IRI 2012 Las Vegas 10-13 August 2012

Information neighbourhoods for visualization and monitoring strategies: Computational aspects

C.T.J. Dodson

School of Mathematics, University of Manchester UK

ctdodson@manchester.ac.uk

Abstract

Aspects of modern information systems that are challenging computational and statistical analysis are dynamic complexity, high dimensionality, and inherent stochasticity.

We outline the use of geometric methods to provide **information neighbourhoods** for visualization and monitoring of algorithms, and dynamics of stochastic behavior trajectories.

Geometrization of models of real phenomena give valuable insights through features that are invariant under the choice of coordinate representation.

Here we look at **computational aspects** relating to the study of real problems, avoiding mathematical details.

MANCHESTER 1824

Introduction

Aspects of modern information systems that have come to the fore and challenged computational and statistical analysis have been dynamic complexity, high dimensionality and inherent stochasticity. Recent themes for information reuse and integration highlighted by IRI Keynote speakers have included uncertain computation [36], and anomaly detection in the context of privacy protection [3].

By 2013, the annual worldwide internet protocol (IP) traffic is predicted to be a zettabyte ($2^{70} \sim 10^{21}$ bytes) for which 90% of consumer IP traffic and 60% of mobile IP traffic will be video.

Digital cameras are merging with smart phones, and visual computing applications for computational photography and augmented reality applications are developing rapidly [23], frequently based on information geometric representation and optimization methods.

Geometrization of models of real phenomena have long been known to give valuable insights because of the established value of analytic geometric features, such as a **natural metric, parallelism, perpendicularity, curvature and geodesic curves** that are invariant under the choice of coordinate representation.

In **information geometry** this corresponds to the **invariance of measure functions of probability distributions under changes of parameters.** Interest of geometers is stimulated by novel applications because these can point to new developments in the geometrical structures available.

The **natural information metric provides distances between states and along state trajectories**, thus facilitating optimization strategies.

Information reuse and integration, utilising information theory within information geometry brings important concepts mirroring physical theory of statistical mechanics: eg. **entropy** (ie the ‘mean log probability density’) and its relation to **maximum likelihood methods** for model optimization.

We outline information geometry methodology to provide **natural neighbourhoods** that respect the intrinsic geometry of the space of states, for visualization and monitoring of algorithms, and dynamics of stochastic behaviour trajectories.

1. Computational Information Geometry For Exponential Families Of Distributions

For a random variable $x \in \mathbb{R}^n$, a set $\{p_\theta\}$ of probability density functions with parameters $\theta = \{\theta_i, i = 1, ..., n\}$ is an **exponential family** if the p_θ express as functions $\{C, F_1, ..., F_n\}$ of $x \in \mathbb{R}^n$ and a function φ of $\theta = \{\theta_1, ..., \theta_n\}$ as:

$$p_\theta(x) = e^{\{C(x) + \sum_i \theta_i F_i(x) - \varphi(\theta)\}}.$$

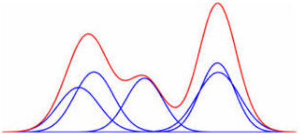
Exponential families of probability density functions and are very important and include Gaussian and gamma distributions. **They admit simple embeddings in** $\mathbb{R}^{n+1}$.

The jMEF package [19], is a Java library with Matlab interface and tutorials to create, process and manage mixtures of exponential families of probability density functions:

http://www.lix.polytechnique.fr/∼nielsen/MEF/

jMEF

A Java library to create, process and manage mixtures of exponential families



What are exponential families?

An exponential family is a generic set of probability distributions that admit the following canonical distribution:

$$p_F(\theta) = \exp\left(<\theta, t(x)> - F(\theta) + k(x)\right)$$

Exponential families are characterized by the log normalizer function *F*, and include the following well-known distributions: Gaussian (generic, isotropic Gaussian, diagonal Gaussian, rectified Gaussian or Wald distributions, lognormal), Poisson, Bernoulli, binomial, multinomial, Laplacian, Gamma (incl. chi-squared), Beta, exponential, Wishart, Dirichlet, Rayleigh, probability simplex, negative binomial distribution, Weibull, von Mises, Pareto distributions, skew logistic, etc.
All corresponding formula of the canonical decomposition are given in the documentation
Mixtures of exponential families provide a generic framework for handling Gaussian mixture models (GMMs also called MoGs for mixture of Gaussians), mixture of Poisson distributions, and Laplacian mixture models as well.

What is jMEF?

jMEF is a Java cross-platform library developped by Vincent Garcia and Frank Nielsen. jMEF allows one to:

- create and manage mixture of exponential families (MEF for short),
- estimate the parameters of a MEF using Bregman soft clustering (equivalent by duality to the Expectation-Maximization algorithm),
- simplify MEFs using Bregman hard clustering (k-means algorithm in natural parameter space),
- define a hierachical MEF using Bregman hierarchical clustering,
- automatically retrieve the *optimal* number of components in the mixture using the hierarchical MEF structure.

A quick overview with slides.
A tutorial for the Matlab interface

Figure: From [19]

Theorems for processes nearly Poisson, or nearly uniform, cf. Arwini and Dodson [1, 16]

Theorem 1
For independent positive random variables with a common probability density function f, having independence of the sample mean and the sample coefficient of variation is equivalent to f being the gamma distribution.

Theorem 2
Every neighbourhood of a Poisson process contains a neighbourhood of processes subordinate to gamma probability density functions.

Theorem 3
Every neighbourhood of a uniform process contains a neighbourhood of processes subordinate to log-gamma probability density functions.

