Kolmogorov Superposition Theorem and Its Applications

A thesis presented for the degree of

Doctor of Philosophy of Imperial College of London

and the

Diploma of Imperial College of London

by

Xing Liu

Department of Mathematics

Imperial College London, UK

Supervisor: Professor Boguslaw Zegarlinski

September 2015

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Copyright

To my mom Fengqing Wang, for her love throughout my life.

Abstract

Hilbert's 13th problem asked whether every continuous multivariate function can be written as superposition of continuous functions of 2 variables. Kolmogorov and Arnold show that every continuous multivariate function can be represented as superposition of continuous univariate functions and addition in a universal form and thus solved the problem positively. In Kolmogorov's representation, only one univariate function (the outer function) depends on and all the other univariate functions (inner functions) are independent of the multivariate function to be represented. This greatly inspired research on representation and superposition of functions using Kolmogorov's superposition theorem (KST).

However, the numeric applications and theoretic development of KST is considerably limited due to the lack of smoothness of the univariate functions in the representation. Therefore, we investigate the properties of the outer and inner functions in detail. We show that the outer function for a given multivariate function is not unique, does not preserve the positivity of the multivariate function and has a largely degraded modulus of continuity. The structure of the set of inner functions only depends on the number of variables of the multivariate function. We show that inner functions constructed in Kolmogorov's representation for continuous functions of a fixed number of variables can be reused by extension or projection to represent continuous functions of a different number of variables.

After an investigation of the functions in KST, we combine KST with Fourier transform and write a formula regarding the change of the outer functions under different inner functions for a given multivariate function. KST is also applied to estimate the optimal cost between measures in high dimension by the optimal cost between measures in low dimension. Furthermore, we apply KST to image encryption and show that the maximal error can be obtained in the encryption schemes we suggested.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor, Professor Boguslaw Zegarlinski, for his patience, continuous support, and encouragement throughout the duration of my Ph.D. study and writing of this thesis. Thanks to Boguslaw for his ideas and advice on many interesting research topics. He also taught me skills in managing life and work. His guidance and support have been invaluable over the four years of my Ph.D. study.

Thanks to my examiners Professor Michael Ruzhansky and Professor Zhongmin Qian for reading and assessing my thesis. They provided very helpful advice and suggestions.

Thanks to my friends for their unfailing mental support in my work. In particular, I would like to thank Weiwei Wang for helping me draw the figures in the thesis.

I would also like to thank China Scholarship Council for the financial support.

Last but not the least, thanks to my family for their accompany, understanding and support.

Table of contents

Chapter 1

An overview of Kolmogorov superposition theorem

1.1 Hilbert's 13th problem and its positive and negative answers

In 1900, Hilbert [21] posted a list of 23 problems which he considered important to the development of mathematics in the 20th century. The 13th problem in Hilbert's problem list is as follows: " The equation of the seventh degree $f^7 + xf^4 + yf^2 + zf + 1 = 0$ is not solvable with the help of any continuous functions of only two arguments." As is known, the solution of an algebraic equation $x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$ of degree $n \leq 4$ is given by a formula involving only algebraic operations, i.e., addition, subtraction, multiplication, division and radicals (fractional power). For $n > 4$, it can be reduced to the form $y^n + b_{n-4}y^{n-4} + \cdots + b_1 y + 1 = 0$ by means of Tschirnhaus transformations [63], which uses only algebraic operations. It follows that the solution of an algebraic equation of degree $n \leq 6$ can be represented by superposition of continuous functions of 2 variable and for $n \geq 7$, the solution can be represented by superposition of continuous functions of $n - 4$ variables. In particular, the solution of an algebraic equation of degree 7 is a superposition of continuous functions of 3 variables and a further reduction seems impossible.

Hilbert's problem list has attracted enormous researchers and stimulated various solutions. Some of the problems on the list have not been solved completely so far. Studies on

Hilbert's 13th problem bloomed around 1960s. Using a technique on functional trees by Kronrod [31], Kolmogorov [28] claimed that every continuous function can be represented as a superposition of continuous functions of 3 variables. Then Arnold [2], Kolmogorov's student, proved that any continuous function of 3 variables can be represented as a superposition of continuous functions of 2 variable in 1957. Kolmogorov and Arnold's results together show that any continuous multivariate function can be represented as a superposition of continuous functions of 2 variables and thus Hilbert's conjecture was incorrect at least for continuous functions. Shortly after this, Kolmogorov [29] found a new construction, avoiding functional trees, and improved the proof significantly, which formed the Kolmogorov superposition theorem (KST): every continuous function f of n variable can be represented as a superposition of continuous functions of one variable and the additive operation:

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g_q \left(\sum_{p=1}^{n} \psi_{pq}(x_p) \right), \tag{1.1.1}$$

where g_q and ψ_{pq} are continuous univariate functions on $\mathbb{R}$ and ψ_{pq}'s are independent of f.

On the other direction, the irrepresentability of multivariate functions by superposition of univariate functions and additive operation was studied at the same time. Hilbert's 13th problem were understood in different ways and it was not sure if it was continuous function class which Hilbert was concerned about. For example, the problem can be understood as to solving algebraic equation of degree 7 by superposition of smooth or analytic functions of two variables. In this sense, Hilbert's conjecture could be right, as there are analytic functions of 3 variables which cannot be represented by finite superposition of analytic functions of 2 variables [66]. In fact, the number of partial derivatives up to order p for a function of 3 variables is proportional to p^3, whereas the number of partial derivatives up to order p of functions of two variables is proportional to p^2. Hence a function of 3 variable f that can be represented by superposition of functions of two variables has to satisfy some algebraic partial differential equation. That is,

$$P(f, \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, ..., \frac{\partial^{p_1+p_2+p_3} f}{\partial x_1^{p_1} \partial x_2^{p_2} \partial x_3^{p_3}}) = 0,$$

where P is a polynomial with constant coefficients in the function f and its partial derivatives up to order p. Therefore there are "more" functions of 3 variables than superposition of functions of 2 variables [66].

Generally, it is showed by Vitushkin [69] that if the complexity of a function is measured by the ratio of the number of variables to the order of smoothness, then almost every function of a given complexity, except for a first-category set of functions, cannot be represented as a superposition of functions of lower complexities. This reveals the inevitable decrease in smoothness of functions as the number of variables of functions decreases in superposition.

Vitushkin's proof [65] used the concept of variations of sets designed by himself. A simpler proof, using ϵ-capacity, is given by Kolmogorov and Tihomirov [60]. For more specific examples of irrepresentability by superpositions of functions from certain classes see [44] [67] [69] and references therein.

1.2 The motivation and structure of the thesis

In the thesis, we focus on the positive answer to Hilbert's 13th problem: Kolmogorov superposition theorem. We study topics around Kolmogorov superposition theorem in both theoretical and applied aspects.

Techniques and methods generalising concepts and results from lower dimensions to higher dimensions have been explored both in theoretical and practical research. In general, mathematical problems in high dimensions are generally more difficult than those in low dimensions. For example, the optimal transport maps between measures in high dimensional spaces are generally not easily obtained, while they are explicitly solved for measures on the real line. Another example is typical partial differential equations (PDEs) describing physical processes involving time and space. When the dimension of spaces involved increases, more complicated techniques and methods are needed to tackle the solutions to the PDEs. In view of Kolmogorov's result, superpositions of functions of one variable and the additive operation exhaust the set of all multivariate functions. One may say that there are essentially no continuous functions of multiple variables except the additive function, $f(x, y) = x + y$, and thus no high-dimensional problems at all. This intuitive prospect looks over optimistic, as we will see later in the thesis. Nevertheless, the evident advantage of KST in representation still motivates us to explore its possible applications to problems in high dimensions.

Next we introduce the structure of the thesis with more detailed motivation in each chapter. In the remaining part of Chapter 1, researches closely related to Kolmogorov superposition theorem from the publication of KST in 1957 till now are briefly reviewed. Specifically, we examine the improvements and generalisations of KST, its topological implications, its numerical implementations, approximative versions of KST (particularly neural network), and applications of Kolmogorov's theorem in image processing and other fields.

In chapter 2, we introduce the background knowledge needed in chapters that follow, such as the main types of proofs of KST, especially the constructive proofs. We also introduce general concepts and theorems from optimal transport theory which will be used in Chapter 6, such as cost functions, the optimal cost between two probability measures and the dual problem of optimal transport problems. The Wasserstein distance induced from optimal transport problem is also introduced, which is used to estimate the difference between multivariate functions obtained from superposition of different inner functions with a shared outer function.

It is notable that the inner functions ψ_{pq} in KST are independent of the functions to be represented and thus KST allows a universal representation for all multivariate continuous functions. In other words, KST separates the topological structure of the domain I^n from the value information of the multivariate function in its representation. This requires strong point-separability properties [59] of the inner functions, which impairs their smoothness significantly. As mentioned in section 1.1, Vitushkin's negative results [66] also leads to this conclusion. Moreover, the special structure of the inner functions also has "bad" effects on the smoothness of the outer function. Since the inner functions are independent of any multivariate function f, information of f is totally stored in its outer function g. Naturally, one expects smoother g for smoother f; however, the domain I^n is mapped by the inner functions to the real line in a highly "non-linear" way such that the neighbourhoods of I^n are no longer preserved in the real line (see figure 7.1). This implies potentially that the analytic properties of f are not well preserved in g in Kolmogorov's representation.

In Chapter 3, we investigate the set of inner functions in Kolmogorov's representation formula. Originally, the inner functions are constructed for continuous functions defined on I^n with a fixed dimension $n \geq 2$. We show that if the inner functions are constructed to separate a sequence of $2n+1$ families of little cubes which covers I^n at least $n+1$ times, then it is possible to extend the inner functions in dimension n to dimension $m > n$ or project the inner functions in dimension m to dimension n. In this way, the inner functions can be reused in different dimensions.

In Chapter 4, we investigate the problem how the analytic properties of f are preserved in the univariate function g. For a given family of inner functions in (1.1.1), we show that the outer function g for a given f is not unique and g does not preserve the positivity for $f \geq 0$. The modulus of continuity of f is drastically lost in g. In particular, we gave a sharp lower bound for the modulus of continuity of g with respect to that of f in Sprecher's constructive proof of KST [7].

In KST, the outer function of a continuous function of several variables depends not only on the multivariate function, but also on the inner functions ψ_{pq} chosen in the representation. In Chapter 5, we try to understand the change of outer functions under different inner functions for a given multivariate function using the techniques from Fourier transform. As shown in (1.1.1), the inner functions are served as the argument of the outer functions g. Fourier transform can separate the argument y of a function $g(y)$ from g by moving y to the exponent of Fourier basis e^{iyt}, that is, formally,

$$g\left(\sum_{p=1}^{n} \psi_{pq}(x_p)\right) = \int_{\mathbb{R}} \hat{g}(t) e^{2\pi i (\sum_{p=1}^{n} \psi_{pq}(x_p))t}.$$

Despite the undesirable performance of KST in preserving analytic properties, KST indeed provides a way to reduce the dimensionality of functions. Probability measures with continuous density functions in n dimension can be transformed to probability measures in one dimension using KST. Thus the transport problems in n dimension can be studied in its equivalent problem formulated in one dimension. In chapter 6, we estimate the optimal transport cost in n dimension with certain cost functions by the corresponding optimal cost in one dimension. Similarly, the Wasserstein distance between two probability measures in n dimension can also be bounded by the Wasserstein distance of their equivalent counterparts in one dimension.

The main disadvantage of KST, its high non-linearity, can be turned into an advantage from a different point of view. In fact, the highly "non-linear" dependence of g on f inspires us to apply KST in encryption. Suppose we have some data to be transmitted and the data is considered as values of a multivariate function f on a discrete set. First, choose a family of inner functions and encode f as its outer function g, then transmit g publicly. The chosen inner functions are served as secret keys. The authorised users with secret keys can reconstruct f from the public data g. It is hard to crack f directly from g, since they are connected in a highly non-linear way. Different versions of KST, such as Lorentz's (1.3.1) and Sprecher's (1.3.2) versions, give different types of keys. It is difficult to construct a key even knowing the type of the correct keys because the construction of the keys are complex due to their high non-linearity. Some experiments and simulations of image coding have been conducted very recently by Leni et. al [34] [33] [35], which we will give more details in section 1.6. In Chapter 7, we estimate the error of reconstructing f from g with wrong keys measured in L_p-norm and Wasserstein distance. It is showed that the error can be maximised in some cases. For example, a greyscale picture can be wrongly restored as any random picture between a purely black and purely white picture. Therefore, the encryption based on KST generally guarantees a high-level security.

1.3 Improvements of Kolmogorov superposition theorem

There have been many comments and refinements of Kolmogorov superposition theorem and its proof. Fridman [15] shows that the inner functions ψ_{pq} in (1.1.1) can be chosen to satisfy a Lipschitz condition*. This turns out to be the best possible smooth property for the inner functions ψ_{pq} in the sense that if ψ_{pq} in (1.1.1) is replaced by some continuously differentiable functions, then there are even analytic functions which cannot be represented by formula (1.1.1) [66]. To simplify the form of representation (1.1.1), Lorentz [40] replaced the $(2n+1)n$ inner functions ψ_{pq} by $2n+1$ functions ϕ_q multiplied by n constants

* We say a real or complex-valued function f on n-dimensional Euclidean space satisfies a Lipschitz condition, or Lipschitz continuous, iff there is a real constants $C > 0$ such that $|f(x) - f(y)| \leq C|x - y|$ for all x, y in the domain of f.

λ_p. He also showed that the dependence of outer function g_q on q can be removed by a shift of their arguments. Thus the number of both inner and outer functions are reduced. More precisely, Lorentz [40] in 1966 showed the existence of continuous and monotonously increasing functions ϕ_q such that for any $f \in C(I^n)$ with $I := [0, 1]$, there exists a continuous function g such that

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_p \phi_q(x_p)\right), \tag{1.3.1}$$

where $\lambda_p > 0, p = 1, ..., n$ are rationally independent †.

Sprecher [48] made a further simplification of the inner functions by replacing ϕ_q in (1.3.1) with one single function ψ and proper shifts in its argument. He [48] proved that there exist positive numbers $a, b, \lambda_p, \lambda_{pq}$, $p = 1, ..., n, q = 0, ..., 2n$, and a real, monotonic increasing function $\psi : I \to I$ satisfying a Hölder condition ‡ with exponent $\log 2/\log(2n+2)$, such that for any given $f \in C(I^n)$ and all $(x_1, ..., x_n) \in I^n$,

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_p \psi(x_p + aq) + bq\right), \tag{1.3.2}$$

or

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_{pq} \psi(x_p + aq)\right) \tag{1.3.3}$$

holds with some continuous function g. Note that the values a and the functions g can differ in the two representations. Moreover, Sprecher [50] also showed that the inner functions in (1.3.2) and (1.3.3) can be Lipschitz continuous. In fact, following the insight by Fridman [15] and Kahane [24], any tuple of $(\phi_0, ..., \phi_{2n})$ in (1.3.1) with ϕ_q continuous and increasing can be re-parametrized to satisfy a Lipschitz condition [26].

Hedberg [20] and Kahane [24] gave a proof of KST totally different from Kolmogorov's. Using the category theory, they show that $\lambda_1, ..., \lambda_n$ can be chosen such that the represen-

---

† $\lambda_1, ..., \lambda_n$ are rationally independent, if for any rational numbers $t_1, ..., t_n$ with $\sum_{p=1}^{n} t_p \lambda_p = 0$ it follows that $t_1 = \cdots = t_n = 0$.

‡ A real or complex-valued function f on n-dimensional Euclidean space satisfies a Hölder condition with exponent α, or is Hölder continuous, iff there are real constants $C > 0$ and $0 < \alpha \leq 1$, such that $|f(x) - f(y)| \leq C|x - y|^\alpha$ for all x, y in the domain of f.

tation in (1.3.1) holds for all tuples $(\phi_0, ..., \phi_{2n})$ with ϕ_q continuous and increasing , except for a subset of first category. The first category here is with respect to the metric space Φ^n with the ordinary maximum metric, where $\Phi \subseteq C(I)$ denotes the metric subspace of $C(I)$ of all increasing functions $\phi : I \to I$. Using their idea, Doss [12] showed that the inner sum $\sum_{p=1}^{n} \psi_{pq}(x_p)$ in formula (1.1.1) can be replaced by a product of univariate functions: $\prod_{p=1}^{n} \phi_{pq}(x_p)$.

1.4 Extensions and generalisations of KST

We now introduce some generalisations and extensions of Kolmogorov superposition theorem which provide a deeper understanding of superposition of functions. Ostrand [43] in 1965 showed that KST holds on compact metric spaces.

Theorem 1.4.1 (Ostrand [43])**.** *Let $X_1, ..., X_l$ be compact metric spaces with X_p having dimension d_p. Let $X = X_1 \times \cdots \times X_l$ and $n = \sum_{p=1}^{l} d_p$. Then there are continuous functions $\psi_{pq} : X_p \to I$, for $p = 1, ..., l$ and $q = 0, ..., 2n$, such that each $f \in C(X)$ is representable in the form*

$$f(x_1, ..., x_l) = \sum_{q=1}^{2n+1} g_q \left(\sum_{p=1}^{l} \psi_{pq}(x_p) \right), \tag{1.4.1}$$

where the functions g_q are real and continuous.

Doss [13] and Demko [10] in 1977 generalised KST to $\mathbb{R}^n$ for unbounded and bounded continuous functions, respectively. Feng [14] in 2010 generalised KST to locally compact and finite dimensional separable metric spaces (or equivalently, spaces homeomorphic to a closed subspace of Euclidean space). He gave a full characterisation of spaces that admitting the superposition formula (1.4.1).

Theorem 1.4.2 (Feng [14])**.** *X is a locally compact, finite dimensional[§] separable metric space, if and only if for every $m, n \in \mathbb{N}$, there is an $r \in \mathbb{N}$ and $\psi_{pq} \in C(X, \mathbb{R}^n)$, for*

[§]Dimension here refers to topological dimension, which is also called Lebesgue covering dimension. A topological space X has Lebesgue covering dimension n if and only if n is the smallest natural number such that for every open cover of X, there exists a refinement of the open cover such that any point in X is covered by at most $n + 1$ sets of the refinement. If a space does not have Lebesgue covering dimension n for any

$q = 1, ..., r$ and $p = 1, ..., m$, such that every $f \in C(X^m, \mathbb{R}^n)$ can be written as

$$f(x_1, ..., x_m) = \sum_{q=1}^{r} g_q \left(\sum_{p=1}^{m} \psi_{pq}(x_p) \right), \tag{1.4.2}$$

for some $g_q \in C(\mathbb{R}^n, \mathbb{R}^n)$.

In particular, for $n = 1$, Feng [14] states Kolmogorov's representation formula for $C(\mathbb{R})$:

Theorem 1.4.3 (KST for $C(\mathbb{R})$ [14])**.** *Fix $m \in \mathbb{N}$. There exists $\psi_{pq} \in C(\mathbb{R})$, for $q = 0, ..., 2m$ and $p = 1, ..., m$, such that for any function $f \in C(\mathbb{R})$, there can be found functions $g_0, ..., g_{2m}$ in $C(\mathbb{R})$ such that:*

$$f(\mathbf{x}) = \sum_{q=0}^{2m} g_q \circ \xi_q(\mathbf{x}), \quad \textit{where } \xi_q(x_1, ..., x_m) = \psi_{1q}(x_1) + \cdots + \psi_{mq}(x_m).$$

Further, one can arrange it so that the co-ordinate functions, $g_0, ..., g_{2m}$, are all identical (say to g), and that ψ_{pq} (and hence ξ_q) are Lipschitz (with Lipschitz constant 1).

Representations of functions from other function classes are also investigated in the literature. For example, the representability of bounded functions or even arbitrary functions are studied in the series of paper of Sternfeld [57] [59][58] and Ismailov [23] respectively. It turns out that the representability depends on the structure of the inner functions with respect to the domain, which has geometrical interpretations (see [27] [62]).

1.5 Cardinality of basic families of a space and its dimension

Kolmogorov superposition theorem implies more topological properties of the spaces in question, in particular the dimension of the space. We first introduce the concept of basic family of a space.

Let X be a metric space. A family of continuous functions, $\{\xi_q : X \to \mathbb{R} | q = 1, ..., m\}$, is said to be **basic** for X iff each $f \in C(X)$ can be written as $f = \sum_{q=1}^{m} g_q \circ \xi_q$, for some

$n \in \mathbb{N}$, it is said to be infinite dimensional. An open cover $\{U_i'\}_{i \in S'}$ of X is a refinement of another open cover $\{U_i\}_{i \in S}$ of X iff $U_i' \subset U_j$ for some $j \in S$ depending on i holds for all $i \in S'$.

$g_1, ..., g_m \in C(\mathbb{R})$. Thus $\{\sum_{p=1}^{n} \psi_{pq}(x_p)\}_{q=0}^{2n}$ in (1.1.1) is a basic family for I^n. The cardinality of basic family relies on the topological structure of the space X. Doss [11] shows that the cardinality of basic family for I^n in (1.1.1), $2n+1$, cannot be reduced when $n = 2$ and ψ_{pq} is monotonously increasing. Sternfeld [58] and Levin [36], with a shorter proof, show that for $n \geq 2$, the dimension of a compact metric space X is no greater than n if and only if the cardinality of the basic family for X is less or equal to $2n+1$. Feng [14] generalised this characterisation to locally compact, separable and metrizable spaces.

Using the duality between function spaces and measure spaces, Sternfeld [59] characterised the basic family by its property in separating points in the topological space X. For example, $\{\xi_1, ..., \xi_m\}$ is a basic family for X if and only if it is a uniformly measure separating family; that is, there exists some $0 < \lambda \leq 1$ such that for every measure μ in the dual space $C(X)^*$ of $C(X)$, $||\mu \circ \xi_q^{-1}||_{TV} \geq \lambda ||\mu||_{TV}$ holds for some $1 \leq q \leq m$, where $(C(X)^*, ||\cdot||_{TV})$ is actually the space of Borel measures on X with total variation norm. We will use this observation to show that the outer function g in Kolmogorov superposition theorem is not unique in Chapter 4.

1.6 Numerical implementation of KST and its application to neural network

From a practical viewpoint, those basic families which are superposition of univariate functions are of our interests.

Let $X_1, ..., X_l$ be locally compact and separable metric spaces with X_p having dimension d_p, p=1,...,l. Let $X = X_1 \times \cdots \times X_l$ and $n = \sum_{p=1}^{l} d_p$. We call a basic family $\{\xi_1, ...\xi_m\}$ for X, $m \geq 2n+1$, a **Kolmogorov basis** for X or **K-basis** for X as abbreviation, iff for every $1 \leq q \leq m$, $\xi_q : X \to \mathbb{R}$ can be written as a sum of continuous univariate functions, that is, there exist continuous $\psi_{pq} : X_p \to \mathbb{R}$, for $p = 1, ..., l$ such that

$$\xi_q(x_1, ..., x_l) = \sum_{p=1}^{l} \psi_{pq}(x_p). \tag{1.6.1}$$

In practice, we consider the special case when $X_1 = \cdots = X_n = I$ and $X = I^n$, which is also the most investigated case of Kolmogorov superposition theorem.

As a matter of fact, most of the proofs of KST are not "implementable" in the sense that the construction is an infinite process. Sprecher contributes a series of papers [53] [52][51][54] concerning the computability and numeric implementation of KST for $C(I^n)$, where the functions in the K-basis are simplified as translations of a single function ψ. Sprecher's construction had minor mistake which was noticed and corrected by Köppen [30] in 2003 without proof. Braun [7] in 2009 implemented the details of the constructive proof and verified Köppen's result, whence a totally constructive proof of KST was presented. Despite the constructive version of KST, the application of KST in real computation as an exact representation is impossible, because the outer function is computed by an infinite number of iterations and the construction of functions in K-bases is also a limit process. Approximative versions of KST are investigated in a series of literatures, such as [6] [22] [32] and so on, where smoother and simpler functions are used to approximate the inner functions or the outer functions or both of them.

KST has applications in various fields, such as non-linear control circuit and system theory, statistical pattern recognition, image and multidimensional signal processing, neural network and so on, see [33] [35] [3], etc.

The application of Kolmogorov superposition theorem to neural network has been mostly discussed since Hecht-Nielsen [19] explained KST as a feedforward neural network in 1987. Briefly speaking, a neural network is a structure to perform computations by a network of interconnected neurons. A neuron produces an output from a certain number of inputs through an activation function. The outputs of some neurons can be sent as inputs to some other neurons. Hecht-Nielsen states that any continuous function $f : I^n \to \mathbb{R}$ can be represented as a neural network with one hidden layer with inputs $(x_1, ..., x_n)$ and activation function ψ (the K-basis), and a single output layer with activation function g (the outer function) that produces the output $f(x_1, ..., x_n)$. Although the highly non-smoothness of the inner functions proposed doubts on the realisation of Hencht-Nielsen's network, Kurkova [32] noticed that instead of exact representation, Hecht-Nielsen's network can be realised in an approximate way by increasing the number of inner functions and approximating both the inner functions and outer functions with sequences of smooth functions. There are various approximation schemes [41] [25] [42] for Hecht-Nielsen's network, for example perceptron type network [41] and projection pursuit algorithm [16]. Particularly, Igelnik and Parikh [22] in 2003 proposed an algorithm of the neural network using spline functions to approximate both the inner function ψ and outer function g. They show that any continuously differentiable function f can be approximated with any given accuracy and the approximation order (the number of net parameters needed for a given approximation error) in their algorithm is better than general approximation schemes for Hencht-Nielsen's network.

In signal processing, most of techniques are applied in 1 dimension or 2 dimension and cannot be easily extended to higher dimensions. Therefore, KST can also be used in image processing in higher dimensions: instead of applying operations directly to signals in high dimensions, such as signal compression, approximative versions of KST allow one to apply the operations to an equivalent representation of the signal in one dimension. Using Igelnik's spline network [22] combined with wavelet transform, Leni, Fougerolle and Truchetet conducted simulations on image reconstruction [34], compression [33] and transmission [35] and produced improved results in image processing. Moreover, they [35] also designed a progressive transmission of secured images by encoding the image as a univariate function with changing K-bases in different resolutions.

Chapter 2

Preliminaries

We introduce two main subjects: the general ideas in the proofs of KST as well as basic concepts and problems in optimal transport theory, which we think are necessary to understand the contents of the thesis.

2.1 Notations

First we introduce some notations and definitions used through this thesis.
$\mathbb{N} := \{1, 2, ...\}$, $\mathbb{N}_0 = \{0, 1, 2, ...\}$.
$\mathbb{R}$ is the set of real numbers, $\mathbb{R}_+$ is the set of non-negative real numbers, and $\mathbb{Q}$ is the set of rational numbers.
$I := [0, 1]$ is the unit interval. I^n, $n \in \mathbb{N}$, is the Cartesian product of I.
$C(A)$: the set of all continuous functions defined on a set $A \subset \mathbb{R}^n$, $n \in \mathbb{N}$.
$C^p(A), p \geq 1$: the set of of functions with continuous p-th partial derivative functions on A.
$C_0(\mathbb{R}) := \{f \in C(\mathbb{R}) : \lim_{x\to\infty} f(x) = 0\}$.
$C_c(\mathbb{R}) := \{f \in C(\mathbb{R})$ with compact support$\}$.
$f[A] := \{f(x) : x \in A\}$ is the image of A under the map f, if $f : A \to B$ is a function defined on set $A \subset \mathbb{R}^n$, $n \in \mathbb{N}$ with values in set $B \subset \mathbb{R}^m$, $m \in \mathbb{N}$.
$P(X)$: the set of probability measures on a Polish space X.
$(\mathcal{X}, \|\cdot\|_\infty)$ is a normed function space and $\|\cdot\|_\infty$ is the maximum norm of functions in $\mathcal{X}$.

A permutation σ of $\{a_1, a_2, ..., a_n\}$ is a bijective map on $\{a_1, a_2, ..., a_n\}$.
$\sigma := (a_1, a_2, ..., a_n)$ denotes a cyclic permutation of $\{a_1, a_2, ..., a_n\}$ defined by

$$\sigma : \{a_1, a_2, ..., a_n\} \mapsto \{a_1, a_2, ..., a_n\}$$
$$\sigma(a_i) = a_{i+1}, \quad i = 1, ..., n-1$$
$$\sigma(a_n) = a_1.$$

Next we introduce three definitions: Kolmogorov basis, Kolmogorov map with respect to a Kolmogorov basis and superposition operator with respect to a family of continuous functions.

