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

References for this talk

Hwang, Sung-Jin, Steven Damelin, Alfred O. Hero, "Shortest path through random points," *Annals of Applied Probability*, Volume 26, Number 5 (2016), 2791-2823.

Calder, Jeff, Selim Esedoglu, and Alfred O. Hero. "A Hamilton–Jacobi Equation for the Continuum Limit of Nondominated Sorting." *SIAM Journal on Mathematical Analysis*, 46, no. 1 (2014): 603-638.

Calder, Jeff, Selim Esedoglu, and Alfred O. Hero. "A PDE-based approach to nondominated sorting." *SIAM Journal on Numerical Analysis*, 53, no. 1 (2015): 82-104.

Outline

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).

Friedman, Jerome H., and Lawrence. C. Rafsky. "Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests." The Annals of Statistics (1979): 697-717.

Small, Christopher G. "Multidimensional medians arising from geodesics on graphs." The Annals of Statistics (1997): 478-494.

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

Tenenbaum, Josh B., V. de Silva, and J. C. Langford. "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science (2000), 290, 2319-2323.

Markowitz, Harry. "Portfolio selection," The journal of finance, no. 1 (1952): 77-91.

Minimal graphs over empirical data

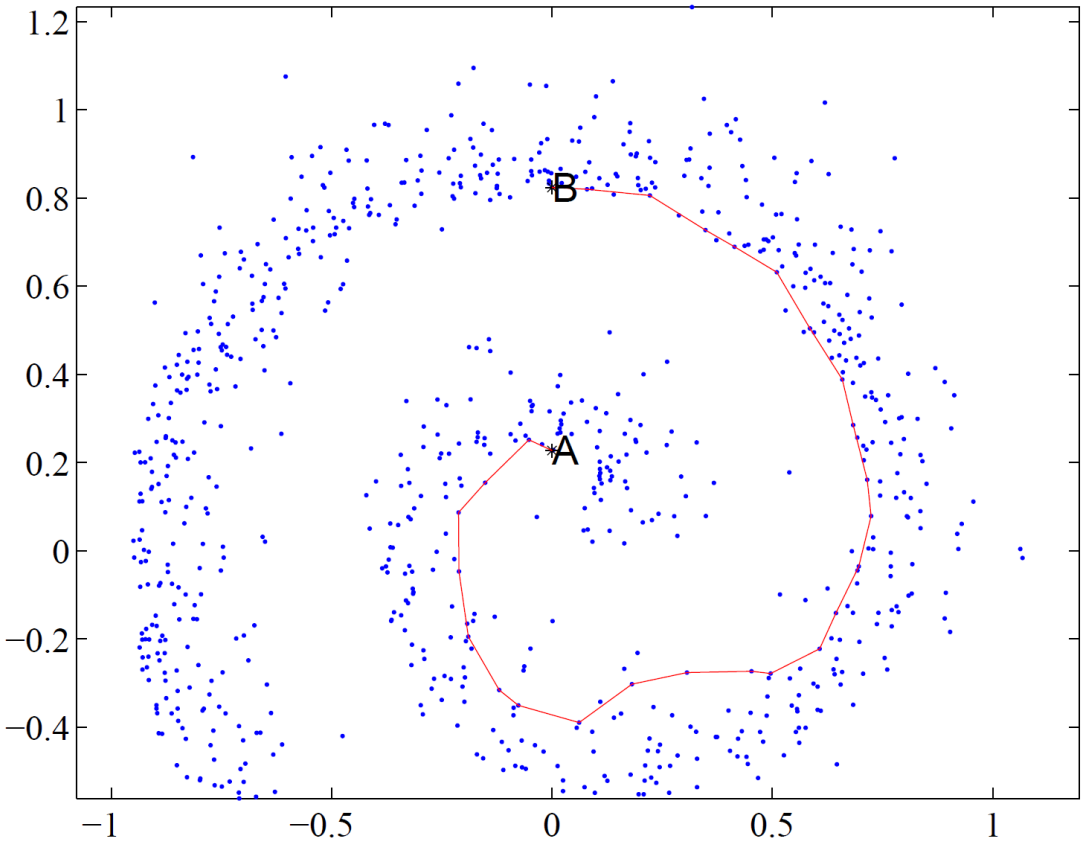


Figure: Shortest path through 2D swiss roll

Lee, Ann B., and Larry Wasserman. "Spectral connectivity analysis." Journal of the American Statistical Association 105, no. 491 (2010): 1241-1255.

Minimal graphs over empirical data

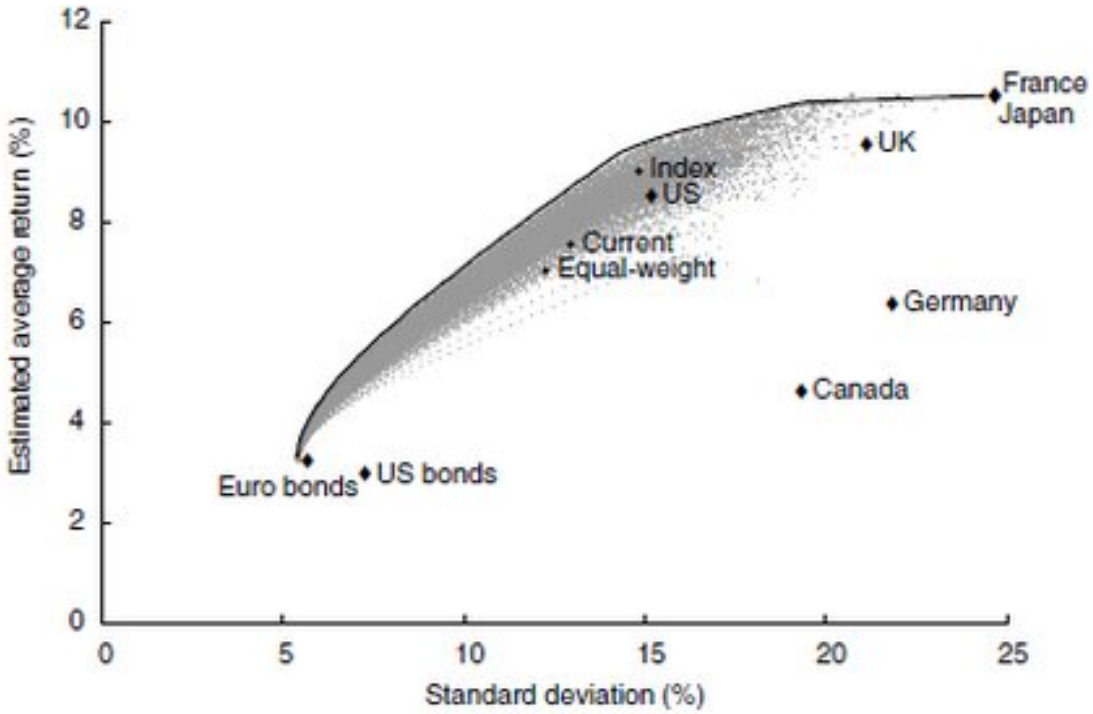


Figure: Efficient frontier portfolio design

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 Length functional**, Beardwood, Halton, Hammersley (1959), Steele (1997), Yukich (2006).
- **MST 2-sample test**: Henze and Penrose (1999).
- **Shortest paths for manifold learning**: Bernstein, et al (2000).

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

Beardwood, Jillian, John H. Halton, and John M. Hammersley. "The shortest path through many points." In Mathematical Proceedings of the Cambridge Philosophical Society, vol. 55, no. 04, pp. 299-327. Cambridge University Press, 1959.

Steele, J. Michael. Probability theory and combinatorial optimization. Society for Industrial and Applied Mathematics, 1997.

Yukich, Joseph E., Probability theory of classical Euclidean optimization problems, Springer, 2006

Henze, Norbert, and Mathew D. Penrose. "On the multivariate runs test." Annals of statistics (1999): 290-298.

Bernstein, M., V. de Silva, J. C. Langford, and J. B. Tenenbaum (2000). Graph approximations to geodesics on embedded manifolds. Technical report, Department of Mathematics, Stanford University.

