Extracting isometric maps and topology from data

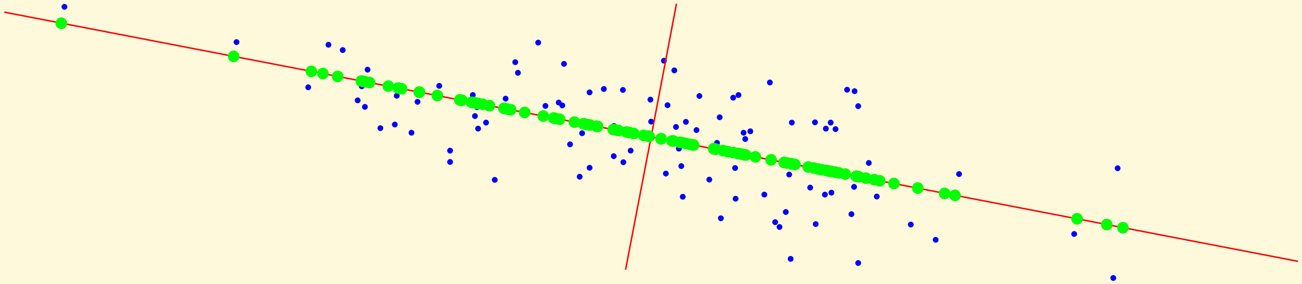
Matthew Brand

MERL

Featuring projects with

- John Hershey, UCSD
- Rahul Bhotika, U Rochester -> GE
- Wojciech Matusik, MIT -> MERL
- Yuechao Zhao, Harvard -> ?

Motivation: Subspace methods



dimensionality reduction: project *into* subspace $\mathbb{R}^D \longrightarrow \mathbb{R}^d$

denoising: project *onto* subspace $\mathbb{R}^D \longrightarrow \mathbb{R}^d \longrightarrow \mathbb{R}^D$

editing: move *on* subspace $\mathbb{R}^D \longrightarrow \mathbb{R}^d \longrightarrow \mathbb{R}^d \longrightarrow \mathbb{R}^D$

MERL subspace sampler

Instant Recommender

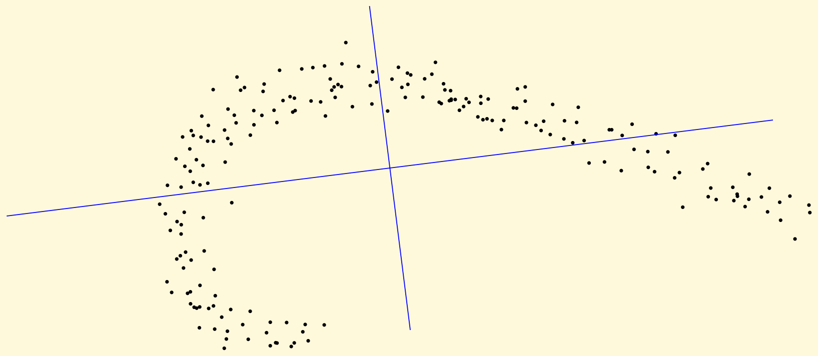
Your Ratings (UserID = 944)

Item Name	Attrib...	Attribute2	Rating (1-5)
Star Wars	1977...	Sci-Fi	
Toy Story	1995...	Animation	
Scream	1996...	Horror	
Liar Liar	1997...	Comedy	
Independence Day	1996...	Sci-Fi	
Fargo	1997...	Drama	
English Patient, The	1996...	Romance	
Raiders of the Lost Ark	1981...	Action	
Willy Wonka and the Chocolate ...	1971...	Family	
Forrest Gump	1994...	Drama	
Terminator, The	1984...	Sci-Fi	
Sound of Music, The	1965...	Family	
Ran	1985...	Drama	
Akira	1988...	Animation	

10 ms to add your 14 ratings to IncSVD.

- 2D ↻ & 3D visual tracking ↻ ↻ ↻
- 3D-from-video ↻
- face & texture synthesis ↻ ↻
- face & pedestrian detection ↻ ↻
- movie recommending ↻

Subspace $\neq$ submanifold

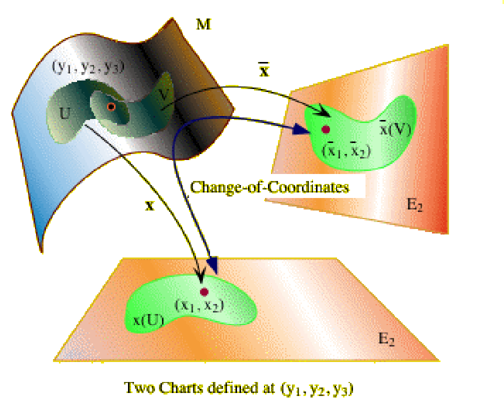


- Subspace methods work for small problems because data distribution is nearly flat.
- In larger datasets, distribution is more clearly curved... ...samples from submanifold, not subspace.

Manifold modeling

NonLinear Dimensionality Reduction (NLDR): Separate

- intrinsic geometry (coordinate system of manifold) from
- extrinsic geometry (its embedding in ambient space).



Discrete analogue: Graph embeddings under metric constraints.

Graph embeddings

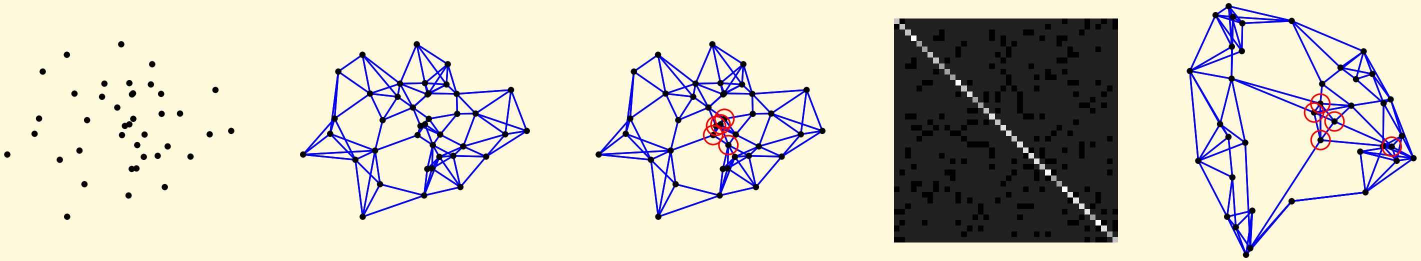
Connectivity: Error metric in $\mathbb{R}^N$ space of immersions = graph Laplacian matrix...

$$\mathbf{L}\text{aplacian} = \operatorname{diag}(\mathbf{d}\text{egree}) - \mathbf{A}\text{djacency}$$

...optimized by its eigenvectors.
[Tutte, 1963; Fiedler, 1975; Mohar, 1991]

Metric structure: Other error metrics can be built from metric constraints of overlapping data neighborhoods.
[Kamada and Kawai, 1989; Tenenbaum et al., 2000]

Generic NLDR framework



1. Impose local connectivity graph on N datapoints.
2. Estimate metric properties in local graph cliques.
3. Combine metric & connectivity constraints in a global $N \times N$ row-space quadratic form (~kernel).
4. Immersion $\longleftarrow$ extremal eigenvectors.

Local NLDR methods

Minimize projection of immersed cliques onto local...

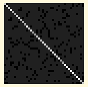
LLE: linear regression weights (~nonconformalities) [Roweis and Saul, 2000]

Laplacian eigenmap: barycentric coordinates (~weighted sum of squared edge lengths) [Belkin and Niyogi, 2002]

Charting, LTSA: nullspace (~local PCA residuals) [Brand, 2003; Zhang and Zha, 2003]

HLLE: Hessian estimator (~Gaussian curvature) [Donoho and Grimes, 2003]

Global NLDR methods

Quadratic form  is itself a global optimization problem:

Isomap: maximize fit to graph shortest-path lengths. [Tenenbaum et al., 2000]

Charting: maximize overlap of adjacent tangent spaces. [Brand, 2003]

SDE: maximize spread of points subject to local angular constraints. [Weinberger et al., 2004]

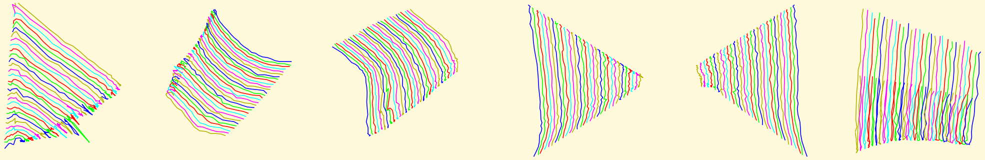
Many error criteria...

...which maximize fidelity to data manifold?

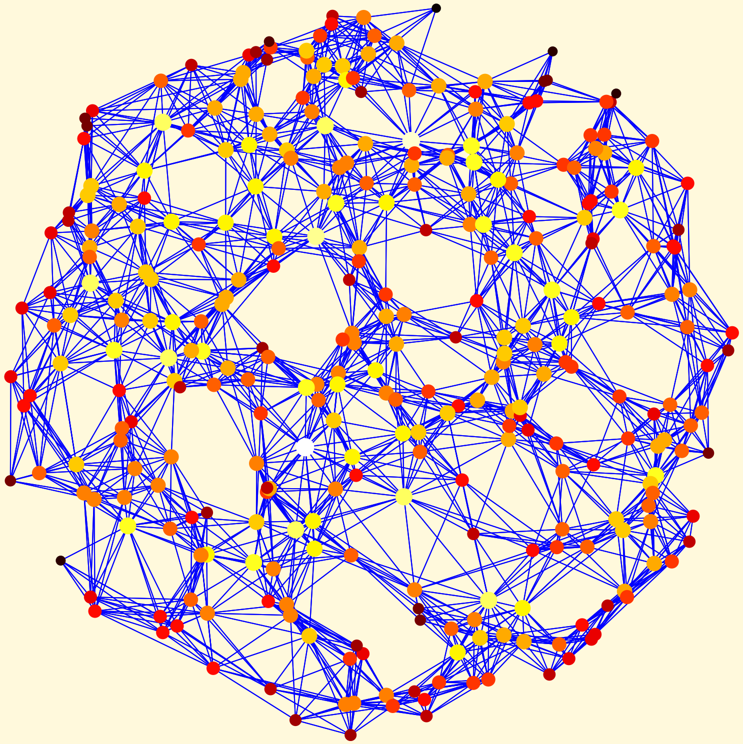
Continua limit: Almost all do.

Finite data:

- Maximize fidelity to invented graph,
- subject to approximate local metric constraints,
- with very optimistic topological assumptions.



Graph density $\neq$ data density

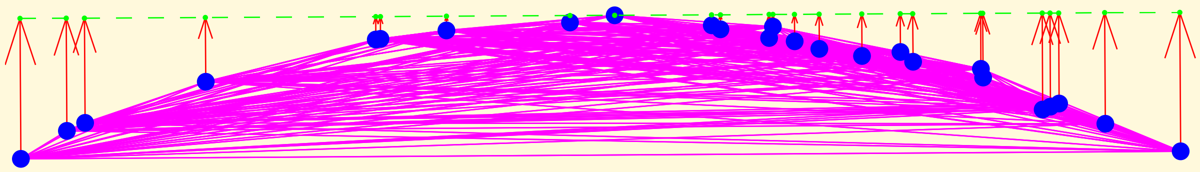


- NLDR solutions distribute error by graph coverage.
- Internal (high-degree) points are more constrained.
- More error near boundaries.

Challenge: Marginalize over all possible graphs.

Today: Maximizing isometry w.r.t. data density.

The locally linear bias



$\mathbb{R}^D$ Euclidean metric used as proxy for geodesic lengths on $\mathcal{M}$.

- Linear distances & projections *always* underestimate geodesic distances on $\mathcal{M}$...
  - ...fish-eye distortion where $\mathcal{M}$ has extrinsic curvature

Today: Exact geodesic parameterizations for broad class of extrinsically curved data manifolds.

Topology

NLDR parameterizations most useful when $\mathcal{M}$ isomorphic to n-ball.

Natural data sets not very accommodating.

Today: Partitioning $\mathcal{M}$ into n-balls for varied n.

Geodesic Nullspace Analysis

Local parameterization: Flattened local coordinate systems.

