Continuum limits of shortest paths

Alfred Hero

Michigan Institute for Data Science (MIDAS)
Dept of Electrical Engineering and Computer Science (EECS)
Dept of Biomedical Engineering (BME)
Dept of Statistics
University of Michigan - Ann Arbor

May 1, 2017
Motivation

Continuum limits

Application to anomaly detection

Summary


Outline

1 Motivation

2 Continuum limits

3 Application to anomaly detection

4 Summary
Minimal graphs over empirical data

Long history in statistics, computer science, social science and finance

- **MST and non-parametric two sample tests**: Friedman and Rafsky (1979).
- **Shortest paths and generalized median**: Small (1997).
- **Shortest paths and social networks**: Sabidussi (1966).
- **Shortest paths and manifold learning**: Tenenbaum, de Silva and Langford (2000).
- **Efficient frontiers (anti-chain graphs) and finance**: Markowitz (1952).

---


Sabidussi, G., "The centrality index of a graph," Psychometrika (1966), 31, 581-603


Minimal graphs over empirical data

Figure: Shortest path through 2D swiss roll

Minimal graphs over empirical data

Figure: Efficient frontier portfolio design

Michaud, Robert O., Efficient asset management, Oxford, 2009
Continuum limits

Q. How to these minimal graphs behave in large sample (node) limit?

A. Asymptotic results have been available in some cases

- **MST 2-sample test**: Henze and Penrose (1999).

These results have provided mathematical justification, intuition and, in some cases, fast computational approximations for minimal graph optimization.


Shortest path (SP)

- Let $G$ be a graph with $m = |E|$ edges on $n$ vertices $V$
- $\pi(X_I, X_F)$ a path over $G$ between points $X_I$ and $X_F$
  \[ \pi(X_I, X_F) = (X_I, X_{i_1}, \ldots, X_{i_l}, X_F) \]
  $X_{i_{j+1}}$ is a neighbor on $G$ of predecessor $X_{i_j}$ and $X_I = X_{i_0}$, $X_F = X_{i_{l+1}}$
- The shortest path is the solution to
  \[ L_{\gamma}^{SP}(V) = \min_{\pi(X_I, X_F)} \sum_{X_i \in \pi(X_I, X_F)} |X_{i_{j+1}} - X_{i_j}|^\gamma \]
- Typical choices of $G$:
  - Complete graph
  - kNN graph
  - MST
- Applications: clustering, manifold learning, image retrieval, efficient network routing, graph classification
- Computational complexity is $O(m + n\log n)$
Shortest paths in manifold learning: ISOMAP geodesic approximation

Fig. 3. The “Swiss roll” data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (A) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect their intrinsic similarity, as measured by geodesic distance along the low-dimensional manifold (length of solid curve). (B) The neighborhood graph $G$ constructed in step one of Isomap (with $K = 7$ and $N =$ 1000 data points) allows an approximation (red segments) to the true geodesic path to be computed efficiently in step two, as the shortest path in $G$. (C) The two-dimensional embedding recovered by Isomap in step three, which best preserves the shortest path distances in the neighborhood graph (overlaid). Straight lines in the embedding (blue) now represent simpler and cleaner approximations to the true geodesic paths than do the corresponding graph paths (red).

Shortest paths with respect to non-Euclidean dissimilarity measures

\[ D(f_i \| f_j) = \int f_i(x) \log \frac{f_i(x)}{f_j(x)} dx + \int f_j(x) \log \frac{f_j(x)}{f_i(x)} dx \quad \text{(KL divergence)} \]

Shortest paths in epidemiology: virus strain genotyping

Define partial order relation "\( \leq \)" between any \( X, Y \in \mathbb{R}^d \):

\[
X \leq Y \iff X_i \leq Y_i, \quad \forall i = 1, \ldots, d
\]

\( X \) a minimal element of \( \mathcal{X} = \{X_1, \ldots, X_n\} \) if

1) \( X \in \mathcal{X} \)
2) \( \{X_i \in \mathcal{X} : X_i \leq X\} = \emptyset \)

Define \( \text{min} \mathcal{X} \) the set (Pareto front) of all minimal elements of \( \mathcal{X} \).

