Useful Facts about RLCT

Shaowei Lin

November 25, 2012

1 Definitions

In this paper, we will be working with rings and ideals of real analytic functions. Given $x \in \mathbb{R}^d$, let $\mathcal{A}_x$ be the ring of real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$ that are analytic at x. When $x = 0$ is the origin, it is useful to think of $\mathcal{A}_0$ as a subring of the formal power series ring $\mathbb{R}[[\omega_1, \ldots, \omega_d]] = \mathbb{R}[[\omega]]$ which consists of power series which are convergent in some neighborhood of the origin. For all $x \in \mathbb{R}^d$, $\mathcal{A}_x$ is isomorphic to $\mathcal{A}_0$ via a translation. Given a subset $\Omega \subset \mathbb{R}^d$, let $\mathcal{A}_\Omega$ denote 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 $f \in \mathcal{A}_\Omega$, we 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 $\mathcal{V}_\Omega(I)$ to be the intersection of $\mathcal{V}_\Omega(f)$ over all $f \in I$. Let ∇f denote the gradient of f and $\nabla^2 f$ its Hessian. Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, we write $A \succ 0$ if A is positive definite, and $A \succeq 0$ if A is positive semidefinite.

Let Ω be a compact subset of $\mathbb{R}^d$. Let us assume that Ω is *semianalytic*, i.e. $\Omega = \{x \in \mathbb{R}^d : g_1(x) \geq 0, \ldots, g_l(x) \geq 0\}$ is defined by real analytic inequalities. Here, the functions $g_i(x)$ only have to be real analytic at points on the boundary where they are active. Let $I = \langle f_1, \ldots, f_r\rangle$ be the ideal generated by functions $f_1, \ldots, f_r$ in the ring $\mathcal{A}_\Omega$. Let φ be *nearly analytic*, i.e. φ is a product $\varphi_a\varphi_s$ of functions where φ_a is real analytic on Ω and φ_s is a smooth and positive on Ω.

Definition 1.1 ([4],[6, §7.1])**.** The following definitions of the real log canonical threshold $\mathrm{RLCT}_\Omega(I;\varphi) = (\lambda, \theta)$ are equivalent.

a. The Laplace integral

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

is asymptotically $CN^{-\lambda/2}(\log N)^{\theta-1}$ for some constant C.

b. The zeta function

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

has a smallest pole λ of multiplicity θ.

c. The volume function

$$V(t) = \int_{\sum_{i=1}^{r} f_i(\omega)^2 \leq t} \varphi(\omega) d\omega$$

is asymptotically $C\, t^{\lambda/2}(-\log t)^{\theta-1}$ for some constant C.

One can show that these definitions are independent of the choice of generators $f_1, \ldots, f_r$. If the variety $\mathcal{V}_\Omega(I)$ is empty, we set $\lambda = \infty$ and leave θ undefined. We define the real log canonical threshold $\mathrm{RLCT}_\Omega(f; \varphi)$ of a real analytic function f to be that of the principal ideal $\langle f \rangle$. □

Two pairs are ordered $(\lambda_1, \theta_1) > (\lambda_2, \theta_2)$ if $\lambda_1 > \lambda_2$, or $\lambda_1 = \lambda_2$ and $\theta_1 < \theta_2$. For $x \in \Omega$, we define $\mathrm{RLCT}_{\Omega_x}(I; \varphi)$ to be the threshold for a sufficiently small neighborhood Ω_x of x in Ω. Let $\mathrm{RLCT}_x(I; \varphi)$ denote the threshold at x in the absence of boundary conditions. We list some basic properties of RLCTs.

Proposition 1.2 ([4])**.** *Let* $(\lambda, \theta) = \mathrm{RLCT}_\Omega(I; \varphi)$.

a. If $\mathcal{V}_\Omega(I)$ *is not empty, then* $(\lambda, \theta) \in \mathbb{Q} \times \mathbb{Z}$ *with* $0 < \lambda \leq d$ *and* $0 < \theta \leq d$.

b. The pair (λ, θ) *is the minimum*

$$\min_{x \in \Omega} \mathrm{RLCT}_{\Omega_x}(I; \varphi).$$

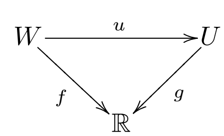
In fact, it is enough to vary x *over* $\mathcal{V}_\Omega(I)$.

c. If $\varphi = \varphi_a \varphi_s$ *as before, then* $\mathrm{RLCT}_\Omega(I; \varphi) = \mathrm{RLCT}_\Omega(I; \varphi_a)$.

d. Let x *be a boundary point of* $\Omega \subset \mathbb{R}^d$*. Then,*

$$\mathrm{RLCT}_x(f; \varphi) \leq \mathrm{RLCT}_{\Omega_x}(f; \varphi).$$

Proposition 1.3 ([5])**.** *Let the minimum of a real analytic function* $f : \Omega \to \mathbb{R}$ *be* f^* *and let* $W \subset \Omega$ *be a neighborhood of the minimum locus* argmin f*. Suppose the restriction of* f *to* W *is the composition of real analytic maps*



where $U \subset \mathbb{R}^k$ *is a neighborhood of a point* u^* *satisfying* $g(u^*) = f^*$, $\nabla g(u^*) = 0$ *and* $\nabla^2 g(u^*) \succ 0$*. If* $u = (u_1, \ldots, u_k)$ *and* $\mathrm{RLCT}_\Omega(f - f^*; \varphi) = (\lambda, \theta)$*, then*

$$(2\lambda, \theta) = \mathrm{RLCT}_W(I; \varphi)$$

where I *is the ideal* $\langle u_1 - u_1^*, \ldots, u_k - u_k^* \rangle$.

