



Ideal-Theoretic Strategies for Asymptotic Approximation of Marginal Likelihood Integrals

Shaowei Lin

Engineering Systems & Design, Singapore University of Technology & Design, Singapore

Abstract. The accurate asymptotic evaluation of marginal likelihood integrals is a fundamental problem in Bayesian statistics. Following the approach introduced by Watanabe, we translate this into a problem of computational algebraic geometry, namely, to determine the real log canonical threshold of a polynomial ideal, and we present effective methods for solving this problem. Our results are based on resolution of singularities. They apply to parametric models where the Kullback-Leibler distance is upper and lower bounded by scalar multiples of some sum of squared real analytic functions. Such models include finite state discrete models.

2000 Mathematics Subject Classifications: Primary 62E99; Secondary 14E15, 62G20

Key Words and Phrases: computational algebra, asymptotic approximation, marginal likelihood, learning coefficient, real log canonical threshold

1. Introduction

The evaluation of marginal likelihood integrals is essential in model selection and has important applications in areas such as machine learning and computational biology. The exact evaluation of such integrals is a difficult problem [9, 20] and classical approximation formulas usually apply only for smooth models. Recent work by Watanabe and his collaborators [1, 27, 26, 28, 29] extended these formulas to a broad class of models with singularities. His work also uncovered interesting connections with resolution of singularities in algebraic geometry. The goal of this paper is to systematically study the algebraic geometry behind Watanabe's formulas, and to develop symbolic algebra tools which allow the user to accurately evaluate the asymptotics of integrals in Bayesian statistics.

Watanabe showed that the key to understanding a singular model is monomializing the Kullback-Leibler function $K(\omega)$ of the model at the true distribution. While general algorithms exist for monomializing any analytic function [7, 4], applying them to non-polynomial functions such as $K(\omega)$ can be computationally expensive. In practice, many singular models are parametrized by polynomials. Therefore, it is natural to ask if this polynomiality can be exploited in the analysis of such models. For simplicity, we explore

Email address: shaowei_lin@sutd.edu.sg (S. Lin)

this question for discrete statistical models. Our point of departure is to describe the asymptotics of the likelihood integral by the real log canonical threshold of an ideal in a polynomial ring. More generally, our results will be proved for rings of analytic functions, and they apply to all parametric models where the Kullback-Leibler distance is upper and lower bounded by scalar multiples of a sum of squared real analytic functions.

Consider a statistical model \mathcal{M} on a discrete space $[k] = \{1, 2, \dots, k\}$ parametrized by a real analytic map $p : \Omega \rightarrow \Delta_{k-1}$ where Ω is a compact subset of \mathbb{R}^d and Δ_{k-1} is the probability simplex $\{x \in \mathbb{R}^k : x_i \geq 0, \sum x_i = 1\}$. We assume that Ω is *semianalytic*, i.e. $\Omega = \{x \in \mathbb{R}^d : g_1(x) \geq 0, \dots, g_l(x) \geq 0\}$ is defined by real analytic inequalities. Let $q \in \Delta_{k-1}$ be a point in the model with non-zero entries. Suppose a sample of size N is drawn from the true distribution q , and let $U = (U_i)$ denote the vector of relative frequencies for this sample. Let $\varphi : \Omega \rightarrow \mathbb{R}$ be *nearly analytic*, i.e. φ is a product $\varphi_a \varphi_s$ of functions where φ_a is real analytic and φ_s is positive and smooth. Consider a Bayesian prior defined by $|\varphi|$. Priors of this form are discussed in Remark 1. We are interested in the asymptotics, for large sample sizes N , of the marginal likelihood integral

$$Z(N) = \int_{\Omega} \prod_{i=1}^k p_i(\omega)^{NU_i} |\varphi(\omega)| d\omega. \quad (1)$$

The first few terms of the asymptotics of the log likelihood integral $\log Z(N)$ was derived by Watanabe. To state his result, we first recall that the Kullback-Leibler distance $K(\omega)$ between q and $p(\omega)$ is

$$K(\omega) = \sum_{i=1}^k q_i \log \frac{q_i}{p_i(\omega)}.$$

This function satisfies $K(\omega) \geq 0$ with equality if and only if $p(\omega) = q$.

Theorem 1 (Watanabe[27, §6]). *Asymptotically as $N \rightarrow \infty$,*

$$\log Z(N) = N \sum_{i=1}^k U_i \log q_i - \lambda \log N + (\theta - 1) \log \log N + \eta_N \quad (2)$$

where the positive rational number λ is the smallest pole of the zeta function

$$\zeta(z) = \int_{\Omega} K(\omega)^{-z} |\varphi(\omega)| d\omega, \quad z \in \mathbb{C}, \quad (3)$$

θ is its multiplicity, and η_N is a random variable whose expectation tends to a constant.

Here, λ is known as the *learning coefficient* of the model at the distribution q . Because formula (2) generalizes the Bayesian information criterion [21, 27], the numbers λ and θ are important in model selection. Indeed, the BIC corresponds to the case $(\lambda, \theta) = (\frac{d}{2}, 1)$ for smooth models. In algebraic geometry, λ is also known as the *real log canonical threshold* [23] of K , a term that is motivated by the more familiar complex *log canonical threshold* (see Remark 2). We denote this algebraic invariant by $(\lambda, \theta) = \text{RLCT}_{\Omega}(K; \varphi)$.

These thresholds may be defined for ideals in rings of real-valued analytic functions as well. Given an ideal $I = \langle f_1, \dots, f_r \rangle$ generated by functions $f_i \neq 0$ which are real analytic on a compact subset $\Omega \subset \mathbb{R}^d$, and a smooth amplitude function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, we consider the zeta function

$$\zeta(z) = \int_{\Omega} \left(f_1(\omega)^2 + \dots + f_r(\omega)^2 \right)^{-z/2} |\varphi(\omega)| d\omega. \quad (4)$$

We show that if φ is nearly analytic, then $\zeta(z)$ has an analytic continuation to the whole complex plane. Its poles are positive rational numbers with a smallest element λ which we call the *real log canonical threshold* of I with respect to φ over Ω . Let θ be the multiplicity of λ as a pole of $\zeta(z)$ and define $\text{RLCT}_{\Omega}(I; \varphi)$ to be the pair (λ, θ) . Order these pairs such that $(\lambda_1, \theta_1) > (\lambda_2, \theta_2)$ if $\lambda_1 > \lambda_2$, or $\lambda_1 = \lambda_2$ and $\theta_1 < \theta_2$. We will show that this pair does not depend on the choice of generators f_1, \dots, f_r for I . In the literature, real log canonical thresholds of ideals are not well-investigated [23]. For this reason, we formally prove many of its properties in Section 3.

With these definitions on hand, we now state our first main theorem. This result expresses the learning coefficient and its multiplicity directly in terms of the functions p_1, \dots, p_k parametrizing the model. Geometrically, it says that the learning coefficient is the real log canonical threshold of the fiber $p^{-1}(q) \subset \Omega$. The theorem is computationally very useful especially when the p_i are polynomials or rational functions, and certain special cases have been applied by Sumio Watanabe and his collaborators [28, 29]. Our proof in Section 3 was inspired by a discussion with him. Recall that $\varphi = \varphi_a \varphi_s$ is nearly analytic.

Theorem 2. *Let (λ, θ) be the learning coefficient and multiplicity of the model \mathcal{M} at $q > 0$. Let I denote the ideal $\langle p(\omega) - q \rangle := \langle p_1(\omega) - q_1, \dots, p_k(\omega) - q_k \rangle$, and let \mathcal{V} be its zero-locus $\{\omega \in \Omega : p(\omega) = q\} = p^{-1}(q)$. Then,*

$$(2\lambda, \theta) = \min_{x \in \mathcal{V}} \text{RLCT}_{\Omega_x}(I; \varphi_a)$$

where each Ω_x is a sufficiently small neighborhood of x in Ω .

More generally, let $K(\omega)$ be any real analytic function on Ω that is bounded for some constants $c_1, c_2 > 0$ and some real analytic $f_i(\omega)$ over Ω , by

$$c_1 \sum_{i=1}^k f_i(\omega)^2 \leq K(\omega) \leq c_2 \sum_{i=1}^k f_i(\omega)^2.$$

Then, the real log canonical threshold $(\lambda, \theta) = \text{RLCT}_{\Omega}(K; \varphi)$ satisfies

$$(2\lambda, \theta) = \text{RLCT}_{\Omega}(I; \varphi_a)$$

where I is the ideal $\langle f_1(\omega), \dots, f_k(\omega) \rangle$.

To prove this theorem and other properties of real log canonical thresholds, we recall Hironaka's theorem on the resolution of singularities [15] and develop useful lemmas in

Section 2. Our treatment differs from that of Watanabe [27] in the following way: we study the *local* behavior of real log canonical thresholds at points x in the parameter space Ω . In particular, we will be interested in the case where x is on the boundary $\partial\Omega$. Example 1 is an illustration of how the threshold is affected by the inequalities $g_i \geq 0$ which are active at x . This issue can be critical in singular model selection because the parameter space of one model is often contained in the boundary of another that is more complex.

After studying the local thresholds, we then show that the real log canonical threshold *globally* over Ω is the minimum of local thresholds at points x in Ω . Identifying where these minimum thresholds occur is a challenging problem which we discuss in Section 2. As a consequence of our results, we write down explicit formulas for the coefficients in asymptotic expansions of Laplace integrals. Our formulas extend those of Arnol'd–Guseĭn-Zade–Varchenko[2] because they apply also to parameter spaces with boundary. Using this expansion to improve statistical inference may be explored in future.

Our next aim is to develop tools for computing or bounding real log canonical thresholds of ideals. Section 3 summarizes useful fundamental properties of real log canonical thresholds. In Section 4, we derive local thresholds in nondegenerate cases using an important tool from toric geometry involving Newton polyhedra. This method was invented by Varchenko [24] and applied to statistical models by Watanabe and Yamazaki [29]. Their formulas were defined for functions, but we develop extensions of these formulas for ideals. We introduce a new notion of nondegeneracy for ideals, known as *sos-nondegeneracy*, and give the following bound for the real log canonical threshold of an ideal with respect to a monomial amplitude function $\omega^\tau := \omega_1^{\tau_1} \cdots \omega_d^{\tau_d}$. These monomial functions occur frequently when we apply a change of variables to resolve the singularities in a model. Newton polyhedra and their τ -distances are defined in Section 4.

Theorem 3. *Let I be a finitely generated ideal in the ring of functions which are real analytic on Ω , and suppose the origin 0 lies in the interior of Ω . Then, for every sufficiently small neighborhood Ω_0 of the origin,*

$$\text{RLCT}_{\Omega_0}(I; \omega^\tau) \leq (1/l_\tau, \theta_\tau)$$

where l_τ is the τ -distance of the Newton polyhedron $\mathcal{P}(I)$ and θ_τ its multiplicity. Equality occurs when I is monomial or, more generally, *sos-nondegenerate*.

This theorem has two main consequences. Firstly, it tells us that the real log canonical threshold of an ideal can be computed by finding a change of variables which monomializes the ideal. Secondly, due to Theorems 1 and 2, upper bounds on real log canonical thresholds translate to asymptotic lower bounds on the likelihood integral of a statistical model, which in turn give upper bounds on the stochastic complexity of the model.

Currently, there are no programs for computing real log canonical thresholds. There are applications which compute resolutions of singularities, but our statistical problems are too big for them. We hope that our work is a step in bridging the gap. Some of our tools are implemented in a SINGULAR library at

<https://w3id.org/people/shaoweilin/public/rlct.html>.

This library computes the Newton polyhedron of an ideal, computes τ -distances, and checks if an ideal is sos-nondegenerate. Instructions and examples on using the library may be found at the above website.

In summary, the learning coefficient of a statistical model is a useful measure of the model complexity and plays an important role in model selection. Because computing this coefficient often requires careful analysis of the Kullback-Leibler function, we propose an ideal-theoretic approach to make this calculation more tractable. This method has several advantages. Firstly, it directly exploits poly-nomiality in the model parametrization. Second, the real log canonical threshold of an ideal is independent of the choice of generators, and this choice provides flexibility to our computations. Thirdly, it is easier to construct Newton polyhedra for polynomial ideals and to check their nondegeneracy (Proposition 3(3)), than for nonpolynomial Kullback-Leibler functions. We demonstrate these ideas in Section 5 by computing the learning coefficients of a discrete mixture model which comes from a study involving 132 schizophrenic patients.

To introduce some notation, given $x \in \mathbb{R}^d$, let $\mathcal{A}_x(\mathbb{R}^d)$ be the ring of real-valued functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ that are analytic at x . We sometimes shorten the notation to \mathcal{A}_x when it is clear that we are working with the space \mathbb{R}^d . When $x = 0$, it is convenient to think of \mathcal{A}_0 as a subring of the formal power series ring $\mathbb{R}[[\omega_1, \dots, \omega_d]] = \mathbb{R}[[\omega]]$. It consists of power series which are convergent in some neighborhood of the origin. For all x , \mathcal{A}_x is isomorphic to \mathcal{A}_0 by translation. Given a subset $\Omega \subset \mathbb{R}^d$, let \mathcal{A}_Ω be the ring of real functions analytic at each point $x \in \Omega$. Locally, each function can be represented as a power series centered at x . Given $f \in \mathcal{A}_\Omega$, define the analytic variety $\mathcal{V}_\Omega(f) = \{\omega \in \Omega : f(\omega) = 0\}$ while for an ideal $I \subset \mathcal{A}_\Omega$, we set $\mathcal{V}_\Omega(I) = \bigcap_{f \in I} \mathcal{V}_\Omega(f)$. Lastly, given a finite multiset $S \subset \mathbb{R}$, let $\#\min S$ denote the number of times the minimum is attained in S .

2. Resolution of Singularities

In this section, we introduce Hironaka's theorem on resolutions of singularities. We derive real log canonical thresholds of monomial functions, and demonstrate how such resolutions allow us to find the thresholds of non-monomial functions. We show that the threshold of a function over a compact set is the minimum of local thresholds, and present an example where the threshold at a boundary point depend on the boundary inequalities. We discuss the problem of locating singularities with the smallest threshold, and end this section with formulas for the asymptotic expansion of a Laplace integral.

Before we explore real log canonical thresholds of ideals, let us study those of functions. Given a compact subset Ω of \mathbb{R}^d , a real analytic function $f \in \mathcal{A}_\Omega$ with $f \not\equiv 0$, and a smooth function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, consider the zeta function

$$\zeta(z) = \int_{\Omega} |f(\omega)|^{-z} |\varphi(\omega)| d\omega, \quad z \in \mathbb{C}. \quad (5)$$

This function is well-defined for $z \in \mathbb{R}_{\leq 0}$. If $\zeta(z)$ can be continued analytically to the whole complex plane \mathbb{C} , then all its poles are isolated points in \mathbb{C} . Moreover, if all its poles are real, then there exists a smallest positive pole λ . Let θ be the multiplicity of this

pole. The pole λ is the *real log canonical threshold* of f with respect to φ over Ω . If $\zeta(z)$ has no poles, we set $\lambda = \infty$ and leave θ undefined. Let $\text{RLCT}_\Omega(f; \varphi)$ be the pair (λ, θ) . By abuse of notation, we sometimes refer to this pair as the real log canonical threshold of f . We order these pairs such that $(\lambda_1, \theta_1) > (\lambda_2, \theta_2)$ if $\lambda_1 > \lambda_2$, or $\lambda_1 = \lambda_2$ and $\theta_1 < \theta_2$. Intuitively, considering the asymptotics of $\log Z(N)$ in Theorem 1, the ordering is defined in this way so that $(\lambda_1, \theta_1) > (\lambda_2, \theta_2)$ if and only if

$$\lambda_1 \log N - (\theta_1 - 1) \log \log N > \lambda_2 \log N - (\theta_2 - 1) \log \log N$$

for sufficiently large N . Lastly, let $\text{RLCT}_\Omega f$ denote $\text{RLCT}_\Omega(f; 1)$ where 1 is the constant unit function.

