Combinatorial continuum limits and their applications

Alfred Hero

†University of Michigan

June 1, 2015

Conference on Scale Space and Variational Methods (SSVM)
Lege, France
1. Geometric graphs in imaging and computer vision

2. Minimal graphs

3. Continuum limits

4. Non-dominated sorting

5. Continuum limits

6. Application to anomaly detection

7. Summary
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
1 Geometric graphs in imaging and computer vision

2 Minimal graphs

3 Continuum limits

4 Non-dominated sorting

5 Continuum limits

6 Application to anomaly detection

7 Summary
Geometric graphs in imaging and computer vision

A geometric graph has nodes $V$ that represent real valued features and edges $E$ 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 $\{X_1, \ldots, X_n\}$ 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 polycrystalline 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 polycrystalline 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

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

1 Geometric graphs in imaging and computer vision

2 Minimal graphs

3 Continuum limits

4 Non-dominated sorting

5 Continuum limits

6 Application to anomaly detection

7 Summary
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 $\mathcal{M}$
- $A = ((a_{ij}))$: adjacency matrix associated with $\mathcal{G}$

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

- $d_i = \sum_j a_{ij}$: 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

\[
L_{kNN}^γ(V) = \min_{\mathcal{E}: A_1 = k} L_γ(V, \mathcal{E}) \\
= \min_{\mathcal{E}: A_1 = k} \sum_{e_{ij} \in \mathcal{E}} |e_{ij}|^γ \\
= \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_k(X_i)} \|X_i - X_j\|^γ
\]

- \(\mathcal{N}_k(X_i)\) 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

k-NNG-based spectral algorithm

- Extract features $\mathcal{X}_n = \{X_1, \ldots, X_n\}$
- Compute similarity matrix $W$ between $X_i$’s
- Use $W$ to construct kNNG graph over $\mathcal{X}_n$
- $(V, \Lambda) = \text{Eigendecomp}(W - D), \ D = \text{diag}(W1)$
  - Dimension reduction: $Y_n = \Lambda_{2 \times 2}^{1/2}[v_1, v_2]^T X_n$
  - Spectral clustering: $K$-means($v_2$)

Adjacency matrix kNNG

Minimal spanning tree (MST)

- MST is solution of the optimization

\[
L_{\gamma}^{MST}(\mathcal{V}) = \min_{\mathcal{E}: A_1 > 0} L_{\gamma}(\mathcal{V}, \mathcal{E}) = \min_{\mathcal{E}: A_1 > 0} \sum_{e_{ij} \in \mathcal{E}} |e_{ij}|^\gamma
\]

- 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 \( O(n^2 \log n) \)
Minimal spanning tree (MST)

- MST is solution of the optimization

\[ L^{MST}_\gamma (\mathcal{V}) = \min_{\mathcal{E}: A_1 > 0} L_\gamma (\mathcal{V}, \mathcal{E}) \]

\[ = \min_{\mathcal{E}: A_1 > 0} \sum_{e_{ij} \in \mathcal{E}} |e_{ij}|^\gamma \]

- 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)

Friedman-Rafsky graph (FR)

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

\[
L_{\gamma}^{MST}(\mathcal{V}) = \min_{\mathcal{E} : A_1 > 0} L_\gamma(\mathcal{V}, \mathcal{E})
= \sum_{e_{ij} \in \mathcal{E}^*} |e_{ij}^{XX}|^\gamma + |e_{ij}^{XY}|^\gamma + |e_{ij}^{YY}|^\gamma
\]

- 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) between 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 $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

$$L_{\gamma}^{MST}(\mathcal{V}) = \min_{\mathcal{E} : A \geq 0} L_{\gamma}(\mathcal{V}, \mathcal{E})$$

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

- 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) between 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 $\mathcal{V} = \mathcal{X}_n \cup \mathcal{Y}_m$

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

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

- 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 $z_i \in \mathbb{R}^2$:
$$X(i) = [I_1(z_i), T(I_2(z_i))]$$

