Found Comput Math (2014) 14:1079–1116
DOI 10.1007/s10208-014-9205-0

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Hypersurfaces and Their Singularities in Partial Correlation Testing

Shaowei Lin · Caroline Uhler · Bernd Sturmfels · Peter Bühlmann

Received: 6 November 2012 / Revised: 3 December 2013 / Accepted: 3 April 2014 /
Published online: 30 May 2014

Abstract An asymptotic theory is developed for computing volumes of regions in the parameter space of a directed Gaussian graphical model that are obtained by bounding partial correlations. We study these volumes using the method of real log canonical thresholds from algebraic geometry. Our analysis involves the computation of the singular loci of correlation hypersurfaces. Statistical applications include the strong-faithfulness assumption for the PC algorithm and the quantification of confounder bias in causal inference. A detailed analysis is presented for trees, bow ties, tripartite graphs, and complete graphs.

Keywords Causal inference · Real log canonical threshold · Resolution of singularities · Gaussian graphical model · Algebraic statistics · Singular learning theory

Mathematics Subject Classification 62H05 · 62H20 · 14Q10

Communicated by Michael Todd.

S. Lin
Institute for Infocomm Research, A*STAR, Singapore, Singapore

B. Sturmfels
Department of Mathematics, University of California Berkeley, Berkeley, CA, USA

C. Uhler (✉)
Institute of Science and Technology Austria, Klosterneuburg, Austria
e-mail: caroline.uhler@ist.ac.at

P. Bühlmann
Seminar for Statistics, ETH Zürich, Zürich, Switzerland

1 Introduction

Extensive theory has been established in recent years for causal inference based on directed acyclic graph (DAG) models. A popular method for estimating a DAG model from observational data employs partial correlation testing to infer the conditional independence relations in the model. In this paper, we apply algebraic geometry and singularity theory to analyze partial correlations in the Gaussian case. The objects of our study are algebraic hypersurfaces in the parameter space of a given graph that encode conditional independence statements.

We begin with definitions for graphical models in statistics. A DAG is a pair $G = (V, E)$ consisting of a set V of nodes and a set E of directed edges with no directed cycle. We usually take $V = \{1, 2, \ldots, p\}$ and associate random variables $X_1, X_2, \ldots, X_p$ with the nodes. Directed edges are denoted by (i, j) or $i \to j$. The *skeleton* of a DAG G is the underlying undirected graph obtained by removing the arrowheads. A node i is an *ancestor* of j if there is a directed path $i \to \cdots \to j$, and a configuration $i \to k$, $j \to k$ is a *collider* at k. Finally, we assume that the vertices are topologically ordered, that is, $(i, j) \in E$ implies $i < j$.

Every DAG G specifies a *Gaussian graphical model* as follows. The adjacency matrix A_G is a strictly upper triangular matrix whose entry in row i and column j is a parameter a_{ij} if $(i, j) \in E$ and is zero if $(i, j) \notin E$. The Gaussian graphical model is defined by the structural equation model $X = A_G^T X + \epsilon$, where $X = (X_1, \ldots, X_p)^T$. We assume that $\epsilon \sim \mathcal{N}(0, I)$, where I is the $p \times p$-identity matrix. Then the *concentration matrix* of this model equals

$$K = (A_G - I)(A_G - I)^T.$$

Since $\det(K) = 1$, the covariance matrix $\Sigma = K^{-1}$ is equal to the adjoint of K. The entries of the symmetric matrices K and Σ are polynomials in the parameters a_{ij}. Our parameter space for this DAG model will always be a full-dimensional subset Ω of $\mathbb{R}^{|E|}$.

For any subset $S \subset V$ and distinct elements i, j in $V \backslash S$, we represent the conditional independence statement $i \perp\!\!\!\perp j \mid S$ by an *almost-principal minor* of either K or Σ. By this we mean a square submatrix whose sets of row and column indices differ in exactly one element. To be precise, $i \perp\!\!\!\perp j \mid S$ holds for the multivariate normal distribution with concentration matrix K if and only if the submatrix $K_{iR,jR}$ is singular, where $R = V \backslash (S \cup \{i, j\})$ and $iR = \{i\} \cup R$. The determinant $\det(K_{iR,jR})$ is a polynomial in $(a_{ij})_{(i,j)\in E}$. We are interested in the hypersurface in $\mathbb{R}^{|E|}$ defined by the vanishing of this polynomial. Indeed, the *partial correlation* $\mathrm{corr}(i, j|S)$ is up to sign equal to the algebraic expression

$$\frac{\det(K_{iR,jR})}{\sqrt{\det(K_{iR,iR}) \cdot \det(K_{jR,jR})}}. \tag{1}$$

Since the principal minors under the square root sign are strictly positive, $\mathrm{corr}(i, j|S) = 0$ if and only if $\det(K_{iR,jR}) = 0$. If this holds for all $a \in \mathbb{R}^{|E|}$, then $i \perp\!\!\!\perp j \mid S$ for G and we say that i is *d-separated* from j *given* S. This translates into a combinatorial condition on the graph G as follows [16, §2.3.4]. An undirected path P from i to j *d-connects* i and j *given* S if

(a) Every noncollider on P is not in S;
(b) Every collider on P is in S or an ancestor of a node in S.

If G has no path that d-connects i and j given S, then i and j are *d-separated given* S, and $\det(K_{iR,jR}) \equiv 0$ as a function of a. The *weight* of a path P is the product of all edge weights a_{rs} along this path. It was shown in [17, Eq. (11)] that the numerator $\det(K_{iR,jR})$ in (1) is a linear combination, as in (5), of the weights of all paths that d-connect i to j given S.

The primary objects of this study are the following subsets of the parameter space:

$$\mathrm{Tube}_{i,j|S}(\lambda) = \{\omega \in \Omega : |\operatorname{corr}(i, j|S)| \leq \lambda\}. \tag{2}$$

Here $\operatorname{corr}(i, j|S)$ is a function of the parameter ω [denoted $(a_{ij})_{(i,j)\in E)}$ earlier] in the space $\Omega \subset \mathbb{R}^{|E|}$, λ is a parameter in [0, 1], and (i, j, S) is a triple, where i and j are d-connected given S. These "tubes" can be seen as hypersurfaces that have been fattened up by a factor that depends on λ and the position on the hypersurface (Fig. 3). The volume of $\mathrm{Tube}_{i,j|S}(\lambda)$ with respect to a given measure, or *prior*, $\varphi(\omega)\,d\omega$ on $\Omega \subset \mathbb{R}^{|E|}$ is represented by the integral

$$\mathrm{V}_{i,j|S}(\lambda) = \int\limits_{\mathrm{Tube}_{i,j|S}(\lambda)} \varphi(\omega)\,\mathrm{d}\omega. \tag{3}$$

In this paper we study the asymptotics of this integral when the parameter λ is close to 0.

Two applications in statistics are our motivation. The first concerns the strong-faithfulness assumption for algorithms that learn Markov equivalence classes of DAG models by inferring conditional independence relations. The PC algorithm [16] is a prominent instance. Our setup is exactly as in [17]. The Gaussian distribution with concentration matrix K is λ-*strong-faithful* to a DAG G if, for any $S \subset V$ and $i, j \notin S$, we have $|\operatorname{corr}(i, j|S)| \leq \lambda$ if and only if i is d-separated from j given S. We write $V_G(\lambda)$ for the volume of the region in Ω representing distributions that are not λ-strong-faithful. In other words, $V_G(\lambda)$ is the volume of the union of all tubes in Ω that correspond to non-d-separated triples (i, j, S).

Zhang and Spirtes [19] proved uniform consistency of the PC algorithm under the strong-faithfulness assumption with $\lambda \asymp 1/\sqrt{n}$, provided the number of nodes p is fixed and the sample size $n \to \infty$. In a high-dimensional, sparse setting, Kalisch and Bühlmann [11] require strong faithfulness with $\lambda \asymp \sqrt{\deg(G)\log(p)/n}$, where $\deg(G)$ denotes the maximal degree (i.e., sum of indegree and outdegree) of nodes in G.

To understand the properties of the PC algorithm for a large sample size n, it is essential to determine the asymptotic behavior of the unfaithfulness volume $V_G(\lambda)$ when λ tends to 0. Given a prior φ over the parameter space, $V_G(\lambda)$ is the prior probability that the true parameter values violate λ-strong faithfulness. Thus $1 - V_G(\lambda)$

for $\lambda \asymp 1/\sqrt{n}$ describes the prior probability that the PC algorithm is able to recover the true graph. We shall see in Example 4.8 that $V_G(\lambda)$ depends on the choice of the parameter space Ω and the prior φ.

We shall address the issue of computing $V_G(\lambda)$ as $\lambda \to 0$. This will be done using the concept of *real log canonical thresholds* (RLCTs) [1,14,18]. Our Sect. 3 establishes the existence of positive constants ℓ, m, C (which depend on G and φ) such that, asymptotically for $\lambda \to 0$,

$$\begin{aligned} \mathrm{V}_G(\lambda) &\approx C \cdot \lambda^{\ell} \cdot (-\ln \lambda)^{m-1}, \\ \mathrm{V}_{i,j|S}(\lambda) &\approx C' \cdot \lambda^{\ell'} \cdot (-\ln \lambda)^{m'-1}. \end{aligned} \tag{4}$$

[See (9) for an exact definition of $\approx$.] This refines the results in [17] on the growth of $V_G(\lambda)$ via the geometry of the *correlation hypersurfaces* $\{\det(K_{iR,jR}) = 0\}$. Whereas [17] focused on developing bounds on $V_G(\lambda)$ for the low-dimensional and the high-dimensional case and showed the importance of the number and degrees of these hypersurfaces, here we analyze the exact asymptotic behavior of $V_G(\lambda)$ for $\lambda \to 0$ and G fixed and demonstrate the importance of the singularities of these hypersurfaces. Singularities get fattened up much more than the smooth parts of the hypersurface, and this increases the volumes (4) substantially.

Our second application concerns *stratification bias in causal inference* (e.g., [7,8]). Here, the volume $V_{i,j|S}(\lambda)$ being large is not a problematic feature but is in fact desired. Suppose we want to study the effect of an exposure E on a disease outcome D. If there is an additional variable C such that $D \to C \leftarrow E$, then stratifying (i.e., conditioning) on C tends to change the association between E and D. This can lead to biases in effect estimation. This is known as *collider bias*. On the other hand, if $D \leftarrow C \to E$ holds, then C is a *confounder*, and stratifying on C corresponds to bias removal. In certain larger graphs, such as Greenland's *bow-tie* example [7], stratifying on C removes confounder bias but at the same time introduces collider bias. To decide whether one should stratify on such a variable C, it is important to understand the partial correlations involved. In this application, the volume $V_{i,j|S}(\lambda)$ can be viewed as the cumulative distribution function of the prior distribution of the partial correlation $\mathrm{corr}(i, j|S)$ implied by the prior distribution on the parameter space, and we compare the two cumulative distribution functions $V_{E,D|C}(\lambda)$ and $V_{E,D}(\lambda)$.

In this paper we examine $V_{i,j|S}(\lambda)$ from a geometric perspective, and we demonstrate how this volume can be calculated using tools from singular learning theory. To derive the asymptotics (4), the main player is the correlation hypersurface, which is the locus in Ω where $\mathrm{corr}(i, j|S)$ vanishes. The first question is whether this hypersurface is smooth and, if not, whether one needs to analyze the nature of its singularities. We study these questions for various classes of interesting causal models using methods from computational algebraic geometry.

The remainder of this paper is organized as follows. In Sect. 2 we introduce the families of DAGs that we will be working with throughout. Example 2.1 illustrates the algebraic computations that are involved in our analysis. We also discuss some simulation results, which indicate the importance of singularities when studying the volume $V_G(\lambda)$ of strong-unfaithful distributions. Sect. 3 presents the connection to singular learning theory [14,18] and explains how this theory can be used to compute the volumes of the tubes $\mathrm{Tube}_{i,j|S}(\lambda)$. Example 3.1 illustrates our theoretical results for some very simple polynomials in two variables.

In Sect. 4 we develop algebraic algorithms for analyzing the singularities of the correlation hypersurfaces. We show that, for the polynomials $\det(K_{iR,jR})$ of interest, the real singular locus is often much simpler than the complex singular locus. For instance, Theorem 4.1 states that these hypersurfaces are always smooth for complete DAGs with up to six nodes. In Sect. 5 we study the singularities and volumes (3) for trees without colliders.

Section 6 focuses on our second application, namely, bias reduction in causal inference. Problems 6.2 and 6.7 offer precise versions of conjectures by Greenland [7], in terms of comparing different volumes $V_{i,j|S}(\lambda)$ for fixed G. We establish some instances of these conjectures.

In Sect. 7 we introduce more advanced methods, based on the resolution of singularities [9,10], for finding the exponents ℓ and m in (4). Finally, in Sect. 8 we present some new results on computing the constants C and C' in our asymptotics (4) for tube volumes.

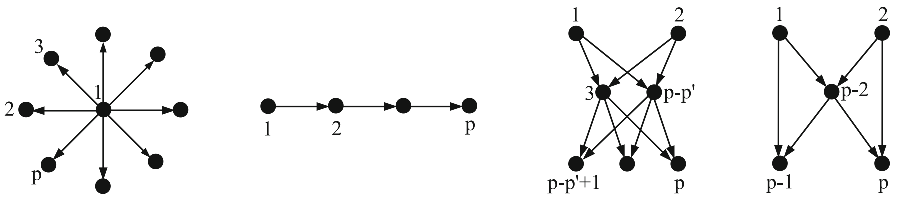
2 Four Classes of Graphs

In this article we will be primarily working with four classes of DAGs:

(i) *Complete graphs*: we denote the complete DAG on p nodes by K_p. The corresponding matrix A_{K_p} is strictly upper triangular, and all $\binom{p}{2}$ parameters a_{ij} are present.

(ii) *Trees*: we call a DAG G a tree graph if the skeleton of G is a rooted tree and all edges point away from the root (i.e., G has no colliders). We are particularly interested in the most extreme trees, namely, star and chainlike graphs. We denote the star graph shown in Fig. 1a by Star_p and the chainlike graph shown in Fig. 1b by Chain_p.

(iii) *Complete tripartite graphs*: let $A, B \subset V$, with $A \cap B = \emptyset$. Then we denote by $A \Rightarrow B$ the complete bipartite graph where $(a, b) \in E$ for all $a \in A$ and $b \in B$. A complete tripartite graph is denoted by $\mathrm{Tripart}_{p,p'}$ with $1 \le p' \le p-3$. It corresponds to the DAG $\{1, 2\} \Rightarrow \{3, \dots, p-p'\} \Rightarrow \{p-p'+1, \dots, p\}$ and is shown in Fig. 1c.



(a) Star_p **(b)** Chain_p **(c)** $\mathrm{Tripart}_{p,p'}$ **(d)** Bow_p

Fig. 1 Various classes of graphs

(iv) *Bow ties*: we define a bow tie as a complete tripartite graph $\mathrm{Tripart}_{p,2}$ with two additional edges, namely $(1, p-1)$ and $(2, p)$. A bow tie is denoted by Bow_p and is shown in Fig. 1d. Bow ties with $p = 5$ feature prominently in Greenland's study [7].

