Multidimensional random sampling for Fourier transform estimation

Mustafa Al-Ani

School of Electronics and Computer Science

Multidimensional Random Sampling for Fourier Transform Estimation

Mustafa Al-Ani

A thesis submitted in partial fulfilment of the requirements of the University of Westminster for the degree of Doctor of Philosophy

January 2013

“Things should be made as simple as possible, but not any simpler.”

Albert Einstein

ABSTRACT

This research considers the Fourier transform calculations of multidimensional signals. The calculations are based on random sampling, where the sampling points are nonuniformly distributed according to strategically selected probability functions, to provide new opportunities that are unavailable in the uniform sampling environment. The latter imposes the sampling density of at least the Nyquist density. Otherwise, alias frequencies occur in the processed bandwidth which can lead to irresolvable processing problems. Random sampling can mitigate Nyquist limit that classical uniform-sampling-based approaches endure, for the purpose of performing direct (with no prefiltering or downconverting) Fourier analysis of (high-frequency) signals with unknown spectrum support using low sampling density. Lowering the sampling density while achieving the same signal processing objective could be an efficient, if not essential, way of exploiting the system resources in terms of power, hardware complexity and the acquisition-processing time.

In this research we investigate and devise novel random sampling estimation schemes for multidimensional Fourier transform. The main focus of the investigation and development is on the aspect of the quality of estimated Fourier transform in terms of the sampling density. The former aspect is crucial as it serves towards the heart objective of random sampling of lowering the sampling density. This research was motivated by the applicability of the random-sampling-based approaches in determining the Fourier transform in multidimensional Nuclear Magnetic Resonance (NMR) spectroscopy to resolve the critical issue of its long experimental time.

ACKNOWLEDGEMENTS

I can no other answer make, but thanks, and thanks

- WILLIAM SHAKESPEARE

It is a great pleasure to thank all the people who helped and supported me through the last few years.

I would like to express my special gratitude to my supervisor, Dr. Andrzej Tarczynski, for giving me the opportunity and introducing me to the field of research, and patiently guiding me during this journey. Through the years, Andrzej has played an important role in my development as a researcher and a person, giving me guidance, support and freedom in pursuing my research directions. His guidance goes far beyond this thesis. Andrzej, in the future I want to be more like you.

I would like to thank Prof. Paulo J. S. G. Ferreira for the time and efforts he put in his thorough evaluation of my report. Thanks to Prof. Izzet Kale for his valuable support over the last few years. He genuinely made me feel like I have a family at the university.

I am indebted to my colleague and friend, Dr. Bashar Ahmad, for providing a stimulating and fun environment in which we learnt and grew. I am grateful for the nice PhD time spent together and the many helpful discussions we had. Bashar is one of the most committed and highly ambitious individual I have ever met. I hope he will achieve all the goals that we talked about.

Last but not least, I want to thank everyone of my family for their unconditional, continuous support. No matter what was going on in their life, they always put me first and made me worried about nothing but my studies. I do not think there are enough words to thank them.

AUTHOR DECLARATION

I declare that all the material contained in this thesis is my own work; and that to the best of my knowledge the thesis does not contain any material previously published or written by another person except where due reference is made in the text.

Mustafa Al-Ani

ASSOCIATED PUBLICATIONS

- **M. Al-Ani** and A. Tarczynski, "Evaluation of Fourier transform estimation schemes of multidimensional signals using random sampling," *Signal processing*, Elsevier, vol. 92, pp. 2484-2496, 2012.
- **M. Al-Ani** and A. Tarczynski, "Efficient multidimensional sampling scheme for Fourier transform estimation," *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '12)*, Seoul, May 2012. pp. 3210-3213.
- **M. Al-Ani**, A. Tarczynski and B. I. Ahmad, "A Novel Fourier transform estimation method using random sampling," *Proceedings of the 19th European Signal Processing Conference (EUSIPCO'11)*, Barcelona, Aug 2011, pp. 859-863.
- B. I. Ahmad, A. Tarczynski and **M. Al-Ani**, "A SARS multiband spectrum sensing method in wideband communication systems using RSG", *Proceedings of the 19th European Signal Processing Conference (EUSIPCO'11*), Barcelona, Aug 2011, pp. 1219-1223.
- **M. Al-Ani** and A. Tarczynski, "The effect of missing samples on the quality of the spectral analysis", *Proceedings of the 9th International Conference on Sampling Theory and Its Applications*, Singapore, May 2011.
- B. I. Ahmad, A. Tarczynski and **M. Al-Ani**, "A DASP multiband spectrum sensing method based on total random sampling on grid without replacement", *Proceedings of the 9th International Conference on Sampling Theory and Its Applications (SAMPTA '11)*, Singapore, May 2011.
- **M. Al-Ani** and A. Tarczynski, "Equidistributed sampling sequences for spectral analysis", *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, California, Nov 2011, pp. 1086-1090.

TABLE OF CONTENTS

GLOSSARY

Acronyms and Abbreviations

DASP	Digital Alias-free Signal Processing
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
DTFT	Discrete-Time Fourier Transform
FFT	Fast Fourier Transform
FID	Free Induction Decay
IID	Independent Identically Distributed
IQ	In-phase and Quadrature
NMR	Nuclear Magnetic Resonance
PDF	Probability Density Function

Special Notations, Operators and Functions

Scalar variables are denoted by lower case letters and vectors are denoted by bold lower case letters:

x	Scalar variable
$\mathbf{x}$	Vector quantity

Operators

$E[x]$	Statistical expectation of x
$\mathrm{var}[x]$	Variance of random variable x
$\mathrm{cov}[x,y]$	Covariance of random variables x and y
$P[x]$	The probability of the event x
$\Re[Z]$	Real part of complex variable Z
$\Im[Z]$	Imaginary part of complex variable Z
$\lvert Z\rvert$	Magnitude of complex variable Z
Z^*	Conjugate of complex variable Z
*	Convolution
$\det[X]$	Determinant of matrix X
$\mathrm{v}_T[x(t)]$	Total variation of function $x(t)$ with respect to t
$\mathcal{F}[x(t)]$	Fourier transform of function $x(t)$ with respect to t
$\mathrm{frac}[x]$	Fractional part of irrational number x
$\max[a,b]$	Maximum number of a and b

Special functions

Φ -function	$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{+\infty}^{x} e^{-\tau^2/2}d\tau$
Rectangular function	$\mathrm{rect}(x)=\begin{cases}1 & \text{for } \lvert x\rvert \le 0.5\\ 0 & \text{elsewhere}\end{cases}$
Sinc function	$\mathrm{sinc}(x)=\sin(\pi x)/\pi x$

Principal Symbols

N	Number of samples
K	Number of dimensions
D	Size of the multidimensional observation window
T_k	The length of the observation window along the k-th dimension
$\mathbf{f}$	The analysed frequency in the K-dimensional domain
f_k	The frequency along the dimension k
$\mathbf{f}_c$	The acceleration frequency in the K-dimensional domain
A	The observation window of the signal
A_i	The i-th subdomain
N_i	The number of samples inside the i-th subdomain
v_i	The weight of the i-th subdomain
$\mathbf{t}_m$	The m-th point on the grid
$\mathbf{t}_n$	The n-th sampling point
M	The total number of the grid points
M_i	The number of grid points in the i-th subdomain
G_q	The q-th interlaced subgrid
$x(\mathbf{t})$	The deterministic, real, finite-energy signal
$w(\mathbf{t})$	Windowing function
$x_w(\mathbf{t})$	The windowed signal, i.e. $x(\mathbf{t})w(\mathbf{t})$

CHAPTER ONE

INTRODUCTION

In signal processing, samples of the signal are usually designed to be captured on an equidistant grid of time instants. In multidimensional signal processing, the most common form of sampling is on a lattice which is the multidimensional equivalent of the equidistant grid in the one-dimensional case. This principle is known as *uniform sampling* and it underlies nearly all signal acquisition protocols used in consumer audio and visual electronics, medical imaging devices, radio receivers, and so on. The main advantage of uniformly sampling the data is the simplicity of the processing algorithms that have been developed in the past century. Nonetheless, uniform sampling (in its classical form) entails some restrictions in its design: the rate of sampling is governed by the Nyquist rate of twice the highest frequency present in the signal (this is for one-dimensional signals. An analogous condition is required in multidimensional domains, which we briefly survey in the next chapter). Violating the Nyquist condition usually results in false process of the signal's samples. In many applications, fulfilling the Nyquist limit is achievable, although for high-frequency signals downconversion and prefiltering are usually needed prior to sampling. However, unfortunately, this is not the case for some applications where maintaining these sampling densities is expensive or even impossible because of system limitations such as hardware complexity, power consumption and the data acquisition-processing time. Generally speaking, reducing the sampling density whilst achieving the aimed signal processing objective is an efficient way of exploiting the system's resources. To this end, a vast number of researchers in several of fields of engineering, mathematics and science have pursued alternatives for uniform sampling in order to lower the operated sampling densities while preserving the ability of achieving the signal processing goal. Indeed, with few

exceptions, it is not expected for the alternative solutions to be uniform nor the classical uniform-based processing approaches to be valid.

Reduction in sampling rates is beneficial in one-dimensional systems and particularly important when dealing with multidimensional signals where maintaining high sampling rates in each direction leads to a significant increase in the number of signal's samples (assuming a Cartesian lattice). These problems are found when dealing with two-dimensional images and three-dimensional video processing; however, they are most notably affecting processing data for Nuclear Magnetic Resonance (NMR) experiments where signals could easily be five, or more, dimensional [1-6]. The latter application has motivated our interest in the artful topic of multidimensional nonuniform sampling and the associated processing algorithms. We must remark that the work presented here is not particularly targeted for this application, and the advantages of this work can be explored in any other applicable scenario. Next, we provide a short survey about multidimensional NMR spectroscopy and its intrinsic bottleneck that is limiting its scope. The main aim of the following section is not to provide the principles of multidimensional NMR spectroscopy, but rather to demonstrate the nature of the NMR signals and highlight the problem from the DSP point of view.

1.1 NMR Spectroscopy

NMR spectroscopy allows the examination of the structure of individual molecules or portions of the molecules within a chemical sample. The NMR spectroscopy works by placing the sample in a strong magnetic field and irradiate it with electromagnetic pulse. Each radiating nucleus in the chemical sample responds with a detectable exponentially-decayed sinusoid wave. Different nuclei, or identical nuclei in a different chemical or structural environment inside the molecule, react differently to the magnetic field and the

pulse, radiating exponentially-decayed sinusoid wave with different parameters. The distinction between the several detected waves provides useful information about the structure of the molecule and its chemical properties. The radiated waves of the nuclei represent the NMR signal (usually called Free Induction Decay (FID)), which can be expressed as the summation of decaying sinusoids with different centre frequencies, amplitude and decaying factors. Thus, if we allow *R* sinusoids to have separate intensities $\{\gamma_r\}_{r=1}^{R}$, resonance frequencies $\{\omega_{0,r}\}_{r=1}^{R}$, and relaxation time constants $\{T_{2,r}\}_{r=1}^{R}$, the NMR signal $x(t)$ takes the form:

$$x(t)=\sum_{r=1}^{R}\gamma_r\cos(\omega_{0,r}t)\exp(-t/T_{2,r}). \tag{1.1}$$

Extracting the embedded parameters from these summed sinusoidal components is unfeasible to be conducted in the time domain. Instead, the signals are transformed to the frequency domain to attain a spectral representation of the signal, where the sinusoids appear as peaks mounted at their centre frequencies, with the highest magnitudes related to their amplitudes, and widths (at half the highest magnitude of the peaks) determined by the decaying coefficients. Accordingly, all the information can be easily interpreted by the user. See Fig. 1.1 for an illustrating example of an NMR signal in the time domain and its frequency representation.

In the case of complex molecules, the spectrum shows an extensive overlap of the peaks because of the large number (hundreds) of radiating nuclei; see Fig. 1.2 for an illustration. In such a complex case, it is very difficult to identify and separate the peaks that are close to each other. A solution was proposed by Jeener [7] introducing the two-dimensional NMR spectroscopy where an additional spectral dimension is inserted to spread the peaks over a two-dimensional representation.

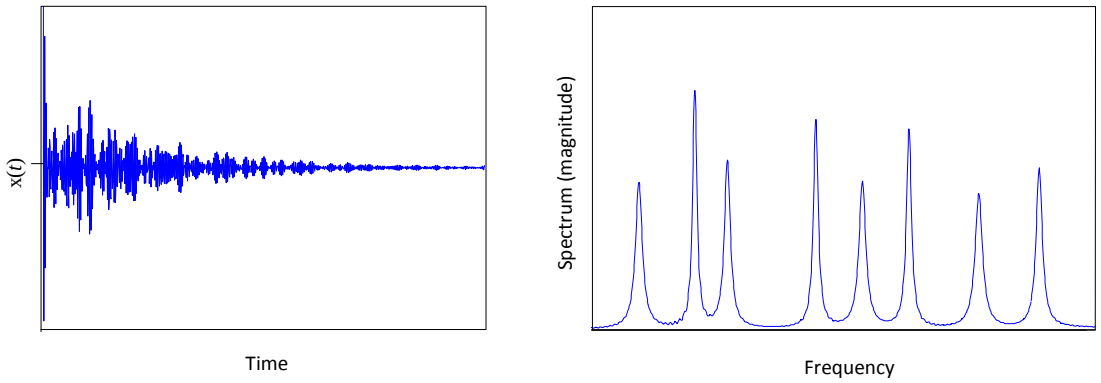

Fig. 1.1. An example of an NMR signal in the time domain (left plot) and frequency domain (right plot).

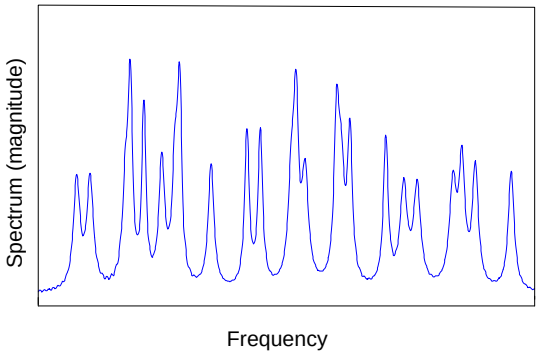

Fig. 1.2. Spectrum of an NMR signal of a complex molecule which shows an overlap of the peaks.

In two dimensional NMR, the signal is recorded as a function of two time variables, t_1 and t_2, and the resulting data is subjected to a two-dimensional Fourier transform to yield a spectrum that is a function of two frequency variables. In the first period, the tested object is excited by an electromagnetic pulse. The resulting magnetization is allowed to *evolve* for the first time period, t_1. Then, another period follows which consists of a further pulse(s). After the second period of pulses, the signal is recorded as a function of the second time variable, t_2. An illustration of a pulse sequence of a two-dimensional experiment is shown in Fig. 1.3.

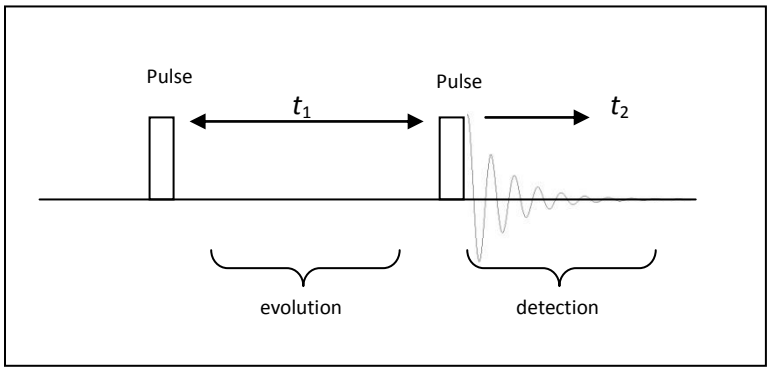

Fig. 1.3. Pulse sequence and recording of an FID of a two-dimensional experiment.

The two dimensional signal is recorded in the following way. First, t_1 is set to zero, the pulse sequence is executed and the resulting radiated signal from the tested chemical object is recorded. Then, the system is allowed to return to equilibrium. After that, t_1 is set to T' and then the sequence is repeated and a signal is recorded and stored separately from the first. Again, the system is allowed to equilibrate, t_1 is set to 2T', the pulse sequence is repeated, and a new signal is recorded and stored. The whole process is repeated again for t_1 = 3T', 4T' and so on until sufficient data is recorded, typically 50 to 500 increments of t_1 (it is usually referred to t_1 and t_2 by the indirect and direct dimension, respectively). Hence, we have a two dimensional data that is discrete along the indirect dimension with distance of T'. By subjecting the recorded signals to a two-dimensional Fourier transform, we can obtain a two-dimensional spectral representation with peaks spread over two dimensions. This concept was straightforwardly generalised later to multiple dimensions by creating multiple indirect dimensions, introducing the multidimensional NMR spectroscopy.

Multidimensional NMR technique provides a unique source of information about biomolecular structure, interactions, and dynamics. But the bottleneck that tends to restrict its scope is the unfavourable experimental time. As we described above the multidimensional experiment is performed by acquiring a series of signals. Hence, for *K*-

dimensional experiment with N_k points in the *k*-th dimension, the total experiment time T_{total} increases exponentially with the number of dimensions, and equals

$$T_{total} = N_k^{K-1} 2^{K-1} T_{rp} \tag{1.2}$$

where T_{rp} is the repetition time that is required to record the signal (as we described above) and for the system to be back to its initial situation, and 2^{K-1} represents the number of data sets required by quadrature detection rules. Thus, for example, to conduct a three-dimensional experiment with 60 samples in each indirect dimension, and 1 second repetition time, the acquisition time will be 4 hour long. And, increasing the experiment dimensionality (with the same number of samples along the added dimension) will increase the experiment time by 120 times.

In addition, the time factor tends to restrict the scale of multidimensional spectroscopy to rule out the study of time-dependent phenomena such as chemical exchange, or the investigation of materials of limited stability [1]. Furthermore, the amount of data we need to store also increases with the dimensionality of the experiments. For example, if a three-dimensional experiment has a size of data we need to store of 4 gigabytes, the four-dimensional experiment with same resolution could have the size of data of 4 tetrabytes.

1.2 The Adopted Sampling Methodology

In the previous section we discussed how the acquisition time is directly proportional to the number of data samples (i.e. FIDs) in the indirect dimensions. An intuitive way to accelerate the experimental time according to (1.2), and reduce the size of data we need to store, is to reduce the number of the collected FIDs. However, conventional multidimensional NMR spectroscopy (which engages classical DSP) requires the distances

between successive samples in each indirect dimension to be fixed. This distance is determined by the Nyquist theorem of uniform-sampling-based DSP. Moreover, the total monitored time along the indirect dimensions is determined by the desired resolution of the constructed spectrum. Although the resolution can be scarified sometimes in experiments with more than three dimensions and, accordingly, the number of FID signals can be decreased, the experiment time can still be too long.

Lowering the density of the collected FID signals below Nyquist leads to the overlap of the signal spectrum aliases that appear as a consequence of the periodicity of sampling points. This is undesirable as it will not be possible to differentiate between the original signal spectrum and its replicas; moreover, the signal spectrum can be distorted as a result of the overlap.

An effective solution to tackle such an issue can come from *random sampling estimation* as methods for *digital alias-free signal processing* (DASP). These methods employ random sampling to mitigate spectral aliasing that is inherent in sub-Nyquist uniform sampling. Accordingly, the signal spectrum can be unambiguously determined whilst using low sampling densities, i.e. reduced number of FIDs. Random sampling simply refers to the sampling process where the sampling points are randomly distributed via strategically chosen probability density functions. NMR spectroscopy is probably the most advanced, suitable technology so far with regards to implementing random sampling. One reason for this is that choosing the location of the sampling points in the indirect dimensions is relatively flexible. Therefore, there is no need for any hardware modification in order to perform the irregular data acquisition.

The recent work on random sampling estimation of Fourier transform [8]-[11] and its application in multidimensional NMR Spectroscopy [2]-[6] prompted our interest in the topic. In [8]-[11] forms of random sampling estimators of the Fourier transform were

introduced and studied for *one dimensional* signals. In [8] and [9] classes of *total random estimators* were considered. The authors showed that the rate of mean-square convergence of the estimates is $1/N$, where N is the number of random samples. A class of *stratified estimation* was considered in [10]. It was demonstrated that the asymptotic convergence rate of the estimates in the mean-square sense is $1/N^3$. This result holds when the signal has a continuous first derivative. In [11], a form of *antithetical stratified estimation* was studied. It was proven that for signals with continuous second order derivatives, the asymptotic mean-square convergence rate of the antithetical stratified estimates is $1/N^5$. Some of these random sampling estimators of the Fourier transform were adapted in multidimensional domains and their performances were evaluated based on NMR experiments, in [2]-[6], with no deep analytical analysis. This triggered us to extend the aforementioned random sampling estimation schemes to multidimensional signals and thoroughly analyse their estimation errors and their rates of decay as a function of the number of samples N and the number of the dimensionality of the signal K. We show that the rate at which the mean-square error of total random estimates decays is independent of the dimensionality of the signal K and equals to $1/N$. We prove that for signals with the continuous derivative along each dimension, the asymptotic mean-square convergence rate of the stratified estimates in K-dimensions can reach $1/N^{1+2/K}$. For signals with the continuous second derivative along each dimension, we show that the antithetical stratified estimates can asymptotically converge at the rate of $1/N^{1+4/K}$. We also establish the asymptotic normality of the estimates which can be used to investigate the probability of their events and find the confidence intervals. Although [8]-[11] prompted us to study and evaluate the performances of these schemes in multidimensional domains, in this thesis we study wider strategies of distributing the sampling points and the associated unbiased estimators under these schemes to provide a deeper evaluation of their efficiency.

In practice the sampling points of the random sampling are always lying on, more or less fine, grid as a result of the rounding of their coordinates caused by the finite step of the system's clock and the limitations of designing capturing devices capable of collecting samples arbitrarily close to each other along the real time dimension. For that reason, we establish theoretical analysis for on-grid random sampling estimates of the Fourier transform of the three aforementioned schemes. We determine the statistical properties of these estimates and provide thorough analysis of their performances.

Then we move to investigate the effect of the signal spectral position on the accuracy of the estimates. This issue is particularly important as it serves towards the key objective of random sampling in signal processing, i.e. using low sampling densities compared with the corresponding Nyquist density of the classical uniform-sampling-based approaches that is governed by the signal position in the frequency domain. The convergence rate of the total random estimates is achieved uniformly for any number of samples across all frequencies and the accuracy of the estimates is independent of the position of the signal spectrum in the frequency domain. Whereas, the quality of the (antithetical) stratified sampling estimates is affected to a certain extent by the signal's spectral position. Their asymptotic fast convergence rates, which hold after a sufficient number of samples, do not appear uniformly across all frequencies We established some limits that demonstrate the fact that the requisite number of samples that is needed for the fast convergence rate to appear depends on the position of the signal spectrum from DC. Hence, the fast convergence takes place at DC and its close neighbourhood for small number of samples, and it spreads to higher frequencies only when the average sampling densities are significantly increased (hence, DC represents the *acceleration frequency*). Thus, to benefit from the fast convergence rates, more samples are required to be collected for processing high-frequency signals than for those placed closer to DC even if their spectral shapes are identical. In fact, the necessary number of samples for signals at high frequencies is comparable to what is required by Nyquist

(otherwise, the user would simply miss out on the advantages of these fast convergence rates). This observation defeats one of the popular objectives of using random sampling. Accordingly, these approaches are most advantageous for signals whose spectrum is concentrated in low frequency ranges.

In this thesis, we develop a novel technique of (antithetical) stratified sampling for Fourier transform estimation which allows moving the acceleration frequency, from zero, to practically any selected point in the frequency domain. Consequently, fast convergence to the targeted Fourier transform can be obtained using small number of samples regardless of the positioning of the signal spectrum in the frequency domain. This is particularly important to process signals which are known to take positions in the region of some high frequencies, where sub-Nyquist random sampling technique is most useful to employ in the first place. This can provide an alternative solution to the classical way of downconverting and filtering the signal to relocate and reshape its spectrum prior to sampling (uniformly or randomly). In addition, this is essential when we only have access to the samples (that we would like to minimise their number) of the signal and hence filtering the signal to eliminate the high frequency replicas resulting from downconversion is not an option. Additionally, the developed methods leverage another principle to provide further reduction in the estimation error (compared with the regular random sampling approaches) that is independent of the error reduction rendered by facilitating the fast convergence rates, as we will see in Chapter 5. In fact, this principle can be even exploited in total random sampling (where no acceleration occurs). We refer to the proposed technique as the "*IQ estimation"*; because of the fact that it estimates the Fourier transform of the signal through its In-phase and Quadrature components, or their equivalents in multidimensional domains.

The theory of random sampling has a relatively long history. The first substantial work in the field can be tracked to 1960 which was followed by considerable amount of research,

including two books [12,13] that have been written in the relatively recent term. Random-sampling-based approaches are significantly important to study not only for their interesting theoretical aspects but also for the wide available implementations of their techniques. Examples range from simple sampling voltmeters and oscilloscopes from 1960s to complex NMR spectroscopy recently. In the next chapter, we demonstrate where the adopted random sampling techniques stand among the others in their family.

Indeed, random sampling techniques are not the only methodology that seeks to reduce the employed sampling densities beyond Nyquist. These techniques share their goals with the wide theory of nonuniform sampling, which is as old as uniform sampling and Nyquist theorem itself. We must remark that it is difficult to build connections between random sampling techniques and other sub-Nyquist strategies as they considerably differ in their ideas and objectives. Nonetheless, in the next chapter we provide a short survey about the most popular sub-Nyquist strategies to help highlighting the main features of the adopted methodology.

1.3 Summary of the Contributions

The original contributions of this research can be summarised by:

Chapter 3

- We extend the three most efficient reported random sampling estimators of the Fourier transform from one dimensional to multidimensional domains. These schemes are total random estimation, stratified estimation and antithetical stratified estimation.
- We thoroughly analyse their performances in terms of the of the mean-square error they introduce to the estimated Fourier transform and their rates of convergence.
- We investigate the dependence of the accuracy of the aforementioned sampling estimates on the distribution of the signal's energy in the frequency domain. We determine that the number of samples required to analyse the signal with fast convergence in the (antithetical) stratified schemes depends on the distance of both of the analysed frequency and the frequencies that carry most of the energy of the signal from the acceleration frequency, i.e. the zero point.
- We establish the joint asymptotic normality of the real and imaginary parts of the aforementioned Fourier transform estimates, which can be used to investigate the probability of events of the constructed random estimates.

Chapter 4

- We provide theoretical analysis for more practical, on-grid versions of the aforementioned Fourier transform estimators, where the sampling points can only take place on a grid of points.

Chapter 5

- We develop a novel Fourier transform estimation method, i.e. IQ estimation, that can locate the acceleration frequency at practically any selected point in the frequency domain, rather than zero point. Choosing this frequency in the locality of the signal spectrum can lead to a substantial reduction in the estimation error or savings on the number of samples, subject to the system requirements.

1.4 Thesis Outline

This thesis is organised as follows. In Chapter 2, we start with providing preliminaries on sampling and Fourier analysis in multidimensional domains, introducing some of the notations that will be adopted throughout the thesis. Then, we provide a concise survey of the most popular sub-Nyquist strategies, focusing on the methods that target the Fourier transform. In Chapter 3, we generalise the total random, stratified and antithetical stratified estimation schemes, and examine their statistical characteristics. The detailed derivations of the developed analytical expressions are included in Appendix A and B. In Chapter 4, we analyse the on-grid random sampling estimation schemes. Chapter 5 focuses on devising the IQ method and study its statistical properties with the derivations in Appendix C and D. Chapter 6 concludes the thesis and includes ideas about potential research based on the presented work.

CHAPTER TWO

SAMPLING TECHNIQUES AND FOURIER ANALYSIS

The Fourier transforms is a powerful mathematical tool which can enable us to view signals or data in a different domain, inside which several difficult (linear) problems become very simple to analyze. Fourier analysis finds applications in many diverse fields. In economics, astronomy, and several other fields, the Fourier analysis may reveal "hidden" periodicities in the collected data, which are to be associated with cyclic activities or recurring processes, and that can help to understand and predict future behaviour of the data. In speech analysis, Fourier analysis of voice signals are useful in better understanding the speech production process and can be used for both speech recognition and compression. In medicine, Fourier analysis of various signals measured from a patient can provide useful material for diagnosis, such as Magnetic Resonance Imaging (MRI). A shifted (delayed) signal in the time domain manifests as a phase change in the frequency domain. This elementary property is widely used in medical applications, especially in imaging and tomography applications. As we described in the previous chapter, the Fourier transform is also used in Nuclear Magnetic Resonance (NMR) and in other kinds of spectroscopy. Fourier analysis appears in other applications such as control systems, seismology, radar systems, vibration monitoring, etc. Its ubiquity in nearly every field of engineering and physical sciences, all for different reasons, makes it impossible to list all its applications. Despite the fact that other theoretical frequency spectrum techniques are used in handling some of these applications, Fourier transform methods are virtually indispensable in all of them.

A fundamental advance in the practicality of Fourier analysis occur owing to the theory of sampling, which made it possible to calculate the Fourier transform of signals using fast digital computers instead of analogue filter banks and spectrum analyzers. The most common form of sampling is the uniform sampling (also known as regular sampling). Its popularity is due to the practicality of designing capturing devices that collect data in a fixed rate. And, the development of the efficient, easy to implement Fast Fourier Transform (FFT) algorithms that are capable of performing the calculations in a considerably short time, gave the Fourier transform of uniform data a different dimension.

