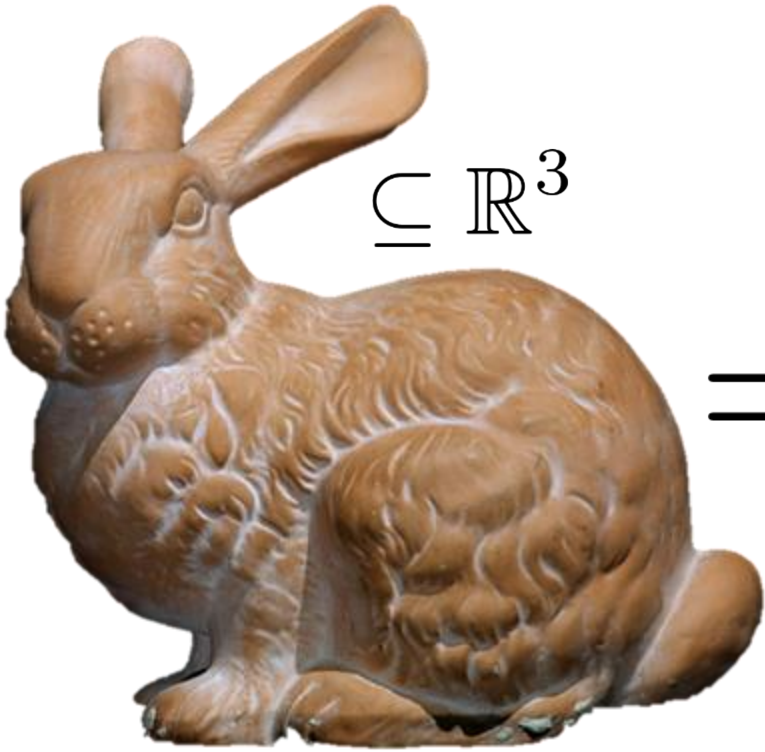
http://thepioneerwoman.com/food-and-friends/chocolate-swiss-roll-cake/

