Summary of habilitation thesis: *Applications of the Euclidean Distance Discriminant*

This habilitation thesis fits in the field of *Metric Algebraic Geometry.* This term is a neologism which joins the names metric geometry and algebraic geometry. It first appeared in the title of M. Weinstein's dissertation. Building on classical foundations, the field embarks towards a new paradigm that combines concepts from algebraic geometry and differential geometry, with the goal of developing practical tools for the 21st century. An important step in Weinstein's dissertation is an article written together with the author of this thesis (namely [12]). This article is at the basis of Chapter 2 of which Chapter 3 is a continuation and conclusion.

The thesis contains six chapters, the first one is an introductory chapter presenting the general research domain of this work by recalling the basic notions of the *Euclidean Distance Degree* (based on the fundamental article of the author [4]) necessary for this thesis to be a self-content reading. The last chapter contains some potential further research development directions. The remaining four chapter are each centred around one of the articles [8, **?**, 11, 12], all written by the author after finishing his doctoral studies.

This work is more precisely on the topic of different applications of the *Euclidean Distance (ED) Discriminant* (**hence the title**) and it has two parts. Part one consists of Chapter 2 and 3 analysing the connection of the ED discriminant to topological notions, such as the offset varieties, the reach, medial axis, bisectors hypersurfaces, centres and radii of curvature of the underlying variety.

Part two consists of Chapter 4 and 5, where it is introduced an inverse problem to the Euclidean Distance Degree problem from the point of view of special *data loci*, namely sets of special parameters, such that a given algebraic optimization problem will have some critical solutions in a predetermined variety. For example one can be interested if the solution will be singular, have symmetry, have coordinates summing to one, etc. The notion of data loci were developed by the research team coordinated by the author during the six years from $2017 - 2023$. Said research materialized in the articles [7] of the author, [10] by the author and Rodriguez, article [11] by the author and Rodriguez which is at the basis of Chapter 4 and article [9] by the author and Turatti which is at the basis of Chapter 5. Since this later article answers an important open question in the field of tensor rank decompositions, it is the paramount point of a more than six years research project. We also remark that special data loci (eg. data singular- and data isotropic loci) are subvarieties of the ED discriminant, hence the connection to the topic.

A short summary of each chapter follows.

Chapter 2: The ED- and offset discriminants and applications to persistent homology

Experimental research is based on collecting and analyzing data. It is very important to understand the background mathematical model that defines a given phenomenon. One of the possibilities is that the data is driven by a geometric model, say an algebraic variety or a manifold. In this case, we would like to "learn the geometric object" from the data. For example, we would like to understand the topological features of the underlying model. A common way to do this is by *persistent homology*, which studies the homology of the set of points within a range of distances from the data set, and considers features to be of interest if they persist through a wide range of the distance parameter.

This chapter is based on the article E. Horobeţ, M. Weinstein, *Offset Hypersurfaces and Persistent Homology of Algebraic Varieties*, Computer Aided Geometric Design, Volume 74 (2019), Page 101767. The present study is at the intersection of computational geometry, geometric design, topology and algebraic geometry, linking all of these topics together. In this chapter we study the *persistent homology of the offset filtration* of an algebraic variety, which we define to be the homology of its offsets.

We show that the indicators (barcodes) of the persistent homology of the offset filtration of a variety defined over the rational numbers are algebraic and thus can be computed exactly (first main theorem of this chapter). Moreover, we connect persistent homology and algebraic optimization (by the Euclidean Distance Degree) through the theory of offsets, bringing insights from each field to the other. Namely, we express the degree corresponding to the distance variable of the offset hypersurface in terms of the Euclidean Distance Degree of the original variety (second main theorem of this chapter), obtaining a new way to compute these degrees. A consequence of this result is a bound on the degree of the ED discriminant and on the degree of the closure of the medial axis. We describe the non-properness locus of the offset construction (Subsection 2.1) and use this to describe the set of points (third main theorem of this chapter) in the ambient space that are topologically interesting and relevant to the computation of persistent homology. Lastly, we show that the reach of a manifold, the quantity used to ensure the correctness of persistent homology computations, is algebraic (fourth main theorem of this chapter).

Chapter 3: The ED discriminant and the critical curvature degree of a variety

We have seen in the previous chapter that sampling an object is an important problem when trying to recover information about its structure. Using geometric information such as the bottlenecks and the local reach in article [3] the authors provide bounds on the density of the sample needed in order to guarantee that the homology of the variety can be recovered from the sample. If the sample is finer than the reach, then the homology can be recovered.