The following example serves as a preview to the topics covered in this paper.

Example 2.1 We illustrate our objects of study for the tripartite graph $G = \mathrm{Tripart}_{6,2}$. Since the error variances are assumed to be fixed at 1, this DAG model has eight free parameters, namely, the unknowns in the matrix

$$A_G = \begin{pmatrix} 0 & 0 & a_{13} & a_{14} & 0 & 0 \\ 0 & 0 & a_{23} & a_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{35} & a_{36} \\ 0 & 0 & 0 & 0 & a_{45} & a_{46} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The covariance matrix Σ equals the inverse (or adjoint) of the concentration matrix

$$K = \begin{pmatrix} a_{13}^2 + a_{14}^2 + 1 & a_{13}a_{23}+a_{14}a_{24} & -a_{13} & -a_{14} & 0 & 0 \\ a_{13}a_{23}+a_{14}a_{24} & a_{23}^2 + a_{24}^2 + 1 & -a_{23} & -a_{24} & 0 & 0 \\ -a_{13} & -a_{23} & a_{35}^2 + a_{36}^2 + 1 & a_{35}a_{45}+a_{36}a_{46} & -a_{35} & -a_{36} \\ -a_{14} & -a_{24} & a_{35}a_{45}+a_{36}a_{46} & a_{45}^2 + a_{46}^2 + 1 & -a_{45} & -a_{46} \\ 0 & 0 & -a_{35} & -a_{45} & 1 & 0 \\ 0 & 0 & -a_{36} & -a_{46} & 0 & 1 \end{pmatrix}.$$

One conditional independence statement of interest is $1 \perp\!\!\!\perp 2 \,|\, \{5, 6\}$. Its correlation hypersurface in $\mathbb{R}^8$ is defined by the almost-principal minor in Σ with rows 1, 5, and 6 and columns 2, 5, and 6, or the almost-principal minor in K with rows 1, 3, and 4 and columns 2, 3, and 4. That determinant equals

$$\begin{aligned} f = {} & (1 + a_{46}^2)a_{13}a_{23}a_{35}^2 + (1 + a_{45}^2)a_{13}a_{23}a_{36}^2 + (1 + a_{35}^2)a_{14}a_{24}a_{46}^2 \\ & + (1+a_{36}^2)a_{14}a_{24}a_{45}^2 + a_{13}a_{24}a_{35}a_{45} + a_{13}a_{24}a_{36}a_{46} \\ & + a_{14}a_{23}a_{35}a_{45} + a_{14}a_{23}a_{36}a_{46} - 2a_{13}a_{23}a_{35}a_{36}a_{45}a_{46} \\ & - 2a_{14}a_{24}a_{35}a_{36}a_{45}a_{46}. \end{aligned} \tag{5}$$

This is a weighted sum of all paths that d-connect nodes 1 and 2 given $\{5, 6\}$. The first term in formula (5) for $f = \det(K_{134,234})$ corresponds to the path $1 \to 3 \to 5 \leftarrow 3 \leftarrow 2$ in $G = \mathrm{Tripart}_{6,2}$, and the last term corresponds to the path $1 \to 4 \to 5 \leftarrow 3 \to 6 \leftarrow 4 \leftarrow 2$.

Let φ denote a prior on the parameter space. For this example we take φ to be the Lebesgue probability measure on the cube $\Omega = [-1, +1]^8$. The expression $V_{1,2|56}(\lambda)$

defined in (3) is the volume of the region of parameters $a \in \Omega$ that satisfy

$$|\operatorname{corr}(1,2 \mid 5,6)| \;=\; \left| \frac{f(a)}{\sqrt{\det(K_{134,134})}\sqrt{\det(K_{234,234})}} \right| \;\leq\; \lambda.$$

As a function in λ, the volume $V_{1,2|56}(\lambda)$ is a cumulative distribution function on $[0,\infty)$. Our aim in this article is to determine the asymptotics of such a function for $\lambda \to 0$.

In Sect. 3 we shall explain the form of the asymptotics that is promised in (4). To find the exponents ℓ and m, the first step is to run the algebraic algorithm in Sect. 4. This answers the question of whether the hypersurface in Ω defined by $f=0$ has any singular points. The set of such points, known as the *singular locus*, is the zero set in Ω of the ideal

$$J \;=\; \left\langle f\,,\, \frac{\partial f}{\partial a_{13}}, \frac{\partial f}{\partial a_{14}}, \frac{\partial f}{\partial a_{23}}, \frac{\partial f}{\partial a_{24}}, \frac{\partial f}{\partial a_{35}}, \frac{\partial f}{\partial a_{36}}, \frac{\partial f}{\partial a_{45}}, \frac{\partial f}{\partial a_{46}} \right\rangle.$$

The tools of Sect. 4 reveal that its *real radical* [15] is the intersection of three prime ideals:

$$\begin{aligned}\sqrt[\mathbb{R}]{J} &= \left\langle \text{entries of } \begin{pmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{pmatrix} \cdot \begin{pmatrix} a_{35} & a_{36} \\ a_{45} & a_{46} \end{pmatrix} \right\rangle \\ &= \langle a_{13}, a_{14}, a_{23}, a_{24} \rangle \,\cap\, \langle a_{35}, a_{36}, a_{45}, a_{46} \rangle \\ &\quad \cap \left\langle 2\times 2 \text{ minors of } \begin{pmatrix} a_{13} & a_{23} & a_{45} & a_{46} \\ a_{14} & a_{24} & -a_{35} & -a_{36} \end{pmatrix} \right\rangle.\end{aligned}$$

Thus the hypersurface $\{f=0\}$ is singular. Its singular locus decomposes into three irreducible varieties, namely, two linear spaces of dimension 4 and one determinantal variety of dimension 5. In Sect. 6 we will return to this example, with a focus on a statistical application of the cumulative distribution function $V_{1,2|56}(\lambda)$. We will then show that (ℓ, m) equals $(1,1)$. □

This paper extends the work of Uhler et al. in [17] on the geometry of the strong-faithfulness assumption in the PC algorithm. Upper and lower bounds on the volume $V_G(\lambda)$ of the unfaithful region $\mathrm{Tube}_G(\lambda)$ for the low- and high-dimensional settings were derived in [17, §5]. These bounds involved only the number $|E|$ of parameters and the degrees of the correlation hypersurfaces $\{\det(K_{iR,jR})=0\}$. The new insight in the current paper is that singularities are essential for the asymptotic behavior of $V_G(\lambda)$ for $\lambda \to 0$.

What led us to this insight was taking a closer look at the simulation results for trees. In [17, §6.1.1] trees were still treated as one single class. We subsequently examined the difference between stars and chains, as depicted in Fig. 1a, b. Our simulation results for Star_p and Chain_p are shown in Fig. 2. We shall now explain the curves in these diagrams.

The left diagram in Fig. 2 is for $p=6$ nodes and the right diagram is for $p=10$. Each curve is the graph of the cumulative distribution function $V_G(\lambda)$, but with the

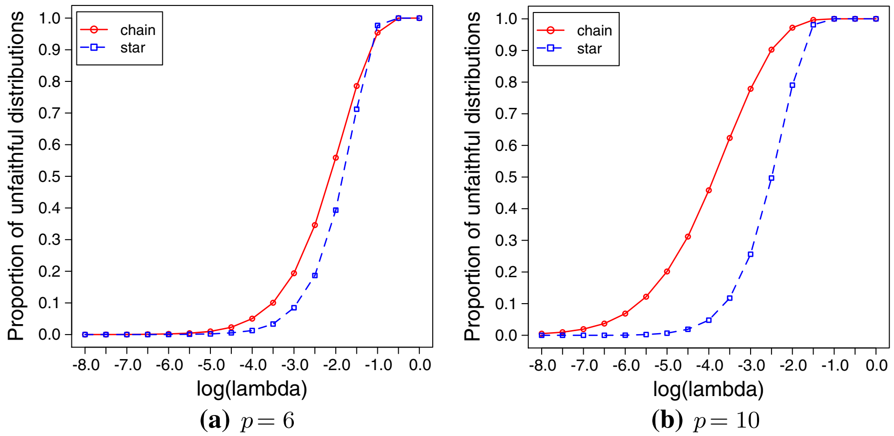


Fig. 2 Proportion of λ-strong-unfaithful distributions for chains compared to stars

x-axis transformed into a logarithmic scale (with base 10). Thus we depict the graph of the function

$$(-\infty, 0] \to [0, 1], \; x \mapsto V_G(10^x). \tag{6}$$

The red curve is for $G = \text{Chain}_p$ and the blue curve for $G = \text{Star}_p$. These curves were computed by simulation: we sampled the parameter a from the uniform distribution on $[-1, 1]^{p-1}$ and recorded the proportion of trials that landed in $\text{Tube}_G(\lambda)$ for various values of λ. The diagrams show clearly that $V_G(\lambda)$ is smaller for star graphs than for chainlike graphs.

A theoretical explanation for these experimental results will be given in Sect. 5. Our asymptotic theory predicts the behavior of these curves as $x = \log(\lambda)$ tends to $-\infty$. The point is that the correlation hypersurfaces for chainlike graphs have deeper singularities than those for star graphs. The equation of any such hypersurface for a tree is the product of a monomial and a strictly positive polynomial. This enables us to apply Proposition 3.5. In Theorem 5.1 and Corollary 5.3 we shall determine the constants ℓ, m, and C of (4) exactly when the graph G is a tree. We shall also address the question of how to obtain ℓ, m, and C from simulations.

Before we get to graphical models, however, we first need to develop the mathematics needed to analyze $V_G(\lambda)$. This will be done, in a self-contained manner, in the next section.

3 Computing the Volume of a Tube

We now introduce the basics regarding the computation of integrals like the one in (3), and we explain why asymptotic formulas like (4) can be expected. While this section is foundational for what is to follow, no reference to any statistical application is made until Theorem 3.8. It can be read from first principles and might be of independent interest to a wider audience.

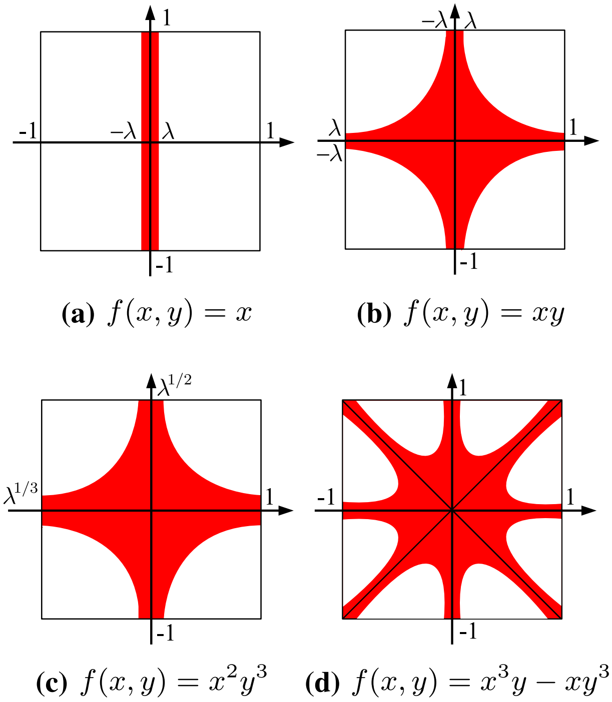


Fig. 3 Tubes for various polynomials in two variables

Let $\Omega \subset \mathbb{R}^d$ be a compact, full-dimensional, semianalytic subset, and consider a probability measure $\varphi(\omega)d\omega$ on Ω, where $d\omega$ is the standard Lebesgue measure and $\varphi : \Omega \to \mathbb{R}$ a real-analytic function. Also, fix an analytic function $f : \Omega \to \mathbb{R}$ whose hypersurface $\{\omega : f(\omega) = 0\}$ has a nonempty intersection with the interior of Ω. We are interested in the volume $V(\lambda)$ with respect to the measure φ of the region

$$\mathrm{Tube}(\lambda) = \big\{\omega \in \Omega : |f(\omega)| \le \lambda\big\}.$$

Here $\lambda > 0$ is a parameter that is assumed to be small. In later sections, we often take Ω to be the cube $[-1,+1]^d$, with φ its Lebesgue probability measure, and f is usually a polynomial.

The asymptotics of the volume function $V(\lambda)$ depends on the singularities of the hypersurface $\{f = 0\}$. This phenomenon is illustrated in Fig. 3. Our measure for the complexity of the singularities of f is a pair (ℓ, m) of nonnegative real numbers. That pair is the real log-canonical threshold (RLCT) of f. It is related to the volume $V(\lambda)$ for small values of λ by the formula

$$V(\lambda) \;\approx\; C\,\lambda^{\ell}(-\ln\lambda)^{m-1}. \tag{7}$$

Here C is a positive real constant whose study we shall defer until Sect. 8.

Example 3.1 Let $d = 2$ and φ be the Lebesgue probability measure on the square $\Omega = [-1, +1]^2$. Our problem is to compute the area of the tube $\{(x, y) \in \Omega : |f(x, y)| \leq \lambda\}$. Here $f(x, y)$ is one of the four simple polynomials below whose tubes are shown in Fig. 3.

(a) $f(x, y) = x$: the corresponding tube is a rectangle, and its area equals

$$V(\lambda) = \lambda.$$

Thus, in this example, we have $(\ell, m) = (1, 1)$ and $C = 1$. For other lines, the value of C will change. Proposition 3.6 below shows that $(\ell, m) = (1, 1)$ for smooth hypersurfaces.

(b) $f(x, y) = xy$: the tube in Fig. 3b consists of four copies of a region that is the union of a small rectangle and a certain area under a hyperbola. Using calculus, we find

$$V(\lambda) = 4\left(\lambda + \int_{\lambda}^{1} \frac{\lambda}{x}\mathrm{d}x\right)\frac{1}{4} = \lambda(-\ln\lambda) + \lambda.$$

The logarithm function appears in this case. We have $(\ell, m) = (1, 2)$ and $C = 1$.

(c) $f(x, y) = x^2y^3$: the corresponding tube is shown in Fig. 3c. Its area equals

$$V(\lambda) = 4\left(\lambda^{1/2} + \int_{\lambda^{1/2}}^{1} \lambda^{1/3}x^{-2/3}\mathrm{d}x\right)\frac{1}{4} = 3\lambda^{1/3} - 2\lambda^{1/2}.$$

Thus, the RLCT equals $(\ell, m) = (\frac{1}{3}, 1)$, and we have $C = 3$. See Proposition 3.5 for a formula for (ℓ, m) when f is a monomial in any number of variables.

(d) $f(x, y) = xy(x + y)(x - y)$: the corresponding tube is shown in Fig. 3d. This example is a slight generalization of Fig. 3b. As in Fig. 3b there is just one singularity at the origin, given by the intersection of lines. Computing the area $V(\lambda)$ is more challenging. In Example 7.3 we shall see that the RLCT equals $(\ell, m) = (\frac{1}{2}, 1)$.

