HILBERT'S 13TH PROBLEM

by

Ziqin Feng

BS, Shandong University, 2001

Submitted to the Graduate Faculty of

the Department of Mathematics in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF MATHEMATICS

This dissertation was presented

by

Ziqin Feng

It was defended on

January 2010

and approved by

Dr. Paul Gartside, Department of Mathematics

Dr. Benjamin Espinoza, Department of Mathematics

Dr. Robert Heath, Department of Mathematics

Dr. Christopher Lennard, Department of Mathematics

Dr. Vladimir Uspenskiy, Department of Mathematics

Dissertation Director: Dr. Paul Gartside, Department of Mathematics

HILBERT'S 13TH PROBLEM

Ziqin Feng, PhD

University of Pittsburgh, 2010

The 13th Problem from Hilbert's famous list [16] asks whether every continuous function of three variables can be written as a superposition (in other words, composition) of continuous functions of two variables. Let X be a space. A family $\Phi \subseteq C(X)$ is said to be **basic** for X if each f in $C(X)$ can be written: $f = \sum_{q=1}^{n} (g_q \circ \phi_q)$, for some $\phi_1, \cdots, \phi_n$ in Φ and $g_1, \ldots, g_n \in C(\mathbb{R})$. For $\psi_1, \psi_2, \ldots, \psi_m \in C(X)$, define $\Sigma : X^m \to \mathbb{R}$ by $\Sigma(x_1, x_2, \ldots, x_m) = \sum_{p=1}^{m} \psi_p(x_p)$. A family Ψ_m of maps $X \to \mathbb{R}$ is **elementary** in dimension m if the family of maps $\Phi_m = \{\Sigma(\psi_1, \psi_2, \ldots, \psi_m) : \psi_1, \ldots, \psi_m \in \Psi_m\}$ is basic for X^m. Kolmogorov and Arnold [18, 4] showed that the closed unit interval has a finite elementary family in every dimension, thereby solving Hilbert's 13th Problem.

Define a new cardinal invariant basic $(X) = \min\{|\Phi| : \Phi \text{ is a basic family for } X\}$. It is established that a space has a finite basic family if and only if it is finite dimensional, locally compact and separable metrizable (or equivalently, homeomorphic to a closed subspace of Euclidean space). Such a space has $\dim(X) \leq n$ if and only if basic $(X) \leq 2n + 1$. Separable metrizable spaces either have finite basic (X) or basic (X) equal to the continuum. The value of basic (K) for a compact space K is closely connected with the cofinality of the countable subsets of a basis $\mathcal{B}$ for K of minimal size ordered by set inclusion.

It is proved that a space has a finite elementary family in every dimension m if and only if it is homeomorphic to a closed subspace of Euclidean space. It is further shown that there is a finite elementary family for the reals in each dimension m consisting of effectively computable functions, and effective procedures for representing any continuous function of m real variables as a superposition of these elementary functions and other univariate maps.

TABLE OF CONTENTS

1.0 INTRODUCTION 1
1.1 Basic and Elementary Families 4
1.2 The Problems 6
1.3 Solutions 6
2.0 SPACES WITH FINITE BASIC FAMILIES 10
2.1 Restrictions Induced by Generating Families 11
2.2 Construction of Finite Basic Families 14
3.0 MINIMAL SIZE OF BASIC FAMILIES 18
3.1 Minimal Size of Finite Basic Families 20
3.2 Separable Metrizable Spaces 20
3.3 Compact Spaces 27
4.0 HILBERT'S 13TH PROBLEM REVISITED 30
4.1 Superpositions 30
4.2 Characterization 38
4.3 An Application to C_p–Theory 39
5.0 CONSTRUCTIVE PROOF AND APPLICATIONS 40
5.1 Construction of the Functions 41
5.2 The Functions are Elementary 45
5.3 Neural Networks 50
6.0 OPEN QUESTIONS AND PROPOSED RESEARCH 53
6.1 Smooth Functions and Analytic Functions 53

6.2 Minimal Basic Families 54
6.3 Construction of Lipschitz Basic or Elementary Functions and Applications 55
APPENDIX A. HILBERT'S 13TH PROBLEM 56
APPENDIX B. PYTHON CODE 58
APPENDIX C. FUNCTION SPACE AND GENERALIZED METRIC PROPERTIES . 68
C.1 Introduction 68
C.2 Stratifiability 69
C.3 M_1 Property 73
C.4 Examples 74
Bibliography 76

LIST OF FIGURES

5.1 Neural Network 52

1.0 INTRODUCTION

The 13th Problem from Hilbert's famous list [16] asks (see Appendix A for the full text) whether every continuous function of three variables can be written as a superposition (in other words, composition) of continuous functions of two variables.

Hilbert motivated his problem from two rather different directions. First he explained that a positive solution would have applications in nomography. Nomography is the use of graphics to do calculations. Before the introduction of digital computers such graphical calculators were widespread, but nomography is now almost a forgotten art.

The second motivation Hilbert gave came from finding roots of polynomial equations. As is well known, polynomials of degree no more than four have roots obtained by applying the standard arithmetical operations along with taking nth roots, but Abel showed that the quintic can not be solved in radicals. However the general quintic equation, $x^5+a_4x^4+a_3x^3+a_2x^2+a_1x+a_0=0$, can be reduced by use of Tschirnhaus transformations, to a quintic equation, $y^5+b_1y+b_0=0$, where $x=x(y,a_4,\ldots,a_0)$ and $b_i=b_i(a_4,\ldots,a_0)$ can be computed using the arithmetical operations and roots. Thus roots of the general quintic can be calculated as a superposition of continuous functions of two or less variables, namely: arithmetical operations, roots and a two place function $y=y(b_1,b_0)$.

Tschirnhaus transformations also allow one to calculate the roots of the general sextic equation as a superposition of continuous functions of two or fewer variables. But applying Tschirnhaus transformations to the general septic equation apparently only reduces the equation to one in *three* parameters, $y^7+b_3y^3+b_2y^2+b_1y+1=0$. Hilbert felt that the difficulties encountered in trying to eliminate an additional coefficient were real — the root function $y=y(b_3,b_2,b_1)$ was *irreducibly* a continuous function of three variables, it could not be written as a superposition of continuous functions of two variables.

Hilbert, then, anticipated a negative answer to his 13th Problem, saying,

> "it is probable that the root of the equation of the seventh degree is a function of its coefficients which [...] cannot be constructed by a finite number of insertions of functions of two arguments. In order to prove this, the proof would be necessary that the equation of the seventh degree $f^7 + xf^3 + yf^2 + zf + 1 = 0$ is not solvable with the help of any continuous functions of only two arguments."

It took over 50 years for significant progress to be made on Hilbert's 13th Problem. Then in 1954 Vitushkin [37] found a result in the direction Hilbert expected: if $m/q > m'/q'$ then there are functions of m variables with all qth order derivatives continuous which can not be written as a superposition of functions of m' variables and all q'th order derivatives continuous. In particular, there are continuously differentiable functions of three variables which can not be written as a superposition of continuously differentiable functions of two variables.

However Kolmogorov and Arnold subsequently proved a series of wonderful, and justly famous, results culminating with Kolmogorov's Superposition Theorem (1957) [2, 4, 18]. (Here, and subsequently, I denotes $[0,1]$, and $C(X)$ denotes the set of continuous real valued functions on the topological space X.)

Kolmogorov's Superposition Theorem.

Step 1 *There exist $\phi_1 \dots, \phi_{2m+1}$ in $C(I^m)$ such that*

$$\forall f \in C(I^m) \quad f = \sum_{i=1}^{2m+1} g_i \circ \phi_i, \qquad \textit{for some } g_i \in C(I).$$

Step 2 *Further, one can choose the $\phi_1, \dots, \phi_{2m+1}$ such that:*

$$\phi_i(y_1, \dots y_m) = \sum_{j=1}^{m} \psi_{ij}(y_j), \qquad \textit{for some } \psi_{ij} \in C(I).$$

Notice that this result really does solve Hilbert's 13th Problem for functions of m variables from $[0, 1]$. The first Step in the theorem says that there are $2m + 1$ continuous functions on I^m (the ϕ_i) so that *every* continuous function on I^m can be simply obtained from these combined with addition and functions of one variable (the g_i). Now we see that Hilbert's 13th Problem has a positive solution if and only if these particular functions, the ϕ_i, of m variables can be written as a superposition of continuous functions of two or fewer variables. Step 2 assures us that is indeed possible, in fact only one function of two variables, addition, is necessary, the other functions (the ψ_{ij}) are functions of just one variable.

Thus every continuous function of m variables from I can be written as a superposition of functions of just one variable along with a single function of two variables, namely addition. This is truly astonishing! It is as if there are no continuous functions of m variables for $m > 1$, except addition, only continuous functions of one variable.

It is natural to wonder how far Kolmogorov's Superposition Theorem can be extended. How smooth can the 'inner' functions (the ψ_{ij}) be chosen? Given a suitably smooth function f how smooth can the 'outer' functions (the g_i) be selected? Is the number of inner functions minimal? Which spaces can be taken as the domain of the given function? An especially natural question in the latter direction is to ask whether the closed unit interval appearing in Kolmogorov's theorem can be replaced with the reals: can every continuous function of m real variables be written as a superposition of continuous functions of two real variables? Can this be done as in the Kolmogorov Superposition Theorem?

Indeed many extensions to Kolmogorov's theorem have been obtained. For example, Fridman showed that the inner functions can be taken to be Lipschitz [9]. That this is the best possible result in this direction follows from work of Vitushkin & Henkin, [15], who established results which imply that the inner functions can not be continuously differentiable. Sprecher established that the inner functions can all be taken to be scaled and translated versions of a single function [31]. The work of Sternfeld described in more detail below shows that the number of inner functions is indeed minimal.

Lorentz showed that the outer functions can be taken to be all equal, and observed that they can be chosen to be absolutely continuous [21].

This thesis is particularly concerned with the domain of the given function, so especially relevant here, is Ostrand's extension in [23] from functions on I^m to finite powers of finite–dimensional compact, metrizable spaces.

Ostrand's Theorem.

Step 1 *Let X be compact, metrizable and of dimension n.*

Then there exist $\phi_1 \ldots, \phi_{2n+1}$ in $C(X)$ such that

$$\forall f \in C(X) \quad f = \sum_{i=1}^{2n+1} g_i \circ \phi_i, \qquad \textit{for some } g_i \in C(I).$$

Step 2 *If $X = Y^m$ one can choose the $\phi_1, \ldots, \phi_{2n+1}$ such that:*

$$\phi_i(y_1, \ldots y_m) = \sum_{j=1}^{m} \psi_{ij}(y_j), \qquad \textit{for some } \psi_{ij} \in C(Y)$$

1.1 BASIC AND ELEMENTARY FAMILIES

Following Sternfeld, let us isolate the behavior of the families of functions ϕ_i and ψ_{ij} appearing in the Superposition Theorems of Kolmogorov and Ostrand. (Here, and below, unless otherwise stated a 'space' is a Tychonoff topological space, and $C^*(X)$ denotes the subset of $C(X)$ consisting of bounded functions.)

Definition 1. *Let X be a space. A family $\Phi \subseteq C(X)$ is said to be* **basic** *(respectively,* **basic****) for X if each f in $C(X)$ (respectively, $C^*(X)$) can be written: $f = \sum_{q=1}^{n} (g_q \circ \phi_q)$,*

for some $\phi_1, \cdots, \phi_n$ in Φ and 'co-ordinate functions' $g_1, \ldots, g_n \in C(\mathbb{R})$.

Note that Step 1 of Ostrand's theorem can now be restated as saying that every compact metrizable space of dimension n has a basic family of size $2n + 1$. In the first step beyond compact domains, and so having to deal with *unbounded* continuous functions, Doss [6] showed that $\mathbb{R}^n$ has a basic family of size $4n$.

Beyond their intrinsic interest, basic functions have proved to be widely useful. Since the use of basic functions reduces calculations of functions simply to addition and evaluation of a fixed finite family of functions, applications to numerical analysis, approximation and function reconstruction are immediately apparent. Also other applications have emerged including to neural networks.

In probably the deepest work on Hilbert's 13th Problem following Kolmogorov's Theorem, Sternfeld [34] (a significantly shorter proof is given by Levin in [20]) showed that if X is a compact, metrizable space of dimension $n \geq 2$, then X does not have a basic family of size $\leq 2n$. (In dimension 1, the minimal size of a basic family can be one ($X = I$), two ($X =$ the tripod) or 3 (the circle). There is no characterization of which one dimensional compact metrizable spaces need precisely two basic functions, but much is known from the work of Sternfeld [34], and Skopenkov [29].) Combining this with Ostrand's Theorem gives a characterization of the dimension of compact metrizable spaces.

Theorem 2. *Let $n \geq 1$, and let X be a compact metrizable space.*

Then $\dim X \leq n$ if and only if X has a basic family with $\leq 2n+1$ members.

To deal with the inner functions from Kolmogorov's and Ostrand's theorems we make the following definitions. For maps $\psi_1, \psi_2, \ldots, \psi_m \in C(X)$, define $\Sigma = \Sigma(\psi_1, \psi_2, \ldots, \psi_m) : X^m \to \mathbb{R}$ by $\Sigma(x_1, x_2, \ldots, x_m) = \sum_{p=1}^{m} \psi_p(x_p)$.

Definition 3. *A family Ψ_m contained in $C(X)$ is* **elementary** *in dimension m if the family of maps $\Phi_m = \{\Sigma(\psi_1, \psi_2, \ldots, \psi_m) : \psi_1, \ldots, \psi_m \in \Psi_m\}$ is basic for X^m.*

Hence Step 2 of Ostrand's Theorem is essentially equivalent to saying that a compact metrizable space of dimension n has an elementary family of size $nm(2m+1)$ in dimension m.

1.2 THE PROBLEMS

The following questions and problem arise naturally from the discussion above of the Arnold–Kolmogorov solution of Hilbert's 13th Problem, Kolmogorov's Superposition Theorem, and the subsequent work of Fridman, Ostrand, Doss, Sternfeld and others. They have all been raised either in full, or in part, by numerous authors.

Question A Which spaces have a finite basic family?

Question B Given a space X, what is the minimal size of a basic family for X?

Question C Which spaces have a finite elementary family in every (or some) dimension?

In particular, does the real line have a finite elementary family in every dimension?

Kolmogorov's Superposition Theorem promises much to numerical analysis, in principle converting frequently intractable multivariate problems into ones involving only univariate functions and addition. However the proof of Kolmogorov's Theorem is highly non–constructive. Only very recently (2007, published 2009) have Braun & Griebel [10] given a rigorous truly constructive version of the Kolmogorov Superposition Theorem.

Problem D In those cases where finite basic or elementary families can be shown to exist, find *constructive* versions, and explore applications.

1.3 SOLUTIONS

Theorem A. *Let X be a space. Then the following are equivalent:*

(1) X has a countable basic family,

(2) X has a finite basic family, and

(3) X is a finite dimensional, locally compact and separable metrizable, or equivalently, is homeomorphic to a closed subspace of Euclidean space.

The proof of Theorem A is given in Chapter 2, where, in fact, a stronger result will be proved where in (1) 'countable basic family' is weakened to 'countable generating* family'. This theorem gives strong and complete solutions to Problems 10, 11 of Sternfeld [34] and questions of Hattori [11], among others.

In order to investigate the minimal size of basic families of a given space, we introduce a new cardinal invariant.

Definition 4. *Let X be a space. Define basic* $(X) = \min\{|\Phi| : \Phi$ *is a basic family for* $X\}$.

Theorem B.

1. *For any space X, basic $(X) \leq 2n+1$ if and only if X is locally compact, separable metrizable and has* $\dim X \leq n$.
2. *For X separable metrizable, then either basic (X) is finite or basic* $(X) = \mathfrak{c}$.
3. *For X compact, then basic $(X) \leq cof([w(X)]^{\aleph_0}, \subseteq)$,*

   and if X contains a discrete subspace D such that $|D| = w(X)$ then:

   either X is finite dimensional, and basic $(X) = cof([w(X)]^{\aleph_0}, \subseteq)$,

   or X is infinite dimensional, and basic $(X) = |C(X)| = w(X)^{\aleph_0}$.

Theorem B is proved in Chapter 3.

Part 1 above answers Problems 12, 13 of Sternfeld [34]), and questions of Hattori, Doss and others [6, 11].

The interest in Part 2 lies in the fact that the dichotomy 'basic (X) is either finite or the continuum, $\mathfrak{c}$' is true *in ZFC* (the standard axioms of set theory), and does not assume the Continuum Hypothesis (for example). Any experienced Set Theorist or Set Theoretic Topologist would find this absoluteness of basic (X), when X is separable metrizable, quite unexpected.

Part 3 yields considerable, if not complete, information on the possible values of basic (X) when X is compact. Note that the 'weight', $w(X)$, of a space X is the minimal size of a basis for X. Further, for a set S, $[S]^{\aleph_0}$ is the set of countably infinite subsets of S, and $cof([S]^{\aleph_0}, \subseteq)$ is the minimal size of a cofinal family in $[S]^{\aleph_0}$ partially ordered by set containment. This leads to some intriguing connections with Shelah's *Potential Cofinalities Theory* (PCF), these are outlined in Chapter 3.3.

Theorem C. *Let X be a space. Then the following are equivalent:*

(1) some power of X has a finite basic family

(2) X has a finite elementary family in some dimension

(3) X has finite elementary families in every dimension

(4) for every $m, n \in \mathbb{N}$, there is an $r \in \mathbb{N}$ and ψ_{pq} from $C(X, \mathbb{R}^n)$, for $q = 1, \ldots, r$ and $p = 1, \ldots, m$, such that every $f \in C(X^m, \mathbb{R}^n)$ can be written

$$f(x_1, \ldots, x_m) = \sum_{q=1}^{r} g\left(\sum_{p=1}^{m} \psi_{pq}(x_p)\right),$$

for some $g \in C(\mathbb{R}^n, \mathbb{R}^n)$;

(5) X is a locally compact, finite dimensional separable metric space, or equivalently, homeomorphic to a closed subspace of Euclidean space.

This Theorem encapsulates our strengthening of the Arnold–Kolmogorov solution to Hilbert's 13th Problem — the answer is, from (5) implies (3), *yes*, any continuous real–valued function of three real variables can be written as a superposition of continuous functions of two or fewer variables. In fact $\mathbb{R}$ has elementary families in every dimension such that only a single co–ordinate function is required and such that the elementary functions are Lipschitz, with Lipschitz constant 1.

From '(5) implies (4)' we know this remains true if we consider complex–valued functions and functions of complex variables. This includes, of course, the solution function f to the septic equation $f^7 + xf^3 + yf^2 + zf + 1 = 0$.

The equivalence of (5) and (3) characterizes those spaces which satisfy a Superposition Theorem of the Kolmogorov type. Theorem C is established in Chapter 4.

Theorem D. *Let m be a natural number. Take any $\gamma \geq 2m + 2$, and let $\mathcal{D} = \{k/\gamma^\ell : k, \ell \in \mathbb{Z}\}$ be the set of all rationals base γ.*

Then $\mathbb{R}^m$ has an elementary family, ψ_{ij} where $i = 1, \ldots, 2m+1$ and $j = 1, \ldots, m$, which are defined constructively *(in fact, recursively) on $\mathcal{D}$.*

Let $\phi_i(x_1, \ldots, x_m) = \sum_j \psi_{ij}(x_j)$. Then, given $f \in C(\mathbb{R}^m)$, there is a constructive algorithm *for computing $g_1, \ldots, g_{2m+1}$ in $C(\mathbb{R})$ such that $f = \sum g_i \circ \phi_i$, to within a specified error $\epsilon > 0$ on any specified compact subset K of $\mathbb{R}^m$.*

Theorem D is proved in Chapter 5. We note that we use the term 'algorithm' a little loosely (as is standard in numerical analysis) to mean a 'procedure' or 'process'. However it would be straightforward to rephrase the algorithms in Chapter 5 to be algorithms in the sense of Blum–Cucker–Shub–Smale *Complexity and Real Computation*, [30], or Weihrauch's *Computable Analysis*, [39]. Computer code in the high level language *Python* implementing the algorithms of Theorem D is given in Appendix B. Some comments on applications to neural networks are given in Chapter 5.3.

Appendix C deals with earlier work of the author on the generalized metric properties of function spaces.

2.0 SPACES WITH FINITE BASIC FAMILIES

This chapter is devoted to proving Theorem 6, which is a strengthening of Theorem A.

To facilitate the proof, and provide full generality we make the following definition allowing more general superposition representations than a 'basic' representation.

Definition 5. *Let X be a space. A family $\Phi \subseteq C(X)$ is said to be* **generating** *(respectively,* **generating****) for X with respect to a 'set of operations' M of continuous functions mapping from subsets of Euclidean space into subsets of Euclidean space, if each $f \in C(X)$ (respectively, $C^*(X)$) can be written as a composition of functions from Φ, M and $C(\mathbb{R})$.*

Clearly a basic family of functions is generating, a basic* family is generating*, and a generating family is generating*.

Theorem 6. *Let X be a space. Then the following are equivalent:*

(1) X has a countable generating* family,

(2) X has a finite basic family, and

(3) X is a finite dimensional, locally compact and separable metrizable, or equivalently, is homeomorphic to a closed subspace of Euclidean space.

In Theorem 6, (2) $\implies$ (1) is immediate. In the next section (Section 2.1) we prove (1) $\implies$ (3), and then in Section 2.2 we establish (3) $\implies$ (2).

2.1 RESTRICTIONS INDUCED BY GENERATING FAMILIES

Lemma 7. *Let X have a generating* family Φ with respect to M. Then $e : X \to \mathbb{R}^\Phi$ defined by $e(x) = (\phi(x))_{\phi \in \Phi}$ is an embedding.*

Proof. Clearly e is continuous (each projection is a ϕ in Φ which is continuous). It is also easy to see e is injective. Take distinct x, x' in X. Pick $f \in C^*(X)$ such that $f(x) = 0$, $f(x') = 1$. Represent f as a composition of $\phi_1, \ldots, \phi_n$ in Φ, members of M and $C(\mathbb{R})$. If $e(x) = e(x')$ then $\phi_i(x) = \phi_i(x')$ for all i, and so $f(x) = f(x')$, which is a contradiction.

It remains to show that the topology induced on X by e contains the original topology. Since X is completely regular it is sufficient to check that for every $f \in C^*(X)$ the map $e(f) : e(X) \to \mathbb{R}$ defined by $e(f)(\mathbf{x}) = f(e^{-1}(\mathbf{x}))$ is continuous. But each $f \in C^*(X)$ can be written as a composition of some $\phi_1, \ldots, \phi_n$ in Φ and members of M and $C(\mathbb{R})$. Note that for each i we have $\phi(e^{-1}(\mathbf{x})) = \pi_{\phi_i}(\mathbf{x})$, where π_{ϕ_i} is the projection map of $\mathbb{R}^\Phi$ onto the ϕ_ith co-ordinate. Hence $e(f) = f \circ e^{-1}$ is the composition of continuous maps, namely the π_{ϕ_i}s and functions in M and $C(\mathbb{R})$, and so is continuous as required. □

Since any subspace of $\mathbb{R}^\mathbb{N}$ is separable metrizable and any subspace of $\mathbb{R}^n$ is finite dimensional, we deduce from Lemma 7:

Corollary 8.

a) A space with a countable generating* family is separable metrizable.

b) A space with a finite generating* family is finite dimensional.

A subspace C of a space X is said to be C^*-embedded in X if every $f \in C^*(C)$ can be extended to a continuous bounded real valued function on X. In a normal space all closed subspaces are C^*-embedded. Compact subspaces are always C^*-embedded. We note the following easy lemma:

Lemma 9. *If Φ is a generating* (respectively, basic*) family for a space X with respect to M, and C is C^*-embedded in X then $\Phi|C = \{\phi|C : \phi \in \Phi\}$ is a generating* (respectively, basic*) family for C.*

Lemma 10. *A space with a countable generating* family is locally compact.*

Proof. Suppose the space X has a countable generating* family Φ with respect to M, but is not locally compact. Since X is metrizable, it follows that the metric fan F (defined below) embeds as a closed subspace in X. Hence by Lemma 9 it suffices to show that F does not admit a countable generating* family (with respect to any set of operations M).

