Combinatorial continuum limits and their applications

Alfred Hero

†University of Michigan

June 1, 2015

Conference on Scale Space and Variational Methods (SSVM)
Lege, France

Acknowledgements

Former students contributing to work

- Kumar Sricharan (Parc Research)
- Jeff Calder (UC Berkeley)
- Ko-Jen Hsiao (Whispers)
- Kevin Xu (University of Toledo)
- John Gorman (DARPA)
- Bing Ma (University of Nevada)
- Huzefa Neemuchwala (Abbott, Inc)
- Jose Costa (Teza Technologies)

Collaborators contributing to work

- Selim Esedoglu (University of Michigan)
- Olivier Michel (INPG Grenoble)
- Alejandro Frangi (U. Sheffield)

Sponsors of work presented here

- ARO Database Indexing and Retrieval Program
- ARO MURI Value of Information Program
- NSF: Theoretical Foundations Program
- NIH Biomedical Imaging Institute

Outline

Geometric graphs in imaging and computer vision

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

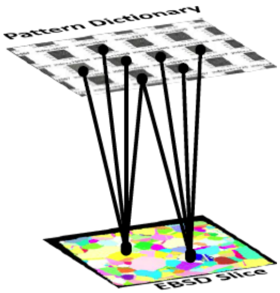
Some applications where geometric graphs arise

- Computer vision, video and image processing
  - Clustering and segmentation (GLap, kNNG, MST, graph cuts)
  - Dimensionality reduction (GLap, kNNG, GMST)
  - Denoising and anomaly detection (kMST, BP-kNNG)
  - Orthoregistration (MST, kNNG)
  - Frame-to-frame registration (TSP)
  - Multi-resolution image representation (MST-based pyramid)
  - Image inpainting interpolation (kNNG)
- Image/video indexing and retrieval
  - Query-reference matching (NNG)
  - Database partitioning (kNNG)
  - Multi-criterion image retrieval (Non-dominated sorting)

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.

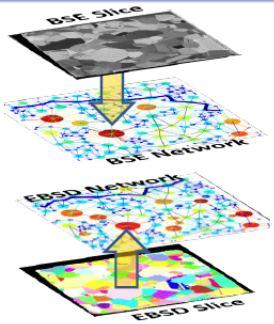
M. Penrose, Random geometric graphs, Oxford University Press 2003

Example: dictionary learning of grain networks in materials science



Bipartite indexing of polycrystaline materials (Park *et al* 2015)

- Nodes: Bottom - spatial locations on slice. Top - patterns in dictionary
- Features: Possible Kikuchi diffraction patterns of crystal planes
- Edges: The top 4 pattern matches between slice and dictionary



Grain-level fusion of polycrystaline materials (Chen *et al* 2015)

- Nodes: spatial locations on slice
- Features: spatial patch intensities (BSE) or Kikuchi patterns (EBSD)
- Edges: feature correlations that exceed a high level

- S.U. Park *et al*, "A dictionary approach to the EBSD indexing problem," Microscopy and Microanalysis, June 2015.
- Y.-H. Chen *et al*,"Parameter estimation in spherical symmetry groups," IEEE Signal Processing Letters, Jan. 2015.
- Y.-H. Chen *et al*, "Coercive region-level registration for multi-modal images," to appear in IEEE Conf. on Image Processing, 2015.

Continuum limits of random geometric graphs

Let $L(\mathcal{X}_n)$ be a function of this graph, e.g., the sum of the edge weights.

Minimizing $L(\mathcal{X}_n)$ over the edge set is a discrete optimization problem and often combinatorial complexity.

Sometimes there is an $\alpha > 0$ such that the continuum limit $\lim_{n\to\infty} \min L(\mathcal{X}_n)/n^\alpha$ exists or has a known probability distribution.

Some benefits of continuum limits

- Provides intuition about asymptotic sensitivity of graph topology to f, e.g., k-point MST
- Reveals new graph-based statistical estimators of mean of continuum limit, e.g. entropy estimation
- Permits setting significance level of hypothesis tests on $P(\mathcal{X}_n)$, e.g., Friedman-Rafsky multivariate runs test
- Leads to scalable continuous relaxations of otherwise combinatorial graph construction problem, e.g., pde solvers for non-dominated sorting problems

Outline

Minimal geometric graphs

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

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

- $\{\mathcal{V}\} = \{X_1, \ldots, X_n\}$: vertices or nodes
- $\{\mathcal{E}\} = \{e_{ij}\}$: 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

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

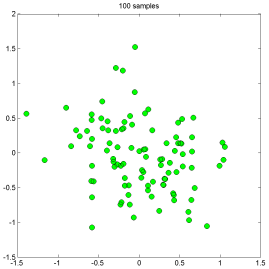
where $\gamma \geq 0$.

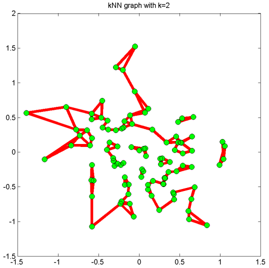
k-nearest neighbor (kNN) graph

- kNN graph is solution of the optimization

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

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





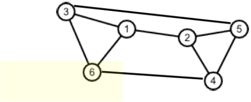
kNNGs in spectral clustering and dimensionality reduction

kNNG clustering for image segementation (Felzenszwalb 2003)

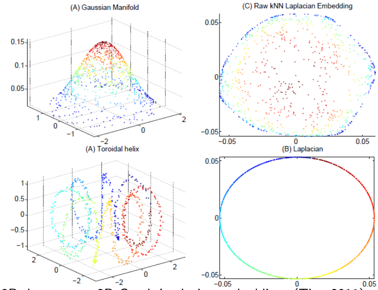
k-NNG-based spectral algorithm

- 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$(\mathbf{W} - \mathbf{D})$, $\mathbf{D} = \mathrm{diag}(\mathbf{W}\underline{1})$
  - 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



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

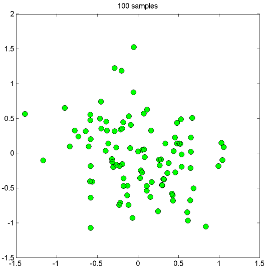
- Belkin, Mikhail, and Partha Niyogi. "Laplacian eigenmaps and spectral techniques for embedding and clustering." NIPS. Vol. 14. 2001.
- Coifman, Ronald R., and Stphane Lafon. "Diffusion maps." Applied and computational harmonic analysis 21.1 (2006): 5-30.

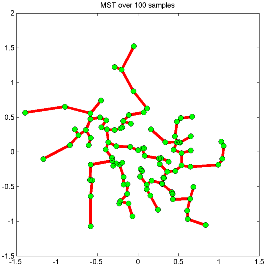
Minimal spanning tree (MST)

- MST is solution of the optimization

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

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



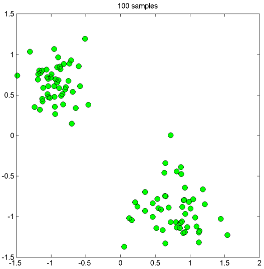


Minimal spanning tree (MST)

- 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
- Vision applications: image segmentation, image registration, clustering
- Computational complexity is $O(n^2 \log n)$



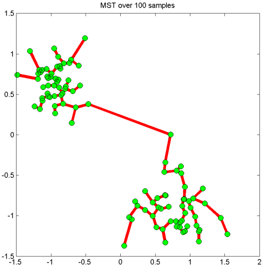
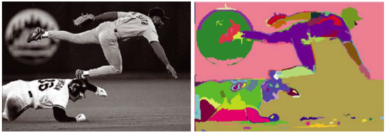
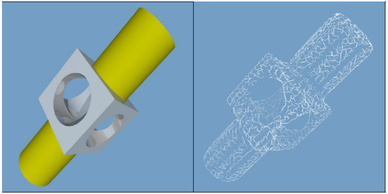


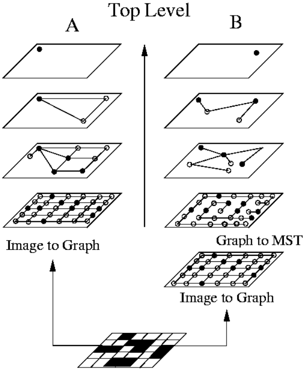
Illustration: MST for image segmentation, representation and rendering



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

Friedman-Rafsky graph (FR)

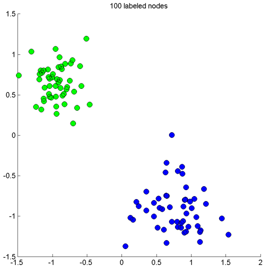
- Two labeled samples $\mathcal{X}_n$, $\mathcal{Y}_m$
- Start with MST over $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

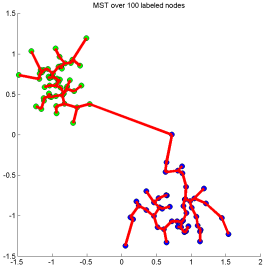
$$L_\gamma^{MST}(\mathcal{V}) = \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E})$$
$$= \sum_{e_{ij}\in\mathcal{E}*} |e_{ij}^{XX}|^\gamma + |e_{ij}^{XY}|^\gamma + |e_{ij}^{YY}|^\gamma$$

