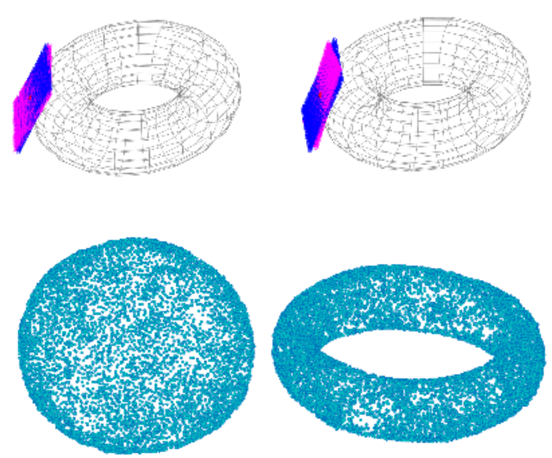
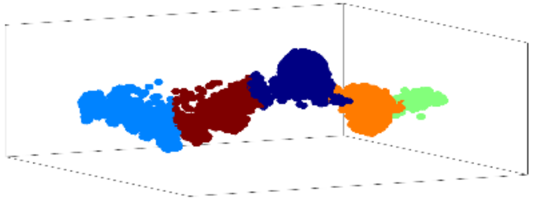
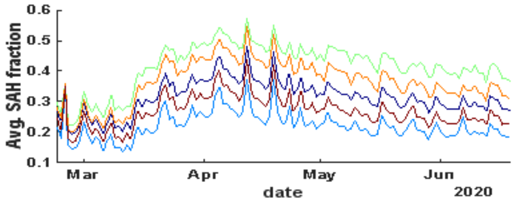
SAND2021-14144C

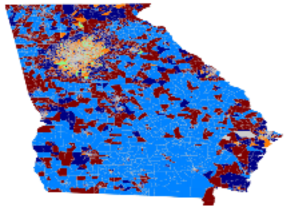
Sandia National Laboratories

Quantum-inspired manifold learning









Mohan Sarovar

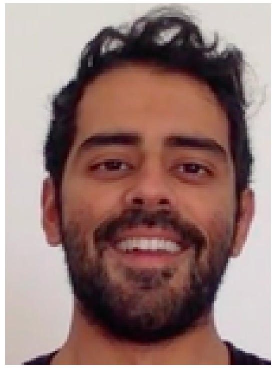
Sandia National Laboratories, Livermore, CA

Quantum Techniques in Machine Learning
November 2021

U.S. DEPARTMENT OF ENERGY
NNSA

Acknowledgements

Collaborator:



Akshat Kumar (Clarkson University)

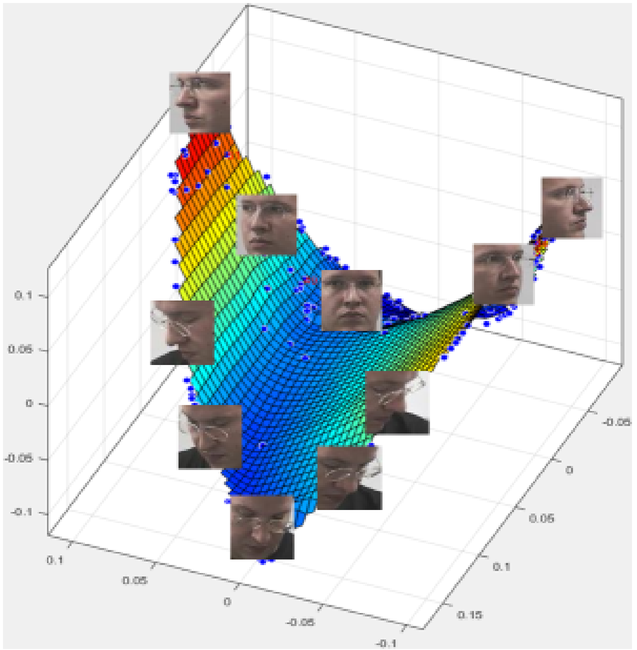
Funding:

Data and the manifold hypothesis

- Geometry can be a powerful tool in making sense of big data.

- **The manifold hypothesis**: "high dimensional data tend to lie in the vicinity of a low dimensional manifold". Fefferman, Mitter, Narayanan. J. Am. Math. Soc., **29**, 983 (2016)
  - *e.g.,* images, randomly generated image of *NxN* pixels will almost surely not correspond to a real world scene.
  - *e.g.,* data generated by a dynamical system will follow some equation of motion.

"Manifold learning" = identifying the geometry and manifold underlying the data



https://www.intechopen.com/books/manifolds-current-research-areas/head-pose-estimation-via-manifold-learning

Some applications of manifold learning

- Identifying the underlying “data manifold” enables:
  - Visualization
  - Representation of data in reduced order coordinates
  - Classification, anomaly detection, image segmentation, autonomous driving, virtual reality, …

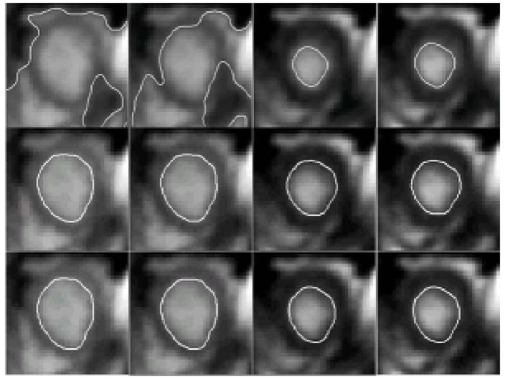
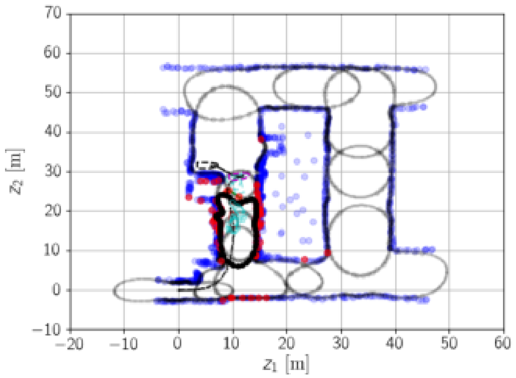
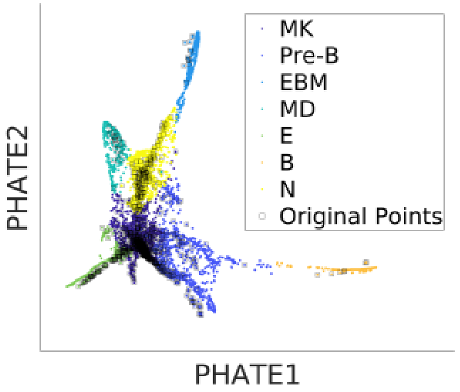


Image segmentation for medical imaging
Qilong Zhang, et al., *2006 IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition (CVPR’06)*, p. 1092.