The metric fan F has underlying set $\{*\} \cup (\mathbb{N} \times \mathbb{N})$ and topology in which all points other than $*$ are isolated and $*$ has basic neighborhoods $B(*, N) = \{*\} \cup ([N, \infty) \times \mathbb{N})$. For a contradiction, let $\Phi = \{\phi_1, \phi_2, \ldots\}$ be a countable generating* family with respect to M.

For each i, let $y_i = \phi_i(x_0)$, and pick basic open U_i containing $*$ such that $\phi_i(U_i) \subseteq (y_1 - 1, y_1 + 1)$. Now for each n let $V_n = \bigcap_{i=1}^n U_i$. So $\phi_i(V_n) \subseteq (y_i - 1, y_i + 1)$ for $i = 1, \ldots, n$. We can write $V_n = \{*\} \cup ([N_n, \infty) \times \mathbb{N})$ and suppose, without loss of generality, that $N_n > N_m$ if $n > m$.

Fix n. Let $D^0 = \{x_k^0 = (N_n, k) : k \in \mathbb{N}\}$. As $\{\phi_1(x_k^0)\}_{k\in\mathbb{N}}$ is a subset of $[y_1 - 1, y_1 + 1]$, which is sequentially compact, there is a $D^1 = \{x_k^1 : k \in \mathbb{N}\} \subseteq D^0$ such that $\{\phi_1(x_k^1)\}_{k\in\mathbb{N}}$ is convergent. As $\{\phi_2(x_k^1)\}_{k\in\mathbb{N}}$ is a subset of $[y_2 - 1, y_2 + 1]$, which is sequentially compact, there is a $D^2 = \{x_n^2 : n \in \mathbb{N}\} \subseteq D^0$ such that $\{\phi_2(x_k^2)\}_{k\in\mathbb{N}}$ is convergent. Inductively we get $D^n = \{x_k^n : k \in \mathbb{N}\}$, which is infinite closed discrete and for each $i = 1, ..., n$ the sequence $\{\phi_i(x_k^n)\}_{k\in\mathbb{N}}$ is convergent, say to z_i^n. Define $D_O^n = \{x_{2k-1}^n : k \in \mathbb{N}\}$ and $D_E^n = \{x_{2k}^n : k \in \mathbb{N}\}$.

Define $f : F \to [0, 1]$ by: f is identically zero outside $\bigcup_n D_O^n$ (in particular, f is zero on each D_E^n), and f is identically $1/n$ on D_O^n. Then f is continuous and bounded.

Hence, for some ℓ, f can be written as the composition of $\phi_1, \ldots, \phi_\ell$ and members of M and $C(\mathbb{R})$. Now, on the one hand $\lim_k \phi_i(x_{2k-1}^\ell) = z_{i.\ell} = \lim_k \phi_i(x_{2k}^\ell)$ so by continuity of the elements of M and $C(\mathbb{R})$ in the compositional representation of f, $\lim_k f(x_{2k-1}^\ell) = \lim_k f(x_{2k}^\ell)$, and on the other hand, $\lim_k f(x_{2k-1}^\ell) = 1/\ell \neq 0 = \lim_k f(x_{2k}^\ell)$. This is our desired contradiction. □

Let Y be a locally compact separable metrizable space. Write $C_k(Y)$ for $C(Y)$ with the compact-open topology. Then $C_k(Y)$ is a Polish (separable, completely metrizable) group. In particular, for any n, $C_k(\mathbb{R})^n$ is a Polish group.

Lemma 11. *If X has a countable generating* family with respect to a countable set of operations, M, then X has a finite generating* family with respect to a finite set of operations M'.*

Proof. Let $\phi_1, \phi_2, \ldots$ be a countable generating* family for X with respect to the countable set of operations M. By Lemma 10 X is locally compact and $C_k(X)$ is a Polish group.

Let $g_1, g_2, \ldots$ be formal letters representing functions from $\mathbb{R}$ to $\mathbb{R}$. Let $\mathcal{W}$ be the set of all formal compositions of ϕ_is, elements of M and g_is. Note that $\mathcal{W}$ is countable.

Fix w in $\mathcal{W}$. Then w induces a map $(g_1, \ldots, g_n) \mapsto w(g_1, \ldots, g_n)$ from $C_k(\mathbb{R})^n \to C_k(X)$ where we substitute actual $g_i \in C(\mathbb{R})$ for the corresponding formal letter. This map is continuous with respect to the compact-open topology. Let $F_w = w(C_k(\mathbb{R})^n)$. It is analytic. Define $G_w = F_w \cap C_k(X, (0,1))$. Since $C_k(X, (0,1))$ is homeomorphic to $C_k(X)$ it is Polish, and hence must be a G_δ subset of $C_k(X)$. So G_w is analytic in $C_k(X, (0,1))$.

Note, by the generating* property, that $C_k^*(X) \subseteq \bigcup_{w \in \mathcal{W}} F_w$. Hence $C_k(X, (0,1)) = \bigcup_{w \in \mathcal{W}} G_w$. By the Baire Category Theorem there must be some particular w in $\mathcal{W}$ such that G_w is not meager.

Fix a homeomorphism $h : \mathbb{R} \to (0,1)$. Via h, addition and subtraction on $\mathbb{R}$ induce (continuous) group operations $\oplus, \ominus : (0,1) \times (0,1) \to (0,1)$. These operations on $(0,1)$ in turn induce operations on $C_k(X, (0,1))$ making this space a Polish group.

Let H_w be the subgroup of $C_k(X, (0,1))$ generated by G_w. By Pettis' Theorem [26], since G_w is non-meager and analytic, $G_w \ominus G_w$ has non-empty interior. Hence the subgroup H_w is open, and so coincides with $C_k(X, (0,1))$ (which is connected).

Set Φ' to be the finite set of ϕ_is appearing in w, and set M' to be $\oplus, \ominus$ and the finite set of elements of M appearing in w. Since $H_w = C(X, (0,1))$, each element of $C(X, (0,1))$ is a composition of members of Φ', M' and $C(\mathbb{R})$.

We check Φ' is a finite generating* family with respect to M'. For if $f \in C^*(X)$, then f maps into some open interval (a, b). Fix a homeomorphism $g_0 : \mathbb{R} \to \mathbb{R}$ taking $(0,1)$ to (a, b). Then $f = g_0 \circ (g_0^{-1} \circ f)$, where $g_0^{-1} \circ f$ is in $C_k(X, (0,1))$. Hence $g_0^{-1} \circ f$ can be expressed as a composition of elements of Φ', M' and some $g_1, \ldots, g_n$ in $C(\mathbb{R})$. But now f is g_0 of this composition and so is also expressible in terms of elements of Φ', M' and $C(\mathbb{R})$, as required. $\square$

We note that the finite generating* family is a subset of the original family, and also that if the original family is generating then we can take $M' \subseteq M \cup \{+, -\}$.

Proof of (1) $\implies$ (3) in Theorem 6. Let X be a space with a countable generating* family with respect to a countable set of operations. By Corollary 8 a) X is separable metrizable. Lemma 10 then says that X is locally compact. From Lemma 11 we deduce that X has a finite generating* family. Hence by Corollary 8 b) X is finite dimensional. □

2.2 CONSTRUCTION OF FINITE BASIC FAMILIES

This section is devoted to proving:

Lemma 12. *If X is a locally compact, separable metrizable space of dimension $\leq n$ then X has a basic family of size $2n+1$.*

The implication '(3) $\implies$ (2)' of Theorem 6 then follows.

We should note the following prior work. Doss extended the first step in Kolmogorov's Theorem to the non-compact case, by showing that $\mathbb{R}^n$ has a finite basic family of size $4n$ [6]. While Hattori [11] showed that every locally compact, separable metrizable space X of dimension n has a finite basic* family of size $2n+1$.

Lemma 12 and its proof improves on Doss' and Hattori's results and proof because: (1) it applies to all functions (not necessarily bounded) on any locally compact, separable metrizable finite–dimensional space (not just $\mathbb{R}^n$), (2) it gives the minimal number of basic functions (Doss does not), (3) it is somewhat constructive (Hattori's argument uses a Baire category argument) and (4) it is considerably shorter than Hattori's. The proof is similar to that of Ostrand for *compact* metric spaces. However difficulties arise because continuous real valued functions on a locally compact space need not be *bounded*.

For this section, fix a locally compact, separable space X of dimension $\leq n$, and with compatible metric d. We can find $\{K_b : b \geq -1\}$ a countable cover of X by compact sets such that $K_{-1} = K_0 = \emptyset$ and $K_b \subseteq K_{b+1}^\circ$ for each $b \geq -1$. For each $b \geq 0$ we put $H_b = K_b \setminus K_{b-1}^\circ$, and set $U_b = K_{b+1}^\circ \setminus K_{b-1}$. Since Ostrand has done the compact case, we can assume that the K_b's are *strictly* increasing. We show X has a basic family of size $2n+1$.

The basic functions ϕ_i are defined to be the limit of approximations f_k^i. The approximations are defined inductively along with some families of 'nice' covers. These 'nice' covers come from Ostrand's Dimension Theorem.

Ostrand's Dimension Theorem. *A metric space X is of dimension $\leq n$ if and only if for each open cover $\mathcal{C}$ of X and each integer $k \geq n+1$ there exist k discrete families of open sets $\mathcal{U}_1, \cdots, \mathcal{U}_k$ such that the union of any $n+1$ of the $\mathcal{U}_i$ is a cover of X which refines $\mathcal{C}$.*

Lemma 13. *Let $\gamma > 0$. There are $2n+1$ many families $\mathcal{S}^1, \ldots, \mathcal{S}^{2n+1}$ of open subsets of X, and $\eta^b > 0$ for $b \geq 0$, satisfying:*

(1) Each $\mathcal{S}^i$ is discrete in X.

(2) For k fixed and each $x \in X$ fixed, $|\{S \in \bigcup_{i=1}^{2n+1} \mathcal{S}^i : x \in S\}| \geq n+1$.

(3) diam $S < \gamma$ for any $S \in \bigcup_{i=1}^{2n+1} \mathcal{S}^i$.

(4) $\bigcup_{i=1}^{2n+1} \mathcal{S}^i$ refines $\{U_b : b \in \omega\}$.

(5) For any $b \in \mathbb{N}$, $\{S : S \in \bigcup_{i=1}^{2n+1} \mathcal{S}^i, S \cap K_b \neq \emptyset\}$ is finite.

(6) $S(H_b, \eta^b) \cap S = \emptyset$ if $H_b \cap \overline{S} = \emptyset$ for any $S \in \bigcup_{i=1}^{n+1} \mathcal{S}^i$.

(7) $\overline{S(H_{b-1}, \eta^{b-1})} \cap \overline{S(H_{b+1}, \eta^{b+1})} = \emptyset$.

In (6) and (7), $S(H_b, \eta^b) = \{x \in X : d(H_b, x) \leq \eta^b\}$

Proof. Let $\mathcal{C} = \{C_a : a \in \mathbb{N}\}$ be a locally finite open cover of X with: diam $(C_a) < \gamma$ and $|\{H_b : H_b \cap \overline{C_a} \neq \emptyset\}| \leq 2$, for each $a \in \mathbb{N}$. Then by Ostrand's covering theorem, there exist $2n+1$ discrete families of open sets $\mathcal{S}_1, \cdots, \mathcal{S}_{2n+1}$ which refines $\mathcal{C}$. Also the union of any $n+1$ of the $\mathcal{S}_i$ is a cover of X. So 1)-4) are easy to verify.

Fix i with $1 \leq i \leq 2n+1$. As $\mathcal{S}^i$ is discrete, $\{S : S \cap K_b \neq \emptyset, S \in \mathcal{S}^i\}$ is finite. Thus condition 5) is satisfied.

Now fix i and b, the discreteness of $\mathcal{S}^i$ guarantees that

$$H_b \cap \overline{\bigcup \{S : S \in \mathcal{S}^i \text{ and } H_b \cap \overline{S} = \emptyset\}} = \emptyset.$$

So $d(H_b, \overline{\bigcup \{S : S \in \mathcal{S}^i \text{ and } H_b \cap \overline{S} = \emptyset\}}) > 0$. Then we can pick η_i^b such that $S(H_b, \eta_i^b) \cap S = \emptyset$ if $H_b \cap \overline{S} = \emptyset$ for any $S \in \mathcal{S}^i$. Let $\eta^b = \min\{\eta_i^b : i = 1, \cdots, 2n+1\}$. This satisfies 6).

Notice that since H_b is compact for each $b \in \mathbb{N}$, we can pick η^b small enough such that $\overline{S(H_{b-1}, \eta^{b-1})} \cap \overline{S(H_{b+1}, \eta^{b+1})} = \emptyset$, giving (7). □

Proof. (Lemma 12) **Step 1: Construction of the approximations**

Again, we generalize the construction of Ostrand, but must find ways around the problem of not having *bounded* functions.

By induction on $k \geq 0$, using Lemma 13, for $i = 1, ..., 2n + 1$, there exist: positive real numbers ϵ_k with $\epsilon_1 < 1/4$, γ_k, η_k^b distinct positive prime numbers r_k^i, discrete families $\mathcal{S}_k^1, ..., \mathcal{S}_k^{2n+1}$ and continuous functions $f_k^i : X \to [0, k+1]$, with the following properties.

For each $k \in \mathbb{N}$, the families $\mathcal{S}_k^1, ..., \mathcal{S}_k^{2n+1}$, γ_k and η_k^b satisfy (1)–(7) of Lemma 13. Further:

(A) $\lim_{k\to\infty} \gamma_k = \lim_{k\to\infty} \epsilon_k = 0$;

(B) $\epsilon_k < 1/\Pi_{i=1}^{2n+1} r_k^i$;

(C) f_k^i is constant on the closure of those members of $\mathcal{S}_k^i$ which have nonempty intersection with K_b for $(b \leq k)$, the constant being an integral multiple of $1/r_k^i$, and takes different values on distinct members. Then we can take a continuous extension of f_k^i to the rest of the space.

(D) For any S in $\mathcal{S}_k^i$ having nonempty intersection with H_b, $b - 1 < f_b^i(S) < b + 1$. Also for $b \geq 2$, by (7), we can make $b - 1 < f_k^i(S(H_b, \eta_k^b) < b + 1$. For each $i \in \mathbb{N}$, if $\overline{S} \cap H_b \neq \emptyset$ and $\overline{S} \cap H_{b+1} \neq \emptyset$, then $b < f_b^i(C) < b+1$;if $\overline{S} \cap H_b \neq \emptyset$ and $\overline{S} \cap H_{b-1} \neq \emptyset$, then $b - 1 < f_b(S)^i < b$;

(E) For each $\ell < j < k$ and $x \in K_\ell$, $f_j^i(x) < f_k^i(x) < f_j^i(x) + \epsilon_j - \epsilon_k$ for any i.

Step 2: Construction of the basic functions

From (E), for any $x \in K_b$ and $k > b$, $f_b^i(x) < f_k^i(x) < f_b^i(x) + \epsilon_1$ for any $i = 1, \ldots, 2n+1$. Thus we can take the uniform limit of f_k^i restricted on K_b. For any $x \in K_b$ let $\phi_i(x) = \lim_{k\to\infty} f_k^i(x)$. So ϕ_i is continuous on K_b for each b. Hence ϕ_i is continuous on X. Also by (D) for $x \in H_b$, $b - 1 < \phi_i(x) < b + 1 + 1/4$.

Let $\mathcal{V}_k^i = \{\phi_i(S) : S \in \mathcal{S}_k^i\}$. Then if $S \cap K_b \neq \emptyset$ and $S \in \mathcal{S}_k^i$ with $k > b$, $\phi_i(S)$ is contained in the interval $[f_k^i(S), f_k^i(S) + \epsilon_k]$ by (E). By (B), these closed intervals are disjoint for each fixed b and k with $k \geq b$. Then each $\mathcal{V}_k^i$ is discrete.

Step 3: Construction of the coordinate functions

Take any function $f \in C(X)$. We find $g_1, \ldots, g_{2n+1} \in C(\mathbb{R})$ such that $f = \sum_{i=1}^{2n+1} g_i \circ \phi_i$.

For each $s \geq 0$, define the compact subset $L_s = K_{s+1} \setminus K_{s-1}^\circ$. Since K_1 is compact and $K_1 \subseteq K_2^\circ$, there exists a function f_1 such that $f_1(x) = f(x)$ for $x \in K_1$ and $f_1(x) = 0$ for $x \in X \setminus K_2^\circ$. Then letting $g_1 = f - f_1$, it is easy to see that $g_1(x) = 0$ for $x \in K_1$. Similarly, there exists f_2 such that $f_2(x) = g_1(x)$ for $x \in K_2$ and $f_2(x) = 0$ for $x \in X \setminus K_3^\circ$. Inductively, f can be written as an infinite sum $\sum_{s=1}^{\infty} f_s$ such that $f_s(x) = 0$ for $x \in X \setminus L_s$.

For each s, f_s is bounded and uniformly continuous. Fix $s \in \mathbb{N}$. Note that for each $x \in L_s$, $s - 2 < \phi_i(x) < s + 2 + 1/4$.

By construction, if we restrict the discrete families $\mathcal{S}_1, \cdots, \mathcal{S}_{2n+1}$ and the functions $\phi_1, \cdots, \phi_{2n+1}$ to K_{s+1}, then the discrete families and functions are exactly those defined by Ostrand [23].

In particular, the functions $\phi_1|L_s, \ldots, \phi_{2n+1}|L_s$ are *basic* for L_s (Lemma 2.1). Thus we can represent $f_s|L_s(x) = \sum_{i=1}^{2n+1} g_i^s(\phi_i|L_s(x))$, for some $g_i^s \in C(\mathbb{R})$. We can redefine g_i^s to be constantly zero outside of $[s-2, s+2+1/4]$ because the image of ϕ_i is contained in $[s-2, s+2+1/4]$ and $f_s(x) = 0$ if $x \in L_s \setminus (L_s)^\circ$. Now $f_s = \sum_{i=1}^{2n+1} g_i^s \circ \phi_i$.

Finally, letting $g_i = \sum_{s=1}^{\infty} g_i^s$, we see that g_i is continuous because $g_i(x)$ is a finite sum of non-zero continuous functions for each $x \in \mathbb{R}$, and $f = \sum_{i=1}^{2n+1} g_i \circ \phi_i$ – as required. $\square$

3.0 MINIMAL SIZE OF BASIC FAMILIES

In this chapter we investigate the minimal size of basic families in a given space. In the process we prove Theorem B (and more) from the Introduction.

The question of the minimal size of finite basic families is considered in Section 3.1.

Then we turn to the case when a space does not have a finite basic family. Since the natural map of X into $\mathbb{R}^\Phi$ is an embedding when Φ is a basic family (Lemma 7) a simple restriction on the size of basic families is: $w(X) \leq \text{basic}\,(X).\aleph_0 \leq |C(X)|$. So further natural questions are: when is basic $(X) \leq w(X)$? when is basic $(X) = |C(X)|$? is it possible to have basic (X) strictly between $w(X)$ and $|C(X)|$?

In this chapter we consider these questions for *separable metrizable* spaces (Section 3.2) and *compact* spaces (Section 3.3). Suppose first that X is separable metrizable. Then from Theorem 6, either basic (X) is finite, and this happens if and only if X is locally compact and finite dimensional, or $\aleph_1 \leq \text{basic}\,(X) \leq \mathfrak{c} = |C(X)|$. Experience of other related cardinal invariants of separable metrizable spaces would suggest that basic (X) should be undetermined by the standard axioms of set theory (ZFC). For example $k(X)$, which is the minimal size of a cofinal family in the set of all compact subsets of X, is undetermined even when X is the rationals or the irrationals. However (Theorem 17) basic (X) *is* determined in ZFC for all separable metrizable X:

either X is locally compact and finite dimensional, and basic $(X) < \aleph_0$,

or X is either infinite dimensional or not locally compact, and basic $(X) = \mathfrak{c}$.

This theme — that basic (X) is remarkably absolute — is continued when we consider compact spaces. Note that if K is compact, then Stone [35] has shown that $|C(K)| = w(K)^{\aleph_0}$. Hence, basic (K) lies between the weight of K and the countable power of the weight. This leads to some intriguing connections with Shelah's *Potential Cofinalities Theory* (PCF).

Let κ be an uncountable cardinal. Shelah observed that $\kappa^{\aleph_0} = cof([\kappa]^{\aleph_0}, \subseteq) \times |\mathbb{P}(\aleph_0)|$. (Here $cof([\kappa]^{\aleph_0}, \subseteq)$ is the minimal size of a cofinal set in the countably infinite subsets of κ ordered by inclusion.) If $\kappa = \aleph_n$ for $n \in \mathbb{N}$, then $cof([\kappa]^{\aleph_0}, \subseteq) = \kappa$, and so $\kappa^{\aleph_0}$ is easily computed — it is $\max(\kappa, \mathfrak{c})$.

However, if κ has countable cofinality then Shelah has shown [28] that interesting things happen. Whereas the value of $|\mathbb{P}(\aleph_0)| = \mathfrak{c}$ is almost entirely unconstrained by the axioms of set theory and can be made arbitrarily large, $cof([\kappa]^{\aleph_0}, \subseteq)$ seems to be almost absolute. For example $\aleph_\omega < cof([\aleph_\omega]^{\aleph_0}, \subseteq) < \aleph_{\omega_4}$, and making $cof([\aleph_\omega]^{\aleph_0}, \subseteq) > \aleph_{\omega+1}$ requires large cardinals.

We prove (Theorem 26) that if K is compact and finite dimensional then basic $(K) \leq cof([w(K)]^{\aleph_0}, \subseteq)$, and deduce (Theorem 28) that if K is suitably 'nice' (contains a discrete subset D with $|D| = w(K)$) then

either K is finite dimensional, and basic $(K) = cof([w(K)]^{\aleph_0}, \subseteq)$,

or K is infinite dimensional, and basic $(K) = |C(K)| = w(K)^{\aleph_0}$.

This gives a lot of information on the possible values of basic (K) for compact K. These are teased out and examples given below.

It is also interesting to note that if K is compact, finite dimensional, 'nice' and of weight κ (for example, $K = 2^\kappa$), and if Φ is a basic family for K of minimal size, then $C(K) \sim \bigcup_{n\in\mathbb{N}} (\Phi^n \times C(\mathbb{R})^n)$ is a natural 'topological realization' of the cardinal identity $\kappa^{\aleph_0} = cof([\kappa]^{\aleph_0}, \subseteq) \times |\mathbb{P}(\aleph_0)|$.

Finally we briefly discuss connections of the above results with Banach algebras. Let K be a compact space. Then $C(K)$ with the supremum norm is a Banach algebra. Sternfeld has observed that for any $\phi \in C(K)$ the set $L(\phi) = \{g \circ \phi : g \in C(\mathbb{R})\}$ is a closed subring of $C(K)$ containing the constants and generated by a single element, and conversely every closed subring with these properties is of the form $L(\phi)$ for some ϕ in $C(K)$.

Thus saying that basic $(K) \leq \kappa$ is the same as saying that $C(K)$ is the sum of no more than κ closed subrings containing the constants and generated by a single element. So the results above imply that the problem of deciding whether the Banach algebra $C(K)$ can be written as a sum of a certain size of 'small' closed subrings is closely linked to $cof([w(K)]^{\aleph_0}, \subseteq)$ and PCF theory.

3.1 MINIMAL SIZE OF FINITE BASIC FAMILIES

Theorem 14. *For any space X, basic $(X) \leq 2n+1$ if and only if X is locally compact, separable metrizable and has $\dim X \leq n$.*

Note that this is Theorem B.1, and can also be read as a characterization of dimension in locally compact, separable metrizable spaces.

Lemma 12 says that a locally compact, separable metrizable space of dimension $\leq n$ has a basic family of size $\leq 2n+1$, giving the reverse implication. For the converse:

Lemma 15. *A space X with a basic* family $\phi_1, \ldots, \phi_N$, where $N \leq 2n+1$, has dimension $\leq n$.*

Proof. Take any compact subset K of X. By Lemma 9, the maps $\Phi_1|K, \ldots, \Phi_N|K$ form a basic* family for K, hence by compactness a basic family. By Sternfeld's result connecting dimension and basic families in compact spaces (Theorem 2), it follows that $\dim K \leq n$.