Pareto front of depth \( k \), an anti-chain denoted by \( \{\mathcal{F}_k\} \), is defined recursively

\[
\mathcal{F}_1 = \text{min} \mathcal{X}
\]

\[
\mathcal{F}_k = \text{min} \left\{ \mathcal{X} / \bigcup_{i=1}^{k-1} \mathcal{F}_i \right\}, \quad k = 1, 2, \ldots
\]

Applications: evolutionary computing, database indexing and retrieval, portfolio design, anomaly detection

Computational complexity is \( O(dn^2) \)
Objective: search a database for images combining concepts of “sea” and “mountain”

Standard image matching is limited

- Produces single rank ordered list of closest matches
- Desired match may be deeply buried in combined lists

Issue: people rarely examine more than a few of the top matches
Objective: search a database for images combining concepts of “sea” and “mountain”

Standard image matching is limited
- Produces single rank ordered list of closest matches
- Desired match may be deeply buried in combined lists

Issue: people rarely examine more than a few of the top matches
Illustration: multiple concept image retrieval in SS dataset

Pareto fronts give high ranks to points that are not highly ranked by linear scalarization.

Red fronts are the first 4 fronts covering around 100 points. Red and green fronts are the first 8 fronts covering around 200 points.

Stanford Scene dataset, SIFT feature, Spatial Pyramid Matching

Outline

1. Motivation
2. Continuum limits
3. Application to anomaly detection
4. Summary
SP through complete graph for uniform points in plane

Euclidean distance ($\gamma = 1$)

(Euclidean distance)$^2$ ($\gamma = 2$)
SP through complete graph for Gaussian points in plane: "Lensing effect"

Shortest path through 2000 nodes. $\gamma = 1$

Shortest path through 2000 nodes. $\gamma = 2$

Euclidean ($\gamma = 1$)  (Euclidean)² ($\gamma = 2$)
Continuum limit of shortest path

Let \( \mathcal{X} = \{X_1, \ldots, X_n\} \) be i.i.d. random vectors in \( \mathbb{R}^d \) with marginal pdf \( f \) with support set \( S \). Fix two points \( x_I \) and \( x_F \) in \( \mathbb{R}^d \).

Define \( G \) as the complete graph spanning \( \mathcal{X} \).

Theorem (Hwang, Damelin and H 2016)

Assume that \( \inf_x f(x) > 0 \) over a compact support set \( S \) with pd metric tensor \( g \). For \( \gamma > 1 \) the shortest path on \( G \) between any two points \( x_I, x_F \in S \) satisfies

\[
L^{SP}_{\gamma} (\mathcal{X}) / n^{(1-\gamma)/d} \to C_{d, \gamma} \inf_{\pi} \int_0^1 f(\pi_t)^{(1-\gamma)/d} \sqrt{g(\dot{\pi}_t, \dot{\pi}_t)} dt \quad \text{(a.s.)}
\]

where the infimum is taken over all smooth curves \( \pi : [0, 1] \to \mathbb{R}^d \) with \( \pi_0 = x_I \) and \( \pi_1 = x_F \) and \( C(d, \gamma) \) is a constant independent of \( f \).

---

Continuum limit of shortest path: archimedean vs relativistic limit

$d = 2$, $\gamma = 2$

Archimedean shortest path

Relativistic shortest path
Continuum limit of shortest path: variational form

Define

\[ F(\pi, \dot{\pi}) = f(\pi)^{(1-\gamma)/d} \sqrt{g(\dot{\pi}, \dot{\pi})} \]

Then normalized shortest path length converges to \( C_{d, \gamma} \inf_{\pi} \int_{0}^{1} F(\pi_t, \dot{\pi}_t) dt \).