Definition 2.1.1 (Kolmogorov basis and Kolmogorov map with respect to Kolmogorov basis)**.** *Let $X_1, ..., X_l$ be locally compact and separable metric spaces with X_p having dimension d_p. Let $X = X_1 \times \cdots \times X_l$ and $n = \sum_{p=1}^{l} d_p$. If there exist continuous functions $\psi_{pq} : X_p \to \mathbb{R}$, $p = 1, ..., l$, $q = 1, ..., m$, such that every $f \in C(X)$ can be represented as*

$$f(x_1, ..., x_l) = \sum_{q=1}^{m} g\left(\sum_{p=1}^{l} \psi_{pq}(x_p)\right)$$

with some $g \in C(\mathbb{R})$. Then we call the family $\xi := \{\xi_1, ..., \xi_m\}$ defined by

$$\xi_q(x_1, ..., x_l) := \sum_{p=1}^{l} \psi_{pq}(x_p), \quad q = 1, ..., m \tag{2.1.1}$$

*a **Kolmogorov basis** or **K-basis** for $C(X)$.*

*For a given K-basis $\xi := \{\xi_1, ..., \xi_m\}$ for $C(X)$, if $K : C(X) \to C(\mathbb{R})$ is a well-defined linear operator such that for all $f \in C(X)$, $f = \sum_{q=1}^{m} (Kf) \circ \xi_q$. Then we call K a **Kolmogorov map** or **K-map** with respect to the K-basis ξ.*

For example, in the constructive proofs of KST [29] [40] [7], a unique outer function g is constructed for any given $f \in C(I^n)$ and thus a Kolmogorov map is well defined.

On the other hand, we can define the superposition operator with respect to a given family of functions, particularly with respect to a K-basis.

Definition 2.1.2 (Superposition operator)**.** *Let $m \in \mathbb{N}$ and X be a locally compact, finite dimensional separable metric space and $\xi := \{\xi_1, ..., \xi_m\}$ with continuous functions $\xi_q : X \to \mathbb{R}$, $q = 1, ..., m$, then we call the linear operator $S_\xi : C(\mathbb{R}) \to C(X)$ defined by*

$$S_\xi(g)(x) := \sum_{q=1}^{m} g\left(\xi_q(x)\right).$$

*the **superposition operator with respect to** ξ or the **superposition operator** when there's no confusion.*

2.2 Proofs of KST

There are two main types of proofs of Kolmogorov's superposition theorem: the proofs using category theory and the constructive proofs.

Kahane [24] and Hedberg [20] around 1970 used the Baire category theorem to show the existence of the inner functions ψ_{pq} in the superposition formula (1.3.1). Despite their proofs were not constructive, they showed more than KST, i.e., $\lambda_1, ..., \lambda_n$ in (1.3.1) can be chosen such that all tuples $(\phi_1, ..., \phi_{2n+1})$ with ϕ_q continuous and increasing make the KST formula hold, except for a subset of first category*.

Based on Kahane's idea [24], Sternfeld [58] in 1979 used a general duality argument in functional analysis and proved KST and Ostrand's [43] generalisation of KST in compact metric spaces. Let X be a compact, separable metric space and $\xi := \{\xi_1, ...\xi_m\}$ be a family of continuous functions defined on X. ξ is a Kolmogorov basis on X if the superposition operator S_ξ with respect to ξ is surjective, which is equivalent to that its dual operator

$$\begin{aligned} S_\xi^* \quad : \quad & C(X)^* \to C(\mathbb{R})^* \\ & \mu \to \sum_{q=1}^{m} \mu \circ \xi_q^{-1}. \end{aligned}$$

is isomprphism into, i.e., S_ξ^* is injective and its inverse, mapping range of S^* onto $C(X)^*$,

*The first category here is with respect to the metric space Φ^n, the corresponding metric product space of Φ with the ordinary maximum metric, where $\Phi \subseteq C(I)$ denote the metric subspace of $C(I)$ of all increasing functions $\phi : I \to I$.

is bounded (see Appendix). Note that $C(X)^*$ is equivalent to the space of Borel measures with bounded total variations on X. Then S_ξ^* is isomorphism into iff $\{\xi_1, ..., \xi_m\}$ uniformly separates the Borel measures on X†. Sprecher then showed by Baire category theorem the existence of families of functions $\{\xi_1, ..., \xi_m\}$ that uniformly separate Borel measures on X and thus proved KST.

The other type of proofs, given by Kolmogorov [29] Lorentz [40], Sprecher [48] and others, is constructive. The constructive proofs consist of two parts: the construction of the independent inner functions and the iterative approximation of the outer function. First, divide the unit cube I^n into disjoint sub-cubes and then shift the sub-cubes along some vectors a certain amount of times such that every point in I^n is covered several times by the union of all the shifted families of sub-cubes. The inner functions are constructed such that the images of all disjoint sub-cubes from all families are mapped into disjoint intervals on the real line. Second, approximate the multivariate function f on these squares by defining proper outer function g_r on the disjoint intervals iteratively at each step r. The infinite series of functions g_r converges to the outer function g for f.

We illustrate the general routine of the constructive proofs in more details with Sprecher and Braun's construction.

Theorem 2.2.1 (Sprecher-Köppen-Braun's constructive version of KST [48][30] [7]). *Let $n \geq 2, m \geq 2n$ and $\gamma \geq m+2$ be given integers and let $\mathbf{x} = (x_1, ..., x_n)$. There exists strictly increasing and continuous function $\psi : I \to \mathbb{R}$ such that for any arbitrary continuous function $f : I^n \to \mathbb{R}$*

$$f(\mathbf{x}) = \sum_{q=0}^{m} g_q \circ \xi_q(\mathbf{x}), \quad \textit{with} \quad \xi_q(\mathbf{x}) = \sum_{p=1}^{n} \lambda_p \psi(x_p + qa),$$

and $g_q : \mathbb{R} \to \mathbb{R}$ continuous, where $a := [\gamma(\gamma-1)]^{-1}$, $\lambda_1 = 1, \lambda_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\beta(r)}$ for $p > 1$ and $\beta(r) = (n^r - 1)/(n-1)$.

The detailed construction of ψ in Theorem 2.2.1 will be discussed in Chapter 4. We

---

†$\xi := \{\xi_1, ..., \xi_m\}$ is said to uniformly separate Borel measures on X, iff there exists some $0 < \lambda \leq 1$ such that for every Borel measure μ in the dual space $C(X)^*$ of $C(X)$, $||\mu \circ \xi_q^{-1}||_{TV} \geq \lambda ||\mu||_{TV}$ holds for some $1 \leq q \leq m$, where $\|\cdot\|_{TV}$ is the total variation of a measure.

assume that ψ is constructed for now. Write the numbers between 0 and 1 in digits with base γ. For each integer $k \in \mathbb{N}$, let $\mathcal{D}_k$ be the set of numbers with k digits:

$$\mathcal{D}_k = \mathcal{D}_k(\gamma) = \left\{ d_k \in \mathbb{Q} : d_k = \sum_{r=1}^{k} i_r \gamma^{-r}, i_r \in \{0, ..., \gamma - 1\} \right\}.$$

Let

$$\delta_k := \frac{\gamma - 2}{(\gamma - 1)\gamma^k}.$$

Then for all $d_k \in \mathcal{D}_k$, the pairwise disjoint intervals

$$E_k^0(d_k) := [d_k, d_k + \delta_k]$$

are mapped by ψ into disjoint image intervals (see Corollary 3.6 in [7]). Furthermore, by the definition of λ_p's the pairwise disjoint cubes

$$S_k^0(\mathbf{d}_k) := \prod_{p=1}^{n} E_k^0(d_{k,p})$$

in I^n are mapped by $\xi_0 = \sum_{p=1}^{n} \lambda_p \psi(x_p)$ in Theorem 2.2.1 into disjoint intervals contained in $Y_0 := \xi_0[I]$ (see Lemma 3.7 in [7]).

Now shift the family of intervals of $E_k^0(d_k)$ by $\frac{q}{\gamma-1}\gamma^k$. For $q = 0, ..., m$, let

$$E_k^q(d_k) := [d_k - \frac{q}{\gamma - 1}\gamma^k, d_k - \frac{q}{\gamma - 1}\gamma^k + \delta_k],$$

and

$$S_k^q(\mathbf{d}_k) := \prod_{p=1}^{n} E_k^q(d_{k,p}).$$

Pairwise disjoint cubes $S_k^q(\mathbf{d}_k)$ are mapped by ξ_q into disjoint intervals contained in $Y_q := \xi_q[I^n]$. For each fixed rank $k \in \mathbb{N}$ and every point $x \in I$, there are at least m values of q's such that x is covered by some q-interval $E_k^q(d_k)$ and thus every $\mathbf{x} \in I^n$ is covered by at least $m + 1 - n$ of q-cubes $S_k^q(\mathbf{d}_k)$.

Let $\|\cdot\|_\infty$ denote the usual maximum norm of functions and $f \in C(I^n)$ be given. Let θ and ϵ be fixed real numbers such that $0 < \frac{m-n+1}{m+1}\epsilon + \frac{2n}{m+1} \leq \theta < 1$, which implies

$\epsilon < 1 - \frac{n}{m-n+1}$. We construct iteratively g_r at step r and Kf will be the sum of the infinite series $g := \sum_{r=1}^{\infty} g_r$.

Starting with $f_0 \equiv f$, for $r = 1, 2, \ldots$, iterate the following step: given the function f_{r-1}, determine an integer k_r such that for any two points $\mathbf{x}, \mathbf{x}' \in I^n$ with $|\mathbf{x} - \mathbf{x}'| \leq \gamma^{-k_r}$, it holds that

$$|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \epsilon \|f_{r-1}\|_\infty.$$

On each interval $\xi_q[S^q_{k_r}(\mathbf{d}_{k_r})]$, take g_r to be constant $\frac{1}{m+1} f(\mathbf{d}_{k_r})$. We can extend g_r linearly into the gaps between $\xi_q[S^q_{k_r}(\mathbf{d}_{k_r})]$ and obtain in this way a continuous function on Y_q for all q such that $\|g_r\|_\infty \leq \frac{1}{m+1}\|f_{r-1}\|_\infty$. Let $f_r(\mathbf{x}) := f(\mathbf{x}) - \sum_{q=0}^{m} \sum_{j=1}^{r} g_j \circ \xi_q(\mathbf{x})$. This completes step r.

Let $\mathbf{x}$ be an arbitrary point of I^n. For at least $m + 1 - n$ values of q, $\mathbf{x} \in S^q_{k_r}(\mathbf{d}_{k_r,q})$ for some $\mathbf{d}_{k_r,q}$'s at every step r. For these q,

$$g \circ \xi_q(\mathbf{x}) = \frac{1}{m+1} f_{r-1}(\mathbf{d}_{k_r,q})$$

and thus

$$\left| \frac{1}{m+1} f_{r-1}(\mathbf{x}) - g \circ \xi_q(\mathbf{x}) \right| \leq \frac{\epsilon}{m+1} \|f_{r-1}\|_\infty.$$

For the remaining values of q's of which $\mathbf{x}$ is not contained in the cubes $S^q_{k_r}(\mathbf{d}_k)$, $g_r \circ \xi_q(\mathbf{x})$ is less that $\frac{1}{m+1}\|f_{r-1}\|_\infty$ in absolute value. Thus it follows that

$$|f_r(\mathbf{x})| = \left| f_{r-1}(\mathbf{x}) - \sum_{q=0}^{m} g_r \circ \xi_q(\mathbf{x}) \right| \leq \left(\frac{m-n+1}{m+1}\epsilon + \frac{2n}{m+1} \right) \|f_{r-1}\|_\infty \leq \theta \|f_{r-1}\|_\infty.$$

Inductively,

$$\|g_r\|_\infty \leq \frac{1}{m+1}\|f_{r-1}\|_\infty \leq \frac{1}{m+1}\theta^{r-1}\|f\|_\infty$$

and

$$\|f_r\|_\infty = \left\| f(\mathbf{x}) - \sum_{q=0}^{m} \sum_{j=1}^{r} g_j \circ \xi_q(\mathbf{x}) \right\|_\infty \leq \theta^r \|f\|_\infty.$$

Therefore $\sum_{r=1}^{\infty} g_r$ converges uniformly and let g be its sum, then $Kf = g$.

2.3 Optimal transport theory

One of the main topics of optimal transport is the optimal cost of transference plans. We first introduce the mathematical formulation of the problem.

Let μ, ν be two probability measures defined on some measure spaces X, Y respectively. A cost function $c(x, y)$ is a measurable map form $X \times Y$ to $\mathbb{R} \cup \{+\infty\}$. $P(X \times Y)$ is the set of probability measures on $X \times Y$. The set of all transference plans

$$\Pi(\mu, \nu) := \{\pi \in P(X \times Y); \pi(A \times Y) = \mu(A), \pi(X \times B) = \nu(B),$$
$$\forall \text{ measurable} \quad A \subset X, B \subset Y.\}$$

is nonempty, since the tensor product $\mu \otimes \nu$ lies in $\Pi(\mu, \nu)$.

$$I[\pi] := \int_{X \times Y} c(x, y) d\pi(x, y), \quad \pi \in \Pi(\mu, \nu)$$

is called the total transportation cost associated to π. The optimal transport problem is to minimize the transportation cost $I[\pi]$ for all $\pi \in \Pi(\mu, \nu)$.

$$\mathcal{T}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} I[\pi]$$

is called the optimal transportation cost between μ and ν. The optimal π's, if they exist, will be called the optimal transference plans. For quadratic cost function $c(x, y) = |x - y|^2$ on $\mathbb{R}^n$, the structure of optimal transference plan is simple and elegant. Here are some results on optimal transference plans.

Theorem 2.3.1 (Optimal transportation theorem for quadratic cost. Theorem 2.12 in [64])**.** *Let μ, ν be probability measures on $\mathbb{R}^n$, with finite second order moments, i.e.,*

$$\int_{\mathbb{R}^n} |x|^2 d\mu(x) + \int_{\mathbb{R}^n} |y|^2 d\nu(y) < \infty.$$

We consider the transportation problem associated with a quadratic cost function $c(x, y) = |x - y|^2$. Then,

(i) (Knott-Smith optimality criterion) $\pi \in \Pi(\mu, \nu)$ is optimal if and only if there exists a

convex lower semi-continuous function ϕ such that

$$\mathrm{Supp}(\pi) \subset \mathrm{Graph}(\partial\phi)$$

or equivalently:

$$\textit{for } d\pi\textit{-almost all } (x, y), \quad y \in \partial\phi(x).$$

Moreover, in that case, the pair (ϕ, ϕ^*) with $\phi^*(y) := \sup_{x\in\mathbb{R}^n}(x \cdot y - \phi(x))$ has to be a minimizer in the problem

$$\inf\left\{\int_{\mathbb{R}^n} \phi d\mu + \int_{\mathbb{R}^n} \psi d\nu : \quad \forall (x, y),\ x \cdot y \leq \phi(x) + \psi(y)\right\}.$$

(ii) (Brenier's theorem) If μ does not give mass to small sets (sets with Lebesgue measure 0), then there is a unique optimal π, which is

$$d\pi(x, y) = d\mu(x)\delta[y = \nabla\phi(x)],$$

or equivalently, $\pi = (\mathrm{Id} \times \nabla\phi)\#\mu$, where $\nabla\phi$ is the unique(i.e., uniquely determined $d\mu$-almost everywhere) gradient of a convex function which pushes μ forward to $\nu : \nabla\phi\#\mu = \nu$. Moreover,

$$\mathrm{Supp}(\nu) = \overline{\nabla\phi(\mathrm{Supp}(\mu))}.$$

(iii) As a corollary, under the assumption of (ii), $\nabla\phi$ is the unique solution to the transportation problem:

$$\int_{\mathbb{R}^n} |x - \nabla\phi(x)|^2 d\mu(x) = \inf_{T\#\mu=\nu} \int_{\mathbb{R}^n} |x - T(x)|^2 d\mu(x)$$

or equivalently,

$$\int_{\mathbb{R}^n} x \cdot \nabla\phi(x) d\mu(x) = \inf_{T\#\mu=\nu} \int_{\mathbb{R}^n} x \cdot T(x) d\mu(x).$$

In the case of real line, $\mathbb{R}$, the solution to the optimal transportation problem can be expressed in terms of cumulative distribution functions. The cumulative distribution function of $\mu \in P(\mathbb{R})$ is $F(x) := \int_{-\infty}^{x} d\mu = \mu[(-\infty, x]]$. One can define the generalised inverse of F on $[0,1]$ by

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) > t\}.$$

Theorem 2.3.2 (Optimal transportation theorem for a quadratic cost on $\mathbb{R}$. Theorem 2.18 in [64]). *Let μ, ν be probability measures on $\mathbb{R}$, with respective cumulative distribution functions F and G. Let π be the probability measure on $\mathbb{R}^2$ with joint two-dimensional cumulative distribution function*

$$H(x, y) = \min(F(x), G(y)).$$

Then, π belongs to $\Pi(\mu, \nu)$, and is optimal in the transference plans between μ and ν for the quadratic cost function $c(x, y) = |x - y|^2$. Moreover, the value of the optimal transportation cost is

$$\mathcal{T}_2(\mu, \nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt.$$

2.3.1 Duality

The optimal transport problem can be studied in its equivalent dual problem.

Theorem 2.3.3 (Kantorovich duality. Theorem 1.3 in [64]). *Let X, Y be Polish spaces, let $\mu \in P(X)$ and $\nu \in P(Y)$, and let $c : X \times Y \to \mathbb{R}_+ \cup \{\infty\}$ be a lower semi-continuous cost function.*

Whenever $\pi \in \Pi(\mu, \nu)$ and $(\phi, \psi) \in L^1(d\mu) \times L^1(d\nu)$, define

$$I[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y), \qquad J(\phi, \psi) = \int_X \phi d\mu + \int_Y \psi d\nu.$$

Define Φ_c to be the set of all measurable functions $(\phi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ satisfying

$$\phi(x) + \psi(y) \leq c(x, y)$$

for all $d\mu$-almost all $x \in X$ and $d\nu$-almost all $y \in Y$.

Then

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \sup_{\Phi_c} J(\phi, \psi). \tag{2.3.1}$$

Moreover, the infimum in the left-hand side of (2.3.1) is attained. Furthermore, it does not change the value of the supremum in the right-hand side of (2.3.1) if one restricts the definition of Φ_c to those functions (ϕ, ψ) which are bounded and continuous.

When the cost function is a metric: $c(x, y) = d(x, y)$ on $X = Y$, then there is more structure in Kantorovich duality principle. Note that this distance need not be the distance defining the topology of the space.

Theorem 2.3.4 (Kantorovich-Rubinstein theorem. Theorem 1.14 in [64])**.** *Let $X = Y$ be a Polish space and d be a lower semi-continuous metric on X. Let $\mathcal{T}_d$ be the cost of optimal transportation for the cost function $c(x, y) = d(x, y)$,*

$$\mathcal{T}_d(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} d(x, y) d\pi(x, y).$$

Let $\mathrm{Lip}(X)$ denote the space of all Lipschitz functions on X, and

$$||\phi||_{\mathrm{Lip}} := \sup_{x \neq y} \frac{|\phi(x) - \phi(y)|}{d(x, y)}.$$

Then

$$\mathcal{T}_d(\mu, \nu) = \sup \left\{ \int_X \phi d(\mu - \nu); \qquad \phi \in L^1(d|\mu - \nu|), \quad ||\phi||_{\mathrm{Lip}} \leq 1 \right\}.$$

Moreover, it does not change the value of the supremum above to impose the additional condition that ϕ be bounded.

2.3.2 Wasserstein distances

When the cost function between two mass points is measured in their distance, the optimal transport cost between two probability measures defines a metric on the paces of probability measures.

Let X be a Polish space. Consider the cost function $c(x,y) = d(x,y)^p$, for $0 \leq p < \infty$ and $d(x,y)^0 := 1_{x \neq y}$. Use the abbreviation $\mathcal{T}_p(\mu,\nu) = \mathcal{T}_{d^p}(\mu,\nu)$. Denote by $P_p(X)$ the set of probability measures with finite moment of order p, namely

$$P_p(X) := \left\{ \mu \in P(X) : \int_X d(x_0,x)^p d\mu(x) < \infty, \text{ for some and thus any } x_0 \in X \right\}.$$

If d is bounded, then $P_p(X)$ coincides with the set $P(X)$.

Theorem 2.3.5 (Wasserstein distances. Theorem 7.3 in [64]). *(i) For all $p \in [1,\infty)$, $W_p := \mathcal{T}_p^{1/p}$ defines a metric on $P_p(X)$.*

(ii) For all $p \in [0,1)$, $W_p = \mathcal{T}_p$ defines a metric on $P_p(X)$.

Wasserstein distance has wide applications in shape recognition [17] and image processing [47].

Chapter 3

Kolmogorov basis

In this Chapter, we investigate the set of Kolmogorov bases and formulate a sufficient condition for a family of continuous functions to be a Kolmogorov basis. Generally, the construction of a K-basis depends on the dimension n of the domain I^n. Let $m > n \geq 2$ be natural numbers. In general, a K-basis on I^m cannot be generated by adding some new functions to a K-basis on I^n and a K-basis on I^n cannot be obtained by subtracting some functions from a K-basis on I^m. Similar construction process of K-bases is repeated for every different $n \geq 2$. Here we show that under certain conditions, a K-basis on I^n can be extended to a K-basis on I^m and a K-basis on I^m can be reduced to a K-basis on I^n.

3.1 Kahane's set of K-basis

Let $\Phi \subseteq C(I)$ denote the metric subspace of $C(I)$, endowed with the ordinary maximum metric, of all increasing functions $\phi : I \to I$. Φ^n is the metric product space of Φ. As mentioned in Section 1.3, the set of continuously increasing functions $(\phi_0, ..., \phi_{2n}) \in \Phi^{2n+1}$ which can form a Kolmogorov basis is of second category.

Theorem 3.1.1 (Kahane [24], Hedberg [20]). *Let $\lambda_p > 0$ with $\sum_{p=1}^{n} \lambda_p \leq 1$ be rationally independent numbers. The set $K \subset \Phi^{2n+1}$ for all such tuples $(\phi_0, ..., \phi_{2n}) \in \Phi^{2n+1}$ which* *has the property that for any $f \in C(I^n)$ there is some $g \in C(\mathbb{R})$ satisfying*

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_p \phi_q(x_p)\right)$$

is of second Baire category *.

To investigate paths between K-bases $\{\xi_q := \sum_{p=1}^n \lambda_p \phi_q\}_{q=0}^{2n}$ with $(\phi_0, ..., \phi_{2n})$ from Kahane's set, we first fix the functions $(\phi_0, ..., \phi_{2n})$ and change the parameter λ_p's. It turns out that the set of parameter λ_p's are dense but not connected. Second, we find that the all K-bases in Kahane's set can be generated by any K-basis in the set.

Proposition 3.1.2. *The set of parameters,*

$$\Lambda := \{(\lambda_1, ..., \lambda_n) : \lambda_1, ..., \lambda_n > 0, \textit{ rationally independent and } \sum_{p=1}^{n} \lambda_p \leq 1\}$$

is dense in the set $\{(\lambda_1, ..., \lambda_n) : \lambda_1, ..., \lambda_n > 0,$ *with* $\sum_{p=1}^n \lambda_p \leq 1\}$. Λ *is not connected and thus not simply connected.*

Any $(\phi_0, ..., \phi_{2n}) \in \Phi^{2n+1}$ *generates all other elements in* Φ^{2n+1}. *Precisely, for any* $(\tilde{\phi}_0, ..., \tilde{\phi}_{2n}) \in \Phi^{2n+1}$, *there exist homeomorphisms* $L_q : \phi_q[I] \to \tilde{\phi}_q[I]$ *such that*

$$\tilde{\phi}_q(x_p) = L_q \circ \phi_q(x_p), \quad \forall q.$$

Remark 3.1.3. *Proposition 3.1.2 implies that one cannot build a continuous path between Kahane's K-bases by continuously changing only the parameter λ_p's.*

Let

$$\Lambda_n := \{(\lambda_1, ..., \lambda_n) : \lambda_1, ..., \lambda_n \text{ rationally independent } \}.$$

When $n = 2$,

$$\Lambda_2 = \{(\lambda_1, \lambda_2) : \lambda_1 \in \mathbb{Q}, \lambda_2 \notin \mathbb{Q}\} \cup \{(\lambda_1, \lambda_2) : \lambda_1 \notin \mathbb{Q}, \lambda_2 \notin \lambda_1\mathbb{Q}\}.$$

---

*A set K is said to be of second category if it is not included in a countable union of nowhere dense sets.

Generally,

$$\Lambda_n = \{(\lambda_1, ..., \lambda_n) : \lambda_1 \in \mathbb{R}, \lambda_2 \notin \lambda_1\mathbb{Q}, \lambda_3 \notin \lambda_1\mathbb{Q} + \lambda_2\mathbb{Q}, ..., \lambda_n \notin \lambda_1\mathbb{Q} + \cdots + \lambda_{n-1}\mathbb{Q}\}.$$

Proof of Proposition 3.1.2. Given any n rationally independent numbers $\lambda_1^0, ..., \lambda_n^0 \in I$, the set $\{(\lambda_1, ..., \lambda_n) : \lambda_p \in I \cap \lambda_p^0\mathbb{Q},\ p = 1, ..., n,\ \text{with } \sum_{p=1}^n \lambda_p \leq 1\} \in \Lambda$ is dense and thus Λ is dense in $\{(\lambda_1, ..., \lambda_n) : \lambda_1, ..., \lambda_n > 0,\ \text{with } \sum_{p=1}^n \lambda_p \leq 1\}$.

Let

$$A := \{(\lambda_1, ..., \lambda_n) \in \Lambda : \lambda_1 < \lambda_2\} \quad \text{and} \quad B := \{(\lambda_1, ..., \lambda_n) \in \Lambda : \lambda_1 > \lambda_2\},$$

then A and B are open in the relative topology induced on the Λ. $\Lambda = A \cup B$. Therefore, Λ is not connected.

For any $(\phi_0, ..., \phi_{2n}), (\tilde{\phi}_0, ..., \tilde{\phi}_{2n}) \in \Phi^{2n+1}$, let $L_q = \tilde{\phi}_q \circ \phi_q^{-1}$, then $\tilde{\phi}_q(x_p) = L_q \circ \phi_q(x_p)$ for all q. $\square$

Moreover, given two K-basis $(\phi_0, ..., \phi_{2n}) \in \Phi^{2n+1}$, $(\tilde{\phi}_0, ..., \tilde{\phi}_{2n}) \in \Phi^{2n+1}$, consider the matrix of homeomorphism L of size $(2n+1) \times (2n+1)$ from Φ^{2n+1} to Φ^{2n+1} such that

$$\begin{pmatrix} \tilde{\phi}_0 \\ \tilde{\phi}_1 \\ \vdots \\ \tilde{\phi}_{2n} \end{pmatrix} = \begin{pmatrix} L_{00} & L_{01} & \cdots & L_{0,2n} \\ L_{10} & L_{11} & \cdots & L_{1,2n} \\ \vdots & \vdots & \ddots & \vdots \\ L_{2n,0} & L_{2n,1} & \cdots & L_{2n,2n} \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_{2n} \end{pmatrix}, \tag{3.1.1}$$

that is,

$$\tilde{\phi}_q = \sum_{q'=0}^{2n} L_{q,q'} \circ \phi_{q'}, \quad \forall q.$$

The matrix L between given two K-basis $(\phi_0, ..., \phi_{2n}), (\tilde{\phi}_0, ..., \tilde{\phi}_{2n}) \in \Phi^{2n+1}$ is not unique.

For example, L can be

$$\begin{pmatrix} \tilde{\phi}_0 \circ \phi_0^{-1} & 0 & \cdots & 0 \\ 0 & \tilde{\phi}_1 \circ \phi_1^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\phi}_{2n} \circ \phi_{2n}^{-1} \end{pmatrix}$$

or

$$L = (a_{ij}\tilde{\phi}_i \circ \phi_j^{-1})_{(2n+1)\times(2n+1)},$$

where $(a_{ij})_{(2n+1)\times(2n+1)}$ is an arbitrary stochastic matrix (see the definition in Appendix).

On the other hand, for any $(\phi_0, ..., \phi_{2n})$ in Kahane's set and matrix of homeomorphisms L, if $L \circ (\phi_0, ..., \phi_{2n})^T$ separates a Kolmogorov cover (see next subsection 3.2 for detailed definition), then it forms a K-basis in Kahane's set. As some simple examples, if L is the elementary matrix obtained by changing two rows of the identity matrix or by multiplying a positive numbers to any row of the identity matrix, then $L \circ (\phi_0, ..., \phi_{2n})^T$ generates a new K-basis.

3.2 Projection and extension of K-bases among dimensions

As a special case, Sprecher [51] constructed a universal function ϕ such that ϕ can be used to form a K-basis for all dimension $n \geq 2$.

Theorem 3.2.1 (Sprecher [51]). *Let $\{\lambda_k\}$ be a sequence of positive rationally independent numbers. There exists a continuous monotonically increasing function $\phi : [0, 1 + \frac{1}{5!}] \to [0, 1 + \frac{1}{5!}]$ having the following property: For every real-valued continuous function $f : I^n \to \mathbb{R}$ with $n \geq 2$ there are continuous functions g_q such that*

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g_q \left(\sum_{p=1}^{n} \lambda_p \phi(x_p + qa_n) \right), \tag{3.2.1}$$

for a suitable constant a_n.

One can separate the domains of definition of the $2n + 1$ outer functions g_q in (3.2.1) by adding a suitable constant to the argument $\xi_q \equiv \sum_{p=1}^{n} \lambda_p \phi(x_p + qa_n)$, and thus replace the g_q's with a single g. This was observed first by Lorentz [39] and remarked in Sprecher [51].

