Understanding Data from Incomplete Inter-Point Distance via Locally Low-rank matrix completion and Geometric PDEs

Rongjie Lai
Rensselaer Polytechnic Institute

Joint with Jia Li, Abiy Tasissa@ RPI
Motivation and problems

- Global configuration from incomplete information

  A network of sensors collaboratively measuring some quantity, the distance matrix is noisy and incomplete

- Protein Structuring

  The nuclear magnetic resonance (NMR) spectroscopy provides distance between pairs of hydrogen atoms in a protein.

  The measure is incomplete and noisy.

  Can we understand the protein structure based on the incomplete measurement without reconstruction?
Motivation and problems

**Assumption:** Given an incomplete distance matrix $D = (d_{ij}^2)$, where $d_{ij} = dist(x_i, x_j)$ for a given point set $\{x_1, \cdots, x_n\}$ sampled on a manifold $\mathcal{M} \subset R^D$. 

![Complete distance](image1.png) ![1%](image2.png)
Motivation and problems
Besides visualization, what geometric information can we have for data **without** having global coordinate reconstruction?

- Global Recognition
- Intrinsic distance
- Classification
- Intrinsic comparison
Challenges

Unlike signals or images

- No global coordinates, only inter-point distance information is given and is with possible missing values and noise.

- No natural or good global parametrization that reveals intrinsic dimensionality and global structure.

- Highly unstructured geometric object in high dimension, difficult to analysis, organize, …. No natural basis for representation.
Our strategy

- Coherent structure inspires us to model distance data for points sampled from manifolds, where local structure can be extracted.

- Geometric PDEs on manifolds can be useful to "connect the dots" and reveal global structure and provide geometric understanding of data and the underlying manifolds.

No global coordinates reconstruction

Local Recon using low-rank matrix completion → Solve geometric PDEs based on local recon → Conduct understanding based on solutions of PDEs
Euclidean distance geometry (EDG) problem and matrix completion

\[ D = (d_{ij}^2)_{n \times n} \text{ Euclidean} \]

EDG

\[ \|x_i - x_j\|_2^2 = d_{ij}^2 \]

\[ B = \sum_i \lambda_i v_i v_i^T, \quad X = \{\sqrt{\lambda_i} v_i\} \]

\[ H = I - \frac{1}{n} \mathbf{1}\mathbf{1}^T \]

Gram matrix \( B \succeq 0 \)

\[ x_1, x_2, \ldots, x_n \in \mathbb{R}^p \]
Let write $\Omega \subset \{(i, j) \mid i, j = 1, \cdots, n\}$ as the index set for available values of $D$. We consider the following matrix completion model to recover the Gram matrix $B$ based on matrix completion theory [Candes-Recht’09, Recht-Fazel-Parrilo’10]

$$
\min_{B \succeq 0} \|B\|_* \quad \text{s.t.} \quad \left\{ \begin{array}{l}
b_{ii} + b_{jj} - 2b_{ij} = d_{ij}^2, \quad (i, j) \in \Omega \\
\sum_j B_{i,j} = 0, \quad \forall 1 \leq i \leq n
\end{array} \right.
$$

- Instead of reconstructing $D$, we consider to reconstruct $B$ as it has lower rank.
- The constraint $\sum_j B_{i,j} = 0$ is to remove possible ambiguity due to translation.
• Let $S = \{X = X^T, X1 = 0\}$, we rewrite the EDG relation $B_{i,i} + B_{jj} - B_{ij} - B_{ji} = D_{ij}, \sum_j B_{ij} = 0$ under an appropriate basis \{ $w_{ij} = e_{ii} + e_{jj} - e_{ij} - e_{ji} \mid i > j$ \}.

• The dual basis of $w_{ij}$ can be written as $v_{ij} = \sum_{kl} H_{ij,kl} w_{kl}$ with $H_{ij,kl} = \langle w_{ij}, w_{kl} \rangle$. Any $X \in S$ can be written as $X = \sum_{ij} \langle X, , w_{ij} \rangle v_{ij}$.