2 Statistical Models

Let X be a random variable with state space $\mathcal{X}$ and Δ be the space of probability distributions on $\mathcal{X}$. Let $\mathcal{M}$ be a statistical model parametrized by a real analytic map $p : \Omega \to \Delta$ where Ω is a compact semianalytic subset of $\mathbb{R}^d$. For each $\omega \in \Omega$, we denote the corresponding probability density function by $p_x(\omega), x \in \mathcal{X}$. Let $\varphi : \Omega \to \mathbb{R}$ 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. We consider the prior on Ω defined by $|\varphi|$. Given a distribution $q \in \Delta$, we study the log loss function

$$\ell_q(\omega) = -\int_{\mathcal{X}} q_x \log p_x(\omega) dx.$$

Definition 2.1. The *learning coefficient* (λ_q, θ_q) is the pair of exponents coming from the asymptotics of the Laplace integral

$$Z(N) = \int_{\Omega} e^{-N\ell_q(\omega)} |\varphi(\omega)| d\omega \approx C e^{-N\ell^*} N^{-\lambda_q} (\log N)^{\theta_q - 1}$$

where ℓ^* is the minimum of ℓ_q over Ω and $C > 0$ is some constant. Consequently, (λ_q, θ_q) is the real log canonical threshold $\mathrm{RLCT}_\Omega(\ell_q - \ell^*; \varphi)$. We call the set of points $\{\omega \in \Omega : \ell_q(\omega) = \ell^*\}$ the *q-locus* of the model. □

When q is the empirical distribution, the log loss function is the negative log likelihood and the q-locus is the set of maximum likelihood estimates. If q is in the model $\mathcal{M}$, then the q-locus is just the fiber $p^{-1}(q)$ of the map p over q.

We say that $\mathcal{M}$ is *regular* at $q \in \Delta$ if the log loss function $\ell_q(\omega)$ is uniquely minimized at some $\omega^* \in \Omega$ and if the Hessian $\nabla^2 \ell_q(\omega^*)$ is positive definite.

Proposition 2.2. *Suppose the map $p : \Omega \to \Delta$ of the model $\mathcal{M}$ can be expressed as the composition of real analytic maps*

$$\begin{array}{ccc} \Omega & \xrightarrow{\;u\;} & \Upsilon \\ & {\scriptstyle p}\searrow \qquad \swarrow{\scriptstyle \tilde{p}} & \\ & \Delta & \end{array}$$

with $\Upsilon \subset \mathbb{R}^k$ and $u = (u_1, \ldots, u_k)$. If the model parametrized by $\tilde{p}$ is regular at $q \in \Delta$ with q-locus $\{u^*\}$ and g maps the q-locus of $\mathcal{M}$ onto u^*, then the learning coefficient (λ_q, θ_q) of $\mathcal{M}$ at q is given by

$$(2\lambda_q, \theta_q) = \mathrm{RLCT}_\Omega(I; \varphi)$$

where I is the fiber ideal $\langle u_1 - u_1^*, \ldots, u_k - u_k^* \rangle$.

Let $X_1, \ldots, X_N$ be independent and identically distributed samples of X. The integrated likelihood of the data is the random variable

$$Z_N = \int_{\Omega} \prod_{i=1}^{N} p_{X_i}(\omega)\, |\varphi(\omega)| d\omega.$$

Note that Z_N is the Laplace integral $Z(N)$ when q is the empirical distribution. Watanabe clarified the asymptotic behavior of Z_N for large samples.

Theorem 2.3 ([6, §6])**.** *Suppose the data is sampled from a distribution $q \in \Delta$, and let Ω, p, φ, q satisfy some mild conditions listed below. Then,*

$$\log Z_N \;=\; \sum_{i=1}^{N} \log q_{X_i} - \lambda_q \log N + (\theta_q - 1) \log\log N + R_N$$

where (λ_q, θ_q) is the learning coefficient of the model at q and the random variable R_N is asymptotically of constant order.

Watanabe refers to the restrictions on Ω, p, φ, q for the above theorem as the Fundamental Conditions [6]. We list these conditions here for convenience. For further discussions on the implications of these conditions and how they can be relaxed, see §6.1, §6.2 and §7.8 of [6].

