Maximum Likelihood Degree of the Small Linear Gaussian Covariance Model

Jane Ivy Coons, Orlando Marigliano, Michael Ruddy

September 11, 2019

Abstract

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 *small 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 Anderson in [1] in the context of the analysis of time series models. They continue to be studied towards this end, for example, in [17]. These models also have applications in a variety of other contexts.

One of the most common types of linear Gaussian covariance models consist 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 [16, 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 [6]. In these networks, genes are connected with an edge if their expressions are sufficiently correlated. The edges and non-edges in the resulting graph dictate the sparsity structure of the covariance matrix. Problems related to estimation of sparse covariance matrices have been studied in [4] and [13].

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 [9]. 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 [12].

Algorithms for computing the maximum likelihood estimate for general linear Gaussian covariance models has been the subject of much study [1, 2, 4, 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 [18].

In this paper, we are concerned with computing maximum likelihood degree of the small 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 [15, 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 [15, Section 5]. For more background on ML-degrees, we refer the reader to [8, 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 zero mean.

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 *small linear Gaussian covariance model* with respect to A and B.

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

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

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:

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

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

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

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

$$\begin{aligned} \underset{x,y}{\mathrm{argmin}} \quad & \ell(xA + yB \mid S) \\ \text{subject to} \quad & xA + yB \in \mathrm{PD}_n. \end{aligned}$$

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 $\ell(xA + yB \mid S)$ with respect to x and y. The *maximum likelihood 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.

One benefit of working with ℓ is that the score equations are rational functions of the data. Let $\Sigma = xA + yB$. For the sake of brevity, we will denote $P(x, y) = \det\Sigma$ and $T(x, y) = \mathrm{tr}(S\,\mathrm{adj}\,\Sigma)$, where $\mathrm{adj}\,\Sigma$ is the classical adjoint. In the small linear Gaussian covariance model, the score equations are

$$\begin{aligned} \ell_x(x, y) &= \frac{P_x}{P} + \frac{PT_x - TP_x}{P^2} \\ \ell_y(x, y) &= \frac{P_y}{P} + \frac{PT_y - TP_y}{P^2}. \end{aligned}$$

Here and throughout, the notation f_x is used for the derivative of a function f 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 ℓ_x and ℓ_y by P^2 to obtain two polynomials,

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

We note that for generic A, B and S, the degrees of each relevant term are as follows:

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

Therefore, the score equations can be written as a sum of a homogeneous degree $2n-1$ form with a homogeneous degree $2n-2$ form.

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

$$J = \mathcal{I}(f,g) : \langle \det \Sigma \rangle^{\infty}.$$

Saturating with $\det \Sigma$ removes all points in $V(f,g)$ where the determinant is zero and ℓ is undefined. The ML-degree of the model is hence the degree of J.

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 small linear Gaussian covariance model $\mathcal{M}_{A,B}$ is $2n-3$.*

To prove this result we will first note that, for generic data and model, saturating the ideal $J = \mathcal{I}(f,g) : \langle \det \Sigma \rangle^{\infty}$ corresponds to removing the origin from the variety $V(f,g)$:

$$V(J) = V(f,g) \setminus V(x,y).$$

This will be proven in Proposition 3.4. Also, for generic data and matrices, the score functions are irreducible polynomials and not constant multiples of each other. This will be proven in Lemmas 3.2 and 3.3.

Next, let $F(x,y,z)$ and $G(x,y,z)$ denote the homogenizations of f and g with respect to z and $q = [x:y:z]$ be a point in $\mathbb{CP}^2$. Then by Bezout'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. For a comprehensive reference on intersection multiplicities of algebraic curves, see [11]. 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:

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

The middle term of the right-hand side of (1) is exactly the degree of J. 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 and 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 (1),

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

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

Example 2.3. Let $n = 3$. We constructed a random model $\mathcal{M}_{A,B}$ defined by the positive definite matrices,

$$A = \begin{pmatrix} 0.912568 & -0.348828 & -0.999112 \\ -0.348828 & 0.999158 & 0.363043 \\ -0.999112 & 0.363043 & 2.03161 \end{pmatrix}$$
$$B = \begin{pmatrix} 0.569108 & 0.580676 & 0.00572083 \\ 0.580676 & 0.779012 & 0.0948078 \\ 0.00572083 & 0.0948078 & 0.204917 \end{pmatrix},$$

using the *Julia* software package `LinearCovarianceModels.jl` [15]. We then generated a matrix with random entries from the interval $[-1, 1]$, and multiplied the matrix by its transpose to obtain the sample covariance matrix:

$$S = \begin{pmatrix} 0.651475 & -0.760286 & -0.20642 \\ -0.760286 & 0.996007 & 0.63946 \\ -0.20642 & 0.63946 & 1.546 \end{pmatrix}.$$

