



Universiteit
Leiden
The Netherlands

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

Muskulus, M.

Citation

Muskulus, M. (2010, February 11). *Distance-based analysis of dynamical systems and time series by optimal transport*. Retrieved from <https://hdl.handle.net/1887/14735>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/14735>

Note: To cite this publication please use the final published version (if applicable).

Chapter 1

General Introduction

Scientific discovery consists in the interpretation for our own convenience of a system of existence which has been made with no eye to our convenience at all. One of the chief duties of a mathematician in acting as an advisor to scientists is to discourage them from expecting too much of mathematicians.

Norbert Wiener

1.1 Distance-based analysis

This section gives a concise overview of the methodology of distance-based analysis and the underlying ideas without undue details.

Systems and measurement devices

The setting is the following: We are given a number of *systems* (S_1, S_2, \dots) that we want to study. Conceptually, we are not concerned with the actual nature of the systems; but we need to specify a way in which to obtain objective (quantitative) information about them. This is achieved by specifying one or more *measuring devices* (D_1, D_2, \dots) that map each system (possibly at a specific point in time or with various other influential factors fixed), arising from a class \mathcal{S} of measurable systems, into a set of numerical measurements: $D_i : \mathcal{S} \rightarrow M$. In the simplest case these measurements will be univariate (a single number) or multivariate (a “vector” of numbers). Let us illustrate this with an example. Consider a time series, i.e., a set of numbers (x_1, x_2, \dots, x_n) sampled at n discrete points in time. This can be considered to be a single measurement of a certain system, e.g., it might represent temperature measurements at noon on consecutive days at a certain point in space, and would be naturally represented by the vector $x = (x_1, x_2, \dots, x_n)^t \in \mathbb{R}^n = M$, where the superscript t denotes the transpose. Other measurements might include some kind of processing and could be more involved. For example, the mean $\bar{x} = 1/n \sum_{i=1}^n x_i$ of the time series can be considered a single measurement.

Generalized measurements

This framework is very general and allows for much flexibility. With a small generalization we can even accommodate “ideal” (not physically realizable) measurements. For example, we might consider the probability distribution of temperature values, i.e., the function $F(x)$ that tells us how likely it is to find a temperature $T \leq x$ on an otherwise unspecified, and therefore random, day. This distribution could only be obtained by an infinitely long time series (and there might still be problems with its definition without some further stationarity assumptions), but we can always approximate it by an estimate obtained from a finite time series. Whether and when at all such an estimate would be sensible is not the point here, but rather that the class of measurement devices should be allowed to also contain potentially infinite objects, namely, *probability measures*. This is a natural enough extension, since each deterministic measurement $x \in M$ can be identified with a probability measure over M where outcomes other than x have zero probability. Let us denote the class of probability measures defined over a common space M by $P(M)$. We think of these as *generalized measurements* of our systems.

The space of measurements

To compare quantitatively two probability measures from $P(M)$, we will require that the space M already comes with a notion of distance, i.e., is a *metric space*. For example, measurements that result in numerical quantities, such that $M \subseteq \mathbb{R}^n$ for some finite $n \in \mathbb{N}$, are usually metric. Only if M is a proper subset of some \mathbb{R}^n there might be problems, but we are usually allowed to assume that, at least potentially, the range M covers the totality of \mathbb{R}^n . Note that the interpretation and usefulness of this metric depends on the measurement apparatus. This expresses the important fact that we only have access to a system through a measurement device D . Properties that are not measured by D can obviously not be restituted later from such measurements. This seems a serious limitation at first, and in a certain way it is: Since it is only possible to analyze measurements, we need to assume that the measurement device captures all the information necessary for the analysis task at hand. It is a fact, however, that this limitation is a *principal* problem that is inherent to all experiments and data analysis and cannot be overcome by *any* method. We are therefore justified, indeed almost compelled, to *identify* the measurements in M with the systems themselves. In practice this is rarely problematic, especially if a lot of data is available (e.g., a long time series recording) that can be assumed to characterize all interesting aspects of a system sufficiently. The main challenge lies in the way we deal with this data and extract meaningful insights from it.