If our sample data is driven by a geometric model, say an algebraic variety X, then the reach can be computed as a minimum of two quantities, more precisely, one of which is the radius of the narrowest bottleneck of X, and the other one is the minimal radius of spherical curvature on X. We will call the spherical curvature simply curvature.

In this chapter we deal with the complexity involved in the computation of the reach and in particular the computation of locally minimal, maximal, etc. curvature points. An analysis of the bottleneck degree of a variety was done in [2, 5], but understanding critical curvature points in arbitrary dimension was open till now. In this chapter we will analyse the set of all points on the Euclidean Distance Discriminant to a real variety X that correspond to centres of curvature at points with critical curvature (locally minimal, maximal, saddle type, etc.). We construct a variety containing these points and we will call its degree the *Critical Curvature Degree* of X. Furthermore in this chapter we give an algorithm to construct the variety of critical curvature points and we analyse their connection to the singular locus of the ED discriminant. This chapter is based on the article E. Horobeţ, *The critical curvature degree of an algebraic variety*, Journal of Symbolic Computation, Volume 121, March–April 2024, 102259.

Chapter 4: The data discriminant: a method to describe subvarieties of the ED discriminant

In many kinds of optimization problems (distance optimization, optimizing communication rate etc.) it is interesting to ask the question of whether the solution is satisfying certain meaningful (polynomial) conditions. For example one can be interested if the solution will be singular, have symmetry, have coordinates summing to one, etc. Equally interesting is to ask the same question not only about the minimizer of the optimization problem, but about all of the local minima and maxima as well. Even stronger, we want conditions for all of the critical points of the problem. In other words, we are considering a generalization of the inverse problem of determining the parameter data that exhibits a certain type of critical point of the objective function. In this chapter we provide examples, methods and algorithms to test the properties of a parametric optimization problem.

Our motivating example for doing the generalization is the *scaled distance function*. A scaled distance function on $\mathbb{R}^n$ is prescribed by a parameter data $u \in \mathbb{R}^n$ and a fixed scaling vector $w \in \mathbb{R}^n$ as

$$d_u^w : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto \sum_{i=1}^{n} w_i(u_i - x_i)^2.$$

For this special case, our main problem is to solve

$$\begin{cases} \text{minimize } d_u^w(x) \\ \text{subject to } x \in X_{\mathbb{R}}, \end{cases} \tag{1}$$

where $X_{\mathbb{R}}$ is an affine variety in $\mathbb{R}^n$. We want to provide the set of parameters $u \in \mathbb{R}^n$ for which at least one of the critical points of Problem 1 satisfy prescribed (polynomial) conditions.

The main result of this chapter gives a description of this special data locus, as well as an algorithm for computing it and we apply our findings to classical distance optimization concerning low rank and structured low rank approximations of matrices and tensors, we cover weighted distance optimization, we relate to the maximum likelihood (ML) degree, from algebraic statistics and finally we show that optimizing communication rate also fits our problem setting. This chapter is based on the article E. Horobeț, J. I. Rodriguez, *Data loci in algebraic optimization*, Journal of Pure and Applied Algebra, Volume 226 (2022), Issue 12, Page 107144.

Chapter 5: When does subtracting a critical rank-one tensor decrease rank?

Low-rank approximation of matrices is used for mathematical modelling and data compression, more precisely in principal component analysis, factor analysis, orthogonal regression, etc. In order to get all critical rank-one approximations of a given matrix one can find all critical points of the distance function from the said matrix to the variety of rank-one matrices. By the Eckhart-Young theorem, this is done by computing singular value decomposition and the number of such critical rank-one approximations is always the minimum of the column and row dimensions of the matrix. Furthermore, by subtracting any such critical rank-one approximation from the matrix we get a drop in the rank, hence obtaining a suitable algorithm to construct any low-rank approximation of the matrix.

Low-rank approximations of tensors have even more application potential, but they are much more challenging both mathematically as well as computationally (tensor rank and many related problems are NP-hard). Despite this fact, many algorithms exist for finding rank-one approximations of a tensor. A way to do this, similarly to the matrix case, is by finding all critical points of the distance function from the said tensor to the variety of rank-one tensors (luckily this is an algebraically closed set). The generic number of such critical approximations was computed in [6] and shows that the degree of complexity of this problem for tensors is substantially higher than for matrices.

