Estimating Riemannian Metric with Noise-Contaminated Intrinsic Distance

Jiaming Qiu[†], Xiongtao Dai[‡]

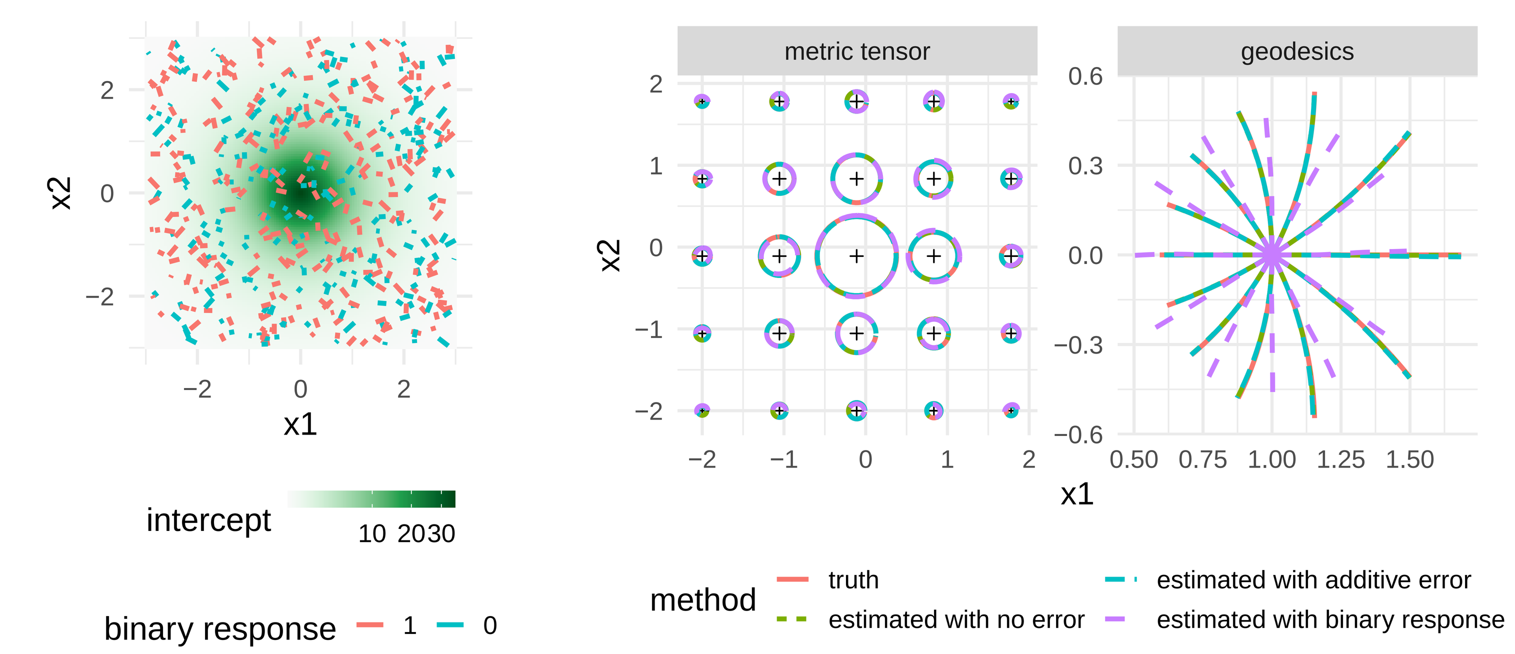
[†]Public Health Sciences, Fred Hutchinson Cancer Center, [‡]School of Public Health, University of California, Berkeley

We extend metric learning to consistently estimate Riemannian metric with noisy similarity measures.

