



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y  
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA.

MÉTODOS DE LA GEOMETRÍA ALGEBRAICA EN REDES NEURONALES.

TESIS.  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN CIENCIAS

PRESENTA:  
MARCO POLO CASTILLO VILLALBA.

DIRECTOR DE LA TESIS.  
DR. ENRIQUE JAVIER ELIZONDO HUERTA. INSTITUTO DE MATEMÁTICAS,  
UNAM, CIRCUITO EXTERIOR, CIUDAD UNIVERSITARIA, 04510.

MÉXICO, D. F. 28 DE OCTUBRE DE 2015.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



## **Agradecimientos**

Quiero agradecer la realización del siguiente trabajo, al profesor Israel Moreno Mejía por el inicio y primeras revisiones de este trabajo, y a mi asesor de tesis Dr. Enrique Javier Elizondo Huerta por darle seguimiento a este trabajo de tesis por su tiempo y atenciones al involucrarse en este tema y tomar la dirección de esta tesis de manera imprevista, le agradezco mucho profesor, muchas gracias. A mis padres Guillermina Hernández Villalba y Reyes Castillo Contreras, los cuales por su perseverancia y enseñanzas, respeto y quiero profundamente, sobretodo por regalarme la experiencia de existir, gracias padres.... Y por último quiero agradecer a mis compañeros de trabajo en general por las noches de proyectos sin fin, donde mi paciencia no era una de mis virtudes, por su tolerancia, gracias equipo, a todos dedico este trabajo.

## Índice general

### Capítulo.

Introducción. . . . .	1
<b>1.</b> Preliminares de geometría convexa . . . . .	5
1.1. Polítopos, conos y abanicos. . . . .	5
1.2. Retículas y monoides . . . . .	13
1.3. Subdivisión estelar, operación estelar regular . . . . .	15
<b>2.</b> Variedades tóricas afines . . . . .	21
2.1. Toro algebraico y morfismos asociados . . . . .	21
2.1.1. Bases de Hilbert . . . . .	23
2.2. Variedades tóricas afines y politopo de Newton . . . . .	25
2.3. Ideales tóricos . . . . .	30
2.4. Morfismos tóricos y pegado de morfismos . . . . .	34
2.4.1. Variedades tóricas y abanicos . . . . .	41
2.5. Normalización de variedades tóricas afines . . . . .	41
2.6. Conjuntos proyectivos y conos normales. . . . .	46
<b>3.</b> Resolución de singularidades . . . . .	51
3.1. Teorema de Hironaka y modificación tórica . . . . .	51
3.2. Blow up . . . . .	64

<b>4.</b>	La estadística de redes neuronales Bayesianas	68
4.1.	R.N.B y teoría del aprendizaje Bayesiano . . . . .	68
4.2.	Máquinas singulares. . . . .	76
4.2.1.	Efecto de las singularidades en R.N.B . . . . .	79
4.3.	El teorema de Watanabe Sumio y curvas de aprendizaje. . . . .	83
<b>5.</b>	Modificación tórica y aplicaciones en R.N.B	94
5.1.	Modificación tórica y curva de aprendizaje en una mezcla binomial . . . . .	94
5.2.	Blow up y un perceptrón de tres capas . . . . .	97
5.3.	Modificación tórica en máquinas de aprendizaje no supervisado. . . . .	101
5.3.1.	Estimación Bayesiana de variables latentes. . . . .	102
5.3.2.	Aplicación en una máquina de aprendizaje de mezclas de modelos. . . . .	104
	Conclusiones . . . . .	112

## **Apéndice.**

<b>A.</b>	Un algoritmo para el cálculo de las bases de Hilbert en una variedad tórica afín.	113
<b>B.</b>	Modificaciones tóricas: Justificación de cálculos.	114
	Lista de símbolos. . . . .	125
	Bibliografía . . . . .	128

## Índice de figuras

### Figura

1.1.	Link de la cara F en un politopo C; ref. [13]. . . . .	19
1.2.	Subdivisión estelar regular; ref. [13]. . . . .	19
1.3.	Subdivisión estelar regular en un cubo; ref. [13]. . . . .	20
2.1.	Conos normales asociados al politopo K; ref. [13]. . . . .	49
2.2.	Conos duales asociados a cada cono asociado a los conjuntos proyectivos; ref. [13]. . . . .	50
3.1.	Modificación tórica resuelto en la superficie cónica; ref. [13]. . . . .	67
5.1.	<p><b>Arquitectura de una red neuronal de una capa.</b> (Donde los <math>\{u_1(t), \dots, u_M(t)\} \subset \Omega</math>, son las funciones de vectores aleatorios que ponderan los nodos de entrada, las funciones <math>\{\varphi_1, \dots, \varphi_n\} \subset \Theta \subset \Omega</math> son las funciones sinápticas de la red o unidades ocultas donde se entrena la red neuronal, las ponderaciones <math>W_{Pn}</math> son las funciones de transición de nuestra red neuronal hacia la capa de salida y en nuestro caso son densidades de probabilidad bayesiana o inferencia verdadera de la máquina entre cada par de nodos <math>P</math> y <math>n</math>, por último tenemos la salida de la red representada por funciones <math>\{y_1(t), \dots, y_P(t)\} \subset \Omega</math> donde <math>\Sigma</math> significa la ponderación de todos los pesos de la red desde la capa oculta, ver [28]). . . . .</p>	110

## Introducción.

Los objetivos del siguiente desarrollo son entender en que consiste la **teoría del aprendizaje singular** propuesta por **S. Watanabe en 2001; ref. [29]**, y su relación con la geometría algebraica, principalmente con la **teoría de resolución de singularidades**, además proponemos en el apéndice **B** de este trabajo una manera de resolver las singularidades en el espacio de parámetros de diferentes redes neuronales Bayesianas a través del **teorema 3.1.7** y el **corolario 4.3.3** propuestos en el capítulo 3 y 4 respectivamente; haciendo uso de las bases Hilbert caracterizamos novedosamente diferentes ejemplos de redes neuronales Bayesianas con variedades tóricas asociadas al espacio de parámetros de las mismas, reproduciendo los resultados obtenidos por **S. Watanabe, ref. [26], [28], [29]**; clarificando estos ejemplos y proponiendo un método más sistemático y riguroso en la teoría del aprendizaje para máquinas singulares; ref. [35].

El siguiente trabajo está dividido en dos partes principalmente. El desarrollo de estas dos partes tienen como objeto exponer de manera introductoria dos teorías muy distintas en esencia como es el caso de algunos métodos de resolución de singularidades desarrollados por los teóricos de la geometría algebraica, [13], [14], [15]; y otra los conceptos básicos de la teoría estadística de máquinas de aprendizaje bayesiano conocida también como redes neuronales bayesianas desarrollada por los teóricos de la inteligencia artificial, [3], [4], [7], [28], [29], [30]; expondremos el problema fundamental que estos últimos presentan y la relación con las técnicas expuestas en la primera parte.

La primera parte consiste de cuatro capítulos enfocados principalmente al estudio de resolución de singularidades mediante la técnica de modificación tórica y blow up usual como consecuencia de este método, en el primer capítulo exponemos los fundamentos de la geometría convexa combinatoria, propiedades y definiciones de conos y polítopos, abanicos y la operación geométrica de subdivisión estelar regular y el puente que vincula estos conceptos con el álgebra mediante el con-

cepto de monoides, estos últimos conceptos de importancia para la construcción y entendimiento geométrico de variedades tóricas y blow up tórico, [11], [13]. En el siguiente capítulo, definimos variedades tóricas afines y sus principales morfismos los cuales estarán definidos a partir de un conjunto de conceptos como el anillo coordinado de polinomios de Laurent, en este anillo asociaremos una familia particular de monomios definidos por medio de un conjunto de conos y politopos también muy particulares a saber (politopos de Newton y conos poliédricos racionales) unido a estos calcularemos los generadores mínimos de dichos objetos mediante la técnica de bases de Hilbert, con dichas bases podremos construir y representar de manera explícita ideales tóricos de variedades toricas afines, [11], [14], [27].

En el capítulo tres expondremos brevemente una introducción a variedades tóricas proyectivas y también en breve calcularemos anillos coordinados mediante cartas afines del cubriente afín asociado a ciertos conjuntos proyectivos. Este hecho se basa también del cálculo de bases de Hilbert de conos poliédricos racionales.

Finalmente en el último capítulo de esta primera parte definimos formalmente una singularidad desde el punto de vista de variedades tóricas y desarrollamos la técnica de **modificación tórica** a partir del cálculo de **bases de Hilbert** de conos poliédricos racionales asociados al politopo de Newton del polinomio a desingularizar. También definimos la técnica del **blow up** usual, como ideal tórico de una variedad tórica afín en particular, [11], [16], [18].

A partir de los métodos explicados en la parte uno, estaremos equipados con la herramienta suficiente para tratar ciertas familias de polinomios que son de interés en la teoría de redes neuronales bayesianas, los cuales surgieron al ser estudiados por el profesor S. Watanabe del Instituto de Tecnología de Tokio al calcular que cuando una red neuronal bayesiana aprende a partir de un conjunto de datos estadísticos ésta tendrá un error en el aprendizaje del modelo predicho por la misma y el modelo verdadero del problema que se esta estudiando, este error es medido a través de una curva de aprendizaje pero esta a su vez depende de que tanto distan el modelo predictivo y el modelo verdadero esta distancia es una métrica llamada distancia Kullback o también es conocida como entropía de la información a través de esta métrica los teóricos de ciencias de la computación desa-

rollaron la llamada geometría de la información, [30], [32], [33], dicha métrica esta representada por polinomios los cuales no tienen singularidad para cierto tipo de redes neuronales pero en su mayoría se ha demostrado que casi todas las redes neuronales bayesianas son singulares, [39], y por consecuencia es difícil el cálculo de curvas de aprendizaje lo cual hace muy difícil el entrenamiento de estas neuronas artificiales en diversos campos como la medicina, la industria del petróleo, reconocimiento de patrones y modelos de aprendizaje bayesiano en general, [36]. Las técnicas de modificación tórica y blow up serán aplicadas a estas familias de polinomios en particular; y como consecuencia el cálculo de la curva de aprendizaje asociada a algunos ejemplos de interés, [40], [41], [42].

A continuación el contexto de la parte dos es comentado resumidamente.

La segunda parte de nuestro trabajo consiste de dos capítulos primordialmente, en el primero definiremos los conceptos respecto que es una máquina estadística de aprendizaje bayesiano, en que clasificación entran estas a partir del concepto de la matriz de información de Fisher y su íntima relación cuando esta es singular a los conceptos de geometría algebraica, [4], [7], [8], [32]. En la siguiente sección veremos la estadística de redes neuronales bayesianas los principales conceptos que estamos interesados en abordar y sus definiciones en términos de ecuaciones explícitas que representan estos temas, abordaremos concretamente los conceptos de **distancia Kullback o entropía de la información** la relación de esta expresión con cierta función zeta y la obtención del polo más grande de esta expresión, la determinación del polo más grande es de utilidad para la obtención de otra expresión que será el tema principal para la teoría de redes neuronales bayesianas y se trata de el cálculo de la **generalización del error o curva de aprendizaje** este último concepto se formalizará en un teorema de central importancia y es de los de mayor relevancia en la estadística de redes neuronales bayesianas a saber el **teorema de curvas de aprendizaje o generalización del error** de Watanabe Sumio, [30], [32], [38].

En la última sección tratamos el tema denominado máquinas singulares donde a su vez las R.N.B, entran también en este contexto de clasificación, y exponemos la importancia y la problemática que generan las singularidades en este tipo de máquinas surgidas a partir del hecho de que la matriz

de Fisher es singular y así el buen desempeño del entrenamiento y aprendizaje de estos modelos de inteligencia artificial se ve afectado, [33], [35], [41].

En el último capítulo de nuestro trabajo nos concentramos en dar diversas aplicaciones de las técnicas de geometría algebraica a la familia de máquinas singulares como es el caso de R.N.B específicamente al conjunto algebraico obtenido del determinante de la matriz de Fisher la cual es singular, se abordarán varios ejemplos explícitos de interés y de reciente publicación de estos últimos años como es el caso de un perceptron de tres capas y máquinas de aprendizaje no supervisado, [40], [44], una vez desingularizado el conjunto algebraico que afecta este modelo se procede al cálculo de su curva de aprendizaje para su implementación y entrenamiento para futuras aplicaciones en la ingeniería artificial, también se desingulariza una máquina de aprendizaje en un ejemplo más corto mediante modificación tórica, [41], [42],[43].

# Capítulo 1

## Preliminares de geometría convexa

Comenzaremos definiendo los conceptos principales para el estudio de politopos convexos y conos, algunas propiedades de interés así como su relación con teoría de monoides mediante un conjunto de proposiciones y teoremas para finalmente definir operaciones estelares regulares mediante abanicos, las pruebas de lemas y teoremas que son probados en este capítulos fueron tomados de, ver referencias, [13] G. Ewald, [11] D. Cox, J. Little, H. Schenck (lema 1.1.3 y lema 1.2.2), las pruebas hechas en la **definición 1.1.6**, son hechas por el autor.

### 1.1. Politopos, conos y abanicos.

1.1.1 **Definición.** Decimos que  $\alpha$  es una combinación convexa de  $\alpha_1 \dots \alpha_r \in \mathbb{R}^n$  si existen  $\lambda_1 \dots \lambda_r \in \mathbb{R}$  tal que

$$(1) \quad \alpha = \alpha_1 \lambda_1 + \dots + \alpha_r \lambda_r$$

$$(2) \quad \lambda_1 + \dots + \lambda_r = 1,$$

$$(3) \quad \lambda_1 \geq 0, \dots, \lambda_r \geq 0.$$

Si la condición (3) es omitida, tenemos una combinación afín de  $\alpha_1 \dots \alpha_r$ , y  $\alpha, \alpha_1 \dots \alpha_r$ , son

llamados afínmente dependientes. Si  $\alpha, \alpha_1, \dots, \alpha_r$  no son afínmente dependientes, decimos que son afínmente independientes.

**1.1 Definición.** El conjunto de todas las combinaciones convexas de elementos de un conjunto  $M \subset \mathbb{R}^n$  es llamado **envolvente convexa** y denotamos

$$\text{conv } M.$$

en particular,  $\text{conv} \emptyset = \emptyset$ . Análogamente, el conjunto de todas las combinaciones afines de elementos de  $M$  es llamado el envolvente afín **aff**  $M$ . Denotaremos **lin**  $M$  como el envolvente lineal de  $M$  y es el espacio lineal más pequeño que contiene a  $M$ .

Sí  $M = \{\alpha_1, \dots, \alpha_r\}$  es un conjunto finito, decimos que  $\mathbf{P} := \text{conv } M$  es un **politopo convexo** o simplemente un **politopo**.

**1.1.2 Definición.** Un **cono poliédrico convexo** en  $M$  es un conjunto de la forma

$$\sigma = \text{Con}(S) = \left\{ \sum_{j=1, \dots, r} \lambda_j u : \lambda_j \in \mathbb{R}^+, u \in S \right\} \subseteq M,$$

donde  $S \subseteq M$  es finito. Decimos que  $\sigma$  es **generado** por  $S$ . También definimos  $\mathbf{Con}(\emptyset) = \mathbf{0}$ , y de aquí en adelante llamaremos a  $\sigma$  solo como cono poliédrico.

Sí  $\alpha_1, \dots, \alpha_r$  son afínmente independientes, decimos que

$$\mathbf{Conv}\{\alpha_1, \dots, \alpha_r\}$$

es un **(r-1)-simplejo** o simplemente un **simplejo**.

En el futuro denotamos el producto escalar usual de dos vectores en el espacio euclídeo  $n$ -dimensional de  $\mathbb{R}^n$  como,  $\langle \cdot, \cdot \rangle$ .

**1.1.3 Definición.** Sea  $m \in M$ ,  $m \neq 0$ , y  $s \in \mathbb{R}$ , el conjunto  $H_p := \{p \in M : \langle p, m \rangle = s\}$  es un hiperplano y los conjuntos  $H_m^+ := \{p \in M : \langle p, m \rangle \geq s\}$  y  $H_m^- := \{p \in M : \langle p, m \rangle \leq s\}$  son los semi-espacios acotados por  $H_p$ . Si  $\sigma \subset H_m^+$  cono poliédrico convexo y  $s = 0$ , decimos que  $\sigma$  tiene un

**ápice en 0** y  $H_p$  es el **hiperplano soporte** del cono  $\sigma$ .

**1.1.4 Definición.** Dado un cono poliédrico  $\sigma \subseteq N$ , su **cono dual** es definido

$$\sigma^\vee = \{m \in \mathbb{R}^n : \langle m, u \rangle \geq 0 \forall u \in \sigma\}.$$

**1.1.5 Definición.** Decimos que  $\tau$  es una **cara** del cono poliédrico  $\sigma$  si existe un hiperplano  $H_p$  tal que  $\tau = H_p \cap \sigma$  para algún  $p \in \sigma^\vee$  y lo escribimos como  $\tau \preceq \sigma$ . Usando  $p = 0$  muestra que  $\sigma$  es una cara de sí mismo,  $\sigma \preceq \sigma$  i.e., ( $\sigma = H_0 \cap \sigma$ ). Las caras  $\tau \neq \sigma$  son llamadas **caras propias** y decimos  $\tau \prec \sigma$ .

Las caras de un cono poliédrico tienen las siguientes propiedades claras.

**Lema 1.1.1.** Sea  $\sigma$  un cono poliédrico, entonces:

- (a) Toda cara de  $\sigma$  es un cono poliédrico.
- (b) Una intersección de dos caras de  $\sigma$  es de nuevo una cara de  $\sigma$ .
- (c) Una cara de una cara de  $\sigma$  es de nuevo una de  $\sigma$ .

**Lema 1.1.2.** Sea  $\tau$  una cara de un cono poliédrico  $\sigma$ . Si  $v, w \in \sigma$  y  $v+w \in \tau$ , entonces  $v, w \in \tau$ .

Observaciones: Una **careta** de  $\sigma$  es una cara  $\tau$  de codimensión 1 i.e.,  $\dim \tau = \dim \sigma - 1$ . Un **vértice** es una cara de dimensión cero y una **arista** de  $\sigma$  es una cara de dimensión 1.

**1.1.6 Definición.** Decimos que un cono poliédrico  $\sigma \subset M$  es **fuertemente convexo** si satisface los enunciados equivalentes siguientes:

- (a)  $\{0\}$  es una cara de  $\sigma$ .
- (b)  $\sigma \cap -\sigma = \{0\}$ .
- (c)  $\dim \sigma^\vee = n$  donde  $n$  es la dimensión del espacio real  $n$ -dimensional real tal que  $M \subset \mathbb{R}^n$ .

**Demostración.** (a) $\Rightarrow$ (b). Si  $0 \preceq \sigma$ , entonces existe un hiperplano  $H_p$  con  $p \in \sigma^\vee$  tal que  $\{0\} = H_p \cap \sigma$ , en particular para el hiperplano soporte cero vector tenemos que;  $\{0\} = H_0 \cap \sigma$  donde

$H_0 = \{0\}$  con  $0 \in \mathbb{R}^n$ , es claro que,  $-0 \in -\sigma$ , de donde se sigue  $\{0\} = H_0 \cap -\sigma$  y obtenemos la siguiente relación de conjuntos,  $H_0 \cap \sigma \cap -\sigma = \{0\}$  escrito de otra manera,  $\{0\} \cap \sigma \cap -\sigma = \{0\}$  lo cual implica que,  $\sigma \cap -\sigma = \{0\}$  de donde se sigue el resultado; **q.e.d.**

(b) $\Rightarrow$ (a). Si  $\sigma \cap -\sigma = \{0\}$ , nuevamente elegimos el hiperplano soporte  $H_0$  tal que  $H_0 \cap \sigma \cap -\sigma = \{0\}$ , esto implica que,  $H_0 \cap \sigma = \{0\}$  y  $H_0 \cap -\sigma = \{0\}$ ; entonces concluimos que  $0 \preceq \sigma$  y también como consecuencia  $0 \preceq -\sigma$ , **q.e.d.**

(b) $\Rightarrow$ (c). Sean  $H_i$  hiperplanos soporte de  $\sigma$  con  $i = 1, \dots, n$ , y sean  $\{b_i\}$  sus rayos generadores de cada hiperplano y elegimos  $p \in \sigma \cap -\sigma = \{0\}$  de modo que siempre podemos escribir este elemento como la siguiente combinación convexa  $\sum_i \lambda_i b_i$  donde  $\lambda_i \in \mathbb{R}^+$  por definición, de la relaciones anteriores obtenemos la siguiente igualdad  $\sum_i \lambda_i b_i = -\sum_i \lambda_i b_i = 0$  siendo que  $b_i \neq 0 \in \mathbb{R}^n \forall i$ , esto implica que  $\lambda_i = 0, \forall i$ ; y muestra que el conjunto  $\{b_i\}$  es l.i.

Puesto que  $H_i$  son hiperplanos soporte de  $\sigma$ , entonces tenemos  $\langle \sigma, b_i \rangle \geq 0$ , donde  $\forall p \in \sigma$  y  $\forall i$ , esto es igual a  $\langle p, b_i \rangle \geq 0$  y también implica que  $p \in \sigma^\vee$ ; así el conjunto  $\{b_i\}$  genera  $\sigma^\vee$ ; por tanto,  $\dim \sigma^\vee = n$ , **q.e.d.**

(c) $\Rightarrow$ (b). Sea  $\dim \sigma^\vee = n$ , y  $\{b_i\}$  el conjunto generador que expande  $\sigma^\vee$ , para todo generador tenemos que  $\langle b_i, \sigma \rangle \geq 0 \Rightarrow \langle b_i, -\sigma \rangle \leq 0 \forall p \in \sigma$ ; de donde conseguimos la siguiente relación,  $\langle b_i, \sigma \rangle = \langle b_i, -\sigma \rangle = 0, \forall p \in \sigma$  y esto implica que;  $\langle b_i, p \rangle = \langle b_i, -p \rangle = 0 \Rightarrow p = 0 \in \mathbb{R}^n$  así  $\{p\} = \sigma \cap -\sigma$ ; así concluimos la prueba, **q.e.d.**

**1.1.7 Definición.** Definimos el complemento ortogonal y la cara dual de  $\tau \preceq \sigma$  del siguiente modo:

$$\begin{aligned} \tau^\perp &= \{m \in M : \langle m, u \rangle = 0 \forall u \in \tau\}, \\ \tau^\vee &= \{m \in \sigma^\vee : \langle m, u \rangle = 0 \forall u \in \tau\}, \\ &= \sigma^\vee \cap \tau^\perp \end{aligned}$$

**1.1.8 Definición.** Sea  $\sigma \subseteq M$  un cono poliédrico contenido en  $\mathbb{R}^n$  definimos el subespacio más pequeño que contiene a  $\sigma$ . También definimos el interior relativo de  $\sigma$ , denotado por  $Relint(\sigma)$ ,

como  $Relint(\sigma) := \sigma^\vee \setminus \sigma^\perp$ .

Si  $m \in \sigma^\vee$ , entonces también se tiene que.

$$m \in Relint(\tau^\vee) \iff \tau = H_m \cap \sigma.$$

**Lema 1.1.3. (Lema de separación).** Sea  $\sigma_1$  y  $\sigma_2$  conos poliédricos en  $M$  que se intersectan a lo largo de una cara común digamos  $\tau = \sigma_1 \cap \sigma_2$ . Entonces

$$\tau = \sigma_1 \cap H_m = \sigma_2 \cap H_m,$$

para cualquier  $m \in Relint\sigma_1^\vee \cap (-\sigma_2)^\vee$ .

**Demostración.** Definimos para los siguientes conos poliédricos  $\sigma_1, \sigma_2 \subseteq M$ ,  $\sigma_1 - \sigma_2 = \{a - b : a \in \sigma_1, b \in \sigma_2\}$ . Un resultado estándar de teoría de conos nos dice que:

$$\sigma_1^\vee \cap (-\sigma_2)^\vee = (\sigma_1 - \sigma_2)^\vee.$$

Ahora fijamos un  $m \in Relint(\sigma_1^\vee \cap (-\sigma_2)^\vee)$ . También probaremos en lema 2.1.1.2 (bases de Hilbert) que  $\sigma \cap (-\sigma)$  es la menor cara de un cono respecto la relación de orden  $\preceq$  definida en este capítulo; y afirmamos que si  $m \in Relint(\sigma)$  entonces la menor cara bajo dicha relación es  $H_m \cap \sigma$ .

Prueba de la afirmación. Sea  $m \in Relint(\sigma)$  entonces por definición tenemos  $\langle m, u \rangle > 0 \forall u \in \sigma^\vee / \sigma^\perp$ . Y sea  $W = \sigma \cap (-\sigma)$  la menor de las caras del cono  $\sigma$  y supongamos que  $W \neq \emptyset$  entonces existe un hiperplano  $H_m$  que se intersecta con ella, así obtenemos que:  $W = H_m \cap (\sigma \cap (-\sigma)) = (H_m \cap \sigma) \cap (H_m \cap (-\sigma))$ ; siendo que  $\langle m, u \rangle > 0 \forall u \in \sigma^\vee / \sigma^\perp$  i.e.  $\langle m, u \rangle \neq 0$  así  $m \notin H_m$  y  $m \notin (-\sigma)$  por tanto  $H_m \cap (-\sigma) = \emptyset$  y como  $W \neq \emptyset$ , solo queda que la menor cara pueda escribirse como  $W = H_m \cap \sigma$ , con esto probamos la afirmación.

Ahora de acuerdo a los hechos anteriores, tomamos  $m \in Relint(\sigma_1^\vee \cap (-\sigma_2)^\vee)$  aplicamos nuestros resultados en el cono poliédrico  $(\sigma_1 - \sigma_2) \subseteq M$  donde la menor cara esta dada por,  $(\sigma_1 - \sigma_2) \cap (\sigma_2 - \sigma_1)$ , i.e;

$$H_m \cap (\sigma_1 - \sigma_2) = (\sigma_1 - \sigma_2) \cap (\sigma_2 - \sigma_1).$$

Mostremos que  $\tau - \tau = (\sigma_1 - \sigma_2) \cap (\sigma_2 - \sigma_1)$ , una inclusión es clara siendo que por hipótesis  $\tau = \sigma_1 \cap \sigma_2$ . Para la otra inclusión, elegimos  $u \in (\sigma_1 - \sigma_2) \cap (\sigma_2 - \sigma_1)$  y lo escribimos de la siguiente manera;

$$u = a_1 - a_2 = b_2 - b_1, \quad a_1, b_1 \in \sigma_1, \quad a_2, b_2 \in \sigma_2.$$

Entonces  $a_1 + b_1 = a_2 + b_2$  implica que este elemento pertenece a  $\tau = \sigma_1 \cap \sigma_2$ . Siendo que  $a_1, b_1 \in \sigma_1$ , tenemos que  $a_1, b_1 \in \tau$  por el lema 1.1.2, y  $a_2, b_2 \in \tau$  se sigue de manera similar. Y de ese modo  $u = a_1 - a_2 \in \tau - \tau$ , como queríamos.

Concluimos que  $H_m \cap (\sigma_1 - \sigma_2) = \tau - \tau$ . Intersectando con  $\sigma_1$ , obtenemos

$$H_m \cap \sigma_1 = (\tau - \tau) \cap \sigma_1 = \tau,$$

donde en la última igualdad usamos nuevamente el lema 1.1.2. Repitiendo el procedimiento anterior pero ahora intersectamos con  $-\sigma_2$ , obtenemos de manera análoga,

$$H_m \cap (-\sigma_2) = (\tau - \tau) \cap (-\sigma_2) = \tau,$$

se sigue que  $H_m \cap \sigma_2 = \tau$ , y con esto probamos el resultado, **q.e.d.**

**1.1.9 Definición.** Decimos que  $\sigma \subseteq M$  es un **cono reticular** si se cumple (a).

(a) Los vectores  $\alpha \in \mathbb{Z}^n \subset \mathbb{R}^n$  serán llamados **vectores reticulares** y si  $\sigma = \text{Con}(\alpha_1, \dots, \alpha_r)$  con  $\alpha_1, \dots, \alpha_r$  vectores reticulares entonces llamamos a  $\sigma$  un **cono reticular**. Similarmente sea  $\text{Conv}(\alpha_1, \dots, \alpha_r)$  y será llamado un **politopo reticular** si los vectores  $\alpha_1, \dots, \alpha_r$  son vectores reticulares.

(b) Un vector reticular es **primitivo o simple** si sus coordenadas son primos relativos a pares. Si  $\sigma = \text{Con}(\alpha_1, \dots, \alpha_k)$  donde  $\alpha_1, \dots, \alpha_k$  son vectores reticulares primitivos, decimos que  $\sigma$  es un **cono regular**.

(c) También decimos que un cono  $\sigma = \text{Con}(\alpha_1, \dots, \alpha_r)$  es **simple o cono simplejo**, si los vectores  $\alpha_1, \dots, \alpha_r$  son linealmente independientes, y se dice que  $\sigma$  es **simplicial** si cada cara propia de  $\sigma$  es un cono simplejo.

**Nota:** Si  $\sigma = \text{Con}(\alpha_1, \dots, \alpha_k)$  es cono regular, como sabemos también es cierto que los elementos  $\alpha_1, \dots, \alpha_k$  son **primitivos**, y además existen vectores primitivos  $\alpha_{k+1}, \dots, \alpha_n \in \mathbb{Z}^n$  tales que,

$$\det(\alpha_1, \dots, \alpha_n) = \pm 1.$$

**Teorema 1.1.1.** Sea  $\sigma$  un cono  $n$ -dimensional en  $\mathbb{R}^n$  con ápice 0, sean  $b_1, \dots, b_r$  las normales internas de las caretas de  $\sigma$ . Entonces

$$\sigma^\vee = \text{Con}(b_1, \dots, b_r).$$

**Demostración.** Cada hiperplano soporte de  $\sigma$  por definición es  $H_i := \{\alpha : \langle \alpha, b_i \rangle = 0\}$  de ahí se sigue que;  $\langle \sigma, b_i \rangle \geq 0$ . Esto implica  $\sigma^\vee \supset \text{Con}(b_1, \dots, b_r)$ . Si  $\langle \alpha, \sigma \rangle \geq 0$ , pero  $\alpha$  no pertenece  $\text{Con}(b_1, \dots, b_r) := \sigma'$ , entonces, ahí existe una cara  $\text{Con}(b_1, \dots, b_r)$  de  $\sigma'$  y  $b_{i_0}$ , tal que  $\alpha = -\lambda_0 b_{i_0} + \dots + \lambda_s b_{i_s}$ , con  $\lambda_0, \dots, \lambda_s$  positivos, y un  $a_k$  tal que  $\langle b_{i_j}, a_k \rangle = 0$  para  $j = 1, \dots, s$  pero  $\langle b_{i_j}, a_k \rangle > 0$ . Ahora,  $\langle b_{i_j}, a_k \rangle = -\alpha \langle b_{i_0}, a_k \rangle < 0$ , lo cual es una contradicción **q.e.d.**

**Lema 1.1.4.** Sean  $\sigma$  y  $\sigma_i$  conos en  $\mathbb{R}^n$ .

- (a) Si  $\dim \sigma = n$  y  $\sigma$  tiene 0 como ápice, entonces  $\dim \sigma^\vee = n$  y  $\sigma^\vee$  tiene 0 como ápice.
- (b)  $\text{cospans} \sigma^\vee = \sigma^\perp$ .
- (c)  $\sigma$  tiene 0 como ápice  $\iff \dim \sigma^\vee = n$ .
- (d)  $(\sigma_1 + \sigma_2)^\vee = (\sigma_1)^\vee \cap (\sigma_2)^\vee$ .
- (e)  $(\sigma_1 \cap \sigma_2)^\vee = (\sigma_1)^\vee + (\sigma_2)^\vee$ .
- (f)  $(\sigma^\vee)^\vee = \sigma$ .

**Demostración.**

(d), (e) y (f) se siguen de la definición de conos duales.

(c) Sea  $\sigma \neq 0$  con un ápice. Entonces, hay un hiperplano  $H_p$  con  $H_p \cap \sigma = 0$  y  $H_p^+ \supset \sigma$ . Sea  $b \in \sigma$  un generador del rayo  $(H_p^+)^\vee$ . En una vecindad suficientemente pequeña de  $b$ , encontramos

vectores li  $b_1, \dots, b_r$  tal que, para  $H_{p_i} = \{\alpha : \langle \alpha, b_i \rangle = 0\}$ . Tenemos también que  $\sigma \cap H_{p_i} = 0$  y  $\sigma \subset (H_{p_i})^+, i = 1, \dots, n$  ( de otra manera  $H_p \cap \sigma \neq 0$  debería seguirse de que  $\sigma$  es un conjunto cerrado). Claramente,  $b_i \in \sigma^\vee$ , y , por tanto,  $Con(b_1, \dots, b_n) \subset \sigma^\vee$  implica que  $\sigma^\vee$  debe ser n-dimensional.

El regreso es análogo.

(a) Sea  $u \in int(\sigma)$ . Entonces, para las caretas normales  $b_1, \dots, b_r$  como en el teorema 1.1.1,

$$\langle u, b_i \rangle > 0, \quad para \quad i = 1, \dots, r.$$

Así, para  $H'_u := \{\alpha : \langle u, \alpha \rangle = 0\}$ , por teorema 1.1.1, obtenemos,

$$H'_u \cap \sigma^\vee = \{0\}.$$

(b) Nosotros obtenemos las siguientes afirmaciones equivalentes que prueban esta última parte:

$$b1. \quad \alpha \in cospan \sigma^\vee = \sigma^\vee \cap (-\sigma^\vee).$$

$$b2. \quad \iff \langle \alpha, \sigma \rangle \geq 0 \quad y \quad \langle -\alpha, \sigma \rangle \leq 0.$$

$$b3. \quad \iff \langle \alpha, \sigma \rangle = 0.$$

$$b4. \quad \iff \alpha \in \sigma^\perp. \quad \mathbf{q.e.d.}$$

**Lema 1.1.5.** Sea  $\tau$  una cara de  $\sigma$  y sea  $m \in Relint(\tau^\perp \cap \sigma^\vee) \setminus \{0\}$ . Entonces,

$$\tau^\vee = \sigma^\vee + \lambda(-m), \quad \forall \lambda \in \mathbb{R}^+.$$

**Demostración.** Si nosotros podemos demostrar las siguientes inclusiones

$$\tau \subset \sigma \cap m^\perp \subset \sigma \cap (-m)^\vee$$

son igualdades, entonces la afirmación se sigue inmediatamente del inciso (e) del resultado anterior.

Para  $\alpha \in \sigma \cap (-m)^\vee$ , sabemos que  $\langle \alpha, m \rangle \leq 0$ , mientras  $m \in \sigma^\vee$  significa que  $\langle \alpha, m \rangle \geq 0$ ; así,  $\sigma \cap (m)^\perp = \sigma \cap (-m)^\vee$ . Para  $\tau \supset \sigma \cap (m)^\perp$  es suficiente verificar lo siguiente:

$$\text{Si } v \in \sigma \setminus \tau \implies \langle m, v \rangle \neq 0.$$

Obviamente podemos asumir que  $v \neq -m$  esta sobre el plano generado por  $0, v, y m$ , asumimos que  $n = 2$ . Entonces,  $\text{lin}\tau = m^\perp$  y  $v$  no pertenece a  $\text{lin}\tau$ , así,  $\langle v, m \rangle \neq 0$ , **q.e.d.**

## 1.2. Retículas y monoides

**1.2.1 Definición.** En este trabajo una **retícula** es un grupo abeliano libre de rango finito. Así una retícula de rango  $n$  es isomorfo a  $\mathbb{Z}^n$ .

**1.2.2 Definición.** Sean  $M$  y  $N$  dos retículas, y en los cuales definimos el siguiente homomorfismo de retículas de la siguiente manera mediante el producto interno usual de  $\mathbb{R}^n$ ,  $\langle, \rangle : M \times N \rightarrow \mathbb{Z}$ , identificamos a  $N$  con  $\text{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$ , y decimos que  $N$  es la **retícula dual** de la retícula  $M$ , y de la misma manera para  $M$  si este es identificado con  $\text{Hom}_{\mathbb{Z}}(N, \mathbb{Z})$ , y escribimos en cualquiera de los casos  $N = M^\vee$ .

**1.2.3 Definición.** Sean  $M$  y  $N$  retículas duales, y asociamos a ellos espacios vectoriales respectivamente de la siguiente manera,  $M_{\mathbb{R}} = M \otimes_{\mathbb{Z}} \mathbb{R}$  y  $N_{\mathbb{R}} = N \otimes_{\mathbb{Z}} \mathbb{R}$  sea  $\sigma = \text{Con}(S) \subseteq M_{\mathbb{R}}$ , para algún conjunto  $S \subseteq M$ , decimos entonces que  $\sigma$  es un **cono poliédrico racional**.

Nota: Es fácil ver que todo cono reticular  $\sigma$  es un cono poliédrico racional.

**1.2.4 Definición.** Un **semi-grupo**, es un conjunto no vacío  $\mathbf{S}$  con una operación asociativa,

$$+ : \mathbf{S} \times \mathbf{S} \rightarrow \mathbf{S}.$$

Un semi-grupo es llamado un **monoide** si a esta operación exigimos que sea conmutativa y existe un neutro, i.e, un elemento  $0 \in \mathbf{S}$  donde el cual

$$s + 0 = s, \forall s \in \mathbf{S},$$

y este satisface una ley de cancelación

$$s + x = t + x \implies s = t, \forall s, t, x \in \mathbf{S}.$$

**Lema 1.2.1.** Si  $\sigma$  es un cono en  $\mathbb{R}^n$ , entonces,  $\sigma \cap \mathbb{Z}^n$  es un monoide.

**Demostración.** De la definición de un cono  $\sigma$ ,  $x + y$  esta en  $\sigma$  si  $x, y \in \sigma$ , en particular  $x + y \in \sigma \cap \mathbb{Z}^n$  si  $x, y \in \sigma \cap \mathbb{Z}^n$ . El vector cero es el cero del monoide, **q.e.d.**

**1.2.5 Definición.** Un monoide se dice **finitamente generado** si existen  $a_1, \dots, a_r \in \mathbf{S}$ , llamados **generadores**, tal que

$$\mathbf{S} = \mathbb{Z}_{\geq 0} a_1 + \dots + \mathbb{Z}_{\geq 0} a_r.$$

Un sistema de generadores es llamado **mínimo** si ni uno de sus elementos es generado por los otros.

**Lema 1.2.2(Lema de Gordan).** Si  $\sigma$  es un cono reticular en  $\mathbb{R}^n$ , entonces, el monoide  $\sigma \cap \mathbb{Z}^n$  es finitamente generado.

**Demostración.** Siendo que  $\sigma$  es un cono poliédrico racional,  $\sigma = \text{Con}(T)$  para un conjunto finito  $T \subseteq M$ . Entonces  $K = \{\sum_{m \in T} \delta_m m : 0 \leq \delta_m < 1\}$  es una región acotada de  $M_{\mathbb{R}} \simeq \mathbb{R}^n$ , así que  $K \cap M$  es finito y así  $M \simeq \mathbb{Z}^n$ . Note que  $T \cup (K \cap M) \subseteq \sigma \cap \mathbb{Z}^n$ .

Afirmamos que  $T \cup (K \cap M)$  genera  $\sigma \cap \mathbb{Z}^n$  como un semi-grupo. Para probar esto, tomamos  $w \in \sigma \cap \mathbb{Z}^n$  y escribimos  $w = \sum_{m \in T} \lambda_m m$  donde  $\lambda_m \geq 0$ . Entonces  $\lambda_m = \lfloor \lambda_m \rfloor + \delta_m$  con  $\lfloor \lambda_m \rfloor \in \mathbb{N}$  y  $0 \leq \delta_m < 1$ , así que

$$w = \sum_{m \in T} \lfloor \lambda_m \rfloor m + \sum_{m \in T} \delta_m m.$$

El segundo sumando esta en  $K \cap M$  (recordemos que  $w \in M$ ). Se sigue que  $w$  es una combinación entera no negativa de elementos de  $T \cup (K \cap M)$  **q.e.d.**

**Lema 1.2.3.** Si  $\sigma$  tiene un ápice, entonces, el monoide  $\sigma \cap \mathbb{Z}^n$  tiene (salvo reenumeración) precisamente un **sistema mínimo de generadores**. ver ref[23].

**1.2.6 Definición.** Un monoide es llamado **saturado** si  $\alpha * x \in S$  para  $\alpha \in \mathbb{Z}_{\geq 0}$  y  $x \in G(S)$  implica que  $x \in S$ .

Donde  $G(S)$  es un grupo que tiene por generadores a  $a_1, \dots, a_k$  los cuales a su vez son generadores del monoide  $S$ .

**Lema 1.2.4.** Para todo cono  $\sigma$ , el monoide  $\sigma \cap \mathbb{Z}^n$  es saturado.

**Demostración.** De la definición misma de  $\sigma \cap \mathbb{Z}^n$  se sigue el resultado **q.e.d.**

### 1.3. Subdivisión estelar, operación estelar regular

**1.3.1 Definición.** Sea  $P'$  un politopo o un cono  $k$ -dimensional y  $\{(P_i)'\}$  el conjunto de todas sus caras propias e impropias, y sea  $\varphi : P' \rightarrow \varphi(P') \subset \mathbb{R}^n$  un homeomorfismo. Entonces  $P = \varphi(P')$  es llamado una  $k$ -celda en  $\mathbb{R}^n$  y son caras de  $\varphi((P_i)')$  si  $\varphi = id$  donde **id** es el homeomorfismo identidad que mapea el politopo de dimensión  $k$  de manera idéntica como un subconjunto del espacio euclídeo real  $n$  dimensional, entonces las celdas  $P$  son llamadas celdas rectas y todos los conos siempre son elegidos de esta manera.

**1.3.2 Definición.** Un conjunto finito  $C$  de celdas de un espacio euclidiano es llamado complejo de celdas si las siguientes condiciones son satisfechas:

- (a) Si  $F \in C$  y  $F_0$  es una cara de  $F$ , entonces  $F_0 \in C$ .
- (b) Si  $F, F' \in C$ , entonces  $F \cap F'$  es una cara común de  $F$  y  $F'$ .

**1.3.3 Definición.** Sea  $C$  un complejo de celdas,  $F \in C$ .

- (a) Definimos  $st(F, C) := \{F' \in C : F \subset F'\}$  y es llamado la **estrella** de  $F$  en  $C$ .
- (b) Y  $clst(F, C) := \{F'' \in C : F'' \subset F' \in st(F, C)\}$  se dice que este conjunto es la **estrella cerrada** de  $F$  en  $C$ .
- (c)  $link(F, C) := \{F' \in clst(F, C) : F' \cap F = \{0\}\}$  si  $C$  consiste de conos o
- (d)  $link(F, C) := \{F' \in clst(F, C) : F' \cap F = \emptyset\}$  si  $C$  no consiste de conos y es llamado link de  $F$  en  $C$  en ambos casos ver fig 1.1.

**1.3.4 Definición.** Sean  $\sigma$  y  $\sigma'$  dos conos con ápice 0, tal que  $lin(\sigma) \cap lin(\sigma') = \{0\}$ . Entonces nosotros llamamos

$$\sigma * \sigma' = Con(\sigma \cup \sigma')$$

la union de  $\sigma$  y  $\sigma'$ .

Estas operaciones son de especial interés en geometría algebraica en la definición tórica de **blowing up y blowing down** que serán abordadas en el capítulo 3.

