Continuum limits: a promising frontier for large scale data analysis

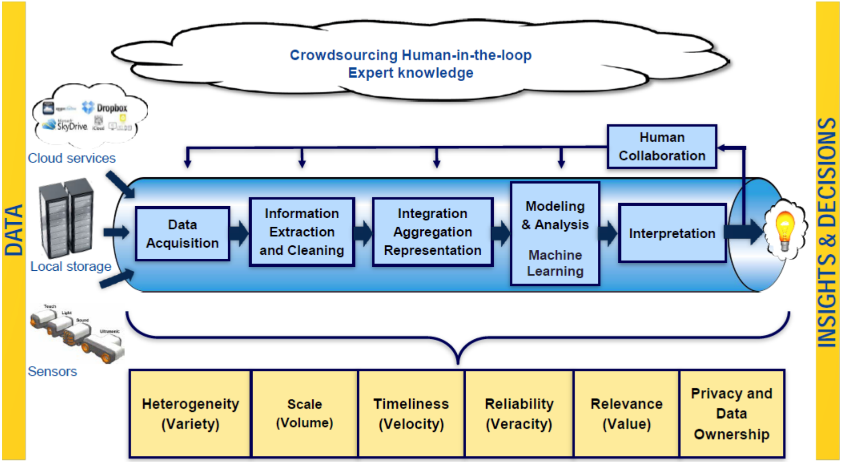
Alfred Hero

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

Sept 23, 2016

Outline

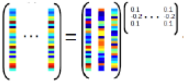
Data science as a pipeline from data to insights and decisions



Data science as a discipline at the interface

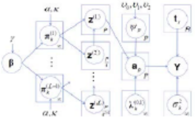
Mathematics:
Data as a matrix

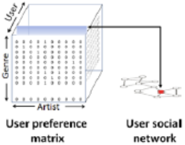
Applied topology
Harmonic analysis
Convex optimization
Num. linear algebra
Applied probability
Random matrix theory



Computer Science:
Data as a list/graph

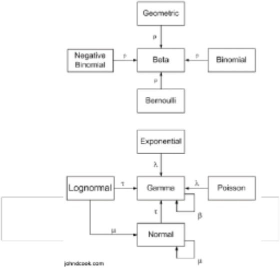
Natural language proc.
Graph theory
Algorithms
Database indexing
Machine learning
Privace and security





Statistics:
Data as a random sample

Sampling theory
Handling missing data
Robust procedures
Experimental design
Multivariate analysis
Graphical models



Data science as a discipline at the interface

Information Science
Data as an interface

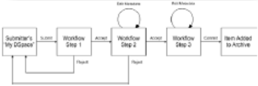
Engineering
Data2Decision

Physics
Data as natural phenomenon

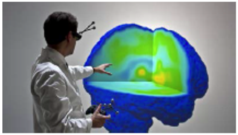
Human Computer Interaction (HCI)
Data sharing and reuse
Process and workflow
Data curation
Visualization

Comm. & info. theory
Signal processing
Sensors and control
Scheduling, RA and OR
Real-time computing
Cyberphysical systems

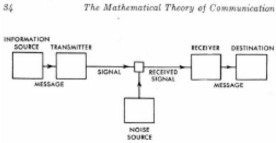
Network science
Complex systems
Statistical physics
Physico-mimetic models for data

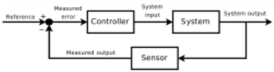


http://dspace.org/sites/dspace.org/

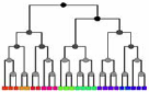


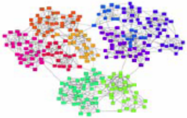
http://um3d.dc.umich.edu/visualization/





http://en.wikipedia.org/wiki/Control_theory





Mark Newman, UM Physics

Continuum limits in physics and applied math

Continuum limits are the basis for many results in applied physics and math

- Riemann integral limits of finite sums

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1=1}^{n} \psi(x_i) = \int_{\mathbb{R}^d} \psi(x) f(x) dx$$

Continuum limits in physics and applied math

Continuum limits are the basis for many results in applied physics and math

- Riemann integral limits of finite sums

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1=1}^{n} \psi(x_i) = \int_{\mathbb{R}^d} \psi(x) f(x) dx$$

- Limits of finite particle systems in statistical mechanics
  - Thermodynamic limit for magnetic systems (Ising 1925, Onsager 1948)
  - Boltzman hydrodynamic limit for dilute gasses (Bardo 1991)
  - Hamilton-Jacobi diffusion limit for non-ideal gases (Rajeev 2008)

Ising, Ernst (1925), Beitrag zur Theorie des Ferromagnetismus. Z. Phys., 31: 253258,

Bardos, C, F. Golse and D. Levermore (1991), Fluid dynamic limits of kinetic equations. J. Stat. Pysics 63, 323 - 344

Rajeev, S.G. (2008), A HamiltonJacobi formalism for thermodynamics. Annals of Physics, 323(9), pp.2265-2285

Continuum limits in physics and applied math

Continuum limits are the basis for many results in applied physics and math

- Riemann integral limits of finite sums

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1=1}^{n} \psi(x_i) = \int_{\mathbb{R}^d} \psi(x) f(x) dx$$

- Limits of finite particle systems in statistical mechanics
    - Thermodynamic limit for magnetic systems (Ising 1925, Onsager 1948)
    - Boltzman hydrodynamic limit for dilute gasses (Bardo 1991)
    - Hamilton-Jacobi diffusion limit for non-ideal gases (Rajeev 2008)

These latter limits often reduce the free energy of a complex system to simpler (maximum entropy) solutions to partial differential equations (Evans 2001).

Ising, Ernst (1925), Beitrag zur Theorie des Ferromagnetismus. Z. Phys., 31: 253258,

Bardos, C, F. Golse and D. Levermore (1991), Fluid dynamic limits of kinetic equations. J. Stat. Pysics 63, 323 - 344

Rajeev, S.G. (2008), A HamiltonJacobi formalism for thermodynamics. Annals of Physics, 323(9), pp.2265-2285

Evans, Lawrence C. (2001). Entropy and partial differential equations. URL math. berkeley. edu/evans

Continuum limits in physics and applied math

Such limits have often motivated discrete approximations to cts operators

- Approximation of integrals by quadrature (Gaussian, Nyström) methods
- Approximation of differential equations by finite differences (Euler, Runge-Kutta)

and construction of asymptotic performance approximations

- Dense network approximations to wireless communication (Gupta and Kumar 2000)
- Fluid approximations to queuing networks (Dai and Meyn 1995)
- High dimensional approximations to eigenspectra of random matrices (Silverstein 1995)

Gupta, Piyush, and PR Kumar (2000). The capacity of wireless networks. IEEE Transactions on information theory 46:2: 388-404.

Dai, Jim G., and Sean P. Meyn (1995). Stability and convergence of moments for multiclass queueing networks via fluid limit models. IEEE Transactions on Automatic Control 40:11: 1889-1904.

Silverstein, Jack W., and Z. D. Bai (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. Journal of Multivariate analysis 54.2: 175-192.

Continuum limits in data science?

Q. Are continuum limits useful for machine learning and data mining?

A. Yes. Continuum limits often reveal scalable approximations for large sample size

Continuum limits in data science?

Q. Are continuum limits useful for machine learning and data mining?

A. Yes. Continuum limits often reveal scalable approximations for large sample size

Some examples

- Nyström low rank approximations for kernel-based learning (Drineas and Mahoney, 2005)
- Information divergence from limit of MST (Henze-Penrose 1999)
- Minimum volume sets from limit of K-point MST (Hero 1998)
- Intrinsic dimension from continuum limit of MST growth rate (Hero 2006)
- Pareto non-dominated sorting from Hamilton-Jacobi continuum limit (Hero 2014)
- Dykstra shortest paths from Euler-Lagrange continuum limit (Hero 2016)

Continuum limits in data science?

Q. Are continuum limits useful for machine learning and data mining?

A. Yes. Continuum limits often reveal scalable approximations for large sample size

Some examples

- Nyström low rank approximations for kernel-based learning (Drineas and Mahoney, 2005)
- Information divergence from limit of MST (Henze-Penrose 1999)
- Minimum volume sets from limit of K-point MST (Hero 1998)
- Intrinsic dimension from continuum limit of MST growth rate (Hero 2006)
- Pareto non-dominated sorting from Hamilton-Jacobi continuum limit (Hero 2014)
- Dykstra shortest paths from Euler-Lagrange continuum limit (Hero 2016)

$\rightarrow$ Euclidean graph continuum limits appear especially promising

Drineas, Peter and Mahoney, Michael W. (2005), On the Nyström method for approximating a Gram matrix for improved kernel-based learning. Journal of Machine Learning Research, 6:12: 2153-2175.

Geometric graphs

A geometric graph has nodes $\mathcal{V}$ that represent real valued features and edges $\mathcal{E}$ that represent similarities between the features (Penrose 2003).

Some data-driven applications where geometric graphs arise

- Data mining
  - Clustering and segmentation (GLap, kNNG, MST, graph cuts)
  - Dimensionality reduction (GLap, kNNG, GMST)
  - Denoising and anomaly detection (kMST, BP-kNNG)
- Imaging and computer vision
  - Orthoregistration (MST, kNNG)
  - Frame-to-frame registration (TSP)
  - Multi-resolution image representation (MST-based pyramid)
  - Image inpainting interpolation (kNNG)
- Database indexing and retrieval
  - Query-reference matching (NNG)
  - Database partitioning (kNNG)
  - Multi-criterion image retrieval (Chain graph)

Such geometric graphs are often modeled as random, having nodal feature vectors $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ drawn from some probability distribution f.

M. Penrose, Random geometric graphs, Oxford University Press 2003

Outline

Minimal Euclidean graphs under constraints

Define $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ a set of points (features) in $\mathcal{M} \subset \mathbb{R}^d$.