Obstacle avoidance in autonomous driving
Diwale et al.,
https://infoscience.epfl.ch/record/265381?ln=en



Synthesis of data in data-constrained scenarios
Lindenbaum et al., NeurIPS 2018

Existing methods for manifold learning

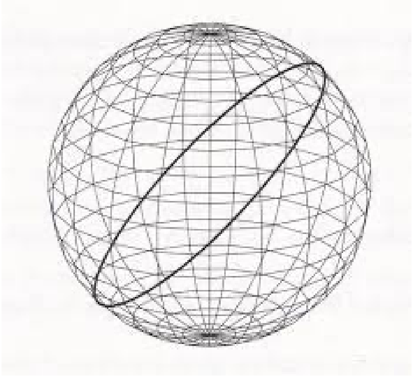
- Existing techniques:
  - Diffusion maps, Laplacian eigenmaps, ISOMAP, Local linear embedding, …
  - Several are *dynamical methods* that rely on the properties of diffusion
    - i.e., at long times heat flow on a manifold distributes heat in a geometrically uniform way
  - Other dynamics useful?

If you want to see something, you send waves in its general direction, you don't throw heat at it.

\- *Attributed to Peter Lax*

A. Cloninger and S. Steinerberger, *Applied and Computational Harmonic Analysis*, 2017.

Geodesics



- Our approach to manifold learning will proceed through learning **geodesic distances on the manifold**.
- Once geodesic distances are known, the intrinsic relationship between the data points is known.
- But isn't calculating geodesic easy? Just do dynamics on the manifold.

Free Hamiltonian (K.E. only) $$\mathcal{H} = |\mathbf{p}|_g^2 = \sum_{i,j} g^{ij}(x) p_i p_j$$

Can solve Hamilton's equations

$$\dot{x}^i = \sum_j g^{ij}(x) p_j$$

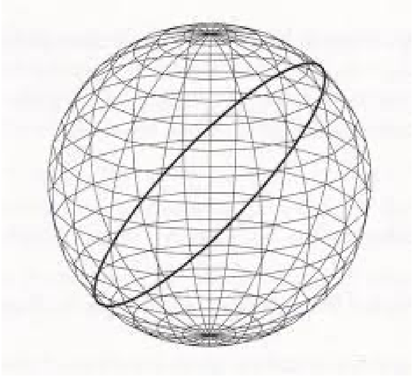
$$\dot{p}_k = -\sum_{ij} g^{ij}_{,k}(x) p_i p_j$$

But our data is just samples of points on the manifold.

Don't know: g_{ij} p^i

Geodesics

- Our approach to manifold learning will proceed through learning **geodesic distances on the manifold**.
- Once geodesic distances are known, the intrinsic relationship between the data points is known.
- But isn't calculating geodesic easy? Just do dynamics on the manifold.



Free Hamiltonian (K.E. only) $$\mathcal{H} = |\mathbf{p}|_g^2 = \sum_{i,j} g^{ij}(x) p_i p_j$$

Can solve Hamilton's equations

$$\dot{x}^i = \sum_j g^{ij}(x) p_j$$

$$\dot{p}_k = -\sum_{ij} g^{ij}_{,k}(x) p_i p_j$$

But our data is just samples of points on the manifold.

Don't know: g_{ij} p^i

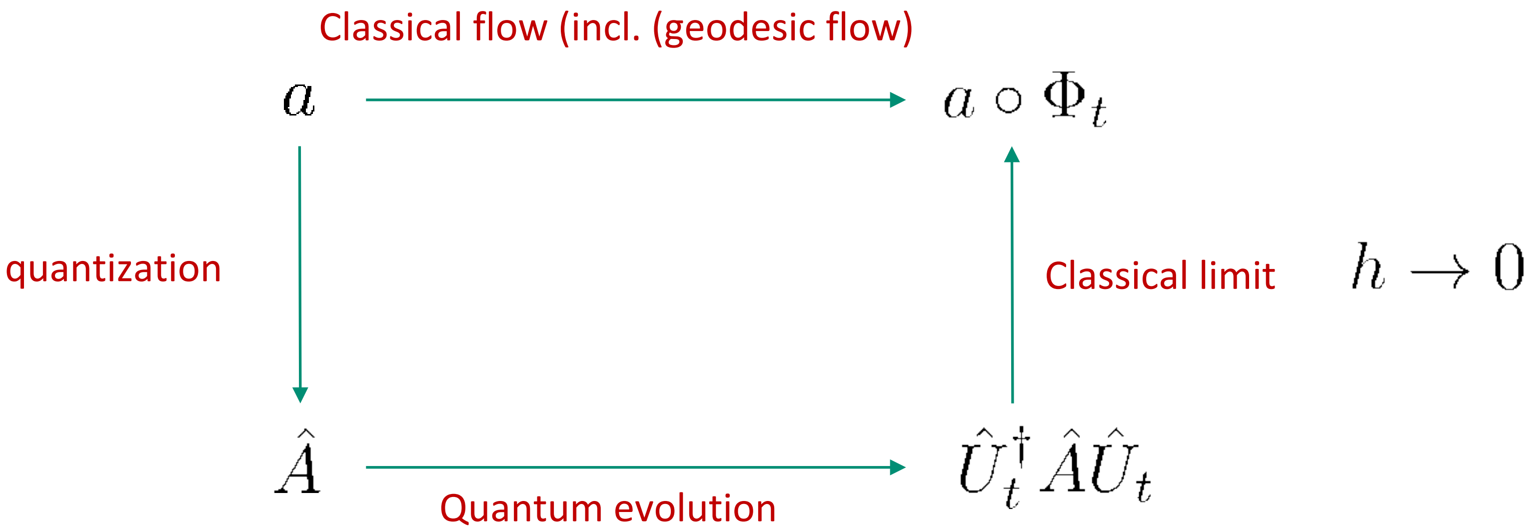
Instead, we will look at quantized dynamics.

- Instead of propagating a classical particle on the manifold, we will propagate a quantum state. Why is that better? We'll see.

The quantum manifold learning program

Central motivation

Egorov's theorem (1969)



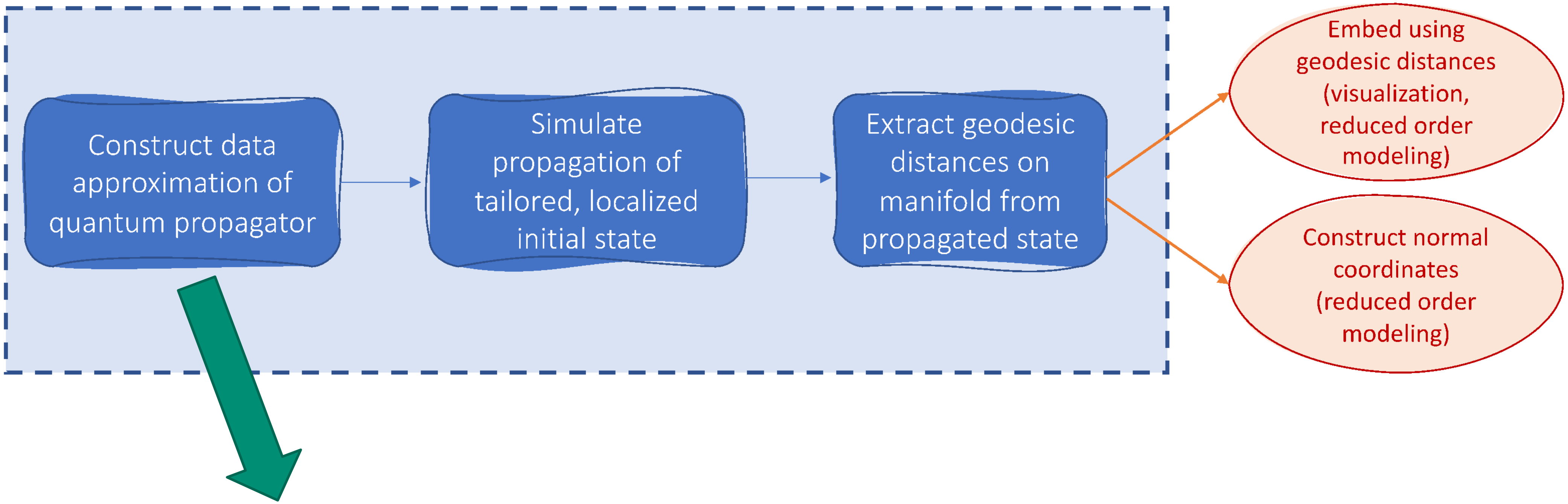
This is the $N \to \infty$ limit in the data context