Shortest path (SP)

- Let $\mathcal{G}$ be a graph with $m = |\mathcal{E}|$ edges on n vertices $\mathcal{V}$
- $\pi(X_I, X_F)$ a path over $\mathcal{G}$ btwn points X_I and X_F

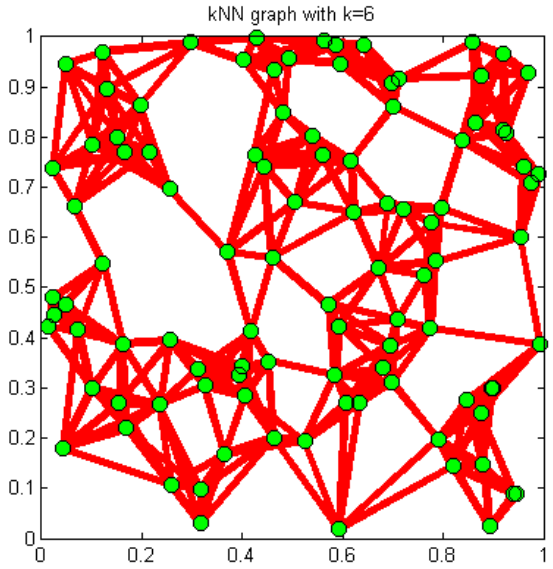
$$\pi(X_I, X_F) = (X_I, X_{i_1}, \ldots, X_{i_l}, X_F)$$

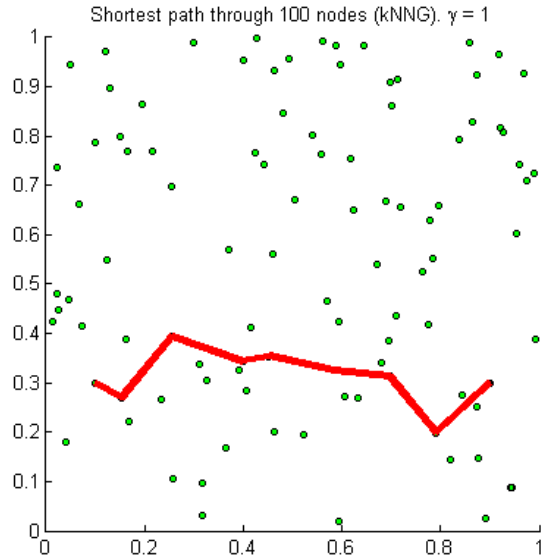
$X_{i_{j+1}}$ is neighbor on $\mathcal{G}$ of predecessor X_{i_j} and $X_I = X_{i_0}$, $X_F = X_{i_{l+1}}$

- The shortest path is the solution to

$$L_\gamma^{SP}(\mathcal{V}) = \min_{\pi(X_i, X_F)} \sum_{X_i \in \pi(X_I, X_F)} |X_{i_{j+1}} - X_{i_j}|^\gamma$$

- Typical choices of $\mathcal{G}$:
  - 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

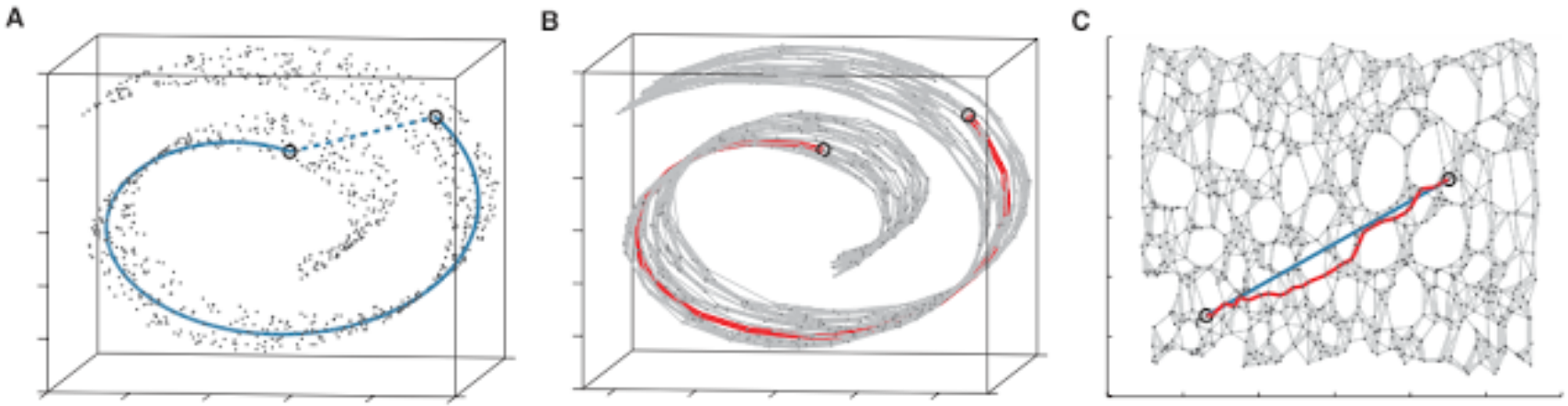
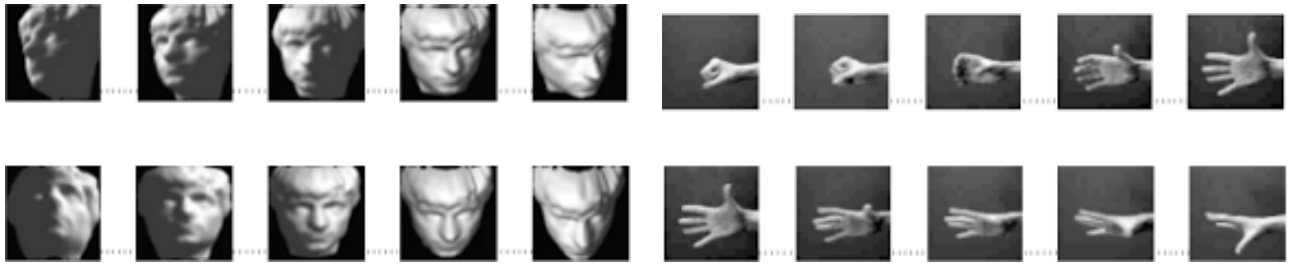


Fig. 3. The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (**A**) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect 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).



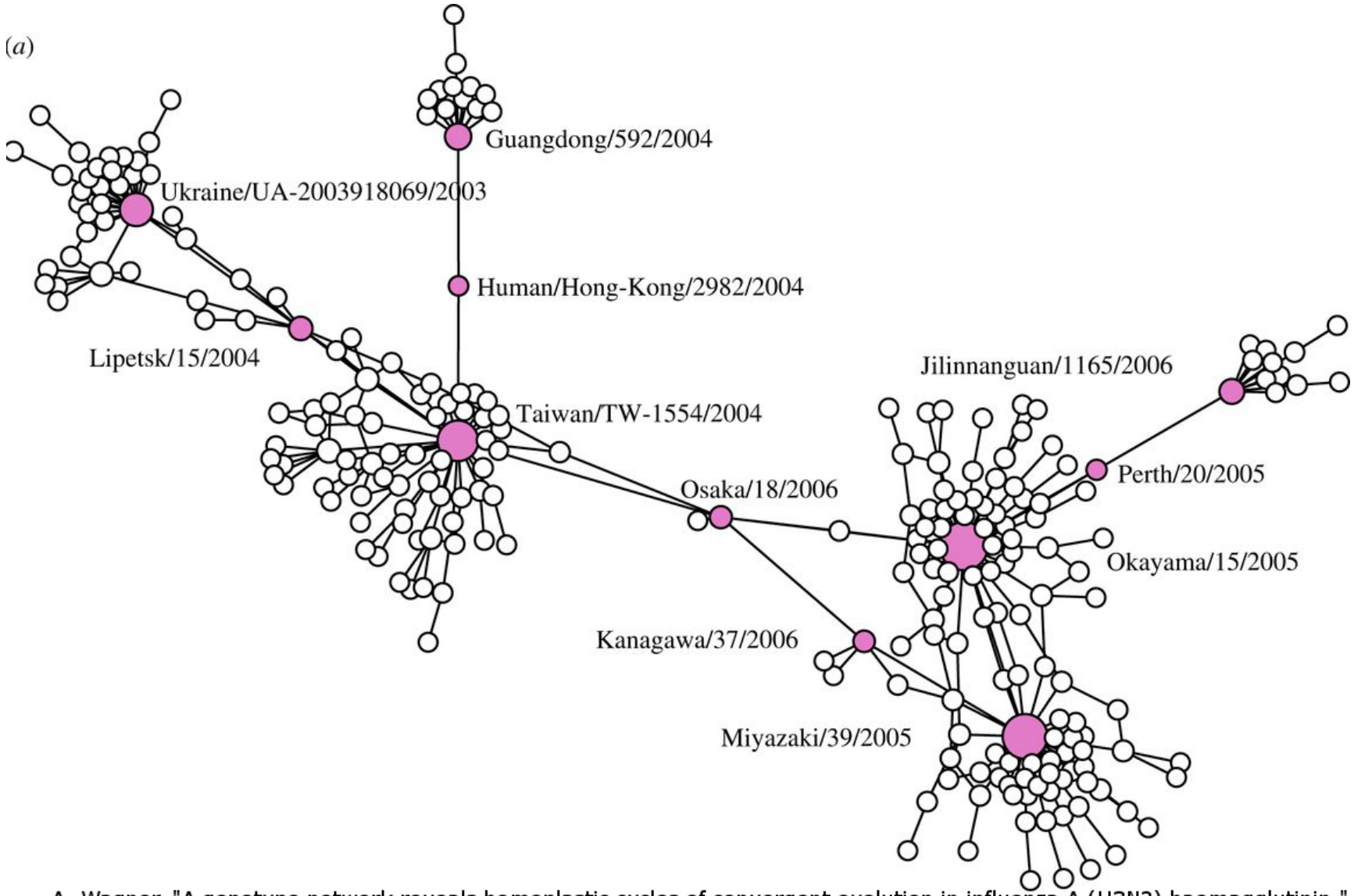
- Tenenbaum, Joshua B., Vin De Silva, and John C. Langford. "A global geometric framework for nonlinear dimensionality reduction." Science 290.5500 (2000): 2319-2323.

