SIGNAL REPRESENTATION AND RECOVERY UNDER MEASUREMENT CONSTRAINTS

A DISSERTATION SUBMITTED TO
THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Ayça Özçelikkale Hünerli

September, 2012

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Dr. Haldun M. Özaktaş (Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Dr. Erdal Arıkan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Dr. Gözde Bozdağı Akar

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Dr. A. Enis Çetin

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Assist. Prof. Dr. Selim Aksoy

Approved for the Graduate School of Engineering and Science:

Prof. Dr. Levent Onural
Director of the Graduate School

ABSTRACT

SIGNAL REPRESENTATION AND RECOVERY UNDER MEASUREMENT CONSTRAINTS

Ayça Özçelikkale Hünerli
Ph.D. in Electrical and Electronics Engineering
Supervisor: Prof. Dr. Haldun M. Özaktaş
September, 2012

We are concerned with a family of signal representation and recovery problems under various measurement restrictions. We focus on finding performance bounds for these problems where the aim is to reconstruct a signal from its direct or indirect measurements. One of our main goals is to understand the effect of different forms of finiteness in the sampling process, such as finite number of samples or finite amplitude accuracy, on the recovery performance. In the first part of the thesis, we use a measurement device model in which each device has a cost that depends on the amplitude accuracy of the device: the cost of a measurement device is primarily determined by the number of amplitude levels that the device can reliably distinguish; devices with higher numbers of distinguishable levels have higher costs. We also assume that there is a limited cost budget so that it is not possible to make a high amplitude resolution measurement at every point. We investigate the optimal allocation of cost budget to the measurement devices so as to minimize estimation error. In contrast to common practice which often treats sampling and quantization separately, we have explicitly focused on the interplay between limited spatial resolution and limited amplitude accuracy. We show that in certain cases, sampling at rates different than the Nyquist rate is more efficient. We find the optimal sampling rates, and the resulting optimal error-cost trade-off curves. In the second part of the thesis, we formulate a set of measurement problems with the aim of reaching a better understanding of the relationship between geometry of statistical dependence in measurement space and total uncertainty of the signal. These problems are investigated in a mean-square error setting under the assumption of Gaussian signals. An important aspect of our formulation is our focus on the linear unitary transformation that relates the canonical signal domain and the measurement domain. We consider measurement set-ups in which a random or a fixed subset of the signal components in the measurement space are erased. We investigate the error performance, both in the average, and also in terms of guarantees that hold with high probability, as a function of system parameters. Our investigation also reveals a possible relationship between the concept of coherence of random fields as defined in optics, and the concept of coherence of bases as defined in compressive sensing, through the fractional Fourier transform. We also consider an extension of our discussions to stationary Gaussian sources. We find explicit expressions for the mean-square error for equidistant sampling, and comment on the decay of error introduced by using finite-length representations instead of infinite-length representations.

Keywords: inverse problems, estimation, signal representation, signal recovery, sampling, spatial resolution, amplitude resolution, coherence, compressive sensing, discrete Fourier transform (DFT), fractional Fourier transform, mixing, wave-propagation, optical information processing.

ÖZET

ÖLÇÜM KISITLARI ALTINDA İŞARET TEMSİLİ VE GERİ KAZANIMI

Ayça Özçelikkale Hünerli
Elektrik ve Elektronik Mühendisliği, Doktora
Tez Yöneticisi: Prof. Dr. Haldun M. Özaktaş
Eylül, 2012

Çeşitli ölçüm kısıtları altında işaret temsili ve geri kazanımı problemleri ile ilgileniyoruz. İşaretlerin doğrudan, ya da dolaylı ölçümlerinden geri kazanılmasının amaçlandığı bu problemler için performans sınırlarını bulmak üstüne yoğunlaşıyoruz. Temel amaçlarımızdan biri sonlu sayıda ölçüm alınması ya da genlik ölçüm hassasiyetinin sonlu olması gibi farklı sonluluk biçimlerinin geri kazanım performansına etkisini anlamaktır. Tezin ilk kısmında, her cihazın sağladığı ölçüm hassasiyetine bağlı bir maliyetle ilişkilendirildiği bir ölçüm cihazı modeli kullanıyoruz: bir ölçüm cihazının maliyeti esas olarak ayırt edebildiği genlik seviyesi sayısı tarafından belirlenir; daha yüksek hassasiyete sahip cihazların maliyetleri daha yüksektir. Ayrıca her noktada yüksek hassasiyetle ölçüm yapmamızı olanaksız kılan bir maliyet bütçemiz olduğunu varsayıyoruz. İşaretin en iyi şekilde kestirilebilmesi için bütçenin ölçüm cihazlarına en iyi şekilde nasıl bölüştürülmesi gerektiğini araştırıyoruz. Örnekleme ve nicemlemeyi ayrı ayrı ele alan yaygın uygulamanın aksine, uzaydaki ve genlikteki çözünürlüklerin arasındaki etkileşime özellikle yoğunlaşıyoruz. Nyquist hızından farklı hızlarda örnekleme yapmanın bazı durumlarda daha etkili olduğunu gösteriyoruz. Eniyi örnekleme hızlarını, ve sonuçta ortaya çıkan hata-maliyet ödünleşim eğrilerini buluyoruz. Tezin ikinci kısmında, ölçüm uzayındaki istatiksel bağımlılığın geometrisi ile işaretin toplam belirsizliği arasındaki ilişkiyi daha iyi anlamayı amaçlayan bir grup ölçüm problemi kuruyoruz. Bu problemleri bilinmeyen sinyalin Gauss istatistiklere sahip olduğu varsayımı altında ortalama karesel hata ölçütü çerçevesinde inceliyoruz. Kurduğumuz çerçevenin önemli özelliklerinden biri sinyal uzayı ile ölçüm uzayını ilişkilendiren birimcil dönüşüme yoğunlaşmış olmamızdır. Sinyalin bileşenlerinden rasgele seçilmiş ya da sabit bir kısmının ölçüm uzayından silindiği ölçüm senaryolarını ele alıyoruz. Hata performansını, sistem parametreleri cinsinden, hem ortalama hata hem de yüksek olasılıkla tutan performans garantileri cinsinden araştırıyoruz. Çalışmamız kesirli Fourier dönüşümü yoluyla, optikte tanımlanmış olan bir rasgele surecin uyumluluk derecesi kavramı ile sıkıştırmalı algılama alanında tanımlanmış olan bir dönüşümün uyumluluk derecesi kavramları arasındaki muhtemel ilişkiyi de ortaya çıkarıyor. Tartışmalarımızın durağan Gauss kaynaklara genişletilmesini de ele alıyoruz. Eşit aralıklı örnekleme icin ortalama karesel hatanın açık ifadesini buluyoruz, ve işaretin temsilinde sonsuz uzunlukta betimlemeler yerine sonlu uzunlukta betimlemenin kullanılması ile ortaya çıkan hatanın azalışı konusunda yorumlar yapıyoruz.

Anahtar sözcükler: ters problemler, kestirim, işaret temsili, işaret geri kazanımı, örnekleme, uzamsal çözünürlük, genlikteki çözünürlük, uyumluluk, sıkıştırmalı algılama, kesirli Fourier dönüşümü, ayrık Fourier dönüşümü (DFT), karıştırma, dalga yayılımı, optik bilgi işleme.

Acknowledgement

I would like to express my sincere gratitude to Prof. Dr. Haldun M. Özaktaş for his supervision, encouragement and guidance throughout the development of this thesis. I would like to thank the members of my thesis monitoring committee Prof. Dr. Erdal Arıkan and Prof. Dr. Gözde Bozdağı Akar for their valuable feedback on my research and their guidance. I would like to thank Assist. Prof. Dr. Serdar Yüksel for his guidance and encouragement which he has been offering continually since our first collaborative work.

I would like to thank my thesis defense committee members Prof. Dr. A. Enis Çetin and Assist. Prof. Dr. Selim Aksoy for accepting to read the manuscript and commenting on the thesis. I would also like to thank Prof. Dr. Orhan Arıkan and Assist. Prof. Dr. Sinan Gezici for their insightful comments on my research.

I would like to acknowledge the support provided by The Scientific and Technological Research Council of Turkey (TUBITAK), in particular the financial support provided through the scholarships BİDEB-2211 and BİDEB-2214.

I would like to thank Kıvanç Köse, Mehmet Köseoğlu, Aslı Ünlügedik, Alexander Suhre, Sinan Taşdelen, Erdem Ulusoy, İpek İstanbulluoğlu, Alican Bozkurt, Erdem Şahin, Ali Özgür Yöntem, Bilge Köse, Gökhan Bora Esmer, Ahmet Serdar Tan, Yiğitcan Eryaman, Esra Abacı, Elif Aydoğdu, Namık Ṣengezer, Prabeen Joshi, and Anne Gut for their friendship and support. I would also like to thank our department secretary Mürüvet Parlakay for all the support and the humor she offers. Kıvanç Köse, Mehmet Köseoğlu and Alexander Suhre deserve a special mention for their friendship, encouragement, and their tolerance, especially during the last months. I would also like to thank Aslı Ünlügedik for our close friendship that has been a silent but huge source of support for me in the last years. Special thanks go to Aslı Köseoğlu for helping us to bring things into perspective, and laugh even in the worst of times.

It is a great pleasure to express my gratitude to my parents, Gülsen, and Levent Özçelikkale, and to my brother Altuğ Özçelikkale. Without their love

and patience, this work would not have been possible. Special thanks go to my mother for her unconditional love and encouragement. It has been a real blessing to feel that she supports me no matter what.

I would like to thank my husband H. Volkan Hünerli for his love, support, encouragement and endless patience. He has believed in me and continually encouraged me to pursue my dreams as a researcher. He has made life joyful for me even in the period of writing the manuscript of this thesis. I can only hope that when the time comes, I could be as supportive as he has been.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The problems addressed in this thesis are centered around sampling and representation of signals under various restrictions. We focus on finding performance bounds for a class of signal representation or recovery problems where one wants to reconstruct a signal from its direct or indirect measurements. One of our main aims is to understand the effect of different forms of finiteness in the sampling process, such as finite number of measurements or finite amplitude accuracy in measurements, on the recovery performance.

1.1 Motivation and Overview

We will now discuss some issues related to sampling of signals that have motivated us to formulate the problems considered in this thesis. When a signal is to be represented with its samples, the Shannon-Nyquist sampling theorem is often used as a guideline. The theorem states that a band-limited signal with maximum frequency $B/2$ Hertz can be recovered from its equidistant samples taken $1/B$ apart [1, Ch. 7]. In practice, signals may not be exactly band-limited, but rather effectively band-limited in the sense that the signal energy beyond a certain frequency is negligible. In such cases, the effective bandwidth is often used to determine a sampling interval. Another practical constraint is the impossibility of taking an infinite number of samples. Thus, it is common to determine an effective spatial extent L in the sense that the signal energy is negligible outside this extent, and use only the samples that fall in this effective spatial extent. This approach leaves us with a finite number LB of samples. This approach may not always be the most appropriate manner in which to use the Shannon-Nyquist sampling theorem; there may be cases where one can do better by incorporating other available information. In particular, consider the practical scenario where the field is to be represented with a finite number of finite accuracy samples. Use of the conventional approach in this scenario raises a number of issues. For one thing, the concept of effective bandwidth and effective spatial extent is intrinsically ambiguous, in that there is some arbitrariness in deciding beyond what point the signal may be assumed negligible. This approach also completely ignores the fact that the samples will have limited amplitude accuracy. When we are required to represent the signal with a prespecified number of bits, the sampling interval dictated by the conventional sampling theorem may not be optimal. For instance, depending on the circumstances, it may be preferable to work with a larger sampling interval and a higher number of amplitude levels. In order to find the optimal values of these parameters, we must abandon the conventional approach and jointly optimize over the sampling interval and amplitude accuracies. Even when the amplitude accuracies are so high that we can assume the sample values to be nearly exact, the conventional sampling theorem may still not predict the optimal sampling interval if we are required to represent the signal with a given finite number of samples (especially when that number is relatively small).

Motivated by these observations, we have formulated a set of signal recovery problems under various restrictions. We now provide a brief overview of these problems.

Firstly, we investigate the effect of restriction of the total number of samples to be finite while representing a random field using its samples. Here we assume that the amplitude accuracies are so high that the sample values can be assumed to be exact. In Chapter 2, we pose this problem as an optimal sampling problem where, for a given number of samples, we seek the optimal sampling interval in order to represent the field with as low error as possible. We obtain the optimum sampling intervals and the resulting trade-offs between the number of samples and the representation error. We deal with questions such as "What is the minimum error that can be achieved with a given number of samples?", and "How sensitive is the error to the sampling interval?" [2].

In Chapter 3, we focus on the effect of limited amplitude accuracy of the measurements in signal recovery. Here we work with a limited amplitude accuracy measurement device model which was proposed in [3–6]. Here each device has a cost that depends on the amplitude accuracy the device provides. The cost of a measurement device is primarily determined by the number of amplitude levels that the device can reliably distinguish; devices with higher numbers of distinguishable levels have higher costs. We also assume that there is a limited cost budget so that it is not possible to make a high amplitude resolution measurement at every point. We investigate the optimal allocation of cost budget to the measurement devices so as to minimize estimation error. Our investigation reveals trade-off curves between the estimation error and the cost budget. This problem differs from standard estimation problems in that we are allowed to "design" the noise levels of the measurement devices subject to the cost constraint. Incorporation of limited amplitude accuracy into our framework through cost constraints reveals an opportunity to make a systematic study. Another important aspect of the formulation here is the cost function we use: while this kind of cost function may come as natural in the context of communication costs, we believe it has not been used to model the cost of measurement devices until [3–6].

We extend the cost budget approach presented in a discrete framework in Chapter 3, to a continuous framework in Chapters 4-5. Here we deal with signals which are functions of continuous independent variables. We consider two main sampling strategies: i) uniform sampling with uniform cost allocation ii) non-uniform sampling with non-uniform cost allocation. In the first of these we consider an equidistant sampling approach, where each sample is taken with the same amplitude accuracy. We seek the optimal number of samples, and sampling interval under a given cost budget in order to recover the signal with as low error as possible. Our investigation illustrates how the sampling interval should be optimally chosen when the samples are of limited amplitude accuracy, in order to achieve best error values possible. We illustrate that in certain cases sampling at rates different than the Nyquist rate is more efficient [7,8]. In the second formulation, which is studied in Chapter 5, we consider a very general scenario where the number, locations and accuracies of the samples are optimization variables. Here the sample locations can be freely chosen, and need not be equally spaced from each other. Furthermore, the measurement accuracy of each sample can vary from sample to sample. Thus this general non-uniform case represents maximum flexibilty in choosing the sampling strategy. We seek the optimal values of the number, locations and accuracies in order to achieve the lowest error values possible under a cost budget. Our investigation illustrates how one can exploit the better optimization opportunity provided by the flexibility of choosing these variables freely, and obtain tighter optimization of the error-cost curves.

An important future of all the above work is the non-stationary signal model. A broad class of physical signals may be better represented with non-stationary models rather than stationary models, which has resulted in increasing interest in these models [9]. Although some aspects of the sampling of non-stationary fields are understood, such as the sampling theorem of [10], our understanding of non-stationary fields falls short of our understanding of stationary fields. One of our goals is to contribute to a better understanding of the trade-offs in the representation of non-stationary random fields.

We study an application of the cost budget approach developed in previous chapters to super-resolution problems in Chapter 6. In a typical super-resolution problem, multiple images with poor spatial resolution are used to reconstruct an image of the same scene with higher spatial resolution [11]. Here we study the effect of limited amplitude resolution (pixel depth) in this problem. In standard super-resolution problems, the researchers mostly focus on increasing resolution in space, whereas in our study both resolution in space and resolution in amplitude are substantial parameters of the framework. We study the trade-off between the pixel depth and spatial resolution of low resolution images in order to obtain the best visual quality in the reconstructed high resolution image. The proposed framework reveals great flexibility in terms of pixel depth and number of low resolution images in super-resolution problem, and demonstrates that it is possible to obtain target visual qualities with different measurement scenarios including images with different amplitude and spatial resolutions [12].

During the above studies, the following two intuitive concepts have been of central importance to our investigations: i) total uncertainty of the signal, ii) geometry of statistical dependence (spread of signal uncertainty) in measurement space. We note that the concepts that are traditionally used in the signal processing and information theory literatures as measures of dependency or uncertainty of signals (such as the degree of freedom, or the entropy) mostly refer to the first of these, which is defined independent of the coordinate system in which the signal is to be measured. As an example one may consider the Gaussian case: the entropy solely depends on the eigenvalue spectrum of the covariance matrix, hence making the concept blind to the coordinate system the signal will be measured.

Our study of the measurement problems described above suggests that although the optimal measurement strategies and signal recovery performance depends substantially on the first of these parameters (total uncertainty of the signal); the second of these concepts (geometry of statistical dependence in measurement space) also plays an important role in the measurement problem. In a measurement scenario, one would typically expect that the optimal measurement strategy (the optimal number, locations, and accuracies of the measurements) depends on how the total uncertainty of the signal source is spread in the measurement domain. For instance, consider these two cases i) most of the uncertainty of the signal is carried by a few components in the measurement domain, ii) the signal uncertainty is somewhat uniformly spread in the measurement domain so that every component in the measurement domain gives some information about the others. For the first of these, one would intuitively expect that the strategy of measuring only these few components with high accuracies will perform well. On the other hand, for the second case, one would expect that measuring a higher number of components with lower accuracies may give better results. Moreover, for the first case one would expect the measurement performance to substantially depend on the locations of the measurements compared to the second case; in the first case it would be important to particularly measure the components that carry most of the uncertainty, whereas in the second case measurements will be, informally speaking, interchangeable.

As illustrated above, the total uncertainty of the signal as quantified by information theoretic measures such as entropy and the geometry of spread of this uncertainty in measurement domain, reflect different aspects of the statistical dependence in a signal. In the second part of this thesis, we have formulated various problems investigating different aspects of this relationship. This line of study also relates to the compressive sensing paradigm, where measurement of sparse signals is considered [13, 14]. The signals that can be represented with a few coefficients after passing through a suitable transform, such as wavelet or Fourier are called sparse. It has been shown that such signals can be recovered from a few randomly located measurements if they are measured after passing through a suitable transform [13, 14]. Contrary to the deterministic signal models commonly employed in compressive sensing, here we work in a stochastic framework based on the Gaussian vector model and minimum mean square error (MMSE) estimation; and investigate the spread of eigenvalue distribution of the covariance matrix as a measure of sparsity. We assume that the covariance matrix of the signal; hence, location of support of the signal is known during estimation.

We first relate the properties of the transformation that relates the canonical signal domain and the measurement domain with the total correlatedness of the field in Chapter 7. In particular, we investigate the relationship between the following two concepts: degree of coherence of a random field as defined in optics and coherence of bases as defined in compressive sensing. Coherence is a concept of central importance in the theory of partially coherent light, which is a well-established area of optics; see for example [15, 16] and the references therein. Coherence is a measure of the overall correlatedness of a random field [15, 16]. One says that a random field is highly coherent when its values at different points are highly correlated with each other. Hence intuitively, when a field is highly coherent, one will need fewer samples to have good signal recovery guarantees. Compressive sensing problems heavily make use of the notion of coherence of bases, for example [13, 14, 17]. The coherence of two bases, say the intrinsic orthogonal signal domain ψ, and the orthogonal measurement system ϕ is measured with $\mu = \max_{i,j} |U_{ij}|$, $U = \phi\psi$ providing a measure of how concentrated the columns of U are. When μ is small, one says the mutual coherence is small. As the coherence gets smaller, fewer samples are required to provide good signal recovery guarantees. In Chapter 7, we illustrate that these two concepts, named exactly the same, but attributes of different things (bases and random fields), important in different areas (compressive sensing and statistical optics), and yet enabling similar type of conclusions (good signal recovery performance) are in fact connected. We also develop an estimation based framework to quantify coherence of random fields; and show that what this concept quantifies is not just a repetition of what more traditional concepts like the degree of freedom or the entropy does. We also study the fractional Fourier transform (FRT) in this setting. The FRT is the fractional operator power of the Fourier transform with fractional order a [18]. When $a = 0$, the FRT matrix reduces to the identity, and when $a = 1$ it reduces to the ordinary DFT matrix. We demonstrate how FRT can be used to generate both bases or statistics for fields with varying degrees of coherence; by changing the order of FRT from 0 to 1, it is possible to generate bases and statistics for fields with varying degree of coherence.

Our work in Chapter 7 can be interpreted as an investigation of basis dependency of MMSE under random sampling. In Chapter 8, we study this problem from an alternative perspective. We consider the transmission of a Gaussian vector source over a multi-dimensional Gaussian channel where a random or a fixed subset of the channel outputs are erased. We consider the setup where the only encoding operation allowed is a linear unitary transformation on the source. For such a setup, we investigate the MMSE performance, both in the average and also in terms of guarantees that hold with high probability, as a function of system parameters. Necessary conditions for optimal unitary encoders are established, and explicit solutions for a class of settings are presented. Although there are observations (including evidence provided by the compressed sensing community) that may suggest the result that the discrete Fourier transform (DFT) matrix may be indeed an optimum unitary transformation for any eigenvalue distribution, we provide a counterexample. Finally, we consider equidistant sampling of circularly wide-sense stationary (c.w.s.s.) signals, and present an upper bound that summarizes the effect of the sampling rate and the eigenvalue distribution. We have presented our findings here in [19, 20].

In Chapter 9, we continue our investigation of dependence in random fields with stationary Gaussian sources defined on $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$. We formulate various problems related to the finite-length representations and sampling of these signals. Our framework here is again based on our vision of understanding the effect of different forms of finiteness in representation of signals, and measures of dependence in random fields, in particular spread of uncertainty. We first consider the decay rates for the error between finite dimensional representations and infinite dimensional representations. Here our approach is based on the notion of mixing which is concerned with dependence in asymptotical sense, that is the dependence between two points of a random process as the distance between these two points increases [21]. Based on this concept, we investigate the decay rates of error introduced by using a finite number of samples instead of an infinite number of samples in representation of these signals. We then consider the MMSE estimation of a stationary Gaussian source from its noisy samples. We first show that for stationary sources, for the purpose of calculating the MMSE based on equidistant samples, asymptotically circulant matrices can be used instead of original covariance matrices, which are Toeplitz. This result suggests that circularly wide-sense stationary signals in finite dimensions are more than an analogy for stationary signals in infinite dimensions: there is an operational relationship between these two signal models. Then, we consider the MMSE associated with estimation of a stationary Gaussian source on $\mathbb{Z}_+ = \{0, 1, \ldots\}$ from its equidistant samples on $\mathbb{Z}_+$. Using the previous result, we give the explicit expression for the MMSE in terms of power spectral density, which explicitly shows how the signal and noise spectral densities contribute to the error.

1.2 Background

The representation and recovery problems considered in this thesis can be related to works in a broad range of fields, including optics, estimation and sampling theory, and information theory. This section provides a brief review of related works in these areas.

One of our main motivations is to contribute to better understanding of information theoretical relationships in propagating wave-fields. The problems discussed in this thesis shed light to different aspects of problems arising in this context. We will first present a review of representative studies in this area. We will then discuss the literature in the general area of distributed estimation, where problems that can be related to our cost budget approach, with different motivations or methods, are considered. Finally, we will review some related work focusing on sampling and finite representations of random fields.

The linear wave equation is of fundamental importance in many areas of science and engineering. It governs the propagation of electromagnetic, optical, acoustic, and other kinds of fields. Although information relationships for wave-fields have been studied in all of these contexts, a substantial amount of work have been done in the context of optics.

One of the most widely used concepts in this area is the concept of degree of freedom (DOF). The terminology of the degree of freedom of a system has been discussed typically with reference to the number of spots in the input of an optical system that can be distinguished in the output of the optical system. This number of spots is called the number of resolvable spots. A resolvable spot can be interpreted to be a communication channel from the input plane of the system to the output plane. Hence the degree of freedom of a system is essentially the number of channels one can use to communicate using this optical system. Reference [22] is an early work that has been important for formulation of this approach, where a Gaussian spot is suggested as the best form for a spot due to its minimum uncertainty property. In this work, it is further suggested that these effectively Gaussian spots can be used to approximate the input field to analyse different optical systems. In [23, 24], the author derives the conclusion that an image formed by a finite pupil has finite degrees of freedom using the sampling theorem; and investigates practical limitations related to the DOF using the theory of the prolate spheroidal functions. In [25], the concepts of DOF and space-bandwidth product are compared, and DOF is concluded to be the fundamental invariant for optical systems. Reference [26] proposes a method for obtaining spatial super resolution by sacrificing of temporal resolution, based on the framework in [25]. Various works have investigated the DOF associated with various particular optical systems or set-ups, such as [27–30].

Reference [31] is a particularly important work which discusses the DOF in a stochastic framework, and proposes a DOF definition based on the coherent mode decomposition of the covariance function. [32] discusses the degree of freedom associated with a transform that can be described by a finite convolution operator in the context of its invertibility, and proposes a measure of ill-conditioning in the presence of noise. Some works have focused on studying different aspects of the space-bandwidth product, such as its proper definition [33], its applications to super-resolution [34], or its generalization to linear canonical transform domains [35]. Super-resolution in optics with special emphasis on the concept of space-bandwidth product is studied in detail in [36].

In [37], MacKay introduces an informal discussion of the concepts of *structural and metrical information*, which has found application in [3,22,38,39]. Mac Kay's informal discussions can be interpreted as a claim that the degree of freedom is intrinsically related to structural information. It is argued that a signal can be approximated as a sum of the structural elements, whose number is given by the degree of freedom of the signal family. This work also introduces the concept of metrical information, which is defined as a measure of amplitude accuracy. It is argued that total information in the signal is given by the sum of metrical information and structural information. It is interesting to note that how this argument resembles how the rate-distortion function for a correlated Gaussian vector is found: the minimum number of bits required to represent such a signal under a given distortion is found by using finite accuracy components in the canonical domain (the domain the components are independent) [40, Ch.13]. Here the concept of metrical information can be said to correspond to finite accuracy in each of these components, and the concept of structural information can be associated with the concept of canonical domain, and the number of components used in the representation (effective degree of freedom).

References [41–43] adopt a particularly interesting approach to understand the limits of information transfer by optical fields: "communication modes expansion". The properties of these type of expansions and applications of them to different optical systems have been studied in many works, such as [44, 45]. This approach is based on appropriately defining so called "communication modes" between two volumes in space. One of these volumes is the volume which contains the scatter, and the other one is the receiving volume in which we want to generate waves. Then these works investigate the number of orthogonal functions that can be generated in the receiving volume as a result of scattering a wave from the scattering volume. The strength of connection between these two volumes is written as a sum of coupling strengths between the modes in scattering volume and the modes in receiving volume. This framework may be interpreted in the light of singular value decomposition of the linear optical system that relates the wave-fields between these two volumes, where communication modes correspond to the left and right singular vectors, and coupling strengths correspond to the eigenvalues. Such an approach brings a novel way to look at diffraction of optical fields based on the connection strengths between two volumes.

A number of works utilizing information theoretic concepts such entropy or channel capacity in different contexts have appeared. [46] studies the information relationships for imaging in the presence of noise with particular emphasis on relating the information theoretical definitions of entropy and mutual information, to intuitive descriptions of information based on physical models. Using the capacity expression for the Gaussian channel, which only depends on the signal-to-noise ratio, and ignoring the possible statistical dependency among pixels, [47] discusses information capacities of two-dimensional optical low-pass channels. [48] adopts a similar approach where the capacity definition is the same, but uses the degree of freedom associated with the system rather than the individual pixels at the input/output image planes. [49, 50], explicitly utilizes the idea of an error threshold, within which the signals are considered to be indistinguishable, in order to asses the information transfer capacity of waves. Under Gaussian signal assumption, [51] discusses the entropy of partially coherent light and its relationship between concepts that are traditionally used in optics to describe light fields, such as degree of polarization and coherence. The concept of entropy has also been studied in the context of acoustical waves [52, 53].

References [54] and [55] provide a general overview of the relationship between optics and information theory. To study optical systems from a communications perspective, these texts provide introductory material on a wide range of fields, including information theory, diffraction theory and signal analysis. The relationship between the concept of entropy in thermodynamics and entropy in information theory is thoroughly discussed. A discussion on the information provided by observations based on the wave nature of light and quantum theory is also presented. Several applications in the area of optical information processing including image restoration, wavelet transforms, pattern recognition, computing with optics and fiber-optic communications are also covered.

While utilizing information theoretic concepts in the study of propagating wave-fields, researchers do not always use concepts and terms exactly as they are traditionally used in the information theory literature. For example, in the context of information theory, entropy is defined as a measure of uncertainty of a random variable and is determined by the probability distribution function of the random source [56, Ch. 2], whereas this is not always how this concept is utilized in some works in optics. For instance, in some works the expression for the entropy of a discrete random vector in terms of its probability mass function is used to provide a measure for the spread of a set quantities one is interested in, such as the spread of eigenvalues associated with the coherent mode decomposition of a source [57, 58]. Other examples include References [59, 60], where the normalized point spread function is treated like a probability distribution function and the entropy is used to calculate the spread of this function providing a measure for its effective area [59], and normalized intensity distribution is used to define the spot size of a laser beam [60].

Some researchers have focused on computational issues, where the aim is to process the signals without losing any significant information, as well as by using as little computational resources as possible, such as [61–64]. Other works have adopted a sampling theory approach [65–68]. Reference [69] provides a review of many approaches used in information optics, including the approaches based on the sampling theory and the concept of DOF. An overview of the history of the subject with special emphasis on research which leads to practical progress can be found in [70, 71].

Historically, the approaches used to study information relationships in propagating wave-fields have commonly been based on scalar and paraxial approximations, or limited to investigating particular systems. Recently, a number of works have extended these approaches by either working with electromagnetic field models or more general system models which consider arbitrary volumes or regions in space. An example is the line of work developed in [41–43], which studies the communication between two volumes in space, and provides a very general framework as discussed above. Among these, with its electromagnetic field model and the extensions to space-variant systems it provides, [43] may be said to provide the most general perspective. Other works which make use of an electromagnetic field model include [49, 72–78]. Among these works, some have put particular emphasis to the restrictions imposed by antennas [74, 78]. For instance, based on a model that takes into account the spatial constraints put by antennas, [74] finds the degrees of freedom associated with a multiple antenna system where the degrees of freedom associated with the time-frequency domain and the spatial angular domain are treated in a unified manner. Unlike this approach, some works prefer to overlook the possible restrictions imposed by the receiving elements, and focus on the limitations imposed by the physical process. An example is [75], which is concerned with the degree of freedom of the system associated with communication with wave-fields where these wave-fields are to be observed in a bounded region in space. Reference [79] is another example where a framework independent of a particular transmismitter-receiver model is considered. This work considers the communication between two volumes in space as in [42], and may be interpreted as a generalization of this work to include the scenario where a scatterer may be present between these two volumes.

We now discuss the relationship of our cost budget framework with some earlier works which also involve estimation of desired quantities from measurements made from multiple sensors transmitting their observations to a decision centre. These works mostly adopt a communications perspective.

The cost constrainted measurements problem we have considered can also be interpreted in the framework of distributed estimation where there are uncooperative sensors transmitting their observations to a decision/fusion center. Such scenarios are quite popular and can be encountered in wireless sensor networks, one of the emerging technologies of recent years, or distributed robotics systems where the agents can only communicate to the fusion center. In a centralized sensor network, sensors with power, complexity and communication constraints sense a physical field and communicate their observations to the fusion/decision center, where the main aim is to reconstruct the field as accurately as possible. In this area, the design of sensor and fusion center strategies is intensively studied under various constraints. A number of works approach this problem as a quantizer design problem where the design of the optimum quantizers to be used by sensors is considered [80–82]. The performance of different distributed estimation systems are evaluated with various approaches, such as estimation of a parameter under a total sum rate constraint by focusing on quantizer bit rate allocation among sensors [83]. A particularly interesting work is the work in [84], where the measurement of one variable through multiple sensors is considered, and estimation performance is analysed under various performance criteria. Here estimation of a scalar variable (or a series of independent and identically distributed variables when time variation is also taken into account) is considered. Among these various scenarios, the one that addresses the problem of finding the optimal power allocation to sensor links to minimize estimation error, can be related to our optimal allocation problem. We note that, contrary to this work which considers estimation of a scalar quantity, in our framework desired quantities are modelled as functions of space where each measurement device has access to a field value only at a particular location. In this respect, we believe that our formulation models the problem of optimal estimation of a physical field from multiple measurements in a more realistic way. Moreover, with our model it is possible to systematically study the effect of coherence of the field on the results, which is a concept of central importance in optics.

A related problem, the distributed source coding problem arises in the framework of multiterminal source coding where the problem is formulated from a coding perspective. In the distributed source coding problem the aim is to determine the best coding strategy when there are uncooperative encoders coding their correlated observations and transmitting the coded versions to a centralized decoder where the observations are jointly decoded. The scheme of uncooperative encoders observing correlated sources was studied in [85] with two encoders and perfect reconstruction constraint. The rate-distortion function for such a scheme when only one of the sources is to be decoded is provided in [86]. A more explicit treatment of the continuous alphabet case is studied in [87]. The distributed source coding problem is widely studied under many constraints [88–91]. This field continues to be a popular area, where the explicit solutions are known only for a few cases; for instance the admissible rate region for two encoder quadratic Gaussian source coding problem is recently provided in [92].

Interpreting the measurement devices as encoders, and assuming the measurement device locations are fixed, we see that in both problems there is a distributed sensing scheme where correlated observations are separately processed and transmitted to a decision center where the messages are used to estimate the unknown variables. Moreover in both problems, the best strategies are determined a priori in a centralized manner, i.e. the coding strategies are based on the knowledge of statistics of what would be available to the all devices, but the encoders act without knowing what is available to the others at a particular instance of coding. Although these problems are closely connected, we now point out some distinctions. In a typical distributed source coding problem, the encoders have the freedom to observe the realizations of variables as long as they need, and they may do arbitrarily complex operations on their observations, whereas the measurement devices are restricted to observe only one realization of the variable to be measured and the message, (the reading of the device output) is restricted by the nature of the actual measurement devices. In source coding scheme there is no cost related to the accuracy of measuring the variable, but there is a communication cost, namely the finite rate related to the transmission of the observations to the decision center. To the contrary, in the measurement problem the cost is related to the accuracy of the measurements and the result of measurements are assumed to be perfectly transmitted to the decoder without any rate restriction. Hence, if the measurement problem is to be considered in a distributed source coding framework, it can be cast as a remote source coding problem where the encoders are constrained to have a policy of amplify and forward, with the cost of resolving power used as a dual for the communication cost.

In our cost-constrained measurement framework, what the measurement devices observe, are not necessarily the variables to be estimated. The fact suggests a connection with the problem of remote/noisy source coding. A simple example for this type of problems is provided in [93, p. 80]. This problem is studied by many authors, for instance [94, 95]. The constraints under which separability principles are applicable in remote source coding problems are also investigated, for instance [96–98]. A related problem, called the Centralized Executive Officer problem is formulated in [99, 100]. In this framework one is interested in estimating a data sequence which one cannot directly observe. The estimation is done based on the outputs of encoders that observe independently corrupted versions of the data sequence and encodes them uncooperatively. Each of these encoders must use noiseless but rate-constrained channels to transmit their observations to the centralized officer. Under a sum-rate constraint, one investigates the trade-off between the sum rate and the estimation error. An important special case of this problem is the so called quadratic Gaussian case, where a Gaussian signal is observed through Gaussian noise and the distortion metric is the mean-square error [100, 101].

The finite accuracy measurements problem is also closely related to analog-to-digital (A/D) conversion problems, where efficient representation of analog sources with finite number of bits is considered. Although in the measurement problem framework the sensors are not necessarily digital devices, they have finite resolving power which in fact corresponds to a finite number of meaningful bits in the readings of the measurement devices. Trade-offs similar to the ones considered in this thesis can also be studied in A/D conversion framework, such as in [102] where the dependence of accuracy of oversampled analog-to-digital conversion on the sampling interval and bit rate is investigated or as in [103], which focuses on the trade-offs between sampling rate and accuracy of the measurements for recovery of a band-limited signal.

To sum up, a number of works studying estimation of desired quantities from multiple measurements share some of the important features of our formulation, or formulate their problems in a context related to ours: The cost function we have proposed in [3–6] has been used to formulate various constrained measurement problems in [104, 105]. In [106, 107], problems related to wave propagation are studied with a statistical signal processing approach. The problem of finding optimal space and frequency coverage of samples for minimum bit representation of random fields is considered in [108] in a framework based on Shannon interpolation formula. Optimal quantizer design has been studied under communication constraints; for instance [109, 110]. A problem of sensor selection is considered in [111] as an estimation problem, and under given sensor performance and costs in [112] as a detection problem. The tradeoff between performance and total bit rate with a special emphasis on quantizer bit rates is studied in [82,83], where the estimation of a single parameter is considered. Trade-offs similar to our cost-error trade-offs are also studied in A/D conversion framework [102]. Although various aspects of the problem of sensing of physical fields with sensors is intensively studied by many authors as distributed estimation and distributed source coding problems, much of this work has loose connections with the underlying physical phenomena. There seems to be a disciplinary boundary between these works and the works that adopt a physical sciences point of view. A notable exception is the line of work developed in [113, 114], where the measurement of random acoustic fields is studied from an information-theoretic perspective with special emphasis on the power spectral density properties of these fields. Further work to bridge these two approaches will help us better understand the information theoretic relationships in physical fields and their measurement from a broader perspective.

Several aspects of sampling of random processes are studied by many researchers. Here we provide a brief overview of results that are pertinent to our work. A fundamental result in this area states that Shannon-Nyquist sampling theorem which is generally expressed for deterministic signals can be generalized to wide-sense stationary (w.s.s.) signals: A band-limited w.s.s. signal can be reconstructed in the mean-square sense from its equally-spaced samples taken at Nyquist rate [115]. In [116] a generalization of this result where possibly multi-band signals are considered is provided. Generalizations of this result where the samples differ from ordinary Nyquist samples are also considered: [117, 118] shows at most how much the sample points may be shifted before the error free recovery becomes impossible. A formal treatment of this subject with a broad view may be found in [118]. [119, 120] offer conditions under which of these generalizations (such as deletion of finitely many samples) error-free recovery is possible. An average sampling theorem for band-limited random signals is presented in [121]. In [122], the mean-square error of approximating a possibly non-bandlimited w.s.s. signal using sampling expansion is considered. [123, 124] focuses on a prediction framework where only the past samples are taken into account while estimating the signal. In [125], signal reconstruction with polynomial interpolators and Poisson sampling is studied. [10] further generalizes the Shannon-Nyquist sampling theorem to non-stationary random fields; [126] clarifies the conditions in [10]. [127, 128] consider problems related to the sampling of varying classes of non-stationary signals. Finite-length truncations in representation of random signals are studied in signal processing community under various formulations. In [129], the truncation error associated with the sampling expansion is studied. [130] focuses on the convergence behaviour of the sampling series. In [131, 132] the difference between the infinite horizon and finite horizon causal MMSE estimators (the estimator based on the last N values) are considered.

Part I

Optimal Representation and Recovery of Non-stationary Random Fields

Chapter 2

Representation and Recovery using Finite Numbers of Samples

In this chapter, we investigate the effect of restriction of the total number of samples to be finite while representing a random field using its samples. Here, we assume that the amplitude accuracies are so high that the sample values can be assumed to be exact. In Chapter 3, we will abandon this simplification, and consider a framework where the effect of limited amplitude accuracies of the samples are also taken into account.

We may summarize our general framework as follows: We consider equidistant sampling of non-stationary signals with finite energy. We are allowed to take only a finite number of samples. For a given number of samples, we seek the optimal sampling interval in order to represent the field with as low error as possible. We obtain the optimum sampling intervals and the resulting trade-offs between the number of samples and the representation error. We present results for varying noise levels and for sources with varying numbers of degrees of freedom. We discuss the dependence of the optimum sampling interval on the problem parameters. We also investigate the sensitivity of the error to the chosen sampling interval.

A crucial aspect of our formulation is the restriction of the total number of samples to be finite. Although several aspects of the sampling of random fields are well understood (mostly for stationary fields and also for non-stationary fields), most studies deal with the case where the number of samples *per unit time* is finite (and the total number of samples are infinite).

In Section 2.1, we present the mathematical model of the problem considered in this chapter. The signal model we use in our experiments, the Gaussian-Schell model, is discussed in Section 2.2. In Section 2.3 we present the numerical experiments. We conclude in Section 2.5.

2.1 Problem Formulation

In the specific measurement scenario under consideration in this chapter, a signal corrupted by noise is sampled to provide a representation of the signal with finite number of samples. More precisely, the sampled signal is of the form

$$g(x) = f(x) + n(x), \tag{2.1}$$

where $x \in \mathbb{R}$, $f : \mathbb{R} \to \mathbb{C}$ is the unknown proper Gaussian random field (random process), $n : \mathbb{R} \to \mathbb{C}$ is the proper Gaussian random field denoting the inherent noise, and $g : \mathbb{R} \to \mathbb{C}$ is the proper Gaussian random field to be sampled in order to estimate $f(x)$. We assume that $f(x)$ and $n(x)$ are statistically independent zero-mean random fields. We consider all signals and estimators in the bounded region $-\infty < x_L \leq x \leq x_H < \infty$. Let $D = [x_L, x_H]$ and $D^2 = [x_L, x_H] \times [x_L, x_H]$. Let $K_f(x_1, x_2) = E[f(x_1)f^*(x_2)]$, and $K_n(x_1, x_2) = E[n(x_1)n^*(x_2)]$ denote the covariance functions of $f(x)$ and $n(x)$, respectively. Here * denotes complex conjugation. We assume that $f(x)$ is a finite energy random field, $\int_{-\infty}^{\infty} K_f(x, x)dx < \infty$, and $K_n(x, x)$, $x \in D$ is bounded.

M samples of $g(x)$ are taken equidistantly with the sampling interval Δ_x at $x = \xi_1, \ldots, \xi_M \in \mathbb{R}$, with $\Delta_x = \xi_{i+1} - \xi_i$, $i = 1, ..., M-1$. Hence we have $g_i \in \mathbb{C}$ observed according to the model $g_i = g(\xi_i)$, for $i = 1, \ldots, M$. By putting the sampled values in vector form, we obtain $\mathbf{g} = [g(\xi_1), \ldots, g(\xi_M)]^{\mathrm{T}}$. Let $\mathbf{K_g} = E[\mathbf{g}\mathbf{g}^\dagger]$ be the covariance matrix of $\mathbf{g}$, † denotes the conjugate transpose.

The vector $\mathbf{g}$ provides a representation of the random field $f(x)$. We do not have access to the true field $f(x)$ but we can find $\hat{f}(x \mid \mathbf{g})$, the minimum mean-square error (MMSE) estimate of $f(x)$ given $\mathbf{g}$. For a given maximum allowed number of sampling points M_b, our objective is to choose the location of the samples ($\xi_1, \ldots, \xi_M \in \mathbb{R}$, $M \leq M_b$), so that the MMSE between $f(x)$ and $\hat{f}(x \mid \mathbf{g})$ is minimum.

This problem can be stated as one of determining

$$\varepsilon(M_b) = \min_{\Delta_x,\, x_0} E\left[\int_D \|f(x) - \hat{f}(x \mid \mathbf{g})\|^2 dx\right], \tag{2.2}$$

subject to $M \leq M_b$. Here the samples are taken around the midpoint $x_0 = 0.5(\xi_1 + \xi_M)$, which along with Δ_x we allow to be optimally chosen.

Noting that the observed values are in vector form, the linear estimator for (2.2) can be written as [133, Ch. 6]

$$\begin{aligned} \hat{f}(x \mid \mathbf{g}) &= \sum_{j=1}^{M} h_j(x) g_j & (2.3)\\ &= \mathbf{h}(x)\mathbf{g} & (2.4) \end{aligned}$$

where the function $\mathbf{h}(x) = [h_1(x), \ldots, h_M(x)]$ satisfies the equation

$$\mathbf{K}_{\mathbf{f}\,\mathbf{g}}(x) = \mathbf{h}(x)\mathbf{K}_{\mathbf{g}}, \tag{2.5}$$

where $\mathbf{K}_{f\,g}(x) = E[f(x)g^\dagger] = [E[f(x)g_1^*], \ldots, E[f(x)g_M^*]]$ is the cross covariance between the input field $f(x)$ and the measurement vector $\mathbf{g}$. To determine the optimal linear estimator, one should solve (2.5) for $\mathbf{h}(x)$.

The error expression can be written more explicitly as follows

$$\begin{aligned} \varepsilon &= E[\int_D \|f(x) - \mathbf{h}(x)\mathbf{g})\|^2 dx] & (2.6)\\ &= \int_D E[\|f(x) - \mathbf{h}(x)\mathbf{g})\|^2] dx & (2.7)\\ &= \int_D (K_f(x,x) - 2\mathbf{K}_{f\,\mathbf{g}}(x)\mathbf{h}(x)^\dagger + \mathbf{h}(x)\mathbf{K}_{\mathbf{g}}\mathbf{h}(x)^\dagger) dx & (2.8)\\ &= \int_D (K_f(x,x) - \mathbf{K}_{f\,\mathbf{g}}(x)\mathbf{h}(x)^\dagger) dx. & (2.9) \end{aligned}$$

Before leaving this section, we would like to comment on the error introduced by estimating $f(x)$ only in the bounded region D. Let us make the following definitions: Let $\hat{f}(x \mid \mathbf{g})$ be shortly denoted as $\hat{f}(x)$. Let us define $\hat{f}_D(x)$ as $\hat{f}_D(x) = \hat{f}(x)$ for $x \in D$ and $\hat{f}_D(x) = 0$ for $x \notin D$. Then, the error of representing $f(x)$ with $\hat{f}_D(x)$ can be expressed as

$$\begin{aligned}
&E[\int_{-\infty}^{\infty} \|f(x) - \hat{f}_D(x)\|^2 dx] \\
&= E[\int_{x \in D} \|f(x) - \hat{f}_D(x)\|^2 dx] + E[\int_{x \notin D} \|f(x) - \hat{f}_D(x)\|^2 dx] && (2.10) \\
&= E[\int_{x \in D} \|f(x) - \hat{f}_D(x)\|^2 dx] + E[\int_{x \notin D} \|f(x)\|^2 dx] && (2.11) \\
&= \varepsilon(M_b) + \int_{x \notin D} K_f(x, x) dx && (2.12)
\end{aligned}$$

Hence (2.12) states that the error of representing a field on the entire line can be expressed as the sum of two terms; the first term expressing the approximation error in this bounded region, and the second term expressing the error due to neglecting the function outside this bounded region (the energy of the field outside region D). Since the field is finite-energy, the second term can be made arbitrarily close to zero by taking a large enough region D and $\varepsilon(C_{\mathrm{B}})$ becomes a good measure of representation performance over the entire space.

2.2 Random Field Model

In our experiments we use a parametric non-stationary signal model known as the Gaussian-Schell model (GSM). This is a random field model widely used in the study of random optical fields with various generalizations and applications. GSM beams have been investigated with emphasis on different aspects such as their coherent mode decomposition [134, 135], or their imaging and propagation properties [136–144].

GSM fields are a special case of Schell model sources. A Schell model source is characterized by the covariance function

$$K(x_1, x_2) = I(x_1)^{0.5} I(x_2)^{0.5} \nu(x_1 - x_2), \qquad (2.13)$$

where $I(x)$ is called the intensity function and $\nu(x_1 - x_2)$ is called the complex degree of spatial coherence in the optics literature. For a Gaussian-Schell model, both of these functions are Gaussian shaped

$$I(x) = A_f \exp(-\frac{x^2}{2\sigma_I^2}) \tag{2.14}$$

$$\nu(x_1 - x_2) = \exp(-\frac{(x_1 - x_2)^2}{2\sigma_\nu^2}) \tag{2.15}$$

where $A_f > 0$ is an amplitude coefficient and $\sigma_I > 0$ and $\sigma_\nu > 0$ determine the width of the intensity profile and the width of the complex degree of spatial coherence, respectively. We note that as a result of the Gaussian shaped intensity profile; as we move away from the $x = 0$, the variances of the random variables decay according to a Gaussian function. We also note that $\nu(x_1 - x_2)$ is simply the correlation coefficient function which may be defined as $\nu(x_1 - x_2) = \rho_f(x_1 - x_2) = \frac{K_f(x_1,x_2)}{K_f(x_1,x_1)^{0.5} K_f(x_2,x_2)^{0.5}}$. Hence, as a result of the Gaussian shaped complex degree of spatial coherence function, the correlation coefficient between two points decays according to a Gaussian function as the distance between these two points increases.

In a more general form, one also includes a phase term in the covariance function. As our signal model, we consider this more general form where GSM source is characterized by the covariance function

$$K_f(x_1, x_2) = A_f \exp\left(-\frac{x_1^2 + x_2^2}{4\sigma_I^2}\right) \exp\left(-\frac{(x_1 - x_2)^2}{2\sigma_\nu^2}\right) \exp\left(-\frac{jk}{2R}(x_1^2 - x_2^2)\right) \tag{2.16}$$

Here $A_f > 0$, $j = \sqrt{-1}$. The parameters $\sigma_I > 0$ and $\sigma_f > 0$ determine the width of the intensity profile and the width of the complex degree of spatial coherence, respectively. R represents the wave-front curvature.

This covariance function may be represented in the form

$$K_f(x_1, x_2) = \sum_{k=0}^{\infty} \lambda_k \phi_k(x_1) \phi_k^*(x_2) \tag{2.17}$$

where λ_k are the eigenvalues and $\phi_k(x)$ are the orthonormal eigenfunctions of the integral equation $\int K_f(x_1, x_2)\phi_k(x_1)dx_1 = \lambda_k \phi_k(x_2)$ [134, 135]. Here we assume that the eigenvalues are indexed in decreasing order as $\lambda_0 \geq \lambda_1 \ldots \lambda_k \leq \lambda_{k+1}, \ldots,$ $k \in Z_+$. In signal processing, this representation is known as the Karhunen-Loève expansion [145]. In statistical optics it is referred to as the coherent mode decomposition, where every eigenfunction is considered to correspond to one fully coherent (fully correlated) mode.

The eigenfunctions $\phi_k(x)$ for GSM sources are Hermite polynomials, whose exact form may be found in [135]. Since the eigenvalue distribution will play a crucial role in our investigations we will discuss them in detail. The ratio of the largest eigenvalue λ_n to the lowest eigenvalue λ_0 is given by $\frac{\lambda_n}{\lambda_0} = \left(\frac{1}{\beta^2+1+\beta[(\beta/2)^2+1]^{0.5}}\right)^n$ where β is defined as [135]

$$\beta = \frac{\sigma_\nu}{\sigma_I}. \tag{2.18}$$

β may be considered as a measure of the degree of (global) coherence of the field [15, 135]. Here β may be considered as a measure of the number of significant eigenvalues, hence the effective number of degrees of freedom (DOF) of the source. The effective DOF of a family of signals may be defined as the effective number of uncorrelated random variables needed to characterize a random signal from that family. The concept of the number of degrees of freedom is central to several works, such as [24, 25, 69, 74, 75]. It is known that the random variables that provide the best characterization of the source under the mean-square error criterion are the random variables with variances given by the eigenvalues associated with the Karhunen-Loève expansion. Hence the spread of eigenvalues can be used to define the DOF of the signals. One can say that the DOF is lower when the eigenvalue distribution is more concentrated, and that the DOF is higher when the eigenvalue distribution is more uniformly spread. This definition may be made more precise, for instance by defining the effective DOF $D(\delta)$ as the smallest number satisfying $\sum_{i=1}^{D} \lambda_i \geq \delta\, \varepsilon_0$, where $\delta \in (0, 1]$ and $\varepsilon_0 = \int_{-\infty}^{\infty} K_f(x, x)dx = \sum_{k\geq 0} \lambda_k < \infty$. Returning to the Gaussian-Schell model and β's relationship to the degree of coherence of the field we recall the following [15, 135]: As β increases, the eigenvalues decay faster, so that the effective number of modes required to represent the field decreases and the field is said to be more coherent. In contrast, as β decreases, the eigenvalues decay slower, so that the effective number of modes required to represent the field increases and the field is said to be more incoherent.

Various aspects of the propagation of the Gaussian-Schell model beams through optical systems have been well studied; see, for instance [15, 136–138, 140, 143]. A fundamental result in this area that we will make use of is the following: Say we have an optical system that may be represented by an $ABCD$ matrix (ray-transfer matrix). When a Gaussian-Schell model beam passes through such an optical system, the output is again a Gaussian-Schell model beam with new parameters σ_I', σ_ν', and R_{out}' [136, 137]. It is known that the ratio $\beta = \sigma_\nu'/\sigma_I'$ doesn't change as the field passes through such systems [136, 137, 146]. Hence σ_ν' is given simply by $\sigma_\nu' = \beta\,\sigma_I'$.

To make it easier for the reader to visualize the propagation of the Gaussian-Schell beams, we now review how the beam parameters change in the case of free-space propagation. Let the field parameters associated with a GSM field that has propagated a distance of z be denoted by $\sigma_I(z)$ and $R(z)$. A convenient parameter in expressing the new beam parameters is the Rayleigh distance z_R. As with deterministic Gaussian beams, z_R can be interpreted as the distance the field can propagate before it begins to diverge significantly. For GSM beams, z_R dependens on σ_I and β, and is given by the following expression:

$$z_R(\beta, \sigma_I) = k\sigma_I^2(\frac{1}{\beta^2} + \frac{1}{4})^{-0.5} \tag{2.19}$$

where k is the wave-number [138, 140]. The new beam parameters for a field after propagation over a distance z can be expressed as follows [136, 137]:

$$\sigma_I(z) = \sigma_I(1 + \frac{z^2}{z_R^2})^{0.5}, \tag{2.20}$$

and

$$R(z) = z(1 + \frac{z_R^2}{z^2}). \tag{2.21}$$

Comparing these with the corresponding formulas for deterministic Gaussian beams, (for instance [147, Chap. 3]), we observe that the expressions relating $\sigma_I(z)$ and $R(z)$ to z_R have the same form. These expressions depend on the degree of coherence of the field through z_R, which depends on β.

Before leaving this section, we would like to make a few remarks about the existence of the expansion in (2.17) for the GSM source. We note that, in general, sources defined on the infinite line do not have expansions with discrete eigenvalue spectrum. To obtain such an expansion, one usually considers the source on a compact region (which in our case corresponds to a bounded region). Then the existence of such a representation is guaranteed by Mercer's Theorem, see for example [148, Ch.7]. In [135], an expansion with discrete eigenvalue spectrum is investigated for the GSM source on the infinite line without discussing the existence of such a decomposition in detail. Nevertheless, we here note that such an expansion is possible for the GSM source due to [149, Thm. 1]. This result states that along with continuity, having $\int_{-\infty}^{\infty} K_f(x,x)dx < \infty$ and $K_f(x,x) \to 0$ as $|x| \to \infty$ is sufficient to ensure such a representation. We note that both of these conditions are plausible in a physical context: the first one is equivalent to the finite energy assumption and the second one requires the intensity of the field to vanish as $|x|$ increases, properties one commonly expects from physically realizable fields. As can be seen from (2.16), the covariance function of a GSM source satisfies these properties. Hence an expansion with a discrete eigenvalue spectrum as in (2.17) is possible for GSM sources.

2.3 Trade-off curves for GSM fields are invariant under propagation through first-order optical systems

We now consider the problem of sampling the output of a first-order optical system in order to represent the input optical field. Such systems are also referred to as $ABCD$ systems or quadratic-phase systems [150]. Mathematically represented by linear canonical transforms [18], these systems encompass arbitrary concatenations of lenses, mirrors and sections of free space, as well as quadratic graded-index media. Here we assume that the parameters A, B, C, D of the $ABCD$ matrix are real with $AD - BC = 1$.

In the next section, we will consider a given number of samples and find the minimum possible representation error for that budget. Varying the bit budget, we will obtain trade-off curves between the error and the number of samples (for instance, look forward to Fig. 4.1 for an example). Here we are concerned with how first-order optical systems change these trade-off curves; in other words, does it make any difference if we represent the signal with samples of the output of such a system, rather than with samples of the input itself? A more general version of this problem, where the samples are of limited accuracy are treated in Section 4.2. To avoid unnecessary repetitions, here we will only review the main results, and postpone the detailed discussions and the proof until Section 4.2.

We first observe that there is no system noise $n(x)$, for GSM fields, the trade-off curves are invariant for different σ_I values. Our second and main observation is the following: the trade-off curves are invariant under passage through arbitrary $ABCD$ systems; that is, the error versus cost trade-offs for the estimation of the input of an optical system based on the samples of the input field are the same as those based on the samples of the output field. In other words, the samples of the output field are as good as the samples of the input field. Moreover, the optimum sampling strategy at the output can be easily found by scaling the optimum sampling strategy at the input. When there is system noise $n(x)$, we observe that the trade-off curves are invariant for different σ_I values and the optimum sampling points can be found by scaling.

2.4 Trade-offs between Error and Number of Samples

We now investigate the trade-off between the error and the number of samples, and the optimum sampling intervals associated with different sampling scenarios.

In our experiments, we choose to work with the equivalent parameters σ_I and β, instead of σ_I and σ_ν. Under fixed β, this choice has the advantage of allowing the results for a given σ_I value to be found by using the results for another σ_I value, by appropriately scaling the coordinate space. Hence in our experiments

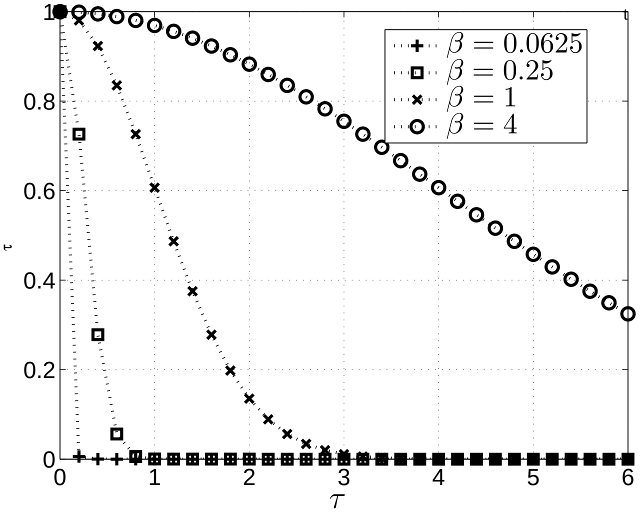


Figure 2.1: Correlation coefficient as a function of distance, β variable.

we fix $\sigma_I = 1$ without loss of generality.

To obtain covariance functions corresponding to random fields with varying DOF, we use different β values: $\beta = 1/16,\ 1/4,\ 1,\ 4$. As stated in Section 2.2, $\sigma_\nu = \beta\sigma_I$ determines the width of the correlation function, which is a Gaussian function. We present the correlation function $\rho(\tau)$ for these values of β in Fig. 2.1.

We choose the noise model similar to the signal model, but with a flat intensity distribution: $I_n(x) = A_n$, $\nu_n(x_1 - x_2) = \exp(-\frac{(x_1-x_2)^2}{2\sigma_{\nu,n}^2})$, where $\sigma_{\nu,n} = \beta_n\sigma_I$, $\beta_n = 1/32$. We consider different noise levels parameterized according to the signal-to-noise ratio, defined as the ratio of the peak signal and noise levels: $\mathrm{SNR} = \frac{A_f}{A_n}$. We consider the values SNR $= 0.1,\ 1,\ 10,\ \infty$ to cover a wide range of situations.

For simplicity in presentation, in our simulations we focus on Δ_x and set the less interesting $x_0 = 0$. We choose the interval D equal to $[x_L, x_H] = [-5\sigma_I, +5\sigma_I]$. With this choice of D, most of the energy of the signal falls inside the interval and the error arising from the fact that only signal values in the region D are estimated is very small ($\leq 10^{-10}$), so that the second term in (2.12) can be ignored.

To compute the error expressions and optimize over the parameters of the representation strategy, we discretize the x space with the spacing Δ_c. For instance, we approximate the integral in (2.2) as $\sum_{k\in D_N} \|f(k\Delta_c) - \hat{f}(k\Delta_c \mid \mathbf{g})\|^2 \Delta_c$ where $D_N = \{k : k\,\Delta_c \in D\}$. The estimates are only calculated at these discrete points: $\hat{f}(k\Delta_c \mid \mathbf{g}) = \mathbf{h}(k\Delta_c)\mathbf{g}$. To determine the estimate functions $\mathbf{h}(k\Delta_c)$, we solve the equation $\mathbf{K}_{\mathbf{f}\,\mathbf{g}}(k\Delta_c) = \mathbf{h}(k\Delta_c)\mathbf{K}_{\mathbf{g}}$ for each $k \in D_N$. We would like to note that the above simple procedure for solving (2.5) for $\mathbf{h}(x)$, corresponds to the following method: we discretize (2.5) and approximate the solutions $h_i(x)$ as $\bar{h}_i(x) = \sum_{j=1}^{N} h_{ji}\,\mathrm{sinc}(\mathrm{x} - \mu_\mathrm{j})$ where $h_{ji} = h_i(x = \mu_j)$. Substitution of the approximate solution $\bar{\mathbf{h}}(x) = [\bar{h}_1(x), \ldots, \bar{h}_M(x)]$ into the right hand side of (2.5) gives an expression that, in general, is not exactly equal to the left hand side. We determine the parameters h_{ji} by requiring (2.5) to hold exactly at N selected points ν_i. Hence (2.5) becomes a system of equations with $N \times M$ unknowns.

To find the optimum sampling intervals, we use a brute force method, where for a given M_b we calculate the error for varying Δ_x, and choose the one providing the best error value. This simple approach has the advantage of enabling us to investigate the effect of Δ_x on the error, and hence the sensitivity of the performance to choosing Δ_x different from the optimal values. (We note that the optimization variable Δ_x and the discretization variable Δ_c are not the same. Δ_x is the sampling interval whose optimal value we seek, whereas Δ_c is the discrete grid spacing we employ in the numerical experiments.)

We report the error as a percentage defined as $100\,\varepsilon(M_b)/\varepsilon_0$ where $\varepsilon_0 = \int_{-\infty}^{\infty} K_f(x,x)dx = A_f\sqrt{2\pi}$.

In the following experiments we will investigate the trade-off between the MSE error $\varepsilon(M_b)$ and M_b, the number of measurements we are allowed to make.

Trade-offs -Variable Noise Level: We first investigate the effect of noise level on the trade-off between $\varepsilon(M_b)$ and M_b. Here SNR takes the values SNR $=$ [0.1, 1, 10, ∞] and two different values of $\beta = [1/16, 1]$ are considered. Fig. 2.2 and Fig. 2.3 correspond to $\beta = 1/16$ (high effective DOF) and $\beta = 1$ (low effective DOF), respectively. As expected, the error decreases with M_b for both cases. We note that for both of cases, $\varepsilon(M_b)$ is very sensitive to increases in M_b

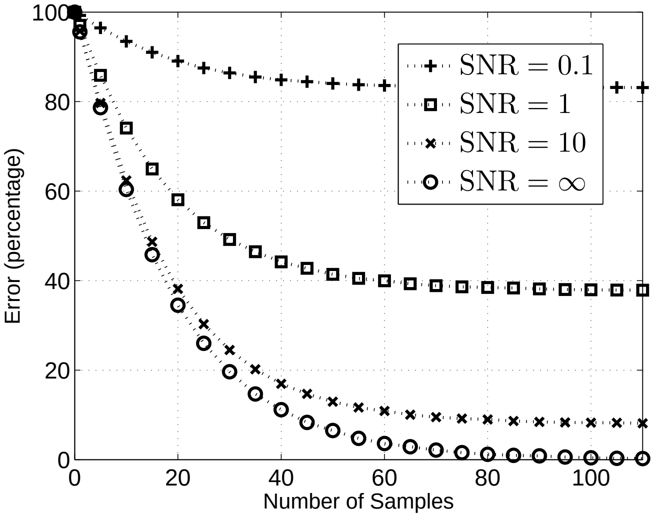


Figure 2.2: Error vs. number of samples, $\beta = 0.0625$, SNR variable.

for smaller M_b. Then it becomes less responsive and eventually saturates. For each value of M_b, the error decreases as SNR increases, and for higher M_b values approaches zero as SNR $\to \infty$. We see that when the field has low effective DOF (Fig. 2.3), we obtain much better trade-off curves for all values of SNR than Fig. 2.2, which represents the relatively high effective DOF case. For instance for SNR $= \infty$, for the high DOF case an error of 20% is obtained when the number of samples is around 30, whereas for the field with low DOF a smaller error value is achieved with only 5 samples. This point is further investigated in the upcoming experiments.

Trade-offs -Variable Effective DOF: We now investigate the effect of the DOF of the unknown field on the trade-off between M_b and $\varepsilon(M_b)$. Here β is varied over $\beta = [1/16,\ 1/4,\ 1,\ 4]$ and two different values of SNR $= [0.1, \infty]$ are considered. Fig. 2.4 and Fig. 2.5 show the results for SNR $= \infty$ and SNR $= 0.1$, respectively. Both of the plots show that for lower values of β (corresponding to higher DOF), it is more difficult to achieve low values of error within a given number of samples. But as β increases, the total uncertainty in the field decreases, and it becomes a lot easier to achieve lower values of error.

In Fig. 2.4, we observe that for all values of β, effectively zero error is obtained

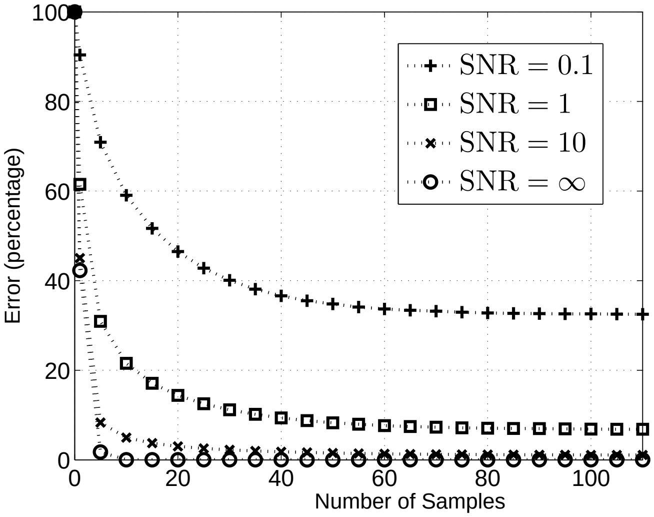


Figure 2.3: Error vs. number of samples, $\beta = 1$, SNR variable.

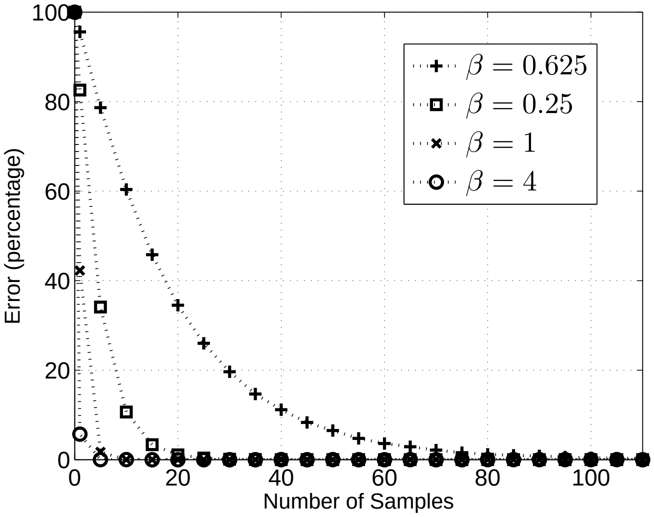


Figure 2.4: Error vs. number of samples, SNR $= \infty$, β variable.

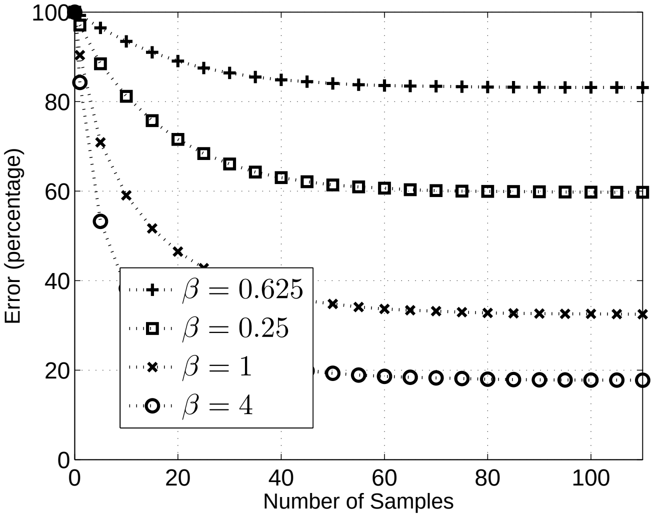


Figure 2.5: Error vs. number of samples, SNR = 0.1, β variable.

as M_b is increased; the field can be represented with effectively 0 error with a finite number of samples. This is not surprising, since the effective DOFs of the signal sources under consideration are finite.

Comparing the performances in Fig. 2.4 and Fig. 2.5 for low and high values of the cost budget, we see that the effect of DOF is more pronounced for different SNR values for different regions of M_b: for low M_b values, the effect of DOF is more strong in the high SNR case; for high M_b values, the effect of DOF is more strong in the low SNR case. For low M_b values, for the high SNR case there is a drastic performance difference between different values of DOF; for the lower DOF values it is possible to obtain very low values of error (≈ 0), a far better performance compared to the higher DOF case. As M_b increases, the difference in performance for different values of DOF decreases, and effectively zero error is obtained for all values of DOF. For high M_b values, the effect of DOF is more pronounced in the low SNR case: the error curves for fields with different DOFs saturate at different values. When the noise level is high, it is not possible to wash out the effect of system noise by taking more samples if the fields have high DOF, hence the curves saturate at relatively high error values. On the other hand, the effect of noise can be cancelled out if the field has relatively low DOF, hence these curves saturate at relatively low values.

Optimum Sampling Intervals: We will now investigate the relationship between the optimum sampling interval Δ_x and the problem parameters M_b, β, SNR.

In general, the optimum policy under a given number of samples can be informally interpreted in the light of two driving forces. The first one is to collect as many effectively uncorrelated samples as possible, so that every sample we have will provide as much new information as possible about the field. The other one is to avoid samples with low variances, since a sample with a low variance is worse than a sample that has higher variance and has the same correlation coefficient with the field values at other points (so that the amount of uncertainty reduction for the other field values due to observation of this sample will be the same). We note that for a GSM source the function that determines the correlation of a field value at a particular point with the field values at other points is the same for a field value at any given location (given by $\nu(x_1, x_2)$), and it is a decreasing function of the distance between the points. Hence when we take a sample at a particular point, we also obtain some information about the field values around that point, but not so much about the field values that are far away. Due to the GSM source structure, low variance samples have relatively low variance neighbours, and hence the decrease in the uncertainty due to observation of field values at these points will be relatively low. This further encourages us to avoid samples with low variances.

Here we investigate the dependence of the optimum sampling interval on β, SNR and M_b. Fig. 2.6 and Fig. 2.7 give the optimum sampling intervals versus number of samples for $\beta = 1/16$ and $\beta = 1$, respectively. We observe that in general the optimum sampling interval decreases with increasing number of samples. When the number of samples one is allowed is low, one tries to obtain as much independent information as possible by choosing the samples apart. As M_b increases and we are allowed to use more samples, one can afford to choose the samples closer so that field values that were considered to give enough information about each other in the former case can be also observed and lower values of error

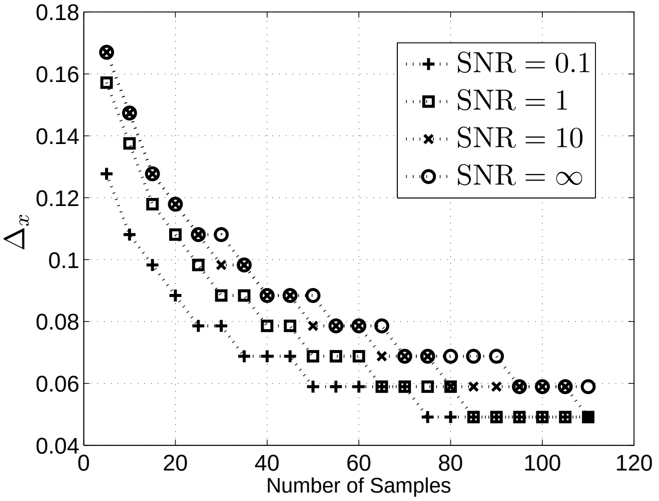


Figure 2.6: Optimum sampling interval vs number of samples, $\beta = 1/16$, SNR variable.

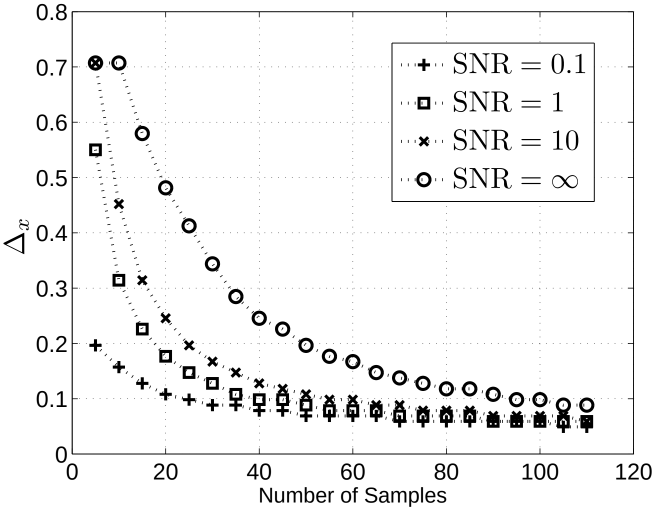


Figure 2.7: Optimum sampling interval vs number of samples, $\beta = 1$, SNR variable.

can be obtained.

For a given β and M_b, the sampling interval increases with increasing SNR. As SNR increases, observing the field at a particular point allows one to estimate the value of the field at this point and its neighbours better. Therefore, to ensure that each sample provides new information, one should increase the sampling interval.

Comparing Fig. 2.6 and Fig. 2.7, we observe that the optimum sampling intervals are smaller for the high DOF case (Fig. 2.6). As DOF increases, that is, the number of uncorrelated random variables required to effectively represent the field increases, and also given the GSM correlation structure, the field value at each point becomes less correlated with its neighbouring points. Hence the reduction in the uncertainty of the field values at the neighbours of a given point due to the observation of the field at a this point is smaller. This, together with the fact that the variances of field values decrease as the samples are placed further away from $x = 0$ point, encourages us to take samples more closely, so that all the effectively uncorrelated samples with high variances can be collected.