We start with a simple class of functions for which it is easy to compute the real log canonical threshold. It is the class of monomials $\omega_1^{\kappa_1} \cdots \omega_d^{\kappa_d} = \omega^\kappa$.

Proposition 1. *Let $\kappa = (\kappa_1, \dots, \kappa_d)$ and $\tau = (\tau_1, \dots, \tau_d)$ be vectors of non-negative integers. If Ω is the positive orthant $\mathbb{R}_{\geq 0}^d$ and $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ is compactly supported and smooth with $\phi(0) > 0$, then $\text{RLCT}_\Omega(\omega^\kappa; \omega^\tau \phi) = (\lambda, \theta)$ where*

$$\lambda = \min_{1 \leq j \leq d} \left\{ \frac{\tau_j + 1}{\kappa_j} \right\}, \quad \theta = \# \min_{1 \leq j \leq d} \left\{ \frac{\tau_j + 1}{\kappa_j} \right\}.$$

Proof. See [2, Lemma 7.3]. The idea is to express $\phi(\omega)$ as $T_s(\omega) + R_s(\omega)$ where T_s is the s -th degree Taylor polynomial and R_s the difference. We then integrate the main term $|f|^{-z} T_s$ explicitly and show that the integral of the remaining term $|f|^{-z} R_s$ does not have smaller poles. This process gives the analytic continuation of $\zeta(z)$ to the whole complex plane, so we have the Laurent expansion

$$\zeta(z) = \sum_{\alpha > 0} \sum_{i=1}^d \frac{d_{i,\alpha}}{(z - \alpha)^i} + P(z) \tag{6}$$

where the poles α are positive rational numbers and $P(z)$ is a polynomial.

For non-monomial $f(\omega)$, Hironaka's celebrated theorem [15] on the *resolution of singularities* tells us that we can always reduce to the monomial case. Here, a d -dimensional real analytic manifold is a topological space (second countable and Hausdorff) that can be covered by charts which are homeomorphic to open balls in \mathbb{R}^d and where the transition maps between charts are real analytic maps.

Theorem 4 (Resolution of Singularities). *Let f be a non-constant real analytic function in some neighborhood $\Omega \subset \mathbb{R}^d$ of the origin with $f(0) = 0$. Then, there exists a triple (M, W, ρ) where*

- a. $W \subset \Omega$ is a neighborhood of the origin,
- b. M is a d -dimensional real analytic manifold,
- c. $\rho : M \rightarrow W$ is a real analytic map

satisfying the following properties.

- i. ρ is proper, i.e. the inverse image of any compact set is compact.
- ii. ρ is a real analytic isomorphism between $M \setminus \mathcal{V}_M(f \circ \rho)$ and $W \setminus \mathcal{V}_W(f)$.
- iii. For any $y \in \mathcal{V}_M(f \circ \rho)$, there exists a chart M_y with coordinates $\mu = (\mu_1, \mu_2, \dots, \mu_d)$ such that y is the origin and

$$f \circ \rho(\mu) = a(\mu)\mu_1^{\kappa_1}\mu_2^{\kappa_2} \cdots \mu_d^{\kappa_d} = a(\mu)\mu^\kappa$$

where $\kappa_1, \kappa_2, \dots, \kappa_d$ are non-negative integers and a is a real analytic function with $a(\mu) \neq 0$ for all μ . Furthermore, the Jacobian determinant equals

$$|\rho'(\mu)| = h(\mu)\mu_1^{\tau_1}\mu_2^{\tau_2} \cdots \mu_d^{\tau_d} = h(\mu)\mu^\tau$$

where $\tau_1, \tau_2, \dots, \tau_d$ are non-negative integers and h is a real analytic function with $h(\mu) \neq 0$ for all μ .

We say that (M, W, ρ) is a *resolution of singularities* or a *desingularization* of f at the origin. The set of points in M where ρ is not one-to-one is the *exceptional divisor*. From properties (i) and (ii), it also follows that ρ is surjective: if $x \in \mathcal{V}_W(f)$, we pick a compact neighborhood V of x and a sequence x_1, x_2, \dots of points in $V \setminus \mathcal{V}_W(f)$ converging to x . The sequence can be chosen off the variety because the variety has measure zero. Then, the preimages $\rho^{-1}(x_1), \rho^{-1}(x_2), \dots$ contain a converging subsequence with limit y , and $\rho(y) = x$ by continuity.

Now, let us desingularize a list of functions simultaneously.

Corollary 1 (Simultaneous Resolutions). *Let f_1, \dots, f_l be non-constant real analytic functions in some neighborhood $\Omega \subset \mathbb{R}^d$ of the origin with all $f_i(0) = 0$. Then, there exists a triple (M, W, ρ) that desingularizes each f_i at the origin.*

Proof. The idea is to desingularize the product $f_1(\omega) \cdots f_l(\omega)$ and to show that such a resolution of singularities is also a resolution for each f_i . See [27, Thm 11] and [13, Lemma 2.3] for more details.

For the rest of this section, let $\Omega = \{\omega \in \mathbb{R}^d, g_1(\omega) \geq 0, \dots, g_l(\omega) \geq 0\}$ be compact and semianalytic. Assume that $f, \varphi \in \mathcal{A}_\Omega$, and that f, g_1, \dots, g_l are not constant functions.

Lemma 1. *For each $x \in \Omega$, there is a neighborhood Ω_x of x in Ω such that for all smooth functions ϕ on Ω_x with $\phi(x) > 0$,*

$$\text{RLCT}_{\Omega_x}(f; \varphi\phi) = \text{RLCT}_{\Omega_x}(f; \varphi).$$

Proof. Let $x \in \Omega$. If $f(x) \neq 0$, then by the continuity of f , there exists a small neighborhood Ω_x where $0 < c_1 < |f(\omega)| < c_2$ for some constants c_1, c_2 . Hence, for all smooth functions ϕ , the zeta functions

$$\int_{\Omega_x} |f(\omega)|^{-z} |\varphi(\omega)\phi(\omega)| d\omega \quad \text{and} \quad \int_{\Omega_x} |f(\omega)|^{-z} |\varphi(\omega)| d\omega$$

do not have any poles, so the lemma follows in this case.

Suppose $f(x) = 0$. By Corollary 1, we have a simultaneous local resolution of singularities (M, W, ρ) for the functions $f, \varphi, g_1, \dots, g_l$ vanishing at x . For each point y in the fiber $\rho^{-1}(x)$, we have a local chart M_y satisfying property (iii) of Theorem 4. Since ρ is proper, the fiber $\rho^{-1}(x)$ is compact so there is a finite subcover $\{M_y\}$. We claim that the image $\rho(\bigcup M_y)$ contains a neighborhood W_x of x in \mathbb{R}^d . Indeed, otherwise, there exists a bounded sequence $\{x_1, x_2, \dots\}$ of points in $W \setminus \rho(\bigcup M_y)$ whose limit is x . We pick a sequence $\{y_1, y_2, \dots\}$ where $\rho(y_i) = x_i$. Since the x_i are bounded, the y_i lie in a compact set so there is a convergent subsequence $\{\tilde{y}_i\}$ with limit y_* . The \tilde{y}_i are not in the open set $\bigcup M_y$ so nor is y_* . But $\rho(y_*) = \lim \rho(\tilde{y}_i) = x$ so $y_* \in \rho^{-1}(x) \subset M_y$, a contradiction.

Now, define $\Omega_x = W_x \cap \Omega$ and let $\{\mathcal{M}_y\}$ be the collection of all sets $\mathcal{M}_y = M_y \cap \rho^{-1}(\Omega_x)$ which have positive measure. Picking a partition of unity $\{\sigma_y(\mu)\}$ subordinate to $\{\mathcal{M}_y\}$ such that σ_y is positive at y for each y [27, Theorem 6.5], we write the zeta function $\zeta(z) = \int_{\Omega_x} |f(\omega)|^{-z} |\varphi(\omega)\phi(\omega)| d\omega$ as

$$\sum_y \int_{\mathcal{M}_y} |f \circ \rho(\mu)|^{-z} |\varphi \circ \rho(\mu)| |\phi \circ \rho(\mu)| |\rho'(\mu)| \sigma_y(\mu) d\mu.$$

For each y , the boundary conditions $g_i \circ \rho(\mu) \geq 0$ become monomial inequalities, so \mathcal{M}_y is the union of closed orthant neighborhoods of y . The integral over \mathcal{M}_y is then the sum of integrals of the form

$$\zeta_y(z) = \int_{\mathbb{R}_{\geq 0}^d} \mu^{-\kappa z + \tau} \psi(\mu) d\mu$$

where κ and τ are non-negative integer vectors while ψ is a compactly supported smooth function with $\psi(0) > 0$. Note that κ and τ do not depend on ϕ nor on the choice of orthant at y . By Proposition 1, the smallest pole of $\zeta_y(z)$ is

$$\lambda_y = \min_{1 \leq j \leq d} \left\{ \frac{\tau_j + 1}{\kappa_j} \right\}, \quad \theta_y = \# \min_{1 \leq j \leq d} \left\{ \frac{\tau_j + 1}{\kappa_j} \right\}.$$

Now, $\text{RLCT}_{\Omega_x}(f; \varphi\phi) = \min_y \{(\lambda_y, \theta_y)\}$. Since this formula is independent of ϕ , we set $\phi = 1$ and the lemma follows.

Proposition 2. *Let $\phi : \Omega \rightarrow \mathbb{R}$ be positive and smooth. Then, for sufficiently small neighborhoods Ω_x , the set $\{\text{RLCT}_{\Omega_x}(f; \varphi) : x \in \Omega\}$ has a minimum and*

$$\text{RLCT}_{\Omega}(f; \varphi\phi) = \min_{x \in \Omega} \text{RLCT}_{\Omega_x}(f; \varphi).$$

Proof. Lemma 1 associates a small neighborhood to each point in the compact set Ω , so there exists a finite subcover $\{\Omega_x : x \in S\}$. Let $\{\sigma_x(\omega)\}$ be a smooth partition of unity subordinate to this subcover where $\sigma_x(x) > 0$ for all x . Then,

$$\int_{\Omega} |f(\omega)|^{-z} |\varphi(\omega)\phi(\omega)| d\omega = \sum_{x \in S} \int_{\Omega_x} |f(\omega)|^{-z} |\varphi(\omega)\phi(\omega)| \sigma_x(\omega) d\omega.$$

From this finite sum, we have

$$\text{RLCT}_\Omega(f; \varphi\phi) = \min_{x \in S} \text{RLCT}_{\Omega_x}(f; \varphi\phi\sigma_x) = \min_{x \in S} \text{RLCT}_{\Omega_x}(f; \varphi).$$

Now, if $y \in \Omega \setminus S$, let Ω_y be a neighborhood of y prescribed by Lemma 1 and consider the cover $\{\Omega_x : x \in S\} \cup \{\Omega_y\}$ of Ω . After choosing a partition of unity subordinate to this cover and repeating the above argument, we get

$$\text{RLCT}_\Omega(f; \varphi\phi) \leq \text{RLCT}_{\Omega_y}(f; \varphi) \quad \text{for all } y \in \Omega.$$

Combining the two previously displayed equations proves the proposition.

Abusing notation, we now let $\text{RLCT}_{\Omega_x}(f; \varphi)$ represent the real log canonical threshold for a sufficiently small neighborhood Ω_x of x in Ω . If x is an interior point of Ω , we denote the threshold at x by $\text{RLCT}_x(f; \varphi)$.

Corollary 2 (See also [27, §4.5]). *Given a compact semianalytic set $\Omega \subset \mathbb{R}^d$, a nearly analytic function $\varphi : \Omega \rightarrow \mathbb{R}$, and $f \in \mathcal{A}_\Omega$ satisfying $f(x) = 0$ for some $x \in \Omega$, the zeta function (5) can be continued analytically to \mathbb{C} . It has a Laurent expansion (6) whose poles are positive rational numbers with a smallest element.*

Proof. The proofs of Lemma 1 and Proposition 2 outline a way to compute the Laurent expansion of the zeta function (5).

Remark 1. *In our definition of real log canonical thresholds, we considered integrals with respect to densities $|\varphi(\omega)| d\omega$ for some nearly analytic function φ , while Watanabe only considers the special case where the density is $\varphi(\omega) d\omega$ for some smooth positive function φ . Our general case includes the situation where the absolute value of a Jacobian determinant is multiplied to the density under a change of variables. To prove the basic properties of real log canonical thresholds, we need to resolve the singularities of the variety $\varphi = 0$ together with those cut out by f, g_1, \dots, g_l , as demonstrated in Lemma 1.*

Example 1. *We now show that the threshold at a boundary point depends on the boundary inequalities. Consider the following two small neighborhoods of the origin.*

$$\begin{aligned} \Omega_1 &= \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq y \leq \varepsilon\} \\ \Omega_2 &= \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq x \leq \varepsilon\} \end{aligned}$$

To compute the real log canonical threshold of the function xy^2 over these sets, we have the corresponding zeta functions below.

$$\begin{aligned} \zeta_1(z) &= \int_0^\varepsilon \int_0^y x^{-z} y^{-2z} dx dy = \frac{\varepsilon^{-3z+2}}{(-z+1)(-3z+2)} \\ \zeta_2(z) &= \int_0^\varepsilon \int_0^x x^{-z} y^{-2z} dy dx = \frac{\varepsilon^{-3z+2}}{(-2z+1)(-3z+2)} \end{aligned}$$

This shows that $\text{RLCT}_{\Omega_1}(xy^2) = 2/3$ while $\text{RLCT}_{\Omega_2}(xy^2) = 1/2$. □

Because the real log canonical threshold over a set $\Omega \subset \mathbb{R}^d$ is the minimum of thresholds at points $x \in \Omega$, we want to know where this minimum is achieved. Let us study this problem topologically. Consider a locally finite collection \mathcal{S} of pairwise disjoint submanifolds $S \subset \Omega$ such that $\Omega = \cup_{S \in \mathcal{S}} S$ and each S is locally closed, i.e. the intersection of an open and a closed subset. Let \bar{S} be the closure of S . We say \mathcal{S} is a *stratification* of Ω if $S \cap \bar{T} \neq \emptyset$ implies $S \subset \bar{T}$ for all $S, T \in \mathcal{S}$. A stratification \mathcal{S} of Ω is a *refinement* of another stratification \mathcal{T} if $S \cap T \neq \emptyset$ implies $S \subset T$ for all $S \in \mathcal{S}$ and $T \in \mathcal{T}$.