1. The parameter space Ω is compact and semianalytic, while the prior φ is nearly analytic.
2. The true distribution q lies in the model $\mathcal{M}$.
3. The distributions q and p have the same support, i.e. for any $\omega \in \Omega$, the sets $\{x \in \mathcal{X} : p_x(\omega) > 0\}$ and $\{x \in \mathcal{X} : q_x > 0\}$ are equal up to closure.
4. The map $f_x(\omega) = \log(q_x/p_x(\omega))$ extends to a complex analytic map $f^{\mathbb{C}}$ on an open set $\Omega^{\mathbb{C}} \subset \mathbb{C}^d$ that contains Ω.
5. If M_x is the supremum of $|f_x^{\mathbb{C}}(\omega)|$ over $\Omega^{\mathbb{C}}$, then $\int_{\mathcal{X}} M_x^2 q_x dx < \infty$.
6. There exists $\varepsilon > 0$ such that $\int_{\mathcal{X}} M_x^2 P_x dx < \infty$ where P_x is the supremum of $p_x(\omega)$ over all ω satisfying $\ell_q(\omega) - \ell^* \le \varepsilon$.

Given any statistical model, it is natural to explore the following questions. First, the log loss function ℓ_q fails to be real analytic at ω when q_x is nonzero and $p_x(\omega)$ is zero for some $x \in \mathcal{X}$. However, the integrand $e^{-N\ell_q(\omega)}$ approaches zero near such points, so the Laplace integral $Z(N)$ could still be well-defined. Moreover, the asymptotics of $Z(N)$ depends only on the behavior of ℓ_q near its locus of minimum points. This locus is far away from points where ℓ_q is non-analytic. Given these considerations, what conditions do p and q have to satisfy in order for learning coefficients to exist in Definition 2.1? Second, for which p and q does Proposition 2.2 allow us to compute learning coefficients via real log canonical thresholds of fiber ideals? Third, what do the fundamental conditions spell for p and q so that we may use Theorem 2.3 in determining the asymptotics of the integrated likelihood? We study these issues for discrete and Gaussian models.

Proposition 2.4. *Let $\mathcal{M}$ be a discrete model parametrized by a real analytic map $p = (p_1, \ldots, p_k) : \Omega \to \Delta$ and let $q \in \Delta$. If there exists $\omega \in \Omega$ such that the support of $p(\omega)$ contains that of q, then the learning coefficient (λ_q, θ_q) exists. Otherwise, the Laplace integral $Z(N)$ is identically zero so we may set $\lambda_q = \infty$.*

Now, we assume $q \in \mathcal{M}$. If all the entries of q are nonzero, then the learning coefficient (λ_q, θ_q) is half the RLCT of the fiber ideal. Otherwise,

$$(\lambda_q, \theta_q) = \mathrm{RLCT}_\Omega(\sum_{i:q_i \neq 0} (p_i(\omega) - q_i)^2 - (p_i(\omega) - q_i); \varphi).$$

Lastly, if q is the true distribution and all the entries of q are nonzero, then the learning coefficient gives the asymptotics of the log integrated likelihood.

Proof. Most of the above claims are easy to check. We will focus on the displayed formula and the last statement. Suppose some of the entries of $q \in \mathcal{M}$ are zero. Then, the q-locus is the variety $\{\omega : p(\omega) = q\}$. Using the Taylor expansion of $\log t$ near $t = 1$, we get the inequality

$$c_1(t-1)^2 \leq -\log t + t - 1 \leq c_2(t-1)^2$$

for some positive constants c_1, c_2. After substituting $t = p_i/q_i$ for each $q_i \neq 0$, multiplying by q_i and summing up over all such i, we get

$$\sum_i \frac{c_1}{q_i}(p_i(\omega) - q_i)^2 \leq \ell_q(\omega) - \ell^* + \sum_i p_i(\omega) - 1 \leq \sum_i \frac{c_2}{q_i}(p_i(\omega) - q_i)^2$$

Now, $1 - \sum_i p_i(\omega)$ is always nonnegative. Therefore,

$$\min_i(\frac{c_1}{q_i}, 1)\, g(\omega) \leq \ell_q(\omega) - \ell^* \leq \max_i(\frac{c_2}{q_i}, 1)\, g(\omega)$$

where $g(\omega)$ is the polynomial $\sum_i (p_i(\omega) - q_i)^2 - (p_i(\omega) - q_i)$. Hence, the learning coefficient is given by $\mathrm{RLCT}_\Omega(g; \varphi)$.

Now, suppose q is the true distribution. When all the entries of q are positive, we may restrict our parameter space to

$$\Omega_\varepsilon = \{\omega \in \Omega : p_i(\omega) \geq \varepsilon \text{ for all } i\}$$

For small $\varepsilon > 0$, the difference between the integrated likelihood over Ω_ε and the integral over Ω is bounded for large N. One can verify that the fundamental conditions for Theorem 2.3 are satisfied over Ω_ε. If some entries of q are zero, then the set $\{w : p(\omega)$ has the same support as $q\}$ does not form a neighborhood of the variety $\{\omega : p(\omega) = q\}$, so Theorem 2.3 cannot be applied. □

Proposition 2.5. *Let $\mathcal{M}$ be a multivariate Gaussian model whose mean $\mu(\omega)$ and covariance $\Sigma(\omega)$ are parametrized by real analytic maps. Then, the learning coefficient exists for all q whose mean μ_q and covariance Σ_q are well-defined.*

Now, we assume $q \in \mathcal{M}$. Then, the learning coefficient (λ_q, θ_q) satisfies

$$(2\lambda_q, \theta_q) = \mathrm{RLCT}_\Omega(\langle \mu(\omega) - \mu_q, \Sigma(\omega) - \Sigma_q \rangle; \varphi).$$

The learning coefficient also gives the asymptotics of the log integrated likelihood.

