on domain-specific functions $G = \{g_{1:p}\}$

- explains large scale structure with domain-relevant functions
- non-parametric; different from **symbolic regression** [Brunton et al. 2016, Rudy et al. 2019]
- transmissible knowledge, compare embeddings from different experiments
- extensions: estimated $\nabla g$, simultaneous explanation of multiple manifolds
Summary of **ManifoldLasso/FunctionalLasso**

Technical contribution

- **non-linear** sparse regression in function spaces
- Method to push/pull vectors through mappings $\phi$
- **ManifoldLasso**: regression of data driven coordinates $\phi_{1:m}$ on domain-specific functions $G = \{ g_{1:p} \}$

- explains large scale structure with domain-relevant functions
- non-parametric; different from symbolic regression [Brunton et al. 2016, Rudy et al. 2019]
- transmissible knowledge, compare embeddings from different experiments
- extensions: estimated $\nabla g$, simultaneous explanation of multiple manifolds
Learning with flows and vector fields [with Yu-chia Chen, Yoannis Kevrekidis]
Outline

Manifold coordinates with Scientific meaning

Machine Learning 1-Laplacians, topology, vector fields
1-Laplacian $\Delta_1(M)$ estimation from samples
Analysis of vector fields – Helmholtz-Hodge decomposition
Harmonic Embedding Spectral Decomposition Algorithm
Spectral Shortest Homologous Loop Detection
Why Laplacians? Why higher order?

The Laplacian $L_0 \in \mathbb{R}^{n \times n}$ is central to Manifold Learning

- embedding data by Diffusion Maps [Coifman, Lafon 2006]
- Spectral Clustering

- $L_0$ related to Riemannian metric – captures geometry of $M$
- Function approximation
- Smoothing, semi-supervised learning (Laplacian regularization) on manifolds

Higher order Laplacians $\Delta_1, \ldots, \Delta_k$ also capture geometry and topology of $M$

This talk

- estimate first order Laplacian (Helmholtzian) $L_1(M)$ from data
- calculate Helmholtz-Hodge decomposition of $L_1(M)$ from data
- Smoothing, function approximation, semi-supervised learning (Laplacian regularization) for vector fields on manifolds
- Manifold prime decomposition ($\approx$ Spectral clustering)
- find short loop bases
Why Laplacians? Why higher order?

The Laplacian $L_0 \in \mathbb{R}^{n \times n}$ is central to Manifold Learning

- embedding data by Diffusion Maps [Coifman, Lafon 2006]
- Spectral Clustering

- $L_0$ related to Riemannian metric – captures geometry of $\mathcal{M}$
- Function approximation
- Smoothing, semi-supervised learning (Laplacian regularization) on manifolds

Higher order Laplacians $\Delta_1, \ldots, \Delta_k$ also capture geometry and topology of $\mathcal{M}$

This talk

- estimate first order Laplacian (Helmholtzian) $L_1(\mathcal{M})$ from data
- calculate Helmholtz-Hodge decomposition of $L_1(\mathcal{M})$ from data
- Smoothing, function approximation, semi-supervised learning (Laplacian regularization) for vector fields on manifolds
- Manifold prime decomposition ($\approx$ Spectral clustering)
- find short loop bases
Why Laplacians? Why higher order?

The Laplacian $\mathcal{L}_0 \in \mathbb{R}^{n \times n}$ is central to Manifold Learning

▶ embedding data by Diffusion Maps [Coifman, Lafon 2006]
▶ Spectral Clustering

▶ $\mathcal{L}_0$ related to Riemannian metric – captures geometry of $\mathcal{M}$
▶ Function approximation
▶ Smoothing, semi-supervised learning (Laplacian regularization) on manifolds

Higher order Laplacians $\Delta_1, \ldots \Delta_k$ also capture geometry and topology of $\mathcal{M}$

This talk

▶ estimate first order Laplacian (Helmholtzian) $\mathcal{L}_1(\mathcal{M})$ from data
▶ calculate Helmholtz-Hodge decomposition of $\mathcal{L}_1(\mathcal{M})$ from data
▶ Smoothing, function approximation, semi-supervised learning (Laplacian regularization) for vector fields on manifolds
▶ Manifold prime decomposition ($\approx$ Spectral clustering)
▶ find short loop bases
Why Laplacians? Why higher order?