The score equations for $\mathcal{M}_{A,B}$ and S are

$$\begin{aligned} f = 1.64549x^5 &+ 8.6051x^4y + 12.0277x^3y^2 + 2.92725x^2y^3 \\ &+ 0.251863xy^4 + 0.00714521y^5 - 1.81193x^4 \\ &- 7.98009x^3y - 12.5563x^2y^2 - 2.11322xy^3 - 0.103053y^4 \end{aligned}$$

and

$$\begin{aligned} g = 1.72102x^5 &+ 6.01385x^4y + 2.92725x^3y^2 + 0.503725x^2y^3 \\ &+ 0.035726xy^4 + 8.91889 \cdot 10^{-4}y^5 - 1.69523x^4 \\ &- 1.32731x^3y - 1.26035^2y^2 \\ &- 0.185787xy^3 - 0.00815292y^4. \end{aligned}$$

We used the numerical polynomial solver package `HomotopyContinuation.jl` [5] to find the solutions to the system of equations $f = 0, g = 0$. The solutions we found were the origin with multiplicity 16 and the three points $(1.17141, -0.0671719), (0.508454 + 1.08924i, 0.803514 - 11.5242i)$ and $(0.508454 - 1.08924i, 0.803514 + 11.5242i)$, as predicted by Theorem 2.2 and Corollary 4.2. Using `LinearConvarianceModels.jl`, we also obtained the real point $(1.17141, -0.0671719)$ as the maximum likelihood estimate of the model $\mathcal{M}_{A,B}$. It corresponds to the positive definite covariance matrix

$$\Sigma = \begin{pmatrix} 1.03077 & -0.447626 & -1.17076 \\ -0.447626 & 1.1181 & 0.418904 \\ -1.17076 & 0.418904 & 2.36609 \end{pmatrix}$$

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

3 Geometry of the Score Equations

In this section, we use Bézout's Theorem to derive a formula for computing $\deg(J)$. The next Lemma will be used throughout the paper for all arguments involving generic choices of A, B and S. We will use Euler's homogeneous function theorem, which says that if $F(x, y)$ is a homogeneous function of degree m, then $mF = xF_x + yF_y$.

Lemma 3.1. *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 non-empty 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.

To show that the projective varieties in the statement are empty, we show that the polynomials defining them have no common factors.

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

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

Assume p divides P and P_x, say $p = x + ky$. We have $P_x \equiv \prod_{j\neq k}(x + jy) \not\equiv 0 \pmod{p}$. This contradiction 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 $V(P_x, P_y)$.

Next let A and B be as before and $u = (1, 0, \ldots, 0)$. 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 $P_x \not\equiv 0 \pmod{p}$ 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. □

Now we will show that the ideal of the score equations is zero-dimensional. This justifies our application of Bezout's theorem to the homogenization of the square equations and allows us to count the points in their variety. To prove this, we must show that f and g do not share a common factor. In fact, these polynomials turn out to be generically irreducible.

Lemma 3.2. *The score equations f and g 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

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

If f decomposes into a product of two polynomials, then at least one of them is homogeneous. 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. Furthermore, any homogeneous divisor of f is a common divisor of F_{2n-1} and F_{2n-2}, thus either divides P or P_x.

In the first case, 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 3.1 this doesn't happen generically. In the second case, it would have to divide either P or T_x, which for the same reason doesn't happen generically. $\square$

Lemma 3.3. *For generic A, B and S, the score equations f and g are not constant multiples of one another.*

Proof. If f and g are constant multiples of each other, then also their highest degree terms PP_x and PP_y. This does not happen generically since by Lemma 3.1 the projective variety $V(P_x, P_y)$ is generically empty. $\square$

Furthermore, we can describe exactly which points are removed from the affine variety of the score equations after we saturate with the determinant. For generic parameters, the only point that is removed after saturation is the origin.

