UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Information-theoretic aspects of signal analysis and reconstruction

Permalink
https://escholarship.org/uc/item/984348j6

Author
Lim, Taehyung

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Information-theoretic aspects of signal analysis and reconstruction

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Electrical Engineering (Communication Theory and Systems)

by

Taehyung Lim

Committee in charge:

Professor Massimo Franceschetti, Chair
Professor Tara Javidi
Professor Todd Kemp
Professor Alon Orlitsky
Professor Paul H. Siegel

2017

Copyright

Taehyung Lim, 2017

All rights reserved.

The Dissertation of Taehyung Lim is approved and is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Foremost, I would like to thank my advisor Professor Massimo Franceschetti for his guidance and mentorship throughout the duration of my graduate studies. He has always treated me as a colleague and given me complete freedom to find and explore areas of research that interest me. I have really enjoyed working with him, and I will be forever grateful.

Next, It has been my good fortune to have known many wonderful colleagues in Franceschetti's group.

Finally, I would like to express my deepest gratitude towards my friend, Harry. Without your help, I could not have completed complete my doctorate degree. Words cannot express how much I value you, and we're going to get along together forever.

Chapter 2, in full, is a reprint of the material as it appears in T. J. Lim and M. Franceschetti, "Information without rolling dice," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1349-1363, 2017. The dissertation author was the primary investigator and co-author of these papers.

Chapter 3, in full, has been submitted for publication of the material as it might appear in T. J. Lim and M. Franceschetti, "Deterministic coding theorems for blind sensing: optimal measurement rate and fractal dimension". The dissertation author was primary investigator and co-author of this paper.

VITA

2004	Bachelor of Science in Electrical Engineering, Seoul National University
2006	Master of Science in Electrical Engineering and Computer Science, Seoul National University
2017	Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems), University of California, San Diego

PUBLICATIONS

T. J. Lim and M. Franceschetti, "Deterministic coding theorems for blind sensing: optimal measurement rate and fractal dimension," submitted to *IEEE Transactions on Information Theory*.

T. J. Lim and M. Franceschetti, "Information without rolling dice," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1349-1363, 2017.

ABSTRACT OF THE DISSERTATION

Information-theoretic aspects of signal analysis and reconstruction

by

Taehyung Lim

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California, San Diego, 2017

Professor Massimo Franceschetti, Chair

The objective of this thesis is to develop a few approaches to *wave theory of information*. Specifically, this dissertation focuses on two special types of waveforms, bandlimited and multi-band signals. In both cases, we investigate the waveforms in the context of signal analysis and reconstructions.

In the first part of this thesis, we derive the amount of information that can be transmitted by bandlimited waveforms under perturbation, and the amount of information required to represent any bandlimited waveforms within a specific accuracy. These goals can be studied using a stochastic approach or a deterministic approach. Despite

their shared goal of mathematically describing communication using the transmission of waveforms, as well as the common geometric intuition behind their arguments, the two approaches to information theory have evolved separately. The stochastic approach flourished in the context of communication, becoming the pillar of modern digital technologies, while the deterministic approach impacted mostly mathematical analysis. Recent interest in deterministic models has been raised in the context of networked control theory. This brings renewed attention to the deterministic approach in information theory. However, in contrast with the stochastic approaches where the tight results are already known, the previous deterministic results only provide the loose bounds. We improve these results by deriving tight results, and compare our results with the stochastic ones, which reveals the intrinsic similarities of two different approaches.

In the second part of this dissertation, we derive the minimum number of measurements to reconstruct multi-band waveforms, without any spectral information aside from the measure of the whole support set in the frequency domain. This problem is called the completely blind sensing problem and has been an open question. Until a recent date, partially blind sensing has been performed commonly instead, assuming to have some partial spectral information available a priori. We provide an answer for the completely blind sensing problem by deriving the minimum number of measurements to guarantee the reconstruction. The blind sensing problem shares some similarities with the compressed sensing problem. Despite these similarities, due to their different settings, the blind sensing problem contains a few additional difficulties which are not included in the compressed sensing problem. We independently develop our own theory to solve the completely blind sensing problem, and compare our results to those of the compressed sensing problem to reveal the similarities and differences between the two problems.

Chapter 1

Introduction

How much information can be carried in a prescribed waveform? This question is of broad mathematical and physical interest, and has numerous engineering applications, including in communications, sensing, imaging, radar detection and classification systems. In order to answer this question, we should determine a way to gauge the amount of information included in waveforms. We firstly use the notions of capacity and entropy to quantify the amount of information in bandlimited waveforms, and secondly derive the number of measurements needed to represent the amount of information in multi-band waveforms.

In chapter 2, the deterministic notions of capacity and entropy are studied in the context of communication and storage of information using square-integrable, bandlimited signals subject to perturbation. The (ε, δ)-capacity, that extends the Kolmogorov ε-capacity to packing sets of overlap at most δ, is introduced and compared to the Shannon capacity. The functional form of the results indicates that in both Kolmogorov and Shannon's settings, capacity and entropy grow linearly with the number of degrees of freedom, but only logarithmically with the signal to noise ratio. This basic insight transcends the details of the stochastic or deterministic description of the information-theoretic model. For $\delta = 0$ the analysis leads to a tight asymptotic expression of the Kolmogorov ε-entropy of bandlimited signals. A deterministic notion of error exponent is introduced. Applications of the theory are briefly discussed.

In chapter 3, a solution for the completely blind sensing problem of recovering multi-band signals from measurements without any spectral information beside an upper bound on the measure of the whole support set in the frequency domain is presented. Determining the number of measurements necessary and sufficient for reconstruction has been an open problem, and usually partially blind sensing is performed, assuming to have some partial spectral information available a priori. In this paper, the minimum number of measurements that guarantees perfect recovery in the absence of measurement error, and robust recovery in the presence of measurement error, is determined in a completely blind setting. Results show that a factor of two in the measurement rate is the price pay for blindness, compared to reconstruction with full spectral knowledge. The minimum number of measurements is also related to the fractal (Minkowski-Bouligand) dimension of a discrete approximating set, defined in terms of the Kolmogorov ε-entropy. These results are analogous to a deterministic coding theorem, where an operational quantity defined in terms of minimum measurement rate is shown to be equal to an information-theoretic one. A comparison with parallel results in compressed sensing is illustrated, where the relevant dimensionality notion in a stochastic setting is the information (Rényi) dimension, defined in terms of the Shannon entropy.

Chapter 2

Information Without Rolling Dice

2.1 Introduction

Claude Shannon introduced the notions of capacity and entropy in the context of communication in 1948 [1], and with them he ignited a technological revolution. His work instantly became a classic and it is today the pillar of modern digital technologies. On the other side of the globe, the great Soviet mathematician Andrei Kolmogorov was acquainted with Shannon's work in the early 1950s and immediately recognized that "*his mathematical intuition is remarkably precise.*" His notions of ε-entropy and ε-capacity [2, 3] were certainly influenced by Shannon's work. The ε-capacity has the same operational interpretation of Shannon's in terms of the limit for the amount of information that can be transmitted under perturbation, but it was developed in the purely deterministic setting of functional approximation. On the other hand, the ε-entropy corresponds to the amount of information required to represent any function of a given class within ε accuracy, while the Shannon entropy corresponds to the average amount of information required to represent any stochastic process of a given class, quantized at level ε. Kolmogorov's interest in approximation theory dated back to at least the nineteen-thirties, when he introduced the concept of N-width to characterize the "massiveness" or effective dimensionality of an infinite-dimensional functional space [4]. This interest also eventually led him to the solution in the late nineteen-fifties, together with his student Arnold, of Hilbert's thirteenth problem [5].

Even though they shared the goal of mathematically describing the limits of communication and storage of information, Shannon and Kolmogorov's approaches to information theory have evolved separately. Shannon's theory flourished in the context of communication, while Kolmogorov's work impacted mostly mathematical analysis. Connections between their definitions of entropy have been pointed out in [6]. The related concept of complexity and its relation to algorithmic information theory has been treated extensively [7, 8]. Kolmogorov devoted his presentation at the 1956 International Symposium on Information Theory [9], and Appendix II of his work with Tikhomirov [3] to explore the relationship with the probabilistic theory of information developed in the West, but limited the discussion "*at the level of analogy and parallelism.*" This is not surprising, given the state of affairs of the mathematics of functional approximation in the nineteen-fifties — at the time the theory of spectral decomposition of time-frequency limiting operators, needed for a rigorous treatment of continuous waveform channels, had yet to be developed by Landau, Pollack and Slepian [10, 11].

Renewed interest in deterministic models of information has recently been raised in the context of networked control theory [12, 13], and in the context of electromagnetic wave theory [14, 15, 16]. Motivated by these applications, in this paper we define the number of degrees of freedom, or effective dimensionality, of the space of bandlimited functions in terms of N-width, and study capacity and entropy in Kolmogorov's deterministic setting. We also extend Kolmogorov's capacity to packing sets of non-zero overlap, which allows a more detailed comparison with Shannon's work.

2.1.1 Capacity and packing

Shannon's capacity is closely related to the problem of geometric packing "billiard balls" in high-dimensional space. Roughly speaking, each transmitted signal, represented by the coefficients of an orthonormal basis expansion, corresponds to a point in the space, and balls centered at the transmitted points represent the probability density of the uncertainty of the observation performed at the receiver. A certain amount of overlap between the balls is allowed to construct dense packings corresponding to codebooks of high capacity, as long as the overlap does not include typical noise concentration regions, and this allows to achieve reliable communication with vanishing probability of error. The more stringent requirement of communication with probability of error equal to zero leads to the notion of zero-error capacity [17], which depends only on the region of uncertainty of the observation, and not on its probabilistic distribution, and it can be expressed as the supremum of a deterministic information functional [13].

Similarly, in Kolmogorov's deterministic setting communication between a transmitter and a receiver occurs without error, balls of fixed radius ε representing the uncertainty introduced by the noise about each transmitted signal are not allowed to overlap, and his notion of 2ε-capacity corresponds to the Shannon zero-error capacity of the ε-bounded noise channel.

In order to represent a vanishing-error in a deterministic setting, we allow a certain amount of overlap between the ε-balls. In our setting, a codebook is composed by a subset of waveforms in the space, each corresponding to a given message. A transmitter can select any one of these signals, that is observed at the receiver with perturbation at most ε. If signals in the codebook are at distance less than 2ε of each other, a decoding error may occur due to the overlap region between the corresponding ε-balls. The total volume of the error region, normalized by the total volume of the ε-balls in the codebook, represents a measure of the fraction of space where the received signal may fall and result in a communication error. The (ε, δ)-capacity is then defined as the logarithm base two of the largest number of signals that can be placed in a codebook having a normalized error region of size at most δ. We provide upper and lower bounds on this quantity, when communication occurs using bandlimited, square-integrable signals, and introduce a natural notion of deterministic error exponent associated to it, that depends only on the communication rate, on ε, on the signals' bandwidth, and on the energy constraint. Our bounds become tight for high values of the signal to noise ratio, and their functional form indicates that capacity grows linearly with the number of degrees of freedom, but only logarithmically with the signal to noise ratio. This was Shannon's original insight, revisited here in a deterministic setting.

For $\delta = 0$ our notion of capacity reduces to the Kolmogorov 2ε-capacity, and we provide bounds on this quantity. By comparing the lower bound for $\delta > 0$ and the upper bound for $\delta = 0$, we also show that a strict inequality holds between the corresponding values of capacity if the signal to noise ratio is sufficiently large. The analogous result in a probabilistic setting is that the Shannon capacity of the uniform noise channel is strictly greater than the corresponding zero-error capacity.

2.1.2 Entropy and covering

Shannon's entropy is closely related to the geometric problem of covering a high-dimensional space with balls of given radius. Roughly speaking, each source signal, modeled as a stochastic process, corresponds to a random point in the space, and by quantizing all coordinates of the space at a given resolution, Shannon's entropy corresponds to the number of bits needed on average to represent the quantized signal. Thus, the entropy depends on both the probability distribution of the process, and the quantization step along the coordinates of the space. A quantizer, however, does not need to act uniformly on each coordinate, and can be more generally viewed as a discrete set of balls covering the space. The source signal is represented by the closest center of a ball covering it, and the distance to the center of the ball represents the distortion measure associated to this representation. In this setting, Shannon's rate distortion function provides the minimum number of bits that must be specified per unit time to represent the source process with a given average distortion.

In Kolmogorov's deterministic setting, the ε-entropy is the logarithm of the minimum number of balls of radius ε needed to cover the whole space and, when taken per unit time, it corresponds to the Shannon rate-distortion function, as it also represents the minimum number of bits that must be specified per unit time to represent any source signal with distortion at most ε. We provide a tight expression for this quantity, when sources are bandlimited, square-integrable signals. The functional form of our result shows that the ε-entropy grows linearly with the number of degrees of freedom and logarithmically with the ratio of the norm of the signal to the norm of the distortion. Once again, this was Shannon's key insight that remains invariant when subject to a deterministic formulation.

The *leitmotiv* of the paper is the comparison between deterministic and stochastic approaches to information theory, and the presentation is organized as follows: In Section 2.2 we informally describe our results, in section 2.3 we present our model rigorously, provide some definitions, recall results in the literature that are useful for our derivations, and present our technical approach. Section 2.4 briefly discusses applications. Section 2.5 provides precise mathematical statements of our results, along with their proofs. A discussion of previous results and the computation of the error exponent in the deterministic setting appear in the Appendixes.

2.2 Description of the results

We begin with an informal description of our results, that is placed on rigorous grounds in subsequent sections.

2.2.1 Capacity

We consider one-dimensional, real, scalar waveforms of a single scalar variable and supported over an angular frequency interval $[-\Omega, \Omega]$. We assume that waveforms are square-integrable, and satisfy the energy constraint

$$\int_{-\infty}^{\infty} f^2(t)dt \leq E. \tag{2.1}$$

These bandlimited waveforms have unbounded time support but are observed over a finite interval $[-T/2, T/2]$, and the distance between any two waveforms is

$$d(f_1, f_2) = \left(\int_{-T/2}^{T/2} |f_1(t) - f_2(t)|^2 dt \right)^{1/2}. \tag{2.2}$$

In this way, and in a sense to be made precise below, any signal can be expanded in terms of a suitable set of basis functions, orthonormal over the real line, and for T large enough it can be seen as a point in a space of essentially

$$N_0 = \Omega T/\pi \tag{2.3}$$

dimensions, corresponding to the number of degrees of freedom of the waveform, and of radius $\sqrt{E}$.

To introduce the notion of capacity, we consider an uncertainty sphere of radius ε centered at each signal point, representing the energy of the noise that is added to the observed waveform. In this model, due to Kolmogorov, the signal to noise ratio is

$$\mathsf{SNR}_K = E/\varepsilon^2. \tag{2.4}$$

A codebook is composed by a subset of waveforms in the space, each corresponding to a given message. Signals in a codebook are 2ε-distinguishable if the distance between any two of them exceeds 2ε.

Definition 1. *The 2ε-capacity is the logarithm base two of the maximum number $M_{2\varepsilon}(E)$ of 2ε-distinguishable signals in the space, namely*

$$C_{2\varepsilon} = \log M_{2\varepsilon}(E) \quad \textit{bits}. \tag{2.5}$$

When taken per unit time, we have

$$\bar{C}_{2\varepsilon} = \lim_{T\to\infty} \frac{\log M_{2\varepsilon}(E)}{T} \quad \textit{bits per second}. \tag{2.6}$$

The operational meaning of the 2ε-capacity is as follows: a transmitter can select any signal in the codebook, that is observed at the receiver with perturbation at most ε. If the signals in the codebook are at distance at least 2ε of each other, the receiver can decode the message without error. The 2ε-capacity is the logarithm base two of the cardinality of the largest codebook which guarantees decoding without error. It geometrically corresponds to the maximum number of disjoint balls of radius ε with their centers situated inside the signals' space.

A similar Gaussian stochastic model, due to Shannon, considers bandlimited signals in a space of essentially N_0 dimensions, subject to an energy constraint over the interval $[-T/2, T/2]$ that scales linearly with the number of dimensions

$$\int_{-T/2}^{T/2} f^2(t)dt \leq PN_0, \tag{2.7}$$

and adds a zero mean Gaussian noise variable of standard deviation σ independently to each coordinate of the space. In this model, the signal to noise ratio on each coordinate is

$$\mathsf{SNR}_S = P/\sigma^2. \tag{2.8}$$

Definition 2. *The Shannon capacity is the logarithm base two of the largest number of messages $M_\sigma^\delta(P)$ that can be communicated with probability of error $\delta > 0$, namely*

$$C(\delta) = \log M_\sigma^\delta(P) \quad \textit{bits}. \tag{2.9}$$

When taken per unit time, we have

$$C = \lim_{T \to \infty} \frac{\log M_\sigma^\delta(P)}{T} \quad \textit{bits per second}, \tag{2.10}$$

and it does not depend on δ.

The definition in (2.10) should be compared with (2.6). The geometric insight on which the two models are built upon is the same. However, while in Kolmogorov's deterministic model packing is performed with "hard" spheres of radius ε and communication in the presence of arbitrarily distributed noise over a bounded support is performed without error, in Shannon's stochastic model packing is performed with "soft" spheres of effective radius $\sqrt{N_0}\sigma$ and communication in the presence of Gaussian noise of unbounded support is performed with arbitrarily low probability of error δ.

Shannon's energy constraint (2.7) scales with the number of dimensions, rather than being a constant. The reason for this should be clear: since the noise is assumed to act independently on each signal's coefficient, the statistical spread of the output, given the input signal, corresponds to an uncertainty ball of radius $\sqrt{N_0}\sigma$. It follows that the norm of the signal should also be proportional to $\sqrt{N_0}$, to avoid a vanishing signal to noise ratio as $N_0 \to \infty$. In contrast, in the case of Kolmogorov the capacity is computed assuming an uncertainty ball of fixed radius ε and the energy constraint is constant. In both cases, spectral concentration ensures that the size of the signals' space is essentially of N_0 dimensions. Probabilistic concentration ensures that the noise in Shannon's model concentrates around its standard deviation, so that the functional form of the results is similar in the two cases.

Shannon's celebrated formula for the capacity of the Gaussian model is [1]

$$C = \frac{\Omega}{\pi} \log(\sqrt{1 + \mathsf{SNR}_S}) \quad \text{bits per second.} \tag{2.11}$$

Our results for Kolmogorov's deterministic model are (Theorem 4)

$$\begin{cases} \bar{C}_{2\varepsilon} \leq \dfrac{\Omega}{\pi} \log\left(1 + \sqrt{\mathsf{SNR}_K/2}\right) \quad \text{bits per second,} & (2.12) \\ \bar{C}_{2\varepsilon} \geq \dfrac{\Omega}{\pi} \left(\log\sqrt{\mathsf{SNR}_K} - 1\right) \quad \text{bits per second.} & (2.13) \end{cases}$$

The upper bound (2.12) is an improved version of our previous one in [**?**]. For high values of the signal to noise ratio, it becomes approximately $\frac{\Omega}{\pi}\left(\log\sqrt{\mathsf{SNR}_K} - 1/2\right)$, i.e. tight up to a term $\Omega/(2\pi)$. Both upper and lower bounds are improvements over the ones given by Jagerman [18, 19], see Appendix 2.6.1 for a discussion. Similar bounds are obtained by Wyner [20] for timelimited, rather bandlimited signals, assuming they are well concentrated inside the bandwidth, namely only a small fraction of their energy falls outside the band.

To provide a more precise comparison between the deterministic and the stochastic model, we extend the deterministic model allowing signals in the codebook to be at distance less than 2ε of each other. We say that signals in a codebook are (ε,δ)-distinguishable if the portion of space where the received signal may fall and result in a decoding error is of measure at most δ.

Definition 3. *The (ε,δ)-capacity is the logarithm base two of the maximum number $M_\varepsilon^\delta(E)$ of (ε,δ)-distinguishable signals in the space, namely*

$$C_\varepsilon^\delta = \log M_\varepsilon^\delta(E) \quad \textit{bits}. \tag{2.14}$$

When taken per unit time, we have

$$\bar{C}_\varepsilon^\delta = \lim_{T\to\infty} \frac{\log M_\varepsilon^\delta(E)}{T} \quad \textit{bits per second}. \tag{2.15}$$

In this case we show (Theorem 5) that for any $\varepsilon,\delta > 0$

$$\begin{cases} \bar{C}_\varepsilon^\delta \leq \dfrac{\Omega}{\pi}\log\left(1+\sqrt{\mathsf{SNR}_K}\right) \quad \text{bits per second,} & (2.16)\\ \bar{C}_\varepsilon^\delta \geq \dfrac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K} \quad \text{bits per second.} & (2.17)\end{cases}$$

As in Shannon's case, these results do not depend on the size of the error region δ. They become tight for high values of the signal to noise ratio.

The lower bound follows from a random coding argument by reducing the problem to the existence of a coding scheme for a stochastic uniform noise channel with arbitrarily small probability of error. The existence of such a scheme in the stochastic setting implies the existence of a corresponding scheme in the deterministic setting as well. Comparing (2.12) and (2.17), it follows that if

$$1+\sqrt{\mathsf{SNR}_K/2} \leq \sqrt{\mathsf{SNR}_K} \tag{2.18}$$

or equivalently

$$\sqrt{\mathsf{SNR}_K} \geq \frac{\sqrt{2}}{\sqrt{2}-1}, \tag{2.19}$$

then $\bar{C}_\varepsilon^\delta$ is strictly larger than $\bar{C}_{2\varepsilon}$. This means that in the high SNR_K regime having a positive error region guarantees a strictly larger capacity, or equivalently that the Shannon capacity of the uniform noise channel strictly greater than the corresponding zero-error capacity.

The analogy between the size of the error region in the deterministic setting and the probability of error in the stochastic setting also leads to a notion of deterministic error exponent. We define the error exponent in a deterministic setting as the logarithm of the size of the error region divided by observation time T. Letting the number of messages in the codebook be $M = 2^{TR}$, where the transmission rate R is smaller than the lower bound (2.17), in Appendix 2.6.3 we show that the size of the error region is at most

$$\delta \leq 2^{-T\left(\frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K}-R\right)}, \tag{2.20}$$

so that the error exponent is

$$\mathsf{Er}(R) = \frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K} - R > 0, \tag{2.21}$$

which depends only on Ω, E, ε, and on the transmission rate R. From (2.21), it follows that for any rate less than the (ε,δ)-capacity, the size of error region can be arbitrarily small as the observation time goes to infinity.

2.2.2 Entropy

We consider the same signals' space as above, corresponding to points of essentially $N_0 = \Omega T/\pi$ dimensions and contained in a ball of radius $\sqrt{E}$. A source codebook is composed by a subset of points in this space, and each codebook point is a possible representation for the signals that are within radius ε of itself. If the union of the ε balls centered at all codebook points covers the whole space, then any signal in the space can be encoded by its closest representation. The radius ε of the covering balls provides a bound on the largest estimation error between any source $f(t)$ and its codebook representation $\hat{f}(t)$. When signals are observed over a finite time interval $[-T/2, T/2]$, this corresponds to

$$d[f,\hat{f}] = \int_{-T/2}^{T/2} [f(t) - \hat{f}(t)]^2 dt \leq \varepsilon^2. \tag{2.22}$$

Following the usual convention in the literature, we call this distortion measure noise, so that the signal to distortion ratio in this source coding model is again $\mathsf{SNR}_K = E/\varepsilon^2$.