Advantages:

1. Linearization: geodesic flow through linear dynamics
   - Must more efficient than approximate solution of geodesic equation (e.g., fast forward marching)

2. Rigorous convergence proofs and hyper-parameter choices

The quantum manifold learning program



Using the samples from the manifold, we want to construct an approximation of the operator:

$$\hat{U}_h(t) = e^{-\frac{i}{h}\hat{H}t} = e^{-i\sqrt{\Delta_g}t}$$

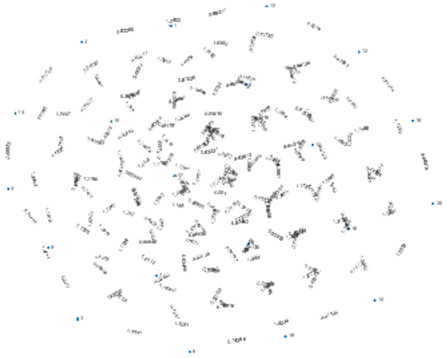
Manifold Laplacian

Data-driven construction of quantum propagator

Data $V = \{v_1, v_2, \ldots, v_N\}$ $v_i \in \mathbb{R}^n$

Samples of points on manifold

car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocati	wheelbase	carlength	carwidth	carheight	curbweight
1	3	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
2	3	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
3	1	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823
4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337
5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824
6	2	audi fox	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507
7	1	audi 100ls	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844
8	1	audi 5000	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954
9	1	audi 4000	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086
10	0	audi 5000s (	gas	turbo	two	hatchback	4wd	front	99.5	178.2	67.9	52	3053
11	2	bmw 320i	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2395
12	0	bmw 320i	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2395
13	0	bmw x1	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2710
14	0	bmw x3	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2765
15	1	bmw z4	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3055
16	0	bmw x4	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3230
17	0	bmw x5	gas	std	two	sedan	rwd	front	103.5	193.8	67.9	53.7	3380
18	0	bmw x3	gas	std	four	sedan	rwd	front	110	197	70.9	56.3	3505



Graph embedding of data

$$[T_{N,\epsilon}]_{i,j} = k\left(\frac{\|v_i - v_j\|^2}{2\epsilon}\right)$$

Scale parameter

Adjacency matrix for symmetric weighted graph

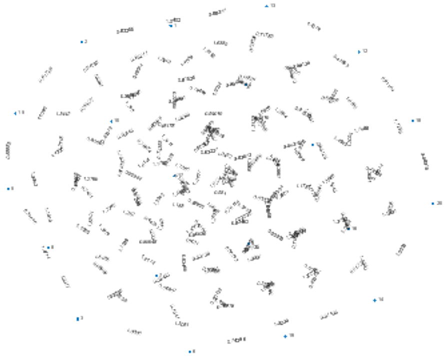
where $k(.)$ is an exponentially decaying function in its argument.

Data-driven construction of quantum propagator

Data $V = \{v_1, v_2, \ldots, v_N\}$ $v_i \in \mathbb{R}^n$

Samples of points on manifold

car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	carlength	carwidth	carheight	curbweight
1	3	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
2	3	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
3	1	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823
4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337
5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824
6	2	audi fox	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507
7	1	audi 100ls	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844
8	1	audi 5000	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954
9	1	audi 4000	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086
10	0	audi 5000s (	gas	turbo	two	hatchback	4wd	front	99.5	178.2	67.9	52	3053
11	2	bmw 320i	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2395
12	0	bmw 320i	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2395
13	0	bmw x1	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2710
14	0	bmw x3	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2765
15	1	bmw z4	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3055
16	0	bmw x4	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3230
17	0	bmw x5	gas	std	two	sedan	rwd	front	103.5	193.8	67.9	53.7	3380
18	0	bmw x3	gas	std	four	sedan	rwd	front	110	197	70.9	56.3	3505



Graph embedding of data

$$[T_{N,\epsilon}]_{i,j} = k\left(\frac{\|v_i - v_j\|^2}{2\epsilon}\right)$$

Scale parameter

Adjacency matrix for symmetric weighted graph

where $k(.)$ is an exponentially decaying function in its argument.

Define the graph Laplacian

$$L_{N,\epsilon} \equiv \frac{I_N - D_{N,\epsilon}^{-1} T_{N,\epsilon}}{\epsilon}, \qquad D_{N,\epsilon} \equiv \mathrm{diag}\left(\sum_{j=1}^{N} [T_{N,\epsilon}]_{i,j}\right)$$

- Also used in spectral methods such as diffusion maps, Laplacian eigenmaps, …
- In general $\epsilon \to 0$ as $N \to \infty$

Data-driven construction of quantum propagator

Define the graph Laplacian

$$L_{N,\epsilon} \equiv \frac{I_N - D_{N,\epsilon}^{-1} T_{N,\epsilon}}{\epsilon}, \qquad D_{N,\epsilon} \equiv \mathrm{diag}\left(\sum_{j=1}^{N} [T_{N,\epsilon}]_{i,j}\right)$$

We prove

$$L_{N,\epsilon} \xrightarrow{N\to\infty} \mathcal{L}_h \approx h^2 \left(\Delta_g + 2\frac{\nabla p \cdot \nabla}{p}\right)$$