• Define $\mathcal{R}_\Omega(X) = \frac{L}{m} \sum_{(ij) \in \Omega} \langle X, , w_{ij} \rangle v_{ij}$, then the EDG nuclear minimization problem as

$$\text{minimize} \quad \|B\| \quad \text{s.t.} \quad \mathcal{R}_\Omega(B) = \mathcal{R}_\Omega(B^T)$$

subject to $B \in S \cap \{X \geq 0\}$.
Local coordinate reconstruction via Non-orthogonal basis sensing

1. Entry sensing is under a special orthonormal basis \( \{e_{ij}\} \) [Candes-Recht’09].

2. Sensing under general orthonormal basis is consider in [Gross’11]

3. Restricted isometry property (RIP) is considered for general linear constraint [Recht-Fazel-Parrilo’10], but hard to check the RIP condition.

**Definition 1.** The \( n \times n \) matrix \( B_T \) has coherence \( \nu \) with respect to basis \( \{w_\alpha\}_{n=1}^L \) and \( \{v_\alpha\}_{n=1}^L \) if the following estimates hold

\[
\max_{ij} \| \mathcal{P}_T w_{ij} \|_F^2 \leq 8 \nu \frac{r}{n}, \quad \text{and} \quad \max_{ij} \| \mathcal{P}_T v_{ij} \|_F^2 \leq 32 \nu \frac{r}{n}
\]

where \( \mathcal{P}_T \) is the projection to the tangent space \( \mathcal{T} = \{UP + QU^T\} \) of the rank \( r \) manifold at \( B_T = UD_\nu U^T \).

**Theorem 1.** If \( |\Omega| \geq O \left( r n \nu (1 + \beta) \log^2 n \right) \), for \( \beta > 1 \), the solution to the above problem is unique and equal to \( B_T \) with probability at least \( 1 - n^{-\beta} \).
Define $A : \mathbb{R}^{n \times n} \to \mathbb{R}^{\Omega} \times \mathbb{R}^{n} : B \mapsto \left( \{ b_{ii} + b_{jj} - 2b_{ij} \}_{(i,j)\in\Omega}, \sum_j B_{i,j} \right)$, and write $\tilde{A}(B) = (\mathcal{P}_\Omega A(B), \sum_j B_{i,j})$ and $D = (\{ d_{ij}^2 \}_{(i,j)\in\Omega}, 0)$. By introducing an auxiliary variable $C = B$, we have the following equivalent version

$$\min_{B,C \succeq 0} \text{Tr}(B), \quad \text{s.t.} \quad AB - D = 0, \quad B = C$$

which can be iteratively solved by

$$
\begin{cases}
B^{k+1} = \arg \min_B \text{Tr}(B) + \frac{\mu_1}{2} \| AB - D + H_1^k \|_2^2 + \frac{\mu_2}{2} \| B - C_k + H_2^k \|_F^2 \\
C^{k+1} = \arg \min_{C \succeq 0} \frac{\mu_2}{2} \| B^{k+1} - C + H_2^k \|_F^2, \\
H_1^{k+1} = H_1^k + \mu_1 (D - AB), \\
H_2^{k+1} = H_2^k + \mu_2 (C - B),
\end{cases}
$$

- Convergence of the above algorithm can be theoretically validated as the problem is convex.
- The most time consumption step is to compute the first $k$ eigen-decomposition with scale $O(n^2k)$.
Examples for coordinates reconstruction via matrix completion

![Examples for coordinates reconstruction via matrix completion](image)

Top: 2%, bottom 3%.

Table 1: Rate of the successful reconstruction $\rho$ and the average relative error $E_B$ out of 50 tests.