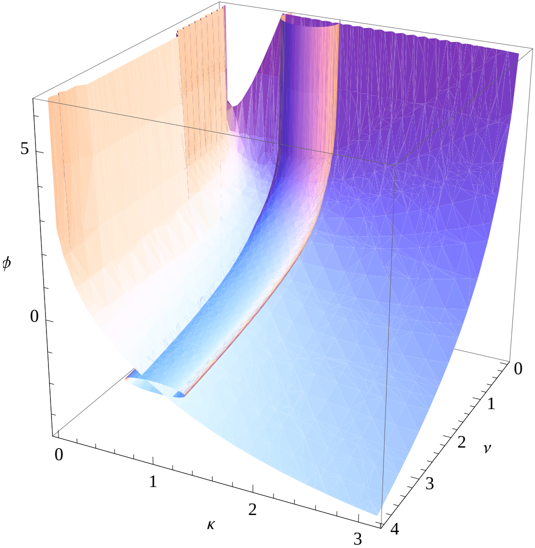


Figure: *A tubular neighbourhood of the curve of all exponential random processes, in the curved surface embedding in $\mathbb{R}^3$ of the 2-manifold of gamma distributions. From Arwini and Dodson 2008 [1].*

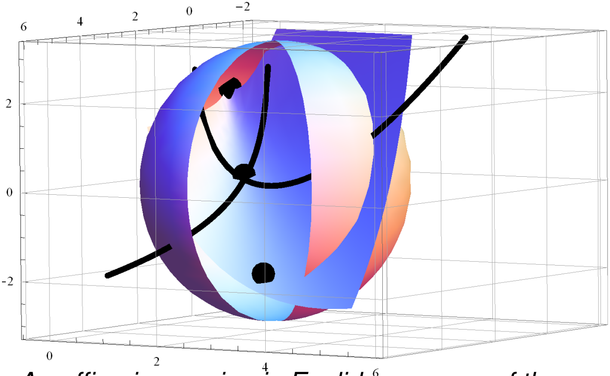


Figure: *An affine immersion in Euclidean space of the curved surface of log-gamma probability densities on* $[0, 1]$, *which includes the uniform distribution as the special case with parameters* $\nu = 1$, $\tau = 1$*. The black curves in the surface represent the log-gamma distributions with* $\nu = 1$ *and* $\tau = 1$, *respectively and the spherical neighbourhood is centred on their intersection. The other two points are for the two cases* $(\nu = 0.1, \tau = 0.289)$ *and* $(\nu = 2.75, \tau = 2.24)$. *From Dodson 2012 [16].*

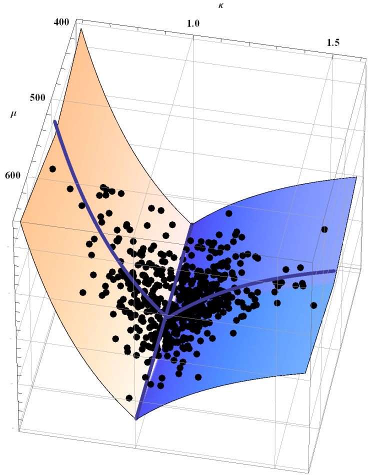


Figure: *Distances in space of gamma models. Surface height represents upper bounds on distances from* $(\mu, \kappa) = (511, 1)$*. Data points from simulations of Poisson random sequences of length* 100000 *for expected separation* $\mu = 511$*. In the limit as the sequence length tends to infinity and the element abundance tends to zero we expect the gamma parameter* $\kappa \to 1$*. From Dodson 2012 [16].*

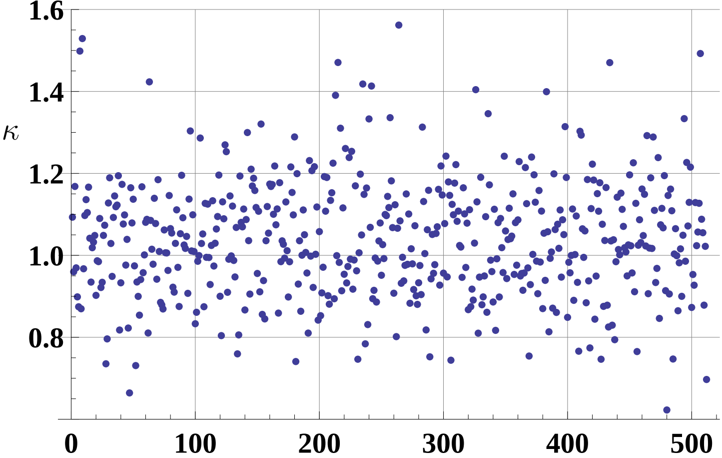


Figure: *Maximum likelihood gamma parameter κ fitted to separation statistics for simulations of Poisson random sequences of length* 100000 *for an element with expected parameters* $(\mu, \kappa) = (511, 1)$. *These simulations used the pseudorandom number generator in Mathematica [32]. From Dodson 2012 [16].*

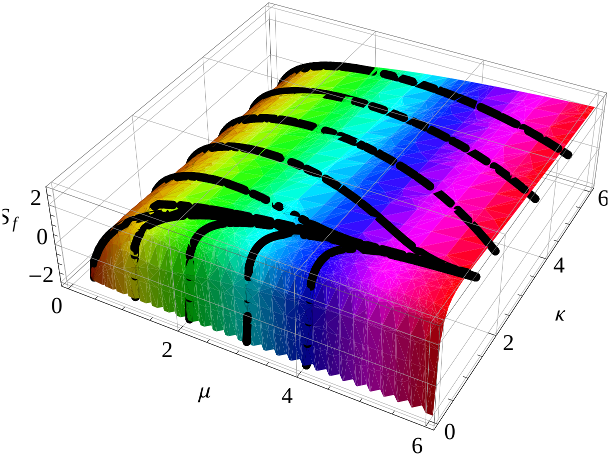


Figure: *Entropy function S_f for the gamma family of distributions with entropy gradient flow and integral curves as a surface. On the surface, the dashed asymptote at $\kappa = 1$ is the exponential case of maximum disorder. From Dodson 2010 [14].*

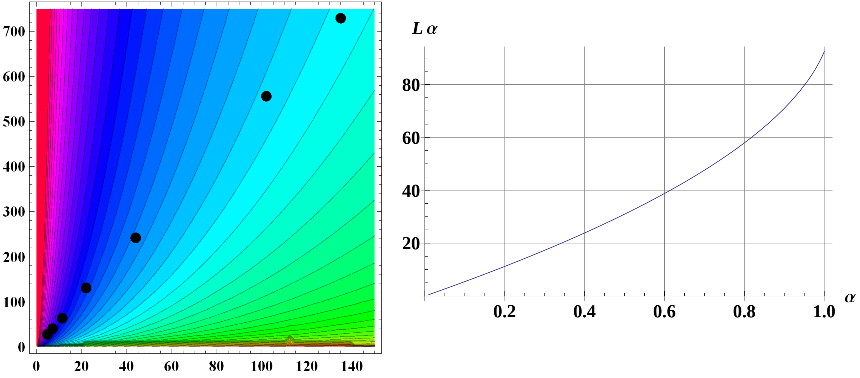


Figure: *Constrained degradation of order: Gamma entropy contour plot (left) with data on progression under noise for surface area distributions of 3D BCC Voronoi cells from Lucarini [21] as perfect crystallinity degenerated from large κ, ν towards $\kappa = 16$, the theoretical Poisson Voronoi limit. Curve gives information length L_α along this disordering trajectory as noise amplitude α increased. The limit is not total disorder, $\kappa = 1$, because the Voronoi cell structure is a constraint. Cf. Dodson 2012 [16].*