A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$

- $\{\mathcal{V}\} = \{X_1, \ldots, X_n\}$, $\mathbf{X}_i \in \mathcal{M} \subset \mathbb{R}^d$: nodes or vertices
- $\{\mathcal{E}\} = \{e_{ij}\}$: edges connecting distinct pairs $\{i, j\}$
- $|e_{ij}| = \|X_i - X_j\|$: edge length wrt to a distance metric on $\mathcal{M}$
- $\mathbf{A} = ((a_{ij}))$: adjacency matrix associated with $\mathcal{G}$

$$a_{ij} = \begin{cases} 1, & e_{ij} \in \mathcal{E} \\ 0, & o.w. \end{cases}$$

- $d_i = \sum_j a_{ij}$: degree of vertex i

Minimal Euclidean graphs under constraints

Define $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ a set of points (features) in $\mathcal{M} \subset \mathbb{R}^d$.

A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$

- $\{\mathcal{V}\} = \{X_1, \ldots, X_n\}$, $\mathbf{X}_i \in \mathcal{M} \subset \mathbb{R}^d$: nodes or vertices
- $\{\mathcal{E}\} = \{e_{ij}\}$: edges connecting distinct pairs $\{i, j\}$
- $|e_{ij}| = \|X_i - X_j\|$: edge length wrt to a distance metric on $\mathcal{M}$
- $\mathbf{A} = ((a_{ij}))$: adjacency matrix associated with $\mathcal{G}$

$$a_{ij} = \begin{cases} 1, & e_{ij} \in \mathcal{E} \\ 0, & o.w. \end{cases}$$

- $d_i = \sum_j a_{ij}$: degree of vertex i

Length functional

$$L(\mathcal{V}, \mathcal{E}) = \sum_{e_{ij} \in \mathcal{E}} |e_{ij}|^\gamma$$

where $\gamma \geq 0$. Given constraint set $\mathcal{C}$ a minimal Euclidean graph $\mathcal{G}^* = \{\mathcal{E}^*, \mathcal{V}\}$

is solution of

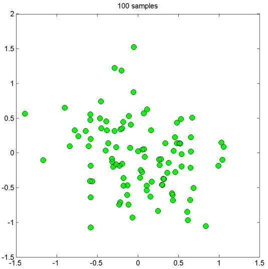
$$\mathcal{E}^* = \mathrm{amin}_{\mathcal{E}:\mathcal{E}\subset\mathcal{C}} \sum_{e_{ij} \in \mathcal{E}} |e_{ij}|^\gamma$$

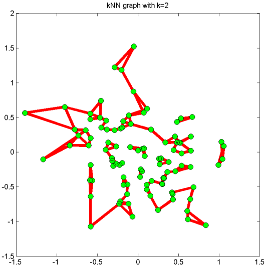
k-nearest neighbor (kNN) graph

- kNN graph is solution of the optimization

$$
\begin{aligned}
L_\gamma^{kNN}(\mathcal{V}) &= \min_{\mathcal{E}:\mathbf{A}\underline{1}\geq k\underline{1}} L_\gamma(\mathcal{V},\mathcal{E}) \\
&= \min_{\mathcal{E}:\mathbf{A}\underline{1}\geq k\underline{1}} \sum_{e_{ij}\in\mathcal{E}} |e_{ij}|^\gamma \\
&= \sum_{i=1}^{n} \sum_{j\in\mathcal{N}_k(X_i)} \|X_i - X_j\|^\gamma
\end{aligned}
$$

- $\mathcal{N}_k(X_i)$ are the k-nearest neighbors of X_i in $\mathcal{X}_n - \{X_i\}$
- Applications: inpainting, feature density estimation, clustering+classification, dimensionality reduction
- Computational complexity is $O(kn\log n)$



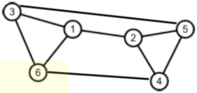


kNNGs in spectral clustering and dimensionality reduction

k-NNG-based spectral algorithm

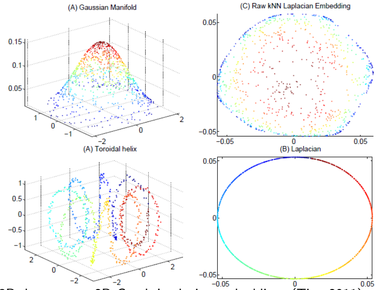
- Extract features $\mathcal{X}_n = \{X_1, \ldots, X_n\}$
- Compute similarity matrix $\mathbf{W}$ btwn X_i's
- Use $\mathbf{W}$ to construct kNN graph over $\mathcal{X}_n$
- $(\mathbf{V}, \mathbf{\Lambda})$ =Eigendecomp$(\mathbf{W} - \mathbf{D})$, $\mathbf{D} = \mathrm{diag}(\mathbf{W}\underline{\mathbf{1}})$
  - Dimension reduction: $\mathbf{Y}_n = \mathbf{\Lambda}^{1/2}_{2\times 2}[\mathbf{v}_1, \mathbf{v}_2]^T \mathbf{X}_n$
  - Spectral clustering: K-means$(\mathbf{v}_2)$

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$



Adjacency matrix kNNG

kNNG clustering for image segementation (Felzenszwalb 2003)



3D data 2D Graph Laplacian embeddings (Ting 2011)

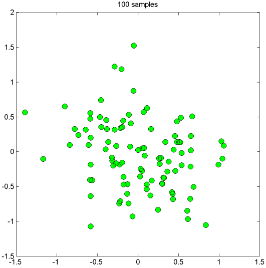
- Belkin, Mikhail, and Partha Niyogi. "Laplacian eigenmaps and spectral techniques for embedding and clustering." NIPS. Vol. 14. 2001.
- Coifman, Ronald R., and Stphane Lafon. "Diffusion maps." Applied and computational harmonic analysis 21.1 (2006): 5-30.

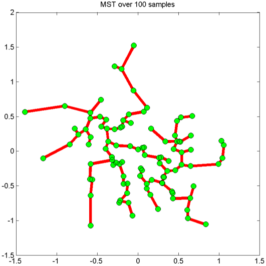
Minimal spanning tree (MST)

- MST is solution of the optimization

$$L_\gamma^{MST}(\mathcal{V}) = \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E})$$
$$= \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} \sum_{e_{ij}\in\mathcal{E}} |e_{ij}|^\gamma$$

- MST spans all of the vertices $\mathcal{V}$ without cycles
- MST has exactly $n-1$ edges
- Applications: image segmentation, image registration, clustering
- Computational complexity is $O(n^2 \log n)$



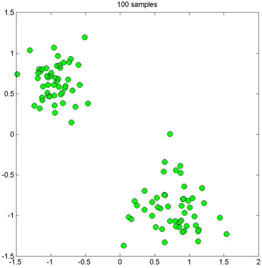


Minimal spanning tree (MST)

- MST is solution of the optimization

$$
\begin{aligned}
L_\gamma^{MST}(\mathcal{V}) &= \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E}) \\
&= \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} \sum_{e_{ij}\in\mathcal{E}} |e_{ij}|^\gamma
\end{aligned}
$$

- MST spans all of the vertices $\mathcal{V}$ without cycles
- MST has exactly $n-1$ edges
- Applications: image segmentation, image registration, clustering
- Computational complexity is $O(n^2 \log n)$



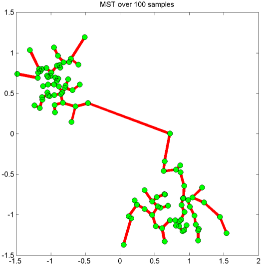
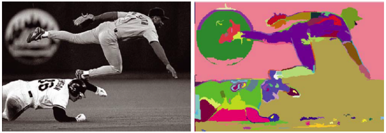
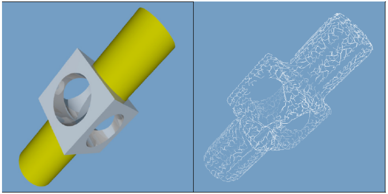


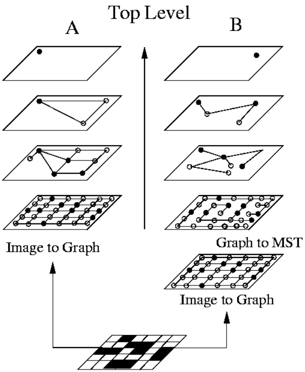
Illustration: MST for image segmentation, representation and rendering



MST-based image segmentation (Zahn 1971, Felzenszwalb 2003)



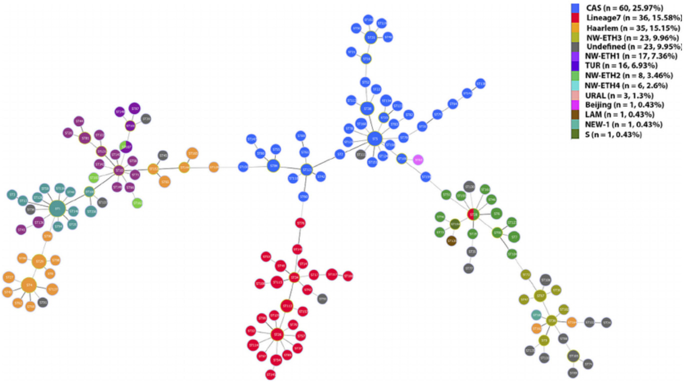
MST for surface rendering (Hoppe 1992))



MST for building image pyramid (Mathieu 1996)

- Zahn, Charles T. "Graph-theoretical methods for detecting and describing gestalt clusters." IEEE Transactions on Computers, 1971
- P. Felzenswalb and D. Huttenlocher, "Efficient graph-based image segmentation," International Journal of Computer Vision, 2004
- H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," SIGRAPH, 1992
- C. Mathieu and I. Magnin, "On the choice of the first level on graph pyramids", Journal of Mathematical Imaging and Vision, 1996

Minimal spanning tree for liineage tracking in epidemiology



Minimum-spanning tree (MST) of Mycobacterium tuberculosis strains based on MIRU-VNTR 24-locus copy numbers. The M. tuberculosis clonal complexes are represented by different colors. Circle size is proportional to the number of MIRU-VNTR types belonging to each complex. Abbreviations: CAS, Central Asian strain; LAM, Latin American-Mediterranean.