Proof. For d-dimensional Gaussian models, the log loss function is

$$\ell_q(\mu, \Sigma) = \frac{d}{2}\log 2\pi + \frac{1}{2}\log\det\Sigma + \frac{1}{2}\operatorname{tr}\Sigma^{-1}\mathbb{E}_q[(X-\mu)(X-\mu)^\top]$$

where $\mathbb{E}_q$ be the expectation of the random variable X over the distribution q. Hence, the learning coefficient exists for all distributions q whose mean μ_q and covariance Σ_q are well-defined. Furthermore, Gaussian models are regular at all such distributions q, so the learning coefficients are given by half the RLCTs of the fiber ideals. Finally, it is easy to check that all the fundamental conditions for Theorem 2.3 is satisfied for all $q \in \mathcal{M}$. □

3 Disjoint Variables and Birational Maps

For the next few results, let $f_1, \ldots, f_r \in \mathcal{A}_X$ and $g_1, \ldots, g_s \in \mathcal{A}_Y$ 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 $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$. Let $\varphi_x : X \to \mathbb{R}$ and $\varphi_y : Y \to \mathbb{R}$ be nearly analytic. Define $(\lambda_x, \theta_x) = \mathrm{RLCT}_X(f_1, \ldots, f_r; \varphi_x)$ and $(\lambda_y, \theta_y) = \mathrm{RLCT}_Y(g_1, \ldots, g_s; \varphi_y)$.

By composing with projections $X \times Y \to X$ and $X \times Y \to Y$, 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 3.1. *The RLCTs for the sum and product of ideals I_x and I_y are*

$$\begin{aligned}
\mathrm{RLCT}_{X\times Y}(I_x + I_y; \varphi_x\varphi_y) &= (\lambda_x + \lambda_y,\ \theta_x + \theta_y - 1),\\
\mathrm{RLCT}_{X\times Y}(I_x I_y; \varphi_x\varphi_y) &= \begin{cases} (\lambda_x,\ \theta_x) & \text{if } \lambda_x < \lambda_y,\\ (\lambda_y,\ \theta_y) & \text{if } \lambda_x > \lambda_y,\\ (\lambda_x,\ \theta_x + \theta_y) & \text{if } \lambda_x = \lambda_y. \end{cases}
\end{aligned}$$

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 $\rho : M \to W$ a proper real analytic map. Then, the pullback $\rho^* I = \{f \circ \rho : f \in I\}$ is an ideal of real analytic functions on M. 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 3.2. *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, $\rho : M \to W$ is a change of variables away from $\mathcal{V}(I)$ and $\mathcal{M} = \rho^{-1}(\Omega \cap W)$, then*

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

Proposition 3.3. *Given a compact semianalytic set $\Omega \subset \mathbb{R}^d$ and a real analytic map $u : \Omega \to \Upsilon$, let $\rho : \mathscr{M} \to \Omega$ be an analytic isomorphism. Then for all $u^* \in \Upsilon$,*

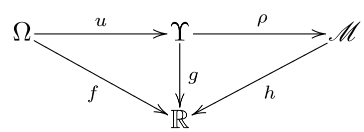
$$\mathrm{RLCT}_{\Omega}(\langle u(\omega) - u^* \rangle; \varphi) = \mathrm{RLCT}_{\mathcal{M}}(\langle u\rho(\mu) - u^* \rangle; \varphi\rho).$$

Proof. Follows from Proposition 3.8 of [4]. □

Proposition 3.4. *Given a compact semianalytic set $\Omega \subset \mathbb{R}^d$ and a real analytic map $u : \Omega \to \Upsilon$, let $\rho : \Upsilon \to \mathscr{M}$ be an analytic isomorphism. Then for all $u^* \in \Upsilon$,*

$$\mathrm{RLCT}_\Omega(\langle u(\omega) - u^* \rangle; \varphi) = \mathrm{RLCT}_\Omega(\langle \rho u(\omega) - \rho(u^*) \rangle; \varphi).$$

Proof. Consider the following commutative diagram of maps



where the sets Υ and $\mathscr{M}$ are k-dimensional and

$$\begin{aligned} f(\omega) &= \textstyle\sum_{i=1}^k (u_i(\omega) - u_i^*)^2, \\ g(u) &= \textstyle\sum_{i=1}^k (u_i - u_i^*)^2, \\ h(\mu) &= \textstyle\sum_{i=1}^k (\rho_i^{-1}(\mu) - u_i^*)^2. \end{aligned}$$

Then, the real log canonical threshold of $\langle u(\omega) - u^* \rangle$ is twice that of $f(\omega)$. Using Proposition 1.3, we can prove that the RLCT of $\langle \rho u(\omega) - \rho(u^*) \rangle$ is also twice that of $f(\omega)$ if we can verify that h is uniquely minimized at $\rho(u^*)$ and that the Hessian at $\rho(u^*)$ is positive definite. The first claim comes from the fact that h is minimized when $\rho^{-1}(\mu) = u^*$. For the second claim, we check that the Hessian is $2A^\top A$ where A is the Jacobian matrix of ρ^{-1} and A is full rank. □