- 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)

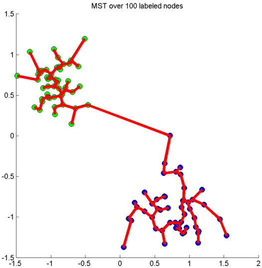
- Two labeled samples $\mathcal{X}_n$, $\mathcal{Y}_m$
- Start with MST over $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

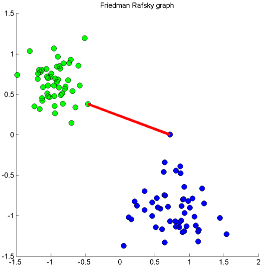
$$
\begin{aligned}
L_\gamma^{MST}(\mathcal{V}) &= \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E}) \\
&= \sum_{e_{ij}\in\mathcal{E}*} |e_{ij}^{XX}|^\gamma + |e_{ij}^{XY}|^\gamma + |e_{ij}^{YY}|^\gamma
\end{aligned}
$$

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

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

- 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)

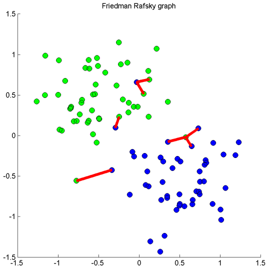
- Two labeled samples $\mathcal{X}_n$, $\mathcal{Y}_m$
- Start with MST over $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

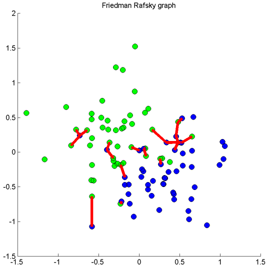
$$L_\gamma^{MST}(\mathcal{V}) = \min_{\mathcal{E}:\mathbf{A}\underline{1}>0} L_\gamma(\mathcal{V},\mathcal{E})$$
$$= \sum_{e_{ij}\in\mathcal{E}*} |e_{ij}^{XX}|^\gamma + |e_{ij}^{XY}|^\gamma + |e_{ij}^{YY}|^\gamma$$

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

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

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

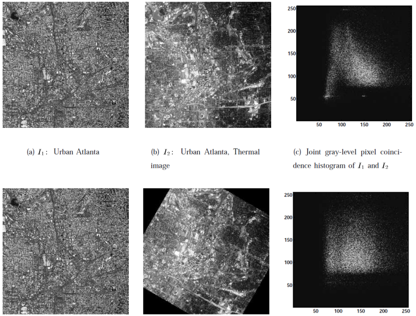
Maximize mutual information (MI)

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

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 $\mathbf{z}_i \in \mathbb{R}^2$:

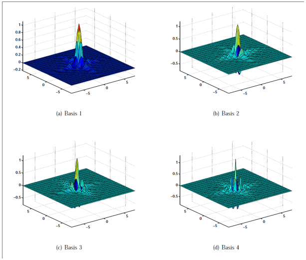
$$\mathbf{X}_1(i) = [\mathbf{W}(\mathbf{z}_i), \mathbf{z}_i], \ \mathbf{X}_2(i) = [\mathbf{W}(\mathbf{z}_i), \mathbf{z}_i]$$

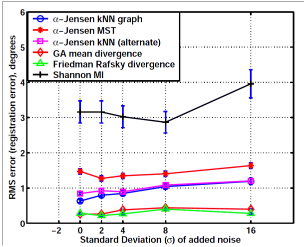
$\mathbf{W}_1(\mathbf{z}_i)$ and $\mathbf{W}_2(\mathbf{z}_i)$ are localized Meyer wavelet coefficients of I_1 and $T(I_2)$

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





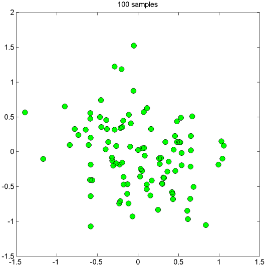
- H. Neemwuchwala and A. Hero, "Entropic Graphs for Registration," in Multi-Sensor Image Fusion and its Applications, Eds. R. S. Blum and Z. Liu, Marcel Dekker, Inc., 2005.

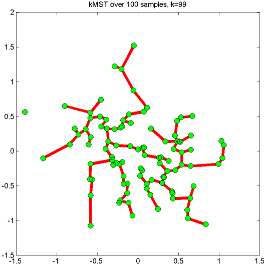
k-Minimal spanning tree (kMST)

- 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 $\mathcal{V}$
- Vision applications: Denoising and outlier detection, robust image registration, robust clustering
- Computational complexity is NP hard
- Greedy approximations are available (Ravi 1994)

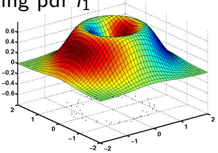




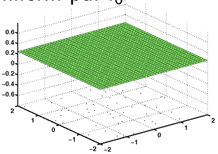
R. Ravi, M. Marathe, D. Rosenkrantz and S. Ravi "Spanning trees short and small," Proc of ACM-SIAM Symp on Discrete Algorithms, 1994.

Denoising illustration of kMST

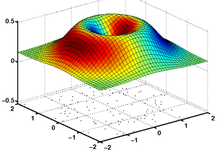
Ring pdf f_1

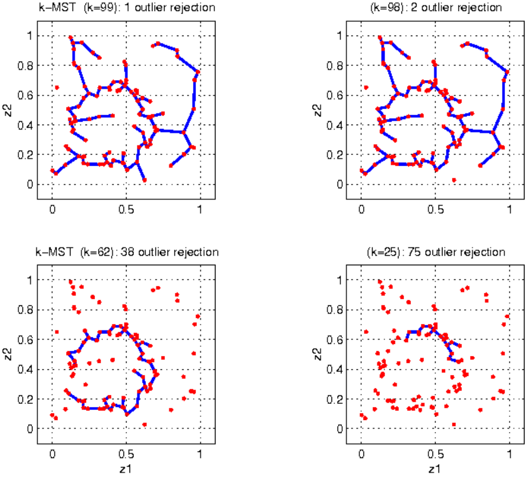


Uniform pdf f_0



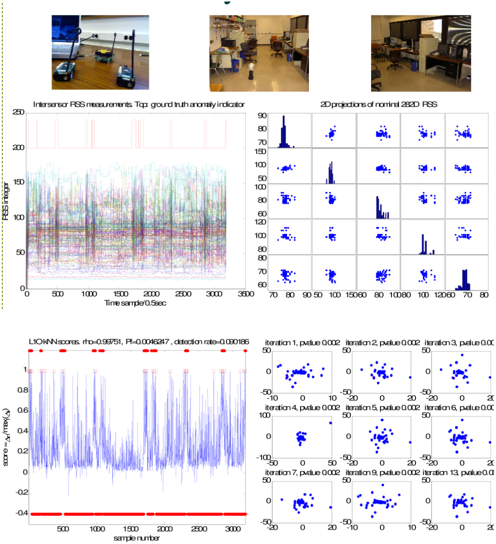
$f = (1 - \epsilon) f_1 + \epsilon f_0$





- 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)

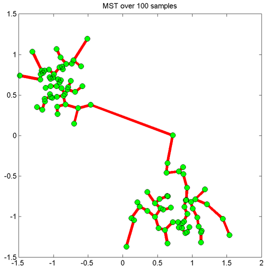
- Let $\mathcal{G}$ be a graph with $m = |\mathcal{E}|$ edges on n vertices $\mathcal{V}$
- $\pi(X_I, X_F)$ a path over $\mathcal{G}$ btwn points X_I and X_F

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

  $X_{i_{j+1}}$ 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 $\mathcal{G}$:
  - kNN graph
  - MST
- Vision applications: dimensionality reduction, manifold learning, image retrieval
- Computational complexity is $O(m + n \log n)$



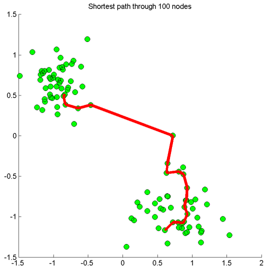


Illustration: kNN and shortest paths ISOMAP dimensionality reduction

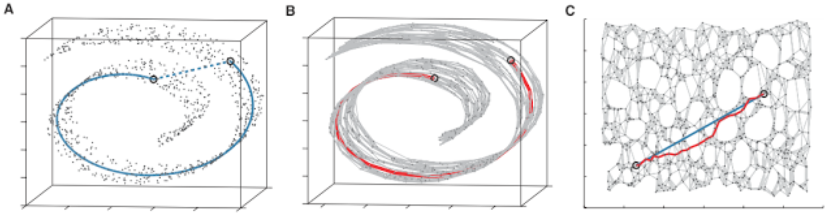
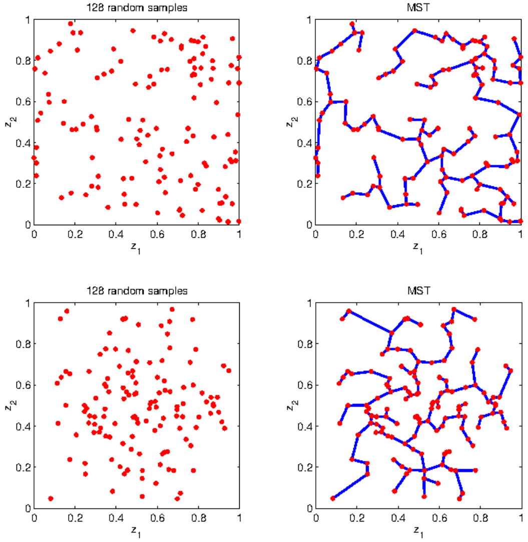