Sensitivity of Performance to the Sampling Intervals: We will now discuss the sensitivity of the performance to the sampling interval. For this purpose we look at the error versus sampling interval curves and observe how much the error deviates from its optimum value as the sampling interval deviates from the optimum sampling interval.

Fig. 2.8, Fig. 2.9, Fig. 2.10 and Fig. 2.11 present the error versus sampling interval curves for $\beta = 1$, SNR $= 0.1$, and $\beta = 1$, SNR $= 10$, and $\beta = 1/16$, SNR $= 10$, and $\beta = 1/16$, SNR $= 0.1$, respectively. We note that in all figures, as M increases, data for fewer numbers of sampling points are plotted. This is due to the fact that we only allow the samples to be taken in the bounded domain D, and as M increases, larger sampling intervals become impermissible.

We observe that in all of these figures, for a given M the error first decreases as we increase the sampling interval, and after reaching the optimum sampling interval it starts to increase again. This behaviour may be interpreted in view

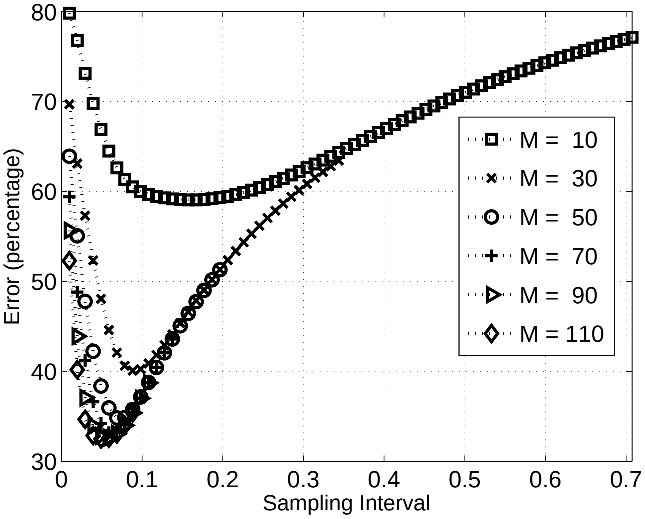


Figure 2.8: Error vs. sampling interval, $\beta = 1$, SNR $= 0.1$, number of samples variable.

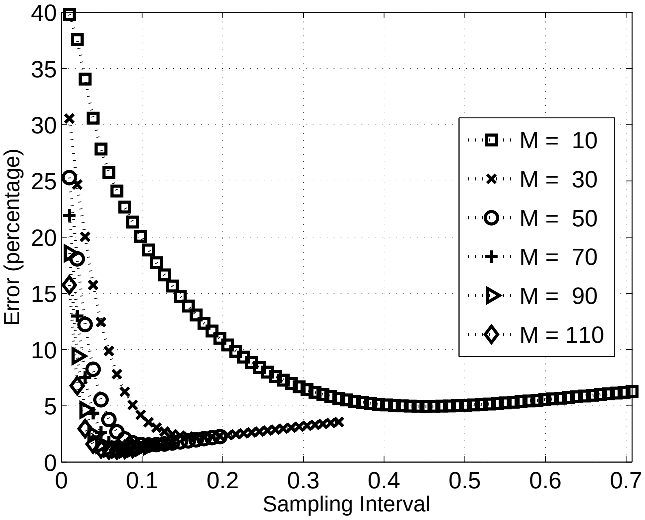


Figure 2.9: Error vs. sampling interval, $\beta = 1$, SNR $= 10$, number of samples variable.

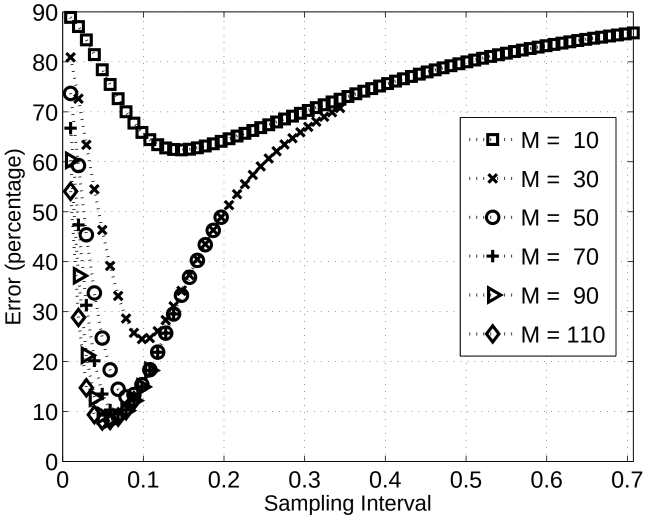


Figure 2.10: Error vs. sampling interval, $\beta = 1/16$, SNR = 10, number of samples variable.

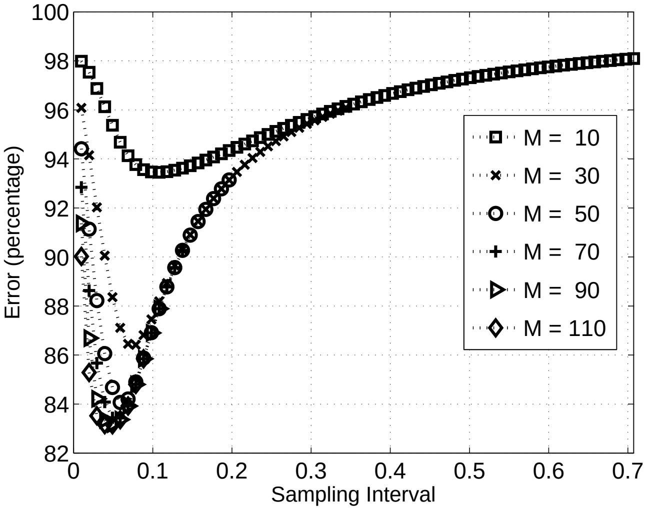


Figure 2.11: Error vs. sampling interval, $\beta = 1/16$, SNR = 0.1, number of samples variable.

of the following observation: We expect that the optimum policy will be the one that takes as many uncorrelated samples with high variances as possible. If we take the samples too close, we acquire random variables close to each other whose correlation will be relatively strong due to the nature of the GSM model. Hence the error will be relatively high, since the samples are spent on obtaining redundant information. On the other hand if we take the samples far apart from each other, we may be missing some of the random variables that contain effectively uncorrelated information with the samples we take. Moreover, we may waste our sample budget on random variables that have relatively low variance (the ones that are outside the main lobe of the Gaussian intensity function). Hence the error may again be relatively high.

While commenting on the sensitivity, we focus on the differences in absolute error in different scenarios. We observe that, for a given β and SNR, as M increases, the achievable error values become more sensitive to the sampling interval. For instance, in Fig. 2.8 for $M = 10$, any sampling interval in the range [0.1 0.25] provides approximately the same error ($\approx 60\%$); whereas for $M = 70$, a similar range of sampling intervals around the optimum sampling interval (such as [0.02 0.15]) produce error values in the range of $\approx 35 - 50\%$. When we are allowed a small number of samples, taking samples with a high enough sampling interval can easily provide effectively uncorrelated samples; avoiding samples with low variances is not a serious issue that requires sensitive design, choosing the sampling interval smaller than a given value is enough. Hence any sampling interval between these lower and higher bounds produces effectively the same error level with the optimum interval. On the other hand, when a larger number of samples are allowed, one has to design the locations of the samples more carefully to find the best trade-off between collecting relatively uncorrelated samples and avoiding samples with low variances.

We observe that when DOF is low, the error may be considered to be more sensitive to the sampling interval for low SNR values. For instance, for $\beta = 1$, SNR $= 10$, and $M = 10$, any sampling interval in the range [0.3 0.6] provide approximately the same error with the optimal sampling strategy ($\approx 5\%$). On the other hand, for $\beta = 1$, SNR $= 0.1$, in order to have approximately the same error with the optimal strategy ($\approx 60\%$), only sampling intervals in the range [0.1 0.25] are allowed. We note that the length of [0.1 0.25] is half of the length of [0.3 0.6]. On the other hand, when DOF is high, the error is more sensitive to the sampling interval for high SNR values. We remind that in these comparisons we consider the variation in absolute error for different scenarios. For instance, for $\beta = 1/16$, SNR = 0.1, and $M_b = 10$, in order to obtain an error that is not worser than the error obtained with the optimal strategy by more than 5% percent ($\approx 93 - 98\%$), it is sufficient to use any sampling interval in the range of [0.01 0.7]. On the other hand for $\beta = 1/16$, SNR = 10, in order to obtain an error that is not worser than the error obtained with the optimal strategy by more than 5% percent, ($\approx 60 - 65\%$), it is necessary to use a sampling interval in the range of [0.1 0.2], a significantly smaller range.

Similar comparisons can be made for the other cases as well: When SNR is high/low, the sensitivity of the error to the sampling interval increases with increasing/decreasing DOF. All of these results concerning the sensitivity can be interpreted in the light of the following observation: In general, we observe that the error becomes more sensitive to our choice of sampling interval when the effect of different problem parameters on the optimum sampling interval conflict: One of the problem parameters requires us to take the samples closer to each other, while the other requires us to take them farther apart. For instance, low DOF requires us to take the samples apart whereas low SNR requires us to take the samples closer. Hence for low DOF, as SNR decreases, the error becomes more sensitive to the sampling interval. Taking a closer look, we observe that when DOF is low, the field values are highly correlated with each other, and for high values of SNR the field values to be observed contain low levels of noise. Hence the samples carry essentially the same information, making the choice of the sampling interval relatively unimportant. As SNR decreases, a compromise between the two conflicting forces is required, making this choice more important: taking samples close enough so that the noise is effectively washed out, and taking samples sufficiently apart from each other so that each sample brings new information.

2.5 Conclusions

We have considered the representation of a finite-energy non-stationary random field with a finite number of samples. By considering a parametric non-stationary field model, namely the Gaussian-Schell model, we obtained the trade-offs between the number of samples and the representation error, for varying noise levels and for sources with varying degrees of freedom (DOF). We have discussed the optimum sampling intervals, and their dependence on the problem parameters. We have observed that increases in either of (i) the number of allowed samples, (ii) DOF, or (iii) the noise level, results in a decrease in the optimum sampling interval. We have also investigated the sensitivity of the error to the chosen sampling interval. We have observed that the error is more sensitive to sampling interval when (i) the number of allowed samples is high, (ii) DOF is high and the noise level is low, or (iii) DOF is low and the noise level is high.

Chapter 3

Representation and Recovery using Limited Accuracy Measurements

In Chapter 2, we have investigated the effect of restriction of the total number of samples to be finite while reconstructing a random field using its samples. We have assumed that the amplitude accuracies are so high that the sample values can be assumed to be exact. For a given number of samples, we have sought the optimal sampling interval in order to represent the field with as low error as possible. In this chapter, we focus on the effect of limited amplitude accuracy of the measurements. Our framework is as follows: We aim to optimally measure an accessible signal, in order to estimate a signal which is not directly accessible. We consider a measurement device model where each device has a cost depending on the number of amplitude levels that the device can reliably distinguish. We also assume that there is a cost budget so that it is not possible to make a high amplitude resolution measurement at every point. We investigate the optimal allocation of cost budget to the measurement devices so as to minimize estimation error. This problem differs from standard estimation problems in that we are allowed to "design" the number and noise levels of the measurement devices subject to the cost constraint. We present the trade-off curves between the best achievable estimation error and the cost budget. In this chapter, we will consider this problem in a discrete framework. In Chapter 4, we will formulate this problem within a continuous framework.

The problem addressed in this chapter was motivated mostly by problems related to measurement of propagating wave-fields. We are concerned with the problem of estimating the values of a wave-field in a certain region from measurements of its values at another region. We consider a very general measurement scenario: Let us consider a wave-field propagating through a system characterized by a linear input-output relationship. We wish to recover the input wave field as economically as possible from noisy measurements of the output field. We are concerned with accuracy both in the sense of spatial resolution and in the sense of the amplitude resolution. We are also concerned with the cost of performing the measurements and the trade-offs between the total cost and estimation accuracy. The cost of a measurement device is primarily determined by the number of amplitude levels that the device can reliably distinguish; devices with higher numbers of distinguishable levels have higher costs. We also assume that there is a limited cost budget so that it is not possible to make a high amplitude resolution measurement at every point. For a given cost budget, we would like to know how to best make the measurements so as to minimize the estimation error, or vice versa, leading to a trade-off. In particular, we are interested in questions such as how many measurements we should make, how the sensitivity of each detector should be chosen, and so forth, in order to obtain the best trade-off. These questions are not merely of interest for practical purposes but can also lead to a better understanding of the information relationships inherent in propagating wave-fields.

While our primary motivation and numerical examples come from wave propagation problems, we emphasize that our formulation is also valid for other measurement problems where similar cost-budget models are applicable, and the observed variables are related to the unknown variables through a linear relation. One such example is the Wiener filtering problem which is a basic problem in signal processing with many practical applications. Another example arises in data communications, where a transmitted signal may suffer intersymbol interference as it passes through a medium, and the equalization problem is to estimate the transmitted signal from the received signal. These problems are of the same general structure as the one we are considering. In digital implemention of such estimators, the usual approach is to work with constant accuracy over all samples of the observation. Our framework introduces great flexibility in terms of the number, positions, and accuracies of these samples. This not only allows better optimization, but also allows us to observe a number of interesting trade-offs and relationships.

The measurement strategy problem we formulate and solve in this chapter arises in a diversity of physical contexts. We are concerned with measurement and estimation of spatially (or temporally) distributed fields modeled as random vectors. An important aspect of our formulation is that it allows sensors with different performances and costs in the model. While the kind of cost function we use may come as natural in the context of communication costs, we believe it has never been used to model the cost of measurement devices. The optimal measurement problem we define differs from standard estimation problems in that we are allowed to "design" the number and noise levels of the measurement devices subject to a total cost constraint. Our main results are presented in the form of trade-off curves between the estimation error and the total cost. We discuss the effects of signal-to-noise ratio (SNR), and the degree of coherence on these trade-offs in wave-propagation problems. The *degree of coherence* is a measure of the amount of correlation among different points of a random wave-field. We are not aware of previous discussion of the effect of degree of coherence in these types of problems. Our conclusions not only yield practical strategies for designing optimal measurement systems under cost constraints, but also provide interesting insights into the nature of information theoretic relationships in wave-fields.

In Section 3.1 of the chapter, we present the mathematical model of the measurement problem discussed above. A fundamental concept in our formulation, the cost of a measurement, is discussed in Section 3.2. Some special cases of our formulation are presented in Section 3.3. In Section 3.4 we propose an iterative algorithm and provide numerical results. We conclude in Section 3.5.

3.1 Problem Formulation

In the specific measurement scenario under consideration in this chapter, noisy measurements are made at the output of a linear system, in order to estimate the input of the system. We study a discrete version of this problem by assuming that the space variable is discretized to a fixed finite set of points. The following development was first proposed in [3–5]. The system we consider may be represented by a matrix equation

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{f} \in \mathbb{R}^N$ is the unknown input random vector, $\mathbf{n} \in \mathbb{R}^M$ is the random vector denoting the inherent system noise, and $\mathbf{g} \in \mathbb{R}^M$ is the output of the linear system. We assume that $\mathbf{f}$ and $\mathbf{n}$ are statistically independent zero-mean random vectors. Here $\mathbf{H}$ is a $M \times N$ matrix denoting the linear system. We put no restrictions on the system matrix $\mathbf{H}$. For instance, in wave propagation applications, the locations of the measurements and the locations of the unknown field values may be quite distant from each other, e.g., we may wish to estimate the field at the outer edges of a region with measurements made in the center.

Measurements are made at the output of the linear system to obtain the measurement vector $\mathbf{s} \in \mathbb{R}^{\mathbb{M}}$ according to the measurement model

$$\mathbf{s} = \mathbf{g} + \mathbf{m}, \tag{3.2}$$

where $\mathbf{m}$ denotes the measurement noise. We assume that $\mathbf{m}$ is independent of $\mathbf{f}$ and $\mathbf{n}$. Further, we assume that the components of $\mathbf{m}$ are indepedent, zero-mean random variables, but not necessarily identically distributed. So, the variance $\sigma^2_{m_i}$ of each component of $\mathbf{m}$, indexed by $i = 1, \ldots, M$, may be different.Here $\mathbf{n}$ is an intrinsic part of the relation between $\mathbf{g}$ and $\mathbf{f}$ which we have no control over, whereas $\mathbf{m}$ is the noise associated with the measurement devices we use and thus depends on the choices we make.

In the following formulation, we assume the knowledge of only second-order statistics of the underlying random variables. We let $\mathbf{K_f}$, $\mathbf{K_n}$, $\mathbf{K_m}$, and $\mathbf{K_s}$ denote the covariance matrices of $\mathbf{f}$, $\mathbf{n}$, $\mathbf{m}$, and $\mathbf{s}$, respectively. Note that $\mathbf{K_s} =$

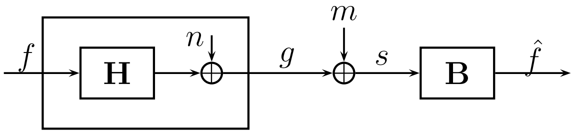


Figure 3.1: Measurement and estimation systems model block diagram.

$\mathbf{H}\mathbf{K_f}\mathbf{H}^{\mathrm{T}} + \mathbf{K_n} + \mathbf{K_m}$. Note also that since we assume that $\mathbf{m}$ has independent components, $\mathbf{K_m} = \mathrm{diag}(\sigma^2_{m_1}, \ldots, \sigma^2_{m_M})$.

We assume that the cost associated with measuring the ith component of $\mathbf{g}$ is $C_i = (1/2)\log\left(\sigma^2_{s_i}/\sigma^2_{m_i}\right)$, where $\sigma^2_{s_i}$ denotes the variance of s_i. The units of C_i are bits. Smaller measurement noise levels result in higher costs whereas larger measurement noise levels allow lower costs. The plausibility of this form for the cost function is discussed in Section 3.2. The cost of measuring $\mathbf{g}$ is defined as $\sum_{i=1}^{M} C_i$, the sum of the costs of measuring all of its components.

The objective is to minimize the mean-square error (MSE) between $\mathbf{f}$ and $\hat{\mathbf{f}}(\mathbf{s})$, the estimate of $\mathbf{f}$ given $\mathbf{s}$. We consider only linear minimum mean-square error (LMMSE) estimators, and $\hat{\mathbf{f}}(\mathbf{s})$ denotes the LMMSE estimator. Hence the estimate may be written as $\hat{\mathbf{f}}(\mathbf{s}) = \mathbf{B}\,\mathbf{s}$ where $\mathbf{B}$ is an N by M matrix. A block diagram illustrating this problem is given in Fig. 3.1.

The problem can be stated as follows: Given the covariance matrices $\mathbf{K_f} \in \mathbb{R}^{N\times N}$, $\mathbf{K_n} \in \mathbb{R}^{M\times M}$, and a system matrix $\mathbf{H} \in \mathbb{R}^{\mathbf{M}\times\mathbf{N}}$, determine

$$\begin{aligned} \varepsilon(C_{\mathrm{B}}) &\triangleq \min_{\mathbf{K_m}} \; E\{\|\mathbf{f} - \hat{\mathbf{f}}(\mathbf{s})\|^{\mathbf{2}}\} && (3.3) \\ &= \min_{\mathbf{K_m}} \; \min_{\mathbf{B}} \; E\left\{\mathrm{tr}\left[(\mathbf{f} - \mathbf{B}\mathbf{s})(\mathbf{f} - \mathbf{B}\mathbf{s})^{\mathrm{T}}\right]\right\} && (3.4) \end{aligned}$$

subject to

$$\sum_{i=1}^{M} \frac{1}{2}\log\left(\frac{\sigma^2_{s_i}}{\sigma^2_{m_i}}\right) \leq C_{\mathrm{B}}. \tag{3.5}$$

where $\mathbf{K_m} = \mathrm{diag}(\sigma^2_{m_1}, \ldots, \sigma^2_{m_M})$ is the covariance of $\mathbf{m}$, E denotes expectation with respect to $\mathbf{f}$, $\mathbf{n}$, and $\mathbf{m}$, $\|\cdot\|$ denotes Euclidean norm, and tr denotes the trace operator.C_{B} is the total cost budget; the sum of the cost of all detectors is not allowed to exceed C_{B}. We go from (3.3) to (3.4) by writing $E\{\|\mathbf{f} - \hat{\mathbf{f}}(\mathbf{s})\|^{\mathbf{2}}\} =$

$E\{\|\mathbf{f}-\mathbf{Bs})\|^{\mathbf{2}}\} = E\{\sum_{i=1}^{N}(\mathbf{f}-\mathbf{Bs})_i^2\} = E\{\mathrm{tr}[(\mathbf{f}-\mathbf{Bs})(\mathbf{f}-\mathbf{Bs})^{\mathrm{T}}]\}$. We let $\sigma_{m_i}^2 \in \mathbb{R}\cup\{\infty\}$, and use min instead of inf in (3.3).

We note that for a given $\mathbf{K_m}$, the minimization over $\mathbf{B}$ in (3.4) is a standard LMMSE problem with solution $\mathbf{B} = \mathbf{K_f}\mathbf{H}^{\mathrm{T}}\mathbf{K}_s^{-1}$. This standard solution may be arrived at using the orthogonality condition $E\left\{(\mathbf{f}-\mathbf{Bs})\,\mathbf{s}^{\mathrm{T}}\right\} = \mathbf{0}\in\mathbb{R}^{N\times M}$, where $E\left\{\mathbf{f}\,\mathbf{s}^{\mathrm{T}}\right\} = \mathbf{K_f}\mathbf{H}^{\mathrm{T}}$. Hence we obtain:

$$\varepsilon(C_{\mathrm{B}}) = \min_{\mathbf{K_m}} \mathrm{tr}\left(\mathbf{K_f}-\mathbf{K_f}\mathbf{H}^{\mathrm{T}}\mathbf{K_s}^{-1}\mathbf{H}\mathbf{K_f}\right), \tag{3.6}$$

subject to (3.5). In other words, our aim is to minimize the estimation error by allocating a given measurement cost budget optimally over the M components of (4.2). This is equivalent to optimally adjusting the measurement noise level for each component, realizing that with a given budget, we cannot measure all components as highly accurately as we wish. Although not explicitly stated, the number of components we actually measure is also an optimization variable. If as the result of our optimization we find that $C_i \approx 0$ for certain components, this means that measuring those components do not usefully contribute to the estimation and therefore need not be measured in the first place.

As seen above, this problem differs from a standard LMMSE estimation problem in that the covariance $\mathbf{K_m}$ of the measurement noise is subject to optimization. We are allowed to "design" the noise levels of the measurement devices subject to a total cost constraint so as to minimize the overall estimation error. To the best of our knowledge this problem is novel.

We would like to note that this minimization problem defined by the objective in (3.6) and the constraint in (3.5) is not convex. The inequality constraints given by (3.5) (and the hidden constraint that $\mathbf{K_m}$ is positive-semidefinite) define convex constraints with respect to the variable $\mathbf{K_m}$, yet the objective function is not a convex function. For convenience, let us consider the vector of noise level variances as the optimization variable instead of the matrix $\mathbf{K_m}$. Let us denote this vector with $\mathbf{k_m}$, where $\mathbf{k_m} = [\sigma_{m_1}^2,\ldots,\sigma_{m_M}^2] = [k_{m1},\ldots,k_{mM}]$. Hence $\mathbf{K_m}$ can be also written as $\mathbf{K_m} = \mathrm{diag}(\mathbf{k_m})$. We first observe that the fact that $\sigma_{m_i}^2 \geq 0$ defines a convex constraint on the optimization variable $\mathbf{k_m}$. Now consider the constraint given in (3.5): $a(\mathbf{k_m}) = \sum_{i=1}^{M} \frac{1}{2} \log(1 + \frac{\sigma_{g_i}^2}{k_{m\,i}}) - C_{\mathrm{B}} \leq 0$. The convexity of this constraint may be seen, for instance by taking the second derivative of $a(\mathbf{k_m})$ and checking whether the Hessian is positive-semidefinite, that is whether $\nabla^2 a(\mathbf{k_m}) \succeq 0$ [151, Ch. 3]. Here this is indeed the case: $\nabla^2 a(\mathbf{k_m}) = \mathrm{diag}(\frac{\sigma_{g_i}^2(\sigma_{g_i}^2 + 2k_{m\,i})}{{k_{m}}_i^2(\sigma_{g_i}^2 + k_{m\,i})^2}) \succeq 0$. The fact that the objective function is concave over $\mathbf{k_m}$ can be seen as follows: Let $\mathbf{K_e} = (\mathbf{K_f} - \mathbf{K_f H^{\mathrm{T}} K_s^{-1} H K_f})$. Then the objective function in (3.6) is given by $\mathrm{tr}(\mathbf{K_e})$. We note that $\mathbf{K_e}$ is the Schur complement of $\mathbf{K_s}$ in $\mathbf{K} = [\mathbf{K_s}\ \mathbf{K_{sf}}; \mathbf{K_{fs}}\ \mathbf{K_f}]$, where $\mathbf{K_{fs}} = \mathbf{K_f H}^{\dagger}$. Schur complement is matrix concave in $\mathbf{K} \succ \mathbf{0}$, for example see [151, Exercise 3.58]. Since trace is a linear operator, $\mathrm{tr}(\mathbf{K_e})$ is concave in K. Since $\mathbf{K}$ is an affine mapping of $\mathbf{k_m}$, and composition with an affine mapping preserves concavity [151, Sec. 3.2.2], $\mathrm{tr}(\mathbf{K_e})$ is concave in $\mathbf{k_m}$.

Our formulation can be easily generalized to allow repeated measurements (more than one measurement of any g_i is allowed); however repeated measurements always yield higher error for a given cost budget hence including them in the model does not provide a better performance. This point is discussed in Section 3.3.0.2.

3.2 Cost Function

We will now discuss the cost of a measurement device. The cost function discussed here were first proposed in [3–6]. What we refer to as a measurement device is an instrument which can measure the value of a scalar physical quantity over some range with some resolution in amplitude. The cost of a measurement device is primarily determined by the number of amplitude levels that the device can reliably distinguish, a notion which is sometimes referred to as its dynamic range (although the term is sometimes also used to refer to an interval). We will assume that the ranges of measurement devices can be chosen freely to match any interval, and that this has no effect on the cost of the measurements provided the number of resolvable levels remains the same (similar to scaling the range of a multimeter). For a given linear system, the ranges of the measurement devices depend only on the given covariances. Therefore, they need to be changed only if the covariances change. Given the variances of g and m in the measurement process $s = g + m$, the number of distinguishable levels can be quantified as

$$\rho = \varrho \frac{\sigma_s}{\sigma_m} = \varrho \sqrt{\left(1 + \frac{\sigma_g^2}{\sigma_m^2}\right)}, \tag{3.7}$$

where $\varrho > 0$ is a constant that depends on how reliably the levels need to be distinguished. In using this expression we are following the same rationale used to define the number of distinguishable signal levels at the receiver of an additive noise channel, which is due to Hartley [152], and further discussed in [153, 154]. The square-root in the expression keeps the number of levels invariant under scaling of the signals by any constant. Clearly, in the limit of very noisy measurements, ϱ should be 1; therefore we set $\varrho = 1$ henceforth.

We now list some properties that any plausible cost function must possess:

1. $C(\rho)$ must be a non-negative, monotonically increasing function of ρ, with $C(1) = 0$ since a device with one measurement level gives no useful information.

2. For any integer $L \geq 1$, we must have $L\,C(\rho) \geq C(\rho^L)$. This is because a measurement device with ρ levels can be used L times in succession with range adjustments between measurements, to distinguish ρ^L levels. (We also note that this property may be expressed in a more general form as $\sum_{i=1}^{L} C(\rho_i) \geq C(\prod_{i=1}^{L} \rho_i)$.)

A function possessing these properties is the logarithm function; therefore we propose

$$C(\rho) = \log \rho = \frac{1}{2} \log \left(\frac{\sigma_s^2}{\sigma_m^2}\right) = \frac{1}{2} \log \left(1 + \frac{\sigma_g^2}{\sigma_m^2}\right) \tag{3.8}$$

as the cost of carrying out a measurement $s = g + m$.

The proposed cost function has the same form as Shannon's formula for the capacity of a Gaussian noise channel. This does not come as a surprise since a measurement process $s = g + m$ is analogous to sending a message g across a communication channel that adds a noise term m to it. On the other hand, the notion of adjusting the range while keeping the number of resolvable levels constant has no direct counterpart in the communication setting; hence, the measurement and communication problems are not identical problems. We believe such a cost function has never been used to model the cost of measurement devices.

We now discuss the proposed cost of a measurement device from a buyer-seller perspective. In the communication problem, the amount of information delivered to the receiver is measured by the mutual information $I(s;g) = h(s) - h(m)$ between the transmitted and received signals. $I(s;g)$ is also an attractive candidate for the cost function in the measurement scenario since the value of a measurement would be quantified most fairly by how many bits of information it actually conveys on the average about the measured quantity. On the other hand, there is a practical difficulty in charging a fee $I(s;g)$ as it depends on the actual probability distribution $p(g)$ of the measured quantity. It is logical that the measurement device manufacturer will try to sell its device at the price $\max_{p(g)} I(s;g)$ where the maximum is calculated subject to a power constraint $E[g^2] \leq \sigma_g^2$. The would-be equipment purchaser on the other hand will not be willing to pay more than $\min_{p(m)} \max_{p(g)} I(s;g)$ since s/he is only assured that $E[m^2] \leq \sigma_m^2$. Shannon [153] shows that this minmax problem is solved by the Gaussian densities for both $p(g)$ and $p(m)$ and the resulting minmax value is the expression $C(\rho)$ given in (3.8). Thus, the cost function we propose has a satisfying economic interpretation: the seller of the equipment assumes that the purchaser will make the best use of the equipment while the purchaser assumes that the equipment will give the worst type of measurement noise (which is Gaussian) for the given level of resolution.

Since Gaussian error is an acceptable model for many types of measurement devices, the cost function that we use makes sense in a wide range of contexts. For problems where measurement noise is known to follow a different distribution, the cost function can be modified accordingly.

Finally we explore the relationship of the measurement problem to rate-distortion theory. It is clear from Fig. 3.1 that, by the data-processing theorem [155], we have the following relationship regarding the mutual information of the related random vectors: $I(\mathbf{f};\hat{\mathbf{f}}) \leq I(\mathbf{g};\mathbf{s})$; i.e., the estimate $\hat{\mathbf{f}}$ can only provide as much information about $\mathbf{f}$ as the measurement devices extract from the observable $\mathbf{g}$. In turn, by standard arguments, we have $I(\mathbf{g};\mathbf{s}) \leq \sum_{i=1}^{M} I(g_i;s_i)$. The cost function $\frac{1}{2}\log(1+\sigma_{g_i}^2/\sigma_{m_i}^2)$ that we use upper-bounds $I(g_i;s_i)$ whenever the measurement noise is Gaussian with a given variance $\sigma_{m_i}^2$ and the variance of the measured quantity is fixed as $\sigma_{g_i}^2$. Thus for Gaussian measurement noise, we have $I(\mathbf{f};\hat{\mathbf{f}}) \leq C_{\mathrm{B}}$ where C_{B} is the total measurement budget.

The goal of measurements is the minimization of the MSE $\varepsilon(C_{\mathrm{B}}) = E[d(\mathbf{f},\hat{\mathbf{f}})]$ within a budget C_{B} where d denotes $\|\mathbf{f} - \hat{\mathbf{f}}\|^{\mathbf{2}}$. From a rate-distortion theory viewpoint, interpreting d as a *distortion measure*, this is similar to minimizing the average distortion in the reconstruction of $\mathbf{f}$ from a representation $\hat{\mathbf{f}}$ subject to a *rate constraint* $I(\mathbf{f};\hat{\mathbf{f}}) \leq C_{\mathrm{B}}$. This viewpoint immediately gives the bound $\varepsilon(C_{\mathrm{B}}) \geq D(C_{\mathrm{B}})$ where $D(C_{\mathrm{B}})$ is the *distortion-rate function* applicable to this situation.

In the rate-distortion framework one is given complete freedom in forming the reconstruction vectors $\hat{\mathbf{f}}$ subject only to a rate constraint, which in measurement terminology would mean the ability to apply arbitrary transformations on the observable $\mathbf{g}$ before performing a measurement (so as to carry out the measurement in the most favorable coordinate system), and not being constrained to linear measurements or linear estimators. Thus, the measurement problem can be seen as a deviation from the rate-distortion problem in which the formation of the reconstruction vector is restricted by various constraints.

3.3 Special Cases

3.3.0.1 Two-input two-output case

We now consider the special case of the problem where the input and the output signals are 2 by 1 vectors, that is $N = 2$, $M = 2$. The main contribution of studying this case will be to reveal the following interesting properties of the cost-error trade-off curves that hold some values of problem parameters: i) the optimal error-cost trade-off curves may follow different curves for different cost budget values, forming a piece-wise trade-off curve ii) as cost budget increases, it may be best to use a device with a small number of distinguishable levels at a point where a more accurate device were used when the cost budget were smaller. In particular, we will illustrate that when the variables to be measured are highly correlated; it is better to use all the cost budget on measuring only one of the components for low values of budget, and start measuring both them only after the cost budget becomes sufficiently large.

We will start by making some general observations on the optimization problem at hand. Let us first express the problem with an alternative but equivalent formulation. We define the following variables $\alpha_i = \frac{1}{\sigma^2_{m_i}}$, $i = 1, \ldots, M$. Then the cost constraint can be expressed as $a_0(\boldsymbol{\alpha}) = \sum_{i=1}^{M} 0.5 \log(1 + \alpha_i k_{g_{ii}}) - C_{\mathrm{B}} < 0$, where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_M]^{\mathrm{T}}$. The non-negativeness of the variances $\sigma^2_{m_i}$ can be expressed with following conditions: $a_i(\boldsymbol{\alpha}) = -\alpha_i \leq 0$, $i = 1, \ldots, M$. Hence the optimization problem at hand can be expressed as

$$\min_{\boldsymbol{\alpha} \in \mathcal{R}^M} e(\boldsymbol{\alpha}) \tag{3.9}$$

such that

$$\mathbf{a}(\boldsymbol{\alpha}) \leq \mathbf{0} \tag{3.10}$$

where $e(\boldsymbol{\alpha}) = \mathrm{tr}\left(\mathbf{K_f} - \mathbf{K_f}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{K_f}\mathbf{H}^{\dagger} + \mathbf{K_n} + \mathrm{diag}(\mathbf{1}/\alpha_{\mathbf{i}}))^{-1}\mathbf{H}\mathbf{K_f}\right)$, and $\mathbf{a}(\boldsymbol{\alpha}) = [a_0(\boldsymbol{\alpha}), \ldots, a_M(\boldsymbol{\alpha})]^{\mathrm{T}}$. Here the inequalities between vectors denote component-wise comparisons. Let $\mathcal{J}$ denote the set of indices of the active constraints, i.e. the constraints that satisfy the inequality constraints with equality. Let $\{\nabla a_i(\boldsymbol{\alpha}) | i \in \mathcal{J}\}$ denote the set of gradients of active constraints.

We recall the following definition:

Definition 3.3.1. *[156, Defn.12.4] Given the point α and the associated active constraint set $\mathcal{J}$, it is said that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients $\{\nabla a_i(\boldsymbol{\alpha})|i \in \mathcal{J}\}$ is linearly independent.*

We now make the following observation: For the minimization problem at hand, LICQ holds at any feasible point. This may be proved by enumerating the different cases for the active constraint sets, and investigating the linear independence of these: i) When the cost constraint is inactive, linear independence of $\{\nabla a_i(\boldsymbol{\alpha})|i \in \mathcal{J}\}$ is trivial, since each element in this set, $\nabla a_i(\boldsymbol{\alpha})$ is a vector with only one nonzero component at ith location. ii) When the cost constraint is active, and none of the nonnegativess constraints are active, we require $\nabla a_0(\boldsymbol{\alpha})$ be different from zero. This is satisfied since $\sigma^2_{s_i} > 0$ iii) When cost constraint is active, and some of the nonnegativess constraints are active, (but not all of them), linear independence of $\{\nabla a_i(\boldsymbol{\alpha})|i \in \mathcal{J}\}$ follows from $\alpha_i < \infty$ and $\sigma^2_{s_i} > 0$. iv) When cost constraint is active, and all of the nonnegativess constraints are active, we have $\alpha_i = 0$, $\forall i$. This case is not meaningful unless $CB = 0$, which in turn is not an meaningful cost budget value.

We now state the necessary conditions for local optimality at point α at which LICQ holds. Suppose $\bar{\boldsymbol{\alpha}}$ is a local minimizer at which LICQ holds. Then $\exists\ \bar{\mathbf{u}} \in \mathcal{R}^m$ such that [156, Thm 12.1]

$$\nabla e(\bar{\boldsymbol{\alpha}}) + \nabla \mathbf{a}(\bar{\boldsymbol{\alpha}})\bar{\mathbf{u}} = 0 \tag{3.11}$$

$$\bar{\mathbf{u}} \geq \bar{0} \tag{3.12}$$

$$\mathbf{a}(\bar{\boldsymbol{\alpha}}) \leq 0 \tag{3.13}$$

$$\bar{\mathbf{u}}^{\mathrm{T}}\mathbf{a}(\bar{\boldsymbol{\alpha}}) = 0 \tag{3.14}$$

These are the Karush-Kuhn-Tucker (KKT) Conditions for a problem with no equality constraints.

Since for the problem at hand LICQ holds at every feasible point, KKT conditions should be satisfied for any local optima, hence any global optima. Hence the feasible points at which KKT conditions are satisfied are candidates for global optima. Thus to find the optimum one can follow the following procedure: first find the set points that satisfy the KKT conditions, find the associated objection function values, and choose the one with the best objective function value as the global optimum. Although this enumeration approach may not be feasible for large M, it is tractable for $N = M = 2$, and yields to important insights about the structure of the problem.

In the rest of this development, we will consider the case $N = M = 2$, $\mathbf{K_n} = \mathbf{0}$, and $\mathbf{H}$ is the identity matrix. Here $\mathbf{K_f}$ can be expressed as follows

$$\mathbf{K_f} = \begin{pmatrix} k_{f_{11}} & k_{f_{12}} \\ k_{f_{21}} & k_{f_{22}} \end{pmatrix} \tag{3.15}$$

where $k_{f_{12}} = k_{f_{21}}$.

We will now enumerate the possible cases regarding the active and inactive constraints.

Cost Constraint is Inactive, i.e. not all the allowed budget is spent: No such feasible point can be a candidate for the global optima, since by using more of the cost budget, one can improve the objective function, i.e. achieve smaller error values. Hence, in this case it is not necessary to study the active constraints sets. Nevertheless, here we present them for the sake of completeness.

(i) $\mathcal{J} = \emptyset$: $\mathbf{u} = [0,\ 0,\ 0]^{\mathrm{T}} \geq \mathbf{0}$. There is no solution for α.

(ii) $\mathcal{J} = \{1\}$: $\alpha_1 = 0$, $\mathbf{u} = [0,\ u_2,\ 0]^{\mathrm{T}} \geq \mathbf{0}$. There is no solution for α.

(iii) $\mathcal{J} = \{2\}$: $\alpha_2 = 0$, $\mathbf{u} = [0,\ 0, u_3]^{\mathrm{T}} \geq \mathbf{0}$. There is no solution for α.

(iv) $\mathcal{J} = \{1, 2\}$: $\alpha_2 = 0$, $\alpha_1 = 0$, $\mathbf{u} = [0,\ u_2,\ u_3]^{\mathrm{T}} \geq \mathbf{0}$. Here $a_0 = 0$ trivially, this case is not meaningful unless $C_{\mathrm{B}} = 0$.

Cost Constraint is Active, i.e. all the cost budget is spent on the measurement points:

(v) $\mathcal{J} = \{0, 1, 2\}$: $\alpha_2 = 0$, $\alpha_1 = 0$, $\mathbf{u} = [u_1,\ u_2,\ u_3]^{\mathrm{T}} \geq \mathbf{0}$. Here $a_0 = 0$ trivially, this case is not meaningful unless $C_{\mathrm{B}} = 0$, which in turn is not an meaningful cost budget value.

(vi) $\mathcal{J} = \{0\}$, i.e. both measurements are done using all the allowed budget: $\mathbf{u} = [u_1,\ 0,\ 0]^{\mathrm{T}} \geq \mathbf{0}$.

(vii) $\mathcal{J} = \{0, 1\}$, i.e. only the second component is measured using all the allowed budget: $\alpha_1 = 0$, $\mathbf{u} = [u_1,\ u_2,\ 0]^{\mathrm{T}} \geq \mathbf{0}$. Such a feasible point $\boldsymbol{\alpha} = [0,\ \alpha_2]^{\mathrm{T}}$ exist if there exist a u such that $\mathbf{u} = [u_1,\ u_2,\ 0]^{\mathrm{T}} \geq \mathbf{0}$ satisfying the following equations

$$\nabla e(\bar{\boldsymbol{\alpha}}) + \nabla a_0(\bar{\boldsymbol{\alpha}}) u_1 + \nabla a_1(\bar{\boldsymbol{\alpha}}) u_2 = 0 \tag{3.16}$$

$$u_i \geq 0 \tag{3.17}$$

$$\alpha = [0,\ \frac{(2^{2C_{\mathrm{B}}} - 1)}{k_{f_{22}}}]^{\mathrm{T}} \tag{3.18}$$

such $\boldsymbol{\alpha}$ vectors will be candidates for local optima.

(viii) $\mathcal{J} = \{0, 2\}$, i.e. only the first component is measured using all the allowed budget Here $\alpha_2 = 0$, $\mathbf{u} = [u_1, 0,\ u_3]^{\mathrm{T}} \geq \mathbf{0}$. This case is similar to the previous case where only the second component is measured. The conditions for such a feasible point to exist can be obtained by rewritting (3.16)-(3.18) with a change of indices.

It is possible to find the conditions imposed on the problem parameters by the requirement that the system of equations and inequalities in (3.16)-(3.18) have a solution. But the resulting equations are algebraically involved, and are not in a form that is open to direct interpretation. Hence here we will first make a few remarks on the nature of these solutions, and then consider the special case where $k_{f_{11}} = k_{f_{22}} = 1$. The conditions in (3.16)-(3.18) yield a second order concave polynomial in $L = 2^{2C_{\mathrm{B}}} - 1 \geq 0$, whose value is required to be non-negative to satisfy (3.17). For it is to be possible to satisfy these equations with some cost budget values, the maximum root of this polynomial should be positive so that for some cost budget values this polynomial can produce positive values. This requirement gives the conditions on $\mathbf{K_f}$ so that we can have “measuring

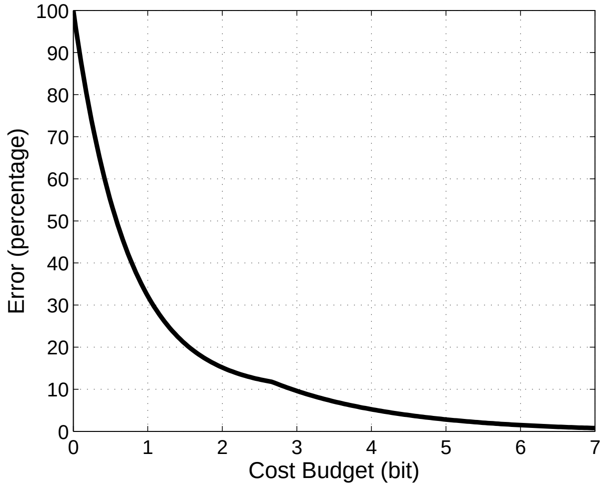


Figure 3.2: Error vs. cost budget, $k_{f_{11}} = k_{f_{22}} = 1$, $k_{f_{12}} = 0.9$.

only one component type" candidates for the local optima. For the polynomial evaluated with a particular value of the cost budget to produce a positive value, the associated L value should be between the roots of the polynomial. This requirement gives the condition on the cost budget.

We now study the special case where $k_{f_{11}} = k_{f_{22}} = 1$. If $|k_{f_{12}}| < \frac{1}{\sqrt{3}} \approx 0.58$, regardless of the allowed cost, it is better to measure both of the components. The optimum α_i's are given by $\alpha_1 = \alpha_2 = 1 + \exp(C_{\mathrm{B}})$. Here Item (vii) or Item (viii) do not provide candidates for global optimum, hence regardless of the allowed cost, we measure both of the components. If $|k_{f_{12}}| > \frac{1}{\sqrt{3}}$, measuring only one of the components with all the cost budget at hand is the globally optimum scheme, if the cost budget satisfies the following equation

$$C_{\mathrm{B}} \leq 0.5 \log(\frac{3k_{f_{12}}^{2} - 1}{(1 - k_{f_{12}}^{2})^2} + 1). \tag{3.19}$$

This threshold is found by solving (3.16)-(3.18). Now the cases described in Items (vii) and (viii) provide candidates for global optimum. Comparing the objective function values provided by the solution of these equations with the one in Item (vi) reveals that measuring only one of the components is the globally optimum scheme for such cost budget values.

To sum up, while measuring a random vector with two components having the same variance, the optimal measurement strategies are found as follows: if the random signal is relatively uncorrelated (the correlation coefficient between two components is less than $\frac{1}{\sqrt{3}}$), for all cost budget values, measuring both of the components with the same number of distinguishable levels is optimal for all cost budget values. If the unknown signal is relatively correlated (the correlation coefficient is larger than $\frac{1}{\sqrt{3}}$), measuring only one component or both of the component can be optimal depending on the cost budget available: for relatively low cost budget values (cost budget values smaller than the bound given (3.19).), measuring only one the components with all the cost budget at hand is optimal. Here any of the components can be measured. On the other hand, if the cost budget is larger than this threshold value, measuring both of the components with equal cost allocation is optimal. Considering the optimal cost allocations for different cost budget values around the treshold in (3.19) we make the following observation: Depending on the problem parameters, it may be necessary to use a measurement device with less number of distinguishable levels at a point where previously a more precise measurement device is assigned.

We now illustrate the piece-wise error-cost curve with an example. Let $k_{f_{12}} = 0.9$. By Eqn. 3.19, for $C_{\mathrm{B}} \lesssim 2.67$ bits, it is better to spend all the cost budget on only one component. For larger cost budget values, measuring both of the components is more better. Fig. 3.2 presents the resulting the cost-error trade-off curve, where the error is reported as a percentage defined as $100\,\varepsilon(C_{\mathrm{B}})/\operatorname{tr}(\mathbf{K_f})$. The curve illustrates the piece-wise behaviour of the cost-error trade-off with the joint point at $C_{\mathrm{B}} \approx 2.67$ bits.

Finally, we would like to note that when the error-cost curves are piece-wise, the region formed by the achievable error and cost budget pairs is not convex. Nevertheless, one can use the following time sharing approach to achieve the error-cost pairs that are in the convex hull of this region, but not reachable with the current setting. Let us choose two cost budget values C_{B}^1 and C_{B}^2, where $C_{\mathrm{B}}^1/\ C_{\mathrm{B}}^2$ is smaller/larger than the bound given in (3.19). For $\theta \in [0, 1]$ of the total time, we employ the strategy of measuring only one of the components with the cost C_{B}^1, and in the remaining $1 - \theta$ of the time, we employ the strategy of measuring both of the components with a total cost of C_{B}^2. Hence, the average cost over time will be given by the following expression: $C_{\mathrm{B}} = \theta C_{\mathrm{B}}^1 + (1-\theta)C_{\mathrm{B}}^2$. Let the error values achieved with these strategies be e^1 and e^2, respectively. Hence, the average error over the total time will be given by the following expression: $e = \theta\, e^1 + (1-\theta)e^2$. Thus, an average error of e is achieved under a cost budget C_{B}. By choosing different C_{B}^1 and C_{B}^2, and different time sharing ratios θ, one will be able to achieve the error-cost values in the convex hull of the region. In particular, by choosing cost budget values relatively close to the threshold in (3.19), one will able to achieve error-cost values that would not have been achievable if we hadn't used the time sharing approach.

3.3.0.2 Repeated Measurements of the Field at a Single Point

As noted, repeated measurements of components of $\mathbf{g}$ are always suboptimal in the sense that doing so results in greater error for given cost. Here we allow more than one measurement of any component of $\mathbf{g}$ and show that this is indeed the case. We assume that different measurements are statistically independent conditional on $\mathbf{g}$ even if repeated measurements of the same component are in question. This result was first proved in [3], here we provide a more compact presentation.

First we consider the simple case in which repeated measurements are made at a single point g_i and the other components of $\mathbf{g}$ are not measured. That is, one is allowed to make P_i measurements on g_i indexed by $j = 1, \ldots, P_i$ as $s_j^i = g_i + m_j^i$ subject to the usual cost constraint. Here the subscript denotes the index of the component of $\mathbf{g}$ where the repeated measurements are made. Since no other component of $\mathbf{g}$ is measured, the total number of measurements is equal to the number of repeated components P_i, the measurement noise vector $\mathbf{m}^i = [m_1^i, \ldots, m_{P_i}^i]^{\mathrm{T}} \in \mathbb{R}^{P_i}$, and the measurement vector $\mathbf{s}^i = [s_1^i, \ldots, s_{P_i}^i]^{\mathrm{T}} \in \mathbb{R}^{P_i}$. We consider the problem of estimation of a single component of the input field f_k where $k \in 1, \ldots, N$. By studying this case, we wish to see which measurement strategy is better: (i) to make one high quality measurement by renting the best device within budget limits, or (ii) to split the budget among multiple lower quality devices. Simple LMMSE analysis shows that the first alternative is better, as we now show.

For any given allocation of noise variances $(\sigma^2_{m_1}, \ldots, \sigma^2_{m_{P_i}})$, the P_i measurements yield the LMMSE estimate $\hat{f}_k(\mathbf{s}) = \mathbf{a}^{\mathrm{T}}\mathbf{s}$ where $\mathbf{a} \in \mathcal{R}^{P_i}$. Here the components of $\mathbf{a}$ are obtained by solving the orthogonality conditions:

$$a_j = \frac{E[f_k g_i]}{\sigma^2_{\text{eq}} + \sigma^2_{g_i}} \frac{\sigma^2_{\text{eq}}}{\sigma^2_{m_j}}, \tag{3.20}$$

where $\sigma^2_{\text{eq}} = \left(\sum\limits_{j=1}^{P_i} \frac{1}{\sigma^2_{m_j}} \right)^{-1}$. The associated MSE is

$$\varepsilon_i = \sigma^2_{f_k} - \frac{E[f_k g_i]^2}{\sigma^2_{\text{eq}} + \sigma^2_{g_i}}. \tag{3.21}$$

The total measurement cost for this scheme is $\sum_{j=1}^{P_i} \frac{1}{2} \log \left(1 + \sigma^2_{g_i}/\sigma^2_{m_j}\right)$. We observe that among all schemes of allocation of noise variances yielding the same σ^2_{eq} (hence giving the same MSE), the cost is minimized by taking $\sigma^2_{m_j} = \sigma^2_{\text{eq}}$ for any one of the indices j and $\sigma^2_{m_j} = \infty$ for the others. This corresponds to making one high quality measurement. Therefore for a given error, the total cost is minimized by making one high quality measurement rather than many low quality ones. The error is a strictly decreasing function of the cost so that we can further conclude that this is also the strategy minimizing error for given cost.

We note that this result trivially holds when one wants to estimate the whole field vector $\mathbf{f} \in \mathbb{R}^{\mathbb{N}}$ instead of a single component f_k of the vector. It also remains true when other components of $\mathbf{g}$ are measured alongside with g_i, as can be shown by noting that the estimation errors for the components of $\mathbf{g}$ do not change as long as σ^2_{eq} is the same, so that the estimator coefficients associated with these components and therefore the estimation error for $\mathbf{f}$ also do not change. Therefore we conclude that allowing repeated measurements of the same point does not provide an opportunity for further optimization, since for every measurement scheme involving more than one measurement of the same component, it is certain that there is another scheme that yields the same error with a lower cost budget.

3.3.0.3 Uncorrelated Case

In order to see the relationship of our formulation with the "water-filling" solutions common in certain information-theoretic problems (e.g., [155, Ch.9], [40, Ch.13]), we consider the special case where $N = M$, the matrix $\mathbf{K_f}$ is diagonal, $\mathbf{K_n} = \mathbf{0}$, and H is the identity matrix. Hence both the components of f, and the components of s are uncorrelated.

In this special case we have N separate LMMSE problems tied together by a total cost constraint. By standard techniques [155, Ch.9], [40, Ch.13], which are illustrated in [3], the optimal detector variances that minimize the estimation error can be obtained as follows

$$\sigma_{m_i}^2 = \begin{cases} \frac{\nu\sigma_{f_i}^2}{\sigma_{f_i}^2 - \nu} & \text{if } \sigma_{f_i}^2 - \nu > 0 \\ \infty & \text{if } \sigma_{f_i}^2 - \nu \leq 0 \end{cases} \tag{3.22}$$

where the parameter ν is selected so that the total cost is C_{B}. Notice that for those components for which there is a non-trivial measurement ($\sigma_{m_i}^2 < \infty$), we have $1/\sigma_{m_i}^2 + 1/\sigma_{f_i}^2 = 1/\nu$, which is reminiscent of the "water-filling" solutions referred to above.

3.3.0.4 Accurate Measurements (High Budget) Case

When the uncertainty introduced by the measurements are small with respect to the range of $\mathbf{g}$, we refer to this case as the accurate measurements case. This is the case where $\mathbf{K_s}$ is near $\mathbf{K_g}$, where $\mathbf{K_g} = \mathbf{H}\mathbf{K_f}\mathbf{H}^{\mathrm{T}} + \mathbf{K_n}$ is the covariance of $\mathbf{g}$. Hence we may use the first order approximation of the inverse of a positive definite symmetric matrix to write $\mathbf{K_s}^{-1} \approx \mathbf{K_g}^{-1} - \mathbf{K_g}^{-1}\mathbf{K_m}\mathbf{K_g}^{-1}$, and using the linearity of the trace operator, the MMSE can be written as

$$\operatorname{tr}\left(\mathbf{K_f} - \mathbf{K_f}\mathbf{H}^{\mathrm{T}}\mathbf{K_g}^{-1}\mathbf{H}\mathbf{K_f}\right) + \operatorname{tr}\left(\mathbf{K_f}\mathbf{H}^{\mathrm{T}}\mathbf{K_g}^{-1}\mathbf{K_m}\mathbf{K_g}^{-1}\mathbf{H}\mathbf{K_f}\right). \tag{3.23}$$

The error is expressed as the sum of two parts. The first part is independent of the accuracy of the measurements. For physical phenomena represented by noninvertible matrices $\mathbf{H}$, this irreducible error remains even if the measurements are perfect, and corresponds to the limited information transfer capability of the physical system. The second additive error component is due to the imperfect measurements. In this case the estimation error is a linear function of $\mathbf{K_m}$, and the resulting optimization problem is convex. Since the objective and constraint functions are differentiable and Slater's condition holds, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality [151, Ch.5]. Hence by solving the KKT conditions, the optimal noise levels can be obtained as [3]

$$\sigma^2_{m_i} = \frac{-\sigma^2_{g_i} + \sqrt{\sigma^4_{g_i} + \frac{4\sigma^2_{g_i}}{\nu d_{ii}}}}{2}, \tag{3.24}$$

where $\nu > 0$ is a parameter chosen so that the total cost is C_{B}, and d_{ii}'s are the diagonal elements of $\mathbf{D} = \mathbf{K_g^{-1}HK_f^{2}H}^{\mathrm{T}}\mathbf{K_g^{-1}}$.

3.4 Trade-offs between Error and Cost Budget

First we present the algorithm we employed for solving the optimization problem (3.6). Our algorithm is based on (3.4) and relies on taking turns in fixing $\mathbf{B}$ and $\mathbf{K_m}$ and minimizing over the other. For fixed $\mathbf{K_m}$, the optimal value of the linear estimator $\mathbf{B}$ that minimizes the error can be analytically written in terms of $\mathbf{K_m}$ as $\mathbf{B} = \mathbf{K_f H}^{\mathrm{T}}\left(\mathbf{HK_fH}^{\mathrm{T}} + \mathbf{K_n} + \mathbf{K_m}\right)^{\mathbf{-1}}$. On the other hand, if we fix $\mathbf{B}$, the problem is to minimize $\operatorname{tr}\left(\mathbf{BK_mB}^{\mathrm{T}}\right)$ over $\mathbf{K_m}$ subject to (3.5). Since the differentiability and Slater's condition hold in this case as well, the optimal noise levels can be found as

$$\sigma^2_{m_i} = \frac{-\sigma^2_{g_i} + \sqrt{\sigma^4_{g_i} + \frac{4\sigma^2_{g_i}}{\eta p_{ii}}}}{2}, \tag{3.25}$$

by solving the KKT conditions. Here $\eta > 0$ is a parameter chosen so that the total cost is C_{B}, and p_{ii}'s are the diagonal elements of $\mathbf{P} = \mathbf{B}^{\mathrm{T}}\mathbf{B}$.

The resulting algorithm is summarized as follows (Fig. 3.3): We initialize the algorithm by setting $t = 0$ and $\mathbf{K_m}^{(t=0)}$ to a random positive-definite diagonal matrix. At each iteration t, first we fix $\mathbf{K_m}$ and set $\mathbf{B}^{(t+1)} = \mathbf{K_f H}^{\mathrm{T}}\left(\mathbf{K}_s^{(t)}\right)^{-1}$ where $\mathbf{K}_s^{(t)} = \mathbf{HK_fH}^{\mathrm{T}} + \mathbf{K_n} + \mathbf{K}_{\mathbf{m}}^{(t)}$, which is the optimum value of $\mathbf{B}$ for $\mathbf{K_m}^{(t)}$. Then we fix $\mathbf{B}$ and minimize over $\mathbf{K_m}$: We obtain $\mathbf{K}_{\mathbf{m}}^{(t+1)}$ by solving equation

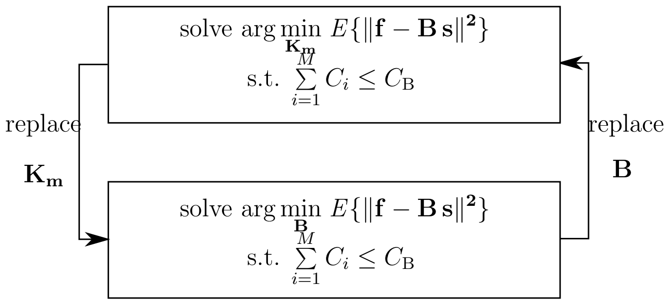

solve $\arg\min_{\mathbf{K_m}} E\{\|\mathbf{f}-\mathbf{B}\,\mathbf{s}\|^{\mathbf{2}}\}$
s.t. $\sum_{i=1}^{M} C_i \leq C_{\mathrm{B}}$
replace
$\mathbf{K_m}$
replace
$\mathbf{B}$
solve $\arg\min_{\mathbf{B}} E\{\|\mathbf{f}-\mathbf{B}\,\mathbf{s}\|^{\mathbf{2}}\}$
s.t. $\sum_{i=1}^{M} C_i \leq C_{\mathrm{B}}$

Figure 3.3: Block diagram of the algorithm.

(3.24) with p_{ii} replaced with $a_i = \sum_{j=1}^{M}\left(b_{j,i}^{(t+1)}\right)^2$. For the stopping criterion, we use the relative error: if $\varepsilon(C_{\mathrm{B}})/\operatorname{tr}(\mathbf{K_f})$ does not change by more than 10^{-4} over 10 consecutive iterations, we stop; otherwise, we increment t and continue. Since the sequence of error values form a monotonically decreasing sequence and the error is bounded from below, the sequence of error values produced by this algorithm is guaranteed to converge to a limit value. In practice, the algorithm stops typically within 10-150 iterations depending on the problem parameters. Details on this type of algorithm may be found, for instance in [157].

The problem we formulate and solve in this chapter was motivated by the physical problem of measuring propagating wave fields at a certain number of points and estimating the values of the field, possibly at other, distant locations. Although our formulation can handle very general cases of this problem, in our numerical examples we will focus on the case where there are two planar or spherical reference surfaces, perpendicular to the axis of propagation and separated by a certain distance. We assume that all measurement probes are placed uniformly on one surface and we desire to estimate the field on the other surface. In this case the measured field is related to the unknown field through a diffraction integral, a convenient approximation of which is the Fresnel diffraction integral or more generally a quadratic-phase integral (linear canonical transform) [158, Ch.8], [159, Ch.2], and [160, 161]. It is well known that these integrals can be expressed in terms of the fractional Fourier transform (FRT), which provides an elegant and pure description of these systems [18, Ch.9], [162], and which has found many applications in signal processing [163–170]. The FRT is the fractional operator power of the Fourier transform with fractional order a. When $a = 0$ the FRT reduces to the identity operation and when $a = 1$ it reduces to the ordinary Fourier transform. Moreover, the transform is index-additive: the a_1th transform of the a_2th transform is equal to the $a_1 + a_2$th transform. Further information on the FRT and its computation may be found in [18, 62]. Essentially, the FRT captures the underlying physics of wave propagation and diffraction phenomena in its purest form and is therefore suitable for modeling wave propagation for our present purposes. Thus in our examples we will take the system matrix $\mathbf{H}$ to be the N by N real equivalent of the $N/2$ by $N/2$ complex FRT matrix. For the generation of FRT matrices of different orders, an implementation of the algorithm presented in [171] and in [18, Ch.6] is used; this implementation is available at [172].

Propagating wave-fields may have different degrees of what is known as *coherence*. Highly coherent fields are those whose values at different points are highly correlated with each other. Highly incoherent fields are those whose values at different points are highly uncorrelated. Since we have observed that our results depend on the degree of coherence of the fields, we will consider several covariance matrices corresponding to different degrees of coherence (correlation between their components). It is known that highly coherent fields have covariance matrices whose eigenvalues are highly unevenly distributed. On the other hand, highly incoherent fields have eigenvalues which are nearly equal to each other [58]. To obtain covariance matrices with different degrees of coherence, we will choose the eigenvalues to be normally distributed with standard deviation equal to N/α pixels. Here the parameter α can also be interpreted as the number of standard deviations of the Gaussian covered by the N samples. In our experiments α takes the values $\alpha = 0.25, 2, 16, 128, 1024$, where $\alpha = 1024$ corresponds to the case where all but one eigenvalue is negligible, and $\alpha = 0.25$ corresponds to the case where all eigenvalues are nearly equal. While α is a convenient parameter to work with, we note that it should not be seen as a linear measure of the degree of coherence [58]. To generate the covariance matrices with these eigenvalues, we use the eigenvalue-eigenvector decomposition of a covariance matrix $\mathbf{K} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{T}}$, where $\mathbf{\Lambda} = \mathrm{diag}(\varsigma_i)$. Here the orthogonal matrix $\mathbf{Q}$ is obtained by QR decomposition of a $N \times N$ matrix with i.i.d. zero-mean Gaussian entries.

For the system noise $\mathbf{n}$, the covariance matrix is generated similarly with $\alpha = 4$ with a different Q matrix.

Another important parameter used in the experiments is

$$\mathrm{SNR} \triangleq \frac{\mathrm{tr}(\mathbf{H}\mathbf{K}_{\mathbf{f}}\mathbf{H}^{\mathrm{T}})}{\mathrm{tr}\,(\mathbf{K}_{\mathbf{n}})} = \frac{\sum_{i=1}^{N} \sigma_{f_i}^2}{\sum_{i=1}^{M} \sigma_{n_i}^2}, \tag{3.26}$$

where the second form follows from $\mathbf{H}^{\mathrm{T}}\mathbf{H} = \mathbf{I}$ which in turn follows from the unitarity of the FRT. SNR measures the ratio of signal power to inherent system noise power, before measurements.

In the following experiments our main purpose will be to investigate the trade-off between the MSE error $\varepsilon(C_{\mathrm{B}})$ and measurement cost budget C_{B} after we have optimized over all possible allocations of cost over the measurement devices. The error will be reported as a percentage defined as $100\,\varepsilon(C_{\mathrm{B}})/\,\mathrm{tr}\,(\mathbf{K}_{\mathbf{f}})$. The cost budget C_{B} is measured in bits by taking logarithms to base 2. Unless otherwise stated all experiments are done with $a = 0.5$ and $N = M = 256$.

Effect of noise level on trade-offs: This experiment investigates the effect of SNR on the trade-off between C_{B} and $\varepsilon(C_{\mathrm{B}})$. In this experiment, SNR was variable, ranging over 0.1, 1, 10, ∞ and two different values of α were considered. Fig. 3.4 and Fig. 3.5 give the curves for low and high α values, respectively. We notice that for both of the cases $\varepsilon(C_{\mathrm{B}})$ is very sensitive to increases in C_{B} for smaller C_{B}. Then it becomes less responsive and eventually saturates to the error value corresponding to zero measurement noise. For each value of cost, the error decreases as SNR increases, and for higher cost values will approach zero as SNR $\to \infty$. We see that when the field is more highly coherent (Fig. 3.5), we obtain much better trade-off curves for all values of SNR than Fig. 3.4 which represents the highly incoherent extreme. For instance for SNR $= \infty$, for the highly incoherent field an error of 10% is obtained at a cost of 400 bits, whereas for the highly coherent field the same error is achieved at a cost lower than 5 bits. This point is further investigated in the experiment.

Effect of degree of coherence on trade-offs: This experiment investigates the effect of degree of coherence of the unknown field on the trade-off between C_{B}

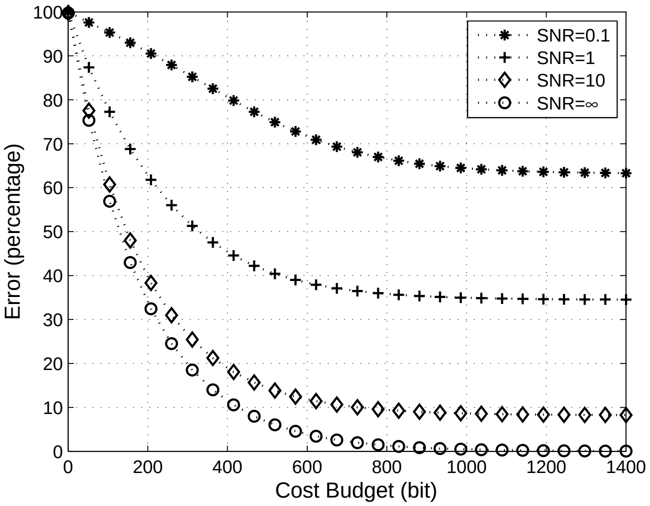


Figure 3.4: Error vs. cost budget for $\alpha = 0.25$, SNR variable.

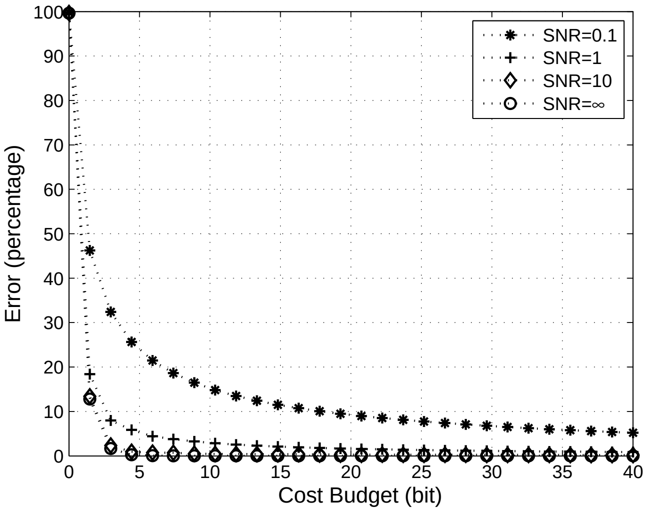


Figure 3.5: Error vs. cost budget for $\alpha = 1024$, SNR variable.

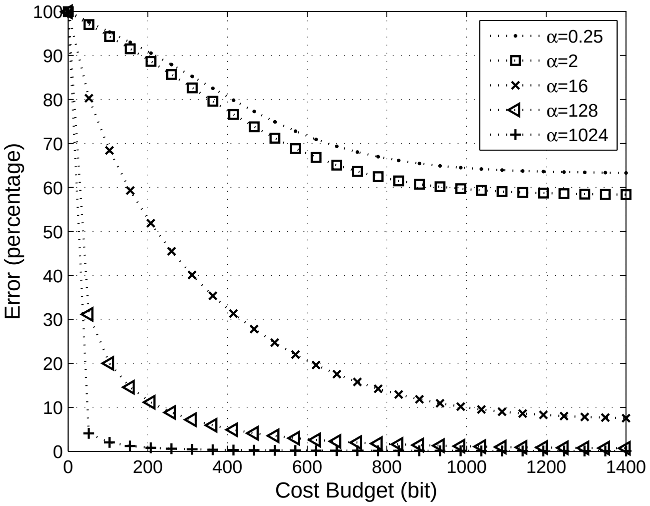


Figure 3.6: Error vs. cost budget for SNR = 0.1, α variable.

and $\varepsilon(C_{\mathrm{B}})$. Fig. 3.6 and Fig. 3.7 show the results for two different SNR values (SNR = 0.1 and SNR = ∞), for $\alpha = 0.25, 2, 16, 128, 1024$. Both of the plots show that for low values of α corresponding to lower degrees of coherence, it is more difficult to achieve low values of error within a given budget. But as α increases, the total uncertainty in the field decreases, and it becomes a lot easier to achieve lower values of error. In fact, for high values of α and for low values of budget, the optimal strategy to minimize error turns out to be to measure the field value at only a few points with more accurate (and costly) measurement devices, rather than spreading the cost budget among many measurement points. This observation is further investigated in the upcoming experiments.

It is interesting to note that in all of the numerical examples we have considered, including the incoherent case, it is possible to reach with an average of 4 bits per component, the same error level that would be achieved with infinite accuracy (and cost).

Comparing the performances in Fig. 3.6 and Fig. 3.7 for low and high values of the cost budget, we see that for low budget values the effect of degree of coherence of the field can be considered more pronounced in the high SNR case, whereas for high budget values this effect is more pronounced in the low SNR case: For high values of cost budget, it is always possible to obtain very low values of error

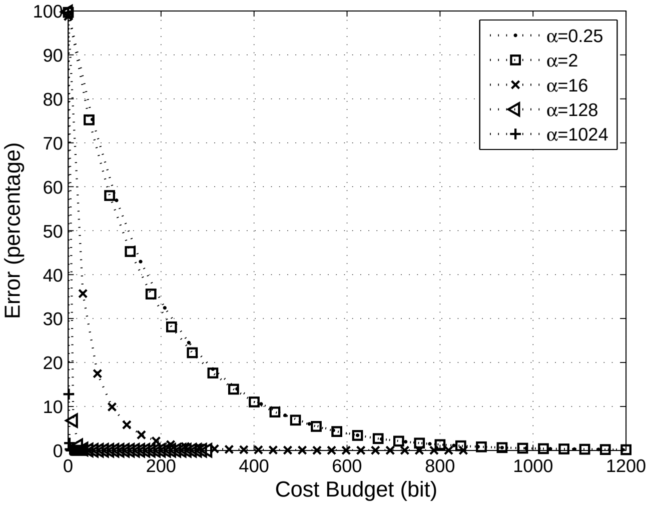


Figure 3.7: Error vs. cost budget for SNR $= \infty$, α variable.

(≈ 0) regardless of degree of coherence, when the SNR is high. But when the SNR is low and the cost budget is high, a substantial performance difference is observed between the correlated and uncorrelated fields, since it is possible to effectively cancel the effect of system noise n when the degree of coherence of the field is high, yielding a better performance. When the budget is small and the SNR is low, although highly correlated fields lead to better performance, this improvement is limited by the presence of noise. When the budget is small but SNR is high, it is possible to obtain very low values of error (≈ 0) when the field is highly correlated, resulting in a far better performance compared to the uncorrelated case.

Effect of noise level on the number of effective measurements: This experiment investigates the effect of SNR on the relationship between the number of effective measurements M_{eff} and the budget C_{B}. We will consider a measurement at a point to be effectively made if the cost of the measurement at this point is greater than $p\,(C_{\text{B}}/N)$ bits. With this choice of threshold, it is guaranteed that the total cost of the measurements that are effectively made is higher than $(1-p)\,C_{\text{B}}$. We use $p = 0.125$. Measurements with less cost are very noisy measurements and do not contribute much either to the quality of the estimate or the total cost, so that it does not make much difference whether we actually perform them or not.

For $\alpha = 0.25$, we see that one has to do measurements at all of the $M = 256$ measurement points for all values of SNR and for all values of C_{B}. This result is plausible since the field values are nearly uncorrelated in this case, and each point can be considered to provide new information.

The case of highly coherent fields is more interesting. Fig. 3.8 shows the results for $\alpha = 1024$ with SNR $= 0.1, 1, 10, \infty$. For low values of SNR, the optimal strategy is to split the budget relatively broadly among the M points. On the other hand, for high values of SNR, the best strategy is to allocate the budget to a smaller number of points. To understand this behavior, we observe that in this experiment the field values are highly correlated, hence the points measured carry nearly the same information. On the hand the system noise is highly uncorrelated. Based on these two observations, we can say that measuring a larger number of points increases the averaging effect and thus suppression of the system noise. Successively measuring highly correlated variables normally adds little information [so that one would prefer fewer but more accurate measurements instead.] However, when there is a lot of noise, the benefits of noise suppression can outweigh this so that a larger number of measurements are preferred.

Although the curves behave as if the number of effective measurements saturate at an asymptote for high values of cost budget, this is in fact not true and the number of effective measurements continue to increase as budget increases. This point is further discussed in the next experiment.

Effect of degree of coherence on the number of effective measurements: This experiment investigates the effect of degree of coherence of the unknown field on the relationship between the number of effective measurements M_{eff} and the budget C_{B}. Fig. 3.9 shows the results for SNR $= 0.1$, $\alpha = 0.25, 2, 16, 128, 1024$. We see that for all values of α, and for low values of cost budget, the best strategy is to measure more accurately a relatively smaller number of points. But as the budget increases, the information that can be gained by measuring the field at a limited number of points with greater and greater accuracy saturates and splitting the budget over a larger number of measurement points become beneficial. For low values of α, this shift in strategy takes place at lower values of cost budget.