Definition 4. *The ε-entropy is the logarithm base two of the minimum number $L_\varepsilon(E)$ of ε-balls covering the whole space, namely*

$$H_\varepsilon = \log L_\varepsilon(E) \quad \textit{bits}. \tag{2.23}$$

When taken per unit time, we have

$$\bar{H}_\varepsilon = \lim_{T\to\infty} \frac{\log L_\varepsilon(E)}{T} \quad \textit{bits per second}. \tag{2.24}$$

An analogous Gaussian stochastic source model, due to Shannon, models the source signal $\mathsf{f}(t)$ as a white Gaussian stochastic process of constant power spectral density P of support $[-\Omega, \Omega]$. This stochastic process has infinite energy, and finite average power

$$\mathbb{E}(\mathsf{f}^2(t)) = R_{\mathsf{f}}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\mathsf{f}}(\omega)d\omega = \frac{P\Omega}{\pi}, \tag{2.25}$$

where R_{f} and S_{f} are the autocorrelation and the power spectral density of $\mathsf{f}(t)$, respectively. When observed over the interval $[-T/2, T/2]$, the process can be viewed as a random point having essentially N_0 independent Gaussian coordinates of zero mean and variance P, and of energy

$$\int_{-T/2}^{T/2} \mathbb{E}(\mathsf{f}^2(t))dt = \frac{P\Omega T}{\pi} = PN_0. \tag{2.26}$$

A source codebook is composed by a subset of points in the space, and each codebook point is a possible representation for the stochastic process. The distortion associated to the representation of $\mathsf{f}(t)$ using codebook point $\hat{\mathsf{f}}(t)$ is defined in terms of mean-squared error

$$d[\mathsf{f}, \hat{\mathsf{f}}] = \int_{-T/2}^{T/2} \mathbb{E}[\mathsf{f}(t) - \hat{\mathsf{f}}(t)]^2 dt. \tag{2.27}$$

Definition 5. *The rate-distortion function is the logarithm base two of the smallest number of codebook points $L_\sigma(P)$ per unit time that can be used to represent the source process with distortion at most $\sigma^2 N_0$, namely*

$$R_\sigma = \lim_{T\to\infty} \frac{\log L_\sigma(P)}{T} \quad \textit{bits per second.} \tag{2.28}$$

In this setting, Shannon's formula for the rate distortion function of a Gaussian source is [1]

$$R_\sigma = \frac{\Omega}{\pi}\log(\sqrt{\mathsf{SNR}_S}) \quad \text{bits per second.} \tag{2.29}$$

We show the corresponding result in Kolmogorov's deterministic setting (Theorem 6)

$$\bar{H}_\varepsilon = \frac{\Omega}{\pi}\log(\sqrt{\mathsf{SNR}_K}) \quad \text{bits per second.} \tag{2.30}$$

Table 2.1. Comparison of stochastic and deterministic models

	Stochastic	Deterministic
Transmitted Signal	$\int_{-T/2}^{T/2} f^2(t)dt \le PN_0$	$\int_{-\infty}^{\infty} f^2(t)dt \le E$
Additive Noise	$\mathbb{E}\sum_{i=1}^{N_0} \mathsf{n}_i^2 = N_0\sigma^2$	$\sum_{i=1}^{\infty} n_i^2 \le \varepsilon^2$
Signal to Noise Ratio	$\mathsf{SNR}_S = P/\sigma^2$	$\mathsf{SNR}_K = E/\varepsilon^2$
Capacity	$C = \frac{\Omega}{\pi}\log(\sqrt{1+\mathsf{SNR}_S})$	$\frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K} \le \bar{C}_\varepsilon^\delta \le \frac{\Omega}{\pi}\log(1+\sqrt{\mathsf{SNR}_K})$
Source Signal	$\int_{-T/2}^{T/2} \mathbb{E}(\mathsf{f}^2(t))dt = PN_0$	$\int_{-\infty}^{\infty} f^2(t)dt \le E$
Distortion	$d[\mathsf{f},\hat{\mathsf{f}}] \le N_0\sigma^2$	$d[\mathsf{f},\hat{\mathsf{f}}] \le \varepsilon^2$
Rate Distortion Function	$R_\sigma = \frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_S}$	$\bar{H}_\varepsilon = \frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K}$

Previously, Jagerman [18, 19] has shown

$$0 \le \bar{H}_\varepsilon \le \frac{\Omega}{\pi}\log\left(1+2\sqrt{\mathsf{SNR}_K}\right), \tag{2.31}$$

see Appendix 2.6.1 for a discussion.

Our result in (2.30) can be derived by combining a theorem of Dumer, Pinsker and Prelov [21, Theorem 2], on the thinnest covering of ellipsoids in Euclidean spaces of arbitrary dimension, our Lemma 1, on the phase transition of the dimensionality of bandlimited square-integrable functions, and an approximation argument given in our Theorem 6. Instead, we provide a self-contained proof.

2.2.3 Summary

Table 2.1 provides a comparison between results in the deterministic and in the stochastic setting. In the computation of the capacity, a transmitted signal subject to a given energy constraint, is corrupted by additive noise. Due to spectral concentration, the signal has an effective number of dimensions N_0. In a deterministic setting, the noise represented by the deterministic coordinates $\{n_i\}$, can take any value inside a ball of radius ε. In a stochastic setting, due to probabilistic concentration, the noise represented by the stochastic coordinates $\{\mathsf{n}_i\}$, can take values essentially uniformly at random inside a ball of effective radius $N_0\sigma^2$. In both cases, the maximum cardinality of the codebook used for communication depends on the error measure $\delta > 0$, but the capacity in bits per unit time does not, and it depends only on the signal to noise ratio. The special case $\delta = 0$ is treated separately, and it does not appear in the table. This corresponds to the Kolmogorov 2ε-capacity, and is the analog of the Shannon zero-error capacity of an ε-bounded noise channel.

In the computation of the rate distortion function, a source signal is modeled as either an arbitrary, or stochastic process of given energy constraint. The distortion measure corresponds to the estimation error incurred when this signal is represented by an element of the source codebook. The minimum cardinality of the codebook used for representation depends on the distortion constraint, and so does the rate distortion function.

2.3 The signals' space

We now describe the signals' space rigorously, mention some classic results required for our derivations, introduce rigorous notions of capacity and entropy, and present the technical approach that we use in the proofs.

2.3.1 Energy-constrained, bandlimited functions

We consider the set of one-dimensional, real, bandlimited functions

$$\mathscr{B}_\Omega = \{f(t) : \mathscr{F}f(\omega) = 0, \text{ for } |\omega| > \Omega\}, \tag{2.32}$$

where

$$\mathscr{F} f(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-j\omega t) dt, \tag{2.33}$$

and j denotes the imaginary unit.

These functions are assumed to be square-integrable, and to satisfy the energy constraint (2.1). We equip them with the $L^2[-T/2, T/2]$ norm

$$\|f\| = \left(\int_{-\frac{T}{2}}^{\frac{T}{2}} f^2(t) dt \right)^{1/2} \tag{2.34}$$

It follows that $(\mathscr{B}_\Omega, \|\cdot\|)$ is a metric space, whose elements are real, bandlimited functions, of infinite duration and observed over a finite interval $[-T/2, T/2]$. The elements of this space can be optimally approximated, in the sense of Kolmogorov, using a finite series expansion of a suitable basis set.

2.3.2 Prolate spheroidal basis set

Given any $T, \Omega > 0$, there exists a countably infinite set of real functions $\{\psi_n(t)\}$, where $1 \leq n \leq \infty$, called prolate spheroidal wave functions (PSWF), and a set of real positive numbers $1 > \lambda_1 > \lambda_2 > \cdots$ with the following properties:

Property 1. The elements of $\{\lambda_n\}$ and $\{\psi_n\}$ are solutions of the Fredholm integral equation of the second kind

$$\lambda_n \psi_n(t) = \int_{-\frac{T}{2}}^{\frac{T}{2}} \psi_n(s) \frac{\sin \Omega(t-s)}{\pi(t-s)} ds. \tag{2.35}$$

Property 2. The elements of $\{\psi_n(t)\}$ have Fourier transform that is zero for $|\omega| > \Omega$.

Property 3. The set $\{\psi_n(t)\}$ is complete in $\mathscr{B}_\Omega$.

Property 4. The elements of $\{\psi_n(t)\}$ are orthonormal in $(-\infty, \infty)$

$$\int_{-\infty}^{\infty} \psi_n(t)\psi_m(t)dt = \begin{cases} 1 & n = m, \\ 0 & \text{otherwise.} \end{cases} \tag{2.36}$$

Property 5. The elements of $\{\psi_n(t)\}$ are orthogonal in $\left(-\frac{T}{2}, \frac{T}{2}\right)$

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \psi_n(t)\psi_m(t)dt = \begin{cases} \lambda_n & n = m, \\ 0 & \text{otherwise.} \end{cases} \tag{2.37}$$

Property 6. The eigenvalues in $\{\lambda_n\}$ undergo a phase transition at the scale of $N_0 = \Omega T/\pi$: for any $\alpha > 0$

$$\lim_{N_0 \to \infty} \lambda_{\lfloor (1-\alpha)N_0 \rfloor} = 1, \tag{2.38}$$

$$\lim_{N_0 \to \infty} \lambda_{\lfloor (1+\alpha)N_0 \rfloor} = 0. \tag{2.39}$$

Property 7. The width of the phase transition can be precisely characterized: for any $k > 0$

$$\lim_{N_0 \to \infty} \lambda_{\lfloor N_0 + k \log(N_0 \pi/2) \rfloor} = \frac{1}{1 + e^{k\pi^2}}. \tag{2.40}$$

For an extended treatment of PSWF see [22]. The phase transition behavior of the eigenvalues is a key property related to the number of terms required for a satisfactory approximation of any square integrable bandlimited function using a finite basis set. Much of the theory was developed jointly by Landau, Pollack, and Slepian, see [11] for a review. The precise asymptotic behavior in (2.40) was finally proven by Landau and Widom [23], after a conjecture of Slepian supported by a non-rigorous computation [24].

2.3.3 Approximation of $\mathscr{B}_\Omega$

Let $\mathscr{X} = L^2[-T/2, T/2]$ be the space of square-integrable signals with the norm (2.34) as metric, the Kolmogorov N-width [25] of $\mathscr{B}_\Omega$ in $\mathscr{X}$ is

$$d_N(\mathscr{B}_\Omega, \mathscr{X}) = \inf_{\mathscr{X}_N \subseteq \mathscr{X}} \sup_{f \in \mathscr{B}_\Omega} \inf_{g \in \mathscr{X}_N} \|f - g\|, \tag{2.41}$$

where $\mathscr{X}_N$ is an N-dimensional subspace of $\mathscr{X}$. For any $\mu > 0$, we use this notion to define the number of degree of freedom at level μ of the space $\mathscr{B}_\Omega$ as

$$N_\mu(\mathscr{B}_\Omega) = \min\{N : d_N(\mathscr{B}_\Omega, \mathscr{X}) \leq \mu\}. \tag{2.42}$$

In words, the Kolmogorov N-width represents the extent to which $\mathscr{B}_\Omega$ may be uniformly approximated by an N-dimensional subspace of $\mathscr{X}$, and the number of degrees of freedom is the dimension of the minimal subspace representing the elements of $\mathscr{B}_\Omega$ within the desired accuracy μ. It follows that the number of degrees of freedom represents the effective dimensionality of the space, and corresponds to the number of coordinates that are essentially needed to identify any one element in the space.

A basic result in approximation theory (see e.g. [25, Ch. 2, Prop. 2.8]) states that

$$d_N(\mathscr{B}_\Omega, \mathscr{X}) = \sqrt{E\lambda_{N+1}}, \tag{2.43}$$

and the corresponding approximating subspace is the one spanned by the PSWF basis set $\{\psi_n\}_{n=1}^N$. It follows that any bandlimited function $f \in \mathscr{B}_\Omega$ can be optimally approximated by retaining a finite number N of terms in the series expansion

$$f(t) = \sum_{n=1}^{\infty} b_n \psi_n(t), \tag{2.44}$$

and that the number of degree of freedom in (2.42) is given by the minimum index N such that $\sqrt{\lambda_{N+1}} \leq \mu/\sqrt{E}$. The phase transition of the eigenvalues ensures that this number is only slightly larger than N_0. More precisely, for any $\mu > 0$, by (2.40) we may choose an integer

$$N = N_0 + \frac{1}{\pi^2}\log\left(\frac{E}{\mu^2} - 1\right)\log\left(\frac{N_0\pi}{2}\right) + o(\log N_0), \tag{2.45}$$

and approximate

$$\mathscr{B}_\Omega = \left\{\mathbf{b} = (b_1, b_2, \cdots) : \sum_{n=1}^{\infty} b_n^2 \leq E\right\}, \tag{2.46}$$

within accuracy μ as $N_0 \to \infty$ using

$$\mathscr{B}'_\Omega = \left\{\mathbf{b} = (b_1, b_2, \cdots, b_N) : \sum_{n=1}^{N} b_n^2 \leq E\right\}, \tag{2.47}$$

equipped with the norm

$$\|\mathbf{b}\|' = \sqrt{\sum_{n=1}^{N} b_n^2 \lambda_n}. \tag{2.48}$$

The energy constraint in (2.47) follows from (2.1) using the orthonormality Property 4 of the PSWF, the norm in (2.48) follows from (2.34) using the orthogonality Property 5 of the PSWF, and the desired level of approximation is guaranteed by Property 7 of the PSWF.

By (2.45) it follows that the number of degrees of freedom is an intrinsic property of the space, essentially dependent on the time-bandwidth product $N_0 = \Omega T/\pi$, and only weakly, i.e. logarithmically, on the accuracy μ of the approximation and on the energy constraint E.

These approximation-theoretic results show that any energy-constrained, bandlimited waveform can be identified by essentially N_0 real numbers. This does not pose a limit on the amount of information carried by the signal. The real numbers identifying the waveform can be specified up to arbitrary precision, and this results in an infinite number of possible waveforms that can be used for communication. To bound the amount of information, we need to introduce a resolution limit at which the waveform can be observed, which allows an information-theoretic description of the space using bits rather than real numbers. This description is given in terms of entropy and capacity.

2.3.4 ε-entropy and ε-capacity

Let $\mathscr{A}$ be a subset of the metric space $\mathscr{X} = L^2[-T/2, T/2]$. A set of points in $\mathscr{A}$ is called an ε-covering if for any point in $\mathscr{A}$ there exists a point in the covering at distance at most ε from it. The minimum cardinality of an ε-covering is an invariant of the set $\mathscr{A}$, which depends only on ε, and is denoted by $L_\varepsilon(\mathscr{A})$. The ε-*entropy* of $\mathscr{A}$ is the base two logarithm

$$H_\varepsilon(\mathscr{A}) = \log L_\varepsilon(\mathscr{A}) \quad \text{bits}, \tag{2.49}$$

see Fig. 2.1-(a). The ε-entropy per unit time is

$$\bar{H}_\varepsilon(\mathscr{A}) = \lim_{T\to\infty} \frac{H_\varepsilon(\mathscr{A})}{T} \quad \text{bits per second}. \tag{2.50}$$

A set of points in $\mathscr{A}$ is called ε-distinguishable if the distance between any two of them exceeds ε. The maximum cardinality of an ε-distinguishable set is an invariant of the set $\mathscr{A}$, which depends only on ε, and is denoted by $M_\varepsilon(\mathscr{A})$. The ε-*capacity* of $\mathscr{A}$ is the base two logarithm

$$C_\varepsilon(\mathscr{A}) = \log M_\varepsilon(\mathscr{A}) \quad \text{bits}, \tag{2.51}$$

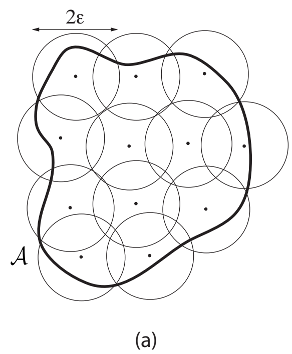

Figure 2.1. Part (a): Illustration of the ε-entropy. Part (b): Illustration of the ε-capacity.

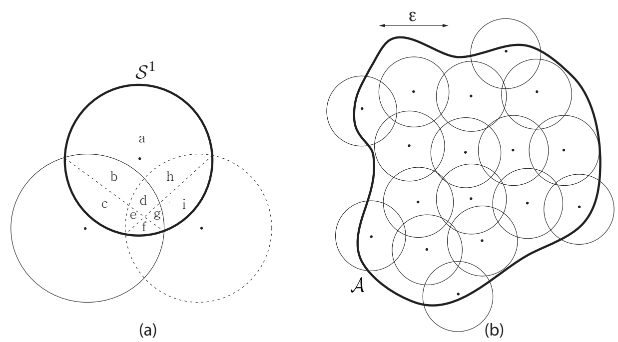

Figure 2.2. Part (a): Illustration of the error region for a signal in the space, where $\Delta_1 = (c+e+f+g+i)/(a+b+c+d+e+f+g+h+i)$. Part (b): Illustration of the (ε, δ)-capacity. An overlap among the ε-balls is allowed, provided that the cumulative error measure $\Delta \leq \delta$.

see Fig. 2.1-(b). The ε-capacity per unit time is

$$\bar{C}_\varepsilon(\mathscr{A}) = \lim_{T\to\infty} \frac{C_\varepsilon(\mathscr{A})}{T} \quad \text{bits per second.} \tag{2.52}$$

The ε-entropy and ε-capacity are closely related to the probabilistic notions of entropy and capacity used in information theory. The ε-entropy corresponds to the rate distortion function, and the ε-capacity corresponds to the zero-error capacity. In order to have a deterministic quantity that corresponds to the Shannon capacity, we extend the ε-capacity and allow a small fraction of intersection among the ε-balls when constructing a packing set. This leads to a certain region of space where the received signal may fall and result in a communication error, and to the notion of (ε, δ)-capacity.

2.3.5 (ε, δ)-capacity

Let $\mathscr{A}$ be a subset of the metric space $\mathscr{X} = L^2[-T/2, T/2]$. We consider a set of points in $\mathscr{A}$, $\mathscr{M} = \{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \cdots, \mathbf{a}^{(M)}\}$. For a given $\mathbf{a}^{(i)}$, $1 \le i \le M$, we let the noise ball

$$\mathscr{S}^i = \{\mathbf{x} \in \mathscr{X} : \|\mathbf{x} - \mathbf{a}^{(i)}\| \le \varepsilon\}, \tag{2.53}$$

where ε is a positive real number, and we let error region with respect to minimum distance decoding

$$\mathscr{D}^i = \{\mathbf{x} \in \mathscr{S}^i : \exists j \neq i : \|\mathbf{x} - \mathbf{a}^{(j)}\| \le \|\mathbf{x} - \mathbf{a}^{(i)}\|\}. \tag{2.54}$$

We define the error measure for the ith signal

$$\Delta_i = \frac{\mathrm{vol}(\mathscr{D}^i)}{\mathrm{vol}(\mathscr{S}^i)}, \tag{2.55}$$

where $\mathrm{vol}(\cdot)$ indicates volume in $\mathscr{X}$, and the cumulative error measure

$$\Delta = \frac{1}{M} \sum_{i=1}^{M} \Delta_i, \tag{2.56}$$

Fig. 2.2-(a) provides an illustration of the error region for a signal in the space. Clearly, we have $0 \le \Delta \le 1$. For any $\delta > 0$, we say that a set of points $\mathscr{M}$ in $\mathscr{A}$ is (ε, δ)-distinguishable set if $\Delta \le \delta$. The maximum cardinality of an (ε, δ)-distinguishable set is an invariant of the space $\mathscr{A}$, which depends only on ε and δ, and is denoted by $M_\varepsilon^\delta(\mathscr{A})$. The (ε, δ)-capacity of $\mathscr{A}$ is the base two logarithm

$$C_\varepsilon^\delta(\mathscr{A}) = \log M_\varepsilon^\delta(\mathscr{A}) \quad \text{bits}, \tag{2.57}$$

see Fig. 2.2-(b). The (ε,δ)-capacity per unit time is

$$\bar{C}_{\varepsilon}^{\delta}(\mathscr{A}) = \lim_{T\to\infty} \frac{C_{\varepsilon}^{\delta}(\mathscr{A})}{T} \quad \text{bits per second.} \tag{2.58}$$

The 2ε-capacity can be regarded as the special case of the (ε,δ)-capacity when $\delta = 0$. Accordingly, from now on we use the notation $C_{\varepsilon}^{0}(\mathscr{A})$ and $\bar{C}_{\varepsilon}^{0}(\mathscr{A})$ to represent $C_{2\varepsilon}(\mathscr{A})$ and $\bar{C}_{2\varepsilon}(\mathscr{A})$.

2.3.6 Technical approach

Our objective is to compute entropy and capacity of square integrable, bandlimited functions. First, we perform this computation for the finite-dimensional space of functions $\mathscr{B}_{\Omega}^{'}$ that approximates the infinite-dimensional space $\mathscr{B}_{\Omega}$ up to arbitrary accuracy $\mu > 0$ in the $L^2[-T/2,T/2]$ norm, as $N_0 \to \infty$. Our results in this setting are given by Theorem 1 for the ε-capacity, Theorem 2 for the (ε,δ)-capacity, and Theorem 3 for the ε-entropy. Then, in Theorems 4, 5, and 6, we extend the computation to the ε-capacity, (ε,δ)-capacity, and ε-entropy of the whole space $\mathscr{B}_{\Omega}$ of bandlimited functions.

When viewed per unit time, results for the two spaces are identical, indicating that using a highly accurate, lower-dimensional subspace approximation leaves only a negligible "information leak" in higher dimensions. We bound this leak in the case of ε-entropy and ε-capacity by performing a projection from the high-dimensional space $\mathscr{B}_{\Omega}$ onto the lower-dimensional one $\mathscr{B}_{\Omega}^{'}$ and noticing that distances do not change significantly when these two spaces are sufficiently close to one another. On the other hand, for the (ε,δ)-capacity the error is defined in terms of volume, which may change significantly, no matter how close the two spaces are. In this case, we cannot bound the (ε,δ) capacity of $\mathscr{B}_{\Omega}$ by performing a projection onto $\mathscr{B}_{\Omega}^{'}$, and instead provide a bound on the capacity per unit time in terms of another finite-dimensional space that asymptotically approximates $\mathscr{B}_\Omega$ with perfect accuracy $\mu = 0$, as $N_0 \to \infty$.

2.4 Possible applications and future work

Recent interest in deterministic models of information has been raised in the context of control theory and electromagnetic wave theory.

Control theory often treats uncertainties and disturbances as bounded unknowns having no statistical structure. In this context, Nair [13] introduced a maximin information functional for non-stochastic variables and used it to derive tight conditions for uniformly estimating the state of a linear time-invariant system over an error-prone channel. The relevance of Nair's approach to estimation over unreliable channels is due to its connection with the Shannon zero-error capacity [13, Theorem 4.1], which has applications in networked control theory [12]. In Appendix 2.6.2 we point out that Nairs' maximum information rate functional, when viewed in our continuous setting of communication with bandlimited signals, is nothing else than $\bar{C}(\mathscr{B}_\Omega)$.

In electromagnetic wave theory, the field measurement accuracy, and the corresponding image resolution in remote sensing applications, are often treated as fixed constants below which distinct electromagnetic fields, corresponding to different images, must be considered indistinguishable. In this framework, much work has been devoted in finding the number of degrees of freedom of radiating fields from their bandlimitation properties [26, 27]. In communication theory, the number of parallel channels available in spatially distributed multiple antenna systems is related to the number of degrees of freedom and can be expressed as the measure of the cut-set boundary separating transmitters and receivers [14]. In this context, our results can be used to obtain a more refined estimate of the size of the signals' space in terms of entropy and capacity, rather than only a first-order characterization in terms of dimensionality. Since electromagnetic signals are functions of space and time, this would require extending results to signals of multiple variables using similar arguments.

Several other applications of the deterministic approach pursued here seem worth exploring, including the analysis of multi-band signals of sparse support. More generally, one could study capacity and entropy under different constrains beside bandlimitation, and attempt, for example, to obtain formulas analogous to waterfilling solutions in a deterministic setting. [28]

2.5 Nothing but proofs

We start with some preliminary lemmas that are needed for the proof of our main theorems. The first lemma is a consequence of the phase transition of the eigenvalues, while the second and third lemmas are properties of Euclidean spaces.