Fig. 3. The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (**A**) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect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

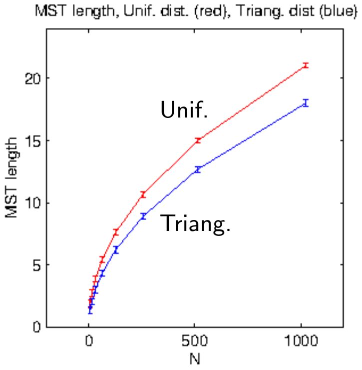
Outline

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

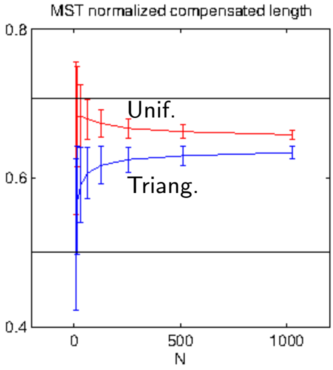


Large n behavior of MST length functional

length(MST)



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



Continuum limit of kNN and MST length functionals

Theorem (Beardwood, Halton&Hammersley 1959)

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

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

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

- RMS rate of convergence (Costa & Hero 2003)

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

J. Beardwood and J. H. Halton and J. M. Hammersley, "The shortest path through many points," Proc. Cambridge Philosophical Society 1959. (BHH proved the limit for the TSP, $f(x)$ uniform, and $\gamma = 1$.)

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

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

  - 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 $[0,1]^d$

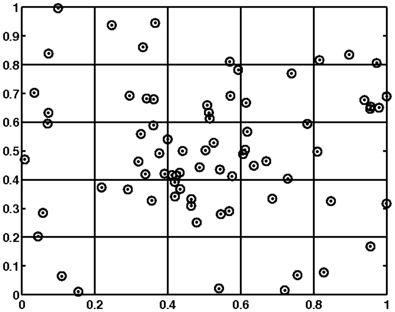
$$|L_\gamma(F \cup G) - L_\gamma(F)| \le C_3\, (\mathrm{card}(G))^{(d-\gamma)/d}$$

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

Comments on BHH continuum limit

Yukich's proof technique applies to TSP, kNNG, Steiner tree, etc.

Each of these quasi-additive continuous length functionals will have its own characteristic constant c.

Extensions beyond BHH

- A CLT for kNN and MST graphs has also been established (Avram and Bertsimas 1993)
- BHH established when $\mathcal{S}$ is smooth Riemannian manifold in $\mathbb{R}^d$ (Costa 2006)

The BHH motivated approximate combinatorial optimization algorithms

- A polynomial nearly optimal approximation to the TSP (Borovkov 1962)
- A partitioning approximations to the the TSP (Karp 1976)
- Partitioning approximations to the k-point MST (Ravi 1994)

The BHH motivated applications of combinatorial optimization algorithms

- MST and kNN graphs for entropy estimation (Hero 1999)
- MST and kNN graphs for pattern matching and image registration (Neemuchwala 2004)
- GMST and kNN graphs for intrinsic dimension estimation (Costa 2003)

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

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

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

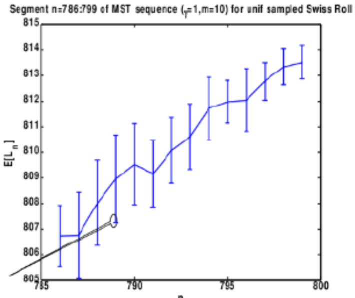
where

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

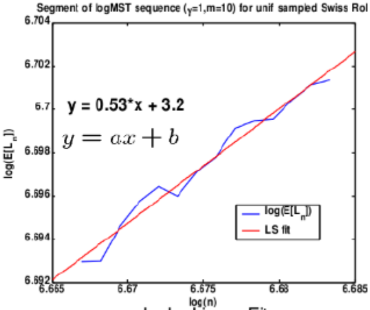
is α-entropy of X and $\varepsilon(n) \to 0$ w.p.1.

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



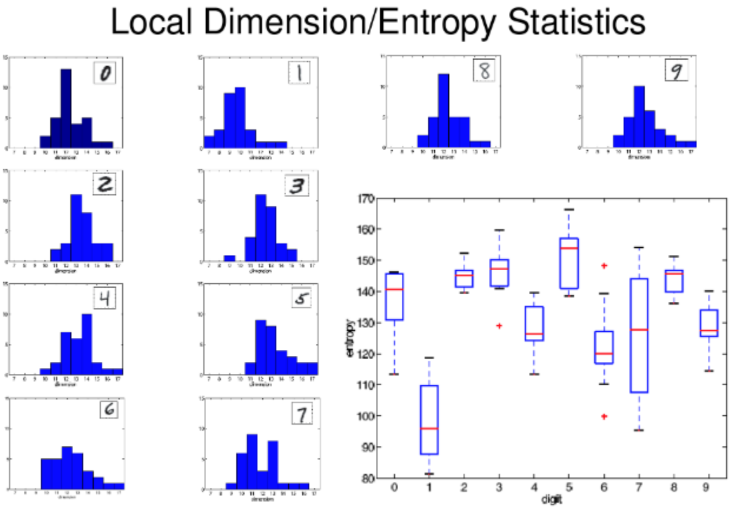
Bootstrap SE bar (83% CI)



loglogLinear Fit

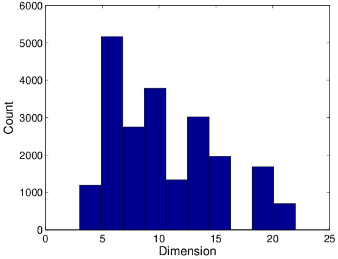
- $\hat{d} = \mathrm{round}\underbrace{\left(\frac{\gamma}{1-a}\right)}_{2.1} = 2$
- $\hat{H}_\alpha(X) = \frac{b-\gamma/2\log\beta_{d,\gamma}}{1-a} = 7.3$
- Ground truth: $H_\alpha(X) = \log(1869) = 7.53$

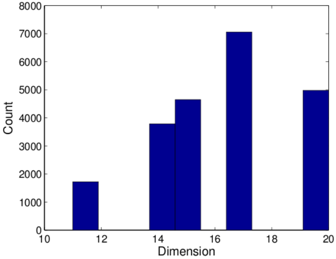
Dimension estimation: MNIST digits



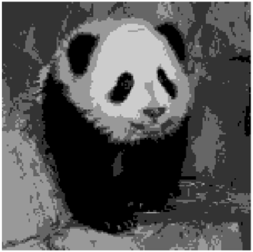
J. Costa and A. Hero, "Learning intrinsic dimension and entropy of shapes," in Statistics and analysis of shape, Eds. H. Krim and T. Yezzi, Birkhauser, 2005

Dimension estimation: dimension-driven image segmentation





Dimension estimation: dimension-driven image segmentation



Carter " Dimensionality Reduction on Statistical Manifolds," PhD Thesis, Univ. of Michigan 2008

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

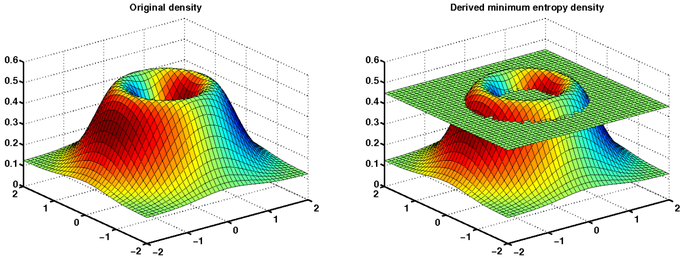
$$H_\alpha(f|x\in A) = \frac{1}{1-\alpha}\ln\int f^\alpha(x|x\in A)dx,$$

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

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 and fix two points x and y in $\mathbb{R}^d$.

Theorem (Hwang, Damelin and Hero 2015)

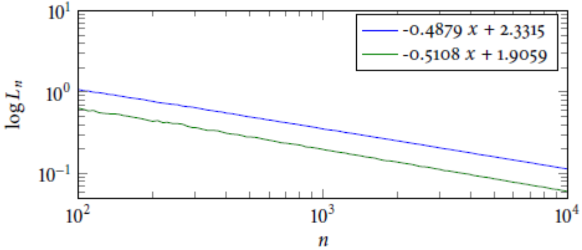
Assume that $\inf_x f(x) > 0$ over a compact support set $\mathcal{S}$ with pd metric tensor g. The shortest path between any two points $x, y \in \mathcal{S}$ satisfies

$$L_\gamma^{SP}(\mathcal{X})/n^{(1-\gamma)/d} \to C_{d,\gamma} \underbrace{\inf_\pi \int_0^1 f(\pi_t)^{(1-\gamma)/d} \sqrt{g(\pi'(t), \pi'(t))} dt}_{dist_\gamma(x,y)} \qquad (a.s.)$$

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

- S.-J. Hwang, A. Hero, "Shortest path through random points," submitted (arXiv:1202.0045) 2012.