Let the amplitude $\varphi : \Omega \rightarrow \mathbb{R}$ be nearly analytic. Let $S_{(\lambda, \theta), 1}, \dots, S_{(\lambda, \theta), r}$ be the connected components of the set $\{x \in \Omega : \text{RLCT}_{\Omega_x}(f; \varphi) = (\lambda, \theta)\}$, and let \mathcal{S} denote the collection $\{S_{(\lambda, \theta), i}\}$ where we vary over all λ, θ and i . Now, define the order $\text{ord}_x f$ to be the smallest degree of a monomial appearing in a series expansion of f at $x \in \Omega$ [10, §3.9]. This number is independent of the choice of local coordinates $\omega_1, \dots, \omega_d$ because it is the largest integer k such that $f \in \mathfrak{m}_x^k$ where $\mathfrak{m}_x = \{g \in A_x : g(x) = 0\}$ is the vanishing ideal of x . Define $T_{l, 1}, \dots, T_{l, s}$ to be the connected components of the set $\{x \in \Omega : \text{ord}_x f = l\}$ and let \mathcal{T} be the collection $\{T_{l, j}\}$ where we vary over all l and j . We conjecture the following relationship between \mathcal{S} and \mathcal{T} . It implies that the minimum real log canonical threshold over a set must occur at a point of highest order.

Conjecture 5. *The collections \mathcal{S} and \mathcal{T} are stratifications of Ω . Furthermore, if the amplitude φ is a positive smooth function, then \mathcal{S} refines \mathcal{T} .*

Laplace integrals such as (1) occur frequently in physics, statistics and other applications. At first, the relationship between their asymptotic expansions and the zeta function (3) seems strange. The key is to write these integrals as

$$\begin{aligned} Z(N) &= \int_{\Omega} e^{-N|f(\omega)|} |\varphi(\omega)| d\omega = \int_0^{\infty} e^{-Nt} v(t) dt \\ \zeta(z) &= \int_{\Omega} |f(\omega)|^{-z} |\varphi(\omega)| d\omega = \int_0^{\infty} t^{-z} v(t) dt \end{aligned}$$

where $v(t)$ is the state density function [27] or Gelfand-Leray function [2]

$$v(t) = \frac{d}{dt} \int_{0 < |f(\omega)| < t} |\varphi(\omega)| d\omega.$$

Formally, $Z(N)$ is the Laplace transform of $v(t)$ while $\zeta(z)$ is its Mellin transform. Note that contrary to its name, $v(t)$ is not strictly a function, but it can be defined as a Schwartz distribution. Next, we study the series expansions

$$Z(N) \approx \sum_{\alpha} \sum_{i=1}^d c_{\alpha, i} N^{-\alpha} (\log N)^{i-1} \quad (7)$$

$$v(t) \approx \sum_{\alpha} \sum_{i=1}^d b_{\alpha, i} t^{\alpha} (\log t)^{i-1} \quad (8)$$

$$\zeta(z) \sim \sum_{\alpha} \sum_{i=1}^d d_{\alpha,i} (z - \alpha)^{-i} \tag{9}$$

where (7) and (8) are asymptotic expansions while (9) is the principal part of the Laurent series expansion. The number d of summands here is the dimension of the parameter space $\Omega \subset \mathbb{R}^d$. Formulas relating the coefficients $b_{\alpha,i}, c_{\alpha,i}$ and $d_{\alpha,i}$ may then be deduced from the Laplace and Mellin transforms of $t^\alpha(\log t)^i$. For detailed expositions, we refer readers to Arnol'd–Guseĭn-Zade–Varchenko [2, §6-7], Watanabe [27, §4] and Greenblatt [14].

Using this strategy, we now give explicit formulas for the asymptotic expansion of an arbitrary Laplace integral. Our formulas generalize those of Arnol'd–Guseĭn-Zade–Varchenko [2, §6-7] because they apply also to parameter spaces Ω with analytic boundary. Watanabe [27, Remark 4.5] gives a similar asymptotic expansion for bounded parameter spaces but we derive precise relationships between the asymptotic coefficients $c_{\alpha,i}$ and the Laurent coefficients $d_{\alpha,i}$ in terms of derivatives $\Gamma^{(i)}$ of Gamma functions.

Theorem 6. *Let $\Omega \subset \mathbb{R}^d$ be a compact semianalytic subset and $\varphi : \Omega \rightarrow \mathbb{R}$ be nearly analytic. If $f \in \mathcal{A}_\Omega$ with $f(x) = 0$ for some $x \in \Omega$, the Laplace integral*

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

has the asymptotic expansion

$$\sum_{\alpha} \sum_{i=1}^d c_{\alpha,i} N^{-\alpha} (\log N)^{i-1}. \tag{10}$$

The α in this expansion range over positive rational numbers which are poles of

$$\zeta(z) = \int_{\Omega_\delta} |f(\omega)|^{-z} |\varphi(\omega)| d\omega \tag{11}$$

for any $\delta > 0$ and $\Omega_\delta = \{\omega \in \Omega : |f(\omega)| < \delta\}$. The coefficients $c_{\alpha,i}$ satisfy

$$c_{\alpha,i} = \frac{(-1)^i}{(i-1)!} \sum_{j=i}^d \frac{\Gamma^{(j-i)}(\alpha)}{(j-i)!} d_{\alpha,j} \tag{12}$$

where $d_{\alpha,j}$ is the coefficient of $(z - \alpha)^{-j}$ in the Laurent expansion of $\zeta(z)$.

Proof. First, set $\delta = 1$. We split the integral $Z(N)$ into two parts:

$$Z(N) = \int_{|f(\omega)| < 1} e^{-N|f(\omega)|} |\varphi(\omega)| d\omega + \int_{|f(\omega)| \geq 1} e^{-N|f(\omega)|} |\varphi(\omega)| d\omega.$$

The second integral is bounded above by Ce^{-N} for some non-negative constant C , so asymptotically it goes to zero more quickly than any $N^{-\alpha}$. For the first integral, we write $\zeta(z)$ as the Mellin transform of the state density function $v(t)$.

$$\zeta(z) = \int_{|f(\omega)| < 1} |f(\omega)|^{-z} |\varphi(\omega)| d\omega = \int_0^1 t^{-z} v(t) dt.$$

By Corollary 2, $\zeta(z)$ has a Laurent expansion (6). Since $|f(\omega)| < 1$, by dominated convergence $\zeta(z) \rightarrow 0$ as $z \rightarrow -\infty$, so the polynomial part $P(z)$ is identically zero. Applying the inverse Mellin transform [3] to $\zeta(z)$, we get a series expansion (8) of the state density function $v(t)$. Applying the Laplace transform to $v(t)$ in turn gives the asymptotic expansion (7) of $Z(N)$. The formulas

$$\int_0^\infty e^{-Nt} t^{\alpha-1} (\log t)^i dt \approx \sum_{j=0}^i \binom{i}{j} (-1)^j \Gamma^{(i-j)}(\alpha) N^{-\alpha} (\log N)^j$$

$$\int_0^1 t^{-z} t^{\alpha-1} (\log t)^i dt = -i! (z - \alpha)^{-(i+1)}$$

from [2, Thm 7.4] and [27, Ex 4.7] give us the relations

$$c_{\alpha,i} = (-1)^{i-1} \sum_{j=i}^d \binom{j-1}{i-1} \Gamma^{(j-i)}(\alpha) b_{\alpha-1,j}, \quad d_{\alpha,j} = -(j-1)! b_{\alpha-1,j}.$$

Equation (12) follows immediately. Finally, for all other values of δ , we write

$$\int_\Omega |f(\omega)|^{-z} |\varphi(\omega)| d\omega = \int_{\Omega_\delta} |f(\omega)|^{-z} |\varphi(\omega)| d\omega + \int_{|f(\omega)| \geq \delta} |f(\omega)|^{-z} |\varphi(\omega)| d\omega.$$

The last integral does not have any poles, so the principal parts of the Laurent expansions of the first two integrals are the same for all δ .

3. Real Log Canonical Thresholds

In this section, we prove basic properties of real log canonical thresholds (RLCTs) which enable us to compute them more efficiently. The learning coefficient of a statistical model is shown to be the RLCT of the ideal generated by its defining equations.

In this section, let $\Omega \subset \mathbb{R}^d$ be a compact semianalytic subset and let $\varphi : \Omega \rightarrow \mathbb{R}$ be nearly analytic. Given functions $f_1, \dots, f_r \in \mathcal{A}_\Omega$, let $\text{RLCT}_\Omega(f_1, \dots, f_r; \varphi)$ be the smallest pole and multiplicity of the zeta function (4). Recall that these pairs are ordered by the rule $(\lambda_1, \theta_1) > (\lambda_2, \theta_2)$ if $\lambda_1 > \lambda_2$, or $\lambda_1 = \lambda_2$ and $\theta_1 < \theta_2$. For $x \in \Omega$, we define $\text{RLCT}_{\Omega_x}(f_1, \dots, f_r; \varphi)$ to be the threshold for a sufficiently small neighborhood of x in Ω .

Remark 2. *The (complex) log canonical threshold may be defined in a similar fashion. It is the smallest pole of the zeta function*

$$\zeta(z) = \int_\Omega \left(|f_1(\omega)|^2 + \dots + |f_r(\omega)|^2 \right)^{-z} d\omega.$$

Note that the f_i^2 have been replaced by $|f_i|^2$ and the exponent $-z/2$ is changed to $-z$. Crudely, this factor of 2 comes from the fact that \mathbb{C}^d is a real vector space of dimension $2d$. The complex threshold is often different from the RLCT [23]. From the algebraic

geometry point of view, more is known about complex log canonical thresholds than about real log canonical thresholds. Many results in this paper were motivated by their complex analogs [6, 16, 17, 19].

Now, we give several equivalent definitions of $\text{RLCT}_\Omega(f_1, \dots, f_r; \varphi)$ which are helpful in proofs of the fundamental properties.

Proposition 3. *Given functions $f_1, \dots, f_r \in \mathcal{A}_\Omega$ such that $\mathcal{V}_\Omega(\langle f_1, \dots, f_r \rangle)$ is nonempty and each $f_i \not\equiv 0$, the pairs (λ, θ) defined below are all equal.*

a. *The logarithmic Laplace integral*

$$\log Z(N) = \log \int_{\Omega} \exp\left(-N \sum_{i=1}^r f_i(\omega)^2\right) |\varphi(\omega)| d\omega$$

is asymptotically $-\frac{\lambda}{2} \log N + (\theta - 1) \log \log N + O(1)$.

b. *The zeta function*

$$\zeta(z) = \int_{\Omega} \left(\sum_{i=1}^r f_i(\omega)^2\right)^{-z/2} |\varphi(\omega)| d\omega$$

has a smallest pole λ of multiplicity θ .

c. *The pair (λ, θ) is the minimum*

$$\min_{x \in \Omega} \text{RLCT}_{\Omega_x}(f_1, \dots, f_r; \varphi).$$

In fact, it is enough to vary x over $\mathcal{V}_\Omega(\langle f_1, \dots, f_r \rangle)$.

Proof. Item (b) is the original definition of the RLCT. The equivalence of (a) and (b) follows from Theorem 6, and that of (b) and (c) from Proposition 2. The last statement of (c) is due to the RLCT being ∞ for $x \notin \mathcal{V}_\Omega(\langle f_1, \dots, f_r \rangle)$. See also [27, Thm 7.1].

Our first property describes the effect of the boundary on the RLCT.

Proposition 4. *Let x be a point on the boundary of $\Omega \subset \mathbb{R}^d$. For every neighborhood W of x in the ambient space \mathbb{R}^d ,*

$$\text{RLCT}_W(f; \varphi) \leq \text{RLCT}_{\Omega_x}(f; \varphi).$$

Proof. For a sufficiently small neighborhood Ω_x of x in Ω , we have $\Omega_x \subset W$, so the corresponding Laplace integrals satisfy $Z_{\Omega_x}(N) \leq Z_W(N)$. By Proposition 3, this gives the opposite inequality on the RLCTs.

If the function whose RLCT we are finding is complicated, we may replace it with a simpler function that bounds it. Given $f, g \in \mathcal{A}_\Omega$, we say that f and g are *equivalent* in Ω if $c_1 f \leq g \leq c_2 f$ in Ω for some $c_1, c_2 > 0$.

Proposition 5 [27, Remark 7.2]). *Given $f, g \in \mathcal{A}_\Omega$, suppose that $0 \leq cf \leq g$ in Ω for some $c > 0$. Then, $\text{RLCT}_\Omega(f; \varphi) \leq \text{RLCT}_\Omega(g; \varphi)$.*

Corollary 3. *If f, g are equivalent in Ω , then $\text{RLCT}_\Omega(f; \varphi) = \text{RLCT}_\Omega(g; \varphi)$.*

$\text{RLCT}_\Omega(f_1^2 + \dots + f_r^2; \varphi) = (\lambda, \theta)$ implies $\text{RLCT}_\Omega(f_1, \dots, f_r; \varphi) = (2\lambda, \theta)$. From this, it seems that we should restrict ourselves to RLCTs of single and not multiple functions. However, as the next proposition shows, multiple functions are important because they allow us to work with ideals for which different generating sets can be chosen. This gives us freedom to switch between single and multiple functions in powerful ways. For instance, special cases of this proposition such as Lemmas 3 and 4 of [1] have been used to simplify computations considerably.

Proposition 6. *If two sets $\{f_1, \dots, f_r\}$ and $\{g_1, \dots, g_s\}$ of functions generate the same ideal $I \subset \mathcal{A}_\Omega$, then*

$$\text{RLCT}_\Omega(f_1, \dots, f_r; \varphi) = \text{RLCT}_\Omega(g_1, \dots, g_s; \varphi).$$

Define this pair to be $\text{RLCT}_\Omega(I; \varphi)$.

Proof. Each g_j can be written as a combination $h_1 f_1 + \dots + h_r f_r$ of the f_i where the h_i are real analytic over Ω . By the Cauchy-Schwarz inequality,

$$g_j^2 \leq (h_1^2 + \dots + h_r^2)(f_1^2 + \dots + f_r^2).$$

Because Ω is compact, the h_i are bounded. Thus, summing over all the g_j , there is some constant $c > 0$ such that,

$$\sum_{j=1}^s g_j^2 \leq c \sum_{i=1}^r f_i^2.$$

By Proposition 5, $\text{RLCT}_\Omega(g_1, \dots, g_s; \varphi) \leq \text{RLCT}_\Omega(f_1, \dots, f_r; \varphi)$ and by symmetry, the reverse is also true, so we are done. See also [23, §2.6].

For the next result, let $f_1, \dots, f_r \in \mathcal{A}_X$ and $g_1, \dots, g_s \in \mathcal{A}_Y$ where $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ are compact semianalytic subsets. This occurs, for instance, when the f_i and g_j are polynomials with disjoint sets of indeterminates $\{x_1, \dots, x_m\}$ and $\{y_1, \dots, y_n\}$. Let $\varphi_X : X \rightarrow \mathbb{R}$ and $\varphi_Y : Y \rightarrow \mathbb{R}$ be nearly analytic. Define $(\lambda_X, \theta_X) = \text{RLCT}_X(f_1, \dots, f_r; \varphi_X)$ and $(\lambda_Y, \theta_Y) = \text{RLCT}_Y(g_1, \dots, g_s; \varphi_Y)$.