Joint intensity histogram
$$p_X(x_1, x_2) = n^{-1} \sum_{i=1}^{n} \chi_{[x_1, x_2]}(X(i))$$

Maximize mutual information (MI)
$$\max_T \sum_{x_1, x_2=0}^{255} p_X(x_1, x_2) \ln \left( \frac{p_X(x_1, x_2)}{p_{X_1}(x_1)p_T(X_2)(x_2)} \right) = \max_T H(I_1, T(I_2)) - H(I_1) - H(T(I_2))$$

Where have defined entropy of $V$
$$H(V) = n^{-1} \sum_v \ln \frac{1}{p_V(v)}$$


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 $z_i \in \mathbb{R}^2$:

$X_1(i) = [W(z_i), z_i], \quad X_2(i) = [W(z_i), z_i]$

$W_1(z_i)$ and $W_2(z_i)$ are localized Meyer wavelet coefficients of $I_1$ and $T(I_2)$

Maximize FR statistic

$$\max_T L^{FR}_{\gamma}(X_1, X_2)$$

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

**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

$$L^{kMST}_\gamma(\mathcal{V}) = \min_{\mathcal{V}_k:|\mathcal{V}_k|=k} L^{MST}_\gamma(\mathcal{V}_k)$$

$$= \min_{\mathcal{V}_k:|\mathcal{V}_k|=k} \min_{\mathcal{E}_k: \mathbf{A}_k \geq 0} \sum_{e_{ij} \in \mathcal{E}_k} |e_{ij}|^\gamma$$

- 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)

Denoising illustration of kMST

\[ f = (1 - \epsilon) f_1 + \epsilon f_0 \]

Illustration: kMST for WSN intruder detection

- A. Hero, "Geometric entropy minimization (GEM) for anomaly detection and localization," NIPS 2006
Shortest path (SP)

- Let $G$ be a graph with $m = |E|$ edges on $n$ vertices $V$
- $\pi(X_I, X_F)$ a path over $G$ between points $X_I$ and $X_F$
  \[ \pi(X_I, X_F) = (X_I, X_{i_1}, \ldots, X_{i_l}, X_F) \]
  $X_{i_{j+1}}$ is a neighbor on $G$ of predecessor $X_{i_j}$ and $X_I = X_{i_0}$, $X_F = X_{i_{l+1}}$
- The shortest path is the solution to
  \[ L_{SP}^\gamma(V) = \min_{\pi(X_I, X_F)} \sum_{X_i \in \pi(X_I, X_F)} |X_{i+1} - X_i|^{\gamma} \]
- Typical choices of $G$:
  - 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."
Outline

1. Geometric graphs in imaging and computer vision
2. Minimal graphs
3. Continuum limits
4. Non-dominated sorting
5. Continuum limits
6. Application to anomaly detection
7. Summary
MST continuum limit: MST length functional captures “spread” of distribution
Large $n$ behavior of MST length functional

\[ \text{length}(\text{MST}) \]

\[ \frac{\log \text{length}(\text{MST})}{\sqrt{n}} \]

MST length, Unif. dist. (red), Triang. dist (blue)

MST normalized compensated length

Unif.

Triang.
Continuum limit of kNN and MST length functionals

Theorem (Beardwood, Halton&Hammersley 1959)

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

$$
\lim_{n \to \infty} L_{\gamma}^{\text{MST},k\text{NN}}(X_n)/n^{(d-\gamma)/d} = \beta_{\gamma,d} \int f(x)^{(d-\gamma)/d} dx, \quad (a.s.)
$$

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

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

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

- RMS rate of convergence (Costa & Hero 2003)