Yimer *et al*, "Mycobacterium tuberculosis Lineage 7 Strains Are Associated with Prolonged Patient Delay in Seeking Treatment for

Friedman-Rafsky graph (FR)

- Two labeled samples $\mathcal{X}_n$, $\mathcal{Y}_m$
- Start with MST over $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

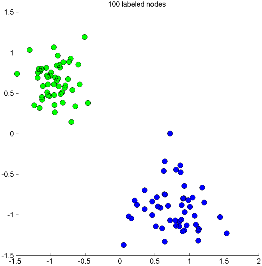
$$L_\gamma^{MST}(\mathcal{V}) = \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E})$$

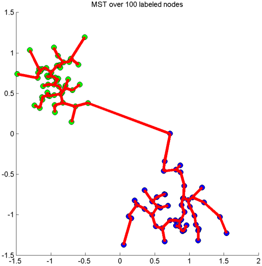
$$= \sum_{e_{ij}\in\mathcal{E}*} |e_{ij}^{XX}|^\gamma + |e_{ij}^{XY}|^\gamma + |e_{ij}^{YY}|^\gamma$$

- FR graph is the set of edges $\{e_{ij}^{XY}\}$
- The length of FR graph is

$$L_\gamma^{FR}(\mathcal{V}) = \sum_{e_{ij}^{XY}\in\mathcal{E}*} |e_{ij}^{XY}|^\gamma$$

- This was proposed as a difference measure (divergence) btwn distributions of $\mathcal{X}_n$ and $\mathcal{Y}_m$ (Friedman and Rafsky, 1979)





J. Friedman, and L. Rafsky. "Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests," Annals of Statistics, 1979.

Friedman-Rafsky graph (FR)

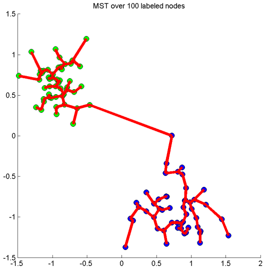
- Two labeled samples $\mathcal{X}_n$, $\mathcal{Y}_m$
- Start with MST over $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

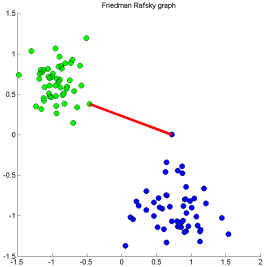
$$L_\gamma^{MST}(\mathcal{V}) = \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E}) = \sum_{e_{ij}\in\mathcal{E}*} |e_{ij}^{XX}|^\gamma + |e_{ij}^{XY}|^\gamma + |e_{ij}^{YY}|^\gamma$$

- FR graph is the set of edges $\{e_{ij}^{XY}\}$
- The length of FR graph is

$$L_\gamma^{FR}(\mathcal{V}) = \sum_{e_{ij}^{XY}\in\mathcal{E}*} |e_{ij}^{XY}|^\gamma$$

- This was proposed as a difference measure (divergence) btwn distributions of $\mathcal{X}_n$ and $\mathcal{Y}_m$ (Friedman and Rafsky, 1979)





Friedman-Rafsky graph (FR)

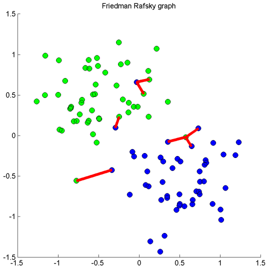
- Two labeled samples $\mathcal{X}_n$, $\mathcal{Y}_m$
- Start with MST over $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

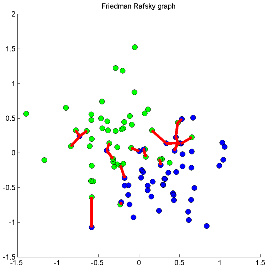
$$\begin{aligned} L_\gamma^{MST}(\mathcal{V}) &= \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E}) \\ &= \sum_{e_{ij}\in\mathcal{E}*} |e_{ij}^{XX}|^\gamma + |e_{ij}^{XY}|^\gamma + |e_{ij}^{YY}|^\gamma \end{aligned}$$

- FR graph is the set of edges $\{e_{ij}^{XY}\}$
- The length of FR graph is

$$L_\gamma^{FR}(\mathcal{V}) = \sum_{e_{ij}^{XY}\in\mathcal{E}*} |e_{ij}^{XY}|^\gamma$$

- Applications: image registration, pattern matching, meta-learning
- Computational complexity is $O((n+m)^2 \log(n+m))$





Application: multimodality image registration using MI

Find transformation T that best aligns images I_1 and I_2

Feature vector at location $\mathbf{z}_i \in \mathbb{R}^2$:
$\mathbf{X}(i) = [I_1(z_i), T(I_2(z_i))]$

Joint intensity histogram
$p_{\mathbf{X}}(x_1, x_2) = n^{-1} \sum_{i=1}^{n} \mathcal{X}_{[x_1, x_2]}(\mathbf{X}(i))$

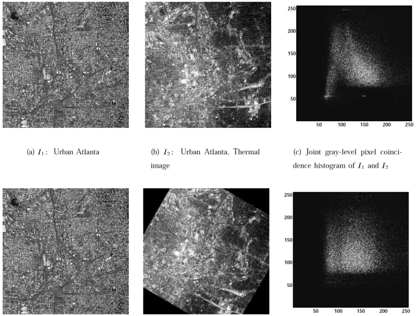
Maximize mutual information (MI)

$$\max_T \sum_{x_1, x_2 = 0}^{255} p_{\mathbf{X}}(x_1, x_2) \ln \left(\frac{p_{\mathbf{X}}(x_1, x_2)}{p_{X_1}(x_1) p_{T(X_2)}(x_2)} \right)$$
$$= \max_T H(I_1, T(I_2)) - H(I_1) - H(T(I_2))$$

Where have defined entropy of $\mathbf{V}$

$$H(\mathbf{V}) = n^{-1} \sum_v \ln \frac{1}{p_V(v)}$$

Mutual information (MI) based registration



(a) I_1: Urban Atlanta

(b) I_2: Urban Atlanta, Thermal image

(c) Joint gray-level pixel coincidence histogram of I_1 and I_2

(d) I_1

(e) $T(I_2)$

(f) Joint gray-level pixel coincidence histogram of I_1 and $T(I_2)$

- W. Wells, P. Viola, P., H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," Medical image analysis, 1996.
- E. Oubel, M. De Craene, A. Hero, A. Pourmorteza, M. Huguet, G. Avegliano, B. Bijnens, A. Frangi, "Cardiac motion estimation by joint alignment of tagged MRI sequences," Med. Image Anal. 2012.

Comparison: multimodality image registration using FR

Find transformation T that best aligns images I_1 and I_2

Feature vectors of I_1 and $T(I_2)$ at location $\mathbf{z}_i \in \mathbb{R}^2$:

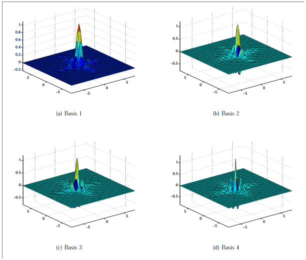
$$\mathbf{X}_1(i) = [\mathbf{W}(\mathbf{z}_i), \mathbf{z}_i], \;\; \mathbf{X}_2(i) = [\mathbf{W}(\mathbf{z}_i), \mathbf{z}_i]$$

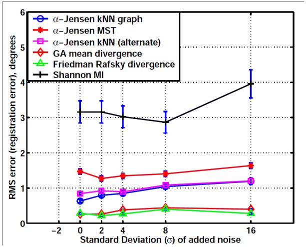
$\mathbf{W}_1(\mathbf{z}_i)$ and $\mathbf{W}_2(\mathbf{z}_i)$ are localized Meyer wavelet coefficients of I_1 and $T(I_2)$

Maximize FR statistic

$$\max_T L_\gamma^{FR}(\mathbf{X}_1, \mathbf{X}_2)$$

FR registration uses higher dimensional (6) features that capture images' local spatial patterns





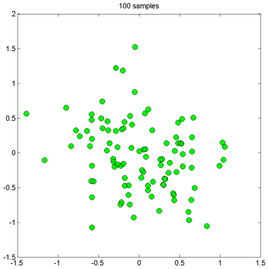
- H. Neemwuchwala and A. Hero, "Entropic Graphs for Registration," in Multi-Sensor Image Fusion and its Applications, Eds. R. S. Blum and Z. Liu, Marcel Dekker, Inc., 2005.

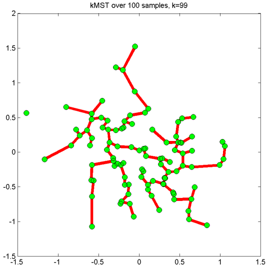
k-Minimal spanning tree (kMST)

- Let $\mathcal{V}_k \subset \mathcal{V}$ and $|\mathcal{V}_k| = k$
- Let $\mathcal{E}_k$ be edges over $\mathcal{V}_k$
- kMST is solution of the optimization

$$
\begin{aligned}
L_\gamma^{kMST}(\mathcal{V}) &= \min_{\mathcal{V}_k : |\mathcal{V}_k| = k} L_\gamma^{MST}(\mathcal{V}_k) \\
&= \min_{\mathcal{V}_k : |\mathcal{V}_k| = k} \; \min_{\mathcal{E}_k : \mathbf{A}_k \underline{1} > 0} \sum_{e_{ij} \in \mathcal{E}_k} |e_{ij}|^\gamma
\end{aligned}
$$

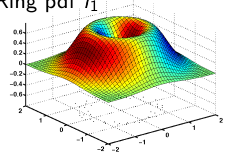
- kMST is the smallest MST that spans any k of the vertices $\mathcal{V}$
- Applications: Denoising and outlier detection, robust image registration, robust clustering
- Computational complexity is NP hard
- Greedy approximations are available (Ravi 1994)



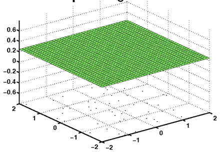


Denoising illustration of kMST

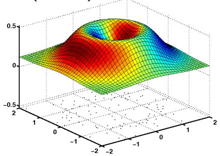
Ring pdf f_1

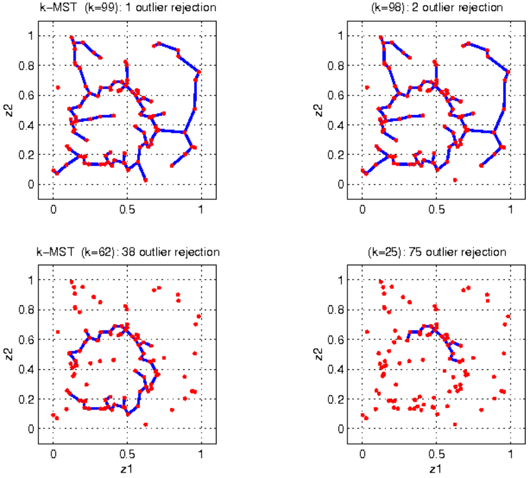


Uniform pdf f_0



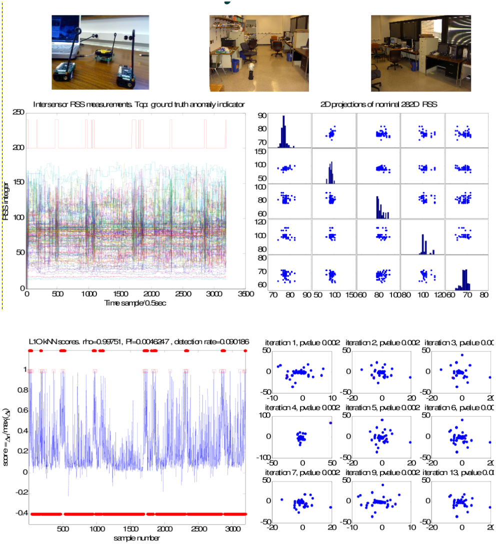
$f = (1 - \epsilon) f_1 + \epsilon f_0$





- A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal K-point random graphs," IEEE Information Theory 1999.