<table>
<thead>
<tr>
<th>Data</th>
<th>$\gamma$</th>
<th>1%</th>
<th>2%</th>
<th>3%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^2$</td>
<td>$E_B$</td>
<td>7.157E-1</td>
<td>1.376E-3</td>
<td>4.791E-4</td>
<td>2.474E-4</td>
<td>1.342E-5</td>
<td>4.262E-5</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0%</td>
<td>92%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Cow</td>
<td>$E_B$</td>
<td>4.9427E-5</td>
<td>3.980E-4</td>
<td>1.837E-4</td>
<td>5.319E-5</td>
<td>1.4072E-5</td>
<td>2.155E-5</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Swiss roll</td>
<td>$E_B$</td>
<td>2.722E-4</td>
<td>2.894E-4</td>
<td>1.633E-4</td>
<td>5.054E-5</td>
<td>1.704E-5</td>
<td>1.114E-5</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Laplace-Beltrami operator: A Bridge from Local to Global

Given a d-dimensional manifold \((M, g)\),

\[-\triangle_M \phi_n = -\frac{1}{\sqrt{G}} \frac{\partial}{\partial x_i} (\sqrt{G} g^{ij} \frac{\partial \phi}{\partial x_j}) = \lambda_n \phi_n, \ n = 0, 1, 2, \cdots\]

- Intrinsicness. Invariant under isometric deformation \(M\).
- Inverse spectrum problem.

\[
Z(t) = \int_M \sum_i e^{-\lambda_i t} \phi_i(x) \phi_i(y) dv = \frac{1}{4\pi t} \left( \sum_{i=0}^{\infty} c_i t^{i/2} \right)
\]

where \(c_0 = \text{area}(M)\), \(c_1 = -\frac{\sqrt{\pi}}{2} \text{length}(B)\), \(c_2 = \frac{1}{3} \int_M K - \frac{1}{6} \int_B J\).

[McKean-Singer’67]

- Generically, LB eigenfunctions are Morse functions [Uhlenbeck’76].
- LB eigenvalues + LB eigenfunctions uniquely fix a manifold up to isometry.
  [Perard-Besson-Gassot’94]
- Estimation of LB eigenvalues and the geometry of underlying manifolds.
Local tangent space approximation

- K-nearest neighborhood (KNN)
- Local principle component analysis (PCA)

\[ P_i = \sum_{k \in N(i)} (p_k - c_i)^T (p_k - c_i) \]

a local coordinate system \( \langle p_i; e^i_1, e^i_2, e^i_3 \rangle \) at each point, where eigenvectors \( (e^i_1, e^i_2, e^i_3) \) of \( P_i \) form an orthogonal frame associated with eigenvalues \( (\lambda^i_1, \lambda^i_2, \lambda^i_3) \) with \( \lambda^i_1 \geq \lambda^i_2 \gg \lambda^i_3 \geq 0.\)
Local manifold approximation using moving least square method

- KNN of $p_i$ have local coordinates $(x^i_k, y^i_k, z^i_k)$

- Local manifold approximation. Find a local degree two bivariate polynomial $z_i(x, y)$

$$\sum_{k \in N(i)} w(\|p_k - p_i\|) (z_i(x^i_k, y^i_k) - z^i_k)^2 \longrightarrow \Gamma_i = (x, y, z_i(x, y)) \& \text{Metric } g$$

- Local function approximation. $\min_{f_{\overline{x}} \in \Pi^d_m} \sum_{k=1}^K w(\|x_k - \overline{x}\|) \|f_{\overline{x}}(x_k) - f_k\|^2$

where $f_{\overline{x}}(x) = b(x)^T c(\overline{x}) = b(x) \cdot c(\overline{x})$ and $b(x)$ is the polynomial basis vector.

$$w(d) = \begin{cases} 1 & \text{if } d = 0 \\ 1/k & \text{if } d \neq 0 \end{cases}, \quad w(d) = \exp(-d^2/\sigma), \quad w(d) = (1 - d/D)^{4d/(4d + 1)}$$

- In the local coordinate system, $M$ and $f$ are well defined function.