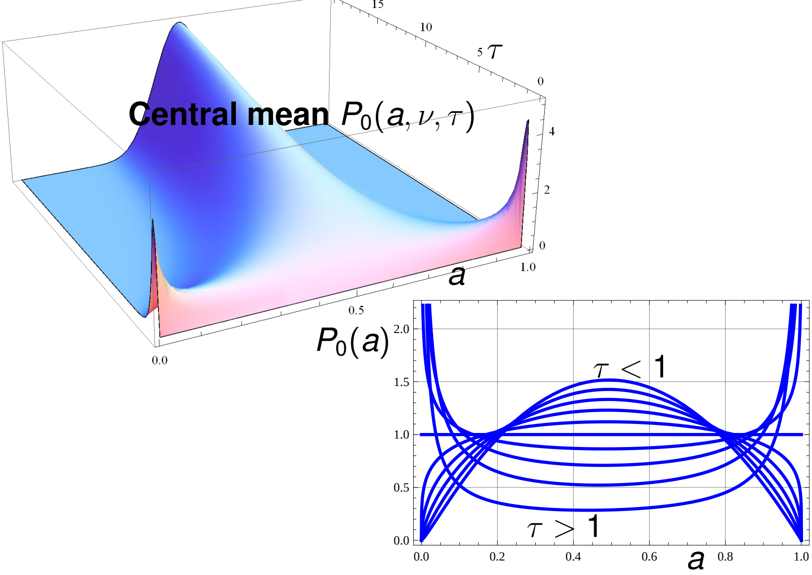


Figure: *The log-gamma family of probability densities $P_0(a, \nu, \tau)$ on $[0, 1]$ as a surface for the case of central mean $E_0(a) = \frac{1}{2}$. From Arwini and Dodson 2008 [1].*

Computational Information Geometry:
2. Interactive Mathematica Notebooks

For analytic geometry, numerical procedures and graphics, we used the computational algebra package *Mathematica* by Wolfram [32] in a wide range of applications of common univariate and bivariate probability density functions in the books [1, 16].

These interactive Mathematica notebooks are available free for download from the author's webpage

http://www.maths.manchester.ac.uk/∼kd
/mmaprogs/InfoGeomMMANotebooks/

Figure: From [32]

Index of /~kd/mmaprogs/InfoGeomMMANotebooks

Name	Last modified	Size	Description
Parent Directory		-	
0ReadMeInformationGeometry.nb	20-Sep-2005 18:49	613K	
AlphaGeomBivNormal.nb	24-Sep-2010 17:30	1.1M	
AlphaGeomFreund.nb	24-Sep-2010 17:27	407K	
AlphaGeomGamma.nb	24-Sep-2010 17:20	67K	
AlphaGeomGaussian.nb	24-Sep-2010 17:28	114K	
Amino&RandomDist.nb	26-Sep-2010 19:59	547K	
CoverGraphic.nb	30-Apr-2008 18:23	759K	
EntropyGradientFlows.nb	26-Sep-2010 20:12	8.6M	
FreundInfoGeom.nb	24-Sep-2010 17:14	491K	
GammaGradFlow.nb	24-Sep-2010 17:18	2.5M	
GaussianGeodesics.nb	24-Sep-2010 17:21	432K	
GaussianInfoGeom.nb	24-Sep-2010 17:15	122K	
LogGamma.nb	24-Sep-2010 17:19	464K	
LogGammaNuTau.nb	24-Sep-2010 17:40	1.0M	
LogGammaRandomWalk.nb	24-Sep-2010 17:57	2.5M	
McKayDistEnergyEstimates.nb	24-Sep-2010 17:56	350K	
MckayInfoGeom.nb	24-Sep-2010 17:31	294K	
NhdIndepBivGaussian.nb	24-Sep-2010 17:23	22K	
NhdIndepExp.nb	25-Sep-2010 06:24	659K	
NhdRandomGamma.nb	24-Sep-2010 17:33	296K	
PoissonGaps.nb	22-Oct-2010 18:25	3.9M	
PrimeGammaIn100000s.nb	24-Sep-2010 17:50	1.1M	
ZetaZeros.nb	24-Sep-2010 17:41	46K	
ZetaZeros1.txt	24-Sep-2010 17:47	1.7M	

Figure: From [1, 16]

3. Computational Geometry Algorithms Library

Software available at CGAL [11] is successful for identifying implementation issues of computational geometry methods.

This Open Source Project offers a C^{++} library of algorithms for common problems in computational geometry: http://www.cgal.org/

Computational Geometry Algorithms Library

The goal of the CGAL Open Source Project is to provide *easy access to efficient and reliable geometric algorithms* in the form of a C++ library. CGAL is used in various areas needing geometric computation, such as: computer graphics, scientific visualization, computer aided design and modeling, geographic information systems, molecular biology, medical imaging, robotics and motion planning, mesh generation, numerical methods... More on the projects using CGAL web page.

The Computational Geometry Algorithms Library (CGAL), offers data structures and algorithms like triangulations (2D constrained triangulations and Delaunay triangulations in 2D and 3D, periodic triangulations in 3D), Voronoi diagrams (for 2D and 3D points, 2D additively weighted Voronoi diagrams, and segment Voronoi diagrams), polygons (Boolean operations, offsets, straight skeleton), polyhedra (Boolean operations), arrangements of curves and their applications (2D and 3D envelopes, Minkowski sums), mesh generation (2D Delaunay mesh generation and 3D surface and volume mesh generation, skin surfaces), geometry processing (surface mesh simplification, subdivision and parameterization, as well as estimation of local differential properties, and approximation of ridges and umbilics), alpha shapes, convex hull algorithms (in 2D, 3D and dD), search structures (kd trees for nearest neighbor search, and range and segment trees), interpolation (natural neighbor interpolation and placement of streamlines), shape analysis, fitting, and distances (smallest enclosing sphere of points or spheres, smallest enclosing ellipsoid of points, principal component analysis), and kinetic data structures.