Distances between measurements

In *distance-based analysis* we calculate an abstract distance $d : P(M) \times P(M) \rightarrow [0, \infty)$ between each pair of (generalized measurements of) systems. To ensure sensible behaviour in a multivariate setting, we require the function d to be a proper distance: It should be a nonnegative number that is zero between two identical systems, $d(P, P) = 0$ (*reflexivity*). Ideally it should only be zero for two identical systems, such that $d(P_1, P_2) > 0$ if $P_1 \neq P_2$, which together with the previous property is called *positive definiteness*. This stronger property is not truly needed for most of our applications, but conceptually important. Another property that we require of d is *symmetry*, $d(P_1, P_2) = d(P_2, P_1)$. Finally, the measure d should be minimal in a certain sense. There should be no “short-cuts” possible between two or more systems that result in a shorter distance than the one measured directly between two systems. This property is guaranteed when d fulfills the *triangle inequality*, $d(P_1, P_2) \leq d(P_1, P_3) + d(P_3, P_2)$ for all systems.

Although there are many candidates for such a distance measure, we prefer the class of optimal transportation distances, also called *Wasserstein distances* in the mathematical literature (Villani, 2003). These quantify the amount of “work” that is needed to transform one probability measure P_1 into a second measure P_2 . Detailed definitions will be given later; let us note here that the Wasserstein distances are, remarkably, true distances that fulfill all metric properties.

The first two steps in the distance-based analysis are therefore:

1. To measure some properties of systems in the form of a probability measure over some numerical space (usually some Euclidean space \mathbb{R}^n).
2. To quantify distances between all pairs of systems under consideration by calculating a Wasserstein distance. This results in a matrix of mutual distances, which will then be analyzed further.

Reconstruction of distances as point configurations

Although there exist quite a few statistical methods to deal with distance matrices, mostly originating from problems in ecology (Legendre and Legendre, 1998), it is very advantageous to represent the distances as actual points in some space with which we are familiar. We will use standard Euclidean space \mathbb{R}^k for this purpose, and furthermore require that $k \ll m$. The latter implies a reduction to a low-dimensional subspace that preserves the most interesting properties of the systems under study. The coordinates of each system in this space \mathbb{R}^k will therefore be derived by the methods of multidimensional scaling (Borg and Groenen, 2005). This is similar to principal component analysis and results in a representation where the first coordinate describes (“explains”) the largest variation in the distances, the second coordinate describes the largest remaining variation in the distances, orthogonal to the first, and so on.

Reconstruction in a Euclidean space

Here the reader might wonder about an important issue. It is not obvious that the distances measured can be represented at all in a Euclidean space. For example, geodetic distances (where the shortest line between two points is not necessarily a straight line) are not obvious to realize in a Euclidean space. Phase distributions are the canonical example here: A phase is a number between 0 and 2π where we identify 2π with 0. Topologically, the space of all phases is a circle, and there are always two paths between each pair of distinct phases. The geodetic distance between two phases φ_1 and φ_2 is the shorter value of these, either $|\varphi_1 - \varphi_2|$ (not crossing 0) or $2\pi - |\varphi_1 - \varphi_2|$ (crossing 0). To represent a set of phases as points in a Euclidean space, where distances are measured by straight lines, usually introduces errors, i.e., misrepresentations of the distances. On the other hand, if the dimensionality k of the ambient space is chosen high enough (e.g., $k = N - 1$ for distances between N points) distances can always be represented perfectly by points in an \mathbb{R}^k . Using a smaller dimension $k' < k$ will usually introduce errors, but it is hoped that these stay reasonably small for a much smaller value $k' \ll k$.

The reasons we advocate the use of Euclidean space, even though it might not be optimal for certain kinds of distances, are threefold. A technical reason is that the reconstruction of Euclidean coordinates from distances is straightforward, whereas representations in other kinds of spaces are much more involved — and bring their own problems in terms of computational efficiency, convergence of algorithms, and so on. Another advantage of Euclidean space is the great familiarity we have with this kind of representation, which is therefore very useful in data exploration and the search for patterns. On top of this, we argue that it is not necessary to obtain a perfect representation anyway. As in the case of mapping the earth, it is often not necessary to consult a globe (an almost distortion-free model of the spatial relationship between points on the earth's surface), but various two-dimensional projections (Mercator, Albers projection, equirectangular projection) suffer for most applications, even though these each of these introduce specific misrepresentations and errors.