By Lemma 10, X is locally compact, separable metrizable. Hence it has a locally finite cover by compact sets – each, by the above, of dimension $\leq n$. By the Locally Finite Sum Theorem for dimension, we deduce that X itself must have dimension $\leq n$. □

3.2 SEPARABLE METRIZABLE SPACES

The following simple lemma is used repeatedly and without further reference. Let Φ be a basic family for a space X, and let C be a C–embedded subspace (every continuous real valued function on C can be extended over X). Then clearly $\Phi \restriction C = \{\phi \restriction C : \phi \in \Phi\}$ is basic for C. Hence:

Lemma 16. *Let C be a C–embedded subspace of a space X — for example if X is normal, and C is closed — then basic $(X) \geq$ basic (C).*

Theorem 17. *Let X be separable metrizable. Then either basic (X) is finite, which occurs if and only if X is locally compact and finite dimensional, or basic $(X) = \mathfrak{c}$.*

Proof. Let X be separable metrizable. Four cases arise.

The first case is when X is locally compact and finite dimensional. Then basic $(X) \leq 2\dim(X)+1$, by Theorem 6.

In all remaining cases we show basic $(X) \geq \mathfrak{c}$, and so equals the continuum.

The second case is when X is not locally compact. Then, as X is first countable and normal, X contains a closed copy of the metric fan, F (defined below). So basic $(X) \geq$ basic $(F) \geq \mathfrak{c}$ by Proposition 23 and Proposition 24.

Case 3 is that X is locally compact, infinite dimensional, but contains no infinite dimensional compact subspaces. Then we can write X as a union of open sets $(U_n)_n$ such that, for all n, compact $\overline{U_n} \subset U_{n+1}$ and $\dim(\overline{U_n}) < \dim(U_{n+1})$. Using the Countable Sum Theorem for dimension, we can extract compact subsets C_n from the 'gaps' $U_{n+1} \setminus \overline{U_n}$ such that $\dim C_n < \dim C_{n+1}$ for all n. Now we see that C, the disjoint union of the C_n's is a closed subspace of X satisfying the conditions of Proposition 21, so we indeed have, basic $(X) \geq$ basic $(C) \geq \mathfrak{c}$.

Finally, suppose X is locally compact and contains an infinite dimensional compact subspace K. It suffices to show basic $(K) \geq \mathfrak{c}$, which is the content of Proposition 22. □

In vector spaces one method of giving a lower bound for the size of a basis is to find large linearly independent sets. We apply the same approach to give lower bounds for basic (X). Note that if V is a vector space, then $L \subseteq V$ is linearly independent if and only if its intersection with any subspace spanned by n members of V contains no more than n elements. This leads us to the correct definition of 'functional independence'.

Let $\mathcal{C}$ be a subset of $C(X)$. We say that $\mathcal{C}$ is (functionally) *independent* if for all n, and any $\phi_1, \ldots, \phi_n \in C(X)$ we have $|\mathcal{C} \cap \{\sum_{i=1}^n g_i \circ \phi_i : g_1, \ldots, g_n \in C(\mathbb{R})\}| \leq n$. (We omit the adjective 'functionally' except when we need to differentiate from linear independence in the vector space sense.)

Further, we say $\mathcal{C}$ is *weakly independent* if for all n, and any $\phi_1, \ldots, \phi_n \in C(X)$ we have $|\mathcal{C} \cap \{\sum_{i=1}^n g_i \circ \phi_i : g_1, \ldots, g_n \in C(\mathbb{R})\}| < \mathfrak{c}$, and we say $\mathcal{C}$ is *strongly independent* if for all n, and any $\phi \in C(X, \mathbb{R}^n)$ we have $|\mathcal{C} \cap \{g \circ \phi : g \in C(\mathbb{R}^n)\}| \leq n$.

Clearly 'independent' implies 'weakly independent'. Further, writing $\sum_{i=1}^n g_i \circ \phi_i$ as $g \circ \phi$ where $\phi(x_1, \ldots, x_n) = (\phi_1(x_1) \ldots, \phi_n(x_n))$ and $g(y_1, \ldots, y_n) = \sum_{i=1}^n g_i(y_i)$, we see that "strongly independent' implies 'independent'.

Lemma 18. *If a space X has a weakly independent family $\mathcal{C}$ of size $\geq \mathfrak{c}$, then basic $(X) \geq \mathfrak{c}$.*

Proof. Let Φ be a basic family for X. For each $f \in \mathcal{C}$, pick $\phi_1, \ldots, \phi_n$ from Φ so that $f = \sum_{i=1}^n g_i \circ \phi_i$. Then as $\mathcal{C}$ is weakly independent, the map taking f in $\mathcal{C}$ to $\{\phi_1, \ldots, \phi_n\}$ in $\bigcup_{m \in \mathbb{N}} [\Phi]^m$ is $< \mathfrak{c}$–to–1. Since $|\mathcal{C}| \geq \mathfrak{c}$, it follows that $|\Phi| \geq \mathfrak{c}$ — as required. □

To create large functionally independent families we will start from large sets generating linearly independent families in the vector space $\mathbb{R}^n$ (with its usual inner product).

Proposition 19. *Fix a natural number n.*

(a) There is a Cantor set C contained in the unit $(n-1)$–sphere of $\mathbb{R}^n$ such that for any distinct $x_1, \ldots, x_n$ in C, the x_i's form a basis of $\mathbb{R}^n$.

(b) Let J be a non–trivial closed bounded interval, and B a homeomorph of the n–cube, J^n. There is a Cantor set D contained in $C(B, J)$ such that for any distinct $d_1, \ldots, d_n$ in D the map $d = (d_1, \ldots, d_n) : B \to J^n$ is an embedding.

Proof (of (a)). The set $A = \{(1, t, \ldots, t^n) : t \in I\}$ is an arc in $\mathbb{R}^n$ such that any n distinct elements from A are linearly independent. Projecting on the unit sphere, and picking a Cantor subset gives what is claimed.

In fact one can show that 'almost all' (in the sense of Baire category applied to the Polish space $\mathcal{K}(\mathbb{R}^n)$) Cantor subsets of $\mathbb{R}^n$ are such that any n distinct elements from the Cantor set are linearly independent. □

Proof (of (b)). First note that if (b) holds for one choice of J and B, then it holds for all. We will use the interval $J = [-1, +1]$, and the closed n–ball, $B^{(n)}$. Also note that we work in the inner product space $\mathbb{R}^n$.

Fix a Cantor set C in the unit sphere of $\mathbb{R}^n$ as in part (a). Let $\hat{C} = \{\hat{c} : c \in C\}$ where $\hat{c}$ is the linear functional on $\mathbb{R}^n$ dual to c, namely $\hat{c}(x) = \langle c, x \rangle$. Then, by duality, $\hat{C}$ is a Cantor set in $\mathbb{R}^* \subseteq C(\mathbb{R}^n, \mathbb{R})$, and any n–many distinct elements of $\hat{C}$ are linearly independent.

Let $D = \{\hat{c} \restriction B^{(n)} : c \in C\}$. Then D is a family of continuous functions mapping $B^{(n)}$ to $[-1, +1]$, with the required properties. □

Proposition 20. *Fix K a compact space of dimension $> n \geq 2$.*

Then there is a Cantor set $C \subseteq C(K, I)$ such that for all $\phi \in C(K, I^n)$ we have $|C \cap \{g \circ \phi : g \in C(I^n, I)\}| \leq n$.

Proof. Recall (see [1], for example) that a normal space, X, has dimension $\leq n$ if and only if every continuous map from a closed subspace into the n–sphere (which is homeomorphic to the boundary of the $(n+1)$–cube) has a continuous extension over X. Hence, as $\dim K > n$, there is a map $p : K \to I^{n+1}$ and closed subspace A, such that $p \restriction A : A \to \partial I^{n+1}$ can not be continuously extended (over K into ∂I^{n+1}). We may suppose that $A = p^{-1}(\partial I^{n+1})$.

By Proposition 19 (b) there is a Cantor set D contained in $C(I^{n+1}, I)$ such that for any distinct $d_1, \ldots, d_{n+1} \in D$ the map $d = (d_1, \ldots, d_{n+1}) : I^{n+1} \to I^{n+1}$ is an embedding. For distinct $d_1, \ldots, d_{n+1} \in D$, and embedding $d = (d_1, \ldots, d_{n+1})$ define $f_d = d \circ p$. Note that p is onto, hence $f_d \neq f_{d'}$ if $d \neq d'$. Let $C = \{f_d : d \in D\}$. This is a Cantor set in $C(K, I^{n+1})$.

Suppose, for a contradiction, for some $\phi \in C(K, I^n)$, there were $(n+1)$ distinct elements $f_1, \ldots, f_{n+1}$ in $C \cap \{g \circ \phi : g \in C(I^n, I)\}$. So, for $i = 1, \ldots, n+1$, we have $f_i = d_i \circ p$ for some (distinct) $d_i \in D$, and $f_i = g_i \circ \phi$ for some $g_i \in C(I^n, I)$.

Let $d = (d_1, \ldots, d_{n+1})$, and $g = (g_1, \ldots, g_{n+1})$. So $p \circ d = g \circ \phi$. Since d is an embedding, we have $p = h \circ \phi$ where $h = (d^{-1} \circ g)$ is in $C(I^n, I^{n+1})$.

Let $A' = h^{-1}(\partial I^{n+1})$. Note that $\phi^{-1}(A') = p^{-1}(\partial I^{n+1}) = A$, so ϕ maps A inside A'. Since $K' = \phi(K)$ is contained in I^n, it has dimension $\leq n$. Hence the map $h \restriction A' : A' \to \partial I^{n+1}$ has a continuous extension $h' : K' \to \partial I^{n+1}$.

But now $p \restriction A : A \to \partial I^{n+1}$ has a continuous extension over K into ∂I^{n+1} — namely $h' \circ \phi$ — contradiction! □

Proposition 21. *Let $(C_n)_n$ be a sequence of compact spaces such that each C_n has finite dimension $> n$. Let $X = \bigoplus_n C_n$, and γX be a compactification of X.*

Then there is a Cantor set C contained in $C(\gamma X, I) \subseteq C(X)$ such that C is strongly independent for $C(X)$ (and hence for $C(\gamma X)$). Hence basic $(X) \geq \mathfrak{c}$ and basic $(\gamma X) \geq \mathfrak{c}$

Proof. For each $n \geq 2$, fix the Cantor set, E_n, guaranteed by Proposition 20 for the $> n$ dimensional space C_n, and fix a homeomorphism h_n from the standard Cantor set $\mathbf{C}$ to E_n. Let $C = \{f_c : c \in \mathbf{C}\}$ where f_c is constantly equal to zero on C_1 and on the remainder $\gamma X \setminus X$, and equals $h_n(c)/n$ on C_n. Note that each f_c is continuous, and so C is a Cantor set in $C(\gamma X, I)$.

Take any $n \geq 2$ and $\phi \in C(X, \mathbb{R}^n)$. Considering the restrictions of ϕ and elements of C to C_n, it is immediate from the properties of E_n, that $|C \cap \{g \circ \phi : g \in C(\mathbb{R}^n)\}| \leq n$. Thus C is strongly independent. □

Proposition 22. *Let K be compact and infinite dimensional. Then there is a Cantor set C contained in $C(K, I)$ which is strongly independent. Hence, basic $(K) \geq \mathfrak{c}$.*

Proof. We show an appropriate, strongly independent, Cantor set C exists. Dowker has shown [7] that if X is a normal space and M is a closed subspace with $\dim \leq n$ then $\dim X \leq n$ if and only if $\dim F \leq n$ for all closed subsets of X disjoint from M. In particular: $(*)$ if M contains a single point, x, then $\dim X > n$ if and only if $\dim F > n$ for some closed subset F of $X \setminus \{x\}$. For each point x in K pick a closed neighborhood of minimal dimension, B_x. By compactness, for some x, B_x is infinite dimensional, and so all neighborhoods of x are infinite dimensional. Let $K_1 = K$. Apply $(*)$ to get a compact subset C_1 of K_1 not containing x with $\dim C_1 > 1$. Pick a closed neighborhood K_2 of x disjoint from C_1. Apply $(*)$ to get a compact subset C_2 of K_2 not containing x with $\dim C_2 > \max(2, \dim C_1)$. Inductively, we get a pairwise disjoint collection, $\{C_n : n \in \mathbb{N}\}$, of compact subsets of K which are either (i) of strictly increasing (finite) dimensions, or (ii) all infinite dimensional. Let K' be the closed subspace $\overline{\bigoplus_n C_n}$.

In the first case we apply Proposition 21 to K' to get a strongly independent Cantor set in $C(K')$ – and hence in $C(K)$ – as required.

In the second case, by Proposition 20, for each n there is a Cantor set $E_n \subseteq C(C_n, I)$ such that for all $\phi \in C(C_n, I^n)$ we have $|E_n \cap \{g \circ \phi : g \in C(I^n, I)\}|$ finite. Fix homeomorphisms h_n between the standard Cantor set $\mathbf{C}$ and E_n.

Define, for $c \in \mathbf{C}$, a map $f_c : K' \to I$ by: f_c is identically zero on $K' \setminus \bigoplus_n C_n$ and $f_c(x') = (1/n)h_n(c)(x')$ if $x' \in C_n$. Then the f_c's are continuous, can be continuously extended over K, and so form a Cantor set C in $C(K, I)$. Further, if $\phi \in C(K, I)$ and $f_1, \ldots, f_{n+1} \in C$, then the f_i's are not all in $\{g \circ \phi : g \in C(I^n, I)\}$, because $f_1 \restriction E_n, \ldots, f_{n+1} \restriction E_n$ are not all in $\{g \circ (\phi \restriction E_n) : g \in C(I^n, I)\}$, by choice of E_n.

Thus the Cantor set C is strongly independent as required. □

Let F be the *metric fan* where $F = (\mathbb{N} \times \mathbb{N}) \cup \{*\}$, points in $\mathbb{N} \times \mathbb{N}$ are isolated and basic neighborhoods of $*$ are $B(*, n) = ([n, \infty) \times \mathbb{N}) \cup \{*\}$. Then a separable metric space is not locally compact if and only if it contains a closed copy of the metric fan. Thus if basic $(F) = \mathfrak{c}$ then basic $(X) = \mathfrak{c}$ for every separable metric space X which is not locally compact.

We first reduce the calculation of basic (F) to that of basic $(\mathbb{N}, [-1, +1])$. Here we say that a family $\hat{\Phi} \subseteq C(\mathbb{N}, [-1. + 1])$ is 'basic for $\mathbb{N}$ into $[-1, 1]$' if $\forall \hat{f} \in C(\mathbb{N}, [-1, +1])$ there are $\hat{\phi}_1, \ldots, \hat{\phi}_n \in \hat{\Phi}$, and $\hat{g}_1, \ldots, \hat{g}_n \in C(\mathbb{R})$ such that $\hat{f} = \sum_{i=1}^n \hat{g}_i \circ \hat{\phi}_i$, and define basic $(\mathbb{N}, [-1, +1]) = \min\{|\hat{\Phi}| : \hat{\Phi}$ is basic for $\mathbb{N}$ into $[-1, 1]\}$.

Proposition 23. *basic* $(F) \geq$ *basic* $(\mathbb{N}, [-1, +1])$.

Proof. Let Φ be basic for F. We will show that there is a $\hat{\Phi}$ with $|\hat{\Phi}| = |\Phi|$ such that $\hat{\Phi}$ is basic for $\mathbb{N}$ into $[-1, +1]$.

For each $\phi \in \Phi$ and n such that ϕ maps $\{n\} \times \mathbb{N}$ into $[-1, +1]$, define $\widehat{\phi_n}$ in $C(\mathbb{N}, [-1, +1])$ by $\widehat{\phi_n}(m) = \phi(n, m)$. Let $\widehat{\Phi_n} = \{\widehat{\phi_n} : \phi \in \Phi\}$ and $\widehat{\Phi} = \bigcup_n \widehat{\Phi_n}$. Note that $|\widehat{\Phi}| = |\Phi|$.

Take any $\hat{f} \in C(\mathbb{N}, [-1, +1])$. Define $f : F \to [-1, +1]$ by $f(*) = 0$ and $f(n, m) = \hat{f}(m)/n$. Note f is continuous. So there are $\phi_1, \ldots, \phi_n$ in Φ and $g_1, \ldots, g_n$ in $C(\mathbb{R})$ such that $f = \sum_i g_i \circ \phi_i$.

By continuity of $\phi_1, \ldots, \phi_n$ at $*$ there is an N such that each ϕ_i maps $\{N\} \times \mathbb{N}$ into a closed bounded interval, say I_i. Fix homeomorphisms h_i of $\mathbb{R}$ with itself carrying I_i to $[-1, +1]$. Now we see that, replacing g_i with $g_i \circ h_i^{-1}$ and ϕ_i with $h_i \circ \phi_i$, we can assume that the ϕ_i all map into $[-1, +1]$.

Thus $\widehat{\phi_1} = \widehat{(\phi_1)_N}, \ldots, \widehat{\phi_n} = \widehat{(\phi_n)_N}$ are in $\widehat{\Phi}_N \subseteq \widehat{\Phi}$. Further, as $\hat{f}(m)/N = f(N,m) = \sum_{i=1}^n g_i(\phi_i(N,m)) = \sum_i g_i(\widehat{\phi_i}(m))$, we have that $\hat{f} = \sum_{i=1}^n \widehat{g_i} \circ \widehat{\phi_i}$ where $\widehat{g_i} = N.g_i$ — as required. □

Proposition 24. *There is a Cantor set C contained in $C(\mathbb{N}, [-1,+1])$ such that $|C \cap \{\sum_{i=1}^n g_i \circ \phi_i : g_1, \ldots, g_n \in C(\mathbb{R})\}| \leq \aleph_0$ for all $\phi_1, \ldots, \phi_n$ from $C(\mathbb{N}, [-1,+1])$.*

Thus C is 'weakly independent' in the sense appropriate for $C(\mathbb{N}, [-1,+1])$, and so basic $(\mathbb{N}, [-1,+1]) = \mathfrak{c}$.

Proof. Define $C = \{f \in C(\mathbb{N}, [-1,+1]) : f(\mathbb{N}) = \{-1,+1\}\}$. Then C is a Cantor set, and we will prove that, for each n, and finite $\Phi' \subseteq C(\mathbb{N}, [-1,+1])$ we have $|C \cap L(\Phi')| = \aleph_0$.

Fix $n \geq 1$. Fix $\phi \in C(\mathbb{N}, [-1,+1]^n)$. As in the argument that 'strongly independent' implies 'independent' to prove the claim it suffices to show that there are only countably many $f \in C$ representable as $g \circ \phi$ for some $g \in C([-1,+1]^n, [-1,+1])$.

Let $K = \overline{\phi(\mathbb{N})}$ — a compact subset of $[-1,+1]^n$. A composition $g \circ \phi : \mathbb{N} \to [-1,+1]$ is determined by the values of g on $\phi(\mathbb{N})$, and so definitely determined by its values on K.

If $g \circ \phi$ is in C, then, by continuity, $g \restriction K$ maps K onto $\{-1,+1\}$. Thus K is partitioned into two non–empty clopen pieces, one of which is mapped by g to -1, and the other to $+1$. But a compact metric space only has countably many clopen subsets. So there are only a countable number of possibilities for g on K, and only countably many $f \in C$ representable as $g \circ \phi$ — as claimed. □

Corollary 25. *Let X be finite dimensional, locally compact, **not compact**, separable metrizable. Then:*

(1) there is a basic family $\Phi \subseteq C(X)$ such that Φ is finite, but

(2) there is no basic family Φ^* consisting of **bounded** functions such that $|\Phi^*| < \mathfrak{c}$.*

Proof. The first claim is just Theorem 6. For the second part, first note that since $\mathbb{N}$ can be embedded as a closed subspace of X, it is sufficient to show that (2) holds for $\mathbb{N}$. Suppose, for contradiction, there exists a basic family Φ^* for $\mathbb{N}$ consisting of bounded function whose cardinality is $< \mathfrak{c}$.

Write $\Phi^* = \bigcup_{n \in \mathbb{N}} \Phi_n$ where $\Phi_n = \{\phi : -n \leq \phi(a) \leq n, \text{ for each } a \in \mathbb{N}\}$. Then $C^*(\mathbb{N}) = \bigcup_{n \in \mathbb{N}} L(\Phi_n)$. Let $\mathcal{F} = \{f \in C(\mathbb{N}, [-1, +1]) : f(\mathbb{N}) = \{-1, +1\}\}$ as in the proof of Proposition 24. There exists an m_0 such that $|\mathcal{F} \cap L(\Phi_{m_0})| = \mathfrak{c}$. But the argument in the proof of Proposition 24 shows $L(\Phi_{m_0}) \leq |\Phi^*| < \mathfrak{c}$ which is the desired contradiction. □

3.3 COMPACT SPACES

Proposition 26. *Suppose K is compact and finite dimensional.*

Then basic $(K) \leq cof([w(K)]^{\aleph_0}, \subseteq)$.

Proof. Let K be compact of dimension n. Then there is a directed set $(\Lambda, \leq)$ where $|\Lambda| = w(K)$, compact metric K_λ with $\dim K_\lambda \leq n$, and for all $\lambda \geq \mu$ a continuous map $f_{\lambda,\mu}$ such that $K = \varprojlim\{K_\lambda : \lambda \in \Lambda\} = \{\langle x_\lambda \rangle \in \prod_\lambda K_\lambda : \lambda \geq \mu \implies f_{\lambda,\mu}(x_\lambda) = x_\mu\}$. For any $\Lambda_0 \subseteq \Lambda$ let $\pi_{\Lambda_0} : \prod_{\lambda \in \Lambda} K_\lambda \to \prod_{\lambda \in \Lambda_0} K_\lambda$ be the natural projection map.

Let $\mathcal{C}$ be cofinal in $([\Lambda]^{\aleph_0}, \subseteq)$. We may suppose that each C in $\mathcal{C}$ is directed. For each $C \in \mathcal{C}$, $K_C = \varprojlim\{K_\lambda : \lambda \in C\}$ is compact, metric of dimension $\leq n$. So K_C has a basic family Φ'_C of size $2n+1$. Define $p_C = \pi_C \restriction \varprojlim\{K_\lambda : \lambda \in \Lambda\}$. Define $\Phi_C = \{\phi' \circ p_C : \phi' \in \Phi'_C\}$. and $\Phi = \bigcup_{C \in \mathcal{C}} \Phi_C$. Then $|\Phi| = |\mathcal{C}|$. We show that Φ is basic – as required.

To this end, take any $f \in C(K)$. The first step is to show that there is a $C \in \mathcal{C}$ and continuous $g : \varprojlim\{K_\lambda : \lambda \in C\} \to \mathbb{R}$ such that $f = g \circ p_C$. We can do so by using the fact that the corresponding property holds for continuous functions on products of compact metrizable spaces [22]. (Alternatively we could proceed more directly by adapting the proof of the theorem for products to the present situation.)

So extend $f : \varprojlim\{K_\lambda : \lambda \in \Lambda\} \to \mathbb{R}$ to continuous $\hat{f} : \prod_{\lambda \in \Lambda} K_\lambda \to \mathbb{R}$. Then there is a countable $\Lambda_0 \subseteq \Lambda$ and continuous $g_0 : \prod_{\lambda \in \Lambda_0} K_\lambda \to \mathbb{R}$ such that $\hat{f} = g_0 \circ \pi_{\Lambda_0}$. Pick $C \in \mathcal{C}$ such that $C \supseteq \Lambda_0$. Note that as C is directed, $\{\langle x_\lambda \rangle_{\lambda \in C} : \lambda \geq \mu \implies f_{\lambda,\mu}(x_\lambda) = x_\mu\} = \varprojlim\{K_\lambda : \lambda \in C\}$, and π_C maps $\varprojlim\{K_\lambda : \lambda \in \Lambda\}$ to $\varprojlim\{K_\lambda : \lambda \in C\}$. We can write $\hat{f} = \hat{g} \circ \pi_C$ where $\hat{g} = g_0 \circ \pi^C_{\Lambda_0}$

is a continuous map $\prod_{\lambda \in C} K_\lambda$ into $\mathbb{R}$. Thus $f = \hat{f} \restriction \varprojlim\{K_\lambda : \lambda \in \Lambda\} = g \circ p_C$ where $p_C = \pi_C \restriction \varprojlim\{K_\lambda : \lambda \in \Lambda\}$ and $g = \hat{g} \restriction \varprojlim\{K_\lambda : \lambda \in C\}$.