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

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
    \[ \forall x \in \mathbb{R}^d, \quad L_\gamma(F + x) = L_\gamma(F), \quad \text{(translation invariance)} \]
    \[ \forall c > 0, \quad L_\gamma(cF) = c^\gamma L_\gamma(F), \quad \text{(homogeneity)} \]
  - Null condition: $L_\gamma(\emptyset) = 0$, where $\emptyset$ is the null set
  - Subadditivity: There exists a constant $C_1$ with the following property: For any uniform resolution $1/m$-partition $Q^m$
    \[ L_\gamma(F) \leq m^{-\gamma} \sum_{i=1}^{m^d} L_\gamma(m[(F \cap Q_i) - q_i]) + C_1 m^{d-\gamma} \]
  - Superadditivity: For same conditions as above, there exists a constant $C_2$
    \[ L_\gamma(F) \geq m^{-\gamma} \sum_{i=1}^{m^d} L_\gamma(m[(F \cap Q_i) - q_i]) - C_2 m^{d-\gamma} \]
  - Continuity: There exists a constant $C_3$ such that for all finite subsets $F$ and $G$ of $[0, 1]^d$
    \[ |L_\gamma(F \cup G) - L_\gamma(F)| \leq C_3 \left( \text{card}(G) \right)^{(d-\gamma)/d} \]

Main ideas behind proof of BHH (Yukich 1998)

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

- Avg distance between $n$ points in $[0, 1]^d$
  \[ |e_i|_{\text{avg}} = n^{-1/d} \]
- Avg length of MST should therefore be
  \[ L_{\gamma}^{\text{MST}} = \sum_{i=1}^{n-1} |e_i|_{\gamma} \approx c n n^{-\gamma/d} = cn^{(d-\gamma)/d} \]
- 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
  \[ L_{\gamma}^{\text{MST}}(\mathcal{X}_n) \approx m^{-\gamma} \sum_{i=1}^{m^d} L_{\gamma}^{\text{MST}}(m(\mathcal{X}_n \cap Q_i)) \]
- From the $[0, 1]^d$ result
  \[ L_{\gamma}^{\text{MST}}(m(\mathcal{X}_n \cap Q_i)) = c(n_i)^{(d-\gamma)/d} \]
- From smoothness of $f$
  \[ n_i/n \approx m^{-d} f(q_i) \]
- Therefore
  \[ L_{\gamma}^{\text{MST}}(m(\mathcal{X}_n \cap Q_i)) \approx cn^{(d-\gamma)/d} (m^{-d} f)^{(d-\gamma)/d} \]
- since
  \[ (m^{-d} f)^{(d-\gamma)/d} = m^{\gamma} m^{-1/d} f^{(d-\gamma)/d}(q_i) \]
  \[ L_{\gamma}^{\text{MST}}(\mathcal{X}_n) \approx n^{(d-\gamma)/d} \cdot c \sum_{i=1}^{m^d} f^{(d-\gamma)/d}(q_i) m^{-1/d} \]
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 $S$ 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)

BHH theorem Riemannian extension

Theorem (Costa 2004, 2005)

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

\[
\lim_{n \to \infty} \frac{L_{\gamma}^{\text{MST}}(X_n)}{n^\alpha} = \beta_{d, \gamma} \int_S f^\alpha(x) d\mu_g
\]

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

Alternative representation For finite \(n\)

\[
\log L_{\gamma}^{\text{MST}}(X_n) = \alpha \log n + (1 - \alpha)H_\alpha(X) + \log \beta_{d, L} + \varepsilon(n)
\]

where

\[
H_\alpha(X) = (1 - \alpha)^{-1} \ln \int_S f^\alpha(x) d\mu_g
\]

is \(\alpha\)-entropy of \(X\) and \(\varepsilon(n) \to 0 \text{ w.p.} 1\).