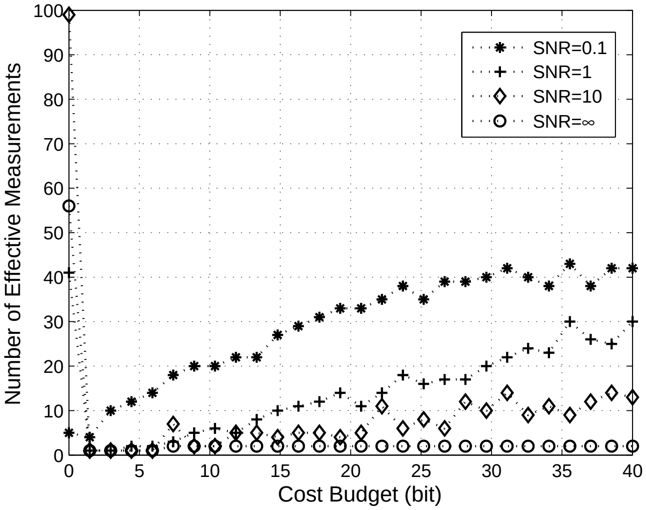


Figure 3.8: Effective number of measurements vs. cost budget for $\alpha = 1024$, SNR variable.

For a highly coherent field, the measurement of the field value at a particular point says much more about the field values at other points, and the benefit of measuring some of the field values with greater accuracy is prevailing.

Comparing this plot with Fig. 3.6 shows that the increase in the number of effective measurements for higher values of budget is not very meaningful since, for these budget values the error has almost reached its saturation value, but the algorithm being blind to this fact, increases the number of effective measurements to achieve tiny decreases in error. For instance, with $\alpha = 1024$, the error reaches a value of almost zero for a cost budget of 200 bits, and beyond this cost budget any increase in the number of measurements is made for the sake of a very small performance improvement.

We have also repeated the above experiment made for SNR = 0.1 for other values of SNR. We have observed that as SNR increases, a similar behavior is observed: the number of effective measurements again increases with increasing budget for all values of α. But this time the rate of increase of the number of effective measurements with increasing budget is smaller. Also, at a given cost budget, the ratio of the number of effective measurements is larger for different values of α. Hence the difference in the optimum cost allocation strategies for

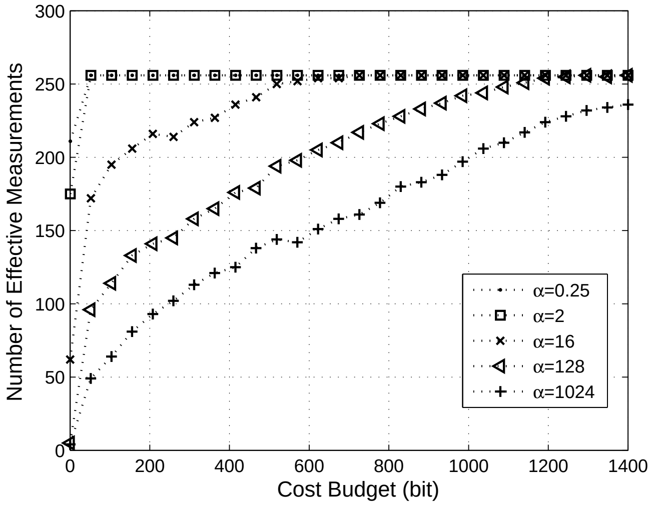


Figure 3.9: Effective number of measurements vs. cost budget for SNR $= 0.1$, α variable.

different values of α is more apparent for higher values of SNR.

Comparison to uniform cost allocation strategy: This experiment aims to demonstrate how applying the optimum cost allocation strategy we have employed up to this point, improves the trade-off between C_{B} and $\varepsilon(C_{\mathrm{B}})$ compared to a simple uniform cost allocation strategy, where the cost budget is equally allocated: $C_i = C_{\mathrm{B}}/M$, $i = 1, \ldots, M$. We expect that use of the optimal cost allocation will make a bigger difference for more highly coherent fields, since previous experiment shows that in this case the optimum cost allocation is drastically different from a uniform cost allocation scheme. Furthermore, Fig. 3.4 suggests this effect should be more pronounced when SNR is high. Fig. 3.10 compares the trade-off curves with optimum and uniform cost allocation schemes with $\alpha = 1024$, SNR $= 0.1, 1, 10, \infty$. The dashed curves and the straight lines show the results for the optimum cost allocation scheme and the uniform cost allocation scheme respectively. As expected, for all values of SNR, the optimum cost allocation scheme gives significantly better trade-offs compared to the uniform cost allocation case. For low C_{B} values, as SNR increases, the ratio of percentage error corresponding to uniform cost allocation to that corresponding to optimum cost allocation increases, showing that when the degree of coherence is high and the system noise is small, it is more important to optimize the allocation of the

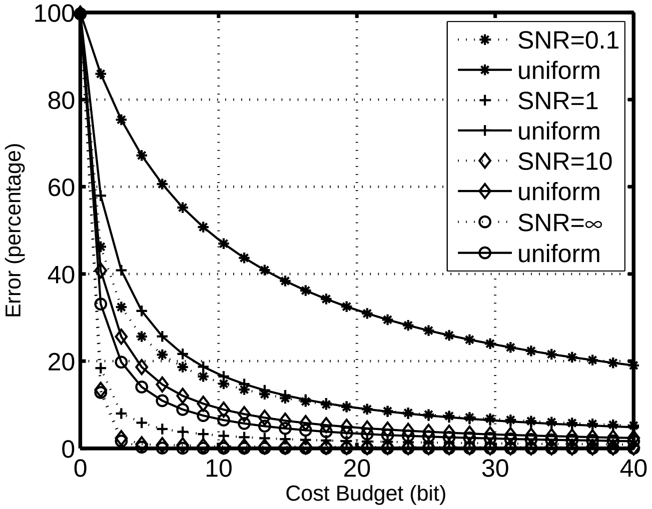


Figure 3.10: Error vs. cost budget for $\alpha = 1024$, SNR variable. The dotted lines are for optimal cost allocation and the corresponding solid lines are for uniform cost allocation.

budget to the measurement points.

Effect of making measurements at a smaller number of points: This experiment investigates the effect of making measurements at a smaller number of points. More specifically, we will examine the dependence of $\varepsilon(C_{\mathrm{B}})$ on $M_{\mathrm{s}} \leq M$ for a fixed C_{B}. Fig. 3.11 shows the results for $a = 0.5$,$N = M = 256$, SNR $= \infty$, $\alpha = 16$ and $M_{\mathrm{s}} = 8, 16, 32, 64, 128, 256$. The measurement locations were chosen as uniformly spaced subgrids of the full 256-point grid (i.e. the grid for $M_{\mathrm{s}} = 32$ was a sub-grid of that for $M_{\mathrm{s}} = 64$ which was a sub-grid of that for $M_{\mathrm{s}} = 128$, etc.). We see that for $M_{\mathrm{s}} = 64$ and $M_{\mathrm{s}} = 128$ roughly the same performance with the $M_{\mathrm{s}} = M = 256$ case is observed, whereas for other values the performance degrades with decreasing M_{s}. This behavior is related to the effective number of nonzero eigenvalues. For $\alpha = 16$, the eigenvalues are samples of a Gaussian with standard deviation 256/16 pixels. Assuming the values of a Gaussian beyond its third standard deviation are negligible, the covariance matrix has about $3 \times 256/16 = 48$ nonzero eigenvalues. Indeed we observe that as long as the number of measurements M_{s} is higher than 48, the trade-off curves are similar to the $M_{\mathrm{s}} = M$ case. But if we restrict ourselves to do measurements at a smaller number of points such as $M_{\mathrm{s}} = 8, 16, 32$, a substantial performance degradation

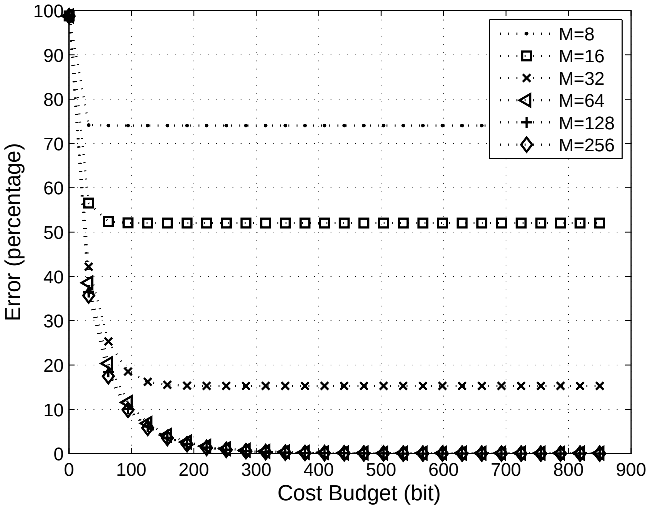


Figure 3.11: Error vs. cost budget for $N = 256$, $\alpha = 16$, SNR $= \infty$, M variable.

is observed.

3.5 Conclusions

Motivated by problems related to measurement of propagating wave-fields, we formulated the problem of optimally measuring observed variables so as to estimate unknown variables under a total cost constraint. We proposed a measurement device model where each device has a cost depending on its resolving power. Based on this cost function we determine the number of measurement devices and their accuracies that minimize estimation error for given total cost. We produce tradeoff curves between the error and the cost budget, corresponding to the optimal measurement strategy. We discuss the effects of SNR, distance of propagation, and the degree of coherence of the wave-fields on these trade-offs.

Specific hardware may deviate from our hardware-independent cost-budget model to varying degrees. However, all measurement devices have finite accuracy and in general their cost is an increasing function of their accuracy. Therefore, we believe that the nature of the tradeoffs observed and the general conclusions and insights will remain useful under a wide variety of circumstances.

We have seen that making measurements with higher quality (and cost) measurement devices, should be preferred over making repeated measurements with lower cost (and quality) devices. This helps explain why it is better to make a limited number of high quality measurements when the field is highly coherent. At the other extreme of coherence, when the fields are uncorrelated, we noted that the best measurement strategy is a reverse-water filling scheme.

As expected, in our numerical experiments we observe that the estimation error decreases with increasing cost budget, and reaches zero error when there is no system noise. Not surprisingly, with increasing system noise levels (decreasing SNRs), poorer trade-offs are observed. The cost-error trade-off is greatly degraded by decreasing SNR for relatively incoherent fields, whereas it can be said to be less sensitive to SNR for coherent fields.

In general, it is possible to obtain better trade-off curves for relatively coherent fields as compared to relatively incoherent fields for all values of SNR. The difference can be quite substantial and in the limit of full coherence/incoherence very large. For instance, for a coherent field, a total cost of a few bits may be sufficient to obtain a certain error, whereas for an incoherent field one may need a total cost which is of the order of N times as large as this to achieve the same error. For relatively incoherent fields the best measurement strategy is to measure a greater number or most of the field components, whereas for relatively coherent light it is better to allocate the cost budget among a smaller number of field components. How small a number also depends on the SNR. It is preferable to measure a somewhat larger number of components when the SNR is low, but still many of the field components remain effectively unmeasured. These observations underline the fact that the degree of coherence (correlation) is a fundamental parameter that can have a significant effect on the results and therefore should be taken into consideration in order to ensure general applicability.

Chapter 4

Joint Optimization of Number of Samples, Sample Locations and Measurement Accuracies: Uniform Case

In Chapter 3, we have introduced a cost budget framework which focuses on the effect of limited amplitude accuracies of the measurements in signal reconstruction. There, we have formulated the problem in a discrete framework, whereas now we will formulate this problem in a continuous framework. We may summarize our approach as follows: we consider the problem of efficient representation of a finite-energy non-stationary field using a finite number of bits. A finite number of samples of the field is used for the representation. Each sample is of finite accuracy; that is, there is a finite number of distinguishable amplitude levels in each sample. Therefore one can use a finite number of bits to represent each sample. The total number of bits used for all of the samples constitutes the bit cost associated with the representation. For a given bit cost budget, we determine the optimum number, locations and the accuracies of the samples in order to represent the field with as low error as possible. We consider two different cases under this framework: i) uniform case: samples are equally spaced and each sample is taken with the same cost. In this case, we determine the optimum number and spacing of the samples. ii) nonuniform case: the more general problem where the number, locations, and accuracies of the samples can be chosen freely. In this case, samples need not be equally spaced from each other, and they can be taken with possibly different accuracies. The first case will be the subject of this chapter, whereas the second case will be investigated in Chapter 5.

One of the questions we ask here is the following: Given that in practice the samples will have limited amplitude accuracy, is it possible to achieve lower reconstruction errors by choosing to sample at a rate different than the Nyquist rate? Although one may expect to compensate for the limited accuracy of the samples by oversampling, the precise relationships between the sampling parameters and the reconstruction error are not immediately evident. In this chapter we give quantitative answers to this question by determining the optimal sampling parameters and the resulting performance bounds for the best achievable error for a given bit budget.

We now present an overview of this chapter. In Section 4.1, we present our general mathematical framework. We show the invariance of our cost error trade-off curves for GSM fields propagating through first order systems in Section 4.2. In Section 4.3, we present the optimum sampling strategies and the trade-off curves between the cost and the error. We compare our optimal trade-off curves with the ones that would be obtained if Shannon-Nyquist sampling theorem was used as the guideline in Section 4.4. We conclude in Section 4.5.

4.1 Problem Formulation

Let the input field $f(x)$ reside in the $z = 0$ plane, which is perpendicular to the optical axis z. Considering only one transverse dimension for simplicity, let $f(x)$ be a zero-mean finite-energy proper complex Gaussian random field (random process). $f(x)$ passes through a possibly noisy linear system to produce the output $g(x)$

$$g(x) = \mathcal{L}\{f(x)\} + n(x), \tag{4.1}$$

where $\mathcal{L}\{.\}$ denotes the linear optical system, and $n(x)$ is a random field denoting the system noise. $n(x)$ is modelled as a zero-mean proper complex Gaussian random field. We assume that the unknown random field $f(x)$ and the system noise $n(x)$ are statistically independent. We consider all signals and estimators over some bounded domain D. Let $K_f(x_1, x_2) = E[f(x_1)f^*(x_2)]$ and $K_n(x_1, x_2) = E[n(x_1)n^*(x_2)]$ denote the covariance functions of $f(x)$ and $n(x)$, respectively. Here * denotes complex conjugation. We assume that $f(x)$ is a finite-energy random field, $\int_{-\infty}^{\infty} K_f(x, x)dx < \infty$, and $K_n(x, x)$ is bounded.

M finite-accuracy samples of $g(x)$ are taken at the sampling locations $x = \xi_1, \ldots, \xi_M \in \mathbb{R}$. The limited amplitude accuracy of the samples is modelled through an additive noise field

$$s_i = g(\xi_i) + m_i, \tag{4.2}$$

We assume that the m_i's are independent, zero mean, proper complex Gaussian random variables. We further assume that the m_i's are statistically independent of $f(x)$ and $n(x)$. By putting s_i in vector form, we obtain the vector of observations $\mathbf{s} = [s_1, \ldots, s_M]^{\mathrm{T}}$.

There is a cost associated with each sample. The cost associated with the ith sample is given by $C_{s_i} = \log_2(\sigma_{s_i}^2/\sigma_{m_i}^2)$ and is measured in bits. Here $\sigma_{s_i}^2 = E[|s_i|^2]$ and $\sigma_{m_i}^2 = E[|m_i|^2]$, so that $\sigma_{s_i}/\sigma_{m_i}$ is essentially the ratio of the spread of the signal to the spread of the uncertainty, which corresponds to the number of distinguishable levels (dynamic range). Hence the logarithm of this number may be considered to provide a measure of the number of bits needed to represent this variable. For a field value at a given location, smaller noise levels (smaller $\sigma_{m_i}^2$) correspond to a sample with higher amplitude accuracy and higher cost. On the other hand, a larger noise level corresponds to lower amplitude accuracy and lower cost. Further discussion of this cost function can be found in Section 3.2.

With the vector $\mathbf{s}$ at hand, one can construct an estimate of the continuous field $f(x)$ given $\mathbf{s}$. How well can $f(x)$ be recovered based on $\mathbf{s}$? To make this question precise, we can find $\hat{f}(x \mid \mathbf{s})$: the minimum mean-square error (MMSE) estimate of $f(x)$ given $\mathbf{s}$. This is the estimate that will minimize the mean-square error between the original field and the reconstructed field given the observations $\mathbf{s}$. The error of this estimate will, of course, depend on the number, locations, and accuracies of the samples. We consider two different problems based on this general framework: i) equidistant sampling with uniform cost allocation ii) non-uniform sampling with non-uniform cost allocation. Here we will investigate the uniform version and the non-uniform version will be investigated in Chapter 5.

Here, the sampling locations $x = \xi_1, \ldots, \xi_M \in D$ are equidistant with the spacing Δ_x, and the midpoint $x_0 = 0.5(\xi_1 + \xi_M)$. The accuracy (hence the cost) associated with each sample is the same; that is $C_{s_i} = C_{s_1}$, $i = 1, \ldots, M$. The total cost of the representation is then simply $C_{\mathrm{T}} = \sum_{i=1}^{M} C_{s_i} = MC_{s_1} = MC_s$. For a given C_{B}, our objective is to choose the number of the samples M and the sampling interval Δ_x, while satisfying $C_{\mathrm{T}} \leq C_{\mathrm{B}}$, with the objective of minimizing the minimum mean-square error between $f(x)$ and $\hat{f}(x \mid \mathbf{s})$. We note that since the cost of each sample is assumed to be the same, by choosing the number of samples we also determine the cost of each sample.

This problem can be stated as one of minimizing

$$E\left[\int_D \|f(x) - \hat{f}(x \mid \mathbf{s})\|^2 dx\right], \tag{4.3}$$

over Δ_x, x_0, and M subject to

$$C_{\mathrm{T}} = MC_s \leq C_{\mathrm{B}}. \tag{4.4}$$

At this point it is worth recalling some of the properties of the MMSE estimation. As noted above, $\hat{f}(x \mid \mathbf{s})$ is the estimate that minimizes the mean-square error between $f(x)$ and $\hat{f}(x \mid \mathbf{s})$ for a given $\mathbf{s}$. The associated mean-square error $E\left[\int_D \|f(x) - \hat{f}(x \mid \mathbf{s})\|^2 dx\right]$ does not depend on the actual value of $\mathbf{s}$, but only on the joint probability distribution of $f(x)$ and $\mathbf{s}$. Under the current problem formulation, for a given cost budget C_{B}, this joint probability distribution is determined by the number of samples M, the sampling interval Δ_x, and the midpoint x_0. The formulation above seek the best choices for these design parameters.

We now provide some details regarding MMSE estimation. The MMSE estimate $\hat{f}(x \mid \mathbf{s})$ can be written as [133, Ch. 6].

$$\hat{f}(x \mid \mathbf{s}) = \sum_{j=1}^{M} h_j(x) s_j = \mathbf{h}(x)\mathbf{s} \tag{4.5}$$

where $\mathbf{h}(x) = [h_1(x), \dots, h_M(x)]$ We note that, given a set of samples, the set of functions $\mathbf{h}(x)$ are the optimal functions that minimize the mean-square error between the actual field and the reconstructed field. Here $\mathbf{h}(x)$ satisfies the equation [133, Ch. 6]

$$\mathbf{K_{f\,s}}(x) = \mathbf{h}(x)\mathbf{K_s}, \tag{4.6}$$

where $\mathbf{K}_{f\,\mathbf{s}}(x) = E[f(x)\mathbf{s}^\dagger] = [E[f(x)s_1^*], \dots, E[f(x)s_M^*]]$ is the cross covariance between the input field $f(x)$ and the representation vector $\mathbf{s}$, and $\mathbf{K_s} = E[\mathbf{s}\mathbf{s}^\dagger]$ is the auto-covariance of $\mathbf{s}$. The symbol † denotes complex conjugate transpose. To determine the optimal linear estimate, one solves this last equation for $\mathbf{h}(x)$. The resulting estimate $\sum_{j=1}^{M} h_j(x) s_j$ can be interpreted as the orthogonal projection of the unknown random field $f(x)$ onto the subspace generated by the samples s_j, with $h_j(x)$ being the projection coefficients. As in (2.7), the error can written more explicitly as follows

$$\varepsilon \;=\; \int_D (K_f(x,x) - \mathbf{K}_{f\,\mathbf{s}}(x)\mathbf{h}(x)^\dagger) dx. \tag{4.7}$$

Finally, we would like to recall that if D is taken large enough, $\varepsilon(C_\mathrm{B})$ becomes a good measure of representation performance for $f(x)$ over the entire space [Sec. 2.1]. More precisely, we have the following (2.12)

$$E[\int_{-\infty}^{\infty} \|f(x) - \hat{f}_D(x)\|^2 dx] = \int_{x \in D} E[\|f(x) - \hat{f}_D(x)\|^2 dx] + \int_{x \notin D} K_f(x,x) dx \tag{4.8}$$

where $\hat{f}(x \mid \mathbf{s})$ is shortly denoted as $\hat{f}(x)$. $\hat{f}_D(x)$ is defined as $\hat{f}_D(x) = \hat{f}(x)$ for $x \in D$ and $\hat{f}_D(x) = 0$ for $x \notin D$. Since $f(x)$ is finite energy, second term can be made arbitrarily close to zero by taking the region D large enough.

4.2 Trade-off curves for GSM fields are invariant under propagation through first-order optical systems

We will discuss some invariance results related to the cost-error trade-off curves for GSM fields propagating through first-order optical systems. These results generalize the results discussed in Section 2.3, where we have assumed that the amplitude accuracies of the samples are so high that the sample values can be assumed to be exact, and commented on the invariance of the trade-off curves between the number of samples and the error.

We consider the problem of sampling the output of a first-order optical system in order to represent the input optical field. Such systems encompass arbitrary concatenations of lenses, mirrors and sections of free space, as well as quadratic graded-index media [18,150]. In the next section, we will consider a given bit budget and find the minimum possible representation error for that budget. Varying the bit budget, we will obtain trade-off curves between the error and the cost budget (for instance, look forward to Fig. 4.1 for an example). Here we are concerned with how first-order optical systems change these trade-off curves. We will show that for GSM fields, the cost-error curves are invariant under passage through arbitrary $ABCD$ systems; that is, these systems have no effect on the error versus cost trade-off curves. Moreover, we show that the optimum sampling strategy at the output can be easily found by scaling the optimum sampling strategy at the input. We assume that the parameters A, B, C, D of the $ABCD$ matrix are real with $AD - BC = 1$. We first consider the case where there is no system noise $n(x)$, and then discuss the effects of noise.

Let us express the covariance function associated with a GSM field with parameters σ_I, β, R as

$$K_{\sigma_I,\beta,R}(x_1,x_2) = A_f \exp\left(-\frac{x_1^2+x_2^2}{4\sigma_I^2}\right) \exp\left(-\frac{(x_1-x_2)^2}{2(\beta\sigma_I)^2}\right) \exp\left(-\frac{jk}{2R}(x_1^2-x_2^2)\right). \tag{4.9}$$

We note the following scaling property for the $R = \infty$ case:

$$K_{\sigma_I', \beta, \infty}(-x_1, x_2) = K_{\sigma_I, \beta, \infty}\left(-x_1 \frac{\sigma_I}{\sigma_I'},\, x_2 \frac{\sigma_I}{\sigma_I'}\right), \tag{4.10}$$

which expresses the fact that the covariance function associated with a given σ_I can be found by scaling that associated with another σ_I'. The error expression depends on the joint distribution of the samples $\mathbf{s}$ and the field $f(x)$, which in turn is determined through the covariance functions. Considering the representation of $f(x)$ in terms of its samples, we also note that for a given set of σ_{m_i}, the cost associated with a set of sampling points remains unchanged if the sampling points are scaled by σ_I'/σ_I. Hence we conclude that for the case $R = \infty$ and $\mathcal{L}$ is the identity, the error does not depend on σ_I, provided the sampling points are scaled appropriately. As a result, the cost-error trade-off curves will be the same for different values of σ_I, and the optimum sampling strategies will be scaled versions of each other.

Here we show that the conclusion of the preceding paragraph continues to remain valid even when $R \neq \infty$. We will first show that for a given set of sampling points $\xi_1, \dots, \xi_M$, and a given covariance matrix K_m, the associated costs and the error for all values of R are the same. This, in fact, stems from the fact that the curvature term corresponds to uncorrelated phase terms. Let the covariance function associated with $f(x)$ be $K_{\sigma_I, \beta, \infty}(x_1, x_2)$. Let $\bar{f}(x)$ be the zero-mean complex proper field with the covariance function

$$\begin{aligned} E[\bar{f}(x)\bar{f}^*(x)] &= K_{\sigma_I, \beta, R}(x_1, x_2) && (4.11)\\ &= K_{\sigma_I, \beta, \infty}(x_1, x_2) \exp\left(-\frac{jk}{2R}(x_1^2 - x_2^2)\right) && (4.12)\\ &= E[f(x)f^*(x)] \exp\left(-\frac{jk}{2R}(x_1^2 - x_2^2)\right). && (4.13) \end{aligned}$$

We first observe that the presence of a curvature does not affect the cost associated with a sample. The cost associated with the ith sample $\bar{s}_i = \bar{f}(\xi_i) + m_i$ is given by $C_{\bar{s}_i} = \log_2(\sigma_{\bar{s}_i}^2/\sigma_{m_i}^2)$, where $\sigma_{\bar{s}_i}^2 = E[|\bar{s}_i|^2] = E[|\bar{f}(\xi_i)|^2] + E[|m_i|^2] = E[|f(\xi_i)|^2] + E[|m_i|^2]$. Hence the cost of a sample with a given $E[|m_i|^2] = \sigma_{m_i}^2$ does not depend on the value of R.

We now show that the error does not depend on the value of R; that is, for a given set of sampling locations and a given set of σ_{m_i}, the errors associated with estimating $f(x)$ and $\bar{f}(x)$ are the same. Let us define the vector $\mathbf{g}$ as $\mathbf{g} = [f(\xi_1), \ldots, f(\xi_M)]^{\mathrm{T}}$, $i = 1, \ldots, M$. Now, the vector of finite accuracy samples of $f(x)$ is given by $\mathbf{s} = \mathbf{g} + \mathbf{m}$, where $\mathbf{m} = [m_1, \ldots, m_M]^{\mathrm{T}}$. Let the $M \times M$ covariance matrix of the finite accuracy samples be denoted by $E[\mathbf{s}\mathbf{s}^\dagger] = K_{\mathbf{s}} = K_{\mathbf{g}} + K_{\mathbf{m}}$, where the element in the ith row and lth column of $K_{\mathbf{g}}$ is given by $K_{\sigma_I, \beta, \infty}(\xi_i, \xi_l)$, $i, l = 1, \ldots, M$. The cross covariance between $f(x)$ and $\mathbf{s}$ is given by the $1 \times M$ row vector $E[f(x)\mathbf{s}^\dagger] = \mathbf{d}(x)$, where the lth element is given by $K_{\sigma_I, \beta, \infty}(x, \xi_l)$. Similarly, we define $\bar{\mathbf{s}} = \bar{\mathbf{g}} + \mathbf{m}$, where $\bar{\mathbf{g}} = [\bar{f}(\xi_1), \ldots, \bar{f}(\xi_M)]^{\mathrm{T}}$. Consequently, we have $K_{\bar{\mathbf{s}}} = K_{\bar{\mathbf{g}}} + K_{\mathbf{m}}$, where the element in the ith row and lth column is given by $K_{\sigma_I, \beta, R}(\xi_i, \xi_l)$, and $E[\bar{f}(x)\bar{\mathbf{s}}^\dagger] = \bar{\mathbf{d}}(x)$, where the lth element is given by $K_{\sigma_I, \beta, R}(x, \xi_l)$. Now, let $T = \mathrm{diag}(t_i)$, $t_i = \exp(-(jk/2R)\xi_i^2))$, $i = 1, \ldots, M$. We observe that

$$\begin{aligned} K_{\bar{\mathbf{s}}} &= K_{\bar{\mathbf{g}}} + K_{\mathbf{m}} && (4.14)\\ &= TK_{\mathbf{g}}T^\dagger + TK_{\mathbf{m}}T^\dagger && (4.15)\\ &= TK_{\mathbf{s}}T^\dagger, && (4.16)\end{aligned}$$

where (4.15) follows from the fact that $TK_{\mathbf{m}}T^\dagger = \mathrm{diag}(t_i)\,\mathrm{diag}(\sigma_{m_i}^2)\,\mathrm{diag}(t_i^*) = \mathrm{diag}(\sigma_{m_i}^2) = K_{\mathbf{m}}$, since $|t_i| = |\exp(-(jk/2R)\xi_i^2))| = 1$. We also observe that

$$\bar{\mathbf{d}}(x) = \exp(-(jk/2R)x^2))\mathbf{d}(x)T^\dagger. \qquad (4.17)$$

Now, using these results, we finally show that the error is independent of the value of R. We consider the error for the field at a given point x. Denoting the MMSE estimate of $\bar{f}(x)$ given $\bar{\mathbf{s}}$ as $\hat{\bar{f}}(x|\bar{\mathbf{s}})$, the associated MMSE can be expressed as [133, Ch. 6]

$$\begin{aligned} E[||\bar{f}(x) - \hat{\bar{f}}(x|\bar{\mathbf{s}})||^2] &= K_{\sigma_I, \beta, R}(x, x) - \bar{\mathbf{d}}(x)K_{\bar{\mathbf{s}}\bar{\mathbf{s}}}^{-1}\bar{\mathbf{d}}(x)^\dagger && (4.18)\\ &= K_{\sigma_I, \beta, \infty}(x, x) - \mathbf{d}(x)K_{\mathbf{s}\mathbf{s}}^{-1}\mathbf{d}(x)^\dagger && (4.19)\\ &= E[||f(x) - \hat{f}(x|\mathbf{s})||^2] && (4.20)\end{aligned}$$

In obtaining (4.19), we used (4.16), (4.17), $TT^\dagger = I$, and $|\exp(-(jk/2R)\xi_i^2))| = 1$, where I is the $M \times M$ identity matrix. Hence we have shown that the value of R does not change the error.

So far we have shown that (i) for $R = \infty$, the error does not depend on σ_I, provided the sampling points are appropriately scaled; (ii) for a given set of sampling points and σ_{m_i}, the associated errors and costs do not depend on R. Thus we conclude that for a given GSM field with a specified value of β, the cost-error trade-off curves associated with the problem of estimating a field based on its own samples do not depend on σ_I and R. Now, recall that GSM fields remain GSM fields with the same β, but different σ_I and R after passing through first-order optical systems [136, 137, 146]. This, combined with the previous observations, show that the error associated with estimating the output field by sampling the output, is the same as the error associated with estimating the input field by sampling the input (under the same cost).

Finally, we consider the problem of sampling the output of a first-order optical system in order to estimate the input field. We first recall that the MMSE is invariant under unitary transformations; that is, the MMSE associated with estimating $f(x)$ based on a random vector s is the same as the MMSE associated with estimating $\mathcal{L}\{f(x)\}$, if $\mathcal{L}$ is a unitary transformation. We also recall that optical systems represented by real A, B, C, D parameters are unitary systems [173, Ch.9]. Hence for any such system, the MMSE associated with estimating the input of the optical system and the output of the optical system based on a given set of samples of the output are the same. Thus, combining this with the observations of the previous paragraph, we conclude that the error versus cost trade-offs for the estimation of the input of an optical system based on the samples of the input field are the same as those based on the samples of the output field. (The same conclusion also holds for estimating the output based on the samples of the input or the output.) In other words, finite-accuracy samples of the output field are as good as finite-accuracy samples of the input field for the broad class of first-order optical systems.

We now discuss the effect of noise $n(x)$. Our system noise model is characterized by the following covariance function: $K_n(x_1 - x_2) = A_n \exp(-(x_1 - x_2)^2/2\sigma_{\nu,n}^2)$ with $\sigma_{\nu,n} = \beta_n \sigma_I$, $\beta_n < \beta$. Here we will show that, as in the noiseless case, when the system $\mathcal{L}$ is identity and $R = \infty$, the error value does not depend on σ_I, provided the sampling points are scaled appropriately.

To show this in the noisy case, we need to show that the associated covariance functions can be scaled with σ_I. (i) The scaling property of $K_{f(x)}$ was already discussed at the beginning of this subsection. (ii) The noise covariance function also scales with σ_I, in a manner similar to (4.10). It follows from (i) and (ii) that the covariance of the observations also scales with σ_I. We also note that, due to statistical independence of $f(x)$ and $n(x)$, the cross covariance of $f(x)$ and $\mathbf{s}$ only depends on the covariance function of $f(x)$, which is known to scale with σ_I. Hence all associated covariances have the scaling property. Thus we can now conclude that the error for a given set of sampling points for a given σ_I, can be found by looking at the error for another σ_I at a scaled set of sampling points. We also note that for a given set of σ_{m_i}, the cost associated with a set of sampling points, remains unchanged under appropriate scaling. This implies that the trade-off curves are invariant for different σ_I values and the optimum sampling points can be found by scaling.

4.3 Trade-offs between Error and Cost Budget

In this section, we present trade-off curves between the error and the cost budget, and the optimum sampling parameters achieving these curves.

Based on the discussion of Section 4.2, we note that in the noiseless case (SNR $= \infty$), the presented cost-error trafe-off curves are valid for any $ABCD$ system with real parameters, $AD - BC = 1$. The optimum sampling points are easily found by scaling in proportion to the ratio of input and output σ_Is. When SNR $\neq \infty$, the curves are obtained for the case of $\mathcal{L}$ is the identity operator and $R = \infty$, and these do not generalize to arbitrary $ABCD$ systems. But the optimum sampling points for one value of σ_I can still be found from those for another by scaling.

To compute the error expressions and optimize over the parameters of the representation strategy, we discretize the x space with the spacing Δ_c as explained in Section 2.3, where more details can be found. In order to find the optimum sampling interval, we use a brute force method, where for a given C_{B} we calculate the error for varying Δ_x and M, and choose the values providing the least error. We note that the optimization variable Δ_x and the discretization variable Δ_c are not the same. Δ_x is the sampling interval whose optimal value we seek, whereas Δ_c is the discrete grid spacing we employ in the numerical experiments.

In our numerical experiments, we use two different β values: $\beta = 1/8$ and $\beta = 1$. We choose $\beta_n = 1/32$. We consider different noise levels parameterized through the signal-to-noise ratio, defined as the ratio of the peak signal and noise levels: $\mathrm{SNR} = A_f/A_n$. We consider the values $\mathrm{SNR} = 1,\ 10,\ \infty$ to cover a wide range of problem instances. For simplicity in presentation, in our simulations we focus on Δ_x and set the less interesting $x_0 = 0$. We choose the interval D equal to $[x_L, x_H] = [-5\sigma_I, +5\sigma_I]$ to ensure that the signal values are safely negligible outside D. We report the error as a percentage defined as $100\,\varepsilon(C_{\mathrm{B}})/\varepsilon_0$ where $\varepsilon_0 = \int_{-\infty}^{\infty} K_f(x,x)dx = A_f\sqrt{2\pi}$.

We would like to note that error-cost trade-off curves do not depend on the total energy of the signal. More precisely, when there is no system noise $n(x)$, the error-cost curves are independent of the constant A_f in (4.9). When there is system noise $n(x)$, the error-cost curves do not depend on the individual values of A_f and A_n, but only on the ratio $\mathrm{SNR} = A_f/A_n$. These are due to the fact that the error is reported as a percentage error which scales with A_f, and the cost is independent of the ranges of the signal values, but only depends on the number of distinguishable levels, that is when A_f changes, m_i's can be scaled without changing the cost.

Fig. 4.1 and Fig. 4.2 present the error vs. bit budget curves for varying SNR for a relatively incoherent field ($\beta = 1/8$) and for a relatively coherent field ($\beta = 1$), respectively. As expected, the error decreases with increasing cost budget in all cases. We note that $\varepsilon(C_{\mathrm{B}})$ is very sensitive to increases in C_{B} for smaller C_{B}. Then it becomes less responsive and eventually saturates.

We observe that in each of these figures, as the noise level becomes higher, it becomes more difficult to obtain low values of error. We observe that for both values of β, when there is no system noise ($\mathrm{SNR} = \infty$), the error goes to zero as

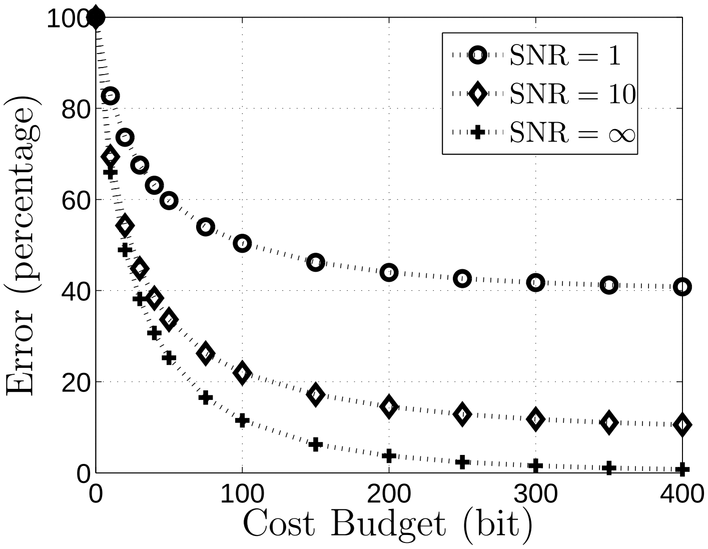


Figure 4.1: Error vs. cost budget, $\beta = 1/8$, SNR variable.

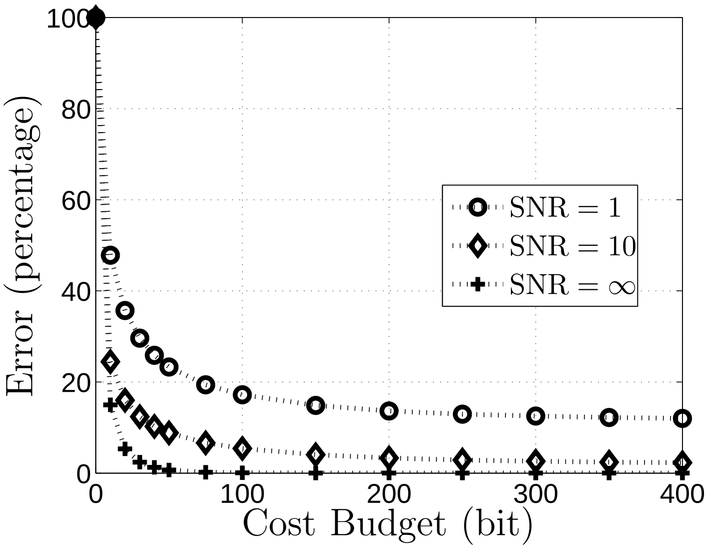


Figure 4.2: Error vs. cost budget, $\beta = 1$, SNR variable.

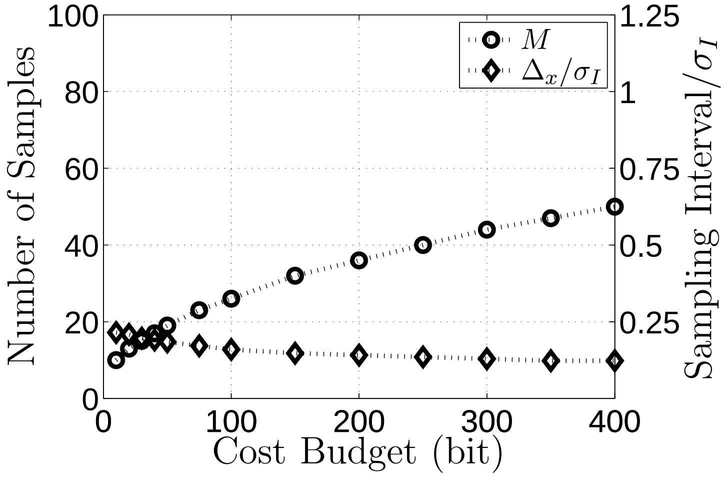


Figure 4.3: Number of samples and optimum sampling interval vs. cost budget, $\beta = 1/8$, $\mathrm{SNR} = \infty$.

we increase the cost. This means that, no matter how small the error tolerance $\varepsilon > 0$ is specified to be, the continuous finite-energy field can be represented with a finite number of bits. This observation is discussed in more detail in Section 5.4.

Comparing these figures, we observe that for the relatively incoherent case (Fig. 4.1), it is more difficult to achieve low values of error for a given bit budget. But as the field becomes more coherent (Fig. 4.2), the field values at different locations become more correlated with each other, the total uncertainty in the field decreases, and it becomes a lot easier to achieve lower values of error.

We now investigate the relationship between the optimum sampling strategies and the problem parameters C_{B}, SNR, and β. The optimum sampling interval Δ_x and the optimum number of samples M that achieve the errors given in Fig. 4.1 are presented in Figs. 4.3 and 4.4 for $\mathrm{SNR} = \infty$ and $\mathrm{SNR} = 1$. The optimum sampling interval Δ_x and the optimum number of samples M that achieve the errors given in Fig. 4.2 are presented in Figs. 4.5 and 4.6 for $\mathrm{SNR} = \infty$ and $\mathrm{SNR} = 1$.

When there is no system noise $n(x)$, the optimum sampling strategies can be informally interpreted in the light of the competition between the following driving forces: i) to have as many effectively uncorrelated samples as possible, ii)

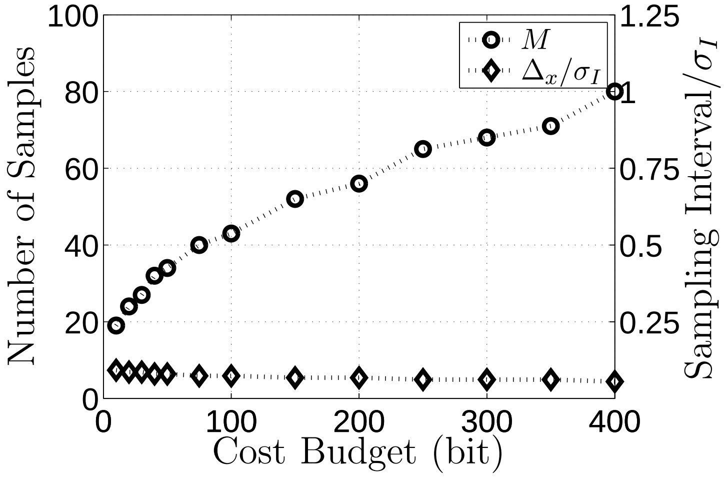


Figure 4.4: Number of samples and optimum sampling interval vs. cost budget, $\beta = 1/8$, SNR $= 1$.

to have samples whose variances are as high as possible, and iii) to have samples which are as highly accurate as possible. When there is system noise $n(x)$, each sample tells less about the value of the field. In order to wash out the effect of noise, one is often willing to take samples at field locations which are considerably correlated, and which one would probably not take samples at, had there been no noise.

We observe that in all cases, in general, as C_{B} increases, the optimum sampling interval decreases and the number of samples increases: when we have more bits to spend, we use a higher number of more closely spaced samples. When C_{B} is low, the optimal strategy is to use a low number of more distantly-spaced samples so that each sample has a reasonable accuracy and each of them provides effectively new information about the field. As the allowed cost increases, we can afford more samples with high enough accuracies and we prefer to use more closely-spaced samples so that we can get more information about field values we previously had to neglect when the allowed cost was lower.

Comparing Fig. 4.3 and Fig. 4.4 (or Fig. 4.5 and Fig. 4.6), we observe that as the noise level increases, the samples should be taken more closer (the sampling interval decreases). When a sample is noisy, one would expect the information provided by that sample to be smaller, encouraging us to take more closely spaced

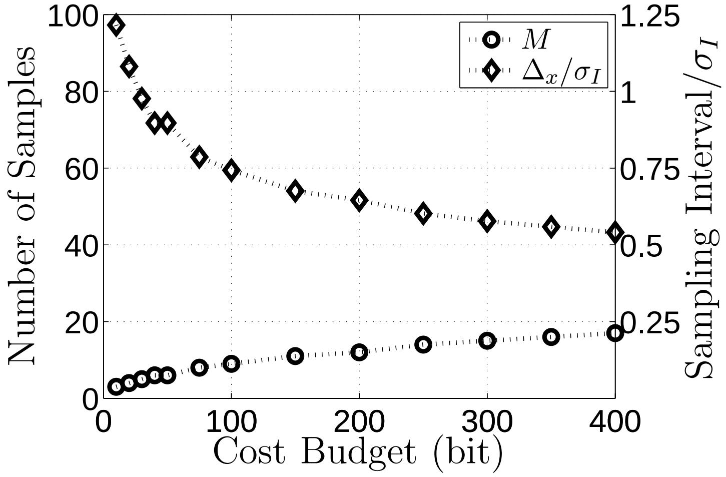


Figure 4.5: Number of samples and optimum sampling interval vs. cost budget, $\beta = 1$, SNR $= \infty$.

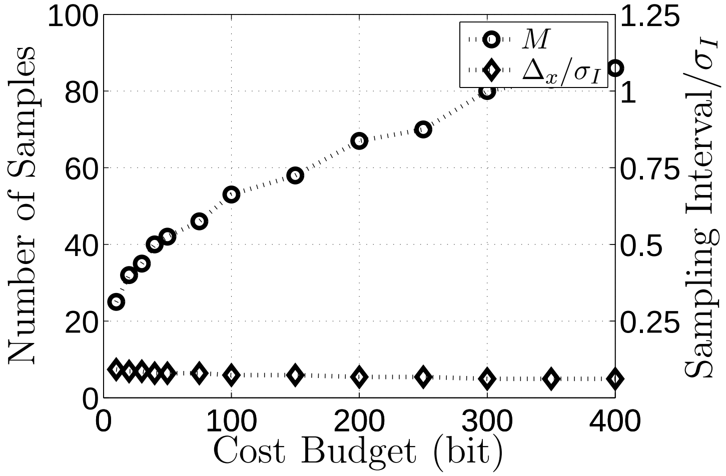


Figure 4.6: Number of samples and optimum sampling interval vs. cost budget, $\beta = 1$, SNR $= 1$.

samples so as to compensate for the effects of noise. We also observe that as the noise level increases, one should take a higher number of samples M. This observation may seem trivial, since decreasing the sampling interval automatically increases the number of samples within a certain spatial range. However, we note that here the range over which samples are taken does not remain constant but also decreases. (The variances of field values decrease as we move away from the $x = 0$ point, so that the field here is highly contaminated by noise. Since samples taken here are of little value for representing the field, it is reasonable to expect that it will be better not to take these samples, thereby decreasing the spatial range the samples are taken over.) However, the decrease in the spatial range is not as much as to compensate the decrease in the sampling interval, so in the end the number of samples taken increases.

Comparing Fig. 4.3 and Fig. 4.5, we see that when the field is more coherent, it is desirable to take a fewer number of samples which are farther apart. When the field is more coherent, under the GSM correlation structure, the field value at each point becomes more correlated with field values farther away. Hence there is a tendency to space the samples well in order to get effectively new information from each sample. Also, the variances of the field values decrease as we move further away from the $x = 0$ point, so we prefer not to waste any of our bit budget on such samples. As a result, the optimum number of samples is smaller, which is consistent with the fact that more coherent fields have a lower number of effective modes (the number of uncorrelated random variables required to effectively represent the field).

4.4 Comparison with Shannon-Nyquist Sampling Based Approaches

A common approach in sampling signals is to use the Shannon-Nyquist sampling theorem as a guideline. As outlined at the beginning of this chapter, in this traditional approach, one determines an effective frequency extent B, and

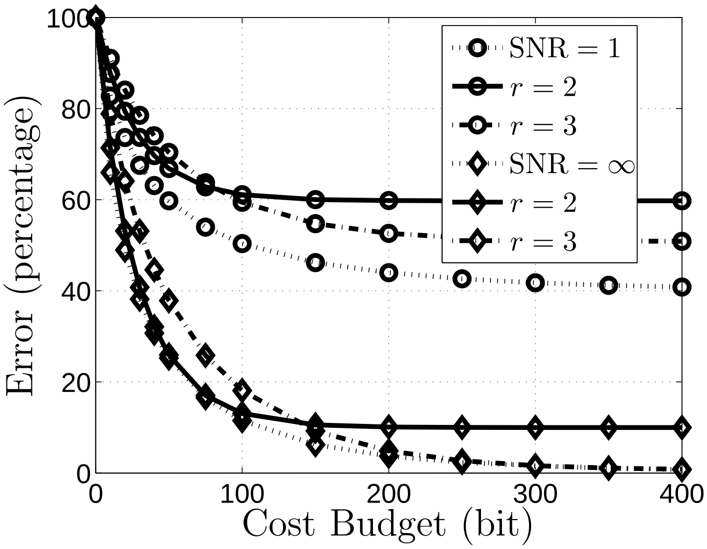


Figure 4.7: Error vs. cost budget, $\beta = 1/8$, SNR variable. The dotted lines are for optimal sampling strategies and the corresponding dashed and solid lines are for sampling theorem based strategies.

an effective spatial extent L which are used to determine the sampling interval, and the spatial extent the samples will be taken over, respectively. Here we will compare the error vs. cost budget curve that is obtained following this traditional approach with the optimal curves obtained with our approach and shown in Figures 4.1 and 4.2. But first we review how the traditional Shannon-Nyquist approach applies to random fields. A fundamental result in this area states that the Shannon-Nyquist sampling theorem can be generalized to wide-sense stationary (WSS) signals: A band-limited WSS signal can be reconstructed in the mean-square sense from its equally-spaced samples taken at the Nyquist rate [115]. [10,126] further generalizes this result to non-stationary random fields: Let $v(x) \in \mathbb{R}$ be a finite-energy random field. Let us consider the covariance function of the Fourier transform of the field defined as $S_v(\nu_1, \nu_2) = E[V(\nu_1)V^*(\nu_2))]$, where $V(\nu)$ is the Fourier transform of $v(x)$. If $S_v(\nu, \nu) = 0$, for $|\nu| > B/2$, then the field can be recovered from its samples in the mean-square sense; that is, $E[||v(x) - \sum_{k=-\infty}^{\infty} v(k/B)\,\mathrm{sinc}(x\,B - k)||^2] = 0$.

We now explicitly work out the conventional sampling approach for GSM

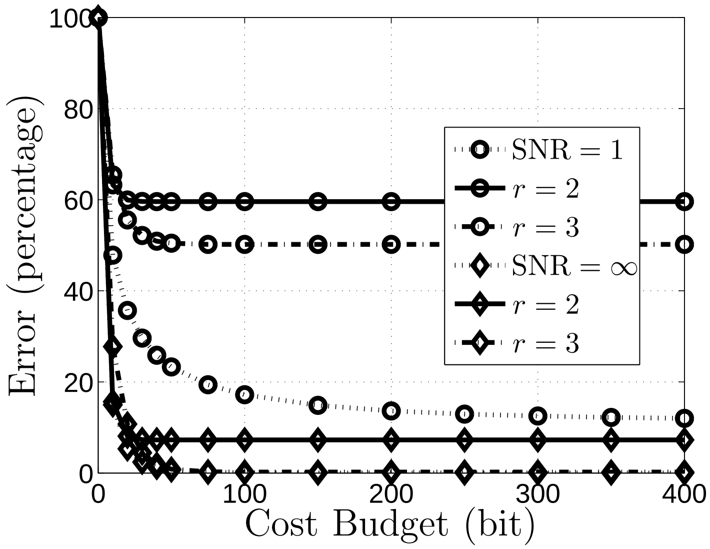


Figure 4.8: Error vs. cost budget, $\beta = 1$, SNR variable. The dotted lines are for optimal sampling strategies and the corresponding dashed and solid lines are for sampling theorem based strategies.

fields. The effective spatial extent of the field will be determined by looking at the intensity distribution $K_f(x, x) = \exp(-x^2/2\sigma_I^2)$, which has a Gaussian profile with standart deviation σ_I. Most of the energy of a Gaussian lies within a few standard deviations so that the effective spatial extent can be taken as $[-r\,\sigma_I, r\,\sigma_I]$; we choose $r = 3$. The intensity of the Fourier transform of the field; that is, the diagonal of the covariance function of the Fourier transform of the field also has a Gaussian profile $S_f(\nu, \nu) \propto \exp(-f^2/2\sigma_{I,F}^2)$, where $S_f(\nu_1, \nu_2) = E[F(\nu_1)F^*(\nu_2)]$, where $F(\nu)$ is the Fourier transform of $f(x)$, and $\sigma_{I,F} = \frac{1}{2\pi}\sqrt{\frac{1}{\beta^2} + \frac{1}{4}}/\sigma_I$ (see, for instance [174]). We take the effective frequency extent as $[-r\,\sigma_{I,F}, r\,\sigma_{I,F}]$, again with $r = 3$. This implies a sampling interval of $1/(2r\sigma_{I,F})$. The number of samples is found by dividing the effective spatial extent to the sampling interval

$$M_s = \frac{2r\sigma_I}{1/(2r\sigma_{I,F})} = \frac{2r^2}{\pi}\left(\frac{1}{\beta^2} + \frac{1}{4}\right)^{0.5}. \tag{4.21}$$

Hence, for each cost budget value C_{B}, the cost associated with each sample will be C_{B}/M_s. To ensure a fair comparison with our approach, we again use the mean-square estimate to estimate the signal from the Nyquist samples.

We now compare the error vs. bit budget trade-offs obtained with the approach presented in this chapter, with those obtained by using the traditional approach described above. We use two different r values; $r = 2$, and $r = 3$. Fig. 4.7 and Fig. 4.8 compare these trade-off curves for $\beta = 1/8$ and $\beta = 1$, respectively. The dotted curves and the dashed/solid lines show the results for the optimal sampling scheme and the sampling theorem based schemes respectively. As expected, for all cases, the optimum sampling strategy gives better trade-offs compared to the sampling strategies based on the sampling theorem.

We note that when there is no system noise $n(x)$, and if we determine the effective extents appropriately, we would expect to obtain error values close to zero for high values of cost budget. We observe that this is indeed the case for $r = 3$, but not for $r = 2$. This suggests that $r = 2$ is a poor choice for defining the effective extents, and illustrates the importance of determining effective extents appropriately.

When $r = 3$ and there is no system noise, for both relatively low and high degrees of coherence, the optimal strategy and the traditional strategy differ by a greater amount for low values of cost budget. This observation may be interpreted as follows: When the cost budget is low, the relatively high number of samples dictated by the sampling theorem will result in the samples being relatively inaccurate, leading to poor performance. (As we have seen earlier, for low cost values, it is better to use a smaller number of samples with relatively better accuracy.) As the cost budget increases, the difference between the two approaches gets smaller, and both strategies achieve error values very close to 0, as expected. For low values of cost budget, the traditional approach with $r = 2$ dictates a sampling strategy closer to the optimal one, compared to $r = 3$, and gives error values closer to the optimal strategy. Yet, as observed above, it gives relatively poor error values for higher values of cost budget, and therefore cannot be considered a good sampling approach for all values of the cost budget.

When the system noise level is high, the difference between the optimal and traditional strategies is pronounced for almost all values of cost budget. The sampling theorem assumes that the samples will be noiseless, and therefore cannot exploit the opportunity for noise elimination through oversampling. (We observe that the traditional strategy with $r = 2$ gives poorer results compared to the $r = 3$ case, which may be attributed to the relatively low number of samples dictated by the former.) We observe that the performance difference between the traditional approaches and the optimal strategy is more pronounced for the coherent case. When the field is more coherent, the sampling theorem based strategy dictates the use of a fewer number of more distantly spaced samples, compared to the incoherent case. However, in the presence of noise, the optimal strategy is not that much different for the incoherent and coherent cases, and dictates that we use a comparably larger number of more closely spaced samples even when the field is coherent. Therefore, the traditional sampling strategies are more markedly inferior than the optimum strategy in the coherent case.

4.5 Conclusions

We focused on various trade-offs in the representation of random fields, mainly: i) the trade-offs between the achievable error and the cost budget, ii) the trade-offs between the accuracy, spacing, and number of samples. We have derived the optimal bounds for simultaneously achievable bit cost and error and obtained the optimal sampling parameters necessary to achieve them. These performance bounds are not only of interest for better understanding of information relationships inherent in propagating wave-fields, but can also lead to guidelines in practical scenarios. We also investigated how these results are affected by the degree of coherence of the field and the noise level. Furthermore, we observed how the optimal sampling parameters change with increasing cost budget.

We also considered the case where the signal is represented by samples taken after the signal passes through a linear system. For the case of Gaussian-Schell model beams, when there is no noise, we have shown that finite-accuracy samples of the output field are as good as finite-accuracy samples of the input field, for the broad class of first-order optical systems. The cost-error trade-off curves obtained turn out to be the same as those obtained for direct sampling of the input, and the optimum sampling points can be found by a simple scaling of the direct sampling results.

Chapter 5

Joint Optimization of Number of Samples, Sample Locations and Measurement Accuracies: Non-uniform Case

In this chapter we will again consider representation of a non-stationary field with a finite number of bits. In Chapter 4, we have focused on the case where the samples are equidistantly spaced, and each sample is taken with the same accuracy. In this chapter, we consider the case where the sample locations can be freely chosen, and need not to be equally spaced from each other. Furthermore, the measurement accuracy of each sample can very from sample to sample. This formulation presents a challenging optimization problem: To solve this problem, one has to find the optimum number of samples, the locations of these samples which take values in a continuum, and the costs associated with each of these samples. Thus this general non-uniform case represents maximum flexibilty in choosing the sampling strategy allowing tighter optimization of error-cost curve.

We now present an overview of this chapter. In Section 5.1, we formulate and discuss the non-uniform sampling problem. In Section 5.2, we present the optimum sampling strategies and the trade-off curves between the cost and the error. We compare the trade-off curves obtained with the non-uniform approach of this chapter with the ones that would be obtained if the uniform scheme of the previous chapter was used in Section 5.3. In Section 5.4, we provide a general discussion on the representation of random fields using finite numbers of bits. We conclude in Section 5.5

5.1 Problem Formulation

In this section, we will formulate and discuss the non-uniform sampling problem described above. The signal and measurement models are the same with those of Section 4.1 of Chapter 4.

We now formulate the problem we will be considering in this chapter; the problem of optimal non-uniform sampling with possibly non-uniform cost allocation. Here, the sampling locations $x = \xi_1, \ldots, \xi_M \in D$ are free. The accuracy (hence the cost) associated with each sample can be different; that is C_{s_i} can have possibly different values. The total cost of the representation is given by $C_{\mathrm{T}} = \sum_{i=1}^{M} C_{s_i}$. For a given C_{B}, our objective is to choose the number of the samples M, the locations ξ_i, and the costs C_{s_i} while satisfying $C_{\mathrm{T}} \leq C_{\mathrm{B}}$, with the objective of minimizing the minimum mean-square error between $f(x)$ and $\hat{f}(x \mid \mathbf{s})$.

Let $\boldsymbol{\xi}^M = [\xi_1, \ldots, \xi_M]^{\mathrm{T}}$ denote the vector of sampling locations. Let $\mathbf{C_s}^M = [C_{s_1}, \ldots, C_{s_M}]$ denote the vector of cost allocations C_{s_i}. The above problem can be stated as one of minimizing

$$E\left[\int_D \|f(x) - \hat{f}(x \mid \mathbf{s})\|^2 dx\right] \tag{5.1}$$

over M, $\boldsymbol{\xi}^M$ and $\mathbf{C_s}^M$ subject to

$$C_{\mathrm{T}} = \sum_{i=1}^{M} C_{s_i} \leq C_{\mathrm{B}}. \tag{5.2}$$

We observe that the optimization space for the sampling locations is very large: we are seeking the best sampling locations over a continuous region of space D. To overcome this difficulty, we will consider a discretization of the optimization region D, where instead of the condition $\xi_i \in D$, we will be considering the condition $\xi_i \in \bar{D}$, where $\bar{D} = \{x_1, \dots, x_N\} \in D$ is a set of N finely chosen equally spaced finite number of points inside D. The number of points N must be chosen large enough to ensure satisfactory optimization: the minimum interval between the points in $\bar{D}$ is taken to be sufficiently smaller than the sampling intervals for the signal $f(x)$ and the noise $n(x)$ dictated by the sampling theorem. We also note the following property that is related to the cost of a measurement: measuring a point with repeated measurements is suboptimal (better errror values are obtained if one high quality measurement is made instead), so including more points in between two adjacent points in $\bar{D}$ does not provide an opportunity for better optimization if the minimum interval between the points in $\bar{D}$ is sufficiently small. Hence we will consider the new constraint $\xi_i \in \bar{D}$ instead of the constraint $\xi_i \in D$ where $i = 1, \dots, M$, and $M \leq N$. We will refer to this optimization problem as Problem P.

We now note that even after this discretization, solution of the optimization problem remains challenging: In most cases, the number of points in $\bar{D}$, will be very large. For instance, optimization with brute force methods will typically require the following steps to be followed: all values of M lying between 0 N will be considered; for each one of these values of M, one will try all possible combinations of the sampling locations in $\bar{D}$ space, that is $\boldsymbol{\xi}^M \subset \bar{D}$; and for each such set of sampling locations, one will optimize over the cost allocations. We note that even this last optimization which optimizes over the cost allocation for a particular fixed set of sampling locations is also in itself a difficult optimization task to be done using brute force methods.

We will now argue that this difficult optimization problem can be, in fact, solved by solving another equivalent optimization problem. We consider the following optimization problem where the aim is to minimize

$$E\left[\int_D \|f(x) - \hat{f}(x \mid \mathbf{s})\|^2 dx\right], \tag{5.3}$$

over ${\mathbf{C_s}}^N = [C_{s_1}, \dots, C_{s_N}]$ such that

$$C_{\mathrm{T}} = \sum_{i=1}^{N} C_{s_i} \le C_{\mathrm{B}} \tag{5.4}$$

where the cost allocation is done over all points in $\bar{D}$, that is $\boldsymbol{\xi}^N = \{x_1, \dots, x_N\}$. We will denote this optimization problem as Problem $\bar{P}$. We first observe that both of these problems have the same objective function, and the cost budget constraint is in the same form. So what we need to show is that the spaces defined by the optimization variables, i.e. the optimization spaces, are the same. These are defined by the number of samples M, the M sampling locations $\boldsymbol{\xi}^M$, and the cost allocation over M sampling points ${\mathbf{C_s}}^M$ in Problem P; and the cost allocation ${\mathbf{C_s}}^N$ over N points in Problem $\bar{P}$. Although at first sight these descriptions seem to refer to different optimization spaces, in fact these spaces are the same. The crucial point here is to observe that the optimization space of Problem $\bar{P}$ includes points where some of the measurements are not made. For instance let the sample at x_i be not measured. Then this point will be described with $\sigma^2_{m_i} = \infty$ and $C_{s_i} = 0$. Hence, in general, any particular point in the optimization space of the first problem described by M, $\boldsymbol{\xi}^M$, and ${\mathbf{C_s}}^M$ can be equivalently described by an appropriate cost vector ${\mathbf{C_s}}^N$. In this longer cost allocation vector ${\mathbf{C_s}}^N$, the individual costs C_{s_i} associated with $\boldsymbol{\xi}^M$ will possibly have $C_{s_i} > 0$, and the other samples will necessarily have $C_{s_i} = 0$. Hence any point in the optimization space of Problem P is also in the optimization space of Problem $\bar{P}$. Similarly, any point in the optimization of Problem $\bar{P}$ is also in the optimization space of Problem P. Thus, we can conclude that solving Problem $\bar{P}$ is sufficient for the purpose of solving Problem P. This means the following: i) the optimum achieved by Problem P cannot be lower than the optimum achieved by Problem $\bar{P}$. ii) any optimum achieved by Problem $\bar{P}$ can also be achieved by Problem P. Both of these assertions are consequences of the fact that the optimization spaces for these two problems are the same.

By putting the problem in the form in Problem $\bar{P}$, we now have the chance to use the numerical approach suggested in Section 3.4. With such a numeric approach at hand, it is now possible to exploit the chance of better optimization offered by non-uniform sampling locations and non-uniform cost allocations.

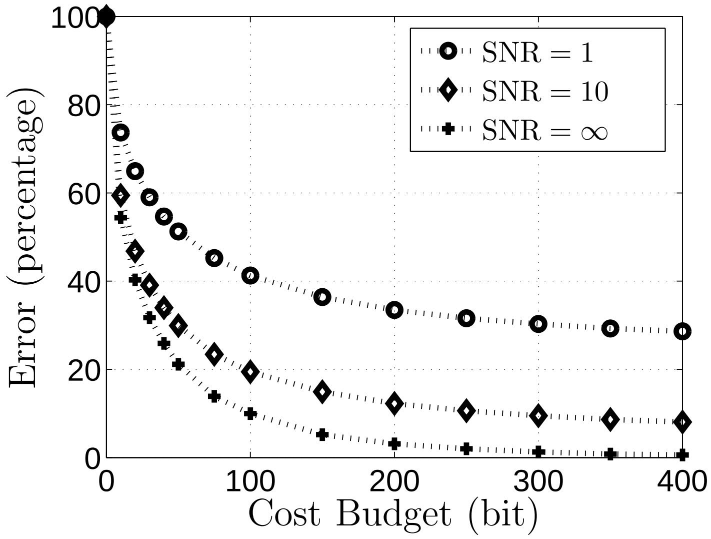


Figure 5.1: Error vs. cost budget, SNR variable.

In our numerical experiments, in evaluating the integrals in the error expressions and solving for the linear estimators, we will use a simple discretization of the space D with Δ_c intervals, which is explained in detail in Section 2.3. We note that the discretization of D into $\bar{D}$ for forming Problem P and this discretization with Δ_c are conceptually different, and they do not necessarily have to be the same. The first one discretizes the optimization space to make the optimization problem tractable and the second one offers a numerical method to take the integrals and solve the estimators. Even if we were to use some other method to evaluate the integrals, we would still want to discretize the optimization space to construct Problem P. Nevertheless, for simplicity, we use the same discretization of D for both of these purposes.

5.2 Trade-offs between Error and Cost Budget

In this section, we present trade-off curves between the error and the cost budget, and the optimum number of samples, sampling locations and measurement accuracy levels achieving these curves.

Based on the discussion of Section 4.2, we recall that in the noiseless case

(SNR $= \infty$), the presented cost-error trafe-off curves are valid for any $ABCD$ system with real parameters, $AD - BC = 1$. The optimum sampling points are easily found by scaling in proportion to the ratio of input and output σ_Is. As discussed in Section 4.3, the error-cost trade-off curves do not depend on the total energy of the signal.

In our experiments, we consider a multiple beam scenario where two statistically independent GSM beams with the same σ_I and R, but different β parameters reside side by side. More precisely, the unknown field has the following covariance function

$$K_f(x_1, x_2) = K_{\beta_a,\sigma_I,R}(x_1 - x_a, x_2 - x_a) + K_{\beta_b,\sigma_I,R}(x_1 - x_b, x_2 - x_b) \qquad (5.5)$$

We choose $-x_a = x_b = 3\sigma_I$, and $\beta_a = 1/8$ and $\beta_b = 1$. We choose $\beta_n = 1/32$. We consider different noise levels parameterized through the signal-to-noise ratio, defined as the ratio of the peak signal and noise levels: SNR $= A_f/A_n$. We consider the values SNR $= 1$, 10, ∞.

We choose the interval D equal to $[x_L, x_H] = [-6\sigma_I, +6\sigma_I]$. We report the error as a percentage defined as $100\,\varepsilon(C_{\mathrm{B}})/\varepsilon_0$ where $\varepsilon_0 = \int_{-\infty}^{\infty} K_f(x,x)dx = 2A_f\sqrt{2\pi}$.

Fig. 5.1 present the error versus bit budget curves for varying SNR. As expected, the error decreases with increasing cost budget in all cases. We note that $\varepsilon(C_{\mathrm{B}})$ is very sensitive to increases in C_{B} for smaller C_{B}. Then it becomes less responsive and eventually saturates. We observe that as the noise level becomes higher, it becomes more difficult to obtain low values of error. We will later further discuss these trade-off curves while comparing them with the ones that would be obtained if the equidistant sampling strategy with uniform cost allocation were used as the sampling strategy.

We now review the optimum measurement strategies, i.e. the number of samples, sampling locations and measurement accuracy levels achieving these error-cost curves. The measurement accuracy levels (i.e. the cost allocations) that achieve the error-cost values given in Fig. 5.1 are presented in Figs. 5.2 and 5.3 for SNR $= \infty$ and SNR $= 1$ for varying cost budget C_{B}. These values are

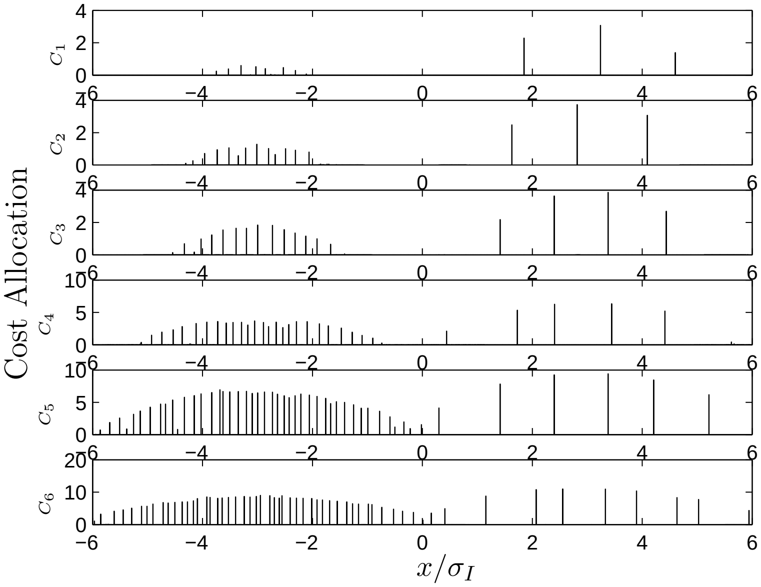


Figure 5.2: Cost allocation, SNR = ∞.

chosen from the C_{B} values used for forming the curves in Fig. 5.1, to illustrate a wide range of situations: $C_1 = 10, C_2 = 20, C_3 = 30, C_4 = 100, C_5 = 250, C_6 = 400$, bits.

Since it will play an important role in our discussions, we now briefly discuss the local coherence structure of the field. The multiple beam structure at hand can be considered to consist of two regions where in one region (around $-3\sigma_I$) the field is incoherent, whereas in the rest (around $+3\sigma_I$) coherent. Although the beams with different β values extend forever, and hence contribute to the coherence structure over the whole space, their contribution is small outside their main lobes due to comparably small intensity values outside these regions. Assuming the values of a Gaussian beyond its third standard deviation are negligible, these incoherent and coherent regions can be assumed to extend from $-3\sigma_I$ to $3\sigma_I$ around the respective beam centers.

As C_{B} increases, the general trend of the cost allocations exhibit the following behaviour for both high and low SNR cases: the optimal samples become more closely spaced, the number of effective measurements increases, and the accuracies of the samples that are among the effective measurements increases. In other

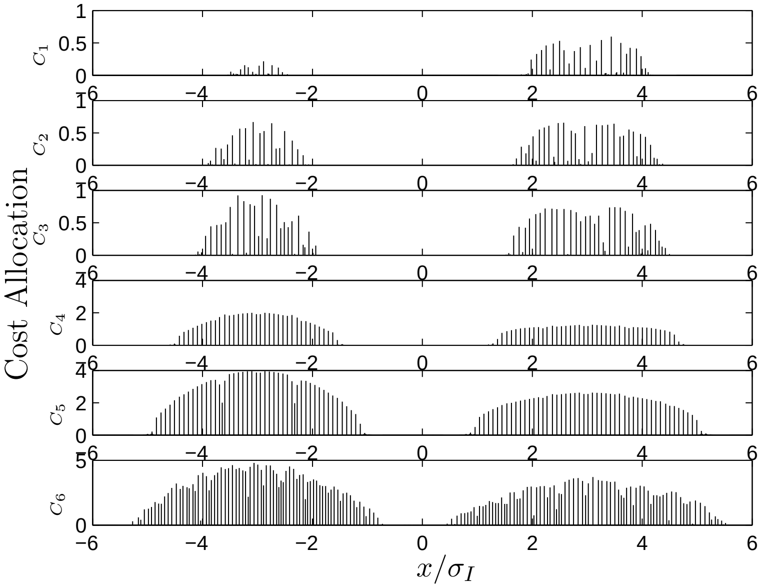


Figure 5.3: Cost allocation, SNR = 1.

words, when there are more bits to spend, one uses a higher number of more closely spaced, more accurate measurements. For a given cost, the general trend of the optimal cost allocation effectively follows two Gaussian-like curves residing side by side. We recall that the intensity distribution of the field is given by two Gaussian curves centred around $-3\sigma_I$ and $3\sigma_I$. The cost allocation is consistent with this structure. The field values that have higher intensity values are sampled with higher costs (higher measurement accuracies). This may be informally interpreted as follows: Let us first consider measurement of a single variable. For a given measurement accuracy (i.e. the cost), the uncertainty reduction due to observing a random variable with a higher variance is higher compared to observing a variable with a smaller variance (although the percentage error for any such variable will be the same). In other words, as depicted in Section 3.3.0.3, if the values to be measured are uncorrelated with each other, it is better to measure field values with higher variances using higher costs. Here the field values are not necessarily uncorrelated, but due to GSM field model the correlation function is the same for all points (given by (2.15)), and the field at locations that are close to a field value with high variances also have comparably high variances (due to Gaussian intensity distribution). These further support the above behaviour of the optimal cost distribution.

We now discuss the effect of local coherence structure of the field on the optimum measurement strategies. We first discuss the case with no system noise $n(x)$, and then discuss the effects of noise level. Looking at Fig. 4.3, we observe that the general trend of the cost allocations reflect the different degrees of coherence associated with the beams centred around $x = -3\sigma_I$ and $x = 3\sigma_I$. The beam centred around $x = -3\sigma_I$ has a smaller β, hence is more incoherent, and its main lobe is sampled with a higher number of more closely spaced samples. Comparing the total cost budget spent here to the cost budget spent around $x = 3\sigma_I$, we observe the following: For low values of cost, a smaller portion of the cost budget is spent around $x = -3\sigma_I$, whereas for high values of cost budget a larger portion of the budget is spent here. For instance, for $C_{\mathrm{B}} = 10$ bits, ≈ 0.32 of C_{B} is spent around $x = -3\sigma_I$, whereas for $C_{\mathrm{B}} = 400$ bits, ≈ 0.8 of C_{B} is spent there. This may be informally interpreted as follows: The beam centred around $3\sigma_I$ has a smaller β, hence the field is more coherent. Hence the uncertainty reduction due to taking a sample around $3\sigma_I$ with a given accuracy is higher than taking a similar sample around $-3\sigma_I$. Thus, when the cost budget is low, one prefers to take samples there. As cost budget increases, the possible error reduction due to observing those field values decrease. The uncertainty reduction that can be obtained by observing relatively incoherent samples becomes higher compared to that which can be obtained by observing more samples from the coherent side. (One may look at the cost-error trade-off curves for beams with varying β values given in Figs. 4.1 - 4.2 in the previous chapter to have a general idea about the size of the gap between the achievable error values for different β values.) Hence one starts to spend larger portions of the cost budget on the incoherent side as cost budget increases. This increase is so large that for high values of cost budget, a larger portion of the cost budget is spent around $-3\sigma_I$ compared to what is spent around $+3\sigma_I$. This is consistent with the fact that for these values of cost, the error associated with estimating the beam with $\beta = 1$, which is centred around $-3\sigma_I$ will become very low, and hence one prefers to spend the cost budget on the beam around $3\sigma_I$ which has not yet achieved such error values.