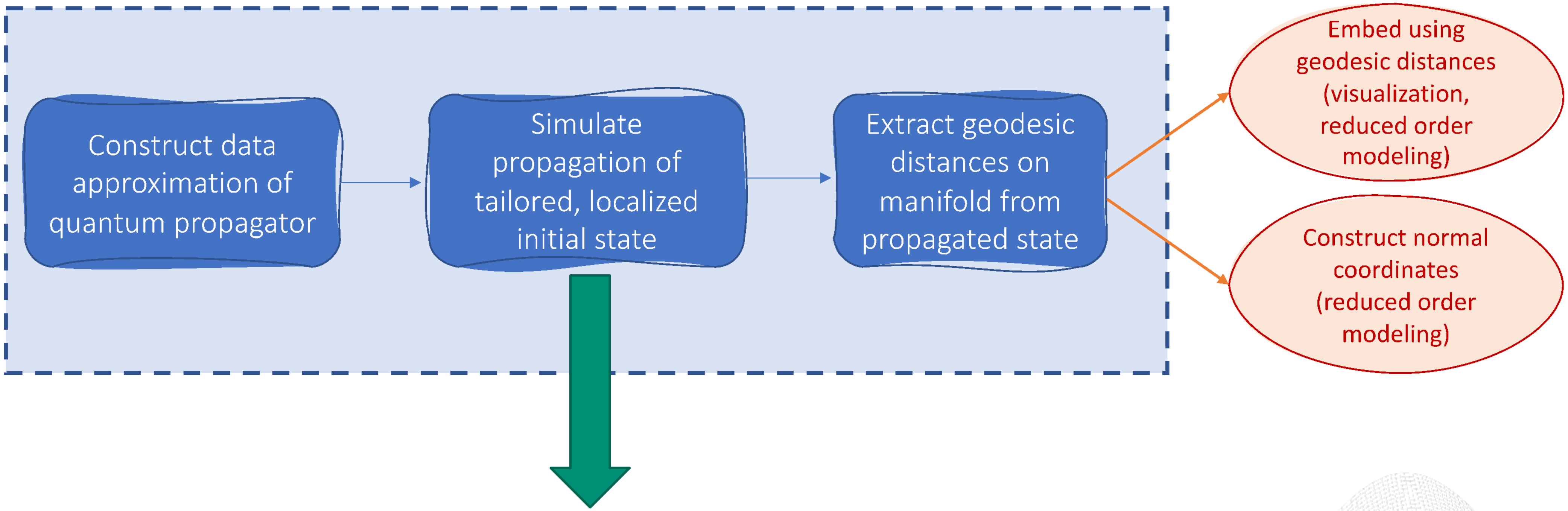
Semiclassical parameter

measure according to which manifold is sampled

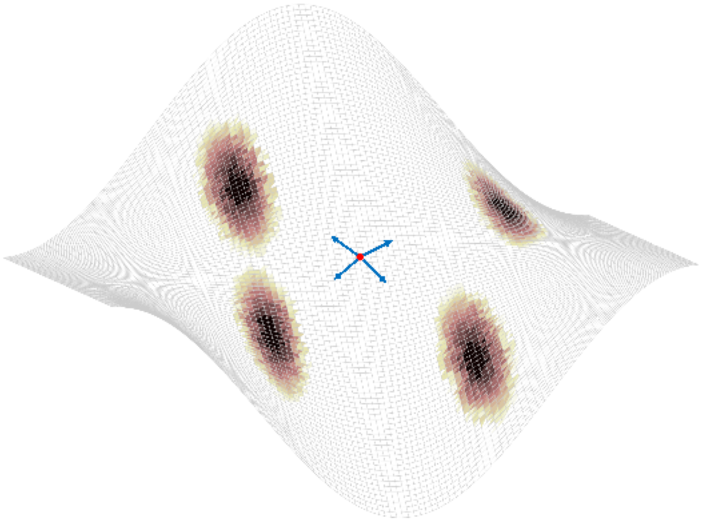
c.f.
- Coifman et al., PNAS, **102,** 7426 (2005)
- Hein, Audibert, von Luxburg, J. Mach. Learn. Res., 8, 1325 (2007)

So we can approximate the quantum propagator $\hat{U}_h(t) = e^{-\frac{i}{h}\hat{H}t}$ with $\breve{U}_h(t) = e^{-\frac{i}{h}\sqrt{L_{N,\epsilon}}\,t}$

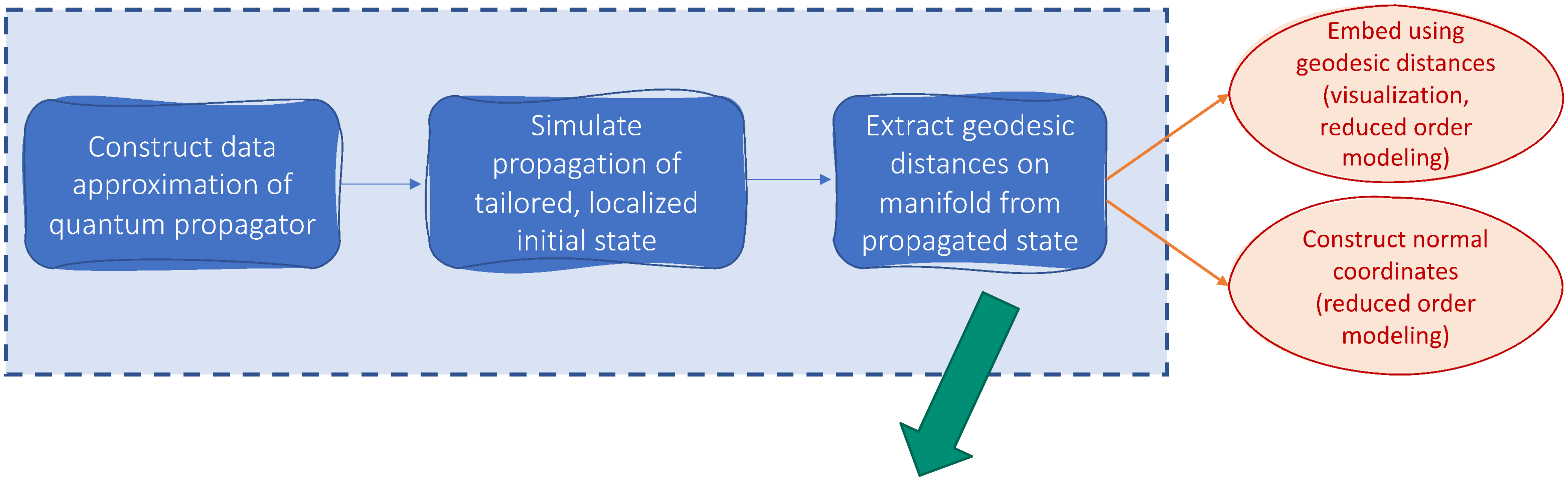
The quantum manifold learning program



- We want to know geodesic distances of sample points from some given point.
- Our strategy will be to propagate a quantum state initially localized at the initial point (like a test particle).
- What localized initial state? A delta-function on the initial point is too localized – its not an L^2 function and will not propagate along geodesics. Instead, use the most classical state, a coherent state.



The quantum manifold learning program



How to extract geodesic distance from propagated state? We prove

$$\lim_{h \to 0} \left| \langle \psi^h_{\zeta_0} | \hat{U}_h(t)^\dagger \hat{x} \hat{U}_h(t) | \psi^h_{\zeta_0} \rangle - x \circ \Phi_t(\mathbf{x}_0, \mathbf{p}_0) \right| = 0$$

Coherent state — Position operator — Geodesic flow

if $h = \epsilon^{\frac{1}{2+\alpha}}, \quad \alpha \approx 1$

Discretization as quantization

$$h = \epsilon^{\frac{1}{2+\alpha}}, \quad \alpha \approx 1$$

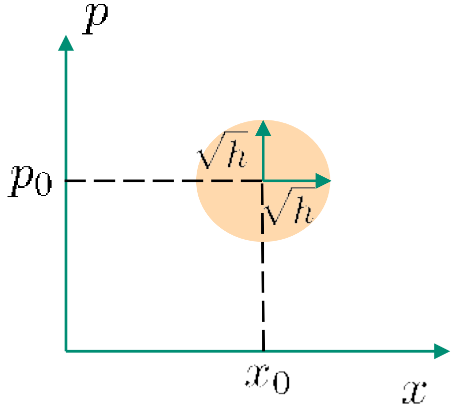
$$L_{N,\epsilon} \equiv \frac{I_N - D_{N,\epsilon}^{-1} T_{N,\epsilon}}{\epsilon},$$