Inverse Distance Problems

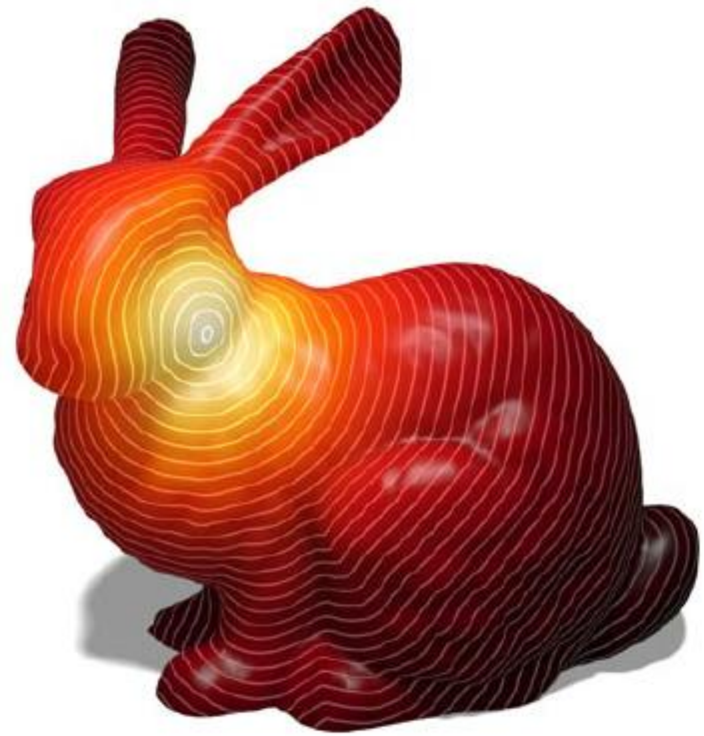
Justin Solomon
MIT, Spring 2017

Last Time



$$\subseteq \mathbb{R}^3$$

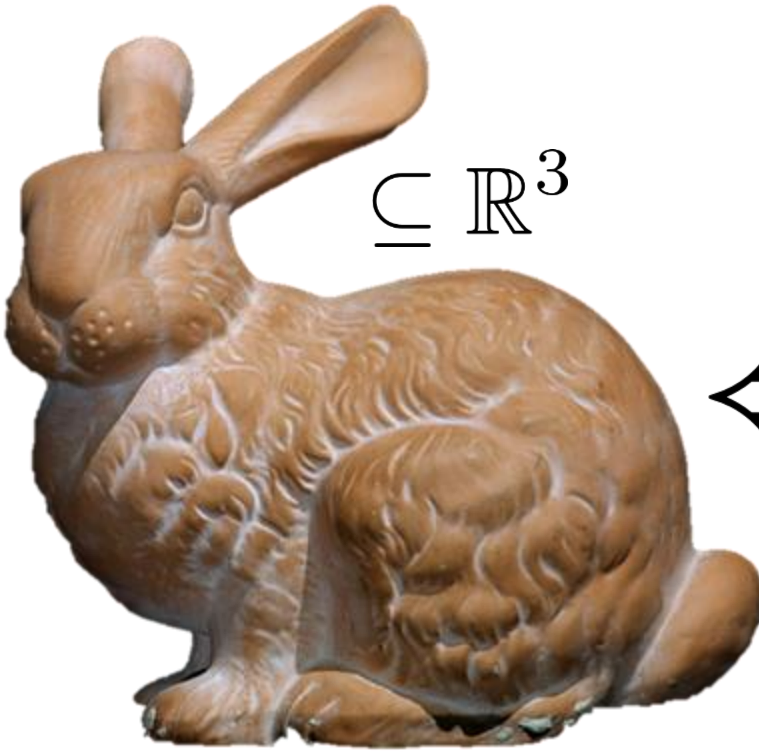
$$\Longrightarrow$$



Embedding | **Geodesic distance**

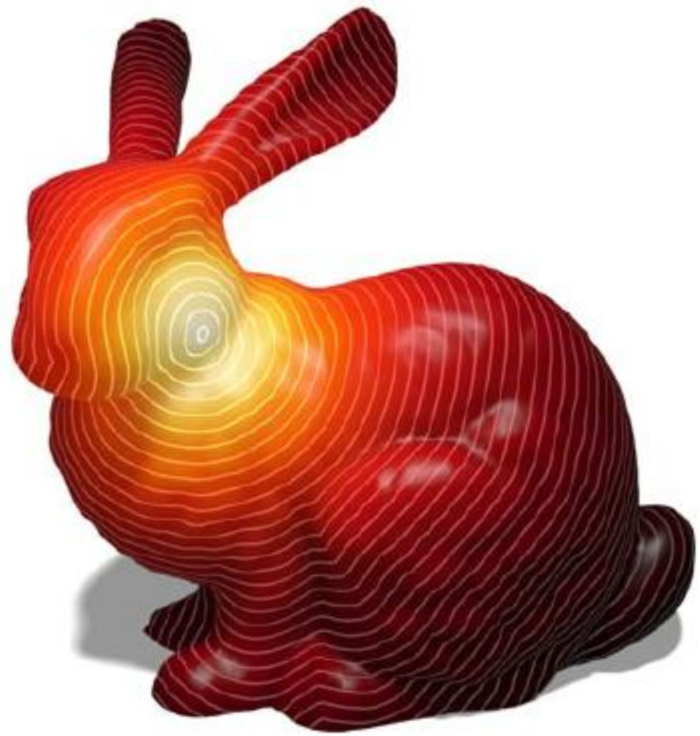
Right bunny from "Geodesics in Heat" (Crane et al.)

Today



$\subseteq \mathbb{R}^3$

$\Longleftarrow$



Embedding

Geodesic distance

Right bunny from "Geodesics in Heat" (Crane et al.)

Many Names

- Dimensionality reduction
- Embedding
- Multidimensional scaling
- Manifold learning

…

Basic Task

Given pairwise distances extract an embedding.

Is it always possible?
What dimensionality?

Metric Space

Ordered pair (M, d) where M is a set and d: $M \times M \to \mathbb{R}$ satisfies

$$d(x,y) \geq 0$$

$$d(x,y) = 0 \iff x = y$$

$$d(x,y) = d(y,x)$$

$$d(x,z) \leq d(x,y) + d(y,z)$$

$$\forall x, y, z \in M$$

Many Examples of Metric Spaces

$$\mathbb{R}^n, d(x,y) := \|x - y\|_p$$

$$S \subset \mathbb{R}^3, d(x,y) := \text{geodesic}$$

$$C^\infty(\mathbb{R}), d(f,g)^2 := \int_{\mathbb{R}} (f(x) - g(x))^2 \, dx$$

Isometry [ahy-**som**-i-tree]**:**

A map between metric spaces that preserves pairwise distances.

Q:

Can you always embed a metric space isometrically in $\mathbb{R}^n$?

Q:

Can you always embed a finite metric space isometrically in $\mathbb{R}^n$?

Disappointing Example

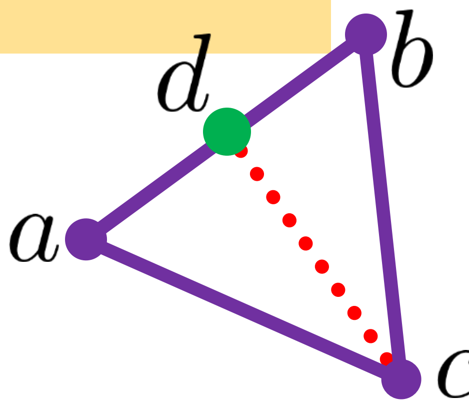
$$X := \{a, b, c, d\}$$

$$d(a, d) = d(b, d) = 1$$

$$d(a, b) = d(a, c) = d(b, c) = 2$$

$$d(c, d) = 1.5$$

Cannot be embedded in Euclidean space!



https://chiasme.wordpress.com/2013/10/07/when-does-a-finite-metric-space-embed-isometrically-into-an-euclidean-space/

Approximate Embedding

$$\text{expansion}(f) := \max_{x,y} \frac{\mu(f(x), f(y))}{\rho(x, y)}$$

$$\text{contraction}(f) := \max_{x,y} \frac{\rho(x, y)}{\mu(f(x), f(y))}$$

$$\text{distortion}(f) := \text{expansion}(f) \times \text{contraction}(f)$$

http://www.cs.toronto.edu/~avner/teaching/S6-2414/LN1.pdf

Well-Known Result

Theorem (Bourgain, 1985).
Let (X,d) be a metric space on n points. Then,

$$(X,d) \xhookrightarrow{O(\log n)} \ell_p^{O(\log^2 n)}$$

$m := 576 \log n)$
for $j = 1$ **to** $\log n$ **do** /* levels of density */
  for $i = 1$ **to** m **do** /* repeat for high probability */
    choose set S_{ij} by sampling each node in X
    independently with probability 2^{-j}
  end
end
$f_{ij}(x) := d(x, S_{ij})$
$f(x) := \bigoplus_{j=1}^{\log n} \bigoplus_{i=1}^{m} f_{ij}(x)$

Of mostly theoretical interest

http://ttic.uchicago.edu/~harry/teaching/pdf/lecture3.pdf

Euclidean Case

$$D_{ij} = \|x_i - x_j\|_2^2, D \in \mathbb{R}^{n\times n}$$

Proposition. $\text{Rank}(D) \leq \min(n, m+2)$.

Proof:

$$D = -2X^\top X + \text{diag}(X^\top X)\mathbf{1}^\top + \mathbf{1}\text{diag}(X^\top X)^\top$$

Embedding via eigenvalue problem (take $x_1 = 0$):

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i \cdot x_j$$

$$\implies x_i \cdot x_j = \frac{1}{2}\left[\|x_i\|_2^2 + \|x_j\|_2^2 - \|x_i - x_j\|_2^2\right]$$

Gram Matrix [gram **mey**-triks]**:**

A matrix of inner products

$$X^\top X$$

Classical Multidimensional Scaling

1. Double centering: $B := -\frac{1}{2}JDJ$
   Centering matrix $J := I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$

2. Find m largest eigenvalues/eigenvectors

3. $X = E_m \Lambda_m^{1/2}$

"MDS"

Torgerson, Warren S. (1958). *Theory & Methods of Scaling*.