Lemma 1. *Let*

$$\zeta(N) = \left(\prod_{i=1}^{N} \lambda_i \right)^{1/(2N)}, \tag{2.59}$$

where $N = N_0 + O(\log N_0)$ as $N_0 \to \infty$. We have

$$\lim_{N_0 \to \infty} \zeta(N) = 1. \tag{2.60}$$

Proof: For any $\alpha > 0$, we have

$$\begin{aligned} \log \zeta(N) &= \frac{1}{2N} \sum_{i=1}^{N} \log \lambda_i \\ &= \frac{1}{2N} \left(\sum_{i=1}^{\lfloor (1-\alpha) N_0 \rfloor} \log \lambda_i \right. \\ &\qquad \left. + \sum_{i=\lfloor (1-\alpha) N_0 \rfloor + 1}^{N} \log \lambda_i \right). \end{aligned} \tag{2.61}$$

From Property 6 of the PSWF and the monotonicity of the eigenvalues it follows that the first sum in (2.61) tends to zero as $N_0 \to \infty$. We turn our attention to the second sum. By the monotonicity of the eigenvalues, we have

$$\sum_{i=\lfloor(1-\alpha)N_0+1\rfloor}^{N} \log \lambda_i \geq (N-(1-\alpha)N_0) \log \lambda_N. \tag{2.62}$$

Since $N = N_0 + O(\log N_0)$ as $N_0 \to \infty$, there exists a constant k such that for N_0 large enough $N \leq N_0 + k \log N_0$ and the right-hand side is an integer. It follows that for N_0 large enough, we have

$$\begin{aligned}\sum_{i=\lfloor(1-\alpha)N_0+1\rfloor}^{N} \log \lambda_i &\geq (\alpha N_0 + k \log N_0) \log \lambda_N \\ &\geq (\alpha N_0 + k \log N_0) \\ &\quad \times \log(\lambda_{N_0+k\log N_0}).\end{aligned} \tag{2.63}$$

Substituting (2.63) into (2.61) it follows that for N_0 large enough

$$\log \zeta(N) \geq \frac{\alpha N_0 + k \log N_0}{2N} \log(\lambda_{N_0+k\log N_0}), \tag{2.64}$$

and since by Property 7 of the PSWF $\log(\lambda_{N_0+k\log N_0})$ tends to $1/(1+e^{\pi^2 k})$ as $N_0 \to \infty$, we have

$$\lim_{N_0 \to \infty} \log \zeta(N) \geq \frac{\alpha}{2} \log\left(\frac{1}{1+e^{\pi^2 k}}\right). \tag{2.65}$$

The proof is completed by noting that α can be arbitrarily small. $\square$

Lemma 2. [29, Lemma 6.1] *Let m be a positive integer and let $\mathbf{x}, \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(m)}$ be* *arbitrary points in n-dimensional Euclidean space,* $(\mathbb{E}^n, \|\cdot\|)$. *We have*

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\|\mathbf{x}^{(j)}-\mathbf{x}^{(k)}\|^2 \leq 2m\sum_{j=1}^{m}\|\mathbf{x}-\mathbf{x}^{(j)}\|^2. \tag{2.66}$$

Lemma 3. **[30, Theorem 2]** *Let L be the cardinality of the minimal* ε*-covering of the* $\sqrt{E}$*-ball* $\mathscr{S}_{\sqrt{E}}$ *in* $\mathbb{E}^n$*. If* $n \geq 9$*, we have*

$$L \leq \frac{4e\cdot n^{3/2}\left(\frac{\sqrt{E}}{\varepsilon}\right)^n}{\ln n - 2}\left[n\cdot\ln n + o(n\cdot\ln n)\right] \tag{2.67}$$

where $1 < \frac{\sqrt{E}}{\varepsilon} < \frac{n}{\ln n}$.

2.5.1 Main theorems for $\mathscr{B}_{\Omega}^{'}$

Although the set $\mathscr{B}_{\Omega}^{'}$ in (2.47) defines an N-dimensional hypersphere, the metric in (2.48) is not Euclidean. It is convenient to express the metric in Euclidean form by performing a scaling transformation of the coordinates of the space. For all n, we let $a_n = b_n\sqrt{\lambda_n}$, so that we have

$$\mathscr{B}_{\Omega}^{'} = \left\{\mathbf{a} = (a_1, a_2, \cdots, a_N) : \sum_{n=1}^{N}\frac{a_n^2}{\lambda_n} \leq E\right\} \tag{2.68}$$

and

$$\|\mathbf{a}\|' = \sqrt{\sum_{n=1}^{N} a_n^2}. \tag{2.69}$$

We now consider packing and covering with ε-balls inside the ellipsoid defined in (2.68), using the Euclidean metric in (2.69).

Theorem 1. *For any $\varepsilon > 0$, we have*

$$\begin{cases} \bar{C}^0_\varepsilon(\mathscr{B}'_\Omega) \geq \frac{\Omega}{\pi}\left(\log\sqrt{\mathsf{SNR}_K} - 1\right), & (2.70) \\ \bar{C}^0_\varepsilon(\mathscr{B}'_\Omega) \leq \frac{\Omega}{\pi}\log\left(1+\sqrt{\mathsf{SNR}_K}/2\right), & (2.71) \end{cases}$$

where $\mathsf{SNR}_K = E/\varepsilon^2$.

Proof: To prove the result it is enough to show the following inequalities for the 2ε-capacity

$$C^0_\varepsilon(\mathscr{B}'_\Omega) \geq N\left[\log\left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right) - 1\right], \tag{2.72}$$

$$C^0_\varepsilon(\mathscr{B}'_\Omega) \leq N\left[\log\left(1+\frac{\sqrt{E}}{\sqrt{2}\varepsilon}\right)\right] + \log\left(1+\frac{N}{2}\right), \tag{2.73}$$

because $\lim_{T\to\infty}\zeta(N) = 1$ and $\log\left(1+\frac{N}{2}\right) = o(T)$.

Lower bound. Let $\mathscr{M}^0_\varepsilon$ be a maximal $(\varepsilon,0)$-distinguishable subset of $\mathscr{B}'_\Omega$ and $M^0_\varepsilon(\mathscr{B}'_\Omega)$ be the number of elements in $\mathscr{M}^0_\varepsilon$. For each point of $\mathscr{M}^0_\varepsilon$, we consider an Euclidean ball whose center is the chosen point and whose radius is 2ε. Let $\mathscr{U}$ be the union of these balls. We claim that $\mathscr{B}'_\Omega$ is contained in $\mathscr{U}$. If that is not the case, we can find a point of $\mathscr{B}'_\Omega$ which is not contained in $\mathscr{M}^0_\varepsilon$, but whose distance from every point in $\mathscr{M}^0_\varepsilon$ exceeds 2ε, which is a contradiction. Thus, we have the chain of inequalities

$$\mathrm{vol}(\mathscr{B}'_\Omega) \leq \mathrm{vol}(\mathscr{U}) \leq M^0_\varepsilon(\mathscr{B}'_\Omega)\mathrm{vol}(\mathscr{S}_{2\varepsilon}), \tag{2.74}$$

where $\mathscr{S}_{2\varepsilon}$ is an Euclidean ball whose radius is 2ε and the second inequality follows from a union bound. Since $\mathrm{vol}(\mathscr{S}_\varepsilon) = \beta_N \cdot \varepsilon^N$, where β_N is the volume of $\mathscr{S}_1$, by (2.74) we have

$$\left(\frac{1}{2}\right)^N \frac{\mathrm{vol}(\mathscr{B}'_\Omega)}{\mathrm{vol}(\mathscr{S}_\varepsilon)} \leq M^0_\varepsilon(\mathscr{B}'_\Omega). \tag{2.75}$$

Since $\mathscr{B}_{\Omega}^{'}$ is an ellipsoid of radii $\{\sqrt{\lambda_i E}\}_{i=1}^{N}$, we also have

$$\mathrm{vol}(\mathscr{B}_{\Omega}^{'}) = \beta_N \prod_{i=1}^{N} \sqrt{\lambda_i E} = \beta_N \left(\zeta(N)\sqrt{E}\right)^N, \tag{2.76}$$

and

$$\frac{\mathrm{vol}(\mathscr{B}_{\Omega}^{'})}{\mathrm{vol}(\mathscr{S}_{\varepsilon})} = \left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right)^N. \tag{2.77}$$

By combining (2.75) and (2.77), we get

$$C_{\varepsilon}^{0}(\mathscr{B}_{\Omega}^{'}) = \log M_{\varepsilon}^{0}(\mathscr{B}_{\Omega}^{'}) \geq N\left[\log\left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right) - 1\right]. \tag{2.78}$$

Upper bound. We define the auxiliary set

$$\bar{\mathscr{B}}_{\Omega}^{'} = \left\{\mathbf{a} = (a_1, a_2, \cdots, a_N) : \sum_{n=1}^{N} a_n^2 \leq E\right\}. \tag{2.79}$$

The corresponding space $(\bar{\mathscr{B}}_{\Omega}^{'}, \|\cdot\|')$ is Euclidean. Since $\mathscr{B}_{\Omega}^{'} \subset \bar{\mathscr{B}}_{\Omega}^{'}$, it follows that $C_{\varepsilon}^{0}(\mathscr{B}_{\Omega}^{'}) \leq C_{\varepsilon}^{0}(\bar{\mathscr{B}}_{\Omega}^{'})$ and it is sufficient to derive an upper bound for $C_{\varepsilon}^{0}(\bar{\mathscr{B}}_{\Omega}^{'})$.

Let $\mathscr{M}_{\varepsilon}^{0} = \{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \cdots, \mathbf{a}^{(M)}\}$ be a maximal $(\varepsilon, 0)$-distinguishable subset of $\bar{\mathscr{B}}_{\Omega}^{'}$, where $M = M_{\varepsilon}^{0}(\bar{\mathscr{B}}_{\Omega}^{'})$. Let $\{\mathbf{a}^{(i_1)}, \mathbf{a}^{(i_2)} \cdots \mathbf{a}^{(i_m)}\}$ be any subset of $\mathscr{M}_{\varepsilon}^{0}$. For any integer $j \neq k$, $j,k \in \{1, \ldots m\}$, we have

$$\|\mathbf{a}^{(i_j)} - \mathbf{a}^{(i_k)}\|' \geq 2\varepsilon, \tag{2.80}$$

and

$$\sum_{j=1}^{m}\sum_{k=1}^{m} \|\mathbf{a}^{(i_j)} - \mathbf{a}^{(i_k)}\|'^2 \geq 4\varepsilon^2 m(m-1). \tag{2.81}$$

By Lemma 2 it follows that

$$\sum_{j=1}^{m} \|\mathbf{a}-\mathbf{a}^{(i_j)}\|'^2 \geq 2\varepsilon^2(m-1). \tag{2.82}$$

We now define the function

$$\gamma(x) = \max\{0, 1-\frac{1}{2\varepsilon^2}x^2\}, \tag{2.83}$$

and for any $\mathbf{a} \in \mathbb{E}^N$, we let $\mathscr{M}_{\mathbf{a}} = \{\mathbf{a}^{(i_1)}, \mathbf{a}^{(i_2)} \cdots \mathbf{a}^{(i_m)}\}$ be a subset of $\mathscr{M}_{\varepsilon}^0$ whose distance from $\mathbf{a}$ is not larger than $\sqrt{2}\varepsilon$. We have

$$\begin{aligned}
\sum_{j=1}^{M} \gamma(\|\mathbf{a}-\mathbf{a}^{(j)}\|') &= \sum_{k=1}^{m} \gamma(\|\mathbf{a}-\mathbf{a}^{(i_k)}\|') \\
&= \sum_{k=1}^{m} \left(1-\frac{1}{2\varepsilon^2}\|\mathbf{a}-\mathbf{a}^{(i_k)}\|'^2\right) \\
&= m-\frac{1}{2\varepsilon^2}\sum_{k=1}^{m} \|\mathbf{a}-\mathbf{a}^{(i_k)}\|'^2 \\
&\leq m-(m-1) \\
&= 1, \qquad (2.84)
\end{aligned}$$

where the last inequality follows from (2.82). If $\mathbf{a} \notin \mathscr{S}_{\sqrt{E}+\sqrt{2}\varepsilon}$, then $\sum_{j=1}^{M} \gamma(\|\mathbf{a}-\mathbf{a}^{(j)}\|') = 0$ because $\mathscr{M}_{\mathbf{a}} = \emptyset$. By using (2.84) and this last observation, we perform the following computation:

$$\begin{aligned}
\mathrm{vol}\left(\mathscr{S}_{\sqrt{E}+\sqrt{2}\varepsilon}\right) &= \int_{\mathscr{S}_{\sqrt{E}+\sqrt{2}\varepsilon}} d\mathbf{a} \\
&\geq \int_{\mathscr{S}_{\sqrt{E}+\sqrt{2}\varepsilon}} \sum_{j=1}^{M} \gamma(\|\mathbf{a}-\mathbf{a}^{(j)}\|')d\mathbf{a} \\
&= \sum_{j=1}^{M} \int_{\mathbb{E}^N} \gamma(\|\mathbf{a}-\mathbf{a}^{(j)}\|')d\mathbf{a} \\
&= M\int_{\mathbb{E}^N} \gamma(\|\mathbf{a}\|')d\mathbf{a} \\
&= M\int_{0}^{\sqrt{2}\varepsilon} \gamma(x)d(\beta_N x^N) \\
&= \beta_N MN\int_{0}^{\sqrt{2}\varepsilon} \gamma(x)x^{N-1}dx \\
&= \frac{2\beta_N M}{N+2}(\sqrt{2}\varepsilon)^N, \qquad (2.85)
\end{aligned}$$

where the inequality follows from (2.84), the third equality follows from the fact that the value of the integral is independent of $\mathbf{a}^{(j)}$, and the last equality follows from (2.83).

Since $\mathrm{vol}\left(\mathscr{S}_{\sqrt{E}+\sqrt{2}\varepsilon}\right) = \beta_N(\sqrt{E}+\sqrt{2}\varepsilon)^N$, we obtain

$$M_\varepsilon^0(\bar{\mathscr{B}}'_\Omega) = M \leq \frac{N+2}{2}\left(1+\frac{\sqrt{E}}{\sqrt{2}\varepsilon}\right)^N. \qquad (2.86)$$

The proof is completed by taking the logarithm. □

Theorem 2. *For any* $0<\delta<1$ *and* $\varepsilon>0$*, we have*

$$\begin{cases}
\bar{C}_\varepsilon^\delta(\mathscr{B}'_\Omega) \geq \dfrac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K}, & (2.87)\\
\bar{C}_\varepsilon^\delta(\mathscr{B}'_\Omega) \leq \dfrac{\Omega}{\pi}\log\left(1+\sqrt{\mathsf{SNR}_K}\right). & (2.88)
\end{cases}$$

where $\mathsf{SNR}_K = E/\varepsilon^2$.

Proof: To prove the result it is enough to show the following inequalities for the (ε,δ)-capacity

$$C_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega}^{'}) \geq N\left[\log\left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right)\right]+\log\delta, \tag{2.89}$$

$$C_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega}^{'}) \leq N\left[\log\left(1+\frac{\sqrt{E}}{\varepsilon}\right)\right]+\log\frac{1}{1-\delta}, \tag{2.90}$$

because $\lim_{T\to\infty}\zeta(N)=1$ and both $\log\delta$ and $\log\frac{1}{1-\delta}$ are $o(T)$.

Lower bound. We show that there exists a codebook $\mathscr{M}=\{\mathbf{a}^{(1)},\mathbf{a}^{(2)},\cdots,\mathbf{a}^{(M)}\}$, where

$$M=\delta\left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right)^{N}, \tag{2.91}$$

that has cumulative error measure $\Delta\leq\delta$. To prove this result, we consider an auxiliary stochastic communication model where the transmitter selects a signal uniformly at random from a given codebook and, given the signal $\mathbf{a}^{(i)}$ is sent, the receiver observes $\mathbf{a}^{(i)}+\mathbf{n}$, with $\mathbf{n}$ distributed uniformly in $\mathscr{S}_{\varepsilon}$. The receiver compares this signal with all signals in the codebook and selects the one that is nearest to it as the one actually sent. Using the definitions of $\mathscr{S}^i$ and $\mathscr{D}^i$ in (2.53) and (2.54), the decoding error probability of this stochastic communication model, averaged over the uniform selection of signals in the codebook, is represented by

$$P_{err}=\frac{1}{M}\sum_{i=1}^{M}\frac{\mathrm{vol}(\mathscr{D}^i)}{\mathrm{vol}(\mathscr{S}^i)}, \tag{2.92}$$

and by (2.55) and (2.56) it corresponds to the cumulative error measure Δ of the deterministic model that uses the same codebook. It follows that in order to prove the desired lower bound in the deterministic model, we can show that there exists a codebook in the stochastic model satisfying (2.91), and whose decoding error probability is at most δ. This follows from a standard random coding argument, in conjunction to a less standard geometric argument due to the metric employed.

We construct a random codebook by selecting M signals uniformly at random inside the ellipsoid $\mathscr{B}'_\Omega$. We indicate the average error probability over all signal selections in the codebook and over all codebooks and by $\bar{P}_{err}$. Since all signals in the codebook have the same error probability when averaged over all codebooks, $\bar{P}_{err}$ is the same as the average error probability over all codebooks when $\mathbf{a}^{(1)}$ is transmitted. Let in this case the received signal be $\mathbf{y}$ and let $\mathscr{S}_\varepsilon^{\mathbf{y}}$ be an Euclidean ball whose radius is ε and center is $\mathbf{y}$.

The probability that the signal $\mathbf{y}$ is decoded correctly is at least as large as the probability that the remaining $M-1$ signals in the codebook are in $\mathscr{B}'_\Omega \setminus \mathscr{S}_\varepsilon^{\mathbf{y}}$. By the union bound, we have

$$\begin{aligned}
1-\bar{P}_{err} &\geq 1-(M-1)\frac{\mathrm{vol}(\mathscr{S}_\varepsilon^{\mathbf{y}})}{\mathrm{vol}(\mathscr{B}'_\Omega)} \\
&\geq 1-M\frac{\mathrm{vol}(\mathscr{S}_\varepsilon^{\mathbf{y}})}{\mathrm{vol}(\mathscr{B}'_\Omega)} \\
&= 1-M\left(\frac{\varepsilon}{\zeta(N)\sqrt{E}}\right)^N, \qquad (2.93)
\end{aligned}$$

where the last equality follows from (2.77). Letting $M = \delta\left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right)^N$, we have $\bar{P}_{err} \leq \delta$. This implies that there exist a given codebook for which the average probability of error over the selection of signals in the codebook given in (2.92) is at most δ. When this same codebook is applied in the deterministic model, we also have a cumulative error measure $\Delta \leq \delta$.

Upper bound. Let $\mathscr{M}_\varepsilon^\delta$ be a maximal (ε,δ)-distinguishable subset of $\mathscr{B}'_\Omega$ and $M_\varepsilon^\delta(\mathscr{B}'_\Omega) = M$ be the number of elements in $\mathscr{M}_\varepsilon^\delta$. Let $\hat{\mathscr{B}}'_\Omega$ be the union of $\mathscr{B}'_\Omega$ and the trace of the inner points of an ε-ball whose center is moved along the boundary of $\mathscr{B}'_\Omega$, as depicted in Fig. 2.3.

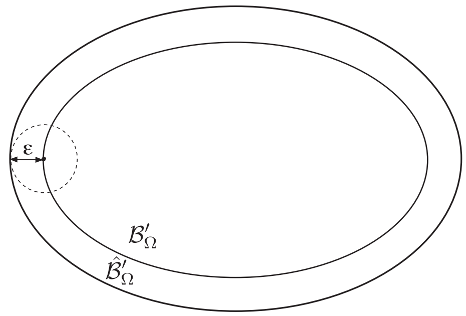

Figure 2.3. Illustration of the relationship between $\mathscr{B}'_\Omega$ and $\hat{\mathscr{B}}'_\Omega$.

Since $\bigcup_{i=1}^{M} \mathscr{S}^i \subset \hat{\mathscr{B}}'_\Omega$, we have

$$\mathrm{vol}\left(\bigcup_{i=1}^{M} \mathscr{S}^i\right) \le \mathrm{vol}\left(\hat{\mathscr{B}}'_\Omega\right). \tag{2.94}$$

Since $\bigcup_{i=1}^{M} \mathscr{S}^i = \bigcup_{i=1}^{M} (\mathscr{S}^i \setminus \mathscr{D}^i)$ and $(\mathscr{S}^i \setminus \mathscr{D}^i) \cap (\mathscr{S}^j \setminus \mathscr{D}^j) = \emptyset$ for $i \neq j$, we obtain

$$\begin{aligned}
\mathrm{vol}\left(\bigcup_{i=1}^{M} \mathscr{S}^i\right) &= \sum_{i=1}^{M} \left[\mathrm{vol}(\mathscr{S}^i) - \mathrm{vol}(\mathscr{D}^i)\right] \\
&= \sum_{i=1}^{M} \mathrm{vol}(\mathscr{S}^i) \left[1 - \frac{\mathrm{vol}(\mathscr{D}^i)}{\mathrm{vol}(\mathscr{S}^i)}\right] \\
&= \sum_{i=1}^{M} \mathrm{vol}(\mathscr{S}^i)\,(1 - \Delta_i) \\
&= M \cdot \mathrm{vol}(\mathscr{S}_\varepsilon)(1 - \Delta).
\end{aligned} \tag{2.95}$$

Since $\hat{\mathscr{B}}'_\Omega \subset \mathscr{S}_{\sqrt{E}+\varepsilon}$, (2.94) can be rewritten as

$$M\mathrm{vol}(\mathscr{S}_\varepsilon)(1 - \Delta) \le \mathrm{vol}\left(\mathscr{S}_{\sqrt{E}+\varepsilon}\right) \tag{2.96}$$

or equivalently

$$M \le \frac{1}{1-\Delta} \frac{\mathrm{vol}\left(\mathscr{S}_{\sqrt{E}+\varepsilon}\right)}{\mathrm{vol}(\mathscr{S}_\varepsilon)} = \frac{1}{1-\Delta}\left(1 + \frac{\sqrt{E}}{\varepsilon}\right)^N. \tag{2.97}$$

Since $C_\varepsilon^\delta(\mathscr{B}_\Omega^{'}) = \log M$ and $\Delta \le \delta$, the result follows. $\square$

Theorem 3. *For any* $\varepsilon > 0$*, we have*

$$\bar{H}_\varepsilon(\mathscr{B}_\Omega^{'}) = \frac{\Omega}{\pi} \log \sqrt{\mathsf{SNR}_K}, \tag{2.98}$$

where $\mathsf{SNR}_K = E/\varepsilon^2$.

Proof: To prove the result it is enough to show the following inequalities for the ε-entropy

$$H_\varepsilon(\mathscr{B}_\Omega^{'}) \ge N \left[\log \left(\zeta(N) \frac{\sqrt{E}}{\varepsilon} \right) \right], \tag{2.99}$$

$$H_\varepsilon(\mathscr{B}_\Omega^{'}) \le N \left[\log \left(\frac{\sqrt{E}}{\varepsilon} \right) \right] + \eta(N), \tag{2.100}$$

where $\eta(N) = o(T)$ and $\lim_{T\to\infty} \zeta(N) = 1$.

Lower bound. Let $\mathscr{L}_\varepsilon$ be a minimal ε-covering subset of $\mathscr{B}_\Omega^{'}$ and $L_\varepsilon(\mathscr{B}_\Omega^{'})$ be the number of elements in $\mathscr{L}_\varepsilon$. Since $\mathscr{L}_\varepsilon$ is an ε-covering, we have

$$\mathrm{vol}(\mathscr{B}_\Omega^{'}) \le L_\varepsilon(\mathscr{B}_\Omega^{'}) \mathrm{vol}(\mathscr{S}_\varepsilon), \tag{2.101}$$

where $\mathscr{S}_\varepsilon$ is an Euclidean ball whose radius is ε. By combining (2.77) and (2.101), we have

$$L_\varepsilon(\mathscr{B}_\Omega^{'}) \ge \left(\zeta(N) \frac{\sqrt{E}}{\varepsilon} \right)^N. \tag{2.102}$$

The proof is completed by taking the logarithm.

Upper bound. We define the auxiliary set

$$\bar{\mathscr{B}}_\Omega^{'} = \left\{ \mathbf{a} = (a_1, a_2, \cdots, a_N) : \sum_{n=1}^{N} a_n^2 \le E \right\}. \tag{2.103}$$

The corresponding space $(\bar{\mathscr{B}}_\Omega^{'}, \|\cdot\|')$ is Euclidean. Since $\mathscr{B}_\Omega^{'} \subset \bar{\mathscr{B}}_\Omega^{'}$, it follows that

$H_\varepsilon(\mathscr{B}_\Omega^{'}) \le H_\varepsilon(\bar{\mathscr{B}}_\Omega^{'})$ and it is sufficient to derive an upper bound for $H_\varepsilon(\bar{\mathscr{B}}_\Omega^{'})$.