Comparing Fig. 5.2 and Fig. 5.3, we observe that as the system noise level increases, a higher number of more closely spaced samples with lower accuracies should be taken to compensate for the effects of noise. The change in the measurement strategy with increasing system noise level is more dramatic for the coherent beam centred around $3\sigma_I$, where a much fewer number of relatively spaced samples were used when there was no noise. We note that the cost allocations exhibit some fluctuations, stronger for the low SNR case, but in effect present for both noise levels. For instance, we observe that it is possible that a sample taken with high accuracy when the cost budget is low, would be taken with lower accuracy when the cost budget increases, and another sample very close to this first one would be taken with higher accuracy to compensate. This non-uniform behaviour suggests that it may be possible to achieve error values close to optimal values with more than one measurement strategy. We now compare the portion of the cost budget spent around $-3\sigma_I$ to the one spent around $3\sigma_I$ when the noise level is high. Here again one starts with spending more cost on the coherent side, and increases the cost budget spent on the incoherent side as cost budget increases. For instance, for $C_{\mathrm{B}} = 10$ bits, ≈ 0.16 of the total cost budget is spent around $x = -3\sigma_I$. Comparing this with the high SNR case, we observe that the portion of the total cost budget spent here is much lower when SNR is low. When SNR is low, one tries to compensate for the effects of noise by taking a higher number of more closely spaced samples. If the field is more coherent, one can reduce the effect of noise more easily, and achieve error values more close to the ones achieved in the noiseless case. On the other hand, if the field is less coherent, it is more difficult to reduce the effect of noise due to uncorrelated field structure. The fact that the field is more locally coherent around $3\sigma_I$ and the above driving forces encourage us to spend even a larger portion of the total cost budget around there, compared to high SNR case. As cost budget increases, the portions of the cost budget spent around $-3\sigma_I$ increase, but this increase is small compared to the increase for high SNR case. As a result, even for high values of cost budget, the portion of the total cost budget spent here is not very high. For instance, for $\mathrm{SNR} = 1$ and $C_{\mathrm{B}} = 400$ bits, ≈ 0.56 of the total cost budget is spent around $-3\sigma_I$, whereas for $\mathrm{SNR} = \infty$, this number is ≈ 0.8.

5.3 Comparison with Uniform Measurement Strategy

In this chapter, we have considered a very general measurement scenario, where the sample locations can be freely chosen, and the measurement accuracy of each sample can vary from sample to sample. In the previous chapter, we have discussed a more simple approach where the samples are equidistantly spaced, and each sample is taken with the same accuracy. Here we will compare the error versus cost budget curves that are obtained following this simple approach of the previous chapter with the curves obtained with the non-uniform approach considered in this chapter and shown in Figure 5.1. Our discussion will illustrate the performance improvement that can be gained by exploiting the flexibility offered by the non-uniform version.

Fig. 5.4 compares these trade-off curves for varying noise levels. The dotted curves and the solid lines show the results for the non-uniform measurement strategy and the uniform measurement strategy, respectively. We note that these comparisons are done after optimization of both of the measurement strategies, hence between the optimal cost-error trade-off curves. For both of the measurement strategies, the error-cost curves show the optimal trade-offs between the error and the cost budget; that is, each curve presents the best achievable error for the given cost budget under the given measurement strategy.

As expected, for all cases, the non-uniform measurement strategy of this chapter give better trade-offs compared to the uniform measurement strategy. The uniform measurement strategy of the previous chapter requires us to take equally spaced samples with uniform cost allocation, hence does not provide any room for the optimization procedure to take into account the space-varying local coherence structure of the field. Although there may be many ways that the coherence structure can change in space, the field considered in this experiment provides a simple example where the local coherence effectively varies from one region in space to another region in space: in one region (around $-3\sigma_I$) the field is incoherent, whereas in the rest (around $3\sigma_I$) is coherent. As illustrated in Figs. 5.2

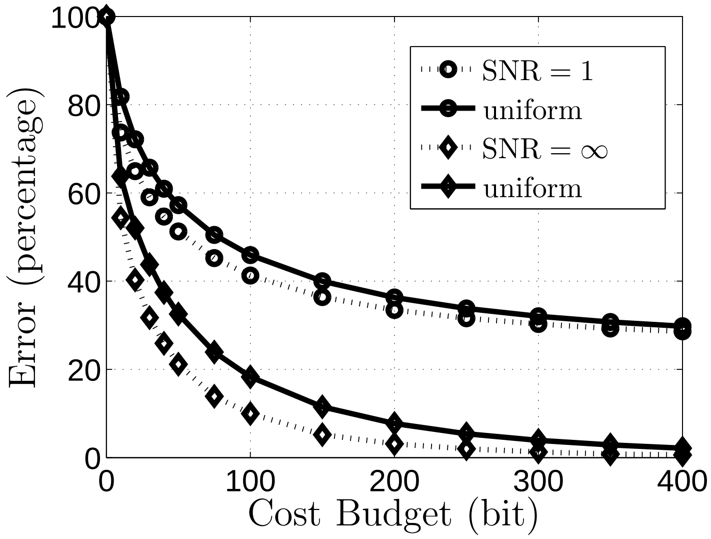


Figure 5.4: Error versus Cost budget C_{B}, varying SNR. The dotted lines are for non-uniform case and the corresponding solid lines are for the uniform case.

and 5.3 the general non-uniform strategy successfully adopts to this change in the local coherence, whereas the uniform measurement strategy cannot, resulting in worse trade-offs.

We observe that for both noise levels, the performance difference between the uniform and the non-uniform strategies are more pronounced for relatively low and moderate cost budget values. For these cost budget values, it is more important to use the limited resources in the best way possible as in the case of non-uniform sampling, without constraining the samples to be equidistantly spaced or to be taken with the same accuracy levels as in the uniform version. Hence the difference is larger. But as the available cost budget further increases, one can take even higher numbers of more and more accurate and closely spaced samples, making the selection of the sampling interval and measurement accuracies less important. As a result, the performance difference becomes comparably small for high values of the cost budget.

We also observe that the performance difference between the uniform and the non-uniform measurement strategies is more pronounced when there is no system noise. This behaviour may be informally interpreted as follows: In our setting, the main factor that creates the performance difference between the uniform and the non-uniform versions is the difference in the local coherence structure of the field. On the other hand, as the noise level increases, the effect of the coherence properties of the noise on the optimum sampling strategies become stronger. Since the coherence of the system noise process is space-invariant, the optimal sampling strategies become alike for the coherent and incoherent parts of the field, resulting in a smaller performance loss due to the use of uniform sampling strategy.

5.4 Discussion

In Sections 4.3 and 5.2, we illustrated that, given an arbitrarily small but non-zero error tolerance, it is possible to represent a finite-energy random field with a finite number of bits without exceeding that error tolerance. At first glance, this may appear as a surprising observation. After all, the random field in question takes continuous amplitude values in continuous and unbounded space, and attempting to use a finite numbers of bits to represent such a field is a severe restriction: such finite representations usually involve a finite number of samples each quantized to a finite number of levels. Therefore here we further discuss this from different perspectives.

First, consider the very simple case of a single sample of the field. Let us assume this sample can assume values between A_{low} and A_{high} and we have agreed to represent this value with an error tolerance of ΔA. Then, it follows that there will be $\sim (A_{\text{high}} - A_{\text{low}})/\Delta A$ distinguisable levels which can then be represented by $\sim \log_2(A_{\text{high}} - A_{\text{low}})/\Delta A$ bits.

Now let us return to the field $f(x)$. We may think of the finite-energy condition as a limitation on how large the amplitude values of the field can be. On the other hand, the specified error tolerance can be considered to determine the minimum separation of two signals such that they are still considered distinguishable. The finite-energy condition restricts the signal to reside within a hypersphere of specific radius, whereas the error tolerance defines a certain volume within which signals are considered indistinguishable. Roughly speaking, the number of distinguishable signals is given by the volume of the hypershere divided by the volume defined by the finite error tolerance. Since this number is finite, the signal can be represented by a finite number of bits.

We now take a somewhat more mathematical, closer look at this issue. In each step of our argument, we introduce a limitation, a form of "finiteness," in the representation of the field (such as limiting the fields to a bounded region), and argue that the error introduced by each of these limitations can be made arbitrarily small. This way, we aim to illustrate how different forms of "finiteness" contribute to the overall picture. Our approach is based on the coherent-mode decomposition. We also note that it is more common to discuss concepts related to "finiteness" in a deterministic setting, and in connection with band-limited approximations, rather than the stochastic setting and approximations based on covariance functions we employ.

Let us consider a finite-energy zero-mean random field that will be approximated using a finite number of bits. For the sake of convenience, let us assume that the random field takes real values. Let us first focus on the error introduced by the limitation of representing the signal in a bounded region D instead of the infinite line. As stated in (4.8), the total error of such an approximation can be expressed as the sum of two terms: one is the approximation error on D, and the other one is the energy outside D. The energy outside D can be made arbitrarily small by taking D large enough. This is the first form of "finiteness" introduced in the representation of the signal.

We now focus on the approximation error on D. The question is whether it is possible to make the approximation error arbitrarily close to zero; that is, whether it is possible to represent the field in a bounded region with a finite number of bits. The answer is not obvious since we are dealing with a field taking continuous amplitude values on a bounded but still continuous space. To give an affirmative answer, we will rely on the existence of the Karhunen-Loéve expansion of the covariance function of the unknown field with a discrete eigenvalue spectrum as

$$K_f(x_1, x_2) = \sum_{k=0}^{\infty} \lambda_k \phi_k(x_1)\phi_k^*(x_2), \tag{5.6}$$

where $\lambda_0 \geq \lambda_1 \ldots \lambda_k \geq \lambda_{k+1}, \ldots$ are the eigenvalues and $\phi_k(x)$ are the orthonormal eigenfunctions, $k \in Z_+$. This is the so called coherent-mode decomposition of the random optical field. Here each λ_i and ϕ_i pair is considered to correspond to one fully coherent mode. Existence of such an expansion for covariance functions on a bounded region is guaranteed by Mercer's Theorem; see for example [148, Ch.7]. Therefore, the signals can be decomposed as

$$f(x) = \sum_{k=1}^{\infty} z_k \phi_k(x), \quad x \in D \tag{5.7}$$

where the random variables z_k are zero-mean random variables with $E[|z_k|^2] = \lambda_k$. Hence a continuous field on the bounded region can be represented with an infinite but at least denumerable number of variables, namely the random variables z_k, $k \in Z_+$. Here it is also known that $\int_D K_f(x_1, x_2) dx = \sum_{k=0}^{\infty} \lambda_k$ [148, Ch.7]. Since $K_f(x_1, x_2)$ is finite-energy, the left hand side of this equation (the energy on the region D), is also finite. Hence the right hand side is also finite and we should have $\lambda_k \to 0$ as $k \to \infty$. Now, let us consider the truncation error

$$\begin{aligned} E[\int_D \|f(x) - \sum_{k=1}^{N} z_k \phi_k(x)\|^2 dx] &= E[\int_D \|\sum_{k=1}^{\infty} z_k \phi_k(x) - \sum_{k=1}^{N} z_k \phi_k(x)\|^2 dx] && (5.8) \\ &= E[\int_D \|\sum_{k=N+1}^{\infty} z_k \phi_k(x)\|^2 dx] && (5.9) \\ &= \sum_{k=N+1}^{\infty} E[|z_k|^2] && (5.10) \\ &= \sum_{k=N+1}^{\infty} \lambda_k && (5.11) \end{aligned}$$

Thus by choosing larger and larger but still finite values of N, we can bring the truncation error below any finite value, no matter how small. This observation shows that finite-energy random fields can be represented by a finite number of variables $(z_1, \ldots, z_N)$ for any given non-zero error tolerance.

Finally, we would like to argue that it is possible to represent the field not only with a finite number of variables, but also with a finite number of bits. Here the question is whether it is possible to represent the finite-variance random variables $z_1, \ldots, z_N$ with a finite number of bits, to meet a given arbitrarily small non-zero error tolerance. The answer is affirmative and a classical result in information theory (rate-distortion theory [40, Ch.13]). Although one would need an infinite number of bits to represent a continuous number perfectly (with zero error), it is possible to represent such a number with a finite number of bits with an arbitrarily small but non-zero error. With this last step, we conclude our argument showing that finite-energy random fields can be represented by a finite number of bits with an arbitrarily small non-zero error tolerance.

In the first step of the argument of this section, we argued that the error introduced by limiting the signal to a bounded region can be made small. Actually, this step can be dispensed with altogether since finite-energy fields have Karhunen-Loéve expansions on the infinite line with a discrete eigenvalue spectrum (and hence coherent-mode decompositions with denumerable modes). Indeed, in the literature authors sometimes write the coherent-mode decomposition of an optical field in the form of a summation without explicit reference to a bounded domain or any detailed discussion of the existence of such an expansion on the infinite line. Here we would like to point out that this practice is supported by mathematical results: [149, Thm. 1] states that along with continuity, having $\int_{-\infty}^{\infty} K_f(x,x)dx < \infty$ and $K_f(x,x) \to 0$ as $|x| \to \infty$ is sufficient to ensure such a representation. We note that both of these conditions are plausible in a physical context: the first one is equivalent to the finite-energy assumption and the second one requires the intensity of the field to vanish as $|x|$ increases, properties one commonly expects from physically realizable fields.

5.5 Conclusions

We have focused on the trade-offs between the achievable error and the cost budget in order to represent a random field with as small a number of bits as possible. Contrary to Chapter 4, where equidistant sampling with uniform cost allocation is considered, here we have addressed the problem of optimal non-uniform sampling with non-uniform cost allocation. In this general case, the sample locations can be freely chosen, and need not to be equally spaced from each other. Furthermore, the measurement accuracy of each sample can very from sample to sample. We have obtained the optimal number of samples, the sampling locations, and the measurement accuracies, and derived the optimal bounds for simultaneously achievable bit cost and error. Our results illustrate that in certain cases, it is possible to reach tighter cost-error trade-off curves with this general approach. We have observed how the local coherence structure of the field affects the optimum measurement strategies and how the optimal sampling parameters change with increasing cost budget. We have also investigated how all these results are affected by the noise level.

Chapter 6

Super-Resolution Using Multiple Limited Accuracy Images

In this chapter, we will study an application of the cost constrained measurement framework proposed in the previous chapters to super-resolution problems. In a typical super-resolution problem, multiple images with poor spatial resolution are used to reconstruct an image of the same scene with higher spatial resolution [11]. Here we study the effect of limited amplitude resolution (pixel depth) in this problem. The problem we address differs from standard super-resolution problems in that in our framework amplitude resolution is considered as important as spatial resolution. In standard super-resolution problems, researchers mostly focus on increasing resolution in space, whereas in our study both resolution in space and resolution in amplitude are substantial parameters of the framework. We study the trade-off between the pixel depth and spatial resolution of low resolution (LR) images in order to obtain the best visual quality in the reconstructed high resolution (HR) image. The proposed framework reveals great flexibility in terms of pixel depth and number of LR images in super-resolution problem, and demonstrates that it is possible to obtain target visual qualities with different measurement scenarios including images with different amplitude and spatial resolutions.

Many applications in image processing will benefit from such a study, for instance applications requiring converting available low resolution content to high definition television (HDTV). This subject is not merely of interest for practical purposes but can also lead to a better understanding of the effect of pixel depth in super-resolution problem. We are concerned with questions such as "To obtain a target resolution, which is better, a high number of coarsely quantized images or a low number of densely quantized images?" or "What is the range of admissible pixel depths at a particular spatial resolution to obtain an image with a target spatial resolution with a target visual quality?". Admitting great flexibility in terms of number and accuracies of the LR images, our framework is similar to other constrained signal acquisition scenarios such as compressed sensing paradigm.

The framework we have presented here can be useful in the area of high dynamic range (HDR) imaging, which is concerned with images with pixel depths greater than the conventional 8-bit pixel depth. A substantial amount of research in this area focuses on reconstruction approaches which processes multiple shots of the same scene captured at different exposures, such as [175, 176]. Each of these shots are taken with low dynamic range, and then processed to reconstruct a high dynamic range image. This approach can be interpreted as an analogy of the standard super-resolution problem. In standard super-resolution problem, multiple shots of the same scene with varying camera motions are used. Each of these shots have low spatial resolution, and then processed to reconstruct an image with high spatial resolution. Hence the above HDR approach does what common super-resolution approaches do in spatial domain, in amplitude domain. In a practical scenario, what one would desire is to combine both of these approaches, that is to increase the resolution both in spatial and amplitude domain. The achievable limits of such a scenario will be of interest for both practical scenarios and understanding the information relationships in such problems. Although our framework has some limitations from the point of view of such a broad and ambitious goal, it still can be considered a step into understanding some aspects of these relationships. In particular, our approach illustrates, under our metric of visual quality, the effect of limited amplitude resolution in the problem of increasing spatial resolution.

We emphasize that since both resolution in space and resolution in amplitude are variables in our framework, the term *low/high resolution* image is, in fact, ambiguous. Nevertheless, we use these terms to refer to images with *low/high spatial resolution* to be consistent with the literature.

6.1 Measurement Model

L low resolution images are obtained from a high resolution image $\mathbf{x}$ according to the model:

$$\mathbf{y}_k = D_k H_k F_k \, \mathbf{x} + \mathbf{v}_k, \qquad k = 1, \ldots, L \tag{6.1}$$

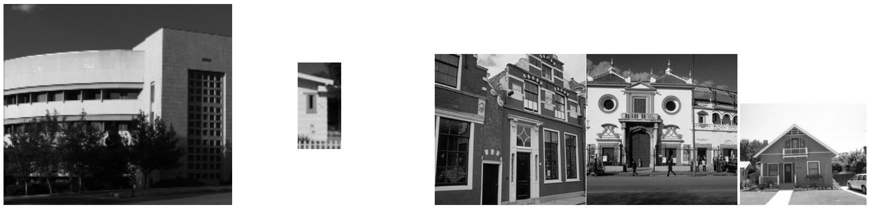
where $\mathbf{y}_k$'s are LR images, v_k's denote the system noise, D_k represents the decimation operator, H_k represents the camera blur, F_k represents the motion operator, L is the number of available LR images. v_k's are independent of each other, and the components of each v_k are i.i.d. All images are rearranged in lexicographic order. Here $\mathbf{x}$ is of size $N_1 \, N_2$, and $\mathbf{y}_k$'s are of size $\bar{N}_1 \, \bar{N}_2$, where $N_1 = r_1 \bar{N}_1$, and $N_2 = r_2 \bar{N}_2$.

We assume that we only have access to quantized LR images;

$$\mathbf{y}_k^{b_{y_k}} = Q_{b_{y_k}}(\mathbf{y}_k), \qquad k = 1, \ldots, L \tag{6.2}$$

where $Q_{b_{\mathbf{y}_k}}$ is the uniform quantizer with $2^{b_{\mathbf{y}_k}}$ levels. In general, b_{y_k} may be different for different LR images. Here, for simplicity, we assume that all LR images are quantized with the same number of bits, i.e. $b_{y_k} = b_y$.

We describe the spatial resolution of each LR image y_k relative to the spatial resolution of target high resolution image $\hat{\mathbf{x}}$, and it is given by $1/(r_1 \, r_2)$. The number of LR images may be thought as a part of spatial resolution, as well as a parameter associated with resolution in time when considered in a spatio-temporal framework. The resolution in amplitude associated with an image I is described by the number of bits used to represent pixel values b_I, which is the pixel depth.



(a) (b) (c)

Figure 6.1: Samples from the image set used in the experiments.

We associate a cost with a particular representation of a scene: cost of a quantized image is given by the total number of bits needed to represent this particular representation, i.e. number of pixels in the image × number of bits used to represent each pixel value. For example the representation cost of the HR image $\mathbf{x}$ is $C_{\mathbf{x}} = N_1 \times N_2 \times b_{\mathbf{x}}$, and similarly the representation cost of a LR image $\mathbf{y}_k^{b_y}$ whose pixel values are quantized with $b_{\mathbf{y}}$ bits is $C_{\mathbf{y}_k^{b_{\mathbf{y}}}} = \bar{N}_1 \times \bar{N}_2 \times b_{\mathbf{y}}$. The total representation cost of L low resolution images is $L \times C_{\mathbf{y}_k^{b_{\mathbf{y}}}}$.

The cost parameter provides a way of expressing the combined effect of the resolution in space, resolution in amplitude, and number of LR images for a given image acquisition scenario (given set of LR images) with a single number. We note that the actual number of bits needed to effectively store or transmit the images may be quite different from C. Our notion of cost should be considered as a part of acquisition rather than the coding of information.

The ratio of the total representation cost of L low resolution images to the representation cost of the target HR image $\hat{\mathbf{x}}$ is a useful parameter and is given by

$$C_r = \frac{L \times \bar{N}_1 \times \bar{N}_2 \times b_{\mathbf{y}}}{N_1 \times N_2 \times b_{\hat{\mathbf{x}}}} = \frac{L \times b_{\mathbf{y}}}{r_1 \times r_2 \times b_{\hat{\mathbf{x}}}}. \qquad (6.3)$$

C may be seen as a measure of information in a particular representation of scene. Hence it may be argued that if $C_r < 1$, there is not as much as information in the LR images as in the target HR image, and the problem is underdetermined in the sense of number of bits available. In a typical image, the values of different pixels are neither independent, nor necessarily identically and uniformly distributed. Yet C provides an upper bound, and still may be useful in interpretation of the results. We finally note that in a typical super-resolution problem effective bit depths of the HR image, and the LR images and achievable bit depths for the target HR image may take different but related values, which puts constraints on the values C_r can take.

6.2 Methodology

To study the trade-off between amplitude resolution and spatial resolution within the given framework, we will consider different image acquisition scenarios and compare their success in generating HR images with a particular super-resolution method.

As super-resolution method, we use the norm approximation method recently proposed in [177]. We note that one could use other image reconstruction methods as well. Although the specifics of these methods may differ, we believe that the nature of the tradeoffs observed and the general conclusions and insights that will be presented in this chapter will remain useful with a wide variety of methods. In [177], the reconstructed image $\hat{\mathbf{x}}$ is given as the following

$$\hat{\mathbf{x}} \;=\; \arg\min_{\mathbf{x}} \left\{ \sum_{k=1}^{L} \|\mathbf{y}_k - D_k H_k F_k \,\mathbf{x}\|_1 + \lambda \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{|m|+|l|} \|\mathbf{x} - S_h^m S_v^l \mathbf{x}\|_1 \right\},$$

where operators S_h^m and S_v^l shift $\mathbf{x}$ by m and l pixels in the horizontal and vertical directions, respectively. We have used $\alpha = 0.6$, and $P = 2$, which are one of the typical values used in [177]. Here $\lambda > 0$ is a scalar parameter used to control the amount of regularization. The method used to determine λ is explained in each experiment.

Structural similarity (SSIM) index [178] and peak signal to noise ratio (PSNR) are used as the quality metrics to report the success of different image acquisition scenarios. SSIM index between two images $\hat{\mathbf{x}}$ and $\mathbf{x}$ are given as the mean of SSIM over aligned image patches, where the SSIM between image patches from $\hat{\mathbf{x}}$ and $\mathbf{x}$ is given as

$$\mathrm{SSIM} = \frac{(2\,\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + C_1)\,(2\,\sigma_{\mathbf{x}\hat{\mathbf{x}}} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + C_1)\,(\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + C_2)}. \tag{6.4}$$

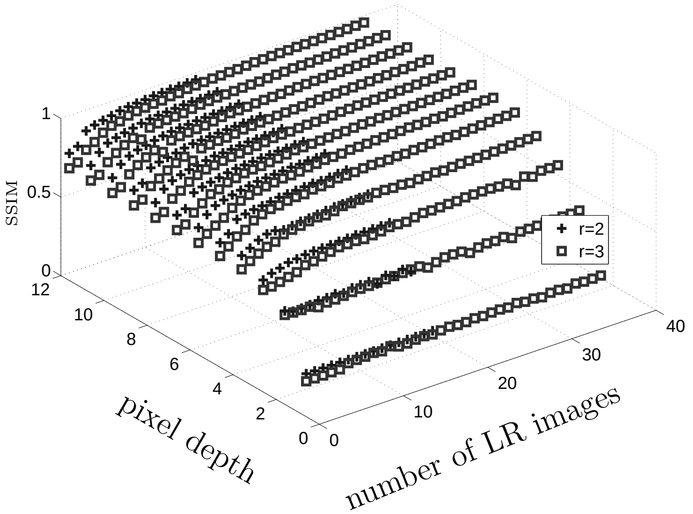


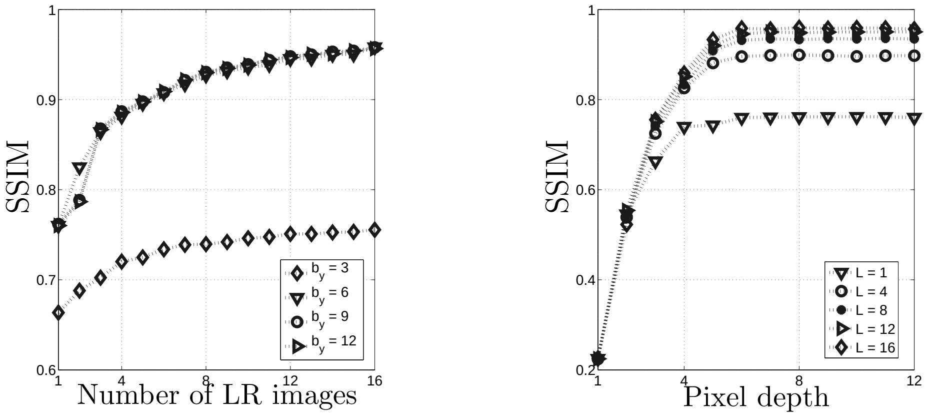
Figure 6.2: SSIM versus the number (L) and pixel depth (b_y) of LR images, upsampling factor r variable

Here $\mu_{\mathbf{x}}$, $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{x}\hat{\mathbf{x}}}$ denote the local estimates of the mean, variance and cross correlation respectively. We have used the implementation offered by [178], and reported SSIM over a dynamic range of 1 using C_1 and C_2 as $(0.01)^2$ and $(0.03)^2$ in accordance with [178].

Finally we give some of the parameters used in the experiments: The upsampling factors in two dimensions are assumed to be the same, i.e. $r_1 = r_2 = r$. Camera point spread function (p.s.f.) is assumed be 3×3 Gaussian filter. Gaussian noise with a standard deviation of 0.02 is used to simulate the system noise. Camera p.s.f. and motion vectors are assumed to be known in the reconstruction.

6.3 Experimental Results

We will now study the relationship between resolution in amplitude and resolution in space in super-resolution scenarios by examining the success of different image acquisition set-ups. This study will also reveal the trade-off between the quality (SSIM of the reconstructed images) and cost (the representation costs of LR images) under the experiment parameters used. We use $C_r = (L \times b_{\mathbf{y}})/(r^2 \times b_{\mathbf{x}})$.
Exp. 1: This experiment investigates the case where HR image is assumed to be



(a) (b)

Figure 6.3: SSIM versus the number (L) and pixel depth (b_y) of LR images (a) SSIM versus the number of LR images for $r = 2$ with varying pixel depth b_y (b) SSIM versus pixel depth for $r = 2$ with varying L.

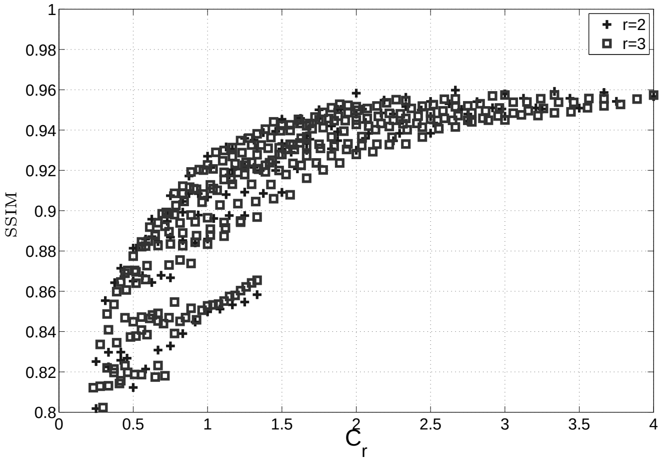
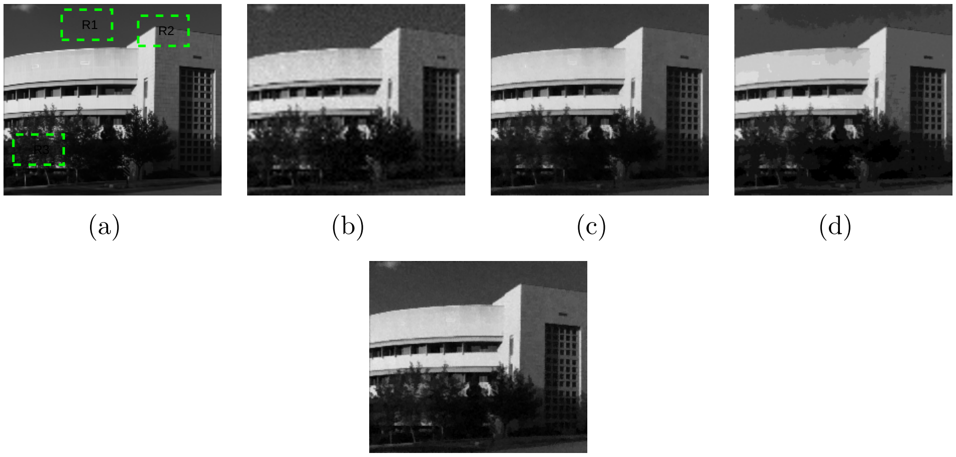


Figure 6.4: SSIM vs C_r, upsampling factor r variable, HR image is used to select λ.



(a) (b) (c) (d)

(e)

Figure 6.5: (a) HR image, (b) bi-cubic interpolation of 1 LR image with 12 bit quantization, Images reconstructed from (c) 6 LR images with 8 bit quantization (P_1) (d) 12 LR images with 4 bit quantization (P_2) (e) 4 LR images with 12 bit quantization (P_3).

known in the reconstruction process and optimum λ to obtain the best SSIM is searched heuristically. This experiment serves the purpose of providing a benchmark for the best performance possible with the reconstruction method used. For this experiment the 12-bit grayscale image, shown in Fig. 6.1(a) is used. This image includes a fair amount of edges as well as textured, and smooth regions. We consider the image acquisition strategies with pixel depths $b_y \in \{1, \ldots, 12\}$ and the number of LR images $L \in \{1, \ldots, 4\,r^2\}$ with upsampling factors $r = 2$, 3.

Figs. 6.2, and 6.3 present the SSIM for different image acquisition scenarios. The associated trade-offs between SSIM and C_r are presented in Fig. 6.4. We see that it is possible to obtain a given SSIM performance with different image acquisition strategies, and possibly different costs. In Fig. 6.4, the boundary of the achievable SSIM-C_r region shows that SSIM is very sensitive to increases in C_r for smaller values of C_r. Then it becomes less responsive, and eventually saturates at an asymptote for high values of C_r. We also note that in all of the measurement scenarios considered in this experiment, for a given pixel depth, if the total number of pixels available are the same for varying upsampling factors, SSIM values turn out to be very close. This also shows that under the image

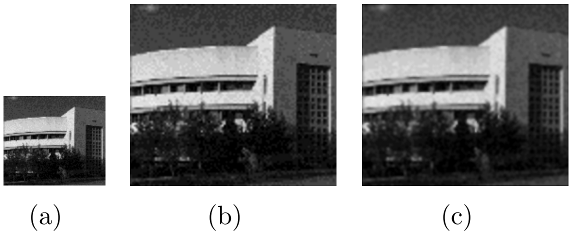


Figure 6.6: (a) LR image with 4 bit quantization ($r = 2$) (b) bi-cubic interpolation (c) after noise removal

acquisition set-ups considered in this experiment, resolution in amplitude, not resolution in space (upsampling factor), is the key factor determining the quality of reconstructed images. This trend is strongly related to the size of camera p.s.f., the size of details in the images as well as the upsampling factors used in the experiment.

We observe that in general for a given pixel depth, SSIM increases as the number of available LR images increases (see for instance Fig. 6.3(a)). We also see that for a given number of available LR images, SSIM increases with increasing pixel depth (see for instance Fig. 6.3(b)). For low values of pixel depth, the information lost due to poor resolution in amplitude can be hardly recovered by acquiring more LR images, resulting in very close SSIM values for all values of L. The increase in SSIM with increasing L is lower for low values of pixel depth compared to high values. As pixel depth increases the number of available images becomes more important in determining the SSIM level that can be reached with a particular pixel depth. However for all values of pixel depth, the increase in SSIM with increasing L gradually becomes lower as L increases.

We now take a closer look on the following data points with $r = 2$: 6 LR images with 8-bit pixel depth (P_1), 12 LR images with 4-bit pixel depth (P_2), and 4 LR images with 12-bit pixel depth (P_3). The costs of these acquisition schemes are the same, so it is reasonable to use them to compare the following different sampling strategies: a high number of images with a coarse resolution in amplitude (P_2), a low number of images with a dense resolution in amplitude (P_3), and the strategy in between (P_1).

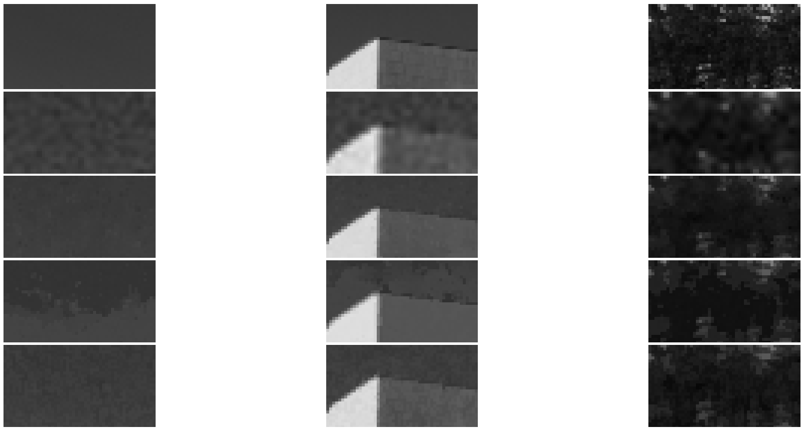


Figure 6.7: Region 1 (Left), Region 2 (Middle), Region 3 (Right): Patches from the images presented in Fig. 6.5

The actual HR image, and reconstructed images for P_1, P_2 and P_3 are shown in Fig. 6.5(a), Fig. 6.5(c), Fig. 6.5(d), and Fig. 6.5(e) respectively. The regions indicated in Fig. 6.5(a) are shown in Fig. 6.7 with the corresponding SSIM and PSNR values in Table 6.1.

We observe that there are quantization artifacts all over the image reconstructed from the set-up in P_2 (Fig. 6.5(d)). Some image details on textured regions are lost, and there are fake borders in smooth regions, which are particularly apparent in the sky region and on the building. After the noise removal, the low pixel depth of LR images causes banding in these regions, in which there is actually a smooth gray level transition. We note that these boundary effects are a result of successful noise removal. To illustrate this point, the LR image and bi-cubic interpolation of one LR image is shown in Fig. 6.6. We observe that with this naive approach the noise removal smoothes the edges and results in a blurred image. For P_3 (Fig. 6.5(e)), we observe that although most of the image details are successfully reconstructed, the image is noisy. In this case the number of available LR images is relatively low, hence they may not be sufficient to successfully remove noise without blurring. The noisy behaviour of the image suggests that using such a high pixel depth is a waste of resources, since the image pixels are already corrupted with a noise whose level is much higher than the quantization interval, and these bits could have been used to acquire more LR images. We note that by adjusting the parameter λ, it may be possible

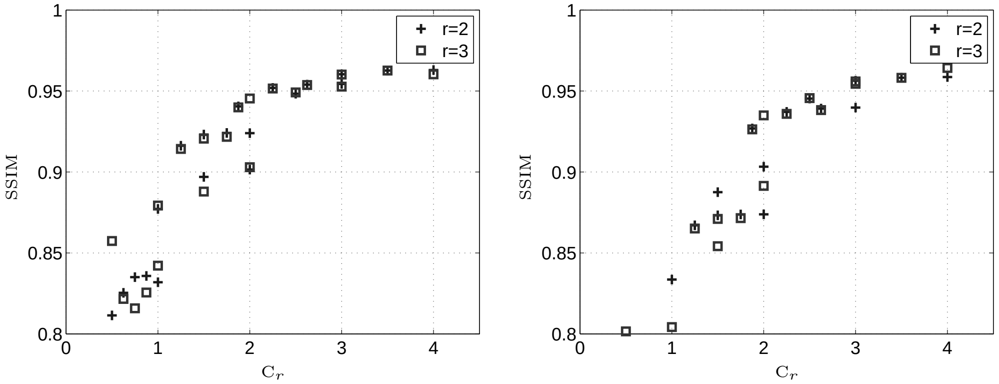
Table 6.1: SSIM and PSNR (dB) values for the image patches extracted from the image shown in Fig. 6.5(a) with different image acquisition scenarios corresponding to P_1, P_2, and P_3

	P_1	P_2	P_3
image	0.9135- 31.30	0.8540 - 29.33	0.8904 - 29.95
region 1	0.9629 - 43.48	0.8712 - 32.95	0.9300 - 40.88
region 2	0.9340 - 37.23	0.9015 - 33.14	0.9187 - 36.40
region 3	0.7879 - 27.86	0.7668 - 27.98	0.7610 - 27.19

to obtain a smoother but blurred image. We also note that if the system noise had been lower, the number of LR images at hand could have been sufficient to construct a less noisy image without blur. Finally, Fig. 6.5(c) (P_1) presents the image reconstructed from the 6 images with 8-bit pixel depth. Among the three measurement strategies, this strategy is the one that gets the highest scores from both of the quality metrics, SSIM and PSNR. We see that there is still some noise in this image, but there are no quantization artifacts similar to the ones present in Fig. 6.5(d).

Exp. 2: In this experiment, we investigate the trade-off when another image with similar characteristics is used to select λ values: The image patch shown in Fig. 6.1(b) which is extracted from an outdoor image is used to learn the optimum λ for different image acquisition schemes. We run the experiments for the first 20 8-bit images in scene categories "CALsuburb" and "MITinsidecity" from the database introduced in [179] (examples shown in Fig. 6.1(c)) and report the mean SSIM values across each image category. We consider the image acquisition strategies with pixel depths $b_y \in \{1, \dots, 8\}$ and the number of LR images $L \in \{1, r^2, 2\,r^2, 3\,r^2, 4\,r^2\}$ with upsampling factors $r = 2$, 3.

Fig. 6.8 shows the trade off between SSIM and C_r. We observe that the nature of these plots are similar to the trade-off curve presented in Fig. 6.4, in which HR image is used to select the best λ is to obtain the best performance. The SSIM values that may be reached with the image acquisition scenarios under consideration does not change significantly. We may conclude that it is possible to reach the benchmark's performance without knowing the HR image in advance,



(a)

(b)

Figure 6.8: SSIM versus C_r: upsampling factor variable, image patch shown in Fig. 6.1(b) is used to select λ. (a) database: "CALsuburban" (b) database: "MITinsidecity"

which is the case for a typical super-resolution application.

6.4 Conclusions

We have studied on understanding the relationship between resolution in amplitude and resolution in space in super-resolution problem. Unlike most previous work, amplitude resolution was considered as important part of the super-resolution problem as spatial resolution. We have studied the success of different measurement set-ups where the resolution in amplitude (pixel depth), resolution in space (upsampling factor) and the number of LR images are variable. Our study has revealed great flexibility in terms of spatial-amplitude resolutions in super-resolution problem. We have seen that it is possible to reach target visual qualities with different measurement scenarios including varying number of images with different amplitude and spatial resolutions. Our results illustrate how coarsely the images with low spatial resolution could be quantized in order to obtain images with high spatial resolution with good visual qualities. We believe that there is a great deal of exciting work to be done to understand the relationship between resolution in amplitude and resolution in space in super-resolution problem.

Part II

Coherence, Unitary Transformations, MMSE, and Gaussian Signals

Chapter 7

Coherence of Bases and Coherence of Random Fields: A Unifying Perspective

Beginning with this chapter, we will discuss a family of problems that aim to provide insight into the correlation structure of random fields. This investigation will help us to explore the relationship between the estimation error and the geometry of statistical dependence in the measurement domain. In these investigations, the unitary transformation that connects the canonical signal space and the measurement space will play an important role. In this chapter, our investigation will be based on two concepts: coherence of bases as defined in compressive sensing and degree of coherence of a random field as defined in optics. One of the main aims of this chapter is to point out the possible relationship between these two seemingly different concepts.

Compressive sensing problems heavily make use of the notion of coherence of bases, for example [13, 14, 17]. The coherence of two bases, say the intrinsic signal domain ψ, and the orthogonal measurement system ϕ is measured with $\mu = \max_{i,j} |U_{ij}|$, $U = \phi\psi$ providing a measure of how concentrated the columns of U are. When μ is small, one says the mutual coherence is small. As the coherence gets smaller, fewer samples are required to provide good signal recovery guarantees.

Theory of partially coherent light is a well-established area of optics, see for example [15, 16] and the references therein. Coherence is the central concept, which describes the overall correlatedness of a random field. One says that a random field is highly coherent when its values at different points are highly correlated with each other. Hence intuitively when a field is highly coherent, one will need fewer samples to have good signal recovery guarantees. (Please see Section 7.2 for a clarification of the naming of extreme points,i.e. full incoherence and full coherence, in compressive sensing framework and in optics.)

Thus we are faced with two concepts: named exactly the same, but attributes of different things (bases and random fields), important in different areas (compressive sensing and statistical optics), and yet enabling similar type of conclusions (good signal recovery performance). One of the main contributions of this study is to explore the relationship between these concepts, and demonstrate that the similarities are more than a coincidence.

In optics, precise quantification of coherence depends on the context: one may talk about notions like coherence length or area; or one may use the covariance matrix itself as a whole. Here we develop an alternative estimation based framework to quantify this qualitative concept, different from the traditional approaches in optics. To do so, we make the additional observation that the estimation error of a field reconstructed from its samples, in essence, should be related to the correlation between its values at different points: when the values of a field at different points are highly correlated with each other, one will need fewer samples to estimate it with low values of error. Hence the estimation error of a field from its samples may be used as a measure of corrrelatedness of a random field, hence the coherence of it. In this chapter we propose an estimation error based framework to develop this intuition and quantify coherence. Such a study is not appealing just because of the importance of coherence concept in optics, but also because of its potential to provide a novel perspective in signal modelling and inverse problems in the field of statistical signal processing.

Since coherence is argued to be a measure of overall correlatedness of the field, one may wonder its relationship with more traditional concepts which measure the total uncertainty of a random source, such as the degree of freedom or the entropy. Our study reveals insights on these relationships; most importantly contrary to what one may suspect, we argue that what coherence quantifies is not just a repetition of what the entropy or the degree of freedom does.

Our study also proposes fractional Fourier transform (FRT) as an intuitively appealing and systematic way to generate bases with varying degree of coherence: by changing the order of the FRT from 0 to 1, it is possible to generate bases whose coherence ranges from most coherent to most incoherent. Moreover, we show that by using these different bases with different FRT orders, it is possible to generate statistics for fields with varying degree of coherence. Hence we also propose the FRT as a systematic way of generating the statistics for fields with varying degree of coherence. This observation also illustrates how definition of coherence of bases in compressive sampling can be used to generate statistical signal models so that the associated fields have varying degrees of correlatedness (coherence).

7.1 Preliminaries

7.1.1 Signal model

We model our signals as zero-mean proper Gaussian vectors. The statistical properties of such a Gaussian random vector x is characterized by its covariance matrix $\mathbf{K}_x = E[xx^\dagger] \succeq 0$. We include the positive-semidefinite matrices except the zero matrix (all entries are zero) in our model. The covariance matrix can be studied through its singular value decomposition: $\mathbf{K}_x = U\Lambda_x U^\dagger$, where U is a $N \times N$ unitary matrix, and $\Lambda_x = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ is the diagonal matrix of eigenvalues, where the eigenvalues are indexed in decreasing order as $\lambda_1 \geq \lambda_2, \dots, \geq \lambda_N$. Here $\dagger$ denotes complex conjugate transpose. Throughout the chapter we assume that the signal dimension N, and $\mathrm{tr}\,(K_x) = P < \infty$ is fixed.

In the field of information theory, entropy is a concept proposed to quantify the uncertainty of a random source. The differential entropy of the above Gaussian source is given by $h(x) = \log((2\pi e)^N |\mathbf{K}_x|)$ bits, where $|.|$ denotes the determinant. Hence it is characterized by the eigenvalues

$$h(x) \propto \log(|\mathbf{K}_x|) = \sum_{i}^{N} \log(\lambda_i). \tag{7.1}$$

We note that among the sources with a given covariance matrix, Gaussian sources have the highest entropy [180, Lemma 2], [56, Thm. 9.6.5]. Hence, in this sense, our signal model can be considered to represent the worst case scenario.

Let $D(\delta)$ be the smallest number satisfying $\sum_{i=1}^{D} \lambda_i \geq \delta P$, where $\delta \in (0, 1]$. Hence for δ sufficiently close to one, $D(\delta)$ can be considered as the effective rank of the covariance matrix and also the effective number of "degrees of freedom" (DOF) of the signal family. For δ close to one, we drop the dependence on δ and use the term effective DOF and the notation D to represent $D(\delta)$.

7.1.2 Coherence as a descriptor of bases

Consider the following decomposition of the matrix $U = \phi\psi$. In compressive sensing framework, the coherence of two bases, the intrinsic orthogonal signal domain ψ, and the orthogonal measurement system ϕ is measured with $\mu = \max_{i,j} |U_{ij}|$ providing a measure of how concentrated the columns of U are (One mostly uses $\mu = \sqrt{N} \max_{i,j} |U_{ij}|$ in compressive sensing, here we drop $\sqrt{N}$, since it is merely a scaling).

If a row of U is such that all the entries of the row vanish except one, then μ gets its maximal value: 1. If all entries have equal magnitude, μ gets its minimal value: $1/\sqrt{N}$. We note that the identity matrix and the DFT matrix are two matrices that are examples of these two extremes.

We observe that the FRT provides an intuitively appealing interpolation between identity matrix and the DFT matrix. The FRT is the fractional operator power of the Fourier transform with fractional order a [18]. When $a = 0$, the

FRT matrix reduces to the identity, and when $a = 1$ it reduces to the ordinary DFT matrix. All FRT matrices are unitary matrices. In the coming sections, we demonstrate that the FRT also provides a satisfying interpolation between the incoherent and coherent limits.

As noted above, the coherence is formally defined between two bases. In this chapter, we sometimes talk about coherence of U, implying coherence of the orthogonal transforms ϕ and ψ forming U, or the coherence of U and the standard basis I (which is in fact a special case of the former with $\phi = I$, and $\psi = U$), depending on the context.

7.1.3 Coherence as a descriptor of random fields

We now identify the limits of full coherence and incoherence of a random field based on its covariance matrix.

7.1.3.1 Full Incoherence

We say that a field is fully incoherent when any two distinct samples of the field is uncorrelated. Hence the covariance matrix of a fully incoherent random field should be diagonal. (In fact under our Gaussian assumption, the field values at different locations are also independent.)

Let $\operatorname{tr}(K_x) = P < \infty$. For a diagonal matrix to be a valid covariance matrix, as long as the power constraint is satisfied the only requirement is that diagonal entries should be nonnegative or positive (corresponding to positive-semidefinite and positive-definite matrices, respectively.) These diagonal values are also the eigenvalues of this matrix. Hence by (7.1) an incoherent field may have varying values for the entropy (hence uncertainty as measured in information theory). Furthermore, any such diagonal matrix can be the eigenvalue matrix of another covariance matrix and these are the only valid eigenvalue distributions a covariance matrix can have. Hence the total uncertainty of a totally incoherent field (as measured with entropy) not only may have varying values, but also these are the only possible values.

At first sight the following reasoning may seem plausible to some: One may think correlatedness as a measure of uncertainty in a signal; for instance when the values of a field at different points are not correlated with each other, the total uncertainty in the source should be high. Hence concept of coherence should be another way to characterize what concepts such as the entropy or the degree of freedom, traditionally associated with uncertainty, characterize. As the argument in the above paragraph shows this is not the case; a fully incoherent field can have all possible entropy and $D(\delta)$ values a covariance matrix may have.

7.1.3.2 Full Coherence

It is reasonable to say that in the fully coherent case, the field values at all points should be fully correlated with the values of the field at all the other points, i.e. the correlation coefficient should have its maximum value, for any pair of points of the field [16]. That is,

$$C_x(i,j) = \frac{K_x(i,j)}{\sqrt{K_x(i,i)K_x(j,j)}} = 1, \quad i,j = 1,\dots,N \tag{7.2}$$

assuming $K_x(i,i) > 0$, $K_x(j,j) > 0$. Hence

$$C = AKA^{\mathrm{T}}, \tag{7.3}$$

where $A = \mathrm{diag}([1/\sqrt{K_x(1,1)},\dots,1/\sqrt{K_x(N,N)}])$. As a result $\mathrm{rank}\{C\} = \mathrm{rank}\{AK_xA^{\mathrm{T}}\} = \mathrm{rank}\{K_x\}$, since A is invertible. We also know that $\mathrm{rank}\{C\} = 1$ (since for example all rows are multiplies of one row). Hence $\mathrm{rank}\{K_x\} = 1$.

Hence in the fully coherent case the covariance matrix should be of rank one. In this case there will be only one independent variable, and all the components of the Gaussian vector will be scaled versions of this one independent variable.

We now discuss whether we should include matrices with $K_x(j,j) = 0$ for some $1 \leq j \leq N$ while characterizing the fully coherent case. Our answer is negative, and we impose the following additional condition for describing the fully coherent case: none of the diagonal entries should be zero, i.e. variances of all of the components should be nonzero. In this way, any one of the field values is just an invertible scaling of any other field value. Hence knowing the value of the field value at any point is equivalent to knowing the value of the field at any other point, and hence at all the other points of the space.

Let us consider what happens when this condition is not imposed: Suppose that we have a rank one covariance matrix with one diagonal entry zero, such as

$$K_x = uu^\dagger = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \tag{7.4}$$

where $u = [1/\sqrt{2},\, 0,\, 1/\sqrt{2}]^\dagger$. Hence the values of f_1, and f_3 do not depend on f_2, and similarly the value of f_2 (which is 0 with probability 1) does not depend on the other two. These random variables are statistically independent. We don't want to call such a field fully coherent.

This condition also prevents the following misleading interpretation in the case of diagonal matrices: Any diagonal matrix with only one nonzero diagonal value will also be rank one. One may try to argue that this field is coherent, since the components are in fact just scaled versions of one variable (the ones with variance 0 are scaled with 0). Hence such a field is called both incoherent (since diagonal) and coherent (since rank 1). The additional nonzero diagonal condition prevents this confusion.

To sum up, we say that a field is fully coherent, when the covariance matrix is rank one and variance of the none of the components is zero, i.e. $K_x = Puu^\dagger$, $u \in \mathbb{C}^N$ $|u_i| > 0$, $||u||^2 = 1$.

7.2 A General Discussion of Coherence Measures

Motivation: Given that two coherence definitions are attributes of different things (bases and random fields), one may be tempted to disregard the fact that both concepts are named the same as a mere coincidence. We now state the main observations that motivated us to investigate whether these concepts are related beyond a similarity in name:

We have observed that both of these concepts, coherence of random fields and bases, are associated with similar type of conclusions; both of them may be used to express conditions for good signal recovery: as the coherence of a basis gets smaller, fewer samples are required to provide good signal recovery guarantees; when a field is highly coherent, its values at different points will be highly correlated with each other and intuitively one will need fewer samples to have good signal recovery guarantees.

Another related observation is the following: As stated in [181], in compressive sensing the good bases are the ones where ”each measurement picks up a little information about each component.” The coherence of two bases (measurement and signal basis) is a measure of this property. Intuitively speaking, if each measurement picks up a little information about each component, the variables that are measured should be highly correlated; hence the total correlatedness or in other words coherence of the random field should be larger. In other words as the bases transforming the signal from its canonical domain to measurement domain become better as measured by coherence of the bases, the resulting field should become more correlated.

These two observations suggest that there may be a fundamental relationship between the concepts of coherence of bases and coherence of random fields. Our study addresses this problem.

Our Approach: We now give a brief overview of our approach to the problem of quantification of coherence of random fields. In optics, precise quantification of coherence depends on the context: one may talk about notions like coherence length or area; or one may use the covariance matrix itself as a whole. Contrary to various approaches for quantification of coherence in general, the characterization of the extreme cases, full incoherence and full coherence is well-understood. Here we would like to propose a scalar measure based on the covariance matrix of the random field, which is consistent with the characterization of the extreme cases, full incoherence and full coherence, as presented in Section 7.1.3. We develop our measures in a estimation framework. Our measure depends on both eigenvalues and eigenvectors of the covariance matrix, compared to approaches focuses on the eigenvalues [182, 183].

When a field is highly coherent, it is understood that its values at different points are highly correlated with each other. Here we make the additional observation that the estimation error of a field reconstructed from its samples is related to the correlation between its values at different points: when the values of a field at different points are highly correlated with each other, one will need fewer samples to estimate it with low values of error. Similarly, when the values of the field are uncorrelated with each other, one will need higher number of samples. Hence the estimation error of a field from its samples may be used as a measure of corrrelatedness of a random field, hence the coherence of it. Hence in this chapter we propose an estimation error based framework to quantify coherence. This framework is developped in Section 7.3.

Naming of Extreme Cases in Different Contexts: We note that the naming of incoherent and coherent limits in the contexts of bases and random fields may be confusing at first. We now clarify this issue.

A point that may cause confusion is the fact that different limits are associated with good performance. For bases, incoherent bases are associated with good performance, whereas for fields coherent fields are so. In compressive sensing, when the coherence of two bases is smaller, (hence incoherent) fewer samples are required to provide good performance guarantees. On the other hand, when we are talking about coherence of random fields, intuitively it is when a field is highly coherent, hence its values at different points are highly correlated with each other, one will need fewer samples to have good performance guarantees.

Another related point that may cause confusion is the fact that the same unitary transform is associated with different extremes: when $U = I$, it is called a coherent base (with a slight abuse of terminology as noted in Section 7.1.2). On the other hand, $U = I$ is the unitary transform in the s.v.d. of a covariance matrix associated with an incoherent field.

We note that both of these choices of naming are meaningful in their respective contexts. For the case of random fields, when the values of the field at different points are correlated with each other, the field is called coherent. In the case of coherence of bases, the situation is much clear when we consider the measurement system matrix ϕ and the signal canonical domain matrix ψ. μ may be represented as $\mu(U) = \mu(\phi\psi) = \max_{i,j} |U_{ij}| = \max_{i,j} | < \phi_i, \psi_j > |$, where ϕ_i are rows of ϕ and ψ_i are columns of ψ. . Hence μ can be said to give ' the largest correlation between rows of ϕ and columns ψ' [184]. Hence if elements of ϕ and ψ are correlated, μ is high [184]. As a result, when the elements of these two bases are 'correlated', the bases are called coherent.

In this chapter, we choose not to rename any of these coherence concepts and remain consistent with the associated fields of research. We hope that the meaning will be clear from the context.

7.3 MMSE based Coherence Measures for Random Fields

In Section 7.2, we have related the coherence of a random field and the estimation error associated with the random field when reconstructed from its samples, and propose to use this relationship to quantify coherence. Here we develop this framework.

We present a framework where degree of coherence is single parameter which describes the estimation performance of a family of measurement systems. To this end, we choose a family of tractable, meaningful set of measurement systems. We choose the MMSE as the performance criterion. We ask ourselves the following question "Is there a single parameter so that the estimation performance of this set is characterized?".

We note that there may be various approaches to characterize the estimation performance over a set of measurement systems. Here we base our approach on the expected performance over the given measurement strategy set $\mathcal{S}$.

We consider the following measurement scenario

$$y = \mathbf{H_s} x, \tag{7.5}$$

where $x \in \mathbb{C}^N$ is the unknown input proper complex Gaussian random vector, and $y \in \mathbb{C}^M$ is the measurement vector. $\mathbf{H_s}$ is the $M \times N$ measurement matrix.

We consider the following measurement strategies:
i) *Random Scalar Channel (S_o):* $H = e_i^T$, $i = 1, \dots, N$ with probability $\frac{1}{N}$.
ii) *'All But One' Channel (S_a):* $H = \{I \setminus e_i^T\}$, $i = 1, \dots, N$ with probability $\frac{1}{N}$. Here $\{I \setminus e_i^T\}$ denotes the matrix formed by deleting the i^{th} row of the identity matrix.
iii) *Bernoulli Sampling Channel ($S_b(p)$):)* $H = \operatorname{diag}(\delta_i)$, where δ_i are i.i.d. Bernoulli random variables with probability of success p.
iv) *Equidistant Sampling ($S_u(\Delta N)$):* Every 1 out of ΔN samples are taken; $H \in \mathbb{R}^{\mathbb{M} \times \mathbb{N}}$ is the sampling matrix formed by taking every 1 out of ΔN rows of the identity matrix. The first sample is taken at one of the first ΔN samples with equal probability.

We note that all of these measurement strategies have an intuitive appeal for characterizing the overall correlatedness of a field: Random scalar channel quantifies -in terms of the MMSE- on average how much the field value at a location tells about the field values at all the other points. 'All but one' channel quantifies on average how much uncertainty is left at particular point on the field when all the other values are known. In Bernoulli sampling channel, the field value at each point contribute to the estimation with the same probability. It is also satisfying to note that for performance guarantees that hold with high probability this Bernoulli sampling model and the uniform random sampling model, where measurement locations chosen uniformly from the set of all subsets of $p\,N$ is equivalent [20, 185–187]. The equidistant sampling strategy is a standard sampling strategy, for which through randomization of the location of the first sample, we achieve an averaging effect over different points in space.

We now propose some MMSE based coherence measures based on these measurement strategies. We first define a set of intermediate variables, which we denote by c', which may have different values corresponding to the incoherent and coherent limits. Then, we finalize the definition by appropriate scaling of the measure so that the resulting measure $c \in [0,1]$, and $c = 0$, and $c = 1$ for the fully incoherent case, and the fully coherent case respectively.

For notational convenience, let average error over the statistics of the signal X be expressed as follows

$$\varepsilon \quad = \quad E_X[||x - E[x|y]||^2] \tag{7.6}$$

Hence $E_{H_s,X}[||x - E[x|y]||^2] = E_{H_s}[\varepsilon]$, where E_{H_s} denotes the expectation over the sampling strategy S_s, $s = o, a, b, u$.

We consider the following family of definitions

$$c'_s = \frac{E_{H_s,X}[||x - E[x|y]||^2]}{\mathrm{tr}(K_x)} = \frac{E_{H_s}[\varepsilon]}{\mathrm{tr}(K_x)}, \tag{7.7}$$

where $s = o, a, b, u$. We note that by definition $0 \leq E_{H_s}[\varepsilon] \leq \mathrm{tr}(K_x)$, hence $c'_s \in [0,1]$.

Another related measure may be based on the following observation: It is appealing to consider the number of measurements that should be done in order to achieve a certain level of performance as a measure of coherence. As the required number of measurements increase, one would like to say the field becomes more incoherent. For instance to have zero error, for a fully incoherent field with variances strictly greater than zero, one would need to measure all of the components. On the other hand, it is sufficient to have one measurement at any location in the coherent case. To have a measure that is not dependent on the locations of the measurements of a given strategy, we consider the Bernoulli sampling channel, and the following definition

$$c_{bt} = \inf_{p \in [0,1]} p, \tag{7.8}$$

such that

$$\frac{E_{H_e,X}[||x - E[x|y]||^2]}{tr(K_x)} \leq R, \tag{7.9}$$

where $0 < R < 1$ is the threshold parameter, which may be interpreted as the desired level of estimation performance. We note that $0 \leq \frac{\varepsilon}{tr(K_x)} \leq 1$ by definition. Furthermore, since in Bernoulli sampling strategy, with probability $(1-p)^N$, none of the components of the vector is measured, the expected error $E_{H_e}[\varepsilon]$ will always greater than zero. Hence $R = 0$ is not a meaningful parameter. On the other hand for $R = 1$, all p values are admissable, (even $p = 0$ case, i.e. doing no measurements) hence $c_{bt} = 0$.

We note that such a definition will not be meaningful for the random scalar and 'all but one' channels, since they do not have such a parameter to minimize over. For the equidistant sampling case, a similar minimization over ΔN can be thought, which by definition can only take a finite number of values because of the discrete nature of ΔN.

We finalize the coherence definitions by scaling them as follows

$$c_s = \frac{c_{s,inc} - c'_s}{c_{s,inc} - c_{s,coh}}, \tag{7.10}$$

where $s = o, a, b, u, bt$. Here $c_{s,inc}$, and $c_{s,coh}$ are the c'_s values for the incoherent field, and the coherent field respectively. These values are discussed at the end of this section.

After scaling, all of the definitions satisfy $c \in [0, 1]$, and $c = 0$, and $c = 1$ for the fully incoherent case, and the fully coherent case respectively. All of these individual definitions, and the associated family of definitions parametrized by p, ΔN or R (whenever applicable) provide possibly different interpolations between these two extremes. This point is further discussed in Section 7.4.

Before leaving this section, we give the expressions for the MMSE estimator and associated error: Under a given measurement matrix H, by standard arguments the MMSE estimate is given by $E[x|y] = \hat{x} = K_{xy}K_y^+ y$, where $K_{xy} = E[xy^\dagger] = K_x H^\dagger$, $K_y = E[yy^\dagger] = HK_xH^\dagger$, and $^+$ denotes the Moore-Penrose pseudo-inverse [188, Ch.2]. The associated MMSE can be expressed as follows [188, Ch.2]

$$E_X[||x - E[x|y]||^2] = \operatorname{tr}(K_x - K_{xy}K_y^{-1}K_{xy}^\dagger). \tag{7.11}$$

Full Incoherence: We now find the $c_{s,inc}$ values for different measurement strategies. In this case the covariance matrix is diagonal. By (7.11) for the random scalar channel, the measure for the incoherent field can be expressed as follows

$$c_{o,inc} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{tr}(K_x) - \sigma^2_{x_i})}{\mathrm{tr}(K_x)} = \frac{N-1}{N}. \tag{7.12}$$

For the 'all but one' channel, the measure for the incoherent field can be expressed as follows

$$c_{o,inc} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\sigma^2_{x_i})}{\mathrm{tr}(K_x)} = \frac{1}{N}. \tag{7.13}$$

For the Bernoulli sampling strategy with probability of success p,the measure for the incoherent field can be expressed as follows

$$c_{b,inc} = \frac{E[\sum_{i=1}^{N}(1-\delta_i)(\sigma^2_{x_i})]}{\mathrm{tr}(K_x)} = 1-p \tag{7.14}$$

For equidistant sampling, the measure for the incoherent field can be expressed as follows

$$\begin{aligned} c_{u,inc} &= \frac{\frac{1}{\Delta N}\sum_{i=1}^{\Delta N}(\mathrm{tr}(K_x) - \sum_{k=0}^{N/\Delta N-1}\sigma^2_{x_k \Delta N+i})}{\mathrm{tr}(K_x)} & (7.15)\\ &= \frac{\Delta N - 1}{\Delta N}. & (7.16)\end{aligned}$$

For c'_{bt}, by (7.9) and (7.14) we have the following condition to be satisfied, $1-p \leq R$, hence $\hat{c}'_{et,inc} = 1-R$. We note that each of the measures satisfy the requirement that the coherence value reported does not depend on the exact statistics of the field as long as the field is incoherent (diagonal covariance matrix).

Full Coherence: Assume that we have a fully coherent field, i.e. a field with a rank one covariance matrix with nonzero variances. Then by measuring one of the components, it is possible to have the estimation error zero, i.e. $E_X[||x - E[x|y]||^2] = 0$. As a result for a field with such covariance matrix, $c'_{o,coh} = 0$, $c'_{a,coh} = 0$ and $c'_{u,coh} = 0$, since in these sampling strategies averages are taken over realizations where in at least one of them, one measurement is guaranteed to be done. For the Bernoulli sampling strategy, with probability $(1-p)^N$, none of the measurements are done, hence the error can be expressed as follows

$$\begin{aligned} E_{H_e}[\varepsilon] &= (1-p)^N\,\mathrm{tr}(K_x) + (1-(1-p)^N)\,0 & (7.17)\\ &= (1-p)^N\,\mathrm{tr}(K_x). & (7.18)\end{aligned}$$

Hence $c'_{b,coh} = (1-p)^N$. For c'_{bt}, by (7.9) and (7.18) we have the following condition to be satisfied, $(1-p)^N \leq R$, hence $c'_{et,coh} = 1 - R^{1/N}$. Since $0 < R < 1$, $c'_{et,inc} < c'_{et,coh}$ for $N > 1$. We note that each of the measures satisfy the requirement that the coherence value reported does not depend on the exact statistics of the field as long as the field is coherent.

7.4 Coherence of Bases and Coherence Measures for Fields

We now investigate the relationships between coherence of bases and coherence of random fields. We first investigate the coherence of the FRT matrices of different orders according to the coherence definition proposed in the compressive sensing. For the generation of the FRT matrices, an implementation of the algorithm presented in [171] and in [18] is used; this implementation is available at [172].

Fractional Fourier Transform and Coherence of Bases as Measured in Compressive Sensing Framework: We now consider the coherence of two bases as defined by compressive sensing framework $\mu = \max_{i,j} |U_{ij}|$ and the FRT. As noted earlier the identity matrix and the DFT matrix are examples of the extreme points of bases, most coherent and incoherent respectively. In this experiment we investigate whether the FRT, which provides an interpolation between the identity operator and Fourier transform, provides a satisfying interpolation between coherent and incoherent limits as measured by this measure.

In Fig. 7.1, the horizontal axis is the order of the $N \times N$ FRT matrix. The vertical axis is the scaled coherence $\bar{\mu}$

$$\bar{\mu} = \frac{\mu - 1}{1 - 1/\sqrt{N}}, \tag{7.19}$$

where $\bar{\mu} \in [0, 1]$. We observe that despite some minor fluctuations, the FRT provides a satisfying interpolation between the incoherent and coherent limits of unitary transforms with respect to coherence measure used in compressive sensing. We see that as N becomes larger, the interpolation becomes more satisfactory,

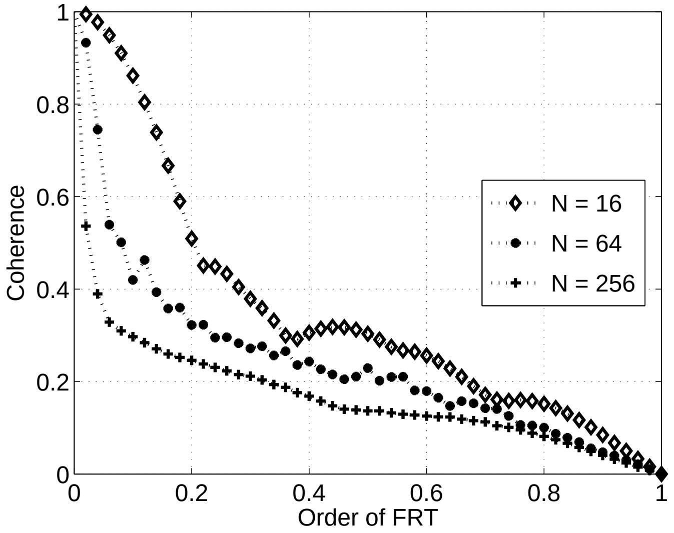


Figure 7.1: Coherence vs. order of the fractional Fourier matrix, N variable.

i.e. there are less fluctuations.

*Coherence of Bases and MMSE based Coherence Measures for Fields:*Here we investigate the coherence of random fields associated with different covariance matrices. To generate the covariance matrices used in the experiments, we consider the singular value decomposition of the covariance matrix: $\mathbf{K}_x = U\Lambda_x U^\dagger$, where $\Lambda_x = \text{diag}(\lambda_1, \dots, \lambda_N)$ is the diagonal matrix of eigenvalues, and U is a $N \times N$ unitary matrix. In a given experiment, we fix the eigenvalue distribution Λ_x, and look at the coherence of the field as we change the unitary basis U. As unitary matrices, we use the FRT matrices. To obtain different unitary matrices, we change the order a of the FRT matrix. Since the eigenvalue distribution is fixed, the entropy and hence the total uncertainty of the source is fixed. As a result, as we change U, what we change is not the total uncertainty of the field, but its distribution among the components of the signal. In Section 7.4, we have illustrated that the FRT order can be considered as a rough measure of coherence of bases. Hence in the upcoming development the FRT order (the horizontal axis in the plots) can be also interpreted as a measure of coherence of bases.

To obtain covariance matrices with different effective ranks, we choose the eigenvalue distribution as follows: Let $\delta \in [0, 1]$ be close to one. Let first D of the eigenvalues be $\frac{\delta P}{D}$, and others $\frac{(1-\delta)P}{N-D}$. As noted in Section 7.1.1, here the parameter D can be interpreted as the effective rank of the covariance matrix for δ sufficiently close to one. We set $\delta = 1 - \epsilon$, where $\epsilon = 10^{-5}$. In our experiments, D takes the values $D = \alpha N$, where $\alpha \in (0, 1]$ takes the values $\alpha = [1/16,\ 8/16,\ 15/16]$. In our experiments, we set the signal dimension as $N = 128$, and $\operatorname{tr}(K_x) = N$.

In the following experiments, we illustrate how different MMSE based definitions quantify total correlatedness of a random field. We consider the sampling strategies, random scalar channel (S_o), all but one channel (S_a), Bernoulli sampling channel (S_b) and equidistant sampling (S_u) and associated definitions as introduced in Section 7.3.

In our plots, we report error as the percentage MMSE defined as $\bar{\varepsilon} \in [0, 100]$, $\bar{\varepsilon} = 100\frac{\varepsilon}{\operatorname{tr}(K_x)}$. We choose p of the S_b as $p = D/N$, where D is the effective rank of the covariance matrix (and also effective DOF of the field). For S_u, we take D samples and distribute the samples over the range of 1 to N in an evenly fashion (as much as D and N values permit). In the case of Bernoulli sampling strategies, S_b, we simulate the expectation over the sampling strategy by taking average over $N_s = 200$ realizations. For c_t, we consider the p values $(1/N) \times [1, 2, \ldots, N]$ in our simulations, and choose the best p value from this set.

Low Effective DOF: Let $\alpha = 1/16$, hence the effective DOF $D = \alpha N = (1/16) \times 128 = 8$. In Fig. 7.2, and Fig. 7.3, we plot the MMSE and the associated measures of coherence, respectively. Since it is instructive to see the both plots together, we present both of them although they carry similar information.

We observe that for each sampling strategy, general trend of the MMSE performance is consistent with FRT order; the MMSE, in general, decreases as the order of the FRT increases from 0 to 1, i.e. as the unitary transform changes from the identity matrix to the DFT matrix. Since as illustrated in Section 7.4, the FRT order can be considered as a rough measure of coherence of bases, we conclude that MMSE performance is, in general, consistent with coherence definition of compressive sensing; as the bases become more incoherent, better MMSE values are obtained for all sampling strategies considered in this experiment. Moreover, as the order of the FRT increases, the coherence values increases from

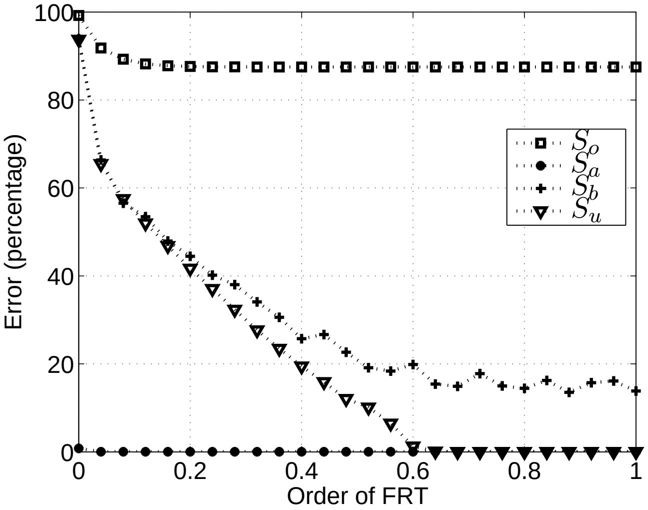


Figure 7.2: Error vs. order of the fractional Fourier matrix, $\alpha = 1/16$, sampling strategy variable.

0 to their saturation values, that is the random fields become more coherent; the FRT provides an interpolation between the incoherent random to coherent random field.

Although the general trend of decreasing MMSE with increasing FRT order is similar for all sampling strategies, the FRT order they saturate and the saturation values are different. As a result we observe that for a given FRT order, the coherence values provided by different measures strongly depend on the measurement strategy the coherence measure is based on.

We observe that for all measures, coherence values act as concave-like functions of the order of the FRT. Moreover, for both scalar channel (S_o) and all but one channel (S_a), the general trend of the error performance is similar; the error saturates at low values of FRT order a, although the saturation values are different. As a result the coherence measures associated with these measurement strategies also show similar behaviour with different saturating values. We also observe that although the exact coherence values provided by c_u, c_b, c_t are possibly different; their general behaviour is quite similar; the range of coherence values reported by these measures and the ways they interpolate in between are

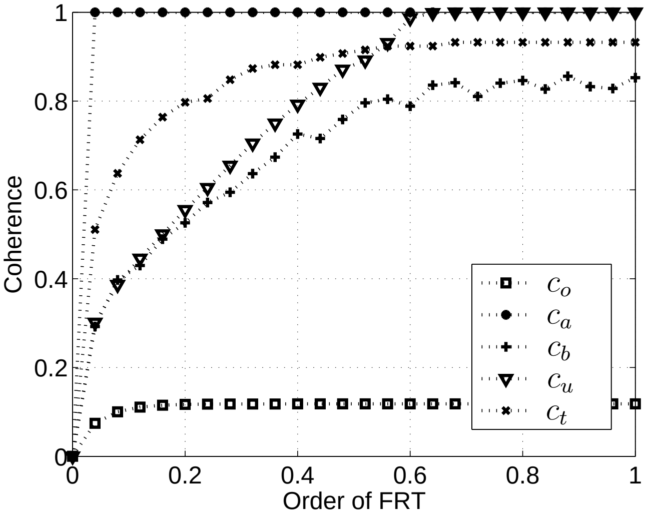


Figure 7.3: Coherence vs. order of the fractional Fourier matrix, $\alpha = 1/16$, coherence measures variable.

not very different.