ϵ captures the limited resolution of the manifold due to finite sampling

The “uncertainty” implied by this limited resolution needs to be distributed in phase space

Equal distribution in configuration and momentum space => $h \sim \epsilon^{\frac{1}{2}}$

$$\zeta_0 = (\mathbf{x}_0, \mathbf{p}_0) \in T^*\mathcal{M}$$

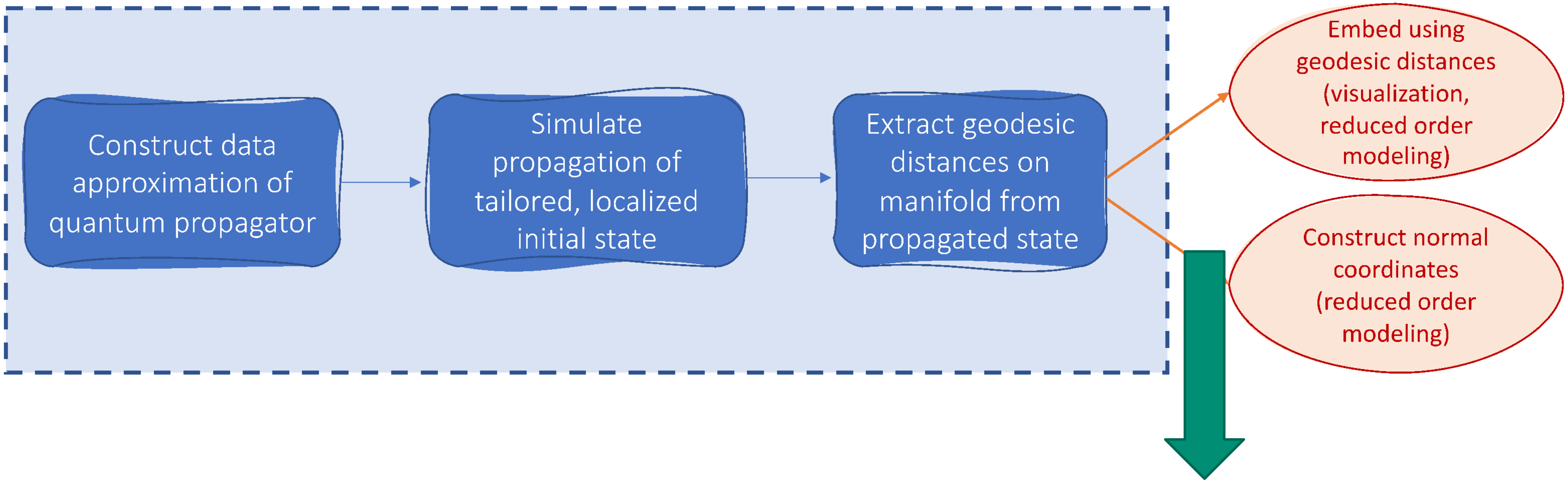


$$\psi_{\zeta_0}(\mathbf{x}) = \langle \mathbf{x} | \psi_{\zeta_0} \rangle = \frac{1}{(\pi h)^{\frac{\nu}{4}}} e^{\frac{i}{h}\langle \mathbf{x}-\mathbf{x}_0, \mathbf{p}_0 \rangle} e^{-\frac{|\mathbf{x}-\mathbf{x}_0|^2}{2h}}$$