Using calculus of variations can show that the asymptotic shortest path \( \pi \) satisfies the system of \( d \) coupled Euler-Lagrange equations

\[
\frac{d}{dt} (\nabla_{\pi} F(\pi, \dot{\pi})) - \nabla_{\pi} F(\pi, \dot{\pi}) = 0, \quad t \in [0, 1]
\]

with boundary conditions \( \pi_0 = x_I, \pi_1 = x_F \). E.g., for \( g(\dot{\pi}, \dot{\pi}) = \langle \dot{\pi}, \dot{\pi} \rangle \)

\[
\frac{1 - \gamma}{d} A(\dot{\pi}) \nabla_{\pi} \ln f(\pi) + \frac{d}{dt} \left( \frac{\dot{\pi}}{\|\dot{\pi}\|} \right) = 0
\]
Continuum limit of shortest path: variational form

Define

\[ F(\pi, \dot{\pi}) = f(\pi)^{(1-\gamma)/d} \sqrt{g(\dot{\pi}, \dot{\pi})} \]

Then normalized shortest path length converges to \( C_{d, \gamma} \inf_{\pi} \int_{0}^{1} F(\pi_t, \dot{\pi}_t) dt \).

Using calculus of variations can show that the asymptotic shortest path \( \pi \) satisfies the system of \( d \) coupled Euler-Lagrange equations

\[
\frac{d}{dt} \left( \nabla_{\dot{\pi}} F(\pi, \dot{\pi}) \right) - \nabla_{\pi} F(\pi, \dot{\pi}) = 0, \quad t \in [0, 1]
\]

with boundary conditions \( \pi_0 = x_I, \pi_1 = x_F \). E.g., for \( g(\dot{\pi}, \dot{\pi}) = \langle \dot{\pi}, \dot{\pi} \rangle \)

\[
\frac{1 - \gamma}{d} A(\dot{\pi}) \nabla_{\pi} \ln f(\pi) + \frac{d}{dt} \left( \frac{\dot{\pi}}{\|\dot{\pi}\|} \right) = 0
\]

Special case of points in the plane \((d = 2)\): \( \pi_t = (t, y_t) \)

\[
\frac{1 - \gamma}{d} \left( \alpha_1(\dot{y}) f_{10}(t, y) + \alpha_2(\dot{y}) f_{01}(t, y) \right) / f(t, y) + \frac{d}{dt} \left( \frac{\dot{y}}{\sqrt{1 + \dot{y}^2}} \right) = 0
\]

\( \alpha_1(\dot{y}) = \dot{y}/\sqrt{1 + \dot{y}^2}, \alpha_2(\dot{y}) = -1/\sqrt{1 + \dot{y}^2} \)
Experimental validation of shortest path continuum limit

Regression equation \( (\alpha = (1 - \gamma)/d) \):

\[
\log L_\gamma(\mathcal{X}) = \alpha \log n + \log \text{dist}_\gamma(x, y) + \log C_{d, \gamma}
\]

Experimental setting

- \( d = 2, \gamma = 2 \) so that slope should be \( (1 - \gamma)/d = -0.5 \)
- \( \mathcal{X}_n \) are \( n \) uniform points on \( S = S^2 \)
- Blue plot: \( x = (1, 0, 0), y = (-1, 0, 0) \)
- Red plot: \( x = (0, 1, 0), y = (0, 0, 1) \)
Continuum limit for non-dominated sorting: Demo for Unif\([0, 1]^2/[0, 0.5]^2\)

Motivation
Continuum limits
Application
Summary
References

Continuum limit for non-dominated sorting: Demo for Unif[0, 1]^2/[0, 0.5]^2

Continuum limit for non-dominated sorting: Demo for Unif[0, 1]^2/[0, 0.5]^2

Continuum limit for non-dominated sorting: Demo for $\text{Unif}[0, 1]^2/[0, 0.5]^2$

Asymptotic theorem for non-dominated sorting

Define $u_n(x)$ the function that counts the number of Pareto fronts in wedge \{\(X_i \leq x\}\}. Assume that \(\text{supp}(f) \subset \Omega \subset \mathbb{R}^d\), \(\Omega\) bounded with Lipshitz \(\partial \Omega\).

