ALGEBRAIC MACHINE LEARNING WITH AN APPLICATION TO CHEMISTRY

EZZEDDINE EL SA1 AND PARKER GARA1 AND MARKUS J. PFLAUM∗1

1Department of Mathematics, University of Colorado, Boulder CO 80309, USA

Abstract. As data used in scientific applications become more complex, studying the geometry and topology of data has become an increasingly prevalent part of data analysis. This can be seen for example with the growing interest in topological tools such as persistent homology. However, on the one hand, topological tools are inherently limited to providing only coarse information about the underlying space of the data. On the other hand, more geometric approaches rely predominantly on the manifold hypothesis which asserts that the underlying space is a smooth manifold. This assumption fails for many physical models where the underlying space contains singularities.

In this paper, we develop a machine learning pipeline that captures fine-grain geometric information without having to rely on any smoothness assumptions. Our approach involves working within the scope of algebraic geometry and algebraic varieties instead of differential geometry and smooth manifolds. In the setting of the variety hypothesis, the learning problem becomes to find the underlying variety using sample data. We cast this learning problem into a Maximum A Posteriori optimization problem which we solve in terms of an eigenvalue computation. Having found the underlying variety, we explore the use of Gröbner bases and numerical methods to reveal information about its geometry. In particular, we propose a heuristic for numerically detecting points lying near the singular locus of the underlying variety.

1. Introduction. Studying the geometry of data can provide insight into high dimensional, highly complex datasets. Applications include dimensionality reduction [25], computer vision [29], chemistry [24] and medicine[26]. Geometric methods in machine learning rely predominantly on the manifold hypothesis [14] which asserts that sample data $\Omega \subset \mathbb{R}^n$ in fact lie on a smooth submanifold $M \subset \mathbb{R}^n$ of dimension $\ll n$. Algorithms based upon the this hypothesis are often called manifold learning methods. Examples include PCA and nonlinear PCA [22], Isomap [36], and UMAP [25]. However, the manifold hypothesis does no always hold, especially when the underlying geometry of the data contains singularities. Singularities are in fact ubiquitous in mathematics and appear often in physical models [13, 31]. This explains the need to go beyond the manifold hypothesis.

To transition from the smooth to the singular setting we replace the language of differential geometry and smooth manifolds with algebraic geometry and algebraic varieties. At a basic level, algebraic geometry studies the geometry of the zeros of systems of polynomials. Those zero sets are called algebraic varieties. At the heart

2020 Mathematics Subject Classification. Primary: 14Q30;68T05 Secondary: 62R07;92E10.

Key words and phrases. statistical learning, real algebraic geometry, varieties, singularities, maximum a posteriori problem, Lojasiewicz inequality, conformation space, cyclooctane.

∗Corresponding author: Markus J. Pflaum.
of algebraic geometry is the duality between geometry and algebra, which allows us to jump back and forth between geometric spaces and computable algebraic procedures. Furthermore, algebraic geometry offers a natural setting for studying and working with singularities. As we shall see, this variety hypothesis provides a great amount of flexibility and it lends itself well for computations.

To examine data in the singular setting we introduce the algebraic machine learning pipeline depicted in Figure 1.1. This pipeline combines ideas from algebraic geometry and machine learning and it is the main subject of this paper.

In Section 2, we give a brief overview of real algebraic geometry, building the language we will be using throughout the paper. In Section 3, we discuss the first stage of learning the underlying variety. We introduce the following learning problem:

\[(LP) \quad \text{Given a dataset } \Omega = (a_1, \ldots, a_m) \text{ of points in } [0,1]^n \text{ sampled from a variety } V \subset \mathbb{R}^n, \text{ find that variety.}\]

We interpret this learning problem as a Maximum A Posteriori (MAP) problem which can be solved in terms of an eigenvalue computation. This gives a more rigorous justification for the algorithm presented in [7]. Building upon this work, we present a closely related algorithm as Algorithm 1 as well as a possible enhancement as Algorithm 2. Let us mention that restricting the sample points to be contained in a unit hypercube might not be an obvious choice, but restriction to some bounded region is necessary for the later statistical arguments to make sense, see Remark 3.3. Indeed, the paper [12] by Dufresne et al. makes similar assumptions.

In Section 4, we introduce the algebraic computations stage. Using the learned variety \( V \) from the previous stage, we explore the use of Gröbner basis computations to reveal information about \( V \). This includes invariants like the dimension and the irreducible decomposition.

In Section 5, we consider the numerical computations stage which involves working directly with the points in \( \Omega \) and additional samples taken from \( V \cap [0,1]^n \). In particular, we introduce a method called the singularity heuristic for detecting points in \([0,1]^n\) lying near the singular locus of \( V \).

In Section 6, we test the algebraic machine learning pipeline on synthetic data and on chemical data sampled from the conformation space of cyclooctane [24].
Finally, in Section 7, we discuss other work related to the algebraic machine learning pipeline.

2. Background in Real Algebraic Geometry. For the reader’s convenience, we collect in this section several fundamental concepts from real and complex algebraic geometry which are used throughout this paper; see [5] and [18] for further details. In this paper, we will be working mostly over the field \( \mathbb{R} \) and its subfields \( \mathbb{Q} \) and \( \mathbb{R}_{\text{alg}} := \overline{\mathbb{Q}} \cap \mathbb{R} \) of rational and real algebraic numbers, respectively. The symbol \( \mathbb{K} \) always denotes a field which we assume to be of characteristic 0 unless stated otherwise.

2.1. Ordered Fields. A total order \( \leq \) on a field \( \mathbb{K} \) is called a field order if it is compatible with the field operations in the following sense:

(M1) If \( a \leq b \), then for all \( c \in \mathbb{K} \) the relation \( a + c \leq b + c \) holds true.

(M2) If \( 0 \leq a \) and \( 0 \leq b \), then \( 0 \leq a \cdot b \).

A field \( \mathbb{K} \) endowed with a field order \( \leq \) is an ordered field. We always assume \( \mathbb{R} \) and \( \mathbb{Q} \) to be equipped with the standard field order \( \leq \). The set of natural numbers \( \mathbb{N} \subset \mathbb{Q} \) carries the induced natural order. Note that the field of complex numbers \( \mathbb{C} \) and finite fields cannot be endowed with the structure of an ordered field. For finite fields this follows from the simple observation that for any ordered field \( \mathbb{K} \) the natural map \( \mathbb{N} \to \mathbb{K} \) is strictly increasing, hence \( \mathbb{K} \) needs to be infinite. For the field of complex numbers see Example 2.1 (a) below.

The order on an ordered field \( (\mathbb{K}, \leq) \) is completely characterized by the positive cone it generates, which is the set \( C = \{ a \in \mathbb{K} \mid 0 \leq a \} \). Let us explain this in some more detail. Recall from e.g. [5, Sec. 1.1] that by a proper cone of \( \mathbb{K} \) one understands a subset \( C \subset \mathbb{K} \) which fulfills the relations

\[
C + C \subset C, \quad C \cdot C \subset C, \quad \mathbb{K}^2 \subset C, \quad \text{and} \quad -1 \notin C,
\]

where \( C + C = \{ a + b \mid a, b \in C \} \), \( C \cdot C = \{ a \cdot b \mid a, b \in C \} \) and \( \mathbb{K}^2 = \{ a \cdot a \mid a \in \mathbb{K} \} \). A proper cone \( C \) of \( \mathbb{K} \) is a positive cone if in addition

\[
\mathbb{K} = C \cup (-C).
\]

A positive cone \( C \) of \( \mathbb{K} \) defines a unique field order \( \leq_C \) by putting \( a \leq_C b \) if and only if \( b - a \in C \). The thus defined relation \( \leq_C \) satisfies property (M1) by definition and (M2) since \( C \cdot C \subset C \). Finally, the cone \( C \) is given by the set \( \{ a \in \mathbb{K} \mid 0 \leq_C a \} \) which concludes the argument that field orders on \( \mathbb{K} \) are in bijective correspondence with positive cones \( C \subset \mathbb{K} \).

Example 2.1. (a) The field of complex numbers \( \mathbb{C} \) does not possess a positive cone since if \( C \subset \mathbb{C} \) were such a cone, then \( -1 = i^2 \in C \) which contradicts (2.1).

(b) For the fields \( \mathbb{K} = \mathbb{R} \) or \( \mathbb{K} = \mathbb{Q} \), the standard field order coincides with the order \( \leq_P \) associated to the positive cone

\[
P = \{ a \in \mathbb{K} \mid \exists a_1, \ldots, a_n \in \mathbb{K} : a = \sum_{i=1}^{n} a_i^2 \}.
\]

An ordered field \( (\mathbb{K}, \leq) \) is called real closed if every polynomial \( f \in \mathbb{K}[x] \) of odd degree has a root in \( \mathbb{K} \) and if for every positive element \( a \in \mathbb{K} \) there exists \( b \in \mathbb{K} \) such that \( a = b^2 \). For example, \( (\mathbb{R}, \leq) \) is real closed, but \( (\mathbb{Q}, \leq) \) is not since the
element 2 is positive but does not have a rational square root. For a real closed
field \((\mathbb{K}, \leq)\), the vector space \(\mathbb{K}^n\) can be endowed with the \(\mathbb{K}\)-valued metric
\[
\|a - b\|_2 = \sqrt{(a_1 - b_1)^2 + \ldots + (a_n - b_n)^2}, \quad a, b \in \mathbb{K}^n.
\] (2.3)
If \(\mathbb{K} \subset \mathbb{R}\) is a real closed subfield, this results in the standard Euclidean metric
restricted to \(\mathbb{K}^n\).