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: $x = (0, 1, 0)$, $y = (0, 0, 1)$

Outline

Motivation of non-dominated sorting

Focus has been on finding *a solution* to a convex optimization problem

- Basis pursuit and dictionary learning find "a best match."
- Parametric estimation produces a ML, MAP, or min MSE estimator.
- Matrix completion gives "the best signal reconstruction."

Motivation of non-dominated sorting

Focus has been on finding *a solution* to a convex optimization problem

- Basis pursuit and dictionary learning find "a best match."
- Parametric estimation produces a ML, MAP, or min MSE estimator.
- Matrix completion gives "the best signal reconstruction."

Emerging area in Imaging, SP and ML: "Learning to rank"

- *Burges, Shaked, Renshaw, Lazier, Deeds, Hamilton, and Hullender, Learning to rank using gradient descent. ICML 2005.*
- *Jamieson, Nowak, Active ranking using pairwise comparisons, NIPS 2011.*
- *Osting, Brune, Osher, Enhanced statistical rankings via targeted data collection, ICML 2013.*
- *Duchi, Mackey, Jordan, The asymptotics of ranking algorithms, Ann. Stat. 2013.*

Motivation of non-dominated sorting

Focus has been on finding *a solution* to a convex optimization problem

- Basis pursuit and dictionary learning find "a best match."
- Parametric estimation produces a ML, MAP, or min MSE estimator.
- Matrix completion gives "the best signal reconstruction."

Emerging area in Imaging, SP and ML: "Learning to rank"

- *Burges, Shaked, Renshaw, Lazier, Deeds, Hamilton, and Hullender, Learning to rank using gradient descent. ICML 2005.*
- *Jamieson, Nowak, Active ranking using pairwise comparisons, NIPS 2011.*
- *Osting, Brune, Osher, Enhanced statistical rankings via targeted data collection, ICML 2013.*
- *Duchi, Mackey, Jordan, The asymptotics of ranking algorithms, Ann. Stat. 2013.*

Driving application: database search and retrieval

- Internet users won't examine more than a few of the top matches
- In many cases there exist multiple criteria, multiple ranking functions
  - cost vs location of hotels in TripAdvisor search
  - recency vs number of citations in Google Scholar search
  - risk vs expected return in financial portfolio selection

Example: retrieve images combining semantic concepts

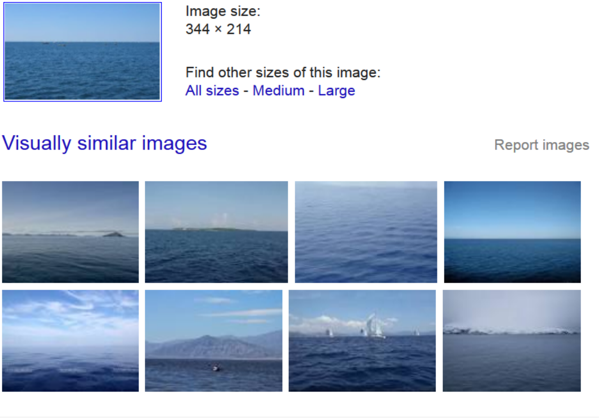
Objective: search a database for images combining semantic concepts of "sea" and "mountain"

Query 1

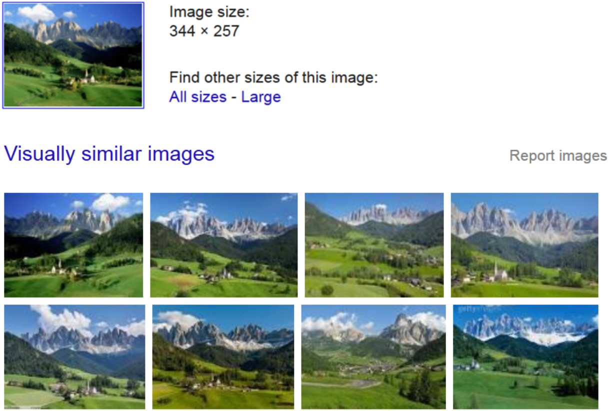
Query 2

Desired match

Example (ctd): Query 1 results from Google images



Example (ctd): Query 2 results from Google images



Problem: single query searches can't combine multiple concepts

Matches to concepts sorted according to scalar ranking functions $r_1, r_2 \geq 0$

$$r_1(i_1) < r_1(i_2) < \ldots < r_1(i_n)$$

$$r_2(j_1) < r_2(j_2) < \ldots < r_2(j_n)$$

Problem: Single matching criterion cannot easily combine different concepts.

The combined concepts will tend to be far down each list.

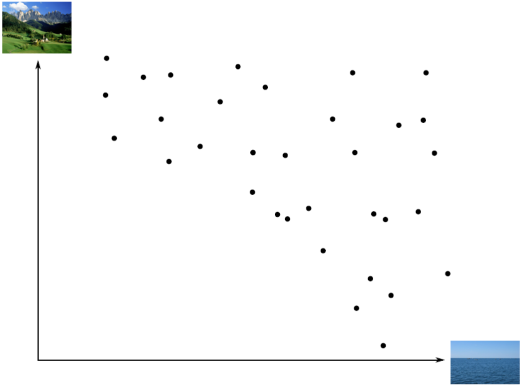
Possible solutions:

- Semantic labeling with text tags: requires human intervention
- Metasearch: search the results of the searches
- Scalarization: convert r_1 and r_2 to single criterion

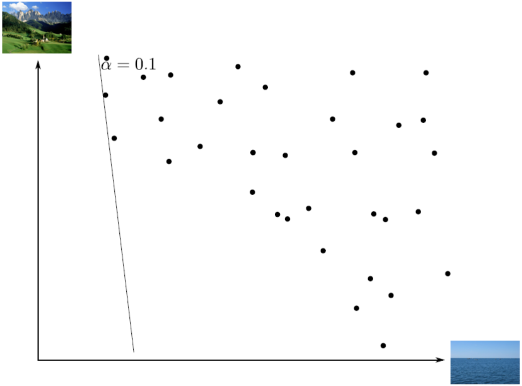
$$r_\alpha = \alpha r_1 + (1 - \alpha) r_2, \quad \alpha \in [0, 1]$$

Scalarization is a convexification of the multiple criteria $[r_1, r_2]$ requiring specification of α

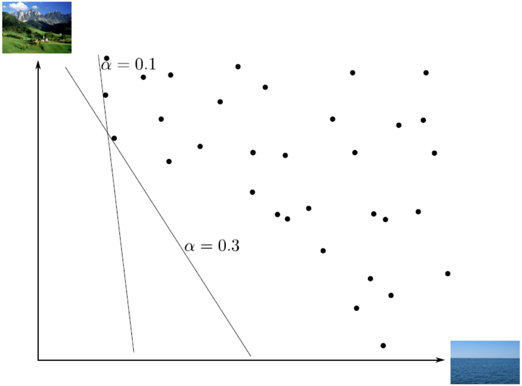
Example (ctd): Set of pairs $\mathcal{X} = \{[r_1(i), r_2(i)]\}_{i=1}^n$



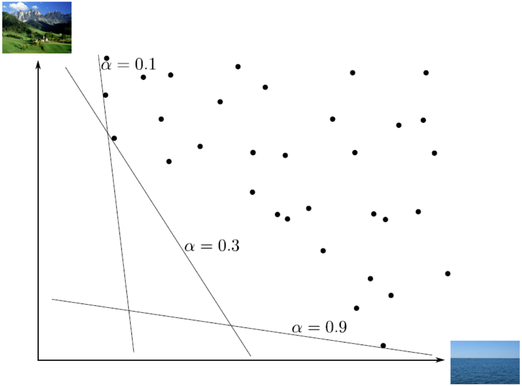
Example (ctd): Scalarization with $\alpha = 0.1$



Example (ctd): Scalarization with $\alpha = 0.3$



Example (ctd): Scalarization with $\alpha = 0.9$



Alternative: Non-dominated ranking combines multiple concepts

Scalarization only guaranteed to highly rank those images on convex hull

Alternative: non-dominated ranking (multi-objective optimization)

Alternative: Non-dominated ranking combines multiple concepts

Scalarization only guaranteed to highly rank those images on convex hull

Alternative: non-dominated ranking (multi-objective optimization)

Let $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be n points in $\mathbb{R}^d$

Alternative: Non-dominated ranking combines multiple concepts

Scalarization only guaranteed to highly rank those images on convex hull

Alternative: non-dominated ranking (multi-objective optimization)

Let $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be n points in $\mathbb{R}^d$

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

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

Alternative: Non-dominated ranking combines multiple concepts

Scalarization only guaranteed to highly rank those images on convex hull

Alternative: non-dominated ranking (multi-objective optimization)

Let $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be n points in $\mathbb{R}^d$

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

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

$\mathbf{X}$ is a minimal element of $\mathcal{X}$ if

1) $\mathbf{X} \in \mathcal{X}$ *and 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 $\mathcal{X}$.

