AStat

Algebraic Statistics

MAXIMUM LIKELIHOOD DEGREE
OF THE TWO-DIMENSIONAL LINEAR GAUSSIAN COVARIANCE MODEL

Jane Ivy Coons, Orlando Marigliano and Michael Ruddy

msp

MAXIMUM LIKELIHOOD DEGREE
OF THE TWO-DIMENSIONAL LINEAR GAUSSIAN COVARIANCE MODEL

JANE IVY COONS, ORLANDO MARIGLIANO AND MICHAEL RUDDY

In algebraic statistics, the maximum likelihood degree of a statistical model is the number of complex critical points of its log-likelihood function. A priori knowledge of this number is useful for applying techniques of numerical algebraic geometry to the maximum likelihood estimation problem. We compute the maximum likelihood degree of a generic two-dimensional subspace of the space of $n \times n$ Gaussian covariance matrices. We use the intersection theory of plane curves to show that this number is $2n - 3$.

1. Introduction

A linear Gaussian covariance model is a collection of multivariate Gaussian probability distributions whose covariance matrices are linear combinations of some fixed symmetric matrices. In this paper, we will focus on the *two-dimensional linear Gaussian covariance model*, in which all of the covariance matrices in the model lie in a two-dimensional linear space. Linear Gaussian covariance models were first studied by 1] in the context of the analysis of time series models. They continue to be studied towards this end, for example, in [23]. These models also have applications in a variety of other contexts.

One of the most common types of linear Gaussian covariance models consists of covariance matrices with some prescribed zeros. Given a Gaussian random vector $(X_1, \ldots, X_n)$ with mean μ and positive-definite covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$, we can discern independence statements from the zeros in Σ. In particular, the disjoint subvectors $(X_{i_1}, \ldots, X_{i_k})$ and $(X_{j_1}, \ldots, X_{j_l})$ are independent if and only if the submatrix of Σ that consists of rows $i_1, \ldots, i_k$ and columns $j_1, \ldots, j_l$ is the zero matrix [22, Proposition 2.4.4].

Maximum likelihood estimation for covariance matrices with a fixed independence structure was studied in [7]. These types of models find applications in the study of gene expression using relevance networks [5]. In these networks, genes are connected with an edge if their expressions are sufficiently correlated. The edges and nonedges in the resulting graph dictate the sparsity structure of the covariance matrix. Problems related to estimation of sparse covariance matrices have been studied in [3; 18].

Linear Gaussian covariance models are also applicable to the field of phylogenetics. In particular, Brownian motion tree models, which model evolution of normally distributed traits along an evolutionary tree, are linear Gaussian covariance models [11]. The covariance matrices of Brownian motion tree models require linear combinations of more than two matrices. However, the authors believe that the

results in this paper will find applications to mixtures of Brownian motion tree models. These apply, for example, to models of trait evolution that consider two genes instead of just one [16].

Algorithms for computing the maximum likelihood estimate for generic linear Gaussian covariance models have been the subject of much study [1; 2; 3; 7]. Zwiernik, Uhler, and Richards have shown that when the number of data points is sufficiently large, maximum likelihood estimation for such models behaves like a convex optimization problem in a large convex region containing the maximum likelihood estimate [24].

In this paper, we are concerned with computing the maximum likelihood degree (ML-degree) of the two-dimensional linear Gaussian covariance model for generic parameters and data. This is the number of complex critical points of the log-likelihood function, and it is considered to be a measurement of the difficulty of computing the maximum likelihood estimate (MLE) [21, Table 3]. Knowledge of the ML-degree of a model is important when applying numerical algebraic geometry methods to solve the MLE problem; in particular, it gives a stopping criterion for monodromy methods [21, §5]. For more background on ML-degrees, we refer the reader to [6; 10, Chapter 2].

2. Preliminaries

Let n be a natural number, and let $\mathrm{PD}_n \subset \mathbb{R}^{\binom{n+1}{2}}$ be the cone of all $n \times n$ symmetric positive-definite matrices. We view PD_n as the space of covariance matrices of all normal distributions $\mathcal{N}(0, \Sigma)$ with mean zero.

In algebraic statistics, a Gaussian statistical model is an algebraic subset of PD_n. In this paper, we consider models of the form

$$\mathcal{M}_{A,B} = \{xA + yB \mid x, y \in \mathbb{R}\} \cap \mathrm{PD}_n$$

for symmetric matrices A and B, whenever the intersection is not empty. That is, $\mathcal{M}_{A,B}$ is the intersection of the positive-definite cone with the linear span of A and B. We call $\mathcal{M}_{A,B}$ the *two-dimensional linear Gaussian covariance model* with respect to A and B.

Given independent, identically distributed (i.i.d.) samples $u_1, \ldots, u_r \in \mathbb{R}^n$ from some normal distribution, the maximum likelihood estimation problem for $\mathcal{M}_{A,B}$ is to find a covariance matrix $\widehat{\Sigma} \in \mathcal{M}_{A,B}$, if one exists, that maximizes the value of the likelihood function

$$L(\Sigma \mid u_1, \ldots, u_r) = \prod_{i=1}^{r} f_\Sigma(u_i),$$

where f_Σ is the density of $\mathcal{N}(0, \Sigma)$. Let S denote the *sample covariance matrix*

$$S = \frac{1}{r} \sum_{i=1}^{r} u_i u_i^T.$$

Since for all Σ the value $L(\Sigma \mid u_1, \ldots, u_r)$ only depends on S, we identify the data given by r i.i.d. samples from a normal distribution with their sample covariance matrix S. The logarithm is a concave function, so the maximizer of the likelihood function is also the maximizer of its natural log, the *log-likelihood*

function. This function can be written in terms of S:

$$\ell(\Sigma \mid S) := \log L(\Sigma \mid S)$$
$$= -\frac{rn}{2} \log(2\pi) - \frac{r}{2} \log \det(\Sigma) - \frac{r}{2} \operatorname{tr}(S\Sigma^{-1}).$$

Note that the maximizer of this function is equal to the minimizer of

$$\tilde{\ell}(\Sigma \mid S) := \log \det(\Sigma) + \operatorname{tr}(S\Sigma^{-1}).$$

When we restrict to the model $\mathcal{M}_{A,B}$, we require that $\Sigma = xA + yB$ for some $x, y \in \mathbb{R}$ such that $xA + yB$ is positive-definite. So the maximum likelihood estimation problem in this case is equivalent to

$$\underset{x,y}{\operatorname{argmin}} \, \tilde{\ell}(xA + yB \mid S) \quad \text{subject to} \quad xA + yB \in \operatorname{PD}_n.$$

To find local extrema of the log-likelihood function, we set its gradient equal to 0 and solve for x and y. The two resulting equations are called the score equations.