For general bivariate polynomials $f(x, y)$ we are facing a hard calculus problem, namely, integrating the function $y = y(x)$ that is defined implicitly by $f(x, y) = \lambda$. We can approach this by expanding y as a Puiseux series in λ whose coefficients depend on x. Integrating these coefficients leads to asymptotic formulas in λ. These are consistent with what follows. □

We now return to the general setting defined at the beginning of this section. Let W be a random variable taking values in Ω with distribution φ. The volume $V(\lambda)$ with respect to the measure φ can then be viewed as the cumulative distribution function of the random variable $f(W)$. The corresponding probability distribution function $v(\lambda) = \mathrm{d}V/\mathrm{d}\lambda$ is called the *state density function*. Its Mellin transform is known as the *zeta function* of f. It is denoted by

$$\zeta(z) = \int_0^\infty \lambda^{-z} v(\lambda) \mathrm{d}\lambda = \int_\Omega |f(\omega)|^{-z} \varphi(\omega) \mathrm{d}\omega \qquad \text{for } z \in \mathbb{C}.$$

According to asymptotic theory [1,14,18], our volume has the complete asymptotic series expansion

$$V(\lambda) \approx \sum_\ell \sum_{m=1}^{d} C_{\ell,m} \lambda^\ell (-\ln \lambda)^{m-1}. \tag{8}$$

Here the index ℓ runs over some arithmetic progression of positive rational numbers and d is the dimension of the parameter space Ω. Equation (8) is valid for sufficiently small $\lambda > 0$. To be precise, writing $V(\lambda) \approx \sum_{i=1}^{\infty} g_i(\lambda)$, where $g_1(\lambda) > g_2(\lambda) > \cdots$ for small λ, means that

$$\lim_{\lambda \to 0} \frac{V(\lambda) - \sum_{i=1}^{k} g_i(\lambda)}{g_k(\lambda)} = 0 \qquad \text{for each positive integer } k. \tag{9}$$

Using the little-o notation, this is equivalent to $V(\lambda) = \sum_{i=1}^{k} g_i(\lambda) + o(g_k(\lambda))$ as $\lambda \to 0$ for each positive integer k. It is a common misconception to think that the infinite series converges to $V(\lambda)$ for each fixed λ when λ is small. Rather, it means that for each fixed k, the k-term approximation for $V(\lambda)$ gets better as $\lambda \to 0$. We will primarily be interested in the first term approximation (7).

Definition 3.2 ([14, §4.1], [18, §7.1]) Here we define the *RLCT* (ℓ, m) of f over Ω with respect to φ. This is a pair in $\mathbb{Q}_+ \times \mathbb{Z}_+$, which we denote by $\mathrm{RLCT}_\Omega(f; \varphi)$. It measures the complexity of the singularities of the hypersurface defined by $f(\omega) = 0$. The following four definitions of $\mathrm{RLCT}_\Omega(f; \varphi) = (\ell, m)$ are known to be equivalent:

(i) For large $N > 0$, the *Laplace integral*

$$Z(N) = \int_\Omega e^{-N|f(\omega)|} \varphi(\omega) \, \mathrm{d}\omega$$

is asymptotically $CN^{-\ell}(\ln N)^{m-1}$ for some constant C.

(ii) The *zeta function*

$$\zeta(z) = \int_\Omega |f(\omega)|^{-z} \varphi(\omega) \, \mathrm{d}\omega$$

has its smallest pole at $z = \ell$, and that pole has multiplicity m.

(iii) For small $\lambda > 0$, the *volume function*

$$V(\lambda) = \int_{|f(\omega)|\le\lambda} \varphi(\omega)\, d\omega$$

is asymptotically $C\lambda^{\ell}(-\ln\lambda)^{m-1}$ for some constant C.

(iv) For small $\lambda > 0$, the *state density function*

$$v(\lambda) = \frac{d}{d\lambda} \int_{|f(\omega)|\le\lambda} \varphi(\omega)\, d\omega$$

is asymptotically $C\lambda^{\ell-1}(-\ln\lambda)^{m-1}$ for some constant C.

If the real analytic hypersurface $\{\omega \in \Omega : f(\omega) = 0\}$ is empty, then we set $\ell = \infty$ and leave m undefined. We say that $(\ell_1, m_1) < (\ell_2, m_2)$ if $\ell_1 < \ell_2$ or if $\ell_1 = \ell_2$ and $m_1 > m_2$. Hence, the pairs are ordered reversely by the size of $\lambda^{\ell}(-\ln\lambda)^{m-1}$ for sufficiently small $\lambda > 0$.

Let us provide some intuition for the ordering of the pairs (ℓ, m). The RLCT is a measure of complexity for singularities. Analytic varieties can be stratified into subsets where this measure is constant. The highest stratum contains the smooth points of the variety. As we go deeper, to strata with lower RLCTs, we encounter singularities of increasing complexity. The volumes of λ-fattenings of deeper singularities will, asymptotically as λ goes to zero, also be larger than those of their less complex counterparts. For instance, in both panels a and b of Fig. 3 the singular locus consists of the origin, but the λ-fattening of the origin in Fig. 3c is larger than in Fig. 3b. See also Example 3.7.

Example 3.3 Let $f(\omega) = \omega_1^2 + \omega_2^2 + \cdots + \omega_d^2$, and let φ be the Lebesgue probability measure on $\Omega = [-1, +1]^d$. Then Tube(λ) is the *standard ball* of radius $\lambda^{1/2}$ whose φ-volume is

$$V(\lambda) = \frac{\pi^{d/2}}{2^d \cdot \Gamma(\frac{d}{2}+1)} \cdot \lambda^{d/2}.$$

By Definition 3.2 (iii), the RLCT equals $\mathrm{RLCT}_\Omega(f;\varphi) = (d/2, 1)$. □

We now list some formulas for computing the RLCT. A first useful fact is that $\mathrm{RLCT}_\Omega(f;\varphi)$ is independent of the underlying measure φ as long as it is positive everywhere. We can thus assume that φ is the uniform distribution on Ω.

Proposition 3.4 *If $\varphi : \Omega \to \mathbb{R}$ is strictly positive and* 1 *denotes the constant unit function on Ω, then*

$$\mathrm{RLCT}_\Omega(f;\varphi) = \mathrm{RLCT}_\Omega(f;1).$$

Proof See [14, Lemma 3.8]. □

Proposition 3.5 *Suppose that* Ω *is a neighborhood of the origin. If* $f(\omega) = \omega_1^{\kappa_1} \cdots \omega_d^{\kappa_d} g(\omega)$*, where the* κ_i *are nonnegative integers and the function* $g : \Omega \to \mathbb{R}$ *has no zeros, then* $\mathrm{RLCT}_\Omega(f; 1) = (\ell, m)$*, where*

$$\ell = \min_i \frac{1}{\kappa_i} \quad \text{and} \quad m = \left| \left\{ \operatorname{argmin}_i \frac{1}{\kappa_i} \right\} \right|.$$

Proof This is a special case of Theorem 7.1 that will be proved later. □

Recall that an analytic hypersurface $\{f(\omega) = 0\}$ is *singular* at a point $\omega \in \Omega$ if ω satisfies

$$f(\omega) = 0 \quad \text{and} \quad \frac{\partial f}{\partial \omega_i}(\omega) = 0 \quad \text{for } i = 1, \dots, d.$$

If the hypersurface is not singular at any point $\omega \in \Omega$, then it is said to be *smooth*.

Proposition 3.6 *If the hypersurface* $\{f(\omega) = 0\}$ *is smooth, then* $\mathrm{RLCT}_\Omega(f; 1) = (1, 1)$.

Proof This is also a special case of Theorem 7.1. □

Example 3.7 Following up on Example 3.1, we now consider an arbitrary monomial function $f(x, y) = x^s y^t$ on the square $\Omega = [-1, 1]^2$. The tube appears as in Fig. 3c. Its area satisfies

$$V(\lambda) \approx \begin{cases} C\lambda^{1/s} & \text{if } s < t, \\ C\lambda^{1/t} & \text{if } s > t, \\ C\lambda^{1/s}(-\ln \lambda) & \text{if } s = t. \end{cases}$$

This formula for the asymptotics (7) follows from Definition 3.2(iii) and Proposition 3.5. □

For the statistical applications in this paper, the relevant functions f are polynomials. They are determinants $f = \det(K_{iR,jR})$, where $R = V \backslash (S \cup \{i, j\})$ as in Sect. 1. Let $\mathrm{RLCT}(i, j|S)$ denote the corresponding RLCT over $\Omega = [-1, 1]^E$ with respect to a positive density φ. The theory developed so far says that the RLCT of the correlation hypersurface gives an asymptotic volume formula for $V_{i,j|S}(\lambda)$.

Theorem 3.8 *If* φ *satisfies the assumptions in Proposition 3.4, then as* λ *tends to zero, the volume of the region* $\mathrm{Tube}_{i,j|S}(\lambda)$ *[see (2)] is asymptotically*

$$V_{i,j|S}(\lambda) \approx C \lambda^{\ell} (-\ln \lambda)^{m-1}$$

for some constant $C > 0$ *(which only depends on* G*) and* $(\ell, m) = \mathrm{RLCT}(i, j|S)$.

Proof By part (iii) in Definition 3.2, the desired pair (ℓ, m) is the RLCT of the partial correlation $f = \mathrm{corr}(i, j|S)$. This is the algebraic (and, hence, analytic) function in (1). This function differs from the polynomial $\det(K_{iR,jR})$ by a denominator that does not vanish over Ω. That denominator is a unit in the ring of real analytic functions over Ω, and multiplying by a unit does not change the RLCT of an analytic function [14, §4.1]. □

We close this section by relating our results to the study of unfaithfulness in [17].

Corollary 3.9 *Under the assumptions in Theorem 3.8, as λ tends to zero, the volume of λ-strong-unfaithful distributions satisfies*

$$V_G(\lambda) \approx C\,\lambda^{\ell}(-\ln\lambda)^{m-1}$$

for some constant $C > 0$. Here (ℓ, m) is the minimum of the pairs $\mathrm{RLCT}(i, j|S)$, *where (i, j, S) runs over all triples in the DAG G such that i is not d-separated from j given S.*

Proof The function $V_G(\lambda)$ is the volume of the union of the regions $\mathrm{Tube}_{i,j|S}(\lambda)$. Thus,

$$\max_{i,j,S} V_{i,j|S}(\lambda) \le V_G(\lambda) \le \sum_{i,j,S} V_{i,j|S}(\lambda).$$

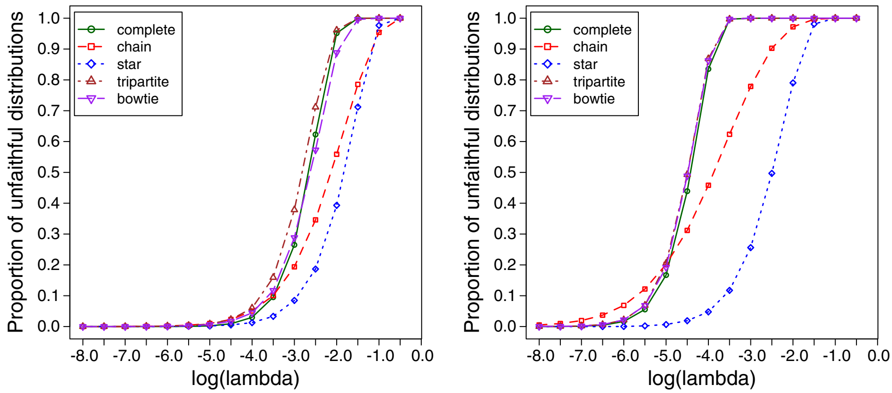
Asymptotically, for small positive values of λ, both the lower and upper bounds vary like a constant multiple of $\lambda^{\ell}(-\ln\lambda)^{m-1}$, where (ℓ, m) is the minimum over all pairs $\mathrm{RLCT}(i, j|S)$. In this minimum, (i, j, S) runs over all triples such that i and j are d-connected given S. □

4 Singular Locus

The asymptotic integration theory in Sect. 3 requires us to analyze the singular locus $\mathrm{Sing}(f)$ of the real algebraic hypersurface determined by a given polynomial f. If $\mathrm{Sing}(f)$ is empty, then the hypersurface is smooth and Proposition 3.6 characterizes the asymptotics of the integral. In this section we return to Gaussian graphical models, develop tools for computing the relevant singular loci, and show that they are empty in many cases. In many of the remaining cases, the singularities are of the monomial type featured in Proposition 3.5.

Consider any almost-principal minor $f = \det(K_{iR,jR})$ of the concentration matrix K of a DAG G. This is a polynomial function on the parameter space $\mathbb{R}^E$. This polynomial and its partial derivatives are elements in the polynomial ring $\mathbb{Q}[a_{ij} : (i, j) \in E]$. The *Jacobian ideal* of f is the ideal in this polynomial ring generated by f and its partials. We denote it by

$$\mathrm{Jacob}_{i,j,R} := \langle f \rangle + \left\langle \frac{\partial f}{\partial a_{ij}} : (i, j) \in E \right\rangle.$$



(a) $p = 6$ **(b)** $p = 10$

Fig. 4 $V_G(\lambda)$ for complete graph K_p compared to Chain_p, Star_p, $\mathrm{Tripart}_{p,2}$, Bow_p

The singular locus $\mathrm{Sing}(f)$ is the subvariety of real affine space $\mathbb{R}^E$ defined by the Jacobian ideal $\mathrm{Jacob}_{i,j,R}$. The structure of the real variety $\mathrm{Sing}(f)$ governs the volume $V_{i,j|S}(\lambda)$ of the set $\mathrm{Tube}_{i,j|S}(\lambda)$ of unfaithful parameters. If $\mathrm{Sing}(f) = \emptyset$, then Proposition 3.6 tells us that $V_{i,j|S}(\lambda)$ asymptotically equals $C\lambda$ for some constant $C > 0$. If the singular locus is not empty, then understanding $\mathrm{Sing}(f)$ is essential for computing its RLCT (ℓ, m).

We conducted a comprehensive study of all DAGs with few nodes by computing the singular locus for every almost-principal minor in their concentration matrix K. Our first result concerns the special case of complete graphs. Noncomplete graphs will be studied later.

Theorem 4.1 *Suppose that φ satisfies the assumptions in Proposition 3.4. For any conditional independence statement on the complete directed graph K_p with $p \le 6$ nodes, we have* $\mathrm{Sing}(f) = \emptyset$, *and hence $V_{i,j|S}(\lambda) \approx C\lambda$ for all triples* (i, j, S).

It is tempting to conjecture that the hypothesis $p \le 6$ can be removed in this theorem. Presently we do not know how to approach this problem other than by direct calculation.