Now we see that $g = \sum_{i=1}^{2n+1} g_i \circ \phi'_I$ where $\phi'_C \in \Phi'_C$ and $g_i \in C(\mathbb{R})$. Thus

$$f = g \circ p_C = \sum_{i=1}^{2n+1} g_i \circ (\phi'_i \circ \pi_C) = \sum_{i=1}^{2n+1} g_i \circ \phi_i,$$

where $\phi_1, \ldots, \phi_{2n+1}$ are in $\Phi_C \subseteq \Phi$ and $g_1, \ldots, g_{2n+1}$ are in $C(\mathbb{R})$. $\square$

Call a space X 'nice' if it contains a discrete subset D with $|D| = w(X)$. Note that there are many examples of compact 'nice' spaces, for example: 2^κ, $I^n \times 2^\kappa$ and I^κ are compact, 'nice' and span the dimensions.

Proposition 27. *If K is compact and 'nice', then basic $(K) \geq cof([w(K)]^{\aleph_0}, \subseteq)$.*

Proof. Let D be discrete in K with $w(K) = |D|$. Let $K' = \overline{D}$, and $K'_c = K' \setminus D$. Since $w(K') = w(K)$ and basic $(K) \geq$ basic (K') it suffices to show basic $(K') \geq cof([w(K')]^{\aleph_0}, \subseteq)$.

Note that D is open in K', so K'_c is compact. Take any function $f \in C(K', \mathbb{R}^n)$. Then $f(K'_c)$ is a compact subset of $\mathbb{R}^n$, so it is a G_δ subset, and we can write $f(K'_c)$ as $\bigcap_{n \in \mathbb{N}} U_n$, where U_n is open set in $\mathbb{R}^n$ for each n. As K' is compact, each $K' \setminus f^{-1}(U_n)$ is closed and discrete, and hence finite. So we can define a countable subset of D for each f by $C_f = \bigcup_{n \in \mathbb{N}} K' \setminus f^{-1}(U_n)$.

Now suppose $\Phi \subseteq C(K')$ with $|\Phi| < cof(|w(K')|^{\aleph_0}, \subseteq)$. We show Φ is not a basic family.

Given $\phi_1, \phi_2, \cdots, \phi_n$ from Φ, let $\hat{\phi} = (\phi_1, \ldots, \phi_n) : K' \to \mathbb{R}^n$, and $C(\phi_1, \ldots, \phi_n) = C_{\hat{\phi}}$. Let $\mathcal{C} = \{C(\phi_1, \ldots, \phi_n) : \phi_1, \ldots, \phi_n \in \Phi\}$. Since $|\Phi| < cof([w(K')]^{\aleph_0}, \subseteq)$, the collection $\mathcal{C}$ is not cofinal in $[D]^{\aleph_0}$. Therefore there exists a countably infinite subset C of D such that for any $\phi_1, \cdots, \phi_n$, C is not a subset of $C(\phi_1, \ldots, \phi_n)$.

Take any $\phi_1, \ldots, \phi_n$ in Φ. Pick x in C but not in $C(\phi_1, \ldots, \phi_n)$. By definition of $C(\phi_1, \ldots, \phi_n)$ there exists $x' \in K'_c$ such that $\hat{\phi}(x) = \hat{\phi}(x')$. Then for any $g_1, \ldots, g_n$ from $C(\mathbb{R})$, $\sum_{i=1}^n g_i \circ \phi_i$ takes the same value at a point in C and at a point in K'_c.

But now we see that if we enumerate $C = \{x_1, x_2, \ldots\}$, and define h by $h(x_n) = 1/n$ and h is identically zero outside C, then h is continuous and $h(C)$ is disjoint from $h(K'_c)$. Thus h can not be represented by any finite collection of Φ, and so Φ is not basic. $\square$

From the identity $w(K)^{\aleph_0} = cof([w(K)]^{\aleph_0}, \subseteq) \times \mathfrak{c}$ and Propositions 22, 26 and 27 we conclude:

Theorem 28. *If K is compact and 'nice' then:*

either *K is finite dimensional and basic $(K) = cof([w(K)]^{\aleph_0}, \subseteq)$,*

or *K is infinite dimensional and basic $(K) = |C(K)| = w(K)^{\aleph_0}$.*

Recall that for a compact space K we have $w(K) \leq$ basic $(K) \leq w(K)^{\aleph_0} = |C(K)|$. Either or both of the inequalities can, at least consistently, be strict.

Taking $K = 2^{\aleph_1}$, we have that $w(K) =$ basic (K) and basic $(K) < w(K)^{\aleph_0}$ if and only if the Continuum Hypothesis fails.

Taking $K = 2^{\aleph_\omega}$ or $K = I^{\aleph_\omega}$ then we have $w(K) <$ basic (K), and while basic $(I^{\aleph_\omega})$ must equal $w(I^{\aleph_\omega})^{\aleph_0}$, it is at least consistent that basic $(2^{\aleph_\omega}) = \aleph_{\omega+1} < \aleph_{\omega+2} = \mathfrak{c} = w(2^{\aleph_\omega})^{\aleph_0}$.

4.0 HILBERT'S 13TH PROBLEM REVISITED

In this chapter, we show that the Kolmogorov Superposition Theorem holds for all continuous functions $f : \mathbb{R}^m \to \mathbb{R}$ (Theorem 29). Further, using earlier work in the previous chapters, we characterize the topological spaces satisfying a superposition result of the Kolmogorov type. It turns out these spaces are precisely the locally compact, finite dimensional separable metrizable spaces, or equivalently, those spaces homeomorphic to a closed subspace of Euclidean space (Theorem 34). Together these results establish Theorem C from the Introduction.

4.1 SUPERPOSITIONS

Note that we always use the max norm. $\|\cdot\|_\infty$, on $\mathbb{R}^m$.

Theorem 29. *Fix m in $\mathbb{N}$. There exist $\psi_{pq} \in C(\mathbb{R})$, for $q = 1, 2, \ldots, 2m+1$ and $p = 1, 2, \ldots, m$, such that for any function $f \in C(\mathbb{R}^m)$, there can be found functions $g_1, \ldots, g_{2m+1}$ in $C(\mathbb{R})$ such that:*

$$f(\mathbf{x}) = \sum_{q=1}^{2m+1} g_q(\phi_q(\mathbf{x})), \quad \textit{where } \phi_q(x_1, \ldots, x_m) = \psi_{1q}(x_1) + \cdots + \psi_{mq}(x_m).$$

Further, one can arrange it so that the co–ordinate functions, $g_1, \ldots, g_{2m+1}$ are all identical (say to g), and that the elementary functions, ψ_{pq}, (and hence the ϕ_q) are Lipschitz (with Lipschitz contstant 1).

Proof. We break the proof into five parts. In the first step we define a family of 'grids', and approximations to the functions ψ_{pq}. Next we define the ψ_{pq} and ϕ_q, and establish certain useful properties of the grids and functions. In the following two steps we show that the functions ϕ_q are basic for $\mathbb{R}^m$ — using just a single co–ordinate function, first for compactly supported functions, and then in general. Finally, we show that the constructed elementary functions can be modified to be Lipschitz with Lipschitz constant 1.

1. Construction of the Grids and Approximations

We establish by induction on k, the existence for each $k \in \mathbb{N}$, $p = 1, 2, \ldots, m$, and $q = 1, 2, \ldots, 2m+1$, of positive ϵ_k, $\gamma_k < 1/10$, distinct positive prime numbers $P_k^{pq} > m + 10$, discrete families ('grids') $\mathcal{S}_k^q$ of open intervals of $\mathbb{R}$ and continuous functions $f_k^{pq} : \mathbb{R} \to \mathbb{R}$ such that:

(1) the sequences of ϵ_k's and γ_k's both strictly decrease to zero (in fact, for all k, $0 < \epsilon_{k+1} < \epsilon_k/6$ and $0 < \gamma_k < 1/k$),

(2) each member of $\mathcal{S}_k^q$ has diameter $\leq \gamma_k$,

for each fixed k any two of the families $\{\mathcal{S}_k^q : q = 1, \ldots, 2m + 1\}$ cover $[-k, k]$, and all cover $\{-k, 0, k\}$;

(3) $\prod_{p=1}^m P_k^{pq} < P_k^{pq'}$ given $q < q'$ for each fixed k;

(4) $m\epsilon_k < 1/\prod_{q=1}^{2m+1} \prod_{p=1}^m P_k^{pq}$;

(5) f_k^{pq} is non–decreasing on $\mathbb{R}^+$, non–increasing on $\mathbb{R}^-$ and constant outside $[-k, k]$;

(6) f_k^{pq} is constant on each member of $\mathcal{S}_k^q$ with value a positive integral multiple of $1/P_k^{pq}$, and $(f_k^{pq}(J_1) - f_k^{pq}(J_2))P_k^{pq} \mod P_k^{pq} \neq 0$ given $J_1, J_2 \in \mathcal{S}_k^{pq}$;

additionally, if J is an interval containing 0, then f_k^{pq} maps J to 0;

(7) $|f_k^{pq}(k) - k| < 1/(m + 1)$ and $|f_k^{pq}(-k) - k| < 1/(m + 1)$;

(8) for each $\ell \leq j < k$ and $x \in [-\ell, \ell]$, $f_j^{pq}(x) \leq f_k^{pq}(x) \leq f_j^{pq}(x) + \epsilon_j - \epsilon_k$.

Base Step: It is straightforward to find discrete collections of open intervals $\mathcal{S}_1^q$ for $q = 1, \ldots, 2m+1$ such that any two of the families $\{\mathcal{S}_1^q : q = 1, 2, \cdots, 2m+1\}$ cover $[-1, 1]$, each of the families covers $\{1, 0, -1\}$, and each interval in the collection has length $\leq \gamma_1 = 1/10$.

Let n_1 be the number of all the open interval in all the collections $\mathcal{S}_1^q$ ($1 \leq p \leq m$, $1 \leq q \leq 2m+1$). For $q = 1, \ldots, 2m+1$ pick distinct primes P_1^{pq} larger than $m \cdot n_1$ and $\prod_{p=1}^m P_1^{pq} < P_1^{pq'}$ given $q < q'$.

Now we define f_1^{pq} on $[-1,1]$. Then for $x > 1$ define $f_1^{pq}(x) = f_1^{pq}(1)$, and for $x < -1$ define $f_1^{pq}(x) = f_1^{pq}(-1)$.

If $J \in \mathcal{S}_1^q$, then define f_1^{pq} such that f_1^{pq} restricted to J is a positive integral multiple of $1/P_1^{pq}$. More specifically, if $0 \in J$ then $f_1^{pq}(J) = 0$; if $1 \in J$ then $f_1^{pq}(J) = 1 - 1/P_1^{pq}$; and if $-1 \in J$ then $f_1^{pq}(J) = 1 - 2/P_1^{pq}$. This can easily be done so that f_1^{pq} (as defined so far) is non–decreasing on $[0,1]$ and non–increasing on $[-1,0]$.

For x in $[-1,1] \setminus \bigcup \mathcal{S}_1^q$, interpolate f_1^{pq} linearly.

Choose $\epsilon_1 > 0$ such that $m\epsilon_1 < 1/\prod_{q=1}^{2m+1}\prod_{p=1}^{m} P_1^{pq}$.

All (applicable) conditions (1)–(8) hold.

Inductive Step: Suppose P_{k-1}^{pq}, ϵ_{k-1}, γ_{k-1}, $\mathcal{S}_{k-1}^q$ and f_{k-1}^{pq} are all given and satisfy the requirements (1)–(8).

By uniform continuity of f_{k-1}^{pq} on $[-(k-1), k-1]$, there exists $\gamma_k < \min\{1/k, \gamma_{k-1}\}$ such that $|f_{k-1}^{pq}(x_1) - f_{k-1}^{pq}(x_2)| < \epsilon_{k-1}/6$ if $|x_1 - x_2| < \gamma_k$ for each $p = 1, \ldots, m$ and $q = 1, \ldots, 2m+1$. Then it is straightforward to find discrete collections of open intervals, $\mathcal{S}_k^q$ for and $1 \le q \le 2m+1$, such that any two of the families $\{\mathcal{S}_k^q : q = 1, 2, \cdots, 2m+1\}$ cover $[-k,k]$, each of the families covers $\{k, 0, -k\}$, each interval in the collection has length $\le \gamma_k$ and the distance between each pair of adjacent intervals is also $\le \gamma_k$.

Let n_k be the total number of open intervals in all the collections $\mathcal{S}_k^q$ for $q = 1, 2, \ldots, 2m+1$. For each p, q select distinct primes P_k^{pq} so that $2n_k/P_k^{pq} < \epsilon_{k-1}/6$. Also, $\prod_{p=1}^{m} P_k^{pq} < P_k^{pq'}$ given $q < q'$.

Next, we give the construction of f_k^{pq} on $[-k,k]$. Outside of $[-k,k]$ extend constantly (as in the Base Step).

- If $J \in \mathcal{S}_k^q$, then $f_k^{pq}(J)$ is a positive integral multiple of $1/P_k^{pq}$. For any $J \in \mathcal{S}_k^q$ with $J \cap [-(k-1), k-1] \neq \emptyset$, we can ensure that $f_{k-1}^{pq}(x) < f_k^{pq}(x) < f_{k-1}^{pq}(x) + \epsilon_{k-1}/3$.

  [i] Since $2n_k/P_k^{pq} < \epsilon_{k-1}/6$ and $|f_{k-1}^{pq}(x_1) - f_{k-1}^{pq}(x_2)| < \epsilon_{k-1}/6$ when $|x_1 - x_2| < \gamma_k$, there are $2n_k$ possible choices for the value of $f_k^{pq}(J)$ ($J \in \mathcal{S}_k^q$) which makes $f_{k-1}^{pq}(x) \le f_k^{pq}(x) \le f_{k-1}^{pq}(x) + \epsilon_{k-1}/3$ for $x \in J \cap [-(k-1), k-1]$. As there are many fewer than $2n_k$ elements in $\mathcal{S}_k^q$, we can select the $f_k^{pq}(J)$'s such that $(f_k^{pq}(J_1) - f_k^{pq}(J_2))P_k^{pq}$

$\mod P_k^{pq} \neq 0$ for any $J_1, J_2 \in \mathcal{S}_k^q$.

[ii] More specifically, if $0 \in J$ then $f_k^{pq}(J) = 0$, if $k \in J$ then $f_k^{pq}(J) = 1 - 1/P_k^{pq}$, and if $-k \in J$ then $f_k^{pq}(J) = 1 - 2/P_k^{pq}$. This can easily be done to make f_k^{pq} (as defined so far) non–decreasing on $[0, k]$ and non-increasing on $[-k, 0]$.

- If $x \notin \bigcup \mathcal{S}_k^q$, let J_L and J_R be the adjacent intervals in $\mathcal{S}_k^{pq}$ such that x lies between them. Let x_L be the right endpoint of J_L and x_R be the left end point of J_R Then f_k^{pq} maps $[x_L, x_R]$ linearly to $[f_{k-1}^{pq}(J_L), f_{k-1}^{pq}(J_R)]$. Since $|x_L - x_R| < \gamma_k$, $|f_{k-1}^{pq}(x_L) - f_{k-1}^{pq}(x_R)| < \epsilon_{k-1}/6$, therefore, $f_k^{pq}(x) - f_{k-1}^{pq}(x) < \epsilon_{k-1}/3 + \epsilon_{k-1}/6 = \epsilon_{k-1}/2$.

Choose ϵ_k such that $m\epsilon_k < \min\{1/\prod_{q=1}^{2m+1}\prod_{p=1}^{m} P_k^{pq}, \epsilon_{k-1}/6\}$.

All requirements (1)–(8) are satisfied.

By conditions (2), (3), (4) and (8), we have the following claim.

Claim 30. *For each k, $|\sum_{p=1}^m f_k^{pq}(J_p) - \sum_{p=1}^m f_k^{pq'}(J'_p)| > m\epsilon_k$, for different $J_p \in \mathcal{S}_k^q$ and $J'_p \in \mathcal{S}_k^{q'}$ where $p = 1, \ldots, m$.*

2. Definition and Useful Properties of the Functions, ψ_{pq} and ϕ_q

For $x \in \mathbb{R}$, let $\psi_{pq}(x) = \lim_{k\to\infty} f_k^{pq}(x)$. Now for a fixed $n \in \mathbb{N}$, and any $x \in [-n, n]$, $f_k^{pq}(x) \leq \psi_{pq}(x) \leq f_k^{pq}(x) + \epsilon_k$ for $k > n + 1$. So ψ_{pq} restricted to $[-n, n]$, being the uniform limit of the f_k^{pq} for $k > n + 1$, is continuous on $[-n, n]$. Therefore, ψ_{pq} is continuous on $\mathbb{R}$.

Also, by construction, the image of $[n, n + 1]$ under ψ_{pq} is a subset of $[|n| - 1/(m + 1), |n| + 1 + 1/(m + 1)]$ for each $n \in \mathbb{Z}$.

Let $\phi_q(x_1, \ldots, x_m) = \psi_{1q}(x_1) + \cdots + \psi_{mq}(x_m)$ for $(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$.

Our eventual goal is to show $\{\phi_q : q = 1, 2, \ldots, 2m + 1\}$ is a basic family of $\mathbb{R}^m$, however first, we establish some useful properties of the grids and functions.

For each q and k, let $\mathcal{J}_k^q = \{C_1 \times C_2 \times \cdots \times C_m : C_p \in \mathcal{S}_k^q$ for each $p = 1, 2, \ldots, m\}$. Then we can say the following about $\mathcal{J}_k^q$.

- For a fixed q and k, $\mathcal{J}_k^q$ is a discrete collection.
- For a fixed k, any element in $\mathbb{R}^m$ belongs to at least $m + 1$ rectangles of $\mathcal{J}_k^q$, i.e. any $m + 1$ of $\{\mathcal{J}_k^q : q = 1, \ldots, 2m + 1\}$ form an open cover of $\mathbb{R}^m$.

Let $\mathcal{U}_k^q = \{\phi_q(C) : C \in \mathcal{J}_k^q\}$. Take $C = C_1 \times C_2 \times \cdots \times C_m \in \mathcal{J}_k^q$, then $\phi_q(C)$ is contained in the interval $[\sum_{p=1}^m f_k^{pq}(C_p), \sum_{p=1}^m f_k^{pq}(C_p) + m\epsilon_k]$. By condition (4) in the construction of the f_k^{pq}, these closed intervals are disjoint for fixed k. Therefore,

Claim 31. *$\bigcup_{q=1}^{2m+1} \mathcal{U}_k^q$ is a discrete collection of subsets of $\mathbb{R}$ for fixed k.*

3. Construct The Co–Ordinate Function for Compactly Supported Functions

We now prove:

Claim 32. *For any compactly supported $h \in C(\mathbb{R}^m)$, there is g in $C(\mathbb{R})$ such that $h = \sum_{q=1}^{2m+1} g \circ \phi_q$.*

Fix a compactly supported $h \in C(\mathbb{R}^m)$. Choose ℓ in $\mathbb{N}$ so that $h(\mathbf{x}) = 0$ for any $\mathbf{x}$ outside $K = [-\ell - 1, \ell + 1]^m$.

For each integer $r \geq 0$, find positive k_r and continuous functions $\chi_r : \mathbb{R} \to \mathbb{R}$ ($k_0 = \ell$ and $\chi_1 = 0$) such that if $h_r(\mathbf{x}) = \sum_{q=1}^{2m+1} \sum_{s=0}^{r} \chi_s(\phi_q(\mathbf{x}))$ and $M_r = \sup_{\mathbf{x} \in \mathbb{R}^m} |(h - h_r)(\mathbf{x})|$, then:

(1) $k_{r+1} > k_r$;

(2) if $\|\mathbf{a} - \mathbf{b}\|_\infty < m/\gamma_{k_{r+1}}$, then $|(h - h_r)(\mathbf{a}) - (h - h_r)(\mathbf{b})| < (2m+2)^{-1} M_r$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$;

(3) χ_{r+1} is constant on each member of $\bigcup_{q=1}^{2m+1} \mathcal{U}_{k_{r+1}}^q$;

(4) if $C \cap (\mathbb{R}^m \setminus K) \neq \emptyset$ for $C \in \bigcup_{q=1}^{2m+1} \mathcal{J}_{k_{r+1}}^q$, then the value of χ_{r+1} on $\phi_q(C)$ is 0, otherwise, its value on $\phi_q(C)$ is $(m+1)^{-1}(h - h_r)(\mathbf{y})$ for some arbitrarily chosen element $\mathbf{y} \in C$; and

(5) $\chi_{r+1}(x) \leq (m+1)^{-1} M_r$ for each $x \in \mathbb{R}$.

We can construct χ_{r+1} which satisfies property (4) by Claim 31.

The k_r and χ_r are defined inductively on r. Also for any $\mathbf{a}, \mathbf{b} \in C \in \mathcal{J}_{k_{r+1}}^q$, $\|\mathbf{a} - \mathbf{b}\|_\infty < m/10^{k_{r+1}}$. Therefore:

(6) for $\mathbf{x} \in \bigcup\{C : C \in \mathcal{J}_{k_{r+1}}\}$,
$|(m+1)^{-1}(h - h_r)(\mathbf{x}) - \chi_{r+1}(\phi_q(\mathbf{x}))| < (m+1)^{-1}(2m+2)^{-1} M_r$.

Also for each $\mathbf{x} \in \mathbb{R}^m$, there are at least $m + 1$ distinct values of q such that $\mathbf{x} \in \bigcup\{C : C \in \mathcal{J}_{k_{r+1}}^q\}$. Then there are $m + 1$ values of q such that (6) is true; for the other m values of q, (5) in the construction holds.

Hence, for $\mathbf{x} \in K$,

$$
\begin{aligned}
|(h - h_{r+1})(\mathbf{x})| &= |(h - h_r)(\mathbf{x}) - \sum_{q=1}^{2m+1} \chi_{r+1}(\phi_q(\mathbf{x}))| \\
&< (m+1) \cdot (m+1)^{-1}(2m+2)^{-1} M_r + m \cdot (m+1)^{-1} M_r \\
&= \frac{2m+1}{2m+2} M_r.
\end{aligned}
$$

While for $\mathbf{x} \notin K$, $\sum_{q=1}^{2m+1} \chi_{r+1}(\phi_q(\mathbf{x})) = 0$ by property (4).

Therefore, $M_{r+1} < (2m+1) \cdot (2m+2)^{-1} \cdot M_r$, so $M_r < ((2m+1) \cdot (2m+2)^{-1})^r \cdot M_0$ for each r, hence $\lim_{r\to\infty} M_r = 0$, and thus $h(\mathbf{x}) = \lim_{r\to\infty} h_r(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^m$.

Moreover, by condition (5), the functions $\sum_{s=0}^{r} \chi_s$ converge uniformly to a continuous function $g : \mathbb{R} \to \mathbb{R}$ and

$$
h(\mathbf{x}) = \lim_{r\to\infty} h_r(\mathbf{x}) = \lim \sum_{q=1}^{2m+1} \sum_{s=0}^{r} \chi_s(\phi_q(\mathbf{x})) = \sum_{q=1}^{2m+1} g(\phi_q(\mathbf{x})).
$$

This complete the proof of the Claim.

4. Construct the Co–Ordinate Function for All Functions

We complete the proof of elementarity by showing:

Claim 33. *For any $f \in C(\mathbb{R}^m)$, there is a g in $C(\mathbb{R})$ such that $f = \sum_{q=1}^{2m+1} g \circ \phi_q$.*

First some preliminary definitions. Let K_n^i be

$$
\{(x_1, x_2, \cdots, x_m) : x_i \in [-n-2, -n] \cup [n, n+2],\ x_j \in [-n-2, n+2] \text{ for } j \neq i\},
$$

and let $\mathcal{K} = \{K_n = \bigcup_{i=1}^{m} K_n^i : n \in \mathbb{N} \cup \{0\}\}$.

For each n, the image of K_n under ϕ_q is $\{[n-1, m(n+2)+1] : n \in \mathbb{N} \cup \{0\}\}$ which is a locally finite collection of subsets of $\mathbb{R}$.