Definition 2.1. The *score equations* for $\mathcal{M}_{A,B}$ are the partial derivatives of the function $\tilde{\ell}(xA + yB \mid S)$ with respect to x and y. The *maximum likelihood degree* or *ML-degree* of $\mathcal{M}_{A,B}$ is the number of complex solutions to the score equations, counted with multiplicity, for a generic sample covariance matrix S.

Definition 2.1 makes reference to a *generic* sample covariance matrix. We give a detailed explanation of this term from algebraic geometry at the end of this section.

One benefit of working with $\tilde{\ell}$ is that the score equations are rational functions of the data. This allows us to use tools from algebraic geometry to analyze their solutions. Let $\Sigma = xA + yB$. For the sake of brevity, we will denote $P(x, y) = \det \Sigma$ and $T(x, y) = \operatorname{tr}(S \operatorname{adj} \Sigma)$, where $\operatorname{adj} \Sigma$ is the classical adjoint. With this notation, the function $\tilde{\ell}$ takes the form

$$\tilde{\ell}(\Sigma \mid S) = \log P + \frac{T}{P}.$$

Accordingly, the score equations are

$$\tilde{\ell}_x(x, y) = \frac{P_x}{P} + \frac{PT_x - TP_x}{P^2},$$
$$\tilde{\ell}_y(x, y) = \frac{P_y}{P} + \frac{PT_y - TP_y}{P^2}.$$

Here and throughout, the notation h_x is used for the derivative of a function h with respect to the variable x. We are concerned with values of $(x, y) \in \mathbb{C}^2$ where both of the score equations are zero. We clear denominators by multiplying $\tilde{\ell}_x$ and $\tilde{\ell}_y$ by P^2 to obtain two polynomials,

$$f(x, y) := PP_x + PT_x - TP_x,$$
$$g(x, y) := PP_y + PT_y - TP_y.$$

(1)

We note the *degrees* of each relevant term for generic A, B, and S. Specifically, their total degrees with respect to their variables x and y are

$$\deg P = n,$$
$$\deg P_x = \deg P_y = \deg T = n - 1,$$
$$\deg T_x = \deg T_y = n - 2.$$

A polynomial h is called a *homogeneous form* if each of its terms has the same degree. The polynomials f and g can be written as a sum of a homogeneous degree $2n - 1$ form with a homogeneous degree $2n - 2$ form.

The critical points of $\tilde{\ell}$ are in the variety $V(f, g)$. However, this variety also contains points at which $\tilde{\ell}$ and the score equations are not defined since we cleared denominators. The ideal whose variety is exactly the critical points of $\tilde{\ell}$ is the saturation

$$J = \mathscr{I}(f, g) : \langle P \rangle^\infty$$
$$:= \{h \in \mathbb{C}[x, y] \mid h P^N \in \mathscr{I}(f, g) \text{ for some } N\}.$$

Saturating with $P = \det \Sigma$ removes all points in $V(f, g)$ where the determinant is zero and $\tilde{\ell}$ is undefined. For more details on the geometric content of saturation, we refer the reader to Chapter 7 of [22]. We will show that $\mathscr{I}(f, g)$ and hence J are zero-dimensional in Lemmas 3.3 and 3.4. The ML-degree of the model is hence the degree of J. This is the number of isolated points in the variety of J counted with multiplicity. For more background on degrees of general varieties, see [15, Lecture 13; 20, Chapter 4, §1.4].

We now state the main result and offer an outline for its proof, which we follow in the remaining sections.

Theorem 2.2. *For generic $n \times n$ symmetric matrices A and B, the maximum likelihood degree of the two-dimensional linear Gaussian covariance model $\mathcal{M}_{A,B}$ is $2n - 3$.*

A key tool used in the proof of Theorem 2.2 is Bézout's theorem, a proof of which can be found in Chapter 5.3 of [13].

Theorem 2.3 (Bézout's theorem). *Let H and K be projective plane curves of degrees d_1 and d_2, respectively. Suppose further that H and K share no common component. Then the intersection of H and K is zero-dimensional and the number of intersection points of H and K, counted with multiplicity, is $d_1 d_2$.*

Let $F(x, y, z)$ and $G(x, y, z)$ denote the homogenizations of f and g with respect to z. Then F and G both define projective plane curves of degree $2n - 1$. Lemmas 3.3 and 3.4 will show that F and G do not share a common component. So we can apply Bézout's theorem to count their intersection points.

Let $q = [x : y : z]$ be a point in $\mathbb{CP}^2$. Then by Bézout's theorem,

$$(2n - 1)^2 = \sum_{q \in V(F,G)} I_q(F, G),$$

where $I_q(F, G)$ denotes the *intersection multiplicity* of F and G at q. The definition and properties of the intersection multiplicity of a pair of algebraic curves at a point can be found in [13, §3, Theorem 3]. For affine points $(x, y) \in V(f, g)$ we sometimes denote the intersection multiplicity as $I_{(x,y)}(f, g) := I_{[x:y:1]}(F, G)$.

We show in Proposition 3.5 that saturating the ideal $\mathscr{I}(f, g)$ with $\det \Sigma$ corresponds to removing only the origin from the affine variety of f and g. This in turn corresponds to removing the point $[0:0:1]$ from the projective variety $V(F, G)$. Since we are only interested in affine intersection points of F and G outside of the origin, we split the sum on the right-hand side of the above equation as follows:

$$(2n - 1)^2 = I_{[0:0:1]}(F, G) + \sum_{\substack{q \in V(F, G) \\ q \notin \{[0:0:1]\} \cup V(F, G, z)}} I_q(F, G) + \sum_{q \in V(F, G, z)} I_q(F, G). \qquad (2)$$

The middle term of the right-hand side of (2) is exactly the degree of the saturated ideal $J = \mathscr{I}(f, g) : \langle \det \Sigma \rangle^\infty$. Thus one can find the degree of J by computing the intersection multiplicities of F and G at the origin and at their intersection points at infinity. We compute the former in Section 4 and the latter in Section 5 to obtain

$$I_{[0:0:1]}(F, G) = (2n - 2)^2 \quad \text{and} \quad \sum_{q \in V(F, G, z)} I_q(F, G) = 2n$$

for generic A, B, and S. Thus, by rearranging (2),

$$\sum_{\substack{q \in V(F, G) \\ q \notin \{[0:0:1]\} \cup V(F, G, z)}} I_q(F, G) = (2n - 1)^2 - (2n - 2)^2 - 2n = 2n - 3,$$

which implies $\deg J = 2n - 3$.