**1.3.4 Definición.** Sea  $C$  un complejo de celdas,  $\sigma \in C, \sigma \neq \emptyset$ . Para un punto  $p \in Relint(\sigma)$ , llamamos la transición

$$C \longrightarrow (C \setminus st(\sigma, C) \cup p * (clst(F, C) \setminus st(\sigma, C))) =: s(p, \sigma)C$$

una **subdivisión estelar (o subdivisión elemental)** de  $C$  en dirección de  $p$ . La operación inversa  $s^{-1}(p; \sigma)$  definida por  $s^{-1}(p; \sigma)(s(p; \sigma)C) = C$  que es la subdivisión estelar inversa.

En palabras mas simples, lanzamos al exterior de una cara la estrella de una cara  $F$  y unimos un punto del interior relativo de  $F$  a la frontera de la estrella. El resultado es la subdivisión estelar, ver fig. 1.2 y fig.1.3.

**1.3.5 Definición.** Sea  $P \subseteq M_{\mathbb{R}}$  un politopo reticular y sea  $\sum_F$  el siguiente conjunto de conos  $\sum_F = \{\sigma_F \mid F \preceq P\}$ , de tal manera que si este conjunto cumple las siguientes propiedades,

- (a)  $\forall \sigma_F \in \sum_F$ , toda cara de  $\sigma_F$  es también una cara de  $\sum_F$ .
- (b) La intersección de cualquiera dos conos  $\sigma_F \cap \sigma_{F'}$  en  $\sum_F$  es una cara común de cada uno de ellos.

Entonces decimos que el conjunto de conos  $\sum_F$ , es un **abanico**.

**1.3.6 Definición.** Un **abanico es llamado regular** si todos sus conos son conos simplejos regulares y simplemente los denotaremos como  $\sum$  en el transcurso de este trabajo.

**1.3.7 Definición.** Sean  $\Sigma$  y  $\Sigma'$  abanicos en  $\mathbb{R}^n$ , y  $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$  un mapeo lineal tal que, para cada  $\sigma' \in \Sigma'$ , existe un  $\sigma \in \Sigma$  que satisface  $L(\sigma' \cap \mathbb{Z}^n) \subset \sigma \cap \mathbb{Z}^n$ . Entonces, decimos que  $L$  es también un **morfismo de abanicos**

$$L : \Sigma' \rightarrow \Sigma.$$

**1.3.8 Definición.** Una subdivisión estelar  $s(\rho, \sigma)$  (en dirección de  $p$  donde  $\rho = \mathbb{R}_{\geq 0}p$ ) es llamada **regular** si esta preserva regularidad. También escribimos  $s(p, \sigma) = s(\rho, \sigma)$ . Su operación inversa, también llamada regular.

**1.3.1 Teorema.** Sea  $P \subseteq M_{\mathbb{R}}$  un politopo reticular y  $\Sigma$  un cono regular asociado a  $P$ , elejimos  $\sigma \in \Sigma$ ,  $\sigma = \text{Con}(\alpha_1, \dots, \alpha_k)$  con  $\alpha_1, \dots, \alpha_k$  simples, y sea  $\rho = \lambda p$ ,  $\forall \lambda \in \mathbb{R}^+$ ,  $p$  simple,  $s(\rho, \sigma)$  es una **subdivisión estelar regular**  $\iff$

$$(*) \quad p = \alpha_1 + \dots + \alpha_k.$$

**Demostración.** Sea  $\sigma$  una cara de un cono regular  $n$ -dimensional  $\sigma'$  (donde  $\sigma'$  no necesariamente está en  $\Sigma$ ). De donde conseguimos

$$\sigma'_1 = \text{Con}(\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n)$$

tal que el  $\det(\sigma') = \pm 1$ . De esa manera  $s(\rho; \sigma)$  se descompone en conos  $n$ -dimensionales,

$$\begin{aligned} \sigma'_1 &= \text{Con}(p, \alpha_2, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n), \dots, \sigma'_k \\ &= \text{Con}(\alpha_1, \dots, \alpha_{k-1}, p, \alpha_{k+1}, \dots, \alpha_n). \end{aligned}$$

Si (1) es cierto, obtenemos

$$\begin{aligned} \det(\sigma'_i) &= \det(\alpha_1, \dots, \alpha_{i-1}, \alpha_1 + \dots + \alpha_k, \alpha_{i+1}, \dots, \alpha_n) \\ &= \det(\sigma') = \pm 1, \quad i = 1, \dots, k. \end{aligned}$$

Así, todos los nuevos conos que se construyen son de nuevo regulares.

Para probar la suficiencia, sean  $\sigma'_1, \dots, \sigma'_k$  conos regulares, y sea

$$p = \lambda_1 \alpha_1 + \dots + \lambda_k \alpha_k, \quad \lambda_1 > 0, \dots, \lambda_k > 0.$$

Entonces

$$\det(\alpha_1, \dots, \alpha_{i-1}, \lambda_1 \alpha_1 + \dots + \lambda_k \alpha_k, \alpha_{i+1}, \dots, \alpha_n) = \lambda_i \det(\sigma') = \pm \lambda_i = \pm 1$$

Siendo que  $\lambda_i > 0$ , esto implica que  $\lambda_i = 1$ ,  $i = 1, \dots, k$ ,  $p = \alpha_1 + \dots + \alpha_k$  es simple, puesto que de otra manera, para  $p = rq$ ,  $r > 1$ ,  $q \in \mathbb{Z}^n$ , tendríamos

$$\pm 1 = \det(rq - \alpha_2 - \dots - \alpha_k, \alpha_2, \dots, \alpha_n) = r \det(q, \alpha_2, \dots, \alpha_n) = rs,$$

para  $s \in \mathbb{Z}$ , lo cual es una contradicción de acuerdo a nuestra última igualdad. Por tanto (1) es cierto **q.e.d.**

Acabamos de revisar dos definiciones de manera muy combinatoria que para el último capítulo serán de interés particular, estas definiciones nos darán una intuición geométrica en la construcción de blow up tórico, pero a su vez daremos la mismas definiciones pero en la terminología de conos y abanicos y es la que oficialmente estaremos usando en el transcurso de este trabajo.

**1.3.9 Definición.(Estrella)** Sea  $\sum_0$  un abanico que pertenece a una retícula  $N$ , y  $\sigma_0$  cono reticular de  $\sum_0$ . Definimos la estrella de  $\sigma_0$  en  $\sum_0$  como

$$\text{st } \sum_0(\sigma_0) := \{\sigma \in \sum_0 : \sigma_0 \preceq \sigma\}.$$

**1.3.10 Definición.(Subdivisión estelar).** Sea  $p$  un vector reticular primitivo (con coordenadas primos relativos). Decimos que un abanico  $\sum_1$  es una subdivisión estelar de  $\sum_0$  en dirección del vector reticular  $p$  si este abanico se puede ver como el siguiente conjunto.

$$\Sigma_1 := (\Sigma_0 \setminus st_{\Sigma_0}(\sigma_0)) \cup \{\tau + Con(p) : \tau \not\subseteq \sigma \in st_{\Sigma_0}(\sigma_0)\}.$$

**Lema 1.2.5.** Sea  $\sigma = Con(\alpha_1, \dots, \alpha_k)$ ,  $k > 1$  un cono regular, y sea  $\sigma$  una cara de un cono regular  $n$ -dimensional  $\sigma' = Con(\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n)$ , y sea  $p = \alpha_1 + \dots + \alpha_k$ . Entonces para  $\sigma_i = Con(\alpha_1, \dots, \alpha_{i-1}, p, \alpha_{i+1}, \dots, \alpha_k, \dots, \alpha_n)$ ,  $i \in \{1, \dots, k\}$ , ponemos  $\sigma^{\vee} = Con(y_1, \dots, y_k, \dots, y_n)$  y  $\sigma_i^{\vee} = Con(y'_1, \dots, y'_{i-1}, y'_i, y'_{i+1}, \dots, y'_k, \dots, y'_n)$  tal que;

$$(*) \quad -y_i + y_j = y'_j, \quad j = 1, \dots, i-1, i+1, \dots, k.$$

**Demostración.** Hacemos los generadores  $\alpha_i = e_i$ ,  $i = 1, \dots, n$ . Entonces,  $\sigma_i^{\vee}$  tiene como generadores los hiperplanos soporte,  $y'_1 = e_1 - e_i, \dots, y'_{i-1} = e_{i-1} - e_i, y_i = e_i, y'_{i+1} = e_{i+1} - e_i, \dots, y'_k = e_k - e_i, y_{k+1} = e_{k+1}, \dots, y_n = e_n$ . Ahora mediante una transformación unimodular conseguimos las relaciones (\*), **q.e.d.**

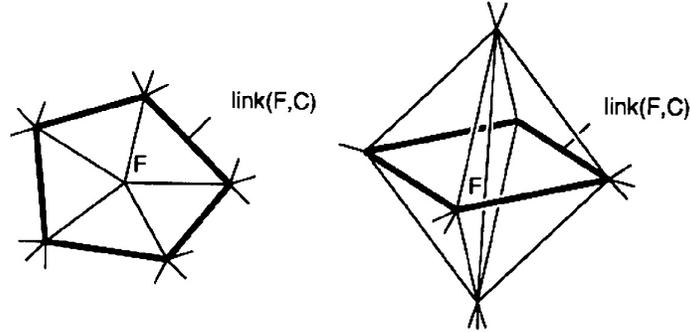


Figura 1.1: Link de la cara F en un politopo C; ref. [13].

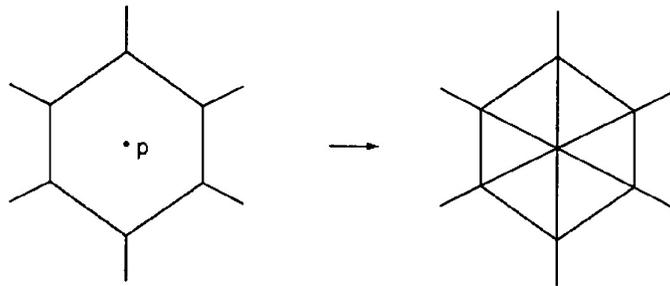


Figura 1.2: Subdivisión estelar regular; ref. [13].

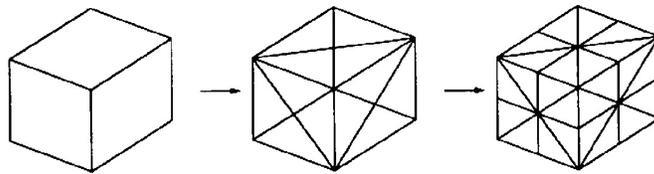


Figura 1.3: Subdivisión estelar regular en un cubo; ref. [13].

# Capítulo 2

## Variedades tóricas afines

En este capítulo una vez desarrollada toda la herramienta necesaria que será de gran importancia junto con los conceptos de geometría algebraica para el fundamento teórico y desarrollo de variedades tóricas, con el objetivo de fundamentar y desarrollar la técnica de modificación tórica como método de resolución de singularidades, también exponemos el blow up usual, con la finalidad de justificar su aplicaciones en el desarrollo de la parte 2, todas las definiciones, teoremas y lemas de este capítulo incluyendo su demostraciones fueron tomadas de los siguientes autores; [13] G. Ewald, [27] B. Sturmfels, [11] D. Cox, J. Little, H. Schenck, a excepción de las pruebas de las afirmaciones que fueron dejadas por estos autores, fueron probadas por el autor de este trabajo así como del lema 2.1.1.1, lema 2.2.1 y la primera parte del teorema 2.4.1.

### 2.1. Toro algebraico y morfismos asociados

**El Toro.** La variedad afín  $(\mathbb{C}^*)^n$  es un grupo bajo la multiplicación compleja coordenada a coordenada. Un **toro**  $\mathbf{T}$  es una variedad afín isomorfa a  $(\mathbb{C}^*)^n$ . A continuación revisamos este hecho de la siguiente manera.

Sea  $R = \mathbb{C}[\zeta_1, \dots, \zeta_{2n}]$  el anillo de polinomios complejos en  $2n$  variables,  $n \geq 1$ . Entonces,

$$\mathbf{a} := R(\zeta_1 \zeta_{n+1} - 1) + \dots + R(\zeta_n \zeta_{2n} - 1),$$

es un ideal en  $R$ . Para  $z_i := \zeta_i + \mathbf{a} \in R/\mathbf{a}$ ,  $i = 1, \dots, 2n$ . Así tenemos que (y donde escribimos  $\mathbf{1}$  simplemente en vez de  $[1] \in R/\mathbf{a}$ )

$$z_j * z_{n+j} = 1,$$

donde  $*$  denota la multiplicación usual en el campo de los números complejos  $\mathbb{C}$  y en lo siguiente solo escribimos el productos de dos de ellos como,  $z_j z_{n+j}$ . Y así  $(z_j)^{-1} = z_{n+j}$  para  $j = 1, \dots, n$ .

De este hecho el conjunto de ecuaciones

$$\zeta_i \zeta_{n+i} = 1, i = 1, \dots, n.$$

definen la variedad afín  $V(\zeta_1 \zeta_{n+1} - 1, \dots, \zeta_n \zeta_{2n} - 1)$ . Para  $n = 1$  obtenemos una hipérbola compleja.

Ahora sea  $\mathbf{T}$  el conjunto de puntos

$$\mathbf{T} := \{(z_1, \dots, z_n) \in \mathbb{C}^n : z_i \neq 0, i = 1, \dots, n\} = (\mathbb{C} \setminus \{0\})^n,$$

afirmamos que este conjunto es isomorfo a  $V(\zeta_1 \zeta_{n+1} - 1, \dots, \zeta_n \zeta_{2n} - 1)$ , bajo la proyección  $\mathbb{C}^{2n} \rightarrow \mathbb{C}^n$  mas explícitamente la proyección restringida  $P_T : V(\zeta_1 \zeta_{n+1} - 1, \dots, \zeta_n \zeta_{2n} - 1) \rightarrow \mathbf{T}$  y morfismo inverso definido por

$$(P_T)^{-1} : (z_1, \dots, z_n) \rightarrow (z_1, \dots, z_n, (z_1)^{-1}, \dots, (z_n)^{-1})$$

lo cual muestra que cada uno son morfismos inversos uno del otro **q.e.d.** Como consecuencia de este hecho tenemos la siguiente definición.

**2.1.1 Definición.** El conjunto  $\mathbf{T} := (\mathbb{C} \setminus \{0\})^n =: (\mathbb{C}^*)^n$  es llamado el **n-toro algebraico complejo**.

Ahora estudiaremos dos morfismos de interés asociados al toro  $\mathbf{T}$ , a saber son su carácter y subgrupos uniparamétricos. Los discutimos brevemente.

**2.1.2 Definición.** Un **carácter** de un toro  $\mathbf{T}$  es un homomorfismo de grupos  $\chi : T \rightarrow \mathbb{C}^*$ . Por ejemplo,  $m = (a_1, \dots, a_n) \in \mathbb{Z}^n$  da un carácter  $\chi : (\mathbb{C}^*)^n \rightarrow \mathbb{C}^*$  definido por

$$\chi^m(t_1, \dots, t_n) = (t_1)^{a_1} * \dots * (t_n)^{a_n}$$

Uno puede mostrar que todos los caracteres de  $(\mathbb{C}^*)^n$  son de esta manera. Así los caracteres de  $(\mathbb{C}^*)^n$  forman un grupo isomorfo a  $\mathbb{Z}^n$ .

Es importante decir que dada la retícula  $M$  y  $m \in M$ , de ésta retícula podemos conseguir el carácter  $\chi^m : T \rightarrow \mathbb{C}^*$ .

**2.1.3 Definición.** Un **subgrupo uniparamétrico** de un toro  $\mathbf{T}$  es un homomorfismo de grupos  $\lambda : \mathbb{C}^* \rightarrow T$ . Por ejemplo,  $u = (b_1, \dots, b_n) \in \mathbb{Z}^n$  produce un subgrupo uniparamétrico  $\lambda^u : \mathbb{C}^* \rightarrow (\mathbb{C}^*)^n$  definido por

$$\lambda^u(t) = (t^{b_1}, \dots, t^{b_n}).$$

todos los subgrupos uniparamétricos de  $(\mathbb{C}^*)^n$  se producen de esta manera. Se sigue que el grupo de subgrupos uniparamétricos de  $(\mathbb{C}^*)^n$  es isomorfo a  $\mathbb{Z}^n$ . Para un toro  $\mathbf{T}$ , los subgrupos uniparamétricos forman un grupo abeliano libre  $N$  de rango igual a la dimensión de  $\mathbf{T}$ . Como con el grupo de caracteres, un elemento  $u \in N$  produce un subgrupo uniparamétrico  $\lambda^u : \mathbb{C}^* \rightarrow T$ .

Hay un apareamiento bilineal  $\langle \cdot, \cdot \rangle : M \times N \rightarrow \mathbb{Z}$  definido como sigue.

(a) (Intrínseco). Dado un carácter  $\chi^m$  y un subgrupo uniparamétrico  $\lambda^u$ , la composición  $\chi^m \circ \lambda^u : \mathbb{C}^* \rightarrow \mathbb{C}^*$  es carácter de  $\mathbb{C}^*$ , que está dado por  $t \mapsto t^l$  para algún  $l \in \mathbb{Z}$ . Entonces  $\langle m, u \rangle = l$ .

(b) (Concreto). Si  $T = (\mathbb{C}^*)^n$  con  $m = (a_1, \dots, a_n) \in \mathbb{Z}^n$ ,  $u = (b_1, \dots, b_n) \in \mathbb{Z}^n$ , entonces uno calcula  $\langle m, u \rangle = \sum_{i=1, \dots, n} a_i b_i$ , i.e., el apareamiento es el producto punto usual.

Se sigue que los caracteres y subgrupos uniparamétricos de un toro  $\mathbf{T}$  forman grupos abelianos libres de rangos finitos. En términos de producto tensorial, uno obtiene un isomorfismo canónico  $N \otimes_{\mathbb{Z}} \mathbb{C}^* \simeq T$  vía  $u \mapsto \lambda^u(t)$ . Así a veces es costumbre escribir un toro como  $T_N$ .

Desde este punto de vista, elegir un isomorfismo  $T_N \simeq (\mathbb{C}^*)^n$  induce bases duales de  $\mathbf{M}$  y  $\mathbf{N}$ , i.e., isomorfismos  $M \simeq \mathbb{Z}^n$  y  $N \simeq \mathbb{Z}^n$ .

### 2.1.1. Bases de Hilbert

Ahora fijemos una retícula  $N \simeq \mathbb{Z}^n$  y su retícula dual  $M = N^\vee$ . Sea  $\sigma$  un cono reticular correspondiente al espacio vectorial  $N_{\mathbb{Q}} \simeq \mathbb{Q}^n$  y sea  $\sigma^\vee$  su cono dual el cual es  $n$ -dimensional en el espacio vectorial dual y denotemos el monoide  $\sigma^\vee \cap M$  por  $S_\sigma$  el cual es finitamente generado (Lema de Gordan).

**2.1.1.1 Lema. (Bases de Hilbert).** El cono  $\sigma \subseteq N$  es  $n$ -dimensional sí y solo sí su cono dual  $\sigma^\vee \subseteq M$  es fuertemente convexo. En este caso el monoide  $S_\sigma$  tiene un único conjunto finito mínimo de generadores  $H \subseteq M \simeq \mathbb{Z}^d$  y estos son irreducibles.

**Demostración.** La necesidad y suficiencia fueron probados en el lema 1.1.4 y la definición 1.1.6. Para probar la existencia del conjunto  $H$  tenemos por hipótesis que el cono  $\sigma^\vee$  es fuertemente convexo, i.e.,  $\sigma^\vee \cap (-\sigma^\vee) = \{0\}$ , intersectando con  $M$  en ambos lados de la igualdad, se sigue que,  $S_\sigma \cap (-S_\sigma) = \{0\}$ ; a continuación demostramos que bajo el orden parcial  $\preceq$  el cono  $S_\sigma$  tiene a  $S_\sigma \cap (-S_\sigma)$  como la menor cara. Sea  $\tau = S_\sigma \cap (-S_\sigma)$  y afirmamos que  $\tau$  es una cara de  $S_\sigma$  ya que  $S_\sigma$  es cara de sí misma y  $-S_\sigma$  también lo es, del lema 1.1.1, la intersección de dos caras es una cara de igual manera. Ahora mostremos que esta es la menor y es única.

Del lema de separación 1.1.3; existe un hiperplano  $H_m$  con  $m \in \text{Relint}(S_\sigma^\vee \cap (-S_\sigma)^\vee)$  tal que,

$$S_\sigma \cap (-S_\sigma) = \tau = H_m \cap S_\sigma = H_m \cap (-S_\sigma).$$

Ahora suponemos que  $\tau'$  es cara de  $S_\sigma$  y suponemos que es menor cara que  $\tau$ , i.e.,  $\tau' \preceq \tau = S_\sigma \cap (-S_\sigma)$ , esto implica que existe un hiperplano  $H_m$  con  $m \in M$ , tal que,

$$\tau' = H_m \cap (S_\sigma \cap (-S_\sigma)) = (H_m \cap S_\sigma) \cap (H_m \cap (-S_\sigma)) = \tau \cap \tau = \tau.$$

Y de aquí conseguimos el resultado; y ahora por el lema de Gordan 1.2.2  $\tau \preceq S_\sigma$  es finitamente generado y este conjunto es mínimo por el resultado anterior; solo nos resta mostrar que estos generadores son irreducibles.

Supogamos que  $m \in \tau$  no es irreducible. Entonces  $m = m_1 + m_2$  donde  $m_1$  y  $m_2$  son elementos no cero de  $S_\sigma$  y tomemos  $u \in \sigma \cap (\mathbb{N} \setminus \{0\})$ . Se sigue que

$$\langle m, u \rangle = \langle m_1, u \rangle + \langle m_2, u \rangle$$

con  $\langle m_1, u \rangle, \langle m_2, u \rangle \in \mathbb{N} \setminus \{0\}$ , tal que

$$\langle m_1, u \rangle < \langle m, u \rangle \text{ y } \langle m_2, u \rangle < \langle m, u \rangle.$$

Usando inducción sobre el elemento  $\langle m, u \rangle$ , concluimos que todo elemento de  $S_\sigma$  es una suma de elementos irreducibles, denotemos por  $H$  a este conjunto de elementos tal que generan a  $S_\sigma$ . Y con

ello concluimos la prueba, **q.e.d.**.

El conjunto mínimo de generadores  $H$  de un monoide fuertemente convexo  $S_\sigma$  es llamado las **bases de Hilbert**. En el apéndice A damos un algoritmo mediante **bases de Grobner** para su cálculo explícito.

## 2.2. Variedades tóricas afines y politopo de Newton

**2.2.1 Definición.** Los elementos de

$$\mathbb{C}[t, t^{-1}] := \mathbb{C}[t_1, \dots, t_n, (t_1)^{-1}, \dots, (t_n)^{-1}] = \mathbb{C}[\zeta_1, \dots, \zeta_{2n}]/\mathbf{a}$$

son llamados **polinomios de Laurent**, mientras que los términos

$$\lambda * t^a = \lambda * t^{a_1} * \dots * t^{a_n}, \text{ para } a = (a_1, \dots, a_n) \in \mathbb{Z}^n, \lambda \in \mathbb{C}^*,$$

se dice que son **monomios de Laurent**.

**2.2.2 Definición.** El **soporte** de un polinomio de Laurent  $f = \sum_{finito} \lambda_a t^a$  es definido como

$$\text{supp}(f) := \{a \in \mathbb{Z}^n : \lambda_a \neq 0\}.$$

**2.2.3 Definición.** Sea  $\omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n$  cualquier vector fijo y para un polinomio  $f = \sum \lambda_i * t^{a_i}$ ; definimos la **forma inicial**  $in_\omega(f)$  como la suma de todos los términos  $\lambda_i * t^{a_i}$  tal que el producto interno  $\langle \omega, a_i \rangle$  es **maximal**. Para un ideal  $I$  definimos el **ideal inicial** como el ideal generado por todas las formas iniciales:

$$in_\omega(I) := \langle in_\omega(f) : f \in I \rangle.$$

**2.2.4 Definición.** A todo polinomio  $f = \sum_{i=1, \dots, m} \lambda_i * t^{a_i}$  en el anillo  $\mathbb{C}[\mathbf{t}]$  podemos asociar el politopo  $New(f) := Conv\{a_i : i = 1, \dots, m\}$  en  $\mathbb{R}^n$  llamado el **politopo de Newton**. La operación algebraica de multiplicación corresponde a la operación de suma de Minkowski en politopos, para mas detalle, ver, ref.[23].

**2.2.1 Lema.**  $New(f * g) = New(f) + New(g)$ .

**Demostración.** Es suficiente mostrar que ambos politopos tienen los mismos vértices. Primero notamos la siguiente relación general entre caras y formas iniciales:

$$\tau_\omega(New(f)) = New(in_\omega(f))$$

La demostración ya es clara para monomios  $f = t^a$  y  $g = t^b$  ya que  $t^a * t^b = t^{a+b}$ . Ahora sean  $f = \sum_{i=1, \dots, n} \lambda_i * t^{a_i}$  y  $g = \sum_{j=1, \dots, n} \lambda_j * t^{b_j}$  polinomios de Laurent. Y a continuación probamos la igualdad de arriba. Donde  $\tau_\omega(New(f))$  denota una cara del politopo de Newton asociado al soporte del polinomio  $f$  y  $\omega \in \mathbb{R}^n$  en particular ahora será un vector normal a esta cara; si la cara es de dimensión cero, i.e., un punto en  $\mathbb{R}^n$  este vector se elige en dirección de este punto. Ahora primero mostramos que  $New(in_\omega(f)) \subseteq \tau_\omega(New(f))$ . Reescribiendo  $New(in_\omega(f)) = New(\max\{\langle \omega, a_i \rangle : a_i \in \text{supp}(f)\}_{i=1, \dots, n})$ . Y solo recordamos que por definición de ideal inicial el conjunto de los elementos  $\langle \omega, a_i \rangle$  son de magnitud máxima y que tengan mismo valor  $\forall i = 1, \dots, n$ , i.e.,  $\langle \omega, a_1 \rangle = \dots = \langle \omega, a_n \rangle$ . Partiendo de este hecho tomamos una combinación convexa  $s_1 a_1 + \dots + s_n a_n \in New(in_\omega(f))$  y realizando producto interno con el vector normal  $\omega$  tenemos,

$$\begin{aligned} \langle \omega, s_1 a_1 + \dots + s_n a_n \rangle &= \langle \omega, s_1 a_1 \rangle + \dots + \langle \omega, s_n a_n \rangle = s_1 \langle \omega, a_1 \rangle + \dots + s_n \langle \omega, a_n \rangle = \\ &\langle \omega, a_i \rangle (s_1 + \dots + s_n) = \langle \omega, a_i \rangle = c \end{aligned}$$

De donde  $s_1 + \dots + s_n = 1$  se sigue de la definición de envolvente convexo; también se tiene  $\langle \omega, a_i \rangle = c$  con  $c \in \mathbb{R}$  y  $\forall a_i$  lo cual define la ecuación de un hiperplano con vector normal  $\omega$  y donde los vectores  $a_i$  pertenecen a este hiperplano, que lo denotaremos por  $H_p$  para algún  $p \in \mathbb{R}^n$ , i.e; existe  $H_p$  tal que,  $H_p \cap New(in_\omega(f))$  donde  $\tau_\omega$  es este conjunto y escribimos  $\tau_\omega = H_p \cap New(in_\omega(f))$ ; de la convexidad de esta cara, cualquier combinación convexa de elementos de ella también pertenece a esta cara, de esto concluimos que  $s_1 a_1 + \dots + s_n a_n \in \tau_\omega$ .

Y viceversa, para probar la otra inclusión; elegimos vectores  $a_1, \dots, a_n \in \tau_\omega = H_p \cap New(in_\omega(f))$  de tal manera que satisfacen la ecuación del hiperplano  $\langle \omega, a_i \rangle = c \forall i, c \in \mathbb{R}$  y  $a_i \in New(in_\omega(f))$ ,

con esto probamos la igualdad.

Ahora utilizando lo probado anteriormente, veamos que:

$$\tau_\omega(New(f * g)) = \tau_\omega(New(f) + New(g))$$

Desarrollando la primera igualdad conseguimos:

$$\begin{aligned} \tau_\omega(New(f * g)) &= New(in_\omega(f * g)) = \\ &New(max\{\langle \omega, a_i + b_j \rangle : a_i + b_j \in supp(f * g)\}_{i,j=1,\dots,n}) = \\ &New(max\{\langle \omega, a_i \rangle + \langle \omega, b_j \rangle : a_i \in supp(f), b_j \in supp(g)\}_{i,j=1,\dots,n}) = \\ &New(max\{\langle \omega, a_i \rangle : a_i \in supp(f)\}_{i=1,\dots,n} + max\{\langle \omega, b_j \rangle : b_j \in supp(g)\}_{j=1,\dots,n}) = \\ &New(max\{\langle \omega, a_i \rangle : a_i \in supp(f)\}_{i=1,\dots,n}) + New(max\{\langle \omega, b_j \rangle : b_j \in supp(g)\}_{j=1,\dots,n}) = \\ &New(in_\omega(f)) + New(in_\omega(g)) = \tau_\omega(New(f)) + \tau_\omega(New(g)) = \tau_\omega(New(f) + New(g)). \end{aligned}$$

Concluimos que  $\tau_\omega$  es cara de los mismos politopos en cuestión y tienen el mismo conjunto de vértices y así ellos son iguales **q.e.d.**

Donde la última igualdad se prueba fácilmente de la definición de suma de Minkowski; ver, ref. [27].

Es importante destacar que a partir de este momento estamos interesados de manera concreta en **politopos de Newton del soporte de polinomios de Laurent** i.e,  $New(supp(f)) = Conv\{a \in \mathbb{Z}^n : \lambda_a \neq 0\}$  y  $f \in \mathbb{C}[t, t^{-1}]$  importantes para nuestra construcción final de variedades tóricas afines.

**2.2.2 Lema y Definición.** Sea el anillo  $\mathbb{C}[t, t^{-1}] = \mathbb{C}[t_1, \dots, t_n, (t_1)^{-1}, \dots, (t_n)^{-1}] = \mathbb{C}[\zeta_1, \dots, \zeta_{2n}]/\mathfrak{a}$  como ya se había definido anteriormente. De modo que para un **cono reticular**, el anillo

$$R_\sigma := \{f \in \mathbb{C}[t, t^{-1}] : supp(f) \subset \sigma\}$$

es una **álgebra monomial finitamente generada**

**Demostración.** Siendo que  $\sigma$  es un cono reticular por tanto es poliédrico racional, así por el lema de

Gordan 1.2.2, es finitamente generado, y por tanto a su vez  $\text{supp}(f) \subseteq \sigma$  es generado por un conjunto finito de elementos,  $\{a_1, \dots, a_r\} \subseteq \sigma$  tomamos  $f \in R_\sigma$  y lo escribimos como  $f = \sum_{i=1, \dots, r} \lambda_i t^{a_i}$ ; ahora sea  $a_i \in \text{supp}(f)$ , de tal manera que los podemos escribir como,  $a_i = a_1 + \dots + a_r$ ; ahora reescribimos a  $f$ , donde siempre podemos ordenarlo de la siguiente manera,

$$f = \sum_i \lambda_i t^{a_i} = \lambda_1 t^{a_1} t^{a_2 + \dots + a_r} + \dots + \lambda_r t^{a_r} t^{a_2 + \dots + a_{r-1}} = t^{a_1} f_1 + \dots + t^{a_r} f_r.$$

donde  $f, f_1, \dots, f_r \in R_\sigma$ ; así concluimos que  $f \in R_\sigma = t^{a_1} * R_\sigma + \dots + t^{a_r} * R_\sigma$ ; por tanto  $R_\sigma$  es un álgebra monomial finitamente generada de la finitud inducida por  $\{a_1, \dots, a_r\} \subseteq \sigma$ , con generadores dados por el siguiente conjunto  $\{t^{a_1}, \dots, t^{a_r}\} \subseteq R_\sigma$ ; y así obtenemos el resultado; **q.e.d.**

Cabe notar que el anillo  $R_\sigma$  es subanillo de  $\mathbb{C}[t, t^{-1}]$  y no tiene divisores de cero. Recordamos que toda  $\mathbb{C}$ -álgebra finitamente generada sin divisores de cero define una **variedad afín abstracta**, a saber su **espectro maximal**.

**2.2.5 Definición.** Una **variedad tórica afín** es una variedad afín irreducible  $X$  que contiene un toro  $T_N \simeq \mathbb{C}^{*n}$  como un subconjunto abierto de Zariski tal que la acción de  $T_N$  sobre sí mismo se extiende a una acción algebraica de  $T_N$  sobre  $X$ . (Por acción algebraica, entenderemos que existe un morfismo definido de  $T_N \times X \rightarrow X$ ).

Como definición alternativa en este trabajo diremos que para un cono reticular  $\sigma$ , el espectro maximal  $X_\sigma = \text{spec}(R_\sigma)$  es llamado **variedad tórica afín abstracta (o encajamiento del toro)**.

A continuación demostramos un teorema relacionado con esta variedad abstracta y algunos ejemplos son desarrollados.

**Teorema 2.2.1** Sea  $\sigma \subseteq N_{\mathbb{R}} \simeq \mathbb{R}^n$  un cono poliédrico racional con monoide  $S_\sigma = \sigma^\vee \cap M$  con  $M$  retícula dual de  $N$ . Entonces

$$X_\sigma = \text{spec}(\mathbb{C}[S_\sigma]) = \text{spec}(\mathbb{C}[\sigma^\vee \cap M])$$

es una variedad tórica afín. Además,

$\dim X_\sigma = n \iff$  el toro de  $X_\sigma$  es  $T_N = N \otimes_{\mathbb{Z}} \mathbb{C}^* \iff \sigma$  es fuertemente convexo.

**Demostración.** Por el lema de Gordan 1.2.2 ,  $X_\sigma$  es una variedad tórica afín cuyo toro tiene retícula de caracteres  $\mathbb{Z}S_\sigma \subseteq M$ . Para estudiar  $\mathbb{Z}S_\sigma$ , notemos que

$$\mathbb{Z}S_\sigma = S_\sigma - S_\sigma = \{m_1 - m_2 : m_1, m_2 \in S_\sigma\}.$$

Ahora suponemos que  $km \in \mathbb{Z}S_\sigma$  para algún  $k > 1$  y  $m \in M$ . Entonces  $km = m_1 - m_2$  para  $m_1, m_2 \in S_\sigma = \sigma^\vee \cap M$ . Así  $m_1$  y  $m_2$  están en el conjunto convexo  $\sigma^\vee$ , tenemos

$$m + m_2 = \frac{1}{k}m_1 + \frac{k-1}{k}m_2 \in \sigma^\vee.$$

se sigue que  $m = (m + m_2) - m_2 \in \mathbb{Z}S_\sigma$ , tal que  $M/\mathbb{Z}S_\sigma$  es libre de torsión. Así

$$\text{el toro de } X_\sigma \text{ es } T_N \iff \mathbb{Z}S_\sigma = M \iff \text{rango } \mathbb{Z}S_\sigma = n.$$

Y de el hecho de que  $\sigma$  es fuertemente convexo  $\iff \dim \sigma^\vee = n$  (definición 1.1.6 y lema 1.1.4), solo resta mostrar que

$$\dim X_\sigma = n \iff \text{rango } \mathbb{Z}S_\sigma = n \iff \dim \sigma^\vee = n.$$

La primera equivalencia se sigue de que la dimensión de una variedad tórica afín es la dimensión de su toro, la cual es la dimensión de su retícula de caracteres. La prueba de la segunda equivalencia se sigue del hecho de que  $\sigma^\vee - \sigma^\vee$  es el menor subespacio que contiene  $\sigma^\vee$  y de la definición 1.1.6 obtenemos el resultado. **q.e.d.**

**Ejemplos de interés: (1)** Fijamos  $0 \leq r \leq n$  y sea  $\sigma$  el cono reticular generado por  $\sigma = \text{Con}(e_1, \dots, e_r) \subseteq \mathbb{R}^n$ . Entonces

$$\sigma^\vee = \text{Con}(e_1, \dots, e_r, \pm e_{r+1}, \dots, \pm e_n),$$

el hecho anterior se sigue de la construcción de conos duales dada en **teorema 1.1.1** y la correspondiente variedad afín tórica es

$$X_\sigma = \mathbf{Spec} \mathbb{C}[x_1, \dots, x_r, (x_{r+1})^{\pm 1}, \dots, (x_{n+r})^{\pm 1}] \simeq \mathbb{C}^r \times (\mathbb{C}^*)^{n-r}.$$

Y esto proviene de los siguientes hechos:

$$\begin{aligned}
& \mathbb{C}[x_1, \dots, x_n, (x_{n+1})^{\pm 1}, \dots, (x_{n+r})^{\pm 1}] \\
& \simeq \mathbb{C}[x_1, \dots, x_n] \otimes_{\mathbb{C}} \mathbb{C}[(x_{n+1})^{\pm 1}, \dots, (x_{n+r})^{\pm 1}] \\
& \simeq \mathbb{C}[x_1, \dots, x_n] \mathbb{C}[(x_{n+1})^{\pm 1}, \dots, (x_{n+r})^{\pm 1}] \\
& = \mathbb{C}[V_1 \times V_2] \Rightarrow \text{Spec}(\mathbb{C}[V_1 \times V_2]) = V_1 \times V_2.
\end{aligned}$$

Los hechos anteriores son ciertos siendo que  $\mathbb{C}[V_1] \otimes_{\mathbb{C}} \mathbb{C}[V_2]$  es una  $\mathbb{C}$ -álgebra sin nilpotentes, entonces  $\mathbb{C}[V_1 \times V_2]$  es su anillo coordenado. Esto implica el hecho general que si  $\sigma \subseteq N_{\mathbb{R}} \simeq \mathbb{R}^n$  es un **cono regular** de dimensión  $r$ , entonces  $X_{\sigma} \simeq \mathbb{C}^r \times (\mathbb{C}^*)^{n-r}$ .

**Ejemplo (2).** Consideremos el plano afín  $\mathbb{C}^2$  como variedad tórica afín  $X_{\sigma}$  para  $\sigma := \text{Con}(e_1, e_2) \subset \mathbb{R}^2$ , calculamos generadores para el monoide  $\sigma \cap \mathbb{Z}^2$  con bases de Hilbert lo cuales son:  $a_1 = e_1, a_2 = e_2, a_3 = e_1 + e_2$  obtenemos la representación del anillo coordenado  $\mathbb{C}[\zeta_1, \zeta_2, \zeta_3]/(\zeta_1\zeta_2 - \zeta_3) = \mathbb{C}[u_1, u_2, u_3]$ , obtenemos la variedad tórica  $X_{\sigma} = \langle u_3 - u_1u_2 \rangle$  la cual es una superficie cuadrática en  $\mathbb{C}^3$ .

### 2.3. Ideales tóricos

Dado un toro  $T_N$  con retícula de caracteres  $M$ , y un conjunto finito  $A = \{m_1, \dots, m_n\} \subseteq \mathbb{Z}^d$  para cada  $m_i$  conseguimos caracteres  $\chi^{m_i} : T_N \rightarrow \mathbb{C}^*$ . Entonces consideramos el homomorfismo

$$\Phi_A : T_N \rightarrow (\mathbb{C}^*)^n$$

definido por

$$\Phi_A(t) = (\chi^{m_1}(t), \dots, \chi^{m_n}(t)) \in (\mathbb{C}^*)^n.$$

Ahora sea  $X_A \subseteq \mathbb{C}^n = \mathbf{Spec}(\mathbb{C}[x_1, \dots, x_n])$  una variedad tórica afín que viene de un conjunto finito  $A = \{m_1, \dots, m_n\} \subseteq M \simeq \mathbb{Z}^d$  y es construida tomando la cerradura de Zariski de la imagen en  $\Phi_A$  como vemos a continuación. Nosotros podemos describir el ideal  $\mathbf{I} \subseteq \mathbb{C}[x_1, \dots, x_n]$  como sigue.  $\Phi_A$

induce un morfismo de retículas

$$\hat{\Phi}_A : \mathbb{Z}^n \longrightarrow M$$

que manda las bases estándar  $e_1, \dots, e_n$  a  $m_1, \dots, m_n$ . Sea  $L$  el núcleo de este morfismo, tal que tenemos la siguiente sucesión exacta

$$0 \longrightarrow L \longrightarrow \mathbb{Z}^n \longrightarrow M.$$

i.e., los elementos de  $l = (l_1, \dots, l_n) \in L$  satisfacen  $\sum_{i=1, \dots, n} l_i m_i = 0$  y notamos las relaciones entre los  $m_i$ .

Dado  $l = (l_1, \dots, l_n) \in L$ , y colocamos

$$l_+ = \sum_{l_i \geq 0} l_i e_i \text{ y } l_- = \sum_{l_i < 0} l_i e_i.$$

Donde  $l$  siempre se puede escribir de la manera  $l = (l_+ - l_-)$ ,  $l_+, l_- \in \mathbb{Z}_+^n$ . Se sigue fácilmente que el binomio

$$x^{l_+} - x^{l_-} = \prod_{l_i > 0} (x_i)^{l_i} - \prod_{l_i < 0} (x_i)^{-l_i}$$

se anula en la imagen de  $\Phi_A$  y así en  $X_A$  de ahí vemos que  $X_A$  es la cerradura de Zariski de la imagen.

**Proposición 2.3.1.** El ideal de la variedad tórica afín  $X_A \subseteq \mathbb{C}^n$  es

$$\mathbf{I}(X_A) = \langle x^{l_+} - x^{l_-} \mid l \in L \rangle = \langle x^\alpha - x^\beta \mid \alpha, \beta \in (\mathbb{Z}_+)^n, \alpha - \beta \in L \rangle.$$

**Demostración.** La prueba de la igualdad derecha se sigue de lo siguiente; sea  $\alpha - \beta \in L$ , entonces  $\hat{\Phi}_A(\alpha - \beta) = 0$ , si escribimos  $l = \alpha - \beta \in L$  y puesto que todo elemento  $l \in \mathbb{Z}^n$  puede ser escrito

como un vector diferencia de su parte positiva y negativa, escribimos  $l = l_+ - l_-$ , entonces conseguimos de lo anterior que  $\hat{\Phi}_A(l_+) = \hat{\Phi}_A(l_-) \implies x^{l_+} = x^{l_-}$  y de esa manera  $x^{l_+} - x^{l_-} = 0$  genera el ideal  $\langle x^{l_+} - x^{l_-} | l \in L \rangle$ . Para mostrar la primera igualdad, denotemos  $I_L$  al ideal  $\langle x^{l_+} - x^{l_-} | l \in L \rangle$ , y notemos que  $I_L \subseteq \mathbf{I}(X_A)$ . Probamos la inclusión opuesta. Escogemos un orden monomial en el anillo  $\mathbb{C}[x_1, \dots, x_n]$  por ejemplo  $<_L$  o  $<_{pos}$  léxicográfico o postlexicográfico y escogemos un isomorfismo de  $T_N \simeq (\mathbb{C}^*)^n$ . Así podemos asumir que  $M = \mathbb{Z}^n$  y el mapeo  $\Phi : (\mathbb{C}^*)^n \rightarrow \mathbb{C}^n$  esta dado por monomios de Laurent  $t^{m_i}$  en las variables  $t_1, \dots, t_n$ . Si  $I_L \neq \mathbf{I}(X_A)$ , entonces podemos elegir un  $f \in \mathbf{I}(X_A) \setminus I_L$  con término líder mínimo de monomios  $x^\alpha = \prod_{i=1, \dots, n} (x_i)^{a_i}$ . Reescalando si es necesario,  $x^\alpha$  se convierte en término líder de  $f$ .