The real closure of an ordered field \((\mathbb{K}, \leq)\) is an ordered field \((\mathbb{F}, \leq\)) which is
real closed and extends \((\mathbb{K}, \leq)\), that is, \(\mathbb{K} \subset \mathbb{F}\) and \(C_\mathbb{K} \subset C_\mathbb{F}\), where \(C_\mathbb{K}\)
and \(C_\mathbb{F}\) are the positive cones in \(\mathbb{K}\) and \(\mathbb{F}\), respectively. Real closures exist and are essentially
unique. For example, the real closure of \((\mathbb{Q}, \leq)\) is \((\mathbb{R}_{alg}, \leq_p)\), where \(P\) is understood
as above to be the set of sums of squares in \(\mathbb{R}_{alg}\).

2.2. The Nullstellensatz. The set of polynomials in \(n\) variables over a field \(\mathbb{K}\)
forms a ring \(\mathbb{K}[x_1, \ldots, x_n]\). Given a subset \(S\) of the polynomial ring \(\mathbb{K}[x_1, \ldots, x_n]\),
we denote by \(Z_\mathbb{K}(S)\) or more briefly by \(Z(S)\) the zero-set of \(S\), that is the set
\[
Z(S) = Z_\mathbb{K}(S) = \{ a \in \mathbb{K}^n | f(a) = 0 \text{ for all } f \in S \}.
\]
By \(\langle S \rangle_\mathbb{K}\) or more shortly by \(\langle S \rangle\) when the ground field is clear, one denotes the ideal
generated by \(S\) which is the intersection of all ideals in \(\mathbb{K}[x_1, \ldots, x_n]\) containing \(S\).
Both the set \(S\) and the ideal \(\langle S \rangle\) have the same zero set \(Z(S) = Z(\langle S \rangle)\) since if
\(f_1(a) = \ldots = f_k(a) = 0\) for polynomials \(f_1, \ldots, f_k\) and \(a \in \mathbb{K}^n\), then \(g_1(a) \cdot f_1(a) +
\ldots + g_k(a) \cdot f_k(a) = 0\) for all \(g_1, \ldots, g_k \in \mathbb{K}[x_1, \ldots, x_n]\).
Since \(\mathbb{K}[x_1, \ldots, x_n]\) is noetherian, any ideal \(I \subset \mathbb{K}[x_1, \ldots, x_n]\) can be written as
\(\langle f_1, \ldots, f_k \rangle\) for some \(k\) and \(f_1, \ldots, f_k \in I\). Therefore, starting with a possibly infinite
set \(S\), \(Z(S)\) can always be rewritten as \(Z(S')\) for some finite set of polynomials
\(S'\).

Over a subfield of \(\mathbb{R}\), this result can be sharpened further. For notional convenience,
we will denote \(Z(\{f\})\) by \(Z(f)\). Note that the following proposition does not hold true over the field \(\mathbb{C}\).

Proposition 2.2. Let \(\mathbb{K}\) be a subfield of \(\mathbb{R}\). If \(S\) is a non-empty subset of the polynomial
ring \(\mathbb{K}[x_1, \ldots, x_n]\), then \(Z(S)\) can be written as \(Z(f)\) for a single polynomial
\(f \in \mathbb{K}[x_1, \ldots, x_n]\).
Proof. First rewrite \(Z(S)\) as \(Z(S')\) for some finite set \(S' \subset \mathbb{K}[x_1, \ldots, x_n]\). Let
\(S' = \{f_1, \ldots, f_k\}\). If \(f_1(a) = \ldots = f_k(a) = 0\), then \(f_1^2(a) + \ldots + f_k^2(a) = 0\) and
\(Z(S') \subset Z(f_1^2 + \ldots + f_k^2)\). Conversely, \(f_1^2(a) + \ldots + f_k^2(a) = 0\) implies \(f_1^2(a) = \ldots = f_k^2(a) = 0\) which then entails \(f_1(a) = \ldots = f_k(a) = 0\). Hence \(Z(\{f_1, \ldots, f_k\}) = Z(f_1^2 + \ldots + f_k^2)\).

A set \(V \subset \mathbb{K}^n\) is called an algebraic variety or a variety if \(V = Z(S)\) for some
set \(S \subset \mathbb{K}[x_1, \ldots, x_n]\). Depending on whether \(\mathbb{K} = \mathbb{R}\) or \(\mathbb{K} = \mathbb{C}\) the variety is called
real or complex.

Similar to the zero-set map \(Z\) one also has the vanishing ideal map \(J\) which maps
a subset \(A \subset \mathbb{K}^n\) to the set of all polynomials vanishing on \(A\) that is to the set
\[
J(A) = J_\mathbb{K}(A) = \{ f \in \mathbb{K}[x_1, \ldots, x_n] \mid f(a) = 0 \text{ for all } a \in A \}.
\]
To see why \(J(A)\) is an ideal, note that if \(f_1(a) = f_2(a) = 0\) for \(a \in A\), then
\(f_1(a) \pm f_2(a) = 0\) and \(g(a) \cdot f_1(a) = 0\) for any polynomial \(g\).

Starting with an ideal \(I \subset \mathbb{K}[x_1, \ldots, x_n]\), it is not necessarily the case that
\(I = J(Z(I))\). This is because there might be other polynomials vanishing on \(Z(I)\)
which were not included in \(I\), so one can only guarantee that \(I \subset J(Z(I))\).
Example 2.3. Consider the ideal $\langle x^2 \rangle \subset \mathbb{C}[x]$. Then $Z(\langle x^2 \rangle) = \{0\}$. However, 

$$J(Z(\langle x^2 \rangle)) = J(\{0\}) = \langle x \rangle ,$$

so $x$ is in $J(Z(\langle x^2 \rangle))$ but not in $\langle x^2 \rangle$.

Over an algebraically closed field, such as $\mathbb{C}$, the missing polynomials are characterized by the radical of $I$, defined as

$$\sqrt{I} = \{ f \mid \exists r \in \mathbb{N}_{>0} : f^r \in I \} .$$

Note that $\sqrt{I}$ is itself an ideal; see e.g. [21, X, §2].

Theorem 2.4 (Hilbert’s Nullstellensatz). Over an algebraically closed field $\mathbb{K}$ the equality

$$J(Z(I)) = \sqrt{I}$$

holds for every ideal $I \subset \mathbb{K}[x_1, \ldots, x_n]$.

An ideal $I$ which satisfies $I = \sqrt{I}$ is called a radical ideal. Over algebraically closed fields, Hilbert’s Nullstellensatz establishes a one-to-one correspondence between varieties and radical ideals in the polynomial ring. For ordered fields there is a similar but more subtle result. To formulate it we need some further notation. Let $(\mathbb{K}, \leq)$ be an ordered field. The real radical of an ideal $I \subset \mathbb{K}[x_1, \ldots, x_n]$ then is defined as

$$\sqrt[\mathbb{R}]{I} = \{ f \in \mathbb{K}[x_1, \ldots, x_n] \mid \exists r \in \mathbb{N}_{>0}, f_1, \ldots, f_k \in \mathbb{K}[x_1, \ldots, x_n] : f^{2r} + \sum_{i=1}^{k} f_i^2 \in I \} .$$

An ideal $I \subset \mathbb{K}[x_1, \ldots, x_n]$ satisfying $I = \sqrt[\mathbb{R}]{I}$ is called real. If $(\mathbb{F}, \leq)$ denotes the real closure of the ordered field $(\mathbb{K}, \leq)$ and if $A$ is a subset of $\mathbb{F}^n$, then the vanishing ideal of $A$ in $\mathbb{K}[x_1, \ldots, x_n]$ is the ideal

$$J_{\mathbb{F}}(A) \cap \mathbb{K}[x_1, \ldots, x_n] .$$

We denote this ideal by $J_{\mathbb{K}}(A)$.

Theorem 2.5 (Real Nullstellensatz [28, Thm. 2.8]). Let $(\mathbb{K}, \leq)$ be an ordered field and $(\mathbb{F}, \leq)$ be its real closure. If $I \subset \mathbb{K}[x_1, \ldots, x_n]$ is an ideal, then $J_{\mathbb{K}}(Z_{\mathbb{F}}(I)) = \sqrt[\mathbb{R}]{I}$. In particular, if $S \subset \mathbb{Q}[x_1, \ldots, x_n]$, then $J_{\mathbb{Q}}(Z_{\mathbb{R}_{\text{alg}}}(S)) = \sqrt[\mathbb{R}]{(S)_{\mathbb{Q}}} .$

Remark 2.6. We are ultimately interested in varieties over $\mathbb{R}$, but for computational purposes we restrict to varieties in $\mathbb{R}^n_{\text{alg}}$ defined by sets of polynomials $S \subset \mathbb{Q}[x_1, \ldots, x_n]$. Therefore, we will mostly work over ideals $I \leq \mathbb{Q}[x_1, \ldots, x_n]$.

