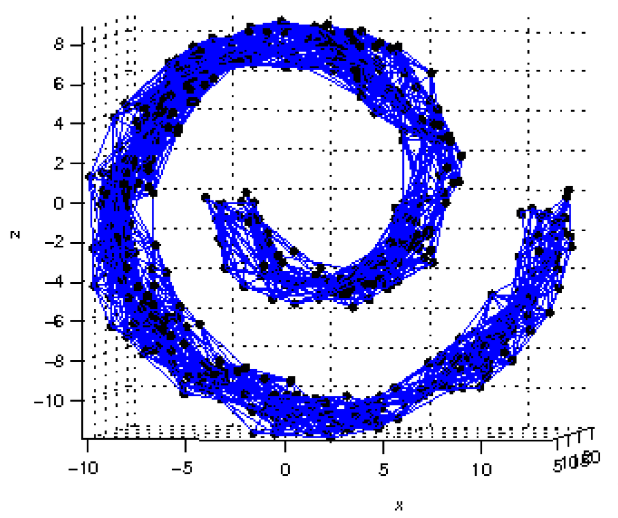
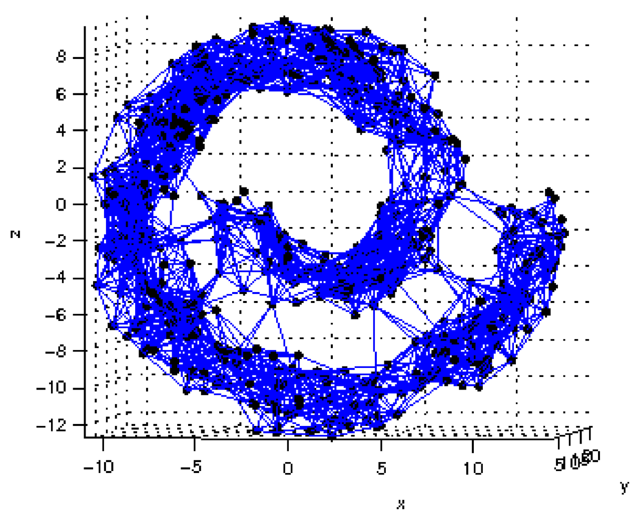
BRIDGE DETECTION AND ROBUST GEODESICS ESTIMATION VIA RANDOM WALKS

Eugene Brevdo, Peter J. Ramadge

Dept. of Electrical Engineering, Princeton Univ., Princeton, NJ USA

Problem and Intuition





Given samples from a Manifold, estimate geodesic distances. Moderate noise leads to bridges, and serious errors. We propose two new bridge detectors. Applications include Semi-Supervised Learning and Inverse Problems.

Notation and an Example Local Classifier

Initial NN Estimates (via k-NN or δ-ball) from manifold $\mathcal{M}$

- **Observed Points** $\mathcal{Y} = \{y_i = x_i + \nu_i\}_{i=1}^n$, $x_i \in \mathcal{M}$, ν_i: noise
- **NN Graph** $G = (\mathcal{Y}, \mathcal{E}, d)$. Edges are $e = (k, l) \in \mathcal{E}$
- **Edge Weight** $d_e = \|y_k - y_l\|_2 \in \mathcal{D}$ for $e \in \mathcal{E}$
- **Neighbors** $\mathcal{F}_k$: set of neighbors of y_k in G
- **Shortest Paths** $\mathcal{P}_{kl}$: minimum weight path between (y_k, y_l)
- **Bridge Estimates** $\mathcal{B} \subset \mathcal{E}$: determined by DR

Example Local Classifier: Jaccard Similarity DR (JDR)

- **JDR** Edge-neighborhood set dissimilarity index:
  $j_e = 1 - |\mathcal{F}_k \cap \mathcal{F}_l| / |\mathcal{F}_k \cup \mathcal{F}_l| \qquad e = (k, l) \in \mathcal{E}$
- **Quantile** Choose "good edge percentage" $0 < q < 1$ (e.g., 99%)
- **Classifier** $\mathcal{B}$: $e \in \mathcal{E}$ with j_e (JDR) above qth quantile

A Global Classifier: NPDR

Neighbor Probability DR (NPDR)

- **Markov Walk** Markov Walk $s(t)$ on G, transition matrix P $(n \times n)$
- **Pr{Stopping}** Stop at time $t = 0, 1, \dots$ w.p. p $(\bar{p} = 1 - p)$
- **Pr{Neighbor}** $N_{ij} = \Pr\{\textit{stopped at } y_j | s(0) = y_i\}$
  $N = p \sum_{t \geq 0} \bar{p}^t P^t = p(I - \bar{p}P)^{-1}$ [also Markov]
- **Classifier** $\mathcal{N} = \{N_e, e \in \mathcal{E}\}$
  $\mathcal{B} = \{e \in \mathcal{E} : N_e \text{ below } (1-q)^{\text{th}} \text{ quantile of } \mathcal{N}\}$
- **Intuition** Likely bridge: edge between low Pr{Neighbor} vertices