Proposition 3.4. *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 $g(q) = T(q)P_x(q)$. 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 3.1, for generic A, B and S, both of these imply $q = (0,0)$. $\square$

Proposition 3.5. *For generic and 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 $J = \langle f, g\rangle : (\det\Sigma)^\infty$. Lemmas 3.2 and 3.3 imply that the variety $V(f,g)$ is zero-dimensional. By the properties of the intersection multiplicity and using Proposition 3.4 we have

$$\begin{aligned}
\deg(J) &= \sum_{V(f,g)\setminus V(\det\Sigma)} I_q(F,G)\\
&= \sum_{V(f,g)\setminus V(0,0)} I_q(F,G)\\
&= \sum_{q\in V(F,G)} I_q(F,G) - I_{[0:0:1]}(F,G) - \sum_{q\in V(F,G,z)} I_q(F,G).
\end{aligned}$$

Both F and G have degree $2n-1$. Applying Bezout's theorem to F and G gives the desired equality. $\square$

4 Multiplicity at the Origin

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

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, see [11, Section 3.1] for more details. Since the score equations can be written as the sum of a homogeneous degree $2n-2$ form with a homogeneous degree $2n-1$ form, we have $m_{(0,0)}(f) = m_{(0,0)}(g) = 2n-2$.

By [11, Section 3.3] we have the identity

$$I_{(0,0)}(f,g) = m_{(0,0)}(f) \cdot m_{(0,0)}(g) \tag{2}$$

if the lowest-degree homogeneous forms of f and g share no common factors. 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. *The polynomials Q and R share no common factors.*

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

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

If Q and R share a common factor p, then $PT \equiv 0 \pmod{p}$. 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 3.1. $\square$

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

Proof. By Proposition 4.1, this follows from (2). $\square$

5 Multiplicity at Infinity

In this section we compute the intersection multiplicity at a point at infinity for the curves $\mathcal{X} = V(f)$ and $\mathcal{Y} = V(g)$ 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.

Let f and g be reduced polynomials in two variables such that f and g vanish at the origin and f_y and g_y do not. By [10, Section 7.11, Corollary 2], there exist infinite series α and β in one variable such that locally around $t = 0$ we have $f(t, \alpha(t)) = 0$ and $g(t, \beta(t)) = 0$. The series α and β are called *series expansions* of f resp. g at $(0,0)$.

Proposition 5.1. *Let f and g be reduced polynomials in two variables such that f and g vanish at $(0,0)$ and f_y and g_y do not. Let α and β be infinite series expansions of f resp. g at $(0,0)$. The intersection multiplicity $I_{(0,0)}(f,g)$ is the valuation of the series $\alpha-\beta$, i.e. the number k such that the first k coefficients of $\alpha-\beta$ are zero and the $(k+1)$-th coefficient is nonzero.*

Proof. By [10, Section 8.7], the intersection multiplicity of f and g at $(0,0)$ is the valuation of the infinite series $f(t,\beta(t))$. We prove that this is the same as the valuation of $\alpha-\beta$. First, let $s(t)=\sum_{\ell=1}^{\infty} s_\ell t^\ell$ be any infinite series and write $f=\sum_{i,j} c_{i,j}x^iy^j$ with $c_{0,1}\neq 0$. We have

$$f(t,s(t))=\sum_{i,j} c_{i,j}t^i\left(\sum_{\ell=1}^{\infty} s_\ell t^\ell\right)^j=\sum_{i,j}\sum_{\nu=1}^{\infty}\sum_{|a|=\nu} c_{i,j}s_{a_1}\cdots s_{a_j}t^{\nu+i}.$$

The coefficient h_ℓ of t^ℓ in this infinite series is a finite sum of products of the form $c_{i,j}s_{a_1}\cdots s_{a_j}$ with $a_j\leq \ell$ and $|a|+i=\ell$. The term s_ℓ only appears in h_ℓ when $j=1$ and $i=0$. Hence, we have $h_\ell=c_{0,1}s_\ell+p(s_1,\ldots,s_{\ell-1})$ for some polynomial p.

Write $\alpha(t)=\sum a_\ell t^\ell$ and $\beta=\sum b_\ell t^\ell$. By the form of h_ℓ it now follows from an inductive argument that a_ℓ and b_ℓ agree up to $\ell=k$ and differ at $\ell=k+1$ if and only if $h_\ell(a_1,\ldots,a_\ell)$ and $h_\ell(b_1,\ldots,b_\ell)$ agree up to $\ell=k$ and differ at $\ell=k+1$. Since $h_\ell(a_1,\ldots,a_\ell)=0$ for all ℓ, the latter is equivalent to $f(t,\beta(t))$ having valuation k. ☐