Fourier analyses appear in various branches of signal processing. However, they bear on a wide aspect with the broad topic of spectral analysis. Typically, spectral analysis of a signal (or data) tells us what frequencies are present in the signal and in what proportions. Generally speaking, spectral analysis includes and focuses more on estimating the power spectrum density of signals (which can be seen as the square magnitude of the Fourier transform) that tells us how much power the signal has at a particular frequency. Spectral analyses approaches can deal with signals that do not possess a direct Fourier transform representation. This is distinct from the problem of the *Fourier transform of deterministic finite-energy real signals* that yields complex-valued spectrum, which is the focus of this thesis. In some applications, the magnitude of the Fourier transform (equivalently, the power spectral density) is adequate to fulfil the task, however for some other applications the phase is as important as or even more vital than the magnitude. There are two types of the Fourier transform based estimates of the power spectral density: the *periodogram* and the *correlogram*. The periodogram in its classical form is a solution to a least square data fitting to complex sinusoids, which results into Fourier transform type of calculations, that can be performed using FFT. Variations and properties of the periodogram have been devised and thoroughly studied in literature; see for example [14] for an introduction. The other type uses the correlation of the observed data via Fourier transform calculations to yield an estimate of

the power spectral density. These techniques are categorised under the broad category of *nonparametric approaches* of spectral estimation. Despite the fact these approaches can suffer from poor resolution and leakage effects, as well as large variance, they are still the most common spectral methods [14]. We note that, although these approaches are based on Fourier transform, the results we present in this thesis are not directly applicable to them.

The second approach to spectral estimation, called the *parametric approach*, is to postulate a model for the data, which provides a means of parameterising the spectrum, and to thereby reduce the spectral estimation problem to that of estimating the parameters in the assumed model. Parametric methods offer more accurate spectral estimates than the nonparametric ones in the cases where the data indeed satisfy the model assumed. However, the nonparametric methods can outperform the parametric ones because of the sensitivity of the latter to model misspecifications [14]. This observation has motivated and renewed interest in the nonparametric approach to spectral estimation. The reader interested in parametric methods can refer to [14] for an introduction.

In many applications, such as engineering, physics, biomedicine, economics, seismology, and, particularly, astronomy the available data is discrete and of irregular nature. Hence, from a different angle from earlier, the discrete nature of the data, considerably complicate the problem of calculating the spectrum of the signal. The usage and the accuracy of the general methods for arbitrary sampling have been reviewed. For example, the periodogram can be readily extended to the irregularly sampled data, see [15] for a to-the-point discussion. Similarly, parametric approaches can cope somehow with the irregular data, a critical review of some of those methods can be found in [16,17].

Several methods have been specially developed in an attempt to relax and eliminate the consequences of the irregularity of samples. For example, modified forms of the periodogram that can deal with the uneven data were introduced, which we will discuss later

because of their popularity in the field. Some of the proposed methods are rather specialised on a particular irregularity of the sampling instants. The most commonly considered schemes correspond to the following situations where the data is available at points of a regular time grid, but not at all points in a given time interval; if the unavailable data samples form large gaps that are interlaced with large available data subsets, then we have a so-called *gapped data* case; on the other hand, if the unavailable samples form a more or less arbitrary pattern, then we simply say that we have a *missing data* case. These two cases have been considered in many papers, for example [18,19], where a nonparametric spectral analysis method for sinusoids in-noise data has been proposed, and in [20] which discussed a parametric approach for data with continuous spectra; many other methods for gapped or missing data can be found in the references of the cited works.

In other application, such as NMR spectroscopy, where the user has the freedom of choosing the location of the sampling points to extract as much information as possible from the signal while using low number of samples, one rarely chooses the uniform strategy. Such problems are extremely common for a variety of reasons. For instance, the number of sensors may be limited. Or the measurements may be extremely expensive as in certain imaging processes via neutron scattering. Or the sensing process may be slow so that one can only measure the object a few times as in multidimensional NMR spectroscopy. And so on. Various strategies of collecting the data and the associated processing algorithms have been proposed in the research literature. To give a quick idea about deliberately nonuniform sampling we mention at this junction that the data can be *randomly* collected with a certain mode and distribution pattern (which is the category we are interested in). Or, it can be collected with a deterministic design of *periodic nonuniform sampling*. Or, there are other schemes that combine aspects of both. More details about these schemes are included in this chapter.

The literature on these topics is diverse, and still considerably expanding because of its theoretical importance and wide applicability in various field of science. In addition, since our target is sampling the signal and process them to extract as much information from the signal, some of the signal sampling-reconstruction strategies share the same goal, and can be easily "converted" to yield the Fourier transform of the signal. There are several superb review papers and books that survey and categorise the various results in the field, e.g. [21-32]. While reviewing the extensive literature on the existing methods for nonuniformly sampled data is beyond the scope of this thesis, we briefly describe and categorise the approaches that have received a lot attention in the field. We highlight their various objectives and paces of addressing the problem. We will see by the end of this chapter the diversity of these methods, which makes them all impossible to compare or build solid bridges in between.

Next, we start with introductory material about multidimensional Fourier transform calculations. Then, we describe some famous families of nonuniform sampling techniques. Indeed, we focus on the category of interest, i.e. random sampling estimation, reviewing the most important results in this field and demonstrating where our contribution fits in the literature. Then, we finish the chapter with a summary.

2.1 Preliminary on Multidimensional Sampling and Fourier Analysis

It is certainly not feasible to provide all the mathematical analysis behind Fourier analysis in here. However, we give a brief review of the representation and notations of multidimensional signals and their Fourier transform, which we will carry out in this thesis. The issues highlighted here can be interpreted in different ways; we choose the way that can provide the best insight from the DSP point of view. The *K*-dimensional Fourier transform

$X(\mathbf{f}) \equiv X(f_1, f_2, ..., f_K)$ of a *K*-dimensional real, deterministic signal $x(\mathbf{t}) \equiv x(t_1, t_2, ..., t_K)$ with finite energy is defined as

$$X(\mathbf{f}) := \int_{\mathbf{t}} x(\mathbf{t}) \exp(-j2\pi \mathbf{f}.\mathbf{t})\, d\mathbf{t} \tag{2.1}$$

where

$$\int_{\mathbf{t}} d\mathbf{t} \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} dt_1 dt_2 ... dt_K$$

and

$$\exp(-j2\pi \mathbf{f}.\mathbf{t}) = \exp(-j2\pi f_1 t_1) \exp(-j2\pi f_2 t_2) \; .... \; \exp(-j2\pi f_K t_K).$$

Hence to find the Fourier transform, the expression mathematically indicates that the signal must be calculated continuously for an infinite-size domain of $\mathbf{t}$, which is not feasible indeed in any (DSP) system. According to Petersen and Middleton theorem [33], which is the generalization of Nyquist theorem to multiple dimensions, the Fourier transform of the signal can be determined using only samples of the signal. At first, loss of information may seem infinite. In fact, simple analysis of the sampling process in the frequency domain shows that the complete information content is conserved provided that the sampling is adapted to the component frequencies of the signal, as we show next.

The sampling operation is performed by multiplying the signal by a multidimensional lattice of deltas. In one dimensional domain, this operation would be equivalent to multiplying the signal by a train of deltas, where the period between successive deltas defines the *sampling rate*. In multiple dimensions, characterising the sampling process is not as straightforward as in one dimensional signals. Here, we have to deal with geometry forms rather than simple intervals, as the multidimensional lattice of deltas is not necessarily

Cartesian. The geometry and the density of the sampling points is defined by the *sampling matrix* which is composed of vectors of size *K*

$$Q := [\mathbf{q}_1 \,|\, \mathbf{q}_2 \,|\, ... \,|\, \mathbf{q}_K]. \tag{2.2}$$

In Fig. 2.1 we show examples of two-dimensional sampling lattices/matrices. The geometry of these sampling points is important as we will see below. Now, we can define the sampled signal by

$$x_d(\mathbf{t}) := x(\mathbf{t}) \sum_{\mathbf{m}} \delta(\mathbf{t} - Q\mathbf{m}) \tag{2.3}$$

where

$$\sum_{\mathbf{m}} \equiv \sum_{m_1=}^{\infty} \sum_{m_2=0}^{\infty} ... \sum_{m_3=0}^{\infty}$$

and

$$\delta(\mathbf{t}) = \delta(t_1)\delta(t_2)...\delta(t_K) \tag{2.4}$$

is the *K*-dimensional delta function. We note the Fourier transform pair

$$\sum_{\mathbf{m}} \delta(\mathbf{t} - Q\mathbf{m}) \leftrightarrow \left|\det[P]\right| \sum_{\mathbf{r}} \delta(\mathbf{f} - P\mathbf{r}) \tag{2.5}$$

where $\det[.]$ denotes the determinant operator, and $P = Q^{-1}$ is the *periodicity matrix* which determines the locations of the deltas of the right hand side of (2.5) in the frequency domain.

Multiplication in the $\mathbf{t}$ domain corresponds to convolution in the $\mathbf{f}$ domain. Accordingly, the transform of (2.3) is

$$X_d(\mathbf{f}) = X(\mathbf{f}) * \left|\det[P]\right| \sum_{\mathbf{r}} \delta(\mathbf{f} - P\mathbf{r}). \tag{2.6}$$

Since convolving a function with $\delta(\mathbf{t} - \mathbf{a})$ generates the same exact function but locates it at $\mathbf{a}$:

$$X(\mathbf{f}) * \delta(\mathbf{f} - \mathbf{a}) = X(\mathbf{f} - \mathbf{a}), \tag{2.7}$$

we conclude that

$$X_d(\mathbf{f}) = \left|\det[P]\right| \sum_{\mathbf{r}} X(\mathbf{f} - P\mathbf{r}). \tag{2.8}$$

By examining the expression above we notice that the Fourier transform of the sampled form of the signal comprises (scaled) replicas of the Fourier transform (2.1) of the original signal that are centred at $\{P\mathbf{m}\}$ (see Fig. 2.1 for some examples). Intuitively, the original signal spectrum can be retrieved if the replicas do not overlap. To this end, the Fourier transform must be zero outside a *K*-dimensional sphere (or any other shape) of finite radius, i.e. *Bandlimited.* Second, the sampling matrix (we defines the density and the geometry of the sampling points) should be properly structured so these spheres are adjacent. If these two conditions are fulfilled the original spectrum can be reconstructed by isolating (filtering) the original spectrum from its aliases. The above calculation of the Fourier transform of a sampled signal is formally known as Discrete-Time Fourier transform (DTFT) in one dimensional domains and also used in the multidimensional case, and it is simply expressed as follows:

$$X_d(\mathbf{f}) = \sum_{\mathbf{m}} x(Q\mathbf{m}) \exp(-j2\pi \mathbf{f}.Q\mathbf{m}). \tag{2.10}$$

The expression above is of high importance; it reduces the continuum infinity of the signal required to obtain its Fourier transform in (2.1), to a countable set. However, the set is still infinitely large and this is not feasible for the two simple facts that no signal can be observed for an infinite-size domain of **t** and no system is capable to handle such data. Alternatively, the Fourier transforms calculation is conveniently performed over a finite-size observation domain A, and (2.10) is usually replaced by

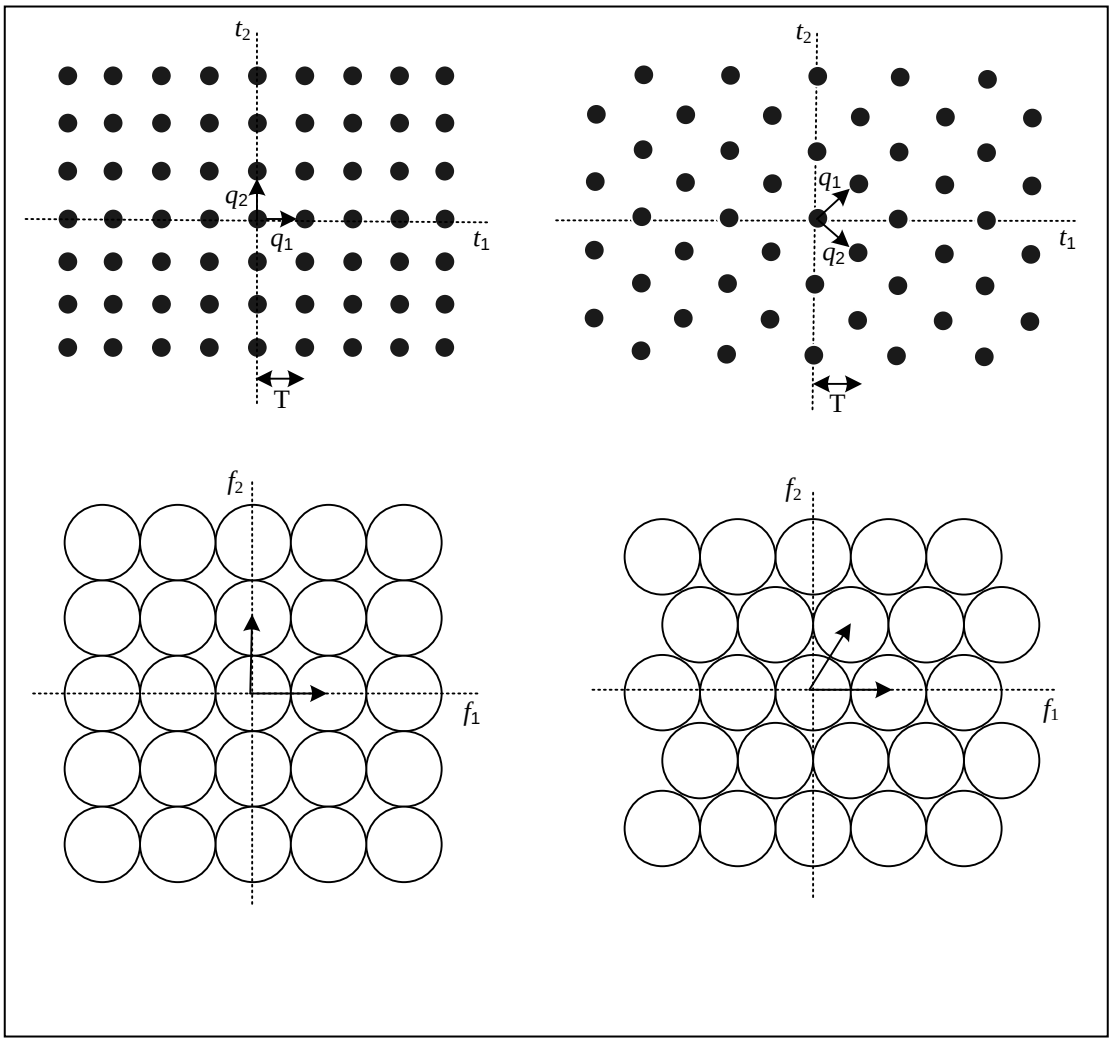

Fig. 2.1. Examples of sampling lattices (on top) and signal spectrum aliases (on the bottom) . On the left, we see the (Cartesian) sampling lattice on top with sampling matrix $Q_1 = \begin{bmatrix} T & 0 \\ 0 & T \end{bmatrix}$, and its results in the frequency domain on the signal spectrum of a circular shape in the bottom. On the right, we see the (Hexagonal) sampling lattice on top with sampling matrix $Q_2 = \begin{bmatrix} T & -T \\ T/\sqrt{3} & T/\sqrt{3} \end{bmatrix}$, and its consequences on the signal spectrum in the bottom.

$$X_D(\mathbf{f}) := \sum_{\mathbf{m}, Q\mathbf{m} \in A} x(Q\mathbf{m}) \exp\left(-j2\pi \mathbf{f}.Q\mathbf{m}\right). \tag{2.11}$$

Simply speaking, the operation of limiting the set of calculated samples can be seen as multiplying the signal by a multidimensional rectangular window and then calculate the Fourier transform of the outcome. This is equivalent to convolving the signal spectrum with a multidimensional sinc function (the Fourier transform of a multidimensional rectangular function). Accordingly, the generated signal spectrum is distorted (with an extent that depends on the size of the observation window) as the result of the convolution with more or less wide main lob of the sinc function as well as its secondary lobes (instead of convolving with delta functions when infinitely large set of samples are collected). This operation limits the resolution of the calculated spectrum and might lead to extra lobes present in the spectrum, especially on the sides of the original signal spectrum. The counter back to this phenomenon is to use a windowing function of non-constant value. There is a vast range of choices of windows that can be used to offer a trade-off between the resolution and the impact of the side lobes. An excellent review of the topic that includes analysis of their properties and smoothing effects can be found in [34, 35].

We conclude that the Fourier transform of signal is usually *approximated* with a sum using a finite number of the signal samples. To complete this section, we remark that in practice nearly all software and electronic devices that generate frequency spectra apply a fast Fourier transform (FFT) algorithms which calculate the Fourier transform on a discrete frequency representation, which corresponds to equally spaced samples of its DTFT. Because of the reversibility (the discrete signal, rather than the continuous signal, can be regenerated from its discrete points of Fourier transform), the generated Discrete Fourier Transform (DFT) is called a frequency representation of the discrete signal.

2.2 Nonuniform Sampling Techniques and Processing Algorithms

In this section we briefly describe some of the most popular alternative paradigms of collecting the samples and their corresponding processing algorithms. We demonstrate these techniques for the one dimensional case for simplicity and refer to references for a generalisation to the multidimensional case, if they exist.

2.2.1 Least-Square Spectral Analysis

Until recent, this category had a major attention in the community of irregular sampling theory. Its merit is in the fact that it does not put assumption on the signal (however, these methods work most satisfactorily when the data is a combination of sinusoids). Typically, it is a method of estimating a frequency spectrum, based on a least squares fit of sinusoids to the data samples. The widely used form of calculations is the one suggested by Scargle [36,37] and Lomb [38] (which can be readily extended to multiple dimensions) that yield the spectral power at the analysed frequencies, and it is known as the Lomb periodogram or the Lomb–Scargle periodogram (although it was shown that straightforward solution to the fitting algorithm is preferable form the computational standpoint). In [15] the reader can find a straightforward explanation of why the least-square fitting of sinusoids is preferable to the least-square fitting (to complex sinusoids) of the classical Fourier periodogram that is developed initially for equally spaced data. We must remark that their fitting procedure is for each frequency at a time (in contrast to some parametric approaches which assume the number of sinusoids and require a nonlinear least-square fitting algorithms to obtain their frequencies and amplitudes).

Their research was motivated by applications where nonuniform sampling is enforced by the experimental circumstances, and freedom in selecting the sampling instants is limited. In

other words, there is no assumption made on the sampling scheme, thereby allowing for arbitrary sampling points. Therefore, it is not possible to tell a lot about the accuracy of spectrum estimation. Indeed the random sampling schemes developed in this thesis can be processed with this means, but there is no guarantee of them serving the purposes that are developed for. Similar to what we described earlier about truncating the observation window in Fourier calculations, this category can suffer from the low resolution and leakage. Despite its inferior performance, the Lomb periodogram is still one of the most commonly used spectral analysis method in applications involving arbitrary sampled data. Recently, a new enhanced method that engages an iterative weighted least-square fitting was developed in [15], extended to multiple dimensions in [39], to outperform the Lomb periodogram, eliminating almost completely its leakage problems.

2.2.2 Compressive Sensing

In the last couple of years a surge of research appeared in the theory of *compressive sensing* (also known as compressed sampling) after the seminal work in [40]. The former is a relatively new data acquisition-processing strategy that exploits the sparsity or compressibility of a signal. Some signal is defined to be sparse if its representation in some orthogonal basis contains only a few nonzero coefficients and is compressible if it can be well approximated by a sparse signal. Mathematically speaking, we have a discrete signal vector $\mathbf{x}$ (or M samples of a continuous signal) which we expand in an orthogonal basis, such as Fourier basis, as follows

$$\mathbf{x} = \mathbf{X}\boldsymbol{\psi} \tag{2.12}$$

where $\boldsymbol{\psi}$ is the representation $M \times M$ basis matrix (e.g. Fourier exponents) and $\mathbf{X}$ is a vector of size M that has $W << M$ nonzero entries. (A signal is sparse in the frequency

representation if its DFT has few nonzero coefficients.) If we are given the values of $\mathbf{x}$ one can obviously construct the coefficients using simple linear algebra.

Indeed, we are interested in the case of observing a subset of the data by acquiring $N << M$ inner product measurements

$$y_k = \langle \varphi_k, x \rangle, \quad k \in P, \tag{2.13}$$

where $P \subset \{1, 2, ..., M\}$ is a subset of cardinality $N << M$, and φ_k is the sensing waveform. The measurement vector can be written in matrix notation,

$$\mathbf{y} = \boldsymbol{\varphi}\mathbf{x} = \boldsymbol{\varphi}\boldsymbol{\psi}\mathbf{X}. \tag{2.14}$$

(If the sensing waveforms are Dirac delta functions for example, then $\mathbf{y}$ is a vector of sampled values of $\mathbf{x}$.) With this information, the reconstruction is achieved by l_1 norm minimization; the proposed reconstruction $\hat{\mathbf{X}}$ is the solution to the convex optimization program

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{X}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \boldsymbol{\varphi}\boldsymbol{\psi}\mathbf{X}. \tag{2.15}$$

That is, among all objects consistent with the data, we pick that whose coefficient sequence has minimal l_1 norm. A remarkable aspect of compressive sensing is that it asserts $\mathbf{x}$ can be accurately reconstructed from the measurements $\mathbf{y}$ even when (2.14) is (perhaps exceedingly) underdetermined, provided $\mathbf{x}$ is sparse enough and the matrix $\boldsymbol{\varphi}\boldsymbol{\psi}$ satisfies certain conditions. These conditions can easily be fulfilled when the entries of the sensing matrix are random numbers (Gaussian or Bernoulli) [40], which are not appealing from the computational and practical point of view. Various works from the mathematical and engineering society develop algorithms in its general context: designing the basis matrix and

the sensing matrix, the optimization criteria to solve the underdetermined system (e.g. (2.15)), and the algorithms to obtain the sparse solution. However, the case of interest (from our prospective) is when compressive sensing is applied to uniform sampling of continuous-time signals, i.e. using a sensing matrix of delta waveforms. In this case, the spectrum is discretised on a fixed frequency grid, then spectral values are estimated by locating a few nonzero values in the corresponding vector, i.e. by reconstructing a sparse vector. For this case, especially with the presence of noise, the conditions that guarantee the unique solutions are not easily satisfied, unless the vector is very sparse [41], and many of the developed techniques in the compressive sensing literature become irrelevant. There are several papers that tackle this exact problem, e.g. [42]. In general, such an approach has become a good alternative to parametric methods to achieve high resolution as their processing algorithm put less assumption on the signal. In fact, the use of such linear model gives more robustness to the estimation compared to parametric methods, especially regarding sampling artifacts [42]. On the other hand, it results in a considerable increase of the number of unknowns, and computationally demanding algorithms. In signal processing, compressed sensing is referred to the process of acquiring and reconstructing a signal that is supposed to be sparse or compressible. However, it is a technique for finding sparse solutions to underdetermined linear systems. And, other sub-Nyquist technique leverages compressive sensing through a different strategy, e.g. [43] and [44] that is reviewed below.

2.2.3 Periodic Nonuniform Sampling

The occurrence of the signal spectrum aliases is the result of the periodicity of the sampling process. As we mentioned before, the density and the geometry of the sampling points dictate the locations of these aliases in the frequency domain. Intuitively, we would like the replicas to not overlap and theoretically the gaps between the replicas to be as small

as possible to reduce the sampling density (downconversion and filtering prior to sampling is a way of reducing these gaps for high frequency signals). However, sampling with Nyquist density can easily produce gaps, which basically means that unnecessarily high sampling density is operated, and the task could be achieved with lower density. This topic has been examined by researchers in areas as diverse as mathematics, signal processing, communications, and circuit design, and the work is spread over many decades in widely differing notations. We will briefly summarise the main results in this category. We start with the work of Mark and Cheng [45], and Cheng [46]. The authors showed that if there are gaps among spectral replications at the Nyquist density then certain periodic samples can be removed with no expenses of losing any information about the signal. Their work targets the gaps that would appear because of the shape of the baseband signal spectrum. For example, let us have a signal with a baseband spectrum of a circular shape. No matter how carefully the sampling matrix is designed, gaps always occur among these circles; see Fig 2.1. Indeed, their work targets multidimensional signals, excluding the one dimensional case. Their approach is to slice the spectrum into narrow cells, and separate those cells which contain signal energy and those which do not.

Sampling rates can be significantly reduced for bandpass signals using *bandpass sampling* (see [47] for a good review on bandpass sampling for one dimensional signal. For a generalization to multidimensional signals, reference [48] provides a concise survey). Although the former scheme is uniform, it shares many aspects with this category. Instead of sampling the signal according to the highest frequency present in the signal, bandpass sampling can significantly reduce the necessary sampling density by tiling the replicas to fill up the entire frequency domain with no gaps. This, of course, requires the signal spectrum of a certain shape to be properly located in certain positions in order for the aliases to be disjoint. Otherwise, either the aliases would overlap, or gaps can appear among them using more than the minimal rate requirement. At this junction we state that the minimal rate

requirement for arbitrary sampling method allowing perfect reconstruction is equal to (the summation of) the bandwidth actually occupied by the signal spectrum, which is known as *Landau rate* [49]. When the signal spectrum is not located "properly", Kohlenberg [50] showed that we can use two sets of uniform sampling to sample bandpass signals with the minimal rate, and therefore it is known as periodic nonuniform sampling of order two. The results were generalised to handle multiband signals in [48], where the order of the periodic nonuniform sampling is larger than two. With all the listed methods, an exact prior knowledge about the spectral support of the signal must be available at the stage of designing the sampling scheme as well as the processing stage. We remark that all the previous approaches can fit or handled under the umbrella of the generalization sampling expansion of Papoulis [51].

A much more challenging, sought problem is to design a *blind* sampling and reconstruction system that does not rely on *a priori* knowledge of the band locations, at asymptotically the minimum rate of Landau. Reconstruction under partial knowledge of the support was addressed in literature. In [52,53] partially blind systems were introduced, where sampling is carried out using (*multi-coset*) periodic nonuniform sampling strategy independent of the bands locations (See [54,55] for a two-dimensional system). However, the recovery is based on some knowledge of the spectral support. A completely blind system was devised in [44] based on the former work and the theory of compressive sensing. This work was slightly modified in [56] to improve the practicality of the system.

All these methods are of significant theoretical aspect, as they can achieve perfect reconstruction of the signal approaching the minimal sampling rates that are required to reconstruct a signal, i.e. (twice for the blind case) Landau. These results are proven for infinitely long sampling sequences and assuming that signal is (bandlimited) multiband.

With the complexity of some of these approaches in the one dimensional case, it makes the high dimensional case is for nothing but theoretical interest.

Another interesting strategy of sampling-reconstruction techniques was initiated in [57], and followed by further research, for processing signals with *finite rate of innovation* (i.e. signals with a finite number of degrees of freedom per second). These techniques can uniformly sample the (prefiltered) signals at a rate slightly above twice their rate of degrees of freedom, which is usually considerably low, and then recover the signal's unknowns to reconstruct the signal using a numerically delicate procedure. Indeed, there are other interesting methodologies that we are not mentioning here, and they vary depending on several factors, such as modelling the signal, previous knowledge about the spectral support of the analysed signal, nonuniform sampling pattern, complexity of calculations, signal processing objective, etc.

2.3 Random Sampling Estimation

Sampling a signal can lead to information loss unless the signal and the sampling process together follow some pattern that has certain conditions. The theory of uniform sampling provides a pattern to achieve the former purpose, which we discussed in Section 2.1. Aliases of the signal spectrum in the uniform sampling environment appear because of the periodicity in the sampling pattern (which can lead to more or less overlapping of these aliases, if the corresponding conditions are not fully fulfilled). Random sampling can avoid the appearance of spectrum aliases, and provide alternative conditions to extract the spectral information from the signal. Random sampling typically focuses on spectral analysis of the signals, i.e. estimating the process's power spectral density or the Fourier transform. The theory of random sampling has been gradually developed. The first paper in the field was by Shapiro and Silverman in 1960 [58] which based their definition of the criterion on the

autocorrelation sequence. Let $\{x(t_n)\}_{n=-\infty}^{\infty}$ be random samples of a wide-sense stationary process with zero-mean and covariance given by

$$r(n)=E\left[x(t_m)x(t_{m+n})\right]=\int_{-\infty}^{+\infty}R(\tau)\,p_n(\tau)d\tau \qquad n=0,\pm1,\pm2, \tag{2.16}$$

where $p_n(\tau)$ is the Probability Density Function (PDF) of $\tau_{m,n}=t_{m+n}-t_m$ and $R(\tau)$ is the autocorrelation function of the underlying continuous ergodic signal. According to their work, the sampling scheme is considered to be sound if there exists only one continuous process with a $R(\tau)$ that would yield $r(n)$ in (2.16). The term "*alias-free*" was given to the sampling pattern that fulfils the former condition. The "alias-free" feature implies that the exact power spectrum density of the continuous stochastic process can be obtained from an infinite number of its irregularly distributed random samples captured at arbitrarily low rates. Random sampling schemes, such as Additive Poisson and Bi-Poisson sampling schemes were found to be alias-free whilst Tri Poisson, jittered sampling and additive random sampling based on uniform distribution are not (the properties of these schemes can be found in the original paper [58] and many other resources such as [12]). Beutler [59] restated Shapiro and Silverman concept to sampling schemes that are alias-free in relation to a $\mathcal{S}$ family of spectra if no two processes with different spectra in $\mathcal{S}$ yielding the same covariance sequence.