We observe that the MMSE performance (and associated coherence measures) for Bernoulli sampling and equidistant sampling channel depend significantly on the FRT order (hence coherence of U measured in compressive sensing). For instance for equidistant sampling strategy (S_u), as the FRT order a changes from $a = 0$ to $a = 0.6$, the percentage error changes from $\bar{\epsilon} \approx 100$ to $\bar{\epsilon} \approx 0$. We also note that c_t, which effectively reports the minimum possible success rate of the sampling strategy to achieve a target performance show a similar strong dependency. These strong dependencies on the FRT order motivates the following conclusion: the importance of the coordinate system the measurements done should not be overlooked while quantifying the uncertainty of a signal in estimation framework; in this experiment the eigenvalue distribution of the covariance matrix is fixed, hence the total uncertainty of the associated random field as measured with entropy (or effective DOF) does not change. On the other hand, for some of the measurement strategies very different estimation performances are obtained as U is changed.

Moderate Effective DOF: Let $\alpha = 1/2$, hence the effective DOF $D = \alpha N =$

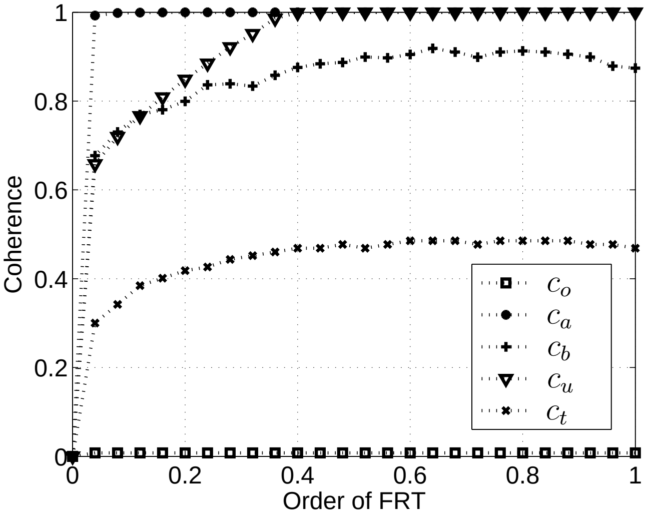


Figure 7.4: Coherence vs. order of the fractional Fourier matrix, $\alpha = 1/2$, coherence measures variable.

$(1/2) \times 128 = 64$. In Fig. 7.4, we plot the associated measures of coherence. We observe that c_o, c_a saturate quite early, and provide very different coherence values. For almost all FRT values, for random scalar channel, coherence is reported as $c_o \approx 0$, and for all but one channel $c_a \approx 1$. On the other hand, similar to the low effective DOF case c_b and c_u behave quite similarly; their ranges and the way they interpolate in between is not much different from each other. We observe that contrary to low DOF case, this time c_t behaves different from these two, and saturates at a significantly different value. These observations are discussed in the experiments for the high effective DOF case, and Section 7.5.

High Effective DOF Let $\alpha = 15/16$, hence the effective DOF $D = \alpha N = (15/16) \times 128 = 120$. In Fig. 7.5, we plot the associated measures of coherence.An intriguing property of this coherence plot is the fact for c_a and c_t there is a strong dependency on the order of the FRT, whereas for the other measures this dependency has vanished. Here c_a shows a strong dependency on the FRT order, and contrary to previous cases (low, and moderate DOF) it does not saturate at very low values of the FRT order; hence distinguishing between different the FRT orders (up to $a \approx 0.6$).

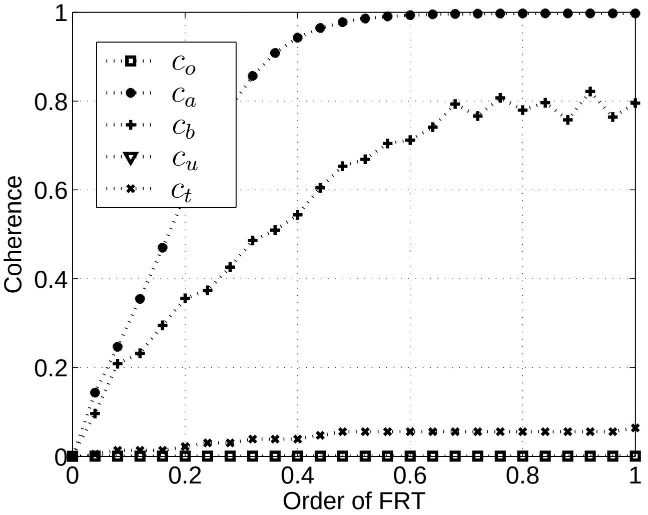


Figure 7.5: Coherence vs. order of the fractional Fourier matrix, $\alpha = 15/16$, coherence measures variable.

Looking at Fig. 7.3, Fig. 7.4 and Fig. 7.5 together, we observe that c_t turns out to be the measure that exhibits the strongest dependency on the order of the FRT, [hence coherence of bases] in the sense that it reports different coherence values for a high range of the FRT orders for a fixed DOF value, and do so for all the DOF levels considered in the experiments, i.e. low, moderate, high.

Effect of Matching of the Eigenvalues and the Columns of the Basis: We now illustrate the effect of matching of the eigenvalues and the columns of the basis on the results. In the previous experiments, the D largest eigenvalues were set to be the first D eigenvalues where the FRT basis vectors are indexed as described in [171] and in [18], which is consistent with the standard representation of the DFT matrix when $a = 1$. We now change the locations of the D largest eigenvalues, and compare the results with the previous case.

Let $\alpha = 1/16$, hence the effective DOF $D = \alpha N = (1/16) \times 128 = 8$. As in the previous cases, the D largest eigenvalues have the value $\frac{\delta P}{D}$ and the remaining has the value $\frac{(1-\delta)P}{N-D}$ with $\delta = 1 - \epsilon$, $\epsilon = 10^{-5}$, $P = N$. Contrary to the previous cases, we choose the locations of the D largest eigenvalues randomly, which are $[26, 35, 55, 64, 81, 88, 90, 119]$ in this experiment. In Fig. 7.6, we plot the MMSE

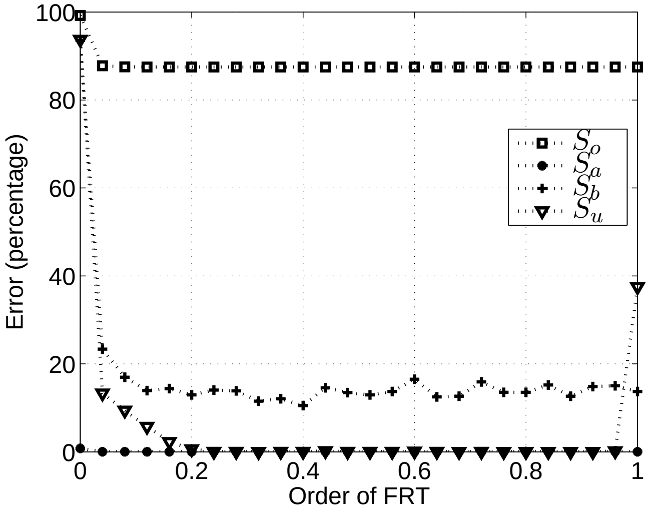


Figure 7.6: Error vs. order of fractional Fourier matrix, $\alpha = 1/16$, random eigenvalue locations, sampling strategy variable.

of varying measurement strategies. We compare these results with Fig. 7.2. We observe that the general behaviour of the MMSE are similar in these two experiments. In particular, the error values associated with random scalar channel and the all but one channel are effectively the same. On the other hand, although the general trend of decreasing error with increasing FRT order is the same with Fig. 7.2, the values that the FRT orders that the error of some of the sample strategies, S_b, S_u, S_t, saturate are quite different. Another interesting point is the unexpected behaviour under S_u when the FRT order is $a = 1$, i.e. the unitary transform is the DFT: although as the FRT order increases, the error in general decreases, when the transform becomes exactly equal to the DFT, the error increases. In the light of all these observations, we conclude that in general for S_b, S_u, S_t the performance of measurement strategies depend on the matching of the eigenvalues and the columns of the basis, determining the unitary transform and the eigenvalue distribution is not sufficient to uniquely determine the error performance. This issue and the relatively robust behaviour of S_o and S_a in this experiment are interesting points to investigate in the future.

Comparison with Measures in Literature: We now investigate the relationship between the FRT order and different coherence definitions presented in [58] in a statistical optics framework. These definitions are based on the following normalized matrix

$$C_x(i,j) = \frac{K_x(i,j)}{\sqrt{K_x(i,i)K_x(j,j)}}, \quad i,j = 1,\dots,N \tag{7.20}$$

assuming $K_x(i,i) > 0$, $K_x(j,j) > 0$. Otherwise if $i = j$, $C_x(i,j) = 1$, if $i \neq j$, $C_x(i,j) = 0$. The diagonal elements of C_x are all 1. We note that this is the matrix of correlation coefficients, where $C_x(i,j)$ is the correlation coefficient between x_i and x_j. Let the eigenvalues of C_x be denoted as $\bar{\lambda}_i$, $i = 1,\dots,N$. We consider the following definitions [58],

$$c_a' = \frac{1}{N}\sum_{i=1}^{N}(\bar{\lambda}_i - \bar{\lambda}_i^0)^2, \tag{7.21}$$

$$c_b' = -\sum_{i=1}^{N}\frac{\bar{\lambda}_i}{N}\log(\frac{\bar{\lambda}_i}{N}), \tag{7.22}$$

where $\bar{\lambda}_i^0 = \frac{\mathrm{tr}(C_x)}{N}$ is the average of the eigenvalues. These definitions are normalized according to (7.10), so that the normalized definitions $c_a, c_b \in [0,1]$. These measures provide an interpolation between the most incoherent and coherent cases, corresponding to 0 and 1 respectively. If one were to apply these definitions to the eigenvalues of the covariance matrix before the normalization, and the above definitions would be independent of the basis U associated with the covariance matrix, and would have been providing alternative ways to quantify the spread of these eigenvalues. As we have discussed earlier, a coherence definition should not be solely depending on eigenvalues. The normalization here makes the measure U dependent.

Let $\alpha = 1/16$. Fig. 7.7 presents the coherence of the field measured by these metrics and the FRT order. We see that with both definitions, the coherence of the field first increase, then decrease as the FRT order increases from 0 to 1. Hence in general, for a given eigenvalue distribution, the FRT provides a poor interpolation between the incoherent and coherent limits (of a random field) according to coherence definitions in [58]. Hence the coherence of the bases measured by

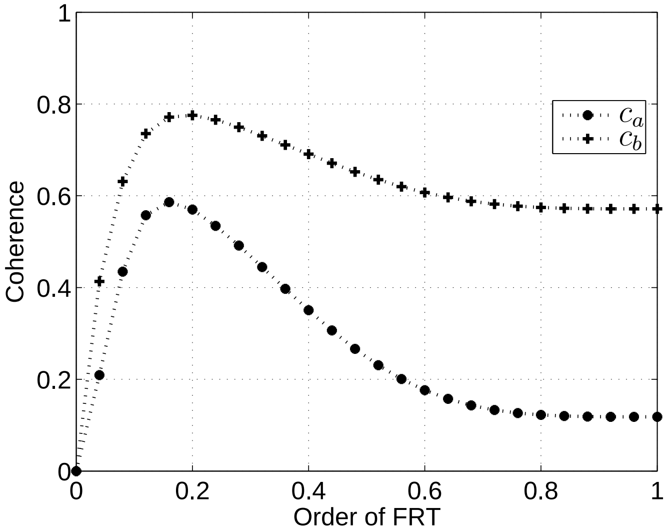


Figure 7.7: Coherence of random field vs. order of the fractional Fourier matrix.

compressive sensing and the coherence of the resulting fields measured by the definitions proposed in [58] are not consistent.

As noted in Section 7.2, one may expect the otherwise. As coherence of the base decreases, we would expect the total correlatedness of the associated field would increase. Although the coherence of the bases shows a general trend of decrease as the order of the FRT increases, the definitions proposed in [58] first increase but then decrease. We note that here the issue is not about the numbers 0 or 1 associated with the extreme points (hence behaviour of increasing/decreasing), but about behaving consistently; constantly showing the same type of behaviour (increasing or decreasing, not both) as the order of the FRT, for example, increases.

Taking a closer look, we make the following observations: The definitions in [58] are designed to wash out the effect of the different power distributions among components of a field; measures are calculated after scaling the power of all components so that all of components have equal power, as shown in (7.20). The definitions quantify the spread of eigenvalues of these normalized matrices. The main motivation behind this normalization is the desire to capture the correlation structure without taking into account the power distributions. For instance for the incoherent case, this idea turns out to be very useful, for all diagonal covariance matrices with nonzero diagonal values, the coherence measures report 0. It is also satisfying to know that whenever the powers of all components are nonzero, the rank of the covariance matrix is preserved by this normalization, since in this case $C = AKA^{\mathrm{T}}$, $A = \mathrm{diag}([1/\sigma_{x_1}, \ldots, 1/\sigma_{x_N}])$, and $\mathrm{rank}\{C\} = \mathrm{rank}\{AKA^{\mathrm{T}}\} = \mathrm{rank}\{K\}$. Hence a rank 1 covariance matrix with nonzero diagonal remains rank 1 after this normalization; coherent fields are reported as coherent as one would desire. On the other hand, in general the relationship between the spread of the eigenvalues of this normalized matrix and the correlation is not clear except these extreme cases. Moreover, this normalization causes some fields associated with different covariance matrices to be reported to have the same coherence value; although whether they should have the same coherence value is questionable. For instance, for a given ρ, the following family of covariance matrices have the same degree of coherence regardless of the value of the constant B

$$K_x(B) = \begin{pmatrix} B & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}. \tag{7.23}$$

On the other hand, these definitions are motivated by statistical optics problems. In statistical optics, coherence is also associated with the concept of fringe visibility, which is a measure of visibility/contrast of interference pattern when two random waves are combined. Most of the time one associates coherence with both of these concepts, fringe visibility and total correlation of the random field, without making the distinction and the relationship between these concepts clear. We have seen that above definitions are not satisfactory for providing a measure for total correlation of the field. Whether they provide a satisfying measure for fringe visibility is an interesting point to investigate in the future.

7.5 Discussion

Coherence of a Random Field and Singular Value Decomposition of the Covariance Matrix:

We now discuss the relationship between coherence of a random field and the s.v.d. of the covariance matrix associated with the field. The s.v.d. of a covariance matrix $\mathbf{K}_x = U\Lambda_x U^\dagger$ has two components: unitary transform U, and the diagonal matrix of the eigenvalues $\Lambda_x = \text{diag}(\lambda_1, \dots, \lambda_N)$. We note that the extreme cases of coherence of the random field are (effectively) solely characterized by different components of this decomposition. Fully incoherent case is solely characterized by the unitary transform; all eigenvalue distributions are allowed as long as the unitary transform is identity (or more generally $U = \text{diag}(U_{ii})$, $|U_{ii}| = 1$). On the other hand, the fully coherent case is (effectively) characterized by the eigenvalue distribution; all unitary transforms (if they satisfy $|U_{ii}| > 0$) are allowed, as long as only one eigenvalue is nonzero.

It is intriguing to investigate what happens in between: how the unitary transform and the eigenvalues interact to determine the total correlatedness of a field. In such a study, it will be educating to study the effect of one while keeping the other constant. We note that under Gaussian assumption, the entropy (hence the uncertainty of the source) is solely characterized by the eigenvalue decomposition. When one fixes the eigenvalue distribution, and varies the unitary transform, one fixes the total uncertainty in the source and only varies how this total uncertainty is spread out through the components of the field. When one fixes the unitary transform, and varies the eigenvalue distribution, the way the uncertainty can be distributed among the components of the signal is fixed, and the total uncertainty varies as dictated by the eigenvalues.

To study these cases, one will need a systematic way to generate eigenvalue distributions representing varying degree of total uncertainty, and a systematic way of generating unitary transforms with varying degrees of power of correlating/mixing the variables it is applied to. It is relatively easier to think of alternative systematic ways to generate such eigenvalue distributions. We also have the advantage of a ground truth: it will be possible to quantify the uncertainty these eigenvalue distributions characterize by notions like entropy or DOF. On the other hand, for the problem of characterizing the unitary transforms according to their correlating power, the situation is not that much clear. In one extreme, we have the identity like transforms ($U = \mathrm{diag}(U_{ii})$, $|U_{ii}| = 1$) which do not correlate the variables it is applied to at all. It is not clear which transform resides in the other extreme, or whether it is eigenvalue distribution specific. Here the idea of coherence of bases used in the compressive sensing provided a possible way to do so. In this chapter we have studied the relationship between this notion, coherence of bases, and the average estimation error of a field which we consider a measure of correlatedness of a field. Still whether notions that provides more suitable characterization of unitary transforms for our purposes is an open problem.

We note that a related issue here is the fact that fixing the eigenvalue distribution and the unitary transform does not fully determine the covariance matrix. The ordering of the eigenvalues and columns of the unitary transform, or in other words, matching of the eigenvalues to the columns of U is also important, and may have an important effect on the total correlatedness of the field.

Coherence and Entropy:

The total uncertainty as measured with entropy in information theory is given by solely the eigenvalue decomposition under our Gaussian assumption. Similarly the bare concept of DOF is characterized by the eigenvalue decomposition. These concepts are designed to quantify the total uncertainty in the signal as number of bits (entropy) or the total number of independent components (DOF). In these frameworks, the assumption is that one can transform the signal before trying to represent it.

On the other hand in understanding the concept of coherence, it is important to also consider the spread of this uncertainty in the basis the signal is observed. As demonstrated before the unitary transform of the covariance matrix is very important. Here we would like to comment on the importance of the eigenvalue distribution. How much total uncertainty can be spread in the measurement domain is closely related to how much uncertainty there is in the signal (eigenvalues). For example, the source has the highest entropy (highest total uncertainty) when all eigenvalues are equal. In that case, no matter what U is, the source is always incoherent resulting in $K_x \propto I$. On the other hand, when the covariance matrix is rank one, this spread of uncertainty can have many forms. All the uncertainty can be in one component, resulting in a matrix for example like $K_x = \text{diag}([1, 0 \dots 0])$. Or the uncertainty may be spread out in all of the components, for example K_x is proportional to matrix of all ones (all entries are one).

7.6 Conclusions

Our work have emphasized a concept that is a measure of dependence, of central importance in statistical optics, but overlooked in signal processing community. We have illustrated that this concept provides a fresh perspective to our understanding of the uncertainty of a signal. Although connected with more traditional concepts like the entropy and the degree of freedom, what this concept quantifies is not just a repetition of what these concepts do. We have also proposed a family of definitions to quantify coherence of random fields in an estimation framework. These definitions are consistent with our qualitative understanding of coherence, and provided a new perspective to our understanding of coherence, hence correlatedness of a field. Through a Gaussian signal model, we have bridged this concept with the concept of coherence in compressive sampling. We have investigated the relationship between these two concepts and proposed the fractional Fourier Transform as a systematic method of generating both bases or statistics for fields with varying degrees of coherence.

Chapter 8

Basis Dependency of MMSE Performance for Random Sampling

In Chapter 7, we have pointed out a possible relationship between the concept of coherence of random fields as defined in optics, and the concept of coherence of bases as defined in compressive sensing, through the fractional Fourier transform. This investigation helped us to explore the relationship between the estimation error and the geometry of the spread of the uncertainty in the measurement domain. In this chapter, we study this relationship from an alternative perspective.

We consider measurement set-ups where a random or a fixed subset of the signal components in the measurement space are erased. We investigate the error performance, both in the average, and also in terms of guarantees that hold with high probability, as a function of system parameters. The unitary transformation that connects the canonical signal domain and the measurement space will play a crucial role throughout this investigation. Contrary to Chapter 7, here we do not restrict the unitary transformation to be a fractional Fourier transform.

We consider the following noisy measurement system

$$y = Hx + n, \tag{8.1}$$

where $x \in \mathbb{C}^N$ is the unknown input proper complex Gaussian random vector, $n \in \mathbb{C}^M$ is the proper complex Gaussian vector denoting the measurement noise, and $y \in \mathbb{C}^M$ is the measurement vector. H is the $M \times N$ measurement matrix.

We assume that x and n are statistically independent zero-mean random vectors with covariance matrices $K_x = E[xx^\dagger]$, and $K_n = E[nn^\dagger]$, respectively. We assume that the components of n are independent and identically distributed (i.i.d.) with $E[n_i n_i \dagger] = \sigma_n^2 > 0$, hence $K_n = \sigma_n^2 I_M \succ 0$, where I_M is the $M \times M$ identity matrix. Let $K_x = U \Lambda_x U^\dagger \succeq 0$ be the singular value decomposition of K_x, where U is a $N \times N$ unitary matrix, and $\Lambda_x = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$. Here $\dagger$ denotes complex conjugate transpose. When needed, we emphasize the random variables the expectations are taken with respect to; we denote the expectation with respect to the random measurement matrix by $E_H[.]$, and the expectation with respect to random signals involved (including x and n) by $E_S[.]$. In all of the problems we assume that the receiver has access to channel realization information.

We now formulate the problems that will be studied in this chapter. Firstly, we will consider equidistant sampling of circularly wide-sense stationary (c.w.s.s.) signals, which may be interpreted as a natural model to represent wide-sense stationary signals in finite dimension.

PROBLEM P1 (Estimation Error of Equidistant Sampling of Circularly Wide-Sense Stationary (c.w.s.s.) Signals): Here the covariance matrix is circulant by assumption, and hence the unitary transform U is fixed and is given by the DFT matrix. We will ask the following questions: "What is the MMSE error associated with equidistant sampling for a c.w.s.s. signal? What is its relationship with the eigenvalue distribution and the rate of sampling?"

This set-up will serve as a benchmark for estimation performance. We will compare the error bounds provided by the high probability results for more general signal models with the error associated with this scheme. We believe that our results here may also be of independent interest, so we state and prove them explicitly.

PROBLEM P2 (Error Bounds For Random Sampling/Support): Here we focus on the case where nonzero eigenvalues all have equal magnitude. Are there any nontrivial lower bounds (i.e. bounds close to 1) on

$$P(E_S[||x - E[x|y]||^2] < f_{P2}(\Lambda_x, U, \sigma_n^2)) \tag{8.2}$$

for some function f_{P2}, where f_{P2} denotes a sufficiently small error level given $\operatorname{tr}(K_x)$, and σ_n^2. In particular, when there is no noise, we will be investigating the probability that the error is zero.

PROBLEM P3 (Error Bounds For Random Projections under a General Eigenvalue Distribution): Let $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$. Are there any nontrivial lower bounds (i.e. bounds close to 1) on

$$P(E_S[||x - E[x|y]||^2] < f_{P3}(\Lambda_x, U, \sigma_n^2)) \tag{8.3}$$

for some function f_{P3} under the scenario of sampling with random projections (entries of H are i.i.d. Gaussian) with fixed eigenvalue distribution? How does the Λ_x and H affect the performance? Here f_{P3} denotes a sufficiently small error level given $\operatorname{tr}(K_x)$ and σ_n^2.

In our investigations, we will see that among the unitary matrices, the DFT matrix (or other unitary matrices satisfying $u_{ij} = 1$, $i, j = 1, \dots, N$) will provide the best performance guarantees, in the sense that fixing the bound on the probability of error, they will require the least number of measurements to have certain error bounds or fixing the bound on the probability of error, it will be possible to obtain tighter error bounds with a given number of measurements. We note that in all these results the performance criterion is of the type "performance guarantees that hold with high probability", but not average, with respect to the random sampling matrix H. (MMSE is of course an average, but it is an average over signals, i.e. the result of expectation of the type $E_S(.)$.) An intriguing question is to investigate the average performance over H. We pay particular attention to the case where U is given by the DFT matrix, since the best guarantees in the previous high probability results are obtained for this matrix.

PROBLEM P4 (Best Unitary Encoder For Random Channels): Let $\mathbb{U}^{\mathbb{N}}$ be the set of $N \times N$ unitary matrices: $\{U \in \mathbb{C}^N : U^\dagger U = I_N\}$. We consider the following minimization problem

$$\inf_{U \in \mathbb{U}^N} E_{H,S}[||x - E[x|y]||^2], \tag{8.4}$$

where the expectation with respect to H is over admissible random measurement strategies: random scalar Gaussian channel (only one of the components is measured each time) or Gaussian erasure channel (each component of the unknown vector is erased independently and with equal probability).

We note that in the context of Problem 3 it is not meaningful to seek for the best orthonormal U (i.e. $U \in \mathbb{R}^{N \times N} : U^\dagger U = I_N$) encoder. This is because the entries of H are i.i.d. Gaussian, and such a random matrix H is left and right 'rotationally invariant': For any orthonormal matrix U, the random matrices UH, HU and H have the same distribution. See [Lemma 5, [180]].

We note that the dependence of signal uncertainty in the signal basis has been considered in different contexts in the information theory literature. The concepts that are traditionally used in the information theory literature as measures of dependency or uncertainty in signals (such as the degree of freedom, or the entropy) are mostly defined independent of the coordinate system in which the signal is to be measured. As an example one may consider the Gaussian case: the entropy solely depends on the eigenvalue spectrum of the covariance matrix, hence making the concept blind to the coordinate system in which the signal lies in. On the other hand, the approach of applying coordinate transformations to orthogonalize signal components takes place in many signal reconstruction and information theory problems. For example the rate-distortion function for a Gaussian random vector is obtained by applying an uncorrelating transform to the source, or approaches such as the Karhunen-Loéve expansion are used extensively. Also, the compressive sensing community heavily makes use of the notion of coherence of bases, see for example [13,14,17]. The coherence of two bases, say the intrinsic signal domain ψ, and the orthogonal measurement system ϕ is measured with $\mu = \max_{i,j} |u_{ij}|$, $U = \phi\psi$ providing a measure of how concentrated the columns of U are. When μ is small, one says the mutual coherence is small. As the coherence gets smaller, fewer samples are required to provide good performance guarantees.

The total uncertainty in the signal as quantified by information theoretic measures such as entropy (or eigenvalues) and the spread of this uncertainty (basis) reflect different aspects of the dependence in a signal. The estimation problems we will consider may be seen as an investigation of the relationship between the MMSE and these two measures.

In the following, we provide a brief overview of the related literature. An important model in this chapter is the Gaussian erasure channel, where each component of the unknown vector is erased independently and with equal probability, and the transmitted components are observed through Gaussian noise. This type of model may be used to formulate various types of transmission with low reliability scenarios, for example Gaussian channel with impulsive noise [189, 190]. This measurement model is also related to the measurement model considered in the compressive sensing framework, where the measurement scenario where each component is erased independently and with equal probability is of central importance [185, 186]. Our work also contributes to the understanding of the MMSE performance of such measurement schemes under noise.

The problem of optimization of precoders or input covariance matrices is formulated in literature under different performance criteria: When the channel is not random, [191] considers a related trace minimization problem, and [192] a determinant maximization problem, which correspond to optimization of the MMSE and mutual information performance respectively in our formulation. [193], [194] formulates the problem with the criterion of mutual information, whereas [195] focuses on the MMSE, and [196] on determinant of the mean-square error matrix. [197, 198] presents a general framework based on Schur-convexity. In these works the channel is known at the transmitter, hence it is possible to shape the input according to the channel. When the channel is a Rayleigh or Rician fading channel, [199] investigates the best linear encoding problem without restricting the encoder to be unitary. [180] focuses on the problem of maximizing the mutual information for a Rayleigh fading channel. [189], [190] consider the erasure channel as in our setting, but with the aim of maximizing the ergodic capacity.

In Problems P2 and P3, we investigate how the results in random matrix theory mostly presented in compressive sampling framework can be used to find bounds on the MMSE associated with the described measurement scenarios. We note that there are studies that consider the MMSE in compressive sensing framework such as [200,201], which focus on the scenario where receiver does not know the location of the signal support. In our case we assume that the receiver has full knowledge of the signal covariance matrix.

Preliminaries and Notation: In the following, we present a few definitions and notations that will be used throughout the chapter. Let $\mathrm{tr}\,(K_x) = P$. Let $D(\delta)$ be the smallest number satisfying $\sum_{i=1}^{D} \lambda_i \geq \delta P$, where $\delta \in (0,1]$. Hence for δ close to one, $D(\delta)$ can be considered as an effective rank of the covariance matrix and also the effective number of "degrees of freedom" (DOF) of the signal family. For δ close to one, we drop the dependence on δ and use the term effective DOF to represent $D(\delta)$. A closely related concept is the (effective) bandwidth. We use the term "bandwidth" for the DOF of a signal whose canonical domain is the Fourier domain, i.e. whose unitary transform is given by the discrete Fourier Transform (DFT) matrix.

Let $\sqrt{-1} = j$. The entries of an $N \times N$ DFT matrix are given by $u_{tk} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}tk}$, where $0 \leq t\,, k \leq N-1$. We note that the DFT matrix is the diagonalizing unitary transform for all circulant matrices [202]. In general, a circulant matrix is determined by its first row and defined by the relationship $C_{tk} = C_{0\,\mathrm{mod}_N(k-t)}$, where rows and columns are indexed by t and k, $0 \leq t\,, k \leq N-1$, respectively.

The transpose, complex conjugate and complex conjugate transpose of a matrix A is denoted by A^{T}, A^* and $A^\dagger$, respectively. The eigenvalues of a matrix A are denoted in decreasing order as $\lambda_1(A) \geq \lambda_2(A), \ldots, \geq \lambda_N(A)$.

We now review the expressions for MMSE estimation. Under a given measurement matrix H, by standard arguments the MMSE estimate is given by $E[x|y] = \hat{x} = K_{xy}{K_y}^{-1}y$, where $K_{xy} = E[xy^\dagger] = K_x H^\dagger$, and $K_y = E[yy^\dagger] = HK_xH^\dagger + K_n$.

We note that since $K_n \succ 0$, we have $K_y \succ 0$, and hence K_y^{-1} exists. The associated MMSE can be expressed as [188, Ch2]

$$\begin{aligned} E_S[||x - E[x|y]||^2] &= \mathrm{tr}(K_x - K_{xy}K_y^{-1}K_{xy}^\dagger) && (8.5)\\ &= \mathrm{tr}(K_x - K_x H^\dagger(HK_xH^\dagger + K_n)^{-1}HK_x) && (8.6)\\ &= \mathrm{tr}(U\Lambda_x U^\dagger - U\Lambda_x U^\dagger H^\dagger(HU\Lambda_x U^\dagger H^\dagger + K_n)^{-1}HU\Lambda_x U^\dagger) && (8.7) \end{aligned}$$

Let $B = \{i : \lambda_i > 0\}$, and let U_B denote the $N \times |B|$ matrix formed by taking the columns of U indexed by B. Similarly, let $\Lambda_{x,B}$ denote the $|B| \times |B|$ matrix by taking the columns and rows of Λ_x indexed by B in the respective order. We note that $U_B^\dagger U_B = I_{|B|}$, whereas the equality $U_B U_B^\dagger = I_N$ is not true unless $|B| = N$. Also note that $\Lambda_{x,B}$ is always invertible. The singular value decomposition of K_x can be written as $K_x = U\Lambda_x U^\dagger = U_B\Lambda_{x,B}U_B^\dagger$. Hence the error may be rewritten as

$$\begin{aligned} &E_S[||x - E[x|y]||^2]\\ &= \mathrm{tr}(U_B\Lambda_{x,B}U_B^\dagger - U_B\Lambda_{x,B}U_B^\dagger H^\dagger(HU_B\Lambda_{x,B}U_B^\dagger H^\dagger + K_n)^{-1}HU_B\Lambda_{x,B}U_B^\dagger) && (8.8)\\ &= \mathrm{tr}(\Lambda_{x,B} - \Lambda_{x,B}U_B^\dagger H^\dagger(HU_B\Lambda_{x,B}U_B^\dagger H^\dagger + K_n)^{-1}HU_B\Lambda_{x,B}) && (8.9)\\ &= \mathrm{tr}\left((\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2}U_B^\dagger H^\dagger HU_B)^{-1}\right) && (8.10) \end{aligned}$$

where (8.9) follows from the identity $\mathrm{tr}(U_B M U_B^\dagger) = \mathrm{tr}(M U_B^\dagger U_B) = \mathrm{tr}(M)$ with an arbitrary matrix M with consistent dimensions. Here (8.10) follows from the fact that $\Lambda_{x,B}$ and K_n are nonsingular and the Sheerman-Morrison-Woodbury identity, which has the following form for our case (see for example [203] and the references therein)

$$K_1 - K_1A^\dagger(AK_1A^\dagger + K_2)^{-1}AK_1 \;=\; (K_1^{-1} + A^\dagger K_2^{-1}A)^{-1}, \qquad (8.11)$$

where K_1 and K_2 are nonsingular.

Here is a brief summary of the rest of the chapter: In Section 8.1, we consider equidistant sampling of a circularly wide-sense stationary signal. We give the explicit expression for the MMSE, and show that two times the total power outside a properly chosen set of indices (a set of indices which do not overlap when shifted by an amount determined by the sampling rate) provides an upper bound for the MMSE. In Section 8.2, we illustrate how some recent results in matrix theory mostly presented in the compressive sampling framework can be used to find performance guarantees for the MMSE estimation that hold with high probability. In Section 8.3, we illustrate how the spread of the eigenvalue distribution and the measurement scheme contribute to obtain performance guarantees that hold with high probability for the case of sampling matrix with i.i.d. Gaussian entries. In Section 8.4, we consider random erasure channels and formulate the problem of finding the most favorable unitary transform under average performance. We investigate the convexity properties of this optimization problem, and obtain conditions of optimality through variational equalities. We identify special cases where the discrete Fourier Transform (DFT)-like unitary transforms turn out to be the best coordinate transforms (possibly along with other unitary transforms). Although there are many observations (including evidence provided by the compressed sensing community) that may suggest the idea that the DFT matrix may be indeed an optimum unitary matrix for any eigenvalue distribution, we provide a counterexample. We conclude in Section 8.5.

8.1 Equidistant Sampling of Circularly Wide-Sense Stationary Random Vectors

We now consider the MMSE associated with equidistant sampling of an important class of signals: circularly wide-sense stationary (c.w.s.s.) signals, which is a way for modelling wide-sense stationary signals in finite dimension. Let $x = [x_t, t \in I = 0, \dots, N-1]$ be a zero-mean, proper, c.w.s.s. Gaussian random vector. We note that the covariance matrix of a c.w.s.s. signal is always circulant, so the eigenvectors of the covariance matrix is given by the columns of the DFT matrix $u_{tk} = \frac{1}{\sqrt{N}} e^{j \frac{2\pi}{N} tk}$, where $0 \leq t\,, k \leq N-1$ [202]. Hence in this section we fix the unitary transform to be the DFT matrix. We denote the associated eigenvalues with λ_k, $0 \leq k \leq N-1$ instead of indexing the eigenvalues in decreasing/increasing order.

In this section, we first consider the noiseless deterministic sampling strategy where every 1 out of ΔN samples are taken. We let $M = \frac{N}{\Delta N} \in \mathbb{Z}$, and assume that the first component is always measured, for convenience. Hence our measurements are in the form

$$y = Hx, \tag{8.12}$$

where $H \in \mathbb{R}^{\mathbb{M}\times\mathbb{N}}$ is the sampling matrix formed by taking the rows of the identity matrix corresponding to the observed variables.

We now present our main result in this section; an explicit expression and an upper bound for the mean-square error associated with the above set-up.

Lemma 8.1.1. *Let the model and the sampling strategy be as described above. Then the MMSE of estimating x from these equidistant samples can be expressed as*

$$E[||x - E[x|y]||^2] \quad = \quad \sum_{k\in J_0}(\sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1}\lambda_{lM+k}}), \tag{8.13}$$

where $J_0 = \{k : \sum_{l=0}^{\Delta N-1} \lambda_{lM+k} \neq 0, 0 \leq k \leq M-1\} \subseteq \{0,\dots,M-1\}$.

In particular, choose a set of indices $J \subseteq \{0,1,\dots,N-1\}$ with $|J| = M$ such that

$$jM + k \in J \Rightarrow iM + k \notin J \qquad \forall i,j,\;\; 0 \leq i,j \leq \Delta N - 1, i \neq j \tag{8.14}$$

with $0 \leq k \leq M-1$. Let $P_J = \sum_{i\in J}\lambda_i$. Then the MMSE is upper bounded by the total power in the remaining eigenvalues

$$E[||x - E[x|y]||^2] \quad \leq \quad 2(P - P_J). \tag{8.15}$$

In particular, if there is such a set J so that $P_J = P$, the MMSE will be zero.

Remark 8.1.1. *The set J essentially consists of the indices which do not overlap when shifted by M.*

Remark 8.1.2. *We note that the choice of the set J is not unique, and each choice of the set of indices may provide a different upper bound. To obtain the* *lowest possible upper bound, one should consider the set with the largest total power.*

Remark 8.1.3. *If there exists such a set J that has the most of power, i.e. $P_J = \delta P$, $\delta \in (0, 1]$, with δ close to 1, then $2(P - P_J) = 2(1 - \delta)P$ is small and the signal can be estimated with low values of error. In particular, if such a set has all the power, i.e. $P = P_J$, the error will be zero. A conventional aliasing free set J may be the set of indices of the band of a band-pass signal with band smaller than M. It is important to note that there may exist other sets J with $P = P_J$, hence the signal may be aliasing free even if the signal is not bandlimited (low-pass, high-pass etc) in the conventional sense.*

Proof: Proof is given in Section A.1 of the Appendix.

We observe that the bandwidth W (or the DOF) turn out to be good predictors of estimation error for this case. On the other hand, the differential entropy of an effectively W-bandlimited Gaussian vector can be very small even if the bandwidth is close to N, hence may not provide any useful information with regards to estimation performance.

We also give the explicit error expression for the noisy case. Here the observations are in the following form

$$y = Hx + n, \tag{8.16}$$

where x and n are statistically independent random vectors, and the components of n are i.i.d. zero mean with $E[n_i n_i\dagger] = \sigma_n^2 > 0$, hence $K_n = \sigma_n^2 I_N \succ 0$, where I_M is the $M \times M$ identity matrix.

Lemma 8.1.2. *The MMSE of estimating x from the equidistant noisy samples as described above is given by the following expression*

$$E[||x - E[x|y]||^2] = \sum_{k=0}^{M-1} \Big(\sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1}(\lambda_{lM+k} + \sigma_n^2)} \Big) \tag{8.17}$$

Proof: We first note that here $K_{xy} = K_x H^\dagger$, as in the noiseless case. We also note that here, K_y is given by $K_y = HK_xH^\dagger + K_n$. Now the result is obtained by retracing the steps of the proof of Lemma 8.1.1, which is given in Section A.1, with K_y replaced by the above expression, that is $K_y = HK_xH^\dagger + K_n$.

A particularly important special case is the error associated with the estimation of a band-pass signal:

Corollary 8.1.1. *Let* $\mathrm{tr}(K_x) = P$*. Let the eigenvalues be given as* $\lambda_i = \frac{P}{D}$*, if* $0 \leq i \leq D-1$*, and* $\lambda_i = 0$*, if* $D \leq i \leq N-1$*. If* $M \geq D$*, then the error can be expressed as follows*

$$E[||x - E[x|y]||^2] = \frac{1}{1 + \frac{1}{\sigma_n^2}\frac{P}{D}\frac{M}{N}} P \tag{8.18}$$

We note that this expression is of the form $\frac{1}{1+\mathrm{SNR}}P$, where $\mathrm{SNR} = \frac{1}{\sigma_n^2}\frac{P}{D}\frac{M}{N}$. This expression will serve as a benchmark in the subsequent sections.

We now compare our error expression with the following results where the signals defined on $\mathbb{R}$ are considered: In [122], mean-square error of approximating a possibly non-bandlimited wide-sense stationary (w.s.s.) signal using sampling expansion is considered and a uniform upper bound in terms of power outside the bandwidth of approximation is derived. Here we are interested in the average error over all points of the N dimensional vector. Our method of approximation of the signal is possibly different, since we use the MMSE estimator. As a result our bound also makes use of the shape of the eigenvalue distribution. [116] states that a w.s.s. signal is determined linearly by its samples if some set of frequencies containing all of the power of the process is disjoint from each of its translates where the amount of translate is determined by the sampling rate. Here for circularly w.s.s. we show a similar result: if there is a set J that consists of indices which do not overlap when shifted by M, and has all the power, the error will be zero. In fact, we show a more general result for our set-up: we give the explicit error expression and show that two times the power outside this set J provides an upper bound for the error, hence putting a bound on the error even if it is not exactly zero.

8.2 Random Sampling/Support at a Fixed Measurement Domain - Error Bounds That Hold with High Probability

In this section we focus on MMSE bounds that hold with high probability. We assume that nonzero eigenvalues are equal, i.e. $\Lambda_{x,B} = \frac{P}{|B|} I_{|B|}$, where $|B| \leq N$. We are interested in the MMSE estimation performance of two set-ups: i) sampling of a signal with fixed support at randomly chosen measurement locations; ii) sampling of a signal with random support at fixed measurement locations. We investigate bounds on the MMSE depending on the support size or the number of measurements. We illustrate how the results in matrix theory mostly presented in compressive sampling framework can provide error bounds for these scenarios. We note that there are studies that consider the MMSE in compressive sensing framework such as [200, 201], which focus on the scenario where receiver does not know the location of the signal support. In our case we assume that the receiver has full knowledge of signal covariance matrix.

We again consider the set-up in (8.1). The sampling operation can be modelled with a $M \times N$ $\mathbf{H}$ matrix, whose rows are taken from the identity matrix as dictated by the sampling operation. We let $U_{MB} = HU_B$ be the $M \times |B|$ submatrix of U formed by taking $|B|$ columns and M rows as dictated by B and H, respectively. The MMSE can be written as (8.10)

$$\begin{aligned} E[||x - E[x|y]||^2] &= \operatorname{tr}\left((\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H^\dagger H U_B)^{-1}\right) && (8.19) \\ &= \sum_{i=1}^{|B|} \frac{1}{\lambda_i(\frac{|B|}{P} I_B + \frac{1}{\sigma_n^2} U_{MB}^\dagger U_{MB})} && (8.20) \\ &= \sum_{i=1}^{|B|} \frac{1}{\frac{|B|}{P} + \frac{1}{\sigma_n^2} \lambda_i(U_{MB}{}^\dagger U_{MB})}. && (8.21) \end{aligned}$$

We see that the estimation error is determined by the eigenvalues of the matrix $U_{MB}^\dagger U_{MB}$. We note that many results in compressive sampling framework make use of the bounds on the eigenvalues of this matrix. We now use some of these results to bound the MMSE performance in different sampling scenarios. We note that different bounds found in the literature can be used, we pick some of the bounds from the literature to make the constants explicit.

Lemma 8.2.1. *Let U be an $N \times N$ unitary matrix with $\sqrt{N}\max_{k,j}|u_{k,j}| = \mu(U)$. Let the signal have fixed support B on the signal domain. Let the sampling locations be chosen uniformly at random from the set of all subsets of the given size M. Let noisy measurements with noise power σ_n^2 be done at these M locations. Then for sufficiently large $M(\mu)$, the error is bounded from above with high probability:*

$$\varepsilon < \frac{1}{1 + \frac{1}{\sigma_n^2}\frac{0.5M}{N}\frac{P}{|B|}} P \tag{8.22}$$

More precisely, if

$$M \geq |B|\mu^2(U)\max(C_1 \log|B|, C_2 \log(3/\delta)) \tag{8.23}$$

for some positive constants C_1 and C_2, then

$$P(\varepsilon \geq \frac{1}{1 + \frac{1}{\sigma_n^2}\frac{0.5M}{N}\frac{P}{|B|}} P) \leq \delta. \tag{8.24}$$

In particular, when the measurements are noiseless, the error is zero with probability at least $1 - \delta$.

Proof: We first note that $\|{U_{MB}}^\dagger U_{MB} - I\| < c$ implies $1 - c < \lambda_i({U_{MB}}^\dagger U_{MB}) < 1 + c$. Consider Theorem 1.2 of [13]. Suppose that M and $|B|$ satisfies (8.23). Now looking at Theorem 1.2, and noting the scaling of the matrix $U^\dagger U = NI$ in [13] , we see that $P(0.5\frac{M}{N} < \lambda_i({U_{MB}}^\dagger U_{MB}) < 1.5\frac{M}{N}) \geq 1-\delta$. By (8.21) the result follows.

For the noiseless measurements case, let $A_{\sigma_n^2}$ be the event $\{\varepsilon < \sigma_n^2 \frac{|B|}{\sigma_n^2\frac{|B|}{P} + \frac{0.5M}{N}}\}$ Hence

$$\begin{aligned} \lim_{\sigma_n^2 \to 0} P(A_{\sigma_n^2}) &= \lim_{\sigma_n^2 \to 0} E[1_{A_{\sigma_n^2}}] && (8.25) \\ &= E[\lim_{\sigma_n^2 \to 0} 1_{A_{\sigma_n^2}}] && (8.26) \\ &= P(\varepsilon = 0) && (8.27) \end{aligned}$$

where we have used Dominated Convergence Theorem to change the order of the expectation and the limit. By (8.24) $P(A_{\sigma_n^2}) \geq 1-\delta$, hence $P(\varepsilon = 0) \geq 1-\delta$. We also note that in the noiseless case, it is enough to have $\lambda_{\min}(U_{MB}^{\dagger}U_{MB})$ bounded away from zero to have zero error with high probability, the exact value of the bound is not important.

We note that when other parameters are fixed, as $\max_{k,j}|u_{k,j}|$ gets smaller, fewer number of samples are required. Since $\sqrt{1/N} \leq \max_{k,j}|u_{k,j}| \leq 1$, the unitary transforms that provide the best guarantees are the ones satisfying $|u_{k,j}| = \sqrt{1/N}$, $k, j = 1, \ldots, N$. We note that for any such unitary transform, the covariance matrix has constant diagonal with $(K_x)_{ii} = P/N$ regardless of the eigenvalue distribution. Hence with any measurement scheme with M noiseless measurements, the reduction in the uncertainty is guaranteed to be at least proportional to the number of measurements, i.e. the error satisfies $\varepsilon \leq P - \frac{M}{N}P$.

We would like to recall that the unitary transform associated with c.w.s.s. signals is the DFT matrix, which satisfies the condition $|u_{k,j}| = \sqrt{1/N}$. Hence Lemma 8.2.1 is also applicable to these signals. Hence among signals with a covariance matrix with a given rectangular eigenvalue spread, c.w.s.s. signals are among the ones that can be estimated with low values of error with high probability with a given number of randomly located measurements.

We now consider a signal sampled at fixed locations with random support uniformly chosen from the set of supports with a given size. We note that in this case the results, such as Theorem 12 of [17] or Theorem 2 of [204] (and the references therein) that explores the bounds on the eigenvalues of random submatrices obtained by uniform column sampling can be used for bounding the estimation error. We assume that the receiver has access to the support set information. In the following we assume the field is real, i.e. $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$. The s.v.d. of K_x is given as $K_x = U\Lambda_x U^{\dagger}$, where U is orthonormal, i.e. $U \in \mathbb{R}^{N\times N}$, $U^{\dagger}U = I_N$. We note that normalized Hadamard matrices satisfy $|u_{i,j}|^2 = \frac{1}{N}$ and orthonormal as required in the lemma. For the proper complex Gaussian case the argument is similar, and Theorem 12 of [17] can be used.

Lemma 8.2.2. *Let U be a $N \times N$ orthonormal matrix such that $|u_{i,j}|^2 = \frac{1}{N}$. Let the M locations at the measurement domain be fixed, and let H be the $M \times N$*

diagonal matrix. Let μ be defined by

$$\mu = \frac{N}{M} \max_{j \neq k} |(HU)_j^\dagger (HU)_k|, \tag{8.28}$$

where $(HU)_j$ denotes the j^{th} column of HU. Let the support of the signal be chosen uniformly from the set of all subsets of the given size $|B| \leq N$. Then for sufficiently small $|B|$, the error is bounded from above with high probability

$$\varepsilon < \frac{1}{1 + (1-r)\frac{1}{\sigma_n^2}\frac{M}{N}\frac{P}{|B|}} P \tag{8.29}$$

where $r \in (0,1)$. More precisely, let $\alpha \geq 1$, and assume that $\mu \leq r/(2(1+\alpha)\log N)$ and $|B| \leq (\frac{r^2}{4(1+\alpha)\exp(1)^2})(\frac{N}{(N/M)||HU||^2 \log N})$. Then

$$P(\varepsilon \geq \frac{1}{1 + (1-r)\frac{1}{\sigma_n^2}\frac{M}{N}\frac{P}{|B|}} P) \leq 216 N^{-\alpha} \tag{8.30}$$

In particular, when the measurements are noiseless, the error is zero with probability at least $1 - 216N^{-\alpha}$.

Proof: We note that $X = \sqrt{N/M} HU$ has unit norm columns and μ given in (8.28) is the coherence of X as defined by equation [1.3] of [204]. We also note that HU is full rank, that is rank of HU is equal to largest possible value i.e. M, since U is orthogonal. We also note that $||X|| = ||\sqrt{N/M} HU|| = \sqrt{N/M}||HU||$. Hence we can use Theorem 3.1 of [204] to bound the singular values of $\sqrt{N/M} HU_B$. As in the proof of the previous lemma, the result follows from (8.21). The noiseless case follows similar to the previous lemma. Again it it is enough to have $\lambda_{\min}(U_{MB}^\dagger U_{MB})$ bounded away from zero to have zero error with high probability. □

We note that the conclusions derived in this section are based on high probability results for the norm of a matrix restricted to random set of coordinates. We note that for the purposes of such results, the uniform random sampling model and the Bernoulli sampling model where each component is taken independently and with equal probability is equivalent [185–187]. For instance, the derivation of Theorem 1.2 of [13], the main step of Lemma 8.2.1, is in fact based on a Bernoulli sampling model. Hence the high probability results presented there also hold for Gaussian erasure channel of Section 8.4 (with possibly different parameters).

We now compare these error bounds found in this section with the error associated with equidistant sampling of a low pass circularly wide-sense stationary (c.w.s.s.) source. The equidistant sampling of a general c.w.s.s. source is studied in Section 8.1. Let us consider the special case where x is a band pass signal with $\lambda_0 = \cdots = \lambda_{|B|-1} = P/|B|$, $\lambda_{|B|} = \ldots = \lambda_{N-1} = 0$. If $M \geq |B|$, the error associated with this scheme can be expressed as follows (8.13):

$$E[||x - E[x|y]||^2] = \frac{1}{1 + \frac{P}{|B|}\frac{1}{\sigma_n^2}\frac{M}{N}} P. \tag{8.31}$$

Comparing (8.22) and (8.29), with this expression, we observe the following: All of these expressions are of the same general form, $\frac{1}{1+c\,\mathrm{SNR}}P$, where $\mathrm{SNR} \triangleq \frac{P}{|B|}\frac{1}{\sigma_n^2}\frac{M}{N}$. Here $0 \leq c \leq 1$ takes different values for different cases. We also note that in (8.22), the choice of $c = 0.5$, which is the constant chosen for the eigenvalue bounds in [13], is for convenience. It could have been chosen differently by choosing a different probability δ in (8.24), similar to the parameterization through r in [204], which is seen here in (8.30) and the conditions there. We also observe that SNR takes its maximum value with $c = 1$ for the deterministic equidistant sampling strategy corresponding to the minimum error value among these expressions. In the other cases c takes possibly smaller values, resulting in larger error expressions. One can choose larger c values in these expressions, but then the probability these error bounds hold decreases, that is better error bounds can be obtained at the expense of lower degrees of guarantees that these results will hold.

8.3 Random Projections - Error Bounds That Hold With High Probability

In this section we consider the measurement strategy where M random projections of the signal are taken, the measurement system matrix H is a $M \times N$,

$M \leq N$ matrix with Gaussian i.i.d. entries. In this section we assume that the field is real. We also assume that Λ_x is positive-definite.

We note that the matrix theory result used in this section is novel, and provides fundamental insights into problem of estimation of signals with small effective number of degrees of freedom. In the previous section we have used some results in compressive sensing literature that are directly applicable only when the signals are known to be exactly sparse (some of the eigenvalues of K_x are exactly equal to zero.) In this section we assume a more general eigenvalue distribution. Our result enables us draw conclusions when some of the eigenvalues are not exactly zero, but small. The method of proof provides us a way to see the effects of the effective number of degree of freedom of the signal (Λ_x) and the incoherence of measurement domain (HU), separately.

Before stating our result, we now make some observations on the related results in random matrix theory. Consider the submatrices formed by restricting a matrix K to random set of its rows, or columns; $R_1 K$ or $K R_2$ where R_1 and R_2 denote the restrictions to rows and columns respectively. The main tool for finding bounds on the eigenvalues of these submatrices is finding a bound on $E||R_1 K - E[R_1 K]||$ or $E||K R_2^\dagger - E[K R_2^\dagger]||$ [17, 204, 205]. We have found this approach unsuitable to our problem in which the matrix we are investigating $\Lambda_x^{-1} + (HU)^\dagger (HU)$ constitutes of two matrices: a deterministic diagonal matrix with possibly different entries on the diagonal and a random restriction. Hence we adopt another method: the approach of decomposing the unit sphere into compressible and incompressible vectors as proposed by M. Rudelson and R. Vershynin [206].

We note that when the eigenvalues of K_x have rectangular spread (the signal is exactly sparse), using the method in Lemma 8.2.1 and for example using Proposition 2.5 of [206], (which is due to [207]), one can prove that it is possible to achieve low values of MMSE with high probability also for random projections. Here we focus on the case where $\Lambda_x \succ 0$ to see the effects of other eigenvalue spreads. We also note that the general methodology in this section can be extended to the case where H has complex entries. In this case the channel will be a Rayleigh fading channel.

We consider the general measurement set-up in (8.1) where $y = Hx + n$, with $K_n = \sigma_n^2 I$, $K_x \succ 0$, and assume the field is real, i.e. $x \in \mathbb{R}^N$ and $n \in \mathbb{R}^M$. The s.v.d. of K_x is given as $K_x = U\Lambda_x U^\dagger$, where $U \in \mathbb{R}^{N\times N}$ is orthonormal and $\Lambda = \mathrm{diag}(\lambda_i)$ with $\sum_i \lambda_i = P$, $\lambda_1 \geq \lambda_2, \dots, \geq \lambda_N$.

Theorem 8.3.1. *Let H be a $M \times N$, $M \leq N$, $M = \beta N$ matrix with Gaussian i.i.d. entries with variances σ_H^2 at least 1. Let $D(\delta)$ be the smallest number satisfying $\sum_{i=1}^{D} \lambda_i \geq \delta P$, where $\delta \in (0, 1]$. Assume that $D(\delta) + M \leq N$, and $\lambda_i < C_\lambda \frac{P}{N}$, $i = D+1, \dots, N$. Then there exist C, C_1, T, T_1 that depend on $\frac{P}{\sigma_n^2}$, σ_H^2, C_λ, β such that if $D(\delta) < T$, and $M > T_1$ the error will satisfy*

$$P(E[||x - E[x|y]||^2] \geq (1-\delta)P + \frac{1}{C}\frac{D}{N}P) \leq e^{-C_1 N} \tag{8.32}$$

Remark 8.3.1. *As we will see in the proof, eigenvalue distribution plays a key role in obtaining stronger bounds: In particular, when the eigenvalue distribution is spread out, the theorem cannot provide bounds for low values of error. As the distribution becomes less spread out, stronger bounds are obtained. We discuss this point in Remark A.2.1, Remark A.2.2, and Remark A.2.3. Effect of noise level is discussed in Remark A.2.4. A special case of problem studied at the end of Section A.2 of the Appendix illustrates these points.*

Proof: Let the eigenvalues of a matrix A be denoted in decreasing order as $\lambda_1(A) \geq \lambda_2(A), \dots, \geq \lambda_N(A)$. We note that by [Lemma 5 , [180]], H and HU have the same probability distribution. Hence we can consider H instead of HU in our arguments. The error can be expressed as follows (8.10)

$$E[||x - E[x|y]||^2]$$

$$= \mathrm{tr}\,((\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)^{-1}) \tag{8.33}$$

$$= \sum_{i=1}^{N} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} \tag{8.34}$$

$$= \sum_{i=1}^{N-D} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} + \sum_{i=N-D+1}^{N} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} \tag{8.35}$$

$$\leq \sum_{i=1}^{N-D} \frac{1}{\lambda_i(\Lambda_x^{-1})} + \sum_{i=N-D+1}^{N} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} \tag{8.36}$$

$$\leq \sum_{i=1}^{N-D} \lambda_{N-i+1}(\Lambda_x) + D \frac{1}{\lambda_{min}(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} \tag{8.37}$$

$$= \sum_{i=D+1}^{N} \lambda_i(\Lambda_x) + D \frac{1}{\lambda_{min}(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} \tag{8.38}$$

where the first inequality follows from case (a) of the following result.

Lemma 8.3.1. *[4.3.3, 4.3.6, [208]] Let $A_1, A_2 \in \mathbb{C}^{N\times N}$ be Hermitian matrices. (a) Let A_2 be positive semi-definite. Then $\lambda_i(A_1 + A_2) \geq \lambda_i(A_1)$, $i = 1, \dots, N$. (b) Let rank of A_2 be at most M, $3M \leq N$. Then $\lambda_{i+M}(A_1 + A_2) \leq \lambda_i(A_1)$, $i = 1, \dots, N - M$.*

Hence the error may be bounded as follows

$$E[||x - E[x|y]||^2] \leq (1-\delta)P + D \frac{1}{\lambda_{min}(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} \tag{8.39}$$

The smallest eigenvalue of $\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H$ is sufficiently away from zero with high probability as noted in the following lemma:

Lemma 8.3.2. *Let H be a $M \times N$, $M \leq N$ matrix with Gaussian i.i.d. entries. Assume that the assumptions of Theorem 8.3.1 holds. Then with the conditions stated in Theorem 8.3.1, the eigenvalues of $\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H$ are bounded from below as follows:*

$$P(\inf_{x \in S^{N-1}} x^\dagger \Lambda_x^{-1} x + \frac{1}{\sigma_n^2} x^\dagger H^\dagger H x \leq C\frac{N}{P}) \leq e^{-C_1 N}. \tag{8.40}$$

Here S^{N-1} denotes the unit sphere where $x \in S^{N-1}$ if $x \in \mathbb{R}^N$, and $||x|| = 1$.

The proof of this lemma is given in Section A.2 of the Appendix.

We now know that $P(\lambda_{min}(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H) > C\frac{N}{P}) \geq 1 - e^{-C_1 N}$, and hence $P(\frac{1}{\lambda_{min}(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)} < \frac{1}{C}\frac{P}{N}) \geq 1 - e^{-C_1 N}$. Together with the error bound in (8.39), we have $P(E[||X - E[X|Y]||^2] < (1-\delta)P + \frac{1}{C}\frac{D}{N}P) \geq 1 - e^{-C_1 N}$, and the result follows. □

8.4 On Average Performance of Random Scalar Gaussian Channel and Gaussian Erasure Channel

In this section, we consider two closely related random channel structures, and focus on the average MMSE performance. We assume that the receiver knows the channel information, whereas the transmitter only knows the channel probability distribution.

We consider the following measurement strategies: a) (*Random Scalar Gaussian Channel:*) $H = e_i^T$, $i = 1, \dots, N$ with probability $\frac{1}{N}$, where $e_i \in \mathbb{R}^N$ is the i^{th} unit vector. We denote this sampling strategy with S_s. b) (*Gaussian Erasure Channel*) $H = diag(\delta_i)$, where δ_i are i.i.d. Bernoulli random variables with probability of success $p \in [0, 1]$. We denote this sampling strategy with S_b.

We are interested in the following problem:

PROBLEM P4 (Best Unitary Encoder For Random Channels): Let K_x denote the covariance matrix of x. Let $K_x = U \Lambda_x U^\dagger$ be the singular value decomposition of K_x, where U is $N \times N$ unitary matrix, and $\Lambda_x = \text{diag}(\lambda_1, \dots, \lambda_N)$. We fix the eigenvalue distribution with $\Lambda_x = \text{diag}(\lambda_i) \succeq 0$, where $\sum_i \lambda_i = P < \infty$. Let $\mathbb{U}^\mathbb{N}$ be the set of $N \times N$ unitary matrices: $\{U \in \mathbb{C}^N : U^\dagger U = I\}$.

We consider the following minimization problem

$$\inf_{U \in \mathbb{U}^N} E_{H,S}[||x - E[x|y]||^2], \tag{8.41}$$

where the expectation with respect to H is over admissible measurement strategies S_s or S_b. Hence we want to determine the best unitary encoder for the random scalar Gaussian channel or Gaussian erasure channel.

We note that [189] and [190] consider the erasure channel model (S_b in our notation) with the aim of maximizing the ergodic capacity. Their formulations let the transmitter also shape the eigenvalue distribution of the source, whereas ours does not.

We note that our problem formulation is equivalent to following unitary encoding problem $\inf_{U\in\mathbb{U}^N} E_{H,S}[||w-E[w|y]||^2]$, where $K_w=\Lambda_x$, $y=HUw+n$. We also note that by solving the Problem P1 for the measurement scheme in (8.1), one also obtains the solution for the generalized the set-up $y=HVx+n$, where V is any unitary matrix: Let U_o denote an optimal unitary matrix for the scheme in (8.1). Then $V^\dagger U_o\in\mathbb{U}^N$ is an optimal unitary matrix for the generalized set-up.

8.4.1 First order conditions for optimality

Let the possible sampling schemes be indexed by the variable k, where $1\leq k\leq N$ for S_s, and $1\leq k\leq 2^N$ for S_b. Let H_k be the corresponding sampling matrix. Let p_k be the probability of the k^{th} sampling scheme.

We can express the objective function as (8.10)

$$\begin{aligned} E_{H,S}[||x-E[x|y]||^2] &= E_H[\operatorname{tr}((\Lambda_{x,B}^{-1}+\frac{1}{\sigma_n^2}U_B^\dagger H^\dagger HU_B)^{-1})] && (8.42)\\ &= \sum_k p_k \operatorname{tr}((\Lambda_{x,B}^{-1}+\frac{1}{\sigma_n^2}U_B^\dagger H_k^\dagger H_kU_B)^{-1}) && (8.43)\end{aligned}$$

We note that the objective function is a continuous function of U_B. We also note that the feasible set defined by $\{U_B\in\mathbb{C}^{N\times|B|}:U_B^\dagger U_B=I_{|B|}\}$ is a closed and bounded subset of $\mathbb{C}^n$, hence compact. Hence the minimum is attained since we are minimizing a continuous function over a compact set (but the optimum U_B is not necessarily unique).

We note that in general, the feasible region is not a convex set. To see this, let $U_1,U_2\in\mathbb{U}^\mathbb{N}$ and $\theta\in[0,1]$. In general $\theta U_1+(1-\theta)U_2\notin\mathbb{U}^\mathbb{N}$. For instance let $N=1$, $U_1=1$, $U_2=-1$, $\theta U_1+(1-\theta)U_2=2\theta-1\notin\mathbb{U}^1,\quad\forall\,\theta\in[0,1]$. Even if the unitary matrix constraint is relaxed, we observe that the objective function is in general neither a convex or a concave function of the matrix U_B. To see this, one can check the second derivative to see if $\nabla^2_{U_B}f(U_B)\succeq 0$ or $\nabla^2_{U_B}f(U_B)\preceq 0$, where $f(U_B)=\sum_k p_k\operatorname{tr}((\Lambda_{x,B}^{-1}+\frac{1}{\sigma_n^2}U_B^\dagger H_k^\dagger H_kU_B)^{-1})$. For example, let $N=1$,

$U \in \mathbb{R}$, $\sigma_n^2 = 1$, $\lambda > 0$, and $p > 0$ for S_b. Then $f(U) = \sum_k p_k \frac{1}{\lambda^{-1} + U^\dagger H_k^\dagger H_k U}$ can be written as $f(U) = (1-q)\lambda + q\frac{1}{\lambda^{-1}+U^\dagger U}$, where $q \in (0,1]$ is the probability that the one possible measurement is done, and $1-q$ is the probability it is not done. Hence $q = 1$ for S_s, and $q = p$ for S_b. Hence $\nabla_U^2 f(U) = q\, 2\, \frac{3U^2 - \lambda^{-1}}{(\lambda^{-1}+U^2)^3}$, whose sign changes depending on λ, and U. Hence neither $\nabla_U^2 f(U) \succeq 0$ nor $\nabla_U^2 f(U) \preceq 0$ holds for all $U \in \mathbb{R}$.

In general, the objective function depends only on U_B, not U. If U_B satifying $U_B^\dagger U_B = I_{|B|}$, with $|B| < N$ is an optimal solution, then unitary matrices satisfying $U^\dagger U$ can be formed by adding column(s) to U_B without changing the value of the objective function. Hence any such unitary matrix U will also be an optimal solution. Therefore it is sufficient to consider the constraint $\{U_B : U_B^\dagger U_B = I_{|B|}\}$, instead of the condition $\{U : U^\dagger U = I_N\}$, while optimizing the objective function. We also note that if U_B is an optimal solution, $\exp(j\theta)U_B$ is also an optimal solution, where $0 \leq \theta \leq 2\pi$.

Let u_i be the i^{th} column of U_B. We can write the unitary matrix constraint as follows:

$$u_i^\dagger u_k = \begin{cases} 1, & \text{if } i = k, \\ 0, & \text{if } i \neq k. \end{cases} \tag{8.44}$$

with $i = 1, \ldots, |B|$, $k = 1, \ldots, |B|$. Since $u_i^\dagger u_k = 0$, iff $u_k^\dagger u_i = 0$, it is sufficient to consider $k \leq i$. Hence this constraint may be rewritten as

$$e_i^{\mathrm{T}}(U_B^\dagger U_B - I_{|B|})e_k = 0, \quad i = 1, \ldots, |B|,\ k = 1, \ldots, i, \tag{8.45}$$

where $e_i \in \mathbb{R}^{|B|}$ is the i^{th} unit vector.

We now consider the first order conditions for optimality. We note that we are optimizing a real valued function of a complex valued matrix $U_B \in \mathbb{C}^{N \times |B|}$. Let $U_{B,R} = \Re\{U_B\} \in \mathbb{R}^{N \times |B|}$, and $U_{B,I} = \Im\{U_B\} \in \mathbb{R}^{N \times |B|}$ denote the real and imaginary parts of the complex matrix U_B, so that $U_B = U_{B,R} + jU_{B,I}$. One may address this optimization problem by considering the objective function as a mapping from these two real components $U_{B,R}$ and $U_{B,I}$ instead of the complex valued U_B. In the following development, we consider this real framework along with the complex framework.

Let $\widetilde{U}_B = \begin{bmatrix} U_{B,R} \\ U_{B,I} \end{bmatrix} \in \mathbb{R}^{2N\times|B|}$. Let us first consider the set of constraint gradients, and investigate conditions for constraint qualification.

Lemma 8.4.1. *The constraints can be expressed as*

$$\begin{aligned} e_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,R} + U_{B,I}^{\mathrm{T}}U_{B,I})e_k &= e_i^{\mathrm{T}} I_{|B|} e_k, \quad (i,k) \in \gamma & (8.46) \\ e_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,I} - U_{B,I}^{\mathrm{T}}U_{B,R})e_k &= 0, \quad (i,k) \in \bar{\gamma} & (8.47) \end{aligned}$$

where $\gamma = \{(i,k)|i = 1,\ldots,|B|,\ k = 1,\ldots,i\}$, *and* $\bar{\gamma} = \{(i,k)|i = 1,\ldots,|B|,\ k = 1,\ldots,i-1\}$. *The set of constraint gradients with respect to* $\widetilde{U}_B$ *is given by*

$$\left\{ \begin{bmatrix} U_{B,R}(e_ie_k^{\mathrm{T}} + e_ke_i^{\mathrm{T}}) \\ U_{B,I}(e_ie_k^{\mathrm{T}} + e_ke_i^{\mathrm{T}}) \end{bmatrix} |(i,k) \in \gamma \right\} \bigcup \left\{ \begin{bmatrix} U_{B,I}(-e_ie_k^{\mathrm{T}} + e_ke_i^{\mathrm{T}}) \\ U_{B,R}(e_ie_k^{\mathrm{T}} - e_ke_i^{\mathrm{T}}) \end{bmatrix} |(i,k) \in \bar{\gamma} \right\} \qquad (8.48)$$

The elements of this set are linearly independent for any matrix U_B *satisying* $U_B^{\dagger}U_B = I_B$.

Proof: Proof is given in Section A.3 of the Appendix.

Since the constraint gradients are linearly independent for any matrix U_B satisying $U_B^{\dagger}U_B = I_B$, the linear independence constraint qualification (LICQ) holds for any feasible U_B [156, Defn.12.4]. Therefore, the first order condition $\widetilde{L}(\widetilde{U}_B, \nu, \upsilon) = 0$ together with the condition $U_B^{\dagger}U_B = I_B$ is necessary for optimality [156, Thm 12.1], where $\widetilde{L}(\widetilde{U}_B, \nu, \upsilon)$ is the Lagrangian for some Lagrangian multiplier vectors ν, and υ. We use the notation $\widetilde{L}$ instead of L to emphasize the function is seen as a mapping from $\widetilde{U}_B$ instead of U_B.

We note that the unitary matrix constraint in (8.45) can be also expressed as

$$\begin{aligned} e_i^{\mathrm{T}}(U_B^{\dagger}U_B - I_{|B|})e_k &= 0, \quad (i,k) \in \bar{\gamma} & (8.49) \\ e_k^{\mathrm{T}}(U_B^{\dagger}U_B - I_{|B|})e_k &= 0, \quad k \in \{1,\ldots,B\} & (8.50) \end{aligned}$$

We note that in general, $e_i^{\mathrm{T}}(U_B^{\dagger}U_B)e_k = u_i^{\dagger}u_k \in \mathbb{C}$, for $i \neq k$ and $e_k^{\mathrm{T}}(U_B^{\dagger}U_B)e_k = u_k^{\dagger}u_k \in \mathbb{R}$. Hence (8.49) and (8.50) expresses the complex and real valued constraints, respectively.

Now we can express the Lagrangian as follows [please see Section A.4 of the Appendix for a discussion]

$$\begin{aligned}\widetilde{L}(\widetilde{U}_B, \nu, \upsilon) = &\sum_k p_k \operatorname{tr}((\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-1}) \\ &+ \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k} e_i^{\mathrm{T}} (U_B^\dagger U_B - I_{|B|}) e_k + \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k}^* e_i^{\mathrm{T}} (U_B^{\mathrm{T}} U_B^* - I_{|B|}) e_k \\ &+ \sum_{k=1}^{|B|} \upsilon_k e_k^{\mathrm{T}} (U_B^\dagger U_B - I_{|B|}) e_k \qquad (8.51)\end{aligned}$$

where $\nu_{i,k} \in \mathbb{C}$, $(i,k) \in \bar{\gamma}$ and $\upsilon_k \in \mathbb{R}$, $k \in \{1, \ldots, N\}$ are Lagrange multipliers.

Let us define $L(U_B, \nu, \upsilon) = \widetilde{L}(\widetilde{U}_B, \nu, \upsilon)$, the Lagrangian seen as a mapping from U_B, instead of $\widetilde{U}_B$. Now we consider finding the stationary points for the Lagrangian, i.e. the first order condition $\nabla_{\widetilde{U}_B} \widetilde{L}(U_B, \nu, \upsilon) = 0$. We note that this condition is equivalent to $\nabla_{U_B} L(U_B, \nu, \upsilon) = 0$ [209, 210]. We can express this last condition explicitly as

$$\begin{aligned}&\sum_k p_k (\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-2} U_B^\dagger H_k^\dagger H_k \\ &= \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k} e_k e_i^{\mathrm{T}} U_B^\dagger + \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k}^* e_i e_k^{\mathrm{T}} U_B^\dagger + \sum_{k=1}^{|B|} \upsilon_k e_k e_k^{\mathrm{T}} U_B^\dagger, \qquad (8.52)\end{aligned}$$

where we absorbed any constants into Lagrange multipliers. In derivation of these expressions, we have used the chain rule, the rules for differentials of products, and the identity $d\operatorname{tr}(X^{-1}) = -\operatorname{tr}(X^{-2}dX)$, see for example [210]. In particular,

$$\begin{aligned}d(\operatorname{tr}(e_k^{\mathrm{T}} U_B^{\mathrm{T}} U_B^* e_i)) &= d(\operatorname{tr}(e_i^{\mathrm{T}} U_B^\dagger U_B e_k)) &(8.53)\\ &= \operatorname{tr}(e_i^{\mathrm{T}} U_B^\dagger dU_B e_k + e_i^{\mathrm{T}} d(U_B^\dagger) U_B e_k) &(8.54)\\ &= \operatorname{tr}(e_k e_i^{\mathrm{T}} U_B^\dagger dU_B + (dU_B^*)^{\mathrm{T}} U_B e_k e_i^{\mathrm{T}}) &(8.55)\\ &= \operatorname{tr}(e_k e_i^{\mathrm{T}} U_B^\dagger dU_B + e_i e_k^{\mathrm{T}} U_B^{\mathrm{T}} dU_B^*). &(8.56)\end{aligned}$$

$$\begin{aligned}&d(\operatorname{tr}(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-1}) \\ &= -\operatorname{tr}((\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-2} d(U_B^\dagger H_k^\dagger H_k U_B)) \qquad (8.57)\\ &= -\operatorname{tr}((\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-2} (U_B^\dagger H_k^\dagger H_k dU_B + d(U_B^\dagger) H_k^\dagger H_k U_B)). \qquad (8.58)\end{aligned}$$

Remark 8.4.1. *For random scalar Gaussian channel, we can analytically show that these conditions are satisfied by the DFT matrix and the identity matrix. It is not surprising that both the DFT matrix and the identity matrix satisfy these equations, since this optimality condition is the same for both minimizing and maximizing the objective function. We show that the DFT matrix is indeed one of the possibly many optimizers for the case where the values of the nonzero eigenvalues are equal in Lemma 8.4.2. The minimizing property of the identity matrix in the noiseless case is investigated in Lemma 8.4.3.*

For Gaussian erasure channel, we show that the observations presented in compressive sensing literature implies that the MMSE is small with high probability for the DFT matrix (see Section 8.2). Although these observations and the other special cases presented in Section 8.4.2 may suggest the result that the DFT matrix may be an optimum solution for the general case, we show that this is not the case by presenting a counterexample where another unitary matrix not satisfying $|u_{ij}|^2 = 1/N$ *outperforms the DFT [Lemma 8.4.6].*

8.4.2 Special cases

In this section, we consider some related special cases. For random scalar Gaussian channel, we will show that when the nonzero eigenvalues are equal any covariance matrix (with the given eigenvalues) having a constant diagonal is an optimum solution [Lemma 8.4.2]. This includes Toeplitz covariance matrices or covariance matrices with any unitary transform satisfying $|u_{ij}|^2 = 1/N$. We note that the DFT matrix satisfies $|u_{ij}|^2 = 1/N$ condition, and always produces circulant covariance matrices. We will also show that for both channel structures, for the noiseless case (under some conditions) regardless of the entropy or the degree of freedom of a signal, the worst coordinate transformation is the same, and given by the identity matrix [Lemma 8.4.3].