Alternative: Non-dominated ranking combines multiple concepts

Scalarization only guaranteed to highly rank those images on convex hull

Alternative: non-dominated ranking (multi-objective optimization)

Let $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be n points in $\mathbb{R}^d$

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

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

$\mathbf{X}$ is a minimal element of $\mathcal{X}$ if

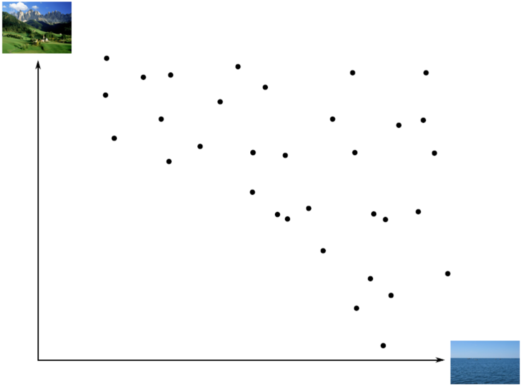
1) $\mathbf{X} \in \mathcal{X}$ *and 2)* $\{\mathbf{X}_i \in \mathcal{X} : \mathbf{X}_i \leqq \mathbf{X}\} = \emptyset$

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

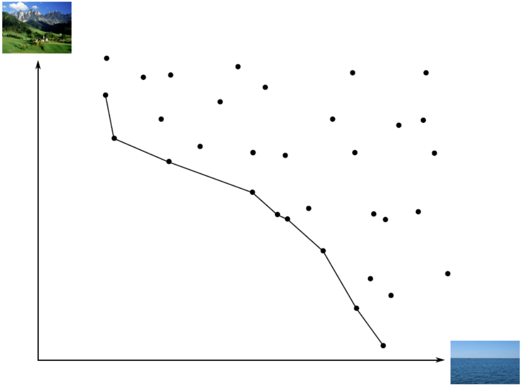
A Pareto front of depth k, denoted $\{\mathcal{F}_k\}$, is defined recursively

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

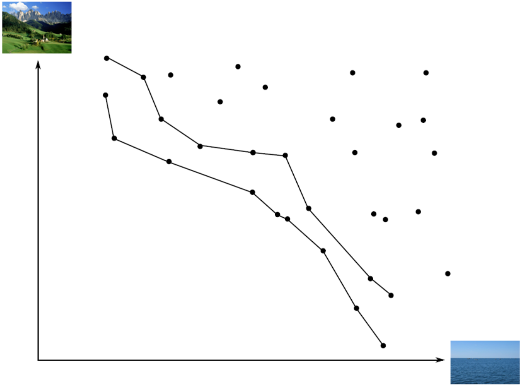
Example (ctd): Set of pairs $\mathcal{F} = \{[f_1(i), f_2(i)]\}_{i=1}^n$



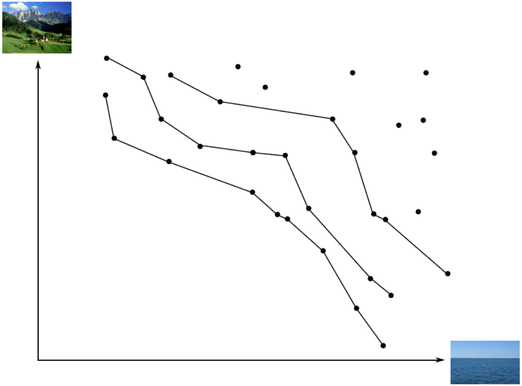
Example (ctd): Pareto front $\mathcal{F}_1$



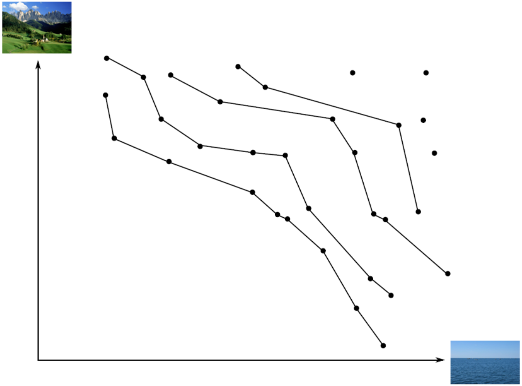
Example (ctd): Pareto fronts $\cup_{i=1}^{2}\mathcal{F}_i$



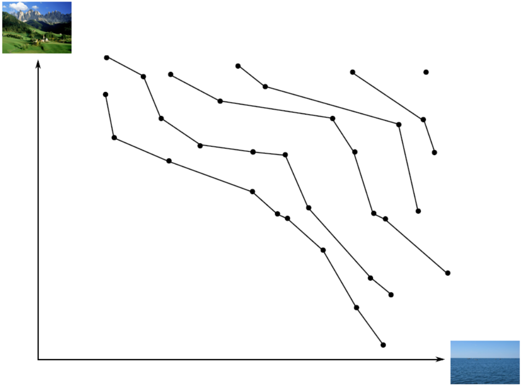
Example (ctd): Pareto fronts $\cup_{i=1}^{3}\mathcal{F}_i$



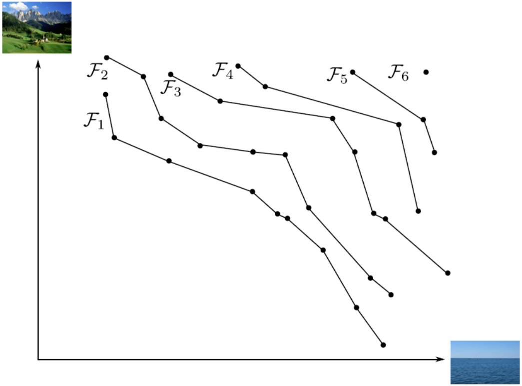
Example (ctd): Pareto fronts $\cup_{i=1}^{4}\mathcal{F}_i$



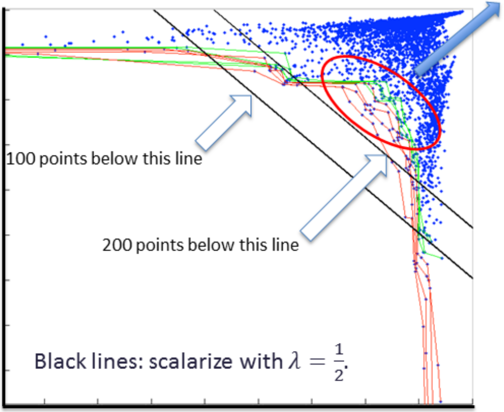
Example (ctd): Pareto fronts $\cup_{i=1}^{5} \mathcal{F}_i$



Example (ctd): Pareto fronts $\cup_{i=1}^{6}\mathcal{F}_i$



Real application: multiple concept image retrieval



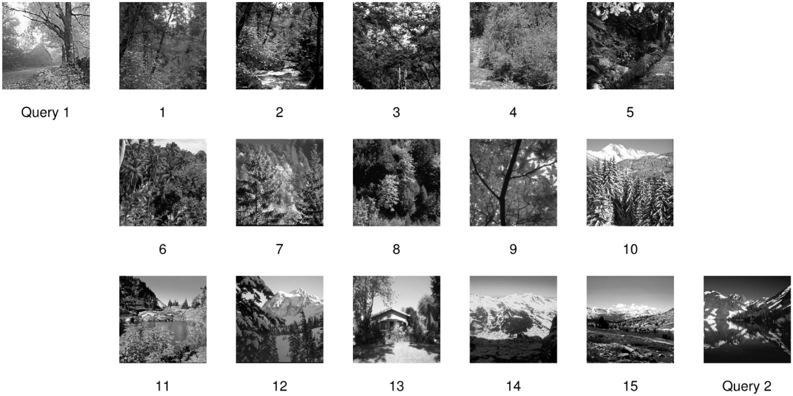
Pareto fronts give high ranks to points that are not highly ranked by linear scalarization.

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

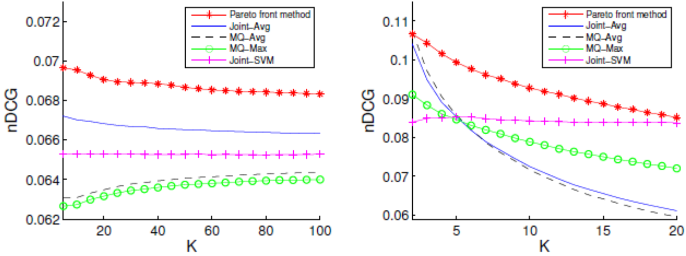
Hsiao, Calder and H, "Multiple-query Image Retrieval using Pareto Front Method," IEEE Trans. on Image Processing 2015.

