BEYOND LINEAR ALGEBRA

BERND STURMFELS

ABSTRACT

Our title challenges the reader to venture beyond linear algebra in designing models and in thinking about numerical algorithms for identifying solutions. This article accompanies the author's lecture at the International Congress of Mathematicians 2022. It covers recent advances in the study of critical point equations in optimization and statistics, and it explores the role of nonlinear algebra for linear PDEs with constant coefficients.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 14Q30; Secondary 13P25, 35A25, 62R01

KEYWORDS

Algebraic variety, polar degrees, maximum likelihood, partial differential equation

INTERNATIONAL CONGRESS OF MATHEMATICIANS 2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 6, pp. 4820–4840
DOI 10.4171/ICM2022/64

Published by EMS Press

1. INTRODUCTION

Linear algebra is ubiquitous in the mathematical universe. It plays a foundational role for many models in the sciences and engineering, and its numerical methods are a driving force behind today's technologies. The power of linear algebra stems from our ability, honed through the practice of calculus, to approximate nonlinear shapes by linear spaces.

Yet, the world is nonlinear. Nonlinear equations are a natural ingredient in mathematical models for the real world. In our view, the true nonlinear nature of a phenomenon should be respected as long as possible. We argue against the common practice of passing to a linear approximation immediately. Of course, in the final step of implementing scalable algorithms, one will always employ the powerful tools of numerical linear algebra. However, in the early phase of exploring and designing a model, there is significant benefit in going beyond linear algebra. Mathematical fields such as algebraic geometry, algebraic topology, combinatorics, commutative algebra, or representation theory furnish practical tools.

The growing awareness of theoretical mathematics in applications has led to a new field called *Nonlinear Algebra*. The textbook **[32]** offers foundations for interested students. The aim of this lecture is to introduce research trends and discuss a few recent results. At the core of many problems lies the study of subsets of $\mathbb{R}^n$ that are defined by polynomials:

$$\begin{aligned}&\{x \in \mathbb{R}^n : f_1(x) = \cdots = f_k(x) = 0, g_1(x) \geq 0, \ldots, g_l(x) \geq 0,\\ &\quad h_1(x) > 0, \ldots, h_m(x) > 0\}.\end{aligned} \tag{1.1}$$

The set (1.1) is a *basic semialgebraic set*. The Positivstellensatz **[32, THEOREM 6.14]** gives a criterion for deciding whether this set is empty. This seemingly theoretical criterion has become a practical numerical method, thanks to sums of squares **[32, §12.3]** and semidefinite programming **[7]**. In addition to this, there are symbolic algorithms for real algebraic geometry (cf. **[5]**). So, the user has a wide range of choices for working with semialgebraic sets.

In this article we disregard the inequalities in (1.1) and retain the equations only:

$$X = \{x \in \mathbb{R}^n : f_1(x) = \cdots = f_k(x) = 0\}. \tag{1.2}$$

This is a *real algebraic variety*. We wish to answer questions about X by reliable numerical computations, in particular using tools such as `Bertini` **[6]** or `HomotopyContinuation.jl` **[10]**. We focus on questions that are addressed by solving auxiliary polynomial systems with finitely many solutions, where the number of complex solutions can be determined a priori.

In Section 2 that number is the Euclidean distance degree (ED degree) of X. This governs the following question: given $u \in \mathbb{R}^n \backslash X$, which point in X is nearest to u in Euclidean distance? We derive the critical equations of this optimization problem (2.1), and we consider all solutions to these equations, both real and complex. These include all local minima and local maxima. Theorem 2.5 expresses the ED degree in terms of the polar degrees of X. Knowing these invariants allows us to find all critical points numerically, along with a proof of correctness **[9]**. We ask our nearest point question also for other norms, notably those given by a polytope. The polar degrees appear again, in Proposition 2.9.

Section 3 concerns algebraic varieties X that serve as models in statistics. Their points represent probability distributions. We focus on models for Gaussian distributions and discrete distributions. In these two scenarios, the ambient space $\mathbb{R}^n$ in (1.2) is replaced by the positive-definite cone PD_n and by the probability simplex Δ_n. Given any data set, we ask whether X is an appropriate model. To this end, maximum likelihood estimation (MLE) is used. This optimization problem is stated in (3.2) and (3.7). We employ nonlinear algebra [32] in addressing it. The number of complex critical points is the maximum likelihood degree (ML degree) of the model X. Theorem 3.7 relates this to the Euler characteristic of the underlying very affine variety. We apply this theory to a class of models arising in particle physics, namely the configuration space of m labeled points in general position in $\mathbb{P}^{k-1}$. Known ML degrees for these models are given in Theorem 3.14.

In Section 4, we turn to an analytic interpretation of the polynomial system in (1.2). The unknowns $x_1, \dots, x_n$ are replaced by differential operators $\frac{\partial}{\partial z_1}, \dots, \frac{\partial}{\partial z_n}$. The polynomials $f_1, \dots, f_k$ are viewed as linear partial differential equations (PDEs) with constant coefficients. The variety X is replaced by the space of functions $\phi(z_1, \dots, z_n)$ that are solutions to the PDE. That space is typically infinite-dimensional. Our task is to compute it. Algorithms are based on differential primary decompositions [2, 17, 18]. We also study linear PDEs for vector-valued functions. These are expressed by modules over a polynomial ring.

This article accompanies a lecture to be given in July 2022 at the International Congress of Mathematicians in St. Petersburg. It encourages mathematical scientists to employ polynomials in designing models and in thinking about numerical algorithms. Sections 2 and 3 are concerned with critical point equations in optimization and statistics. Section 4 offers a glimpse on how nonlinear algebra interfaces with the study of linear PDEs.

2. NEAREST POINTS ON ALGEBRAIC VARIETIES

We consider a model X that is given as the zero set in $\mathbb{R}^n$ of a collection $\{f_1, \dots, f_k\}$ of nonlinear polynomials in n unknowns $x_1, \dots, x_n$. Thus, X is a real algebraic variety. We assume that X is irreducible, that $I_X = \langle f_1, \dots, f_k \rangle$ is its prime ideal, and that the set of nonsingular real points is Zariski dense in X. The $k \times n$ Jacobian matrix $\mathcal{J} = (\partial f_i / \partial x_j)$ has rank at most c at any point $x \in X$, where $c = \mathrm{codim}(X)$, and x is *nonsingular* on X if the rank is exactly c. Explanations of these hypotheses are found in Chapter 2 of the textbook [32].

The following optimization problem arises in many applications. Given a data point $u \in \mathbb{R}^n \backslash X$, compute the distance to the model X. Thus, we seek a point x^* in X that is closest to u. The answer depends on the chosen metric. One might choose the Euclidean distance, a p-norm [29], or polyhedral norms, such as those arising in optimal transport [15]. In all of these cases, the solution x^* can be found by solving a system of polynomial equations.

We begin by discussing the *Euclidean distance (ED) problem*, which is as follows:

$$\text{minimize} \sum_{i=1}^{n} (x_i - u_i)^2 \text{ subject to } x \in X. \tag{2.1}$$

We now derive the critical equations for (2.1). The *augmented Jacobian matrix* $\mathcal{A}\mathcal{J}$ is the $(k+1)\times n$ matrix obtained by placing the row $(x_1-u_1,\ldots,x_n-u_n)$ atop the Jacobian matrix $\mathcal{J}$. We form the ideal generated by its $(c+1)\times(c+1)$ minors, we add the ideal of the model I_X, and we then saturate [19, (2.1)] that sum by the ideal of $c\times c$ minors of $\mathcal{J}$. The result is the *critical ideal* $\mathcal{C}_{X,u}$ of the model X with respect to the data u. The variety of $\mathcal{C}_{X,u}$ is the set of critical points of (2.1). For random data u, this variety is finite and it contains the optimal solution x^*, provided the latter is attained at a nonsingular point of X.

The algebro-geometric approach to the ED problem was pioneered in a project with Draisma, Horobeţ, Ottaviani, and Thomas [19]. That article introduced the *ED degree* of X. This is the cardinality of the complex algebraic variety in $\mathbb{C}^n$ defined by the critical ideal $\mathcal{C}_{X,u}$. The ED degree of a model X measures the difficulty of solving the ED problem for X.