Illustration: kMST for WSN intruder detection



- A. Hero, "Geometric entropy minimization (GEM) for anomaly detection and localization," NIPS 2006
- K. Sricharan and A. Hero, "Efficient anomaly detection using bipartite k-NN graphs," NIPS 2011.

Shortest path (SP)

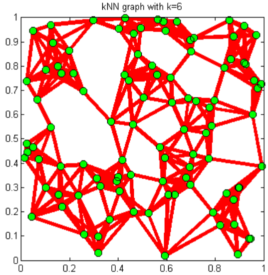
- Let $\mathcal{G}$ be a graph with $m = |\mathcal{E}|$ edges on n vertices $\mathcal{V}$
- $\pi(X_I, X_F)$ a path over $\mathcal{G}$ btwn points X_I and X_F

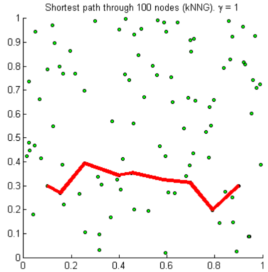
  $$\pi(X_I, X_F) = (X_I, X_{i_1}, \ldots, X_{i_l}, X_F)$$

  $X_{i_{j+1}}$ 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

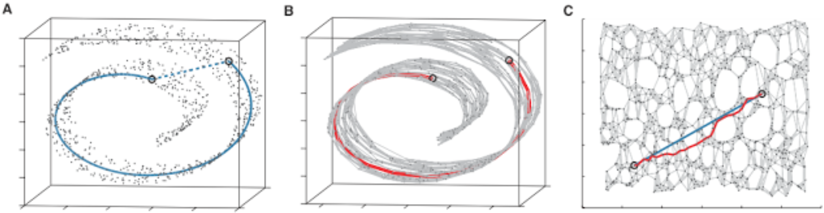


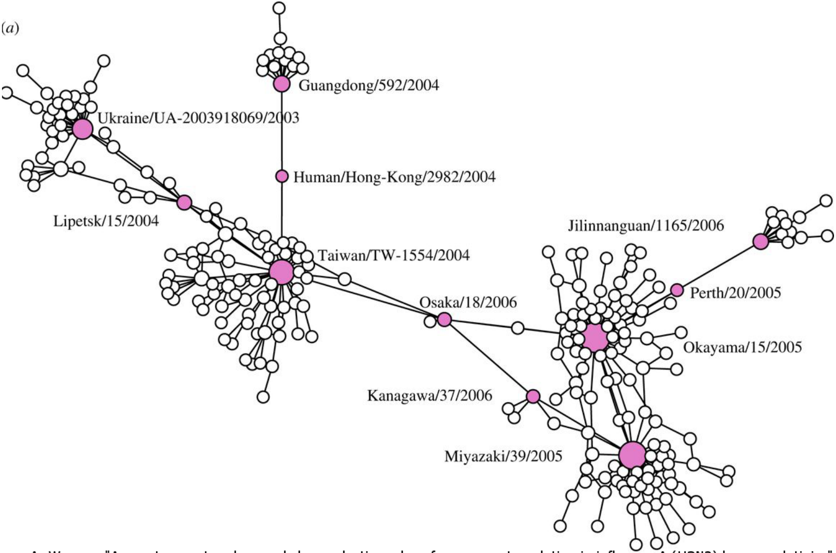
Fig. 3. The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (**A**) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect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

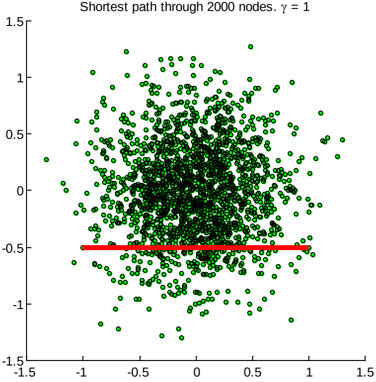
Averbuch-Elor, Cohen-Or and Kopf, "Smooth Image Sequences for Data Driven Morphing," Computer Graphics Forum, 35(6), 2016

Shortest paths in epidemiology: virus strain genotyping

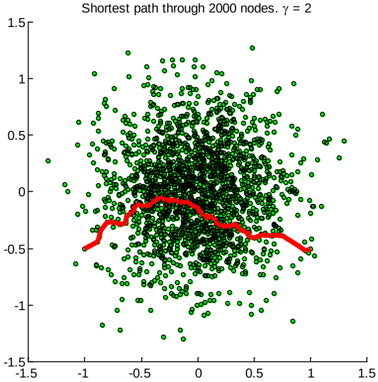


A. Wagner, "A genotype network reveals homoplastic cycles of convergent evolution in influenza A (H3N2) haemagglutinin," Proc. Royal Soc. B, May 2014.

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

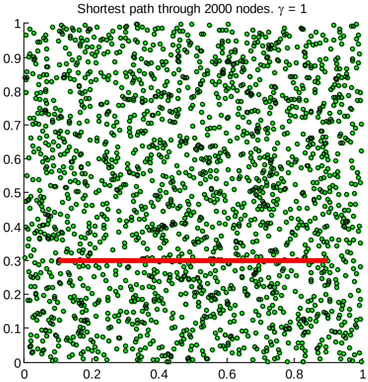


Euclidean ($\gamma = 1$)

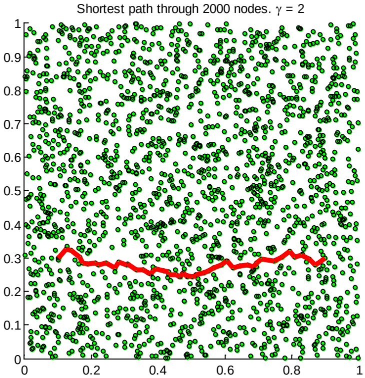


(Euclidean)^2 ($\gamma = 2$)

No lensing effect: SP through complete graph for uniform points in plane



Euclidean distance ($\gamma = 1$)



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

Non-dominated ranking in multiple dimensions

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

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

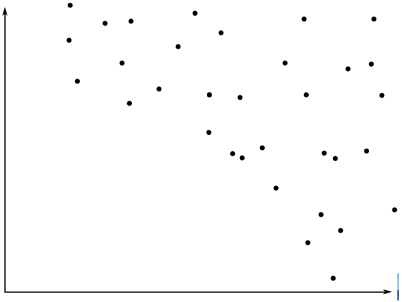
- $\mathbf{X}$ a minimal element of $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ if

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

- Define $\min \mathcal{X}$ the set (Pareto front) of all minimal elements of $\mathcal{X}$.
- Pareto front of depth k, denoted $\{\mathcal{F}_k\}$, 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 $O(dn^2)$



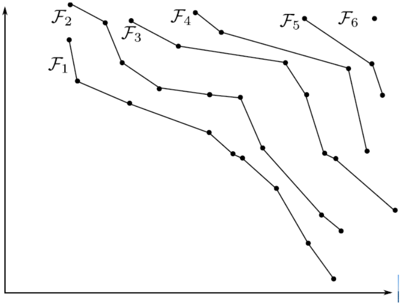


Illustration: Image retrieval combining multiple semantic concepts

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

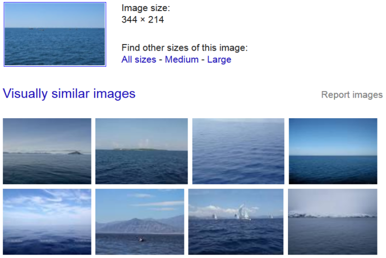

Query 2


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



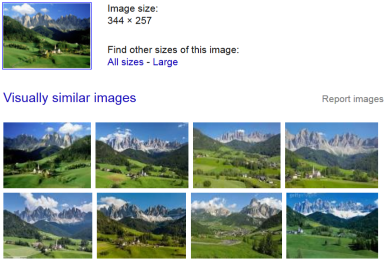
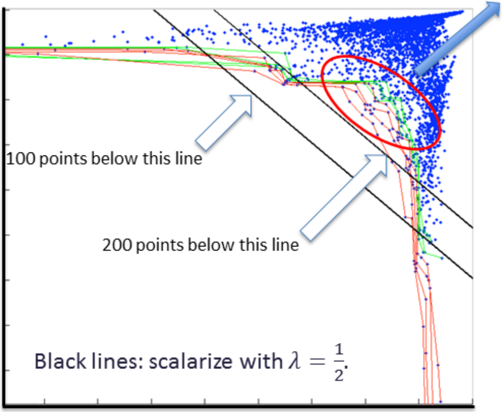


Illustration: multiple concept image retrieval in SS dataset



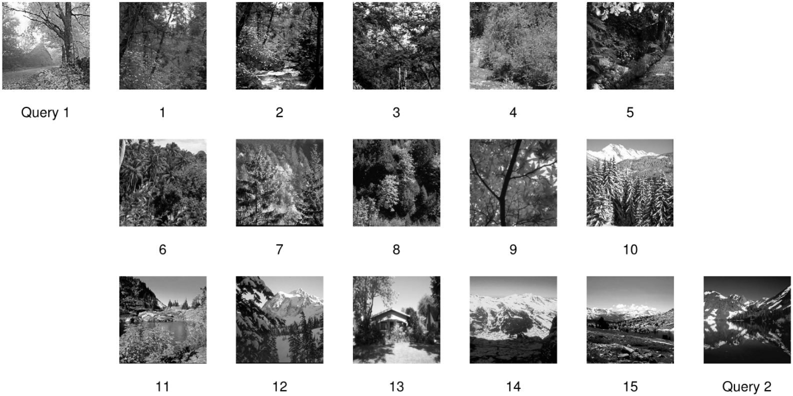
Pareto fronts give high ranks to points that are not highly ranked by linear scalarization.