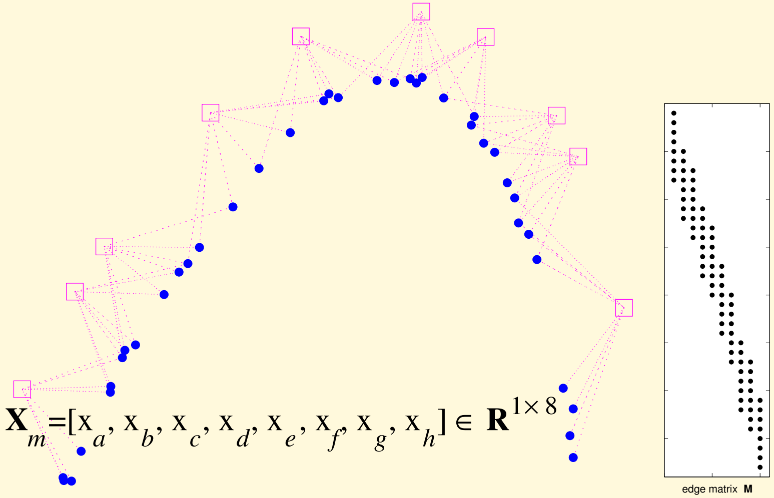
Global coordination: Construct operator $\mathbf{K}$ that separates linear (intrinsic) from nonlinear (extrinsic) component of the data.

- *Exact* isometry from *finite* samples.
- Generalizations for new points & their mappings.

GNA coordination

Input:

1. Graph $\mathcal{G}$ assigns N points to M overlapping cliques via edge matrix $\mathbf{M} \in \mathbb{R}^{N\times M} \geq 0$ (binary or weighted).
2. Local low-dimensional parameterization $\mathbf{X}_m \in \mathbb{R}^{d\times k}$ of each k-point clique on $\mathcal{M}$.



Output: Immersion $\mathbf{Y}_{\text{iso}} \in \mathbb{R}^{d\times N}$ having local isometry to $\mathcal{M}$ (MMSE w.r.t. the empirical density).

The local row-nullspace

Let $\mathbf{P}_m$ be orthogonal basis for rows of m^{th}

homogeneous clique coordinates $\begin{bmatrix} \mathbf{X}_m \\ \mathbf{1}^\top \end{bmatrix} \in \mathbb{R}^{(d+1)\times k}$.

- $\mathbf{P}_m$ spans the space of affine functions of $\mathbf{X}_m$.

Row-nullspace $\mathbf{Q}_m \doteq \mathbf{P}_m^\perp$.

$$\forall_{\mathbf{A}\in\mathbb{R}^{(d+1)\times(d+1)}}\ \mathbf{A}\begin{bmatrix} \mathbf{X}_m \\ \mathbf{1}^\top \end{bmatrix}[\mathbf{P}_m, \mathbf{Q}_m] = \left[\mathbf{A}\begin{bmatrix} \mathbf{X}_m \\ \mathbf{1}^\top \end{bmatrix}, \mathbf{0}\right]$$

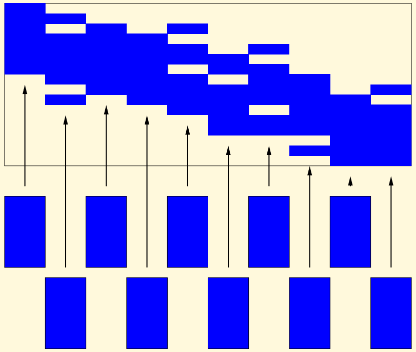
Measuring immersion error

Let $\mathbf{Z} \in \mathbb{R}^{z\times k}$ be a candidate immersion of clique in $\mathbb{R}^{z}$.

- $\|\mathbf{Z}\mathbf{P}_m\|_F$ measures component of $\mathbf{Z}$ that is an affine function of $\mathbf{X}_m$ (~PCA preserved variance)
- $\|\mathbf{Z}\mathbf{Q}_m\|_F$ measures component of $\mathbf{Z}$ that is a nonlinear function of $\mathbf{X}_m$. (~PCA residual)

Joint nullspace constraints

Concatenate nullspaces:

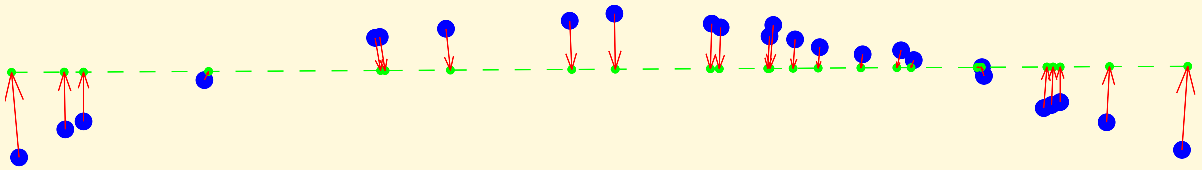


- Perfect $\mathbf{X}_m$: Locally affine immersion lies in column nullspace.
  - Gram matrix is error metric of LTSA and charting.
- Imperfect $\mathbf{X}_m$:
  - Error is distributed w.r.t. clique density, *not* data density.

Nullspace projectors

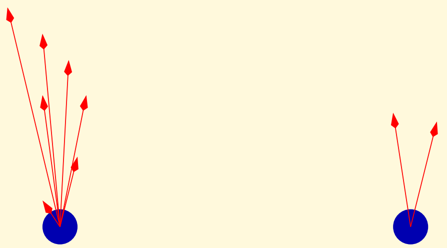
Idempotent matrix $\mathbf{Q}_m\mathbf{Q}_m^\top = \mathbf{I} - \mathbf{P}_m\mathbf{P}_m^\top \in \mathbb{R}^{k\times k}$.

- Let $\mathbf{Z} \in \mathbb{R}^{z\times k}$ be a candidate immersion of m^{th} clique in $\mathbb{R}^z$.



- $\mathbf{Z}\mathbf{P}_m\mathbf{P}_m^\top$ is component of $\mathbf{Z}$ that is an affine function of $\mathbf{X}_m$ (~local PCA denoising).
- $\mathbf{Z}\mathbf{Q}_m\mathbf{Q}_m^\top$ is component of $\mathbf{Z}$ that is a nonlinear function of $\mathbf{X}_m$. (~local extrinsic curvature+noise vectors).

The linearizing operator K



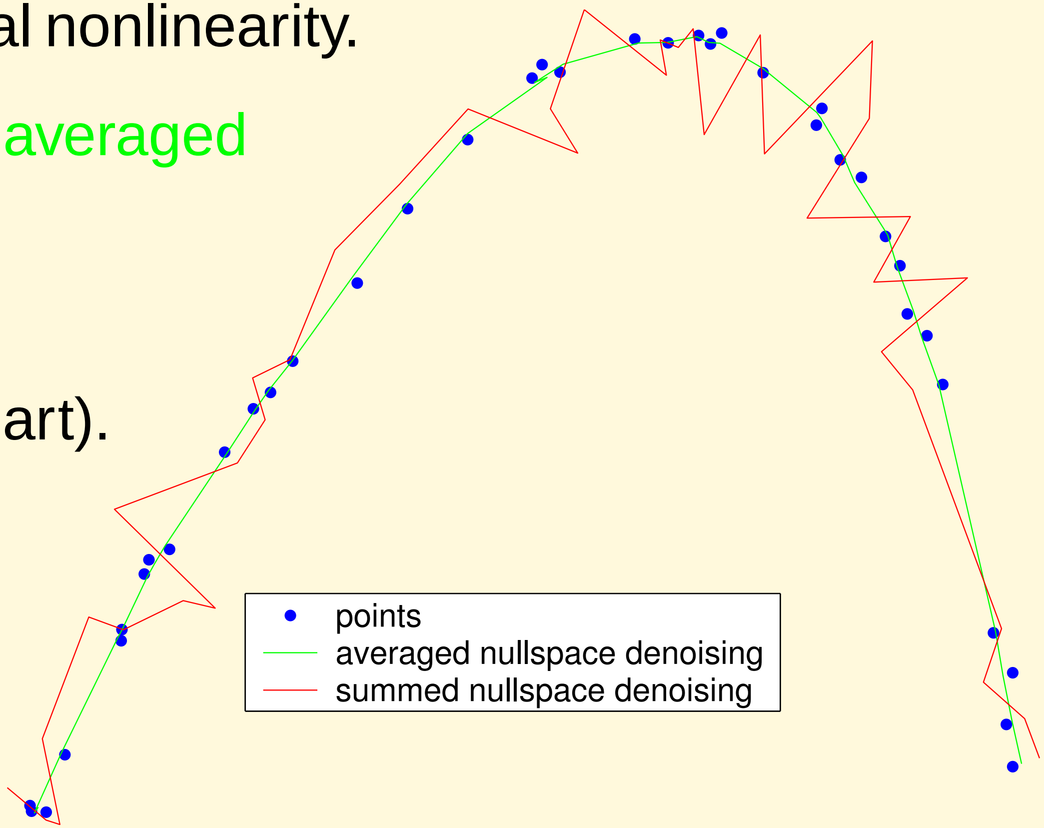
Each clique projector contributes an error vector for point i; average to remove graph coverage bias.

$$\mathbf{K} \doteq \{\sum_m \underbrace{\mathbf{F}_m \mathbf{Q}_m \mathbf{Q}_m^\top \mathbf{F}_m^\top}_{\text{projectors}} \underbrace{\operatorname{diag}(\mathbf{m}_m)\}\operatorname{diag}(\mathbf{M1})^{-1}}_{\text{averaging}} \in \mathbb{R}^{N \times N}$$

- $\mathbf{F}_m \in \{0,1\}^{N \times k}$ selects the m^{th} clique from the full data set.
- $\operatorname{diag}(\cdot)$ terms perform desired averaging.

Data denoising

- $\mathbf{XK} =$ average local nonlinearity.
- $\mathbf{X}(\mathbf{I} - \mathbf{K}) =$ averaged linearization.
- Compare summed linearization (prior art).



Properties of K

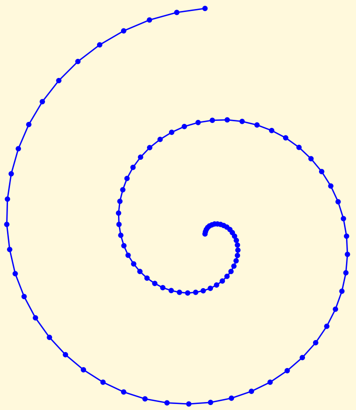
1. Invariant to rigid motions & rescalings of the data.
2. $\mathbf{KK}^\top$ is an error metric; $(\mathbf{I}-\frac{1}{N}\mathbf{1}\mathbf{1}^\top)-\mathbf{KK}^\top$ is a kernel.
3. Bounded spectral radius: $0 \leq \lambda(\mathbf{K}) \leq 1$.
4. $\mathbf{X}(\mathbf{I}-\mathbf{K})$ smooths and denoises data at clique scale.
5. $\mathbf{X}(\mathbf{I}-\mathbf{K})^n$ removes noise and curvature at larger scales. ↻
    - $\lim_{n\to\infty}\mathbf{X}(\mathbf{I}-\mathbf{K})^n \in \mathrm{span}(\mathrm{null}(\mathbf{K}))^\top$

Immersion properties

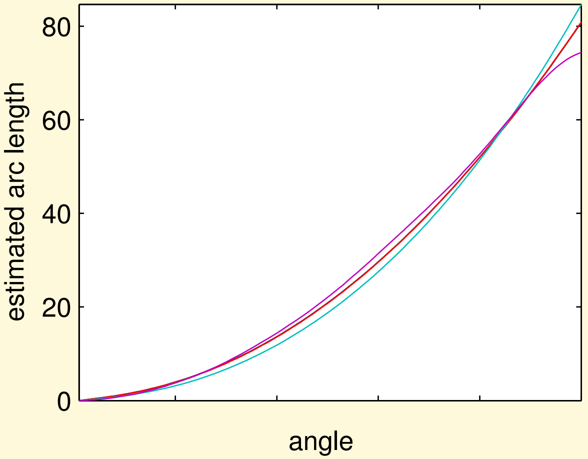
Let $\mathbf{Y}_{\text{aff}}^{\top} \doteq \text{null}(\mathbf{K}) \backslash \mathbf{1}$.