Example 2.1 (Space curves). Fix $n=3$ and let X be the curve in $\mathbb{R}^3$ defined by two general polynomials f_1 and f_2 of degrees d_1 and d_2 in x_1,x_2,x_3. The augmented Jacobian matrix is

$$\mathcal{A}\mathcal{J}=\begin{pmatrix} x_1-u_1 & x_2-u_2 & x_3-u_3\\ \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \partial f_1/\partial x_3\\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \partial f_2/\partial x_3\end{pmatrix}. \tag{2.2}$$

For random data $u\in\mathbb{R}^3$, the ideal $\mathcal{C}_{X,u}=\langle f_1,f_2,\det(\mathcal{A}\mathcal{J})\rangle$ has $d_1d_2(d_1+d_2-1)$ zeros in $\mathbb{C}^3$, by Bézout [32, THEOREM 2.16]. Hence the ED degree of X equals $d_1d_2(d_1+d_2-1)$. This can also be seen using the general formula from algebraic geometry in [19, COROLLARY 5.9]. If X is a general smooth curve of degree d and genus g, then $\mathrm{EDdegree}(X)=3d+2g-2$. The above curve in 3-space has degree $d=d_1d_2$ and genus $g=d_1^2d_2/2+d_1d_2^2/2-2d_1d_2+1$.

Here is a general upper bound on the ED degree in terms of the given polynomials.

Proposition 2.2. *Let X be a variety of codimension c in $\mathbb{R}^n$ whose ideal I_X is generated by polynomials $f_1,f_2,\ldots,f_c,\ldots,f_k$ of degrees $d_1\geq d_2\geq\cdots\geq d_c\geq\cdots\geq d_k$. Then*

$$\mathrm{EDdegree}(X)\leq d_1d_2\cdots d_c\cdot\sum_{i_1+i_2+\cdots+i_c\leq n-c}(d_1-1)^{i_1}(d_2-1)^{i_2}\cdots(d_c-1)^{i_c}. \tag{2.3}$$

Equality holds when X is a generic complete intersection of codimension c (hence $c=k$).

This appears in [19, PROPOSITION 2.6]. We can derive it as follows. Bézout's Theorem ensures that the degree of the variety X is at most $d_1d_2\cdots d_c$. The entries in the ith row of the matrix $\mathcal{A}\mathcal{J}$ are polynomials of degrees d_i-1. The degree of the variety of $(c+1)\times(c+1)$ minors of $\mathcal{A}\mathcal{J}$ is at most the sum in (2.3). The intersection of that variety with X is our set of critical points, and the cardinality of that set is bounded by the product of the two degrees. Generically, that intersection is a complete intersection and inequality (2.3) is attained.

Formulas or a priori bounds for the ED degree are important when studying exact solutions to the optimization problem (2.1). The paradigm is to compute all complex critical points, by either symbolic or numerical methods, and to then extract one's favorite real solutions among these. This leads, for instance, to all local minima in (2.1). The ED degree is an upper bound on the number of real critical points. This bound is generally not tight.

Example 2.3. Consider the case $n = 2$, $c = 1$, $d_1 = 4$ in Proposition 2.2, where X is a quartic curve in the plane $\mathbb{R}^2$. The number of complex critical points is $\mathrm{EDdegree}(X) = 16$. But, they cannot be all real. For an illustration, consider the *Trott curve* $X = V(f)$, defined by

$$f = 144(x_1^4 + x_2^4) - 225(x_1^2 + x_2^2) + 350x_1^2x_2^2 + 81.$$

For general data $u = (u_1, u_2)$ in $\mathbb{R}^2$, we find 16 complex solutions to the critical equations $f = \frac{\partial f}{\partial x_2}(x_1 - u_1) - \frac{\partial f}{\partial x_1}(x_2 - u_2) = 0$. For u near the origin, eight of them are real. For $u = (\frac{7}{8}, \frac{1}{100})$, which is inside the rightmost oval, there are 10 real critical points. The two scenarios are shown in Figure 1. Local minima are green, while local maxima are purple. For $u = (2, \frac{1}{100})$, to the right of the rightmost oval, the number of real critical points is 12.

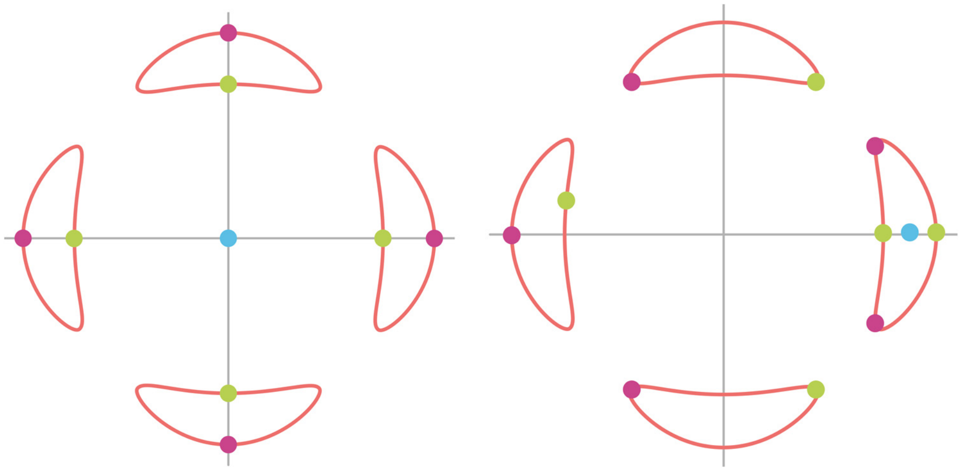

FIGURE 1
ED problems on the Trott curve: configurations of eight (left) or ten (right) critical points.

In general, our task is to compute the zeros of the critical ideal $C_{X,u}$. Algorithms for this computation can be either symbolic or numerical. Symbolic methods usually rest on the construction of a Gröbner basis, to be followed by a floating-point computation to extract the solutions. In recent years, numerical methods have become increasingly popular. These are based on homotopy continuation. Two notable packages are `Bertini` [6] and `HomotopyContinuation.jl` [10]. The ED degree is important here because it indicates how many paths need to be tracked to solve (2.1). We next illustrate current capabilities.

Example 2.4. Suppose X is defined by $c = k = 3$ random polynomials in $n = 7$ variables, for a range of degrees d_1, d_2, d_3. The table below lists the ED degree in each case, and the times used by `HomotopyContinuation.jl` to compute and certify all critical points in $\mathbb{C}^7$.

$d_1\, d_2\, d_3$	3 2 2	3 3 2	3 3 3	4 2 2	4 3 2	4 3 3	4 4 2	4 4 3
EDdegree	1188	3618	9477	4176	10152	23220	23392	49872
Solving (s)	3.849	21.06	61.51	31.51	103.5	280.0	351.5	859.3
Certifying (s)	0.390	1.549	4.653	2.762	7.591	17.16	21.65	50.07

Here we represent $C_{X,u}$ by a system of 10 equations in 10 variables. In addition to the three equations $f_1 = f_2 = f_3 = 0$ in $x_1, \dots, x_7$, we take the seven equations $(1, y_1, y_2, y_3) \cdot \mathcal{A}\mathcal{J} = 0$. Here y_1, y_2, y_3 are new variables. These ensure that the 4×7 matrix $\mathcal{A}\mathcal{J}$ has rank ≤ 3. In all cases the timings include the certification step **[9]** that proves correctness and completeness. These computations were performed using `HomotopyContinuation.jl` v2.5.6 on a 16 GB MacBook Pro with an Intel Core i7 processor working at 2.6 GHz. They suggest that our critical equations can be solved fast and reliably, with proof of correctness, when the ED degree is less than 50000. For even larger numbers of solutions, success with numerical path tracking will depend on the specific structure of the problem. If the discriminant is well-behaved, then larger ED degrees are feasible. An example of this appears in **[34, TABLE 1]**.

We next present a general formula for ED degrees in terms of projective geometry.

Theorem 2.5. *If X meets both the hyperplane at infinity and the isotropic quadric transversally, then* EDdegree(X) *equals the sum of the polar degrees of the projective closure of X.*

The *projective closure* of $X \subset \mathbb{R}^n$ is its Zariski closure in the complex projective space $\mathbb{P}^n$, which we will also denote by X. Theorem 2.5 appears in **[19, PROPOSITION 6.10]**. The hypothesis is stated precisely in **[19, EQUATION (6.4)]**. It holds for all X after a general linear change of coordinates. We now explain what the polar degrees of a variety $X \subset \mathbb{P}^n$ are. Points h in the dual projective space $(\mathbb{P}^n)^\vee$ represent hyperplanes $\{x \in \mathbb{P}^n : h_0 x_0 + \cdots + h_n x_n = 0\}$. We are interested in all pairs (x, h) in $\mathbb{P}^n \times (\mathbb{P}^n)^\vee$ such that x is a nonsingular point of X and h is tangent to X at x. The Zariski closure of this set is the *conormal variety* $N_X \subset \mathbb{P}^n \times (\mathbb{P}^n)^\vee$.