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.

Illustration: first Pareto front for (forest, mountan) query

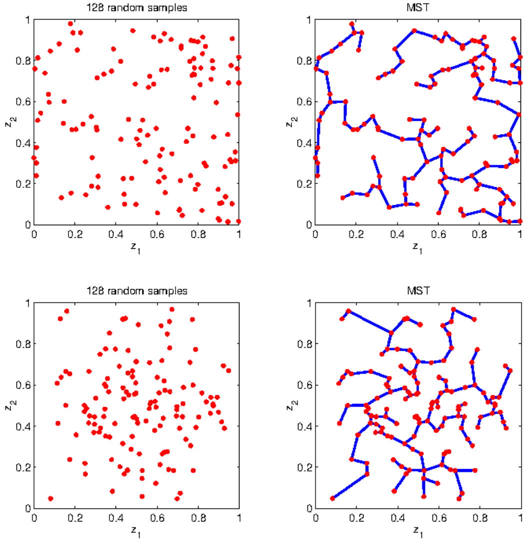


Stanford Scene dataset, SIFT feature, Spatial Pyramid Matching

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

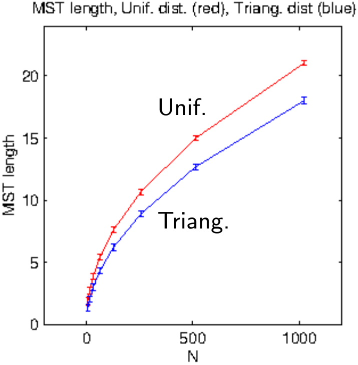
Outline

MST continuum limit: MST length functional captures "spread" of distribution

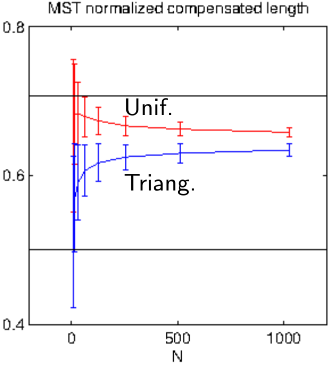


Large n behavior of MST length functional

length(MST)



(log length(MST))/$\sqrt{n}$



Continuum limit of kNN and MST length functionals

Theorem (Beardwood, Halton&Hammersley 1959, Steele 1997)

Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be an i.i.d. realization from a Lebesgue density f supported on compact subset of $\mathbb{R}^d$. If $0 < \gamma < d$

$$\lim_{n\to\infty} L_\gamma^{MST,kNN}(\mathcal{X}_n)/n^{(d-\gamma)/d} \quad = \quad \beta_{\gamma,d} \int f(x)^{(d-\gamma)/d} dx, \qquad (a.s.)$$

Alternatively, letting $\alpha = (d-\gamma)/d$ and defining the entropy function

$$H_\alpha(f) = \frac{1}{1-\alpha} \ln \int f^\alpha(x)dx,$$

$$\frac{1}{1-\alpha} \ln L_\gamma(\mathcal{X}_n)/n^\alpha \quad \to \quad H_\alpha(f) + c \qquad (a.s.)$$

- RMS rate of convergence (Costa & Hero 2003)

$$\sup_{f\in\mathcal{H}_{\beta,K}} E\left[\left|\beta_{\gamma,d} \int_{\mathcal{S}} f(x)^{(d-\gamma)/d} dx - L_\gamma^{MST}(\mathcal{X}_n)/n^{(d-\gamma)/d}\right|^2\right]^{1/2} \geq cn^{-\frac{\beta}{\beta+1}\frac{1}{d}}$$

Steele, *Probability theory and combinatorial optimization*, SIAM 1997.

Beardwood and Halton and Hammersley, "The shortest path through many points," Proc. Cambridge Philosophical Society 1959.

Continuum limit for Euclidean length functionals (Yukich 1998)

- BHH theorem holds generally for any quasi-additive continuous Euclidean length functional $L_\gamma(F)$ (Yukich 1998) - kNN, Steiner tree, TSP
  - Translation invariant and homogeneous

$$\begin{aligned} \forall\, x \in \mathbb{R}^d,\ \ L_\gamma(F+x) &= L_\gamma(\mathcal{F}), \quad \text{(translation invariance)} \\ \forall\, c>0,\ \ L_\gamma(cF) &= c^\gamma L_\gamma(\mathcal{F}), \quad \text{(homogeneity)} \end{aligned}$$

  - Null condition: $L_\gamma(\phi)=0$, where ϕ is the null set
  - Subadditivity: There exists a constant C_1 with the following property: For any uniform resolution $1/m$-partition $\mathcal{Q}^m$

$$L_\gamma(F) \le m^{-\gamma} \sum_{i=1}^{m^d} L_\gamma(m[(F \cap Q_i) - q_i]) + C_1 m^{d-\gamma}$$

  - Superadditivity: For same conditions as above, there exists a constant C_2

$$L_\gamma(F) \ge m^{-\gamma} \sum_{i=1}^{m^d} L_\gamma(m[(F \cap Q_i) - q_i]) - C_2 m^{d-\gamma}$$

  - Continuity: There exists a constant C_3 such that for all finite subsets F and G of $[0,1]^d$

$$|L_\gamma(F \cup G) - L_\gamma(F)| \le C_3\,(\mathrm{card}(G))^{(d-\gamma)/d}$$

J. Yukich, "Probability theory of classical Euclidean optimization problems," Springer Lecture Notes in Mathematics, 1998.

Main ideas behind proof of BHH (Yukich 1998)

Start with $f(x)$ uniform over $[0,1]^d$

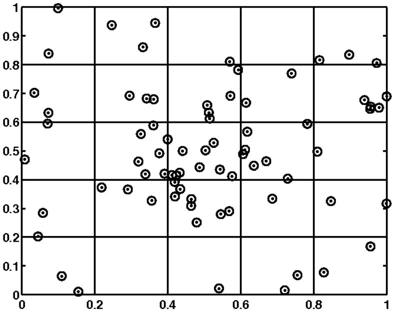
- Avg distance between n points in $[0,1]^d$

$$|e_i|_{avg} = n^{-1/d}$$

- Avg length of MST should therefore be

$$L_\gamma^{MST} = \sum_{i=1}^{n-1} |e_i|_{avg}^\gamma \approx c\ nn^{-\gamma/d} = cn^{(d-\gamma)/d}$$

- The constant c in front is $\beta_{d,\gamma}$



Next apply partitioning heuristic

- Dissect $[0,1]^d$ into m^d cubes $\{Q_i\}$ each with center q_i.
- From translation invariance, homogeneity, quasi-additivity of MST

$$L_\gamma^{MST}(\mathcal{X}_n) \approx m^{-\gamma} \sum_{i=1}^{m^d} L_\gamma^{MST}(m(\mathcal{X}_n \cap Q_i))$$

- From the $[0,1]^d$ result

$$L_\gamma^{MST}(m(\mathcal{X}_n \cap Q_i)) = c(n_i)^{(d-\gamma)/d}$$

- From smoothness of f

$$n_i/n \approx m^{-d} f(q_i)$$

- Therefore

$$L_\gamma^{MST}(m(\mathcal{X}_n \cap Q_i)) \approx cn^{(d-\gamma)/d}(m^{-d}f)^{(d-\gamma)/d}$$

- since

$$(m^{-d}f)^{(d-\gamma)/d} = m^\gamma m^{-1/d} f^{(d-\gamma)/d}(q_i)$$

$$L_\gamma^{MST}(\mathcal{X}_n) \approx n^{(d-\gamma)/d} \cdot c \sum_{i=1}^{m^d} f^{(d-\gamma)/d}(q_i) m^{-1/d}$$

BHH theorem Riemannian extension

Theorem (Costa 2004, 2005)

Let $(\mathcal{S}, g)$ be a compact smooth Riemannian d-dimensional manifold in $\mathbb{R}^D$. Suppose $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ is a random sample on $\mathcal{S}$ with density f relative to μ_g and $d \geq 2$, $1 \leq \gamma < d$. Then

$$\lim_{n\to\infty} \frac{L_\gamma^{MST}(\mathcal{X}_n)}{n^\alpha} = \beta_{d,\gamma} \int_{\mathcal{S}} f^\alpha(x) d\mu_g$$

where $\alpha = (d - \gamma)/d$.

Alternative representation For finite n

$$\log L_\gamma^{MST}(\mathcal{X}_n) = \alpha \log n + (1-\alpha) H_\alpha(X) + \log \beta_{d,L} + \varepsilon(n)$$

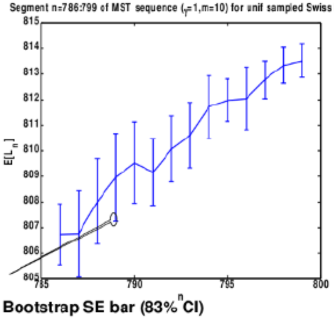
where

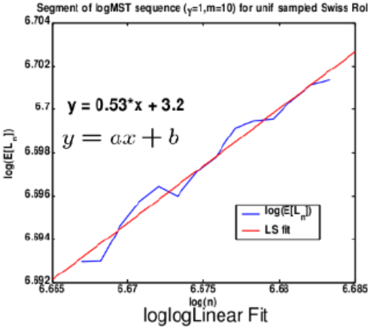
$$H_\alpha(X) = (1-\alpha)^{-1} \ln \int_{\mathcal{S}} f^\alpha(x) d\mu_g$$

is α-entropy of X and $\varepsilon(n) \to 0$ w.p.1.