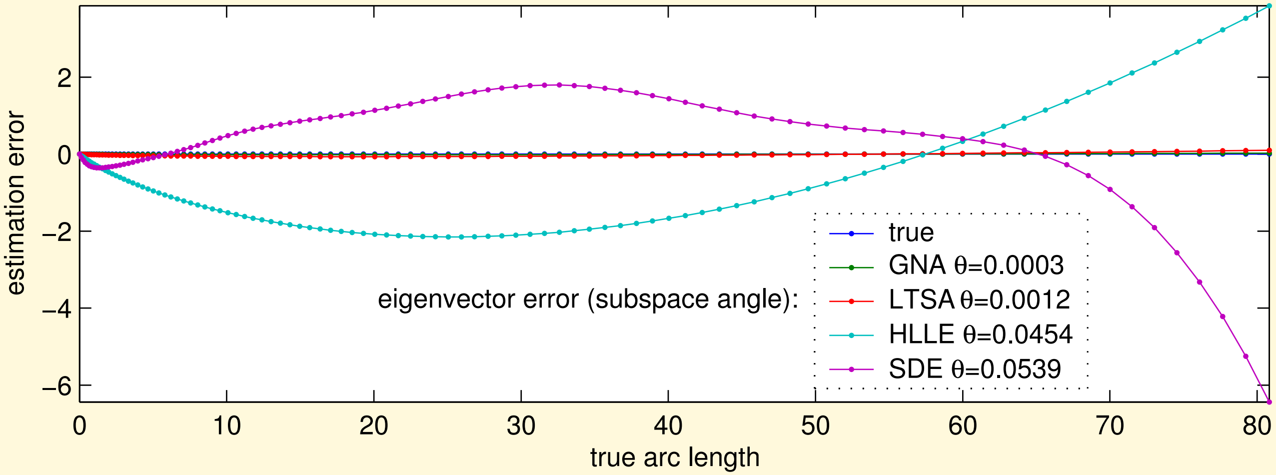
1. Perfect $\mathbf{X}_m$, $\mathcal{M}$ immersible in $\mathbb{R}^d$ w/ local isometry: $\mathbf{Y}_{\text{iso}}$ obtained from $\mathbf{Y}_{\text{aff}}$ by a global linear shear in $\mathbb{R}^d$.
2. Perfect $\mathbf{X}_m$, $\mathcal{M}$ *not* isometrically immersible in $\mathbb{R}^d$: $\mathbf{Y}_{\text{aff}}$ minimally nonaffine to $\mathcal{M}$ at each point.
3. $\mathbf{X}_m + \mathcal{N}(0, \sigma)$ noise: $\mathbf{Y}_{\text{aff}}$ is MMSE w.r.t. data density.
   - On average, each point minimally displaced from a configuration that is linear in the local configuration on $\mathcal{M}$.

Imperfect constraints example

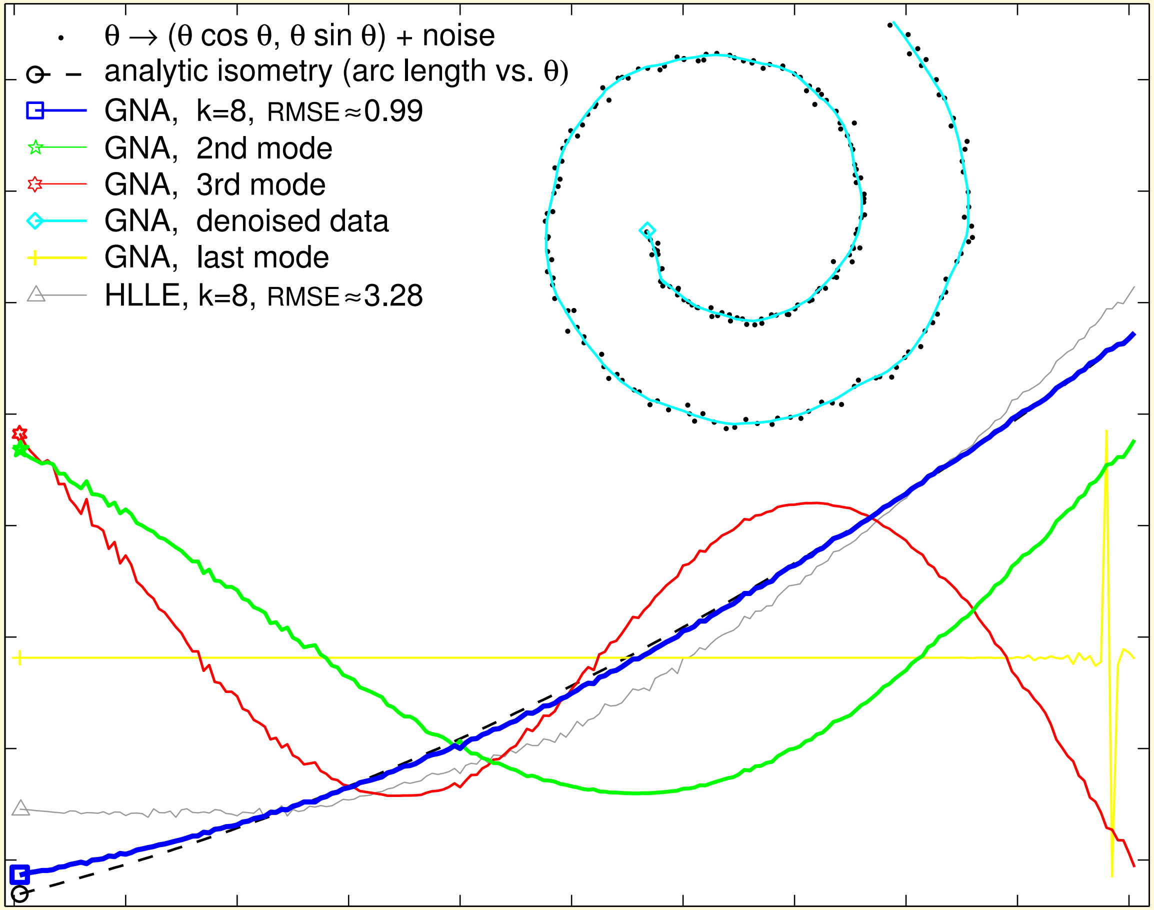


spiral manifold



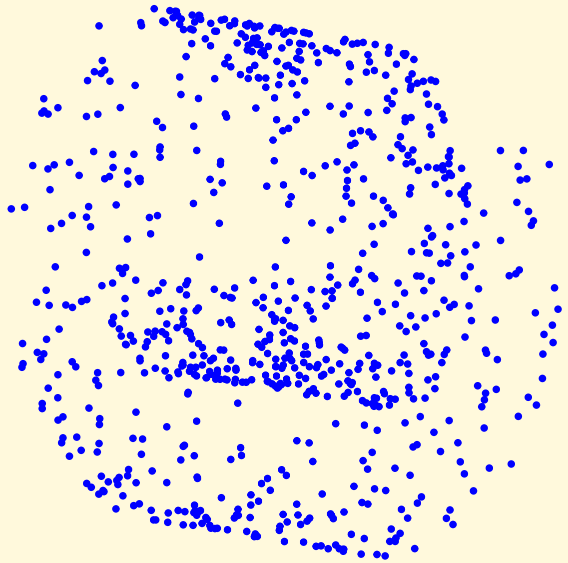


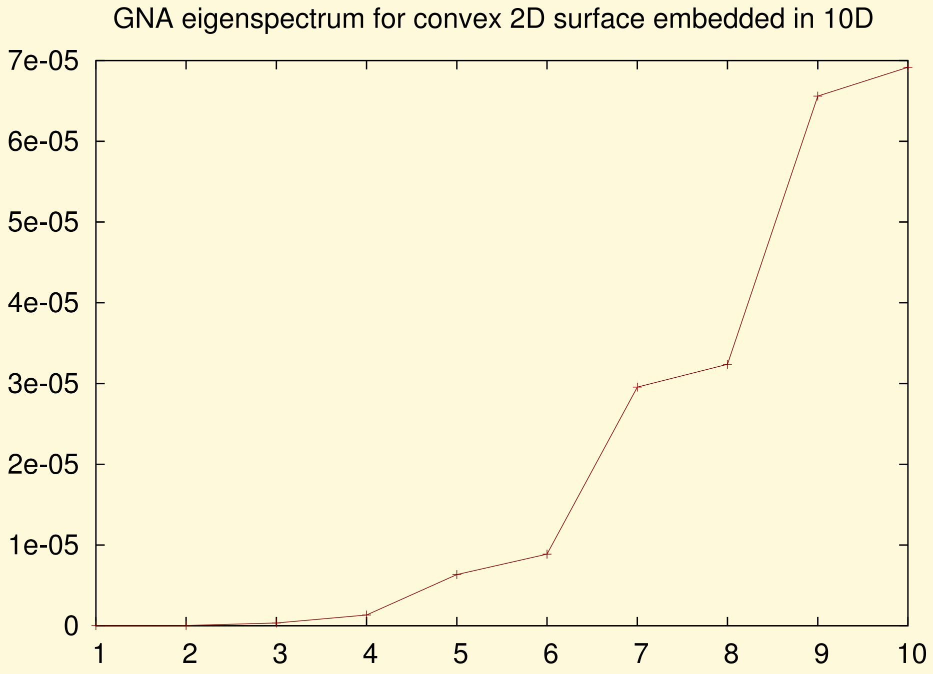
Harmonic View



Remaining eigenvectors of Laplacian-like $\mathbf{K}\mathbf{K}^\top$ give harmonic basis for curvature & noise in extrinsic embedding.

Harmonic View II





Nonzero eigenvalues = bending energy at higher frequencies...stairstep spectrum for convex boundary.

Generalization

- NLDR good for batch visualization...
  ...signal-processing requires inferences about new points.

Two routes:

Extensions: Map new point w.r.t. existing immersion.

Updates: Change immersion w.r.t. new point.

Out-of-sample extensions

Given thin SVD $\mathbf{U}^\top\mathbf{X}\mathbf{V} = \mathbf{S}\ \ (\mathbf{U}\mathbf{S}\mathbf{V}^\top \approx \mathbf{X})$,

PCA projection of $\mathbf{X}$ into column-subspace $\mathbf{U}$:

$$\mathbf{U}^\top\mathbf{X} = \mathbf{S}\mathbf{V}^\top = (\mathbf{S}^+\mathbf{V}^\top\mathbf{V}\mathbf{S})\mathbf{S}\mathbf{V}^\top = \mathbf{S}^+\mathbf{V}^\top(\mathbf{V}\mathbf{S}^2\mathbf{V}^\top) = \mathbf{S}^+\mathbf{V}^\top\mathbf{X}\mathbf{X}^\top$$

= row-space encoding trick. Used in Eigenfaces [Turk and Pentland, 1991], kernel PCA [Schölkopf et al., 1998].

Extend kernel to new point $\mathbf{x}$, and compute vector

$$\mathbf{k}(\mathbf{x}) \in \mathbb{R}^N : \quad k_i = \phi(\mathbf{x})^\top\phi(\mathbf{x}_i).$$

$$\text{Then} \quad \mathbf{y}_\phi(\mathbf{x}) = \mathbf{S}^+\mathbf{V}^\top\mathbf{k}.$$

a.k.a. Nyström extension [Bengio et al., 2003].

Out-of-sample GNA

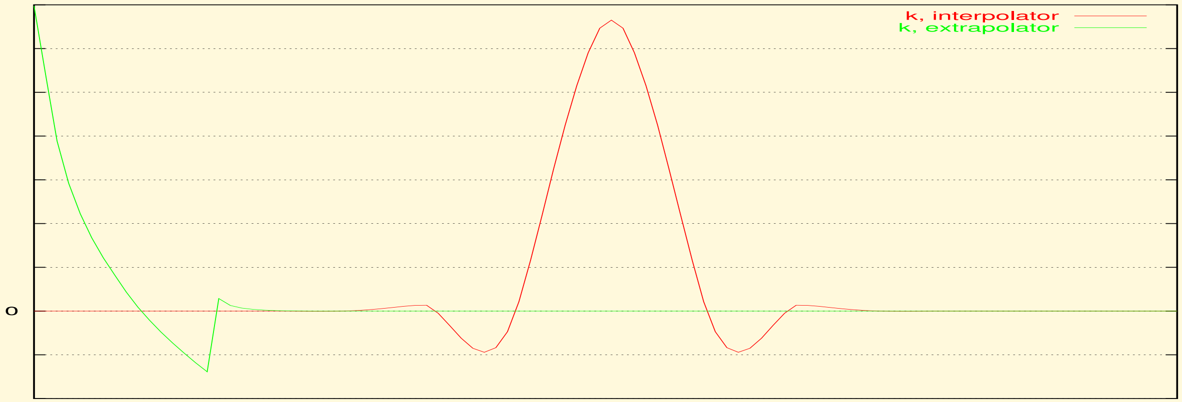
Assign new $\mathbf{x}$ to nearby existing cliques, then build new smaller $\mathbf{K}'$ from $\mathbf{x}$ and original clique points $\mathbf{x}_i, \mathbf{x}_j, \cdots$.

$$\text{Define} \quad \mathbf{k}(\mathbf{x}) \doteq -\mathbf{K}'\mathbf{k}_\circ'^\top / \|\mathbf{k}_\circ'\|^2,$$

with $\mathbf{k}_\circ'$ = row of $\mathbf{K}'$ corresponding to $\mathbf{x}$.