In general, the univariate functions ϕ_q constructed in a K-basis for a given dimension n cannot be reused in another dimension $m \neq n$. In fact, for a given dimension $n \geq 2$, the minimal number of K-basis for $C(I^n)$ is $2n+1$ [57], which clearly depends on n. The construction of the functions $(\phi_0, ..., \phi_{2n})$ in the corresponding K-basis $\xi_q(x_1, ..., x_n) = \sum_{p=1}^{n} \lambda_p \phi_q(x_p)$ also depends on n. To investigate the conditions under which ϕ_q's constructed for dimension n can be reused in another dimension $m \neq n$, we first examine the properties needed for $\{\xi_q\}_{q=0}^{2n}$ to be a Kolmogorov basis in dimension $n \geq 2$.

3.2.1 Kolmogorov cover

Definition 3.2.2. *Let $n \geq 2$ and D_k^n be the Cartesian product of D_k, where $D_k \subset \mathbb{N}$ is a finite set. We call a sequence of $2n+1$ q-families of disjoint closed sets*

$$\{S_{q,\mathbf{i}}^k : \mathbf{i} \in D_k^n, q = 0, ..., 2n\}_{k \in \mathbb{N}}$$

***Kolmogorov cover** of I^n, iff*

(i) Diameter of $S_{q\mathbf{i}}^k$ goes to 0 with $k \to \infty$,

(ii) For any fixed $k \in \mathbb{N}$, the family of cubes $\{S_{q\mathbf{i}}^k : \mathbf{i} \in D_k^n; q = 0, ..., 2n\}$ covers I^n at least $n+1$ times.

*We say that a family of continuous functions $\{\xi_0, ..., \xi_{2n}\}$ with $\xi_q : I^n \to \mathbb{R}$ **separates a Kolmogorov cover** $\{S_{q\mathbf{i}}^k\}_{k \in \mathbb{N}}$ iff for any fixed $k \in \mathbb{N}$, the image of cubes*

$$\{S_{q\mathbf{i}}^k, \mathbf{i} \in D_k^n, q = 0, ..., 2n\}$$

under $\{\xi_0, ..., \xi_{2n}\}$ are all mutually disjoint.

As the routine of the constructive proofs of KST, one first sets up a Kolmogorov cover, then constructs a family of functions that separates a Kolmogorov cover and this family of functions is naturally a Kolmogorov basis (see section 2.2).

Lemma 3.2.3. *Let $n \geq 2$ be a natural number. For $q = 0, ..., 2n$, let $\xi_q : I^n \to \mathbb{R}$ be continuous. If $\{\xi_q\}_{q=0}^{2n}$ separates a Kolmogorov cover for I^n, then $\{\xi_q\}_{q=0}^{2n}$ is a Kolmogoorv basis on I^n.*

Moreover, let $\lambda_1, ..., \lambda_n > 0$ and rationally independent, and $\phi_q : I \to \mathbb{R}$ continuous and strictly increasing. If $\xi := \{\sum_{p=1}^{n} \lambda_p \phi_q\}_{q=0}^{2n}$ separates a Kolmogorov cover, then ξ belongs to Kahane's set of K-basis, e.g., Lorentz's K-basis.

Let $D_k \subset \mathbb{N}$ be a finite set. Suppose that there exists a sequence of $2n + 1$ families of disjoint sub-intervals

$$\{I_{q,i}^k \subset I : i \in D_k, q = 0, ..., 2n\}_{k \in \mathbb{N}}$$

such that for any $x \in I$, x is not covered by q-intervals $I_{q,i}^k$ for at most one of $q \in \{0, ..., 2n\}$. One can generate a Kolmogorov cover of I^n by the Cartesian products of $I_{q,i}^k$:

$$S_{q,\mathbf{i}}^k = \prod_{p=1}^{n} I_{q,i_p}^k, \quad \mathbf{i} := (i_1, ..., i_n) \in D_k^n. \tag{3.2.2}$$

Since for any $\mathbf{x} \in I^n$ there exists at most n of q's such that $\mathbf{x}$ is not covered by q-cubes $S_{q,\mathbf{i}}^k$, it means that $\mathbf{x}$ is covered by at least the other $n+1$ q-cubes. Therefore, $\{S_{q,\mathbf{i}}^k\}$ defined as in (3.2.2) is indeed a Kolmogorov cover for I^n.

A Kolmogorov cover constructed as in (3.2.2) is called a **Kolmogorov cover of Cartesian type**. For example, the Kolmogorov covers constructed by Kolmogorov [29], Lorentz[40] and Sprecher [48] are all of Cartesian type.

Lemma 3.2.4. *If Let $m > n \geq 2$ be natural numbers and*

$$\{S_{q,\mathbf{i}}^k(m) := \prod_{p=1}^{m} I_{q,i_p}^k : \mathbf{i} \in D_k^m, q = 0, ..., 2m\}_{k \in \mathbb{N}}$$

be a Kolmogorov cover of Cartesian type for I^m, then

$$\{S_{q,\mathbf{i}}^k(n) := \prod_{p=1}^{n} I_{q,i_p}^k : \mathbf{i} \in D_k^n, q = 0, ..., 2n\}_{k \in \mathbb{N}}$$

is a Kolmogorov cover of Cartesian type for I^n.

Proof. By the property of $I^k_{q,i}$, $q = 0, ..., 2n$, for any $\mathbf{x} \in I^n$, there exists at most n of q's such that $\mathbf{x}$ is not covered by q-cubes $S^k_{q,\mathbf{i}}(n)$, then $\mathbf{x}$ is covered by at least the remaining $n+1$ q-cubes. Therefore, $\{S^k_{q,\mathbf{i}} := \prod_{p=1}^n I^k_{q,i_p} : \mathbf{i} \in D^n_k, q = 0, ..., 2n\}_{k\in\mathbb{N}}$ is a Kolmogorov cover of Cartesian type for I^n. □

On the other hand, a Kolmogorov cover of Cartesian type in a low dimension cannot always be extended to high dimension by adding more shifted families of the sub-cubes. We take Lorentz's [40] construction of Kolmogorov cover in dimension 2 for example.

Example 3.2.5. *(1) For $k \in \mathbb{N}$, $i = 0, ..., 10^{k-1}$ and $q = 0, ..., 4$, let*

$$I^k_{0i} = [i \cdot 10^{-k+1} + 10^{-k}, i \cdot 10^{-k+1} + 9 \cdot 10^{-k}]$$

and

$$I^k_{qi} = I^k_{0i} - 2q \cdot 10^{-k}.$$

Replace by $I \cap I^k_{qi}$ those intervals I^k_{qi} that are not entirely contained in I. Then any $x \in I$ is not contained by at most one q-interval. Therefore

$$S^k_{q,\mathbf{i}}(2) := I^k_{qi_1} \times I^k_{qi_2} \quad \text{with } \mathbf{i} = (i_1, i_2) \in \{0, ..., 10^{k-1}\}^2$$

is a Kolmogorov cover of Cartesian type for I^2. However, $S^k_{q,\mathbf{i}}(2)$ cannot be extended to a Kolmogorov cover to dimension 3. If we allow $q = 5, 6$, then $I^k_{5i} = I^k_{0i} - 10 \cdot 10^{-k} = I^k_{0,i-1}$ and $I^k_{6i} = I^k_{0i} - 12 \cdot 10^{-k} = I^k_{1,i-1}$, which are repetitions of q-intervals for $q = 0, 1$ respectively.

(2) If we let

$$I^k_{0i} = [i \cdot 14^{-k+1} + 14^{-k}, i \cdot 14^{-k+1} + 13 \cdot 14^{-k}], \quad k \in \mathbb{N},\ i = 0, ..., 14^{k-1}$$

and

$$I^k_{qi} = I^k_{0i} - 2q \cdot 14^{-k}, \quad q = 0, ..., 4.$$

Again replace by $I \cap I_{qi}^k$ those intervals I_{qi}^k that are not entirely contained in I. Then

$$S_{q,\mathbf{i}}^k(2) := I_{qi_1}^k \times I_{qi_2}^k \quad \textit{with } \mathbf{i} = (i_1, i_2) \in \{0, ..., 14^{k-1}\}^2$$

is a Kolmogorov cover of Cartesian type for I^2.

Moreover, adding $I_{qi}^k := I_{0i}^k - 2q \cdot 14^{-k}, q = 5, 6$ to $\{I_{qi}^k,\ q = 0, ..., 4\}$,

$$S_{q,\mathbf{i}}^k(3) = I_{qi_1}^k \times I_{qi_2}^k \times I_{qi_3}^k \quad \textit{with } \mathbf{i} = (i_1, i_2, i_3) \in \{0, ...14^{k-1}\}^3$$

is a Kolmogorov cover of Cartesian type for I^3 extended from $\{S_{q,\mathbf{i}}^k(2)\}$.

(3) More general, let $m > n \geq 2$ and for $k \in \mathbb{N}, i = 0, ..., (4m+2)^{k-1}$,

$$I_{0i}^k = [i \cdot (4m+2)^{-k+1} + (4m+2)^{-k}, i \cdot (4m+2)^{-k+1} + (4m+1) \cdot (4m+2)^{-k}]$$

and

$$I_{qi}^k = I_{0i}^k - 2q \cdot (4m+2)^{-k}, \quad q = 0, ..., 2n.$$

In a similar way, the Kolmogorov cover for dimension n,

$$S_{q,\mathbf{i}}^k(n) := \prod_{p=1}^{n} I_{qi_p}^k \quad \textit{with } \mathbf{i} = (i_1, ..., i_n) \in \{0, ...(4m+2)^{k-1}\}^n,\ q = 0, ..., 2n$$

can be extend to a Kolmogorov cover for $I^l, n < l \leq m$ by adding $2(l-n)$ new q-shifts of I_{qi}^k.

3.2.2 Projection and extension of K-basis

From the projection property of Kolmogorov cover of Cartesian type in Lemma 3.2.4, the projection of K-basis from high dimensions to low dimensions follows:

Theorem 3.2.6 (K-basis projection theorem)**.** *Let $m > n \geq 2$ be natural numbers and*

$$\{S_{q,\mathbf{i}}^k(m) := \prod_{p=1}^{m} I_{q,i_p}^k, \mathbf{i} \in D_k^m, q = 0, ..., 2m\}_{k \in \mathbb{N}}$$

be a Kolmogorov cover of Cartesian type for I^m*. Let* $\lambda_1, ..., \lambda_m > 0$ *be rationally independent numbers, and* $\phi_q : I \to I$*,* $q = 0, ..., 2m$*, be monotonously increasing, continuous functions such that* $\{\sum_{p=1}^{m} \lambda_p \phi_q(x_p) : q = 0, ..., 2m\}$ *forms a K-basis on* I^m *that separates* $\{S_{q,\mathbf{i}}^k(m)\}$*. Then* $\{\sum_{p=1}^{n} \lambda_p \phi_q(x_p) : q = 0, ..., 2n\}$ *also forms a K-basis on* I^n *that separates*

$$\{S_{q,\mathbf{i}}^k(n) := \prod_{p=1}^{n} I_{q,i_p}^k, \mathbf{i} \in D_k^n, q = 0, ..., 2n\}_{k\in\mathbb{N}}. \tag{3.2.3}$$

Proof. For any $k \in \mathbb{N}, q = 0, ..., 2n$ and $\mathbf{i} \in D_k^n$, since the image of $S_{q,\mathbf{i}}^k(n)$ under $\sum_{p=1}^{n} \lambda_p \phi_q(x_p)$ after a shift $\sum_{p=n+1}^{m} \lambda_p \phi_q(0)$ is contained in the image of $S_{q,\mathbf{i}}^k(n) \times (I_{q,0}^k)^{m-n}$ under $\sum_{p=1}^{m} \lambda_p \phi_q(x_p)$, the disjointness of the images of $S_{q,\mathbf{i}}^k(n) \times (I_{q,0}^k)^{m-n}$ implies the disjointness of the images of $S_{q,\mathbf{i}}^k(n)$ after the shift $\sum_{p=n+1}^{m} \lambda_p \phi_q(0)$, and thus the disjointness of the images of $S_{q,\mathbf{i}}^k(n)$.

More precisely, for any fixed $k \in \mathbb{N}$, $(i_1, ..., i_n) \in D_k^n$, and any cubes $S_{q,\mathbf{i}}^k(n) = \prod_{p=1}^{n} I_{q,i_p}^k$, let

$$S_{q,\mathbf{i}}^k(m) = \prod_{p=1}^{n} I_{q,i_p}^k \times \prod_{p=n+1}^{m} I_{q,0}^k,$$

then

$$\sum_{p=1}^{n} \lambda_p \phi_q[I_{q,i_p}^k] + \sum_{p=n+1}^{m} \lambda_p \phi_q(0) \subset \sum_{p=1}^{m} \lambda_p \phi_q[I_{q,i_p}^k],$$

where $i_{n+1} = \cdots = i_m = 0$ and the plus "+" in the left hand side means a shift of the set $\sum_{p=1}^{n} \lambda_p \phi_q[I_{q,i_p}^k]$ by the constant $\sum_{p=n+1}^{m} \lambda_p \phi_q(0)$.

Therefore, $\{\sum_{p=1}^{n} \lambda_p \phi_q(x_p) : q = 0, ..., 2n\}$ separates the new Kolmogorov cover (3.2.3) and thus forms a Kolmogorov basis. $\square$

The extension of K-basis requires more on the K-basis. Let

$$\{S_{q,\mathbf{i}}^k(n) := \prod_{p=1}^{n} I_{q,i_p}^k : \mathbf{i} \in D_k^n, q = 0, ..., 2n\}_{k\in\mathbb{N}}$$

be a Kolmogorov cover of Cartesian type for I^n. $\{\sum_{p=1}^{n} \lambda_p \phi_q(x_p) : q = 0, ..., 2n\}$ is a K-basis on I^n that separates $\{S_{q,\mathbf{i}}^k(n)\}$. For any fixed $k \in \mathbb{N}$ and $q = 0, ..., 2n$, write

$$I_{q,i}^k := [\alpha_{qi}^k, \beta_{qi}^k], \quad i \in D_k.$$

Define for $i \in D_k, 0 \leq q \leq 2n$ and $k \in \mathbb{N}$,

$$\epsilon_{qi}^k := \frac{1}{2}(\phi_q(\beta_{qi}^k) - \phi_q(\alpha_{qi}^k)), \text{ and } \epsilon_k := \max_{\substack{i \in D_k \\ 0 \leq q \leq 2n}} \{\epsilon_{qi}^k\}. \tag{3.2.4}$$

Theorem 3.2.7 (K-basis extension theorem)**.** *Let $m > n \geq 2$ be natural numbers and $\{S_{q,\mathbf{i}}^k(n) := \prod_{p=1}^n I_{q,i_p}^k : \mathbf{i} \in D_k^n, q = 0, ..., 2n\}_{k \in \mathbb{N}}$ be a Kolmogorov cover of Cartesian type for I^n. Let $\lambda_1, ..., \lambda_n > 0$ be rationally independent numbers, and $\phi_q : I \to I$, $q = 0, ..., 2n$, be monotonously increasing, continuous functions such that $\{\sum_{p=1}^n \lambda_p \phi_q(x_p) : q = 0, ..., 2n\}$ forms a K-basis on I^n that separates $\{S_{q,\mathbf{i}}^k(n)\}$. Suppose that $\{I_{qi}^k, i \in D_k, q = 2n+1, ..., 2m\}$ can be added to $\{I_{qi}^k, i \in D_k, q = 0, ..., 2n\}$ such that every point $x \in I$ is not covered by at most one of the $2m+1$ q-intervals I_{qi}^k.*

Let $\epsilon_{qi}^k, \epsilon_k$ be as in (3.2.4). If there exists $\lambda_{n+1}, ..., \lambda_m > 0$ such that $\lambda_1, ..., \lambda_m$ are rationally independent with $\sum_{p=1}^m \lambda_p \leq 1$ and

$$\min_{\substack{0 \leq q, q' \leq 2n \\ \sum_{p=1}^m |i_p - i'_p| \neq 0}} \left\{ \left| \sum_{p=1}^m \lambda_p \left[(\phi_q(\alpha_{q,i_p}^k) + \epsilon_{q,i_p}^k) - (\phi_{q'}(\alpha_{q',i'_p}^k) + \epsilon_{q',i'_p}^k) \right] \right| \right\} > 2\epsilon_k \tag{3.2.5}$$

holds for infinitely many $k \in \mathbb{N}$, then we can add continuous functions $\phi_q : I \to I, q = 2n+1, ..., m$, monotonously increasing such that $\{\sum_{p=1}^m \lambda_p \phi_q(x_p) : q = 0, ..., 2m\}$ also forms a K-basis on I^m that separates

$$\{S_{q,\mathbf{i}}^k(m) := \prod_{p=1}^m I_{q,i_p}^k ; \mathbf{i} \in D_k^m; q = 0, ..., 2m\}_{k \in \mathbb{N}}.$$

We first show the idea of the proof of Theorem 3.2.7 by the extension of Lorentz's K-basis from dimension 2 to dimension 3.

Example 3.2.8. *In Lorentz's K-basis in dimension 2, suppose the Kolmogorov cover for I^2 is*

$$\{S_{q,\mathbf{i}}^k(2) = I_{qi_1}^k \times I_{qi_2}^k : \mathbf{i} = (i_1, i_2) \in D_k^2 := \{0, ..., 14^{k-1}\}^2, \ q = 0, ..., 4\}_{k \in \mathbb{N}}.$$

As shown in Example 3.2.5 (2), it can be extended to a K-cover in dimension 3*:*

$$\{S^k_{q,\mathbf{i}}(3) = I^k_{qi_1} \times I^k_{qi_2} \times I^k_{qi_3} : \mathbf{i} = (i_1, i_2, i_3) \in D^3_k, q = 0, ..., 6\}_{k\in\mathbb{N}}. \tag{3.2.6}$$

We first show that under some assumption on the structure of ϕ_q*'s, the family of functions* $\{\lambda_1\phi_q(x_1) + \lambda_2\phi_q(x_2) + \lambda_3\phi_q(x_2), q = 0, ..., 4\}$ *separates the cubes*

$$\{S^k_{q,\mathbf{i}}(3) : \mathbf{i} \in D^3_k, q = 0, ..., 4\}$$

for any $k \in \mathbb{N}$*.*

Denote the endpoint of intervals by $I^k_{q,i} := [\alpha^k_{qi}, \beta^k_{qi}]$*. From Theorem 1 in Chapter 11, [40],* $\phi_q, q = 0, ..., 4$*, are defined iteratively on the set of endpoints of* I^k_{qi} *at step* $k \in \mathbb{N}$ *and then extended continuously to the whole interval* I*.*

Suppose for all $l < k$*,* $\{\phi_q(\alpha^l_{qi}), \phi_q(\beta^l_{qi}) : i \in D_k, q = 0, ..., 4\}$ *have been defined. At step k, first let* ϕ_q *be a constant on intervals* $I^k_{q,i} = [\alpha^k_{qi}, \beta^k_{qi}]$*, i.e.,* $\phi_q(\alpha^k_{qi}) = \phi_q(\beta^l_{qi}) := \phi^k_{qi}$ *such that* $\phi^k_{qi}, i \in D_k, q = 0, ..., 4$ *are rational and distinct and* ϕ_q*'s are monotonously increasing on the defined set. Since* λ_1, λ_2 *are rationally independent, the values*

$$\lambda_1\phi^k_{qi} + \lambda_2\phi^k_{qj}, \quad q = 0, ..., 4, \quad i, j \in D_k \tag{3.2.7}$$

are distinct. Choose $\delta^k_2 > 0$ *so small that the* $2\delta^k_2$*-neighbourhood of the points (3.2.7) are disjoint and contained in* $[\phi_q(0), \phi_q(1)]$*. This allows one to amend the values on* ϕ_q *on the endpoints* $\alpha^k_{qi}, \beta^k_{qi}$ *to make it strictly increasing. For each* $I^k_{q,i}$*, select* $\phi_q(\alpha^k_{qi}), \phi_q(\beta^k_{qi})$ *in the* δ^k_2*-neighbourhood of* ϕ^k_{qi} *such that* $\phi_q(\alpha^k_{qi}) < \phi^k_{qi} < \phi_q(\beta^k_{qi})$*. Then the images of any squares* $\{S^k_{q,\mathbf{i}}(2), q = 0, ..., 4\}$ *for* $\mathbf{i} = (i, j)$*:*

$$[\lambda_1\phi_q(\alpha^k_{qi}) + \lambda_2\phi_q(\alpha^k_{qj}), \lambda_1\phi_q(\beta^k_{qi}) + \lambda_2\phi_q(\beta^k_{qj})]$$

are contained in the $2\delta^k_2$*-neighbourhood of* $\lambda_1\phi^k_{qi} + \lambda_2\phi^k_{qj}$ *and thus disjoint.*

For any fixed $\lambda_3 > 0$ *such that* $\lambda_1, \lambda_2, \lambda_3$ *are rationally independent with* $\lambda_1+\lambda_2+\lambda_3 < 2$*, the points*

$$\lambda_1\phi^k_{qi} + \lambda_2\phi^k_{qj} + \lambda_3\phi^k_{ql}, \quad q = 0, ..., 4, \quad i, j, l \in D_k \tag{3.2.8}$$

are distinct.

Now we assume that δ_2^k *chosen before is so small that the* $(\lambda_1+\lambda_2+\lambda_3)\delta_2^k$*-neighbourhoods of point (3.2.8) are disjoint. That is, condition (3.2.5) is satisfied with* $\epsilon_k = 2\delta_2^k$*.*

The image of cube $S_{q,\mathbf{i}}^k(3)$ *with* $\mathbf{i} = (i,j,l)$ *under* $\lambda_1\phi_q(x_1) + \lambda_2\phi_q(x_2) + \lambda_3\phi_q(x_3), q = 0, ..., 4$*:*

$$[\lambda_1\phi_q(\alpha_{qi}^k) + \lambda_2\phi_q(\alpha_{qj}^k) + \lambda_3\phi_q(\alpha_{ql}^k), \lambda_1\phi_q(\beta_{qi}^k) + \lambda_2\phi_q(\beta_{qj}^k) + \lambda_3\phi_q(\beta_{ql}^k)]$$

is contained in the $(\lambda_1 + \lambda_2 + \lambda_3)\delta_2^k$*-neighbourhood of and thus the* $2\delta_2^k$*-neighbourhood of* $\lambda_1\phi_{qi}^k + \lambda_2\phi_{qj}^k + \lambda_3\phi_{ql}^k$*. Thus, the images of cubes* $\{S_{q,\mathbf{i}}^k(3), \mathbf{i} \in D_k^3, q = 0, ..., 4\}$ *under* $\lambda_1\phi_q(x_1) + \lambda_2\phi_q(x_2) + \lambda_3\phi_q(x_3)$ *are disjoint.*

Second, we need to construct $\phi_q : I \to \mathbb{R}, q = 5, 6,$ *continuous and monotonously increasing such that* $\{\lambda_1\phi_q(x_1) + \lambda_2\phi_q(x_2) + \lambda_3\phi_q(x_2), q = 5, 6\}$ *separates the cubes* $\{S_{q,\mathbf{i}}^k(3) : \mathbf{i} \in D_k^3, q = 5, 6\}$*.*

Since for $q = 0, ..., 4$*,* $0 \le \phi_q(0) < \phi_q(1) < 1$*, one can define* $0 \le \phi_q(0) < \phi_q(1) < 1, q = 5, 6$*, in such a way that* $\{\phi_q(0), \phi_q(1) : q = 5, 6\}$ *are rational and distinct and*

$$(\cup_{q=5,6}[\phi_q(0), \phi_q(1)]) \cap (\cup_{0\le q\le 4}[\phi_q(0), \phi_q(1)]) = \emptyset. \tag{3.2.9}$$

Applying Lemma 1 in Chapter 11 of [40], there exist strictly increasing, continuous functions $\phi_q : I \to I, q = 5, 6$*, such that for each fixed rank* $k \in \mathbb{N}$*, the intervals*

$$\Delta_{q,\mathbf{i}}^k(3) = \left[\sum_{p=1}^{3} \lambda_p\phi_q(\alpha_{q,i_p}^k), \sum_{p=1}^{3} \lambda_p\phi_q(\beta_{q,i_p}^k)\right], \quad q = 5, 6,\ \mathbf{i} \in D_k^2$$

are all disjoint.

For $q = 0, ..., 6$*,* ϕ_q *is strictly increasing on* I*. Then by (3.2.9),* $\phi_q[I] = [\phi_q(0), \phi_q(1)], q = 5, 6$ *are disjoint from* $\cup_{0\le q\le 4}\phi_q[I]$*. Therefore,* $\Delta_{q,\mathbf{i}}^k(3), q = 5, 6$ *are also disjoint from* $\Delta_{q,\mathbf{i}}^k(3), q = 0, ..., 4$*.*

In summary, we find λ_3 *and construct* ϕ_5, ϕ_6 *such that*

$$\{\lambda_1\phi_q(x_1) + \lambda_2\phi_q(x_2) + \lambda_3\phi_q(x_2), q = 0, ..., 6\}$$

separates the Kolmogorov cover (3.2.6) and thus is a Kolmogorov basis in dimension 3.

Now we show the extension of K-basis in general case.

Proof of Theorem 3.2.7. If $\{I_{qi}^k, i \in D_k, q = 2n+1, ..., 2m\}$ can be added to $\{I_{qi}^k, i \in D_k, q = 0, ..., 2n\}$ such that every point $x \in I$ is not covered by at most one of the $2m+1$ q-intervals I_{qi}^k, then the family of cubes

$$\{S_{q,\mathbf{i}}^k(m) := \prod_{p=1}^{m} I_{q,i_p}^k : \mathbf{i} \in D_k^m, q = 0, ..., 2m\}_{k \in \mathbb{N}}$$

forms a Kolmogorov cover of Cartesian type for I^m.

Suppose there exists $\lambda_{n+1}, ..., \lambda_m > 0$ such that $\lambda_1, ..., \lambda_m$ are rationally independent with $\sum_{p=1}^m \lambda_p \leq 1$ and condition (3.2.5) is satisfied. The image of $S_{q,\mathbf{i}}^k(m) = \prod_{p=1}^m I_{q,i_p}^k$ is

$$\xi_q[S_{q,\mathbf{i}}^k(m)] = \left[\sum_{p=1}^{m} \lambda_p \phi_q(\alpha_{q,i_p}^k), \sum_{p=1}^{m} \lambda_p \phi_q(\beta_{q,i_p}^k)\right], \quad q = 0, ..., 2n.$$

For any two cubes $S_{q,\mathbf{i}}^k(m) = \prod_{p=1}^m I_{q,i_p}^k$ and $S_{q',\mathbf{i}'}^k(m) = \prod_{p=1}^m I_{q',i'_p}^k$, condition (3.2.5) implies the distance between the middle points of intervals $\xi_q[S_{q,\mathbf{i}}^k(m)]$ and $\xi_{q'}[S_{q',\mathbf{i}'}^k(m)]$ is greater than sum of the half lengths of the two intervals. Therefore, the image intervals $\xi_q[S_{q,\mathbf{i}}^k(m)]$ and $\xi_{q'}[S_{q',\mathbf{i}'}^k(m)]$ are disjoint.

For $q = 2n+1, ..., 2m$, ϕ_q can be constructed as in Lorentz' version of KST such that the image of $\{S_{q,\mathbf{i}}^k(m), q = 2n+1, ..., 2m\}$ are mutually disjoint under the map of $\{\sum_{p=1}^m \lambda_p \phi_q(x_p), q = 2n+1, ..., 2m\}$ and also disjoint with the images of $\{S_{q,\mathbf{i}}^k(m), q = 0, ..., 2n\}$.

More precisely, by Lemma 1 Chapter 11 [40], there exists $2(m-n)$ strictly increasing, continuous functions $\phi_q : I \to I$, $q = 2n+1, ..., 2m$, such that for each fixed rank $k \in \mathbb{N}$, the intervals

$$\Delta_{q,\mathbf{i}}^k = \left[\sum_{p=1}^{m} \lambda_p \phi_q(\alpha_{q,i_p}^k), \sum_{p=1}^{m} \lambda_p \phi_q(\beta_{q,i_p}^k)\right], \quad q = 2n+1, ..., 2m, \ \mathbf{i} \in D_k^m$$

are all disjoint.