Shortest paths with respect to non-Euclidean dissimilarity measures

Averbuch-Elor, Cohen-Or and Kopf, "Smooth Image Sequences for Data Driven Morphing," Computer Graphics Forum, 35(6), 2016

$$D(f_i \| f_j) = \int f_i(x) \log \frac{f_i(x)}{f_j(x)} + \int f_j(x) \log \frac{f_j(x)}{f_i(x)} dx \quad (\textit{KL divergence})$$

Shortest paths in epidemiology: virus strain genotyping



A. Wagner, "A genotype network reveals homoplastic cycles of convergent evolution in influenza A (H3N2) haemagglutinin," Proc. Royal Soc. B, May 2014.

Non-dominated ranking in multiple dimensions

- Define partial order relation "$\leqq$" between any $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^d$:

$$\mathbf{X} \leqq \mathbf{Y} \Leftrightarrow X_i \leq Y_i, \quad \forall i = 1, \ldots, d$$

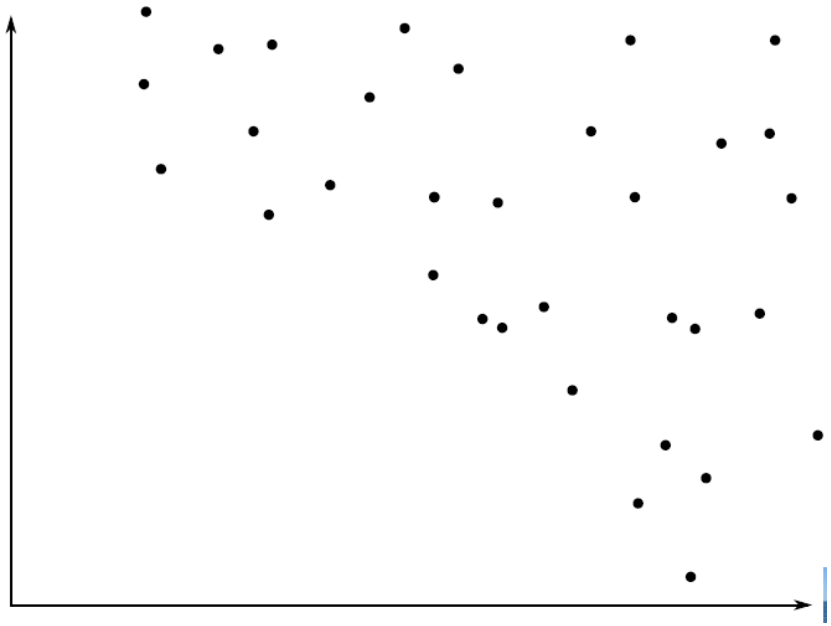
- $\mathbf{X}$ a minimal element of $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ if

$$\begin{aligned} &1)\ \mathbf{X} \in \mathcal{X} \\ &2)\ \{\mathbf{X}_i \in \mathcal{X} : \mathbf{X}_i \leqq \mathbf{X}\} = \emptyset \end{aligned}$$

- Define $\min \mathcal{X}$ 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

$$\begin{aligned} \mathcal{F}_1 &= \min \mathcal{X} \\ \mathcal{F}_k &= \min \left\{ \mathcal{X} / \cup_{i=1}^{k-1} \mathcal{F}_i \right\}, \quad k = 1, 2, \ldots \end{aligned}$$

- Applications: evolutionary computing, database indexing and retrieval, portfolio design, anomaly detection
- Computational complexity is $O(dn^2)$



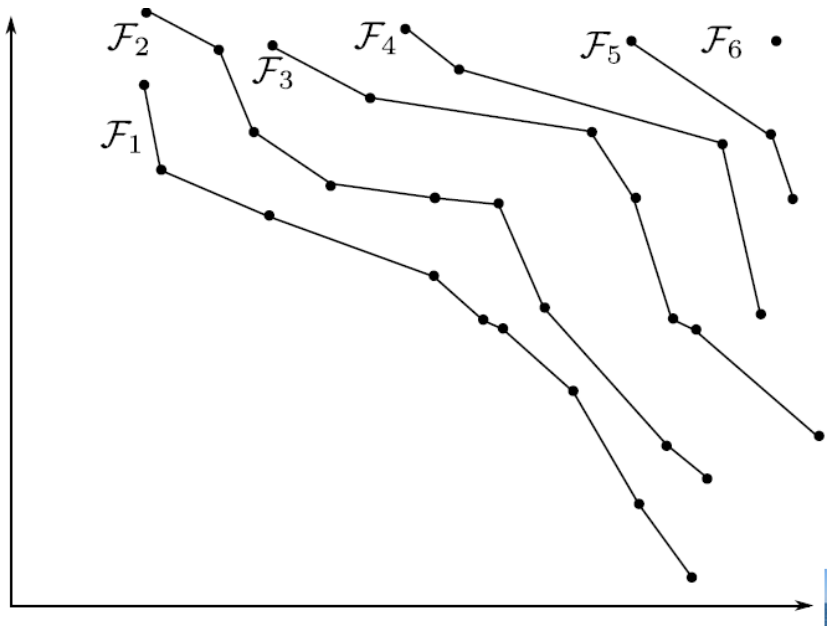


Illustration: Image retrieval combining multiple semantic concepts

Objective: search a database for images combining concepts of "sea" and "mountain"

Query 1

Query 2

Desired match

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: Image retrieval combining multiple semantic concepts

Objective: search a database for images combining concepts of "sea" and "mountain"