For Gaussian erasure channel, we will show that when only one of the eigenvalues is nonzero (i.e. rank of the covariance matrix is one), any unitary transform satisfying $|u_{ij}|^2 = 1/N$ is an optimizer [Lemma 8.4.4]. We will also show that under the relaxed condition $\mathrm{tr}(K_x^{-1}) = R$, the best covariance matrix is circulant, hence the best unitary transform is the DFT matrix [Lemma 8.4.5]. Furthermore in Section 8.2, we show that the observations presented in compressive sensing literature implies that the MMSE is small with high probability when $|u_{ij}|^2 = 1/N$. Although all these observations may suggest the idea that the DFT matrix may be an optimum solution in the general case, we will show that this is not the case by presenting a counterexample where another unitary matrix not satisfying $|u_{ij}|^2 = 1/N$ outperforms the DFT matrix [Lemma 8.4.6].

Before moving on, we note the following relationship between the eigenvalue distribution and the MMSE. Let $H \in \mathbb{R}^{M \times N}$ be a given sampling matrix which formed by taking $1 \leq 3M \leq N$ rows from the identity matrix. Assume that $\Lambda_x \succ 0$. Let the eigenvalues of a matrix A be denoted in decreasing order as $\lambda_1(A) \geq \lambda_2(A), \dots, \geq \lambda_N(A)$. The MMSE can be expressed as (8.10)

$$E[||x - E[x|y]||^2] = \mathrm{tr}\,((\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U^\dagger H^\dagger H U)^{-1}) \tag{8.59}$$

$$= \sum_{i=1}^{N} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U^\dagger H^\dagger H U)} \tag{8.60}$$

$$= \sum_{i=M+1}^{N} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U^\dagger H^\dagger H U)} + \sum_{i=1}^{M} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U^\dagger H^\dagger H U)} \tag{8.61}$$

$$\geq \sum_{i=M+1}^{N} \frac{1}{\lambda_{i-M}(\Lambda_x^{-1})} + \sum_{i=1}^{M} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma_n^2} U^\dagger H^\dagger H U)}, \tag{8.62}$$

$$\cdot \geq \sum_{i=M+1}^{N} \frac{1}{\lambda_{i-M}(\Lambda_x^{-1})} + \sum_{i=1}^{M} \frac{1}{\frac{1}{\lambda_{N-i+1}(\Lambda_x)} + \frac{1}{\sigma_n^2}}, \tag{8.63}$$

$$= \sum_{i=M+1}^{N} \lambda_{N-i+M+1}(\Lambda_x) + \sum_{i=N-M+i}^{N} \frac{1}{\frac{1}{\lambda_i(\Lambda_x)} + \frac{1}{\sigma_n^2}}, \tag{8.64}$$

$$= \sum_{i=M+1}^{N} \lambda_i(\Lambda_x) + \sum_{i=N-M+1}^{N} \frac{1}{\frac{1}{\lambda_i(\Lambda_x)} + \frac{1}{\sigma_n^2}}, \tag{8.65}$$

where we have used case (b) of Lemma 8.3.1 in (8.62), and the fact that $\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} U^\dagger H^\dagger H U) \leq \lambda_i(\Lambda_x^{-1}) + \frac{1}{\sigma^2}\lambda_1(U^\dagger H^\dagger H U) = \lambda_i(\Lambda_x^{-1}) + \frac{1}{\sigma^2}$ in (8.63).

This lower bound is consistent with our intuition: If the eigenvalues are well-spread, that is $D(\delta)$ is large in comparison to N for δ close to 1, the error cannot be made small without large number of measurements.

The first term in (8.65) may be obtained by the following intuitively appealing alternative argument: The energy compaction property of Karhunen-Loève expansion guarantees that the best representation of this signal with M variables in mean-square error sense is obtained by first decorrelating the signal with $U^\dagger$ and then using the random variables that correspond to the highest M eigenvalues. The mean-square error of such a representation is given by the sum of the remaining eigenvalues, i.e. $\sum_{i=M+1}^{N} \lambda_i(\Lambda_x)$. Here we make measurements before decorrelating the signal, and each component is measured with noise. Hence the error of our measurement scheme is lower bounded by the error of the optimum scheme, which is exactly the first term in (8.65). The second term is the MMSE associated with the measurement scheme where M independent variables with variances given by the M smallest eigenvalues of Λ_x are observed through i.i.d noise.

Lemma 8.4.2. *Let* $\mathrm{tr}(K_x) = P$. *Assume that the nonzero eigenvalues are equal, i.e.* $\Lambda_{x,B} = \frac{P}{|B|} I_B$. *Let* $K_n = \sigma_n^2 I$. *Then the minimum average error for random scalar Gaussian channel (*$H = e_i^T$, $i = 1, \dots, n$ *with probability* $\frac{1}{N}$*) is*

$$P - \frac{P}{|B|} + \frac{1}{1 + \frac{P}{N}\frac{1}{\sigma_n^2}} \frac{P}{|B|}, \tag{8.66}$$

which is achieved by covariance matrices with constant diagonal. In particular, covariance matrices whose unitary transform is the DFT matrix satisfy this.

Proof: Note that if none of the eigenvalues are zero, $K_x = I$ regardless of the unitary transform, hence the objective function value does not depend on it.) The objective function may be expressed as (8.43)

$$\begin{aligned} E_{H,S}[||x - E[x|y]||^2] &= \sum_{k=1}^{N} \frac{1}{N} \mathrm{tr}\, (\frac{|B|}{P} I_B + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-1} &(8.67)\\ &= \frac{P}{|B|} \sum_{k=1}^{N} \frac{1}{N} (|B| - 1 + (1 + \frac{P}{|B|}\frac{1}{\sigma_n^2} H_k U_B U_B^\dagger H_k^\dagger)^{-1}) &(8.68)\\ &= \frac{P}{|B|}(|B| - 1) + \sum_{k=1}^{N} \frac{P}{|B|}\frac{1}{N}(1 + \frac{P}{|B|}\frac{1}{\sigma_n^2} e_k^\dagger U_B U_B^\dagger e_k)^{-1}, &(8.69) \end{aligned}$$

where in (8.68) we have used Lemma 2 of [199]. We now consider the minimization of the following function

$$\sum_{k=1}^{N}(1+\frac{P}{|B|}\frac{1}{\sigma_n^2}e_k^{\dagger}U_BU_B^{\dagger}e_k)^{-1} = \sum_{k=1}^{N}\frac{1}{1+\frac{P}{|B|}\frac{1}{\sigma_n^2}\frac{|B|}{P}z_k} \tag{8.70}$$

$$= \sum_{k=1}^{N}\frac{1}{1+\frac{1}{\sigma_n^2}z_k}, \tag{8.71}$$

where $(U_BU_B^{\dagger})_{kk} = \frac{|B|}{P}(K_x)_{kk} = \frac{|B|}{P}z_k$ with $z_k = (K_x)_{kk}$. Here $z_k \geq 0$ and $\sum_k z_k = P$, since $\operatorname{tr}(K_x) = P$. We note that the goal is the minimization of a convex function over a convex region. Since the objective and constraint functions are differentiable and Slater's condition is satisfied, we consider the Karush-Kuhn-Tucker (KKT) conditions which are necessary and sufficient for optimality [151]:

$$\nabla_z(\sum_{k=1}^{N}\frac{1}{1+\frac{1}{\sigma_n^2}z_k}+\mu(\sum_{k=1}^{N}z_k)-\sum_{k=1}^{N}\nu_kz_k)=0 \tag{8.72}$$

where μ, ν are Lagrange multipliers with $\nu_i \geq 0$, and $\nu_iz_i = 0$, for $i = 1,\dots,N|$. Solving for the KKT conditions and investigating the set of active constraints for the best objective function value reveals that best z_i is given by $z_i = P/N$. We observe that this condition is equivalent to require that the covariance matrix has constant diagonal. This condition can be always satisfied; for example with a Toeplitz covariance matrix or with any unitary transform satisfying $|u_{ij}|^2 = 1/N$. We note that the DFT matrix satisfies $|u_{ij}|^2 = 1/N$ condition, and always produces circulant covariance matrices.

Lemma 8.4.3. *We now consider the random scalar channel without noise, and consider the following maximization problem which searches for the worst coordinate system for a signal to lie in: Let $x \in \mathbb{C}^N$ be a zero-mean proper Gaussian random vector. Let $\Lambda_x = \operatorname{diag}(\lambda_i)$, with $\operatorname{tr}(\Lambda_x) = P$ be given.*

$$\sup_{U\in\mathbb{U}^{\mathbb{N}}} E[\sum_{t=1}^{N}[(x_t - E[x_t|y])^2]], \tag{8.73}$$

where

$$y = x_i \quad \text{with probability } \frac{1}{N}, \qquad i=1,\dots,N \tag{8.74}$$

$$K_x = U\Lambda_xU^{\dagger}. \tag{8.75}$$

The solution to this problem is as follows: The maximum value of the objective function is $\frac{N-1}{N}P$. $U = I$ achieves this maximum value.

Remark 8.4.2. *We emphasize that this result does not depend on the eigenvalue spectrum Λ_x.*

Remark 8.4.3. *We note that when some of the eigenvalues of the covariance matrix are identically zero, the eigenvectors corresponding to the zero eigenvalues can be chosen freely (of course as long as the resulting transform U is unitary).*

Proof: The objective function may be written as

$$\begin{aligned} E[\sum_{t=1}^{N}[||x_t - E[x_t|y]||^2]] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{N} E[||x_t - E[x_t|x_i]||^2]] && (8.76)\\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{N}(1-\rho_{i,t}^2)\sigma_{x_t}^2 && (8.77)\end{aligned}$$

where $\rho_{i,t} = \frac{E[x_t x_i^\dagger]}{(E[||x_t||^2]E[||x_i||^2])^{1/2}}$ is the correlation coefficient between x_t and x_i, assuming $\sigma_{x_t}^2 = E[||x_t||^2] > 0$, $\sigma_{x_i}^2 > 0$. (Otherwise one may set $\rho_{i,t} = 1$ if $i = t$, and $\rho_{i,t} = 0$ if $i \neq j$.) Now we observe that $\sigma_t^2 \geq 0$, and $0 \leq |\rho_{i,t}|^2 \leq 1$. Hence the maximum value of this function is given by $\rho_{i,t} = 0$, $\forall\, t, i$ s.t. $t \neq i$. We observe that any diagonal unitary matrix $U = \text{diag}(u_{ii})$, $|u_{ii}| = 1$ (and also any $\bar{U} = U\Pi$, where Π is a permutation matrix) achieves this maximum value. In particular, the identity transform $U = I_N$ is an optimal solution.

We note that a similar result hold for Bernoulli sampling scheme: Let $y = Hx$. $\sup_{U \in \mathbb{U}^N} E_{H,S}[||x - E[x|y]||^2]$, where the expectation with respect to H is over admissible measurement strategies S_b is $(1-p)\,\text{tr}\,(K_x)$, which is achieved by any $U\Pi$, $U = \text{diag}(u_{ii})$, $|u_{ii}| = 1$, Π is a permutation matrix.

Lemma 8.4.4. *Suppose $|B| = 1$, i.e. $\lambda_k = P > 0$, and $\lambda_j = 0$, $j \neq k, j \in 1, \ldots, N$. Let the channel be the Gaussian erasure channel, i.e. $y = Hx + n$, where $H = \text{diag}(\delta_i)$, where δ_i are i.i.d. Bernoulli random variables, and $K_n = \sigma_n^2 I_N$. Then the minimum error is given by*

$$E[\frac{1}{\frac{1}{P} + \frac{1}{\sigma_n^2}\frac{1}{N}\sum_{i=1}^{N}\delta_i}], \tag{8.78}$$

where this optimum is achieved by any unitary matrix with entries of k^{th} column satisfying $|u_{ik}|^2 = 1/N$, $i = 1, \dots, N$.

Proof: Let $v = [v_1, \dots, v_n]^{\mathrm{T}}$, $v_i = |u_{ki}|^2$, $i = 1, \dots, N$, where T denotes transpose. We note the following

$$E[\mathrm{tr}\,(\frac{1}{P} + \frac{1}{\sigma_n^2} U_B^{\dagger} H^{\dagger} H U_B)^{-1}] = E[\frac{1}{\frac{1}{P} + \frac{1}{\sigma_n^2}\sum_{i=1}^{N} \delta_i |u_{ki}|^2}] \tag{8.79}$$

$$= E[\frac{1}{\frac{1}{P} + \frac{1}{\sigma_n^2}\sum_{i=1}^{N} \delta_i v_i}]. \tag{8.80}$$

The proof uses an argument in the proof of [180, Thm. 1], which is also used in [199]. Let $\Pi_i \in \mathbb{R}^{N \times N}$ denote the permutation matrix indexed by $i = 1, \dots, N!$. We note that a feasible vector v satisfies $\sum_{i=1}^{N} v_i = 1$, $v_i \geq 0$, which forms a convex set. We observe that for any such v, weighted sum of all permutations of v, $\bar{v} = \frac{1}{N!}\sum_{i=1}^{N!} \Pi_i v = (\frac{1}{N}\sum_{i=1}^{N} v_i)[1, \dots, 1]^T = [\frac{1}{N}, \dots, \frac{1}{N}]^T \in \mathbb{R}^N$ is a constant vector and also feasible. We note that $g(v) = E[\frac{1}{\frac{1}{P} + \frac{1}{\sigma_n^2}\sum_i \delta_i v_i}]$ is a convex function of v over the feasible set. Hence $g(v) \geq g(\bar{v}) = g([1/N, \dots, 1/N])$ for all v, and $\bar{v}$ is the optimum solution. Since there exists a unitary matrix satisfying $|u_{ik}|^2 = 1/N$ for any given k (such as any unitary matrix whose k^{th} column is any column of the DFT matrix), the claim is proved.

Lemma 8.4.5. *Let $K_x^{-1} \succ 0$. Instead of fixing the eigenvalue distribution, let us consider the relaxed constraint $\mathrm{tr}(K_x^{-1}) = R$. Let $K_n \succ 0$. Let the channel be the Gaussian erasure channel, i.e. $y = Hx + n$, $H = \mathrm{diag}(\delta_i)$, where δ_i are i.i.d. Bernoulli random variables with probability of success p. Then*

$$\arg\min_{K_x^{-1}} E_{H,S}[||x - E[x|y]||^2] = \arg\min_{K_x^{-1}} E_H[(\mathrm{tr}(K_x^{-1} + \frac{1}{\sigma_n^2} H^{\dagger} K_n^{-1} H)^{-1}] \tag{8.81}$$

is a circulant matrix.

Proof: The proof uses an argument in the proof of [190, Thm. 12], [189]. Let Π be the following permutation matrix,

$$\Pi = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & 1 & 0\cdots \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}. \tag{8.82}$$

We observe that Π and Π^l (l^{th} power of Π) are unitary matrices. We form the following matrix $\bar{K}_x^{-1} = \frac{1}{N}\sum_{l=0}^{N-1} \Pi^l K_x^{-1} (\Pi^l)^\dagger$, which also satisfies the power constraint $\mathrm{tr}\,(\bar{K}_x^{-1}) = R$. We note that since $K_x^{-1} \succ 0$, so is $\bar{K}_x^{-1} \succ 0$, hence $\bar{K}_x^{-1}$ is well-defined.

$$\begin{aligned}
& E[(\mathrm{tr}(\frac{1}{N}\sum_{l=0}^{N-1} \Pi^l K_x^{-1}(\Pi^l)^\dagger + \frac{1}{\sigma_n^2} H^\dagger K_n^{-1} H)^{-1}] \\
& \leq \frac{1}{N}\sum_{l=0}^{N-1} E[\mathrm{tr}(\Pi^l K_x^{-1}(\Pi^l)^\dagger + \frac{1}{\sigma_n^2} H^\dagger K_n^{-1} H)^{-1}] && (8.83) \\
& = \frac{1}{N}\sum_{l=0}^{N-1} E[\mathrm{tr}(\Pi^l (K_x^{-1} + \frac{1}{\sigma_n^2} (\Pi^l)^\dagger H^\dagger K_n^{-1} H \Pi^l)(\Pi^l)^\dagger)^{-1}] && (8.84) \\
& = \frac{1}{N}\sum_{l=0}^{N-1} E[\mathrm{tr}(K_x^{-1} + \frac{1}{\sigma_n^2} (\Pi^l)^\dagger H^\dagger K_n^{-1} H \Pi^l)^{-1}] && (8.85) \\
& = \frac{1}{N}\sum_{l=0}^{N-1} E[\mathrm{tr}(K_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger K_n^{-1} H)^{-1}] && (8.86) \\
& = E[\mathrm{tr}(K_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger K_n^{-1} H)^{-1}] && (8.87)
\end{aligned}$$

We note that $\mathrm{tr}((M + K_n^{-1})^{-1})$ is a convex function of M over the set $M \succ 0$, since $\mathrm{tr}(M^{-1})$ is a convex function (see for example [151, Exercise 3.18]), and composition with an affine mapping preserves convexity [151, Sec. 3.2.2]. Hence the first inequality follows from Jensen's Inequality. (8.85) is due to the fact that Π^ls are unitary and trace is invariant under unitary transforms. (8.86) follow from the fact that $H\Pi^l$ has the same distribution with H. Hence we have shown that $\bar{K}_x^{-1}$ provides a lower bound for arbitrary K_x^{-1} satisfying the power constraint. Since $\bar{K}_x^{-1}$ is circulant and also satisfies the power constraint $\mathrm{tr}\,(\bar{K}_x^{-1}) = R$, the optimum K_x^{-1} should be circulant.

We note that we cannot follow the same argument for the constraint $\mathrm{tr}(K_x) = P$, since the objective function is concave in K_x over the set $K_x \succ 0$. This fact was proved for a slightly different setting in Section 3.1, here we repeat the argument for convenience: $E[||x - E[x|y]||^2] = \mathrm{tr}\,(K_e)$, where $K_e = K_x - K_{xy}K_y^{-1}K_{xy}^\dagger$. We note that K_e is the Schur complement of K_y in $K = [K_y\ K_{yx}; K_{xy}\ K_x]$, where $K_y = HK_xH^\dagger + K_n$, $K_{xy} = K_xH^\dagger$. Schur complement is matrix concave in $K \succ 0$, for example see [151, Exercise 3.58]. Since trace is a linear operator, $\mathrm{tr}(K_e)$ is concave in K. Since K is an affine mapping of K_x, and composition with an affine mapping preserves concavity [151, Sec. 3.2.2], $\mathrm{tr}(K_e)$ is concave in K_x.

Lemma 8.4.6. *The DFT matrix is, in general, not an optimizer of Problem P4 for Gaussian erasure channel.*

Proof: We provide a counterexample to prove the claim of the lemma: An example where a unitary matrix not satisfying $|u_{ij}|^2 = 1/N$ outperforms the DFT matrix. Let $N = 3$. Let $\Lambda_x = \mathrm{diag}(1/6, 2/6, 3/6)$, and $K_n = I$. Let U be

$$U_0 = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \tag{8.88}$$

Hence K_x becomes

$$K_x = \begin{bmatrix} 1/3 & 0 & 1/6 \\ 0 & 1/3 & 0 \\ 1/6 & 0 & 1/3 \end{bmatrix} \tag{8.89}$$

We write the average error as a sum conditioned on the number of measurements as $J(U) = \sum_{M=0}^{3} p^M (1-p)^{3-M} e_M(U)$, where e_M denotes the total error of all cases where M measurements are done. Let $e(U) = [e_0(U), e_1(U), e_2(U), e_3(U)]$. The calculations reveal that $e(U_0) = [1, 65/24, 409/168, 61/84]$ whereas $e(F) = [1, 65/24, 465/191, 61/84]$, where F is the DFT matrix. We see that all the entries are the same with the DFT case, except $e_2(U_0) < e_2(F)$, where $e_2(U_0) = 409/168 \approx 2.434524$ and $e_2(F) = 465/191 \approx 2.434555$. Hence U_0 outperforms the DFT matrix.

We note that our argument covers any unitary matrix that is formed by changing the order of the columns of the DFT matrix, i.e. any matching of the given eigenvalues and the columns of the DFT matrix: U_0 provides better performance than any K_x formed by using the given eigenvalues and any unitary matrix formed with columns from the DFT matrix. The reported error values hold for all such K_x.

8.4.3 Rate-distortion bound

We note that by combining the rate distortion theorem and the converse to the channel coding theorem, one can see that the rate-distortion function lower bounds the channel capacity for a given channel structure [211]. We now show that this rate-distortion bound is not achievable with the channel structure we have.

We consider the scalar real channel: $y = au\alpha + n$, where $a = 1$ with probability p, and $a = 0$ with probability $1 - p$. Let $u\alpha = x$. Let α, and n be independent zero mean Gaussian random variables. When needed, we emphasize the random variables the expectations are taken with respect to; we denote the expectation with respect to the random channel gain by $E_a[.]$, and the expectation with respect to random signals involved (including x and n) by $E_s[.]$ Assuming the knowledge of realization of a at the receiver, but not at the transmitter, the capacity of this channel with power constraint $P_x < \infty$ is given by

$$\bar{C} = \max_{E_s[x^2] \leq P_x} E_a[I(x;y)] \tag{8.90}$$

$$= \max_{E_s[x^2] \leq P_x} [pI(u\alpha + n; x) + (1-p)I(0;x)] \tag{8.91}$$

$$= p\, 0.5 \log(1 + \frac{P_x}{\sigma_n^2}). \tag{8.92}$$

Here we have used the fact that the capacity of an additive Gaussian channel with noise variance σ_n^2 and power constraint P_x is $0.5 \log(1 + \frac{P_x}{\sigma_n^2})$.

The rate-distortion function of a Gaussian random variable with variance σ_α^2 is given as

$$R(D) = \min_{f_{\hat{\alpha}|\alpha},\, E[(\alpha - \hat{\alpha})^2] \leq D} I(\alpha; \hat{\alpha}) = \max\{0.5 \log(\frac{\sigma_\alpha^2}{D}), 0\}. \tag{8.93}$$

We note that by the converse to the channel coding theorem, for a given channel structure with capacity C, we have $R(D) \leq C$, which provides $D(C) \leq E[(\alpha - \hat{\alpha})^2]$

[211]. Hence

$$\begin{aligned} E_{a,s}[(\alpha-\hat{\alpha})^2] &= p\,E_{\alpha}[(\alpha-\hat{\alpha})^2|a=1]+(1-p)\,E_{\alpha}[(\alpha-\hat{\alpha})^2|a=0] && (8.94)\\ &\geq pD(R)+(1-p)D(R) && (8.95)\\ &= \sigma_{\alpha}^2\;2^{-2R} && (8.96)\\ &\geq \sigma_{\alpha}^2\;2^{-p\log(1+\frac{P_x}{\sigma_n^2})} && (8.97)\\ &= \sigma_{\alpha}^2\;(\frac{\sigma_n^2}{\sigma_n^2+P_x})^p && (8.98) \end{aligned}$$

where we have used the fact that $C(a) \geq R(D)$ for each realization of the channel, hence $\bar{C} = p\,C(a=1)+(1-p)C(a=0) \geq pR(D)+(1-p)R(D) = R(D)$. On the other hand the average error of this system with Gaussian input α, $\sigma_{\alpha}^2 u^2 = \sigma_x^2 = P_x$ is

$$\begin{aligned} E_{a,s}[(\alpha-\hat{\alpha})^2] &= (1-p)\sigma_{\alpha}^2+p(\sigma_{\alpha}^2-\frac{\sigma_{\alpha}^2u^2\sigma_{\alpha}^2}{P_x+\sigma_n^2}) && (8.99)\\ &= (1-p)\sigma_{\alpha}^2+p\frac{\sigma_{\alpha}^2\;\sigma_n^2}{P_x+\sigma_n^2} && (8.100) \end{aligned}$$

We observe that (8.100) is strictly larger than the bound in (8.98) for $0 < p < 1$, $\sigma_{\alpha}^2 > 0$. (This follows from the fact that $f(x) = b^x$, $b \neq 0, 1$ is a strictly convex function so that $f((1-p)x_1+px_2) < (1-p)f(x_1)+pf(x_2)$ for $0 < p < 1$, $x_1 \neq x_2$. Hence with $b = \frac{\sigma_n^2}{\sigma_n^2+P_x}$, $0 < P_x < \infty$, $x_1 = 0$, $x_2 = 1$, the inequality follows.)

8.5 Discussion and Conclusions

We have considered the transmission of a Gaussian vector source over a multidimensional Gaussian channel where a random or a fixed subset of the channel outputs are erased. The unitary transformation that connects the canonical signal domain and the measurement space played a crucial role in our investigation. Under the assumption the estimator knows the channel realization, we have investigated the MMSE performance both in average and in terms of guarantees that hold with high probability as a function of system parameters.

In addition to providing insights into the importance of unitary transformation in transmission of signals through Gaussian erasure channels, our work also contributed to our understanding of the relationship between the MMSE and the total uncertainty in the signal as quantified by information theoretic measures such as entropy (eigenvalues) and the spread of this uncertainty (basis). We believe that through this relationship our work here also sheds light on how to properly characterize the concept of "coherence", and complements our work in Chapter 7.

In Section 8.1, we have considered circularly wide-sense stationary signals, which is a natural way to model wide-sense stationary signals in finite dimension. In this section the covariance matrix was circulant by assumption, hence the unitary transform was fixed and given by the DFT matrix. In this part, we have focused on equidistant sampling and gave the explicit expression for the MMSE. We have also shown that two times the total power outside a properly chosen set of indices (a set of indices which do not overlap when shifted by an amount determined by the sampling rate) provides an upper bound for the MMSE. We have observed that the notion of such a set of indices generalizes the conventional sense of bandlimited signals. Our results showed that the error will be zero if there is such a set of indices that contains all of the power even if the signal is not band-limited (low-pass, high-pass) in the conventional sense. We have also noted that the results of Section 8.2 are applicable to c.w.s.s. signals. For instance, when these signals have a flat nonzero eigenvalue spectrum, they can be estimated with zero MMSE with high probability with a given number of noiseless measurements whose locations are chosen uniformly random.

In Section 8.2 and Section 8.3, we have illustrated how some recent results in matrix theory mostly presented in compressive sampling framework can be used to find performance bounds for the MMSE estimation. In this part we have provided performance guarantees that hold with high probability. We have considered three set-ups: i) sampling of a signal with fixed support at uniformly random chosen measurement locations at a fixed domain; ii) sampling of a signal with uniformly random support at fixed measurement locations at a fixed measurement domain; iii) random projections (random channel matrix with i.i.d. Gaussian entries) where the eigenvalue distribution of the covariance matrix is arbitrary. For the first two cases, we have investigated bounds on the MMSE depending on the support size and the number of measurements. For the third case, we have illustrated the interplay between the amount of information in the signal, and the spread of this information in the measurement domain for providing performance guarantees.

We now make a few remarks on our MMSE based sparse signal recovery approach and computational constraints. In a standard compressive sensing problem, for finding the unknown signal a l_1 minimization problem can be formulated [185, 186]. Efficient methods for the solution of such problems is known, for instance the linear programming approach of [212]. In our formulation, we solve for the MMSE estimator whose direct implementation requires inversion of a matrix, which is a computationally heavy operation. Nevertheless we observe the following: the mean-square error is a convex function of the estimator matrix B, where $E[x|y] = By$, (for instance see (3.4)), so that an approximate numerical solution may be found by using convex programming methods. Hence, there may exist some room for improvement in implementation of the MMSE approach. Whether the approximate solutions provided by these methods will perform well, or these algorithms (together with the implementation of the multiplication operation By) can be customized to be as efficient as the approaches in compressive sensing literature are interesting research directions to pursue in the future.

In Section 8.4, we have focused on the average performance. We have considered two channel structures: i) random Gaussian scalar channel where only one measurement is done through Gaussian noise and ii) Gaussian erasure channel where measurements are done through parallel Gaussian channels with a given channel erasure probability. Under these channel structures, we have formulated the problem of finding the most favorable unitary transform under average performance criterion. We have investigated the convexity properties of this optimization problem, and obtain conditions of optimality through variational equalities. We were not able to solve this problem in its full setting, but we have solved some related special cases. Among these we have identified special cases where DFT-like unitary transforms (unitary transforms with $|u_{ij}|^2 = \frac{1}{N}$) turn out to be the best coordinate transforms, possibly along with other unitary transforms. Although these observations and the observations of Section 8.2 (which are based on compressive sensing results) may suggest the idea that the DFT matrix may be indeed an optimum unitary matrix for any eigenvalue distribution, we have provided a counterexample.

Chapter 9

Sampling and Finite Dimensional Representations of Stationary Gaussian Processes

One of the main motivations of the work in Chapter 7 and Chapter 8 were to provide insight into statistical dependence in random fields; in particular geometric properties of the spread of uncertainty. The problems studied in these chapters were formulated in a finite dimensional framework. In this chapter, we continue our investigation with stationary Gaussian sources defined on $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$. We formulate various problems related to the finite-length representations and sampling of these signals, which will shed light on different aspects of statistical dependence in random fields.

We first consider the decay rates for the error between finite dimensional representations and infinite dimensional representations. Here our approach is based on the notion of mixing which is concerned with dependence in asymptotical sense, that is the dependence between two points of a random process as the distance between these two points increases. The concept of mixing is proposed as a measure of dependence for random processes with many variants, see for example [21] and the references therein. There is a vast literature on the notion of mixing in the fields of information theory and applied mathematics, but this notion does not seem to have been utilized in signal processing community. Providing several alternative ways to quantify dependence in random processes, this family of notions may provide new perspectives in signal processing problems where one needs to quantify the dependence in a signal family. Our work constitutes an example for these potential directions of research. In Section 9.1, based on this concept, we investigate the difference between using finite window and infinite window length representations of a random process. We show that for exponentially mixing sequences, for different representations and estimators, the error difference between using a finite-length representation and infinite-length representation is upper bounded by an exponentially decreasing function of the finite window length.

We then consider the MMSE estimation of a stationary Gaussian source from its noisy samples. In Section 9.2.2, we first show that for stationary sources for the purpose of calculating the MMSE based on equidistant samples, asymptotically circulant matrices can be used instead of original covariance matrices, which are Toeplitz. This result suggests that circularly wide-sense stationary signals in finite dimensions are more than an analogy for stationary signals in infinite dimensions: there is an operational relationship between these two signal models. To show convergence of the error expressions in this section, we make use of our results in Section 9.1 regarding finite-length representations. In Section 9.2.3, we consider the MMSE associated with estimation of a stationary Gaussian source on $\mathbb{Z}_+$ from its equidistant samples on $\mathbb{Z}_+$. Using the previous result, we give the explicit expression for the MMSE in terms of power spectral density. An important aspect of our framework is the fact that we consider the sampling of the source on the half infinite line $\mathbb{Z}_+$ instead of the infinite line $\mathbb{Z}$. This framework makes direct usage of stationary arguments difficult, and makes the arguments more challenging.

In Section 9.1, we consider decay rates of error for finite-length truncations based on the notion of mixing. In Section 9.2 we focus on the problem of the MMSE estimation of a stationary Gaussian source from its noisy samples, and the sequences of finite dimensional models therein. We conclude in Section 9.3.

9.1 Finite-length Representations

Let $\{X_t\}$ be a real valued zero-mean stationary Gaussian random process defined on $I = \mathbb{Z}$. We use $r_x(t_1 - t_2) = E[X_{t_1} X_{t_2}]$ to denote the auto-covariance function. We assume that $r_x \in l_1(\mathbb{Z})$, i.e. the auto-correlation function is absolutely-summable.

We assume that $\{X_t\}$ has a moving average representation

$$X_t = \sum_{k=0}^{\infty} c_k W_{t-k}, \quad \forall t \tag{9.1}$$

where W_t's are i.i.d real valued zero-mean Gaussian random variables with variance $\sigma_w^2 < \infty$. Here $\{c_k\} \in l_2$. We note that the infinite summation is guaranteed to be mean-square convergent to some limit with $\sigma_{X_t} < \infty$, which can be proven using for example [213, Sec. 7.11, pr.11].

We further assume that $\{X_t\}$ may be represented as an autoregressive process as follows:

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + W_t, \quad \forall t \tag{9.2}$$

Here $a_k \in \mathcal{R}$ are not t dependent. We assume that $\{a_k\}$ is absolutely summable, $\{a_k\} \in l_1$, so that with $\sigma_{X_{t-k}} < \infty$, $k > 0$, $E[|\sum_{k=1}^{\infty} a_k X_{t-k}|] < \infty$, and X_t has finite variance.

We assume that the source is exponentially mixing; the decay of statistical dependence between X_{t_1} and X_{t_2} upper bounded by an exponential function as $|t_1 - t_2|$ increases. Of course, here one needs to make the notion of statistical dependence clear. We present a precise definition of exponentially mixing source in Definition 9.1.1.

We now take a brief look at the problems we investigate in this section. We will be interested in decay rates of errors introduced by the following different truncations:

- $\{\tilde{X}_t\}$ the N-truncated representation of $\{X_t\}$

$$\tilde{X}_t = \sum_{k=1}^{N} a_k X_{t-k} + W_t, \quad \forall t \tag{9.3}$$

- $\{\tilde{\tilde{X}}_t\}$ the finite-length estimator associated with causal MMSE estimation of $\{X_t\}$ from its equidistant samples

$$\tilde{\tilde{X}}_t = \sum_{k=1}^{\lfloor N/\tau \rfloor} b_k X_{t-\tau k}, \quad \forall t \tag{9.4}$$

  where b_k are the optimal coefficients for the MMSE estimation. Here the samples which fall within the length N window preceding X_t contribute to the estimation.

- $\{\hat{\hat{X}}_t\}$ the finite-length estimator associated with acausal MMSE estimation of $\{X_t\}$ from its equidistant samples

$$\hat{\hat{X}}_t = \sum_{k=-\lfloor N/\tau \rfloor}^{\lfloor N/\tau \rfloor} d_k X_{t-\tau k}, \quad \forall t \tag{9.5}$$

  where d_k are the optimal coefficients for the MMSE estimation. Here the samples which fall within the length $2N+1$ window around X_t contribute to the estimation.

We also comment on the decay of the mutual information between the current value of the random process and the remaining values of the random process, given the values of the process in a finite window of length N.

We now give some technical details about the existence of the above representations. $\{X_t\}$ has a nonnegative measure F_x on $(-\pi, \pi]$ called the spectral measure such that $r_x(\tau) = \int_{-\pi}^{\pi} \exp^{j\tau\theta} dF_x(\theta)$. The derivative of F with respect to θ is called the spectral density and denoted by $f_x(\theta)$. We note that the Gaussian stationary process admits the causal representation in (9.1) if and only if the spectral measure $F_x(\theta)$ is absolutely continuous and the spectral density $f_x(\theta)$ satisfies the following condition [214, pg. 112]

$$\int_{-\pi}^{\pi} \log f_x(\theta) d\theta > -\infty. \tag{9.6}$$

The conditions on the spectral density for the process to have infinite order autoregressive representation can be found in [215, Ch.7].

The integrability condition in (9.6) guarantees that the process is non-deterministic, the process cannot be determined from its past values [132, Ch 10.6]. We also note that this assumption implies $f_{x,\inf} = \operatorname{ess\,inf} f_x > 0$. It is worth emphasizing that this means the process $\{X_t\}$ cannot be band-limited or similar (multi-pass. etc). Note that we have $r_x \in l_1(\mathbb{Z})$, so we also have the following: $f_{x,\sup} = \operatorname{ess\,sup} f_x < \infty$.

We now provide a brief overview of our results in this section:

(i) The exponentially mixing sequence has exponentially decaying AR model coefficients.

(ii) The error associated with the truncation of the AR model coefficients is exponentially decreasing with the window length N.

(iii) We consider an equidistant sampling scenario, where the signal is to be estimated from its samples taken equidistantly. The difference between the best estimator for the finite window and the best estimator for infinite horizon decays exponentially. These results are true for both causal estimation and non-causal estimation.

(iv) We also show that given the past values of the process in a finite window of length N, the decay of mutual information between the current value of the random process and the remaining values of the random process decays exponentially with the window length.

The results presented in Item (ii) and Item (iii) can be related to the following findings in the literature: In [131,132], the difference between the infinite horizon and finite horizon causal estimators (the estimator based on the last N values) is found to decay at least exponentially, $f(.) > 0$. In [131,132], no assumptions are explicitly made on the mixing behaviour; [131] assumes particular forms for the spectral density. We approach the problem with methods different from [131,132] and we obtain the following results which were not shown in these works: In these works the best causal estimators are considered, here we consider the truncation of the AR coefficients (Item (ii)), which may be considered as suboptimal causal estimator coefficients. With Item (iii), we consider MMSE estimators based on the equidistant samples in a window of length N. In [131,132] all samples within a finite-length causal window are considered. Our work mentioned in Item (iii) generalizes this to equidistant samples in the finite window and covers the former case where all samples in the window are used in the estimation.

We now introduce some further notation. Let $\mathbb{Z}_+ = \{0, 1, \ldots\}$ denote the set of non-negative integers. The transpose, complex conjugate and complex conjugate transpose of a matrix A is denoted by A^{T}, A^* and $A^\dagger$, respectively.

9.1.1 Mixing rate and decay of the causal autoregressive representation coefficients

In this section, we will relate the decay of the autoregressive representation coefficients of a stationary Gaussian source to its mixing rate. Consider the following autoregressive representation of the source

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + W_t, \quad \forall t. \tag{9.7}$$

We first review the definition of mixing:

Definition 9.1.1. *For a stationary source $\{X_t\}$ the strong or α-mixing coefficient is defined as follows*

$$\alpha(\tau) = \sup_{A \in \mathcal{F}_{-\infty}^{k},\, B \in \mathcal{F}_{k+\tau}^{\infty},\, k \in \mathbb{Z}} |P(A \cap B) - P(A)P(B)|, \tag{9.8}$$

where $\mathcal{F}_{t_1}^{t_2}$ is the following sigma-field

$$\mathcal{F}_{t_1}^{t_2} = \sigma(X_t, t_1 \leq t \leq t_2, t \in Z) \tag{9.9}$$

We will say the process is exponentially mixing if $\alpha(\tau) \leq ce^{-\gamma\tau}$ for some $\gamma > 0$, and some constant $0 < c < \infty$.

Our main result in this subsection is the following:

Theorem 9.1.1. *For an exponentially mixing sequence* $\{X_t\}$*, the AR coefficients* a_k *in* (9.7) *decays at least exponentially*

$$\alpha(\tau) \le ce^{-\gamma\tau} \;\Rightarrow\; |a_k| \le c_2 e^{-\mu k}, \quad \mu < \gamma. \tag{9.10}$$

Proof: We first relate the mixing coefficient to the correlation coefficients associated with $\{X_t\}$.

Lemma 9.1.1. *For a stationary Gaussian process exponentially mixing with coefficient* γ*, decay of correlation function* $|r_x(\tau)|$ *is also upper-bounded exponentially with the same coefficient, i.e.*

$$\alpha(\tau) \le ce^{-\gamma\tau} \;\Rightarrow\; |r_x(\tau)| \le c_1 e^{-\gamma|\tau|}. \tag{9.11}$$

Proof is given in Section B.1.

We now relate the correlation coefficients and the autoregressive representation coefficients. Multiplying both sides of (9.7) with X_{t-l}, $l \ge 0$ and taking expectations yield the following expression

$$\begin{aligned} E[X_t X_{t-l}] &= E[\sum_{k=1}^{\infty} a_k X_{t-k} X_{t-l}] + E[W_t X_{t-l}] && (9.12)\\ &= \sum_{k=1}^{\infty} a_k E[X_{t-k} X_{t-l}] + E[W_t X_{t-l}] && (9.13) \end{aligned}$$

Here (9.13) can be justified as in the proof of Lemma 9.1.3, given in Appendix B.2. We note that if $l = 0$, $E[W_t X_{t-l}] = \sigma_w^2$, and if $l > 0$, $E[W_t X_{t-l}] = 0$.

Hence we have the following semi-infinite system of equations

$$\begin{bmatrix} r_0 & r_1 & r_2 & \dots \\ r_1 & r_0 & r_1 & \\ r_2 & & \ddots & \vdots \\ & \dots & & \dots \end{bmatrix} \begin{bmatrix} 1 \\ -a_1 \\ -a_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix} \tag{9.14}$$

We can write this system of equations as

$$Ta = b \tag{9.15}$$

where T is the infinite Toeplitz matrix with $T = [T_{kl}]_{k,l=0}^{\infty}$, $T_{kl} = r_x(k-l)$, $k, l \in \mathbb{Z}_+$, $a = [1, -a_1, -a_2, \ldots]$ and $b = [\sigma_w^2, 0, \ldots, 0]$.

We note that eigenvalues of finite sections of T satisfy $f_{x,inf} \leq \lambda(T_N) \leq f_{x,\sup}$, see for example [202, Lemma 4.1]. The solution of this system may be found by the following

$$\begin{aligned} \hat{a_N} &= \lim_{N\to\infty} T_N^{-1} b_N & (9.16) \\ &= \lim_{N\to\infty} \sigma_w^2 [T_N^{-1}]_{k0}, & (9.17) \end{aligned}$$

Here $[T_N^{-1}]_{k0}, k \in \mathbb{Z}^{\mathbb{N}}$ denotes the first column of T_N^{-1}. We note that off diagonal decay properties of T^{-1} imply decay properties a_k: for instance if off diagonal elements of T^{-1} were exponentially decaying, a_k would be at least exponentially decaying.

To relate the correlation function to the off-diagonal decay of T^{-1}, we use the following result, which relates the off-diagonal decay properties of T^{-1} to that of T. We note here that the original result is due to [216], this is the form reported in [217].

Definition 9.1.2. *[216] Let $A : l_2(\mathcal{F}) \to l_2(\mathcal{F})$ be an invertible matrix, where $\mathcal{F} = \mathbb{Z}, \mathbb{Z}_+$ or $\{0, \ldots, N-1\}$. A belongs to the space $\mathcal{E}_\gamma$, $\gamma > 0$ if $|A_{kl}| \leq ce^{-\gamma|k-l|}$, for some constant $0 < c < \infty$.*

Lemma 9.1.2. *[216] Let $A : l_2(\mathcal{F}) \to l_2(\mathcal{F})$ be an invertible matrix, where $\mathcal{F} = \mathbb{Z}, \mathbb{Z}_+$ or $\{0, \ldots, N-1\}$. If $A \in \mathcal{E}_\gamma$, then $A^{-1} \in \mathcal{E}_{\gamma_i}$ for some $\gamma_i < \gamma$.*

We now complete our argument: Since decay of $|r_x(\tau)|$ is upper-bounded exponentially, the covariance matrix T has exponential off-diagonal decay, i.e. $T \in \mathcal{E}_\gamma$. By Lemma 9.1.2, T^{-1} also has exponential off-diagonal decay, i.e. $T^{-1} \in \mathcal{E}_\mu$, $\mu < \gamma$. Now by (9.17), $|a_k|$ is also exponentially decaying

$$\alpha(\tau) \leq ce^{-\gamma\tau} \;\Rightarrow\; |a_k| \leq c_2 e^{-\mu k}, \quad \mu < \gamma. \tag{9.18}$$

□

We note that the result of [216] regarding the decay type preservation in inverses (here stated as Lemma 9.1.2) also includes the polynomial type decays. Hence our arguments can be also used to derive conclusions for the polynomial type mixing case, which we skip here for the simplicity of presentation.

9.1.2 Mixing rate and decay of the truncation error in finite-length autoregressive representation

In this section we consider the following truncation of the AR representation coefficients

$$\tilde{X}_t = \sum_{k=1}^{N} a_k X_{t-k} + W_t \quad \forall t. \tag{9.19}$$

A measure of goodness of this representation will be the mean-square error between the truncated representation and the infinite-length representation, which may be written as follows

$$E[||X_t - \tilde{X}_t||^2] \quad = \quad E[|| \sum_{k=N+1}^{\infty} a_k X_{t-k}||^2] \tag{9.20}$$

We will show that this error is upper bounded by an exponentially decreasing function of N without t dependence, i.e. decay of the error introduced by the truncation is at least exponential.

We first note the following result:

Lemma 9.1.3.

$$E[|| \sum_{k=N+1}^{\infty} a_k X_{t-k}||^2] \quad = \quad \sum_{k=N+1}^{\infty} \sum_{l=N+1}^{\infty} a_k a_l r_{k-l}. \tag{9.21}$$

The proof is given in Appendix B.2.

We also have the following result:

Lemma 9.1.4. *$|\sum_{k=N+1}^{\infty}\sum_{l=N+1}^{\infty} a_k a_l r_{k-l}| < \infty$, since $r_x \in l_1(\mathbb{Z})$, and $\{a_k\} \in l_1$.*

The proof is given in Appendix B.3.

We now note that (B.6) can be rewritten as

$$\lim_{L\to\infty} \sum_{k=N+1}^{L} \sum_{l=N+1}^{L} a_k a_l r_{k-l} \quad = \quad \lim_{L\to\infty} \bar{a}_L^{\dagger} T_L \bar{a}_L \tag{9.22}$$

where the length $L > N$ vector $\bar{a}_L$ is defined as

$$\bar{a}_L \quad = \quad [0, \ldots, 0, a_N + 1, \ldots, a_i, \ldots a_L] \tag{9.23}$$

whose first $N+1$ components are zero.

Our main result in this section is the following:

Theorem 9.1.2. *The approximation error for an exponentially mixing sequence with rate γ decays exponentially with some rate 2ν where $\nu > 0$ is strictly smaller than the mixing rate, $\nu < \gamma$*

$$E[||X_t - \tilde{X}_t||^2] \quad \leq \quad \bar{c}e^{-2\nu N}. \tag{9.24}$$

Proof:

$$\begin{aligned}
\sum_{k=N+1}^{L} \sum_{l=N+1}^{L} a_k a_l r_{k-l} = \bar{a}^{\dagger} R_L \bar{a} \quad &= \quad ||T_L^{1/2}\bar{a}||^2 && (9.25)\\
&\leq \quad ||T_L^{1/2}||^2\, ||\bar{a}||^2 && (9.26)\\
&= \quad \lambda_{max}(T_L)\, ||\bar{a}||^2 && (9.27)\\
&\leq \quad f_{x,\sup} \sum_{i=N+1}^{L} |a_i|^2 && (9.28)\\
&\leq \quad c\, f_{x,\sup} \sum_{i=N+1}^{L} e^{-2\nu i} && (9.29)\\
&\leq \quad c\, f_{x,\sup} \frac{e^{-2\nu(N+1)} - e^{-2\nu(L+1)}}{1 - e^{-2\nu}} && (9.30)
\end{aligned}$$

(9.28) follows from the fact that $\sigma_{max}(T_L) \leq f_{x,\sup} < \infty$, where $f_{x,\sup} = ess \sup f$ [202], [132]. (9.29) follows from the fact that AR coefficients decay exponentially, i.e. Theorem 9.1.1. We finally take the limit $L \to \infty$, and absorb all constants into some constant $\bar{c} < \infty$. □

9.1.3 Mixing rate and decay of the truncation error in finite-length MMSE Estimation

In this section, we investigate the decay of the truncation error introduced by using finite-length windows in acausal and causal MMSE estimation of a stationary Gaussian source from its samples and show that this decay is at least exponential.

Finite section method – doubly infinite system: With a doubly infinite system of equations we associate the below finite section method. Consider the infinite dimensional system of equations given by the following equation:

$$Tz = d \tag{9.31}$$

Let P_N be the projection onto the $2N+1$ dimensional space as follows:

$$P_N d = [\ldots, 0, d_{-N}, \ldots, d_{+N}, 0, \ldots] \tag{9.32}$$

Let the associated finite dimensional section of (9.31) be defined by the following expressions:

$$T_N = P_N T (P_N)^T \quad d^N = P_N d, \tag{9.33}$$

Let z^N be the solution of the resulting finite dimensional system of equations:

$$T_N z^N = d^N \tag{9.34}$$

Finite section method – semi-infinite system: Similarly for a semi-infinite system we associate a similar finite section method where, now, P_N is the projection onto the N dimensional space as follows:

$$P_N d = [d_1, \ldots, d_N, 0, \ldots] \tag{9.35}$$

We note that the above projections may be interpreted as mappings to $\mathbb{Z}$ / $\mathbb{Z}_+$, or $2N+1$ / N finite dimensional spaces. The inverses (ex. T_N^{-1}) and such are considered in the finite dimensional spaces.

We note that eigenvalues of finite sections of a Toeplitz matrix T satisfy $f_{x,inf} \leq \lambda(T_N) \leq f_{x,\sup}$, see for example [202, Lemma 4.1]. We also note that the eigenvalues of the principal sub-matrices of T_N (the matrices obtained by taking a certain set of columns and rows from T_N) are also in the range $[f_{x,inf}, f_{x,\sup}]$, since the eigenvalues of principal sub-matrices of a Hermitian matrix are bounded by eigenvalues of the original matrix [148, Theorem 4.3.15].

Theorem 9.1.3. *[217, Thm. 3.1] Let $Tz = d$ be given, where $T_{i,j} = r_{i-j}$ is Hermitian positive definite doubly infinite Toeplitz matrix and let $z^N = T_N^{-1} d^N$ be the finite section solution. If there exist constants c, c' such that*

$$|r_k| \leq c \exp(-\gamma|k|) \quad \textit{and} \quad |d_k| \leq c' \exp(\gamma|k|) \quad \gamma > 0, \tag{9.36}$$

then there exists a γ_1 with $0 < \gamma_1 < \gamma$, and a constant c'' depending only on γ_1 and condition number of T such that

$$||z - z^N|| \leq c'' \exp(-\gamma_1 N) \tag{9.37}$$

This result is also correct for semi-infinite-Toeplitz matrices [217, Remark 3.2] .

We have the following Corollary to Theorem 9.1.3:

Corollary 9.1.1. *Let the setting be the same with previous lemma. Then we have the following:*

$$|d^T z - (d^N)^T z^N| \leq c_1 \exp(-\gamma_1 N) \tag{9.38}$$

for some constant $c_1 > 0$.

Proof:

$$\begin{aligned}
|d^T z - (d^N)^T z^N| &= |d^T z - d^T z^N + d^T z^N - (d^N)^T z^N| && (9.39)\\
&\leq |d^T (z - z^N)| + |(d^T - (d^N)^T) z^N| && (9.40)\\
&= |d^T (z - z^N)| && (9.41)\\
&\leq ||d|| \, ||z - z^N|| && (9.42)\\
&\leq c_1 \exp(-\gamma_1 N) && (9.43)
\end{aligned}$$

Here we have used $|(d^T - (d^N)^T) z^N|$ is zero, since $[z^N]_k = 0$ for $|k| > n$, and $[(d^T - (d^N)^T) z^N]_k = 0$ for $|k| \leq n$. Since $|d_k|$ is exponentially bounded, $c_1 < \infty$.

Let us introduce the following notation to express the MMSE as follows

$$e_t(L_1,L_2) = E[||X_t - E[X_t|X_{k\tau}, k\tau \in [L_1,L_2]||^2] \tag{9.44}$$

We also denote the estimators using infinite number of observations as $\lim_{L_1\rightarrow -\infty} e_t(L_1,L_2) = e_t(-\infty,L_2)$, and
$\lim_{L_2\rightarrow -\infty} e_t(L_1,L_2) = e_t(L_1,\infty)$, and $\lim_{L_1,L_2\rightarrow -\infty} e_t(L_1,L_2) = e_t(-\infty,\infty)$.

Our main result in this section is the following:

Theorem 9.1.4. *Consider an equidistant sampling scenario, where X_t, $t \in \mathbb{Z}$ given, is to be estimated from equidistant samples $\{Y_k\} = \{X_{\tau k}, k \in \mathbb{Z}\}$. For an exponentially mixing sequence with rate γ, the difference in the MMSE introduced by using the samples within a finite window decays exponentially with rate $\gamma_1 > 0$, where $\gamma_1 < \gamma$. More precisely, we have the following:*
i) $e_t(t-L/2, t+L/2) - e_t(-\infty,\infty) \leq c'' \exp(-\gamma_1 L)$.
ii) $e_t(t, t+L) - e_t(t,\infty) \leq c'' \exp(-\gamma_1 L)$.
iii) $e_t(t-L, t) - e_t(-\infty,t) \leq c'' \exp(-\gamma_1 L)$.
iv) $e_t(0, L) - e_t(0,\infty) \leq c'' \exp(-\gamma_1 (L-t))$, $t \in [0,L]$.
c'' and γ_1 take possibly different values for the different cases (i)-(iv).

Proof: We first prove the case (i). The one sided cases (ii)-(iii) are similar to the case (i), and uses the version of [217, Thm. 3.1] (Theorem 9.1.3 above) for semi-infinite Toeplitz matrices. Proof of case (iv), which is based on (ii) is given at the end.

Let $\{Y_k\} = \{X_{\tau k}\}$ be the sampled process. We note that if the Toeplitz covariance matrix associated with the process $\{X_t\}$, $K_X = T(f_x)$ satisfies $K_X = T(f_x) \in \mathcal{E}_\gamma$, then the covariance matrix associated with the process $\{Y_k\}$ satisfies $K_y = T(f_y) \in \mathcal{E}_{\tau\gamma}$. The correlation sequence between X_t and the observations in window centered around t is also exponentially bounded, i.e. $k_{X_tY} = E[X_t(\ldots, X_{l\tau}, X_{(l+1)\tau}, \ldots)] \leq c \exp(-\gamma\tau)$, $c > 0$, where $l = \min\{k, k \in \mathbb{Z}, l\tau \leq t \leq (l+1)\tau\}$.

We recall that the generating function of $K_X = T(f_x)$ is real and assumed to have $f_{x,inf} > 0$. Since rows of $K_Y = T(f_y)$ are obtained by sampling the rows of K_X, the generating function of K_Y, f_y is an aliased form of generating function of f_x. Hence the generating function of K_Y is also bounded below $f_{y,min} > 0$. Hence K_y is a Hermitian positive-definite matrix.

Let the MMSE estimate for estimating X_t from the observations $\{Y_k, k \in \mathbb{Z}\}$ be given as $\hat{X}_t = b^T Y$. b^T can be found by solving the following equation [133, Ch. 6]

$$K_Y \, b = k_{X_iY}^T \tag{9.45}$$

The associated MMSE is given by the following expression [133, Ch. 6]

$$e_t(-\infty, \infty) \;\; = \;\; k_{X_t} - k_{X_tY} K_Y^{-1} k_{X_tY}^T \tag{9.46}$$

Now consider the case where we only use the samples within the $L + 1$ length window around time t, that is we are interested in $e_t(-L/2, L/2)$. Let $\bar{L} = \lceil L/2 \rceil$, where $\lceil . \rceil$ denotes the ceiling function. The coefficients for the finite-length estimator, that is $b^{\bar{L}}$, can be found by solving the following equation

$$K_{Y\bar{L}} b^{\bar{L}} = ({k_{X_tY}}^{\bar{L}})^T, \tag{9.47}$$

As defined through (9.33), $K_{Y\bar{L}}$ and ${k_{X_tY}}^{\bar{L}}$ are the size $(2\bar{L}+1) \times (2\bar{L}+1)$ and $1 \times (2\bar{L}+1)$ finite sections of K_Y, and ${k_{X_tY}}^{\bar{L}}$ respectively. $b^{\bar{L}}$ is the solution to this system of equations with $2\bar{L}+1$ unknowns.

We observe that since $K_y \in \mathcal{E}_{\tau\gamma}$ and $k_{X_tY} = E[X_t(\ldots, X_{l\tau}, X_{(l+1)\tau}, \ldots)] \leq c \exp(-\gamma\tau)$, by Theorem 9.1.3, the norm of the difference between the finite-length estimator and the infinite-length estimator decays exponentially, $||b - b^{\bar{L}}|| \leq c_1' \exp(-\gamma_1 \tau \bar{L}) \leq c_1'' \exp(-\gamma_1 L)$, $c_1'' > 0$, $\gamma_1 < \gamma$.

The MMSE associated with the finite-length estimation is given by following expression

$$e_t(-L/2, L/2) \;\; = \;\; k_{X_i} - (k_{X_iY,\bar{L}}) K_{Y,\bar{L}}^{-1} (k_{X_iY,\bar{L}})^T \tag{9.48}$$

The difference between the errors for infinite horizon case and the finite horizon case is also exponentially bounded as follows

$$\begin{aligned} |e_t(-L/2, L/2) - e_t(-\infty, \infty)| &= |k_{X_iY}^T K_Y^{-1} k_{X_iY} - (k_{X_iY,\bar{L}}) K_{Y,\bar{L}}^{-1} (k_{X_iY,\bar{L}})^T| && (9.49) \\ &\leq c' \exp(-\tau\gamma_1 \bar{L}) && (9.50) \\ &\leq c'' \exp(-\gamma_1 L) && (9.51) \end{aligned}$$

where the first step follows by Corollary 9.1.1. This proves (i).

We now prove (iv). We define the following for $t \in [0, L]$

$$e_t^1 = e_t(t, L) - e_t(0, L), \tag{9.52}$$

$$e_t^2 = e_t(t, \infty) - e_t(0, \infty). \tag{9.53}$$

Hence we have the following:

$$\begin{aligned} e_t(0, L) - e_t(0, \infty) &= e_t(t, L) - e_t(t, \infty) - (e_t^1 - e_t^2) && (9.54) \\ &\leq c_1 \exp(-\gamma_1 (L - t)) - (e_t^1 - e_t^2) && (9.55) \\ &\leq c_1 \exp(-\gamma_1 (L - t)) && (9.56) \end{aligned}$$

Here (9.55) follows from part (iii). (9.56) follows from the fact that $e_t^1 - e_t^2 \geq 0$; the uncertainty reduction due to observing the samples before time t given the observations in the finite window after t ($X_{k\tau}$, $k\tau \in [t+1, L]$, $k \in \mathbb{Z}$) is greater than the uncertainty reduction due to observing the samples before time t given the observations on the semi-infinite line after time t ($X_{k\tau}$, $k\tau \in [t+1, \infty)$). □

We now consider Theorem 9.1.3 again. We note that the fact that T is Hermitian positive-definite is sufficient for $||z^N - z||$ go to zero for any $d \in l_2(\mathbb{Z})$ (or $d \in l_2(\mathbb{Z}_+)$ if T is semi-infinite), see for example the discussion on [217, pg.327]. We note that in that case the expression in Corollary 9.1.1 $|d^T z - (d^N)^T z^N|$ is also guaranteed to go to zero as $N \to \infty$. Theorem 9.1.3, and Corollary 9.1.1 describe how fast the decay is. Hence for any Toeplitz matrix with $f_{min} > 0$, the estimators and the associated errors are guaranteed to converge. The above theorem specifies how fast this convergence is.

9.1.4 Mixing rate and the mutual Information associated with the past values of the process

Lemma 9.1.5. *Given the values of the process in a finite window of length N, the mutual information between the current value of the random process and* *the remaining values of the random process decays exponentially with the window length N for an exponentially mixing sequence*

$$I(X_t; X_{-\infty}^{t-N-1} | X_{t-N}^{t-1}) \leq 0.5 \log(1 + \frac{c \exp(-\gamma_1 N)}{|e_t|}), \tag{9.57}$$

Proof: The mutual information between the observations in the far past $X_{-\infty}^{t-N-1} = [X_{t-N-1}, X_{t-N-2}, \ldots]$ and the current value X_t, given the observations in the finite-length N window $X_{t-N}^{t-1} = [X_{t-1}, \ldots, X_{t-N}]$ is

$$\begin{aligned} I(X_t; X_{-\infty}^{t-N-1} | X_{t-N}^{t-1}) &= h(X_t | X_{t-N}^{t-1}) - h(X_t | X_{-\infty}^{t-N-1}, X_{t-N}^{t-1}) && (9.58) \\ &= h(X_t | X_{t-N}^{t-1}) - h(X_t | X_{-\infty}^{t-1}) && (9.59) \\ &= 0.5 \log(|e_t^N|) - 0.5 \log(|e_t|) && (9.60) \\ &= 0.5 \log(\frac{|e_t^N|}{|e_t|}) && (9.61) \end{aligned}$$

Here $e_t^N = E[(X_t - E[X_t | X_{t-N}^{t-1}])^2]$ and $e_t = E[(X_t - E[X_t | X_{-\infty}^{t-N-1}])^2]$. We note that e_t cannot be zero, because the process is non-deterministic.

We note the following relationship

$$|e_t| |\frac{|e_t^N|}{|e_t|} - 1| = ||e_t^N| - |e_t|| \leq |e_t - e_t^N| \leq c \exp(-\gamma_1 N) \tag{9.62}$$

where the first inequality is due to triangle inequality, and the second inequality is due to Theorem 9.1.4. Here $0 < \gamma_1 < \gamma$. Hence we arrive at the desired result

$$\begin{aligned} I(X_t; X_{-\infty}^{t-N-1} | X_{t-N}^{t-1}) &= 0.5 \log(\frac{|e_t^N|}{|e_t|}) && (9.63) \\ &\leq 0.5 \log(1 + \frac{c \exp(-\gamma_1 N)}{|e_t|}). && (9.64) \end{aligned}$$

□

9.2 Measurement of Stationary Gaussian Sources

We now consider the problem of estimation of a stationary Gaussian source from its samples. We will show how the associated estimation error can be calculated using a sequence of finite dimensional models. We will also show that these errors can be calculated using circulant covariance matrices instead of the original matrices, which are Toeplitz. We will then use this result to find the explicit expression for the MMSE associated with equidistant sampling of a stationary source on $\mathbb{Z}_+$.

We now present the sampling problem we will consider. Let $\mathbb{Z}_+ = \{0, 1, \ldots\}$ denote the set of non-negative integers. Let Γ_N denote the following index set $\Gamma_N = \{0, \ldots, N-1\} \subset \mathbb{Z}_+$. Let $\{X_t\}$ be a real valued zero-mean stationary Gaussian random process defined on $\mathbb{Z}$. We start observing samples of $\{X_t\}$ at $t = 0$ as dictated by the $\{0,1\}$-valued sampling process $\{S_t : t \in \mathbb{Z}_+\}$ under noise. We obtain the following noisy samples

$$Y_t = S_t X_t + Z_t, \quad t \in \mathbb{Z}_+ \tag{9.65}$$

where $\{Z_t \in \mathbb{R} : t \in \mathbb{Z}_+\}$ i.i.d. zero-mean Gaussian noise with variance $0 < \sigma_z^2 < \infty$. We assume that $\{Z_t\}$, $\{X_t\}$ are statistically independent. We assume that $\{S_t\}$ is the equidistant sampling process with the sampling interval τ.

We denote the auto-covariance function with $r_x(t_1 - t_2) = E[X_{t_1} X_{t_2}]$. We assume that $r_x \in l_1(\mathbb{Z})$, i.e. the auto-correlation function is absolutely-summable. Let $f_x(\theta)$ be the power spectral density function defined as

$$f_x(\theta) = \sum_{m=-\infty}^{\infty} r_x(m) e^{-j\theta m}, \quad \theta \in [-\pi, \pi] \tag{9.66}$$

with the inverse relation

$$r_x(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_x(\theta) e^{j\theta m} d\theta, \quad m \in \mathbb{Z}. \tag{9.67}$$

Since we have $r_x \in l_1(\mathbb{Z})$, $f_x(\theta)$ and the inverse relation are well-defined; furthermore, $f_x(\theta)$ is a continuous function of $\theta \in [-\pi, \pi]$, except at a possibly countable number of points [202, Sec. 4]. We also note that since $\{X_t\}$ is a real valued process, $f_x(\theta)$ is an even function. In general, we will again assume that the process is exponentially mixing. For some special cases that will be pointed out through the text, we won't need this assumption. In these cases, the above assumptions on the auto-correlation function will be needed.

The MMSE associated with the estimation of X_t from the observations $Y_l, l \in \Gamma_N$, $\Gamma_N = \{0, \dots, N-1\} \subset \mathbb{Z}_+$ can be expressed as $E[||X_t - E[X_t|Y_l, l \in \Gamma_N]||^2]$. We are interested in the average MMSE associated with estimation of $X_t, t \in \mathbb{Z}_+$ from the observations in $Y_t, t \in \mathbb{Z}_+$. This error may be expressed as the following:

$$\varepsilon \;=\; \lim_{L\to\infty} \frac{1}{L} \sum_{t\in\Gamma_L} \lim_{N\to\infty} E[||X_t - E[X_t|Y_l, l \in \Gamma_N]||^2] \tag{9.68}$$

We observe the following:

Lemma 9.2.1. *The error expression given in* (9.68) *has a finite limit.*

The proof is given in Section B.4.

We now introduce some notation. $[A]_{k,l}$ denotes the k^{th} row, l^{th} column entry of the matrix A. In general, a circulant matrix is determined by its first row and defined by the relationship $C_{tk} = C_{0\,\mathrm{mod}_N(k-t)}$, where rows and columns are indexed by t and k, $0 \leq t\,, k \leq N-1$. We note that the DFT matrix is the diagonalizing transform for all circulant matrices [202]. Let $\sqrt{-1} = j$. The entries of the $N \times N$ DFT matrix A are given by $A_{tk} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}tk}$, where $0 \leq t\,, k \leq N-1$. The transpose, complex conjugate and complex conjugate transpose of a matrix A is denoted by A^{T}, A^* and $A^\dagger$, respectively. The eigenvalues of a $N \times N$ matrix A are shown by $\lambda_k(A)$, $0 \leq k \leq N-1$.

Let $T(f_x)$ denote the semi-infinite Toeplitz matrix associated with the spectrum $f_x(\theta)$. The autocovariance matrix of $\{X_t : t \in \mathbb{Z}_+\}$ is given by $T(f_x)$. Hence the entries of $T(f_x)$ are given by the auto-correlation function $[T(f_x)]_{t_1,t_2} = R_x(t_1 - t_2)$, $t_1, t_2 \in \mathbb{Z}_+$. Let x^N denote the finite-length truncation of $\{X_t\}$, i.e. $x^N = [X_t : t \in \Gamma_N] \in \mathbb{R}^N$. The auto-covariance matrix of x^N is denoted by $K_{x^N} = E[x^N(x^N)^{\mathrm{T}}]$, which is a finite section of the autocorrelation matrix of $\{X_t\}$: $K_{x^N} = T_N(f_x)$. Here $T_N(f_x)$ denote the $N \times N$ finite section of the matrix T with the entries $[T]_{k,l}$, $k, l \in \Gamma_N$

9.2.1 Preliminaries

We now review some definitions and key results that will used in the coming sections. An important ingredient in our study is the exchange of large Toeplitz and circulant matrices. A thorough review for the relationship between large Toeplitz matrices and circulant matrices can be found in [132, 202], where some of the many applications of this relationship in signal processing and information theory are also presented.

We first recall the following definition from [202].

Definition 9.2.1. *[202, Sec. 2.2] The weak norm of a $N \times N$ matrix A is defined by*

$$|A| = (\frac{1}{N}\sum_{i=1}^{N}\sum_{i=1}^{N}|a_{i,j}|^2)^{1/2} = (\frac{1}{N}\operatorname{tr}(A^\dagger A))^{1/2}. \tag{9.69}$$

We also recall that the strong norm $\|A\|$ is defined by the following:

$$\|A\|^2 = \max_k \lambda_k(A^\dagger A). \tag{9.70}$$

A weak asymptotic equivalence of two sequences of matrices is defined as follows:

Definition 9.2.2. *[202, Sec 2.3] Two sequences of $N \times N$ matrices A_N and B_N are "asymptotically equivalent" if*

1. *A_N and B_N are uniformly bounded in strong (and hence in weak) norm: $||A_N||,\ ||B_N|| \leq M < \infty$, N=1, 2, ...,*

2. *$A_N - B_N$ goes to zero in weak norm as $N \to \infty$: $\lim_{N\to\infty} |A_N - B_N| = 0$.*

Asymptotic equivalence of the two sequences A_N and B_N will be abbreviated as $A_N \sim B_N$.

We immediately have the following.

Lemma 9.2.2. *[202, Theorem 2.1] Let $A_N \sim B_N$, and $C_N \sim D_N$. Then (a) $A_N C_N \sim B_N D_N$. (b) $A_N + C_N \sim B_N + D_N$. (c) If $||A_N^{-1}||, ||B_N^{-1}|| \leq K < \infty$, $\forall N$, then ${A_N}^{-1} \sim {B_N}^{-1}$.*

We note the following special cases of the Lemma 9.2.2-(a,b). Let the sampling matrix be defined as $H = \mathrm{diag}(S_t), t \in \mathbb{Z}_+$. Let $H_N = \mathrm{diag}(S_t)$, $t \in \Gamma_N$ denote the $N \times N$ finite section of it. Let $A_N \sim B_N$. Then the following holds a) $H_N A_N H_N^{\mathrm{T}} \sim H_N B_N H_N^{\mathrm{T}}$, b) $A_N + H_N^{\mathrm{T}} H_N \sim B_N + H_N^{\mathrm{T}} H_N$.

We note that if $A_N \sim B_N$, then there exist finite numbers m and M such that $m \leq \lambda_i(A_N), \lambda_i(B_N) \leq M, \quad i = 0, \dots, N-1$. We also recall the following result

Lemma 9.2.3. *[202, Theorem 2.4] If $A_N \sim B_N$ with $m \leq \lambda_i(A_N), \lambda_i(B_N) \leq M, \quad i = 0, \dots, N-1$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N} F(\lambda_t(A_N)) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N} F(\lambda_t(B_N)) \tag{9.71}$$

for an arbitrary function F continuous on $[m, M]$, provided either of the limits exits.

The next result states that sequences of Toeplitz and properly defined circulant matrices are asymptotically equivalent.

Lemma 9.2.4. *[202, Lemma 4.6] Let $T_N(f_x)$ be a sequence of Toeplitz matrices with $[T_N]_{il} = r_x(i-l)$, $r_x \in l_1(\mathbb{Z})$. Then*

$$T_N(f_x) \sim C_N(f_x), \tag{9.72}$$

where $C_N(f_x)$ is the circulant matrix with the eigenvalues $\lambda_k(C_N(f_x)) = f_x(\frac{2\pi k}{N})$, $k = 0, \dots, N-1$.

Another important result in our derivations will be the following.

Lemma 9.2.5. *[202, Theorem 4.2] Let $T_N(f)$ be defined as above. Assume that $f_x(\theta)$ is real. Then for any function F continuous on $[ess\inf f_x,\ ess\sup f_x]$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} F(\lambda_k^N) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(f_x(\theta)) d\theta \tag{9.73}$$

where λ_k^N, $k = 0, \ldots, N-1$ *are the eigenvalues of* $T_N(f_x)$.

9.2.2 Finite dimensional models in MMSE estimation of a stationary source

In this section we discuss finite dimensional models for calculation of error in the MMSE estimation. We first express the error in terms of errors associated with a sequence of finite dimensional models.

Lemma 9.2.6. *Let* $\{X_t\}$ *be an exponentially mixing source. The MMSE can be found by using a sequence of finite dimensional models with dimension* N *and taking the limit as* $N \to \infty$*. More precisely, we have the following*

$$\varepsilon \;=\; \lim_{N\to\infty} \frac{1}{N} E[||x^N - E[x^N|y^N]||^2]. \tag{9.74}$$

where $x^N = [X_t : t \in \Gamma_N] \in \mathbb{R}^N$*, and* $y^N = [Y_t : t \in \Gamma_N] \in \mathbb{R}^N$.

Proof: Let us define the following:

$$e_t(0, N) = E[||X_t - E[X_t|Y_k, k \in \Gamma_N]||^2] \tag{9.75}$$

$$e_t(0, \infty) = \lim_{N\to\infty} e_t(0, N) \tag{9.76}$$

Hence the error defined in (9.68) can be expressed as follows:

$$\begin{aligned}
\varepsilon &= \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} \lim_{N\to\infty} e_t(0, N) && (9.77)\\
&= \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} e_t(0, \infty) - \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} e_t(0, L) + \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} e_t(0, L) && (9.78)\\
&= \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} (e_t(0, \infty) - e_t(0, L)) + \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} e_t(0, L) && (9.79)\\
&= \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} e_t(0, L), && (9.80)
\end{aligned}$$

where (9.80) follows from the fact that the first term goes to zero since we have the following:

$$\begin{aligned}
\lim_{L\to\infty}\frac{1}{L}\sum_{t=0}^{L-1}(e_t(0,L)-e_t(0,\infty)) &\leq \lim_{L\to\infty}\frac{1}{L}\sum_{t=0}^{L-1}c_1\exp(-\gamma_1(L-t)) &(9.81)\\
&= \lim_{L\to\infty}\frac{1}{L}\exp(-\gamma_1 L)\frac{1-\exp(\gamma_1 L)}{1-\exp(\gamma_1)} &(9.82)\\
&= 0 &(9.83)
\end{aligned}$$

where (9.81) follows from case (iv) of Theorem 9.1.4. We note that Theorem 9.1.4 relies on the assumption that the source is mixing.

Hence using (9.80), we can express the error in (9.68) as follows:

$$\begin{aligned}
\varepsilon &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N-1}e_t(0,N) &(9.84)\\
&= \lim_{N\to\infty}\frac{1}{N}E[||x^N-E[x^N|y^N]||^2]. &(9.85)
\end{aligned}$$

□

The MMSE associated with a N dimensional truncation can be expressed in terms of covariance matrices as follows:

$$\begin{aligned}
&E[||x^N-E[x^N|y^N]||^2]\\
&=\mathrm{tr}(K_{x^N}-K_{x^Ny^N}K_{y^N}^{-1}K_{x^Ny^N}^{\mathrm{T}}) &(9.86)\\
&=\mathrm{tr}(K_{x^N}-K_{x^N}H_N^{\mathrm{T}}(H_NK_{x^N}H_N^{\mathrm{T}}+K_{z^N})^{-1}H_NK_{x^N}) &(9.87)\\
&=\mathrm{tr}(T_N(f_x)-T_N(f_x)H_N^{\mathrm{T}}(H_NT_N(f_x)H_N^{\mathrm{T}}+K_{z^N})^{-1}H_NT_N(f_x)) &(9.88)
\end{aligned}$$

where (9.88) follows from the fact that $K_{x,N}=T_N(f_x)$.