Stress Majorization

$$\min_X \sum_{ij} \left(d^0_{ij} - \|x_i - x_j\|_2 \right)^2$$

Nonconvex!

SMACOF:

Scaling by Majorizing a Complicated Function

de Leeuw, J. (1977), "Applications of convex analysis to multidimensional scaling" *Recent developments in statistics*, 133–145.

SMACOF Potential Terms

$$\min_X \sum_{ij} \left(d_{ij}^0 - \|x_i - x_j\|_2\right)^2$$

$$\sum_{ij} (d_{ij}^0)^2 = \text{const.}$$

$$\sum_{ij} \|x_i - x_j\|_2^2 = \text{tr}(XVX^\top), \text{ where } V = 2nI - \mathbf{2}\mathbf{1}\mathbf{1}^\top$$

$$\sum_{ij} d_{ij}^0 \|x_i - x_j\|_2 = \text{tr}(XB(X)X^\top)$$

$$\text{where } b_{ij}(X) := \begin{cases} -\frac{2d_{ij}^0}{\|x_i - x_j\|_2} & \text{if } x_i \neq x_j, i \neq j \\ 0 & \text{if } x_i = x_j, i \neq j \\ -\sum_{j \neq i} b_{ij} & \text{if } i = j \end{cases}$$

SMACOF Lemma

$$\sum_{ij} (d_{ij}^0)^2 = \text{const.}$$

$$\sum_{ij} \|x_i - x_j\|_2^2 = \text{tr}(XVX^\top)$$

$$\sum_{ij} d_{ij}^0 \|x_i - x_j\|_2 = \text{tr}(XB(X)X^\top)$$

$$\text{where } b_{ij}(X) := \begin{cases} -\frac{2d_{ij}^0}{\|x_i - x_j\|_2} & \text{if } x_i \neq x_j, i \neq j \\ 0 & \text{if } x_i = x_j, i \neq j \\ -\sum_{j \neq i} b_{ij} & \text{if } i = j \end{cases}$$

Lemma. Define

$$\tau(X, Z) := \text{const.} + \text{tr}(XVX^\top) - 2\text{tr}(XB(Z)Z^\top)$$

Then,

$$\tau(X, X) \leq \tau(X, Z) \ \forall Z$$

with equality exactly when $X=Z$.

Proof on board using Cauchy-Schwarz.

See *Modern Multidimensional Scaling* (Borg, Groenen)

SMACOF: Single Step

$$X^{k+1} \leftarrow \min_X \tau(X, X^k)$$

$$\begin{aligned}
\tau(X,Z) &:= \text{const.} + \text{tr}(XVX^\top) - 2\text{tr}(XB(Z)Z^\top) \\
\implies 0 &= \nabla_X[\tau(X,X^k)] \\
&= 2XV - 2X^kB(X^k) \\
\implies X^{k+1} &= X^kB(X^k)V^+ \\
V^+ &= (2nI - 2\mathbf{1}\mathbf{1}^\top)^+ \\
&= \frac{1}{2n}\left(I - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right)^+ \\
&= \frac{1}{2n}\left(I - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right)
\end{aligned}$$

Majorization-Minimization algorithm

Objective convergence:
$\tau(X^{k+1}, X^{k+1}) \leq \tau(X^k, X^k)$

Recent SMACOF Application

DOI: 10.1111/cgf.12558
EUROGRAPHICS 2015 / O. Sorkine-Hornung and M. Wimmer
(Guest Editors)

Volume 34 (2015), Number 2

Shape-from-Operator: Recovering Shapes from Intrinsic Operators

Davide Boscaini, Davide Eynard, Drosos Kourounis, and Michael M. Bronstein

Università della Svizzera Italiana (USI), Lugano, Switzerland

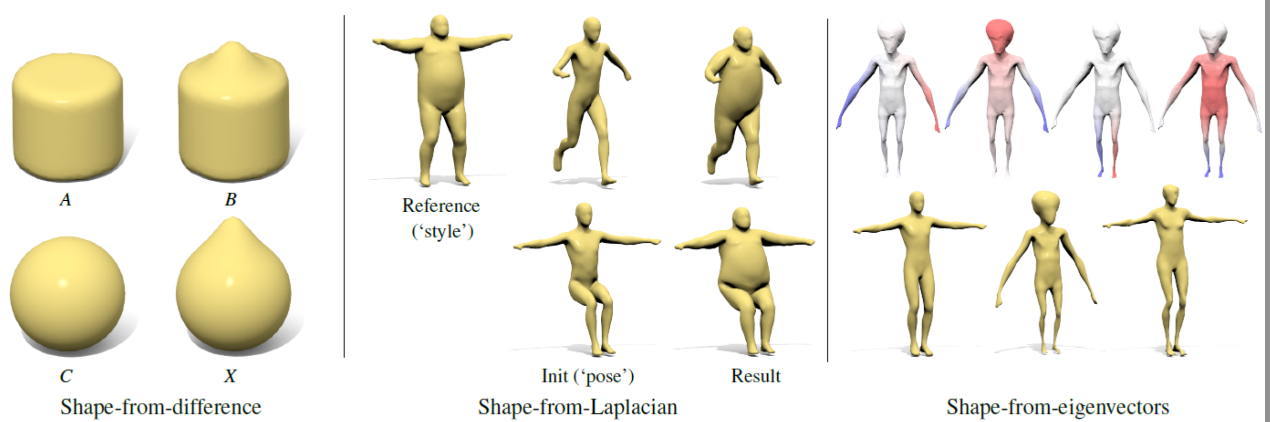
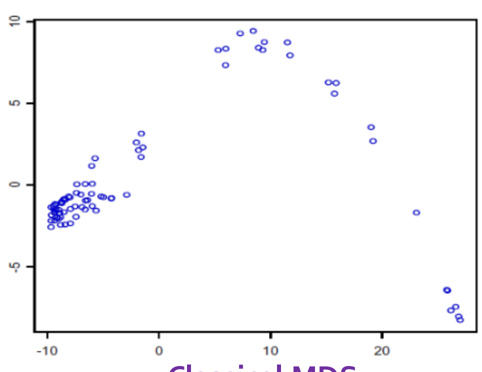