The following density result is concerned with the metric given by Eq. (2.3) and is an immediate consequence of [5, Prop. 5.3.5].

Theorem 2.7. If $\mathbb{K} \subset \mathbb{F}$ are both real closed fields subfields of $\mathbb{R}$ and $S \subset \mathbb{K}[x_1, \ldots, x_n]$, then $Z_{\mathbb{F}}(S)$, viewed as a subset of $\mathbb{F}^n$, is dense in $Z_{\mathbb{F}}(S)$ with respect to the metric $\| \cdot \|_2$. In particular, if $S \subset \mathbb{Q}[x_1, \ldots, x_n]$, then $Z_{\mathbb{R}_{\text{alg}}}(S)$ is dense in $Z_{\mathbb{R}}(S)$. 
Intuitively, this means that we can approximate points in $\mathbb{Z}_R(S)$ arbitrarily closely by points in $\mathbb{Z}_R(\text{alg})$. If $f$ vanishes on $\mathbb{Z}_R(\text{alg})$ then since $f$ continuous on $\mathbb{R}^n$ and $\mathbb{Z}_R(\text{alg})$ is dense in $\mathbb{Z}_R(S)$, $f$ must also vanish on $\mathbb{Z}_R(S)$. Therefore, by the real Nullstellensatz

$$\sqrt{\langle S \rangle_Q} = J_Q(\mathbb{Z}_R(\text{alg})(S)) = J_Q(\mathbb{Z}_R(S)) .$$

2.3. Properties of Algebraic Varieties. Real varieties inherit topological and differential structures based on their embedding in $\mathbb{R}^n$ and form a very large class of spaces. For example, a famous result due to John Nash [27] says that every closed manifold is diffeomorphic to a real algebraic variety. Moreover, by the fundamental work [37] of Hassler Whitney, every algebraic variety carries a unique stratification which is minimal among all stratifications fulfilling Whitney’s condition (b). This means that every algebraic variety can be decomposed into finitely many pairwise disjoint smooth manifolds called strata such that two natural regularity conditions are fulfilled, namely the condition of frontier and Whitney’s condition (b). Moreover, there is a smallest among all such decompositions which we call the canonical stratification of an algebraic variety. In this paper, we do not need the subtleties of the construction of the canonical stratification of an algebraic variety, just its existence, and therefore refer the interested reader to [30, Chpt. 1] for the technical details on stratified spaces, the condition of frontier and the Whitney conditions. Since manifolds of different dimensions may appear in the canonical stratification, varieties can also possess singularities and show non-smooth behavior. This makes the variety hypothesis, i.e. the claim that the underlying space of the data is an algebraic variety, a very reasonable assumption especially for scientific and computational purposes.

Example 2.8. Consider the sphere of radius $\frac{1}{2}$ centered at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and the plane $x - y = 0$. Their union is a variety which we can represent as the zero set of the polynomial

$$f(x, y, z) = \left((x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 + (z - \frac{1}{2})^2 - \frac{1}{4}\right) \cdot (x - y) .$$

The variety $V = Z(f)$ is illustrated in Figure 2.1a. Its singular points are the points in the intersection of the sphere with the plane. Note that any point in the intersection must satisfy $y = x$, so one may set $y = x$ in the equation of the sphere and then observes that the singular locus of the variety $V$ is given by the circle $Z((2x^2 - 2x + z^2 - z + \frac{1}{2})) \cap Z(\langle x - y \rangle)$ illustrated in Figure 2.1b.
In what follows we review the concepts of dimension, the Jacobians, and the singular locus of a variety. Let \( V \) be a variety and let \( J(V) \subset \mathbb{K}[x_1, \ldots, x_n] \) be its vanishing ideal. The coordinate ring of \( V \), denoted \( \mathbb{K}(V) \), is the quotient ring \( \mathbb{K}[x_1, \ldots, x_n]/J(V) \). It contains all the information about the variety \( V \). In fact, \( V \) can be reconstructed from \( \mathbb{K}(V) \) using its spectrum; see [18]. The Krull dimension of a commutative ring \( R \), denoted \( \dim(R) \), is the length \( n \) of the longest chain \( \langle 0 \rangle = P_0 \leq P_1 \leq \ldots \leq P_n = \mathbb{K}(V) \) of prime ideals in \( \mathbb{K}(V) \). In case the variety is smooth, the Krull dimension of its coordinate ring coincides with its dimension as a real or complex manifold, respectively. The following result gives a geometric interpretation of the Krull dimension of a coordinate ring. It is an immediate consequence of [5, Prop. 2.8.5].

**Proposition 2.9.** For every real algebraic variety \( V \), \( \dim(\mathbb{K}(V)) = \dim(V) \), where \( \dim(V) \) is the maximal dimension of the manifolds in the canonical Whitney stratification of \( V \).

Even though a variety \( V \) is completely characterized by the ideal \( J_R(Z_R(S)) \), the subset \( \sqrt{\langle S \rangle} \) still carries useful information.

**Proposition 2.10.** If \( S \subset \mathbb{Q}[x_1, \ldots, x_n] \), then

\[
\dim(Z_R(S)) = \dim(\mathbb{R}_{\text{alg}}(Z_{\mathbb{R}_{\text{alg}}}(S))) = \dim(\mathbb{Q}[x_1, \ldots, x_n]/\sqrt{\langle S \rangle}_\mathbb{Q}) .
\]

**Proof.** By [3, Prop. 7], \( \dim(\mathbb{R}_{\text{alg}}(Z_{\mathbb{R}_{\text{alg}}}(S))) = \dim(\mathbb{Q}[x_1, \ldots, x_n]/\sqrt{\langle S \rangle}) \). Furthermore, the Krull dimension is preserved under field extensions, see e.g. [18, II. Ex. 3.20]. Hence \( \dim(\mathbb{R}_{\text{alg}}(Z_{\mathbb{R}_{\text{alg}}}(S))) = \dim(\mathbb{R}(Z_R(S))) \), which coincides with \( \dim(Z_R(S)) \).

Related to the dimension are the singularities of a variety \( V \subset \mathbb{K}^n \). Express \( J(V) = \langle f_1, \ldots, f_k \rangle \) with appropriate polynomials \( f_1, \ldots, f_k \in \mathbb{K}[x_1, \ldots, x_n] \). The Jacobian of \( V \) at the point \( a \in V \) then is given by the matrix

\[
\text{Jac}_a(f_1, \ldots, f_k) = \begin{pmatrix}
\partial_{x_1} f_1(a) & \ldots & \partial_{x_n} f_1(a) \\
\vdots & \ddots & \vdots \\
\partial_{x_1} f_k(a) & \ldots & \partial_{x_n} f_k(a)
\end{pmatrix} .
\]
A singular point or singularity of $V$ now is a point $a \in V$ such that
$$\text{rk} \left( \text{Jac}_a(f_1, \ldots, f_k) \right) < n - \dim(V).$$
Note that this condition does not depend on the choice of generators $f_1, \ldots, f_k$ for $J(V)$. Conversely, a non-singular point is a point $a \in V$ where
$$\text{rk} \left( \text{Jac}_a(f_1, \ldots, f_k) \right) = n - \dim(V).$$
The set of all singular points of $V$ is called the singular locus of $V$ and is denoted $\text{Sing}(V)$. One immediately checks that $\text{Sing}(V)$ is again a variety by observing that
$$\text{rk} \left( \text{Jac}_a(f_1, \ldots, f_k) \right) < n - r$$
if and only if all $(n - r)$-minors of $\text{Jac}_a(f_1, \ldots, f_k)$ vanish.

For our purposes, it will be more convenient to use the following characterization of singularities which follows from [5, Prop. 3.3.10].

**Theorem 2.11.** If $V \subset \mathbb{R}^n$ is a variety such that

(i) $\dim(V) = n - 1$, and

(ii) $V = Z(f)$ for an irreducible polynomial $f$,

then $a \in V$ is a singular point if and only if $\left( \frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a) \right) = (0, \ldots, 0)$.

By subvariety of a variety $V$ one understands a subset $W \subset V$ which is itself a variety. The subvariety is proper if $W \subsetneq V$. A variety $V$ is called reducible if it is the union of two or more proper non-empty subvarieties, otherwise it is called irreducible. This suggests decomposing varieties into their irreducible components i.e. expressing $V$ as a union $V_1 \cup \ldots \cup V_s$ where each $V_i$ is irreducible. The irreducible decomposition of a variety is finite and unique. It can be obtained by finding the minimal prime ideals $P_1, \ldots, P_s$ over $J(V)$. In this case, the irreducible decomposition of $V$ is given by $V = Z(P_1) \cup \ldots \cup Z(P_s)$.