Real application: first Pareto front (Skyline)



Stanford Scene dataset, SIFT feature, Spatial Pyramid Matching

Multi-query retrieval performance comparisons



(a) **Mediamill** dataset (b) **LAMDA** dataset

nDCR is normalized discounted cumulative relevance score (Hsiao 2015) that measures the relevance of top K matches to the queries.

Mediamill and LAMBDA are widely used multi-concept benchmarking datasets

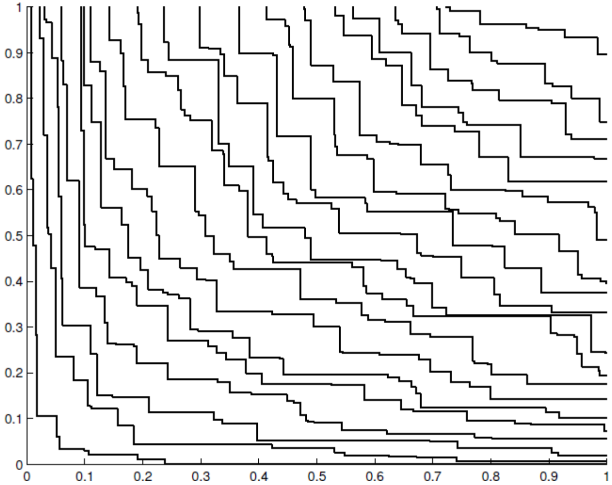
- Mediamill has 29800 videos and 101 semantic concept labels (Snoek 2006)
- LAMBDA has 2000 images with 5 class labels: desert, sea, sunset, mountains, trees (Zhou 2006)

• C. Snoek, M. Worring, J. Van Gemert, J. Geusebroek, and A. Smeulders, The challenge problem for automated detection of 101 semantic concepts in multimedia, in Proceedings of the 14th annual ACM Intern. Conf. on Multimedia, 2006

• Z.-H. Zhou and M.-L. Zhang, Multi-instance multi-label learning with application to scene classification, Advances in Neural Information Processing Systems,2006.

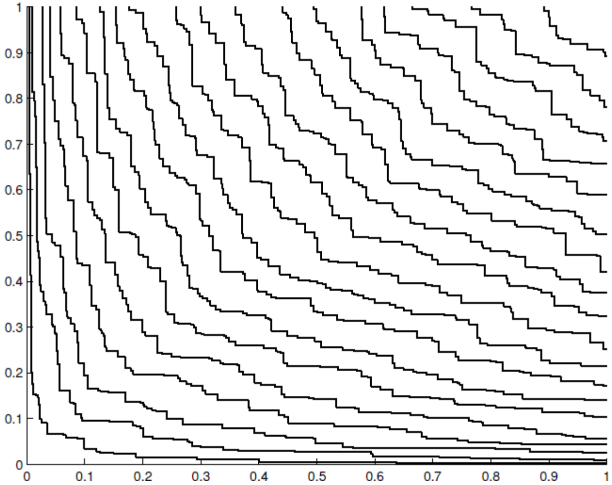
Outline

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



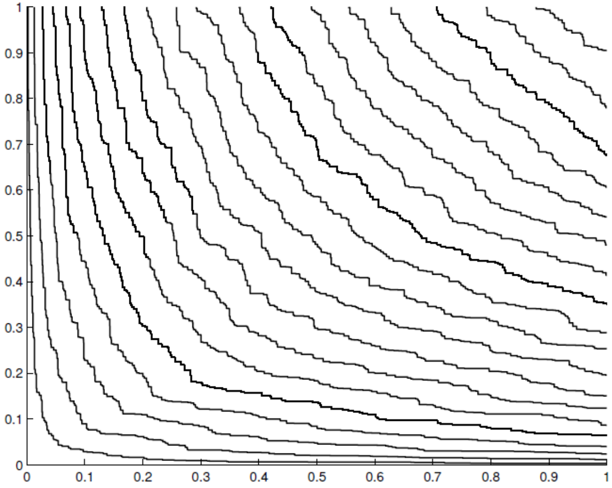
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



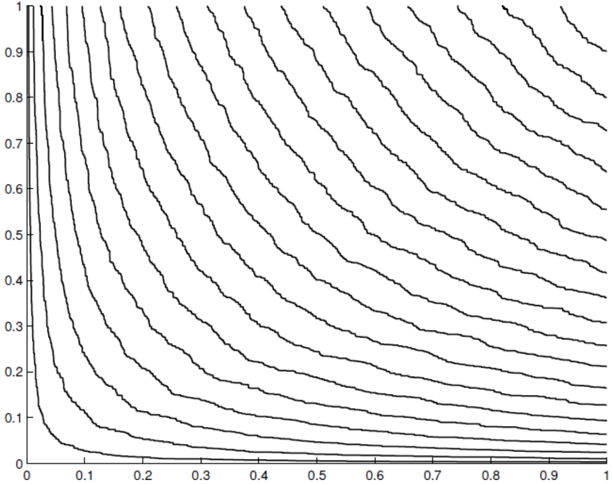
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



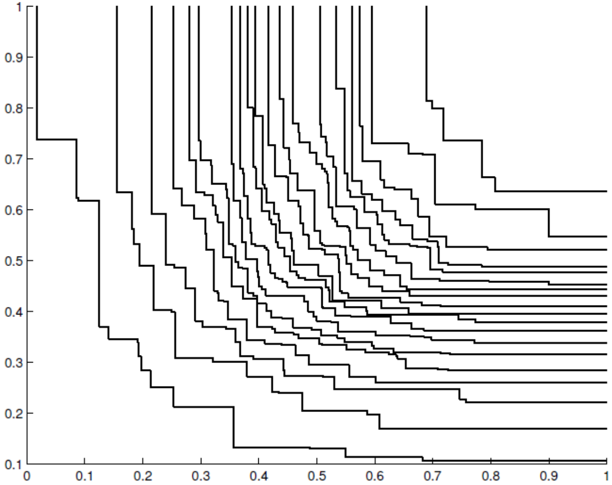
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Continuum limit: Demo for Unif$[0,1]^2$



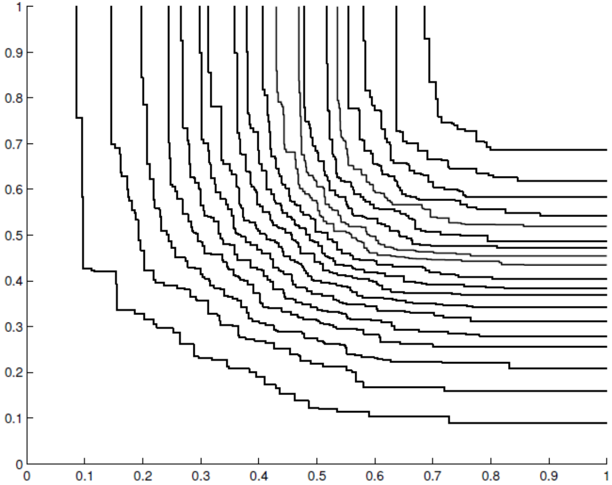
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Continuum limit: Demo for Gaussian$_2(0, I)$



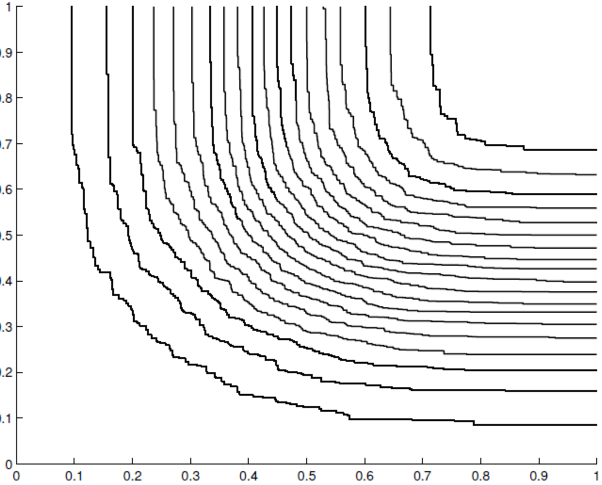
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Continuum limit: Demo for Gaussian$_2(0, I)$



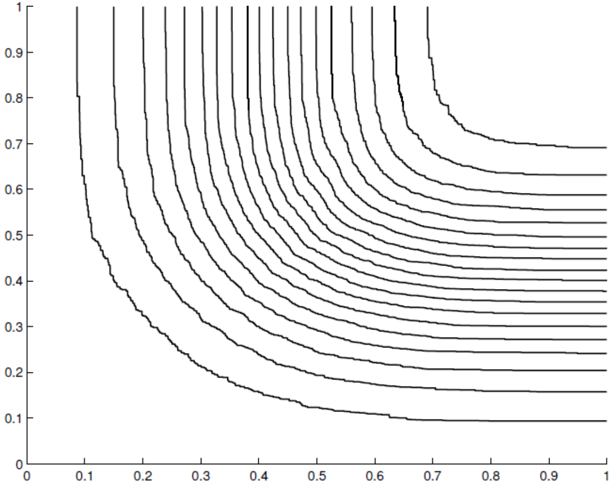
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Continuum limit: Demo for Gaussian$_2(0, I)$