Applying Corollary 3.9, this means that the volume of λ-strong-unfaithful distributions for the complete graph satisfies $V_{K_p}(\lambda) \approx C\lambda$ for $\lambda \to 0$, which is the best possible behavior regarding strong faithfulness. This may be counterintuitive, but it has been confirmed in simulations. In Fig. 4 we plot, via (6), the proportion of strong-unfaithful distributions $V_G(\lambda)$ for the five graphs in Sect. 2 for varying values of λ. Especially in the plot for $p = 10$ it becomes apparent that the behavior for $\lambda \to 0$ is very different than, say, for $\lambda = 0.001$. For $\lambda \to 0$ we have $V_{\mathrm{complete}}(\lambda) < V_{\mathrm{chain}}(\lambda)$, although the chainlike graph is much sparser than the complete graph. Note also that the complete graph K_{10} has $\sum_{k=2}^{10} \binom{10}{k}\binom{k}{2} = 11520$ relevant triples (i, j, S), whereas for Chain_{10} there are only $\sum_{k=1}^{9} k2^{k-1} = 4097$ such triples.

In what follows we explain the algebraic computations that led to Theorem 4.1. We used ideal-theoretic methods from [3] in their implementation in the Gröbner-based software packages `Macaulay 2` [5] and `Singular` [4]. An important point to note at the outset is that the ideal $\mathrm{Jacob}_{i,j,R}$ is almost never the unit ideal. By Hilbert's Nullstellensatz, this means that the hypersurfaces $f = \det(K_{iR,jR})$ have plenty of singular points over the field $\mathbb{C}$ of complex numbers. What Theorem 4.1 asserts is that, in many of the cases of interest to us here, none of those singular points have their coordinates in the field $\mathbb{R}$ of real numbers.

To study the real variety of an ideal, techniques from real algebraic geometry are needed. A key technique is to identify a sum of squares (SOS). Indeed, the *real Nullstellensatz* [15] states that the real variety is empty if and only if the given ideal contains a certain type of SOS. To apply this to directed Gaussian graphical models, we shall use the fact that every principal minor of the covariance matrix or the concentration matrix furnishes such a SOS.

Lemma 4.2 *Every principal minor* $\det(K_{R,R})$ *of the concentration matrix* K *of a DAG is equal to* 1 *plus a sum of squares in* $\mathbb{Q}[a_{ij} : (i,j) \in E]$. *In particular, its real variety is empty.*

Proof We can write the principal submatrix $K_{R,R}$ as the product $(A - I)_{R,*} \cdot ((A - I)_{R,*})^T$, where $(\ )_{R,*}$ refers to the submatrix with row indices R. Thus $K_{R,R}$ is the product of an $|R| \times p$ matrix and its transpose. By the Cauchy–Binet formula, $\det(K_{R,R})$ equals the sum of squares of all maximal minors of the $|R| \times p$ matrix $(A - I)_{R,*}$. One of these maximal minors is the identity matrix. Hence the polynomial $\det(K_{R,R})$ has the form $1 + \mathrm{SOS}$. In particular, the matrix $K_{R,R}$ is invertible for all parameter values in $\mathbb{R}^E$. $\square$

We note that Lemma 4.2 holds more generally also in the case of unequal noise variances. In the context of commutative algebra, it now makes sense to introduce the *saturations*

$$\mathrm{Singu}^*_{i,j,R} = \left(\mathrm{Singu}_{i,j,R} : \left(\prod_{(i,j)\in E} a_{ij} \right)^{\infty} \right).$$

These are also ideals in $\mathbb{Q}[a_{ij} : (i,j) \in E]$. By definition, $\mathrm{Singu}_{i,j,R}$ consists of all polynomials that get multiplied into the Jacobian ideal by some power of the determinant of $K_{R,R}$, and $\mathrm{Singu}^*_{i,j,R}$ consists of polynomials that get multiplied into $\mathrm{Singu}_{i,j,R}$ by some monomial. By [3, §4.4], the variety of $\mathrm{Singu}_{i,j,R}$ is the Zariski closure of the set-theoretic difference of the variety of $\mathrm{Jacob}_{i,j,R}$ and the hypersurface $\{\det(K_{R,R}) = 0\}$. We saw in Lemma 4.2 that the latter hypersurface has no real points. The ideal $\mathrm{Singu}^*_{i,j,R}$ represents singularities in $(\mathbb{R}\backslash\{0\})^E$.

Corollary 4.3 *The singular locus of the real algebraic hypersurface* $\{\det(K_{iR,jR}) = 0\}$ *in* $\mathbb{R}^E$ *coincides with the set of real zeros of the ideal* $\mathrm{Singu}_{i,j,R}$. *The set of real zeros of* $\mathrm{Singu}^*_{i,j,R}$ *is the Zariski closure of the subset of all singular points whose coordinates are nonzero.*

Proof of Theorem 4.1 We computed the ideals $\mathrm{Jacob}_{i,j,R}$ and $\mathrm{Singu}_{i,j,R}$ for every almost-principal minor $K_{iR,jR}$ in the concentration matrices of the graphs $G = K_3, K_4, K_5, K_6$. In all cases the ideal $\mathrm{Singu}_{i,j,R}$ was found to equal the unit ideal $\langle 1 \rangle$. These exhaustive computations were carried out using the software `Singular` [4]. This establishes Theorem 4.1. □

We briefly discuss our computations for the complete directed graph on six nodes.

Example 4.4 Fix the complete directed graph $G = K_6$. We tested all 240 conditional independence statements and computed the corresponding ideal $\mathrm{Singu}_{i,j,R}$. We discuss one interesting instance, namely, $i = 1$, $j = 3$, $R = \{2,4\}$. The almost-principal minor $K_{241,243} =$

$$\begin{pmatrix} a_{23}^2 + a_{24}^2 + a_{25}^2 + a_{26}^2 + 1 & a_{25}a_{45} + a_{26}a_{46} - a_{24} & a_{24}a_{34}+a_{25}a_{35}+a_{26}a_{36}-a_{23} \\ a_{25}a_{45} + a_{26}a_{46} - a_{24} & a_{45}^2 + a_{46}^2 + 1 & a_{35}a_{45} + a_{36}a_{46} - a_{34} \\ a_{13}a_{23}+a_{14}a_{24}+a_{15}a_{25}+a_{16}a_{26}-a_{12} & a_{15}a_{45} + a_{16}a_{46} - a_{14} & a_{14}a_{34} + a_{15}a_{35} + a_{16}a_{36} \end{pmatrix}$$

contains all 15 parameters except a_{56}. Its determinant is a polynomial of degree 6. Of its 14 partial derivatives, 13 have degree 5. The derivative with respect to a_{12} has degree 4. Thus $\mathrm{Jacob}_{1,3,\{2,4\}}$ is generated by 15 polynomials of degrees $4, 5, \ldots, 5, 6$. The matrix $K_{24,24}$ is the upper left 2×2-block in the preceding matrix. The square of its determinant is a polynomial of degree 8 that happens to lie in the ideal $\mathrm{Jacob}_{1,3,\{2,4\}}$. This proves $\mathrm{Singu}_{1,3,\{2,4\}} = \langle 1 \rangle$. □

For graphs G that are not complete, $\mathrm{Singu}_{i,j,R}$ may not be the unit ideal. We already saw one nonobvious instance of this for the tripartite graph in Example 2.1. Here is an even smaller example where the Jacobian ideal and its saturations are equal and not the unit ideal.

Example 4.5 Let $p = 4$, and take G to be the almost-complete graph with adjacency matrix

$$A_G \quad = \quad \begin{pmatrix} 0 & 0 & a_{13} & a_{14} \\ 0 & 0 & a_{23} & a_{24} \\ 0 & 0 & 0 & a_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The conditional independence statement $1 \perp\!\!\!\perp 2 \,|\, 4$ is represented by the almost-principal minor

$$K_{31,32} = \begin{pmatrix} a_{34}^2 + 1 & a_{24}a_{34} - a_{23} \\ a_{14}a_{34} - a_{13} & a_{13}a_{23} + a_{14}a_{24} \end{pmatrix}$$

of the concentration matrix. The determinant of this minor factors into two binomials:

$$\det(K_{31,32}) = (a_{13}a_{34} + a_{14})(a_{23}a_{34} + a_{24}). \tag{10}$$

Table 1 RLCT for all DAGs with three nodes

	(1, 1)	(1, 2)	Subtotal
Monomial	21	3	24
Smooth	3		3
Subtotal	24	3	27

Table 2 RLCT for all DAGs with four nodes

	(1, 1)	(1, 2)	(1, 3)	(1/2, 1)	Subtotal
Monomial	568	145	14	1	728
Smooth	198				198
Normal crossing		22	2		24
Blowup	12				12
Special	2	1			3
Subtotal	780	168	16	1	965

The Jacobian ideal is the prime ideal generated by these factors:

$$\mathrm{Jacob}_{1,2,3} = \mathrm{Singu}_{1,2,3} = \mathrm{Singu}^{*}_{1,2,3} = \langle\, a_{13}a_{34} + a_{14}\, ,\ a_{23}a_{34} + a_{24}\, \rangle.$$

The left equality holds because $\det(K_{3,3}) = a_{34}^2 + 1$ is a non-zero-divisor modulo $\mathrm{Jacob}_{1,2,3}$. The singular locus of (10) is the three-dimensional real variety defined by this binomial ideal in the parameter space $\mathbb{R}^5$. Its RLCT is found to be $(\ell, m) = (1, 2)$. □

This example inspired us to analyze the partial correlations of all small DAGs with $p \leq 4$ nodes. In our experiments, we found that $\det(K_{iR,jR})$ is frequently the product of a monomial with a strictly positive sum of squares. This is the case when there is a unique path which d-connects nodes i and j given S. For instance, this holds for trees. Such cases are denoted as *Monomial* in Tables 1 and 2. For these, the RLCT is read off directly from Proposition 3.5. The rows labeled *Smooth* cover cases that are not monomial but where $\mathrm{Singu}_{i,j,R}$ is the unit ideal, so Proposition 3.6 gives us the RLCT. The next theorem summarizes the complete results. The trivial case $p = 2$ is excluded because there is only one graph $1 \to 2$, with $\mathrm{RLCT}(1, 2|\emptyset) = (1, 1)$. Here and in Tables 1 and 2 we enumerate unlabeled DAGs.

Theorem 4.6 *Under the assumptions in Theorem 3.8, for all DAGs with $p \leq 4$ nodes and all triples (i, j, S), the value $\mathrm{RLCT}(i, j|S)$ is given in Tables 1 and 2. In all cases but one, we have $\mathrm{RLCT}(i, j|S) = (1, m)$, where $m < p$.*

To establish Theorem 4.6, we listed every DAG G and every triple (i, j, S) that is not d-separated in G. The rows *Monomial* and *Smooth* were discussed earlier. On three nodes there are only three partial correlations that correspond to the weighted sum of more than one d-connecting path, namely, the partial correlations $\mathrm{corr}(1, 2 \mid 3)$, $\mathrm{corr}(1, 3)$, $\mathrm{corr}(2, 3)$ in the complete DAG $1 \to 2$, $2 \to 3$, $1 \to 3$. These are the

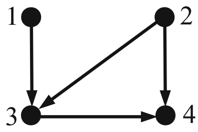


Fig. 5 Four-node DAG

three cases of smooth RLCTs in Table 1. The row *Normal crossing* refers to cases covered by Theorem 7.1. The *Special* cases are treated in Examples 4.5 and 4.8. Lastly, the row *Blowup* represents instances where the real singular locus is a linear space. Our computation of $\mathrm{RLCT}(i, j|S) = (\ell, m)$ for such instances uses the method in Example 7.4. We now examine the unique exceptional case where $\ell \neq 1$.

Example 4.7 Let $p = 4$ and $G = \mathrm{Tripart}_{4,1}$. Then the corresponding concentration matrix may be obtained from Example 4.5 by setting $a_{14} = a_{24} = 0$. The partial correlation for $1 \perp\!\!\!\perp 2 \,|\, 4$ is now given by

$$\det(K_{13,23}) = a_{13}a_{23}a_{34}^2.$$

For this monomial, Proposition 3.5 tells us that $(\ell, m) = \mathrm{RLCT}(1, 2|4) = (1/2, 1)$. □

Here is an case where the RLCT depends in a subtle way on the choice of Ω.

Example 4.8 Consider the conditional independence statement $1 \perp\!\!\!\perp 3 \,|\, 4$ for the DAG in Fig. 5. The partial correlation is represented by the almost-principal minor

$$\det(K_{12,23}) = a_{13} \cdot g, \qquad \text{where} \quad g = a_{23}a_{24}a_{34} + a_{24}^2 + 1.$$

The component $\{g = 0\}$ is smooth in $\mathbb{R}^4$. However, it is disjoint from the cube $\Omega = [-1, 1]^4$. To see this, note that $-1 \leq a_{23}a_{24}a_{34}$ in Ω. With this, $g = 0$ would imply $a_{24} = 0$ and, hence, $g = 1$, a contradiction. Consequently, if Ω is the cube $[-1, 1]^4$, then the correlation hypersurface is simply $\{a_{13} = 0\}$, and the RLCT equals $(1, 1)$ by Proposition 3.5. The other special case with $\mathrm{RLCT} = (1, 1)$ in Table 2 comes from swapping the labels of nodes 1 and 2.

Now, if we enlarge the parameter space Ω, then the situation changes. For instance, suppose $(a_{13}, a_{23}, a_{24}, a_{34}) = (0, -2, 1, 1)$ is in the interior of Ω. This is a singular point of $\det(K_{12,23}) = a_{13} \cdot g$. The RLCT can be computed by applying Theorem 7.1. It is now $(1, 2)$ instead of $(1, 1)$. This example shows that the asymptotics of $\mathrm{V}_{i,j|S}(\lambda)$ depends on Ω. However, it is possible to choose Ω in such a way that further enlargement will not cause the asymptotics of $V_G(\lambda)$ to change. Such a choice could be used as a worst-case analysis for $V_G(\lambda)$, but to avoid complicating the paper, we will not explore this any further. □

Remark 4.9 We briefly return to the issue of faithfulness in the PC algorithm. Zhang and Spirtes [20] introduced a variant known as the *conservative PC algorithm*. As the name suggests, this algorithm is more conservative and may decide not to orient certain edges. The conservative PC algorithm only requires *adjacency faithfulness* for correct inference, which is simply strong faithfulness restricted to the edges of G:

$$|\operatorname{corr}(i, j|S)| > \lambda \qquad \text{for all } (i, j) \in E \text{and} S \subset V \backslash \{i, j\}.$$

If $\{i, j\}$ is not adjacent to R, then the relevant minor equals $\det(K_{iR,jR}) = a_{ij}\det(K_{R,R}) + f(\bar{a})$, where f is a polynomial in $\bar{a} = \{a_{st} \mid (s,t) \neq (i, j)\}$, the correlation hypersurface is smooth, and $(\ell, m) = (1, 1)$. If $\{i, j\}$ is adjacent to R, then the behavior can be more complicated, as seen in Example 4.8.

5 Asymptotics for Trees

In [17] trees were treated as one class. However, as noted in the discussion of Fig. 2, there is a striking difference between the volume $V_G(\lambda)$ for chainlike graphs compared to stars. In this section we give an explanation for this difference based on RLCTs.