All these data structures and algorithms operate on geometric objects like points and segments, and perform geometric tests on them. These objects and predicates are regrouped in CGAL Kernels.

Finally, the Support Library offers geometric object generators and spatial sorting functions, as well as a matrix search framework and a solver for linear and quadratic programs. It further offers interfaces to third party software such as the GUI libraries Qt, Geomview, and the Boost Graph Library.

License

CGAL is distributed under a dual-license scheme. CGAL can be used together with Open Source software free of charge. Using CGAL in other contexts can be done by obtaining a commercial license from GeometryFactory. For more details see the License page.

Figure: From [11]

First order inhomogeneous rate processes

Let $l_t(a)$ represent the frequency at the a-cohort, then we have

$$N(t) = \int_0^\infty l_t(a)\,da \quad \text{and} \quad P_t(a) = \frac{l_t(a)}{N(t)}$$
$$\frac{dl_t(a)}{dt} = -a l_t(a) \quad \text{so} \quad l_t(a) = l_0(a)e^{-at}.$$

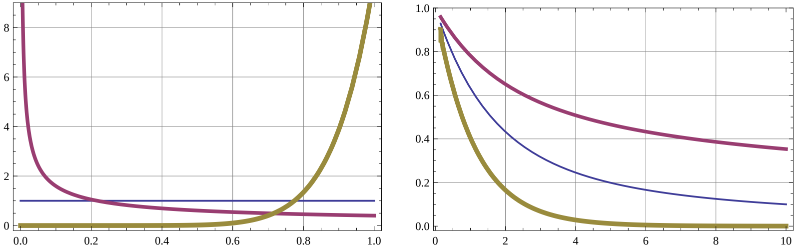


Figure: *Initial log-gamma densities $P_0(a, \nu, 1)$ are shown in the left panel for the uniform density $\tau = \nu = 1$, and also for $\tau = 1$ with $\nu = 0.4$ and 10. The right panel shows the corresponding fractional decline with time of the population $N(t)/N(0)$ for these initial densities. From Dodson 2012 [16].*

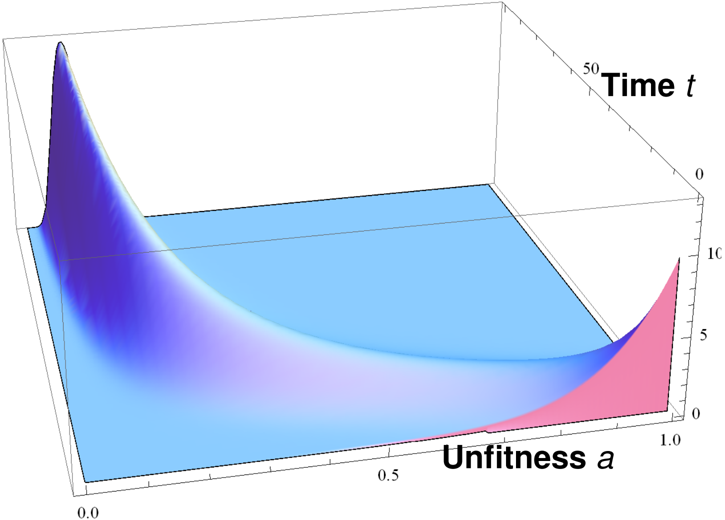


Figure: *The evolution of the probability density $P_t(a, 10, 1)$ of unfitness, under a first order inhomogeneous rate process, from an initial log-gamma density with $\nu = 10, \kappa = 1$. Showing how the initially high population unfitness is reduced over time.*
From Dodson 2012 [16].

4. Computational Information Geometry: Interpreting data

Extracting and using rich information from massive data sets is a serious challenge. Visual data abounds, so computer vision and computer graphics are increasingly relying on machine learning and information-theoretic methods.

Computational information geometry effectively performs high-fidelity data analysis using the language and invariant features of geometry, Rabin et al. [26, 27].

This allows mapping of the data in a suitable space for efficient processing and retrieval of intrinsic information using coordinate-free operations on the data.

Image Processing

The geometrization of statistics has provided novel algorithms for manipulating statistical models such as Gaussian mixture models, Nielsen et al. [25], Takatsu [30], that are commonly used in image processing.

An image pixel at position (x, y) with colour attributes (red, green, blue) is embedded into a 5D space so that a 2D colour image is interpreted as a 5D spatial point cloud.

We then seek a compact generative statistical representation of the image point set. Such statistical methods are useful for explaining human cognitive and learning skills, Tenenbaum et al. [31], and analyzing emerging phenomena of complex systems using hierarchical Bayesian models.

Amino Acid Clustering in Genome

Molecular biologists have large data sets of amino acid sequences. Cai et al. analysed [4] for each of the 20 amino acids X, the statistics of spacings between consecutive occurrences of X within the *Saccharomyces cerevisiae* genome, from 6294 protein chains with sequence lengths up to $n = 4092$. The spacing distributions were well approximated by gamma distributions.

Expect, with 20 types of amino acids distributed at different abundances along a protein chain, that some would more clustered ($\kappa < 1$) and others more evenly spread ($\kappa > 1$) compared to Poisson process—which has exponential spacings and $\kappa = 1$. In fact none of the amino acids was distributed with such a low variance, all clustered and so had greater variance than would result from a Poisson process.