Query 1

Query 2

Desired match

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

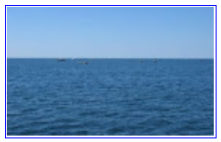


Image size:
344 × 214

Find other sizes of this image:
All sizes - Medium - Large

Visually similar images

Report images

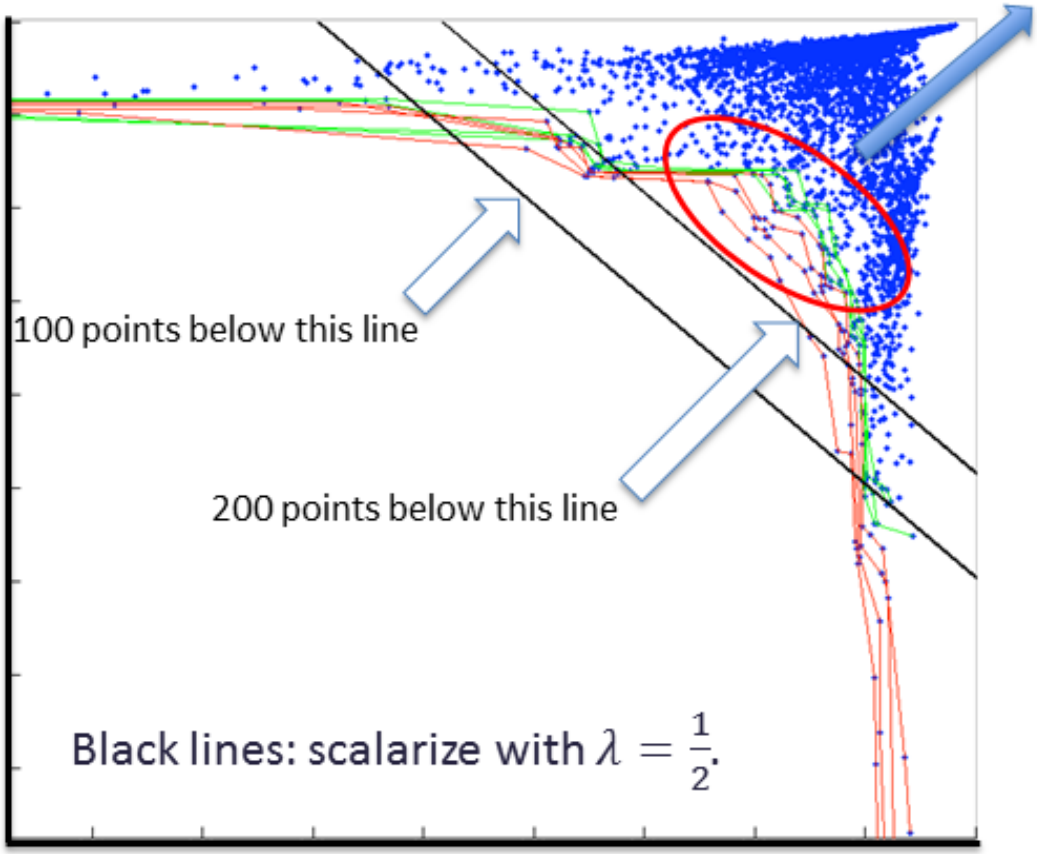
Image size:
344 × 257

Find other sizes of this image:
All sizes - Large

Visually similar images

Report images

Illustration: multiple concept image retrieval in SS dataset



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

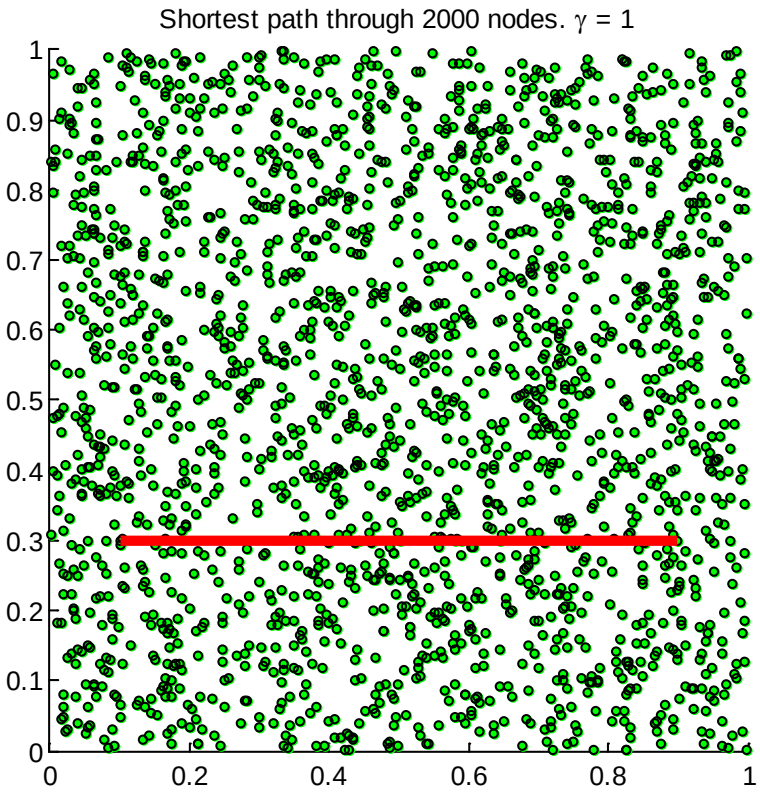
Red fronts are the first 4 fronts covering around 100 points.
Red and green fronts are the first 8 fronts covering around 200 points

Stanford Scene dataset, SIFT feature, Spatial Pyramid Matching

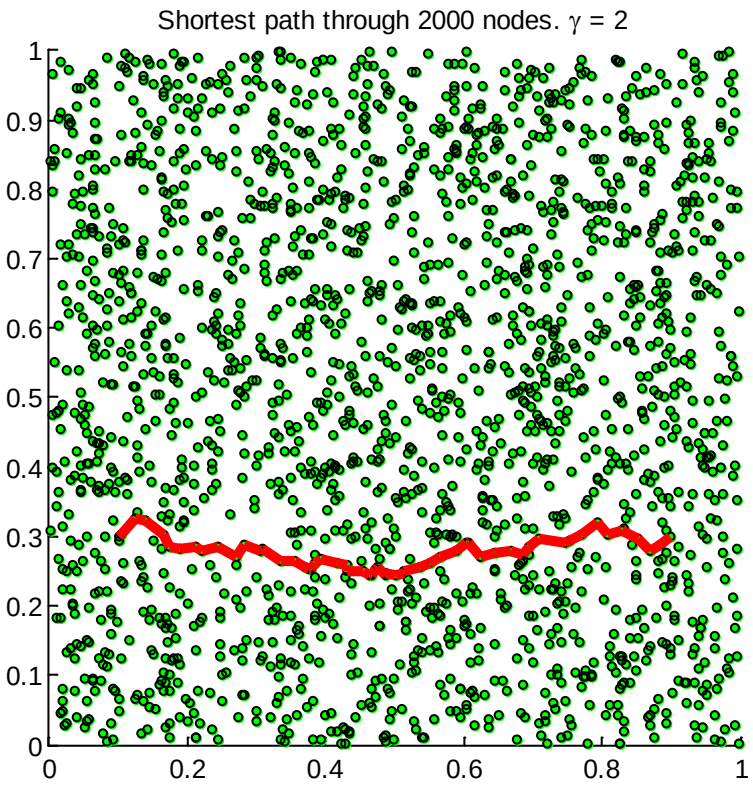
Hsiao, Calder and H, "Multiple-query Image Retrieval using Pareto Front Method," IEEE Trans. on Image Processing 2015.

Outline

SP through complete graph for uniform points in plane

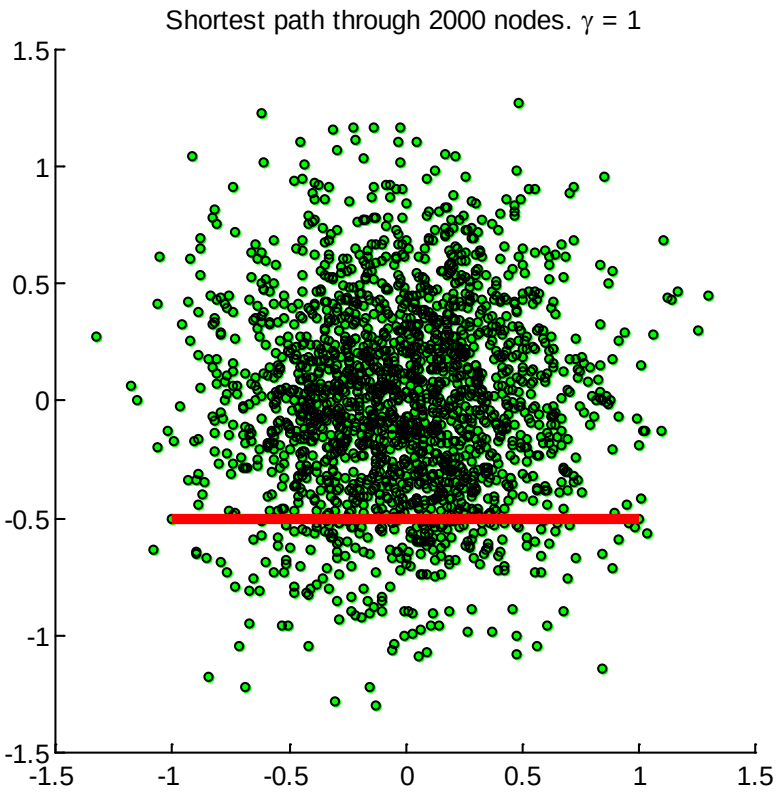


Euclidean distance ($\gamma = 1$)

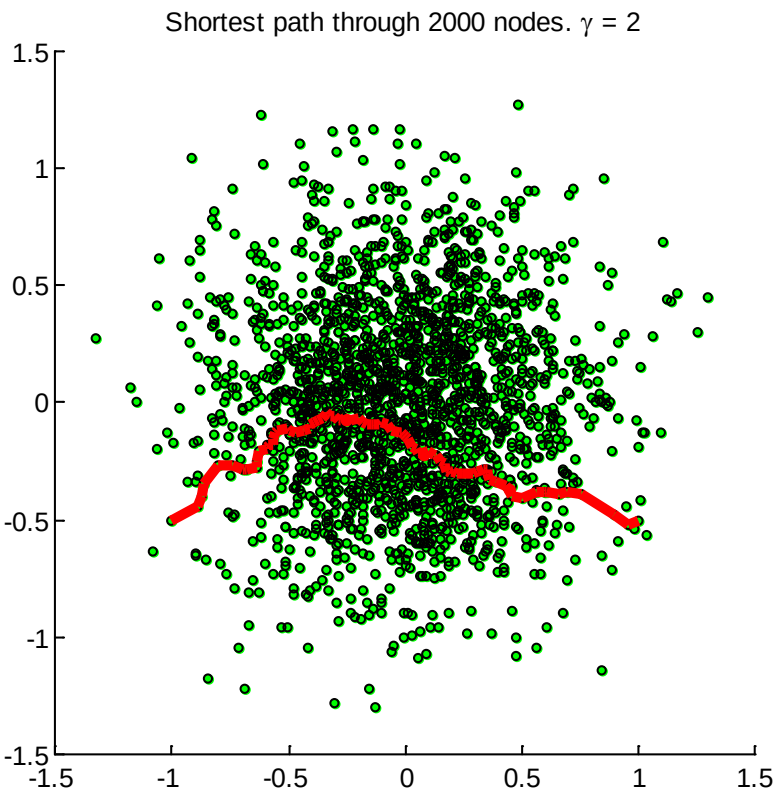


(Euclidean distance)^2 ($\gamma = 2$)

SP through complete graph for Gaussian points in plane: "Lensing effect"



Euclidean ($\gamma = 1$)



$(\text{Euclidean})^2$ ($\gamma = 2$)

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 $\mathcal{S}$. Fix two points x_I and x_F in $\mathbb{R}^d$.

