

STATISTICS ON NON-SMOOTH SPACES: A SELECTION OF LITERATURE

TOM M. W. NYE

This document gives a selective literature review covering statistics on non-smooth spaces, focusing mainly on statistics for data sets of phylogenetic trees. This list has been prepared for use by students on the International PhD course in Nonlinear Statistics, Copenhagen, June 2017. The references are ordered roughly as they appear in the lectures.

1. INTRODUCTION: TREE-SPACE AND FUNDAMENTAL GEOMETRY

Billera, L., Holmes, S., & Vogtmann, K. (2001). Geometry of the space of phylogenetic trees. *Adv. Appl. Math.*, 27, 733–767.

The original paper parametrizing phylogenetic trees and proving that the space is CAT(0).

Skwerer, S., Bullitt, E., Huckemann, S., Miller, E., Oguz, I., Owen, M., ... Marron, J. (2014). Tree-oriented analysis of brain artery structure. *J. Math. Imaging Vis.*, 50(1-2), 126–143.

Analysis of brain artery data.

Feragen, A., Owen, M., Petersen, J., Wille, M., Thomsen, L., Dirksen, A., & de Bruijne M. (2013). Tree-space statistics and approximations for large-scale analysis of anatomical trees. In *23rd biennial International Conference on Information Processing in Medical Imaging (IPMI)*.

Analysis of lung image data.

Bridson, M. R., & Haefliger, A. (2011). *Metric spaces of non-positive curvature* (Vol. 319). Springer-Verlag.

Rigorous description of CAT(0) spaces and other geometrical background.

Miller, E., Owen, M., & Provan, J. S. (2015). Polyhedral computational geometry for averaging metric phylogenetic trees. *Adv. Appl. Math.*, 68, 51–91.

Definition of orthant spaces, together with lots of material relevant to lecture 2.

Owen, M., & Provan, J. S. (2011). A fast algorithm for computing geodesic distances in tree space. *IEEE ACM T. Comput. Bi.*, 8(1), 2–13.

Algorithm for constructing geodesics in tree-space.

Ardila, F., Owen, M., & Sullivant, S. (2012). Geodesics in CAT(0) cubical complexes. *Adv. Appl. Math.*, 48(1), 142–163.

Computation of geodesics on cubical complexes – a generalization of the algorithm on tree-space.

Devadoss, S., & Petti, S. (2016). A space of phylogenetic networks. *arXiv preprint arXiv:1607.06978*.

Geometry of a space of phylogenetic networks.

Gavryushkin, A., & Drummond, A. J. (2016). The space of ultrametric phylogenetic trees. *J. Theor. Bio.*, *403*, 197–208.

A space of time-like trees.

Maclagan, D., & Sturmfels, B. (2015). *Introduction to tropical geometry* (Vol. 161). American Mathematical Soc.

Tropical geometry for phylogenetic trees.

Miller, E. (2015). Fruit flies and moduli: interactions between biology and mathematics. *Not. Am. Math. Soc.*, *62*(10), 1178–1184.

Descriptive article covering statistics on non-smooth spaces.

Groisser, D., Jung, S., Schwartzman, A., et al. (2017). Geometric foundations for scaling-rotation statistics on symmetric positive definite matrices: minimal smooth scaling-rotation curves in low dimensions. *Elec. J. of Stat.*, *11*(1), 1092–1159.

Non-smooth statistics for covariance matrices.

2. DESCRIBING VARIATION: MEAN, VARIANCE AND CORRELATION

Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In A. Pascal, T. Coulhon, & A. Grigor'yan (Eds.), *Heat kernels and analysis on manifolds, graphs, and metric spaces* (pp. 357–390). Providence, Rhode Island: American Mathematical Society.

Existence of Fréchet means and algorithms for calculation.

Bačák, M. (2014). Computing medians and means in Hadamard spaces. *SIAM J. Optimiz.*, *24*(3), 1542–1566.

Algorithms for computing the Fréchet mean in CAT(0) spaces.

Hotz, T., Huckemann, S., Le, H., Marron, J., Mattingly, J., Miller, E., ... Skwerer, S. (2013). Sticky central limit theorems on open books. *Ann. Appl. Probab.*, *23*(6), 2238–2258.

Central limit theorem for the open book.

Barden, D., Le, H., Owen, M., et al. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab.*, *18*(25), 1–25.

Central limit theorem for tree-space.

Willis, A. (2016). Confidence sets for phylogenetic trees. *arXiv preprint arXiv:1607.08288*.

Tangent-space approximation for tree-space.

Nye, T. M. (2011). Principal components analysis in the space of phylogenetic trees. *Ann. Stat.*, 39(5), 2716–2739.

Tree-space PCA by firing geodesics.

Feragen, A., Owen, M., Petersen, J., Wille, M., Thomsen, L., Dirksen, A., & de Bruijne M. (2013). Tree-space statistics and approximations for large-scale analysis of anatomical trees. In *23rd biennial International Conference on Information Processing in Medical Imaging (ipmi)*.

‘Empirical’ PCA using geodesic segments.

Nye, T. (2014). An algorithm for constructing principal geodesics in phylogenetic treespace. *IEEE ACM T. Comput. Bi.*, 11(2), 304–315.

Unconstrained PCA using geodesic segments on tree-space.

Nye, T. M. W., Tang, X., Weyenberg, G., & Yoshida, R. (2016). Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *arXiv preprint arXiv:1609.03045*.

Higher-order PCA using a type of ‘minimal surface’.

Penneç, X. (2016). Barycentric subspace analysis on manifolds. *arXiv preprint arXiv:1607.02833*.

Higher order PCA for manifold-valued data.

Lin, B., Sturmfels, B., Tang, X., & Yoshida, R. (2015). Convexity in tree spaces. *arXiv:1510.08797*.

Convexity in tree-space, with material on tropical geometrical view-point.

Lubiw, A., Maftuleac, D., & Owen, M. (2016). Shortest paths and convex hulls in 2d complexes with non-positive curvature. *arXiv preprint arXiv:1603.00847*.

Convexity in tree-space, with lots of counter-examples.

3. DISTRIBUTIONS AND STOCHASTIC PROCESSES ON TREE-SPACE

Weyenberg, G., Huggins, P. M., Schardl, C. L., Howe, D. K., & Yoshida, R. (2014). KDE-trees: non-parametric estimation of phylogenetic tree distributions. *Bioinformatics*, 30(16), 2280–2287.

Application of Gaussian kernels in tree-space.

Nye, T., & White, M. (2014). Diffusion on some simple stratified spaces. *J. Math. Imaging Vis.*, *50*, 115–125.

Brownian motion on cones and metric graphs.

Brin, M., & Kifer, Y. (2001). Brownian motion, harmonic functions and hyperbolicity for Euclidean complexes. *Math. Z.*, *237*(3), 421–468.

Existence and properties of Brownian motion on 2d complexes.

Enriquez, N., & Kifer, Y. (2001). Markov chains on graphs and Brownian motion. *J. Theor. Probab.*, *14*(2), 495–510.

Convergence of random walks to Brownian motion on metric graphs.

Nye, T. M. W. (2015). Convergence of random walks to Brownian motion in phylogenetic tree-space. *arXiv preprint arXiv:1508.02906*.

Convergence of random walks to Brownian motion.

E-mail address: tom.nye@ncl.ac.uk

SCHOOL OF MATHEMATICS AND STATISTICS, NEWCASTLE UNIVERSITY, NEWCASTLE UPON TYNE, NE1 7RU, UK