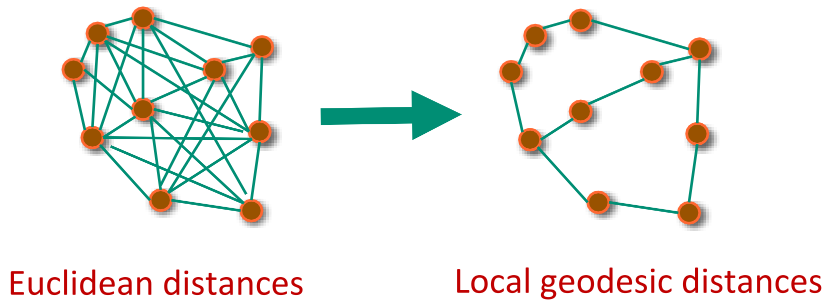
“Classical limit” $h \to 0$

is the large data limit $N \to \infty$

The quantum manifold learning program

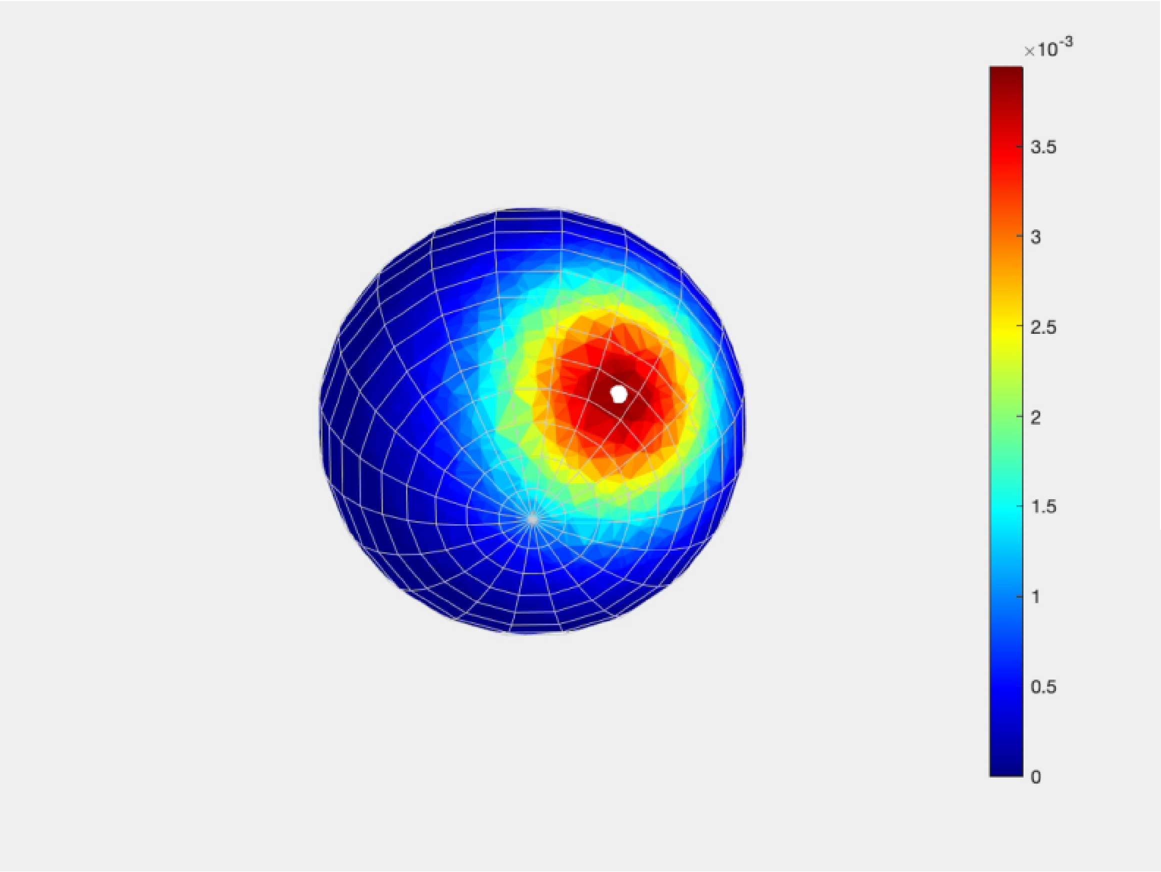


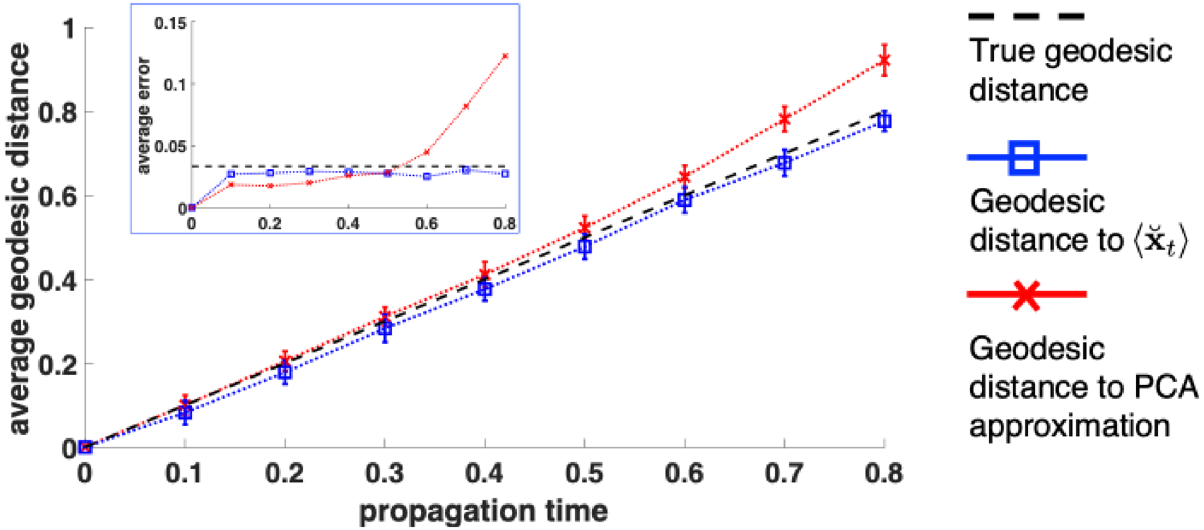
By converting our input Euclidean distances to approximated geodesic distances in neighborhoods, we get a sparse, and more accurate graph embedding of the data



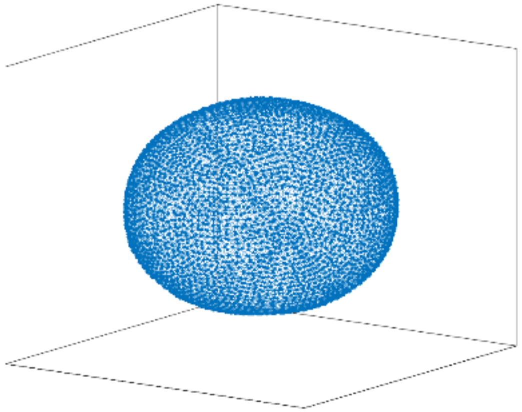
Example: sphere

N=3000 points, uniformly sampled on unit sphere





Force-based embedding based on extracted geodesic distances



Example: COVID-19 mobility data

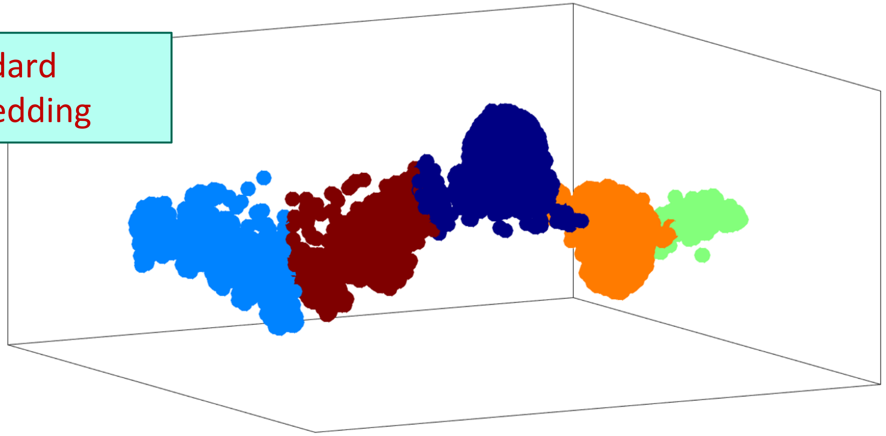
Social Distancing Metric dataset from SafeGraph Inc.
https://docs.safegraph.com/docs/social-distancing-metrics

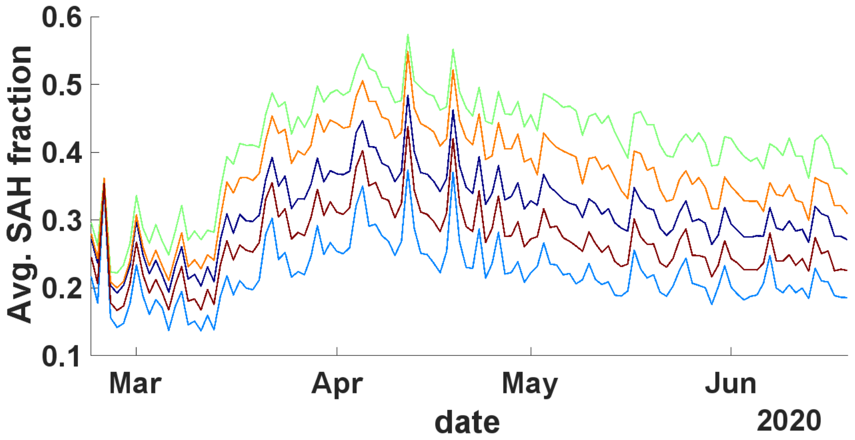
- Dataset collects user location information (from cellphone GPS data) over the course of the initial 3 months of the COVID-19 pandemic (Feb 23, 2020 – June 19, 2020: 117 days).
- Aggregated at the census block group (CBG) level.
- Understanding patterns in mobility behavior can help tune public health policy.
- We compute a “stay-at-home” fraction which represents the fraction of devices that stayed at their home location on a day.
- We concentrate on the data for Georgia (GA), which has 5509 CBGs.
- Dataset: 5509 x 117

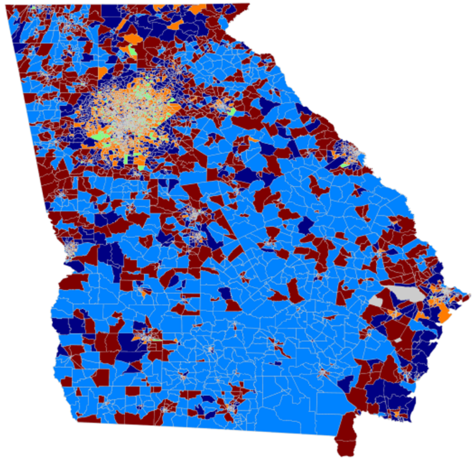
Apply manifold learning through geodesics and embed in 3 dimensions (**reduction from 117 dimensions**) and then perform clustering using K-means.