We use the notation $SOS(a)$ for any polynomial that is a sum of squares of polynomials in the model parameters $(a_{ij})_{(i,j)\in E}$. Suppose that G is a tree on $V = \{1, 2, \dots, p\}$, and let m be the longest length of an undirected path in G. It was shown in [17, Corollary 4.3(a)] that any nonzero almost-principal minor of the concentration matrix K has the form

$$\det(K_{iR,jR}) = (1 + SOS(a)) \cdot a_{i\to j}, \tag{11}$$

where $a_{i\to j}$ is the monomial of degree $\leq m$ formed by multiplying the parameters a_{rs} along the unique path between i and j. Specifically, for the two trees in Fig. 1 we have

$$\det(K_{iR,jR}) = \begin{cases} (1 + SOS(a)) \prod_{k=i}^{j-1} a_{k,k+1} & \text{if } \ G = \text{Chain}_p, \\ (1 + SOS(a)) \cdot a_{1,i}a_{1,j} & \text{if } \ G = \text{Star}_p, \ \text{ and } i, j > 1. \end{cases}$$

In both cases, the term $SOS(a)$ disappears when i and j are leaves of the tree G; cf. (13) and (14).

Since the correlation hypersurfaces for trees are essentially given by monomials, we can apply Proposition 3.5. The minimal RLCT is $(1, m)$, where m is the largest degree of any of the monomials in (11). Corollary 3.9 implies the following result.

Theorem 5.1 *Under the assumptions in Theorem 3.8, if G is a tree, then the volume of λ-strong-unfaithful distributions satisfies*

$$V_G(\lambda) \approx C\lambda(-\ln\lambda)^{m-1},$$

where m is the length of the longest path in the tree G, and C is a suitable constant.

In the case of stars we have $m = 2$, whereas for chainlike graphs we have $m = p-1$.

Corollary 5.2 *Under the assumptions in Theorem 5.1, the volume* $V_G(\lambda)$ *of strong-unfaithful distributions satisfies*

$$V_G(\lambda) \approx \begin{cases} C_{\text{chain}} \cdot \lambda(-\ln\lambda)^{p-2} & \text{if} \quad G = \text{Chain}_p, \\ C_{\text{star}} \cdot \lambda(-\ln\lambda) & \text{if} \quad G = \text{Star}_p, \end{cases} \tag{12}$$

where C_{chain} *and* C_{star} *are suitable positive constants.*

As a consequence, the volume $V_G(\lambda)$ is asymptotically larger for chains compared to stars, and the difference increases with an increasing number of nodes p. This furnishes an explanation for Fig. 2, at least for small values of λ. In that figure we saw the curve for the chain lying clearly above the curve for the star tree. However, one subtle issue is the size of the constants C_{chain} and C_{star}. These need to be understood in order to make accurate comparisons.

In Sect. 8, we develop new theoretical results regarding the computation of the constant C in (7). Theorem 8.5 gives an integral representation for C when the partial correlation hypersurface is essentially defined by a monomial. In Example 8.7 we shall then derive the following corollary.

Corollary 5.3 *The two constants in (12) are*

$$C_{\text{chain}} = \frac{1}{(p-2)!} \quad \text{and} \quad C_{\text{star}} = \binom{p}{3}.$$

This result surprised us at first. It establishes the counterintuitive fact that as p grows, the constant for the lower curve in Fig. 2 is exponentially larger than that for the upper curve. Therefore, to fully explain the relative position of the two curves for a wider range of values of $\lambda > 0$, it does not suffice to just consider the first order asymptotics (7). Instead, we need to consider some of the higher-order terms in the series expansion (8).

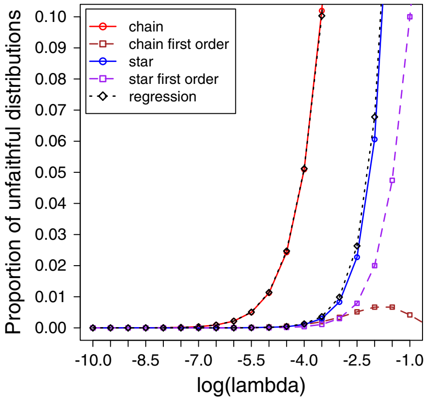
As we shall see in Sect. 8, it is difficult to determine the constants $C_{\ell,m}$ in (8) analytically. In the remainder of this section, we propose a procedure based on simulation and linear regression for estimating the constants $C_{\ell,m}$ in the asymptotic explanations of the volumes $V_G(\lambda)$ and $V_{i,j|S}(\lambda)$. For simplicity we focus on the latter case and we take $f = \det(K_{iR,jR})$.

Suppose that G is a DAG for which the RLCTs (ℓ, m) in Theorem 3.8 and Corollary 3.9 are known. This is the case for all trees by Theorem 5.1. Our procedure goes as follows. We first sample n points uniformly from Ω and compute the proportion of points ω that lie in $\text{Tube}_{i,j|S}(\lambda)$ for different values of λ. We then fit a linear model to

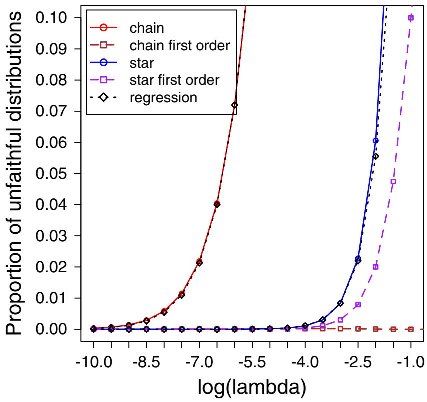
$$\frac{V_{i,j|S}(\lambda)}{\lambda^{\ell}} \approx C_{m-1}(-\ln\lambda)^{m-1} + C_{m-2}(-\ln\lambda)^{m-2} + \cdots + C_0,$$

where (ℓ, m) is the known RLCT.

In what follows, we illustrate this procedure for chains and stars. We analyze two specific examples of partial correlation volumes, namely, those corresponding to the



(a) $p=6$



(b) $p=10$

Fig. 6 Regression-based asymptotics for chains and stars

longest paths in each graph, that is, $V_{1,p|\emptyset}(\lambda)$ for Chain_p, and $V_{2,3|\emptyset}(\lambda)$ for Star_p. For chainlike graphs

$$\operatorname{corr}(1,p) = \frac{(-1)^p \prod_{i=1}^{p-1} a_{i,i+1}}{\sqrt{1 + a_{p-1,p}^2 \left(1 + a_{p-2,p-1}^2 \left(\cdots (1 + a_{12}^2)\right)\right)}}, \tag{13}$$

whereas for star graphs

$$\operatorname{corr}(2,3) = -\frac{a_{12}\, a_{13}}{\sqrt{(1 + a_{12}^2)(1 + a_{13}^2)}}. \tag{14}$$

We first approximate $V_{1,p|\emptyset}(\lambda)$ for chainlike graphs and $V_{2,3|\emptyset}(\lambda)$ for star graphs by simulation for various values of λ. This means that we sample n points uniformly in the $(p-1)$-dimensional parameter space Ω, and we count how many of them are $\leq \lambda$. The results for $p=6$ and $p=10$ are shown in Fig. 6. These are based on a sample size of $n =$1,000,000. We then fit a linear model

$$\frac{V_{1,p|\emptyset}(\lambda)}{\lambda} \approx C_{p-2}(-\ln\lambda)^{p-2} + C_{p-3}(-\ln\lambda)^{p-3} + \cdots + C_0$$

for chainlike graphs. The curve resulting from the regression estimates is shown in black in Fig. 6. The curve resulting from the first-order approximation with the constants computed using Corollary 5.3 is shown in gray in Fig. 6. We note that, especially for chainlike graphs, where the true constant in Corollary 5.3 is small, the first-order approximation is very bad.

The approximation by regression, on the other hand, is a fast way to get pretty accurate estimates of all constants. The same was done with star graphs, but with the linear model

$$\frac{V_{2,3|\emptyset}(\lambda)}{\lambda} \approx C_1(-\ln\lambda) + C_0.$$

Figure 6 shows that the first-order approximation is more accurate for stars than for chains.

6 Volume Inequalities for Bias Reduction in Causal Inference

We now discuss the problem of quantifying bias in causal models. Our point of departure is Greenland's paper [7], where the problem of quantifying bias is discussed for binary variables. In contrast to the previous sections, in the situation discussed here, a large tube volume is in fact desired since it corresponds to small bias. In this section we use the notation $K_{i,j|S}$ for the almost-principal minor $K_{iR,jR}$ of the concentration matrix.

We are interested in estimating the direct effect of an exposure E on a disease outcome D (i.e., the coefficient on the edge $E \to D$) from the partial correlation $\mathrm{corr}(E, D \mid S)$, where S is a subset of the measurable variables. This partial correlation is a weighted sum over all *open paths* (i.e., paths that d-connect E to D) given S (the direct path a_{ED} being just one of them). For estimating the direct effect a_{ED} from $\mathrm{corr}(E, D \mid S)$, all open paths other than the direct path are thus considered as bias. We shall analyze two forms of bias that are of particular interest in practice, namely, confounding bias and collider-stratification bias. We start by defining collider-stratification bias.

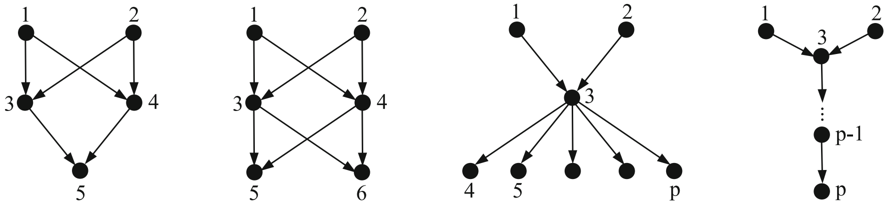
Suppose we are given a DAG G with $D, E \in V$ and there is another node C such that

$$E \to V_1 \to \cdots \to V_s \to C \leftarrow W_1 \leftarrow \cdots \leftarrow W_t \leftarrow D.$$

This says that C is a collider on a path from D to E. Stratifying (i.e., conditioning) on C opens a path between E and D leading to bias when estimating a_{ED}. The partial correlation corresponding to the opened path between E and D is known as *collider-stratification bias*. It arises, for example, in the context of discrete variables, where instead of obtaining a random sample from the full population, a random sample is obtained from the subpopulation of individuals with a particular level of C.

Example 6.1 We illustrate collider-stratification bias for the tripartite graph $G = \mathrm{Tripart}_{5,1}$ shown in Fig. 7a. Let node 1 represent the exposure E and node 2 the disease outcome D. In this example, node 5 is a collider C for multiple paths between E and D. When stratifying on $C = 5$, node $E = 1$ is d-connected to node $D = 2$ via the following paths:

$$1 \to 3 \to 5 \leftarrow 4 \leftarrow 2, \quad 1 \to 4 \to 5 \leftarrow 3 \leftarrow 2, \quad 1 \to 3 \to 5 \leftarrow 3 \leftarrow 2,$$
$$1 \to 4 \to 5 \leftarrow 4 \leftarrow 2.$$



(a) $\mathrm{Tripart}_{5,1}$ **(b)** $\mathrm{Tripart}_{6,2}$ **(c)** $\mathrm{Tripart}_{p,p-3}$ **(d)** Almost-chain graph

Fig. 7 Various tripartite and almost tripartite graphs

The bias introduced for estimating the direct effect of E on D when conditioning on C is

$$\mathrm{corr}(1,2\,|\,5) = \frac{a_{13}a_{35}a_{45}a_{24} + a_{14}a_{45}a_{35}a_{23} + a_{13}a_{35}^2a_{23} + a_{14}a_{45}^2a_{24}}{\sqrt{\det(K_{134,134})\det(K_{234,234})}}. \tag{15}$$

The numerator $\det(K_{134,234})$ is the weighted sum of all open paths between E and D. Similarly, nodes 3 and 4 are colliders for multiple paths. The bias when conditioning on these is

$$\mathrm{corr}(1,2\,|\,34) = \mathrm{corr}(1,2\,|\,345) = \frac{a_{13}a_{23} + a_{14}a_{24}}{\sqrt{(a_{13}^2 + a_{14}^2 + 1)(a_{23}^2 + a_{24}^2 + 1)}}. \tag{16}$$

Problem 6.2 is about comparing the tube volume for (15) with the tube volume for (16). □

A question of practical interest in causal inference is to understand the situations in which stratifying on a collider leads to a particularly large bias. It is widely believed that collider-stratification bias tends to attenuate when it arises from more extended paths [2,7]. What follows is our interpretation of this statement as a precise mathematical conjecture.

Problem 6.2 Let $D, E \in V$ and $\mathcal{C} = \{C \in V \mid \exists$ path P from E to D with C as a collider$\}$. We introduce a partial order on the *collider set* $\mathcal{C}$ by setting $C \leq C'$ if all paths on which C is a collider also go through C'. Given subsets $S, S' \subset \mathcal{C}$, we set $S \leq S'$ if for all $C \in S$ there exists $C' \in S'$ such that $C \leq C'$. If this holds, then the bias introduced when conditioning on S should be smaller than when conditioning on S'. To make this precise, we conjecture:

$$V_{D,E|S}(\lambda) \geq V_{D,E|S'}(\lambda) \qquad \text{for all } S \leq S' \text{ and all } \lambda \in [0,1]. \tag{17}$$

We now study this conjecture for the tripartite graphs $\mathrm{Tripart}_{p,p'}$. For tripartite graphs the conjecture says that says that the collider-stratification bias introduced

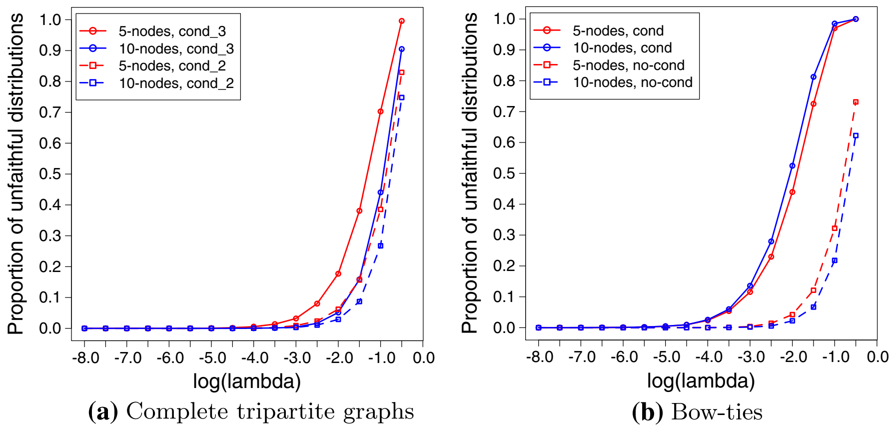


Fig. 8 Effect of collider bias on complete tripartite graphs and bow ties

when conditioning on the third level $\{p - p' + 1, \dots, p\}$ is in general smaller than when conditioning on the second level of nodes $\{3, \dots, p - p'\}$, i.e.,

$$V_{1,2|p-p'+1,\dots,p}(\lambda) \;\ge\; V_{1,2|3,\dots,p-p'}(\lambda). \tag{18}$$

This inequality is confirmed by the simulations shown in Fig. 8a. Here $p = 5$, $p' = 2$ is shown in red and $p = 10$, $p' = 2$ is shown in blue. The solid lines correspond to the volume $V_{1,2|p-p'+1,\dots,p}(\lambda)$, whereas the dashed lines correspond to the volume $V_{1,2|3,\dots,p-p'}(\lambda)$.