The Laplacian $\mathcal{L}_0 \in \mathbb{R}^{n \times n}$ is central to Manifold Learning

- embedding data by Diffusion Maps [Coifman, Lafon 2006]
- Spectral Clustering

- $\mathcal{L}_0$ related to Riemannian metric – captures geometry of $\mathcal{M}$
- Function approximation
- Smoothing, semi-supervised learning (Laplacian regularization) on manifolds

Higher order Laplacians $\Delta_1, \ldots \Delta_k$ also capture geometry and topology of $\mathcal{M}$

This talk

- estimate first order Laplacian (Helmholtzian) $\mathcal{L}_1(\mathcal{M})$ from data
- calculate Helmholtz-Hodge decomposition of $\mathcal{L}_1(\mathcal{M})$ from data
- Smoothing, function approximation, semi-supervised learning (Laplacian regularization) for vector fields on manifolds
- Manifold prime decomposition ($\approx$ Spectral clustering)
- find short loop bases
Estimating the 1-Laplacian with samples from $\mathcal{M}$

$SC_2 = (V, E, T)$

$e = (s, t) \in E$
if $\|X_s - X_t\| < \delta$

$t = (s, t, u) \in T$
if $(s, t), (t, u), (s, u) \in E$

VR Complex

$w_T(x, y, z) = \kappa_\epsilon(x, y)\kappa_\epsilon(y, z)\kappa_\epsilon(x, z)$

$w_E(x, y) = |B_T|w_T$

$w_E(x, y) + w_E(x, z) + w_E(x, z') + \ldots$

$\mathcal{L}_1 = a \cdot \mathcal{L}_1^{\text{down}} + b \cdot \mathcal{L}_1^{\text{up}}$

$\mathcal{L}_1^{\text{down}} = B_E^T W_V^{-1} B_E W_E$

$\mathcal{L}_1^{\text{up}} = W_E^{-1} B_T W_T B_E^T$

$\mathcal{L}_1 = a \cdot \mathcal{L}_1^{\text{down}} + b \cdot \mathcal{L}_1^{\text{up}}$

$C_1 \cong \mathbb{R}^{n_E} = \text{gradient} \oplus \text{harmonic} \oplus \text{curl}$
$L_1$ estimation for Molecular Dynamics data (malonaldehyde)

graph Laplacian $\omega_t = 1$, [Berry, Giannakis 2020], [Chen, M, Kevrekidis 2020]
Outline

Manifold coordinates with Scientific meaning

Machine Learning 1-Laplacians, topology, vector fields
  1-Laplacian $\Delta_1(M)$ estimation from samples
  Analysis of vector fields – Helmholtz-Hodge decomposition
  Harmonic Embedding Spectral Decomposition Algorithm
  Spectral Shortest Homologous Loop Detection
Eigenfunctions of $\mathcal{L}_1$ – what are they useful for?

- **Helmholtz-Hodge Decomposition** classifies eigenfunctions of $\mathcal{L}_1$

$$\mathcal{C}_1 \cong \mathbb{R}^{nE} \cong \text{Im} \mathcal{L}_1^{\text{down}} \oplus \text{Null} \mathcal{L}_1 \oplus \text{Im} \mathcal{L}_1^{\text{up}}$$

- **Analysis of vector fields on $\mathcal{M}$**
  - Decompose onto harmonic, gradient, curl
  - Smooth, predict, extend, complete a flow

- **Analysis of $\mathcal{M}$**
  - $\mathcal{H}_1 = \text{Null} \mathcal{L}_1$ Space of loops on $\mathcal{M}$ (1st co-homology space)
  - $\dim \mathcal{H}_1 = \beta_1$ number of (independent loops)
  - Find shortest loop basis
Helmholtz-Hodge decomposition for ocean buoys data

simplicial complex \((V, E, T)\)
Flow Smoothing

1. Smoothed flow: $\omega = (I + \alpha \mathcal{L}_e)^{-1} \omega$
2. Obtain vertex-wise vector field by solving a linear system

A 

B $\alpha = 5$
C $\alpha = 50$
D $\alpha = 500$
Flow Completion – Semi-Supervised Learning (SSL)
Outline