$$\nabla_M f = \sum_{i,j=1}^d g^{ij} \frac{\partial f}{\partial x_i} \partial x_j, \quad \text{Div}_M V = \frac{1}{\sqrt{G}} \sum_i \frac{\partial}{\partial x_i} (\sqrt{G} v^i), \quad \Delta_M \phi = \frac{1}{\sqrt{G}} \sum_{i,j=1}^d \frac{\partial}{\partial x_i} (\sqrt{G} g^{ij} \frac{\partial \phi}{\partial x_j})$$

**Theorem.** Assume $f \in C^{m+1}(R^d)$, $I \leq K$, $w(\cdot) > 0$. Let $h = \max_k \|x_k - \overline{x}\|$ then

$$\left| c_i - \frac{1}{\alpha_i!} D^{\alpha_i} f(\overline{x}) \right| = C h^{m+1-|\alpha_i|}$$

where $C$ is a constant depends on $w$, $f$ and $\alpha_i$. (Liang-Zhao, natural extension of results in [Levin’98, Lipman et al’06])
Local Mesh Method [Lai-Liang-Zhao]

1. K nearest neighbor (KNN)

2. Local principal component analysis (PCA) on KNN

\[ P_i = \sum_{k \in N(i)} (\mathbf{p}_k - \mathbf{c}_i)^T (\mathbf{p}_k - \mathbf{c}_i) \]

3. Projection on tangent planes.

4. Inherit triangle structure from the tangent space.

With the local connectivity \( \{ p_i; \mathcal{V}(i), \mathcal{R}(i) \} \), we have:

\[ \nabla_{\mathcal{P}} f(p_i) \approx \frac{1}{W} \sum_{T \in \mathcal{R}(i)} \text{Area}(T) \nabla_T f(p_i), \quad \text{div}_{\mathcal{P}} \mathbf{V}(p_i) \approx \frac{1}{W} \sum_{T \in \mathcal{R}(i)} \text{Area}(T) \text{div}_T \mathbf{V}(p_i) \]

- Only use the first ring structure, more accurate approximation can be obtained using the second ring structure.

- Alternatively, we can also combine the local mesh with the moving least square approximation to obtain better approximation.
Solving PDEs on data represented by incomplete inter-point distance

Key features of our methods:

- No global coordinates or global parameterization is needed.

- Only local information such as K nearest neighbors are needs, which can be reconstructed by matrix completion.

- Our methods works for points sampled from manifolds with representation by incomplete distance. It naturally can work with data in any dimensions and co-dimension.
Solve LB eigenvalue problem based on distance of points on unit sphere

Figure 1: LB eigenfunctions corresponding to $\lambda = 2, 6, 12$ from 80% local distance (1002 points).

<table>
<thead>
<tr>
<th>sample size</th>
<th>1002</th>
<th>1962</th>
<th>4002</th>
<th>7842</th>
<th>16002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise free ($K = 9$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 20$</td>
<td>0.0469</td>
<td>0.0280</td>
<td>0.0175</td>
<td>0.0108</td>
<td>0.0046</td>
</tr>
<tr>
<td>$\lambda = 72$</td>
<td>0.1292</td>
<td>0.0720</td>
<td>0.0420</td>
<td>0.0256</td>
<td>0.0161</td>
</tr>
<tr>
<td>Noise free ($K = 18$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 20$</td>
<td>0.0482</td>
<td>0.0250</td>
<td>0.0126</td>
<td>0.0065</td>
<td>0.0032</td>
</tr>
<tr>
<td>$\lambda = 72$</td>
<td>0.3643</td>
<td>0.1178</td>
<td>0.0614</td>
<td>0.0328</td>
<td>0.0174</td>
</tr>
<tr>
<td>local distance with Gaussian error of $\sigma = 5% \cdot d_{\text{max}}$ ($K = 18$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 20$</td>
<td>0.0469</td>
<td>0.0216</td>
<td>0.0133</td>
<td>0.0081</td>
<td>0.0043</td>
</tr>
<tr>
<td>$\lambda = 72$</td>
<td>0.3624</td>
<td>0.1123</td>
<td>0.0625</td>
<td>0.0280</td>
<td>0.0187</td>
</tr>
<tr>
<td>Noise free ($K = 30$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 20$</td>
<td>0.0850</td>
<td>0.0454</td>
<td>0.0224</td>
<td>0.0115</td>
<td>0.0057</td>
</tr>
<tr>
<td>$\lambda = 72$</td>
<td>0.6146</td>
<td>0.3452</td>
<td>0.1041</td>
<td>0.0563</td>
<td>0.0283</td>
</tr>
<tr>
<td>local distance with Gaussian error of $\sigma = 10% \cdot d_{\text{max}}$ ($K = 30$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 20$</td>
<td>0.1023</td>
<td>0.0619</td>
<td>0.0393</td>
<td>0.0274</td>
<td>0.0232</td>
</tr>
<tr>
<td>$\lambda = 72$</td>
<td>0.6248</td>
<td>0.3653</td>
<td>0.1147</td>
<td>0.0668</td>
<td>0.0395</td>
</tr>
</tbody>
</table>