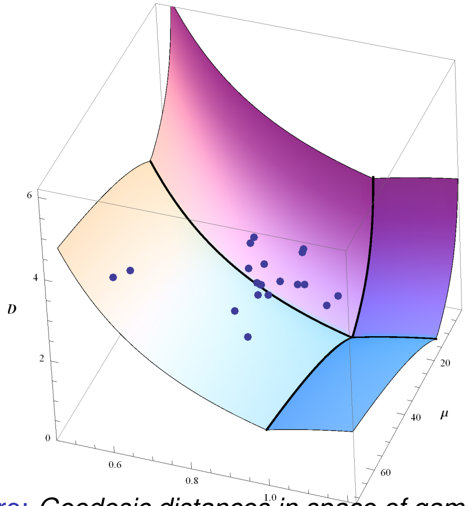
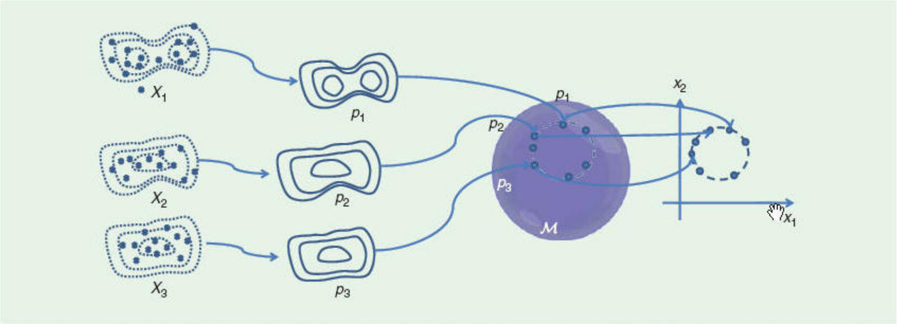


Figure: *Geodesic distances in space of gamma models, [1, 16]. Surface of upper bounds on distance D from grand mean point* $(\mu, \kappa) = (18, 1)$, *the Poisson case. The 20 data points are for the amino acid sequences from a large database. All amino acids show clustering by lying to the left of the Poisson random line* $\kappa = 1$.

Dimensionality reduction

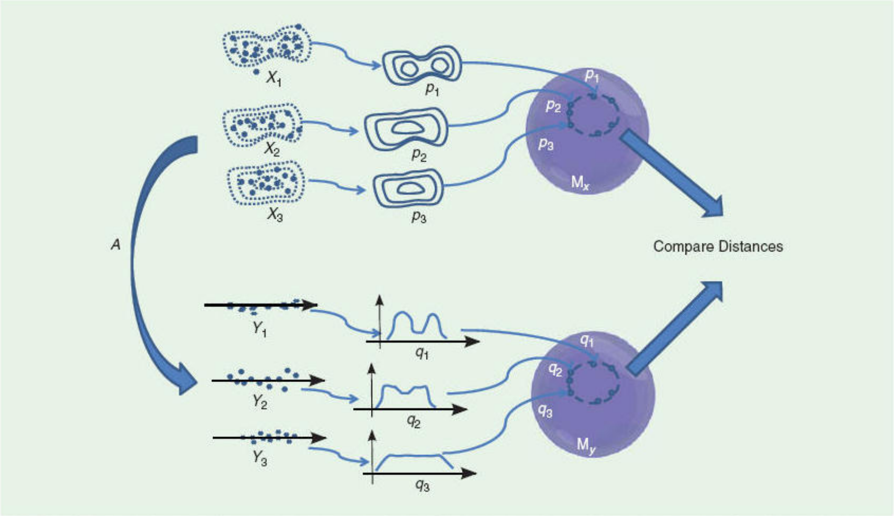
In many real world problems we encounter high dimensionality in large data sets and often do not have the luxury of knowing the optimal net probability density function family for the features represented in the data. A fundamental problem in the identification of probability densities from large multidimensional data sets, that of efficient dimensionality reduction, was addressed by Carter et al. [7, 8, 9, 10].

They used information geometry to obtain nearest neighbour distances by means of geodesic estimates subordinate to a Fisher information metric giving a non-parametric embedding (FINE). Also, data dimensionality reduction using information-preserving component analysis (IPCA) and information-maximizing component analysis (IMCA).



[FIG4] **FINE: first, a PDF p_i is estimated for each data set X_i. Then, an information-geometric metric is used to learn the geometry of the manifold of PDFs from pairwise distance measurements. Finally, a Euclidean embedding from the manifold $\mathcal{M}_x$ to $\mathbb{R}^d$ is obtained, associating each original data set X_i with its embedded point in Euclidean space x_i.**

Figure: From Carter 2011 [10]



[FIG6] IPCA/IMCA: first, a PDF p_i is estimated for each data set X_i. Simultaneously, a PDF q_i is estimated for each data set $Y_i = AX_i$. Then, an information-geometric metric is used to learn the geometry of the manifold $\mathcal{M}_x$ of PDFs p_is and manifold $\mathcal{M}_y$ from PDFs q_is from pairwise distance measurements. Finally, an objective is calculated to compare the geometry of the two manifolds $\mathcal{M}_x$ and $\mathcal{M}_y$. For IPCA, we consider the minimization of the sum of squared differences between each pairwise distance on $\mathcal{M}_x$ and its equivalent in $\mathcal{M}_y$. For IMCA, we consider the maximization of the sum of distances in $\mathcal{M}_y$.

Figure: From Carter 2011 [10]

This method takes account of the true curved geometry of the data set, rather than displaying as uncurved in a Euclidean geometry cf.Carter [8], Figure 3.2. The significance is that the non-obvious global topology of frequency connectivity in the data is revealed by the geodesics. An illustration using router traffic on subdomains (comp., rec., sci., talk.) of the Abilene network showed how anomalous behavior unseen by local methods could be picked up through dimensionality changes cf. [8], Figure 3.10.

In document classification, the information metric approach outperforms standard PCA and Euclidean embeddings (LEM), Carter et al. [7], and it outperforms traditional approaches to video indexing and retrieval with real world data, Chen and Hero [12].

3.4.2 Network Anomaly Detection

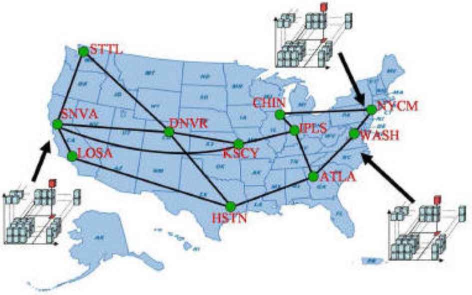


Figure 3.9: Map of Abilene router network

As illustrated in Figure 3.9, the Abilene Network is the set of routers which is the backbone of the '.edu' network. When an anomaly occurs on the network, there are changes in the correlation between traffic traces at different points in the network,

Figure: From Carter 2009 [8]