Let $\mathscr{L}_\varepsilon$ be a minimal ε-covering subset of $\bar{\mathscr{B}}_\Omega^{'}$ and $L_\varepsilon(\bar{\mathscr{B}}_\Omega^{'})$ be the number of elements in $\mathscr{L}_\varepsilon$. By applying Lemma 3, we have

$$L_\varepsilon(\bar{\mathscr{B}}_\Omega^{'}) \le \frac{4eN^{3/2}\left(\frac{\sqrt{E}}{\varepsilon}\right)^N}{\ln N - 2}\left[N\ln N + o(N\ln N)\right], \tag{2.104}$$

for $N \ge 9$ and $1 < \frac{\sqrt{E}}{\varepsilon} < \frac{N}{\ln N}$. By taking the logarithm, we have

$$\begin{aligned} H_\varepsilon(\bar{\mathscr{B}}_\Omega^{'}) \le N\log\left(\frac{\sqrt{E}}{\varepsilon}\right) \\ + \log\left(\frac{4eN^{3/2}}{\ln N - 2}\left[N\ln N + o(N\ln N)\right]\right). \end{aligned} \tag{2.105}$$

Letting $\eta(N)$ be equal to the second term of (2.105) the result follows. □

2.5.2 Main theorems for $\mathscr{B}_\Omega$

We now extend results to the full space $\mathscr{B}_\Omega$. We define the auxiliary set

$$\underline{\mathscr{B}_\Omega} = \left\{\mathbf{b} = (b_1, \cdots, b_N, 0, 0, \cdots) : \sum_{n=1}^{N} b_n^2 \le E\right\} \tag{2.106}$$

whose norm is defined by

$$\|\mathbf{b}\| = \sqrt{\sum_{n=1}^{\infty} b_n^2 \lambda_n}. \tag{2.107}$$

We also use another auxiliary set

$$\mathscr{B}_\Omega^{''} = \left\{\mathbf{b} = (b_1, b_2, \cdots, b_{N'}) : \sum_{n=1}^{N'} b_n^2 \le E\right\} \tag{2.108}$$

equipped with the norm

$$\|\mathbf{b}\|'' = \sqrt{\sum_{n=1}^{N'} b_n^2 \lambda_n}. \tag{2.109}$$

where $N' = (1+\alpha)N_0$ for an arbitrary $\alpha > 0$.

Theorem 4. *For any $\varepsilon > 0$, we have*

$$\begin{cases} \bar{C}_\varepsilon^0(\mathscr{B}_\Omega) \geq \dfrac{\Omega}{\pi}\left(\log\sqrt{\mathsf{SNR}_K} - 1\right) & (2.110) \\ \bar{C}_\varepsilon^0(\mathscr{B}_\Omega) \leq \dfrac{\Omega}{\pi}\log\left(1+\sqrt{\mathsf{SNR}_K/2}\right), & (2.111) \end{cases}$$

where $\mathsf{SNR}_K = E/\varepsilon^2$.

Proof: By the continuity of the logarithmic function, to prove the upper bound it is enough to show that for any $\varepsilon > \mu > 0$

$$\bar{C}_\varepsilon^0(\mathscr{B}_\Omega) \leq \frac{\Omega}{\pi}\left[\log\left(1+\frac{\sqrt{E}}{(\varepsilon-\mu)\sqrt{2}}\right)\right], \tag{2.112}$$

and in order to prove (2.110) and (2.112), it is enough to show the following inequalities for the 2ε-capacity: for any $\varepsilon > \mu > 0$

$$C_\varepsilon^0(\mathscr{B}_\Omega) \geq C_\varepsilon^0(\mathscr{B}'_\Omega), \tag{2.113}$$

$$C_\varepsilon^0(\mathscr{B}_\Omega) \leq C_{\varepsilon-\mu}^0(\mathscr{B}'_\Omega), \tag{2.114}$$

and then apply Theorem 1.

Lower bound. Let $\mathscr{D}$ be a maximal $(\varepsilon,0)$-distinguishable subset of $\mathscr{B}_\Omega$ whose cardinality is $2^{C_\varepsilon^0(\mathscr{B}_\Omega)}$. Similarly, let $\mathscr{E}$ be a maximal $(\varepsilon,0)$-distinguishable subset of $\underline{\mathscr{B}}_\Omega$ whose cardinality is $2^{C_\varepsilon^0(\underline{\mathscr{B}}_\Omega)}$. Note that $\mathscr{E}$ is also a $(\varepsilon,0)$-distinguishable subset of $\mathscr{B}_\Omega$. Thus, we have

$$2^{C_\varepsilon^0(\underline{\mathscr{B}}_\Omega)} = |\mathscr{E}| \leq |\mathscr{D}| = 2^{C_\varepsilon^0(\mathscr{B}_\Omega)}. \tag{2.115}$$

From which it follows that

$$C_\varepsilon^0(\underline{\mathscr{B}}_\Omega) \leq C_\varepsilon^0(\mathscr{B}_\Omega). \tag{2.116}$$

Since $C_\varepsilon^0(\underline{\mathscr{B}}_\Omega) = C_\varepsilon^0(\mathscr{B}_\Omega^{'})$, the result follows.

Upper bound. For any $\varepsilon > \mu > 0$, we consider a projection map $\beta_\mu : \mathscr{B}_\Omega \to \underline{\mathscr{B}}_\Omega$. Let $\mathscr{D}$ be a maximal $(\varepsilon, 0)$-distinguishable subset of $\mathscr{B}_\Omega$ whose cardinality is $2^{C_\varepsilon^0(\mathscr{B}_\Omega)}$. Similarly, let $\mathscr{E}$ be a maximal $(\varepsilon-\mu, 0)$-distinguishable subset of $\underline{\mathscr{B}}_\Omega$ whose cardinality is $2^{C_{\varepsilon-\mu}^0(\underline{\mathscr{B}}_\Omega)}$.

We define $\mathscr{E}' = \beta_\mu(\mathscr{D})$. In general, β_μ is not one-to-one correspondence, however $|\mathscr{D}| = |\mathscr{E}'|$. If this is not the case, then there exist a pair of points $\mathbf{b}^{(1)}, \mathbf{b}^{(2)} \in \mathscr{D}$ satisfying $\beta_\mu(\mathbf{b}^{(1)}) = \beta_\mu(\mathbf{b}^{(2)}) = \mathbf{a}$, and we have

$$\begin{aligned} \|\mathbf{b}^{(1)} - \mathbf{b}^{(2)}\| &= \|\mathbf{b}^{(1)} - \mathbf{a} + \mathbf{a} - \mathbf{b}^{(2)}\| \\ &\leq \|\mathbf{b}^{(1)} - \mathbf{a}\| + \|\mathbf{a} - \mathbf{b}^{(2)}\| \\ &\leq \mu + \mu \\ &\leq 2\varepsilon, \end{aligned} \tag{2.117}$$

which is a contradiction. Thus, we have

$$2^{C_\varepsilon^0(\mathscr{B}_\Omega)} = |\mathscr{D}| = |\mathscr{E}'|. \tag{2.118}$$

The distance between any pair of points in $\mathscr{E}'$ exceeds $2(\varepsilon - \mu)$. If this is not the case, then there exist a pair of points in $\mathscr{E}'$ whose distance is smaller than $2(\varepsilon - \mu)$. These two point can be represented by $\mathbf{a}^{(1)} = \beta_\mu(\mathbf{b}^{(1)})$ and $\mathbf{a}^{(2)} = \beta_\mu(\mathbf{b}^{(2)})$, where $\mathbf{b}^{(1)}, \mathbf{b}^{(2)} \in \mathscr{D}$.

It follows that

$$\begin{aligned}\|\mathbf{b}^{(1)}-\mathbf{b}^{(2)}\| &= \|\mathbf{b}^{(1)}-\mathbf{a}^{(1)}+\mathbf{a}^{(1)}-\mathbf{a}^{(2)}+\mathbf{a}^{(2)}-\mathbf{b}^{(2)}\| \\ &\leq \|\mathbf{b}^{(1)}-\mathbf{a}^{(1)}\|+\|\mathbf{a}^{(1)}-\mathbf{a}^{(2)}\| \\ &\quad +\|\mathbf{a}^{(2)}-\mathbf{b}^{(2)}\| \\ &\leq \mu+2(\varepsilon-\mu)+\mu \\ &\leq 2\varepsilon, \end{aligned} \tag{2.119}$$

which is a contradiction. Thus, $\mathscr{E}'$ is a $(\varepsilon-\mu,0)$-distinguishable subset of $\underline{\mathscr{B}}_\Omega$, and we have

$$|\mathscr{E}'| \leq |\mathscr{E}| = 2^{C^0_{\varepsilon-\mu}(\underline{\mathscr{B}}_\Omega)}. \tag{2.120}$$

By combining (2.118) and (2.120), we obtain

$$2^{C^0_\varepsilon(\mathscr{B}_\Omega)} = |\mathscr{D}| = |\mathscr{E}'| \leq |\mathscr{E}| = 2^{C^0_{\varepsilon-\mu}(\underline{\mathscr{B}}_\Omega)}. \tag{2.121}$$

From which it follows that

$$C^0_\varepsilon(\mathscr{B}_\Omega) \leq C^0_{\varepsilon-\mu}(\underline{\mathscr{B}}_\Omega). \tag{2.122}$$

Since $C^0_{\varepsilon-\mu}(\underline{\mathscr{B}}_\Omega) = C^0_{\varepsilon-\mu}(\mathscr{B}'_\Omega)$, the result follows. □

Theorem 5. *For any* $0<\delta<1$ *and* $\varepsilon>0$*, we have*

$$\begin{cases} \bar{C}^\delta_\varepsilon(\mathscr{B}_\Omega) \geq \dfrac{\Omega}{\pi}\log\left(\sqrt{\mathsf{SNR}_K}\right) & (2.123) \\ \bar{C}^\delta_\varepsilon(\mathscr{B}_\Omega) \leq \dfrac{\Omega}{\pi}\log\left(1+\sqrt{\mathsf{SNR}_K}\right), & (2.124) \end{cases}$$

where $\mathsf{SNR}_K = E/\varepsilon^2$.

Proof: In this case, while the lower bound follows from a corresponding inequality on the (ε,δ)-capacity, the upper bound follows from an approximation argument and holds for the (ε,δ)-capacity per unit time only.

Lower bound. Let $\mathscr{E}$ be a maximal (ε,δ')-distinguishable subset of $\mathscr{B}_\Omega^{'}$ whose cardinality is $2^{C_\varepsilon^{\delta'}(\mathscr{B}_\Omega^{'})}$. We define a map $\alpha : \mathscr{B}_\Omega^{'} \to \mathscr{B}_\Omega$ such that, for $\mathbf{b} = (b_1,\cdots,b_N) \in \mathscr{B}_\Omega^{'}$, we have

$$\alpha(\mathbf{b}) = (b_1,\cdots,b_N,0,0,\cdots) \in \mathscr{B}_\Omega. \tag{2.125}$$

It follows that $\alpha(\mathscr{E})$ is a (ε,δ'')-distinguishable subset of $\mathscr{B}_\Omega$, where $0 \le \delta'' \le 1$ and δ'' tends to zero as $\delta' \to 0$. For any $\delta > 0$, we can now choose δ' so small that $\delta'' < \delta$. In this case, we have

$$2^{C_\varepsilon^{\delta'}(\mathscr{B}_\Omega^{'})} = |\mathscr{E}| \le 2^{C_\varepsilon^{\delta''}(\mathscr{B}_\Omega)}. \tag{2.126}$$

Also, since $\delta'' < \delta$, we have

$$C_\varepsilon^{\delta''}(\mathscr{B}_\Omega) \le C_\varepsilon^{\delta}(\mathscr{B}_\Omega). \tag{2.127}$$

By combining (2.126) and (2.127), we obtain

$$C_\varepsilon^{\delta'}(\mathscr{B}_\Omega^{'}) \le C_\varepsilon^{\delta}(\mathscr{B}_\Omega). \tag{2.128}$$

The result now follows from Theorem 2.

Upper bound. We define

$$d(\mathscr{B}_\Omega^{''},\mathscr{B}_\Omega) = \sup_{f\in\mathscr{B}_\Omega} \inf_{g\in\mathscr{B}_\Omega^{''}} \|f-g\| \tag{2.129}$$

which is a measure of distance between $\mathscr{B}_\Omega^{''}$ and $\mathscr{B}_\Omega$. From the Property 6 of the PSWF, we have

$$d(\mathscr{B}_\Omega^{''},\mathscr{B}_\Omega) \to 0 \quad \text{as} \quad N_0 \to \infty. \tag{2.130}$$

which implies

$$\bar{C}_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega}) = \bar{C}_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega}^{''}). \tag{2.131}$$

Thus, in order to prove the upper bound of $\bar{C}_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega})$, it is sufficient to derive an upper bound for $\bar{C}_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega}^{''})$.

By using the same proof technique as the one in Theorem 2, we obtain

$$C_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega}^{''}) \leq N' \left[\log\left(1+\frac{\sqrt{E}}{\varepsilon}\right)\right] + \log\frac{1}{1-\delta} \tag{2.132}$$

which implies

$$\bar{C}_{\varepsilon}^{\delta}(\mathscr{B}_{\Omega}^{''}) \leq (1+\alpha)\frac{\Omega}{\pi}\left[\log\left(1+\frac{\sqrt{E}}{\varepsilon}\right)\right]. \tag{2.133}$$

Since α is an arbitrary positive number, the result follows. □

Theorem 6. *For any $\varepsilon > 0$, we have*

$$\bar{H}_{\varepsilon}(\mathscr{B}_{\Omega}) = \frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K}, \tag{2.134}$$

where $\mathsf{SNR}_K = E/\varepsilon^2$.

Proof: By the continuity of the logarithmic function, to prove the result it is enough to show that for any $\varepsilon > \mu > 0$

$$\bar{H}_{\varepsilon}(\mathscr{B}_{\Omega}) \geq \frac{\Omega}{\pi}\left[\log\left(\frac{\sqrt{E}}{\varepsilon}\right)\right], \tag{2.135}$$

$$\bar{H}_{\varepsilon}(\mathscr{B}_{\Omega}) \leq \frac{\Omega}{\pi}\left[\log\left(\frac{\sqrt{E}}{\varepsilon-\mu}\right)\right], \tag{2.136}$$

and in order to prove (2.135) and (2.136), it is enough to show the following inequalities for the ε-entropy: for any $\varepsilon > \mu > 0$

$$H_\varepsilon(\mathscr{B}_\Omega) \geq H_\varepsilon(\mathscr{B}_\Omega^{'}) \tag{2.137}$$

$$H_\varepsilon(\mathscr{B}_\Omega) \leq H_{\varepsilon-\mu}(\mathscr{B}_\Omega^{'}), \tag{2.138}$$

and then apply Theorem 3.

Lower bound. For any $\varepsilon > \mu > 0$, we consider a projection map $\beta_\mu : \mathscr{B}_\Omega \to \underline{\mathscr{B}}_\Omega$. Let $\mathscr{D}$ be a minimal ε-covering subset of $\mathscr{B}_\Omega$ whose cardinality is $2^{H_\varepsilon(\mathscr{B}_\Omega)}$. Similarly, let $\mathscr{E}$ be a minimal ε-covering subset of $\underline{\mathscr{B}}_\Omega$ whose cardinality is $2^{H_\varepsilon(\underline{\mathscr{B}}_\Omega)}$.

We define $\mathscr{E}' = \beta_\mu(\mathscr{D})$. We claim that $\mathscr{E}'$ is also a ε-covering subset of $\underline{\mathscr{B}}_\Omega$. Let $\mathbf{p}$ be a point of $\underline{\mathscr{B}}_\Omega$. Since $\mathscr{D}$ is an ε-covering subset of $\mathscr{B}_\Omega$ and $\underline{\mathscr{B}}_\Omega \subset \mathscr{B}_\Omega$, there exists a point $\mathbf{b} \in \mathscr{D}$ such that $\|\mathbf{b} - \mathbf{p}\| \leq \varepsilon$. Note that $\|\beta_\mu(\mathbf{b}) - \mathbf{p}\| \leq \|\mathbf{b} - \mathbf{p}\|$ and $\beta_\mu(\mathbf{b}) \in \mathscr{E}'$. This means that, for any point $\mathbf{p} \in \underline{\mathscr{B}}_\Omega$, there exists a point in $\mathscr{E}'$ whose distance from $\mathbf{p}$ is equal or less than ε, which implies $\mathscr{E}'$ is a ε-covering subset of $\underline{\mathscr{B}}_\Omega$. Thus, we have

$$|\mathscr{E}'| \geq |\mathscr{E}| = 2^{H_\varepsilon(\underline{\mathscr{B}}_\Omega)}. \tag{2.139}$$

Since $|\mathscr{D}| \geq |\mathscr{E}'|$, we obtain the following chain of inequalities:

$$2^{H_\varepsilon(\mathscr{B}_\Omega)} = |\mathscr{D}| \geq |\mathscr{E}'| \geq |\mathscr{E}| = 2^{H_\varepsilon(\underline{\mathscr{B}}_\Omega)}. \tag{2.140}$$

From which it follows that

$$H_\varepsilon(\underline{\mathscr{B}}_\Omega) \leq H_\varepsilon(\mathscr{B}_\Omega). \tag{2.141}$$

Since $H_\varepsilon(\underline{\mathscr{B}}_\Omega) = H_\varepsilon(\mathscr{B}_\Omega^{'})$, the result follows.

Upper bound. Let $\mathscr{D}$ be a minimal ε-covering subset of $\mathscr{B}_\Omega$ whose cardinality is $2^{H_\varepsilon(\mathscr{B}_\Omega)}$. Similarly, let $\mathscr{E}$ be a minimal $(\varepsilon - \mu)$-covering subset of $\underline{\mathscr{B}}_\Omega$ whose cardinality is $2^{H_{\varepsilon-\mu}(\underline{\mathscr{B}_\Omega})}$.

We claim that $\mathscr{E}$ is also an ε-covering subset of $\mathscr{B}_\Omega$. Let $\mathbf{p}$ be a point of $\mathscr{B}_\Omega$. Since $\mathscr{E}$ is an $(\varepsilon-\mu)$-covering subset of $\underline{\mathscr{B}_\Omega}$ and $\beta_\mu(\mathbf{p}) \in \underline{\mathscr{B}_\Omega}$, there exists a point $\mathbf{a} \in \mathscr{E}$ such that $\|\mathbf{a}-\beta_\mu(\mathbf{p})\| \leq \varepsilon-\mu$. Then,

$$\begin{aligned}
\|\mathbf{a}-\mathbf{p}\| &= \|\mathbf{a}-\beta_\mu(\mathbf{p})+\beta_\mu(\mathbf{p})-\mathbf{p}\| \\
&\leq \|\mathbf{a}-\beta_\mu(\mathbf{p})\|+\|\beta_\mu(\mathbf{p})-\mathbf{p}\| \\
&\leq \varepsilon-\mu+\mu \\
&= \varepsilon. \qquad (2.142)
\end{aligned}$$

This means that, for any point $\mathbf{p} \in \mathscr{B}_\Omega$, there exists a point in $\mathscr{E}$ whose distance from $\mathbf{p}$ is equal or less than ε, which implies $\mathscr{E}$ is an ε-covering subset of $\mathscr{B}_\Omega$. Thus, we have

$$2^{H_{\varepsilon-\mu}(\underline{\mathscr{B}_\Omega})} = |\mathscr{E}| \geq |\mathscr{D}| = 2^{H_\varepsilon(\mathscr{B}_\Omega)}. \qquad (2.143)$$

From which it follows that

$$H_{\varepsilon-\mu}(\underline{\mathscr{B}_\Omega}) \geq H_\varepsilon(\mathscr{B}_\Omega). \qquad (2.144)$$

Since $H_{\varepsilon-\mu}(\underline{\mathscr{B}_\Omega}) = H_{\varepsilon-\mu}(\mathscr{B}'_\Omega)$, the result follows. $\square$

2.6 Appendix

2.6.1 Comparison with Jagerman's results

A basic relationship between ε-entropy and ε-capacity, given in [3], is

$$C_{2\varepsilon}(\mathscr{A}) \leq H_\varepsilon(\mathscr{A}). \qquad (2.145)$$

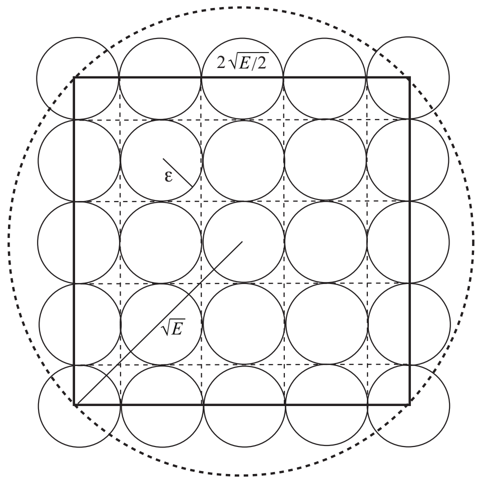

Figure 2.4. Lattice packing in Jagerman's lower bound.

It follows that a typical technique to estimate entropy and capacity is to find a lower bound for $C_{2\varepsilon}$ and an upper bound for H_ε, and if these are close to each other, then they are good estimates for both capacity and entropy.

Following this approach, Jagerman provided a lower bound on the 2ε-capacity and an upper bound on the ε-entropy of bandlimited functions. In our notation, the lower bound [18, Theorem 6] can be written as

$$C_{2\varepsilon} \geq N_0 \log\left(\frac{2}{\sqrt{10}}\sqrt{\frac{\mathsf{SNR}_K}{N_0}}+1\right), \tag{2.146}$$

where the result is adapted here to real signals.

Jagerman's proof roughly follows the codebook construction corresponding to the lattice packing depicted in Figure 2.4. In higher dimensions the side length of the hypercube corresponding to the square in Figure 2.4 becomes $2\sqrt{E/N_0}$, which divided by the diameter 2ε of the noise sphere gives the leading term $\sqrt{\mathsf{SNR}_K/N_0}$ inside the logarithm. The precise result requires a more detailed analysis of the asymptotic dimensionality of the space. This lower bound becomes very loose as $N_0 \to \infty$. In this case, by using the Taylor expansion of $\log(1+x)$ for x near zero in (2.146), it follows that $C_{2\varepsilon}$ grows only as $\sqrt{N_0}$ and, as a consequence, we have the trivial lower bound on the 2ε-capacity per unit time

$$\bar{C}_{2\varepsilon} \geq 0. \tag{2.147}$$

Geometrically, this is due to the volume of the high-dimensional sphere tending to concentrate on its boundary. For this reason, the packing in the inscribed hypercube in Figure 2.4 captures only a vanishing fraction of the volume available in the sphere. In contrast, our lower bound in Theorem 1 is non-constructive, and it gives the correct scaling order of the number of bits that can be reliably communicated over the channel, namely N_0 rather than $\sqrt{N_0}$, yielding a non-trivial lower bound on the 2ε-capacity per unit time.

In the same paper, Jagerman derives an upper bound on the ε-entropy [18, Theorem 8] by applying Mitjagin's theorem [31], which relates entropy to the Kolmogorov N-width. This standard technique is also illustrated in [32, Theorem 8]. For bandlimited signals, Jagerman further improves Mityagin's bound in a subsequent paper [19, Theorem 1], obtaining in our notation

$$H_\varepsilon \leq N \log\left(\frac{2\sqrt{E}}{\varepsilon-\mu} + \frac{\varepsilon+\mu}{\varepsilon-\mu}\right), \tag{2.148}$$

where $0 < \mu < \varepsilon$ and N is defined in (2.45). Since μ is an arbitrary positive number, (2.148) can be approximated by

$$H_\varepsilon \leq N \log\left(2\sqrt{\mathsf{SNR}_K} + 1\right). \tag{2.149}$$

The ε-entropy per unit time is then bounded as

$$\bar{H}_\varepsilon \leq \frac{\Omega}{\pi}\log(2\sqrt{\mathsf{SNR}_K}+1). \tag{2.150}$$

By combining (2.145),(2.147) and (2.150), Jagerman obtains

$$0 \leq \bar{H}_\varepsilon \leq \frac{\Omega}{\pi}\log\left(2\sqrt{\mathsf{SNR}_K}+1\right), \tag{2.151}$$

while we provide a tight characterization of the same quantity in Theorem 6 of this paper. If we use this tight result to bound the 2ε-capacity using the classic approach of (2.145), we obtain

$$\bar{C}_{2\varepsilon} \leq \frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K}, \tag{2.152}$$

while our direct bounds yield, for high values of SNR_K,

$$\begin{cases} \bar{C}_{2\varepsilon} \geq \frac{\Omega}{\pi}\left(\log\sqrt{\mathsf{SNR}_K}-1\right) & (2.153) \\ \bar{C}_{2\varepsilon} \leq \frac{\Omega}{\pi}\left(\log\sqrt{\mathsf{SNR}_K}-1/2\right), & (2.154) \end{cases}$$

These are essentially the same bounds obtained by Wyner [20] in a slightly different context.

2.6.2 Relationship with Nair's work

Nair defined the peak maximum information rate R_* in [13, Lemma 4.2] and showed R_* equals the zero-error capacity [13, Theorem 4.1]. In his paper, Nair defined R_* for a discrete time channel, but this definition can be modified for a continuous time channel as follows:

$$R_* = \lim_{T\to\infty}\sup_{X:X\subset\mathscr{B}_\Omega}\frac{I_*(X;Y)}{T}, \tag{2.155}$$

where Y is the uncertain output signal yielded by X.