**Proposition 2.12** (cf. [5, Thm 2.8.3]). Given $S \subset \mathbb{Q}[x_1, \ldots, x_n]$ let $V = Z_{\mathbb{R}}(S)$. If $P_1, \ldots, P_s$ are the minimal prime ideals over $\sqrt{S}$, then each prime ideal $P_i$ is equal to $J_{\mathbb{Q}}(V_i)$ for some variety $V_i \subset \mathbb{R}^n$. Furthermore, $V_1 \cup \ldots \cup V_s$ is the irreducible decomposition of $V$.

3. **Learning the Underlying Variety.** In this section, we consider the following learning problem:

(LP) Given a dataset $\Omega = (a_1, \ldots, a_m)$ of points in $[0,1]^n$ sampled from a variety $V \subset \mathbb{R}^n$, find that variety.

We interpret this as the optimization problem:

(OP) For a fixed class of varieties $V$ and an objective function $A : V \times ([0,1]^n)^m \to \mathbb{R}^+$, given a dataset $\Omega = (a_1, \ldots, a_m)$ of points in $[0,1]^n$ find $\arg \max_{V \in \mathcal{V}} A(V, \Omega)$.

One way to find a suitable objective function $A$ is by following the Bayesian machine learning paradigm [2]. Instead of working with varieties directly, we use the observation from Proposition 2.2 that every real variety $V$ can be expressed as $Z(f)$ for a single polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$. Therefore, one can single out a class of polynomials $\Theta \subset \mathbb{R}[x_1, \ldots, x_n]$ such that each variety $V \in \mathcal{V}$ is defined by some polynomial $f \in \Theta$. Given a posterior probability distribution $p(f|\Omega)$ over $\Theta$, we can define an objective function
$$A(V, \Omega) := \max_{f \in \Theta, Z(f)=V} p(f|\Omega).$$
The value $A(V, \Omega)$ is to be interpreted as being proportional to the probability that $V$ is the variety from which $\Omega$ was sampled. We need to take a maximum in this definition of $A$ since different polynomials in $\Theta$ may define the same variety $V$. With this choice of $A$, the learning problem $\arg \max_{V \in V} A(V, \Omega)$ reduces to the Maximum A Posteriori (MAP) problem

$$\arg \max_{f \in \Theta} p(f|\Omega).$$

Assume that we are given a likelihood distribution $p(x|f)$ over $[0,1]^n$ and a prior distribution $p(f)$ over $\Theta$ so that $\frac{1}{\kappa} p(x|f)p(f)$ for a normalization constant $\frac{1}{\kappa}$. If furthermore the samples $\Omega = (a_1, \ldots, a_m)$ are independent and identically distributed (IID) [2], then the MAP problem is explicitly given by

$$\arg \max_{f \in \Theta} \frac{1}{\kappa} \prod_{i=1}^m p(a_i|f)p(f)^m.$$

The likelihood $p(x|f)$ should be roughly thought of as the probability of sampling $x \in [0,1]^n$ if the true underlying variety is $Z(f)$. To be robust towards noise and outliers, we allow points sampled from $Z(f)$ to be near $Z(f)$ even if they are not exactly on it. So instead of being supported on $Z(f)$, $p(x|f)$ should ideally depend on the distance of $x$ to $Z(f)$.

**Remark 3.1.** In the remainder of this section we explore one way to obtain these probabilities. It should be emphasized that this is only one particular approach within the general theory discussed in this paper. Further approaches and more detailed comparions will be followed up in future work.

The most obvious notion of distance here is the **geometric distance**

$$d_G(x, f) = \inf_{y \in Z(f)} \|x - y\|_2.$$

However, working with $d_G$ can be intractable, especially for optimization purposes. For a discussion on the complexity of this problem see [15].

Instead, we use a relaxation called the **algebraic distance** [16] given by

$$d_A(x, f) = |f(x)|.$$

By Lojasiewicz’s inequality [23, Sec. 18, Thm. 2] (see also [4, 10]), the algebraic distance is an upper bound on the geometric distance. More precisely, Lojasiewicz’s inequality entails the following result.

**Proposition 3.2.** Given a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ such that $Z(f) \cap [0,1]^n \neq 0$, there exist $a, C > 0$ with $0 < a < \frac{1}{2}$ such that

$$d_G(x, f) \leq C d_A(x, f)^a \quad \text{for all } x \in [0,1]^n.$$

**Remark 3.3.** (a) The compactness requirement in Lojasiewicz’s inequality is the main reason why we restrict our datasets to be contained in the hypercube $[0,1]^n$.

(b) By Bierstone-Milman’s version of Lojasiewicz’s inequality [4, Thm. 6.4] and subanalyticity of the Euclidean distance to a subanalytic set [4, page 6] one concludes that the algebraic and the geometric distances are actually equivalent. This means that in addition to the constants $a, C > 0$ of the proposition there exist constants $c, B > 0$ such that

$$d_A(x, f) \leq B d_G(x, f)^c \quad \text{for all } x \in [0,1]^n.$$
The degree
tion constant
where
the magnitude of the coefficients, we restrict to the class
the normalization factor
10 EZZEDDINE EL SAI AND PARKER GARA AND MARKUS J. PFLAUM
The space \( \Theta \)
N
f
is a polynomial
This operation can be expressed in matrix notation as
are elements of
U
where
R
c
as a vector in terms of its coefficients (\( f(a_1), \ldots, f(a_m) \)). This operation can be expressed in matrix notation as
\[
\begin{pmatrix}
f(a_1) \\
\vdots \\
f(a_m)
\end{pmatrix}
= U \begin{pmatrix}
c_1 \\
\vdots \\
c_N
\end{pmatrix},
\]
where \( U \) is the multivariate Vandermonde matrix [7] defined by \( U_{ij} = x^{\alpha_i}(a_j) \).
One way to capture the inverse relationship between probability and distance is by taking the likelihood to be
\[
p(x|f) = \frac{1}{r(f)} e^{-d_\Lambda(x,f)^2} = \frac{1}{r(f)} e^{-f(x)^2}
\]
where \( r(f) \) is the normalization factor \( \int_{[0,1]^n} e^{-f(x)^2} dx \). Since the algebraic distance depends on the magnitude of the coefficients, we restrict to the class
\[
\Theta_D := \{ f \in \mathbb{R}[x_1, \ldots, x_n] | \text{Deg}(f) \leq D, \| F^{-1}(f) \| = 1 \},
\]
where \( \| F^{-1}(f) \| \) coincides with the norm on the coefficients \( \sqrt{[c_1, \ldots, c_N)(c_1, \ldots, c_N)^T} \).
The space \( \Theta_D \) has an obvious parametrization \( F_{|S^{N-1}} : S^{N-1} \to \Theta_D \) which takes a point \( (c_1, \ldots, c_N) \) in the sphere \( S^{N-1} \) to the polynomial \( c_1 x^{\alpha_1} + \ldots + c_{N-1} x^{\alpha_{N-1}} + c_N \in \Theta_D \).
The degree bound \( D \) should be treated as a hyperparameter. Choosing the hypothesis class \( \Theta_D \) restricts us to the class \( \mathcal{V}_D \) of varieties defined by a single polynomial of degree \( \leq D \) whose coefficients have norm 1. As with the standard machine learning set-up, there is a trade-off where setting \( D \) higher leads to better approximation but higher risk of over-fitting. This issue is explored in Section 6.
We define a prior distribution on \( \Theta_D \) by \( p(f) = \frac{1}{\beta} r(f) \) where \( \beta \) is the normalization constant
\[
\beta = \int_{S^{N-1}} r(F(y))dy = \int_{S^{N-1}} \int_{[0,1]^n} e^{-(F(y)(x))^2} dxdy.
\]
This prior disfavors polynomials with low values of \( r(f) \). Intuitively, such polynomials have high values of \( \int_{[0,1]^n} |f(x)| \, dx \), which means that, with respect to the algebraic distance, these polynomials have a large total distance from the domain \([0,1]^n\). With this prior, the joint distribution simplifies to \( p(x)p(f) = \frac{1}{\beta} e^{-f(x)^2} \).

With this choice of likelihood and prior distributions, the MAP problem is given by:

\[
\arg\max_{f \in \Theta_D} \frac{1}{K} \prod_{i=1}^{m} \frac{1}{\beta} e^{-f(a_i)^2} = \arg\max_{f \in \Theta_D} \frac{1}{K} \left( \frac{1}{\beta^m} e^{-\sum_{i=1}^{m} f(a_i)^2} \right)
\]

\[
= \arg\max_{f \in \Theta_D} \log \left( \frac{1}{K} \left( \frac{1}{\beta^m} e^{-\sum_{i=1}^{m} f(a_i)^2} \right) \right)
\]

\[
= \arg\max_{f \in \Theta_D} -\log(K) - \log(\beta^m) - \sum_{i=1}^{m} f(a_i)^2
\]

\[
= \arg\min_{f \in \Theta_D} \sum_{i=1}^{m} f(a_i)^2.
\]

This is equivalent to the problem

\[
\arg\min_{c \in \mathbb{R}^{N-1}} (c_1, \ldots, c_N)^t U^t U (c_1, \ldots, c_N)^t
\]

which is a convex Quadratically Constrained Quadratic Program (QCQP) \([6]\). The solutions are the normalized elements of the eigenspace \( E_\lambda \) where \( \lambda \) is the smallest eigenvalue of \( U^t U \).