It is known that N_X has dimension $n-1$, and if X is irreducible then so is N_X. The image of N_X under projection onto the second factor is the dual variety $X^\vee$. The role of $x \in \mathbb{P}^n$ and $h \in (\mathbb{P}^n)^\vee$ can be swapped. The following biduality relations **[22, §I.1.3]** hold:

$$N_X = N_{X^\vee} \quad \text{and} \quad (X^\vee)^\vee = X.$$

The class of N_X in the cohomology ring $H^*(\mathbb{P}^n \times (\mathbb{P}^n)^\vee, \mathbb{Z}) = \mathbb{Z}[s,t]/\langle s^{n+1}, t^{n+1}\rangle$ has the form

$$[N_X] = \delta_1(X) s^n t + \delta_2(X) s^{n-1} t^2 + \delta_3(X) s^{n-2} t^3 + \cdots + \delta_n(X) s t^n.$$

The coefficients $\delta_i(X)$ of this binary form are nonnegative integers, known as *polar degrees*.

Remark 2.6. The polar degrees satisfy $\delta_i(X) = \#(N_X \cap (L \times L'))$, where $L \subset \mathbb{P}^n$ and $L' \subset (\mathbb{P}^n)^\vee$ are general linear subspaces of dimensions $n+1-i$ and i, respectively. This geometric interpretation implies that $\delta_i(X) = 0$ for $i < \mathrm{codim}(X^\vee)$ and for $i > \dim(X)+1$.

Example 2.7. Let X be a general surface of degree d in $\mathbb{P}^3$. Its dual $X^\vee$ is a surface of degree $d(d-1)^2$ in $(\mathbb{P}^3)^\vee$. The conormal variety N_X is a surface in $\mathbb{P}^3 \times (\mathbb{P}^3)^\vee$, with class

$$[N_X] = d(d-1)^2 s^3 t + d(d-1)s^2t^2 + d\, st^3.$$

The sum of the three polar degrees equals $\mathrm{EDdegree}(X) = d^3 - d^2 + d$; see Proposition 2.2.

Theorem 2.5 allows us to compute the ED degree for many interesting varieties, e.g., using Chern classes [19, THEOREM 5.8]. This is relevant for applications in machine learning [11] which rest on low-rank approximation of matrices and tensors with special structure [33].

The discussion so far was restricted to the Euclidean norm. But, we can measure distances in $\mathbb{R}^n$ with any other norm $\|\cdot\|$. Our optimization problem (2.1) extends naturally:

$$\text{minimize } \|x-u\| \text{ subject to } x \in X. \tag{2.4}$$

The unit ball $B = \{x \in \mathbb{R}^n : \|x\| \le 1\}$ is a centrally-symmetric convex body. Conversely, every centrally-symmetric convex body B defines a norm, and we can paraphrase (2.4) as follows:

$$\text{minimize } \lambda \text{ subject to } \lambda \ge 0 \text{ and } (u+\lambda B) \cap X \neq \emptyset. \tag{2.5}$$

If the boundary of B is smooth and algebraic then we express the critical equations as a polynomial system. This is derived as before, but we now replace the first row of the augmented Jacobian matrix $\mathcal{AJ}$ with the gradient of the map $\mathbb{R}^n \to \mathbb{R}$, $x \mapsto \|x-u\|$.

Another case of interest arises when $\|\cdot\|$ is a *polyhedral norm*. This means that B is a centrally-symmetric polytope. Familiar examples of polyhedral norms are $\|\cdot\|_\infty$ and $\|\cdot\|_1$, where B is the cube and the crosspolytope, respectively. In optimal transport theory, one uses a Wasserstein norm [15] whose unit ball B is the polar dual of a Lipschitz polytope.

To derive the critical equations, a combinatorial stratification of the problem is used, given by the face poset of the polytope B. Suppose that X is in general position. Then $(u + \lambda^* B) \cap X = \{x^*\}$ is a singleton for the optimal value λ^* in (2.5). The point $\frac{1}{\lambda^*}(x^* - u)$ lies in the relative interior of a unique face F of the unit ball B. Let L_F denote the linear span of F in $\mathbb{R}^n$. We have $\dim(L_F) = \dim(F)+1$. Let ℓ be any linear functional on $\mathbb{R}^n$ that attains its minimum over the polytope B at the face F. We view ℓ as a point in $(\mathbb{P}^n)^\vee$.

Lemma 2.8. *The optimal point x^* in (2.4) is the unique solution to the optimization problem*

$$\textit{minimize } \ell(x) \textit{ subject to } x \in (u+L_F) \cap X. \tag{2.6}$$

Proof. The general position hypothesis ensures that $u + L_F$ intersects X transversally, and x^* is a smooth point of that intersection. Moreover, x^* is a minimum of the restriction of ℓ to the variety $(u+L_F) \cap X$. By our hypothesis, this linear function is generic relative to the variety, so the number of critical points is finite and the function values are distinct. ∎

Problem (2.6) amounts to linear programming over a real variety. We now determine the algebraic degree of this optimization task when F is a face of codimension i.

Proposition 2.9. *Let L be a general affine-linear space of codimension $i-1$ in $\mathbb{R}^n$ and ℓ a general linear form. The number of critical points of ℓ on $L \cap X$ is the polar degree $\delta_i(X)$.*

Proof. This result is [15, THEOREM 5.1]. The number of critical points of a linear form is the degree of the dual variety $(L \cap X)^\vee$. That degree coincides with the polar degree $\delta_i(X)$. ■

Example 2.10. Consider (2.4) and (2.5) where X is a general surface of degree d in $\mathbb{R}^3$. The optimal face F of the unit ball B depends on the location of the data point u. This is shown for $d = 2$ and $\|\cdot\|_\infty$ in Figure 2. The algebraic degree of the solution x^* equals $\delta_3(X) = d$ if $\dim(F) = 0$, it is $\delta_2(X) = d(d-1)$ if $\dim(F) = 1$, and it is $\delta_1(X) = d(d-1)^2$ if $\dim(F) = 2$.

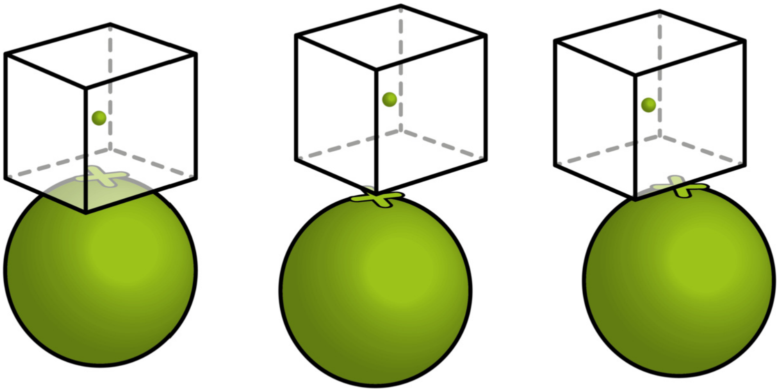

FIGURE 2
The cube is the $\|\cdot\|_\infty$ ball $\lambda^* B$ around the green point u. The variety X is the sphere. The contact point x^* is marked with a cross. The optimal face F is a facet, a vertex, or an edge.

We conclude that the conormal variety N_X and its cohomology class $[N_X]$ are key players when it comes to reliably solving the distance minimization problem for a variety X. The polar degrees $\delta_i(X)$ reveal precisely how many paths need to be tracked by numerical solvers like [6,10] in order to find and certify [9] the optimal solution x^* to (2.1) or (2.4).

3. LIKELIHOOD GEOMETRY

The previous section was concerned with minimizing the distance from a given data point u to a model X that is described by polynomial equations. In what follows, we consider the analogous problem in the setting of algebraic statistics [36], where the model X represents a family of probability distributions. Distance to u is replaced by the log-likelihood function.