Example 2.4. Let $n = 3$ and consider the model $\mathcal{M}_{A,B}$ defined by the positive-definite matrices

$$A = \begin{pmatrix} 5 & 1 & 0 \\ 1 & 3 & -2 \\ 0 & -2 & 6 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 1 \end{pmatrix}.$$

Using the *Julia* software package `LinearCovarianceModels.jl` [21] we find that the maximum likelihood degree of $\mathcal{M}_{A,B}$ is indeed $2 \cdot 3 - 3 = 3$, meaning that for a generic sample covariance matrix there will be three solutions over the complex number to the score equations. If we take the sample covariance matrix,

$$S = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 6 & -7 \\ -2 & -7 & 9 \end{pmatrix},$$

then the equations f, g from (1) for $\mathcal{M}_{A,B}$ and S are given explicitly by

$$f = 12288x^5 + 57600x^4 y + 74272x^3 y^2 + 20172x^2 y^3 + 1729xy^4 + 37y^5$$
$$- 10496x^4 - 33792x^3 y - 45484x^2 y^2 - 7232xy^3 - 513y^4$$

and

$$g = 11520x^5 + 37136x^4y + 20172x^3y^2 + 3458x^2y^3 + 185xy^4 + 3y^5$$
$$- 12624x^4 - 9448x^3y - 6480x^2y^2 - 528xy^3 - 21y^4.$$

We used the numerical polynomial solver package `HomotopyContinuation.jl` [4] to find the solutions to the system of equations $f = 0$, $g = 0$. The solution set consisted of the origin (with multiplicity 16) and three points corresponding to the critical points of the log-likelihood function,

$$\{(0.6897, 0.1773), (0.2655 + 0.3071i, 0.9865 - 2.4601i), (0.2655 - 0.3071i, 0.9865 + 2.4601i)\}.$$

The number of critical points and the multiplicity at the origin are predicted by Theorem 2.2 and Corollary 4.2, respectively. This fits into (2), which for $n = 3$ becomes $5^2 = 16 + 3 + 6$. Thus the maximum likelihood estimate for $\mathcal{M}_{A,B}$ and S is the real point in the list above, which corresponds to the positive-definite covariance matrix

$$\Sigma = \begin{pmatrix} 3.6257 & 0.5124 & 0 \\ 0.5124 & 3.1329 & -1.7340 \\ 0 & -1.7340 & 4.3154 \end{pmatrix}$$

that maximizes the likelihood function $L(\Sigma \mid S)$.

Properties that hold generically. In Example 2.4, it was important to choose the matrices A, B, and S to be "generic enough." We explain the precise notion of genericity in classical algebraic geometry below.

Let X be an algebraic variety and $\mathcal{P}$ a property of the points of X. One says that $\mathcal{P}(x)$ *holds for generic* $x \in X$, or *holds generically* on X, if there exists a nonempty Zariski open set U of X such that $\mathcal{P}(x)$ holds for all $x \in U$.

Consider the case $X = \mathbb{C}^N$. A Zariski open set in $\mathbb{C}^N$ is the complement of a set $V = V(f_1, \ldots, f_k)$ of common zeros of a collection of polynomials $f_1, \ldots, f_k$ in N variables.

Thus, to verify that some property $\mathcal{P}$ holds generically on $\mathbb{C}^N$, we first have to find such a set V with the property that for all x, if $\mathcal{P}(x)$ does not hold, then $x \in V$. This verifies that $\mathcal{P}(x)$ holds for all $x \in U$, where $U = \mathbb{C}^N \setminus V$. We also have to verify that U is nonempty, which amounts to finding a specific element x_0 such that $x_0 \notin V$.

Note that $\dim V$ is at most $N - 1$, which justifies the term "generic". In particular it is expected that a point $x \in \mathbb{C}^N$ taken at random[1] will lie in U.

Suppose that $\mathcal{Q}$ is another property of the points of X and we want to show that both $\mathcal{P}(x)$ and $\mathcal{Q}(x)$ hold generically on X. Then it is enough to show separately that $\mathcal{P}(x)$ holds for generic x and that $\mathcal{Q}(x)$ holds for generic x. This follows from the fact that the intersection of two nonempty Zariski open sets U_1, U_2 is always a nonempty Zariski open set. In practice, this means that after finding U_1 and U_2 it is enough to find two separate elements $x_1 \in U_1$ and $x_2 \in U_2$, which could be easier than finding an element $x_0 \in U_1 \cap U_2$.

---

[1] Say, according to the multivariate normal distribution.

In this article, the notion of a property holding generically is important for two reasons. First, it is needed for the definition of the ML-degree. Indeed, the number (with multiplicities) of solutions $(\hat{x}, \hat{y})$ to the score equations $\tilde{\ell}_i(x, y)$ given an empirical covariance matrix S could vary with S. Nevertheless, it is constant for generic S, which justifies the use of a single number. Second, we consider a family of models $\mathcal{M}_{A,B}$ parametrized by pairs of symmetric matrices (A, B) and compute its ML-degree only for generic A, B. To perform the computation, we use several properties that hold for generic A, B, and S. We prove this separately for each one of them and use them together at the same time, as explained above.

3. Geometry of the score equations

In this section, we use Bézout's theorem to derive a formula for computing $\deg J$. Lemma 3.2 will be used throughout the paper for all arguments involving generic A, B, and S. We will use Euler's homogeneous function theorem, which says that if $H(x, y)$ is a homogeneous function of degree m, then $mH = xH_x + yH_y$. We will also use the following fact about binary forms.

Proposition 3.1. *Let $H(x, y) \in \mathbb{C}[x, y]$ be a homogeneous polynomial in two variables. Then $H(x, y)$ factors as a product of linear forms.*

Proof. Suppose that $H(x, y)$ is homogeneous of degree d. Let $h(x) := H(x, 1)$ have degree c. Since $\mathbb{C}$ is algebraically closed, by the fundamental theorem of algebra, $h(x) = a \prod_{i=1}^{c}(x - r_i)$ for some $a, r_1, \ldots, r_c \in \mathbb{C}$. Then $H(x, y) = ay^{d-c} \prod_{i=1}^{c}(x - r_i y)$.   □

We further note that a generic $h \in \mathbb{C}[x, y]$ factors as a product of *distinct* linear forms. A binary form has a multiple root if and only if its discriminant vanishes, which is a closed condition on the space coefficients [17, §0.12].