Define $\mathcal{G}$ 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 $\mathcal{S}$ with pd metric tensor g. For $\gamma > 1$ the shortest path on $\mathcal{G}$ between any two points $x_I, x_F \in \mathcal{S}$ satisfies

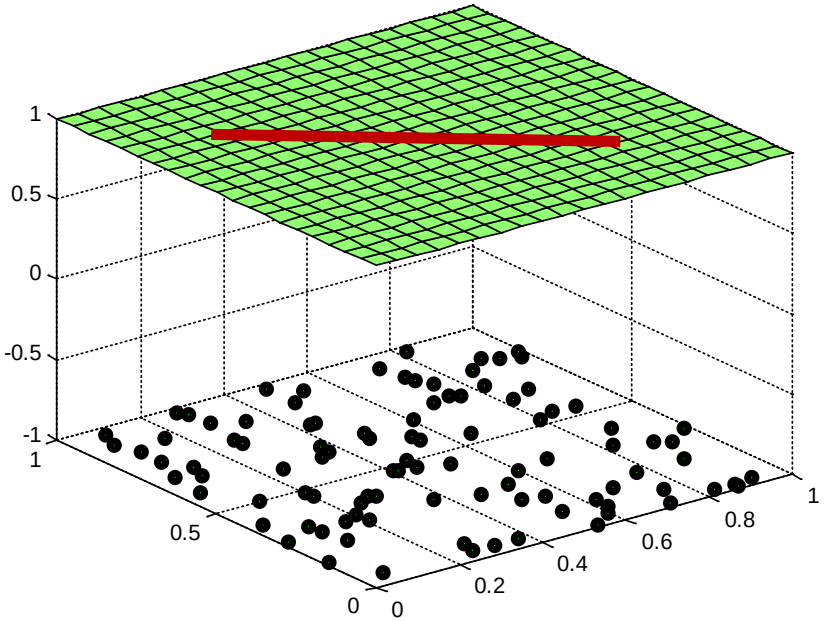
$$L_\gamma^{SP}(\mathcal{X})/n^{(1-\gamma)/d} \to C_{d,\gamma} \underbrace{\inf_\pi \int_0^1 f(\pi_t)^{(1-\gamma)/d} \sqrt{g(\dot{\pi}_t, \dot{\pi}_t)} dt}_{dist_\gamma(x_I, x_F)} \qquad (a.s.)$$

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.

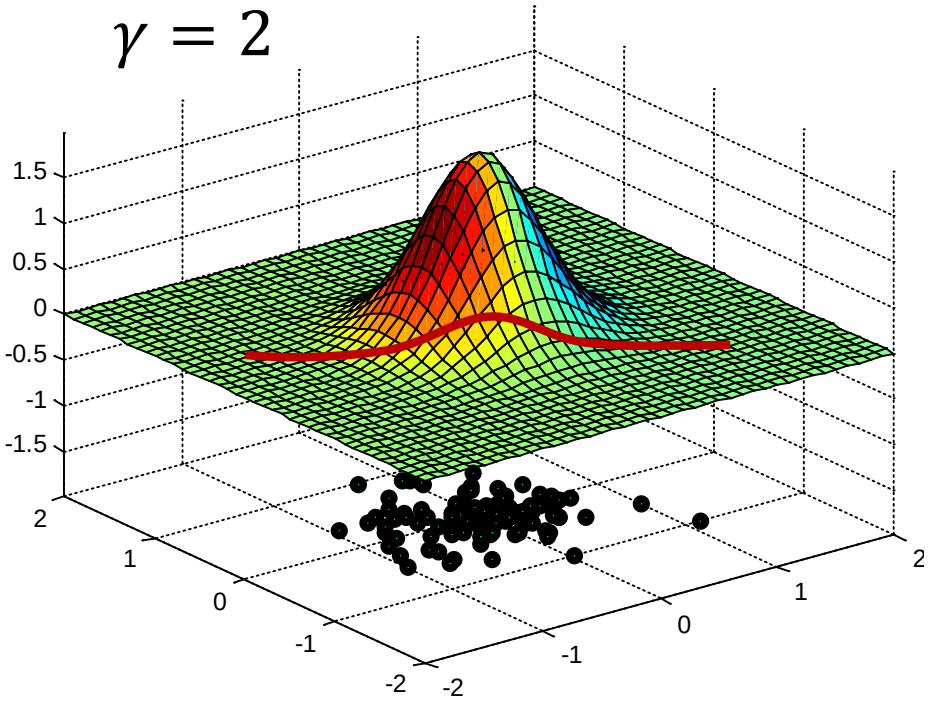
• S.-J. Hwang, S. Damelin, A. Hero, "Shortest path through random points," Annals of Applied Probability, 2016 (arXiv:1202.0045).

Continuum limit of shortest path: archimedean vs relativistic limit

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



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 π satisfies the system of d coupled Euler-Lagrange equations

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

with boundary conditions $\pi_0 = \mathbf{x}_I$, $\pi_1 = \mathbf{x}_F$. E.g., for $g(\dot{\pi}, \dot{\pi}) = \langle \dot{\pi}, \dot{\pi} \rangle$

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

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 π satisfies the system of d coupled Euler-Lagrange equations

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

with boundary conditions $\pi_0 = \mathbf{x}_I$, $\pi_1 = \mathbf{x}_F$. E.g., for $g(\dot{\pi}, \dot{\pi}) = \langle \dot{\pi}, \dot{\pi} \rangle$

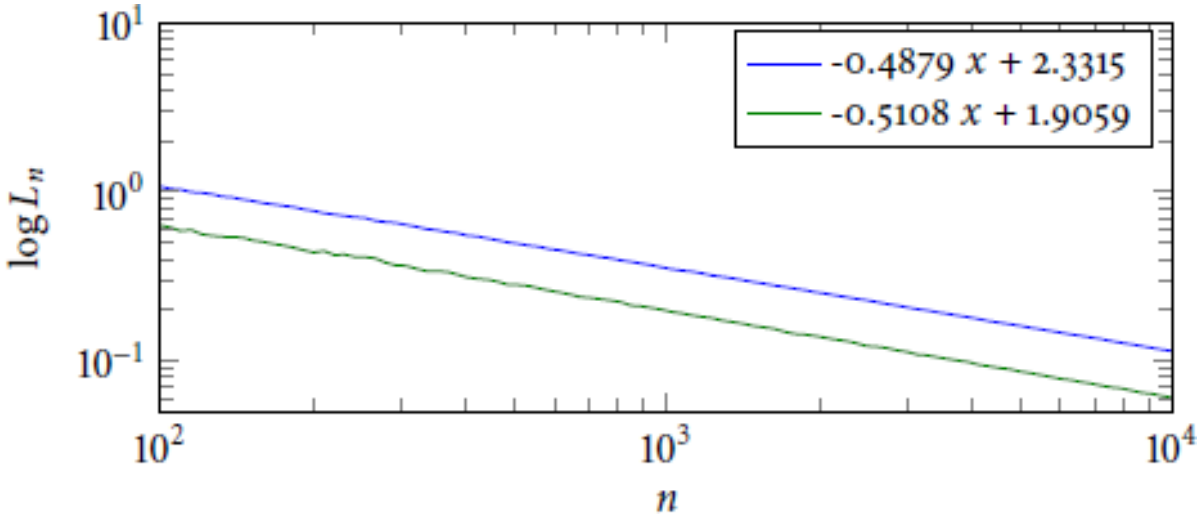
$$\frac{1-\gamma}{d} \mathbf{A}(\dot{\pi}) \nabla_\pi \ln f(\pi) + \frac{d}{dt} \left(\frac{\dot{\pi}}{\|\dot{\pi}\|} \right) = 0$$