By composing with projections $X \times Y \rightarrow X$ and $X \times Y \rightarrow Y$, we may regard the f_i and g_j as functions analytic over $X \times Y$. Let I_X and I_Y be ideals in $\mathcal{A}_{X \times Y}$ generated by the f_i and g_j respectively. Recall that the sum $I_X + I_Y$ is generated by all the f_i and g_j while the product $I_X I_Y$ is generated by $f_i g_j$ for all i, j .

Proposition 7. *The RLCTs for the sum and product of ideals I_X and I_Y are*

$$\begin{aligned} \text{RLCT}_{X \times Y}(I_X + I_Y; \varphi_X \varphi_Y) &= (\lambda_X + \lambda_Y, \theta_X + \theta_Y - 1), \\ \text{RLCT}_{X \times Y}(I_X I_Y; \varphi_X \varphi_Y) &= \begin{cases} (\lambda_X, \theta_X) & \text{if } \lambda_X < \lambda_Y, \\ (\lambda_Y, \theta_Y) & \text{if } \lambda_X > \lambda_Y, \\ (\lambda_X, \theta_X + \theta_Y) & \text{if } \lambda_X = \lambda_Y. \end{cases} \end{aligned}$$

Proof. Define $f(x) = f_1^2 + \dots + f_r^2$ and $g(y) = g_1^2 + \dots + g_s^2$, and let $Z_X(N)$ and $Z_Y(N)$ be the corresponding Laplace integrals. By Proposition 3,

$$\begin{aligned} \log Z_X(N) &= -\frac{1}{2}\lambda_X \log N + (\theta_X - 1) \log \log N + O(1) \\ \log Z_Y(N) &= -\frac{1}{2}\lambda_Y \log N + (\theta_Y - 1) \log \log N + O(1) \end{aligned}$$

asymptotically. If $(\lambda, \theta) = \text{RLCT}_{X \times Y}(I_X + I_Y; \varphi_X \varphi_Y)$, then

$$\begin{aligned} &-\frac{1}{2}\lambda \log N + (\theta - 1) \log \log N + O(1) \\ &= \log \int_{X \times Y} e^{-Nf(x) - Ng(y)} |\varphi_X| |\varphi_Y| dx dy \\ &= \log \left(\int_X e^{-Nf(x)} |\varphi_X| dx \right) \left(\int_Y e^{-Ng(y)} |\varphi_Y| dy \right) \\ &= \log Z_X(N) + \log Z_Y(N) \\ &= -\frac{1}{2}(\lambda_X + \lambda_Y) \log N + (\theta_X + \theta_Y - 2) \log \log N + O(1) \end{aligned}$$

and the first result follows. For the second result, note that

$$f(x)g(y) = f_1^2 g_1^2 + f_1^2 g_2^2 + \dots + f_r^2 g_s^2.$$

Let $\zeta_X(z)$ and $\zeta_Y(z)$ be the zeta functions corresponding to $f(x)$ and $g(y)$. According to Proposition 3, (λ_X, θ_X) and (λ_Y, θ_Y) are the smallest poles of $\zeta_X(z)$ and $\zeta_Y(z)$. Meanwhile, $\text{RLCT}_{X \times Y}(I_X I_Y; \varphi_X \varphi_Y)$ is the smallest pole of

$$\begin{aligned} \zeta(z) &= \int_{X \times Y} (f(x)g(y))^{-z/2} |\varphi_X| |\varphi_Y| dx dy \\ &= \left(\int_X f(x)^{-z/2} |\varphi_X| dx \right) \left(\int_Y g(y)^{-z/2} |\varphi_Y| dy \right) = \zeta_X(z) \zeta_Y(z). \end{aligned}$$

The second result then follows from the relationship between the poles.

Our last property tells us the behavior of RLCTs under a change of variables. Consider an ideal $I \subset \mathcal{A}_W$ where W is a neighborhood of the origin. Let M be a real analytic manifold and $\rho : M \rightarrow W$ be a proper real analytic map. Then, the *pullback* ρ^*I is locally the ideal of real analytic functions on M that is generated by $f \circ \rho$ for all $f \in I$ (also called the inverse image ideal sheaf [18, §3.3]). If ρ is an isomorphism between $M \setminus \mathcal{V}(\rho^*I)$ and $W \setminus \mathcal{V}(I)$, we say that ρ is a *change of variables away from* $\mathcal{V}(I)$. Let $|\rho'|$ denote the Jacobian determinant of ρ . We call $(\rho^*I; (\varphi \circ \rho)|\rho'|)$ the *pullback pair*.

Proposition 8. *Let W be a neighborhood of the origin and $I \subset \mathcal{A}_W$ a finitely generated ideal. If M is a real analytic manifold, $\rho : M \rightarrow W$ is a change of variables away from $\mathcal{V}(I)$ and $\mathcal{M} = \rho^{-1}(\Omega \cap W)$, then*

$$\text{RLCT}_{\Omega_0}(I; \varphi) = \min_{x \in \rho^{-1}(0)} \text{RLCT}_{\mathcal{M}_x}(\rho^*I; (\varphi \circ \rho)|\rho'|).$$

Proof. Let f_1, \dots, f_r generate I and let $f = f_1^2 + \dots + f_r^2$. Then, $\text{RLCT}_{\Omega_0}(I; \varphi)$ is the smallest pole and multiplicity of the zeta function

$$\zeta(z) = \int_{\Omega_0} f(\omega)^{-z/2} |\varphi(\omega)| d\omega$$

where $\Omega_0 \subset W$ is a sufficiently small neighborhood of the origin in Ω . Applying the change of variables ρ , we have

$$\zeta(z) = \int_{\rho^{-1}(\Omega_0)} f \circ \rho(\mu)^{-z/2} |\varphi \circ \rho(\mu)| |\rho'(\mu)| d\mu.$$

The proof of Lemma 1 shows that if Ω_0 is sufficiently small, there are finitely many points $y \in \rho^{-1}(0)$ and a cover $\{\mathcal{M}_y\}$ of $\mathcal{M} = \rho^{-1}(\Omega_0)$ such that

$$\zeta(z) = \sum_y \int_{\mathcal{M}_y} f \circ \rho(\mu)^{-z/2} |\varphi \circ \rho(\mu)| |\rho'(\mu)| \sigma_y(\mu) d\mu$$

where $\{\sigma_y\}$ is a partition of unity subordinate to $\{\mathcal{M}_y\}$. Furthermore, the $f_i \circ \rho$ generate the pullback ρ^*I and $f \circ \rho = (f_1 \circ \rho)^2 + \dots + (f_r \circ \rho)^2$. Therefore,

$$\text{RLCT}_{\mathcal{M}_y}(f \circ \rho; (\varphi \circ \rho)|\rho'|\sigma_y) = \text{RLCT}_{\mathcal{M}_y}(\rho^*I; (\varphi \circ \rho)|\rho'|)$$

and the result follows from the two previously displayed equations.

We are now ready to prove Theorem 2 which was inspired by Watanabe.

Proof of Theorem 2. Let $Q(\omega) = \sum_{i=1}^k (p_i(\omega) - q_i)^2$. The learning coefficient is the RLCT of the Kullback-Leibler distance $K(\omega)$, so it is enough to show that $\text{RLCT}_{\Omega_x} K = \text{RLCT}_{\Omega_x} Q$ for each $x \in \mathcal{V}(K) = \mathcal{V}(Q)$. By Corollary 3, we only need to show that K and Q are equivalent in a sufficiently small neighborhood of x . Now, the Taylor expansion $-\log t = (1-t) + \frac{1}{2}(1-t)^2 + \dots$ implies there are $c_1, c_2 > 0$ such that near $t = 1$,

$$c_1(t-1)^2 \leq -\log t + t - 1 \leq c_2(t-1)^2. \tag{13}$$

Choosing a sufficiently small W_x such that $p_i(\omega)/q_i$ is near 1, we have

$$c_1 \left(\frac{p_i(\omega)}{q_i} - 1 \right)^2 \leq -\log \frac{p_i(\omega)}{q_i} + \frac{p_i(\omega)}{q_i} - 1 \leq c_2 \left(\frac{p_i(\omega)}{q_i} - 1 \right)^2$$

for all $\omega \in W_x$. Multiplying by q_i , summing from $i = 1$ to k and observing that the p_i and the q_i add up to 1, we get

$$c_1 \sum_{i=1}^k q_i \left(\frac{p_i(\omega)}{q_i} - 1 \right)^2 \leq K(\omega) \leq c_2 \sum_{i=1}^k q_i \left(\frac{p_i(\omega)}{q_i} - 1 \right)^2.$$

Again, using the fact that the q_i are non-zero, we have

$$\frac{c_1}{\max_i q_i} \sum_{i=1}^k (p_i(\omega) - q_i)^2 \leq K(\omega) \leq \frac{c_2}{\min_i q_i} \sum_{i=1}^k (p_i(\omega) - q_i)^2$$

which completes the claim. The more general statement for a real analytic $K(\omega)$ which is bounded by scalar multiples of a sum of squared functions follows from Proposition 2, Corollary 3 and the definition of $\text{RLCT}_{\Omega}(I; \varphi)$.

4. Newton Polyhedra and Nondegeneracy

Given an analytic function $f \in \mathcal{A}_0(\mathbb{R}^d)$, we pick local coordinates $\{\omega_1, \dots, \omega_d\}$ in a neighborhood of the origin. This allows us to represent f as a power series $\sum_{\alpha} c_{\alpha} \omega^{\alpha}$ where $\omega = (\omega_1, \dots, \omega_d)$ and each $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$. Let $[\omega^{\alpha}]f$ denote the coefficient c_{α} of ω^{α} in this expansion. Define its *Newton polyhedron* $\mathcal{P}(f) \subset \mathbb{R}^d$ to be the convex hull

$$\mathcal{P}(f) = \text{conv} \{ \alpha + \alpha' : [\omega^{\alpha}]f \neq 0, \alpha' \in \mathbb{R}_{\geq 0}^d \}.$$

A subset $\gamma \subset \mathcal{P}(f)$ is a *face* if there exists $\beta \in \mathbb{R}^d$ such that

$$\gamma = \{ \alpha \in \mathcal{P}(f) : \alpha \cdot \beta \leq \alpha' \cdot \beta \text{ for all } \alpha' \in \mathcal{P}(f) \}.$$

where \cdot is the standard dot product. Dually, the *normal cone* at γ is the set of all $\beta \in \mathbb{R}^d$ satisfying the above condition. Each β lies in the non-negative orthant $\mathbb{R}_{\geq 0}^d$ so that $\alpha \cdot \beta$ achieves a minimum as α varies over the unbounded set $\mathcal{P}(f)$. As a result, the union of all the normal cones gives a partition $\mathcal{F}(f)$ of the non-negative orthant called the *normal fan*. Now, given a compact subset $\gamma \subset \mathbb{R}^d$, define the *face polynomial*

$$f_{\gamma} = \sum_{\alpha \in \gamma \cap \mathbb{N}^d} c_{\alpha} \omega^{\alpha}.$$

Recall that f_{γ} is singular at a point $x \in \mathbb{R}^d$ if $\text{ord}_x f_{\gamma} \geq 2$, i.e.

$$f_{\gamma}(x) = \frac{\partial f_{\gamma}}{\partial \omega_1}(x) = \dots = \frac{\partial f_{\gamma}}{\partial \omega_d}(x) = 0.$$

We say that f is *nondegenerate* if f_{γ} is non-singular at all points in the torus $(\mathbb{R}^*)^d$ for all compact faces γ of $\mathcal{P}(f)$, otherwise we say f is *degenerate*. Now, we define the *distance* l of $\mathcal{P}(f)$ to be the smallest $t \geq 0$ such that $(t, t, \dots, t) \in \mathcal{P}(f)$. Let the *multiplicity* θ of l be the codimension of the lowest-dimensional face of $\mathcal{P}(f)$ at this intersection of the diagonal with $\mathcal{P}(f)$. However, if $l = 0$, we leave θ undefined. These notions of nondegeneracy, distance and multiplicity were first coined and studied by Varchenko [24].

We now extend the above notions to ideals. For any ideal $I \subset \mathcal{A}_0$, define

$$\mathcal{P}(I) = \text{conv} \{ \alpha \in \mathbb{R}^d : [\omega^{\alpha}]f \neq 0 \text{ for some } f \in I \}.$$

Related to this geometric construction is the monomial ideal

$$\text{mon}(I) = \langle \omega^{\alpha} : [\omega^{\alpha}]f \neq 0 \text{ for some } f \in I \rangle.$$

Note that I and $\text{mon}(I)$ have the same Newton polyhedron, and if I is generated by f_1, \dots, f_r , then $\text{mon}(I)$ is generated by monomials ω^{α} appearing in the f_i . One consequence is that $\mathcal{P}(f_1^2 + \dots + f_r^2)$ is the scaled polyhedron $2\mathcal{P}(I)$. More importantly, the threshold of I is bounded by that of $\text{mon}(I)$. To prove this result, we need the following lemma. Recall that by the Hilbert Basis Theorem or by Dickson's Lemma [10], $\text{mon}(I)$ is finitely generated.

Lemma 2. *Given $f \in \mathcal{A}_0(\mathbb{R}^d)$, let S be a finite set of exponents α of monomials ω^α which generate $\text{mon}(\langle f \rangle)$. Then, there is a constant $c > 0$ such that*

$$|f(\omega)| \leq c \sum_{\alpha \in S} |\omega|^\alpha$$

in a sufficiently small neighborhood of the origin.

Proof. Let $\sum_{\alpha} c_{\alpha} \omega^{\alpha}$ be the power series expansion of f . Because f is analytic at the origin, there exists $\varepsilon > 0$ such that

$$\sum_{\alpha} |c_{\alpha}| \varepsilon^{\alpha_1 + \dots + \alpha_d} < \infty.$$

Let $S = \{\alpha^{(1)}, \dots, \alpha^{(s)}\}$ where the monomials $\omega^{\alpha^{(i)}}$ generate $\text{mon}(\langle f \rangle)$. Then,

$$f(\omega) = \omega^{\alpha^{(1)}} g_1(\omega) + \dots + \omega^{\alpha^{(s)}} g_s(\omega)$$

for some power series $g_i(\omega)$. Each $g_i(\omega)$ is absolutely convergent in the ε -neighborhood U of the origin because f is absolutely convergent in U . Thus, the $g_i(\omega)$ are analytic. Their absolute values are bounded above by some constant c in U , and the lemma follows.

Below, $\text{RLCT}_0(I; \varphi)$ denotes the RLCT of I with respect to φ at the origin.

Proposition 9. *Let $I \subset \mathcal{A}_0$ be a finitely generated ideal and $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be nearly analytic at the origin. Then,*

$$\text{RLCT}_0(I; \varphi) \leq \text{RLCT}_0(\text{mon}(I); \varphi).$$

Proof. Given $f \in \mathcal{A}_0(\mathbb{R}^d)$, let S be a finite set of generating exponents α for $\text{mon}(\langle f \rangle)$. By Lemma 2 and the Cauchy-Schwarz inequality, there exist constants $c, c' > 0$ such that

$$f^2 \leq \left(c \sum_{\alpha \in S} |\omega|^\alpha \right)^2 \leq c' \sum_{\alpha \in S} \omega^{2\alpha}$$

in a sufficiently small neighborhood of the origin. Therefore, if f_1, \dots, f_r generate I , then $f_1^2 + \dots + f_r^2$ is bounded by a constant multiple of the sum of squares of monomials generating $\text{mon}(I)$. The result now follows from Proposition 5.