In fact, at step $k \in \mathbb{N}$, ϕ_q's are first defined as rational constants on the intervals $I^k_{qi} := [\alpha^k_{qi}, \beta^k_{qi}]$, i.e.,

$$\phi_q(\alpha^k_{qi}) := \phi_q(\alpha^k_{qi}) := \phi^k_{qi}.$$

$\phi^k_{qi}, q = 2n+1, ...2m$, can be so chosen that they are distinct from each other and also disjoint from $\phi_q[I^k_{qi}], i \in D_k, q = 0, ...2n$. Since λ_p's are rationally independent, the values

$$\sum_{p=1}^{m} \lambda_p \phi^k_{q,i_p}, \quad q = 0, ..., 2m, \ (i_1, ..., i_m) \in D^m_k \tag{3.2.10}$$

are distinct. Choose δ_k so small that the $2\delta_k$-neighbourhoods of the points (3.2.10) are mutually disjoint and disjoint from $\sum_{p=1}^{m} \lambda_p \phi_q[I^k_{q,i_p}], (i_1, ..., i_p) \in D^n_k, q = 0, ...2n$. Then amend the values of ϕ_q at the end points $\alpha^k_{qi}, \beta^k_{qi}$ to make ϕ_q strictly increasing. Select $\phi_q(\alpha^k_{qi}), \phi_q(\beta^k_{qi})$ in the δ_k-neighbourhood of $\phi^k_{q,i}$ in such a way that

$$\phi_q(\alpha^k_{qi}) < \phi^k_{q,i} < \phi_q(\beta^k_{qi}), \ q = 2n+1, ..., 2m.$$

If $\beta^k_{qi} = 1$, then the inequality should be $\phi_q(\alpha^k_{qi}) < \phi^k_{q,i} = \phi_q(\beta^k_{qi})$. Similarly, if $\alpha^k_{qi} = 0$, then define $\phi_q(\alpha^k_{qi}) = \phi^k_{q,i} < \phi_q(\beta^k_{qi})$. Then the images of the cubes, $\Delta^k_{q,\mathbf{i}}$ with $\mathbf{i} = (i_1, ..., i_m)$, are contained in the $2\delta_k$-neighbourhoods of the point $\sum_{p=1}^{m} \lambda_p \phi^k_{q,i_p}$ and thus are mutually disjoint.

Therefore, we show that the extended family of functions,

$$\{\sum_{p=1}^{m} \lambda_p \phi_q(x_p) : q = 0, ..., 2m\},$$

forms a K-basis on I^m that separates

$$\{S^k_{q,\mathbf{i}}(m) := \prod_{p=1}^{m} I^k_{q,i_p}; \mathbf{i} \in D^m_k; q = 0, ..., 2m\}_{k \in \mathbb{N}}.$$

□

Chapter 4

Kolmogorov map

In this chapter, we investigate how the properties of f are preserved in its outer function g under given K-bases. We show that for a given K-basis, the outer function g for a given f is generally not unique and g does not preserve the positivity of f. Taking Sprecher's K-basis as a particular case, we also show that the modulus of continuity of f is significantly lost in its outer function g.

4.1 Shape-preserving properties of Kolmogorov maps

The universal presentation formula in KST is not obtained without a price. Fridman's result [15] with Vitushkin's counterexample [66] gives a sharp upper bound for the smoothness of the inner functions; that is, the inner functions ψ_{pq} in (1.1.1) can be at most Lipschitz continuous. If we consider the set of superposition of continuous univariate functions with continuously differentiable inner functions, then the set is nowhere dense shown by the following statement of Vitushkin and Khenkin:

Theorem 4.1.1 (Vitushkin and Khenkin [69]). *For any continuous functions $p_m(x, y)$ and continuously differentiable functions $q_m(x, y)$, $m = 1, ..., N$ and any region $D \subset \mathbb{R}^2$, the set of superposition of the form*

$$\sum_{m=1}^{N} p_m(x, y) f_m(q_m(x, y)),$$

where f_m are arbitrary continuous functions, is nowhere dense in the space of all continuous functions in D with uniform convergence.

Moreover, Vitushkin's [68] and Kolmogorov's result [60] also gives a general upper bound for the smoothness of all functions involved in representing functions of given smoothness in superposition. For example, for the superposition of continuously differentiable functions, we have

Theorem 4.1.2 (Vitushkin [68] and Kolmogorov [60])**.** *Let $\Lambda^n_{p\alpha}$ be the set of all continuous functions of n variables that have uniformly bounded continuous partial derivatives of orders $\leq p$, and additionally those of order p have moduli of continuity not exceeding Mh^α with some constant $M > 0$.*

Let $\chi_0 > 0$ be given and L be the union of all classes $\Lambda^n_{p\alpha}$ with $\chi := (p+\alpha)/n > \chi_0$ and $p + \alpha \geq 1$. Then not all functions of a class $\Lambda^{n_0}_{p_0\alpha_0}$ with $\chi_0 = (p_0+\alpha_0)/n_0$ can be represented as superposition of functions of L.

Roughly speaking, not all functions of a given "complexity" can be represented as superposition of functions with less "complexity".

Back to the superposition in KST, from many graphs simulated in the numerical implementation of KST, we can see that g is highly oscillating. For example, Figure 6.2 and 6.3 in Bryant [8]. This motivates us to have a close examination of the outer function g: given a K-basis, what is the best possibility of g for a given f? If g is not unique in the presentation formula, can we then choose relatively better ones from all available g's?

4.1.1 Non-uniqueness of outer functions

Let X be a compact separable metric space. Let S_ξ be the superposition operator with respect to a K-basis $\{\xi_0, ..., \xi_m\}$ on X. For a given $f \in C(X)$, there are generally more than one $g \in C(\mathbb{R})$ such that

$$S_\xi g := \sum_{q=0}^{m} g \circ \xi_q = f.$$

Theorem 4.1.3 (Non-uniqueness of outer functions)**.** *Let X be a compact separable metric space. Suppose $\{\xi_q : X \to \mathbb{R} |\ q = 0, ..., 2n\}$ is a K-basis such that $Y := \cup_{q=0}^{2n} Y_q :=$*

$\cup_{q=0}^{2n}\xi_q[X]$ is not connected in $\mathbb{R}$, then the superposition operator S with respect to $\{\xi_q\}$ is not injective.

Remark 4.1.4. *For Sprecher's [7] K-basis $\{\xi_q : I^n \to \mathbb{R} | \, q = 0, ..., 2n\}$, the image of I^n under ξ_q are mutually disjoint, i.e., $Y_q \cap Y_{q'} = \emptyset$ for any $q \neq q'$. Thus $Y = \cup_{q=0}^{2n} Y_q$ is not connected.*

In fact, for any K-basis $\{\xi_q\}_{q=0}^{2n}$ on I^n with $Y := \cup_{q=0}^{2n}\xi_q[I^n]$, we can choose a real constant a_0 such that $\tilde{Y}_0 := \xi_0[I^n] - a_0$ is disjoint with $\cup_{q=1}^{2n}\xi_q[I^n]$. Let $\tilde{\xi}_0 := \xi_0 - a_0$, then $\tilde{\xi}_0[I^n] = \tilde{Y}_0$ and $\{\tilde{\xi}_0, \xi_1, ..., \xi_{2n}\}$ is also a K-basis on I^n. Since $\{\xi_q\}_{q=0}^{2n}$ is a K-basis on I^n, for any given $f \in C(I^n)$, there exists a $g \in C(Y)$ such that $f = \sum_{q=0}^{2n} g \circ \xi_q$. Define $\tilde{g}(y) = g(y)$ for $y \in \cup_{q=1}^{2n}\xi_q[I^n]$, $\tilde{g}(y) = g(y - a_q)$ for $y \in \tilde{Y}_0$ and extend $\tilde{g}$ linearly into the gap between $\tilde{Y}_0$ and $\cup_{q=1}^{2n}\xi_q[I^n]$. Then $f = \tilde{g} \circ \tilde{\xi}_0 + \sum_{q=1}^{2n} \tilde{g} \circ \xi_q$. Hence $\{\tilde{\xi}_0, \xi_1, ..., \xi_{2n}\}$ is a K-basis on I^n with $\tilde{Y}_0 \cup Y_1 \cup \cdots \cup Y_{2n}$ not connected.

Proof of Theorem 4.1.3. To show the corresponding superposition operator $S : C(Y) \to C(X)$ is not injective, by duality of adjoint operators (see Appendix), it is equivalent to show that the range of its adjoint operator

$$\begin{aligned} S^* : C(I^n)^* &\longmapsto C(Y)^* \\ \mu &\longmapsto \sum_{q=0}^{2n} \mu \circ \xi_q^{-1} \end{aligned}$$

is not dense in $C(Y)^*$, where $C(X)^*$ and $C(Y)^*$ are the dual spaces of $C(X)$ and $C(Y)$ respectively. We identify $C(X)^*$ and $C(Y)^*$ with the spaces of regular Borel measures with the total variation norm on X and Y respectively. Since Y is not connected, then Y must contain at least two disjoint intervals. Without loss of generality, assume that $Y_0 \cup Y_1$ is disjoint with $\cup_{q=2}^{2n} Y_q$.

Let ν be any Borel measure supported in $Y_0 \cup Y_1$ with $\nu(Y) \neq 0$. We show that there does not exist any $\mu \in C(X)^*$ such that $S^*(\mu) = \nu$. Assume that there exists some

$\mu \in C(X)^*$ such that $S^*(\mu) = \nu$. Since $(Y_0 \cup Y_1) \cap (\cup_{q=2}^{2n} Y_q) = \emptyset$,

$$\begin{aligned} 0 &= \nu(\cup_{q=2}^{2n} Y_q) = S^*(\mu)(\cup_{q=2}^{2n} Y_q) \\ &= \sum_{q=0}^{2n} \mu \circ \xi_q^{-1}(\cup_{q=2}^{2n} Y_q) = \sum_{q=2}^{2n} \mu \circ \xi_q^{-1}(Y_q) = (2n-1)\mu(X). \end{aligned} \tag{4.1.1}$$

On the other hand,

$$\begin{aligned} \nu(Y) &= \nu(Y_0 \cup Y_1) = S^*(\mu)(Y_0 \cup Y_1) \\ &= \sum_{q=0}^{2n} \mu \circ \xi_q^{-1}(Y_0 \cup Y_1) = \mu \circ \xi_0^{-1}(Y_0) + \mu \circ \xi_1^{-1}(Y_1) = 2\mu(X) \end{aligned} \tag{4.1.2}$$

By (4.1.1) and (4.1.2), $0 = \mu(X) = \frac{1}{2}\nu(Y) \neq 0$, which is a contradiction. Hence the set of all $\nu \in C(Y)^*$ supported in $Y_0 \cup Y_1$ with $\nu(Y) \neq 0$, which is an open set, is not included in the range of S^*. Therefore, the range of S^* is not dense and thus S is not injective. $\square$

Actually, the non-uniqueness of outer functions does not essentially depends on the disjointness of the images Y_q, as we will see in Remark 4.2.6 for the outer functions with respect to Sprecher's K-basis (1.3.2). We think that the outer functions are not unique for any K-basis, which we leave as an open problem.

The non-uniqueness of outer functions does not contradict with the definition of Kolmogorov maps, because for every continuous multivariate function, one can restrict its outer function to be chosen in a definite way. For example, in the proofs of Kolmogorov [29], Lorentz [40] and Sprecher [48], for a given continuous function f of several variables, a unique outer function g is constructed for f in a specific way described by the authors. Moreover, for the Kolmogorov maps in their proofs, we have

Proposition 4.1.5. *The Kolmogorov map with respect to a K-basis defined in the constructive proofs of KST (see section 2.2)*

$$\begin{aligned} K : C(I^n) &\longrightarrow C(\mathbb{R}) \\ f &\longrightarrow Kf \end{aligned}$$

are bounded and continuous with respect to the maximum norms. That is, there exists a constant $c > 0$ *such that* $\|Kf\|_\infty \leq c\|f\|_\infty$.

4.1.2 Non-positivity-preserving of K-maps

In an attempt to study Markov semigroups in high dimensions defined by the corresponding semigroups in one dimension through KST, we checked the positivity preserving property of K-maps. It turns out that K-maps do not preserve the positivity, the prove of which is simple but gives some insights about the distribution of the values of g.

Theorem 4.1.6 (Non-positivity-preserving of K-maps)**.** *Let* X *be a connected compact metric space and* K *be a K-map with respect to K-basis* $\{\xi_q : X \to \mathbb{R}|\ q = 0, ..., 2n\}$*, then* K *does neither preserve positivity nor strict positivity; that is,* $Kf \geq 0$ *does not follow from* $f \geq 0$*, for some* $f \in C(X)$*, and* $Kf > 0$ *does not follow from* $f > 0$*, for some* $f \in C(X)$*.*

Proof of Theorem 4.1.6. Assume that K preserves positivity, i.e. for any $f \geq 0$, we have $Kf \geq 0$. Since $f(\mathbf{x}) = \sum_{q=0}^{2n}(Kf)(\xi_q(\mathbf{x}))$,

$$(Kf)(\xi_q(\mathbf{x})) = 0 \quad \forall q, \text{ whenever } f(\mathbf{x}) = 0. \tag{4.1.3}$$

For $q = 0, ..., 2n$, choose any continuous curve C_q connecting the maximum and minimum points of ξ_q on X, then define

$$f_0(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in \cup_{q=0}^{2n} C_q, \\ \text{distance}(\mathbf{x}, \cup_{q=0}^{2n} C_q), & \text{otherwise.} \end{cases}$$

Then $f_0 \in C(I^n)$. Thus by (4.1.3)

$$(Kf_0)(\xi_q(\mathbf{x})) = 0, \text{ for } \mathbf{x} \in \cup_{q=0}^{2n} C_q.$$

By continuity of ξ_q, it maps connected set X and C_q to closed intervals $\xi_q[X]$ and $\xi_q[C_q]$ respectively. Since $\xi_q[C_q]$ contains the maximal and minimal points of $\xi_q[X]$ and $\xi_q[C_q] \subseteq$

$\xi_q[X]$, $\xi_q[X] = \xi_q[C_q]$ for all q. That is

$$(Kf_0)(y) \equiv 0, \quad \forall y \in \cup_{q=0}^{2n} \xi_q[I^n],$$

which implies $f_0(\mathbf{x}) := \sum_{q=0}^{2n}((Kf_0)(\xi_q(\mathbf{x})) \equiv 0$. This is a contradiction. Therefore, K-map does not preserve positivity.

Again assume K is strictly positivity preserving. For any $0 \leq f \in C(X)$, there exists a sequence of $f_n > 0$ converging to f as $n \to \infty$ in the uniform norm. By assumption, $Kf_n > 0$. Using the continuity of K, we have

$$Kf = \lim_{n\to\infty} Kf_n \geq 0.$$

This means K is positivity preserving, which is a contradiction. Therefore K does not preserve strict positivity . □

4.2 Moduli of continuity of outer functions

Next we investigate the analytical properties of the outer functions, such as differentiability and moduli of continuity. Superposition of Sprecher's K-basis with a differentiable outer function g results in a singular (with partial derivatives 0 almost everywhere) or constant multivariate function. Therefore, it is reasonable to study the modulus of continuity of outer functions instead of their differentiability.

4.2.1 Analytic properties of Kf

First we mention the Sprecher's version of KST:

Theorem 4.2.1 (Sprecher-Köppen-Braun's constructive version of KST [53][30][7])**.** *Let $n \geq 2, m \geq 2n$ and $\gamma \geq m+2$ be given integers and let $\mathbf{x} = (x_1, ..., x_n) \in I^n$. There exists continuous and monotonously increasing function $\psi : \mathbb{R} \to \mathbb{R}$ such that for any arbitrary* *continuous function $f : I^n \to \mathbb{R}$,*

$$f(\mathbf{x}) = \sum_{q=0}^{m} g_q \circ \xi_q(\mathbf{x}), \quad \textit{with} \quad \xi_q(\mathbf{x}) = \sum_{p=1}^{n} \lambda_q \psi(x_p + qa),$$

hold for some continuous functions $g_q : \mathbb{R} \to \mathbb{R}$, where $a := [\gamma(\gamma-1)]^{-1}$, $\lambda_1 = 1, \lambda_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\beta(r)}$ for $p > 1$ and $\beta(r) = (n^r - 1)/(n-1)$.

In order to substitute g_q with one g, we modify ξ_q to be

$$\xi_q(\mathbf{x}) = \sum_{p=1}^{n} \lambda_q \psi(x_p + qa) + bq, \tag{4.2.1}$$

where $b > \sum_{p=1}^{n} \lambda_p$ is a constant. Thus the images of I^n under ξ_q's are pairwise disjoint and g can be defined separately on each $Y_q := \xi_q[I^n]$. In [6], ξ_q is modified to $\xi_q(\mathbf{x}) = \sum_{p=1}^{n} \alpha_q \psi(x_p + qa) + \Delta_q$, although Δ_q defined there is actually not a constant as the author stated and it tended to 0 as the iterative step r goes to infinity (see Section 2.2).

$\psi(x)$ constructed by Sprecher [53] in 1997 is neither continuous nor monotonously increasing. This was noticed by Köppen [30] in 2002 and corrected without proof. Braun [7] in 2009 proved that the modified $\psi(x)$ by Köppen is indeed continuous and strictly monotonously increasing on I. Moreover, the corrected ψ satisfies a Hölder condition with exponent $\log_\gamma 2$ by Theorem 1 in [48]. With a minor modification of Theorem 1 in [49], one can show that $\psi'(x) = 0$ for all $x \in I$ excluding a set with 0 Lebesgue measure. Therefore, we have

Theorem 4.2.2. *Let $\xi := \{\xi_0, ..., \xi_m\}$ be Sprecher's K-basis in (4.2.1), and $Y := \cup_{q=0}^{m} Y_q := \cup_{q=0}^{m} \xi_q[I^n]$. Under the superposition operator S with respect to ξ, we have:*

(i) If $g \in C(Y)$ is Hölder continuous with exponent α, $0 < \alpha \leq 1$, then Sg is Hölder continuous with exponent $\alpha \log_\gamma 2$.

(ii) If g is almost everywhere differentiable with bounded derivatives, then Sg is almost everywhere differentiable with partial derivatives 0.

In particular, for all $p \geq 1$ and $g \in C^p(Y)$, Sg is almost everywhere differentiable with all partial derivatives 0.

Proof of Theorem 4.2.2. Given Sprecher's K-basis $\xi_q(\mathbf{x}) = \sum_{p=1}^{n} \lambda_q \psi(x_p + qa) + bq$, $q = 0, ..., m$.

1. Assume g is Hölder continuous of exponent α, $0 < \alpha < 1$. Since ξ_q is Hölder continuous of exponent $\log_\gamma 2$ for each q, $g \circ \xi_q$ is Hölder continuous of exponent $\alpha \log_\gamma 2$. Hence $S_g := \sum_{q=0}^{m} g \circ \xi_q$ is Hölder continuous of exponent $\alpha \log_\gamma 2$.

2. Assume g is almost everywhere differentiable with bounded derivative, then by chain rule and the singularity of ψ,

$$\frac{\partial g \circ \xi_q}{\partial x_p} = \frac{\partial g \circ \xi_q}{\partial \xi_q} \cdot \frac{\lambda_p d\psi(x_p + aq)}{dx_p} = \frac{\partial g \circ \xi_q}{\partial \xi_q} \cdot 0 = 0$$

almost everywhere and hence also $\partial(Sg)/\partial x_p = 0$ almost everywhere. □

Next, we illustrate the moduli of continuity of the outer functions with respect to Sprecher's K-basis.

Theorem 4.2.3. *Let $\omega_f(\delta)$, $0 < \delta < 1$ be the modulus of continuity of $f \in C(I^n)$. Under the Kolmogorov map K with respect to Sprecher's K-basis with $\gamma > m + 2$, the modulus of continuity of $g := Kf$, $\omega_g(\delta)$, can be at most $\omega_f(\gamma(\log_{\gamma^{-1}} \delta^{n+1} + 1)^{-\rho^{-1}}))$, where $\rho := \log_\gamma n < 1$, i.e.,*

$$\omega_g(\delta) \geq \mathcal{O}(\omega_f(\gamma(\log_{\gamma^{-1}} \delta^{(n-1)} + 1)^{-\rho^{-1}})).$$

For example, if $\omega_f(\delta) = \delta$ and $\delta = \gamma^{-k}$ for $k > 2$, $\omega_g(\delta) = \omega_g(\gamma^{-k}) \geq \mathcal{O}(\gamma((n+1)k + 1)^{-\rho^{-1}})$. With Theorem 4.2.3, one can obtain the following corollaries by investigating the moduli of continuity of different function classes.

Corollary 4.2.4. *Under the K-map with respect to Sprecher's K-basis with $\gamma > m + 2$,*

(i) There exists differentiable $f \in C(I^n)$ such that Kf is not differentiable.

(ii) For all $p \geq 1$, there exists $f \in C^p(I^n)$, such that Kf is not differentiable.

(iii) For all $0 < \alpha \leq 1$, there exists $f \in C(I^n)$ Hölder continuous of exponent α such that Kf is not Hölder continuous of any exponent.

4.2.2 Proof of the modulus of continuity of Kf

To prove Theorem 4.2.3, we introduce the construction of inner function ψ in [7] [30]. Sprecher-Köppen's K-basis is first defined on a dense set $\mathcal{D} := \cup_{k\in\mathbb{N}}\mathcal{D}_k$, where

$$\mathcal{D}_k = \mathcal{D}_k(\gamma) = \left\{ d_k \in \mathbb{Q} : d_k = \sum_{r=1}^{k} i_r\gamma^{-r}, i_r \in \{0, ..., \gamma - 1\} \right\},$$

by

$$\psi_k(d_k) = \begin{cases} d_k & \text{for } k = 1, \\ \psi_{k-1}(d_k - \frac{i_k}{\gamma^k}) + \frac{i_k}{\gamma^{\beta(k)}} & \text{for } k > 1 \text{ and } i_k < \gamma - 1, \\ \frac{1}{2}\left(\psi_{k-1}(d_k - \frac{i_k}{\gamma^k}) + \psi_{k-1}(d_k + \frac{1}{\gamma^k}) + \frac{i_k}{\gamma^{\beta(k)}}\right) & \text{for } k > 1 \text{ and } i_k = \gamma - 1, \end{cases} \tag{4.2.2}$$

where $\beta(k) := \frac{n^k-1}{n-1}$. Then extend ψ continuously to all $x \in I$. Every $x \in I$ has a representation

$$x = \sum_{r=1}^{\infty} i_r\gamma^{-r} = \lim_{k\to\infty}\sum_{r=1}^{k} i_r\gamma^{-r} = \lim_{k\to\infty} d_k.$$

For such an x, define

$$\psi(x) := \lim_{k\to\infty}\psi_k(d_k) = \lim_{k\to\infty}\psi_k(\sum_{r=1}^{k} i_r\gamma^{-r}).$$

ψ thus defined is continuous and monotonously increasing on I [7]. See Figure 4.1.

Assume $\gamma > m + 2$, i.e. $\gamma - m - 2 > 0$. For all $k \in \mathbb{N}$ and $d_k \in \mathcal{D}_k$, recall in section 2.2 that the q-intervals are defined as:

$$E_k^q(d_k) := [d_k - \frac{q}{\gamma - 1}\gamma^k, d_k - \frac{q}{\gamma - 1}\gamma^k + \delta_k], \quad q = 0, ..., m.$$

Let

$$E_k(d_k) := \cap_{q=0}^{m} E_k^q(d_k) = \left[d_k, d_k + \frac{\gamma - m - 2}{\gamma - 1}\gamma^{-k}\right],$$

and

$$S_k(\mathbf{d}_k) := E_k(d_{k,1}) \times \cdots \times E_k(d_{k,n})$$

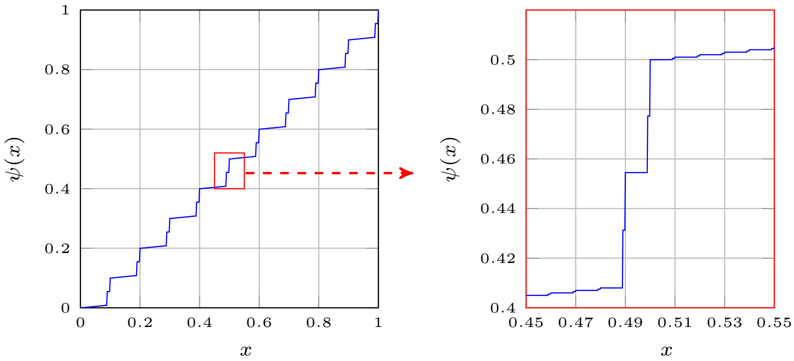

Figure 4.1: Image of ψ with $n = 2$ and $\gamma = 10$. The left graph is ψ_3 and the right is ψ_4. Similar graphs can be found in Braun [7].

with $\mathbf{d}_k := (d_{k,1} \cdots d_{k,n})$.

For any given increasing sequence of natural numbers $\{k_r\}_{r=1}^{\infty}$, define

$$F := \{x \in I | \exists \text{ sequence } \{d_{k_r}\}_{r=1}^{\infty} \text{ such that } x \in \cap_{r=1}^{\infty} E_{k_r}(d_{k_r})\},$$

and

$$F^n = \{\mathbf{x} \in I^n | \exists \text{ sequence of vectors } \{\mathbf{d}_{k_r}\}_{r=1}^{\infty} \text{ such that } \mathbf{x} \in \cap_{r=1}^{\infty} S_{k_r}(\mathbf{d}_{k_r})\}.$$

We claim the following lemma.

Lemma 4.2.5. *For any given increasing sequence $\{k_r\}_{r=1}^{\infty}$ of natural numbers, F is a Cantor set, that is, the Lebesgue measure of F, m(F), is 0 and F contains uncountably many points. Hence $F^n \subset I^n$ is also a Cantor set.*

Moreover, for any given $f \in C(I^n)$, let $\{k_r\}_{r=1}^{\infty}$ be the integers determined at iterative steps r in the construction of the outer function g for f (see Section 2.2), then

$$g \circ \xi_q(\mathbf{x}) = \frac{1}{m+1} f(\mathbf{x}),$$

for all $\mathbf{x} \in F^n$ and all $0 \leq q \leq m$.

Remark 4.2.6. *(i) Let $\{k_r\}_{r=1}^{\infty}$ be an increasing sequence of natural numbers. For any given $x \in F$, suppose $x \in \cap_{r=1}^{\infty} E_{k_r}(d_{k_r})$ for some $\{d_{k_r}\}_{r=1}^{\infty}$. Then $\{E_{k_r}(d_{k_r})\}_{r=1}^{\infty}$*

forms a sequence of nested intervals with diminishing diameter and thus

$$x = \cap_{r=1}^{\infty} E_{k_r}(d_{k_r}).$$

Moreover, from the definition of F, d_{k_r}, $d_{k_r} + \frac{\gamma-m-2}{\gamma-1}\gamma^{-k_r} \in F$ *for all* r. *Actually, For any* $k \in \mathbb{N}$ *and any* $d_k \in \mathcal{D}_k$, *let* $k_r = k + r - 1$ *and* $d_{k_r} = d_k$ *for* $r \in \mathbb{N}$, *then* $d_k \in \cap_{r=1}^{\infty} E_{k_r}(d_{k_r})$ *and thus* $d_k \in F$. *For any* $k \in \mathbb{N}$, *and* $d_k + \frac{\gamma-m-2}{\gamma-1}\gamma^{-k}$ *with* $d_k \in \mathcal{D}_k$, *let* $k_r = k + r - 1$, $d_{k_1} = d_k$ *and* $d_{k_r} = d_k + (\gamma - m - 2)\sum_{l=1}^{r-1}\gamma^{-(k+l)}$ *for* $r \geq 2$. *Then* $d_k + \frac{\gamma-m-2}{\gamma-1}\gamma^{-k} \in \cap_{r=1}^{\infty} E_{k_r}(d_{k_r})$ *and thus* $d_k + \frac{\gamma-m-2}{\gamma-1}\gamma^{-k} \in F$.

(ii) From the construction process in [7] and [53], the outer function g *for* $f \in C(I^n)$ *depends on the choice of* $\{k_r\}_{r=1}^{\infty}$.

For any given $f \in C(I^n)$, *observe that at step* $r = 1$, *for all* $\mathbf{d}_{k_1} \in (\mathcal{D}_{k_1})^n$ *and all* q,

$$g_1 \circ \xi_q(\mathbf{d}_{k_1}) = \frac{1}{m+1} f(\mathbf{d}_{k_1}) \quad \textit{and} \quad f(\mathbf{d}_{k_1}) - \sum_{q=0}^{m} g_1 \circ \xi_q(\mathbf{d}_{k_1}) = 0.$$

Then the residue at $\mathbf{d}_{k_1}$, $f_2(\mathbf{d}_{k_1})$, *equals* 0 *and so* $g_2 \circ \xi_q(\mathbf{d}_{k_1}) = \frac{1}{m+1} f_2(\mathbf{d}_{k_1}) = 0$. *Inductively,* $g_r(\mathbf{d}_{k_1}) = 0$ *for all* $r > 1$. *Hence*

$$g \circ \xi_q(\mathbf{d}_{k_1}) := \sum_{r=1}^{\infty} g_r(\mathbf{d}_{k_1}) = g_1(\mathbf{d}_{k_1}) = \frac{1}{m+1} f(\mathbf{d}_{k_1}).$$

Now assume that g *does not depend on the choice of* k_1, *then*

$$g \circ \xi_q(\mathbf{x}) = \frac{1}{m+1} f(\mathbf{x}) \tag{4.2.3}$$

for any $\mathbf{x} \in \left(\cup_{k=k_1}^{\infty} \mathcal{D}_k\right)^n$, *which is a dense set of* I^n. *Since* f *and* $g \circ \xi_q$ *is continuous, (4.2.3) holds for all* $\mathbf{x} \in I^n$, *which is not true. For example, there exists* $\mathbf{x}_1 \neq \mathbf{x}_2$ *but* $\xi_q(\mathbf{x}_1) = \xi_q(\mathbf{x}_2)$ *for some* q. *Choose a* $f \in C(I^n)$ *with* $f(\mathbf{x}_1) = f(\mathbf{x}_2)$. *If (4.2.3) holds for all* $\mathbf{x} \in I^n$, *then*

$$\frac{1}{m+1} f(\mathbf{x}_1) = g \circ \xi_q(\mathbf{x}_1) = g \circ \xi_q(\mathbf{x}_2) = \frac{1}{m+1} f(\mathbf{x}_2),$$

which is a contradiction. Therefore, Kf depends on the choice of k_1 at least, for every f except constant function.