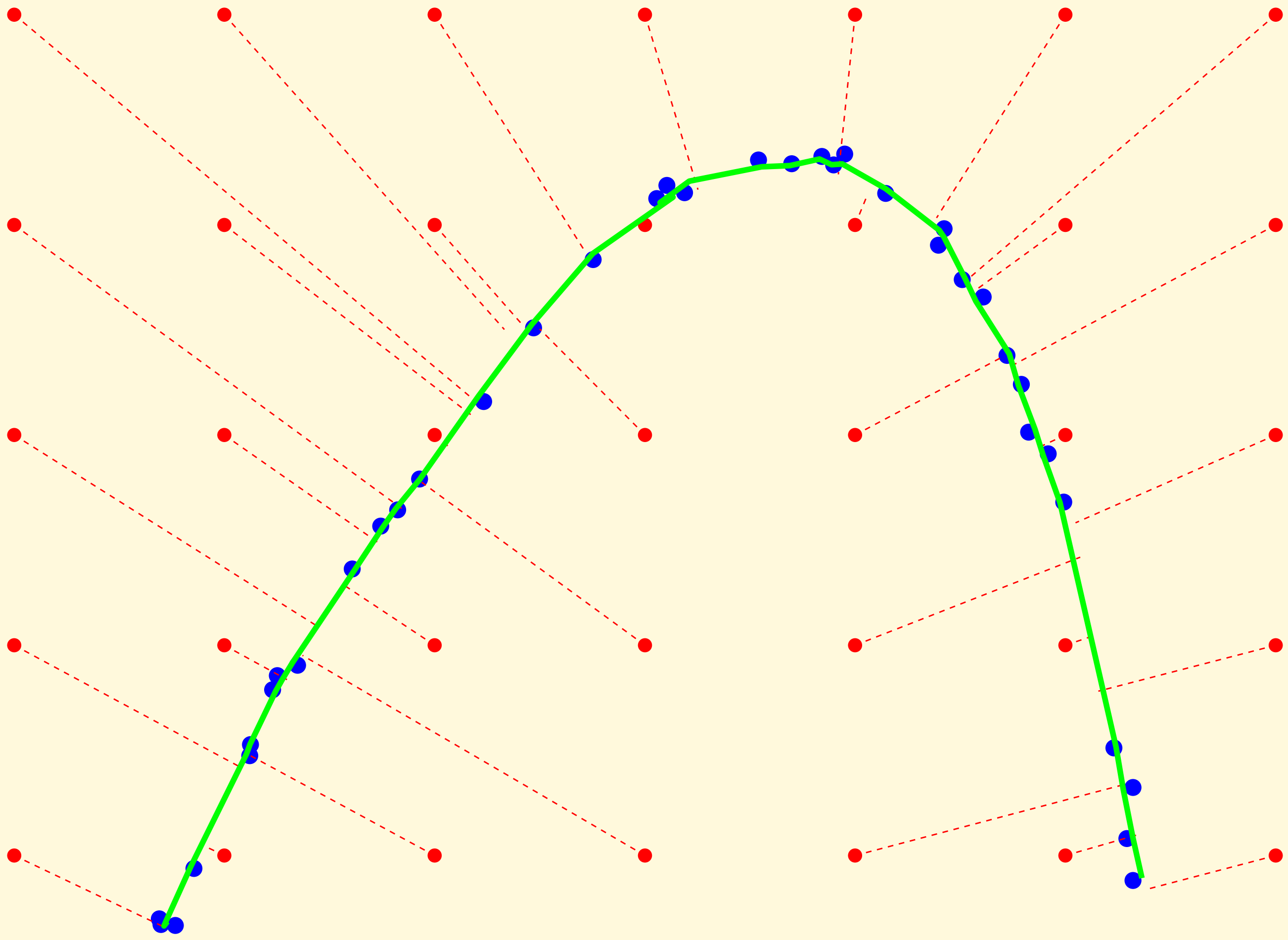
$$\text{Then} \quad \mathbf{y}(\mathbf{x}) = [\mathbf{0}, \mathbf{y}_i, \mathbf{y}_j, \cdots]\mathbf{k}(\mathbf{x}).$$

$\mathbf{k}$ sets $\mathbf{y}$ so Laplacian of immersion w.r.t. local coordinates$= 0$:



Out-of-sample denoising

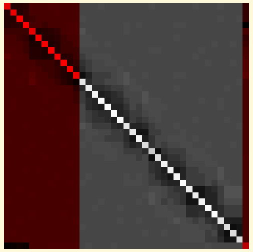
Under local linearity assumption, $\mathbf{k}$ also suitable for denoising and reverse maps, e.g., $\hat{\mathbf{x}} = [\mathbf{0}, \mathbf{x}_i, \mathbf{x}_j, \cdots]\mathbf{k}(\mathbf{x})$:



Updating

Lanczos methods for computing thin EVD well suited to rank-1 updates, e.g., of $\mathbf{K}\mathbf{K}^\top + \mathbf{v}\mathbf{v}^\top$

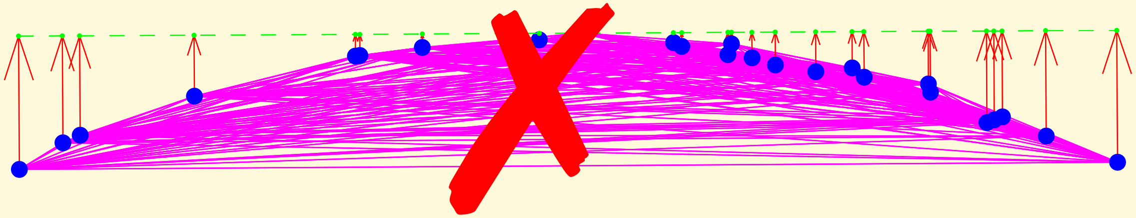
Proposition: To add a new point to cliques containing a total of k points, no more than $2k$ updates needed.



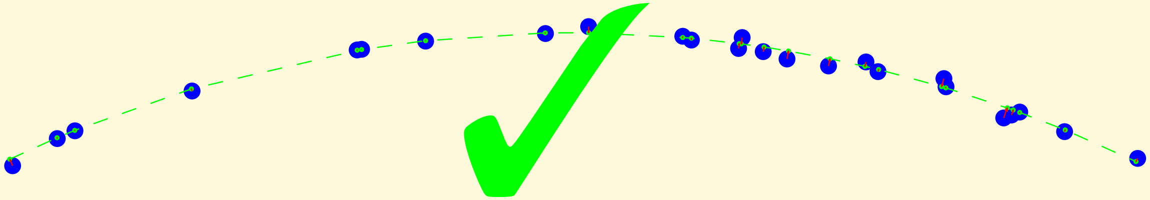
Caution: Should keep around extra eigenvectors if geometry changes substantially.

Local Parameterizations

- 1st-order (linear) models cause distortion.



- 2nd-order algebraic surfaces can fit curved cliques exactly.

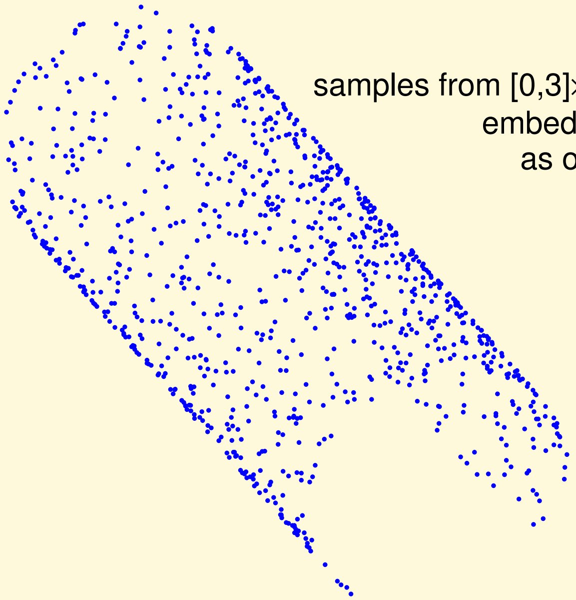


- $\exists$ subclass of quadric manifolds with local isometry to $\mathbb{R}^d$ that can be recovered from samples.

Product manifolds

Example: tube = circle × line

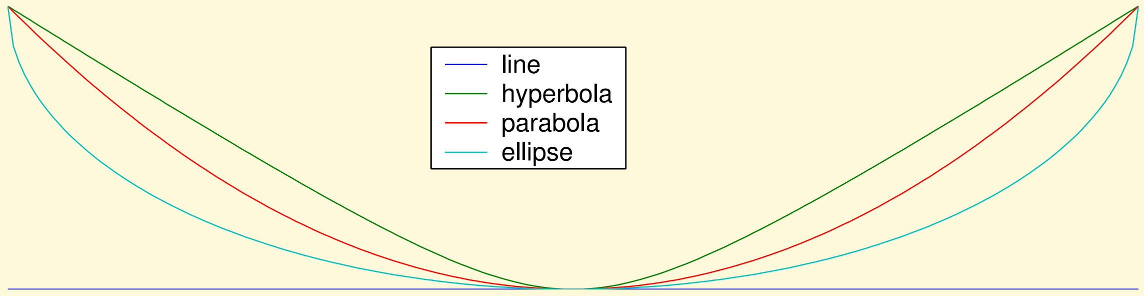
- Inherits coordinate system from factors.
- Arc-lengths are geodesic.



samples from [0,3]×[0,4]⊂R
embedded in R
as open tube

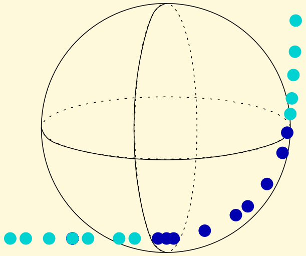
Product of planar quadrics

- Orthogonal product of d planar curves is isometric to $\mathbb{R}^d$.
  - Use quadric curves.

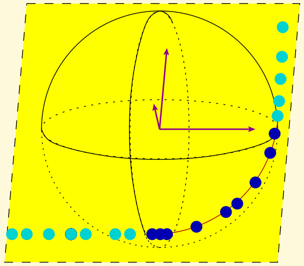


- Most developable surfaces are locally PPQ.
- Data factorization reveals planes.

PPQ factorization from data



Fit osculating quadric $Q_{\mathbf{p}} \subset \mathbb{R}^{2d} : \begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix}^{\top} \mathbf{F}_{\mathbf{p}} \begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix} = 0$

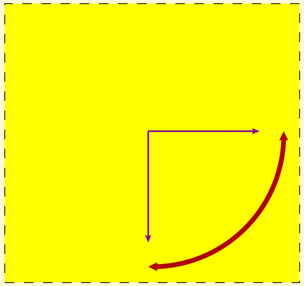


Rotate $\mathbf{x}' \longrightarrow \mathbf{x}''$ to sum-of-quadratics form:

$$Q_{\mathbf{p}} : \sum_{i \in \mathrm{dims}(\mathbf{x}'')} a_i x_i''^2 + 2 b_i x_i'' = c.$$

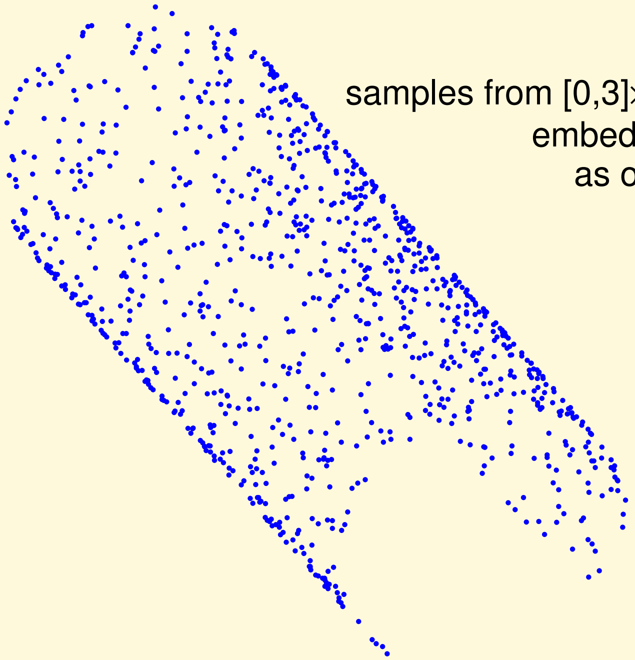
Thm: Data reveals pairing of dimensions into planar quadrics:

$$a_i x_i''^2 + b_i x_i'' + a_j x_j''^2 + b_j x_j'' = c_{ij}.$$

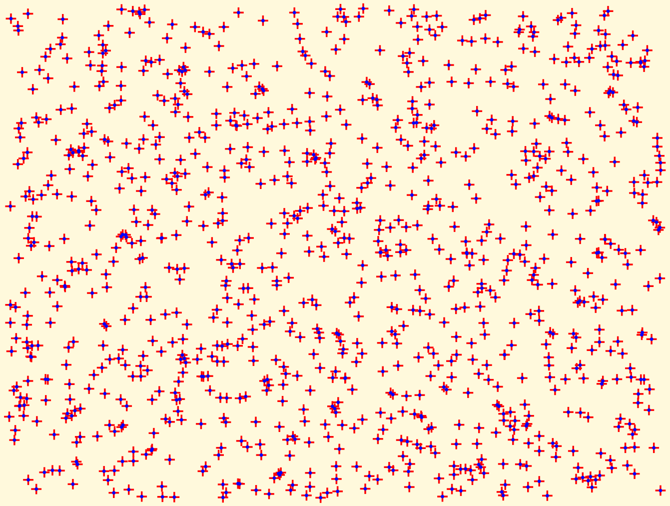


Separate planes & integrate arc lengths.