Given a compact subset $\gamma \subset \mathbb{R}^d$, define the *face ideal*

$$I_{\gamma} = \langle f_{\gamma} : f \in I \rangle.$$

The next result tells us how to compute I_{γ} for an ideal $I = \langle f_1, \dots, f_r \rangle$.

Proposition 10. *For all compact faces $\gamma \in \mathcal{P}(I)$, $I_{\gamma} = \langle f_{1\gamma}, \dots, f_{r\gamma} \rangle$.*

Proof. By definition, $\langle f_{1\gamma}, \dots, f_{r\gamma} \rangle \subset I_\gamma$. For the other inclusion, it is enough to show that $f_\gamma \in \langle f_{1\gamma}, \dots, f_{r\gamma} \rangle$ for all $f \in I$. First, we claim that if $\omega^\alpha = \omega^{\alpha'} \omega^{\alpha''}$ with $\alpha \in \gamma$ and $\omega^{\alpha'} \in \text{mon}(I)$, then $\omega^{\alpha''} = 1$. Indeed, for all $\beta \in \mathbb{R}_{\geq 0}^d$ in the normal cone dual to γ , we have $\alpha \cdot \beta = \alpha' \cdot \beta + \alpha'' \cdot \beta$, but $\alpha \cdot \beta \leq \alpha' \cdot \beta$ so $\alpha'' \cdot \beta = 0$. This implies $\alpha'' + k\alpha'' \in \gamma$ for all integers $k > 0$. Since γ is compact, α'' must be the zero vector so $\omega^{\alpha''} = 1$.

Now, if $f \in I$, then $f = h_1 f_1 + \dots + h_r f_r$ for some analytic functions h_1, \dots, h_r . Clearly, $f_\gamma = (h_1 f_1)_\gamma + \dots + (h_r f_r)_\gamma$. By the above claim, $(h_i f_i)_\gamma = h_{i0} f_{i\gamma}$ where h_{i0} is the constant term in h_i . Hence, $f_\gamma = h_{10} f_{1\gamma} + \dots + h_{r0} f_{r\gamma} \in \langle f_{1\gamma}, \dots, f_{r\gamma} \rangle$ as required.

Remark 3. Let β be a vector in the normal cone dual to the face γ of $\mathcal{P}(I)$. Consider the weight order associated to β , and let $\text{in}_\beta f$ be the sum of all the terms of f that are maximal with respect to this (partial) order [10, §15]. Let $\text{in}_\beta I$ be the initial ideal $\text{in}_\beta I = \langle \text{in}_\beta f : f \in I \rangle$. A set of functions $f_1, \dots, f_r \in I$ is a Gröbner basis for I if and only if

$$\text{in}_\beta I = \langle \text{in}_\beta f_1, \dots, \text{in}_\beta f_r \rangle.$$

Comparing this statement with the previous result, one could ask why the generators f_1, \dots, f_r of I need not be a Gröbner basis for Proposition 10 to hold. This confusion comes from incorrectly equating the face ideal I_γ with the initial ideal $\text{in}_\beta I$ when we only have containment $I_\gamma \subset \text{in}_\beta I$. For instance, if

$$I = \langle f_1, f_2, f_3 \rangle = \langle xy - z^3, xz - y^3, yz - x^3 \rangle$$

and γ is the convex hull of $\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$, then $I_\gamma = \langle xy, xz, yz \rangle$. Meanwhile, $\beta = (1, 1, 1)$ and $\text{in}_\beta I$ contains $y^4 - z^4 = z f_1 - y f_2$ but $y^4 - z^4 \notin I_\gamma$.

Lastly, we give several equivalent definitions of nondegeneracy for ideals. If an ideal I satisfies them, we say that I is *sos-nondegenerate*, where *sos* stands for *sum-of-squares*.

Proposition 11. Let $I \subset \mathcal{A}_0$ be an ideal. The following are equivalent:

1. For some generating set $\{f_1, \dots, f_r\}$ for I , $f_1^2 + \dots + f_r^2$ is nondegenerate.
2. For all generating sets $\{f_1, \dots, f_r\}$ for I , $f_1^2 + \dots + f_r^2$ is nondegenerate.
3. For all compact faces $\gamma \subset \mathcal{P}(I)$, the variety $\mathcal{V}(I_\gamma) \subset \mathbb{R}^d$ does not intersect $(\mathbb{R}^*)^d$.

Proof. Let f_1, \dots, f_r generate I and let $f = f_1^2 + \dots + f_r^2$. If γ is a compact face of $\mathcal{P}(I)$, then the set (2γ) is a compact face of $\mathcal{P}(f) = 2\mathcal{P}(I)$. Moreover, $f_{(2\gamma)} = f_{1\gamma}^2 + \dots + f_{r\gamma}^2$ and

$$\frac{\partial f_{(2\gamma)}}{\partial \omega_i} = 2f_{1\gamma} \frac{\partial f_{1\gamma}}{\partial \omega_i} + \dots + 2f_{r\gamma} \frac{\partial f_{r\gamma}}{\partial \omega_i}.$$

Now, $f_{1\gamma}^2 + \dots + f_{r\gamma}^2 = 0$ if and only if $f_{1\gamma} = \dots = f_{r\gamma} = 0$. It follows that f is nondegenerate if and only if $\mathcal{V}(\langle f_{1\gamma}, \dots, f_{r\gamma} \rangle) \cap (\mathbb{R}^*)^d = \mathcal{V}(I_\gamma) \cap (\mathbb{R}^*)^d = \emptyset$ for all compact faces $\gamma \subset \mathcal{P}(I)$. This proves (1) \Leftrightarrow (3) and (2) \Leftrightarrow (3).

Remark 4. *The nondegeneracy of a function f need not imply the sos-non-degeneracy of the ideal $\langle f \rangle$, e.g. $f = x + y$.*

Remark 5. *After finishing this paper, we discovered another notion of nondegeneracy for ideals of complex formal power series due to Saia [22], which was shown to be equivalent to the complex version of Proposition 11(3) [5, §2].*

We recall some basic facts about toric varieties. We say a polyhedral cone σ is generated by vectors $v_1, \dots, v_k \in \mathbb{R}^d$ if $\sigma = \{\sum_i \lambda_i v_i : \lambda_i \geq 0\}$. If σ is generated by lattice vectors $v_i \in \mathbb{Z}^d$, then σ is *rational*. If the origin is a face of σ , then σ is *pointed*. A *ray* is a pointed one-dimensional cone. Every rational ray has a lattice generator of minimal length called the *minimal generator*, and every pointed rational polyhedral cone σ is generated by the minimal generators of its one-dimensional faces. If these minimal generators are linearly independent over \mathbb{R} , then σ is *simplicial*. A simplicial cone is *smooth* if its minimal generators also form part of a \mathbb{Z} -basis of \mathbb{Z}^d . A collection \mathcal{F} of pointed rational polyhedral cones in \mathbb{R}^d is a *fan* if the faces of every cone in \mathcal{F} are in \mathcal{F} and the intersection of any two cones in \mathcal{F} are again in \mathcal{F} . The *support* of \mathcal{F} is the union of its cones as subsets of \mathbb{R}^d . If the support of \mathcal{F} is the non-negative orthant, then \mathcal{F} is *locally complete*. If every cone of \mathcal{F} is simplicial (resp. smooth), then \mathcal{F} is *simplicial* (resp. *smooth*). A fan \mathcal{F}_1 is a *refinement* of another fan \mathcal{F}_2 if the cones of \mathcal{F}_1 come from partitioning the cones of \mathcal{F}_2 . See [12] for more details.

Given a smooth simplicial locally complete fan \mathcal{F} , we have a smooth toric variety $\mathbb{P}(\mathcal{F})$ covered by open charts $U_\sigma \simeq \mathbb{R}^d$, one for each cone σ of \mathcal{F} that is maximal under inclusion. Furthermore, the *blow-up* $\rho_{\mathcal{F}} : \mathbb{P}(\mathcal{F}) \rightarrow \mathbb{R}^d$ is defined as follows: for each maximal cone σ of \mathcal{F} minimally generated by v_1, \dots, v_d with $v_i = (v_{i1}, \dots, v_{id})$, we have monomial maps $\rho_\sigma : U_\sigma \rightarrow \mathbb{R}^d$ on the open charts.

$$\begin{aligned} (\mu_1, \dots, \mu_d) &\mapsto (\omega_1, \dots, \omega_d) \\ \omega_1 &= \mu_1^{v_{11}} \mu_2^{v_{21}} \dots \mu_d^{v_{d1}} \\ \omega_2 &= \mu_1^{v_{12}} \mu_2^{v_{22}} \dots \mu_d^{v_{d2}} \\ &\vdots \\ \omega_d &= \mu_1^{v_{1d}} \mu_2^{v_{2d}} \dots \mu_d^{v_{dd}} \end{aligned}$$

Let $v = v_\sigma$ be the matrix (v_{ij}) where each minimal generator v_i forms a row of v . We represent the above monomial map by $\omega = \mu^v$. If v_{i+} represents the i -th row sum of v , the Jacobian determinant of this map is

$$(\det v) \mu_1^{v_{1+}-1} \dots \mu_d^{v_{d+}-1}.$$

We will now connect these concepts. The next two theorems are due to Varchenko, see [24] and [2, §8.3]. His notion of degeneracy is weaker than ours because it does not include the condition $f_\gamma = 0$, but his proof [2, Lemma 8.9] actually supports the stronger notion. The set up is as follows: suppose f is analytic in a neighborhood W of the origin. Let \mathcal{F} be any smooth simplicial refinement of the normal fan $\mathcal{F}(f)$ and $\rho_{\mathcal{F}}$ be the blow-up associated to \mathcal{F} . Set $M = \rho_{\mathcal{F}}^{-1}(W)$. Let l be the distance of $\mathcal{P}(f)$ and θ its multiplicity.

Theorem 7. $(M, W, \rho_{\mathcal{F}})$ desingularizes f at 0 if f is nondegenerate.

Theorem 8. $\text{RLCT}_0 f = (1/l, \theta)$ if $(M, W, \rho_{\mathcal{F}})$ desingularizes f at 0.

We extend Theorem 8 to compute $\text{RLCT}_0(f; \omega^\tau)$ for monomials ω^τ . Given a polyhedron $\mathcal{P}(f) \subset \mathbb{R}^d$ and a vector $\tau = (\tau_1, \dots, \tau_d)$ of non-negative integers, let the τ -distance l_τ be the smallest $t \geq 0$ such that $t(\tau_1 + 1, \dots, \tau_d + 1) \in \mathcal{P}(f)$ and let the multiplicity θ_τ be the codimension of the face at this intersection.

Theorem 9. $\text{RLCT}_0(f; \omega^\tau) = (1/l_\tau, \theta_\tau)$ if $(M, W, \rho_{\mathcal{F}})$ desingularizes f at 0.

Proof. We follow roughly the proof in [2, §8] of Theorem 8. Let σ be a maximal cone of \mathcal{F} . Because \mathcal{F} refines $\mathcal{F}(f)$, σ is a subset of some maximal cone σ' of $\mathcal{F}(f)$. Let $\alpha \in \mathbb{R}^d$ be the vertex of $\mathcal{P}(f)$ dual to σ' . Let v be the matrix whose rows are minimal generators of σ and ρ the monomial map $\mu \mapsto \mu^v$. Under this map, the term $c_\alpha \omega^\alpha$ in f becomes the leading monomial, so $f(\rho(\mu)) = g(\mu)\mu^{v\alpha}$ for some non-vanishing function g . Then,

$$\begin{aligned} |f(\omega)|^{-z} |\omega^\tau| d\omega &= |f(\rho(\nu))|^{-z} |\rho(\mu)|^\tau |\rho'(\mu)| d\mu \\ &= (\det v) |g(\mu)|^{-z} |\mu|^{-v\alpha z} |\mu^{v\tau} \mu_1^{v_1+1} \cdots \mu_d^{v_d+1}| \end{aligned}$$

Thus, for the cone σ ,

$$(\lambda_\sigma, \theta_\sigma) = (\min S, \#\min S), \quad S = \left\{ \frac{v_i \cdot (\tau + 1)}{v_i \cdot \alpha} : 1 \leq i \leq d \right\}$$

where $\tau + 1 = (\tau_1 + 1, \dots, \tau_d + 1)$. We now give an interpretation for the elements of S . Fixing i , let P be the affine hyperplane normal to v_i passing through α . Then, $(v_i \cdot \alpha)/(v_i \cdot (\tau + 1))$ is the distance of P from the origin along the ray $\{t(\tau + 1) : t \geq 0\}$. Since $\text{RLCT}_0(f; \omega^\tau) = \min_\sigma (\lambda_\sigma, \theta_\sigma)$, the result follows.

Remark 6. After finishing this paper, the author discovered that a similar result was proved by Vasil'ev [25] for complex analytic functions. \square

Monomial ideals play in special role in the theory of real log canonical thresholds of ideals. The proof of this next result is due to Piotr Zwiernik.

Proposition 12. Monomial ideals are sos-nondegenerate.

Proof. Let $f = f_1^2 + \cdots + f_r^2$ where f_1, \dots, f_r are monomials generating I . For each face γ of $\mathcal{P}(I)$, f_γ is also a sum of squares of monomials, so f_γ does not have any zeros in $(\mathbb{R}^*)^d$ and the result now follows from Proposition 11(3).

Our tools now allow us to prove Theorem 3. As a special case, we have a formula for the RLCT of a monomial ideal with respect to a monomial amplitude function. The analogous formula for complex log canonical thresholds of monomial ideals was discovered and proved by Howald [16].

Proof of Theorem 3. If the ideal I is sos-nondegenerate, then the equality follows from Proposition 11, Theorem 7 and Theorem 9. For all other ideals, the inequality is the result of Proposition 9 and Proposition 12.

Remark 7. Define the principal part $f_{\mathcal{P}}$ of f to be $\sum_{\alpha} c_{\alpha} \omega^{\alpha}$ where the sum is over all α lying in some compact face γ of $\mathcal{P}(f)$. The above theorems imply that if f is nondegenerate, then $\text{RLCT}_0 f = \text{RLCT}_0 f_{\mathcal{P}}$. However, the latter is not true in general. For instance, if $f = (x + y)^2 + y^4$, then $f_{\mathcal{P}} = (x + y)^2$ but $\text{RLCT}_0 f = (3/4, 1)$ and $\text{RLCT}_0 f_{\mathcal{P}} = (1/2, 1)$.

Our first corollary shows that the BIC is a special case of Theorem 1.

Corollary 4. If $f \in \mathcal{A}_0(\mathbb{R}^d)$ has a local minimum at the origin with $f(0) = 0$ and its Hessian $(\partial^2 f / \partial \omega_i \partial \omega_j)$ is full rank, then $\text{RLCT}_0 f = (d/2, 1)$.

Proof. Because its Hessian is full rank, there is a linear change of variables such that $f = \omega_1^2 + \dots + \omega_d^2 + O(\omega^3)$. Thus, f is nondegenerate and the Newton polyhedron $\mathcal{P}(f)$ has distance $l = 2/d$ with $\theta = 1$.