When we consider our channel model, it is clear that the supremum is achieved when X is a maximal 2ε-distinguishable set, $\mathscr{M}_{2\varepsilon}$. In this case, $I_*(X;Y) = \log|X| = \log M_{2\varepsilon}(\mathscr{B}_\Omega)$. Thus (2.155) can be rewritten as follows:

$$R_* = \lim_{T\to\infty} \frac{\log M_{2\varepsilon}(\mathscr{B}_\Omega)}{T}. \tag{2.156}$$

The right-hand side of (2.156) is the definition of $\bar{C}_{2\varepsilon}(\mathscr{B}_\Omega)$. Thus, we conclude that $\bar{C}_{2\varepsilon}(\mathscr{B}_\Omega)$ is a peak maximum information rate and equals the zero-error capacity in our setting.

2.6.3 Derivation of the error exponent

By (2.93), we have

$$\Delta = P_{err} \leq M\left(\frac{\varepsilon}{\zeta(N)\sqrt{E}}\right)^N. \tag{2.157}$$

Let $M = 2^{TR}$, where the transmission rate R is smaller than the lower bound on $\bar{C}_\varepsilon^\delta$. Then, (2.157) can be rewritten as

$$\Delta = P_{err} \leq 2^{-T\left[\frac{N}{T}\log\left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right)-R\right]}. \tag{2.158}$$

In a stochastic setting the error exponent is defined as the logarithm of the error probability. It follows that we may also define the error exponent in our deterministic model

$$\mathsf{Er}(R) = \frac{N}{T}\log\left(\zeta(N)\frac{\sqrt{E}}{\varepsilon}\right) - R. \tag{2.159}$$

Since N/T tends to Ω/π and $\zeta(N)$ tends to 1 as $T \to \infty$, we can approximate the error exponent when N_0 is sufficiently large by

$$\mathsf{Er}(R) = \frac{\Omega}{\pi} \log\left(\frac{\sqrt{E}}{\varepsilon}\right) - R. \tag{2.160}$$

2.7 Acknowledgments

Chapter 2, in full, is a reprint of the material as it appears in T. J. Lim and M. Franceschetti, "Information without rolling dice," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1349-1363, 2017. The dissertation author was the primary investigator and co-author of these papers.

Chapter 3

Deterministic coding theorems for blind sensing: optimal measurement rate and fractal dimension

3.1 Introduction

3.1.1 Problem set-up

Let $f : \mathbb{R} \to \mathbb{R}$ be square-integrable and such that

$$\mathfrak{F}f(\omega) = 0, \text{ for } \omega \notin \mathscr{Q}, \tag{3.1}$$

where $\mathfrak{F}$ indicates Fourier transform, ω indicates angular frequency, and $\mathscr{Q}$ is a subset of the interval $[-\Omega, \Omega]$ of measure

$$m(\mathscr{Q}) \leq 2\Omega'. \tag{3.2}$$

A typical example occurs when $\mathscr{Q}$ is the union of a finite number of disjoint sub-intervals of $[-\Omega, \Omega]$ and $\Omega' \ll \Omega$, see Figure 3.1.

These kind of signals arise in many applications, ranging from radio, to audio, and biological communication and sensing systems. A natural question is what is the minimum number of measurements that can be performed over a given time interval and

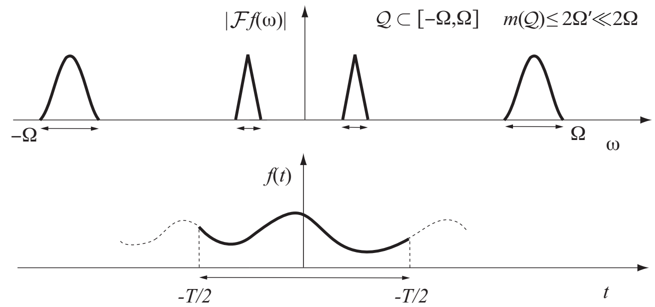

Figure 3.1. Illustration of a sparse multi-band signal observed over a single time interval.

that guarantees reconstruction with a minimum amount of error.

To address this question, we consider a measurement vector $\mathbf{y} \in \mathbb{R}^M$

$$\mathbf{y} = \mathfrak{M} f(t) + \mathbf{e}, \tag{3.3}$$

where $\mathfrak{M}$ is an operator from multi-band signals to M-dimensional vectors and $\mathbf{e} \in \mathbb{R}^M$ is the measurement error.

We assume each measurement $y_n \in \mathbf{y}$ results from observing the signal over the interval $[-T/2, T/2]$ through the inner product with a bandlimited kernel, plus an error term.

Definition 6. (Measurements) *For all $n \in \{1, \ldots, M\}$, we have*

$$y_n = \int_{-T/2}^{T/2} f(t)\varphi_n(t)dt + e_n, \tag{3.4}$$

where

$$\mathscr{F}\varphi_n(\omega) = 0 \, \mathit{for} \; \omega \notin [-\Omega, \Omega]. \tag{3.5}$$

This set-up covers a wide range of real measurements. Possible bandlimited ker-

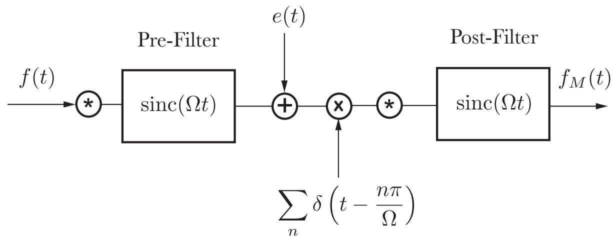

Figure 3.2. Block diagram for sampling measurement and reconstruction. The symbol $*$ indicates convolution.

nels that fall in this framework include the Shannon cardinal basis $\text{sinc}(\cdot)$ functions [33], the Slepian prolate spheroidal wave functions (PSWF) [11], as well as other bandlimited functions of practical interest, such as wavelets, and splines. The measurements are functionals of the signal over the entire observation interval, but in some cases they can reduce to the sampled signal values. For example, for the cardinal basis the measurements in (3.4) also correspond to low-pass filtering and sampling, and the signal can be recovered by low-pass filtering the sampled signal values [34]. This special case is illustrated in Figure 3.2. The general case is illustrated in Figure 3.3.

In the general setting, our aim is to determine the smallest measurement rate

$$\bar{M} = \lim_{T \to \infty} \frac{M}{T} \tag{3.6}$$

for which it is possible to obtain an approximation f_M of f from $\mathbf{y}$, such that the energy of the reconstruction error is at most proportional to the energy of the measurement error, as the size of the observation interval $T \to \infty$. This corresponds to determining the scaling of the minimum number of measurements $M = M(T)$ that guarantees *robust recovery* of any multi-band signal, namely a small perturbation in the measurement does not lead to a large reconstruction error.

Definition 7. (Robust recovery). *There exists a universal constant $c \geq 0$, such that for T*

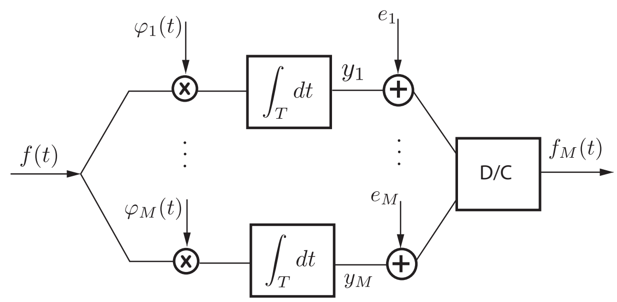

Figure 3.3. Block diagram for general measurement and reconstruction. The box D/C stands for discrete-to-continuous transformation and performs the reconstruction of the signal from the discrete measurements.

large enough

$$\begin{aligned} \|f - f_M\|^2 &= \int_{-T/2}^{T/2} [f(t) - f_M(t)]^2 dt \\ &\leq c \sum_{n=1}^{M} e_n^2. \end{aligned} \tag{3.7}$$

When the measurement error tends zero, robust recovery reduces to *perfect recovery* of the signal. Namely,

Definition 8. (Perfect recovery).

$$\lim_{T\to\infty} \|f - f_M\|^2 = 0. \tag{3.8}$$

3.1.2 Bandlimited signals

Since our signals are assumed to be bandlimited to Ω, one may readily observe that in the absence of measurement error they can be perfectly recovered from a number of measurements slightly above the Nyquist number

$$N_0 = \Omega T/\pi. \tag{3.9}$$

For any f satisfying (3.1) and (3.2), and $\nu > 0$, we can construct an approximation f_N of f from a measurement vector $\mathbf{y}$ of size

$$N = (1+\nu)\Omega T/\pi, \tag{3.10}$$

and such that

$$\lim_{T\to\infty} \|f - f_N\|^2 = 0. \tag{3.11}$$

This classic result is equivalent to stating that a measurement rate strictly above Ω/π is sufficient for reconstruction of any bandlimited singnal, and constitutes one of the milestones of electrical and communication engineering.

For bandlimited signals, the rate Ω/π is also optimal, in the following approximation theoretic sense. Consider performing signal reconstruction by a linear interpolation of a number $N > 0$ of orthogonal basis functions

$$f_N(t) = \sum_{n=1}^{N} y_n \varphi_n(t), \tag{3.12}$$

and let the Kolmogorov N-width be the smallest approximation error achievable for all signals in the space, over all possible choices of basis sets. This minimum error is achieved by measurements that provide the coefficients of the interpolation through the integrals

$$y_n = \frac{1}{\lambda_n} \int_{-T/2}^{T/2} f(t)\varphi_n(t)dt, \quad n \in \{1,\ldots,N\}, \tag{3.13}$$

where $\{\lambda_n\}$ are the eigenvalues of a Fredholm integral equation of the second kind arising from Slepian's concentration problem [11], and the basis functions $\{\varphi_n\}$ are the corresponding eigenfunctions, called PWSF [22]. The measurement rate Ω/π corresponds to the critical threshold at which the Kolmogorov N-width transitions from strictly positive values to zero, as $T \to \infty$ [25]. This phase transition behavior of the approximation error

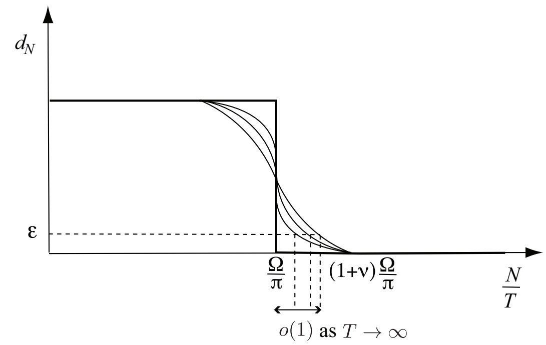

Figure 3.4. Phase transition of the Kolmogorov N-width d_N of bandlimited signals.

is illustrated in Figure 3.4. With a number of measurements $(1+\nu)\Omega T$ the error tends to zero as $T \to \infty$, while with a number of measurements $\Omega T/\pi + o(T)$ the error remains positive as $T \to \infty$.

3.1.3 Multi-band signals

For bandlilmtied signals that are supported over disjoint sub-bands, an important extension of the results above, due to Landau and Widom [23], states that if we have a priori knowledge of the size and positions of all the sub-bands, then signal reconstruction with vanishing error as $T \to \infty$ is also possible using the smaller number of measurements

$$S = (1+\nu)\Omega' T/\pi. \tag{3.14}$$

A simple way to achieve this result is to demodulate each sub-band down to baseband, isolate it through low-pass filtering, and then sample each sub-band separately. The key contribution of Landau and Widom is to consider the optimal subspace approximation, and show a phase transition of the error expressed in terms of Kolmogorov N-width. As in the single-band case, a subspace approximation with vanishing error for all multi-band signals of a given frequency allocation is obtained with a number of measurements $(1+\nu)\Omega' T$, while a subspace approximation with vanishing error is not possible for all multi-band signals using a number of measurements $\Omega' T/\pi + o(T)$, and the value of the error is controlled by the pre-constant in the $o(T)$ term. It follows that for multi-band signals the Nyquist number $N_0 = \Omega T/\pi$ can be replaced by the "sparsity number"

$$S_0 = \Omega' T/\pi, \tag{3.15}$$

and the *occupied portion* of the frequency bandwidth determines the critical measurement rate Ω'/π required for reconstruction. In the case of sampling measurements, Landau [35] also showed that a rate Ω'/π is necessary for reconstruction, regardless of the reconstruction strategy being linear or not.

The results of above rely on two critical assumptions. First, they need a priori knowledge of the spectral occupation, since the eigenvalues and the optimal eigenfunctions used for reconstruction are solutions of an integral equation that depends on the spectral support set. In practice, it might be difficult to know the exact number of sub-bands, their location, and their widths prior to the measurements. A second critical assumption is the absence of measurement error. In practice, the measurement process always carries a certain amount of error and its impact on the reconstruction error should be taken into account.

3.1.4 Completely blind sensing

In this paper, we consider robust signal reconstruction in the presence of measurement error and without any a priori knowledge of the sub-bands beside an upper bound on the measure of the whole support set of the signal in the frequency domain. We call this *robust, completely blind sensing.* The blindness requirement is important when detecting the sub-bands is impossible or too expensive to implement. The robustness requirement is important to guarantee stability in the reconstruction process.

Partially blind sensing, where some partial spectral information is assumed, has been studied extensively. First key results were given in a series of papers by Bresler and co-authors [36, 37, 38]. Later extensions [39, 40] reduced the number of a priori assumptions, but still require knowledge of the number of sub-bands, and of their widths. The same assumptions are made in [41, 42, 43]. The main result in this setting is that the price to pay for partial blindness is a factor of two in the measurement rate. Several reconstruction strategies have been proposed using a measurement rate above $2\Omega'/\pi$, all assuming some partial spectral knowledge, and lacking an information-theoretic converse.

We remove these assumptions, show that a measurement rate $2\Omega'/\pi$ is sufficient for robust reconstruction in a completely blind setting, and provide a tight converse result. We also provide a deterministic coding theorem for continuous analog sources, giving an interpretation of the minimum number of measurements in terms of the "effective" Minkowski-Boulingand dimension of the infinite-dimensional set of multi-band signals, expressed in terms of the Kolmogorov ε-entropy. This is compared with an analogous interpretation arising in the framework of compressed sensing, where the objective is the lossless source coding of a discrete, analog, stochastic process [44, 45]. In that case, an analogous coding theorem has been given in terms of the Rény dimension, expressed in terms of the Shannon entropy.

Finally, we remark that while in the case of multi-band signals of a given sub-band allocation the results of Landau and Widom provide an optimal subspace approximation in terms of a linear interpolation of eigenfunctions supported over multiple sub-bands, and having the highest energy concentration over the observation domain, our results only provide an answer to the question of whether recovery is possible or not, without giving an explicit approximation procedure. In our case, the discrete-to-continuous block in Figure 3.3 remains unknown.

Nevertheless, from an information-theoretic perspective one is primarily interested in the possibility of recovery using any discrete to continuous transformation, and does not wish to restrict reconstruction to a linear approximation strategy. The explicit construction of practical blind recovery strategies is certainly of interest, and these should be compared with the information-theoretic optimum determined here.

The rest of this chapter is organized as follows: In section 3.2 we describe our results. In section 3.3 we compare our results with compressed sensing and illustrate coding theorems in deterministic and stochastic settings. In section 3.4 we provide some definitions and preliminaries that are useful for our derivation. Proofs are given in section 3.5 and 3.6. Section 3.7 draws conclusions and discusses future work.

3.2 Description of the results

3.2.1 Noiseless Case

Theorem 7. (Direct). *In the absence of measurement error, we can perfectly recover any signal f satisfying (3.1) and (3.2) using a measurement rate*

$$\bar{M} > \frac{2\Omega'}{\pi}. \tag{3.16}$$

Theorem 8. (Converse). *In the absence of measurement error, we cannot perfectly recover all signals f satisfying (3.1) and (3.2) using a measurement rate*

$$\bar{M} \leq \frac{2\Omega'}{\pi}. \tag{3.17}$$

These results can interpreted in terms of the effective dimensionality of the signals' space, leading to a coding theorem. For bandlimited signals, the effective number of dimensions can be identified with the Nyquist number $N_0 = \Omega T/\pi$. For multi-band signals for which the location and widths of all the sub-bands is fixed a priori, as in the

Landau-Widom case, it can be identified with the sparsity number $S_0 = \Omega' T/\pi$. On the other hand, without any a priori knowledge, we need to account for the additional degrees of freedom of allocating the sub-bands in the frequency domain, and our results indicate that the effective dimensionality increases to $2S_0$.

To make these considerations precise, we consider an information-theoretic quantity that measures the dimensionality of a set in metric space, namely its fractal (Minkowski-Bouligand) dimension, which corresponds to the rate of growth of the Kolmogorov ε-entropy of successively finer discretizations of the space, and represents the degree of fractality of the set [46].

Definition 9. (Fractal dimension). *For any subset $\mathscr{X}$ of a metric space, the fractal dimension is*

$$\dim_F(\mathscr{X}) = \lim_{\varepsilon \to 0} \frac{H_\varepsilon(\mathscr{X})}{-\log \varepsilon}, \tag{3.18}$$

where H_ε is the Kolmogorov ε-entropy [3].

If this limit does not exist, then the corresponding upper and lower fractal dimensions are defined using lim sup and lim inf, respectively.

We also define the dilation

Definition 10. (Minkowski sum).

$$\mathscr{X} \oplus \mathscr{X} = \{\mathbf{x_1} + \mathbf{x_2} : \mathbf{x_1}, \mathbf{x_2} \in \mathscr{X}\}. \tag{3.19}$$

Consider now the set of all bandlimited signals whose energy is at most one. These signals can be approximated by an infinite set $\mathscr{X}_\mathrm{B}$ of vectors, each containing $N = (1+\nu)\Omega T/\pi$ real coefficients. Using the PSWF as a basis for interpolation, every assignment of coefficients satisfying the given energy constraint approximates, with vanishing error as $T \to \infty$, a bandlimited signal.

In the appendix, we show that

$$\dim_F(\mathscr{X}_{\mathrm{B}}) = \dim_F(\mathscr{X}_{\mathrm{B}} \oplus \mathscr{X}_{\mathrm{B}}), \tag{3.20}$$

and letting the fractal dimension rate of the approximating set be

$$R_F(\mathscr{X}_{\mathrm{B}}) = \lim_{T \to \infty} \frac{\dim_F(\mathscr{X}_{\mathrm{B}})}{T}, \tag{3.21}$$

we have

$$R_F(\mathscr{X}_{\mathrm{B}}) = \Omega/\pi, \tag{3.22}$$

which coincides with the measurement rate needed for reconstruction.

Next, we quantize the bandwidth at level $\Delta > 0$ and let

$$\mathscr{J} = \{-\Omega, -\Omega + \Delta, -\Omega + 2\Delta, \cdots, \Omega\}. \tag{3.23}$$

We consider the subset of all multi-band signals of a given sub-band allocation, whose energy is at most one, and such that the extremal points of all sub-bands belong to $\mathscr{J}$. This subset of signals approximates, with vanishing energy error as $\Delta \to 0$, the one of all multi-band signals of a given sub-band allocation and of energy at most one. It can also be approximated, with vanishing error as $T \to \infty$, by an infinite set $\mathscr{X}_{\mathrm{MB}}(\Delta)$ of vectors, each containing $N = (1+\nu)\Omega T/\pi$ real coefficients of a PSWF interpolation. Compared to the previous case, the choice of the coefficients is now restricted by the given sub-band allocation, so that we have

$$\mathscr{X}_{\mathrm{MB}}(\Delta) \subset \mathscr{X}_{\mathrm{B}}. \tag{3.24}$$

Following the same argument used to derive (3.20), we obtain

$$\dim_F[\mathscr{X}_{\mathrm{MB}}(\Delta)] = \dim_F[\mathscr{X}_{\mathrm{MB}}(\Delta) \oplus \mathscr{X}_{\mathrm{MB}}(\Delta)]. \tag{3.25}$$

In this case, however, the N-dimensional prolate spheroidal approximation is somewhat redundant, and following the same argument used to derive (3.22), we obtain

$$\lim_{\Delta\to 0} R_F[\mathscr{X}_{\mathrm{MB}}(\Delta)] = \Omega'/\pi, \tag{3.26}$$

which coincides with the Landau-Widom rate [23, 35] needed for reconstruction.

Finally, consider the subset of all multi-band signals whose energy is at most one, having an arbitrary sub-band allocation of measure at most $2\Omega'$, and such that the extremal points of all sub-bands belong to $\mathscr{J}$. These signals can be approximated, as $T \to \infty$, by an infinite set $\mathscr{X}(\Delta)$ of vectors, each containing $N = (1+\nu)\Omega T/\pi$ real coefficients of a PSWF interpolation. The choice of the coefficients is now restricted only by the measure of the occupied portion of the spectrum and not by a specific sub-band allocation, and we have

$$\mathscr{X}_{\mathrm{MB}}(\Delta) \subset \mathscr{X}(\Delta) \subset \mathscr{X}_{\mathrm{B}}. \tag{3.27}$$

By combining Theorems 7 and 8 with Theorems 9 and 10 below, we obtain

$$\lim_{\Delta\to 0} R_F[\mathscr{X}(\Delta)] = \Omega'/\pi. \tag{3.28}$$

Theorem 9. (Direct). *In the absence of measurement error, we can perfectly recover any signal f satisfying (3.1) and (3.2) using a measurement rate*

$$\bar{M} > 2\lim_{\Delta\to 0} R_F[\mathscr{X}(\Delta)]. \tag{3.29}$$

Theorem 10. (Converse). *In the absence of measurement error, we cannot perfectly recover all signals f satisfying (3.1) and (3.2) using a measurements rate*

$$\bar{M} \leq 2 \lim_{\Delta \to 0} R_F[\mathscr{X}(\Delta)]. \tag{3.30}$$

In section 3.4, we also show that

$$R_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)] = 2R_F[\mathscr{X}(\Delta)], \tag{3.31}$$

which also implies

$$\lim_{T \to \infty} \frac{\dim_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)]}{\dim_F[\mathscr{X}(\Delta)]} = 2. \tag{3.32}$$

We now give a geometric interpretation of these results. The set of all multi-band signals is the union of infinitely many subsets, each corresponding to the multi-band signals of a given sub-band allocation. The Minkowski sum in (3.19) takes into account the additional degrees of freedom of allocating the sub-bands in the frequency domain. Within any subset, any multi-band signal is specified by essentially $\dim_F[\mathscr{X}(\Delta)]$ coordinates, but when considering the union of all subsets, it is specified by essentially $2\dim_F[\mathscr{X}(\Delta)]$ coordinates. By (3.31) it then follows that the relevant information-theoretic quantity that characterizes the possibility of reconstruction is the fractal dimension rate of the dilation, rather than the fractal dimension rate of the set itself.

Finally, it is useful introduce the *sparsity fraction* as the ratio of the fractal dimension of the approximating set and its ambient dimension:

Definition 11. (Sparsity fraction).

$$\sigma = \inf_{\nu > 0} \lim_{\Delta \to 0} \lim_{T \to \infty} \frac{\dim_F[\mathscr{X}(\Delta)]}{N}. \tag{3.33}$$

By the results above, it is easy to see that the sparsity fraction is equal to the fraction of occupied bandwidth, namely substituting $N = (1+\nu)\Omega T/\pi$ into (3.33) we get

$$\sigma = \inf_{\nu>0} \lim_{\Delta\to 0} \frac{R_F[\mathscr{X}(\Delta)]}{\Omega} \frac{\pi}{(1+\nu)} = \frac{\Omega'}{\Omega}, \tag{3.34}$$

and twice the sparsity fraction corresponds to the critical number of measurements per unit ambient dimension necessary and sufficient for reconstruction.

3.2.2 General Case

Results generalize to the noisy case. The critical threshold for the number of measurements is not affected by the presence of a measurement error, provided that we ask for robust, rather than perfect reconstruction.

Theorem 11. (Direct). *We can robustly recover all signals f satisfying (3.1) and (3.2) using a measurements rate*

$$\bar{M} > 2 \lim_{\Delta\to 0} R_F[\mathscr{X}(\Delta)] = \frac{2\Omega'}{\pi}. \tag{3.35}$$

Theorem 12. (Converse). *We cannot robustly recover all signals f satisfying (3.1) and (3.2) using a measurements rate*

$$\bar{M} \leq 2 \lim_{\Delta\to 0} R_F[\mathscr{X}(\Delta)] = \frac{2\Omega'}{\pi}. \tag{3.36}$$

A factor of two is the price to pay for blindness for both robust recovery and perfect recovery of multi-band signals, and in virtue of (3.31) the relevant dimensionality notion is the one associated to the dilation of the set.