The two scenarios of most interest for statisticians are Gaussian models and discrete models. We shall discuss them both, beginning with the Gaussian case. Let PD_n denote the open convex cone of positive-definite symmetric $n \times n$ matrices. Given a mean vector $\mu \in \mathbb{R}^n$ and a covariance matrix $\Sigma \in \mathrm{PD}_n$, the associated *Gaussian distribution* on $\mathbb{R}^n$ has the density

$$f_{\mu,\Sigma}(x) := \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right).$$

We fix a model $Y \subset \mathbb{R}^n \times \mathrm{PD}_n$ that is defined by polynomial equations in (μ, Σ). Suppose we are given N samples $U^{(1)}, \dots, U^{(N)}$ in $\mathbb{R}^n$. These are summarized in the *sample mean* $\bar{U} = \frac{1}{N}\sum_{i=1}^N U^{(i)}$ and in the *sample covariance matrix* $S = \frac{1}{N}\sum_{i=1}^N (U^{(i)} - \bar{U})(U^{(i)} - \bar{U})^T$. Given these data, the log-likelihood is the following function in the unknowns (μ, Σ):

$$\ell(\mu, \Sigma) = -\frac{N}{2} \cdot \left[\log\det\Sigma + \mathrm{trace}(S\Sigma^{-1}) + (\bar{U}-\mu)^T \Sigma^{-1}(\bar{U}-\mu)\right]. \tag{3.1}$$

The task of likelihood inference is to minimize this function subject to $(\mu, \Sigma) \in Y$.

There are two extreme cases. First, consider a model where Σ is fixed to be the identity matrix Id_n. Then $Y = X \times \{\mathrm{Id}_n\}$ and we are supposed to minimize the Euclidean distance from the sample mean $\bar{U}$ to the variety X in $\mathbb{R}^n$. This is precisely our problem (2.1).

We instead focus on the second case, the family of *centered Gaussians*, where μ is fixed at zero. The model has the form $\{0\} \times X$, where X is a variety in the space $\mathrm{Sym}_2(\mathbb{R}^n)$ of symmetric $n \times n$ matrices. Following [36, PROPOSITION 7.1.10], our task is now as follows:

$$\text{minimize the function } \Sigma \mapsto \log\det\Sigma + \mathrm{trace}(S\Sigma^{-1}) \quad \text{subject to } \Sigma \in X. \tag{3.2}$$

Using the concentration matrix $K = \Sigma^{-1}$, we can write this equivalently as follows:

$$\text{maximize the function } \Sigma \mapsto \log\det K - \mathrm{trace}(SK) \quad \text{subject to } K \in X^{-1}. \tag{3.3}$$

Here the variety X^{-1} is the Zariski closure of the set of inverses of all matrices in X.

The critical equations of the optimization problem (3.3) can be written as polynomials, since the partial derivatives of the logarithm are rational functions. These equations have finitely many complex solutions. Their number is the *ML degree* of the model X^{-1}.

Let $\mathcal{L} \subset \mathrm{Sym}_2(\mathbb{R}^n)$ be a linear space of symmetric matrices (LSSM), whose general element is assumed to be invertible. We are interested in the models $X^{-1} = \mathcal{L}$ and $X = \mathcal{L}$. It is convenient to use primal–dual coordinates (Σ, K) to write the respective critical equations.

Proposition 3.1. *Fix an LSSM $\mathcal{L}$ and its orthogonal complement $\mathcal{L}^\perp$ for the inner product $\langle X, Y\rangle = \mathrm{trace}(XY)$. The critical equations for the* linear concentration model $X^{-1} = \mathcal{L}$ *are*

$$K \in \mathcal{L}, \quad K\Sigma = \mathrm{Id}_n, \quad \textit{and} \quad \Sigma - S \in \mathcal{L}^\perp. \tag{3.4}$$

The critical equations for the linear covariance model $X = \mathcal{L}$ *are*

$$\Sigma \in \mathcal{L}, \quad K\Sigma = \mathrm{Id}_n, \quad \textit{and} \quad KSK - K \in \mathcal{L}^\perp. \tag{3.5}$$

Proof. This is well known in statistics. For proofs see [35, PROPOSITIONS 3.1 AND 3.3]. ■

The system (3.4) is linear in K, but the last group of equations in (3.5) is quadratic in K. The numbers of complex solutions are the *ML degree* of $\mathcal{L}$ and the *reciprocal ML degree* of $\mathcal{L}$. The former is smaller than the latter, and (3.4) is easier to solve than (3.5).

Example 3.2. Let $n = 4$ and let $\mathcal{L}$ be a generic LSSM of dimension k. Our degrees are as follows:

$k = \dim(\mathcal{L})$	2	3	4	5	6	7	8	9
ML degree	3	9	17	21	21	17	9	3
reciprocal ML degree	5	19	45	71	81	63	29	7

These numbers and many more appear in [35, TABLE 1].

ML degrees and the reciprocal ML degrees have been studied intensively in the recent literature, both for generic and special spaces $\mathcal{L}$. See [3,8,21] and the references therein. We now present an important result due to Manivel, Michałek, Monin, Seynnaeve, Vodička, and Wiśniewski. Theorem 3.3 paraphrases highlights from their articles [30, 31].

Theorem 3.3. *The ML degree of a generic linear subspace $\mathcal{L}$ of dimension k in $\mathrm{Sym}_2(\mathbb{R}^n)$ is the number of quadrics in $\mathbb{P}^{n-1}$ that pass through $\binom{n+1}{2} - k$ general points and are tangent to $k-1$ general hyperplanes. For fixed k, this number is a polynomial in n of degree $k-1$.*

Proof. The first statement is [31, COROLLARY 2.6 (4)], here interpreted classically in terms of Schubert calculus. For a detailed discussion, see the introduction of [30]. The second statement appears in [30, THEOREM 1.3 AND COROLLARY 4.13]. It proves a conjecture of Sturmfels and Uhler. ■

Example 3.4 ($n = 4$). Fix $10 - k$ points and $k - 1$ planes in $\mathbb{P}^3$. We seek quadratic surfaces containing the points and tangent to the planes. This imposes 9 constraints on $\mathbb{P}(\mathrm{Sym}_2(\mathbb{C}^4)) \simeq \mathbb{P}^9$. Passing through a point is a linear equation. Being tangent to a plane is a cubic equation. Bézout's Theorem suggests that there could be 3^{k-1} solutions. This is correct for $k \leq 3$ but it overcounts for $k \geq 4$. Indeed, in Example 3.2 we see $17, 21, 21, \ldots$ instead of $27, 81, 243, \ldots$

The intersection theory in [30, 31] leads to formulas for the ML degrees of linear Gaussian models. From this we obtain provably correct numerical methods for maximum likelihood estimation. Namely, after computing critical points as in [35], we can certify them as in [9]. Since the ML degree is known, one can be sure that all solutions have been found.

We now shift gears and turn our attention to discrete statistical models. We take the state space to be $\{0, 1, \ldots, n\}$. The role of the cone PD_n is played by the probability simplex

$$\Delta_n = \{p = (p_0, p_1, \ldots, p_n) \in \mathbb{R}^{n+1} : p_0 + p_1 + \cdots + p_n = 1 \text{ and } p_0, p_1, \ldots, p_n > 0\}. \tag{3.6}$$

Our model is a subset X of Δ_n defined by polynomial equations. As before, for venturing beyond linear algebra, we identify X with its Zariski closure in complex projective space $\mathbb{P}^n$.

We shall present the algebraic approach to maximum likelihood estimation (MLE). See [14,20,25,27,28,36] and references therein. Suppose we are given N i.i.d. samples. These are summarized in the data vector $u = (u_0, u_1, \ldots, u_n)$ where u_i is the number of times state i was observed. Note that $N = u_0 + \cdots + u_n$. The associated log-likelihood function equals

$$\ell_u : \Delta_n \to \mathbb{R}, \quad p \mapsto u_0 \cdot \log(p_0) + u_1 \cdot \log(p_1) + \cdots + u_n \cdot \log(p_n).$$

Performing MLE for the model X means solving the following optimization problem:

$$\text{maximize } \ell_u(p) \text{ subject to } p \in X. \tag{3.7}$$

The *ML degree* of X is the number of complex critical points of (3.7) for generic data u. The optimal solution is denoted $\hat{p}$ and called the *maximum likelihood estimate* for the data u.

The critical equations for (3.7) are similar to those of (2.1). Let $I_X = \langle f_1, \ldots, f_k \rangle + \langle p_0 + p_1 + \cdots + p_n - 1 \rangle$ be the defining ideal of the model. Let $\mathcal{J} = (\partial f_i / \partial p_j)$ denote the Jacobian matrix of size $(k+1) \times (n+1)$, and set $c = \operatorname{codim}(X)$. The augmented Jacobian $\mathcal{A}\mathcal{J}$ is obtained by prepending one more row, namely the gradient of the objective function

$$\nabla \ell_u = (u_0/p_0,\ u_1/p_1,\ \ldots,\ u_n/p_n).$$

To obtain the critical equations, enlarge I_X by the $c \times c$ minors of the $(k+2) \times (n+1)$ matrix $\mathcal{A}\mathcal{J}$, then clear denominators, and finally remove extraneous components by saturation.