For a given $f \in C(I^n)$, there are different outer functions g depending on choices of k_r. Thereom 4.2.3 shows that all these outer functions g constructed with different k_r lose their modulus of continuity from f drastically.

Proof of Theorem 4.2.3. We investigate the modulus of continuity of $g(u)$ when $u \in \cup_{q=0}^{m} \xi_q[F^n]$.

For any $\mathbf{x} \in F^n$, there exists a sequence of numbers $\{\mathbf{d}_{k_r}\}_{r=1}^{\infty}$ and the corresponding nested closed cubes $\{S_{k_r}(\mathbf{d}_{k_r})\}_{r=1}^{\infty}$ such that

$$\mathbf{x} = \cap_{r=1}^{\infty} S_{k_r}(\mathbf{d}_{k_r}).$$

Then these $\mathbf{d}_{k_r}, \mathbf{d}_{k_r} + \frac{\gamma-m-2}{\gamma-1}\gamma^{-k_r}\mathbf{v} \in F^n$, where $\mathbf{v} := (1, ..., 1)$ is an n dimensional vector.

Write $\mathbf{x} := (x_1, ..., x_n)$ and $\mathbf{x}_r := (x_{r,1}, ..., x_{r,n}) \in I^n$, $r \in \mathbb{N}$.

Case 1. If $x_p \notin \cup_{k=1}^{\infty} \mathcal{D}_k$ for all $1 \le p \le n$, let $\mathbf{y}_r = \mathbf{d}_{k_r}$. There exists a subsequence of $\mathbf{y}_r$, $\mathbf{x}_r := \mathbf{y}_{i_r}$ such that

$$x_1 - x_{r,1} := x_1 - y_{i_r,1} := x_1 - d_{k_{i_r},1} \ge \gamma^{-(k_{i_r}+1)}.$$

Notice that

$$0 < x_p - x_{r,p} \le \frac{\gamma - m - 2}{\gamma - 1}\gamma^{-k_{i_r}},\ 1 \le p \le n.$$

Then

$$\gamma^{-(k_{i_r}+1)} \le |\mathbf{x} - \mathbf{x}_r| \le \sqrt{n}\frac{\gamma - m - 2}{\gamma - 1}\gamma^{-k_{i_r}},$$

and by the definition of $\xi_q(\mathbf{x}) := \sum_{p=1}^{n} \lambda_p \psi(x_p + aq) + bq$,

$$\lambda_1 \gamma^{-\beta(k_{i_r}+1)} \le \xi_q(\mathbf{x}) - \xi_q(\mathbf{x}_r) \le \left(\sum_{p=1}^{n} \lambda_p^2\right)^{1/2} (\gamma - m - 2) \left(\sum_{l=k_{i_r}+1}^{\infty} \gamma^{-\beta(l)}\right). \quad (4.2.4)$$

Case 2. Else if there exists some $P \subseteq \{1, ..., n\}$ and some $k \in \mathbb{N}$ such that $x_p \in \mathcal{D}_k$ for all $p \in P$. For these $p \in P$, let $x_{r,p} = d_{k_r} + \frac{\gamma-m-2}{\gamma-1}\gamma^{-k_r}$. For $p \notin P$, let $x_{r,p} = x_p$. Define

$\mathbf{x}_r := (x_{r,1}, ..., x_{r,n})$, then $\mathbf{x}_r \in F^n$. For all r with $k_r \geq k$,

$$|\mathbf{x} - \mathbf{x}_r| = \sqrt{|P|}\frac{\gamma - m - 2}{\gamma - 1}\gamma^{-k_r},$$

where $|P|$ denotes the cardinality of set P and thus

$$\xi_q(\mathbf{x}_r) - \xi_q(\mathbf{x}) = \left(\sum_{p\in P}\lambda_p^2\right)^{1/2}(\gamma - m - 2)\left(\sum_{l=k_r+1}^{\infty}\gamma^{-\beta(l)}\right). \tag{4.2.5}$$

In both cases, it holds that

$$|\mathbf{x} - \mathbf{x}_r| = \mathcal{O}(\gamma^{-j_r}) \quad \text{and} \quad |\xi_q(\mathbf{x}) - \xi_q(\mathbf{x}_r)| = \mathcal{O}(\gamma^{-\beta(j_r+1)}),$$

for some some increasing sequence of natural numbers $\{j_r\}_{r\in\mathbb{N}}$.

By Lemma 4.2.5, since $\mathbf{x}, \mathbf{x}_r \in F^n$,

$$|g \circ \xi_q(\mathbf{x}_r) - g \circ \xi_q(\mathbf{x})| = \frac{1}{m+1}|f(\mathbf{x}_r) - f(\mathbf{x})|. \tag{4.2.6}$$

Therefore, if $|f(\mathbf{x}_r) - f(\mathbf{x})| = \mathcal{O}(\omega_f(|\mathbf{x}_r - \mathbf{x}|)) = \mathcal{O}(\omega_f(\gamma^{-j_r}))$, then

$$\omega_g(\gamma^{-\beta(j_r+1)}) \geq \mathcal{O}(|g \circ \xi_q(\mathbf{x}_r) - g \circ \xi_q(\mathbf{x})|) = \mathcal{O}(\omega_f(\gamma^{-j_r})).$$

Note that $\beta(k) := (n^k - 1)/(n-1)$. When $\delta = \gamma^{-\beta(k+1)}$, $\gamma^{-k} = \gamma(\log_{\gamma^{-1}}\delta^{(n-1)} + 1)^{-\rho^{-1}})$, where $\rho := \log_\gamma n$. Hence,

$$\omega_g(\delta) \geq \mathcal{O}(\omega_f(\gamma(\log_{\gamma^{-1}}\delta^{(n-1)} + 1)^{-\rho^{-1}})).$$

□

Proof of Lemma 4.2.5. For any given increasing sequence $\{k_r\}_{r=1}^{\infty}$ of natural numbers, we show $m(F) = 0$ by computing the length of all gaps contained in $\cap_{l=1}^{r-1}E_{k_l}(d_{k_l})$ for some $\{d_{k_l}\}_{l=1}^{r-1}$ at each step r, the sum of which amounts to 1.

For $r \in \mathbb{N}$, denote by a_r the total length of intervals E_{k_r} such that $E_{k_r} \subseteq \cap_{l=1}^{r-1}E_{k_l}(d_{k_l})$ for some $\{d_{k_l}\}_{l=1}^{r-1}$. Denote by b_r the total length of gaps between E_{k_r} and contained in

$\cap_{l=1}^{r-1} E_{k_l}(d_{k_l})$. The number of E_{k_r} contained in each $E_{k_{r-1}}$ depends only on $k_{r+1}-k_r$. We denote it by $N_{k_{r+1}-k_r}$ and

$$N_{k_{r+1}-k_r} = \frac{\gamma-m-2}{\gamma-1}(\gamma^{k_r-k_{r-1}}-1)+1, \quad r \geq 2.$$

r=1, the length of $E_{k_1}(d_{k_1})$ is $\frac{\gamma-m-2}{\gamma-1}\gamma^{-k_1}$ and the number of E_{k_1} intervals is γ^{k_1}. Then

$$a_1 = \frac{\gamma-m-2}{\gamma-1}\gamma^{-k_1}\cdot\gamma^{k_1} = \frac{\gamma-m-2}{\gamma-1},$$

and

$$b_1 = 1-a_1 = \frac{m+1}{\gamma-1}.$$

$r=2$, the length of E_{k_2} is $\frac{\gamma-m-2}{\gamma-1}\gamma^{-k_2}$. Hence

$$a_2 = \frac{\gamma-m-2}{\gamma-1}\gamma^{-k_2}\cdot N_{k_2-k_1}\gamma^{k_1},$$

$$b_2 = a_1-a_2 = \frac{\gamma-m-2}{\gamma-1}(1-N_{k_2-k_1}\gamma^{-(k_2-k_1)}).$$

At step $r \geq 3$, the length of E_{k_r} is $\frac{\gamma-m-2}{\gamma-1}\gamma^{-k_r}$. Then

$$\begin{aligned}
a_r &= \frac{\gamma-m-2}{\gamma-1}\gamma^{-k_r}\cdot N_{k_r-k_{r-1}}N_{k_{r-1}-k_{r-2}}\cdots N_{k_2-k_1}\gamma^{k_1},\\
b_r &= a_{r-1}-a_r > 0\\
&= \frac{\gamma-m-2}{\gamma-1}\cdot N_{k_{r-1}-k_{r-2}}N_{k_{r-2}-k_{r-3}}\cdots N_{k_2-k_1}\gamma^{-(k_{r-1}-k_1)}(1-N_{k_r-k_{r-1}}\gamma^{-(k_r-k_{r-1})}).
\end{aligned}$$

Notice that $a_r \to 0$ as $r \to \infty$, then

$$\sum_{r=1}^{\infty} b_r = b_1 + \sum_{r=2}^{\infty}(a_{r-1}-a_r) = 1.$$

That is $m(F^c) = 1$ and thus the Lebesgue measure of F is 0.

Next, we show that F is not countable by establishing a surjective map T from F onto I. From the definition of F, $x = \sum_{i=1}^{\infty} c_i\gamma^{-i} \in F$ iff for all $r \in \mathbb{N}$,

(i) $0 \leq c_{k_r+1} \leq \gamma - m - 2$.

(ii) If there exists l_r such that $k_r < l_r < k_{r+1}$ and $c_{l_r} > \gamma - m - 2$, then there exists some $k_r < r < l_r$ such that $c_r < \gamma - m - 2$.

Define

$$T : F \to I$$
$$T\left(\sum_{i=1}^{\infty} c_i \gamma^{-i}\right) = \sum_{i=1}^{k_1} c_i \gamma^{-i} + \sum_{r=1}^{\infty} \sum_{i=k_r+2}^{k_{r+1}} c_i \gamma^{-(i-r)}.$$

Since $0 \leq c_i \leq \gamma - 1$ for all i, the infinite series above convergent absolutely. In fact, express $x = 0.c_1 c_2 ... c_{k_1} ... c_{k_2} ...$ in base γ, then $T(x) = 0.c_1 c_2 ... c_{k_1} c_{k_1+2} ... c_{k_2} c_{k_2+2} ...$, obtained by removing the $k_r + 1$-th digits from x.

T defined above is a surjective but not injective. First, notice that for all $x = \sum_{i=1}^{\infty} c_i \gamma^{-i}$ with $c_{k_r+1} < \gamma - m - 2$ for all $r \in \mathbb{N}$, they belong to F and their images $T(x) = 0.c_1 c_2 ... c_{k_1} c_{k_1+2} ... c_{k_2} c_{k_2+2} ...$ cover all points in I. Thus it is surjective. Second, let $x = \sum_{i=1}^{k_1} c_i \gamma^{-i} + (\gamma - m - 2)\gamma^{-(k_r+1)} + (\gamma - m - 2)\gamma^{-(k_r+2)}$ and $x' = \sum_{i=1}^{k_1} c_i \gamma^{-i} + 0 \cdot \gamma^{-(k_r+1)} + (\gamma - m - 2)\gamma^{-(k_r+2)}$, then $x \neq x'$, $x, x' \in F$ and $T(x) = T(x') = \sum_{i=1}^{k_1} c_i \gamma^{-i} + (\gamma - m - 2)\gamma^{-(k_r+2)}$, so T is not injective.

F is a subset of I. By Schröder-Bernstein theorem (see appendix), cardinality of F equals the cardinality of I and thus F is a uncountable.

Finally, for any given $f \in C(I^n)$, let $\{k_r\}_{r=1}^{\infty}$ be the integers determined at iterative steps r in the construction of the outer function g for f (see Section 2.2).

For any $\mathbf{x} = \cap_{r=1}^{\infty} S_{k_r}(\mathbf{d}_{k_r}) \in F^n$,

$$g \circ \xi_q(\mathbf{x}) = \sum_{r=1}^{\infty} g_r \circ \xi_q(\mathbf{d}_{k_r}) = \frac{1}{m+1} \sum_{r=1}^{\infty} f_r(\mathbf{d}_{k_r})$$

holds for all $0 \leq q \leq m$.

Since $\sum_{q=0}^{m} g \circ \xi_q(\mathbf{x}) = f(\mathbf{x})$,

$$g \circ \xi_q(\mathbf{x}) = \frac{1}{m+1} f(\mathbf{x}), \quad \forall\, 0 \leq q \leq m.$$

□

Chapter 5

Kolmogorov-Fourier transform

In KST, a continuous function of several variables has different outer functions corresponding to different Kolmogorov bases used in the representations. In this chapter, we investigate the change of outer functions using Fourier transform. For any given continuous function of several variables f and two different K-bases ξ and η, let g_ξ and g_η be the corresponding outer functions of f. We obtain a formula to transform g_ξ to g_η. We also combine Kolmogorov bases in n dimension with bases of function spaces defined on 1 dimensional domain $\mathbb{R}$ and thus obtain new bases for function spaces defined on n dimensional domain I^n.

5.1 Kolmogorov-Fourier transform

5.1.1 Kolmogorov-Fourier kernel

In KST, f is represented by sums of compositions of an f-dependent outer function g and an independent K-basis ξ_q's. It would help us to understand the dependence of g on f by separating the independent K-basis from the argument of g. Fourier transform is one technique to serve the purpose.

In the following we investigate the combined operator of inverse Fourier transform and

Kolmogorov map. We denote the Fourier transform of a function g by $\mathscr{F}g$ or $\hat{g}$:

$$\mathscr{F}g(t) = \hat{g}(t) := \int_{\mathbb{R}} g(x)e^{-2\pi itx}dx$$

whenever the integral converges in Lebesgue sense.

Definition 5.1.1 (Kolmogorov-Fourier transform)**.** *For a given K-basis $\xi := \{\xi_q\}_{q=0}^{2n}$ on I^n, let $k_\xi(\mathbf{x}, t) := \frac{1}{2n+1}\sum_{q=0}^{2n} e^{2\pi i\xi_q(\mathbf{x})t}$. Define the Kolmogorov-Fourier (K-F) transform with respect to K-basis ξ,*

$$\begin{aligned} KF_\xi : L_1(\mathbb{R}) &\to C(I^n) \\ \hat{g}(t) &\to KF_\xi(\hat{g})(\mathbf{x}) := \int_{\mathbb{R}} k_\xi(\mathbf{x}, t)\hat{g}(t)dt. \end{aligned}$$

We also write $KF(\hat{g})$ when there is no confusion about the underlying K-basis.

Note that the kernel $k_\xi(\mathbf{x}, t)$ is uniformly continuous in $(\mathbf{x}, t) \in I^n \times \mathbb{R}$ and $|k_\xi(\mathbf{x}, t)| \leq 1$. We describe some properties of K-F transform.

(i) If $g \in C(Y)$ is real and $\hat{g} \in L^1(\mathbb{R})$ then $KF\hat{g}$ is a real function, since

$$\begin{aligned} \overline{KF(\hat{g})} &= \int_{\mathbb{R}} \overline{\hat{g}(t)k(\mathbf{x}, t)}dt = \int_{\mathbb{R}} \overline{\hat{g}(t)}k(\mathbf{x}, -t)dt \\ &= \int_{\mathbb{R}} \hat{g}(-t)k(\mathbf{x}, -t)dt = \int_{\mathbb{R}} \hat{g}(t)k(\mathbf{x}, t)dt = KF(\hat{g}). \end{aligned}$$

(ii) If $\hat{g} \in L^1(\mathbb{R})$, then $KF\hat{g}$ is uniformly continuous.

In fact, by the uniform continuity of $k_\xi(\mathbf{x}, t)$ with respect to $\mathbf{x}$, for any $\epsilon > 0$ and any $\mathbf{x}, \mathbf{x}' \in I^n$ within $|\mathbf{x} - \mathbf{x}'| < \delta$, there exists a $\delta > 0$ such that $|k_\xi(\mathbf{x}, t) - k_\xi(\mathbf{x}', t)| < \epsilon$. Therefore, for $|\mathbf{x} - \mathbf{x}'| < \delta$,

$$|KF(\mathbf{x}) - KF(\mathbf{x}')| = \left| \int_{\mathbb{R}} [k_\xi(\mathbf{x}, t) - k_\xi(\mathbf{x}', t)]\hat{g}(t)dt \right| < \epsilon\|\hat{g}\|_1.$$

(iii) KF is a bounded linear operator from $L^1(\mathbb{R})$ into $L^\infty(I^n)$ and thus into $L^p(I^n)$ with $1 \leq p \leq \infty$, since $|k_\xi(\mathbf{x}, t)| \leq 1$.

The classical Fourier transform can be extended to L^p, $1 \leq p \leq 2$. First we mention the Fourier transform on L^2.

Theorem 5.1.2 (Plancherel Theorem [56])**.** *The Fourier transform $\mathcal{F} : L^1(\mathbb{R}) \to C_0(\mathcal{R})$ is an unitary on $L^2(\mathbb{R})$; that is, $\mathcal{F}$ maps onto $L^2(\mathbb{R})$ and $\|\hat{g}\|^2 = \|g\|^2$. Furthermore,*

$$\hat{g}(\xi) = \lim_{R\to\infty} \int_{|x|<R} g(x) e^{-2\pi i x\cdot\xi} dx$$

and

$$g(x) = \lim_{R\to\infty} \int_{|x|<R} \hat{g}(\xi) e^{2\pi i x\cdot\xi} dx,$$

where the limits are in L^2.

Then $\mathcal{F}$ can be defined on $L^1(\mathbb{R}) + L^2(\mathbb{R}) := \{g = g_1 + g_2 : g_1 \in L^1(\mathbb{R}), g_2 \in L^2(\mathbb{R})\}$ by $\mathcal{F}(g_1 + g_2) = \mathcal{F}g_1 + \mathcal{F}g_2$. If $g_1' + g_2' = g_1 + g_2$ with $g_1' \in L^1(\mathbb{R}), g_2' \in L^2(\mathbb{R})$, then $g_1' - g_1 = g_2 - g_2' \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. The definition of $\mathcal{F}$ coincides with that on $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and thus $\hat{g}_1' - \hat{g}_1 = \hat{g}_2 - \hat{g}_2'$. That is, $\hat{g}_1 + \hat{g}_2 = \hat{g}_1' + \hat{g}_2'$. Hence, $\mathcal{F}$ is well-defined on $L^1(\mathbb{R}) + L^2(\mathbb{R})$. Since $L^p(\mathbb{R}) \subseteq L^1(\mathbb{R}) + L^2(\mathbb{R})$, $1 \leq p \leq 2$, $\mathcal{F}$ is well defined for all $g \in L^p(\mathbb{R})$, $1 \leq p \leq 2$.

Moreover, applying Riesz-Thorin interpolation theorem (see Appendix), we have the following Hausdorff-Young inequality.

Theorem 5.1.3 (Hausdorff-Young Inequality, Corollary 1.20 in [70])**.** *If $g \in L^p(\mathbb{R})$, $1 \leq p \leq 2$, then $\hat{g} \in L^{p'}(\mathbb{R})$ and*

$$\|\hat{g}\|_{p'} \leq \|g\|_p,$$

where $1/p + 1/p' = 1$.

Furthermore, $\mathcal{F}$ can be defined as a linear functional acting on tempered distributions.

The extension of classical Fourier transform from $L^1(\mathbb{R})$ to $L^2(\mathbb{R})$ depends on its nice properties on convolution and multiplication of functions as a consequence of the simplicity of the exponential kernel e^{-ixt}. If $g_1, g_2 \in L^1(\mathbb{R})$, then $\mathcal{F}(f_1 * f_2) = (\mathcal{F}f_1)(\mathcal{F}f_2)$ (see Theorem 1.4 in [56]). However, in the case of K-F transform, the kernel $k(\mathbf{x}, t) := \frac{1}{2n+1} \sum_{q=0}^{2n} e^{2\pi i \xi_q(\mathbf{x})t}$. $\xi_q(\mathbf{x})$ is not linear in $\mathbf{x}$, and the sum over q in $k(\mathbf{x}, t)$ also makes the relation between convolution and multiplication more complicated. For example, $KF(\hat{g}_1 * \hat{g}_2) = Kf(\hat{g}_1)KF(\hat{g}_2)$ no longer holds in K-F transform.

We introduce a new norm on $L^1(\mathbb{R})$ for K-F transform. Let $\xi := \{\xi_q\}_{q=0}^{2n}$ be a K-basis on I^n and

$$A_\xi(t,s) := \int_{I^n} k_\xi(\mathbf{x},t)\overline{k_\xi(\mathbf{x},s)}d\mathbf{x} := \frac{1}{(2n+1)^2}\sum_{q,q'=0}^{2n}\int_{I^n} e^{2\pi i(\xi_q(\mathbf{x})t-\xi_{q'}(\mathbf{x})s)}d\mathbf{x}. \quad (5.1.1)$$

Proposition 5.1.4. *For any $\hat{g}_1, \hat{g}_2 \in L^1(\mathbb{R})$, let $A_\xi(t,s)$ be as in (5.1.1), define*

$$< \hat{g}_1, \hat{g}_2 >_\xi := \int_{\mathbb{R}^2} \hat{g}_1(t)A_\xi(t,s)\overline{\hat{g}_2(s)}dtds.$$

The sesquilinear operator $< \cdot, \cdot >_\xi$ defines a semi norm on $L^1(\mathbb{R})$ by

$$\|\hat{g}\|_\xi := < \hat{g}, \hat{g} >_\xi^{1/2} .$$

Proof of Proposition 5.1.4. By Fubini's theorem (see Appendix), since for any $\hat{g}_1, \hat{g}_2 \in L^1(\mathbb{R})$,

$$\int_{\mathbb{R}^2}\int_{I^n} |\hat{g}_1(t)k_\xi(\mathbf{x},t)\overline{\hat{g}_2(s)k_\xi(\mathbf{x},s)}|d\mathbf{x}dtds \le \|\hat{g}_1\|_1\|\hat{g}_2\|_1,$$

we have

$$\begin{aligned}
& \int_{I^n} KF\hat{g}_1(\mathbf{x})\overline{KF\hat{g}_2(\mathbf{x})}d\mathbf{x} \\
:= & \int_{I^n}\left(\int_{\mathbb{R}} k_\xi(\mathbf{x},t)\hat{g}_1(t)dt\right)\left(\overline{\int_{\mathbb{R}} k_\xi(\mathbf{x},s)\hat{g}_2(s)ds}\right)d\mathbf{x} \\
= & \int_{\mathbb{R}^2}\hat{g}_1(t)\left(\int_{I^n} k_\xi(\mathbf{x},t)\overline{k_\xi(\mathbf{x},s)}d\mathbf{x}\right)\overline{\hat{g}_2(s)}dtds \\
= & < \hat{g}_1, \hat{g}_2 >_\xi .
\end{aligned}$$

Thus $\|\hat{g}\|_\xi = \|KF\hat{g}\|_2 \ge 0$.

By Hölder inequality (see Appendix),

$$\begin{aligned}
< \hat{g}_1, \hat{g}_2 >_\xi &= \int_{I^n} KF\hat{g}_1(\mathbf{x})\overline{KF\hat{g}_2(\mathbf{x})}d\mathbf{x} \\
&\le \|KF\hat{g}_1\|_2\|KF\hat{g}_2\|_2 = \|\hat{g}_1\|_\xi\|\hat{g}_2\|_\xi.
\end{aligned}$$

Then

$$\begin{aligned} & < \hat{g}_1 + \hat{g}_2, \hat{g}_1 + \hat{g}_2 >_\xi \\ = & < \hat{g}_1, \hat{g}_1 >_\xi + < \hat{g}_2, \hat{g}_2 >_\xi + < \hat{g}_1, \hat{g}_2 >_\xi + < \hat{g}_2, \hat{g}_1 >_\xi . \\ \leq & \|\hat{g}_1\|_\xi^2 + 2\|\hat{g}_1\|_\xi \|\hat{g}_2\|_\xi + \|\hat{g}_2\|_\xi^2. \end{aligned}$$

Therefore

$$\|\hat{g}_1 + \hat{g}_2\|_\xi \leq \|\hat{g}_1\|_\xi + \|\hat{g}_2\|_\xi. \tag{5.1.2}$$

Finally, for any complex number a, $\|a\hat{g}\|_\xi = \|\widehat{(ag)}\|_\xi = |a| \cdot \|\hat{g}\|_\xi$.

Therefore, $\| \cdot \|_\xi$ is a semi norm on $L^1(\mathbb{R})$. □

5.1.2 Change of K-bases

Given two different K-bases $\xi := \{\xi_q\}_{q=0}^{2n}$ and $\eta := \{\eta_q\}_{q=0}^{2n}$ on I^n, let $Y := \cup_{q=0}^{2n}(\xi_q[I^n] \cup \eta_q[I^n])$. Let $k_\xi(\mathbf{x}, t)$ and $k_\eta(\mathbf{x}, t)$ be the kernels of K-F transform corresponding to ξ and η respectively. By KST, for $k_\xi(\mathbf{x}, t)$, let $b_t = K_\eta(k_\xi(\cdot, t))$, then $b_t \in C(Y)$ and

$$k_\xi(\mathbf{x}, t) = \frac{1}{2n+1} \sum_{q=0}^{2n} b_t(\eta_q(\mathbf{x})), \quad \forall \mathbf{x} \in I^n. \tag{5.1.3}$$

$b_t(u)$ is defined on $u \in Y$. We make an extension of $b_t(u)$, still denoted by $b_t(u)$, such that $b_t(u)$ is supported on a bounded open interval $\tilde{Y} \supset Y$ and $b_t(u) \in C(\mathbb{R})$. There are many such extensions and suppose that we can pick one of them such that $\hat{b}_t(s) := \int_{\tilde{Y}} b_t(u) e^{-2\pi i s u} du \in L^1(\mathbb{R})$. Since both $b_t, \hat{b}_t \in L^1(\mathbb{R})$, by Corollary 1.21 in [56] (see Appendix), we have

$$b_t(u) = \int_{\mathbb{R}} \hat{b}_t(s) e^{2\pi i u s} ds$$

for almost all $u \in \mathbb{R}$ in Lebesgue sense. Hence,

$$k_\xi(\mathbf{x}, t) = \int_{\mathbb{R}} \hat{b}_t(s) k_\eta(\mathbf{x}, s) ds, \tag{5.1.4}$$

for almost all $\mathbf{x} \in I^n$ in Lebesgue sense and thus all $\mathbf{x} \in (0,1)^n$ since both sides of (5.1.4) are continuous functions on $(0,1)^n$.

Because $|k_\xi(\mathbf{x}, t)| \leq 1$, by Proposition 4.1.5, $|b_t| \leq c\|k_\xi(\mathbf{x}, t)\|_\infty \leq c$ for some constant $c > 0$. Since $\hat{b}_t(s) := \int_{\tilde{Y}} b_t(r)e^{-2\pi isr}dr$ with $\tilde{Y}$ as a bounded open interval, $\hat{b}_t(s)$ is continuous and bounded in both t and s. Then we can define the outer function transform operator, which transforms the outer function g_ξ under K-basis ξ to the outer function g_η under η for a given multivariate function f:

$$\begin{aligned} B : L^1(\mathbb{R}) &\to L^\infty(\mathbb{R}) \\ \hat{g}(t) \to B\hat{g}(s) &:= \int_{\mathbb{R}} \hat{b}_t(s)\hat{g}(t)dt. \end{aligned}$$

B is a bounded operator by the boundedness of $\hat{b}_t(s)$ with respect to t.