Figure 1: *Examples of three different shape-from-operator problems considered in the paper. Left: shape analogy synthesis as shape-from-difference operator problem (shape X is synthesized such that the intrinsic difference operator between C,X is as close as possible to the difference between A,B). Center: style transfer as shape-from-Laplacian problem. The Laplacian of the*

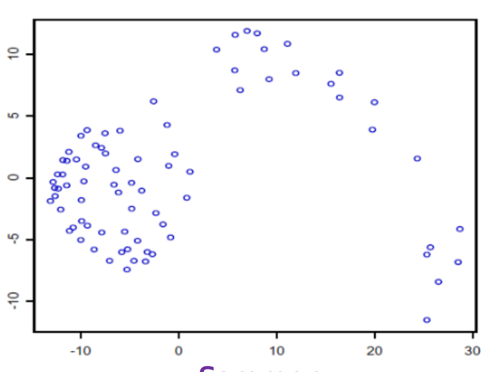
Related Method

$$\min_X \sum_{ij} \frac{\left(d^0_{ij} - \|x_i - x_j\|_2\right)^2}{d^0_{ij}}$$

Cares more about preserving small distances



Classical MDS



Sammon

"Sammon mapping"

Sammon (1969). "A nonlinear mapping for data structure analysis." IEEE Transactions on Computers 18.

http://www.stat.pitt.edu/sungkyu/course/2221Fall13/lec8_mds_combined.pdf

Intrinsic-to-Extrinsic: ISOMAP

- **Construct neighborhood graph**

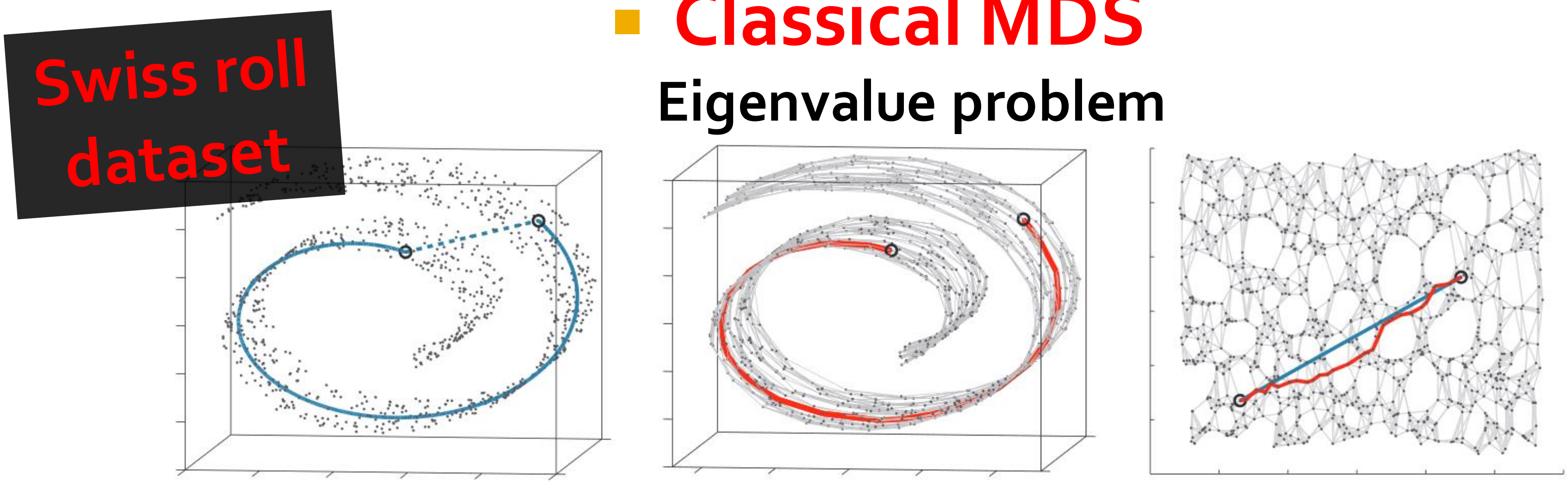
  k-nearest neighbor graph or ε-neighborhood graph

- **Compute shortest-path distances**

  Floyd-Warshall algorithm or Dijkstra

- **Classical MDS**

  Eigenvalue problem



Tenenbaum, de Silva, Langford.
"A Global Geometric Framework for Nonlinear Dimensionality Reduction." Science (2000).

Floyd-Warshall Algorithm

```
let dist be a |V| × |V| array of minimum distances initialized to ∞ (infinity)
for each vertex v
   dist[v][v] ← 0
for each edge (u,v)
   dist[u][v] ← w(u,v)  // the weight of the edge (u,v)
for k from 1 to |V|
   for i from 1 to |V|
      for j from 1 to |V|
         if dist[i][j] > dist[i][k] + dist[k][j]
             dist[i][j] ← dist[i][k] + dist[k][j]
         end if
```

https://en.wikipedia.org/wiki/Floyd%E2%80%93Warshall_algorithm

Landmark ISOMAP

- **Construct neighborhood graph**

  k-nearest neighbor graph or ε-neighborhood graph

- **Compute some shortest-path distances**

  Dijkstra: $O(kn\,N\log N)$, n landmarks, N points

- **MDS on landmarks**

  Smaller $n \times n$ problem

- **Closed-form embedding formula**

  $\delta(x)$ vector of squared distances from x to landmarks

$$\mathrm{Embedding}(x)_i = -\frac{1}{2}\frac{v_i^\top}{\sqrt{\lambda_i}}\left(\delta(x) - \delta_{\mathrm{average}}\right)$$

Projection

Locally Linear Embedding (LLE)