Masry [60-62] pointed out the deficiency in the alias-free definition in [58,59]. He highlighted the fact that the Shapiro and Silverman criterion does not guarantee a consistent estimate of the power spectral density of the underlying continuous-time signal for a finite number of signal samples. He reformulated the definition in order to accommodate the latter practical constraint to yield consistent estimators from the samples (he also showed that additive Tri-Poisson sampling is alias-free, which contradicts [58]). In the recent terms, Bilinskis and Mikelsons [12] enunciated a completely new criterion for alias-free sampling:

a scheme is alias-free if all time instants are equally likely to become sampling instants. This ensures that all parts of the signal are sampled with equal probability. They showed that the (Fourier transform or the power spectral density) spectrum of a randomly sampled (deterministic or random stationary) signal can give the underlying continuous-time spectrum using infinite number of samples, provided that the sampling scheme is alias-free according to their definition. This can be shown here by simple analysis. Let us consider the problem of estimating the Fourier transform of a continuous-time deterministic signal from a set of its nonuniformly distributed samples. Using the Fourier transform estimator defined by:

$$\hat{X}(f) := \frac{1}{\alpha} \sum_{n=-\infty}^{+\infty} x(t_n) \exp\left(-j2\pi f t_n\right), \tag{2.17}$$

where α is the average sampling density. It can be noticed that: $x(t_n) = \int_{-\infty}^{+\infty} X(f) \exp\left(j2\pi f t_n\right) df$, where $X(f)$ is Fourier transform of the signal. The expected value of the estimator is

$$E\left[\hat{X}(f)\right] = \frac{1}{\alpha} \int_{-\infty}^{+\infty} X(\nu) \sum_{n=-\infty}^{+\infty} E\left[\exp\left(-j2\pi(f-\nu)t_n\right)\right] d\nu . \tag{2.18}$$

Since $E\left[\exp\left(-j2\pi(f-\nu)t_n\right)\right] = \int_{-\infty}^{+\infty} p_n(t) \exp\left(-j2\pi(f-\nu)t\right) dt$, where $p_n(t)$ is the probability density function of the n-th sampling instant, the expression in (2.18) emerges as:

$$E\left[\hat{X}(f)\right] = X(f) * \mathcal{F}\{p_S(t)\} \tag{2.19}$$

where $\mathcal{F}\{p_S(t)\}$ is the Fourier transform of the sample-point density function $p_S(t) = \sum_{n=1}^{N} p_n(t)$. When $p_S(t)$ is a constant and equal to α, this means that $\mathcal{F}\{p_S(t)\}$ is a

delta and the convolution yields $X(f)$. Hence, $\hat{X}(f)$ is an unbiased estimator of $X(f)$ regardless of the average sampling rate. It was noted that for some sampling schemes a constant sample-point density function might not be observed until a certain time period passes. And, these random sampling schemes have been developed/modified, in the literature, to quickly reach the constant state. In [63-65] the interested reader can find a straightforward discussion and experimental studies on the shapes of the sample-point density functions for a number of randomised sampling techniques.

We recall that such analysis imposes sampling the signal for infinitely long periods using infinite number of samples. In [12] Bilinskis and Mikelsons noted that under the practical constraint of having a finite number of signal samples that cover a finite period of time, their alias-free criterion cannot be totally fulfilled, and "the aliasing is more or less suppressed". And, they noticed that white-noise-like spectral components appear across the whole frequency range with no clear dominant spectral parts, known by them and others as *smeared aliasing*. Most recently, Tarczynski [8,9] followed by Masry [10,11], focused on the practical case of estimating the Fourier transform over a *finite-size* observation domain using a *finite* number of samples. They developed and thoroughly analyse schemes for the former purpose, characterising and quantifying the smeared aliasing that appear with no dominant spectral parts. In fact, the main quest in their work is to devise sampling schemes (and associated processing algorithms) that can minimise the smeared aliasing for a finite set of data. The sampling instants in their case are distributed across the whole finite size observation domain but not necessarily with equal probability (for the goal of minimising the introduced smeared components). Effectively, alias-free criterion in their work which is also adopted in this thesis simply refers to the ability of the appropriately randomised sampling scheme to attenuate spectrum aliasing (in its classical form) within a wide frequency range. This suppression permits the unambiguous identification of the spectral

components of the signal within a bandwidth independent of and much wider than the employed sampling rates. In Chapter 3, the reader will be toured through a more formal and descriptive definition of the alias-free characteristics of the adopted schemes. The interesting work of [8-11] and its applicability in applications, such as multidimensional NMR spectroscopy, promoted our research to investigate and devise random sampling estimation schemes of the Fourier transform over a finite-size *multidimensional* observation domain using a finite number of samples. However, in this thesis we study and develop the Fourier transform estimation schemes in a wider view than the NMR scenario or the original papers [8-11].

2.4 Chapter Summary

Nonuniform sampling is a wide theory. A big part of this theory seeks the interesting target of lowering the sampling density, and various categories of sampling and processing techniques have been proposed in this regard. In lieu of the fact that they go in the same direction, each of these categories formulate the problem in a different way. These categories differ on the type of the signal and its spectrum they deal with as well as their access to it (for example, in applications where the user has to deal with a finite discrete set of data, all the techniques that involve filtering the signal prior to sampling are not applicable). They also vary in terms of the aspects of the data acquisition-processing they strive to improve, as well as the signal processing objective. Accordingly, it is not possible to compare these methodologies, and each one could be more or less appealing depending on its relevance to the particular scenario.

In conclusion, the merit of the adopted methodology, in comparison with other sub-Nyquist techniques can be summarised as follows. First, no restrictive assumption is required to be made about the signal or the shape of its spectral support. Second, the

calculations at the processing stage (i.e. modified forms of the traditional discrete-time Fourier transform calculations) are relatively undemanding. Third, there are no strict limits on the employed sampling densities; the accuracy of the constructed spectrum increases at fast rates with the sampling density, so their values can be determined by the system's specifications and the signal processing objective. Thus, for example, for detection purposes, where sensing the occurrence of spectral activities in a wide range of frequencies is the main aim rather than obtaining the exact features of the signal spectrum, the operated sampling rates can be significantly reduced, see [66]-[68]. Also, the techniques we consider here focus on the practical issue of dealing with a finite set of data. These characteristics and others put this methodology to be the first choice for some applications, such as high dimensional NMR spectroscopy [6]. Additionally, this methodology could be combined with other techniques. That is, the Fourier transform calculated via the random sampling methodology could be the input to other techniques developed in the nonuniform sampling theory which can iteratively enhance the quality of the observed spectrum. Examples of these techniques are the CLEAN methods in their different flavours [69,70] and sequential component extraction method (SECOEX) [71].

CHAPTER THREE

RANDOM SAMPLING ESTIMATION OF MULTIDIMENSIONAL FOURIER TRANSFORMS

In the previous chapter we provided preliminary material about Fourier analysis in multidimensional domains. In particular, we discussed the trivial fact that (real, deterministic, finite energy) signals $x(\mathbf{t})$ are conveniently observed over a finite-size *K*-dimensional window, denoted by *A*, to obtain a form of its Fourier transform $X_w(\mathbf{f}) \equiv X_w(f_1, f_2, ..., f_K)$. This Fourier transform is simply defined by

$$X_w(\mathbf{f}) := \int_A x(\mathbf{t}) w(\mathbf{t}) \exp(-j2\pi \mathbf{f}.\mathbf{t}) d\mathbf{t} \tag{3.1}$$

where $w(\mathbf{t}) \equiv w(t_1, t_2, ..., t_K)$ is a windowing function. As we mentioned in Section 2.1, the windowing function is imposed herein to reduce the effect of truncation, so the Fourier transform (3.1) can be as close as possible to the actual Fourier transform that is calculated over an infinitely large observation domain. We note that choosing the size and the function of windowing in (3.1) is a deterministic problem that is not related to the estimation problem we are tackling in this chapter. And, for notational simplicity we use $x_w(\mathbf{t}) := x(\mathbf{t}) w(\mathbf{t})$ henceforth.

In classical DSP, (the integral in) the Fourier transform (3.1) is *approximated* with a summation using samples of the signal collected in a uniform fashion (2.11), that has its advantages and limitations. As we previously mentioned with random sampling estimation schemes, the sampling points are distributed nonuniformly and randomly in the observation window. Then, these sampling points are used to construct an estimate of the Fourier

transform (3.1) via tailored forms of discrete-time Fourier transform type of calculations. A relatively large variety of random process can, in principle, be applied to the sampling points (and hence the form of the associated Fourier transform calculations). However, there are only few schemes that can lead to alias-free process in the adopted criterion; namely, the sampling process must generate no dominant alias frequencies. In this chapter, we analyse the three most efficient (to the best of our knowledge) random sampling estimation schemes that fulfil the criterion, in multidimensional domains. These schemes are the total random estimation, stratified estimation and antithetical stratified estimation. Since we are studying Fourier transform estimation using random sampling in its general concept and not limited to the scenario of NMR applications, we provide a broad investigation of these schemes. On the other hand, the analysis of error is the topic of interest wherever approximation is involved and given that the main sought target of random sampling is to employ low sampling density, our investigation goes around the critical point of the quality of the generated Fourier transform as a function of the number of samples.

This chapter is organised as follows. We start with describing and analysing the three schemes individually in the following three sections. For each scheme we depict the distribution of the sampling points in the observation window and present an unbiased estimator of the Fourier transform defined in (3.1). The accuracy of the estimates using N samples of the signal are deeply studied and evaluated. In Section 3.4, we study the dependence of their accuracy on the frequency in particular. In Section 3.5, we provide numerical analysis for demonstration based on the provided analytical expressions. We discuss in Section 3.6 the selection of PDFs and the ways of investing prior information about the signal to reduce the error of the estimates. In Section 3.7, we analyse the effect of the observation error on the estimates. Then, we finish the chapter with a summary in Section 3.8.

3.1 Total Random Estimation

In this scheme all the sampling points $\{\mathbf{t}_n\}_{n=1}^{N}$ are Independent Identically Distributed (IID) random variables with a continuous, bounded away from zero PDF $p(\mathbf{t}) \equiv p(t_1, t_2, ..., t_K)$ aligned with the observation window. The estimator of the Fourier transform (3.1) is defined by

$$\hat{X}_{wTR}(\mathbf{f}) := \frac{1}{N}\sum_{n=1}^{N}\frac{x_w(\mathbf{t}_n)\exp(-j2\pi\mathbf{f}.\mathbf{t}_n)}{p(\mathbf{t}_n)}. \tag{3.2}$$

Indeed, the expectation of the estimator is identical with the targeted Fourier transform (2.1):

$$E\left[\hat{X}_{wTR}(\mathbf{f})\right] = \frac{1}{N}\sum_{n=1}^{N}\int_A x_w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})\,d\mathbf{t}$$

$$= \int_A x_w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})\,d\mathbf{t} = X_w(\mathbf{f}). \tag{3.3}$$

Hence, the estimator is unbiased and its variance represents the mean-square estimation error. By exploiting the fact that the sampling points $\mathbf{t}_n$ are IID, the variance of the estimator (3.2) can be derived in the following way:

$$\mathrm{var}\left[\hat{X}_{wTR}(\mathbf{f})\right] = \frac{1}{N}\left(E\left[\left|\frac{x_w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t}_n)}{p(\mathbf{t}_n)}\right|^2\right] - \left|E\left[\frac{x_w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t}_n)}{p(\mathbf{t}_n)}\right]\right|^2\right)$$

$$= \frac{B_{TR}(\mathbf{f})}{N} \tag{3.4}$$

where

$$B_{TR}(\mathbf{f}) = \int_A \frac{x_w{}^2(\mathbf{t})}{p(\mathbf{t})}d\mathbf{t} - \left|X_w(\mathbf{f})\right|^2. \tag{3.5}$$

We note from (3.4) that the mean-square estimation error decays at the rate of $1/N$. This rate is not affected by the number of dimensions or the smoothness of the signal (unlike the subsequent schemes). We also note that the variance varies with the frequency **f**. However, the variance depends actually on the absolute value of the Fourier transform $\left|X_w(\mathbf{f})\right|$ at the frequency **f,** rather than the value of the frequency itself. This observation is of particular interest as it indicates that the quality of estimating the Fourier transform of the signal is *independent* of the position of the signal spectrum in the frequency domain. This observation clearly illustrates the profound difference between uniform sampling and the considered random sampling schemes, and the alias-free characteristics of the latter. In Section 3.6, we include a discussion about the selection of the PDF $p(\mathbf{t})$ that affects the variance (3.5).

Since the sampling points $\{\mathbf{t}_n\}_{n=1}^{N}$ are IID random variables, it follows from the standard multivariate central limit theorem that the scaled real and imaginary parts of the total random estimate: $\sqrt{N}\Re\left[\hat{X}_{wTR}(\mathbf{f})-X_w(\mathbf{f})\right]$ and $\sqrt{N}\Im\left[\hat{X}_{wTR}(\mathbf{f})-X_w(\mathbf{f})\right]$, respectively, are jointly asymptotically normal with zero means and covariance matrix:

$$\Sigma=\begin{bmatrix}\sigma_R^2 & \sigma_{RI}\\ \sigma_{RI} & \sigma_I^2\end{bmatrix} \tag{3.6}$$

where

$$\sigma_R^2 := \left(\int_A \frac{{g_R}^2(\mathbf{t})}{p(\mathbf{t})}d\mathbf{t}-\Re^2\left[X_w(\mathbf{f})\right]\right), \tag{3.7}$$

$$\sigma_I^2 := \left(\int_A \frac{{g_I}^2(\mathbf{t})}{p(\mathbf{t})}d\mathbf{t}-\Im^2\left[X_w(\mathbf{f})\right]\right), \tag{3.8}$$

and

$$\sigma_{RI} := \left(\int_A \frac{g_R(\mathbf{t}) g_I(\mathbf{t})}{p(\mathbf{t})} d\mathbf{t} - \Re\left[X_w(\mathbf{f}) \right] \Im\left[X_w(\mathbf{f}) \right] \right) \tag{3.9}$$

with

$$g_R(\mathbf{t}) := x_w(\mathbf{t}) \cos(2\pi \mathbf{f}.\mathbf{t}) \tag{3.10}$$

where $\cos(2\pi \mathbf{f}.\mathbf{t}) = \cos(2\pi f_1 t_1 + 2\pi f_2 t_2 + ... + 2\pi f_K t_K)$, and

$$g_I(\mathbf{t}) := x_w(\mathbf{t}) \sin(2\pi \mathbf{f}.\mathbf{t}) \tag{3.11}$$

where $\sin(2\pi \mathbf{f}.\mathbf{t}) = \sin(2\pi f_1 t_1 + 2\pi f_2 t_2 + ... + 2\pi f_K t_K)$. These results are particularly important for investigating the distribution of the estimation error for moderate/large N number of samples:

$$P\left[\left| \hat{X}_{wTR}(\mathbf{f}) - X_w(\mathbf{f}) \right| \le \varepsilon \right] \simeq 2\Phi\left(\varepsilon \sqrt{\frac{N}{B_{TR}(\mathbf{f})}} \right) - 1 \tag{3.12}$$

where $\Phi(.)$ is the zero mean unit variance Gaussian distribution function. Hence, with probability $1-\alpha$ the estimation error satisfies (we obtain the confidence intervals)

$$\left| \hat{X}_{wTR}(\mathbf{f}) - X_w(\mathbf{f}) \right| \le \sqrt{\frac{B_{TR}(\mathbf{f})}{N}} \Phi^{-1}(1 - \alpha/2).$$

3.2 Stratified Random Estimation

In this scheme the observation domain A is divided into *S* non-overlapping subdomains A_i:

$$\bigcup_{i=1}^{S} A_i = A. \tag{3.13}$$

Then N_i sampling points are chosen within each subdomain A_i. The sampling points $\mathbf{t}_{i.j} \equiv \left(t_{i,j,1}, t_{i,j,2}, ..., t_{i,j.K}\right)$ $j = 1, 2, ..., N_i$ in the *i*-th subdomain are IID random variables with PDF $p(\mathbf{t})/v_i$ inside the *i*-th subdomain and zero elsewhere, where $v_i := \int_{A_i} p(\mathbf{t}) d\mathbf{t}$ is the weight of the subdomain and $p(\mathbf{t}) \equiv p(t_1, t_2, ..., t_K) = \prod_{d=1}^{K} p_d(t_d)$ is a separable, continuous, positive PDF aligned with the observation domain. The stratified estimator of the Fourier transform (3.1) is given by

$$\hat{X}_{wS}(\mathbf{f}) := \sum_{i=1}^{S} v_i \hat{X}_{S,i}(\mathbf{f}) \tag{3.14}$$

where

$$\hat{X}_{S,i}(\mathbf{f}) = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{x_w(\mathbf{t}_{i,j}) \exp(-j2\pi \mathbf{f}.\mathbf{t}_{i,j})}{p(\mathbf{t}_{i,j})}. \tag{3.15}$$

To prove that the estimator is unbiased, we show that its expectation is identical with the Fourier transform (3.1). First we note that the sampling points in each subdomain are independent of the sampling points in other subdomains and

$$E\left[\hat{X}_{S,i}(\mathbf{f})\right] = \frac{1}{N_i} \sum_{j=1}^{N_i} \int_{A_i} \frac{x_w(\mathbf{t}) \exp(-j2\pi \mathbf{f}.\mathbf{t})}{v_i} d\mathbf{t}$$

$$= \frac{1}{v_i} \int_{A_i} x_w(\mathbf{t}) \exp(-j2\pi \mathbf{f}.\mathbf{t}) d\mathbf{t}. \tag{3.16}$$

Hence,

$$E\left[\hat{X}_{wS}(\mathbf{f})\right] = \sum_{i=1}^{S} v_i E\left[\hat{X}_{S,i}(\mathbf{f})\right]$$

$$= \sum_{i=1}^{S} \int_{A_i} x_w(\mathbf{t}) \exp(-j2\pi \mathbf{f}.\mathbf{t}) d\mathbf{t} = X_w(\mathbf{f}). \tag{3.17}$$

The variance of the estimator is

$$\operatorname{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \sum_{i=1}^{S} v_i^2 \operatorname{var}\left[\hat{X}_{S,i}(\mathbf{f})\right]$$

$$= \sum_{i=1}^{S} \frac{v_i^2}{N_i} \left(\int_{A_i} \frac{x_w^2(\mathbf{t})}{v_i p(\mathbf{t})} d\mathbf{t} - \left| E\left[\hat{X}_{S,i}(\mathbf{f})\right] \right|^2 \right) \tag{3.18}$$

where $\operatorname{var}\left[\hat{X}_{S,i}(\mathbf{f})\right]$ is found by performing the same calculations we went through to find the variance of the total random estimator in Section 3.1. We observe that the way the observation domain is subdivided and the number of samples N_i taken within each subdomain affect the value of the variance of the estimator. A strategy of choosing the number of samples within each subdomain to minimise the variance is through *proportional allocation* [72]. This is done by choosing the number of samples within each subdomain proportionally to its weight v_i:

$$N_i = v_i N \tag{3.19}$$

where N is the total number of the signal samples we seek to collect. We note that when proportional allocation is used the relation between the variance of stratified estimator (3.18) and total random estimator (3.4), when both estimators use the same PDF and the same number of samples, is

$$\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \mathrm{var}\left[\hat{X}_{wTR}(\mathbf{f})\right] + \frac{1}{N}\left|X_w(\mathbf{f})\right|^2 - \sum_{i=1}^{S}\frac{v_i^2}{N_i}\left|E\left[\hat{X}_{S,i}(\mathbf{f})\right]\right|^2 \tag{3.20}$$

with simple calculations, it can be written as follows

$$\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \mathrm{var}\left[\hat{X}_{wTR}(\mathbf{f})\right] - \frac{1}{N}\sum_{i=1}^{S} v_i\left|E\left[\hat{X}_{S,i}(\mathbf{f})\right] - X_w(\mathbf{f})\right|^2 . \tag{3.21}$$

The second term in the above expression cannot be negative; therefore, stratified estimation with proportional allocation offers an improvement over total random estimation. Now, if we increase the stratification, i.e. divide every subdomain into N_i (sub)subdomains $A_{i,j}$ of weight $v_{i,j}$ such that each one contain one sampling point, the last term of (3.20) becomes $\sum_{i=1}^{S}\sum_{j=1}^{N_i} {v_{i,j}}^2\left|E\left[\hat{X}_{S,i,j}(\mathbf{f})\right]\right|^2$. It follows from Cauchy inequality:

$$\left|\sum_{j=1}^{N_i} a_j b_j\right|^2 \le \sum_{j=1}^{N_i}\left|a_j\right|^2 \sum_{j=1}^{N_i}\left|b_j\right|^2 \tag{3.22}$$

with $a_j = \int_{A_{i,j}} x(\mathbf{t})w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})d\mathbf{t}$ and $b_j = 1$, that

$$\frac{1}{N_i}\left|\sum_{j=1}^{N_i}\int_{A_{i,j}} x(\mathbf{t})w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})d\mathbf{t}\right|^2 \le \sum_{j=1}^{N_i}\left|\int_{A_{i,j}} x(\mathbf{t})w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})d\mathbf{t}\right|^2 . \tag{3.23}$$

Now, using (3.16) we see that

$$\frac{v_i^2}{N_i}\left|E\left[\hat{X}_{S,i}(\mathbf{f})\right]\right|^2 \le \sum_{j=1}^{N_i} {v_{i,j}}^2\left|E\left[\hat{X}_{S,i,j}(\mathbf{f})\right]\right|^2 . \tag{3.24}$$

This indicates that full stratification increases the last term of (3.20) and, hence, reduces the variance of $\hat{X}_{wS}(\mathbf{f})$ even further. We conclude that, using *N* samples, the best choice is to divide the observation window into subdomains of the same weight, i.e. $v_i = 1/N$. The variance is clearly

$$\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \mathrm{var}\left[\hat{X}_{wTR}(\mathbf{f})\right] - \frac{1}{N^2}\sum_{i=1}^{N}\left|E\left[\hat{X}_{S,i}(\mathbf{f})\right] - X_w(\mathbf{f})\right|^2. \tag{3.25}$$

Hence,

$$\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \frac{1}{N}\left(\int_A \frac{x_w{}^2(\mathbf{t})}{p(\mathbf{t})}d\mathbf{t} - \frac{1}{N}\sum_{i=1}^{N}\left|E\left[\hat{X}_{S,i}(\mathbf{f})\right]\right|^2\right). \tag{3.26}$$

The rate of convergence of the stratified estimates cannot be determined for a finite number of *N*, from the exact expression of the variance above. As an alternative, it is standard in statistical estimation to determine the rate of convergence from an asymptotic expression of the variance as *N* tends towards infinity. Accordingly, we prove that the variance satisfies

$$\lim_{N\to\infty} N^{1+\frac{2}{K}}\,\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = B_S(\mathbf{f}) \tag{3.27}$$

where

$$B_S(\mathbf{f}) := \frac{1}{12}\sum_{d=1}^{K}\int_A \frac{\left[\dfrac{\partial\left(x(\mathbf{t})w(\mathbf{t})\right)}{\partial t_d}\right]^2 + (2\pi f_d)^2\left[x(\mathbf{t})w(\mathbf{t})\right]^2}{p_d{}^2(t_d)\prod_{k=1}^{K} p_k(t_k)}d\mathbf{t}. \tag{3.28}$$

The proof of (3.28) is in the Appendix A. It follows from above that for sufficiently large *N* the variance of $\hat{X}_{wS}(\mathbf{f})$ is asymptotically equal to

$$\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] \simeq \frac{B_S(\mathbf{f})}{N^{1+\frac{2}{K}}}. \tag{3.29}$$

(In the derivation of (3.28) we assume that $x(\mathbf{t})$ has a continuous derivative along each dimension, and when N goes to infinity, the number of subdomains inside the observation domain goes to infinity with equal number of subdomains along each dimension.) We note from (3.29) that the mean-square error decays asymptotically at the rate of $1/N^{1+2/K}$. The convergence rate of these stratified estimates depends on the number of dimensions; the higher the dimension of the signal is, the slower the rate of the convergence becomes. However, the asymptotic convergence rate of these estimates is always faster than the rate of total random estimates regardless of how large *K* is. A question arises here about the scale of the samples that is required for the asymptotic expression to be valid and, hence, the fast convergence rates to appear. The answer to this question is in Section 3.4, where we examine the dependence of the variance on the frequency.

In Appendix A, we prove the joint asymptotic normality of the real and imaginary parts of the stratified estimates: $N^{0.5+1/K}\Re\left[\hat{X}_{wS}(\mathbf{f}) - X_w(\mathbf{f})\right]$ and $N^{0.5+1/K}\Im\left[\hat{X}_{wS}(\mathbf{f}) - X_w(\mathbf{f})\right]$, which they have zero mean and asymptotic covariance matrix of the following:

$$\Sigma = \begin{bmatrix} \sigma_R^2 & \sigma_{RI} \\ \sigma_{RI} & \sigma_I^2 \end{bmatrix} \tag{3.30}$$

with

$$\sigma_R^2 := \sum_{d=1}^{K} \int_A \frac{\left[\frac{\partial g_R(\mathbf{t})}{\partial t_d}\right]^2}{p_d{}^2(t_d)\prod_{k=1}^{K} p_k(t_k)} d\mathbf{t}, \tag{3.31}$$

$$\sigma_I^2 := \sum_{d=1}^{K} \int_A \frac{\left[\frac{\partial g_I(\mathbf{t})}{\partial t_d}\right]^2}{p_d{}^2(t_d)\prod_{k=1}^{K} p_k(t_k)} d\mathbf{t}, \tag{3.32}$$

and

$$\sigma_{RI} := \sum_{d=1}^{K} \int_A \frac{\frac{\partial g_R(\mathbf{t})}{\partial t_d} \times \frac{\partial g_I(\mathbf{t})}{\partial t_d}}{p_d{}^2(t_d)\prod_{k=1}^{K} p_k(t_k)} d\mathbf{t} \tag{3.33}$$

where $g_R(\mathbf{t})$ and $g_I(\mathbf{t})$ are defined in (3.10) and (3.11), respectively. As we discussed in the previous section, these results give distributional information about the estimation error:

$$P\left[\left|\hat{X}_{wS}(\mathbf{f}) - X_w(\mathbf{f})\right| \le \varepsilon\right] \simeq 2\Phi\left(\varepsilon \frac{N^{0.5+\frac{1}{K}}}{\sqrt{B_S(\mathbf{f})}}\right) - 1 \tag{3.34}$$

which holds for sufficient *N* number of samples.

3.3 Antithetical Stratified Random Estimation

In this scheme the observation domain is also subdivided into *S* non-overlapping subdomains, then select N_i sampling points within each subdomain. Half of the sampling points $\mathbf{t}_{i,j}\ \ j=1,2,...,N_i/2$ within the *i*-th subdomain are IID random variables selected in the same way as in stratified sampling with PDF $p(\mathbf{t})/v_i$ inside the *i*-th subdomain and zero elsewhere, where $p(\mathbf{t}) \equiv p(t_1,t_2,...,t_K) = \prod_{d=1}^{K} p_d(t_d)$ is a continuous, positive, separable PDF aligned with the observation domain and $v_i = \int_{A_i} p(\mathbf{t})d\mathbf{t}$ is the weight of the subdomain. The remaining sampling points are selected as "reflections" of the random ones around the centre points of the subdomains, i.e. random sampling point $\mathbf{t}_{i,j}$ is accompanied by another sampling point at $2\mathbf{c}_i - \mathbf{t}_{i.j}$, where $\mathbf{c}_i \equiv (c_{i,1}, c_{i,2}, ..., c_{i,K})$ is the centre point of the *i*-th subdomain. We emphasise the fact that only one sampling point is selected as the reflection, regardless of the dimensionality of the domain. The antithetical stratified estimator of Fourier transform (3.1) is defined as

$$\hat{X}_{wANT}(\mathbf{f}) := \sum_{i=1}^{S} v_i \hat{X}_{ANT,i}(\mathbf{f}) \tag{3.35}$$

where

$$\hat{X}_{ANT,i}(\mathbf{f}) = \frac{1}{N_i}\sum_{j=1}^{N_i/2}\left(\frac{x_w(\mathbf{t}_{i,j})\exp(-j2\pi\mathbf{f}.\mathbf{t}_{i,j})}{p(\mathbf{t}_{i,j})} + \frac{x_w(2\mathbf{c}_i - \mathbf{t}_{i.j})\exp(-j2\pi\mathbf{f}.(2\mathbf{c}_i - \mathbf{t}_{i.j}))}{p(\mathbf{t}_{i.j})}\right). \tag{3.36}$$

This estimator is also an unbiased estimator of the targeted Fourier transform:

$$E\left[\hat{X}_{ANT,i}(\mathbf{f})\right] = \frac{1}{N_i}\sum_{j=1}^{N_i/2}\int_{A_i}\frac{x(\mathbf{t})w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})}{v_i}d\mathbf{t}$$

$$+\frac{1}{N_i}\sum_{j=1}^{N_i/2}\int_{A_i}\frac{x(2\mathbf{c}_i-\mathbf{t})w(2\mathbf{c}_i-\mathbf{t})\exp(-j2\pi\mathbf{f}.(2\mathbf{c}_i-\mathbf{t}))}{v_i}d\mathbf{t}$$

$$=\frac{1}{v_i}\int_{A_i}x(\mathbf{t})w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})d\mathbf{t}\,. \tag{3.37}$$

Hence,

$$E\left[\hat{X}_{wANT}(\mathbf{f})\right] = \sum_{i=1}^{S}v_i E\left[\hat{X}_{ANT,i}(\mathbf{f})\right]$$

$$=\sum_{i=1}^{S}\int_{A_i}x(\mathbf{t})w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})d\mathbf{t}$$

$$=X_w(\mathbf{f})\,. \tag{3.38}$$

The variance of the estimator is

$$\mathrm{var}\left[\hat{X}_{wANT}(\mathbf{f})\right] = \sum_{i=1}^{S}v_i^{\,2}\,\mathrm{var}\left[\hat{X}_{ANT,i}(\mathbf{f})\right] \tag{3.39}$$

where

$$\mathrm{var}\left[\hat{X}_{ANT,i}(\mathbf{f})\right] = \frac{1}{N_i}\left(\int_{A_i}\frac{x^2(\mathbf{t})w^2(\mathbf{t})}{v_i p(\mathbf{t})}d\mathbf{t} - 2\left|E\left[X_{ANT,i}(\mathbf{f})\right]\right|^2\right.$$

$$\left.+\int_{A_i}\frac{x(\mathbf{t})w(\mathbf{t})x(2\mathbf{c}_i-\mathbf{t})w(2\mathbf{c}_i-\mathbf{t})\cos(4\pi\mathbf{f}.(\mathbf{t}-\mathbf{c}_i))}{v_i p(\mathbf{t})}d\mathbf{t}\right)\,. \tag{3.40}$$

Similarly to the stratified case, proportional allocation with maximum stratification reduces the variance of the estimator. Thus, having subdomains of equal weights, $v_i = 2/N$, with two sampling points in each of them is the best choice to reduce the estimation error.