Next we inductively define a sequence of continuous functions α_n on $\mathbb{R}^m$ for $n \in \mathbb{N} \cup \{0\}$, as follows:

Base step: $\alpha_0(\mathbf{x}) = 1$ for $\mathbf{x} \in [-1, 1]^m$, $\alpha_0(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathbb{R}^m \setminus K_0$.

Inductive step: $\alpha_n(\mathbf{x}) = 1 - \alpha_{n-1}(\mathbf{x})$ for $\mathbf{x} \in K_n \cap K_{n-1}$, $\alpha_n(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathbb{R}^m \setminus K_n$.

To prove the Claim, take any $f \in C(\mathbb{R}^m)$. Then $f(\mathbf{x}) = \sum_{i=0}^{\infty} \alpha_i(\mathbf{x}) \cdot f(\mathbf{x})$. Also $\alpha_i(\mathbf{x}) \cdot f(\mathbf{x}) = 0$ if $\mathbf{x} \notin K_i$.

From the Claim in the previous Step, for each $i \in \mathbb{N} \cup \{0\}$, there exist continuous functions g^i such that $\alpha_i(\mathbf{x}) \cdot f(\mathbf{x}) = \sum_{q=1}^{2m+1} g^i(\phi_q(\mathbf{x}))$.

Then let $g = \sum_{i=0}^{\infty} g^i$. This function is well-defined and continuous because $\{x : g^i(x) \neq 0\} \subseteq [i-1, m(i+2)+1]$, which means there are only finitely many i with $g^i(x) \neq 0$ for each $x \in \mathbb{R}$.

Then we have

$$f(\mathbf{x}) = \sum_{i=0}^{\infty} \alpha_i(\mathbf{x}) \cdot f(\mathbf{x}) = \sum_{i=0}^{\infty} \sum_{q=1}^{2m+1} g^i(\phi_q(\mathbf{x})) = \sum_{q=1}^{2m+1} g(\phi_q(\mathbf{x})),$$

— as claimed.

5. Lipschitz Elementary Functions

We conclude the proof by showing that the elementary functions, ψ_{pq} constructed above, can be modified so as to be Lipschitz, with Lipschitz constant 1. Recall that the ψ_{pq} are monotone increasing on $\mathbb{R}^+$ and monotone decreasing on $\mathbb{R}^-$.

Fix, for the moment, p between 1 and m. Define $\boldsymbol{\psi}_p : \mathbb{R} \to \mathbb{R}^{2m+1}$ by $\boldsymbol{\psi}_p(t) = (\psi_{p,1}(t), \ldots, \psi_{p,2m+1}(t))$. Then $\boldsymbol{\psi}_p$ is continuous. Let $C_p^+ = \boldsymbol{\psi}_p([0,\infty))$, $C_p^- = \boldsymbol{\psi}_p((-\infty, 0])$ and $C_p = C_p^+ \cup C_p^-$. Since the co-ordinates of $\boldsymbol{\psi}_p$ are monotone on $\mathbb{R}^+$ and $\mathbb{R}^-$, C_p^+, C_p^- and C_p are rectifiable. Let $ArcL(t)$ be the arc length along the curve C_p from $\boldsymbol{\psi}_p(0)$ to $\boldsymbol{\psi}_p(t)$. Then $\lambda_p : \mathbb{R} \to \mathbb{R}$ defined by $\lambda_p(t) = ArcL(t)$ for $t \geq 0$, otherwise, $\lambda_p(t) = -ArcL(t)$ is continuous and monotone. Let λ_p^{-1} be the (continuous, monotone) inverse of λ_p. Observe that, given $t, t' \in \mathbb{R}$, the distance between $\boldsymbol{\psi}_p(t)$ and $\boldsymbol{\psi}_p(t')$ along the curve C_p is $|\lambda_p(t) - \lambda_p(t')|$.

We verify that the functions $\psi_{pq} \circ \lambda_p^{-1}$ (for $p = 1, \ldots, m$ and $q = 1, \ldots, 2m+1$) are Lipschitz with Lipschitz constant 1.

To see this, fix q, fix p again, and take any s, s'. Without loss of generality, suppose $s' \leq s$. Let $t' = \lambda_p^{-1}(s')$ and $t = \lambda_p^{-1}(s)$. The distance along the curve C_p from $\boldsymbol{\psi}_p(\lambda_p^{-1}(s'))$ to $\boldsymbol{\psi}_p(\lambda_p^{-1}(s))$ is the distance along the curve from $\boldsymbol{\psi}_p(t')$ to $\boldsymbol{\psi}_p(t)$, which is $|\lambda_p(t) - \lambda_p(t')|$ (by definition of λ_p), and that equals $s - s' = |s - s'|$.

On the other hand, the distance along the curve C_p from $\boldsymbol{\psi}_p(\lambda_p^{-1}(s'))$ to $\boldsymbol{\psi}_p(\lambda_p^{-1}(s))$ is at least as large as the change from $\boldsymbol{\psi}_p(\lambda_p^{-1}(s'))$ to $\boldsymbol{\psi}_p(\lambda_p^{-1}(s))$ in just the qth coordinate. And the change in the qth coordinate, is $|(\psi_{pq}\circ\lambda_p^{-1})(s)-(\psi_{pq}\circ\lambda_p^{-1})(s')|$. So $|(\psi_{pq}\circ\lambda_p^{-1})(s)-(\psi_{pq}\circ\lambda_p^{-1})(s')| \leq |s-s'|$, as claimed.

It remains to show that the functions are elementary for $\mathbb{R}$ in dimension m (using just a single co–ordinate function).

Take any $f \in C(X^m)$. Let $f'(x_1', \ldots, x_m') = f(\lambda_1(x_1'), \ldots, \lambda_m(x_m'))$. Then, as the ψ_{pq} are elementary using a single co–ordinate function, there is a g in $C(\mathbb{R})$ such that $f'(x_1', \ldots, x_m') = \sum_{q=1}^{2m+1} g(\sum_{p=1}^{m} \psi_{pq}(x_p))$.

Hence $f(x_1, \ldots, x_m) = f'(\lambda_1^{-1}(x_1), \ldots, \lambda_m^{-1}(x_m)) = \sum_{q=1}^{2m+1} g(\sum_{p=1}^{m} \psi_{pq}(\lambda_p^{-1}(x_p)))$ $= \sum_{q=1}^{2m+1} g(\sum_{p=1}^{m}(\psi_{pq} \circ \lambda_p^{-1})(x_p))$, as required. $\square$

Remark 1: The theorem shows that for the space $X = \mathbb{R}$, in each dimension m there is an elementary family ψ_{pq} so that every f in $C(\mathbb{R}^m)$ can be written in the form $f = \sum_q g \circ \phi_q$ using a *single* co–ordinate function g. The same, of course, is true for $X = I$, but in this case it is essentially trivial as sketched below. This easy argument does not work for $X = \mathbb{R}$.

Suppose the maps ψ_{pq} are elementary for the closed unit interval, I. For each q, ϕ_q maps I to some $[a_q, b_q]$. Scaling and translating the original elementary functions we may suppose, without loss of generality, that the intervals $[a_q, b_q]$ are pairwise disjoint and contained in I. For each q, let $h_q : [a_q, b_q] \to I$ be a homeomorphism.

Take any f in $C(I^m)$. Then there are $g_1, \ldots, g_{2m+1}$ in $C(I)$ so that $f = \sum_q g_q \circ \phi_q$. Define g to be $g_q \circ h_q$ on $[a_q, b_q]$ and extend to a continuous function on I (this step is not, in general, possible for $X = \mathbb{R}$). Then $f = \sum_q g \circ \phi_q$, as required.

Remark 2: The argument given in Step 5 modifying the original elementary functions (which are definitely not Lipschitz) via arc length so as to become Lipschitz, is an elaboration of an idea of Kahane [17].

4.2 CHARACTERIZATION

Theorem 34. *Let X be a Tychonoff space. Then the following are equivalent:*

(1) some power of X has a finite basic family;

(2) for every $m, n \in \mathbb{N}$, there is an $r \in \mathbb{N}$ and ψ_{pq} from $C(X, \mathbb{R}^n)$, for $q = 1, \ldots, r$ and $p = 1, \ldots, m$, such that every $f \in C(X^m, \mathbb{R}^n)$ can be written

$$f(x_1, \ldots, x_m) = \sum_{q=1}^{r} g\left(\sum_{p=1}^{m} \psi_{pq}(x_p)\right),$$

for some $g \in C(\mathbb{R}^n, \mathbb{R}^n)$;

(3) X is a locally compact, finite dimensional separable metric space, or equivalently, homeomorphic to a closed subspace of Euclidean space.

Proof. It was shown in Theorem 6 that a Tychonoff space has a finite basic family if and only if it is a locally compact, finite dimensional separable metrizable space. Hence (1) implies (3), and (2) implies (1).

Now suppose (3) holds and X is a locally compact, finite dimensional separable metric space. Fix m. Then X is (homeomorphic to) a closed subspace of some $\mathbb{R}^\ell$. We establish (2) when $n = 1$. The general case follows easily by working co–ordinatewise.

According to Theorem 29 there exist ψ_{pq} for $p = 1, 2, \ldots, \ell m$ and $q = 1, 2, \ldots, 2\ell m + 1$ such that any $f \in C(\mathbb{R}^{\ell m})$ can be written as $f(x_1, \ldots, x_{\ell m}) = \sum_{q=1}^{2\ell m+1} g(\sum_{p=1}^{\ell m} \psi_{pq}(x_p))$ for some $g \in C(\mathbb{R})$.

Let $r = 2\ell m + 1$. Let $\Psi_{pq} = \sum_{i=1+(p-1)m}^{m+(p-1)m} \psi_{iq}$ for $p = 1, \ldots, m$ and $q = 1, \ldots, r$. Since X is a closed subset of $\mathbb{R}^\ell$, any continuous function on X can be continuously extended to $\mathbb{R}^\ell$. Then $\{\Psi_{pq} \upharpoonright X : p = 1, \ldots, m, \text{ and } q = 1, \ldots, r\}$ are as required. □

Note that from Theorem 34 (2) it follows that every continuous function of three complex variables can be written as a superposition of addition and continuous functions of one complex variable.

4.3 AN APPLICATION TO C_P–THEORY

Call two spaces X and Y ℓ–equivalent if there is a linear homeomorphism between $C_p(X)$ and $C_p(Y)$, and say that X ℓ–dominates Y if there is a continuous linear surjection of $C_p(X)$ onto $C_p(Y)$.

A beautiful result of Pestov [25] is that if two spaces X and Y are ℓ–equivalent then the (covering) dimension of X equals the (covering) dimension of Y. Arhangelskii asked whether it was true that if a space X ℓ–dominates another space Y, then $\dim(X) \geq \dim(Y)$. This natural conjecture was refuted by Leiderman et al. [24] who showed that the closed unit interval I ℓ–dominates every n–cube, I^n, using basic functions and a single co–ordinate function (very similarly to the argument below for $\mathbb{R}$).

Recently Gartside (private communication) has characterized the spaces ℓ–dominated by I as those which are compact, metrizable and strongly countable dimensional. Towards characterizing those spaces ℓ–dominated by the reals, we note the following consequence of Theorem 29.

Theorem 35. *There is a continuous linear surjection of $C_p(\mathbb{R})$ onto $C_p(\mathbb{R}^m)$ for any $m \in \mathbb{N}$. In other words, $\mathbb{R}$ ℓ-dominates $\mathbb{R}^m$ for every $m \in \mathbb{N}$.*

The same linear surjection is also continuous as a function of $C_k(\mathbb{R})$ to $C_k(\mathbb{R}^m)$,

Proof. Fix $m \in \mathbb{N}$. Then by the Theorem 29, there exist $\phi_1, \phi_2, \ldots, \phi_{2m+1} \in C(\mathbb{R}^m)$ such that any $f \in C(\mathbb{R}^m)$ can be represented as $f = \sum_{q=1}^{2m+1} g \circ \phi_q$ for some $g \in C(\mathbb{R})$. Hence we define the map $L : C_p(\mathbb{R}) \to C_p(\mathbb{R}^m)$ as $L(g) = \sum_{q=1}^{2m+1} g \circ \phi_q$. Obviously L is linear, and is surjective by the particular properties of the ϕ_q.

It is also easy to verify that L is continuous when the function spaces are either both given the topology of pointwise convergence, or both given the compact–open topology. □

5.0 CONSTRUCTIVE PROOF AND APPLICATIONS

Theorem 29 from the previous chapter says that every continuous real–valued function of m–real variables can be written as a superposition of continuous functions of one variable along with addition. From a theoretical point of view this is absolutely unexpected, and quite remarkable. However the proof of Theorem 29, as with Kolmogorov's proof of his Superposition Theorem does not give a computable algorithm.

The purpose of this Chapter is to present a genuinely computable variant of the Superposition Theorem for $\mathbb{R}^m$, and in doing so establish Theorem D from the Introduction. In Section 5.1 a family of effectively computable functions of the reals to the reals is given (Algorithm 36). Continuity and other properties of these functions are then verified. In the following Section 5.2 it is established that these functions are elementary, and moreover algorithms are presented and justified (Algorithm 45, Theorem 46 and Algorithm 50, Theorem 51) which given a continuous function $f : \mathbb{R}^m \to \mathbb{R}$ computes the corresponding co–ordinate functions in the Kolmogorov representation of f accurate to within a given error $\epsilon > 0$ on any specified compact subset of $\mathbb{R}^m$. These results are encapsulated in Theorem 53, which extends Theorem D of the Introduction.

These algorithms are such that if the Kolmogorov representation is calculated to within ϵ on compact set K, then if extra accuracy is required on K, or the error needs to be controlled on a larger compact set, then the existing approximation can be reused (so no unnecessary recalculation occurs). In Appendix B Python code is given implementing these algorithms for functions of two real variables. The Algorithms given here build on the proof of Theorem 29 and earlier work by Sprecher [32, 33], and Braun & Griebel [10] who gave constructive versions of Kolmogorov's Superposition Theorem.

To conclude in Section 5.3 we give a brief introduction to neural networks, and explain how the results of this Chapter have applications, both theoretical and practical, to the understanding and use of neural networks.

5.1 CONSTRUCTION OF THE FUNCTIONS

Fix the dimension m, and $\gamma \geq 2m+2$. Define for any $k \in \mathbb{N}$, $\mathcal{D}_k(\gamma)^+ = \{d_k = i_{1,k}/\gamma + \sum_{r=2}^{k} i_r \cdot \gamma^{-r} \in \mathbb{Q} : 0 \leq i_{1,k} \leq k \cdot \gamma^k - 1$ and $0 \leq i_j \leq \gamma - 1$ for $j \neq 1$ and $d_k \leq k\}$, $\mathcal{D}_k(\gamma)^- = \{d_k = i_{1,k}/\gamma + \sum_{r=2}^{k} i_r \cdot \gamma^{-r} \in \mathbb{Q} : -(k \cdot \gamma^k - 1) \leq i_{1,k} \leq 0$ and $-(\gamma-1) \leq i_j \leq 0$ for $j \neq 1$ and $-k \leq d_k\}$, and $\mathcal{D}_k = \mathcal{D}_k(\gamma) = \mathcal{D}_k(\gamma)^+ \cup \mathcal{D}_k(\gamma)^-$. Note that $\mathcal{D}_k \subseteq [-k, k]$.

Then the set of all rational numbers base γ, $\mathcal{D} = \{k/\gamma^\ell : k, \ell \in \mathbb{Z}\}$ (which is dense in $\mathbb{R}$) is the union over k of all the $\mathcal{D}_k$'s.

We define, functions from the reals to the reals, $\psi_1, \ldots, \psi_m$, first recursively in k on the set $\mathcal{D}_k$, and then extend over the whole of $\mathbb{R}$ by taking limits. At the same time, a sequence of positive numbers $(\epsilon_k)_k$ and sequences of natural numbers $(n_k)_k$, $(a_k)_k$, $(b_{k,s})_k$ ($s = 1, 2, \ldots, m$) are also introduced to control the functions. (The ϵ_k's are only needed for the following proofs, but the n_k's, a_k's and $b_{k,s}$'s play a key role in the definition of the functions ψ_q.)

Algorithm 36. *Define recursively numbers ϵ_k, a_k, n_k, $b_{k,1}, \ldots, b_{k,m}$ and functions on $\mathcal{D}_k$, $\psi_1, \ldots, \psi_m$.*

Base Step k=1: *Let $n_1 = 2$, $a_1 = 2$, and let $b_{1,s} = n_1 + (s-1)a_1$ for $s = 1, 2, \ldots, m$.*

Take any $d_1 = i_{1,1}/\gamma$ from $\mathcal{D}_1$

and set $\psi_s(d_1) = \begin{cases} 2 \cdot i_{1,1}/\gamma^{b_{1,s}} & \text{for } i_{1,1} \geq 0 \\ (-2 \cdot i_{1,1} + 1)/\gamma^{b_{1,s}} & \text{for } i_{1,1} < 0. \end{cases}$

Let $\epsilon_1 = 1/\gamma^{n_1 + (m-1)a_1 + 1}$.

Inductive Step: *Now suppose we have defined a_{k-1}, n_{k-1}, $b_{k-1,s}$ and ψ_s for $s = 1, 2, \ldots, m$, on $\mathcal{D}_{k-1}$.*

Let $a_k = k + \ulcorner \log_\gamma(2k) \urcorner + 1$ and $n_k = n_{k-1} + (m-1)a_{k-1} + 1 + a_k$, and let $b_{k,s} = n_k + (s-1)a_k$ for $s = 1, 2, \ldots, m$.

Take any $d_k = i_{1,k}/\gamma + \cdots + i_k/\gamma^k$ from $\mathcal{D}_k$. Set $d_{k-1} = d_k - i_k/\gamma^k$, and define some indexes of d_k by

$$\hat{i}_{d_k} = \begin{cases} 2 \cdot i_k & \text{for } d_k \geq 0 \\ 2 \cdot (-i_k) + 1 & \text{for } d_k < 0 \end{cases} \qquad C_{d_k} = \begin{cases} 2 \cdot i_{1,1} & \text{for } i_{1,1} \geq 0 \\ -2 \cdot i_{1,1} + 1 & \text{for } i_{1,1} < 0 \,, \end{cases}$$

and $$I_{d_k} = \begin{cases} 2 \cdot (\gamma^{k-1} i_{1,k} + \gamma^{k-2} i_2 + \cdots + i_k) & \text{for } d_k \geq 0 \,, \\ 2 \cdot (-\gamma^{k-1} i_{1,k} - \gamma^{k-2} i_2 - \cdots - i_k) + 1 & \text{for } d_k < 0 \end{cases}$$

Define ψ_s, for $s = 1, 2, \ldots, m$ in three cases depending on d_k.

1. $i_k \neq \pm(\gamma - 1) \wedge d_k \in \mathcal{D}_k \cap [-(k-1), k-1]$

For $s = 1, 2, \ldots, m$, we define $\psi_s(d_k) = \psi_s(d_k - i_k/\gamma^k) + \hat{i}_{d_k}/\gamma^{b_{k,s}}$

2. $d_k \in \mathcal{D}_k \cap [-k, -(k-1)) \cup (k-1, k]$

For $s = 1, 2, \ldots, m$, we define $\psi_s(d_k) = C_{d_k}/\gamma^{2s-1} + I_{d_k}/\gamma^{b_{k,s}}$

3. $i_k = \pm(\gamma - 1) \wedge d_k \in \mathcal{D}_k \cap (-(k-1), k-1)$

For each s, define $\psi_s(d_k) = 1/2(\psi_s(d_{k-1}) + \psi_s(d_k + d_k/(|d_k|\gamma^k))) + (I_{d_{k-1}} + \gamma)/\gamma^{b_{k,s}}$

Let $\epsilon_k = 1/\gamma^{n_k+(m-1)a_k+1}$.

The functions $\psi_1, \ldots, \psi_m$ are now defined on $\mathcal{D} = \bigcup_{k\in\mathbb{N}} \mathcal{D}_k$, we extend them over $\mathbb{R}$.

Every real number $x \in \mathbb{R}$ has a representation $x = \sum_{r=1}^{\infty} i_r/\gamma^r = \lim_k \sum_1^k i_r/\gamma^r$, and define, for $s = 1, \ldots, m$:

$$\psi_s(x) := \lim_{k\to\infty} \psi_s(\sum_{r=1}^{k} i_r/\gamma^r).$$

Notice that in this construction: if $a, b \in \mathcal{D}_k \cap ([-k, -(k-1)] \cup [k-1, k])$ are distinct, then $|\psi_s(a) - \psi_s(b)| > 1/(\gamma^{n_k+(s-1)a_k})$.

Proposition 37. *The functions $\psi_s, s = 1, 2, \ldots, m$ are monotonic increasing on $\mathbb{R}^+$, monotonic decreasing on $\mathbb{R}^-$ and continuous. (In particular they are well defined.)*

Proof. First, we will show the monotonicity properties of ψ_s for some $s = 1, 2, \ldots, m$.

By the definition in Algorithm 36, for each $k > n + 1$, ψ_s is strictly increasing on $\mathcal{D}_k \cap [0, n]$.

Then let $x = \sum_{r=1}^{\infty} i_r/\gamma^r$ and $x' = \sum_{r=1}^{\infty} i'_r/\gamma^r$. And suppose $x < x'$, then there exists $r_0 > n+1$ such that $\sum_{r=1}^{\ell} i_r/\gamma^r < \sum_{r=1}^{\ell} i'_r/\gamma^r$ for each $\ell \geq r_0$. Hence for $\ell \geq r_0$, $\psi_s(\sum_{r=1}^{\ell} i_r/\gamma^r) < \psi(\sum_{r=1}^{\ell} i'_r/\gamma^r)$. So

$$\psi_s(x) = \lim_{\ell\to\infty} \psi_s(\sum_{r=1}^{\ell} i_r/\gamma^r) \leq \lim_{\ell\to\infty} \psi_s(\sum_{r=1}^{\ell} i'_r/\gamma^r) = \psi_s(x').$$

Therefore, ψ_s is monotonic increasing on $[0, n]$ for each n, hence monotonic increasing on $\mathbb{R}^+$. Similarly, we can prove that ψ_s is monotonic decreasing on $\mathbb{R}^-$.

Now we will show the continuity of ψ_s for some $s = 1, 2, \ldots, m$. Fix $n \geq 2$, it is enough to prove ψ_s is continuous on $[0, n)$. Fix $k \geq n+1$. Then define $d^+_{k+j} = d_{k+j} + 1/\gamma^{k+j}$ for $j \in \mathbb{N}$, and define $\tau_j = \max\{\psi_s(d^+_{k+j}) - \psi_s(d_{k+j}) : d_{k+j}, d^+_{k+j} \in \mathcal{D}_{k+j}\}$. Then by the Algorithm 36 , we see that $\tau_{j+1} \leq \tau_j/2$ for $j \in \mathbb{N}$. Therefore, $\tau_j \leq \tau_0/2^j$.

Now take $x = \sum_{r=1}^{\infty} i_r/\gamma^r \in [0, n)$. Given arbitrary $\varepsilon > 0$, we need to find an open interval U containing x such that for any $y \in U$, $|\psi_s(x) - \psi_s(y)| < \varepsilon$.

Pick J such that $\tau_j \leq \tau_0/2^j < \varepsilon$ for $j \geq J$. Then because $\psi_s(x) = \lim_{\ell\to\infty} \psi_s(\sum_{r=1}^{\ell} i_r/\gamma^r)$, we can find $A > J$ such that $|\psi_s(\sum_{r=1}^{\ell} i_r/\gamma^r) - \psi_s(x)| < \varepsilon$ for $\ell \geq A+k$ and $\sum_{r=1}^{A+k} i_r/\gamma^r + 1/\gamma^{A+k} \in (0, n)$. Take U to be the interval $(\sum_{r=1}^{A+k} i_r/\gamma^r, \sum_{r=1}^{A+k} i_r/\gamma^r + 1/\gamma^{A+k}) \cap [0, n]$. It is easy to see that $x \in U$, and

$$\psi(\sum_{r=1}^{A+k} i_r/\gamma^r + 1/\gamma^{A+k}) - \psi(\sum_{r=1}^{A+k} i_r/\gamma^r) < \tau_0/2^{A+k} < \varepsilon.$$

Therefore, for any $y \in U$, $|\psi_s(x) - \psi_s(y)| < \varepsilon$, by the monotonicity properties of ψ_s. Hence ψ_s is continuous on $[0, n)$. Then ψ_s is continuous on $\mathbb{R}^+$. Similarly, we can prove that ψ_s is continuous $\mathbb{R}^-$. □

Definition 38. *Define ϕ in $C(\mathbb{R}^m)$ by $\phi(x_1, \ldots, x_m) = \psi_1(x_1) + \cdots + \psi_m(x_m)$.*

Lemma 39. *For distinct $\mathbf{d}$ and $\mathbf{d}'$ from $\mathcal{D}^m_k$, $|\phi(\mathbf{d}) - \phi(\mathbf{d}')| \geq 1/\gamma^{n_k+(m-1)a_k}$,* *Proof.* We will prove this by induction on k.