**Theorem (Calder, Esedoglu and H, 2014)**

There exists a \(c_d > 0\) such that w.p.1

\[
n^{-1/d} u_n \to c_d U, \text{ in } L^\infty(\mathbb{R}^d_+)\]

where

1. **\(U\) is the Pareto monotone \(^a\) solution of the variational problem**

\[
U(x) = \sup_{\gamma \in A} \int_0^1 f^{1/d}(\gamma(t))(\gamma'(t) \cdots \gamma'_d(t))^{1/d} dt
\]

where \(A = \left\{ \gamma \in C^1(0, 1; \mathbb{R}^d) : \gamma'(t) \geq 0 \ \forall t \in [0, 1] \right\}\)

2. **\(U\) is the unique viscosity solution to the Hamilton-Jacobi p.d.e**

\[
\frac{\partial U}{\partial x_1} \cdots \frac{\partial U}{\partial x_d} = \frac{1}{d^d} f \text{ in } \Omega
\]

\[
U = 0 \text{ on } \partial \Omega
\]

\(^a U(x) \leq U(y) \text{ if } x \leq y\)
Demonstration: theory vs experiment for Unif\([0, 1]/[0, 0.5]^2\)

Outline

1 Motivation

2 Continuum limits

3 Application to anomaly detection

4 Summary
Motivation: Detect anomalous pedestrian trajectories.
Question: Which one of these groups of trajectories are anomalous?

Anomalous trajectories  Nominal trajectories

Curve features: curve length, shape, walking speed.

Run-time comparisons

Curve similarity sorting for 50,000 curves ($10^9$ pairs)

- Curve pairs are represented by two similarity measures
  - Similarity measure 1: similarity of speed histograms along curves
  - Similarity measure 2: similarity between shapes of curves
Run-time comparisons

Curve similarity sorting for 50,000 curves ($10^9$ pairs)

- Curve pairs are represented by two similarity measures
  - Similarity measure 1: similarity of speed histograms along curves
  - Similarity measure 2: similarity between shapes of curves

- Grid size used by pde solver: $250 \times 250$
Run-time comparisons

Curve similarity sorting for 50,000 curves ($10^9$ pairs)

- Curve pairs are represented by two similarity measures
  - Similarity measure 1: similarity of speed histograms along curves
  - Similarity measure 2: similarity between shapes of curves

- Grid size used by PDE solver: $250 \times 250$

- Non-dominated Pareto sorting by PDE increases in #samples as $O(1)$
Outline

1. Motivation
2. Continuum limits
3. Application to anomaly detection
4. Summary
Summary

- Continuum limit analysis can lead to useful tools and insights for combinatorial problems in data science
  - Analysis connects the geometry of data to the structure of the underlying data distribution
  - Can lead to scalable pde-based algorithms for solving minimal path and non-dominated sorting problems
Summary

- Continuum limit analysis can lead to useful tools and insights for combinatorial problems in data science
  - Analysis connects the geometry of data to the structure of the underlying data distribution
  - Can lead to scalable pde-based algorithms for solving minimal path and non-dominated sorting problems

- Some open problems
  - Quantification of bias-variance tradeoffs for continuum approximations
  - Minimal paths on sparse graphs, directed paths, multigraphs, hypergraphs
  - Non-dominated sorting extensions to data depth and convex hull peeling
• Continuum limit analysis can lead to useful tools and insights for combinatorial problems in data science
  • Analysis connects the geometry of data to the structure of the underlying data distribution
  • Can lead to scalable pde-based algorithms for solving minimal path and non-dominated sorting problems

• Some open problems
  • Quantification of bias-variance tradeoffs for continuum approximations
  • Minimal paths on sparse graphs, directed paths, multigraphs, hypergraphs
  • Non-dominated sorting extensions to data depth and convex hull peeling

• Broader questions
  • New frontier: combinatorial analysis by density estimates and pde's?
  • New primitive: state-of-the-art numerical pde solvers in pipeline?