Unfortunately, apart from some special cases, there is no analytical way to compare the variance of antithetical stratified estimator with those derived earlier. However, a profound evaluation of this scheme can still be obtained. It can be shown that if the analysed signal has continuous second derivatives along each dimension, the antithetical stratified estimates with proportional allocation and maximum stratification can have a faster asymptotic convergence rate than any of the other two schemes. In fact, if the observation domain is subdivided into $N/2$ subdomains with equal number of subdomains along each dimension and two samples in each of them, then (the proof is in Appendix B)

$$\lim_{N\to\infty} N^{1+\frac{4}{K}} \operatorname{var}\left[\hat{X}_{wSANT}(\mathbf{f})\right]$$

$$= \left[\frac{2^{1+\frac{4}{K}}}{288} \sum_{d=1}^{K} \sum_{\substack{r=1 \\ r\neq d}}^{K} \int_A \frac{\left[\frac{\partial^2\left(x(\mathbf{t})w(\mathbf{t})\right)}{\partial t_d \partial t_r} - (2\pi)^2 f_d f_r x(\mathbf{t}) w(\mathbf{t})\right]^2 + \left[2\pi f_r \frac{\partial\left(x(\mathbf{t})w(\mathbf{t})\right)}{\partial t_d} + 2\pi f_d \frac{\partial\left(x(\mathbf{t})w(\mathbf{t})\right)}{\partial t_r}\right]^2}{{p_d}^2(t_d)\,{p_r}^2(t_r)\prod_{k=1}^{K} p_k(t_k)} d\mathbf{t} \right.$$

$$\left. + \frac{2^{1+\frac{4}{K}}}{720} \sum_{d=1}^{K} \int_A \frac{\left[\frac{\partial^2\left(x(\mathbf{t})w(\mathbf{t})\right)}{\partial {t_d}^2} - (2\pi f_d)^2 x(\mathbf{t}) w(\mathbf{t})\right]^2 + (4\pi f_d)^2 \left[\frac{\partial\left(x(\mathbf{t})w(\mathbf{t})\right)}{\partial t_d}\right]^2}{{p_d}^4(t_d)\prod_{k=1}^{K} p_k(t_k)} d\mathbf{t} \right] =: B_{SANT}(\mathbf{f}). \tag{3.41}$$

Hence, for sufficiently high number of samples

$$\operatorname{var}\left[\hat{X}_{wSANT}(\mathbf{f})\right] \simeq \frac{B_{SANT}(\mathbf{f})}{N^{1+\frac{4}{K}}}. \tag{3.42}$$

It follows that the asymptotic mean-square convergence rate of these estimates is $1/N^{1+4/K}$. Thus, antithetical stratified estimation increases the asymptotic convergence rate noticeably

in low-dimensional domains, but the effect reduces in high-dimensional domains. However, these asymptotic convergence rates are always faster than the rates for the total random and stratified estimates. Expression (3.41) also shows that the variance is highly affected by the value of the analysed frequency **f**. A further discussion on the effect of the frequency and asymptotic rate of convergence is included in Section 3.4.

The scaled real and imaginary parts of the antithetical stratified estimate: $N^{0.5+2/K}\Re\left[\hat{X}_{wSANT}(\mathbf{f})-X_w(\mathbf{f})\right]$ and $N^{0.5+2/K}\Im\left[\hat{X}_{wSANT}(\mathbf{f})-X_w(\mathbf{f})\right]$, are jointly asymptotically normal with zero mean and covariance matrix: (see Appendix B for the proof)

$$\Sigma=\begin{bmatrix}\sigma_R^2 & \sigma_{RI}\\ \sigma_{RI} & \sigma_I^2\end{bmatrix} \tag{3.43}$$

with $g_R(\mathbf{t})$ and $g_I(\mathbf{t})$ defined in (3.10) and (3.11), respectively

$$\sigma_R^2 := 2^{1+\frac{4}{K}}\left[\frac{1}{288}\sum_{d=1}^{K}\sum_{\substack{r=1\\ r\neq d}}^{K}\int_A \frac{\left[\frac{\partial^2 g_R(\mathbf{t})}{\partial t_d \partial t_r}\right]^2}{p_d{}^2(t_d)\,p_r{}^2(t_r)\prod_{k=1}^{K}p_k(t_k)}d\mathbf{t}+\frac{1}{720}\sum_{d=1}^{K}\int_A \frac{\left[\frac{\partial^2 g_R(\mathbf{t})}{\partial t_d{}^2}\right]^2}{p_d{}^4(t_d)\prod_{k=1}^{K}p_k(t_k)}d\mathbf{t}\right], \tag{3.44}$$

$$\sigma_I^2 := 2^{1+\frac{4}{K}}\left[\frac{1}{288}\sum_{d=1}^{K}\sum_{\substack{r=1\\ r\neq d}}^{K}\int_A \frac{\left[\frac{\partial^2 g_I(\mathbf{t})}{\partial t_d \partial t_r}\right]^2}{p_d{}^2(t_d)\,p_r{}^2(t_r)\prod_{k=1}^{K}p_k(t_k)}d\mathbf{t}+\frac{1}{720}\sum_{d=1}^{K}\int_A \frac{\left[\frac{\partial^2 g_I(\mathbf{t})}{\partial t_d{}^2}\right]^2}{p_d{}^4(t_d)\prod_{k=1}^{K}p_k(t_k)}d\mathbf{t}\right], \tag{3.45}$$

$$\sigma_{RI} := 2^{1+\frac{4}{K}}\left[\frac{1}{288}\sum_{d=1}^{K}\sum_{\substack{r=1\\ r\neq d}}^{K}\int_A \frac{\frac{\partial^2 g_R(\mathbf{t})}{\partial t_d \partial t_r}\times\frac{\partial^2 g_I(\mathbf{t})}{\partial t_d \partial t_r}}{p_d{}^2(t_d)\,p_r{}^2(t_r)\prod_{k=1}^{K}p_k(t_k)}d\mathbf{t}+\frac{1}{720}\sum_{d=1}^{K}\int_A \frac{\frac{\partial^2 g_R(\mathbf{t})}{\partial t_d{}^2}\times\frac{\partial^2 g_I(\mathbf{t})}{\partial t_d{}^2}}{p_d{}^4(t_d)\prod_{k=1}^{K}p_k(t_k)}d\mathbf{t}\right]. \tag{3.46}$$

One can use the results to examine the distribution of the estimation error for sufficient N number of samples:

$$P\left[\left|\hat{X}_{wSANT}(\mathbf{f}) - X_w(\mathbf{f})\right| \le \varepsilon\right] \simeq 2\Phi\left(\varepsilon \frac{N^{0.5+\frac{2}{K}}}{\sqrt{B_{SANT}(\mathbf{f})}}\right) - 1 .$$

3.4 The Effect of the Frequency on the Rate of Convergence

In the previous sections we profoundly analysed the statistical properties of the estimates to demonstrate and compare their performances. In this section, we study the important phenomenon of the influence of the value of the analysed frequency $\mathbf{f}$ on the accuracy of the estimates. This is particularly important because of its relation to the main objective of random sampling estimates of lowering the sampling density below what is required by the uniform case whose densities are usually governed by the value of the frequencies we seek to analyse.

We start with total random estimates. By examining the derived expression of the variance (3.4), we notice that the variance at a certain frequency is not an explicit function of the frequency. Rather, it depends on the magnitude of the sought Fourier transform. This simply means that a signal of certain spectral shape can be estimated with the same accuracy regardless of its position in the frequency domain. Besides, the rate of convergence is fixed for all frequencies. With stratified scheme (with maximum stratification for its importance), the exact expression of the variance (3.26) depends on the value of the frequency, but it does not explicitly show how. Whereas, the asymptotic expression of the variance (3.27) shows a quadratic dependence on the frequency. Our primary investigation at this point is to determine the scale of the samples that is needed for (3.27) to be potentially valid and, accordingly, the fast convergence rates to appear at a frequency $\mathbf{f}$. By addressing the raised

investigation, we can demonstrate the effect of the frequency on the actual mean-squared estimation error and its decay rate. For easy demonstration we focus on one-dimensional estimates with uniform stratification. Before embarking upon the investigation, we remind the reader about what we remarked earlier about stratified estimates: at any frequency, the mean-square error of the stratified estimates is upper limited by that for the total random estimates, when both schemes use the same number of samples. Thus, the asymptotic expression (3.27) can only hold at frequency f when

$$\begin{aligned}\mathrm{var}\left[\hat{X}_{wTR}(f)\right] &> \frac{B_S(f)}{N^3} \\ &> \frac{D^3}{12N^3}\int_0^T\left[\left(\frac{dx_w(t)}{dt}\right)^2+(2\pi f)^2 x_w{}^2(t)\right]dt\,. \end{aligned} \tag{3.47}$$

Now, the target is to find limits on the number of samples required for (3.47) to hold at frequency *f*. To this end, we rewrite the variance of the total random estimator (with uniform PDF for simplicity) using Parseval's theorem:

$$\begin{aligned}\mathrm{var}\left[\hat{X}_{wTR}(f)\right] &= \frac{D}{N}\left(2\int_0^\infty |X_w(v)|^2\,dv - |X_w(f)|^2\right) \\ &\le \frac{2D}{N}\int_0^\infty |X_w(v)|^2\,dv =: J\,. \end{aligned} \tag{3.48}$$

Similarly, the asymptotic expression can be rewritten in the following form:

$$\begin{aligned}\mathrm{var}\left[\hat{X}_S(f)\right] &\simeq \frac{D^3}{6N^3}\int_0^\infty |(2\pi v)X_w(v)|^2\,dv + \frac{T^3(2\pi f)^2}{6N^3}\int_0^\infty |X_w(v)|^2 dv \\ &=: J_1 + J_2\,. \end{aligned} \tag{3.49}$$

We now define a frequency point f_0 in the spectrum where half of the energy of the signal resides above, i.e.

$$\int_{f_0}^\infty |X_w(v)|^2\,dv = \frac{1}{2}\int_0^\infty |X_w(v)|^2\,dv. \tag{3.50}$$

(Hence, for example, if the signal has a rectangular spectrum that spans the frequency range $\left[f_L, f_H\right]$, the defined frequency f_0 will accordingly be $f_L + 0.5\left(f_H - f_L\right)$.) It follows that the term J_1 in (3.49) satisfies

$$J_1 > \frac{D^3}{6N^3} \int_{f_0}^{\infty} \left|(2\pi v)\, X_w(v)\right|^2 dv$$

$$> \frac{D^3 \left(2\pi f_0\right)^2}{6N^3} \int_{f_0}^{\infty} \left|X_w(v)\right|^2 dv$$

$$= \frac{D^3 \left(2\pi f_0\right)^2}{12N^3} \int_{0}^{\infty} \left|X_w(v)\right|^2 dv =: J_3 . \tag{3.51}$$

We can observe that the smallest number of samples required for

$$J > J_2 + J_3 \tag{3.52}$$

to be true, can only be less than the smallest number that is necessary for (3.47) to be valid. Thus, by replacing the terms in (3.52) with their definitions and simplifying the relation, we conclude that the fast convergence rates can only appear for the number of samples satisfying

$$N > 1.81T\sqrt{f^2 + 0.5 f_0^{\,2}} . \tag{3.53}$$

It follows for the density of the random samples

$$\frac{N}{D} > 1.81\sqrt{f^2 + 0.5 f_0^{\,2}} . \tag{3.54}$$

The expressions above indicate that the number of samples/sampling density required to analyse the signal with fast convergence increases with the distance of the analysed frequency f and the distance of the frequencies that carry the bulk of the energy of the signal, from DC. In the case of analysing high-frequency signals, where using sub-Nyquist random sampling is particularly meaningful, this fact leads to unreasonably high sampling

rates (comparable to what is demanded by the traditional uniform-based methods), or simply missing out on the high convergence rate. Thus, the reported Fourier transform estimates based on stratified techniques do not outperform the other total random sampling approaches for nonbaseband signals. With antithetical stratified estimation, similar limits can be determined which indicate the requirement of a large number of samples in order for the fast convergence rates to be present if the signal spectrum is concentrated far from DC.

3.5 Numerical Analysis

Now, we use the analytical results of the previous sections to illustrate the performance of the estimators numerically. We use a two-dimensional decaying cosine function as a test signal. We choose such signal for easy computations and because of its relation to (NMR) Spectroscopy.

$$x(t_1,t_2)=\cos(2\pi\alpha_1 t_1)\exp\left(\frac{-t_1}{\lambda_1}\right)\cos(2\pi\alpha_2 t_2)\exp\left(\frac{-t_2}{\lambda_2}\right). \tag{3.55}$$

We set $\alpha_1,\alpha_2=3\times10^3$ and $\lambda_1,\lambda_2=10^{-3}$ in (4.1). The signal is observed over a window of the same length along both dimensions 4×10^{-3} with a rectangular windowing function $w(t_1,t_2)$. We use the three estimators to estimate the Fourier transform of the truncated test signal. All the estimators use a uniform PDF. And since we have a full control/freedom on the design of the schemes, we use proportional allocation and maximum stratification where applicable in this numerical example.

To depict the rate of convergence of the schemes, in Fig. 3.1, we show the mean-square error of all the estimators at specific frequency point $(f_1, f_2)=(3,3)\text{kHz}$ as a function of the relative number of random samples to the requisite number of samples to practically achieve

Nyquist if uniform sampling was employed, i.e. 36×36 samples. With log-log scale plot we can manifest the rate of convergence of the estimates by the slope of the curve. The mean-square errors of the estimators shown in Fig. 3.1 are obtained by computational results of the provided exact expressions of the variances. Whereas, the asymptotic mean-square error of stratified and antithetical stratified estimators are calculated using the derived expressions (3.27) and (3.42). We observe that the asymptotic convergence rates we provided theoretically are reached in practice after collecting a sufficient number of samples. These convergence rates are $1/N^{1+2/K}$ and $1/N^{1+4/K}$ with $K=2$ for stratified and antithetical stratified estimates, respectively.

We notice that all the estimators show similar performance when the number of samples N is below what is required to achieve Nyquist density. Also, stratified and antithetical stratified estimates converge at the same rate of total random estimates, $1/N$ in the mean-square sense, and they only show their accelerated convergence rates when the number of samples is higher than what is required to achieve Nyquist. Similar behaviour of the error is observed at other frequency points; however, the acceleration in the convergence rates appears earlier (in term of N) for frequencies closer to DC, and the opposite is true as expected from the limits (3.54). To illustrate the influence of the frequency on the convergence rates of the estimates, we conduct another numerical analysis. We use the same form of the test signal of (3.55). But, we vary its centre frequency (α_1,α_2). In Fig. 3.2, we plot the mean-square error of the three estimates against (α_1,α_2), which we increase its two values α_1 and α_2 simultaneously from 3 kHz to 30 kHz, using the fixed number of samples 100×100.

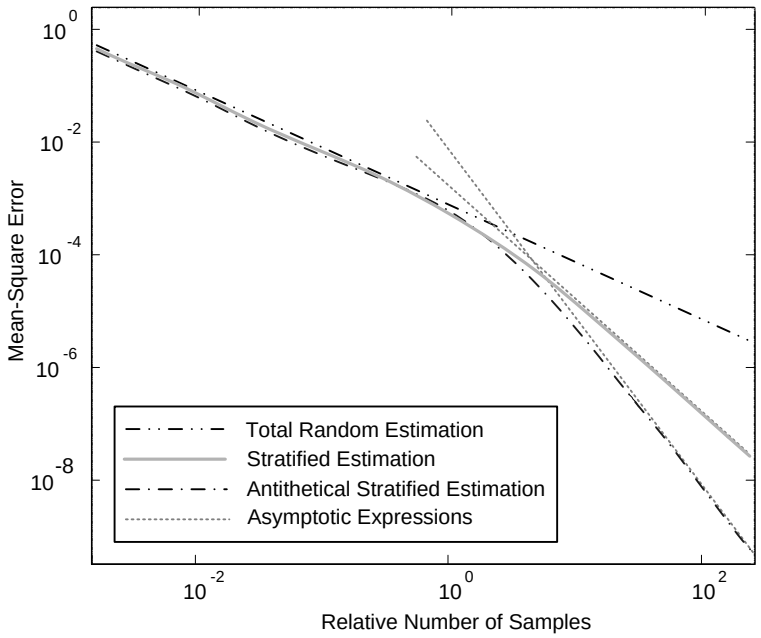

Fig. 3.1. The mean-square error of total random, stratified, antithetical stratified schemes at frequency $(f_1, f_2) = (3,3)$ kHz

In plain term, we effectively vary both of the analysed frequency and the frequencies that carry the bulk of the energy of the signal. Hence, we can see the relation between the signal position in the frequency domain and the accuracy of the estimates. Total random estimation has the same performance across all the frequencies. Whereas, stratified and antithetical stratified estimates deliver excelling results when the signal energy is near DC and degrade with increasing the frequency to reach that of total random estimation.

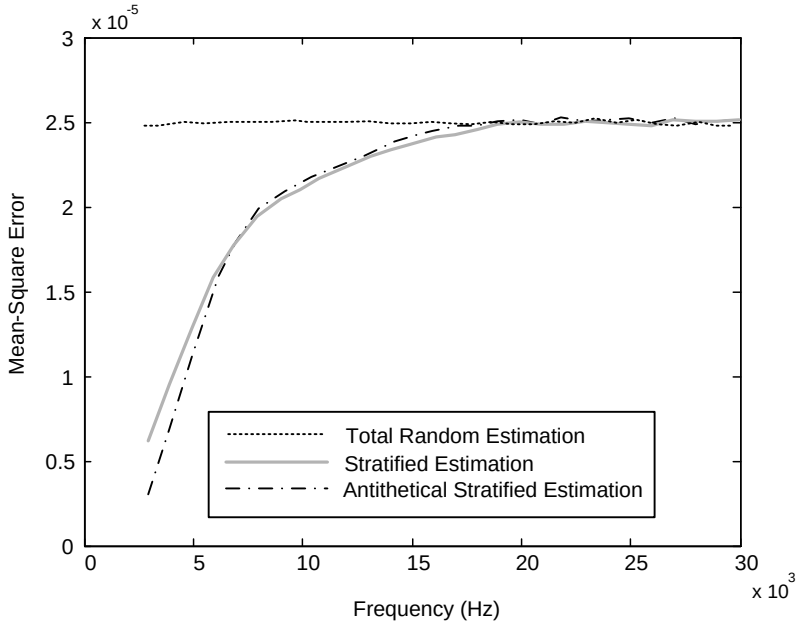

Fig. 3.2. The mean-square error of total random, stratified, antithetical stratified schemes against the frequency α_1 and α_2 simultaneously.

3.6 Probability Density Function (PDF) Design

We can notice from the derived expressions of the variances of the studied estimators that the design of the PDF of the sampling points has an effect on the accuracy of the constructed estimates. The freedom of choosing the PDF could be used to reduce the variance of the estimator. This can be achieved by choosing a density function proportional to $\left|x_w(\mathbf{t})\right|$ to smooth it and bring to a nearly constant function to lessen the variance. Mathematically, this can reduce the first term of the variance (3.5) of total random scheme, i.e. $\int_A x_w^{\,2}(\mathbf{t}) / p(\mathbf{t}) d\mathbf{t}$. This principle is known in Monte Carlo methods (which are numerical integration methods that exploit random assessing points of functions), as *importance sampling* [73]. It is obvious that the best choice is to put the density function in total random scheme as follows (see [9] for the proof)

$$p_{optimal}(\mathbf{t}) := C\left|x_w(\mathbf{t})\right| \tag{3.56}$$

where the constant of proportionality

$$C := \frac{1}{\int_A \left|x_w(\mathbf{t})\right| d\mathbf{t}}. \tag{3.57}$$

Unfortunately, to properly implement this method one has to have *a priori* knowledge of the signal $x(\mathbf{t})$. However, more pragmatic solutions may involve the use of the envelope of the signal or using a density function proportional to the designed window function $w(\mathbf{t})$ to reduce the variance. Nonetheless, having the density function proportional to the window function may or may not smooth the integrand and reduce the variance. This will depend on the shape of the signal and the chosen window function. In [8], the reader can find a comparison between using a PDF that is proportional to the windowing function and a uniform PDF. Their comparison was for some special cases of one dimensional signals. This PDF (which effectively means the way of subdividing the observation window in maximum stratification) can also be used to reduce the error in stratified estimates, which decides the sizes of subdomains in the maximum stratification. However, PDF designs were devised in [10,11] to minimise the mean-square asymptotic error or its integration over a range of frequencies, which also requires the exact signal itself in advance. However, with no previous knowledge about the signal, it is more practical to deploy a uniform PDF in all the schemes.

We have observed from above that some knowledge about the proportionality (which could be along only some of the dimensions) of the signal can be used to smooth the signal and reduce the variance. This brings us to another way of exploiting prior (different type of) information about the signal to decrease the estimation error. The idea is borrowed from

numerical integration literature; it is known as *control variants* [73] and it can be exploited herein as follows. If we have some knowledge about part of an additive signal with computationally obtainable Fourier transform, (for example, if some of the decaying sinusoids in an NMR spectroscopy signal: $x(t)=\sum_{r=1}^{R}\gamma_r\cos(\omega_{0,r}t)\exp(-t/T_{2,r})$ are known in advance), denoted by $s(\mathbf{t})$, we can modify the estimator to create the unbiased estimator of the form (we only show the total random estimator as the other estimators can be modified analogously):

$$\hat{X}_{wTR}(\mathbf{f})=\frac{1}{N}\sum_{n=1}^{N}\frac{\left[x_w(\mathbf{t}_n)-s(\mathbf{t}_n)\right]\exp(-j2\pi\mathbf{f}.\mathbf{t}_n)}{p(\mathbf{t}_n)}+\int_A s(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t}), \qquad (3.58)$$

which can have a lower variance than the basic one, i.e. the first term of the variance (3.4) satisfies

$$\int_A\left[x_w(\mathbf{t})-s(\mathbf{t})\right]^2/p(\mathbf{t})d\mathbf{t}<\int_A {x_w}^2(\mathbf{t})/p(\mathbf{t})d\mathbf{t}.$$

We deduce that depending on the type of the knowledge we have about the signal, we can invests it to reduce the variance in two different ways, and either one of them could be the best way.

3.7 The Effect of the Observation Errors

Here we briefly investigate the statistical properties of the studied random sampling estimates with the presence of the observations error: let $\{\varepsilon_n\}_{n=1}^{N}$ IID random variables with zero mean and ${\sigma_\varepsilon}^2$ variance, and $\{x(\mathbf{t}_n)+\varepsilon_n\}_{n=1}^{N}$ are captured instead of $\{x(\mathbf{t}_n)\}_{n=1}^{N}$. The total random estimator hence takes the following form:

$$\hat{X}_{wTR}(\mathbf{f}) = \frac{1}{N}\sum_{n=1}^{N}\frac{\left(x(\mathbf{t}_n)+\varepsilon_n\right)w(\mathbf{t}_n)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_n\right)}{p\left(\mathbf{t}_n\right)} \tag{3.56}$$

The expected value of the estimator is identical with the targeted Fourier transform (3.1):

$$E\left[\hat{X}_{wTR}(\mathbf{f})\right] = X_w(\mathbf{f}). \tag{3.57}$$

And, the variance is

$$\mathrm{var}\left[\hat{X}_{wTR}(\mathbf{f})\right] = \frac{B_{TR}(\mathbf{f})}{N} + \frac{{\sigma_\varepsilon}^2}{N}\int_A \frac{w^2(\mathbf{t})}{p(\mathbf{t})}d\mathbf{t}. \tag{3.58}$$

We observe that the first term is the same as in (3.4). Whereas, the second term which is due to the observation error is independent of the frequency and decreases at the rate of $1/N$.

For stratified estimation, we assume that $\varepsilon_{i,j}$ $j=1,2,...,N_i$ and $i=1,2,...,S$ are IID random variables with zero mean and ${\sigma_\varepsilon}^2$ variance. And, we are observing $x\left(\mathbf{t}_{i,j}\right)+\varepsilon_{i,j}$ instead of $x\left(\mathbf{t}_{i,j}\right)$. The stratified estimator is now given by

$$\hat{X}_{wS}(\mathbf{f}) = \sum_{i=1}^{S} v_i \hat{X}_{S,i}(\mathbf{f}) \tag{3.59}$$

where

$$\hat{X}_{S,i}(\mathbf{f}) = \frac{1}{N_i}\sum_{j=1}^{N_i}\frac{\left(x\left(\mathbf{t}_{i,j}\right)+\varepsilon_{i,j}\right)w\left(\mathbf{t}_{i,j}\right)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_{i,j}\right)}{p\left(\mathbf{t}_{i,j}\right)}. \tag{3.60}$$

We can show that the estimator is still unbiased:

$$E\left[\hat{X}_{wS}(\mathbf{f})\right] = X_w(\mathbf{f}). \tag{3.61}$$

And, the variance of the stratified estimator with maximum stratification (which is the most interesting case) is

$$\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \frac{1}{N}\left(\int_A \frac{{x_w}^2(\mathbf{t})}{p(\mathbf{t})}d\mathbf{t} - \sum_{i=1}^{N}\left|\int_{A_i} x_w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t})d\mathbf{t}\right|^2\right) + \frac{{\sigma_\varepsilon}^2}{N}\int_A \frac{w^2(\mathbf{t})}{p(\mathbf{t})}d\mathbf{t}. \quad (3.62)$$

We also see that the first term is equivalent to the variance of the stratified estimator with no noise (3.26). Last, the antithetical stratified estimator with the presence of additive error is given by

$$\hat{X}_{wANT}(\mathbf{f}) = \sum_{i=1}^{S} v_i \hat{X}_{ANT,i}(\mathbf{f}) \quad (3.63)$$

where

$$\hat{X}_{ANT,i}(\mathbf{f}) = \frac{1}{N_i}\sum_{j=1}^{N_i/2}\left(\frac{\left(x(\mathbf{t}_{i,j})+\varepsilon_{i,j}\right)w(\mathbf{t}_{i,j})\exp(-j2\pi\mathbf{f}.\mathbf{t}_{i,j})}{p(\mathbf{t}_{i,j})}\right.$$
$$\left. + \frac{\left(x(2\mathbf{c}_i-\mathbf{t}_{i.j})+\varepsilon'_{i,j}\right)w(2\mathbf{c}_i-\mathbf{t}_{i.j})\exp\left(-j2\pi\mathbf{f}.(2\mathbf{c}_i-\mathbf{t}_{i.j})\right)}{p(\mathbf{t}_{i.j})}\right). \quad (3.64)$$

And, $\{\varepsilon_{i,j}\}$ and $\{\varepsilon'_{i,j}\}$ are IID random variables with zero mean and variance ${\sigma_\varepsilon}^2$. Second-order calculations show that the estimate remains unbiased and that its variance has an additional term equal to the one of total random and stratified sampling schemes.

Thus, all the estimators have an additional term because of the observation error that is independent of the frequency and decreases at the rate of $1/N$. For such observation error in uniform-sampling-based applications, the signal is observed (i.e. a sample is captured) multiple times to improve the level of error in the generated Fourier transform. Whereas, in random sampling approaches we can acquire additional random samples (at different

sampling positions) which allows the error level to decrease in the calculated spectrum as well as more spectral information to be captured to reduce the mean-squared error that appear because of the sampling process, allowing the overall quality of the estimated Fourier transform to be improved.