**Remark 3.4.** More generally, we can think of likelihood distributions of the form \( e^{-\lambda f(x)^2} \) where \( \lambda \) accounts for a variance-like term with \( \lambda \in (0, \infty) \). In this case, by taking derivatives with respect to \( \lambda \), optimization with respect to \( \lambda \) is obtained by solving the following equation for \( \lambda \):

\[
\frac{\int_{S^{N-1}} \int_{[0,1]^n} (F(y)(x))^2 e^{-\lambda(F(y)(x))^2} \, dx \, dy}{\int_{S^{N-1}} \int_{[0,1]^n} e^{-\lambda(F(y)(x))^2} \, dx \, dy} = \frac{\sum_{i=1}^{m} f(a_i)^2}{m}.
\]

Intuitively, this says that \( \lambda \) should be chosen so that over all the distributions coming from \( \Theta_D \), the average variance agrees with the empirical variance \( \frac{\sum_{i=1}^{m} f(a_i)^2}{m} \).

On the other hand, optimization with respect to \( f \) is again given by

\[
\arg\min_{f \in \Theta_D} \sum_{i=1}^{m} f(a_i)^2.
\]

We are currently not aware of a computationally efficient way to infer \( \lambda \) from the data, so we leave a more detailed consideration of this issue for later work.

In the case where \( \lambda > 0 \), the matrix \( U^t U \) is positive definite, in which case the QCQP is strictly convex \([6]\), so there is essentially one unique MAP solution \( \hat{f} \). In the case where \( \lambda = 0 \), the MAP solutions are the normalized elements of \( E_\lambda = \ker(A^t A) \). Here, every MAP solution \( \hat{f} \) vanishes exactly on \( \Omega \) and \( p(\hat{f}|\Omega) = \frac{1}{\beta^m} \).

Therefore, under the assumptions that the hypothesis class is \( \mathcal{V}_D \), the objective function is \( A(V, \Omega) = \max_{f \in \Theta_D} \mathcal{V}(f|\Omega) \), and the posterior distribution is \( p(\hat{f}|\Omega) = \frac{1}{\pi} \left( \frac{1}{\beta^m} e^{-\sum_{i=1}^{m} f(a_i)^2} \right) \) on \( \Theta_D \), the solutions to the learning problem are the varieties
Z(\hat{f}) for every normalized element \( \hat{f} \in E_\lambda \). We call this the MAP model and summarize it in the algorithm below.

\begin{algorithm}[h]
\caption{MAP Model}
\begin{algorithmic}
\State \textbf{Input:} a dataset \( \Omega \in ([0, 1]^n)^m \) and a degree bound \( D \).
\State \textbf{Output:} a polynomial \( \hat{f} \) in \( \Theta_D \), such that \( Z(\hat{f}) \) solves the learning problem under the above assumptions.
\State Fix an ordering and list all homogeneous monomials \( x^{\alpha_1}, \ldots, x^{\alpha_N} \) of degree \( \leq D \).
\State Compute the multivariate Vandermonde matrix \( U_{ij} = x^{\alpha_i}(a_j) \).
\State Find the smallest eigenvalue \( \lambda \) of \( U^T U \) and its corresponding eigenspace \( E_\lambda \).
\State \textbf{return:} any normalized element \( \hat{f} \) of \( E_\lambda \).
\end{algorithmic}
\end{algorithm}

3.1. Expanding on the MAP Model. One way to refine the result for the case \( \lambda = 0 \) is to enlarge our hypothesis class. If \( \lambda = 0 \) and \( f_1, \ldots, f_k \) is a normalized basis for \( \ker(U^T U) \), then \( \hat{f} := f_1^2 + \ldots + f_k^2 \) is a degree 2D polynomial whose zero-set satisfies
\[ Z(\hat{f}) = Z(\{f_1, \ldots, f_k\}) = \bigcap_{f \in \ker(U^T U)} Z(f). \]
That is, \( \hat{f} \) defines the smallest variety given by a set of polynomials of degree \( \leq D \). This method of taking intersections changes the hypothesis class and may no longer yield an MAP solution over \( \Theta_{2D} \). However, this method yields a less redundant variety than the MAP solutions over \( \Theta_D \) without the need to preform any further optimization. We call this the intersected MAP model. It is summarized in the algorithm below.

\begin{algorithm}[h]
\caption{Intersected MAP Model}
\begin{algorithmic}
\State \textbf{Input:} a dataset \( \Omega \in ([0, 1]^n)^m \) and a degree bound \( D \).
\State \textbf{Output:} a polynomial \( \hat{f} \) in \( \Theta_{2D} \), such that \( Z(\hat{f}) \) is the intersection of all solutions in \( V_D \) to the learning problem under the previous assumptions.
\State Fix an ordering and list all homogeneous monomials \( x^{\alpha_1}, \ldots, x^{\alpha_N} \) of degree \( \leq D \).
\State Compute the multivariate Vandermonde matrix \( U_{ij} = x^{\alpha_i}(a_j) \).
\State Find the smallest eigenvalue \( \lambda \) of \( U^T U \) and its corresponding eigenspace \( E_\lambda \).
\If {\( \lambda > 0 \)}
\State \textbf{return:} the (essentially) unique normalized element \( \hat{f} \) of \( E_\lambda \)
\Else
\State Find a basis \( f_1, \ldots, f_k \) for \( \ker(U^T U) \).
\State \textbf{return:} \( \hat{f} := f_1^2 + \ldots + f_k^2 \).
\EndIf
\end{algorithmic}
\end{algorithm}

4. Algebraic Computations. Assume that \( Z(f) \) for \( f \) a polynomial in \( n \) real variables is the true underlying variety for the data set \( \Omega \). We can use tools from commutative algebra to reveal information about the geometry of \( Z(f) \subset \mathbb{R}^n \). This analysis can be automated with the use of Gröbner bases which are particular and computationally powerful types of generating sets for ideals \( I \subset \mathbb{K}[x_1, \ldots, x_n] \). We do not state the precise definition of a Gröbner basis here, but refer the reader to the excellent exposition in [17, Sec. 1.6]. Moreover, [17] provides a general overview on
computational commutative algebra. To use Gröbner basis methods in a computer algebra system like SINGULAR [11], we have to change the base field from \( \mathbb{R} \) to \( \mathbb{Q} \). If \( f \) was obtained through a numerical procedure such as the MAP learning model, then the floating point coefficients of \( f \) can be interpreted as rational numbers.

With Gröbner basis methods one can compute a generating set \( f_1, \ldots, f_k \) for the ideal
\[
\sqrt[\mathbb{Q}]{\langle f \rangle} \subset \mathbb{Q}[x_1, \ldots, x_n].
\]
Using this generating set we can construct the ring \( \mathbb{Q}[x_1, \ldots, x_n] / \sqrt[\mathbb{Q}]{\langle f \rangle} \) which is a subring of the coordinate ring \( \mathbb{R}(Z_\mathbb{R}(f)) \). By Proposition 2.10,
\[
\dim(Z_\mathbb{R}(f)) = \dim(\mathbb{Q}[x_1, \ldots, x_n] / \sqrt[\mathbb{Q}]{\langle f \rangle})
\]
which can be computed using Gröbner bases. Similarly, we can compute the minimal prime ideals over \( \sqrt[\mathbb{Q}]{\langle f \rangle} \). By Proposition 2.12, these are the ideals \( J_\mathbb{Q}(V_1), \ldots, J_\mathbb{Q}(V_s) \), where \( V_1, \ldots, V_s \) are the irreducible components of \( V \).

It should be noted, however, that even over \( \mathbb{Q} \), Gröbner basis computations are in general very costly and may only be feasible for small \( n \) and \( D \), hence the need for numerical computations. For more details on the complexity of finding Gröbner bases see [20].

**Example 4.1.** We can apply these concepts to the variety \( V \) from Example 2.8 using the following SINGULAR code.

```sage
// Define R = QQ[x,y,z] with lexicographic ordering.
ring R = 0,(x,y,z),lp;
poly f = ((x-1/2)^2 + (y-1/2)^2 + (z-1/2)^2 - 1/4)*(x-y);
ideal I = f;
LIB "realrad.lib";
ideal I2 = realrad(I);
size(reduce(I,I2));
//->0
size(reduce(I2,I));
//->0
LIB "primdec.lib";
minAssGTZ(I2);
//->[1]:
//-> _[1]=2x2-2x+2y2-2y+2z2-2z+1
//->[2]:
//-> _[1]=x-y
dim(I2);
//->2
```

First we define the ideal \( I = \langle ((x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 + (z - \frac{1}{2})^2 - \frac{1}{4}) \cdot (x-y) \rangle \) over \( \mathbb{Q} \). We then compute the real radical using the library `realrad.lib` [34] and observe that in fact \( I = \sqrt[\mathbb{Q}]{T_0} \). As expected, we find that \( \dim(\mathbb{Q}[x_1, \ldots, x_n] / \sqrt[\mathbb{Q}]{T_0}) = 2 \) which coincides with \( \dim(V) \). Using the library `primdec.lib` we also determine the minimal primes above \( \sqrt[\mathbb{Q}]{T_0} \) to be the vanishing ideal of the sphere of radius \( \frac{1}{2} \) centered around \( (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \) and the vanishing ideal of the plane \( x - y \). This agrees with the decomposition of \( V \).
5. **Numerical Computations.** Again, assume that \( Z(f) \) is the true underlying variety for \( \Omega \). We can also use numerical methods to study the geometry of \( Z(f) \) by sampling new points from \( Z(f) \) and working directly with those samples. This approach has the advantage of being computationally more tractable than the algebraic computations relying on Gröbner bases.