Manifold coordinates with Scientific meaning

Machine Learning 1-Laplacians, topology, vector fields
- 1-Laplacian $\Delta_1(M)$ estimation from samples
- Analysis of vector fields – Helmholtz-Hodge decomposition
  Harmonic Embedding Spectral Decomposition Algorithm
- Spectral Shortest Homologous Loop Detection
Connected sum and manifold (prime) decomposition

The connected sum $\mathcal{M} = \mathcal{M}_1 \# \mathcal{M}_2$:

1. removing two $d$-dimensional “disks” from $\mathcal{M}_1$ and $\mathcal{M}_2$ (shaded area)
2. gluing together two manifolds at the boundaries

Existence of prime decomposition: factorize a manifold $\mathcal{M} = \mathcal{M}_1 \# \cdots \# \mathcal{M}_\kappa$ into $\mathcal{M}_i$’s so that $\mathcal{M}_i$ is a prime manifold

- $d = 2$: classification theorem of surfaces
- $d = 3$: the uniqueness of the prime decomposition was shown by Kneser-Milnor theorem
- $d \geq 5$: proved the existence of factorization (but they might not be unique)
The connected sum $\mathcal{M} = \mathcal{M}_1 \# \mathcal{M}_2$:

1. removing two $d$-dimensional “disks” from $\mathcal{M}_1$ and $\mathcal{M}_2$ (shaded area)

2. gluing together two manifolds at the boundaries

Existence of prime decomposition: factorize a manifold $\mathcal{M} = \mathcal{M}_1 \# \cdots \# \mathcal{M}_\kappa$ into $\mathcal{M}_i$’s so that $\mathcal{M}_i$ is a prime manifold

- $d = 2$: classification theorem of surfaces
- $d = 3$: the uniqueness of the prime decomposition was shown by Kneser-Milnor theorem
- $d \geq 5$: proved the existence of factorization (but they might not be unique)
The decomposition of the higher-order homology embedding constructed from the $k$-Laplacian [Chen, M NeurIPS 2021]

- $\mathcal{L}_1$ is $n_E \times n_E$, operates on edges flows of neighborhood graph
- Null $\mathcal{L}_1 = \mathcal{H}_1$ harmonic space, $\beta_1 = \dim \mathcal{H}_1$
- $Y$ is basis of $\mathcal{H}_1$ harmonic flows
- $Y$ NOT UNIQUE

Harmonic Eigenfunctions $Y$ (raw) vs. $Z$ (decoupled)
Connected sum as a matrix perturbation: Assumptions

1. Points are sampled from a decomposable manifold
   - $\kappa$-fold connected sum: $\mathcal{M} = \mathcal{M}_1 \# \cdots \# \mathcal{M}_\kappa$
   - $H_k(\text{SC})$ (discrete) and $H_k(\mathcal{M}, \mathbb{R})$ (continuous) are isomorphic. Also for every $\mathcal{M}_i$
     - Works for any consistent method to build $\mathcal{L}_k$
     - We use our prior work $\beta$ for $\mathcal{L}_1$

2. No $k$-homology class is created/destroyed during the connected sum
   - If $\dim(\mathcal{M}) > k$, then $H_k(\mathcal{M}_1 \# \mathcal{M}_2) \cong H_k(\mathcal{M}_1) \oplus H_k(\mathcal{M}_2)$
   - [Technical] The eigengap of $\mathcal{L}_k$ is the min of each $\hat{\mathcal{L}}_k^{(ii)}$:
     $$\delta = \min \{ \delta_1, \cdots, \delta_\kappa \}$$

3. Sparsely connected manifold
   - Not too many triangles are created/destroyed during connected sum (for $k = 1$)
   - Empirically, the perturbation is small even when $\mathcal{M}$ is not sparsely connected
   - [Technical] Perturbations of $\ell$-simplex set $\Sigma_\ell$ are small ($\epsilon_\ell$ and $\epsilon'_\ell$ are small) for $\ell = k, k - 1$
Connected sum as a matrix perturbation: Assumptions