Isometric parameterization



samples from $[0,3]\times[0,4]\subset R^2$
embedded in R^3
as open tube

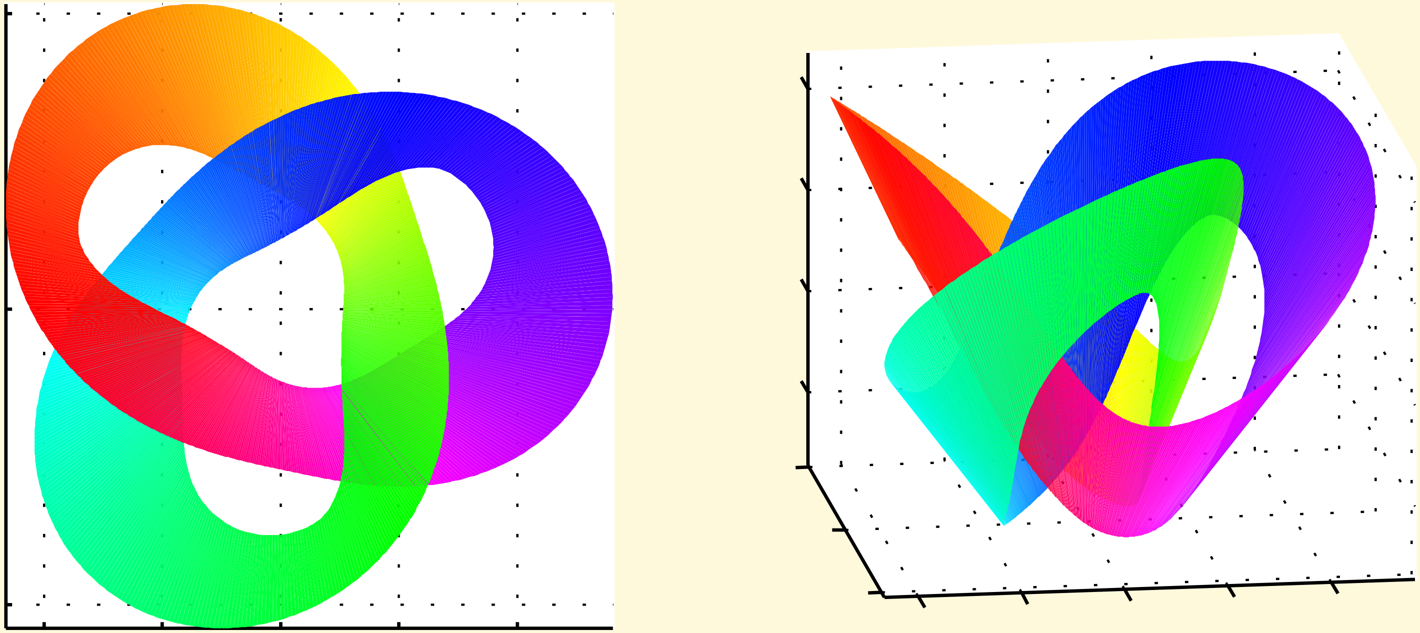


+original & •recovered parameterizations

PPQ arc-length parameterizations + nullspace coordination.

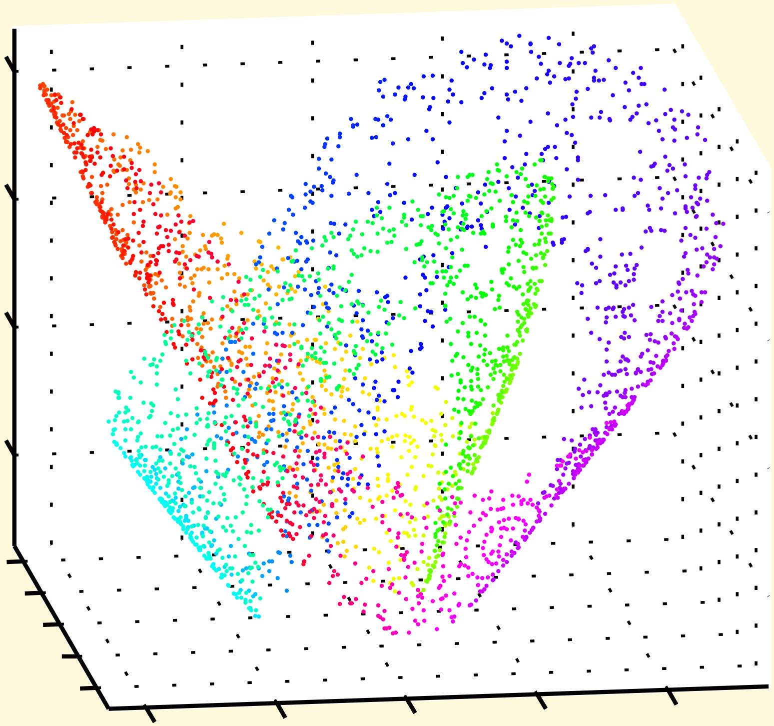
- Recovers geodesic pre-image *exactly*.

Example: Trefoil knot

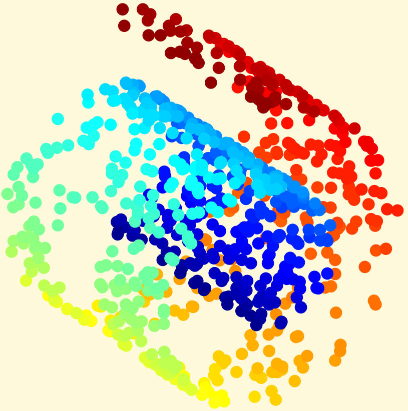


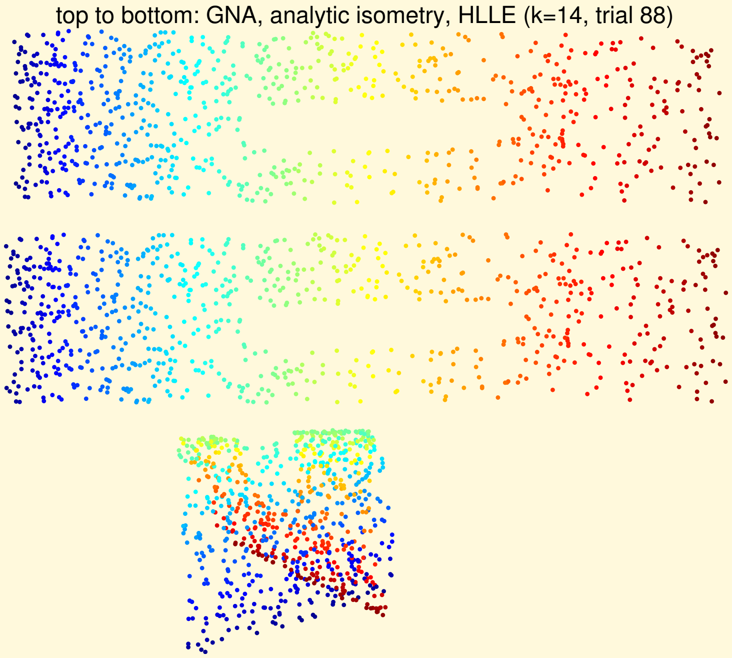
Proposition: Any knot $\mathbb{S}^1 \subset \mathbb{R}^3$ can be thickened to a ribbon manifold having a locally isometric embedding in $\mathbb{R}^2$.

Trefoil knot embedding

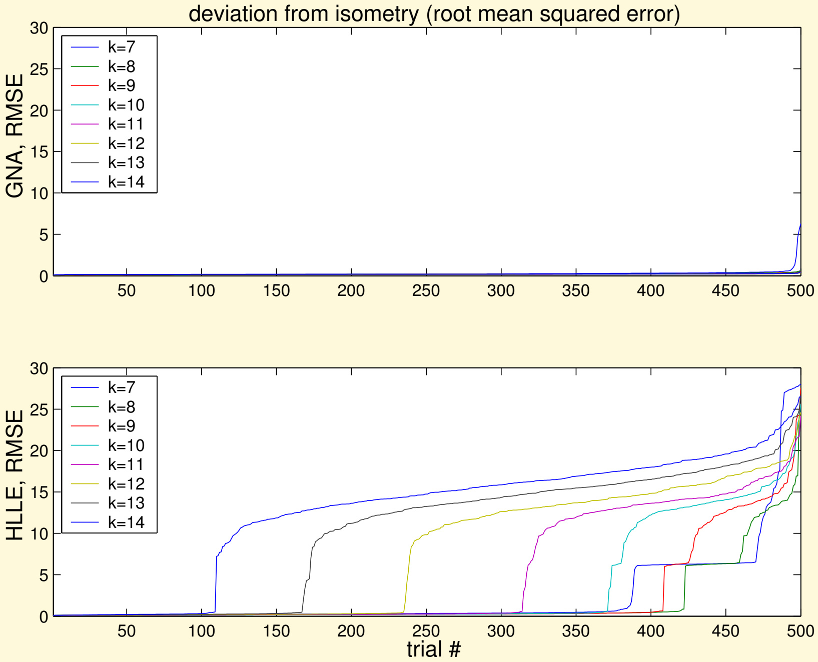


Perforated swiss roll, 1% noise



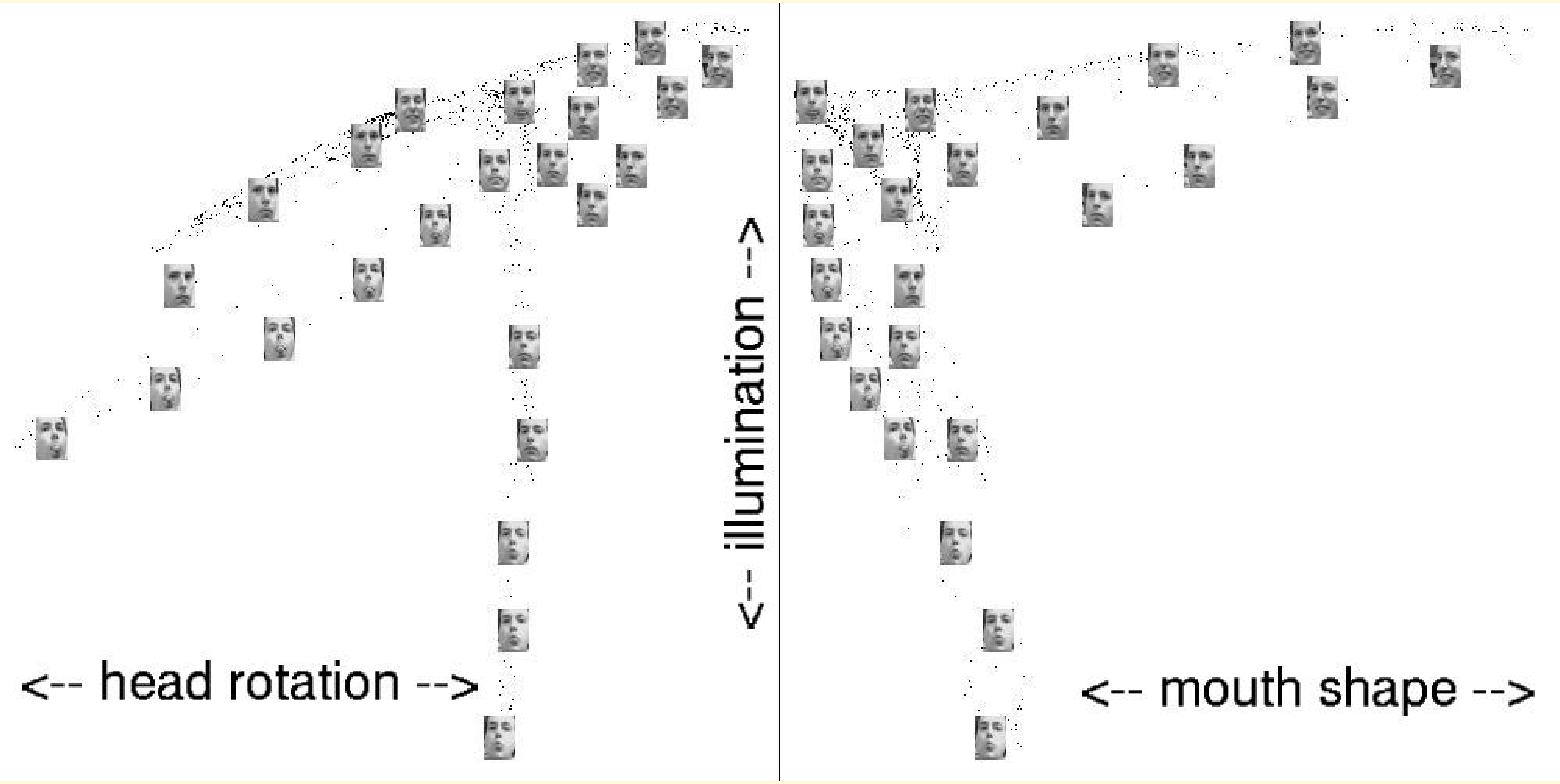


top to bottom: GNA, analytic isometry, HLLE (k=14, trial 88)