Así  $f(t^{m_1}, \dots, t^{m_n})$  es idénticamente cero como un polinomio en  $t_1, \dots, t_n$ , ahí debe haber una cancelación involucrando los términos que vienen de  $x^\alpha$ . En otras palabras,  $f$  debe contener un monomio  $x^\beta = \prod_{i=1, \dots, n} (x_i)^{b_i} < x^\alpha$  tal que

$$\prod_{i=1, \dots, n} (t^{m_i})^{b_i} = \prod_{i=1, \dots, n} (t^{m_i})^{a_i}.$$

Esto implica que

$$\sum_{i=1, \dots, n} a_i m_i = \sum_{i=1, \dots, n} b_i m_i,$$

tal que  $\alpha - \beta = \sum_{i=1, \dots, n} (a_i - b_i) * e_i \in L$ . Entonces  $x^\alpha - x^\beta \in I_L$  por la segunda descripción de  $I_L$ . Se sigue que  $f - x^\alpha + x^\beta$  también esta en  $\mathbf{I}(X_A) \setminus I_L$  que tiene término líder estrictamente mas pequeño; esto nos lleva a una contradicción y con ello completamos la prueba, **q.e.d.**

**2.3.1. Definición.** Sea  $L \subseteq \mathbb{Z}^n$  una subretícula.

(a) El ideal  $I_L = \langle x^\alpha - x^\beta | \alpha, \beta \in \mathbb{Z}_+^n \text{ y } \alpha - \beta \in L \rangle$  es llamado un **ideal reticular**.

(b) Un ideal reticular primo es llamado un **ideal tórico**.

**Proposición 2.3.2.** Un ideal  $I \subseteq \mathbb{C}[x_1, \dots, x_n]$  es tórico sí y solo sí este es primo y generado por binomios.

**Demostración.** Una dirección es clara, de la definición anterior  $I$  es tórico entonces es ideal reticular primo; i.e, primo y generado por binomios, se sigue de inciso (a). Ahora supongamos que

$I$  es primo y generado por binomios  $x_i^\alpha - x_i^\beta$ . Entonces observamos que  $V(I) \cap (\mathbb{C}^*)^n$  es no vacío (este contiene a la coordenada  $(1, \dots, 1)$ ) y es un subgrupo de  $(\mathbb{C}^*)^n$  un elemento en la intersección se ve de la siguiente manera  $z_i = x_i^\alpha$  en la  $i$ -ésima coordenada, estos elementos forman un subgrupo bajo la multiplicación compleja coordenada a coordenada, con elementos identidad e inverso dados por  $z_i z_i^{-1} = 1$  en cada coordenada, y de la misma manera para  $z_i^{-1} = x_i^{-\alpha}$  con  $\alpha \in \mathbb{Z}^+$  en cada coordenada, también claramente esta operación es cerrada en  $V(I) \cap (\mathbb{C}^*)^n$ . Siendo que  $V(I) \subseteq \mathbb{C}^n$  es irreducible, se sigue que  $V(I) \cap (\mathbb{C}^*)^n$  es una subvariedad irreducible de  $(\mathbb{C}^*)^n$  y es también un subgrupo, vemos que  $T = V(I) \cap (\mathbb{C}^*)^n$  es un toro; mediante este toro construimos la siguiente retícula de caracteres asociada, como se muestra a continuación.

Proyectando en la  $i$ -ésima coordenada de  $(\mathbb{C}^*)^n$  nos da un carácter  $T \hookrightarrow (\mathbb{C}^*)^n \rightarrow \mathbb{C}^*$ , que por nuestra convención usual escribimos  $\chi^{m_i} : T \rightarrow \mathbb{C}^*$  para cada  $m_i \in A = \{m_1, \dots, m_n\} \subseteq M \simeq \mathbb{Z}^d$ . Se sigue fácilmente que  $V(I) = X_A$ ; lo cual se prueba como sigue, con cada carácter  $\chi^{m_i}(t)$  obtenemos el mapeo  $\Phi_A(t)$ , definido al principio de la sección, la imagen  $\Phi_A(T_N) = T$  es un toro que es cerrado en  $(\mathbb{C}^*)^n$  y de esto último se tiene que  $X_A \cap (\mathbb{C}^*)^n = T$ , y siendo  $I$  primo de la hipótesis, tenemos  $I = \mathbf{I}(X_A)$  por el teorema de los ceros de Hilbert. Entonces  $I$  es tórico por la proposición anterior, **q.e.d.**

Todas las variedades tóricas afines se construyen a partir de ideales tóricos como vemos en el siguiente ejemplo a continuación.

**Ejemplo (3).** Dado el cono  $\sigma$  generado por  $(3, 1), (1, 2)$ , i.e.,  $\sigma = \text{Con}((3, 1), (1, 2))$  en el plano  $\mathbb{Q}^2$ . Calculamos la bases de Hilbert asociadas al monoide  $S_\sigma = \sigma^\vee \cap \mathbb{Z}^2$ , siguiendo el algoritmo del apéndice A o algún programa de álgebra computacional como Nórmliz, conseguimos las bases

$$H_{\sigma^\vee} = \{(-1, 3), (0, 1), (1, 0), (2, -1)\}.$$

La variedad tórica normal  $X_{\sigma^\vee}$  es el espectro del álgebra tórica  $\mathbb{C}[S_\sigma] = \mathbb{C}[H] = \mathbb{C}[x]/I_H$ . Concluimos que  $X_{\sigma^\vee}$  esta encajada en un 4-espacio como es el conjunto de ceros del ideal tórico.

$$I_H = \langle x_2 x_4 - (x_3)^2, x_1 x_4 - (x_2)^2 x_3, x_1 x_3 - (x_2)^3 \rangle.$$

## 2.4. Morfismos tóricos y pegado de morfismos

**2.4.1. Definición.** Sea  $\Phi : \mathbb{C}^k \longrightarrow \mathbb{C}^m$  es un morfismo monomial es decir (toda componente no cero de  $\Phi$  es un monomio en la coordenadas de  $\mathbb{C}^k$ ), y sean  $X_\sigma \hookrightarrow \mathbb{C}^k$  y  $X_{\sigma'} \hookrightarrow \mathbb{C}^m$  inclusiones de variedades tóricas afines. Si  $\Phi(X_\sigma) \subset X_{\sigma'}$ , entonces  $\varphi := \Phi|_{X_\sigma}$  es un morfismo llamado **morfismo tórico afín** de  $X_\sigma$  a  $X_{\sigma'}$ . Si  $\varphi$  es biyectivo y el mapeo inverso  $\varphi^{-1} : X_{\sigma'} \longrightarrow X_\sigma$  es de nuevo un morfismo tórico, entonces, llamamos a  $\varphi$  un **isomorfismo tórico**, decimos que  $X_\sigma$  y  $X_{\sigma'}$  son isomorfos, y escribimos  $X_\sigma \simeq_{torico} X_{\sigma'}$  o brevemente,  $X_\sigma \simeq X_{\sigma'}$ .

**Proposición 2.4.1.** Todo morfismo tórico  $\varphi : X_\sigma \longrightarrow X_{\sigma'}$ , determina únicamente un homomorfismo monomial  $\varphi^* : R_{\sigma'} \longrightarrow R_\sigma$  y viceversa.

**Demostración.** Sea  $\varphi$  un morfismo tórico y sean  $a \in \mathbb{C}[x_1, \dots, x_k]$ ,  $a' \in \mathbb{C}[y_1, \dots, y_m]$  ideales primos. Entonces, por definición, tenemos el siguiente diagrama conmutativo:

$$\begin{array}{ccccc} X_\sigma & \simeq & V(a) & \hookrightarrow & \mathbb{C}^k \\ \downarrow \varphi & & \downarrow & & \downarrow \Phi \\ X_{\sigma'} & \simeq & V(a') & \hookrightarrow & \mathbb{C}^m \end{array}$$

donde  $\Phi$  es monomial, i.e., de la forma  $\Phi = (\lambda_1 x^{e_1}, \dots, \lambda_m x^{e_m})$ . El correspondiente homomorfismo de álgebras

$$\Phi^* : \mathbb{C}[y_1, \dots, y_m] \longrightarrow \mathbb{C}[x_1, \dots, x_k], \quad g \longrightarrow g \circ \Phi$$

que mapea  $y_j$  a  $\lambda_j x^{e_j}$  y así, es monomial. Se puede probar la siguiente afirmación

$$(*) \quad \Phi(V(a)) \subset V(a') \iff \Phi^*(a') \subset a.$$

**Prueba de la afirmación:**  $\Rightarrow$ ) Suponemos que  $\Phi(V(a)) \subset V(a')$  aplicamos el conjunto algebraico  $I$  en ambos lados, lo cual implica que,  $a' = \sqrt{a'} \subseteq I(V(a')) \subset I(\Phi(V(a)))$ . Ahora sea  $f' \in a'$  y a su

vez  $f' \in I(\Phi(V(a))) \Rightarrow f' \circ \Phi(x) = 0$  donde  $x \in V(a) \ni \Phi(x) = p \forall p \in \Phi(V(a))$ ; notamos también que,  $f' \circ \Phi(x) = (f' \circ \Phi)(x) = \Phi^*(x) = 0$  siendo que  $x \in V(a) \Rightarrow \Phi^* \in a$ , esto es cierto  $\forall f' \in a'$ , por tanto  $\Phi^*(a') \subset a$ .

De ahí,  $\Phi^*$  induce un homomorfismo monomial

$$\begin{array}{ccc} \varphi^* : R_{\sigma'} = & \mathbb{C}[y]/a' & \longrightarrow & \mathbb{C}[x]/a = & R_{\sigma} \\ & \parallel & & \parallel & \\ & \mathbb{C}[v] & \longrightarrow & \mathbb{C}[u] & \end{array}$$

Y viceversa, un homomorfismo monomial  $\vartheta : R'_{\sigma} \longrightarrow R_{\sigma}$  puede ser levantado a un homomorfismo monomial  $\Theta : \mathbb{C}[y] \longrightarrow \mathbb{C}[x]$  con  $\Theta(a') \subset a$ . Esto corresponde a un homomorfismo monomial

$$\Theta : \mathbb{C}^k \longrightarrow \mathbb{C}^m, x \longrightarrow (\Theta(y_1), \dots, \Theta(y_m))$$

con  $\Phi^* = \Theta$ . Tenemos  $\Theta(a') \subset a$ , así,  $\Phi(V(a)) \subset V(a')$  por (\*), de ese modo  $\Phi$  induce un morfismo tórico  $\varphi : X_{\sigma} \longrightarrow X'_{\sigma}$ , con esto se prueba la proposición **q.e.d.**

**Lema 2.4.1.** Si  $R_{\sigma} \simeq \mathbb{C}[u_1, \dots, u_k] = \mathbb{C}[\zeta]/\mathbf{a}$  y  $R_{\sigma'} \simeq \mathbb{C}[v_1, \dots, v_m] = \mathbb{C}[\eta]/\mathbf{a}'$  son dos representaciones del anillo coordenado  $R_{\sigma}$ , entonces, existe una transformación de coordenadas entre  $(u_1, \dots, u_k)$  y  $(v_1, \dots, v_m)$ .

**Demostración.** Sea  $A = (a_1, \dots, a_k)$  y  $B = (b_1, \dots, b_m)$  dos sistemas de generadores del monoide  $\sigma \cap \mathbb{Z}^n$ , y sea  $u_i := t^{a_i}$  y  $v_j := t^{b_j}$  y recordando que  $\sigma \cap \mathbb{Z}^n \subset \mathbb{R}^n$  entonces  $a_i, b_j \in \mathbb{R}^n$  siendo que es un espacio vectorial real; entonces existe una transformación lineal de coordenadas que denotamos por  $L$  donde  $L : \sigma \cap \mathbb{Z}^n \rightarrow \sigma \cap \mathbb{Z}^n$  tal que  $L(a_i) = b_j$  y viceversa mediante  $L^{-1}$ ; de esa manera existe una transformación de coordenadas inducida  $T : \mathbb{C}^k \rightarrow \mathbb{C}^m$ , que definimos como  $T(u_i) = t^{L(a_i)} = t^{b_j} = v_j$  y transformación inversa definida por  $T^{-1}(v_j) = t^{L^{-1}(b_j)} = t^{a_i} = u_i$  y claramente son transformaciones lineales. Esto prueba la proposición **q.e.d.**

**2.4.1. Definición.** Para dos conos reticulares, donde  $\sigma \subset \mathbb{R}^n = \text{lin}\sigma$  y  $\sigma' \subset \mathbb{R}^m = \text{lin}\sigma'$ , decimos que  $\sigma$  y  $\sigma'$  son isomorfos y escribimos esto como  $\sigma \simeq \sigma'$ , si  $m = n$  y existe una transformación

unimodular  $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$  con  $L(\sigma) = \sigma'$ . Entonces, los monoides  $\sigma \cap \mathbb{Z}^n$  y  $\sigma' \cap \mathbb{Z}^n$  son isomorfos también.

**2.4.2. Definición.** Decimos que dos álgebras  $R_\sigma$  y  $R_{\sigma'}$  son monomialmente isomorfas y denotamos como  $R_\sigma \simeq R_{\sigma'}$ , si existen homomorfismos monomiales mutuamente inversos  $R_\sigma \longleftrightarrow R_{\sigma'}$ .

**Teorema 2.4.1.** Para conos reticulares  $\sigma \subset \mathbb{R}^n = \text{lin}\sigma$  y  $\sigma' \subset \mathbb{R}^n = \text{lin}\sigma'$ , las siguientes condiciones son equivalentes:

$$(a) \sigma \simeq \sigma' \quad (b) R_\sigma \simeq R_{\sigma'} \quad (c) X_\sigma \simeq X_{\sigma'}$$

**Demostración.** Las implicaciones a)  $\Rightarrow$  b)  $\Rightarrow$  c) son probadas mediante el siguiente diagrama y probamos que es conmutativo.

$$\begin{array}{ccccc} \sigma & \longrightarrow & R_\sigma & \hookrightarrow & X_\sigma = \text{Spec}(R_\sigma) \\ \downarrow \uparrow L^{-1} & & \downarrow \uparrow \psi^{-1} & & \downarrow \uparrow \varphi^{-1} \\ \sigma' & \longrightarrow & R_{\sigma'} & \hookrightarrow & X_{\sigma'} = \text{Spec}(R_{\sigma'}) \end{array}$$

Donde definimos los homomorfismos monomiales  $h_\sigma : \sigma \rightarrow R_\sigma$ ,  $h_{\sigma'} : \sigma' \rightarrow R_{\sigma'}$ ,  $j_\sigma : R_\sigma \rightarrow X_\sigma$ ,  $j_{\sigma'} : R_{\sigma'} \rightarrow X_{\sigma'}$ ; también por hipótesis sabemos que  $\sigma \simeq \sigma'$  entonces existe una transformación unimodular (definición 2.4.1)  $L$  tal que  $L(\sigma) = \sigma'$  y que a su vez  $L^{-1}(\sigma') = \sigma$  esta bien definido; con este hecho construimos los homomorfismos monomiales siguientes:

$$\begin{aligned} h_\sigma(a) &= \sum \lambda_a t^a \in R_\sigma \text{ y } a \in \text{supp}(h_\sigma) \subset \sigma, \\ h_{\sigma'}(a') &= \sum \lambda_{a'} t^{a'} \in R_{\sigma'} \text{ y } a' \in \text{supp}(h_{\sigma'}) \subset \sigma', \\ \psi(h_\sigma(a)) &= \sum \lambda_a t^{L(a)} \in R_{\sigma'} \text{ y } L(a) = a' \in \text{supp}(\psi) \subset \sigma', \\ \psi^{-1}(h_{\sigma'}(a')) &= \sum \lambda_{a'} t^{L^{-1}(a')} \in R_\sigma \text{ y } L^{-1}(a') = a \in \text{supp}(\psi^{-1}) \subset \sigma. \end{aligned}$$

También elegimos generadores primos  $t^a \in R_\sigma$ ,  $a \in \sigma$  y  $t^{a'} \in R_{\sigma'}$ ,  $a' \in \sigma'$  y definimos:

$$\begin{aligned} j_\sigma(t^a) &= \langle t^a \rangle \in X_\sigma = \text{Spec}(R_\sigma), \\ j_{\sigma'}(t^{a'}) &= \langle t^{a'} \rangle \in X_{\sigma'} = \text{Spec}(R_{\sigma'}), \\ \varphi(\langle t^a \rangle) &= \langle t^{L(a)} \rangle \in X_{\sigma'} = \text{Spec}(R_{\sigma'}), \\ \varphi^{-1}(\langle t^{a'} \rangle) &= \langle t^{L^{-1}(a')} \rangle \in X_\sigma = \text{Spec}(R_\sigma). \end{aligned}$$

con  $\langle t^{L(a)} \rangle$  y  $\langle t^{L^{-1}(a')} \rangle$  ideales primos en los respectivos espectros de anillos coordenados  $R_\sigma$  y  $R_{\sigma'}$ . Vemos fácilmente que estos homomorfismos monomiales cumplen las siguientes identidades y sin pérdida de generalidad podemos tomar  $\lambda_a = \lambda_{a'} = 1$  así:  $L \circ L^{-1} = id_\sigma$ ,  $L^{-1} \circ L = id_{\sigma'}$ ,  $\psi \circ \psi^{-1} = id_{R_{\sigma'}}$ ,  $\psi^{-1} \circ \psi = id_{R_\sigma}$ ,  $\varphi^{-1} \circ \varphi = id_{X_\sigma}$ ,  $\varphi \circ \varphi^{-1} = id_{X_{\sigma'}}$ , aclaramos también que los isomorfismos  $L$ ,  $\psi$ ,  $\varphi$ , son isomorfismos de, conos, álgebras de anillos coordenados, e isomorfismos de variedades tóricas (morfismos tóricos) respectivamente, los dos primeros están bien definidos; uno es transformación unimodular y el otro homomorfismo de álgebras; solo resta ver que  $\varphi$  es morfismo tórico. Definimos el homomorfismo monomial  $\Phi : \mathbb{C}^n \rightarrow \mathbb{C}^n$  como  $\Phi(\langle t^a \rangle) = \langle t^{a'} \rangle \ni \Phi(X_\sigma) \subset X_{\sigma'}$ , este homomorfismo induce el morfismo  $\varphi$  el cual es biyectivo, ya que para el generador  $t^0 = 1_{X_\sigma}$  con vector reticular  $a = 0 \in \sigma$ ,  $\varphi(t^0) = t^{L(0)} = 1_{X_{\sigma'}}$ , así  $\varphi$  es inyectivo, ahora tomamos el generador  $t^{a'} \in X_{\sigma'}$ , siendo que  $L(a) = a' \Rightarrow \exists t^a \in X_\sigma \ni \varphi(t^a) = t^{L(a)} = t^{a'}$ , así  $\varphi$  es sobreyectivo; notamos que  $\varphi$  podemos escribirlo como  $\varphi = \Phi|_{X_\sigma}$ ; a su vez podemos ver también que  $\varphi^{-1}$  es de nuevo morfismo tórico, por tanto  $\varphi$  es un isomorfismo tórico. Por otro lado también,  $\psi(h_\sigma(a)) = h_{\sigma'}(L(a))$ ,  $\varphi(j_\sigma(t^a)) = j_{\sigma'}(t^{a'})$ ; y con ello probamos que el diagrama conmuta y conseguimos los isomorfismos deseados.

Ahora nos concentramos en mostrar c)  $\Rightarrow$  a). Ya hemos visto en la proposición 2.4.1 que todo morfismo tórico  $\varphi : X_\sigma \rightarrow X_{\sigma'}$  induce un homomorfismo monomial de  $\mathbb{C}$ -álgebras  $\varphi^* : R_\sigma \rightarrow R_{\sigma'}$ . Primero mostramos que  $\varphi^*$  se extiende a un único homomorfismo

$$\Theta : \mathbb{C}[\omega_1, (\omega_1)^{-1}, \dots, \omega_n, (\omega_n)^{-1}] \rightarrow \mathbb{C}[z_1, (z_1)^{-1}, \dots, z_n, (z_n)^{-1}],$$

de álgebras de monomios de Laurent con  $R_{\sigma'} \subset \mathbb{C}[\omega, \omega^{-1}]$  y  $R_\sigma \subset \mathbb{C}[z, z^{-1}]$ : Elegimos un vector reticular  $b \in \sigma'$  tal que todos los vectores reticulares  $b + e_j$  para  $j = 1, \dots, n$  también pertenece a  $\sigma'$  (esto es posible del hecho de que  $\sigma'$  es convexo). Con  $g_0 = \omega^b$  y  $g_j = \omega^{b+e_j} = \omega_j g_0$  colocamos  $\Theta(\omega_j) = \varphi^*(g_j)/\varphi^*(g_0)$ , el cual es un monomio de Laurent bien definido de  $\mathbb{C}[z, z^{-1}]$  así  $\varphi^*(g_0)$  y  $\varphi^*(g_j)$  son monomios de Laurent. De ahí, tenemos que  $\Theta(\omega_j) = \lambda_j z^{c_j}$  con  $\lambda_j \in \mathbb{C}^*$  y  $c_j = (c_{j1}, \dots, c_{jn}) \in \mathbb{Z}^n$  para  $j = 1, \dots, n$ .

Como  $\varphi : X_\sigma \rightarrow X_{\sigma'}$  es un isomorfismo tórico por hipótesis, inmediatamente encontramos que  $\Theta$  es un isomorfismo de  $\mathbb{C}$ -álgebras de monomios de Laurent, y que mapea inversamente  $z_i$  a  $\mu_j \omega^{d_i}$  con  $\mu_j \in \mathbb{C}^*$  y  $d_i = (d_{i1}, \dots, d_{im}) \in \mathbb{Z}^m$  para  $i = 1, \dots, n$ . Hacemos las siguientes observaciones de nuestra notación, tenemos que  $\Theta(\omega_j) = \lambda_j z^{c_j}$  donde  $\omega_j = \omega^{e_j} = \omega^{(0, \dots, 1, \dots, 0)} = \omega_0 * \dots * \omega^1 * \omega^0 \in \mathbb{C}[\omega_1, (\omega_1)^{-1}, \dots, \omega_m, (\omega_m)^{-1}]$  así para  $j = 1, \dots, n$  tenemos  $\omega^{e_1} * \dots * \omega^{e_n} = \omega^{(e_1, \dots, e_n)}$  y ordenando como columnas los vectores reticulares  $e_j^s \in \mathbb{R}^n$  vemos que  $\delta_{ik} = (e_1, \dots, e_n) \in M_{n \times n}(\mathbb{R})$  es la matriz identidad, escribimos  $\omega^{(e_1, \dots, e_n)} = \omega^{\delta_{ik}}$ . Y utilizando el hecho de que  $\Theta^{-1} \circ \Theta(\omega_j) = \omega_j \Rightarrow \Theta^{-1}(\lambda_j z^{c_j}) = \omega_j$ , siguiendo el cálculo de abajo,

$$\begin{aligned} \omega_j &= \Theta^{-1}(\lambda_j z^{c_j}) = \lambda_j * \prod_{i=1, \dots, n} \Theta^{-1}(z_i)^{c_{ji}} \\ &= \lambda_j * \prod_{i=1, \dots, n} (\mu_i)^{c_{ji}} \omega^{c_{ji} d_i} \text{ para } j = 1, \dots, n \end{aligned}$$

de donde conseguimos las siguientes relaciones, y corriendo sobre el índice  $j = 1, \dots, n$  se tiene  $\omega^{(e_1, \dots, e_n)} = \omega^{(c_{1i} d_i, \dots, c_{ni} d_i)}$  o de otra manera  $\omega^{\delta_{ik}} = \omega^{\sum c_{ji} d_{ik}}$  con  $\lambda_j * \prod_{i=1, \dots, n} (\mu_i)^{c_{ji}} = 1$  también obtenemos usando un cálculo análogo para  $z_i = \Theta(\mu_i \omega^{d_i})$  las ecuaciones siguientes:

$$\sum d_{ij} c_{jk} = \delta_{ik},$$

$$\mu_i \prod_{j=1, \dots, n} (\lambda_j)^{d_{ij}} = 1.$$

Consecuentemente las matrices  $c_{ij}$  y  $d_{ji}$  son inversas una de la otra.

Por construcción de  $\Theta$ ,  $\Theta(R_{\sigma'}) \subset R_\sigma$ . Sea  $A = (a_1, \dots, a_k)$  y  $B = (b_1, \dots, b_l)$  denotan sistemas de generadores de los monoides  $\sigma \cap \mathbb{Z}^n$  y  $\sigma' \cap \mathbb{Z}^n$ , respectivamente. Entonces,  $\Theta$  mapea cada generador  $v_s = \omega^{b_i}$  de  $R_{\sigma'}$  a un monomio  $\kappa_s z^{A p_s}$  en generadores  $u_i = z^{a_i}$  de  $R_\sigma$ , donde  $\kappa_s \in \mathbb{C}^*$  y  $p_s \in \mathbb{Z}_{>0}^k$  para  $s = 1, \dots, l$ . De otra manera,  $\Theta(\omega^{b_i}) = \prod_{j=1, \dots, n} (\lambda_j z^{c_j})^{b_{sj}}$ , y  $\sum_{j=1, \dots, n} c_j b_{sj} = \sum_{i=1, \dots, k} a_i p_{si}$  donde los  $c_j$  son filas de la matriz  $c_{ji}$  la cual ya se probó que es invertible. Junto con la afirmación análoga para  $\Theta^{-1}$ , esto implica que  $\sigma$  y  $\sigma'$  son conos reticulares isomorfos, **q.e.d.**

**Pegado de morfismos tóricos.** Para construir variedades tóricas en general las cuales se obtienen pegando ciertos conjuntos afines que veremos a continuación. La información del pegado

esta codificada en un abanico  $\Sigma$  en  $\mathbb{R}^n$ , pero no es el más simple modo. Para el propósito presente, regresamos al concepto esencial de conos duales de  $\sigma \in \Sigma$  y consideramos la variedades tóricas afines asociadas  $X_{\sigma^\vee}$ . La condición de que los conos de  $\Sigma$  tienen **0 como ápice garantiza que sus conos duales son todos de la misma dimensión n**. Si  $\tau$  es una cara de un cono  $\sigma$ , entonces, tenemos una inclusión  $\sigma^\vee \subset \tau^\vee$ . Dado esto podemos ver que el morfismo inducido tórico de esta inclusión  $\psi_{\tau,\sigma} : X_{\tau^\vee} \rightarrow X_{\sigma^\vee}$  es un isomorfismo dado por una transformación de coordenadas sobre las copias del n-toro algebraico  $T$  incluido en ambas variedades como subconjuntos abiertos. Queremos mostrar que  $\psi_{\tau,\sigma}$  es una inclusión abierta. Aquello permitirá pegar cualquiera dos variedades tóricas  $X_\sigma, X_{\sigma'}$  a lo largo de subconjuntos abiertos correspondientes a la cara común  $\tau = \sigma \cap \sigma'$ , ver fig. 2.1.

**Lema 2.4.2.** Con lo dicho anteriormente arriba. Sea  $\tau$  una cara de  $\sigma$  entonces tenemos la identificación natural.

$$X_{\tau^\vee} \simeq X_{\sigma^\vee} \setminus \{u_k = 0\}.$$

donde  $u_k$  es el último generador de la representación en el anillo coordinado asociado a  $X_{\sigma^\vee}$ .

**Demostración.** Sea  $\tau^\vee = \sigma^\vee + \mathbb{R}_{\geq 0}(-m)$  la expresión entre generadores que relaciona a  $\sigma^\vee$  y  $\tau^\vee$  esta relación fue probada anteriormente en el lema 1.1.4 del primer capítulo.

De ahí conseguimos la siguiente expresión,

$$\tau^\vee \cap \mathbb{Z}^n = (\sigma^\vee \cap \mathbb{Z}^n) + \mathbb{R}_{\geq 0}(-m).$$

El monoide  $\tau^\vee \cap \mathbb{Z}^n$ , es de esta manera obtenido de  $\sigma^\vee \cap \mathbb{Z}^n$  por la introducción del generador adicional  $-m$  que pertenece al espacio lineal de  $\tau^\vee$ . Podemos asumir que  $a_k := m$  ocurre en el último miembro de los generadores  $a_1, \dots, a_k$  de  $\sigma^\vee \cap \mathbb{Z}^n$ . El monoide  $\tau^\vee \cap \mathbb{Z}^n$ , es así generado por  $a_1, \dots, a_k, a_{k+1} := -m$ . las relaciones entre estos generadores son  $\sum_{i=1}^{k+1} v_i a_i = \sum_{j=1}^{k+1} \mu_j a_j$  con  $v_i, \mu_j \in \mathbb{Z}_{\geq 0}, v_{k+1} \geq \mu_{k+1}$ , y provienen del núcleo del morfismo de monomios que sirvió para la construcción de ideales tóricos  $\hat{\Phi}(\sum_{i,j=1}^{k+1} v_i a_i - \mu_j a_j) = 0$ , también se tiene la siguiente relación al

haber incluido el generador extra  $a_k + a_{k+1} = 0$ , ahora restando  $\mu_{k+1}a_{k+1}$  de ambos lados se reduce  $\mu_{k+1} = 0$ . Y sumando  $v_{k+1}a_k$  en ambos lados y usando el hecho de que  $a_k + a_{k+1} = 0$ , se sigue que,

$$\begin{aligned} \sum_{i=1}^k v_i a_i - \sum_{j=1}^k \mu_j a_j + v_{k+1} a_{k+1} + v_{k+1} a_k &= v_{k+1} a_k, \\ \sum_{i=1}^k v_i a_i - \sum_{j=1}^k \mu_j a_j + v_{k+1} (a_{k+1} + a_k) &= v_{k+1} a_k. \end{aligned}$$

Entonces nos queda una relación que únicamente involucra  $a_1, \dots, a_k$ .

Del lado de las álgebras monomiales asociadas  $R_\sigma^\vee$  y  $R_\tau^\vee$  con generadores correspondientes  $u_i := t^{a_i}$ , la relación aditiva entre  $a_k + a_{k+1} = 0$  corresponde a la versión multiplicativa de los monomios coordenados a  $u_k * u_{k+1} = 1$  en  $R_{\tau^\vee}$  específicamente. Como aquello es lo único nuevo que se definió en las relaciones al pasar de  $R_{\sigma^\vee}$  a  $R_{\tau^\vee}$  y como los generadores  $u_i$  son solo las funciones coordenadas en las variedades tóricas  $X_{\sigma^\vee}$  y  $X_{\tau^\vee}$ , así vemos que la proyección

$$(x_1, \dots, x_k, x_{k+1}) \longmapsto (x_1, \dots, x_k)$$

identifica  $X_{\tau^\vee}$  con subconjuntos abiertos de  $X_{\sigma^\vee}$  dado por  $(x_k \neq 0)$  (llamado **subconjunto principal abierto de Zariski**), y con la inclusión como mapeo inverso se ha probado el resultado, **q.e.d.**

Para dos conos  $\sigma$  y  $\sigma' \in \Sigma$ , sea  $\tau := \sigma \cap \sigma'$  una cara común entre ellos. Elegimos un sistema de coordenadas apropiado  $v_1, \dots, v_l$  para  $X_{\sigma'^\vee}$ , de acuerdo al lema anterior, tenemos isomorfismos

$$X_{\sigma^\vee} \setminus \{u_k = 0\} \simeq X_{\tau^\vee} \simeq X_{\sigma'^\vee} \setminus \{v_l = 0\}$$

su composición produce un isomorfismo tórico

$$\psi_{\sigma, \sigma'} : X_{\sigma^\vee} \setminus \{u_k = 0\} \longrightarrow X_{\sigma'^\vee} \setminus \{v_l = 0\}$$

correspondiente a la transformación de coordenadas

$$(u_1, \dots, u_k, u_{k+1}) \longmapsto (v_1, \dots, v_l, v_{l+1})$$

de  $X_{\tau^\vee}$ .

**2.4.3. Definición.** El isomorfismo de arriba  $\psi_{\sigma, \sigma'}$  es llamado el **morfismo de pegado** que pega juntos las variedades  $X_{\sigma^\vee}$  y  $X_{\sigma'^\vee}$  a lo largo de  $X_{\tau^\vee}$ .

### 2.4.1. Variedades tóricas y abanicos

**2.4.1.1 Definición.** Sea  $\Sigma$  un abanico en  $\mathbb{R}^n$ . En la unión disjunta  $\bigcup_{\sigma \in \Sigma} X_{\sigma^\vee}$ , identificamos dos puntos  $x \in X_{\sigma^\vee}$  y  $x' \in X_{\sigma'^\vee}$  que son mapeados uno con el otro por el morfismo de pegado  $\psi_{\sigma, \sigma'}$ . El conjunto de puntos, así obtenido es llamado **variedad tórica** o encajamiento del toro  $X_\Sigma$  determinado por el abanico  $\Sigma$ .

$X_\Sigma$  es un espacio topológico cubierto por cubrientes abiertos de variedades complejas que intersectan subvariedades abiertas de Zariski. Tal cubriente esta dado por copias de variedades tóricas afines  $X_{\sigma^\vee}$ , para  $\sigma \in \Sigma$ , las cuales son naturalmente incluidas en  $X_\Sigma$ , así queda definida la topología. Y estos datos son provistos por construcción.

## 2.5. Normalización de variedades tóricas afines

En esta sección caracterizaremos cuando una variedad tórica es normal su relación con conos poliédricos y bases de Hilbert asociadas.

**Teorema 2.5.1.** Sea  $V$  una variedad tórica afín con toro  $T_N$ . Entonces los siguientes enunciados son equivalentes:

- (a)  $V$  es normal.
- (b)  $V = \mathbf{Spec}(\mathbb{C}[S])$ , donde  $S \subseteq M$  es un monoide saturado (definición 1.2.6).
- (c)  $V = \mathbf{Spec}(\mathbb{C}[S_\sigma])$ , donde  $S_\sigma = \sigma^\vee \cap M$  y  $\sigma \subseteq N$  es un cono reticular.

**Demostración.** El conjunto  $V = \mathbf{Spec}(\mathbb{C}[S_\sigma])$ , para un monoide  $S$  contenido en una retícula, el toro de  $V$  tiene caracter reticular  $M = \mathbb{Z}S$ . También  $n = \dim V$ , tal que  $M \simeq \mathbb{Z}^n$ .

- (a)  $\implies$  (b): Si  $V$  es normal, entonces  $\mathbb{C}[S] = \mathbb{C}[V]$  es de cerradura entera en su campo de

fracciones  $\mathbb{C}(V)$ . Supongase que  $km \in S$  para algún  $k \neq 0 \in \mathbb{N}$  y  $m \in M$ . Entonces  $\chi^m$  es una función polinomial en  $T_N$  y así una función racional en  $V$  y  $T_N \subseteq V$  es un abierto de Zariski. Tenemos  $\chi^{km} \in \mathbb{C}[S]$  siendo que  $km \in S$ . Se sigue que  $\chi^m$  es una raíz del polinomio mónico  $Y^k - \chi^{km}$  con coeficientes en  $\mathbb{C}[S]$ . Por definición de normal, obtenemos  $\chi^m \in \mathbb{C}[S]$  i.e,  $m \in S$ . Así  $S$  es saturado.

(b)  $\implies$  (c). Sea  $A \subseteq S$  un conjunto finitamente generado de  $S$ . Entonces  $S$  está contenido en el cono poliédrico racional  $Con(A) \subseteq M_{\mathbb{R}}$  y  $\text{rango} \mathbb{Z}A = n$ , implica que  $\dim Con(A) = n$ . Se sigue que  $\sigma = Con(A)^\vee \subseteq N_{\mathbb{R}}$  es un cono poliédrico racional tal que  $S \subseteq \sigma^\vee \cap M$ . Y la igualdad se sostiene cuando  $S$  es saturado, veamos a continuación  $\forall m \in \sigma^\vee \cap M \Rightarrow m \in M$  y  $\forall k \neq 0 \in \mathbb{N} \Rightarrow km \in S \Rightarrow m \in S$  siendo que  $S$  es saturado. Entonces  $S = S_\sigma$ .

(c)  $\implies$  (a): Necesitamos mostrar que  $\mathbb{C}[S_\sigma] = \mathbb{C}[\sigma^\vee \cap M]$  es normal cuando  $\sigma \subseteq N_{\mathbb{R}}$  es un cono poliédrico racional. Sean  $\rho_1, \dots, \rho_r$  rayos de  $\sigma$ . Y puesto que  $\sigma$  es generado por sus rayos teorema 1.1.1, tenemos

$$\sigma^\vee = \bigcap_{i=1, \dots, r} \rho_i^\vee.$$

Intersectando con  $M$  conseguimos  $S_\sigma = \bigcap_{i=1, \dots, r} S_{\rho_i}$ , que implica claramente

$$\mathbb{C}[S_\sigma] = \bigcap_{i=1, \dots, r} \mathbb{C}[S_{\rho_i}].$$

Si  $\mathbb{C}[S_\sigma]$  es normal si cada  $\mathbb{C}[S_{\rho_i}]$  es normal, así es suficiente probar que  $\mathbb{C}[S_\rho]$  es normal cuando  $\rho$  es un rayo racional en  $N_{\mathbb{R}}$ . Sea  $u_\rho \in \rho \cap N$  el rayo generador de  $\rho$ . Así  $u$  es **primitivo** i.e,  $\frac{1}{k}u_\rho$  no pertenece a  $N \forall k > 1$ , podemos encontrar bases  $e_1, \dots, e_n$  de  $N$  con  $u_\rho = e_1$ . Esto nos permite asumir que  $\rho = Con(e_1)$ , tal que

$$\mathbb{C}[S_\rho] = \mathbb{C}[x_1, x_2^{\pm 1}, \dots, x_n^{\pm 1}].$$

Pero  $\mathbb{C}[x_1, \dots, x_n]$  es normal (esto es dominio de factorización única), y afirmamos que su localización,

$$\mathbb{C}[x_1, \dots, x_n]_{(x_2 \dots x_n)} = \mathbb{C}[x_1, x_2^{\pm 1}, \dots, x_n^{\pm 1}]$$

es también normal.

**Prueba de la afirmación.** Sea  $R$  un dominio normal con campo de fracciones  $K$  y  $S \subseteq R$  un subconjunto multiplicativo, entonces la localización  $R_S$  es normal. Sea  $q = \frac{t}{s} \in K$ ,  $t \in R$  y en particular tomemos  $s \in S$  y siendo que  $R$  es normal, existe un polinomio mónico  $p(x) \in R[x] \ni p(q) = 0$  y sea  $p(x) = (x - q)$  dicho polinomio, notemos que  $\frac{p(x)}{1_S} \in R_S[x]$  con  $1_S$  la identidad de  $S$ , ahora escribiendo de otra manera  $\frac{p(x)}{1_S} = \left(\frac{x}{1_S} - \frac{q}{1_S}\right) = \left(\frac{x}{1_S} - \frac{t}{s}\right) = 0$ ; por tanto existe un polinomio mónico  $\frac{p(x)}{1_S} \in R_S[x] \ni \frac{p(\frac{t}{s})}{1_S} = 0$  con  $\frac{t/s}{1_S} \in C(R_S)$  donde  $C(R_S)$  denota el campo de fracciones de  $R_S$ . Con esto se completa la prueba **q.e.d.**

**Ejemplo (a).** El conjunto  $V = \mathbf{V}(xy - zw)$  es la variedad tórica afín  $X_\sigma$  del cono poliédrico  $\sigma = \text{Con}(e_1, e_2, e_1 + e_3, e_2 + e_3)$ . Entonces por el teorema anterior implica que la variedad  $V$  es normal.

La normalización de una variedad es fácil de describir. Sea  $V = \mathbf{Spec}(\mathbb{C}[S])$  para un monoide  $S$ , tal que el toro de  $V$  tiene carácter reticular  $M = \mathbb{Z}S$ . Sea  $\text{Con}(S)$  denote el cono de cualquier conjunto finitamente generado de  $S$  y sea  $\sigma = \text{Con}(S)^\vee \subseteq N_{\mathbb{R}}$ . Entonces se tiene el siguiente resultado.

**Proposición 2.5.1.** El cono de arriba es reticular en  $N_{\mathbb{R}}$  y la inclusión  $\mathbb{C}[S] \subseteq \mathbb{C}[S^\vee \cap M]$  induce un morfismo  $X_\sigma \rightarrow V$  el cual es **el morfismo de normalización de  $V$** .

**Demostración.** De manera general;  $V' = \text{Spec}(\mathbb{C}[V]')$  y llamamos a  $V'$  la **normalización** de  $V$ , la inclusión natural  $\mathbb{C}[V] \subseteq \mathbb{C}[V]' = \mathbb{C}[V']$  induce el morfismo  $V' \rightarrow V$ , y el cual este es el morfismo de normalización que queremos, aplicado de manera particular para nuestro caso llegamos al resultado, **q.e.d.**

**Variedades tóricas afines no singulares.** Nuestro objetivo ahora es caracterizar cuando una variedad es no singular, es importante notar que variedades no singulares implica que son normales. Consideraremos únicamente variedades  $X_\sigma$  que se construyen a partir de conos poliédricos racionales  $\sigma \subseteq N_{\mathbb{R}}$ .

Estudiamos  $X_\sigma$  donde  $\sigma$  tiene dimensión maximal. Entonces  $\sigma^\vee$  es fuertemente convexo tal que

$S_\sigma = \sigma^\vee \cap M$ . Es cierto también que la acción del toro  $T_N$  sobre  $X_\sigma$  tiene un único **punto fijo** denotado por  $p_\sigma \in X_\sigma$  este hecho junto con las bases de Hilbert se relacionan en el siguiente resultado importante para el estudio de variedades no singulares.

**Lema 2.5.1.** Sea  $\sigma \subseteq N_{\mathbb{R}}$  es un cono poliédrico racional de dimensión maximal y sea  $T_{p_\sigma}(X_\sigma)$  el espacio tangente de Zariski a la variedad tórica afín  $X_\sigma$  en el punto  $p_\sigma$ . Entonces  $\dim T_{p_\sigma}(X_\sigma) = |H|$  donde  $H$  son las bases de Hilbert asociadas al cono  $\sigma$ .