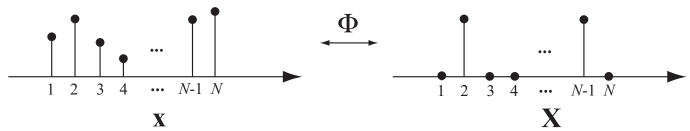

Figure 3.5. Illustration of a discrete vector with a sparse representation.

3.3 Comparison with compressed sensing

There are analogies between our results and the ones in compressed sensing. We illustrate similarities and differences in deterministic and stochastic settings. For simplicity, we only consider the case of zero measurement error, but the same considerations apply to the case of non-zero measurement error.

3.3.1 Deterministic setting

Consider an N-dimensional vector $\mathbf{x}$ such that

$$\mathbf{x} = \Phi\mathbf{X}, \tag{3.37}$$

where Φ is an $N \times N$ orthogonal matrix and $\mathbf{X}$ has at most S non-zero elements. If $S \ll N$ we say that $\mathbf{X}$ is a sparse representation of $\mathbf{x}$. An example is illustrated in Figure 3.5.

We define a measurement vector

$$\mathbf{y} = \mathrm{A}\mathbf{x}, \tag{3.38}$$

where A is an $M \times N$ matrix, and M is the number of measurements. Cleary, $\mathbf{x}$ can be recovered from N measurements by observing all the elements of $\mathbf{x}$. In this case, the $N \times N$ measurement matrix A is diagonal. If we know the position of the nonzero elements of $\mathbf{X}$, then S measurements are also enough to perfectly reconstruct $\mathbf{x}$. In this case, each measurement extracts the nth coefficient of $\mathbf{X}$ from $\Phi^{-1}\mathbf{x}$, and the signal is recovered by performing a final multiplication by Φ. However, if we only know that $\mathbf{x}$ has a sparse representation, but we do not know the positions of the nonzero elements of $\mathbf{X}$, without further investigation we can only conclude that that the minimum number M of measurements sufficient for reconstruction is $S \leq M \leq N$. The objective of compressed sensing is to reconstruct any sparse, discrete signal $\mathbf{x}$ using $M \ll N$ measurements [47].

Without worrying about an explicit reconstruction procedure, a simple linear algebra argument [44, Remark 2], [47, Section 2.2] shows that the necessary and sufficient number of measurements for reconstruction is $2S$. It follows that in both the continuous and discrete settings, the number of linear measurements necessary and sufficient for reconstruction is equal to twice the sparsity level of the signal. The main differences between the two settings are as follows: the compressed sensing formulation assumes knowledge of the matrix Φ, corresponding to the basis where the discrete signal is sparse. In the case of blind sensing, it is only assumed that the signal does not occupy the whole frequency spectrum, but the discrete basis set required for the optimal representation is unknown a priori. A more extreme situation is the blind compressed sensing set-up [48, 49], where there is a complete lack of knowledge about the signal. In this case, the basis must either be learned from data, or selected from a restricted set. Finally, in blind sensing the reconstruction error tends to zero as $T \to \infty$, while in compressed sensing perfect reconstruction is possible for all N.

3.3.2 Stochastic setting

The problem of compressed sensing can also be formulated in a probabilistic setting. In this case, the discrete signal to be recovered is modeled as a stochastic process and the objective is to reconstruct the signal with arbitrarily small probability of error, given a sufficiently long observation sequence. Viewing the measurement operator as an encoder and the reconstruction operator as a decoder acting on a sequence of independent,

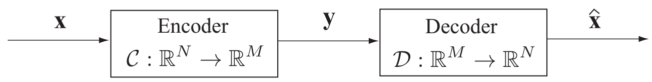

Figure 3.6. Source coding view of compressed sensing.

identically distributed (i.i.d.), real-valued random variables, the compressed sensing set-up corresponds to lossless source coding of analog memoryless sources when the encoding operation $\mathscr{C}: \mathbb{R}^N \to \mathbb{R}^M$ is the multiplication by a real-valued matrix, see Figure 3.6.

Compared to the deterministic setting, where reconstruction is required for all possible source signals, here the performance is measured on a probabilistic basis by considering long block lengths and averaging with respect to the distribution of the source signal. Compared to the continuous setting, probabilistic concentration is used to bound the error performance as $N \to \infty$, instead than spectral concentration as $T \to \infty$.

Modeling $\mathbf{x}$ in (3.38) as a random vector composed of N independent random variables $(\mathsf{X}_1, \mathsf{X}_2, \ldots \mathsf{X}_N)$, all distributed as X, to capture the notion of sparsity in a probabilistic setting we may consider the following mixture distribution for the source sequence

$$p_{\mathsf{X}}(x) = (1-\gamma)\delta(x) + \gamma p'(x), \tag{3.39}$$

where $\delta(\cdot)$ is Dirac's distribution, $0 \leq \gamma \leq 1$, and p' is an absolutely continuous probability measure [1].

By the law of large numbers, the parameter γ in (3.39) represents, for large values of N, the level of sparsity of the signal in terms of the fraction of its nonzero elements. Given this source model, a basic result for probabilistic reconstruction by Wu and Verdú [44, 45] shows that the threshold for the smallest measurement rate that

[1]Results hold more generally for discrete-continuous mixtures, not only when the discrete part is a Dirac's distribution.

guarantees reconstruction with vanishing probability of error is independent of the prior distribution of the non-zero elements p', and equals the sparsity level γ. Comparing this result with the deterministic case, it follows that probabilistic reconstruction, rather than reconstruction for all signals in the space, yields an improvement in the number of measurements of a factor of two.

Wu and Verdú also showed that their result can be viewed in terms of the infomation (Rényi) dimension of the source process. This is somewhat analogous to a coding theorem, where an operational quantity, such as the smallest rate for recovery, is shown to be equal to an information-theoretic one. Consider the quantized version X^{ε} of X obtained from the discrete probability measure induced by partitioning the real line into intervals of size ε and assigning to the quantized variable the probability of lying in each interval. The Rényi dimension of X is defined as [50]

Definition 12. (Information dimension).

$$\dim_I(\mathsf{X}) = \lim_{\varepsilon \to 0} \frac{H_{\mathsf{X}^{\varepsilon}}}{-\log \varepsilon}, \tag{3.40}$$

where $H_{\mathsf{X}^{\varepsilon}}$ indicates the Shannon entropy of X^{ε}.

In the case the limit in (3.40) does not exist, then lower and upper information dimensions are defined by taking lim inf and lim sup, respectively.

The definition immediately extends to a sequence of N i.i.d. random variables

$$\dim_I(\mathsf{X}_1, \mathsf{X}_2, \ldots \mathsf{X}_N) = N \dim_I(\mathsf{X}), \tag{3.41}$$

and should be compared with Definition 9 for continuous signals in a deterministic setting.

We can also give an information-theoretic definition of the sparsity fraction in the stochastic setting that is analogous to Definition 11.

Definition 13. (Sparsity fraction —stochastic setting).

$$\gamma = \frac{\dim_I(\mathsf{X}_1, \mathsf{X}_2, \ldots \mathsf{X}_N)}{N}. \tag{3.42}$$

For a mixture distribution such as (3.39), assuming $H(\lfloor \mathsf{X} \rfloor) < \infty$, Rényi showed that [50]

$$\dim_I(\mathsf{X}) = \gamma. \tag{3.43}$$

Combining this result with (3.41) it follows that the sparsity fraction is also equal to γ, and the fraction of non-zero elements of the signal coincides with the information dimension per unit ambient dimension. In the analogous deterministic setting, the fraction of occupied bandwidth plays the role of the fraction of non-zero elements of the discrete-time signal, and this coincides with the fractal dimension per unit ambient dimension of its prolate spheroidal approximation.

3.3.3 Coding theorems

The results of Wu and Verdú combined with Rényi's one in (3.43) yield the following general coding theorem:

Theorem 13. (Coding theorem —stochastic setting).
The minimum number of measurement per unit dimension sufficient for reconstruction with vanishing probability of error of an analog, γ-sparse, memoryless, discrete-time process coincides with the information dimension per unit ambient dimension of the space, which is equal to γ.

The analogous deterministic coding theorem in our continuous setting is obtained by combining Theorems 9 and 10, and using Definition 11:

Theorem 14. (Coding theorem —deterministic setting).
The minimum number of measurement per unit dimension sufficient for reconstruction with vanishing error of any σ-sparse, continuous-time signal coincides with twice the fractal dimension per unit ambient dimension of its prolate spheroidal approximation, which is equal to 2σ.

A factor of two appears in the deterministic formulation, due to the worst case reconstruction scenario.

3.4 Technical Preliminaries

3.4.1 Metric spaces

We begin our proofs by defining the metric spaces associated to the bandlimited and multi-band signals satisfying (3.1) and (3.2). Let $f \in L^2(-\infty,\infty)$, $2\Omega' < \Omega$, and

$$\mathscr{B}_\Omega = \{f(t) : \mathfrak{F}f(\omega) = 0, \text{ for } |\omega| > \Omega\}, \tag{3.44}$$

$$\mathscr{B}_\mathscr{Q} = \{f(t) : \mathfrak{F}f(\omega) = 0, \text{ for } \omega \notin \mathscr{Q}\}, \tag{3.45}$$

$$\mathscr{Q}' = \{\mathscr{Q} : \mathscr{Q} \subset [-\Omega,\Omega] \text{ and } m(\mathscr{Q}) \leq 2\Omega'\}, \tag{3.46}$$

$$\mathscr{B}_{\mathscr{Q}'} = \bigcup_{\mathscr{Q}\in\mathscr{Q}'} \mathscr{B}_\mathscr{Q}. \tag{3.47}$$

It follows that $\mathscr{B}_{\mathscr{Q}'} \subset \mathscr{B}_\Omega$. We equip $\mathscr{B}_\Omega$ and $\mathscr{B}_{\mathscr{Q}'}$ with the $L^2[-T/2,T/2]$ norm

$$\|f\| = \left(\int_{-T/2}^{T/2} f^2(t)dt\right)^{1/2}. \tag{3.48}$$

It follows that $(\mathscr{B}_\Omega, \|\cdot\|)$ and $(\mathscr{B}_{\mathscr{Q}'}, \|\cdot\|)$ are metric spaces, whose elements are square-integrable, real, bandlimited or multi-band signals, of infinite duration and observed over the finite interval $[-T/2,T/2]$.

3.4.2 Optimal representations

Let $\mathscr{Q}$ be a measurable subset of $\mathbb{R}$ and $\mathscr{T} = [-T/2, T/2]$. We define the following time-limiting and band-limiting operators

$$\mathfrak{T}_{\mathscr{T}} f(t) = \mathbb{1}_{\mathscr{T}} f(t) \tag{3.49}$$

$$\mathfrak{B}_{\mathscr{Q}} f(t) = \mathfrak{F}^{-1} \mathbb{1}_{\mathscr{Q}} \mathfrak{F} f(t), \tag{3.50}$$

where $\mathbb{1}_{(\cdot)}$ is the indicator function. We consider the following eigenvalues equation

$$\mathfrak{T}_T \mathfrak{B}_{\mathscr{Q}} \mathfrak{T}_{\mathscr{T}} \psi^{\mathscr{Q}}(t) = \lambda^{\mathscr{Q}} \psi^{\mathscr{Q}}(t). \tag{3.51}$$

There exists a countably infinite set of real functions $\{\psi_n^{\mathscr{Q}}(t)\}_{n=1}^{\infty}$ and a set of real positive numbers $1 > \lambda_1^{\mathscr{Q}} > \lambda_2^{\mathscr{Q}} > \cdots > 0$ with the following properties, see [51].

Property 1. The elements of $\{\lambda_n^{\mathscr{Q}}\}$ and $\{\psi_n^{\mathscr{Q}}(t)\}$ are solutions of (3.51).

Property 2. The elements of $\{\psi_n^{\mathscr{Q}}(t)\}$ are in $\mathscr{B}_{\mathscr{Q}}$.

Property 3. $\{\psi_n^{\mathscr{Q}}(t)\}$ is complete in $\mathscr{B}_{\mathscr{Q}}$.

Property 4. The elements of $\{\psi_n^{\mathscr{Q}}(t)\}$ are orthonormal in $(-\infty, \infty)$.

Property 5. The elements of $\{\psi_n^{\mathscr{Q}}(t)\}$ are orthogonal in $(-T/2, T/2)$

$$\int_{-T/2}^{T/2} \psi_n^{\mathscr{Q}}(t) \psi_m^{\mathscr{Q}}(t) dt = \begin{cases} \lambda_n^{\mathscr{Q}} & n = m, \\ 0 & \text{otherwise.} \end{cases} \tag{3.52}$$

We write $\psi(t)$ and λ instead of $\psi^{\mathscr{Q}}(t)$ and $\lambda^{\mathscr{Q}}$ when $\mathscr{Q} = [-\Omega, \Omega]$. In this special case, the eigenfunctions $\{\psi_n(t)\}$ are the prolate spheroidal wave functions (PWSF) [22].

Lemma 4. (Slepian [11]). *For any $\nu > 0$, $N = (1+\nu)\Omega T/\pi$, and $f \in \mathscr{B}_{\Omega}$, there exist* *real coefficients* $\{x_n\}$, *such that the approximation*

$$f_N(t) = \sum_{n=1}^{N} x_n \psi_n(t) \tag{3.53}$$

has vanishing error norm $\|f - f_N\|$, *as* $T \to \infty$.

Lemma 5. (Landau and Widom [23]). *For any* $\nu > 0, S = (1+\nu)\Omega' T/\pi$, *and* $f \in \mathscr{B}_{\mathscr{Q}'}$, *there exist real coefficients* $\{\alpha_n\}$, *such that the approximation*

$$f_S(t) = \sum_{n=1}^{S} \alpha_n \psi_n^{\mathscr{Q}}(t), \tag{3.54}$$

has vanishing error norm $\|f - f_S\|$, *as* $T \to \infty$.

3.4.3 Measurement vector

We consider the measurements of $f(t) \in \mathscr{B}_{\mathscr{Q}'} \subset \mathscr{B}_\Omega$

$$y_n = \int_{-T/2}^{T/2} f(t)\varphi_n(t)dt + e_n, \quad n \in \{1, \ldots, M\}, \tag{3.55}$$

where e_n is the measurement error and each measurement kernel φ_n is a bandlimited function. Since φ_n is bandlimited, this can be represented by a linear combination of the "canonical" PSWF basis of $\mathscr{B}_\Omega$, namely

$$\varphi_n(t) = \sum_{k=1}^{\infty} a_{nk} \psi_k(t) \tag{3.56}$$

Using the completeness of the $\{\psi_n(t)\}$ in $\mathscr{B}_\Omega$, and their orthogonality property, it follows that the n-th measurement can also be expressed as

$$\begin{aligned} y_n &= \int_{-T/2}^{T/2} f(t)\varphi_n(t)dt + e_n \\ &= \int_{-T/2}^{T/2} \sum_{j=1}^{\infty} x_j \psi_j(t) \sum_{k=1}^{\infty} a_{nk}\psi_k(t)dt + e_n \\ &= \sum_{j=1}^{N} a_{nj}x_j\sqrt{\lambda_j} + \sum_{j=N+1}^{\infty} a_{nj}x_j\sqrt{\lambda_j} + e_n. \end{aligned} \tag{3.57}$$

Letting $N = (1+\nu)\Omega T/\pi$, we have

$$\lim_{T\to\infty} \sum_{j=N+1}^{\infty} a_{nj}x_j\sqrt{\lambda_j} = 0. \tag{3.58}$$

It follows that as $T \to \infty$ the measurements become

$$y_n = \sum_{j=1}^{N} a_{nj}x_j\sqrt{\lambda_j} + e_n + o(1). \tag{3.59}$$

Letting $\mathbf{y} = (y_1, \cdots, y_M)$, $\mathbf{x} = (x_1\sqrt{\lambda_1}, \cdots, x_N\sqrt{\lambda_N})$, and A be an $M \times N$ matrix such that $[A]_{nj} = a_{nj}$, we define

$$\mathbf{y} = \mathrm{A}\mathbf{x} + \mathrm{e}, \tag{3.60}$$

and consider the set

$$\overline{\mathscr{X}} = \left\{\mathbf{x} : \mathbf{x} = \left(x_1\sqrt{\lambda_1}, \cdots, x_N\sqrt{\lambda_N}\right)\right\}. \tag{3.61}$$

In virtue of Lemma 4, there exists a one-to-one correspondence between $\mathscr{B}_{\mathscr{Q}'}$ and $\overline{\mathscr{X}}$, as $T \to \infty$. By (3.59) it then follows that to complete our proofs we can derive lower and upper bounds on the number of rows of A required to recover $\mathbf{x} \in \overline{\mathscr{X}}$ from $\mathbf{y} = \mathrm{A}\mathbf{x} + \mathrm{e}$ in

(3.60), and then evaluate their order of growth as $T \to \infty$.

3.5 Proofs of Theorems 7 and 8

We consider a function $\zeta(t)$ such that

$$f_N(t) = f_S(t) + \zeta(t), \tag{3.62}$$

where $f_N(t)$ and $f_S(t)$ are given in (3.53) and (3.54), and let

$$\zeta_k = \int_{-T/2}^{T/2} \zeta(t)\psi_k(t)dt. \tag{3.63}$$

It follows that for all $1 \leq k \leq N$, we have

$$\begin{aligned}\sqrt{\lambda_k}x_k &= \int_{-T/2}^{T/2} \sum_{n=1}^{S} \alpha_n \psi_n^{\mathscr{Q}}(t)\psi_k(t)dt + \zeta_k \\ &= \sum_{n=1}^{S} \alpha_n \int_{-T/2}^{T/2} \psi_n^{\mathscr{Q}}(t)\psi_k(t)dt + \zeta_k.\end{aligned} \tag{3.64}$$

We now define

$$\varphi_{k,n}^{\mathscr{Q}} = \int_{-T/2}^{T/2} \psi_n^{\mathscr{Q}}(t)\psi_k(t)dt, \tag{3.65}$$

so that we have

$$\sqrt{\lambda_k}x_k = \sum_{n=1}^{S} \alpha_n \varphi_{k,n}^{\mathscr{Q}} + \zeta_k. \tag{3.66}$$

We rewrite (3.66) in vector form as

$$\mathbf{x} = \Phi_{\mathscr{Q}}\alpha + \zeta, \tag{3.67}$$

where $\mathbf{x} \in \overline{\mathscr{X}}$, $\alpha = (\alpha_1, \cdots, \alpha_S)$, $\zeta = (\zeta_1 \cdots, \zeta_N)$, and $\Phi_{\mathscr{Q}}$ is an $N \times S$ matrix such that

$$[\Phi_{\mathscr{Q}}]_{k,n} = \varphi_{kn}^{\mathscr{Q}}. \tag{3.68}$$

By Lemmas 4 and 5, we have that ζ tends to the all zero vector as $T \to \infty$. Therefore, it is enough to determine the minimum number of measurements to recover

$$\mathbf{x} = \Phi_{\mathscr{Q}}\alpha. \tag{3.69}$$

Let us define the following set

$$\mathscr{D} = \{\Phi_{\mathscr{Q}} : \mathscr{Q} \in \mathscr{Q}'\}. \tag{3.70}$$

We rewrite $\overline{\mathscr{X}}$ in (3.61) as follows:

$$\overline{\mathscr{X}} = \{\mathbf{x} : \mathbf{x} = \Phi_{\mathscr{Q}}\alpha \text{ where } \Phi_{\mathscr{Q}} \in \mathscr{D} \text{ and } \alpha \in \mathbb{R}^S\}. \tag{3.71}$$

Lemma 6. *For all $\Phi_1, \Phi_2 \in \mathscr{D}$, there exists an $m \times N$ matrix* A *s.t.* $\operatorname{rank}(A[\Phi_1, \Phi_2]) = \operatorname{rank}[\Phi_1, \Phi_2]$*, provided that*

$$m \geq \max_{\Phi_1, \Phi_2 \in \mathscr{D}} (\operatorname{rank}[\Phi_1, \Phi_2]). \tag{3.72}$$

Proof: It is enough to show that for all $\Phi_1, \Phi_2 \in \mathscr{D}$, if A is an i.i.d Gaussian random matrix of size $m \times N$, then $\operatorname{rank}(\mathrm{A}[\Phi_1, \Phi_2]) = \operatorname{rank}([\Phi_1, \Phi_2])$ with probability 1. Since $\operatorname{rank}(\mathrm{A}[\Phi_1, \Phi_2]) \leq \operatorname{rank}([\Phi_1, \Phi_2])$, it is enough to show that $\operatorname{rank}(\mathrm{A}[\Phi_1, \Phi_2]) \geq \operatorname{rank}([\Phi_1, \Phi_2])$. For convenience, we let $[\Phi_1, \Phi_2] = \Phi$ and we will show $\operatorname{rank}(\mathrm{A}\Phi) \geq \operatorname{rank}(\Phi)$.

Note that Φ is an $N \times 2S$ matrix with $\text{rank}(\Phi) = r \leq m$. Collect r independent columns of Φ and compose an $N \times r$ matrix Φ'. Using the Gram-Schmidt process, we can transform Φ' into Φ_G, an $N \times r$ matrix, whose columns are orthonormal. By adding redundant $N - r$ orthonormal columns followed by the original r columns of Φ_G, we obtain an $N \times N$ orthogonal matrix $\overline{\Phi}_G$.

Let us define $\sigma(\mathrm{X})$ as the smallest number of linearly dependent columns of a matrix X. It is well known that, if A is an i.i.d. Gaussian random matrix of size $m \times N$, where $m < N$, then $\sigma(\mathrm{AP}) = m + 1$ with probability 1 for any fixed orthogonal matrix P, see for example [48, Proposition 1] for a proof. Therefore, the first r columns of $\mathrm{A}\overline{\Phi}_G$ are independent. Thus, we have $\text{rank}(\mathrm{A}\Phi_G) = r$, which implies $\text{rank}(\mathrm{A}\Phi') = r$. We can then conclude that $\mathrm{A}\Phi$ contains at least r independent columns, which implies $\text{rank}(\mathrm{A}\Phi) \geq r = \text{rank}(\Phi)$. $\square$

Lemma 7. *A number of measurements*

$$m \geq \max_{\Phi_1, \Phi_2 \in \mathscr{D}} \left(\text{rank}[\Phi_1, \Phi_2]\right), \tag{3.73}$$

is sufficient to recover all the elements of $\overline{\mathscr{X}}$.

Proof: From Lemma 6 it follows that for all $\Phi_1, \Phi_2 \in \mathscr{D}$ there exists an $m \times N$ matrix A such that $\text{rank}(\mathrm{A}[\Phi_1, \Phi_2]) = \text{rank}[\Phi_1, \Phi_2]$. Let us assume $\mathrm{A}\mathbf{x_1} = \mathrm{A}\mathbf{x_2}$ where $\mathbf{x_1} = \Phi_1 \alpha_1$ and $\mathbf{x_2} = \Phi_2 \alpha_2$. The expression $\mathrm{A}\mathbf{x_1} = \mathrm{A}\mathbf{x_2}$, can be rewritten as

$$\mathrm{A}[\Phi_1, \Phi_2] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = 0, \tag{3.74}$$

namely $[\alpha_1, \alpha_2]^T$ belongs to the null space of $\mathrm{A}[\Phi_1, \Phi_2]$. Since $\text{rank}(\mathrm{A}[\Phi_1, \Phi_2]) = \text{rank}[\Phi_1, \Phi_2]$, the null space of $\mathrm{A}[\Phi_1, \Phi_2]$ is the same as the null space of $[\Phi_1, \Phi_2]$. It follows that $[\alpha_1, \alpha_2]^T$ belongs to the null space of $[\Phi_1, \Phi_2]$, or equivalently

$$[\Phi_1, \Phi_2] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = 0. \tag{3.75}$$

This means $\Phi_1\alpha_1 = \Phi_2\alpha_2$, namely $\mathbf{x_1} = \mathbf{x_2}$. Therefore, A is one-to-one on $\overline{\mathscr{X}}$, which implies that the elements of $\overline{\mathscr{X}}$ can be recovered. □

Lemma 8. *A number of measurements*

$$m < \max_{\Phi_1, \Phi_2 \in \mathscr{D}} (\mathrm{rank}[\Phi_1, \Phi_2]) \tag{3.76}$$

is not sufficient to recover all the elements of $\overline{\mathscr{X}}$.