Lemma 3.2. *For generic A, B, and S, the following projective varieties are empty*:

(1) $V(P, P_x), V(P, T), V(P, P_y)$,

(2) $V(P_x, P_y)$,

(3) $V(P_x, T_x), V(P_y, T_y), V(T, T_x), V(T, T_y)$.

Proof. The emptiness of the varieties in the statement is an open condition in the space of parameters (A, B, S). For instance, the subset of the parameter space $\mathbb{A}_{(A,B,S)}$ where $V(P, P_x)$ is nonempty is the image of the variety defined by P and P_x in the space $\mathbb{A}_{(A,B,S)} \times \mathbb{P}^1_{[x:y]}$ under the first projection. This is a Zariski-closed subset of the parameter space by the projective elimination theorem [8, Chapter 8.5].

To show that the projective varieties in the statement are empty, we show that the polynomials defining them have no common factors. This makes use of Proposition 3.1, which states that every homogeneous form in two variables factors as a product of linear forms.

First, consider the case where A is the $n \times n$ identity matrix, B is the diagonal matrix with diagonal entries $1, \ldots, n$, and $S = uu^T$ where u is the vector of all ones. We have

$$P = \prod_{k=1}^{n}(x + ky) \quad \text{and} \quad P_x = \sum_{k=1}^{n} \prod_{j \neq k}(x + jy).$$

From this we deduce that if $p = x + ky$ is a linear form that divides P, then it does not divide P_x. This shows that $V(P, P_x)$ is empty. The variety $V(P, T)$ is empty as well since $P_x = T$ in this case. Similarly, one shows that $V(P, P_y)$ is empty.

Euler's homogeneous function theorem applied to P says that $nP = xP_x + yP_y$. Since $V(P, P_x)$ is generically empty, the same holds for $V(P_x, P_y)$.

To prove the rest of the statements, we switch to an element (A, B, S) that makes the form of T particularly simple. Let A and B be as before and $u = (1, 0, \dots, 0)$. This is allowed when combining generic properties as we explained at the end of Section 2. In this case we have

$$T = \prod_{k \neq 1}(x + ky) \quad \text{and} \quad P_x = T + (x + y)T_x.$$

Assume p divides P_x and T_x. Then p divides T; hence, we may assume $p = x + ky$ with $k \neq 1$. However, we have $p \nmid P_x$ as before. This contradiction shows that $V(P_x, T_x)$ is empty. Similarly, $V(P_y, T_y)$ is empty. This example also has T with no common roots; hence, $V(T, T_x)$ and $V(T, T_y)$ are generically empty. $\quad\square$

Now we will show that the projective curves defined by F and G satisfy the hypothesis of Bézout's theorem, that is, that they do not share a common component. This justifies our application of Bézout's theorem and allows us to count the points in their variety. To prove this, we show that the polynomials f and g in (1) generically are irreducible and do not share a common factor.

Lemma 3.3. *The polynomials f and g in* (1) *are irreducible for generic A, B, and S.*

Proof. We prove the statement for f. The proof for g is analogous. Write $f = F_{2n-1} + F_{2n-2}$, where

$$F_{2n-1} = PP_x \quad \text{and} \quad F_{2n-2} = PT_x - TP_x$$

and $\deg F_i = i$. If f decomposes into a product of two polynomials, then at least one of them is homogeneous and we call it h. Indeed, otherwise the degrees of F_{2n-1} and F_{2n-2} would be at least two apart, when in fact they differ by one. Since h is homogeneous and divides a nonzero sum of homogeneous polynomials, h divides each of the summands F_{2n-1} and F_{2n-2}. Using Proposition 3.1, let h_0 be a linear factor of h. Since h_0 divides F_{2n-1} and is irreducible, h_0 divides P or P_x. In the first case, since h_0 divides F_{2n-2}, it would have to divide either T or P_x. This would imply that one of the projective varieties $V(P, T)$ and $V(P, P_x)$ is nonempty. By Lemma 3.2 this does not happen generically. In the second case, it would have to divide either P or T_x, which for the same reason does not happen generically. $\quad\square$

Lemma 3.4. *For generic A, B, and S, the polynomials f and g in* (1) *are not constant multiples of one another.*

Proof. If f and g are constant multiples of each other, then so are their highest-degree terms PP_x and PP_y. This does not happen generically since by Lemma 3.2 the projective variety $V(P_x, P_y)$ is generically empty. $\quad\square$

Furthermore, we can describe exactly which points are removed from the affine variety $V(f, g)$ after we saturate with the determinant. For generic parameters, the only point that is removed after saturation is the origin.

Proposition 3.5. *For generic A, B, and S, we have*

$$V(f, g) \setminus V(\det \Sigma) = V(f, g) \setminus \{(0, 0)\}.$$

Proof. Let $q \in V(P, f, g)$. Then $f(q) = -T(q)P_x(q)$ and $g(q) = -T(q)P_y(q)$. In order to have $f(q) = g(q) = 0$, we must either have both $P_x(q) = P_y(q) = 0$ or $T(q) = 0$. By Lemma 3.2, for generic A, B, and S, both of these imply $q = (0, 0)$. □

Proposition 3.6. *For generic A and B, the* ML*-degree of the model* $\mathcal{M}_{A,B}$ *is*

$$(2n - 1)^2 - I_{[0:0:1]}(F, G) - \sum_{q \in V(F, G, z)} I_q(F, G).$$

Proof. The ML-degree of $\mathcal{M}_{A,B}$ is defined as the degree of the ideal $J = \langle f, g \rangle : (\det \Sigma)^\infty$. The affine variety $V(J)$ embeds in projective space as

$$V(F, G) \setminus (V(F, G, z) \cup V(\det \Sigma)).$$

By Bézout's theorem (Theorem 2.3), Lemmas 3.3 and 3.4 imply that the variety $V(F, G)$ is zero-dimensional. Using Proposition 3.5 we have

$$\deg J = \sum_{\substack{q \in V(F, G) \\ q \notin \{[0:0:1]\} \cup V(F, G, z)}} I_q(F, G)$$

$$= \sum_{q \in V(F, G)} I_q(F, G) - I_{[0:0:1]}(F, G) - \sum_{q \in V(F, G, z)} I_q(F, G).$$

Both F and G have degree $2n - 1$. Applying Theorem 2.3 to F and G gives the desired equality.   □

4. Multiplicity at the origin

In this section we compute the intersection multiplicity of the polynomials f, g in (1) at the origin, denoted by $I_{[0:0:1]}(F, G)$ and also $I_{(0,0)}(f, g)$.