Going beyond simulations, we now present an algebraic proof of our conjecture when λ is small for the tripartite graphs in Fig. 7c, where the second level has only one node.

Example 6.3 For $G = \text{Tripart}_{p,p-3}$ the left-hand side of (18) is given by

$$\det(K_{1,2|4,5,\dots,p}) = a_{13}a_{23}\left(\sum_{k=4}^{p} a_{3k}^2\right).$$

Depending on the values of p, the corresponding RLCT is given by

$$\mathrm{RLCT}(1, 2|4, \dots, p) = \begin{cases} (\frac{1}{2}, 1) & \text{if } p = 4, \\ (1, 3) & \text{if } p = 5, \\ (1, 2) & \text{if } p \ge 6. \end{cases} \tag{19}$$

For $p = 4$ this was Example 4.7. To prove (19) for $p \ge 5$, we need two ingredients. Firstly, if the polynomial is a product of factors with disjoint variables, then the RLCT is the minimum of the RLCT of the factors, taken with multiplicity [e.g., if the RLCTs are (ℓ, m_1) and (ℓ, m_2), then the combined RLCT is $(\ell, m_1 + m_2)$, just like in the case of a monomial]. Secondly, the RLCT of a sum of squares of d unknowns is equal to $(d/2, 1)$. We saw this in Example 3.3.

For the right-hand side of (18), we condition on node 3. Now, the defining polynomial is

$$\det(K_{1,2|3}) = a_{13}a_{23}.$$

By Proposition 3.5, this has RLCT (1, 2), which is larger than or equal to all values of (ℓ, m) in (19). To compare the behavior of $V_{1,2|3}(\lambda)$ and $V_{1,2|4,\dots,p}(\lambda)$ for small λ, we will need to derive the constant C in (7). In Example 8.8, we will show that if $p \geq 6$ and the parameter space is

$$\Omega = \{a \in \mathbb{R}^{p-1} : |a_{12}| \leq 1, |a_{23}| \leq 1, a_{34}^2 + \cdots + a_{3p}^2 \leq 1\}, \tag{20}$$

then the asymptotic constants are given by $C_{1,2|3} = 1$ and $C_{1,2|4,\dots,p} = 2+2/(p-5)$. We conclude that $V_{1,2|3}(\lambda) \leq V_{1,2|4,\dots,p}(\lambda)$ for small values of λ, as conjectured in Problem 6.2. □

Example 6.4 A slight twist to Example 6.3 is the almost-chain graph shown in Fig. 7d, with edges $E = \{(1,3), (2,3), (3,4), \dots, (p-1, p)\}$. For such graphs, Problem 6.2 asks whether

$$V_{1,2|s}(\lambda) \leq V_{1,2|t}(\lambda) \qquad \text{if } s \leq t.$$

This holds for small λ because $\det(K_{1,2|s}) = a_{13}a_{23}\prod_{k=3}^{s-1} a_{k,k+1}^2$. By Proposition 3.5,

$$\mathrm{RLCT}(1,2|s) = \begin{cases} (1,2) & \text{if } s = 3, \\ (\frac{1}{2}, s-3) & \text{if } s \geq 4, \end{cases}$$

so $\mathrm{RLCT}(1,2|s) \geq \mathrm{RLCT}(1,2|t)$ for $s \leq t$. □

In Example 6.3 we resolved Problem 6.2 for tripartite graphs whose middle level consists of one node. We next consider the case $\mathrm{Tripart}_{p,1}$, where the third level has one node.

Example 6.5 The graph $\mathrm{Tripart}_{5,1}$ shown in Fig. 7a was discussed in Example 6.1. We focus on the numerators in (15) and in (16). The polynomial (15) will be studied in Example 7.5, where we prove that $\mathrm{RLCT}(1, 2\,|\,5) = (1, 3)$. Using the same method for $\mathrm{Tripart}_{p,1}$ gives

$$\begin{aligned} \mathrm{RLCT}(1,2|p) &= \begin{cases} (\frac{1}{2}, 1) & \text{if } p = 4, \\ (1,3) & \text{if } p = 5, \\ (1,2) & \text{if } p \geq 6, \end{cases} \quad \mathrm{RLCT}(1,2|3,\dots,p-1) \\ &= \begin{cases} (1,2) & \text{if } p = 4, \\ (1,1) & \text{if } p \geq 5. \end{cases} \end{aligned}$$

Thus, we conclude that $V_{1,2|p}(\lambda) \geq V_{1,2|3,\dots,p-1}(\lambda)$ for small values of $\lambda > 0$. □

Example 6.6 For the graph $\mathrm{Tripart}_{6,2}$ in Fig. 7b we check whether $V_{1,2|56}(\lambda) \geq V_{1,2|34}(\lambda)$ for small λ. As before, $\mathrm{RLCT}(1,2\,|\,3,4) = (1,1)$, but a hard computation using the tools of Sect. 7 reveals that now $\mathrm{RLCT}(1,2\,|\,5,6) = (1,1)$. Thus, knowledge of the RLCT is not sufficient to establish (17). What is needed is a finer analysis along the lines of Sect. 8. □

The second form of bias studied by Greenland [7] is confounder bias. In the context of a directed graphical model G, a *confounder* for the effect of E on D is a node C such that

$$E \leftarrow V_1 \leftarrow \cdots \leftarrow V_s \leftarrow C \rightarrow W_1 \rightarrow \cdots \rightarrow W_t \rightarrow D.$$

The partial correlation introduced by the path from E to D passing through C is referred to as *confounder bias*. In such situations, stratifying on C blocks the path between E and D (i.e., C d-separates E from D) and therefore corresponds to bias removal.

In certain graphs, such as the bow-tie example in [7], there are variables where stratifying removes confounder bias but at the same time introduces collider-stratification bias. For instance, consider the graph $G = \mathrm{Bow}_5$, where node 4 corresponds to exposure E and node 5 corresponds to disease outcome D. Then conditioning on node 3 blocks the paths

$$4 \leftarrow 3 \rightarrow 5, \quad 4 \leftarrow 1 \rightarrow 3 \rightarrow 5, \quad 4 \leftarrow 3 \leftarrow 2 \rightarrow 5, \tag{21}$$

and therefore reduces confounder bias, but opens the path

$$4 \leftarrow 1 \rightarrow 3 \leftarrow 2 \rightarrow 5, \tag{22}$$

and therefore introduces collider-stratification bias. This tradeoff is of particular interest in situations where one cannot condition on 1 and 2, for example because these variables were unmeasured. It is believed that in such examples the bias removed by conditioning on the confounders is larger than the collider-stratification bias introduced, and one should therefore stratify. We translate this statement into the following mathematical problem.

Problem 6.7 Let $D, E \in V$, and we denote by $\mathcal{D}$ the confounder-collider subset, i.e.,

$$\mathcal{D} = \mathcal{C} \cap \{C \in V \mid \exists \text{ path } \pi \text{ from } E \text{ to } D \text{ having } C \text{ as a confounder}\}.$$

We conjecture the following inequality for the relevant tube volumes:

$$V_{D,E|S}(\lambda) \geq V_{D,E|\emptyset}(\lambda), \qquad \text{for all } S \subset \mathcal{D} \text{ and all } \lambda \in [0,1].$$

This conjectural inequality is interesting for the bow-tie graphs Bow_p. It means that conditioning on the nodes in the second level reduces bias since the bias removed by conditioning on the confounders is larger than the collider-stratification bias introduced by conditioning:

$$V_{p-1,p|3,\ldots,p-2}(\lambda) \;\geq\; V_{p-1,p|\emptyset}(\lambda). \tag{23}$$

This is confirmed by our simulations in Fig. 8b for $p = 5$ in red and $p = 10$ in blue. The solid line corresponds to the volume $V_{p-1,p|3,\ldots,p-2}(\lambda)$ and the dashed line corresponds to $V_{p,p-1|\emptyset}(\lambda)$. In the following example we prove inequality (23) for $p = 5$ and small $\lambda > 0$.

Example 6.8 Let $G = \mathrm{Bow}_5$ as in Fig. 1d. The left-hand side of (23) is represented by

$$\det(K_{4,5|3}) = a_{13}a_{14}a_{23}a_{25}.$$

This monomial is the path in (22). The corresponding RLCT is $(1, 4)$. The polynomial representing the right-hand side of(23) is a weighted sum of the paths in (21):

$$\det(K_{4,5|\emptyset}) = a_{34}a_{35}(1 + a_{13}^2 + a_{23}^2) + a_{23}a_{25}a_{34} + a_{13}a_{14}a_{35}.$$

We derive its RLCT using the blowups described in Sect. 7. We find that it is $(1, 1)$. Since $(1, 4) < (1, 1)$, we conclude $V_{4,5|3}(\lambda) \geq V_{4,5|\emptyset}(\lambda)$. □

7 Normal Crossing and Blowing Up

In this section we develop more refined techniques for computing RLCTs. The following theorem combines the monomial case of Proposition 3.5 with the smooth case of Proposition 3.6. As promised in Sect. 3, this furnishes the proofs for these two propositions.

Theorem 7.1 *Suppose $\varphi(\omega) = \omega_1^{\tau_1}, \ldots, \omega_d^{\tau_d}$ and $f(\omega) = \omega_1^{\kappa_1}, \ldots, \omega_r^{\kappa_r} g(\omega)$ where $\tau_1, \ldots, \tau_d$ are nonnegative integers, $\kappa_1, \ldots, \kappa_r$ are positive integers, and the hypersurface $g(\omega) = 0$ is either empty or smooth and normal crossing (see definition in what follows) with $\omega_1, \ldots, \omega_r$. We write $\omega_0 = g$, $\kappa_0 = 1$, and $\tau_0 = 0$, and we define*

$$\ell = \min_{i \in \mathcal{I}} \frac{\tau_i + 1}{\kappa_i}, \quad \mathcal{J} = \left\{ \operatorname*{argmin}_{i \in \mathcal{I}} \frac{\tau_i + 1}{\kappa_i} \right\}, \quad m = |\mathcal{J}|,$$

where $\mathcal{I}$ is the set of all indices $0 \leq i \leq r$ such that ω_i has a zero in Ω. Then we have

$$\mathrm{RLCT}_\Omega(f; \varphi) = (\ell, m),$$

provided the equations $\omega_i = 0$ for $i \in \mathcal{J}$ have a solution in the interior of Ω.

The normal crossing hypothesis in Theorem 7.1 means that the system

$$f = \omega_1 \frac{\partial f}{\partial \omega_1} = \cdots = \omega_r \frac{\partial f}{\partial \omega_r} = \frac{\partial f}{\partial \omega_{r+1}} = \cdots = \frac{\partial f}{\partial \omega_d} = 0$$

has no solutions in Ω. See [12] to learn more about normal crossing singularities.

We begin with a technical lemma establishing that the RLCT can be computed locally.

Lemma 7.2 *For every $x \in \Omega$ there exists a neighborhood $\Omega_x \subset \Omega$ of x such that*

$$\mathrm{RLCT}_{\Omega_x}(f;\varphi) = \mathrm{RLCT}_U(f;\varphi)$$

for all neighborhoods $U \subset \Omega_x$ of x. Moreover,

$$\mathrm{RLCT}_{\Omega}(f;\varphi) = \min_x \mathrm{RLCT}_{\Omega_x}(f;\varphi),$$

where we take the minimum over all x in the real analytic hypersurface $\{\omega \in \Omega : f(\omega) = 0\}$.

Proof This comes from [14, Lemma 3.8, Proposition 3.9]. □

Proof of Theorem 7.1 Lemma 7.2 states that $\mathrm{RLCT}_{\Omega}(f;\varphi)$ is the minimum of $\mathrm{RLCT}_{\Omega_x}(f;\varphi)$ as x varies over Ω. Writing each subset Ω_x as $R_x \cap \Omega$, where R_x is a sufficiently small neighborbood of x in $\mathbb{R}^d$, we claim that $\mathrm{RLCT}_{\Omega}(f;\varphi) = \min_{x\in\Omega} \mathrm{RLCT}_{R_x}(f;\varphi)$ if this minimum is attained in the interior of Ω. Indeed, for x in the interior of Ω, we get $\mathrm{RLCT}_{R_x}(f;\varphi) = \mathrm{RLCT}_{\Omega_x}(f;\varphi)$. Otherwise, the volume of $\{\omega \in \Omega_x : f(\omega) \leq \lambda\}$ is less than that of $\{\omega \in R_x : f(\omega) \leq \lambda\}$ for all λ. Hence, $\mathrm{RLCT}_{R_x}(f;\varphi) \leq \mathrm{RLCT}_{\Omega_x}(f;\varphi)$, and the claim follows.

Now, to prove Theorem 7.1, it suffices to show that for each $x \in \Omega$ we have

$$\mathrm{RLCT}_{R_x}(f;\varphi) = \left(\min_{i\in\mathcal{I}_x} \frac{\tau_i + 1}{\kappa_i}, \left| \left\{ \operatorname{argmin}_{i\in\mathcal{I}_x} \frac{\tau_i + 1}{\kappa_i} \right\} \right| \right)$$

where $\mathcal{I}_x$ is the set of all indices $0 \leq i \leq r$ that satisfy $\omega_i(x) = 0$. Without loss of generality, suppose $x = (x_1, \ldots, x_d)$, where $x_1 = \cdots = x_s = 0$ and $x_{s+1}, \ldots, x_r$ are nonzero. If $g(x) \neq 0$, we may divide $f(\omega)$ by $g(\omega)$ without changing the RLCT in a sufficiently small neighborhood R_x of x. The RLCT of the remaining monomial is determined by [14, Proposition 3.7]. Now, let us suppose $g(x) = 0$. Because $g(\omega)$ is normal crossing with $\omega_1, \ldots, \omega_r$, one of the derivatives $\partial g/\partial\omega_j$ must be nonzero at x for some $s + 1 \leq j \leq d$. We assume R_x is sufficiently small, so that this derivative and $\omega_{s+1}, \ldots, \omega_r$ do not vanish. Consider the map $\sigma : R_x \to \mathbb{R}^d$ given by

$$\sigma_j(\omega) = g(\omega), \quad \sigma_i(\omega) = \omega_i \quad \text{for } i \neq j.$$

The Jacobian matrix of σ is nonsingular, so this map is an isomorphism onto its image. Set $U = \mu(R_x)$ and $\rho = \sigma^{-1} : U \to R_x$. Then for all $\mu \in U$ we have

$$(f \circ \rho)(\mu) = \mu_1^{\kappa_1}, \ldots, \mu_s^{\kappa_s} \mu_j \cdot a(\mu) \quad \text{and} \quad (\varphi \circ \rho)(\mu) = \mu_1^{\tau_1}, \ldots, \mu_s^{\tau_s} \cdot b(\mu),$$

where the factors $a(\mu)$ and $b(\mu)$ do not vanish in U. Using the chain rule [14, Proposition 4.6], we get $\mathrm{RLCT}_{R_x}(f;\varphi) = \mathrm{RLCT}_U(f \circ \rho; \varphi \circ \rho)$. The latter RLCT can be computed once again by dividing out the nonvanishing factors and applying [14, Proposition 3.7]. □

The hypersurface $\{f(\omega) = 0\}$ may not satisfy the hypothesis in Theorem 7.1. In that case, we can try to simplify its singularities via a change of variables $\rho : U \to \Omega$. With some luck, the transformed hypersurface $\{(f \circ \rho)(\mu) = 0\}$ will be described locally by monomials and the RLCT can be computed using Theorem 7.1. More precisely, let U be a d-dimensional real analytic manifold and $\rho : U \to \Omega$ a real analytic map that is *proper*, i.e., the preimage of any compact set is compact. Then ρ *desingularizes* $f(\omega)$ if it satisfies the following conditions:

1. The map ρ is an isomorphism outside the variety $\{\omega \in \Omega : f(\omega) = 0\}$.
2. Given any $y \in U$, there exists a local chart with coordinates $\mu_1, \dots, \mu_d$ such that

$$(f \circ \rho)(\mu) = \mu_1^{\kappa_1}, \dots, \mu_d^{\kappa_d} \cdot a(\mu), \quad \det \partial\rho(\mu) = \mu_1^{\tau_1}, \dots, \mu_d^{\tau_d} \cdot b(\mu),$$

where $\det \partial\rho$ is the Jacobian determinant, the exponents κ_i, τ_i are nonnegative integers, and the real analytic functions $a(\mu), b(\mu)$ do not vanish at y.