One reason to obtain a new sample set \( \Omega' \) from \( Z(f) \cap [0,1]^n \) is that if \( Z(f) \) is only an approximate fit, then the sample \( \Omega' \) will reflect the geometry of \( Z(f) \) more closely than \( \Omega \). Hence, if we are specifically studying the model \( Z(f) \), generating a sample set \( \Omega' \) can lead to more accurate results.

First notice that the likelihood distributions \( p(x|f) \) on \([0,1]^n\) can be used to generate data. This can be achieved using rejection sampling which is outlined in Algorithm 3. This sampling process reveals the underlying assumptions that the model makes about how the original data set \( \Omega \) was generated and how noise was introduced.

However, capturing the model’s assumptions also captures the noisy process through which \( \Omega \) was supposedly generated. If the likelihood distribution depends on the algebraic distance, such as the distribution used in Section 3, then this noise can be avoided by fixing a small \( \eta > 0 \) and accepting a point \( a \in [0,1]^n \) if and only if \( d_A(a, Z(f)) = |f(a)| < \eta \). We call this direct sampling and summarize it in Algorithm 4.

Setting a smaller \( \eta \) threshold reduces the sampling noise, however this comes at the cost of increasing the efficiency of the sampling. This still works reasonably well for low values of \( n \), but it does not scale well to higher dimensions. For higher dimensions, a more effective sampling method is given in [12]. Alternatively, the variety could be sampled using Homotopy Continuation [8], which is a numerical method for computing the zero-set of a system of polynomials. In Homotopy Continuation, one begins with a simple system of polynomials whose roots are known, and then defines a homotopy, i.e. a continuous deformation, from the simple system to the system of polynomials that one is trying to solve. This method relies on tracking the paths that the roots take as the homotopy is being applied.

**Algorithm 3: Rejection Sampling**

- **Input:** a polynomial \( f \), a likelihood distribution \( p(x|f) \), and a target number of samples \( m \).
- **Output:** a set \( \Omega' \) of \( m \) points sampled from \( Z(f) \cap [0,1]^n \) according to \( p(x|f) \).

Initialize \( \Omega' \) to \( \emptyset \).

while \( |\Omega'| < m \) do
  Draw a random point \( a \) from the uniform distribution on \([0,1]^n\).
  Draw a random number \( \alpha \) from \([0,1]\).
  if \( \alpha < p(a|f) \) then
    Accept: \( \Omega' \leftarrow \Omega' \cup \{a\} \).
  else
    Reject.
  end
end

return: \( \Omega' \)
Let $\Omega'$ be a set of samples from $Z(f) \cap [0,1]^n$ obtained using direct sampling with an accuracy threshold of $\eta$. Under the hypothesis that the distribution $p(x|f)$ depends on $d_2^2(x,f)$, we propose a method for finding the points in $\Omega'$ near $\text{Sing}(Z(f))$. First, by Theorem 2.11, if $V$ is $n-1$ dimensional and $f$ is irreducible, then every point $b \in [0,1]^n$ which lies in the singular locus $\text{Sing}(Z(f))$ satisfies

$$f(b) = \partial_{x_1} f(b) = \ldots = \partial_{x_n} f(b) = 0.$$ 

By continuity of the map $\|\nabla f(x)\|_2 = \sqrt{(\partial_{x_1} f(x))^2 + \ldots + (\partial_{x_n} f(x))^2}$ there exists for every $\varepsilon > 0$ an open neighborhood $U$ of $b$ such that all points $a \in U \cap \Omega'$ satisfy $\|\nabla f(a)\|_2 < \varepsilon$.

Note, however, that the converse of this need not be true meaning that a point $a \in \Omega'$ satisfying $\|\nabla f(a)\|_2 < \varepsilon$ is not necessarily near a point $b \in \text{Sing}(Z(f))$. For example, consider the polynomial $f \in \mathbb{R}[x,y]$ given by $f(x,y) = x^2$. The gradient $\nabla f$ vanishes exactly on the line $Z(x) = \{(x,y) \in \mathbb{R}^2 \mid x = 0\}$ which is smooth and coincides with the variety $Z(f)$. Hence the zero locus of the gradient is nowhere close to the singular locus of the variety $Z(f)$ which is empty by smoothness of $Z(f) = Z(x)$.

Nevertheless, assuming the converse appears to be justified for computational purposes and it provides a powerful heuristic method for detecting singularities. More specifically, we can heuristically assume that regardless of the dimension of $V$ and the reducibility of $f$, if $\varepsilon > 0$ is small enough, then $\{a \in \Omega' \mid \|\nabla f(a)\|_2 < \varepsilon\}$ is a set of points at or near $\text{Sing}(Z(f)) \cap [0,1]^n$. We will denote this set by $\text{Sing}(\Omega')$ and call the described method the singularity heuristic. Clearly, the accuracy of this method depends on the magnitudes of $\eta$ and $\varepsilon$ and the density of the sample set $\Omega'$. We explore the efficacy of the singularity heuristic in Section 6.

Another crucial numerical task is to test if two varieties $Z(f)$ and $Z(g)$ are equal. Note that the equality

$$Z(f) = Z(g) \quad (5.1)$$

does in general not entail the polynomials $f$ and $g$ to be equal as for example the choice $f = x - y$ and $g = (x^2 + y^2) \cdot (x - y)$ shows. Furthermore, testing the equality (5.1) is obviously not possible if one of the polynomials, say $g$, is unknown and we only have access to samples $\Lambda$ from $Z(g) \cap [0,1]^n$. One solution is to work entirely in terms of samples. Namely, take a set of samples $\Omega'$ from $Z(f) \cap [0,1]^n$ and directly compare $\Lambda$ with $\Omega'$. To compare $\Lambda$ with $\Omega'$, we use the Wasserstein distance [32],

### Algorithm 4: Direct Sampling

**Input:** a polynomial $f$ and a target number of samples $m$.  
**Output:** a set $\Omega'$ of $m$ points sampled from $Z(f) \cap [0,1]^n$.  

Initialize $\Omega'$ to $\emptyset$.  

while $|\Omega'| < m$ do  

Draw a random point $a$ from the uniform distribution on $[0,1]^n$.  

if $|f(a)| < \eta$ then  

Accept: $\Omega' \leftarrow \Omega' \cup \{a\}$.  
else  

Reject.  
end  

end  

return: $\Omega'$.
Algorithm 5: Singularity Heuristic

**Input:** a polynomial \( f \), a set of samples \( \Omega' \) from \( Z(f) \cap [0,1]^n \), and a singularity threshold \( \varepsilon \).

**Output:** \( \text{Sing}(\Omega') \), a subset of \( \Omega' \), heuristically assumed to be near \( \text{Sing}(Z(f)) \).

Initialize \( \text{Sing}(\Omega') \) to \( \emptyset \).

Find the partial derivatives \( \partial_{x_1} f(x), \ldots, \partial_{x_n} f(x) \), for \( a \in \Omega' \) do

- if \( \|\nabla f(a)\|_2 < \varepsilon \) then
  - Accept \( a \)
  - \( \text{Sing}(\Omega') \leftarrow \text{Sing}(\Omega') \cup \{a\} \).
- else
  - Reject.

end

Return: \( \text{Sing}(\Omega') \).

Figure 6.1. Samples from \( V \) and from the learned variety

which measures the cost of the optimal transport taking the point cloud \( \Lambda \) to the point cloud \( \Omega' \). If we have \( N \) samples from \( \mathbb{R}^n \) this is simply

\[
W(\Omega', \Lambda) = \min_{\pi \in S_N} \left\{ \sum_{i=1}^{N} \|\Omega'_i - \Lambda_{\pi(i)}\|_2 \right\},
\]

where \( S_N \) denotes the group of permutations of the integers \( \{1, 2, \ldots, N\} \). We can similarly apply the Wasserstein distance in order to compare a set of samples from \( \text{Sing}(Z(g)) \cap [0,1]^n \) with the set \( \text{Sing}(\Omega') \).