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



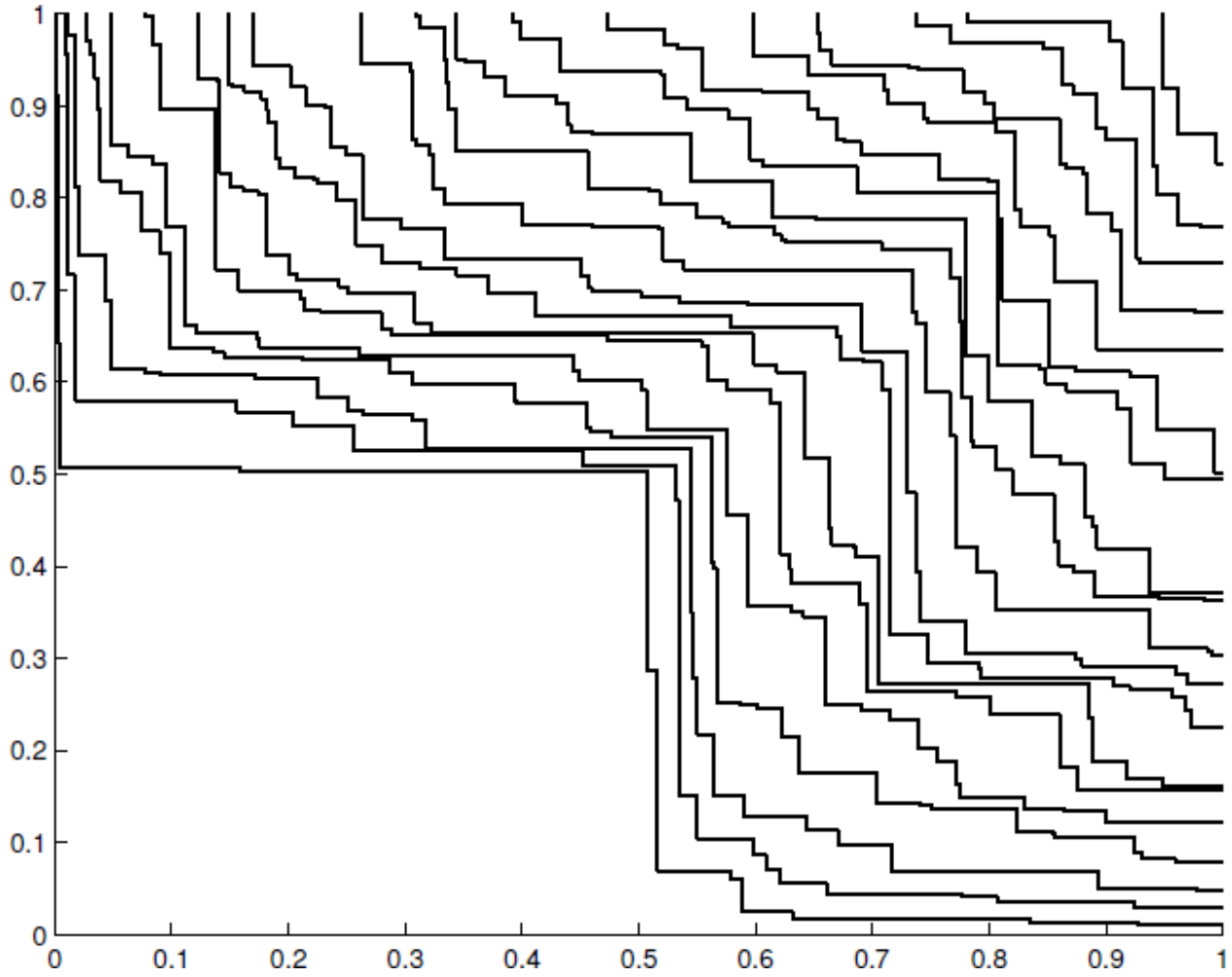
Regression equation ($\alpha = (1-\gamma)/d$):

$$\log L_\gamma(\mathcal{X}) = \alpha \log n + \log dist_\gamma(x,y) + \log C_{d,\gamma}$$

Experimental setting

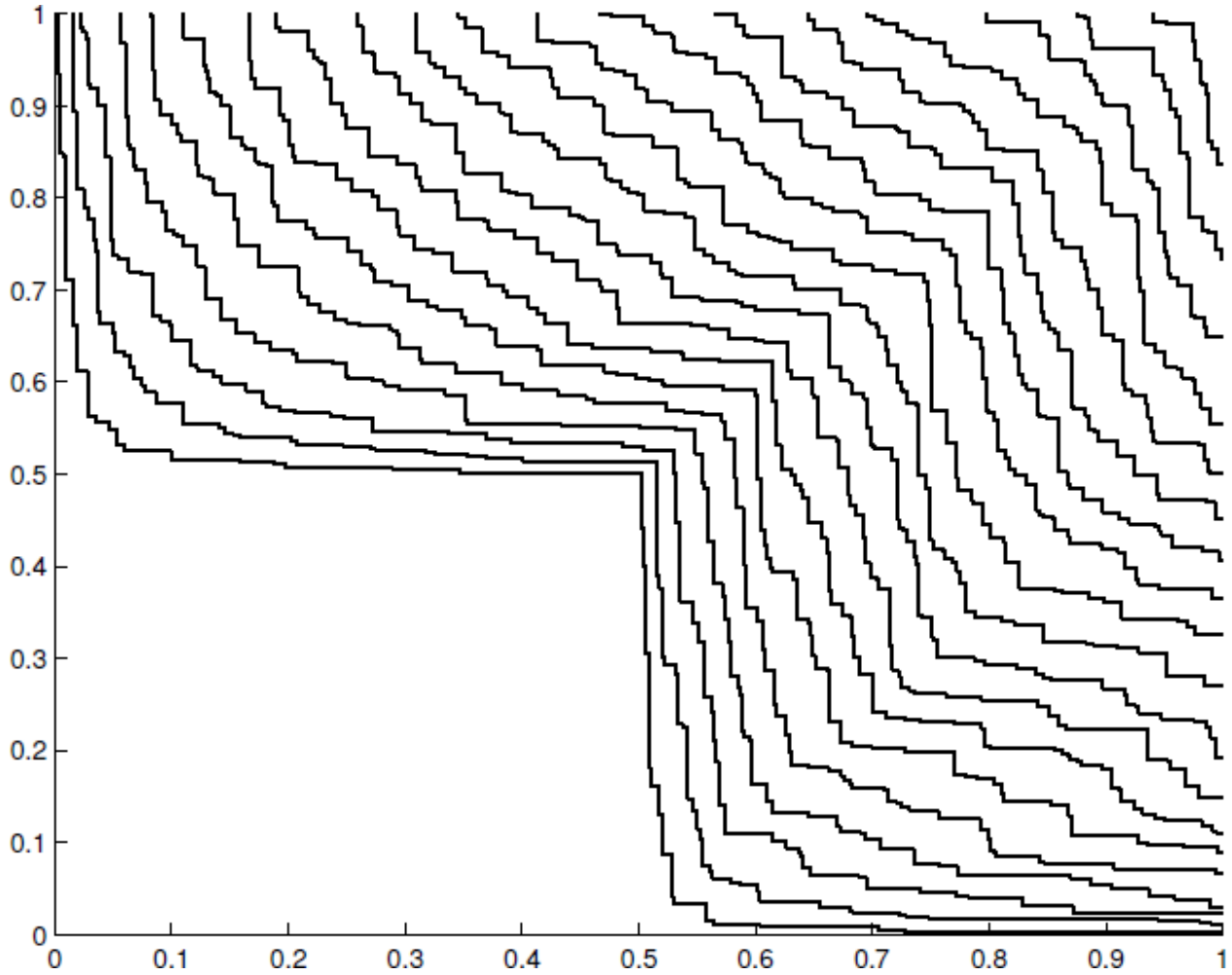
- $d = 2$, $\gamma = 2$ so that slope should be $(1-\gamma)/d = -0.5$
- $\mathcal{X}_n$ are n uniform points on $\mathcal{S} = S^2$
- Blue plot: $x = (1,0,0)$, $y = (-1,0,0)$
- Red plot: $x = (0,1,0)$, $y = (0,0,1)$