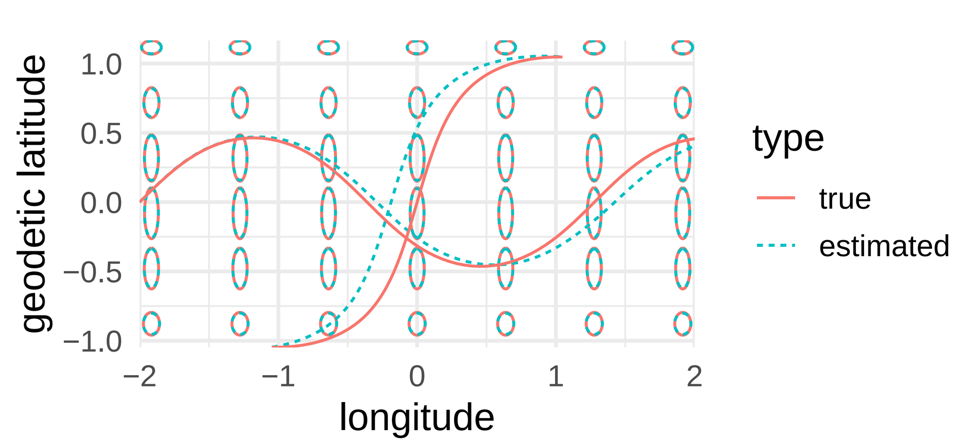
- Probabilistic modeling of pairwise similarity measures via intrinsic distances.
- Geometric interpretation for the similarity induced data space structure.
- Works for continuous, binary, or comparative similarity measures.
- Based on local regression and Taylor expansion for the squared distances.
- Justify the metric learning from a geometrical perspective.

Simulation Examples

Binary and continuous similarity on 2D sphere with stereographic coordinates.



On a 2D ellipsoid: metric and geodesics.



Binary relative comparison (X is more like Y than Z) on double spirals.

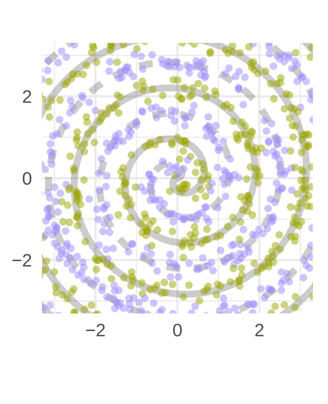


Fig. 4: The points and their latent classes

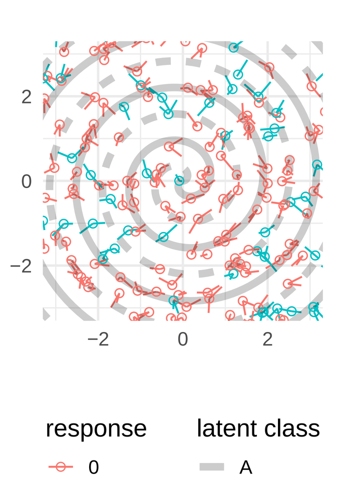


Fig. 5: Relative comparison.

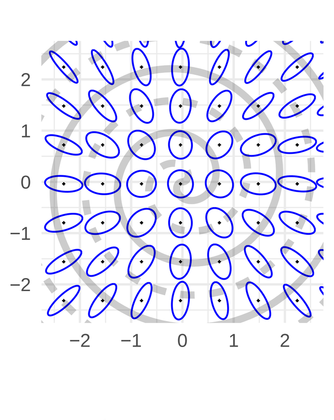


Fig. 6: Estimated metric.

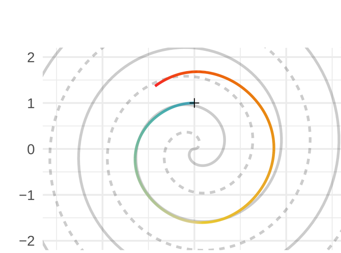


Fig. 7: Estimated geodesic.

New York Travel Time Metric via Taxi Trips

(X_{u0}, X_{u1}, Y_u) are pickup/dropoff locations, and trip duration.