**Demostración.** El ideal maximal de  $\mathbb{C}[S_\sigma]$  correspondiente a  $p_\sigma$  es  $\mathfrak{m} = \langle \chi^m : m \in S_\sigma \setminus \{0\} \rangle$ . Y  $\{\chi^m\}_{m \in S_\sigma}$  es una base de  $\mathbb{C}[S_\sigma]$ , obtenemos

$$\mathfrak{m} = \bigoplus_{m \neq 0} \mathbb{C}\chi^m = \bigoplus_{\text{irreducible}} \mathbb{C}\chi^m \oplus \bigoplus_{\text{reducible}} \mathbb{C}\chi^m = (\bigoplus_{m \in H} \mathbb{C}\chi^m) \oplus \mathfrak{m}^2,$$

la suma de  $\mathbb{C}$ -álgebras anterior implica el isomorfismo  $\mathfrak{m}/\mathfrak{m}^2 \simeq \bigoplus_{m \in H} \mathbb{C}\chi^m$  de donde se sigue que,  $\dim \mathfrak{m}/\mathfrak{m}^2 = |H|$ . Para relacionar este hecho con el ideal maximal  $\mathfrak{m}_{X_\sigma, p_\sigma}$  en el anillo local  $\mathbf{O}_{X_\sigma, p_\sigma}$ , usamos el morfismo natural

$$\mathfrak{m}/\mathfrak{m}^2 \longrightarrow \mathfrak{m}_{X_\sigma, p_\sigma} / \mathfrak{m}_{X_\sigma, p_\sigma}^2$$

y como otro resultado estándar de álgebra conmutativa se puede probar que siempre es un isomorfismo. Así  $T_{p_\sigma}(X_\sigma)$  es el espacio dual de  $\mathfrak{m}_{X_\sigma, p_\sigma} / \mathfrak{m}_{X_\sigma, p_\sigma}^2$ , y vemos que  $\dim T_{p_\sigma}(X_\sigma) = |H|$ . Consiguiendo lo que se quería probar **q.e.d.**

Las bases de Hilbert  $H$  de  $S_\sigma$  producen  $X_\sigma = Y_H \subseteq \mathbb{C}^n$  donde  $n = |H|$ . Este encajamiento afín es especialmente agradable. Dado cualquier encajamiento afín  $X_\sigma \rightarrow \mathbb{C}^l$ , tenemos  $\dim T_{p_\sigma}(X_\sigma) \leq l$ . En otras palabras,  $\dim T_{p_\sigma}(X_\sigma)$  es la cota inferior más pequeña en la dimensión de un encajamiento afín. El lema anterior muestra que cuando  $\sigma$  tiene dimensión maximal, las bases de Hilbert de  $S_\sigma$  nos dan el más eficiente encajamiento afín de  $X_\sigma$ . A continuación el resultado principal de variedades no singulares, recordemos que un cono es no singular o regular si es generado por un subconjunto de unas bases de la retícula.

**Teorema 2.5.2.** Sea  $\sigma \subseteq N_{\mathbb{R}}$  cono reticular. Entonces  $X_\sigma$  es **no singular** sí y solo sí  $\sigma$  es regular

o no singular, además, todas las variedades tóricas afines son de esta forma.

**Demostración.** Si una variedad tórica afín es no singular, entonces es normal, y por tanto de la forma  $X_\sigma$ . También, si  $\sigma$  es regular como un cono, entonces  $X_\sigma$  es no singular como variedad. Solo resta probar la necesidad. Fijamos  $\sigma \subseteq N_{\mathbb{R}}$  tal que  $X_\sigma$  es no singular. Y sea  $n = \dim X_\sigma = \dim N_{\mathbb{R}}$ . Primero supóngase que  $\sigma$  tiene dimensión  $n$  y sea  $p_\sigma \in X_\sigma$  el mismo punto del resultado anterior, así  $p_\sigma$  es no singular en  $X_\sigma$ , el espacio tangente de Zariski  $T_{p_\sigma}(X_\sigma)$  tiene dimensión  $n$  por definición. De otra manera por el lema anterior implica que  $\dim T_{p_\sigma}(X_\sigma)$  es la cardinalidad de las bases de Hilbert  $H$  de  $S_\sigma = \sigma \vee \cap M$ . De ese modo

$$n = |H| \geq |\{\text{aristas } \rho \subseteq \sigma^\vee\}| \geq n,$$

donde la primera desigualdad se sostiene (cada arista  $\rho \subseteq \sigma^\vee$  contribuye a un elemento de  $H$ ) y la segunda se sostiene del hecho de que  $\dim \sigma^\vee = n$ . Se sigue que  $\sigma$  tiene  $n$  aristas y  $H$  consiste de rayos generadores de estas aristas. Y siendo que  $M = \mathbb{Z}S_\sigma$ , las  $n$  aristas generadoras de  $\sigma^\vee$  generan la retícula  $M = \mathbb{Z}^n$  y por tanto son bases de  $M$ . De esa manera  $\sigma^\vee$  es no singular, y entonces  $\sigma = (\sigma^\vee)^\vee$  es regular siendo que la dualidad preserva regularidad.

Ahora supóngase que  $\dim \sigma = r < n$ . Regresamos al caso anterior como sigue. Sea  $N_1 \subseteq N$  la más pequeña subretícula saturada que contiene los generadores de  $\sigma$ . Entonces  $N/N_1$  es libre de torsión, y se puede probar que esto implica la existencia de un subretícula  $N_2 \subseteq N$  con  $N = N_1 \oplus N_2$ . Note que  $\text{rango}(N_1) = r$  y el  $\text{rango}(N_2) = n - r$ .

El cono  $\sigma$  pertenece a ambos  $(N_1)_{\mathbb{R}}$  y  $N_{\mathbb{R}}$ . Esto nos produce variedades tóricas afines  $X_{\sigma, N_1}$  y  $X_{\sigma, N}$  de dimensión  $r$  y  $n$  respectivamente. Además,  $N = N_1 \oplus N_2$  induce  $M = M_1 \oplus M_2$ , tal que  $\sigma \subseteq (N_1)_{\mathbb{R}}$  y  $\sigma \subseteq N_{\mathbb{R}}$  ahora intersectando  $\sigma^\vee$  en ambas inclusiones se inducen monoides  $S_{\sigma, N_1} \subseteq M_1$  y  $S_{\sigma, N} \subseteq M$ , respectivamente. Ahora utilizando las equivalencias mostradas en el teorema 2.4.1, y escribiendo,

$$S_{\sigma, N} = S_{\sigma, N_1} \oplus M_2,$$

y en términos de álgebras de semigrupos puede ser escrito

$$\mathbb{C}[S_{\sigma,N}] \simeq \mathbb{C}[S_{\sigma,N_1}] \otimes_{\mathbb{C}} \mathbb{C}[M_2].$$

El lado derecho es el anillo coordenado de  $X_{\sigma,N_1} \times T_{N_2}$  y usando la prueba del ejemplo (1) sección 2.2 y el teorema 2.4.1 nuevamente. Tenemos,

$$X_{\sigma,N} \simeq X_{\sigma,N_1} \times T_{N_2},$$

lo cual implica que

$$X_{\sigma,N} \simeq X_{\sigma,N_1} \times (\mathbb{C}^*)^{n-r} \subseteq X_{\sigma,N_1} \times \mathbb{C}^{n-r}.$$

puesto que asumimos que  $X_{\sigma,N}$  es no singular, se sigue que  $X_{\sigma,N_1} \times \mathbb{C}^{n-r}$  es no singular en cualquier punto  $(p, q) \in X_{\sigma,N_1} \times \mathbb{C}^{n-r}$ . Y afirmamos que

$$X_{\sigma,N_1} \times \mathbb{C}^{n-r} \text{ es no singular en } (p, q) \implies X_{\sigma,N_1} \text{ es no singular en } p.$$

Para probar esta afirmación solo hacemos uso de la definición de singularidad dada en la definición 3.1.2. Y ahora sea  $p = p_{\sigma} \in X_{\sigma,N_1}$ , de el anterior caso  $\sigma$  es no singular en  $N_1$  y  $\dim \sigma = \dim(N_1)_{\mathbb{R}}$ . De ahí vemos que  $\sigma$  es claramente no singular en  $N = N_1 \oplus N_2$  y con esto probamos el resultado **q.e.d.**

## 2.6. Conjuntos proyectivos y conos normales.

**2.6.1 Definición.** Un **n-espacio proyectivo complejo**  $\mathbb{C}P^n$  es el espacio de clases de equivalencia de pares de puntos tales que consisten de líneas en  $\mathbb{C}^{n+1}$ . La relación entre los puntos es de la siguiente manera dado cualquier vector  $v := (\eta_0, \dots, \eta_n)$  define una línea  $\mathbb{C} * v$  y dos de tales vectores  $v, v' \in \mathbb{C}^{n+1} \setminus \{0\}$  define la misma línea sí y solo sí es un múltiplo escalar de la otra.

Podemos así asociar con cualquier elemento  $\mathbb{C} * v$  de  $\mathbb{C}P^n$  sus **coordenadas homogéneas**  $[\eta_0, \dots, \eta_n]$

donde al menos una componente  $\eta_j$  es no cero, y todas las componentes son determinadas salvo un factor escalar común (no cero). El subconjunto  $U_j := \{[\eta_0, \dots, \eta_n] \in \mathbb{C}P^n : \eta_j \neq 0\}$  puede ser identificado con el  $n$ -espacio afín  $\mathbb{C}^n$  por el significado de los siguientes morfismos biyectivos

$$U_j \longrightarrow \mathbb{C}^n, \\ [\eta_0, \dots, \eta_n] \longrightarrow (\eta_0/\eta_j, \dots, \eta_{j-1}/\eta_j, \dots, \eta_n/\eta_j)$$

esto define un  $j$ -ésimo sistema de coordenadas inhomogéneas o coordenadas afines  $(\zeta_{j,i})_{i=1, \dots, n}$  en  $\mathbb{C}P^n$ . El  $n$ -espacio proyectivo es, así, cubierto por un cubriente de  $n+1$  copias del  $n$ -espacio afín. Para  $0 \leq j < k \leq n$  las transiciones de los sistemas coordenados  $(\zeta_{j,i})$  a  $(\zeta_{k,l})$  proporcionan el pegado de las cartas afines  $U_j$  y  $U_k$  que ya hemos visto y esta dado por una transformación de monomios. Para ver que  $\mathbb{C}P^n$  es una **variedad tórica**, notamos que la intersección  $\bigcap_{i=1, \dots, n} U_i$  es inmediatamente identificada con  $(\mathbb{C}^*)^n$ , esto es, el toro  $T$  encajado en  $\mathbb{C}P^n$ . Sin embargo, la acción natural del toro en  $\mathbb{C}^n = U_0$  por multiplicación compleja coordenada a coordenada se extiende de la manera siguiente,

$$((t_1, \dots, t_n), [\eta_0, \dots, \eta_n]) \longrightarrow [\eta_0, t_1 * \eta_1, \dots, t_n * \eta_n].$$

**Ejemplo (a).** Sea el plano proyectivo  $\mathbb{C}P^2 = \{[\eta_0, \eta_1, \eta_2] : \text{no todos } \eta_i = 0\}$ , las coordenadas homogéneas  $\eta_0, \eta_1, \eta_2$  son determinados salvo múltiplo común.

Este plano tiene un cubriente definido por tres cartas afines a saber,

$$A_0 = \{(1, \eta_1(\eta_0)^{-1}, \eta_2(\eta_0)^{-1}) : \eta_0 \neq 0\}, \quad A_1 = \{(\eta_0(\eta_1)^{-1}, 1, \eta_2(\eta_1)^{-1}) : \eta_1 \neq 0\} \text{ y } A_2 = \\ \{(\eta_0(\eta_2)^{-1}, \eta_1(\eta_2)^{-1}, 1) : \eta_2 \neq 0\}.$$

y colocamos  $z_1 = \eta_1(\eta_0)^{-1}$ ,  $z_2 = \eta_2(\eta_0)^{-1}$ , obtenemos

$$A_0 = \{(z_1, z_2)\}, \quad A_1 = \{((z_1)^{-1}, z_2(z_1)^{-1})\} \text{ y } A_2 = \{((z_2)^{-1}, z_1(z_2)^{-1})\}.$$

construimos anillos coordenados isomorfos a cada una de estas cartas afines, de la siguiente manera. A los monomios de Laurent que definen las coordenadas de estos afines asociamos sus politopos de Newton, la unión de ellos define el abanico regular como se ve abajo, también ver fig. 2.2.

Los politopos de Newton asociados son,  $\sigma_0 = \text{Con}(e_1, e_2)$ ,  $\sigma_1 = \text{Con}(e_2, -e_1 - e_2)$ ,  $\sigma_2 = \text{Con}(e_1, -e_1 - e_2)$  de otra manera,

$$\begin{aligned}\sigma_0 &= \mathbb{R}_{\geq 0}e_1 + \mathbb{R}_{\geq 0}(e_2), \\ \sigma_1 &= \mathbb{R}_{\geq 0}e_2 + \mathbb{R}_{\geq 0}(-e_1 - e_2), \\ \sigma_2 &= \mathbb{R}_{\geq 0}e_1 + \mathbb{R}_{\geq 0}(-e_1 - e_2),\end{aligned}$$

mediante el algoritmo de bases de Hilbert calculamos los conos duales de cada cono poliédrico de donde:

$$\begin{aligned}\sigma_0^\vee &= \text{Con}(e_1, e_2), \\ \sigma_1^\vee &= \text{Con}(-e_1, -e_1 + e_2), \\ \sigma_2^\vee &= \text{Con}(e_1 - e_2, -e_2),\end{aligned}$$

mediante el abanico  $\Sigma = \{\sigma_0^\vee, \sigma_1^\vee, \sigma_2^\vee\}$  obtenemos

$$\begin{aligned}R_{\sigma_0^\vee} &= \mathbb{C}[z_1, z_2] \text{ asociado a la variedad } X_{\sigma_0^\vee} = A_0, \\ R_{\sigma_1^\vee} &= \mathbb{C}[z_1^{-1}, z_1^{-1} * z_2] \text{ asociado a la variedad } X_{\sigma_1^\vee} = A_1 \\ R_{\sigma_2^\vee} &= \mathbb{C}[z_1 * z_2^{-1}, z_2^{-1}] \text{ asociado a la variedad } X_{\sigma_2^\vee} = A_2.\end{aligned}$$

**Ejemplo (b).** La **superficie de Hirzebruch**  $H_k$ . Consideramos un a hipersuperficie en  $\mathbb{C}P^1 \times \mathbb{C}P^2 = \{([\eta_0, \eta_1], [\zeta_0, \zeta_1, \zeta_2]) : (\eta_0, \eta_1) \neq (0, 0), (\zeta_0, \zeta_1, \zeta_2) \neq (0, 0, 0)\}$  determinada por la ecuación

$$\eta_0^k \zeta_0 = \eta_1^k \zeta_1, \quad k \in \mathbb{Z}.$$

Similarmente que el ejemplo anterior al determinar los politopos de Newton asociados así como sus conos duales, se encuentran cuatro planos como cartas afines y sus pegados dependen de  $k$  se sigue entonces:

$$\begin{aligned}
R_{\sigma_0^\vee} &= \mathbb{C}[z^{e_1}, z^{e_2}] = \mathbb{C}[z_1, z_2], \\
R_{\sigma_1^\vee} &= \mathbb{C}[z^{-e_1}, z^{e_1+ke_2}] = \mathbb{C}[z_2^{-1}, z_1 z_2^k], \\
R_{\sigma_2^\vee} &= \mathbb{C}[z^{-e_1}, z^{e_2}] = \mathbb{C}[z_1^{-1}, z_2], \\
R_{\sigma_3^\vee} &= \mathbb{C}[z^{-e_1-ke_2}, z^{-e_2}] = \mathbb{C}[z_1^{-1} z_2^{-k}, z_2^{-1}],
\end{aligned}$$

de donde conseguimos las siguientes variedades tóricas asociadas,

$$\begin{aligned}
X_{\sigma_0^\vee} &= \mathbf{Spec}(\mathbb{C}[z_1, z_2]); \\
X_{\sigma_1^\vee} &= \mathbf{Spec}(\mathbb{C}[z_2^{-1}, z_1 z_2^k]); \\
X_{\sigma_2^\vee} &= \mathbf{Spec}(\mathbb{C}[z_1^{-1}, z_2]); \\
X_{\sigma_3^\vee} &= \mathbf{Spec}(\mathbb{C}[z_1^{-1} z_2^{-k}, z_2^{-1}]).
\end{aligned}$$

Este enfoque es útil para la resolución de polinomios con este tipo de parametrización y usaremos este enfoque ampliamente a través de los desarrollos del apéndice B.

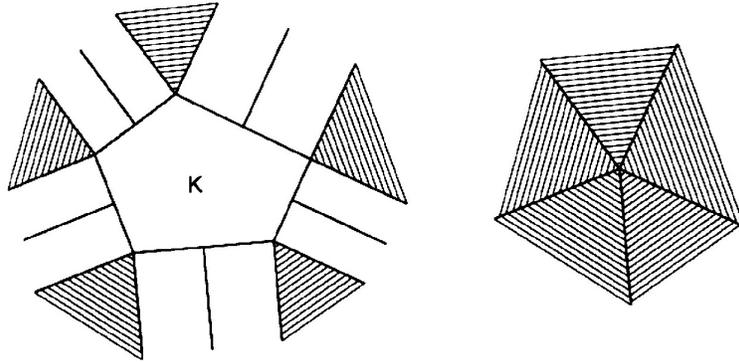


Figura 2.1: Conos normales asociados al politopo K; ref. [13].

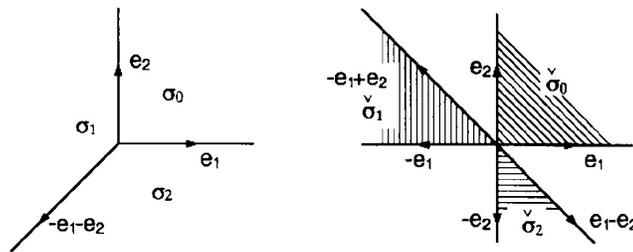


Figura 2.2: Conos duales asociados a cada cono asociado a los conjuntos proyectivos; ref. [13].

# Capítulo 3

## Resolución de singularidades

En este capítulo definiremos una singularidad desde el contexto de variedades tóricas. Enseguida mostraremos la existencia de resolución de singularidades enunciando sin demostración el famoso teorema de Hironaka. Posteriormente definiremos el blow up tórico o modificación tórica, a partir de retomar el concepto definido en la última sección del capítulo primero, y se trata de la subdivisión estelar regular, y este será junto con la técnica usual de blow up los métodos de resolución de singularidades que se aplicarán en la última parte de nuestro trabajo, las definiciones lemas y teoremas así como sus demostraciones de este capítulo fueron tomadas de los trabajos referenciados en la bibliografía de los siguientes autores, [13] **G. Ewald**, [15] **H. Hironaka**, [16] **R. Hartshorne**, [32] **S. Watanabe**, a excepción de los **ejemplos (a) y (b)** y el **teorema 3.1.7**, los cuales fueron probados por el autor de este trabajo.

### 3.1. Teorema de Hironaka y modificación tórica

**Definición 3.1.1.** Sea  $\sigma$  un cono reticular  $n$ -dimensional; cualquier cono se puede descomponer siempre en suma directa de espacios vectoriales, de la siguiente manera,

$$\sigma = \sigma_0 + U,$$

donde  $\sigma_0 \subset \mathbb{R}^n$  es un cono reticular con ápice 0 y  $U \subset \mathbb{R}^n$  se dice que es el subespacio lineal maximal de  $\mathbb{R}^n$  incluido en  $\sigma$ , también  $U$  se define como  $U := \sigma \cap (-\sigma)$  ya hemos estado familiarizados con este cono reticular en el caso de que  $\sigma \cap (-\sigma) = \{0\}$ , y en ese caso hemos dicho que  $\sigma$  es

como fuertemente convexo;  $U$  será llamado el **coespacio lineal** de  $\sigma$  y lo denotamos como **cospan** $\sigma$ .

Sea  $\sigma$  un cono reticular no trivial, esto es, este puede ser descompuesto como  $\sigma = \sigma_0 + U$  como ya se ha visto anteriormente. Esto puede ser reducido al caso de conos con ápice cero.

**Teorema 3.1.1.** Para un cono reticular  $n$ -dimensional  $\sigma \in \mathbb{R}^n$ , sea  $d := n - \dim\sigma_0$ . Entonces, la variedad tórica afín  $X_\sigma$  tiene estructura de un **producto cartesiano**

$$X_\sigma \simeq X_{\sigma_0} \times (\mathbb{C}^*)^d.$$

En particular, si el cono  $\sigma$  es regular, entonces,

$$X_\sigma \simeq \mathbb{C}^{n-d} \times (\mathbb{C}^*)^d.$$

**Demostración.** Primero elegimos un conjunto de generadores  $a_1, \dots, a_{k'}$  para  $\sigma_0 \cap \mathbb{Z}^n$  y  $b_1, \dots, b_{k''}$  para  $U \cap \mathbb{Z}^n$ , respectivamente. Su unión genera el monoide  $\sigma \cap \mathbb{Z}^n$ . Toda relación lineal de la forma  $\sum v_i * a_i = \sum \mu_j * b_j = 0$  es trivial, siendo que los conos  $\sigma_0 \subset \mathbb{R}^n$  y  $U \subset \mathbb{R}^n$  se encuentran en suma directa como subespacios vectoriales sobre  $\mathbb{R}$ , así que cualquier relación que involucre ambos conjuntos de generadores se descompone únicamente como la suma de relaciones que involucran  $a_1, \dots, a_{k'}$  y otras que involucran a  $b_1, \dots, b_{k''}$ . Ahora sean  $u_1, \dots, u_{k'}$  y  $v_1, \dots, v_{k''}$  funciones coordenadas que son representaciones de  $V'$  de  $X_{\sigma_0}$  en  $\mathbb{C}^{k'}$  y  $V''$  de  $X_U \simeq (\mathbb{C}^*)^d$  en  $\mathbb{C}^{k''}$  y pueden ser escritas como  $u_i = t^{a_i}$ ,  $i = 1, \dots, k'$  y  $v_j = t^{b_j}$ ,  $j = 1, \dots, k''$ , respectivamente. Entonces,  $u_1, \dots, u_{k'}, v_1, \dots, v_{k''}$  son las funciones coordenadas para la subvariedad afín  $V$  de  $\mathbb{C}^{k'+k''}$  que representan a  $X_\sigma$ . Aquí no hay relaciones algebraicas de la forma  $f(u) = g(v)$  con polinomios no constantes  $f$  y  $g$  puesto que distintos sistemas de generadores en distintas dimensiones en este caso  $k'$  y  $k''$  inducen distintas construcciones geométricas de variedades tóricas y por tanto distintas funciones coordenadas en ellas, i.e.,  $X_{\sigma_0} \cap X_U = \emptyset$ , así al tomar un punto  $(x, y) \in \mathbb{C}^{k'} \times \mathbb{C}^{k''}$  este pertenece a  $V$  sí solo sí  $x \in V'$  y  $y \in V'' \simeq (\mathbb{C}^*)^d$ , i.e.,  $V = V' \times V''$ , lo cual prueba la afirmación, **q.e.d.**

**Definición 3.1.2.(Singularidad)** Sea  $X_\Sigma$  una variedad tórica n-dimensional y  $\Sigma$  un abanico regular. Un punto  $x \in X$  es llamado **singular o una singularidad** de  $X_\Sigma$  si alguna carta afín  $X_{\sigma^\vee}$ ,  $\sigma \in \Sigma$ , a la que este pertenece, no es de la forma  $\mathbb{C}^k \times (\mathbb{C}^*)^{n-k}$  de acuerdo al teorema anterior.

A continuación enunciamos el teorema principal debido a Heisuke Hironaka, sin demostración y posteriormente enunciamos una variante de este importante teorema debido a Atiyah y el cual es con el que estaremos trabajando.

**Teorema Principal I (H. Hironaka 1963.) 3.1.2** Sea  $\mathbf{B}$  un campo de característica cero. Si  $X$  es un  $\mathbf{B}$ - esquema algebraico, reducible e irreducible, entonces existe un sub-esquema  $D$  de  $X$  tal que,

- (i) el conjunto de puntos de  $D$  es exactamente los lugares singulares de  $X$ , y
- (ii) si  $f : \tilde{X} \rightarrow X$  es la transformación de monomios de  $X$  con centro  $D$ , entonces  $\tilde{X}$  es no singular.

Vemos que el subconjunto abierto complementario de  $X$  a  $D$ , lo denotamos como  $\mathbf{U}$ , es denso en el espacio base de  $X$  y  $f$  induce un isomorfismo de  $X|_{f^{-1}(U)}$  a  $X|_U$ . Reemplazando la condición (i) por estas condiciones sobre  $D$ , obtenemos el teorema principal I en la forma débil, para ver detalles de la demostración ver [13].

**Teorema 3.1.3 (Teorema de Hironaka-Atiyah).** Sea  $f$  una función analítica real en una vecindad de  $\omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n$  con  $f(\omega) = 0$ . Entonces ahí existe un conjunto abierto  $V$ , una variedad analítica real  $U$  y un mapeo analítico propio  $g$  de  $U$  a  $V$  tal que:

- (a)  $g : U - \epsilon \rightarrow V - f^{-1}(0)$  es un isomorfismo, donde  $\epsilon = g^{-1}(f(0))$ ,
- (b) Para cada  $u \in U$ , existen coordenadas analíticas locales  $(u_1, \dots, u_n)$  tal que  $f(g(u)) = \pm u_1^{s_1} u_2^{s_2} \dots * u_n^{s_n}$ , donde  $s_1, \dots, s_n$  son enteros no negativos; ver ref. [29].

**Comentarios.** Este Teorema es una versión del bien conocido teorema en geometría algebraica de resolución de singularidades debido a Hironaka. Donde  $f(\omega)$  no es un polinomio pero es una función analítica, las singularidades de  $f$  pueden ser localmente resueltas, la prueba de Hironaka es completamente constructiva, y el mapeo  $g$  puede ser construido de procedimientos finitos recursivos

de blowing ups de variedades no singulares contenidas en conjuntos singulares. En el teorema ambos  $U$  y  $W$  son espacios localmente Hausdorff y  $g$  es continua. En este caso  $g$  es un mapeo propio sí y solo sí para un conjunto compacto arbitrario  $K$ ,  $g^{-1}(K)$  es compacto, [30], [32], [33].

### Modificación tórica.

**Teorema 3.1.4.** El mapeo definido como

$$T \times X_{\sigma^\vee} \longrightarrow X_{\sigma^\vee}, \quad (t, x) \mapsto tx = (t^{a_1} * x_1, \dots, t^{a_k} * x_k)$$

es una acción del toro sobre la variedad tórica afín  $X_{\sigma^\vee}$ , **llamado la acción natural del toro.** Esta se extiende de manera natural al toro en sí misma. El toro encajado es una órbita de la acción, llamada **la gran órbita.**

**Demostración.** Sea  $\sigma$  un cono en  $\mathbb{R}^n$  con ápice cero, y sea  $A = (a_1, \dots, a_k)$  un sistema de generadores del monoide  $\sigma^\vee \cap \mathbb{Z}^n$ . Ya sabemos que la elección de dicho sistema de generadores determina la construcción geométrica de la variedad tórica afín  $X_{\sigma^\vee}$  como una subvariedad algebraica  $V$  de  $\mathbb{C}^k$ . Sin embargo, sabemos que el mapeo  $T \longrightarrow \mathbb{C}^k$  dado por  $t \longrightarrow (t^{a_1}, \dots, t^{a_k})$  mapea a  $T$  biyectivamente sobre el subconjunto abierto denso  $V \cap (\mathbb{C}^*)^k$  de  $V$ , i.e., el encajamiento del toro, y ahora para  $x = (x_1, \dots, x_k) \in V$  y  $t \in T$ , tenemos que  $(t^{a_1} * x_1, \dots, t^{a_k} * x_k)$  define una acción de  $T$  como grupo en  $V$  por multiplicación compleja coordenada a coordenada. Y el encajamiento del toro es una órbita de la acción, llamada **la gran órbita.**

Para ver que la acción es independiente de la elección de  $A$ , damos dos sistemas de coordenadas que generan la variedad abstracta afín  $X_{\sigma^\vee}$ , correspondiente a elecciones diferentes  $A$  y  $A'$  de generadores de monoides y así diferentes construcciones geométricas  $V = V(a)$  en  $\mathbb{C}^k$  y  $V' = V'(a')$  en  $\mathbb{C}^m$ , existe una transformación de coordenadas  $\varphi : V \longrightarrow V'$ ,  $\varphi$  induce una biyección entre las copias encajadas de los toros en  $V$  y  $V'$ , respectivamente. Entonces es fácil ver que, para todo punto dado  $x = (x_1, \dots, x_k) \in V$  su imagen  $\varphi(x) = (y_1, \dots, y_m) \in V'$  y  $\forall t \in T$ ,

$$\varphi(t^{a_1} * x_1, \dots, t^{a_k} * x_k) = (t^{a'_1} * y_1, \dots, t^{a'_m} * y_m),$$

la identificación de la variedad abstracta afín  $X_{\sigma^\vee}$  con la construcción geométrica  $V \subset \mathbb{C}^k$ , es, de esa manera compatible con la acción del toro sobre  $V$ , **q.e.d.**

**Lema 3.1.1.** Si una singularidad  $x \in X_\Sigma$  pertenece a una órbita  $T_k$  de  $X_\Sigma$ , entonces todos los puntos de la órbita  $T_k$  son singularidades. El gran toro no contiene singularidades, ver ref. [5].

**Demostración.** Si un punto de  $T_k$  está en  $X_{\sigma^\vee}$ , entonces así lo están todos los puntos. Por tanto el lema es cierto por la definición de singularidad. En el gran toro  $T = (\mathbb{C}^*)^n$ , igualmente por definición de singularidad, no hay singularidad, **q.e.d.**

**Definición 3.1.3.** Sean  $X_\Sigma, X_{\Sigma'}$  variedades con encajamiento de toros  $T \subset X_\Sigma, T' \subset X_{\Sigma'}$ . Sea  $\Psi : X_\Sigma \rightarrow X_{\Sigma'}$  un mapeo, y  $\phi(c) : T \rightarrow T'$  un homomorfismo tal que

$$\Psi(cx) = \phi(c) * \Psi(x), \forall c \in T.$$

Entonces, llamamos a  $\Psi$  equivariante respecto a  $\phi$ .

**Definición 3.1.4. (Modificación tórica)** Sea  $\Sigma$  un abanico regular, y sea  $s(p, \sigma)$  una subdivisión estelar regular de  $\Sigma$ . Entonces, el morfismo tórico

$$\Psi_\sigma : X_{s(p, \sigma)\Sigma} \rightarrow X_\Sigma$$

inducido por el mapeo identidad  $id$ ,  $\Psi_\sigma = id'$ , es llamado un **blow down equivariante** de  $X_{s(p, \sigma)\Sigma}$ .

La operación inversa  $\Psi_\sigma^{-1}$  se dice que es un **blow up equivariante o blow up tórico** de  $X_\Sigma$ .

**Definición 3.1.5.** Sea  $\tau$  una cara propia de  $\sigma$  y sea  $a_1, \dots, a_k$  un sistema de generadores de el monoide  $\tau \cap \mathbb{Z}^n$  y  $a_1, \dots, a_k, a_{k+1}, \dots, a_q$  generadores del monoide  $\sigma^\vee \cap \mathbb{Z}^n$ . Los cuales inducen sistemas de funciones coordenadas  $(u_1, \dots, u_k)$  y  $(u_1, \dots, u_k, u_{k+1}, \dots, u_q)$  con  $u_i = t^{a_i}$ . Entonces hay un mapeo natural definido por

$$\begin{aligned} \varphi : X_\tau &\rightarrow X_{\sigma^\vee} \\ (u_1, \dots, u_k) &\mapsto (u_1, \dots, u_k, 0, \dots, 0). \end{aligned}$$

Decimos que una subvariedad cerrada de una variedad tórica afín  $X_{\sigma^\vee}$  es llamada una **subvariedad invariante afín (subvariedad tórica cerrada)** si esta es de la forma  $\varphi(X_\tau) := X_{\sigma^\vee} \cap (u_{k+1} = \dots = u_q = 0)$  de  $X_{\sigma^\vee}$ .

**Teorema 3.1.5.** Sea  $X_\Sigma$  una variedad tórica regular, y sea  $X_{\Sigma_0}$  una subvariedad tórica invariante definida por la estrella  $st(\sigma, \Sigma) \simeq \Sigma_0$  de  $\sigma$  en  $\Sigma$ ;  $1 < k := \dim \sigma \leq n$ .

(a) Bajo el blow up tórico  $\Psi_\sigma^{-1}$ , cualquier punto  $x \in X_{\Sigma_0}$  es reemplazado por un **(k-1)-Espacio proyectivo k-dimensional**.

(b) El blow down  $\Psi_\sigma$  es un morfismo tórico que es biyectivo en el exterior de  $\Psi_\sigma^{-1}$ .

**Demostración.** Sea  $\sigma = \text{Con}(a_1, \dots, a_k)$  y sea  $p := a_1 + \dots + a_k$ . Primero asumimos que  $\dim \sigma' = n$  para un  $\sigma' \in st(\sigma, \Sigma)$ , y  $\sigma' = \text{Con}(a_1, \dots, a_k, a_{k+1}, \dots, a_n)$ . Entonces por subdivisión estelar regular en dirección de  $p$ ,  $\sigma'$  es descompuesto en conos n-dimensionales  $\sigma_i := \text{Con}(a_1, \dots, a_{i-1}, p, a_{i+1}, \dots, a_k, \dots, a_n)$ ,  $i = 1, \dots, k$  y por el teorema 1.1.1 podemos escribir que

$$(\sigma^\vee)^\vee = \text{Con}(b_1, \dots, b_k, \dots, b_n), \text{ y}$$

también calculamos los conos duales de los conos  $\sigma_i$ , y se obtienen aplicando los resultados del lema 1.2.5., así,

$$\sigma_i^\vee = \text{Con}(b_1 - b_i, \dots, b_{i-1} - b_i, b_i, b_{i+1} - b_i, \dots, b_k - b_i, b_{k+1}, \dots, b_n),$$

donde  $i = 1, \dots, k$ . Por tanto,  $X_{(\sigma^\vee)^\vee}$  tiene coordenadas  $(u_1, \dots, u_n) = (t^{b_1}, \dots, t^{b_n})$  y  $X_{\sigma_i^\vee}$  tiene coordenadas  $(v_1, \dots, v_n) = (t^{b_1 - b_i}, \dots, t^{b_{i-1} - b_i}, t^{b_i}, t^{b_{i+1} - b_i}, \dots, t^{b_k - b_i}, t^{b_{k+1}}, \dots, t^{b_n})$ . Sustituyendo las coordenadas  $u_i$  respectivamente como aparecen en las coordenadas  $v_i$ , conseguimos el morfismo  $\Psi_\sigma|_{\sigma_i^\vee}$  esta dado por las parametrizaciones,

$$u_l = v_l v_i, \dots, u_{i-1} = v_{i-1} v_i, u_i = v_i,$$

$$u_{i+1} = v_{i+1} v_i, \dots, u_k = v_k v_i, u_{k+1} = v_{k+1}, \dots, u_n = v_n.$$

Su inverso esta definido por  $v_i \neq 0$ . Si  $v_i = 0$ , las imagenes de todos los puntos  $(v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_k, v'_{k+1}, \dots, v'_n)$  son  $(0, \dots, 0, v'_{k+1}, \dots, v'_n)$ , para cualquier conjunto fijo  $v' := \{v'_{k+1}, \dots, v'_n\}$  de coordenadas, esto es, la fibra  $X_0^{(i)}$  de  $\Psi_\sigma|_{X_{\sigma_i^\vee}}$ , arriba  $(0, \dots, 0, v'_{k+1}, \dots, v'_n)$  es un  $(k-1)$ -espacio afín  $X_{v'}^{(i)}$ . Cualquiera de dos tales espacios  $X_{v'}^{(i)}$  son isomorfos a cada uno de los otros en  $X_0^{(i)}$ . Pero los espacios  $X_0^{(i)}$ ,  $i = 1, \dots, k$ , pueden ser considerados como las cartas afines del  $(k-1)$ -espacio proyectivo definido por los generadores,  $b_1 - b_2, b_2 - b_3, \dots, b_{k-1} - b_k, b_k - b_1$  ellos están en  $(\text{Con}\{p\})^\perp$  y suman el vector 0; i.e., son una subdivisión estelar regular en dirección de  $p$ , de acuerdo al teorema 1.3.1; y todo este procedimiento, hemos mostrado anteriormente es similar al haber encontrado las bases de Hilbert del cono inicial  $\sigma'$  y posteriormente las bases de Hilbert de los conos  $\sigma_i$ , y este procedimiento es muy útil para conos de alta dimensión; la subdivisión estelar regular la utilizamos en la prueba como un método constructivo para el blow up tórico  $\Psi_\sigma|_{\sigma_i^\vee}$ .

Ahora si la dimensión maximal de un cono  $\sigma' \in \text{st}(\sigma, \Sigma)$  es menor que  $n$ , decimos  $r = \dim \sigma' < n$ ,  $(\sigma')^\vee$  tiene  $\text{cospan} U \neq 0$ . Así  $\sigma_i = \sigma_{0i} + U$  nuevamente aplicando el lema 1.2.5. construimos los conos  $\sigma_i^\vee$ , reemplazamos  $n$  por  $r$ ,  $i = 1, \dots, k$ . Entonces se puede ver fácilmente que para cada  $\text{cospan} \sigma_i^\vee$ ,

$$\text{cospan}(U^\vee) = \text{cospan}(\sigma')^\vee = \text{cospan} \sigma_1^\vee = \dots = \text{cospan} \sigma_k^\vee,$$

tal que  $X_{(\sigma')^\vee}, X_{\sigma_1^\vee}, \dots, X_{\sigma_k^\vee}$  tienen las mismas coordenadas  $u_j$  para  $j > k$ . De ese modo todos los argumentos de arriba pueden ser aplicados a cualquier variedad  $X_{\sigma_k^\vee}$ , y obtenemos el mismo resultado para  $r < n$ , así terminamos la prueba **q.e.d.**

**Definición 3.1.6.** Sea  $X_\Sigma$  una variedad tórica que tiene singularidades, y sea  $X_{\Sigma'}$  una variedad tórica regular para la que  $|\Sigma| = |\Sigma'|$ , y un morfismo tórico

$$\Psi : X_\Sigma \longrightarrow X_{\Sigma'}$$

existe un isomorfismo en el toro. Entonces, llamamos a  $\Psi$  **una resolución de singularidades de  $X_\Sigma$** .

**Teorema 3.1.6.** Una variedad tórica  $X_\Sigma$  con singularidades tiene una resolución  $\Psi$ . Podemos elegir  $\Psi$  como una composición de morfismos  $\Psi_1, \dots, \Psi_q$  que provienen de las subdivisiones estelares  $s_1, \dots, s_q$ :

$$\begin{array}{ccccccc}
 \Sigma' & & \longleftarrow_{s_q} & \cdots & \longleftarrow_{s_1} & & \Sigma \\
 \cdot & & & & & & \uparrow \\
 \cdot & & & & & & \cdot \\
 \cdot & & & & & & \cdot \\
 \downarrow & & & & & & \cdot \\
 X_{\Sigma'} & \longrightarrow_{\Psi_{s_q}} & \cdots & \longrightarrow_{\Psi_{s_1}} & & & X_\Sigma
 \end{array}$$

**Demostración.** Primero, por subdivisiones estelares, regresamos a  $\Sigma$  dentro de un abanico simplicial  $\Sigma'$ . Sea  $\sigma \in \Sigma'$  un cono maximal no regular,  $\dim \sigma = k$ . Aplicamos inducción sobre  $k$  y asumimos que una  $(k-1)$ -cara  $\sigma_0$  de  $\sigma$  ha sido obtenida por subdivisiones estelares de  $\Sigma$ . Así  $\sigma$  es descompuesto dentro de  $k$ -símplexos. Sea  $\tau$  uno de ellos,  $\tau = \tau_0 + \varrho$ , donde  $\tau_0 \subset \sigma_0$  y  $\tau_0$  es regular. Salvo una transformación unimodular, podemos asumir que  $\tau_0 = \text{Con}(e_1, \dots, e_{k-1})$ . Entonces,  $\varrho = \mathbb{R}_{\geq 0}a$  con  $a$  primitivo,  $a = \alpha_1 e_1 + \dots + \alpha_{k-1} e_{k-1} + \alpha e_k$  (en  $\mathbb{R}^k$  expandido por  $\tau$ ) y  $\alpha_1, \dots, \alpha_{k-1}, \alpha \in \mathbb{Z}_{\geq 0}$ .

Si  $\alpha = 1$ ,  $\tau$  es regular. Si  $\alpha > 1$ , ahí existe un punto reticular  $b = \beta_1 e_1 + \dots + \beta_{k-1} e_{k-1} + \frac{1}{\alpha} a$ ,  $0 \leq \beta_i < 1$ ,  $i = 1, \dots, k-1$ . Obtenemos

$$\alpha' := |\det(e_1, \dots, e_{i-1}, b, e_{i+1}, \dots, e_{k-1}, a)| < \alpha = \det(e_1, \dots, e_{k-1}, a).$$

Aplicamos a  $\Sigma$  la subdivisión estelar  $s(b; \Sigma)$  en dirección de  $b$ . Todo cono  $k$ -dimensional afectado por  $s(b; \Sigma)$  se descompone dentro de conos con menor determinante que la matriz de generadores. Así, después de un número finito de pasos conseguimos conos de tal manera que al elegir cualquiera

de ellos, digamos  $\sigma = \text{Con}(\alpha_1, \dots, \alpha_r)$  obtengamos  $|\det(\alpha_1, \dots, \alpha_r)| = \pm 1$ , lo cual siempre es posible siendo que  $\sigma \subset \mathbb{Z}^n$ , y que de acuerdo a la definición 1.1.9, estos conos, son conos regulares, con esto concluimos la prueba, **q.e.d.**

**Ejemplo (a).** Consideremos la siguiente superficie cónica  $z_1 z_2 = z_3^2$  la cual tiene singularidad en  $(0, 0, 0) \in \mathbb{R}^3$ . Y sea  $\sigma = \text{Con}((1, 0), (1, 2))$  el cono poliédrico que pertenece a cualquier abanico;  $\sigma^\vee = \text{Con}((0, 1), (2, -1))$ , donde los generadores del cono dual fueron calculados como vectores normales a los generadores de  $\sigma$ . De esa manera obtenemos geoméricamente la siguiente variedad tórica afín  $X_\sigma^\vee$ .

$$X_{\sigma^\vee} = \mathbf{Spec} \mathbb{C}[z_1, z_2, z_3] = \mathbb{C}[v, u^2 v^{-1}, u]$$

las relaciones entre los monomios generadores de esta variedad parametrizan la superficie cónica. Ahora subdividimos  $\sigma$  en la dirección del rayo  $\mathbb{R}_{\geq 0}(e_1 + e_2)$ . Sea  $\sigma_1 = \text{Con}(e_1, e_1 + e_2)$  y  $\sigma^\vee = \text{Con}(e_2, e_1 - e_2)$ . Entonces, podemos conseguir la variedad  $X_{\sigma_1^\vee} = \mathbf{Spec} \mathbb{C}[v_1, v_2] = \mathbb{C}[t_2, t_1 t_2^{-1}]$ , y el blow down  $\Psi_1$  esta dado por

$$(z_1, z_2, z_3) \longrightarrow (v_1, v_1 v_2^2, v_1 v_2).$$

Así primero, mapeamos cada punto  $(v_1, v_2)$  sobre el punto  $(v_1, v_2, v_1 v_2)$  de una superficie regular cuadrática, y entonces proyectamos sobre la superficie cónica, y entonces realizamos el blow up tórico al proyectar esta superficie cónica  $(v_1, v_1 v_2^2, v_1 v_2)$ . a la línea  $\{(0, v_2)\}$ , después sobre la línea  $\{(0, v_2, 0)\}$  y entonces sobre el punto  $(0, 0, 0)$ , mientras que las líneas  $\{(c, v_2)\}$  para  $c \neq 0$  son mapeadas sobre las líneas  $(c, v_2, c v_2)$  y entonces sobre las parábolas  $\{(c, c v_2^2, c v_2)\}$ .