4 Newton Polyhedra

Given an analytic function $f \in \mathcal{A}_0(\mathbb{R}^d)$, we pick local coordinates $\{w_1, \ldots, w_d\}$ in a neighborhood of the origin. This allows us to represent f as a power series $\sum_\alpha c_\alpha \omega^\alpha$ where $\omega = (\omega_1, \ldots, \omega_d)$ and each $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$. 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

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

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

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

where $\langle\ ,\ \rangle$ is the standard dot product. Now, given a compact subset $\gamma \subset \mathbb{R}^d$, define the *face polynomial*

$$f_\gamma = \sum_{\alpha \in \gamma} c_\alpha \omega^\alpha.$$

Recall that f_γ is singular at a point $x \in \mathbb{R}^d$ if $\mathrm{ord}_x f \geq 2$, i.e.

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

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*.

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 $f_1, \ldots, f_r$, then $\text{mon}(I)$ is generated by monomials ω^α appearing in the f_i. One consequence is that $\mathcal{P}(f_1^2 + \cdots + f_r^2)$ is the scaled polyhedron $2\mathcal{P}(I)$. 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 $I = \langle f_1, \ldots, f_r\rangle$.

Proposition 4.1. *For all compact faces $\gamma \in \mathcal{P}(I)$, $I_\gamma = \langle f_{1\gamma}, \ldots, f_{r\gamma}\rangle$.*

Lastly, we give several equivalent definitions of nondegeneracy for ideals. If an ideal I satisfies these conditions, then we say that I is *sos-nondegenerate*, where *sos* stands for *sum-of-squares*. Note that the nondegeneracy of a function f need not imply the sos-nondegeneracy of the ideal $\langle f\rangle$, e.g. $f = x + y$.

Proposition 4.2. *Let $I \subset \mathcal{A}_0$ be an ideal. The following are equivalent:*

1. *For some generating set $\{f_1, \ldots, f_r\}$ for I, $f_1^2 + \cdots + f_r^2$ is nondegenerate.*
2. *For all generating sets $\{f_1, \ldots, f_r\}$ for I, $f_1^2 + \cdots + f_r^2$ 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 the torus $(\mathbb{R}^*)^d$.*

Given a polyhedron $\mathcal{P} \subset \mathbb{R}^d$ and a vector $\tau = (\tau_1, \ldots, \tau_d)$ of non-negative integers, let the *τ-distance* l_τ be the smallest $t \geq 0$ such that $t(\tau_1+1, \ldots, \tau_d+1) \in \mathcal{P}$ and let the multiplicity θ_τ be the codimension of the face at this intersection.

Theorem 4.3. *Let the function $f : \Omega \to \mathbb{R}$ be real analytic at the origin. If f is nondegenerate and has a minimum at 0, then* $\text{RLCT}_0(f; \omega^\tau) = (1/l_\tau, \theta_\tau)$ *where l_τ is the τ-distance of $\mathcal{P}(f)$ and θ_τ its multiplicity.*

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 4.4. *Monomial ideals are sos-nondegenerate.*

Theorem 4.5. *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.

5 Regular Parameters

Let $I \subset \mathbb{R}[x_1, \dots, x_d]$ be an ideal. Recall that the variety $Y = \mathcal{V}(I)$ is regular at a point p if the rank of the Jacobian matrix $(\partial f_i / \partial x_J)$ is the codimension of Y at p. Equivalently, Y is regular at p if the local ring $\mathcal{O}_{Y,p}$ is regular.

Theorem 5.1. *If Y is regular at a point p, then the RLCT of I at p is $(r, 1)$ where r is the codimension of Y in $\mathbb{R}^n$.*

Proof. Without loss of generality, let us assume that p is the origin. Let the ring $\mathcal{R}$ be the localization of $\mathbb{R}[x_1, \dots, x_d]$ at the origin. By Theorem 8.17 of Hartshorne, the ideal $I\mathcal{R}$ is generated by r elements $f_1, \dots, f_r$ in $\mathcal{R}$. Because the Jacobian matrix $(\partial f_i / \partial \bar{x}_j)$ has rank r, the set $\{f_i\}$ may be extended to a set of local coordinates at the origin. In these coordinates, we see that the RLCT of the ideal $I = \langle f_1, \dots, f_r \rangle$ is $(r, 1)$. □

Remark 5.2. Compare this to Watanabe's theorem in his book about the RLCT being the codimension if the variety is locally isomorphic to a linear subspace. His theorem gives a geometric condition, while ours gives an algebraic condition. With this theorem, the RLCT at regular points are easily computed. As for points in the singular locus, we will need to do more analysis.

6 Working with Boundaries

These results are stated and proved in [3].