Example: COVID-19 mobility data

Standard embedding



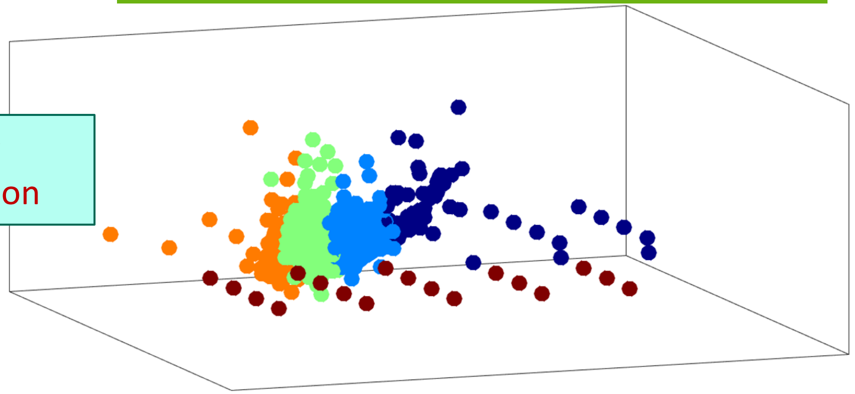


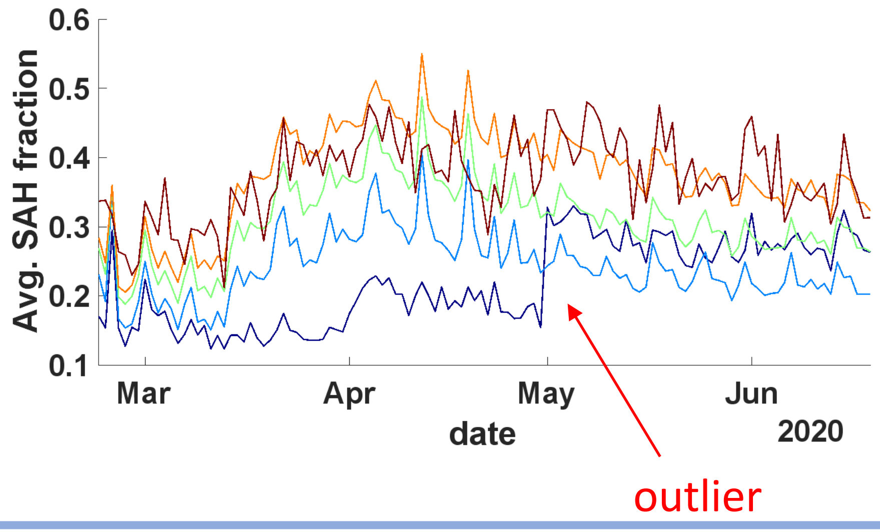


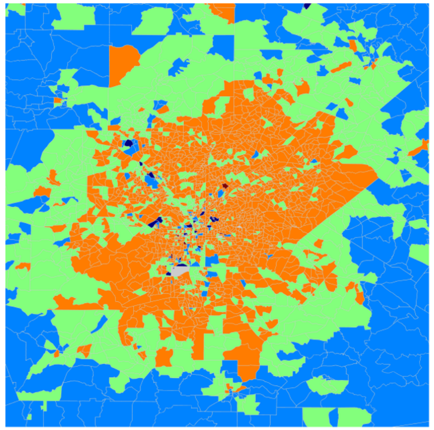
Embedding in 3D and k-means using these 3D coordinates

Average SAH fraction time series for each cluster

Outlier detection







c.f. Levin et al., “Cell Phone Mobility Data and Manifold Learning.” https://doi.org/10.1101/2020.10.31.20223776

Summary

- A new approach to discovering geodesics on data manifolds, and manifold learning, based on quantum dynamics. c.f.
  - Diffusion => Monte Carlo methods
  - Annealing => optimization
  - Quantum and wave propagation => geodesics

- Alternative to some common data processing algorithms:
  - Dijkstra’s algorithm
  - Fast forward marching

- Many directions for follow-on research:
  - Develop a quantum algorithm for manifold learning on quantum computers
  - Make existing approach more efficient
  - Develop methods for embedding point cloud data based on geodesic distances

Backup slides

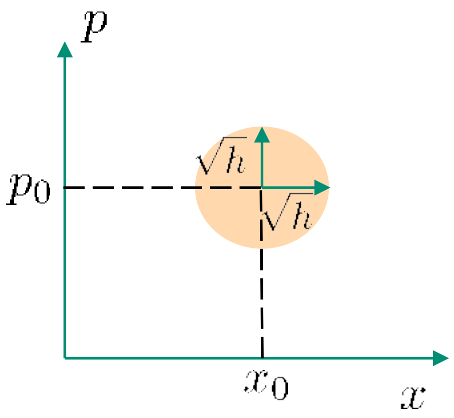
Initial states

- What localized initial state? A delta-function on the initial point is too localized – its not an L^2 function and will not propagate along geodesics. Instead, use the most classical state, a coherent state.

- Coherent states can be characterized in various ways. One useful one:
  - Have equal and minimal uncertainty in position and momentum.

$$\psi_{\zeta_0}(\mathbf{x}) = \langle \mathbf{x} | \psi_{\zeta_0} \rangle = \frac{1}{(\pi h)^{\frac{\nu}{4}}} e^{\frac{i}{h}\langle \mathbf{x}-\mathbf{x}_0, \mathbf{p}_0 \rangle} e^{-\frac{\|\mathbf{x}-\mathbf{x}_0\|^2}{2h}}$$

$$\zeta_0 = (\mathbf{x}_0, \mathbf{p}_0) \in T^*\mathcal{M}$$



Combescure & Robert. *Coherent States and Applications in Mathematical Physics.* Springer Netherlands, 2012.
Gazeau, *Coherent States in Quantum Physics.* Wiley-VCH, 2009.

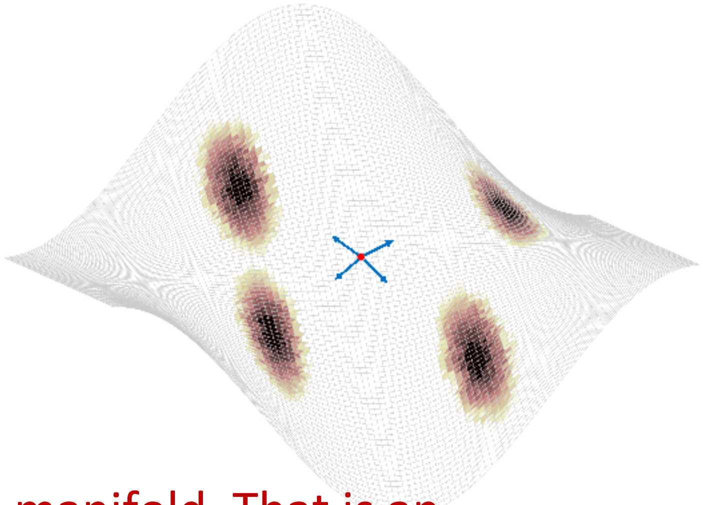
- Can approximate this using the data we have

$$\left[|\tilde{\psi}_{\zeta_0}\rangle \right]_i = e^{\frac{i}{h}(v_i - v_0)^\top p_0} e^{-\frac{\|v_i - v_0\|^2}{2h}}, \quad 1 \leq i \leq N.$$

- Initial momentum approximated using a principal component analysis (PCA)

Propagation of coherent states

- We propagate a collection of such coherent states to determine the *geodesic spray* of points a certain geodesic distance from initial point.