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



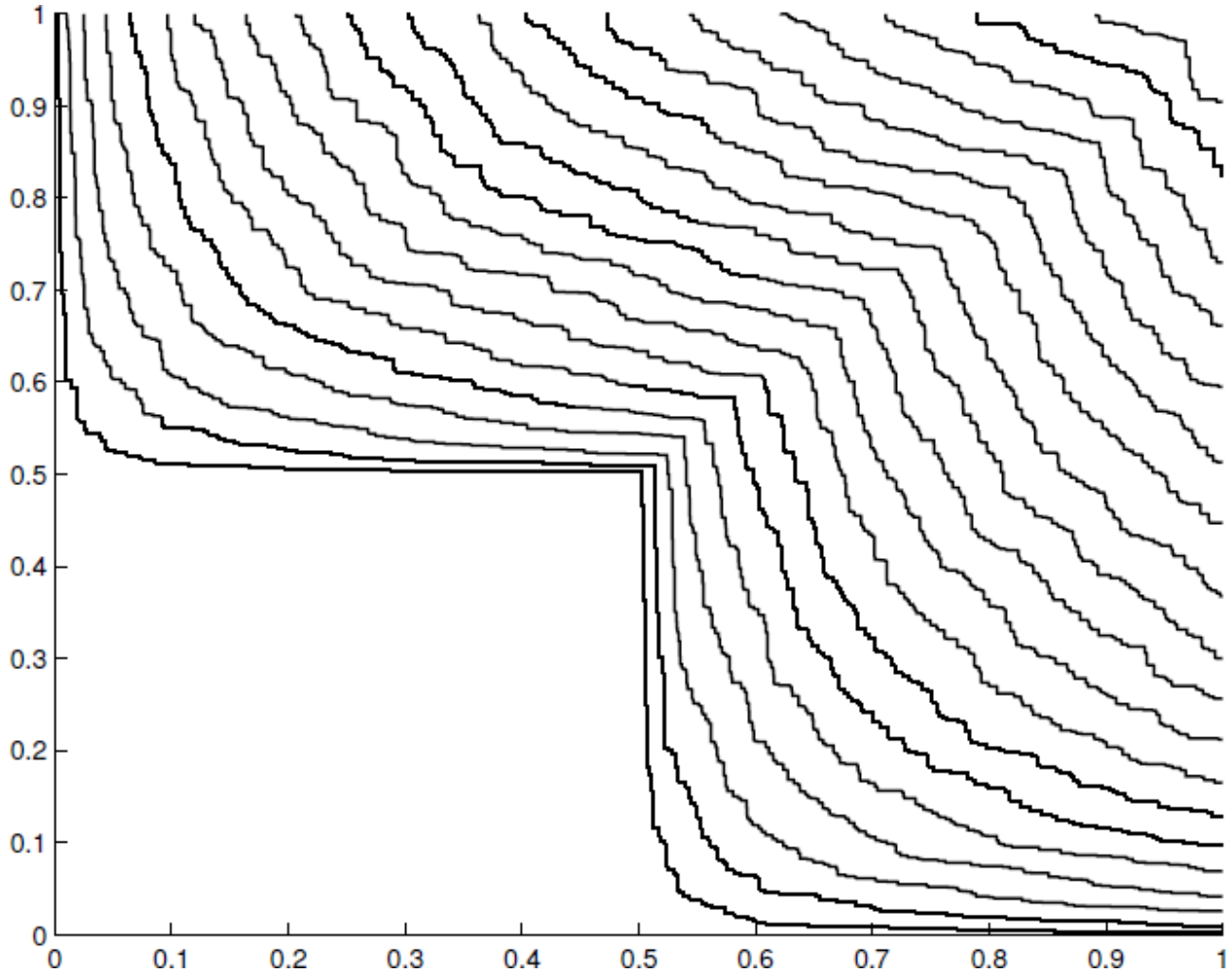
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



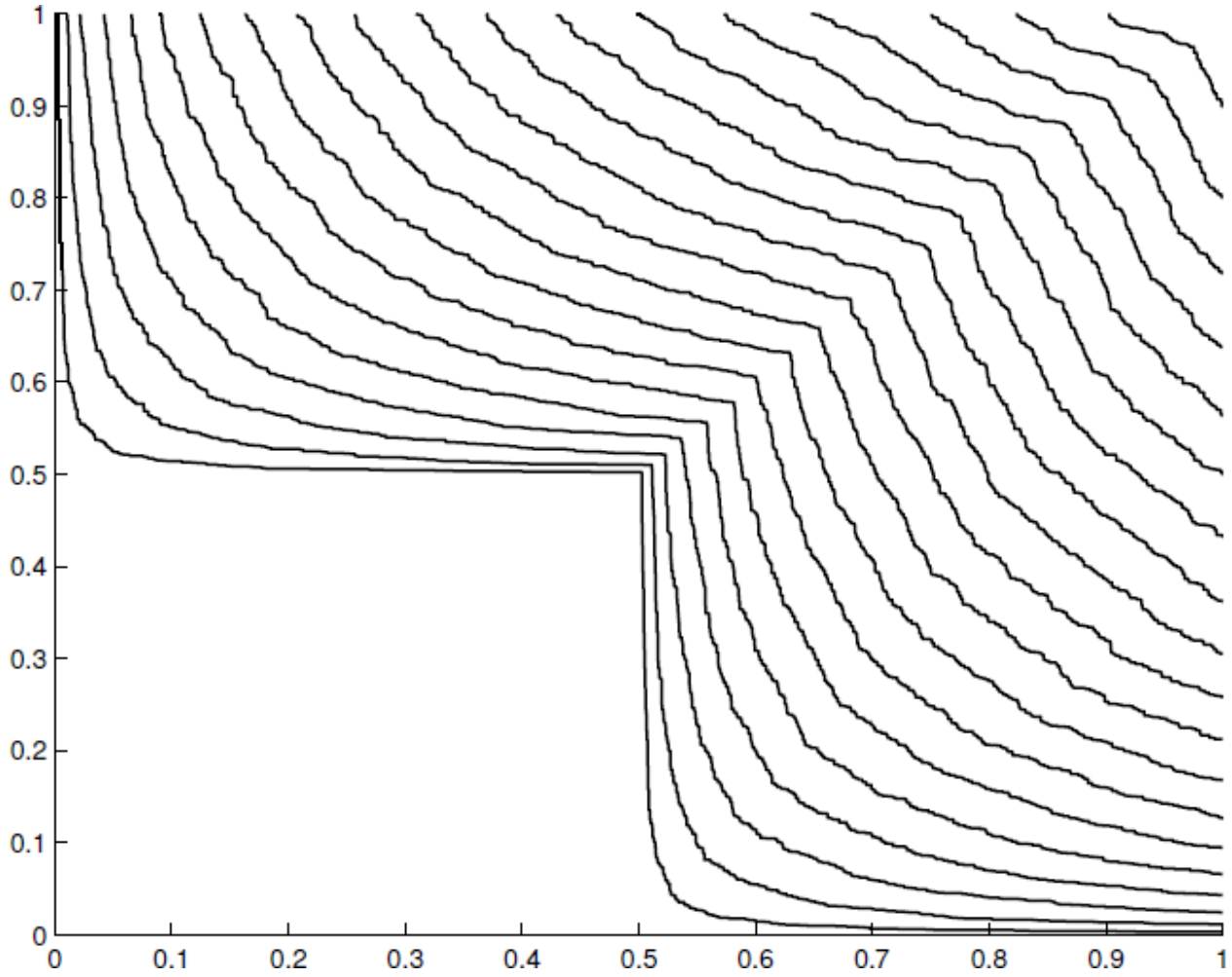
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Asymptotic theorem for non-dominated sorting

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

Theorem (Calder, Esedoglu and H, 2014)

There exists a $c_d > 0$ such that w.p.1

$$n^{-1/d} u_n \;\rightarrow\; c_d U, \quad \text{in } L^\infty(\mathbb{R}^d_+)$$

where

1. *U is the Pareto monotone [a] solution of the variational problem*

$$U(\mathbf{x}) = \sup_{\gamma \in \mathcal{A}} \int_0^1 f^{\frac{1}{d}}(\gamma(t))(\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt$$