We now introduce some shorthand notation. Let us denote the matrix inside the trace expression as a function of the covariance matrix as follows

$$\xi(K_{x,N})=\xi(T_N(f_x)), \qquad (9.89)$$

Hence (9.87) and (9.88) can be written as $\mathrm{tr}(\xi(K_{x^N}))$ and $\mathrm{tr}(\xi(T_N(f_x)))$, respectively. Hence the MMSE we are interested in can be expressed as follows:

$$\begin{aligned} \varepsilon &= \lim_{N\to\infty} \frac{1}{N} E[||x^N - E[x^N|y^N]||^2] && (9.90) \\ &= \lim_{N\to\infty} \frac{1}{N} \mathrm{tr}(\xi(T_N(f_x))). && (9.91) \end{aligned}$$

We now prove that for the purposes of calculating the MMSE associated with length N truncations, one can use circulant matrices instead of Toeplitz matrices.

Lemma 9.2.7. *The limit of the MMSE's associated with length N truncations as $N \to \infty$ can be calculated by using circulant matrices instead of Toeplitz matrices, that is we have the following:*

$$\lim_{N\to\infty} \frac{1}{N} E[||x^N - E[x^N|y^N]||^2] = \lim_{N\to\infty} \frac{1}{N} \mathrm{tr}(\xi(C_N(f_x))) \qquad (9.92)$$

where $C_N(f_x)$ is the $N \times N$ circulant matrix with the eigenvalues $\lambda_k(C_N(f_x)) = f_x(\frac{2\pi k}{N})$, $k = 0, \dots, N-1$.

Proof: The proof follows from the fact that $T_N(f_x) \sim C_N(f_x)$ [202, Lemma 4.6], and a series of application of properties of asymptotically equivalent matrices. We have the following:

$$\begin{aligned} &E[||x^N - E[x^N|y^N]||^2] \\ &= \mathrm{tr}(T_N(f_x) - T_N(f_x)H_N^{\mathrm{T}}(H_N T_N(f_x) H_N^{\mathrm{T}} + K_{z^N})^{-1} H_N T_N(f_x)) && (9.93) \\ &= \mathrm{tr}(T_N(f_x)) - \mathrm{tr}((H_N T_N(f_x) H_N^{\mathrm{T}} + K_{z^N})^{-1} H_N T_N(f_x)^2 H_N^{\mathrm{T}}) && (9.94) \end{aligned}$$

where the last line follows from the identity $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for arbitrary matrices A, B with consistent dimensions.

We have $H_N T_N(f_x) H_N^{\mathrm{T}} + K_{z^N} \sim H_N C_N(f_x) H_N^{\mathrm{T}} + K_{z^N}$ by Lemma 9.2.2, and the fact that $T_N(f_x) \sim C_N(f_x)$. Then the inverses of these matrices are also asymptotically equivalent since the eigenvalues of both inverses are bounded in strong norm for all N due to the relation $K_{z^N} = \sigma_z^2 I_N$. We will then have the following:

$$(H_N T_N(f_x) H_N^{\mathrm{T}} + K_{z^N})^{-1} H_N T_N^2(f_x) H_N^{\mathrm{T}} \sim (H_N C_N(f_x) H_N^{\mathrm{T}} + K_{z^N})^{-1} H_N C_N^2(f_x) H_N^{\mathrm{T}}, \qquad (9.95)$$

by the fact that multiplication of asymptotically equivalent matrices create an asymptotically equivalent sequence of matrices (see for example [202, Thm 2.1]). Now we can apply [202, Theorem 2.4] (Lemma 9.2.3 in the preceding section) with F simply as $F = \lambda_t$. Hence the error can be expressed as follows:

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} E[||x^N - E[x^N|y^N]||^2] \\
&= \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}(\xi(T_N(f_x))) && (9.96)\\
&= \lim_{N\to\infty} \frac{1}{N} (\operatorname{tr}(T_N(f_x)) - \operatorname{tr}((H_N T_N(f_x) H_N^{\mathrm{T}} + K_{z^N})^{-1} H_N T_N(f_x)^2 H_N^{\mathrm{T}})), && (9.97)\\
&= \lim_{N\to\infty} \frac{1}{N} (\operatorname{tr}(C_N(f_x)) - \operatorname{tr}((H_N C_N(f_x) H_N^{\mathrm{T}} + K_{z^N})^{-1} H_N C_N(f_x)^2 H_N^{\mathrm{T}})) && (9.98)\\
&= \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}(\xi(C_N(f_x))) && (9.99)
\end{aligned}$$

□

Theorem 9.2.1. *Let $\{X_t\}$ be an exponentially mixing source. The MMSE for estimating X_t from the observations $Y_t = S_t X_t + Z_t$, $t \in \mathbb{Z}_+$ with S_t and Z_t as described before is given by the following expression*

$$\begin{aligned}
\varepsilon &= \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} \lim_{N\to\infty} E[||X_t - E[X_t|Y_l, l \in \{0,\dots,N-1\}]||^2] && (9.100)\\
&= \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}(\xi(C_N(f_x))) && (9.101)
\end{aligned}$$

where $C_N(f_x)$ is the $N \times N$ circulant matrix with the eigenvalues $\lambda_k(C_N(f_x)) = f_x(\frac{2\pi k}{N})$, $k = 0, \dots, N-1$.

Proof: The result follows from Lemma 9.2.6 and Lemma 9.2.7. □

We wish to emphasize that one should be careful while attempting to replace Toeplitz matrices with associated circulant matrices; the legitimacy of such an exchange depends crucially on the application. Some discussion along this direction is presented in [218]. Here we have showed that for the purposes of the noisy sampling problem at hand, a Toeplitz and a circulant matrix are operationally equivalent. In Section 9.2.3, we will use this result to find an explicit expression for the MMSE associated with equidistant sampling.

Remark 9.2.1. *We have shown for the purpose of calculating the MMSE on $\mathbb{Z}_+$, one can assume that the covariance matrix is circulant. Hence the geometric spread of uncertainty is given by the DFT matrix, which is the diagonalizing unitary transform for all circulant matrices (see for example [202]). This result implies that for the purposes of calculating the MMSE for infinite dimensional stationary sources on $\mathbb{Z}_+$ with a given power spectrum , the uncertainty can be spread in the measurement domain in only one way; the way as dictated by the DFT matrix.*

Remark 9.2.2. *If we were concerned with sources over the entire line, i.e. $\mathbb{Z}$, this result, i.e. one can use the DFT matrix to calculate the MMSE, could have been natural, since in this case using stationarity of the field, Fourier transform methods become easily applicable to calculate MMSE for equidistant sampling. (This approach is illustrated in Section B.5.) The fact that in our case the source is considered on $\mathbb{Z}_+$ makes the result more intriguing.*

Remark 9.2.3. *We now make an observation related to the finite dimensional models in stationary signal models. Circularly wide-sense stationary signals are considered to be a natural way to model wide-sense stationary signals in finite dimension. In this case, by definition, the covariance matrix is circulant. The result of this lemma suggest that circularly w.s.s. signals may be more than an analogy of w.s.s. signals; there is an operational relationship between these two. The lemma shows that for the purposes of calculating the MMSE one may use the sequence of associated circulant matrices (hence the c.w.s.s. models) instead of the original model. In Section 9.2.3, we use this lemma to find the MMSE associated with the equidistant sampling of a stationary source using the result for the equidistant sampling of a circularly w.s.s. signal.*

9.2.3 MMSE estimation of a stationary Gaussian source from its equidistant samples

We now present the MMSE associated with estimation of a stationary Gaussian Source from its equdistant samples on $\mathbb{Z}_+$. We prove the result by the following method: we first use a finite dimensional model and find the associated error; then using Theorem 9.2.1, we extend this result to the infinite dimensional source.

We now compare our error result with the following results where the signals defined on $\mathbb{R}$ are considered: In [122], the mean-square error of approximating a possibly non-bandlimited wide-sense stationary (w.s.s.) signal using sampling expansion is considered and a uniform upper bound in terms of power outside the bandwidth of approximation is derived. Here we are interested in the average error over all points of the sequence on $\mathbb{Z}_+$. Our method of approximation of the signal is possibly different, since we use the MMSE estimator. As a result our error expression also makes use of the shape of the power spectrum. Another related result is [116]'s result which states that a w.s.s. signal is determined linearly by its samples if some set of frequencies containing all of the power of the process is disjoint from each of its translates where the amount of translate is determined by the sampling rate. We note that the notion of such a set of frequencies provides a generalization of the standard band-limitedness (low-pass, band-pass etc.) concept. Here for a w.s.s. signal defined on $\mathbb{Z}_+$, under a set of conditions, we arrive at the same conclusion: if there is such a set of frequencies, the signal will be linearly determined from its samples, hence the MMSE will be zero. Moreover, we provide the MMSE expression for the other cases where the MMSE is not exactly zero. Our expression shows explicitly how the signal and noise spectral densities contribute to the error.

Let us recall the equidistant sampling problem. We consider the problem of estimation of $\{X_t, t \in \mathbb{Z}_+\}$ from its equidistant noisy samples $\{Y_t, t \in \mathbb{Z}_+\}$. Let the samples be taken every τ points, i.e. $Y_t = S_t X_t + Z_t$, where $S_t = 1$, if $t = \tau k$, $k \in \mathbb{Z}_+$ otherwise zero. As before, $\{Z_t, t \in \mathbb{Z}_+\}$ is i.i.d. zero-mean Gaussian noise with variance $0 < \sigma_z^2 < \infty$. We assume that $\{Z_t\}$, and $\{X_t\}$ are statistically independent.

Our main result in this section is the following:

Theorem 9.2.2. *Consider the MMSE estimation of* $\{X_t, t \in \mathbb{Z}_+\}$ *from* $\{Y_t, t \in \mathbb{Z}_+\}$ *as described above. The estimation error is given by the following expression:*

$$E[\lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{N-1} (X_t - \hat{X}_t)^2] = \lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} \lim_{N\to\infty} E[||X_t - E[X_t|Y_l, l \in \Gamma]||^2] \quad (9.102)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} (f_x(\theta) - \frac{1}{\tau^2} \sum_{i=0}^{\tau-1} \frac{(f_x(\frac{\theta+2\pi i}{\tau}))^2}{\frac{1}{\tau}\sum_{l=0}^{\tau-1} f_x(\frac{\theta+2\pi l}{\tau}) + \sigma_z^2}) d\theta. \quad (9.103)$$

Proof: This proof is based on a sequence of finite dimensional models. We use the result for equidistant sampling of a circularly wide-sense stationary signal defined on the finite interval $[0, \dots, N-1]$ to find the MMSE associated with equidistant sampling of a stationary signal on $\mathbb{Z}_+$. As pointed out in Remark 9.2.3, Theorem 9.2.1 shows that there is an operational relationship between these two models: under conditions of the theorem, circulant matrices, hence circularly w.s.s. models, can be used to evaluate the MMSE associated with sampling of a stationary processes on $\mathbb{Z}_+$.

Let us assume that the conditions of Theorem 9.2.1 hold. Theorem 9.2.1 states that the MMSE can be expressed as follows:

$$\lim_{L\to\infty} \frac{1}{L} \sum_{t=0}^{L-1} \lim_{N\to\infty} E[||X_t - E[X_t|Y_l, l \in \Gamma]||^2] = \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}(\xi(C_N(f_x))) \quad (9.104)$$

where $C_N(f_x)$ is the $N \times N$ circulant matrix with the eigenvalues $\lambda_k(C_N(f_x)) = f_x(\frac{2\pi k}{N})$, $k = 0, \dots, N-1$. Without loss of generality, we will assume that $M = N/\tau \in \mathbb{Z}$, and take the limits accordingly. (Since (9.68) converges, any subsequence converges to the same limit.) We recall that $\operatorname{tr}(\xi(C_N(f_x)))$ can be expressed as follows (9.88), (9.89)

$$\operatorname{tr}(C_N(f_x)) - \operatorname{tr}((H_N C_N(f_x) H_N^{\mathrm{T}} + K_{z^N})^{-1} H_N C_N(f_x)^2 H_N^{\mathrm{T}})) \quad (9.105)$$

Here H_N is the sampling matrix. We note that the error does not change whether we consider the measurements that are zero or discard them. In other words, the error does not change whether H_N is interpreted as the $N \times N$ matrix with 0 rows for the unmeasured components ($H_N = \mathrm{diag}(S_t)$, $t = 0, \ldots, N-1$), or it is a $M \times N$ matrix formed with only the nonzero rows. For convenience we will use the latter.

Let us first consider the first term in (9.105) as $N \to \infty$

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \mathrm{tr}(C_N(f_x)) &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \lambda_k(C_N(f_x)) && (9.106)\\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \lambda_k(T_N(f_x)) && (9.107)\\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f_x(\theta) d\theta && (9.108)\\
&= r_x(0) && (9.109)
\end{aligned}$$

where in (9.107) we went back to using the asymptotically equivalent Toeplitz matrix $C_N(f_x) \sim T_N(f_x)$ [202, Theorem 2.4] (Lemma 9.2.3). (9.108) follows from [202, Theorem 4.2] (Lemma 9.2.5).

To evaluate the second term in (9.105), we use the following facts a) $C_{N/\tau}(\bar{f}_x) + \sigma_z^2 I_{N/\tau} = C_{N/\tau}(\bar{f}_x + f_z)$, where $f_z(\theta) = \sigma_z^2$ for $\theta \in [-\pi, \pi]$; b) $H_N C_N(f_x) H_N^{\mathrm{T}}$ is a circulant matrix with dimension $N/\tau \times N/\tau$ and the eigenvalues

$$\begin{aligned}
\lambda_k(H_N C_N(f_x) H_N^{\mathrm{T}}) &= \frac{1}{\tau} \sum_{i=0}^{\tau-1} \lambda_{i\frac{N}{\tau}+k}(C_N(f_x)), \quad k = 0, \ldots, N/\tau - 1 && (9.110)\\
&= \frac{1}{\tau} \sum_{i=0}^{\tau-1} f_x(\frac{2\pi(i\frac{N}{\tau} + k)}{N}) && (9.111)\\
&= \frac{1}{\tau} \sum_{i=0}^{\tau-1} f_x(\frac{2\pi i}{\tau} + \frac{2\pi k}{N}) && (9.112)
\end{aligned}$$

Here (9.110) is based on the fact that equidistant column and row sampling of the DFT matrix gives another DFT matrix with a smaller dimension (These eigenvalues are calculated explicitly in (A.12) in Section A.1). (9.111) follows from the fact that $\lambda_t(C_N(f_x)) = f_x(\frac{2\pi t}{N})$, $t \in 0, \ldots, N-1$.

We can now express the second term in (9.105) as follows

$$\begin{aligned}
&\mathrm{tr}((H_N C_N(f_x) H_N^{\mathrm{T}} + \sigma_z^2 I_{N/\tau})^{-1} H_N C_N(f_x) C_N(f_x) H_N^{\mathrm{T}}) \\
&= \mathrm{tr}(H_N C_N(f_x) H_N^{\mathrm{T}} + \sigma_z^2 I_{N/\tau})^{-1} H_N C_N(f_x^2) H_N^{\mathrm{T}}) && (9.113)\\
&= \mathrm{tr}((C_{N/\tau}(\bar{f}_x) + \sigma_z^2 I_{N/\tau})^{-1} H_N C_N(f_x^2) H_N^{\mathrm{T}}) && (9.114)\\
&= \mathrm{tr}((C_{N/\tau}(\bar{f}_x + \sigma_z^2))^{-1} C_{N/\tau}(\hat{f}_x)) && (9.115)\\
&= \mathrm{tr}((C_{N/\tau}((\bar{f}_x + f_z))^{-1}) C_{N/\tau}(\hat{f}_x)) && (9.116)\\
&= \mathrm{tr}(C_{N/\tau}(\frac{\hat{f}_x}{\bar{f}_x + f_z})) && (9.117)
\end{aligned}$$

where $\bar{f}_x = \frac{1}{\tau}\sum_{i=0}^{\tau-1} f_x(\frac{\theta+2\pi i}{\tau})$ and $\hat{f}_x = \frac{1}{\tau}\sum_{i=0}^{\tau-1} f_x^2(\frac{\theta+2\pi i}{\tau})$. In (9.114) and (9.115), we have used the observation (b) and (a) given above, respectively. We have used the following property of the circulant matrices $C_N(f_1)C_N(f_2) = C_N(f_1 f_2)$ in (9.113) and (9.117), and $C_N^{-1}(f_1) = C_N(1/f_1)$ for $\operatorname{ess\,inf} f_1 > 0$ in (9.117) [202, Lemma 4.5].

Hence as $N \to \infty$ the second term in (9.105) can be expressed as follows

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} \mathrm{tr}(C_{N/\tau}(\frac{\hat{f}_x}{\bar{f}_x + \sigma_z^2})) && (9.118)\\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{M-1} \lambda_t(C_{N/\tau}(\frac{\hat{f}_x}{\bar{f}_x + \sigma_z^2})) && (9.119)\\
&= \lim_{M\to\infty} \frac{1}{\tau M} \sum_{t=0}^{M-1} \lambda_t(C_{N/\tau}(\frac{\hat{f}_x}{\bar{f}_x + \sigma_z^2})) && (9.120)\\
&= \frac{1}{\tau^2} \sum_{i=0}^{\tau-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(f_x(\frac{\theta+2\pi i}{\tau}))^2}{\frac{1}{\tau}\sum_{l=0}^{\tau-1} f_x(\frac{\theta+2\pi l}{\tau}) + \sigma_z^2}) d\theta. && (9.121)
\end{aligned}$$

Here, similar to the derivation of (9.108), we have used the fact that $C_N(f_x) \sim T_N(f_x)$, and [202, Theorem 2.4] (Lemma 9.2.3) together with [202, Theorem 4.2] (Lemma 9.2.5).

We note that both of the expressions in (9.108) and (9.121) are finite. We putting these into (9.105), together with (9.104), we obtain the expression in (9.103), as desired. □

9.2.4 Discussion on autoregressive sources

In this section, we provide an alternative form of Theorem 9.2.1 for stationary autoregressive sources. Suppose X_t is a stationary Gaussian AR source defined by

$$X_t = \begin{cases} -\sum_{k=1}^{\infty} a_k X_{t-k} + W_t, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases} \tag{9.122}$$

where W_t's are i.i.d real valued zero-mean Gaussian random variables with variance $\sigma_W^2 = 1$ with $\sum_{k=0}^{\infty} |a_k| < \infty$. With the convention $a_0 = 1$, we assume that the zeros of the polynomial $\sum_{k=0}^{\infty} a_k z^{-k}$ lie inside the unit circle, so that the process is asymptotically stationary.

We note that although the process is asymptotically stationary, the covariance matrix of the process is not exactly Toeplitz; due to the initialization at $t = 0$. Although we can use the fact that the sequence of the covariance matrices is asymptotically similar to a sequence of Toeplitz matrices, and use Theorem 9.2.1 directly, we adopt a slightly different approach which highlights some of the intrinsic properties of the AR source.

Lemma 9.2.8. *The MMSE for estimating X_t from the observations $Y_t = S_t X_t + Z_t$ with S_t and Z_t as described above can be expressed as follows:*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} \frac{1}{\lambda_t (C_{x,N} + H_N^{\mathrm{T}} H_N)} \tag{9.123}$$

where $C_{x,N}$ is the circulant matrix with eigenvalues $\lambda_k(C_{x,N}) = |a(k2\pi/N)|^2$ where $a(\theta) = \sum_{k=0}^{\infty} a_k e^{ik\theta}$.

Proof:

The inverse covariance matrix of this AR Gaussian source is asymptotically equivalent to a Toeplitz matrix with spectral density $|a(\theta)|^2$, i.e. $K_{x,N}^{-1} \sim T(|a(\theta)|^2)$ [202, Thm. 6.2]. We also note that $T_N(|a(\theta)|^2) \sim C_N(|a(\theta)|^2)$ under the condition $\sum_{k=-\infty}^{\infty} |[T_N]_{0,k}| < \infty$ [202, Lemma 4.6]. Hence $K_{x,N}^{-1} \sim C_N(|a(\theta)|^2)$. By [132, Section 1.13], $\int_{-\pi}^{\pi} \log |a(\theta)|^2 d\theta > -\infty$, hence we must have

$\operatorname{ess\,inf} |a(\theta)|^2 = m > 0$. Thus, the eigenvalues of T_N are guaranteed to be away from zero.

We note that when $T_N(f_x)$ is non-singular, the error expression can be also written as

$$\begin{aligned} E[||x^N - E[x^N|y^N]||^2] &= \operatorname{tr}\left((T_N^{-1}(f_x) + \frac{1}{\sigma_z^2} H_N^{\mathrm{T}} H_N)^{-1}\right) && (9.124) \\ &= \sum_{k=0}^{N-1} \frac{1}{\lambda_k(T_N^{-1}(f_x) + \frac{1}{\sigma_z^2} H_N^{\mathrm{T}} H_N)} && (9.125) \end{aligned}$$

This follows from the fact that $K_{x,N}$ and $K_x = \sigma_z^2 I_N$ are nonsingular and the Sheerman-Morrison-Woodbury identity, which has the following form for our case (see for example [203] and the references therein)

$$K_1 - K_1 A^\dagger (A K_1 A^\dagger + K_2)^{-1} A K_1 \;=\; (K_1^{-1} + A^\dagger K_2^{-1} A)^{-1}, \qquad (9.126)$$

where K_1 and K_2 are nonsingular. Since eigenvalues of $T_N(f_x)$ are away from zero, we can apply [202, Theorem 2.4] to the error expression in (9.125) with $K_{x,N}^{-1} + H_N^{\mathrm{T}} H_N \sim C_N(|a(\theta)|^2) + H_N^{\mathrm{T}} H_N$ with $F(\lambda_t) = 1/\lambda_t$. □

Discussion On Nonstationary AR Sources: Even when the AR source is nonstationary, the inverse covariance matrix satisfies $K_{x,N}^{-1} \sim T(|a(\theta)|^2)$. But in this case, the eigenvalues $\lambda_t(K_{x,N}^{-1})$ approach zero [219, 220] . In general we only know $\lambda_{min}(K_{x,N}^{-1} + H_N^{\mathrm{T}} H_N) \geq \lambda_{min}(K_{x,N}^{-1}) + \lambda_{min}(H_N^{\mathrm{T}} H_N) = \lambda_{min}(K_{x,N}^{-1})$. On the other hand, the function $F(x) = 1/x$ is discontinuous at $x = 0$, making direct application of Theorem 2.4 of [202] impossible.

Nevertheless, some aspects of sampling of non-stationary sources are well-understood. Consider a causal estimation scenario where a Kalman filter is used. Let us consider the case of Bernoulli sampling, with success rate p. The estimation error is unbounded if $|\lambda_{max}(A)|^2(1-p) > 1$ [221], where p is the A is $q \times q$ state transition matrix obtained by expressing the finite dimensional AR source as a vector Markov source. Since the largest eigenvalue of the state transition matrix provides a measure for boundedness of estimation, it could be associated with degree of overall correlatedness of the field.

9.2.5 First order stationary Markov source and Bernoulli sampling strategy

We now consider a different sampling scheme for measurement of a particular family of stationary sources: we address the problem of estimating a first order stationary Markov Source on Z_+ under Bernoulli sampling scheme. Under Bernoulli sampling scheme, the value of the random sequence at a point is observed with probability p independent of the other points.

Our signal model can be expressed as follows

$$X_t = a_1 X_{t-1} + W_t, \quad t \geq 0 \tag{9.127}$$

where $X_{-1} = 0$, W_t is zero mean i.i.d Gaussian source with variance σ_w^2. Let $E[X_{t_1} X_{t_2}] = r_x(t_1 - t_2) = a_1^{|k|} \frac{\sigma_w^2}{1-|a_1|^2} = r^{|k|}$, where for notational convenience we fix $\frac{\sigma_w^2}{1-|a_1|^2} = 1$, and denote a_1 with $a_1 = r$.

We consider the following measurement scenario: We sample $\{X_t\}$ as dictated by the i.i.d. $\{0, 1\}$-valued sampling process $\{S_t\}$, that is we observe Y_t formed as follows:

$$Y_t = S_t X_t, , \quad t \geq 0 \tag{9.128}$$

where $\{S_t\}$ is a sequence of independent Bernoulli random variables, i.e. $S_t = 1$ with probability p, $S_t = 0$ with probability $1-p$. We assume that $\{X_t\}$ and $\{S_t\}$ are statistically independent, and the realization of the sampling process $\{S_t\}$ is known at the estimator. Hence the information available to the estimator can expressed as the following sequence $\{I_t, t \in \mathbb{Z}\}$, where $I_t = [S_t, Y_t]$.

Lemma 9.2.9. *The estimation error associated with the above model can be expressed as follows*

$$\begin{aligned} \varepsilon(p, r) &= \lim_{L \to \infty} E[\frac{1}{L} \sum_{t=0}^{L} [(X_t - E[X_t | I_t, t \in \mathbb{Z}])^2] && (9.129) \\ &= -1 + p - \frac{2p}{1-r^2} + 2pE[\frac{T_1}{1-|r|^{2T_1}}] && (9.130) \\ &= -1 + p - \frac{2p}{1-r^2} + 2p^2 \sum_{k=0}^{\infty} \frac{r^{2k}}{(1-(1-p)r^{2k})^2} && (9.131) \end{aligned}$$

where $1 > p > 0$*, and* $1 > |r| > 0$*. Here* T_1 *is the time of the first success of Bernoulli sampling that is* $T_1 = \min(k > 0 : S_k = 1)$*.*

The proof is given in Section B.6.

Corollary 9.2.1. *The above error is lower bounded as follows:*

$$\varepsilon(p, r) \geq -1 + p - \frac{2p}{1 - r^2} + 2\frac{1}{1 - |r|^{2/p}} \tag{9.132}$$

Proof: We observe that $f(T_1) = T_1/(1 - r^{2T_1})$ is a convex function of $T_1 \geq 0$. This can be proven, for instance by using the fact that if the second derivative of a function defined on a convex domain is non-negative, the function is convex [151, Sec. 3.1.4]. Hence we have

$$E[\frac{T_1}{1 - |r|^{2T_1}}] \geq \frac{1/p}{1 - |r|^{2/p}} \tag{9.133}$$

where we have used the fact that $E[T_1] = 1/p$ and Jensen's Inequality [151, Sec. 3.1.8]. The result follows by (9.130).

9.3 Conclusions

In this chapter we have worked on finite-length models and representations of stationary Gaussian sequences. We have discussed the decay of the error in finite-length representations/estimation of these sources. We have showed that for exponentially mixing sequences, for various representations and estimators, the error difference between using a finite-length representation and an infinite-length representation is upper bounded by an exponentially decreasing function of the finite window length. For stationary Gaussian signals, it is known that the presence of strong mixing may prevent a signal from being precisely bandlimited, but otherwise puts comparably loose restrictions on the spectral density, hence the effective bandwidth and the entropy. Nevertheless, the above results shows that mixing rate is pertinent to the geometric spread of uncertainty in the signal in the sense that it determines how the error difference between finite and infinite-length representations decays. In the second part, we have used the finite dimensional circularly wide-sense stationary signal model to find MMSE associated with noisy equidistant sampling of stationary Gaussian source on $\mathbb{Z}_+$. Our expression explicitly shows how the signal and noise spectral densities contribute to the error.

Chapter 10

Conclusions

In this thesis, we have studied on a family of signal representation and recovery problems under various measurement restrictions. In each of the problems formulated, we have focused on different aspects of information transfer in the measurement process. In particular we paid attention to different forms of finiteness, such as finite number of measurements or finite amplitude accuracy in measurements.

Our work has contributed to better understanding of information theoretic relationships in physical fields, in particular propagating waves, such as optical fields. Although these fields are usually represented by functions of continuous variables, in effect they carry a finite amount of information. This finiteness is intrinsically related to the finiteness of the energy and the specified non-zero error tolerance or noise in the system. To quantify how these come into the picture in recovery of the signal from its measurements, we have set ourselves the goal of representing the field as efficiently as possible; that is, with as small a number of samples as possible or as small a number of bits as possible.

We have formulated a family of optimal measurements problems to answer these questions. In the first one of these, we have focused on the finite number of samples restriction. We have investigated the optimal sampling interval in order to represent the field with as low error as possible for a given number of samples. Here we have focused on the following two trade-offs i) the trade-offs between the achievable error and the number of samples, ii) the trade-off between the spatial coverage and the frequency coverage of the samples. Our results reveal how, for a given number of samples, we should choose the space and frequency coverage. That is, we have illustrated whether it is better to take more closely spaced samples (with wider frequency coverage but smaller spatial coverage), or to take more distant samples (with smaller frequency coverage but larger spatial coverage). One of our contributions is to show that in certain cases, sampling at rates different than the Nyquist rate is more efficient, and to find the optimal sampling rates.

Motivated by the fact that we often use digital systems to process information, we have also considered the problem of representing a signal with its samples using as small a number of bits as possible. Formulating and solving this problem is one of the major contributions of this thesis. Here we focused on various trade-offs in the representation of random fields, mainly: i) the trade-offs between the achievable error and the cost budget, ii) the trade-offs between the accuracy, spacing, and number of samples. In contrast to common practice which often treats sampling and quantization separately, we have explicitly focused on the interplay between limited spatial resolution and limited amplitude accuracy. Under a given cost budget, we have investigated whether it is better to take a higher number of samples with relatively lower cost per sample (hence with lower amplitude accuracy), or a lower number of samples with relatively higher cost per sample (hence with higher amplitude accuracy).

We have considered two versions of the above problem: i) the uniform version where the samples are equidistantly spaced, and all the samples are taken with the same level of measurement accuracy, ii) the non-uniform version where the sample locations can be freely chosen, and need not be equally spaced from each other. Furthermore, the measurement accuracy of each sample can vary from sample to sample. For the first, uniform version, we have found the optimal number of samples and sampling interval under a given cost budget in order to recover the field with as low error as possible. We have again illustrated that, in some cases, sampling at rates different than the Nyquist rate is more efficient, and found the optimum sampling intervals. We note that although one may expect to compensate for the limited accuracy of the samples by oversampling, the precise relationships between the sampling parameters and the reconstruction error are not immediately evident. Here we gave quantitative answers to this question by determining the optimal sampling parameters and the resulting performance bounds for the best achievable error for a given bit budget. The second, general, non-uniform version represents maximum flexibility in choosing the sampling strategy; the number, locations and accuracies are all free variables. Here we have found the optimal values of these in order to achieve the lowest error values possible under a cost budget. Here we have illustrated how one can exploit the better optimization opportunity provided by the flexibility of choosing these parameters freely, and obtain tighter optimization of the error-cost curves. Our results illustrate that sampling with this more general scheme provides greater improvements when the uncertainty of the signal is not spread uniformly in space (that is, when the uncertainty reduction due to sampling of the field at different parts of the space are substantially different).

The degree of coherence, which is defined as a measure of total correlatedness of an optical field, is a concept of central importance in statistical optics. In all of the above work, this concept played a major role. We have systematically investigated the effect of coherence, as well as the effect of signal-to-noise ratio on cost-error trade-offs and optimal cost allocations.

The field at one part of a system is not independent from the field at another part of the system. In other words, knowledge of the field at one part of the system gives us a certain degree of information about the field at other parts. Thus we also considered the case where the signal is represented by samples taken after the signal passes through a linear system. For the case of Gaussian-Schell model beams, when there is no noise, we have shown that samples of the output field are as good as samples of the input field, for the broad class of first-order optical systems. This class includes arbitrary concatenations of lenses, mirrors and sections of free space, as well as quadratic graded-index media. We have shown that the cost-error trade-off curves obtained turn out to be the same as those obtained for direct sampling of the input, and the optimum sampling points can be found by a simple scaling of the direct sampling results.

Although various aspects of the problem of sensing of physical fields have been widely studied as estimation problems, much of this work has loose connections with both the underlying physical phenomena and the physical aspects of the sensors employed. There seems to be a disciplinary boundary between the works that look at this problem from an estimation or coding point of view and a physical sciences point of view. By utilizing a cost budget approach to measurement of these fields, our work has contributed to bridging this gap, and has helped us to better understand the information theoretic relationships in physical fields and their measurement from a broader perspective.

We have also considered an application of the above cost based measurement framework to super-resolution problems; and have studied the effect of limited amplitude resolution (pixel depth). Unlike most previous work, amplitude resolution was considered as a just as important aspect of the super-resolution problem as spatial resolution. The cost budget approach mentioned above made it possible to study this problem systematically. We have studied the success of different measurement strategies where the resolution in amplitude (pixel depth), resolution in space (upsampling factor) and the number of low resolution images are variable. The proposed framework has revealed great flexibility in terms of spatial-amplitude resolutions in super-resolution problem. We have seen that it is possible to reach target visual qualities with different measurement scenarios including varying number of images with different amplitude and spatial resolutions.

Our study of the measurement problems described above suggests that although the optimal measurement strategies and signal recovery performance depends substantially on total uncertainty of the signal, the geometry of the spread of uncertainty in measurement space also plays an important role in the signal recovery problem. We note that the concepts that are traditionally used in the signal processing and information theory literatures as measures of dependency or uncertainty of signals (such as the degree of freedom or the entropy) mostly refer to total uncertainty of the signal. In the second part of this thesis, we have formulated various problems investigating different aspects of the relationship between total uncertainty of the signal and its spread in the measurement domain, and their effects on signal recovery performance. We have considered this problem in a mean-square error setting under the assumption of Gaussian signals. This framework makes it possible to approach the problem in terms of second-order statistics. Entropy, which is a measure of total uncertainty, solely depends on the eigenvalue spectrum of the covariance matrix; hence the concept is blind to the coordinate system in which the signal will be measured. The spread of uncertainty in the measurement domain depends on both the total uncertainty, and the coordinate system the signal will be measured. This line of study also relates to the compressive sensing paradigm. Contrary to the deterministic signal models commonly employed in compressive sensing, here we work in a stochastic framework based on the Gaussian vector model and minimum mean-square error (MMSE) estimation; and investigate the spread of the eigenvalue distribution of the covariance matrix as a measure of sparsity. In our framework, we have assumed that the covariance matrix of the signal, hence location of support of the signal is known during estimation.

We have first investigated the relationship between the following two concepts: degree of coherence of a random field as defined in optics and coherence of bases as defined in compressive sensing. Degree of coherence of a basis is a concept from compressive sensing which provides a ranking of bases. In compressive sensing the good bases are the ones where "each measurement picks up a little information about each component" [181]. Coherence of bases is a measure of this property. We have observed that these concepts are named exactly the same, but attributes of different things (bases and random fields), and yet enable similar type of conclusions (good signal recovery performance). One of the main contributions of this study is to explore the relationship between these concepts, and demonstrate that the similarities are more than a coincidence. Our study proposes the fractional Fourier transform (FRT) as an intuitively appealing and systematic way to generate bases with varying degree of coherence: we illustrate that by changing the order of the FRT from 0 to 1, it is possible to generate bases whose coherence ranges from most coherent to most incoherent. We have also developed an estimation based framework to quantify coherence of random fields and have illustrated that what this concept quantifies is not just a repetition of what more traditional concepts like the degree of freedom or the entropy does. Moreover, we have shown that by using these different bases with different FRT orders, it is possible to generate statistics for fields with varying degree of coherence. Hence we also propose the FRT as a systematic way of generating the statistics for fields with varying degree of coherence.

Our above work can be interpreted as an investigation of basis dependency of the MMSE under random sampling. We have also studied this problem from an alternative perspective. We have considered the transmission of a Gaussian vector source over a multi-dimensional Gaussian channel where a random or a fixed subset of the channel outputs are erased. We have focused on the setup where the only encoding operation allowed is a linear unitary transformation on the source. For such a setup, we have investigated the MMSE performance both in average and in terms of guarantees that hold with high probability as a function of system parameters. For the average error criterion necessary conditions for optimal unitary encoders are established, and explicit solutions for a class of settings are presented. Although there are observations (including evidence provided by the compressed sensing community) that may suggest the result that the discrete Fourier transform (DFT) matrix may be indeed an optimum unitary transformation for any eigenvalue distribution, we provide a counterexample. Most of this work is based on a measurement model where each component is erased independently and with equal probability. This measurement model is of central importance in compressive sensing. Our work also contributes to the understanding of the MMSE performance of such measurement schemes under noise. For guarantees that hold with high probability, we have first considered the case where the covariance matrix has a flat eigenvalue distribution (nonzero eigenvalues all have the same value). We have illustrated how the random matrix results in compressive sensing can be directly applied to the MMSE expression to provide error bounds. Here we have considered both the case that the sampling locations are random and the eigenvalue distribution is fixed, and the case that the sampling locations are fixed and the locations of the nonzero eigenvalues are random. For a more general eigenvalue distribution, we have used a more complicated argument to obtain error bounds for measurement through random projections. Here our main contribution is to illustrate the interplay between the total uncertainty of the signal (different eigenvalue distributions) and the coordinate space transform that relates the canonical signal domain and the measurement domain to form error bounds. Finally, we have considered equidistant sampling of circularly wide-sense stationary (c.w.s.s.) signals, for which the coordinate transformation between the canonical signal domain and the measurement domain is given by the DFT. Here we have provided an explicit error expression that shows how the sampling rate and the eigenvalue distribution contribute to the error.

We have then continued our investigation of dependence in random fields with stationary Gaussian sources defined on $\mathbb{Z}$. We have formulated a family of problems related to the finite-length representations and sampling of these signals. Our framework here is again based on our vision of understanding the effect of different forms of finiteness in representation of signals, and measures of statistical dependence in random fields, in particular geometry of spread of uncertainty. We have first considered the decay rates for the error between finite dimensional representations and infinite dimensional representations. Our approach is based on the notion of mixing which is concerned with dependence in asymptotical sense. There is a vast literature on the notion of mixing in the fields of information theory and applied mathematics, but this notion does not seem to have been utilized in signal processing community. Providing several alternative ways to quantify dependence in random processes, this family of notions may provide new perspectives in signal processing problems where one needs to quantify the dependence in a signal family. Our work constitutes an example for these potential directions of research. We believe that it will be useful to researchers who would like to understand in what kind of problems this notion can be utilized. We have showed that for exponentially mixing sequences, for various representations and estimators, the error difference between using a finite-length representation and an infinite-length representation is upper bounded by an exponentially decreasing function of the finite window length. For stationary Gaussian signals, it is known that the presence of strong mixing may prevent a signal from being precisely bandlimited, but otherwise puts comparably loose restrictions on the spectral density, hence the effective bandwidth and the entropy. Nevertheless, the above results shows that mixing rate is pertinent to the geometric spread of uncertainty in the signal in the sense that it determines how the error difference between the finite and infinite-length representations decays. We have then considered the MMSE estimation of a stationary Gaussian source from its noisy samples. We have first showed that for stationary sources, for the purpose of calculating the MMSE based on equidistant samples, asymptotically circulant matrices can be used instead of original covariance matrices, which are Toeplitz. This result suggests that circularly wide-sense stationary signals in finite dimensions are more than an analogy for stationary signals in infinite dimensions: there is an operational relationship between these two signal models. Then, we have considered the MMSE associated with estimation of a stationary Gaussian source on $\mathbb{Z}_+$ from its equidistant samples on $\mathbb{Z}_+$. Using the previous result and our result on c.w.s.s. signals in our earlier work, we gave the explicit expression for the MMSE in terms of power spectral density of the source. An important aspect of our framework is the fact that we consider the sampling of the source on the half infinite line $\mathbb{Z}_+$ instead of the infinite line $\mathbb{Z}$. This framework makes direct usage of stationary arguments difficult, and makes the arguments more challenging. We note that contrary to much previous work which considers the Shannon-Nyquist interpolation formula as the means for the reconstruction of the signal, our performance criterion here is the MMSE, which, by definition, gives the minimum mean-square error achievable with the given samples. In this sense, our error expression provides performance limits for estimation of such a source from its samples. It is also important that our expression is explicit; in the sense that it does not just state the conditions under which the MMSE will be zero, but also shows exactly how the sampling rate, and signal and noise spectrums contribute to the error if these conditions are not met.

In this thesis, we were concerned with signal recovery and representation under various measurement constraints. We have investigated the effect of different forms of finiteness, such as finite number of samples or finite amplitude accuracy, on the signal recovery performance. An important concept in our investigations was the concept of spread of uncertainty in the measurement space, as opposed to the total uncertainty in the signal. In our belief, our work provides valuable insight for understanding different aspects of information transfer in the measurement process. We believe that our results are not only useful for better understanding of fundamental limits in signal recovery problems, but can also lead to guidelines in practical scenarios. Our general framework will be useful in a wide range of situations where inverse problems with similar constraints are encountered.

APPENDIX A

A.1 Proof of Lemma 8.1.1

We remind that in this section $u_{tk} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}tk}$, $0 \leq t\,, k \leq N{-}1$ and the associated eigenvalues are denoted with λ_k without reindexing them in decreasing/increasing order. We first assume that $K_y = E[yy^\dagger] = HK_xH^\dagger$ is non-singular. The generalization to the case where K_y may be nonsingular is presented at the end of the proof.

The MMSE error for estimating x from y is given by [188, Ch.2]

$$\begin{aligned} E[\|x - E[x|y]\|^2] &= \mathrm{tr}(K_x - K_{xy}K_y^{-1}K_{xy}^\dagger) && \text{(A.1)} \\ &= \mathrm{tr}(U\Lambda_xU^\dagger - U\Lambda_xU^\dagger H^\dagger(HU\Lambda_xU^\dagger H^\dagger)^{-1}HU\Lambda_xU^\dagger) && \text{(A.2)} \\ &= \mathrm{tr}(\Lambda_x - \Lambda_xU^\dagger H^\dagger(HU\Lambda_xU^\dagger H^\dagger)^{-1}HU\Lambda_x). && \text{(A.3)} \end{aligned}$$

We now consider $HU \in \mathbb{C}^{M\times N}$, and try to understand its structure

$$(HU)_{lk} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}(\Delta N l)k} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{M}lk}, \tag{A.4}$$

where $0 \leq l \leq \frac{N}{\Delta N} - 1, \;\; 0 \leq k \leq N-1$. We now observe that for a given l, $e^{j\frac{2\pi}{M}lk}$ is a periodic function of k with period $M = \frac{N}{\Delta N}$. So l^{th} row of HU can be expressed as

$$\begin{aligned} (HU)_{l:} &= \frac{1}{\sqrt{N}}\,[e^{j\frac{2\pi}{M}l[0\ldots N-1]}] && \text{(A.5)} \\ &= \frac{1}{\sqrt{N}}\,[e^{j\frac{2\pi}{M}l[0\ldots M-1]}|\ldots|e^{j\frac{2\pi}{M}l[0\ldots M-1]}]. && \text{(A.6)} \end{aligned}$$

Let U_M denote the $M \times M$ DFT matrix, i.e. $(U_M)_{lk} = \frac{1}{\sqrt{M}} e^{j\frac{2\pi}{M}lk}$ with $0 \leq l \leq M-1,\;\; 0 \leq k \leq M-1$. Hence HU is the matrix formed by stacking ΔN $M \times M$ DFT matrices side by side

$$HU = \frac{1}{\sqrt{\Delta N}}[U_M| \dots |U_M]. \tag{A.7}$$

Now we consider the covariance matrix of the observations $K_y = HK_xH^\dagger = HU\Lambda_xU^\dagger H^\dagger$. We first express Λ_x as a block diagonal matrix as follows

$$\Lambda_x = \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_{N-1} \end{bmatrix} = \begin{bmatrix} \Lambda^0 & 0 & \cdots & 0 \\ 0 & \Lambda^1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \Lambda^{\Delta N-1} \end{bmatrix}. \tag{A.8}$$

Hence $\Lambda_x = \mathrm{diag}(\Lambda_x^i)$ with $\Lambda_x^i = \mathrm{diag}(\lambda_{iM+k}) \in \mathbb{R}^{M\times M}$, where $0 \leq i \leq \Delta N - 1$, $0 \leq k \leq M-1$. We can write K_y as

$$\begin{aligned} K_y &= HU\Lambda_xU^\dagger H^\dagger & \text{(A.9)} \\ &= \frac{1}{\sqrt{\Delta N}}[U_M|\dots|U_M]\,\mathrm{diag}(\Lambda_x^i)\begin{bmatrix} U_M^\dagger \\ \vdots \\ U_M^\dagger \end{bmatrix}\frac{1}{\sqrt{\Delta N}} & \text{(A.10)} \\ &= \frac{1}{\Delta N}\;U_M(\sum_{i=0}^{\Delta N-1}\Lambda_x^i)U_M^\dagger & \text{(A.11)} \end{aligned}$$

We note that $\sum_{i=0}^{\Delta N-1}\Lambda_x^i \in \mathbb{R}^{M\times M}$ is formed by summing diagonal matrices, hence also diagonal. Since U_M is the $M \times M$ DFT matrix, K_y is again a circulant matrix whose k^{th} eigenvalue is given by

$$\frac{1}{\Delta N}\sum_{i=0}^{\Delta N-1}\lambda_{iM+k}. \tag{A.12}$$

Hence $K_y = U_M\Lambda_yU_M^\dagger$ is the eigenvalue-eigenvector decomposition of K_y, where $\Lambda_Y = \frac{1}{\Delta N}\sum_{i=0}^{\Delta N-1}\Lambda_x^i = \mathrm{diag}(\lambda_{y,k})$ with $\lambda_{y,k} = \frac{1}{\Delta N}\sum_{i=0}^{\Delta N-1}\lambda_{iM+k},\;\; 0 \leq k \leq M-1$. We note that there may be aliasing in the eigenvalue spectrum of K_y depending on the eigenvalue spectrum of K_x and ΔN. We also note that K_y may be aliasing free even if it is not bandlimited (low-pass, high-pass, etc.) in the conventional sense. Now K_y^{-1} can be expressed as

$$\begin{aligned} K_y^{-1} &= (U_M \Lambda_y U_M^\dagger)^{-1} && \text{(A.13)} \\ &= U_M \operatorname{diag}(\frac{1}{\lambda_{y,k}}) U_M^\dagger && \text{(A.14)} \\ &= U_M \operatorname{diag}(\frac{\Delta N}{\sum_{i=0}^{\Delta N-1} \lambda_{iM+k}}) U_M^\dagger. && \text{(A.15)} \end{aligned}$$

We note that since K_y is assumed to be non-singular, $\lambda_{y,k} > 0$. We are now ready to consider the error expression in (A.3). We first consider the second term $\operatorname{tr}(\Lambda_x U^\dagger H^\dagger K_y^{-1} H U \Lambda_x)$

$$\operatorname{tr}(\frac{1}{\sqrt{\Delta N}} \begin{bmatrix} \Lambda_x^0 U_M^\dagger \\ \vdots \\ \Lambda_x^{\Delta N-1} U_M^\dagger \end{bmatrix} (U_M \Lambda_y^{-1} U_M^\dagger) \frac{1}{\sqrt{\Delta N}} [U_M \Lambda_x^0 | \ldots | U_M \Lambda_x^{\Delta N-1}])$$

$$\begin{aligned} &= \sum_{i=0}^{\Delta N-1} \frac{1}{\Delta N} \operatorname{tr}(\Lambda_x^i \Lambda_y^{-1} \Lambda_x^i) && \text{(A.16)} \\ &= \sum_{i=0}^{\Delta N-1} \sum_{k=0}^{M-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}} && \text{(A.17)} \end{aligned}$$

Hence the MMSE becomes

$$\begin{aligned} E[||x - E[x|y]||^2] &= \sum_{t=0}^{N-1} \lambda_t - \sum_{i=0}^{\Delta N-1} \sum_{k=0}^{M-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}} && \text{(A.18)} \\ &= \sum_{k=0}^{M-1} \sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \sum_{k=0}^{M-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}} && \text{(A.19)} \\ &= \sum_{k=0}^{M-1} (\sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}}) && \text{(A.20)} \end{aligned}$$

We note that we have now expressed the MMSE as the sum of the errors in M frequency bands. Let us define the error at k^{th} frequency band as

$$e_k^w = \sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}}, \qquad 0 \leq k \leq M-1 \qquad \text{(A.21)}$$

Example A.1.1. *Before moving on, we study a special case: Let $\Delta N = 2$. Then*

$$\begin{aligned} e_k^w &= \lambda_k + \lambda_{\frac{N}{2}+k} - \frac{\lambda_k^2 + \lambda_{\frac{N}{2}+k}^2}{\lambda_k + \lambda_{\frac{N}{2}+k}} && \text{(A.22)} \\ &= \frac{2\lambda_k \lambda_{\frac{N}{2}+k}}{\lambda_k + \lambda_{\frac{N}{2}+k}}. && \text{(A.23)} \end{aligned}$$

Hence $\frac{1}{e_k^w} = \frac{1}{2}(\frac{1}{\lambda_{\frac{N}{2}+k}} + \frac{1}{\lambda_k})$. *We note that this is the MMSE error for the following single output multiple input system*

$$z^k = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} s_0^k \\ s_1^k \end{bmatrix}, \tag{A.24}$$

where $s^k \sim \mathcal{N}(0, K_{s^k})$, *with* $K_{s^k} = \operatorname{diag}(\lambda_k, \lambda_{\frac{N}{2}+k})$. *Hence the random variables associated with the frequency components at* k, *and* $\frac{N}{2}+k$ *act as interference for estimating the other one. We observe that for estimating* x *we have* $\frac{N}{2}$ *such channels in parallel.*

We may bound e_k^w *as*

$$\begin{aligned} e_k^w = \frac{2\lambda_k \lambda_{\frac{N}{2}+k}}{\lambda_k + \lambda_{\frac{N}{2}+k}} &\leq \frac{2\lambda_k \lambda_{\frac{N}{2}+k}}{\max(\lambda_k, \lambda_{\frac{N}{2}+k})} && \text{(A.25)} \\ &= 2\min(\lambda_k, \lambda_{\frac{N}{2}+k}) && \text{(A.26)} \end{aligned}$$

This bound may be interpreted as follows: Through the scalar channel shown in (A.24), *we would like to learn two random variables* s_0^k *and* s_1^k. *The error of this channel is upper bounded by the error of the scheme where we only estimate the one with the largest variance, and don't try to estimate the variable with the small variance. In that scheme, one first makes an error of* $\min(\lambda_k, \lambda_{\frac{N}{2}+k})$, *since the variable with the small variance is ignored. We may lose another* $\min(\lambda_k, \lambda_{\frac{N}{2}+k})$, *since this variable acts as additive noise for estimating the variable with the large variance, and the MMSE error associated with such a channel may be upper bounded by the variance of the noise.*

Now we choose the set of indices J *with* $|J| = N/2$ *such that* $k \in J \Leftrightarrow \frac{N}{2}+k \notin J$ *and* J *has the most power over all such sets, i.e.* $k + \arg\max_{k_0 \in \{0, N/2\}} \lambda_{k_0+k} \in J$, *where* $0 \leq k \leq N/2-1$. *Let* $P_J = \sum_{k \in J} \lambda_k$.

Hence

$$E[||x - E[x|y]||^2] = \sum_{k=0}^{N/2-1} e_k^w \leq 2\sum_{k=0}^{N/2-1} \min(\lambda_k, \lambda_{\frac{N}{2}+k}) = 2(P - P_J). \tag{A.27}$$

We observe that the error is upper bounded by 2× *(the power in the "ignored band").*

We now return to the general case. Although it is possible to consider any set J that satisfies the assumptions stated in (8.14), for notational convenience we choose the set $J = \{0, \ldots, M-1\}$. Of course in general one would look for the set J that has most of the power in order to have a better bound on the error.

We now consider

$$e_k^w = \sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}}, \qquad 0 \le k \le M-1 \tag{A.28}$$

We note that this is the MMSE of estimating S^k from the output of the following single output multiple input system

$$z^k = \left[\begin{array}{ccc} 1 & \cdots & 1 \end{array}\right] \left[\begin{array}{c} s_1^k \\ \vdots \\ s_{\Delta N-1}^k \end{array}\right], \tag{A.29}$$

where $s^k \ \sim \mathcal{N}(0, K_{s^k})$, with K_{s^k} as follows

$$\begin{aligned} K_{s^k} &= \mathrm{diag}(\sigma^2_{s_i^k}) && \text{(A.30)}\\ &= \mathrm{diag}(\lambda_k, \ldots, \lambda_{iM+k}, \ldots, \lambda_{(\Delta N-1)M+k}) && \text{(A.31)} \end{aligned}$$

We define

$$P^k = \sum_{l=0}^{\Delta N-1} \lambda_{lM+k}, \qquad 0 \le k \le M-1 \tag{A.32}$$

We note that $\sum_{k=0}^{M-1} P^k = P$.

We now bound e_k^w as in the $\Delta N = 2$ example

$$\begin{aligned} e_k^w &= \sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}}, && \text{(A.33)}\\ &= \sum_{i=0}^{\Delta N-1} (\lambda_{iM+k} - \frac{\lambda_{iM+k}^2}{P^k}), && \text{(A.34)}\\ &= (\lambda_k - \frac{\lambda_k^2}{P^k}) + \sum_{i=1}^{\Delta N-1} (\lambda_{iM+k} - \frac{\lambda_{iM+k}^2}{P^k}), && \text{(A.35)}\\ &\le (P^k - \lambda_k) + \sum_{i=1}^{\Delta N-1} \lambda_{iM+k} && \text{(A.36)}\\ &= (P^k - \lambda_k) + P^k - \lambda_k && \text{(A.37)}\\ &= 2(P^k - \lambda_k) && \text{(A.38)} \end{aligned}$$

where we've used $\lambda_k - \frac{\lambda_k^2}{P^k} = \frac{\lambda_k(P^k-\lambda_k)}{P^k} \leq P^k - \lambda_k$ since $0 \leq \frac{\lambda_k}{P^k} \leq 1$ and $\lambda_{iM+k} - \frac{\lambda_{iM+k}^2}{P^k} \leq \lambda_{iM+k}$ since $\frac{\lambda_{iM+k}^2}{P^k} \geq 0$. This upper bound may interpreted similar to the Example A.1.1: The error is upper bounded by the error of the scheme where one estimates the random variable associated with λ_k, and ignore the others.

The total error is bounded by

$$\begin{aligned} E[||x - E[x|y]||^2] = \sum_{k=0}^{M-1} e_k^w &\leq \sum_{k=0}^{M-1} 2(P^k - \lambda_k) && \text{(A.39)} \\ &= 2(\sum_{k=0}^{M-1} P^k - \sum_{k=0}^{M-1} \lambda_k) && \text{(A.40)} \\ &= 2(P - P_J) && \text{(A.41)} \end{aligned}$$

Remark A.1.1. *We now consider the case where K_y may be singular. In this case, it is enough to use K_y^+ instead of K_y^{-1}, where $^+$ denotes the Moore-Penrose pseudo-inverse [188, Ch.2]. Hence the MMSE may be expressed as $\mathrm{tr}(K_x - K_{xy}K_y^+K_{xy}^\dagger)$. We have $K_y^+ = (U_M\Lambda_y U_M^\dagger)^+ = U_M\Lambda_y^+U_M^\dagger = U_M \,\mathrm{diag}(\lambda_{y,k}{}^+)U_M^\dagger$, where $\lambda_{y,k}^+ = 0$ if $\lambda_{y,k} = 0$ and $\lambda_{y,k}^+ = \frac{1}{\lambda_{y,k}}$ otherwise. Going through calculations with K_y^+ instead of K_y^{-1} reveals that the error expression remain essentially the same*

$$E[||x - E[x|y]||^2] \quad = \quad \sum_{k\in J_0}(\sum_{i=0}^{\Delta N-1} \lambda_{iM+k} - \sum_{i=0}^{\Delta N-1} \frac{\lambda_{iM+k}^2}{\sum_{l=0}^{\Delta N-1} \lambda_{lM+k}}), \qquad \text{(A.42)}$$

where $J_0 = \{k : \sum_{l=0}^{\Delta N-1} \lambda_{lM+k} \neq 0, 0 \leq k \leq M-1\} \subseteq \{0,\dots,M-1\}$. We note that $\Delta N \lambda_{y,k} = \sum_{l=0}^{\Delta N-1} \lambda_{lM+k} = P^k$.

A.2 Proof of Lemma 8.3.2

Our aim is to show that the smallest eigenvalue of $A = \Lambda_x^{-1} + \frac{1}{\sigma_n^2}H^\dagger H$ is bounded from below with a sufficiently large number with high probability. That is we are interested in

$$\inf_{x\in S^{N-1}} x^\dagger \Lambda_x^{-1} x + \frac{1}{\sigma_n^2} x^\dagger H^\dagger H x \qquad \text{(A.43)}$$

To lower bound the smallest eigenvalue, we adopt the approach proposed by [206]: We consider the decomposition of the unit sphere into two sets, compressible vectors and incompressible vectors. We remind the following definitions from [206].

Definition A.2.1. *[pg.14, [206]] Let $|supp(x)|$ denote the number of elements in the support of x. Let $\eta, \rho \in (0,1)$. $x \in \mathbb{R}^{\mathbb{N}}$ is sparse, if $|supp(x)| \leq \eta N$. The set of vectors sparse with a given η is denoted by $Sparse(\eta)$. $x \in S^{N-1}$ is compressible, if x is within an Euclidean distance ρ from the set of all sparse vectors, that is $\exists\, y \in Sparse(\eta), d(x,y) \leq \rho$. The set of compressible vectors is denoted by $Comp(\eta, \rho)$. $x \in S^{N-1}$ is incompressible if it is not compressible. The set of incompressible vectors is denoted by $Incomp(\eta, \rho)$.*

Lemma A.2.1. *[Lemma 3.4, [206]] Let $x \in Incomp(\eta, \rho)$. Then there exists a set of $\psi \subseteq 1, ..., N$ of cardinality $|\psi| \geq 0.5\rho^2\eta N$ such that*

$$\frac{\rho}{\sqrt{(2N)}} \leq |x_k| \leq \frac{1}{\sqrt{\eta N}} \qquad for\, all\, k \in \psi \tag{A.44}$$

We note that the set of compressible and incompressible vectors provide a decomposition of the unit sphere, i.e. $S^{N-1} = Incomp(\eta, \rho) \bigcup Comp(\eta, \rho)$ [206]. We will show that the first/second term in (A.43) is sufficiently away from zero for $x \in Incomp(\eta, \rho) /\ x \in Comp(\eta, \rho)$ respectively.

As noted in [206]

$$\begin{aligned} &P(\inf_{x \in S^{N-1}} x^\dagger A x \leq C_0 N) \\ &\leq P(\inf_{x \in Comp(\eta,\rho)} x^\dagger A x \leq C_0 N) + P(\inf_{x \in Incomp(\eta,\rho)} x^\dagger A x \leq C_0 N) \end{aligned} \tag{A.45}$$

We also note that

$$\inf_{x \in Incomp(\eta,\rho)} x^\dagger \Lambda_x^{-1} x + x^\dagger \frac{1}{\sigma_n^2} H^\dagger H x \geq \inf_{x \in Incomp(\eta,\rho)} x^\dagger \Lambda_x^{-1} x \tag{A.46}$$

$$= \inf_{x \in Incomp(\eta,\rho)} ||\Lambda_x^{-1/2} x||^2 \tag{A.47}$$

and

$$\inf_{x \in Comp(\eta,\rho)} x^\dagger \Lambda_x^{-1} x + x^\dagger \frac{1}{\sigma_n^2} H^\dagger H x \geq \inf_{x \in Comp(\eta,\rho)} x^\dagger \frac{1}{\sigma_n^2} H^\dagger H x \tag{A.48}$$

$$= \frac{1}{\sigma_n^2} (\inf_{x \in Comp(\eta,\rho)} ||Hx||^2) \tag{A.49}$$

where inequalites are due to the fact that Λ_x^{-1}, $H^\dagger H$ are both positive-semidefinite.

We first consider the following special case of [206, Lemma 3.3]:

Lemma A.2.2. *[206, Lemma 3.3] Let H be a $M = \beta N \times N$ random matrix with i.i.d Gaussian entries with variances at least 1. Then there exist $\eta, \rho, C_2, C_1 > 0$ that does not depend on N such that*

$$P(\inf_{x \in Comp(\eta,\rho)} ||Hx||^2 \leq C_2 N) \leq e^{-C_1 N} \tag{A.50}$$

To see the relationship between the number of measurements and the parameters of the lemma, we take a closer look at the proof of this lemma: We observe that here H is a $M = \beta N \times N$ matrix, hence [206, Proposition 2.5] requires $\eta N < \delta_0 M$ where $0 < \delta_0 < 0.5$ is a parameter of [206, Proposition 2.5]. Hence M should satisfy $M > T'$ where $T' = \frac{1}{\delta_0} \eta N$.

We now look at $\inf_{x \in Incomp(\eta,\rho)} ||\Lambda_x^{-1/2} x||^2$. We note that none of the entities in this expression is random. We note the following

$$\inf_{x \in Incomp(\eta,\rho)} ||\Lambda_x^{-1/2} x||^2 = \inf_{x \in Incomp(\eta,\rho)} \sum_{i=1}^{N} \frac{1}{\lambda_i} |x_i|^2 \tag{A.51}$$

$$\geq \sum_{i \in \psi} \frac{1}{\lambda_i} \frac{\rho^2}{2N}, \tag{A.52}$$

where the inequality is due to Lemma A.2.1. We observe that to have this expression sufficiently bounded away from zero, the distribution of $\frac{1}{\lambda_i}$ should be spread enough.

Different approaches to quantify the spread of the eigenvalue distribution can be adopted. One may directly quantify the spread of $\frac{1}{\lambda_i}$ distribution, for example by requiring $[\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_N}] / \sum_i \frac{1}{\lambda_i} \in Incomp(\bar{\eta}, \bar{\rho})$, where $\bar{\eta}$, $\bar{\rho}$ are new parameters.

Since it is more desirable to have explicit constraints on the λ_i distribution itself instead of constraints on the distribution of $\frac{1}{\lambda_i}$, we consider another approach.

Let us assume that $\lambda_i < C_\lambda \frac{P}{N}$, for $i \geq \kappa|\psi|$, where $\kappa \in (0,1)$, $0 < C_\lambda < \infty$. Then we have

$$\begin{aligned}
\inf_{x \in Incomp(\eta,\rho)} ||\Lambda_x^{-1/2} x||^2 &\geq \sum_{i \in \psi} \frac{1}{\lambda_i} \frac{\rho^2}{2N} && \text{(A.53)}\\
&> (|\psi| - \kappa|\psi|) \frac{1}{C_\lambda P} \frac{\rho^2}{2} && \text{(A.54)}\\
&\geq (1-\kappa) 0.5 \rho^2 \eta N \frac{1}{C_\lambda P} \frac{\rho^2}{2} && \text{(A.55)}\\
&= (1-\kappa) 0.25 \rho^4 \eta \frac{1}{C_\lambda P} N && \text{(A.56)}\\
&= \frac{1}{P} C_3 N && \text{(A.57)}
\end{aligned}$$

where we have used $|\psi| \geq 0.5\rho^2 \eta N$. Here $C_3 = (1-\kappa)0.25\rho^4\eta\frac{1}{C_\lambda}$.

We will now complete the argument to arrive at $P(\inf_{x \in S^{N-1}} x^\dagger A x \leq C\frac{N}{P}) \leq e^{-C_1 N}$ as claimed in the Lemma we are proving, and then discuss the effect of different eigenvalue distributions, noise level and M on this result. Let $C = P\min(\frac{1}{\sigma_n^2}C_2, \frac{1}{P}C_3) = \min(\frac{P}{\sigma_n^2}C_2, C_3)$. By (A.47) and (A.57), $P(\inf_{x \in Incomp(\eta,\rho)} x^\dagger A x \leq C\frac{N}{P}) = 0$. By (A.49), Lemma A.2.2, $P(\inf_{x \in Comp(\eta,\rho)} x^\dagger A x \leq C\frac{N}{P}) \leq e^{-C_1 N}$. The result follows by (A.45).

Up to now, we have not considered the admissibility of C to provide guarantees for low values of error. We note that as observed in Remark A.2.1, and Remark A.2.2, the error bound expression in Theorem 8.3.1 cannot provide bounds for low values of error when the eigenvalue distribution is spread. Hence while stating the result of Lemma 8.3.2, hence Theorem 8.3.1, we consider the other case, the case where the eigenvalue distribution is not spread out, as discussed in Remark A.2.3.

Remark A.2.1. *We note that as $C = P\min(\frac{1}{\sigma_n^2}C_2, \frac{1}{P}C_3) = \min(\frac{P}{\sigma_n^2}C_2, C_3)$ gets larger, the lower bound on the eigenvalues of $\Lambda_x^{-1} + \frac{1}{\sigma_n^2}H^\dagger H$ gets larger, and the bound on the MMSE (see for example* (8.39)*) gets smaller. To have guarantees for low values of error for a given M, we want to have have C as large as possible.*

For a given number of measurements M, we have a C_2 and associated η, ρ, C_1. For a given P and σ_n^2, to have guarantees for error levels as low as this C_2, P and σ_n^2 permit, we should have $\frac{P}{\sigma_n^2} C_2 \leq C_3$ so that the overall constant is as good as the one coming from Lemma A.2.2. We note that to have C_3 large, C_λ must be small.

Remark A.2.2. *Let us assume that all the eigenvalues are approximately equal, i.e. $|\lambda_i - \frac{P}{N}| \leq \bar{q}\frac{P}{N}$, $\bar{q} \in [0,1]$ where $\bar{q}$ is close to 0. We have*

$$\begin{aligned}\inf_{x \in Incomp(\eta,\rho)} ||\Lambda_x^{-1/2} x||^2 &\geq \sum_{i \in \psi} \frac{1}{1+\bar{q}} \frac{N}{P} \frac{\rho^2}{2N} &\text{(A.58)}\\ &\geq 0.5\rho^2 \eta N \frac{1}{1+\bar{q}} \frac{1}{P} \frac{\rho^2}{2} &\text{(A.59)}\\ &= \frac{1}{1+\bar{q}} 0.25 \rho^4 \eta N \frac{1}{P}, &\text{(A.60)}\end{aligned}$$

Hence $C_3 = \frac{1}{1+\bar{q}} 0.25\rho^4\eta > 0$. In this case (8.39) *will not provide guarantees for low values of error. In fact, with $3M \leq N$ the error may be lower bounded as follows*

$$\begin{aligned}E[||x - E[x|y]||^2] &= \operatorname{tr}((\Lambda_x^{-1} + \frac{1}{\sigma_n^2} H^\dagger H)^{-1}) &\text{(A.61)}\\ &= \sum_{i=1}^{N} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} H^\dagger H)} &\text{(A.62)}\\ &= \sum_{i=M+1}^{N} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} H^\dagger H)} + \sum_{i=1}^{M} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} H^\dagger H)} &\text{(A.63)}\\ &\geq \sum_{i=M+1}^{N} \frac{1}{\lambda_{i-M}(\Lambda_x)} + \sum_{i=1}^{M} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} H^\dagger H)}, &\text{(A.64)}\\ &= \sum_{i=M+1}^{N} \lambda_{N-i+M+1}(\Lambda_x) + \sum_{i=1}^{M} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} H^\dagger H)}, &\text{(A.65)}\\ &= \sum_{i=M+1}^{N} \lambda_i(\Lambda_x) + \sum_{i=1}^{M} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} H^\dagger H)}, &\text{(A.66)}\\ &\geq (1-\bar{q}) \frac{N-M}{N} P + \sum_{i=1}^{M} \frac{1}{\lambda_i(\Lambda_x^{-1} + \frac{1}{\sigma^2} H^\dagger H)} &\text{(A.67)}\end{aligned}$$

where in (A.64), *we have used case (b) of Lemma 8.3.1 and the fact that $H^\dagger H$ is at most rank M. We note that as $\bar{q}$ gets closer to* 0, *the first term gets closer to* $\frac{N-M}{N}P$.

Remark A.2.3. *Let $D(\delta)$ be the smallest number satisfying $\sum_{i=1}^{D} \lambda_i \geq \delta P$, where $\delta \in (0,1]$. Let $D(\delta) = \alpha N$, $\alpha \in (0,1]$. Let $D(\delta)$ be sufficiently small for δ sufficiently large, more precisely $D(\delta) = \alpha N < \kappa|\psi|$, $\kappa \in (0,1)$, $\lambda_i < C_\lambda \frac{P}{N}$, for $i \geq \kappa|\psi|$ with $C_\lambda = q\frac{(1-\delta)}{(1-\alpha)}$, with $1 > q > 0$. Hence we have $\lambda_i < q\frac{(1-\delta)P}{(1-\alpha)N}$, $i \geq \kappa\alpha N$. We observe that other parametes fixed, as admissible $\alpha > 0$ gets closer to 0, or $\delta > 0$ gets close to 1, C_λ gets smaller as desired. We note that the inequality $D(\delta) < 0.5\kappa\rho^2\eta N = T$ together with the inequality $M > T' = \frac{1}{\delta_0}\eta N$ relates the spread of the eigenvalues to the admissible number of measurements.*

Remark A.2.4. *We now discuss the effect of noise level. We note that the total signal power is given by $\operatorname{tr}(K_x) = P$, whereas each measurement is done with noise whose variance is σ_n^2. We want to have $C = P\min(\frac{1}{\sigma_n^2}C_2, \frac{1}{P}C_3) = \min(\frac{P}{\sigma_n^2}C_2, C_3)$ as large as possible. Let us assume that other parameters of the problem are fixed and focus on the ratio $\frac{P}{\sigma_n^2}$. For constant P, as noise level increases, $\frac{P}{\sigma_n^2}$ decreases. After some noise level, the minimum will be given by $\frac{P}{\sigma_n^2}C_2$. Hence the lower bound on the eigenvalues of $\Lambda_x^{-1} + \frac{1}{\sigma_n^2}H^\dagger H$ will get smaller, and the upper bound on the MMSE will get larger. Hence Theorem 8.3.1 will not provide guarantees for low values of error for high levels of noise.*

Example: We now study a special case to illustrate the nature of error bounds this result can provide. We assume that we have the following eigenvalue distribution structure: $\lambda_i = \delta\frac{P}{D}$, if $i \in \mathcal{D}$, and $\lambda_i = (1-\delta)\frac{P}{N-D}$, if $i \notin \mathcal{D}$, where $\delta \approx 1$, for a set of indices $\mathcal{D} \subset \{1,\dots,N\}$ with $D = |\mathcal{D}|$. Let us assume that $\sigma_H^2 = 1/N$. We note that this scaling of the variance of the components of H can be obtained by a simple scaling of the measurement matrix H. Let $\eta N = vD$, for $v > 1$. If $M \geq C\bar{\gamma}^{-2}(v\,D\ln(N/(v\,D))) + \ln(\epsilon^{-1}))$ (for a universal constant $C > 0$), then with probability at least $1-\epsilon$, we have the following (see for instance [222, Thm. 2.12])

$$\inf_{x \in Sparse(\eta)} ||Hx||^2 \geq (1-\bar{\gamma})\frac{M}{N} \tag{A.68}$$

As in the proof of Lemma A.2.2, this result can be extended to compressible vectors. In particular, we have the following bound

$$\inf_{x \in Comp(\eta,\rho)} ||Hx||^2 \geq (1-\gamma)\frac{M}{N} \tag{A.69}$$

with probability at least $1 - \epsilon - \epsilon'$. Here γ depends on $\bar{\gamma}$, ρ and $C_s = 1 + \sqrt{M/N} + t/\sqrt{N}$, $t \geq 0$ and $\epsilon' = 2\exp(-0.5t^2)$. Whether $\gamma > 0$ is small enough is determined by the choice of ρ, and the values of these parameters. Smaller choices of ρ result in better γ values which come at the expense of larger set of incompressible vectors to deal with. Here C_s and ϵ' comes from the upper bound on the singular values of a $M \times N$ random matrix with Gaussian i.i.d entries given in Corollary 5.35 of [223], which can be stated as follows

$$P(\sup_{x \in \mathcal{S}^{N-1}} ||Hx|| \geq C_s) \leq \epsilon'. \tag{A.70}$$

Suppose that $\lceil 0.5\rho^2 vD \rceil > D$. Now (A.53) can be expressed as follows

$$\inf_{x \in Incomp(\eta,\rho)} ||\Lambda_x^{-1/2} x||^2 \geq (0.5\rho^2 v - 1)D \frac{1}{(1-\delta)\frac{P}{N-D}} \frac{0.5\rho^2}{N} \tag{A.71}$$

Following the same steps in the general proof, we combine (A.69) and (A.71) to obtain the following bound on the error

$$E[||x - E[x|y]||^2] \leq (1-\delta)P + \max(C_e(1-\delta)P, \frac{1}{\frac{1}{\delta} + (1-\gamma)\,\mathrm{SNR}} P) \tag{A.72}$$

which holds with probability at least $1 - \epsilon - \epsilon'$. Here we have used the notation $C_e^{-1} = (0.5\rho^2 v - 1)0.5\rho^2 \frac{N-D}{N}$, and $\mathrm{SNR} = \frac{1}{\sigma_n^2}\frac{P}{D}\frac{M}{N}$. C_e will take small values for large values of v, that is when one uses significantly less sparse signals (signals with support size ηN) in the proof than the number of significant eigenvalues associated with the signal (D) resulting in a higher number of measurements requirement or guarantees that hold with lower probabilities.

Let us take a closer look at this error bound. The first $(1-\delta)P$ term is the total power in the insignificant eigenvalues (i.e. λ_i such that $i \notin \mathcal{D}$). This term is an upper bound for the error that would have been introduced if we had preffered not estimating the random variables corresponding to these insignificant eigenvalues. Since in our setting we are interested in signals with low degree of freedom, hence δ close to 1, this term is guaranteed to be small. Let us now look at the term that will come out of the maximum function. When the noise level is relatively low, the $C_e(1-\delta)P$ term comes out of the max term. This term may be interpreted as a scaled version of the upper bound on the error due to the insignificant eigenvalues acting as noise for estimating of the random variables corresponding to the significant eigenvalues (i.e. λ_i such that $i \in \mathcal{D}$). Hence in the case where the noise level is relatively low, the insignificant eigenvalues become the dominant source of error in estimation. When the noise level is relatively high, the second argument comes out of the max term. Hence for high levels of noise, system noise rather than the insignificant eigenvalues becomes the dominant source of error in the estimation. We note that this term has the same form with the error expressions in Section 8.2, where the case that the insignificant eigenvalues are exactly zero were considered. We observe that there is again a loss of effective signal-to-noise ratio through a multiplicative factor appearing in front of SNR, compared to the error expression associated with the deterministic equidistant scenario of Corollary 8.31.