Example 3.5 (Space curves). Let $n = 3$ and X the curve in Δ_3 defined by two general polynomials f_1 and f_2 of degrees d_1 and d_2 in p_0, p_1, p_2, p_3. The augmented Jacobian matrix is

$$\mathcal{A}\mathcal{J} = \begin{pmatrix} u_0/p_0 & u_1/p_1 & u_2/p_2 & u_3/p_3 \\ 1 & 1 & 1 & 1 \\ \partial f_1/\partial p_0 & \partial f_1/\partial p_1 & \partial f_1/\partial p_2 & \partial f_1/\partial p_3 \\ \partial f_2/\partial p_0 & \partial f_2/\partial p_1 & \partial f_2/\partial p_2 & \partial f_2/\partial p_3 \end{pmatrix}. \tag{3.8}$$

Clearing denominators amounts to multiplying the ith column by p_i, so the determinant contributes a polynomial of degree $d_1 + d_2 + 1$ to the critical equations. Since the generators of I_X have degrees $d_1, d_2, 1$, we conclude that the ML degree of X equals $d_1 d_2 (d_1 + d_2 + 1)$.

The following MLE analogue to Proposition 2.2 is established in [25, THEOREM 5].

Proposition 3.6. *Let X be a model of codimension c in Δ_n whose ideal I_X is generated by polynomials $f_1, f_2, \ldots, f_c, \ldots, f_k$ of degrees $d_1 \geq d_2 \geq \cdots \geq d_c \geq \cdots \geq d_k$. Then*

$$\operatorname{MLdegree}(X) \leq d_1 d_2 \cdots d_c \cdot \sum_{i_1 + i_2 + \cdots + i_c \leq n-c} d_1^{i_1} d_2^{i_2} \cdots d_c^{i_c}. \tag{3.9}$$

Equality holds when X is a generic complete intersection of codimension c (hence $c = k$).

We next present the MLE analogue to Theorem 2.5. The role of the polar degrees is now played by the Euler characteristic. Consider X in the complex projective space $\mathbb{P}^n$, and let X^o be the open subset of X that is obtained by removing $\{p_0 p_1 \cdots p_n (\sum_{i=0}^n p_i) = 0\}$. We recall from [26,27] that a *very affine variety* is a closed subvariety of an algebraic torus $(\mathbb{C}^*)^r$.

Theorem 3.7. *Suppose that the very affine variety X^o is nonsingular. The ML degree of the model X equals the signed Euler characteristic $(-1)^{\dim(X)} \cdot \chi(X^o)$ of the manifold X^o.*

Proof and discussion. This was proved with a further smoothness assumption in [14, THEOREM 19], and in full generality in [26, THEOREM 1]. If X^o is singular then the Euler characteristic can be replaced by the Chern–Schwartz–MacPherson class, as shown in [26, THEOREM'2]. ■

Of special interest is the case when the ML degree is equal to one. This means that the estimate $\hat{p}$ is a rational function of the data u. Here are two examples where this happens.

Example 3.8 ($n = 3$). The independence model for two binary random variables is a quadratic surface X in the tetrahedron Δ_3. This model is described by the constraints

$$\det \begin{bmatrix} p_0 & p_1 \\ p_2 & p_3 \end{bmatrix} = 0 \quad \text{and} \quad p_0 + p_1 + p_2 + p_3 = 1 \quad \text{and} \quad p_0, p_1, p_2, p_3 > 0.$$

Consider data $u = \left[\begin{smallmatrix} u_0 & u_1 \\ u_2 & u_3 \end{smallmatrix}\right]$ of *sample size* $|u| = u_0 + u_1 + u_2 + u_3$. The ML degree of the surface X equals one because the MLE $\hat{p}$ is a rational function of the data, namely

$$\begin{aligned} \hat{p}_0 &= |u|^{-2}(u_0+u_1)(u_0+u_2), & \hat{p}_1 &= |u|^{-2}(u_0+u_1)(u_1+u_3), \\ \hat{p}_2 &= |u|^{-2}(u_2+u_3)(u_0+u_2), & \hat{p}_3 &= |u|^{-2}(u_2+u_3)(u_1+u_3). \end{aligned} \tag{3.10}$$

In words, we multiply the row sums with the column sums in the empirical distribution $\frac{1}{|u|}u$.

Example 3.9 ($n = 2$). Given a biased coin, we perform the following experiment: *Flip a biased coin. If it shows heads, flip it again*. The outcome is the number of heads: 0, 1, or 2. This simple model is visualized in Figure 3.

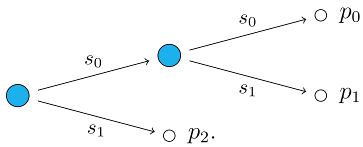

FIGURE 3
Probability tree that describes the coin toss model in Example 3.9.

If s is the bias of the cone, then the model is the parametric curve X given by

$$(0,1) \to X \subset \Delta_2, \quad s \mapsto \big(s^2, s(1-s), 1-s\big).$$

This model is the conic $X = V(p_0 p_2 - (p_0 + p_1)p_1) \subset \mathbb{P}^2$. The MLE is given by the formula

$$(\hat{p}_0, \hat{p}_1, \hat{p}_2) = \left(\frac{(2u_0 + u_1)^2}{(2u_0+2u_1+u_2)^2}, \frac{(2u_0+u_1)(u_1+u_2)}{(2u_0 + 2u_1 + u_2)^2}, \frac{u_1 + u_2}{2u_0+2u_1+u_2} \right). \tag{3.11}$$

Since the coordinates of $\hat{p}$ are rational functions, the ML degree of X is equal to one.

The following theorem explains what we saw in equations (3.10) and (3.11):

Theorem 3.10. *If $X \subset \Delta_n$ is a model of ML degree one, so $\hat{p}$ is a rational function of u, then each coordinate $\hat{p}_i$ is an alternating product of linear forms with positive coefficients.*

Proof and discussion. This was shown for very affine varieties in [27]. It was adapted to statistical models in [20]. These articles offer precise statements via Horn uniformization for A-discriminants [22], i.e., hypersurfaces dual to toric varieties. See also [28, COROLLARY 3.12]. ■

This section concludes with a connection to scattering amplitudes in particle physics that was discovered recently in [34]. We consider the *CEGM model*, due to Cachazo and his collaborators [12, 13]. The role of the data vector u is played by the Mandelstam invariants. This theory rests on the space X^o of m labeled points in general position in $\mathbb{P}^{k-1}$, up to projective transformations. Consider the action of the torus $(\mathbb{C}^*)^m$ on the Grassmannian $\mathrm{Gr}(k,m) \subset \mathbb{P}^{\binom{m}{k}-1}$. Let $\mathrm{Gr}(k,m)^o$ be the open Grassmannian where all Plücker coordinates are nonzero. The CEGM model is the $(k-1)(m-k-1)$-dimensional manifold

$$X^o = \mathrm{Gr}(k,m)^o / \big(\mathbb{C}^*\big)^m. \qquad (3.12)$$

Proposition 3.11. *The variety X^o is very affine, with coordinates given by the $k \times k$ minors of*

$$M_{k,m} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & (-1)^k & 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & (-1)^{k-1} & 0 & 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,m-k-1} \\ \vdots & \vdots & \vdots & \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -1 & \cdots & 0 & 0 & 1 & x_{k-3,1} & x_{k-3,2} & \cdots & x_{k-3,m-k-1} \\ 0 & 1 & 0 & \cdots & 0 & 0 & 1 & x_{k-2,1} & x_{k-2,2} & \cdots & x_{k-2,m-k-1} \\ -1 & 0 & 0 & \cdots & 0 & 0 & 1 & x_{k-1,1} & x_{k-1,2} & \cdots & x_{k-1,m-k-1} \end{bmatrix}. \qquad (3.13)$$

To be precise, the coordinates on $X^o \subset (\mathbb{C}^*)^{\binom{m}{k}}$ are the nonconstant minors $p_{i_1 i_2 \cdots i_k}$.

Following [1, EQUATION (4)], the antidiagonal matrix in the left $k \times k$ block of $M_{k,m}$ is chosen so that each unknown $x_{i,j}$ is precisely equal to $p_{i_1 i_2 \cdots i_k}$ for some $i_1 < i_2 < \cdots < i_k$. The *scattering potential* for the CEGM model is the following multivalued function on X^o:

$$\ell_u = \sum_{i_1, i_2, \ldots, i_k} u_{i_1 i_2 \cdots i_k} \cdot \log(p_{i_1 i_2 \cdots i_k}). \qquad (3.14)$$

The critical point equations, known as *scattering equations* [1, EQUATION (7)], are given by

$$\frac{\partial \ell_u}{\partial x_{i,j}} = 0 \quad \text{for } 1 \leq i \leq k-1 \text{ and } 1 \leq j \leq m-k-1. \qquad (3.15)$$

These are equations of rational functions. Solving these equations is the agenda in [12, 13, 34].