Table 1: $E_{\text{max}}$ errors for Gaussian perturbed distance of uniformed distributed point clouds of a unit sphere. Assume the inaccurate distance comes from 80% of distance information and corrupted by some type of noise.

Figure 1: $E_{\text{max}}(\lambda = 20)$ for Gaussian perturbation corrupted distance of uniformed distributed point clouds on the unit sphere.
Local vs. Global: Time consumption comparisons

\[ O(n^2m) \text{ vs. } O(nl^2m) \]

<table>
<thead>
<tr>
<th>number of points</th>
<th>1002</th>
<th>1962</th>
<th>4002</th>
<th>7842</th>
<th>16002</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma = 100% ), ( l = 6 ), available distance = ( \gamma l/n )</td>
<td>( \gamma = 80% ), ( l = 9 ), available distance = ( \gamma l/n )</td>
<td>( \gamma = 50% ), ( l = 18 ), available distance = ( \gamma l/n )</td>
<td>( \gamma = 30% ), ( l = 30 ), available distance = ( \gamma l/n )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.26</td>
<td>0.51</td>
<td>1.01</td>
<td>2.03</td>
<td>4.05</td>
<td></td>
</tr>
<tr>
<td>2.28</td>
<td>5.60</td>
<td>11.17</td>
<td>22.28</td>
<td>45.02</td>
<td></td>
</tr>
<tr>
<td>4.03</td>
<td>8.09</td>
<td>16.14</td>
<td>32.44</td>
<td>64.71</td>
<td></td>
</tr>
<tr>
<td>15.13</td>
<td>30.19</td>
<td>60.42</td>
<td>120.95</td>
<td>241.63</td>
<td></td>
</tr>
</tbody>
</table>

**Global** reconstruction using 3% distance (\( l = 6 \) for MLS)

| 2.09 | 9.86 | 40.13 | 154.40 | 597.06 |

Table 1: Comparisons of time consumption (minutes) of solving the LB eigenvalue problem based on local/global reconstruction methods.
Solve LB eigenvalues for distance data from high-D manifolds

LB eigenkproblem for a 2 dimensional flat torus in $\mathbb{R}^4$.

Figure 1: a 2D torus $\mathbb{R}^4$ with 2500 points. Bottom: Relative errors. Top: full distance. Bottom: 60% distance. Left: The largest 4 eigenvalues of inner-product matrix. Middle: Relative error for the first 100 eigenvalues. Right: Convergence curves.
Solve LB eigenvalues for distance data from high-D manifolds

LB eigenkproblem for a 3 dimensional flat torus in $\mathbb{R}^6$.