A.3 Proof of Lemma 8.4.1

The left hand side of the unitary matrix constraint in (8.45) may be rewritten as

$$\begin{aligned} & e_i^{\mathrm{T}}(U_B^{\dagger}U_B - I_{|B|})e_k \\ &= e_i^{\mathrm{T}}((U_{B,R} + jU_{B,I})^{\dagger}(U_{B,R} + jU_{B,I}) - I_{|B|})e_k && \text{(A.73)} \\ &= e_i^{\mathrm{T}}((U_{B,R}^{\mathrm{T}} - jU_{B,I}^{\mathrm{T}})(U_{B,R} + jU_{B,I}) - I_{|B|})e_k && \text{(A.74)} \\ &= e_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,R} + U_{B,I}^{\mathrm{T}}U_{B,I})e_k + je_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,I} - U_{B,I}^{\mathrm{T}}U_{B,R})e_k - e_i^{\mathrm{T}}I_{|B|}e_k. && \text{(A.75)} \end{aligned}$$

Hence the constraint becomes

$$e_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,R} + U_{B,I}^{\mathrm{T}}U_{B,I})e_k + je_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,I} - U_{B,I}^{\mathrm{T}}U_{B,R})e_k = e_i^{\mathrm{T}}I_{|B|}e_k. \quad \text{(A.76)}$$

By considering the real and imaginary parts of the equality separately, these constraints may be expressed as

$$\begin{aligned} e_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,R} + U_{B,I}^{\mathrm{T}}U_{B,I})e_k &= e_i^{\mathrm{T}}I_{|B|}e_k, \quad (i,k) \in \gamma && \text{(A.77)} \\ e_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}}U_{B,I} - U_{B,I}^{\mathrm{T}}U_{B,R})e_k &= 0, \quad (i,k) \in \bar{\gamma} && \text{(A.78)} \end{aligned}$$

where $\gamma = \{(i,k) | i = 1, \ldots, |B|,\ k = 1, \ldots, i\}$, and $\bar{\gamma} = \{(i,k) | i = 1, \ldots, |B|,\ k = 1, \ldots, i-1\}$. For the $i = k$ case, we only consider the real part of the constraint since the imaginary part necessarily vanishes, i.e. $e_i^{\mathrm{T}}(U_B^{\dagger}U_B)e_i = u_i^{\dagger}u_i \in \mathbb{R}$.

The set of constraint gradients with respect to $\begin{bmatrix} U_{B,R} \\ U_{B,I} \end{bmatrix}$ can be expressed as

$$\left\{ \begin{bmatrix} U_{B,R}(e_i e_k^{\mathrm{T}} + e_k e_i^{\mathrm{T}}) \\ U_{B,I}(e_i e_k^{\mathrm{T}} + e_k e_i^{\mathrm{T}}) \end{bmatrix} |(i,k) \in \gamma \right\} \bigcup \left\{ \begin{bmatrix} U_{B,I}(-e_i e_k^{\mathrm{T}} + e_k e_i^{\mathrm{T}}) \\ U_{B,R}(e_i e_k^{\mathrm{T}} - e_k e_i^{\mathrm{T}}) \end{bmatrix} |(i,k) \in \bar{\gamma} \right\} \tag{A.79}$$

where we have used the following identities [224]

$$\begin{aligned} d(\mathrm{tr}(A_1 X^{\mathrm{T}} A_2)) &= d(\mathrm{tr}(A_2^{\mathrm{T}} X A_1^{\mathrm{T}})) && \text{(A.80)} \\ &= \mathrm{tr}(A_2^{\mathrm{T}} dX A_1^{\mathrm{T}}) && \text{(A.81)} \\ &= \mathrm{tr}(A_1^{\mathrm{T}} A_2^{\mathrm{T}} dX) && \text{(A.82)} \end{aligned}$$

and

$$\begin{aligned} d(\mathrm{tr}(X^{\mathrm{T}} A_2 X A_1)) &= d(\mathrm{tr}(X A_1 X^{\mathrm{T}} A_2)) && \text{(A.83)} \\ &= \mathrm{tr}(dX A_1 X^{\mathrm{T}} A_2 + X A_1 d(X^{\mathrm{T}}) A_2) && \text{(A.84)} \\ &= \mathrm{tr}(A_1 X^{\mathrm{T}} A_2 dX + d(X^{\mathrm{T}}) A_2 X A_1) && \text{(A.85)} \\ &= \mathrm{tr}(A_1 X^{\mathrm{T}} A_2 dX + A_1^{\mathrm{T}} X^{\mathrm{T}} A_2^{\mathrm{T}} dX) && \text{(A.86)} \end{aligned}$$

where X is the matrix variable defined on real numbers and A_1 and A_2 are constant real matrices. For instance, with $U_{B,R}$ as the variable $d(\mathrm{tr}(e_i^{\mathrm{T}}(U_{B,R}^{\mathrm{T}} U_{B,R}) e_k)) = d(\mathrm{tr}(U_{B,R}^{\mathrm{T}} U_{B,R} e_k e_i^{\mathrm{T}}) = \mathrm{tr}((e_i e_k^{\mathrm{T}} + e_k e_i^{\mathrm{T}}) U_{B,R}^{\mathrm{T}} dU_{B,R})$ with $A_1 = e_k e_i^{\mathrm{T}}$, and $A_2 = I_N$.

The linear independence of the elements of this set follows from the following fact: For any matrix $U_B \in \mathbb{C}^{N \times B}$ satisfying $U_B^{\dagger} U_B = I_{|B|}$, the matrix $\hat{U}_B = \begin{bmatrix} U_{B,R} & -U_{B,I} \\ U_{B,I} & U_{B,R} \end{bmatrix} \in \mathbb{R}^{2N \times 2B}$ satisfies $\hat{U}_B^{\mathrm{T}} \hat{U}_B = I_{2|B|}$ [180]. Hence the columns of $\hat{U}_B$ form an orthonormal set of vectors. We observe that the elements of the constraint gradient set given in (A.79) are matrices with zero entries except at $k^{th}$and i^{th} columns, where at these two (or one if $i = k$) column(s), we have columns from $\hat{U}_B$. For instance consider $\begin{bmatrix} U_{B,R}(e_i e_k^{\mathrm{T}} + e_k e_i^{\mathrm{T}}) \\ U_{B,I}(e_i e_k^{\mathrm{T}} + e_k e_i^{\mathrm{T}}) \end{bmatrix}$ for some $(i,k) \in \gamma$, and let $i \neq k$. This is a matrix of zeros except at k^{th} column we have i^{th} column of $\hat{U}_B$ and at i^{th} column we have k^{th} column of $\hat{U}_B$. Now since $\hat{U}_B$ has orthonormal columns, it is not possible to form the values at k^{th} and i^{th} columns using other columns of U_B, and hence other elements of the set given in (A.79). Similar arguments hold for all the other elements of the set in (A.79). Hence the constraint gradients are linearly independent for any matrix $U_B \in \mathbb{C}^{N\times B}$ satisfying $U_B^\dagger U_B = I_{|B|}$.

A.4 A note on the Lagrangian in Section 8.4

We now clarify the form of the Lagrangian in (8.51). We note that here we are concerned with Lagrangian for optimizing a real valued function of a matrix variable with complex entries under equality constraints. Let $\widetilde{f}_0(\widetilde{U}_B)$ be the function to be optimized with complex equality constraints $\widetilde{f}_{i,k}(\widetilde{U}_B) = 0 \in \mathbb{C}$, $(i,k) \in \bar{\gamma}$, with $|\bar{\gamma}| = N_1 = 0.5|B|(|B|-1)$ and the real equality constraints $\widetilde{h}_k(\widetilde{U}_B) = 0 \in \mathbb{R}$, $k = 1, \ldots, N_2 = |B|$. The N_1 complex equality constraints can be expressed equivalently as $2N_1$ real equality constraints $\Re\{\widetilde{f}_{i,k}(\widetilde{U}_B)\} = 0 \in \mathbb{R}$, and $\Im\{\widetilde{f}_{i,k}(\widetilde{U}_B)\} = 0 \in \mathbb{R}$ for $(i,k) \in \bar{\gamma}$. Then the Lagrangian can be expressed as

$$\begin{aligned}
&\widetilde{L}(\widetilde{U}_B, \nu, \upsilon) \\
&= \widetilde{f}_0(\widetilde{U}_B) + \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k,R} \Re\{\widetilde{f}_{i,k}(\widetilde{U}_B)\} + \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k,I} \Im\{\widetilde{f}_{i,k}(\widetilde{U}_B)\} + \sum_{k=1}^{N_2} \upsilon_k \widetilde{h}_k(\widetilde{U}_B) \qquad \text{(A.87)} \\
&= \widetilde{f}_0(\widetilde{U}_B) + \sum_{(i,k)\in\bar{\gamma}} \Re\{\nu_{i,k}\{\widetilde{f}_{i,k}(\widetilde{U}_B)\}\} + \sum_{k=1}^{N_2} \upsilon_k \widetilde{h}_k(\widetilde{U}_B) \qquad \text{(A.88)} \\
&= \widetilde{f}_0(\widetilde{U}_B) + 0.5 \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k} \widetilde{f}_{i,k}(\widetilde{U}_B) + 0.5 \sum_{(i,k)\in\bar{\gamma}} \nu_{i,k}^* \widetilde{f}_{i,k}^*(\widetilde{U}_B) + \sum_{k=1}^{N_2} \upsilon_k \widetilde{h}_k(\widetilde{U}_B) \qquad \text{(A.89)}
\end{aligned}$$

where $\nu_{i,k} \in \mathbb{C}$, with $\Re\{\nu_{i,k}\} = \nu_{i,k,R}$, $\Im\{\nu_{i,k}\} = \nu_{i,k,I}$, and $\upsilon_k \in \mathbb{R}$ are Lagrange multipliers. Now (8.51) is obtained with $\widetilde{f}_0(\widetilde{U}_B) = \sum_k p_k \operatorname{tr}((\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-1})$, $\widetilde{f}_{i,k}(\widetilde{U}_B) = e_i^{\mathrm{T}}(U_B^\dagger U_B - I_{|B|})e_k$, $\widetilde{h}_k(\widetilde{U}_B) = e_k^{\mathrm{T}}(U_B^\dagger U_B - I_{|B|})e_k$ and absorbing any constants into Lagrange multipliers.

APPENDIX B

B.1 Proof of Lemma 9.1.1

We first give the definition of maximal correlation coefficient.

Definition B.1.1. *For a Gaussian stationary zero-mean source $\{X_t\}$ the maximal correlation coefficient is defined as the following:*

$$\rho(\tau) = \sup_{\eta,\, \xi} \frac{|E[\eta\xi]|}{(E[|\eta|^2]E[|\xi|^2])^{1/2}}. \tag{B.1}$$

Here the random variables η and ξ are finite variance random variables measurable with respect to $\mathcal{F}_{-\infty}^{k}$ and $\mathcal{F}_{k+\tau}^{\infty}$, $k \in \mathbb{Z}$ respectively.

The following result relates the α-mixing coefficient and the maximal correlation coefficient.

Lemma B.1.1. *[225] For Gaussian processes, the following holds:*

$$\alpha(\tau) \leq \rho(\tau) \leq 2\pi\alpha(\tau). \tag{B.2}$$

We note that the correlation function $r(\tau)$ and the maximal correlation coefficient has the following relation $\frac{r(\tau)}{r(0)} \leq \rho(\tau)$. Hence by Lemma B.1.1, we have $r(\tau) \leq 2\,\pi\, r(0)\alpha(\tau)$. We conclude that when the process is exponentially mixing, decay of $|r(\tau)|$ is also upper-bounded exponentially. This proves the claim of Lemma 9.1.1 as desired.

B.2 Proof of Lemma 9.1.3

We note the following

$$\begin{aligned} E[|| \sum_{k=N+1}^{\infty} a_k X_{t-k}||^2] &= E[\lim_{K\to\infty} (\sum_{k=N+1}^{K} a_k X_{t-k})(\sum_{l=N+1}^{K} a_l X_{t-l})] && \text{(B.3)} \\ &= E[\lim_{K\to\infty} \sum_{k=N+1}^{K} \sum_{l=N+1}^{K} a_k a_l X_{t-l} X_{t-k}] && \text{(B.4)} \\ &= \lim_{K\to\infty} \sum_{k=N+1}^{K} \sum_{l=N+1}^{K} a_k a_l E[X_{t-l} X_{t-k}] && \text{(B.5)} \\ &= \sum_{k=N+1}^{\infty} \sum_{l=N+1}^{\infty} a_k a_l r_{k-l} && \text{(B.6)} \end{aligned}$$

We now provide the detailed steps for the justification of the step from (B.4) to (B.5). The relevant assumptions are the following: X_t's are Gaussian with $E[X_t] = 0$, $E[X_t^2] = \sigma_x^2 < \infty$, $\{a_k\} \in l_1$.

Let us introduce the following notation

$$f_K = \sum_{k=N+1}^{K} \sum_{l=N+1}^{K} a_k a_l X_{t-l} X_{t-k}, \qquad \text{(B.7)}$$

$$h_K = \sum_{k=0}^{K} \sum_{l=0}^{K} |a_k||a_l||X_{t-l} X_{t-k}|, \qquad \text{(B.8)}$$

$$g = \lim_{K\to\infty} h_K = \lim_{K\to\infty} \sum_{k=0}^{K} \sum_{l=0}^{K} |a_k||a_l||X_{t-l} X_{t-k}|. \qquad \text{(B.9)}$$

We want to prove that $E[\lim_{K\to\infty} f_K] = \lim_{K\to\infty} E[f_K]$, which can be accomplished by making the following observations:

Remark 1: $|f_K| \leq g$

Remark 2: $E[g] < \infty$

Remark 3: The desired result, i.e. $E[\lim_{K\to\infty} f_k] = \lim_{K\to\infty} E[f_K]$ follows by Remark 1 and Remark 2 and the Dominated Convergence Theorem.

We now prove the important steps in the proof:

Proof of Remark 1:

$$
\begin{aligned}
|f_K| = |\sum_{k=N+1}^{K}\sum_{l=N+1}^{K} a_k a_l X_{t-l}X_{t-k}| &\le \sum_{k=N+1}^{K}\sum_{l=N+1}^{K} |a_k a_l X_{t-l}X_{t-k}| && \text{(B.10)}\\
&\le \sum_{k=0}^{K}\sum_{l=0}^{K} |a_k a_l X_{t-l}X_{t-k}| && \text{(B.11)}\\
&\le \lim_{K\to\infty}\sum_{k=0}^{K}\sum_{l=0}^{K} |a_k a_l X_{t-l}X_{t-k}| && \text{(B.12)}
\end{aligned}
$$

Proof of Remark 2: We note that $0 \le h_K \le h_{K+1}$. Hence by the Monotone Convergence Theorem we can write $E[\lim_{K\to\infty} h_K] = \lim_{K\to\infty} E[h_K]$. Thus we have the following:

$$
\begin{aligned}
E[g] &= \lim_{K\to\infty} E[h_K] && \text{(B.13)}\\
&= \lim_{K\to\infty} E[\sum_{k=0}^{K}\sum_{l=0}^{K} |a_k||a_l||X_{t-l}X_{t-k}|] && \text{(B.14)}\\
&= \lim_{K\to\infty} \sum_{k=0}^{K}\sum_{l=0}^{K} |a_k||a_l|E[|X_{t-l}X_{t-k}|] && \text{(B.15)}\\
&\le \lim_{K\to\infty} \sum_{k=0}^{K}\sum_{l=0}^{K} |a_k||a_l|\sqrt{3}\sigma_x^2 && \text{(B.16)}\\
&< \infty && \text{(B.17)}
\end{aligned}
$$

where the last strict inequality follows from the fact that $a_l \in l_1$, and $\sigma_x^2 < \infty$. Here (B.16) follows from the fact that $E[|X_{t-l}X_{t-k}|] \le \sqrt{3}\sigma_x^2$, which can be proven as follows:

$$
\begin{aligned}
E[|X_{t-l}X_{t-k}|] &= E[\sqrt{(X_{t-l}X_{t-k})^2}] && \text{(B.18)}\\
&\le \sqrt{E[(X_{t-l}X_{t-k})^2]} && \text{(B.19)}\\
&\le \sqrt[1/4]{E[X_{t-l}^4]E[X_{t-k}^4]} && \text{(B.20)}\\
&= \sqrt[1/4]{3(E[X_{t-l}^2])^2 3(E[X_{t-k}^2])^2} && \text{(B.21)}\\
&= \sqrt[1/4]{9(\sigma_x^2)^4} && \text{(B.22)}\\
&= \sqrt{3}\,\sigma_x^2 && \text{(B.23)}
\end{aligned}
$$

Here (B.19) follows from the Jensen's Inequality, (B.20) follows from the Cauchy–Schwarz inequality, (B.21) follows from the recursive identities for higher order moments of Gaussian random variables, in particular $E[X_t^4] = 3(E[X_t^2])^2$, where $E[X_t] = 0$.

B.3 Proof of Lemma 9.1.4

We now prove $|\sum_{k=N+1}^{\infty}\sum_{l=N+1}^{\infty} a_k a_l r_{k-l}| < \infty$, since $r_x \in l_1(\mathbb{Z})$, and $\{a_k\} \in l_1$.

i) Consider a fixed $k \in \mathbb{N}$. Then $r_x \in l_1 \Rightarrow \quad \{r_{|k-l|}\} \in l_1$, $\forall k \in \mathbb{N}$ since we have the following:

$$\sum_{l=0}^{\infty} |r_{|k-l|}| \quad = \quad \sum_{l=0}^{k} |r_{k-l}| + \sum_{l=k+1}^{\infty} |r_{l-k}| \leq \sum_{l=0}^{k} |r_\tau| + \sum_{l=0}^{\infty} |r_\tau| < \infty \qquad \text{(B.24)}$$

ii) Consider a fixed $k \in \mathbb{N}$. Then $\{a_l\} \in l_1, \{r_{k-l}\} \in l_1 \Rightarrow \{a_l r_{k-l}\} \in l_1$, $\forall k \in \mathbb{N}$ since we have the following: $\{a_l\} \in l_1 \Rightarrow |a_l| \leq |A| < \infty$ and

$$\sum_{l=0}^{\infty} |a_l r_{k-l}| \leq \sum_{l=0}^{\infty} |A||r_{k-l}| \leq |A| \sum_{l=0}^{\infty} |r_{k-l}| < \infty \qquad \text{(B.25)}$$

iii) $\sum_{k=N+1}^{\infty}\sum_{l=N+1}^{\infty} |a_k a_l r_{k-l}| < \infty$ since we have the following:

$$\begin{aligned} \sum_{k=N+1}^{\infty}\sum_{l=N+1}^{\infty} |a_k a_l r_{k-l}| &\leq (\sum_{k=N+1}^{\infty} |a_k|)(\sum_{l=N+1}^{\infty} |a_l r_{k-l}|) && \text{(B.26)} \\ &\leq (\sum_{k=N+1}^{\infty} |a_k|)(\sup_k \sum_{l=N+1}^{\infty} |a_l r_{k-l}|) && \text{(B.27)} \\ &< \infty, && \text{(B.28)} \end{aligned}$$

where $\sup_k \sum_{l=N+1}^{\infty} |a_l r_{k-l}| \leq S < \infty$. We note that by (ii), there exist an $S < \infty$ not dependent on k. The last line follows from $a_l \in l_1$. □

B.4 Proof of Lemma 9.2.1

Here we will prove that the error expression given in (9.68) has a finite limit. We first introduce some shorthand notation. Let us express the MMSE associated with the estimation of X_t from the observations $Y_l, l \in [0, \dots N-1]$ as follows:

$$\varepsilon_t(0, N-1) = E[||X_t - E[X_t|Y_l, l \in [0, \dots N-1]]||^2] \tag{B.29}$$

Hence the MMSE associated with the estimation of X_t based on the observations over $\mathbb{Z}_+$ can be expressed as the following limit:

$$\bar{\varepsilon}_t = \lim_{N \to \infty} \varepsilon_t(0, N-1). \tag{B.30}$$

We note that $\varepsilon_t(0, N-1)$ is always non-negative, and as N increases, the number of Y_l contributing to estimation does not decrease, hence the error do not increase. Hence the limit exists by an application of the monotone convergence theorem; a non-increasing bounded sequence has a finite limit.

For equidistant sampling with sampling interval τ, it is convenient to define the average error over a period, which can be expressed as follows:

$$\varepsilon_l^p(0, N-1) = \frac{1}{\tau} \sum_{t=l\tau}^{(1+l)\tau-1} \varepsilon_t(0, N-1) \tag{B.31}$$

Hence the average MMSE associated with the estimation of X_t in a period based on the observations over $\mathbb{Z}_+$ can be expressed as the following limit:

$$\bar{\varepsilon}_l^p = \lim_{N \to \infty} \varepsilon_l^p(0, N-1). \tag{B.32}$$

Thus, the error expression in (9.68) can be expressed as follows:

$$\begin{aligned} \varepsilon &= \lim_{M \to \infty} \frac{1}{M} \sum_{t=0}^{M-1} \lim_{N \to \infty} E[||X_t - E[X_t|Y_l, l \in [0, \dots N-1]]||^2] && \text{(B.33)} \\ &= \lim_{L \to \infty} \frac{1}{L} \sum_{l=0}^{L} \frac{1}{\tau} \sum_{t=l\tau}^{(1+l)\tau-1} \lim_{N \to \infty} \varepsilon_t(0, N-1) && \text{(B.34)} \\ &= \lim_{L \to \infty} \frac{1}{L} \sum_{l=0}^{L-1} \bar{\varepsilon}_l^p. && \text{(B.35)} \end{aligned}$$

We note that $\{\bar{\varepsilon}_l^p\}$, $l \in \mathbb{Z}_+$ form a non-increasing sequence, which can be proved as follows:

$$\begin{aligned}
\bar{\varepsilon}_l^p &= \lim_{N\to\infty} \varepsilon_l^p(0, N-1) && \text{(B.36)}\\
&= \lim_{N\to\infty} \varepsilon_{l+1}^p(\tau, \tau+N-1) && \text{(B.37)}\\
&\geq \lim_{N\to\infty} \varepsilon_{t+1}^p(0, \tau+N-1) && \text{(B.38)}\\
&= \lim_{N\to\infty} \varepsilon_{l+1}^p(0, N) && \text{(B.39)}\\
&= \bar{\varepsilon}_{l+1}^p && \text{(B.40)}
\end{aligned}$$

Here (B.37) is due to stationarity, (B.38) is due to the fact that possibly increasing number of observations cannot increase error, and (B.39) is due to the fact that we take the limit as $N \to \infty$.

We now consider the following error expression (B.35)

$$\varepsilon = \lim_{L\to\infty} \frac{1}{L} \sum_{l=0}^{L-1} \bar{\varepsilon}_l^p. \qquad \text{(B.41)}$$

As noted above $\{\bar{\varepsilon}_l^p\}$ is a non-increasing sequence. Hence $\frac{1}{L}\sum_{l=0}^{L-1} \bar{\varepsilon}_t^p$, which is the average of a non-increasing sequence, is also non-increasing. So the limit above is guaranteed to exist by monotone convergence theorem. Therefore, the expression for the error given in (9.68) which is the same as (B.41) is guaranteed to converge. □

B.5 Theorem 9.2.2 for Sampling on $\mathbb{Z}$

Here we provide the proof of counterpart of Theorem 9.2.2 (which is for a source on $\mathbb{Z}_+$) for a source on $\mathbb{Z}$. We base our proof directly on a model on $\mathbb{Z}$, instead of taking limits of errors associated with a sequence of finite dimensional models.

Let us first define the equidistant sampling problem on $\mathbb{Z}$. We consider the problem of estimation of stationary zero mean Gaussian source $\{X_t, t \in \mathbb{Z}\}$ from its equidistant noisy samples $\{Y_t, t \in \mathbb{Z}\}$. Let the samples be taken every τ points, i.e. $Y_k = X_{\tau k}$, where $k \in \mathbb{Z}$. As before, $\{Z_t, t \in \mathbb{Z}\}$ is i.i.d. zero-mean Gaussian noise with variance $0 < \sigma_z^2 < \infty$. We assume that $\{Z_t\}$, and $\{X_t\}$ are statistically independent.

As before let $E[X_tX_t] = r_x(k,t) = r_x(k-t)$, $E[X_tY_k] = R_{xy}(t,k)$. We note that since $r_x \in l^1$, so is $R_y(k) = r_x(\tau k) + R_z(\tau k)$. The power spectral density of $\{Y_k\}$, $f_y(\theta), \theta \in [-\pi, \pi]$ can be expressed as follows

$$f_y(\theta) = \sum_m r_y(m)e^{-j\theta m} = \frac{1}{\tau}\sum_{l=0}^{\tau-1} f_x(\frac{\theta + 2\pi l}{\tau}) + \sigma_z^2, \tag{B.42}$$

where $f_z(\theta) = \sum_m r_z(m)e^{-j\theta m} = \sigma_z^2$.

Lemma B.5.1. *Consider the MMSE estimation of $\{X_t, t \in \mathbb{Z}\}$ from $\{Y_t, t \in \mathbb{Z}\}$ as described above. The estimation error is given by the following expression:*

$$E[\lim_{L\to\infty}\frac{1}{L}\sum_{t=0}^{N-1}(X_t - \hat{X}_t)^2] = \lim_{L\to\infty}\frac{1}{L}\sum_{t=0}^{L-1}\lim_{N\to\infty}E[||X_t - E[X_t|Y_l, l \in \Gamma]||^2] \tag{B.43}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}(f_x(\theta) - \frac{1}{\tau^2}\sum_{i=0}^{\tau-1}\frac{(f_x(\frac{\theta+2\pi i}{\tau}))^2}{\frac{1}{\tau}\sum_{l=0}^{\tau-1}f_x(\frac{\theta+2\pi l}{\tau}) + \sigma_z^2})d\theta \tag{B.44}$$

where $\Gamma = \{0, \dots, N-1\}$

Proof: Let the estimator be expressed as follows

$$\hat{X}_t = \sum_{k=-\infty}^{\infty} h_k^t Y_k. \tag{B.45}$$

Here h_k^t is the kth coefficient for estimating the process at time t, that is X_t. The estimator is found by the orthogonality principle, that is the following condition

$$E[(X_t - \sum_{k=-\infty}^{\infty} h_k^t Y_k)Y_m] = 0, \quad \forall\, m \in \mathbb{Z} \tag{B.46}$$

The orthogonality principle can be expressed as follows

$$\sum_{k=-\infty}^{\infty} h_k^t r_y(k-m) \;=\; r_{xy}(t,m) = r_x(t - m\tau). \tag{B.47}$$

We take the discrete time Fourier transform (DTFT) of both sides with the time variable m as follows

$$\begin{aligned}
\sum_m \sum_k h_k^t r_y(k-m) e^{-j\theta m} &= \sum_m r_x(t-m\tau) e^{-j\theta m} && \text{(B.48)}\\
\sum_k h_k^t f_y(\theta) e^{-j\theta k} &= \frac{1}{\tau} e^{-j\frac{\theta}{\tau}} \sum_{i=0}^{\tau-1} e^{-j\frac{2\pi}{\tau}ti} f_x(\frac{\theta+2\pi i}{\tau}) && \text{(B.49)}\\
H_t(\theta) f_y(\theta) &= f_{x_t y}(\theta). && \text{(B.50)}
\end{aligned}$$

Here we have denoted the DTFT of $r_x(t-m\tau)$ with variable m as follows $f_{x_t y}(\theta) = \sum_m r_x(t-m\tau)e^{-j\theta m}$.

The error at time t is given by the following expression

$$\begin{aligned}
e_t &= E[(X_t - \hat{X}_t)^2] && \text{(B.51)}\\
&= E[(X_t - \sum_k h_k^t Y_k)^2] && \text{(B.52)}\\
&= E[(X_t - \sum_k h_k^t Y_k) X_t] && \text{(B.53)}\\
&= E[(X_t - \sum_k h_k^t X_{k\tau}) X_t] && \text{(B.54)}\\
&= r_x(0) - \sum_k h_k^t r_x(k\tau - t) && \text{(B.55)}
\end{aligned}$$

where we have used orthogonality principle to obtain (B.53), and the fact that $E[Z_t X_t] = 0$ to obtain (B.54).

The average error can be expressed as follows

$$\begin{aligned}
\lim_{L\to\infty} \frac{1}{2L+1} \sum_{t=-L}^{L} E[(X_t - \hat{X}_t)^2] &= \lim_{L\to\infty} \frac{1}{2L+1} \sum_{t=-L}^{L} e_t && \text{(B.56)}\\
&= \lim_{M\to\infty} \frac{1}{(2M+1)\tau} \sum_{m=-M}^{M} \sum_{t=m\tau}^{(m+1)\tau-1} e_t && \text{(B.57)}\\
&= \frac{1}{\tau} \sum_{t=0}^{\tau-1} e_t && \text{(B.58)}
\end{aligned}$$

Substituting the expressions for h_k^t and $r_{yx}(k-t)$ results in the following expression

$$\begin{align}
\frac{1}{\tau}\sum_{t=0}^{\tau-1} e_t &= \frac{1}{\tau}\sum_{t=0}^{\tau-1}(r_x(0) - \sum_{k=\infty}^{\infty} h_k^t r_x(k\tau - t)) \tag{B.59}\\
&= \frac{1}{\tau}\sum_{t=0}^{\tau-1}(\frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(\theta) - \frac{|f_{x_t y}(\theta)|^2}{f_y(\theta)} d\theta) \tag{B.60}\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}(f_x(\theta) - \frac{1}{\tau}\sum_{t=0}^{\tau-1}\frac{|\sum_{i=0}^{\tau-1}\frac{1}{\tau}e^{-j\frac{2\pi}{\tau}ti}f_x(\frac{\theta+2\pi t}{\tau}))|^2}{f_y(\theta)})d\theta \tag{B.61}\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(\theta) - \frac{1}{\tau}\sum_{t=0}^{\tau-1}\frac{|\sum_{i=0}^{\tau-1}\frac{1}{\tau}e^{-j\frac{2\pi}{\tau}ti}f_x(\frac{\theta+2\pi t}{\tau}))|^2}{\frac{1}{\tau}\sum_{l=0}^{\tau-1} f_x(\frac{\theta+2\pi l}{\tau}) + \sigma_z^2})d\theta \tag{B.62}\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}(f_x(\theta) - \frac{1}{\tau^2}\sum_{t=0}^{\tau-1}\frac{(f_x(\frac{\theta+2\pi t}{\tau}))^2}{\frac{1}{\tau}\sum_{l=0}^{\tau-1} f_x(\frac{\theta+2\pi l}{\tau}) + \sigma_z^2})d\theta \tag{B.63}
\end{align}$$

where in (B.62) we have used the following fact $f_y(\theta) = \frac{1}{\tau}\sum_{l=0}^{\tau-1} f_x(\frac{\theta+2\pi l}{\tau}) + \sigma_z^2$. (B.63) follows from the following equality

$$\begin{align}
&\frac{1}{\tau}\sum_{t=0}^{\tau-1}|\sum_{l=0}^{\tau-1}\frac{1}{\tau}e^{-j\frac{2\pi}{\tau}tl}f_x(\frac{\theta+2\pi l}{\tau}))|^2 \nonumber\\
&= \frac{1}{\tau^3}\sum_{t=0}^{\tau-1}\sum_{k=0}^{\tau-1}\sum_{l=0}^{\tau-1} e^{-j\frac{2\pi}{\tau}t(k-l)} f_x(\frac{\theta+2\pi k}{\tau}) f_x^{\dagger}(\frac{\theta+2\pi l}{\tau}) \tag{B.64}\\
&= \frac{1}{\tau^3}\sum_{k=0}^{\tau-1}\sum_{l=0}^{\tau-1} f_x(\frac{\theta+2\pi k}{\tau}) f_x^{\dagger}(\frac{\theta+2\pi l}{\tau}) \sum_{t=0}^{\tau-1} e^{-j\frac{2\pi}{\tau}t(k-l)} \tag{B.65}\\
&= \frac{1}{\tau^2}\sum_{k=0}^{\tau-1} f_x^2(\frac{\theta+2\pi k}{\tau}) \tag{B.66}
\end{align}$$

where in (B.66) we have used the following equality

$$\sum_{t=0}^{\tau-1} e^{-i\frac{2\pi}{\tau}(k-l)t} = \begin{cases} \tau, & \text{if } k-l=0,\\ 0, & \text{if } k-l\neq 0.\end{cases} \tag{B.67}$$

B.6 Proof of Lemma 9.2.9

Let $\{T_k\}$ denote the sequence of sampling times defined as follows

$$T_n = \min(k > T_{n-1} : S_k = 1), \tag{B.68}$$

where $T_0 = 0$. We note that if the sampling times were deterministic, by Markov property we would have the following relationship

$$E[X_t|I_t] = E[X_t|X_{T_n}, X_{T_{n+1}}]. \tag{B.69}$$

Let $p > 0$. We make the following important observation: The Markov property can be extended to the Bernoulli sampling scheme by the strong Markov property: conditioned on $T_n < \infty$, and X_{T_n}, $\{X_{T_n+t}, t \geq 0\}$ is again a Markov process. Hence whenever $T_n \leq t \leq T_{n+1} - 1$, we again have the following:

$$E[X_t|I_t] = E[X_t|X_{T_n}, X_{T_{n+1}}]. \tag{B.70}$$

Hence our objective function may be expressed as follows:

$$\begin{aligned}
\varepsilon(p, r) &= \lim_{L\to\infty} E[\frac{1}{L}\sum_{t=0}^{L}[(X_t - E[X_t|I_t])^2] && \text{(B.71)}\\
&= \lim_{L\to\infty} \frac{1}{L}\sum_{m=0}^{L} E[\sum_{t=0}^{L}[(X_t - E[X_t|I_t])^2]|M_L = m]\, P(M_L = m) && \text{(B.72)}\\
&= \lim_{L\to\infty} \frac{1}{L}\sum_{m=1}^{L} E[\sum_{t=0}^{L}[(X_t - E[X_t|I_t])^2]|M_L = m]\, P(M_L = m) && \text{(B.73)}\\
&= \lim_{L\to\infty} \frac{1}{L}\sum_{m=1}^{L} E[\sum_{n=0}^{m-1}\sum_{t=T_n}^{T_{n+1}-1} E[(X_t - E[X_t|X_{T_n}, X_{T_{n+1}}])^2]|M_L = m]\, P(M_L = m) && \text{(B.74)}\\
&= \lim_{L\to\infty} \frac{1}{L}\sum_{m=1}^{L} E[m\sum_{t=0}^{T_1-1} E[(X_t - E[X_t|X_0, X_{T_1}])^2]|M_L = m]\, P(M_L = m) && \text{(B.75)}\\
&= \lim_{L\to\infty} \frac{1}{L}\sum_{m=1}^{L} m\, P(M_L = m)(E[\sum_{t=0}^{T_1-1}[(X_t - E[X_t|X_0, X_{T_1}])^2]]) && \text{(B.76)}\\
&= p\ E[\sum_{t=0}^{T_1-1}[(X_t - E[X_t|X_0, X_{T_1}])^2]], && \text{(B.77)}
\end{aligned}$$

where M_L is the random variable denoting the number of measurements done out of L measurements. Here we have adopted the following convention $T_n = \min(L, \min(k > T_{n-1} : S_k = 1))$. The argument is as follows: In (B.72), we have conditioned on disjoint events. In (B.73) we have changed the limits of summation, since we have $\lim_{L\to\infty} \frac{1}{L}P(M_L = 0) = 0$ and error for any X_t, is uniformly bounded, that is $E[(X_t - E[X_t|I_t, t \in \mathbb{Z}])^2] \leq \sigma_{X_t}^2 = 1$. To obtain

(B.74), we have used the strong Markov property. In (B.75), we have used fresh start property. To obtain (B.77), we have used the fact that the mean of a binomial random variable with probability of succes p and the number of trials L is pL.

We now note the following :

$$\begin{aligned}
&= E[\sum_{k=0}^{T_1-1}[(X_k - E[X_k|X_0, X_{T_1}])^2]|T_1 = t_1] \\
&= E[\sum_{k=0}^{t_1-1}(1 - \frac{1}{1-r^{2t_1}}(r^{2k} - 2r^{2t_1} + r^{(2t_1-2k)}))|T_1 = t_1] \qquad \text{(B.78)} \\
&= \sum_{k=0}^{t_1-1}(1 - \frac{1}{1-r^{2t_1}}(r^{2k} - 2r^{2t_1} + r^{(2t_1-2k)})), \qquad \text{(B.79)}
\end{aligned}$$

where we have used $r_x(t_1 - t_2) = r_x(k) = r^{|k|}$, and $|r| < 1$.

Hence using law of iterated expectations, we can write the following:

$$E[\sum_{k=0}^{T_1-1}[(X_k - E[X_k|X_0, X_{T_1}])^2]] = E[\sum_{k=0}^{T_1-1}(1 - \frac{1}{1-r^{2T_1}}(r^{2k} - 2r^{2T_1} + r^{(2T_1-2k)}))] \qquad \text{(B.80)}$$

If $r = 0$, we note that $\varepsilon(p, r) = 1 - p$. If $p = 0$, $\varepsilon(p, r) = 1 - p = 1$. Now assuming $|r| > 0$, $p > 0$, the error can be expressed as follows:

$$\begin{aligned}
\varepsilon(p, r) &= pE[\sum_{k=0}^{T_1-1}(1 - \frac{1}{1-r^{2T_1}}(r^{2k} - 2r^{2T_1} + r^{(2T_1-2k)}))] \qquad \text{(B.81)} \\
&= p(1/p - \frac{1}{1-r^2} + -2/p + 2E[\frac{T_1}{1-r^{2T_1}}] + \frac{-1}{1-r^{-2}}) \qquad \text{(B.82)} \\
&= p(-1/p - \frac{1+r^2}{1-r^2} + 2E[\frac{T_1}{1-r^{2T_1}}]) \qquad \text{(B.83)} \\
&= -1 + p - \frac{2p}{1-r^2} + 2pE[\frac{T_1}{1-r^{2T_1}}] \qquad \text{(B.84)}
\end{aligned}$$

While evaluating these expressions, we have used the following:

$$E[\sum_{k=0}^{T_1-1} 1] = 1/p \qquad \text{(B.85)}$$

$$E[\sum_{k=0}^{T_1-1}(\frac{1}{1-r^{2T_1}}r^{2k})] = E[\frac{1}{1-r^{2T_1}}\frac{1-r^{2T_1}}{1-r^2}] = \frac{1}{1-r^2} \qquad \text{(B.86)}$$

$$E[\sum_{k=0}^{T_1-1}\frac{1}{1-r^{2T_1}}2r^{2T_1}] = E[T_1\frac{1}{1-r^{2T_1}}2r^{2T_1}] \tag{B.87}$$

$$= E[2T_1(-1+\frac{1}{1-r^{2T_1}})] \tag{B.88}$$

$$= -2/p + E[2T_1(\frac{1}{1-r^{2T_1}})] \tag{B.89}$$

$$E[\sum_{k=0}^{T_1-1}(\frac{1}{1-r^{2T_1}}(r^{(2T_1-2k)}))] = E[\frac{r^{2T_1}}{1-r^{2T_1}}\sum_{k=0}^{T_1-1}r^{-2k}] = \frac{-1}{1-r^{-2}} \tag{B.90}$$

We can express the error more explicitly by rewriting the term with expectation in (B.84) as follows

$$\varepsilon(p,r) = -1+p-\frac{2p}{1-r^2}+2pE[\frac{T_1}{1-r^{2T_1}}] \tag{B.91}$$

$$= -1+p-\frac{2p}{1-r^2}+2p\sum_{t_1=1}^{\infty}\frac{t_1}{1-r^{2t_1}}(1-p)^{(t_1-1)}p \tag{B.92}$$

$$= -1+p-\frac{2p}{1-r^2}+2p^2\sum_{k=0}^{\infty}\frac{r^{2k}}{(1-(1-p)r^{2k})^2} \tag{B.93}$$

To obtain (B.93), we have used the following

$$E[\frac{T_1}{1-r^{2T_1}}] = \sum_{t_1=1}^{\infty}\frac{t_1}{1-r^{2t_1}}(1-p)^{(t_1-1)}p \tag{B.94}$$

$$= \sum_{t_1=1}^{\infty}t_1(1-p)^{(t_1-1)}p(\sum_{k=0}^{\infty}r^{2t_1k}) \tag{B.95}$$

$$= p\sum_{k=0}^{\infty}r^{2k}\sum_{t_1=1}^{\infty}t_1(1-p)^{(t_1-1)}(r^{2k})^{(t_1-1)} \tag{B.96}$$

$$= p\sum_{k=0}^{\infty}\frac{r^{2k}}{(1-(1-p)r^{2k})^2}, \tag{B.97}$$

where we've used the following property

$$\sum_{t_1=1}^{\infty}t_1a^{(t_1-1)} = \frac{1}{(1-a)^2}. \tag{B.98}$$

Here $a=(1-p)r^{2k}$, $|a|<1$. □

Bibliography

[1] A. V. Oppenheim and A. S.Willsky, *Signals and Systems*, 2nd ed. Prentice Hall, 1997.

[2] A. Özçelikkale and H. M. Ozaktas, "Optimal representation of non-stationary random fields with finite numbers of samples," Submitted.

[3] A. Özçelikkale, "Structural and metrical information in linear systems," Master's thesis, Bilkent University, Ankara, Turkey, 2006.

[4] A. Özçelikkale, H. M. Özaktaş, and E. Arıkan, "Measurement strategies for input estimation in linear systems (In Turkish: Doğrusal sistemlerde girdi kestirimi için ölçüm yöntemleri)," in *Proc. 2007 IEEE Signal Process. and Commun. App. Conf.*

[5] A. Özçelikkale, H. M. Ozaktas, and E. Arıkan, "Optimal measurement under cost constraints for estimation of propagating wave fields," in *Proc. 2007 IEEE Int. Symp. Information Theory*, pp. 696–700.

[6] ——, "Signal recovery with cost constrained measurements," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3607–3617, Jul. 2010.

[7] A. Özçelikkale and H. M. Ozaktas, "Representation of optical fields using finite numbers of bits," *Opt. Lett.*, vol. 37, no. 12, pp. 2193–2195, Jun. 2012.

[8] ——, "Beyond Nyquist sampling: A cost-based framework," 2012, Submitted.

[9] P. Flandrin, A. Napolitano, H. M. Ozaktas, and D. J. Thomson, “Recent advances in theory and methods for nonstationary signal analysis,” *Special Issue of EURASIP Journal on Advances in Signal Processing*, 2011.

[10] W. A. Gardner, “A sampling theorem for nonstationary random processes,” *IEEE Trans. Inf. Theory*, vol. 18, no. 6, pp. 808 – 809, 1972.

[11] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21 – 36, May 2003.

[12] A. Özçelikkale, G. B. Akar, and H. M. Ozaktas, “Super-resolution using multiple quantized images,” *2010 IEEE Int. Conf. on Image Processing*, pp. 2029–2032.

[13] E. J. Candes and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007.

[14] D. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845 –2862, Nov. 2001.

[15] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics.* Cambridge University Press, 1995.

[16] J. W. Goodman, *Statistical Optics.* Wiley, 2000.

[17] J. A. Tropp, “On the conditioning of random subdictionaries,” *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 1 – 24, 2008.

[18] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing.* Wiley, 2001.

[19] A. Özçelikkale, S. Yüksel, and H. M. Özaktaş, “Average error in recovery of sparse signals and discrete Fourier transform (In Turkish: Seyrek işaretlerin geri kazanımında ortalama hata ve ayrık Fourier dönüşümü),” in *2012 IEEE Signal Process. and Commun. App. Conf.*, Apr. 2012.

[20] A. Özçelikkale, S. Yüksel, and H. M. Ozaktas, “Unitary precoding and basis dependency of MMSE performance for Gaussian erasure channels,” *preprint*, Nov. 2011, available as arXiv:1111.2451v1 [cs.IT].

[21] R. C. Bradley, “Basic properties of strong mixing conditions. a survey and some open questions,” *Probability Surveys*, vol. 2, pp. 107–144, 2005.

[22] D. Gabor, “Light and information,” in *Progress In Optics*, E. Wolf, Ed. Elsevier, 1961, vol. I, ch. 4, pp. 109–153.

[23] G. Toraldo Di Francia, “Resolving power and information,” *J. Opt. Soc. Am.*, vol. 45, no. 7, pp. 497–501, Jul. 1955.

[24] ——, “Degrees of freedom of an image,” *J. Opt. Soc. Am.*, vol. 59, no. 7, pp. 799–804, Jul. 1969.

[25] W. Lukozs, “Optical systems with resolving powers exceeding the classical limit,” *J. Opt. Soc. Am.*, vol. 56, no. 11, pp. 1463–1472, Nov. 1966.

[26] ——, “Optical systems with resolving powers exceeding the classical limit II,” *J. Opt. Soc. Am.*, vol. 57, no. 7, pp. 932–941, Jul. 1967.

[27] F. Gori and G. Guattari, “Effects of coherence on the degrees of freedom of an image,” *J. Opt. Soc. Am.*, vol. 61, no. 1, pp. 36–39, Jan. 1971.

[28] ——, “Shannon number and degrees of freedom of an image,” *Opt. Commun.*, vol. 7, no. 2, pp. 163–165, Feb. 1973.

[29] L. Ronchi and F. Gori, “Degrees of freedom for spherical scatterers,” *Opt. Lett.*, vol. 6, no. 10, pp. 478–480, Oct. 1981.

[30] R. Pierri and F. Soldovieri, “On the information content of the radiated fields in the near zone over bounded domains,” *Inv. Probl.*, vol. 14, no. 2, pp. 321–337, Jan. 1998.

[31] A. Starikov, “Effective number of degrees of freedom of partially coherent sources,” *J. Opt. Soc. Am.*, vol. 72, no. 11, pp. 1538–1544, 1982.

[32] G. Newsam and R. Barakat, “Essential dimension as a well-defined number of degrees of freedom of finite-convolution operators appearing in optics,” *J. Opt. Soc. Am. A*, vol. 2, no. 11, pp. 2040–2045, Jan. 1985.

[33] A. Lohmann, R. Dorsch, D. Mendlovic, Z. Zalevsky, and C. Ferreira, “Space-bandwidth product of optical signals and systems,” *J. Opt. Soc. Am. A*, vol. 13, no. 3, pp. 470–473, Jan. 1996.

[34] D. Mendlovic and A. W. Lohmann, “Space-bandwidth product adaptation and its application to superresolution: Fundamentals,” *J. Opt. Soc. Am. A*, vol. 14, no. 3, pp. 558–562, Mar. 1997.

[35] F. S. Oktem and H. M. Ozaktas, “Equivalence of linear canonical transform domains to fractional Fourier domains and the bicanonical width product: a generalization of the space–bandwidth product,” *J. Opt. Soc. Am. A*, vol. 27, no. 8, pp. 1885–1895, Aug 2010.

[36] Z. Zalevsky and D. Mendlovic, *Optical Superresolution.* Springer-Verlag, 2003.

[37] D. MacKay, “Quantal aspects of scientific information,” *IEEE Trans. Inf. Theory*, vol. 1, no. 1, pp. 60–80, Feb. 1953.

[38] J. T. Winthrop, “Propagation of structural information in optical wave fields,” *J. Opt. Soc. Am.*, vol. 61, no. 1, pp. 15–30, Jan. 1971.

[39] T. W. Barret, “Structural information theory,” *J. Acoust. Soc. Am.*, vol. 54, no. 4, pp. 1092–1098, Oct. 1973.

[40] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 1991.

[41] D. A. B. Miller, “Spatial channels for communicating with waves between volumes,” *Opt. Lett.*, vol. 23, no. 21, pp. 1645–1647, Nov 1998.

[42] ——, “Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths,” *Appl. Opt.*, vol. 39, no. 11, pp. 1681–1699, Jan. 2000.

[43] R. Piestun and D. A. B. Miller, "Electromagnetic degrees of freedom of an optical system," *J. Opt. Soc. Am. A*, vol. 17, no. 5, pp. 892–902, May 2000.

[44] A. Thaning, P. Martinsson, M. Karelin, and A. T. Friberg, "Limits of diffractive optics by communication modes," *J. Opt. A: Pure Appl. Opt.*, vol. 5, no. 3, pp. 153–158, May 2003.

[45] A. Burvall, P. Martinsson, and A. T. Friberg, "Communication modes in large-aperture approximation," *Opt. Lett.*, vol. 32, no. 6, pp. 611–613, Mar. 2007.

[46] D. Blacknell and C. J. Oliver, "Information-content of coherent images," *J. Phys. D*, vol. 26, no. 9, pp. 1364–1370, Jan. 1993.

[47] M. A. Neifeld, "Information, resolution, and space-bandwidth product," *Opt. Lett.*, vol. 23, no. 18, pp. 1477–1479, Sep 1998.

[48] A. Stern and B. Javidi, "Shannon number and information capacity of three-dimensional integral imaging," *J. Opt. Soc. Am. A*, vol. 21, no. 9, pp. 1602–1612, Sep. 2004.

[49] M. Migliore, "On electromagnetics and information theory," *IEEE Trans. Antennas Propag.*, vol. 56, no. 10, pp. 3188–3200, Oct. 2008.

[50] E. D. Micheli and G. A. Viano, "Inverse optical imaging viewed as a backward channel communication problem," *J. Opt. Soc. Am. A*, vol. 26, no. 6, pp. 1393–1402, Jan. 2009.

[51] P. Réfrégier and J. Morio, "Shannon entropy of partially polarized and partially coherent light with Gaussian fluctuations," *J. Opt. Soc. Am. A*, vol. 23, no. 12, pp. 3036–3044, Dec. 2006.

[52] M. S. Hughes, "Analysis of digitized wave-forms using Shannon entropy," *J. Acoust. Soc. Am.*, vol. 93, no. 2, pp. 892–906, Jan. 1993.

[53] A. Morozov and J. Colosi, "Entropy of acoustical beams in a random ocean," in *Proc. of Oceans 2003*, pp. 558–563.

[54] F. T. Yu, *Optics and Information Theory.* Wiley , 1976.

[55] ——, *Entropy and Information Optics.* Marcel Dekker, 2000.

[56] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 1991.

[57] M. J. Bastiaans, “Uncertainty principle and informational entropy for partially coherent light,” *J. Opt. Soc. Am. A*, vol. 3, no. 8, pp. 1243–1246, Aug. 1986.

[58] H. M. Ozaktas, S. Yüksel, and M. A. Kutay, “Linear algebraic theory of partial coherence: discrete fields and measures of partial coherence,” *J. Opt. Soc. Am. A*, vol. 19, no. 8, pp. 1563–1571, Aug. 2002.

[59] R. Barakat, “Some entropic aspects of optical diffraction imagery,” *Opt. Commun.*, vol. 156, no. 6, pp. 235–239, Nov. 1998.

[60] M. A. Porras and R. Medina, “Entropy-based definition of laser beam spot size,” *Applied Optics*, vol. 34, no. 36, pp. 8247–8251, December 1995.

[61] H. M. Ozaktas, A. Koç, I. Sari, and M. A. Kutay, “Efficient computation of quadratic-phase integrals in optics,” *Opt. Lett.*, vol. 31, no. 1, pp. 35–37, Jan 2006.

[62] A. Koç, H. M. Ozaktas, C. Candan, and M. A. Kutay, “Digital computation of linear canonical transforms,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2383–2394, Jun. 2008.

[63] F. Oktem and H. Ozaktas, “Exact relation between continuous and discrete linear canonical transforms,” *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 727 –730, Aug. 2009.

[64] J. J. Healy and J. T. Sheridan, “Space–bandwidth ratio as a means of choosing between Fresnel and other linear canonical transform algorithms,” *J. Opt. Soc. Am. A*, vol. 28, no. 5, pp. 786–790, May 2011.

[65] F. S. Roux, “Complex-valued Fresnel-transform sampling,” *Appl. Opt.*, vol. 34, no. 17, pp. 3128–3135, Jan. 1995.

[66] A. Stern and B. Javidi, "Analysis of practical sampling and reconstruction from Fresnel fields," *Opt. Eng.*, vol. 43, no. 1, pp. 239–250, Jan. 2004.

[67] L. Onural, "Exact analysis of the effects of sampling of the scalar diffraction field," *J. Opt. Soc. Am. A*, vol. 24, no. 2, pp. 359–367, Jan. 2007.

[68] J. J. Healy, B. M. Hennelly, and J. T. Sheridan, "Additional sampling criterion for the linear canonical transform," *Opt. Lett.*, vol. 33, no. 22, pp. 2599–2601, Nov. 2008.

[69] F. Gori, "Advanced topics in Shannon sampling and interpolation theory," R. J. M. II, Ed. Springer-Verlag, 1993, ch. 2, pp. 37–83.

[70] E. N. Leith, "The evaluation of information optics," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 6, pp. 1297–1304, November/December 2000.

[71] ——, "Some highlights in the history of information optics," *Information Sciences*, vol. 149, pp. 271–275, November/December 2003.

[72] O. Bucci and G. Franceschetti, "On the degrees of freedom of scattered fields," *IEEE Trans. Antennas Propag.*, vol. 37, no. 7, pp. 918–926, Jul. 1989.

[73] O. Bucci, C. Gennarelli, and C. Savarese, "Representation of electromagnetic fields over arbitrary surfaces by a finite and nonredundant number of samples," *IEEE Trans. Antennas Propag.*, vol. 46, no. 3, pp. 351 –359, Mar. 1998.

[74] A. Poon, R. Brodersen, and D. Tse, "Degrees of freedom in multiple-antenna channels: a signal space approach," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 523– 536, Feb. 2005.

[75] R. Kennedy, P. Sadeghi, T. Abhayapala, and H. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2542–2556, Jun. 2007.

[76] J. Xu and R. Janaswamy, “Electromagnetic degrees of freedom in 2-D scattering environments,” *IEEE Trans. Antennas Propag.*, vol. 54, no. 12, pp. 3882–3894, Dec. 2006.

[77] F. Gruber and E. Marengo, “New aspects of electromagnetic information theory for wireless and antenna systems,” *IEEE Trans. Antennas Propag.*, vol. 56, no. 11, pp. 3470–3484, Nov. 2008.

[78] M. Jensen and J. Wallace, “Capacity of the continuous-space electromagnetic channel,” *IEEE Trans. Antennas Propag.*, vol. 56, no. 2, pp. 524–531, Feb. 2008.

[79] L. Hanlen and M. Fu, “Wireless communication systems with-spatial diversity: a volumetric model,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 133– 142, Jan. 2006.

[80] J. Gubner, “Distributed estimation and quantization,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.

[81] W.-M. Lam and A. Reibman, “Design of quantizers for decentralized estimation systems,” *IEEE Trans. Commun.*, vol. 41, no. 11, pp. 1602–1605, Nov. 1993.

[82] S. Marano, V. Matta, and P. Willett, “Quantizer precision for distributed estimation in a large sensor network,” *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 4073–4078, Oct. 2006.

[83] J. Li and G. AlRegib, “Rate-constrained distributed estimation in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1634–1643, May 2007.

[84] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. Poor, “Estimation diversity and energy efficiency in distributed sensing,” *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[85] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471 – 480, Jul. 1973.

[86] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[87] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder II: General sources," *Information and Control*, vol. 38, no. 1, pp. 60–80, July 1978.

[88] P. Viswanath, "Sum rate of a class of Gaussian multiterminal source coding problems," in *Advances in Network Information Theory.* American Mathematical Society, 2004, pp. 43–60.

[89] E. Haroutunian, "Multiterminal source coding achievable rates and reliability," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 2094–2101, Nov. 1996.

[90] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.

[91] P. Ishwar, R. Puri, K. Ramchandran, and S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 765–775, Apr. 2005.

[92] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938 –1961, May 2008.

[93] T. Berger, *Rate-distortion theory: A mathematical basis for data compression.* Prentice Hall, 1971.

[94] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 773–787, Nov. 1987.

[95] H. Yamamoto, "Wyner - Ziv theory for a general function of the correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 5, pp. 803–807, Sep. 1982.

[96] R. L. Dobrushin and B. S. Tsybakov, “Information transmission with additional noise,” *IEEE Trans. Inf. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[97] H. S. Witsenhausen, “Indirect rate distortion problems,” *IEEE Trans. Inf. Theory*, vol. IT-26, no. 5, pp. 518–521, Sep. 1980.

[98] E. Ayanoglu, “On optimal quantization of noisy sources,” *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1450–1452, Nov 1990.

[99] T. Berger, Z. Zhang, and H. Viswanathan, “The CEO problem,” *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[100] H. Viswanathan and T. Berger, “The quadratic Gaussian CEO problem,” *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.

[101] Y. Oohama, “The rate-distortion function for the quadratic Gaussian CEO problem,” *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[102] Z. Cvetkovic and M. Vetterli, “On simple oversampled A/D conversion in L2(R),” *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 146–154, Jan. 2001.

[103] A. Kumar, P. Ishwar, and K. Ramchandran, “High-resolution distributed sampling of bandlimited fields with low-precision sensors,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 476 –492, Jan. 2011.

[104] B. Dulek and S. Gezici, “Average Fisher information maximisation in presence of cost-constrained measurements,” *Electronics Lett.*, vol. 47, no. 11, pp. 654 –656, 26 2011.

[105] ——, “Cost minimization of measurement devices under estimation accuracy constraints in the presence of Gaussian noise,” *Digital Signal Processing*, vol. 22, no. 5, pp. 828 – 840, 2012.

[106] S. Nordebo and M. Gustafsson, “Statistical signal analysis for the inverse source problem of electromagnetics,” *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2357– 2361, Jun. 2006.

[107] T. Oliphant, “On parameter estimates of the lossy wave equation,” *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 49–60, Jan. 2008.

[108] T. C. Güçlü, “Finite representation of finite energy signals,” Master’s thesis, Bilkent University, Ankara, Turkey, 2011.

[109] J.-J. Xiao and Z.-Q. Luo, “Decentralized estimation in an inhomogeneous sensing environment,” *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.

[110] A. Ribeiro and G. Giannakis, “Bandwidth-constrained distributed estimation for wireless sensor networks -part I: Gaussian case,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[111] S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[112] M. Lazaro, M. Sanchez-Fernandez, and A. Artes-Rodriguez, “Optimal sensor selection in binary heterogeneous sensor networks,” *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1577–1587, Apr. 2009.

[113] R. Konsbruck, E. Telatar, and M. Vetterli, “On the multiterminal rate-distortion function for acoustic sensing,” in *Proc. 2006 IEEE Int. Conf. Acous., Speech and Signal Proc.*, vol. 4, 2006, pp. 701–704.

[114] ——, “On sampling and coding for distributed acoustic sensing,” *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3198 –3214, may 2012.

[115] A. Balakrishnan, “A note on the sampling principle for continuous signals,” *IRE Trans. Inf. Theory*, vol. 3, no. 2, pp. 143 –146, Jun. 1957.

[116] S. P. Lloyd, “A sampling theorem for stationary (wide-sense) stochastic processes,” *Transactions of the American Mathematical Society*, vol. 92, no. 1, pp. pp. 1–12, Jul. 1959.

[117] M. I. Kadec, “The exact value of the Paley-Wiener constant,” *Dokl. Akad. Nauk SSSR 155*, pp. 1253 – 1254, 1964.

[118] R. M. Young, *An Introduction to Nonharmonic Fourier Series.* New York: Academic Press, 1980.

[119] F. J. Beutler, "Sampling theorems and bases in a Hilbert space," *Information and Control*, vol. 4, no. 2-3, pp. 97–117, 1961.

[120] ——, "Error-free recovery of signals from irregularly spaced samples," *SIAM Review*, vol. 8, no. 3, pp. 328 – 335, 1966.

[121] Z. Song, W. Sun, X. Zhou, and Z. Hou, "An average sampling theorem for bandlimited stochastic processes," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4798 –4800, Dec. 2007.

[122] J. L. Brown, "On mean-square aliasing error in cardinal series expansion of random processes," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 2, pp. 254 – 256, Mar. 1978.

[123] A. Papoulis, "A note on the predictability of band-limited processes," *Proceedings of the IEEE*, vol. 73, no. 8, pp. 1332 – 1333, aug. 1985.

[124] J. Medina and B. Cernuschi-Frias, "On the prediction of a class of wide-sense stationary random processes," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 70 –77, Jan. 2011.

[125] E. Masry, "Polynomial interpolation and prediction of continuous-time processes from random samples," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 776–783, 1997.

[126] F. Garcia, I. Lourtie, and J. Buescu, "$L_2(R)$ nonstationary processes and the sampling theorem," *IEEE Signal Processing Letters,*, vol. 8, no. 4, pp. 117 –119, Apr. 2001.

[127] A. Napolitano, "Sampling theorems for Doppler-stretched wide-band signals," *Signal Processing*, vol. 90, no. 7, pp. 2276 – 2287, 2010.

[128] ——, "Sampling of spectrally correlated processes," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 525 –539, Feb. 2011.

[129] E. Masry, "On the truncation error of the sampling expansion for stationary bandlimited processes," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2851 –2853, Oct. 1994.

[130] H. Boche and U. J. Mönich, "Convergence behavior of non-equidistant sampling series," *Signal Processing*, vol. 90, no. 1, pp. 145 – 156, 2010.

[131] L. D. Davisson, "Prediction of time series from finite past," *Journal of the Society for Industrial and Applied Mathematics*, vol. 13, no. 3, pp. pp. 819–826, 1965.

[132] U. Grenander and G. Szegö, *Toeplitz forms and their applications.* University of California Press, 1958.

[133] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I.* Wiley, 2001.

[134] F. Gori, "Collett-Wolf sources and multimode lasers," *Opt. Commun.*, vol. 34, no. 3, pp. 301 – 305, 1980.

[135] A. Starikov and E. Wolf, "Coherent-mode representation of Gaussian-Schell model sources and of their radiation fields," *J. Opt. Soc. Am. A*, vol. 72, no. 7, pp. 923–928, 1982.

[136] A. T. Friberg and R. J. Sudol, "Propagation parameters of Gaussian Schell-model beams," *Opt. Commun.*, vol. 41, no. 6, pp. 383 – 387, 1982.

[137] A. T. Friberg and J. Turunen, "Imaging of Gaussian-Schell model sources," *J. Opt. Soc. Am. A*, vol. 5, no. 5, pp. 713–720, Jan 1988.

[138] P. Jixiong, "Waist location and Rayleigh range for Gaussian Schell-model beams," *J. Opt.*, vol. 22, no. 3, pp. 157–159, 1991.

[139] H. Yoshimura and T. Iwai, "Properties of the Gaussian Schell-model source field in a fractional Fourier plane," *Opt. Act.*, vol. 14, no. 12, pp. 3388–3393, Dec 1997.

[140] G. Gbur and E. Wolf, "The Rayleigh range of Gaussian Schell-model beams," *J. Mod. Opt.*, vol. 48, no. 11, pp. 1735–1741, Sep 2001.

[141] Q. Lin and Y. Cai, “Fractional Fourier transform for partially coherent Gaussian–Schell model beams,” *Opt. Lett.*, vol. 27, no. 19, pp. 1672–1674, Oct 2002.

[142] T. Shirai, A. Dogariu, and E. Wolf, “Directionality of Gaussian Schell-model beams propagating in atmospheric turbulence,” *Opt. Lett.*, vol. 28, no. 8, pp. 610–612, Apr 2003.

[143] S. Zhu, Y. Cai, and O. Korotkova, “Propagation factor of a stochastic electromagnetic Gaussian Schell-model beam,” *Opt. Express*, vol. 18, no. 12, pp. 12 587–12 598, Jun 2010.

[144] Y. Dan, S. Zeng, B. Hao, and B. Zhang, “Range of turbulence-independent propagation and Rayleigh range of partially coherent beams in atmospheric turbulence,” *J. Opt. Soc. Am. A*, vol. 27, no. 3, pp. 426–434, Mar 2010.

[145] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. Mcgraw-Hill, 1991.

[146] E. Collett and E. Wolf, “Beams generated by Gaussian quasi-homogeneous sources,” *Opt. Commun.*, vol. 32, no. 1, pp. 27 – 31, 1980.

[147] B. E. A. Saleh and M. C. Teich, *Fundamental of Photonics.* Wiley, 1991.

[148] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1990.

[149] J. Buescu, “Positive integral operators in unbounded domains,” *Journal of Mathematical Analysis and Applications*, vol. 296, no. 1, pp. 244 – 255, 2004.

[150] M. J. Bastiaans, “Wigner distribution function and its application to first-order optics,” *J. Opt. Soc. Am. A*, vol. 69, no. 12, pp. 1710–1716, Dec 1979.

[151] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[152] R. V. L. Hartley, “Transmission of information,” *Bell Syst. Tech. J.*, vol. 7, pp. 535–563, Jul. 1928.

[153] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 , 623–656, Jul. - Oct. 1948.

[154] B. M. Oliver, J. R. Pierce, and C. E. Shannon, “The philosophy of PCM,” in *Proc. of the I.R.E.*, vol. 36, Nov. 1948, pp. 1324–1331.

[155] R. G. Gallager, *Information Theory and Reliable Communication.* Wiley, 1968.

[156] J. Nocedal and S. J. Wright, *Numerical Optimization.* Springer, 2006.

[157] J. Gorski, F. Pfeuffer, and K. Klamroth, “Biconvex sets and optimization with biconvex functions - a survey and extensions,” *Mathematical Methods of Operations Research*, vol. 66, no. 3, p. 373–408, Jun. 2007.

[158] P. M. Morse and K. U. Ingard, *Theoretical Acoustics.* Princeton University Press, 1986.

[159] Constantine A. Balanis, *Antenna Theory: Analysis and Design.* Wiley, 2005.

[160] L. Onural and H. M. Ozaktas, “Signal processing issues in diffraction and holographic 3DTV,” *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 169–177, Feb. 2007.

[161] M. J. Bastiaans, “Applications of the Wigner distribution function in optics,” in *The Wigner Distribution: Theory and Applications in Signal Processing*, W. Mecklenbrauker and F. Hlawatsch, Eds. Elsevier, 1997, pp. 375–426.

[162] H. M. Ozaktas, B. Barshan, D. Mendlovic, and L. Onural, “Convolution, filtering, and multiplexing in fractional Fourier domains and their relation to chirp and wavelet transform,” *J. Opt. Soc. Am. A*, vol. 11, no. 2, pp. 547–559, Feb. 1994.

[163] M. A. Kutay, H. Özaktaş, H. M. Ozaktas, and O. Arıkan, "The fractional Fourier domain decomposition," *Signal Process.*, vol. 77, no. 1, pp. 105–109, Aug. 1999.

[164] M. F. Erden, M. A. Kutay, and H. M. Ozaktas, "Repeated filtering in consecutive fractional Fourier domains and its application to signal restoration," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1458–1462, May 1999.

[165] H. M. Ozaktas and U. Sümbül, "Interpolating between periodicity and discreteness through the fractional Fourier transform," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4233–4243, Nov. 2006.

[166] S. Qazi, A. Georgakis, L. K. Stergioulas, and M. Shikh-Bahaei, "Interference suppression in the Wigner distribution using fractional Fourier transformation and signal synthesis," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3150–3154, Jun. 2007.

[167] S.-C. Pei and J.-J. Ding, "Relations between Gabor transforms and fractional Fourier transforms and their applications for signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4839–4850, Oct. 2007.

[168] A. S. Amein and J. J. Soraghan, "Fractional chirp scaling algorithm: Mathematical model," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4162–4172, Aug. 2007.

[169] K. K. Sharma and S. D. Joshi, "Uncertainty principle for real signals in the linear canonical transform domains," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2677–2683, Jul. 2008.

[170] R. Tao, X.-M. Li, Y.-L. Li, and Y. Wang, "Time-Delay Estimation of Chirp Signals in the Fractional Fourier Domain," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2852–2855, Jul. 2009.

[171] Ç. Candan, M. A. Kutay, and H. M. Ozaktas, "The discrete fractional Fourier transform," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1329–1337, May 2000.

[172] Ç. Candan, “Discrete fractional Fourier transform matrix generator,” http://www.ee.bilkent.edu.tr/∼haldun/dFRT.m, July 1998.

[173] K. B. Wolf, *Integral Transforms in Science and Engineering.* Plenum Publ. Corp, 1979.

[174] W. H. Carter and E. Wolf, “Correlation theory of wavefields generated by fluctuating, three-dimensional, scalar sources II: Radiation from isotropic model sources,” *J. Opt.*, vol. 28, no. 2, pp. 245–259, 1981.

[175] S. Mann and R. Picard, “Being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures,” *Proc. of IST’s 48th Annual Conference*, pp. 442 – 448, May 1995.

[176] T. Mitsunaga and S. Nayar, “Radiometric self calibration,” in *Proc. of 1999 IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 637–663.

[177] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, “Fast and robust multi-frame super resolution,” *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327 – 1344, Oct. 2004.

[178] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[179] L. Fei-Fei and P. Perona, “A Bayesian hierarchical model for learning natural scene categories,” in *Proc. of 2005 IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 524 – 531.

[180] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. on Telecommunications*, vol. 10, pp. 585–595, 1999.

[181] J. Romberg and M. Wakin, “Compressed sensing: A tutorial,” *IEEE Statistical Signal Processing Workshop*, Aug. 2007.

[182] H. Gamo, “Intensity matrix and degree of coherence,” *J. Opt. Soc. Am.*, vol. 47, no. 10, pp. 976–976, Oct. 1957.

[183] ——, "Matrix treatment of partial coherence," in *Progress in Optics Volume III*, E. Wolf, Ed., 1964, pp. 187–332.

[184] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21 –30, Mar. 2008.

[185] E. J. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Found. Comput. Math.*, vol. 6, pp. 227–254, Apr. 2006.

[186] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489 – 509, Feb. 2006.

[187] J. A. Tropp, "The random paving property for uniformly bounded matrices," *Studia Mathematica,*, vol. 185, no. 1, pp. 67–82, 2008.

[188] B. D. O. Anderson and J. B. Moore, *Optimal filtering.* Prentice-Hall, 1979.

[189] A. Tulino, S. Verdu, G. Caire, and S. Shamai, "The Gaussian erasure channel," in *IEEE International Symposium on Inf. Theory, 2007*, Jun. 2007, pp. 1721 –1725.

[190] ——, "The Gaussian erasure channel," *preprint*, Jul. 2007.

[191] T. Basar, "A trace minimization problem with applications in joint estimation and control under nonclassical information," *Journal of Optimization Theory and Applications*, vol. 31, no. 3, pp. 343–359, 1980.

[192] H. S. Witsenhausen, "A determinant maximization problem occurring in the theory of data communication," *SIAM Journal on Applied Mathematics*, vol. 29, no. 3, pp. 515–522, 1975.

[193] Y. Wei, R. Wonjong, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145 – 152, Jan. 2004.

[194] F. Perez-Cruz, M. Rodrigues, and S. Verdu, “MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation,” *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070 –1084, Mar. 2010.

[195] K.-H. Lee and D. Petersen, “Optimal linear coding for vector channels,” *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1283 – 1290, Dec. 1976.

[196] J. Yang and S. Roy, “Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback,” *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1334 –1347, Sep. 1994.

[197] D. Palomar, J. Cioffi, and M. Lagunas, “Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization,” *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381 – 2401, Sep. 2003.

[198] D. Palomar, “Unified framework for linear MIMO transceivers with shaping constraints,” *IEEE Commun. Lett.*, vol. 8, no. 12, pp. 697 – 699, Dec. 2004.

[199] A. Kashyap, T. Basar, and R. Srikant, “Minimum distortion transmission of Gaussian sources over fading channels,” in *Proc. of 2003 IEEE Conf. on Decision and Control*, vol. 1, Dec., pp. 80 – 85.

[200] M. Elad and I. Yavneh, “A plurality of sparse representations is better than the sparsest one alone,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4701–4714, Oct. 2009.

[201] M. Protter, I. Yavneh, and M. Elad, “Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary,” *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3471–3484, Jul. 2010.

[202] R. M. Gray, “Toeplitz and circulant matrices: a review,” *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–329, 2006, *Available as a paperback book from* Now Publishers Inc.

[203] H. V. Henderson and S. R. Searle, “On deriving the inverse of a sum of matrices,” *SIAM Review*, vol. 23, no. 1, pp. 53–60, 1981.

[204] S. Chrétien and S. Darses, "Invertibility of random submatrices via tail-decoupling and a matrix Chernoff Inequality," *ArXiv e-prints*, Mar. 2012.

[205] J. A. Tropp, "Norms of random submatrices and sparse approximation," *C. R. Math. Acad. Sci. Paris*, vol. 346, pp. 1271–1274, 2008.

[206] M. Rudelson and R. Vershynin, "The Littlewood-Offord problem and invertibility of random matrices," *Advances in Mathematics*, vol. 218, pp. 600 – 633, 2008.

[207] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, "Smallest singular value of random matrices and geometry of random polytopes," *Adv. Math*, vol. 195, pp. 491–523, 2005.

[208] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[209] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proceedings,*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[210] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740 –2746, Jun. 2007.

[211] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems.* Akadémiai Kiadó, 1997.

[212] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[213] G. Grimmet and D. Stirzaker, *Probability and Random Processes.* Oxford University Press, 2009.

[214] I. Ibragimov and Y. A. Rozanov, *Gaussian Random Processes.* Springer-Verlag, 1978.

[215] L. H. Koopmans, *The spectral analysis of time series.* Academic Press, 1995.

[216] S. Jaffard, “Propriétés des matrices “bien localisées” près de leur diagonale et quelques applications,” *Ann. Inst. H. Poincaré Anal. Non Linéaire*, vol. 7, pp. 461–476, 1990.

[217] T. Strohmer, “Four short stories about Toeplitz matrix calculations,” *Linear Algebra and its Applications*, vol. 343-344, pp. 321 – 344, 2002.

[218] F.-W. Sun, Y. Jiang, and J. Baras, “On the convergence of the inverses of Toeplitz matrices and its applications,” *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 180 – 190, Jan. 2003.

[219] T. Hashimoto and S. Arimoto, “On the rate-distortion function for the nonstationary Gaussian autoregressive process,” *IEEE Trans. Inf. Theory*, vol. 26, no. 4, pp. 478 – 480, Jul. 1980.

[220] R. Gray and T. Hashimoto, “A note on rate-distortion functions for nonstationary Gaussian autoregressive processes,” *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1319 –1322, Mar. 2008.

[221] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, “Kalman filtering with intermittent observations,” in *Proc. of 42. IEEE Conf. on Decision and Control*, vol. 1, pp. 701 – 708.

[222] H. Rauhut, “Compressive sensing and structured random matrices,” in *Theoretical Foundations and Numerical Methods for Sparse Recovery, Radon Series Comp. Appl. Math.*, M. Fornasier, Ed., 2010, vol. 9, pp. 1–92.

[223] R. Vershynin, “Introduction to the non-asymptotic analysis of random matrices,” in *Compressed sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012, available as arXiv:1011.3027v7.

[224] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics.* Wiley, 1988.

[225] A. Kolmogorov and Y. A. Rozanov, “On strong mixing conditions for stationary Gaussian processes,” *Theory of Probability and Its Applications*, vol. 5, no. 2, pp. 204–208, 1960.