Corollary 5. Let I be the ideal $\langle f_1, \dots, f_s \rangle$, and suppose the Jacobian matrix $(\partial f_i / \partial \omega_j)$ has rank r at 0. Then, $\text{RLCT}_0 I \leq (\frac{1}{2}(r + d), 1)$.

Proof. Because the rank of $(\partial f_i / \partial \omega_j)$ is r , there is a linear change of variables such that the only linear monomials appearing in I are $\omega_1, \dots, \omega_r$. It follows that $\mathcal{P}(I)$ lies in the halfspace $\alpha_1 + \dots + \alpha_r + \frac{1}{2}(\alpha_{r+1} + \dots + \alpha_d) \geq 1$ and its distance is at least $1/(r + \frac{d-r}{2}) = 2/(r + d)$.

5. Applications to Statistical Models

In this section, we use our tools to compute the learning coefficients of a naïve Bayesian network \mathcal{M} with two ternary random variables and two hidden states. It was designed by Evans, Gilula and Guttman [11] for investigating connections between the recovery time of 132 schizophrenic patients and the frequency of visits by their relatives. Their data is summarized in the 3×3 contingency table

	$2 \leq Y < 10$	$10 \leq Y < 20$	$20 \leq Y$	Totals
Visited regularly	43	16	3	62
Visited rarely	6	11	10	27
Visited never	9	18	16	43
Totals	58	45	29	132

which we store as a 3×3 matrix \hat{q} of relative frequencies. The model is given by

$$\begin{aligned}
 p: \quad \Omega &= \Delta_1 \times \Delta_2 \times \Delta_2 \times \Delta_2 \times \Delta_2 \quad \rightarrow \quad \Delta_8 \\
 \omega &= (t, a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2) \quad \mapsto \quad (p_{ij}) \\
 p_{ij} &= t a_i b_j + (1 - t) c_i d_j, \quad i, j \in \{0, 1, 2\}
 \end{aligned}$$

where $a_0 = 1 - a_1 - a_2$, $a = (a_0, a_1, a_2) \in \Delta_2$ and similarly for b, c and d . Hence, a 3×3 matrix in the model is a convex combination of two rank one matrices, so it has rank at

most two. The marginal likelihood of the data (after discarding a constant multiplicative factor) is the integral

$$\mathcal{I} = \int_{\Omega} p_{00}^{43} p_{01}^{16} p_{02}^3 p_{10}^6 p_{11}^{11} p_{12}^{10} p_{20}^9 p_{21}^{18} p_{22}^{16} d\omega$$

which was computed exactly by Sturmfels, Xu and the author [20].

We now estimate this integral using Watanabe's asymptotic formula for the log likelihood integral in Theorem 1. We assume that the data \hat{q} was generated by some true distribution $q = (q_{ij}) \in \mathbb{R}^{3 \times 3}$ in the model. Ideally, we want q to be equal to the matrix \hat{q} of relative frequencies, but in general, the data \hat{q} rarely lies in the model. In this example, the matrix \hat{q} is not in the model because it is full rank. However, we *should* be able to find a distribution q in the model that is close to \hat{q} , because in practice, we want to study models which describe the data well. A good candidate for q is the maximum likelihood distribution. Using the EM algorithm, this distribution is

$$q = \frac{1}{132} \begin{pmatrix} 43.00153927 & 15.99813189 & 3.000328847 \\ 5.979732739 & 11.12298188 & 9.897285383 \\ 9.018728012 & 17.87888620 & 16.10238577 \end{pmatrix}$$

which comes from the maximum likelihood estimate

$$\begin{aligned} t &= 0.5129202328 \\ (a_1, a_2) &= (0.09139459898, 0.3457903589), \\ (b_1, b_2) &= (0.1397061214, 0.4386217768), \\ (c_1, c_2) &= (0.8680689680, 0.05580725171), \\ (d_1, d_2) &= (0.7549807403, 0.2380125694). \end{aligned}$$

Note that the ML distribution q is indeed very close to the data \hat{q} .

Our next theorem summarizes how the asymptotics of $\log Z(N)$ depend on q . Let S_i denote the set of rank i matrices in $p(\Omega)$, and $S_i^* \subset S_i$ be the matrices with positive entries. Before we prove this theorem, let us apply it to our statistical problem. Using the exact value of \mathcal{I} computed by Lin–Sturmfels–Xu [20], we get

$$(\log \mathcal{I})_{\text{exact}} = -273.1911759.$$

Meanwhile, if the BIC was erroneously applied with the dimension $d = 9$ of the parameter space, we would have

$$(\log \mathcal{I})_{\text{BIC}} = -280.7992160.$$

On the other hand, by calculating the real log canonical threshold of the polynomial ideal $\langle p(\omega) - q \rangle$, we find that the learning coefficient of the model at the ML distribution q is $(\lambda, \theta) = (7/2, 1)$. This gives us the approximation

$$(\log \mathcal{I})_{\text{RLCT}} \approx -275.9164140$$

which is closer than the BIC to the exact value of $\log \mathcal{I}$.

Our proposal to use the ML distribution as the true distribution q is admittedly simplistic, given that noise in the data will almost surely bring us to some $q \in S_2^*$. Nonetheless, our next theorem proves that the learning coefficient is always smaller than the $(9/2, 1)$ prescribed by the BIC. For deeper statistical discussions, the reader should turn to Drton and Plummer [8] where they addressed the paradox of circular reasoning in requiring true parameter values for the asymptotic approximation of Bayesian integrals. They also proposed a novel algorithm where the marginal likelihood is estimated as a weighted average of the contributions from all true distributions. We hope that mathematical analyses such as our next theorem will help inform these kinds of discussions, and provide useful estimates and bounds for a variety of statistical computations.

Theorem 10. *The learning coefficient (λ, θ) of the model at $q > 0$ is given by*

$$(\lambda, \theta) = \begin{cases} (5/2, 1) & \text{if } q \in S_1^*, \\ (7/2, 1) & \text{if } q \in S_2^*. \end{cases}$$

Therefore, asymptotically as $N \rightarrow \infty$,

$$\log Z(N) = N \sum_{i,j} \hat{q}_{ij} \log q_{ij} - \lambda \log N + (\theta - 1) \log \log N + \eta_N$$

where \hat{q} is the matrix of relative frequencies of the data and η_N is a random variable whose expectation $\mathbb{E}[\eta_N]$ converges to a constant.

We postpone the proof of this theorem to the end of the section. Let us begin with a few remarks about our approach to this problem. Firstly, Theorem 2 states that the learning coefficient (λ, θ) of the statistical model is given by

$$(2\lambda, \theta) = \min_{\omega^* \in \mathcal{V}} \text{RLCT}_{\Omega_{\omega^*}} \langle p(\omega) - q \rangle$$

where \mathcal{V} is the fiber $p^{-1}(q) = \{\omega \in \Omega : p(\omega) = q\}$ over q . Instead of focusing on a fixed q and its fiber \mathcal{V} , let us vary the parameter ω^* over all of Ω . For each $\omega^* \in \Omega$, we translate Ω so that ω^* is the origin and compute the RLCT of the ideal $\langle p(\omega + \omega^*) - p(\omega^*) \rangle$. This is the content of Proposition 13. The proof of Theorem 10 will then consist of minimizing these RLCTs over the fiber \mathcal{V} for each q in the model.

Secondly, in our computations, we will often be choosing different generators for our ideal and making appropriate changes of variables. Generators with few terms and small total degree are often highly desired. Another useful trick is to multiply or divide the generators by functions $f(\omega)$ satisfying $f(0) \neq 0$. Such functions are units in the ring \mathcal{A}_0 of real analytic functions so this multiplication or division will not change the ideal.

We will perform many of the computations by hand to demonstrate how the various properties from Section 3 can be applied. At points in the proof where RLCTs of monomial ideals are required, the `Singular` library from Section 1 comes in useful. We hope that some day the computation of learning coefficients for statistical models will be automated.

Thirdly, for the full proof of Proposition 13, we will have to analyze interactions between the model singularities and the boundary of the parameter space. Some of these interactions are messy. To improve the readability of the paper, we moved the detailed proof to the appendix while retaining some interesting computations in this section.

Finally, we come to our main proposition. Let us define the following subsets of Ω . These subsets stratify Ω according to the real log canonical threshold in the manner described in Conjecture 5.

$$\begin{aligned}
 \Omega_u &= \{\omega^* \in \Omega : t^* \in \{0, 1\}\} \\
 \Omega_m &= \{\omega^* \in \Omega : t^* \notin \{0, 1\}\} \\
 \Omega_{m0} &= \{\omega^* \in \Omega_m : a^* = c^*, b^* = d^*\} \\
 \Omega_{m0kl} &= \{\omega^* \in \Omega_{m0} : \#\{i : a_i^* = 0\} = k, \#\{i : b_i^* = 0\} = l\} \\
 \Omega_{m1} &= \{\omega^* \in \Omega_m : (b^* \neq d^*, a^* = c^*) \text{ or } (a^* \neq c^*, b^* = d^*)\} \\
 \Omega_{m10} &= \{\omega^* \in \Omega_{m1} : (a^* = c^*, \exists i a_i^* = 0) \text{ or } (b^* = d^*, \exists i b_i^* = 0)\} \\
 \Omega_{m2} &= \{\omega^* \in \Omega_m : a^* \neq c^*, b^* \neq d^*\} \\
 \Omega_{m2ad} &= \{\omega^* \in \Omega_{m2} : \exists i, j a_i^* = d_j^* = 0, c_i^* \neq 0, b_j^* \neq 0\} \\
 \Omega_{m2bc} &= \{\omega^* \in \Omega_{m2} : \exists i, j b_i^* = c_j^* = 0, d_i^* \neq 0, a_j^* \neq 0\} \\
 \Omega_{m21} &= \Omega_{m2ad} \cup \Omega_{m2bc} \\
 \Omega_{m22} &= \Omega_{m2ad} \cap \Omega_{m2bc}.
 \end{aligned}$$

Proposition 13. *Given $\omega^* \in \Omega$, let I be the ideal $\langle p(\omega + \omega^*) - p(\omega^*) \rangle$. Then,*

$$\text{RLCT}_0 I = \begin{cases} (5, 1) & \text{if } \omega^* \in \Omega_u, \\ (6, 2) & \text{if } \omega^* \in \Omega_{m000}, \\ (6, 1) & \text{if } \omega^* \in \Omega_{m010} \cup \Omega_{m001} \cup \Omega_{m020} \cup \Omega_{m002}, \\ (7, 2) & \text{if } \omega^* \in \Omega_{m011}, \\ (7, 1) & \text{if } \omega^* \in \Omega_{m012} \cup \Omega_{m021}, \\ (8, 1) & \text{if } \omega^* \in \Omega_{m022}, \\ (6, 1) & \text{if } \omega^* \in \Omega_{m1} \setminus \Omega_{m10}, \\ (7, 1) & \text{if } \omega^* \in \Omega_{m10}, \\ (7, 1) & \text{if } \omega^* \in \Omega_{m2} \setminus \Omega_{m21}, \\ (8, 1) & \text{if } \omega^* \in \Omega_{m21} \setminus \Omega_{m22}, \\ (9, 1) & \text{if } \omega^* \in \Omega_{m22}. \end{cases}$$

Proof Idea. We give a shortened analysis that ignores the effect of the boundary of Ω on the RLCTs. The derived RLCTs will be smaller than the actual ones by Proposition 4. A full proof involving boundary effects is given in the appendix.

Our ideal I is generated by $g_{ij} = f_{ij}(\omega + \omega^*) - f_{ij}(\omega^*)$ where

$$f_{ij} = ta_i b_j + (1 - t)c_i d_j, \quad i, j \in \{0, 1, 2\}$$

and $a_0 = b_0 = c_0 = d_0 = 1$. One can check that I is also generated by $g_{10}, g_{20}, g_{01}, g_{02}$,

and $g_{ij} - (d_j + d_j^*)g_{i0} - (a_i + a_i^*)g_{0j}, i, j \in \{1, 2\}$ which expand to give

$$\begin{aligned} & c_1(t_1^* - t) + a_1(t_0^* + t) + tu_1^* \\ & c_2(t_1^* - t) + a_2(t_0^* + t) + tu_2^* \\ & d_1(t_1^* - t) + b_1(t_0^* + t) + tv_1^* \\ & d_2(t_1^* - t) + b_2(t_0^* + t) + tv_2^* \\ & a_1d_1 - a_1t_0^*v_1^* + d_1t_1^*u_1^* \\ & a_1d_2 - a_1t_0^*v_2^* + d_2t_1^*u_1^* \\ & a_2d_1 - a_2t_0^*v_1^* + d_1t_1^*u_2^* \\ & a_2d_2 - a_2t_0^*v_2^* + d_2t_1^*u_2^* \end{aligned}$$

where $t_0^* = t^*, t_1^* = 1 - t^*, u_i^* = a_i^* - c_i^*, v_i^* = b_i^* - d_i^*$. Note that $\sum(a_i + a_i^*) = 1$ and $\sum a_i^* = 1$ so $\sum a_i = 0$ and similarly for b, c, d . Also, $\sum u_i^* = \sum a_i^* - c_i^* = 0$. The same is true for v^* . We now do a case-by-case analysis.

Case 1: $\omega^* \in \Omega_m$. This implies $t_0^* \neq 0$ and $t_1^* \neq 0$. Since the indeterminates b_1, b_2, c_1, c_2 appear only in the first four polynomials, this suggests the change of variables

$$\begin{aligned} c_i &= (c'_i - tu_i^* - a_i(t_0^* + t))/(t_1^* - t), \quad i = 1, 2 \\ b_i &= (b'_i - tv_i^* - d_i(t_1^* - t))/(t_0^* + t), \quad i = 1, 2 \end{aligned}$$

with new indeterminates $t, a_1, a_2, b'_1, b'_2, c'_1, c'_2, d_1, d_2$. In view of Proposition 8, the Jacobian determinant of this substitution is a constant, while the pullback ideal can be written as $I_1 + I_2$ where $I_1 = \langle b'_1, b'_2, c'_1, c'_2 \rangle$ and I_2 is generated by

$$\begin{aligned} & a_1d_1 - a_1t_0^*v_1^* + d_1t_1^*u_1^*, \\ & a_1d_2 - a_1t_0^*v_2^* + d_2t_1^*u_1^*, \\ & a_2d_1 - a_2t_0^*v_1^* + d_1t_1^*u_2^*, \\ & a_2d_2 - a_2t_0^*v_2^* + d_2t_1^*u_2^*. \end{aligned}$$

The indeterminates in I_1 and I_2 are disjoint, so we may apply Proposition 7. The RLCT of I_1 is $(4, 1)$. Now, we focus on computing the RLCT of I_2 .

Case 1.1: $\omega^* \in \Omega_{m1}$. This implies $u^* \neq 0, v^* = 0$ or $u^* = 0, v^* \neq 0$. Since $u_1^* + u_2^* + u_3^* = 0$, we assume without loss of generality that $v^* = 0, u_1^* \neq 0, u_2^* \neq 0$ and substitute

$$d_i = (d'_i + a_1t_0^*v_i^*)/(t_1^*u_1^* + a_1), \quad i = 1, 2.$$

The resulting pullback of I_2 is $\langle d'_1, d'_2 \rangle$. If ω^* lies in the interior of Ω , we use either Newton polyhedra or Proposition 7 to show that the RLCT of this monomial ideal is $(2, 1)$. If ω^* lies on the boundary of Ω , the situation is more complicated and we analyze it in greater detail in the appendix.