Constructing Markov Matrices P and N

1. Let $(A_\epsilon)_{lk} = \begin{cases} \exp\left(-d_{lk}^2/\epsilon\right) & (l,k) \in \mathcal{E} \text{ or } l = k \\ 0 & \text{otherwise} \end{cases}$
2. Let D_ϵ be diagonal: $(D_\epsilon)_{ii} = e_i^T A_\epsilon 1$ (row sums of A_ϵ)
3. Set $P_\epsilon = D_\epsilon{}^{-1} A_\epsilon$. Associated $N_\epsilon = p(I - \bar{p}P_\epsilon)^{-1}$.

Theorem (N_ϵ defines a heat-type diffusion operator on $\mathcal{M}$)

$$\lim_{\epsilon \downarrow 0} \lim_{n \uparrow \infty} \frac{I - N_\epsilon}{\epsilon} = c' \Delta_{\mathcal{M}}$$

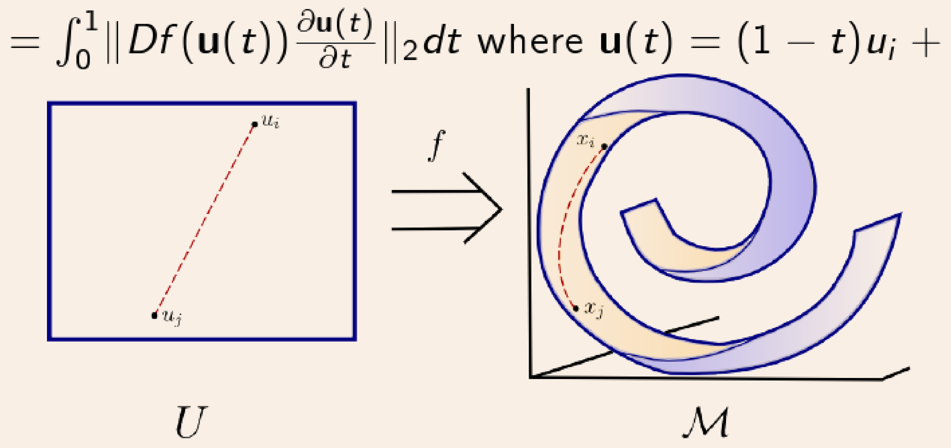
Proof Sketch:

1. *Woodbury Identity:* $\frac{I - N_\epsilon}{\epsilon} \propto \frac{I - P_\epsilon}{\epsilon}(I - \bar{p}P_\epsilon)^{-1}$
2. *Normalized Laplacian Convergence:* $\frac{I - P_\epsilon}{\epsilon} \to c\Delta_{\mathcal{M}}$ *[Lafon et al.]*