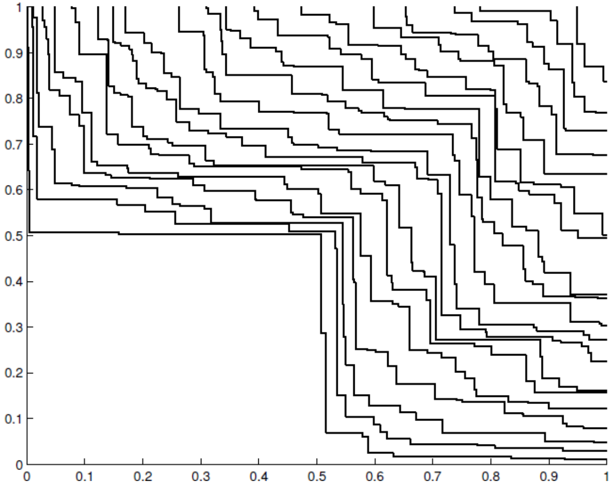
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Continuum limit: Demo for $\text{Gaussian}_2(0, I)$



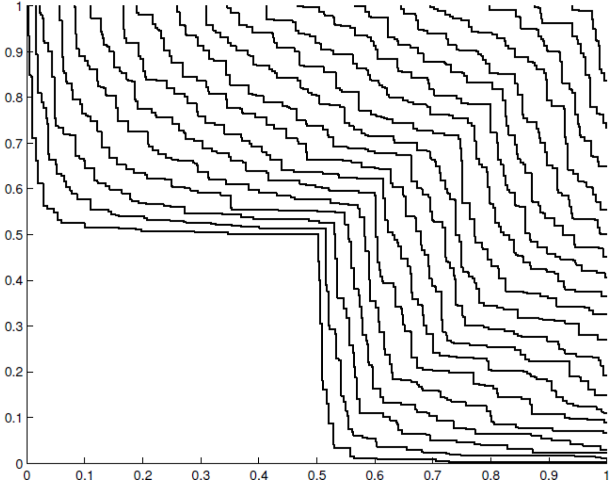
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

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



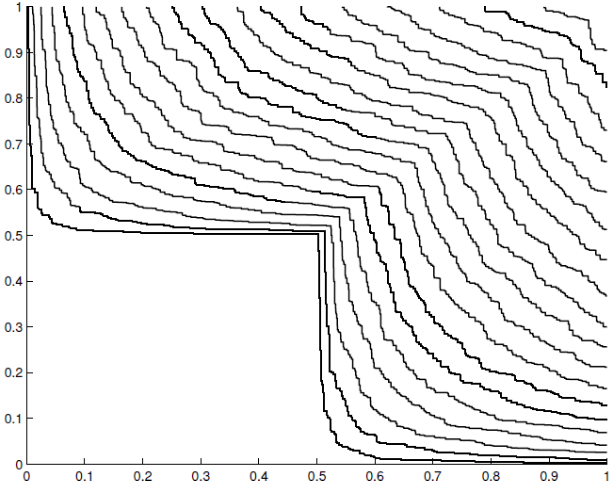
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Continuum limit: Demo for Unif$[0,1]^2/[0,0.5]^2$



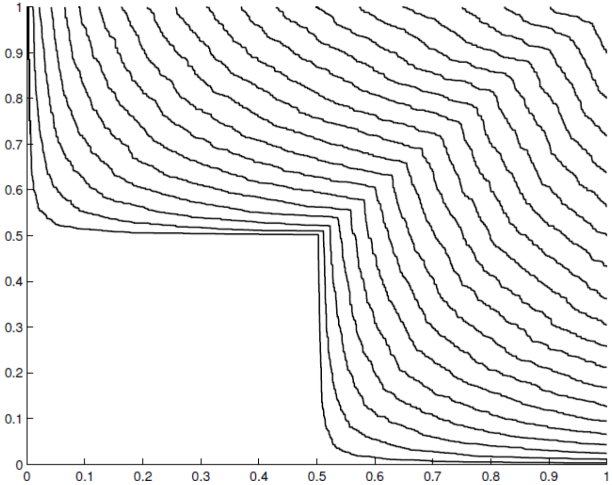
J. Calder, "Hamilton-Jacobi equations for sorting and percolation problems", PhD thesis Univ Michigan 2014.

Continuum limit: Demo for 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: Demo for 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

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

Theorem (Calder et al. [2014])

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

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

where

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

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

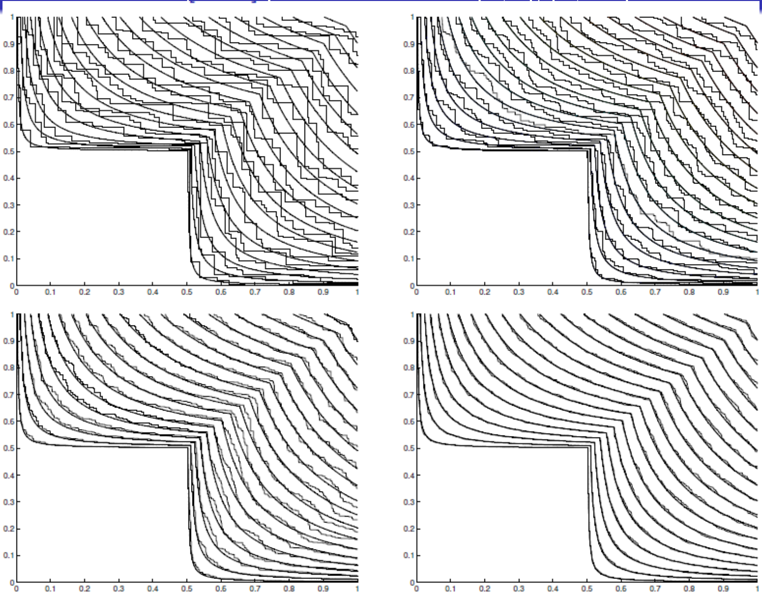
where $\mathcal{A} = \left\{\gamma \in C^1(0,1;\mathbb{R}^d) : \gamma'(t) \geqq 0\ \forall t \in [0,1]\right\}$

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

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

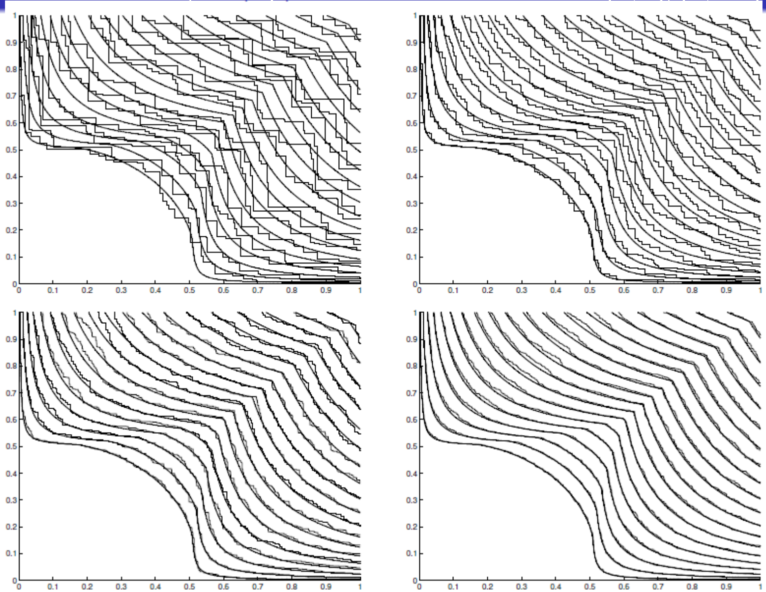
[a] $U(\mathbf{x}) \leq U(\mathbf{y})$ if $\mathbf{x} \leqq \mathbf{y}$

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.

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



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

Discussion

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

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

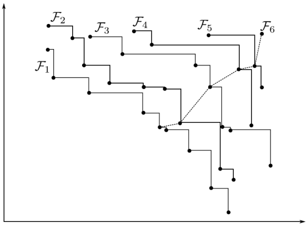
- Superadditivity implies convergence of $n^{-1/d} u_n$
- Smoothness of f implies convergent limit obeys Hamiltonian-Jacobi p.d.e.