3.8 Chapter Summary

Three random sampling estimation schemes of the multidimensional Fourier transform were considered. The schemes use finite numbers of randomly selected samples of windowed signals to estimate their Fourier transform via tailored forms of the simple calculation of discrete-time Fourier transform. The main difference between these schemes is the pattern of distributing the sampling points, which impacts the accuracy of the produced estimates. The main advantage of these methods is that the estimators approximate the Fourier transform of the signal in infinitely large frequency domain using any number of samples with quality improves with increasing the number of samples N. Unlike classical uniform sampling based approaches, the size of the processing frequency domain and the dimensionality of the signal do not govern the employed sampling density. In fact, it was shown that the accuracy of the Fourier transform estimates is independent of the position of the signal in the frequency domain (the stratification technique can provide an improvement if the signal is located near DC; however, their accuracy is lower bounded by that for total random estimation which is independent of the value of the analysed frequency). The former quality is the splendour of random-sampling-based approaches, as opposed to uniform sampling which is usually dictated by the dimensionality and the spectral position of the signal and hence requires high sampling densities and/or downconversion and filtering (if possible) prior to sampling if the signal is located far from DC

We have demonstrated that the standard mean-square convergence rate of $1/N$ for the total random estimates is maintained by all the studied estimates for small number of samples N. Also, total random sampling estimates deliver the rate of convergence of $1/N$ regardless of the dimensionality of the signal, the frequency, or the number of samples. Whereas, stratified and antithetical stratified schemes deliver convergence rates that can accelerate with increasing *N* and reach the fast rates that depend on the dimensionality K, i.e. $1/N^{1+2/K}$ and $1/N^{1+4/K}$ respectively, for sufficiently smooth signals. Also, the required number of sampling in order for the acceleration to emerge depends on the distance of the signal spectrum from zero.

We conclude that total random sampling scheme can be appealing for applications where the signal is of high dimensionality, not enough smoothness or very high frequency compared with the prospective sampling density, and there are extra complications involved with practising stratification (for example, in deploying random sensors in fields). Whereas, stratified and antithetical stratified schemes shine if the signal is of low dimensionality and has its spectrum near the acceleration frequency, i.e. DC. We have also shown that the antithetical stratified scheme can outperform its regular stratified counterpart. However, for the case where choosing the reflection sampling points in antithetical sampling can be inaccurate, it is more effective to employ further (double) stratification and collect one sample from each subdomain.

We must mention here that there are other random schemes that can be developed from the numerical integration literature which would produce the same performance of total random sampling for Fourier transform estimation. Example of these schemes is *Latin hypercubic sampling* [73]. According to this scheme the observation domain is divided into N^K subdomains, and then N subdomains are randomly selected such that only one of them occurs in each row and column of the subdomains. Afterwards, one sampling point is

randomly and independently selected from each chosen subdomain. With these further complications and no improvement in the performance, it seems pointless for us to study this scheme here.

CHAPTER FOUR

RANDOM ON GRID SAMPLING ESTIMATION OF MULTIDIMENSIONAL FOURIER TRANSFORMS

Three random sampling estimation schemes of the Fourier transform were considered in the previous chapter. These schemes employ sampling points with freely chosen coordinates. In fact, any point in the observation domain can be a sampling point. This means that the sampling points can be arbitrarily close to each other. This is not feasible because of, for example, the finite steps of the system's clock and the limitations of building acquisition devices that are infinitely fast to capture such successive samples (in the real time dimension). For example, see [76] for a hardware implementation of random sampling, where the sampling instants are conveniently captured on a multiple of some time interval. And, see [6] where the sampling points are rounded to their coordinates, which effectively means that the sampling points are randomly selected from a Cartesian grid of points. For the calculations of spectra acquired with random on grid sampling it is possible to use the FFT-type of algorithms instead of performing the relatively slow direct calculation of the random (off-grid) schemes studied in the previous chapter, or recasting the sampled data into a Cartesian grid to employ the fast algorithms. Hence, the numerical efficiency of random sampling estimation can be increased with introducing the grid, especially in multidimensional signals where computational savings gained through FFT are very important because of the great volume of data.

The above drove us to provide an analytical and numerical study of the on-grid versions of the random sampling schemes studied in the previous chapter (which we refer to in this chapter by the off-grid schemes for distinction). Some of the results presented in this chapter

might seem straightforward, but the aim is to provide a complete survey of their sampling patterns, estimators and accuracy (In fact, we have engaged these schemes in devising spectrum sensing algorithms, and used the analysis we providing here to obtain guidelines for the user [77,78]). In addition, this chapter facilitates demonstrating some aspects in Chapter 5. Indeed, the Fourier transform the off-grid estimation schemes theoretically seek is not valid for the on-grid estimation. The targeted Fourier transform here can be written as follows:

$$X_{wD}(\mathbf{f}) \coloneqq \frac{D}{M}\sum_{m=1}^{M} x_w(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_m\right) \tag{4.1}$$

where M is the total number of the Cartesian grid points, D is the size of the observation domain and $\mathbf{t}_m \equiv \left(t_{m,1}, t_{m,2}, ..., t_{m,K}\right)$, $m = 1, 2, ..., M$ are the grid points. Indeed, by introducing the Cartesian grid, the above Fourier transform consist of replicas of the Fourier transform (3.1) (that for the off-grid random sampling estimates of Chapter 3) and the area that is free of alias frequencies is limited. Accordingly, in this case a conservative prior knowledge about the highest frequency present in the signal is required. And, whether the grid is deliberately designed by the user or enforced because of practical consideration, its density must be high enough such that the replicas are shifted away from the region of interest and (4.1) serves a good approximation of the Fourier transform (3.1).

In the next sections, we look at total random on grid estimation, stratified on grid estimation and antithetical stratified on grid estimation, in multidimensional space, and provide expressions for their variances. Then we use these derived expressions in a numerical test to evaluate the schemes and compare them with each other as well as their off-grid counterparts, providing a better insight into their performances.

4.1 Total Random on Grid Estimation

In this estimation scheme the sampling points are selected as an N-element subset of M grid points. There are $\binom{M}{N}$ different possibilities of selecting the set of sampling points each with the same probability. To achieve this scenario and analyse the estimates, we use a random binary sequence $\{a_m\}_{m=1}^{M}$ that we associate with $\{\mathbf{t}_m\}_{m=1}^{M}$ such that $a_m = 1$ when the grid point $\mathbf{t}_m$ is selected and $a_m = 0$ otherwise. These random coefficients have the following probabilities

$$P[a_m = 1] = \frac{N}{M} \tag{4.2}$$

$$P[a_m = 0] = \frac{M-N}{M} \ . \tag{4.3}$$

We note that the variables $\{a_m\}_{m=1}^{M}$ are not independent from each other. When estimating the Fourier transform (4.1) of the signal we use the following formula

$$\hat{X}_{TRG}(\mathbf{f}) := \frac{D}{N}\sum_{m=1}^{M} a_m x_w(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_m\right). \tag{4.4}$$

The mean of each random variable a_m is given by

$$E[a_m] = \frac{N}{M} \ . \tag{4.5}$$

The expected value of the estimator (4.4) is

$$\begin{aligned} E\left[\hat{X}_{TRG}(\mathbf{f})\right] &= \frac{D}{N}\sum_{m=1}^{M} E[a_m]\, x_w(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_m\right) \\ &= \frac{D}{N}\sum_{m=1}^{M}\frac{N}{M} x_w(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_m\right) = X_{wD}(\mathbf{f}) \, . \end{aligned} \tag{4.6}$$

This proves that total random on grid estimator is an unbiased estimator of the Fourier transform (4.1). The variance of total random on grid estimator can be found as follows

$$\operatorname{var}\left[\hat{X}_{TRG}(\mathbf{f})\right]=\frac{D^2}{N^2}\left[\sum_{m=1}^{M} x_w{}^2(\mathbf{t}_m)\operatorname{var}[a_m] + \sum_{m=1}^{M}\sum_{k=1,k\neq m}^{M} x_w(\mathbf{t}_m)\exp(-j2\pi\mathbf{f}.\mathbf{t}_m)x_w(\mathbf{t}_k)\exp(j2\pi\mathbf{f}.\mathbf{t}_k)\operatorname{cov}[a_m,a_k]\right] \tag{4.7}$$

where

$$\operatorname{var}[a_m]=\left(1-\frac{N}{M}\right)\frac{N}{M}=\frac{N(M-N)}{M^2}. \tag{4.8}$$

Since the probability that two grid points are selected in a set of sampling points is

$$P[a_m a_k=1]=\frac{N}{M}\left(\frac{N-1}{M-1}\right), \tag{4.8}$$

the covariance is given by

$$\operatorname{cov}[a_m,a_k]=\frac{N}{M}\left(\frac{N-1}{M-1}\right)-\left(\frac{N}{M}\right)^2=\frac{-N(M-N)}{M^2(M-1)}. \tag{4.9}$$

Thus,

$$\operatorname{var}\left[\hat{X}_{TRG}(\mathbf{f})\right]=\frac{D^2(M-N)}{M^2N}\left[\sum_{m=1}^{M} x_w{}^2(\mathbf{t}_m) - \frac{1}{M-1}\sum_{m=1}^{M}\sum_{k=1,k\neq m}^{M} x_w(\mathbf{t}_m)\exp(-j2\pi\mathbf{f}.\mathbf{t}_m)\, x_w(\mathbf{t}_k)\exp(j2\pi\mathbf{f}.\mathbf{t}_k)\right]. \tag{4.10}$$

By completing the square of the second term of (4.10), we obtain

$$\operatorname{var}\left[\hat{X}_{TRG}(\mathbf{f})\right]=\frac{D^2(M-N)}{M^2N}\left[\frac{M}{M-1}\sum_{m=1}^{M} x_w{}^2(\mathbf{t}_m)-\frac{M^2}{D^2(M-1)}\left|X_{wD}(\mathbf{f})\right|^2\right]. \tag{4.11}$$

Hence, the variance of the estimator is

$$\mathrm{var}\left[\hat{X}_{TRG}(\mathbf{f})\right]=\frac{B_{TRG}(\mathbf{f})}{N}\left(1-\frac{N}{M}\right) \tag{4.12}$$

where

$$B_{TRG}(\mathbf{f})=\left[\frac{D^2}{M-1}\sum_{m=1}^{M}x_w{}^2(\mathbf{t}_m)-\frac{M}{M-1}\left|X_{wD}(\mathbf{f})\right|^2\right]. \tag{4.13}$$

By examining the expression above we notice that the constant $B_{TRG}(\mathbf{f})$ is similar to the total random constant $B_{TR}(\mathbf{f})$. What differs the variances of the two estimators is the factor $1-N/M$. This factor plays the decisive role in shaping $\mathrm{var}\left[\hat{X}_{TRG}(\mathbf{f})\right]$ when the number of samples *N* approaches the number of grid points M. Hence, if the number of samples we aim to collect is close to the number of grid points, total random on grid scheme is more accurate at estimating its target than its off-grid counterpart.

4.2 Stratified on Grid Estimation

Similarly to off-grid stratified estimation, this estimation scheme relies on dividing the observation domain into *S* non-overlapping subdomains $\{A_i\}_{i=1}^{S}$. Then within the *i*-th subdomain we select N_i sampling points in the same manner as in Section 4.1. We estimate the Fourier transform (4.1) using the following estimator

$$\hat{X}_{SG}(\mathbf{f}):=\sum_{i=1}^{S}v_i\hat{X}_{SG,i}(\mathbf{f}) \tag{4.14}$$

where

$$\hat{X}_{SG,i}(\mathbf{f})=\frac{D}{N_i}\sum_{\mathbf{t}_m\in A_i}a_m x_w(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_m\right) \tag{4.15}$$

where M_i is the number of grid points inside the *i*-th subdomain and $v_i = M_i / M$ is its weight. The random variables a_m inside the *i*-th subdomain have the probability to be one or zero according to

$$P[a_m = 1] = \frac{N_i}{M_i} \tag{4.16}$$

$$P[a_m = 0] = \frac{M_i - N_i}{M_i}. \tag{4.17}$$

The random variables $\{a_m\}$ are dependent within each subdomain but they are independent of the other variables in other subdomains, and their unconditional probabilities are the same in each subdomain, defined by (4.16) and (4.17). The mean of the random variables a_m inside each subdomain are identical and equal to

$$E[a_m] = \frac{N_i}{M_i}. \tag{4.18}$$

Hence,

$$E\left[\hat{X}_{SG,i}(\mathbf{f})\right] = \frac{D}{N_i} \sum_{\mathbf{t}_m \in A_i} E[a_m] x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m)$$

$$= \frac{D}{M_i} \sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m). \tag{4.19}$$

The stratified on grid estimator is an unbiased estimator of the Fourier transform (4.1):

$$E\left[\hat{X}_{SG}(\mathbf{f})\right] = \sum_{i=1}^{S} v_i E\left[\hat{X}_{SG,i}(\mathbf{f})\right]$$

$$= \sum_{i=1}^{S} \frac{v_i D}{M_i} \sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m)$$

$$= \frac{D}{M} \sum_{i=1}^{S} \sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m)$$

$$= X_{wD}(\mathbf{f}). \tag{4.20}$$

The variance of the estimator is

$$\operatorname{var}\left[\hat{X}_{SG}(\mathbf{f})\right] = \sum_{i=1}^{S} v_i^2 \operatorname{var}\left[\hat{X}_{SG,i}(\mathbf{f})\right] \tag{4.21}$$

where

$$\operatorname{var}\left[\hat{X}_{SG,i}(\mathbf{f})\right] = \frac{\left[\frac{D^2}{M_i - 1}\sum_{\mathbf{t}_m \in A_i} x_w^{\,2}(\mathbf{t}_m) - \frac{M_i}{M_i - 1}\left|E\left[\hat{X}_{SG,i}(\mathbf{f})\right]\right|^2\right]}{N_i}\left(1 - \frac{N_i}{M_i}\right). \tag{4.22}$$

When the proportional allocation is used to distribute the sampling points between the subdomains, the relation between the variance of the stratified on grid estimator and total random on grid estimator is

$$\operatorname{var}\left[\hat{X}_{SG}(\mathbf{f})\right] = \operatorname{var}\left[\hat{X}_{TRG}(\mathbf{f})\right] - \frac{1 - \left(\frac{N}{M}\right)}{N}\left[\sum_{i=1}^{S}\frac{M_i}{(M-1)}\left|E\left[\hat{X}_{SG,i}(\mathbf{f})\right] - X_{wD}(\mathbf{f})\right|^2 - \frac{1}{M(M-1)}\sum_{i=1}^{S}(M - M_i)\operatorname{var}\left[\hat{X}_{SG,i}(\mathbf{f})\right]\right]. \tag{4.22}$$

If M is considerably large, the expression becomes

$$\operatorname{var}\left[\hat{X}_{SG}(\mathbf{f})\right] = \operatorname{var}\left[\hat{X}_{TRG}(\mathbf{f})\right] - \frac{1 - \left(\frac{N}{M}\right)}{N}\sum_{i=1}^{S} v_i \left|E\left[\hat{X}_{SG,i}(\mathbf{f})\right] - X_{wD}(\mathbf{f})\right|^2. \tag{4.23}$$

We note that, unlike off-grid stratified estimators, the second term in (4.22) can be negative and hence the stratified on grid estimator with proportional allocation can have a larger variance than the total random on grid estimator. On the other hand, if the grid is significantly dense, formula (4.23) shows that the stratified on gird estimator delivers better

results compared to the total random on grid estimator. By designing the subdomains of equal sizes and collecting one sample from each subdomain, the variance is clearly

$$\mathrm{var}\left[\hat{X}_{SG}(\mathbf{f})\right] = \mathrm{var}\left[\hat{X}_{TRG}(\mathbf{f})\right] - \frac{1-\left(\frac{N}{M}\right)}{N^2}\sum_{i=1}^{N}\left|E\left[\hat{X}_{SG,i}(\mathbf{f})\right] - X_{wD}(\mathbf{f})\right|^2 . \qquad (4.24)$$

The performance of the stratified on grid estimator compared with total random on grid estimator is more evident in the numerical results in Section 4.4.

4.3 Antithetical Stratified on Grid Estimation

The Cartesian gird in this scheme must be divided such that each subdomain contains an even number of the grid points. As we showed above, maximum stratification reduces the estimation error. And since this scheme is mainly for applications where the user has complete freedom of designing the subdomains and the number/location of the sampling points within, we focus here on studying the scheme with maximum stratification. That is, only two sampling points are selected per subdomain. One sampling point is selected randomly from the grid points. The other sampling point is selected as reflection of the random one around the centre point of the subdomain. Explicitly, for each randomly selected grid point $\mathbf{t}_m$, we allocate another grid point as a sampling point at $2\mathbf{c}_i - \mathbf{t}_m$, where $\mathbf{c}_i \equiv \left(c_{i,1}, c_{i,2}, ..., c_{i,K}\right)$ is the centre point between the first and the last grid points in the *i*-th subdomain. The Fourier transform is estimated using the following estimator:

$$\hat{X}_{AG}(\mathbf{f}) := \sum_{i=1}^{S} v_i \hat{X}_{A,i}(\mathbf{f}) \qquad (4.25)$$

where $v_i = M_i / 2M$, and

$$\hat{X}_{A,i}(\mathbf{f}) = D \sum_{\mathbf{t}_m \in A_i} a_m \Big[x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m)$$

$$+ x_w(2\mathbf{c}_i - \mathbf{t}_m) \exp\left(-j2\pi \mathbf{f}.(2\mathbf{c}_i - \mathbf{t}_m)\right) \Big]. \tag{4.26}$$

The random variables $\{a_m\}$ that belong to the *i*-th subdomain, have the probability to be one or zero according to

$$P[a_m = 1] = \frac{1}{M_i} \tag{4.27}$$

$$P[a_m = 0] = \frac{M_i - 1}{M_i}. \tag{4.28}$$

The means of the random variables $\{a_m\}$ inside each subdomain are equal to

$$E[a_m] = \frac{1}{M_i}. \tag{4.29}$$

Hence,

$$E\left[\hat{X}_{A,i}(\mathbf{f})\right] = D \sum_{\mathbf{t}_m \in A_i} E[a_m] \left[x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m) + x_w(2\mathbf{c}_i - \mathbf{t}_m) \exp\left(-j2\pi \mathbf{f}.(2\mathbf{c}_i - \mathbf{t}_m)\right) \right]$$

$$= \frac{D}{M_i} \sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m). \tag{4.30}$$

The above estimator is an unbiased estimator of the Fourier transform (4.1):

$$E\left[\hat{X}_{AG}(\mathbf{f})\right] = \sum_{i=1}^{S} v_i E\left[\hat{X}_{A,i}(\mathbf{f})\right]$$

$$= \sum_{i=1}^{S} \frac{2v_i D}{M_i} \sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m)$$

$$= \frac{D}{M}\sum_{i=1}^{S}\sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m)\exp(-j2\pi\mathbf{f}.\mathbf{t}_m)$$

$$= X_{wD}(\mathbf{f}). \tag{4.31}$$

The variance of the estimator is

$$\mathrm{var}\left[\hat{X}_{AG}(\mathbf{f})\right] = \frac{D^2}{M^2}\sum_{=1}^{N/2}\left(\frac{M_i}{2}\sum_{\mathbf{t}_m \in A_i} x_w{}^2(\mathbf{t}_m) + \frac{M_i}{2}\sum_{A_i} x_w(\mathbf{t}_m)x_w(2c_n' - \mathbf{t}_m)\cos\left(2\pi\mathbf{f}.(2c_n' - 2\mathbf{t}_m)\right)\right.$$
$$\left. - \left|\sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m)\exp(-j2\pi\mathbf{f}.\mathbf{t}_m)\right|^2\right). \tag{4.32}$$

In the next section, we use the expression of the variance derived above to demonstrate the performance of this scheme in a numerical test.

4.4 Numerical Results

We use the test signal of Section 3.5 to illustrate the performance of the on-grid estimators described in this chapter. A practical operation that we would like to address in this test is to compare the performance of all the estimators we discussed so far in Chapter 3 and 4. However, there is an issue with the fact that the target (3.1) for the off-grid estimation is not the same as the target (4.1) for the on-grid estimation. The difference between them could be considered as the bias of the on-grid approaches. When the density of the gird is relatively high, the bias is very small compared to the variance, and hence the bias does not affect significantly any comparisons. In this numerical test we therefore compare only the variances of the estimators. Here, we set the grid density as $f_T = 40\times10^3$ along both dimensions. For frequencies $\{(f_1, f_2): -f_T/2 < f_1 < f_T/2 \text{ and } -f_T/2 < f_2 < f_T/2\}$, the targeted Fourier transforms, (3.1) and (4.1), almost represent the same quantity and a

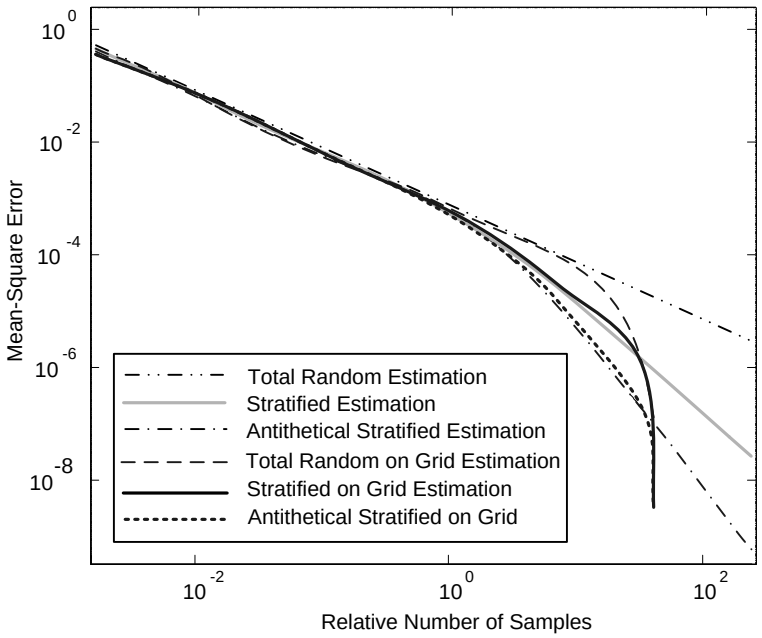

Fig. 4.1. The mean-square error of total random, stratified, antithetical stratified, total random on grid, stratified on grid, and antithetical stratified on grid estimators at frequency $(f_1, f_2) = (3,3)\,\text{kHz}$

comparison between their estimation errors is justified. We use uniform PDF/partitioning and maximum stratification, where applicable, to extract the best performance out of these schemes. Fig. 4.1 shows the mean-square error of all the estimators at $(f_1, f_2) = (3,3)\,\text{kHz}$ versus the relative number of random samples to the required number of samples to practically achieve Nyquist if uniform sampling was deployed, i.e. 36×36 samples. We see that the error of the on-grids schemes show similar behaviour to the error of their counterparts. However, when the number of collected samples N approaches the number of gird points M, the on-grid schemes tend to introduce less error. We conclude that the on-grid estimators do not perform less than their off-grid counterparts. And, the on-grid estimators do accelerate and deliver the same fast convergence rate at the same number of

samples the off-grid estimates do. These results are consistent for different frequencies and different signals, given that the grid is sufficiently dense. And, the (antithetical) stratified on-grid estimation schemes have their acceleration frequency at zero and require the same number of samples in order for their convergence rates to accelerate as their off-grid counterparts.

CHAPTER FIVE

THE IQ ESTIMATION

In the previous chapters we deeply analysed random sampling estimates of the Fourier transform, using N number of random samples. The focus of the previous chapters, as well as of [8]-[11], is on the error of the estimated Fourier transform for N number of samples. We found that the total random (on-grid and off-grid) sampling estimates have a fixed mean-square error convergence rate of $1/N$, regardless of the dimensions of the signal or its location in the frequency domain. Whereas, stratified sampling and antithetical stratified sampling show a different behaviour. These advanced random sampling estimates deliver a mean-square convergence rate that undergoes an acceleration with increasing the number of samples. It was shown that for sufficiently smooth K-dimensional signals, the convergence rates can accelerate to reach $1/N^{1+2/K}$ for the stratified (on-grid and off-grid) estimates and $1/N^{1+4/K}$ for the antithetical (on-grid and off-grid) stratified estimates. These convergence rates are considerably fast for low-dimensional domain, compared with the rate these estimates deliver before they accelerate (i.e. $1/N$) or the rates rendered by total random schemes. With these fast convergence rates a substantial reduction in the estimation error can be achieved for N number of samples.

In Addition, we found an important fact about that the (antithetical) stratified schemes, namely, the acceleration of their convergence rates does not start uniformly (in terms of the number of samples) across all frequencies. First, it takes place at their acceleration frequency, i.e. DC, and its close neighbourhood, and it spreads to higher frequencies only when the average sampling densities significantly increase. In details, the requisite number

of samples in order to benefit from the fast convergence increases with the distance of the analysed frequency and the frequencies that carry most of the energy of the signal, from the acceleration/DC. This observation defeats the main objective of random sampling of lowering the sampling density.

The main focus of this chapter is to develop a novel technique of (antithetical) stratified sampling for Fourier transform estimation where we have the freedom of choosing the location of the acceleration frequency. Consequently, the acceleration and hence the fast convergence to the targeted Fourier transform can be obtained for small numbers of samples regardless of the positioning of the signal spectrum in the frequency domain. This is important for high-frequency signals, where random sampling is particularly meaningful to engage in the first place.

The above developed technique provides an alternative solution to downconversion and filtering the analogue signal prior to sampling at no additional cost. And, it is a vital solution to a wide range of scenarios, especially in multidimensional applications, where downconversion-filtering before sampling is not feasible. These scenarios are met in applications where we only have access to discrete samples of the signal while the target is to minimise the size of the collected data. For such applications, the IQ estimation serves an excellent tool. Examples include minimising the number of sensors (i.e. sampling points in our definition) in spatial sampling.

Additionally, the developed methods leverage another principle to provide further reduction in the estimation error (compared with the regular random approaches) that is independent from the error reduction rendered by facilitating the fast convergence rates. In fact, this error reduction mechanism is gained through symmetry of the Fourier transform, and can be employed for total random estimation that does not undergo any acceleration in its convergence rates, reducing its error by a considerable extent. Although it might seem

pointless to introduce/study the IQ total random estimation, we briefly describe its form and statistical properties for theoretical interest and to clearly illustrate the principle of error reduction through symmetry that the IQ (antithetical) stratified schemes also benefit from.

As we mentioned in Chapter 1, we refer to the proposed technique as the IQ estimation because of the fact that it estimates the Fourier transform of the signal through its In-phase and Quadrature components, or their equivalents in higher dimensions. This chapter commences by restating the considered problem and stressing the intention on the estimation target in Section 5.1. In Sections 5.2 and 5.3, we describe the proposed stratified and antithetical stratified schemes separately, which we refer to as the IQ stratified scheme and the IQ antithetical stratified scheme, respectively. Also, we provide statistical analysis of the introduced estimates, and most importantly, we explain how the developed technique locates the acceleration frequency at the selected point in the frequency domain. In Section 5.4 we discuss the other mechanism of error reduction of the proposed schemes, and also we introduce the IQ total random estimation which can also leverage such technique. A numerical example is included to illustrate the performance of the proposed estimates.

5.1 The Targeted Fourier Transform of the IQ Estimation

The IQ estimation uses sampling points selected randomly from a dense Cartesian grid of uniformly distributed points. Unlike the traditional random on grid estimation, the density of the grid is uniquely determined by the value of the preselected acceleration frequency $\mathbf{f}_c \equiv \left(f_{c,1}, f_{c,2}, ..., f_{c,K} \right)$, and precisely equal to $4\mathbf{f}_c$ with density of $4 f_{c,k}$ along the *k*-th dimension. The form of the targeted Fourier transform is same of the on-grid schemes, i.e.

$$X_{wD}(\mathbf{f}) = \frac{D}{M} \sum_{m=1}^{M} x_w(\mathbf{t}_m) \exp\left(-j2\pi \mathbf{f}.\mathbf{t}_m\right) , \tag{5.1}$$

but indeed with a specific grid. Since we usually would move the acceleration frequency far from zero, the grid is dense and hence spectral analysis can be performed over a very wide range of frequencies with an extent that is independent of the employed (low) average sampling density.

5.2 IQ Stratified Estimation

After designing the Cartesian grid as we described above, in order to capture N samples of the signal, we divide the observation domain into $N/2^K$ adjacent subdomains $\{A_i\}_{i=1}^{N/2}$. Following the division of the domain, 2^K sampling points are selected from the grid points within each subdomain in the following way. First, we extract 2^K interlaced subgrids $\{G_q\}_{q=1}^{2^K}$ from the original grid, then we select one sampling point from each subgrid within each subdomain. That is, for one dimensional signal, one sampling point is selected randomly from the odd-order grid points, and another sampling point is selected randomly from the even-order grid points within each subdomain. Each sampling point is selected independently of the other sampling points inside or outside the subdomain. Fig. 5.1 demonstrates the sampling scheme for two-dimensional domains.

The Fourier transform (5.1) is estimated using the IQ stratified estimator:

$$\hat{X}_{IQS}(\mathbf{f}) := \sum_{i=1}^{N/2^K} v_i \hat{X}_{S,i}(\mathbf{f}) \tag{5.2}$$

where $v_i = M_i / 2^K M$ and

$$\hat{X}_{S,i}(\mathbf{f}) = D \sum_{q=1}^{2^K} \sum_{\mathbf{t}_m \in A_i, G_q} a_m x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m). \tag{5.3}$$

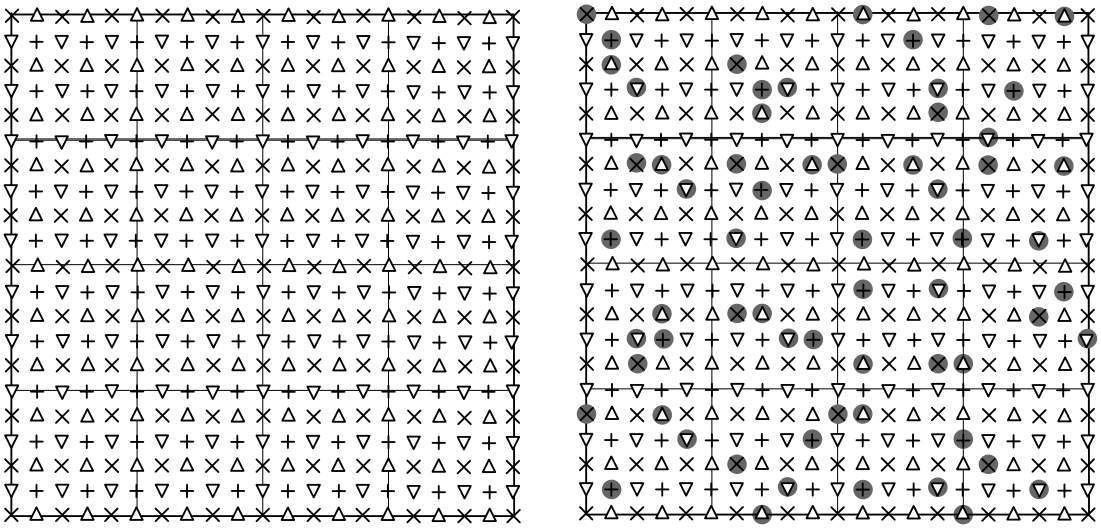

Fig 5.1. A two-dimensional observation window, with four interlaced subgrids labelled with (x, ∇, Δ, +) as grid points, and the shadowed points are the randomly sampling points.