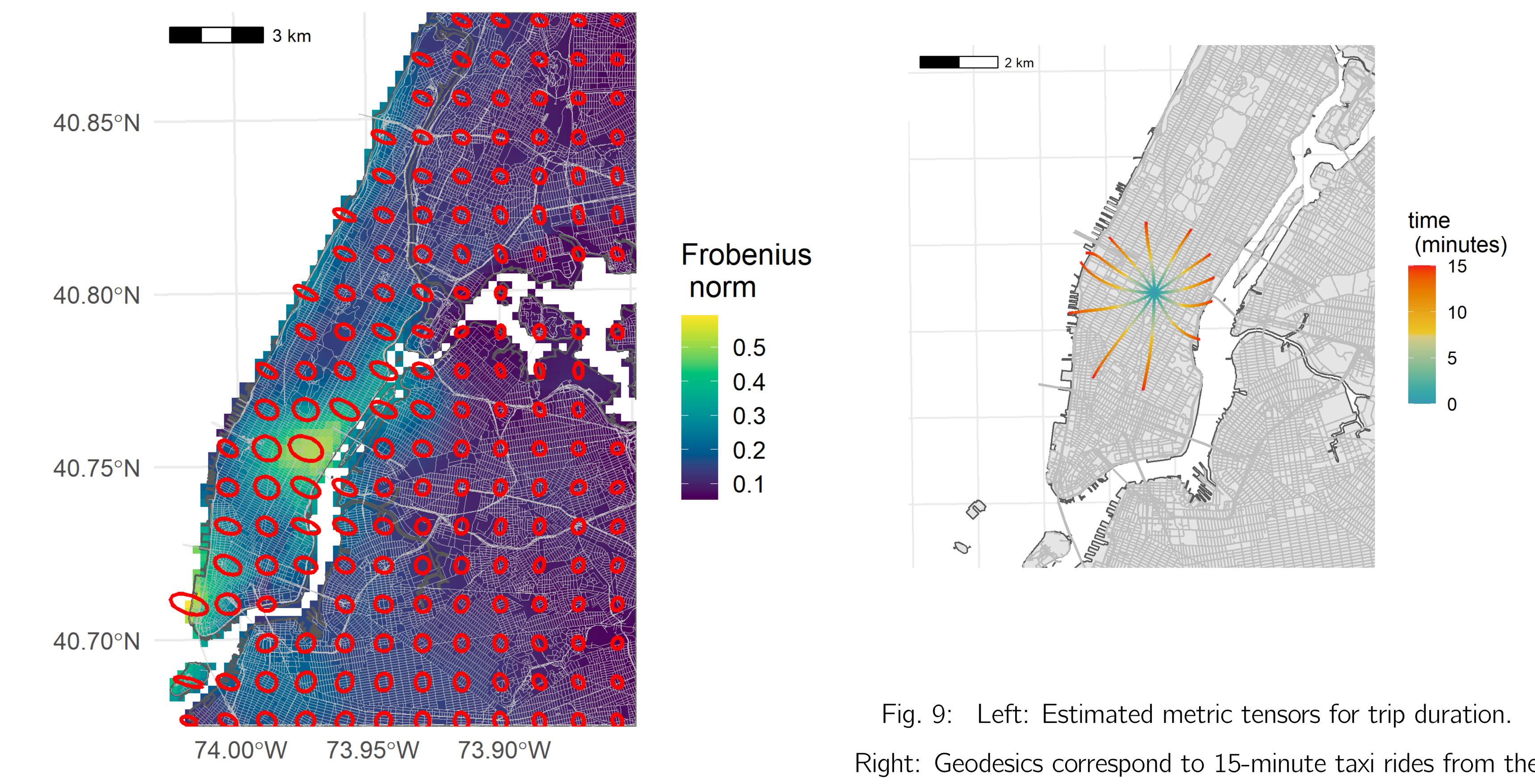


Fig. 9: Left: Estimated metric tensors for trip duration. Right: Geodesics correspond to 15-minute taxi rides from the Empire State Building.