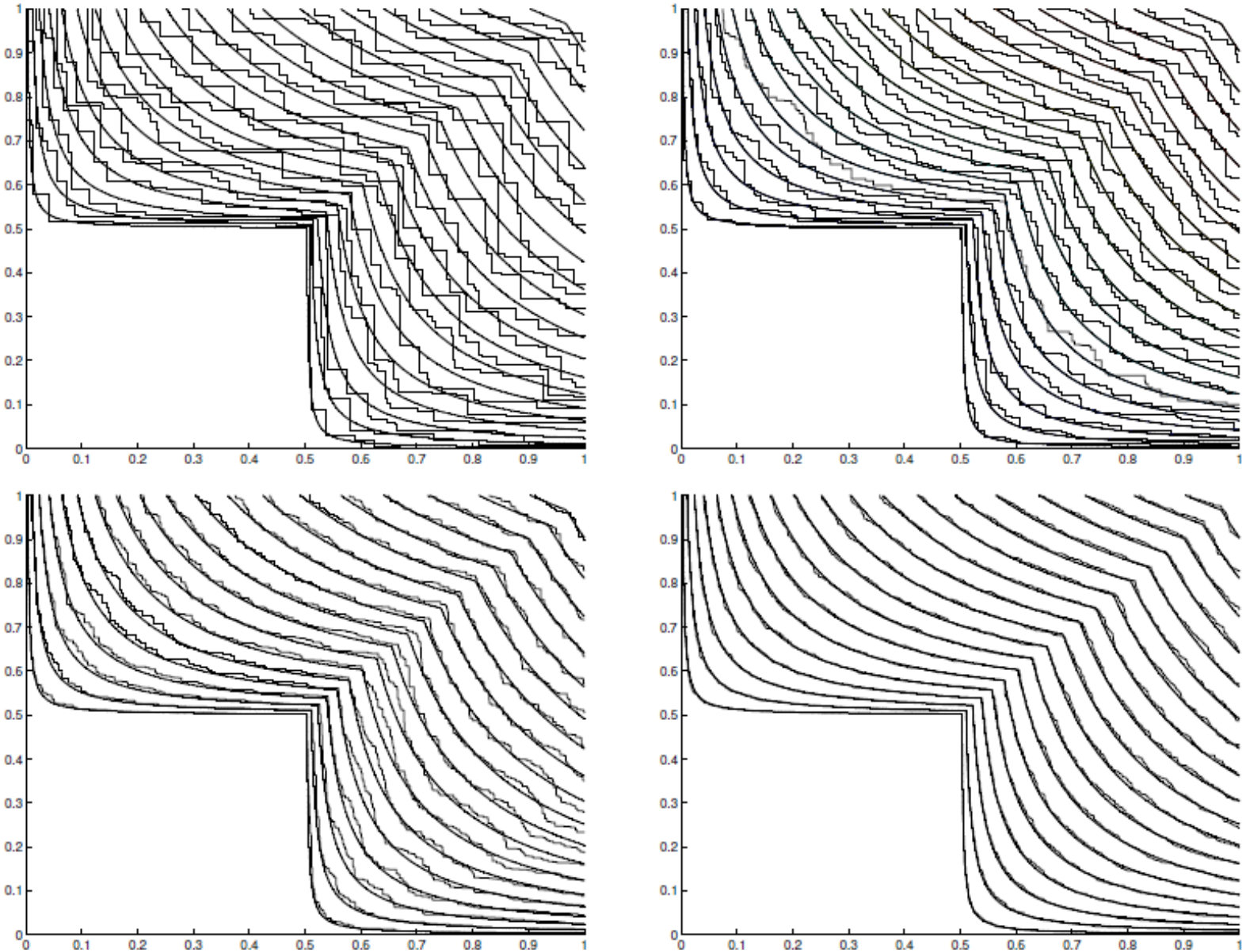
where $\mathcal{A} = \left\{\gamma \in C^1(0,1;\mathbb{R}^d) : \gamma'(t) \geqq 0 \;\forall t \in [0,1]\right\}$

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

$$\begin{aligned} \frac{\partial U}{\partial x_1} \cdots \frac{\partial U}{\partial x_d} &= \frac{1}{d^d} f \text{ in } \Omega \\ U &= 0 \text{ on } \partial\Omega \end{aligned}$$

[a] $U(\mathbf{x}) \leq U(\mathbf{y})$ if $\mathbf{x} \leqq \mathbf{y}$

Demonstration: theory vs experiment for Unif$[0,1]/[0,0.5]^2$

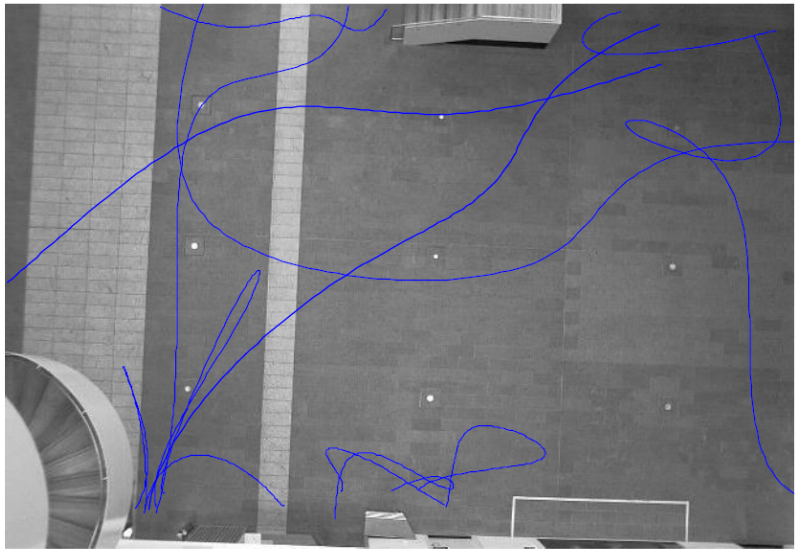


J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Outline

Multicriteria anomaly detection

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.

K.-J. Hsiao, K. Xu, J. Calder and A. Hero, "Multi-criteria anomaly detection using Pareto depth analysis," NIPS 2012.

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

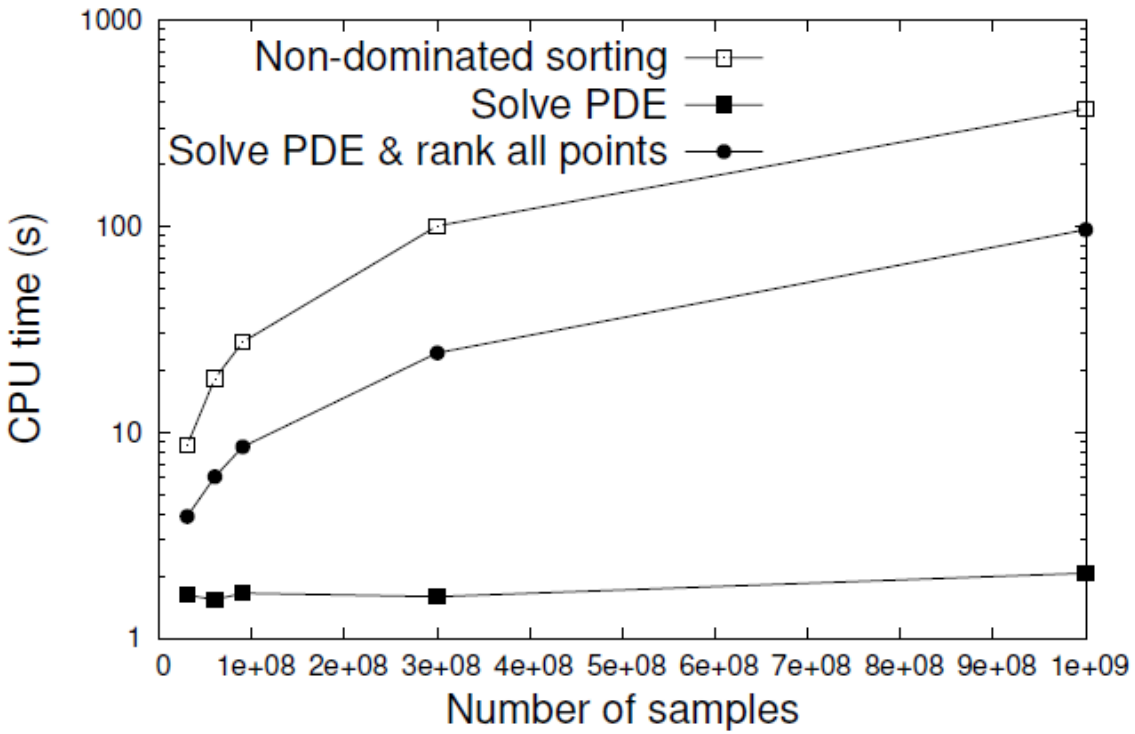
Curve similarity sorting for 50,000 curves (10^9 pairs)

- 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: 250×250

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: 250×250
- Non-dominated Pareto sorting by pde increases in #samples as O(1)



Calder, Esedoglu, H, "A PDE-based approach to non-dominated sorting," SIAM Numerical Analysis, 2015. Calder et al. [2015]

Outline

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

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

J. Calder, S. Esedoglu, and AO Hero. A hamilton-jacobi equation for the continuum limit of non-dominated sorting. *SIAM Journ on Mathematical Analysis (arXiv:1302.5828)*, 46(1):603–638, 2014.

Jeff Calder, Selim Esedoglu, and Alfred O Hero. A pde-based approach to non-dominated sorting. *SIAM Numerical Analysis (arXiv:1310.2498)*, 53(1):82–104, 2015.

K.J. Hsiao, K.S. Xu, and A.O. Hero III. Multi-criteria anomaly detection using pareto depth analysis. In *Proceeedings of NIPS 2012, also available as Arxiv preprint arXiv:1110.3741*, 2012.

Ko-Jen Hsiao, Jeff Calder, and Alfred O Hero III. Pareto-depth for multiple-query image retrieval. *arXiv preprint arXiv:1402.5176*, 2014.

Sung Jin Hwang, Steven B Damelin, and Alfred O Hero III. Shortest path through random points. *arXiv preprint arXiv:1202.0045*, 2012.

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