The coefficients $\{a_m\}$, which are binary random variables that identify the associated selected/nonselected grid points $\{\mathbf{t}_m\}$ according to the described pattern, have their statistical characteristics described in Appendix C, where we also show that the expected value of $\hat{X}_{IQS}(\mathbf{f})$ satisfies:

$$E\left[\hat{X}_{IQS}(\mathbf{f})\right] = X_{wD}(\mathbf{f}). \tag{5.4}$$

This proves that $\hat{X}_{IQS}(\mathbf{f})$ is an unbiased estimator of the Fourier transform (5.1). The variance of the estimator at the analysed frequency $\mathbf{f}$ is described by (see Appendix C for the proof)

$$\mathrm{var}\left[\hat{X}_{IQS}(\mathbf{f})\right] = \frac{D^2}{M^2}\sum_{i=1}^{N/2}\left(\frac{M_i}{2^K}\sum_{\mathbf{t}_m\in A_n} x_w{}^2(\mathbf{t}_m) \;\; -\sum_{q=1}^{2^K}\left|\sum_{\mathbf{t}_m\in A_i, G_q} x_w(\mathbf{t}_m)\exp(-j2\pi\mathbf{f}.\mathbf{t}_m)\right|^2\right). \tag{5.5}$$

We notice from the expression above that the way the grid is partitioned, i.e. the sizes of the subgrids, has an influence on the accuracy of the constructed estimates. As we have mentioned before, designing the sizes of the subdomains to reduce the variance of the estimator requires *a priori* knowledge about the signal itself. Hence, a practical, neutral operation is to divide the grid into subdomains of equal sizes, resulting in an IQ stratified estimator of the following simple form:

$$\hat{X}_{IQS}(\mathbf{f}) = \frac{D}{N}\sum_{m=1}^{M} a_m x_w(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_m\right) . \tag{5.6}$$

Further discussion on the variance of the estimates is included in Section 5.4. Now, we demonstrate the rationale behind the devised technique and how it can locate the acceleration frequency at the selected point $\mathbf{f}_c$. For easy demonstration, we use one dimensional, equal partitioning estimator. We rewrite the IQ estimator as

$$\hat{X}_{IQS}(f) = \frac{D}{N}\sum_{\substack{m=1,\\ m \text{ odd}}}^{M} a_m x_w(t_m)\exp\left(-j2\pi f t_m\right) + \frac{D}{N}\sum_{\substack{m=2,\\ m \text{ even}}}^{M} a_m x_w(t_m)\exp\left(-j2\pi f t_m\right) . \tag{5.7}$$

We now express the signal $x(t)$ by its in-phase component $x_I(t)$ and its quadrature component $x_Q(t)$ as follows:

$$x(t) = x_I(t)\cos 2\pi f_c t - x_Q(t)\sin 2\pi f_c t . \tag{5.8}$$

According to the defined grid, the signal samples captured at the odd-order points are samples of the in-phase component of the signal:

$$x(t_m) = j^{m-1}x_I(t_m), \qquad m \text{ odd}, \tag{5.9}$$

whereas the even-order grid points capture samples of the quadrature component:

$$x(t_m) = j^{m}x_Q(t_m), \qquad m \text{ even}. \tag{5.10}$$

With $\exp\left(j2\pi f_c t_m\right) = j^{m-1}$, $m = 1,2,...,M$ and $f_v := f - f_c$, the IQ stratified estimator can be rewritten in the form

$$\hat{X}_{IQS}(f) = \frac{D}{N} \sum_{m=1,\ m\ \text{odd}}^{M} a_m x_I(t_m) w(t_m) \exp\left(-j2\pi f_v t_m\right)$$

$$+ j\frac{D}{N} \sum_{m=2,\ m\ \text{even}}^{M} a_m x_Q(t_m) w(t_m) \exp\left(-j2\pi f_v t_m\right). \qquad (5.11)$$

By examining the expression above we observe that the proposed estimator consists of two Fourier transform estimators of the in-phase and quadrature components. Each of these two estimators is a regular stratified on grid estimator, which estimates the Fourier transform of one of the components at f_v. Also, it is known that with a reasonably selected f_c, each component is a low-frequency signal with spectrum concentrated around $f_v = 0$. (This means that the IQ estimator $\hat{X}_{IQS}(f)$ estimates the Fourier transform (of arbitrarily high frequency signal) at f through estimating the Fourier transform of low-frequency components at the low values of f_v.) We conclude that the acceleration in the convergence rates of the Fourier transform estimates of these two components originates at $f_v = 0$, and it can spread to the frequencies where their spectral parts reside after collecting a low number of samples. This effectively means that the acceleration in the convergence rates of the IQ stratified estimates starts at $f = f_c$, and it can rapidly spread to the frequencies where the signal spectrum is situated for small numbers of samples. Likewise, the IQ stratified estimator in multiple dimensions with equal partitioning can be expressed as follows:

$$\hat{X}_{IQS}(\mathbf{f}) = \frac{D}{N} \sum_{q=1}^{2^K} j^{q-1} \sum_{\mathbf{t}_m \in G_q} a_m x_{c,q}(\mathbf{t}_m) w(\mathbf{t}_m) \exp\left(-j2\pi \mathbf{f}_v . \mathbf{t}_m\right) \qquad (5.11)$$

where $x_{c,q}(\mathbf{t})$ is the in-phase or quadrature components, or their equivalents in higher dimensions (the reader interested about the extension of the concept of the in-phase or quadrature components to multiple dimensions can refer to [79,80]). The expression (5.11) basically means that $\hat{X}_{IQS}(\mathbf{f})$ consist of 2^K regular stratified on grid estimators that approximate the Fourier transform of low frequency signals $\{x_{c,q}(\mathbf{t})\}_{q=1}^{2^K}$ at the low values of $\mathbf{f}_v := \mathbf{f} - \mathbf{f}_c$. Hence, similar conclusion can be conducted, i.e. the acceleration in the convergence rates in the multidimensional domain starts at $\mathbf{f} = \mathbf{f}_c$.

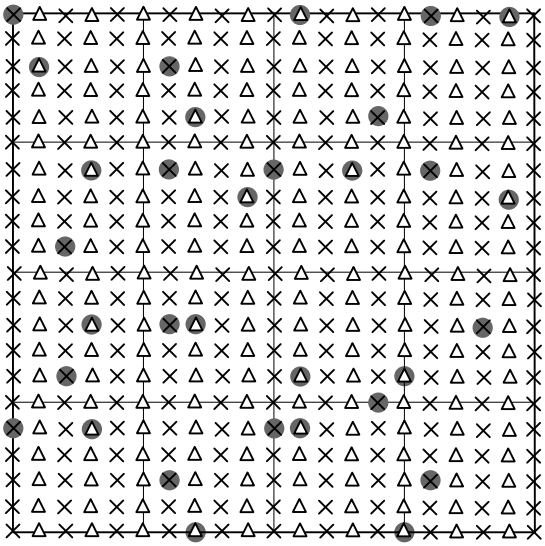

Fig 5.2. A two-dimensional observation window, with two interlaced subgrids labelled with (x, Δ) as grid points, and the shadowed points are the randomly selected sampling points.

Before we finish this section we note an important remark about selecting $\mathbf{f}_c$ For *K*-dimensional signals that vary merely or mainly along L dimensions, where $L < K$ (in other words, if the user would like to choose $\mathbf{f}_c \equiv \left(f_{c,1}, f_{c,2}, ..., f_{c,K} \right)$ such that $\left\| \mathbf{f}_c \right\|_0 = L$). For such case, the grid is designed in the same way described earlier along the "nonzero" dimension, i.e. with density of $4 f_{c,k}$ along the *k*-th dimension. Whereas, any point can be a sampling point along the "zero" dimension, or the sampling points can be selected from an arbitrarily dense grid (the grid should be dense enough to shift the replicas far from the region of interest). And, following the subdivision of the observation domain, only $\left\{ G_q \right\}_{q=1}^{2^L}$ are selected with each subdomain. An example of a two-dimensional domain where $\mathbf{f}_c = \left(f_{c,1}, 0 \right)$ is in the Figure 5.2.

5.3 IQ Antithetical Stratified Estimation

In this section we introduce and study the IQ antithetical stratified sampling and the associated estimates. This scheme divides the observation domain into $N/4$ non-overlapping subdomains $\left\{ A_i \right\}_{i=1}^{N/4}$ with $\left\{ M_i \right\}_{i=1}^{N/4}$ grid points inside. We extract 2^K interlaced subgrids $\left\{ G_q \right\}_{q=1}^{2^K}$ from the original grid. Each subdomain must contain an even number of each of the $\left\{ G_q \right\}_{q=1}^{2^K}$ grid points. Then, 2^{K+1} sampling points are selected within each subdomain. Half of these sampling points are selected in the same way as in the IQ stratified scheme, namely, one sampling point is randomly selected from each of $\left\{ G_q \right\}_{q=1}^{2^K}$ within each subdomain. The other half of the sampling points are selected as reflections of the random ones around the centre points of the subdomains with respect to each subgrid, denoted by

$\{\mathbf{c}_{i,q}\}_{q=1}^{2^K}$ (such that $\mathbf{c}_{i,q} = (\mathbf{t}' + \mathbf{t}'')/2$, where $\mathbf{t}'$ and $\mathbf{t}''$ are the closest and the furthest grid points from the origin that belong to the *q*th subgrid inside the *i*th subdomain). For example, in one dimensional case if the odd-order grid point $\mathbf{t}_m$ is randomly selected within the subdomain A_i, another sampling point will be assigned at the odd-order grid point of position $2\mathbf{c}_{i,1} - \mathbf{t}_m$, where $\mathbf{c}_{i,1}$ is the centre between the first and the last odd-order grid points in A_i. Similarly, for each randomly selected even grid point $\mathbf{t}_m$, another even grid point is allocated as a sampling point at $2\mathbf{c}_{i,2} - \mathbf{t}_m$, where $\mathbf{c}_{i,2}$ corresponds to the middle point between the first and the last even-order grid points in the *n*th subdomain. The Fourier transform (5.1) is estimated using the IQ antithetical stratified estimator described by

$$\hat{X}_{IQA}(\mathbf{f}) := \sum_{i=1}^{N/2^{k+1}} v_i \hat{X}_{A,i}(\mathbf{f}) \tag{5.12}$$

where $v_i = M_i / 2^{k+1} M$ and

$$\hat{X}_{A,i}(\mathbf{f}) = D \sum_{q=1}^{2^K} \sum_{\mathbf{t}_m \in A_i, G_q} a_m \left[x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m) + x_w(2\mathbf{c}_{i,q} - \mathbf{t}_m) \exp\left(-j2\pi \mathbf{f}.(2\mathbf{c}_{i,q} - \mathbf{t}_m)\right) \right]. \tag{5.13}$$

The PDFs and the dependency of these binary random coefficients $\{a_m\}$ are discussed in Appendix D. The estimator $\hat{X}_{IQA}(\mathbf{f})$ is an unbiased estimator of the discrete-time Fourier transform (5.1) (the proof is in Appendix D):

$$E\left[\hat{X}_{IQA}(\mathbf{f})\right] = X_{wD}(\mathbf{f}). \tag{5.14}$$

The variance of the estimates is (see Appendix D for the proof)

$$\operatorname{var}\left[\hat{X}_{IQA}(\mathbf{f})\right] = \frac{D^2}{M^2}\sum_{i=1}^{N/2^{K+1}}\left(\frac{M_i}{2^K}\sum_{\mathbf{t}_m\in A_i} x_w{}^2(\mathbf{t}_m)\right.$$

$$+\sum_{q=1}^{2^K}\left[\frac{M_i}{2^K}\sum_{\mathbf{t}_m\in A_i,G_q} x_w(\mathbf{t}_m)x_w(2\mathbf{c}_{i,k}-\mathbf{t}_m)\cos\left(2\pi f\left(2\mathbf{c}_{i,k}-2\mathbf{t}_m\right)\right)\right.$$

$$\left.\left.-\left|\sum_{\mathbf{t}_m\in A_i,G_q} x_w(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}.\mathbf{t}_m\right)\right|^2\right]\right). \qquad (5.15)$$

The way the grid is divided affects the variance, and as we discussed in the previous section, the most sensible practice is the uniform division. Further discussion about the derived variance is included in Section 5.4.

As with the IQ stratified estimator, the IQ antithetical stratified estimator comprises regular antithetical stratified estimators of the Fourier transform of the inphase-quadrature multidimensional equivalent components $\left\{x_{c,q}(\mathbf{t})\right\}_{q=1}^{2^K}$, and it can be expressed (for equal partitioning and uniform windowing for simplicity)

$$\hat{X}_{IQA}(\mathbf{f}) = \frac{D}{N}\sum_{q=1}^{2^K} j^{q-1}\sum_{\mathbf{t}_m\in G_q}\sum_{i=1}^{N/2^{k+1}} a_m\left[x_{c,q}(\mathbf{t}_m)\exp\left(-j2\pi\mathbf{f}_v.\mathbf{t}_m\right)\right.$$

$$\left.+x_{c,q}\left(2\mathbf{c}_{i,q}-\mathbf{t}_m\right)\exp\left(-j2\pi\mathbf{f}_v.\left(2\mathbf{c}_{i,q}-\mathbf{t}_m\right)\right)\right]. \qquad (5.16)$$

Hence, with a reasonably selected $\mathbf{f}_c$, the spectrum of these components concentrates around $\mathbf{f}_v := \mathbf{f}-\mathbf{f}_c = 0$, and hence the estimates $\hat{X}_{IQA}(\mathbf{f})$ at (arbitrarily high) frequency $\mathbf{f}$ are basically low-frequency estimates at $\mathbf{f}_v$. It follows from these facts that the acceleration in the convergence rate of $\hat{X}_{IQA}(\mathbf{f})$ starts at $\mathbf{f}_v = 0$, i.e. $\mathbf{f} = \mathbf{f}_c$, and it can spread rapidly to the frequencies where the signal spectral parts reside for small numbers of samples.

5.4 IQ Error Reduction Through Spectrum Symmetry and IQ Total Random Estimation

The main advantage of the IQ estimation is to provide a key solution to the stratified technique estimation to relocate the acceleration frequency to the most prominent points in the spectrum. As we mentioned in the introduction of this chapter, the IQ estimation can provide another mechanism of error reduction that is independent of facilitating the acceleration in the convergence rates. By carefully examining the expressions of the variances of the estimates derived in this thesis we notice that at a given frequency point, the magnitude of the Fourier transform affects the variance. The variance at a specific frequency $\mathbf{f}$ tends to be smaller if $\left|X_w(\mathbf{f})\right|$ is strong, as opposed to that if $\left|X_w(\mathbf{f})\right|$ is weak. Interestingly, the IQ estimation provides another advantage in this regard. All the estimation schemes leverage this technique, including total random scheme. Therefore, we introduce herein the IQ total random estimation. Indeed, there is no stratification with total random sampling. Instead, as we did with the previous schemes, we extract 2^K interlaced subgrids $\{G_q\}_{q=1}^{2^K}$ from the original grid. Then, $N/2^K$ sampling points are selected from each subgrid. The IQ total random estimator is given by

$$\hat{X}_{IQTR}(\mathbf{f}) := \frac{D}{N}\sum_{q=1}^{2^K}\sum_{\mathbf{t}_m \in G_q} a_m x_w(\mathbf{t}_m)\exp(-j2\pi\mathbf{f}.\mathbf{t}_m). \tag{5.17}$$

The random variables within the *q*-th subgrid have the same probability to be one/zero, given by (assuming/designing the number of the grid points of each subgrid to be identical)

$$P[a_m = 1] = \frac{2^K}{M},$$

$$P[a_m = 0] = \frac{M - 2^K}{M}\ . \tag{5.18}$$

These random variables are dependent, but they are independent of all the other random variables outside their subgrids. Indeed, the estimator $\hat{X}_{IQTR}(\mathbf{f})$ is an unbiased estimator of the discrete-time Fourier transform (5.1):

$$E\left[\hat{X}_{IQTR}(\mathbf{f})\right] = X_{wD}(\mathbf{f}). \tag{5.19}$$

The variance of the estimates is

$$\mathrm{var}\left[\hat{X}_{IQTR}(\mathbf{f})\right] = \frac{D^2}{M^2}\left(\sum_{m=1}^{M} x_w{}^2(\mathbf{t}_m) - \frac{M}{4}\sum_{k=1}^{2^K}\left|\sum_{\mathbf{t}_m \in G_k} x_w(\mathbf{t}_m)\exp(-j2\pi\mathbf{f}.\mathbf{t}_m)\right|^2\right). \tag{5.20}$$

We remarked earlier that the IQ estimates consist of regular estimates of the Fourier transform of the in-phase and quadrature components (or their equivalents in multidimensional domains). As a result, the variance of the IQ estimators at a frequency point is influenced by the magnitude of the Fourier transform of these components (this remark can also be clearly conducted from (5.20)). As we know, the in-phase and quadrature components (and their equivalents in multidimensional domains) are real signals with symmetrical Fourier transform, in magnitude, around $\mathbf{f}_v = 0$ (i.e. $\mathbf{f} = \mathbf{f}_c$). Accordingly, weak parts of the estimated Fourier transform (at $\mathbf{f}_c - \boldsymbol{\upsilon}$, where $\boldsymbol{\upsilon}$ is a real-valued vector) could be the result of estimating strong spectral parts of its in-phase and quadrature components (if there is a strong spectral part of the signal residing at $\mathbf{f}_c + \boldsymbol{\upsilon}$). We illustrate this phenomenon when it is most beneficial with the following simple example: consider a one dimensional signal with relatively strong and weak spectral elements, as shown in the upper plot of Fig. 5.3. The variances of the regular and IQ total random schemes are shown at the lower plot of Fig. 5.3. With the IQ total random estimation which does not undergo acceleration in the convergence rate, this phenomenon is manifested. According to the regular approach, the variance at the weak part is high compared with the variance at the strong part. Whereas, with the IQ estimation $f_c = 10^3$ Hz, the variance at the weak part is as relatively low as the

variance at the strong part owing to the spectral shapes of the in-phase and quadrature components of the signal. We remark that the positive effect of the strong parts on the variance spreads to their adjacent frequencies for a higher number of samples, especially for the stratified-based estimation schemes

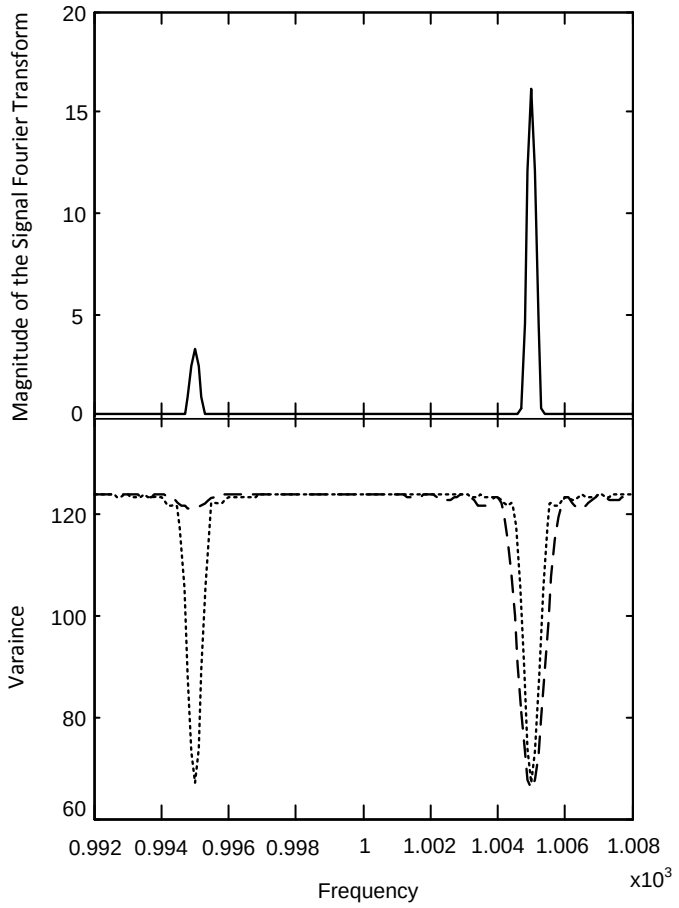

Fig. 5.3. The magnitude of the Fourier transform of the example signal in the upper plot. The variance of the regular total random estimator (dashed line) and the IQ total random estimator (dotted line) are shown in the lower plot.

Indeed, not every signal leverages the described property of the IQ estimation in such a way. However, for any other signal, by selecting $\mathbf{f}_c$ in the middle of the frequency range where the user intends to search and estimate the signal spectrum, the IQ estimation exploits the strong spectral parts of the signal to reduce the error over more of the pertinent frequency points, where the signal spectrum is relatively weak or there is no signal activity. We remind the reader that this reduction in the error of the IQ estimates is independent of and supplementary to the main reduction caused by the facilitated acceleration of their convergence rates for the (antithetical) stratified estimation schemes.

5.5 Numerical Illustration

Here we present a numerical example to demonstrate the efficiency of the introduced technique. We compare the proposed IQ methods with the regular on-grid approaches and show how the proposed methods outperform their counterparts. For clear illustration of the estimation errors against the frequency, we consider a simple one-dimensional test signal:

$$x(t) = 2B\text{sinc}(2Bt)\cos\left(2\pi \times 10^9 t\right). \tag{5.20}$$

The signal is centred at 10^9 Hz with $B = 10^6$ Hz. The targeted Fourier transform is set for the signal observed over $D = 10^{-6}$s and averaged with a Blackman window.

Designing the grid, i.e. selecting f_c, is a key factor for the IQ methods as it represents the frequency where the acceleration starts. Certainly, we would like to bring the advantages of fast convergence to the frequency points where the signal spectrum is positioned. Therefore, we would ideally like to choose f_c in the middle of the signal spectrum so the acceleration starts at and rapidly spreads to the signal's spectral parts. However, it is not always the case that the location of the centre frequency of the signal spectrum is known in advance.

Nonetheless, as we will show here, choosing f_c in the locality of the signal spectrum can still facilitate the acceleration in the convergence rates at the pertinent frequencies for a low number of samples, in contrast to the regular approaches which have an acceleration that starts at DC and does not spread to these high frequencies unless a considerably large number of samples are collected. Throughout this example, we use equal partitioning in all the schemes. The results in this numerical example are obtained by computing the derived expressions of the variances of the IQ estimation schemes, (5.5) and (5.15), and the exact expressions of the regular on grid approaches, i.e. (4.5) and (4.16).

In Fig. 5.4, we show the mean-square estimation error of the regular stratified on-grid scheme and the IQ stratified scheme with various values of f_c, where all the estimators use the same number of samples, i.e. $N = 20$. The regular random schemes use the same grid of the IQ ones. The error is displayed for a frequency range of size $4B$ which covers the signal spectrum. It is evident how the IQ estimators outperform their counterpart. To manifest the convergence rates of the estimates and the acceleration they undergo, we plot the mean-square estimation error at a frequency point in the signal spectrum versus the number of random samples. In Fig. 5.5, we demonstrate the mean-square error of the IQ and regular stratified on-grid estimation at $f = 10^9\,\text{Hz}$ (for the IQ method, f_c is chosen to be $1.001\times10^9\,\text{Hz}$). The thin lines in Fig. 5.5 represent reference curves with the labelled order of decay with respect to the number of samples *N*. Since the plots are on a log-log scale, these reference curves are lines of slopes equal to the order of convergence. We observe that the mean-square error of the regular method shows a rate of decay of $1/N$ and does not accelerate for the shown numbers of samples, whilst the IQ stratified estimates render the fast convergence rate of $1/N^3$ for small numbers of samples.

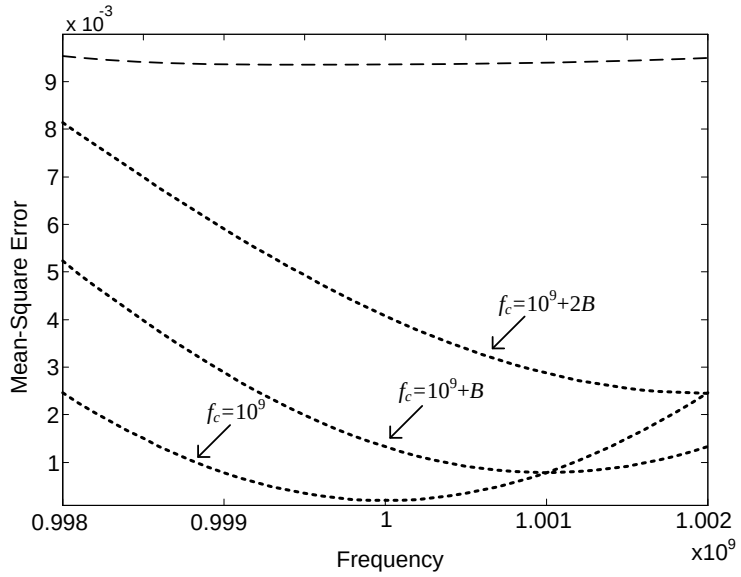

Fig. 5.4. Mean-square estimation error of the regular stratified scheme (dashed line) and the IQ stratified scheme (dotted lines) with different f_c using $N = 20$.

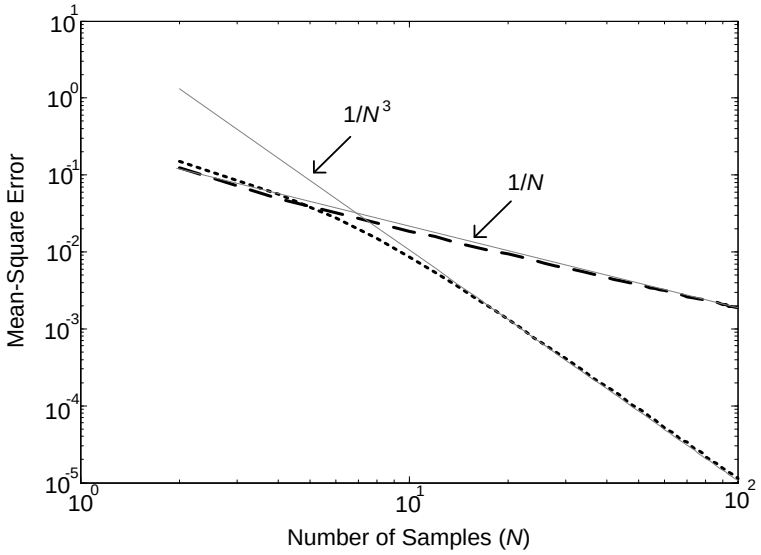

Fig. 5.5. Mean-square estimation error of the regular stratified scheme (dashed line) and the IQ stratified scheme (dotted line) at frequency point $f = 10^9$Hz. The fine lines are reference curves with the labelled order of decay.

Fig. 5.6 demonstrates the mean-square error in the estimated Fourier transform of the IQ and regular antithetical stratified schemes with $N = 40$, for a frequency range that covers the signal spectrum. We show the estimation error for the IQ estimates with various f_c, and observe that the outperformance of the IQ method is prominent. In Fig. 5.7, we show the mean- square estimation error of the IQ antithetical scheme ($f_c = 10^9$ Hz) and the regular antithetical approach at frequency $f = 0.999 \times 10^9$ Hz versus the number of samples *N*. We discern that the decay rate in the mean-square error of the regular antithetical stratified estimates is $1/N$ for the shown numbers of samples, whereas the IQ antithetical stratified estimates show a rate of $1/N^5$ for small values of *N*.

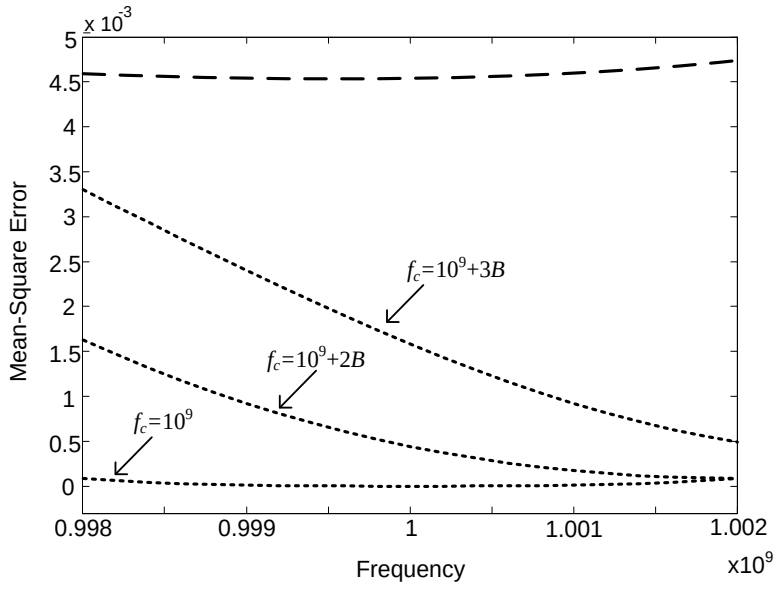

Fig. 5.6. Mean-square error of the regular antithetical stratified estimator (dashed line) and the IQ antithetical stratified estimator (dotted lines) with various f_c for $N = 40$.