For a polynomial in two variables h there is a notion of *multiplicity* of h at the origin, denoted $m_{(0,0)}(h)$. This is the degree of the lowest-degree summand in the decomposition of h as a sum of homogeneous polynomials (for details, see [13, §3.1]). Since the polynomials f, g can be written as the sum of a homogeneous degree $2n - 2$ form with a homogeneous degree $2n - 1$ form, we have $m_{(0,0)}(f) = m_{(0,0)}(g) = 2n - 2$. We have the identity

$$I_{(0,0)}(f, g) = m_{(0,0)}(f) \cdot m_{(0,0)}(g) \tag{3}$$

if the lowest-degree homogeneous forms of f and g share no common factors [13, §3.3]. The degree $2n - 2$ parts of f and g are $Q = PT_x - TP_x$ and $R = PT_y - TP_y$, respectively.

Proposition 4.1. *For generic A, B, and S, the polynomials Q and R share no common factor.*

Proof. By the definition of Q and R and two applications of Euler's homogeneous function theorem we have

$$
\begin{aligned}
xQ + yR &= (xT_x + yT_y)P - (xP_x + yP_y)T \\
&= (2n-2)TP - (2n-1)PT \\
&= -PT.
\end{aligned}
$$

Assume that Q and R share a common factor p, which we may assume is irreducible. Then p divides PT. So p divides P or p divides T, but not both by Lemma 3.2. If p divides P, then since $Q = PT_x - TP_x$ and p is a factor of Q, p also divides TP_x. Similarly if p divides T, then p also divides PT_x. But then either P and TP_x share a common factor, or T and PT_x do. Each of the resulting four further cases does not occur generically by Lemma 3.2. $\square$

Corollary 4.2. *For generic A, B, and S, the intersection multiplicity of f and g at the origin is $(2n-2)^2$.*

Proof. By Proposition 4.1, this follows from (3). $\square$

5. Multiplicity at infinity

In this section we compute the intersection multiplicity at a point at infinity for the curves $V(f)$ and $V(g)$ defined by the polynomials in (1) for generic A, B, and S. To do this we use the connection between intersection multiplicity of curves and their series expansions about an intersection point.

Consider an irreducible polynomial h in two variables such that $h(0, 0) = 0$ and $h_y(0, 0) \neq 0$. By [12, §7.11, Corollary 2], there exists an infinite series $\alpha = \sum_{m=1}^{\infty} a_m t^m$ and an open neighborhood $U \subset \mathbb{C}$ containing $t = 0$ such that $h(t, \alpha(t)) = 0$ for all $t \in U$. The series α is called the *series expansion* of h at the origin. The *valuation* of a series is the number M such that $a_M \neq 0$ and $a_m = 0$ for all $m < M$.

Proposition 5.1. *Let h and k be irreducible polynomials in two variables such that h and k vanish at $(0, 0)$ and h_y and k_y do not. Let α and β be infinite series expansions of h and k at $(0, 0)$, respectively. The intersection multiplicity $I_{(0,0)}(h, k)$ is the valuation of the series $\alpha - \beta$.*

Proof. By [12, §8.7], the intersection multiplicity of h and k at $(0, 0)$ is the valuation of the infinite series $h(t, \beta(t))$. We prove that this is the same as the valuation of $\alpha - \beta$. First, let $s(t) = \sum_{m=1}^{\infty} s_m t^m$ be any infinite series and write $h = \sum_{i,j} c_{i,j} x^i y^j$, where the sum ranges over the pairs (i, j) with $0 < i+j \leq \deg h$. We have

$$
h(t, s(t)) = \sum_{i,j} c_{i,j} t^i \left(\sum_{m=1}^{\infty} s_m t^m \right)^j = \sum_{i,j} c_{i,j} t^i \left(\sum_{\nu=0}^{\infty} \left(\sum_{|a|=\nu} s_{a_1} \cdots s_{a_j} \right) t^\nu \right)
$$

$$
= \sum_{i,j} \sum_{\nu=0}^{\infty} \sum_{|a|=\nu} c_{i,j} s_{a_1} \cdots s_{a_j} t^{\nu+i}.
$$

The coefficient r_m of t^m in this infinite series is a finite sum of products of the form $c_{i,j} s_{a_1} \cdots s_{a_j}$ with $a_j \leq m$ and $|a| + i = m$. The term s_m only appears in r_m when $j = 1$ and $i = 0$. Hence, we have $r_m = c_{0,1} s_m + p(s_1, \ldots, s_{m-1})$ for some polynomial p, where $c_{0,1} \neq 0$ since $h_y(0, 0) \neq 0$. For example,

the coefficient r_0 is zero since h, k vanishing at the origin implies that $c_{0,0}$ and s_0 are zero, and the coefficient of the first nonzero term is given by $r_1 = c_{0,1} s_1 + c_{1,0}$.

Write $\alpha(t) = \sum_{m=1}^{\infty} a_m t^m$ and $\beta(t) = \sum_{m=1}^{\infty} b_m t^m$. Suppose that the valuation of the series $\alpha - \beta$ is M. Then $a_M \neq b_M$ and $a_m = b_m$ for all $m < M$. We show that this is equivalent to $h(t, \beta(t)) = \sum_{m=1}^{\infty} r_m t^m$ having valuation M. Suppose that $M = 1$; then $a_1 \neq b_1$. Since $h(t, \alpha(t))$ is identically zero in a neighborhood of $t = 0$, we have $r_m(a_1, \ldots, a_m) = 0$ for all m. In particular $r_1(a_1) = c_{0,1} a_1 + c_{1,0} = 0$. Since $a_1 \neq b_1$ this implies that $r_1(b_1) = c_{0,1} b_1 + c_{1,0} \neq 0$ and $h(t, \beta(t))$ has valuation one. Similarly if $h(t, \beta(t))$ has valuation one, then $r_1(a_1) \neq r_1(b_1)$ implying $a_1 \neq b_1$. Thus $\alpha - \beta$ has valuation one if and only if $h(t, \beta(t))$ has valuation one.