1. Points are sampled from a decomposable manifold
   - $\kappa$-fold connected sum: $\mathcal{M} = \mathcal{M}_1 \# \cdots \# \mathcal{M}_\kappa$
   - $\mathcal{H}_k(\text{SC})$ (discrete) and $H_k(\mathcal{M}, \mathbb{R})$ (continuous) are isomorphic. Also for every $\mathcal{M}_i$
     - Works for any consistent method to build $\mathcal{L}_k$
     - We use our prior work ? for $\mathcal{L}_1$

2. No $k$-homology class is created/destroyed during the connected sum
   - If $\dim(\mathcal{M}) > k$, then $\mathcal{H}_k(\mathcal{M}_1 \# \mathcal{M}_2) \cong \mathcal{H}_k(\mathcal{M}_1) \oplus \mathcal{H}_k(\mathcal{M}_2)$?
   - [Technical] The eigengap of $\mathcal{L}_k$ is the min of each $\hat{\mathcal{L}}_k^{(ii)}$:
     \[ \delta = \min\{\delta_1, \cdots, \delta_\kappa\} \]

3. Sparsely connected manifold
   - Not too many triangles are created/destroyed during connected sum (for $k = 1$)
   - Empirically, the perturbation is small even when $\mathcal{M}$ is not sparsely connected
   - [Technical] Perturbations of $\ell$-simplex set $\Sigma_\ell$ are small ($\epsilon_\ell$ and $\epsilon'_\ell$ are small) for $\ell = k, k - 1$
Connected sum as a matrix perturbation: Assumptions

1. Points are sampled from a decomposable manifold
   ▶ $\kappa$-fold connected sum: $\mathcal{M} = \mathcal{M}_1 \# \cdots \# \mathcal{M}_\kappa$
   ▶ $H_k(\text{SC})$ (discrete) and $H_k(\mathcal{M}, \mathbb{R})$ (continuous) are isomorphic. Also for every $\mathcal{M}_i$
     ▶ Works for any consistent method to build $L_k$
     ▶ We use our prior work $?^?$ for $L_1$

2. No $k$-homology class is created/destroyed during the connected sum
   ▶ If $\dim(\mathcal{M}) > k$, then $H_k(\mathcal{M}_1 \# \mathcal{M}_2) \cong H_k(\mathcal{M}_1) \oplus H_k(\mathcal{M}_2)$?
   ▶ [Technical] The eigengap of $L_k$ is the min of each $\hat{L}_k^{(ii)}$:
     $\delta = \min\{\delta_1, \cdots, \delta_\kappa\}$

3. Sparsely connected manifold
   ▶ Not too many triangles are created/destroyed during connected sum (for $k = 1$)
   ▶ Empirically, the perturbation is small even when $\mathcal{M}$ is not sparsely connected
   ▶ [Technical] Perturbations of $\ell$-simplex set $\Sigma_\ell$ are small ($\epsilon_\ell$ and $\epsilon'_\ell$ are small) for $\ell = k, k - 1$
Subspace perturbation

**Theorem 1**
*Under Assumptions 1–3*

\[
\|\text{DiffL}^\text{down}_k\|^2 \leq \left[ 2 \sqrt{\epsilon'_k + \epsilon'_k} + (1 + \sqrt{\epsilon'_k})^2 \sqrt{\epsilon'_{k-1} + 4 \epsilon_{k-1}} \right]^2 (k + 1)^2; \quad \text{and}
\]

\[
\|\text{DiffL}^\text{up}_k\|^2 \leq \left[ 2 \sqrt{\epsilon'_k + \epsilon'_k + 2 \epsilon_k + 4 \epsilon_k} \right]^2 (k + 2)^2,
\]

*and there exists a unitary matrix* \( \mathbf{O} \in \mathbb{R}^{\beta_k \times \beta_k} \) *such that*

\[
\|\mathbf{Y}_{N_k} - \hat{\mathbf{Y}}_{N_k} : \mathbf{O}\|^2_F \leq \frac{8 \beta_k \left[ \|\text{DiffL}^\text{down}_k\|^2 + \|\text{DiffL}^\text{up}_k\|^2 \right]}{\min\{\delta_1, \cdots, \delta_\kappa\}}.
\]  

1. **Assu. 2**: no topology is destroyed/created
2. **Assu. 3**: sparsely connected
3. **N_k**: bound only simplexes that are not altered during connected sum
Subspace perturbation