Corollary 3.12. *The number of complex solutions to (3.15) is the ML degree of the CEGM model X^o. This number equals the signed Euler characteristic $(-1)^{(k-1)(m-k-1)} \cdot \chi(X^o)$.*

Example 3.13 ($k = 2, m = 6$). The very affine threefold X^o is embedded in $(\mathbb{C}^*)^9$ via

$$p_{24} = x_1, \quad p_{25} = x_2, \quad p_{26} = x_3, \quad p_{34} = x_1 - 1, \quad p_{35} = x_2 - 1,$$
$$p_{36} = x_3 - 1, \quad p_{45} = x_2 - x_1, \quad p_{46} = x_3 - x_1, \quad p_{56} = x_3 - x_2.$$

These nine coordinates on $X^o \subset (\mathbb{C}^*)^9$ are the nonconstant 2×2 minors of our matrix

$$M_{2,6} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & x_1 & x_2 & x_3 \end{bmatrix}.$$

The scattering potential is the analogue to the log-likelihood function in statistics:

$$\ell_u = u_{24} \log(p_{24}) + u_{25} \log(p_{25}) + \cdots + u_{56} \log(p_{56}).$$

This function has six critical points in X^o. Hence $\text{MLdegree}(X^o) = -\chi(X^o) = 6$.

We now examine the number of critical points of the scattering potential (3.14).

Theorem 3.14. *The known values of the ML degree for the CEGM model* (3.12) *are as follows. For $k = 2$, the ML degree equals* $(m-3)!$ *for all $m \geq 4$. For $k = 3$, it equals* $2, 26, 1272, 188112, 74570400$ *for $m = 5, 6, 7, 8, 9$, respectively, and for $k = 4$, $m = 8$ it equals* 5211816.

Proof. We refer to [1, EXAMPLE 2.2], [1, THEOREM 5.1] and [1, THEOREM 6.1] for $k = 2, 3, 4$. ■

Knowing these ML degrees helps in solving the scattering equations reliably. We demonstrated in [1,34] how this can be done in practice with `HomotopyContinuation.jl` [9,10]. For instance, we see in [34, TABLE 1] that the $10! = 3628800$ solutions for $k = 2, m = 13$ are found in under one hour. See [1, SECTION 6] for the solution in the challenging case $k = 4$, $m = 8$.

4. NONLINEAR ALGEBRA MEETS LINEAR PDES

In his 1938 article on the foundations of algebraic geometry, Wolfgang Gröbner introduced differential operators to characterize membership in a polynomial ideal. He solved this for zero-dimensional ideals using Macaulay's inverse systems [24]. Gröbner wanted this for all ideals, ideally with algorithmic methods. This was finally achieved in the article [18].

Analysts made substantial contributions to this subject. In the 1960s, Leon Ehrenpreis and Victor Palamodov studied solutions to linear partial differential equations (PDEs) with constant coefficients. A main step was the characterization of membership in a primary ideal by Noetherian operators. This led to their celebrated *Fundamental Principle*. That result is presented in Theorem 4.4. For background reading, see [2,17,18] and their references.

Example 4.1 ($n = 3$). We give an illustration by exploring a progression of four questions.

Question 1: What are the solutions to the system of equations $x_1^2 = x_2^2 = x_1x_3 - x_2x_3^2 = 0$?

Question 2: Determine all functions $\phi(z_1, z_2, z_3)$ that satisfy the following three linear PDEs:

$$\frac{\partial^2 \phi}{\partial z_1^2} = \frac{\partial^2 \phi}{\partial z_2^2} = \frac{\partial^2 \phi}{\partial z_1 \partial z_3} - \frac{\partial^3 \phi}{\partial z_2 \partial z_3^2} = 0.$$

Question 3: Which polynomials lie in the ideal

$$I = \langle x_1^2, x_2^2, x_1 - x_2x_3 \rangle \cap \langle x_1^2, x_2^2, x_3 \rangle? \tag{4.1}$$

Question 4: Describe the geometry of the subscheme $V(I)$ of affine 3-space given by (4.1).

Here are our answers to these four questions. Notice how they are intertwined:

Answer 1: Assuming that $x_i^2 = 0$ implies $x_i = 0$, the equations are equivalent to $x_1 = x_2 = 0$. Their solution set is a line through the origin in 3-space, namely the x_3-axis.

Answer 2: The solutions to these PDEs are precisely the functions $\phi(z)$ that have the form

$$\phi(z_1, z_2, z_3) = \xi(z_3) + \big(z_2\psi(z_3) + z_1\psi'(z_3)\big) + \alpha z_1 z_2 + \beta z_1, \tag{4.2}$$

where α, β are constants, and ξ and ψ are differentiable functions in one variable.

Answer 3: A polynomial f is in the ideal I if and only if the following four conditions hold: Both f and $\frac{\partial f}{\partial x_2} + x_3 \frac{\partial f}{\partial x_1}$ vanish on the x_3-axis, and both $\frac{\partial^2 f}{\partial x_1 x_2}$ and $\frac{\partial f}{\partial x_1}$ vanish at the origin.

Answer 4: This scheme is a double x_3-axis together with an embedded point of length two at the origin. Hence I has arithmetic multiplicity four: two for the line and two for the point.

Answer 4 reveals the multiplicity structure on the naive solution set in Answer 1. This is characterized by four features, one for each differential condition in Answer 3. These are in natural bijection with the four summands of the general solution (4.2) in Answer 2.

We now turn to ideals I in the polynomial ring $\mathbb{C}[x] = \mathbb{C}[x_1, \ldots, x_n]$. We identify the n variables with differential operators $x_i = \partial_{z_i}$ that act on functions $\phi(z) = \phi(z_1, \ldots, z_n)$. In this manner, each I is a system of linear homogeneous PDEs with constant coefficients. This role of polynomials is the topic of Section 3.3 in the textbook **[32]**. The story begins in **[32, LEMMA 3.25]** with the following encoding of the variety $V(I)$ in the solutions to the PDE.

Lemma 4.2. *A point $a \in \mathbb{C}^n$ lies in the variety $V(I)$ if and only if the exponential function* $\exp(a \cdot z) = \exp(a_1 z_1 + \cdots + a_n z_n)$ *is a solution to the system of linear PDE given by I.*

Since our PDEs are linear, their solution sets are linear spaces. Arbitrary $\mathbb{C}$-linear combinations of solutions are again solutions. The following proposition makes this precise.

Proposition 4.3. *Given any measure μ on the variety $V(I)$, here is a solution to our PDEs:*

$$\phi(z) = \int_{V(I)} \exp(a \cdot z)\, d\mu(a). \tag{4.3}$$

If I is a prime ideal then every solution to the PDEs admits such an integral representation.

The first part of Proposition 4.3 is straightforward. Recall that an ideal Q is *primary* if it has only one associated prime P. The second part is a special case of the following result.

Theorem 4.4 (Ehrenpreis–Palamodov). *Fix a prime ideal P in $\mathbb{C}[x]$. For any P-primary ideal Q in $\mathbb{C}[x]$, there exist polynomials $B_1, \ldots, B_m$ in $2n$ unknowns such that the function*

$$\phi(z) = \sum_{i=1}^{m} \int_{V(P)} B_i(x, z) \exp(x \cdot z)\, d\mu_i(x) \tag{4.4}$$

is a solution to the PDEs given by Q, for any measures $\mu_1, \dots, \mu_m$ on the variety $V(P)$. Conversely, every solution $\phi(z)$ of the PDEs given by Q admits such an integral representation.

Proof. See [17, THEOREM 3.3] and the pointers to the analysis literature given there. ■

The polynomials $B_1(x,z), \dots, B_m(x,z)$ are known as *Noetherian multipliers*. They depend only on the primary ideal Q, and not on the function $\phi(z)$. They encode the scheme structure imposed by Q on the irreducible variety $V(P)$. The Noetherian multipliers furnish a finite representation of a vector space that is usually infinite-dimensional, namely the space of all solutions to the PDE, within a suitable class of scalar-valued functions on n-space.