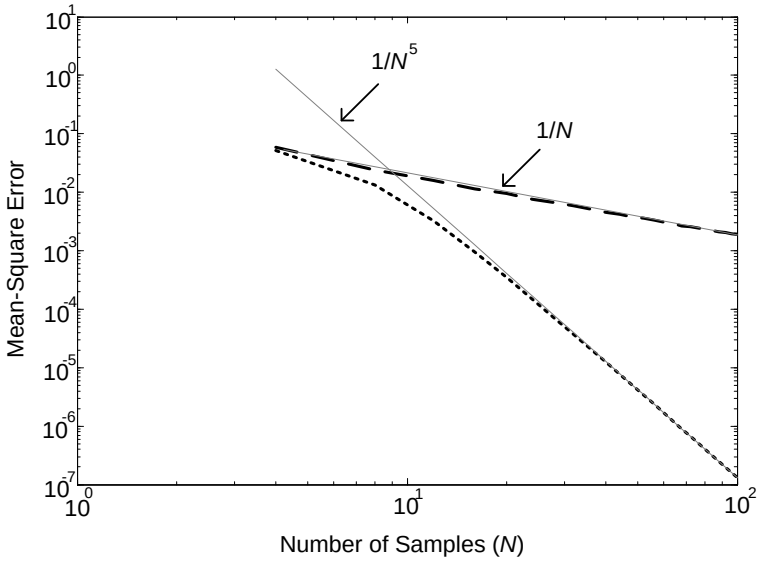

Fig. 5.7. Mean-square error of the regular antithetical stratified estimator (dashed line) and the IQ antithetical stratified estimator (dotted line) at frequency point $f = 0.999 \times 10^9 \text{Hz}$. The thin lines are reference curves with the labelled order of decay.

5.6 Chapter Summary

The developed IQ Fourier transform estimation can facilitate fast convergence rates using low sampling densities regardless of the positioning of the signal spectrum. By appropriately distributing the sampling points with no extra complications in the procedure of acquiring/processing the samples, the IQ estimates have convergence rates that start accelerating at a frequency point $\mathbf{f}_c$ that is selected by the user. Choosing $\mathbf{f}_c$ in the locality of the signal spectrum can lead to a substantial reduction in the estimation error or saving on the number of samples, subject to the system requirements, especially in low-dimensional domains. The proposed methods provide great advantages over the existing Fourier transform estimation approaches that engage (antithetical) stratified sampling, which have

convergence rates that start accelerating at DC and, hence, require high numbers of samples to generate estimates with a satisfactory error if the signal spectrum is distant from zero.

The IQ estimation constitutes a smart means to tackle the Fourier transform of high-frequency signals with no requirement for high sampling rates or/and downconversion-filtering the analogue signals (if accessible) prior to sampling. In particular, the devised approach is an excellent tool for processing signals with energy concentrated around some high frequency and have undesired spectral tails that spread over long frequency ranges. The reasons for stressing the former, besides the fact that these tails usually appear in the Fourier transform of truncated signals, are as follows. Many other "competing" methods (which were categorised under the periodic nonuniform sampling family in Chapter 2) are sensitive to signals with long tails. The other reason is that these tails do not carry a lot of the energy of the signal and hence they do not negatively affect the acceleration in the convergence rates at the prominent frequencies. As we showed in Section 3.7 for the regular estimators, the IQ estimators analogously deliver unbiased estimation with the presence of additive error (its characteristics is included in Section 3.7). And, the variances are equal to the ones without error plus a term that depends on the size of observation domain and the variance of error, which decays at the rate of $1/N$, see (3.58) and (3.68).

CHAPTER SIX

CONCLUDING REMARKS

In this thesis we considered Fourier transform calculation of multidimensional signals using random sampling. The key advantage of using random sampling here, over other sampling schemes, is its capability to avoid aliasing. Various categories of random sampling have been examined here, and throughout this thesis we have sought to highlight their performances. All the categories deliver a performance that is lower bounded by the performance of total random sampling scheme. That is, the error in the calculated Fourier transform is independent of the location of the signal spectrum in the frequency domain. If the signal has certain smoothness, the performance can be improved with some of the studied techniques. We showed that all these techniques lose their effectiveness with increasing the dimensionality of the signal as well as the distance of its spectrum to the acceleration frequency. The IQ estimation widens the class of the signals that can capitalise upon these techniques by creating the unique chance of relocating the acceleration frequency.

After the thorough analysis of the performances of the several techniques studied here, we conclude that there is no one optimal scheme. Each of the studied schemes can be in a certain scenario the scheme of choice in terms of the quality of the Fourier transform it yields as well as the way of implementing its distribution of the sampling points. As examples for the latter, stratified sampling gives a good prospect of building two, or more, interlaced relatively slow analogue-to-digital converters, which makes it rather practical in real time applications of one-dimensional signals as opposed to total random sampling. And,

total random sampling is effortless to deploy in field estimation. Also, adding more sampling points can be performed efficiently as opposed to stratified-based estimation. And, so on.

The random sampling estimation of the Fourier transform can be used for various signal processing objectives. These methods can/have been used for spectral analysis via reduced sampling density in multidimensional NMR spectroscopy, which is nowadays one of the most efficient spectroscopic techniques, providing insight into molecular structure and dynamics, and it has opened an avenue to a large number of new emerging applications in chemistry and biochemistry. Random sampling allows performing higher dimensional experiments (i.e. increasing the dimensionality of a multidimensional experiment) that are not conventionally obtainable. In many cases increasing the dimensionality of an experiment can be very interesting from a scientific point of view as it reveals more effects by spreading the spectral components over further dimensions. This approach is referred to by the magnetic resonance community as “Accurate NMR” [6].

Many interesting methods can be applied when NMR is considered in general. However, many of these methods usually engage heavy computations which make them intractable for the high dimensions and great amount of data that appear in “Accurate NMR” applications. For such applications, the simple data acquisition and processing algorithms of random sampling estimation are more suitable [6]. These random sampling approaches leverage the reasonable computational requirement of a discrete-time Fourier transform type of calculations. The other interesting characteristic of these approaches is that the dimensionality of the signal or the size of the observation window do not dedicate the required sampling density. These features qualify the random sampling approaches to be the methods of choice in the “Accurate NMR” applications.

Higher Order Approximation

If the signal has a *p*-order (where $p > 2$) continuous derivative, it is possible to construct a higher convergence approximation of the Fourier transform, compared with the studied approaches. This can be achieved by employing more sophisticated estimation schemes (based on the Newton-Cotes formulas). In fact, there are groups of numerical integration methods that can be employed in principles to approximate the Fourier transform with high order convergence rates. Examples are Newton-Cotes methods (such as the Simpson rule) that are based on equidistant evaluation points, and Gauss methods which use unequal-spaced points. Although, these methods can theoretically provide fast convergence rate, it is safe to say that they are not useful for our purpose. The reason for that is that they provide "wrong" results when using relatively low number of samples. And, they can only provide acceptable error when the sampling density is higher than the corresponding Nyquist density. This can be anticipated from their mechanism of approximating the integration using high order polynomials, which is not appropriate for the highly oscillated function that we deal with here.

We remind the reader about the remarkable following fact about the studied random sampling schemes that provide improved results over total random sampling. Although we showed that these methods can lose their effectiveness depending on several factors, the loss means that these methods perform no better than total random sampling and no less.

Future Work

- Spectrum sensing entails scanning predefined parts of the spectrum in search of meaningful activity such as an ongoing transmission or the occurrence of an event. Spectrum sensing has a vast range of application areas including astronomy, seismology, communication systems and many others. In addition, the conceptualised Cognitive Radio technology has recently revived and intensified the research into new effective sensing techniques. The majority of these DSP-based methods employ uniform sampling. The remarkably high sampling rates imposed by classical DSP (especially for wideband radio-frequency applications) can be an obstacle for systems whose resources (such as power, complexity, and size) are limited. Approaches based on random sampling can accomplish the sensing task using considerably low sampling rates. In fact, since the spectrum sensing task does not seek the exact shape of the signal spectrum, the sampling rates can be significantly reduced. Most importantly, general guidelines can be developed based on the statistical properties of these estimates to ensure that the approaches satisfy certain detection probabilities predefined by the user. Some of the studied schemes have been employed and rigorously analysed for this purpose in [66-68,77,78]. However, other schemes, such as the IQ estimation schemes, can be considered and capitalised upon in spectrum sensing. And, efficient spectrum sensing algorithms are expected to be devised based on these advanced schemes.
- Fluctuation of sampling instants is a common occurrence in all sorts of sampling. This practical limitation has been profoundly addressed in literature for uniform sampling. Hence, it can be an important aspect to analyse the performance of all the random sampling estimation schemes studied here with the presence of this limitation, and study the possibility of modifying their features to relax its effect.

- In spite of the wide research reported in the nonuniform sampling literature, not many papers address the sampling pattern design problem. It is rather understandable as many of the processing algorithms developed in this field can cope with arbitrary sampling schemes (or there is not an analytical way of associating their performance with the sampling pattern to optimise the design of the latter). In addition, the problem of arbitrarily distributed data is wider; in many applications the user has no control over the location of the sampling points. However, as we have seen, this thesis focuses on the distribution of the sampling points. And, it is worth to investigate the performance of these schemes with other processing tools (apart from the tailored forms of DTFT) and explore their advantages. Also, in the various studied schemes we have shown some mechanisms of improving the quality of the estimated Fourier transform by investing some knowledge about the signal through adapting the sampling process. This can be achieved by, for example, re-designing the PDF of the sampling points or re-locating the acceleration frequency to the most important part of the signal spectrum in the IQ case. Hence, a promising future research is to develop a sort of adaptive sampling procedure to provide a reduced-error Fourier transform estimation.

APPENDIX A

PROOFS FOR STRATIFIED ESTIMATION

We find the asymptotic expression of the variance of the stratified estimator with maximum stratification for a *K*-dimensional signal with a continuous derivative along each dimension. We let

$$g(\mathbf{t}) = x_w(\mathbf{t})\exp(-j2\pi\mathbf{f}.\mathbf{t}). \tag{A.1}$$

With $\mathbf{t}_i \quad i = 1, 2, \ldots, N$ random sampling points, we define a new random variable within each subdomain

$$Z_i = \frac{g(\mathbf{t}_i)}{p(\mathbf{t}_i)} v_i. \tag{A.2}$$

Taylor expansion of $g(\mathbf{t})\ (\mathbf{t} \in A_i)$ around the centre of the subdomain $\mathbf{c}_i \equiv (c_{i,1}, c_{i,2}, ..., c_{i,K})$, assuming that $g(\mathbf{t})$ has a continuous derivative along each dimension, is

$$g(\mathbf{t}) = g(\mathbf{c}_i) + \sum_{d=1}^{K} \frac{\partial g(\mathbf{c}_i)}{\partial t_d}(t_d - c_{i,d}) + R \tag{A.3}$$

where $R = \sum_{d=1}^{K} o(|t_d - c_{i,d}|)$. The expected value of Z_i is

$$E[Z_i] = \int_{A_i} g(\mathbf{t}) d\mathbf{t}. \tag{A.4}$$

First, we note that

$$\int_{A_i} (t_d - c_{i,d})^p \, d\mathbf{t} = \begin{cases} 0 & ,p \text{ odd} \\ \dfrac{(\tau_{i,d})^p \prod_{k=1}^{K} \tau_{i,k}}{2^p (p+1)} & ,p \text{ even} \end{cases} \tag{A.5}$$

and

$$\int_{A_i} |t_d - c_{i,d}|^p \, d\mathbf{t} = \frac{(\tau_{i,d})^p \prod_{k=1}^{K} \tau_{i,k}}{2^p (p+1)} \tag{A.6}$$

where $\tau_{i,d}$ and $\tau_{i,k}$ are the lengths of the subdomain A_i along the dimension d and k respectively. Hence,

$$E[Z_i] = g(\mathbf{c}_i) \prod_{d=1}^{K} \tau_{i,d} + \int_{A_i} R \, d\mathbf{t} \tag{A.7}$$

where $\int_{A_i} R \, d\mathbf{t} = \sum_{d=1}^{K} o\left(\tau_{i,d} \prod_{k=1}^{K} \tau_{i,k} \right)$ and

$$|E[Z_i]|^2 = |g(\mathbf{c_i})|^2 \prod_{d=1}^{K} (\tau_{i,d})^2 + \sum_{d=1}^{K} \sum_{r=1}^{K} o\left(\tau_{i,d} \tau_{i,r} \prod_{k=1}^{K} (\tau_{i,k})^2 \right) + U_1 \tag{A.8}$$

where $U_1 = 2\Re\left[g(\mathbf{c}_i) \prod_{d=1}^{K} \tau_{i,d} \int_{A_i} R \, d\mathbf{t} \right]$. The expected value of $|Z_i|^2$ is

$$E\left[|Z_i|^2 \right] = \int_{A_i} \frac{|g(\mathbf{t})|^2}{p(\mathbf{t})} v_i d\mathbf{t} \, . \tag{A.9}$$

Since $|g(\mathbf{t})|^2$ and $p_i(\mathbf{t})$ are continuous, according to the mean value theorem for integration

$$E\left[|Z_i|^2 \right] = \frac{v_i}{p(\mathbf{c}_i')} \int_{A_i} |g(\mathbf{t})|^2 \, d\mathbf{t} \tag{A.10}$$

where $\mathbf{c}_i$' is a point in A_i. For now, we assume that $p(\mathbf{t})$ is uniform. This assumption will simplify the derivation and not affect the final expression where $p(\mathbf{t})$ can be any permissible density function. Hence,

$$E\left[|Z_i|^2\right] = \prod_{d=1}^{K} \tau_{i,d} \int_{A_i} |g(\mathbf{t})|^2 \, d\mathbf{t}$$

$$= |g(\mathbf{c_i})|^2 \prod_{d=1}^{K} (\tau_{i,d})^2 + \frac{1}{12} \sum_{d=1}^{K} \left| \frac{\partial g(\mathbf{c}_i)}{\partial t_d} \right|^2 (\tau_{i,d})^2 \prod_{k=1}^{K} (\tau_{i,k})^2$$

$$+ \sum_{d=1}^{K} \sum_{r=1}^{K} o\left(\tau_{i,d} \tau_{i,r} \prod_{k=1}^{K} (\tau_{i,k})^2 \right) + U_2 \qquad \text{(A.11)}$$

where $U_2 = 2\Re\left[g(\mathbf{c}_i) \prod_{d=1}^{K} \tau_{i,d} \int_{A_i} R \, d\mathbf{t} \right]$. It is seen that $U_2 = U_1$. Using (A.8) and (A.11), the variance of Z_i is

$$\text{var}[Z_i] = E\left[|Z_i|^2\right] - \left|E[Z_i]\right|^2$$

$$= \frac{1}{12} \sum_{d=1}^{K} \left| \frac{\partial g(\mathbf{c}_i)}{\partial t_d} \right|^2 (\tau_{i,d})^2 \prod_{k=1}^{K} (\tau_{i,k})^2 + \sum_{d=1}^{K} \sum_{r=1}^{K} o\left(\tau_{i,d} \tau_{i,r} \prod_{k=1}^{K} (\tau_{i,k})^2 \right). \qquad \text{(A.12)}$$

The variance of the stratified estimator is

$$\text{var}\left[\hat{X}_{wS}(\mathbf{f}) \right] = \sum_{i=1}^{N} \text{var}[Z_i]$$

$$= \frac{1}{12} \sum_{i=1}^{N} \sum_{d=1}^{K} \left| \frac{\partial g(\mathbf{c}_i)}{\partial t_d} \right|^2 (\tau_{i,d})^2 \prod_{k=1}^{K} (\tau_{i,k})^2 + \sum_{i=1}^{N} \sum_{d=1}^{K} \sum_{r=1}^{K} o\left(\tau_{i,d} \tau_{i,r} \prod_{k=1}^{K} (\tau_{i,k})^2 \right). \qquad \text{(A.13)}$$

For any continuous positive $p(\mathbf{t}) = \prod_{d=1}^{K} p_d(t_d)$ with the observation domain divided equally along each dimension, i.e. having the same number of subdomains along each dimension, the following holds:

$$\int_{A_{i,d}} p_d(t_d)\,dt_d = \left[\frac{1}{N}\right]^{\frac{1}{K}} \tag{A.14}$$

where $\int_{A_{i,d}} dt_d$ is the integration along dimension *d* of the subdomain A_i with respect to t_d .

Since $p_d(t_d)$ is continuous, using the mean value theorem for integral

$$\int_{A_{i,d}} p_d(t_d)\,dt_d = p_d\left(c'_{i,d}\right)\tau_{i,d} \tag{A.15}$$

where $c'_{i,d} \in A_{i,d} \equiv \{t_d : t_d \in A_i\}$. Thus

$$\tau_{i,d} = \left[\frac{1}{N}\right]^{\frac{1}{K}} \times \frac{1}{p_d\left(c'_{i,d}\right)}. \tag{A.16}$$

Hence,

$$\mathrm{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \frac{1}{12N^{1+\frac{2}{K}}} \sum_{i=1}^{N}\sum_{d=1}^{K} \frac{\left|\frac{\partial g(\mathbf{c}_i)}{\partial t_d}\right|^2 \prod_{k=1}^{K}\tau_{i,k}}{{p_d}^2\left(c'_{i,d}\right)\prod_{k=1}^{K} p_k\left(c'_{i,d}\right)} + o\left(\frac{1}{N^{1+\frac{2}{K}}}\right). \tag{A.17}$$

Using Riemann integration, we find

$$\lim_{N\to\infty} N^{1+\frac{2}{K}} \operatorname{var}\left[\hat{X}_{wS}(\mathbf{f})\right] = \sum_{d=1}^{K} \int_{A} \frac{\left|\frac{\partial g(\mathbf{t})}{\partial t_d}\right|^2}{p_d{}^2(t_d)\prod_{k=1}^{K} p_k(t_k)} d\mathbf{t}. \tag{A.18}$$

Now, we establish the joint asymptotic normality of the real and imaginary parts of the stratified estimate. Since these parts are the sum of N independent random variables which are not identically distributed, the classical central limit theorem does not apply. Instead, we prove that the Lyapunov condition [74] is satisfied. For this purpose we use the Cramér-Wold theorem which allows proving multivariate convergence in distribution by establishing univariate convergence in distribution for arbitrary linear combinations of the vector components [75]. Accordingly, we re-define $g(\mathbf{t})$ as

$$g(\mathbf{t}) := \lambda_1 g_R(\mathbf{t}) + \lambda_2 g_I(\mathbf{t}) \tag{A.19}$$

where λ_1 and λ_2 are arbitrary real numbers, and $g_R(\mathbf{t})$ and $g_I(\mathbf{t})$ are defined in (3.9) and (3.10). If for some $r > 2$, the Lyapunov condition

$$\lim_{N\to\infty} L_N = \lim_{N\to\infty} \frac{\sum_{i=1}^{N} E\left[\left|Z_i - E[Z_i]\right|^r\right]}{\left[\sum_{i=1}^{N} \operatorname{var}[Z_i]\right]^{r/2}} = 0 \tag{A.20}$$

is fulfilled, the random variable

$$S = \sum_{i=1}^{N} \left[Z_i - E[Z_i]\right] \tag{A.21}$$

converges in distribution to normal random variable as N goes to infinity. Now, we determine the behaviour of the nominator and dominator of L_N as N goes to infinity. Calling (A.7) we have

$$Z_i - E[Z_i] = \left[g(\mathbf{t}_i) - g(\mathbf{c}_i) \right] \prod_{d=1}^{K} (\tau_{i,d}) + \sum_{d=1}^{K} o\left(\tau_{i,d} \prod_{k=1}^{K} \tau_{i,k} \right). \tag{A.22}$$

By using Hölder's inequality, each element of the nominator follows

$$E\left[\left| Z_i - E[Z_i] \right|^r \right] \le 2^{r-1} \left(E\left[\left| g(\mathbf{t}_i) - g(\mathbf{c}_i) \right|^r \right] \prod_{d=1}^{K} (\tau_{i,d})^r + \left| \sum_{d=1}^{K} o\left(\tau_{i,d} \prod_{k=1}^{K} \tau_{i,k} \right) \right|^r \right). \tag{A.23}$$

The expectation of the first term on the right side of the inequality

$$E\left[\left| g(\mathbf{t}_i) - g(\mathbf{c}_i) \right|^r \right] = \frac{1}{\prod_{d=1}^{K} \tau_{i,d}} \int_{A_i} \left| g(\mathbf{t}) - g(\mathbf{c}_i) \right|^r d\mathbf{t}$$

$$= \frac{1}{\prod_{d=1}^{K} \tau_{i,d}} \int_{A_i} \left| \sum_{d=1}^{K} \frac{\partial g(\mathbf{c}_i)}{\partial t_d} (t_d - c_{i,d}) \right|^r + \left| \sum_{d=1}^{K} o\left(\left| t_d - c_{i,d} \right| \right) \right|^r d\mathbf{t} . \tag{A.24}$$

Using (A.5), (A.6) and Hölder's inequality,

$$E\left[\left| g(\mathbf{t}_i) - g(\mathbf{c}_i) \right|^r \right] \le \frac{1}{\prod_{d=1}^{K} \tau_{i,d}} \left[\frac{K^{r-1}}{2^r (r+1)} \sum_{d=1}^{K} \left| \frac{\partial g(\mathbf{c}_i)}{\partial t_d} \right|^r (\tau_{i,d})^r \prod_{k=1}^{K} (\tau_{i,k}) + \sum_{d=1}^{K} o\left((\tau_{i,d})^r \prod_{k=1}^{K} \tau_{i,k} \right) \right].$$

(A.25)

Hence, with the aid of Hölder's inequality one more time with the second term of (A.23) on the right hand side, we have

$$E\left[\left| Z_i - E[Z_i] \right|^r \right] \le \left(\frac{K^{r-1}}{2(r+1)} \sum_{d=1}^{K} \left| \frac{\partial g(\mathbf{c}_i)}{\partial t_d} \right|^r (\tau_{i,d})^r \prod_{d=1}^{K} (\tau_{i,d})^r + \sum_{d=1}^{K} o\left((\tau_{i,d})^r \prod_{k=1}^{K} (\tau_{i,k})^r \right) \right).$$

(A.27)

Using (A.16),

$$E\left[\left|Z_i - E\left[Z_i\right]\right|^r\right] \le \frac{C}{N^{r+r/K}} \tag{A.28}$$

where C is a positive constant, thus

$$\sum_{n=1}^{N} E\left[\left|Z_i - E\left[Z_i\right]\right|^r\right] \le \frac{C}{N^{r-1+r/K}}\,. \tag{A.29}$$

Subsequently,

$$L_N \le \frac{C}{N^{r/2-1}\left(N^{1+2/K}\sum_{i=1}^{N}\mathrm{var}\left[Z_i\right]\right)^{r/2}} \tag{A.30}$$

calling (A.18), (A.21) is proven for $r > 2$. The asymptotic covariance matrix follows by identifying the coefficients of $\lambda_1^{\,2}$, $\lambda_2^{\,2}$ and $\lambda_1\lambda_2$ in the right hand side of (A.18) with $g\left(\mathbf{t}\right)$ of definition (A.19).

APPENDIX B

PROOFS FOR ANTITHETICAL STRATIFIED ESTIMATION

Here, we calculate the asymptotic expression of the variance of the antithetical stratified estimator with full stratification for *K*-dimensional signals with a continuous second-order derivative along each dimension. Young's form of the Taylor expansion of $g(\mathbf{t}) := x(\mathbf{t})w(\mathbf{t})\exp(-j2\pi t\mathbf{f}.\mathbf{t}) \quad (\mathbf{t} \in A_i)$ around the centre of the subdomain $\mathbf{c}_i \equiv (c_{i,1}, c_{i,2}, ..., c_{i,K})$, assuming that $g(\mathbf{t})$ has a continuous second-order derivative along each dimension is

$$g(\mathbf{t}) = g(\mathbf{c}_i) + \sum_{d=1}^{K} \frac{\partial g(\mathbf{c}_i)}{\partial t_d}(t_d - c_{i,d}) + \frac{1}{2}\sum_{d=1}^{K} \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d{}^2}(t_d - c_{i,d})^2$$

$$+ \frac{1}{2}\sum_{d=1}^{K}\sum_{\substack{r=1 \\ r\neq d}}^{K} \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d \partial t_r}(t_d - c_{i,d})(t_r - c_{i,r}) + \sum_{d=1}^{K} o\left(|t_d - c_{i,d}|^2\right). \qquad \text{(B.1)}$$

With $g(\mathbf{t})$ and $\mathbf{t}_i \quad i = 1, 2, \ldots, N/2$ random sampling points, we define a random variable Z_i $i = 1, 2, \ldots, N/2$ as follows

$$Z_i = \left(\frac{g(\mathbf{t}_i) + g(2\mathbf{c}_i - \mathbf{t}_i)}{p(\mathbf{t}_i)} \right) \frac{v_i}{2}. \qquad \text{(B.2)}$$

The expected value of Z_i is

$$E[Z_i] = \int_{A_i} \frac{g(\mathbf{t}) + g(2\mathbf{c}_i - \mathbf{t})}{2} d\mathbf{t}$$

$$= \int_{A_i} \left(g(\mathbf{c}_i) + \frac{1}{2}\sum_{d=1}^{K} \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d{}^2} \left(t_d - c_{i,d}\right)^2 + \frac{1}{2}\sum_{d=1}^{K}\sum_{\substack{r=1 \\ r\neq d}}^{K} \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d \partial t_r}^2 \left(t_d - c_{i,d}\right)\left(t_r - c_{i,r}\right) + R \right) d\mathbf{t}$$

(B.3)

where $R = \sum_{d=1}^{K} o\left(\left|t_d - c_{i,d}\right|^2\right)$. Using (A.5) and (A.6) we have

$$E[Z_i] = g(\mathbf{c}_i) \prod_{d=1}^{K} \tau_{i,d} + \frac{1}{24} \sum_{d=1}^{K} \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d{}^2} (\tau_{i,d})^2 \prod_{k=1}^{K} (\tau_{i,k}) + \int_{A_i} R \, d\mathbf{t} \tag{B.4}$$

where $\int_{A_i} R \, d\mathbf{t} = \sum_{d=1}^{K} o\left((\tau_{i,d})^2 \prod_{k=1}^{K} \tau_{i,k} \right)$. Hence,

$$\left|E[Z_i]\right|^2 = |g(\mathbf{c}_i)|^2 \prod_{d=1}^{K} (\tau_{i,d})^2 + \frac{1}{576}\sum_{d=1}^{K} \left| \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d{}^2} \right|^2 (\tau_{i,d})^4 \prod_{k=1}^{K} (\tau_{i,k})^2$$

$$+ \sum_{d=1}^{K}\sum_{r=1}^{K} o\left((\tau_{i,d})^2 (\tau_{i,r})^2 \prod_{k=1}^{K} (\tau_{i,k})^2 \right) + U_1 \, . \tag{B.5}$$

where $U_1 = 2\Re\left[g(\mathbf{c}_i) \prod_{d=1}^{K} \tau_{i,d} \int_{A_i} R d\mathbf{t} \right]$. For easy demonstration of the derivations, we assume that $p(\mathbf{t})$ is uniform in these intermediate steps since this will not affect the final expression of the asymptotic variance where $p(\mathbf{t})$ can be of any appropriate shape. With $v_i = \prod_{d=1}^{K} \tau_{i,d} / D$, the expected value of $|Z_i|^2$

$$E\left[|Z_i|^2 \right] = \prod_{d=1}^{K} \tau_{i,d} \int_{A_i} \left| \frac{g(\mathbf{t}) + g(2\mathbf{c}_i - \mathbf{t})}{2} \right|^2 d\mathbf{t}$$

$$= \left|g(\mathbf{c_i})\right|^2 \prod_{d=1}^{K} (\tau_{i,d})^2 + \frac{1}{288} \sum_{d=1}^{K} \sum_{\substack{r=1 \\ r \neq d}}^{K} \left| \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d \partial t_r} \right|^2 \left(\tau_{i,d}\right)^2 \left(\tau_{i,r}\right)^2 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2$$

$$+ \frac{1}{320} \sum_{d=1}^{K} \left| \frac{\partial^2 g(\mathbf{c}_i)}{\partial {t_d}^2} \right|^2 \left(\tau_{i,d}\right)^4 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2 + \sum_{d=1}^{K} \sum_{r=1}^{K} o\left(\left(\tau_{i,d}\right)^2 \left(\tau_{i,r}\right)^2 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2 \right) + U_2$$