Base Case: $k = 1$. Here $\mathcal{D}_1 = \{i_{1,1}/\gamma : -(\gamma - 1) \leq i_{1,1} \leq (\gamma - 1)\}$, and the conclusion of the lemma follows immediately from the definition of the ψ_s.

Inductive Step. Suppose the conclusion is true for $k-1$. Next we will show this is also true for k. Suppose $\mathbf{d}, \mathbf{d}' \in \mathcal{D}_k$ are distinct. Then for some s, the sth coordinates of $\mathbf{d}$ and $\mathbf{d}'$, say d_s and d'_s, are distinct.

Case: $d_s, d'_s \in \mathcal{D}_k \cap [-(k-1), k-1]$

Suppose $d_s = i_1^k/\gamma + \sum_{j=2}^{k} i_j/\gamma^k$ and $d'_s = i'_{1,k}/\gamma + \sum_{j=2}^{k} i'_j/\gamma^k$.

If $i'_k = (\gamma - 1)$ or $i_k = (\gamma - 1)$ or $i'_k \neq i_k$, then $|\psi_s(d_s) - \psi_s(d'_s)| > 1/\gamma^{n_k+(s-1)a_k}$ by construction from which the claim follows.

If $i'_k = i_k$ and $i_k \neq (\gamma - 1)$, then $\psi_s(d_s) = \psi_s(d_s - i_k/\gamma^k) + i_k/\gamma^{n_k+(s-1)a_k}$. Therefore, $|\psi_s(d_s) - \psi_s(d'_s)| = |\psi_s(d_s - i_k/\gamma^k) - \psi_s(d'_s - i'_k/\gamma^k)| > 1/\gamma^{n_{k-1}+(m-1)a_{k-1}} > 1/\gamma^{n_k+(m-1)a_k}$ by hypothesis.

Case: $d_s, d'_s \in \mathcal{D}_k \cap [-k, -(k-1)) \cup (k-1, k]$

In this case, $|\psi_s(d_s) - \psi_s(d'_s)| > 1/\gamma^{n_k+(m-1)a_k}$ follows directly from the construction.

Case: $d_s \in \mathcal{D}_k \cap (-(k-1), k-1) \wedge d'_s \in \mathcal{D}_k \cap [-k, -(k-1)) \cup (k-1, k]$

In this case, $|\psi_s(d_s) - \psi_s(d'_s)| > 1/\gamma^{n_k+(m-1)a_k}$ follows directly from the construction.

□

Lemma 40. *For each integer $k \in \mathbb{N}$, let $\rho_k = (\gamma - 2)/((\gamma - 1) \cdot \gamma^k) = (\gamma - 2)/\gamma^k \cdot \sum_{j=1}^{\infty} 1/\gamma^j$. Then for all $d \in \mathcal{D}_k$ and $s = 1, 2, \ldots, m$, we have*

$$\psi_s(d + \rho_k) = \psi_s(d) + (\gamma - 2) \sum_{j=k+1}^{\infty} 1/\gamma^{b_{j,s}} < \psi_s(d) + \epsilon_k/2$$

A direct consequence of this lemma is given in the next lemma.

Lemma 41. *For fixed $k \in \mathbb{N}$ and $\mathbf{d} = (d_1, d_2, \cdots, d_m) \in \mathcal{D}_k^m$, the pairwise disjoint cubes*

$$S_k(\mathbf{d}) = E_k(d_1) \times E_k(d_2) \times \ldots \times E_k(d_m) \text{ where}$$

$$E_k(d_s) = [d_s, d_s + \rho_k] \text{ for } s = 1, 2, \ldots, m,$$

are mapped by ϕ into the pairwise disjoint intervals $T_k(\mathbf{d}) = [\phi(\mathbf{d}), \phi(\mathbf{d}) + \epsilon_k]$.

Definition 42. *Define* $\delta = \frac{1}{\gamma(\gamma-1)} = \sum_{r=2}^{\infty} \frac{1}{\gamma^r}$ *and* $\boldsymbol{\delta} = (\delta, \dots, \delta) \in \mathbb{R}^m$.

For $q = 0, 1, \dots, 2m$*, define* $\phi_q(\mathbf{x}) = \phi(\mathbf{x} + q\boldsymbol{\delta})$ *for* $\mathbf{x} \in \mathbb{R}^m$.

Fix k*. Define* $\delta_k = \sum_{r=2}^{k} 1/\gamma^r$*, and* $\boldsymbol{\delta}_k = (\delta_k, \dots, \delta_k) \in \mathbb{R}^m$.

For $q = 0, 1, \dots, 2m$*, define* $\phi_q^k(\mathbf{x}) = \phi(\mathbf{x} + q\boldsymbol{\delta}_k)$ *for* $\mathbf{x} \in \mathbb{R}^m$.

Definition 43. *For* $q = 0, 1, \dots, 2m$ *and* $\mathbf{d} = (d_1, d_2, \cdots, d_m) \in \mathcal{D}_k^m$*, define*

$$E_k^q(d_s) = [d_s + q\delta_k - q\delta, d_s + q\delta_k - q\delta + \rho_k]$$

for $s = 1, 2, \dots, m$ *and* $\rho_k = (\gamma - 2)/(\gamma - 1) \cdot 1/\gamma^k = (\gamma - 2)/\gamma^k \cdot \sum_{j=1}^{\infty} 1/\gamma^j$.

Define $S_k^q(\mathbf{d}) = E_k^q(d_1) \times E_k^q(d_2) \times \cdots \times E_k^q(d_m)$*, and* $T_k^q(\mathbf{d}) = [\phi_q^k(\mathbf{d}), \phi_q^k(\mathbf{d}) + \epsilon_k]$.

For $s = 1, 2, \dots, m$, we can see that $E_k^q(d_s)$ are separated by gaps $G_k^q(d_s) = (d_s + q\delta_k - q\delta + \rho_k, d_j + q\delta_k + \gamma^{-k})$ with width $1/(\gamma - 1) \cdot \gamma^{-k}$ for $d_s \in \mathcal{D}_k$. Further, the image of $S_k^q(\mathbf{d})$ for $\mathbf{d} \in \mathcal{D}_k^m$ under the mapping $\phi_q(\mathbf{x})$ is a subset of $T_k^q(\mathbf{d})$. It follows from Lemma 41, that $\{T_k^q(\mathbf{d}) : \mathbf{d} \in \mathcal{D}_k^m\}$ is a collection of disjoint closed intervals.

5.2 THE FUNCTIONS ARE ELEMENTARY

We now present the algorithm which implements the representation of an arbitrary continuous function f with support contained in the cube $[-N + 1, N - 1]^m$ as a superposition of single variable functions. Let $\| \cdot \|$ denote the maximum norm of bounded functions. Furthermore, let η be a fixed real number satisfying $1 > \eta > 2m/(2m + 1)$. Let $\xi = ((2m + 1)\eta - 2m)/(m + 1)$. Note that $0 < \frac{m+1}{2m+1}\xi + \frac{2m}{2m+1} \leq \eta < 1$.

Definition 44. *Fix* $\mathbf{d}$ *in* $\mathcal{D}_k^m$*. Define* $\omega(y; \mathbf{d}, q, k)$ *to be the piecewise linear function in the variable* y *which is identically equal to zero outside* $U_k(\mathbf{d}, q) = (\phi_q^k(\mathbf{d}) - \epsilon_{k+1}, \phi_q^k(\mathbf{d}) + \epsilon_k + \epsilon_{k+1})$ *and identically equal to one on* $T_k^q(\mathbf{d}) = [\phi_q^k(\mathbf{d}), \phi_q^k(\mathbf{d}) + \epsilon_k]$.

Algorithm 45. *Set* $f_0 = f$*, and* $g_0^0, \dots, g_{2m}^0 \equiv 0$.

For $r = 1, 2, 3, \dots$*, iterate the following steps until* $\eta^r \|f\|$ *is less than the desired error in the Kolmogorov representation* $\sum_{q=0}^{2m} g_q \circ \phi_q$ *of* f *where* $g_q = \sum_{i=0}^{r} g_q^i$*:*

I. Given the function f_{r-1}, determine an integer $k_r > N+2$ such that any two points $\mathbf{x}, \mathbf{x}' \in [-N,N]^m$ which satisfy $\|\mathbf{x}-\mathbf{x}'\| \le \gamma^{-k_r}$, it is true that $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \le \xi \|f_{r-1}\|$.

II. Set $g_0^r, \ldots, g_{2m}^r \equiv 0$.

For each $\mathbf{d}$ from $(\mathcal{D}_{k_r} \cap [-N,N])^m$:

- *Calculate $\tilde{f} = f_{r-1}(\mathbf{d})$.*
- *For $q = 0, 1, \ldots, 2m$:*

  (a) Compute $\phi_q^{k_r}(\mathbf{d}) - \epsilon_{k_r+1}, \phi_q^{k_r}(\mathbf{d}), \phi_q^{k_r}(\mathbf{d}) + \epsilon_{k_r}$, and $\phi_q^{k_r}(\mathbf{d}) + \epsilon_{k_r} + \epsilon_{k_r+1}$, and so compute the function $\omega(y; \mathbf{d}, q, k_r)$.

  (b) Add the term $\frac{1}{2m+1}\tilde{f} \cdot \omega(y; \mathbf{d}, q, k_r)$ to g_q^r.

Thus for each q,

$$g_q^r(y) = \frac{1}{2m+1} \sum f_{r-1}(\mathbf{d}) \omega(y; \mathbf{d}, q, k_r),$$

where the sum is over all $\mathbf{d} \in (\mathcal{D}_{k_r} \cap [-N,N])^m$.

III. Compute the function

$$f_r = f_{r-1} - \sum_{q=0}^{2m} g_q^r \circ \phi_q = f - \sum_{q=0}^{2m} \sum_{j=0}^{r} g_q^j \circ \phi_q.$$

That this algorithm does its job is established by the following result.

Theorem 46. *For $r = 1, 2, 3, \cdots$, there hold the following estimates:*

$$\|g_q^r\| \le \frac{1}{2m+1} \cdot \eta^{r-1} \|f\| \quad \textit{and} \quad \|f_r\| = \left\| f - \sum_{q=0}^{2m} \sum_{j=1}^{r} g_q^j \circ \phi_q \right\| \le \eta^r \|f\|.$$

Hence the functions $g_q = \sum_j g_q^j$ are well defined, continuous and satisfy $f = \sum_{q=0}^{2m} g_q \circ \phi_q$, and to calculate the Kolmogorov approximation $\sum_{q=0}^{2m} \left(\sum_{j=0}^{r} g_q^j \right) \circ \phi_q$ to f within an error $\epsilon > 0$ it suffices to iterate until $\eta^r \|f\| < \epsilon$.

From the definition of ω, we easily see:

Lemma 47. *For each q and r, g_q^r is continuous and the following estimate holds: $\|g_q^r\| \le \frac{1}{2m+1} \|f_{r-1}\|$.*

Thus Theorem 46 follows by induction from Lemma 47 and:

Theorem 48. *For the approximations f_r, $r = 0, 1, 2, \ldots$, defined in Algorithm 45, there holds the estimate*

$$\|f_r\| = \left\| f - \sum_{q=0}^{2m} \sum_{j=1}^{r} g_q^j \circ \phi_q \right\| \leq \eta \|f_{r-1}\|$$

.

Proof. Here f_{r-1}, f_r, N, r are as in Algorithm 45. Let $k_r > N + 2$ be the integer given in step I, so if $\|\mathbf{x} - \mathbf{x}'\|_{\max} \leq \gamma^{-k_r}$ then $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \xi \|f_{r-1}\|$. Fix q, for $\mathbf{d} \in \mathcal{D}_{k_r}^m$, the mapping ϕ_q associates to each $S_{k_r}^q(\mathbf{d})$ a unique image $T_{k_r}^q(\mathbf{d})$ on the real line and the images of any two squares form the set $\{S_{k_r}^q(\mathbf{d}) : \mathbf{d} \in \mathcal{D}_{k_r}^m\}$ have empty intersections. Now consider step I of Algorithm 45. Remember that $0 < \frac{m+1}{2m+1}\xi + \frac{2m}{2m+1} = \eta < 1$ where ξ and η are fixed.

Let $\mathbf{x} \in [-N, N]^m$ be an arbitrary point, then there are $m + 1$ values of q in $\{0, \ldots, 2m\}$ such that there is some $\mathbf{d} \in (\mathcal{D}_{k_r} \cap [-N, N])^m$ such that $\mathbf{x} \in S_{k_r}^q(\mathbf{d})$. List these q as $\tilde{q}^j$ for $j = 1, 2, \ldots, m + 1$, and let $\mathbf{d}^j$ be the corresponding elements in $(\mathcal{D}_{k_r} \cap [-N, N])^m$.

Now fix j. Since $\mathbf{d}^j \in S_{k_r}^{\tilde{q}_j}(\mathbf{d}^j)$, it follows that $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{d}^j))| \leq \xi \|f_{r-1}\|$.

Also, for this $\mathbf{x}$, we have that $\phi_{\tilde{q}_j}(\mathbf{x}) \in T_k^{\tilde{q}_j}(\mathbf{d}^j)$, so by definition of ω, $\omega(y; \mathbf{d}^j, \tilde{q}_j, k_r) \equiv 1$ on $T_k^{\tilde{q}_j}(\mathbf{d}^j)$. Therefore, $g_{\tilde{q}_j}^r(\phi_{\tilde{q}_j}(\mathbf{x})) = \frac{1}{2m+1} f_{r-1}(\mathbf{d}^j)$. This shows $|\frac{1}{2m+1} f_{r-1}(\mathbf{x}) - g_{\tilde{q}_j}^r(\phi_{\tilde{q}_j}(\mathbf{x}))| \leq \frac{\xi}{2m+1} \|f_{r-1}\|$ for $j = 1, 2, \ldots, m + 1$.

Note that this estimate does not hold for the remaining values of q for which there might not exist $\mathbf{d} \in (\mathcal{D}_{k_r} \cap [-N, N])^m$ such that $\mathbf{x} \in S_{k_r}^q(\mathbf{d})$. Let us now denote these values by $\bar{q}_i$, $i = 1, 2, \ldots, m$. By Lemma 47, we have $\|g_{\bar{q}_i}^r\| \leq \frac{1}{2m+1} \|f_{r-1}\|$.

Then with the special choice of the values ξ and η we obtain the estimate

$$\begin{aligned}
\|f_r\| &= \left\| f_{r-1} - \sum_{q=0}^{2m} g_q^r \circ \phi_q \right\| \\
&= \left\| \sum_{q=0}^{2m} \frac{1}{2m+1} f_{r-1} - \sum_{j=1}^{m+1} g_{\tilde{q}_j}^r \circ \phi_{\tilde{q}_j} - \sum_{i=1}^{m} g_{\bar{q}_i}^r \circ \phi_{\bar{q}_i} \right\| \\
&\leq \left\| \frac{m}{2m+1} f_{r-1} + \sum_{j=1}^{m+1} \frac{1}{2m+1} f_{r-1} - \sum_{j=1}^{m+1} g_{\tilde{q}_j}^r \circ \phi_{\tilde{q}_j} \right\| + \frac{m}{2m+1} \|f_{r-1}\| \\
&\leq [\frac{m+1}{2m+1}\xi + \frac{2m}{2m+1}] \|f_{r-1}\| \\
&\leq \eta \|f_{r-1}\|.
\end{aligned}$$

This complete the proof of Theorem 48. □

Next, we can use this algorithm to implement the representation of an arbitrary continuous multivariate function f defined on $\mathbb{R}^m$ as superposition of single variable functions. First some definitions.

Definition 49. *Let* $K_n = \bigcup_{s=1}^{m}\{(x_1, x_2, \cdots, x_m) : -n-1 \leq x_j \leq n+1$ *for* $j \neq s; n-1 \leq x_s \leq n+1$ *or* $-n-1 \leq x_s \leq -n+1\}$ *where* $n > 0$*.*

Define $\alpha_n : \mathbb{R}^m \to \mathbb{R}$ *by:*

$$\alpha_1(\mathbf{x}) = \begin{cases} 1 & \textit{for } \mathbf{x} \in [-1,1]^m \\ 2 - \|\mathbf{x}\| & \textit{for } \mathbf{x} \in [-2,2]^m \setminus [-1,1]^m \\ 0 & \textit{for } \mathbf{x} \notin [-2,2]^m \end{cases},$$

and for $n > 1$*:*

$$\alpha_n(\mathbf{x}) = \begin{cases} \|\mathbf{x}\| - n + 1 & \textit{for } \mathbf{x} \in K_n \cap K_{n-1} \\ n + 1 - \|\mathbf{x}\| & \textit{for } \mathbf{x} \in K_n \setminus K_{n-1} \\ 0 & \textit{for } \mathbf{x} \notin K_n \end{cases}.$$

Algorithm 50. *Given* $f \in C(\mathbb{R}^m)$*,* $\epsilon > 0$ *and* N*, construct* $g_0, \ldots, g_{2m}$ *in* $C(\mathbb{R})$ *such that* $|f(\mathbf{x}) - \sum_{q=0}^{2m} g_q(\phi_q(\mathbf{x}))| < \epsilon$ *for all* $\mathbf{x}$ *in* $[-N, N]^m$*, as follows:*

I. Compute $f_n = \alpha_n \cdot f$ *for* $n = 1, \ldots, N+1$*.*

II. For each $n \leq N+1$*, apply Algorithm 45 to* f_n *on* $[-(n+1), n+1]^m$ *to get continuous functions* g_q^n *(*$q = 0, \ldots, 2m$*) so that* $\|f_n - \sum_{q=0}^{2m} g_q^n \circ \phi_q\| < \epsilon/(N+1)$ *on* $[-(n+1), n+1]^m$*.*

III. Calculate $g_q = \sum_{n=1}^{N+1} g_q^n$*.*

This algorithm does what is claimed.

Theorem 51. *In the notation of Algorithm 50 above, we have that the* g_q *are well–defined, continuous and are such that* $|f(\mathbf{x}) - \sum_{q=0}^{2m} g_q(\phi_q(\mathbf{x}))| < \epsilon$ *for all* $\mathbf{x}$ *in* $[-N, N]^m$*.*

Proof. First note that, for every n, α_n is continuous and has support contained in K_n, so f_n is continuous and has support contained in K_n, and so contained in $[-(n+1), n+1]^m$. Thus we can indeed apply Algorithm 45 in step II, to get the g_q^n with the claimed properties.

Hence the g_q in step III are continuous. Since $\sum_{n=1}^{\infty} \alpha_n \equiv 1$ everywhere, and $\sum_{n=1}^{N+1} \alpha_n \equiv 1$ on $[-N, N]^m$. Hence, on $[-N, N]^m$, $f = \sum_{n=1}^{N+1} f_n$, and so for $\mathbf{x}$ in $[-N, N]^m$,

$$\left| f(\mathbf{x}) - \sum_{q=0}^{2m} g_q(\phi_q(\mathbf{x})) \right| = \left| \sum_{n=1}^{N+1} \left(f_n(\mathbf{x}) - \sum_{q=0}^{2m} g_q^n(\phi_q(\mathbf{x})) \right) \right| < (N+1)\epsilon/(N+1) = \epsilon,$$

— as required. □

Lemma 52. *For any* $\mathbf{x} \in K_n$, $\frac{20(n-1)-2}{\gamma^{2m}} \leq \phi_q(\mathbf{x}) \leq \sum_{s=1}^{m} \frac{20(n+1)+3}{\gamma^{2s}}$.

Proof. Since $K_n = \bigcup_{s=1}^{m} \{(x_1, x_2, \cdots, x_m) : -n-1 \leq x_j \leq n+1$ for $j \neq s; n-1 \leq x_s \leq n+1$ or $-n-1 \leq x_s \leq -n+1\}$ where $n \geq 1$, the minimal value of ϕ on K_n is obtained at the point $(0, 0, \cdots, 0, n-1)$ where it has value $\frac{20(n-1)}{\gamma^{2m}}$.

The maximal value of ϕ on K_n is obtained at the point $(-n-1, -n-1, \cdots, -n-1, -n-1)$, where it has value $\sum_{s=1}^{m} \frac{20(n+1)+2}{\gamma^{2s}}$.

Then the claim follows immediately from the definition of ϕ_q, and the monotonicity properties the of ψ_s for $s = 1, 2, \ldots, m$. □

Finally, we are in the position to prove the main theorem of this Chapter:

Theorem 53. *Let* $m \geq 2$ *and* $\gamma \geq 2m+2$. *Set* $\delta = \frac{1}{\gamma(\gamma-1)}$ *and* $\mathcal{D} = \{k/\gamma^{\ell} : k, \ell \in \mathbb{Z}\}$.

Then there are functions, given by Algorithm 36, $\psi_1, \psi_2, \ldots, \psi_m$ *in* $C(\mathbb{R})$ *which are effectively computable on the dense set* $\mathcal{D}$ *of* $\mathbb{R}$*, such that: for an arbitrary continuous* $f \in C(\mathbb{R})$*, there exist* $2m+1$ *continuous functions* g_q*,* $q = 0, \ldots, 2m$ *such that*

$$f = \sum_{q=0}^{2m} g_q \circ \phi_q, \quad \textit{where } \phi_q(x_1, \ldots, x_m) = \sum_{s=1}^{m} \psi_s(x_s + q\delta).$$

Further the functions g_q *can be effectively computed to within any given error* $\epsilon > 0$ *on any specified compact subset of* $\mathbb{R}^m$*, by applying Algorithm 50 (and Algorithm 45).*

Proof. Everything claimed has already been established in Theorems 46 and 51 — except that the functions g_q exist, are continuous and are such that $f = \sum_{q=0}^{2m} g_q \circ \phi_q$.

Given a function $f \in C(\mathbb{R}^m)$, we can write f as a sum of compactly supported family of functions f_n where $f_n = \alpha_n \cdot f$. For each n, we can find functions g_q^n from Theorem 46 so that $f_n = \sum_{q=0}^{2m} g_q^n \circ \phi_q$. Define $g_q = \sum_{n=1}^{\infty} g_q^n$.

By the Lemma 52, $g_q^n(y) \equiv 0$ if $y > \sum_{s=1}^{m} \frac{20(n+1)+3}{\gamma^{2s}}$ or $y < \frac{20(n-1)-2}{\gamma^{2m}}$. So $g_q(y)$ is a finite sum for each value of $y \in \mathbb{R}$. Then by the continuity of each g_q^n, it follows that g_q exists is continuous at every point y. And since, $\sum \alpha_n \equiv 1$, $f = \sum f_n = \sum_{q=0}^{2m} g_q \circ \phi_q$, as required. □

5.3 NEURAL NETWORKS

A neural network is a way to perform computations using networks of interconnected computational units vaguely analogous to neurons simulating how our brain solves them. A 'neuron' in a neural net is a device with m real inputs $x_1, \ldots, x_m$ and an output $y = g(w_1x_1 + \ldots + w_mx_m + w_0)$. Here, $g(x)$ is a function that is called an activation function, and parameters w_i are called weights (w_0 is also called a threshold). If we send the output of some neurons as inputs to others, we get a neural network.

Two fundamental questions about neural networks arise, in essence they ask how powerful a neural network can be in theory, and in practice. Let X be a subset of $\mathbb{R}$. Let us say that neural networks are *universal for* X if every continuous function $f : X^m \to \mathbb{R}$ can be exactly computed by a neural network, and they are *approximately universal for* X if every continuous function $f : X^m \to \mathbb{R}$ can be computed arbitrarily well by a neural network.

The history of neural networks started with a lot of hype and excitement, as researchers started investigating two layer neural networks (also known as perceptrons). This period came to an abrupt end when it was shown that perceptrons were extremely limited in the functions they could compute.