Proof: If all elements $\mathbf{x} \in \overline{\mathscr{X}}$ can be recovered from the measurements $\mathbf{y} = \mathrm{A}\mathbf{x}$, where A is an $m \times N$ matrix, this means A is one-to-one on $\overline{\mathscr{X}}$. Therefore, for all $\mathbf{x_1}$ and $\mathbf{x_2}$, $\mathrm{A}\mathbf{x_1} = \mathrm{A}\mathbf{x_2}$ implies $\mathbf{x_1} = \mathbf{x_2}$. Let us assume $\mathbf{x_1} = \Phi_1\alpha_1$ and $\mathbf{x_2} = \Phi_2\alpha_2$, then $\mathrm{A}\Phi_1\alpha_1 = \mathrm{A}\Phi_2\alpha_2$ implies $\Phi_1\alpha_1 = \Phi_2\alpha_2$. This is equivalent to saying that

$$\mathrm{A}[\Phi_1, \Phi_2] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = 0 \tag{3.77}$$

implies

$$[\Phi_1, \Phi_2] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = 0. \tag{3.78}$$

Namely, the null space of $\mathrm{A}[\Phi_1\Phi_2]$ is contained in the null space of $[\Phi_1\Phi_2]$. By the rank-nullity theorem, we have

$$\text{rank}(\text{A}[\Phi_1, \Phi_2]) \geq \text{rank}[\Phi_1, \Phi_2]. \tag{3.79}$$

Since $m \geq \text{rank}(\text{A}[\Phi_1, \Phi_2])$ and (3.79) holds for all $\Phi_1, \Phi_2 \in \mathscr{D}$, the result follows. □

Lemma 9. *We have*

$$\lim_{T\to\infty} \frac{\max_{\Phi_1,\Phi_2\in\mathscr{D}}(\text{rank}[\Phi_1, \Phi_2])}{2S} = 1. \tag{3.80}$$

Proof: Let $\mathscr{Q}_1, \mathscr{Q}_2 \in \mathscr{Q}'$, and consider the multi-band signals

$$f_1(t) = \sum_{n=1}^{S} \alpha_n \psi_n^{\mathscr{Q}_1}(t), \tag{3.81}$$

$$f_2(t) = \sum_{n=1}^{S} \beta_n \psi_n^{\mathscr{Q}_2}(t), \tag{3.82}$$

and

$$f_S(t) = f_1(t) + f_2(t). \tag{3.83}$$

Consider the N-dimensional vector

$$\mathbf{z} = [\Phi_{\mathscr{Q}_1}, \Phi_{\mathscr{Q}_2}] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_S \\ \beta_1 \\ \vdots \\ \beta_S, \end{bmatrix}. \tag{3.84}$$

whose elements, by (3.65) and (3.68), and then using (3.81), (3.82), (3.83), are

$$\begin{aligned}
z_n &= \sum_{j=1}^{S} \alpha_j \varphi_{n,j}^{\mathscr{Q}_1} + \sum_{j=1}^{S} \beta_j \varphi_{n,j}^{\mathscr{Q}_2} \\
&= \sum_{j=1}^{S} \alpha_j \int_{-T/2}^{T/2} \psi_j^{\mathscr{Q}_1}(t)\psi_n(t)dt \\
&+ \sum_{j=1}^{S} \beta_j \int_{-T/2}^{T/2} \psi_j^{\mathscr{Q}_2}(t)\psi_n(t)dt \\
&= \int_{-T/2}^{T/2} f_1(t)\psi_n(t)dt + \int_{-T/2}^{T/2} f_2(t)\psi_n(t)dt \\
&= \int_{-T/2}^{T/2} f_S(t)\psi_n(t)dt. \qquad (3.85)
\end{aligned}$$

We consider the case when $\mathbf{z}$ is the all zero vector. In this case, since by (3.85) the elements $\{z_n\}$ are also the PSWF coefficients of $f_S(t)$, it follows that

$$\lim_{T\to\infty} f_S(t) = 0. \qquad (3.86)$$

We now choose $\mathscr{Q}_1$ and $\mathscr{Q}_2$ such that $\mathscr{Q}_1 \cap \mathscr{Q}_2 = \emptyset$, so that (3.86) implies

$$\lim_{T\to\infty} f_1(t) = \lim_{T\to\infty} f_2(t) = 0. \qquad (3.87)$$

It follows that all coefficients $\{\alpha_n\}$ and $\{\beta_n\}$ in (3.81) and (3.82) must tends to zero as $T \to \infty$, the columns of $[\Phi_{\mathscr{Q}_1}, \Phi_{\mathscr{Q}_2}]$ become independent, and we have

$$\lim_{T\to\infty} \frac{\text{rank}[\Phi_{\mathscr{Q}_1}, \Phi_{\mathscr{Q}_2}]}{2S} = 1. \qquad (3.88)$$

On the other hand, $\text{rank}[\Phi_1, \Phi_2] \leq 2S$ for all $\Phi_1, \Phi_2 \in \mathscr{D}$ because the number of columns of $[\Phi_1, \Phi_2]$ is $2S$. It follows that our choice $\Phi_1 = \Phi_{\mathscr{Q}_1}$ and $\Phi_2 = \Phi_{\mathscr{Q}_2}$ achieves the maximum rank and the result follows. $\square$

By combining Lemmas 7 and 9 it follows that with $2S = 2(1+\nu)\Omega' T/\pi$ measurements we can recover any vector $\mathbf{x}$ in (3.69) with vanishing error as $T \to \infty$, and since the vector ζ tends zero we can also recover any vector $\mathbf{x}$ in (3.67). It follows that we can recover the coefficients representing any signal in $\mathscr{B}_{\mathscr{Q}'}$ with vanishing error using a measurement rate

$$\bar{M} = \frac{2\Omega'}{\pi} + 2\nu\frac{\Omega'}{\pi} > \frac{2\Omega'}{\pi}, \tag{3.89}$$

and the proof of Theorem 7 is complete.

On the other hand, by combining Lemmas 8 and 9 it follows that with less than $2S = 2(1+\nu)\Omega' T/\pi$ measurements we cannot recover all possible vectors $\mathbf{x}$ in (3.69) with vanishing error as $T \to \infty$. This also means that we cannot recover all possible vectors $\mathbf{x}$ in (3.67). It follows that with a number of measurements $M = 2\Omega' T/\pi + o(T)$, and hence a measurement rate

$$\bar{M} = 2\Omega'/\pi \tag{3.90}$$

we cannot recover all signals in $\mathscr{B}_{\mathscr{Q}'}$, and the proof of Theorem 8 is also complete.

3.6 Proofs of Theorems 9-12

In the following, we use $||\cdot||$ to denote the Euclidean norm for vectors in $\mathbb{R}^N$

$$||\mathbf{x}|| = \sqrt{\sum_{n=1}^{N} x_n^2}, \tag{3.91}$$

and the spectral norm for matrices

$$||\mathrm{A}|| = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{||\mathrm{A}\mathbf{x}||}{||\mathbf{x}||}. \tag{3.92}$$

We also use the usual notation for signals defined in (3.48).

3.6.1 The key lemmas

Let $\mathscr{B}_\Delta$ be the collection of all elements in $\mathscr{B}_{\mathscr{Q}'}$ such that the extremal points of all sub-bands belong to the discrete set $\mathscr{J}$ defined in (3.23). For any signal $f \in \mathscr{B}_{\mathscr{Q}'}$, let $f_\Delta \in \mathscr{B}_\Delta$ such that

$$f_\Delta = \arg\min_{f' \in \mathscr{B}_\Delta} \|f - f'\|. \tag{3.93}$$

Since all $f \in \mathscr{B}_{\mathscr{Q}'}$ are square-integrable, it follows that

$$\lim_{\Delta \to 0} \|f - f_\Delta\| = 0. \tag{3.94}$$

Hence, if f_Δ can be recovered using a measurement rate $\bar{M}_\Delta$, then f can be recovered using a measurement rate

$$\bar{M} = \lim_{\Delta \to 0} \bar{M}_\Delta. \tag{3.95}$$

Consider now the set $\overline{\mathscr{X}}(\Delta)$ of vectors of $N = (1+\nu)\Omega T/\pi$ real coefficients, such that every element of $\mathscr{B}_\Delta$ is approximated, with vanishing error as $T \to \infty$, by an element of $\overline{\mathscr{X}}(\Delta)$. We also consider $\mathscr{X}(\Delta) \subset \overline{\mathscr{X}}(\Delta)$ containing all elements of $\overline{\mathscr{X}}(\Delta)$ that have norm at most one. To prove Theorems 9-12, it is enough to prove following two lemmas.

Lemma 10. *We can robustly recover all signals $f \in \mathscr{B}_\Delta$ using a measurements rate*

$$\bar{M}_\Delta > R_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)]. \tag{3.96}$$

Lemma 11. *In the absence of measurement error, we cannot perfectly recover all signals $f \in \mathscr{B}_\Delta$ using a measurement rate*

$$\bar{M}_\Delta < 2R_F[\mathscr{X}(\Delta)]. \tag{3.97}$$

To see that Theorems 9-12 follow from these two lemmas, first note that the lemmas imply

$$R_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)] \geq 2R_F[\mathscr{X}(\Delta)], \tag{3.98}$$

on the other hand, we have

$$\dim_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)] \leq 2\dim_F[\mathscr{X}(\Delta)], \tag{3.99}$$

which implies

$$R_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)] \leq 2R_F[\mathscr{X}(\Delta)]. \tag{3.100}$$

Combining (3.98) and (3.100) it follows that

$$R_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)] = 2R_F[\mathscr{X}(\Delta)]. \tag{3.101}$$

Theorem 11 now follows from Lemma 10 and (3.101) by taking the limit for $\Delta \to 0$, and Theorem 9 follows directly from Theorem 11. On the other hand, from Lemma 11 it follows by taking the limit for $\Delta \to 0$ that with a measurement rate

$$\bar{M} < \lim_{\Delta \to 0} 2R_F[\mathscr{X}(\Delta)] \tag{3.102}$$

we cannot perfectly recover all signals $f \in \mathscr{B}_{\mathscr{Q}'}$. As for the equality, combining this result with Theorems 7, 8, and 9, we conclude that

$$\lim_{\Delta \to 0} 2R_F[\mathscr{X}(\Delta)] = \frac{2\Omega'}{\pi}, \tag{3.103}$$

which completes the proof of Theorem 10. Theorem 12 follows directly from Theorem 10.

3.6.2 Proof of Lemma 10

Definition 14. (Inverse Lipschitz condition.) *A matrix* A *satisfies the inverse Lipschitz condition on a set* $\mathscr{U}$ *if there exists a constant* $\beta > 0$ *such that for all* $\mathbf{u_1}, \mathbf{u_2} \in \mathscr{U}$*, we have*

$$\beta \|\mathbf{u}_1 - \mathbf{u}_2\| \leq \|\mathrm{A}\mathbf{u}_1 - \mathrm{A}\mathbf{u}_2\|. \tag{3.104}$$

We claim that if A satisfies the inverse Lipschitz condition on $\overline{\mathscr{X}}(\Delta)$, then every $\mathbf{x} \in \overline{\mathscr{X}}(\Delta)$ can be robustly recovered from $\mathbf{y} = \mathrm{A}\mathbf{x} + \mathbf{e}$. To prove this claim, consider the following two cases: (a) $\mathbf{y} \in \mathrm{A}\overline{\mathscr{X}}(\Delta)$, where $\mathrm{A}\overline{\mathscr{X}}(\Delta)$ is the set $\{\mathrm{A}\mathbf{x} : \mathbf{x} \in \overline{\mathscr{X}}(\Delta)\}$, and (b) $\mathbf{y} \notin \mathrm{A}\overline{\mathscr{X}}(\Delta)$. In the first case, let $\mathbf{x}'$ be a solution of $\mathbf{y} = \mathrm{A}\mathbf{x}'$ and let $\mathbf{x}'$ be the vector used to recover $\mathbf{x}$. Then, the recovery error is bounded as

$$\beta \|\mathbf{x} - \mathbf{x}'\| \leq \|\mathrm{A}\mathbf{x} - \mathrm{A}\mathbf{x}'\| = \|\mathbf{e}\|, \tag{3.105}$$

which guarantees robust recovery. On the other hand, if $\mathbf{y} \notin \mathrm{A}\overline{\mathscr{X}}(\Delta)$, let $\mathbf{x}'' \in \overline{\mathscr{X}}(\Delta)$ such that $\mathrm{A}\mathbf{x}''$ is the closest to $\mathbf{y}$ among all the elements of $\mathrm{A}\overline{\mathscr{X}}(\Delta)$. By letting $\mathbf{x}''$ be the vector used to recover $\mathbf{x}$, we can bound the recovery error as

$$\begin{aligned}
\beta \|\mathbf{x} - \mathbf{x}''\| &\leq \|\mathrm{A}\mathbf{x} - \mathrm{A}\mathbf{x}''\| && (3.106)\\
&\leq \|\mathrm{A}\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathrm{A}\mathbf{x}''\| && (3.107)\\
&\leq 2\|\mathbf{e}\|, && (3.108)
\end{aligned}$$

which guarantees robust recovery. The claim now follows and we can proceed to derive a sufficient condition that ensures A satisfies the inverse Lipschitz condition on the set $\overline{\mathscr{X}}(\Delta)$.

By letting $\overline{\mathscr{Z}}(\Delta) = \overline{\mathscr{X}}(\Delta) \oplus \overline{\mathscr{X}}(\Delta)$, the inverse Lipschitz condition is equivalent to stating that for all $\mathbf{z} \in \overline{\mathscr{Z}}(\Delta)$

$$\beta\|\mathbf{z}\| \leq \|\mathrm{A}\mathbf{z}\|. \tag{3.109}$$

Consider the normalized set $\mathscr{Z}^{'}(\Delta) \subset \overline{\mathscr{Z}}(\Delta)$ containing all the elements of $\overline{\mathscr{Z}}(\Delta)$ that are vectors of unit norm, and let $k' = \dim_F[\mathscr{Z}^{'}(\Delta)]$. If (3.109) holds for all $\mathbf{z} \in \overline{\mathscr{Z}}(\Delta)$, then it also holds for all $\mathbf{z} \in \mathscr{Z}^{'}(\Delta)$, and vice versa. In the following, we consider $\mathscr{Z}^{'}(\Delta)$ instead than $\overline{\mathscr{Z}}(\Delta)$.

Let $\mathscr{L}_\varepsilon[\mathscr{Z}^{'}(\Delta)]$ be a minimal ε-covering set of $\mathscr{Z}^{'}(\Delta)$, namely a minimum cardinality set such that any point in $\mathscr{Z}^{'}(\Delta)$ is within distance ε from at least one point of $\mathscr{L}_\varepsilon[\mathscr{Z}^{'}(\Delta)]$. Let $L_\varepsilon[\mathscr{Z}'(\Delta)] = |\mathscr{L}_\varepsilon[\mathscr{Z}^{'}(\Delta)]|$. We need the following preliminary results.

Lemma 12. [53, Fact 2.1.]

$$\begin{aligned}\dim_F[\mathscr{Z}^{'}(\Delta)] = \inf\{d : \forall\varepsilon \in (0,1)\,\exists\gamma > 0 :\\ L_\varepsilon[\mathscr{Z}^{'}(\Delta)] \leq \gamma\left(\frac{1}{\varepsilon}\right)^d\right\}.\end{aligned} \tag{3.110}$$

Let $\mathscr{G}$ be the space of all orthogonal projections in $\mathbb{R}^N$ of rank m, and μ be the invariant measure on $\mathscr{G}$ with respect to orthogonal transformations.

Definition 15. (Shadow of a set). *The shadow of a set $\mathscr{B}$ in $\mathbb{R}^N$ is*

$$S(\mathscr{B}) = \{\mathrm{P} \in \mathscr{G} : 0 \in \mathrm{P}\mathscr{B}\}. \tag{3.111}$$

Lemma 13. [53, Theorem 5.1.] *The measure of the shadow of a ρ-ball $\mathscr{B}$ centered at a distance r from the origin is bounded as*

$$\mu(S(\mathscr{B})) \leq \delta\left(\frac{\rho}{r}\right)^m, \tag{3.112}$$

where δ is a positive constant.

We now provide a key lemma.

Lemma 14. *For almost every projection* P *of rank $m > k'$, there exists a constant c such that, for all $\mathbf{z} \in \mathscr{X}^{'}(\Delta)$*

$$\|\mathrm{P}\mathbf{z}\| > c\|\mathbf{z}\|. \tag{3.113}$$

Proof: From Lemma 12, it follows that for any $0 < \varepsilon < 1$ there exists a constant $\gamma > 0$ such that

$$L_{\varepsilon}[\mathscr{X}^{'}(\Delta)] \leq \gamma \left(\frac{1}{\varepsilon}\right)^{k'}. \tag{3.114}$$

By definition of ε-covering, for any $\mathbf{z} \in \mathscr{X}^{'}(\Delta)$, there exists a vector $\mathbf{l} \in \mathscr{L}_{\varepsilon}[\mathscr{X}^{'}(\Delta)]$ such that

$$\|\mathbf{z} - \mathbf{l}\| \leq \varepsilon. \tag{3.115}$$

Letting $\mathbf{v} = \mathbf{z} - \mathbf{l}$, we have

$$\begin{aligned} \|\mathrm{P}\mathbf{z}\| &= \|\mathrm{P}(\mathbf{l} + \mathbf{v})\| \\ &\geq \|\mathrm{P}\mathbf{l}\| - \|\mathrm{P}\mathbf{v}\| \\ &\geq \|\mathrm{P}\mathbf{l}\| - \varepsilon, \end{aligned} \tag{3.116}$$

where the last inequality follows from

$$\begin{aligned} \|\mathrm{P}\mathbf{v}\| &\leq \|\mathrm{P}\|\,\|\mathbf{v}\| \\ &= \|\mathbf{v}\| \\ &\leq \varepsilon. \end{aligned} \tag{3.117}$$

From (3.116) we have that if for all $\mathbf{l} \in \mathscr{L}_{\varepsilon}[\mathscr{X}^{'}(\Delta)]$ we have $\|\mathrm{P}\mathbf{l}\| > 2\varepsilon$, then we also have $\|\mathrm{P}\mathbf{z}\| > \varepsilon = \varepsilon\|\mathbf{z}\|$, and letting $c = \varepsilon$ the result follows. What remains to be shown then, is that for almost every projection P of rank m, and for all $\mathbf{l} \in \mathscr{L}_\varepsilon[\mathscr{Z}^{'}(\Delta)]$, we have $\|\mathrm{P}\mathbf{l}\| > 2\varepsilon$.

We let

$$\mathscr{L}_\varepsilon[\mathscr{Z}^{'}(\Delta)] = \{\mathbf{l}_1, \cdots, \mathbf{l}_L\}, \tag{3.118}$$

where $L = L_\varepsilon[\mathscr{Z}^{'}(\Delta)]$, and for all $1 \leq i \leq L$ we define

$$\mathscr{H}_i = \{\mathrm{P} \in \mathscr{G} : \|\mathrm{P}\mathbf{l}_i\| \leq 2\varepsilon\}. \tag{3.119}$$

We also let

$$\mathscr{H} = \bigcup_{i=1}^{L} \mathscr{H}_i, \tag{3.120}$$

so that

$$\mu(\mathscr{H}) = \mu\left(\bigcup_{i=1}^{L} \mathscr{H}_i\right) \leq \sum_{i=1}^{L} \mu(\mathscr{H}_i). \tag{3.121}$$

We claim that if $\|\mathrm{P}\mathbf{l}\| \leq 2\varepsilon$, then $0 \in \mathrm{P}\mathscr{B}^{\mathbf{l}}_{2\varepsilon}$, where $\mathscr{B}^{\mathbf{l}}_{2\varepsilon}$ is a 2ε-ball whose center is $\mathbf{l}$. This can be shown as follows: let $\mathbf{b} = \mathbf{l} - \mathrm{P}\mathbf{l}$, then $\mathbf{b} \in \mathscr{B}^{\mathbf{l}}_{2\varepsilon}$ and $\mathrm{P}\mathbf{b} = \mathrm{P}\mathbf{l} - \mathrm{P}^2\mathbf{l} = 0$. It follows that

$$\begin{aligned} \mu(\mathscr{H}_i) &\leq \mu\left(\{\mathrm{P} \in \mathscr{G} : 0 \in \mathrm{P}\mathscr{B}^{\mathbf{l}_i}_{2\varepsilon}\}\right) \\ &= \mu(S(\mathscr{B}^{\mathbf{l}_i}_{2\varepsilon})) \\ &\leq \delta\,(2\varepsilon)^m, \end{aligned} \tag{3.122}$$

where the last inequality follows from Lemma 13. We now have

$$\begin{aligned} \mu(\mathscr{H}) &\leq \sum_{i=1}^{L} \mu(\mathscr{H}_i) \\ &\leq L\delta(2\varepsilon)^m \\ &\leq \gamma\delta 2^m \varepsilon^{m-k'}, \end{aligned} \tag{3.123}$$

where the last inequality follows from (3.114). By taking a sufficiently small ε, we can now make $\mu(\mathscr{H})$ arbitrary close to 0, and the proof is complete. □

By Lemma 14, there exists a projection P of rank m such that for all $\mathbf{z} \in \mathscr{Z}^{'}(\Delta)$ we have $\|\mathrm{P}\mathbf{z}\| > c\|\mathbf{z}\|$. By applying Gaussian elimination to such a projection and selecting the non-zero rows of it, we obtain an $m \times N$ matrix A. Since $\|\mathrm{P}\mathbf{z}\| = \|\mathrm{A}\mathbf{z}\|$, it follows that any $\mathbf{x} \in \overline{\mathscr{X}}(\Delta)$ can be robustly recovered from $\mathbf{y} = \mathrm{A}\mathbf{x} + \mathbf{e}$ with a number of measurements larger than k'.

What remains to be done is to show that $k' = \dim_F[\mathscr{Z}^{'}(\Delta)] \leq \dim_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)]$. Let $\mathscr{Z}(\Delta) \subset \overline{\mathscr{Z}}(\Delta)$ containing all elements of $\overline{\mathscr{Z}}(\Delta)$ that have norm at most one. Since $\mathscr{Z}^{'}(\Delta) \subset \mathscr{Z}(\Delta)$, we have $k' \leq \dim_F[\mathscr{Z}(\Delta)]$. It is then enough to show that $\dim_F[\mathscr{Z}(\Delta)] = \dim_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)]$.

Lemma 15. *We have*

$$\dim_F[\mathscr{Z}(\Delta)] = \dim_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)] \tag{3.124}$$

Proof: Let $\mathbf{z} \in \mathscr{Z}(\Delta)$ be a vector of coefficients of a multi-band function $f_{\mathbf{z}}$ whose spectral support is bounded by $4\Omega'$ and whose energy is bounded by one. It follows that $f_{\mathbf{z}}$ can be represented as

$$f_{\mathbf{z}} = f_{\mathbf{x}_1} + f_{\mathbf{x}_2} \tag{3.125}$$

where $f_{\mathbf{x}_i}, i \in \{1,2\}$ is a multi-band signal whose spectral support is bounded by $2\Omega'$ and whose energy is bounded by one. Let $\mathbf{x}_i$ be a vector of coefficients for $f_{\mathbf{x}_i}$, $i \in \{1,2\}$. Then, we have

$$\mathbf{z} = \mathbf{x}_1 + \mathbf{x}_2 \tag{3.126}$$

where $\mathbf{x}_i \in \mathscr{X}(\Delta)$. Since $\mathscr{Z}(\Delta) \subset \mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)$, we conclude that

$$\dim_F[\mathscr{Z}(\Delta)] \leq \dim_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)]. \tag{3.127}$$

Conversely, let us consider $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{X}(\Delta)$. Then, we have

$$\frac{\mathbf{x}_1 + \mathbf{x}_2}{2} \in \mathscr{Z}(\Delta), \tag{3.128}$$

which implies $\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta) \subset 2\mathscr{Z}(\Delta)$, where $2\mathscr{Z}(\Delta)$ indicates the set $\{2\mathbf{z} : \mathbf{z} \in \mathscr{Z}(\Delta)\}$. Therefore, we conclude that

$$\dim_F[\mathscr{Z}(\Delta)] \geq \dim_F[\mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)]. \tag{3.129}$$

By combining (3.127) and (3.129), we obtain the desired result. □

3.6.3 Proof of Lemma 11

If all vectors $\mathbf{x} \in \mathscr{X}(\Delta)$ can be recovered from $\mathbf{y} = \mathrm{A}\mathbf{x}$, then all vectors $\mathbf{x} \in \overline{\mathscr{X}}(\Delta)$ can also be recovered from $\mathbf{y} = \mathrm{A}\mathbf{x}$, and vice versa. In the following, we consider $\mathscr{X}(\Delta)$ rather than $\overline{\mathscr{X}}(\Delta)$.