Key observation: can use representation of \(\log L_{\gamma}^{\text{MST}}\) to estimate intrinsic dimension \(d\) of \(S\) in addition to entropy of \(f(x)\).
\[ \hat{d} = \text{round} \left( \frac{\gamma}{1 - a} \right) = 2 \]
\[ \hat{H}_\alpha(X) = \frac{b - \gamma/2 \log \beta_{d, \gamma}}{1 - a} = 7.3 \]
\[ \text{Ground truth: } H_\alpha(X) = \log(1869) = 7.53 \]
Dimension estimation: MNIST digits

Local Dimension/Entropy Statistics

Dimension estimation: dimension-driven image segmentation

![Image of black bear and panda](image1.jpg)

![Graphs showing distribution of dimension counts](graphs.jpg)

- Dimension estimation: dimension-driven image segmentation
- Geometric graphs
- Minimal graphs
- Continuum limits
- Non-dominated sorting
- Continuum limits
- Application
- Summary
- References
Dimension estimation: dimension-driven image segmentation

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

$$L_{\gamma}^{k\text{MST}}(X_n)/\left(\rho n\right)^\alpha \to \beta_{\gamma,d} \inf_{A:Pr(A) \geq \rho} \int f^\alpha(x|x \in A)dx \quad (a.s.)$$

$$Pr(A) = \int_A f.$$  

Alternatively, defining the conditional entropy function

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

$$\frac{1}{1-\alpha} \ln \left(L_{\gamma}^{k\text{MST}}(X_n)/\left(\rho n\right)^\alpha\right) \to \beta_{\gamma,d} \inf_{A:Pr(A) \geq \rho} H_\alpha(f|x \in A) + c \quad (a.s.)$$

Solution to variational problem is a level set $A = A_\circ$ of $f$.

Continuum limit of $k$MST 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

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

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

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)

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

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 $S$ with pd metric tensor $g$. The shortest path between any two points $x, y \in S$ satisfies

$$L_{\gamma}^{SP}(\mathcal{X})/n^{(1-\gamma)/d} \rightarrow C_{d, \gamma} \inf_{\pi} \int_0^1 f(\pi_t)^{(1-\gamma)/d} \sqrt{g(\pi'(t), \pi'(t))} dt$$

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

Experimental validation of shortest path continuum limit

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

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

Experimental setting

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

1. Geometric graphs in imaging and computer vision
2. Minimal graphs
3. Continuum limits
4. Non-dominated sorting
5. Continuum limits
6. Application to anomaly detection
7. Summary
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”

- Jamieson, Nowak, Active ranking using pairwise comparisons, NIPS 2011.
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”

- Jamieson, Nowak, Active ranking using pairwise comparisons, NIPS 2011.

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

Image size:
344 x 214

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

Visually similar images
Example (ctd): Query 2 results from Google images

Image size:
344 × 257

Find other sizes of this image:
All sizes - Large

Visually similar 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 $\alpha$
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 $\mathcal{X} = \{X_1, \ldots, X_n\}$ 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 $\mathcal{X} = \{X_1, \ldots, X_n\}$ be $n$ points in $\mathbb{R}^d$

Define partial order relation "$\leq$" between any $X, Y \in \mathbb{R}^d$: $X \leq Y \iff X_i \leq Y_i, \ \forall i = 1, \ldots, d$
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 $\mathcal{X} = \{X_1, \ldots, X_n\}$ be $n$ points in $\mathbb{R}^d$

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

$$X \leq Y \iff X_i \leq Y_i, \ \forall i = 1, \ldots, d$$

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

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

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 $\mathcal{X} = \{X_1, \ldots, X_n\}$ be $n$ points in $\mathbb{R}^d$

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

$$X \leq Y \iff X_i \leq Y_i, \text{ } \forall i = 1, \ldots, d$$

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

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

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

A Pareto front of depth $k$, denoted $\{F_k\}$, is defined recursively

$$\begin{align*}
F_1 &= \min \mathcal{X} \\
F_k &= \min \left\{ \mathcal{X} / \bigcup_{i=1}^{k-1} F_i \right\}, \text{ } k = 1, 2, \ldots
\end{align*}$$
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 $\bigcup_{i=1}^{2} \mathcal{F}_i$
Example (ctd): Pareto fronts $\bigcup_{i=1}^{3} \mathcal{F}_i$
Example (ctd): Pareto fronts $\bigcup_{i=1}^{4} \mathcal{F}_i$
Example (ctd): Pareto fronts $\bigcup_{i=1}^{5} F_i$
Example (ctd): Pareto fronts $\bigcup_{i=1}^{6} F_i$
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.