Key observation: can use representation of $\log L_\gamma^{MST}$ to estimate intrinsic dimension d of $\mathcal{S}$ in addition to entropy of $f(x)$.

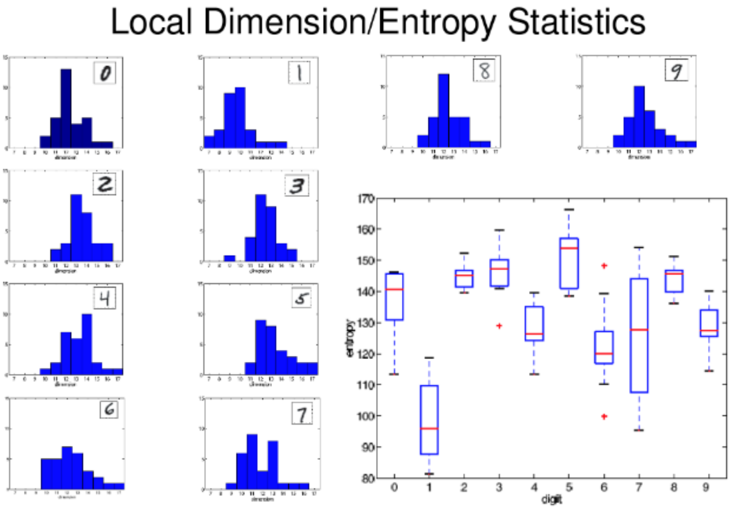
Dimension and entropy estimation for unif density on swiss roll





- $\hat{d} = \text{round}\underbrace{\left(\frac{\gamma}{1-a}\right)}_{2.1} = 2$
- $\hat{H}_\alpha(X) = \frac{b-\gamma/2\log\beta_{d,\gamma}}{1-a} = 7.3$
- Ground truth: $H_\alpha(X) = \log(1869) = 7.53$

Dimension estimation: MNIST digits



J. Costa and A. Hero, "Learning intrinsic dimension and entropy of shapes," in Statistics and analysis of shape, Eds. H. Krim and T. Yezzi, Birkhauser, 2005

Continuum limit of greedy kMST length functional

Ravi (1996) proposed a greedy partitioning approximation to kMST.

Theorem (Hero and Michel 1999)

Fix $\rho \in [0,1]$. If $k/n \to \rho$ then the length of Ravi's greedy partitioning k-MST satisfies

$$L_\gamma^{kMST}(\mathcal{X}_n)/(\rho n)^\alpha \to \beta_{\gamma,d} \inf_{A:Pr(A)\geq\rho} \int f^\alpha(x|x\in A)dx \qquad (a.s.)$$

$Pr(A) = \int_A f$.

Alternatively, defining the conditional entropy function

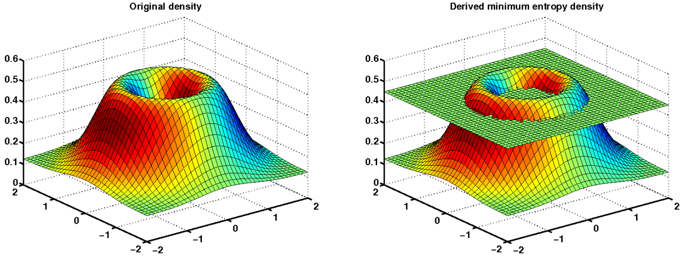
$$H_\alpha(f|x\in A) = \frac{1}{1-\alpha} \ln \int f^\alpha(x|x\in A)dx,$$

$$\frac{1}{1-\alpha} \ln \left(L_\gamma^{kMST}(\mathcal{X}_n)/(\rho n)^\alpha\right) \to \beta_{\gamma,d} \inf_{A:Pr(A)\geq\rho} H_\alpha(f|x\in A) + c \qquad (a.s.)$$

Solution to variational problem is a level set $A = A_o$ of f.

- A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal K-point random graphs," IEEE Information Theory 1999.

Continuum limit of kMST length functional



Here level set A_0 satisfies $P(X \in A_0) = \rho$.

Level set can be estimated empirically from data $\mathcal{X}_n$ by

- Empirical kernel estimation of f by $\hat{f}(x) = G(x) * \sum_{i=1}^{n} \delta(X_i)$
- Solve for level-set of $\hat{f}$ by variational pde

• S. Osher and R. Fedkiw, "Level set methods: an overview and some recent results," Journal of Computational physics, 2001

• J. Sethian, "Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science," Vol. 3. Cambridge university press, 1999

Continuum limit of FR length functional

Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ and $\mathcal{Y} = \{Y_1, \ldots, Y_m\}$ be independent sets of i.i.d. random vectors in $\mathbb{R}^d$ with marginal pdfs f_x and f_y, respectively.

Theorem (Henze and Penrose, 1999)

Let n, m converge to infinity in such a way that $n/(n+m) = \epsilon$, $\epsilon \in [0,1]$. Then the FR length functional satisfies

$$L_1^{FR}(\mathcal{X} \cup \mathcal{Y})/(n+m) \to \int \frac{f_x(x) f_y(x)}{\epsilon f_x(x) + (1-\epsilon) f_y(x)} dx \qquad (a.s.)$$

Alternatively, define the f-divergence

$$D_\epsilon(p,q) = (4\epsilon(1-\epsilon))^{-1} \left(\int \frac{(\epsilon p(x) - (1-\epsilon) q(x))^2}{\epsilon p(x) + (1-\epsilon) q(x)} dx - (2\epsilon - 1)^2 \right)$$

then (Berisha and Hero 2015)

$$1 - L_1^{FR}(\mathcal{X} \cup \mathcal{Y}) \frac{n+m}{2nm} \to D_\epsilon(f_x, f_y) \qquad (a.s.)$$

- N. Henze and M. Penrose, "On the multivariate runs test," Ann. of Statistics, 1999.
- V. Berisha and A. Hero, "Empirical non-parametric estimation of the Fisher Information," IEEE Signal Processing Letters, 2015.

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*

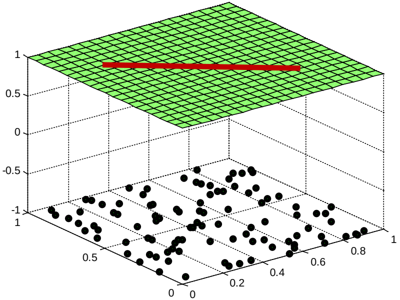
$$L_\gamma^{SP}(\mathcal{X})/n^{(1-\gamma)/d} \to C_{d,\gamma} \underbrace{\inf_\pi \int_0^1 f(\pi_t)^{(1-\gamma)/d} \sqrt{g(\dot{\pi}_t, \dot{\pi}_t)} dt}_{dist_\gamma(x_I, x_F)} \qquad (a.s.)$$

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.

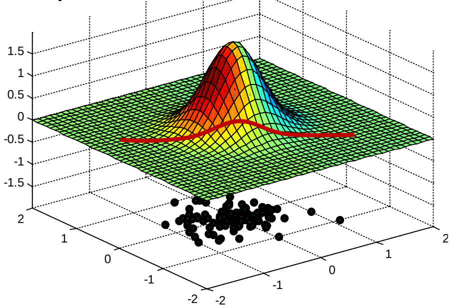
• S.-J. Hwang, S. Damelin, A. Hero, "Shortest path through random points," Annals of Applied Probability, 2016 (arXiv:1202.0045).

Continuum limit of shortest path: archimedean vs relativistic limit

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



Archimedean shortest path



Relativistic shortest path

Main ideas behind proof of SP (Hwang, Damelin, H 2016)

Start with $\{\mathbf{X}_i\}_{i=1}^n \sim f(x) = U([0,1]^d)$

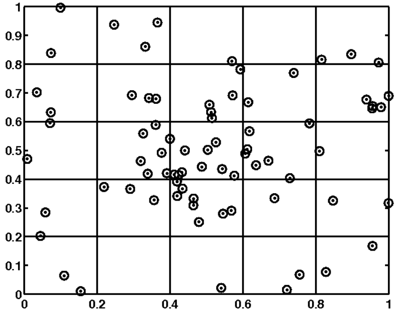
- Avg. interpoint distance is

$$|e_i|_{avg} = n^{-1/d}$$

- Avg # points in a short path π: $cn^{1/d}$
- Avg length of π should therefore be

$$L_\gamma^\pi = n^{1/d}\ |e_i|_{avg}^\gamma \approx c\ n^{1/d} n^{-\gamma/d} = cn^{(1-\gamma)/d}$$

- Contant is $c = c_{\dot{\pi}} = C_{d,\gamma} \int_0^1 \|\dot{\pi}\|$



Next apply partitioning heuristic

- Dissect $[0,1]^d$ into m^d cubes $\{Q_i\}$ each with center q_i.
- Let π be any short path crossing through $O(m)$ cubes. Then, length of path satisfies

$$L_\gamma^\pi(\mathcal{X}_n) \approx m^{-\gamma} \sum_{i=1}^m L_\gamma^\pi(m(\mathcal{X}_n \cap Q_i))$$

- From the $[0,1]^d$ result, with $\pi_i = \pi \cap Q_i$

$$L_\gamma^\pi(m(\mathcal{X}_n \cap Q_i)) = c_{\dot{\pi}_i}\ \|(n_i)^{(1-\gamma)/d}$$

- From smoothness of f

$$n_i/n \approx m^{-d} f(q_i)$$

- Therefore

$$L_\gamma^\pi(m(\mathcal{X}_n \cap Q_i)) \approx c_{\dot{\pi}}\ n^{(1-\gamma)/d}(m^{-d}f)^{(1-\gamma)/d}$$

- since $(m^{-d}f)^{(1-\gamma)/d} = m^\gamma m^{-1} f^{(1-\gamma)/d}$

$$L_\gamma^\pi(\mathcal{X}_n) \approx n^{(1-\gamma)/d} \cdot \sum_{i=1}^m c_{\dot{\pi}}\ f^{(1-\gamma)/d}(q_i) m^{-1}$$