Interest returned to neural networks when Hecht–Nielsen [12, 13, 14] noticed that Kolmogorov's Superposition Theorem shows that four layer neural networks are universal for compact intervals.

Later Kurkova [19], among many others, developed approximate versions of Kolmogorov's Superposition Theorem which give algorithms for constructing neural nets approximating a given function. Neural nets are now very actively studied and used.

As we remarked before when discussing the restriction in Kolmogorov's Superposition Theorem to functions on a compact cube, it makes little sense, and may well be very inconvenient, to restrict neural nets to only have inputs from a compact interval.

Theorem 29 and the algorithms of this Chapter remove this unnatural restriction:

Theorem 54. *Let X be any closed subset of the reals.*

- *Four layer neural networks are universal for X.*
- *There is a constructive algorithm witnessing that four layer neural networks are approximately universal for X.*

To prove this theorem we simply sketch how, given a continuous function f of two variables, to connect together a four layer neural network computing the Kolmogorov representation of f. The more general results are immediate.

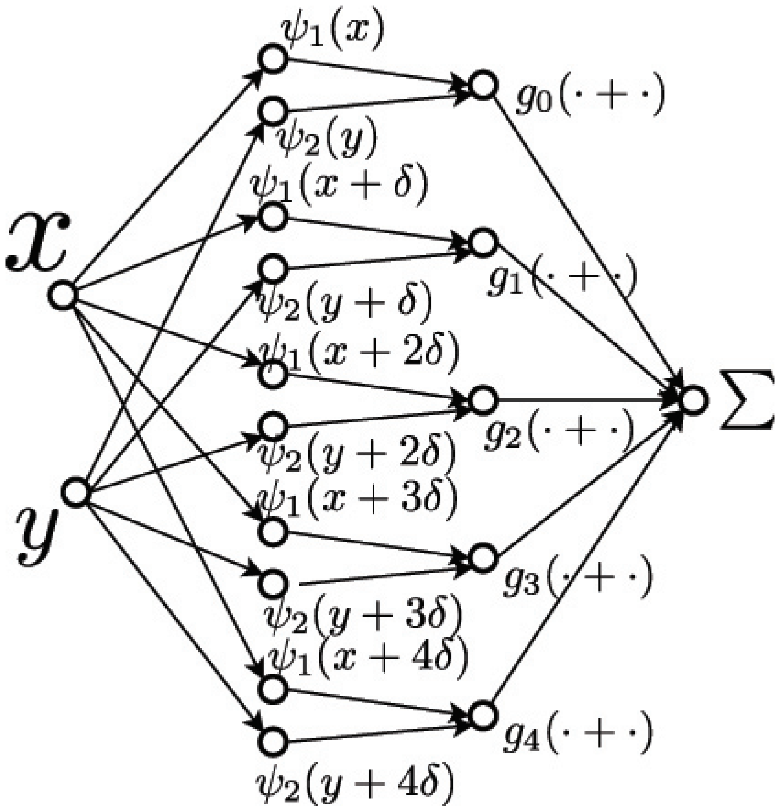

Figure 5.1: Neural Network

6.0 OPEN QUESTIONS AND PROPOSED RESEARCH

My results presented in this thesis on spaces with a finite basic or elementary families are complete. Theorem 29 shows that a space has a finite basic family if and only if it has a finite elementary family, and this occurs if and only if the space is homeomorphic to a closed subspace of Euclidean space.

However, a number of open problems remain. In this chapter, I will present some interesting open problems related to Hilbert's 13th problem along with my future research plan.

6.1 SMOOTH FUNCTIONS AND ANALYTIC FUNCTIONS

Hilbert in posing the 13th Problem remarked that there is an analytic function of 3 variables which can not be represented as a superposition of analytic functions of 2 variables. Ostrowski subsequently proved that the analytic function $\xi(x,y) = \sum_{n=1}^{\infty} \frac{x^n}{n^y}$ can not be represented as a superposition of infinitely differentiable functions of one variable and algebraic functions of arbitrarily many variables. In the 1930's Hilbert studied the algebraic aspect of his 13th Problem, showing, for example, that the solution of the general equation of degree 9 can be represented as a superposition of algebraic functions of 4 variables (down from the 5 obtained by applying Tschirnhaus transformations). Later, in 1954, Vitushkin gave partial confirmation to Hilbert's intuition that some functions are *irreducibly* of 3 or more variables. Let f be an r-times continuously differentiable function of n–variables. Vitushkin [37] showed that the characteristic $\chi = r/n$ can be used to measure the complexity of a class of functions as follows:

Theorem 55. *If $\chi = r/n > r_0/n_0 = \chi_0 > 0$ with $r \geq 1$, then there are functions of characteristic χ which can not be represented as a superposition of functions of characteristic χ_0.*

The following questions are very natural:

Question 1 (Vitushkin)**.** *Can every analytic, or C^∞, function of 2 variables can be represented as a superposition of continuously differentiable functions of one variable and the operation of addition?*

Question 2 (Arnold [3])**.** *Is the converse of Vitushkin's Theorem true, namely: if $\chi = r/n \leq r_0/n_0 = \chi_0 > 0$ with $r \geq 1$, then can every function of characteristic χ be represented as a superposition of functions of characteristic χ_0?*

Vitushkin repeated his question in a very recent paper [38]. My conjectural answer is 'no' to both these questions.

6.2 MINIMAL BASIC FAMILIES

The results on basic (X) are complete when X is separable metrizable, but there is an inconvenient gap for compact X — is the restriction to 'nice' compacta in Proposition 27 necessary?

Question 3. *Is it true that basic $(K) \geq cof([w(K)]^{\aleph_0}, \subseteq)$ for all compact spaces K?*

The proofs of the results for compact spaces clearly rely on facts and techniques that only apply to compact spaces. But it seems possible that the results could be extended to larger classes of spaces.

Question 4. *Do the results for basic (K) for compact K hold for (1) locally compact, $Lindel\ddot{o}f$ spaces or even (2) all $Lindel\ddot{o}f$ spaces?*

In a different direction, what about discrete spaces?

Question 5. *Is basic $(D(\aleph_1)) = \aleph_1$? $= 2^{\aleph_0}$?*

6.3 CONSTRUCTION OF LIPSCHITZ BASIC OR ELEMENTARY FUNCTIONS AND APPLICATIONS

In the construction of Chapter 5, the elementary functions ψ_{pq} are not Lipschitz. This reduces their value for applications. Further the co–ordinate functions, g_q appear to be highly irregular, how bad are they?

Problem 6. *In the constructive versions of basic or elementary families, improve the basic or elementary functions to be Lipschitz and then explore more applications.*

For a smooth function f how wild are the co-ordinate functions produced by the Contructive Algorithm? Can they be made to be Lip $-\alpha$?

Also, in the constructive proof of Theorem 29, the elementary functions are well-defined and continuous. However the co–ordinate functions g_q are given by infinite number of iterations. It would be very useful to fix g_q at a dense set of $\mathbb{R}$ in finite steps. This will also enhance the application of the Theorem 29 enormously.

One Application: Wavelet image decompositions

Most of the signal processing techniques are applied in 1D or 2D and they can not easily extended to higher dimensions. Using the Theorem 29, any multivariate function can be decomposed into two types of univariate functions, inner and external functions.

APPENDIX A

HILBERT'S 13TH PROBLEM

David Hilbert presented a lecture to the International Congress of Mathematicians at Paris in 1900, titled *Mathematical Problems.* In this lecture he laid out his famous list of 23 'Hilbert Problems'. The lecture was published in the Göttinger Nachrichten, 1900, pp. 253-297, and in the Archiv der Mathernatik und Physik, 3d ser., vol. 1 (1901), pp. 44-63 and 213-237, and subsequently translated from the original German by Dr Mary Newson for the Bulletin of the American Math Society.

Here is the text for the 13th Problem:

13. IMPOSSIBILITY OF THE SOLUTION OF THE GENERAL EQUATION OF THE 7TH DEGREE BY MEANS OF FUNCTIONS OF ONLY TWO ARGUMENTS.

Nomography[1] deals with the problem: to solve equations by means of drawings of families of curves depending on an arbitrary parameter. It is seen at once that every root of an equation whose coefficients depend upon only two parameters, that is, every function of two independent variables, can be represented in manifold ways according to the principle lying at the foundation of nomography. Further, a large class of functions of three or more variables can evidently be represented by this principle alone without the use of variable elements, namely all those which can be generated by forming first a function of two arguments, then equating each of these arguments to a function of two arguments, next replacing each of those arguments in their turn by a function

[1] d'Ocagne, Traité de Nomographie, Paris, 1899.

of two arguments, and so on, regarding as admissible any finite number of insertions of functions of two arguments. So, for example, every rational function of any number of arguments belongs to this class of functions constructed by nomographic tables; for it can be generated by the processes of addition, subtraction, multiplication and division and each of these processes produces a function of only two arguments. One sees easily that the roots of all equations which are solvable by radicals in the natural realm of rationality belong to this class of functions; for here the extraction of roots is adjoined to the four arithmetical operations and this, indeed, presents a function of one argument only. Likewise the general equations of the 5th and 6th degrees are solvable by suitable nomographic tables; for, by means of Tschirnhausen transformations, which require only extraction of roots, they can be reduced to a form where the coefficients depend upon two parameters only.

Now it is probable that the root of the equation of the seventh degree is a function of its coefficients which does not belong to this class of functions capable of nomographic construction, i.e., that it cannot be constructed by a finite number of insertions of functions of two arguments. In order to prove this, the proof would be necessary that the equation of the seventh degree $f^7 + xf^3 + yf^2 + zf + 1 = 0$ is not solvable with the help of any continuous functions of only two arguments. I may be allowed to add that I have satisfied myself by a rigorous process that there exist analytical functions of three arguments x, y, z which cannot be obtained by a finite chain of functions of only two arguments.

By employing auxiliary movable elements, nomography succeeds in constructing functions of more than two arguments, as d'Ocagne has recently proved in the case of the equation of the 7th degree.

APPENDIX B

PYTHON CODE

In this Appendix Python code implementing the Algorithms of Chapter 5 are presented. Only functions of two variables will dealt with ($m = 2$), and γ will be taken to be 10.

Python was used as it is a high level language, treating functions as first–order objects, and has a succinct and descriptive notation. Additionally Python has a built–in module for exact decimal arithmetic. (This is important because standard floating point arithmetic is inexact and would cause the algorithms to fail.)

We start, then, by importing the decimal and math packages.

```
from decimal import *
from math import log, ceil
```

Next some useful functions for dealing with functions. The first is the function which is identically zero. Then there is a function which adds two functions, another which multiplies two functions. Lastly there is a function which takes a list of pairs of decimals and creates the function which is the piecewise linear interpolate through these points.

```
def identically_zero_fn(x): return Decimal('0.0')


def add_fn(f,g):      return lambda x: f(x)+g(x)
```

```
def multiply_fn(f,g):    return lambda x: f(x)*g(x)


def piecewise_fn_from_list(g):
    def fn_g(x):
        i0, i1=0, len(g)-1
        while (1<>0):
            sm_pt, sm_val= g[i0]
            if (x<=sm_pt):  return sm_val
            lg_pt, lg_val= g[i1]
            if (x>=lg_pt):  return  lg_val
            if ((i1-i0)==1):
                if (sm_val == lg_val):  return sm_val
                return (x-sm_pt)*(lg_val-sm_val)/(lg_pt-sm_pt)+
                    sm_val
            i_mid=i0+(i1-i0)/2
            mid_pt, mid_val=g[i_mid]
            if (x <= mid_pt):    i1=i_mid
            if (x> mid_pt):      i0=i_mid
    return fn_g
```

The following function takes a decimal d, and returns a pair whose first component is the minimal k so d is in $\mathcal{D}_k$ and second component is d's representation as an element of $\mathcal{D}_k$ (see the definition of $\mathcal{D}_k$).

```
def decimal2dk(d):
    d=d.normalize()
    if (d==0):  return (0,)
    dt=d.as_tuple()
    k=max(abs(int(d))+int(ceil(abs(d-int(d)))), -(dt[2]))
    sn=1-2*(dt[0])
```

```
    front_dgs=int(d*10)
    if (k==1):  return (1, (front_dgs,))
    rem_str=str(((d*10)-front_dgs).normalize())
    if (rem_str=='0'):  rem_dgs=()
    else:   rem_dgs=tuple(sn*int(dg) for dg in (rem_str.split('.
        '))[1])
    pad=k-len(rem_dgs)-1
    pad_dgs=tuple(0 for n in range(pad))
    all_dgs=(front_dgs,)+rem_dgs+pad_dgs
    return (k,all_dgs)
```

Now we can get down to implementing Algorithm 36. First the sequences of a_k's and the n_k's. (The $b_{n,k}$'s are subsumed in the following definition of ψ_1 and ψ_2.) Then the functions `psi_one` (ψ_1) and `psi_two` (ψ_2), both of which are functions of decimals to decimals.

```
def a(k):
    if (k==1):  return 2
    else:       return k+int(ceil(log(2*k,10)))+1
def n(k):
    if (k==1):  return 2
    else:       return n(k-1)+a(k-1)+a(k)+1


def psi_one(d):
    if (d==0):  return Decimal("0")
    # otherwise d<>), and have more work to do
    k,i = decimal2dk(d)
    if (k==1):
        if (i[0]<0):    return Decimal(-2*i[0]+1)/(10**n(1))
        else:       return Decimal(2*i[0])/(10**n(1))
    # otherwise k>1, and proceed inductively...
```

```
ik=i[k-1]
dp=d-Decimal(ik)/(10**k)
if (d<0):      i_hat=1+2*(-ik)
else:          i_hat=2*ik
if (d<0):      big_I=2*sum(-i[j]*(10**(k-1-j)) for j in range(k)
    )+1
else:          big_I=2*sum(i[j]*(10**(k-1-j)) for j in range(k))
    if (d<0):
               big_I_one=(-2*i[0]+1)*(10**(k-1))+2*sum(-i[j
                  +1]*(10**(k-j-2)) for j in range(k-1))+1
    else:      big_I_one=2*sum(i[j]*(10**(k-1-j)) for j in range
        (k))
if (abs(d)<(k-1)):
    if (abs(ik)<>9):
                    return (psi_one(dp)+Decimal(i_hat)/(10**n
                       (k)))
               elif (d<0):
                    return ((psi_one(dp)+psi_one(d-Decimal
                       (1)/(10**k))))/2+Decimal(big_I+10)
                       /(10**n(k))
               elif (d>0):
                    return ((psi_one(dp)+psi_one(d+Decimal
                       (1)/(10**k))))/2+Decimal(big_I+10)
                       /(10**n(k))
else:
    # otherwise d not in ( -(k-1), (k-1) )
               return (Decimal(big_I_one)/(10**(k+1))+Decimal(
                  big_I)/(10**n(k)))
```

```
def psi_two(d):
    if (d==0):  return Decimal("0")
    k,i = decimal2dk(d)
    if (k==1):
        if (i[0]<0):      return Decimal(-2*i[0]+1)/(10**(n(1)+a(1)
          ))
        else:          return Decimal(2*i[0])/(10**(n(1)+a(1)))
    # otherwise k>1, and proceed inductively...
    ik=i[k-1]
    dp=d-Decimal(ik)/(10**k)
    if (d<0):    i_hat=1+2*(-ik)
    else:        i_hat=2*ik
    if (d<0):    big_I=2*sum(-i[j]*(10**(k-1-j)) for j in range(k)
      )+1
    else:        big_I=2*sum(i[j]*(10**(k-1-j)) for j in range(k))
        if (d<0):
                 big_I_one=(-2*i[0]+1)*(10**(k-1))+2*sum(-i[j
                   +1]*(10**(k-j-2)) for j in range(k-1))+1
        else:    big_I_one=2*sum(i[j]*(10**(k-1-j)) for j in range
          (k))
    if (abs(d)<(k-1)):
        if (abs(ik)<>9):
            return psi_two(dp)+Decimal(i_hat)/(10**(n(k)+a(k)))
            if (d<0):
                        return  ((psi_two(dp)+psi_two(d-Decimal
                          (1)/(10**k))))/2+Decimal(big_I+10)
                          /(10**(n(k)+a(k)))
```

```
                else:
            return  ((psi_two(dp)+psi_two(d+Decimal(1)/(10**k))))
               /2+Decimal(big_I+10)/(10**(n(k)+a(k)))
    else:
        # otherwise d not in ( -(k-1), (k-1) )
                return (Decimal(big_I_one)/(10**(k+3))+Decimal(
                    big_I)/(10**(n(k)+a(k))))
```

Now for the implementation of Algorithm 45. This is broken into three parts: first calculate one step of the iteration (`one_iteration_step`), second the computation of the new function (`new_f`), and third a complete implementation of the algorithm finding the Kolmogorov approximation to a compactly supported function (`cptly_supp_k`).

`one_iteration_step(A,f,k)` takes as its inputs a function f taking two decimals and returning a decimal, which is supported on the square $[-A, A]^2$, and an integer k. It returns the 5 functions g_0, g_1, g_2, g_3, g_4 as in the iterative step of Algorithm 45.

```
def one_iteration_step(A,f,k):
    g0,g1,g2,g3,g4=[],[],[],[],[]

    eps_big=Decimal('1.0')/(10**(n(k)+a(k)+1))
    eps_small=Decimal('1.0')/(10**(n(k+1)+a(k+1)+1))

    delta=sum(Decimal('1.0')/(10**r) for r in range(2,k+1))
    Delta=Decimal('1.0')/(10**k)
    d1=-Decimal(A)
    while (d1 <Decimal(A)):
        psi1d=psi_one(d1)
        psi1d1=psi_one(d1+delta)
        psi1d2=psi_one(d1+2*delta)
        psi1d3=psi_one(d1+3*delta)
```

```
        psi1d4=psi_one(d1+4*delta)
        d2=-Decimal(A)
        while (d2 < Decimal(A)):
            fd=f(d1,d2)
            phi_d0=psi1d+psi_two(d2)
            g0[len(g0):]=[(phi_d0-eps_small,0), (phi_d0,fd/5), (
                phi_d0+eps_big, fd/5), (phi_d0+eps_big+eps_small
                ,0)]
            phi_d1=psi1d1+psi_two(d2+delta)
            g1[len(g1):]=[(phi_d1-eps_small,0), (phi_d1,fd/5), (
                phi_d1+eps_big, fd/5), (phi_d1+eps_big+eps_small
                ,0)]
            phi_d2=psi1d2+psi_two(d2+2*delta)
            g2[len(g2):]=[(phi_d2-eps_small,0), (phi_d2,fd/5), (
                phi_d2+eps_big, fd/5), (phi_d2+eps_big+eps_small
                ,0)]
            phi_d3=psi1d3+psi_two(d2+3*delta)
            g3[len(g3):]=[(phi_d3-eps_small,0), (phi_d3,fd/5), (
                phi_d3+eps_big, fd/5), (phi_d3+eps_big+eps_small
                ,0)]
            phi_d4=psi1d4+psi_two(d2+4*delta)
            g4[len(g4):]=[(phi_d4-eps_small,0), (phi_d4,fd/5), (
                phi_d4+eps_big, fd/5), (phi_d4+eps_big+eps_small
                ,0)]
            d2=d2+Delta
        d1=d1+Delta
    return (g0,g1,g2,g3,g4)
```

```
def new_f(f,G0,G1,G2,G3,G4,d):
    return lambda x,y: f(x,y)-G0(psi_one(x)+psi_two(y))-G1(
        psi_one(x+d)+psi_two(y+d))-G2(psi_one(x+2*d)+psi_two(y+2*d
        ))-G3(psi_one(x+3*d)+psi_two(y+3*d))-G4(psi_one(x+4*d)+
        psi_two(y+4*d))
```

`cptly_supp_k(A,f,delta,M,error)` takes a positive integer A, a function f taking pairs of decimals to a decimal, which has support contained in $[-A, A]^2$, a function `delta` taking decimals to decimals which is a 'delta of uniform continuity of f on $[-A, A]^2$', an upper bound M (decimal) on the norm of f (on $[-A, A]^2$), and strictly positive decimal `error`. It returns functions $G0, G1, .., G4$ from decimals to decimals such that $|f - \sum_i Gio\phi_i| <$ `error`.

```
def cptly_supp_k(A,f,delta,M,error):
    r, k, d, F=0, [], [], [f]
    G0=[identically_zero_fn]
    G1=[identically_zero_fn]
    G2=[identically_zero_fn]
    G3=[identically_zero_fn]
    G4=[identically_zero_fn]
    while (M>= error):
        k.append(int(ceil(log(float(1/delta(M/18)),10))))
        g0,g1,g2,g3,g4 = one_iteration_step(A,F[r],k[r])
        G0.append(add_fn(G0[r],piecewise_fn_from_list(g0)))
        G1.append(add_fn(G1[r],piecewise_fn_from_list(g1)))
        G2.append(add_fn(G2[r],piecewise_fn_from_list(g2)))
        G3.append(add_fn(G3[r],piecewise_fn_from_list(g3)))
        G4.append(add_fn(G4[r],piecewise_fn_from_list(g4)))
        #new delta fn XXXXX
        delta = delta
        #new d
```

```
        d.append(sum(Decimal('1.0')/(10**s) for s in range(2,k[r
            ]+1)))
        #new f
        F.append(new_f(F[r],G0[r],G1[r],G2[r],G3[r],G4[r],d[r]))
        #new upper bound, M
        M=(5*M)/6
        # increase r, go round again
        r=r+1
    return (G0[r],G1[r],G2[r],G3[r],G4[r])
```

Towards implementing Algorithm 50 define the functions α_n as `alpha(n)`.

```
def alpha(n):
    if (n==1):
        def alpha_n(x1,x2):
            if ((abs(x1)>2) or (abs(x2)>2)):     return 0
            elif ((abs(x1)<1) and (abs(x2)<1)):      return 1
            else:                return 2-max(abs(x1),abs(x2))
    else:
        def alpha_n(x1,x2):
            if (abs(x1)>n+1) or (abs(x2)>n+2) or (abs(x1)<n-1) or
                (abs(x2)<n-1):
                return 0
            elif ((abs(x1)>n) or (abs(x2)>n)):
                return n+1-max(abs(x1),abs(x2))
            else:    return max(abs(x1),abs(x2))-n+1
    return alpha_n
```

Finally implement Algorithm 50. The function `gen_k(f,delta,M,N,error)` takes as inputs: a function f of pairs of decimals to decimals, integer N, uniform delta of continuity on

$[-(N+1), N+1]^2$ called `delta`, an upper bound M on the same square, and error bound, `error`. The outputs are the functions g_0, g_1, g_2, g_3, g_4 in the Kolmogorov approximation of f on $[-N, N]^2$ to within `error` given by Algorithm 50.

```
def gen_k(f, delta ,M,N, error):
    g0, g1=identically_zero_fn , identically_zero_fn
    g2,g3,g4=identically_zero_fn , identically_zero_fn ,
       identically_zero_fn
    for n in range(1,N+2):
        f_n=multiply_fn(f,alpha(n))
        g0_n,g1_n,g2_n,g3_n,g4_n=cptly_supp_k(n+1,f_n,delta , M,
           error/(N+1))
        g0=add_fn(g0,g0_n)
        g1=add_fn(g1,g1_n)
        g2=add_fn(g2,g2_n)
        g3=add_fn(g3,g3_n)
        g4=add_fn(g4,g4_n)
    return (g0,g1,g2,g3,g4)
```

APPENDIX C

FUNCTION SPACE AND GENERALIZED METRIC PROPERTIES

C.1 INTRODUCTION

In [27] Gartside & Reznichenko showed that the space $C_k(X)$ of continuous real valued functions on a Polish (i.e., separable, completely metrizable) space X is stratifiable (definition below). Interestingly it remains unknown if these function spaces are necessarily M_1 (have a σ-closure preserving base), and $C_k(\text{irrationals})$ is a prime candidate for a counter-example to the $M_3 \Rightarrow M_1$ question whether every M_3-space is an M_1-space or not.

Here in this note we expand the class of function spaces known to be stratifiable by showing: if X is a compact-covering image of a closed subspace of product of a σ-compact Polish space and a compact space, then $C_k(X, M)$, the space of continuous maps of X into M with the compact-open topology, is stratifiable for any metric space M.