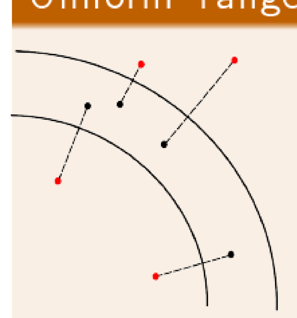
Denoising Geodesic Estimates

Swiss Roll Geodesics

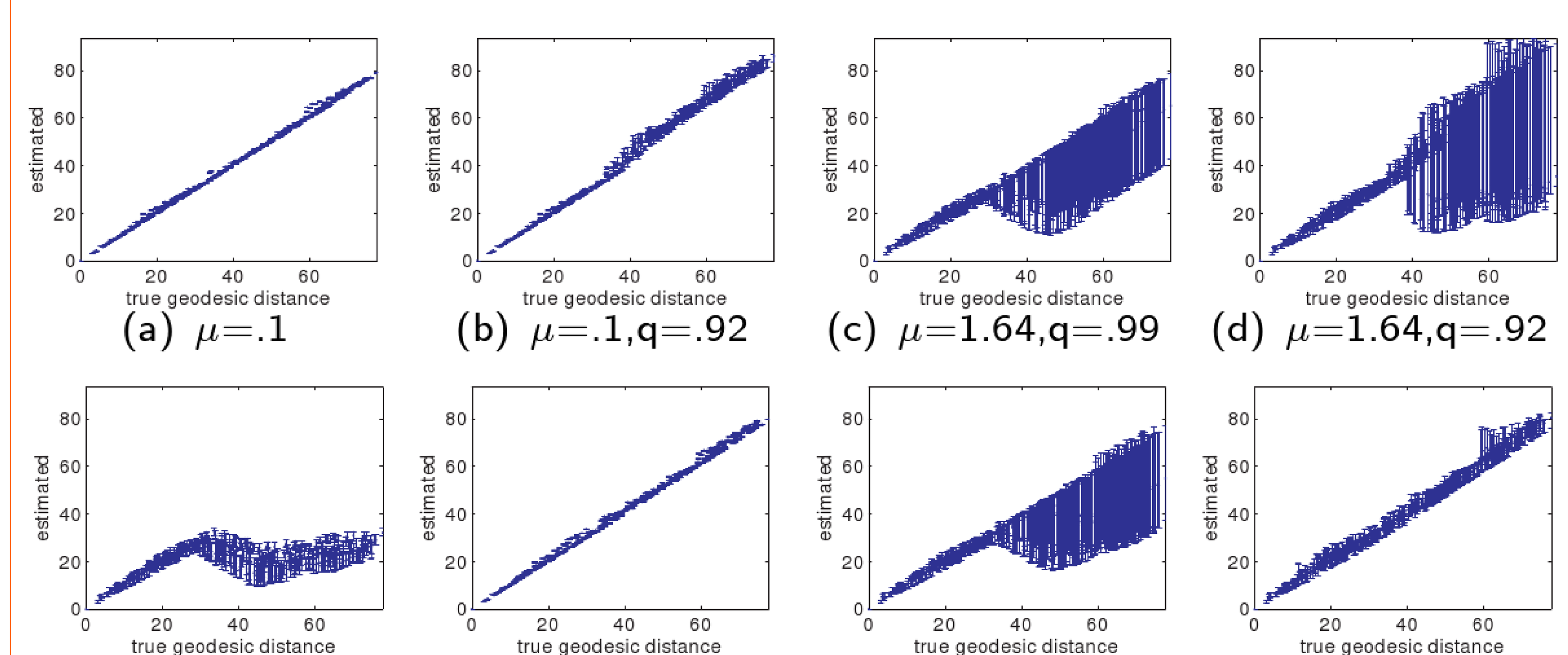
$g_{ij} = \int_0^1 \|Df(\mathbf{u}(t)) \frac{\partial \mathbf{u}(t)}{\partial t}\|_2 dt$ where $\mathbf{u}(t) = (1-t)u_i + tu_j$



Uniform Tangent Noise: $y_{ti} = x_i + \mu u_{ti} \mathbf{n}_i$



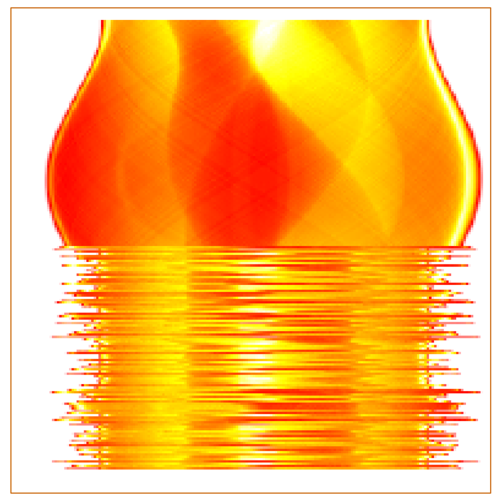
- $T = 100$ iterations, $t = 1, 2, \cdots, T$
- $\mathbf{n}_i$: normal to x_i
- u_{ti}: uniform noise on $[-1, 1]$
- Study geodesics estimates as μ increases



(a) μ=.1 (b) μ=.1,q=.92 (c) μ=1.64,q=.99 (d) μ=1.64,q=.92
(e) μ=1.64 (f) μ=.1,q=.92 (g) μ=1.64,q=.99 (h) μ=1.64,q=.92

Geodesic estimates vs. truth (from x_1). (a,e): SP, (b,c,d) ECDR, (f,g,h): NPDR

Random Projection CT[1]



Observations: y_i, $i = 1, \dots, 1000$

- $R_\theta(I)$: Radon transform of I
- $f(\theta) = R_\theta(I)$, $\theta \sim \text{Unif}[0, 2\pi]$
- $y_i = f(\theta_i) + \nu_i$ $(\nu_i \sim N(0, \sigma^2))$

Pruning Bridges (SNR: $-2db$)

