Distance-based analysis of dynamical systems and time series by optimal transport

Michael Muskulus

1. Distances are a powerful tool to study complex systems, especially in a multivariate setting when they are reconstructed by multidimensional scaling. This thesis, chs. 2, 3, 4, 6.

2. There is more clinically relevant information in respiratory impedance time series than is conveyed by the average and standard variation of its magnitude over time. Quantified by distances, this information allows to distinguish reliably between asthma and chronic obstructive pulmonary disease. This thesis, ch. 3.

3. Magnetic resonance images, interpreted as empirical probability distributions, can be compared with each other quantitatively to detect subtle changes in tissue composition. Although more involved than traditional histogram analysis, this leads to higher diagnostic accuracy, e.g., for systemic lupus erythematosus. This thesis, ch. 4.

4. A solution to the problem described below will provide the exact significance probability for the two-sample generalized Moore-Rayleigh test, i.e., allow for the detection and localization of differences in local brain shape:

Given a bound \( r > 0 \) and a finite number of maximal step sizes \( c_n \geq c_{n-1} \geq \cdots \geq c_1 > 0 \), what is the symmetric random flight, whose \( i \)-th step is bounded in magnitude by \( c_i \), that maximizes the probability to result in a distance farther than \( r \) from the origin?

This thesis, ch. 5.
5. Many measures in the analysis of electrophysiological signals are bivariate and non-metric. These are difficult to interpret and should not be trusted in a multivariate setting. The Wasserstein distances overcome this limitation, and lead to more meaningful and interpretable results. *This thesis, ch. 6.*

6. It is expected that repeated measures of regional MR parameter distributions will allow to detect Alzheimer’s disease earlier than current single time point relaxometric or brain atrophy measurements.

7. Displacement interpolation along optimal transportation plans allows to find intermediate states for each pair of systems. It is possible to determine a representative “centroid” for a finite number of systems by a probabilistic algorithm using only such pairwise interpolation along lines.

8. Simple and well-understood methods such as canonical discriminant analysis should be the first choice for classification problems, even if their assumptions are not met. Since fewer parameters need to be estimated, they often outperform more complicated methods.

9. Concepts that are beautiful and elegant in theory do not necessarily have satisfactory practical properties. Nonlinear time series analysis is an important example.

10. There is a huge difference between a good method and a good method for which there exists a good computer code.

11. The biggest problem in science is not to find a problem worth working on, but to find one that has the right degree of difficulty.

12. There is no scientific definition of what science is, so the evaluation of science (e.g., by citation analysis) has to be considered unscientific and ultimately irrational.