6. Results. In this section, we test the MAP model from Section 3 and the singularity heuristic from Section 5.

Starting with the variety \( V \) from Example 2.8, we used the parametric form of \( V \) to produce a sample set \( \Omega \) consisting of 1600 points sampled from \( V \cap [0,1]^3 \) (Figure 6.1a). We also used the parametric form of \( \text{Sing}(V) \) to produce a sample set, call it \( \text{Sing}(\Omega) \), of 400 points sampled from \( \text{Sing}(V) \) (Figure 6.2a). Plotted in Figure 6.1c is the learned variety.
(a) Figure 6.2. Samples from the singular locus of $V$ and results from the singularity heuristic

(b) Figure 6.3. MAP model and singularity heuristic performance (noise-free)

We applied the MAP model with $D = 3$. After rounding and scaling, we obtained exactly the target polynomial from Example 2.8

$$x^3 - x^2 y + x^2 + xy^2 + xz^2 - xz + \frac{1}{2} x^3 - y^2 + y^2 - yz^2 + yz - \frac{1}{2} y =$$

$$= (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 + (z - \frac{1}{2})^2 - \frac{1}{4} \cdot (x - y) = f.$$

By direct sampling with $\eta = 0.001$ we produced a data set $\Omega'$ consisting of 1600 points sampled from $Z(f) \cap [0,1]^3$ (Figure 6.1b). Using these samples, we applied the singularity heuristic with value $\varepsilon = 0.02$ and obtained a set $\text{Sing}(\Omega')$ consisting of 232 points (Figure 6.2b).

To test the MAP model under more general conditions, we used the dataset $\Omega$ as above and tested the MAP model for different values of $D$. For a more quantitative picture, we computed the Wasserstein distance between $\Omega$ and 1600 points obtained through direct sampling from $Z(f) \cap [0,1]^3$ for each learned polynomial $f$ and for different values of $\eta$. We repeated each test 3 times. The average values are given in Figure 6.3a, where $\log_{10}(\eta)$ is the variable along the horizontal axis and the Wasserstein distance $W(\Omega, \Omega')$ is the variable along the vertical axis.
As one can see, increasing $\eta$ largely results in an increase in the Wasserstein distance. However, very low values of $\eta$ seem to perform worse than higher values. One explanation for this is that direct sampling does not produce a uniform set of samples over $Z(f)/[0,1]^3$. This can be seen for example in Figure 6.1b. Nonetheless, we see that the overall minimal distance is attained with $D = 3$ which is consistent with the fact that 3 is the lowest degree of any single polynomial defining $V$.

To test the singularity heuristic, we used the above samples and learned polynomials. For each value of $D$, we selected the best performing value of $\eta$ and applied the singularity heuristic to the 3 sample sets corresponding to that value of $D$ and $\eta$. Then we measured the Wasserstein distance between $\text{Sing}(\Omega)$ and the set of points $\text{Sing}(\Omega')$ which passed the singularity heuristic. The average values are given in Figure 6.3b. The missing results are the cases where $\text{Sing}(\Omega')$ is empty. In Figure 6.3c we give the numbers of points that do pass the singularity heuristic. Once again, we see that the overall minimal distance is attained with $D = 3$.

To examine the effect of noise on the MAP model, we repeated the same process but added Gaussian noise with a standard deviation of $0.025$ to the original data $\Omega$. The result of this is illustrated in Figure 6.1c. Samples from the learned polynomials were compared to the original noise-free data in order to test the robustness towards noise. The results are given in Figures 6.4a, 6.4b, and 6.4c.

The values $D = 1, 2, 3$ show similar trends with the overall minimal distances still attained at $D = 3$. However the values $D = 4, 5$ show significantly worse performance which is due to over-fitting to noise. This is to be expected since the models for $D = 4, 5$ contain many more additional parameters which lead to higher model complexity.

**Example 6.1.** Cyclooctane is a cyclic molecule with chemical formula $(\text{CH}_2)_8$ (Figure 6.5). The full energy landscape of cyclooctane in the sense of Born-Oppenheimer contains the positions of the carbon and hydrogen atoms and therefore has 72 dimensions which is too high dimensional to be studied effectively by computational or numerical means. The **conformation space of cyclooctane** is a reduced energy landscape which consists of the chemically allowable positions for the eight carbon atoms differing from the lowest energy state only by rotation around single bonds. By geometric reasons, see e.g. [7], the conformation space therefore is the variety...
Figure 6.5. Cyclooctane

cut out by a set of 16 equations in \( \mathbb{R}[x_1, y_1, z_1, \ldots, x_8, y_8, z_8] \):

\[
(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2 - c = 0 \quad \text{and} \quad
(x_i - x_{i+2})^2 + (y_i - y_{i+2})^2 + (z_i - z_{i+2})^2 - \frac{8}{3}c = 0 \quad \text{for all } i \in \mathbb{Z}_8.
\]

The constant \( c \) is the bond length between two neighboring carbon atoms. By computational means, it has been determined to have the value \( c \approx 2.21 \) \cite{7}.

A dataset of 6040 points sampled from this variety was produced by \cite{24}. The variety was further reduced to a subspace of \( \mathbb{R}^8 \). The reduction results in a compact 2-dimensional singular surface which we will refer to as the reduced conformation space of cyclooctane. The specific dimensionality reduction map consists of trigonometric functions so in particular is analytic; see \cite{24} for details. We are not aware of any proof that the reduced conformation space is also a variety, however, it is clear from the construction that the reduced conformation space is at least a compact subanalytic set by analyticity of the reduction map \cite{4}. Applying the reduction to the original 6040 points, one obtains a set \( \Omega \) of points lying on the reduced conformation space of cyclooctane.

We are particularly interested in the singular locus of the reduced conformation space. Based upon geometric considerations and the sample tests discussed in \cite{24}, the reduced conformation space of cyclooctane is believed to be topologically equivalent to two Möbius bands and a sphere glued together along two disjoint circles. Those two circles form the boundaries of the two Möbius bands and constitute the singular locus of the reduced conformation space. Chemically, the conformations represented by the two circles are called peak (P) and saddle (S) and correspond to maximal and minimal energy conformations, respectively, \cite{24}. Using this information, we applied an adhoc method to extract the set of singular points from \( \Omega \) resulting in a set \( \text{Sing}(\Omega) \) of 233 points shown in Figures 6.6a, 6.6b and 6.6c for different coordinate axes.

We applied the MAP model to the reduced conformation space for various values of \( D \) but \( D = 4 \) seemed to preform best. From the learned polynomial, we sampled a set \( \Omega' \) of 40000 points at threshold value \( \eta = 10^{-7} \). We applied the singularity heuristic at a threshold value of \( \varepsilon = 0.0003 \) and obtained a set \( \text{Sing}(\Omega') \) of 235 points. The set \( \text{Sing}(\Omega') \) is shown in Figures 6.7a, 6.7b and 6.7c.

The Wasserstein distance between \( \text{Sing}(\Omega) \) and \( \text{Sing}(\Omega') \) was found to be 0.022. We can see that despite the presence of noise coming from the sampling, the singularity heuristic succeeds to capture the overall geometry of the singular locus of the reduced conformation space of cyclooctane. This is further evidence that the
reduced conformation space of cyclooctane is likely an algebraic variety and is likely defined by a polynomial of degree 4.

![Figure 6.6](image1)

**Figure 6.6.** Sampled singular locus of the conformational space of $(\text{CH}_2)_8$

![Figure 6.7](image2)

**Figure 6.7.** Learned singular locus of the conformational space of $(\text{CH}_2)_8$

**Remark 6.2.** The conformation space of cyclooctane has also been analyzed by Breiding et al. in [7], but the approach taken there is different. In [7], the authors do not consider the reduced conformation space but instead apply a mix of algebraic and topological data analysis methods to the unreduced conformation space. Based upon Martin et al.'s dataset [24] sampled from the unreduced conformation space, they determine in particular the dimension of the conformation space and compute the persistent homology of the dataset. This approach returns the dimension of the conformation space to be 2 and Betti numbers indicating that the conformation space indeed should have a singular locus homeomorphic to the disjoint union of two circles. Note, however, that this method does not directly provide the particular location or coordinates of the singular points. Furthermore, Betti numbers do not always encode information as to whether the underlying space possesses singularities or not. Finally, we remark that the method of learning varieties presented in [7] has been applied there to the unreduced conformation space of cyclooctane, only.
Remark 6.3. Let us briefly explain what has been achieved by considering the conformational space of cyclooctane as an algebraic variety and comment on the potential broader use of singularity theory in chemistry. According to the theoretical results by Hendrickson [19] and Anet-Krane [1], cyclooctane has 10 conformations: chair, crown, boat, boat-chair, twist-boat-chair, boat-boat, twist-boat, twist-chair, chair-chair, and twist-chair-chair. The existence of all these conformations has been verified experimentally. The peak and the saddle conformations described above by the singular locus of the reduced conformational space were detected much later by the ground-breaking work [24], where sophisticated methods from topological data analysis were used to study energy landscapes and conformational spaces. We expect that to be a more general phenomenon. Singular parts of (reduced) energy landscapes or other chemically relevant surfaces such as the electron density or the electron localization function are very small and difficult to find. But they might play an important role to theoretically detect and describe new conformations or possibly even new reactions. Further example studies from chemistry are therefore intended.

7. Related Work. The MAP and intersected MAP model we developed in Section 3 are inspired by the learning method given in [7]. The method presented in [7] aims to find the kernel of the Vandermonde matrix $U$ in a numerically stable manner. However, this is presented as a purely numerical method without a given statistical basis. As such, the authors focus on the case where $\Omega$ is sampled from a variety without any anomalies present.