- Traffic is slower along the longer axes of the cost ellipses. I.e., along the east–west direction (narrower streets) compared to the north–south direction (wider avenues).
- Brighter regions (midtown and the financial district) are more congested.

MNIST Digits

- Embed MNIST images to a 2-dimensional space via tSNE.
- Similarity $= C\,dist_{wass}\left(\text{pic}_{u0}, \text{pic}_{u1}\right) + \mathbb{1}_{\{\text{lbl}_{u0}\neq\text{lbl}_{u1}\}}$ for some constant C.

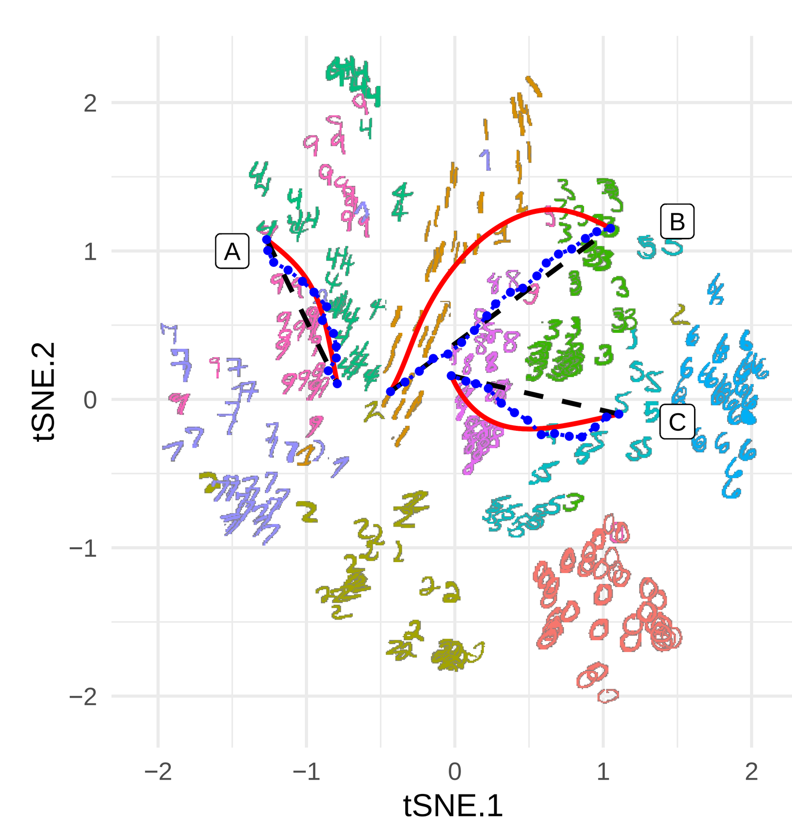


Fig. 10: The estimated geodesic curves (solid red), shortest path along kNN graph (dotted blue), and straight lines on the chart (dashed black).