Más claramente definimos el blowing up  $\Psi_1^{-1}$  de  $X_{\tau^\vee}$  donde  $\tau^\vee = \text{Con}(e_1, e_2, e_3)$  es una cara de  $\sigma^\vee$  y una de las cartas de  $\Psi^{-1}(X_{\tau^\vee})$  esta dada por  $X_{\sigma_0^\vee}$  donde  $\sigma_0 = \text{Con}(e_1, e_1 + e_2, e_3)$ , y, de ese modo conseguimos, que  $\sigma_0^\vee = \text{Con}(e_2, e_1 - e_2, e_3)$ . La construcción geométrica de  $X_{\sigma_0^\vee}$  es entonces,  $(v_1, v_2, v_3) = (t_2, t_1 t_2^{-1}, t_3)$  de esta igualdad se sigue  $(z_1 z_2, z_3) = (v_1, v_1 v_2^2, v_3)$ .

Ahora,  $v_3 = v_1v_2$  de nuevo representa la superficie de arriba, esta vez obtenida del cono cuadrático por blowing up tórico  $\mathbb{C}^3$  a lo largo de  $\{(0, z_2, 0)\}$ , ver fig. 3.1.

**Observación.** El procedimiento de desingularización desarrollado mediante subdivisión estelar arriba; se puede resumir y desarrollar para cálculos posteriores mediante el cálculo de las bases de Hilbert sobre los generadores del monoide que definen el cono poliédrico del soporte definido por el polinomio de la superficie puesto que los elementos de la base de Hilbert definen conos regulares minimales y siendo que sus generadores generan todas las cartas afines de la variedad, lo cual es misión de la subdivisión estelar regular, de ese modo tenemos procedimientos equivalentes de construcción uno combinatorio y otro más algebraico. Veamos a continuación.

Consideremos nuevamente la superficie cónica anterior como el siguiente polinomio  $h(z_1, z_2, z_3) = z_1z_2 - z_3^2$  elegimos una cara dual del cono definido por la superficie cónica  $\sigma^\vee = \text{Con}(e_1, e_1 + 2e_3)$  estos generadores se encuentran contenidos en el cono definido por el conjunto  $\text{supp}(h) = \{e_1 + e_2, 2e_3\}$ , y calculamos las bases de Hilbert asociadas a este cono mediante el programa **Nórmaliz** o el algoritmo sugerido en el apéndice A, con lo cual conseguimos la base de Hilbert  $H_{\sigma^\vee} = \{e_1, e_1 + e_3, e_1 + 2e_3\}$ , ahora podemos conseguir una construcción de la variedad tórica afín definida por el cono poliédrico mediante la siguiente carta afín  $X_{\sigma^\vee} = \{(u_1, u_2, u_3)\} = \mathbf{Spec}\mathbb{C}[u_1, u_1u_3, u_1u_3^2]$ , de donde conseguimos el mapeo de resolución

$$g : (z_1, z_2, z_3) \longrightarrow (u_1, u_1u_3, u_1u_3^2).$$

aquí la parametrización es similar a la obtenida anteriormente pero con posición en el anillo coordenado distinta, esto es debido a la elección de la cara contenida en el cono poliédrico elegida, pero para nuestros propósitos de desingularización realizan el mismo efecto y ya que debido al teorema de Hironaka-Atiyah se elige que  $u = (u_1, u_2, u_3) \neq (0, 0, 0) \in \mathbb{R}^3$ , obtenemos

$$h(g(u)) = u_1^2u_3(1 - u_3^3).$$

con lo cual se obtiene un polinomio con características establecidas por el teorema de Hironaka-Atiyah, también hemos conseguido utilizar un técnica basada netamente por el cálculo de las bases de Hilbert la cual implícitamente contiene la técnica de subdivisión estelar regular y donde esta conlleva a un modo muy extenso de cálculos pero sin embargo, nos funciona muy bien para formalizar de manera amplia la definición de **blow up tórico o modificación tórica**.

**Ejemplo (b).** Consideremos el polinomio  $H(a, b, c) = a^2b^2 + 2abc + c^2 + 3a^2b^4$ ,  $(a, b, c) \in \mathbb{R}^3$  el cual es singular en  $(0, 0, 0) \in \mathbb{R}^3$ . Construimos el politopo de Newton definido por  $\text{supp}(H)$ , del cual obtenemos el siguiente como poliédrico definido por,  $\sigma = \text{Con}((2, 2, 0), (1, 1, 1), (0, 0, 2), (2, 4, 0))$ .

Ahora conseguimos el siguiente como dual asociado de acuerdo al teorema 1.1.1 tenemos que

$$\sigma^\vee = \text{Con}(2e_1 - e_2 - e_3, -e_1 + e_2, e_3)$$

del cual obtenemos las bases de Hilbert asociadas al monoide  $\sigma \cap \mathbb{Z}^3$  de acuerdo al apéndice A o el programa Nórmliz, así conseguimos

$$\mathbb{H}_{\sigma^\vee} = \{e_1 + e_2, e_1 + e_2 + e_3, e_1 + 2e_2\},$$

ahora de acuerdo a nuestros resultados expuestos en el teorema 2.2.1 podemos obtener geométricamente la siguiente variedad tórica afín  $X_{\sigma^\vee}$  donde,

$$X_{\sigma^\vee} = \mathbf{Spec}(\mathbb{C}[S_{\sigma^\vee} \cap \mathbb{Z}^3]) = \mathbf{Spec}(\mathbb{C}[u_1u_2, u_1u_2u_3, u_1u_2^2]).$$

Y que de acuerdo a la demostración del teorema 3.1.5, elegimos el conjunto de generadores de  $X_{\sigma^\vee}$  los cuales forman una matriz unimodular  $A$  este proceso fue equivalente a obtener la mejor subdivisión estelar en dirección del vector reticular  $p = (2, 2, 0) + (1, 1, 1) + (0, 0, 2) + (2, 4, 0)$ , en este caso de acuerdo a los generadores  $A = \text{Columnas}((1, 1, 0)', (1, 1, 1)', (1, 2, 0)'),$  y parametrizamos mediante este sistema de monomios de Laurent.

Por el teorema de Hironak-Atiyah, obtenemos el siguiente mapeo de resolución  $g_1 : (a, b, c) \rightarrow (u_1u_2, u_1u_2u_3, u_1u_2^2)/(0, 0, 0)$  tal que,

$$\begin{aligned}
H(g_1(u_1, u_2, u_3)) &= u_1^4 u_2^4 u_3^2 + 2u_1^3 u_2^4 u_3 + u_1^2 u_2^4 + 3u_1^6 u_2^6 u_3^4 \\
&= u_1^2 u_2^4 (u_1^2 u_3^2 + 2u_1 u_3 + 1 + 3u_1^4 u_2^2 u_3^4) \\
&= u_1^2 u_2^4 ((u_1 u_3 + 1)^2 + 3u_1^4 u_3^4 u_2^2) \\
&= c_1^2 ((b_1 + 1)^2 + 3b_1^4 d_1^2) \\
&= c_1^2 (b_1'^2 + 3(b_1' - 1)^4 d_1^2) \\
&= c_1^2 (b_1'^2 + 3e_1^4 d_1^2)
\end{aligned}$$

volvemos aplicar por segunda vez la técnica de resolución mediante bases de Hilbert al polinomio definido por  $h(b_1', d_1, e_1) = b_1'^2 + 3e_1^4 d_1^2$ , de donde conseguimos la base de Hilbert asociada al politopo de Newton del  $\text{supp}(h)$ , de ese modo se tiene usando el programa Nórmaiz o el algoritmo del apéndice A,  $H_{\sigma^v} = \{e_1, e_1 + e_2, e_1 + 2e_2\}$ , de la cual conseguimos el siguiente mapeo de resolución

$$g_2 : (b_1', d_1, e_1) \longrightarrow (s_1, s_1 s_2, s_1 s_2^2) / (0, 0, 0).$$

del cual se obtiene la siguiente variedad tórica afín;  $X_{\sigma^v} = \text{Spec} \mathbb{C}[s_1, s_1 s_2, s_1 s_2^2]$ , de donde conseguimos

$$\begin{aligned}
H(g_1(g_2(s_1, s_2, s_3))) &= c_1^2 s_1^2 (1 + 3s_1^4 s_2^8 s_2^2) \\
&= c_1^2 s_1^2 (1 + 3s_1^4 s_2^{10}).
\end{aligned}$$

el cual no es singular en  $(0, 0, 0) \in \mathbb{R}^3$ , además nuestro resultado reproduce el mismo cálculo para la máquina de aprendizaje asociada a este polinomio el cual veremos a profundidad en la parte dos de este trabajo, ver ref. [32], [39].

**En resumen.** La interpretación geométrica a todo esto es la siguiente, la estrella del abanico construido a partir del cono poliédrico inicial  $st(\Sigma, \sigma)$  tiene que ver con el conjunto de normales en  $\sigma$  y detecta el conjunto de conos cara que no son regulares en el abanico  $\Sigma$  construido a partir de  $\sigma$  la forma de resolver esta irregularidad es mediante una subdivisión estelar regular, la manera

de realizarla algebraicamente es encontrando los generadores asociados a los monoides de los conos reticulares involucrados (bases de Hilbert), al hacer esto la subdivisión estelar se presenta como la unión de los conos regulares que ya existían y sustituyendo las partes irregulares por subespacios lineales minimales como  $\tau + \text{Con}(p)$  donde esta definido en el teorema 1.3.1 y  $\text{Con}(p)$  es el rayo vector en la dirección en la que se realiza la subdivisión estelar,  $\tau$  es una cara regular que no es cara de las partes irregulares en la estrella, la regularidad es muy importante en resolución de singularidades tóricas puesto que esta suma de espacios es regular, y que de acuerdo al teorema 2.2.1 entonces el conjunto algebraico  $X_{U+\tau+\text{Con}(p)}$  asociado al cono poliédrico racional  $U + \tau + \text{Con}(p)$  regular implica la obtención de una variedad tórica afín, y de acuerdo a esta regularidad de poliédros también aplica el teorema 3.1.1 de donde esta variedad tórica afín es posible verla como el siguiente producto cartesiano de espacios  $\mathbb{C}^{n-d} \times (\mathbb{C}^*)^d$  ( $n$  dimensión del cono reticular y  $d = n - \dim(\tau + \text{Con}(p))$ ) donde es claro el encajamiento de nuestras variedades tóricas afines en variedades algebraicas no singulares, a todo este procedimiento geométrico algebraico le llamamos blow up tórico o modificación tórica, ver. Ref.[11], [13], [14].

A continuación resumimos el desarrollo que hasta el momento llevamos en nuestro trabajo mediante el siguiente teorema.

**Teorema 3.1.7.** Sea  $\Sigma \subset \mathbb{R}^n$  un abanico singular o no regular, y  $\sigma \in \Sigma$  cualquier cono no regular, sea  $H_{\sigma^\vee}$  las bases del Hilbert asociadas a su cono dual. Entonces el proceso de encontrar las bases de Hilbert del cono  $\sigma$  es equivalente a realizar el proceso de subdivisión estelar  $\text{st}(\sigma, p)$  en dirección del vector reticular  $p$ .

**Demostración.** Hacemos uso del siguiente algoritmo de resolución singular para su prueba; consultar blow ups de Nash, ver ref. [1].

1. Dado el cono  $\sigma$ , encontramos las bases de Hilbert  $H_{\sigma^\vee}$  asociadas a su monoide  $\sigma^\vee \cap \mathbb{Z}^d$  mediante cualquier algoritmo como el sugerido en el apéndice A.
2. Encontramos el conjunto  $P = \{h_1 + \dots + h_d \mid h_i \in H_{\sigma^\vee}\}$ , donde cada  $p \in P$  es un vector reticular.
3. Calculamos el envolvente convexo de  $P$ , i.e.,  $\mathbf{conv}(P)$  como se dijo en la definición 1.1.
4. Encontrar la suma de Minkowski  $\sigma^\vee + \mathbf{conv}(P)$ , i.e., todas las combinaciones y desigualdades

que definen este nuevo como reticular.

**5.** Encontrar la localización o monoide del nuevo cono reticular  $\sigma' = (\sigma^\vee + \text{conv}(P))_v$  para cada vértice; lo cual es equivalente a encontrar sus hiperplanos soporte (se puede probar del teorema 1.1.1; que los hiperplanos soporte siempre se encuentran contenidos en las bases de Hilbert).

**6.** Calculamos el determinante asociado a los vectores reticulares que generan el nuevo cono  $\sigma'$  y el cual denotamos por  $\det\sigma'$ , si  $\det(\sigma') = 1$  detemos el proceso por definición de cono regular, si  $\det(\sigma') \neq 1$  aplicamos el algoritmo entero a este nuevo cono  $\sigma'$ ; hasta cumplir con la condición de regularidad, i.e.,  $\det(\sigma') = 1$ , y con ello concluimos la prueba, **q.e.d.**

### 3.2. Blow up

A continuación estudiamos un importante mapeo birracional. El blow up. Primero construiremos el **blow up** de  $A^n$  en el punto  $O = (0, \dots, 0)$ . Consideramos el producto  $A \times \mathbb{C}P^{n-1}$ , que es una variedad quasi-proyectiva. Si  $x_1, \dots, x_n$  son coordenadas afines de  $A^n$ , y si  $y_1, \dots, y_n$  son las coordenadas homogéneas de  $\mathbb{C}P^{n-1}$ , entonces los subconjuntos cerrados de  $A \times \mathbb{C}P^{n-1}$  son definidos por polinomios en  $x_i, y_j$ , que son homogéneos con respecto a los  $y_j$ .

Ahora definiremos el blowing up de  $A^n$  en el punto  $O$  es el subconjunto cerrado  $X$  de  $A \times \mathbb{C}P^{n-1}$  definido por las ecuaciones  $\{x_i y_j = x_j y_i \mid i, j = 1, \dots, n\}$ .

$$\begin{array}{ccc} X & \hookrightarrow & A \times \mathbb{C}P^{n-1} \\ & \searrow \varphi & \downarrow \\ & & A^n \end{array}$$

Tenemos un morfismo natural  $\varphi : X \rightarrow A^n$  obtenido por la restricción del mapeo proyección de  $A^n \times \mathbb{C}P^{n-1}$  sobre el primer factor. Estudiaremos las propiedades de  $X$ .

(1) Si  $p \in A^n, p \neq O$ , entonces  $\varphi^{-1}(O)$  consiste de un solo punto. De hecho,  $\varphi$  da un isomorfismo de  $X - \varphi^{-1}(O)$  sobre  $A^n - O$ . Ahora sea  $p = (a_1, \dots, a_n)$  con algún  $a_i \neq 0$ . Ahora si  $p \times (y_1, \dots, y_n) \in \varphi^{-1}(p)$ , entonces para cada  $j, y_j = (a_j/a_i)y_i$ , y  $(y_1, \dots, y_n)$  es únicamente determinado como un pun-

to en  $\mathbb{C}P^{n-1}$ . Si colocamos  $y_i = a_i$ , podemos tomar  $(y_1, \dots, y_n) = (a_1, \dots, a_n)$ . De esa manera  $\varphi^{-1}(p)$  consiste de un único punto. Además, para  $p \in A^n - O$ , colocamos  $\psi(p) = (a_1, \dots, a_n) \times (a_1, \dots, a_n)$  define un morfismo inverso a  $\varphi$ , mostrando que  $X - \varphi^{-1}(O)$  es isomorfo a  $A^n - O$ .

(2)  $\varphi^{-1}(O) \simeq \mathbb{C}P^{n-1}$ . Y  $\varphi^{-1}(O)$  consiste de todos los puntos  $O \times Q$ , con  $Q = (y_1, \dots, y_n) \in \mathbb{C}P^{n-1}$ , sujeto a que no hay restricción.

(3) Los puntos de  $\varphi^{-1}(O)$  están en correspondencia 1-1 con el conjunto de líneas a través de  $O$  en  $A^n$ . Claramente, una línea  $L$  a través de  $O$  en  $A^n$  puede estar dado por ecuaciones paramétricas  $x_i = a_i t, i = 1, \dots, n$  donde los  $a_i \in k$  son todos no cero, y  $t \in A^1$ . Ahora consideramos la línea  $L' = \varphi^{-1}(L - O)$  en  $X - \varphi^{-1}(O)$ . Esta dado paraméricamente por  $x_i = a_i t, y_i = a_i t$ , con  $t \in A^1 - O$ . Pero los  $y_i$  son coordenadas homogéneas en  $\mathbb{C}P^{n-1}$ , y podemos igualmente describir bien  $L'$  con las ecuaciones  $x_i = a_i t, y_i = a_i$  para  $t \in A^1 - O$ . Estas ecuaciones tienen sentido también para  $t = 0$  y dada la cerradura  $cl(L')$  de  $L'$  en  $X$ . Ahora  $cl(L')$  intersecta  $\varphi^{-1}(O)$  en el punto  $Q = (a_1, \dots, a_n) \in \mathbb{C}P^{n-1}$ , así vemos que mandar  $L$  a  $Q$  da una correspondencia 1-1 entre líneas a través de  $O$  en  $A^n$  y puntos de  $\varphi^{-1}(O)$ .

(4)  $X$  es irreducible. Es claro que si,  $X$  es la unión de  $X - \varphi^{-1}(O)$  y  $\varphi^{-1}(O)$ . La primera pieza es isomorfa a  $A^n - O$ , por tanto irreducible. De otra manera, hemos solo visto que todo punto de  $\varphi^{-1}(O)$  esta en la cerradura de algún subconjunto (la línea  $L'$ ) de  $X - \varphi^{-1}(O)$ . Así  $X - \varphi^{-1}(O)$  es denso en  $X$ , y  $X$  es irreducible.

**Definición 3.2.1.** Si  $Y$  es una variedad de  $A^n$  que pasa a través  $O$ . Definimos el **blow up** de  $Y$  en el punto  $O$  como  $\tilde{Y} = (\varphi^{-1}(Y - O))^-$ , donde  $\varphi : X \rightarrow A^n$  es el blowing up de  $A^n$  en el punto  $O$  descrita como arriba. Denotamos también por  $\varphi : \tilde{Y} \rightarrow Y$  el morfismo obtenido por restricción  $\varphi : X \rightarrow A^n$  a  $\tilde{Y}$ . Para el blow up en cualquier otro punto  $p$  de  $A^n$ , hacemos un cambio lineal de coordenadas que mandan  $p$  a  $O$ .

Notemos que  $\varphi$  induce un isomorfismo de  $\tilde{Y} - \varphi^{-1}(O)$  a  $Y - O$ , tal que  $\varphi$  es un morfismo birracional de  $\tilde{Y}$  a  $Y$ . Notemos también que esta definición aparentemente depende del encajamiento de  $Y$  en  $A^n$ , pero de hecho se puede probar que el blowing up es intrínseco.

El efecto de blowing up en un punto de  $Y$  es pull apart  $Y$  cerca de  $O$  de acuerdo a las diferentes

direcciones de líneas a través de  $O$ .

**Ejemplo (c).** Sea  $Y$  una curva cúbica plana dado por las ecuaciones  $y^2 = x^2(x+1)$ . Haremos blow up de  $Y$  en  $O$ . Sea  $t, u$  coordenadas homogéneas para  $\mathbb{C}P^1$ . Entonces  $X$ , el blowing up de  $A^2$  en  $O$ , es definido por la ecuación  $xu = ty$  dentro de  $A^2 \times \mathbb{C}P^1$ . Este se mira como  $A^2$ , excepto que el punto  $O$  ha sido reemplazado por un  $\mathbb{C}P^1$  correspondiente a las pendientes de las líneas a través de  $O$ , llamaremos a  $\mathbb{C}P^1$  la **curva excepcional** y es denotada por  $E$ .

Obtenemos la imagen inversa total de  $Y$  en  $X$  considerando las ecuaciones  $y^2 = x^2(x+1)$  y  $xu = ty$  en  $A^2 \times \mathbb{C}P^1$ . Ahora  $\mathbb{C}P^1$  es cubierto por conjuntos abiertos  $t \neq 0$  y  $u \neq 0$  que consideramos separadamente. Si  $t \neq 0$ , podemos colocar  $t = 1$ , y usamos  $u$  como parámetro afín. Entonces tenemos las ecuaciones

$$\begin{aligned}y^2 &= x^2(x+1), \\y &= xu,\end{aligned}$$

en  $A^3$  con coordenadas  $x, y, u$ . Substituyendo, conseguimos  $x^2u^2 - x^2(x+1) = 0$ . Así, obtenemos dos componentes irreducibles, uno definido por  $x = 0, y = 0, u$  arbitrario, el cual es  $E$  y el otro definido por  $u^2 = x+1, y = xu$ . Este es  $\tilde{Y}$ , notamos que  $\tilde{Y}$  intersecta  $E$  en los puntos  $u = \pm 1$ . Estos puntos corresponden a las pendientes de las dos ramas de  $Y$  en  $O$ .

Similarmente uno puede revisar que la imagen inversa total de los ejes  $x$  consiste de  $E$  y otra curva irreducible, que nosotros llamamos transformación estricta de los ejes  $x$  (esta es la curva  $cl(L')$  descrita arriba correspondiente a la línea eje- $L = x$ ). Esta transformación estricta intersecta  $E$  en el punto  $u = 0$ . Considerando el otro conjunto abierto  $u \neq 0$  en  $A^2 \times \mathbb{C}P^1$ , uno puede observar que la transformación estricta en del eje- $y$  intersecta  $E$  en el punto  $t = 0, u = 1$ .

El efecto del blowing up es separar las ramas de curvas que pasan a través de  $O$  de acuerdo a sus pendientes. Si las pendientes son diferentes, su transformación estricta no se encontrará fuera del conjunto cerrado  $X$ . Ellas se encuentran en  $E$  en puntos correspondientes a diferentes pendientes,

ver, ref. [16].

**Ejemplo (d).** (Máquina de aprendizaje). Para la siguiente máquina estadística regular definida más adelante de nuestro trabajo, usando la coordenadas apropiadas para este modelo  $(\omega_1, \dots, \omega_d) \in \mathbb{C}^d$  se define la función de pérdida promedio  $f(\omega)$  que es localmente escrita por

$$f(\omega) = \sum_{i=1, \dots, d} \omega_i^2$$

en el abierto  $W = \{\omega : \|\omega\| < 1\}$  este modelo tiene singularidad en el origen  $p = (0, \dots, 0) \in \mathbb{C}^d$ .

Utilizando **blow up** en este punto mediante la siguiente parametrización definida por el mapeo

$$g : (u_1, \dots, u_d) \longrightarrow (\omega_1, \dots, \omega_d),$$

$$\omega_1 = u_1 \quad \omega_i = u_1 u_i, \quad 2 \leq i \leq d.$$

en esta parametrización por el teorema de Hironaka-Atiyah, escribimos

$f(g(u_1, \dots, u_d)) = u_1^2 * (1 + \sum_{i=1, \dots, d} u_i^2)$ . La cual ya no es singular en el origen, ver; ref. [32].

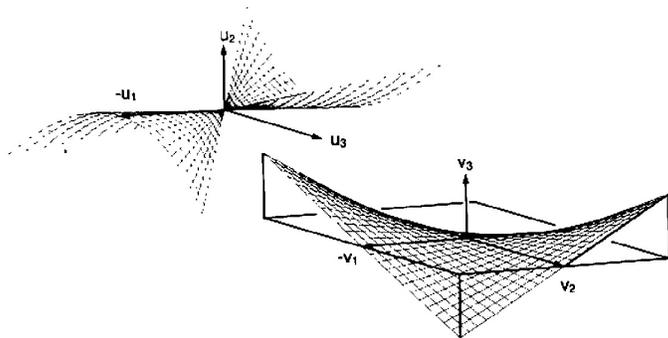


Figura 3.1: Modificación tórica resuelto en la superficie cónica; ref. [13].

# Capítulo 4

## La estadística de redes neuronales Bayesianas

A continuación seguimos con la última parte de este trabajo, dando una noción breve del concepto de redes neuronales Bayesianas (R.N.B.), definición de estas mismas, su importancia teórica como modelos eficaces de la inteligencia artificial en la toma de decisiones, pronósticos de datos, entre otros. Posteriormente abordaremos de manera introductoria los conceptos más centrales de la teoría del aprendizaje estadístico Bayesiano incluyendo uno de los teoremas importantes de este tópico, el teorema de Watanabe Sumio para curvas de aprendizaje, posteriormente la definición en breve de máquinas singulares, con estos conceptos definimos la problemática del aprendizaje por presencia de singularidades en el espacio de parámetros de la red neuronal, viendo este conjunto como un conjunto algebraico. Terminamos la exposición de nuestro trabajo con ejemplos de interés en su implementación, y de reciente publicación en estos últimos años, como es el caso de un perceptrón de tres capas entre otros dos ejemplos más. La teoría de este capítulo que incluye, definiciones, lemas y teoremas así como sus demostraciones fue tomada de los siguientes autores, [32], [33], [37], [39] S. Watanabe; [24], [25] A.S. Poznyak; [12] L.D. García; [8] D.M. Chickering; a excepción del corolario 4.3.3, el cual es una conclusión y resultado de la teoría vista en los capítulos anteriores y es escrito por el autor de este trabajo.

### 4.1. R.N.B y teoría del aprendizaje Bayesiano

Ahora consideramos  $(\Omega, \mathbb{F})$  un espacio de medida y sea  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$  un sistema de conjuntos de Borel de la línea real tal que  $(\Omega, \mathbb{F})$  es  $\mathbb{F}$ -medible, y elegimos el triplete  $(\Omega, \mathbb{F}, P)$  de tal ma-

nera que defina un espacio probabilístico, con  $\Omega$  denotando el espacio muestral de eventos,  $\mathbb{F}$  una  $\sigma$ -álgebra definida sobre este espacio, y  $P$  la probabilidad de Kolmogorov; tomemos un conjunto de observación de eventos que será denotado como  $X = \{x_1, \dots, x_n\} \subset \Omega$ , y diremos que la variable  $x_i = k \in X$ , son variables aleatorias, i.e., existe una función  $\xi$  valuada real, tal que  $\xi(x_i) \in B$  para cada conjunto de Borel  $B \in \mathbb{B}$ , donde la  $i$ -ésima variable se encuentra en el estado  $k$ ; en lo sucesivo de los siguientes capítulos trabajaremos con densidades de probabilidad  $P$  definidas en este espacio probabilístico  $(\Omega, \mathbb{F}, P)$ . Ahora consideramos un dominio  $U$  de  $n$  variables discretas  $x_1, \dots, x_n$  donde cada  $x_i$  tienen un número finito de estados. Para este conjunto  $U$  tenemos la siguiente definición.

**Definición 4.1.1. (R.N.B)** Sea  $(\Omega, \mathbb{F}, P)$  un espacio de probabilidad; una **red neuronal Bayesiana** para el conjunto  $U \subset \Omega$ , representa una unión de distribuciones de probabilidad en este espacio donde cada una cumple el criterio de independencia condicional, de manera más formal es una tupla de dos elementos  $(B_s, \Theta)$ , donde  $B_s$  es la **estructura de la red**, y  $\Theta \subset \Omega$  es el **conjunto de parámetros que codifica las distribuciones de probabilidad local**. La estructura  $B_s$  tiene dos componentes: La **estructura global**  $G$ , y la **estructura local**  $M$ . Donde  $G$  es un grafo dirigido acíclico que contiene un nodo para cada variable  $x_i \in U$ . Las aristas en  $G$  denotan las dependencias probabilísticas entre las variables de  $U$  y son las probabilidades de transición  $p_{ij}$  del estado  $i$  al estado  $j$  (probabilidades condicionales bayesianas entre par de estados) y usamos  $par(x_i)$  para denotar el conjunto de nodos pariente de  $x_i \in G$ , usamos  $x_i$  para denotar tanto elementos en  $U$  como los nodos del grafo  $G$ . El conjunto de estructuras locales  $M = \{M_1, \dots, M_n\}$  es un conjunto de  $n$  mapeos, uno para cada variable  $x_i$ , tal que  $M_i : \{par x_i, x_i\} \mapsto \Theta$ , ver, ref. [8].

**Arquitectura de (R.B.N).** Para este tipo de redes neuronales el conjunto de nodos de la estructura global forman tres conjuntos distintos, el primer conjunto de nodos son los nodos de entrada y es la conexión a la base de datos a analizar, es segundo grupo de nodos son la capa intermedia de la red y se les conoce como **unidades ocultas o funciones sinápticas** y es donde se realizan los procesos cognitivos del aprendizaje sobre el conjunto de parámetros de la red, estas funciones sinápticas sin entrar en mas detalle puesto que dependen del problema a resolver se clasifican en, funciones sigmoides, treshold, o funciones lineales, estos nodos se activan mediante alguna regla de

disparo para iniciar su aprendizaje, a veces es utilizado la técnica de la métrica de Hamming, ver ref.[6], también estos nodos solo pueden tener una única salida que se conecta al tercer grupo de nodos. El tercer grupo de nodos son los nodos de salida, los cuales a su vez es posible conectarlos a otra red neuronal formando redes neuronales multicapa. Las aristas del grafo  $G$  juegan un papel determinante para nosotros y son etiquetadas por funciones de ponderación y que para el caso de R.N.B. son las probabilidades de transición entre estados  $p_{ij}$ , ver fig. 5.1, ver [7], [28], [38].

A partir de aquí nos enfocaremos más a fondo en que consisten estas probabilidades de transición  $p_{ij}$  y su relación con el espacio de parámetros  $\Theta$ .

Ahora nos enfocamos en la estructura más formal de R.N.B, y su relación con la geometría algebraica, pero antes resumiremos la terminología básica de R.N.B.

**Teorema de Bayes 4.1.1.** Sea  $(\Omega, \mathbb{F}, P)$  un espacio de probabilidad y  $\{\theta_1, \dots, \theta_n\} \subset \Omega$  un conjunto de sucesos mutuamente excluyentes y exhaustivos, y tales que la probabilidad de cada uno de ellos es distinta de cero. Sea  $x_i$  un suceso cualquiera del que se conocen las probabilidades condicionales  $p(x_i|\theta_j)$ , entonces, la probabilidad  $P(\theta_j|x_i)$  esta dada por la siguiente expresión.

$$P(\theta_j|x_i) = \frac{p(x_i|\theta_j)p(\theta_j)}{\sum_{j=1}^n p(x_i|\theta_j)p(\theta_j)}.$$

Para ver detalles de la prueba, ver ref. [24], de la expresión anterior decimos que:

(a)  $p(x_i) =$  Probabilidad de la variable  $x_i \in U$ , en R.N.B. podrá ser calculada explícitamente por distribuciones de probabilidad Gaussianas, distribuciones binomiales etc.

(b)  $p(x_i|\theta_j) =$  Probabilidad condicional de  $x_i$  dado el parámetro  $\theta_j$  y coinciden con las probabilidades de transición así  $p_{ij} = p(x_i|\theta_j)$ .

(c)  $p(x_i, \theta_j) =$  Unión de la probabilidad de  $x_i$  y  $\theta_j$ ,  $p(x_i, \theta_j) = p(x_i)P(\theta_j|x_i) = p(\theta)P(x_i|\theta_j)$ , además,  $p(x_i) = \int p(x_i, \theta_j)d\theta_j$  (**marginalización**).

**Regla de Bayes aplicada a máquinas de aprendizaje.**

**Definición 4.1.2.** Sea  $(\Omega, \mathbb{F}, P)$  un espacio de probabilidad y  $D = \{x_n\} \subset U \subset \Omega$  y  $n = 1, 2, \dots, N$  el conjunto de nodos de entrada de una R.N.B., también este conjunto de datos puede ser visto como una sucesión de vectores aleatorios, por ello también una R.B.N. puede ser considerada como un proceso estocástico ahora elegimos  $\theta \in \Theta$  un parámetro. Entonces decimos que para;  $P(\theta|D) = \frac{P(D|\theta)p(\theta)}{p(D)}$ ,  $P(D|\theta)$  es la **probabilidad condicional** del conjunto de datos  $D$  dada la probabilidad del parámetro  $\theta$ ; y  $p(\theta)$  es la **probabilidad priori o previa** de  $\theta$ ; y  $P(\theta|D)$  es la **probabilidad posterior o futura** de  $\theta$  dado el conjunto  $D$ .

Si el conjunto  $m$  es otro modelo de otro conjunto de datos, tenemos que;  $P(m|D) = \frac{P(D|m)p(m)}{p(D)}$ ,

### Modelo de comparación.

Donde  $P(D|m) = \int P(D|\theta, m)P(\theta|m)$ , esta última probabilidad también se conoce como **probabilidad marginal** (evidencia): y es la probabilidad de que al tomar parámetros de la probabilidad previa seleccionados aleatoriamente generen al conjunto  $D$ .

### Predicción de la variable $x$ .

Dado el conjunto de datos  $D = \{x_n\} \subset U$  y  $n = 1, 2, \dots, N$ , la predicción de la variable  $x$  la realizamos mediante el siguiente cómputo  $P(x|D, m) = \int P(x|\theta, D, m)P(\theta|D, m)d\theta$ , y  $P(x|D, m) = \int P(x|\theta)P(\theta|D, m)d\theta$ , (para muchos modelos).

### Probabilidades previas hereditarias.

Con las definiciones anteriores podemos calcular probabilidades previas del parámetro  $\theta$ , de modelos que dependen de muchos parámetros, llamados hiperparámetros mediante las siguientes expresiones:

$$\begin{aligned} p(\theta) &= \int p(\theta|\alpha)p(\alpha)d\alpha \\ &= \int p(\theta|\alpha)d\alpha \int p(\alpha|\beta)p(\beta) \\ &= \int p(\theta|\alpha)d\alpha \int p(\alpha|\beta) \int p(\beta|\gamma)p(\gamma)d\gamma. \end{aligned}$$

De manera más específica y de acuerdo a la primera definición. En resumen una **R.N.B.** y de acuerdo a su definición 4.1.1 esta caracterizada de los siguientes mapeos y conjuntos, explícitamente.

(a)**Datos:**  $D = \{x^{(i)}, y^{(i)}\}_{i=1, \dots, N} = (X, Y) \subset \mathbb{R}^N \times \mathbb{R}^N$ .

(b)**Parámetros (ponderaciones o pesos):**  $\Theta = \{\{p_{ij}\}, \{v_k\}\}$ .

(c)**Probabilidad previa:**  $P(\Theta|\alpha)$ .

(d)**Probabilidad posterior o futura:**  $P(\Theta|\alpha, D) \propto P(Y|X, \Theta)P(\Theta|\alpha)$ .

(e)**Evidencia o probabilidad marginal:**  $P(Y|X, \alpha) = \int P(Y|X, \Theta)P(\Theta|D, \alpha)d\Theta$ .

(f)**Predicción:**  $P(y'|D, x', \alpha) = \int P(y'|x', \Theta)P(\Theta|D, \alpha)d\Theta$ .

Cabe mencionar que el cálculo de estas últimas integrales son difíciles de realizar y computacionalmente caras. Pero la inferencia Bayesiana tiene muchas ventajas; como una aparato matemático formal para el entendimiento de la adquisición artificial del conocimiento y su implementación en máquinas o servomecanismos. Además de cumplirse a través de esta teoría axiomas básicos comparados a la inteligencia humana como son:

Siendo que si la probabilidad de un evento  $p(X)$  en la probabilidad de Kolmogorov se encuentra definida en  $0 \leq p(x) \leq 1$

Si  $p(X) = 0$  decimos que el evento  $X$  es definitivamente falso.

Si  $p(X) = 1$  decimos que el evento  $X$  es definitivamente verdadero.

Si  $p(X|Y)$  es el rango de creencia de que el evento  $X$  suceda siempre que el evento  $Y$  sea verdadero.

**Definición 4.1.3.** Decimos que una R.N.B. es un **modelo paramétrico**, dado un número fijo de parámetros  $\Theta$ , y registrado el tamaño de el conjunto de datos. Dado  $\Theta$ , las predicciones son independientes de los datos  $D$ ,  $P(x, \Theta|D) = P(x|\Theta)P(\Theta|D)$ . Los parámetros son un resumen finito de los datos. Podemos también llamar a esto **modelo basado en aprendizaje** (ejemplo de ello, mezcla de  $k$  Gaussianas). Si permitimos que el número de parámetros crezca con el tamaño del conjunto, o alternativamente las predicciones son dependientes de los datos, y los parámetros  $\alpha$  son usualmente un conjunto pequeño. Y  $P(x|D, \alpha)$  es llamado **aprendizaje basado en memoria** (ejemplo de ello la estimación de la densidad de kernel).

**De un caso paramétrico a uno no paramétrico.** Consideremos el conjunto de datos  $D = \{x^{(1)}, \dots, x^{(n)}\}$ , y sea  $x$  la predicción de un nuevo punto, y  $\Theta$  parámetros, e hiperparámetros  $\alpha$ .

Integrando sobre el espacio de parámetros  $\Theta$  de un modelo paramétrico obtenemos un modelo no-paramétrico.

$$P(x, D, \Theta | \alpha) = P(x | \Theta) P(D | \Theta) P(\theta | \alpha) \text{ (paramétrico)}.$$

$$P(x | D, \alpha) \propto \int P(x | \Theta) P(D | \Theta) P(\theta | \alpha) d\Theta \text{ (no paramétrico)}.$$

**Definición 4.1.4.** Sea  $D = \{(x^{(i)}, y^{(i)}) \in \Omega\}_{i=1, \dots, N}$  un conjunto de datos de una R.N.B. y son independientemente tomados con probabilidad  $q(y|x)q(x)$ . Esta probabilidad esta bien definida y se ajusta a los datos del conjunto  $D$ , de ese modo decimos que la probabilidad  $q(x)$  es la **probabilidad verdadera** de los datos de entrada  $x$ , y  $q(y|x)$  es la **probabilidad de inferencia verdadera** de la red neuronal.

**Definición 4.1.5.** Sea  $(\Omega, \mathbb{F}, P)$  un espacio de probabilidad y elegimos  $\omega \in \Theta \subset \Omega$ , un parámetro que representa una probabilidad de inferencia de la red neuronal y sea  $\varphi(\omega)$  su densidad de probabilidad a priori o previa, definimos la densidad de probabilidad futura del parámetro  $\omega$  como

$$\rho_n = \frac{1}{Z_n} \varphi(\omega) \prod_{i=1, \dots, n} p_{ii}(\omega) \text{ con } p_{ii}(\omega) = P(x_i | y_i, \omega).$$

Notemos la evaluación como ponderación en toda la red. Donde  $Z_n$  es la constante de normalización que asegura que  $\int \rho_n(\omega) d\omega = 1$

**Definición 4.1.6.** De la misma manera definimos la **inferencia estadística** de la R.N.B como el promedio del ensamble de probabilidades futuras del parámetro  $\omega$  y se escribe

$$P_n(y|x) = \int P(y|x, \omega) \rho_n(\omega) d\omega.$$

Donde  $P(y|x, \omega)$  es la probabilidad condicional de una salida o predicción de un vector  $y \in \mathbb{R}^N$  dado un vector de entrada  $x \in \mathbb{R}^M$ .

**Definición 4.1.7 (Máquinas identificables y no identificables).** Sea  $(\Omega, \mathbb{F}, P)$  un espacio de probabilidad y  $P(y|x, \omega)$  la inferencia probabilística o probabilidad de predicción de una red neuronal del vector de salida  $y \in \mathbb{R}^N$  y  $x \in \mathbb{R}^M$  un vector de entrada. Dado el mapeo  $\omega \mapsto P(y|x, \omega)$  y este es inyectivo decimos que la R.N.B es una **máquina identificable** en otras palabras, un modelo es identificable sí y solo sí su parámetro está determinado por su conducta. Si el mapeo no es uno a uno entonces decimos que la R.N.B es una **máquina no identificable** y en lo sucesivo estaremos interesados en este tipo de modelos.

Hay muchos modelos de redes neuronales que entran en este tipo de modelos y lo cual el concepto de no identificable complica su implementación de manera que es difícil su entrenamiento y obtener buenos resultados que aseguren el correcto aprendizaje de este tipo de máquinas y están asociadas íntimamente a máquinas singulares, ejemplos de ellas son: perceptrón de tres capas, funciones de bases radiales, máquinas de Boltzman, mezclas Gaussianas, entre otras de interés analizaremos dos de este tipo en el próximo capítulo. Esta situación de modelos no identificables es la causa del problema de aprendizaje en R.N.B, como veremos y son la principal causa de que entren en juego ciertos conjuntos algebraicos, como se ve a continuación.

### Geometría algebraica de R.N.B.

En lo subsiguiente siempre consideraremos que las densidades de probabilidad de las máquinas de aprendizaje: **inferencia estadística**  $P(y|x, \omega)$  o predicción del vector  $y \in \mathbb{R}^M$ ; **inferencia verdadera de la red neuronal**  $q(y|x)$ ; y  $q(y|x)q(x)$  es la **distribución de probabilidad con la que son tomados y entrenados los ejemplos** de la máquina de inferencia de manera independiente. Todas estas densidades estarán definidas a través del espacio de probabilidad  $(\Omega, \mathbb{F}, P)$ .

**Definición 4.1.8.** Sea  $\omega_0 \in \Theta \subset \Omega$  un parámetro tal que  $P(y|x, \omega_0) = q(y|x)$  es decir el parámetro que realiza que la inferencia verdadera de la red neuronal  $q(y|x)$  sea igual a la densidad de probabilidad predictiva del vector de salida  $y \in \mathbb{R}^N$ . Para máquinas no identificables este parámetro no es único y es un conjunto de parámetros, escribimos,

$$W_0 = \{\omega_0 \in \Theta \subset \Omega : P(y|x, \omega_0) = q(y|x)\},$$

decimos que este conjunto es el **espacio de parámetros verdaderos** y este conjunto es una subvariedad no vacía que consiste de puntos singulares, cuando estas densidades de probabilidad son funciones analíticas decimos que este conjunto es un **conjunto analítico**, y si son polinomios decimos que es un **conjunto algebraico**. Estos conjuntos son los conjuntos en los que estamos interesados, aunque usaremos una definición alternativa para este conjunto que es de interés práctico para el estudio de nuestras técnicas de resolución de singularidades en máquinas no identificables.

**Definición 4.1.9.** Definimos la **distancia Kullback** de una R.N.B, también conocida en la literatura como **entropía de la información** y es una forma de cuantificar cuánto dista la probabilidad predictiva  $P(y|x, \omega)$  de la variable de salida  $y \in \mathbb{R}^N$  a la inferencia estadística verdadera de la red neuronal  $q(y|x)$ , escribimos

$$\text{(distancia Kullback)} \quad H(\omega) = \int \log \frac{q(y|x)}{P(y|x, \omega)} q(y|x) q(x) dx dy.$$

donde  $q(x)$  es la probabilidad verdadera de la variable de entrada  $x$ .

**Definición 4.1.10.** La distancia Kullback induce otra importante definición y la de mayor relevancia en este capítulo. Definimos la **curva de aprendizaje** de una R.N.B. o **generalización del error** conocida así en la literatura también, como

$$K(n) = E_n \left\{ \int \log \frac{q(y|x)}{P_n(y|x, \omega)} q(y|x) q(x) dx dy \right\}.$$

Donde  $E_n\{\bullet\}$  denota el valor esperado de todos los conjuntos de pares de ejemplos entrenados en la red neuronal. Y  $P_n(y|x, \omega)$  es la densidad de probabilidad promedio de todas las probabilidades futuras o posteriores de las salidas de la red neuronal, i.e., la expresión del integrando de arriba, es la distancia Kullback de la inferencia estadística verdadera de la red neuronal de la variable de salida  $y \in \mathbb{R}^N$ , a  $P_n$ .

**Definición 4.1.11.** Es importante notar que de esta última definición el **conjunto algebraico**  $W_0$  definido en la definición 4.1.8. es equivalente al siguiente conjunto,

$$W_0 = \{H(\omega) = 0 : \omega \in \Theta\}.$$

y finalmente este conjunto algebraico el cual es no vacío, será estudiando en los ejemplos del capítulo siguiente mediante resolución de singularidades.