Remark 5.2. In the context of Proposition 5.1, consider instead polynomials f and g defining the curves $\mathcal{X}$ resp. $\mathcal{Y}$ such that $\mathcal{X}$ and $\mathcal{Y}$ meet at a point q. Also, let v be a vector such that the directional derivatives f_v and g_v do not vanish at q. Choose an affine-linear transformation $\varphi\colon \mathbb{C}^2\to\mathbb{C}^2$ sending $(0,0)$ to q and $(1,0)$ to v. Then $I_q(f,g)=I_{(0,0)}(f\circ\varphi,g\circ\varphi)$ and we can hence compute the intersection multiplicity using Proposition 5.1. When the series $\alpha-\beta$ has valuation k, one says that f and g have *contact order* or *order of tangency* $k-1$ at q. For more on contact order of algebraic curves see [14, Chapter 5].

Remark 5.3. The fact that the curves $\mathcal{X}$ and $\mathcal{Y}$ have intersection multiplicity one at q if and only if the gradients of f and g at q are linearly independent arises as a special case of Proposition 5.1 once one computes the first terms of the series α and β.

Recall that F and G denote the homogenizations of f and g with respect to the new variable z. The intersection of $\mathcal{X}$ and $\mathcal{Y}$ at infinity is the variety $V(F,G,z)$.

Lemma 5.4. *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

$$\begin{aligned}F&=PP_x+z(PT_x-TP_x)\\ G&=PP_y+z(PT_y-TP_y),\end{aligned}$$

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 3.1, for generic A, B and S the variety $V(P_x, P_y)$ is empty. By the same lemma, $V(P, P_x)$ is empty, meaning $P(x,y)$ factors into n unique linear forms and thus there are n distinct points in $V(F,G,z)$. $\square$

Lemma 5.5. *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 Euler's homogeneous function theorem applied first to P then to T, we have

$$nT_xP - yH = P_x(yT_y + xT_x) = (n-1)P_xT.$$

If P and H share a root p, then p is also a root of P_xT. This does not happen generically by Lemma 3.1. $\square$

Lemma 5.6. *For generic A, B and S, if $q \in V(F,G,z)$ then $I_q(F,G) = 2$.*

Proof. By Lemma 5.4, 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. Since intersection multiplicity at a point is a local quantity, we may dehomogenize with respect to x and consider the curves $\overline{\mathcal{X}} = V(F(1,y,z))$ and $\overline{\mathcal{Y}} = V(G(1,y,z))$. We can compute the partial derivatives with respect to y and z:

$$\begin{aligned} F_y &= P_yP_x + PP_{xy} + z\left(\frac{d}{dy}(PT_x - TP_x)\right), & F_z &= PT_x - TP_x, \\ G_y &= P_y^2 + PP_{yy} + z\left(\frac{d}{dy}(PT_y - TP_y)\right), & G_z &= PT_y - TP_y. \end{aligned} \tag{3}$$

Evaluating the determinant of the Jacobian of $\overline{\mathcal{X}} \cap \overline{\mathcal{Y}}$ at $q = [1 : q_2 : 0]$ we obtain

$$(F_yG_z - F_zG_y)(q) = (-P_yP_xTP_y + TP_xP_y^2)(1, q_2) = 0,$$

meaning the gradients of $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ are linearly dependent, hence $I_q(F,G) > 1$.

Note that $F_z(q) = (-TP_x)(1, q_2)$ and $G_z(q) = (-TP_y)(1, q_2)$. Since $P(1, q_2) = 0$ we have $F_z(q), G_z(q) \neq 0$ using Lemma 3.1. Thus we can find local parametrizations of the curves $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$,

$$F\left(1, t + q_2, \sum_{n=1}^{\infty} a_nt^n\right) = 0 \quad \text{and} \quad G\left(1, t + q_2, \sum_{n=1}^{\infty} b_nt^n\right) = 0,$$

for t in a neighborhood of 0. Differentiating these expressions once with respect to t yields $a_1 = -\frac{F_y}{F_z}(q)$ and similarly for b_1. Using these values and differentiating again yields

$$a_2 = \left(\frac{-F_{yy}F_z^2 + 2F_{yz}F_yF_z - F_{zz}F_y^2}{2F_z^3} \right)\Bigg|_q$$
$$b_2 = \left(\frac{-G_{yy}G_z^2 + 2G_{yz}G_yG_z - G_{zz}G_y^2}{2G_z^3} \right)\Bigg|_q .$$