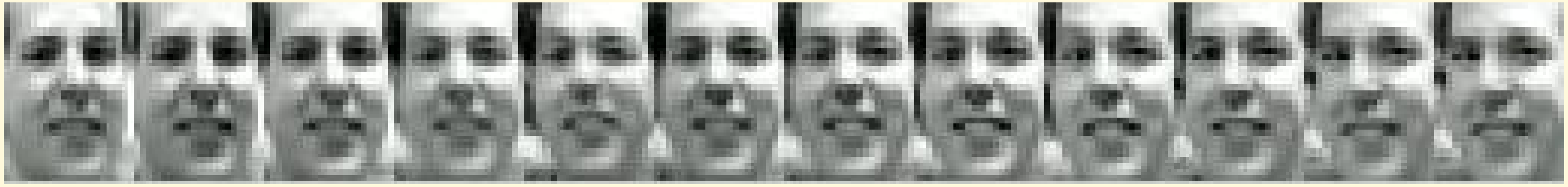
deviation from isometry (root mean squared error)

Frey faces



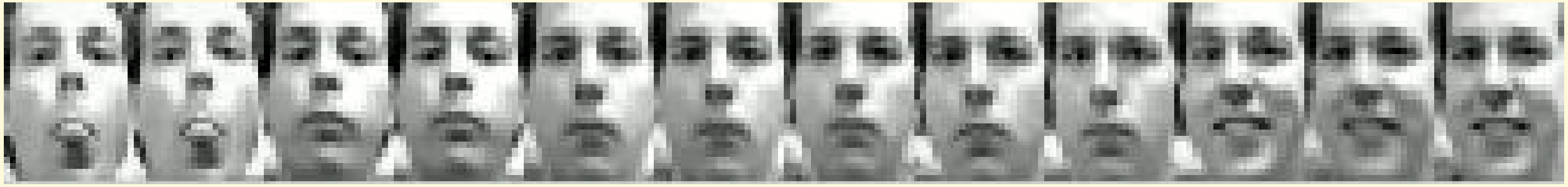
Synthetic face images

pose



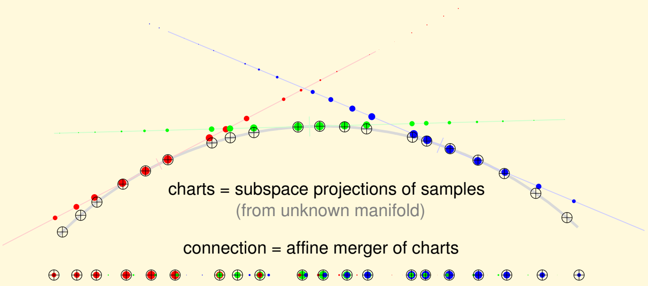
brightness (tilt?)

shape



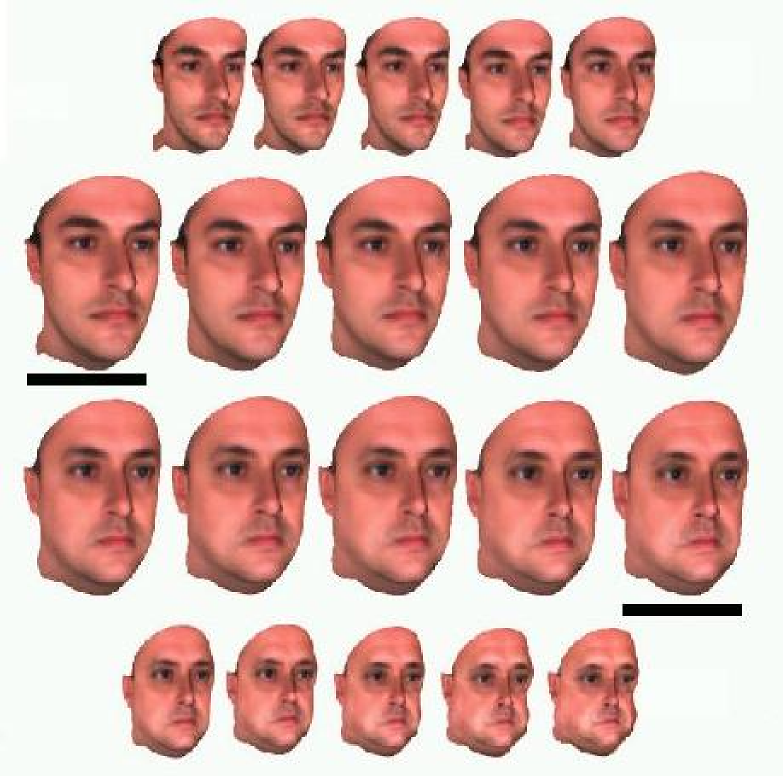
USF 3D Cyberscanned faces

Charting (=gaussian-weighted mixture of linear projections minimizing GNA-like error).



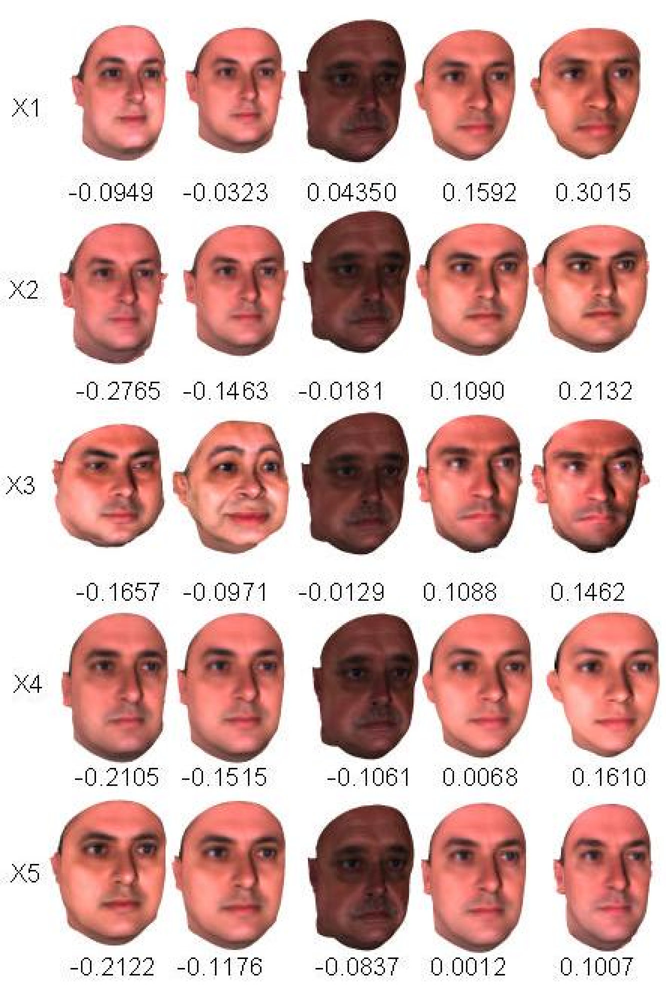
charts = subspace projections of samples
(from unknown manifold)

connection = affine merger of charts



Interpolation and extrapolation along a line on face manifold.

Degrees of freedom



triangularity ↻

wideness

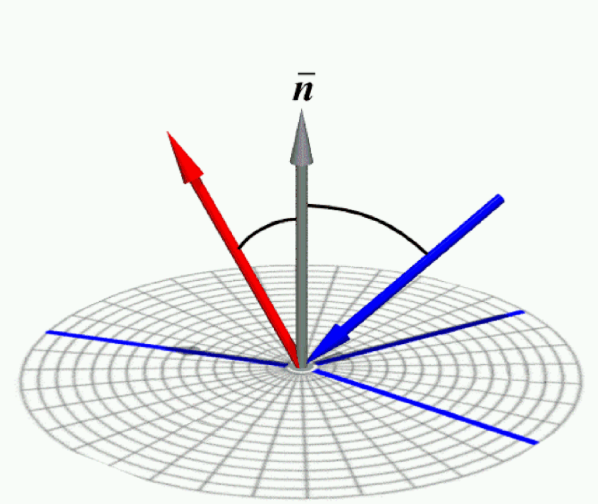
jowliness

gender

roundness

Bidirectional Reflectance Functions

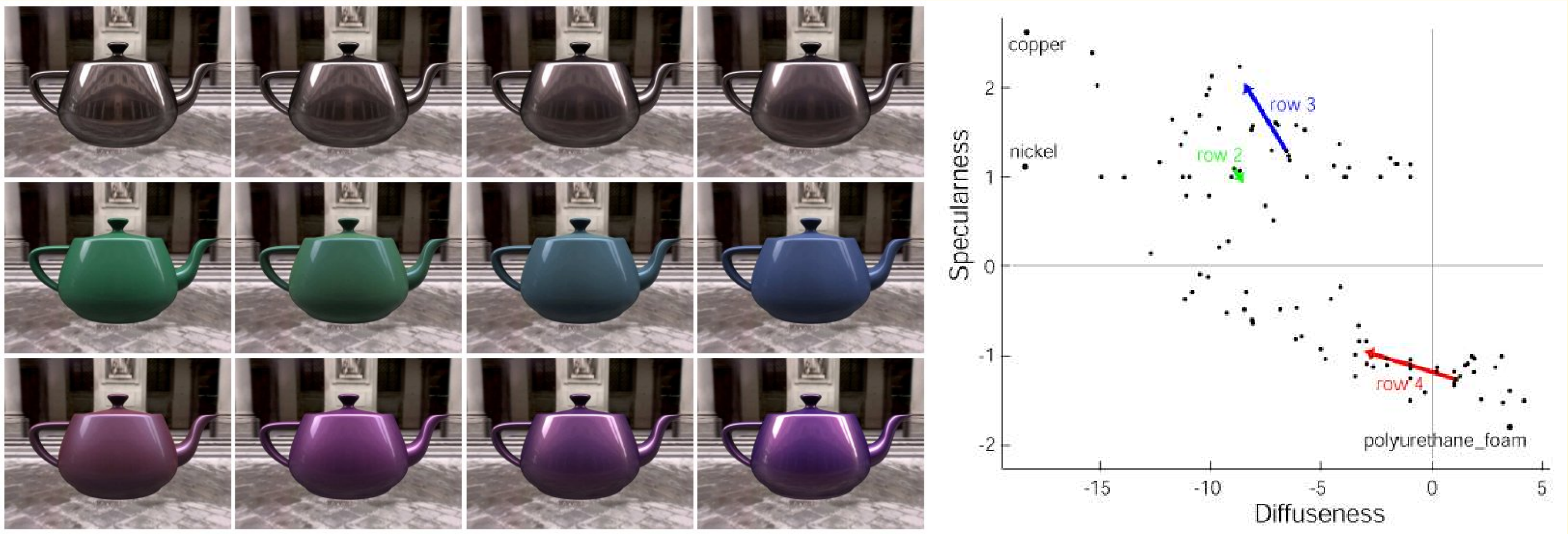
- Each sample an $\mathbb{R}^{10^7}$ vector of material reflectance values.



- 15D nonlinear model has reconstruction error < 40D PCA.