- **Construct neighborhood graph**

k-nearest neighbor graph or *ε*-neighborhood graph

- **Compute weights W_{ij}**

$$\min_{\mathbf{1}^\top W_i = 1} \|x_i - \sum_{x_j \in \mathcal{N}(x_i)} W_{ij} x_j\|_2^2$$

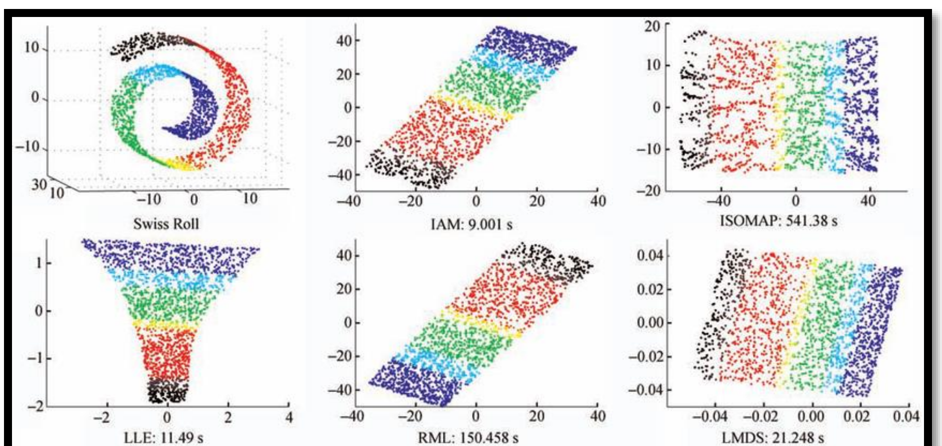
- **Minimum eigenvalue problem**

$$\min_{YY^\top = I} \|y_i - \sum_{x_j \in \mathcal{N}(x_i)} W_{ij} y_j\|_2^2$$

Derive on board

Comparison: ISOMAP vs. LLE

ISOMAP	LLE
Global distances	Local averaging
k-NN graph distances	*k*-NN graph weighting
Largest eigenvectors	Smallest eigenvectors
Dense matrix	Sparse matrix



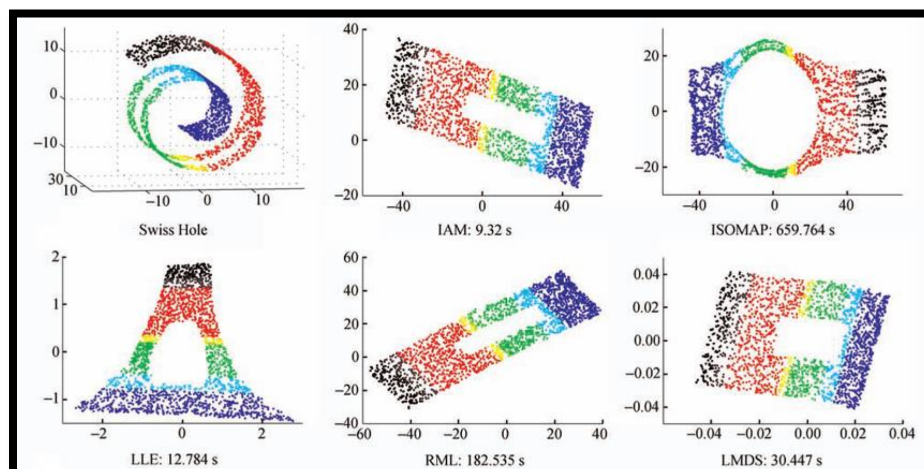


Image from "Incremental Alignment Manifold Learning." Han et al. JCST 26.1 (2011).

Other option:

Diffusion Maps

- **Construct similarity matrix**

Example: $K(x, y) := e^{-\|x-y\|^2/\varepsilon}$

- **Normalize rows**

$$M := D^{-1}K$$

- **Embed from *k* largest eigenvectors**

$$(\lambda_1\psi_1, \lambda_2\psi_2, \ldots, \lambda_k\psi_k)$$

(more later)

Coifman, R.R.; S. Lafon. (2006). "Diffusion maps." *Applied and Computational Harmonic Analysis*. 21: 5–30.

Embedding from Geodesic Distance

On reconstruction of non-rigid shapes with intrinsic regularization

Yohai S. Devir   Guy Rosman   Alexander M. Bronstein   Michael M. Bronstein
Ron Kimmel
{yd|rosman|bron|mbron|ron}@cs.technion.ac.il
Department of Computer Science
Technion - Israel Institute of Technology

Abstract

Shape-from-X *is a generic type of inverse problems in computer vision, in which a shape is reconstructed from some measurements. A specially challenging setting of this problem is the case in which the reconstructed shapes are non-rigid. In this paper, we propose a framework for intrinsic regularization of such problems. The assumption is that we have the geometric structure of a shape which is intrinsically (up to bending) similar to the one we would like to reconstruct. For that goal, we formulate a variation with respect to vertex coordinates of a triangulated mesh approximating the continuous shape. The numerical core of the proposed method is based on differentiating the fast marching update step for geodesic distance computation.*

1. Introduction

In many tasks, both in human and computer vision, one tries to deduce the shape of an object given an observa-

many other problems, in which an object is reconstructed based on some measurement, are known as *shape reconstruction problems*. They are a subset of what is called *inverse problems*. Most such inverse problems are underdetermined, in the sense that measuring different objects may yield similar measurements. Thus, in the above illustration, the essence of the shadow theater is that it is hard to distinguish between shadows cast by an animal and shadows cast by hands. Therefore prior knowledge about the unknown object is needed.

Of particular interest are reconstruction problems involving non-rigid shapes. The world surrounding us is full with objects such as live bodies, paper products, plants, clothes etc., which may be deformed to different postures. These objects may be deformed to an infinite number of different postures. While bending, though, objects tends to preserve their internal geometric structure. Two objects differing by a bending are said to be *intrinsically similar*. In many cases, while we do not know the measured object, we have a prior on its intrinsic geometry. For example, in the shadow theater, though we do not know which exact posture of the hand

Take-Away

Huge zoo

of embedding techniques.