- It's pretty neat that we're able to do everything with just samples from the manifold. That is an advantage of quantization.

- In the continuum (asymptotic) setting, computing $|\psi_t\rangle = \breve{U}_t|\psi_{\zeta_0}\rangle$ corresponds to computing

$$\psi(\mathbf{x}, t) = \int_{\mathcal{M}} d\mathbf{x}' \; G(\mathbf{x}, \mathbf{x}'; t)\psi(\mathbf{x}', 0)$$

Dynamics completely in terms of position coordinates only.

Example: COVID-19 mobility data

Google COVID-19 Community Mobility Reports

https://www.google.com/covid19/mobility/

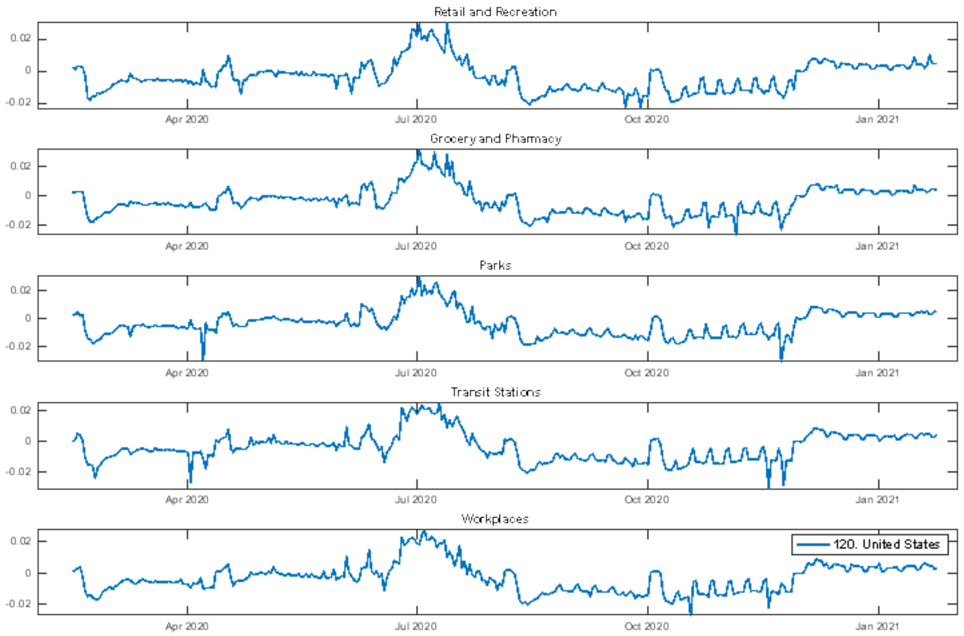
Dataset collects user mobility information (% change in mobility from baseline) over 6 categories for 132 countries and regions within these countries.

Timeframe: ~1 year (Feb 15, 2020 – Jan 24, 2021)
Baseline: Jan 3 – Feb 6, 2020
Categories: Retail and recreation, grocery and pharmacy, parks, transit stations, workplaces.

After pre-processing:

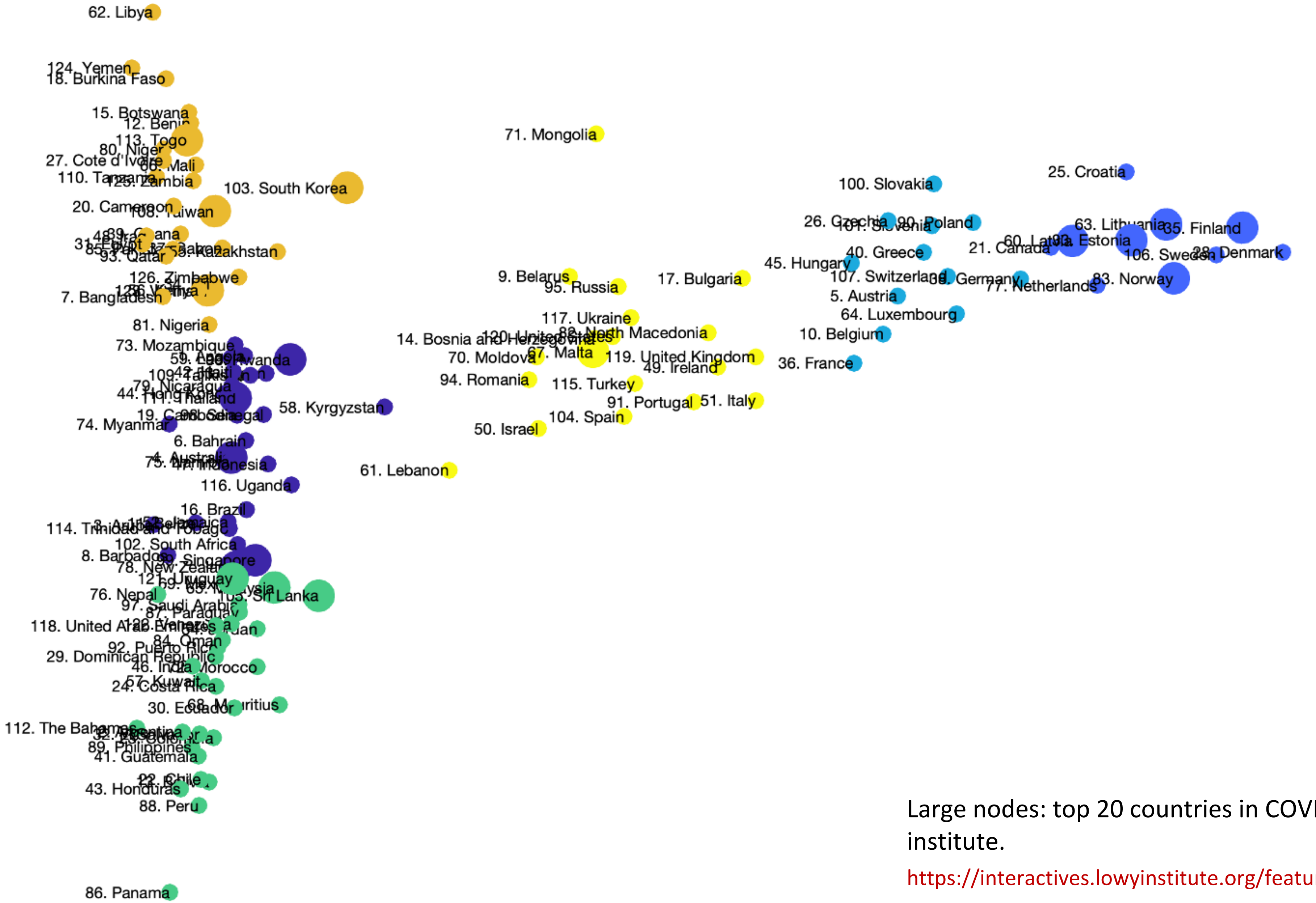
For each country, **345*5=1725 columns (features)** that represent a time series of mobility changes across 5 categories.

e.g.,



Apply manifold learning through geodesics and embed in 3 dimensions (**reduction from 1725 dimensions**)

Example: COVID-19 mobility data



Large nodes: top 20 countries in COVID response according to Lowy institute.

https://interactives.lowyinstitute.org/features/covid-performance/

Example: COVID-19 mobility data