**Theorem 1**

*Under Assumptions 1–3*

\[
\|\text{DiffL}_{k}^{\text{down}}\|^2 \leq \left[ 2\sqrt{\epsilon'_k + \epsilon'_k} + \left( 1 + \sqrt{\epsilon'_k} \right)^2 \sqrt{\epsilon'_{k-1}} + 4\sqrt{\epsilon_{k-1}} \right]^2 (k + 1)^2; \quad \text{and}
\]

\[
\|\text{DiffL}_{k}^{\text{up}}\|^2 \leq \left[ 2\sqrt{\epsilon'_k + \epsilon'_k + 2\epsilon_k + 4\sqrt{\epsilon_k}} \right]^2 (k + 2)^2,
\]

*and there exists a unitary matrix* \(O \in \mathbb{R}^\beta_k \times \beta_k\) *such that*

\[
\|Y_{N_k}: - \hat{Y}_{N_k}:O\|_F^2 \leq \frac{8\beta_k \left[ \|\text{DiffL}_{k}^{\text{down}}\|^2 + \|\text{DiffL}_{k}^{\text{up}}\|^2 \right]}{\min\{\delta_1, \cdots, \delta_{\kappa}\}}. \quad (1)
\]

*Assu. 2:* no topology is destroyed/created

*Assu. 3:* sparsely connected

*N_k:* bound only simplexes that are not altered during connected sum
Subspace perturbation

**Theorem 1**

*Under Assumptions 1–3*

\[
\left\| \text{Diff}L_{k}^{\text{down}} \right\|^{2} \leq \left[ 2\sqrt{\epsilon_{k}'} + \epsilon_{k}' + (1 + \sqrt{\epsilon_{k}'})^{2} \sqrt{\epsilon_{k-1}'} + 4\sqrt{\epsilon_{k-1}'} \right]^{2} (k + 1)^{2}; \text{ and}
\]

\[
\left\| \text{Diff}L_{k}^{\text{up}} \right\|^{2} \leq \left[ 2\sqrt{\epsilon_{k}'} + \epsilon_{k}' + 2\epsilon_{k} + 4\sqrt{\epsilon_{k}} \right]^{2} (k + 2)^{2},
\]

and there exists a unitary matrix \( O \in \mathbb{R}^{\beta_{k} \times \beta_{k}} \) such that

\[
\left\| Y_{N_{k}}; - \hat{Y}_{N_{k}}; O \right\|_{F}^{2} \leq \frac{8\beta_{k} \left[ \left\| \text{Diff}L_{k}^{\text{down}} \right\|^{2} + \left\| \text{Diff}L_{k}^{\text{up}} \right\|^{2} \right]}{\min\{\delta_{1}, \ldots, \delta_{\kappa}\}}. \quad (1)
\]

- **Assu. 2:** no topology is destroyed/created
- **Assu. 3:** sparsely connected
- **\( N_{k} \):** bound only simplexes that are not altered during connected sum
Theorem 1
Under Assumptions 1–3

\[ \| \text{DiffL}_k^{\text{down}} \|^2 \leq \left[ 2\sqrt{\epsilon_k' + \epsilon_k'} + \left( 1 + \sqrt{\epsilon_k'} \right)^2 \sqrt{\epsilon_{k-1}' + 4\sqrt{\epsilon_{k-1}'}} \right]^2 (k + 1)^2; \text{ and} \]

\[ \| \text{DiffL}_k^{\text{up}} \|^2 \leq \left[ 2\sqrt{\epsilon_k' + \epsilon_k'} + 2\epsilon_k + 4\sqrt{\epsilon_k} \right]^2 (k + 2)^2, \]

and there exists a unitary matrix \( O \in \mathbb{R}^{\beta_k \times \beta_k} \) such that

\[ \left\| \mathbf{Y}_{N_k} - \hat{\mathbf{Y}}_{N_k} : O \right\|_F^2 \leq \frac{8\beta_k \left[ \| \text{DiffL}_k^{\text{down}} \|^2 + \| \text{DiffL}_k^{\text{up}} \|^2 \right]}{\min\{\delta_1, \cdots, \delta_\kappa\}}. \]  (1)

▶ **Assu. 2:** no topology is destroyed/created

▶ **Assu. 3:** sparsely connected

▶ **\( N_k \):** bound only simplexes that are not altered during connected sum
Harmonic Embedding Spectral Decomposition Algorithm