In the case when there is no sample noise, the method of [7] coincides with the MAP model from Section 3 since in this case $\lambda = 0$ and the MAP solutions would be given by $E_\lambda = \ker(U^T U) = \ker(U)$. In the case when there is noise, the authors use a noise tolerance parameter $\tau > 0$. In one instance, they compute $\ker(U)$ by finding the singular vectors of $U$ whose singular values are $\leq \tau$. This is similar to the QCQP from Section 3 since the eigenvectors of $U^T U$ are the right-singular vectors of $U$ and the eigenvalues of $U^T U$ are the squares of the singular values of $U$. The difference is that we only accept the smallest eigenvalue (or singular value). This can have a significant effect if the noise levels are high or the underlying space is only approximately a variety, as shown in Example 7.1 below. Intuitively, this effect appears when we have two varieties $V_1$ and $V_2$ which fit the data $\Omega$ well but not exactly. In this case, the intersection $V_1 \cap V_2$ might not be close to the data set $\Omega$ at all as shown in the following example.

Example 7.1. Consider the noisy data $\Omega$ in Figure 7.1a sampled from the plane $y - z = 0$ with noise applied across the $z$ axis. Setting $D = 1$, we find that the planes defined by the polynomials $\ell_1 = 0.73y - 0.68z - 0.03$ (7.1b) and $\ell_2 = 0.17x + 0.76y - 0.61z - 0.17$ (7.1c) are both in $\bigoplus_{\lambda \leq 0.003} E_\lambda$. Moreover, both planes are close to the original data. However, the intersection $Z(\ell_1) \cap Z(\ell_2)$ is given by the noisy line in 7.1d which has the wrong dimension.
A related approach for singularity detection is taken in [35]. This approach involves local cohomology to test how well different regions of the data set can be approximated by Euclidean space. In contrast to algebraic varieties, the authors of [35] assume that the underlying geometry is a stratified space. However, their method does not rely on explicitly learning the underlying stratified space, and instead it uses persistent cohomology [9] to detect the singular regions. This comes at a computational advantage, however, it also captures less information about the space itself. Furthermore, this method only applies to singularities where the space fails to be a topological manifold, and not the singularities where smoothness fails.

Acknowledgments. The authors gratefully acknowledge support by the NSF under award OAC 1934725 for the project DELTA: Descriptors of Energy Landscape by Topological Analysis. M.J. Pflaum thanks David Jonas and David Walba for crucial explanations on the conformations of cylooctane. The authors also thank Henry Adams, Howy Jordan and Trevor Sesnic for helpful discussions.

REFERENCES

[1] F. Anet and J. Krane, Strain energy calculation of conformations and conformational changes in cyclooctane, *Tetrahedron Letter*, 8 (1973), 5029–5052.
[2] D. Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, USA, 2012.
[3] E. Becker and R. Neuhaus, Computation of real radicals of polynomial ideals, in Computational Algebraic Geometry, vol. 109 of Progr. Math., Birkhäuser Boston, 1993, 1–20, URL https://doi.org/10.1007/978-1-4612-2752-6_1.

[4] E. Bierstone and P. D. Milman, Semianalytic and subanalytic sets, Publications mathématiques de l’IHÉS, 67 (1988), 5–42, URL https://doi.org/10.1007/bf02699126.

[5] J. Bochnak, M. Coste and M.-F. Roy, Real algebraic geometry, vol. 36 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], Springer-Verlag, Berlin, 1998, Translated from the 1987 French original, Revised by the authors.

[6] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[7] P. Breiding, S. Kališnik, B. Sturmfels and M. Weinstein, Learning algebraic varieties from samples, Revista Matemática Complutense, 31 (2018), 545–593, URL https://doi.org/10.1007/s13163-018-0273-6.

[8] P. Breiding and S. Timme, Homotopy continuation.jl: A package for homotopy continuation in julia, in Mathematical Software – ICMS 2018 (eds. J. H. Davenport, M. Kauers, G. Labahn and J. Urban), Springer International Publishing, Cham, 2018, 458–465.

[9] G. Carlsson, Topology and data, Bulletin of the American Mathematical Society, 46 (2009), 255–308.

[10] T. H. Colding and W. P. Minicozzi II, Lojasiewicz inequalities and applications, in Surveys in Differential Geometry 2014. Regularity and evolution of nonlinear equations, vol. 19 of Surv. Differ. Geom., Int. Press, Somerville, MA, 2015, 63–82, https://doi.org/10.4310/SDG.2014.v19.n1.a3.

[11] W. Decker, G.-M. Greuel, G. Pfister and H. Schönemann, Singular 4-2-0 - A computer algebra system for polynomial computations, http://www.singular.uni-kl.de, 2020.

[12] E. Dufresne, P. Edwards, H. Harrington and J. Hauenstein, Sampling real algebraic varieties for topological data analysis, in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019, 19197–19209, URL https://doi.org/10.1021/jp962746i.

[13] M. I. Dykman, V. N. Smelyanskiy, R. S. Maier and M. Silverstein, Singular features of large fluctuations in oscillating chemical systems, The Journal of Physical Chemistry, 100 (1996), 19197–19209, URL https://doi.org/10.1021/jp962746i.

[14] C. Fefferman, S. Mitter and H. Narayanan, Testing the manifold hypothesis, Journal of the American Mathematical Society, 29 (2016), 983–1049, URL https://doi.org/10.1090/jams/852.

[15] O. Gäfvert, Computational complexity of learning algebraic varieties, Advances in Applied Mathematics, 121 (2020), 102100, URL https://doi.org/10.1016/j.aam.2020.102100.

[16] W. Gander, G. H. Golub and R. Strebel, Least-squares fitting of circles and ellipses, BIT, 34 (1994), 558–578, URL https://doi.org/10.1007/bf01934268.

[17] G.-M. Greuel and G. Pfister, A Singular Introduction to Commutative Algebra, 2nd edition, Springer Publishing Company, Incorporated, 2007.

[18] R. Hartshorne, Algebraic Geometry, Graduate Texts in Mathematics, Springer New York, 2013.

[19] J. B. Hendrickson, Molecular geometry. V. Evaluation of functions and conformations of medium rings, J. Am. Chem. Soc., 89 (1967), 7047–7061.

[20] D. T. Huynh, A superexponential lower bound for Gröbner bases and Church-Rosser commutative thue systems, Information and Control, 68 (1986), 196–206, URL https://www.sciencedirect.com/science/article/pii/S0019995886800353.

[21] S. Lang, Algebra, vol. 211 of Graduate Texts in Mathematics, 3rd edition, Springer-Verlag, New York, 2002, in https://doi.org/10.1007/978-1-4613-0041-0.

[22] J. Leeuw, History of nonlinear principal component analysis, UCLA: Department of Statistics, UCLA (2013), https://escholarship.org/uc/item/1vp9f9kx.

[23] S. Lojasiewicz, Ensemble semi-analytique, 1965, Mimeographié, Institute des Hautes Études Scientifique, Bures-sur-Yvette, France.

[24] S. Martín, A. Thompson, E. A. Coutsias and J.-P. Watson, Topology of cyclo-octane energy landscape, The Journal of Chemical Physics, 132 (2010), 234115, URL https://doi.org/10.1063/1.3445267.

[25] L. McInnes, J. Healy and J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, arXiv:1802.03426 (2020).

[26] F. A. Moughari and C. Eslahchi, ADRML: Anticancer drug response prediction using manifold learning, Scientific Reports, 10.
[27] J. Nash, Real algebraic manifolds, Ann. of Math. (2), 56 (1952), 405–421.
[28] R. Neuhaus, Computation of real radicals of polynomial ideals. II, J. Pure Appl. Algebra, 124 (1998), 261–280, URL https://doi.org/10.1016/S0022-4049(96)00103-X.
[29] J. Park, Manifold Learning in Computer Vision, PhD thesis, USA, 2005, AAI3193223.
[30] M. J. Pflaum, Analytic and geometric study of stratified spaces, vol. 1768 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2001.
[31] A. Reiman, Singular surfaces in the open field line region of a diverted tokamak, Physics of Plasmas, 3 (1996), 906–913.
[32] L. Rüschendorf, The Wasserstein distance and approximation theorems, Probability Theory and Related Fields, 70 (1985), 117–129.
[33] P. Solernó, Effective Lojasiewicz inequalities in semialgebraic geometry, Applicable Algebra in Engineering, Communication and Computing, 2 (1991), 1–14, URL http://dx.doi.org/10.1007/BF01810850.
[34] S. J. Spang, On the computation of the real radical., PhD thesis, Technische Universität Kaiserslautern, 2007.
[35] B. J. Stolz, J. Tanner, H. A. Harrington and V. Nanda, Geometric anomaly detection in data, Proceedings of the National Academy of Sciences, 117 (2020), 19664–19669, URL https://www.pnas.org/content/117/33/19664.
[36] J. Tenenbaum, V. D. Silva and J. Langford, A global geometric framework for nonlinear dimensionality reduction., Science, 290 5500 (2000), 2319–23.
[37] H. Whitney, Local properties of analytic varieties, in Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse), Princeton Univ. Press, Princeton, N.J., 1965, 205–244.