Lemma 6.1. Given a compact semialgebraic set Ω and an ideal I in the ring of real analytic functions over Ω, let M be the locus of points x in Ω where the boundary-less local $\text{RLCT}_x(I)$ is minimized. Let (λ, θ) be this minimum RLCT. Then, the RLCT of I over Ω is lower bounded by (λ, θ). If M is not contained in the boundary of Ω, then this RLCT is precisely (λ, θ).

Lemma 6.2. Let Ω be a compact semialgebraic set where the origin is in the interior of Ω. Then, the RLCT of any monomial ideal I over Ω is given by the RLCT at the origin, i.e. we may use the Newton polyhedra method to compute the RLCT.

Lemma 6.3 (Localization)**.** Let Ω be a compact semialgebraic set, R the ring of real analytic functions over Ω and I an ideal in R. Given any multiplication subset S of R, let I_S be the localization of I in the local ring R_S (in other words, all elements of S become units in the ideal I). Let V be the variety defined by I, and U be the set of points x in Ω such that $s(x)$ is nonzero for all s in S. Then, $\text{RLCT}_\Omega I = \min(a, b)$ where

$$a = \min_{x \in U} \text{RLCT}_{\Omega_x} I_S,$$
$$b = \min_{x \in V \setminus U} \text{RLCT}_{\Omega_x} I.$$

Lemma 6.3 has some interesting corollaries. First, if we let S be the complement of a maximal ideal of a point u in Ω, then $U = \{u\}$ and R_S is the local ring of the point u. Second, if we let S be the set of functions that do not vanish anywhere on $\mathcal{V}(I)$, then $V \setminus U$ is the empty set so $\mathrm{RLCT}_\Omega I = \mathrm{RLCT}_\Omega I_S$. This can be used to in our Gaussian trees problem to separate the variables into the two sets we talked about. Third, if we let S be the complement of the prime ideal of an irreducible component of V, then U is this irreducible component so we can compute RLCTs at points on U and at points not on U. If our variety V is simple normal crossing, we can use this trick to localize components away to get monomial ideals and then apply Lemma 6.2 to prove that our favorite point is the deepest singularity. This last idea can be useful for hyperplane arrangements. More generally, if we can have candidate for the deepest singularity, we can localize unimportant components away and then perform a blowup. Now, repeat this localization and blowup process again till we get our desired RLCT. After which, we can backtrack this process to find the region where the original candidate point was a deepest singularity.

7 Linear and Monomial Maps

In this section, we study semianalytic subsets $\Omega \subset \mathbb{R}^d$ which are not necessarily compact. We define $\mathrm{RLCT}_\Omega(I;\varphi)$ of an ideal I with respect to a function φ to be the minimum of local thresholds $\mathrm{RLCT}_{\Omega_x}(I;\varphi)$ over all $x \in \Omega$, if this minimum exists. We say that Ω is a *product of intervals* if $\Omega = [a_1, b_1]\times[a_2, b_2]\times\cdots\times[a_d, b_d]$ where each $a_i, b_i \in \mathbb{R} \cup \{-\infty, \infty\}$. Here, Ω need not be compact.

These results are also stated and proved in [3].

Lemma 7.1. If the subset $\Omega \subset \mathbb{R}^d$ is a product of intervals, then the RLCT of any monomial ideal I over Ω is given by the RLCT at the origin.

We say an ideal $I \subset \mathbb{R}[\omega_1, \ldots, \omega_d]$ is linear if it is generated by degree one polynomials. If Ω is a product of intervals, then the variety $\mathcal{V}_\Omega(I)$ is a polyhedron in $\mathbb{R}^d$. If a polyhedron $\mathcal{P}$ is the empty set, we define its codimension to be ∞.

Proposition 7.2. *Let I be a linear ideal and Ω a product of intervals. Then,*

$$\mathrm{RLCT}_\Omega(I;1) = (\mathrm{codim}\mathcal{V}_\Omega(I), 1).$$

Example 7.3. Let $\Omega = [0,\infty) \times [0,\infty) \subset \mathbb{R}^2$ and consider the linear ideals $I_1 = \langle x - y\rangle$ and $I_2 = \langle x + y\rangle$ and their corresponding varieties $\mathcal{V}_1$ and $\mathcal{V}_2$. Both varieties are codimension one in $\mathbb{R}^2$. However, in the orthant Ω, $\mathcal{V}_1$ has codimension one while $\mathcal{V}_2$ has codimension two. Therefore, $\mathrm{RLCT}_\Omega(I_1;1) = (1,1)$ but $\mathrm{RLCT}_\Omega(I_2;1) = (2,1)$.