Each with different theoretical properties: Try them all!

But what if the distance matrix is incomplete or noisy?

Euclidean Matrix Completion

$$\min_{G} \|H \circ (\mathcal{D}(G) - D_{\text{input}})\|_{\text{Fro}}^2$$
$$\text{s.t. } G \succeq 0$$

Convex program

Related method: "Maximum variance unfolding"

Alfakih, Khandani, and Wolkowicz. "Solving Euclidean distance matrix completion problems via semidefinite programming." Comput. Optim. Appl., 12 (1999).

More General: Metric Nearness

$$\min_{X \in \mathcal{M}_{N\times N}} \|X - D\|_{\text{Fro}}^2$$

TRIANGLE_FIXING(D, ϵ)
Input: Input dissimilarity matrix D, tolerance ϵ
Output: $M = \operatorname{argmin}_{X \in \mathscr{M}_N} \|X - D\|_2$.
for $1 \le i < j < k \le n$
  $(z_{ijk}, z_{jki}, z_{kij}) \leftarrow 0$
for $1 \le i < j \le n$
  $e_{ij} \leftarrow 0$
$\delta \leftarrow 1 + \epsilon$
while $(\delta > \epsilon)$ {convergence test}
  foreach triangle (i, j, k)
    $b \leftarrow d_{ki} + d_{jk} - d_{ij}$
    $\mu \leftarrow \frac{1}{3}(e_{ij} - e_{jk} - e_{ki} - b)$
    $\theta \leftarrow \min\{-\mu, z_{ijk}\}$ {Stay within half-space of constraint}
    $e_{ij} \leftarrow e_{ij} - \theta$, $e_{jk} \leftarrow e_{jk} + \theta$, $e_{ki} \leftarrow e_{ki} + \theta$
    $z_{ijk} \leftarrow z_{ijk} - \theta$ {Update correction term}
  end *foreach*
  $\delta \leftarrow$ sum of changes in the e
end *while*
return $M = D + E$

In other words, the vector e is projected orthogonally onto the constraint set $\{e' : e'_{ij} - e'_{jk} - e'_{ki} \le b_{ijk}\}$. This is tantamount to solving

$$\min_{e'} \ \tfrac{1}{2}\left[(e'_{ij} - e_{ij})^2 + (e'_{jk} - e_{jk})^2 + (e'_{ki} - e_{ki})^2)\right], \quad \text{subject to} \quad e'_{ij} - e'_{jk} - e'_{ki} = b_{ijk}. \tag{3.2}$$

It is easy to check that the solution is given by

$$e'_{ij} \leftarrow e_{ij} - \mu_{ijk}, \qquad e'_{jk} \leftarrow e_{jk} + \mu_{ijk}, \qquad \text{and} \qquad e'_{ki} \leftarrow e_{ki} + \mu_{ijk}, \tag{3.3}$$

where $\mu_{ijk} = \frac{1}{3}(e_{ij} - e_{jk} - e_{ki} - b_{ijk}) > 0$.

Iterative projection

Dhillon, Sra, Tropp. "Triangle Fixing Algorithms for the Metric Nearness Problem." NIPS.

Challenging Computational Problems

- Is my data **embeddable**?
- Can you compute intrinsic **dimensionality**?
- Are two metric spaces **isometric**?
- How **similar** are two metric spaces?
- What is the **average** of two metric spaces?
- Can I embed into **non-Euclidean** spaces?

NP-Hardness Result

Robust Euclidean Embedding

Lawrence Cayton LCAYTON@CS.UCSD.EDU
Sanjoy Dasgupta DASGUPTA@CS.UCSD.EDU
Department of Computer Science and Engineering, University of California, San Diego
9500 Gilman Dr. La Jolla, CA 92093

Abstract

We derive a robust Euclidean embedding procedure based on semidefinite programming that may be used in place of the popular classical multidimensional scaling (cMDS) algorithm. We motivate this algorithm by arguing that cMDS is not particularly robust and has several other deficiencies. General-purpose semidefinite programming solvers are too memory intensive for medium to large sized applications, so we also describe a fast subgradient-based implementation of the robust algorithm. Additionally, since cMDS is often used for dimensionality reduction, we provide an in-depth look at reducing dimensionality with embedding procedures. In particular, we show that it is NP-hard to find optimal low-dimensional embeddings under a variety of cost functions.

choice for embedding seems to be
sional scaling (cMDS). Its popul
ing relatively fast, parameter-fre
and optimal for its cost functio
look carefully at the algorithm and
has some problematic features as w
we argue that the cost function is
conceptually awkward.

We propose a robust alternative to
clidean embedding (REE), that ret
desirable features of cMDS, but a
pitfalls. We show that the global
REE cost function can be found
nite program (SDP). Though this is
dard SDP-solvers can only manage t
gram for around 100 points. So
used on more reasonably sized dat
a subgradient-based implementation

Dimensionality reduction is an imp

ℓ_1 EUCLIDEAN EMBEDDING
Input: A dissimilarity matrix $D = (d_{ij})$.
Output: An embedding into the line: $x_1, x_2, \ldots \in \mathbf{R}$
Goal: Minimize $\sum_{i,j} |d_{ij} - |x_i - x_j||$.

We show that this problem is NP-hard by reducing from a variant of not-all-equal 3SAT.