In practice, we will therefore use multidimensional scaling to obtain a low-dimensional Euclidean representation of a given distance matrix. Various diagnostic measures allow us to assess the misrepresentation error introduced, and to obtain a viable compromise with regard to the dimensionality used. This representation of the original systems in an abstract Euclidean space is called the *functional* or *behaviour* representation, since it illustrates the relationships between the systems measured and is obtained from their function or behaviour (as defined by the measurement device).

Classification in functional space

This functional space supplies us with coordinates that are then amenable to statistical analysis by the usual tools of multivariate analysis (Härdle and Simar, 2003; Webb, 2002). We will primarily be interested in the classification of different subtypes of systems, and the main tool we will employ is linear discriminant analysis (LDA). Although there exist many more involved methods, LDA is easy to implement, well understood both theoretically and in practice, and relatively robust: even if its assumptions (normal distribution for the subgroups in functional space, with equal variance) are not met, it is usually the method of choice for small sample sizes (number of systems studied), since further degrees of freedom (e.g., in quadratic discriminant analysis which discards with the equal variance assumption) need large datasets to improve upon LDA. The same Caveat also applies to truly nonparametric methods such as kernel density or nearest-neighbour classification, and these methods will not be used in our applications (where sample sizes are on the order of 10 to 100 typically).

As will be seen later, this rather simple sounding setup (distances between probability measures, reconstruction of coordinates in a functional space, classification of systems by normal models) allows for a surprisingly effective classification of complex systems, improving on currently established methods in many cases. Moreover, this approach is completely modular. Each of these steps can be individually refined, and we will indeed describe some of these possibilities in more detail later.

1.2 Reader's guide

Interdisciplinary research is faced with two dilemmas, both of which are mirrored in this thesis. The first is concerned with the *relevance* of findings. Ideally, problems from one discipline, when transferred into another domain, are more easily solved, and result in nontrivial insights in the original setting. Moreover, questions in the former should also lead to interesting problems and insights in the second domain. In practice, this balance is seldomly achieved and usually one domain benefits the most. Here we will be mostly concerned with two domains: the application domain (medicine, neuroscience), and the mathematical domain (analysis, statistics). We have tried to balance the text such that readers from both domains will find the thesis interesting. However, as the actual application of distance-based methods can be quite involved, the text is biased toward the applications, and the mathematical side of the story is not fully developed.

The second dilemma is concerned with the necessary level of *exposition*. Since this thesis should be both readable by mathematicians and non-mathematicians alike, compromises had to be made on both sides. On the one hand, we will not state mathematical results in their most elegant or general form, but discuss a setting that

is natural for most applications and refer to further results in notes at the end of the thesis (starting at page 229). On the other hand, we will also discuss some subtle mathematical problems and actually prove a few theorems.

Most of these technical details have been delegated to two appendices that can be consulted later or in parallel with the rest of the thesis, and there are therefore three entry points to this thesis:

- The reader can start reading with Chapter 2, which gives a detailed overview of distance-based analysis (without too many distracting technicalities),
- he/she can directly skip to the applications (starting with Chapter 3),
- or he/she can first read appendices A and B for a more mathematical exposition, and then continue with Chapter 2.

In the rest of this section we will briefly describe the contents of the remainder of this thesis.

Chapter 2: Dynamical systems and time series

Dynamical systems as a general framework for time-varying phenomena are discussed in Chapter 2. Optimal transportation distances for dynamical systems are defined, and their usefulness is discussed by way of examples. The main theoretical problem is the dependence of the distances on the measurement apparatus (projection from the space of systems to the metric space of measurements). In dynamical systems theory this is avoided by focussing on properties that are invariant under diffeomorphisms, i.e., smooth changes of coordinates. This solution is not available when metric properties are important. On the other hand, in practice one often faces the situation where a number of systems are measured by one and the same measurement apparatus, avoiding this difficulty. Other ways to alleviate this problem are discussed in Section 2.7.