Now suppose $M > 1$. By the form of r_m it now follows from an inductive argument on m that a_m and b_m agree up to $m = M$ and differ at $m = M + 1$ if and only if $r_m(a_1, \ldots, a_m)$ and $r_m(b_1, \ldots, b_m)$ agree up to $m = M$ and differ at $m = M + 1$. Since $r_m(a_1, \ldots, a_m) = 0$ for all m, the latter is equivalent to $h(t, \beta(t))$ having valuation M. $\qquad\square$

Remark 5.2. In the context of Proposition 5.1, consider instead polynomials h and k defining the curves $\mathcal{X}$ and $\mathcal{Y}$, respectively, such that $\mathcal{X}$ and $\mathcal{Y}$ meet at a nonsingular point q. Also, let v be a vector such that the directional derivatives h_v and k_v do not vanish at q. Choose an affine-linear transformation $\varphi : \mathbb{C}^2 \to \mathbb{C}^2$ sending $(0, 0)$ to q and $(0, 1)$ to v. Then $I_q(h, k) = I_{(0,0)}(h \circ \varphi, k \circ \varphi)$ and the polynomials $h \circ \varphi$ and $k \circ \varphi$ satisfy the hypotheses of Proposition 5.1. Thus we can compute the intersection multiplicity at any nonsingular intersection point of h, k using Proposition 5.1.

Remark 5.3. When the series $\alpha - \beta$ has valuation M, one says that h and k have *contact order* or *order of tangency* $M - 1$ at q. Therefore the contact order of two curves at an intersection point is always one less than the intersection multiplicity. For more on contact order of algebraic curves see [19, Chapter 5].

Remark 5.4. The fact that the curves $\mathcal{X}$ and $\mathcal{Y}$ have intersection multiplicity one at q if and only if the gradients of h and k at q are linearly independent (see, e.g., [13, §3.3]) arises as a special case of Proposition 5.1 once one computes the first terms of the series α and β.

Returning to the expressions in (1), recall that F and G denote the homogenizations of f and g with respect to the new variable z. The intersection points of $V(f)$ and $V(g)$ at infinity are exactly the variety $V(F, G, z)$.

Lemma 5.5. *For generic A, B, and S, the projective variety $V(F, G, z)$ consists of n points of the form $[q_1 : q_2 : 0]$ such that $P(q_1, q_2) = 0$.*

Proof. Let $q = [q_1 : q_2 : 0]$ be a projective point of $V(F, G)$. We have

$$F = PP_x + z(PT_x - TP_x),$$
$$G = PP_y + z(PT_y - TP_y),$$

and hence $V(F, G, z)$ consists of points q where $[q_1 : q_2] \in V(PP_x, PP_y)$. Clearly if $P(q_1, q_2) = 0$, then $q \in V(F, G, z)$. These are the only such points since, by Lemma 3.2, for generic A, B, and S the variety $V(P_x, P_y)$ is empty. By Proposition 3.1 $P(x, y)$ factors in n linear forms. These forms are distinct,

since a repeated factor would divide both P and P_x, while $V(P, P_x)$ is empty by Lemma 3.2. Thus there are n distinct points in $V(F, G, z)$. □

Lemma 5.6. *For generic A, B, and S, the projective variety $V(P, P_yT_x - P_xT_y)$ is empty.*

Proof. Let $H = P_yT_x - P_xT_y$. By applying Euler's homogeneous function theorem twice in the following chain of equalities, we have

$$nT_xP - yH = T_x(nP - yP_y) + yP_xT_y = P_x(yT_y + xT_x) = (n-1)P_xT.$$

If P and H have an irreducible common factor p, then $p \mid P_xT$. This does not happen generically by Lemma 3.2. □

Lemma 5.7. *For generic A, B, and S, if $q \in V(F, G, z)$, then $I_q(F, G) = 2$.*

Proof. By Lemma 5.5, such points are of the form $q = [q_1 : q_2 : 0]$ where $P(q_1, q_2) = 0$. Fix such a point q and assume for simplicity that $q_1 \neq 0$. This is not a restriction since the conditions $q_1 = 0$ and $P(q) = 0$ imply $\det B = 0$, which is a closed condition on the parameter space. Thus we can assume q is of the form $[1 : q_2 : 0]$.

Since intersection multiplicity at a point is a local quantity, we may dehomogenize with respect to x and consider the intersection multiplicity of the affine curves $V(F(1, y, z))$ and $V(G(1, y, z))$ at q. We can compute the partial derivatives with respect to y and z:

$$F_y(x, y, z) = P_yP_x + PP_{xy} + z\left(\frac{d}{dy}(PT_x - TP_x)\right), \qquad F_z(x, y, z) = PT_x - TP_x,$$

$$G_y(x, y, z) = P_y^2 + PP_{yy} + z\left(\frac{d}{dy}(PT_y - TP_y)\right), \qquad G_z(x, y, z) = PT_y - TP_y. \tag{4}$$

Consider the translated polynomials obtained by translating q to $[1 : 0 : 0]$ given by $\widetilde{F} = F(1, y + q_2, z)$ and $\widetilde{G} = G(1, y + q_2, z)$. Then $\widetilde{F}_z(1 : 0 : 0)$, $\widetilde{G}_z(1 : 0 : 0) \neq 0$ if and only if $F_z(q)$, $G_z(q) \neq 0$, and from (4), we have that

$$F_z(q) = (-TP_x)(1, q_2) \quad \text{and} \quad G_z(q) = (-TP_y)(1, q_2).$$

Since $P(1, q_2) = 0$, Lemma 3.2 implies that $F_z(q)$, $G_z(q) \neq 0$. Thus there exist series expansions $\alpha = \sum_{m=1}^{\infty} a_m t^m$ and $\beta = \sum_{m=1}^{\infty} b_m t^m$ such that, for all t in a neighborhood of $t = 0$,

$$\widetilde{F}\left(1, t, \sum_{m=1}^{\infty} a_m t^m\right) = 0 \quad \text{and} \quad \widetilde{G}\left(1, t, \sum_{m=1}^{\infty} b_m t^m\right) = 0,$$

and hence

$$F\left(1, t + q_2, \sum_{m=1}^{\infty} a_m t^m\right) = 0 \quad \text{and} \quad G\left(1, t + q_2, \sum_{m=1}^{\infty} a_m t^m\right) = 0, \tag{5}$$