In Simplicial complex $(V, E, T)$, weights $W_V, W_E, W_T$

1. Compute $\mathcal{L}_1$
2. Eigendecomposition

$$\beta_1, Y \leftarrow \text{Null}(\mathcal{L}_1)$$

3. Independent Component Analysis

$$Z \leftarrow \text{ICAnoprewhite}(Y)$$

Out $Z$
Outline

Manifold coordinates with Scientific meaning

Machine Learning 1-Laplacians, topology, vector fields
  1-Laplacian $\Delta_1(\mathcal{M})$ estimation from samples
  Analysis of vector fields – Helmholtz-Hodge decomposition
  Harmonic Embedding Spectral Decomposition Algorithm
  Spectral Shortest Homologous Loop Detection
Spectral Shortest Homologous Loop Detection

\[ Z = [z_1, \ldots z_{\beta_1}], (V, E), \text{ edge lengths } d_E \]
for \( l = 1 : \beta_1 \)

1. Remove edges \( e \) with low \( |Z_{le}| \), keep top \( 1/\beta_1 \) fraction \( E_{\text{keep}} \)
2. Construct \( G_l = (V, E_{\text{keep}}) \), edge weights \( d_E \)
3. Repeat for a lot of edges in \( E_{\text{keep}} \)
   3.1 select \( e = (t, s_0) \in E_{\text{keep}} \)
   3.2 find shortest path \( s_0 \) to \( t \)
      \( P_e \leftarrow \text{DIJKSTRA}(V, E_{\text{keep}} \setminus \{e\}, s_0, t, d_E) \)
4. \( C_l \leftarrow \text{argmin}_e \text{ length}(\text{loop}(P_e)) \)

Out loops \( C_{1:\beta_1} \)
Shortest loop basis on real data

RNA single cell  sculpture  ocean buoys  retina
Summary – Manifold Learning beyond embedding algorithm

- Manifolds, vector fields, …
  - historically used for modeling scientific data
  - represented analytically
  **NOW** representations learned from data
    - machine learning needs to handle new mathematical concepts
    - need to output results in scientific language

- Generic method for Interpretation in the language of the domain
  - by finding coordinates from among domain-specific functions
  - non-parametric and non-linear

- Extended manifold learning from scalar functions to vector fields
  - first 1-Laplacian estimator
  - continuous limit derived
  - natural extensions of smoothing, semi-supervised learning to vector field data
  - perturbation result for prime manifold decomposition
  - algorithm for shortest loop basis
Summary – Manifold Learning beyond embedding algorithm

- Manifolds, vector fields, ...
  - historically used for modeling scientific data
  - represented analytically
  NOW representations learned from data
  • machine learning needs to handle new mathematical concepts
  • need to output results in scientific language

- Generic method for Interpretation in the language of the domain
  - by finding coordinates from among domain-specific functions
  - non-parametric and non-linear

- Extended manifold learning from scalar functions to vector fields
  - first 1-Laplacian estimator
  - continuous limit derived
  - natural extensions of smoothing, semi-supervised learning to vector field data
  - perturbation result for prime manifold decomposition
  - algorithm for shortest loop basis
Summary – Manifold Learning beyond embedding algorithm

- Manifolds, vector fields, ...
  - historically used for modeling scientific data
  - represented analytically

NOW representations learned from data
  - machine learning needs to handle new mathematical concepts
  - need to output results in scientific language

- Generic method for Interpretation in the language of the domain
  - by finding coordinates from among domain-specific functions
  - non-parametric and non-linear

- Extended manifold learning from scalar functions to vector fields
  - first 1-Laplacian estimator
  - continuous limit derived
  - natural extensions of smoothing, semi-supervised learning to vector field data
  - perturbation result for prime manifold decomposition
  - algorithm for shortest loop basis
Samson Koelle, Yu-Chia Chen, Hanyu Zhang, Alon Milchgrub

Hugh Hillhouse (UW), Jim Pfaendtner (UW), Chris Fu (UW)
Stefan Chmiela, Cecilia Clementi, John Maddocks, Klaus Müller, Frank Noé, Christof Schütte
Alexandre Tkatchenko, Alvaro Vasquez-Mayagoitia

Thank you