This chapter is based on:

Muskulus M, Verduyn-Lunel S: 2009 — Wasserstein distances in the analysis of dynamical systems and time series. Technical Report MI-2009-12, Mathematical Institute, Leiden University. Extended journal version submitted.

Chapter 3: Lung diseases

In Chapter 3 we will analyze experimental data that consists of time series containing information on mechanical properties of the lungs. We will see that optimal transportation distances allow to successfully distinguish different lung diseases and might potentially allow to track airway status over the course of time. For completeness, the complementary approach of fluctuation analysis is also discussed.

This chapter is based on:

Muskulus M, Slats AM, Sterk PJ, Verduyn-Lunel S — Fluctuations and determinism of respiratory impedance in asthma and chronic obstructive pulmonary disease. Submitted.

Chapter 4: Structural brain diseases

Chapter 4 considers an application to brain imaging. Instead of tracking dynamical processes in time, the tissue properties of the brain are evaluated at one or more points in time by magnetic resonance (MR) imaging, and the goal is the quantification and detection of brain diseases from the distribution of MR parameters (relaxometry). The classical quantitative approach to quantitative MR imaging is outlined, including a discussion and critique of commonly used histogram analysis methods (Tofts, 2004). Wasserstein distances are then tested in the detection of subtle tissue changes in patients suffering from systemic lupus erythematosus (with respect to healthy controls), and in the detection of Alzheimer's disease.

This chapter is based on:

Luyendijk J, Muskulus M, van der Grond J, Huizinga TWJ, van Buchem MA, Verduyn-Lunel S — Diagnostic application of a new method of magnetization transfer ratio analysis in systemic lupus erythematosus: initial results. In preparation.
and also

Muskulus M, Scheenstra AEH, Braakman N, Dijkstra J, Verduyn-Lunel S, Alia A, de Groot HJM, Reiber JHC: 2009 — Prospects for early detection of Alzheimer's disease from serial MR images in transgenic mouse models. *Current Alzheimer Research* 6, 503–518.

Chapter 5: Deformation morphometry

For completeness, Chapter 5 is included which discusses the related problem of deformation morphology, i.e., how to find regions in the brain that are deformed with respect to an average brain image.

This chapter is based on:

Muskulus M, Scheenstra AEH, Verduyn-Lunel S: 2009 — A generalization of the Moore-Rayleigh test for testing symmetry of vector data and two-sample problems. Technical Report MI-2009-05, Mathematical Institute, Leiden University. Extended journal version submitted.

Chapter 6: Electrophysiology of the brain

In contrast to the application to imaging data is the electrophysiological approach of Chapter 6, where time series with high temporal resolution are obtained. Even on the sensor niveau, i.e., on the brain surface, optimal transportation distances reveal interesting phenomena (Section 6.5). Ideally, this application would need to validate

the distances by a forward model, or to solve the inverse problem of source localization (e.g., by beamforming) and then compare source-related time series of activation. Unfortunately, many details of these exciting ideas still need to be worked out, and we therefore restrict ourselves to a proof of principles. A different use is discussed in Section 6.4, where Wasserstein distances are used to qualitatively compare distributions of instantaneous phases obtained from magnetoencephalographic recordings by the Hilbert transform. These offer interesting insights into the functional organization of neuronal networks.

This chapter is based on:

Muskulus M, Houweling S, Verduyn-Lunel S, Daffertshofer A: 2009 — Functional similarities and distance properties. *Journal of Neuroscience Methods* **183**, 31–41.

and also

Muskulus M, Verduyn-Lunel S: 2008 — Reconstruction of functional brain networks by Wasserstein distances in a listening task. In: Kakigi R, Yokosawa K, Kurik S (eds): *Biomagnetism: Interdisciplinary Research and Exploration*. Hokkaido University Press. Sapporo, Japan, pp. 59–61.