PCA versus NLDR

Synthetic BRDF progressions



Rust

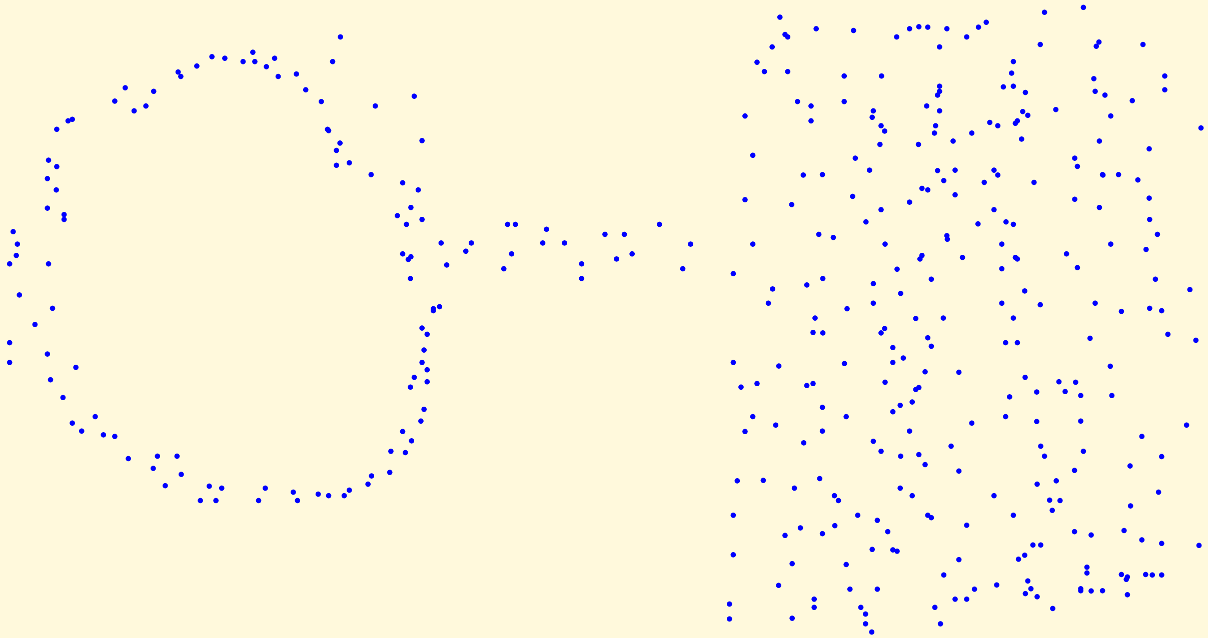
Silver + fingerprints

Features & Caveats

- Geodesic Nullspace Analysis offers almost all of the functionality of subspace methods.
  - reduction, denoising, resynthesis, updating
- Linear in N & D; cubic in d & k.
- Exact local isometry for intrinsically flat piecewise product of planar quadric manifolds.
  - Minimal local nonlinear distortion otherwise.
- Assumes: intrinsic dimension d, approximate local isometry

BUT

How useful is an $\mathbb{R}^3$ parameterization of this manifold?



The CW complex

CW-complex = decomposition of a manifold into cells, each isomophic to an d-ball (varied d).

- A d-cell is connected to $d-1$-cells at its boundary.
- Each cell allows an $\mathbb{R}^d$ parameterization.

CW-approximation theorem: Every space is a CW-complex.

CW-complex of a torus

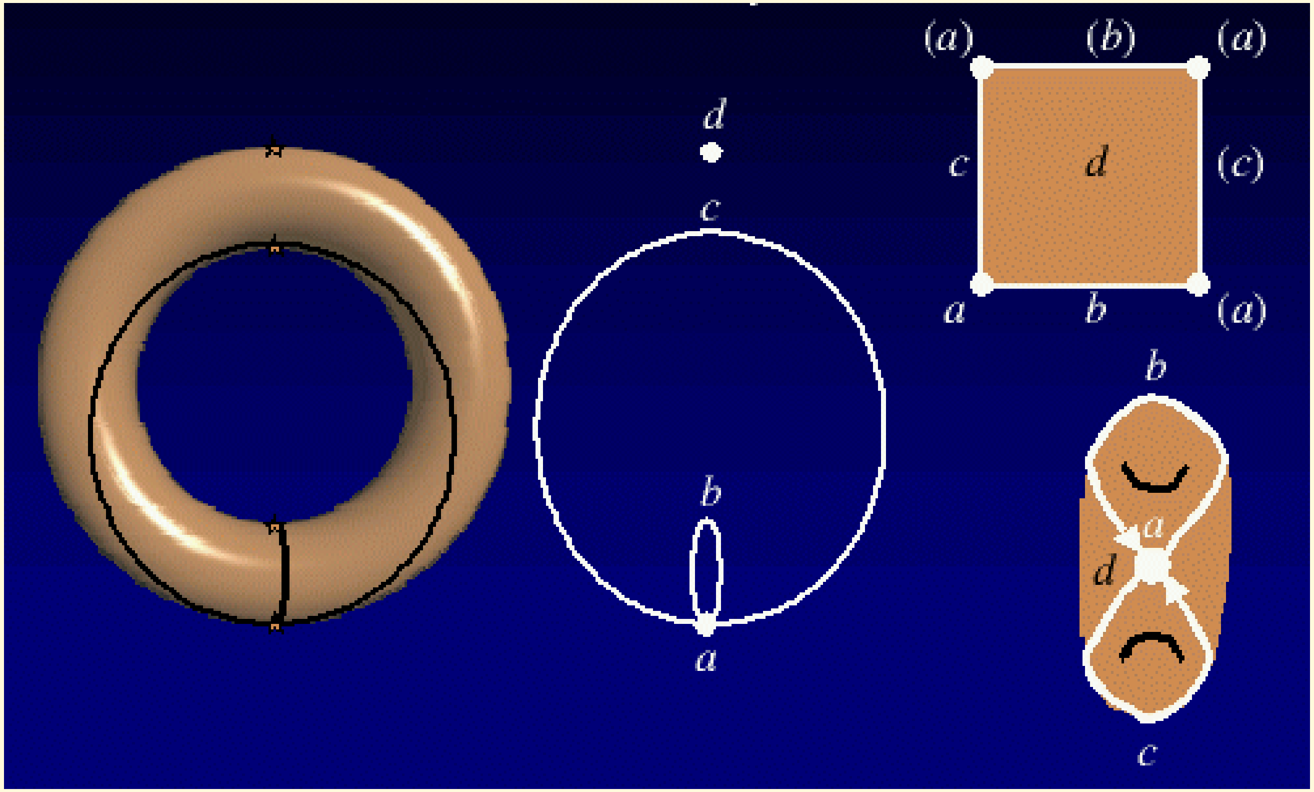


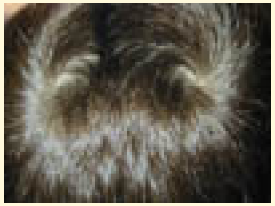
Image courtesy John Hart.

Morse theory

- Consider a smooth C^2 function f on a manifold.

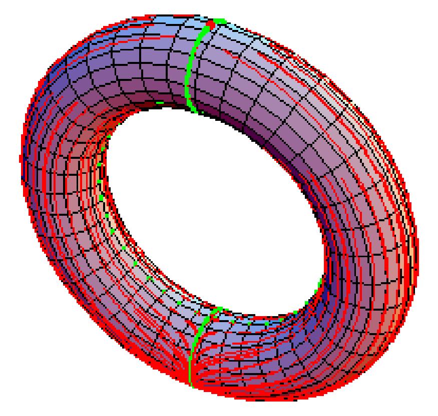
- 

  E.g.,
  $f =$ distance to your crown;
  gradient field $\bigtriangledown f \approx$ your hair.

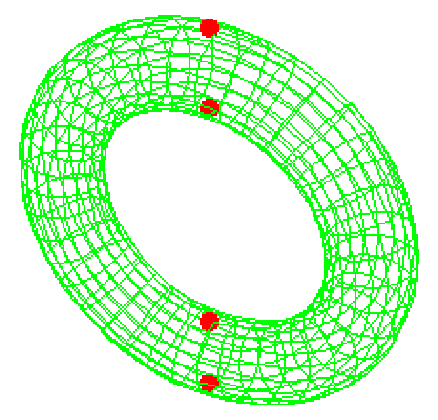
  

- Key elements of a CW-complex are 1-to-1 with nondegenerate critical points of f.
  - =zero gradient, nonzero det(Hessian)

Morse function on a torus



gradient field



critical points

Images courtesy of Andrzej Kozlowski.

Missing in our setting

Point samples do not specify

- the manifold $\mathcal{M}$,
- the Morse function f,
- its parameterization,
- the graph connectivity of the points.

Markov chains on point sets

Consider a random walk on the points.

- Let transition probability

  $$P_{ij} = p(\mathbf{x}_j | \mathbf{x}_i) \propto g(d_{ij})$$

  for exponentially decreasing $g(\cdot) \geq 0$.
  - $P_{ij} \doteq$ probability points i, j locally connected on $\mathcal{M}$.
  - “Short-circuit” exponentially improbable.
- Expected hitting time $H_{ij} =$ mean random walk time from i to j
  - $\mathbf{H}$ marginalizes over all possible connectivities.

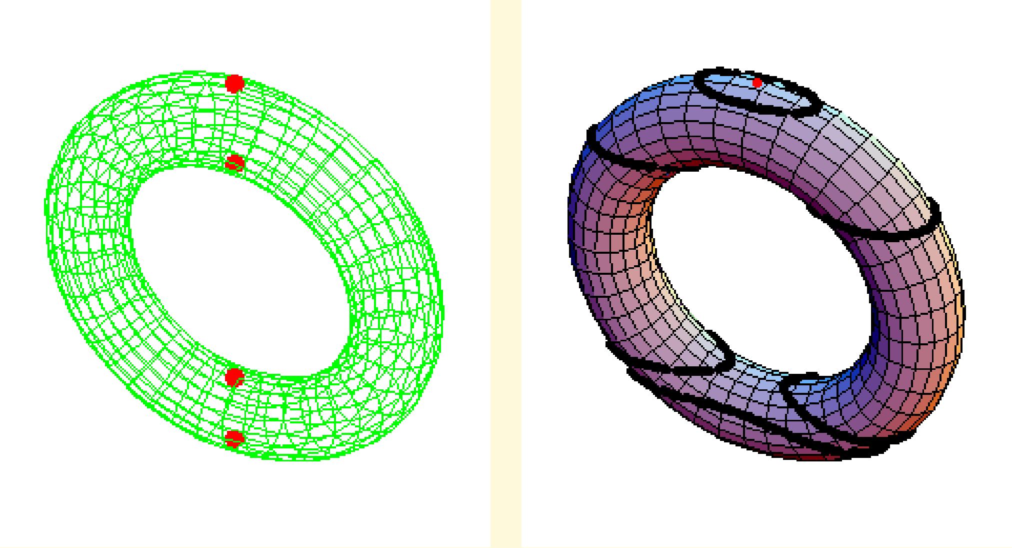
Morse functions on points

Proposition: Any row of $\mathbf{H}$ samples a Morse function on $\mathcal{M}$.

Easy to construct C^2 function $f_{\mathbf{H}}$ via radial basis functions.

Intuition: The level sets of $f_{\mathbf{H}}$ specifies wavefronts on $\mathcal{M}$.

Wavefront topology changes at critical points.



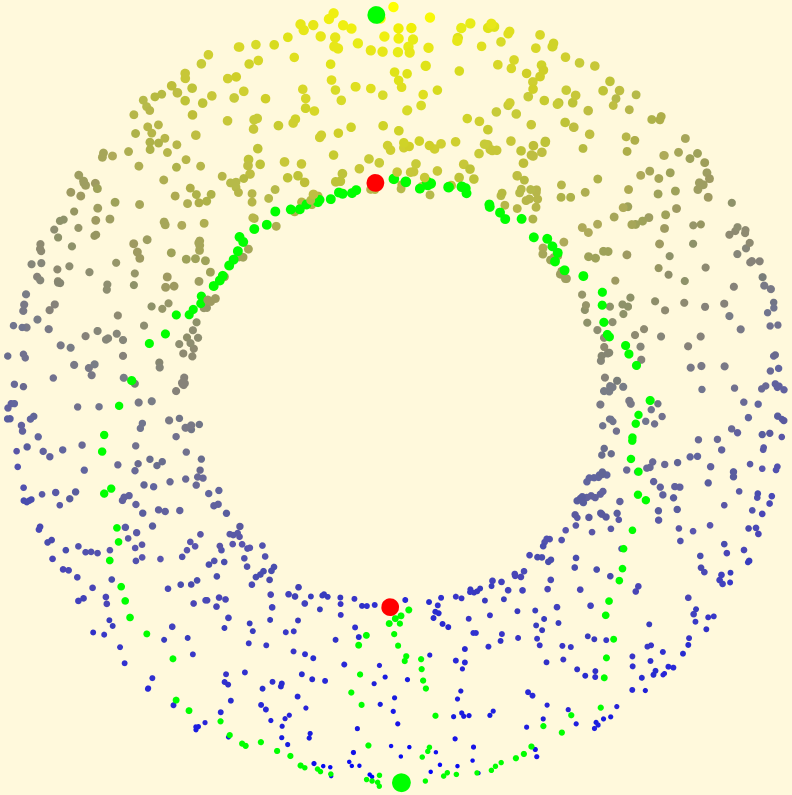
critical points wavefront

Algorithm

- Source point is 0-cell.
- Follow wavefront and mark critical points where connectivity changes.
- Estimate indices (=#(eig(Hessian)<0)) of critical points from $\mathbf{H}$ & local flattenings.
- A d-cell connects $d-1$-cells an index-d critical point via most probable paths...