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 $g(\dot{\pi},\dot{\pi}) = \langle\dot{\pi},\dot{\pi}\rangle$

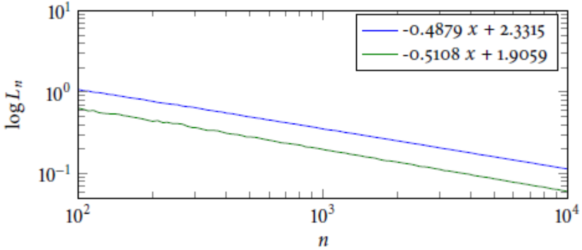
$$\frac{1-\gamma}{d}\mathbf{A}(\dot{\pi})\nabla_{\pi}\ln f(\pi) + \frac{d}{dt}\left(\frac{\dot{\pi}}{\|\dot{\pi}\|}\right) = 0$$

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



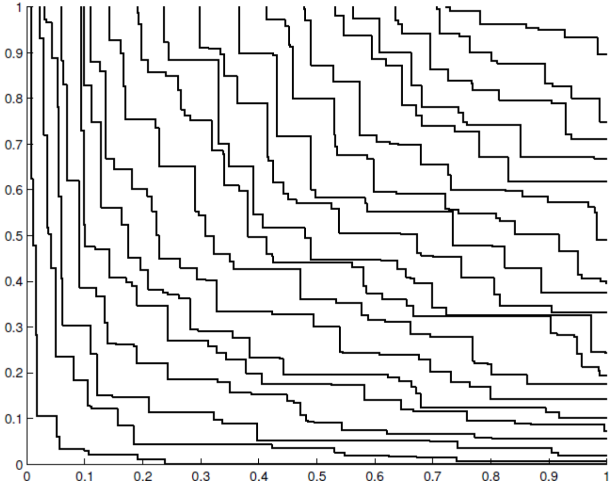
Regression equation ($\alpha = (1-\gamma)/d$):

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

Experimental setting

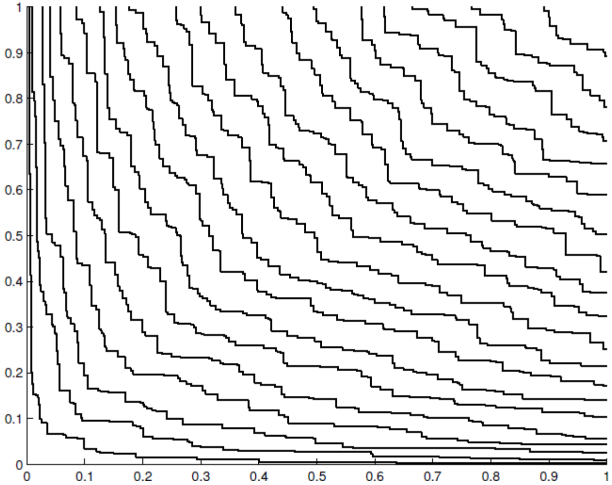
- $d = 2$, $\gamma = 2$ so that slope should be $(1-\gamma)/d = -0.5$
- $\mathcal{X}_n$ are n uniform points on $\mathcal{S} = S^2$
- Blue plot: $x = (1,0,0)$, $y = (-1,0,0)$
- Red plot: $x = (0,1,0)$, $y = (0,0,1)$

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



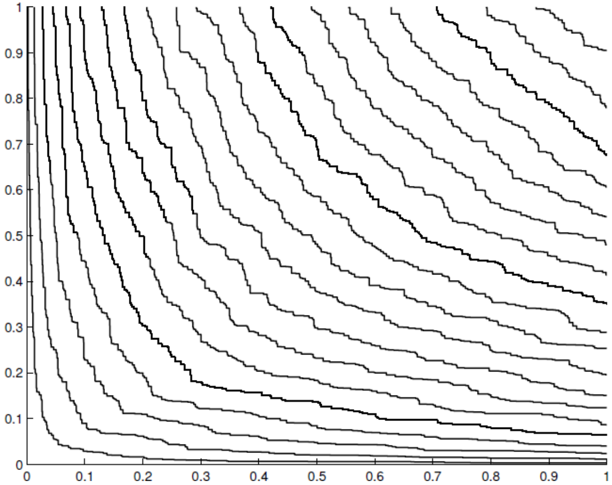
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



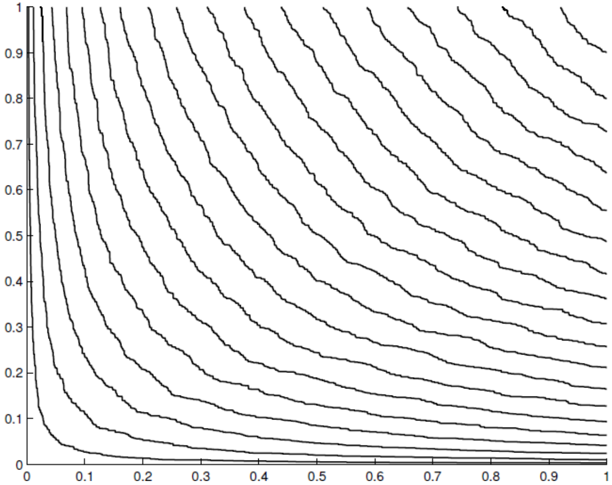
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



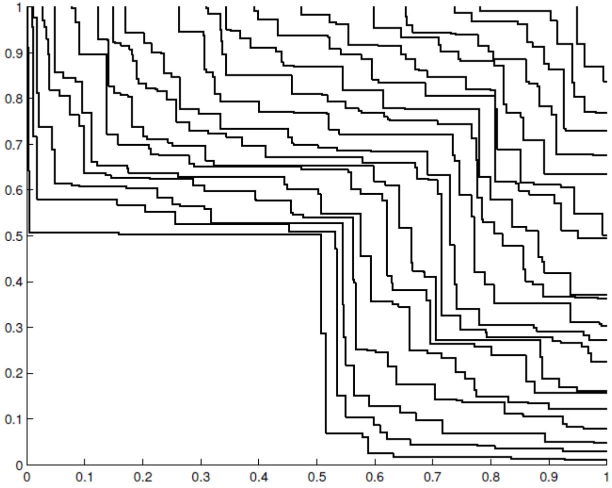
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



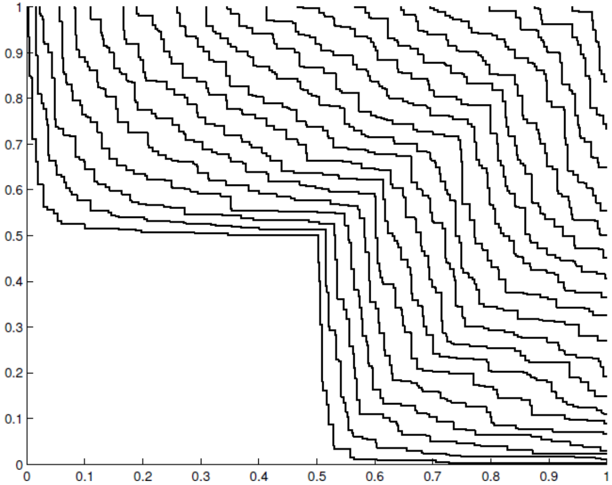
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



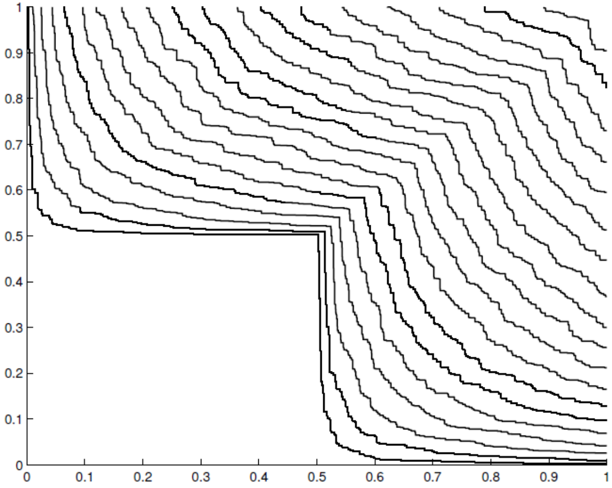
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



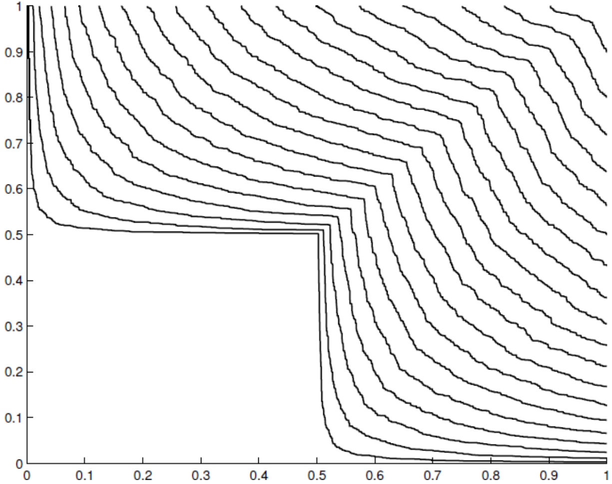
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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.

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 $\mathrm{supp}(f) \subset \Omega \subset \mathbb{R}^d$, Ω bounded with Lipshitz $\partial\Omega$.