in this same neighborhood. Since $I_{[1:0:0]}(\widetilde{F}, \widetilde{G}) = I_q(F, G)$, by Proposition 5.1 we can compute the valuation of the series $\alpha - \beta$ to determine $I_q(F, G)$. Differentiating the expressions in (5) with respect to t, then substituting $t = 0$ yields

$$F_y(q) + F_z(q)a_1 = 0 \quad \text{and} \quad G_y(q) + G_z(q)b_1 = 0.$$

Thus $a_1 = -F_y(q)/F_z(q)$ and $b_1 = -G_y(q)/G_z(q)$, and $(F_y G_z - F_z G_y)(q) = 0$ implies that $a_1 - b_1 = 0$. By differentiating (5) twice with respect to t and substituting these values for a_1 and b_1, one can similarly show that

$$
\begin{aligned}
a_2 &= \left(\frac{-F_{yy}F_z^2 + 2F_{yz}F_y F_z - F_{zz}F_y^2}{2F_z^3} \right)\Bigg|_q , \\
b_2 &= \left(\frac{-G_{yy}G_z^2 + 2G_{yz}G_y G_z - G_{zz}G_y^2}{2G_z^3} \right)\Bigg|_q .
\end{aligned}
\tag{6}
$$

Since we know that $a_1 - b_1 = 0$, the valuation of $\alpha - \beta = \sum_{m=1}^{\infty}(a_m - b_m)t^m$ is *at least* two; we now show that the valuation is *exactly* two for generic A, B, and S. We verified that $a_2 - b_2 \neq 0$ with the help of the computer algebra system *Maple* by the following steps. First, we computed all second-order derivatives of F and G with respect to y and z in terms of partial derivatives of P and T, by differentiating the expressions in (4). Then, we substituted $P = 0$ and $z = 0$ in these expressions, which corresponds to evaluation at q. Thus from (6) we can obtain expressions for a_2 and b_2 evaluated at q in terms of partial derivatives of P and T. Next we cleared denominators in the resulting expression for $a_2 - b_2$, yielding

$$
(a_2 - b_2)(q) = (T^4 P_x^2 P_y^4 (P_y T_x - P_x T_y))(1, q_2).
$$

Since $P(1, q_2) = 0$, this expression does not generically evaluate to 0 by Lemmas 3.2 and 5.6. $\qquad\square$

Corollary 5.8. *For generic A, B, and S, we have $\sum_{q \in V(F,G,z)} I_q(F, G) = 2n$.*

Proof. This follows from Lemmas 5.5 and 5.7. $\qquad\square$

Now we can prove our main result that $\deg J = 2n - 3$:

Proof of Theorem 2.2. Combining Proposition 3.6 with Corollaries 4.2 and 5.8 shows that the ML-degree of $\mathcal{M}_{A,B}$ for generic A and B is

$$
(2n - 1)^2 - (2n - 2)^2 - 2n = 2n - 3. \qquad\square
$$

6. Discussion

Sturmfels, Timme, and Zwiernik [21] use numerical algebraic geometry methods implemented in the Julia package `LinearGaussianCovariance.jl` to compute the ML-degrees of linear Gaussian covariance models for several values of n and m, where n is the size of the covariance matrix and m is the dimension of model. We have proven that for $m = 2$ and arbitrary n, the ML-degree is $2n - 3$, which agrees with the computations in Table 1 of [21].

For higher-dimensional linear spaces, where $m > 2$, the score equations consist of the partial derivatives of $\tilde{\ell}$ with respect to the m parameters of the linear space. Again, in this case, these are rational functions of the data and the parameters. For instance when $m = 3$, we can consider the linear span of three $n \times n$

matrices A, B, and C. Then if $\Sigma = xA + yB + zC$, $P = \det \Sigma$, and $T = \operatorname{tr}(S \operatorname{adj} \Sigma)$, the score equations are

$$\tilde{\ell}_x(x, y, z) = \frac{P_x}{P} + \frac{PT_x - TP_x}{P^2},$$

$$\tilde{\ell}_y(x, y, z) = \frac{P_y}{P} + \frac{PT_y - TP_y}{P^2},$$

$$\tilde{\ell}_z(x, y, z) = \frac{P_z}{P} + \frac{PT_z - TP_z}{P^2},$$

and we can similarly define polynomials

$$f(x, y, z) := PP_x + PT_x - TP_x,$$

$$g(x, y, z) := PP_y + PT_y - TP_y,$$

$$h(x, y, z) := PP_z + PT_z - TP_z,$$

such that the ML-degree of the model is the degree of $J = \mathcal{I}(f, g, h) : \langle \det \Sigma \rangle^\infty$. 21] conjecture that the ML-degree in this case is $3n^2 - 9n + 7$. To prove this conjecture as we did for $m = 2$, one might turn to a higher-dimensional generalization of Bézout's theorem, which says that the number of solutions to $V(f, g, h)$ counted with multiplicity is the product $\deg(f) \deg(g) \deg(h)$ *provided that $V(f, g, h)$ is zero-dimensional* (see for example [9, §3, Chapter 3] or [20, §2.1, Chapter 3]). This zero-dimensionality restriction is necessary for equality; otherwise the product of the degrees in this case simply gives an upper bound for the number of zero-dimensional solutions counted with multiplicity [14, Theorem 12.3].

Indeed the variety $V(f, g, h)$ contains the one-dimensional affine variety $V(P, T)$ as well as a "curve at infinity" corresponding to the vanishing of P. When $m = 2$, the variety $V(P, T) \subset \mathbb{C}^2$ consisted of only the origin and the elements at infinity were points whose multiplicity we were able to ascertain using properties of curves. This illustrates the added difficulties in counting solutions when moving from planar intersection theory to higher-dimensional intersections.

21] also consider the *generic diagonal model*, in which the linear space that comprises the model consists of diagonal matrices. Their computations show that for $m = 2$, the ML-degree of the generic diagonal model for the first several values of n is also $2n - 3$ [21, Table 2]. It follows from the proof of our result that this ML-degree is indeed $2n - 3$ for all n, as the witnesses for the nonemptiness of the open dense sets that we produced in the proof of Lemma 3.2 were all diagonal matrices. For $m > 2$ and $n > 3$, the ML-degree of the generic diagonal model is conjectured in [21] to be strictly less than the corresponding generic linear Gaussian covariance model. This suggests that the study of linear Gaussian covariance models of arbitrary dimension will require us to look beyond diagonal matrices as witnesses to the nontriviality of some open conditions.