Figure 1: a 3D torus $\mathbb{R}^6$ with 12167 points. Bottom: Relative errors. Top: full distance. Bottom: 60% distance. Left: The largest 4 eigenvalues of inner-product matrix. Middle: Relative error for the fist 100 eigenvalues. Right: Convergence curves.
More examples for solving LB eigenvalue problem

1st and 2nd LB eigenvalues based on 50% local distance matrix.
(For the first 2, mesh is only used for visualization)
A non diffusion type: Eikonal equation from distance

The Eikonal equation for the distance map on $\mathcal{M}$:

$$\begin{cases} 
|\nabla_{\mathcal{M}} d(x)| = 1 \\
\quad d(x) = 0, \quad x \in \Gamma \subset \mathcal{M}
\end{cases}$$

<table>
<thead>
<tr>
<th>Uniform sampling on $S^2$</th>
<th>(\text{sample size})</th>
<th>1002</th>
<th>1962</th>
<th>4002</th>
<th>7842</th>
<th>16002</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{Dijkstra})</td>
<td></td>
<td>0.008615</td>
<td>0.008606</td>
<td>0.008296</td>
<td>0.010642</td>
<td>0.011501</td>
</tr>
<tr>
<td>(\text{our method})</td>
<td></td>
<td>0.008100</td>
<td>0.005890</td>
<td>0.004110</td>
<td>0.002877</td>
<td>0.002158</td>
</tr>
<tr>
<td>Non-uniform sampling on $S^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{Dijkstra})</td>
<td></td>
<td>0.011209</td>
<td>0.016090</td>
<td>0.018380</td>
<td>0.016391</td>
<td>0.019953</td>
</tr>
<tr>
<td>(\text{our method})</td>
<td></td>
<td>0.012016</td>
<td>0.008792</td>
<td>0.003742</td>
<td>0.001736</td>
<td>0.002765</td>
</tr>
<tr>
<td>Uniform sampling on swiss roll</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{Dijkstra})</td>
<td></td>
<td>0.013104</td>
<td>0.021242</td>
<td>0.024560</td>
<td>0.024311</td>
<td>0.026004</td>
</tr>
<tr>
<td>(\text{our method})</td>
<td></td>
<td>0.003127</td>
<td>0.001637</td>
<td>0.001130</td>
<td>0.000783</td>
<td>0.000620</td>
</tr>
<tr>
<td>Non-Uniform sampling on swiss roll</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{Dijkstra})</td>
<td></td>
<td>0.016612</td>
<td>0.015779</td>
<td>0.014573</td>
<td>0.016587</td>
<td>0.018649</td>
</tr>
<tr>
<td>(\text{our method})</td>
<td></td>
<td>0.004754</td>
<td>0.005189</td>
<td>0.003087</td>
<td>0.005171</td>
<td>0.007246</td>
</tr>
</tbody>
</table>

Table 1: Relative error of geodesic distances from north pole to south pole reconstructed from 60% of local distances in each point’s 20 nearest neighbourhoods.
Construction of Skeleton

Reeb graph and skeleton structure obtained from LB eigenfunction $\phi$:

Quotient space: $\mathcal{M}/\sim: x \sim y \iff \phi(x) = \phi(y)$. This can be used to medical image analysis and data analysis.
Shape DNA [Reuter’06]

Eigenvalues vs. N-th eigs
Intrinsic comparisons using LB eigenmaps + optimal transportation [Lai-Zhao’16]

Rotation-Invariant sliced-Wasserstein distance for registration:

$$\text{RSWD}\left( (\mathcal{P}, \mu^P), (Q, \mu^Q) \right)^2 = \min_{R \in O(n)} \int_{S^{n-1}} \min_{\sigma \in \text{ADM}(\pi^R, \mu^P, \pi^R_\# \mu^Q)} \int_{\mathbb{R} \times \mathbb{R}} \|x - y\|^2_2 \, d\sigma(x, y) \, d\theta$$

**Theorem** (Lai-Zhao). **RSWD**(·, ·) defines a distance on the space $\mathcal{M}_n / \sim$. 

![Image of David2 to David1 correspondence](image1.png) 

![Image of TOSCA - Correspondences](image2.png) 

R. Lai@ RPI  Geo. Understanding  Data from Distance
Manifold stitching from distance

- \( P \approx \sum_{i=1}^{N} \phi_i \alpha_i \leftrightarrow \min_{\alpha} \|P - \Phi \alpha\|_F^2. \)

The coefficients \( \alpha \) do not depend on location.