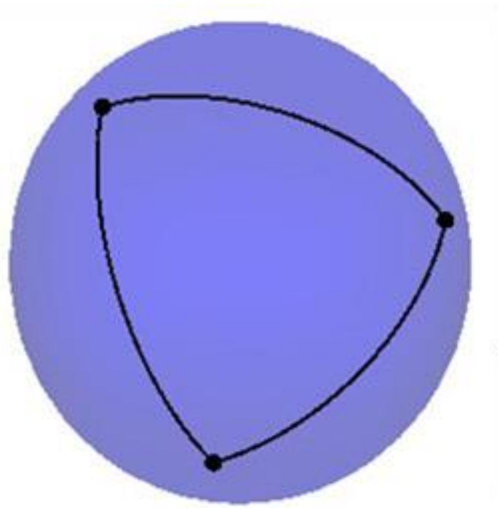
The hardness result can be extended to distortion functions of the form $\sum_{i,j} g\Big(f(d_{ij}) - f(|x_i - x_j|)\Big)$ We assume that f, g are

1. symmetric;
2. monotonically increasing in the absolute values of their arguments;
3. Lipschitz on $[0, 1]$ with constant λ_U, that is, for $x, y \in [0, 1]$, $|f(x) - f(y)| \leq \lambda_U |x - y|$; and
4. similarly lower-bounded: for some $\lambda_L > 0$, for any $x, y \in [0, 1]$, $|f(x) - f(y)| \geq \lambda_L |x - y| \max\{x, y\}$.

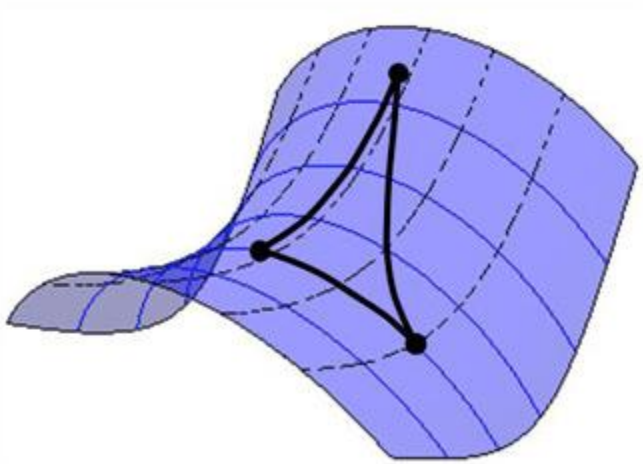
Notice that $f(x), g(x) \in \{x, x^2\}$ satisfy these conditions with $\lambda_U = 2, \lambda_L = 1$, meaning that $\|D - D^*\|_1$ and $\|D - D^*\|_2$ are both hard to minimize over one-dimensional embeddings.

Preview:

Dependence on Curvature



positively curved space
sphere



negatively curved space
saddle

https://www.learner.org/courses/physics/unit/text.html?unit=3&secNum=6

Q:

What are some applications of this machinery?

Applications

- Reduce algorithmic runtime
- Compression
- Visualize data
- Interpolate
- Sample

…

Visualization Examples

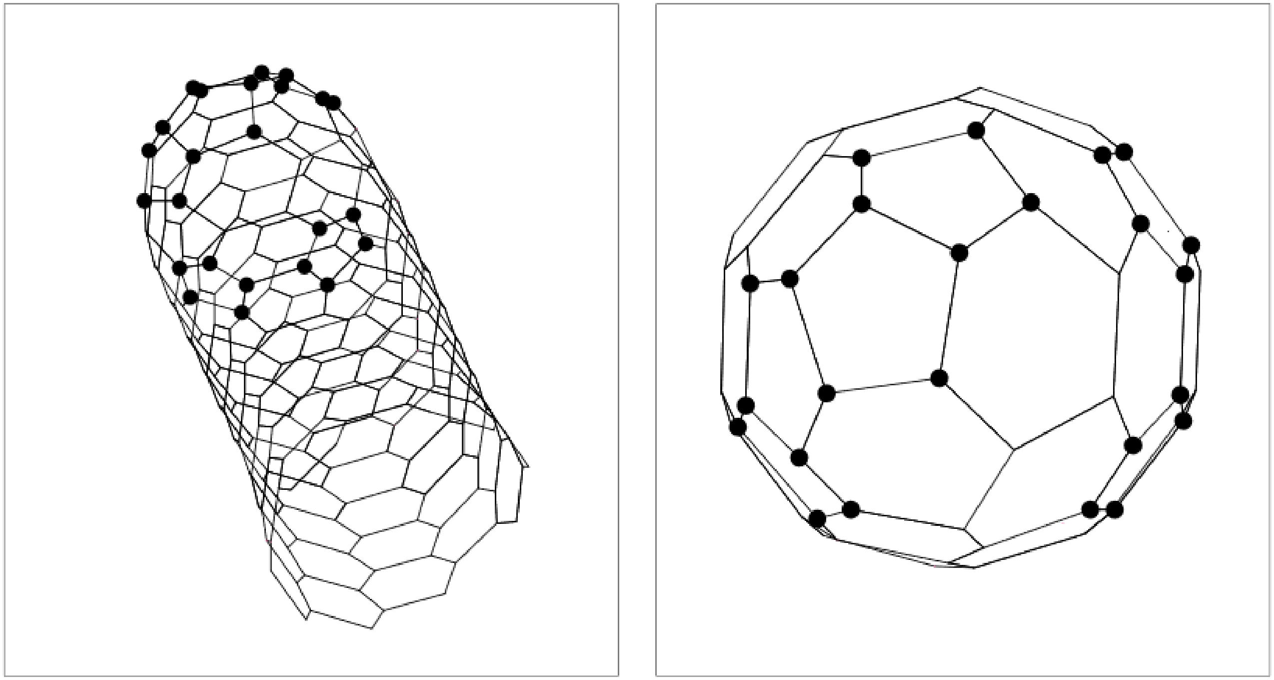


Figure 10: *Nanotube Embedding. One of Asimov's graphs for a nanotube is rendered with MDS in 3-D (Stress=0.06). The nodes represent carbon atoms, the lines represent chemical bonds. The right hand frame shows the cap of the tube only. The highlighted points show some of the pentagons that are necessary for forming the cap.*