Our proof of stratifiability is necessarily completely different from the argument of [27] where essential use was made of the separability of $C_k(X)$ when X is Polish. There are two kinds of differences. First, instead of making $\sigma-$cushioned pair base, we demonstrate the existence of g-functions as in the definition of stratifiability: a space Z is stratifiable if for every point z of Z there is a decreasing sequence $g(n, z)$ of open sets with intersection $\{z\}$ such that if z is in an open set U, then there exists an open W and integer N such that $z \in W \subseteq U$ and if $y \notin U$ then $g(N, y) \cap W = \emptyset$.

Second, we apply the argument due to Gruenhage & Tamano [36] who showed, if X is a σ-compact Polish space then there are two collections, $\mathcal{K}$ and $\mathcal{P}$, of compact sets with the following properties:

1) $\mathcal{K}$ is dominating (in the family of all compact subsets of X), closure-preserving and:

($*$) whenever $x_n \in K_n \in \mathcal{K}$, and $x_n \notin \bigcup_{j \neq n} K_j$, then the set $\{x_n\}_{n \in \omega}$ has a limit point;

2) $\mathcal{P} = \{P_n : n \in \omega\}$ is an increasing collection whose union is X and:

($**$) for any $n \in \omega$ and $K \in \mathcal{K}$, $P_n \setminus P_{n-1} \subset K$ or $(P_n \setminus P_{n-1}) \bigcap K = \emptyset$.

Their proof then proceeds by induction on C-scattered rank. If X is σ-compact Polish then define $X^{(0)} = X$, and inductively $X^{(\alpha+1)} = X^{(\alpha)} \setminus$ (all points of $X^{(\alpha)}$ with a compact neighborhood) and $X^{(\lambda)} = \bigcap_{\beta<\lambda} X^{(\beta)}$ for limit λ.

For some minimal $\alpha < \omega_1$, called the C-scattered rank, $X^{(\alpha)} = \emptyset$. Gruenhage & Tamano used these collections to show that if X is σ-compact Polish then $C_k(X)$ is M_1. In Section 3 we will similarly show that: if X is σ-compact Polish, K is compact and M metric then $C_k(X \times K, M)$ is m_1 (every point has a closure preserving local base) and hence is M_1.

In the final Section we give some relevant examples.

Let B be a Banach space with norm $\| \cdot \|$. For any $f \in C_k(X, B)$, compact set K, and $\epsilon > 0$, let $B(f, K, \epsilon) = \{g \in C_k(X, B) : \| g(x) - f(x) \| < \epsilon\}$.

C.2 STRATIFIABILITY

Theorem 56. *Suppose X is a σ-compact Polish space and B is a Banach space with norm $\| \cdot \|$. Then $C_k(X, B)$ is stratifiable.*

Proof: The proof is by induction on the C-scattered rank.

Case 1. X is locally compact. This corresponds to X having C-scattered rank one. Write X as an increasing union of compact sets $L_n, n \in \omega$, where L_n is contained in the interior of L_{n+1}. Then $\{B(f, L_n, 1/(m+1)) : n \in \omega, m \in \omega\}$ is a countable local base of $f \in C_k(X, B)$.

Therefore, $C_k(X, B)$ is metrizable, and hence stratifiable, since it is a first countable topological group.

Case 2. X has a locally-finite cover $\mathcal{G} = \{G_m : m \in \omega\}$ by closed sets such that $C_k(G_m, B)$ is stratifiable for each $m \in \omega$. Note that this case is satisfied if X has C-scattered rank a limit ordinal.

Fix a g-function g_m for each $C_k(G_m, B)$. Then, for any $f \in C_k(X, B)$, we can consider $f|_{G_m}$ the restriction of f on G_m, and the $\{g_m(n, f|_{G_m}) : n \in \omega\}$ satisfy the requirements of a g-function of stratifiable spaces.

Also, we may assume that $g_m(n, f|_{G_m})$ is of the form $B(f|_{G_m}, K, \epsilon) = \{g \in C_k(G_m, B) :\parallel g(x) - f(x) \parallel < \epsilon$ for any $x \in K\}$ for some compact set $K \subset G_m$ and $\epsilon > 0$. Then $g_m(n, f|_{G_m})$ can be considered as an open subset $\hat{g}_m(n, f|_{G_m}) = B(f, K, \epsilon) = \{g \in C_k(X, B) :\parallel g(x) - f(x) \parallel < \epsilon$ for any $x \in K\}$ of $C_k(X, B)$. So we can denote $g_m(n, f|_{G_m})$ as $g_m(n, f)$, and define $g(n, f) = \bigcap_{i \leq n, j \leq n} g_i(j, f)$. Then, by the local finiteness of $\mathcal{G}$ and the definition of $g(n, f)$, it is easy to check $g(n, f)$ a g-function for $C_k(X, B)$.

Case 3. The C-scattered rank of X is a successor ordinal $\alpha + 1$ (where $\alpha \geq 1$). In this case, suppose $A = X^{(\alpha)}$. By case 2), it is sufficient to prove this when A is compact. Then by Borges-Dugundji Extension Theorem [5],[8], $C_k(X, B)$ can be embedded as a subspace of $C_k(A, B) \times C_{k,0}(X/A, B)$ by taking f to $(f|_A, f - e(f|_A))$. Here e is the extension map, and $C_{k,0}(X/A, B)$ is the subspace of $C_k(X/A, B)$ consisting of all maps assigning the point A to the zero element in B. Since it is obvious that $C_k(A, B)$ is metrizable, we just need to show that $C_{k,0}(X/A, B)$ is stratifiable.

In the following, we will first give the definition of the g-function of $C_{k,0}(X/A, B)$, then verify it has the requisite properties.

1. Definition of the g-function.

By above remark, it suffices to show that the space $C_{k,0}(X, B) = \{f \in C_k(X, B) : f(*) = \theta\}$ has a g-function in case that $X^{(\alpha)} = \{*\}$(one point set).

Fix a non increasing local base $(U_k)_{k\in\omega}$ at $*$ $U_0 = X$. Let $V_k = \overline{U_k} \backslash U_{k+1}$. Hence, $r(V_k) < \alpha+1$ for each k. So, by our Inductive Hypothesis, for any $f \in C_k(X, B)$, $f|_{V_k}$ has g-function for each $k \in \omega$ denoted by $G_k(n, f|_{V_k})$. Notice that $G_k(n, f|_{V_k})$ can be considered as an open neighborhood of f in $C_{k,0}(X, B)$. So, here we can denote this open neighborhood by $G_k(n, f)$.

Let $\mathcal{K}$ and $\{P_n : n \in \omega\}$ be the collection of the compact subsets of X satisfying $(*)$ and $(**)$ from the Introduction.

For each $k \in \omega$ and $K \in \mathcal{K}$, let $K^k = K \cap V_k$ and $\mathcal{K}^k = \{K^k : K \in \mathcal{K}\}$, and $P_n^k = P_n \cap V_k$. Then since V_k is closed for each $k \in \omega$, it is obvious that $\mathcal{K}^k$ and $\{P_n^k : n \in \omega\}$ also have the properties $(*)$ and $(**)$ with respect to each V_k.

Let q be any positive rational, and let $q_n = (1 - 1/2^{n+1})q$. For each $L \in \mathcal{K}$, define

$$B_q(L) = \{f \in C_{k,0}(X, B) : \forall n \forall x \in L \cap P_n(\| f(x) \| < q_n)\}$$

Claim: $B_q(L)$ is open in $C_{k,0}(X, B)$.

Proof of Claim: Fix $f \in B_q(L)$. Since L is compact, there exists $x \in L$ such that $\| f(x) \| = \sup\{\| f(y) \| : y \in L\}$. Then $x \in P_n^k$ for some $n \in \omega$. Hence, $\| f(x) \| < q_n$. Let $\epsilon_i = \min\{q_i - \| f(y) \| : y \in L \bigcap P_i^k\}$, if $L \bigcap P_i^k \neq \emptyset$. Finally let $\epsilon = \min\{\epsilon_i : 1 \leq i \leq n, L \cap P_i \neq \emptyset\}$.

Then we can check $B(f, K, \epsilon) \subseteq B_q(K^k)$.

Fix $a \in \omega$. Since f is continuous and $f(*) = \theta$, we can get $M_f^a \in \omega$, such that $\| f(x) \| < (1 - 1/2)10^{-(a+1)}$ for any $x \in V_m$ with $m \geq M_f^a$.

In the following, set $q = 10^{-(a+1)}$ and $q_\ell = (1 - 1/2^{\ell+1})10^{-(a+1)}$, and let $\mathcal{K}_f^k = \{K^k \in \mathcal{K}^k, f \notin \overline{B_q(K^k)}\}$.

Call $x \in V_k$ a *bad point* of f if there exits $\ell \in \omega$ such that $x \in P_\ell \cap V_k$ but $\| f(x) \| > q_\ell$. (This terminology, and the following proof is similar to the argument in [36].) It is easy to see that f has a bad point in every $K^k \in \mathcal{K}_f^k$. Also, we can see $\mathcal{K}_f^k = \emptyset$ if $k \geq M_f^a$.

Fix $k \in \omega$ with $\mathcal{K}_f^k \neq \emptyset$.

Let ℓ_0 be the least such that there is a bad point $x_0 \in P_{\ell_0}^k$ of f which is in some $K_0^k \in \mathcal{K}_f^k$. Then there exists ϵ_0^k such that $B(f, \{x_0\}, \epsilon_0^k) \cap B_q(K^k) = \emptyset$, for any K^k with $x_0 \in K^k \in \mathcal{K}_f^k$.

Then take $\mathcal{K}_{1,f}^k = \{K^k \in \mathcal{K}_f^k : x_0 \notin K^k\}$. If $\mathcal{K}_{1,f}^k \neq \emptyset$, we can get $x_1, \ell_1, \epsilon_1^k$, and K_1^k.

Here, ℓ_1 is the least number such that there is a bad point $x_0 \in P_{\ell_0}^k$ of f in some $K_1^k \in \mathcal{K}_{1,f}^k$ and $B(f, \{x_1\}, \epsilon_1^k) \cap B_q(K^k) = \emptyset$, for any K^k with $x_1 \in K^k \in \mathcal{K}_{1,f}^k$.

Then we can take $\mathcal{K}_{2,f}^k = \{K^k \in \mathcal{K}_f^k : x_1 \notin K^k\}$.

Inductively we get $x_i \in K_i^k \in \mathcal{K}_{i,f}^k$, where x_i is in $P_{\ell_i} \setminus P_{\ell_{i-1}}$ and is a bad point of f, $\ell_0 < \ell_1 <$ ..., and K_i^k contains no bad points of f in $P_{\ell_i - 1}$.

In particular, this implies $x_i \notin K_j^k$ if $i < j$. We show, by contradiction that this process must terminate after a finite number of steps.

If not, suppose that we get an infinite sequence $\{x_i : i \in \omega\}$. We claim the x_i's form a closed discrete set. For suppose they have a limit point y, say $y \in P_L$. Then y is a bad point of f (note $f(y) \geq q$). For sufficiently large j, $\ell_j > L$, it follows that y is not in K_j^k. Then by closure-preserving, the set $\bigcup\{K_j^k : \ell_j > L\}$ is closed, contains all but finitely many x_i's and misses y – a contradiction. Since $\{x_j : j \in \omega\}$ is discrete, we can pass to an infinite subset A of ω such that, for $i \neq j \in A$, we have x_i not in K_j. Then by the convergence property $(*)$ of $\mathcal{K}^k$, $\{x_i\}$ must have a limit point – contradiction.

Therefore, we can suppose the above stops in $\ell_f^{k,a}$ steps. Take $\epsilon_f^k = \min\{\epsilon_0^k, ..., \epsilon_{\ell_f^{k,a}}^k\}$ and $F_f^{k,a} = \{x_0, ..., x_{\ell_f^{k,a}}\}$. Now $B(f, F_f^{k,a}, \epsilon_f^{k,a}) \cap \overline{B_{10^{-(a+1)}}(K^k)} = \emptyset$, for any $K^k \in \mathcal{K}_f^k$.

Finally we can give the definition of the g-function at f.

$$g(n, f) = (\bigcap_{i=0}^{n} G_i(n, f)) \cap (\bigcap_{a=1}^{n} \bigcap_{k=0}^{M_f^a} B(f, F_f^{k,a}, \epsilon_f^{k,a})).$$

2. Verification of the g-Function.

Take $\psi \in C_0(X, B)$, $K \in \mathcal{K}$, $n \in \omega$ and let $U = B(\psi, K, 10^{-n})$. Since $\psi(*) = \theta$ and ψ is continuous, there exists M_ψ such that $\| \psi(x) \| < 10^{-(n+1)}$ for any $x \in \overline{U_{M_\psi}}$. So, we can see $\psi \in B_{10^{-(n+1)}}(K \cap \overline{U_{M_\psi}})$.

For each $V_i, i \leq M_\psi$, we have n_i and W_i which contains $\psi|_{V_i}$ satisfying that $G_i(n_i, h) \cap W_i = \emptyset$ for any $h \in C_{k,0}(V_i, B) \setminus B(\psi|_{V_i}, K \cap V_i, 10^{-n})$.

Define $N = \max\{n_1, ..., n_{M_\psi}, n\}$ and $W = W_1 \cap ... \cap W_{M_\psi} \cap B_{10^{-(n+1)}}(K \cap \overline{U_{M_\psi}})$.

It remains to check the $g(n, f)$'s, N and W satisfy the conditions in the definition of stratifiability.

Take $f \notin U$, which means there exists $x \in K$ such that

(1): $\| f(x) - \psi(x) \| > 10^{-n}$. Two cases arise.

Case 1, $x \in V_i$ *and* $1 \leq i \leq M_\psi$. Then easily, we get $g(N, f) \cap W = \emptyset$.

Case 2, $x \in V_i$ *and* $i > M_\psi$. Then since $\| \psi(x) \| < 10^{-(n+1)}$, from inequality (1), we get $\| f(x) \| > 9 \cdot 10^{-(n+1)}$. Hence $f \notin \overline{B_{10^{-(n+1)}}(K^i)}$, so $K^i \in \mathcal{K}_f^i$ Then we know $B(f, F_n^i, \backslash\epsilon_f^{i,n}) \cap B_{10^{-(n+1)}}(K^i) = \emptyset$. Now $g(N, f)$ is a subset of the first term and W is a subset of the second one, and hence $g(N, f) \cap W = \emptyset$. □

More generally, we have the following theorem.

Theorem 57. *Suppose Y is a σ-compact Polish space, K is a compact space, and M is a metric space. If X is a compact-covering image of a closed subspace of $Y \times K$, then $C_k(X, M)$ is stratifiable.*

This follows directly from the theorem above and the following observations: stratifiability is hereditary, and for X, Y, K and M as in the theorem $C_k(X, M)$ embeds in $C_k(Y, C_k(K) \times B)$ for any Banach space B containing M.

C.3 M_1 PROPERTY

Theorem 58. *Suppose X is a σ-compact Polish space and B is a Banach space with norm $\| \cdot \|$. Then $C_k(X, B)$ is an m_1-space, and hence M_1.*

Hence, if K is a compact space, then $C_k(X \times K)$ is m_1 and M_1.

Proof: First recall that a stratifiable m_1 space is M_1[5]. So it is sufficient to show $C_k(X, B)$ is m_1. Further, since $C_k(X, B)$ is a topological group, we only need to construct a closure preserving base for the zero function **0**.

Let $q > 0$, and let $q_n = (1/2^{n+1})q$. As in [36], for each $K \in \mathcal{K}$, define $B_q(K) = \{f \in C(X,B) : \forall n \forall x \in K \cap P_n(\| f(x) \| < q_n)\}$. Then the same proof as in [36] shows that $\{B_q(K) : K \in \mathcal{K}\}$ is closure-preserving (Note that the difference is only between the absolute value and the norm). Take an increasing cover $\{K_n\}_{n\in\omega}$ of X consisting of elements of $\mathcal{K}$. Then $\{B(\mathbf{0}, K_n, 1/2^n) : n \in \omega\}$ is an open family of $C_k(X,B)$ which is locally finite outside $\{\mathbf{0}\}$. Now define $\mathcal{B}_n = \{B_{1/2^n}(K) : K_n \subseteq K\}$ and $\mathcal{B} = \{\mathcal{B}_n\}$. Then $\mathcal{B}$ is a closure-preserving open neighborhood base of $\mathbf{0}$. $\square$

C.4 EXAMPLES

Observe that if we take any σ-compact Polish space, Y, which is not locally compact, for example an open disc in the plane along with one boundary point, or the metric fan (see below), and any non-metrizable compactum, K, say $[0,1]^{\omega_1}$, then $C_k(Y \times K)$ is non-separable, stratifiable but not metrizable.

Now we give an example of a non-metrizable space X which is the compact-covering image of a σ-compact Polish space. Then $C_k(X)$ is (separable) stratifiable but not metrizable.

Let $X = \mathbb{F}$ be the metric fan and σ be the metric fan topology. So $\mathbb{F}$ has underlying set $(\omega \times \omega) \cup \{*\}$, points in $\omega \times \omega$ are isolated, and a basic neighborhood of $*$ has the form $\{*\} \cup ((N, \infty) \times \omega)$ for some $N \in \omega$. This is indeed σ-compact Polish, but not locally compact.

Fix $\mathcal{P}$ a non-principal ultrafilter on ω. Define a new topology τ as follows: points of $\omega \times \omega$ are isolated, and basic neighborhoods of $*$ are of the form $\{*\} \cup ((N,\infty) \times \omega) \cup (\bigcup_{n \le N} \{n\} \times F)$ where $F \in \mathcal{P}$ and $N \in \omega$.

Claim: The compact subsets of $(\mathbb{F}, \tau)$ coincides with the compact subsets of $(\mathbb{F}, \sigma)$.

Proof of Claim: First observe that $\overline{\{n\} \times \omega}^{\tau} = \{n\} \times \omega \cup \{*\}$.

Take any compact subset $K \subseteq (\mathbb{F}, \tau)$. Then for each $n \in \omega$, $K \cap (\{n\} \times \omega) \subseteq K \cap (\{n\} \times \omega \cup \{*\})$ which is finite. Therefore, K is compact in $(\mathbb{F}, \sigma)$.

Since $\tau \subseteq \sigma$, it is clear that sets compact in $(\mathbb{F}, \sigma)$ are τ-compact.

Therefore $(\mathbb{F}, \tau)$ is a (continuous) compact-covering image of $(\mathbb{F}, \sigma)$ by the identity mapping. Since $*$ has no countable local base in $(\mathbb{F}, \tau)$, $(\mathbb{F}, \tau)$ is not metrizable.

Bibliography

[1] P. Alexandrov. On the dimension of normal spaces. *Proc. Roy. Soc. London*, 189:11–39, 1947.

[2] V.I. Arnold. On functions of three variables. *Dokl. Aka. Nauk. SSSR*, (114):679–681, 1957.

[3] V.I. Arnold. Some questions of approximation and representation of functions (Russian). *Proc. International Congress Math.*, pages 339–348, 1958.

[4] V.I. Arnold. On the Representation of continuous functions of three variables by superpositions of continuous functions of two variables. *Math. Sb. (N. S.)*, 48(90):3–74, 1959.

[5] C. R. Borges. On stratifiable spaces. *Pacific J. Math.*, 17:1–16, 1996.

[6] R. Doss. A superposition theorem for unbounded continuous functions. *Trans. Amer. Math. Soc.*, 233:197–203, 1977.

[7] C.H. Dowker. Local dimension of normal spaces. *Quart. J. Math.*, 6:101–120, 1955.

[8] J. Dungundji. An extension of Tietze's theorem. *Pacific J. Math.*, 1:353–367, 1951.

[9] B. L. Fridman. An improvement in the smoothness of the functions in A. N. Kolmogorov's theorem on superpositions. *Dokl. Akad. Nauk SSSR 177 (1967), 1019–1022; English transl., Soviet Math. Dokl. 8 (1967), 1550–1553*.

[10] J. Braun & M. Griebel. On a constructive proof of Kolmogorov's superposition theorem. *Constructive Approximation*, 2009.

[11] Y. Hattori. Dimension and superposition of bounded continuous functions on locally compact, separable metric spaces. *Topology Appl.*, 54(1-3):123–132, 1993.

[12] R. Hecht-Nielsen. Counter propagation networks. *Proceedings of the International Conference on Neural Networks II*, pages 19–32, 1987.

[13] R. Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem. *In Proceedings IEEE International Conference On Neural Networks III, New York, IEEE Press*, pages 11–14, 1987.

[14] R. Hecht-Nielsen. *Neurocomputing*. Addison-Wesley, Reading, 1990.

[15] A. G. Vitushkin & G. M. Henkin. Linear superpositions of functions . *UMN*, 22:77–124, 1967.

[16] D. Hilbert. *Mathematische Probleme*. Nachr. Akad. Wies. Gottingen (1900), 253 297, Gesammelte Abhandlungen, Bd. 3, Springer, Berlin, 1935.

[17] J-P. Kahane. Hilbert's 13th problem: an intersection of algebra, analysis and geometry. *Proceedings of the Seminar on the History of Mathematics*, 3:1–25, 1982.

[18] A. Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. (Russian). *Dokl. Akad. Nauk SSSR*, 114:953–956, 1957.

[19] V. Kurkova. Kolmogorov's theorem and multilayer neural networks. *Neural Networks*, 5:501–506, 1992.

[20] M. Levin. Dimension and superposition of continuous functions. *Israel Journal of Math.*, 70(2):205–218, 1990.

[21] G. G. Lorentz. *Approximation of Functions*. Athena Series, Selected Topics in Mathematics. Holt, Rinehart and Winston, Inc., New York, 1966.

[22] Y. Mibu. On Baire functions on infinite product spaces. *Proc. Imp. Acad.*, 20(9):661–663, 1944.

[23] P. Ostrand. Dimension of metric spaces and Hilbert's problem 13. *Bull. Amer. Math. Soc.*, 71:619–622, 1965.

[24] A. Leidermana & M. Levin & V.G. Pestov. On linear continuous open surjections of the spaces $C_P(X)$. *Top. and its Appl.*, 81(3):269–279, 1997.

[25] V.G. Pestov. The coincidence of the dimension dim of l-equivalent topological spaces. *Soviet Math. Dokl.*, (26):380–383, 1982.

[26] B.J. Pettis. On continuity and openness of homomorphisms in topological groups. *Ann. of Math. (2)*, 52:293–308, 1950.

[27] P.M. Gartside & E. Reznichenko. Near metric properties of function spaces. *Fund. Math.*, 164:97–114, 2000.

[28] S. Shelah. Cardinal arithmetic. *Oxford University Press*, 1994.

[29] A. Skopenkov. A description of continua basically embeddable in $\mathbb{R}^2$. *Top. and its Appl.*, 65(1):29–48, 1995.

[30] L. Blum & F. Cucker & M. Shub & S. Smale. Complexity and real computation: A manifesto. *International Journal of Bifurcation and Chaos*, 6:3–26, 1995.

[31] D. A. Sprecher. On the structure of continuous functions of several variables. *Trans. Amer. Math. Soc.*, 115(3):340–355, 1965.

[32] D. A. Sprecher. A numerical implementation of Kolmogorov's Superpositions. *Neural Networks*, 9(5):765–772, 1996.

[33] D. A. Sprecher. A numerical implementation of Kolmogorov's Superpositions II. *Neural Networks*, 10(3):447–457, 1997.

[34] Y. Sternfeld. Hilbert's 13th problem and dimension. *Geometric aspects of functional analysis (1987–88),Lecture Notes in Math., 1376,Springer, Berlin*, pages 1–49, 1989.

[35] A. Stone. Cardinals of closed sets. *Mathematatika*, 6:99–107, 1959.

[36] G. Gruenhage & K. Tamano. If X is a $\sigma-$compact Polish, then $C_k(X)$ has a σ-closure-preserving base. *Top. and its Appl.*, 151:99–106, 2005.

[37] A. G. Vitushkin. On Hilbert's thirteenth problem (Russian). *Dokl. Akad. Nauk SSSR*, 96:701–704, 1954.

[38] A. G. Vitushkin. On Hilbert's thirteenth problem and related questions. *Russian Math. Surveys*, 59:11–25, 2004.

[39] K. Weihrauch. *Computable analysis: an introduction*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2000.