## 4.2. Máquinas singulares.

**Definición 4.2.1.** Sea  $(\Omega, \mathbb{F}, P)$  un espacio de probabilidad con  $y \in \mathbb{R}^N$  un vector aleatorio sobre este espacio, definimos la **matriz de información de Fisher**, de la siguiente manera

$$I(\omega) = E_n \{ \nabla_\omega \log P_n(y|x, \omega) \nabla_\omega^\top \log P_n(y|x, \omega) \} = \int_{\omega \in \mathbb{R}^M} \{ \nabla_\omega \log P_n(y|x, \omega) \nabla_\omega^\top \log P_n(y|x, \omega) \} P_n(y|x, \omega) q(x) dx dy,$$

donde  $P_n$  es como se definió en 4.1.6, en general esta expresión puede ser entendida como una métrica en el espacio de parámetros siempre que esta matriz sea definida positiva, de ahí que en los inicios de teoría estadística del aprendizaje mediante este concepto se desarrollo la **geometría de la información**, como lo es ahora para redes neuronales la geometría algebraica.

**Definición 4.2.2.** Una máquina de aprendizaje es llamada **máquina de aprendizaje regular** si la matriz de información de Fisher es definida positiva, de otra manera diremos que es una **máquina de aprendizaje singular** i.e., existe un parámetro  $\omega \in \Theta$  tal que  $\det I(\omega) = 0$  tal parámetro es llamado singularidad de la matriz de Fisher. Cabe mencionar que la singularidad de la matriz de Fisher la probabilidad del parámetro  $\omega$  no puede ser aproximado por cualquier forma cuadrática, como en el caso de máquinas estadísticas regulares, ver ref. [39].

**Ejemplo (b).** Consideremos la siguiente función sináptica de una R.N.B, que emplea el modelo de regresión de series de Fourier la cual es regular (matriz de Fisher definida positiva). Sin embargo, si una función con un parámetro  $\omega = \{a_k, b_k\}$ , es escrita como

$$f(x, \omega) = \sum_{k=1, \dots, N} a_k \tanh(b_k x), \quad (0 \leq x \leq 1)$$

es usada en un modelo de regresión, entonces el conjunto

$$\left\{ \frac{\partial}{\partial a_k} f(x, \omega), \frac{\partial}{\partial b_k} f(x, \omega) \right\}$$

no es linealmente independiente si al menos uno de  $a_k$  o  $b_k$  es igual a cero. De esa manera este modelo de regresión es singular. Se muestra que (ver ref. [39]), que el conjunto de todas las singularidades de la matriz de información de Fisher de esta máquina

$$S = \{\omega \in \Theta : \det I(\omega) = 0\}$$

es una variedad algebraica que es determinada por el ideal generado por

$$\langle \alpha \rangle = \langle \sum_{k=1, \dots, N} a_k b_k^{2h-1} = 0 \rangle, \quad (h = 1, \dots, N).$$

Esta variedad algebraica contiene singularidades.

**Ejemplo (c).** La distribución normal

$$P(x|b) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\|x - b\|^2}{2}\right),$$

con  $a \in \mathbb{R}$ ,  $x, b, c \in \mathbb{R}^N$ , la cual es máquina estadística regular, mientras la mezcla de dos Gaussianas

$$P(x|a, b, c) = \frac{1}{\sqrt{2\pi}} \left\{ a \exp\left(-\frac{\|x - b\|^2}{2}\right) + (1 - a) \exp\left(-\frac{\|x - c\|^2}{2}\right) \right\},$$

es singular. En general, si una máquina de aprendizaje tiene estructura de capas o variables ocultas como es este último caso, entonces es singular, en general. Casi todas las máquinas de aprendizaje de procesos de información neuronal (R.N.B) son singulares. También consultar ref., [29], [39], [40],

[41].

**La distancia Kullback y teorema de Hironaka-Atiyah.** A continuación definimos la distancia Kullback empírica promedio como

$$H_n = \frac{1}{n} \sum_{i=1}^n \log \frac{q(y_i|x_i)}{P(y_i|x_i, \omega)},$$

y sea  $H(\omega)$  la distancia Kullback como la definimos anteriormente, si existe un parámetro  $\omega_0$  tal que  $H(\omega_0) = 0$ , entonces  $H(\omega)$  satisface las afirmaciones del teorema de resolución de Hironaka-Atiyah, i.e., ahí existen una variedad  $U$  y un mapeo de resolución  $g : U \mapsto W$ , tal que

$$H(g(u)) = A(u)^2 \text{ con } A(u) = u_1^{k_1} * \dots * u_d^{k_d}.$$

usando el hecho de que la distancia empírica puede ser escrita como, para mayor detalle ver ref. [30], [32], [33].

$$H_n(g(u)) = nA(u)^2 + \sqrt{n}A(u)\xi_n(u), \dots (*)$$

donde

$$\xi_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1, \dots, n} \psi(x_i, y_i, u).$$

de ahí la función sináptica  $\psi(x, y, u)$  es definida del siguiente modo por

$$\psi(x, y, u) = \frac{1}{A(u)} \left( H(g(u)) - \log \frac{q(y|x)}{P(y|x, g(u))} \right).$$

Donde la ecuación (\*) es la representación estándar de función de probabilidad de una máquina de aprendizaje singular.

**Observación.** La función  $\psi(x, y, u)$  puede ser escrita como  $\psi(x, y, g^{-1}(u))$  si  $H(\omega) \neq 0$ . Sin embargo sigue bien definida ahora si  $H(\omega) = 0$  en general. De otra manera probamos, que  $\psi(x, y, u)$  es una función analítica de  $u$  aún cuando  $H(g(u)) = 0$ . El hecho de que  $\psi(x, y, u)$  este bien definida en  $u$  puede ser probado de la propiedad de cruces normales de  $A(u)$ , ver ref., [16], [20], [21], [39].

#### 4.2.1. Efecto de las singularidades en R.N.B

Veremos cual es el efecto de las singularidades en R.N.B mediante un ejemplo y posteriormente en la siguiente sección, generalizamos el modo de corregir este efecto con un teorema probado por S.Watanabe en R.N.B.

Consideremos la siguiente R.N.B caracterizada por los siguientes modelos de aprendizaje.

Probabilidad predictiva de la máquina definida por:

$$P(y|x, a, b) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - af(b, x))^2\right).$$

Inferencia estadística verdadera de la máquina:

$$q(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2}\left(y - \frac{a_0 f(b_0, x)}{\sqrt{n}}\right)^2\right).$$

Donde  $y \in \mathbb{R}$ , es una variable de salida,  $x \in \mathbb{R}^n$  es una variable de entrada con distribución de probabilidad verdadera  $q(x)$ . El espacio de parámetros esta definido por  $W_0 = \{(a, b) \in \mathbb{R} \times \mathbb{R}^N\}$ .

En este ejemplo asumiremos que  $f(b, x) = 0$  es equivalente a  $b = 0$ . Entonces

$$W^* = \{(a, b); P(y|x, a, b) = P(y|x, 0, 0), \forall x, y\} = \{a = 0\} \cup \{b = 0\}$$

Por tanto,  $W^*$  no es una variedad pero una unión de variedades y  $(0, 0)$  es una singularidad de  $W^*$ .

Enfatizamos que la distancia entre la probabilidad verdadera y singularidades es medida por  $\frac{1}{n}$ .

En estadística este escalamiento es frecuentemente usado para comparar el poder de diferentes hipótesis de prueba.

La distancia Kullback de la distribución  $P(y|x, 0, 0)$  a la distribución verdadera  $q(y|x)$  en proporción a  $\frac{1}{n}$ ,

$$\int q(x)P(y|x, 0, 0) \log \frac{P(y|x, 0, 0)}{q(y|x)} dx dy = \frac{a_0^2}{2n} \int f(b_0, x)^2 q(x) dx.$$

Ahora sea  $\varphi(a, b)$  la distribución previa del espacio de parámetros, es una función continuamente diferenciable de  $a$  y además  $\psi(b) = \varphi(0, b)$ , es una función de soporte compacto de  $b$  y  $\psi(0) > 0$ . Sea también  $D_n = \{(x_i, y_i); i = 1, \dots, n\}$  es un conjunto de ejemplos entrenados y que son tomados independientemente de  $q(x)q(y|x)$ . Ambos la distribución posterior o futura  $P(a, b|D_n)$ , y distribución predictiva  $P(y|x, D_n)$  son respectivamente definidas por

$$P(a, b|D_n) = \frac{1}{C} \varphi(a, b) \prod_{i=1, \dots, n} P(y_i|x_i, a, b),$$

$$P(y|x, D_n) = \int P(y|x, a, b) P(a, b|D_n) da db,$$

donde  $C$  es una constante de normalización. También afirmamos que esta R.N.B satisface la condición

$$f(b, x) = \sum_{j=1}^J f_j(b) e_j(x)$$

donde  $\{e_j(x); j = 1, 2, \dots, J\}$  es un conjunto de funciones ortonormales. Entonces se sigue

$$\|f(b)\|^2 = \sum_{i=1}^J f_j(b)^2 = \int f(b, x)^2 q(x) dx.$$

Tenemos el siguiente teorema .

**Teorema 4.2.1.1 (Coeficiente de aprendizaje).** Con la notación de arriba, asumimos que

$$\int \frac{\psi(b) db}{\|f(b)\|} < \infty.$$

Entonces, la curva de aprendizaje de esta máquina o generalización del error de Bayes puede ser expandido asintóticamente como

$$G(n) = \frac{\lambda(a_0, b_0)}{n} + o\left(\frac{1}{n}\right).$$

y definimos el **coeficiente de aprendizaje** que es independiente de  $n$  y esta dado por

$$\lambda(a_0, b_0) = \frac{1}{2} \left\{ 1 + a_0^2 \|f(b_0)\|^2 - \sum_{j=1}^J a_0 f_j(b_0) E_g \left[ \frac{\partial}{\partial g_j} \log Z(g) \right] \right\},$$

donde  $g = \{g_j\}$  es una variable aleatoria que esta sujeta a la distribución Gaussiana  $J$  dimensional cuyo promedio y matriz de covarianza son cero y la identidad respectivamente. Y  $E_g$  denota los valores esperados sobre  $g$ , y

$$Z(g) = \int \exp(L(g)) \frac{\psi(b) db}{\|f(b)\|},$$

$$L(b) = \frac{((g + a_0 f(b_0)) * f(b))^2}{2 \|f(b)\|^2}.$$

**Demostración.** Usando de la definición de distancia Kullback hacemos

$$r(y|x, a, b) = \log \frac{q(y|x)}{P(y|x, a, b)},$$

en la curva de aprendizaje nos queda

$$G(n) = E \left[ -\log \frac{\int \exp \left( -\sum_{i=1}^{n+1} r(y_i|x_i, a, b) \right) \varphi(a, b) da db}{\int \exp \left( -\sum_{i=1}^n r(y_i|x_i, a, b) \right) \varphi(a, b) da db} \right].$$

Introducimos el parámetro de reescalamiento  $\alpha = \sqrt{n}a$  y definimos el promedio  $\langle S(\alpha, b) \rangle$  dado por

$$\langle S(\alpha, b) \rangle = \frac{\int e^{-E_1(\alpha, b)} S(\alpha, b) \varphi\left(\frac{\alpha}{\sqrt{n}}, b\right) d\alpha db}{\int e^{-E_1(\alpha, b)} \varphi\left(\frac{\alpha}{\sqrt{n}}, b\right) d\alpha db}$$

donde

$$E_1(\alpha, b) = \frac{1}{2n} \sum_{i=1}^n d(\alpha, b, x_i)^2 - \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i d(\alpha, b, x_i),$$

$$d(\alpha, b, x) = \alpha f(b, x) - a_0 f(b_0, x)$$

Aquí  $\epsilon_i = y_i - a_0 f(b_0, x_i)/\sqrt{n}$  los cuales son ejemplos independientes tomados de la distribución normal estándar. Entonces la curva de aprendizaje  $G(n)$  esta dada por

$$G(n) = E \left[ -\log \left\langle \exp \left( -\frac{e(\alpha, b)}{\sqrt{n}} \right) \right\rangle \right].$$

Donde

$$e(\alpha, b) = \frac{1}{\sqrt{2n}} d(\alpha, b, x_{n+1})^2 - \epsilon_{n+1} d(\alpha, b, x_{n+1}).$$

Se sigue que

$$G(n) = E \left[ -\log \left( 1 - \frac{\langle e(\alpha, b) \rangle}{\sqrt{n}} + \frac{\langle e(\alpha, b)^2 \rangle}{2n} \right) \right] + o \left( \frac{1}{n^{3/2}} \right),$$

entonces por la ecuación

$$\frac{1}{\sqrt{n}} E[\langle e(a, b) \rangle] = \frac{1}{2n} E[\langle e(a, b)^2 \rangle] + o \left( \frac{1}{n^{3/2}} \right),$$

obtenemos

$$G(n) = \frac{1}{2n} E[\langle e(a, b) \rangle^2] + o \left( \frac{1}{n^{3/2}} \right).$$

En otras palabras, el coeficiente de aprendizaje esta dado por

$$2\lambda = \lim_{n \rightarrow \infty} E[\langle e(a, b) \rangle^2],$$

si la convergencia de arriba se sostiene. Probemos su convergencia y valor límite. Usando el hecho de que  $f(b, x) = \sum_{j=1}^J f_j(b) e_j(x)$ , y el teorema del límite central la siguiente convergencia en ley se sostiene,

$$E_1(\alpha, b) \rightarrow \frac{1}{2} \sum_{j=1}^J (\alpha f_j(b) - a_0 f_j(b_0))^2 - \sum_{j=1}^J g_j(\alpha f_j(b)) - a_0 f_j(b_0) = E_2(\alpha, b) + c_0$$

donde  $\{g_j\}$  es la variable aleatoria Gaussiana J-dimensional cuyo promedio y matriz de covarianza respectivamente son cero y la identidad,  $E_2(\alpha, b)$  es una función definida por

$$E_2(\alpha, b) = \frac{1}{2} \sum_{j=1}^J \alpha^2 f_j(b)^2 - \sum_{j=1}^J \alpha f_j(b) (g_j + a_0 f_j(b_0)),$$

y  $c_0$  es una variable aleatoria pero una constante como función de  $(\alpha, b)$ . También definimos el promedio  $\langle \bullet \rangle_\infty$  de  $S(\alpha, b)$  por

$$\langle S(\alpha, b) \rangle_\infty = \frac{\int e^{-E_2(\alpha, b)} S(\alpha, b) \psi(b) d\alpha db}{\int e^{-E_2(\alpha, b)} \psi(b) d\alpha db}.$$

Entonces se sigue que,

$$\begin{aligned} 2\lambda(a_0, b_0) &= \lim_{n \rightarrow \infty} E[\langle e(\alpha, b) \rangle^2] = \lim_{n \rightarrow \infty} E[\langle d(\alpha, b, x_{n+1}) \rangle^2] \\ &= \lim_{n \rightarrow \infty} E \left[ \sum_{j=1}^J \langle j(b) - a_0 f_j(b_0) \rangle^2 \right] = \sum_{j=1}^J E_g \left[ \{ \langle \alpha f_j(b) \rangle_\infty - a_0 f_j(b_0) \}^2 \right], \end{aligned}$$

donde  $E_g$  muestra la esperanza sobre  $g$ . De ese modo

$$\begin{aligned} 2\lambda(a_0, b_0) &= \sum_{j=1}^J E_g \left[ \left\{ \frac{1}{Z} \frac{\partial Z}{\partial g_j} - a_0 f_j(b_0) \right\}^2 \right] \\ &= \sum_{j=1}^J E_g \left[ \left( \frac{1}{Z} \frac{\partial Z}{\partial g_j} \right)^2 \right] - \sum_{j=1}^J 2a_0 f_j(b_0) E_g \left[ \frac{1}{Z} \frac{\partial Z}{\partial g_j} \right] + \sum_{j=1}^J a_0^2 f_j(b_0)^2, \end{aligned}$$

donde,

$$Z(g) = \int e^{-E_2(\alpha, b)} \psi(b) d\alpha db = \int \exp \left( \frac{\{ \sum_{j=1}^J f_j(b) (g_j + a_0 f_j(b_0)) \}^2}{2 \|f(b)\|^2} \right) \frac{\psi(b) db}{\|f(b)\|}.$$

Y usando las identidades.

$$\left\{ \frac{1}{Z} \frac{\partial Z}{\partial g_j} \right\}^2 = \frac{1}{Z} \frac{\partial^2 Z}{\partial g_j^2} - \frac{\partial}{\partial g_j} \left\{ \frac{1}{Z} \frac{\partial Z}{\partial g_j} \right\}, \quad \sum_{j=1}^J \frac{\partial^2 Z}{\partial g_j^2} = Z + \sum_{j=1}^J (g_j + a_0 f_j(b_0)) \frac{\partial Z}{\partial g_j},$$

donde esta última ecuación es demostrada por la integración parcial de  $\alpha$ . Finalmente usando la identidad

$$E_g \left[ \frac{\partial}{\partial g_j} \left\{ \frac{1}{Z} \frac{\partial Z}{\partial g_j} \right\} \right] = E_g \left[ g_j \frac{1}{Z} \frac{\partial Z}{\partial g_j} \right],$$

obtenemos la prueba, **q.e.d.** ref.[33].

Ahora consideramos el principio de nuestro ejemplo escribiendo la función sináptica  $f(b, x)$  en términos de su expansión en bases ortonormales  $e_j$ . Tenemos que

$$P(y|x, a, b) = \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-1}{2} (y - \sum_{j=1}^N ab_j e_j(x))^2 \right).$$

**Efecto de las singularidades en el espacio de parámetros.** En este caso para la función sináptica. Por ejemplo al estudiar el caso  $N \geq 2$ , para un modelo de regresión verdadero, ver ref. [36], este modelo tiene espacio de parámetros singular  $W_0 = \{(a, b), a = 0, b = 0, : a \in \mathbb{R}, b \in \mathbb{R}^N\}$ . Utilizando la técnica del blow up parametrizamos mediante  $\omega_i = ab_i$ , sustituyendo en el modelo de la máquina conseguimos

$$P(y|x, a, b) = \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-1}{2} (y - \sum_{j=1}^N \omega_j e_j(x))^2 \right).$$

Se puede mostrar que ahora esta máquina de aprendizaje es un modelo regular estadístico con coeficiente de aprendizaje  $\lambda(\omega_0) = N/2$  para un arbitrario parámetro  $\omega_0$ . Mientras que al no utilizar las técnicas de resolución de singularidades el coeficiente de aprendizaje estará dado por la fórmula del teorema anterior  $\lambda(a_0, b_0)$  y claramente  $\lambda(a_0, b_0) \neq \lambda(\omega_0)$ . Con ello dislumbramos con más claridad el efecto de las singularidades y su efecto en la curva de aprendizaje como se muestra en el teorema 4.2.1.1.

### 4.3. El teorema de Watanabe Sumio y curvas de aprendizaje.

En este capítulo nos concentramos en probar el teorema de Watanabe, y su implicación más fuerte en un corolario para el cálculo explícito de curvas de aprendizaje de redes neuronales Bayesianas, y la relación estrecha que tienen estos resultados con resolución de singularidades. El

objeto del desarrollo de la prueba también tiene como finalidad clarificar más a detalle la estructura matemática de la teoría de R.N.B.

**(1) Teorema de Watanabe 4.3.1.** Asumimos que  $f(\omega)$  es una función analítica y  $\varphi(\omega)$  es una función de densidad de probabilidad en  $\mathbb{R}^d$ . Entonces, existe una constante real  $C$  tal que para cualquier número natural  $n$

$$G(n) \leq \lambda_1 \log n - (m_1 - 1) \log(\log n) + C,$$

donde el número racional  $-\lambda_1$  ( $\lambda_1 > 0$ ) y el número natural  $m_1$  son los polos más grandes y su multiplicidad de la función meromórfica que es continuación analítica, y esta dada por,

$$J(\lambda) = \int_{f(\omega) < \epsilon} f(\omega)^\lambda \varphi'(\omega) d\omega, \quad (Re(\lambda) > 0).$$

Donde  $\epsilon > 0$  es una constante suficientemente pequeña, y  $\varphi'(\omega)$  es una función de clase  $C_0^\infty$  que satisface  $0 \leq \varphi'(\omega) \leq \varphi(\omega)$ .

**Demostración.**

**Lema 4.3.1.** Asumimos que  $f(\omega)$  y  $f_n(\omega)$  son funciones continuas y  $\varphi(\omega)$  una función de densidad de probabilidad. Entonces la desigualdad

$$G(n) \leq -\log \int \exp(-nf(\omega)) \varphi(\omega) d\omega$$

se sostiene para un número natural arbitrario.

**Demostración.** De la desigualdad de Jensen, ref. [33], [32], tenemos que

$$-\log \int \exp(a(\omega)) b(\omega) d\omega \leq -\int a(\omega) b(\omega) d\omega$$

se sostiene para una función continua arbitraria  $a(\omega)$  y una densidad de probabilidad arbitraria de soporte compacto  $b(\omega)$ . Primero, asumimos que  $\varphi(\omega)$  es una función de soporte compacto. Aplicando la desigualdad de arriba al caso especial,

$$a(\omega) = -n(f_n(\omega) - f(\omega)), \quad b(\omega) = \frac{1}{Y} \exp(-nf(\omega)) \varphi(\omega),$$

donde  $Y = \int \exp(-nf(\omega)) \varphi(\omega) d\omega$ , obtenemos

$$-\log \frac{1}{Y} \int \exp(-nf_n(\omega))\varphi(\omega)d\omega \leq \frac{1}{Y} \int n\{f_n(\omega) - f(\omega)\} \exp(-nf(\omega))\varphi(\omega)d\omega$$

para una función arbitraria de soporte compacto  $\varphi(\omega)$ . Siendo que el conjunto de todas las funciones de soporte compacto es denso en el conjunto de todas funciones de densidad de probabilidad por la norma  $L^1$ , se sigue la prueba del lema **q.e.d.**

Para un  $\epsilon > 0$ , un conjunto de parámetros  $W_\epsilon$  es definido por

$$W_\epsilon = \{\omega \in W = \text{supp}(\varphi); f(\omega) < \epsilon\}.$$

Introducimos a continuación la función de Sato-Berstein.

**Teorema (Sato, Berstein, Bjork, Kashiwara) 4.3.2.** Asumimos que existe  $\epsilon_0 > 0$  tal que  $f(\omega)$  es una función analítica en  $W_{\epsilon_0}$ . Entonces existe una tupla de tres elementos  $(\epsilon, P, b)$  donde (1)  $\epsilon < \epsilon_0$  es una constante positiva,

(2)  $P = P(\lambda, \omega, \partial\omega)$  es un operador diferencial el cual es un polinomio para  $\lambda$ , y

(3)  $b(\lambda)$  es un polinomio tal que

$$P(\lambda, \omega, \partial\omega)f(\omega)^{\lambda+1} = b(\lambda)f(\omega)^\lambda (\forall \omega \in W_\epsilon, \forall \lambda \in C).$$

Los ceros de la ecuación algebraica  $b(\lambda) = 0$  son reales, racionales y números negativos.

En el teorema  $P = P(\lambda, \omega, \partial\omega)$  es un operador diferenciable para  $\omega$  cuyos coeficientes son funciones analíticas para  $\omega$ .  $P$  puede ser entendido como un mapeo  $C^\infty$  a  $C^\infty$ . Los operadores adjuntos son formalmente definidos de la misma manera para operadores parciales diferenciales usuales,

$$\int \phi(\omega)P(\lambda, \omega, \partial\omega) = \int P(\lambda, \omega, \partial\omega)^*\phi(\omega)\varphi(\omega)$$

para cualquier  $\phi \in C^\infty$ ,  $\varphi \in C^\infty$ . La racionalidad de los ceros de  $b(\lambda) = 0$  es probada por medio de resolución de singularidades (Atiyah,1970; Kashiwara, 1976). El más pequeño orden del polinomio  $b(\lambda)$  que satisface las relaciones de arriba es llamado polinomio de Sato-Berstein o  $b$ -función. Ver ref. [33], para la prueba.

Ahora sea  $\epsilon > 0$  suficientemente pequeño, podemos asumir que

$$\|\nabla f(\omega)\| > 0, \forall \omega \in W \setminus W_0,$$

de el hecho de que  $f(\omega)$  es analítica.  $\epsilon > 0$  es tomada suficientemente pequeña tal que el teorema anterior se pueda sostener. También podemos asumir que  $\epsilon < 1$  sin pérdida de generalidad.

Para una función analítica  $f(\omega)$ , definimos una función compleja  $J(\lambda)$  de  $\lambda \in \mathbb{C}$  por

$$J(\lambda) = \int_{W_\epsilon} f(\omega)^\lambda \varphi(\omega) d\omega.$$

De la cual tenemos el siguiente lema.

**Lema 4.3.2.** Sea  $f(\omega)$  una función analítica en  $W_\epsilon$  y que  $\varphi(\omega)$  es una función de clase  $C_0^\infty$ . Entonces,  $J(\lambda)$  puede ser continuación analítica de una función meromórfica en todo el espacio complejo, en otras palabras,  $J(\lambda)$  tiene únicamente polos en  $|\lambda| < \infty$ . Sin embargo  $J(\lambda)$  satisface las siguientes condiciones.

(1) Los polos de  $J(\lambda)$  son racionales, reales, y números negativos.

(2) Para un arbitrario  $a \in \mathbb{R}$ ,  $J(\infty + a\sqrt{-1}) = 0$  y  $J(a \pm \infty * \sqrt{-1}) = 0$ .

**Demostración.** Sea  $J(\lambda)$  una función analítica en la región  $Re(\lambda) > 0$ . Primero,  $J(\infty + a\sqrt{-1}) = 0$  para  $a > 0$  fijo. Para  $t > 0$  definimos una función

$$\hat{J} = \exp(at) \int_{W_\epsilon} \delta(t - \log f(\omega)) \varphi(\omega) d\omega,$$

donde  $\hat{J}_a(t)$  es bien definida, porque  $\delta(g(\omega))$  es bien definida si  $\|\nabla g(\omega)\| > 0$  para cualquier  $\omega$  que satisface  $g(\omega) = 0$ , ref. [33]. De la definición, de  $\hat{J}_a(t)$  es una función  $L^1$ , y

$$J(a \pm b\sqrt{-1}) = \int_{-\infty}^{\log \epsilon} \hat{J}_a(t) \exp(\pm bt\sqrt{-1}) dt,$$

que muestra que podemos aplicar el teorema de convergencia de Riemann-Lebesgue, resultando que  $J(a \pm \infty\sqrt{-1}) = 0$  para  $a > 0$  fijo. Por último usando el operador adjunto  $P^*$ ,

$$J(\lambda) = \frac{1}{b(\lambda)} \int_{W_\epsilon} P f(\omega)^{\lambda+1} \varphi(\omega) d\omega = \frac{1}{b(\lambda)} \int_{W_\epsilon} f(\omega)^{\lambda+1} P^* \varphi(\omega) d\omega.$$

Donde usamos la propiedad de la **b-función**. Porque  $P^* \varphi \in C_0^\infty$ ,  $J(\lambda)$  puede ser una continuación analítica a  $J(\lambda - 1)$  si  $b(\lambda) \neq 0$ . Usando continuación analítica,  $J(a \pm \infty\sqrt{-1}) = 0$  aún para  $a < 0$ . Si  $b(\lambda) = 0$ , entonces  $\lambda$  es a lo más un polo que se encuentra en la parte negativa del

eje real. **q.e.d.**

**Definición 4.3.1.** Los polos de la función  $J(\lambda)$  están en la parte negativa del eje real y contenidos en el conjunto  $\{m + \nu; m = 0, -1, -2, \dots, b(\nu) = 0\}$ . Los cuales son ordenados del más grande al más pequeño y los escribimos como,  $-\lambda_1, -\lambda_2, -\lambda_3, \dots, \lambda_k$  donde  $\lambda_k$  es un número racional y la multiplicidad de  $-\lambda_k$  es denotada por  $m_k$ .

Ahora definimos la función  $I(t)$  valuada realmente por

$$I(t) = \int_{W_\epsilon} \delta(t - f(\omega)) \varphi(\omega) d\omega \text{ con, } 0 < t < \epsilon.$$

Como,  $\|\nabla f(\omega)\| > 0$  para  $\omega \in W_\epsilon \setminus W_0$ ,  $\delta(t - f(\omega))$  es bien definida. Para  $t \geq \epsilon$  o  $t < \epsilon$  entonces definimos  $I(t) = 0$ .

**Lema 4.3.3.** Asumimos que  $f(\omega)$  es una función analítica en  $W_\epsilon$  y que  $\varphi(\omega)$  es una función de clase  $C_0^\infty$ . Entonces  $I(t)$  tiene una expansión asintótica para  $t \rightarrow 0$

$$I(t) \cong \sum_{k=1}^{\infty} \sum_{m=0}^{m_k-1} c_{k,m+1} t^{\lambda_k-1} (-\log t)^m.$$

donde  $(m! * c_{k,m+1})$  es el coeficiente de los  $(m+1)$ -ésimo orden en la expansión de Laurent de  $J(\lambda)$  en el polo  $\lambda = -\lambda_k$ .

**Demostración.** El caso especial de este lema es probado en la teoría de distribuciones. Ahora sea

$$I_k(t) \equiv I(t) - \sum_{k=1}^K \sum_{m=0}^{m_k-1} c_{k,m+1} t^{\lambda_k-1} (-\log t)^m.$$

es suficiente mostrar que, para un arbitrario  $K$  fijo,

$$\lim_{t \rightarrow 0} I_K(t) t^\lambda = 0, \quad \forall \lambda > -\lambda_{k+1} + 1.$$

De la definición de  $J(\lambda)$ ,  $J(\lambda) = \sum_0^1 I(t) t^\lambda dt$ . El simple cálculo muestra

$$\int_0^1 t^{\lambda+\lambda_k-1} (-\log t)^m dt = \frac{m!}{(\lambda + \lambda_k)^{m+1}}.$$

Por tanto,

$$\int_0^1 I_K(t) t^\lambda dt = J_K(\lambda),$$

donde  $J_K(\lambda)$  es definida por

$$J_K(\lambda) = J(\lambda) - \sum_{k=1}^K \sum_{m=0}^{m_k-1} \frac{m! * c_{k,m+1}}{(\lambda + \lambda_k)^{m+1}},$$

que es holomórfica en la región  $Re(\lambda) > -\lambda_{K+1}$ . Y haciendo  $t = e^{-x}$  y usando la transformada inversa de Laplace, se sigue que,

$$I_K(e^{-x})e^{-x} = \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} J_K(u) e^{ux} du,$$

se sostiene para cualquier real  $\tau > 0$ . Por el lema anterior,  $J(a \pm i\infty) = 0$ , así  $J_K(a \pm i\infty) = 0$  para un real  $a$  arbitrario. Y siendo que  $J_K(\lambda)$  es holomórfico en la región  $Re(\lambda) > -\lambda_{K+1}$ , la integral de trayectoria compleja anterior puede ser de  $\tau > 0$  a  $\tau > -\lambda_{K+1}$ . Por tanto,

$$I_K(t)(e^{-x})e^{-x-\tau x} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} J_K(\tau + iu) e^{iux} du,$$

se sostiene para cualquier real  $x$ . De ese modo, haciendo  $x = 0$ ,  $J_K(\tau + iu) \in L^1$ . Los términos del lado derecho de esta última ecuación tienden a cero cuando  $x \rightarrow \infty$  se sigue del teorema de convergencia de Riemann-Lebesgue. Y haciendo  $t = -\log x$ , obtenemos el resultado al que queríamos llegar, **q.e.d.** Finalmente podemos probar el teorema de Watanabe, juntando todos estos argumentos como sigue.

$$\begin{aligned} G(n) &\leq -\log \left\{ \int_W \exp(-nf(\omega)) \varphi(\omega) d\omega \right\} \leq \\ &-\log \left\{ \int_{W_\epsilon} \exp(-nf(\omega)) \varphi'(\omega) d\omega \right\} = -\log \int_0^1 e^{-nt} I(t) dt = -\log \int_0^n e^{-t} I\left(\frac{t}{n}\right) \frac{dt}{n} \approx \\ &-\log \left\{ \sum_{k=1}^{\infty} \sum_{m=0}^{m_k-1} \sum_{j=0}^m \frac{c_{k,m+1} (\log n)^j (m * C_j)}{n^{\lambda_k}} \int_0^n e^{-t} t^{\lambda_k-1} (-\log t)^{m-j} dt \right\} = \\ &\lambda_1 \log(n) - (m_1 - 1) \log(\log n) + o(1), \end{aligned}$$

donde  $I(t)$  es definida por definición 4.3.1. con  $\varphi'(\omega)$  en vez de  $\varphi(\omega)$ , y con esto concluye la prueba, **fin de la prueba del teorema 4.3.1. q.e.d.**

Consecuencias importantes de este resultado se presentan a continuación en el siguiente teorema.

**(2) Teorema de Watanabe 4.3.3.** Sea  $\sigma > 0$  un valor constante. Asumimos que  $\varphi(\omega)$  es una densidad de probabilidad de clase  $C_0^\infty$ . Consideremos el caso de todas las R.N.B caracterizadas por la siguiente máquina de aprendizaje con inferencia de máquina dada por el modelo

$$P(y|x, \omega) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{-\|y - \psi(x, \omega)\|^2}{2\sigma^2}\right),$$

donde ambos  $\psi(x, \omega)$  y  $\|\psi(x, \omega)\|^2$  satisface la condición (A) (que se explica abajo). Entonces, existe una constante  $C' > 0$  tal que para cualquier número natural  $n$

$$|G(n) - \lambda_1 \log n + (m_1 - 1) \log \log n| \leq C',$$

donde el número racional  $-\lambda_1$  ( $\lambda_1 > 0$ ) y un número natural  $m_1$  son los polos más grandes y su multiplicidad de la función meromórfica que es continuación analítica de

$$J(\lambda) = \int_{f(\omega) < \epsilon} f(\omega)^\lambda \varphi(\omega) d\omega, (Re(\lambda) > 0),$$

donde  $\epsilon > 0$  es un constante suficientemente pequeña.

Condición (A) del teorema 4.3.3. Sea  $\psi(x, \omega)$  una función vectorial valuada realmente de  $(x, \omega) \in \mathbb{R}^M \times \mathbb{R}^d$ . Definimos las condiciones siguientes para esta función.

(1)  $\psi(x, \omega)$  es una función analítica de  $\omega \in W = \text{supp}(\varphi) \subset \mathbb{R}^d$  que puede ser extendida como una función holomórfica sobre algún conjunto abierto  $W^*$ , donde  $W \subset W^* \subset \mathbb{C}^d$  y  $W^*$  es independiente de  $x \in \text{supp}(q) \subset \mathbb{R}^M$ .

(2)  $\psi(x, \omega)$  es una función medible de  $x \in \mathbb{R}^M$  que satisface la condición

$$\int \sup_{\omega \in W^*} \|\psi(x, \omega)\|^2 q(x) dx < \infty.$$

donde  $\|\bullet\|$  es la norma del vector  $\psi(x, \omega)$ .

**Bosquejo de la prueba.** Con los hechos dados por los lemas que se enuncian a continuación y el cálculo que se desarrolla hasta el final obtenemos la prueba del teorema para mas detalle de la prueba, ver ref.[33]. Lema 4.3.4. Sea  $\{x_i, \eta_i\}_{i=1}^n$  es un conjunto de ejemplos independientes tomados de  $q(x)q_0(\eta)$ , donde  $q(x)$  es una densidad de probabilidad y  $q_0(\eta)$  es la distribución normal estándar. Afirmamos que una función  $\xi(x, \omega)$  satisface la condición (A) y la expansión de Taylor de  $\xi(x, \omega)$  en  $\hat{\omega}$  absolutamente converge en una región  $T = \{\omega; |\omega_j - \hat{\omega}_j| < r_j\}$ . Para una constante dada  $0 < a < 1$ , definimos una región  $T_a \equiv \{\omega; |\omega_j - \hat{\omega}_j| < ar_j\}$ . Entonces, lo siguiente se sostiene.

(1) Si  $\int \xi(x, \omega)q(x)dx = 0$ , existe una constante  $c'$  tal que para un arbitrario  $n$ ,

$$A_n = E_n \left[ \sup_{\omega \in T_a} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi(x_i, \omega) \right|^2 \right] < c' < \infty.$$

(2) Donde existe una constante  $c''$  tal que para un arbitrario  $n$ ,

$$B_n = E_n \left[ \sup_{\omega \in T_a} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i \xi(x_i, \omega) \right|^2 \right] < c'' < \infty. \text{ ver ref. [29].}$$

**Teorema 4.3.4.** Definimos la fluctuación del aprendizaje como

$$\zeta_n(\omega) = \frac{\sqrt{n}(f(\omega) - f_n(\omega))}{\sqrt{f(\omega)}}.$$

esta puede ser discontinua en  $\omega_0 \in W_0$ , pero este teorema asegura que es acotada en el promedio. Seguimos asumiendo que se cumple la condición (A). Entonces, existe una constante  $c'''$  tal que para un arbitrario  $n$

$$E_n \left[ \sup_{\omega \in W \setminus W_0} |\zeta_n(\omega)|^2 \right] < c'''.$$

Ahora definamos  $\alpha_n = \sup_{\omega \in W \setminus W_0} |\zeta_n(\omega)|$  y por el teorema anterior, tenemos  $E_n[\alpha_n^2] < \infty$ . La generalización del error satisface

$$\begin{aligned} G(n) &= -E_n \left[ \log \int_W \exp(-nf(\omega)) \varphi(\omega) d\omega \right] \\ &= -E_n \left[ \log \int_W \exp(-nf(\omega)) - \sqrt{nf(\omega)} \zeta_n(\omega) \varphi(\omega) d\omega \right] \\ &\geq -E_n \left[ \log \int_W \exp(-nf(\omega)) + \alpha \sqrt{nf(\omega)} \varphi(\omega) d\omega \right] \\ &\geq -\log \left\{ \int_W \exp\left(\frac{-n}{2} f(\omega)\right) \varphi(\omega) d\omega \right\} - \frac{1}{2} E_n[\alpha_n^2], \end{aligned}$$

donde usamos una desigualdad dada por  $\alpha_n \sqrt{nf(\omega)} \leq (\alpha_n^2 + nf(\omega))/2$ . Y definimos para  $i = 1, 2$

$$Z_i(n) = \int_{W(i)} \exp\left(\frac{-n}{2} f(\omega)\right) \varphi(\omega) d\omega,$$

donde  $W(1) = W_\epsilon$  y  $W(2) = W \setminus W_\epsilon$ . Entonces por el mismo método como la prueba del teorema (1),

$$Z_1 \approx c_{1,m_1} \frac{(\log n)^{m_1-1}}{n^{\lambda_1}}.$$

De otra manera,

$$Z_2 \leq \int_{W \setminus W_\epsilon} \exp\left(-\frac{n\epsilon}{2}\right) \varphi(\omega) d\omega \leq \exp\left(-\frac{n\epsilon}{2}\right).$$

Por tanto

$$G(n) \geq -\log \left[ c_{1,m_1} \frac{(\log n)^{m_1-1}}{n^{\lambda_1}} + \exp\left(-\frac{n\epsilon}{2}\right) \right] + \text{const.}$$

Con esto completamos nuestra prueba , **q.e.d.**

De este hecho se desprende el resultado fundamental de este trabajo, y es el cálculo de la curva de aprendizaje y la relación con resolución de singularidades que ya se ha comentado en las pruebas anteriores.

En el teorema (2), la función  $c(n)$  definida por

$$c(n) = G(n) - \lambda_1 \log n + (m_1 - 1) \log(\log n),$$

es una función acotada,  $|c(n)| < C''$ . Si  $c(n+1) - c(n) = o\left(\frac{1}{n \log n}\right)$  entonces de aquí se sigue que la curva de aprendizaje esta dada por

$$K(n) = \frac{\lambda_1}{n} + \frac{m_1 - 1}{n \log n} + o\left(\frac{1}{n \log n}\right).$$

### Curvas de aprendizaje y resolución de singularidades.

También se sabe que, modelos estadísticos regulares tienen  $\lambda_1 = d/2$  y  $m_1 = 1$ , que puede ser mostrado como caso especial del teorema (2) Watanabe. Para modelos de máquinas no identificables tal como redes neuronales Bayesianas tienen diferentes valores  $\lambda_1 \leq d$  y  $m_1 \geq 1$ , en general. De la función meromórfica  $J(\lambda)$  definida en teorema (1) y (2)(Watanabe), es tarea de la geometría algebraica encontrar los polos  $\lambda_1$  y  $m_1$  aplicando técnicas de resolución de singularidades como

las sugeridas por S.Watanabe, modificación tórica y blow up, en el conjunto algebraico  $\{\lambda \in W : H(\lambda) = J(\lambda) = 0\}$ . el cual ya se había definido una sección anterior.

**Corolario 4.3.1.** Asumimos que se cumplen las hipótesis del teorema (2). Si  $c(n+1) - c(n) = o\left(\frac{1}{n \log n}\right)$ , entonces la **curva de aprendizaje** esta dada por

$$K(n) = \frac{\lambda_1}{n} + \frac{m_1 - 1}{n \log n} + o\left(\frac{1}{n \log n}\right).$$

De manera que para máquinas estadísticas regulares tienen  $\lambda_1 = d/2$  y  $m_1 = 1$ , que puede ser mostrado como caso especial del teorema (2) (Watanabe). Y para modelos no identificables tales como R.N.B, tienen valores  $\lambda_1 \leq d/2$  y  $m_1 \geq 1$ , en general.

**Corolario 4.3.2.** Afirmamos que las condiciones del teorema (1)(Watanabe) se cumplen. Si  $\varphi'(\omega)$  puede ser tomada como  $\varphi'(\omega) > 0$  para  $\omega_0 \in W_0$  arbitrario, entonces  $\lambda_1 \leq d/2$  donde  $d$  es la **dimensión del espacio de parámetros**. Ver ref. [32], [33].

**Corolario 4.3.3.** Sea una máquina de aprendizaje estadístico  $P(y|x, \omega)$  no singular (definición 4.2.2) definida en el espacio de probabilidad  $(\Omega, \mathbb{F}, P)$  con distancia Kullback asociada a su espacio de parámetros, i.e., el conjunto algebraico  $\{\lambda \in W \subset \Omega : H(\lambda) = J(\lambda) = 0\}$  y sea el siguiente polinomio una reparametrización tal que para un conjunto  $\Theta \subset \Omega$ ,  $\exists \omega \in \Theta \subset \Omega \subset \mathbb{C}^n$ ,  $\ni$  como hemos visto  $H(\omega) = \sum_{i=1}^n c_i \omega^a$  con  $c_i \in \mathbb{R}$ ,  $\omega^a \in \mathbb{C}^*$ ,  $\forall i$ , y  $a \in \mathbb{Z}^n$  vector reticular (ver definición 2.2.4, politopo de Newton). Si y solo sí el cono reticular  $\sigma \subseteq \text{supp}(H(\omega))$  generado por los vectores reticulares del soporte de  $H(\omega)$  es no singular; es decir,  $\det(\sigma) = 1$ .

**Demostración.**  $\Rightarrow$ ) La necesidad se sigue de los siguientes hechos; siendo que  $H(\omega)$  es una reparametrización del polinomio singular  $H(\lambda)$  entonces  $\exists$  un mapeo de resolución  $g : H(\lambda) \rightarrow H(\omega)$  por el teorema de Hironaka-Atiyah 3.1.3, tal que  $H(\omega)$  es no singular; ahora sean  $\sigma' \subset \text{supp}(H(\lambda))$  y  $\sigma \subset \text{supp}(H(\omega))$  conos reticulares generados por los vectores reticulares del soporte de  $H(\lambda)$  y  $H(\omega)$  respectivamente; afirmamos que  $g$  se levanta a un morfismo de  $\mathbb{C}$ -álgebras monomiales finitamente generadas  $R_{\sigma'} = \{f | \text{supp}(f) \subset \sigma'\}$  y  $R_{\sigma} = \{f | \text{supp}(f) \subset \sigma\}$  tal que  $H(\lambda) \in R_{\sigma'}$  y  $H(\omega) \in R_{\sigma}$ , la prueba de esto se sigue de su definición como mapeo de resolución. Entonces por el teorema 2.4.1, siendo que  $R_{\sigma'} \simeq R_{\sigma} \implies \sigma' \simeq \sigma$ . Entonces  $\exists$  una transformación unimo-

dular  $L \in \mathbb{Z}^n \times \mathbb{Z}^n$  la cual tiene una matriz asociada a las bases de Hilbert  $H_{(\sigma')^\vee}$  del cono dual  $(\sigma')^\vee \implies L(\sigma') = \sigma$  y de la unimodularidad de esta transformación tenemos  $\mathbf{det} L(\sigma') = \mathbf{det} \sigma = 1 \implies \sigma$  es no singular.

( $\Leftarrow$ . Ahora tomamos un cono reticular  $\sigma \subset \text{supp}(H(\omega)) \ni \mathbf{det} \sigma = 1 \implies \exists L \in \mathbb{Z}^n \times \mathbb{Z}^n \ni L(\sigma') = \sigma$  con  $L$  transformación unimodular, para algún cono  $\sigma' \subset \mathbb{Z}^n \implies \sigma' \simeq \sigma$ , entonces por el teorema 2.4.1, este isomorfismo se levanta a un morfismo tórico  $\Psi$  tal que  $R_{\sigma'} \simeq R_\sigma$  para hacer compatible con la notación anterior podemos hacer  $g = \Psi$  siendo que los morfismos de  $\mathbb{C}$ -álgebras monomiales f.g. son tóricos como ya se ha dicho, entonces  $g$  es un mapeo de resolución de  $H(\omega)$  y por tanto este polinomio es no singular, afirmamos que si esto sucede la matriz de información de Fisher  $I(\omega)$  asociada al modelo es no singular, i.e.,  $\mathbf{det} I(\omega) \neq 0$ .

**Prueba de la afirmación.** Si  $H(\omega)$  es singular la máquina de inferencia  $P(y|x, \omega)$  es un modelo no identificable por definición 4.1.8, de donde se sigue  $P(y|x, \omega) = q(y|x)$  con  $q(y|x)$  la inferencia estadística verdadera del modelo; de aquí si aplicamos el operador  $\nabla_\omega$  a  $P(y|x, \omega)$  de la definición de matriz de información de Fisher (definición 2.4.2); tenemos que  $\nabla_\omega P(y|x, \omega) = \nabla_\omega q(y|x) = 0 \implies \mathbf{det} I(\omega) = 0$  y con esto concluimos la prueba, **q.e.d.**

# Capítulo 5

## Modificación tórica y aplicaciones en R.N.B

En este capítulo veremos algunas aplicaciones de resolución de singularidades en R.N.B. para el cálculo de sus curvas de aprendizaje mediante los resultados del teorema de Watanabe y los corolarios que de éste se derivaron aplicados a modelos conocidos en la literatura de la inteligencia artificial, como mezclas binomiales y perceptrón de tres capas. Mediante estas curvas es posible implementar estos modelos en sistemas de cómputo inteligente. Los ejemplos y sus demostraciones de este capítulo, fueron tomados de los siguientes autores, [29], [30], [32], [33] **S. Watanabe.**, [43] **K. Yamazaki, M. Aoyagui, S. Watanabe**, [44] **K. Yamazaki**. Las máquinas de inferencia estadística tratadas en este capítulo, sus densidades de probabilidad estarán definidas en el espacio de probabilidad  $(\Omega, \mathbb{F}, P)$  con  $\omega \in \Theta \subset \Omega$  espacio de parámetros.

### 5.1. Modificación tórica y curva de aprendizaje en una mezcla binomial

**Aplicación (1). Mezcla de distribuciones binomiales.** Consideramos la siguiente R.N.B. caracterizada por la siguiente máquina de aprendizaje definida por las siguientes densidades de probabilidad.

Probabilidad verdadera de  $x$ ,  $q(x = k) = Bin_N(x; p^*) = \binom{N}{x} p^{*x} (1 - p^*)^{N-x}$ .

Inferencia probabilística verdadera del modelo  $P(x = k|w) = a Bin_N(x, p_1) + (1 - a) Bin_N(x, p_2)$ .

Donde el espacio de parámetros  $W$  esta definido por

$$\omega = (\{a_i\}_{i=1}^K, \{p_i\}_{i=1}^{K+1})$$

donde las coordenadas de los parámetros están definidas por  $0 < p_i < 1/2$  y,

$$a_{K+1} = 1 - \sum_{i=1}^K a_i.$$

Este tipo de R.N.B es utilizado para el análisis de génes y análisis espectral de mutaciones. Este consiste de  $K + 1$  componentes. Resumimos el cálculo de la curva de aprendizaje en el siguiente teorema.

**Teorema 5.1.1.** Sea una R.N.B. caracterizada por la máquina de aprendizaje definida por las probabilidades dadas arriba, entonces para un número suficiente grande  $n$  de ejemplos entrenados, se muestra de acuerdo al corolario 4.3.1, que la curva de aprendizaje esta dada por

$$K(n) = \frac{3}{4} \log(n) + C,$$

donde  $C$  es una constante independiente de  $n$ . El teorema se cumple para el siguiente modelo de dos mezclas binomiales. La distancia Kullback o entropía de la información esta dada por,

$$\begin{aligned} H(x,a,b_1, b_2) &= \sum_{x=0}^N q(x) \log \left( \frac{q(x)}{P(x|\omega)} \right) = (ap_1 + (1-a)p_2)^2 + (ap_1^2 + (1-a)p_2^2)^2 + \dots + \\ &= b_2^2 + (ab_1^2 + (b_2 - ab_1)^2)^2 + \dots + \text{términos de orden mayor.} \end{aligned}$$

la cual es singular en  $(0,0,0) \in \mathbb{R}^3$  de lo anterior obtenemos el siguiente desarrollo, recordemos que de acuerdo al teorema Watanabe (2), solo requerimos el cálculo del polo más grande  $\lambda_1$  y su multiplicidad  $m_1$  de la función zeta  $J(z)$  para así poder obtener la curva de aprendizaje de la máquina, de ello solo consideramos el primer polinomio de esta suma. De acuerdo a nuestra técnica de modificación tórica notamos que el desarrollo del polinomio anterior puede ser generado a través del siguiente ideal, con  $I \subset \mathbb{C}[a, b_1, b_2]$

$$I := \langle b_2^2; ab_1^2; ab_1b_2 \rangle.$$

De los generadores de este ideal conseguimos el siguiente politopo de Newton o cono poliédrico racional

$$\sigma = \text{Con}((0, 0, 2); (1, 2, 0); (1, 1, 1)),$$

ahora procedemos a construir una construcción geométrica de una variedad tórica afín  $X_{\sigma^\vee}$  asociada a dicho cono poliédrico como vemos a continuación.

Primero calculamos las bases de Hilbert asociadas al monoide  $S_\sigma = \sigma^\vee \cap \mathbb{Z}^3$ , mediante el programa Nórmaiz o el algoritmo del apéndice A, de donde conseguimos, este programa implícitamente calcula el cono dual asociada al cono a tratar

$$H_{\sigma^\vee} = \{e_3; e_1 + e_2 + e_3; e_1 + 2e_2\}$$

La variedad tórica afín asociada es

$$X_{\sigma^\vee} = \mathbf{Spec}\mathbb{C}[w_3, w_1w_2w_3, w_1w_2^2],$$

de donde se tiene el sistema coordinado

$$a = w_3;$$

$$b_1 = w_1w_2w_3;$$

$$b_2 = w_1w_2^2.$$

en esta parametrización obtenemos un mapeo de resolución  $g : X_{\Sigma'} \rightarrow X_\Sigma$  de tal modo  $H(g(w))$ ,  $w = (w_1, w_2, w_3) \in \mathbb{R}^3$  es no singular en  $(0, 0, 0)$  de acuerdo al teorema de Hironaka-Atiyah

$$\begin{aligned} H(g(w)) &= w_1^2w_2^4 + (w_3w_1^2w_2^2w_3^2 + (w_1w_2^2 - w_1w_2w_3^2)^2) + \dots + \text{términos de orden mayor}, \\ &= w_1^2w_2^4 + [w_1^2w_2^2 + (w_3^3 + (w_2 - w_3^2)^2)]^2 + \dots + \text{términos de orden mayor} \\ &= w_1^2w_2^4 + w_1^4w_2^4[w_3^3(w_2 - w_3^2)^2] + \dots + \text{términos de orden mayor}. \\ &= w_1^2w_2^4(1 + w_1^2(w_3^6 + 2w_3^3(w_2 - w_3^2)^2 + (w_2 - w_3)^4) + \dots + \text{términos de orden mayor}). \end{aligned}$$

de aquí es fácil ver que al escribir los términos del integrando de  $J(z)$  como vemos

$$J(z) = \int H(g(w))^z |g'(w)| du = \int ((1 + w_3^2w_1^2 + \dots)w_2^4w_1^2)^z |w_2^2w_1| dw_1 dw_2 dw_3 = \frac{f(z)}{4z + 3}.$$

de donde se sigue que el polo mas grande de  $J(z)$  es  $\lambda_1 = \frac{3}{4}$  y multiplicidad  $m_1 = 1$  y curva de aprendizaje dada por

$$K(n) = \frac{3}{4} \log(n) + C$$

**q.e.d.** Ver ref. [42].

## 5.2. Blow up y un perceptrón de tres capas

Ahora aplicamos nuestra teoría a un perceptrón de tres capas, un perceptrón es un modelo de red neuronal que surge de la analogía de el sistema de visión del ojo humano donde los nodos de entrada son los receptores de la información que capta el ojo (colores, formas etc.), este conjunto de datos es llevado a una capa que son la funciones sinápticas de unidades ocultas que procesan la información considerado como segunda capa de nodos en nuestra red, análogo al ojo humano hay una tercera capa que lleva la información al cerebro y que es la forma en la que percibimos el mundo, y esta última esta representada por los nodos de salida de nuestra red neuronal. Cada etapa del proceso de información del perceptrón esta representada por funciones de densidad de probabilidad como ya se han definido en la teoría. A continuación encontramos la curva de aprendizaje para este modelo caracterizado por las siguientes densidades.

Inferencia estadística verdadera de la red:

$$P(y|x, \omega) = \frac{1}{(2\pi s^2)^{N/2}} \exp\left(\frac{-1}{2s^2} \|y - f_K(x, \omega)\|^2\right),$$

con unidades ocultas definidas por la función,

$$f_K(x, \omega) = \sum_{k=1}^K a_k \sigma(b_k x + c_k),$$

donde  $y$ ,  $f(x, \omega)$ , y  $a_h$  son vectores  $N$  dimensionales,  $x$  y  $b_h$  son vectores  $M$  dimensionales,  $c_h \in \mathbb{R}$ , y  $\sigma(x) = \tanh(x)$  funciones sinápticas. Aquí  $M$ ,  $N$  y  $K$  son los números de unidades de entrada, unidades de salida, y unidades ocultas. Esta máquina no estima la desviación estándar  $s > 0$  constante, y consideremos que los ejemplos se escogen con distribución verdadera

$$q(y|x)q(x) = \frac{1}{(2\pi s^2)^{N/2}} \exp\left(-\frac{1}{2s^2}\|y\|^2\right) q(x).$$

**Teorema 5.1.2** Dada la R.N.B definida por la máquina estadística anterior, la cual es entrenada tomando ejemplos independientes de la distribución  $q(y|x)q(x)$ . Si una distribución previa satisface que  $\varphi(\omega) > 0$  en la vecindad del origen  $\omega = 0$ , entonces el polo más grande de la función zeta  $J(z)$  definida en teorema (2) Watanabe, es

$$\lambda_1 \leq \frac{K}{2} \min\{M + 1, N\}.$$

**Demostración.** Sea  $\omega = (a, b, c) \in \Theta$  el vector parámetro definido por

$$a = \{a_k \in \mathbb{R}^N; k = 1, 2, \dots, K\}$$

$$b = \{b_k \in \mathbb{R}^M; k = 1, 2, \dots, K\}$$

$$c = \{c_k \in \mathbb{R}; k = 1, 2, \dots, K\},$$

$$a_k = \{a_{kp} \in \mathbb{R}; p = 1, 2, \dots, N\}$$

$$b_k = \{b_{kp} \in \mathbb{R}; p = 1, 2, \dots, M\}.$$

La distancia Kullback o entropía de la información esta dada por

$$H(a, b, c) = \frac{1}{2s^2} \int \|f_K(x, a, b, c)\|^2 q(x) dx = \sum_{p=1}^N \sum_{h,k=1}^K B_{hk}(b, c) a_{hp} b_{kp}.$$

donde definimos  $B_{hk}(b, c)$ , para mas detalle de la obtención de estos resultados; ver el apéndice B.

$$B_{hk}(b, c) = \frac{1}{2s^2} \int \sigma(b_h * x + c_h) \sigma(b_k * x + c_k) q(x) dx$$

Nuestro propósito es encontrar el polo de la función

$$J(z) = \int_{W_\epsilon} H(a, b, c)^z \varphi(a, b, c) da db dc,$$

donde

$$W_\epsilon = \{(a, b, c); H(a, b, c) < \epsilon; \varphi(a, b, c) > 0\}.$$

Aplicamos la técnica de **blow up** a la distancia Kullback  $H(a, b, c)$ , introduciendo el mapeo de resolución

$$g : \{u_{hp}; 1 \leq h \leq K; 1 \leq p \leq N\} \mapsto \{a_{hp}; 1 \leq h \leq K; 1 \leq p \leq N\}$$

que esta definido por

$$\begin{aligned} a_{11} &= u_{11}; \\ a_{hp} &= u_{11}u_{hp}, \quad (h \neq 1 \text{ ó } p \neq 1). \end{aligned}$$

Sea  $u = (u_{11}, u')$  tal que

$$H(u, b, c) = u_{11}^2 H_1(u', b, c)$$

$$H_1(u', b, c) = B_{11}(b, c) + \sum_{h=2}^K 2B_{h1}(b, c) + \sum_{p=2}^N \sum_{h,k=1}^K B_{hk}(b, c)u_{hp}u_{kp},$$

y el jacobiano  $|g'(u)|$  del mapeo  $g$  es

$$|g'(u)| = \left| \det \left( \frac{\partial a_{ij}}{\partial u_{kl}} \right) \right| = |u_{11}^{NK-1}|.$$

definimos un conjunto de parámetros para  $\delta > 0$

$$U(\delta) = \{(u_{11}, u', b, c);_{11} | \leq \delta, |H_1(u', b, c)| < 1\}.$$

por hipótesis existe  $\delta > 0$  tal que

$$g(U(\delta)) \subset W_\epsilon.$$

En orden de obtener una cota superior de la curva de aprendizaje, restringimos la región de integración del espacio de parámetros, de la siguiente manera

$$J(z) = \int_{g(U(\delta))} H(a, b, c)^z \varphi(a, b, c) da db dc.$$

Del hecho de que  $\varphi(\omega) > 0$  en  $g(U(\delta))$ . En el cálculo del polo de  $J(z)$ , podemos asumir que  $\varphi(\omega) = \varphi_0$  constante en  $g(U(\delta))$ .

$$J(z) = \int_{U(\delta)} H(g(u), b, c)^z \varphi_0 |g'(u)| \lambda du' db dc$$

$$\begin{aligned}
&= \varphi_0 \int_0^\delta u_{11}^{2z} u_{11}^{NK-1} du_{11} \int H_1(u', b, c)^z du' dbdc \\
&= \frac{\varphi_0 \delta^{2z+NK}}{2z + NK} \int H_1(u', b, c)^z du' dbdc.
\end{aligned}$$

El polo de la función  $\frac{\delta^{2z+NK}}{(2z + NK)}$  es  $z = -NK/2$ . Sea  $-\lambda_1$  y  $-\lambda'_1$  respectivamente los polos más grandes de  $J(z)$  y

$$J_1(z) = \int H_1(u', b, c) du' dbdc.$$

Entonces, ya que  $H_1(u', b, c) \geq 0$ ,  $J_1(z)$  no tiene punto cero en el intervalo  $(-\lambda_1, \infty)$ . Si  $z = -NK/2$  es mayor que  $-\lambda_1$ , entonces  $z = -NK/2$  es un polo de  $J(z)$ , de otra manera  $J(z)$  tiene un polo mayor que  $-NK/2$ . Así  $\lambda_1 \leq NK/2$ .

En segundo lugar, consideramos otro blowing up  $g$ ,

$$g : \{u_{kp}, v_k; 1 \leq k \leq K; 1 \leq p \leq M\} \mapsto \{b_{kp}, c_k; 1 \leq k \leq K; 1 \leq p \leq M\}.$$

que es definido

$$\begin{aligned}
b_{11} &= u_{11} \\
b_{kp} &= u_{11} u_{kp}, \quad (k \neq 1) \text{ o } (p \neq 1), \\
c_k &= u_{11} v_k.
\end{aligned}$$

Entonces como el mismo método de la primera parte, existe una función analítica  $H_2(a, u', v)$ , tal que

$$H(a, b, c) = u_{11}^2 H(a, u', v),$$

que implica

$$J(z) = \varphi_0 \int_0^\delta u_{11}^{2z} u_{11}^{(M+1)K-1} du_{11} \int H_2(a, u', v)^z du' dv$$

$$= \varphi_0 \frac{\delta^{2z+(M+1)K}}{2z+(M+1)K} \int H_2(a, u', v)^z du' dv.$$

Por tanto  $\lambda_1 \leq (M+1)K/2$ .

Combinando los dos resultados de arriba, el polo más grande de  $-\lambda_1$  de los polos de  $J(z)$  satisface la desigualdad,

$$\lambda_1 \leq \frac{K}{2} \min\{N, M+1\},$$

lo cual se completa la prueba, **q.e.d.**

Con esta información y de acuerdo al corolario 4.3.1, tenemos que la curva de aprendizaje esta dada por

$$K(n) \leq \frac{K}{2} \min\{N, M+1\} \log(n) + o(\log(n)).$$

Y una desigualdad para la generalización del error,

$$G(n) \leq \frac{K}{2n} \min\{N, M+1\} + o\left(\frac{1}{n}\right).$$

Cuando consideramos una densidad normal de probabilidad previa  $\varphi$  es fácil generalizar el resultado de arriba en el caso de que la red neuronal tenga,  $M$  unidades de entrada,  $K_1$  primeras unidades ocultas,  $K_2$  segundas unidades ocultas, ...,  $K_p$  unidades ocultas, y  $N$  unidades de salida. Asumimos que las unidades ocultas y unidades de salida tienen parámetros bias. Entonces usando blow up, podemos generalizar el teorema, ver ref. [32].

$$\lambda_1 \leq \min\{(M+1)K_1, (K_1+1)K_2, (K_2+1)K_3, \dots, (K_{p-1}+1)K_p, K_p N\}.$$

Este resultado se sostiene cuando la función de regresión  $f_K$  es el caso especial de  $y = 0$ .

### 5.3. Modificación tórica en máquinas de aprendizaje no supervisado.

Por último resolvemos el problema de una familia de máquinas singulares en particular y son los llamados modelos paramétricos hereditarios empleados para resolver tareas de aprendizaje no

supervisado como es el caso de sistemas informáticos en minería de datos para el reconocimiento de patrones estadísticos, y sistemas de software de clusterización para la clasificación del tratamiento masivo de datos y predicción de datos no observables en el sistema.

Un ejemplo de este tipo de máquinas es el tratamiento masivo de información que viene de los estudios de sismica en yacimientos petroleros, en estos datos se recolecta información en cierta área cúbica considerable de volumen del yacimiento, como es el caso de, porosidades, permeabilidad del medio, viscosidades de aceites contenidos, presiones, mojabilidad del medio, coeficientes de compresibilidad entre otros, todos ellos serían considerados parámetros observables, una cantidad fundamental e involucrada en últimos desarrollos de vanguardia en esta tecnología es la estimación de la fractura del medio para la salida de flujo de aceites(petróleo) la cual es posible estimarla a través de densidades de probabilidad Bayesiana, este tipo de variable será una variable latente del modelo la cual no es observable, y es posible pronosticarla mediante aprendizaje no supervisado.

Este tipo de máquinas de aprendizaje no supervisado, consisten en su espacio de parámetros de dos tipos de variables, las variables observables del sistema que son parámetros que lo caracterizan desde el principio, y las variables latentes que son parámetros no observables como se dijo en el ejemplo anterior y aplicamos redes neuronales Bayesianas para su predicción. El desarrollo de este ejemplo viene dado de manera mas amplia en, ref. [44].

### 5.3.1. Estimación Bayesiana de variables latentes.

Consideramos la inferencia estadística verdadera del conjunto de variables observables  $X^n = \{x_1, \dots, x_n\}$ , que son independientes e idénticamente distribuidas donde

$$P(x|\omega) = \sum_{y=1}^K P(x, y|\omega),$$

con  $\omega \in \mathbb{R}^d$  un parámetro,  $x \in \mathbb{R}^n$  es una variable observable,  $y \in \{1, \dots, K\}$ , una variable latente.

Y la probabilidad verdadera de la variable  $x$ , esta dada por.

$$q(x) = \sum_{y=1}^{K^*} q(x, y),$$

donde  $K^* < K$ . Notamos que el modelo de aprendizaje tiene dimensión redundante en las variables latentes comparadas con la inferencia verdadera. Asumimos la existencia de un espacio de parámetros verdaderos por,

$$W = \{\omega^* : P(x, y|\omega^*) = q(x, y)\}.$$

Denotemos ahora el conjunto de variables no observables por  $Y^n = \{y_1, \dots, y_n\}$ , que corresponden a las variables latentes, así la forma completa de los datos es el par  $(X^n, Y^n)$ , el objetivo principal es la estimación bayesiana de las variables latentes como  $P(X^n|Y^n)$ . La cual estará definida por,

$$\begin{aligned} P(X^n|Y^n) &= \frac{Z(X^n, Y^n)}{Z(X^n)} \dots (*), \\ Z(X^n, Y^n) &= \int \prod_{i=1}^n P(x_i, y_i|\omega) \varphi(\omega, \eta) d\omega, \\ Z(X^n) &= \sum_{Y^n} Z(X^n, Y^n) = \int \prod_{i=1}^n P(x_i|\omega) \varphi(\omega, \eta) d\omega, \end{aligned}$$

donde  $\varphi(\omega, \eta)$  es una distribución previa con el hiperparámetro  $\eta$ . La probabilidad verdadera de  $Y^n$  será únicamente determinada y calculado por,

$$q(Y^n|X^n) = \frac{q(X^n, Y^n)}{q(X^n)} = \prod_{i=1}^n \frac{q(x_i, y_i)}{q(x_i)} \dots (**).$$

La exactitud de la estimación bayesiana de las variables latentes es medida a través de la generalización del error o curva de aprendizaje.

$$K(n) = \frac{1}{n} E_{XY} \left[ \ln \left( \frac{q(Y^n, X^n)}{P(Y^n, X^n)} \right) \right].$$

El siguiente teorema resume el cálculo de este tipo de máquinas de aprendizaje no supervisado.

**Teorema 5.3.1.** Sea la distribución verdadera de las variables latentes y la distribución estimada por las ecuaciones (\*) y (\*\*) respectivamente, y asumimos que la distancia Kullback  $H_{XY}(\omega)$  es analítica y la distribución previa  $\varphi(\omega, \eta)$  de los parámetros es de clase  $C_0^\infty$ , la curva de aprendizaje de estos modelos es expresada como:

$$K(n) = (\lambda_{XY} - \lambda_X) \frac{\ln(n)}{n} - (m_{XY} - m_X) \frac{\ln(\ln(n))}{\ln(n)} + o\left(\frac{\ln(\ln(n))}{n}\right).$$

donde  $\lambda_{XY}$  es el polo más grande de la función zeta  $J_{XY}(Z)$  definida en el capítulo anterior y  $m_{XY}$  es la multiplicidad de este polo, para ver detalles de la prueba ver [44].

### 5.3.2. Aplicación en una máquina de aprendizaje de mezclas de modelos.

Consideramos el siguiente modelo de aprendizaje de la siguiente mezcla de distribuciones dada por,

$$P(x|\omega) = \sum_{k=1}^K a_k f(x|b_k),$$

con función de distribución  $f$  de  $x$  y  $\omega$  consiste de mezcla de radios  $\{a_2, \dots, a_K\}$  y parámetros componentes  $b_k \in \mathbb{R}^{d_c}$ . Las mezcla de radios tienen restricciones  $a_k \geq 0$  y  $\sum_{k=1}^K a_k = 1$ . Escribimos a  $a_1$  en función de los parámetros  $a_1 = 1 - \sum_{k=1}^K a_k$ . El modelo verdadero es expresado como

$$q(x) = \sum_{k=1}^{K^*} a_k^* f(x|b_k^*),$$

donde  $a_k^*$  y  $b_k^*$  son todas constantes. Los números de componentes son asumidos que sean  $K^* < K$ . El modelo verdadero genera datos  $(X^n, Y^n)$ , donde  $Y^n = \{y_1, \dots, y_n\} \in \{1, \dots, K^*\}^n$  son las componentes etiquetadas que son estimadas en las bases de los datos incompletos  $X^n$ . La distribución

previa de los parámetros esta definida por

$$\begin{aligned}\varphi(\omega, \eta) &= \varphi(a, \eta_1)\varphi(b, \eta_2), \\ \varphi(a, \eta_1) &= \text{Dir}(a, \eta_1),\end{aligned}$$

donde  $a = \{a_1, \dots, a_K\}$ ,  $b = \{b_1, \dots, b_K\}$ , y  $\eta = \{\eta_1, \eta_2\}$ , y Dir es una distribución simétrica de Dirichlet incluyendo los hiperparámetros comunes  $\eta_1$  para todo  $a_k$ .

Ahora definimos la función zeta y la distancia Kullback para todos los datos completos de este modelo, variables observables y latentes  $(X, Y)$ ,

$$\begin{aligned}J_{XY}(z) &= \int H_{XY}(\omega)^z \varphi(\omega, \eta) d\omega, \\ H_{XY}(\omega) &= \sum_{y=1}^{K^*} \int q(x, y) \ln \left( \frac{q(x, y)}{P(x, y|\omega)} \right) dx.\end{aligned}$$

Para la máquina de aprendizaje de una mezcla de modelos la curva de aprendizaje Bayesiano esta acotada por **teorema 5.3.2.1**. Asumimos el las ecuaciones definidas en la parte de arriba para una mezcla de modelos. Sin embargo supongamos que existe una función  $M(x)$  tal que

$$H_X(\omega) \leq \int [q(x) - P(x|\omega)]^2 M(x) q(x) dx,$$

donde el lado derecho es la distancias kullback y es analítica. La estimación de Bayes (generalización de error o curva de aprendizaje) de las variables latentes definidas arriba con distribución previa tiene las siguientes cotas de error asintótico:

$$K(n) \geq \frac{(K - K^*)\eta_1}{2} \frac{\ln(n)}{n} + o\left(\frac{\ln(n)}{n}\right).$$

El teorema se demuestra mediante los siguientes lemas los cuales hacen uso de la técnicas de modificación tórica en las pruebas de los últimos dos lemas.

**Lema 5.3.2.1** Sean los polos de las funciones zeta  $\int H_1(\omega)^z \varphi(\omega) d\omega$  y  $\int H_2(\omega)^z \varphi(\omega) d\omega$ ,  $z = -\lambda_1$  y  $z = -\lambda_2$ , respectivamente. Se sostiene que  $\lambda_1 \leq \lambda_2$  cuando  $H_1(\omega) = H_2(\omega)$  en el soporte de  $\varphi(\omega)$ .

**Lema 5.3.2.2.** Bajo las mismas condiciones como lema anterior, se sostiene que  $\lambda_1 = \lambda_2$  si existen constantes positivas  $C_1$  y  $C_2$  tal que  $C_1 H_2(\omega) \leq H_1(\omega) \leq C_2 H_2(\omega)$ .

**Lema 5.3.2.3.** El polo más grande de la función zeta  $J_{XY}(z)$  es

$$\lambda_{XY} = \frac{K^* - 1 + K^* d_c}{2} + (K - K^*) \eta_1,$$

$$m_{XY} = 1.$$

**Lema 5.3.2.4.** El polo mas grande de la función zeta  $J_X(z)$  tiene la cota,

$$\lambda_X \leq \frac{K^* - 1 + K^* d_c}{2} + \frac{(K - K^*) \eta_1}{2}.$$

**Demostración del lema 5.3.2.3.** Consideramos el espacio de parámetros restringido  $W_1$ , que es una vecindad del conjunto dado por,

$$a_k = a_k^* \quad (2 \leq k \leq K^*),$$

$$a_k = 0 \quad (k > K^*),$$

$$b_k = b_k^* \quad (1 \leq k \leq K^*),$$

por que los polos de la función zeta dependen únicamente de la vecindad del espacio de parámetros verdaderos  $W_t$ , la distancia Kullback de acuerdo al modelo de máquina planteado en esta sección es la expresión.

$$H_{XY} = \sum_{k=1}^{K^*} a_k^* \left\{ \ln \left( \frac{a_k^*}{a_k} \right) + \int f(x|b_k^*) \ln \left( \frac{f(x|b_k^*)}{f(x|b_k)} \right) dx \right\} q = \left( 1 - \sum_{k=2}^{K^*} a_k^* \right) \ln \left( \frac{1 - \sum_{k=2}^{K^*} a_k^*}{1 - \sum_{k=2}^K a_k} \right) + \int f(x|b_1^*) \ln \left( \frac{f(x|b_1^*)}{f(x|b_1)} \right) dx + \sum_{k=2}^{K^*} \left\{ a_k^* \ln \left( \frac{a_k^*}{a_k} \right) + \int f(x|b_k^*) \ln \left( \frac{f(x|b_k^*)}{f(x|b_k)} \right) \right\}.$$

Hacemos un cambio de coordenadas  $\Phi_1(\omega)$  donde las singularidades de la expresión anterior son trasladadas al origen, donde queda:

$$a'_k = a_k - a_k^* \quad (2 \leq k \leq K^*),$$

$$\begin{aligned}
a'_k &= a_k \quad (k > K^*), \\
b'_{km} &= b_{km} - b_{km}^* \quad (1 \leq k \leq K^*, 1 \leq m \leq d_c), \\
b'_{km} &= b_{km} \quad (k > K^*, 1 \leq m \leq d_c).
\end{aligned}$$

Y la expansión de Taylor de  $\ln(1 + \Delta x)$  alrededor de  $|\Delta x| = 0$ , obtenemos

$$H_{XY}(\Phi_1(\omega)) = \sum_{k=2}^{K^*} a_k'^2 + \sum_{k=K^*+1}^K a'_k + \sum_{k=1}^{K^*} b_k'^2.$$

Con esta expresión formamos la función zeta  $J_1(z)$  de acuerdo a los teoremas de Watanabe quedando la siguiente expresión,

$$J_1(z) = \int H_{XY}(\Phi_1(\omega))^z \varphi(\Phi_1(\omega); \eta) d\Phi_1.$$

De acuerdo al lema 5.3.2.2 las posiciones de los polos de  $J_1(z)$  son las mismas que los de  $J_{XY}(\omega)$ . Ahora construimos la modificación tórica  $\Phi_2$  para el desarrollo de la expresión en  $H_{XY}(\Phi_1(\omega))$ , donde la transformación de monomios nos queda como,

$$\begin{aligned}
u_2 &= a'_2, \\
u_2 u_k &= a'_k \quad (2 < k < K^*), \\
u_2^2 u_k &= a'_k \quad (k > K^*), \\
u_2 v_{km} &= b'_{km} \quad (1 \leq k \leq K^*, 1 \leq m \leq d_c), \\
v_{km} &= b'_{km} \quad (k > K^*, 1 \leq m \leq d_c).
\end{aligned}$$

Y obtenemos la siguiente expresión para la función zeta en esta parametrización,

$$J_1(z) = \int_{\mathbb{F}} {}_2\Phi_1 f_1({}_2\Phi_1(\omega)) u_2^{2z} \varphi({}_2\Phi_1(\omega); \eta) |u_2|^{K^* - 2 + K^* d_c + 2(K - K^*)} d{}_2\Phi_1(\omega) + \dots,$$

donde  $f_1$  es una función que consiste de los parámetros excepto para  $u_2$  y un factor  $|u_2|$  que es derivado del jacobiano del mapeo de resolución  $\Phi_2$ . Es fácil confirmar que la distribución previa de Dirichlet tiene un factor  $u_2^{2(K - K^*)(\eta_1 - 1)}$ , lo cual indica que  $J_1(z)$  tiene un polo en  $z = -(K^* - 1 + K^* d_c) \eta_1$ . Considerando la simetría de los parámetros en  $H_{XY}(\omega)$ , determinamos que este es el polo más grande y su orden es  $m_{XY} = 1$  con lo que se prueba el lema, **q.e.d.**

**Demostración del lema 5.3.2.4.** Basados en la desigualdad dada en el teorema 5.3.2.1, existe una constante positiva tal que

$$H_X(\omega) \leq C_1 \int \{P(x|\omega) - q(x)\}^2 dx.$$

Usando la parametrización  $\Phi_1(\omega)$ , obtenemos

$$\begin{aligned} H_X(\Phi_1(\omega)) &= C_1 \int \left\{ \sum_{k=2}^{K^*} a'_k (f(x|b'_k + b_k^*) - f(x|b'_1 + b_1^*)) + \sum_{k=2}^{K^*} a_k^* (f(x|b'_k + b_k^*) - f(x|b_k^*)) \right. \\ &+ \left. (1 - \sum_{k=2}^{K^*} a_k^*) (f(x|b'_1 + b_1^*) - f(x|b_1^*)) + \sum_{k>K^*}^K a'_k (f(x|b'_k) - f(x|b'_1 + b_1^*)) \right\}^2 dx. \end{aligned}$$

Puesto que  $f(x|b_k)$  es un modelo regular, la expansión de Taylor en  $b'_k$  nos queda:

$$f(x|b'_k + b_k^*) = b_k'^T \frac{\partial}{\partial b'_k} f(x|b_k^*) + \dots$$

Entonces en el espacio de parámetros  $W_1$ , existe una constante positiva  $C_2$  tal que,

$$H(\Phi_1(\omega)) \leq C_2 \left\{ \sum_{k=2}^{K^*} a_k'^2 + \sum_{k=2}^{K^*} b_k'^2 + b_1^2 + \sum_{k>K^*}^K a_k'^2 \right\}.$$

Sea  $H_X(\Phi_1(\omega))$  igual a la expresión a la que tenemos arriba y consideramos la función zeta dada por

$$J_2(z) = \int_{\Phi_1(W_1)} H_X(\Phi_1(\omega))^z \varphi(\Phi_1(\omega); \eta) d\Phi_1(\omega).$$

De acuerdo al lema 5.3.2.1, un polo  $z = -\mu$  de esta función proporciona una cota del polo más grande de  $J_X(z)$ , tal que  $z = -\lambda_X \geq -\mu$ . Realizamos una modificación tórica  $\Phi_3$  en la expresión para  $H(\Phi_1(\omega))$  definida por,

$$\begin{aligned} u_2 &= a'_2, \\ u_2 u_k &= a'_k \quad (2 < k < K^*), \\ u_2 u_k &= a'_k \quad (k > K^*), \\ u_2 v_{km} &= b_{km} \quad (1 \leq k \leq K^*, 1 \leq m \leq d_c), \\ v_{km} &= b'_{km} \quad (k > K^*, 1 \leq m \leq d_c). \end{aligned}$$

Donde obtenemos para  $J_2(z)$

$$J_2(z) = \int_{\Phi_3} \Phi_1(W_1) f_2(\Phi_3 \Phi_1(\omega)) u_2^{2z} \varphi(\Phi_3 \Phi_1(\omega); \eta) |u_2|^{K^* - 2 + K^* d_c + (K - K^*)} d\Phi_3 \Phi_1(\omega),$$

donde  $f_2$  es una función de los parámetros excepto para  $u_2$ , y el factor  $|u_2|$  es debido al Jacobiano del mapeo de resolución teórico  $\Phi_3$ . Confirmamos fácilmente que la distribución previa de Dirichlet tiene un factor de  $u_2^{(K - K^*)(\eta_1 - 1)}$ . Por tanto, que  $J_2(z)$  tiene un polo en  $z = -\mu = -(K^* - 1 + K^* d_c)/2 - (K - K^*)\eta_1/2$ , y con ello probamos el lema, **q.e.d.**

**Demostración del teorema 5.3.2.1.** De acuerdo al teorema 5.3.1, se sostiene que

$$K(n) = (\lambda_{XY} - \lambda_X) \frac{\ln(n)}{n} + o\left(\frac{\ln(n)}{n}\right).$$

Combinando los dos últimos lemas obtenemos,

$$\begin{aligned} K(n) &\geq \left\{ \frac{K^* - 1 + K^* d_c}{2} + (K - K^*)\eta_1 - \frac{K^* - 1 + K^* d_c}{2} - \frac{(K - K^* d_c)\eta_1}{2} \right\} \frac{\ln(n)}{n} + o\left(\frac{\ln(n)}{n}\right) \\ &= \frac{(K - K^*)\eta_1 \ln(n)}{2n} + \frac{\ln(n)}{n}. \end{aligned}$$

Con esto completamos la prueba del teorema, **q.e.d.**

**Conclusiones del ejemplo.** Hemos podido desarrollar las pruebas de tres resultados importantes para el caso de máquinas de aprendizaje no supervisado en una mezcla de modelos, donde estos presentan singularidad dentro de su espacio de parámetros lo cual no permite vizualizar el cálculo del polo más grande y su multiplicidad de la función zeta  $J(z)$  asociada a su modelo y que de acuerdo a los teoremas de Watanabe probados en el capítulo anterior el cálculo de estos valores, nos permite obtener la curva de aprendizaje o estimar el error del aprendizaje del modelo de red neuronal estudiado. De ese modo se ha empleado la técnica de modificación tórica para hacer una resolución en el espacio de parámetros de la red neuronal y resolver el problema de aprendizaje en estos ejemplos.

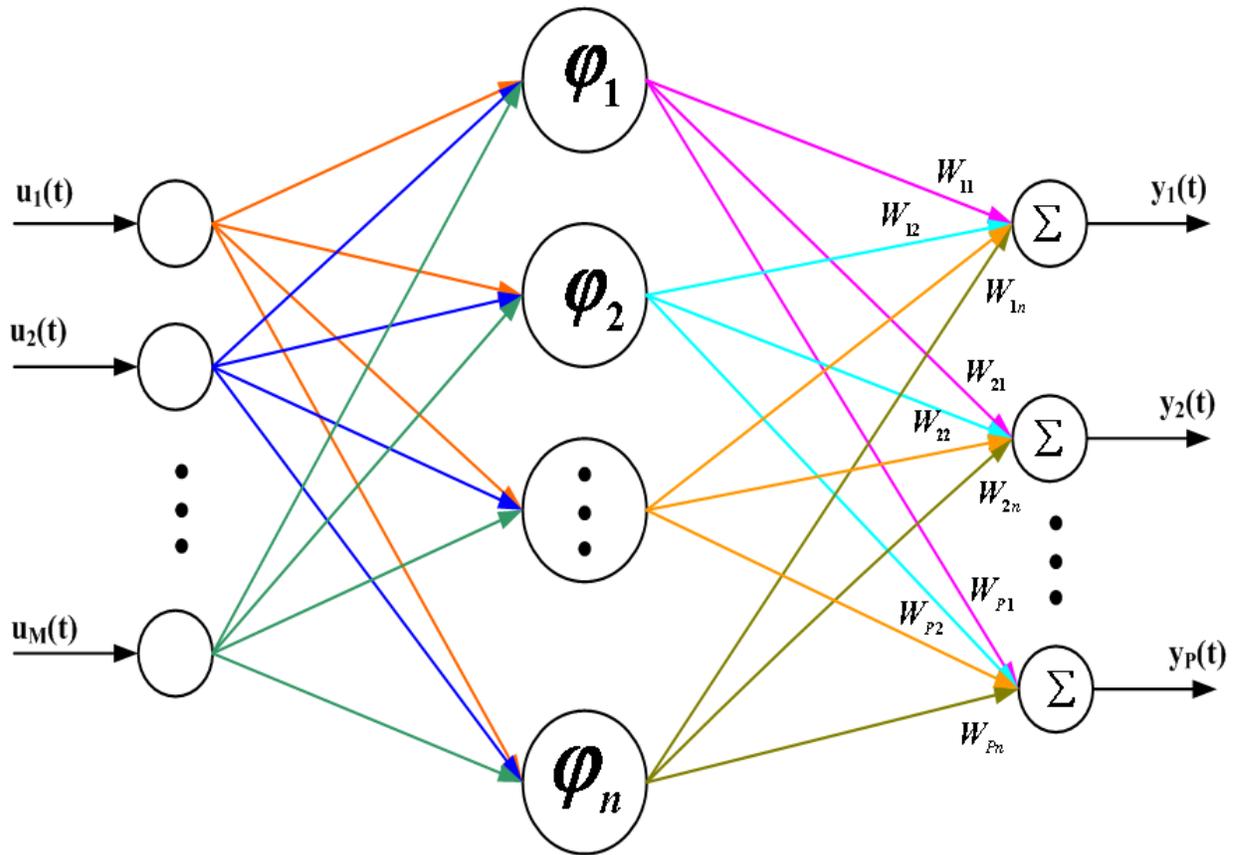


Figura 5.1: **Arquitectura de una red neuronal de una capa.** (Donde los  $\{u_1(t), \dots, u_M(t)\} \subset \Omega$ , son las funciones de vectores aleatorios que ponderan los nodos de entrada, las funciones  $\{\varphi_1, \dots, \varphi_n\} \subset \Theta \subset \Omega$  son las funciones sinápticas de la red o unidades ocultas donde se entrena la red neuronal, las ponderaciones  $W_{Pn}$  son las funciones de transición de nuestra red neuronal hacia la capa de salida y en nuestro caso son densidades de probabilidad bayesiana o inferencia verdadera de la máquina entre cada par de nodos  $P$  y  $n$ , por último tenemos la salida de la red representada por funciones  $\{y_1(t), \dots, y_P(t)\} \subset \Omega$  donde  $\Sigma$  significa la ponderación de todos los pesos de la red desde la capa oculta, ver [28]).

## Conclusiones

Hemos estudiado de manera amplia dos técnicas de resolución de singularidades mediante variedades tóricas y su importante aplicación para el desarrollo teórico del problema del aprendizaje y su solución en redes neuronales Bayesianas en particular el problema de las máquinas de aprendizaje singular que aparecen en una gran variedad de ejemplos de interés práctico en la industria, estos modelos han sido estudiados en la última década por Sumio Watanabe desarrollando resultados y teoremas de gran importancia en esta área como el teorema para curvas de aprendizaje (Watanabe 2001.), [29],[30], con el cual es posible estimar el error de entrenamiento de este tipo de redes neuronales, donde el ingrediente fundamental para el cálculo de este tipo de curvas de aprendizaje se encuentra en calcular el polo más grande de la función zeta y su multiplicidad definida en el teorema 4.3.1 a través de los métodos de la geometría algebraica expuestos en este trabajo, en particular vimos tres ejemplos de interés, uno de reciente publicación como es el caso del aprendizaje no supervisado en sistemas informáticos de minería de datos y clusterización [44]. Con ello nace un puente entre los teoristas la inteligencia artificial y la geometría algebraica, quedando abierto el cálculo de la resolución de singularidades en el espacio de parámetros de otras familias de máquinas de aprendizaje, con lo que se queda la pregunta, de que tipo de variedades algebraicas generalizan las redes neuronales singulares. Resumimos nuestro trabajo de la siguiente manera.

a). Revisamos de manera amplia la teoría de resolución de singularidades mediante variedades tóricas.

b). Entendimos y revisamos en gran detalle la **teoría del aprendizaje singular**, propuesta por

S. Watanabe en 2001 ref. [29], para máquinas de aprendizaje hereditario y como esta se relaciona con la geometría algebraica.

**c).** Proponemos el **teorema 3.7.1**, y el **corolario 4.3.3** como criterios alternativos para máquinas singulares y como una técnica de resolución de singularidades en sus espacios de parámetros de ellas; además de ser una metodología con más formalización matemática que las mostradas en Watanabe [29], [30], [32], [33], también es una técnica mas directa mediante el cálculo de la bases de Hilbert, computacionalmente podría resultar mas eficiente para espacios de parámetros de alta dimensión asociados a familias de redes neuronales, como se muestra en el apéndice B para algunos ejemplos.

**d).** A partir del uso de las bases de Hilbert pudimos caracterizar varios ejemplos de interés de máquinas de aprendizaje mediante variedades tóricas como se muestra en el apéndice B, lo que nos da una luz para un futuro, de como podríamos caracterizar familias de redes neuronales más generales mediante variedades algebraicas.

# Apéndice A

Un algoritmo para el cálculo de las bases de Hilbert en una variedad tórica afín.

El siguiente algoritmo puede consultarse en, [27] **B. Sturmfels**.

**Entrada:** Un conjunto que expanda el cono poliédrico convexo  $\sigma$ , en la retícula  $N \simeq \mathbb{Z}^d$ .

**Salida:** Las bases de Hilbert  $H$  del monoide  $S_\sigma \subset M = N^\vee$ .

**0.** Identificamos los monoides  $M$  y  $N$  con  $\mathbb{Z}^d$ .

**1.** Reemplazamos los generadores dados del cono  $\sigma$  por un nuevo conjunto generador  $\{v_1, \dots, v_d, v_{d+1}, \dots, v_m\}$

que consisten únicamente de puntos reticulares tal que  $\{v_1, \dots, v_d\}$  es una base de la retícula de  $\mathbb{Z}^d$ .

Y sea  $V$  la matriz de  $m \times d$  cuyas filas consisten de los vectores  $v_1, \dots, v_m$ .

**2.** La imagen de  $V$  en  $\mathbb{Z}^m$  es una subretícula saturada, i.e,  $\mathbb{Z}^m / \text{im}_{\mathbb{Z}}(V)$  es abeliano libre. Calculamos una matriz entera  $(m - d) \times m$  denotada por  $B$  cuyo núcleo sea igual a  $\text{im}_{\mathbb{Z}}(V)$ .

**3.** Calculamos las bases Graver  $Gr_B \subset \mathbb{Z}^m$  de la matriz  $B$ .

**4.** Para cada vector no negativo  $s = (s_1, \dots, s_m)$  en las bases Graver  $Gr_B$  determinamos y obtenemos los vectores únicos  $u \in \mathbb{Z}^d$ , tal que  $u * v_i = s_i$ . Estos vectores son la única preimagen bajo  $V$  de  $s$  en  $S_\sigma$ , y ellos constituyen la base de Hilbert del monoide  $S_\sigma$ .

Para otras alternativas del cálculo de bases de Hilbert de conos poliédricos, puede consultarse, [10]

**D. Cox, J. Little, D. O Shea, [18] M. Henk, R. Weismantel, [22] D. Maclagan, R.R.**

**Thomas, S. Faridi, L. Gold, A.V. Jayanthan, A. Khetan, T. Puthenpurakal.**

# Apéndice B

## Modificaciones tóricas: Justificación de cálculos.

En este apéndice desarrollaremos a mas detalle los resultados obtenidos en los desarrollos del teorema 5.1.2, del lema 5.3.2.3 y lema 5.3.2.4, principalmente obtendremos las transformaciones de monomios de estos tres ejemplos obtenidos de sus respectivas modificaciones tóricas. Todas las máquinas de inferencia estadística tratadas en este apéndice estarán definidas en el espacio de probabilidad  $(\Omega, \mathbb{F}, P)$ ; también todos los desarrollos planteados están justificados mediante el **teorema 3.1.7** y el **corolario 4.3.3** demostrados en el capítulo 3 y 4 respectivamente.

En el teorema 5.1.2 afirmamos que,

$$\begin{aligned} H(a, b, c) &= \frac{1}{2s^2} \int \|f_K(x, a, b, c)\|^2 q(x) dx \\ &= \sum_{p=1}^N \sum_{h,k=1}^K B_{hk}(b, c) a_{hp} b_{kp}. \end{aligned}$$

lo cual se sigue del siguiente desarrollo y tomando en cuenta las densidades de probabilidad previa y posterior dadas en el teorema tenemos.

$$\begin{aligned} H(a,b,c) &= \int q(y, x) * \ln \left( \frac{q(y, x)}{P(y, a, b, c)} \right) * q(x) dx dy. \\ &= \frac{1}{(2\pi s^2)^{N/2}} \int \exp \left( \frac{-\|y\|^2}{2s^2} \right) * q(x) * \ln \left[ \frac{1}{(2\pi s^2)^{N/2}} \exp \left( \frac{-\|y\|^2}{2s^2} \right) \right] - \ln P(y, a, b, c) dx dy \\ &= \frac{1}{(2\pi s^2)^{N/2}} \int \exp \left( \frac{-\|y\|^2}{2s^2} \right) q(x) * \\ &\quad \left[ \frac{-1}{2s^2} \|y\|^2 - \ln(2\pi s^2)^{N/2} + \frac{1}{2s^2} \|y - f_k(x, a, b, c)\|^2 - \ln \left( \frac{1}{(2\pi s^2)^{N/2}} \right) \right] dx dy. \end{aligned}$$

$$= \frac{1}{2s^2} \int \|y - f_k(x, a, b, c)\|^2 * q(x)dx.$$

y para el caso de que  $y = 0$  obtenemos el modelo de regresión el cual es explicada esta cuestión en ref.[32], para mayor detalle de esta igualdad, de ese modo conseguimos la igualdad a la que se quería llegar, para la obtención de las sumas se había definido la función  $f_K(x, \omega) = \sum_{k=1}^K a_k \sigma(b_k x + c_k)$ , sustituyendo en la integral anterior, recordamos también que  $\sigma(x) = \tanh(x)$ , puesto que todas estas funciones son analíticas en el plano complejo y tienen desarrollo en serie de potencias así obtenemos las sumas de arriba de donde definimos la función  $B_{hk}(b, c) = \frac{1}{2s^2} \int \sigma(b_h * x + c_h) \sigma(b_k * x + c_k) q(x) dx$ , de donde  $a \in \mathbb{R}^N$  y  $b \in \mathbb{R}^M$ .

Ahora construimos la modificación tórica utilizando la noción de conjuntos proyectivos y conos normales visto al final del capítulo 2 de la primera parte; primero desarrollamos las sumas que calculamos arriba de donde tenemos

$$\begin{aligned} H(a, b, c) &= \sum_{p=1}^N (B_{11}(b, c)a_{1p}a_{1p} + B_{22}(b, c)a_{2p}a_{2p} + \dots + B_{KK}a_{Kp}a_{Kp}) \\ &= B_{11}a_{11}^2 + B_{22}a_{21}^2 + \dots + B_{KK}a_{K1}^2 \\ &\quad + B_{11}a_{12}^2 + B_{22}a_{22}^2 + \dots + B_{KK}a_{K2}^2 \\ &\quad + B_{11}a_{13}^2 + B_{22}a_{23}^2 + \dots + B_{KK}a_{K3}^2 \\ &\quad + B_{11}a_{14}^2 + B_{22}a_{24}^2 + \dots + B_{KK}a_{K4}^2 + \\ &\quad \dots + \dots + \dots \\ &\quad + B_{11}a_{1N}^2 + B_{22}a_{2N}^2 + \dots + B_{KK}a_{KN}^2. \end{aligned}$$

Este polinomio visto en la coordenadas  $a_{hk} \in \mathbb{R}$ , es singular en  $(0, 0, \dots, 0) \in \mathbb{R}^N$ , construimos una modificación tórica en estas coordenadas utilizando el concepto de conjuntos proyectivos como se había dicho anteriormente, definimos las cartas afines utilizando el siguiente conjunto proyectivo,

$$U_j = \{[a_{11}, \dots, a_{1N}, a_{21}, \dots, a_{2N}, \dots, a_{K1}, \dots, a_{KN}] \in \mathbb{R}P^{KN-1} : a_{jj} \neq 0\}.$$

Donde  $\mathbb{R}P^{KN-1}$  es el espacio proyectivo real  $KN - 1$ -dimensional, y también existe una biyección como vimos en el capítulo 2, al espacio afín  $KN - 1$ -dimensional real  $\mathbb{R}^{KN-1}$ , dada por,

$$U_j : \mathbb{R}P^{KN-1} \mapsto \mathbb{R}^{KN-1}$$

$$U_j : [a_{11}, \dots, a_{1N}, a_{21}, \dots, a_{2N}, \dots, a_{K1}, \dots, a_{KN}] \mapsto \\ (1, a_{12}a_{11}^{-1}, \dots, a_{1N}a_{11}^{-1}, a_{21}a_{11}^{-1}, \dots, a_{2N}a_{11}^{-1}, \dots, a_{K1}a_{11}^{-1}, \dots, a_{KN}a_{11}^{-1})$$

ahora en estas coordenadas proyectivas redefinimos  $H(a, b, c) = u_{11}^2 H_1(a, b, c)$ , ya que:

$$H(a,b,c) = a_{12}^2 u_{11}^2 B_{11} + a_{13}^2 u_{11}^2 B_{11} + \dots + a_{1N}^2 u_{11}^2 B_{11} + a_{21}^2 u_{11}^2 B_{22} + \dots + a_{2N}^2 u_{11}^2 B_{22} + \dots + \\ a_{k1}^2 u_{11}^2 B_{KK} + \dots + a_{KN}^2 u_{11}^2 B_{KK}.$$

donde las nuevas coordenadas en el espacio afín  $\mathbb{R}^{KN-1}$  son  $(u_{11}, a_{12}, \dots, a_{1N}, \dots, a_{K1}, \dots, a_{KN}) \in \mathbb{R}^{KN-1}$ , y en este nuevo anillo coordenado construimos el cono poliédrico de Newton asociado al polinomio reparametrizado, lo cual nos queda como  $\sigma = \text{Con}(2e_1 + 2e_2, \dots, 2e_1 + 2e_{1N}, \dots, 2e_1 + 2e_{K1}, \dots, 2e_1 + 2e_{KN})$ , que de manera matricial nos quedaría el siguiente arreglo asociado al cono,

$$A_\sigma =$$

$$\begin{array}{cccccccc} 2 & 2 & 0 & . & . & . & 0 & \\ 2 & 0 & 2 & . & . & . & 0 & \\ 2 & 0 & 0 & 2 & . & . & 0 & \\ . & . & . & . & . & . & . & \\ . & 0 & . & . & . & 2 & 0 & \\ 2 & 0 & . & . & . & 0 & 2 & \end{array}$$

Se puede mostrar de manera inductiva y utilizando el programa Nórmaiz o el algoritmo del apéndice A, que las bases de Hilbert asociadas a este cono poliédrico racional, están dadas por el siguiente arreglo

$$H_{\sigma^v} =$$

$$\begin{array}{ccccccc} 1 & 0 & 0 & . & . & . & 1 \\ 1 & 0 & . & . & . & 1 & 0 \\ 1 & 0 & . & . & 1 & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & 1 & 0 & . & . & 0 \\ 1 & 1 & 0 & . & . & . & 0 \end{array}$$

donde los vectores reticulares de este arreglo nos representan un cono poliédrico regular, y como indican nuestros resultados del capítulo 3, tenemos por consiguiente una subdivisión estelar regular, y por tanto una modificación tórica que nos permite obtener la siguiente variedad tórica afín, tomando como exponente los elementos de esta base obtenemos la siguiente transformación de monomios

$$\begin{aligned} a_{11} &= v_{11}; \\ u_{11} &= v_{11}^{-1} \\ a_{hp} &= u_{11} * u_{hp}; \quad (h \neq 1 \text{ ó } p \neq 1). \end{aligned}$$

Con lo que conseguimos la primera modificación tórica construida en los resultados del teorema 5.1.2. Además cabe notar que agregamos una coordenada extra  $u_{11}$  siendo que trabajamos con el conjunto proyectivo  $U_j$ , de donde construimos la carta afín

$$A_0 = \{(1, a_{12}a_{11}^{-1}, \dots, a_{1N}a_{11}^{-1}, a_{21}a_{11}^{-1}, \dots, a_{2N}a_{11}^{-1}, \dots, a_{K1}a_{11}^{-1}, \dots, a_{KN}a_{11}^{-1}) \in \mathbb{R}^{KN-1}\},$$

que de acuerdo a los resultados del capítulo 3, es suficiente realizar la modificación tórica en esta carta, siendo que los morfismos tóricos son propios y se extienden a toda la variedad.

A continuación construimos la modificación tórica desarrollada en la prueba del lema 5.3.2.3, de ese modo trabajamos directamente con el polinomio obtenido en el lema,

$$H_{XY}(\Psi_1(\omega)) = \sum_{k=2}^{K^*} a'_k{}^2 + \sum_{k=K^*+1}^K a'_k + \sum_{k=1}^{K^*} b_{km}^2.$$

donde  $\omega = (a'_k, b_{km}) \in \mathbb{R}^{(K-1) \times d}$ ,  $a = \{a'_k\} \in \mathbb{R}^{K-1}$ ,  $b = \{b_{km}\} \in \mathbb{R}^d$ , y  $1 \leq m \leq d$ , también recordamos que  $\Psi_1(\omega)$  es una transformación que se hizo en el espacio de parámetros previa a la modificación tórica para trasladar la singularidad al origen, lo cual nos queda el polinomio de arriba.

Desarrollando estas sumas conseguimos

$$\begin{aligned} H_{XY}(\Psi_1(\omega)) &= a'_2{}^2 + \dots + a'_{K^*}{}^2 + a'_{K^*+1} + a'_{K^*+2} + \dots + a'_K + b_{1m}^2 + \dots + b_{K^*m}^2. \\ &= a'_2{}^2 + \dots + a'_{K^*}{}^2 + a'_{K^*+1}{}^2 + a'_{K^*+2} + \dots + a'_K + b_{11}^2 + \dots + b_{1d}^2 + b_{21}^2 + \dots + b_{2d}^2 + \dots + b_{K^*1}^2 + \dots + b_{K^*d}^2. \end{aligned}$$

Este se encuentra definido en las siguientes coordenadas afines

$$(a'_2, a'_3, \dots, a'_{K^*}, a'_{K^*+1}, \dots, a'_K, b_{11}, \dots, b_{1d}, b_{21}, \dots, b_{2d}, \dots, b_{K^*1}, \dots, b_{K^*d}) \in \mathbb{R}^{(K-1) \times K^*d}.$$

Y definimos las cartas afines mediante el siguiente conjunto proyectivo, análogo al ejemplo anterior,

$$\begin{aligned} U_j &= \{[a'_2, a'_3, \dots, a'_{K^*}, a'_{K^*+1}, \dots, a'_K, b_{11}, \dots, b_{1d}, b_{21}, \dots, b_{2d}, \dots, b_{K^*1}, \dots, b_{K^*d}] \in \mathbb{R}P^{(K-1) \times K^*d-1} : \\ &\quad a_j \text{ ó } b_{ij} \neq 0\} \end{aligned}$$

Donde  $\mathbb{R}P^{(K-1) \times K^*d-1}$  es el espacio proyectivo real  $(K-1) \times K^*d-1$ -dimensional, y también existe una biyección al espacio afín  $(K-1) \times K^*d$ -dimensional real  $\mathbb{R}^{(K-1) \times K^*d}$ , dada por,

$$U_j : \mathbb{R}P^{(K-1) \times K^*d-1} \longmapsto \mathbb{R}^{(K-1) \times K^*d}$$

$$\begin{aligned} U_j : [a'_2, a'_3, \dots, a'_{K^*}, a'_{K^*+1}, \dots, a'_K, b_{11}, \dots, b_{1d}, b_{21}, \dots, b_{2d}, \dots, b_{K^*1}, \dots, b_{K^*d}] &\longmapsto \\ (1, a'_3 a'_2{}^{-1}, \dots, a'_{K^*} a'_2{}^{-1}, a'_{K^*+1} a'_2{}^{-1}, \dots, a'_K a'_2{}^{-1}, \dots, b_{11} a'_2{}^{-1}, \dots, b_{1d} a'_2{}^{-1}, b_{21} a'_2{}^{-1}, \dots, \\ b_{2d} a'_2{}^{-1}, \dots, b_{K^*1} a'_2{}^{-1}, \dots, b_{K^*d} a'_2{}^{-1}). \end{aligned}$$

Reparametrizamos nuestro polinomio con esta biyección que llamamos  $\Psi_2$  en este nuevo sistema coordenado nos queda de la siguiente manera nuestro polinomio.

$$\begin{aligned} H_{XY}(\Psi_2(\Psi_1(\omega))) &= a'_3{}^2 u_2^2 + a'_4{}^2 u_2^2 + \dots + a'_{K^*}{}^2 u_2^2 + a'_{K^*+1} u_2^2 + \dots + a'_K u_2^2 + b_{11}^2 u_2^2 + \dots \\ &+ b_{1d}^2 u_2^2 + b_{21}^2 u_2^2 + \dots + b_{2d}^2 u_2^2 + \dots + b_{K^*1}^2 u_2^2 + \dots + b_{K^*d}^2 u_2^2. \end{aligned}$$

de donde ahora construimos el politopo de Newton asociado a esta expresión, y así el cono asociado es:

$$\begin{aligned} \sigma = \text{Con}(2e_2 + 2e_3, \dots, 2e_2 + 2e_{K^*}, 2e_2 + e_{K^*+1}, 2e_2 + e_{K^*+1}, \dots, 2e_2 + e_K, 2e_2 + 2e_{11}, \dots, \\ 2e_2 + 2e_{1d}, 2e_2 + 2e_{21}, \dots, 2e_2 + 2e_{2d}, \dots, 2e_2 + 2e_{K^*1}, \dots, 2e_2 + 2e_{K^*d}). \end{aligned}$$

donde los vectores  $e_K$  y  $e_{K^*d}$ , son vectores reticulares unitarios del espacio afín  $\mathbb{R}^{(K-1) \times K^*d}$ , la notación de doble índice solo es para hacerla congruente con el último factor de este producto cartesiano. Escribimos el arreglo matricial asociado a este cono lo cual nos queda

$$A_\sigma =$$

$$\begin{array}{cccccccc}
 2 & 2 & 0 & . & . & . & . & . & 0 \\
 . & 0 & 2 & 0 & . & . & . & . & 0 \\
 . & . & . & 2 & 0 & . & . & . & 0 \\
 2 & 0 & . & . & 1 & 0 & . & . & 0 & 0 \\
 . & . & . & . & . & . & . & . & . & . \\
 . & . & . & . & . & . & 1 & 0 & . & . \\
 2 & 0 & . & . & . & . & . & 2 & 0 & 0 \\
 2 & 0 & . & . & . & . & . & . & 2 & 0 \\
 2 & 0 & . & . & . & . & . & . & . & 2
 \end{array}$$

Del mismo modo que el ejemplo anterior se puede mostrar de manera inductiva y utilizando el programa Nórmaiz o el algoritmo del apéndice A, las bases de Hilbert asociadas a este cono y calculando primero su cono dual el cual hace implícitamente el programa Nórmaiz en nuestro caso obtenemos el siguiente arreglo de vectores reticulares.

$H_{\sigma^\vee} =$

$$\begin{array}{cccccccc}
 1 & 0 & . & . & . & . & . & . & 1 \\
 . & 0 & . & . & . & . & . & 1 & 0 \\
 . & . & . & . & . & . & 1 & . & 0 \\
 2 & 0 & . & . & . & . & 1 & 0 & . & . \\
 . & . & . & . & . & . & . & . & . & . \\
 2 & . & . & . & 1 & 0 & . & . & . & . \\
 . & 0 & . & 1 & 0 & . & . & . & . & 0 \\
 1 & 0 & 1 & 0 & . & . & . & . & . & 0 \\
 1 & 1 & 0 & . & . & . & . & . & . & 0
 \end{array}$$

De donde obtenemos la siguiente transforamción de monomios para la modificación tórica  $\Psi_2$ , y de donde se ha introducido la coordenada proyectiva  $u_2 = a_2'^{-1}$ , la cual esta bien definida para el conjunto proyectivo  $U_2$ .

$$\begin{aligned}
a'_k &= u_2 u_k; & (2 < k < K^*) \\
a'_k &= u_2^2 u_k; & (K^* < k \leq K) \\
b_{km} &= u_2 v_{km}; & (1 \leq k \leq K^*; 1 \leq m \leq d).
\end{aligned}$$

Si la suma para el caso de las coordenadas  $b_{km}$  se extendiera para  $k > K^*$ , en este caso se define la transformación  $b_{km} = v_{km}$ , y siendo que la matriz identidad siempre es la representación de un cono regular, la transformación sigue siendo una modificación tórica. De el conjunto proyectivo  $U_2$  se obtiene la siguiente carta afín

$$\begin{aligned}
A_2 = \{ & (1, a'_3 a'_2{}^{-1}, \dots, a_{K^*} a'_2{}^{-1}, a'_{K^*+1} a'_2{}^{-1}, \dots, a'_K a'_2{}^{-1}, b_{11} a'_2{}^{-1}, \dots, b_{1d} a'_2{}^{-1}, \\
& b_{21} a'_2{}^{-1}, \dots, b_{2d} a'_2{}^{-1}, \dots, b_{K^*1} a'_2{}^{-1}, \dots, b_{K^*d} a'_2{}^{-1}) \in \mathbb{R}^{(K-1) \times K^* d} \}.
\end{aligned}$$

Y de acuerdo a los resultados del capítulo dos, se obtiene la variedad tórica afín en notación de dicho capítulo  $X_{\sigma_2^\vee} = \mathbf{Spec} \mathbb{C}[\sigma^\vee \cap \mathbb{Z}^{(K-1) \times K^* d}] = A_2$  a través de esta carta afín y siendo que la acción del toro  $T_N$  la extiende sobre toda la variedad como se vio en cap. 3; también hacemos notar que se reproducen los resultados para la modificación tórica de este ejemplo como en la ref.[28].

Y por último construimos la modificación tórica para la prueba del lema 5.3.2.4, donde volvemos a utilizar la transformación  $\Psi_1$ , para trasladar la singularidad al origen, de ese modo

$$H(\Psi_1(\omega)) = \sum_{k=2}^{K^*} a'_k{}^2 + \sum_{k>K^*}^K a'_k{}^2 + b_1^2 + \sum_{k=2}^{K^*} b'_{km}{}^2.$$

donde  $\omega = (a'_k, b_1, b'_{km}) \in \mathbb{R}^{(K-1) \times (d+1)}$ ,  $a = \{a'_k\} \in \mathbb{R}^{K-1}$ ,  $b = \{(b'_{km}, b_1) \in \mathbb{R}^{d+1}$ , y  $1 \leq m \leq d\}$ , desarrollando estas sumas conseguimos,

$$\begin{aligned}
H(\Psi_1(\omega)) &= a'_2{}^2 + \dots + a'_{K^*}{}^2 + a'_{K^*+1}{}^2 + a'_{K^*+2}{}^2 + \dots + a'_K{}^2 + b_1^2 + b'_{2m}{}^2 + \dots + b'_{K^*m}{}^2. \\
&= a'_2{}^2 + \dots + a'_{K^*}{}^2 + a'_{K^*+1}{}^2 + a'_{K^*+2}{}^2 + \dots + a'_K{}^2 + b_1^2 + b'_{21}{}^2 + \dots + b'_{2d}{}^2 + b'_{31}{}^2 + \dots \\
&+ b'_{3d}{}^2 + \dots + b'_{K^*1}{}^2 + \dots + b'_{K^*d}{}^2.
\end{aligned}$$

Este polinomio se encuentra definido en las siguientes coordenadas afines.

$$(a'_2, a'_3, \dots, a'_{K^*}, a'_{K^*+1}, \dots, a'_K, b_1, b'_{21}, \dots, b'_{2d}, b'_{31}, \dots, b'_{3d}, \dots, b'_{K^*1}, \dots, b'_{K^*d}) \in \mathbb{R}^{(K-1) \times (K^*-1)d+1}.$$

Y definimos las cartas afines mediante el siguiente conjunto proyectivo, análogo al ejemplo anterior,

$$U_j = \{[a'_2, a'_3, \dots, a'_{K^*}, a'_{K^*+1}, \dots, a'_K, b_1, b'_{21}, \dots, b'_{2d}, b'_{31}, \dots, b'_{3d}, \dots, b'_{K^*1}, \dots, b'_{K^*d}] \in \mathbb{R}P^{(K-1) \times (K^*-1)d} : \\ a_j \text{ ó } b_1, b'_{ij} \neq 0\}$$

Donde  $\mathbb{R}P^{(K-1) \times (K^*-1)d}$  es el espacio proyectivo real  $(K-1) \times (K^*-1)$  d-dimensional, y también existe una biyección al espacio afín  $(K-1) \times (K^*-1)d$ -dimensional real  $\mathbb{R}^{(K-1) \times (K^*-1)d}$ , dada por,

$$U_j : \mathbb{R}P^{(K-1) \times (K^*-1)d} \mapsto \mathbb{R}^{(K-1) \times (K^*-1)d}$$

$$U_j : [a'_2, a'_3, \dots, a'_{K^*}, a'_{K^*+1}, \dots, a'_K, b_1, b'_{21}, \dots, b'_{2d}, b'_{31}, \dots, b'_{3d}, \dots, b'_{K^*1}, \dots, b'_{K^*d}] \mapsto \\ (1, a'_3 a'_2^{-1}, \dots, a'_{K^*} a'_2^{-1}, a'_{K^*+1} a'_2^{-1}, \dots, a'_K a'_2^{-1}, b_1 a'_2^{-1}, b'_{21} a'_2^{-1}, \dots, b'_{2d} a'_2^{-1}, \\ b'_{31} a'_2^{-1}, \dots, b'_{3d} a'_2^{-1}, \dots, b'_{K^*1} a'_2^{-1}, \dots, b'_{K^*d} a'_2^{-1}).$$

Reparametrizamos nuestro polinomio con esta biyección que llamamos  $\Psi_3$  en este nuevo sistema coordenado nos queda de la siguiente manera nuestro polinomio.

$$\mathbb{H}_{XY}(\Psi_3(\Psi_1(\omega))) = a'_3{}^2 u_2^2 + a'_4{}^2 u_2^2 + \dots + a'_{K^*}{}^2 u_2^2 + a'_{K^*+1}{}^2 u_2^2 + \dots + a'_K{}^2 u_2^2 + b_1^2 u_2^2 \\ + b'_{21}{}^2 u_2^2 + \dots + b'_{2d}{}^2 u_2^2 + b'_{31}{}^2 u_2^2 + \dots + b'_{3d}{}^2 u_2^2 + \dots + b'_{K^*1}{}^2 u_2^2 + \dots + b'_{K^*d}{}^2 u_2^2.$$

de donde  $u_2 = a'_2{}^{-1}$  y ahora construimos el politopo de Newton asociado a esta expresión, y así el cono asociado es:

$$\sigma = \text{Con}(2e_2 + 2e_3, \dots, 2e_2 + 2e_{K^*}, 2e_2 + e_{K^*+1}, 2e_2 + e_{K^*+1}, \dots, 2e_2 + e_K, 2e_2 + 2e_{11}, 2e_2 + 2e_{21}, \dots, \\ 2e_2 + 2e_{2d}, 2e_2 + 2e_{31}, \dots, 2e_2 + 2e_{3d}, \dots, 2e_2 + 2e_{K^*1}, \dots, 2e_2 + 2e_{K^*d}).$$

donde los vectores  $e_K$  y  $e_{K^*d}$ , son vectores reticulares unitarios del espacio afín  $\mathbb{R}^{(K-1) \times (K^*-1)d}$ , la notación de doble índice solo es para hacerla congruente con el último factor de este producto cartesiano. Escribimos el arreglo matricial asociado a este como lo cual nos queda

$$A_\sigma =$$

$$\begin{array}{cccccccc} 2 & 2 & 0 & . & . & . & . & . & 0 \\ . & 0 & 2 & 0 & . & . & . & . & 0 \\ . & . & . & 2 & 0 & . & . & . & 0 \\ 2 & 0 & . & . & 2 & 0 & . & . & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & 2 & 0 & . & . \\ 2 & 0 & . & . & . & . & . & 2 & 0 & 0 \\ 2 & 0 & . & . & . & . & . & . & 2 & 0 \\ 2 & 0 & . & . & . & . & . & . & . & 2 \end{array}$$

Del mismo modo que el ejemplo anterior se puede mostrar de manera inductiva y utilizando el programa Nórmliz o el algoritmo del apéndice A, las bases de Hilbert asociadas a este cono y calculando primero su cono dual el cual hace implícitamente el programa Nórmliz en nuestro caso obtenemos el siguiente arreglo de vectores reticulares.

$$H_{\sigma^\vee} =$$

$$\begin{array}{cccccccc}
1 & 0 & . & . & . & . & . & 1 \\
. & 0 & . & . & . & . & . & 1 & 0 \\
. & . & . & . & . & . & . & 1 & . & 0 \\
1 & 0 & . & . & . & . & 1 & 0 & . & . \\
. & . & . & . & . & . & . & . & . & . \\
1 & . & . & . & 1 & 0 & . & . & . & . \\
. & 0 & . & 1 & 0 & . & . & . & . & 0 \\
1 & 0 & 1 & 0 & . & . & . & . & . & 0 \\
1 & 1 & 0 & . & . & . & . & . & . & 0
\end{array}$$

De donde obtenemos la siguiente transformación de monomios para la modificación tórica  $\Psi_3$ , se ha introducido la coordenada proyectiva  $u_2 = a'_2{}^{-1}$ , la cual está bien definida para el conjunto proyectivo  $U_2$ .

$$\begin{aligned}
a'_k &= u_2 u_k; & (2 < k < K^*) \\
a'_k &= u_2 u_k; & (K^* < k \leq K) \\
b'_{km} &= u_2 v_{km}; & (1 \leq k \leq K^*; 1 \leq m \leq d).
\end{aligned}$$

Si la suma para el caso de las coordenadas  $b'_{km}$  se extendiera para  $k > K^*$ , en este caso se define la transformación  $b'_{km} = v_{km}$ , y siendo que la matriz identidad siempre es la representación de un cono regular, la transformación sigue siendo una modificación tórica. De el conjunto proyectivo  $U_2$  se obtiene la siguiente carta afín

$$\begin{aligned}
A_2 = \{ & (1, a'_3 a'_2{}^{-1}, \dots, a'_{K^*} a'_2{}^{-1}, a'_{K^*+1} a'_2{}^{-1}, \dots, a'_{K'} a'_2{}^{-1}, b_1 a'_2{}^{-1}, b'_{21} a'_2{}^{-1}, \dots, b'_{2d} a'_2{}^{-1}, \\
& b'_{31} a'_2{}^{-1}, \dots, b'_{3d} a'_2{}^{-1}, \dots, b'_{K^*1} a'_2{}^{-1}, \dots, b'_{K^*d} a'_2{}^{-1}) \in \mathbb{R}^{(K-1) \times (K^*-1)d} \}.
\end{aligned}$$

Y de acuerdo a los resultados del capítulo dos, se obtiene la variedad tórica afín en notación de dicho capítulo  $X_{\sigma_Y} = \mathbf{Spec} \mathbb{C}[\sigma^\vee \cap \mathbb{Z}^{(K-1) \times (K^*-1)d}] = A_2$  a través de esta carta afín y siendo que la acción del toro  $T_N$  la extiende sobre toda la variedad como se vio en cap. 3; también hacemos notar que se reproducen los resultados para la modificación tórica de este ejemplo como en la ref.[40].

## Lista de símbolos.

$st$ . Subdivisión estelar regular.

$Link$ . El link asociado a una estrella.

$\sigma$ . Cono reticular.

$\sigma^\vee$ . Cono reticular dual.

$A_\sigma$ . Arreglo matricial respecto los vectores retícula del cono  $\sigma$ .

$H_{\sigma^\vee}$ . Bases de Hilbert generadas por el monoide  $M$ .

$H_\sigma$ . Arreglo matricial asociado a las bases de Hilbert del cono  $\sigma$ .

$\Sigma$ . Abanico regular asociado a un politopo convexo.

$\sigma_Q$ . Cono poliédrico racional fuertemente convexo.

$\mathbb{Z}^n$ . Espacio reticular de enteros n-dimensional.

$\mathbb{Q}$ . Campo de los números racionales.

$M, N$ . Monoides isomorfos a  $\mathbb{Z}^n$ .

$S_\sigma$ . Monoide generado por el cono poliédrico racional  $\sigma_Q$ .

$\mathbb{C}$ . Campo de los números complejos.

$\mathbb{C}^*$ . Grupo multiplicativo de los números complejos.

$\mathbb{R}$ . Campo de los números reales.

$\mathbb{R}P^n$ . Espacio proyectivo real  $n$  dimensional.

$\mathbb{C}P^n$ . Espacio proyectivo complejo  $n$  dimensional.

$\mathbb{C}[M]$ . Anillo coordinado asociado al monoide  $M$ .

$\mathbb{C}(V)$ . Campo de funciones de la variedad  $V$ .

$T$ . Toro algebraico.

$X_\sigma$ . Variedad tórica afín asociada al cono reticular  $\sigma$ .

$R_\sigma$ . Algebra de monomios o anillo coordinado abstracto asociado al cono  $\sigma$ .

$\Psi$ . Morfismo tórico blow down.

$\Psi^{-1}$ . Morfismo tórico blow up o modificación tórica.

R.N.B. Red neuronal Bayesiana.

$P(X)$ . Densidad de probabilidad del evento  $X$ .

$P(X|Y)$ . Probabilidad condicional del evento  $X$  dado el evento  $Y$ .

$\Theta = \{\omega_\alpha\}$ . Conjunto de parámetros de una red neuronal.

$I_{ij}(\omega)$ . Matriz de información de Fisher.

$H(\omega)$ . Distancia Kullback o entropía de la información.

$K(n)$ . Curva de aprendizaje o generalización del error.

$C_0^\infty$ . Conjuntos de funciones continuas en todos sus órdenes de derivación.

$Re(z)$ . Parte real de un número complejo.

## Bibliografía

- [1] **A. Atanasov, C. López, A. Perry, N. Proudfoot, M. Thaddeus**, *Resolving Toric Varieties with Nash Blowups*, *Experimental Mathematics*, 20(3): 288 a 303, 2011.
- [2] **M. Atiyah and I.G MacDonald**, *Introduction to Commutative Algebra*, Addison-Wesley, Reading, MA, 1969.
- [3] **S. Amari**, *A Universal Theorem On Learning Curves*, *Neural Networks*, Vol.6, pp. 161-166, 1993.
- [4] **M. Aoyagui, S. Watanabe**, *Resolution Of Singularities And Generalization Error With Bayesian Estimation For Layered Neural Network*, Vol.J88-D-II, No.10, pp.2112-2124, 2005.
- [5] **M. Aoyagui, S. Watanabe**, *Generalization Error Of Three Layered Learning Model In Bayesian Estimation*, *Proceedings Of The IASTED International Conference Computational Intelligence*, San Francisco USA, pp.20-22, Nov. 2006.
- [6] **A. Brøndsted**, *An Introduction to Convex Polytopes*, Springer-Verlag, New York 10010, 1983.
- [7] **C.M. Bishop**, *Neural Networks And Machine Learning*, Serie F: Computer and Systems Sciences, Springer, Vol. 168, Cambridge CB2 3NH, U.K, 1997.
- [8] **D.M. Chickering, D. Heckerman, C. Meek**, *A Bayesian Approach to Learning Bayesian Networks With Local Structure*, Technical Report MSR-TR-97-07, Microsoft Research, Mi-

crosoft Corporation, On Microsoft Way, Redmond WA 98052, August 1997.

[9] **D. Cox, J. Little, D. O'Shea**, *Using Algebraic Geometry*, Springer, 2nd. Edition, August, 2004.

[10] **D. Cox, J. Little, D. O'Shea**, *Ideals, Varieties, and Algorithms (An Introduction to Computational Geometry and Commutative Algebra)*, Springer Science + Bussines Media, LLC, 3rd. Edition, New York 10013, 2007.

[11] **D. Cox, J. Little, H. Schenck**, *Toric Varieties*, Departement Of Mathematics, Amherst College, Amherst MA 01002, 2010.

[12] **L.D. García, M. Stillman, B. Sturmfels**, *Algebraic geometry of Bayesian networks*; Ph.D. dissertation, Department of Mathematics, Virginia Polytechnic Institute and State University, 2004.

[13] **G. Ewald**, *Combinatorial Convexity and Algebraic Geometry*, Spriger-Verlag, New York, Inc., 1996.

[14] **T. Fukui**, *Introduction to Toric Modifications With an Application to Real Singularities*, Departament Of Mathematics, Faculty Of Science, Saitama University, 255 Shimo-Okubo, Urawa 338-8570, Japan, pag. 96-114, 2000.

[15] **H. Hironaka**, *Resolution Of Singularities Of An Algebraic Variety Over A Field Of Zero Characteristic*, Annals Of Math., Vol.79, pp.109-326, 1964.

[16] **R. Hartshorne**, *Algebraic Geometry*, Springer, New York, August 1977.

- [17] **K. Hulek**, *Elementary Algebraic Geometry*, American Mathematical Society (AMS), Rhode Island 02904-2294, USA, 2003.
- [18] **M. Henk, R. Weismantel**, *On Hilbert Bases Of Polyhedral Cones*, Mathematical Foundations of Computer Science, 1999.
- [19] **E. Kunz**, *Introduction to Plane Algebraic Curves*, Birkhäuser, 2005.
- [20] **H. Matsumura**, *Commutative Algebra*, The Benjamin/Cummings Publishing Company, Inc., 2nd Edition, 1980.
- [21] **H. Matsumura**, *Commutative Ring Theory*, Cambridge University Press, Nagoya, Japan, 1980.
- [22] **D. Maclagan, R.R. Thomas, S. Faridi, L. Gold, A.V. Jayanthan, A. Khetan, T. Puthenpurakal**, *Computational Algebra and Combinatorics Of Toric Ideals*, Harisch Chandra Research Institute, Allahabad India, Dec. 2003.
- [23] **K. Nagata, S. Watanabe**, *A Method To Estimate The Generalization Error Of Singular Learning Machines By Decomposition Of Kullback Information*, Vol.J88-II, No.6, pp.994-1002, 2005.
- [24] **A.S. Poznyak**, *Advanced Mathematical Tools for Automatic Control Engineers; Vol. 2, Stochastic Techniques*, Elsevier, Oxford, UK, 2009.
- [25] **A.S. Poznyak, E.N. Sánchez, W. Yu**, *Differential Neural Networks For Robust Non-*

*linear Control: Identification, State Estimation and Trajectory Tracking*, World Scientific, Farrer Road, Singapore 912805, 2001.

[26] **I.R. Shafarevich**, *Algebraic Geometry I (Algebraic Curves, Algebraic Manifolds and Schemes)*, Springer-Verlag.

[27] **B. Sturmfels**, *Grobner Bases and Convex Polytopes*, Amer. Math. Soc; Providence, RI, 1996.

[28] **D. Saad**, *On-Line Learning In Neural Networks*, Cambridge University Press, New York 10011-4211 USA, 1998.

[29] **S. Watanabe**, *Algebraic Geometry Of Learning Machines With Singularities And Their Prior Distributions*, Journal Of Japanese Society For Artificial Intelligence, Vol. 16, No. 2; Pages 308-315(2001).

[30] **S. Watanabe**, *Algebraic Information Geometry For Learning Machines With Singularities*, Neural Networks, Vol. 14 Issue 8, pages 1049-1060, January 2001.

[31] **S. Watanabe**, *Training And Generalization Error Of Learning Machines With Algebraic Singularities*. IEICE Transactions, Vol. J84-A, No.1, pp.99-108, Jan. 2001.

[32] **S. Watanabe**, *Algebraic Geometrical Methods For Hierarchical Learning Machines*, Neural Networks, Vol. 14 Issue 8, pages 1049-1060, Oct. 2001.

[33] **S. Watanabe**, *Algebraic Analysis For Non-Identifiable Learning Machines*, Neural Computation, Vol. 13, No. 4, pp.899-933, 2001.

- [34] **S. Watanabe**, *Learning Efficiency Of Redundant Neural Networks In Bayesian Estimation*, IEEE Transactions On Neural Networks, Vol.12, No.6, pp.1475-1486, 2001.
- [35] **S. Watanabe, K. Yamazaki**, *A Probabilistic Algorithm to Calculate The Learning Curves Of Hierarchical Learning Machines With Singularities*, Trans. On IEICE, Vol. J85-D-II, No.3, pp.363-372, Mar. 2002.
- [36] **S. Watanabe, S.I. Amari**, *Learning Coefficients Of Layered Models When The True Distribution Mismatches The Singularities*, Neural Computation, Vol.15, No.5, 1013-1033, 2003.
- [37] **S. Watanabe**, *Algebraic Geometry Of Singular Learning Machines And Symmetry Of Generalization And Training Errors*, Neurocomputing, Vol. 67, Pages 198-213, August 2005.
- [38] **S. Watanabe**, *Algebraic Geometry And Statistical Learning Theory*, Cambridge Monographs On Applied And Computational Mathematics, Cambridge CB2 8RU, U.K, 2009.
- [39] **S. Watanabe**, *Almost All Learning Machines Are Singular*, Invited Paper in FOCI2007.
- [40] **S. Watanabe, M. Aoyagui**, *Generalization Error Of Three Layered Learning Model In Bayesian Estimation*, Proceedings of the IASTED International Conference, Computational Intelligence, pp. 20-22, Nov. 2006.
- [41] **K. Yamazaki, S. Watanabe**, *Singularities In Mixture Models And Upper Bounds Of Stochastic Complexity*, Neural Networks, Vol. 16, Issue 7, pages 1029-1038, September 2003.
- [42] **K. Yamazaki, M. Aoyagui, S. Watanabe**, *Stochastic Complexity And Newton Dia-*

*gram*, International Symposium On Information Theory And Its Applications, Parma Italy, pp.10-13, Oct. 2004.

[43] **K. Yamazaki, M. Aoyagui, S. Watanabe**, *Asymptotic Analysis Of Bayesian Generalization Error With Newton Diagram Neural Networks*, Neural Networks, Vol.23, No.1,pp.35-43, 2010.

[44] **K. Yamazaki**, *Asymptotic Accuracy of Bayes Estimation for Latent Variables with Redundancy*, Statistics Machine Learning, arXiv:1205.3234, May. 2012.