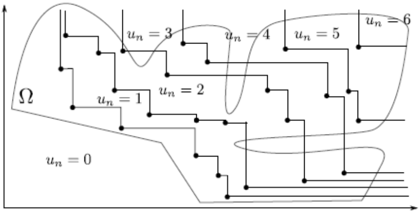
Low complexity (linear) numerical p.d.e. solver proposed (Calder et al. [2013])

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

J. Calder, S. Esedoglu, A. O. Hero, "A PDE-based approach to non-dominated sorting," SIAM Numerical Analysis, 2015

Relation of Pareto fronts to longest chain problem





A chain is a sequence $x_1, \ldots, x_n$ such that

$$x_1 \leqq \ldots \leqq x_l$$

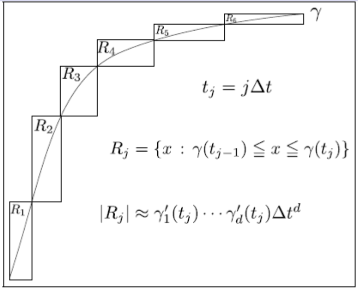
Equivalent definition of counting function $u_n(x)$

$u_n(x)$ *is length of longest chain in* $\{X_i : X_i \leqq x\}$

Note: Number of points on a front $u_n(x) = k$ is of order $n^{\frac{d-1}{d}}$
$\Rightarrow$ Number of fronts is of order $n^{\frac{1}{d}}$

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

Proof concept behind integral form of continuum limit



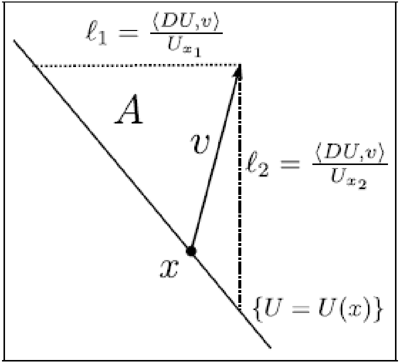
$$\begin{aligned}\ell(\{X_1,\ldots,X_n\}\cap R_j) &\approx c_d(f(\gamma(t_j))|R_j|n)^{\frac{1}{d}}\\ &\approx c_d f(\gamma(t_j))^{\frac{1}{d}}\left(\gamma_1'(t_j)\cdots\gamma_d'(t_j)\right)^{\frac{1}{d}}\Delta t\, n^{\frac{1}{d}}\end{aligned}$$

By monotonicity of γ, can connect the chains within R_j's

$$\begin{aligned}\ell(\{X_1,\ldots,X_n\}) &\geq \sum_j \ell(\{X_1,\ldots,X_n\}\cap R_j)\\ &\approx c_d\left(\sum_j f^{\frac{1}{d}}(\gamma(t_j))(\gamma_1^{'}(t_j)\cdots\gamma_d^{'}(t_j))^{\frac{1}{d}}\Delta t\right) n^{\frac{1}{d}}\end{aligned}$$

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

Proof concept behind pde form of continuum limit



For small $|v|$

$$\begin{aligned}\langle DU, v\rangle &\approx U(x+v) - U(x)\\ &\approx (\#\text{ fronts in } A)n^{-\frac{1}{d}}\\ &\approx (\#\text{ samples in } A)^{\frac{1}{d}} n^{-\frac{1}{d}}\\ &\approx (n|A|f(x))^{\frac{1}{d}} n^{-\frac{1}{d}}.\end{aligned}$$

Using $|A| \approx \frac{\langle DU, v\rangle^d}{U_{x_1}\cdots U_{x_d}}$ we have

$$\langle DU, v\rangle \approx \left(\frac{f(x)}{U_{x_1}\cdots U_{x_d}}\right)^{\frac{1}{d}} \langle DU, v\rangle$$

$$\boxed{U_{x_1}\cdots U_{x_d} = f}$$

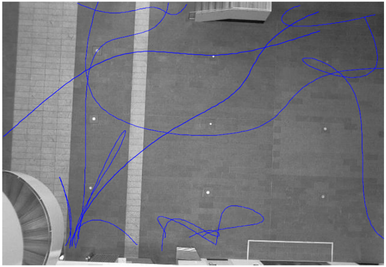
Rigorous proof is more complicated since

- on this slide we assumed $U \in C^1$, which is not generally true
- on this slide we assumed $n^{-1/d} u_n \to U$
- norm of HJ p.d.e. is non-standard and may not have unique soln

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

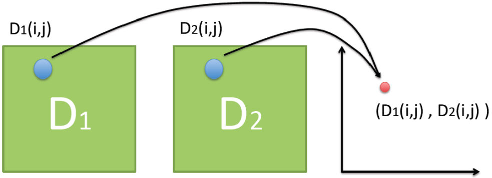
Speed and shape similarity between trajectories $T_i(t), T_j(t) \in \mathbb{R}^2$:

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

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

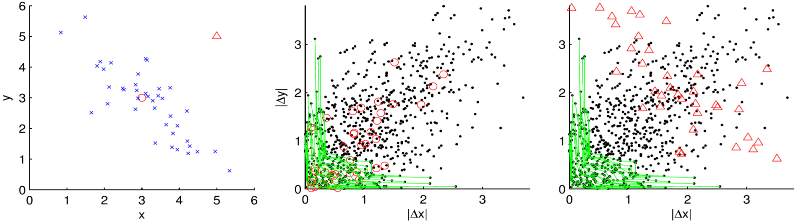
1. Scalarization:
   $D_\lambda(i,j) = \lambda D_1(i,j) + (1-\lambda) D_2(i,j)$

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.

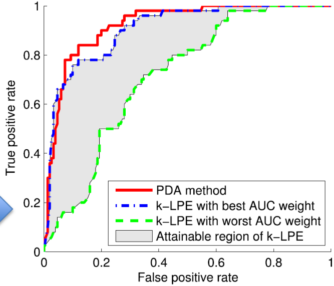
Detection performance of multicriteria anomaly detection



PDA Algorithm:

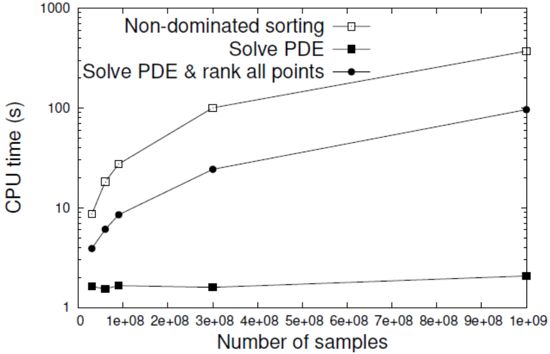
- Embed N choose 2 dyads onto plane
- Build Pareto fronts of non-dominated dyads.
- Compute anomaly scores = depth of front.

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

J. Calder, S. Esedoglu, A. O. Hero, "A PDE-based approach to non-dominated sorting," SIAM Numerical Analysis, 2015. Calder et al. [2013]

References for continuum limits of non-dominated sorting

References for continuum approximations to non-dominated sorting:

1. J. Calder, S. Esedoglu and A.O. Hero, "A Hamilton-Jacobi equation for the continuum limit of non-dominated sorting", SIAM Mathematical Analysis, Feb 2014. arXiv:1302.5828.

2. J. Calder, S. Esedoglu, A. O. Hero, "A PDE-based approach to non-dominated sorting," SIAM Numerical Analysis, Jan 2015. arxiv:1320.2498.

A gentle introduction to the theory:

1. J. Calder, S. Esedoglu and A.O. Hero, "A continuum limit for non-dominated sorting," Conference on Information Theory and Applications(ITA), Feb. 2014.

References for applications of non-dominated sorting:

1. K.-J. Hsiao, J. Calder and A.O. Hero, "Pareto-depth for multiple-query image retrieval," IEEE Trans. Image Processing (in press) 2015. arxiv:1402.5176

2. K.-J. Hsiao, K. S. Xu, J. Calder and A. O. Hero, "Multi-criteria Anomaly Detection using Pareto Depth Analysis", Proc. of Neural Information Processing Systems (NIPS), Dec. 2012. arxiv:1110.3741 .

Outline

Summary

- Asymptotic continuum limits can be useful for combinatorial problems in imaging
- These limits are related to entropy and divergence and motivate
    - New methods for geometric combinatorial optimization problems
    - New approaches to clustering, classification, registration in high dimension
    - Continuous optimization interpretations for certain discrete optimization problems
- There are other fruitful continuum limit applications: directed graphs, multigraphs, hypergraphs.

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

J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. *Proc. Cambridge Philosophical Society*, 55: 299–327, 1959.

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. *arXiv preprint arXiv:1310.2498*, 2013.

K Carter. *Dimensionality Reduction on Statistical Manifolds*. PhD thesis, University of Michigan, Dept of EECS, 2008.

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.

John C Duchi, Lester Mackey, Michael I Jordan, et al. The asymptotics of ranking algorithms. *The Annals of Statistics*, 41(5):2292–2323, 2013.

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.

Kevin G Jamieson and Robert D Nowak. Active ranking using pairwise comparisons. *arXiv preprint arXiv:1109.3701*, 2011.

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

H. Neemuchwala, A. O. Hero, and P.L. Carson. Image matching using alpha-entropy measures and entropic graphs. *European Journal of Signal Processing (Special Issue on Content-based Visual Information Retrieval)*, 85:277–296, 2005b.

Braxton Osting, Christoph Brune, and Stanley Osher. Enhanced statistical rankings via targeted data collection. In *Proceedings of the 30th International Conference on Machine Learning*, pages 489–497, 2013.

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.

SU Park, D Wei, M De Graef, M Shah, J Simmons, and AO Hero. A physics-based pattern dictionary for ebsd image segmentation. *Microscopy and Microanalysis*, 19(S2):734–735, 2013.

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.