Theorem (Calder, Esedoglu and H, 2014)

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

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

where

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

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

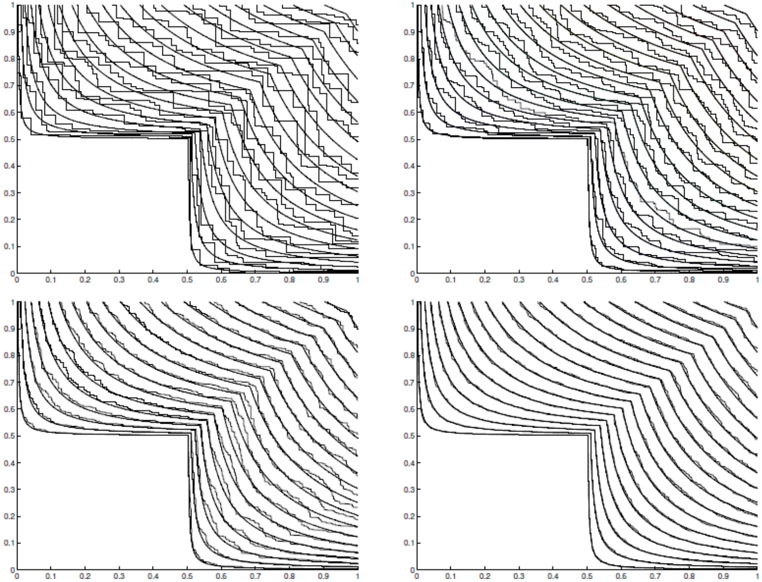
where $\mathcal{A} = \left\{ \gamma \in C^1(0,1;\mathbb{R}^d) : \gamma'(t) \geqq 0 \ \forall t \in [0,1] \right\}$

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 $\mathrm{Unif}[0,1]/[0,0.5]^2$



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

Relation of Pareto fronts to longest chain problem

Proof of theorem relies on connection to longest chain problem (Ulam [1961]),(Hammersley et al. [1972]), (Aldous and Diaconis [1995])

- $u_n(\mathbf{x})$ is the length of longest chain in $\{\mathbf{X}_i \in \mathcal{X} : \mathbf{X}_i \leqq \mathbf{x}\}$.
- $\mathcal{F}_k$ is anti-chain containing $\{\mathbf{X}_i \in \mathcal{X} : u_n(\mathbf{X}_i) = k\}$
- $u_n = u_{\{X_1,\ldots,X_n\}}$ is a superadditive functional in the sense that

$$u_{\{X_1,\ldots,X_n\}}(\mathbf{x}) \geq \sum_{i=1}^{m} u_{\{X_1,\ldots,X_n \cap R_i\}}(\mathbf{x})$$

- Superadditivity implies convergence of $n^{-1/d} u_n$
- Smoothness of f implies convergent limit obeys Hamiltonian-Jacobi p.d.e.

Low complexity (linear) numerical p.d.e. solver proposed (Calder et al. [2015])

$$\prod_{i=1}^{d} [U(\mathbf{x}) - U(\mathbf{x} - h\mathbf{e}_i)] = h^d d^{-d} f(\mathbf{x}), \;\; \mathbf{x} \in \{h, 2h, \ldots, Mh\}^d$$

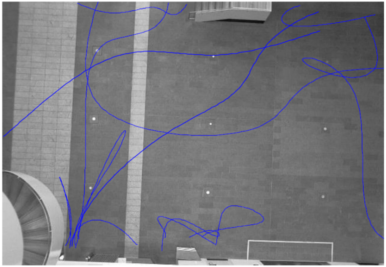
Calder, Esedoglu and H, "A Hamilton-Jacobi equation for the continuum limit of non-dominated sorting", SIAM Mathematical Analysis, Feb 2014

Calder, Esedoglu, H, "A PDE-based approach to non-dominated sorting," SIAM Numerical Analysis, 2015

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.

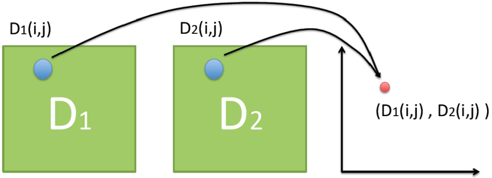
Multicriteria anomaly detection

Speed and shape similarity between trajectories $T_i(t), T_j(t) \in \mathbb{R}^2$:

$$D_1(i,j) = \|\mathrm{hist}(\Delta T_i) - \mathrm{hist}(\Delta T_j)\|,$$

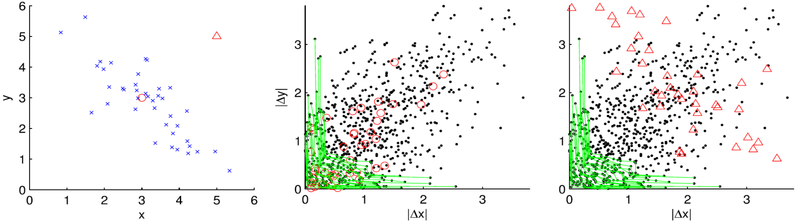
$$D_2(i,j) = \|T_i - T_j\|$$

1. Scalarization:
   $D_\lambda(i,j) = \lambda D_1(i,j) + (1-\lambda) D_2(i,j)$
2. Pareto depth analysis:
   $(D_1(i,j), D_2(i,j)) \rightarrow$ one *dyad*



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

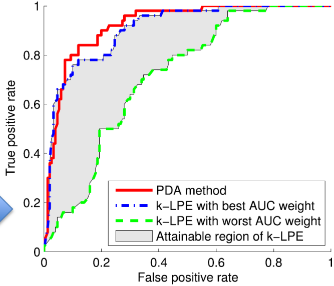
Detection performance of multicriteria anomaly detection



PDA Algorithm:

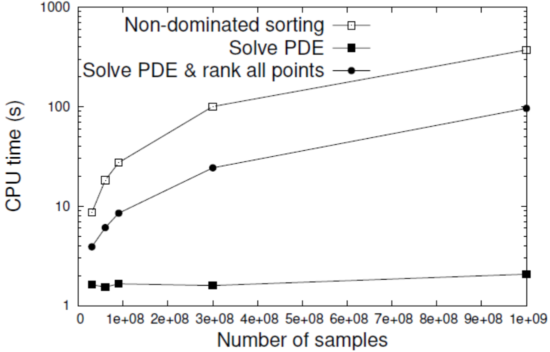
- Embed N choose 2 dyads onto plane
- Build Pareto fronts of non-dominated dyads.
- Compute anomaly scores = depth of front.

PDA outperforms scalarization



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

Run-time comparisons



- Performed on $50,000$ trajectories (a total of 10^9 Pareto points)
- Grid size used 250×250

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 data science
    - They lie at the interface between statistical physics, machine learning, combinatorial optimization, probability, and applied math
    - Scalable pde-based algorithms for solving minimal path and non-dominated sorting problems
    - Graph-based methods for estimating information measures (entropy, divergence, mutual information)

Summary

- Continuum limit analysis can lead to useful tools and insights for data science
    - They lie at the interface between statistical physics, machine learning, combinatorial optimization, probability, and applied math
    - Scalable pde-based algorithms for solving minimal path and non-dominated sorting problems
    - Graph-based methods for estimating information measures (entropy, divergence, mutual information)
- Some related open problems
    - 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 data science
  - They lie at the interface between statistical physics, machine learning, combinatorial optimization, probability, and applied math
  - Scalable pde-based algorithms for solving minimal path and non-dominated sorting problems
  - Graph-based methods for estimating information measures (entropy, divergence, mutual information)
- Some related open problems
  - Minimal paths on sparse graphs, directed paths, multigraphs, hypergraphs
  - Non-dominated sorting extensions to data depth and convex hull peeling
- Broader questions
  - New frontier: statistical mechanics of big data and data analysis?
  - New primitive: state-of-the-art numerical pde solvers in pipeline?

David Aldous and Persi Diaconis. Hammersley's interacting particle process and longest increasing subsequences. *Probability theory and related fields*, 103(2):199–213, 1995.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, volume 14, pages 585–591, 2001.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6): 1373–1396, 2003.

Visar Berisha and A Hero. Empirical non-parametric estimation of the fisher information. *IEEE Signal Processing Letters*, 22(7), 2014.

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.

Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

J. Costa and A. O. Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. on Signal Process.*, SP-52(8):2210–2221, August 2004.

J. Costa and A. O. Hero. Learning intrinsic dimension and entropy of shapes. In H. Krim and T. Yezzi, editors, *Statistics and analysis of shapes*. Birkhauser, 2005.

Jerome H. Friedman and Lawrence C. Rafsky. Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Annals of Statistics*, 7(4):697–717, 1979.

Laurent Galluccio, Olivier Michel, Pierre Comon, Mark Kliger, and Alfred O Hero. Clustering with a new distance measure based on a dual-rooted tree. *Information Sciences*, 251:96–113, 2013.

John Michael Hammersley et al. A few seedlings of research. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*. The Regents of the University of California, 1972.

N. Henze and M. Penrose. On the multivariate runs test. *Annals of Statistics*, 27:290–298, 1999.

A.O. Hero and O. Michel. Asymptotic theory of greedy approximations to minimal k-point random graphs. *IEEE Trans. on Inform. Theory*, IT-45(6):1921–1939, Sept. 1999.

Alfred O Hero III. Geometric entropy minimization (gem) for anomaly detection and localization. *Ann Arbor*, 1001:48109–2122, 2006.

Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. *Surface reconstruction from unorganized points*, volume 26. SIGGRAPH, ACM, 1992.

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.

H. Neemuchwala, A. O. Hero, and Paul Carson. Entropic graphs for registration. In *Image fusion*, pages 185–235. Marcel Dekker, 2005.

Dimitris Papadias, Yufei Tao, Greg Fu, and Bernhard Seeger. An optimal and progressive algorithm for skyline queries. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 467–478. ACM, 2003.

M. Penrose. *Random geometric graphs*. Oxford University Press, 2003.

R. Ravi, M.V. Marathe, D.J. Rosenkrantz, and S.S. Ravi. Spanning trees short or small. In *Proc. 5th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 546–555, Arlington, VA, 1994.

K. Sricharan, R. Raich, and A.O. Hero III. Efficient anomaly detection using bipartite k-nn graphs. In *Proc of Conf on Neural Information Processing Systems (NIPS)*, Granada, Dec. 2011.

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

Daniel Ting, Ling Huang, and Michael Jordan. An analysis of the convergence of graph laplacians. *arXiv preprint arXiv:1101.5435*, 2011.

Stanislaw M Ulam. Monte carlo calculations in problems of mathematical physics. *Modern Mathematics for the Engineers*, pages 261–281, 1961.

P. Viola and W. M. Wells. Alignment by maximization of mutual information. In *Proceedings of IEEE International Conference on Computer Vision*, pages 16–23, Los Alamitos, CA, Jun. 1995.

Ying Xu and Edward C Uberbacher. 2d image segmentation using minimum spanning trees. *Image and Vision Computing*, 15(1):47–57, 1997.

J. E. Yukich. *Probability theory of classical Euclidean optimization*, volume 1675 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1998.