Theorem 5.1.5. *Let $\xi := \{\xi_q\}_{q=0}^{2n}$ and $\eta := \{\eta_q\}_{q=0}^{2n}$ be K-bases on I^n such that ξ_q, η_q are strictly increasing with respect to x_p, $p = 1, ..., n$. Let $Y := \cup_{q=0}^{2n}(\xi_q[I^n] \cup \eta_q[I^n])$ and $b_t(s)$ be as in (5.1.3). For $f \in C(I^n)$, there exists $g_\xi \in C(Y)$ such that $f = S_\xi g_\xi$. Suppose that $\hat{g}_\xi \in L^1(\mathbb{R})$ and $\hat{b}_t(s) \in L^1(\mathbb{R}, ds)$ with $\|\hat{b}_t\|_1$ uniformly bounded in $t \in \mathbb{R}$. Then*

$$f(\mathbf{x}) = KF_\xi(\hat{g}_\xi)(\mathbf{x}) = KF_\eta(B\hat{g}_\xi)(\mathbf{x}), \quad \forall \mathbf{x} \in (0,1)^n.$$

Proof of Theorem 5.1.5. Since $\|\hat{b}_t\|_1$ uniformly bounded in $t \in \mathbb{R}$, there exists a $c > 0$ such that

$$\int_{\mathbb{R}} |\hat{b}_t(s)|ds \leq c.$$

Hence

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |\hat{g}_\xi(t)\hat{b}_t(s)k_\eta(\mathbf{x}, s)|ds \right) dt \leq c \int_{\mathbb{R}} |\hat{g}_\xi(t)|dt,$$

By (5.1.4) and then Fubini-Tonelli theorem (see Appendix), we have:

$$\begin{aligned}
KF(\hat{g}_\xi)(\mathbf{x}) &:= \int_{\mathbb{R}} \hat{g}_\xi(t) k_\xi(\mathbf{x}, t) dt \\
&= \int_{\mathbb{R}} \hat{g}_\xi(t) \left(\int_{\mathbb{R}} \hat{b}_t(s) k_\eta(\mathbf{x}, s) ds \right) dt \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \hat{b}_t(s) \hat{g}_\xi(t) dt \right) k_\eta(\mathbf{x}, s) ds \\
&=: \int_{\mathbb{R}} B(\hat{g}_\xi)(s) k_\eta(\mathbf{x}, s) ds \\
&=: KF_\eta(B\hat{g}_\xi)(\mathbf{x})
\end{aligned}$$

By Corollary 1.21 in [56], for almost all $u_q \in Y_q$, $q = 0, ..., 2n$,

$$g_\xi(u_q) = \int_{\mathbb{R}} \hat{g}_\xi(t) e^{2\pi i u_q t} dt. \tag{5.1.5}$$

Notice that the left hand side of (5.1.5), $g(u_q)$, is continuous in the interior of Y_q, and the right hand side is continuous on $\mathbb{R}$. Therefore (5.1.5) holds for all u_q in the interior of Y_q. For $u_q = \xi_q(\mathbf{x})$, since ξ_q is strictly increasing in x_p, $p = 1, ..., n$, $\xi_q[(0,1)^n]$ is included in the interior of Y_q for $q = 0, ..., 2n$. Thus,

$$\frac{1}{2n+1} \left(\sum_{q=0}^{2n} g_\xi(\xi_q(\mathbf{x})) \right) = \int_{\mathbb{R}} \hat{g}_\xi(t) \frac{1}{2n+1} \left(\sum_{q=0}^{2n} e^{2\pi i \xi_q(\mathbf{x}) t} \right) dt$$

holds for all $\mathbf{x} \in (0,1)^n$. That is,

$$f(\mathbf{x}) = KF_\xi(\hat{g}_\xi)(\mathbf{x}) = KF_\eta(B\hat{g}_\xi)(\mathbf{x}), \quad \forall \mathbf{x} \in (0,1)^n.$$

$\square$

5.2 KST combined with Fourier basis and wavelet basis

5.2.1 Kolmogorov-Fourier basis

Let $\{\xi_q\}_{q=0}^{2n}$ be a K-basis on I^n such that $\cup_{q=0}^{2n}\xi_q[I^n] \subseteq I$. For any $f \in C(I^n)$, the domain of its outer functions g will be the unit interval I. Extend g periodically to $\mathbb{R}$ by

$$g(u) := g(u - m), \text{ for } m \leq u < m + 1, \; m \in \mathbb{Z}.$$

Thus we can consider the Fourier series of $g(u)$:

$$\sum_{m\in\mathbb{Z}} c_m(g)e^{2\pi imu}, \quad \text{with} \quad c_m(g) := \int_0^1 g(u)e^{-2\pi imu}du.$$

Definition 5.2.1 (Kolmogorov-Fourier series)**.** *For $g \in C(I)$, its Kolmogorov-Fourier series (K-F series) with respect to a K-basis $\{\xi_q\}_{q=0}^{2n}$ is defined as*

$$\frac{1}{2n+1}\left(\sum_{m\in\mathbb{Z}} c_m(g)\sum_{q=0}^{2n} e^{2\pi i\xi_q(\mathbf{x})m}\right).$$

Let $S_\xi(g) := \frac{1}{2n+1}\left(\sum_{q=0}^{2n} g \circ \xi_q\right)$, then by the convergence of Fourier series of g, we have the convergence of K-F series of g.

Proposition 5.2.2. *Let $g \in C(I)$.*

(i) Dini test: By Theorem 6.8 [71], if g is Holder continuous with exponent $0 < \alpha \leq 1$ on I, then its K-F series converges uniformly to $S_\xi(g)$ on I^n.

(ii) Dirichlet-Jordan test: By Theorem 8.6 [71], if g is of bounded variation on I, then its K-F series converges everywhere to $S_\xi(g)$ on I^n.

5.2.2 KST with wavelet basis

In [34], Leni, Dougerolle and Truchetet design an image compression scheme using KST and wavelet decomposition. Consider a grey-scale image as a function of two variables:

the horizontal and vertical coordinates of a pixel. The value of the function is the grey level of each pixel, which changes between white, $f = 0$, and black, $f = 1$. They first decompose the image into four sub-images using wavelet transform. One sub-image contains low frequencies and the other three contain high frequencies of the original image. Then each sub-image is represented by a univariate function through KST. Ingelnik's [22] spline network approximation scheme is used to implement Kolmogorov's representation. Sub-images of high frequencies are represented with relative high accuracy by using less pixels of the sub-images, while the sub-image of low frequencies are represented with the highest accuracy by using more pixels of the original. In this way, they compress the original image.

In this subsection, we combine wavelet basis in 1 dimension with Kolmogorov basis in n dimension to develop a new basis in n dimension. In particular, we take Haar wavelets as an example.

Let $a_0 > 1$, $b_0 > 0$ and $\psi \in L^2(\mathbb{R})$. Suppose that

$$\psi_{n,m}(x) = |a_0|^{-m/2}\psi(a_0^{-m}x - nb_0) \quad (n,m) \in \mathbb{Z} \times \mathbb{Z},$$

is an orthonormal wavelet basis for $L^2(\mathbb{R})$. For a fixed g, the discrete wavelet coefficients of g are given by

$$< g, \psi_{n,m} >= |a_0|^{-m/2} \int g(x)\overline{\psi(a_0^{-m}x - nb_0)}dx$$

and then

$$g(x) = \sum_{(n,m)\in\mathbb{Z}\times\mathbb{Z}} < g, \psi_{n,m} > \psi_{n,m}(x) \tag{5.2.1}$$

in L^2 sense.

Given a K-basis $\xi := \{\xi_q\}_{q=0}^{2n}$ on I^n. Let $Y := \cup_{q=0}^{2n}\xi_q[I^n]$. By KST, for any $f \in C(I^n)$, there exists $g \in C(Y) \subseteq L^2(\mathbb{R})$ such that

$$f = \sum_{q=0}^{2n} g \circ \xi_q.$$

Thus if $\{e_k\}_{k=0}^{\infty}$ is an orthonormal basis for $L^2(\mathbb{R})$, then $\{\sum_{q=0}^{2n} e_k \circ \xi_q\}_{k=0}^{\infty}$ is possibly a Schauder basis for proper function spaces defined on I^n. In particular, any $f \in C(I^n)$ could be approximated by finite linear combinations of $\{E_k := \sum_{q=0}^{2n} e_k \circ \xi_q\}_{k=0}^{\infty}$:

$$\sum_{k=1}^{N} < K_\xi f, e_k > E_k := \sum_{k=1}^{N} < K_\xi f, e_k > \left(\sum_{q=0}^{2n} e_k \circ \xi_q\right),$$

in proper topologies.

Example 5.2.3. *Take the Haar wavelets for example [9]. The Haar function is*

$$\psi(x) = \begin{cases} 1 & 0 \le x < \frac{1}{2} \\ -1 & \frac{1}{2} \le x < 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Taking $a_0 = 2, b_0 = 1$*, then*

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n), \quad m, n \in \mathbb{Z}$$

constitutes an orthonormal basis for $L^2(\mathbb{R})$*, which is called Haar wavelet basis.*

We claim that for any given K-basis $\{\xi_q\}_{q=0}^{2n}$ *on* $\mathbb{R}^n$ *(see Theorem 1.4.2), if for any fixed* $J_1 \in \mathbb{N}$*, the Lebesgue measure*

$$m(\{\mathbf{x} : \xi_q(\mathbf{x}) \in [-2^{J_1+K}, 2^{J_1+K})\}) < \mathcal{O}(2^{pK}), \quad q = 0, ..., 2n,$$

then $C(\mathbb{R}^n)$ *can be approximated by linear combinations of*

$$\Psi_{j,k} := \sum_{q=0}^{2n} \psi_{j,k} \circ \xi_q, \; j, k \in \mathbb{Z},$$

in L^p *sense,* $2 \le p < \infty$*.*

Since any $g \in L^p(\mathbb{R}), 2 \le p < \infty$*, can be approximated by a function with compact support which is piecewise constant on* $[l2^{-j}, (l+1)2^{-j})$*, we can restrict ourselves to consider piecewise constant functions only. Assume* g *is supported on* $[-2^{J_1}, 2^{J_1}]$ *and* *is piecewise constant on* $[l2^{-J_0}, (l+1)2^{-J_0})$*, where* J_1, J_0 *can be arbitrarily large. Using the techniques of multiresolution analysis and notations in section 1.3.3 [9], we approximate* g *with linear combinations of Haar wavelets. Denote the constant value* $g^0 = g$ *on* $[l2^{-J_0}, (l+1)2^{-J_0})$ *by* g^0_l*. Represent* g^0 *as two parts,* $g^0 = g_1 + \delta^1$*, where* g^1 *is the approximation to* g^0 *which is piecewise constant over intervals twice as large as originally; that is,* $g^1|_{[k2^{-J_0+1},(k+1)2^{-J_0+1})} \equiv$ *constant* $= g^1_k := \frac{1}{2}(g^0_{2k} + g^0_{2k+1})$*. The function* δ^1 *is a piecewise constant with the same stepwidth as* g^0 *and*

$$\delta^1_{2l} = g^0_{2l} - g^1_l = \frac{1}{2}(g^0_{2l} - g^0_{2l+1})$$

and

$$\delta^1_{2l+1} = g^0_{2l+1} - g^1_l = \frac{1}{2}(g^0_{2l+1} - g^0_{2l}) = -\delta^1_{2l.}$$

It follows that δ_1 *is a linear combination of Haar wavelets:*

$$\delta^1 = \sum_{l=-2^{J_1+J_0-1}}^{2^{J_1+J_0-1}-1} \delta^1_{2l}\psi(2^{J_0-1}x - l).$$

We have

$$g = g^0 = g_1 + \sum_{l=-2^{J_1+J_0-1}}^{2^{J_1+J_0-1}-1} c_{-J_0+1,l}\psi_{-J_0+1,l},$$

where g^1 *is of the same type of* g^0 *, but with stepwidth twice as large. We can repeat the procedure till we have* $g = g^{J_0+J_1+K} + \sum_{m=-J_0+1}^{J_1+K} \sum_l c_{m,l}\psi_{m,l}$*, where support of* $g^{J_0+J_1+K}$ *is* $[-2^{J_1+K}, 2^{J_1+K}]$*, and*

$$g^{J_0+J_1+K}|_{[0,2^{J_1+K})} = 2^{-K} g^{J_0+J_1}_0,$$
$$g^{J_0+J_1+K}|_{[-2^{J_1+K},0)} = 2^{-K} g^{J_0+J_1}_{-1},$$

and $g^{J_0+J_1}_0$ *is the average of* g *over* $[0, 2^{J_1})$ *and* $g^{J_0+J_1}_{-1}$ *is the average of* g *over* $[-2^{J_1}, 0)$*.*

Then

$$\begin{aligned}
& \int_{\mathbb{R}^n} \left| g(\xi_q(\mathbf{x})) - \sum_{m=-J_0+1}^{J_1+K} \sum_l c_{m,l}\psi_{m,l}(\xi_q(\mathbf{x})) \right|^p d\mathbf{x} && (5.2.2)\\
= & \int_{\mathbb{R}^n} \left| g^{J_0+J_1+K}(\xi_q(\mathbf{x})) \right|^p d\mathbf{x} \\
= & \; 2^{-pK} \left(\int_{\{\mathbf{x}:\xi_q(\mathbf{x})\in[0,2^{J_1+K})\}} |g_0^{J_0+J_1}|^p d\mathbf{x} + \int_{\{\mathbf{x}:\xi_q(\mathbf{x})\in[-2^{J_1+K},0)\}} |g_{-1}^{J_0+J_1}|^p \right) d\mathbf{x} \\
\leq & \; 2^{-pK} m(\{\mathbf{x} : \xi_q(\mathbf{x}) \in [-2^{J_1+K}, 2^{J_1+K})\})(|g_0^{J_0+J_1}|^p + |g_{-1}^{J_0+J_1}|^p). && (5.2.3)
\end{aligned}$$

The sum over l *in (5.2.2) depends on* m*:* $l \in \{-1, 0\}$*, if* $m \geq J_1$*, and* $l \in \{-2^{J_1-m}, ..., 2^{J_1-m} - 1\}$*, if* $m < J_1$*.*

Therefore, if the Lebesgue measure

$$m(\{\mathbf{x} : \xi_q(\mathbf{x}) \in [-2^{J_1+K}, 2^{J_1+K})\}) < \mathcal{O}(2^{pK}),$$

then (5.2.3) can be made arbitrarily small by taking sufficiently large K*. Hence* $g \circ \xi_q$ *and thus* $f = \sum_{q=0}^{2n} g \circ \xi_q$ *can be approximated in* L^p *norm to any precision by a finite combination of* $\Psi_{m,n}$*.*

In the case of $C(I^n)$*, let* $\{\xi_q\}_{q=0}^{2n}$ *be a K-basis on* I^n *and* $Y := \cup_{q=0}^{2n}\xi_q[I^n]$*, then* $C(I^n)$ *can be approximated by linear combinations of* $\Psi_{m,n} := \sum_{q=0}^{2n} \psi_{m,n} \circ \xi_q$*,* $m, n \in \mathbb{Z}$ *in* $L^p(I^n)$ *for* $2 \leq p < \infty$*, since* $m(I^n) = 1 < \mathcal{O}(2^{pK})$*.*

Chapter 6

KST and optimal transport problems

In this chapter, we introduce 1 dimensional measures induced from $n \geq 2$ dimensional measures and relate the optimal transport cost between measures in n dimension with the corresponding optimal cost between measures in one dimension.

6.1 1 dimensional measures induced from n dimensional measures

Let μ be any Borel measure defined on the Borel sets of I^n and $\{\xi_q\}_{q=0}^{2n}$ be a K-basis on I^n. For $q = 0, ..., 2n$, let

$$\nu_q := \mu \circ \xi_q^{-1}$$

be the measure defined on the algebra of subsets of $Y_q := \xi_q[I^n]$ consisting of all sets $E \subseteq Y_q$ such that $\xi_q^{-1}[E]$ is a Borel subset of I^n.

Let

$$\nu := \frac{1}{2n+1} \sum_{q=0}^{2n} \nu_q := \frac{1}{2n+1} \sum_{q=0}^{2n} \mu \circ \xi_q^{-1}. \tag{6.1.1}$$

We list some properties of ν_q and ν induced by a Borel measure μ on the Borel sets of I^n:

(i) Let $\|\cdot\|_{TV}$ denote the total variation of a measure. We have $\|\nu_q\|_{TV} := \|\mu\circ\xi_q^{-1}\|_{TV} \leq \|\mu\|_{TV}$ and thus $\|\nu\|_{TV} \leq \|\mu\|_{TV}$.

(ii) By Theorem 1.1 in [59], if $Y_q := \xi_q[I^n]$, $0 \leq q \leq 2n$, are mutally disjoint, then there is a constant $0 < c \leq 1$ depending only on the K-basis $\{\xi_q\}_{q=0}^{2n}$, such that for any

Borel measure μ, there is some $0 \leq q \leq 2n$ such that $\|\nu_q\|_{TV} \geq c\|\mu\|_{TV}$.

If g is measurable on $Y := \cup_{q=0}^{2n} Y_q$ with respect to ν, then $f := \frac{1}{2n+1}\sum_{q=0}^{2n} g \circ \xi_q$ is measurable with respect to μ and

$$\begin{aligned}\int_{I^n} f(\mathbf{x})d\mu(\mathbf{x}) &:= \frac{1}{2n+1}\int_{I^n}\sum_{q=0}^{2n} g(\xi_q(\mathbf{x}))d\mu(\mathbf{x}) \\ &= \frac{1}{2n+1}\int_{Y_q}\sum_{q=0}^{2n} g(y)d\nu_q(y) = \int_Y g(y)d\nu(y).\end{aligned}$$

For a K-basis $\xi_q(x_1, ..., x_n) := \sum_{p=1}^{n} \psi_{pq}(x_p)$ on I^n, if the n-dimensional measure

$$\mu(x_1, ..., x_n) = \mu^1(x_1)\cdots\mu^n(x_n)$$

is a product measure with Borel measures μ^p on I, $p = 1, ..., n$, then for any bounded ν_q-measurable function g,

$$\begin{aligned}\int_{Y_q} g(y)d\nu_q(y) &= \int_{I^n} g(\xi_q(x_1, ..., x_n))d\mu(x_1, ..., x_m) \\ &= \int_I\cdots\int_I g(\psi_{1q}(x_1) + \cdots + \psi_{nq}(x_n))d\mu^1(x_1)\cdots d\mu^n(x_n) \\ &= \int_{Y_q}\cdots\int_{Y_q} g(y_{1q} + \cdots + y_{nq})d(\mu^1 \circ \psi_{1q}^{-1})\cdots d(\mu^n \circ \psi_{nq}^{-1}).\end{aligned}$$

Thus, $\nu_q := \mu \circ \xi_q^{-1}$ is a convolution of measures $\mu^p \circ \psi_{pq}^{-1}$ (see page 237 in [46]). Namely, for $q = 0, ..., 2n$,

$$\nu_q = (\mu^1 \circ \psi_{1q}^{-1}) * \cdots * d(\mu^n \circ \psi_{nq}^{-1}).$$

6.2 Two variations of KST

In this section, we introduce two varied versions of KST, which we will use in the next section to compare the optimal transport cost between measures in high dimension and measures in 1 dimension.

This first version of KST enables one to represent a continuous function on I^{2n} using a K-basis on I^n.

Proposition 6.2.1. *Let $n \geq 2$ be natural number and $\{\xi_q\}_{q=0}^{2n}$ be an arbitrary K-basis on I^n, then for any $f \in C(I^{2n})$, there exists a $g \in C(\mathbb{R}^2)$ such that*

$$f(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) = \frac{1}{(2n+1)^2} \left(\sum_{q=0}^{2n} \sum_{q'=0}^{2n} g(\xi_q(x_1, ..., x_n), \xi_{q'}(x_{n+1}, ..., x_{2n})) \right) \tag{6.2.1}$$

holds for all $(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) \in I^{2n}$.

Proof of Proposition 6.2.1. For any $\mathbf{x} \in I^{2n}$, write $\mathbf{x} := (\mathbf{x}_1, \mathbf{x}_2)$ with $\mathbf{x}_1, \mathbf{x}_2 \in I^n$. For any $f(\mathbf{x}) \in C(I^{2n})$, $f(\mathbf{x}) := f(\mathbf{x}_1, \mathbf{x}_2) = f_{\mathbf{x}_2}(\mathbf{x}_1) \in C(I^n)$. Since $\{\xi_q\}_{q=0}^{2n}$ is a K-basis on $C(I^n)$, there exists $h_{\mathbf{x}_2} \in \mathbb{R}$ such that

$$f_{\mathbf{x}_2}(\mathbf{x}_1) = \frac{1}{2n+1} \left(\sum_{q=0}^{2n} h_{\mathbf{x}_2}(\xi_q(\mathbf{x}_1)) \right).$$

By Proposition 4.1.5, the Kolmogorov map with respect to $\{\xi_q\}_{q=0}^{2n}$ is strongly continuous, and thus $h_{\mathbf{x}_2}(\xi_q(\mathbf{x}_1)) = h_{\xi_q(\mathbf{x}_1)}(\mathbf{x}_2) \in C(I^n)$. Then using Kolmogorov's representation again,

$$h_{\xi_q(\mathbf{x}_1)}(\mathbf{x}_2) = \frac{1}{2n+1} \left(\sum_{q'=0}^{2n} g_{\xi_q(\mathbf{x}_1)}(\xi_{q'}(\mathbf{x}_2)) \right)$$

holds for some $g_{\xi_q(\mathbf{x}_1)} \in C(I^n)$, all $\mathbf{x}_1 \in I^n$ and $q = 0, ..., 2n$. Therefore,

$$\begin{aligned} f(\mathbf{x}) := f(\mathbf{x}_1, \mathbf{x}_2) &= \frac{1}{(2n+1)^2} \left(\sum_{q=0}^{2n} \sum_{q'=0}^{2n} g_{\xi_q(\mathbf{x}_1)}(\xi_{q'}(\mathbf{x}_2)) \right) \\ &:= \frac{1}{(2n+1)^2} \left(\sum_{q=0}^{2n} \sum_{q'=0}^{2n} g(\xi_q(\mathbf{x}_1), \xi_{q'}(\mathbf{x}_2)) \right), \end{aligned}$$

with $g \in C(\mathbb{R}^2)$. $\square$

The second version KST tries to represent a continuous function on I^{2n} by a redundant K-basis on I^n. The number of summed items in the second version is $4n + 1$, compared to $(2n + 1)^2$ in the first version.

Proposition 6.2.2. *Let $n \geq 2$ be a natural number and $D_k \subset \mathbb{N}, k \in \mathbb{N}$ be finite sets. Let*

$$\{S^k_{q,\mathbf{i}} : \mathbf{i} \in D^n_k, q = 0, ..., 4n\}_{k\in\mathbb{N}} \tag{6.2.2}$$

be a Kolmogorov cover of I^n which covers I^n at least $4n + 1 - n = 3n + 1$ times and $\{\xi_q\}^{4n}_{q=0}$ be a K-basis on I^n which separates the Kolmogorov cover (6.2.2). Then for any $f \in C(I^{2n})$, there exists a $g \in C(\mathbb{R}^2)$ such that for all $(x_1, ..., x_{2n}) \in I^{2n}$,

$$f(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) = \frac{1}{4n + 1}\left(\sum_{q=0}^{4n} g(\xi_q(x_1, ..., x_n), \xi_q(x_{n+1}, ..., x_{2n}))\right). \tag{6.2.3}$$

Remark 6.2.3. *In fact, if we take a Lorentz's K-basis $\{\eta_q\}^{4n}_{q=0}$ on I^{2n}:*

$$\eta_q(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) = \sum_{p=1}^{n} \lambda_p\phi_q(x_p) + \sum_{p=n+1}^{2n} \lambda_p\phi_q(x_p),$$

then KST, for any $f \in C(I^{2n})$, there is a $g_\eta \in C(\mathbb{R})$ such that

$$f(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) = \frac{1}{4n + 1}\left(\sum_{q=0}^{4n} g_\eta(\sum_{p=1}^{n} \lambda_p\phi_q(x_p) + \sum_{p=n+1}^{2n} \lambda_p\phi_q(x_p))\right).$$

Define $g_2 \in C(\mathbb{R}^2)$ by $g_2(y_1, y_2) := g_\eta(y_1 + y_2)$, then

$$f(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) = \frac{1}{4n + 1}\left(\sum_{q=0}^{4n} g_2(\sum_{p=1}^{n} \lambda_p\phi_q(x_p), \sum_{p=n+1}^{2n} \lambda_p\phi_q(x_p))\right). \tag{6.2.4}$$

If we take a redundant Lorentz's K-basis $\xi_q := \sum_{p=1}^{n} \lambda_p\phi_q(x_p), q = 0, ..., 4n$ on I^n, (6.2.3) becomes

$$f(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) = \frac{1}{4n + 1}\left(\sum_{q=0}^{4n} g(\sum_{p=1}^{n} \lambda_p\phi_q(x_p), \sum_{p=1}^{n} \lambda_p\phi_q(x_{p+n}))\right). \tag{6.2.5}$$

The number of summed items in (6.2.5) and (6.2.4) are both $4n + 1$, but the latter has more parameter λ_p's.

Proof of Proposition 6.2.2. For any $(x_1, ..., x_n) \in I^n$, there are at most n of q's such that $(x_1, ..., x_n)$ is not covered by these q-cubes $S^k_{q,\mathbf{i}}$. Therefore, for any $(x_1, ..., x_n, x_{n+1}, ..., x_{2n}) \in I^{2n}$, there are at most $2n$ of q's such that $(x_1, ..., x_n)$ is not covered by the q-cubes $S^k_{q,\mathbf{i}} \times S^k_{q,\mathbf{j}}$. In other words,

$$\{S^k_{q,\mathbf{i}} \times S^k_{q,\mathbf{j}} : \mathbf{i}, \mathbf{j} \in D^n_k, q = 0, ..., 4n\}_{k \in \mathbb{N}} \tag{6.2.6}$$

is a Kolmogorov cover of I^{2n} and covers I^{2n} at least $4n + 1 - 2n = 2n + 1$ times. Notice that $\{\xi_q(x_1, ..., x_n) + \xi_q(x_{n+1}, ..., x_{2n}))\}_{q=0}^{4n}$ does not necessarily separate the Kolmogorov cover (6.2.6) and thus not necessarily a K-basis on I^{2n}. However, for any fixed $k \in \mathbb{N}$, the images of the Kolmogorov cover (6.2.6) under (ξ_q, ξ_q),

$$\xi_q[S^k_{q,\mathbf{i}}] \times \xi_q[S^k_{q,\mathbf{j}}], \quad (\mathbf{i}, \mathbf{j}) \in D^{2n}_k;\ q = 0, ..., 4n,$$

are all disjoint in $\mathbb{R}^2$.

Now for any given $f \in C(I^{2n})$, we construct iteratively a $g \in C(\mathbb{R}^2)$ such that (6.2.3) holds in a similar way as Sprecher's proof of KST [48].

For any small enough $\epsilon > 0$, let $\theta > 0$ such that

$$0 < \frac{2n+1}{4n+1}\epsilon + \frac{4n}{4n+1} \leq \theta < 1.$$

Let $f_0 := f$. At step $r \in \mathbb{N}$, choose $k_r \in \mathbb{N}$ such that for any two points $\mathbf{x}, \mathbf{x}' \in I^{2n}$ contained in one cube $S^k_{q,\mathbf{i}} \times S^k_{q,\mathbf{j}}$, $(\mathbf{i}, \mathbf{j}) \in D^{2n}_{k_r}, q = 0, ..., 4n$, it holds that

$$|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \epsilon \|f_{r-1}\|_\infty. \tag{6.2.7}$$

This can be done by the uniform continuity of f_{r-1} on I^{2n}, since the diameter of $S^k_{q,\mathbf{i}} \times S^k_{q,\mathbf{j}}$ goes to 0 as k goes to infinity.

Let $f^{k_r}_{q,\mathbf{ij}}$ be the value of f at any point of $S^k_{q,\mathbf{i}} \times S^k_{q,\mathbf{j}}$. On each square $\xi_q[S^k_{q,\mathbf{i}}] \times \xi_q[S^k_{q,\mathbf{j}}]$, we take $g_r(y_1, y_2)$ constant and equal to $f^{k_r}_{q,\mathbf{ij}}$. We can extend g_r linearly into the gaps between the $\xi_q[S^k_{q,\mathbf{i}}] \times \xi_q[S^k_{q,\mathbf{j}}]$'s such that $\|g\|_\infty \leq \|f_{r-1}\|_\infty$. In this way, we obtain a continuous g_r. Write $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \in I^{2n}$ with $\mathbf{x}_1, \mathbf{x}_2 \in I^n$. Let $f_r(\mathbf{x}) := f_{r-1}(\mathbf{x}) -$

$\frac{1}{4n+1}\sum_{q=0}^{4n} g(\xi(\mathbf{x}_1), \xi(\mathbf{x}_2))$ for all $\mathbf{x} \in I^{2n}$. Then replace r with $r+1$ and repeat the process.