If such a desingularization exists, then we may apply ρ to the volume function (7) to calculate the RLCT. Care must be taken to multiply the measure φ by the Jacobian determinant $|\det \partial\rho|$ in accordance with the change-of-variables formula for integrals.

Hironaka's celebrated theorem on the resolution of singularities [9,10] guarantees that such a desingularization exists for all real analytic functions $f(\omega)$. The proof employs transformations known as *blowups* to simplify the singularities. We now describe the blowup $\rho : U \to \mathbb{R}^d$ of the origin in $\mathbb{R}^d$. The manifold U can be covered by local charts $U_1, \dots, U_d$ such that each chart is isomorphic to $\mathbb{R}^d$ and each restriction $\rho_i : U_i \to \mathbb{R}^d$ is the monomial map

$$(\mu_1, \dots, \mu_{i-1}, \xi, \mu_{i+1}, \dots, \mu_d) \mapsto (\xi\mu_1, \dots, \xi\mu_{i-1}, \xi, \xi\mu_{i+1}, \dots, \xi\mu_d).$$

Here, the coordinate hypersurface $\xi = 0$, also called the *exceptional divisor*, runs through all the charts. If the origin is locally the intersection of many smooth hypersurfaces with distinct tangent hyperplanes, then these hypersurfaces can be separated by blowing up the origin [9].

Example 7.3 Consider the curve $\{f(x, y) = xy(x + y)(x - y) = 0\}$ in Fig. 3d. To resolve this singularity, we blow up the origin. In the first chart, the map is $\rho_1 : (\xi, y_1) \mapsto (\xi, \xi y_1)$, so

$$f \circ \rho_1 = \xi^4 y_1(1 + y_1)(1 - y_1) \quad \text{and} \quad \det \partial\rho_1 = \xi.$$

The lines $\{y = 0\}$, $\{x + y = 0\}$ and $\{x - y = 0\}$ are transformed into $\{y_1 = 0\}$, $\{y_1 = -1\}$ and $\{y_1 = 1\}$ respectively in this chart, thereby separating them. The line $\{x = 0\}$ does not show up here, but it appears as $\{x_1 = 0\}$ in the second chart, where $\rho_2 : (x_1, \xi) \mapsto (\xi x_1, \xi)$ and

$$f \circ \rho_2 = \xi^4 x_1(x_1 + 1)(x_1 - 1), \quad \det \partial\rho_2 = \xi.$$

Since the curve $\{(1+y_1)(1-y_1)=0\}$ is normal crossing with ξy_1 in the first chart, we can now apply Theorem 7.1. The chain rule [14, Proposition 4.6] shows that $\mathrm{RLCT}_\Omega(f;1)$ is the minimum of $\mathrm{RLCT}_{U_i}(f\circ\rho_i;\det\partial\rho_i)$ for $i=1,2$. In both charts, this RLCT equals $(\frac{1}{2},1)$. □

Example 7.4 Let $p=4$ and G be an almost complete DAG with $a_{13}=0$. We consider the conditional independence statement $1 \perp\!\!\!\perp 3\,|\,4$. The correlation hypersurface is defined by

$$f=\det(K_{12,23})=a_{14}a_{23}^2a_{34}+a_{14}a_{23}a_{24}+a_{12}a_{24}a_{34}-a_{12}a_{23}+a_{14}a_{34}.$$

The real singular locus is a line in the parameter space $\mathbb{R}^5$ since $\mathrm{Singu}_{1,3,2}=\langle a_{12},a_{14},a_{23},a_{34}\rangle$. Blowing up this line in $\mathbb{R}^5$ creates four charts: U_1,U_2,U_3,U_4. For instance, the first chart has

$$\begin{aligned}\rho_1\ :\ U_1\to\mathbb{R}^5,\ \ &(\xi,\mu_{14},\mu_{23},a_{24},\mu_{34})\mapsto(\xi,\xi\mu_{14},\xi\mu_{23},a_{24},\xi\mu_{34}),\\ \det\partial\rho_1&=\xi^3.\end{aligned}$$

Then f transforms into $f\circ\rho_1=\xi^2\cdot g$, where $g=\mu_{14}\mu_{23}^2\mu_{34}\xi^2+\mu_{14}\mu_{34}+\mu_{14}\mu_{23}a_{24}+\mu_{34}a_{24}-\mu_{23}$. The hypersurface $\{g=0\}$ has no real singularities, so it is smooth in U_1. We can thus apply Theorem 7.1 with $\mathcal{I}=\{0,1\}$ to find $\mathrm{RLCT}_{U_1}(\xi^2\cdot g,\xi^3)=(1,1)$. The behavior is the same on U_2, U_3, and U_4, and we conclude that $\mathrm{RLCT}(1,3|4)=(\ell,m)=(1,1)$. This example is one of the 12 cases that were labeled *Blowup* in Table 2. The other 11 cases are similar. □

Example 7.5 In Example 6.5 we claimed that $\mathrm{RCLT}(1,2\,|\,5)=(1,3)$ for $G=\mathrm{Tripart}_{5,1}$. We now prove this claim by using the blowup method. The polynomial in question is

$$f=\det(K_{1,2|5})=(a_{13}a_{35}+a_{14}a_{45})(a_{23}a_{35}+a_{24}a_{45}).$$

The singular locus of the hypersurface $\{f=0\}$ is given by

$$\mathrm{Singu}_{1,2,34}=\langle\, a_{35}\,,\ a_{45}\,\rangle\ \cap\ \left\langle 2\times 2 \text{ minors of } \begin{pmatrix} a_{13} & a_{23} & a_{45}\\ a_{14} & a_{24} & -a_{35}\end{pmatrix}\right\rangle.$$

We blow up the linear subspace $\{a_{35}=a_{45}=0\}$ in $\mathbb{R}^6$. This creates two charts. The map for the first chart is $\rho_1:(a_{13},a_{14},a_{23},a_{24},\xi,\mu_{45})\mapsto(a_{13},a_{14},a_{23},a_{24},\xi,\xi\mu_{45})$. This map gives

$$f\circ\rho_1=\xi^2(a_{13}+a_{14}\mu_{45})(a_{23}+a_{24}\mu_{45}),\quad \det\partial\rho_1=\xi.$$

Now, by setting $a_{13}=x-a_{14}\mu_{45}$ and $a_{23}=y-a_{24}\mu_{45}$, the transformed function $f\circ\rho_1$ is the monomial ξ^2xy. Then Theorem 7.1 can be employed to evaluate $\mathrm{RLCT}_{U_1}(\xi^2xy,\xi)=(1,3)$. The calculation in the second chart is completely analogous. □

The same approach as in Example 7.5 can be applied to the polynomial $f = \det(K_{1,2|56})$ in Example 2.1. A lengthy calculation, involving many charts and multiple blowups, eventually reveals that $G = \text{Tripart}_{6,2}$ satisfies $\text{RCLT}(1, 2|56) = (1, 1)$. This was stated in Example 6.6.

8 Computing the Constants

We now describe a method for finding the constant C in the formula $V(\lambda) \approx C\lambda^{-\ell}(-\ln\lambda)^{m-1}$ in (7). The two theorems in this section are new, and they extend the results of Greenblatt [6] and Lasserre [13] on the volumes of sublevel sets. Unless stated otherwise, all measures used in this section are the standard Lebesgue measures. We begin by showing that the constant C is a function of the highest-order term in the Laurent expansion of the zeta function of f.

Lemma 8.1 *Given real analytic functions $f, \varphi : \Omega \to \mathbb{R}$, consider the Laurent expansion of*

$$\zeta(z) := \int_\Omega |f(\omega)|^{-z}\varphi(\omega)\mathrm{d}\omega = \frac{a_{\ell,m}}{(\ell - z)^m} + \frac{a_{\ell,m-1}}{(\ell - z)^{m-1}} + \cdots,$$

where ℓ is the smallest pole and m its multiplicity. Then, asymptotically as λ tends to zero,

$$V(\lambda) := \int_{|f(\omega)|\le\lambda} \varphi(\omega)\mathrm{d}\omega \;\approx\; \frac{a_{\ell,m}}{\ell(m-1)!}\,\lambda^\ell(-\ln\lambda)^{m-1}.$$

Proof According to the proof of [18, Theorem 7.1], the volume function $V(\lambda)$ equals $\int_0^\lambda v(s)\mathrm{d}s$, where $v(s) = \int_\Omega \delta(s - f(\omega))\,\varphi(\omega)\mathrm{d}\omega$ is the state density function and δ the delta function. Now, using the proof of [14, Theorem 3.16], we obtain

$$v(s) = \frac{a_{\ell,m}}{(m-1)!}\,s^{\ell-1}(-\ln s)^{m-1} + o\left(s^{\ell-1}(-\ln s)^{m-1}\right) \quad \text{as } s \to 0.$$

Here we used the little-o notation. Finally, using integration by parts, we find that

$$V(\lambda) = \frac{a_{\ell,m}}{\ell(m-1)!}\,\lambda^\ell(-\ln\lambda)^{m-1} + o\left(\lambda^\ell(-\ln\lambda)^{m-1}\right) \quad \text{as } \lambda \to 0. \qquad \square$$

Example 8.2 In Example 3.3, we saw that the volume of the d-dimensional ball defined by $|\omega_1^2 + \cdots + \omega_d^2| \le \lambda$ was equal to $V(\lambda) = C\lambda^{-d/2}$ for some positive constant C. Here we show how to compute that constant using asymptotic methods. By Lemma 8.1, $C = 2\alpha/(d\,2^d)$, where α is the coefficient of $(d/2 - z)^m$ in the Laurent expansion of the zeta function

$$\zeta(z) = \int_{\mathbb{R}^d} |\omega_1^2 + \ldots + \omega_d^2|^{-z}\mathrm{d}\omega.$$

Computing this Laurent coefficient from first principles is not easy. Instead, we derive α using the asymptotic theory of Laplace integrals. The connection between such integrals and volume functions was alluded to in Definition 3.2. By [14, Proposition 5.2], the Laplace integral

$$Z(N) = \int_{\mathbb{R}^d} \mathrm{e}^{-N(\omega_1^2+\cdots+\omega_d^2)} \mathrm{d}\omega$$

is asymptotically $\alpha\,\Gamma(\frac{d}{2})N^{-d/2}$ for large N. But this Laplace integral also decomposes as

$$Z(N) = \int_{\mathbb{R}} \mathrm{e}^{-N\omega_1^2} \mathrm{d}\omega_1 \cdots \int_{\mathbb{R}} \mathrm{e}^{-N\omega_d^2} \mathrm{d}\omega_d = (\sqrt{\pi} N^{-1/2})^d,$$

where each factor is the classical Gaussian integral. Solving for α leads to the formula

$$C = \frac{\pi^{d/2}}{2^d \cdot \Gamma(\frac{d}{2}) \cdot \frac{d}{2}} = \frac{\pi^{d/2}}{2^d \cdot \Gamma(\frac{d}{2}+1)}.$$

□

In Sect. 3, we saw how the RLCTs of smooth hypersurfaces and of hypersurfaces defined by monomial functions can be computed. The following two theorems and their accompanying examples demonstrate how the asymptotic constant C can also be evaluated in those instances. Here, we say that two hypersurfaces *intersect transversally* in $\mathbb{R}^d$ if the points of intersection are smooth on the hypersurfaces and if the corresponding tangent spaces at each intersection point generate the tangent space of $\mathbb{R}^d$ at that point.

Theorem 8.3 *Let $\{f = 0\}$ be a smooth hypersurface, and let $\varphi : \Omega \to \mathbb{R}$ be positive. Suppose $\partial f/\partial\omega_1$ is nonvanishing in Ω. Let W be the projection of the hypersurface $\{f = 0\} \subset \Omega$ onto the subspace $\{(\omega_2, \ldots, \omega_d) \in \mathbb{R}^{d-1}\}$, and let $\rho : \Omega \to \mathbb{R}^d$ be the map $\omega \mapsto (f(\omega), \omega_2, \ldots, \omega_d)$. If the boundary of Ω intersects transversally with the hypersurface $\{f = 0\}$, then*

$$V(\lambda) := \int_{\{\omega\in\Omega:|f(\omega)|\leq\lambda\}} \varphi(\omega)\mathrm{d}\omega \;\approx\; C\lambda$$

asymptotically (as $\lambda \to 0$), where

$$C = 2\int_W \frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(0, \omega_2, \ldots, \omega_d)\, \mathrm{d}\omega_2, \ldots, \mathrm{d}\omega_d.$$

Proof The asymptotics of the volume $V(\lambda)$ depends only on the region $\{\omega \in \Omega : |f(\omega)| \le \lambda\}$. So we may assume that Ω is a small neighborhood of the hypersurface $\{f(\omega) = 0\}$. As we saw in the proof of Theorem 7.1, the map ρ is an isomorphism onto its image. Thus, after changing variables, the zeta function associated to $V(\lambda)$ becomes

$$\zeta(z) = \int_{\rho(\Omega)} |f|^{-z} \frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(f, \omega_2, \dots, \omega_d)\, \mathrm{d}f \mathrm{d}\omega_2, \dots, \mathrm{d}\omega_d$$

$$= \int_W \int_{\varepsilon_1(\omega_2,\dots,\omega_d)}^{\varepsilon_2(\omega_2,\dots,\omega_d)} |f|^{-z} \frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(f, \omega_2, \dots, \omega_d)\, \mathrm{d}f \mathrm{d}\omega_2, \dots, \mathrm{d}\omega_d.$$

Here, the lower and upper limits $\varepsilon_1, \varepsilon_2$ straddle zero because the boundary of Ω is transversal to the hypersurface. By substituting the Taylor series

$$\frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(f, \omega_2, \dots, \omega_d) = \frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(0, \omega_2, \dots, \omega_d) + O(f)$$

and the exponential series $\varepsilon_2^{1-z} = 1 + O(1-z)$, we get the Laurent expansion

$$\int_0^{\varepsilon_2} |f|^{-z} \frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(f, \omega_2, \dots, \omega_d)\, \mathrm{d}f$$

$$= \left[\frac{|f|^{1-z}}{1-z} \cdot \frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(0, \omega_2, \dots, \omega_d) \right]_0^{\varepsilon_2} + \cdots$$

$$= \frac{1}{1-z} \cdot \frac{\varphi}{|\partial_{\omega_1} f|} \circ \rho^{-1}(0, \omega_2, \dots, \omega_d) + \cdots .$$

The same is true for the integral from ε_1 to 0. The result now follows from Lemma 8.1. □

Example 8.4 By Theorem 4.1, all conditional independence statements in small complete graphs lead to smooth hypersurfaces. Here we analyze the statement $1 \perp\!\!\!\perp 2 \,|\, 3$ in the complete three-node DAG. This example was studied in [17, §2]. The corresponding partial correlation is

$$\mathrm{corr}(1, 2\,|\,3) = \frac{a_{13}a_{23} - a_{12}}{\sqrt{1 + a_{23}^2}\sqrt{1 + a_{12}^2 + a_{13}^2}}.$$

This partial correlation hypersurface lives in $\mathbb{R}^3$ and is depicted in [17, Fig. 2b].