Indeed, many of the projective varieties in Lemma 3.2 are nonempty for diagonal matrices when $m > 2$. For example, when $m \geq 3$, the determinant of P for a diagonal Σ has a nonempty singular locus. Let $m = 3$ and let

$$\Sigma = xA + yB + zC$$

where A, B, and C are the diagonal matrices with diagonal entries $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$, and $(c_1, \ldots, c_n)$, respectively. Then we have

$$P = \prod_{i=1}^{n}(a_i x + b_i y + c_i z).$$

The derivatives of P are of the form

$$P_x = \sum_{i=1}^{n} a_i \prod_{\substack{j=1 \\ j \neq i}}^{n}(a_j x + b_j y + c_j z),$$

and similarly for P_y and P_z. So any projective point in the intersection of linear spaces of the form

$$V(a_i x + b_i y + c_i z) \cap V(a_j x + b_j y + c_j z)$$

for $i \neq j$ is a singular point of P. When $m > 2$, these intersections are nonempty, so such singular points exist.

Thus, when Σ is not defined by diagonal matrices, the problem of finding witnesses to the emptiness of the varieties in Lemma 3.2 for arbitrary n is nontrivial, which adds another layer of difficulty for establishing the ML-degree when $m > 3$. Nevertheless we believe that examining the structure of the score equations for $m = 2$ provides a possible blueprint for approaching the problem for $m > 2$, although it will require different tools from intersection theory.

For the purposes of statistical inference, one is most interested in *real* solutions to the score equations, as these are the ones that may have statistical meaning. Furthermore, it would be nice to understand whether there are truly $2n - 3$ distinct (complex) solutions to the score equations, as opposed to some having higher multiplicity. Based on the examples we have computed, we conjecture an affirmative answer. We thus still have the following open questions regarding the $m = 2$ case.

Problem. How many real solutions can the score equations of a generic two-dimensional linear Gaussian covariance model have?

Conjecture. For generic values of A, B, and S, the score equations of $\mathcal{M}_{A,B}$ with sample covariance matrix S have $2n - 3$ *distinct* solutions.

Acknowledgements

References

[1] T. W. Anderson, "Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices", pp. 1–24 in *Essays in probability and statistics*, edited by R. C. Bose et al., University of North Carolina, Chapel Hill, NC, 1970.

[2] T. W. Anderson, "Asymptotically efficient estimation of covariance matrices with linear structure", *Ann. Statist.* **1** (1973), 135–141.

[3] J. Bien and R. J. Tibshirani, "Sparse estimation of a covariance matrix", *Biometrika* **98**:4 (2011), 807–820.

[4] P. Breiding and S. Timme, "HomotopyContinuation.jl: a package for homotopy continuation in Julia", pp. 458–465 in *Mathematical software: ICMS 2018*, edited by J. H. Davenport et al., Lecture Notes in Computer Science **10931**, Springer, 2018.

[5] A. J. Butte, P. Tamayo, D. Slonim, T. R. Golub, and I. S. Kohane, "Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks", *Proc. Nat. Acad. Sci.* **97**:22 (2000), 12182–12186.

[6] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels, "The maximum likelihood degree", *Amer. J. Math.* **128**:3 (2006), 671–697.

[7] S. Chaudhuri, M. Drton, and T. S. Richardson, "Estimation of a covariance matrix with zeros", *Biometrika* **94**:1 (2007), 199–216.

[8] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, Springer, 1992.

[9] D. Cox, J. Little, and D. O'Shea, *Using algebraic geometry*, Graduate Texts in Mathematics **185**, Springer, 1998.

[10] M. Drton, B. Sturmfels, and S. Sullivant, *Lectures on algebraic statistics*, Oberwolfach Seminars **39**, Birkhäuser, Basel, 2009.

[11] J. Felsenstein, "Maximum-likelihood estimation of evolutionary trees from continuous characters", *Am. J. Hum. Genet.* **25**:5 (1973), 471–492.

[12] G. Fischer, *Plane algebraic curves*, Student Mathematical Library **15**, American Mathematical Society, Providence, RI, 2001.

[13] W. Fulton, *Algebraic curves: an introduction to algebraic geometry*, Addison-Wesley, Redwood City, CA, 1989.

[14] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **2**, Springer, 1998.

[15] J. Harris, *Algebraic geometry: a first course*, Graduate Texts in Mathematics **133**, Springer, 1995.

[16] H. Jiang, *Modeling trait evolutionary processes with more than one gene*, Ph.D. thesis, University of New Mexico, 2017, available at https://digitalrepository.unm.edu/math_etds/106/.

[17] V. L. Popov and E. B. Vinberg, "Invariant theory", pp. 123–278 in *Algebraic geometry*, vol. IV, edited by A. N. Parshin and I. R. Shafarevich, Encyclopaedia of Mathematical Sciences **55**, Springer, 1994.

[18] A. J. Rothman, E. Levina, and J. Zhu, "A new approach to Cholesky-based covariance regularization in high dimensions", *Biometrika* **97**:3 (2010), 539–550.

[19] J. W. Rutter, *Geometry of curves*, Chapman & Hall, Boca Raton, FL, 2000.

[20] I. R. Shafarevich, *Basic algebraic geometry*, vol. 1: Varieties in projective space, 3rd ed., Springer, 2013.

[21] B. Sturmfels, S. Timme, and P. Zwiernik, "Estimating linear covariance models with numerical nonlinear algebra", *Alg. Stat.* **11**:1 (2020), 31–52.

[22] S. Sullivant, *Algebraic statistics*, Graduate Studies in Mathematics **194**, American Mathematical Society, Providence, RI, 2018.

[23] W. B. Wu and H. Xiao, "Covariance matrix estimation in time series", Chapter 8, pp. 187–209 in *Time series analysis: methods and applications*, edited by T. Subba Rao et al., Handbook of Statistics **30**, North-Holland, Amsterdam, 2012.

[24] P. Zwiernik, C. Uhler, and D. Richards, "Maximum likelihood estimation for linear Gaussian covariance models", *J. R. Stat. Soc. B. Stat. Methodol.* **79**:4 (2017), 1269–1292.

JANE IVY COONS: jicoons@ncsu.edu
Department of Mathematics, North Carolina State University, Raleigh, NC, United States

ORLANDO MARIGLIANO: orlando.marigliano@mis.mpg.de
Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany

MICHAEL RUDDY: michael.ruddy@mis.mpg.de
Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany