Singular Learning Theory 14 : From analytic to algebraic

SLT14

0

10/5/21

updated 3/6/21

Watanabe claims in [W] a deep role for algebraic geometry in statistical learning theory, which is at first quite surprising since the main function of interest $K(w)$ is analytic but only rarely a polynomial. So where do the polynomials come from?

Recall from [W, §7] that given a triple (p, q, φ) with Kullback-Leibler distance

$$K(w) = \int q(x) \log \frac{q(x)}{p(x|w)} dx$$

and zeta function

$$\zeta(z) = \int K(w)^z \varphi(w) dw$$

if the largest pole of ζ is $(-\lambda)$ then λ is called the learning coeffient of (p, q, φ), and as long as φ does not vanish on W_0 this agrees with the RLCT of K [W, Def^n 2.7], which recall is computed as the minimum of $(h_j + 1)/k_j$ over all $1 \leq j \leq d$ and coordinate patches in the resolution that put $K(g(u))$ into normal crossing form u^{2k} with Jacobian $u^h du$. From [W, Remark 7.2]

Def^n Two analytic functions $H(w), K(w)$ are equivalent if there exist $c_1, c_2 > 0$ with

$$c_1 H(w) \leq K(w) \leq c_2 H(w) \qquad \forall w \in W.$$

Lemma If H, K are equivalent they have the same learning coefficient.

shows $K \leq c_2 H$ implies $\lambda_K \leq \lambda_H$

Proof (Sketch) We may assume H, K are simultaneously resolved, and then $K \leq c_2 H$ means $K(g(u)) \leq c_2 H(g(u))$ or $u^{2k} \leq c_2 f(u) u^{2k'}$ hence by [W, Theorem 2.6] there is a real analytic function $g(u)$ with $u^{2k} = c_2 g(u) u^{2k'}$ and from the Taylor expansion $k_j \geq k_j'$ for all j. Hence $\dfrac{h_j+1}{2k_j'} \geq \dfrac{h_j+1}{2k_j}$. $\square$

Def^n We write $H \sim K$ if H, K are equivalent analytic functions. It is easy to see that this is an equivalence relation. We sometimes write $H \overset{c_1,c_2}{\sim} K$ for the situation in (1.3).

In Watanabe's book and papers [AW, W2] the following strategy is employed to compute the learning coefficient of a statistical model:

1. Compute a Taylor series expansion (in x) of $f(x,w) = \log \frac{q(x)}{p(x|w)}$ and use it to construct a polynomial function $H(w)$ equivalent to $K(w)$.

2. Perform resolution of singularities on H to determine its learning coefficient (and thus the learning coefficient of K).

In this note we focus on 1, following [W, Remark 7.6, p. 227].

Setup

We assume (p, q, φ) satisfy Fundamental Condition (I) of [W, Def^n 6.1] for some $s \geq 2$. In particular $F(x,w) = \log\left(q(x)/p(x,w) \right)$ is represented by an absolutely convergent power series in the neighborhood of an arbitrary $w^* \in W$

$$F(x,w) = \sum_{\alpha} a_\alpha(x)(w - w^*)^\alpha \qquad (1.1)$$

with $a_\alpha(x) \in L^s(X, q)$ (see [W, §5.2]). Then in the neighborhood of w^*

$$\begin{aligned} K(w) &= \int q(x) F(x,w)\, dx \\ &= \sum_{\alpha} (w - w^*)^\alpha \int a_\alpha(x)\, q(x)\, dx \end{aligned} \qquad (1.2)$$

is an absolutely convergent series, so $K(w)$ is analytic.

<u>Remark</u> Adapting this to conditional distributions works as follows: firstly replace x by x, y and assume $p(x,y|w) = p(y|x,w)q(x)$, $q(x,y) = q(y|x)q(x)$. Then

$$F(x,y,w) = \log\left(q(x,y) / p(x,y|w) \right) \\ = \log\left(q(y|x) / p(y|x,w) \right) \tag{2.1}$$

is represented by an absolutely convergent power series in the neighborhood of any $w^* \in W$

$$F(x,y,w) = \sum_\alpha a_\alpha(x,y)(w-w^*)^\alpha \tag{2.2}$$

with $a_\alpha(x,y) \in L^s(X \times Y, q)$. Integrating over Y is continuous and linear (see e.g. [MHS, Lemma L 17-½]) and hence induces a continuous linear map $L^s(X \times Y, q) \longrightarrow L^s(X, q)$. Applying this to (7.2) yields

$$\mathcal{T}(x,w) = \int q(y|x) F(x,y,w)\, dy \\ = \sum_\alpha \left\{ \int q(y|x) a_\alpha(x,y)\, dy \right\} (w-w^*)^\alpha \tag{2.3}$$

with $K(w) = \int \mathcal{T}(x,w) q(x)\, dx$. Moreover $b_\alpha(x) = \int q(y|x) a_\alpha(x,y)\, dy \in L^s(X, q)$.

<u>In this note</u> We assume that we are in the special case where there exists $G(x,w)$ with

$$K(w) = \int q(x) G(x,w)^2 dx = \| G(x,w) \|^2 \tag{2.4}$$

where the norm is in $L^2(X,q)$ and we assume that $G(x,w)$ is represented by a <u>polynomial</u> $G(x,w) = \sum_\alpha a_\alpha(x) w^\alpha$ with $a_\alpha(x) \in L^2(X,q)$, i.e. $a_\alpha(x) \equiv 0$ for $|\alpha|$ sufficiently large.

We assume given a linearly independent set $(e_j)_{j=1}^{\infty}$ in $L^2(X,q)$ such that

(A) The sequence $(\|e_j\|)_{j=1}^{\infty}$ is square-summable $\sum_{j=1}^{\infty}\|e_j\|^2<\infty$ and the induced bounded linear map $\ell^2(\mathbb{R})\longrightarrow L^2(X,q)$ is injective and has closed image (see [LIL]).

with absolutely convergent series $a_\alpha(x)=\sum_{j=1}^{\infty} c_{j,\alpha}\, e_j(x)$ for all α, with coefficients $c_{j,\alpha}\in\mathbb{R}$. Then with $f_j(w)=\sum_\alpha c_{j,\alpha} w^\alpha$ we have

$$\begin{aligned} G(x,w) &= \sum_\alpha a_\alpha(x)\, w^\alpha \\ &= \sum_{j=1}^{\infty}\sum_\alpha c_{j,\alpha} w^\alpha e_j(x) \qquad (3.1)\\ &= \sum_{j=1}^{\infty} f_j(w)\, e_j(x) \end{aligned}$$

with <u>polynomial</u> coefficients $f_j(w)$.

<u>Example</u> In Remark 7.6 (p. 225) Watanabe gives the following example of a statistical model $p(y|x,w)q(x)$ and true distribution given by (we assume $q(x)$ is given)

$$\begin{aligned} p(y|x,w) &= \tfrac{1}{2}\exp\left(-\tfrac{1}{2}(y-f(x,w))^2\right)\\ q(y|x) &= \tfrac{1}{2}\exp\left(-\tfrac{1}{2}(y-f_0(x))^2\right)\end{aligned}$$

so that $p(x,y|w)=p(y|x,w)q(x)$, $q(x,y)=q(y|x)q(x)$ and

$$\begin{aligned} K(w) &= \tfrac{1}{2}\int\left(f(x,w)-f_0(x)\right)^2 q(x)\,dx\\ &= \tfrac{1}{2}\| f(x,w)-f_0(x)\|^2\end{aligned}$$

So we take $G(x,w)=\frac{1}{\sqrt{2}}(f(x,w)-f_0(x))$. It remains to be checked G satisfies the hypotheses.

Hypothesis (A) implies $T : \ell^2(\mathbb{F}) \longrightarrow L^2(X, q)$ is bounded and bounded below, so there exist $c_1, c_2 > 0$ such that for all $a = (a_j)_{j=1}^{\infty}$ in $\ell^2(\mathbb{F})$

$$c_1\left(\sum_{j=1}^{\infty} |a_j|^2\right) \leqslant \left\| \sum_{j=1}^{\infty} a_j e_j \right\|^2 \leqslant c_2\left(\sum_{j=1}^{\infty} |a_j|^2\right) \qquad (4.1)$$

In particular, applying this to $\sum_{j=1}^{\infty} f_j(w) e_j(x) = G(x, w)$ we have

$$c_1\left(\sum_{j=1}^{\infty} f_j(w)^2\right) \leqslant K(w) \leqslant c_2\left(\sum_{j=1}^{\infty} f_j(w)^2\right) \qquad (4.2)$$

(pages were skipped)

The upper bound

Let $I \subseteq \mathbb{R}[w]$ denote the ideal generated by the polynomials $\{f_j\}_{j=1}^{\infty}$. By the Hilbert basis theorem $I = (f_1, \dots, f_J)$ for some integer J. Let $>$ denote the graded lex monomial order on $\mathbb{Z}^n_{\geq 0}$ where $\mathbb{R}[w] = \mathbb{R}[w_1, \dots, w_n]$, see [GGB, p.17]. We assume $f_1, \dots, f_J$ is a Gröbner basis of I (see [CLO] for background), and we introduce the following notation from [GGB] clarify

$$D_\alpha = \{ j \mid 1 \leq j \leq J \text{ and } LT(f_j) \mid w^\alpha \} \qquad \alpha \in \mathbb{Z}^n_{\geq 0} \tag{7.1}$$

where LT denotes the leading term with respect to $>$. We write $d_\alpha = |D_\alpha|$ and for $\alpha > \beta$ and an index $j \in D_\alpha$ we write (see [GGB, p. 11])

$$\mathcal{J}_{\alpha,\beta,j} = \left(\frac{w^\alpha}{LT(f_j)} f_j \right)_\beta \in \mathbb{R} \tag{7.2}$$

where $(-)_\beta$ denotes the coefficient of w^β. Then for $\alpha > \beta$

$$\mathcal{J}_{\alpha,\beta} = \sum_{j \in D_\alpha} \mathcal{J}_{\alpha,\beta,j} \tag{7.3}$$

The proposition on [GGB, p. 14] shows that if $f \in I$ then (using a "generic" form of the division algorithm, called Algorithm II in [GGB])

$$f = \sum_{m=1}^{\infty} \sum_{\alpha_1 > \dots > \alpha_m} \sum_{j \in D_{\alpha_m}} \frac{(-1)^{m+1}}{d_{\alpha_1} \cdots d_{\alpha_m}} \cdot f_{\alpha_1} \underbrace{\mathcal{J}_{\alpha_1,\alpha_2} \mathcal{J}_{\alpha_2,\alpha_3} \cdots \mathcal{J}_{\alpha_{m-1},\alpha_m}}_{\text{this is } 1 \text{ if } m=1, \text{ and } f_{\alpha_1} = f_{\alpha_m}} \frac{w^{\alpha_m}}{LT(f_j)} f_j \tag{7.4}$$

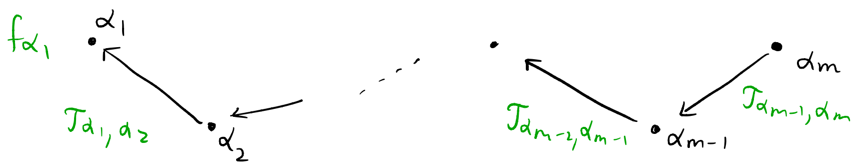
i.e. $m=1$ has $\sum_\alpha$ and $\alpha_1 = \alpha_m = \alpha$.

where the summand is zero if $d_{\alpha_i} = 0$ for any $1 \leq i \leq m$. This is a "sum over paths". We have used the observation on p. 2 of ggb5 which allows us to avoid fixing a downward closed set Λ.

In particular this means we may write $f = \sum_{k=1}^{J} a^k f_k$ with polynomials a^k given by

$$a^k = \sum_{m=1}^{\infty} \sum_{\substack{\alpha_1 > \cdots > \alpha_m \\ k \in D_{\alpha_m}}} \frac{(-1)^{m+1}}{d_{\alpha_1} \cdots d_{\alpha_m}} f_{\alpha_1} \underbrace{J_{\alpha_1,\alpha_2} J_{\alpha_2,\alpha_3} \cdots J_{\alpha_{m-1},\alpha_m}}_{\text{all constants}} \frac{w^{\alpha_m}}{LT(f_k)} \tag{8.1}$$

As above, for the $m=1$ term we have $\sum_\alpha$ and $f_\alpha \frac{w^\alpha}{LT(f_k)}$. Despite the "$\infty$" this sum is finite. We associate each summand to a path in the oriented graph which has $\mathbb{Z}^n_{\geq 0}$ as vertices and an edge $\beta \longrightarrow \alpha$ if $J_{\alpha,\beta} \neq 0$,



(8.2)

<u>Def^n</u> We call (8.2) the <u>division graph</u> of the Gröbner basis $f_1, \ldots, f_J$.

Note that, as opposed to the standard division algorithm, which "branches" depending on f, (8.1) depends on f only via the coefficients f_{α_1} and in this sense is "generic" in f. We can write this even more manifest by defining for $\alpha > \beta$

$$K(\alpha,\beta) = \sum_{m=1}^{\infty} \sum_{\substack{\alpha_1 > \cdots > \alpha_m \\ \alpha_1 = \alpha, \alpha_m = \beta}} \frac{(-1)^{m+1}}{d_{\alpha_1} \cdots d_{\alpha_m}} J_{\alpha_1,\alpha_2} J_{\alpha_2,\alpha_3} \cdots J_{\alpha_{m-1},\alpha_m} \tag{8.3}$$

the $m=1$ contribution is

FIX

so that

$$a^k = \sum_\alpha \sum_{\substack{\beta \\ k \in D_\beta}} f_\alpha K(\alpha,\beta) \frac{w^\beta}{LT(f_k)} \tag{8.4}$$

Since $>$ is total we can think of $\sum_\alpha$ as a sum over $\mathbb{N}$, and think of (8.4) as a dot product of a sequence $(f_\alpha)_\alpha$ and $\left(\sum_{\beta \text{ s.t. } k \in D_\beta} K(\alpha,\beta) w^\beta / LT(f_k) \right)_\alpha$. We set

$$\Omega^k_\alpha = \sum_{\substack{\beta < \alpha \\ k \in D_\beta}} K(\alpha,\beta) \frac{w^\beta}{LT(f_k)} \tag{8.5}$$

so that $\Omega^k_\alpha \in \mathbb{R}[w]$ and

$$a^k = \sum_\alpha f_\alpha \Omega^k_\alpha \qquad (9.1)$$

If we now have a sequence of polynomials $(F_i)_{i=1}^\infty$ in place of f then $F_i = \sum_{k=1}^J a_i^k f_k$ where $a_i^k = \sum_\alpha (F_i)_\alpha \Omega^k_\alpha$. Suppose we wish to construct an upperbound for $\sum_{i=1}^\infty F_i^2$ in terms of $\sum_{j=1}^J f_j^2$. Then we will want to use Cauchy-Schwartz as follows

$$\begin{aligned}\sum_{i=1}^N F_i^2 &= \sum_{i=1}^N \left(\sum_{k=1}^J a_i^k f_k\right)^2 \qquad (9.1)\\ &\leq \sum_{i=1}^N \left(\sum_{k=1}^J (a_i^k)^2\right)\left(\sum_{k=1}^J f_k^2\right)\\ &= \left(\sum_{j=1}^J f_j^2\right)\cdot \sum_{k=1}^J \sum_{i=1}^N (a_i^k)^2\end{aligned}$$

so the existence of the upper bounded hinges on convergence of $\sum_{i=1}^\infty (a_i^k)^2$.

<u>Remark</u> If there is no path from β to α in the division graph then $K(\alpha,\beta)=0$.

<u>Theorem</u> Suppose in addition to Hypothesis (A) of p2.5 we additionally assume

(B) The sequence $(c_{j,\alpha})_{j=1}^\infty$ is square-summable $\sum_{j=1}^\infty |c_{j,\alpha}|^2 < \infty$ for all α (recall that $f_j = \sum_\alpha c_{j,\alpha} w^\alpha$). We write $\| f_\alpha \| = \{\sum_{j=1}^\infty |c_{j,\alpha}|^2\}^{1/2}$.

(C) $\sum_\alpha \| f_\alpha \|_2^2 (\Omega^k_\alpha)^2 < \infty$ for $1 \leq k \leq J$.

Then K is equivalent to $\sum_{j=1}^J f_j^2$.

<u>Proof</u> The lower bound follows from (4.2) so it suffices to show the upperbound. Set $F_i = f_{J+i}$. Then by hypothesis (B), the sequence $((F_i)_\alpha)_{i=1}^\infty$ is square-summable for each $\alpha \in \mathbb{Z}^n_{\geq 0}$. Hence by Cauchy-Schwartz, in the above notation

Contains mistake

$$\sum_{i=1}^{N} (a_i^k)^2 = \sum_{i=1}^{N} \Big(\sum_\alpha (F_i)_\alpha \Omega^k_\alpha \Big)^2$$

error

$$\leqslant \sum_{i=1}^{N} \sum_\alpha (F_i)_\alpha^2 (\Omega^k_\alpha)^2 \qquad (10.1)$$

$$= \sum_\alpha \Big(\sum_{i=1}^{N} (F_i)_\alpha^2 \Big)(\Omega^k_\alpha)^2$$

The sum (9.1) is finite because f is polynomial, but in (9.2) we must keep in mind that as N increases the number of α being summed over may also increase without bound. For each N let $\alpha_N \in \mathbb{Z}^n_{\geqslant 0}$ be sufficiently large in the monomial order that

$$\sum_\alpha \Big(\sum_{i=1}^{N} (F_i)_\alpha^2 \Big)(\Omega^k_\alpha)^2 = \sum_{\alpha \leq \alpha_N} \sum_{i=1}^{N} (F_i)_\alpha^2 (\Omega^k_\alpha)^2 \qquad (10.2)$$

Let $\| F_\alpha \|_2 = \Big\{ \sum_{i=1}^{\infty} (F_i)_\alpha^2 \Big\}^{1/2}$ which we have assumed is finite. Then (10.2) gives

$$\sum_{i=1}^{N} (a_i^k)^2 \leqslant \sum_{\alpha \leq \alpha_N} \| F_\alpha \|_2^2 (\Omega^k_\alpha)^2 \qquad (10.3)$$

Hypothesis (C) says the RHS is bounded above and hence the LHS converges. This is uniform convergence, so the limit is a continuous function $A^k(w) = \sum_{i=1}^{\infty} (a_i^k)(w)^2$. Since W is compact $A^k(w) \leq M^k$ for some constant M^k. By (4.2) there exists $C > 0$ with

$$K(w) \leqslant C \sum_{j=1}^{\infty} f_j(w)^2$$

$$= C \sum_{j=1}^{J} f_j(w)^2 + C \sum_{i=1}^{\infty} F_i(w)^2 \qquad (10.4)$$

$$\overset{(9.2)}{\leqslant} C \sum_{j=1}^{J} f_j(w)^2 + C \Big(\sum_{j=1}^{J} f_j^2 \Big) \cdot \sum_{k=1}^{J} A^k(w)$$

$$\leqslant \Big\{ C + C(\textstyle\sum_k M^k) \Big\} \sum_{j=1}^{J} f_j(w)^2$$

as claimed. □

$$\sum_{i=1}^{N} (a_i^k)^2 = \sum_{i=1}^{N} \Big(\sum_\alpha (F_i)_\alpha \Omega^k_\alpha \Big)^2$$ those α appearing in F_i.

$$\leqslant \sum_{i=1}^{N} \Big(\sum_\alpha (F_i)_\alpha^2 \Big)\Big(\sum_\alpha (\Omega^k_\alpha)^2 \Big) \qquad (10.1)$$

$$= \Big(\sum_\alpha (\Omega^k_\alpha)^2 \Big) \sum_\alpha \sum_{i=1}^{N} (F_i)_\alpha^2$$

assuming this converges

The sum (9.1) is finite because f is polynomial, but in (9.2) we must keep in mind that as N increases the number of α being summed over may also increase without bound. For each N let $\alpha_N \in \mathbb{Z}^n_{\geq 0}$ be sufficiently large in the monomial order that

$$\sum_\alpha \Big(\sum_{i=1}^{N} (F_i)_\alpha^2 \Big)(\Omega^k_\alpha)^2 = \sum_{\alpha \leq \alpha_N} \sum_{i=1}^{N} (F_i)^2_\alpha (\Omega^k_\alpha)^2 \qquad (10.2)$$

Let $\| F_\alpha \|_2 = \Big\{ \sum_{i=1}^{\infty} (F_i)_\alpha^2 \Big\}^{1/2}$ which we have assumed is finite. Then (10.2) gives

$$\sum_{i=1}^{N} (a_i^k)^2 \leqslant \sum_{\alpha \leq \alpha_N} \| F_\alpha \|_2^2 (\Omega^k_\alpha)^2 \qquad (10.3)$$

Hypothesis (C) says the RHS is bounded above and hence the LHS converges. This is uniform convergence, so the limit is a continuous function $A^k(w) = \sum_{i=1}^{\infty} (a_i^k)(w)^2$. Since W is compact $A^k(w) \leq M^k$ for some constant M^k. By (4.2) there exists $C > 0$ with

$$K(w) \leqslant C \sum_{j=1}^{\infty} f_j(w)^2$$

$$= C \sum_{j=1}^{J} f_j(w)^2 + C \sum_{i=1}^{\infty} F_i(w)^2 \qquad (10.4)$$

$$\overset{(9.2)}{\leqslant} C \sum_{j=1}^{J} f_j(w)^2 + C \Big(\sum_{j=1}^{J} f_j^2 \Big) \cdot \sum_{k=1}^{J} A^k(w)$$

$$\leqslant \Big\{ C + C \big(\textstyle\sum_k M^k \big) \Big\} \sum_{j=1}^{J} f_j(w)^2$$

as claimed. □

<u>Example</u> In [W, Example 7.1] we have, up to a factor of ½ we will ignore,

$$G(x,w) = \sum_{j=1}^{\infty} \frac{x^j}{j!}\left(ab^j + cd^j\right) \tag{11.1}$$

where $\mathbb{R}[w] = \mathbb{R}[a,b,c,d]$. As is typical, there is some choice of how to allocate the factor $1/j!$ between $e_j(x)$ and $f_j(w)$. Let us choose $s_j, r_j > 0$ such that $s_j r_j = 1/j!$ and set $e_j(x) = s_j x^j$, $f_j(w) = r_j(ab^j + cd^j)$. The $\{e_j\}$ are linearly independent (under any reasonable choice of X, q) and Spencer's note shows that e.g. if $s_j = 1/\sqrt{j!}$ then $\sum_{j=1}^{\infty} \| e_j \|^2 < \infty$ if $X = [-1,1]$ with $q(x)$ uniform, so (A) is satisfied. Hence we get the lower bound $cH \le K$ of p. 6 with $H = \sum_{j=1}^{J} f_j^2$.

NO

We concentrate our attention here on (B), (C). For the moment take $f_j = ab^j + cd^j$. A Gröbner basis for I is f_1, g_2 where $g_2 = c^2d^2$. Note that if we can upper bound K by a constant multiple of $f_1^2 + g_2^2$ we can certainly upper bound it by a constant multiple of $f_1^2 + f_2^2$ (by Cauchy-Schwartz) so we just now assume $f_2 = g_2$ in the above. Note that for $|\alpha| \geqslant 5$ (to avoid f_1, f_2)

$$\| f_\alpha \| = \left\{ \sum_{j=1}^{\infty} |c_{j,\alpha}|^2 \right\}^{1/2} = \begin{cases} 1 & \alpha = (1,j,0,0) \text{ or } (0,0,1,j) \\ 0 & \text{otherwise} \end{cases}$$

Hence (B) holds and for (C) it suffices to show that

$$\sum_{\alpha \in \Lambda} (\Omega_\alpha^k)^2 < \infty \qquad \Lambda = \{(1,j,0,0) \mid j \geqslant 4\} \cup \{(0,0,1,j) \mid j \geqslant 4\} \tag{11.2}$$

For this we analyse the constants of (7.2) and paths of (8.2). For $\alpha > \beta$, if $LT(f_1) \mid w^\alpha$ and say $\alpha = \gamma + (1,1,0,0)$ then

$$\begin{aligned} J_{\alpha,\beta,1} &= \left(\frac{w^\alpha}{ab}(ab + cd) \right)_\beta = \left(w^\alpha + w^\gamma cd \right)_\beta \\ &= (w^\gamma cd)_\beta = \delta\big(\beta = \gamma + (0,0,1,1)\big) \\ &= \delta\big(\beta = \alpha + (-1,-1,1,1)\big) \end{aligned} \tag{11.3}$$

If $LT(f_2) \mid w^\alpha$ say $\alpha = \gamma + (0,0,2,2)$ then for $\alpha > \beta$

$$J_{\alpha,\beta,2} = \left(\frac{w^\alpha}{c^2d^2}(c^2d^2) \right)_\beta = (w^\alpha)_\beta = 0 \tag{12.1}$$

Hence $J_{\alpha,\beta} = \sum_{j\in D_\alpha} J_{\alpha,\beta,j}$ is zero if $ab \nmid w^\alpha$ and otherwise it is equal to $\delta(\beta = \alpha + (-1,-1,1,1))$. Thus the division graph (8.2) consists of edges from $\beta \in \mathbb{Z}^n_{\geq 0}$ to $\beta + (1,1,-1,-1)$ whenever this makes sense, i.e. belongs to $\mathbb{Z}^n_{\geq 0}$. Now

$$\Omega^k_\alpha = \sum_{\substack{\beta<\alpha \\ k\in D_\beta}} K(\alpha,\beta) \frac{w^\beta}{LT(f_k)}$$

and $LT(f_1) \mid w^\beta \iff \beta_1 \geq 1, \beta_2 \geq 1$, $LT(f_2) \mid w^\beta \iff \beta_3 \geq 2, \beta_4 \geq 2$. But $K(\alpha,\beta) \neq 0$ implies $\alpha = \beta + r\eta$ where $\eta = (1,1,-1,-1)$ and $r \geq 1$ is an integer. If $\alpha \in \Lambda$ then $\beta = \alpha - r\eta$ can only be in $\mathbb{Z}^n_{\geq 0}$ if $\alpha = (1,j,0,0)$ for some $j \geq 4$ in which case the only possibility for β is $\alpha - \eta = (0,j-1,1,1)$. But $LT(f_1) \nmid w^\beta$, $LT(f_2) \nmid w^\beta$ hence

$$\Omega^k_\alpha = 0 \qquad \forall k \ \forall \alpha \in \Lambda \tag{12.2}$$

proving (c). Hence by the Theorem

$$\begin{aligned} K &\sim f_1^2 + f_2^2 = (ab+cd)^2 + (ab^2+cd^2) \\ &\sim f_1^2 + g_2^2 = (ab+cd)^2 + c^4d^4 \end{aligned} \tag{12.3}$$

References

[W] S. Watanabe "Algebraic geometry and statistical learning theory" 2009.

[AW] M. Aoyagi, S. Watanabe "Resolution of singularities and the generalization error with Bayesian estimation for layered neural networks" 2007

[W2] S. Watanabe "Algebraic geometrical methods for hierarchical learning machines" Neural networks 2001.

[B1] http://therisingsea.org/notes/mast90132/background1.pdf

[CLO] D. Cox, J. Little, D. O'Shea "Ideals, Varities and algorithms", evil Springer.

[GGB] D. Murfet "Generic Gröbner bases".

[GGB2] D. Murfet "Generic Gröbner bases 2"

[LIL] D. Murfet "Linear independence and Lipschitz"

Now

$$\left(\Omega^k_\alpha\right)^2 = \sum_{\substack{\beta,\beta'<\alpha \\ k\in D_\beta\cap D_{\beta'}}} K(\alpha,\beta)K(\alpha,\beta')\frac{w^{\beta+\beta'}}{LT(f_k)^2} \tag{12.1}$$

entrywise not grevlex

Hence for $\gamma\in\mathbb{Z}^n_{\geq 0}$ (if we have $LT(f_k)=hw^\zeta$ then $k\in D_\beta$ iff. $w^\zeta \mid w^\beta$ iff. $\beta-\zeta \overset{e}{\geq} 0$)

$$\left(\Omega^k_\alpha\right)^2_\gamma = \sum_{\zeta \overset{e}{\leq} \beta,\beta'<\alpha} K(\alpha,\beta)K(\alpha,\beta')\left[\frac{w^{\beta+\beta'}}{LT(f_k)^2}\right]_\gamma$$

$$= \sum_{\zeta \overset{e}{\leq} \beta,\beta'<\alpha} K(\alpha,\beta)K(\alpha,\beta')\cdot\frac{1}{h^2}\delta(\beta+\beta'-2\zeta=\gamma)$$

$$= \frac{1}{h^2}\sum_{\substack{\zeta \overset{e}{\leq} \beta,\beta'<\alpha \\ \beta+\beta'=\gamma+2\zeta}} K(\alpha,\beta)K(\alpha,\beta') \tag{12.2}$$

Hence

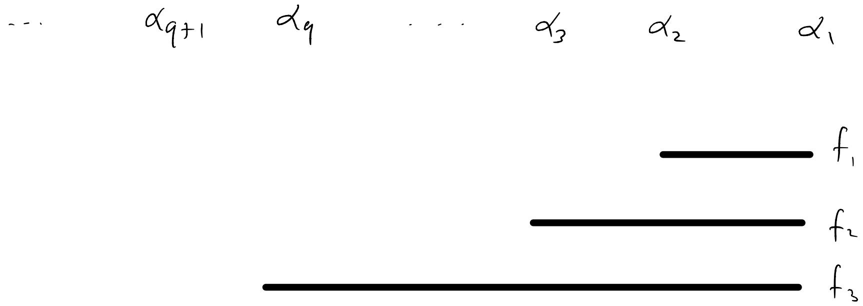
$$\sum_{\alpha\in\Lambda}\left(\Omega^k_\alpha\right)^2_\gamma = \frac{1}{h^2}\sum_{\alpha\in\Lambda}\sum_{\substack{\beta+\beta'=\gamma+2\zeta \\ \zeta \overset{e}{\leq}\beta,\beta'}} K(\alpha,\beta)K(\alpha,\beta') \tag{12.3}$$

Now

$$d_{\alpha - i\eta} = \begin{cases} 0 & ab \nmid \omega^{\alpha - i\eta} \text{ and } c^2d^2 \nmid \omega^{\alpha - i\eta} \\ 1 & ab \mid \omega^{\alpha - i\eta} \text{ or } c^2d^2 \mid \omega^{\alpha - i\eta} \\ 2 & ab \mid \omega^{\alpha - i\eta} \text{ and } c^2d^2 \mid \omega^{\alpha - i\eta} \end{cases} \tag{13.4}$$

Hence $d_\alpha = 1$, $d_{\alpha - \eta} = 1$ so $K(\alpha, \alpha - \eta) = -1$. Now $\beta = \alpha - r\eta = (0, j-1, 1, 1)$ so $LT(f_k) \mid \omega^\beta$ is impossible for $k \in \{1,2\}$. Hence

$$\sum_{\alpha \in \Lambda} (\Omega^k_\alpha)^2_\gamma = 0 \tag{13.5}$$



• $J_{a,b} = 0$ for $\alpha_a >> \alpha_b$
(i.e. $b >> a$)

so paths between distant α's involve many steps, hence if we can bound J's < 1 we can probably get convergence?

• assume $f_1, \dots, f_J$ are monomial?

$f_k = ab^k + cd^k$

$LT(f_j) \mid x^\alpha \iff \alpha_1 \geq 1$ & $\alpha_2 \geq 1$

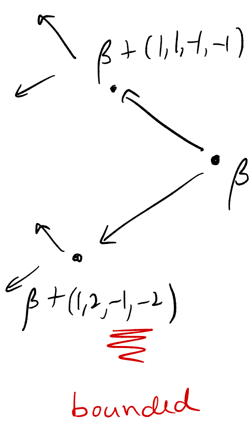
$\alpha = (1,1,0,0) + \gamma$

$$J_{\alpha,\beta,1} = \left(\frac{x^\alpha}{LT(f_1)} f_1 \right)_\beta$$

$\alpha > \beta$

$j = 1$

$$= \left(x^\gamma f_1 \right)_\beta = \left(x^\gamma ab + x^\gamma cd \right)_\beta$$
$$= \delta_{\alpha=\beta} + \delta_{\beta = \gamma + (0,0,1,1)}$$
$$= \delta(\beta = \alpha + (-1,-1,1,1))$$
$$= \delta(\alpha = \beta + (1,1,-1,-1))$$



$LT(f_2) \mid x^\alpha \iff \alpha_1 \geq 1, \alpha_2 \geq 2$

$\iff \alpha = (1,2,0,0) + \gamma$

$$J_{\alpha,\beta,2} = \left(\frac{x^\alpha}{LT(f_2)} f_2 \right)_\beta$$
$$= \left(x^\gamma ab^2 + x^\gamma cd^2 \right)_\beta = \delta(\beta = \alpha + (-1,-2,1,2))$$
$$= \delta(\alpha = \beta + (1,2,-1,-2))$$

In (12.3) we can rewrite the sum as being indexed by $r, r' \geqslant 1$ such that $\beta = \alpha - r\eta$, $\beta' = \alpha - r'\eta$ belong to $\mathbb{Z}^n_{\geqslant 0}$ and satisfy the required conditions. These are

$$\begin{aligned} 2\alpha - (r+r')\eta &= \gamma + 2\zeta \\ \zeta + r\eta,\ \zeta + r'\eta &\overset{e}{\leqslant} \alpha \end{aligned} \qquad (*)$$

Hence α only contributes if 2α is on a path in the division graph starting at $\gamma + 2\zeta$. There are only finitely many such α, so $\sum_\alpha (\Omega^k_\alpha)^2_\gamma$ is a finite sum.

$$\sum_\alpha (\Omega^k_\alpha)^2_\gamma \leqslant \frac{1}{h^2} \sum_{\substack{\alpha \\ 2\alpha = \gamma + 2\zeta + t\eta \\ \text{for some } t \geqslant 2}} \sum_{\substack{r+r'=t \\ r,r' \geqslant 1}} |K(\alpha, \alpha - r\eta) K(\alpha, \alpha - r'\eta)|$$

$$\leqslant \frac{1}{h^2} \sum_{\substack{\alpha \\ 2\alpha = \gamma + 2\zeta + t\eta \\ \text{for some } t \geqslant 2}} \sum_{\substack{r+r'=t \\ r,r' \geqslant 1}} 4 \cdot 2^{-(r+r')}$$

$$= \frac{4}{h^2} \sum_{\substack{\alpha \\ 2\alpha = \gamma + 2\zeta + t\eta \\ \text{for some } t \geqslant 2}} (t-1) 2^{-t}$$

$$\leqslant \frac{4}{h^2} \sum_{t=2}^{M^k_\gamma} (t-1) 2^{-t}$$

where $M^k_\gamma = \min\{ (\gamma + 2\zeta)_3, (\gamma + 2\zeta)_4 \}$.

The upper bound

We now assume $(\|e_j\|)_{j=1}^{\infty} \in \ell^2(\mathbb{R})$ and let $C>0$ be such that $K(w) = C\sum_{j=1}^{\infty} f_j(w)^2$.
Let $A_n(w) = \sum_{j=n+1}^{\infty} f_j(w)^2$ which is analytic since $A_n(w) = K(w) - \sum_{j=1}^{n} f_j(w)^2$, and clearly $A_{n+1}(w) \leq A_n(w)$ for all $w \in W$.

<u>Lemma</u> For all $w \in W$ there exists N_w such that $A_n(w) \leq \sum_{j=1}^{n} f_j(w)^2$ for all $n \geq N_w$.

<u>Proof</u> If $w \in W_0$ this is vacuous, since both sides are zero. If $w \notin W_0$ then $f_{j_0}(w) \neq 0$ for some j_0, and since $\lim_{n\to\infty} A_n(w) = 0$ there exists N such that $A_n(w) < f_{j_0}(w)^2$ for all $n \geq N$. Set $N_w = \max\{N, j_0\}$ then for $n \geq N_w$

$$\sum_{j=1}^{n} f_j(w)^2 \geq f_{j_0}(w)^2 > A_n(w)$$

as claimed. $\square$

By the Lemma the following quantity is well-defined:

$$M(w) := \inf\{ N \mid N \geq 1,\ A_n(w) \leq \sum_{j=1}^{n} f_j(w)^2 \text{ for all } n \geq N\}$$

Note that $M(w) = 1$ for all $w \in W_0$.

<u>Lemma</u> The function M is uppersemi-continuous: for every $w \in W$ there is an open neighborhood U of w such that $M(u) \leq M(w)$ for all $u \in U$.

<u>Proof</u> If $M = M(w)$ then $\sum_{j=1}^{n} f_j(w)^2 - A_n(w) \geq 0$

Hence by [MHS, Thm L21-10]

$$K(w) = \tfrac{1}{2}\| f(x,w) - f_0(x) \|^2$$

$$= \tfrac{1}{2}\sum_{j=1}^{\infty} |\langle f(x,w) - f_0(x), c_j(x)\rangle|^2 \qquad \text{(Parseval)}$$

$$= \tfrac{1}{2}\sum_{j=1}^{\infty} \Big|\sum_{k=1}^{\infty} f_k(w)\langle e_k, c_j\rangle\Big|^2$$

$$= \tfrac{1}{2}\sum_{j=1}^{\infty} \lim_{m\to\infty} \Big|\sum_{k=1}^{m} f_k(w)\langle e_k, c_j\rangle\Big|^2 \qquad (4.1)$$

$$\leq \tfrac{1}{2}\sum_{j=1}^{\infty} \lim_{m\to\infty} \Big(\sum_{k=1}^{m} f_k(w)^2\Big)\Big(\sum_{k=1}^{m} \langle e_k, c_j\rangle^2\Big)$$

(?)

$$= \tfrac{1}{2}\lim_{m\to\infty} \Big(\sum_{k=1}^{m} f_k(w)^2\Big)\Big(\sum_{k=1}^{m}\sum_{j=1}^{\infty} \langle e_k, c_j\rangle^2\Big)$$

$$= \tfrac{1}{2}\lim_{m\to\infty} \Big(\sum_{k=1}^{m} f_k(w)^2\Big)\Big(\sum_{k=1}^{m} \| e_k\|^2\Big)$$

We can rescale the $e_j(x)$ to ensure that $\sum_{k=1}^{\infty}\|e_k\|^2$ converges (for example if $e_j(x) = \frac{1}{(j-1)!}x^{j-1}$ on $X = [-1,1]$ with $q(x) = \frac{1}{2}$ uniform, $\|e_j\|^2 = \frac{1}{2}\Big[\frac{1}{(j-1)!}\Big]^2\int_{-1}^{1} x^{2j-2}dx$ $= \frac{1}{2}[1/(j-1)!]^2\Big[\frac{1}{2j-1}x^{2j-1}\Big]_{-1}^{1} = [1/(j-1)!]^2\, 1/2j-1$ and

$$\frac{\|e_{j+1}\|^2}{\|e_j\|^2} = \frac{[1/j!]^2\, 1/2j+1}{[1/(j-1)!]^2\, 1/2j-1} = \frac{1}{(j!)^2}\frac{1}{2j+1}\big((j-1)!\big)^2(2j-1)^2$$

$$= \Big(\frac{1}{j}\Big)^2\frac{(2j-1)^2}{2j+1} = \frac{1}{2j+1}\Big(2-\frac{1}{j}\Big)^2$$

hence $\lim_{j\to\infty}\frac{\|e_{j+1}\|^2}{\|e_j\|^2} = 0$ so the series $\sum_{k=1}^{\infty}\|e_k\|^2$ converges). Then

$$\boxed{K(w) \leq \tfrac{1}{2}C\sum_{k=1}^{\infty} f_k(w)^2 \qquad C = \sum_{k=1}^{\infty}\|e_k\|^2} \qquad (4.2)$$

Suppose $e_j(x) = \frac{1}{j!} x^j$ so that (2.3) is for any fixed j a globally convergent Taylor series expansion of the LHS at zero (we switch to $j \geq 0$). It is necessarily the case that this function is differentiable and that its derivative may be computed term-by-term, so

$$\begin{aligned} \frac{d^a}{dx^a}\Big(\sum_{j=0}^\infty f_j(w) e_j(x) \Big) &= \sum_{j=0}^\infty f_j(w) \frac{d^a}{dx^a} e_j(x) \\ &= \sum_{j=0}^\infty f_j(w) e_{j-a}(x) \qquad (5.1) \\ &= \sum_{j=a}^\infty f_{j+a}(w) e_j(x) \end{aligned}$$

If $w \in W_0$ then the LHS is identically zero, hence so is the RHS, and evaluating at $x=0$ gives $f_a(w) = 0$. This holds for any $a \geq 0$ and $w \in W$, so w is in the vanishing locus of the $(f_j)_{j \in J}$.

Hence under some mild hypotheses on $q(x)$ and with $e_j(x) = \frac{1}{j!} x^j$ we have

$$W_0 = \{ w \in W \mid f_j(w) = 0 \text{ for all } j \geq 0 \}$$

Let I be the ideal generated by $\{f_j\}_{j=0}^\infty$ in $\mathbb{R}[w]$, where w stands for a list of variables $w_1, \dots, w_d$ for some d. By the Hilbert basis theorem we can find J such that $I = (f_1, \dots, f_J)$. For any $j > J$ we have for some polynomials $a_j^1, \dots, a_j^J$ an equation

$$f_j = \sum_{k=1}^J a_j^k f_k \qquad (5.2)$$

and hence by Cauchy-Schwartz

$$\begin{aligned} f_j(w)^2 &= \Big(\sum_{k=1}^J a_j^k(w) f_k(w) \Big)^2 \qquad (5.3) \\ &\leq \Big(\sum_{k=1}^J a_j^k(w)^2 \Big)\Big(\sum_{k=1}^J f_k(w)^2 \Big) \end{aligned}$$

That is,

$$f_j^2 \leq \sum_{k=1}^{J} (a_j^k)^2 \sum_{k=1}^{J} f_k^2 \qquad (6.1)$$

Hence with $H = \sum_{k=1}^{J} f_k^2$

$$\sum_{j=1}^{m} f_j(\omega)^2 \leq \sum_{k=1}^{J} \left(\sum_{j=1}^{m} a_j^k(\omega)^2 \right) H(\omega) \qquad (6.2)$$

Assuming that $\sum_{j=1}^{\infty} a_j^k(\omega)^2$ converges for each k (I do not know how to show this) to a continuous function $A^k(\omega)$ on W

$$\sum_{j=1}^{\infty} f_j(\omega)^2 \leq \left[\sum_{k=1}^{J} A^k(\omega) \right] H(\omega) \qquad (6.3)$$

Now set $\alpha^k = \sup\{ A^k(\omega) \mid \omega \in W \}$, which is finite since W is compact and A^k is continuous. Then $\sum_{k=1}^{J} \alpha^k > 0$ since if $\alpha^k = 0$ for all k then $A^k(\omega) = 0$ for all k and ω, hence $a_j^k(\omega) = 0$ for all j, k, ω hence $f_j(\omega) = 0$ for all j, ω and so $K \equiv 0$. Except in this trivial case $D = \sum_{k=1}^{J} \alpha^k > 0$ and

$$\boxed{\sum_{j=1}^{\infty} f_j(\omega)^2 \leq D\, H(\omega)} \qquad (6.4)$$

From this and (4.2) we have

$$K \leq \tfrac{1}{2} C D H$$

To prove that K is equivalent to H we still need to establish $C' H \leq K$ for some $C' > 0$.

Equivalence via norm equivalence

Set $V_n = \mathrm{span}_{\mathbb{R}}\{e_1,\dots,e_n\} = \mathrm{span}_{\mathbb{R}}\{c_1,\dots,c_n\}$ as a subspace of $L^2(X,q)$.
Since the e_j are linearly independent $V_n \cong \mathbb{R}^n$ and we can define a norm on V_n for $v = \sum_{j=1}^n a_j e_j$ by

$$\|v\|_2 = \left\{ \sum_{j=1}^n |a_j|^2 \right\}^{1/2}$$

We let $\|v\|$ denote the restriction of the L^2-norm to V_n. Any two norms on a finite-dimensional normed space are Lipschitz equivalent (see e.g. [B1]) so we can find $c_1(n), c_2(n) > 0$ such that

$$c_1(n)\,\|v\|_2 \leq \|v\| \leq c_2(n)\,\|v\|_2 \qquad \forall v \in V_n \qquad (5.0)$$

in fact we may take

$$\begin{aligned} c_1(n) &= \inf\{ \|v\| \mid v \in V_n,\ \|v\|_2 = 1\} \\ c_2(n) &= \sup\{ \|v\| \mid v \in V_n,\ \|v\|_2 = 1\}. \end{aligned} \qquad (5.1)$$

Now set $s_n = \sum_{j=1}^n f_j(\omega) e_j(x)$, $r_n = \sum_{j=1}^n g_j(\omega) c_j(x)$ so $r_n, s_n \in V_n$ and $\|s_n - r_n\| \to 0$ as $n \to \infty$ since both series converge to $f(x,\omega) - f_0(x)$ in $L^2(X,q)$. We have

$$\begin{aligned} \left\{ \sum_{j=1}^n f_j(\omega)^2 \right\}^{1/2} &= \|s_n\|_2 \leq \frac{1}{c_1(n)}\|s_n\| \leq \frac{1}{c_1(n)}\|s_n - r_n + r_n\| \\ &\leq \frac{1}{c_1(n)}\|s_n - r_n\| + \frac{1}{c_1(n)}\|r_n\| \qquad (5.2) \\ &= \frac{1}{c_1(n)}\|s_n - r_n\| + \frac{1}{c_1(n)}\left\{ \sum_{j=1}^n g_j(\omega)^2 \right\}^{1/2} \end{aligned}$$

Hence

$$c_1(n)\left\{\sum_{j=1}^{n} f_j(w)^2\right\}^{1/2} \leq \|s_n - r_n\| + \left\{\sum_{j=1}^{n} g_j(w)^2\right\}^{1/2} \qquad (6.1)$$

Similarly

$$\left\{\sum_{j=1}^{n} g_j(w)^2\right\}^{1/2} = \|r_n\| \leq \|r_n - s_n\| + \|s_n\|$$
$$\leq \|s_n - r_n\| + c_2(n)\|s_n\|_2 \qquad (6.2)$$

Hence

$$\left\{\sum_{j=1}^{n} g_j(w)^2\right\}^{1/2} \leq \|s_n - r_n\| + c_2(n)\left\{\sum_{j=1}^{n} f_j(w)^2\right\}^{1/2} \qquad (6.3)$$

We cannot naively take $n \to \infty$ in (5.1), (5.2) because a priori $c_1(n)$ could converge to zero and $c_2(n)$ to ∞ as $n \to \infty$, rendering the inequality useless. Dealing with $\|s_n - r_n\|$ is awkward, so we can use $g_j^n(w)$ of p. 4 instead: from (5.0)

$$c_1(n)\left\{\sum_{j=1}^{n} f_j(w)^2\right\}^{1/2} \leq \left\{\sum_{j=1}^{n} g_j^n(w)^2\right\}^{1/2} \leq c_2(n)\left\{\sum_{j=1}^{n} f_j(w)^2\right\}^{1/2} \qquad (6.4)$$

That is, we have

$$\sum_{j=1}^{n} f_j(w)^2 \overset{c_1(n)^2,\, c_2(n)^2}{\sim} \sum_{j=1}^{n} g_j^n(w)^2 \qquad (6.5).$$

<u>Lemma</u> If two sets $\{a_1,\dots,a_r\}$, $\{b_1,\dots,b_s\}$ generate the same ideal in the ring of analytic functions on a compact set Ω then

$$\sum_{i=1}^r a_i^2 \sim \sum_{j=1}^s b_j^2 \qquad (7.1)$$

<u>Proof</u> By Cauchy-Schwartz, see Shaowei Lin's thesis Prop 4.3. Suppose $a_i = \sum_{j=1}^s h_j b_j$ then $a_i^2 \le (h_1^2 + \dots + h_s^2)(b_1^2 + \dots + b_s^2)$ and so with $C_i = \sup\{\sum_{j=1}^s h_j(w)^2 \mid w \in W\}$ we have $a_i^2 \le C_i \sum_{j=1}^s b_j^2$ hence $\sum_{i=1}^r a_i^2 \le (\sum_i C_i)(\sum_j b_j^2)$. If $C_i = 0$ for all i $\sum_{i=1}^r a_i^2 \equiv 0$ hence $a_i \equiv 0$ for all i so also $b_j \equiv 0$ for all j, so (7.1) is vacuous. □

By the Hilbert basis theorem we have that the ideal $I \subseteq \mathbb{R}[w]$ generated by the set $\{f_j\}_{j=1}^\infty$ can be generated by $f_1,\dots,f_J$ for some $J \geqslant 1$. Hence for $n \geqslant J$ we have by the Lemma that $\sum_{j=1}^n f_j^2 \sim H$ where we set $H(w) = \sum_{j=1}^J f_j(w)^2$. Hence

$$\sum_{j=1}^n g_j^n(w)^2 \overset{(6.5)}{\sim} \sum_{j=1}^n f_j(w)^2 \sim H \qquad (7.2)$$

In particular for $n \geqslant J$

$$\begin{aligned} H(w) &\le \sum_{j=1}^n f_j(w)^2 \le \frac{1}{c_1(n)^2} \sum_{j=1}^n g_j^n(w)^2 \\ \sum_{j=1}^n g_j^n(w)^2 &\le c_2(n)^2 \sum_{j=1}^n f_j(w)^2 \le c_2(n)^2 C H \end{aligned} \qquad (7.3)$$

where $C = \sum_{i=1}^n C_i$ where $f_i^2 \le C_i \sum_{j=1}^J f_j^2$. We can obviously take $C_i = 1$ for $1 \le i \le J$ and for $i > J$ we have to write $f_i = \sum_{j=1}^J a_{ij} f_j$ and $C_i = \sup\{\sum_{j=1}^J a_{ij}^2(w) \mid w \in W\}$. In particular

$$H(w) \le c_1(J)^{-2} \sum_{j=1}$$

The $g_j(w)$ involve potentially infinitely many $f_k(w)$'s, but we can let $g_j^n(w)$ denote the polynomial function of w with

$$\sum_{j=1}^{n} f_j(w)\, e_j(x) = \sum_{j=1}^{n} g_j^n(w)\, c_j(x) \qquad (4.1)$$

Since $\lim_{n\to\infty} \sum_{j=1}^{n} g_j^n(w) c_j(x) = G(x,w)$ we have

$$\begin{aligned} K(w) &= \| G(x,w) \|^2 \\ &= \lim_{n\to\infty} \left\| \sum_{j=1}^{n} g_j^n(w) c_j(x) \right\|^2 \qquad (4.2) \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} g_j^n(w)^2 \end{aligned}$$

Moreover since

$$\begin{aligned} g_j^n(w) &= \left\langle \sum_{\ell=1}^{n} f_\ell(w) e_\ell(x),\ c_j(x) \right\rangle \qquad (4.3) \\ &= \sum_{\ell=1}^{n} f_\ell(w) \langle e_\ell, c_j \rangle \end{aligned}$$

we have $\lim_{n\to\infty} g_j^n(w) = g_j(w)$ for all j.

Note that since $\{e_j\}_{j=1}^n$, $\{c_j\}_{j=1}^n$ are LI and span the same space there must be an invertible matrix $A^{(n)} \in M_n(\mathbb{C})$ with $A^{(n)}\underline{c} = \underline{e}$ where $\underline{e} = (e_1, \dots, e_n)^T$, $\underline{c} = (c_1, \dots, c_n)^T$.
Clearly $A^{(n)}_{\ell j} = \langle e_\ell, c_j \rangle$. We write $R^{(n)} = (A^{(n)})^T$. From (4.3), $\underline{g}^n = R^{(n)} \underline{f}$.
Hence $\underline{f} = (R^{(n)})^{-1} \underline{g}^n$ and hence

$$f_j(w) = \sum_{\ell=1}^{n} (R^{(n)})^{-1}_{j\ell}\, g_\ell^n(w) \qquad (4.4)$$

$$\sum_{j=1}^{n+1} f_j(w) e_j(x) = \sum_{j=1}^{n+1} g_j^{n+1}(w) c_j(x)$$

$$- \sum_{j=1}^{n} f_j(w) e_j(x) \qquad - \sum_{j=1}^{n} g_j^{n}(w) c_j(x)$$

$$f_{n+1}(w) e_{n+1}(x) = \sum_{j=1}^{n} (g_j^{n+1} - g_j^{n}) c_j + g_{n+1}^{n+1}(w) c_{n+1}(x)$$

$$\langle c_j, f_{n+1}(w) e_{n+1}(x) \rangle = g_j^{n+1} - g_j^{n}$$

$$f_{n+1}(w) \langle c_j, e_{n+1}(x) \rangle = g_j^{n+1} - g_j^{n}$$

<u>Paper</u>
- Deals $\mathbb{R}^n$
- deep networks
- always $K \sim H$

- <u>Check</u> g_j's are <u>not</u> polynomial
- <u>Check</u> higher dim input

$$\sum_{j=1}^{\infty} g_j(w)^2 \geq \sum_{j=1}^{\infty} \langle \Sigma_k f_k e_k, c_j \rangle$$

$$v = \sum_{j=1}^{\infty} f_j(w) e_j(x) = \sum_{j=1}^{\infty} g_j(w) c_j(x)$$

$$\|v\|^2 = \sum_{j=1}^{\infty} g_j(w)^2 \geq \sum_{k=1}^{\infty} |\langle v, c_k \rangle|^2$$

$$= \sum_{k=1}^{\infty} \left| \left\langle \sum_{j=1}^{\infty} f_j e_j, c_k \right\rangle \right|^2$$

$$= \sum_{k=1}^{\infty} \left| \sum_{j=1}^{k} f_j(w) \langle e_j, c_k \rangle \right|^2$$

$$= \sum_{k=1}^{\infty} (g_k^k)^2$$

$$\geq \sum_{k=1}^{J} (g_k^k)^2$$

From (4.3) we obtain (using $\|-\|$ to denote the operator norm on $\mathcal{B}(V)$, $V = \mathbb{R}^n$ with $\|-\|_2$)

$$\sum_{j=1}^n g_j^n(w)^2 = \| \underline{g} \|_2^2 \leq \| R^{(n)} \|^2 \| \underline{f} \|_2^2 = \| R^{(n)} \|^2 \sum_{j=1}^n f_j(w)^2 \qquad (5.1)$$

and from (4.4)

$$\sum_{j=1}^n f_j(w)^2 = \| \underline{f} \|_2^2 \leq \| (R^{(n)})^{-1} \|^2 \| \underline{g} \|_2^2 = \| (R^{(n)})^{-1} \|^2 \sum_{j=1}^n g_j^n(w)^2 \qquad (5.2)$$

Hence

$$\| (R^{(n)})^{-1} \|^{-2} \sum_{j=1}^n f_j(w)^2 \leq \sum_{j=1}^n g_j^n(w)^2 \leq \| R^{(n)} \|^2 \sum_{j=1}^n f_j(w)^2 \qquad (5.3)$$

Suppose we can show $\lim_{n\to\infty} \| R^{(n)} \| = C < \infty$. Since $1 = \| R^{(n)} (R^{(n)})^{-1} \| \leq \| R^{(n)} \| \cdot \| (R^{(n)})^{-1} \|$ we have $\| (R^{(n)})^{-1} \|^{-1} \leq \| R^{(n)} \|$ so also $\lim_{n\to\infty} \| (R^{(n)})^{-1} \|^{-1} < \infty$. So it will follow from (5.3) that $\sum_{j=1}^\infty f_j(w)^2 < \infty$ and ← provided $\lim_{n\to\infty} \| (R^{(n)})^{-1} \|^{-1} > 0$ (see p. 6.5)

$$K(w) \sim \sum_{j=1}^\infty f_j(w)^2 \qquad (5.4)$$

Note that

$$\begin{aligned} \| R^{(n)} \|^2 &= \sup\left\{ \frac{\| R^{(n)} \underline{x} \|_2^2}{\| \underline{x} \|_2^2} \;\middle|\; \underline{x} \in \mathbb{R}^n \setminus \{0\} \right\} \\ &= \sup\left\{ \sum_{\ell=1}^n (R^{(n)} \underline{x})_\ell^2 \Big/ \sum_{\ell=1}^n x_\ell^2 \;\middle|\; \underline{x} \neq 0 \right\} \\ &= \sup\left\{ \sum_{\ell=1}^n \Big\{ \sum_{j=1}^n R^{(n)}_{\ell j} x_j \Big\}^2 \Big/ \sum_{\ell=1}^n x_\ell^2 \;\middle|\; \underline{x} \neq 0 \right\} \end{aligned} \qquad (5.5)$$

But $R^{(n)}_{\ell j} x_j = \langle e_j, c_\ell \rangle x_j = \langle e_j x_j, c_\ell \rangle$ so (by Parseval's identity)

$$\sum_{\ell=1}^n \Big\{ \sum_{j=1}^n R^{(n)}_{\ell j} x_j \Big\}^2 = \Big\| \sum_{j=1}^n x_j e_j \Big\|^2 \qquad (5.6)$$

Hence (5.5) gives

$$\| R^{(n)} \|^2 = \sup\left\{ \left\| \textstyle\sum_{j=1}^{n} x_j e_j \right\|^2 / \|\underline{x}\|_2^2 \;\middle|\; \underline{x} \neq 0 \right\} = \| \mu_n \|^2 \qquad (6.1)$$

where $\mu_n : \mathbb{R}^n \longrightarrow L^2(X, q)$ is the linear map $\mu_n(u_i) = e_i$ where u_i is the standard basis, and $\|\mu_n\|$ denotes the operator norm with respect to $\|-\|_2$ on $\mathbb{R}^n$ and $\|-\|$ of $L^2(X,q)$. Hence $\|R^{(n)}\| = \|\mu_n\|$. It is clear that $(\|\mu_n\|)_{n=1}^{\infty}$ is an increasing function, so $\lim_{n\to\infty} \|R^{(n)}\| = \sup_n \|\mu_n\|$ and so it suffices to show the set $\{ \|\mu_n\| \mid n \geq 1 \}$ is bounded.

<u>Lemma</u> There is a well-defined linear map $\mu : \ell^2(\mathbb{R}) \longrightarrow L^2(X, q)$ defined by $\mu(\underline{a}) = \sum_{j=1}^{\infty} a_j e_j$. Moreover μ is bounded, and $\|\mu_n\| \leq \|\mu\|$ for all n.

<u>Proof</u> If $\underline{a} \in \ell^2(\mathbb{R})$ then $\sum_{j=1}^{\infty} \| a_j e_j \| = \sum_{j=1}^{\infty} |a_j| \, \|e_j\|$ converges by Hölder's inequality, since $|\underline{a}| \in \ell^2(\mathbb{R})$ and $(\|e_j\|)_{j=1}^{\infty} \in \ell^2(\mathbb{R})$. Hence $\sum_{j=1}^{\infty} a_j e_j$ converges [B1, Lemma B1-4]. We have

Hypothesis (A)

$$\| \mu(\underline{a}) \| = \| \lim_{n\to\infty} \textstyle\sum_{j=1}^{n} a_j e_j \|$$

$$\leq \lim_{n\to\infty} \textstyle\sum_{j=1}^{n} |a_j| \, \|e_j\|$$

$$\overset{\text{Hölder}}{\leq} \|\underline{a}\|_2 \, \|\underline{e}\|_2 \qquad \underline{e} = (\|e_j\|)_{j=1}^{\infty}$$

It is clear $\|\mu_n\| \leq \|\mu\|$. □

Hence $\|\mu\| \leq \|\underline{e}\|_2$ and so μ is bounded. Since $\|\mu_n\|$ is a supremum over a subset of the set $\|\mu\|$ is the supremum of, $\|\mu_n\| \leq \|\mu\|$. □

<u>Lemma</u> Suppose $(\|e_j\|)_{j=1}^{\infty} \in \ell^2(\mathbb{R})$. Then $K(w) \sim \sum_{j=1}^{\infty} f_j(w)^2$, and hence $W_0 = \{ w \in W \mid f_j(w) = 0 \text{ for all } j \geq 1 \}$.

<u>Proof</u> Immediate from (5.4) and the previous lemma. □

Details on lower bound

The matrix $A^{(n)}$ is lower triangular by construction: $A^{(n)}_{\ell j} = \langle e_\ell, c_j \rangle$ (see (3.2)).
Recall that

$$\tilde{c}_k = e_k - \sum_{i=1}^{k-1} \frac{\langle e_k, \tilde{c}_i \rangle}{\langle \tilde{c}_i, \tilde{c}_i \rangle} \tilde{c}_i$$

$$\therefore \quad \|\tilde{c}_k\|^2 = \langle e_k, e_k \rangle - 2 \sum_{i=1}^{k-1} \frac{\langle e_k, \tilde{c}_i \rangle^2}{\langle \tilde{c}_i, \tilde{c}_i \rangle} + \sum_{i=1}^{k-1} \frac{\langle e_k, \tilde{c}_i \rangle^2}{\langle \tilde{c}_i, \tilde{c}_i \rangle}$$

$$= \|e_k\|^2 - \sum_{i=1}^{k-1} \langle e_k, c_i \rangle^2 = \|e_k\|^2 - \left(\|e_k\|^2 - \langle e_k, c_k \rangle^2 \right)$$

$$= \langle e_k, c_k \rangle^2$$

By definition

$$c_1 = \frac{1}{\|e_1\|} e_1$$

$$c_2 = \frac{1}{\|\tilde{c}_2\|} \left(e_2 - \langle e_2, c_1 \rangle c_1 \right)$$

$$c_k = \frac{1}{\|\tilde{c}_k\|} \left(e_k - \sum_{i=1}^{k-1} \langle e_k, c_i \rangle c_i \right)$$

$$= \frac{1}{|\langle e_k, c_k \rangle|} e_k - \sum_{i=1}^{k-1} \frac{\langle e_k, c_i \rangle}{|\langle e_k, c_k \rangle|} c_i$$

Since $(A^{(n)})^{-1} \underline{e} = \underline{c}$ we see that $(A^{(n)})^{-1}$ is lower triangular and has diagonal entries $1/\|\tilde{c}_k\|$.

Read as μ^{-1}

$$I + N$$

$$(I+N)^{-1} = I - N + N^2 - N^3 + \cdots N^k$$

- identify $(R^{(n)})^{-1}$ with inverse μ^{-1}

- "bounded inverse theorem" $\Longrightarrow \|\mu^{-1}\| < \infty$

  deduce $\|(R^{(n)})^{-1}\| \leq \|\mu^{-1}\|$

- Replacing K by sum of squares is related to "obvious positivity"

$$\|\mu^{-1}_{*}\| = \sup\left\{ \frac{\quad}{\|a\|} \right.$$

$$\mu^{-1}(c) =$$

Gram-Schmidt

Applying Gram-Schmidt we may produce an orthonormal basis for the span of $\{e_j(x)\}_{j=1}^{\infty}$ in $L^2(X,q)$, call it $(c_j)_{j=1}^{\infty}$. Recall that

$$\tilde{c}_1 = e_1$$

$$\tilde{c}_2 = e_2 - \frac{\langle e_2, \tilde{c}_1\rangle}{\langle \tilde{c}_1, \tilde{c}_1\rangle}\tilde{c}_1 \tag{3.1}$$

$$\vdots$$

$$\tilde{c}_k = e_k - \sum_{i=1}^{k-1} \frac{\langle e_k, \tilde{c}_i\rangle}{\langle \tilde{c}_i, \tilde{c}_i\rangle}\tilde{c}_i$$

and $c_i = \frac{1}{\|\tilde{c}_i\|}\tilde{c}_i$ so that by construction $e_j \in \mathrm{span}(c_1,\dots,c_j)$ for $j \geq 1$. Indeed we have $e_k = \tilde{c}_k + \sum_{i=1}^{k-1} \frac{\langle e_k, \tilde{c}_i\rangle}{\langle \tilde{c}_i, \tilde{c}_i\rangle}\tilde{c}_i$. Hence

$$\langle e_k, c_j\rangle = \frac{1}{\|\tilde{c}_j\|}\langle e_k, \tilde{c}_j\rangle = \begin{cases} 0 & j > k \\ \|\tilde{c}_k\| & j = k \\ \langle e_k, \tilde{c}_j\rangle / \|\tilde{c}_j\| & j < k \end{cases} \tag{3.2}$$

The coefficients of $G(x,w)$ in this new basis are (note the $g_j(w)$ need not be polynomial)

$$\begin{aligned} g_j(w) := \langle G(x,w), c_j(x)\rangle &= \sum_{k=1}^{\infty} \langle f_k(w) e_k(x), c_j(x)\rangle \\ &= \sum_{k=1}^{\infty} f_k(w)\langle e_k, c_j\rangle \end{aligned} \tag{3.3}$$

and we have

$$G(x,w) = \sum_{j=1}^{\infty} g_j(w) c_j(x) \tag{3.4}$$

$$K(w) = \sum_{j=1}^{\infty} g_j(w)^2$$