For higher-rank approximations, though, we have that tensors of bounded rank do not form a closed subset, so the best low-rank approximation of a tensor on the boundary does not exist (see [1]). From this results that subtracting a rank-one approximation from a tensor might even increase its rank (see [13]).

In this chapter, to resolve this obstacle for higher-rank approximations we turn our attention to the definition of *border rank* of tensors and we ask the question: what is the closure of the set of those tensors for which subtracting a rank-one approximation does result in lowering the (border) rank?

We approach this problem by constructing the variety DL_1 of tensors for which subtracting a critical rank-one approximation yields a rank-one tensor. Then we construct the variety DL_2 of tensors for which subtracting a critical rank-one approximation yields an element of DL_1, and so on. The main finding of this chapter can be formulated as follows.

Theorem 0.1. *Let $X \subset \mathcal{S}^d\mathbb{C}^n$ be the cone over the Veronese variety and our inner product in $\mathcal{S}^d\mathbb{C}^n$ is the Bombieri-Weyl inner product. Then the chain $X = \mathrm{DL}_1 \subset \cdots \subset \mathrm{DL}_r \subset \ldots$ stabilizes. This limit DL_N (for some sufficiently large N) is the closure of all tensors T in V, for which subtracting a critical rank-one approximations of T from itself and repeating this process finitely many times we will eventually get to zero.*

Furthermore, we study in depth the variety DL_2 of tensors for which subtracting one of its critical rank-one approximations we get a rank-one tensor. We will see that this variety is determined by the *bottleneck points* of the variety of rank-one tensors and is related to the nodal singularities of the *hyperdeterminant*. We also show the relation between the DL_i's and orthogonally decomposable tensors.

This chapter is based on the recent article E. Horobeț, E. T. Turatti, *When does subtracting a rank-one approximation decrease tensor rank?*, arXiv:2303.14985, 2023. Since this article answers an important open question in the field of tensor rank decompositions, it is paramount point of a six years research project coordinated by the author.

References

[1] V. De Silva, L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM Journal on Matrix Analysis and Applications, (2008), 30(3), 1084-1127.

[2] S. Di Rocco, D. Eklund, M. Weinstein, *The bottleneck degree of algebraic varieties*, SIAM Journal on Applied Algebra and Geometry 4.1 (2020), 227-253.

[3] S. Di Rocco, D. Eklund, O. Gäfvert, *Sampling and homology via bottlenecks*, Mathematics of Computation, (2022), 91(338), 2969-2995.

[4] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels and R.R. Thomas: *The Euclidean distance degree of an algebraic variety*, Foundations of Computational Mathematics **16** (2016), 99–149.

[5] D. Eklund, *The numerical algebraic geometry of bottlenecks*, Advances in Applied Mathematics, (2023), 142, 102416.

[6] S. Friedland, G. Ottaviani, *The number of singular vector tuples and uniqueness of best rank-one approximation of tensors*, Foundations of Computational Mathematics, (2014), 14, 1209-1242.

[7] E. Horobeț, *The Data Singular and the Data Isotropic Loci for Affine Cones*, Comm. Algebra, Volume 45 (2017), Issue 3,1177 – 1186

[8] E. Horobeț, *The critical curvature degree of an algebraic variety*, Journal of Symbolic Computation, Volume 121, March–April 2024, 102259.

[9] E. Horobeț, E. T. Turatti, *When does subtracting a rank-one approximation decrease tensor rank?*, arXiv:2303.14985, 2023.

[10] E. Horobeț, J. I. Rodriguez, *The Maximum Likelihood Data Singular Locus*, J. Symbolic Comput., Volume 79 (2017), Part 1, 99–107

[11] E. Horobeţ, J. I. Rodriguez, *Data loci in algebraic optimization*, Journal of Pure and Applied Algebra 226, no. 12 (2022): 107144.

[12] E. Horobeţ, M. Weinstein, *Offset hypersurfaces and persistent homology of algebraic varieties*, Computer Aided Geometric Design 74 (2019) : 101767.

[13] A. Stegeman, P. Comon, *Subtracting a best rank-1 approximation does not necessarily decrease tensor rank*, Linear Algebra and its Applications, 2010, 433 (7), pp.1276–1300.