Proposition 7.4. *Given a product of intervals $\Omega \subset \mathbb{R}^d$, let $f : \Omega \to \mathbb{R}^n$ be the monomial map $f(\omega) = \omega^v$, i.e.*

$$f_i(\omega) = \omega_1^{v_{1i}}\omega_2^{v_{2i}}\cdots\omega_d^{v_{di}}, \quad \textit{for } 1 \le i \le n$$

and let $\varphi : \Omega \to \mathbb{R}$ be a monomial function $\varphi(\omega) = \omega^\tau$. Given $u^* \in f(\Omega)$, suppose $u_1^*, \dots, u_r^*$ are the nonzero entries of u^*, and suppose also that $\omega_1, \dots, \omega_s$ are the parameters appearing in the monomials $f_1(\omega), \dots, f_r(\omega)$. Let I_1 be the ideal generated by binomials

$$\omega_1^{v_{1i}} \omega_2^{v_{2i}} \cdots \omega_s^{v_{si}} - u_i^*, \quad \textit{for } 1 \le i \le r$$

and I_0 be generated by the monomials

$$\omega_{s+1}^{v_{(s+1)i}} \omega_{s+2}^{v_{(s+2)i}} \cdots \omega_d^{v_{di}}, \quad \textit{for } r+1 \le i \le n.$$

Let $\tau_0 = (\tau_{s+1}, \dots, \tau_d)$ and $\Omega = W_1 \times W_0 \subset \mathbb{R}^s \times \mathbb{R}^{d-s}$. Then,

$$\mathrm{RLCT}_\Omega(\langle f(\omega) - u^* \rangle; \varphi) = (\mathrm{codim}\mathcal{V}_{W_1}(I_1) + 1/l_{\tau_0}, \theta_{\tau_0})$$

where l_{τ_0} is the τ_0-distance of the polyhedron $\mathcal{P}(I_0)$ and θ_{τ_0} its multiplicity.

8 Normalized Models

These results are not yet published.

Lemma 8.1. Given a discrete model with probabilities $p_i(\omega) = f_i(\omega)/Z(\omega)$ where i varies over $\{1, ..., k\}$, the $f_i(\omega)$ are polynomials and the normalization factor $Z(\omega) = \sum_i f_i(\omega)$ is positive over all points ω in the parameter space Ω. Then the learning coefficient of the model over the true distribution $p(\omega^*)$ is half the RLCT of the ideal generated by the two-by-two minors of the matrix

$$\begin{pmatrix} f_1(w) & f_2(w) & \cdots & f_k(w) \\ f_1(w^*) & f_2(w^*) & \cdots & f_k(w^*) \end{pmatrix}.$$

Proof. By Proposition 2.4, the learning coefficient of this discrete model is half the RLCT of the ideal $I = \langle p(\omega) - p(\omega^*) \rangle$. We claim that this ideal is generated by the minors described above. Treating the f_i as indeterminates, I is the ideal of vectors in $\mathbb{R}^k$ which are parallel to $(f_1(\omega^*), \dots, f_k(\omega^*))$. This occurs if and only if the above matrix is rank one, and the result follows. □

The next lemma is due to Aoyagi [1, Lemma 7].

Lemma 8.2. Assume the situation in Lemma 8.1 and suppose that for each i, we have $f_i(\omega) = e^{g_i(\omega)}$ for some real analytic functions $g_i(\omega)$. Then the learning coefficient of the model over the true distribution $p(\omega^*)$ is half the RLCT of the ideal generated by

$$[g_i(\omega) - g_i(\omega^*)] - [g_j(\omega) - g_j(\omega^*)]$$

where i, j varies over all distinct pairs in $\{1, \dots, k\}$.

Proof. We give an alternative proof. The functions $f_i(\omega)$ are positive and thus invertible over Ω so the ideal generated by minors in Lemma 8.1 is equal to

$$\left\langle \frac{f_2(\omega)}{f_1(\omega)} - \frac{f_2(\omega^*)}{f_1(\omega^*)}, \ \frac{f_3(\omega)}{f_1(\omega)} - \frac{f_3(\omega^*)}{f_1(\omega^*)}, \ \dots, \ \frac{f_k(\omega)}{f_1(\omega)} - \frac{f_k(\omega^*)}{f_1(\omega^*)} \right\rangle.$$

We now apply Proposition 3.4 using the log function and the result follows. □

9 Counterexamples

Example 9.1. When the boundary near a point x is a smooth hypersurface, the RLCT is not necessarily equal to the boundaryless RLCT.

Let $f = y - x^2$ and $\Omega = \{y \leq 0\}$. We compute the RLCT of f at the origin. Ignoring the boundary, since f is smooth at the origin, the RLCT of f is $(1, 1)$. Now, considering the boundary, we need to blowup the origin.

Chart 1: $x = x_1y_1, y = y_1$. Here, $f = y_1(1 - x_1^2 y_1)$ so the strict transform does not intersect the exceptional divisor in this chart, giving $\text{RLCT}(y_1; y_1) = (2, 1)$.

Chart 2: $x = x_1, y = x_1y_1$. Here, $f = x_1(y_1 - x_1)$ so the strict transform intersects the exceptional divisor at the origin and Ω is the union of the orthants $\{x_1 \leq 0, y_1 \geq 0\}$ and $\{x_1 \geq 0, y_1 \leq 0\}$. We blow up the origin once more.

Chart 2.1: $x_1 = x_2, y_1 = x_2y_2$. Here, $f = x_2^2(y_2 - 1)$ and $\Omega = \{y_2 \leq 0\}$. The strict transform does not intersect the exceptional divisor in Ω so the RLCT is just that of the exceptional divisor which is $\text{RLCT}(x_2^2; x_2^2) = (3/2, 1)$.