Then by Proposition 5.1 after shifting the origin to $(q_2, 0)$, we have $I_q(F, G) = 2$ if and only if $a_2 - b_2 \neq 0$. We verified the latter relation with the help of the computer algebra system *Maple* [3] by the following steps. First, we computed all second-order derivatives of F and G with respect to y and z, by taking derivatives of Equations (3). Then, we substituted $P = 0$ and $z = 0$ in these expressions, which corresponds to evaluating the expressions at q. Next, we cleared denominators in the expression $a_2 - b_2$ and evaluated it at the point q, 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 cannot evaluate to 0 by Lemmas 3.1 and 5.5. $\square$

Corollary 5.7. *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 5.4 and 5.6. $\square$

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

Proof of Theorem 2.2. Combining Proposition 3.5 with Corollaries 4.2 and 5.7 shows that the ML-degree of $\mathcal{M}_{A,B}$ for generic A and B is

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

6 Discussion

In [15], Sturmfels, Timme and Zwiernik 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 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 [15]. The authors of [15] further conjecture that for 3-dimensional models, the ML-degree is $3n^2 - 9n + 7$ and that for 4-dimensional models, the ML-degree is $11/3n^3 - 18n^2 + 85/3n - 15$. We believe that the methods used here will be useful for proving these conjectures as well as for approaching generic linear Gaussian covariance models of arbitrary dimension. The authors of [15] also compute dual ML-degrees of several linear Gaussian covariance models. The study of the dual ML-degree is a possible future application of methods from intersection theory.

The authors of [15] 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 $2n - 3$, see [15, 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 non-emptiness of the open dense sets that we produced in the proof of Lemma 3.1 were all diagonal matrices.

Acknowledgements

The authors would like to thank Carlos Amendola, Irina Kogan, Emre Sertöz, Bernd Sturmfels, Seth Sullivant, Sascha Timme, Caroline Uhler, Cynthia Vinzant and Piotr Zwiernik for many helpful conversations. All three authors were supported by the Max-Planck-Institute for Mathematics in the Sciences. Jane Coons was supported by the US National Science Foundation (DMS 1615660).

References

[1] Anderson, T. W. Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices.

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

[3] Bernardin, L., Chin, P., DeMarco, P., Geddes, K. O., Hare, D. E. G., Heal, K. M., Labahn, G., May, J. P., McCarron, J., Monagan, M. B., Ohashi, D., and Vorkoetter, S. M. *Maple Programming Guide*. Maplesoft, 2019.

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

[5] Breiding, P., and Timme, S. HomotopyContinuation.jl: A package for homotopy continuation in Julia. In *International Congress on Mathematical Software* (2018), Springer, pp. 458–465.

[6] Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R., and Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Sciences 97*, 22 (2000), 12182–12186.

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

[8] Drton, M., Sturmfels, B., and Sullivant, S. *Lectures on algebraic statistics*, vol. 39. Springer Science & Business Media, 2008.

[9] Felsenstein, J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American journal of human genetics 25*, 5 (1973), 471.

[10] Fischer, G. *Plane algebraic curves*, vol. 15 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2001. Translated from the 1994 German original by Leslie Kay.

[11] Fulton, W. *Algebraic curves: an introduction to algebraic geometry*. Addison-Wesley, 1989.

[12] Jiang, H. *Modeling Trait Evolutionary Processes with More Than One Gene*. PhD thesis, University of New Mexico, 2017.

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

[14] Rutter, J. W. *Geometry of curves*. Chapman & Hall/CRC Mathematics. Chapman & Hall/CRC, Boca Raton, FL, 2000.

[15] Sturmfels, B., Timme, S., and Zwiernik, P. Estimating linear covariance models with numerical nonlinear algebra. arXiv:1909.00566.

[16] Sullivant, S. *Algebraic Statistics*, vol. 194. American Mathematical Soc., 2018.

[17] Wu, W. B., and Xiao, H. Covariance matrix estimation in time series. In *Handbook of Statistics*, vol. 30. Elsevier, 2012, pp. 187–209.

[18] Zwiernik, P., Uhler, C., and Richards, D. Maximum likelihood estimation for linear Gaussian covariance models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*, 4 (2017), 1269–1292.

Authors' affilitiations:

Jane Ivy Coons, North Carolina State University jicoons@ncsu.edu

Orlando Marigliano, MPI MiS Leipzig orlando.marigliano@mis.mpg.de

Michael Ruddy, MPI MiS Leipzig michael.ruddy@mis.mpg.de