- Stitching patches using LB eigenfunctions

\[
\min_{\alpha, \{R_j\}, \{b_j\}} \sum_{j=1}^{N_p} \|Q_j - \Phi_j \alpha R_j - 1b_j\|_F^2,
\]

s.t. \( R_j^\top R_j = I_d \)

**Figure 1:** 50% of local Euclidean distance. Armadillo (16519 points), Kitten (2884 points)

<table>
<thead>
<tr>
<th>methods</th>
<th>data</th>
<th>Armadillo</th>
<th>Kitten</th>
<th>Swiss roll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global recon</td>
<td>35321.84</td>
<td>1315.50</td>
<td>622.65</td>
<td></td>
</tr>
<tr>
<td>Stitching</td>
<td>760.18</td>
<td>138.76</td>
<td>199.75</td>
<td></td>
</tr>
</tbody>
</table>

**Computation**
Dimension reduction using geodesic distance

Figure 1: Top: the swiss roll surface (left) and its dimensional reduction result (right) from randomly 3% of pair-wise geodesic distance. Bottom: local and global coordinates reconstruction of the swiss roll from its 80% local geodesic distance.
Some extensions to other problems
Extension to Manifold low-rank for EDG

\[
\min_P \sum_i \text{rank}(R_{\Omega_i}(P)), \quad \text{s.t. } \mathcal{A}(PP^\top) = D.
\]

\[
\min_P \sum_i \|R_{\Omega_i}(P)\|_*, \quad \text{s.t. } \mathcal{A}(PP^\top) = D.
\]
we define the cluster functions \( \{ \phi_i(x) \} \) which is partially assigned from the training data \( S \).

\[
\phi_i(x) = \begin{cases} 
1, & L(x) = i, \\
0, & \text{otherwise}, \\
\end{cases}, \quad x \in S, \quad i = 0, 1, 2, \ldots, l.
\]

\[
\min_{\Phi} \sum_{x \in \mathcal{I}} \| (R_{\mathcal{M},x}) \Phi \|_\ast, \quad \text{s.t.} \quad P \subset \mathcal{M}, \quad \Phi(x, i) \big|_{x \in S} = \begin{cases} 
1, & L(x) = i, \\
0, & \text{otherwise}. \\
\end{cases}
\]
Semi-supervised learning: MNIST, 70K images

Figure 1: Success rate of label estimation by graph Laplacian, weighted graph Laplacian, and proposed MLR methods.
Extension to Image Processing: Patch Manifold and LDMM [Osher-Shi-Zhu’16]

\[
\min_f \ dim(M(f)) + \frac{\mu}{2} \|Af - f_0\|^2
\]

\[
\min_{f \in \mathbb{R}^{m \times n}, D, \mathcal{M} \subset \mathbb{R}^d} \frac{1}{2} \sum_{i=1}^{d} \int_{\mathcal{M}} |\nabla_{\mathcal{M}} a_i(p)|^2 \, d\mathcal{M} + \frac{\mu}{2} \|Af - f_0\|^2, \quad \text{s.t.} \quad \mathcal{P}(f) \subset \mathcal{M}
\]
A Patch manifold based low-rank regularization model [Lai-Li’17]

$$\min_{\mathcal{M} \subset \mathbb{R}^{\tau^2}, f} \sum_{x \in \mathcal{I}} \text{rank}((R_{\mathcal{M},x})(\mathcal{P}(f))), \quad \text{s.t.} \quad \mathcal{P}(f) \subset \mathcal{M}, \quad Af = g.$$ 