We apply Theorem 8.3 by setting $\Omega := [-1, 1]^3$, $f := \mathrm{corr}(1, 2\,|\,3)$, and $\varphi := 1/2^3$, the uniform distribution on Ω. We choose ω_1 to be a_{12}. Then $\rho^{-1}(0, a_{13}, a_{23}) = (a_{13}a_{23}, a_{13}, a_{23})$. The projection W of the surface $\{a_{12} = a_{13}a_{23}\}$ onto $\{(a_{13}, a_{23}) \in$

$[-1, 1]^2\}$ is the whole square $[-1, 1]^2$. The formula for the constant C now simplifies to

$$C = \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} \sqrt{1+a_{13}^2}\sqrt{1+a_{13}^2+a_{13}^2a_{23}^2}\, da_{13}\, da_{23} \;\approx\; 5.4829790759.$$

This two-dimensional integral was evaluated numerically using `Mathematica`. □

We now come to the monomial case that was discussed in Theorem 7.1.

Theorem 8.5 *Let $g : \Omega \to \mathbb{R}$ and $\varphi : \Omega \to \mathbb{R}$ be positive, and let $f(\omega) = \omega_1^{\kappa_1} \cdots \omega_d^{\kappa_d}\, g(\omega)$, where the κ_i are nonnegative integers. Suppose that $1/\ell = \kappa_1 = \cdots = \kappa_m > \kappa_{m+1} \geq \cdots \geq \kappa_d$ and that the boundary of Ω is transversal to the subspace L defined by $\omega_1 = \cdots = \omega_m = 0$. Let $\bar{\omega}$ and $\bar{\kappa}$ denote the vectors $(\omega_{m+1}, \ldots, \omega_d)$ and $(\kappa_{m+1}, \ldots, \kappa_d)$, respectively. Then*

$$V(\lambda) := \int_{\{\omega\in\Omega:|f(\omega)|\leq\lambda\}} \varphi(\omega)\mathrm{d}\omega \;\approx\; C\lambda^{\ell}(-\ln\lambda)^{m-1}$$

asymptotically as λ tends to zero, where

$$C \;=\; \frac{(2\ell)^m}{\ell(m-1)!}\int_{\Omega\cap L} \bar{\omega}^{-\ell\bar{\kappa}} g(0, \ldots, 0, \bar{\omega})^{-\ell}\varphi(0, \ldots, 0, \bar{\omega})\mathrm{d}\bar{\omega}. \tag{24}$$

Proof Let us suppose for now that Ω is the hypercube $[0, \varepsilon]^d$. Our goal is to apply Lemma 8.1 by computing the Laurent coefficient $a_{\ell,m}$ of the zeta function $\zeta(z)$. We first study the Taylor series expansion of the integrand about $\omega_1 = \cdots = \omega_m = 0$. This gives

$$\begin{aligned}\left(\omega^{\kappa} g(\omega)\right)^{-z}\varphi(\omega) = \omega^{-z\kappa}\big(&g(0, \ldots, 0, \bar{\omega})^{-z}\varphi(0, \ldots, 0, \bar{\omega})\\ &+O(\omega_1) + \cdots + O(\omega_m)\big).\end{aligned}$$

The higher-order terms in this expansion contribute larger poles to $\zeta(z)$, so we only need to compute the coefficient of $(\ell - z)^{-m}$ in the Laurent expansion of

$$\begin{aligned}&\int_{\Omega} \omega_1^{-z\kappa_1} \cdots \omega_m^{-z\kappa_m}\, \bar{\omega}^{-z\bar{\kappa}}\, g(0, \ldots, 0, \bar{\omega})^{-z}\, \varphi(0, \ldots, 0, \bar{\omega})\; \mathrm{d}\omega_1, \ldots, \mathrm{d}\omega_m\, \mathrm{d}\bar{\omega}\\ &= \left(\frac{\varepsilon^{1-z/\ell}}{1-z/\ell}\right)^m \int_{\Omega\cap L} \bar{\omega}^{-z\bar{\kappa}}\, g(0, \ldots, 0, \bar{\omega})^{-z}\, \varphi(0, \ldots, 0, \bar{\omega})\; \mathrm{d}\bar{\omega}.\end{aligned}$$

Because g is positive, the last integral in the preceding expression has no poles near $z = \ell$, so the constant term in its Laurent expansion comes from substituting $z = \ell$. Hence,

$$a_{\ell,m} = \ell^m \int_{\Omega\cap L} \bar{\omega}^{-\ell\bar{\kappa}}\, g(0,\ldots,0,\bar{\omega})^{-\ell}\, \varphi(0,\ldots,0,\bar{\omega})\, \mathrm{d}\bar{\omega}.$$

Now suppose Ω is not the hypercub $[0,\varepsilon]^d$. Since the boundary of Ω is transversal to the subspace L, we decompose Ω into small neighborhoods that are isomorphic to orthants. Summing up the contributions from these orthants gives the desired result. □

Remark 8.6 We revisit the planar tubes shown in Fig. 3a–c. Using the formula (24) in Theorem 8.5, one can easily check the constants C we saw in Example 3.1, namely,

$$C = \begin{cases} 1 & \text{for } f(x,y) = x, \\ 1 & \text{for } f(x,y) = xy, \\ 3 & \text{for } f(x,y) = x^2y^3. \end{cases}$$

Example 8.7 We apply Theorem 8.5 to find the constants in Corollary 5.3 for chains and stars. In both cases we set $\Omega = [-1,1]^{p-1}$ and $\varphi = 2^{1-p}$. For chains we have $(\ell,m) = (1,p-1)$, and L is the subspace $a_{12} = \cdots = a_{p-1,p} = 0$. Then the integral in (24) is the evaluation of the denominator of (13) at the origin multiplied by φ, so $C_{\text{chain}} = 1/(p-2)!$, as claimed.

For stars, $(\ell,m) = (1,2)$ is achieved by $1 < i < j$ and $S \subset \bar{S} := \{2,\ldots,p\}\backslash\{i,j\}$, with

$$\operatorname{corr}(i,j|S) = -\frac{a_{1i}a_{1j}}{\sqrt{1+SOS(S)+a_{1i}^2}\,\sqrt{1+SOS(S)+a_{1j}^2}}, \quad SOS(S) = \sum_{s\in S} a_{1s}^2.$$

Since $|\operatorname{corr}(i,j|S)| \geq |\operatorname{corr}(i,j|\bar{S})|$, the quantity $V_G(\lambda)$ is the volume of the union of the tubes $\{|\operatorname{corr}(i,j|\bar{S})| \leq \lambda\}$ over all $1<i<j$. By application of formula (24), the asymptotic volume of each tube computes to $p\lambda(-\ln\lambda)/3$. Meanwhile, the volumes of the intersections of these tubes become negligible as $\lambda \to 0$. After summing over all $1<i<j$, we get $C_{star} = \binom{p-1}{2}\frac{p}{3} = \binom{p}{3}$. □

Example 8.8 We compute the constant C of the volume $V_{1,2|4,\ldots,p}(\lambda)$ for $G =$ $\text{Tripart}_{p,p-3}$ as in Example 6.3. Let $p \geq 6$ and Ω be given by (20). We are interested in the tube

$$\left| a_{13}a_{23}\frac{g(\bar{a})}{h(a)} \right| \leq \lambda, \quad \text{where } g(\bar{a}) = \sum_{k=4}^{p} a_{3k}^2$$
$$\text{and } \; h(a) = \sqrt{1+g(\bar{a})(a_{13}^2+1)}\cdot\sqrt{1+g(\bar{a})(a_{23}^2+1)}.$$

The measure on Ω is $\varphi(a)\, da_{12}\, da_{23}\, d\bar{a}$, where $\varphi(a) = 1/4$ and $d\bar{a}$ is the Lebesgue probability measure on the ball $\{g(\bar{a}) \leq 1\}$. According to Theorem 8.5,

$$C = \int_{\{g(\bar{a})\leq 1\}} \left(\frac{g(\bar{a})}{1+g(\bar{a})}\right)^{-1} \mathrm{d}\bar{a}.$$

By substituting spherical coordinates for the integration, this expression simplifies to

$$1 + \int_{\{g(\bar{a})\leq 1\}} \frac{1}{g(\bar{a})} \mathrm{d}\bar{a} = 2 + \frac{2}{p-5},$$

yielding the constant $C_{1,2|4,\ldots,p}$ needed for the bias reduction analysis in Example 6.3. □

9 Discussion

In this paper we examined the volume of regions in the parameter space of a directed Gaussian graphical model that are given by bounding partial correlations. We established a connection to singular learning theory, and we showed that these volumes can be computed by evaluating the RLCT of the partial correlation hypersurfaces. Throughout the paper we made the simplifying assumption of equal noise, i.e., $\epsilon \sim \mathcal{N}(0, I)$. Ideally, one would like to allow for different noise variances. This would increase the dimension of the parameter space Ω. It would be very interesting to study this more difficult situation and understand how the asymptotic volumes change or, more generally, how the asymptotics depends on our choice of the parameter space Ω. This issue was discussed briefly in Example 4.8.

This paper can be seen as a first step toward developing a theory that would make it possible to compute the complete asymptotic expansion of particular volumes. We have concentrated on computing only the leading coefficients of these expansions, and even this question is still open in many cases (e.g., Example 6.6). An interesting extension would be to better understand how to use properties of the graph to compute the coefficients $C_{l,m}$ in the asymptotic expansion. Finally, another interesting problem for future research would be to ascertain all values of (l, m) for which $C_{l,m}$ is nonzero in terms of the intrinsic properties of the underlying graph.

Acknowledgments This project began at the workshop *Algebraic Statistics in the Alleghenies*, which took place at Pennsylvania State University in June 2012. We are grateful to the organizers for a very inspiring meeting. We thank Thomas Richardson for pointing out the connection to the problem of bias reduction in causal inference. We also thank all reviewers for thoughtful comments and one of the reviewers in particular for numerous detailed comments and for spotting a mistake in a proof. This work was supported in part by the US National Science Foundation (DMS-0968882) and the Defense Advanced Research Projects Agency (DARPA) Deep Learning program (FA8650-10-C-7020).

References

1. V. I. Arnol'd, S. M. Guseĭn-Zade and A. N. Varchenko: *Singularities of Differentiable Maps*, Vol. II, Birkhäuser, Boston, 1985.
2. S. Chaudhuri and T.S. Richardson: Using the structure of d-connecting paths as a qualitative measure of the strength of dependence, *Proceedings of the Nineteenth Conference on Uncertainty in, Artificial Intelligence*, pages 116–123, 2003.
3. D. Cox, J. Little and D. O'Shea: *Ideals, Varieties and Algorithms*, Springer Undergraduate Texts, Third edition, Springer Verlag, New York, 2007.
4. W. Decker, G.-M. Greuel, G. Pfister, and H. Schönemann: SINGULAR 3-1-5 - A computer algebra system for polynomial computations, 2012, http://www.singular.uni-kl.de.
5. D. Grayson and M. Stillman: MACAULAY 2, a software system for research in algebraic geometry, available at http://www.math.uiuc.edu/Macaulay2/.
6. M. Greenblatt: Oscillatory integral decay, sublevel set growth, and the Newton polyhedron, *Math. Annalen* **346**, pages 857–890, 2010.
7. S. Greenland: Quantifying biases in causal models: classical confounding vs collider-stratification bias, *Epidemiology* **14**, pages 300–306, 2003.
8. S. Greenland and J. Pearl: Adjustments and their consequences: collapsibility analysis using graphical models, *International Statistical Review* **79**, pages 401–426, 2011.
9. H. Hauser: The Hironaka theorem on resolution of singularities (or: a proof that we always wanted to understand), *Bull. Amer. Math. Soc.* **40**, pages 323–403, 2003.
10. H. Hironaka: Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II. *Ann. of Math.* (2) **79**, pages 109–326, 1964.
11. M. Kalisch and P. Bühlmann: Estimating high-dimensional directed acyclic graphs with the PC-algorithm, *Journal of Machine Learning Research* **8**, pages 613–636, 2007.
12. J. Kollár: *Lectures on Resolution of Singularities*, Annals of Mathematics Studies 166, Princeton University Press, Princeton, NJ, 2007.
13. J. B. Lasserre: Level sets and non Gaussian integrals of positively homogeneous functions, *International Game Theory Review.* http://homepages.laas.fr/lasserre/forthcoming.html.
14. S. Lin: A*lgebraic Methods for Evaluating Integrals in Bayesian Statistics*, Ph.D. dissertation, University of California, Berkeley, May 2011.
15. M. Marshall: *Positive Polynomials and Sums of Squares, Mathematical Surveys and Monographs* **146**, American Mathematical Society, Providence, RI, 2008.
16. P. Spirtes, C. Glymour and R. Scheines: *Prediction and Search*, MIT Press, second edition, 2001.
17. C. Uhler, G. Raskutti, P. Bühlmann and B. Yu: Geometry of faithfulness assumption in causal inference, *Annals of Statistics* **41**, pages 436–463, 2013.
18. S. Watanabe: *Algebraic Geometry and Statistical Learning Theory*, Monographs on Applied and Computational Mathematics **25**, Cambridge University Press, 2009.
19. J. Zhang and P. Spirtes: Strong faithfulness and uniform consistency in causal inference, *Uncertainty in Artificial Intelligence (UAI)*, pages 632–639, 2003.
20. J. Zhang and P. Spirtes: Detection of unfaithfulness and robust causal inference, *Minds and Machines* **18**, pages 239–271, 2008.