Example 4.5 ($n = 3$). Let $Q = \langle x_1^2, x_2^2, x_1 - x_2x_3\rangle$ be the first primary ideal in (4.1). Here $m = 2$, $B_1 = 1$, and $B_2 = x_3z_1 + z_2$. Solutions to Q are given by the two summands in (4.4):

$$\phi_1(z) = \int 1 \cdot \exp(0z_1 + 0z_2 + x_3z_3)\, d\mu_1(x) = \xi(z_3)$$

and

$$\begin{aligned}\phi_2(z) &= \int (z_2 + z_1x_3) \cdot \exp(0z_1 + 0z_2 + x_3z_3)\, d\mu_2(x)\\ &= z_2 \int \exp(0z_1+0z_2+x_3z_3) d\mu_2(x) + z_1 \int x_3 \exp(0z_1+0z_2+x_3z_3) d\mu_2(x)\\ &= z_2\psi(z_3) + z_1\psi'(z_3).\end{aligned}$$

We conclude that our solution $\phi_1(z) + \phi_2(z)$ agrees with the first two summands in (4.2).

Switching the roles of x and z, we now set $z_1 = \partial_{x_1}, \dots, z_n = \partial_{x_n}$ in the Noetherian multipliers. Here it is important that the x-variables occur to the left of the z-variables in the monomial expansion of each $B_i(x,z)$. This results in the *Noetherian operators* $B_i(x, \partial_x)$. These operators are elements in the Weyl algebra and they act on polynomials in $\mathbb{C}[x]$. We use $\bullet$ to denote the action of differential operators on polynomials and other functions.

Proposition 4.6. *The Noetherian operators determine membership in the primary ideal Q. Namely, a polynomial $f(x)$ lies in Q if and only if $B_i(x, \partial_x) \bullet f(x)$ lies in P for $i = 1, \dots, m$.*

Proof. This is the content of [2, PROPOSITION 4.8]. See also [17, THEOREMS 3.2 AND 3.3]. ■

Example 4.7. From B_1 and B_2 in Example 4.5, we obtain the Noetherian operators 1 and $x_3\partial_{x_1} + \partial_{x_2}$. A polynomial f lies in Q if and only if f and $(x_3\partial_{x_1} + \partial_{x_2}) \bullet f$ are in $P = \langle x_1, x_2\rangle$.

We have seen that Noetherian multipliers and Noetherian operators are two sides of the same coin. While the latter characterize the membership in a primary ideal, as envisioned by Gröbner [24], the former furnish the general solution to the associated PDEs. A next step is the extension from primary to arbitrary ideals in the polynomial ring $R = \mathbb{C}[x]$. To be more general, we consider an arbitrary submodule M of the free module R^k. Such a submodule represents a system of linear PDEs as before, but for vector-valued functions $\phi : \mathbb{C}^n \to \mathbb{C}^k$.

For a vector $m \in R^k$, the quotient $(M : m)$ is the ideal $\{f \in R : fm \in M\}$. A prime ideal $P_i \subseteq R$ is *associated to* the module M if $(M : m) = P_i$ for some $m \in R^k$. The list of all associated primes of M is finite, say $P_1, \dots, P_s$. If $s = 1$ then M is P_1-primary. A *primary decomposition* of M is a list of primary submodules $M_1, \dots, M_s \subseteq R^k$ where M_i is P_i-primary and $M = M_1 \cap M_2 \cap \cdots \cap M_s$. The contribution of the primary module M_i to M is quantified by a positive integer m_i, called the arithmetic length of M along P_i. To define this, we consider the localization $(R_{P_i})^k / M_{P_i}$. This is a module over the local ring R_{P_i}. The *arithmetic length* is the length of the largest submodule of finite length in $(R_{P_i})^k / M_{P_i}$. The sum $m_1 + \cdots + m_s$ is denoted $\mathrm{amult}(M)$ and called the *arithmetic multiplicity* of M.

Example 4.8 ($n = 3, k = 1$). The ideal I in (4.1) has arithmetic multiplicity 4. The arithmetic length is $m_1 = m_2 = 2$ along each of the associated primes $P_1 = \langle x_1, x_2 \rangle$ and $P_2 = \langle x_1, x_2, x_3 \rangle$.

We now present an extension of Theorem 4.4 to PDEs for vector-valued functions. Let $V_i = V(P_i) \subset \mathbb{C}^n$ be the irreducible variety defined by the ith associated prime P_i of M.

Theorem 4.9 (Ehrenpreis–Palamodov for modules). *For any submodule $M \subset R^k$, there exist* $\mathrm{amult}(M) = \sum_{i=1}^s m_i$ *Noetherian multipliers: these are vectors $B_{ij} \in \mathbb{C}[x, z]^k$ such that*

$$\phi(z) = \sum_{i=1}^{s} \sum_{j=1}^{m_i} \int_{V_i} B_{ij}(x, z) \exp(x \cdot z) d\mu_{ij}(x) \tag{4.5}$$

is a solution to the PDE given by M. Here μ_{ij} are measures that are supported on the variety V_i. Conversely, every solution to that PDE admits such an integral representation.

Proof. This statement appears in [2, THEOREM 2.2]. Differential primary decomposition [18, THEOREM 4.6 (I)] shows that the number of inner summands equals the arithmetic length m_i. ■

As before, we can pass from Noetherian multipliers $B_{ij}(x, z)$ to Noetherian operators $B_{ij}(x, \partial_x)$ and obtain a differential primary decomposition of M; see [18] and [2, §4]. We write $\bullet$ for the application of a vector of differential operators to a vector of functions. This is done coordinatewise and followed by summing the coordinates. The result is a function.

Corollary 4.10. *The Noetherian operators determine membership in the module M. Namely, a vector $m \in R^k$ lies in M if and only if $B_{ij}(x, \partial_x) \bullet m(x)$ vanishes on V_i for all i, j.*

The package `NoetherianOperators` [16] in the software `Macaulay2` [23] is a convenient tool for solving the PDE given by a submodule `M` of R^k. Typing `amult(M)` gives the arithmetic multiplicity of `M`. The command `solvePDE(M)` lists all associated primes P_i along with their Noetherian multipliers $B_{ij}(x, z)$. These features are described in [2, §5].

What is intended with the command `solvePDE` vastly generalizes the problem of solving systems of polynomial equations, which is central to nonlinear algebra. That point is argued in [32, CHAPTER 3], which culminates with writing polynomials as PDEs. First steps towards a numerical version of `solvePDE` are discussed in [2, §7.5] and [16]. It is instructive to revisit [32, THEOREM 3.27] through the lens of Theorem 4.9. The solution space of an ideal I is finite-dimensional if and only if each V_i is a point. If, furthermore, $s = 1$ and $V_1 = \{0\}$, then the Noetherian multipliers $B_1(z), \dots, B_{m_1}(z)$ form a basis for the solution space of I.

If we pass from ideals to modules then even the case $s = 1$, $V_1 = \mathbb{C}^n$ is quite rich and interesting, especially in connection with the theory of wave cones [4]. We close with a nontrivial example which shows what wave solutions are and how they can be constructed.

Example 4.11 ($n = 4, k = 7$). Let $R = \mathbb{C}[x]$ and let $M \subset R^7$ be the module generated by $(x_1, x_2, x_3, x_4, 0, 0, 0)$, $(0, x_1, x_2, x_3, x_4, 0, 0)$, $(0, 0, x_1, x_2, x_3, x_4, 0)$, and $(0, 0, 0, x_1, x_2, x_3, x_4)$. This module is primary with $V_1 = \mathbb{C}^4$ and $\mathrm{amult}(M) = 3$. It represents a first-order PDE for unknown functions $\phi : \mathbb{R}^4 \to \mathbb{R}^7$. To explore solutions of M, we apply the `Macaulay2` command `solvePDE`. The code outputs three Noetherian multipliers, namely the rows of

$$\begin{bmatrix} x_2^4 - 3x_1x_2^2x_3 + x_1^2x_3^2 + 2x_1^2x_2x_4 & 2x_1^2x_2x_3 - x_1x_2^3 - x_1^3x_4 & x_1^2x_2^2 - x_1^3x_3 & -x_1^3x_2 & x_1^4 & 0 & 0 \\ x_2^3x_3 - 2x_1x_2x_3^2 - x_1x_2^2x_4 + 2x_1^2x_3x_4 & x_1^2x_3^2 - x_1x_2^2x_3 + x_1^2x_2x_4 & x_1^2x_2x_3 - x_1^3x_4 & -x_1^3x_3 & 0 & x_1^4 & 0 \\ x_2^3x_4 - 2x_1x_2x_3x_4 + x_1^2x_4^2 & -x_1x_2^2x_4 + x_1^2x_3x_4 & x_1^2x_2x_4 & -x_1^3x_4 & 0 & 0 & x_1^4 \end{bmatrix}.$$