Epilogue, Boxes & Notes

In the epilogue, we look back on the experiences obtained with the optimal transportation distances and end with an outlook to the future. This is followed by a number of appendices that have been included for completeness (see below) and then a Notes section where additional topics are discussed, references to additional or complementary works are given, and other details are discussed that would disrupt the flow of the main text. To increase readability, scattered throughout the text are also *boxes* that highlight and summarize important points, see Box 1 for an example.

Appendix A: Distances

Appendix A recalls the basic facts and properties of distances. After introducing metric spaces as the abstract mathematical object in which a notion of distance is defined, the question of embeddability in Euclidean spaces is discussed, culminating in the classical solution of Theorem 4. This is followed by an exposition of multidimensional scaling, which is the main tool that allows for the reconstruction of a metric space from its distances, when these are influenced by (small) errors and numerical inaccuracies. As discussed in Section 1.1, this reconstruction will be the vantage point for statistical analysis. Multivariate methods can be successfully applied in the reconstructed space, and we provide the interested reader with details about diagnostic measures, linear classification methods, cross-validation, and permutation tests for distance matrices.

Box 1. Additional typographic elements

- Summaries of main points and results can be found in boxes such as these, scattered throughout the text.
- Pointers to further topics and additional references can be found in the Notes at the end of the thesis.

Appendix B: Optimal transportation distances

Appendix B recalls facts about the optimal transportation distances that are used throughout the rest of the thesis. These distances are motivated by considering a few other possible distances and their shortcomings, and introduced in a general setting that shows the elegant theoretical foundation. For the application in practice this will be considerably restricted, and the theory will attain a discrete and more combinatorial flavour.

Additional appendices in the electronic version

In the electronic version, additional appendices complete the presentation, in which the two software packages are documented that were developed for the computations in this thesis.

1.3 Major results & discoveries

Due to the many topics touched upon in this thesis, it seems worthwhile to stress the main innovations and results in a central place. Let us begin with the major general achievements:

- This thesis provides the main tools for the multivariate analysis of complex systems by distances. Although these are all more or less well-known techniques in their relative disciplines, this is the first time that they are *combined* in such a way.
- *Extensive software* has been written that allows for a streamlined application of the methods introduced and discussed here.
- All of the necessary background information is available *in one place* in this thesis, with extensive references also covering further, advanced topics.

Minor innovations that are introduced, but not actually worked out in detail, include the following:

- Wasserstein distances be used in the numerical analysis of dynamical systems to visualize their response to changes in parameters. This numerical *bifurcation analysis* is quantitative.
- The Wasserstein distances can be defined *relative to transformations* of the data, e.g., a translation or rotation. This offers an alternative way to solve the Procrustes problem, and quantifies differences in the shape of distributions that are invariant with respect to location or orientation.
- Wasserstein distances can always be interpolated along geodesics between two distributions. An iterative stochastic algorithm then allows to find approximations to centroids by these bivariate interpolations only. Thereby, it becomes possible to determine *characteristic representations* even for very complex kinds of data.

With regard to actual applications, the following are main results discussed in the text:

- Lung diseases are assessed routinely by the forced oscillation technique, but mostly only time averages are used. It is shown that there is potentially much more information contained in the dynamics of these signals, and that this information allows to discriminate between healthy and diseased lungs with very high accuracy.
- There is indeed evidence for power-law-like scaling behaviour of the fluctuations of respiratory impedance. This has been conjectured previously, but due to methodological problems the evidence in previous work has to be judged unreliable and inconclusive. Our findings provide evidence by state-of-the-art maximum-likelihood estimation that supports the power-law hypothesis..
- Systemic lupus erythematosus is a neurodegenerative disease that affects the brain. It has previously been shown that it significantly alters the distribution of magnetization transfer ratios in the brain, which is evaluated in so-called “histogram analysis” by quantifying changes in the location and height of its mode. The distance-based analysis improves on this and allows for better classification of individual patients.
- The distance-based analysis of distributions of relative phases, obtained from magnetoencephalographic signals in a bimanual coordination task, revealed and quantified interactions (crosstalk) between motor areas in a robust way. From the functional representation of these distances it now becomes possible to formulate hypotheses regarding the underlying neural networks and to set-up mathematical models.