Inspired by matrix completion theory, we use nuclear norm to approximate rank which provides:

$$\min_{\mathcal{M} \subset \mathbb{R}^{\tau^2}, f} \sum_{x \in \mathcal{I}} \| (R_{\mathcal{M},x})(\mathcal{P}(f)) \|_* \quad \text{s.t.} \quad \mathcal{P}(f) \subset \mathcal{M}, \quad Af = g.$$ 

with diffusion:

$$\min_{\mathcal{M} \subset \mathbb{R}^{\tau^2}, f} \sum_{x \in \mathcal{I}} \| R_{\mathcal{M},x}(\mathcal{P}(f)) \|_* + \frac{\lambda}{2} \| \nabla_{\mathcal{M}} f \|_2^2, \quad \text{s.t.} \quad \mathcal{P}(f) \subset \mathcal{M}, \quad D(f) = g,$$

- Need to update manifold and f both;
- It is a non-convex problem;
- For each point, SVD is only applied to a small size matrix;
Example: Image inpainting

Figure 1: Image inpainting results of $256 \times 256$ Barbara image from 10% random available pixels using different methods.
### Image inpainting

<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>Incomplete Image</th>
<th>Fingerprint</th>
<th>Boat</th>
<th>Baboon</th>
<th>Peppers</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- **Fingerprint**: 5.04dB
- **Boat**: 5.70dB
- **Baboon**: 5.38dB
- **Peppers**: 6.02dB

**Figure 1**: Image inpainting for different images from 10% available pixels.
Image inpainting

<table>
<thead>
<tr>
<th>Method</th>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
<th>Image 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDMM</td>
<td>20.25 dB</td>
<td>25.51 dB</td>
<td>19.79 dB</td>
<td>24.58 dB</td>
</tr>
<tr>
<td>LDMM+WGL</td>
<td>20.24 dB</td>
<td>25.77 dB</td>
<td>20.05 dB</td>
<td>24.29 dB</td>
</tr>
<tr>
<td>MRL</td>
<td>19.32 dB</td>
<td>25.08 dB</td>
<td>19.43 dB</td>
<td>23.39 dB</td>
</tr>
</tbody>
</table>

R. Lai @RPI
Geo. Understanding Data from Distance
Image inpainting

Incomplete image

Wavelet model [cai-chan-shen’10]

MLR method

12.73dB
24.91dB
29.97dB
9.97dB
21.24dB
25.74dB
Super-resolution

Figure 1: Super resolution from average. Down sample rate $4 \times 4$ and $8 \times 8$. 
X-ray CT reconstruction

\[
\min_{\mathcal{M} \subset \mathbb{R}^r^2, f} \sum_{x \in \mathcal{I}} \|(R_{\mathcal{M}, x})(\mathcal{P}(f))\|_* \\
\text{s.t. } \mathcal{P}(f) \subset \mathcal{M}, \quad Af = g.
\]

where \(g_i = \int_{\ell_i} \mu(\mathbf{r})\,d\ell \approx \sum_{j=1}^{N^J} a_{ij}f_j = [Af]_i\),

Figure 1: Fan-beam imaging for a clinical X-ray scanned chest slice from 15, 30 and 60 projection views. The second row: wavelet tight frame [DongLiShen2012]. The third row: the proposed MLR based method.
Summary

- We propose to use solutions of PDEs to understand geometric structure of data represented as incomplete inter-point distance.

- We develop a systematic way of computing PDEs on distance data sampled from manifolds.

- We also propose to use solutions of geometric PDEs to conduct global analysis, examples include global skeleton extraction, parameterization construction, and multi-scaled registration.

- We also consider extensions to image processing based on manifold low-rank regularization.
Rongjie Lai, Jia Li and Abiy Tasissa, Exact Reconstruction of Distance Geometry Problem Using Low-rank Matrix Completion, preprint, 2017

Rongjie Lai and Jia Li, Manifold Based Low-rank Regularization for Image Restoration and Semi-supervised Learning, submitted, 2017

Rongjie Lai and Jia Li, Solving Partial Differential Equations on Manifolds From Incomplete Inter-Point Distanc. submitted, 2017.


Thanks for your attention!

Rongjie Lai (lair@rpi.edu)
http://www.rpi.edu/~lair/
Rensselaer Polytechnic Institute