Next we show that

$$\|f_r\|_\infty = \left\| f_{r-1}(\mathbf{x}) - \frac{1}{4n+1}\sum_{q=0}^{4n} g_r(\xi(\mathbf{x}_1), \xi(\mathbf{x}_2)) \right\|_\infty \leq \theta \|f_{r-1}\|_\infty.$$

Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ be an arbitrary point in I^{2n}. For at least $2n+1$ values of q, denoted by q', $\mathbf{x} \in \xi_{q'}[S^k_{q',\mathbf{i}}] \times \xi_{q'}[S^k_{q',\mathbf{j}}]$ for some $(\mathbf{i}, \mathbf{j}) \in D^{2n}_{k_r}$. For these q',

$$g_r(\mathbf{x}) = g_r(\xi_{q'}(\mathbf{x}_1), \xi_{q'}(\mathbf{x}_2)) = f^{k_r}_{q',\mathbf{ij}}$$

and by (6.2.7)

$$|f_{r-1}(\mathbf{x}) - g_r(\xi_{q'}(\mathbf{x}_1), \xi_{q'}(\mathbf{x}_2))| \leq \epsilon \|f_{r-1}\|_\infty.$$

For the remaining q's, denoted by q'', $|g_r(\xi_{q''}(\mathbf{x}))| \leq \|f_{r-1}\|_\infty$. Therefore,

$$\begin{aligned}
& |f_r(\mathbf{x})| \\
= & \left| f_{r-1}(\mathbf{x}) - \frac{1}{4n+1}\sum_{q=0}^{4n} g_r(\xi_q(\mathbf{x}_1), \xi_q(\mathbf{x}_2)) \right| \\
= & \frac{1}{4n+1}\left| \sum_{q=0}^{4n}(4n+1) f_{r-1}(\mathbf{x}) - \sum_{q'} g_r(\xi_{q'}(\mathbf{x}_1), \xi_{q'}(\mathbf{x}_2)) - \sum_{q''} g_r(\xi_{q''}(\mathbf{x}_1), \xi_{q''}(\mathbf{x}_2)) \right| \\
\leq & \frac{1}{4n+1}\left| 2n f_{r-1}(\mathbf{x}) + \sum_{q'} \left(f_{r-1}(\mathbf{x}) - g_r(\xi_{q'}(\mathbf{x}_1), \xi_{q'}(\mathbf{x}_2))\right) \right| + \frac{2n}{4n+1}\|f_{r-1}\|_\infty \\
\leq & \left(\frac{2n+1}{4n+1}\epsilon + \frac{4n}{4n+1} \right) \|f_{r-1}\| \leq \theta \|f_{r-1}\|_\infty.
\end{aligned}$$

Thus,

$$\left| f(\mathbf{x}) - \frac{1}{4n+1}\sum_{l=1}^{r}\sum_{q=0}^{4n} g_r(\xi_q(\mathbf{x}_1), \xi_q(\mathbf{x}_2)) \right| \leq \theta \|f_{r-1}\|_\infty \leq \theta^r \|f\|_\infty$$

and

$$\|g_r\|_\infty \leq \|f_{r-1}\|_\infty \leq \theta \|f_{r-2}\|_\infty \leq \cdots \leq \theta^{r-1}\|f\|_\infty.$$

The series $\sum_{r=1}^{\infty} g_r$ converges uniformly and thus we define $g := \sum_{r=1}^{\infty} g_r$ and (6.2.3) holds. $\square$

6.3 Optimal transport cost

For a fixed K-basis $\{\xi_q\}_{q=0}^{2n}$ on I^n and two probability measures $\mu_1, \mu_2 \in P(I^n)$, let ν_1, ν_2 be defined as in (6.1.1). For any given cost function $c_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$, there exists a cost function $c_n(\mathbf{x}_1, \mathbf{x}_2) : I^n \times I^n \to \mathbb{R}_+$ such that the optimal transport cost between μ_1, μ_2 with cost function c_n is greater or equal to the optimal transport cost between ν_1, ν_2 with cost function c_1. On the other hand, for any given cost function $c_n(\mathbf{x}_1, \mathbf{x}_2) : I^n \times I^n \to \mathbb{R}_+$, there exists a cost function $c_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ such that the optimal transport cost between μ_1, μ_2 with cost function c_n is greater or equal to the optimal transport cost between ν_1, ν_2 with cost function c_1.

Recall the definitions and notations in optimal transport problems introduced in section 2.3. Let $\mu_1, \mu_2 \in P(I^n)$ and $c(\mathbf{x_1}, \mathbf{x_2}) : I^n \times I^n \to \mathbb{R}_+$ be a continuous cost function. Denote

$$\begin{aligned}\Pi(\mu_1, \mu_2) := \{&\pi \in P(I^n \times I^n); \pi(A \times I^n) = \mu_1(A), \pi(I^n \times B) = \mu_2(B),\\ &\forall \text{ measurable} \quad A, B \subset I^n.\}\end{aligned}$$

Then

$$\mathcal{T}_c(\mu_1, \mu_2) = \inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_{I^n \times I^n} c(\mathbf{x_1}, \mathbf{x_2}) d\pi(\mathbf{x_1}, \mathbf{x_2}).$$

is called the optimal transport cost between μ_1 and μ_2.

Theorem 6.3.1. *Given a K-basis $\{\xi_q\}_{q=0}^{2n}$ on I^n and any two measures $\mu_1, \mu_2 \in P(I^n)$, define ν_1, ν_2 by (6.1.1). Denote by $Y_q := \xi_q[I^n]$, the image of I^n under ξ_q, and $Y := \cup_{q=0}^{2n} Y_q$. For any continuous cost function $c_1(y_1, y_2) : Y^2 \to \mathbb{R}_+$, define*

$$c_n(\mathbf{x}_1, \mathbf{x}_2) := \frac{1}{(2n+1)^2} \sum_{q,q'=0}^{2n} c_1(\xi_q(\mathbf{x}_1), \xi_{q'}(\mathbf{x}_2)), \quad \forall (\mathbf{x}_1, \mathbf{x}_2) \in I^n \times I^n. \tag{6.3.1}$$

Then the corresponding optimal transport cost in 1 *dimension is less or equal to the cost in*

n dimension:

$$\mathcal{T}_{c_1}(\nu_1, \nu_2) \le \mathcal{T}_{c_n}(\mu_1, \mu_2). \tag{6.3.2}$$

Moreover, given any continuous cost function $c_n(\mathbf{x}_1, \mathbf{x}_2) : I^n \times I^n \to \mathbb{R}_+$, by Proposition 6.2.1, there is a continuous $\tilde{c}_1 \in C(\mathbb{R}^2)$ such that

$$c_n(\mathbf{x}_1, \mathbf{x}_2) := \frac{1}{(2n+1)^2} \sum_{q,q'=0}^{2n} \tilde{c}_1(\xi_q(\mathbf{x}_1), \xi_{q'}(\mathbf{x}_2)).$$

Then for this $\tilde{c}_1$,

$$\mathcal{T}_{\tilde{c}_1}(\nu_1, \nu_2) \le \mathcal{T}_{c_n}(\mu_1, \mu_2).$$

Proof of Theorem 6.3.1. Firstly, for any $\pi_n(\mathbf{x_1}, \mathbf{x_2}) \in \Pi(\mu_1, \mu_2)$, we have

$$\pi_1(y_1, y_2) := \frac{1}{(2n+1)^2} \left(\sum_{q,q'=0}^{2n} \pi_n[\xi_q^{-1}(y_1), \xi_{q'}^{-1}(y_2)] \right) \in \Pi(\nu_1, \nu_2). \tag{6.3.3}$$

In fact, for any measurable sets $C, D \in Y := \cup_{q=0}^{2n} \xi_q[I^n]$,

$$\begin{aligned} \pi_1(C, Y) &= \frac{1}{(2n+1)^2} \left(\sum_{q,q'=0}^{2n} \pi_n[\xi_q^{-1}[C], \xi_{q'}^{-1}[Y]] \right) \\ &= \frac{1}{(2n+1)^2} \left(\sum_{q,q'=0}^{2n} \pi_n[\xi_q^{-1}[C], I^n] \right) \\ &= \frac{1}{2n+1} \left(\sum_{q=0}^{2n} \mu_1(\xi_q^{-1}[C]) \right) =: \nu_1(C). \end{aligned}$$

Similarly, $\pi_1(Y, D) = \nu_2(D)$.

Secondly,

$$\begin{aligned} &\int_{I^n \times I^n} c_n(\mathbf{x_1}, \mathbf{x_2}) d\pi_n(\mathbf{x_1}, \mathbf{x_2}) \\ = &\int_{I^n \times I^n} \frac{1}{(2n+1)^2} \left(\sum_{q,q'=0}^{2n} c_1(\xi_q(\mathbf{x_1}), \xi_{q'}(\mathbf{x_2})) \right) d\pi_n(\mathbf{x_1}, \mathbf{x_2}) \\ = &\int_{Y \times Y} c_1(y_1, y_2) d\pi_1(y_1, y_2) \end{aligned} \tag{6.3.4}$$

Combining (6.3.3) and (6.3.4), we have

$$\inf_{\pi_n \in \Pi(\mu_1,\mu_2)} \int_{I^n \times I^n} c_n(\mathbf{x_1}, \mathbf{x_2}) d\pi(\mathbf{x_1}, \mathbf{x_2}) \geq \inf_{\pi_1 \in \Pi(\nu_1,\nu_2)} \int_{Y \times Y} c_1(y_1, y_2) d\pi_1(y_1, y_2).$$

That is,

$$\mathcal{T}_{c_n}(\mu_1, \mu_2) \geq \mathcal{T}_{c_1}(\nu_1, \nu_2).$$

□

Using Kantorovich's duality theorem 2.3.3, we can define a simpler cost function c_n for any given continuous cost function c_1 such that (6.3.2) holds.

Theorem 6.3.2. *Given a K-basis $\{\xi_q\}_{q=0}^{2n}$ on I^n and any two measures $\mu_1, \mu_2 \in P(I^n)$, define ν_1, ν_2 by (6.1.1). Denote by $Y_q := \xi_q[I^n]$, the image of I^n under ξ_q, and $Y := \cup_{q=0}^{2n} Y_q$. For any continuous cost function $c_1(y_1, y_2) : Y^2 \to \mathbb{R}_+$, define the cost function $c_n : I^n \times I^n \to \mathbb{R}_+$,*

$$c_n(\mathbf{x}_1, \mathbf{x}_2) := \frac{1}{2n+1} \sum_{q=0}^{2n} c_1(\xi_q(\mathbf{x}_1), \xi_q(\mathbf{x}_2)). \tag{6.3.5}$$

Then

$$\mathcal{T}_{c_1}(\nu_1, \nu_2) \leq \mathcal{T}_{c_n}(\mu_1, \mu_2). \tag{6.3.6}$$

Proof of Theorem 6.3.2. First note that $\nu_1, \nu_2 \in P(Y)$ and $c_n(x_1, x_2)$ is continuous. By Kantorovich duality theorem 2.3.3, it is sufficient to show

$$\sup_{(g_1,g_2) \in \Phi_{c_1}} \left(\int_Y g_1 d\nu_1 + \int_Y g_2 d\nu_2 \right) \leq \sup_{(f_1,f_2) \in \Phi_{c_n}} \left(\int_{I^n} f_1 d\mu_1 + \int_{I^n} f_2 d\mu_2 \right),$$

where Φ_{c_1} contains all $(\phi, \psi) \in L^1(d\nu_1) \times L^1(d\nu_2)$ with $\phi(y_1) + \psi(y_2) \leq c(y_1, y_2)$, for all $d\nu_1$-almost all $y_1 \in Y$ and $d\nu_2$-almost all $y_2 \in Y$. Φ_{c_n} is defined similarly. One can restrict the functions in Φ_{c_1}, Φ_{c_n} to continuous and bounded functions. Thus for $g_1, g_2 \in C(Y)$

with $g_1(y_1) + g_2(y_2) \leq c_1(y_1, y_2)$, we have

$$\begin{aligned} & \frac{1}{2n+1}\left(\sum_{q=0}^{2n} g_1(\xi_q(\mathbf{x}_1)) + \sum_{q=0}^{2n} g_2(\xi_q(\mathbf{x}_2))\right) \\ \leq \; & \frac{1}{2n+1}\sum_{q=0}^{2n} c_1(\xi_q(\mathbf{x}_1), \xi_q(\mathbf{x}_2)) =: c_n(\mathbf{x}_1, \mathbf{x}_2) \end{aligned}$$

That is,

$$\left(\frac{1}{2n+1}\sum_{q=0}^{2n} g_1 \circ \xi_q,\ \frac{1}{2n+1}\sum_{q=0}^{2n} g_2 \circ \xi_q\right) \in \Phi_{c_n}. \tag{6.3.7}$$

By the definition of ν_1, ν_2 and (6.3.7),

$$\begin{aligned} & \sup_{(g_1,g_2)\in\Phi_{c_1}} \left(\int_Y g_1 d\nu_1 + \int_Y g_2 d\nu_2\right) \\ = \; & \sup_{(g_1,g_2)\in\Phi_{c_1}} \frac{1}{2n+1}\left(\int_{I^n}\sum_{q=0}^{2n} g_1 \circ \xi_q d\mu_1 + \int_{I^n}\sum_{q=0}^{2n} g_2 \circ \xi_q d\mu_2\right) \\ \leq \; & \sup_{(f_1,f_2)\in\Phi_{c_n}} \left(\int_{I^n} f_1 d\mu_1 + \int_{I^n} f_2 d\mu_2\right). \end{aligned}$$

That is, $\mathcal{T}_{c_1}(\nu_1, \nu_2) \leq \mathcal{T}_{c_n}(\mu_1, \mu_2)$. $\square$

If the cost function c_1 is a distance function on $Y \times Y$, c_n defined by (6.3.5) is also a distance function on $I^n \times I^n$. Then one can also use Theorem 2.3.4 to show inequality (6.3.6).

The cost function c_n in (6.3.5) for a given cost function c_1 is simpler than the one in (6.3.1). We can also obtain a simpler cost functions c_1 for a given c_n such that (6.3.2) holds by using the redundant version of KST.

We need to modify the definition of the 1-dimensional measures ν first. Given a redundant K-basis $\{\xi_q\}_{q=0}^{4n}$ on I^n and two probability measures $\mu_1, \mu_2 \in P(I^n)$, let

$$\tilde{\nu}_i := \frac{1}{4n+1}\sum_{q=0}^{4n} \mu_i \circ \xi_q^{-1} \tag{6.3.8}$$

be measures defined on the algebra of the subsets $E \subseteq Y$ such that $\xi_q^{-1}[E]$ is μ_i-measurable, $i = 1, 2$.

Corollary 6.3.3. *Let $\mu_1, \mu_2 \in P(I^n)$ and $\{\xi_q\}_{q=0}^{4n}$ be a K-basis on I^n. ν_1, ν_2 are defined as in (6.3.8). Denote by $Y_q := \xi_q[I^n]$, the image of I^n under ξ_q, and $Y := \cup_{q=0}^{4n} Y_q$. For any continuous cost function $c_1(y_1, y_2) : Y^2 \to \mathbb{R}_+$, define*

$$c_n(\mathbf{x}_1, \mathbf{x}_2) := \frac{1}{4n+1} \sum_{q=0}^{4n} c_1(\xi_q(\mathbf{x}_1), \xi_q(\mathbf{x}_2)), \quad \forall (\mathbf{x}_1, \mathbf{x}_2) \in I^n \times I^n. \tag{6.3.9}$$

Then $\mathcal{T}_{c_1}(\tilde{\nu}_1, \tilde{\nu}_2) \leq \mathcal{T}_{c_n}(\mu_1, \mu_2)$.

Moreover, given any continuous cost function $c_n(\mathbf{x}_1, \mathbf{x}_2) : I^n \times I^n \to \mathbb{R}_+$, by Proposition 6.2.2, there is a continuous c_1 such that

$$c_n(\mathbf{x}_1, \mathbf{x}_2) := \frac{1}{4n+1} \sum_{q=0}^{4n} c_1(\xi_q(\mathbf{x}_1), \xi_q(\mathbf{x}_2)).$$

Then for this c_1,

$$\mathcal{T}_{c_1}(\tilde{\nu}_1, \tilde{\nu}_2) \leq \mathcal{T}_{c_n}(\mu_1, \mu_2).$$

Apply Theorem 6.3.1 and Theorem 6.3.2, one can prove Proposition 6.3.3.

Chapter 7

KST applied to image processing

In this chapter, we investigate an encryption scheme developed by using Kolmogorov's superposition theorem. We estimate the maximum error caused by decoding the original image by using wrong K-bases. The error are measured in L^p-norm and Wasserstein distance respectively. Another modified encryption scheme using K-basis and additionally embedding maps is also presented.

The independence of K-basis inspires us to use KST in encryption. Suppose there is some information stored in the form of a multivariate function. For example, a piece of video is a functions of three variables: time and two coordinate variables. Now we want to send the information to someone else confidentially. Instead of encrypting the original information as a multivariate function f directly, we choose a K-basis ξ and represent f by a univariate function g through the corresponding K-map , then send g publicly and keep ξ as secret keys. When the authorized users receive g, the original information f can be reconstructed from g with the authorized secrete keys ξ.

A natural question is how secure is the KST cryptography described above? First, there are infinitely many (a set of second Baire category) K-bases of the form (1.3.1). Therefore it is impossible to crack the keys by exhausting the set of all K-bases. Second, due to the high "non-linearity" of K-basis (see figure 4.1 and 7.1), the construction of a K-basis takes most of the time in the numeric implementation of approximate versions of KST, while the iteration of outer function g converges rapidly. Third, the error caused by decrypting the message with wrong keys could be large.

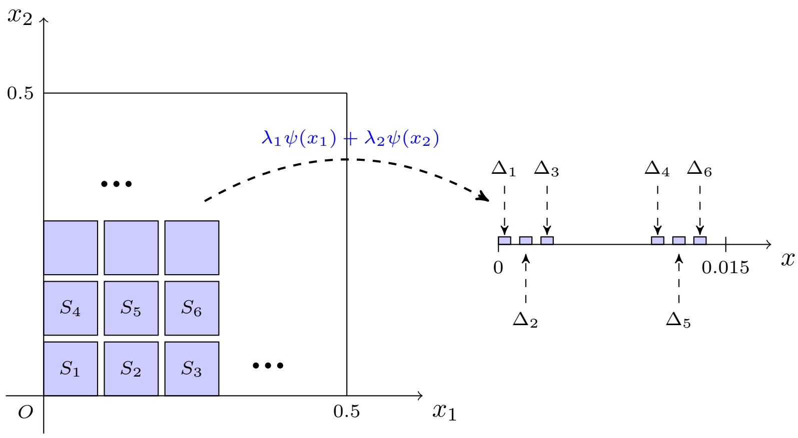

Figure 7.1: The image of $q = 0$ family of squares under Sprecher's K-basis for $n = 2, \lambda_1 = \sqrt{2}/101$ and $\lambda_2 = \sum_{r=1}^{\infty}\gamma^{-(2^n-1)}$. (Here we do not choose $\lambda_1 = 1$ because the distance between Δ_1 and Δ_4 would be much bigger than the distance between Δ_1 and Δ_2 and it is difficult to show them on one axis properly.) We see that under the map of the K-basis $\lambda_1\psi(x_1) + \lambda_2\psi(x_2)$, some neighbourhoods in I^2 are not neighbourhoods on $\mathbb{R}$ any more, e.g, S_1 and S_4.

We mentioned some algorithms to implement the approximate version of KST in section 1.6. Now we give more details on Kolmogorov's spline network designed by Igelnik and Parikh [22].

Theorem 7.0.4 (Estimate of the rate of convergence of Kolmogorov Spline network to the target function [22])**.** *For any function $f \in C^1(I^n)$ and any natural number N, there exists a Kolmogorov spline network defined by*

$$f^s_{N,W}(\mathbf{x}) = \sum_{q=1}^{N} g^s_q\left[\sum_{p=1}^{n}\lambda_p\psi^s_{pq}(x_p,\gamma^{pq}),\gamma^q)\right]$$

where $g^s_q(\cdot,\gamma^1), \psi^s_{pq}(\cdot,\gamma^{pq})$ are cubic spline univariate functions defined on I with parameters γ^1,γ^{pq}, and $\lambda_1,...,\lambda_n > 0$ are rationally independent numbers with $\sum_{p=1}^{n}\lambda_p \leq 1$, such that

$$\|f - f^s_{N,W}\|_\infty = \mathcal{O}(\frac{1}{N}).$$

The number P of net parameters $W := \{\gamma^{pq}, \gamma^q, N\}$ satisfies

$$P = \mathcal{O}(N^{3/2}).$$

This result compares favourably with the approximation order $\mathcal{O}(1/\sqrt{N})$ for general neural networks $f_{N,W}$ and class of functions $C^1(I^n)$ requires $\mathcal{O}(N^2)$ degrees of freedom to achieve the same error. One can compare also [1] [4] [61] and the references cited therein.

Leni et al. [34] implement the idea of encrypting data using Kolmogorov network and approximate the inner and outer functions in KST by splines. Using Igelnik's Kolmogorov Spline network, Leni et al. [35] conducted experiments on greyscale pictures and shows that if the keys are incorrect, the reconstructed pictures have random gray values at every pixel.

7.1 Error in L^p-norm

The property of being a K-basis is a topological property.

Lemma 7.1.1. *Let $\{\xi_q(\mathbf{x})\}_{q=0}^{2n}$ be a Kolmogorov basis on I^n and T be a homeomorphism on I^n, then*

$$\{\tilde{\xi}_q(\mathbf{x})\}_{q=0}^{2n} := \{\xi_q(T(\mathbf{x})\}_{q=0}^{2n}$$

is also a Kolmogorov basis.

For any $f \in C(I^n)$, there exist $\tilde{f} \in C(I^n)$ such that $f = \tilde{f} \circ T$. For this $\tilde{f}$, by KST. there exists a $\tilde{g}$ continuous such that $\tilde{f}(\mathbf{x}) = \sum_{q=0}^{2n} \tilde{g}(\xi_q(\mathbf{x}))$. Thus,

$$f(\mathbf{x}) = \tilde{f}(T(\mathbf{x})) = \sum_{q=0}^{2n} \tilde{g}(\xi_q(T(\mathbf{x}))) = \sum_{q=0}^{2n} \tilde{g}(\tilde{\xi}_q(\mathbf{x})).$$

By the definition of K-basis, $\{\tilde{\xi}_q(\mathbf{x})\}_{q=0}^{2n}$ is a K-basis.

If we decode images encoded by $\{\xi_q(\mathbf{x})\}_{q=0}^{2n}$ with the wrong key $\{\tilde{\xi}_q(\mathbf{x})\}_{q=0}^{2n}$, what is the possible error between f and $\tilde{f}$? We show that the error can be maximised when measured in L^p-norm and Wasserstein distance respectively.

Given a homeomorphism T and a probability measure μ on I^n, we say T **admits a measure decomposition with respect to** μ, iff there exist μ-measurable sets A, B such that

$$I^n = A \cup T[A] \cup B, \quad \text{with } \mu(B) = 0 \quad \text{and} \quad A \cap T[A] = \emptyset.$$

Theorem 7.1.2. *Let* $\{\xi_q(\mathbf{x})\}_{q=0}^{2n}$ *be a K-basis,* μ *be a probability measure on* I^n *that is absolutely continuous to Lebesgue measure, and* T *be a homeomorphism on* I^n*. Let* $\{\tilde{\xi}_q(\mathbf{x}) := \xi_q(T(\mathbf{x})\}_{q=0}^{2n}$ *, then*

$$\sup_{\substack{f \in C(I^n) \\ 0 \le f \le 1}} \left\| \sum_{q=0}^{2n} g_f(\xi_q(\mathbf{x})) - \sum_{q=0}^{2n} g_f(\tilde{\xi}_q(\mathbf{x})) \right\|_{L^p(I^n, \mu)} = 1, \quad 1 \le p \le \infty \tag{7.1.1}$$

if and only if T *admits a measure decomposition with respect to* μ*.*

Theorem 7.1.2 implies that the function reconstructed can be just the "opposite" of the original function in an extreme case. In general cases, all intermediate error is possible, which explains the observation of random greyscale pictures in the reconstruction in [35].

Proof of Theorem 7.1.2. Suppose T admits a measure decomposition of I^n with respect to μ:

$$I^n = A \cup T[A] \cup B, \quad \text{with } \mu(B) = 0 \quad \text{and} \quad A \cap T[A].$$

Then $T[A] \cap T[T[A]] = \emptyset$, since $A \cap T[A] = \emptyset$ and T is a bijection.

Define

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in A \cup B, \\ 0 & \text{otherwise.} \end{cases}$$

f defined above is a measurable function and thus theres exist a sequence of $f_r \in C(I^n)$ such that

$$\lim_{n \to \infty} \int_{I^n} |f_r(\mathbf{x}) - f(\mathbf{x})| d\mu(\mathbf{x}) = 0.$$

Let $g_r := K_\xi f_r, r \in \mathbb{N}$.

$$\begin{aligned}
& \lim_{n\to\infty} \int_{I^n} \left| \sum_{q=0}^{2n} g_r(\xi_q(\mathbf{x})) - \sum_{q=0}^{2n} g_r(\tilde{\xi}_q(\mathbf{x})) \right| d\mu(\mathbf{x}) \\
= & \lim_{n\to\infty} \int_{I^n} |f_r(\mathbf{x}) - f_r(T(\mathbf{x}))| d\mu(\mathbf{x}) \\
= & \int_{I^n} |f(\mathbf{x}) - f(T(\mathbf{x}))| d\mu(\mathbf{x}) \\
= & \int_A |f(\mathbf{x}) - f(T(\mathbf{x}))| d\mu(\mathbf{x}) + \int_{T(A)} |f(\mathbf{x}) - f(T(\mathbf{x}))| d\mu(\mathbf{x}) \\
= & \int_A |1-0| d\mu(\mathbf{x}) + \int_{T(A)} |0-1| d\mu(\mathbf{x}) \\
= & \int_{A\cup T(A)} 1 d\mu(\mathbf{x}) = 1
\end{aligned}$$

Therefore, the supreme error in L^1 is maximised. Similarly, the supreme error is maximised in L^p norm for $1 \leq p \leq \infty$.

On the other hand, suppose (7.1.1) holds, then $|f(x) - \tilde{f}(x)| = 1$ almost everywhere. That is, for mostly all $\mathbf{x}$, either $f(\mathbf{x}) = 0$ and $f(T(\mathbf{x})) = 1$, or $f(\mathbf{x}) = 1$ and $f(T(\mathbf{x})) = 0$. Thus define $A := \{\mathbf{x} \in I^n | f(\mathbf{x}) = 0 \text{ and } f(T(\mathbf{x})) = 1\}$ and $B := I^n \backslash (A \cup T[A])$. Then $A \cap T[A] = \emptyset$ and $\mu(A) + \mu(T[A]) = \mu(I^n)$. That is, T admits a measure decomposition of I^n with respect to μ. □

We illustrate the decomposing property of T with respect to Lebesgue measure m when T is the homeomorphism that permutes the coordinates in dimension 2.

Example 7.1.3. *$T : I^2 \to I^2$ such that $T(x_1, x_2) = (x_2, x_1)$. Then let $A := \{(x_1, x_2) \in I^n | x_1 < x_2\}$ and $B = \{(x_1, x_2) \in I^n | x_1 = x_2\}$, then $A \cap T[A] = \emptyset$ and $\mu(B) = 0$. See figure 7.2.*

For dimension $n \geq 3$, let σ be a permutation of $\{1, ..., n\}$, then $T_\sigma(x_1, ..., x_n) := (x_{\sigma(1)}, ..., x_{\sigma(n)})$ is the homeomorphism on I^n. If $\xi_q(\mathbf{x}) = \sum_{p=1}^n \lambda_p \phi_q(x_p)$ is a Kolmogorov basis on I^n, then by Lemma 7.1.1, $\xi_q \circ T_\sigma = \sum_{p=1}^n \lambda_p \phi_q(x_{\sigma(p)}) = \sum_{p=1}^n \lambda_{\sigma^{-1}(p)} \phi_q(x_p)$ is also a Kolmogorov basis on I^n. There are $n!$ of permutations of the coordinates $(x_1, ..., x_n)$ and thus $n!$ homeomorphisms T_σ. Not all these homeomorphisms maximise the error in

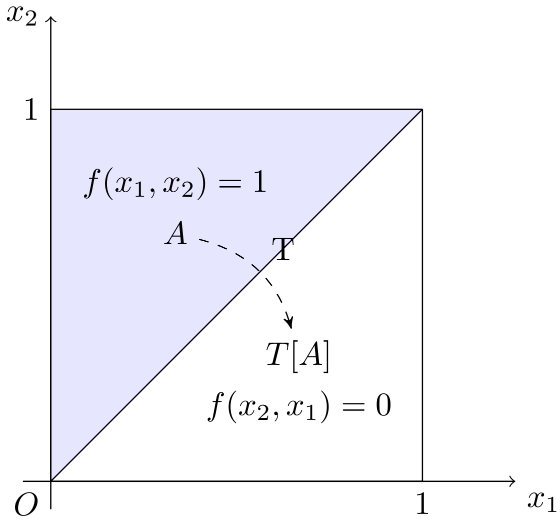

Figure 7.2: Measure decomposition of I^2 under the homeomorphism $T(x_1, x_2) = (x_2, x_1)$.

(7.1.1). If a permutation σ admits a cycle with even length in its cycle decomposition, then T_σ admits a decomposition of I^n with respect to Lebesgue measure.

For example, if $\sigma = (1, 2)\alpha$, where α is a permutation of $\{3, ..., n\}$. The cycle $(1, 2)$ is of length 2. Then let

$$A = \{(x_1, ..., x_n) \in I^n | x_1 > x_2\} \quad \text{and} \quad B = \{(x_1, ..., x_n) \in I^n | x_1 = x_2\}.$$

One can verify that $m(B) = 0$, $T_\sigma(A) = \{(x_1, ..., x_n) \in I^n | x_1 < x_2\}$ and thus $A \cap T[A] = \emptyset$ and $I^n = A \cup T[A] \cup B$. Else if, σ does not admits a even cycle in its decomposition, then the maximal error cannot be obtained. For example, for $\sigma = (1, 2, 3)$, by Theorem 7.1.2, to obtain the maximal error, for almost all $\mathbf{x} := (x_1, x_2, x_3) \in I^3$, $(f(\mathbf{x}), f \circ T_\sigma(\mathbf{x}), f \circ T_\sigma \circ T_\sigma(\mathbf{x}), f \circ T_\sigma \circ T_\sigma \circ T_\sigma(\mathbf{x}) = (0, 1, 0, 1)$ or $(1, 0, 1, 0)$. Notice that $T_\sigma \circ T_\sigma \circ T_\sigma =$ I. Thus $f(\mathbf{x}) = f \circ T_\sigma \circ T_\sigma \circ T_\sigma(\mathbf{x})$ for all $\mathbf{x} \in I^3$, which is a contradiction. See Figure 7.3.