These rows are syzygies of M. They span all syzygies as a vector space over the function field $\mathbb{R}(x)$. Solutions ϕ to the PDE can be constructed from any syzygy by applying that differential operator to any function $f(z_1, z_2, z_3, z_4)$. For instance, writing subscripts for differentiation, the first row of the matrix above gives the following solution to our PDE M:

$$\phi = (f_{2222} - 3f_{1223} + f_{1133} + 2f_{1124}, 2f_{1123} - f_{1222} - f_{1114}, f_{1122} - f_{1113}, \\ -f_{1112}, f_{1111}, 0, 0).$$

Next, we show how nonlinear algebra makes waves. Consider the Hankel matrix

$$H(u) = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 \\ u_2 & u_3 & u_4 & u_5 \\ u_3 & u_4 & u_5 & u_6 \\ u_4 & u_5 & u_6 & u_7 \end{bmatrix}.$$

We identify the four entries of $x \cdot H(u)$ with the generators of M. The wave cones of [4] are the determinantal varieties $\{u \in \mathbb{P}^6 : \mathrm{rank}(H(u)) \le r\}$. For $r = 1$, this is the rational normal curve in $\mathbb{P}^6$. For $r = 2$, it is the secant variety to the curve, of dimension 3. For $r = 3$, it is the variety of secant planes. The latter is the quartic hypersurface $\{u \in \mathbb{P}^6 : \det(H(u)) = 0\}$. The span of our three Noetherian multipliers furnishes a parametrization of that hypersurface.

Any $u \in \mathbb{P}^6$ with $H(u)$ of low rank yields wave solutions to M. For an illustration, let

$$u = (1, 2, 4, 8, 16, 32, 64).$$

Here $H(u)$ has rank 1. Its kernel is spanned by $2e_1 - e_2, 2e_2 - e_3, 2e_3 - e_4$. For any scalar function ψ in three variables, we obtain a function that satisfies the PDE given by M, namely

$$\phi(z) = \psi(2z_1 - z_2, 2z_2 - z_3, 2z_3 - z_4) \cdot u.$$

This vector is an example of a wave solution. If we take ψ to be the Dirac distribution at the origin in $\mathbb{R}^3$ then ϕ is a distributional solution that is supported on a line in $\mathbb{R}^4$. Characterizing such low-dimensional supports of solutions is the objective of the article [4].

ACKNOWLEDGMENTS

Many thanks to Simon Telen for the computation in Example 2.4. Helpful comments on draft versions of this paper were provided by Yulia Alexandr, Claudia Fevola, Marc Härkönen, Yelena Mandelshtam, Chiara Meroni, and Charles Wang.

REFERENCES

[1] D. Agostini, et al., Likelihood degenerations. 2021, arXiv:2107.10518.

[2] R. Ait El Manssour, M. Härkönen, and B. Sturmfels, Linear PDE with constant coefficients. *Glasg. Math. J.* (2022), published online.

[3] C. Améndola, et al., The maximum likelihood degree of linear spaces of symmetric matrices. *Matematiche* **76** (2021) 535–583.

[4] A. Arroyo-Rabasa, G. De Philippis, J. Hirsch, and F. Rindler, Dimensional estimates and rectifiability for measures satisfying linear PDE constraints. *Geom. Funct. Anal.* **29** (2019), 639–658.

[5] P. Aubry, F. Rouillier, and M. Safey El Din, Real solving for positive dimensional systems. *J. Symbolic Comput.* **34** (2002), 543–560.

[6] D. Bates, J. Hauenstein, A. Sommese, and C. Wampler, *Numerically solving polynomial systems with Bertini*. Software Environ. Tools 25, SIAM, Philadelphia, 2013.

[7] G. Blekherman, P. Parrilo, and R. Thomas, *Semidefinite optimization and convex algebraic geometry*. MOS-SIAM Ser. Optim. 13, SIAM, Philadelphia, 2013.

[8] T. Boege, et al., Reciprocal maximum likelihood degrees of Brownian motion tree models. *Matematiche* **76** (2021) 383–398.

[9] P. Breiding, K. Rose, and S. Timme, Certifying zeros of polynomial systems using interval arithmetic. 2020, arXiv:2011.05000.

[10] P. Breiding and S. Timme, HomotopyContinuation.jl: A package for homotopy continuation in Julia. In *Math. Software – ICMS 2018*, pp. 458–465, Springer, 2018.

[11] J. Bruna, K. Kohn, and M. Trager, Pure and spurious critical points: a geometric study of linear networks. In *Internat. Conf. on Learning Representations*, 2020.

[12] F. Cachazo, N. Early, A. Guevara, and S. Mizera, Scattering equations: from projective spaces to tropical Grassmannians. *J. High Energy Phys.* **2019** (2019), no. 6, 039.

[13] F. Cachazo, B. Umbert, and Y. Zhang, Singular solutions in soft limits. *J. High Energy Phys.* **2020** (2020), no. 5, 148.

[14] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels, The maximum likelihood degree. *Amer. J. Math.* **128** (2006), 671–697.

[15] T. Çelik, et al., Wasserstein distance to independence models. *J. Symbolic Comput.* **104** (2021), 855–873.

[16] J. Chen, et al., Noetherian operators in Macaulay2. 2021, arXiv:2101.01002.

[17] Y. Cid-Ruiz, R. Homs, and B. Sturmfels, Primary ideals and their differential equations. *Found. Comput. Math.* **21** (2022) 1363–1399.

[18] Y. Cid-Ruiz and B. Sturmfels, Primary decomposition with differential operators. 2021, arXiv:2101.03643.

[19] J. Draisma, et al., The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.* **16** (2016), 99–149.

[20] E. Duarte, O. Marigliano, and B. Sturmfels, Discrete statistical models with rational maximum likelihood estimator. *Bernoulli* **27** (2021), 135–154.

[21] C. Eur, T. Fife, J. Samper, and T. Seynnaeve, Reciprocal maximum likelihood degrees of diagonal linear concentration models. *Matematiche* **76** (2021) 447–459.

[22] I. Gel'fand, M. Kapranov, and A. Zelevinsky, *Discriminants, resultants and multidimensional determinants*. Birkhäuser, Boston, 1994.

[23] D. Grayson and M. Stillman, Macaulay2, a software system for research in algebraic geometry, available at http://www.math.uiuc.edu/Macaulay2/.

[24] W. Gröbner, On the Macaulay inverse system and its importance for the theory of linear differential equations with constant coefficients. *ACM Commun. Comput. Algebra* **44** (2010), 20–23. [*Abh. Math. Semin. Univ. Hambg.* **12** (1937), 127–132].

[25] S. Hoşten, A. Khetan, and B. Sturmfels, Solving the likelihood equations. *Found. Comput. Math.* **5** (2005), 389–407.

[26] J. Huh, The maximum likelihood degree of a very affine variety. *Compos. Math.* **149** (2013), 1245–1266.

[27] J. Huh, Varieties with maximum likelihood degree one. *J. Algebr. Stat.* **5** (2014), 1–17.

[28] J. Huh and B. Sturmfels, Likelihood geometry, In *Combinatorial algebraic geometry*, pp. 63–117, Lecture Notes in Math. 2108, Springer, 2014.

[29] K. Kubjas, O. Kuznetsova, and L. Sodomaco, Algebraic degree of optimization over a variety with an application to p-norm distance degree. 2021, arXiv:2105.07785.

[30] L. Manivel, et al., Complete quadrics: Schubert calculus for Gaussian models and semidefinite programming. 2020, arXiv:2011.08791.

[31] M. Michałek, L. Monin, and J. Wiśniewski, Maximum likelihood degree, complete quadrics, and $\mathbb{C}^*$-action. *SIAM J. Appl. Algebra Geom.* **5** (2021), 60–85.

[32] M. Michałek and B. Sturmfels, *Invitation to nonlinear algebra*. Grad. Stud. Math. 211, American Mathematical Society, Providence, 2021.

[33] G. Ottaviani, P-J. Spaenlehauer, and B. Sturmfels, Exact solutions in structured low-rank approximation. *SIAM J. Matrix Anal. Appl.* **35** (2014), 1521–1542.

[34] B. Sturmfels and S. Telen, Likelihood equations and scattering amplitudes. *Algebr. Stat.* **12** (2021), no. 2, 167–186. 2020, arXiv:2012.05041.

[35] B. Sturmfels, S. Timme, and P. Zwiernik, Estimating linear covariance models with numerical nonlinear algebra. *Algebr. Stat.* **1** (2020), 31–52.

[36] S. Sullivant, *Algebraic statistics*. Grad. Stud. Math. 194, American Mathematical Society, Providence, 2018.

BERND STURMFELS

Max-Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany, bernd@mis.mpg.de, and University of California, Berkeley CA 94720, USA, bernd@berkeley.edu