Chart 2.2: $x_1 = x_2y_2, y_1 = y_2$. Here, $f = x_2y_2^2(1 - x_2)$ and $\Omega = \{x_2 \leq 0\}$. Again, the strict transform does not intersect the exceptional divisor in Ω so the RLCT is given by $\text{RLCT}(x_2y_2^2; x_2y_2^2) = (3/2, 1)$.

Thus, we see that even though Ω is a half-space near the origin, the RLCT is $(3/2, 1)$ which is different from the boundaryless RLCT $(1, 1)$. □

Example 9.2. Proposition 2.7 and Lemma 2.8 of [9]. I think Lemma 2.8 may not be entirely correct when the point is at the boundary.

Let $f = x^2 + y^2$ and $\Omega = \{y^2 - x^3 < 0, x^2 + y^2 < \varepsilon\}$. The fiber $\mathcal{F}$ is the origin. I want to show that the RLCT is strictly more than 1. Blow up the origin in the plane.

Chart 1: $x = x_1y_1, y = y_1$. Here, the proper transform of $\mathcal{F}$ is $\{y_1 = 0\}$ while the proper transform of Ω is $\{1 - x_1^3 y_1 < 0\}$ which does not intersect the fiber.

Chart 2: $x = x_1, y = x_1y_1$. Here, $\mathcal{F}$ is $\{x_1 = 0\}$ while Ω is the parabolic region $\{y_1^2 - x_1 < 0\}$. They intersect at the origin. Thus, f is equivalent to x_1^2 while the Jacobian determinant is $J = x_1$. Blow up the origin in the plane.

Chart 2.1: $x_1 = x_2y_2, y_1 = y_2$. Here, $\mathcal{F}$ is the union of the axes $\{x_2 = 0 \text{ or } y_2 = 0\}$ while $\Omega = \{y_2^2 - x_2y_2 < 0\}$. They intersect at $\{x_2 = 0\}$. Also, $f = x_2^2y_2^2$ and $J = x_2y_2^2$. Along every point but the origin, f is equivalent to x_2^2 so the RLCT is $(3/2, 1)$. At the origin, we blow up again.

Chart 2.1.1: $x_2 = x_3y_3, y_2 = y_3$. All interesting points covered in Chart 2.1.2.

Chart 2.1.2: $x_2 = x_3, y_2 = x_3y_3$. Covers all points except $\{x_2 = 0, y_2 \neq 0\}$. Here, $\mathcal{F}$ is the union of the axes while $\Omega = \{0 \leq y_3 \leq 1\}$. All points have higher or equal RLCT than the origin. At the origin, we have $f = x_3^4y_3^2$ and $J = x_3^4y_3^2$ so the RLCT is $(5/4, 1)$.

Chart 2.2: $x_1 = x_2, y_1 = x_2y_2$. Here, $\mathcal{F}$ is $\{x_2 = 0\}$ while Ω is $\{x_2(x_2y_2^2 - 1) \leq 0\}$. The only point we did consider in the other charts is the origin. Here, $f = x_2^2$ while $J = x_2^2$, so the RLCT is $(3/2, 1)$.

Therefore, the RLCT in this boundary example is $(5/4, 1)$. □

10 Miscellaneous

Other Results

1. removing units: primary decomposition, Mora's algorithm
2. simplifying regular parameters
3. binomial ideals
4. homogeneous ideals

Counterexamples

1. nonreduced varieties
2. boundaries

Remark 10.1. When we do a blowup, the charts overlap so we could end up analyzing the same point in different charts. It is useful to consider the points in one chart that is not in any of the other charts. We call these points *essential*. One way is to keep track of the points we have done, and strike them off when we encounter them again in a later chart. Another way is to study the geometry of the variety before blowing up. The variety intersects the essential points in one chart if it is tangent to the image of these essential points before the blowup.

If the variety of the ideal is completely contained in the boundary, I think we can also say something about the RLCT? Example to try is the cusp reversed and the positive orthant. Not really. Need to resolve the singularities first, then see which points the bounded region picks out.

References

[1] M. Aoyagi: *Stochastic Complexity and Generalization Error of a Restricted Boltzmann Machine in Bayesian Estimation*, Journal of Machine Learning Research **11**: 1243–1272, 2010.

[2] Lin, Steele: *Lessons from Singular Learning Theory*, in preparation.

[3] Drton, Lin, Sullivant, Zwiernik: *Learning Coefficients of Gaussian Trees*, in preparation.

[4] S. Lin: *Asymptotic Approximation of Marginal Likelihood Integrals*, `arXiv: 1003.5338`, November 2011. submitted.

[5] S. Lin: *Algebraic Methods for Evaluating Integrals in Bayesian Statistics,*, Ph.D. dissertation, University of California, Berkeley, May 2011.

[6] S. Watanabe: *Algebraic Geometry and Statistical Learning Theory*, Cambridge Monographs on Applied and Computational Mathematics **25**, Cambridge University Press, Cambridge, 2009.

[7] A. Critch: Geometry of Binary Hidden Markov Models (in preparation).

[8] Smith, Zwiernik: Tree-Cumulants and the Geometry of Binary Tree Models

[9] Zwiernik: An Asymptotic Approximation of the Marginal Likelihood for General Markov Models