Case 1.2: $\omega^* \in \Omega_{m2}$. This implies $u^* \neq 0, v^* \neq 0$. Without loss of generality, suppose that $u_1^* \neq 0$. If $\omega^* \in \Omega_{m21}$, we further assume that $a_1^* = d_1^* = 0, u_1^* \neq 0, v_1^* \neq 0$. Substituting

$$\begin{aligned} d_i &= (d'_i + a_1t_0^*v_i^*)/(a_1 + t_1^*u_1^*), \quad i = 1, 2 \\ a_2 &= (a'_2 + a_1u_2^*)/u_1^*, \end{aligned}$$

the pullback ideal is $\langle a'_2, d'_1, d'_2 \rangle$ so the RLCT at an interior point is $(3, 1)$.

Case 1.3: $\omega^* \in \Omega_{m0}$. This implies $u_i^* = v_i^* = 0$ for all i . The pullback ideal is

$$\langle a_1, a_2 \rangle \langle d_1, d_2 \rangle$$

whose RLCT over an interior point of Ω is $(2, 2)$ by Proposition 7.

Case 2: $\omega^* \in \Omega_u$. Without loss of generality, assume $t^* = 0$ and substitute

$$\begin{aligned} c_i &= (c'_i - t(a_i + u_i^*)) / (1 - t) & i = 1, 2 \\ d_i &= (d'_i - t(b_i + v_i^*)) / (1 - t) & i = 1, 2. \end{aligned}$$

The pullback ideal is the sum of $\langle c'_1, c'_2, d'_1, d'_2 \rangle$ and

$$\langle t \rangle \langle a_1 + u_1^*, a_2 + u_2^* \rangle \langle b_1 + v_1^*, b_2 + v_2^* \rangle.$$

The RLCT of the first summand is $(4, 1)$. The RLCT of $\langle t \rangle$ is $(1, 1)$ while that of $\langle a_1 + u_1^*, a_2 + u_2^* \rangle$ and $\langle b_1 + v_1^*, b_2 + v_2^* \rangle$ are at least $(2, 1)$ each. By Proposition 7, the RLCT of their product is $(1, 1)$ and that of the pullback ideal is $(5, 1)$.

Proof of Theorem 10. Given a matrix $q = (q_{ij})$, the learning coefficient (λ, θ) of the model at q is the minimum of RLCTs at points $\omega^* \in \Omega$ where $p(\omega^*) = q$. The theorem then follows from Proposition 13, Theorem 1 and the claims

$$p(\Omega_u) = S_1, \quad p(\Omega_{m0}) \subset S_1, \quad p(\Omega_{m1}) \subset S_1, \quad p(\Omega_{m21}) \notin S_2^*.$$

These four claims are easy to check from the definitions of the subsets of Ω .

Acknowledgements

The author wishes to thank Christine Berkesch, Mathias Drton, Anton Leykin, Bernd Sturmfels, Zach Teitler, Sumio Watanabe and Piotr Zwiernik, as well as the anonymous reviewers for the useful discussions, corrections and advice.

References

- [1] Miki Aoyagi and Sumio Watanabe. Stochastic complexities of reduced rank regression in Bayesian estimation. *Neural Netw.*, 18(7):924–933, 2005.
- [2] V. I. Arnol'd, S. M. Guseĭn-Zade, and A. N. Varchenko. *Singularities of differentiable maps. Vol. II*, volume 83 of *Monographs in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1988.
- [3] Jacqueline Bertrand, Pierre Bertrand, and Jean-Philippe Ovarlez. The mellin transform. In Alexander D. Poularikas, editor, *Transforms and applications handbook*, The Electrical Engineering Handbook Series, chapter 12, pages xii+898. CRC Press, Boca Raton, FL, third edition, 2010.

- [4] Edward Bierstone and Pierre D. Milman. Resolution of singularities. In *Several complex variables*, volume 37 of *Math. Sci. Res. Inst. Publ.*, pages 43–78. Cambridge Univ. Press, Cambridge, 1999.
- [5] Carles Bivià-Ausina. Nondegenerate ideals in formal power series rings. *Rocky Mountain J. Math.*, 34(2):495–511, 2004.
- [6] Manuel Blickle and Robert Lazarsfeld. An informal introduction to multiplier ideals. In *Trends in commutative algebra*, volume 51 of *Math. Sci. Res. Inst. Publ.*, pages 87–114. Cambridge Univ. Press, Cambridge, 2004.
- [7] Ana María Bravo, Santiago Encinas, and Orlando Villamayor U. A simplified proof of desingularization and applications. *Rev. Mat. Iberoamericana*, 21(2):349–458, 2005.
- [8] Mathias Drton and Martyn Plummer. A bayesian information criterion for singular models. *Journal of the Royal Statistical Society: Series B*, 79:1–38, 2017.
- [9] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on algebraic statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel, 2009.
- [10] David Eisenbud. *Commutative algebra*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. With a view toward algebraic geometry.
- [11] Michael J. Evans, Zvi Gilula, and Irwin Guttman. Latent class analysis of two-way contingency tables by Bayesian methods. *Biometrika*, 76(3):557–563, 1989.
- [12] William Fulton. *Introduction to toric varieties*, volume 131 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1993.
- [13] Michael Greenblatt. An elementary coordinate-dependent local resolution of singularities and applications. *J. Funct. Anal.*, 255(8):1957–1994, 2008.
- [14] Michael Greenblatt. Resolution of singularities, asymptotic expansions of integrals and related phenomena. *J. Anal. Math.*, 111:221–245, 2010.
- [15] Heisuke Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II. *Ann. of Math. (2)*, 79:109–326, 1964.
- [16] J. A. Howald. Multiplier ideals of monomial ideals. *Trans. Amer. Math. Soc.*, 353(7):2665–2671, 2001.
- [17] János Kollár. Singularities of pairs. In *Algebraic geometry—Santa Cruz 1995*, volume 62 of *Proc. Sympos. Pure Math.*, pages 221–287. Amer. Math. Soc., Providence, RI, 1997.
- [18] János Kollár. *Lectures on resolution of singularities*, volume 166 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2007.

- [19] Robert Lazarsfeld. *Positivity in algebraic geometry. I*, volume 48 of *A Series of Modern Surveys in Mathematics*. Springer-Verlag, Berlin, 2004. Classical setting: line bundles and linear series.
- [20] Shaowei Lin, Bernd Sturmfels, and Zhiqiang Xu. Marginal likelihood integrals for mixtures of independence models. *J. Mach. Learn. Res.*, 10:1611–1631, 2009.
- [21] Dmitry Rusakov and Dan Geiger. Asymptotic model selection for naive Bayesian networks. *J. Mach. Learn. Res.*, 6:1–35, 2005.
- [22] Marcelo J. Saia. The integral closure of ideals and the Newton filtration. *J. Algebraic Geom.*, 5(1):1–11, 1996.
- [23] Morihiko Saito. On real log canonical thresholds. *arXiv:0707.2308*, 2007.
- [24] Alexander N Varchenko. Newton polyhedra and estimation of oscillating integrals. *Functional analysis and its applications*, 10(3):175–196, 1976.
- [25] V. A. Vasil’ev. Asymptotic behavior of exponential integrals in the complex domain. *Funktsional. Anal. i Prilozhen.*, 13(4):1–12, 96, 1979.
- [26] Sumio Watanabe. Algebraic analysis for nonidentifiable learning machines. *Neural Comput.*, 13(4):899–933, 2001.
- [27] Sumio Watanabe. *Algebraic geometry and statistical learning theory*, volume 25 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2009.
- [28] Keisuke Yamazaki and Sumio Watanabe. Singularities in mixture models and upper bounds of stochastic complexity. *Neural Netw.*, 16(7):1029–1038, 2003.
- [29] Keisuke Yamazaki and Sumio Watanabe. Newton diagram and stochastic complexity in mixture of binomial distributions. In *Algorithmic learning theory. 15th international conference, ALT 2004, Padova, Italy, October 2–5, 2004. Proceedings.*, pages 350–364. Berlin: Springer, 2004.

Appendices

In this section, we give a full proof of Proposition 13 that considers the effect of the boundary of the parameter space Ω on the RLCTs. The next lemma comes in handy in dealing with boundary issues. It helps us in computing the RLCTs of monomial ideals at boundary points where the parameter space contains a nice neighborhood $\Omega_1 \times \Omega_2$. Here, Ω_1 is an orthant in the coordinates involved in the monomials of I , while Ω_2 is a small cone in the remaining coordinates.

Lemma 3. Let $\Omega \subset \{(x_1, \dots, x_d) \in \mathbb{R}^d\}$ be semianalytic. Let I be a monomial ideal and φ a monomial function in x_1, \dots, x_r . If there exists a vector $\xi \in \mathbb{R}^{d-r}$ such that $\Omega_1 \times \Omega_2 \subset \Omega$ for sufficiently small ε ,

$$\begin{aligned}\Omega_1 &= \{(x_1, \dots, x_r) \in [0, \varepsilon]^r\} \\ \Omega_2 &= \{(x_{r+1}, \dots, x_d) = t(\xi + \xi') \text{ for } t \in [0, \varepsilon], \xi' \in [-\varepsilon, \varepsilon]^{d-r}\},\end{aligned}$$

then $\text{RLCT}_{\Omega_0}(I; \varphi) = \text{RLCT}_0(I; \varphi)$.

Proof. Because I and $|\varphi|$ remain unchanged by the flipping of signs of x_1, \dots, x_r , their threshold does not depend on the choice of orthant, so $\text{RLCT}_{\Omega_1}(I; \varphi) = \text{RLCT}_0(I; \varphi)$. The lemma now follows from Proposition 7 and the fact that the threshold of the zero ideal over the cone neighborhood Ω_2 is $(\infty, -)$.

Detailed Proof of Proposition 13. Recall that the ideal I is generated by

$$\begin{aligned}c_1(t_1^* - t) + a_1(t_0^* + t) + tu_1^* \\ c_2(t_1^* - t) + a_2(t_0^* + t) + tu_2^* \\ d_1(t_1^* - t) + b_1(t_0^* + t) + tv_1^* \\ d_2(t_1^* - t) + b_2(t_0^* + t) + tv_2^* \\ a_1d_1 - a_1t_0^*v_1^* + d_1t_1^*u_1^* \\ a_1d_2 - a_1t_0^*v_2^* + d_2t_1^*u_1^* \\ a_2d_1 - a_2t_0^*v_1^* + d_1t_1^*u_2^* \\ a_2d_2 - a_2t_0^*v_2^* + d_2t_1^*u_2^*\end{aligned}$$

We do a case-by-case analysis of the structure of I and the boundary of Ω .

Case 1: $\omega^* \in \Omega_m$. This implies $t_0^* \neq 0$ and $t_1^* \neq 0$. Since the indeterminates b_1, b_2, c_1, c_2 appear only in the first four polynomials, this suggests the change of variables

$$\begin{aligned}c_i &= (c'_i - tu_i^* - a_i(t_0^* + t))/(t_1^* - t), \quad i = 1, 2 \\ b_i &= (b'_i - tv_i^* - d_i(t_1^* - t))/(t_0^* + t), \quad i = 1, 2\end{aligned}$$

with new indeterminates $t, a_1, a_2, b'_1, b'_2, c'_1, c'_2, d_1, d_2$. In view of Proposition 8, the Jacobian determinant of this substitution is a constant.

Case 1.1: $\omega^* \in \Omega_{m1}$. This implies $u^* \neq 0, v^* = 0$ or $u^* = 0, v^* \neq 0$. Without loss of generality, we assume $v^* = 0, u_1^* > 0$ and substitute

$$d_i = (d'_i + a_1t_0^*v_i^*)/(t_1^*u_1^* + a_1), \quad i = 1, 2.$$

The resulting pullback ideal is $\langle b'_1, b'_2, c'_1, c'_2, d'_1, d'_2 \rangle$. If ω^* lies in the interior of Ω , we use either Newton polyhedra or Proposition 7 to show that the RLCT of this monomial ideal is $(6, 1)$. If ω^* lies on the boundary of Ω , the situation is more complicated. Since we are considering a subset of a neighborhood of ω^* , the corresponding Laplace integral from Proposition 3a is smaller so the threshold is at least $(6, 1)$. To compute it exactly, we need blowups to separate the coordinate hyperplanes and the boundary hypersurfaces.

Because $-u_1^* = u_2^* + u_3^*$, we cannot have $u_2^* = u_3^* = 0$. Suppose $u_2^* \neq 0$ and $u_3^* \neq 0$. We consider a blowup where one of the charts is given by the monomial map $t = s, a_i = sa'_i, c'_1 = rs, c'_2 = rsc'_2, b'_i = rsb'_i, d'_i = rsd'_i$. Here, the pullback pair is $(\langle rs \rangle; r^5s^8)$. Now, we study the inequalities which are *active* at ω^* . For instance, if $b_1^* = 0$, then ω^* lies on the boundary defined by $0 \leq b_1 + b_1^*$. After the various changes of variables, the inequalities are as shown below, where $b_3^* = -b_1^* - b_2^*$ and similarly for c_3^*, d_3^* and a_3^* . Note that the inequality for $a_1^* = 0$ is omitted because $a_1^* = 0$ implies $u_1^* = -c_1^* \leq 0$. Similar conditions on the u_i^*, v_i^* hold for the other inequalities.

$$\begin{aligned}
b_i^* = 0 &: 0 \leq rs(b_i'' - d_i''(t_1^* - s)/(t_1^*u_1^* + sa'_1))/(t_0^* + s) \\
d_i^* = 0 &: 0 \leq rsd_i''/(t_1^*u_1^* + sa'_1) \\
c_1^* = 0 &: 0 \leq s(-u_1^* + a'_1(t_0^* + s) + r)/(t_1^* - s) \\
c_2^* = 0 &: 0 \leq s(-u_2^* + a'_2(t_0^* + s) + rc_2'')/(t_1^* - s) & u_2^* > 0 \\
c_3^* = 0 &: 0 \leq s(-u_3^* + a'_3(t_0^* + s) - r - rc_2'')/(t_1^* - s) & u_3^* > 0 \\
a_2^* = 0 &: 0 \leq sa'_2 & u_2^* < 0 \\
a_3^* = 0 &: 0 \leq sa'_3 & u_3^* < 0
\end{aligned}$$

In applying Lemma 3, the choice of coordinates is important. For instance, if $b_2^* = b_3^* = 0$, we choose coordinates b_2'' and b_3'' and set $b_1'' = -b_2'' - b_3''$. The same is done for the d_i'' . The pullback pair is unchanged by these choices. Now, with coordinates (r, s) and $(b_{i_1}'', b_{i_2}'', d_{j_1}'', d_{j_2}'', c_2'', a_2', a_3')$, we apply the lemma with the vector $\xi = (2, 2, u_1^*, u_1^*, 1, 1, 1)$, so the threshold is $\text{RLCT}_0(rs; r^5s^8) = (6, 1)$.