Reconstruction: $\hat{I}$

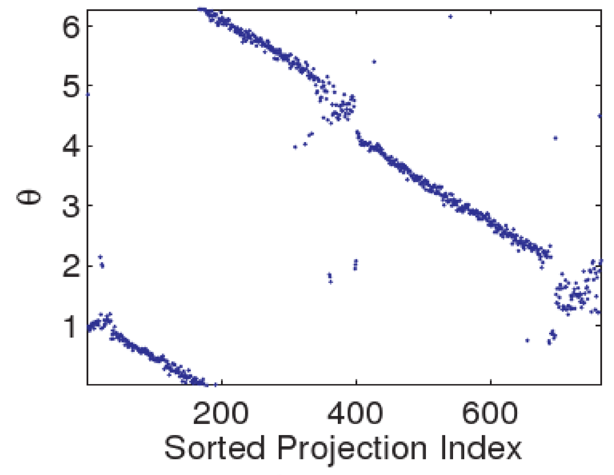
1. Preprocess data (denoise)
2. Build NN graph ($k = 50$)
3. *Prune bridges*
4. Solve eigenvalue problem: Angular ordering
5. Reconstruct $\hat{I}$ via R^{-1}

Approach of [1]: JDR

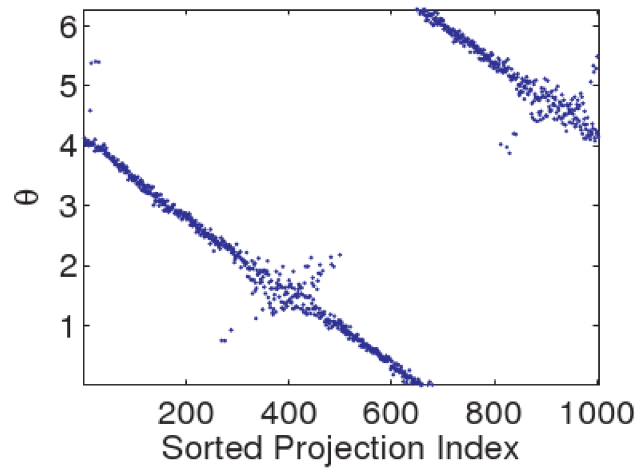
Optimal $q = .78$
(found via cross-validation)

Our approach: NPDR

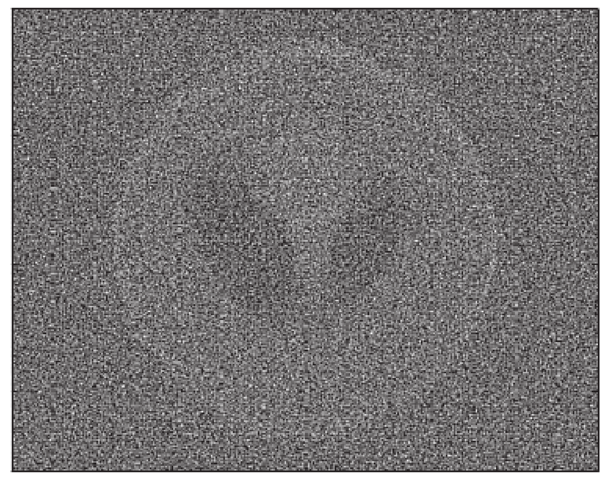
$q = .8, p = .01$, on $\hat{P}_0$
(all edges in G have weight 1)

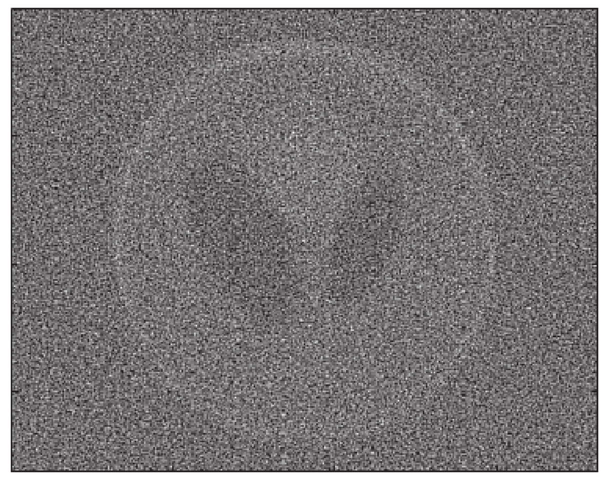


(i) $\hat{\theta}$, JDR



(j) $\hat{\theta}$, NPDR





(k) $\hat{I}$, JDR (l) $\hat{I}$, NPDR

Comparing Reconstruction Quality: 25% Improvement

Metric: $\rho = \frac{I^T \hat{I}}{\|I\| \|\hat{I}\|}$ ($\hat{I}$ aligned with I)

JDR removes 277 nodes, $\rho_{JDR} = 0.12$

NPDR removes 21 nodes, $\rho_{NPDR} = 0.15$

1. A. Singer, H. Wu, 2009