(B.6)

where $U_2 = 2\Re\left[g\left(\mathbf{c}_i\right) \prod_{d=1}^{K} \tau_{i,d} \int_{A_i} R d\mathbf{t} \right]$. With $U_2 = U_1$ the variance of Z_i is

$$\mathrm{var}\left[Z_i\right] = E\left[\left|Z_i\right|^2 \right] - \left| E\left[Z_i\right] \right|^2$$

$$= \frac{1}{288} \sum_{d=1}^{K} \sum_{\substack{r=1 \\ r \neq d}}^{K} \left| \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d \partial t_r} \right|^2 \left(\tau_{i,d}\right)^2 \left(\tau_{i,r}\right)^2 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2$$

$$+ \frac{1}{720} \sum_{d=1}^{K} \left| \frac{\partial^2 g(\mathbf{c}_i)}{\partial {t_d}^2} \right|^2 \left(\tau_{i,d}\right)^4 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2 + \sum_{d=1}^{K} \sum_{r=1}^{K} o\left(\left(\tau_{i,d}\right)^2 \left(\tau_{i,r}\right)^2 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2 \right) . \quad \text{(B.7)}$$

Since Z_i are independent, the variance of the antithetical stratified estimator is

$$\mathrm{var}\left[\hat{X}_{wSANT}(\mathbf{f}) \right] = \sum_{i=1}^{N/2} \mathrm{var}\left[Z_i\right]$$

$$= \frac{1}{288} \sum_{i=1}^{N/2} \sum_{d=1}^{K} \sum_{\substack{r=1 \\ r \neq d}}^{K} \left| \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d \partial t_r} \right|^2 \left(\tau_{i,d}\right)^2 \left(\tau_{i,r}\right)^2 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2$$

$$+ \frac{1}{720} \sum_{i=1}^{N/2} \sum_{d=1}^{K} \left| \frac{\partial^2 g(\mathbf{c}_i)}{\partial {t_d}^2} \right|^2 \left(\tau_{i,d}\right)^4 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2 + \sum_{i=1}^{N/2} \sum_{d=1}^{K} \sum_{r=1}^{K} o\left(\left(\tau_{i,d}\right)^2 \left(\tau_{i,r}\right)^2 \prod_{k=1}^{K} \left(\tau_{i,k}\right)^2 \right).$$

(B.8)

With the observation domain divided equally along each dimension, for any continuous positive $p(\mathbf{t}) = \prod_{d=1}^{K} p_d(t_d)$ the following is valid

$$\int_{A_{i,d}} p_d(t_d)\, dt_d = \left[\frac{2}{N}\right]^{\frac{1}{K}} \tag{B.9}$$

where $\int_{A_{i,d}} dt_d$ is the integration over the subdomain A_i with respect to t_d . By (A.15) we have

$$\tau_{i,d} = \left[\frac{2}{N}\right]^{\frac{1}{K}} \times \frac{1}{p_d\left(c'_{i,d}\right)} \tag{B.10}$$

where $c'_{i,d} \in A_{i,d} \equiv \{t_d : t_d \in A_i\}$. This leads to

$$\mathrm{var}\left[\hat{X}_{wSANT}(\mathbf{f})\right] = \left(\frac{2}{N}\right)^{1+\frac{4}{K}} \left[\frac{1}{288} \sum_{i=1}^{N/2} \sum_{d=1}^{K} \sum_{\substack{r=1 \\ r\neq d}}^{K} \frac{\left|\frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d \partial t_r}\right|^2 \prod_{k=1}^{K} \tau_{i,k}}{{p_d}^2\left(c'_{i,d}\right) {p_r}^2\left(c'_{i,d}\right) \prod_{k=1}^{K} p_k\left(c'_{i,k}\right)} \right.$$

$$\left. + \frac{1}{720} \sum_{i=1}^{N/2} \sum_{d=1}^{K} \frac{\left|\frac{\partial^2 g(\mathbf{c}_i)}{\partial {t_d}^2}\right|^2 \prod_{k=1}^{K} \tau_{i,k}}{{p_d}^4\left(c'_{i,d}\right) \prod_{k=1}^{K} p_k\left(c'_{i,k}\right)} \right] + o\left(\frac{1}{N^{1+\frac{4}{K}}}\right). \tag{B.11}$$

Using Riemann integration

$$\lim_{N\to\infty} N^{1+\frac{4}{K}}\, \mathrm{var}\left[\hat{X}_{wSANT}(\mathbf{f})\right]$$

$$= 2^{1+\frac{4}{K}} \left[\frac{1}{288} \sum_{d=1}^{K} \sum_{\substack{r=1 \\ r \neq d}}^{K} \int_A \frac{\left| \frac{\partial^2 g(\mathbf{t})}{\partial t_d \partial t_r} \right|^2}{{p_d}^2(t_d)\, {p_r}^2(t_r) \prod_{k=1}^{K} p_k(t_k)} d\mathbf{t} + \frac{1}{720} \sum_{d=1}^{K} \int_A \frac{\left| \frac{\partial^2 g(\mathbf{t})}{\partial {t_d}^2} \right|^2}{{p_d}^4(t_d) \prod_{k=1}^{K} p_k(t_k)} d\mathbf{t} \right]. \quad \text{(B.12)}$$

Analogously to stratified estimation, Lyapunov condition has to be satisfied to prove the normality of the antithetical stratified estimate. To establish this condition, we redefine $g(\mathbf{t})$ as in (A.19) in order to use Cramér-Wold device for the proof of a multivariate distribution. For the antithetical stratified case the *i*-th element in the nominator of L_N , defined in (A.20), is as follows

$$Z_i - E[Z_i] = \frac{1}{2} \prod_{k=1}^{K} \tau_{i,k} \sum_{d=1}^{K} \frac{\partial^2 g(\mathbf{c}_i)}{\partial {t_d}^2} \left[(t_d - c_{i,d})^2 - \frac{1}{12} (\tau_{i,d})^2 \right]$$

$$+ \frac{1}{2} \prod_{k=1}^{K} \tau_{i,k} \sum_{d=1}^{K} \sum_{\substack{r=1 \\ r \neq d}}^{K} \frac{\partial^2 g(\mathbf{c}_i)}{\partial t_d \partial t_r} (t_d - c_{i,d})(t_r - c_{i,r}) + \prod_{k=1}^{K} \tau_{i,k} \sum_{d=1}^{K} o\left(\left| t_d - c_{i,d} \right|^2 \right)$$

$$+ \sum_{d=1}^{K} o\left((\tau_{i,d})^2 \prod_{k=1}^{K} \tau_{i,k} \right). \quad \text{(B.13)}$$

Following the same steps we took in the previous section, with further use of Hölder's inequality, we find that the nominator of L_N :

$$\sum_{n=1}^{N} E\left[\left| Z_i - E[Z_i] \right|^r \right] \leq \frac{C}{N^{r-1+2r/K}} \quad \text{(B.14)}$$

where *C* denotes a positive constant. The behaviour of the dominator of L_N for $N \to \infty$ follows from (B.12). Thus,

$$L_N \leq \frac{C}{N^{r/2-1} \left(N^{1+4/K} \sum_{i=1}^{N} \mathrm{var}[Z_i] \right)^{r/2}} \to 0 \ ,$$

as $N \to \infty$ for $r > 2$. The asymptotic covariance matrix is directly obtained by identifying the coefficients of λ_1^2, λ_2^2 and $\lambda_1\lambda_2$ on the right hand side of (B.12) with the redefined $g(\mathbf{t})$.

APPENDIX C

PROOFS FOR IQ STRATIFIED ESTIMATION

To achieve the described scenario of distributing the sampling points and study the characteristics of the proposed estimator, we define sequences of random variables a_m's which take either zero or one. We associate these random variables $\{a_m\}_{m=1}^{M}$ with the grid points $\{\mathbf{t}_m\}_{m=1}^{M}$ one by one, as shown in (5.3) (the IQ stratified estimates can be written in different forms. We chose the stated form for a clear demonstration of the scheme). The random variables within the *q*-th subgrid of the *i*-th subdomain have the same probability to be one/zero, given by (assuming/designing the number of the grid points of each subgrid to be identical inside the *i*-th subdomain)

$$P[a_m = 1] = \frac{2^K}{M_i},$$
$$P[a_m = 0] = \frac{M_i - 2^K}{M_i}\ . \tag{C.1}$$

These random variables are dependent, but they are independent of all the other random variables inside outside their subdomain. Recall that the dependence of the random variables only affects their conditional probability. And, according to the law of total probability, the probability of the variables $\{a_m\}$ to be one/zero is constant, defined in (C.1). Now, since the random variables $\{a_m\}$ within the *q*-th subgrid of the *i*-th subdomain are binary and all have the same probability of being one/zero, their expectations are identical and given by

$$E[a_m] = \frac{2^K}{M_i}\,. \tag{C.2}$$

Now, the expected value of the IQ stratified estimator (5.2) is

$$E\left[\hat{X}_{IQS}(\mathbf{f})\right] = \sum_{i=1}^{N/2^K} \frac{M_i D}{2^K M} \sum_{q=1}^{2^K} \sum_{\mathbf{t}_m \in A_i, G_q} E[a_m] x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m)$$

$$= \frac{D}{M} \sum_{m=1}^{M} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m)$$

$$= X_{wD}(\mathbf{f}). \tag{C.3}$$

Hence, (5.4) is proven. Next, we derive the expression of the variance of the developed estimator:

$$\text{var}\left[\hat{X}_{IQS}(\mathbf{f})\right] = \sum_{i=1}^{N/2^K} v_i \, \text{var}\left[\hat{X}_{S,i}(\mathbf{f})\right]. \tag{C.4}$$

We find the variance of the *i*-th component as follows:

$$\text{var}\left[\hat{X}_{S,i}(\mathbf{f})\right] = D^2 \sum_{q=1}^{2^K} \left(\sum_{\mathbf{t}_m \in A_i, G_q} a_m x_w{}^2(\mathbf{t}_m) \text{var}[a_m] + \sum_{\mathbf{t}_m \in A_i, G_q} \sum_{\substack{\mathbf{t}_l \in A_i, G_q \\ l \neq m}} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m) \; x_w(\mathbf{t}_l) \exp(j2\pi \mathbf{f}.\mathbf{t}_l) \text{cov}[a_m, a_l] \right) \tag{C.5}$$

The variances of the random variables $\{a_m\}$ are identical and given by

$$\text{var}[a_m] = \left(1 - \frac{2^K}{M_i}\right) \frac{2^K}{M_i} = \frac{2^K (M_i - 2^K)}{M_i{}^2}. \tag{C.6}$$

And,

$$\text{cov}[a_m, a_l] = -E[a_m] E[a_l]$$

$$= -\frac{2^{2K}}{M_i{}^2}. \tag{C.7}$$

Thus,

$$\text{var}\left[\hat{X}_{S,i}(\mathbf{f})\right] = \frac{2^{2K} D^2}{M_i{}^2} \sum_{q=1}^{2^K} \left(\frac{M_i - 2^K}{2^K} \sum_{\mathbf{t}_m \in A_i, G_q} x_w{}^2(\mathbf{t}_m) - \sum_{\mathbf{t}_m \in A_i, G_q} \sum_{\substack{\mathbf{t}_l \in A_i, G_q \\ l \neq m}} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m) x_w(\mathbf{t}_l) \exp(j2\pi \mathbf{f}.\mathbf{t}_l) \right). \tag{C.8}$$

By completing the squares of the last term, we obtain

$$\operatorname{var}\left[\hat{X}_{S,i}(\mathbf{f})\right] = \frac{2^{2K} D^2}{M_i^{\,2}} \sum_{q=1}^{2^K} \left(\frac{M_i}{2^K} \sum_{\mathbf{t}_m \in A_i, G_q} x_w^{\,2}(\mathbf{t}_m) - \left| \sum_{\mathbf{t}_m \in A_i, G_q} x_w^{\,2}(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m) \right|^2 \right). \quad \text{(C.9)}$$

Using (C.4) we obtain the variance of the IQ stratified estimates.

APPENDIX D

PROOFS FOR ANTITHETICAL STRATIFIED ESTIMATION

In this scheme we also design the binary random variables $\{a_m\}$ within the *q*-th subgrid of the *i*-th subdomain with probabilities (assuming/designing the number of the grid points of each subgrid to be the same within the *i*-th subdomain)

$$P[a_m = 1] = \frac{2^K}{M_i},$$

$$P[a_m = 0] = \frac{M_i - 2^K}{M_i}. \tag{D.1}$$

The random variables within the *q*-th subgrid of the *i*-th subdomain are dependent, but they are independent of the variables elsewhere. Now, let

$$g(\mathbf{t}) := x_w(\mathbf{t}) \exp(-j2\pi \mathbf{f}.\mathbf{t}). \tag{D.2}$$

The expected value of $\hat{X}_{IQA}(\mathbf{f})$ can be determined in the following way:

$$\begin{aligned} E\left[\hat{X}_{A,i}(\mathbf{f})\right] &= \sum_{i=1}^{N/2^{K+1}} \frac{DM_i}{2^{K+1}M} \sum_{q=1}^{2^K} \sum_{\mathbf{t}_m \in A_i, G_q} E[a_m]\left[g(\mathbf{t}_m) + g(2\mathbf{c}_{i,q} - \mathbf{t}_m)\right] \\ &= \frac{D}{M} \sum_{i=1}^{N/2^{K+1}} \sum_{\mathbf{t}_m \in A_i} g(\mathbf{t}_m) \\ &= \frac{D}{M} \sum_{i=1}^{N/2^{K+1}} \sum_{\mathbf{t}_m \in A_i} x_w(\mathbf{t}_m) \exp(-j2\pi \mathbf{f}.\mathbf{t}_m) = X_{wD}(\mathbf{f}). \end{aligned} \tag{D.3}$$

Now, we find the variance of the estimator:

$$\operatorname{var}\left[\hat{X}_{IQA}(\mathbf{f})\right] = \sum_{i=1}^{N/2^{K+1}} v_i^2 \operatorname{var}\left[\hat{X}_{A,i}(\mathbf{f})\right]. \tag{D.4}$$

The variance of each component can be obtained as follows:

$$\operatorname{var}\left[\hat{X}_{A,i}(\mathbf{f})\right] = D^2 \sum_{q=1}^{2^K} \left(\sum_{\mathbf{t}_m \in A_i, G_q} \left| g(\mathbf{t}_m) + g(2\mathbf{c}_{i,q} - \mathbf{t}_m) \right|^2 \operatorname{var}[a_m] + \sum_{\mathbf{t}_m \in A_i, G_q} \sum_{\substack{\mathbf{t}_l \in A_i, G_q \\ l \neq m}} \left[g(\mathbf{t}_m) + g(2\mathbf{c}_{i,q} - \mathbf{t}_m) \right] \left[g(\mathbf{t}_l) + g(2\mathbf{c}_{i,q} - \mathbf{t}_l) \right]^* \operatorname{cov}[a_m, a_l] \right). \tag{D.5}$$

We note that the variance is given by

$$\operatorname{var}[a_m] = \frac{4(M_n - 4)}{{M_n}^2} \tag{D.6}$$

and the covariance

$$\operatorname{cov}[a_m, a_l] = -\frac{16}{{M_n}^2}. \tag{D.7}$$

Accordingly,

$$\begin{aligned} &\operatorname{var}\left[\hat{X}_{A,i}(\mathbf{f})\right] \\ &= \frac{2^{2K} D^2}{{M_i}^2} \sum_{q=1}^{2^K} \left(\frac{M_i - 2^K}{2^K} \sum_{\mathbf{t}_m \in A_i, G_q} \left(\left| g(\mathbf{t}_m) \right|^2 + \left| g(2\mathbf{c}_{i,q} - \mathbf{t}_m) \right|^2 + 2\Re\left[g(\mathbf{t}_m) g^*(2\mathbf{c}_{i,q} - \mathbf{t}_m) \right] \right) \right. \\ &\left. - \sum_{\mathbf{t}_m \in A_i, G_q} \sum_{\substack{\mathbf{t}_l \in A_i, G_q \\ l \neq m}} \left[g(\mathbf{t}_m) g^*(\mathbf{t}_l) + g(\mathbf{t}_m) g^*(2\mathbf{c}_{i,q} - \mathbf{t}_l) + g(2\mathbf{c}_{i,q} - \mathbf{t}_m) g^*(\mathbf{t}_l) + g(2\mathbf{c}_{i,q} - \mathbf{t}_m) g^*(2\mathbf{c}_{i,q} - \mathbf{t}_l) \right] \right) \end{aligned} \tag{D.8}$$

By completing the squares and substitute $g(\mathbf{t})$ by its definition, we obtain

$$\operatorname{var}\left[\hat{X}_{A,i}(\mathbf{f})\right] = \frac{2^{2K+2} D^2}{{M_i}^2} \sum_{q=1}^{2^K} \left(\frac{M_i}{2^{K+1}} \sum_{\mathbf{t}_m \in A_i, G_q} {x_w}^2(\mathbf{t}_m) + x_w(\mathbf{t}_m) x_w(2\mathbf{c}_{i,k} - \mathbf{t}_m) \cos\left(2\pi f (2\mathbf{c}_{i,k} - 2\mathbf{t}_m)\right) - \left| \sum_{\mathbf{t}_m \in A_i, G_q} x_w(\mathbf{t}_m) \exp(-j 2\pi \mathbf{f}.\mathbf{t}_m) \right|^2 \right). \tag{D.9}$$

REFERENCES

[1] James Keeler, *Understanding NMR spectroscopy,* John Wiley & Sons, Chichester, 2005.

[2] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, and I. Zhukov, "Random sampling of evolution time space and Fourier transform processing*," Journal of Biomolecular NMR,* vol. 36, pp. 157-168, 2006.

[3] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, and I. Zhukov, "Lineshapes and artifacts in multidimensional Fourier transform of arbitrary sampled NMR data sets," *Journal of Magnetic Resonance*, vol. 188, pp. 344-356, 2007.

[4] K. Kazimierczuk, A. Zawadzka, and W. Koźmiński, "Optimization of random time domain sampling in multidimensional NMR," *Journal of Magnetic Resonance,* vol. 192, 2008.

[5] K. Kazimierczuk, A. Zawadzka, and W. Kozminski, "Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling," *Journal of Magnetic Resonance*, vol. 197, pp. 219-228, 2009.

[6] K. Kazimierczuk, J. Stanek, and W. Koźmiński, "Random sampling in multidimensional NMR spectroscopy," *Journal of Magnetic Resonance,* vol 57, pp 420-434, 2011.

[7] J. Jeneer, Lecture presented at Ampere International Summer School, Basko Poljie, Yugoslavia.

[8] A. Tarczynski and N. Allay, "Spectral analysis of randomly sampled signals: suppression of aliasing and sampler jitter," *IEEE Trans. on Signal Processing*, vol. SP-52, pp. 3324-3334, 2004.

[9] A. Tarczynski and D. Qu, "Optimal random sampling for spectrum estimation in DASP applications," *International Journal of Applied Mathematics and Computer Science*, vol.15, pp. 463-469, 2005.

[10] E. Masry, "Random sampling of deterministic signals: statistical analysis of Fourier transforms estimates," *IEEE Trans. on Signal Processing*, vol. 54, pp. 1750-1761, 2006.

[11] E. Masry, "Random sampling estimates of Fourier transforms: antithetical stratified Monte Carlo," *IEEE Trans. Signal Processing*, vol. 57, pp. 149-204, 2009.

[12] I. Bilinskis and M. Mikelsons, *Randomised Signal Processing*. London: Prentice Hall, 1992.

[13] I. Bilinskis, *Digital Alias-free Signal Processing*. New York: John Wiley and Sons, 2007.

[14] P Stoica and R Moses, *Spectral analysis of signals*, Prentice Hall, NJ, 2005.

[15] P. Stoica, J. Li, and H. He, "Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram," *IEEE Trans. Signal Processing*, vol.57, pp.843-858, 2009.

[16] R.J. Martin, Irregularly sampled signals: Theories and techniques for analysis, Ph.D. thesis, University of London, 1998.

[17] R.J. Martin, "Autoregression and irregular sampling: Spectral estimation," *Signal Processing*, vol. 77, pp. 139–157, 1999.

[18] P. J. S. G. Ferreira, "Incomplete sampling and the recovery of missing samples from oversampled band-limited signals," IEEE Trans. Signal Processing, vol. 40, pp. 225–27, Jan. 1992.

[19] Y. Wang, J. Li and P. Stoica, *Spectral Analysis of Signals—The Missing Data Case*, Morgan & Claypool, San Rafael, CA, 2005.

[20] P.M.T. Broersen, S. de Waele and R. Bos, "Autoregressive spectral analysis when observations are missing," *Automatica*, vol. 40, pp. 1495–1504, 2004.

[21] Y. Eldar and T. Michaeli, "Beyond Bandlimited Sampling," *IEEE Signal Processing Magazine*, vol.26, pp.48-68, 2009.

[22] P. P. Vaidyanathan, "Generalizations of the sampling theorem: seven decades after Nyquist," *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, vol.48, pp.1094-1109, 2002.

[23] J. Benedetto and P. J. S. G. Ferreira, *Modern Sampling Theory: Mathematics and Applications*, Birkhauser, 2001.

[24] M. Unser, "Sampling-50 years after Shannon," *Proceedings of the IEEE*, vol.88, pp.569-587, 2000.

[25] J. R. Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundation*, London, Oxford University Press, 1996.

[26] A. I. Zayed, *Advanced in Shannon's Sampling Theory*, Boca Rotan, CRC Press, 1993.

[27] R. J. Marks, *Advanced Topics in Shannon Sampling and Interpolation Theory*, New York, Springer-Verlag, 1993.

[28] R. J. Marks, *Introduction to Shannon Sampling and Interpolation Theory*, New York, Springer-Verlag, 1991.

[29] A. J. Jerri, "Shannon sampling theorem - its various extensions and applications: A tutorial review," *Proceedings of the IEEE*, vol.65, pp.1565-1596, 1977.

[30] F. Marvasti, *Nonuniform Sampling Theory and Practice*, New York, Kluwer Academic, 2001.

[31] A. Aldroubi and K. Gronhenig, "Nonuniform sampling and reconstruction in shift invariant spaces," *SIAM Reviews*, vol.43, pp.585-620, 2001.

[32] F. Marvasti, *A Unified Approach to Zero-crossing and Nonuniform Sampling*, Illinois, Oak Park, 1987.

[33] D. P. Petersen and D. Middleton, "Sampling and Reconstruction of wave-number-limited functions in N-dimensional euclidean spaces", *Information and Control*, vol. 5, pp. 279–323, 1962.

[34] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol.66, pp.51-83, 1978.

[35] A. Antoniou, *Digital Filters: Analysis, Design and Applications*, New York, McGraw-Hill, 1993.

[36] J. D. Scargle, "Studies in astronomical time series analysis. III: Fourier transforms, autocorrelation functions and crosscorrelation functions of unevenly spaced data," *The Astronomical Journal*, vol.343, pp.874-887, 1989.

[37] J. D. Scargle, "Studies in astronomical time series analysis. II: statistical aspects of spectral analysis of unevenly spaced data," *The Astronomical Journal*, vol.263, pp.835-853, 1982.

[38] N. R. Lomb, "Least-squares frequency analysis of un-equally spaced data," *Astrophysics and Space Science*, vol.39, pp.447-462, 1967.

[39] N.R. Butt and A. Jakobsson, "High-resolution estimation of multidimensional spectra from unevenly sampled data," *Proc. Conf. on Digital Signal Processing (DSP),* pp.1-4, 6-8 July 2011.

[40] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[41] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[42] S. Bourguignon, H. Carfantan, and J. Idier. “A sparsity-based method for the estimation of spectral lines from irregularly sampled data”, IEEE Journal of Selected Topics in Signal Processing, vol. 1, pp. 575–585, 2007.

[43] J. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss, "Random Sampling for Analog-to-information Conversion of Wideband Signals," *proceedings of IEEE Dallas Circuits and Systems Workshop*, 2007, pp.119-122.

[44] M. Mishali and Y.C. Eldar, "Blind multiband signal reconstruction: compressed sensing for analog signals," *IEEE Trans. on Signal Processing*, vol. 57, no.3, pp. 993-1009, 2009.

[45] K. F. Cheung and R. J. Marks II, "Image sampling below the Nyquist density without aliasing," J. Opt. Soc. Amer., vol. 7, pp. 92–105, Jan. 1990.

[46] K. F. Cheung, "A multidimensional extension of Papoulis' generalized sampling expansion with the application in minimum density sampling," in R. J. Marks II, Advanced Topics in Shannon Sampling and Interpolation Theory. Berlin, Germany: Springer-Verlag, 1992.

[47] R. G. Vaughan, N. L. Scott, and D. R. White, "The theory of bandpass sampling," *IEEE Trans. on Signal Processing*, vol.39, pp.1973-1984, 1991.

[48] Y. P. Lin and P. P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Trans. on Circuits Systems II*, vol. 45, pp. 340-351, 1998.

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

[50] A. Kohlenberg, "Exact interpolation of band-limited functions," *J. Appl. Phys.*, vol. 24, pp. 1432–1435, Dec. 1953.

[51] A. Papoulis "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 652–654, Nov. 1977.

[52] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," *Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1688–1691, May 1996.

[53] R. Venkataramani and Y. Bresler, “Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals,” *IEEE Trans. on Information Theory*, vol. 46, pp. 2173-2183, 2000.

[54] Y. Bresler and P. Feng, “Spectrum-blind minimum-rate sampling and reconstruction of 2-D multiband signals,” *Proc. IEEE Inter. Conf. Image Process.*, Sep. 1996, vol. 1, pp. 701–704.

[55] R. Venkataramani and Y. Bresler, “Further results on spectrum blind sampling of 2D signals,” *Proc. IEEE Int. Conf. Image Process.*, Oct. 4–7, 1998, vol. 2, pp. 752–756.

[56] M. Mishali and Y.C. Eldar, “From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals,” Selected Topics in Signal Processing, IEEE Journal of, vol. 4, no. 2, pp. 375–391, Apr 2010.

[57] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1417-1428, 2002.

[58] H. S. Shapiro and R. A. Silverman, "Alias-free sampling of random noise," *SIAM Journal of Applied Mathematics*, vol. 8, pp.225-236, 1960.

[59] F. Beutler, "Alias-free randomly timed sampling of stochastic processes," *IEEE Trans. on Information Theory*, vol. IT-16, pp. 147-152, 1970.

[60] E. Masry, "Random sampling and reconstruction of spectra," *Information and Control* vol.19, pp. 275-288, 1971.

[61] E. Masry, "Alias-free sampling: an alternative conceptualization and its applications," *IEEE Trans. on Information Theory*, vol. 24, pp. 317-324, 1978.

[62] E. Masry, "Non-parametric Covariance Estimation from Irregular-spaced Data," *Advances in Applied Probability*, vol.15, pp.113-132, 1983.

[63] J. J. Wojtiuk, Randomised Sampling for Radio Design, PhD Thesis, University of South Australia, 2000.

[64] N. Allay, Applications of Nonuniform Sampling Techniques in Digital Signal Processing and Communication Systems, PhD Thesis, University of Westminster, London, 2006.

[65] D. D. Qu, Nonuniform Sampling Algorithms for DASP-based Instrumentation , PhD Thesis, University of Westminster, London, 2008.

[66] B. I. Ahmad and A. Tarczynski, "Reliable wideband multichannel spectrum sensing using randomized sampling schemes," *Signal Processing*, Elsevier, vol. 90, pp. 2232-2242, 2010.

[67] B. I. Ahmad and A. Tarczynski, "A SARS method for reliable spectrum sensing in multiband communication systems," *IEEE Trans. on Signal Processing*, vol.59, no.12, pp.6008-6020, 2011.

[68] B. I. Ahmad and A. Tarczynski, "Spectral analysis of stratified sampling: a means to perform efficient multiband spectrum sensing," *IEEE Trans. on Wireless Communications*, vol.11, no.1, pp.178-187, 2012.

[69] D. H. Roberts, J. Lehar, and J. W. Dreher, "Time series analysis with CLEAN. I. Derivation of a spectrum," *Astronom. J.*, vol. 93, no. 4, pp. 968–989, Apr. 1987.

[70] G. Foster, "The cleanest Fourier spectrum," *Astronom. J.*, vol. 109, no. 4, pp. 1889–1902, Apr. 1995.

[71] I. Mednieks, "*Methods for spectral analysis of nonuniformly sampled signals*", *Proc. SAMPTA'99*, Loen, August 1999, pp. 190-193.

[72] K. D. Tocher, *The Art of Simulation*, D. Van Nostrand, Princeton, New Jersey, 1963.

[73] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, John Wiley & Sons, New York, 1986.

[74] H. Urkowitz, *Signal Theory and Random Processes*, MA, USA, Artech House, 1983.

[75] A.W. Van Der Vaart, *Asymptotic Statistics*, Cambridge University Press, New York, 1998.

[76] F. Papenfuss, Digital signal processing of nonuniform sampled signals, Ph.D. Thesis, University of Rostock, Germany, 2007.

[77] B. I. Ahmad, A. Tarczynski and M. Al-Ani, "A SARS multiband spectrum sensing method in wideband communication systems using RSG", *Proceedings of the 19th European Signal Processing Conference (EUSIPCO'11*), Barcelona, Aug 2011, pp. 1219-1223.

[78] B. I. Ahmad, A. Tarczynski and M. Al-Ani, "A DASP multiband spectrum sensing method based on total random sampling on grid without replacement", *Proceedings of the 9th International Conference on Sampling Theory and Its Applications (SAMPTA '11)*, Singapore, May 2011.

[79] S.L. Hahn, "Multidimensional complex signals with single-orthant spectra," *Proceedings of the IEEE* , vol.80, no.8, pp.1287-1300, Aug 1992.

[80] T. Bulow and G. Sommer, "Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case," *IEEE Trans. on Signal Processing*, vol.49, no.11, pp.2844-2852, Nov 2001.