In order to prove Lemma 11, it is enough to show that a number of measurements

$$m < 2 \dim_F[\mathscr{X}(\Delta)] \tag{3.130}$$

is not sufficient to recover all the elements of $\mathscr{X}(\Delta)$ as $T \to \infty$.

Let us define the set $\mathscr{W}(\Delta) = \mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)$. If all $\mathbf{x} \in \mathscr{X}(\Delta)$ can be recovered from $\mathbf{y}$, then A is a one-to-one map on $\mathscr{X}(\Delta)$, and vice versa. Also, if A is a one-to-one map on $\mathscr{X}(\Delta)$, then

$$\ker(\mathrm{A}) \cap \mathscr{W}(\Delta) = \{\mathbf{0}\}, \tag{3.131}$$

and vice versa, where $\ker(\mathrm{A})$ indicates the kernel of A. We then need to show that (3.130) violates (3.131). For convenience of notation, we define $k = 2\,\dim_F[\mathscr{X}(\Delta)]$.

Let us assume that $\mathscr{W}(\Delta)$ contains a k-dimensional Euclidean ball. Note that (3.130) implies $\mathrm{rank}(\mathrm{A}) < k$. Since $\mathrm{rank}(\mathrm{A}) + \mathrm{nullity}(\mathrm{A}) = N$, it follows that $\mathrm{nullity}(\mathrm{A})$ is greater than $N - k$. This means that the dimension of $\ker(\mathrm{A})$ is larger than $N - k$, which violates (3.131) because $\mathscr{W}(\Delta)$ contains a k-dimensional Euclidean ball.

It follows that in order to prove Lemma 11, it is enough to show that $\mathscr{W}(\Delta)$ contains a k-dimensional Euclidean ball. We will show that this is the case when $T \to \infty$.

We need some additional definitions, followed by a preliminary result.

Definition 16. (Diameter). *For any $\mathscr{S} \subset \mathbb{R}^N$, we let*

$$\mathrm{diam}(\mathscr{S}) = \sup_{\mathbf{x},\mathbf{y} \in \mathscr{S}} \|\mathbf{x} - \mathbf{y}\|. \tag{3.132}$$

Definition 17. (Hausdorff measure). *Let $\mathscr{U} \subset \mathbb{R}^N$ and $\{\mathscr{S}_i\}$ be a cover of $\mathscr{U}$ formed by balls of radius $r < \mu$. We let*

$$\zeta_\mu^s(\mathscr{U}) = \inf_{\{\mathscr{S}_i\}} \sum_i [\mathrm{diam}(\mathscr{S}_i)]^s. \tag{3.133}$$

The s-dimensional Hausdorff measure of $\mathscr{U}$ is given by the limit

$$\zeta^s(\mathscr{U}) = \lim_{\mu \to 0} \zeta_\mu^s(\mathscr{U}). \tag{3.134}$$

Definition 18. (Hausdorff dimension). *For any $\mathscr{U} \subset \mathbb{R}^N$, the Hausdorff dimension of $\mathscr{U}$*

is

$$\dim_H(\mathscr{U}) = \sup\{s \geq 0 : \zeta^s(\mathscr{U}) = \infty\}. \tag{3.135}$$

The Hausdorff dimension has the following two important properties, see [46].

Property 1. (Unit ball). For any integer d such that $0 \leq d \leq N$, the Hausdorff dimension of the unit ball $B^d(0,1) \subset \mathbb{R}^d \subset \mathbb{R}^N$ is d.

Property 2. (Countable stability). Let $\mathscr{U}_i \subset \mathbb{R}^N$. Then, $\dim_H(\bigcup_{i=1}^{\infty} \mathscr{U}_i) = \sup_i\{\dim_H(\mathscr{U}_i)\}$

From these definitions it follows that

$$\dim_H(\mathscr{U}) \leq \dim_F(\mathscr{U}). \tag{3.136}$$

However, by Lemma 16 below, if a set satisfies a quasi self-similar property, then the Hausdorff dimension is equal to the fractal dimension.

Definition 19. (Quasi self-similarity) *Let $\mathscr{U} \subset \mathbb{R}^N$. If for all $\mathbf{x}, \mathbf{y} \in \mathscr{U} \cap \mathscr{B}$, there exist $a, r_0 > 0$ such that for any ball $\mathscr{B}$ with radius $r < r_0$, there is a mapping $\phi : \mathscr{U} \cap \mathscr{B} \to \mathscr{U}$ satisfying*

$$a \cdot \|\mathbf{x} - \mathbf{y}\| \leq r \cdot \|\phi(\mathbf{x}) - \phi(\mathbf{y})\|, \tag{3.137}$$

then, we say that $\mathscr{U}$ is quasi self-similar.

Lemma 16. [52, Theorem 3.] *Let $\mathscr{U}$ be a nonempty compact subset of $\mathbb{R}^N$ that is quasi self-similar. Then,*

$$\dim_H(\mathscr{U}) = \dim_F(\mathscr{U}). \tag{3.138}$$

We are now ready to show our final step.

Lemma 17. *For sufficiently large T, the set $\mathscr{W}(\Delta) = \mathscr{X}(\Delta) \oplus \mathscr{X}(\Delta)$ contains a k-dimensional Euclidean ball.*

Proof: We have

$$\mathscr{X}(\Delta) = \bigcup_i \mathscr{X}_i \tag{3.139}$$

where $\mathscr{X}_i$ is the set of coefficient vectors of all multi-band signals of a fixed sub-band allocation of measure at most $2\Omega'$ and norm at most one. Since $\mathscr{X}(\Delta)$ is a countable union, by Property 2 of the Hausdorff dimension we have

$$\dim_H[\mathscr{X}(\Delta)] = \sup_i\{\dim_H(\mathscr{X}_i)\}. \tag{3.140}$$

Since the Hausdorff dimension of $\mathscr{X}_i$ does not depend on i, we also have that for all i

$$\dim_H[\mathscr{X}(\Delta)] = \dim_H(\mathscr{X}_i). \tag{3.141}$$

Since $\mathscr{X}(\Delta)$ is a nonempty compact subset of $\mathbb{R}^N$ that is also quasi self-similar with $a = r_0 = 1$ and $\phi(\mathbf{x}) = \mathbf{x}/r$, it follows that

$$\dim_H[\mathscr{X}(\Delta)] = \dim_F[\mathscr{X}(\Delta)]. \tag{3.142}$$

Next, we consider two sets of coefficient vectors $\mathscr{X}_1$ and $\mathscr{X}_2$, whose sub-bands do not have any intersection. We have

$$\mathscr{X}_i = \{\mathbf{x} : \mathbf{x} = \Phi_i\alpha \text{ where } \|\mathbf{x}\| \leq 1 \text{ and } \alpha \in \mathbb{R}^S\}, \tag{3.143}$$

for $i = 1, 2$. By the same argument used in the proof of Lemma 9, it follows that for T large enough the columns of Φ_1 and Φ_2 are independent. Also, note that $\mathscr{X}_i$ is an Euclidean ball and by Property one of the Hausdorff dimension it follows that $\mathscr{X}_i$ is a $\dim_H(\mathscr{X}_i)$-dimensional Euclidean ball. Now, by definition, $\mathscr{W}(\Delta)$ includes $\mathscr{X}_1 \oplus \mathscr{X}_2$, and since $\mathscr{X}_1$ and $\mathscr{X}_2$ are $\dim_H(\mathscr{X}_i)$-dimensional Euclidean balls and the columns of

Φ_1 and Φ_2 are independent, it follows that $\mathscr{W}(\Delta)$ contains a $2\dim_H(\mathscr{X}_i)$-dimensional Euclidean ball. Using (3.141) and (3.142), it follows that for T large enough $\mathscr{W}_\Delta$ contains a $2\dim_F[\mathscr{X}(\Delta)]$ Euclidean ball, or equivalently, a k-dimensional Euclidean ball. $\square$

3.7 Conclusion

We have investigated the phase-transition threshold of the minimum measurement rate sufficient for completely blind reconstruction of any multi-band signal of given spectral support measure. This threshold has been shown to coincide with twice the fractal dimension per unit ambient dimension of the space spanned by the optimal approximation for bandlimited signals. This result provides an operational characterization of the fractal dimension, and parallels an analogous coding theorem for the compression of discrete-time, analog, i.i.d. sources, where the critical threshold is shown to be equal to the information dimension per unit ambient dimension of the source [44, 45]. Advantages of the deterministic approach include being oblivious to a priori assumptions on the source distribution, and providing recovery guarantees for all signals, rather than for a large fraction of them. In both cases, fundamental limits apply to the asymptotic regime of large signal dimension. In the stochastic case, probabilistic concentration is achieved exploiting the ergodicity of the process, while in the deterministic case vanishing error energy is achieved exploiting spectral concentration. Despite both results can be viewed at the high level as an instance of dimensional reduction due to regularity constraints, the tools required in the deterministic setting are quite different from those used in traditional information theory, and include machinery from approximation theory, and geometry of functional spaces. The systematic study of these techniques is clearly desirable, and this recommendation dates back to Kolmogorov [3]. Exploiting some of our recent results [55], we have shown that the price to pay to obtain deterministic guarantees of reconstruction for all signals is only a factor of two in the measurement rate, compared to probabilistic reconstruction. It is also the case that the absence of additional spectral information such as the one assumed in [39, 40, 43], does not lead to any penalty in the measurement rate.

Practical achievability schemes for blind reconstruction of continuous signals that come close to the information-theoretic optimum remain an open problem, while much progress has been made for both discrete-time and continuous-time settings, under various assumptions on what information about the signal is available a priori [39, 40, 43, 44, 45, 56]. Another interesting open question is the determination of the critical threshold for linear approximation schemes. In this case, without any knowledge of the spectral support it is not possible to set-up the eigenvalue equation leading to the optimal subspace approximation [23], and the challenge is to infer the basis functions directly from the measurements. Investigation of sampling schemes for blind reconstruction is also of interest, due to their relevance for practical applications. Our results provide an information-theoretic baseline for performance assessment in all of these cases. Finally, extensions to signals of multiple variables would be of interest in various settings, for example in the context of remote sensing. In this case, a desirable outcome would be the computation of the fractal dimension of signals radiated from a bounded domain, generalizing the notion of number of degrees of freedom for bandlimited signals studied in [14], to signals that are sparse in both the frequency and the wavenumber spectrum.

3.8 Appendix

3.8.1 Proofs of (3.20) and (3.22)

First let us consider (3.20). Since $\mathscr{X}_{\mathrm{B}} \subset \mathscr{X}_{\mathrm{B}} \oplus \mathscr{X}_{\mathrm{B}}$, we have

$$\dim_F(\mathscr{X}_{\mathrm{B}}) \leq \dim_F(\mathscr{X}_{\mathrm{B}} \oplus \mathscr{X}_{\mathrm{B}}). \tag{3.144}$$

For any $\mathbf{x} \in \mathscr{X}_{\mathrm{B}} \oplus \mathscr{X}_{\mathrm{B}}$, we have $\mathbf{x}/2 \in \mathscr{X}_{\mathrm{B}}$, or equivalently $\mathbf{x} \in 2\mathscr{X}_{\mathrm{B}}$, where $2\mathscr{X}_{\mathrm{B}}$ indicates the set $\{2\mathbf{x} : \mathbf{x} \in \mathscr{X}_{\mathrm{B}}\}$. This implies $\mathscr{X}_{\mathrm{B}} \oplus \mathscr{X}_{\mathrm{B}} \subset 2\mathscr{X}_{\mathrm{B}}$, and we have

$$\dim_F(\mathscr{X}_{\mathrm{B}}) \geq \dim_F(\mathscr{X}_{\mathrm{B}} \oplus \mathscr{X}_{\mathrm{B}}). \tag{3.145}$$

Combining (3.144) and (3.145), we obatin (3.20).

Next, we consider (3.22). We let $\mathscr{X}'_{\mathrm{B}}$ be a set of vectors such that for any $\mathbf{x} = (x_1, \cdots, x_N) \in \mathscr{X}_{\mathrm{B}}$, $\mathbf{x}' \in \mathscr{X}'$ is the vector of its first N' components, namely $\mathbf{x}' = (x_1, \cdots, x_{N'})$ where $N' = \Omega T/\pi + o(T)$. From inequality (137) in Theorem 6 of [55], we have

$$H_\varepsilon(\mathscr{X}_{\mathrm{B}}) \geq H_\varepsilon(\mathscr{X}'_{\mathrm{B}}). \tag{3.146}$$

By inequality (99) of Theorem 3 in [55], we have

$$H_\varepsilon(\mathscr{X}'_{\mathrm{B}}) \geq N' \left[\log\left(\zeta(N')\frac{1}{\varepsilon}\right)\right], \tag{3.147}$$

where $\zeta(N')$ is independent of ε. Combining (3.146) and (3.147), we obtain

$$H_\varepsilon(\mathscr{X}_{\mathrm{B}}) \geq N' \left[\log\left(\zeta(N')\frac{1}{\varepsilon}\right)\right]. \tag{3.148}$$

Similarly, by inequality (138) of Theorem 6 in [55] we have

$$H_\varepsilon(\mathscr{X}_{\mathrm{B}}) \leq H_{\varepsilon-\mu}(\mathscr{X}'_{\mathrm{B}}), \tag{3.149}$$

and using inequality (100) of Theorem 3 in [55], we have

$$H_{\varepsilon-\mu}(\mathscr{X}'_B) \leq N' \log\left(\frac{1}{\varepsilon-\mu}\right) + \eta(N'), \tag{3.150}$$

where $0 < \mu < \varepsilon$, and $\eta(N')$ is independent of ε. Combining (3.149) and (3.150), we obtain

$$H_\varepsilon(\mathscr{X}_\mathrm{B}) \leq N' \log\left(\frac{1}{\varepsilon - \mu}\right) + \eta(N'). \tag{3.151}$$

Since μ can be arbitrarily small and the logarithm is a continuous function, it follows that

$$H_\varepsilon(\mathscr{X}_\mathrm{B}) \leq N' \log(1/\varepsilon) + \eta(N'). \tag{3.152}$$

Putting together (3.148) and (3.152), we finally obtain

$$\begin{cases} H_\varepsilon(\mathscr{X}_\mathrm{B}) & \geq N' \log[\zeta(N')1/\varepsilon], \\ H_\varepsilon(\mathscr{X}_\mathrm{B}) & \leq N' \log(1/\varepsilon) + \eta(N'). \end{cases} \tag{3.153}$$

Dividing both sides of (3.153) by $-\log\varepsilon$ and taking the limit for $\varepsilon \to 0$, we have

$$\dim_F(\mathscr{X}_\mathrm{B}) = N', \tag{3.154}$$

so that

$$\lim_{T\to\infty} \frac{\dim_F(\mathscr{X}_\mathrm{B})}{T} = \Omega/\pi. \tag{3.155}$$

3.9 Acknowledgments

Chapter 3, in full, has been submitted for publication of the material as it might appear in T. J. Lim and M. Franceschetti, “Deterministic coding theorems for blind sensing: optimal measurement rate and fractal dimension”. The dissertation author was primary investigator and co-author of this paper.

Bibliography

[1] C. Shannon, “A mathematical theory of communication,” *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 1948.

[2] A. N. Kolmogorov, “On certain asymptotic characteristics of completely bounded metric spaces,” *Uspekhi Matematicheskikh Nauk*, 108(3), pp. 385–388, 1956. (In Russian)

[3] A. N. Kolmogorov, V. M. Tikhomirov, “ε-entropy and ε-capacity of sets in functional spaces,” *Uspekhi Matematicheskikh Nauk*, 14(2), pp. 3–86, 1959. English translation: *American Mathematical Society Translation Series*, 2(17), pp. 277–364, 1961.

[4] A. N. Kolmogorov, “ Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. ”*Annals of Mathematics* 37(1), pp. 107-110, 1936. (In German).

[5] A. N. Kolmogorov, “On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition.” *Doklady Akademii Nauk SSSR*, 114, pp. 953-956, 1957. (In Russian).

[6] D. L. Donoho, “Wald lecture I: counting bits with Shannon and Kolmogorov, Technical Report, Stanford University, 2000.

[7] T. M. Cover, J. Thomas, “Elements of information theory.” Second edition. *J. Wiley and sons*, 2006.

[8] T. M. Cover, P. Gacs, R. Gray, “Kolmogorov’s contributions to information theory and algorithmic complexity.” *The Annals of Probability*, 17(3), pp. 840-865, 1989.

[9] A. N. Kolmogorov, “On the Shannon theory of information transmission in the case of continuous signals.” *IRE Transactions on Information Theory*, 2(4), pp.102-108, December 1956.

[10] D. Slepian, “On Bandwidth.” *Proceedings of the IEEE*, 64(3), pp. 292-300, March 1976.

[11] D. Slepian, “Some comments on Fourier analysis, uncertainty and modeling,” *SIAM Review*, vol. 25, no. 3, pp. 379–393, 1983.

[12] A. S. Matveev and A. V. Savkin, "Shannon zero-error capacity in the problems of state estimation and stabilization via noisy communication channels." *International Journal of Contro*l, 80, pp. 241–255, 2007.

[13] G. Nair, "A non-stochastic information theory for communication and state estimation," *IEEE Transactions on automatic control*, vol. 58, pp. 1497–1510, 2013.

[14] M. Franceschetti, "On Landau's eigenvalue theorem and information cut-sets." *IEEE Transactions on Information Theory*, 61(9), pp. 5042-5051, 2015.

[15] M. Franceschetti, M. D. Migliore, P. Minero, and F. Schettino, "The degrees of freedom of wireless networks via cut-set integrals." *IEEE Transactions on Information Theory*, 57(11), pp. 3067-3079, 2011.

[16] M. Franceschetti, M. D. Migliore, and P. Minero, "The capacity of wireless networks: information-theoretic and physical limits." *IEEE Transactions on Information Theory*, 55(8), pp. 3413–3424, 2009.

[17] C. Shannon, "The zero-error capacity of a noisy channel," *IRE Transactions on Information Theory*, 2(3), pp. 8–19, 1956.

[18] D. Jagerman, "ε-entropy and approximation of bandlimted functions," *SIAM Journal of Applied Mathematics.*, vol. 17, no. 2, pp. 362–376, 1969.

[19] D. Jagerman, "Information theory and approximation of bandlimted functions," *Bell Systems Technical Journal*, 49(8), pp. 1911–1941, 1970.

[20] A. Wyner . A bound on the number of distinguishable functions which are time-limited and approximately band-limited. *SIAM Journal of Applied Mathematics* 24(3), pp. 289-297, 1973.

[21] I. Dumer, M. S. Pinsker and V. V. Prelov, "On coverings of ellipsoids in Eucliden spaces." *IEEE Transactions on Information Theory*, 50(10), pp. 2348–2356, 2004.

[22] C. Flammer, "Spheroidal wave functions," *Stanford University Press,* 1957. Reprinted 2005, *Dover.*

[23] H. J. Landau, H. Widom, "The eigenvalue distribution of time and frequency limiting," *Journal of Mathematical Analysis and Applications*, vol. 77, pp. 469–481, 1980.

[24] D. Slepian. "Some asymptotic expansions for prolate spheroidal wave functions." *Journal of mathematics and physics*, 44, pp. 99-140, 1965.

[25] A. Pinkus, "n-Widths in Approximation Theory," *Springer-Verlag*, 1985.

[26] O. Bucci, G. Franceschetti. On the spatial bandwidth of scattered fields. *IEEE Trans. on Antennas and Propagation,* 35(12), pp. 1445-1455, December 1985.

[27] O. Bucci, G. Franceschetti. On the degrees of freedom of scattered fields. *IEEE Trans. on Antennas and Propagation,* 37(7), pp. 918-926, July 1989.

[28] R. G. Gallager. "Information theory and reliable communication." *J. Wiley and sons*, 1968.

[29] C. Zong, *Sphere Packings*, Springer, 2013.

[30] C. A. Rogers, "Covering a sphere with spheres", Mathematika, 10, pp. 157–164, 1963.

[31] B. S. Mitjagin. "The approximative dimension and bases in nuclear spaces," Uspekhi Matematicheskikh Nauk, 16(4), pp. 63-132, 1961. English translation: *Russian Mathematical Surveys*. 16(4). pp. 59-127, 1961.

[32] G. Lorentz, "Approximation of Functions," *AMS Chelsea Publishing*, second edition, 1986.

[33] A. Jerri. "The Shannon sampling theorem —its various extensions and applications: a tutorial review." *Proceedings of the IEEE*, 65(11), pp. 1565–1596, 1977.

[34] M. Unser. "Sampling —50 years after Shannon." *Proceedings of the IEEE*, 88(4), pp. 569-587, 2000.

[35] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions" *Acta Math.*, vol. 117, pp. 37-52, Feb. 1967.

[36] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 3, pp. 1688–1691, May 1996.

[37] Y. Bresler and P. Feng, "Spectrum-blind minimum-rate sampling and reconstruction of 2-d multiband signals," in *Proc. IEEE Int. Conf. Image Process*, vol. 1, pp. 701–704, Sep. 1996.

[38] R. Venkataramani and Y. Bresler, "Further results on spectrum blind sampling of 2D signals," in *Proc. IEEE Int. Conf. Image Process*, vol. 2, pp. 752–756, Oct. 1998.

[39] M. Mishali and Y. Eldar, "Blind multi-band signal reconstruction: compressed sensing for analog signals," in *IEEE Trans. Signal Process*, 57 (3), pp. 993–1009, 2009.

[40] M. Mishali and Y. Eldar, "From theory to practice: sub-Nyquist sampling of sparse wide-band analog signals," in *IEEE J. Select. Top. Signal Process.*, 4 (2), pp. 375–391, 2010.

[41] S. Izu and J. Lakey, "Time-frequency localization and sampling of multiband signals," in *Acta Appl. Math.*, 107 (1), pp. 399–435, 2009.

[42] E. Sejdic, M. Luccini, S. Primak, K. Baddour, T. Willink, "Channel estimation using DPSS based frames," in *Proc. IEEE Conf. Accoust., Speech, and Signal Processing (ICASSP)*, Las Vegas, Nevada, March 2008.

[43] M. A. Davenport and M. B. Wakin, "Compressive sensing of analog signals using Discrete Prolate Spheroidal Sequences," in *Appl. Comput. Harmon. Anal.,*, 33, pp. 438-472, 2012.

[44] Y. Wu and S. Verdu, "Renyi information dimension: Fundamental limits of almost lossless analog compression." *IEEE Trans. on Information Theory*, 56(8), pp. 3721-3748, 2010.

[45] Y. Wu and S. Verdu, "Optimal phase transitions in compressed sensing," *IEEE Trans. on Information Theory*, 58,(10), pp. 6241-6263, 2012.

[46] K. J. Falconer, "Fractal geometry: mathematical foundations and applications," *John Wiley & Sons*, 2004.

[47] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," *Springer*, New York, 2013.

[48] S. Gleichman and Y. Eldar, "Blind compressed sensing," in *IEEE Trans. on Information Theory,* 57(10), pp. 6958-6975, 2011.

[49] S. Ravishankar and Y. Bresler, "Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to MRI," in *SIAM J. Imaging Sciences*, 8(4), pp. 2519-2557, 2015.

[50] A. Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Hungarica*, 10(12), pp. 193-215,

[51] H. J. Landau, "An Overview of Time and Frequency Limiting." Fourier techniques and applications," J. F. Price (editor), Plenum, New York, pp. 201-220, 1985.

[52] K. J. Falconer, "Dimensions and measures of quasi self-similar sets." *Proceedings of the American mathematical society* 106.(2), pp. 543-554, 1989.

[53] E. Olson, "Bouligand dimension and almost Lipschitz embeddings." *Pacific journal of mathematics*, 202.(2), pp. 459-474, 2002

[54] J.-L. Verger-Gaugry, “Covering a ball with smaller equal balls in $\mathbb{R}^n$”. Discrete and Computational Geometry,” Springer Verlag, 33, pp.143-155, 2005.

[55] T. J. Lim and M. Franceschetti, “Information without rolling dice,” *IEEE Trans. on Information Theory,* 63(3), pp. 1349-1363, 2017.

[56] D. L. Donoho, A. Javanmard, A. Montanari. “Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing”. *IEEE Trans. on Information Theory,* 59(11), pp. 7434 - 7464, 2013.