http://www.stat.yale.edu/~lc436/papers/JCGS-mds.pdf

Visualization Examples

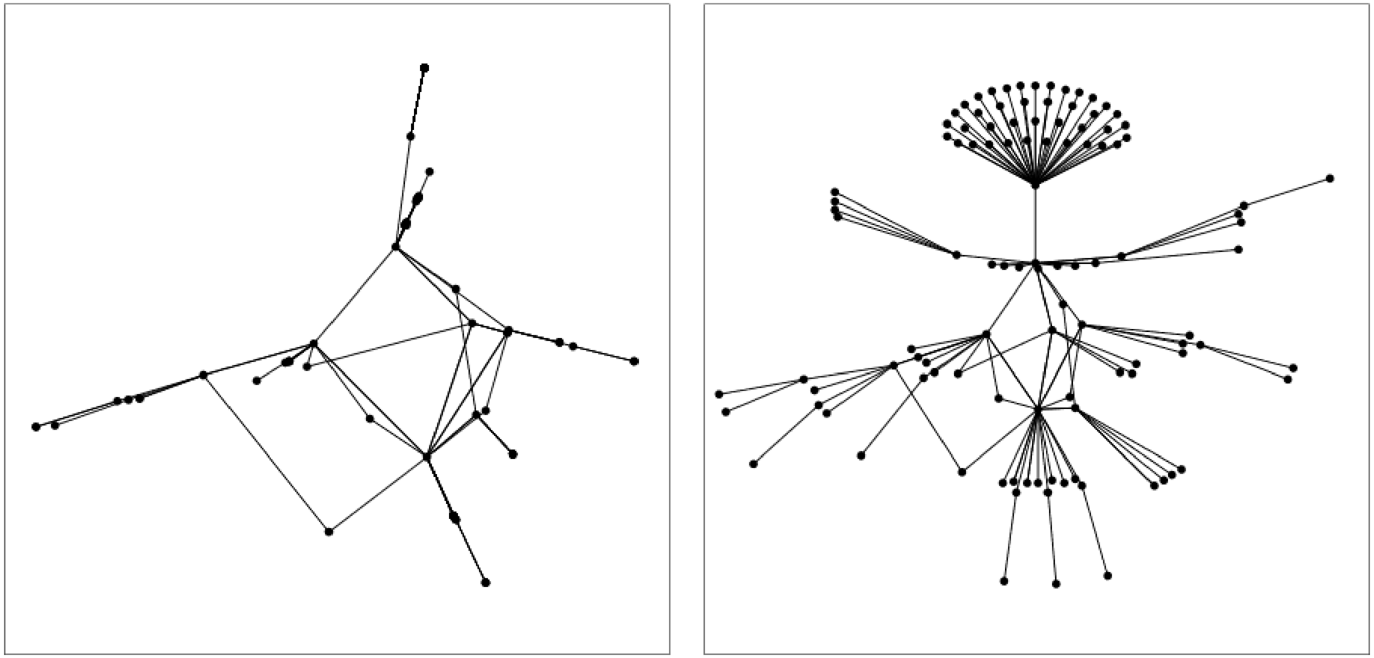


Figure 9: *A Telephone Call Graph, Layed Out in 2-D. Left: classical scaling (Stress=0.34); right: distance scaling (Stress=0.23). The nodes represent telephone numbers, the edges represent the existence of a call between two telephone numbers in a given time period.*

http://www.stat.yale.edu/~lc436/papers/JCGS-mds.pdf

Visualization Examples

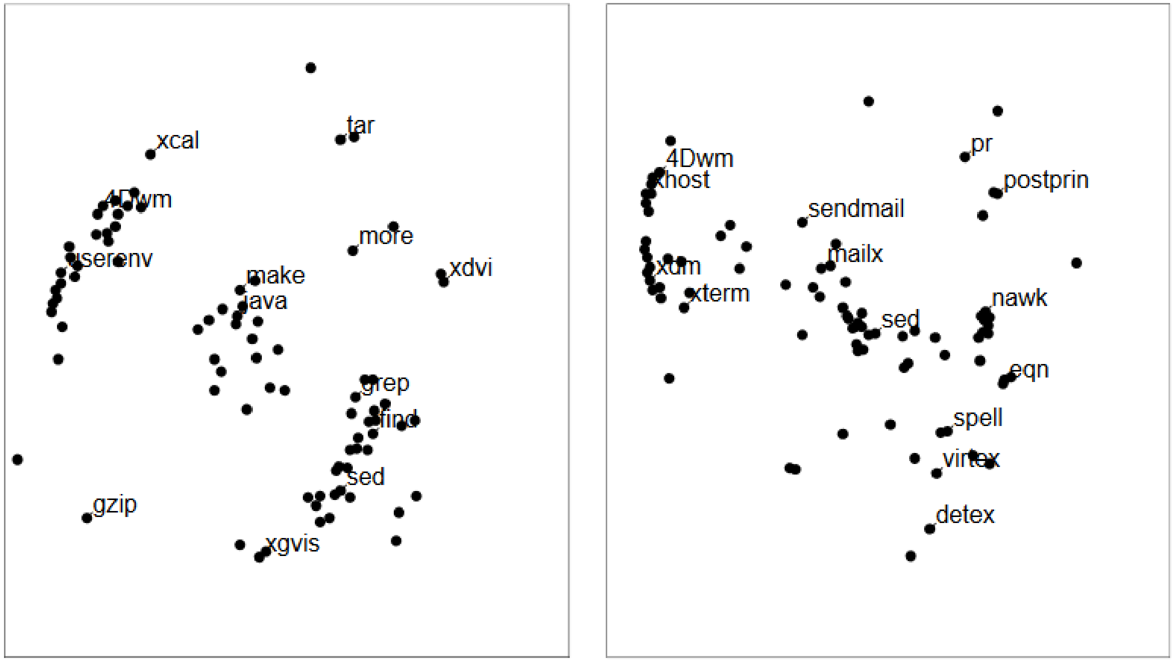


Figure 7: *Maps of Computer Commands for Two Individuals.*
Left: a member of technical staff who programs and manipulates data (Stress=0.29).
Right: an administrative assistant who does e-mail and word processing (Stress=0.34).
The labels show selected operating system commands used by these individuals.

http://www.stat.yale.edu/~lc436/papers/JCGS-mds.pdf

http://thepioneerwoman.com/food-and-friends/chocolate-swiss-roll-cake/

Inverse Distance Problems

Justin Solomon

MIT, Spring 2017