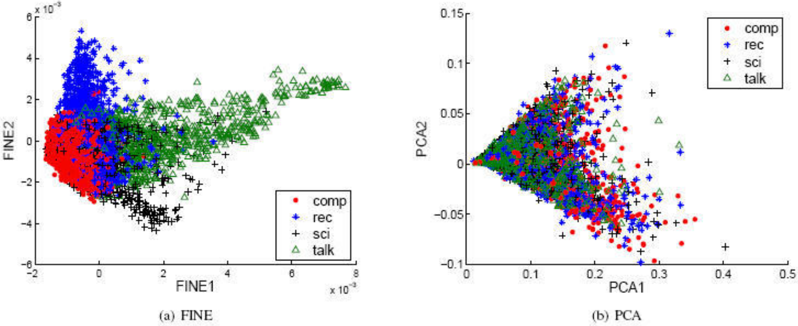
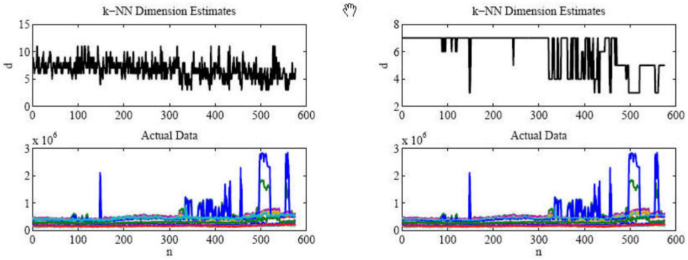


Fig. 1. 2-dimensional embeddings of 20Newsgroups data. The data displays some natural clustering, in the information based embedding, while the PCA embedding does not distinguish between classes.

Figure: From Carter et al. 2008 [7]



(a) Before Smoothing (b) After Smoothing

Figure 3.10: Neighborhood smoothing applied to Abilene Network traffic data dimension estimation results. Anomalous activity is preserved and more easily observed.

[11, 12]. Specifically, when only a few of the routers contribute disproportionably large amounts of traffic, the intrinsic dimension of the entire network decreases. Using

Figure: From Carter et al. 2007 [6] and 2009 [8]. Near n=244 was revealed anomalous increased traffic from a single IP address, not visible in the raw data.

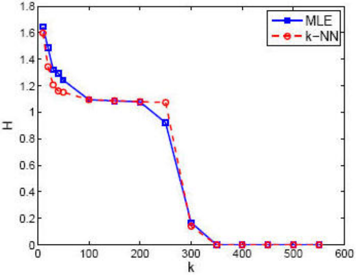


Figure 3.12: The entropy of the local dimension estimates changes as a function of neighborhood size k. As k increases to the size of the differing regions ($k = 200$ samples each), the entropy becomes constant and the data is properly clustered. As the neighborhood incorporates samples from differing manifolds, the entropy decrease until all points estimate at the same value ($k = 350$).

Figure: From Carter 2009 [8]

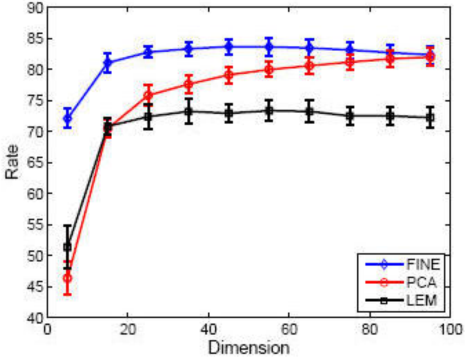


Fig. 2. Classification rates for low-dimensional embedding using different methods for dimensionality reduction

Figure: From Carter et al. 2008 [7]

Anomaly Detection

Information geometric methods extend to anomaly detection using large sample size data sets derived from an underlying probability distribution of unknown parameterization. A comparison of relevant information theoretic measures that are intrinsic to information geometry can be found in Lee and Xiang [20].

They used information-theoretic measures: **entropy, conditional entropy, relative conditional entropy, information gain, and information cost for anomaly detection.** Described the characteristics of an audit data set, identify appropriate anomaly detection model to be built, and explain the performance of the model. Illustrated with case studies on Unix system call data, BSM data, and network tcpdump.

Cf. also Gu et al. Measuring **intrusion detection capability**: An information theoretic approach: http://dl.acm.org/citation.cfm?id=1128834

An important problem in naval studies is the extraction and analysis of seasonal information from large data sets of wave height measurements obtained by satellite observations.

A new approach based on information geometrical techniques has yielded useful results, Galanis et al. [18].

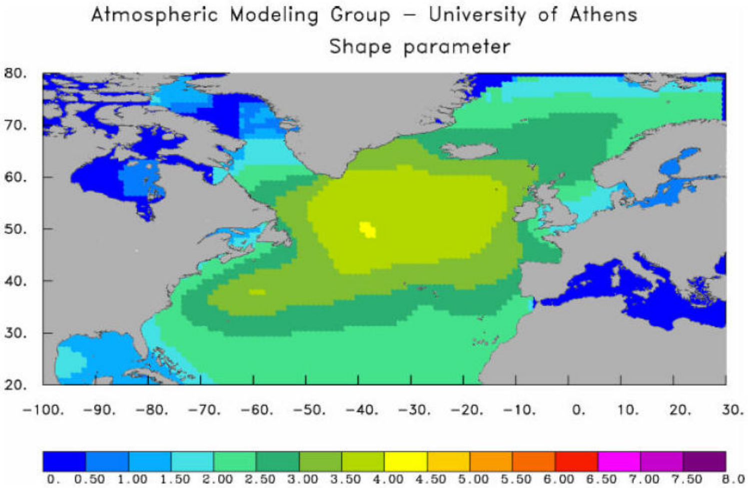
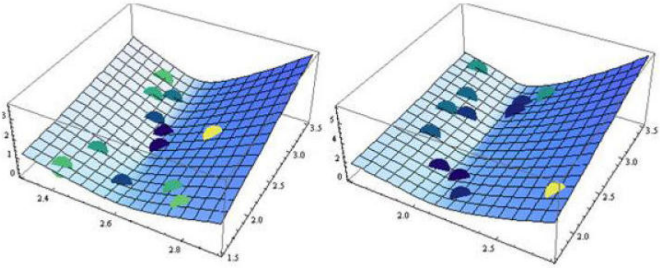


Figure 4. The shape parameter of the Weibull distributions that fit to the significant wave height satellite data over the North Atlantic Ocean for the months March-May.