Black lines: scalarize with $\lambda = \frac{1}{2}$.

Stanford Scene dataset, SIFT feature, Spatial Pyramid Matching

Real application: first Pareto front (Skyline)

Stanford Scene dataset, SIFT feature, Spatial Pyramid Matching
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)

Outline

1 Geometric graphs in imaging and computer vision
2 Minimal graphs
3 Continuum limits
4 Non-dominated sorting
5 Continuum limits
6 Application to anomaly detection
7 Summary
Continuum limit: Demo for Unif[0, 1]^2

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

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

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

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

Continuum limit: Demo for Gaussian_2(0, I)

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

Continuum limit: Demo for Unif\([0, 1]^2/ [0, 0.5]^2\)

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

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

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

Asymptotic theorem

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

**Theorem (Calder et al. [2014])**

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

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

where

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

\[
U(x) = \sup_{\gamma \in A} \int_0^1 f^{\frac{1}{d}}(\gamma(t))(\gamma'_1(t) \cdots \gamma'_d(t))^{\frac{1}{d}} dt
\]

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

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

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

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

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

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

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

- $u_n(x)$ is the length of longest chain in $\{X_i \in \mathcal{X} : X_i \leq x\}$.
- $\mathcal{F}_k$ is anti-chain containing $\{X_i \in \mathcal{X} : u_n(X_i) = k\}$
- $u_n = u\{x_1,\ldots,x_n\}$ is a superadditive functional in the sense that

$$u\{x_1,\ldots,x_n\}(x) \geq \sum_{i=1}^{m} u\{x_1,\ldots,x_n \cap R_i\}(x)$$

- 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])

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

Relation of Pareto fronts to longest chain problem

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

$$x_1 \leq \ldots \leq x_l$$

Equivalent definition of counting function $u_n(x)$

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

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 $n^{\frac{1}{d}}$

Proof concept behind integral form of continuum limit

By monotonicity of $\gamma$, can connect the chains within $R_j$’s

\begin{align*}
\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}} (\gamma_1'(t_j) \cdots \gamma_d'(t_j))^{\frac{1}{d}} \Delta t n^{\frac{1}{d}}
\end{align*}

\begin{align*}
\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{align*}

Proof concept behind pde form of continuum limit

For small $|v|$

$$\langle DU, v \rangle \approx U(x + v) - U(x)$$

$$\approx (\# \text{ fronts in } A) n^{-\frac{1}{d}}$$

$$\approx (\# \text{ samples in } A) \frac{1}{d} n^{-\frac{1}{d}}$$

$$\approx (n |A| f(x))^{\frac{1}{d}} n^{-\frac{1}{d}}.$$

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

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

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

1. Geometric graphs in imaging and computer vision
2. Minimal graphs
3. Continuum limits
4. Non-dominated sorting
5. Continuum limits
6. Application to anomaly detection
7. Summary
Motivation: Detect anomalous pedestrian trajectories.
Question: Which one of these groups of trajectories are anomalous?

Anomalous trajectories
Nominal trajectories

Curve features: curve length, shape, walking speed.

Multicriteria anomaly detection

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

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

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

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

2. Pareto depth analysis:
$$(D_1(i, j), D_2(i, j)) \rightarrow \text{one dyad}$$

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

Run-time comparisons

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

References for continuum limits of non-dominated sorting

References for continuum approximations to non-dominated sorting:


A gentle introduction to the theory:


References for applications of non-dominated sorting:


Outline

1. Geometric graphs in imaging and computer vision
2. Minimal graphs
3. Continuum limits
4. Non-dominated sorting
5. Continuum limits
6. Application to anomaly detection
7. Summary
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.