  ...dynamic programming problem.

Computed torus example



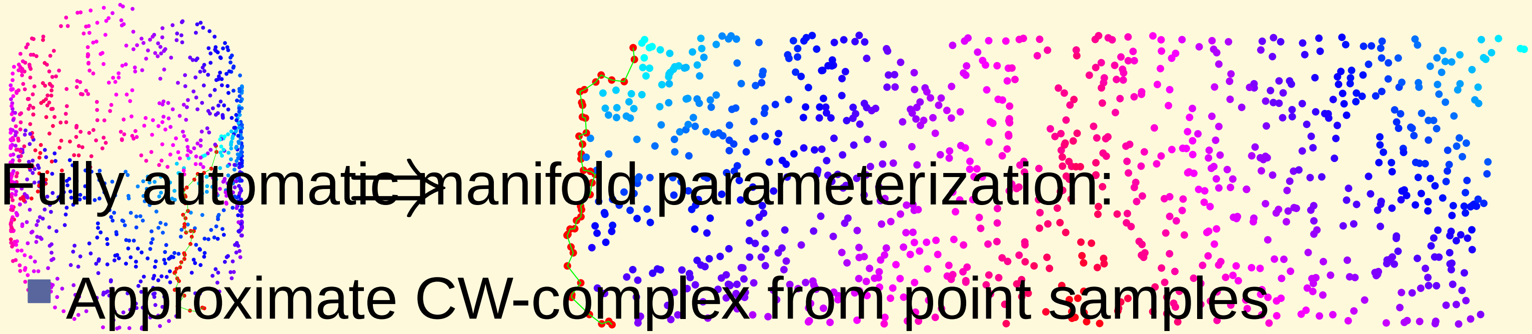
g is a Gaussian kernel.

- CW-complex not necessarily “simplest” form

✔ ∃ full catalog of transformation rules...

✘ ...but simplification may be NP-hard.

Where we’re heading

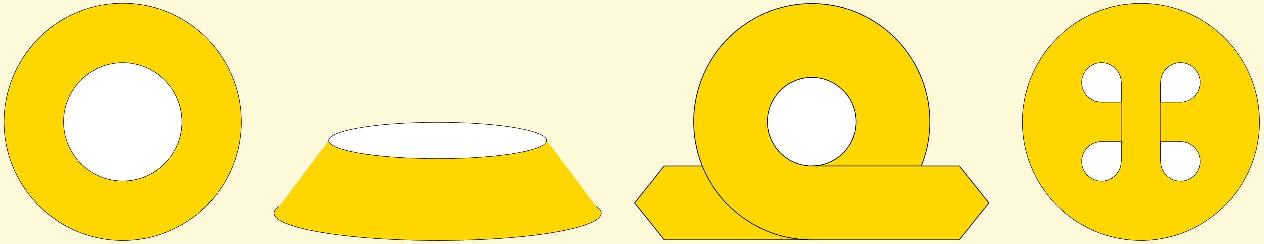


Fully automatic manifold parameterization:

- Approximate CW-complex from point samples
  ...embedding graph partitioned according to cells.
- GNA immersions of subgraphs
  ...signal processing on genuine $\mathbb{R}^d$ submanifolds.

Immersions vs. embeddings

Some manifolds with local isometry to $\mathbb{R}^2$:



immersible...	√	√	√	√
...isometrically	√		√	√
embeddable...	√	√	√	
...isometrically	√			

- Embeddability determined by manifold, not algorithm.
- Local isometry usually needed to preserve data density.

See also

References

Aldous, D. and Fill, J. (In prep.). Reversible markov chains and random walks on graphs. Manuscript, http://www.stat.berkeley.edu/users/aldous/RWG/book.html.

Belkin, M. and Niyogi, P. (2002). Laplacian eigenmaps for dimensionality reduction and data representation. volume 14 of *Advances in Neural Information Processing Systems*.

Bengio, Y., Paiement, J., and Vincent, P. (2003). Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering. In *Advances in Neural Information Processing Systems*, volume 15.

Brand, M. (2003). Charting a manifold. In *Advances in Neural Information Processing Systems*, volume 15.

Donoho, D. L. and Grimes, C. (2003). Hessian eigenmaps. *Proceedings, National Academy of Sciences*.

Fiedler, M. (1975). A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czech. Math. Journal*, 25:619–633.

Kamada, T. and Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, (31):7–15.

Lovasz, L. (1995). *Handbook of Combinatorics*, chapter Random Walks, Eigenvalues and Resistance, pages 1740–1748. MIT Pres.

Mohar, B. (1991). The laplacian spectrum of graphs. In Alavi, Y., editor, *Graph Theory, Combinatorics and Applications*, pages 871–898. J. Wiley, New York.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326.

Schölkopf, B., Smola, A., and Müller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Tutte, W. (1963). How to draw a graph. *Proc. London Mathematical Society*, 13:743–768.

Weinberger, K. Q., Sha, F., and Saul, L. K. (2004). Learning a kernel matrix for nonlinear dimensionality reduction. In *Proc. 21st ICML*.

Zhang, Z. and Zha, H. (2003). Nonlinear dimension reduction via local tangent space alignment. In *Proc., Conf. on Intelligent Data Engineering and Automated Learning*, number 2690 in Lecture Notes on Computer Science, pages

Extra slides begin here

Graph embeddings

Theorem: [Tutte, 1963] Any planar graph can be embedded inside a convex boundary in $\mathbb{R}^2$ w/ straight uncrossed edges.

- Least-squares solution: place vertices at barycenters of their neighbors.

Theorem: [Fiedler, 1975] First nonuniform eigenvector of graph Laplacian matrix gives optimal $\mathbb{R}^1$ embedding.

- [Mohar, 1991]: Use more eigenvectors for more dimensions.

Social networks, spectral clustering literatures: No graph? Invent one.

NLDR: Can mix metric constraints with connectivity.

Some context

Interpret immersion error metric $\mathbf{KK}^\top$ as graph Laplacian $\mathbf{L}$. $\mathbf{L}$ has same eigenvectors as centered commute-time matrix [Lovasz, 1995; Aldous and Fill, prep]

$$(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top)\mathbf{C}(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top)$$

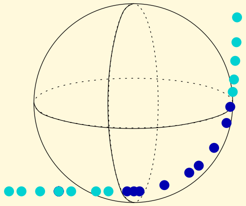
on associated Markov chain, & kernel matrix

$$(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top) - \mathbf{KK}^\top$$

$\therefore$ commute times $\sim d(\cdot,\cdot)^2$ in kernel feature space $\sim d(\cdot,\cdot)^2$ in isometric immersion.

PPQ factorization from data I

Locally, PPQ $\mathcal{M}$ fits into $\mathbb{R}^{2d}$ affine subspace.



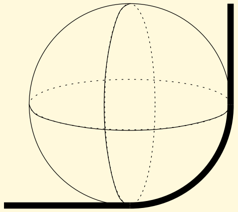
1. Local PCA: $\mathbf{x} \longrightarrow \mathbf{x}' \in \mathbb{R}^{2d}$

2. Fit quadric manifold $Q_{\mathbf{p}}$: $\begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix}^{\top} \mathbf{F}_{\mathbf{p}} \begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix} = 0$ to data:

$$\mathrm{vec}(\mathbf{F}_{\mathbf{p}}) = \mathrm{null}\left[\begin{bmatrix} \mathbf{x}'_1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{x}'_1 \\ 1 \end{bmatrix}, \begin{bmatrix} \mathbf{x}'_2 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{x}'_2 \\ 1 \end{bmatrix}, \cdots \right]$$

PPQ factorization from data II

Locally $\mathcal{M}$ is submanifold of $Q_{\mathbf{p}}$.



3. Rotate $\mathbf{x}' \longrightarrow \mathbf{x}''$ to partially diagonalize $\mathbf{F}_{\mathbf{p}} \longrightarrow \begin{bmatrix} \mathrm{diag}(\mathbf{a}) & \mathbf{b} \\ \mathbf{b}^\top & -c \end{bmatrix}$

giving implicit equation

$$Q_{\mathbf{p}} : \sum_{i \in \mathrm{dims}(\mathbf{x}'')} a_i x_i''^2 + 2 b_i x_i'' = c$$

4. Thm: With $O(d^2)$ points, one can group dimensions to $Q_{\mathbf{p}}$ into a set of independent implicit equations

e.g., $a_i x_i''^2 + b_i x_i'' + a_j x_j''^2 + b_j x_j'' = c_{ij},$

revealing local factorization of $\mathcal{M}$ into product of lines, planar quadric curves, & cones.

PPQ downsides

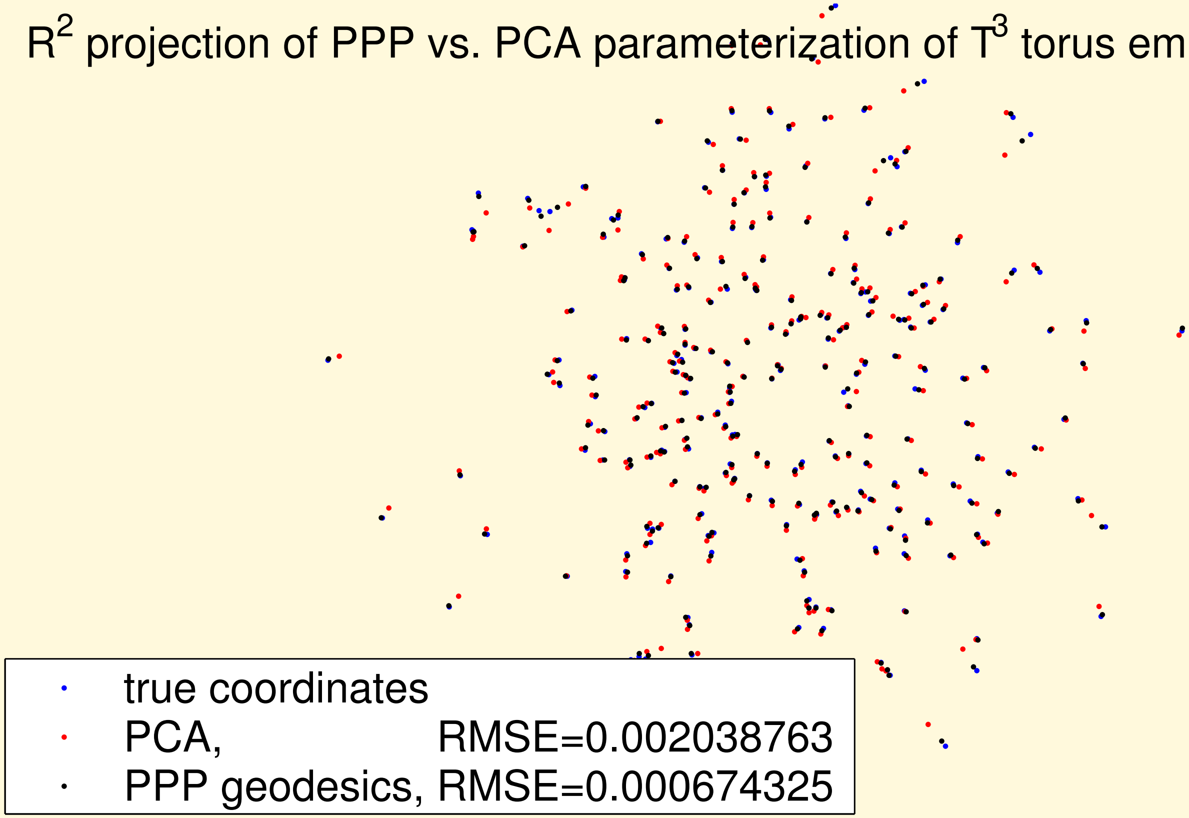
- Error is algebraic – robustness to noise unknown.
- More square root operations (EVDS) reduces significant bits.
- Arc-length integrals analytic only for circles & parabolas (resort to PPP factorization).

PPQ upsides

- Offers *exact* local parameterizations for a large and industrially useful class of manifolds.
- Can parameterize large cliques ... more stability.

Noise

- Outperforms PCA in low-noise, non-PPQ problems.



R^2 projection of PPP vs. PCA parameterization of T^3 torus embedded in R^{1000}

- Dimensionality estimation, local & global.
- Eliminate (or marginalize out) the graph.
- Embeddings versus immersions.