Figure: From Galanis et al. 2012 [18]



(a) (b)

Figure 26. The statistical manifolds formed by the monthly values of the satellite records (a) and WAM outputs (b) as elements of the non-Euclidean space of all Weibull distributions. A classical "BlueGreenYellow" color palette has been used depending on their approximate divergence from annual averages

Figure: From Galanis et al. 2012 [18]. Such methods lend themselves to automated algorithms to optimize efficiency.

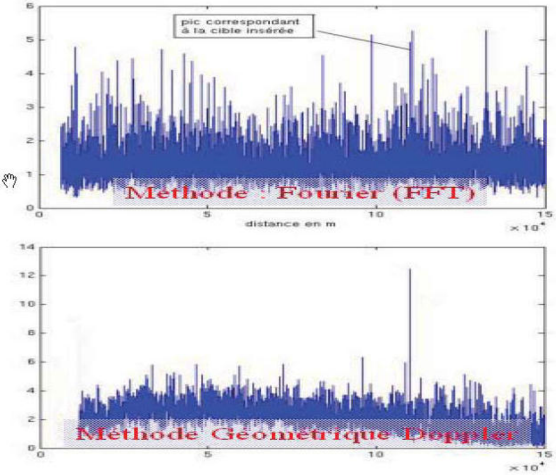


Figure 6. Doppler detection with sea clutter (x axis : range) for one azimuth. (at top) classical method based on FFT, (at bottom) new detector

Figure: From Barbaresco 2008 [2]. The information geometric method correctly and unambiguously detects the anomalous signal.

Target tracking has been shown to be enhanced using information geometric methods by Cheng et al. [13]:

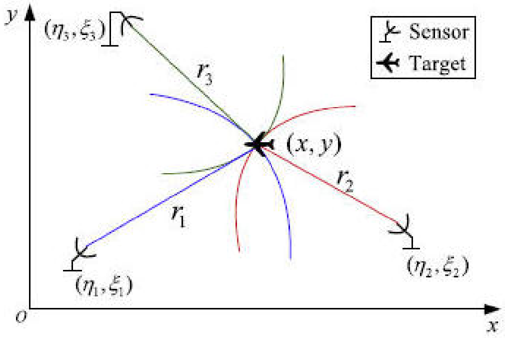
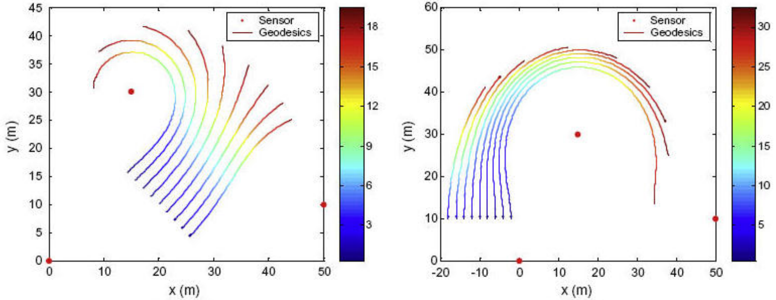


Fig. 6. Example of sensor network of three range only sensors for target localization.



(a) Case1 (b) Case2

Fig. 11. Bundle of geodesics of identical FIDs with parallel initial tangent vectors for the sensor network of three ranges-only sensors. Case 1 illustrates the situation of divergent geodesic bundle and a convergent bundle example is shown in Case 2.

Figure: From Cheng et al. 2012 [13]. Bundles of information geodesics of the same length in the statistical manifold for three sensors.

From Cheng et al. 2012 [13]. Affine immersion of the statistical manifold for three sensors.

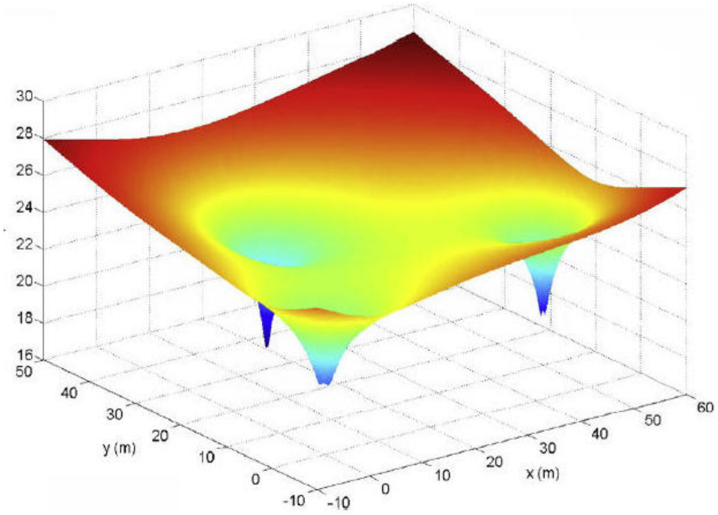


Fig. 16. Affine immersion for the manifold of three ranges-only sensor network in Example 3.

References

[1] Khadiga Arwini and C.T.J. Dodson. *Information Geometry Near Randomness and Near Independence*. Lecture Notes in Mathematics, Springer-Verlag, New York, Berlin, 2008. http://www.springer.com/mathematics/geometry/book/978-3-540-69391-8

[2] F. Barbaresco. Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices and information geometry. IEEE Radar Conference Rome, 26-30 May 2008. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4720937

[3] E. Bertino. Protecting Information Systems from Insider Threats - Concepts and Issues. IRI 2011 Keynote Lecture.

[4] Y. Cai, C.T.J. Dodson, O. Wolkenhauer and A.J. Doig. Gamma Distribution Analysis of Protein Sequences shows that Amino Acids Self Cluster. *J. Theoretical Biology* 218, 4 (2002) 409-418.

[5] E. Candes and M. Wakin. An introduction to compressive sampling. *IEEE Trans. Signal Processing*, 25 (2008) 21-30.

[6] K.M. Carter, A.O. Hero III and R. Raich. De-biasing for intrinsic dimension estimation. In Proc. 2007 IEEE Workshop on Statistical Signal Processing, pp 601-605.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4301329

[7] K.M. Carter, R. Raich and A.O. Hero III. FINE: Information embedding for document classification. In Proc. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, March 2008.
http://tbayes.eecs.umich.edu/kmcarter/fine_doc

[8] K.M. Carter. Dimensionality reduction on statistical manifolds. PhD thesis, University of Michigan, 2009.
http://tbayes.eecs.umich.edu/kmcarter/thesis

[9] K.M. Carter, R. Raich,W.G. Finn and A.O. Hero III. Fisher information nonparametric embedding. *IEEE Trans. Pattern. Anal. Mach. Intell.* 31 (2009) 2093Ű2098.
http://www.computer.org/portal/web/csdl/doi/10.1109/TPAMI.2009.67

[10] K.M. Carter, R. Raich,W.G. Finn and A.O. Hero III. Information-geometric dimensionality reduction. *IEEE Signal Processing Mag.* 99 (2011) 89-99.