7.2 Error in Wasserstein distance

Next we show that the error between the original functions and the functions reconstructed with wrong keys can also be maximised in Wasserstein distance.

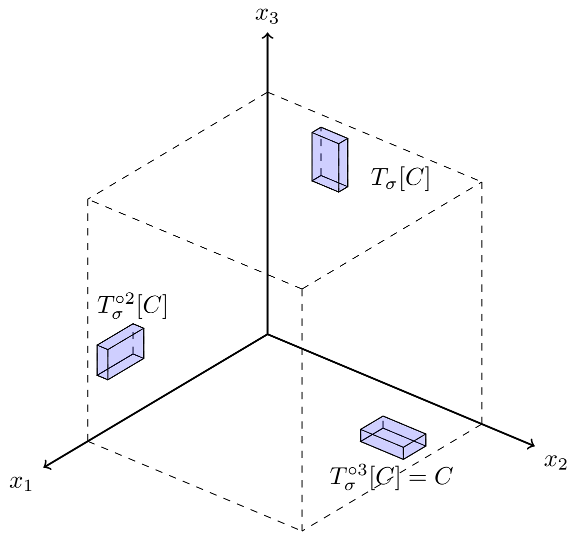

Figure 7.3: The image of a cubic C under the permutative homeomorphism $T(x_1, x_2, x_3) = (x_2, x_3, x_1)$ goes back to itself in odd steps, i.e. 3 steps.

Recall that on a Polish space (X, d), the Wasserstein distance [64] is defined as the optimal transport cost

$$W_p(\mu, \nu) := \mathcal{T}_{d^p}^{1/p}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d^p(x, y) d\pi(x, y) \right)^{1/p}$$

where d is a distance function and μ, ν are probability measures in

$$P_p(X) := \left\{ \mu \in P(X) : \int_X d(x_0, x)^p d\mu(x) < \infty, \text{ for some and thus any } x_0 \in X \right\}.$$

Notice that when the distance d is bounded, then $P(\mathbb{R}^n) = P_p(\mathbb{R}^n),\ p \geq 1$. In the following, we assume that the distance function d is bounded.

Let $\{\xi_q\}_{q=0}^{2n}$ be K-basis for $C(\mathbb{R}^n)$ (see Theorem 1.4.3). We use $\{\xi_q\}_{q=0}^{2n}$ to encode images $f \in C(\mathbb{R}^n)$. Suppose the wrong key is of the type $\tilde{\xi}_q = \xi_q \circ T$ for $q = 0, ..., 2n$, where T is an area-preserving homeomorphism on $\mathbb{R}^n$. Then images decoded with the wrong key $\{\tilde{\xi}_q\}_{q=0}^{2n}$ will be $\tilde{f} = f \circ T$.

Let $(\mathbb{R}^n, d)$ be a Polish space and

$$P(\mathbb{R}^n) := \left\{ f \in C(\mathbb{R}^n) | f \geq 0 \text{ and } \int_{\mathbb{R}^n} f(x)dx = 1 \right\}.$$

If $f \in P(\mathbb{R}^n)$, then $\tilde{f}(x) := f(T(x)) \in P(\mathbb{R}^n)$.

Theorem 7.2.1. *Let T be an area-preserving homeomorphism on $\mathbb{R}^n$ and d be a bounded distance function on $\mathbb{R}^n$. For any $f \in P(\mathbb{R}^n)$ and $\tilde{f} := f \circ T$,*

$$\sup_{f \in P(\mathbb{R}^n)} W_p(f, \tilde{f}) = \sup_{x \in \mathbb{R}^n} d(x, T(x)).$$

The supremum is attained when f is supported on the set $\{x \in \mathbb{R}^n : d(x, T(x)) = \sup_{y \in \mathbb{R}^n} d(y, T(y))\}$.

Proof. We first claim that: for any δ_x, $x \in \mathbb{R}^n$, there exists a sequence $\{f_l \in P(\mathbb{R}^n)\}_{l \in \mathbb{N}}$ such that

$$\lim_{l \to \infty} W_p(f_l, \delta_x) = 0. \tag{7.2.1}$$

In fact, by Theorem 7.12 in [64],

$$W_p(\mu_k, \mu) \to 0, \quad \text{as } k \to \infty,$$

is equivalent to $\mu_k \to \mu$ in weak sense and $\{\mu_k\}$ satisfies the following tightness condition: for some $x_0 \in X$,

$$\lim_{R \to \infty} \limsup_{k \to \infty} \int_{d(x_0, x) \geq R} d(x_0, x)^p d\mu_k(x) = 0.$$

For δ_x, choose $f_l \in P(\mathbb{R}^n)$ such that f is supported in a neighbourhood of x with radius 2^{-l}, then f_l converges to δ_x in weak sense and

$$\lim_{R \to \infty} \limsup_{l \to \infty} \int_{d(x,y) \geq R} d(x, y)^p f_l(y) dy = 0.$$

Thus f_l converges to δ_x in Wasserstein distance.

For any given $x \in \mathbb{R}^n$, let $\{f_l \in P(\mathbb{R}^n)\}_{l \in \mathbb{N}}$ such that (7.2.1) holds. Let $\tilde{f}_l := \tilde{f}_l \circ T$,

then

$$\lim_{l\to\infty} W_p(\tilde{f}_l, \delta_{T(x)}) = 0.$$

By triangle inequality,

$$W_p(f_l, \tilde{f}_l) \le W_p(f_l, \delta_x) + W_p(\delta_x, \delta_{T(x)}) + W_p(\tilde{f}_l, \delta_{T(x)})$$

and

$$W_p(f_l, \tilde{f}_l) \ge W_p(\delta_x, \delta_{T(x)}) - W_p(f_l, \delta_x) - W_p(\tilde{f}_l, \delta_{T(x)}).$$

Then

$$\lim_{l\to\infty} W_p(f_l, \tilde{f}_l) = W_p(\delta_x, \delta_{T(x)}). \tag{7.2.2}$$

For the homeomorphism T, there exists a sequence $\{x_k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^n$ such that

$$\lim_{k\to\infty} d(x_k, T(x_k)) = \sup_{x\in\mathbb{R}^n} d(x, T(x)),$$

and thus

$$\lim_{k\to\infty} W_p(\delta_{x_k}, \delta_{T(x_k)}) = \sup_{x\in\mathbb{R}^n} d(x, T(x)).$$

For each x_k, $k \in \mathbb{N}$, by (7.2.2), there exists a sequence $\{f_{k,l}\}_{l\in\mathbb{N}}$ in $P(\mathbb{R}^n)$ such that

$$\lim_{l\to\infty} W_p(f_{k,l}, \tilde{f}_{k,l}) = W_p(\delta_{x_k}, \delta_{T(x_k)}),$$

where $\tilde{f}_{k,l} := f_{k,l} \circ T$. Then

$$\lim_{k\to\infty} W_p(f_{k,k}, \tilde{f}_{k,k}) = \lim_{k\to\infty} W_p(\delta_{x_k}, \delta_{T(x_k)}) = \sup_{x\in\mathbb{R}^n} d(x, T(x)).$$

Therefore,

$$\sup_{f\in P(\mathbb{R}^n)} W_p(f, \tilde{f}) = \sup_{x\in\mathbb{R}^n} d(x, T(x)).$$

□

7.3 Encryption involving embedding

Consider the embedding of $C(I^n)$ into $C(I^m)$, $m \geq n \geq 2$, by a continuous injective map $U : I^n \hookrightarrow I^m$. For every $f \in C(I^n)$, there exists $F \in C(I^m)$ such that

$$F(y_1, ..., y_m) = f(U(x_1, ..., x_n)), \quad \forall (x_1, ..., x_n) \in I^n.$$

This provides another way to encode f. First, embed f as F, then encode F as an its outer function under a K-basis in dimension m. As in the direct KST encryption method, keep the K-basis in dimension m as secret keys and send the outer function of F publicly. Then to obtain the original function, both the secret keys and the embedding map are needed, which makes the coding scheme safer.

Suppose T is an embedding map from dimension n into dimension m,

$$\begin{aligned} U : I^n &\hookrightarrow I^m \\ (x_1, ..., x_n) &\hookrightarrow (u_1(x_1, ..., x_n), ..., u_m(x_1, ..., x_n)). \end{aligned}$$

Given $\lambda_1, ..., \lambda_m$ and $(\phi_0, ..., \phi_{2m})$ such that $\{\sum_{p=0}^{n} \lambda_p \phi_q(x_p)\}_{q=0}^{2n}$ and $\{\sum_{p=0}^{m} \lambda_p \phi_q(x_p)\}_{q=0}^{2m}$ are K-bases on I^n and I^m respectively. This is feasible by Theorem 3.2.6. Choose any $F \in C(I^m)$ such that

$$F(u_1(x_1, ..., x_n), ..., u_m(x_1, ..., x_n)) = f(x_1, ..., x_n), \quad \forall (x_1, ..., x_n) \in I^n.$$

By KST, there exist g_n depending on f and g_m depending on the chosen F such that

$$f(x_1, ..., x_n) = \frac{1}{2n+1} \sum_{q=0}^{2n} g_n \left(\sum_{p=0}^{n} \lambda_p \phi_q(x_p) \right)$$

and

$$\begin{aligned} & F(u_1(x_1, ..., x_n), ..., u_m(x_1, ..., x_n)) \\ = \ & \frac{1}{2m+1} \sum_{q=0}^{2m} g_m \left(\sum_{p=0}^{m} \lambda_p \phi_q(t_p(x_1, ..., x_n)) \right). \end{aligned}$$

Suppose one receives the public signal g_m and decodes $F(u_1, ..., u_m)$ successfully with the right key $\{\xi_q\}_{q=0}^{2m}$. Now he or she needs to choose a right slice of hypersurface $R \subset I^m$ and map the value of F on R to values of f on $I^n = U^{-1}[R]$. If he or she chooses the wrong embedding map $\tilde{U}$ with wrong hypersurface $\tilde{R} := \tilde{U}[I^n]$and reconstruct wrong $\tilde{f} = F \circ \tilde{U}$, then

$$\begin{aligned} |f(x_1, ..., x_n) - \tilde{f}(x_1, ..., x_n)| &= |F(U(x_1, ..., x_n)) - F(\tilde{U}(x_1, ..., x_n))| \\ &= |F(u_1, ..., u_m) - F(\tilde{u}_1, ..., \tilde{u}_m)| \\ &= \frac{1}{2m+1} \left| \sum_{q=0}^{2m} \left(g_m(\sum_{p=1}^{m} \lambda_p \phi_q(u_p)) - g_m(\sum_{p=1}^{m} \lambda_p \phi_q(\tilde{u}_p)) \right) \right| . \end{aligned}$$

The error between f and $\tilde{f}$ could be large. For example, for any $f \in C(I^n)$ and $\tilde{f} = f \circ T$ with any homeomorphism T on I^n, there is a wrong embedding map $\tilde{U} = U \circ T$ such that $F \circ \tilde{U} = \tilde{f}$. Figure 7.4 compares the original image and the reconstructed image with the right hypersurface but wrong embedding map.

Figure 7.4: Left is the image f of a historic site of Michigan. Right is the reconstructed image with wrong embedding map $\tilde{U} = U \circ T$ with $T(x_1, x_2) = (\frac{x_1+x_2}{2}, \frac{e^{x_1-1}+x_2}{2})$.

Chapter 8

Open problems

There are several open problems related to our research, which we would like to list here:

- K-basis $\{\xi_q\}_{q=0}^m$ can also be in the form of a product, that is, $\xi_q(x_1, ..., x_n) = \prod_{p=1}^n \phi_{pq}(x_p)$ [12]. This provided a possibility to consider KST on spaces endowed with certain group structure. For example, Kolmogorov representation on compact Lie groups.

  Suppose $(G, \circ)$ is a compact Lie group and $C(G \times G)$ is the space of continuous functions on $G \times G$. For each $f \in C(G \times G, \mathbb{R})$, one can try to construct a $g \in C(G, \mathbb{R})$ such that

  $$f(x_1, x_2) = \sum_{q=0}^{m} g(\phi_{1q}(x_1) \circ \phi_{2q}(x_2)) \quad \forall (x_1, x_2) \in G \times G,$$

  with some $m \in \mathbb{N}$ and $m+1$ continuous inner functions $\phi_q : G \to G$, $q = 0, ..., m$.

- Although Sternfeld [59] gave a necessary and sufficient condition of a family of continuous functions $\{\xi_q\}_{q=0}^n$ to be a K-basis for $C(I^n)$, it cannot be readily checked if $\{\xi_q\}_{q=0}^{2n}$ satisfies his condition, i.e., separating the Borel measures on I^n. In chapter 3, we only have a sufficient condition for a family of continuous functions $\{\xi_q\}_{q=0}^{2n}$ to be a K-basis. Namely, if $\{\xi_q\}_{q=0}^{2n}$ separates a Kolmogorov cover, then it is a K-basis. All the constructive proofs of KST by now are based on this property. It is an open question whether this condition is necessary.

- The extension or projection of K-bases among domains of different dimensions is proved only for special cases. The open problem is whether there is a general way to extend or project any K-basis from one dimension to another.

- Vitushkin [68] [67] [66] presented negative results of representation of one function classes by another function class in superposition. He only answers the question when such a representation does not hold, but does not give any description on when does it hold. This is still an open area.

- The non-uniqueness of the outer function in Kolmogorov's representation is only proved when the image of I^n under K-basis ξ is not a connected interval. We infer that it also holds for any K-basis.

- The KST encryption schemes with public outer function introduced Chapter 7 can be implemented. A quantitative description of the error between original image and images constructed with wrong keys can be further discussed. The author and Zegarlinski are working on this [38].

  In the neural networks established using KST, the inner functions and outer functions are approximated with smooth function. The approximate error partially depends on the analytic property of the outer functions in KST. The moduli of outer functions can be used to estimate the approximate error in Kolmogorov's neural network. See [37].

References

[1] A. Tsoi and F. Scarselli. Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results. *Neural networks : the official journal of the International Neural Network Society*, 11(1):15–37, 1998.

[2] V. I. Arnol'd. On functions of three variables. *Doklady Akademii Nauk SSSR*, 114(4):679–681, 1957.

[3] S. P. Banks. Non-linear systems and Kolmogorov's representation theorem. *International Journal of Systems Science*, 22(1):173–180, 1991.

[4] A. R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945, 1993.

[5] F. Bernstein. Untersuchungen aus der Mengenlehre. *Mathematische Annalen*, 61:117–155, 1905.

[6] J. Braun. *An Application of Kolmogorov's Superposition Theorem to Function Reconstruction in Higher Dimensions*. PhD thesis, Univerisity of Bonn, 2009.

[7] J. Braun and M. Griebel. On a constructive proof of Kolmogorov's superposition theorem. *Constructive Approximation*, 30(3):653–675, 2009.

[8] D. W. Bryant. *Analysis of Kolmogorov's superposition theorem and its implementation in applications with low and high dimensional data*. PhD thesis, University of Central Florida, 2008.

[9] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992.

[10] S. Demko. A Superposition Theorem for Bounded Continuous Functions. *Proceedings of the American Mathematical Society*, 66(1):75–78, 1977.

[11] R. Doss. On the representation of continuous functions of two variables by means of addition and continuous functions of one variable. *Colloquium Mathematicum*, X(2):249–259, 1963.

[12] R. Doss. Representations of continuous functions of several variables. *American Journal of Mathematics*, 98(2):375–383, 1976.

[13] R. Doss. A superposition theorem for unbounded continuous functions. *Transactions of the American Mathematical Society*, 233:197–203, 1977.

[14] Z. Feng. *Hilbert's 13th problem*. PhD thesis, University of Pittsburgh, 2010.

[15] B. L. Fridman. An improvement in the smoothness of the functions in Kolmogorov's theorem on superpositions. *Doklady Akademii Nauk SSSR*, 117:1019–1022, 1967.

[16] J. H. Friedman and W. Stuetzle. Projection Pursuit Regression. *Journal of the American Statistical Association*, 76(376):817–823, 1981.

[17] W. Gangbo and R. McCann. Shape recognition via Wasserstein distance. *Quarterly of Applied Mathematics*, 58(4):705–738, 2000.

[18] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, 1952.

[19] R. Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem. In *Proceedings of the international conference on Neural Networks*, volume 3, pages 11–14. New York: IEEE Press, 1987.

[20] T. Hedberg. The Kolmogorov superposition theorem, Appendix II in Topics in Approximation Theory, H Shapiro. *Lecture Notes*, 187, 1971.

[21] D. Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 8(10):437–480, 1902.

[22] B. Igelnik and N. Parikh. Kolmogorov's spline network. *IEEE Transactions on Neural Networks*, 14(4):725–733, 2003.

[23] V. E. Ismailov. On the representation by linear superpositions. *Journal of Approximation Theory*, 151(2):113–125, 2008.

[24] J.-P. Kahane. Sur le Théorème de Superposition de Kolmogorov. *Journal of Approximation Theory*, 13:229–234, 1975.

[25] H. Katsuura and D. Sprecher. Computational aspects of Kolmogorov's superposition theorem. *Neural Networks*, 7(3):455–461, 1994.

[26] S. Khavinson. *Best approximation by linear superpositions (approximate nomography)*, volume 159. American Mathematical Society, 1997.

[27] A. Klopotowski, M. G. Nadkarni, and K. B. Rao. When is $f(x_1, x_2, ..., x_n) = u_1(x_1) + u_2(x_2) + ... + u_n(x_n)$? *Proceedings of the Indian Academy of Sciences (Mathematical Sciences)*, 113(1):77–86, 2003.

[28] A. Kolmogorov. The representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. *Doklady Akademii Nauk SSSR*, 108(2):179–182, 1956.

[29] A. Kolmogorov. The representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, 114:953–956, 1957.

[30] M. Koppen. On the Training of a Kolmogorov Network. *Lecture Notes in Computer Science*, 2415:474–479, 2002.

[31] A. S. Kronrod. On functions of two variables. *Uspekhi Matematicheskikh Nauk*, 5(1):24–134, 1950.

[32] V. Kurkova. Kolmogorov's theorem is relevant. *Neural Computation*, 3(4):617–622, 1991.

[33] P.-E. Leni, Y. Fougerolle, and F. Truchetet. Kolmogorov superposition theorem for image compression. *IET Image Processing*, 6(8):1114–1123, 2012.

[34] P.-E. Leni, Y. D. Fougerolle, and F. Truchetet. Kolmogorov superposition theorem and its application to wavelet image decompositions. *SPIE Proceedings*, 7248:0401–0412, 2009.

[35] P.-E. Leni, Y. D. Fougerolle, and F. Truchetet. Progressive transmission of secured images with authentication using decompositions into monovariate functions. *Journal of Electronic Imaging*, 23(3):033006, 2014.

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

[37] X. Liu. Approximate errors in Kolmogorov neural netwrok approximation . *In preparation*, 2015.

[38] X. Liu and B. Zegarlinski. Error estimates in encryption schemes using Kolmogorov superposition theorem. *In preparation*, 2015.

[39] G. G. Lorentz. Metric entropy, widths, and superpositions of functions. *The American Mathematical Monthly*, 69(6):469–485, 1962.

[40] G. G. Lorentz. *Approximation of Functions*. AMS Chelsea Publishing Series. AMS Chelsea, 2005.

[41] M. Nakamura, R. Mines, and V. Kreinovich. Guaranteed intervals for Kolmogorov's theorem (and their possible relation to neural networks). *Interval Computations*, pages 183–199, 1993.

[42] M. Nees. Approximative versions of Kolmogorov's superposition theorem, proved constructively. *Journal of Computational and Applied Mathematics*, 54(2):239–250, 1994.

[43] P. A. Ostrand. Dimension of metric spaces and Hilbert's problem 13. *Bulletin of American Mathematical Society*, 71(4):619–622, 1965.

[44] A. Ostrowski. Über Dirichletsche Reihen und algebraische Differentialgleichungen. *Mathematische Zeitschrift*, 8(3-4):241–298, 1920.

[45] W. Rudin. *Real and complex analysis*. Mathematics series. McGraw-Hill Inc., 1987.

[46] W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill Inc., 1991.

[47] B. Schmitzer and C. Schnörr. Object segmentation by shape matching with Wasserstein modes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8081 of the series Lecture Notes in Computer Science:123–136, 2013.

[48] D. Sprecher. On the structure of continuous functions of several variables. *Transactions of the American Mathematical Society*, 115:340–355, 1965.

[49] D. Sprecher. On the structure of representation of continuous functions of several variables as finite sums of continuous functions of one variable. *Proceedings of the American Mathematical Society*, 17(1):98–105, 1966.

[50] D. Sprecher. An improvement in the superposition theorem of Kolmogorov. *Journal of Mathematical Analysis and Applications*, 38(1):208–213, 1972.

[51] D. Sprecher. A universal mapping for Kolmogorov's superposition theorem. *Neural Networks*, 6(8):1089–1094, 1993.

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

[53] D. Sprecher. A Numerical Implementationof Kolmogorov's Superpositions II. *Neural Networks*, 10(3):447–457, 1997.

[54] D. Sprecher and S. Draghici. Space-filling curves and Kolmogorov superposition-based neural networks. *Neural networks : the official journal of the International Neural Network Society*, 15(1):57–67, 2002.

[55] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Monographs in harmonic analysis. Princeton University Press, 1970.

[56] E. M. Stein and G. L. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Mathematical Series. Princeton University Press, 1971.

[57] Y. Sternfeld. Uniformly separating families of functions. *Israel Journal of Mathematics*, 29(1):61–91, 1978.

[58] Y. Sternfeld. Superpositions of continuous functions. *Journal of Approximation Theory*, 25:360–368, 1979.

[59] Y. Sternfeld. Uniform separation of points and measures and representation by sums of algebras. *Israel Journal of Mathematics*, 55(3):350–362, 1986.

[60] V. Tikhomorov. Kolmogorov's Work on ϵ-Entropy of Functional Classes and the Superposition of Functions. *Russian Mathematical Surveys*, 18(5):51–87, 1963.

[61] N. W. Townsend and L. Tarassenko. Estimations of error bounds for neural-network function approximators. *IEEE Transactions on Neural Networks*, 10(2):217–230, 1999.

[62] E. Trenklerova. Constructive decomposition of functions of two variables using functions of one variable. *Proceedings of the American Mathematical Society*, 137(2):607–614, 2007.

[63] E. W. Tschirnhaus. A method for removing all intermediate terms from a given equation. *ACM SIGSAM Bulletin*, 37(1):1, 2003.

[64] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.

[65] A. G. Vitushkin. Variations of functions of several variables and sufficient conditions for the variations to be finite. *Doklady Akademii Nauk SSSR*, 96:1089–1091, 1954.

[66] A. G. Vitushkin. Proof of the existence of analytic functions of several complex variables which are not representable by linear superpositions of continuously differentiable functions of fewer variables. *Soviet Mathematics*, 5:793–796, 1964.

[67] A. G. Vitushkin. *On representation of functions by means of superpositions and related topics: lectures*. Monographie de l'Enseignement mathématique. Université, L'Enseignement mathématique, 1978.

[68] A. G. Vitushkin. On Hilbert's thirteenth problem. *Doklady Akademii Nauk SSSR*, 95(4):701–704, 1954.

[69] A. G. Vitushkin and G. M. Khenkin. Linear superpositions of functions. *Russian Mathematical Surveys*, 22(1):77–125, 1967.

[70] J. D. Zuazo. *Fourier Analysis*. Graduate Studies in Mathematics. American Mathematical Society, 2001.

[71] A. Zygmund. *Trigonometric Series*. Cambridge University Press, 2002.

Appendix

In the appendix, we list the concepts and theorems used in the main content of the thesis in alphabetical order.

Concepts

Hölder continuity

Let (X, d_X) and (Y, d_Y) be two metric spaces. Let $f : X \to Y$ be a function. f is said to be Hölder continuous or satisfy a Hölder condition, if there exists $0 < \alpha \leq 1$ and $c > 0$ such that $d_Y(f(x), f(x')) \leq c(d_X(x, x'))^\alpha$ for all $x, x' \in X$. The number α is called the exponent of the Hölder condition.

Lipschitz continuity

A function that is Hölder continuous with exponent $\alpha = 1$ is said to be Lipschitz continuous or satisfy a Lipschitz condition.

Stochastic matrix

$P = (p_{ij})_{n \times n}$ is called a stochastic matrix iff for all $1 \leq i, j \leq n$, $0 \leq p_{ij} \leq 1$ and $\sum_{j=1}^{n} p_{ij} = 1$. A stochastic matrix describes a Markov Chain X_t over a finite state space S. p_{ij} is the probability of moving from state i to state j in one step.

Theorems

Adjoint operators (Corollary after Theorem 4.12 [46])

Suppose X and Y are Banach spaces and $T : X \to Y$ is a bounded linear operator.

Then

(i) The range of T is dense in Y if and only if T^* is injective.

(ii) T is injective if and only if range of T^* is weak*-dense in X^*.

(iii) The range of T is all of Y if and only if T^* is isomorphism into, i.e., T^* is injective and its inverse, mapping range of T^* onto Y^*, is bounded.

Fubini-Tonelli Theorem (Theorem 8.8 [45])

Let $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \lambda)$ be σ-finite measure spaces, and let f be an $(\mathcal{S} \times \mathcal{T})$-measurable function.

(i) If $0 \leq f \leq \infty$, and if

$$\phi(x) = \int_Y f(x,y) d\lambda(y), \quad \psi(y) = \int_X f(x,y) d\mu(x) \quad x \in X, y \in Y,$$

then ϕ is $\mathcal{S}$-measurable, ψ is $\mathcal{T}$-measurable, and

$$\int_X \phi d\mu = \int_{X \times Y} f d(\mu \times \lambda) = \int_Y \psi d\lambda. \tag{.0.1}$$

(ii) If f is complex and if

$$\phi^*(x) = \int_Y |f(x,y)| d\lambda(y) \quad \text{and} \quad \int_X \phi^* d\mu \leq \infty,$$

then $f \in L^1(\mu \times \lambda)$.

(iii) If $f \in L^1(\mu \times \lambda)$, then $f(x,y) \in L^1(Y, \lambda(y))$ for almost all $x \in X$, $f(x,y) \in L^1(X, \mu(x))$ for almost all $y \in Y$; the functions ϕ and ψ almost everywhere are in $L^1(X, \mu(x))$, $L^1(Y, \lambda(y))$ respectively and (.0.1) holds.

Hölder inequality (Theorem 3.5 [45])

Let p and q be conjugate expoents, i.e., $\frac{1}{p} + \frac{1}{q} = 1$. Let X be a measure space with measure μ. Let f and g be measurable functions on X with range in $[0, \infty]$. Then

$$\int_X fg d\mu \leq \left(\int_X f^p d\mu \right)^{1/p} \left(\int_X g^q d\mu \right)^{1/q}. \tag{.0.2}$$

When $p = 2$, (.0.2) is known as the Schwartz inequality.

Minkowski's integral inequality (A.1 [55] or Inequality 202 [18])

Let $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \lambda)$ be σ-finite measure spaces, and let f be an $(\mathcal{S} \times \mathcal{T})$-measurable function. Let $1 \leq p < \infty$, then

$$\left(\int_Y \left(\int_X |f(x,y)| d\mu(x)\right)^p d\lambda(y)\right)^{1/p} \leq \int_X \left(\int_Y |f(x,y)|^p d\lambda(y)\right)^{1/p} d\mu(x).$$

Pointwise convergence of inverse Fourier transform (Corollary 1.21 in [56])

If both f and $\hat{f}$ are integrable on $\mathbb{R}^n$ then

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(t) e^{2\pi i x \cdot t} dt$$

for almost every x.

Riesz-Thorin Interpolation Theorem (Theorem 1.19 [70])

Let $1 \leq p_0, p_1, q_0, q_1 \leq \infty$, and for $0 < \theta < 1$ define p and q by

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \quad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

If T is a linear operator from $L^{p_0} + L^{p_1}$ to $L^{q_0} + L^{q_1}$ such that

$$\|Tf\|_{q_0} \leq M_0 \|f\|_{p_0} \quad \text{for } f \in L^{p_0}$$

and

$$\|Tf\|_{q_1} \leq M_1 \|f\|_{p_1} \quad \text{for } f \in L^{p_1},$$

then

$$\|Tf\|_q \leq M_0^{1-\theta} M_1^{\theta} \|f\|_p \quad \text{for } f \in L^p.$$

Schröder-Bernstein Theorem (Bernstein [5])

If there exist injective functions $f : A \longrightarrow B$ and $g : B \longrightarrow A$ between the sets A and B, then there exists a bijective function $h : A \longrightarrow B$. In terms of the cardinality of the two sets, this means that if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.