Now, if only one of u_2^*, u_3^* is zero, suppose $u_2^* = 0, u_3^* \neq 0$ without loss of generality. If $a_2^* = c_2^* \neq 0$, then the arguments of the previous paragraph show that the RLCT is again $(6, 1)$. If $a_2^* = c_2^* = 0$, we blow up the origin in \mathbb{R}^7 and consider the chart where $a_2 = s, c'_i = sc_i'', b'_i = sb_i'', d'_i = sd_i''$. The pullback pair is $(\langle sb_1'', sb_2'', sc_1'', sc_2'', sd_1'', sd_2'' \rangle; s^6)$. The active inequalities for $a_2^* = c_2^* = 0$ are

$$\begin{aligned}
c_2^* = 0 &: 0 \leq s(c_2'' - t_0^* + t)/(t_1^* - t) \\
a_2^* = 0 &: 0 \leq s.
\end{aligned}$$

Near the origin in $(s, b_1'', b_2'', c_1'', c_2'', d_1'', d_2'') \in \mathbb{R}^7$, these inequalities imply $s = 0$ so the new region \mathcal{M} defined by the active inequalities is not full at the origin. Thus, we can ignore the origin in computing the RLCT. All other points on the exceptional divisor of this blowup lie on some other chart of the blowup where the pullback pair is $(s; s^6)$, so the RLCT is at least $(7, 1)$. In the chart where $c_2 = s, c_1 = sc_1'', a_2 = sa_2', b'_i = sb_i'', d'_i = sd_i''$, we have the active inequalities below. Note that $c_3^* \neq 0$ because $u_3^* = -u_1^* < 0$.

$$\begin{aligned}
b_i^* = 0 &: 0 \leq s(b_i'' - d_i''(t_1^* - t)/(t_1^*u_1^* - (sa_2' + a_3)))/(t_0^* + t) \\
d_i^* = 0 &: 0 \leq sd_i''/(t_1^*u_1^* - (sa_2' + a_3)) \\
c_1^* = 0 &: 0 \leq (sc_1'' - tu_1^* + (sa_2' + a_3)(t_0^* + t))/(t_1^* - t) \\
c_2^* = 0 &: 0 \leq s(1 - a_2'(t_0^* + t))/(t_1^* - t) \\
a_2^* = 0 &: 0 \leq sa_2' \\
a_3^* = 0 &: 0 \leq a_3
\end{aligned}$$

Again, choosing suitable coordinates in the b_i'' and d_i'' , we find that the RLCT is (7, 1) by using Lemma 3 with $\xi = (2, 2, u_1^*, u_1^*, 1, 1, 1, -1)$ in coordinates $(b_{i_1}'', b_{i_2}'', d_{j_1}'', d_{j_2}'', a_2', a_3, c_1', t)$.

Case 1.2: $\omega^* \in \Omega_{m2}$. This implies $u^* \neq 0, v^* \neq 0$. Without loss of generality, suppose that $u_1^* \neq 0$. If $\omega^* \in \Omega_{m21}$, we further assume that $a_1^* = d_j^* = 0, u_1^* \neq 0, v_j^* \neq 0$. Substituting

$$\begin{aligned} d_i &= (d_i' + a_1 t_0^* v_i^*) / (a_1 + t_1^* u_1^*), \quad i = 1, 2 \\ a_2 &= (a_2' + a_1 u_2^*) / u_1^*, \end{aligned}$$

the pullback ideal is $\langle a_2', b_1', b_2', c_1', c_2', d_1', d_2' \rangle$ so the RLCT is at least (7, 1). Note that $a_i = (a_2' w_i^* + a_1 u_i^*) / u_1^*$ for $i = 1, 2, 3$ where $w_i^* = 0, 1, -1$ respectively. If ω^* is not in Ω_{m21} , we consider the blowup chart $a_2' = s, b_i' = s b_i'', c_i' = s c_i'', d_i' = s d_i''$. The active inequalities are as follows. The symbol $v-$ denotes $v_i^* \leq 0$.

$$\begin{aligned} b_i^* = 0 : & \quad 0 \leq [s b_i'' - t v_i^* - (s d_i'' + a_1 t_0^* v_i^*)(t_1^* - t) / (t_1^* u_1^* + a_1)] / (t_0^* + t) & v- \\ c_i^* = 0 : & \quad 0 \leq [s c_i'' - t u_i^* - (s w_i^* + a_1 u_i^*)(t_0^* + t) / u_1^*] / (t_1^* - t) & u+ \\ a_i^* = 0 : & \quad 0 \leq (s w_i^* + a_1 u_i^*) / u_1^* & u- \\ d_i^* = 0 : & \quad 0 \leq (s d_i'' + a_1 t_0^* v_i^*) / (t_1^* u_1^* + a_1) & v+ \end{aligned}$$

The crux to understanding the inequalities is this: if $a_i^* = d_j^* = 0, u_i^* \neq 0, v_j^* \neq 0$, the coefficient of a_1 appears with different signs in the inequalities for $a_i^* = 0$ and $d_j^* = 0$. This makes it difficult to choose a suitable vector ξ for Lemma 3. Similarly, if $b_i^* = c_j^* = 0, v_i^* \neq 0, u_j^* \neq 0$, the coefficient of $u_1^* t + t_0^* a_1$ appears with different signs. Fortunately, since $\omega^* \notin \Omega_{m21}$, we do not have such obstructions and it is an easy exercise to find the vector ξ . Thus, the RLCT is (7, 1).

If $\omega^* \in \Omega_{m21} \setminus \Omega_{m22}$, we blow up $a_1 = s, a_2' = s a_2'', b_i' = s b_i'', c_i' = s c_i'', d_i' = s d_i''$. The active inequalities for $a_1^* = d_j^* = 0$ imply that the new region \mathcal{M} is not full at the origin of this chart. Thus, we shift our focus to the other charts of the blowup where the pullback pair is $(s; s^7)$, so the RLCT is at least (8, 1). In the chart where $a_2' = s, a_1 = s a_1', b_i' = s b_i'', c_i' = s c_i'', d_i' = s d_i''$, we do not have obstructions coming from any $b_i^* = c_j^* = 0, v_i^* \neq 0, u_j^* \neq 0$ so it is again easy to find the vector ξ for Lemma 3. The threshold is exactly (8, 1).

If $\omega^* \in \Omega_{m22}$, consider the following two charts in the blowup of the origin in \mathbb{R}^9 .

$$\begin{aligned} \text{Chart 1: } & \quad a_1 = s, t = s t', a_2' = s a_2'', b_i' = s b_i'', c_i' = s c_i'', d_i' = s d_i'' \\ \text{Chart 2: } & \quad t = s, a_1 = s a_1', a_2' = s a_2'', b_i' = s b_i'', c_i' = s c_i'', d_i' = s d_i'' \end{aligned}$$

The inequalities for $a_i^* = d_j^* = 0, u_i^* \neq 0, v_j^* \neq 0$ and $b_i^* = c_j^* = 0, v_i^* \neq 0, u_j^* \neq 0$ imply that the new region \mathcal{M} is not full at points outside of the other seven charts, so we may ignore these two charts in computing the RLCT. Indeed, for Chart 1, the active inequalities

$$\begin{aligned} a_i^* = 0 : & \quad 0 \leq s(a_2'' w_i^* + u_i^*) / u_1^* & u- \\ d_i^* = 0 : & \quad 0 \leq s(d_i'' + t_0^* v_i^*) / (t_1^* u_1^* + s) & v+ \end{aligned}$$

tell us that a_2'' or d_2'' must be non-zero for \mathcal{M} to be full. In Chart 2, suppose \mathcal{M} is full at some point x where $a_2'' = b_1'' = b_2'' = c_1'' = c_2'' = d_1'' = d_2'' = 0$. Then,

$$\begin{aligned} a_i^* = 0 : & \quad 0 \leq s(a_2'' w_i^* + a_1' u_i^*) / u_1^* & u- \\ d_i^* = 0 : & \quad 0 \leq s(d_i'' + a_1' t_0^* v_i^*) / (t_1^* u_1^* + s a_1') & v+ \end{aligned}$$

imply that $a'_1 = 0$ at x . However, if this is the case, the inequalities

$$\begin{aligned} b_i^* = 0 &: 0 \leq s[b_i'' - v_i^* - (d_i'' + a'_1 t_0^* v_i^*)(t_1^* - s)/(t_1^* u_1^* + s a'_1)]/(t_0^* + s) & v- \\ c_i^* = 0 &: 0 \leq s[c_i'' - u_i^* - (a_2'' w_i^* + a'_1 u_i^*)(t_0^* + s)/u_1^*]/(t_1^* - s) & u+ \end{aligned}$$

forces b_i'' or c_i'' to be non-zero for some i , a contradiction. Thus, we shift our focus to the other seven charts where the pullback pair is $(s; s^8)$ and the RLCT is at least $(9, 1)$. In the chart for $a'_2 = s, a_1 = s a'_1, t = s t', b'_i = s b_i'', c'_i = s c_i'', d'_i = s d_i''$, note that we cannot have both $a_2^* = 0$ and $a_3^* = 0$ because we assumed $a_1^* = 0$. It is now easy to find the vector ξ for Lemma 3, so the threshold is $(9, 1)$.

Case 1.3: $\omega^* \in \Omega_{m0}$. This implies $u_i^* = v_i^* = 0$ for all i . The pullback ideal is

$$\langle b'_1, b'_2, c'_1, c'_2 \rangle + \langle a_1, a_2 \rangle \langle d_1, d_2 \rangle$$

whose RLCT over an interior point of Ω is $(6, 2)$ by Proposition 7. This occurs in Ω_{m000} where none of the inequalities are active. Now, suppose the only active inequalities come from $a_1^* = c_1^* = 0$. We blow up the origin in $\{(a_1, c_1) \in \mathbb{R}^2\}$. In the chart given by $a_1 = a'_1, c_1 = a'_1 c'_1$, the new region \mathcal{M} is not full at the origin, so we only need to study the chart where $c'_1 = c''_1, a_1 = c''_1 a'_1$. The pullback pair becomes $(\langle c''_1 \rangle + \langle b'_1, b'_2, c'_2 \rangle + \langle a_2 \rangle \langle d_1, d_2 \rangle; c''_1)$, and a simple application of Lemma 3 and Proposition 7 shows that the threshold is $(6, 1)$.

In this fashion, we study the different scenarios and summarize the pullback pairs and thresholds in the table below.

Inequalities	Pullback pair	RLCT
–	$(\langle b'_1, b'_2, c'_1, c'_2 \rangle + \langle a_1, a_2 \rangle \langle d_1, d_2 \rangle; 1)$	$(6, 2)$
$a_1^* = 0$	$(\langle b'_1, b'_2, c''_1, c'_2 \rangle + \langle a_2 \rangle \langle d_1, d_2 \rangle; c''_1)$	$(6, 1)$
$a_1^* = 0, b_1^* = 0$	$(\langle b''_1, b'_2, c''_1, c'_2 \rangle + \langle a_2 \rangle \langle d_2 \rangle; b''_1 c''_1)$	$(7, 2)$
$a_1^* = 1$	$(\langle b'_1, b'_2, c''_1, c''_2 \rangle; c''_1 c''_2)$	$(6, 1)$
$a_1^* = 1, b_1^* = 0$	$(\langle b''_1, b'_2, c''_1, c''_2 \rangle; b''_1 c''_1 c''_2)$	$(7, 1)$
$a_1^* = 1, b_1^* = 1$	$(\langle b''_1, b''_2, c''_1, c''_2 \rangle; b''_1 b''_2 c''_1 c''_2)$	$(8, 1)$

For example, the case $a_3^* = c_3^* = 1$ corresponds to $a_1^* = a_2^* = c_1^* = c_2^* = 0$. Here, we blow up the origins in $\{(a_1, c_1) \in \mathbb{R}^2\}$ and $\{(a_2, c_2) \in \mathbb{R}^2\}$. As before, we can ignore the other charts and just consider the one where $a_1 = c''_1 a'_1, c_1 = c''_1, a_2 = c''_2 a'_2, c_2 = c''_2$. The pullback pair is $(\langle c''_1 \rangle + \langle c''_2 \rangle + \langle b'_1, b'_2 \rangle, c''_1 c''_2)$. If $b_i^* \neq 0$ for all i , the RLCT is $(6, 1)$ by Lemma 3 and Proposition 7.

Case 2: $\omega^* \in \Omega_u$. Without loss of generality, assume $t^* = 0$ and substitute

$$\begin{aligned} c_i &= (c'_i - t(a_i + u_i^*)) / (1 - t) & i = 1, 2 \\ d_i &= (d'_i - t(b_i + v_i^*)) / (1 - t) & i = 1, 2. \end{aligned}$$

The pullback ideal is the sum of $\langle c'_1, c'_2, d'_1, d'_2 \rangle$ and $\langle t \rangle \langle a_1 + u_1^*, a_2 + u_2^* \rangle \langle b_1 + v_1^*, b_2 + v_2^* \rangle$. Since $c'_3 = -c'_1 - c'_2$ and similarly for the d'_i, a_i, b_i, u_i^* and v_i^* , it is useful to write this ideal

more symmetrically as the sum of $\langle c'_1, c'_2, c'_3 \rangle$, $\langle d'_1, d'_2, d'_3 \rangle$ and

$$\langle t \rangle \langle a_1 + u_1^*, a_2 + u_2^*, a_3 + u_3^* \rangle \langle b_1 + v_1^*, b_2 + v_2^*, b_3 + v_3^* \rangle.$$

Meanwhile, the inequalities are

$$\begin{aligned} a_i^* = 0 &: 0 \leq a_i \\ c_i^* = 0 &: 0 \leq (c'_i - t(a_i + u_i^*)) / (1 - t) \quad u_i^* \geq 0 \\ b_j^* = 0 &: 0 \leq b_j \\ d_j^* = 0 &: 0 \leq (d'_j - t(b_j + v_j^*)) / (1 - t) \quad v_j^* \geq 0. \end{aligned}$$

We now relabel the indices of the a_i and c'_i , without changing the b_j and d'_j , so that the active inequalities are among those from $a_1^* = 0, a_2^* = 0, c_{i_1}^* = 0, c_{i_2}^* = 0$. The b_j and d'_j are thereafter also relabeled so that the inequalities come from $b_1^* = 0, b_2^* = 0, d_{j_1}^* = 0, d_{j_2}^* = 0$. We claim that the new region \mathcal{M} contains, for small ε , the orthant neighborhood

$$\{(a_1, a_2, b_1, b_2, c_{i_1}, c_{i_2}, d_{j_1}, d_{j_2}, -t) \in [0, \varepsilon]^9\}.$$

Indeed, the only problematic inequalities are

$$\begin{aligned} c_3^* = 0 &: 0 \leq (c'_3 - t(-a_1 - a_2 + u_i^*)) / (1 - t) \quad u_3^* = 0 \\ d_3^* = 0 &: 0 \leq (d'_3 - t(-b_1 - b_2 + v_j^*)) / (1 - t) \quad v_3^* = 0. \end{aligned}$$

However, these inequalities cannot occur because for instance, $u_3^* = 0$ and $c_3^* = 0$ implies $a_3^* = 0$, a contradiction since the a_i were relabeled to avoid this. Finally, the threshold of $\langle t \rangle$ is $(1, 1)$ while that of $\langle a_1 + u_1^*, a_2 + u_2^* \rangle$ and $\langle b_1 + v_1^*, b_2 + v_2^* \rangle$ are at least $(2, 1)$ each. By Proposition 7, the RLCT of their product is $(1, 1)$ and that of the pullback ideal we were originally interested in is $(5, 1)$.