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5714382

[11] CGAL: Computational Geometry Algorithms Library. http://www.cgal.org/

[12] X. Chen and A. Hero. Fisher Information Embedding for Video Indexing and Retrieval. SPIE Electronic Imaging Conference, San Jose, 2011. http://tbayes.eecs.umich.edu/xhen/indexing_spie2011

[13] Y. Cheng, X. Wang, M. Morelande and W. Moran. Information geometry of target tracking sensor networks. *Information Fusion* (2012) 10.1016/j.inffus.2012.02.005. http://www.sciencedirect.com/science/article/pii/S1566253512000310

[14] C.T.J. Dodson. On the entropy flows to disorder. In C. H. Skiadas and I. Dimotikalis, (Eds.), *Chaotic Systems: Theory and Applications* World Scientific, Singapore, 2010 pp 75-84. http://arxiv.org/abs/0811.4318

[15] C.T.J. Dodson. An inhomogeneous stochastic rate process for evolution from states in an information geometric neighbourhood of uniform fitness. Invited paper at 3rd

Conference on Information Geometry and its Application, Leipzig 2-6 August 2010. Cf also: http://arxiv.org/abs/1001.4177v1

[16] C.T.J. Dodson. Some illustrations of information geometry in biology and physics. In *Handbook of Research on Computational Science and Engineering: Theory and Practice* Eds. J. Leng, W. Sharrock, IGI-Global, Hershey, PA, 2012, pp 287-315.
http://www.maths.manchester.ac.uk/~kd/PREPRINTS/DodsonCSE.pdf
http://www.maths.manchester.ac.uk/~kd/mmaprogs/InfoGeomMMANote

[17] D. Donoho. Compressed sensing. *IEEE Trans. Inf Theory*, 52 (2006) 1289-1306.

[18] G. Galanis, P.C. Chu, G. Kallos, Yu-Heng Kuo and C.T.J. Dodson. Wave Height Characteristics in North Atlantic Ocean: A new approach based on statistical and geometrical techniques. *Stochastic Environmental Research and Risk Assessment (SERRA)* 26, 1 (2012) 83-103.

[19] jMEF: A Java library to create, process and manage mixtures of exponential families.
http://www.lix.polytechnique.fr/~nielsen/MEF/

[20] W. Lee and D. Xiang. Information-theoretic measures for anomaly detection. In Proc. **IEEE Symposium Security and Privacy** (2001) pp 130-143.
http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=924294
Cf. also Gu et al. Measuring intrusion detection capability: An information-theoretic approach:
http://dl.acm.org/citation.cfm?id=1128834

[21] Valerio Lucarini. Three-dimensional Random Voronoi Tessellations: From Cubic Crystal Lattices to Poisson Point Processes. *Journal Statistical Physics* 134, 1 (2009) 1047-1062.
http://www.springerlink.com/content/221l62x008444723/

[22] N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi. Information geometry of U-boost and Bregman divergence. *Neural Computation*, 16 (2004) 1437-1481.

[23] F. Nielsen. **Visual Computing: Geometry, Graphics, and Vision** Charles River Media, ISBN: 1-58450-427-7, 2005.

[24] F. Nielsen, J-D. Boissonnat and R. Nock On Bregman Voronoi diagrams. In Proc ACM-SIAM Symposium on Data Mining, 2007, 746-755.

ftp://ftp-sop.inria.fr/geometrica/boissonnat/Papers/bregman.pdf

[25] F. Nielsen and S. Boltz. The Burbea-Rao and Bhattacharyya Centroids. *IEEE Transactions on Information Theory* 57, 8 (2011) 5455-5466.

[26] J. Rabin and G. Peyré. Wasserstein Regularization of Imaging Problem. In Proc. ICIP'11 (2011) http://hal.archives-ouvertes.fr/hal-00591279

[27] J. Rabin, G. Peyré, J. Delon and M. Bernot. Wasserstein Barycenter and its Applications to Texture Mixing. In Proc. SSVM'11 (2011). http://hal.archives-ouvertes.fr/hal-00476064/

[28] F. Reverter and J.M. Oller. Computing the Rao distance for Gamma distributions. *Journal of Computational and Applied Mathematics* 157, 1, (2003) 155-167.
http://www.sciencedirect.com/science/article/pii/S037704270300387X#sec1

[29] H. Ruiz, I.H. Jarman, J.D. Martin and P.J. Lisboa. The role of Fisher information in primary data space for neighbourhood mapping. In Proc. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium 27-29 April 2011.

http://www.i6doc.com/en/livre/?GCOI=28001100817300

[30] Asuka Takatsu. On Wasserstein geometry of the space of Gaussian measures. (2011) http://arxiv.org/pdf/0801.2250.pdf

[31] J.B. Tenenbaum, C. Kemp, T.L. Griffiths, and N.D. Goodman. How to Grow a Mind: Statistics, Structure, and Abstraction *Science* 331(6022), (2011) 1279-1285.

[32] Wolfram, S. (1996). *The Mathematica Book* 3rd edition, Cambridge University Press, Cambridge.

[33] S. Wu and S. Amari. Conformal Transformation of Kernel Functions: A Data-Dependent Way to Improve Support Vector Machine Classifiers. *Neural Processing Letters* 15 (2002) 59-67.

[34] S. Wu, S. Amari and H. Nakahara. Information processing in a neuron ensemble with the multiplicative correlation structure. *Neural Networks* 17 (2004) 205-214.

[35] K.S. Xu, M. Kliger, A.O. Hero III. Identifying spammers by their resource usage patterns. In Proc. Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS), 2010. http://tbayes.eecs.umich.edu/xukevin/spam_ceas2010

[36] L.A. Zadeh. A Note on Z-numbers. IRI 2011 Keynote Lecture, *Information Sciences* 181, 14 (2011) 2923-2932.
http://www.sciencedirect.com/science/article/pii/S0020025511001046