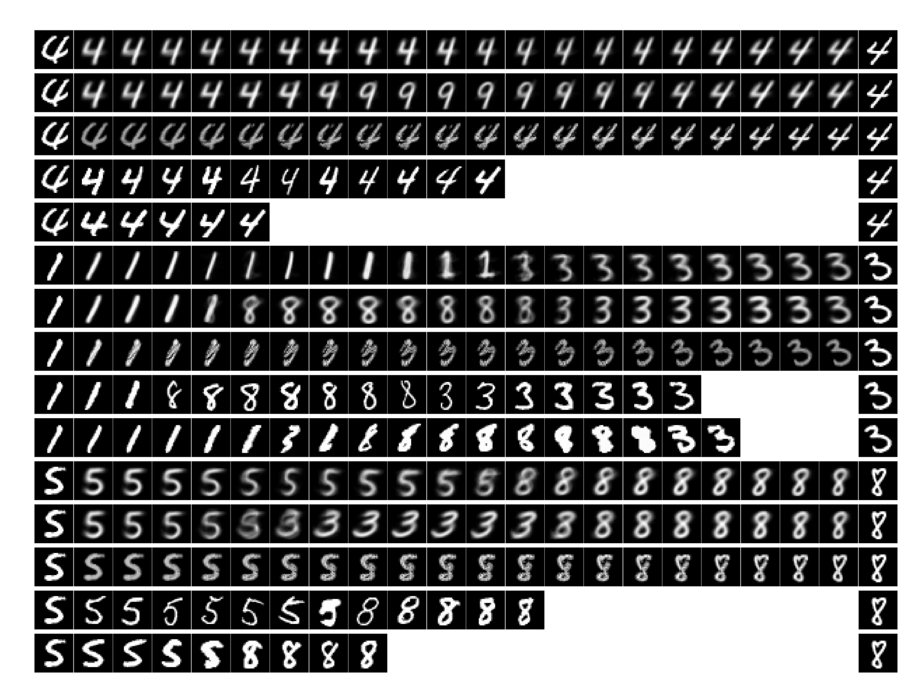


Fig. 11: Image transitions. Every 5 rows form a group. The 1st – 5th row correspond to the estimated geodesics, the straight lines on the chart, the optimal transport (path not shown), the paths along kNN graph on the tSNE chart, and along kNN graph in the Euclidean space (path not shown), respectively.

- Similar if they are the same digit and look the same.
- Estimated geodesics try to stay within same digit class and go through similar images.

A Geometric Perspective

Observe N independent triplets (Y_u, X_{u0}, X_{u1}), where for $u = 1, \dots, N$,

- X_{uj} are points on the manifold $\mathcal{M}$ with coordinates $\left(X_{uj}^1, \dots, X_{uj}^d\right) \in \mathbb{R}^d$.
- Y_u measures similarity between X_{u0} and X_{u0}.
- Suppose $\mathbb{E}\left(Y_u | X_{u0}, X_{u1}\right) = g^{-1}\left(dist\left(X_{u0}, X_{u1}\right)^2\right)$ in a small neighborhood $\mathcal{U}_p \subset \mathcal{M}$ of some target point $p \in \mathcal{M}$ with some given link function g.

To connect squared distances to metric tensor and more:

- $dist\left(X_{u0}, X_{u1}\right)^2 \approx \delta^i_{0-1}\delta^j_{0-1}G_{ij} + \delta^i_{0-1}\left(\delta^k_0\delta^l_0 - \delta^k_1\delta^l_1\right)\Gamma^j_{kl}G_{ij}$,
- $\delta^i_0 = X^i_{u0} - p^i$, $\delta^i_1 = X^i_{u1} - p^i$, and $\delta^i_{0-1} = \delta^i_0 - \delta^i_1$ are differences in coordinates;
- G_{ij} and Γ^j_{kl} are the metric tensor and Christoffel symbols at p.

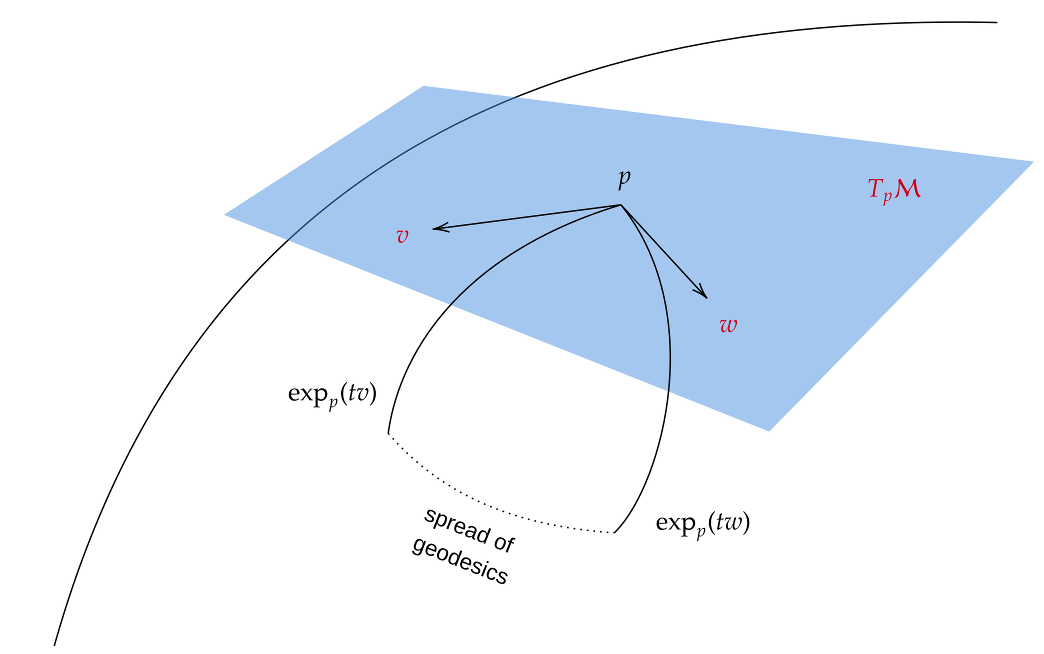


Fig. 12: Spread of geodesics. The tangent space (blue plane) and tangent vectors annotated in red.

Squared distance is approximated by a quadratic form (metric) plus a cubic form of the difference in coordinates. Higher order terms also possible (relate to curvature).

Motivates local regression estimation.

- Linear predictor: $\eta_u := \beta^{(0)} + \delta^i_{u,0-1}\delta^j_{u,0-1}\beta^{(1)}_{ij} + \delta^k_{u,0-1}\left(\delta^i_{u0}\delta^j_{u0} - \delta^i_{u1}\delta^j_{u1}\right)\beta^{(2)}_{ijk}$. The intercept $\beta^{(0)}$ is optional.
- Find

$$\left(\hat\beta^{(0)}, \hat\beta^{(1)}_{ij}, \hat\beta^{(2)}_{ijk}\right) = \underset{\beta^{(0)},\beta^{(1)}_{ij},\beta^{(2)}_{ijk};i,j,k}{\arg\min} \sum_{u=1}^{N} Q\left(Y_u,\ g^{-1}(\eta_u)\right) w_u,$$

  subject to symmetric constraints $\beta^{(1)}_{ij} = \beta^{(1)}_{ji}$, $\beta^{(2)}_{ijk} = \beta^{(2)}_{jik}$ for $i, j, k, l = 1, \dots, d$. Here Q is a loss function (e.g., likelihood), and w_u are non-negative weights.
- $\hat G_{ij} = \hat\beta^{(1)}_{ij}$, $\hat\Gamma^l_{ij} = \hat\beta^{(2)}_{ijk}\hat G^{kl}$ estimate the metric tensor and Christoffel symbols, where $\hat G^{kl}$ is the matrix inverse of $\hat G$ satisfying $\hat G^{kl}\hat G_{kj} = 1_{\{j=l\}}$.

Proposition 1 (Consistency). Suppose $g(\mu) = \mu$ and $Q(\mu, y) = (\mu - y)^2$. Under conditions similar to those in a local regression setting, bias$\left(\hat\beta | \mathbf{X}\right) = O_p\left(h^2\right)$, var$\left(\hat\beta | \mathbf{X}\right) = O_p\left(N^{-1}h^{-4-2d}\right)$, as $h \to 0$ and $Nh^{2+2d} \to \infty$, where $\mathbf{X} = \{(X_{u0}, X_{u1})\}_{u=1}^N$ is the collection of observed endpoints.

OpenReview